**PROJECT PROPOSAL**

Data Explorers

Group 19

Advanced Data Analytics, University of North Texas

ADTA 5130: Data Analytics I

Dr. Mehmet Orhan

March 28, 2024

**PROJECT TITLE – STRATEGIC ASSET DIVERSIFICATION AND OPTIMIZATION(SADO)**

**SADO** is a Data-Driven Approach to Minimize Risk and Maximize Returns." This title reflects both the analytical depth and strategic focus of your work, suggesting a comprehensive, data-informed methodology for investment decision-making.

**TEAM MEMBERS:**

* Hasritha Polamarasetti – 11706930 (Team Leader)
* Pavan Sai Teja Kokkura – 11704857
* Nitesh Runku – 11736022
* Jesse Jackson Yajjala – 11765561

**ABSTRACT:**

The financial markets are rapidly evolving, so making well-informed investment selections is critical. This project, initiated by an innovative finance firm, harnesses the power of data analytics to navigate through the maze of the most complex investment potential industries. Our mandate will be to select stocks that promise a high return while maintaining tolerable risk levels. We propose using the Capital Asset Pricing Model (CAPM) to create evidence-based recommendations that will assist the organization in meeting its profit objectives while avoiding or limiting unnecessary risk.

**TABLE OF CONTENTS:**

* Purpose and Motivation
* Data Acquisition and Selection
* Business Understanding / Analytics Questions
* Methodology
  + Data Understanding and Data Preparation
  + Exploratory Data Analysis (EDA)
  + ANOVA
  + CAPM Analysis
* Key Findings
* Investment Recommendations
* Conclusion
* Bibliography

**PURPOSE AND MOTIVATION:**

Basically, this activity's main objective is to find investment opportunities that support the company's goals by using statistical and analytical techniques that concentrate on facts. This drive is a result of the evolving financial markets, where traditional investment approaches frequently fall short of expected returns. Additionally, the CAPM model offers a mechanism to connect a security's predicted return with the market, as well as a potent way to assess a security's risk. The goal of this research project is to apply this model to direct our investment strategy and guarantee that choices are grounded on reliable empirical data.

**DATA ACQUISITION AND SELECTION:**

**Companies Selected:**

For this project, we have selected stocks from seven companies: Tesla (TSLA), Amazon (AMZN), Bitcoin (BTC-USD), Snap Inc. (SNAP), Google (GOOG), Netflix (NFLX), and Meta Platforms, Inc. (META). These companies have been chosen based on their market relevance, trading history of at least five years, and their potential to contribute to a diversified investment portfolio.

**Rationale Behind Company Selection:**

* Tesla (TSLA) and Amazon (AMZN) are two of the largest brands in their respective industries: one can only imagine the company's growth graph and the potential rewards.
* Bitcoin (BTC-USD), the major cryptocurrency, provides a unique potential for diversification. Despite the volatility, it nevertheless provides an opportunity to participate in a new asset class with different risk-reward characteristics than traditional equities.
* Snap Inc. (SNAP) The company is a major participant in the business, which is typical of the high-risk, high-reward side of the technology sector. The high level of relevance to a younger population may indicate significant long-term growth opportunities.
* Google (GOOG) and Netflix (NFLX) are technology heavyweights and streaming service pioneers, respectively. In a qualitative sense, if we argue that Google leads the world of digital advertising and technology, its main competitor is Netflix, which rules the streaming sector.
* Meta Platforms, Inc. (META) (previously Facebook): Meta is distinguished by advanced social media technology, such as virtual reality, a large user base, and an innovative desire to balance stability and growth.

**BUSINESS UNDERSTANDING / ANALYTICS QUESTIONS:**

This project's main objective is to develop an investment strategy that is resilient enough to endure the financial markets' current volatility. The current exercise builds a portfolio that will yield the highest returns while taking reasonable risks using data analytics in the form of Anova and the Capital Asset Pricing Model (CAPM). As a result, the Project acknowledges the significance of flexible investing techniques in the extremely complex and volatile financial markets.

**Analytical Questions Covered in this Project:**

* **Asset Correlation and diversification:** “How do various assets correlate, and how does this correlation affect our diversification strategy?”
* **Risk Assessment:** “What risks are associated with each asset, and how can we mitigate them in our portfolio?”
* **Trend Analysis:** “Are there any emerging market trends that could influence our investment decisions?”

