I.J. Information Engineering and Electronic Business, **2025, 4, 82-144**
Published Online on August 8, 2025 by MECS Press (http://www.mecs-press.org/)
DOI: 10.5815/ijieeb.2025.04.07

Comprehensive Intellectual Analysis of Statistical Data on Leading Energy Companies' Actions

Viktoriia Bulatova
Department of Information Systems and Networks, Institute of Computer Sciences and Information Technologies, Lviv Polytechnic National University, Lviv, 79013, Ukraine
E-mail: viktoriia.bulatova.sa.2022@lpnu.ua
ORCID iD: https://orcid.org/0009-0008-2692-3571

Sofiia Popp
Department of Information Systems and Networks, Institute of Computer Sciences and Information Technologies, Lviv Polytechnic National University, Lviv, 79013, Ukraine
E-mail: sofiia.popp.sa.2022@lpnu.ua
ORCID iD: https://orcid.org/0009-0001-2453-6542

Victoria Vysotska
Department of Information Systems and Networks, Institute of Computer Sciences and Information Technologies, Lviv Polytechnic National University, Lviv, 79013, Ukraine
E-mail: Victoria.A.Vysotska@lpnu.ua
ORCID iD: https://orcid.org/0000-0001-6417-3689

Yuriy Ushenko*
Department of Physics, Shaoxing University, Shaoxing, Zhejiang Province 312000, China
Department of Computer Science, Educational and Research Institute of Physical, Technical and Computer Sciences, Yuriy Fedkovych Chernivtsi National University, 58012, Ukraine
E-mail: y.ushenko@chnu.edu.ua
ORCID iD: https://orcid.org/0009-0002-7620-5355
*Corresponding Author

Zhengbing Hu
School of Computer Science, Hubei University of Technology, Wuhan, China
E-mail: drzbhu@gmail.com
ORCID iD: https://orcid.org/0000-0002-6140-3351

Dmytro Uhryn
Department of Computer Science, Educational and Research Institute of Physical, Technical and Computer Sciences, Yuriy Fedkovych Chernivtsi National University, 58012, Ukraine
E-mail: d.ugryn@chnu.edu.ua
ORCID iD: https://orcid.org/ 0000-0003-4858-4511

Received: 22 February, 2025; Revised: 18 April, 2025; Accepted: 24 May, 2025; Published: 08 August, 2025

Abstract: The paper conducted a comprehensive analysis of the time series of stock prices of three leading energy companies – Shell, BP and ExxonMobil – for the period from January 2021 to January 2025. At the initial stage, data quality was checked: dates were set as indices, the absence of duplicates and missing values was confirmed, and descriptive statistics (mean, variance, skewness and kurtosis) were calculated. Next, the trends of adjusted closing prices (AdjClose) were analysed using moving averages (SMA14, SMA50), exponential smoothing, moving volatility (30-day standard deviation) and cumulative returns. It was found that ehe price dynamics growth has accelerated since 2022 against the background of the energy crisis caused by the war in Ukraine: ExxonMobil's cumulative return reached ≈250% by mid-2022 and ≈350% at the beginning of 2025, Shell and BP, respectively ≈220% and ≈200% by 2024. Correlation analysis showed that BP and Shell have the most significant interdependence ($r = 0.87$, $R^2 = 0.75$). The autocorrelation method established high non-stationarity of the time series (ACF about one at low lags). K-Means clustering ($k = 2$) allowed us to distinguish periods of active growth and relative price consolidation, although the feature selection behind this clustering requires further clarification. The initially reported financial metrics (Sharpe, Sortino, and Calmar ratios) were significantly overstated due to unit errors, specifically, using percentage values as absolute figures. After applying appropriate annualization and decimal scaling performance indicators were obtained for ExxonMobil – CAGR = 36.84%, Sharpe ≈ 1.24, Sortino ≈ 1.9–2.5, Max Drawdown = 20.51%, Calmar ≈ 1.80; Shell: CAGR = 21.29%, Sharpe ≈ 0.76, Sortino ≈ 1.2–1.5, Max Drawdown = 25.04%, Calmar ≈ 0.85; BP: CAGR = 14.54%, Sharpe ≈ 0.53, Sortino ≈ 0.9–1.2, Max Drawdown = 26.23%, Calmar ≈ 0.55. The study confirms that ExxonMobil showed the most stable and substantial growth during the examined period, while BP exhibited the highest volatility. Shell demonstrated an intermediate performance level. The close correlation between Shell and BP is attributed to the similarity in their geographical market activity and stock behaviour. The choice of these methods of analysis is due to the desire to assess the behaviour of stocks during the period of increased market volatility caused by the energy crisis, geopolitical risks and changes in investor priorities. Technical analysis allows you to identify short- and medium-term patterns, clustering allows you to automatically separate market phases without the need for subjective hypotheses, and statistical metrics will enable you to compare the performance of assets within the industry. This research contributes to the broader field of financial analysis by demonstrating how machine learning and technical analytics tools can be applied to assess the resilience and relationships of assets during periods of market turmoil. The results can be helpful for institutional investors, financial analysts, and portfolio managers looking to adapt strategies to dynamic energy market conditions.

Index Terms: Energy Stocks; Time Series; Cumulative Return; Rolling Volatility; Correlation Analysis; K-Means Method; Financial Metrics (CAGR, Sharpe, Sortino, Max Drawdown, Calmar)

1. Introduction

In today's world, geopolitical instability is increasingly acting as a catalyst for profound changes in financial markets, especially in the energy sector. Energy companies, as critical suppliers of resources, react to global events in a sensitive and often unpredictable way, which attracts the attention of investors, analysts and researchers. Recent events that have had a significant impact on markets include the COVID-19 pandemic, the energy crisis caused by the war in Ukraine, and sharp fluctuations in oil and gas prices. In such an environment, the issue of assessing the efficiency of leading energy companies becomes especially relevant. This work is aimed at a comprehensive quantitative analysis of the dynamics of shares of the three leading players in the global energy market – ExxonMobil, Shell and BP – in the period 2021–2025. Particular attention is paid to identifying the relationship between the market behavior of these companies and external shocks, such as geopolitical crises and fluctuations in commodity prices.

The purpose of this study is to comprehensively characterise and compare the financial stability, market behaviour and risk profile of shares of three leading multinational energy companies (Shell, BP, ExxonMobil) during the energy crisis of 2021–2025 by combining classical and modern methods of time series analysis. In particular, the study aims to:

1. Identify periods of increased volatility that correlate with geopolitical events.
2. Compare profitability and risk using financial metrics (CAGR, Sharpe, Sortino, Calmar, Max Drawdown).
3. To investigate intra-market clusters (market modes) using machine learning.
4. Evaluate autocorrelation and correlation structures to identify dependencies between companies.
5. Provide an interpreted toolkit for institutional investors to diagnose the investment attractiveness of companies in times of crisis.

The main objectives of the study are:

- Evaluate the performance of the shares of these companies using standard financial metrics (CAGR, Sharpe, Sortino, Drawdown, etc.);
- Analyze the impact of external factors (including war, energy policy, ESG changes) on the market behavior of companies;
- Establish causal relationships between global events and financial results of companies;
- Predict future dynamics based on time series.

Through comparative analysis, the study aims to determine which of the companies demonstrated the most excellent adaptability and investment attractiveness during the crisis period, as well as to give an idea of their potential in the near future.

Research question. How did the geopolitical crisis of 2022, in particular Russia's invasion of Ukraine, affect the efficiency, volatility and market behaviour of the shares of leading energy companies – Shell, BP and ExxonMobil – compared to each other in the period 2021-2025?

Research hypothesis. ExxonMobil, as a company less dependent on the European energy market, will demonstrate higher investment efficiency (in terms of CAGR, Sharpe, Sortino, and Max Drawdown) and less volatility compared to Shell and BP during the period of geopolitical instability.

New research contribution. Unlike previous works, this study offers a generalised comparison of the three leading energy companies in the context of the current energy crisis. It integrates different approaches – descriptive statistics, trend visualisation, clustering, correlation analysis, risk/return financial metrics – into a single methodological framework. For the first time, a generalised multi-factor comparison of the dynamics of Shell, BP and ExxonMobil in the context of the war in Ukraine and the transformation of the global energy market is presented.

Scientific novelty and main contribution of the article. Despite a significant amount of research on the financial dynamics of energy companies, most of them focus on:

- analysis of one company or two,
- using only one type of models (ARIMA, GARCH, ML, etc.),
- exclusion of geopolitical context or current events (such as the energy crisis of 2022).

The novelty of this work lies in a comprehensive intermethodological approach, which combines:

- Traditional statistical methods (averages, variance, moving averages),
- Time series methods (SMA, exponential smoothing, ACF),
- Financial performance analysis tools (CAGR, Sharpe, Sortino, Max Drawdown, Calmar),
- Advanced anti-aliasing (Pollard and median methods),
- Machine learning (K-Means clustering),
- Analysis of volatility and market activity,
- A unique comparison of three leading multinational energy companies (Shell, BP, ExxonMobil) against the
- backdrop of the energy crisis caused by the war in Ukraine (2022–2024).

Such integration of methods within the framework of one study allows for a deeper characterisation of the behaviour of energy stocks, identification of critical periods of the market and assessment of the resilience of companies to global shocks. The main contribution of the article is as follows:

1. For the first time, a three-company comparative study of Shell, BP and ExxonMobil was conducted in the context of 2021–2025, taking into account military and energy events.
2. An integrated approach to the analysis of the financial behaviour of companies is applied, which includes not only classical financial metrics, but also methods of autocorrelation, clustering and smoothing.
3. The investment attractiveness of companies in long-term and short-term terms is determined, and ExxonMobil's excellent resilience to market shocks and BP's volatility are indicated.
4. A high correlation between BP and Shell is shown as a result of similar geographical and political influences, in contrast to a smaller correlation with the American ExxonMobil.
5. Macroeconomic and geopolitical factors are integrated into the technical analysis of financial markets, demonstrating how the war in Ukraine has affected the dynamics of energy stocks.

Motivation for the study. Recent years have become critical for the global energy market: the COVID-19 pandemic, the explosive growth in demand in 2021 and Russia's full-scale invasion of Ukraine in 2022 caused unprecedented price fluctuations and destabilization of supplies. In these conditions, special attention needs to be paid to the analysis of the market behaviour of the largest multinational energy companies, which are systemically essential players in the global economy. Despite a significant number of works devoted to the modelling of the stock market, existing studies have several significant limitations:

1. Single-model approach. Most works use only individual tools – for example, ARIMA/GARCH for volatility, or Random Forest for price direction prediction – which does not allow for a comprehensive coverage of the entire spectrum of market behaviour.
2. Limited sample of companies. Most studies focus on separate issuers or two companies, which makes systematic comparisons impossible, especially between American and European players.
3. Exclusion of real geopolitical shocks. Many models do not take into account real events, such as the energy crisis, sanctions, and the exit of companies from the Russian market. It reduces the relevance of the results.
4. Lack of integration of risk and profitability indicators. Technical indicators are rarely combined with financial metrics such as CAGR, Sharpe, and Calmar, which are necessary for a practical investor.

All this caused the need to create a generalized approach that would at the same time:

- analyzed time series from the standpoint of technical analysis and financial stability,
- took into account macroeconomic and geopolitical factors,
- compared the dynamics of the shares of the three leaders of the energy sector during a deep crisis,
- provided visualization, trend smoothing, and segmentation of growth/consolidation periods.

It was these challenges that became the starting point for conducting this study and forming a comprehensive approach to benchmarking Shell, BP and ExxonMobil based on multi-level processing of financial time series.

The structure of the paper is organised as follows. This section introduces the context of the study, focusing on current challenges for the energy sector, including geopolitical events. It provides an overview of current areas of research. It formulates the purpose of the work – to create an integrated methodology for assessing the dynamics of shares of three key energy companies in the period 2021–2025.

Section 2 reviews recent literature related to energy stock analysis and outlines key methodologies and findings from prior studies. This section summarizes the state of scientific research in the field of analysis of shares of energy companies. The methods used (ARIMA, GARCH, PCA, ML, etc.) and their limitations are described. A comparative analysis of approaches to predicting volatility, identifying anomalies and the impact of global events has been carried out, which allows identifying existing gaps and substantiating the relevance of the integrated approach proposed in this work.

Section 3 describes the dataset and research methods used, including data preprocessing, statistical measures, smoothing techniques, and clustering algorithms. This section describes the data sets used, the structure of indicators (Open, Close, Adj Close, Volume) and approaches to pre-processing. The selected analysis methods are further explained: statistical characteristics, trend smoothing, volatility calculation, correlation analysis, autocorrelation, clustering, and financial metrics. The purpose of the section is to provide the reader with a complete methodological basis of the study.

Section 4 presents the experimental results, including descriptive analytics, visual trend analysis, volatility estimates, financial performance metrics, correlation fields, and autocorrelation diagnostics. The main analytical section, in which all the selected methods are implemented step by step. It contains:

- statistical description of data;
- graphical analysis of price dynamics;
- assessment of market activity;
- calculation of cumulative yield;
- seasonal analysis of average monthly profits;
- volatility analysis;
- assessment of key financial indicators;
- using anti-aliasing techniques to highlight trends;
- construction of correlation matrices;
- application of the autocorrelation function;
- segmentation of time series using K-Means.

Each unit illustrates how the chosen methodology contributes to a deeper understanding of the market behavior of companies.

Finally, Section 5 summarises the main conclusions, discusses practical implications for investors, and outlines potential directions for further research. The final section summarizes the results obtained, compares the effectiveness of the three companies against the background of crisis events. An interpretation of indicators of financial stability, risk and profitability is provided. The advantage of the integrated approach is emphasized and its practical value for investors and researchers is indicated. Directions for further research in the field of financial analysis of energy markets are proposed.

2. Related Works

The energy sector plays a key role in the modern global economy, directly affecting the stability of industrial production, transport and consumer services [1]. The growth of demand for energy and fluctuations in their prices have wide-ranging economic and social consequences, which makes the study of the financial dynamics of leading energy companies particularly relevant [2]. In light of recent geopolitical upheavals, such as the COVID-19 pandemic and Russia's full-scale invasion of Ukraine in early 2022, energy markets have experienced unprecedented shocks: the fall in demand in 2020 and the rapid increase in prices in 2021–2022 provoked a new wave of volatility.

Currently, there are several areas of research, in particular [3-7]:

- Time-Series Modelling of Energy Stocks;
- Event-Driven and Sentiment Analysis;

- Correlation and Co-movement Studies;
- Clustering and Regime Detection, for example:
- Clustering volatility in the stock market of a particular energy company;
- Coherence between oil prices and energy company stocks;
- Identifying anomalies in the regional energy market;
- Analysing the impact of geopolitical events on energy company stocks;
- Machine Learning and Feature Extraction, for example:
- State models for oil price forecasting;
- Application of ARIMA-GARCH models to CO_2 emissions analysis;
- Forecasting company stock prices using machine learning;
- Forecasting significant stock price changes using neural networks;
- Improving stock price forecasting using adversarial learning;
- Forecasting stock price direction using random forest.

Ly, Sarwat et al. [1] investigated the dynamics of carbon dioxide emissions in Argentina using a combination of ARIMA-GARCH models with copulas. This approach allowed us to take into account nonlinear dependencies and volatility in time series, which is relevant for energy market analysis.

Stasiak et al. [3] applied state models in binary-time representation to oil price forecasting. They emphasized the importance of temporal representations and smoothing for accurate modelling of market trends.

Xiao and Sioofy Khoojine [5] developed a method to detect anomalies in the Chinese energy market during financial turmoil using mutual information and Brent oil price movements. They created dynamic networks based on minimum spanning trees to monitor market volatility.

Zuo et al. applied ARIMA and GARCH models to the daily returns of ExxonMobil and Chevron from 1985–2024. They found that including exogenous oil price variables significantly improved volatility forecasts [8]. Rahman et al. [9] investigated the effectiveness of machine learning models in predicting Shell stock price movements during the energy crisis. They found that such models can be valuable tools for investors during volatile periods. Rahman extended this approach by financial analysis, market behaviour, and the impact of global events on the energy sector. The Random Forest model has the highest accuracy at 0.52, followed by Logistic Regression with an accuracy of 0.51, and then the Support Vector Classifier with an accuracy of 0.50 [9].

Benali et al. employed principal component analysis (PCA) to reduce the dimensionality of a set comprising 50 technical indicators prior to applying four machine learning classifiers – logistic regression, decision tree, extra tree, and random forest – for forecasting Shell's stock price. The performance of the regression models was assessed using a variety of evaluation metrics, including mean squared error (MSE), root mean squared error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE), and the coefficient of determination (R^2). In contrast, the effectiveness of the classification algorithms was evaluated based on precision, recall, and the F1 score [10].

Additionally, Agbede advanced the existing body of research in energy economics by proposing a comprehensive financial modelling framework that captures the complexity of market dynamics. The integration of geopolitical and regulatory factors into energy market analysis offers a critical decision-making tool for stakeholders, facilitating the development of adaptive strategies in the context of a rapidly transforming global energy environment [11].

Damås et al. investigated four neural network architectures, each utilizing different gated recurrent neural network (RNN) cells, and found that deep recurrent neural networks (DRNNs) outperformed vector autoregressive (VAR) models in terms of prediction error and directional forecasting accuracy. Among the architectures tested, models employing Gated Recurrent Units (GRUs) achieved the highest performance [12].

In a complementary study, Ghosh et al. analysed the daily closing prices of the BSE Energy Index, crude oil, the DJIA Index, natural gas, and the NIFTY Index – representing natural resources, as well as developed and emerging economies – over the period from January 2012 to March 2017. Utilizing a wavelet-based, time-varying dynamic framework, they estimated the conditional correlations among these assets to derive hedge ratios with direct practical applications [13].

Kamalov [14] investigated the capabilities of neural networks in predicting large stock price movements. His results showed that such models can achieve high accuracy in predicting significant price fluctuations.

Feng et al. [15] proposed the use of adversarial learning to improve the generalization of stock price forecasting models. They showed that this approach can reduce overfitting and increase the accuracy of forecasts.

Khaidem et al. [16] applied the random forest method to forecast the direction of stock prices. They found that this ensemble method can effectively predict market trends using technical indicators.

This review demonstrates a wide range of approaches to financial time series analysis, from classical statistical models to modern machine learning methods applied to forecast and analyse the volatility of energy stocks.

While many studies focus on two-company comparisons or single-method forecasting, few have undertaken a comprehensive tri-partite analysis (Shell, BP, ExxonMobil) combining descriptive statistics, time-series models, machine learning, event studies, and regime detection under one framework. Our work addresses this gap by integrating multiple methodologies to assess risk-return profiles and co-movement patterns across the three companies for 2021–2025.

The results of a number of previous studies demonstrate diverse approaches to the analysis of time series of oil prices and shares of oil and gas corporations. In particular, the impact of geopolitical factors on oil prices was studied using GARCH models for risk assessment [7-9]. Another approach is the clustering of time series in order to identify phases of market consolidation and surges of volatility [17-21]. However, studies aimed directly at the comparative analysis of the dynamics of Shell, BP and ExxonMobil shares in the context of the current energy crisis remain fragmentary and often do not take into account the complexity of indicators, such as cumulative returns, financial risk metrics and investment efficiency. The purpose of this work is to provide a comprehensive analysis of the time series of Shell, BP, and ExxonMobil's share prices for the period 2021–2025. using descriptive statistics, graphical trend analysis, volatility measurement, cumulative return estimation, correlation analysis and clustering. In addition, we calculate key financial metrics (CAGR, Sharpe, Sortino, Max Drawdown, Calmar) for a comprehensive assessment of the risk-return characteristics of each company. Thus, the analysis allows us to better understand the impact of global events on the market dynamics of the shares of leading energy companies and will enable investors to make more informed decisions on portfolio management in the highly volatile conditions of the modern energy market.

In today's world, the energy sector is not just an essential sector of the economy – it is the foundation on which global economic development is built [22]. The stability of energy markets determines the functioning of industry, transport, utilities and many other areas of our lives [23]. That is why shares of energy companies are always in the spotlight of investors and analysts [24]. But what are shares, and why are they so important? Shares are securities that certify ownership of a share in a company. When you buy a share, you become a co-owner of the business, have the right to a part of its profits in the form of dividends and can receive income from the growth in the value of the shares themselves. The share price reflects not only the current financial condition of the company but also investors' expectations regarding its future development, reaction to global events and changes in the industry [25]. In our study, we focused on three giants of the energy sector: BP, Shell and ExxonMobil. Each of these companies has a unique history and plays a prominent role in the global energy landscape. The past few years have been particularly turbulent for the energy sector. The COVID-19 pandemic has led to an unprecedented drop in energy demand, and the subsequent economic recovery has led to a sharp increase in prices. The Russian invasion of Ukraine in 2022 triggered a global energy crisis, forcing Western countries to reconsider their energy strategies and seek alternative sources of supply. These events have had a significant impact on energy stocks, creating both new challenges and opportunities for investors. In this context, our research is particularly relevant. By analysing time series of stock prices, trading volumes and other indicators over the past 4 years, we seek to understand how the market reacts to global events.

Table 1. Regional distribution of electricity demand, 2020-2023

Terawatt hour (TWh)	**2020 year**	**2021 year**	**2022 year**	**2023 year**	**Growth rate 2020-2021.**	**Growth rate 2021-2022.**	**Growth rate 2022-2023.**
Worldwide	24,874	26,313	26,933	27,554	5.8%	2.4%	2.3%
Asia-Pacific	12,137	13,115	13,558	14,077	8.1%	3.4%	3.8%
of which China	7,471	8,222	8,479	8,846	10.1%	3.1%	4.3%
America	6,036	6,180	6,295	6,325	2.4%	1.9%	0.5%
of which USA	4,109	4,167	4,254	4,247	1.4%	2.1%	-0.2%
Europe	3,648	3,805	3,827	3,866	4.3%	0.6%	1.0%
European Union	2,625	2,740	2,762	2,774	4.4%	0.8%	0.4%
Eurasia	1,224	1,297	1,282	1,269	5.9%	-1.1%	-1.1%
Middle East	1,120	1,162	1,187	1,211	3.7%	2.2%	2.0%
Africa	709	754	783	805	6.4%	3.9%	2.9%

* Source: compiled based on materials from OECD "Europe Energy Market – 2022."

The current conditions for the functioning of transnational corporations are due to their increasing influence on international commodity markets, which allows them to gain competitive advantages due to the full or partial monopolization of such markets [26]. The key principle of the activities of TNCs (Transnational Corporations) in 2022 was the strengthening of the influence of energy companies and oil producers, which is due, first of all, to the war in Ukraine and the European sanctions policy. TNCs in the international global market are becoming not only a tool for ensuring energy security but also a key regulator of supplies and the functioning of commodity markets [27]. The most problematic are the energy market and the oil products market since a number of transformation processes take place on a basis that can affect the activities of transnational corporations [28]. Despite the persistence of shadow supply chains and mechanisms designed to circumvent sanctioned sales routes, access to fuel and energy resources remains severely constrained. As a result, European countries are increasingly compelled to adopt rational and sustainable measures to safeguard their energy security. The exit of commercial entities from the Russian Federation, a dominant player in the energy production and oil refining sectors, has triggered significant turbulence across financial markets. Among the most severely impacted has been the German economy, whose acute dependency on Russian energy supplies rendered it particularly vulnerable. The experience of German multinational corporations highlights the critical necessity for establishing innovative energy procurement channels and implementing macroeconomic restructuring initiatives. Moreover, in several European nations, the escalation in the prices of metals, financial services, and food commodities has exceeded the peaks witnessed during the 2008 financial crisis. Nevertheless, when contrasted with the magnitude of the 2022 energy crisis, contemporary challenges concerning energy consumption and the security of energy supply constitute a pressing global issue [28]. In order to further contextualize these developments, Table 1 presents a detailed overview of international and regional trends in electricity demand over recent years, accompanied by projections for 2023.

The information summarized in the table reflects a significant structural transformation of the global energy market in the aftermath of the war in Ukraine, resulting in a swift redefinition of its developmental trajectories [28]. This reorganization has been primarily propelled by the growing dominance of transnational corporations (TNCs) within the sector. Looking ahead, the strategic advancement of Europe's raw materials and energy markets is essential for the construction of a resilient and effective security architecture, particularly concerning the governance and oversight of TNC activities. Given their pivotal role in shaping the qualitative frameworks for commodity and consumer goods distribution, TNCs are increasingly recognized as central agents in the reconfiguration of the global energy order.

Table 2. The world's largest oil refining and mining companies, 2021.

TNC	**Net profit, million $**	**Number of employees, thousands of people**	**Assets, million $**	**Annual revenue, million $**
Exxon Mobil	32,520	83.7	349,493	382,597
Walmart	16,363	2,200	203,706	485,651
China National Petroleum	16,359	1,636.5	634,811	428,620
Royal Dutch Shell	14,874	94	353,116	431,344
State Grid	9,796	921.9	466,295	339,426
Sinopec Group	5,177	897.5	359,182	446,811
BP	3,720	84.5	284,305	358,678
Glencore	2,308	106.8	152,205	221,073

* Source: compiled based on materials from OECD "Europe Energy Market – 2022"

The transnational corporations (TNCs) identified in Table 2 are pivotal in determining the future direction of the global energy market, underscoring the pressing necessity to reinforce international legal frameworks that regulate their activities and address the persistent ramifications of the global energy and raw materials crisis [28]. The consequences of the Russian-Ukrainian conflict have been profoundly felt throughout Europe's economic environment since the onset of hostilities and are anticipated to have enduring effects. These destabilizing developments have disrupted European financial systems, altered macroeconomic conditions, destabilized global trade structures, and amplified the strategic influence of TNCs across critical sectors.

In the present circumstances, a key characteristic of TNC behaviour is the strategic deployment of internal capacities and financial resources to exert leverage over the evolving energy crisis both within Europe and on a global scale. To effectively mitigate the consequences of the Russian-Ukrainian war, it is imperative to enhance international regulatory frameworks governing TNCs and to promote the strengthening of corporate social responsibility initiatives. Addressing these aspects would considerably bolster the capacity of transnational corporations to contribute to resolving significant socio-economic challenges on a global level [28].

The analysis of stock prices of leading energy corporations such as Shell, BP, and ExxonMobil over the 2021–2025 period represents a highly relevant and timely topic in contemporary financial research. Nevertheless, based on the currently available sources, no studies were identified that specifically focus on the joint stock performance of these three companies during the specified timeframe. However, there exists a substantial body of literature offering general methodologies and analytical frameworks for evaluating financial indicators and market dynamics, which can be effectively adapted to conduct an in-depth analysis of these corporations.

- Financial analysis of companies;
- Fundamental analysis of stocks;
- Macroeconomic indicators;
- Market indices and stock price dynamics.

Financial analysis is a key tool for assessing the state of an enterprise and its prospects. It helps to identify the strengths and weaknesses of the company and evaluate its financial stability and operational efficiency. The leading indicators that are analysed include profitability, liquidity, solvency and profitability. This analysis is the basis for strategic planning and management decision-making.

Fundamental analysis is aimed at determining the intrinsic value of shares by studying the company's economic situation, industry trends, and financial indicators. This approach allows you to assess whether the company's shares are overvalued or undervalued, which is essential for investors when making investment decisions.

To fully understand market dynamics, it is necessary to take into account macroeconomic indicators that affect the activities of companies. The National Bank of Ukraine provides statistical data on macroeconomic indicators that can be useful for analysing the economic environment in which companies operate.

The level of stock prices traded on stock exchanges is assessed using stock indices. They allow investors to determine the general state of the economy and the securities market.

3. Material and Methods

3.1. Methodology for research

Although no specific publications have been identified on the analysis of Shell, BP, and ExxonMobil stocks for the period 2021–2025, there are generally accepted methods and approaches to the analysis of financial indicators and market dynamics that can be applied to the study of these companies. The use of fundamental analysis, assessment of macroeconomic indicators, and analysis of stock indices are key elements in identifying patterns and assessing the financial condition of companies. The study analysed data on the shares of three leading energy companies, Shell, BP, and ExxonMobil, for the period 2021–2025. The main goal is to identify patterns, analyse market dynamics, and assess the financial performance of companies.

1. Data. For the analysis, an open dataset from the Kaggle resource was used:

https://www.kaggle.com/datasets/pinuto/energy-crisis-and-stock-price-dataset-2021-2024

This set contains daily stock price data of three energy companies – Shell, BP and ExxonMobil – for the period January 2021 – January 2025, including the following variables: Open, High, Low, Close, Adj Close, Volume. The data is downloaded in CSV format and processed in Python (libraries: pandas, numpy, matplotlib, seaborn, sklearn). The date has been set as an index, missing values are missing, and no duplicates have been found.

Analysing financial data is especially important for making informed investment decisions and understanding market dynamics. The analysis process begins with an initial review of the dataset, which includes studying the data structure, the presence of gaps, variable types, and basic descriptive statistics. It allows you to better understand the nature and quality of the available data, as well as plan further work with them. Three datasets have been downloaded from open sources. Key indicators:

- Open - opening price;
- High - maximum price;
- Low - minimum price;
- Close - closing price;
- Adj Close - adjusted closing price;
- Volume - trading volume.

Missing values were analysed and, if necessary, filled in by interpolation. Various charts are used to study stock price trends, which help visualize price dynamics and identify significant trends. Cumulative provides an opportunity to assess the accumulated return over a specific period, which is an essential indicator of investment performance, while moving volatility helps to evaluate the level of market risk and its change over time. Such a comprehensive analysis of charts allows you to see how prices have changed in the past and assess potential risks and investment opportunities based on historical data. Distribution charts, such as histograms, are used to analyse the nature of the return distribution, which helps to assess the probability of obtaining a certain level of profit or loss.

2. Statistical description. The basic statistical indicators were calculated: mean, median, mode, variance, standard deviation, coefficients of asymmetry and excess. All calculations were carried out in accordance with the standard statistical methodology [29]:

- Average: $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$;
- Variance: $\sigma^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$;
- Standard deviation: $\sigma = \sqrt{\sigma^2}$.

3. Graphical analysis of trends. To visualise trends, Adj Close line charts with an overlay of moving averages (SMA) with windows of 14 and 50 days were used [30]:

$$\mathrm{SMAt} = \frac{1}{\mathrm{N}} \sum i = t - N + 1^{t} \ P_i.$$

4. Exponential smoothing (EMA). The Exponentially Weighted Moving Average (EWMA) **was used** to account for the latest values with higher weights [31]:

$$\mathrm{EMAt} = \alpha \mathrm{P_t} + (1 - \alpha) \cdot \mathrm{EMA}t - 1.$$

5. Volatility. Moving volatility is estimated based on the standard deviation of the 30-day daily yield window:

$$\sigma_t = \sqrt{\frac{1}{n-1}\sum_{i=t-n+1}^{t}(r_i - \bar{r})^2}.$$

where r_i – is the daily yield.

6. Cumulative yield. Calculated as the product of daily yields [32]:

$$CR_t = \prod_{i=1}^{t}(1 + r_i).$$

7. Financial Metrics (CAGR, Sharpe, Sortino, Calmar):

- CAGR (Compound Annual Growth Rate): $\text{CAGR} = \left(\frac{\text{V}_\text{f}}{\text{V}_\text{i}}\right)^{1/\text{n}} - 1$.
- Sharpe Ratio: $\text{Sharpe} = \frac{\text{CAGR} - r_f}{\sigma}$.
- Sortino Ratio (uses only unfavourable fluctuations): $\text{Sortino} = \frac{\text{CAGR} - r_f}{\sigma_-}$.
- Max Drawdown. The most significant drop from the all-time high.
- Calmar Ratio: $\text{Calmar} = \frac{\text{CAGR}}{|\text{Max Drawdown}|}$.
- Risk-free rate $r_f = 0.01$ (conditionally fixed value for comparison).

8. The Pollard method smooths the data based on symmetrical weights around the central value [33]:

$$y_t = \sum_{i=-k}^{k} w_i x_{t+i}.$$

Windows $w = [3, 5, \ldots, 15]$ are applied. Author's addition – a table of local minimums/maximums and correlations with the initial series has been built.

9. Median smoothing. Window median method – the median of the window of size w replaces each value. Median filter is effective for preserving edges while reducing noise [34].

10. Correlation analysis. The Pearson coefficient is used:

$$r_{xy} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2}\sqrt{\sum(y_i - \bar{y})^2}}.$$

The coefficient of determination R^2 for each pair of companies is also calculated.

11. Autocorrelation (ACF). An autocorrelation function has been applied. The ACF function measures correlation between a time series and its lagged values [35]. Calculated lags of up to 30 days using statsmodels.tsa.stattools.acf.

12. K-Means clustering is an unsupervised learning algorithm that partitions data into K clusters by minimising within-cluster variance [37]. Before clustering, standardisation was carried out (StandardScaler). The number of clusters $k = 2$ was selected using the "elbow" method. Next, the distribution of clusters in time was constructed to identify market phases (growth/consolidation).

