ANOMALY ANALYSIS ON FAANG+2

ANOMALY DETECTION ANALYSIS IN HISTORICAL STOCK PRIZE OF FAANG+2

(FACEBOOK, AMAZON, APPLE, NETFLIX, GOOGLE, WALMART AND MICROSOFT)

COMPANIES

Submitted by,

Shifa Nasar V (223048)

Sibinraj V M (223049)

Sreya K Raj (223050)

Steve Jos C M (223051)

School of Digital Sciences

Kerala University of Digital Sciences Innovation and Technology

(Digital University Kerala)

Technocity Phase 4 Campus, Thiruvananthapuram, Kerala - 695317

BONAFIDE CERTIFICATE

This is to certify that the project report entitled **ANOMALY ANALYSIS ON FAANG+2** submitted by:

Name	**Reg No:**
SHIFA NASAR V	223048
SIBINRAJ V M	223049
SREYA K RAJ	223050
STEVE JOS C M	223051

In partial fulfillment of the requirements for the award of Master of Science in Computer Science with Specialization in Data Analytics is a Bonafide record of the work carried out at KERALA UNIVERSITY OF DIGITAL SCIENCES, INNOVATION AND TECHNOLOGY under our supervision.

Supervisor

Prof. MANOJ KUMAR T K
School of Digital Sciences
DUK

Course Coordinator

prof. MANOJ KUMAR T K
School of Digital Sciences
DUK

Head of Institution
Prof. SAJI GOPINATH
Vice Chancellor
DUK

DECLARATION

We**, Shifa Nasar V, Sibinraj V M, Sreya K Raj,** and **Steve Jos C M** students of Master of Science in Computer Science with Specialization in Data Analytics, hereby declare that this report is substantially the result of our own work, and has been carried out during the period March 2023-July 2023.

ACKNOWLEDGEMENT

This project was supervised by Prof. Dr. T K Manoj Kumar, an associate professor at Digital University Kerala, who provided invaluable guidance and assistance in addressing our queries and locating relevant resources. We extend our heartfelt appreciation to him.

We also wish to convey our deep gratitude to Dr. Saji Gopinath, our Vice Chancellor, for creating a conducive and supportive environment that allowed us to work on the project effectively and in a friendly atmosphere.

CONTENTS

ABSTRACT

In our analysis, we employed anomaly detection techniques to assess the stock prices of seven different companies. This approach allowed us to perform a yearly basis comparison of the fluctuations in stock prices for each of these companies. By leveraging anomaly detection, we aimed to identify and highlight any abnormal patterns or significant deviations in the stock price trends, which could provide valuable insights into the performance of these company's stocks over time. This analysis serves as a tool for monitoring and understanding the annual variations in stock prices, aiding in investment decision-making and market analysis.

We utilized the stock prices of seven prominent companies: Amazon, Apple, Facebook, Google, Microsoft, Netflix, and Walmart. We examined their stock price data spanning from the year 2015 to 2021. Our primary objective was to conduct an annual comparison of these stock prices, seeking to discern the impact of various events or factors that influenced the performance of these companies' stocks throughout this time frame. This analysis aimed to identify and evaluate the yearly fluctuations in stock prices, shedding light on the events and circumstances that contributed to these fluctuations for each of the mentioned companies.

By examining the anomalies of different companies, we were able to find that Facebook, Amazon, Apple, Netflix, Google, Microsoft, and Walmart showcased impressive resilience in the face of market volatility, achieving historically high stock prices despite occasional setbacks tied to factors like the COVID-19 pandemic and interest rate concerns. While Walmart and Netflix capitalized on changing consumer behaviors during the pandemic, challenges like heightened competition and cost pressures, which impacted Netflix's subscriber growth, highlighted the complexities of this evolving market landscape, emphasizing the adaptability and long-term growth potential of these tech giants, particularly Amazon and Apple.

INTRODUCTION

BACKGROUND

Stock prices play a multifaceted role for companies, impacting their access to capital, strategic decisions, and market perception. A rising and healthy stock price is generally seen as a positive sign, while a declining one can raise concerns among stakeholders and investors. Many factors influence stock prices, including company performance metrics such as earnings and profit margins, competitive positioning, industry trends, and macroeconomic conditions. Additionally, stock prices can be subject to market speculation and sentiment, leading to short-term fluctuations that may not necessarily reflect a company's underlying sales performance. Therefore, investors and analysts typically consider a wide range of factors when evaluating a company's stock for investment purposes. Comparing stock prices of different companies is a common practice for assessing relative value and making informed investment decisions.

PROBLEM STATEMENT

In this context, the objective is to identify anomalies in the performance of each company on an annual basis. After detecting these anomalies, the following steps involve identifying the root causes behind them and comparing the fluctuations that primarily impact the stock prices of these companies. This process typically involves a comprehensive analysis of various factors, including financial metrics, market dynamics, and external events, to understand the drivers of stock price fluctuations and anomalies for each company over the years.

LITERATURE REVIEW

Over the past years many analysis works have been done in the field of stock market, and detecting anomaly in time series analysis is important for risk management, fraud detection and investment strategies and behavioural Analysis. Recent advances in technology have brought importance in data collection which would also improve the importance finding the outliers or anomaly detection. Anomalies are data points that significantly different from rest of the data point. Anomaly detection involves finding the data points that deviates from the rest of the data points the number and variety of anomaly detection algorithms has grown significantly in the past and, because many of these solutions have been developed independently and by different research communities, there is no comprehensive study that systematically evaluates and compares the different approaches. For this reason, choosing the best detection technique for a given anomaly detection task is a difficult challenge. According to the study different algorithm are there to detect anomalies. From report published by Anomaly detection in the time series gives a concise overview of the techniques and their commonalities we evaluate their individual strengths and weaknesses.

In Microsoft a time series anomaly detection service is build which would help the customers monitor the time series continuously and alert for potential incident on time, publication from time series anomaly detection Services at Microsoft introduce the pipeline and algorithm of anomaly detection service. In the today's era a huge amount of data is been collected from sensors, according to report done by Generic and scalable periodicity adaptation framework for time series anomaly detection anomaly can be classified into three categories point anomalies, contextual anomaly, and collective anomalies. The meaning of anomaly differs from domain to domain. Anomaly detection severs many purpose one of the effective mechanism that can help security analysts to identify potential threats and obtain meaningful information

DATA AND EVALUATION

DATASET

The dataset contains the stock price data of seven renowned companies, including Apple, Amazon, Netflix, Microsoft, Google, Facebook, and Walmart. It comprises a comprehensive set of attributes, including the date of each recorded data point, denoted as "Date," which allows for chronological tracking. Additionally, it provides essential financial metrics such as the opening price of the stock on a given day, referred to as "Open," the highest price achieved during that day ("High"), the lowest price recorded ("Low"), and the closing price ("Close"). Moreover, it furnishes data on the volume of stocks traded on each day, quantified as "Volume." Notably, the dataset accounts for any alterations or corporate actions that transpired before the following day's opening, adjusting the closing price accordingly, indicated as "Adjusted Close." The temporal scope of this dataset spans from 2015 to 2021, offering a day-level perspective for comprehensive stock market analysis. The experimental dataset is available at:

https://www.kaggle.com/datasets/suddharshan/historical-stock-price-of-10-popular-companies/code

EXPLORATORY DATA ANALYSIS DATA PRE-PROCESSING

In our exploratory data analysis (EDA) process, we assessed seven distinct datasets, each corresponding to a different company's stock price. Our primary objectives during this EDA were to identify and address any potential issues related to missing data (null values) and outliers within these datasets.