Understanding the correlation of assets can help us create a portfolio that reduces risks and maximizes rewards. To be able to apply effective risk management techniques that guarantee stable and profitable investments, one must be in a position to know and be conscious of the risks. Early pattern recognition avails us opportunities to make our minds up and also steer clear of the risks that may not have arisen at that time, thus optimizing our investment plans.Effective development and execution of our data-driven investment strategy will depend on answering these analytical questions.

**METHODOLOGY:**

**Data Understanding / Data Preparation:**

* **Data Collection:**

Each of the seven data sets used was sourced from Yahoo Finance's historical records. This kind of data makes making informed decisions and doing comprehensive analysis much easier. Since the market will fluctuate over time, one can make judgments by studying these trends. The data for the present dataset was collected between the year 2019 and the present. It includes the historical data of the seven selected companies.

* **Initial Data Examination:**

Upon initial examination, the dataset includes the following columns:

* **Date**: The day of trade when the information is obtained.
* **Open**: The opening price of the stock for the day's trading.
* **High**: The stock price at which it peaked throughout the trading day.
* **Low**: The stock price was at its lowest point throughout the trading day.
* **Close**: The price at which the stock closes on a trading day.
* **Adj Close**: Adjusted Close price.
* **Volume**: The total quantity of shares that are traded during a trading day.
* **Ticker**: Stock symbol.
* **Data Preparation:**

In this stage, throughout this process of data preparation, a number of critical steps were taken in order to make it clear, well-organized, and fit for high-quality analysis.

* **Date Conversion and Indexing**: The text 'Date' column was converted using Python's datetime format to allow for precise time series analysis. The data frame's index was created using the generated datetime column, which made it simpler to slice and dice the data by era.
* **Missing Data Handling**: We have scrutinized column by column for any missing values throughout the process of data processing. Fortunately, we have identified one missing record for most financial indicators. This implies the data is of high quality and hence probably little need for imputation or elimination. For quality work, we removed rows with missing values. We believe that the small amount of missing data will not impact our conclusions.
* **Data Integrity Checks**: It was checked that the numerical columns (Open, High, Low, Close, Adj Close, Volume) are applied with the respective numeric operations having the correct float codes. Also, the column 'Ticker' is checked to ensure consistency in identifying the stock and appropriate formatting.
* **Descriptive Statistics**: Financial measures heavily use descriptive statistics since they help to give much insight into the distribution, variability, and most importantly, the underlying trends of stock prices and transaction volumes. This phase is very critical since it helps to point to the typical results, abnormalities, and, most importantly, the outliers that can compromise the precision of further analysis.

We have reviewed the dataset carefully and organized it, resulting in a sound footing for the exploratory data analysis and modeling phases. It was guaranteed that the data used was well-structured, comprehensive, and with accuracy to better the precision of the study. This enhanced the validity and reliability of the findings while reducing the probability of errors and biases.

**Exploratory Data Analysis (EDA):**

Exploratory Data Analysis Cornerstone of every data analytics research lies exploratory data analysis, in search of underlying trends, patterns, and correlations existing in the data. Below are explanations for each chart.

* **Closing Price Trends by Stock:**

This line graph depicts historic data regarding the closing prices of several equities, including some famous tech companies, and another one for Bitcoin (BTC-USD) cryptocurrency. It clearly shows that during the period examined, the price of Bitcoin was not stable but had some great fluctuations, which shows big volatility. And Tesla (TSLA) is really not such a good example, particularly in times of very rapid growth and then correction. On the other hand, for companies like Google (GOOG) and Amazon (AMZN), closing prices demonstrate a much more balanced and systematic tendency to increase in value. This can be a very good help in understanding the various risk profiles that are available under each kind of asset and help apply the same into the strategy of portfolio diversification for balancing growth and stability.



* **Market Share by Trade Volume:**

The pie chart above shows the market share of some selected assets relative to the volume of their transactions. From the chart, 99.3% of all the asset exchange volume represents Bitcoin. But while volatility concerns are intrinsic, interest from investors remains high. However, this large share of the market ensures a continuous level of liquidity for Bitcoin. The way in which trade volume is distributed over the assets would show how effective the volume analysis in portfolio risk and investment plans is being assessed. The heatmap of correlation helps to investigate the relations of different trade data elements against each other within a single stock.