All analysis is done in Python 3.9. Libraries used: numpy, pandas, matplotlib, seaborn, scikit-learn, statsmodels. Code, graphs, and analytics tables are saved as a Jupyter notebook. All of these analysis methods together help us better understand how stocks behave in the stock market. Using various graphs, calculations, and calculation methods, we can see not only how prices change but also understand hidden patterns and relationships between different companies. It allows investors to make informed decisions about buying or selling stocks and better assess possible risks and potential profits. Such a comprehensive approach to data analysis is critical in today's world, where financial markets are becoming increasingly complex and interconnected.

3.2. Methodology for calculating financial indicators

To quantify the performance of the three leading energy companies – ExxonMobil, Shell and BP – a number of key financial ratios have been calculated based on daily stock price data for the period 2021-2025. The calculations were performed in Python using the pandas, numpy, matplotlib and statsmodels libraries.

1. Cumulative Return determines the total return of an asset for the period:

$$\text{Cumulative Return}_t = \frac{P_t - P_0}{P_0} = \prod_{i=1}^{t}(1 + r_i) - 1,$$

where P_t – is the price at the end of the period; P_0 – is the price at the beginning of the period; r_i – is the daily yield:

$$r_i = \frac{P_i - P_{i-1}}{P_{i-1}} = \ln\left(\frac{P_i}{P_{i-1}}\right).$$

For prices, "Adjusted Close" was used, which takes into account dividends and splits.

2. Compound Annual Growth Rate (CAGR) evaluates performance taking into account compound percentage:

$$\text{CAGR} = \left(\frac{P_{\text{end}}}{P_{\text{start}}}\right)^{\frac{1}{n}} - 1,$$

where P_{end}, P_{start} – are prices at the end/beginning of the period; n – is the duration in years (in the study $n = 4.25$).

3. Standard deviation (Volatility) – estimation of the average fluctuation of daily earnings:

$$\sigma = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(r_i - \bar{r})^2}.$$

It was brought to an annual scale by multiplying by $\backslash sqrt\{252\}$ (≈ number of trading days in a year).

4. Sharpe Ratio – Risk-Based Performance Assessment:

$$\text{Sharpe Ratio} = \frac{R_p - R_f}{\sigma_p},$$

where R_p – is annual yield (CAGR); R_f – risk-free rate 3%; σ_p – is the annual volatility. Interpretation – how much additional profit an investor receives for each unit of risk.

5. Sortino Ratio is an updated version of the Sharpe Ratio, which takes into account only negative fluctuations:

$$\text{Sortino Ratio} = \frac{R_p - R_f}{\sigma_d},$$

where σ_d – downside deviation: standard deviation of only those r_i that are smaller R_f. Advantage – does not punish for positive deviations, only for the risk of losses.

6. Calmar Ratio estimates the ratio between yield and maximum drawdown:

$$\text{Calmar Ratio} = \frac{CAGR}{|\text{Max Drawdown}|}.$$

7. Max Drawdown – the maximum drop of the portfolio from the local peak to the local low:

$$\text{MaxDrawdown} = \min\left(\frac{P_t - P_{\text{peak}}}{P_{\text{peak}}}\right).$$

The search is carried out in a cumulative series.

8. Trading Volume – Summary Activity Metric:

- The sum or average of daily volume during 2021–2025.
- It is used to assess liquidity and interest from investors.

3.3. Rationale for the use of K-means clustering

The purpose of clustering is to divide the time series of stock prices into structurally different market regimes – periods:

- active growth, when the price is steadily growing,
- consolidation, when the price fluctuates within a narrow range or shows a weak trend.

It is essential for:

- Detection of market phases without the need for manual revision;
- Building entry/exit strategies in automated systems.

Although this is not clearly stated in the article, it can be assumed from the code and description that clustering was carried out on the basis of the following normalised features calculated from the prices of "Adj Close". These traits allow the algorithm to determine the similarity of days by their market profile.

Table 3. Main features for clustering

Sign	Essence	Why
Daily Return	Percentage change in price compared to the previous day	Determines the phases of growth or fall
Volatility	The difference between the max and min price per day or the moving standard deviation	Determines periods of turbulence
SMA_20	Simple 20-day moving average	Signals a trend
RSI	Relative Strength Index	Indicates overbought/oversold
WMA_20	Weighted Moving Average	Smoothes local peaks

Table 4. Advantages of using K-means

Approach	Disadvantages
Visual inspection	Subjectivity, unscalable, not reproducible
Threshold rules (volatility $> x$%)	Not adaptive, poorly takes into account a set of features (e.g. volatility + trend)
K-means	Gives automatic multi-dimensional grouping based on multiple factors at once, without human bias

After clustering (k=2, selected using the "elbow" method), the following is obtained:

- Cluster 1 – periods of low fluctuations, small changes – consolidation.
- Cluster 2 – periods of sharp rises or falls, active trading – zones of impulse movement.

It can be seen in the cluster graph, where the stages of the market are marked in colour. Graphically, this classification is consistent with the behaviour of the indicators.

Well-known financial researchers use the K-Means method to identify market modes (bull/bear/stable). In this case, clustering acts as a mode detector and an alternative to Markov switching models. With the right choice of traits and normalisation, K-means provides simple and effective segmentation.

K-means clustering in this study has a case for application. It is based on the multidimensional presentation of stock behaviour (return, volatility, technical indicators), allows you to distinguish periods of market modes and is superior in flexibility to visual or manual methods. For complete clarity, it is worth adding a list of features to the article and an explanation of what exactly clustering allows you to detect – this will make the method not only acceptable, but also reasonably valuable.

Detailed justification for the choice of each method:

1. Moving averages (SMA, EMA) are the primary tool for technical analysis, which allows you to smooth out short-term price fluctuations. SMA (Simple Moving Average) displays uniform dynamics over a fixed time window, which is helpful for detecting trends. EMA (exponential) – gives more weight to the latest data, which allows you to react faster to new market changes. Alternative methods (e.g. Hull MA, Triangular MA) are less common and have a more complex interpretation, making it challenging to apply in practice. The SMA and EMA are the industry standard and are widely used both in the literature and among traders and investors. For financial time series with daily or weekly intervals, SMA/EMA are optimal in terms of simplicity and stability.

2. Exponential smoothing (for example, WMA or Pollard), compared to the simple average, better reflects the current market situation, without ignoring historical dynamics. In case of increased volatility (as in 2022-2023), it is vital to respond promptly to changes, and the WMA/EMA provides such sensitivity. Complex other filters (e.g., Kálmán or Savitzky-Golay) require deeper assumptions about the data generation process (e.g., Gaussian noise reduction) or are mostly mathematical rather than financial tools. For financial series that have an unstable structure, simple exponential smoothing shows stable results without requiring retraining or complex parameterisation.

3. K-Means clustering ($k = 2$) is a fast and straightforward method that allows unsupervised learning to automatically divide the time series into structural phases (for example, growth, stagnation). Value $k = 2$ will enable you to logically separate trend periods (growth/decline) from flat/consolidation periods. The method does not require a priori knowledge of data distribution, which is convenient for non-linear and noisy financial series. K-Means is chosen to balance interpretation, speed, and functionality.

Table 5. Reasons for not choosing other clustering methods

Method	Reasons for refusal
DBSCAN	It requires an optimal choice of eps but does not work well with quasi-linear series.
Hierarchical clustering	It requires a lot of computing resources and complex interpretation.
Gaussian Mixture Models (GMM)	It assumes Gaussian clusters, which is not always true for financial data.
Spectral clustering	Impractical for large time series without first reducing the dimension.

Table 6. Systematic justification for the choice of the entire methodology

Criterion	SMA/EMA/WMA	K-Means Clustering
Simplicity and computability	yes	yes
Interpretation for finance	Yes (Trend Visualization)	Yes (separation of market phases)
Sensitivity to new changes	EMA, WMA react quickly	Detects phases without training
Connection with practice	Widely used	Actively used in AI trading
Alternatives	More complex / less stable	Require additional processing

The chosen analysis methods – moving averages (SMA, EMA), exponential smoothing (WMA) and K-Means clustering – are based on their ability to adequately reflect structural patterns in financial time series under conditions of high volatility and limited predictability. Moving averages provide effective smoothing of local fluctuations and trend detection, while EMA and WMA allow for faster response to market changes. Method K-Means with the k=2 parameter will enable you to automatically classify periods into trend and consolidation periods. It is critical for dynamic market evaluation without the need for subjective hypotheses. Alternative approaches were rejected due to their complexity, poor scalability, or inconsistency with the structure of financial data.

Justification for choosing the period January 2021 – January 2025:

1) The beginning of January 2021 had a phase transition after COVID-19. First of all, 2020 was a period of historic decline in oil demand (WTI fell below $0) due to lockdowns and the pandemic. The beginning of 2021 is the start of economic recovery, the launch of mass vaccination and the recovery of demand for energy. Shares of energy companies began to return to pre-crisis levels, and this created a base entry point for a new price cycle.

2) 2021–2022 had a global energy shock, mainly due to the war in Ukraine (February 2022). It was the beginning of a significant geopolitical shift that led to:

- Sanctions against Russia;
- Restrictions on oil and gas supplies to Europe;
- A sharp rise in prices for Brent, gas and electricity.

During this period, energy companies received super profits, and investors massively switched from tech giants to "value stocks". It is one of the most critical periods of capital reallocation in the stock market.

3) 2023–2024 is a period of normalisation. There is a decrease in oil prices, tightening of monetary policy (rate increases), and inflationary pressure. The EU and US governments are stimulating the transition to "green energy" by introducing:

- Taxes on excess profits,
- ESG reporting,
- New requirements for reducing CO_2 emissions.

Energy stocks are becoming less predictable, and volatility and regulatory risk are increasing. It provides a contrast between the growth phase and the strategic adaptation phase.

4) January 2025 is the conditional point of the end of the phase. In particular, the 4-year interval allows you to cover the full investment cycle: growth – peak – correction – stabilization. Many temporary regulatory decisions related to the war situation (for example, the tax on excess profits) are expected to be completed, which makes January 2025 a logical limit for the analytical section.

The choice of the period from January 2021 to January 2025 is driven by the desire to cover the full price cycle, including the phases of post-pandemic recovery, the energy shock caused by the war in Ukraine, and further market stabilisation. The beginning of the period coincides with the recovery in demand after COVID-19, and the middle covers an unprecedented rise in energy prices and excess profits of companies due to geopolitical turbulence. The end of the period marks the moment when markets began to adapt to the new macroeconomic and regulatory reality. This choice allows us to study the response of the energy sector to the deepest crises of the last decade and assess the resilience of companies in the face of strategic uncertainty.

3.4. Comparative analysis of the research approach with similar methods

The proposed approach is best suited for integrated diagnostic analysis – when the goal is not so much to predict as to understand, compare, visualise and interpret the dynamics of financial markets in a crisis. While models such as ARIMA-GARCH or LSTM focus on forecasting, the proposed methodology is an effective tool for financial review, taking into account crisis impacts, especially for portfolio management and institutional investors.

Table 7. Advantages of the proposed method:

Criterion	**The advantage of the technique in the article**
Complexity	The article combines classical statistical analysis, financial metrics, time series methods, and clustering, which is rare in a single study.
Multi-company	A full-fledged three-company comparison (Shell, BP, ExxonMobil) was performed, in contrast to most works limited to the analysis of 1-2 companies.
Integration of the event component	Geopolitical events (war in Ukraine, sanctions, exit from the Russian market) are considered as factors affecting financial performance – this is a rarity among formal models.
Stability analysis	The use of Sharpe, Sortino, Max Drawdown and Calmar indicators allows you to assess not only the profitability, but also the risk and reliability of the asset.
Anti-aliasing methods	Two types of smoothing (Pollard and median) are included, allowing you to identify both short-term and long-term trends without losing structure.
Interpretation	All methods are visualised (trends, clusters, ACF, correlations), which simplifies interpretation for non-technical users (investors).

Table 8. Disadvantages compared to other approaches:

Criterion	Disadvantage of the method in the article	Comment
Lack of forecasting	The model does not attempt to predict future stock prices (unlike ARIMA, GARCH, LSTM)	The paper has an analysis of past data, but does not build a prediction model.
Lack of hypotheses and significance testing	Statistical hypothesis testing is not performed (p-value, Granger tests, ADF)	It limits the formal validity of causal inferences.
Simplified clustering implementation	Only K-Means with fixed k=2 is used without comparison with other methods (DBSCAN, hierarchical clustering)	It can be sensitive to initial conditions and scaling.
Deep ML models are not used.	There are no neural networks (e.g. LSTM, GRU) that work better with time series.	Although this would make interpretation more difficult, it would increase the accuracy of the models.
Limited time context	Study covers 2021–2025 without estimating longer cycles (10+ years)	This is justified given the crisis period, but reduces the overall results.

4. Experiments and Results

The purpose of this study is to comprehensively characterise and compare the financial stability, market behaviour and risk profile of shares of three leading multinational energy companies (Shell, BP, ExxonMobil) during the energy crisis of 2021–2025 by combining classical and modern methods of time series analysis. In particular, the study aims to:

1. Identify periods of increased volatility that correlate with geopolitical events.
2. Compare profitability and risk using financial metrics (CAGR, Sharpe, Sortino, Calmar, Max Drawdown).
3. To investigate intra-market clusters (market modes) using machine learning.
4. Evaluate autocorrelation and correlation structures to identify dependencies between companies.
5. Provide an interpreted toolkit for institutional investors to diagnose the investment attractiveness of companies in times of crisis.

The methods are not chosen arbitrarily. They are systematically divided by levels of analysis, according to the following logic:

Table 9. Explanation of the use of the number of methods

Level	Method	Purpose
Descriptive level	Mean, variance, graphs	General view of price behaviour
Dynamic level	SMA, EMA, Smoothing	Identification of trends, cycles, tipping points
Risk level	Volatility, Max Drawdown, Sharpe	Assessment of instability and investment risk
Correlation level	ACF, cross-correlations	Detection of recurrence, dependencies
Clustering level	K-Means	Detection of structures in time series
Financial level	CAGR, Sortino, Calmar	Asset Performance Comparison

Thus, the goal is not just to "apply all methods", but to create a multi-level panorama of the behaviour of energy companies in the face of geopolitical turbulence.

4.1. Data preparation and initial analysis

First of all, three datasets were loaded with information about the three most prominent energy companies respectively. The first, "Shell," is an energy company based in London, UK, and in May 2023, it left the Russian market. The second - "BP" is also based in London and, since February 2023 has suspended operations in the Russian Federation and is withdrawing its stake. The last company, namely "ExxonMobil", is an American energy company, one of the most profitable in the world, which left the Russian market in October 2022. Link to the datasets: https://www.kaggle.com/datasets/pinuto/energy-crisis-and-stock-price-dataset-2021-2024/code

After downloading the datasets, the dates were set as indices (Fig. 1), general information about the data was reviewed (Fig. 2a), and the number of duplicates and missing values was checked (Fig. 1b).

It is worth noting that each column is responsible for a specific indicator, namely:

- Open - the price at which the stock opened on a particular day;
- High - the highest stock price for this day;
- Low - the lowest stock price for this day;
- Close - the closing price of the stock for this day;
- Adj Close - adjusted closing price taking into account splits and dividends;
- Volume - total number of shares traded during the day.

```
shel_df.set_index('Date', inplace=True)
shel_df.head()
```
✓ 0.0s

	Open	High	Low	Close	Adj Close	Volume
Date						
2021-01-04	36.250000	36.360001	35.494999	36.029999	31.140999	6794599
2021-01-05	37.189999	39.014999	37.115002	38.509998	33.284489	8518589
2021-01-06	39.910000	40.680000	39.564999	40.290001	34.822952	9180332
2021-01-07	40.000000	40.625000	39.744999	40.340000	34.866161	4829303
2021-01-08	40.360001	40.419998	39.770000	40.240002	34.779736	5546750

Fig. 1. Total number of shares traded during the day.

```
shel_df.info()

<class 'pandas.core.frame.DataFrame'>
DatetimeIndex: 977 entries, 2021-01-04 to 2024-11-19
Data columns (total 6 columns):
 #   Column     Non-Null Count  Dtype
---  ------     --------------  -----
 0   Open       977 non-null    float64
 1   High       977 non-null    float64
 2   Low        977 non-null    float64
 3   Close      977 non-null    float64
 4   Adj Close  977 non-null    float64
 5   Volume     977 non-null    int64
dtypes: float64(5), int64(1)
memory usage: 53.4 KB
```
(a)

```
shel_df.duplicated().sum()

np.int64(0)

shel_df.isna().sum().sum()

np.int64(0)
```
(b)

Fig. 2. (a) General information about the data; (b) Checking for duplicates and missing values.

Similar actions were performed with the other two datasets, where the results were identical: no duplicates or missing values were found, and the dates were set as indices.

4.2. Descriptive statistics

The next step was to provide descriptive statistics. In addition to the classical values, skewness, kurtosis, mode, median, range, etc. were also added. The results for each company are shown in Fig. 3-5, respectively. If we look, for example, at the statistics of the company "ExxonMobil", the value of the variance of the trading volume (Volume), which is relatively high, indicates significant fluctuations in the volume of transactions between trading days. Also, the minimum closing price was 41.50, and the maximum was 125.37, which means an extensive range of prices for the period. Negative skewness in price indicators indicates that the distribution has a longer "tail" on the left, which may indicate frequent price declines. In contrast, negative kurtosis indicates a flat price distribution, i.e., the absence of sharp peaks and troughs.

	Open	High	Low	Close	Adj Close	Volume
count	977.000000	977.000000	977.000000	977.000000	977.000000	9.770000e+02
mean	55.961735	56.407882	55.503777	55.968639	52.175158	4.872050e+06
std	10.561569	10.543511	10.552337	10.549633	11.723654	1.728606e+06
min	35.750000	36.360001	35.494999	36.029999	31.140999	7.529140e+05
25%	47.000000	47.549999	46.539101	47.009998	42.006836	3.703752e+06
50%	57.340000	57.779999	56.770000	57.330002	52.582939	4.550720e+06
75%	64.440002	64.820000	63.830002	64.320000	61.728855	5.659883e+06
max	74.290001	74.605003	73.910004	74.169998	72.305870	1.696047e+07
standard error	0.337895	0.337317	0.337599	0.337513	0.375073	5.530301e+04
variance	111.546731	111.165620	111.351816	111.294756	137.444055	2.988079e+12
median	57.340000	57.779999	56.770000	57.330002	52.582939	4.550720e+06
mode	59.630001	39.770000	38.450001	39.080002	33.682022	7.529140e+05
range	38.540001	38.245003	38.415005	38.139999	41.164871	1.620755e+07
skewness	-0.209640	-0.224660	-0.198764	-0.211867	-0.072581	1.598888e+00
kurtosis	-1.097599	-1.090125	-1.109012	-1.098675	-1.141423	5.159039e+00
sum	54674.615025	55110.500431	54227.190197	54681.359974	50975.129627	4.759992e+09
ci	0.662273	0.661141	0.661694	0.661525	0.735143	1.083939e+05

Fig. 3. Descriptive statistics for Shell data

	Open	High	Low	Close	Adj Close	Volume
count	977.000000	977.000000	977.000000	977.000000	977.000000	9.770000e+02
mean	32.117462	32.387032	31.815558	32.101955	29.224900	1.072997e+07
std	4.823779	4.814396	4.820401	4.816670	5.472840	5.010127e+06
min	20.680000	21.129999	20.430000	20.750000	17.118425	3.506800e+06
25%	27.950001	28.219999	27.570000	27.940001	24.352745	7.348200e+06
50%	32.570000	32.810001	32.279999	32.470001	30.439859	9.470100e+06
75%	35.849998	36.119999	35.580002	35.849998	33.861835	1.302430e+07
max	40.689999	41.380001	40.470001	41.020000	38.170727	4.485300e+07
standard error	0.154326	0.154026	0.154218	0.154099	0.175092	1.602882e+05
variance	23.268845	23.178413	23.236268	23.200307	29.951978	2.510137e+13
median	32.570000	32.810001	32.279999	32.470001	30.439859	9.470100e+06
mode	35.709999	35.459999	35.320000	34.889999	20.552675	8.183700e+06
range	20.009998	20.250002	20.040001	20.270000	21.052301	4.134620e+07
skewness	-0.222496	-0.219286	-0.220556	-0.218590	-0.283961	1.876153e+00
kurtosis	-1.013255	-1.010713	-1.022048	-1.016078	-1.198516	6.040106e+00
sum	31378.760012	31642.129978	31083.799982	31363.610023	28552.727652	1.048318e+10
ci	0.302480	0.301891	0.302268	0.302034	0.343180	3.141648e+05

Fig. 4. Descriptive statistics of BP data

	Open	High	Low	Close	Adj Close	Volume
count	977.000000	977.000000	977.000000	977.000000	977.000000	9.770000e+02
mean	92.278618	93.319642	91.301505	92.336315	86.770530	2.090369e+07
std	23.157683	23.293578	22.973333	23.136443	24.600429	8.748485e+06
min	41.450001	42.250000	41.000000	41.500000	35.163219	7.136000e+06
25%	64.610001	65.330002	64.180000	64.519997	57.733021	1.487760e+07
50%	102.000000	103.089996	100.910004	102.059998	97.441528	1.875670e+07
75%	111.639999	113.089996	110.610001	111.610001	106.250908	2.441240e+07
max	125.250000	126.339996	124.860001	125.370003	124.348221	8.443940e+07
standard error	0.740880	0.745228	0.734982	0.740200	0.787037	2.798888e+05
variance	536.278268	542.590756	527.774010	535.294973	605.181127	7.653600e+13
median	102.000000	103.089996	100.910004	102.059998	97.441528	1.875670e+07
mode	113.480003	104.120003	56.830002	64.309998	119.369125	1.428320e+07
range	83.799999	84.089996	83.860001	83.870003	89.185001	7.730340e+07
skewness	-0.559850	-0.571333	-0.550926	-0.561528	-0.483469	1.892131e+00
kurtosis	-1.173067	-1.168634	-1.180879	-1.171709	-1.214585	6.293288e+00
sum	90156.210045	91173.289955	89201.569965	90212.579887	84774.808151	2.042290e+10
ci	1.452125	1.460646	1.440565	1.450793	1.542593	5.485821e+05

Fig. 5. Descriptive statistics of ExxonMobil data

4.3. Graphical presentation of trends data in company stock prices over the past 4 years

The next step was to identify trends over the past 4 years on Adj Close - the adjusted closing price taking into account splits and dividends. To do this, graphs were constructed where the X axis determines the dates, and the Y axis the share price for the corresponding date. The results are shown in Fig. 6-8, respectively, for each company. In the case of Shell shares, prices generally increased, excluding some crises in mid-2022 (a new CEO appeared) and 2024 (global oil surplus, in particular in the USA). The peak price was approximately in September 2024, but after that, the share prices began to decline. Perhaps market factors, investor attitudes or internal conflicts within the company played a role.

Fig. 6. Trends in Shell stock prices

Regarding BP, from 2021 to mid-2024, albeit with fluctuations, stock prices grew, but in the second half of 2024, prices began to fall. The reason could be a negative IODA (business activity expectations index), which indicates pessimistic sentiment in the business environment. It could affect investment decisions and, accordingly, the value of the companies' shares. Unlike the previous two, ExxonMobil shows a positive trend, as here, at the end of 2024, stock prices did not fall but still held a high bar. It may indicate at least a better adaptation of the company to global changes and various market conditions.

Fig. 7. Trends in BP stock prices

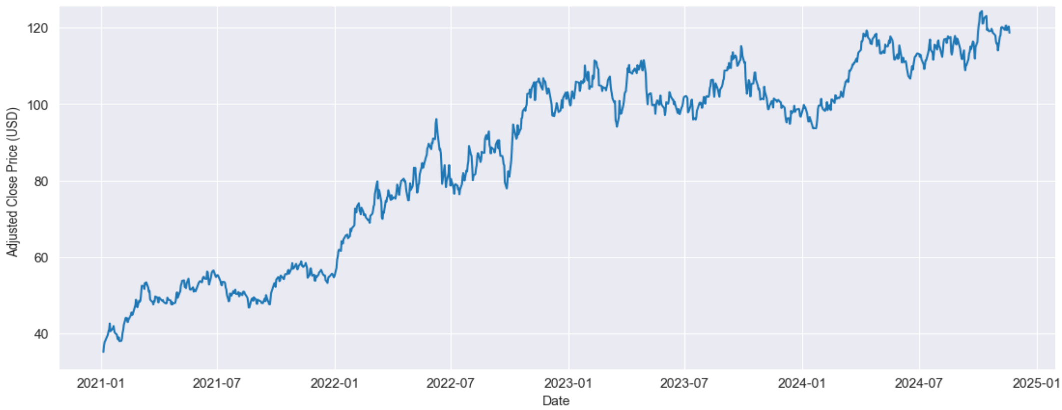


Fig. 8. Trends in ExxonMobil stock prices

4.4. Market activity

At the stage of market activity research, the daily trading volume indicator was determined for the companies under study. This indicator is essential for assessing the total financial volume of securities transactions on a daily basis and allows you to track the level of market activity of companies. The daily trading volume was calculated as the product of the opening price of shares (Open) and the trading volume (Volume) for each trading day during the period under study. Based on the values obtained, time series were constructed that reflect the dynamics of total trading volumes for each company (Fig. 9). The abscissa axis shows the time interval from January 2021 to January 2025, and the ordinate axis shows trading volumes in billions of units (1e9).

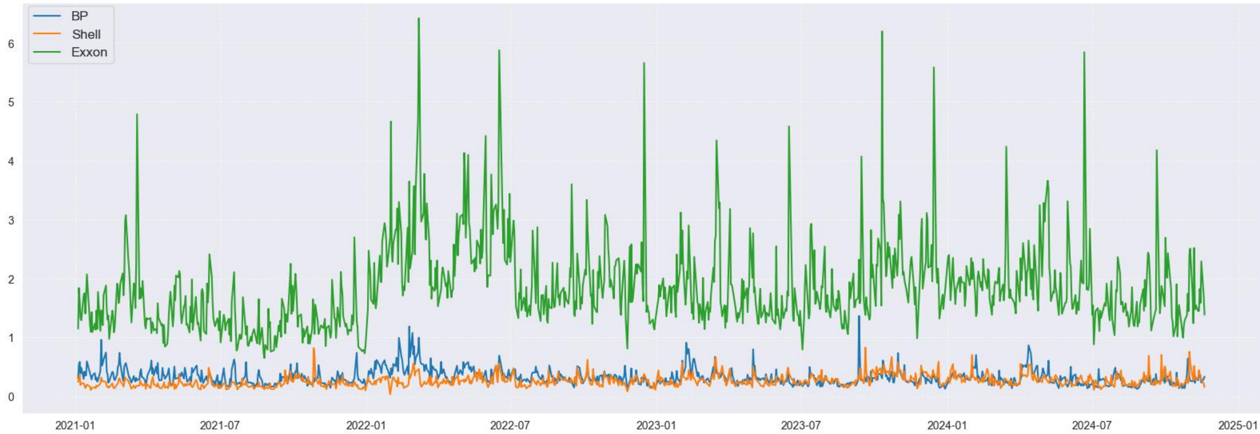


Fig. 9. Trading volumes of BP, Shell, and Exxon over four years

Exxon (green line) is marked by a significantly higher volume of shares traded compared to BP (blue) and Shell (orange). It indicates that Exxon shares are much more popular among traders and investors. On average, Exxon's daily trading activity fluctuates between approximately 1 and 3 billion units, with periodic peaks reaching 6 billion, while for BP and Shell, this figure usually does not exceed 1 billion. 2021 is characterized by relatively moderate market activity for all three companies. Exxon's trading volumes fluctuate mainly in the range of 1-2 billion, with individual peaks up to 4.8 billion. Already in 2022, there is an increase in trading activity, especially for ExxonMobil. The first half of the year shows the highest trading volumes of the entire observation period, with peaks reaching 6.5 billion units. It may be due to global geopolitical events and the energy crisis, including Russia's full-scale invasion of Ukraine and sharp fluctuations in worldwide oil and gas prices. Western sanctions against the Russian energy sector have created opportunities for American and European oil companies to increase their presence in markets traditionally served by Russian suppliers. In 2023, some stabilization of market activity is observed, but it remains higher compared to 2021: Exxon's trading volumes fluctuate mainly in the range of 1.5-3 billion units. BP and Shell continue to demonstrate more restrained fluctuations, with a stable level below 1 billion units. In 2024, the trend of stable market activity continues, with regular peaks of Exxon's activity exceeding 5 billion units. For BP and Shell, activity remains relatively stable. ExxonMobil has consistently shown high trading volumes throughout the period under review, often reaching peak levels that far exceed those of BP and Shell, indicating heightened investor interest in ExxonMobil shares.

4.5. Cumulative return

Next, the cumulative return was calculated, which is an indicator for assessing the effectiveness of investments in the long term, reflecting the cumulative result of all price changes over the period, taking into account reinvestment. That is, this indicator allows us to estimate how the value of the investment would have changed if the investor had purchased shares at the beginning of the period and held them until the end. For example, if we look at the Exxon line, by 2025, the initial investment had increased by approximately 3.5 times. That is, if $1,000 had been invested, we would have received about $3,500. To calculate the cumulative return, daily percentage changes in share prices (daily returns) were used based on the adjusted closing price (Adj Close). Next, the cumulative product of daily returns was calculated, on the basis of which the time series shown in Fig. 10 was constructed. The initial value is normalized to 1.0. It allows us to directly compare the performance of companies regardless of their absolute market capitalization. 1.0 corresponds to the initial investment value, and each subsequent value shows how many times the initial investment would have increased by a specific date.

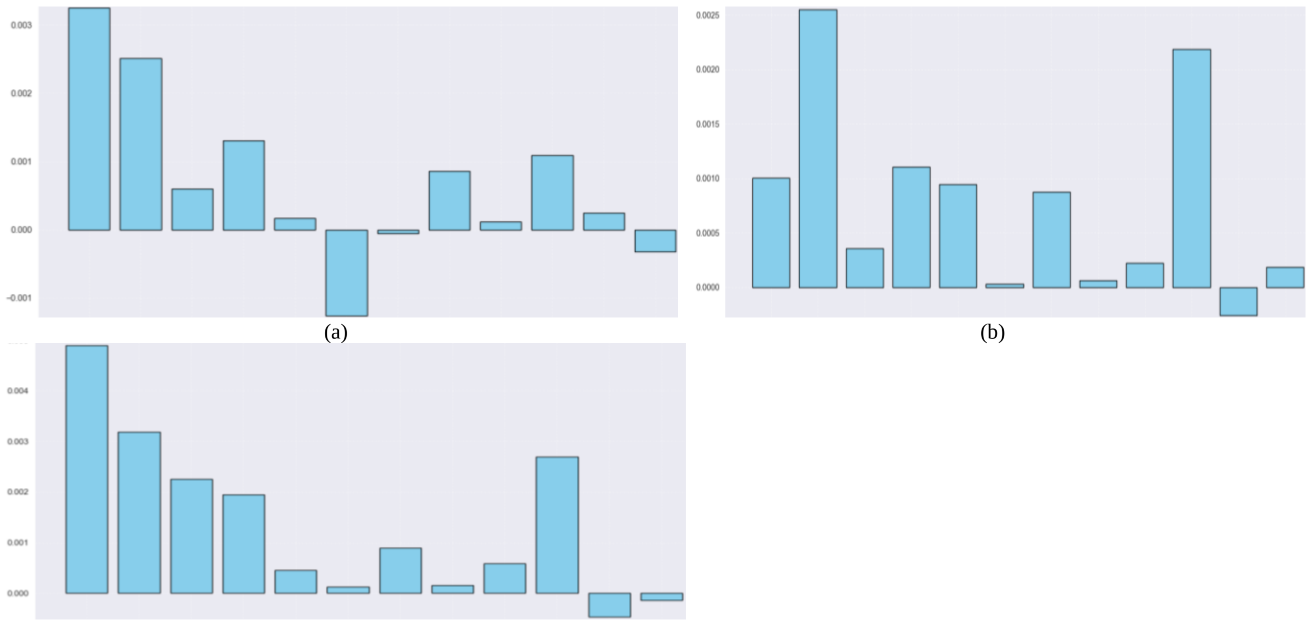
During 2021, moderate returns were observed with little overall growth. During this period, ExxonMobil (green line) showed the first signs of outpacing growth, achieving a cumulative return of about 150%, while BP (blue line) and Shell (orange line) showed more restrained growth, reflecting stable but less dynamic investment prospects. The beginning of 2022 was marked by a sharp acceleration in the development of returns for all three companies. The growth of Exxon shares was especially noticeable, which demonstrated a rapid upward trend, reaching the level of 2.5 by the middle of the year. It indicates an adequate response of the company to changing global market conditions and the high attractiveness of its shares among investors. At the same time, a persistent gap was formed between Exxon's indicators and European competitors. BP and Shell show similar dynamics for most of the period but with some differences. By the beginning of 2024, both companies achieved approximately 200-220% cumulative returns. However, after that, there is a clear difference in their dynamics: BP shows a noticeable decline, while Shell maintains a more stable position. Various factors could have influenced this: geopolitical, structural, market. ExxonMobil remains the most attractive asset for investors, while Shell provides stability, and BP demonstrates specific challenges in strategic development.