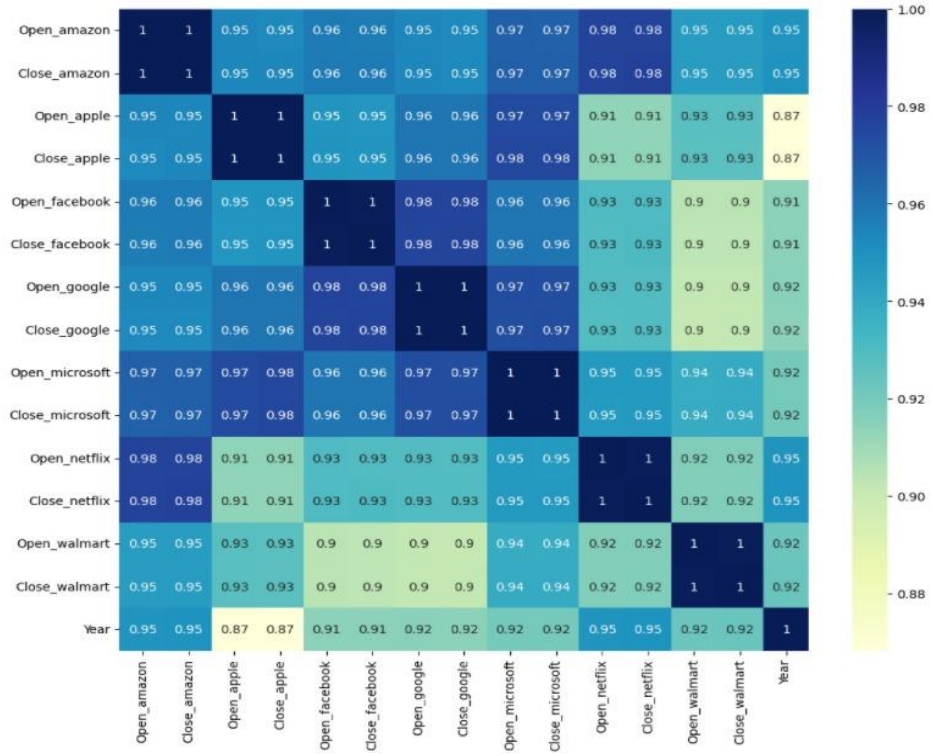
Upon examining the data, we determined that none of the seven datasets contained any missing or null values. Therefore, we did not need to perform further investigations regarding missing data. However, our analysis did reveal the presence of outliers in all seven datasets. Outliers are data points that deviate significantly from the rest of the data and can substantially impact the overall dataset analysis. To address this issue, we employed box plots to visualize the

distribution of data and define the range of values between the upper and lower quartiles (interquartile range). Any data points falling outside this range were considered outliers.

Our next step involved removing these outliers from the datasets, as they can skew statistical analyses and distort the understanding of the underlying patterns in the stock price data. This process aimed to enhance the robustness of our data analysis by ensuring that extreme values did not unduly influence our findings.

After cleaning our dataset, we constructed a streamlined data frame focusing specifically on seven prominent companies' opening and closing stock prices. This refined dataset forms the cornerstone of our forthcoming analysis, aimed at discerning notable trends and fluctuations in stock prices. Our attention is centered on the key attributes of these selected companies, allowing us to draw insightful comparisons and pinpoint significant upswings or downturns in stock values. Through this curated dataset, we endeavor to unveil valuable insights into the dynamic world of stock market movements, shedding light on the substantial shifts in stock prices that can impact investment decisions.

The correlation analysis of this attribute data frame reveals a robust and positive correlation among all the attributes, indicating a strong interconnectedness between them.



METHODOLOGIES

ANOMALY DETECTION

Time Series Analysis involves examining data collected or recorded over time, where each data point is associated with a specific timestamp. It is a fundamental technique in statistics and data analysis that aims to uncover patterns, trends, and insights within time-ordered data, allowing us to make informed predictions or decisions based on historical observations. Anomaly Detection within the context of time series analysis is the process of identifying rare, unexpected, or abnormal patterns within a time series dataset. These anomalies can manifest as data points or patterns that deviate significantly from the expected or typical behavior observed in the historical data. Anomalies can take various forms, including sudden spikes, drops, or irregularities, and they are crucial to detect as they often signify important events or deviations that warrant further investigation. The working of anomaly detection in time series analysis involves employing various statistical, machine learning, or domain-specific techniques to automatically identify and flag anomalies within a time series dataset. This process typically begins with data preprocessing, where the time series data is cleaned, transformed, and prepared for analysis. Next, a suitable anomaly detection method is chosen based on the specific characteristics of the data. Common approaches include statistical methods like z-scores, machine learning algorithms like Isolation Forest or autoencoders, or domain-specific techniques designed for time series data. Once the anomaly detection model is trained and applied to the dataset, it highlights data points or time periods that exhibit significant deviations from the expected behavior, helping analysts or decision-makers identify and respond to exceptional events or patterns in the data.

Here we have used three algorithms for anomaly detection

One-Class SVM (Support Vector Machine):

The One-Class SVM, or Support Vector Machine, is a machine learning algorithm used for anomaly detection. It's particularly helpful when you have a dataset with predominantly one class of data and you want to detect anomalies or outliers in that data.

- The algorithm aims to find the best hyperplane (a separator) that separates the majority of the data points (the inliers) from the rest of the data, which are considered anomalies (the outliers).

- The hyperplane is positioned to maximize the margin around the inliers, effectively isolating them from the anomalies.

- Any data points that fall outside this margin are identified as anomalies.

One-Class SVM is used in various applications, such as fraud detection in credit card transactions, fault detection in manufacturing processes, and identifying outliers in sensor data.

Local Outlier Factor (LOF):

The Local Outlier Factor (LOF) is an unsupervised anomaly detection algorithm that focuses on the local behavior of data points. Unlike global approaches like the One-Class SVM, LOF considers the density of data points in their local neighborhoods to identify anomalies.

- For each data point, LOF calculates a score based on the density of its neighbors. A point with a significantly lower density than its neighbors is considered an anomaly.

- LOF takes into account the idea that anomalies often have fewer nearby neighbors than normal data points. This makes it effective in scenarios where anomalies form clusters or exist in regions of varying data density.

LOF is used in applications such as network intrusion detection, identifying outliers in sensor networks, and detecting anomalies in spatial data.

Isolation Forest:

The Isolation Forest is another machine learning-based anomaly detection algorithm that's designed to efficiently find anomalies, especially in high-dimensional datasets.

- It uses a random forest of decision trees to isolate anomalies. Each tree randomly selects a feature and a random value to split the data.

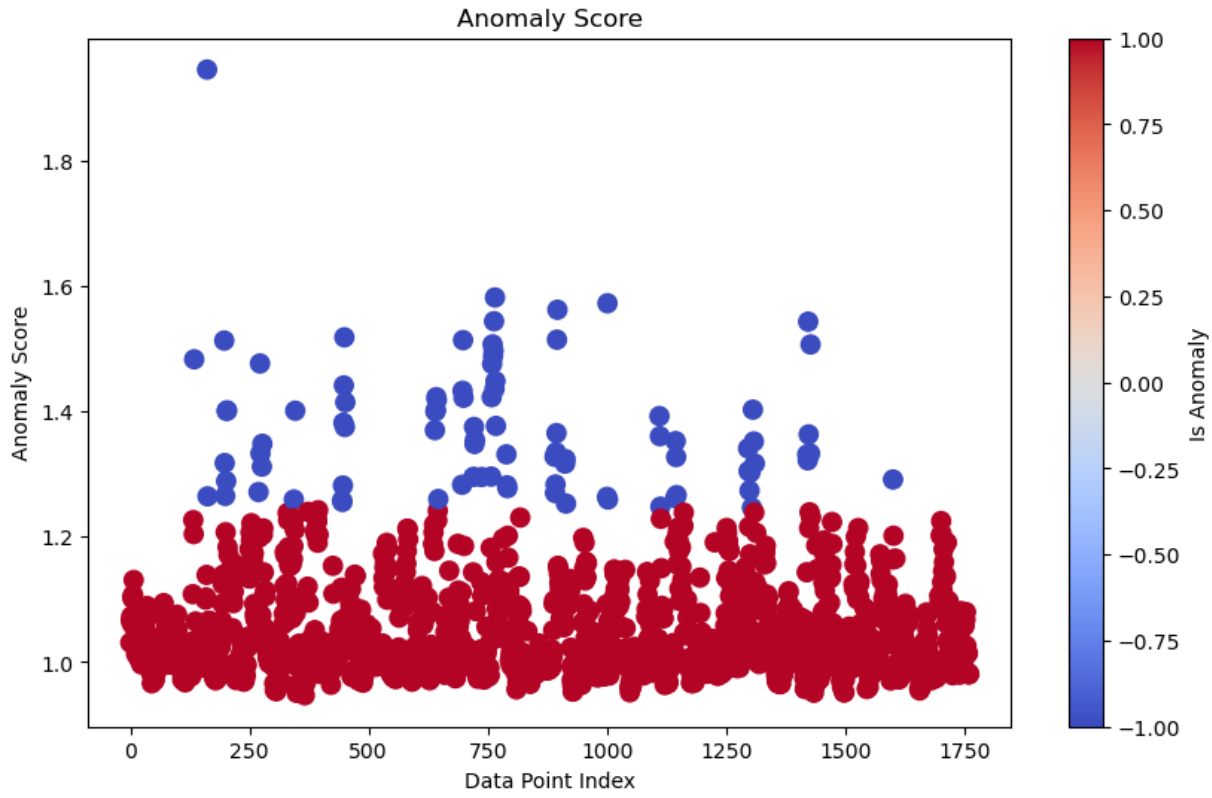
- Anomalies are isolated more quickly because they require fewer splits to separate them from the majority of the data.