* **Correlation Heatmap of Stock Features:**

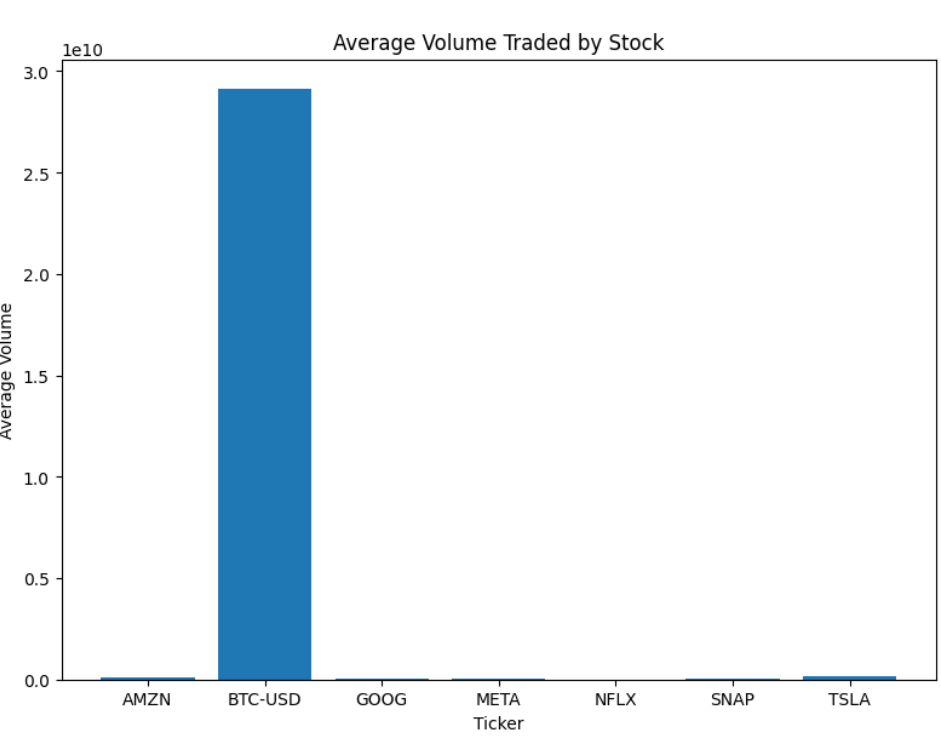
From this angle, there are only strong pieces of evidence reflecting the fact that Open, High, Low, or Close prices tend to move in tandem throughout a typical trading day. The only marginally less correlation with the volume was for these prices, which might be explained by factors such as market announcements or occurrences that have a bearing on trading behavior despite the absence of price fluctuation. This heatmap highlights the dynamics of measurements of stock trading in a manner that has an impact on choice of algorithmic trading methods and risk management



* **Average Volume Traded by Stock:**

The following bar graph reveals the trading volume for each stock averaging over a day, hence portraying the how liquid the asset is:

Indeed, according to an average volume exchanged per market day basis, Bitcoin (BTC-USD) is almost 20 times more popular than Amazon (AMZN) or Google (GOOG). Such information would be useful to a portfolio liquidity manager, as he or she would know in the first place that assets with a larger average volume allow traders to enter and exit positions much more easily than those with lower average volumes. First, the average traded volume could reflect the interest and visibility of investors on the stock, traits that could have a bearing with sentiment.