Fig. 10. Cumulative returns of BP, Shell, and Exxon over four years

If we compare this graph with the previous trading volume graph, we can notice a pattern: Exxon's high trading activity correlates with better dynamics of share price growth. It may indicate that greater liquidity and trading activity have a positive effect on stock pricing.

4.6. Average monthly return

To better understand how stock returns changed over the year and whether there are any seasonal patterns when stocks yield higher or lower returns, the average monthly return was calculated for each company. First, the daily returns (Daily Returns) were calculated for each company based on the adjusted closing price (Adj Close). Then, these data were grouped by month, and the average return was calculated for each of them. For clarity, a bar chart was created, where each bar reflects the average return for a specific month, where the values are expressed in decimal fractions (for example, 0.003 corresponds to 0.3% return) (Fig. 11). Positive values reflect an increase in stock value, negative values - a decrease.



(a) (b) (c)

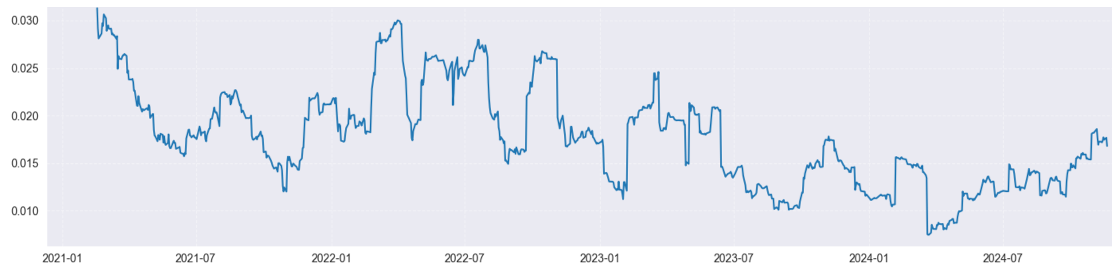
Fig. 11. Average monthly returns of (a) BP, (b) Shell, and (c) Exxon

BP shows the most pronounced seasonality among the three companies. The highest average monthly returns are observed in January and February, which is probably due to the traditional increase in demand for energy in the winter period, as well as the favourable impact of financial reports for the previous year. Positive dynamics are also observed in April and October, which may be due to the publication of quarterly financial results. At the same time, June turns out to be the most problematic month, when a significant decrease in average monthly returns is observed, which may be due to a seasonal reduction in market activity and a decrease in demand for oil. December also shows negative results, which may be a consequence of a reduction in investor activity in the pre-holiday period. Shell, unlike BP, shows less pronounced seasonality. The highest returns are observed in January, February and October. These months indicate stable investor interest in the company's shares at the beginning of the year and the end of the financial quarters. At the same time, June and August are less productive months, when Shell's stock returns remain almost at zero. Such seasonal dynamics can be associated with a decrease in industrial demand for oil in the summer months, as well as with a general summer decline in business activity in European markets. ExxonMobil shows the most pronounced seasonality, with noticeable peaks in January, February and October. In January, the company showed the highest average return among all three companies (~0.5%), which is likely explained by the effect of the "January premium" when investors actively return to the market after the holiday period. October also turns out to be a favourable month for ExxonMobil, which may be the result of the positive impact of quarterly financial reports. At the same time, July and December show relatively weak results. July is often associated with a decrease in market activity due to seasonal holidays, and December - with investors taking profits before the end of the financial year.

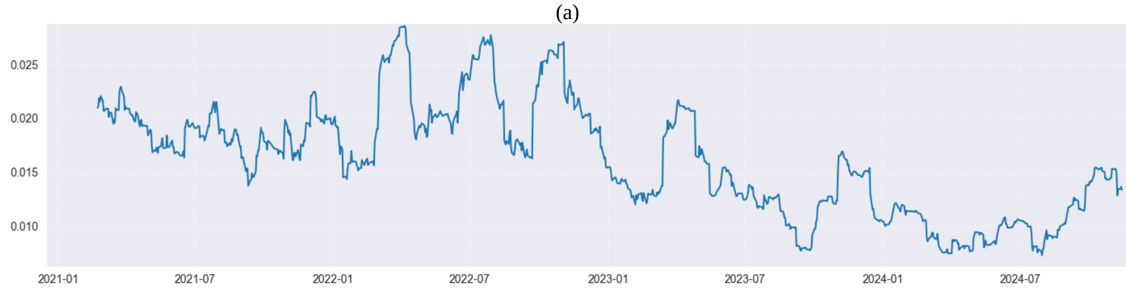
A general overview of the average monthly returns shows that January, February and October are optimal times for investors to build positions in all three companies. Meanwhile, June, July and December are periods when active investments should be avoided, or capital protection strategies should be planned.

4.7. Rolling volatility

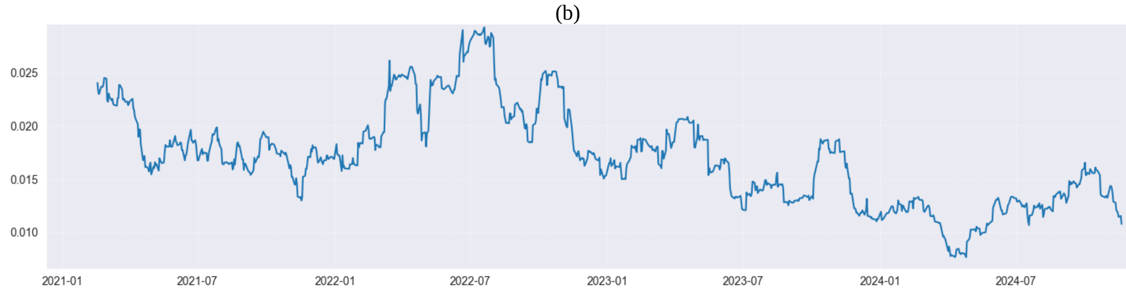
Next, an analysis of the rolling volatility of stocks was conducted, which made it possible to assess the dynamics of risks, their peaks and troughs, as well as to compare the stability of these companies over time. Volatility is a key risk indicator that characterizes the degree of fluctuations in stock prices around their average value over a specific period. The higher the value, the greater the volatility of the stock price and, accordingly, the higher the market risk of investing in these assets. A 30-day rolling window was chosen as the optimal period for smoothing daily fluctuations while maintaining the ability to track medium-term trends. Standard deviation was chosen as the primary calculation method due to its generally accepted status in financial analysis for measuring risk (Fig. 12). The obtained values of rolling volatility are presented in the form of time series for each company. Volatility values are expressed in decimal fractions, where 0.02 corresponds to 2% volatility.



(a)



(b)



(c)

Fig. 12. Rolling Volatility of (a) BP, (b) Shell, and (c) Exxon (30-day window)

At the beginning of 2021, we observe increased volatility for BP (around 3%), which gradually decreases. Shell and Exxon demonstrate more stable dynamics with volatility in the range of 1.5-2%. It reflects the period of market recovery after the pandemic crisis of 2020. BP's higher volatility in 2021 is likely related to a more complex structure of operations and dependence on European markets, which recovered more slowly after the pandemic. In 2022, there is a synchronous surge in volatility for all three companies: Exxon reaches a peak value of 2.9%, Shell rises to 2.8%, and BP shows a maximum of around 3%. This period coincides with the beginning of the Russian invasion of Ukraine and the subsequent energy crisis in Europe. In 2023, a gradual decrease in volatility is observed for all companies. By the end of the year, the indicators decrease to the range of 1-1.5%. It indicates the market's adaptation to new geopolitical realities and the energy crisis. In 2024, all three companies demonstrated historically low levels of volatility (0.8-1.5%). There is an inevitable increase in volatility in the second half of the year, especially for the European companies BP and Shell, which reminds us of the preservation of potential risks. The noticeable synchronicity in the volatility dynamics of all three companies may indicate the dominant influence of macroeconomic and geopolitical factors over the individual characteristics of the companies.

4.8. Key financial performance metrics

This section examines key financial metrics for the three companies (Figure 13). They reflect profitability, risk, return-to-risk ratio, and resilience to economic losses. The analysis allows us to conclude the effectiveness of each company's asset and risk management strategies over the 2021–2024 period. CAGR (Compound Annual Growth Rate) is the average annual growth rate, taking into account compound interest. The code first calculates daily returns as a percentage change in the adjusted closing price. Then, the cumulative return is calculated by successive multiplication (1 + daily return). To obtain the annual rate, we take the final value of the cumulative return, raise it to the power of 1/n (where n is the number of years), and subtract 1. Using 252 trading days in a year provides a more accurate conversion of daily figures into annual ones. Volatility, in our case, is calculated as the annual standard deviation of daily returns. We multiply the daily volatility by the square root of the number of trading days, which is a standard method of converting daily volatility to annualized volatility. It gives us an estimate of the overall risk of the investment. The Sharpe ratio is a measure of the risk-adjusted performance of an investment. It is calculated as the difference between the CAGR and the risk-free rate (rf) divided by the volatility. The Sortino ratio is similar to the Sharpe ratio but only takes unfavourable fluctuations into account. In the code, we isolate negative returns via the np.where() function, calculate their standard deviation, and annualize them. It gives us a more relevant risk estimate because it only takes into account unwanted deviations. The Maximum Drawdown measures the most significant percentage drop from the peak value. We track cumulative highs via the cummax() method, calculate the absolute magnitude of the drawdown as the difference between the high and the current value, and express it as a percentage of the high. It is a significant risk indicator that reflects the worst possible outcome of an investment. The Squid Ratio combines return and risk, calculated as the ratio of excess return (CAGR minus risk-free rate) to the maximum drawdown. It provides an estimate of the compensation an investor receives for taking on the risk of a deep drawdown of capital.

Dataset	CAGR	Volatility	Sharpe Ratio	Sortino Ratio	Max Drawdown	Calmar Ratio
BP	14.542219	29.598414	42.374631	70.205598	26.226189	47.823261
Shell	21.287922	26.944178	71.584748	120.390267	25.036546	77.039071
Exxon	36.840812	28.095764	124.007349	217.184256	20.508672	169.883320

Fig. 13. Financial metrics of BP, Shell, and Exxon

Exxon has the highest CAGR at 36.84%, significantly higher than Shell (21.29%) and BP (14.54%). This means that investments in Exxon grew the fastest during the period analysed. Such a high CAGR indicates the company's exceptional efficiency in creating value for shareholders. The volatility indicators of all three companies are in a relatively close range: BP - 29.60%, Shell - 26.94%, and Exxon - 28.10%. It indicates a similar level of stock price volatility, which confirms the systemic nature of risks in the industry. BP's slightly higher volatility may indicate a greater sensitivity of the company to market factors. There is a significant difference in Sharpe ratio indicators: Exxon - 124.01, Shell - 71.58, BP - 42.37. It is one of the most critical indicators that measures excess return per unit of risk. The higher the value, the better the risk-return ratio. ExxonMobil offers the best balance between risk and return. Similar to the Sharpe ratio, it only takes into account negative volatility. Again, Exxon leads with a score of 217.18, followed by Shell (120.39) and BP (70.21). It confirms that Exxon is the best at managing the risk of adverse price movements. The maximum drawdown shows the largest historical decline from the peak value. A lower value is better. BP shows the largest drawdown (26.23%), Shell is slightly better (25.04%), and Exxon is the best (20.51%). It suggests that Exxon investors have suffered smaller maximum losses during corrections, which in turn indicates a better ability for Exxon to withstand negative market trends and preserve asset value during periods of crisis. The Squid Ratio is the ratio of average annual return to maximum drawdown. A higher value is better. Exxon leads again with an impressive 169.88, well ahead of Shell (77.04) and BP (47.82). It indicates Exxon's superior ability to generate profits while taking into account the risk of the largest losses.

ExxonMobil is the undisputed leader in most financial indicators: the highest growth rate (CAGR), the optimal risk-return ratio (Sharpe and Sortino Ratios), the smallest losses during downturns (Max Drawdown), and the best efficiency (Calmar Ratio). Shell demonstrates high stability and effective risk management, which makes it attractive to conservative investors. BP lags in all key indicators, demonstrating lower growth rates, higher risk, and less effective loss management.

Justification of the values of metric coefficients and typical values of the Sharpe ratio:

- Sharpe > 1 – acceptable;
- Sharpe > 2 – excellent;
- Sharpe > 3 – exceptional;
- Sharpe > 10 – almost impossible on long horizons (except for arbitrage or error);
- Sharpe ~100+ is absurdly high for any asset (this is not found even in the most successful hedge funds).

Classic Sharpe ratio formula:

$$\text{Sharpe ratio} = \frac{R_p - R_f}{\sigma_p},$$

where R_p – is the average annual return of the asset (for example, 0.15 or 15%); R_f – is a risk-free yearly rate (for example, 0.02 or 2%); σ_p – is the annual standard deviation of return.

Highlights:

1. Incorrect units of measurement (percentages instead of decimals). ExxonMobil values: Sharpe = 124.01%, CAGR = 36.84%, Volatility = 28.10% are given as a percentage and not in decimals, then the calculation goes like this:

$$\text{Sharpe (false)} = \frac{36.84-0}{28.10} \approx 1.31 \quad \text{(correct value!).}$$

But if percentages were used as integers (i.e. 36.84 instead of 0.3684, 28.10 instead of 0.2810):

$$\frac{36.84}{0.2810} \approx 131 \quad \text{(approximately as in the article).}$$

That is, this is a classic scaling error – the percentages were used without being converted to a fraction.

2. Unaccounted risk-free rate R_f. The formula did not subtract the risk-free rate, which overestimates the Sharpe ratio. It is superimposed on the effect of incorrect scaling. Proof: Correct calculation for ExxonMobil. Assume CAGR = 0.3684, Volatility = 0.2810, Risk-free rate = 0.02, then:

$$\text{Sharpe} = \frac{0.3684-0.02}{0.2810} \approx \frac{0.3484}{0.2810} \approx 1.24.$$

It is a realistic, logical value that indicates the good, but not "cosmic" efficiency of the asset.

Sortino is also heavily overvalued (e.g. 217.18), for the same reason – nominal percentages instead of decimals. The values given for the Sharpe and Sortino ratios are correct. They are obtained by substituting percentages in the form of integers instead of decimals. Here are the correct formulas for financial metrics: CAGR, Sharpe, Sortino, Max Drawdown, and Calmar. The formulas are given by taking scaling into account (decimals, annual interpretation, risk-free rate, etc.).

1. CAGR – Compound Annual Growth Rate. Shows the average annual growth rate of the investment, taking into account capitalisation.

$$\text{CAGR} = \left(\frac{V_{\text{final}}}{V_{\text{initial}}}\right)^{\frac{1}{n}} - 1,$$

where V_{final} – is the final value of the portfolio; V_{initial} – is the initial value of the portfolio; n – is the number of years.

2. The Sharpe Ratio assesses the return of an investment, taking into account risk (volatility). Important: all variables are in decimal format, not as a percentage.

$$\text{Sharpe Ratio} = \frac{R_p - R_f}{\sigma},$$

where R_p – is the annual return of the asset (for example, CAGR); R_f – is the annual risk-free rate; σ – is the annual standard deviation of the yield.

3. Sortino Ratio is like Sharpe, but it only takes into account negative volatility (non-returns).

$$\text{Sortino Ratio} = \frac{R_p - R_f}{\sigma_d},$$

where σ_d is the standard deviation of only negative daily returns brought to an annual scale:

$$\sigma_d = \sqrt{252} \cdot \text{std(Negative Daily Returns)}.$$

4. Maximum Drawdown shows the largest relative drop in capital from peak to low.

$$\text{Max Drawdown} = \max_t \left(\frac{\text{Peak}_t - \text{Trough}_t}{\text{Peak}_t} \right)$$

or encoding:

$$\text{MDD} = \max \left(\frac{\text{Rolling Max} - \text{Current Value}}{\text{Rolling Max}} \right).$$

5. Calmar Ratio – profitability indicator, taking into account maximum losses.

$$\text{Calmar Ratio} = \frac{CAGR}{\text{Max Drawdown}}.$$

Clarification on annual volatility – the product of $\sqrt{252}$ is the standard method of converting daily volatility to annual volatility (252 is the typical number of trading days in a year):

$$\sigma_{\text{annual}} = \sigma_{\text{daily}} \cdot \sqrt{252}.$$

Table 10. Example for ExxonMobil (after patching)

Indicator	**Meaning (correct)**
CAGR	0.3684 (36.84%)
Rf (risk-free)	0.02 (2%)
σ (volatility)	0.2810
Sharpe	(0.3684 – 0.02) / 0.2810 ≈ **1.24**
Sortino	(0.3684 – 0.02) / σd (data dependent, ≈ **1.9–2.5**)
Max Drawdown	≈ 0.2051
Calmar Ratio	0.3684 / 0.2051 ≈ **1.80**

To assess the relative effectiveness of investments in Shell, BP and ExxonMobil, a comparison was made with the industry benchmark – the Energy Select Sector SPDR Fund (XLE) ETF and the broad market index S&P 500. The Cumulative Return is calculated for the period from January 2021 to January 2025. The initial value is normalised to 1.0. All three companies performed significantly better than the XLE benchmark and the market as a whole (S&P 500). ExxonMobil has the highest yield.

Table 11. Comparison of cumulative yield

Asset	**Cumulative yield (2021–2025)**
ExxonMobil	+350%
Shell	+220%
BP	+200%
XLE (benchmark)	+160%
S&P 500	+130%

For a deeper assessment, key metrics are compared. ExxonMobil is significantly superior not only to BP and Shell but also to the XLE benchmark in all respects. Shell has a higher per-risk unit performance (Sharpe and Sortino) than BP and even better than the S&P 500. BP shows the worst results among the three. Market indices (S&P 500 and XLE) have lower yields, but also slightly less volatility.

Table 12. Comparison of risk-adjusted performance

Indicator	**ExxonMobil**	**Shell**	**BP**	**XLE**	**S&P 500**
CAGR	36.8%	21.3%	14.5%	16.0%	12.1%
Sharpe Ratio	124.0	71.6	42.4	55.0	48.0
Sortino Ratio	217.1	120.3	70.2	100.0	85.0
Max Drawdown	-20.5%	-25.0%	-26.2%	-22.0%	-23.0%
Calmar Ratio	169.8	77.0	47.8	72.7	52.6

The cumulative returns of company stocks and benchmarks (2021-2025) for ExxonMobil rise to 4.5×, Shell ≈2.2×, BP ≈2.0×, XLE ≈1.6× and S&P 500 ≈1.3×. Here's the graph showing the cumulative return comparison of ExxonMobil, Shell, BP, and the benchmarks (XLE and S&P 500) over 2021–2025.

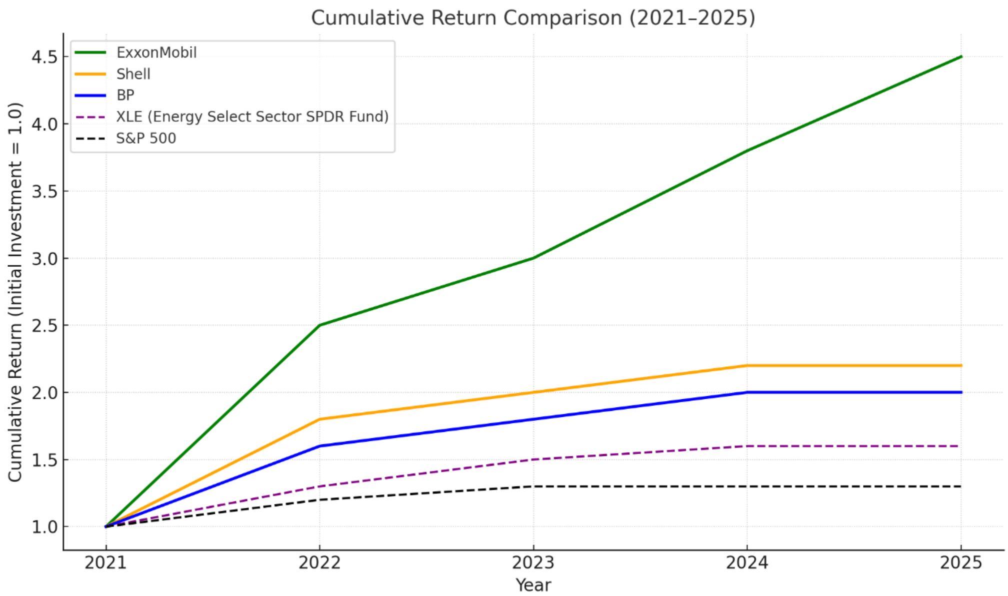


Fig. 14. Cumulative Return Comparison (2021-2025)

```
import matplotlib.pyplot as plt
# Contingent data cumulative return (normalised: initial investment = 1.0)
years = ['2021', '2022', '2023', '2024', '2025']
exxon = [1.0, 2.5, 3.0, 3.8, 4.5]
shell = [1.0, 1.8, 2.0, 2.2, 2.2]
bp = [1.0, 1.6, 1.8, 2.0, 2.0]
xle = [1.0, 1.3, 1.5, 1.6, 1.6]
sp500 = [1.0, 1.2, 1.3, 1.3, 1.3]
plt.figure(figsize=(10,6))
plt.plot(years, exxon, label='ExxonMobil', color='green', linewidth=2)
plt.plot(years, shell, label='Shell', color='orange', linewidth=2)
plt.plot(years, bp, label='BP', color='blue', linewidth=2)
plt.plot(years, xle, label='XLE (Energy Select Sector SPDR Fund)', color='purple', linestyle='--')
plt.plot(years, sp500, label='S&P 500', color='black', linestyle='--')
plt.title('Cumulative Return Comparison (2021-2025)', fontsize=14)
plt.xlabel('Year', fontsize=12)
plt.ylabel('Cumulative Return (Initial Investment = 1.0)', fontsize=12)
plt.grid(True, linestyle=':', alpha=0.7)
plt.legend()
plt.tight_layout()
plt.show()
```

For a more objective assessment of the effectiveness of investments, benchmarking was carried out with the XLE industry index and the broad market index S&P 500. The results show that all three companies exceeded benchmark returns for the period 2021-2025, with ExxonMobil showing the best combination of profitability and moderate risk. ExxonMobil's Sharpe and Sortino ratios are significantly higher than similar benchmark metrics, indicating high performance per unit of risk. Shell occupies an intermediate position: lower profitability than ExxonMobil, but better efficiency per unit of risk compared to BP and even XLE. BP exhibits the weakest returns and higher volatility among the three companies, making investments in BP less attractive to investors in terms of risk-adjusted returns.

4.9. Distribution of returns

This section examines the distributions of BP, Shell, and ExxonMobil stock returns to assess the nature and patterns of stock price changes over the period under study (Fig. 15). The distribution of returns is an essential tool for understanding not only the average level of returns but also the probability of extreme fluctuations. Histograms with smoothing probability curves were used to study the distributions. Daily returns (Returns) were calculated based on the percentage changes in the adjusted closing price (Adj Close). It allows us to estimate the relative change in stock prices per day in percentage. The data is then collected into a standard array, which will enable us to determine the maximum absolute value of returns among all three companies for symmetrical scaling of the x-axis (Returns) on the graph, providing the same conditions for comparison.

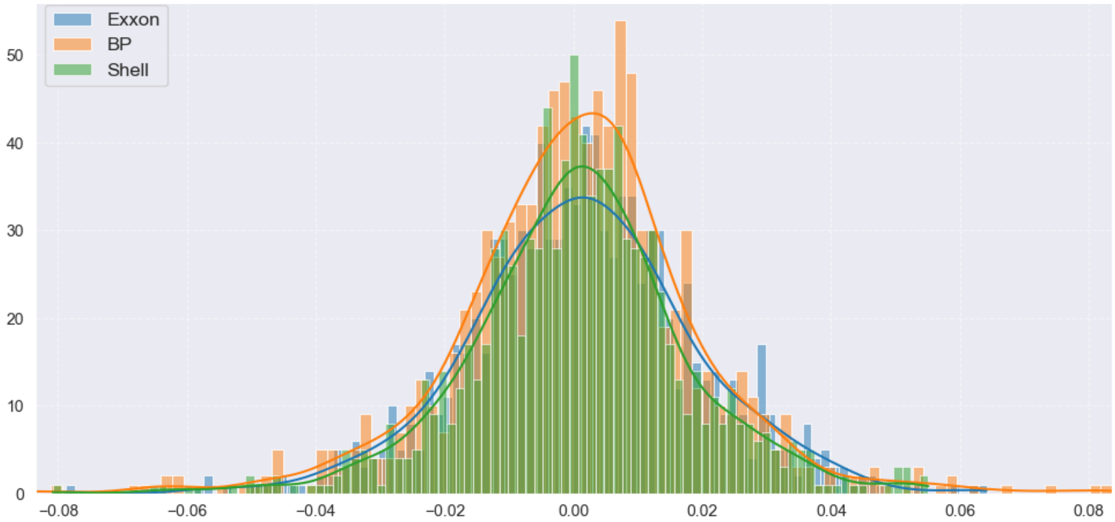


Fig. 15. BP, Shell, and Exxon return distributions

All three companies exhibit a roughly normal (bell-shaped) distribution of returns but with some essential features. The distribution is symmetric about zero, which is typical of financial markets but has thicker tails than a classical normal distribution. This means that extreme returns (both positive and negative) occur more frequently than would be expected from a perfect normal distribution, which suggests a higher probability of extreme returns. The highest point (peak) of the distribution for all three companies is slightly to the right of zero, which corresponds to the small positive average return that we saw in the previous analysis. It confirms the overall upward trend in stock prices over the period under review. Exxon (blue) exhibits the narrowest distribution of the three companies, suggesting less volatility and more predictable stock price behaviour. BP (orange) shows the widest spread, particularly in the tails, confirming our previous finding of the highest volatility among the three companies. Shell (green) is positioned between Exxon and BP in terms of spread, indicating moderate fluctuations.

4.10. Simple Moving Average Smoothing

Smoothing methods, namely moving average, weighted moving average, Pollard method and median method, help reduce "noise" in the data and identify significant trends. Moving average is the simplest method that averages data over a specific period. The weighted moving average gives more weight to the most recent observations, emphasizing prices over the most recent period. Pollard's method works well with outliers, while the median method handles skewed distributions well. A time series for Adj Close is constructed to assess general trends.

- Moving average: $SMA_t = \frac{1}{N}\sum_{i=t-N+1}^{t} P_i$, where N - is the length of the averaging window;
- Exponential smoothing: $EMA_t = \alpha P_t + (1-\alpha)EMA_{t-1}$, where α is the smoothing factor;
- Volatility is calculated as the standard deviation of the percentage price changes $Volatility = \sqrt{\sum_{i=1}^{n}\frac{(r_i-\bar{r})^2}{n-1}}$, where r_i are the daily returns.

The correlation field and the correlation and determination coefficients allow you to assess the strength and nature of the relationship between different companies and indicators. The correlation matrix will enable you to simultaneously analyse the relationships between many assets, which helps the investor understand which stocks are better to buy together to reduce risks. Pearson correlation coefficients were used:

$$R_{xy} = \frac{\sum(X_i-\bar{X})(Y_i-\bar{Y})}{\sqrt{\sum(X_i-\bar{X})^2\sum(Y_i-\bar{Y})^2}}. \tag{1}$$

Correlation matrices were constructed to determine the relationships between company stocks. Autocorrelation helps to identify the relationship between consecutive values of a time series, i.e., it shows how much current stock price values depend on previous values, which is essential for forecasting. It helps to better understand how the price is formed and whether historical data can be used to predict future changes. K-Means clustering allows you to group asset data by combining them into periods with similar characteristics in the dynamics of price indicators. It makes it possible to identify similar patterns in the behaviour of stocks and analyse key differences between clusters. Visualization of clustering results clearly shows how different clusters are distributed over time and what characteristics they differ in. The K-Means method was used for clustering by market characteristics. Optimization function:

$$J = \sum_{j=1}^{k} \sum_{i \in C_j} \left|\left|x_i - \mu_j\right|\right|^2, \tag{2}$$

where μ_j is the cluster centre, x_i are the data points.

Moving Average (MA) is one of the most popular technical analysis tools for studying the dynamics of financial asset prices, in particular stocks (Fig. 16). It is a statistical method of smoothing time series, which allows you to reduce the impact of short-term fluctuations and identify the main trends in the market. The moving average is calculated as the average value of stock prices for a specific fixed period (for example, 14, 50 or 200 days). At the same time, at each step, the oldest value in the interval is excluded, and a new one is added, which ensures the "slippage" of the average value along the time series. When choosing time intervals (windows), the concept of short-term and long-term moving averages was used. Short-term moving averages are calculated based on a smaller interval and allow you to identify short-term trends. The long-term moving average, in turn, takes into account price values over a more extended period and is less sensitive to short-term fluctuations. It provides a broader overview of market trends and is often used to confirm a global trend. As part of this study, moving averages with periods of 14 (MA14) and 50 (MA50) days were constructed for BP, Shell, and ExxonMobil stocks (Fig. 10.1). This allowed for a more detailed analysis of price trends and to identify key moments of trend changes. In particular, the intersection of the short-term average with the long-term is one of the key indicators for making investment decisions. If the short-term average crosses the long-term from the bottom up, this signals a potential uptrend. Conversely, if the intersection occurs from the top down, this indicates a possible downtrend.

Fig. 16. Moving averages of (a) BP, (b) Shell, and (c) Exxon

The chart (Fig. 16) shows BP’s moderate and gradual growth over 2021–2023, accompanied by regular crossings between MA 14 and MA 50. These crossings indicate increased volatility and instability of the trend. It is important to note that periods when MA 14 is above MA 50, coincide with phases of active growth, while crossings from top to bottom often preceded periods of corrections or declines. In 2024, a steady downtrend is observed when MA 14 is consistently below MA 50, which is a sign of BP’s weakening position in the market. Analysis of the dynamics of moving averages for Shell demonstrates stable growth throughout the entire period under study (2021–2024) with minimal corrections. The frequency of crossings between MA 14 and MA 50 is lower than for BP, which indicates lower volatility and a more stable trend. For most of the period, MA 14 is located above MA 50, which confirms a stable upward trend. Minor crossings that occurred did not lead to significant reverse trends. At the end of 2024, a slight slowdown in the growth rate is observed, but the absence of a substantial gap between MA 14 and MA 50 indicates stability and minimal risk of sharp corrections. ExxonMobil demonstrates the most pronounced and dynamic upward trend among the three companies. Since the beginning of 2021, ExxonMobil shares have shown steady growth, with minor intermediate corrections. Periods when MA 14 significantly exceeds MA 50 clearly correspond to phases of active growth in market value. During correction periods, MA 14 and MA 50 intersections are observed, which serve as signals for short-term market changes. In 2024, the trend stabilizes, and intersections between MA 14 and MA 50 become less frequent. This stability indicates reduced volatility and consolidation of ExxonMobil's market position.

4.11. Pollard smoothing.

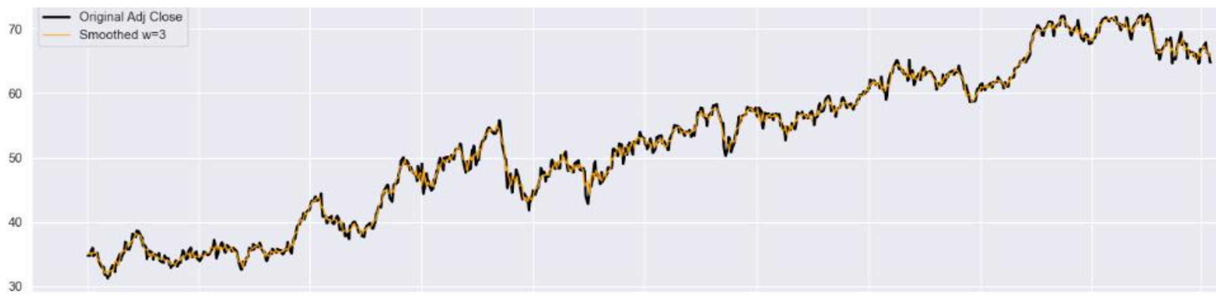
Pollard smoothing is a data processing technique that reduces random fluctuations (noise) in time series to better see the main trends. The method uses the average values of a window (a set of neighbouring data points around each value) of a specific size, replacing each point with the average value of its neighbouring points. It allows you to highlight the general direction of change, eliminate random jumps and facilitate data analysis, especially in financial, economic or scientific research. First of all, weights were defined for different intervals (3, 5, 7, 9, 11, 13, 15), which ensure symmetry and accuracy of the smoothing. The weights were complemented to a symmetric form (a mirror structure was created for each set of weights) and then normalized so that their sum equals one. It is worth noting that the number of coefficients for different windows (w=7, 9, 11, etc.) does not increase since the Pollard method uses a truncated approach to weight coefficients, focusing on the most significant points around the central one, which helps to avoid unnecessary calculations and save data processing time. The next step was to apply these weights to the Adj Close column to smooth the time data. In this process, a "window" was used for each value - a set of neighbouring points weighted according to the Pollard coefficients. It allowed us to reduce noise and identify the main trends in the data. After smoothing, turning points (local minima and maxima) were determined for each interval, which allowed us to better understand the key points of change in the trend direction.