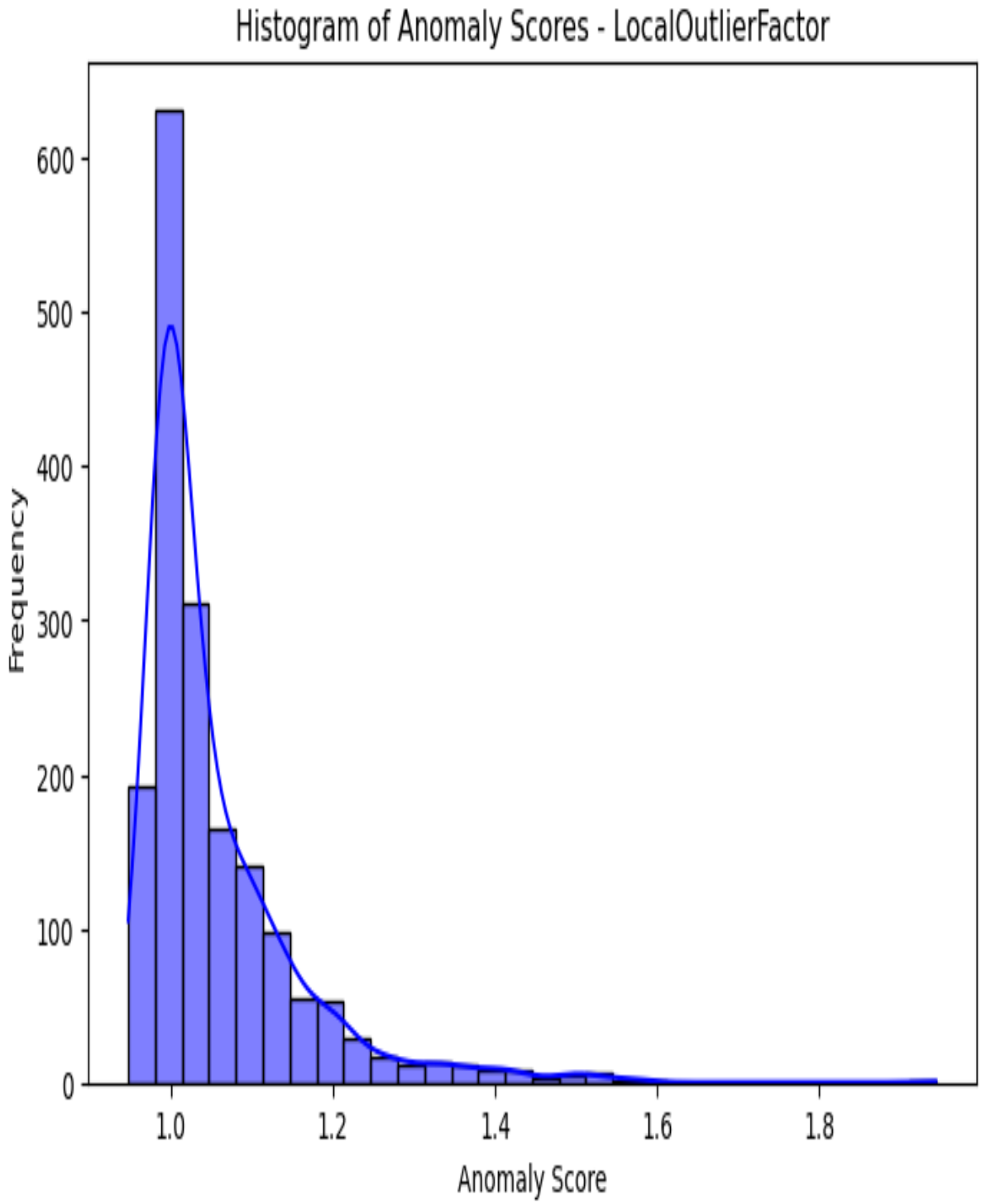
- By averaging the results from multiple trees, the algorithm produces a final anomaly score for each data point, and those with higher scores are considered anomalies.

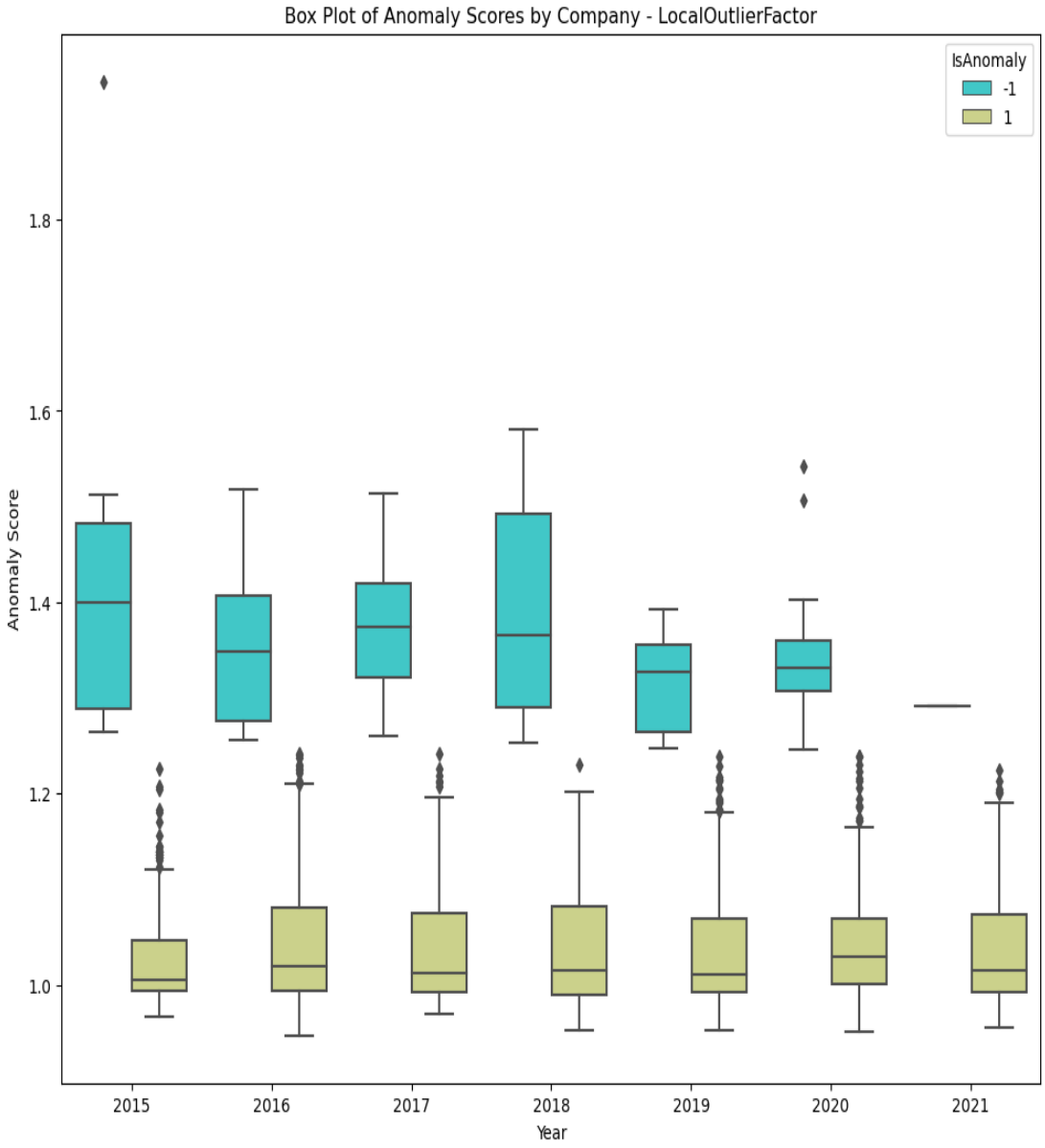
Isolation Forest is efficient, scalable, and capable of handling datasets with many features. It's used in various applications, including identifying anomalies in cybersecurity (e.g., network intrusion detection) and quality control in manufacturing processes.

EXPERIMENTAL ANALYSIS

We utilized the mentioned algorithms to detect anomalies in our dataset. After conducting an in-depth analysis, we concluded that the Local Outlier Factor (LOF) algorithm outperformed the others. This decision was based on a thorough comparison, including visualization of the results. LOF identified a higher number of anomalies when compared to the alternative algorithms we explored.