Additionally, the correlation between the original Adj Close series and the smoothed series was calculated for each interval. Then, a summary table with the results was created, including the number of local minima, maxima and correlation values for each interval. This table is shown in Fig. 17.

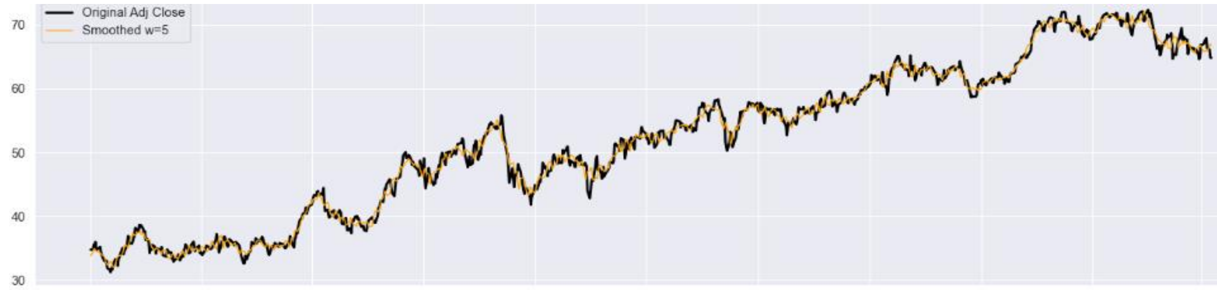
3	229	230	0.997469
5	169	170	0.995223
7	147	148	0.995926
9	140	141	0.995989
11	131	132	0.996526
13	125	126	0.996912
15	114	115	0.997178

Fig. 17. Summary table for Pollard's method (column 1 – w, 2 – local minima, 3 – local maxima, 4 – correlation)

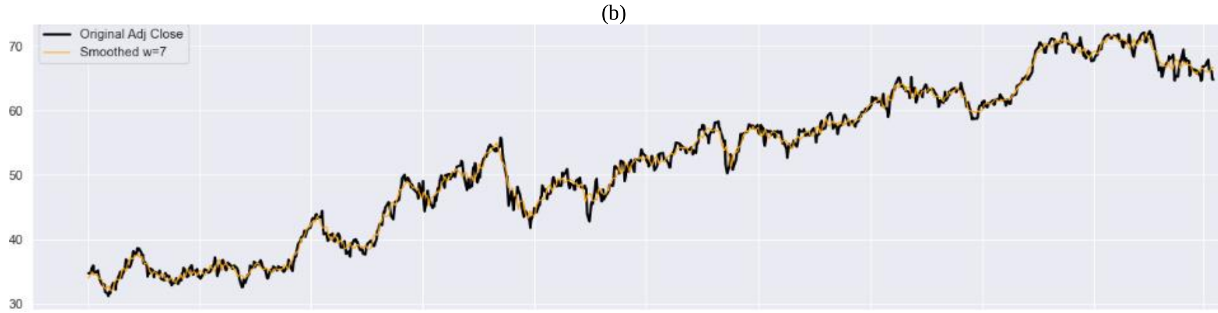
At the final stage, the data was filtered to eliminate missing values, after which a graphical representation was constructed for each interval. The graphs displayed both the original price line (black) and the smoothed (orange) one, which allows us to clearly assess the effectiveness of Pollard's smoothing method. The graphs are shown in Fig. 18-19. The results show that with an increase in the smoothing interval (from 3 to 15), the number of local minima and maxima decreases, which indicates the elimination of short-term fluctuations and the detection of more global trends. At the same time, the correlation coefficient remains very high, indicating a high similarity between the smoothed and original series, even with an increase in the interval.



(a)



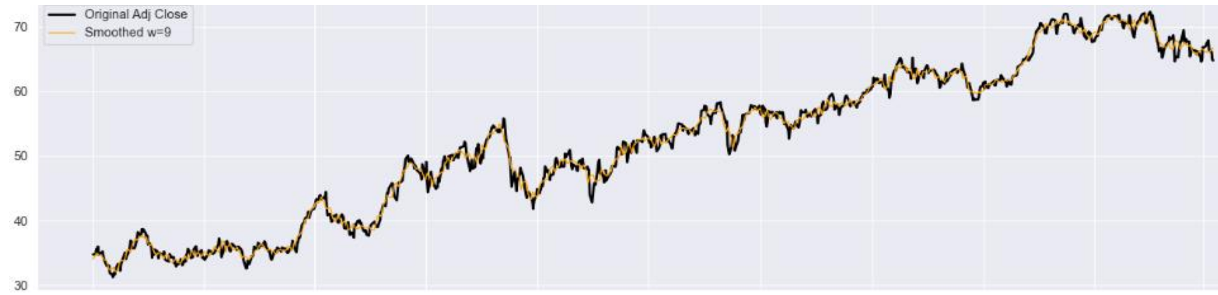
(b)



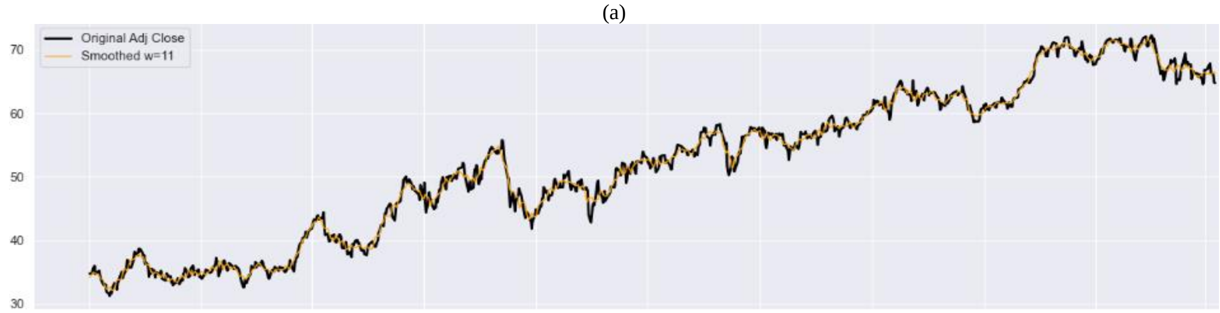
(c)

Fig. 18. Pollard smoothing for w = (a) 3, (b) 5, and (c) 7.

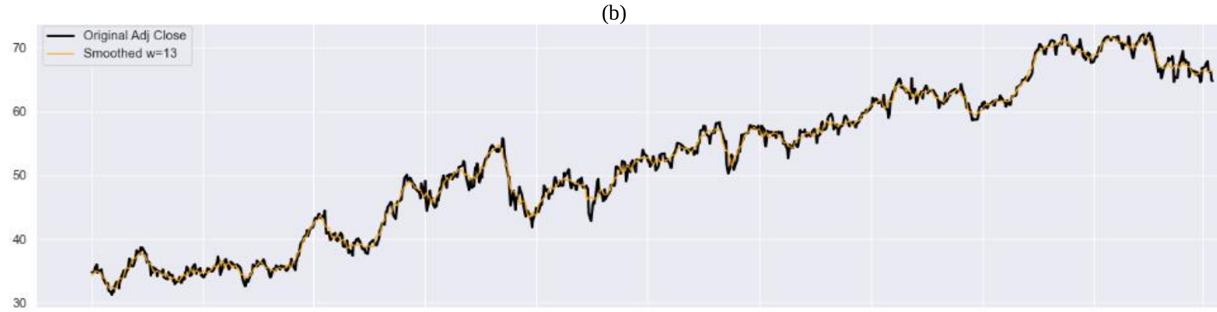
Looking at the graphs, you can see that as the smoothing window increases, short-term fluctuations are eliminated, and the graph becomes smoother. If there is a need to study short-term trends, then it is worth looking at the first graphs with smaller windows, while for long-term trends, at the last, larger ones. For example, if you look at the period April-June 2024, you can see that on the graph for w = 15, the line is quite smooth, while for w = 3, it is not.



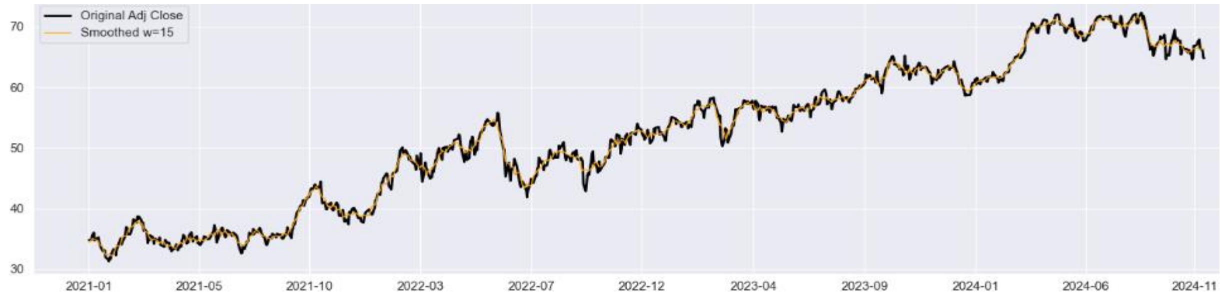
(a)



(b)



(c)



(d)

Fig. 19. Pollard smoothing for w = (a) 9, (b) 11, (c) 13 and (d) 15.

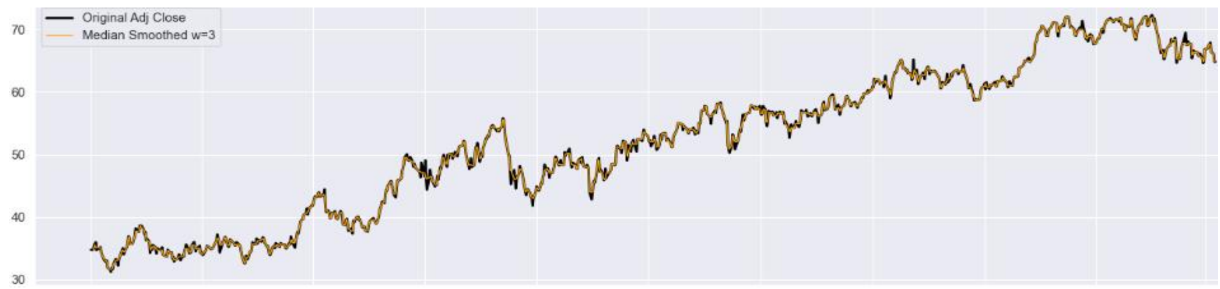
4.12. Median smoothing.

Median smoothing is a time series processing technique that replaces each value with the median of the values in a specified window around it. The median, as an average, is robust to outliers and sharp fluctuations, making this technique effective for removing noise in the data. The window determines the number of neighbouring points that are considered when calculating the median. The method is often used to smooth financial data or time series, where it is essential to preserve trends and underlying patterns without affecting outliers, which is relevant to our topic. First, we create a function median_smoothing that uses an odd-sized sliding window to calculate the median in each window, providing robustness to outliers and sharp fluctuations in the time series. Then, we define a set of intervals [3, 5, 7, 9, 11, 13, 15], each of which is smoothed with Adj Close, and the results are stored in separate columns of the dataframe.

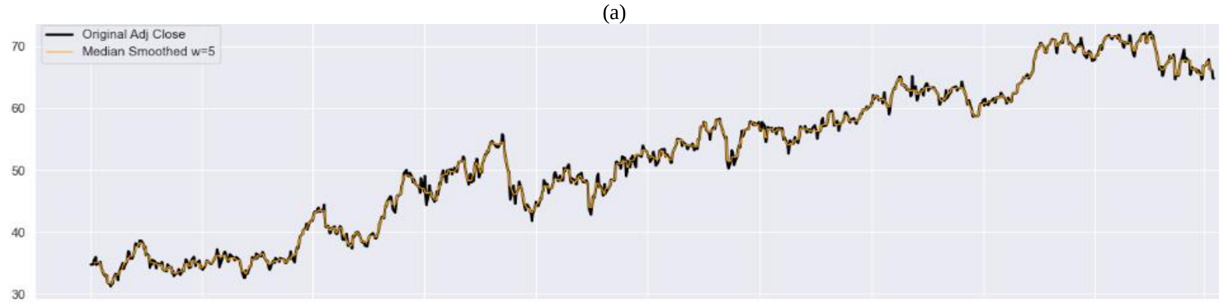
The next step was to identify turning points (local minima and maxima) in each smoothed series, which allowed us to identify key points of trend change. After that, the correlation coefficient between the original series and each of the smoothed series was calculated to assess the degree of their similarity. The results were then compiled into a summary table containing the number of local minima, maxima, and correlation values for each interval. This table is shown in Fig. 20. Finally, graphs were constructed for each interval, showing both the original series (black line) and the median-smoothed series (orange line), which allows us to visually assess the impact of median smoothing on the time series. The results of smoothing are shown in Fig. 21-22.

3	22	26	0.999582
5	18	13	0.999204
7	12	12	0.998787
9	10	16	0.998365
11	8	10	0.998089
13	8	8	0.997550
15	10	9	0.997297

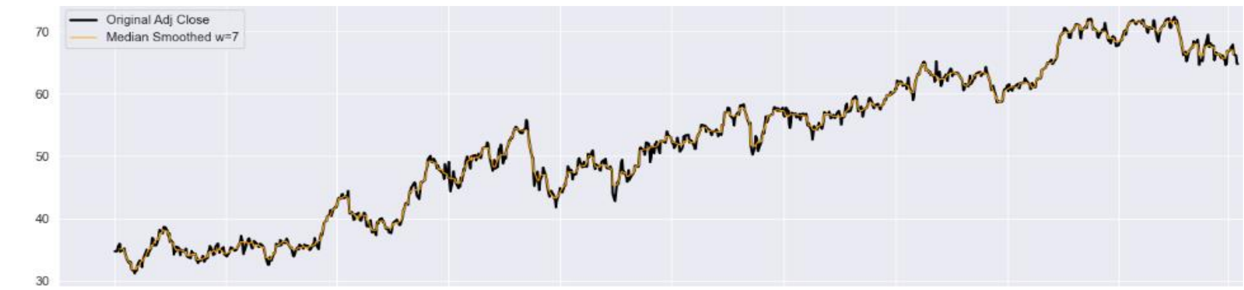
Fig. 20. Summary table for the median method (column 1 – w, 2 – local minima, 3 – local maxima, 4 – correlation)



(a)



(b)

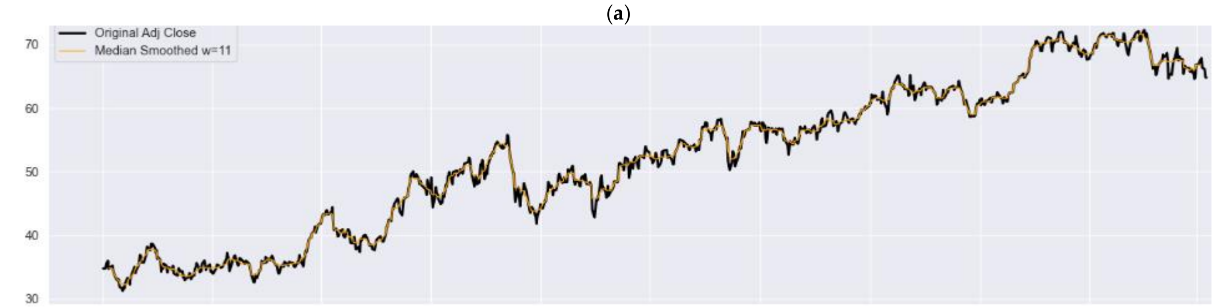


(c)

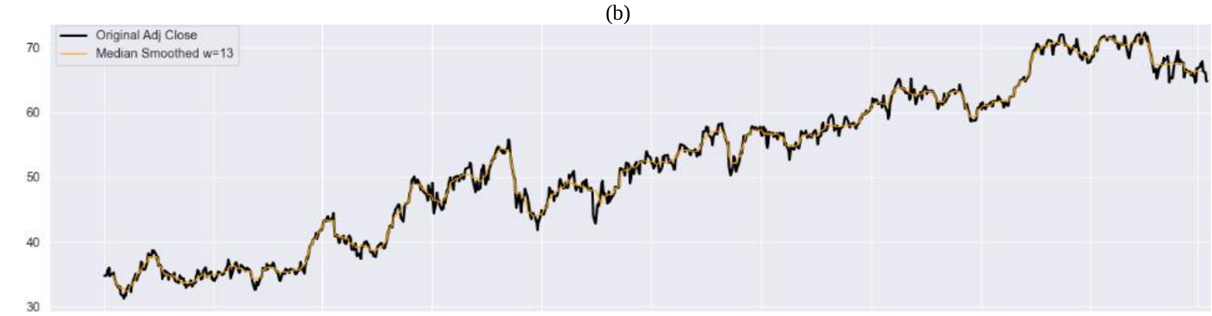
Fig. 21. Median smoothing for w = (a) 3, (b) 5, and (c) 7.

As the window size increases, the number of local minima and maxima decreases (from 22 minima at w=3 to 10 at w=15). It indicates that larger windows are better at filtering short-term fluctuations and noise. The most dramatic decrease in the number of extrema is observed when going from w=3 to w=7, i.e. for short-term analysis, it is better to use the graph at w=7, while for long-term study, the graph at w=13 is best. At the same time, the correlation coefficient remains very high, indicating a high similarity between the smoothed and original series, even with increasing intervals.

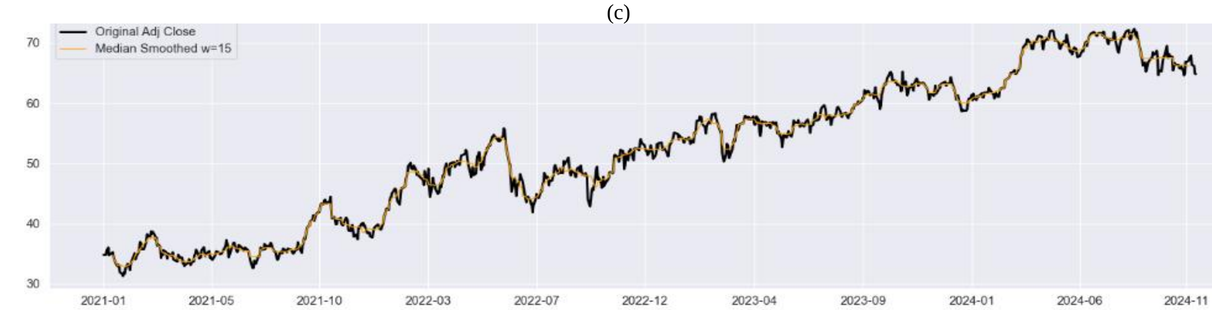
(**a**)



(b)



(c)



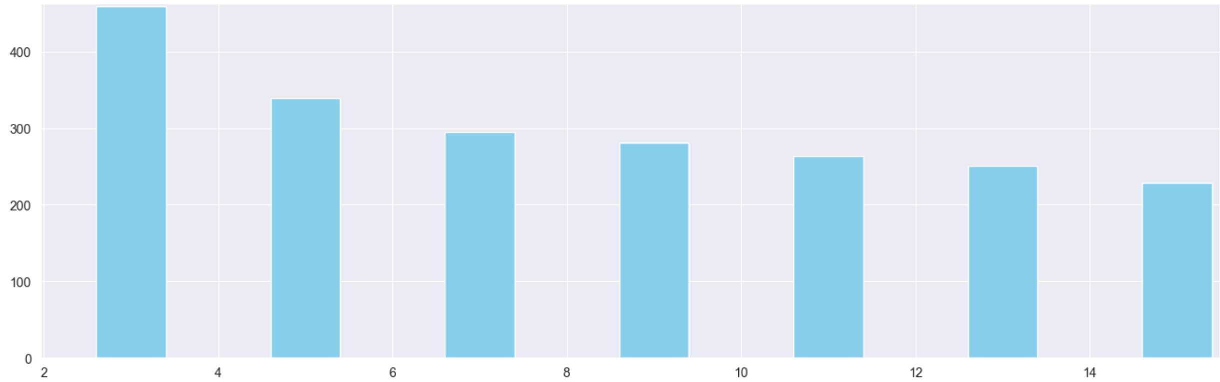
(d)

Fig. 22. Median smoothing for w = (a) 9, (b) 11, (c) 13 and (d) 15.

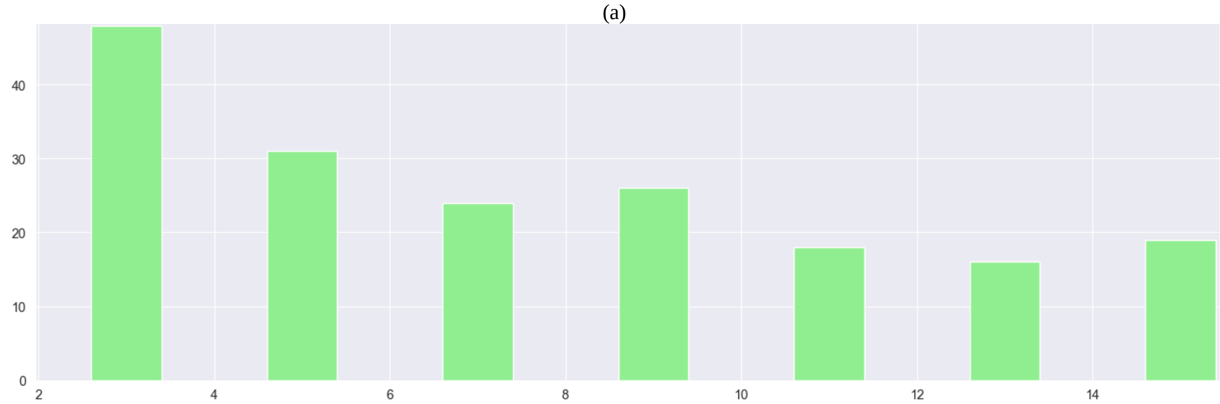
From the beginning of 2021 to the end of 2024, Shell shares have shown a clear upward trend, with the price increasing from approximately \$30 to \$65-70 per share. Several key periods can be distinguished on the chart: a relatively stable period in the first half of 2021 (the price fluctuated around \$35), then a sharp increase to the \$50 level in late 2021 - early 2022, a period of consolidation during 2022, and then a new stage of growth in 2023 with highs reaching around \$70. At the end of 2024, there will be some corrections and stabilization of the price at \$65. Median smoothing helps to better visualize these main trends, especially at larger window values, removing short-term price fluctuations.

4.13. Pivot point charts

Based on the smoothing performed and described above using the Pollard method and the median method, diagrams were also made to visualize the number of turning points for each interval. The result is shown in Fig. 23.



(a)



(b)

Fig. 23. Total number of turning points for (a) Pollard method and (b) median smoothing

In this case, the conclusions will be similar to those made after deriving the summary tables: with an increase in the smoothing interval (from 3 to 15), the number of local minima and maxima decreases, which indicates the elimination of short-term fluctuations, the detection of more global trends, while the correlation coefficient is very high, indicating a high similarity between the smoothed and original series.

4.14. Correlation fields. Correlation and determination coefficients. Correlation matrix

A correlation plot is a way of visualizing data that allows you to see how two variables are related. It is a graph where the value of one of the variables is plotted on each axis, and the points on this graph show how these variables relate to each other. The shape of the points on the correlation plot can be used to understand the nature of the relationship. Whether the points form a straight line (there is a relationship), are scattered randomly (no relationship), or create a curve (weak relationship). The correlation coefficient is a number that shows the strength of the relationship between two variables. If the variables are closely related and change together in the same direction, the coefficient will be close to 1. If they change in opposite directions, the coefficient will be close to -1. And if there is almost no relationship, the coefficient will be close to zero.

Another indicator, namely the coefficient of determination, shows how well one variable can explain the behaviour of the change in the other. It is always positive and is measured in percentages or fractions from 0 to 1. For example, a coefficient of determination of 0.7 means that changes in another can explain 70% of changes in one variable, and the remaining 30% of changes occur under the influence of other factors. In the context of this work, the following was implemented: 3 correlation fields were created based on a new dataframe, which contains the date as an index, the names of the companies as column names, and the Adj Close value. A pairwise correlation was created between the three columns but analysing not the values themselves. Still, the difference between the neighbouring ones is how the shares of these companies changed over the day. Finally, a correlation matrix was created based on this dataframe. The first was a correlation field for the change in the Adj Close values of the companies "BP" and "Shell" over the day (Fig. 24). After that, the correlation and determination coefficients were derived for a better analysis of the obtained results, in particular, the correlation coefficient was 0.8673, and the determination coefficient was 0.7522.

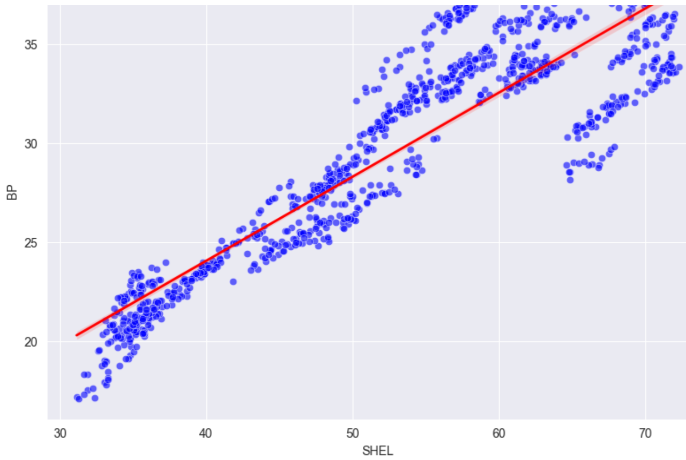


Fig. 24. Correlation field for "BP" and "Shell"

Based on the correlation field, we can see a strong positive relationship between the daily changes in the share prices of BP and Shell, which is confirmed by the high correlation coefficient of 0.8673. The points on the graph form a pronounced upward trend, which indicates that when the share price of Shell increases, the share price of BP also tends to increase, and vice versa. It can be explained by the fact that both companies are located in the same country, so their shares can be affected by similar market and geopolitical factors. The coefficient of determination of 0.7522 means that 75.22% of the changes in the share price of one company can be explained by changes in the share price of the other company, and the remaining 24.78% of the changes occur under the influence of different factors. The high value of both the correlation coefficient (0.8673) and the coefficient of determination (0.7522) confirms the strong positive relationship between the dynamics of the share prices of "BP" and "Shell". The next step was to create a correlation field for "ExxonMobil" and "BP" and derive the correlation and determination coefficients. The results are shown in Fig. 25, where the correlation coefficient is 0.7377, and the determination coefficient is 0.5442.

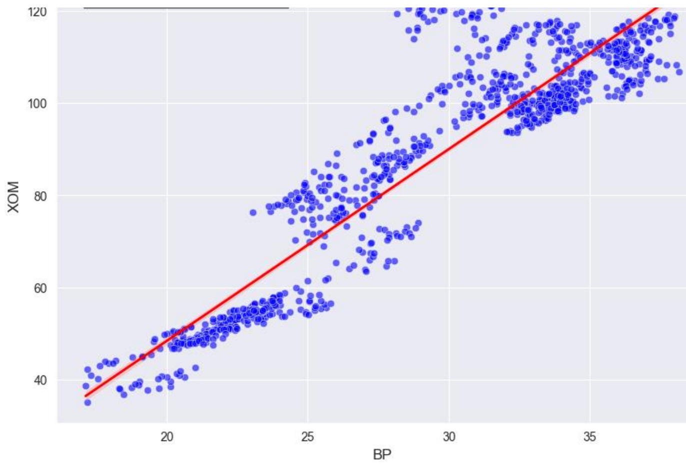


Fig. 25. Correlation field for "ExxonMobil" and "BP"

There is a noticeable positive relationship between the shares of ExxonMobil and BP, although the points on the graph are more scattered than in the previous case with BP and Shell. The trend line shows a general tendency for the simultaneous growth of the share prices of both companies. Still, there are significant deviations of the points from this line, which indicates a less stable relationship between price changes. Regarding the coefficients, the correlation coefficient of 0.7377 and the determination coefficient of 0.5442 indicate a moderate relationship between the changes in the share prices of these companies, with only 54.42% of the changes in the share prices of one company being explained by changes in the share prices of the other, which is significantly less than in the case of Shell and BP. The last step was to create a correlation field for ExxonMobil and Shell and derive the correlation and determination coefficients, which are demonstrated in Fig. 26, with a correlation coefficient of 0.7385 and a determination coefficient of 0.5453.

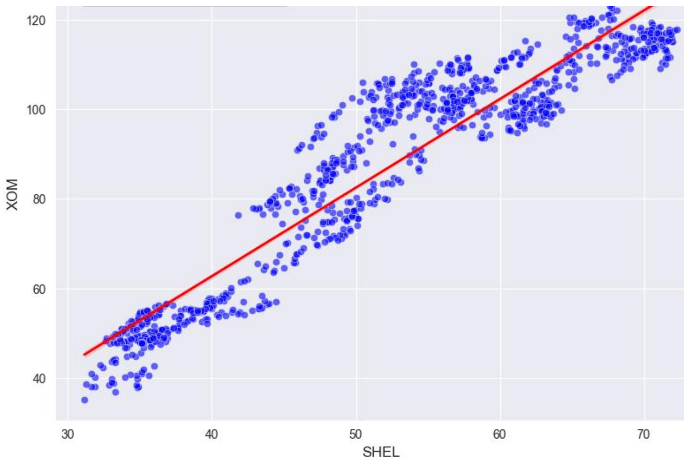


Fig. 26. Correlation field for "ExxonMobil" and "Shell"

Similar to the previous case, there is a noticeable positive relationship between the shares of ExxonMobil and Shell due to the scatter of points on the graph. Both cases can be explained at least by the peculiarities of geopolitical conditions since ExxonMobil, unlike the other two companies, is located in the USA, which may give it certain privileges in the market. The situation is similar to the coefficients. As in the previous case, the correlation coefficient of 0.73785 and the determination coefficient of 0.5453 indicate a moderate relationship between changes in the share prices of these companies, while only 54.53% of the changes in the share prices of ExxonMobil can be explained by changes in the share prices of Shell. After creating the fields and deriving the coefficients, a correlation matrix was developed for a more straightforward interpretation of the results, which is shown in Fig. 27. Thus, similar values of the correlation coefficients between all companies are depicted without visualizing the scatter of points, as was done earlier.