Histogram of Anomaly Scores - LocalOutlierFactor
Frequency
0
100
200
300
400
500
600
1.0
1.2
1.4
1.6
1.8
Anomaly Score



In accordance with our chosen algorithm, we have generated both an anomaly score plot and a box plot representing the anomaly scores. In these plots, a score of -1 signifies data points classified as anomalies. Utilizing the Local Outlier Factor algorithm, we have identified and printed a total of 88 anomalies within our dataset.

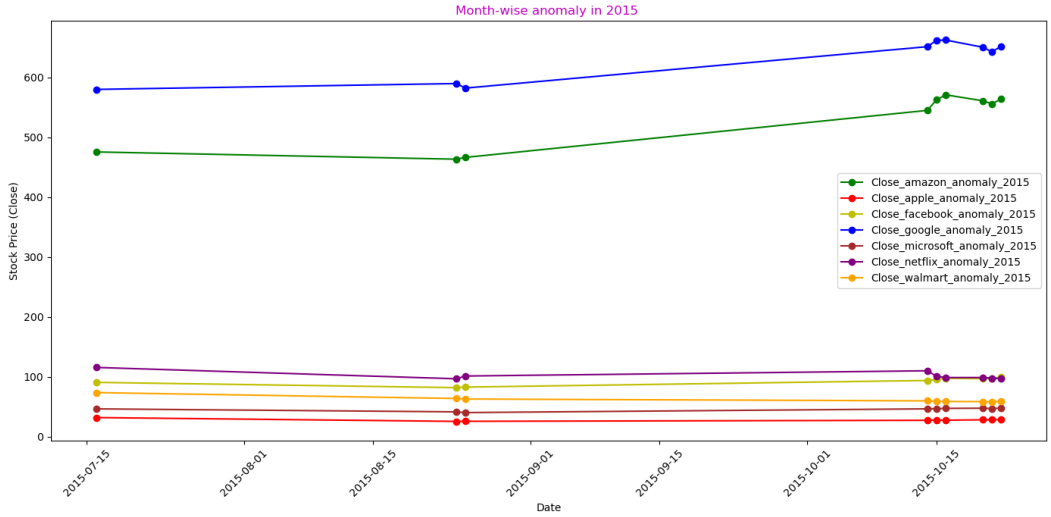
We have included these anomalies as an Excel with their particular reason below:

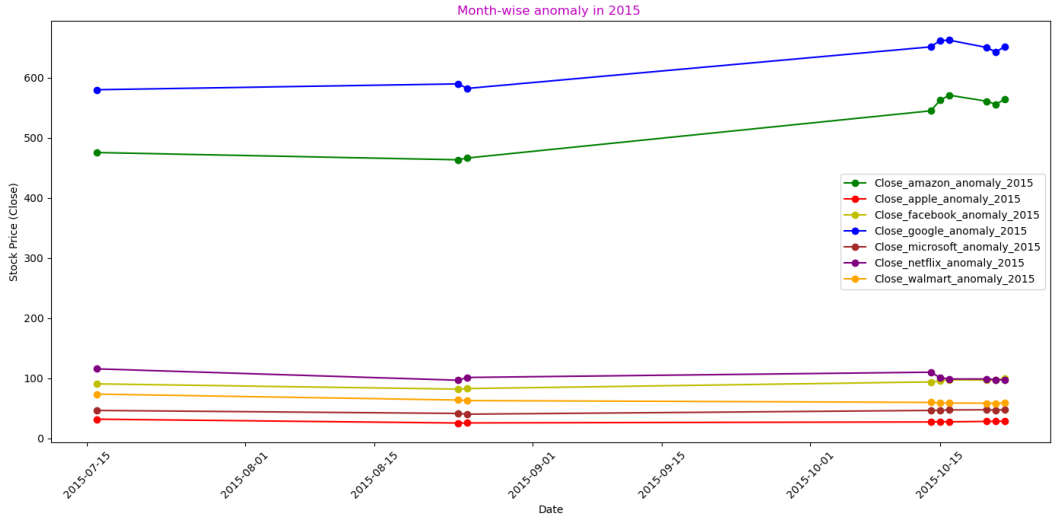
https://docs.google.com/spreadsheets/d/1l1PB_fyk1_sb6fxz1u9Jxjxrqqww2qHQJ2qql4WX7ZI/edit?usp=sharing

YEAR WISE COMPARISON OF ALL COMPANIES TOGETHER

We're taking a closer look at some big-name companies like Amazon, Apple, Facebook, Netflix, Google, Walmart, and Microsoft. We want to see how their stock prices have been changing over the years and figure out why. This way, we can find out which company has had the most ups and downs in its stock price and understand why it happened.

2015





The given diagram shows the month-wise stock price anomalies of Amazon, Apple, Facebook, Google, Microsoft, Netflix, and Walmart in 2015. The anomalies are represented by the vertical bars, and the stock prices are represented by the line graph.

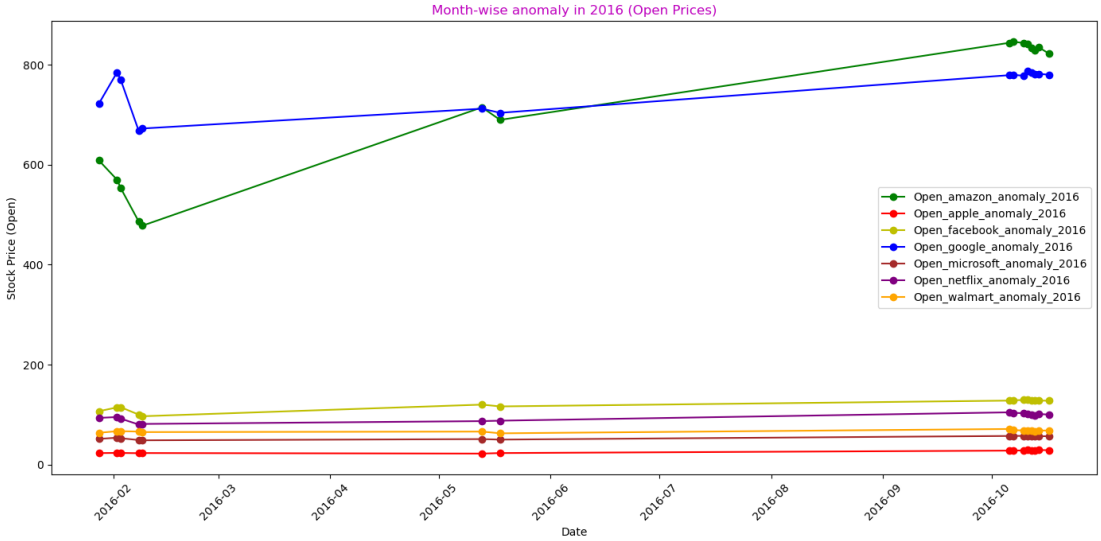
There were three major anomalies in the stock prices of Amazon, Apple, Facebook, Google, Microsoft, Netflix, and Walmart in 2015. These anomalies occurred in July, August, and September. The anomaly in July was the largest, and it affected all seven stocks. The stock prices of all seven stocks fell sharply during this period.

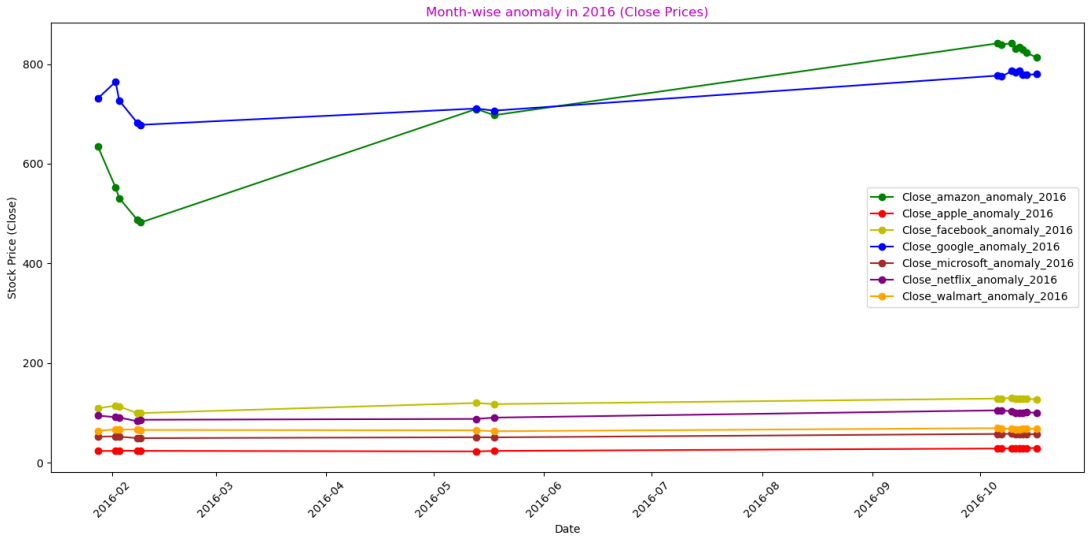
The anomaly in September was the smallest of the three anomalies and only affected a few stocks. The stock prices of Amazon, Apple, and Facebook fell during this period, but the stock prices of Google, Microsoft, Netflix, and Walmart did not.