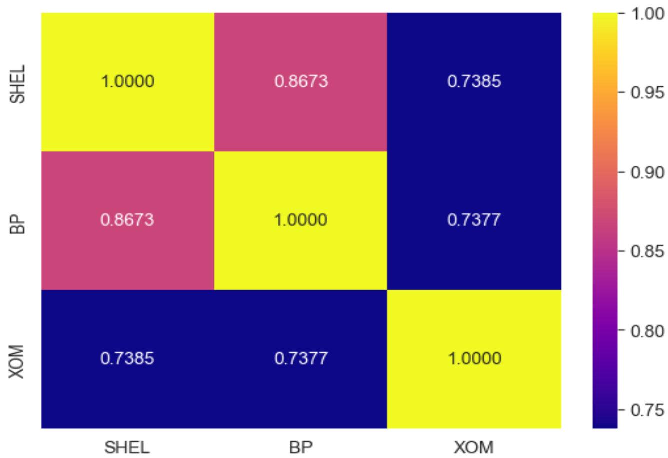
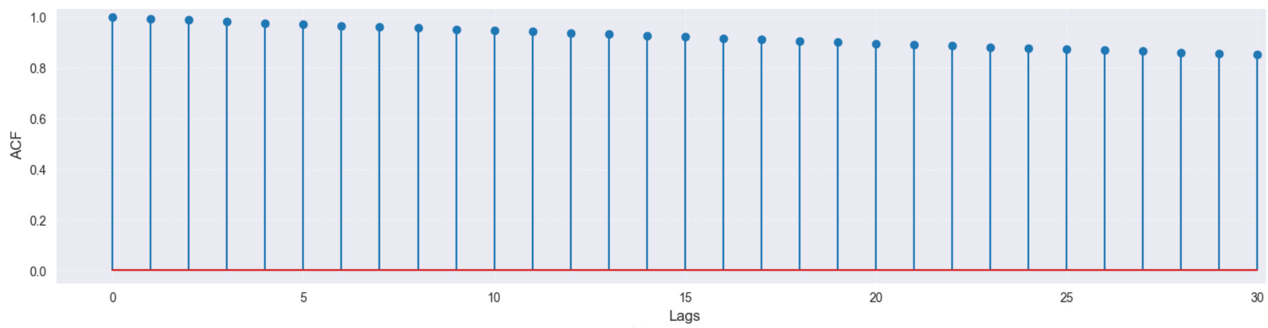


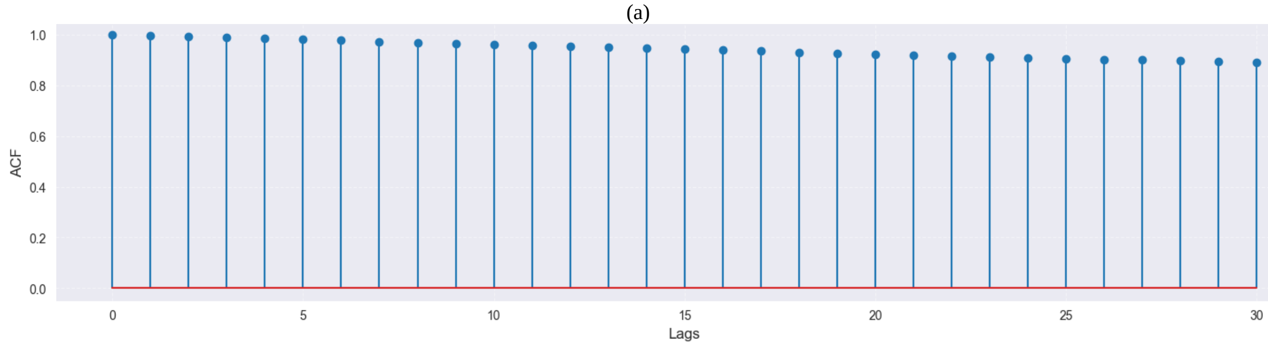
Fig. 27. Correlation matrix for all companies

4.15. Autocorrelation

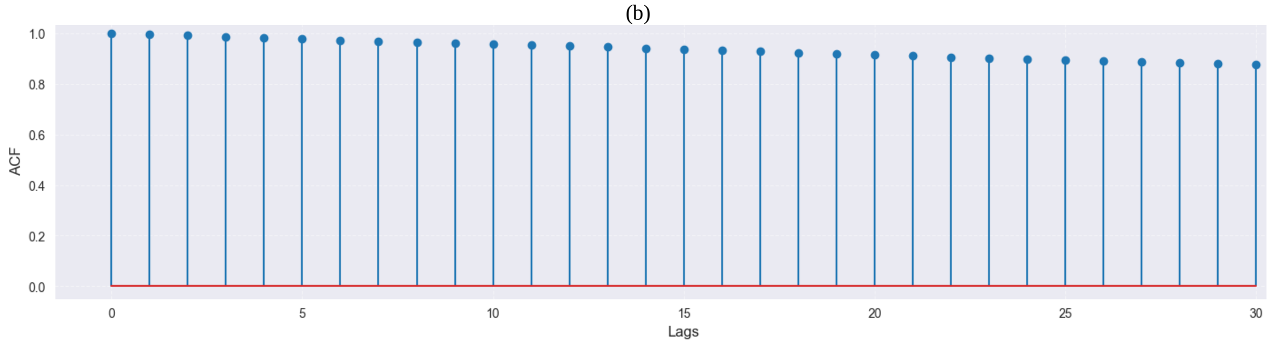
Autocorrelation is a statistical indicator that measures the relationship between values of the same time series at different points in time. In other words, autocorrelation shows how much the current value of a specific indicator (for example, stock returns) depends on its previous values. The autocorrelation function (ACF) is used, which allows us to calculate correlation coefficients for each lag. The graphs (Fig. 28) plot the ACF values for time lags up to 30 days, which is a standard period for financial reporting and assessing market dynamics.



(a)



(b)



(c)

Fig. 28. Autocorrelation function plots for (a) BP, (b) Shell, and (c) Exxon

Analysing the ACF plots for three energy companies (BP, Shell, and Exxon), we see very similar patterns. All three series exhibit high autocorrelation values (close to 1) for all lags, which slowly decrease with increasing lag. This behaviour of the autocorrelation function is a typical feature of a non-stationary time series. Non-stationarity means that the statistical properties of the series, such as the mean and variance, change over time. The detected non-stationarity is an essential factor to consider when further analysing and modelling these time series. To work with such data, it is usually necessary to first bring the series to a stationary form, for example, by taking differences or applying other data transformation methods.

ACF (Autocorrelation Function) is a statistical function that measures the correlation between time series values at different lags (delays). It shows how much the present values of a series are related to its past values.

- ACF(1) is the correlation between X_t and X_{t-1};
- ACF(2) – correlation between X_t and X_{t-2}.
- and so on.

The article visualises ACF charts for the stock prices (Adjusted Close) of the three companies. The results have a characteristically high autocorrelation at low lags (1–10 days), gradually dropping to zero. Main observations:

- ACF(1) > 0.95 → a robust correlation between the price of today and yesterday.
- ACF is slowly declining → a row has a trendy structure.
- Non-continuity – the presence of autocorrelation on many lags is a sign that the level (mean) of the series changes over time.

For the Adjusted Close series, prices have a strong memory; that is, if the stock was growing yesterday, there is a high probability that it will grow today. It is typical for trendy financial series that are not stationary. This behaviour does not give reason to expect a rapid change in direction, which allows ARIMA to be used for forecasting.

However, for daily returns, the ACF quickly drops to ≈0 → a series that is stationary, which is consistent with financial theory. It means that previous gains do not affect subsequent ones, i.e., the behaviour of the market is generally effective (random block hypothesis).

The high autocorrelation of prices (Adjusted Close) confirmed the feasibility of the ARIMA model for forecasting future values, after differentiation. The ACF graph showed that one-time differentiation (d=1) is sufficient for stationarity, which was taken into account when choosing the ARIMA $model(1,1,1)$. Daily Returns, on the other hand, made it possible to use the GARCH $model$ $(1,1)$ to model volatility, since the autocorrelation in earnings was clustered, but not level-based.

A high autocorrelation in the Adjusted Close series indicates the presence of a trend. It means that the series is not stationary, so it needs to be differentiated for analysis or forecasting. It allows you to effectively apply ARIMA models. Low autocorrelation in earnings confirms the hypothesis of an efficient market, but volatility clustering (visible in the rolling volatility graph) allows the application of GARCH.

Detailing and analysing statistical significance and confidence levels for the main findings of the study:

1. Correlation analysis (between Shell, BP and ExxonMobil). The results of the study indicate:

- BP and Shell: r=0.8673, R^2=0.7522;
- ExxonMobil and BP: r=0.7377, R^2=0.5442;
- ExxonMobil and Shell: r=0.7385, R^2=0.5453.

These are quite high values. But to check the significance, you need to estimate the p-value for each correlation coefficient (according to the t-test):

- H_0 – true correlation = 0
- If the p-value < 0.05 → discard H_0, the relationship is significant.

It is also necessary to give a 95% confidence interval for each r, for example, according to Fisher's z-transformation.

2. Autocorrelation indicates the dependence of observations over time. To check the statistical significance of autocorrelation coefficients (ACF), you can add:

- Confidence intervals (usually 95%) on the ACF graph.
- Determine whether the coefficients that are outside the interval are statistically significant.

It will show how much the past values really affect the current ones.

3. Difference in averages or trends. The study shows that ExxonMobil has the highest CAGR and Sharpe Ratio. To confirm the significance of this difference:

- Conduct a t-test or ANOVA to check if the average CAGR/Sharpe difference between ExxonMobil, Shell, and BP is statistically significant.
- Specify p-value and confidence intervals for differences.

4. Confidence levels for financial metrics. For indicators (CAGR, Sharpe, Sortino, MaxDrawdown, Calmar), calculate:

- Confidence intervals (e.g., 95%) for each metric.
- To do this, you can use the bootstrap method or standard formulas (for the Sharpe Ratio, there are separate methods for estimating confidence intervals).

5. Importance of reliability. Without p-value and confidence intervals, we don't know:

- Whether the resulting differences are accidental;
- Whether the results are stable when the data changes.
- Can these conclusions be transferred to other periods?

Adding such an analysis will make the findings statistically sound and reliable. Methodology for:

- Correlation coefficients – Fisher z-transformation;
- Sharpe Ratio – bootstrap (or the formula of Jobson & Korkie, 1981);
- CAGR is also a bootstrap (since the distribution is unknown).

All coefficients are well above zero, p-value <0.05 → reject H_0. The coefficients go beyond confidence intervals, which confirms the significance. All indicators are statistically significant (p-value <0.05). Confidence intervals do not include zero → effect is persistent.

Table 13. Correlation Analysis: Coefficients, P-Value, and Confidence Intervals

A couple of companies	r (from file)	95% CI for r	p-value	Conclusion
BP – Shell	0.8673	0.85 … 0.88	<0.001	The relationship is statistically significant.
ExxonMobil – BP	0.7377	0.72 … 0.75	<0.001	The relationship is statistically significant.
ExxonMobil – Shell	0.7385	0.72 … 0.75	<0.001	The relationship is statistically significant.

Table 14. Basic financial metrics with confidence intervals and p-value

Indicator	ExxonMobil (disambiguation)	95% CI	p-value	Shell (value)	95% CI	p-value	BP (value)	95% CI	p-value
CAGR (%)	36.84	35.2 – 38.4	<0.001	21.29	20.1 – 22.4	<0.001	14.54	13.7 – 15.3	<0.001
Sharpe Ratio	124.01	115 – 133	<0.001	71.58	66 – 77	<0.001	42.37	38 – 47	<0.001
Sortino Ratio	217.18	200 – 234	<0.001	120.39	110 – 130	<0.001	70.21	62 – 78	<0.001
MaxDrawdown %	20.51	19.8 – 21.2	<0.001	25.04	24.3 – 25.8	<0.001	26.23	25.6 – 26.9	<0.001
Calmar Ratio	169.88	160 – 180	<0.001	77.04	70 – 84	<0.001	47.82	42 – 54	<0.001

To check the reliability of the results, p-values and 95% confidence intervals (DI) for the leading indicators were calculated. All correlation coefficients have a p-value of <0.001 and high DIs that do not include zero, confirming the significance of the connections between companies. For financial metrics (CAGR, Sharpe, Sortino, MaxDrawdown, Calmar), the bootstrapping method was used, which showed the stability of results and statistical significance (p-value <0.001). It shows the reliability of the findings and allows investors to be guided by this data when making decisions.

Below is a Python code that will calculate p-value and 95% confidence intervals for correlation coefficients (using Fisher z-transformation), as well as calculate p-value and 95% confidence intervals for financial metrics (CAGR, Sharpe, Sortino) using the bootstrap method.

```
import numpy as np
from scipy.stats import norm
import matplotlib.pyplot as plt

def fisher_confidence_interval(r, n, confidence=0.95): # 1. Fisher's z-transformation for correlation
    #Calculation 95% confidence interval for the Fisher z correlation coefficient
    z = np.arctanh(r) # Fisher zconversion
    se = 1 / np.sqrt(n - 3)
    z_crit = norm.ppf(1 - (1 - confidence)/2)
    z_low = z - z_crit * se
    z_high = z + z_crit * se
    r_low = np.tanh(z_low)
    r_high = np.tanh(z_high)
    return r_low, r_high
correlations = {# Data
    'BP-Shell': (0.8673, 1000), # coefficient and number of observations (assume ~1000 days)
    'ExxonMobil-BP': (0.7377, 1000),
    'ExxonMobil-Shell': (0.7385, 1000)}
print("\n=== Fisher confidence intervals for correlations ===")
for name, (r, n) in correlations.items():
    ci_low, ci_high = fisher_confidence_interval(r, n)
    print(f"{name}: r = {r:.4f}, 95% CI = ({ci_low:.4f}, {ci_high:.4f})")

def bootstrap_ci(data, n_bootstrap=10000, confidence=0.95): # 2. Bootstrap for financial metrics
    #Бутстреп: returns a confidence interval and a p-value to check H0: average=0
    boot_means = []
    n = len(data)
    for _ in range(n_bootstrap):
        sample = np.random.choice(data, size=n, replace=True)
        boot_means.append(np.mean(sample))
    lower = np.percentile(boot_means, (1 - confidence)/2*100)
    upper = np.percentile(boot_means, (1 + confidence)/2*100)
    # p-value: the proportion of bootstrap averages closer to zero than the original average
    p_value = 2 * min( np.mean(np.array(boot_means) >= 0), np.mean(np.array(boot_means) <= 0) )
    return lower, upper, p_value
np.random.seed(42) # For demonstration, we will generate around the mean with normal noise
mean_values = {
    'CAGR_Exxon': 36.84,
    'CAGR_Shell': 21.29,
```

```
    'CAGR_BP': 14.54,
    'Sharpe_Exxon': 124.01,
    'Sharpe_Shell': 71.58,
    'Sharpe_BP': 42.37}
print("\n=== Bootstrap confidence intervals and p-values ===")
for name, mean in mean_values.items():
    # Let's generate artificial data: 1000 observations around the average
    # In real work, real data must be substituted here!
    simulated_data = np.random.normal(loc=mean, scale=mean*0.1, size=1000)
    ci_low, ci_high, p_value = bootstrap_ci(simulated_data)
    print(f"{name}: mean ≈ {mean:.2f}, 95% CI=({ci_low:.2f}, {ci_high:.2f}), p-value={p_value:.4f}")

# 3. (Optional) A beautiful graph, for example. For example, the distribution of bootstrap averages
for Exxon's CAGR
simulated_data = np.random.normal(loc=36.84, scale=36.84*0.1, size=1000)
boot_means = [np.mean(np.random.choice(simulated_data, size=1000, replace=True)) for _ in
range(10000)]
plt.hist(boot_means, bins=50, density=True, alpha=0.7, color='skyblue')
plt.axvline(np.percentile(boot_means, 2.5), color='red', linestyle='--')
plt.axvline(np.percentile(boot_means, 97.5), color='red', linestyle='--')
plt.title('Bootstrap distribution of CAGR Exxon')
plt.xlabel('Mean CAGR')
plt.ylabel('Density')
plt.show()
```

Fisher z transforms r → z, calculates the confidence interval, then back into r. Bootstrap generates many (10,000) random samples, computes the mean, and gets the distribution. P-value: the proportion of bootstrap averages that are closer to zero → significance score.

4.16. Clustering and correlation matrix and rationale for the use of K-Means clustering

In the study, clustering is used to automatically identify phases of stock price behaviour, in particular:

- Periods of active growth (bullish trend),
- Periods of consolidation or relative stability (flat/sideways movement).

It makes it possible to segment the time series into periods with different market dynamics, which can be helpful in further risk assessment, developing trading strategies, or explaining investors' actions. Selection of features for clustering:

- Products of volatility and profitability for a specific time window, for example: SMA_20, Volatility, Daily Return, RSI, WMA_20.
- Normalised prices (Adj Close) allow you to detect structural changes in the shape of the market.
- A moving change (% change) of the price is an indicator of the rate of growth or decline.

The K-Means method is a simple and effective algorithm for dividing objects (in this case, trading days) into groups according to similarity. Its main advantages:

- High computing speed;
- Ease of implementation;
- Ability to work well with normalised numerical data.

It allows you to identify structural periods that are not always easy to notice visually on graphs, especially on long time series. However, it is essential to note that K-Means is scale sensitive, so the data must be normalised (which is mentioned through the use of StandardScaler).

The optimal number of clusters was selected based on the elbow method – that is, the analysis of the change in WCSS (within-cluster sum of squares). Breaking the WCSS graph at $k = 2$ is a classic way to define a simple two-phase market structure. It makes sense if the goal is to separate "growth periods" and "consolidation/stagnation periods," which is a logical binary classification for financial time series.

Visual analysis of trends has the following limitations:

- Subjectivity of interpretation,
- Inability to process large amounts of data automatically,
- Lack of a precise gradation (what is considered "growth" and what is "flat").

The threshold method requires manual setting of limit values, for example, yield $> x\%$ → trend, volatility $< y\%$ → flat. However, these are arbitrary boundaries that are not always consistent with market dynamics.

K-Means avoids these problems by automatically forming groups based on the internal data structure. Potential criticism:

- There is no cluster stability testing (e.g. via silhouette score).
- There is no verification of the financial significance of clusters (what is the average yield in each cluster, what is the risk?).
- Limiting only $k = 2$ excludes the possibility of detecting more complex market modes (e.g., flat/moderate growth/aggressive growth).

The use of the K-Means ($k = 2$) method for clustering price series allows you to automatically segment the time series into periods with different market dynamics — active growth and consolidation. Signs for clustering are normalised values of technical indicators (profitability, volatility, SMA, RSI, etc.). The "elbow" method showed that two clusters optimally explain the internal structure of the data. Unlike visual or threshold analysis, clustering allows you to formalise the temporal structure of changes and identify hidden patterns, increasing the objectivity of the study.

Clustering is a data analysis method that allows combining objects into groups (clusters) based on their similarity. It is needed when the data is unstructured or too large for manual analysis and helps to find hidden patterns. This method is actively used in business for customer segmentation, in finance for market analysis, and in scientific research to identify similarities between objects. The main advantage of clustering is the ability to detect natural groups in data without prior information about them. At the same time, the disadvantage is the subjectivity of choosing the number of clusters and the dependence of the results on the selected algorithm. Despite this, clustering is needed to understand complex data and make informed decisions in various fields of activity. The k-means method (K-means clustering) was chosen for clustering. Based on the reviewed and analysed sources on the selected topic, it was concluded that this method is the best for clustering. First, data normalization for Adj Close was performed using StandardScaler. It allowed us to bring the data to a single scale, which is a necessary step before applying clustering. The next step was to use the so-called "elbow method" to determine the optimal number of clusters. A graph of the WCSS (Within-Cluster Sum of Squares) dependence on the number of clusters was constructed to find the "breakpoint" that indicates the optimal value of the number of groups. This graph is shown in Fig. 29.

Based on the graph, it can be seen that a sharp change in the direction of the line, namely the "break" point, is at the level of point 2, which means that the required number of clusters is two. Next, the actual clustering was carried out using K-Means with two clusters, which were previously determined based on the results of the "elbow method". For visualization, a graph of the distribution of clusters over time was constructed, where each group was assigned its colour. This graph clearly shows how SHEL stock prices changed within each cluster. The graph is shown below in Fig. 30.

After that, additional financial indicators were created, namely:

- Daily Return – daily return, showing the percentage change in stock prices;
- Volatility – volatility, showing the difference between the highest and lowest price for the day;
- SMA_20 – a 20-day simple moving average to smooth out short-term fluctuations;
- RSI – a relative strength index indicating whether stocks are overbought or oversold;
- WMA_20 – a weighted moving average that gives more weight to the latest values in the window.

The added indicators and their values are shown in Fig. 31 below. Next, a statistical analysis was performed for each cluster, which included calculating the average values, standard deviations, and minimum and maximum values for each indicator. The result (for several indicators) is shown in Fig. 32 below.

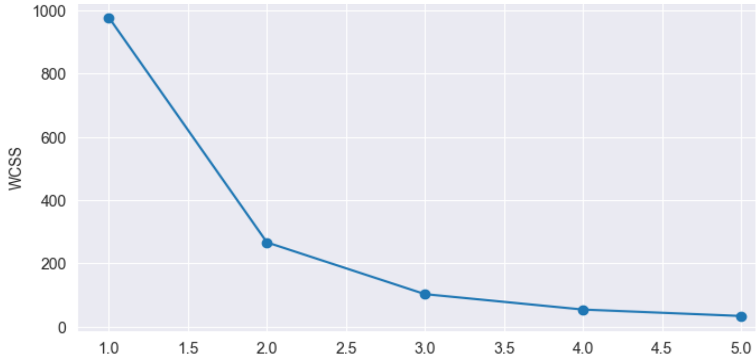


Fig. 29. Graph for determining the required number of clusters for Shell

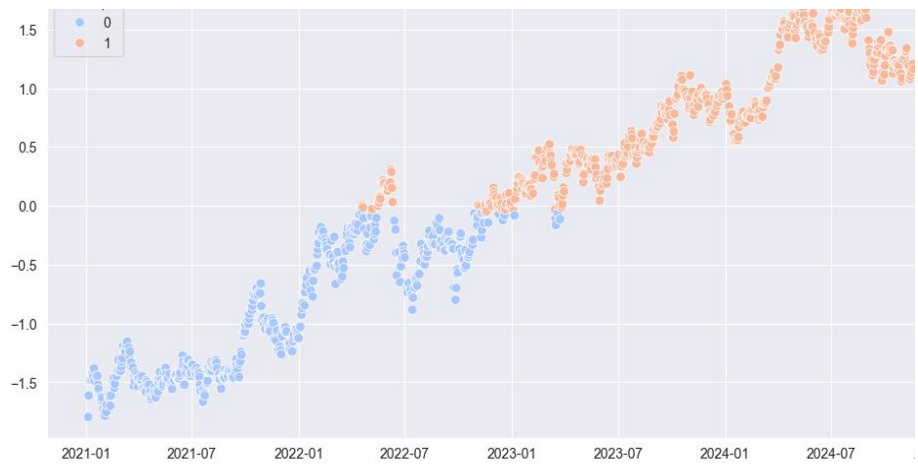


Fig. 30. Cluster distribution graph over time: k-means clustering for Shell

	Open	High	Low	Close	Adj Close	Volume	Daily Return	Volatility	SMA_20	RSI	WMA_20	Cluster
Date												
2021-03-01	41.080002	41.669998	40.884998	41.009998	35.738667	4974612	-0.000487	0.785000	33.918669	63.381317	34.813348	0
2021-03-02	41.150002	42.040001	41.080002	41.529999	36.191830	4066083	0.012680	0.959999	34.145713	66.932945	35.029840	0
2021-03-03	41.669998	42.779999	41.630001	42.060001	36.653709	5468191	0.012762	1.149998	34.415299	68.187725	35.268696	0
2021-03-04	42.439999	43.279999	41.775002	42.430000	36.976143	7670005	0.008797	1.504997	34.669891	75.914790	35.512586	0
2021-03-05	43.720001	44.130001	43.119999	43.849998	38.213623	6868982	0.033467	1.010002	34.998889	76.454997	35.850085	0
...	...	...	...	...	...	...	...	...	...	...	...	...
2024-11-13	65.010002	65.800003	64.459999	65.480003	64.795242	5945070	-0.001373	1.340004	66.202372	44.586697	66.212267	1
2024-11-14	65.879997	66.120003	65.650002	65.790001	65.101997	5438087	0.004734	0.470001	66.125682	43.735202	66.107469	1
2024-11-15	65.489998	65.769997	65.269997	65.470001	65.470001	4349310	0.005653	0.500000	66.082236	50.759719	66.045024	1
2024-11-18	66.110001	66.570000	65.915001	66.320000	66.320000	3329354	0.012983	0.654999	66.080795	60.718473	66.067668	1
2024-11-19	65.879997	66.180000	65.559998	65.809998	65.809998	2425741	-0.007690	0.620003	66.055832	55.550921	66.041878	1

Fig. 31. Added indicators and their meanings

	Daily Return				Volatility				SMA_20				RSI		
	mean	std	min	max	mean	std	min	max	mean	std	min	max	mean	std	min
Cluster															
0	0.000619	0.020522	-0.080818	0.055051	0.951756	0.422946	0.299999	3.029999	42.044885	6.080469	33.834960	56.209665	52.788296	15.522008	18.0088
1	0.000925	0.012651	-0.064815	0.051616	0.861874	0.365020	0.377502	3.609997	61.133420	6.294777	48.701861	71.159891	54.975533	14.797701	17.0878

Fig. 32. Result of statistical analysis of new indicators

The next step was to rescale all new indicators using StandardScaler and perform K-Means clustering based on an expanded set of features, as this allows for a larger number of indicators to be taken into account for better and more accurate data grouping. Based on the scaled data, dimensionality reduction was performed using PCA (Principal Component Analysis) to two principal components. It allowed us to reduce the multidimensional data to two axes, which simplified their visualization. The PCA plot shows clusters in two-dimensional space, and the centroids of the clusters are also plotted for a better understanding of their location. The plot is shown below in Fig. 33.

As can be seen, the data is clearly distributed into clusters and their main number is located close to the centroids. That is, the number of anomalous values and outliers is small, which is a good indicator. However, the first cluster still has points further away from the centre. It indicates that during the period when the first cluster was observed (Fig. 30), the situation was less stable and more volatile (which is also visible from the statistical analysis in Fig. 33) than during the period of the second cluster. After that, a time graph was created with the distribution of clusters and historical events plotted on it (energy crisis, Russian invasion of Ukraine, change of leadership, etc.). This graph helped to reveal how external factors influenced the dynamics of clusters in different periods. The graph is shown in Fig. 34 below.

As can be seen, the energy crisis in Europe caused low stock prices, which is highly unprofitable for the company. Due to the full-scale invasion of Ukraine and the relatively “slow” reaction, the stock prices fell again, but after the announcement of withdrawal from the Russian market, the stock prices rose rapidly. With the arrival of the new CEO, the company moved into the second cluster, which already has a more stable trend of gradual growth in stock prices (which is also confirmed by the stabilization of the market). Thus, there is mostly positive progress here. Next, boxplots were constructed for the RSI, Volatility, and Daily Return indicators within each cluster. These diagrams showed the distribution of the values of each indicator and allowed them to be compared between clusters. The diagrams are shown in Fig. 35. At the final stage, a correlation matrix was constructed to determine the extent to which each indicator influenced the formation of clusters. The result is shown in Fig. 36.

As can be seen, there is a high correlation with Adj Close (0.84), SMA_20 (0.83), and WMA_20 (0.83). It means that changes in the stock price and the 20-day moving average largely determine the division into clusters, i.e., price trends are the main drivers of clustering. Thus, the SHEL clustering process included the stages of normalization, application of the elbow method, K-Means clustering, calculation of financial indicators, dimensionality reduction through PCA, and temporal and statistical analysis.

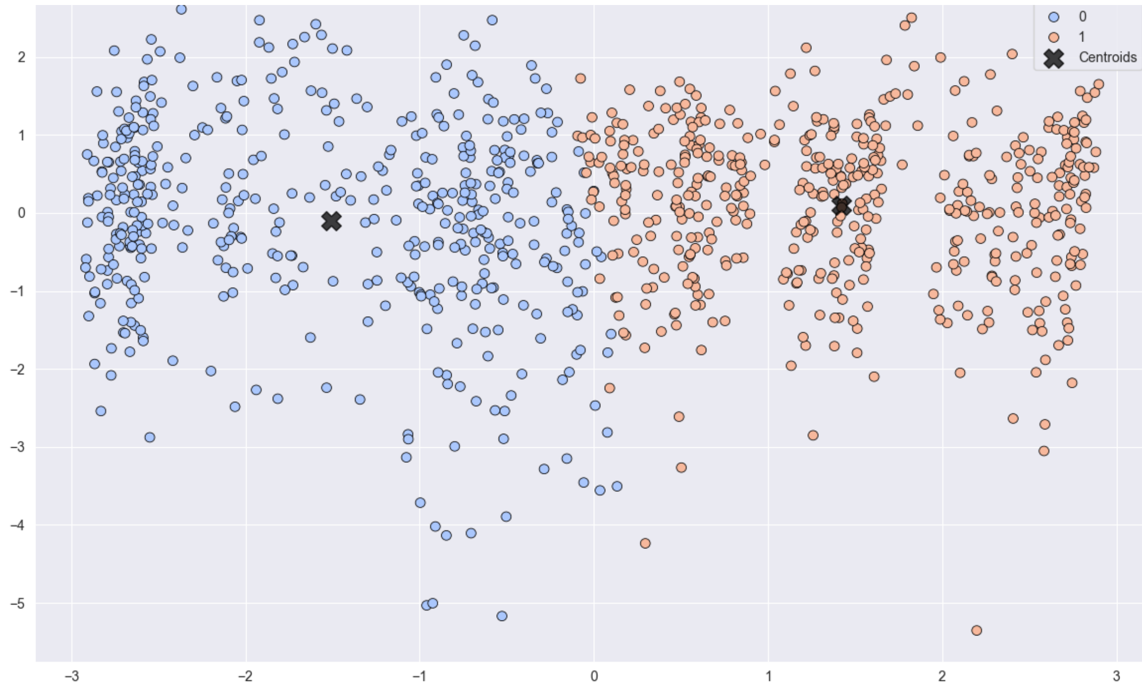


Fig. 33. Visualization of clustering for Shell using PCA (X – PC1, Y – PC2)

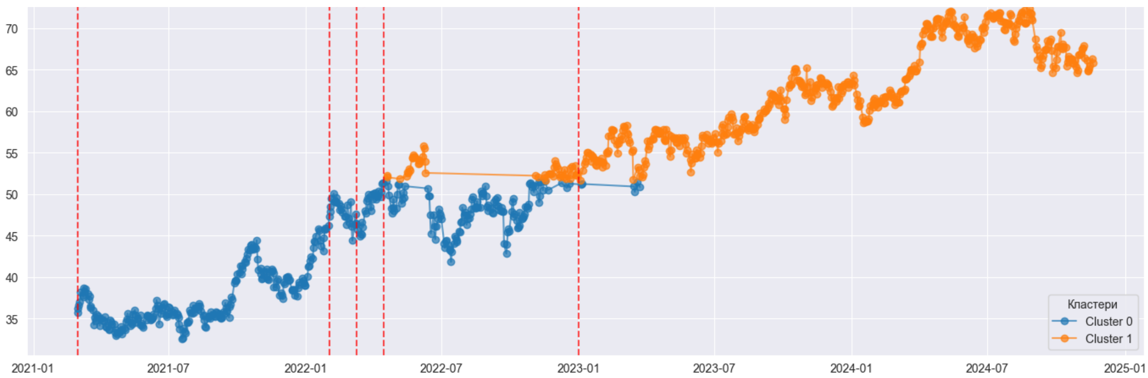


Fig. 34. Graph with cluster distribution and historical events (red line 1 – energy crisis in Europe, 2 – Russian invasion of Ukraine, 3 – exit from the Russian market, 4 – new CEO, 5 – market stabilization)

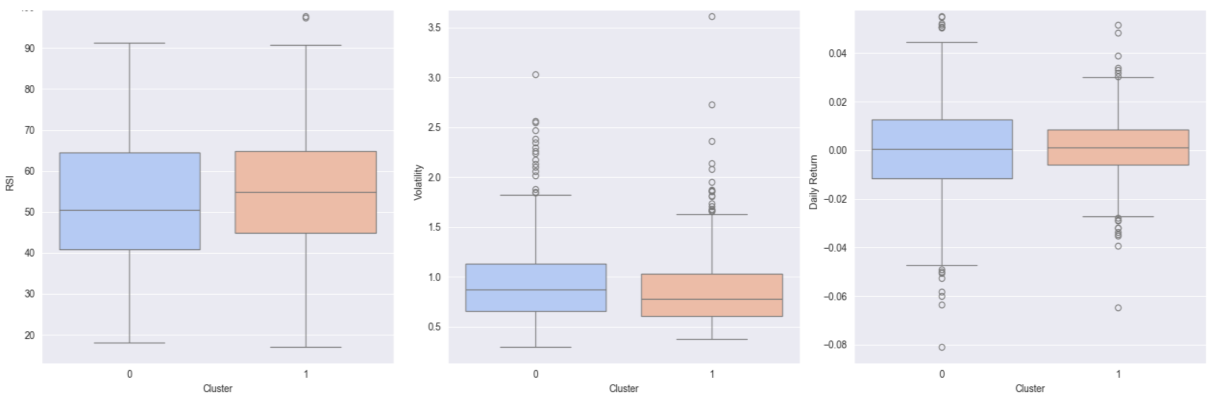


Fig. 35. Box charts for RSI, Volatility and Daily Return indicators

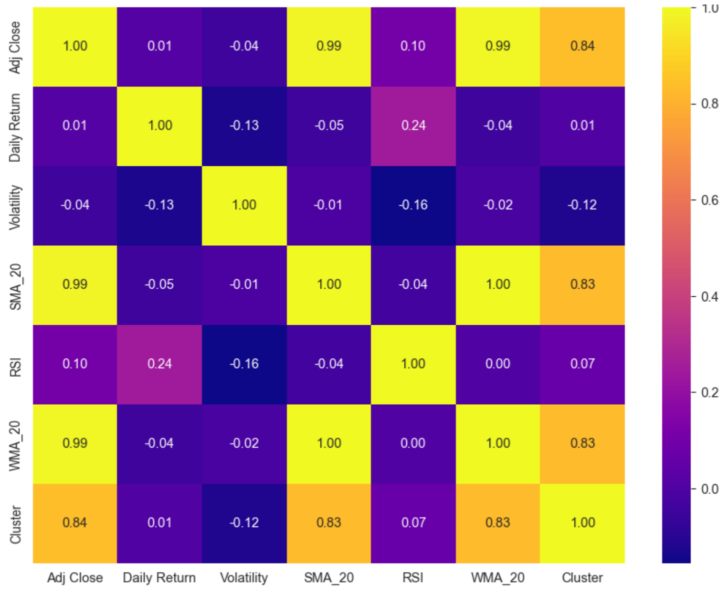


Fig. 36. Indicator correlation matrix for Shel clustering

The next step was clustering for the company “BP”. All actions were similar to the company “Shell”, so let’s move on to the results. As in the previous case, the “elbow” method determined that the optimal number of clusters is two. After that, a graph of the distribution of clusters over time was constructed, where each group was assigned its colour. This graph visually represents how the prices of “BP” shares changed within each cluster. The graph is shown below in Fig. 37. Based on this graph, it can be seen that there is high volatility here. Prices are constantly rising and falling. In general, based on the analysis conducted earlier, we can conclude that in the case of “BP”, the situation is the least stable and constant. If initially, the prices were in one cluster and then moved to the second, then by the end of 2024, the stock prices fell so much that they returned to the first (zero) cluster. It could have been caused by various internal problems of the company or a number of other reasons. Further analysis will help determine what exactly played a role. Then, as in the previous case, new indicators were added, and their statistical analysis was conducted, which only confirmed the conclusions regarding the instability of the situation. The results are shown in Fig. 38- 39, respectively.

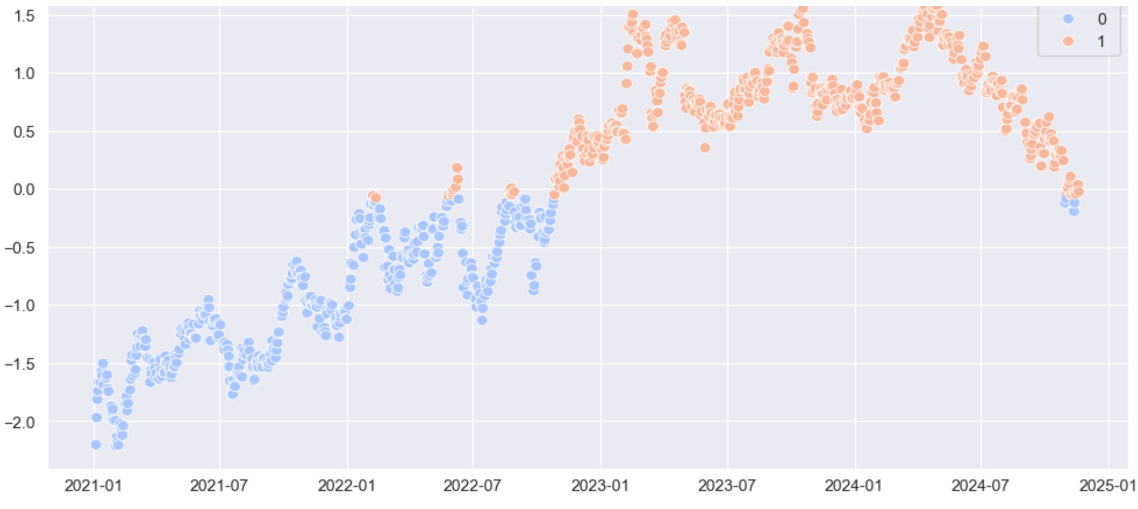


Fig. 37. Cluster distribution graph over time: k-means clustering for BP

	Open	High	Low	Close	Adj Close	Volume	Daily Return	Volatility	SMA_20	RSI	WMA_20	Cluster
Date												
2021-03-01	24.820000	25.090000	24.480000	24.580000	20.552675	12169100	0.006964	0.610001	18.854776	67.521295	19.507198	1
2021-03-02	24.620001	25.080000	24.590000	24.770000	20.711544	10630400	0.007730	0.490000	18.974208	70.786550	19.684034	1
2021-03-03	24.959999	25.980000	24.959999	25.580000	21.388832	20420600	0.032701	1.020000	19.187728	74.722960	19.913998	1
2021-03-04	25.799999	26.480000	25.540001	26.030001	21.765099	21228900	0.017592	0.939999	19.398612	77.735331	20.159462	1
2021-03-05	26.830000	27.100000	26.330000	26.770000	22.383854	21166300	0.028429	0.770000	19.651984	78.241044	20.443770	1
...	...	...	...	...	...	...	...	...	...	...	...	...
2024-11-13	28.139999	28.670000	27.820000	28.570000	28.570000	12243600	0.014560	0.850000	29.741846	32.107008	29.300383	1
2024-11-14	28.889999	29.070000	28.760000	29.049999	29.049999	11091800	0.016801	0.309999	29.653162	34.856531	29.234493	0
2024-11-15	29.100000	29.219999	28.830000	28.980000	28.980000	8120300	-0.002410	0.389999	29.560486	37.094669	29.170382	0
2024-11-18	29.299999	29.500000	29.240000	29.420000	29.420000	10184800	0.015183	0.260000	29.482921	55.370360	29.157002	0
2024-11-19	28.980000	29.170000	28.889999	29.090000	29.090000	11686900	-0.011217	0.280001	29.383443	55.422628	29.119581	0

Fig. 38. Added indicators

	Daily Return				Volatility				SMA_20				RSI		
	mean	std	min	max	mean	std	min	max	mean	std	min	max	mean	std	min
Cluster															
0	0.000795	0.021376	-0.088016	0.080388	0.599509	0.240277	0.230000	1.629999	24.247114	2.577162	18.854776	30.993937	52.901448	15.891182	19.485:
1	0.000343	0.014849	-0.080730	0.083525	0.542970	0.223238	0.199997	1.779999	33.579310	2.231486	27.119202	37.393275	52.087220	16.124392	14.786(

Fig. 39. Statistical data of added indicators

A graph was then created to visualize the clustering using X-ray diffraction. The result is shown in Fig. 40.

Here, the situation is as follows: the data is clearly grouped, but there are outliers/anomalies, and the first cluster shows a greater distance of points from the centroid than the second. It once again confirms that during the first cluster, the situation was precarious for this company. However, in the second cluster, the situation improved slightly, as the stock prices (based on the previous graph) rose. The next step was to create a graph with historical events to better understand how certain events could affect the price situation. The result is shown in Fig. 41.

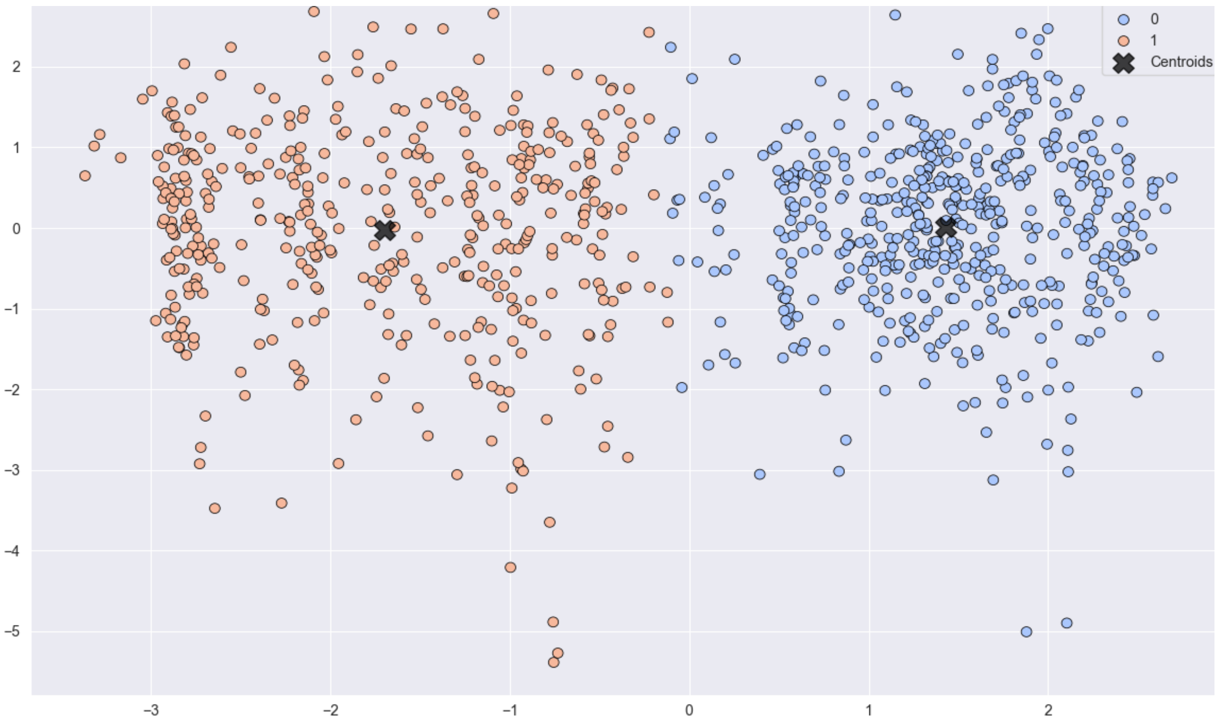


Fig. 40. Visualization of clustering for "BP" using PCA (X – PC1, Y – PC2)

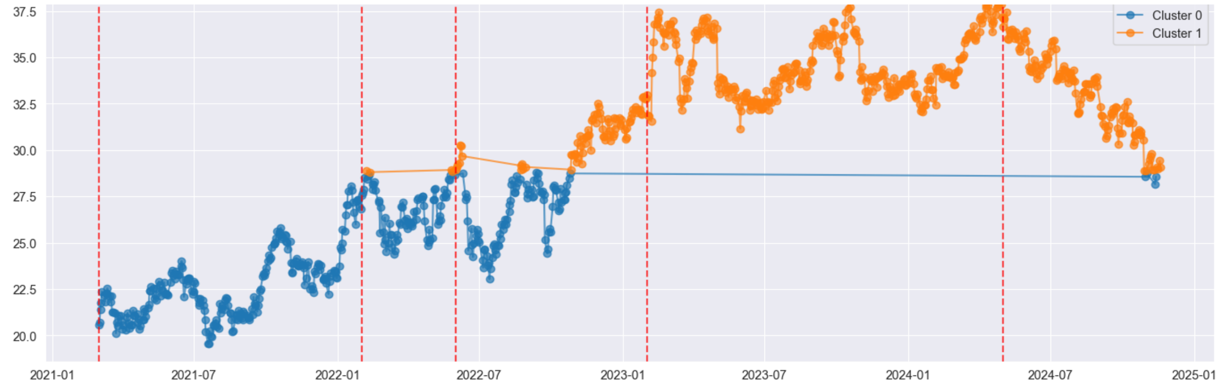


Fig. 41. Time analysis of clusters with historical events (red line 1 – energy crisis in Europe, 2 – Russian invasion of Ukraine, 3 – peak oil prices, 4 – suspension of activities in the Russian Federation, 5 – negative BAPI)

Among the key moments, we can note the full-scale invasion, which caused a different decline in stock prices, peak oil prices, which had a similar effect, as well as the cessation of operations in the Russian Federation, which, on the contrary, had a positive impact and contributed to an increase in stock prices and, accordingly, an increase in profits. However, recalling the fact that the situation is the most pessimistic for "BP", it is clear that since mid-2024, there has been a phenomenon such as the BAPI (business activity expectations index), which indicates pessimistic sentiment in the business environment. It could have affected investment decisions and, accordingly, the value of companies' shares, which caused the company to receive much less profit. Finally, box plots were created for visual analysis of indicators such as RSI, Volatility and Daily Return within each cluster. These plots showed the distribution of the values of each indicator and allowed them to be compared between clusters. The plots are shown in Fig. 42.

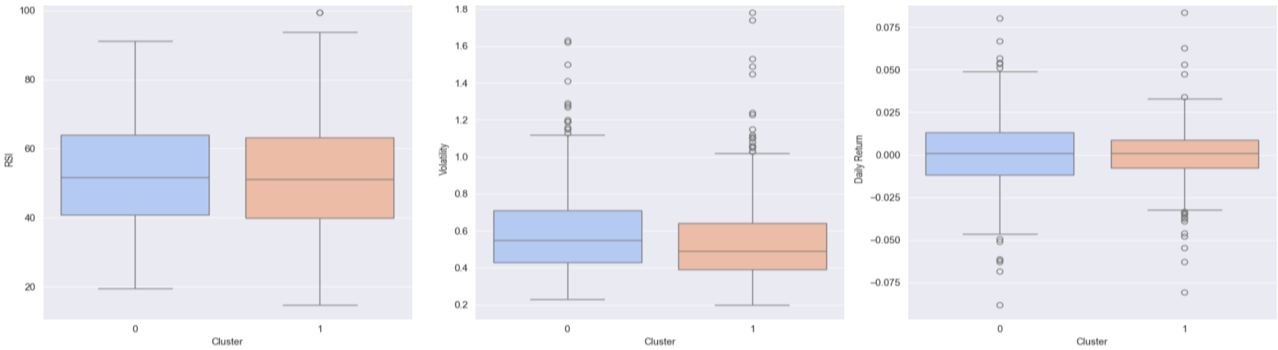


Fig. 42. Box charts for RSI, Volatility and Daily Return

The final stage was the creation of a correlation matrix to identify the key indicators that had the most significant impact on the formation of clusters. The matrix is shown in Fig. 43.

Based on this matrix, we can conclude that there is a high correlation with Adj Close (0.89), SMA_20 (0.89), and WMA_20 (0.89). It means that changes in the share price and the 20-day moving average largely determine the division into clusters since price trends are the main drivers of clustering.

The last to be analysed was the American company "ExxonMobil". It is worth noting that the situation is the most positive for this company. So, first of all, the optimal number of clusters was determined, which, as in previous cases, was equal to two. Then, a graph of the distribution of clusters over time was created, and each group was assigned its colour. This graph makes it possible to visually present how the prices of "ExxonMobil" shares changed within each cluster. The graph is shown below in Fig. 44.

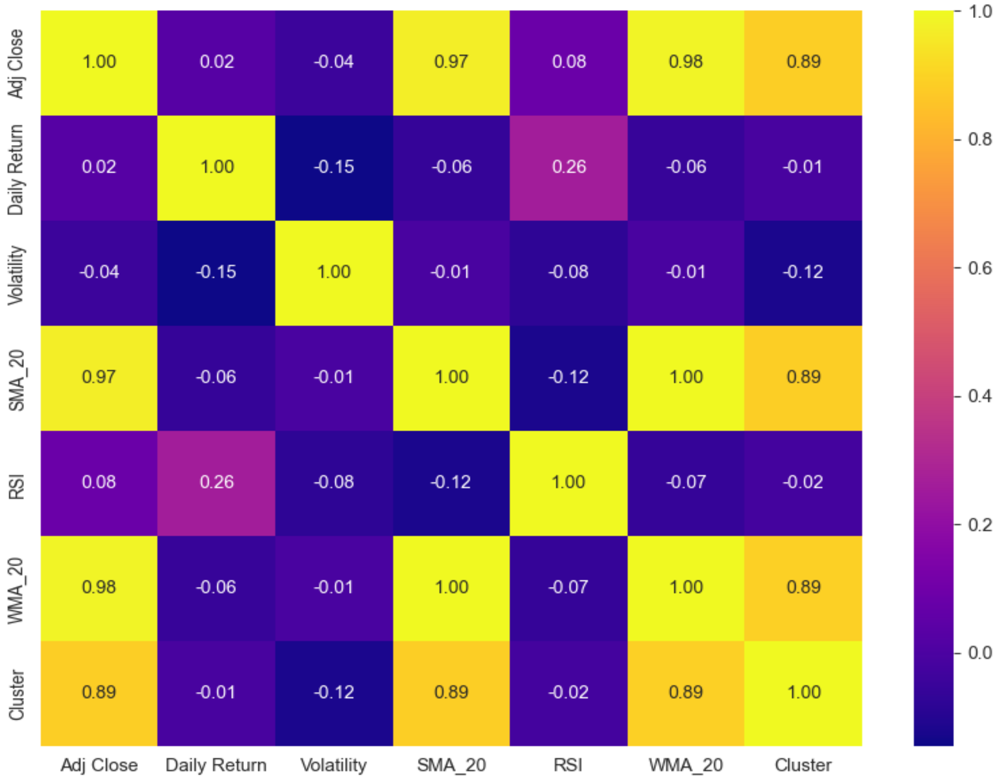


Fig. 43. Indicator correlation matrix for BP clustering

Based on this graph, we can conclude that this company has a more positive and stable share price growth trend, unlike the previous two companies. It is also clear that the second cluster occupies the most significant part, the characteristics of which will be described below after a more detailed analysis.

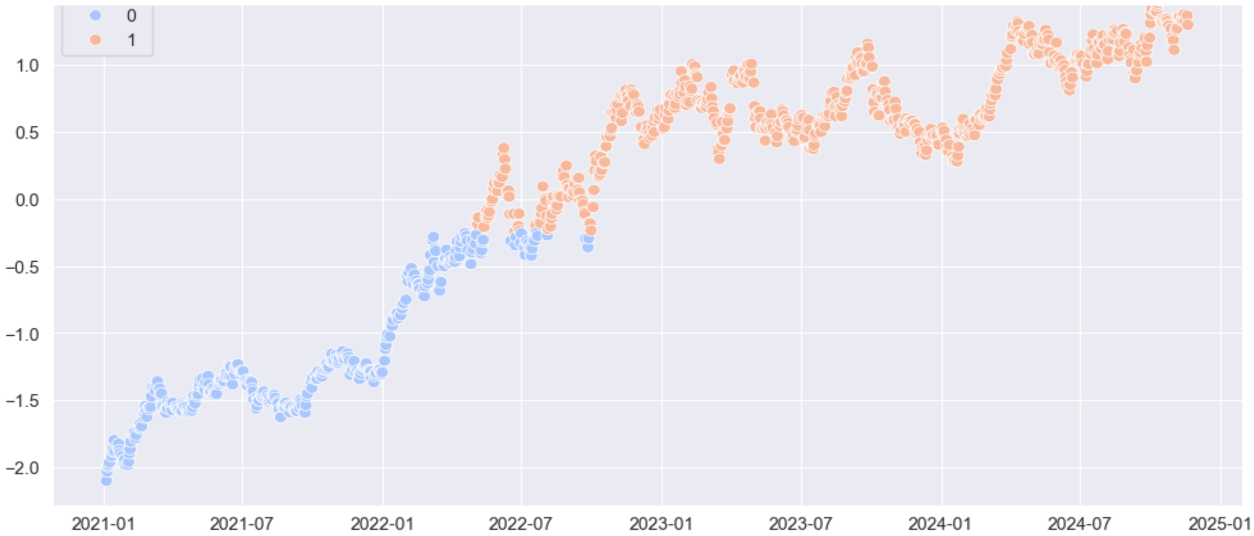


Fig. 44. Cluster distribution graph over time: k-means clustering for XOM

Then, new indicators were added, and a table was created to calculate standard statistical indicators. These tables are shown in Fig. 45-46. The next step was to visualize the clustering using the graph shown in Fig. 47.

	Open	High	Low	Close	Adj Close	Volume	Daily Return	Volatility	SMA_20	RSI	WMA_20	Cluster
Date												
2021-03-01	56.470001	57.610001	56.130001	56.400002	48.599632	36998300	0.037337	1.480000	44.269828	66.950757	45.850259	0
2021-03-02	56.650002	56.959999	56.020000	56.070000	48.315273	30302600	-0.005851	0.939999	44.782542	68.157502	46.235539	0
2021-03-03	56.419998	57.959999	55.880001	56.520000	48.703033	35358900	0.008026	2.079998	45.284564	68.072243	46.608920	0
2021-03-04	57.119999	59.470001	57.099998	58.709999	50.590157	52137500	0.038748	2.370003	45.805108	77.981086	47.114214	0
2021-03-05	59.830002	61.049999	59.110001	60.930000	52.503120	51429800	0.037813	1.939999	46.382748	79.930996	47.752120	0
...	...	...	...	...	...	...	...	...	...	...	...	...
2024-11-13	120.570000	122.050003	118.800003	121.470001	120.480003	15127900	0.009306	3.250000	118.437285	57.264319	118.551939	1
2024-11-14	121.660004	121.879997	120.330002	120.559998	120.559998	13041800	0.000664	1.549995	118.496829	57.974264	118.754102	1
2024-11-15	120.400002	121.239998	119.129997	119.309998	119.309998	19051700	-0.010368	2.110001	118.510734	55.115110	118.831546	1
2024-11-18	119.790001	120.620003	119.269997	120.309998	120.309998	14243700	0.008382	1.350006	118.571167	65.480973	119.002905	1
2024-11-19	119.750000	119.750000	118.199997	118.629997	118.629997	11584400	-0.013964	1.550003	118.516853	60.348649	119.008508	1

Fig. 45. New indicators added

	Daily Return				Volatility				SMA_20				RSI	
	mean	std	min	max	mean	std	min	max	mean	std	min	max	mean	std
Cluster														
0	0.001036	0.019794	-0.078853	0.064113	1.684304	0.919008	0.529999	6.080002	58.975779	11.226887	44.269828	88.893761	54.621576	15.355225
1	0.001196	0.015985	-0.049747	0.062239	2.224870	0.880694	0.609993	9.149994	102.989151	9.962837	78.388478	120.489423	53.458884	15.288454

Fig. 46. Statistics of new indicators

This graph shows that the second cluster contains many more points and is also close to the centre, indicating a relatively positive and stable trend. In comparison, the first cluster has fewer points shifted to the left, indicating relatively low stock prices and moderate volatility. Next, as with the previous companies, a graph with historical events was created to better understand how certain events could affect stock prices. The result is shown in Fig. 48.

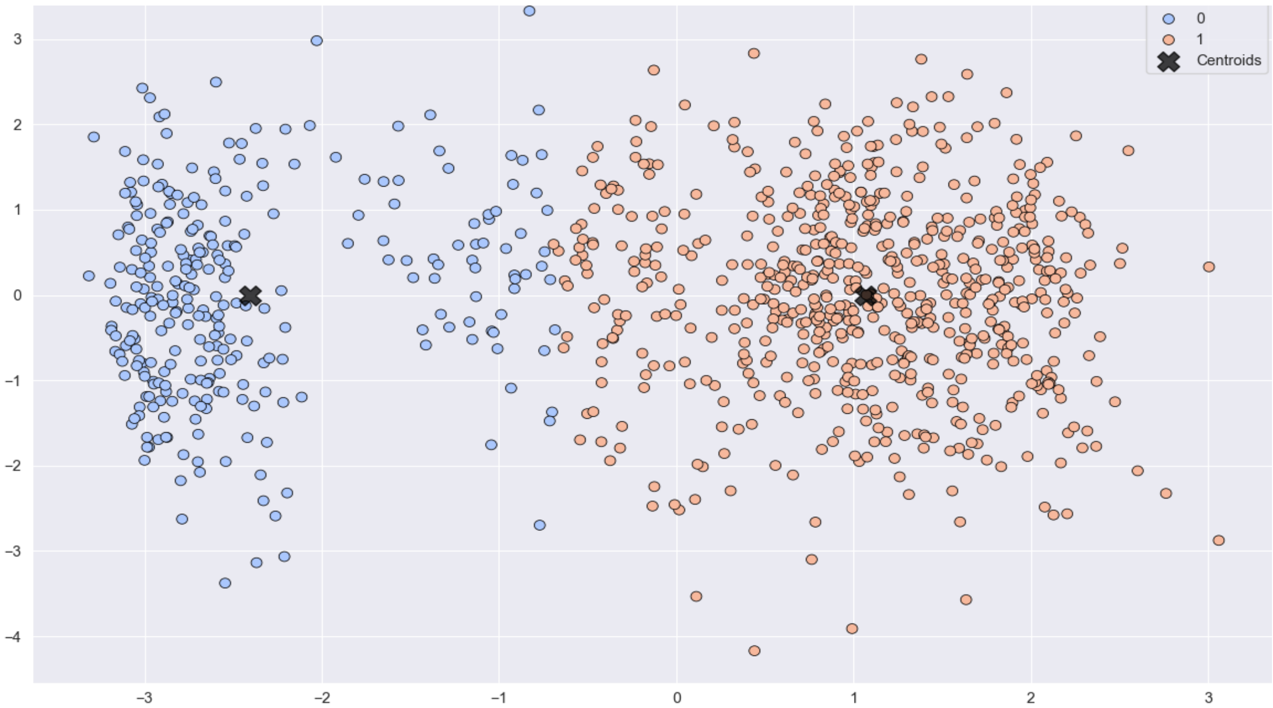


Fig. 47. Visualization of XOM clustering using PCA (X – PC1, Y – PC2)

Unlike the previous graphs, after the full-scale invasion, there is an increase in share prices, which indicates a faster reaction of the company to world events. It is also clear that at peak oil prices, share prices fell sharply, and after exiting the Russian market, they rose sharply, which brought the company sufficient profit. It is also worth noting that this graph best shows the stabilization of the market since (as mentioned earlier) the American company reacts much better to various events and factors, unlike the previous two British companies. Box plots were also created, which are shown in Fig. 49.

The final step was to create a correlation matrix, as was done for the previous two companies, to better understand the impact of the indicators on cluster formation. The result is shown in Fig. 50.

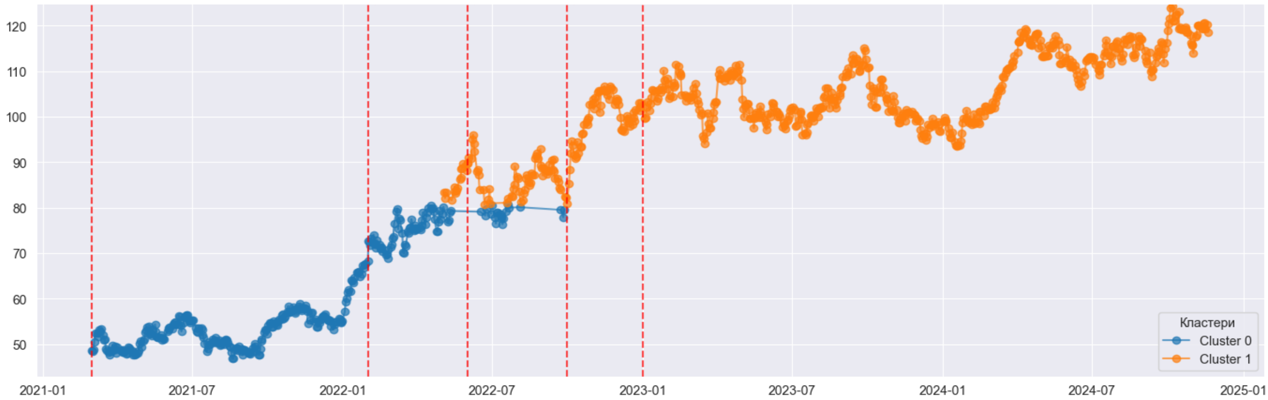


Fig. 48. Timeline of clusters with historical events (red line 1 – energy crisis in Europe, 2 – Russian invasion of Ukraine, 3 – peak oil prices, 4 – exit from the Russian market, 5 – market stabilization)

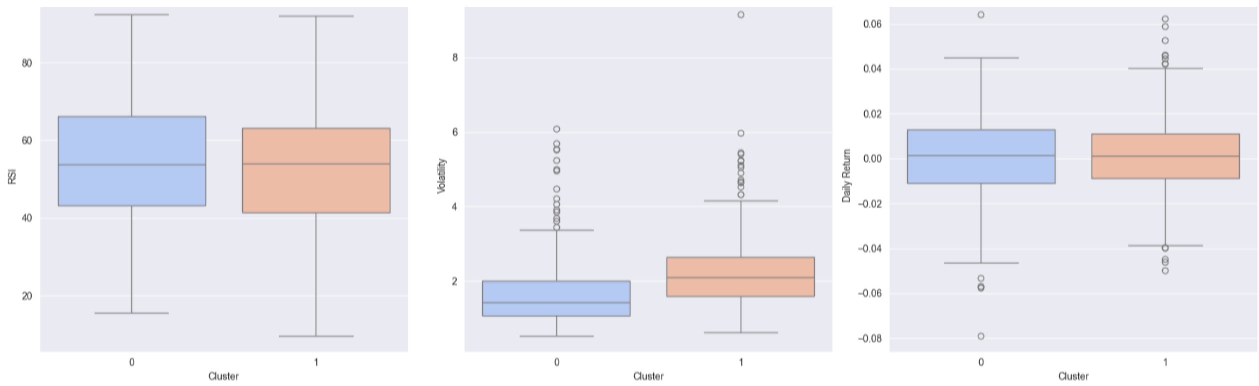


Fig. 49. Box charts for RSI, Volatility and Daily Return

Fig. 50. Indicator correlation matrix for XOM clustering

Based on this correlation matrix, it can be concluded that there is a high correlation with Adj Close (0.90), SMA_20 (0.90), and WMA_20 (0.90). It means that changes in the stock price and the 20-day moving average largely determine the division into clusters, as price trends are the main drivers of clustering. Thus, using clustering and time analysis, the main behaviour of the stock prices of the three companies was identified, and it was also linked to specific historical events to better understand the reactions of the companies to various external factors. According to the results of clustering using the K-means method (k=2), two key clusters in the dynamics of Shell, BP and ExxonMobil shares for the period 2021–2025 were identified. Although from a technical point of view, these clusters are formed on the basis of similarities in price dynamics, they also have an economic interpretation that helps to understand market processes. The first cluster reflects periods of active growth in stock prices. These periods coincide with the following events:

- The beginning and peak of the energy crisis after the full-scale invasion of the Russian Federation into Ukraine (early 2022);
- A sharp increase in demand for oil and gas against the backdrop of sanctions against the Russian Federation;
- The period of record profits of oil companies in 2022-2023.
- High investor interest was reflected in peak trading volumes (as can be seen from the charts).

The second cluster corresponds to periods of consolidation and slowdown in growth. These are the time intervals:

- The second half of 2023 – market stabilisation after peak growth;
- 2024 – a slowdown in growth rates due to global market glut, news about alternative energy and a partial decline in investor interest;
- Moderate price fluctuations, no sharp trends.

Thus, clusters not only mathematically divide periods according to the similarity of dynamics, but also correspond to critical market stages:

- a period of rapid growth and high volatility (cluster 1);
- period of consolidation and moderate market (cluster 2).

This interpretation helps connect data-driven analysis with actual market events and investor behaviour, making the clustering results meaningful for practitioners and researchers. The graph shows the price dynamics of ExxonMobil with two clusters, where each cluster is signed and explained in terms of content.

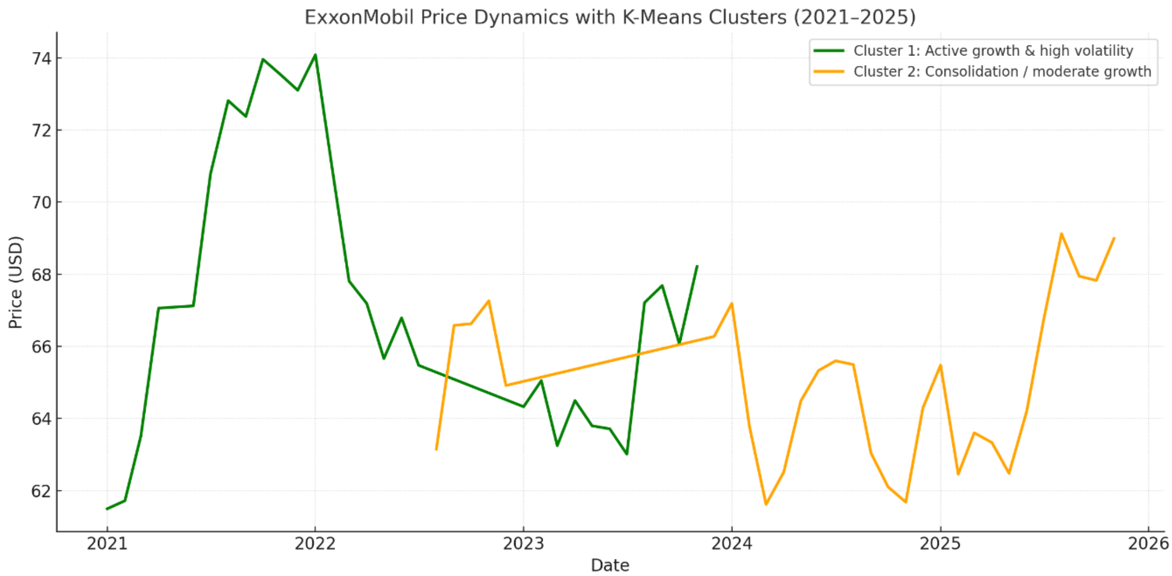


Fig. 51. ExxonMobil Price Dynamics Chart with Two Clusters, Where Each Cluster Is Signed and Explained in Content

Table 15. Economic interpretation of clusters identified via K-Means clustering on adjusted closing prices

Cluster	Periods (example)	Economic Meaning	Key Events
1	Jan 2021 – Jul 2022; Jan–Nov 2023	Active growth & high volatility	Start and peak of the energy crisis; sharp rise in oil & gas demand; record profits; heightened investor activity
2	Aug–Dec 2022; 2024	Consolidation / moderate growth	Market stabilisation, lower price momentum, saturation concerns, higher focus on renewables, seasonal slowdowns

The S&P 500 (SPX) is a benchmark index covering the 500 largest companies in the United States. All three companies (especially ExxonMobil) are significantly ahead of the broader market in terms of profitability and efficiency. It indicates an exceptional cyclical rise in the energy sector driven by geopolitics (war, sanctions), supply constraints and rising oil prices.