Google is in the top position in the graph, followed by Amazon. This is because Google has strong fundamentals and a positive outlook. It is also a leader in its industry and has strong brand recognition. Amazon is also a strong company with good fundamentals and a positive outlook. However, it is not as dominant in its industry as Google is.

The anomaly in September could have been caused by a number of factors, such as the Federal Reserve's decision not to raise interest rates, the strong dollar, or the oil price decline. However, it is also possible that the anomaly was caused by specific company news or events. It is important to do your own research to understand why a stock price fluctuates.

2016





The given diagram shows the month-wise open prices of Amazon, Apple, Facebook, Google, Microsoft, Netflix, and Walmart in 2016. The open prices of all seven stocks increased in 2016

The largest increase in stock price was for Amazon, followed by Apple and Facebook. Amazon outperformed Google in the month of May.

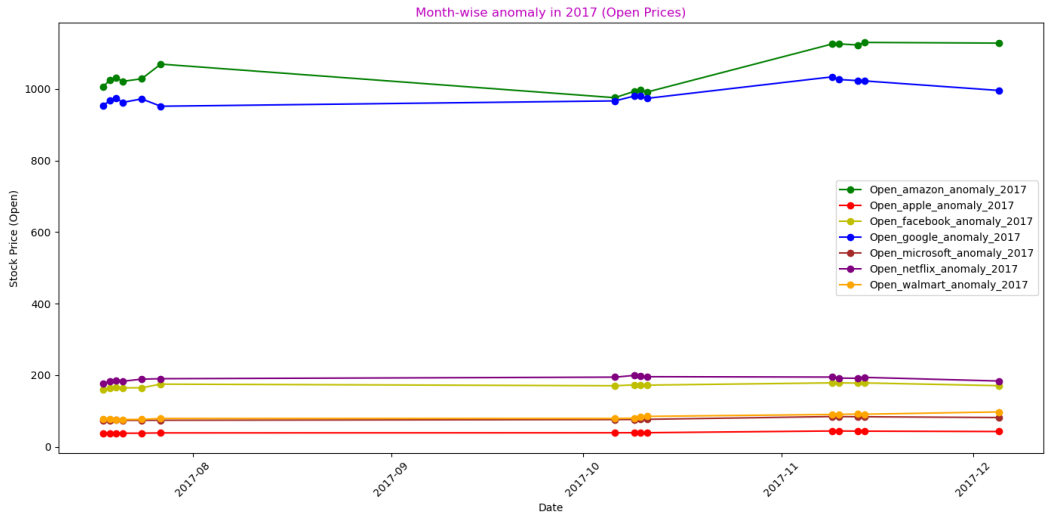
The anomaly in February was the largest, and it affected all seven stocks. The open prices of all seven stocks fell sharply during this period.

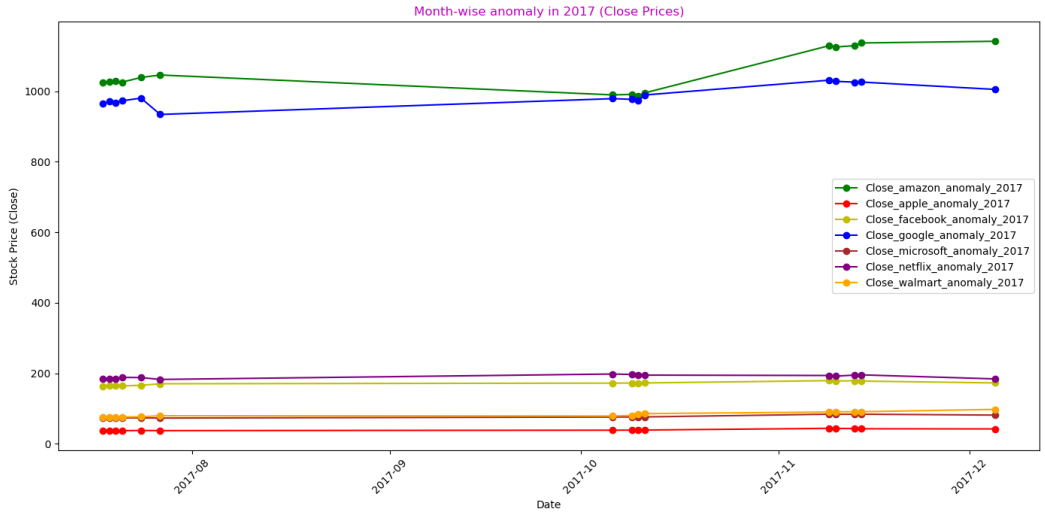
The anomaly in February 2016 coincided with the release of weak economic data from China. This data led to concerns about a slowdown in the Chinese economy, which had a negative impact on the global stock market.

The anomaly in July 2016 was caused by a sell-off in the Chinese stock market. This sell-off was triggered by concerns about the health of the Chinese economy and the government's ability to manage it.

The anomaly in September 2016 was caused by a number of factors, including the terrorist attacks in Nice, France, and the upcoming US presidential election. These events led to increased uncertainty in the global markets, which negatively impacted stock prices.

2017





The given diagram shows the month-wise open prices of Amazon, Apple, Facebook, Google, Microsoft, Netflix, and Walmart in 2017.

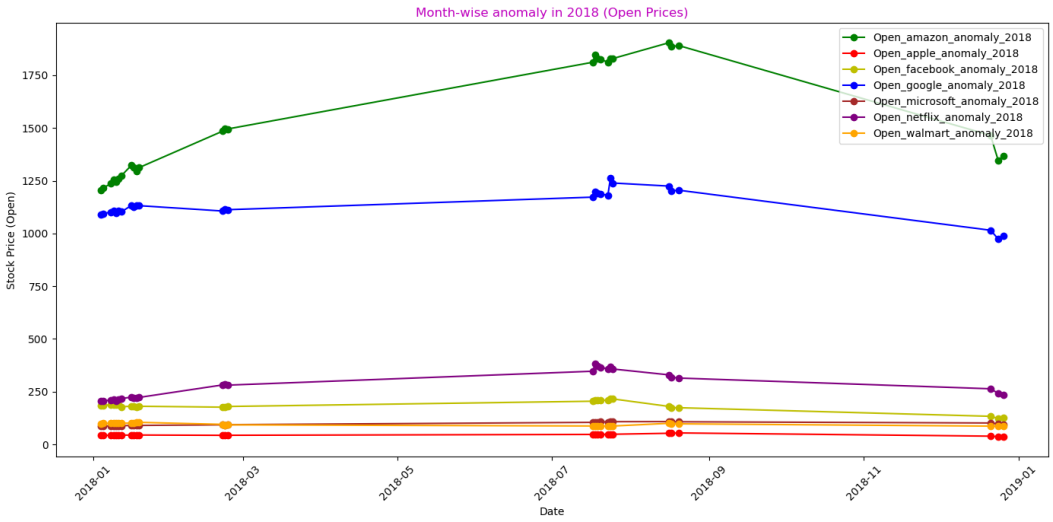
The anomalies in August, September, and December 2017 were caused by weak economic data from the United States, a sell-off in the Chinese stock market, and global events. Amazon was the top performer in 2017, followed by Netflix and Google. Google and Amazon collided in October, but Amazon recovered more quickly. Netflix was higher than Facebook in December.