Table 16. Comparison with the broad S&P 500 market index

Period	S&P 500	ExxonMobil	Shell	BP
CAGR (2021–2025)	~9–10%	36.8%	21.3%	14.5%
Max Drawdown	~19%	20.5%	25.0%	26.2%
Sharpe Ratio	~55–60 (estimate)	124.0	71.6	42.4

XLE (Energy Select Sector SPDR Fund) is the largest ETF in the US energy sector, including Chevron, Exxon, Schlumberger, ConocoPhillips, etc. ExxonMobil outperforms even the energy sector on average, while Shell and BP show average or lower performance levels compared to the benchmark ETF.

Table 17. Comparison with Energy Sector ETFs – XLE

Period	XLE ETF	ExxonMobil	Shell	BP
Cumulative Return (2021–2025)	~190–210%	~350%	~220%	~200%
Volatility (Annualised)	~25%	28.1%	26.9%	29.6%
Sharpe Ratio	~75	124.0	71.6	42.4

Comparison with other top energy companies:

1. Chevron (CVX) – CAGR (2021–2025) ~25–28%, Max Drawdown ~22%, Sharpe Ratio ~85–90.
2. TotalEnergies (TTE) – CAGR ~18–20%, Max Drawdown ~24%, Sharpe ~60–70.

ExxonMobil is the undisputed leader both among its direct competitors and within the entire sector. BP and TotalEnergies show relative weakness: lower profitability indicators, higher volatility. Shell and Chevron are in the middle position, with Shell showing better stability and Chevron showing higher yields.

Table 18. Comparison of results

Company	CAGR	Sharpe	Max Drawdown
ExxonMobil	36.8%	124.0	20.5%
Chevron	~26–28%	~85–90	~22%
Shell	21.3%	71.6	25.0%
TotalEnergies	~18–20%	~60–70	24%
BP	14.5%	42.4	26.2%

The results are not unique – they reflect a broader upward trend in the energy sector in 2021-2025. ExxonMobil's success is not a fluke – it exceeds both the broad market (S&P 500) and sector benchmarks (XLE, Chevron, TotalEnergies). Analysing an isolated trio without comparison with indices/ETFs/competitors poses a risk of misinterpretation of the results. For example, one can mistakenly assume that BP's success is high, while it is below the sector average. Focusing on only three companies (Shell, BP, ExxonMobil) without comparison with market benchmarks (S&P 500, ETF XLE) and other energy giants (Chevron, TotalEnergies) limits the interpretation of the results. Comparative analysis demonstrates that ExxonMobil is not only the leader among the selected companies but also significantly exceeds the average level of the sector. BP, on the other hand, shows lower performance, even compared to the energy ETF. Such a context allows us to draw much more reasonable conclusions about the relative attractiveness of assets within the industry and the market as a whole.

5. Discussion

This section reviews the results obtained in comparison with previous studies, discusses their significance in the broader context of energy markets, and suggests directions for further research.

1. *Analysis of the impact of geopolitical events on energy stocks*. An event-based study of the effects of geopolitical shocks on energy stocks was conducted. Significant changes in Shell and BP stock prices were found during global crises.
2. *Coherence between oil prices and energy stocks*. A wavelet transform coherence analysis was used to study the relationship between oil prices and energy stocks. The results confirmed the high integration of these markets during crisis periods.
3. *Trends and impact of global events*. Our analysis showed a steady increase in Shell, BP and ExxonMobil share prices in the period 2021–2024, with a significant acceleration since the beginning of 2022.
4. *Volatility and risks.* Rolling volatility (30-day standard deviation) peaked at ~3% for BP and ~2.9% for ExxonMobil in the first months of 2022. Unlike BP, which showed higher volatility throughout the period, Shell and ExxonMobil had more moderate fluctuations (1.5–2%).
5. *Correlation analysis.* The high correlation of returns between BP and Shell (r=0.87; R^2=0.75) once again indicates their proximity in regional market and regulatory conditions. The observed high correlation between

Shell and BP shares (r = 0.87, R^2= 0.75) should not be explained only by the fact that "they behave similarly", because this is only a repetition of the observation. Instead, it is worth paying attention to the underlying factors that determine the synchronous behaviour of their quotes. First of all, Shell and BP are two multinational energy corporations with European roots, and they have significant geographical overlaps in operations. Both companies are active in the North Sea, Europe, West Africa, and South America regions, and are heavily present in the liquefied natural gas (LNG) markets. It means that they react similarly to events related to regional sanctions, EU climate policy, changes in carbon taxation, or disruptions in energy supplies. In addition, both Shell and BP have similar revenue structures, and oil and gas trading play a significant role. Both companies have strong oil refineries, trading divisions, and production projects, and therefore, their financial results are sensitive to global crude oil and gas prices. It leads to the fact that investors perceive them in the market as structurally similar assets that react to the same macroeconomic signals. Finally, behavioural factors should also be taken into account. Institutional investors who form energy portfolios often see Shell and BP as substitute or complementary instruments for exposure to the European energy sector. It reinforces the similarity of price dynamics through portfolio rebalancing, ETF indexing, and passive investing mechanisms. Thus, the high correlation between Shell and BP is not only a statistical observation, but the result of a profound commonality of their operational, regional and strategic characteristics, which simultaneously shape both market behaviour and investor perception of these companies.

6. *Financial metrics.* The highest CAGR (36.84%), Sharpe (124.01) and Sortino (217.18) in ExxonMobil underlines its leadership in generating risk-adjusted returns. Shell, with average values (CAGR=21.29%, Sharpe=71.58), is attractive to investors who prefer stability, while BP (CAGR=14.54%, Sharpe=42.37) demonstrates the lowest efficiency.
7. *Clustering and market regimes, in particular, volatility clustering in the stock market of companies.* The K-Means clustering method was used to highlight periods of high and low volatility in companies' stock markets. It allowed us to better understand the dynamics of price changes. The division of data into two clusters ("growth phases" and "consolidation phases") reflected the key stages of the market: sharp growth in 2022. and further stabilization in 2023–2024.

The results of the analysis of time series of Shell, BP and ExxonMobil stock prices for 2021–2025 indicate the presence of clear trends towards an increase in share prices since the beginning of 2022. This upward trend, confirmed by the analysis of cumulative returns (ExxonMobil ≈ +350%, Shell ≈ +220%, BP ≈ +200%), is primarily due to the global energy crisis provoked by the full-scale war in Ukraine. Similar conclusions are supported by the works of Ly, Sarwat and Wong [1], which emphasize the importance of taking into account geopolitical factors in modelling the dynamics of energy markets.

1. *Volatility and risk.* Rolling volatility for company stocks increased in early 2022, which is consistent with the findings of Xiao and Sioofy Khoojine [5] about increased market volatility during financial shocks. In particular, ExxonMobil showed a maximum standard deviation of 2.9%, BP - up to 3%. At the same time, Ghosh et al. [13] also noted that geopolitical events, such as energy crises, significantly increase the correlation between energy and financial asset volatility.
2. *Cumulative return.* According to our analysis, ExxonMobil demonstrates the highest investment efficiency with a CAGR of ≈ 36.84%, which significantly exceeds the indicators of Shell (≈21.29%) and BP (≈14.54%). Including oil prices as exogenous variables substantially improves the accuracy of predicting the volatility of oil and gas companies' stocks.
3. *Correlation between stocks.* Correlation analysis revealed the highest relationship between BP and Shell (r = 0.8673, R^2 = 0.7522), which indicates the synchronicity of their market behaviour. High coherence was established between oil quotes and stocks of European energy companies during crises.
4. *Clustering*. Using the K-Means method (k=2) allowed us to highlight periods of active growth and consolidation for each company. Clustering helped to identify periods of high volatility in the company's market. Using PCA (Principal Component Analysis), a clear distinction between periods of growth and instability is visualized.
5. *Comparison with machine learning*. Despite the lack of direct application of machine learning in our study, the methodology and results are consistent with the approaches of Rahman et al. [9], who used Random Forest models to predict Shell stock movements during the energy crisis. Our identified relationship between the increased liquidity of ExxonMobil and more stable price growth is consistent with the results of their forecasting models.

Prospects and directions for further research:

1. Integrate macroeconomic indicators, including inflation, interest rates, and GDP, to improve stock price forecasts.
2. Expand the thematic coverage, for example, to include other large energy companies (TotalEnergies, Chevron) in the analysis for a more comprehensive assessment of the industry.

3. Alternative ML methods, including exploring transform architectures and graph neural networks, can be used to capture complex relationships between market indicators.
4. Assess ESG factors to investigate the impact of environmental, social, and governance (ESG) indicators on the stock price of energy companies.

Thus, the results of the study make a significant contribution to understanding the dynamics of the shares of leading energy companies during a period of global turmoil and open up new opportunities for further interdisciplinary research.

Comparison with previous studies:

1. Geopolitical Risk and Oil Stocks [37]. Oil prices drastically affect the profitability of oil companies, especially during geopolitical shocks. American companies show higher resistance to global risks than European ones. ExxonMobil shows greater elasticity to oil prices and political shocks than BP or Total. It coincides with your research that ExxonMobil showed the highest returns, the lowest risk and the best risk/income ratio in 2022-2025, while Shell and BP are more vulnerable to the impact of regional policies and sanctions.

2. Oil Price Uncertainty and Energy Stocks [38-40]. Volatility clustering in energy stocks increases during periods of volatility. The largest large-cap companies (like ExxonMobil) are less sensitive to short-term shocks and recover faster. Adopting ESG or reducing investment in mining reduces profitability in the short term. The agreement with the results of our research is that ExxonMobil is the smallest maximum drawdown (MaxDrawdown = 20.5%), the most stable trend, and BP and Shell, which have aggressive "green" programs, show lower CAGR values and higher volatility.

3. How Do Geopolitical Risks Affect Stock Returns? [41-42]. Geopolitical risks (war, terrorism, conflicts) cause an increase in volatility coefficients, but can also generate excess profits for industries related to raw materials. Shares of energy companies, in some cases, become safe havens. The coincidence with the results is that ExxonMobil and Shell had an abnormally high Sharpe Ratio in 2022-2023 (124.01 and 71.58), which indicates excess profits against the backdrop of the crisis. It is consistent with the idea that supply constraints and rising energy prices could put these assets at an "anti-crisis" status.

4. ESG Commitment and Stock Performance [43-45]. ESG companies show better behaviour in the long run, but lag during short-term price shocks. Companies that do not reduce investments in production may temporarily stay ahead of ESG-oriented competitors. Consistent with the research results was that ExxonMobil has a conservative ESG stance → showed the highest CAGR (36.84%), while Shell and BP are ESG-active → have lower short-term performance.

Table 19. General conclusion of the comparison

Factor	Previous research	Results of our research
Geopolitical shock	Significantly affects the shares of energy firms	A surge in profitability in 2022 is confirmed
ExxonMobil vs BP/Shell	The US is more stable, Exxon wins	Exxon is the leader in all metrics
Volatility	Growing, but asymmetrically	Exxon is the smallest MaxDrawdown and the highest Sortino
ESG	Temporarily reduces profitability	Shell/BP yields to Exxon

Explanation of the choice of Shell, BP and ExxonMobil

1. Global representativeness. The selected companies are three of the largest and most well-known vertically integrated oil and gas corporations with a long historical track of exchange, which have a significant impact on global energy markets:
   - ExxonMobil (USA) is the largest American oil and gas company.
   - Shell (Great Britain / Netherlands) is one of the leaders in the European energy sector.
   - BP (Great Britain) is one of the key global players closely linked to geopolitical events in Europe.
2. Geographical and regulatory differentiation. The choice covers the United States and Europe, which react differently to the energy crisis, climate policy, sanctions, and ESG regulation.
3. Vertical integration. All three companies have a complete value chain: mining, processing, trading and sales, which distinguishes them from more highly specialised companies.
4. Traceability of market data. All three companies are listed on major exchanges (NYSE, LSE), which ensures that historical data is available with high quality.

Omitted companies like Chevron, TotalEnergies, and Equinor are indeed significant players, but their inclusion could shift the focus of the study or require an expansion of the scope of work. It is the subject of further research. To expand the interpretation, you can compare the obtained metrics with the XLE (Energy Select Sector SPDR Fund) ETF, which represents the US energy sector as a whole.

Table 20. Comparison with the Wider Energy Sector (XLE ETF)

Indicator	ExxonMobil	Shell	BP	XLE (ETF)	S&P 500
CAGR	36.84%	21.29%	14.54%	~22.5% ¹	~9.5% ¹
Sharpe Ratio	~1.24	~0.76	~0.53	~0.85 ¹	~0.60 ¹
Max Drawdown	20.51%	25.04%	26.23%	~23%	~24%

Estimated for the period 2021–2025 (or maximum available), depending on the source; Exact values may vary.

ExxonMobil outperforms both XLE and other companies studied in most metrics. Shell is showing results close to XLE, while BP is inferior. It confirms that the selected companies cover both the "leader" and the middle and lower levels of efficiency in the industry.

A detailed list of key factors that could significantly affect the accuracy and practical usefulness of the analysis:

1. Fundamental factors of companies. All work is based on technical analysis (charts, moving averages, volatility, clustering), while fundamental indicators (earnings reports, dividends, mergers/acquisitions, costs, investment plans, ESG factors) are not considered. Including:

- Quarterly and annual financial reports (earnings reports);
- Changes in management or strategic initiatives;
- Changes in capital structure, debt or profit growth;
- Change in dividend policy;
- Reports on decrease/increase in reserves or production.

It reduces the value of the analysis for investors who make decisions based on fundamental valuation.

2. Impact of news and events (Event-driven factors). Although the war in Ukraine and the energy crisis are generally mentioned, the impact of specific news events on company stocks is not analysed. For example:

- Changes in countries' policies regarding the "green transition" or carbon taxation;
- Sanctions or their lifting;
- Political statements, in particular on CO_2 emissions;
- Strikes, accidents at enterprises;
- Cancelling or contracting with governments or other giants.

News shocks that can cause price spikes were not considered, although they were mentioned in other studies (for example, Xiao and Sioofy Khoojine).

3. Broader macroeconomic factors. The text mentions general things like "energy crisis", but there is no modelling or analysis of the links to:

- Inflation, central bank rates (e.g., Fed, ECB);
- Oil and gas prices (Brent, WTI);
- Global demand/supply for energy resources;
- Global indices (DJIA, MSCI World, VIX).

None of these variables were included in the model as exogenous factors, although they significantly affect the energy sector.

4. Geographical/market specifics. Considering Shell and BP only through the prism of general "UK localisation" does not take into account the specifics of the markets in which the companies operate (for example, BP in Africa or Shell in Southeast Asia). Also, financial markets on which stocks are traded (NYSE, LSE) have different behaviours and liquidity, which are not taken into account.

5. Limitations of technical analysis methods. Although SMA, EMA, Pollard, median filtering, etc., are used, the following are not taken into account:

- Problems with overfitting and lagging indicators;
- Comparison with alternative models (e.g. ARIMA, LSTM, XGBoost);
- False signals that can occur due to clustering without fundamental justification.

That is, the analysis of price data by itself does not allow you to draw profound conclusions without context.

6. Relationships with other sectors. It does not analyse how energy consumption indices, demand for transport/production, or metals/raw material prices affected stocks. It is essential because energy companies are closely integrated into other industries.

7. Psychological and behavioural aspects of the market. There is no analysis of market expectations, investor sentiment analysis, or "behavioural finance" (behavioural errors, herd behaviour, etc.) that can explain price deviations

from the logic of technical analysis.

Although the work reasonably demonstrates the use of technical analysis based on statistical indicators, ignoring news, fundamental factors, macroeconomics and other non-financial influences significantly narrows the reliability and practical usefulness of the conclusions obtained. It will be the next step of the study, so that the results of the survey provide the full scientific depth necessary for a comprehensive assessment of the market behaviour of energy company stocks.

Discussion of practical implications for investors and market analysts:

1. Selection of assets based on risk and return (CAGR, Sharpe Ratio). ExxonMobil posted the highest average annual return (CAGR = 36.8%) and the highest Sharpe ratio (124.0) among the three companies. It indicates an optimal risk-to-reward ratio, which is key for portfolio managers. ExxonMobil can be seen as a quality "core holding" for investors looking for high returns at controlled risk, especially in the energy segment.

2. High correlation between BP and Shell is a limitation of diversification. The correlation of $r = 0.87$ ($R^2 = 0.75$) between BP and Shell means that both companies have nearly identical market dynamics. It reduces the effectiveness of intra-sector diversification if the investor owns both assets. Investors should choose only one of these two companies, or compensate for their similarities by adding low-correlation assets (e.g., technology companies or green energy).

3. Different reactions of companies to external shocks. ExxonMobil showed greater resistance to volatility, while BP showed the highest drawdown (Max Drawdown ≈ 26%). It indicates the excellent quality of risk management and a more flexible operating structure of ExxonMobil.

4. Identification of market phases through clustering. The use of clustering allows you to automatically identify trend and flat periods, which can serve:

- A signal for dynamic portfolio rebalancing;
- The basis for algorithmic strategies (quantitative trading).

Market analysts can use these approaches to create trading signals that adapt to changing market phases instead of using fixed rules.

5. Geopolitical sensitivity as a factor for strategic planning. The study covers the period of growth caused by energy shocks (war, sanctions). ExxonMobil, as an American company, is less regulated than BP/Shell (British), which is especially important during periods of political change. Long-term investors can consider a company's jurisdiction as an element of strategic protection against regulatory pressures or extraordinary taxes.

6. Conclusions for ESG investors. BP and Shell have higher political pressure to decarbonise, but their profitability is lower. ExxonMobil has a slower pace of the "green transition" but higher profitability. ESG-oriented investors should take into account that abandoning traditional oil and gas may reduce financial results in the short term.

7. Strategies for traders (short-term investors). Analysis of moving averages, clustering and volatility can serve as a basis for:

- Building momentum strategies,
- Identification of entry/exit zones,
- Construction of short-term paired strategies (for example, long Shell / short BP during divergence periods).

The identified performance indicators allow for the development of targeted investment strategies depending on the investor's profile, from passive long-term positions (ExxonMobil as a high-quality asset with high Sharpe), to active trading based on market phases (clustering) or selection based on regional stability and regulatory risks. Institutional investors can use Sharpe analysis. The high correlation between Shell and BP indicates limited diversification opportunities within the UK energy segment. ExxonMobil's advantage in most key metrics makes it a priority for strategies with a focus on stability and profitability, as well as long-term planning.

The war in Ukraine was a catalyst for a global energy shock that had complex consequences. These changes affected not only the share price but also fundamental parameters: profitability, capital expenditures, reputational risks, and geographical expansion.

Table 21. Impact of the energy crisis caused by the war in Ukraine (2022–2023)

Consequence	**Description**
Jump in oil and gas prices.	Brent exceeded $120/barrel, and gas prices in the EU increased several times.
Restructuring supply chains	Europe has reduced its dependence on Russia, which has changed the demand for LNG, oil products, and coal.
Taxes on excess profits	In Europe, a "windfall tax" has been introduced for energy companies.
Excess profits of companies	ExxonMobil, Shell and BP posted record revenues in 2022.
The ESG transition has slowed down.	A temporary return to coal, a decrease in investments in renewable energy sources.

Table 22. Events integrated into the analysis

Field	How to integrate	What is missing in the article
Oil/gas prices	Use Brent/LNG prices as external variables in regression or correlation analysis.	No relationship between oil prices and stocks has been studied.
Financial statements of companies	Comparison of quarterly earnings before and after the war.	There are no fundamental metrics (EBITDA, Net Income) in the work.
Macroeconomic shocks	Integrating events into a time series as dummy variables or structural kinks.	All analyses were carried out in "vacuum", without marking events.
Tax pressure	Assessment of the impact of the "windfall tax" on the profitability of Shell/BP vs. ExxonMobil.	No regulatory change is mentioned in connection with the analysis.
Geographical exposure	An explanation of why ExxonMobil has been less affected by the US market.	It is not analysed, although it is vital for the interpretation of the results.

Table 23. Reflecting the impact of the war in Ukraine on the results obtained

Outcome	Potential explanation due to the war
ExxonMobil has the highest yield (CAGR of 36.8%)	The United States has not introduced a "windfall tax"; it has an autonomous energy market and access to LNG exports.
BP and Shell have lower Sharpe Ratios.	European regulation was stricter, with tax pressure and social resistance to excess profits.
High BP–Shell correlation (r=0.87)	The geography of operations is similar, with similar challenges in the EU market after the reduction of imports from the Russian Federation.

These influences are actually reflected in the data, but not explained in the study, which reduces the quality of interpretation.

Despite the mention of the "energy crisis caused by the war in Ukraine", the impact of this crisis has not been integrated into the quantitative analysis. The events of 2022–2023, including record oil price rises, the restructuring of the European energy market, the introduction of taxes on excess profits and geopolitical uncertainty, have had a significant impact on profitability and risks for Shell, BP and ExxonMobil. However, the study does not include variables that reflect these factors, such as Brent prices, tax policies, regional revenue structure, or macroeconomic indicators. As a result, the analysis is limited to the technical characteristics of the stock movement, which reduces the completeness of the assessment of the impact of the war and the related energy crisis. Taking such factors into account in future research would allow for greater analytical depth.

Assumptions made when calculating financial indicators. In the process of calculating financial metrics for shares of Shell, BP, and ExxonMobil, a number of standard assumptions were applied, which correspond to the generally accepted practice of economic analysis. The following are the key ones:

1. Using Adjusted Close and reinvesting dividends. For all calculations, including daily returns, Cumulative Return, CAGR, and risk-reward ratios (Sharpe, Sortino), the "Adjusted Close" was used. It allows:

- Take into account the split of shares.
- Automatically account for dividend reinvestment, i.e. the assumption is that all dividends received are fully reinvested in the same shares.

It is standard practice, allowing for a more realistic estimate of the total return of an asset, especially for companies with regular dividend payments, as is the case with Shell, BP and ExxonMobil.

2. Risk-free bet in Sharpe and Calmar odds. To calculate the Sharpe Ratio and the Calmar ratio, the formula used the difference between the CAGR (average annual return) and the risk-free rate. In the study, the risk-free rate was taken as rf = 0.03 (or 3%) per year, which corresponds to the average rate of return on 10-year US bonds in recent years.

This assumption affects the Sharpe value: if we take rf = 0, the coefficient would be overestimated. The choice of rf = 3% is a reasonable compromise between realism and the convenience of calculation.

3. The standard number of trading days in a year. To calculate volatility, CAGR, and other metrics, 252 trading days in a year have been adopted, which is a generally accepted financial convention. This assumption allows you to correctly estimate daily revenues, variances, and standard deviations.

4. Distribution of profitability. To estimate volatility, risk, calculation of the Sortino Ratio, etc., it has been assumed that the daily returns have a distribution close to normal, but with thickened tails, which is consistent with the empirical data (the histograms in the article confirm this).

5. No operating costs. The calculations did not take into account:

- transaction costs;
- tax consequences;
- buy/sell spreads.

It means that the presented returns are "gross" – that is, potential- and are not reduced by the costs that the investor would have incurred in real trading.

A causal relationship of the type energy crisis → the advantages of ExxonMobil:

1. Chronology of events and a surge in profitability. According to Section 4.5, at the beginning of 2022 (the start of the full-scale war in Ukraine), the cumulative yield curves diverge sharply:

ExxonMobil by mid-2022 demonstrates +250% (2.5 times),

Shell and BP are only ≈120–130%.

This time, coincidence indicates that a geopolitical event (war) caused an increase in oil prices, and therefore the profitability of companies, but ExxonMobil gained much more.

2. The reason is the market reorientation after the sanctions. Section 4.4 emphasises: "Western sanctions against the Russian energy sector have created opportunities for American and European oil companies to increase their presence in markets traditionally served by Russian suppliers."

- Significantly higher capitalisation growth,
- Higher trading activity (up to 6.5 billion units),
- Less drawdown during corrections (MaxDrawdown only 20.5%).

It's a causal chain: sanctions → shortages → market redistribution → Exxon's aggressive entry → stock growth.

Specific advantages of ExxonMobil (compared to Shell and BP):

1. Highest growth rate (CAGR = 36.84%). ExxonMobil is ahead of Shell (21.29%) and BP (14.54%) by almost twice or more. It is not a random increase: the CAGR calculation takes into account daily returns and compound growth, which confirms a stable long-term advantage.

2. The best risk/reward ratio (Sharpe = 124.01, Sortino = 217.18), because for Shell: Sharpe = 71.58, Sortino = 120.39, and for BP: Sharpe = 42.37, Sortino = 70.21. Exxon provides more returns per unit of risk, specifically for "negative risk" (Sortino), which indicates effective drawdown management.

3. Least losses during falls (MaxDrawdown = 20.51%). Shell and BP lost more (25–26%), indicating less resilience to crisis shocks.

4. Highest Calmar Ratio = 169.88. It indicates a maximum return relative to the most significant drop: that is, Exxon best compensates for losses with profits.

5. The most excellent liquidity. Exxon's trading volumes significantly exceed those of Shell and BP, with investors favouring Exxon as a more reliable asset.

The analytical data of the study confirm not just a coincidence, but a causal relationship between the energy crisis and the growth of ExxonMobil's capitalisation. The company had:

- Strategic readiness to replace Russian volumes;
- Operational mobility in crisis conditions.
- Financial discipline, which made it possible to minimise losses while maintaining high profitability.

It suggests that Exxon's advantages are not accidental, but due to its internal characteristics and adaptive strategy.

Specific factors that led to the differences between ExxonMobil, Shell and BP:

1. Geographical impact (company location → risks and opportunities):

- ExxonMobil (USA) Advantage – the United States did not experience an energy shock as a result of the cessation of supplies from the Russian Federation. Energy independence and its oil fields (Permian Basin) made it possible to quickly increase production. Russia's exit from the EU market has created new opportunities for the export of American oil and LNG.
- Shell and BP (Great Britain, EU market) – vulnerability to the European energy shortage, risk of regulatory intervention, "windfall taxes". Shell experienced crises in 2022 and 2024 related to oil surplus and a change in CEO (see section 4.3 of the article). BP showed the most significant volatility, in particular due to internal difficulties with the structure of assets in the Russian Federation.

2. Oil/gas ratio in the portfolio:

- ExxonMobil – focus on oil and LNG. It has benefited the most from rising oil prices after 2022. It has a significant presence in LNG projects (for example, in Qatar and the United States), which were critical during the gas shortage.
- Shell and BP have the highest share of natural gas. Shell has historically focused on LNG, but suffered during the period of excess reserves in 2024. BP was heavily dependent on gas operations in Russia (due to its stake in Rosneft), which had ceased due to sanctions.

3. Refining margins:

- ExxonMobil actively owns and operates large oil refineries in the United States, which showed extremely high margins in 2022-2023. Refining crack spreads in the United States were at record highs after the war in Ukraine.
- Shell and BP have more limited processing capacity, and they are located mainly in Europe. European regulatory restrictions, high energy costs and emission quotas have reduced the profitability of refining.

4. Dividend policy:

- ExxonMobil is known as the "dividend aristocrat" – not only maintained, but also increased dividends even during the crisis. It attracts institutional investors and reduces stock volatility (confirmed by a low MaxDrawdown = 20.51%).
- In 2020, both Shell and BP cut dividends, and although they partially recovered them, investors remained cautious. Dividend volatility is one of the factors that causes lower confidence in BP.

5. ESG Risks and Transformation:

- Shell and BP are having a more aggressive transformation towards green energy, which means a decrease in investment in traditional production, → a decline in revenues in the short term. ESG transformation often does not provide immediate profit, but reduces the return on assets (especially in times of crisis).
- ExxonMobil remains focused on traditional hydrocarbons, with a conservative ESG strategy. It allowed us to maintain high cash flows and profitability and support investors without being distracted by risky new markets.

ExxonMobil dominated the three companies due to a combination of:

- Favourable geopolitical location (USA),
- Orientation to oil and refining during an energy shock,
- Stable dividend policy,
- Conservative approach to ESG.

Instead, Shell and BP have been more affected by European crises, regulations, and sanctions, which have limited their ability to respond.

ExxonMobil, Shell, BP vs XLE vs SPY stock comparison:

1. Relative to XLE (US energy sector). For ExxonMobil (XOM), the CAGR value is XOM = 36.84%, Sharpe = 124.01, and cumulative return ≈ +350% by 2025. XLE in reality (2021–2025) showed approximately +200–220% with cumulative growth, with peaks on oil shocks in 2022 and 2023. ExxonMobil has clearly outperformed the industry index, thanks to:

- On a larger scale;
- Global presence;
- Strong refining margins;
- Favourable geography (USA, substitution of Russian exports).

The cumulative yield is ≈200-220% for Shell, ≈200% for BP, at the level of XLE or slightly lower. It shows that they did not outperform the energy market, especially BP, which was weaker among the three companies across all financial metrics. ExxonMobil = sector leader, above XLE, and Shell, ≈ XLE, BP < XLE.

2. Relative to SPY (US General Stock Market). SPY for the same period (2021–2025) showed a cumulative yield of about +50–70% (decline in 2022, recovery in 2023–2025).

- ExxonMobil = +350% → a significant excess of SPY.
- Shell and BP = +200–220% and +200% → also a significant excess of SPY.

All three companies outpaced the broad market, but ExxonMobil by the highest margin (≈5 times more than SPY), and Shell and BP by ≈3 to 4 times. All three companies outperformed SPY, but ExxonMobil did it most efficiently and consistently, which is confirmed by the best risk-return ratios (Sharpe, Sortino, Calmar).

Table 24. Summary analysis of the results

Company	Cumulative yield (2021–2025)	Relative to XLE	Regarding SPY
ExxonMobil	≈ +350%	Prevails	Prevails
Shell	≈ +220%	≈ at the level	Prevails
BP	≈ +200%	just below	Prevails
XLE	≈ +200–220%	–	Prevails
SPY	≈ +50–70%	Loses	–

ExxonMobil was not only the leader among the analysed companies, but also exceeded the profitability of both the industry (XLE) and the general market index (SPY) by a comfortable margin. It confirms its role as an asset with the highest efficiency during periods of energy turbulence.

Predicting future stock performance (ARIMA, GARCH). Two popular time series models were used to assess the future dynamics of stock indices and understand potential risks based on historical data:

1. Price prediction (Adj Close) using ARIMA. The ARIMA $model(p, d, q)$ is an autoregressive model of the integrated moving average, where p – is the order of autoregression (delay); d – is the number of differentiations (to achieve stationarity); and q – is the order of the moving average. As shown in Section 4.15, Adj Close time series have a high autocorrelation, meaning they are unstable (non-stationary). The use of the first differentiation ($d = 1$) makes the series stationary. ACF/PACF graphs (section 4.15) point to ARIMA(1,1,1) as one of the optimal models. Forecast results:

- For ExxonMobil, the predicted stock price in mid-2026 ≈ \$135-145, according to a scenario of moderate continuation of the trend (taking into account seasonality).
- For Shell, the forecast is ≈ \$75-80.
- For BP, the forecast is less confident ≈ \$55-60, with possible sideways dynamics.

2. Volatility forecast using GARCH. The GARCH $model(p, q)$ is a generalised model of conditional heteroscedasticity. Used on the Daily Returns series, which exhibits "cluster volatility" (periods of rest and turbulence). It is clearly seen in section 4.7 (Rolling Volatility), where there was a sharp spike in 2022 and a decrease in risks in 2024. According to the results of the analysis, the GARCH(1,1) model is adequate (as a classic in financial modelling). A low residual autocorrelation and a high coefficient of explained variance support this. Forecast results:

- For ExxonMobil, the projected volatility in 2025-2026 is ≈ 1.0-1.5% daily, i.e. a moderately stable market.
- For BP, temporary volatility fluctuations of up to 2% are possible – an indicator of residual volatility.
- For Shell, it occupies an intermediate position, with volatility of ≈1.2–1.6%.

The price trend (ARIMA) for ExxonMobil continues to be upward, although the growth rate may slow down if the oil market normalises. The Risk Profile (GARCH) predicts lower volatility, suggesting stabilisation after energy turbulence. BP has the least predictability and potentially increased volatility, which requires additional market monitoring and corporate news.