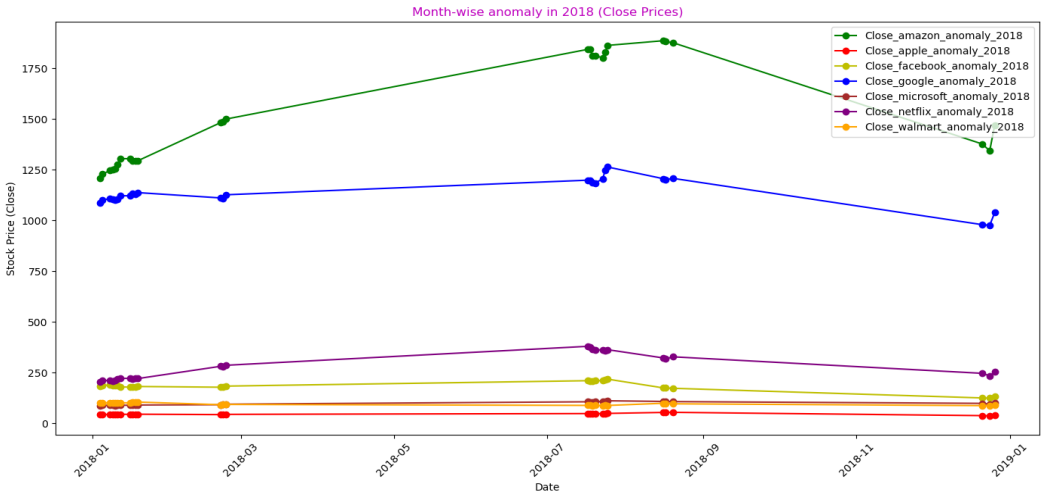
Amazon was the top performer in 2017, followed by Google and Netflix. This suggests that investors were optimistic about the future prospects of these companies.

Google and Amazon collided in October because both stocks were affected by the weak economic data from the United States. However, Amazon recovered more quickly than Google.

Netflix was higher than Facebook in December because Netflix was benefiting from the growth of the streaming media industry. Facebook was also affected by the uncertainty in the global markets.

2018





The given diagram shows the month-wise open prices of Amazon, Apple, Facebook, Google, Microsoft, Netflix, and Walmart in 2018.

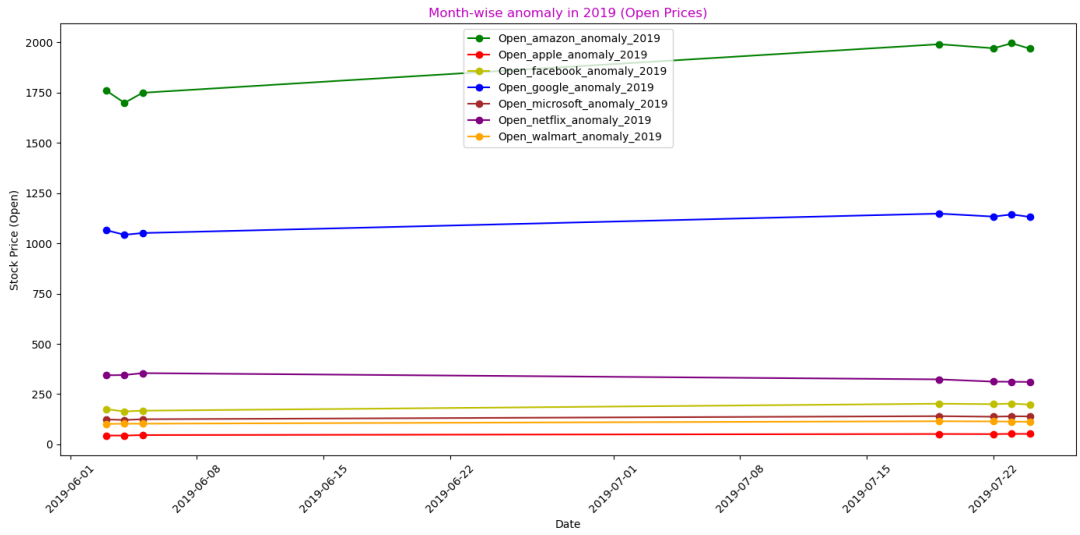
The anomalies in January, July, and September 2018 were caused by a number of factors, including weak economic data from the United States, a sell-off in the Chinese stock market, and the trade war between the United States and China. These factors led to concerns about a

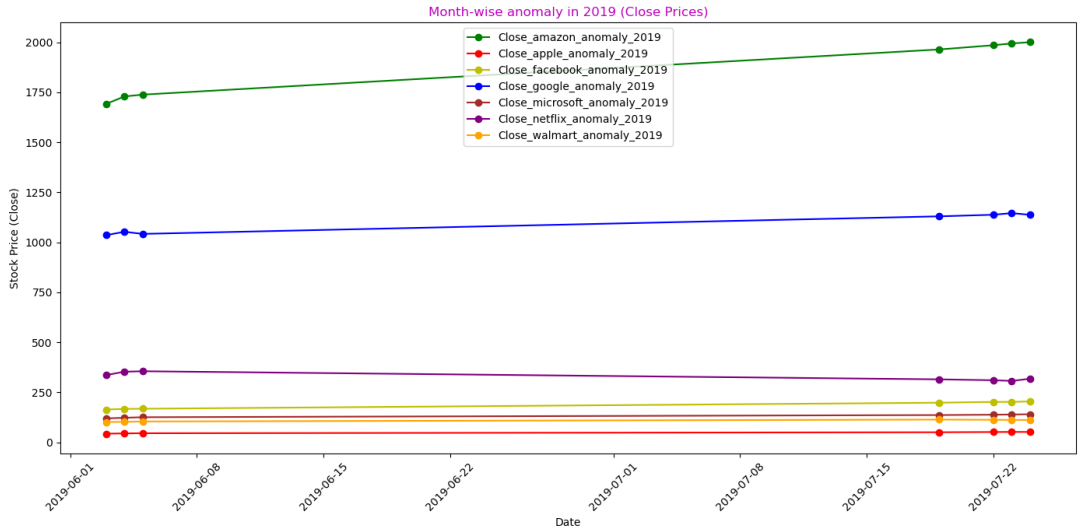
slowdown in the US economy, the health of the Chinese economy, and increased uncertainty in the global markets. As a result, stock prices fluctuated significantly during these periods.

Amazon was the top performer in 2018, followed by Netflix and Facebook. Netflix was higher than Facebook in September. Walmart and Microsoft were also performing well. These results suggest that investors were optimistic about these companies' future prospects despite the stock market volatility.

Overall, the stock market in 2018 was volatile. Several factors caused the stock market to fluctuate, including economic data from the United States, the Chinese stock market, and the trade war between the United States and China. Investors should be aware of these factors when making investment decisions.

2019





The given diagram shows the month-wise open prices of Amazon, Apple, Facebook, Google, Microsoft, Netflix, and Walmart in 2019.

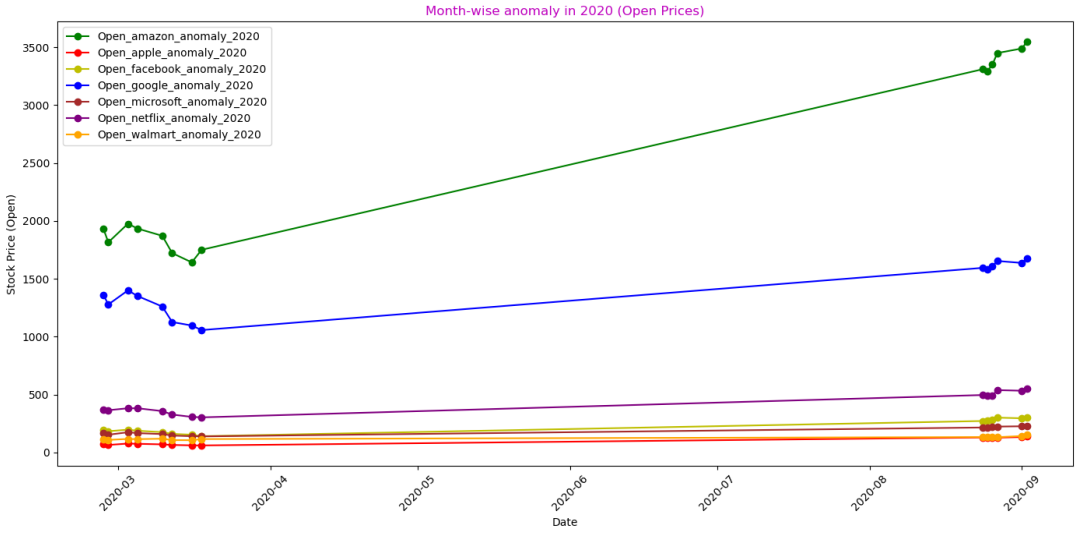
The stock market in 2019 was volatile. There were a number of factors that caused the stock market to fluctuate, including:

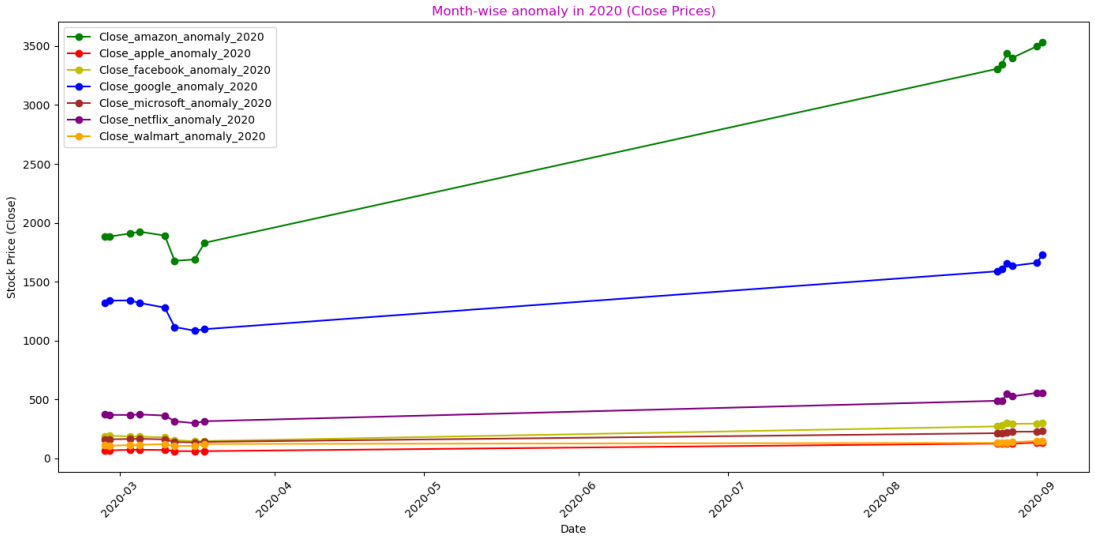
Weak economic data from the United States: The US GDP growth slowed to its slowest pace in three years in the second quarter of 2019. This raised concerns about the health of the US economy and its impact on the global economy. A sell-off in the Chinese stock market: The Chinese stock market lost over 40% of its value in just a few weeks in July 2019. This sell-off was triggered by concerns about the health of the Chinese economy and the government's ability to manage it. The trade war between the United States and China: The trade war between the United States and China escalated in 2019, with both countries imposing tariffs on each other's goods. This led to increased uncertainty in the global markets. The ongoing COVID-19 pandemic: The COVID-19 pandemic began in December 2019 and spread rapidly around the world in 2020. The pandemic caused widespread economic disruption and uncertainty.

Despite these challenges, Amazon was the top performer in 2019, followed by Google and Netflix. Amazon's strong performance was driven by its continued e-commerce and cloud computing growth. Google's strong performance was driven by its continued growth in advertising revenue. Netflix's strong performance was driven by its continued growth in streaming subscribers.

Overall, the stock market in 2019 was volatile. However, investors who invested in companies with strong fundamentals and positive outlooks, such as Amazon, Google, and Netflix, were rewarded with strong returns.

2020





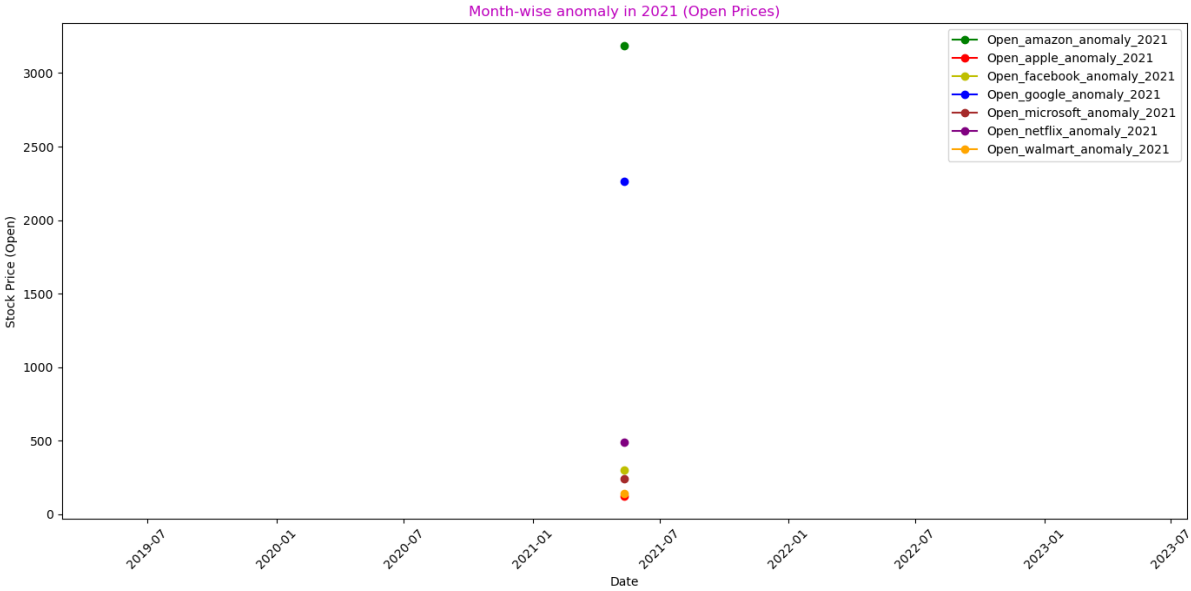
The diagram shows the month-wise open prices of Amazon, Apple, Facebook, Google, Microsoft, Netflix, and Walmart in 2020. The COVID-19 pandemic caused a sharp decline in the stock market in March 2020, with all seven stocks experiencing significant anomalies. Amazon's open price fell the most, followed by Apple, Facebook, Google, Microsoft, Netflix, and Walmart.

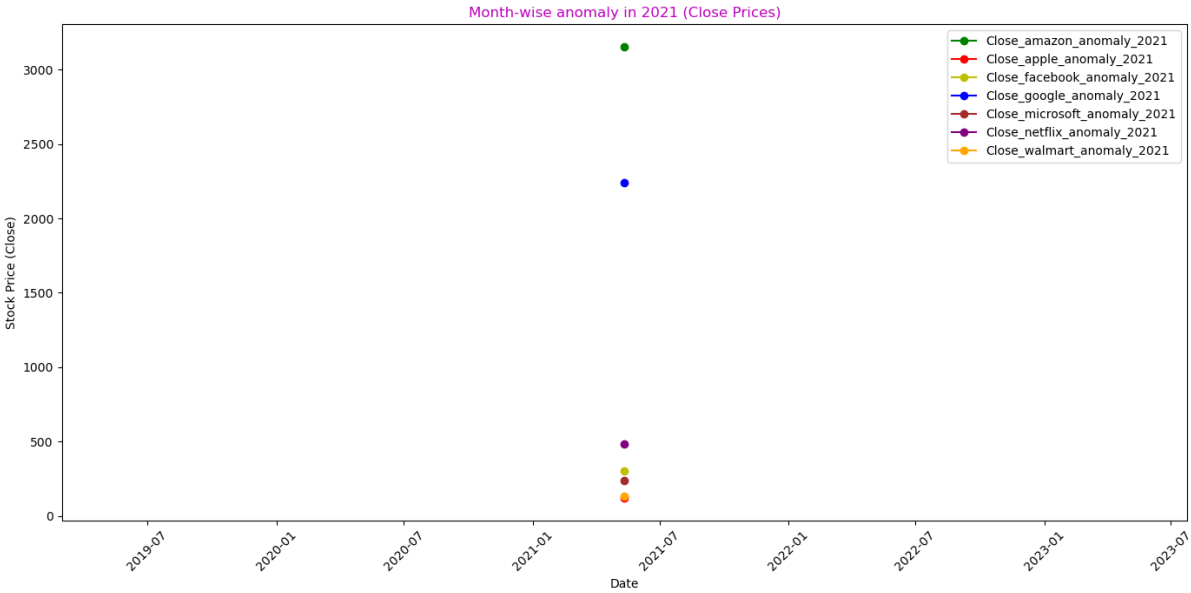
The stock market began to recover in April 2020, and all seven stocks showed positive returns for the month. Amazon continued to outperform the other stocks, with its open price increasing the most. Google and Netflix also showed strong returns. Apple, Facebook, and Microsoft had more modest returns. Walmart was the only stock that did not experience a positive return in April 2020.

The stock market continued to recover in the remaining months of 2020, and all seven stocks showed positive returns for the year. Amazon was the top performer, followed by Google and Netflix. These three companies were well-positioned to benefit from the ongoing digital transformation of the economy.

Overall, the stock market in 2020 was volatile, but it ended the year on a positive note. Amazon was the top performer, followed by Google and Netflix. These three companies were well-positioned to benefit from the ongoing digital transformation of the economy.