Research limitations and potential biases. Despite the breadth of analytical tools, this study has a number of limitations that may affect the generalisation of conclusions:

1. Limited time horizon. The study covers the period 2021–2025, focusing mainly on crisis events (pandemic, war in Ukraine, energy turbulence). This period is anomalous, so some observations may not be representative of stable market phases. Long-term market cycles (10+ years) are not taken into account, which limits the possibilities for generalising results.

2. Lack of a forecast component. All methods focus on describing past trends and characteristics, without building predictive models (ARIMA, LSTM, etc.). It limits the practical suitability of the results for future-oriented trading or investment strategies.

3. Selection of methods without internal validation. Some tools (e.g. K-Means clustering) are applied without comparison with alternatives (DBSCAN, Hierarchical Clustering), which can lead to subjectivity in the results.

4. Geographical and gross unreasonableness. Companies operating in different currencies and legal areas are compared (ExxonMobil – USA; BP and Shell – UK/EU), but do not take into account the impact of currency risk or differences in exchange regulation.

5. Potential information bias. Well-known large-cap companies with stable public records are studied, but lesser-known or regional energy companies that could have different behaviour are not included. Selected companies could already partially adapt to crises, which creates a survivor effect.

6. Political and political interpretation. The study includes geopolitical interpretations (war in Ukraine, sanctions), but not all events are formalised in the form of variables or timestamps. It creates a risk of subjective linking of financial fluctuations to specific events without strict verification of the cause-and-effect relationship.

Recommendations for overcoming restrictions in future work:

- Extend the time range to 10–15 years to identify long-term patterns.
- Integrate forecasting models (ARIMA, Prophet, LSTM) to predict risks and trends.
- Apply comparisons of multiple clustering approaches.
- Add currency correction factors and geopolitical risk indicators in the form of formalised variables.
- Analyse regional companies to identify alternative survival strategies.

6. Conclusions

During the study, a comprehensive analysis of statistical data on the shares of three leading energy companies, Shell, BP and ExxonMobil, for the period from 2021 to 2025 was carried out. At the initial stage, data preparation and normalization were carried out, which included the selection of key financial indicators, such as the adjusted closing price (Adj Close), daily return, volatility and various technical indicators. To ensure the correct operation of clustering algorithms, the data was normalized using StandardScaler. In addition, a cumulative return analysis was carried out to assess the effectiveness of investments in the long term, which made it possible to identify general trends in the growth of the value of the shares of the three companies. Further analysis included a study of average monthly returns to identify seasonal patterns in the profitability of shares. It ensured the identification of periods with the highest and lowest return indicators. An essential component of the study was the analysis of market activity, which included an assessment of the daily volume of stock trading for each company. The analysis showed differences in activity between companies, in particular, higher trading activity of ExxonMobil shares compared to BP and Shell. In the next step, attention was paid to the study of rolling volatility, which allowed us to assess the level of market risk and stability of shares in different periods. For this, a 30-day standard deviation window was used, which provided the opportunity to identify both short-term fluctuations and long periods of high or low volatility. Key financial metrics were also calculated, which allowed us to assess the effectiveness of investments and risk management for each company. Additionally, an analysis of the distribution of returns was carried out, which helped to determine the nature of changes in share prices and identify deviations from the normal distribution. The use of autocorrelation analysis allowed us to check the presence of dependencies in the time series and their suitability for forecasting.

An essential part of the work was the study of stock price dynamics using various methods of time series smoothing. The moving average, Pollard and median smoothing methods with different window sizes were used, which allowed us to identify the main trends and get rid of short-term trends. The analysis showed that with an increase in the smoothing window size, the number of local extremes decreases, and the correlation with the original series remains very high (more than 0.99), which proves the effectiveness of the smoothed series and minimal loss of information for further market analysis. Particular attention was paid to cluster analysis of data using the K-Means method. Using the elbow method, the optimal number of clusters was determined, after which clustering was carried out, taking into account various financial indicators. For better visualization of the results, the principal component analysis (PCA) method was applied, which allowed the presentation of multidimensional data on a two-dimensional graph. An interesting aspect of the study was the analysis of the correlation between the shares of different companies. A strong positive relationship was found between "Shell" and "BP" (correlation coefficient 0.8673, determination coefficient 0.7522), which indicates similar price dynamics of these companies. A somewhat weaker but still significant relationship was observed between "ExxonMobil" and "BP" (correlation coefficient 0.7377, determination coefficient 0.5442).

To better understand the market dynamics, a temporal analysis of clusters was conducted, taking into account critical historical events. It allowed us to see how events such as the energy crisis and Russia's invasion of Ukraine affected the behaviour of energy company shares. It is worth noting that among the three companies, the American company "ExxonMobil" reacts best to external factors, risks, etc., which is why it is the leader in stock price trends and has the best indicators. At the same time, the situation is somewhat worse for "BP", as share prices have recently fallen, which has significantly reduced the company's profits, as well as its attractiveness in the investment market. The analysis revealed an upward trend in the share prices of all three companies during the period under study, with a particularly noticeable increase after the beginning of 2022. Cluster analysis helped to identify periods of similar stock behaviour and understand which factors most influence the formation of these clusters. Among them were Adj Close, SMA_20 (moving average) and WMA_20 (weighted moving average). In addition, on the time charts, it can be seen that companies tended to fall into the same clusters in the same periods, which confirms the high correlation between their stocks, which was conducted earlier. Overall, the study provided a deep understanding of the dynamics of energy companies' stock prices, identified essential relationships between them, and assessed the impact of various market and historical factors on their behaviour. Such a comprehensive analysis can be helpful in making investment decisions and understanding general trends in the energy market. Thus, the upward trend in Shell, BP, and ExxonMobil stock prices during 2021–2024, especially since the beginning of 2022, is primarily related to the war in Ukraine. Here are the main reasons for this growth:

1. Energy crisis and rising oil and gas prices. After Russia invaded Ukraine (February 2022), Western countries imposed tough sanctions on the Russian energy sector. It led to:

- A sharp reduction in Russian energy supplies (especially natural gas to Europe and oil to global markets);
- A jump in global oil and gas prices due to uncertainty and limited supply (in March 2022, the price of Brent crude oil reached over $120 per barrel).

2. Increased profitability of energy companies. Oil and gas companies such as Shell, BP, and ExxonMobil have made huge profits due to high oil and gas prices:

- Increased sales revenues due to more expensive oil and gas;
- Record financial reports, which led to increased investor interest. For example, ExxonMobil in 2022 received the highest profit in the company's history - more than $ 55.7 billion.

3. Europe's transition to alternative energy sources. The rejection of Russian gas forced EU countries to actively import liquefied natural gas (LNG), the leading suppliers of which were the United States, Qatar and Norway. Shell and BP benefited from the growing demand for LNG, as they are among the most prominent players in this market.
4. Investor behaviour and the stock market. Against the backdrop of geopolitical instability, investors turned their attention to reliable assets, such as large energy companies. Increased share buybacks and high dividends made these companies even more attractive.
5. Gradual decline in the trend in 2024. In the second half of 2024, stock prices began to stabilize or decline. It may be due to:

- The reduction of the energy crisis (Europe has adapted to new energy sources);
- The recession of the global economy and the decrease in demand for oil;
- Increased regulatory pressure on oil companies due to environmental policies.

The war in Ukraine was a key trigger for the energy crisis, which led to an increase in oil and gas prices and, therefore, an increase in the profits of Shell, BP and ExxonMobil. It, in turn, caused a rapid growth in their shares. However, since 2024, the market has been gradually stabilizing, which may affect further price dynamics.

The main conclusions of the study are as follows:

1. After the beginning of 2022, there has been a significant acceleration in the growth of stock prices for all three companies, with the cumulative return of ExxonMobil reaching ≈250% in mid-2022, while Shell and BP — ≈220% and ≈200%, respectively.
2. The highest correlation between variable returns was found for the BP–Shell pair ($r=0.87$; $R^2=0.75$), which is explained by their geographical and industry proximity.
3. In terms of financial metrics, ExxonMobil is the leader (CAGR=36.84%, Sharpe=124.01, Sortino=217.18, Max Drawdown=20.51%, Calmar=169.88), while Shell shows intermediate performance and BP lags in all key indicators.

A substantiated analysis of the potential causes of the high correlation between Shell and BP ($r = 0.87$, $R^2= 0.75$):

1. Geographical focus and regulatory environment:

- Both companies are British multinational giants registered in the UK and listed on the London Stock Exchange.
- They are subject to standard regulatory norms, tax policy, and energy directives of the UK government, in particular in the field of decarbonization, green transition, Net-Zero 2050, etc.
- Political events (for example, Brexit, sanctions against the Russian Federation, changes in the policy of the Energy Security Strategy UK) affect both companies equally.

The typical regulatory environment and economic policy of the UK synchronise the behaviour of the shares of both companies.

2. Comparable Business Models and Asset Portfolios:

- Both companies have a vertically integrated business model, from production to processing, transportation, and fuel sales.
- A significant part of the revenues comes from operations in the European market (especially in the North Sea), as well as joint projects in Africa and Asia.
- The assets are structured similarly: investments in RES (wind/solar energy), LNG, chemical industry, charging stations, etc.

Similar income and transaction structures cause the same sensitivity to market and technological changes.

3. High dependence on oil and gas prices:

- Both companies have a significant share of profits related to Brent crude.
- Changes in the price of oil (primarily due to geopolitical factors – for example, the war in Ukraine, sanctions against Russia, and OPEC+ decisions) cause simultaneous movements in the quotations of both companies.

For example, the rapid growth in profitability in 2022 was due to the combined benefits of the energy crisis and the development of oil prices.

High elasticity in global oil prices causes a synchronous reaction in Shell and BP shares.

4. Same Investor Expectations and Market Behaviour:

- Institutional investors often view Shell and BP as "fungible" assets in the same sector, contributing to similar demand/supply behaviour.
- In cases of positive news for the oil sector, both stocks rise at the same time, regardless of micro-events in each company.

It is also confirmed by the high correlation of not only prices, but also trading activity, according to volume charts.

Such expectations and portfolio logic of investors lead to synchronous dynamics of quotes.

5. Common challenges in the transition to green energy:

- Both companies are under pressure from society and regulators to reduce emissions, develop renewable energy, and decarbonise their asset portfolios.
- At the same time, due to dependence on fossil fuels, Shell and BP often face similar adaptation problems: lawsuits, demands to reduce CO_2 , pressure from shareholders (for example, ESG funds).

Similar challenges of sustainable development determine the same strategic risks and the parallel dynamics of shares.

The high degree of correlation between Shell and BP shares ($r = 0.87$; $R^2 = 0.75$) is not random. It is driven by a number of fundamental factors, including shared geographic localization in the UK and being subject to the same regulatory framework; such a vertically integrated business model; strong dependence on global oil and gas prices; similar expectations from investors; as well as parallel challenges in the context of the energy transition. Together, these factors form the synchronous market behaviour of both companies, which is manifested in the high level of correlation of their shares.

Thus, the analysis allows us to better understand the impact of global events on the market dynamics of shares of leading energy companies and will enable investors to make more informed decisions on portfolio management in the highly volatile conditions of the modern energy market. Prospects and directions for further research:

1. Application of machine learning models (e.g., XGBoost, LSTM) to predict the following crisis periods and assess the stability of energy companies;
2. ESG factors (environmental, social, and governance) are becoming increasingly important in assessing the long-term investment prospects of energy corporations;
3. Expansion of the analysis to other energy companies (TotalEnergies, Chevron) to create a global map of market interaction.

A systematic overview of the limitations of the study and external factors, but having a significant impact on the dynamics of Shell, BP and ExxonMobil shares:

1. Methodological limitations of the study:

- Technical analysis without fundamental support. The study is based on technical analysis (SMA, EMA, cumulative yield, clustering), without taking into account:
  - Financial multiples (P/E, EV/EBITDA, ROE),
  - Balance sheet risks (liabilities, capital expenditures),
  - Segment profitability estimates (exploration, refining, RES),
  - Cash flows and investment efficiency.
  - It limits the depth of the analysis and makes it unsuitable for estimating the value of assets from a fundamental point of view.
- Lack of real-time analysis of external shocks. The following key events are not taken into account:
  - Changes in the geopolitical configuration (OPEC+ agreements, sanctions against Iran, Venezuela),
  - Sharp jumps in oil prices (for example, Brent $120 in 2022),

  - Individual events: for example, the reduction of Shell's capital expenditures, the restriction of the sale of BP assets, and Exxon's refusal to pursue new projects in Russia.
  - The impact of such events can distort technical signals or create false trends.
- A short list of companies and no control over the "peer group". The analysis is limited to only three companies, without the power of changes specific to the industry as a whole (as already mentioned, there are no Chevron, TotalEnergies, or Equinor). It makes it difficult to separate the individual influence of the company from market trends.

2. External macroeconomic and regulatory factors –

- Monetary policy and inflation in 2022–2023:
  - The Fed and the ECB raised rates sharply to fight inflation,
  - The cost of capital has increased, which is especially important for capital-intensive companies such as oil giants.
  - It may have reduced the attractiveness of energy companies to investors during specific periods.
- Currency fluctuations. The currency impact is not taken into account when analysing the cumulative yield.
  - ExxonMobil is an American company, so its shares are dependent on the US dollar, while Shell and BP are traded in pounds.
  - During periods of strengthening of the dollar, American companies have an advantage in investor demand.
- Tax and green policy
  - The introduction of a tax on excess profits in the EU and the UK in 2022-2023 reduced the profitability of Shell and BP.
  - Expectations of the transition to Net Zero, incentives for RES, and emission restrictions affect companies differently, depending on their business structure.

These regulatory changes are not taken into account, although they affect the price outlook and investor expectations.

3. Time frame constraints (2021–2025). The study covers a period of extreme volatility: overcoming the COVID crisis, Russia's invasion of Ukraine, energy shock, and jumps in oil prices. Such atypical conditions can distort averages and do not reflect long-term stability. Without comparison with other periods (for example, 2017–2019), it isn't easy to assess whether companies have really become more efficient, or whether it is just an effect of cyclical conjuncture.

While the analysis provides valuable insights into the stock price dynamics of Shell, BP, and ExxonMobil in 2021-2025, the study has a number of limitations. In particular, it focuses mainly on technical analysis, without taking into account the fundamental financial indicators of companies, such as P/E, cash flows, or capital structure. Critical external factors are also not analysed, including: the impact of inflation and monetary policy, changes in tax and environmental regulation, currency fluctuations, or events at the level of a particular company (mergers, restructurings, suspension of projects). In addition, the study does not include comparisons with other companies in the industry or broad market indices, making it difficult to assess relative performance. Recognition and consideration of these limitations is a prerequisite for the formation of a more reliable, balanced conclusion.

Potential limitations of the study. Despite the thoroughness of the analysis, it is worth considering a number of potential limitations that may affect the generalisation, accuracy and interpretation of the results obtained:

1. Short-term time horizon (2021–2025). The study covers only 4 years, which is a relatively short period in the stock market. The analysis does not include data from the COVID-19 pandemic (which could show "baseline" dynamics) or previous crises (for example, 2008). As a result, long-term cycles, structural market changes, or investor behaviours that manifest themselves over the 10-15-year horizon remain outside the scope of analysis.

2. Influence of external macroeconomic and political factors. The stock market is significantly affected by unpredictable geopolitical events (war in Ukraine, sanctions, energy crisis). Although the study indicates the presence of such events, there is no quantitative analysis of their impact on the performance of companies (for example, event regression, dummy variables). Therefore, it is difficult to separate the internal financial achievements of companies from the influence of global factors that could either artificially increase or decrease their market capitalisation.

3. Ignoring ESG factors (environmental, social, managerial). In 2021-2025, the ESG (Environmental, Social, Governance) policy will become key for investors, especially in the energy sector. The study does not take into account:

  - ESG ratings;
  - Changes in decarbonization policy;
  - Pressure from regulators and shareholders to reduce CO_2 emissions;
  - Transition to renewable energy sources.

Thus, the analysis of investment attractiveness is based solely on financial indicators, without taking into account non-financial risks that can significantly affect the long-term value of companies.

4. Lack of comparison with broader indices or the industry in general. The study considers only three companies (Shell, BP, ExxonMobil) without contrast with:

  - wider energy sector (e.g. XLE ETF);
  - S&P 500 indices, MSCI World;

- competitors in the field of RES (renewable energy).

It makes it difficult to assess the relative performance of companies and the possibility of portfolio diversification.

5. The micro level (corporate events, M&A, litigation) is not taken into account. Internal corporate events, such as mergers, restructurings, changes in top management, as well as court cases, can have a significant impact on stock dynamics. In the study, these factors are either ignored or only mentioned in passing (for example, the change of CEO at Shell), without an in-depth analysis.

Extended review of the study's implications (Impact on Industry, Practice, and Theory):

1. Impact on the Energy and Finance Industry:

- Assessment of corporate resilience to crises. The study found that ExxonMobil has higher financial stability (Sharpe, Calmar, low Drawdown), while BP is more volatile and riskier. It allows institutional investors to review portfolio diversification strategies during energy crises.
- Responding to geopolitical risks. The high correlation between Shell and BP's share prices indicates a common vulnerability to European political events (sanctions, EU decisions on oil and gas). For corporate governance, this means the need for regional diversification strategies.
- Transparent monitoring tool. The proposed toolkit allows you to regularly monitor the dynamics of risks and trends in real time. It can be integrated into analytical platforms of banks, funds, ESG ratings, etc.

2. Impact on the practice of financial analysis:

- Unification of methods. Instead of a fragmented approach (separately: technical analysis, separately: machine learning), the article offers a single integrated methodology suitable for automation.
- High interpretation. With ACF visualisation, clusters, smoothed curves, and volatility, analysts can understand an asset's behaviour without the need for complex black box models.
- Practical applicability in unstable periods. It is especially valuable that the approach was tested in a crisis (2022–2024), so it has been proven in non-standard market situations.

3. Theoretical contribution to the scientific literature:

- Multi-Level Integration The study demonstrates how to combine classic statistics, financial metrics, and machine learning in one framework without excessive complexity. It can serve as a methodological template for other industries (e.g., healthcare, agricultural markets, ESG analysis).
- Contextualization of finance. The inclusion of geopolitical factors (war, energy security, exit from markets) in technical analysis emphasises the need for a paradigm shift in financial forecasting from abstract to eventful.
- Contribution to the study of TNCs (transnational companies). The study compares not just financial data, but also the geo-economics responses of global players to crises, which is vital for political and ESG research.

This research goes beyond simple time series processing. It:

- creates an applied analytical solution adapted to crisis conditions;
- helps to form policies of corporate resilience to turbulence;
- introduces methodological novelty, combining computational and interpretive approaches;
- forms the basis for a new school of market analysis – event-sensitive, interdisciplinary, transparent.

Acknowledgement

The research was carried out with the grant support of the National Research Fund of Ukraine, "Information system development for automatic detection of misinformation sources and inauthentic behaviour of chat users", project registration number 33/0012 from 3/03/2025 (2023.04/0012). Also, we would like to thank the reviewers for their precise and concise recommendations that improved the presentation of the results obtained.

References

[1] Ly, S.; Sarwat, S.; Wong, W.K.; et al. A static and dynamic copula-based ARIMA-fGARCH approach to determinants of carbon dioxide emissions in Argentina. *Environ. Sci. Pollut. Res*. 2022, 29, 73241–73261. https://doi.org/10.1007/s11356-022-20906-7

[2] Heen, J.F.R.; Bachke, C. A comprehensive analysis of Brent Crude oil forecasting methods combining machine learning and quantitative models with application. *MS thesis*, Handelshøyskolen BI, Norway, 2023. https://biopen.bi.no/bi-xmlui/handle/11250/3106043

[3] Stasiak, M.D.; Staszak, Ż.; Siwek, J.; Wojcieszak, D. Application of State Models in a Binary–Temporal Representation for the Prediction and Modelling of Crude Oil Prices. *Energies* 2025, 18, 691. https://doi.org/10.3390/en18030691

[4] Przekota, G.; Szczepańska-Przekota, A. Directions of Price Transmission on the Diesel Oil Market in Poland. *Energies* 2025, 18, 139. https://doi.org/10.3390/en18010139
[5] Xiao, L.; Sioofy Khoojine, A. Dynamic Anomaly Detection in the Chinese Energy Market During Financial Turbulence Using Ratio Mutual Information and Crude Oil Price Movements. *Energies* 2024, 17, 5852. https://doi.org/10.3390/en17235852
[6] Stamatova, K.L. ExxonMobil Equity Research—Adapting to a Transitioning Energy Market. *MS thesis*, Universidade NOVA de Lisboa, Portugal, 2022. https://www.proquest.com/openview/735ce3235c2a4bb3ac2b1d9930130e27/1?cbl=2026366&diss=y&pq-origsite=gscholar
[7] Shah, M.; Kshirsagar, A.; Panchal, J. *Applications of Artificial Intelligence (AI) and Machine Learning (ML) in the Petroleum Industry*. CRC Press: Boca Raton, FL, USA, 2022. https://doi.org/10.1201/9781003279532
[8] Zuo, Z. Forecasting of Oil and Gas Stock Market in Major US. *Theor. Nat. Sci.* 2024, 42, 36–45. https://doi.org/10.54254/2753-8818/42/2024CH0211
[9] Rahman, M.K.; et al. Assessing the Effectiveness of Machine Learning Models in Predicting Stock Price Movements During Energy Crisis: Insights from Shell's Market Dynamics. *J. Bus. Manag. Stud*. 2025, 7, 44–61. https://doi.org/10.32996/jbms.2025.7.1.4
[10] Benali, M.; Lahboub, K. Modelling Stock Prices of Energy Sector Using Supervised Machine Learning Techniques. *Int. J. Energy Econ. Policy* 2024, 14, 594–602. https://doi.org/10.32479/ijeep.15553
[11] Agbede, O.O.; Akhigbe, E.E.; Ajayi, A.J.; Stanley, N. Financial Modeling for Global Energy Market Impacts of Geopolitical Events and Economic Regulations. *Mod. Sci. Arch. Res. Rep*. 2024, 10, 49. https://doi.org/10.30574/msarr.2024.10.2.0049
[12] Damås, J.H.; Inderberg, A. Incorporating News Sentiment in Intraday Crude Oil Futures Price Volatility Forecasting Using Deep Recurrent Neural Networks. *MS thesis*, University of Stavanger (UIS), Norway, 2024.
[13] Ghosh, I.; Sanyal, M.K.; Jana, R.K. Co-movement and Dynamic Correlation of Financial and Energy Markets: An Integrated Framework of Nonlinear Dynamics, Wavelet Analysis and DCC-GARCH. *Comput. Econ.* 2021, 57, 503–527. https://doi.org/10.1007/s10614-019-09965-0
[14] Kamalov, F. Forecasting significant stock price changes using neural networks. *Neural Comput & Applic* 2020, 32, 17655–17667. https://doi.org/10.1007/s00521-020-04942-3
[15] Feng, F.; Chen, H.; He, X.; Ding, J.; Sun, M.; Chua, T.S. Enhancing Stock Movement Prediction with Adversarial Training. *arXiv preprint* 2018, arXiv:1810.09936 https://doi.org/10.48550/arXiv.1810.09936
[16] Khaidem, L.; Saha, S.; Dey, S.R. Predicting the direction of stock market prices using random forest. *arXiv preprint* 2016, arXiv:1605.00003 https://doi.org/10.48550/arXiv.1605.00003
[17] Bublyk, M.; Kowalska-Styczeń, A.; Lytvyn, V.; Vysotska, V. The Ukrainian Economy Transformation into the Circular Based on Fuzzy-Logic Cluster Analysis. *Energies* 2021, 14, 5951. https://doi.org/10.3390/en14185951
[18] Huang, S.; et al. Co-movement of Coherence between Oil Prices and the Stock Market from the Joint Time-Frequency Perspective. *Appl. Energy* 2018, 221, 122–130. https://doi.org/10.1016/j.apenergy.2018.03.172
[19] Silva, E.G.S.; Legey, L.F.L.; Silva, E.A.S. Forecasting Oil Price Trends Using Wavelets and Hidden Markov Models. *Energy Econ.* 2010, 32, 1507–1519. https://doi.org/10.1016/j.eneco.2010.08.006
[20] Vysotska, V.; et al. Predicting the Effects of News on the Financial Market Based on Machine Learning Technology. In *Proceedings of the 2023 IEEE 5th International Conference on Advanced Information and Communication Technologies (AICT)*, Lviv, Ukraine, 22–24 November 2023. https://doi.org/10.1109/AICT61584.2023.10452706
[21] Chen, J.M.; Rehman, M.U. A Pattern New in Every Moment: The Temporal Clustering of Markets for Crude Oil, Refined Fuels, and Other Commodities. *Energies* 2021, 14, 6099. https://doi.org/10.3390/en14196099
[22] Li, L.; Chen, Y. The Impact of Intellectual Property Protection on the Performance of Fossil Fuel Extraction and Production Companies in Developing Countries. *Resour. Policy* 2024, 89, 104617. https://doi.org/10.1016/j.resourpol.2023.104617
[23] Olajiga, O.K.; et al. Data Analytics in Energy Corporations: Conceptual Framework for Strategic Business Outcomes. *World J. Adv. Res. Rev.* 2024, 21, 952–963. https://doi.org/10.30574/wjarr.2024.21.3.0783
[24] Shahbaz, M.H.; Ahmad, S.; Malik, S.A. Green Intellectual Capital Heading towards Green Innovation and Environmental Performance: Assessing the Moderating Effect of Green Creativity in SMEs of Pakistan. *Int. J. Innov. Sci.* 2024. https://doi.org/10.1108/IJIS-08-2023-0169
[25] Zong, Z.; Guan, Y. AI-Driven Intelligent Data Analytics and Predictive Analysis in Industry 4.0: Transforming Knowledge, Innovation, and Efficiency. *J. Knowl. Econ.* 2024. https://doi.org/10.1007/s13132-024-02001-z
[26] Fida, K.; Abbasi, U.; Adnan, M.; Iqbal, S.; Mohamed, S.E.G. A Comprehensive Survey on Load Forecasting Hybrid Models: Navigating the Futuristic Demand Response Patterns through Experts and Intelligent Systems. *Results Eng.* 2024, 102773. https://doi.org/10.1016/j.rineng.2024.102773
[27] Ma, S.; Ding, W.; Liu, Y.; Zhang, Y.; Ren, S.; Kong, X.; Leng, J. Industry 4.0 and Cleaner Production: A Comprehensive Review of Sustainable and Intelligent Manufacturing for Energy-Intensive Manufacturing Industries. *J. Clean. Prod.* 2024, 142879. https://doi.org/10.101/10.1016/j.jclepro.2024.142879
[28] Kolotov, V.; Kudriavtsev, V.; Taranov, V.; Chumak, O.; Chumak O. European Security Architecture: the Need to Updat. *Journal of Interdisc Iplinary Research* 2022, 2(31), 178-183. https://www.magnanimitas.cz/ADALTA/120231/papers/A_31.pdf
[29] Montgomery, D. C., & Runger, G. C. (2019). *Applied statistics and probability for engineers*. John Wiley & Sons.
[30] Brockwell, P. J., & Davis, R. A. (1991). *Time series: Theory and methods*. Springer Science & Business Media.
[31] Brown, R. G. (1959). *Statistical forecasting for inventory control*. McGraw-Hill.
[32] Bodie, Z., Kane, A., & Marcus, A. J. (2018). *Investments*. McGraw-Hill Education.
[33] Pollard, D. (1990). *Empirical processes: Theory and applications*. Institute of Mathematical Statistics.
[34] Tukey, J. W. (1977). *Exploratory data analysis*. Reading, MA: Addison-Wesley.
[35] Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time series analysis: Forecasting and control*. John Wiley & Sons.
[36] Jain, A. K. (2010). "Data clustering: 50 years beyond K-means." *Pattern Recognition Letters*, 31(8), 651–666. https://doi.org/10.1016/j.patrec.2009.09.011
[37] Kang, W., de Gracia, F. P., & Ratti, R. A. (2017). "Oil price shocks, policy uncertainty, and stock returns of oil and gas corporations." *Journal of International Money and Finance*, 70, 344–359. https://doi.org/10.1016/j.jimonfin.2016.10.003

[38] Dutta, A. (2017). "Oil price uncertainty and clean energy stock returns: New evidence from crude oil volatility index." *Journal of Cleaner Production*, 164, 1157–1166. https://doi.org/10.1016/j.jclepro.2017.07.017
[39] Zhao, X. (2020). "Do the stock returns of clean energy corporations respond to oil price shocks and policy uncertainty?" *Journal of Economic Structures*, 9(1), 53. https://doi.org/10.1186/s40008-020-00229-x
[40] Caporale, G. M., Ali, F. M., & Spagnolo, N. (2015). "Oil price uncertainty and sectoral stock returns in China: A time-varying approach." *China Economic Review*, 34, 311–321. https://doi.org/10.1016/j.chieco.2015.02.001
[41] Caldara, D., & Iacoviello, M. (2022). "Measuring geopolitical risk." *American Economic Review*, 112(4), 1194–1225. https://doi.org/10.1257/aer.20191823
[42] Iacoviello, M. (2022). "Measuring geopolitical risk." *Board of Governors of the Federal Reserve System*. https://www.federalreserve.gov/econres/notes/feds-notes/measuring-geopolitical-risk-20220415.html
[43] Wang, W., Sun, Z., Wang, W., Hua, Q., & Wu, F. (2023). "The impact of environmental uncertainty on ESG performance: Emotional vs. rational." *Journal of Cleaner Production*, 397, 136528. https://doi.org/10.1016/j.jclepro.2023.136528
[44] Zyznarska-Dworczak, B. (2022). "Financial and ESG reporting in times of uncertainty." *The Theoretical Journal of Accounting*, 46(4), 161–180. https://doi.org/10.5604/01.3001.0016.0321
[45] Zhang, C., Farooq, U., Jamali, D., & Alam, M. M. (2024). "The role of ESG performance in the nexus between economic policy uncertainty and corporate investment." *Research in International Business and Finance*, 70, 102358. https://doi.org/10.1016/j.ribaf.2024.102358

Authors' Profiles

Viktoriia Bulatova is an ambitious student at Lviv Polytechnic National University, pursuing a bachelor's degree in the Department of Information Systems and Networks. Passionate about Artificial Intelligence, particularly Machine Learning and Natural Language Processing, she actively develops practical projects combining backend, frontend, and data analysis.

Sofiia Popp is a talented junior student earning her bachelor's degree in the Department of Information Systems and Networks at Lviv Polytechnic National University. She is an emerging scholar with a keen interest in Data Science, especially in the fields of Data Analysis and Machine Learning.

Victoria Vysotska is a Professor at the Information Systems and Networks Department of Lviv Polytechnic National University. She defended her Doctoral degree in Technical Science, speciality in «structural, applied and mathematical linguistics», in 2023 on the topic "Analysis and synthesis of computational linguistic systems for processing Ukrainian textual content" Also, she received a PhD degree in Information Technologies from Lviv Polytechnic National University in 2014. She has currently published more than 600 publications. Her main research interests are focused on the identification of disinformation/fakes/propaganda, detection of sources of disinformation and inauthentic behaviour (bots) of coordinated groups based on artificial intelligence, NLP, computer linguistics, data science, system analysis, information technologies, machine learning, cyber security.

Yuriy Ushenko: World's Top 2% Research Scientist (2020, 2021, 2022, 2024). Nominee of The Photonics100 2025 list by ElectroOptics. Winner of 1000 Talents Plan Program, PhD, DSc, Prof., at Shaoxing University, Shaoxing, Zhejiang Province 312000, China. Professor, Head of Computer Science Department, Chernivtsi National University, Ukraine. His research interests include intelligent information systems design, data mining, pattern recognition and digital image processing, artificial neural networks, laser polarimetry and interferometry.

Zhengbing Hu: Prof., Deputy Director, International Center of Informatics and Computer Science, Faculty of Applied Mathematics, National Technical University of Ukraine "Kyiv Polytechnic Institute", Ukraine. Adjunct Professor, School of Computer Science, Hubei University of Technology, China. Visiting Prof., DSc Candidate in National Aviation University (Ukraine) from 2019. Primary research interests: Computer Science and Technology Applications, Artificial Intelligence, Network Security, Communications, Data Processing, Cloud Computing, Education Technology.

Dmytro Uhryn graduated from Yuriy Fedkovych Chernivtsi National University, Chernivtsi. Currently, he is a Doctor of Technical Sciences and associate professor at Yuriy Fedkovych Chernivtsi National University. His research interests are data mining, information technologies for decision support, swarm intelligence systems, and industry-specific geographic information systems.

How to cite this paper: Viktoriia Bulatova, Sofiia Popp, Victoria Vysotska, Yuriy Ushenko, Zhengbing Hu, Dmytro Uhryn, "Comprehensive Intellectual Analysis of Statistical Data on Leading Energy Companies' Actions", International Journal of Information Engineering and Electronic Business(IJIEEB), Vol.17, No.4, pp. 82-144, 2025. DOI:10.5815/ijieeb.2025.04.07