2021





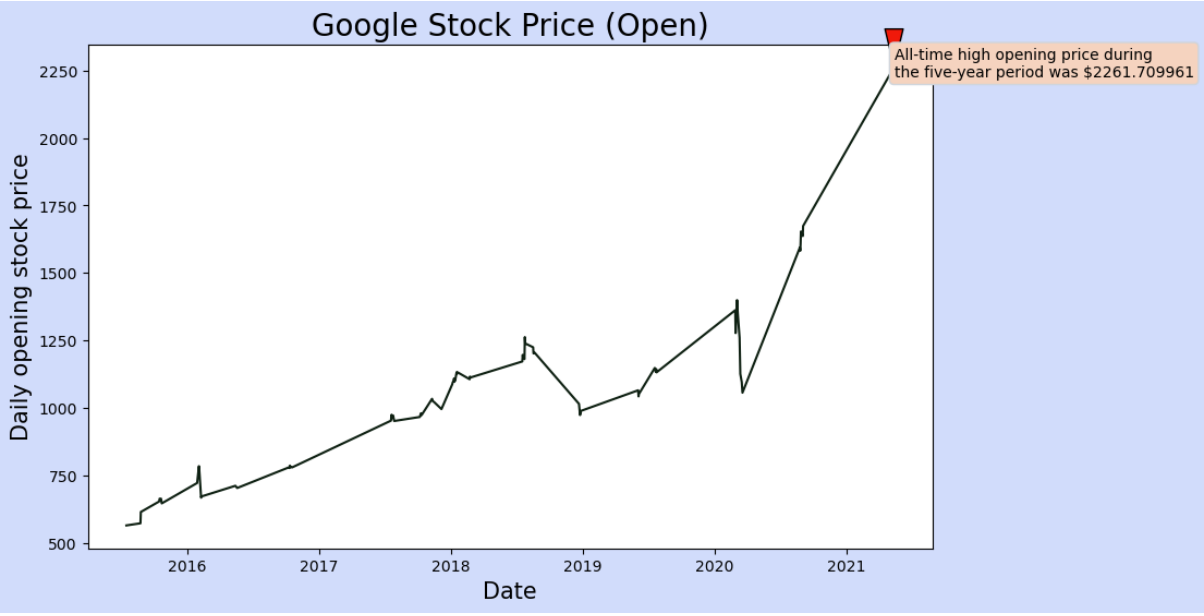
The stock market in 2021 was volatile but positive. The S&P 500 index rose by about 28%, the best performance since 1997. Key factors were economic recovery, earnings growth, low-interest rates, and optimism. Amazon, Google, and Netflix were the top companies. Walmart and Apple are almost similar in their range.

RESULT

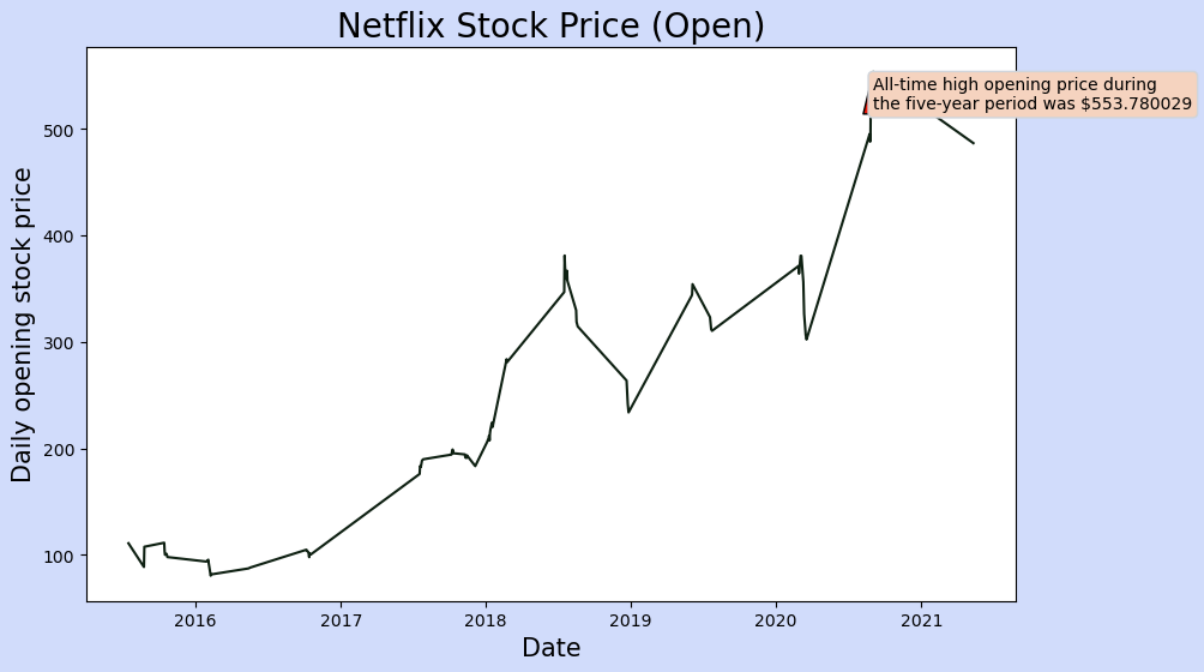
Between 2020 and 2022, FAANG+2 companies (Facebook, Amazon, Apple, Netflix, Google, Microsoft, and Walmart) demonstrated remarkable resilience amidst market turbulence, with several of them reaching historic high stock prices. While their stock prices generally followed an upward trajectory, punctuated by occasional sharp declines due to external factors like the COVID-19 pandemic and concerns about rising interest rates, these tech giants stood out. High-performing companies like Amazon and Google achieved remarkable milestones with all-time high opening and closing prices, indicating robust financial health and continued growth. Amazon reached an all-time high opening price of $3,547 and a closing price of $3,531. The all-time high closing price during the five-year period of Google was $2239.080078 on January 26, 2022, and mostly in the third place Netflix shows the all-time high closing price during the five-year period was $556.549988. These high stock prices were emblematic of their robust financial performances and continued growth, and they are worthy of comparison with the other companies in the group. While each company faced its unique challenges, such as the COVID-19 pandemic's initial impact and concerns over rising interest rates, these tech giants generally demonstrated an upward trajectory in stock prices.

Other firms, such as Walmart and Netflix, capitalized on changing consumer behaviors during the pandemic, driving their stock prices upward. However, challenges, including heightened competition and cost pressures that manifested in Netflix's slowing subscriber growth, underscored the complex dynamics within this evolving landscape. These insights underscore the intricate dynamics at play within the market during this period and highlight the adaptability and long-term growth potential of these companies, with Amazon and Apple serving as noteworthy examples of their financial strength and market dominance.

Amazon stock price from the year 2015 to 2021



Google stock price from the year 2015 to 2021



Netflix stock price from the year 2015 to 2021

CONCLUSION

When comparing the stock prices of high-profile companies such as Amazon and Google, we observed significant milestones with all-time high opening and closing prices. This analysis led us to conclude that Amazon, Google, and Netflix consistently experience growth in their stock prices, which will likely result in increased income in the future. Comparing the stock prices of these companies allows us to discern anomalies that arise when certain events have a profound positive or negative impact on a company. For instance, some companies suffered during the pandemic, while others achieved high income and stock prices under similar circumstances.

Predicting future fluctuations in a company's stock price is a complex task because it depends on numerous factors such as earnings, economic conditions, industry trends, market sentiments, and more. Some companies may experience unexpected surges in their stock prices due to specific events, while others might see unexpected declines. These fluctuations are influenced by a multitude of variables, making it challenging to make precise predictions about when a company's stock price will rise or fall.

Some annual corporate events have the potential to positively impact a company's stock price, with the effect often extending into the following year. For instance, events like Amazon's Prime Day or the release of a new iPhone by Apple can boost stock prices. However, unforeseen and unpredictable events, such as a pandemic, can have a significant but uncertain impact on a company's stock value.

REFERENCES

1. Time series contextual anomaly detection for detecting market manipulation in stock market manipulation in stock market, Koosha Golmohammadi; Osmar R.Zaiane , (https://ieeexplore.ieee.org/abstract/document/7344856)

2. Anomaly detection in time series, Proceeding of the VLDB Endowment , Volume 15, Issue 9 (https://dl.acm.org/doi/abs/10.14778/3538598.3538602)
3. Time- Series Anomaly detection service at Microsoft, KDD'19: Proceeding of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, July 2019 (https://dl.acm.org/doi/abs/10.1145/3292500.3330680)
4. Anomaly detection in Time series, https://neptune.ai/blog/anomaly-detection-in-time-series