****

* **Portfolio Risk- Return Profile:**

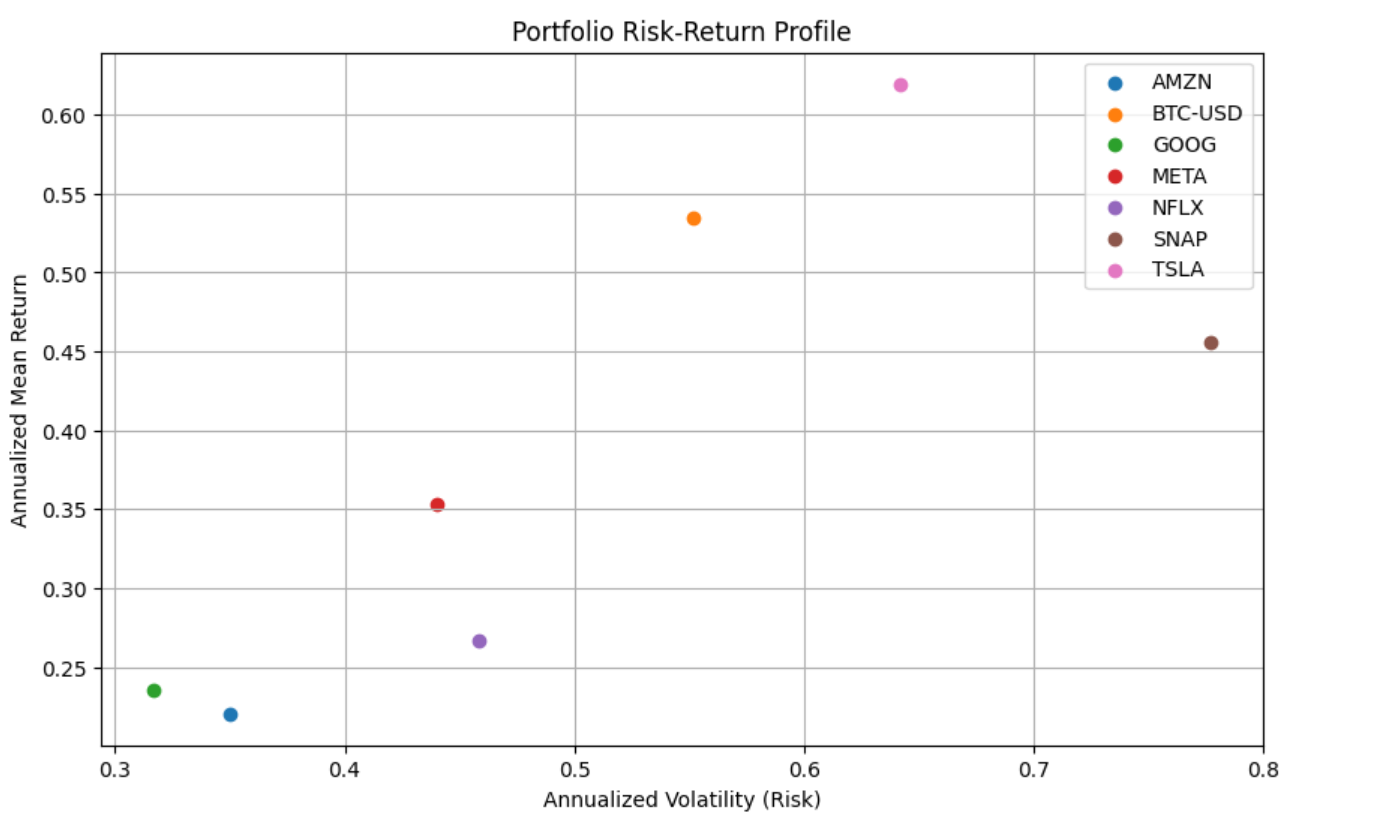
Our approach allows for an analysis and presentation of the possible investment performance in a contextually comparative manner. Risk-return profiles of a wide selection of assets, including stock and cryptocurrency holdings, mean annualized return, and corresponding annualized volatility, will be visually represented by us. It is clearly visible that each asset has a differentiated risk and return profile. For instance, NFLX and SNAP are placed in the bottom left position, which indicates lower expected profits with lower price fluctuation.

This clearly indicates that the risk and return generated from each asset are different. For example, NFLX and SNAP are towards the bottom left, indicating lower expected returns, but also less risk due to price volatility. The upper right corner of the chart is the place where lie assets with a superior risk-return profile, such as TSLA and AMZN, indicating more volatility but a higher predicted return. It brings clarity for the investors who stand to gain from this strategy, as it clears which assets would fit their goals for return and risk tolerance.

Perhaps because of their distinct market dynamics and associated risks, those assets, like BTC, with lesser returns and more pronounced price swings than other technology enterprises, seem to defy ordinary notions of investment. This clearly proves the fact that the different kinds of investments can coexist in the same business. Indeed, this chart outlines the risk and return profiles for several technology companies in this space: from the highly variable TSLA to the comparatively stable META.

The risk-return scatter map is very helpful as it makes it simple for the investors to assess likely investments within seconds. In this manner, with the help of this visual analysis, the investor can rebalance his portfolio, taking into consideration both his return requirements and the level of risk that is acceptable to him through strategic asset allocation. The fact is that its distance from the origin on the map represents the full risk profile of the asset, with a steeper angle meaning a bigger risk-reward ratio. The figure underscores the importance of diversifying by combining those assets that can maximize returns for a given level of risk and improve portfolio performance. With the help of this graph, the investor can form an adequately diversified or a well-balanced portfolio.

Most investors would like to obtain the highest risk-return profile by selecting an optimum combination of assets to build a diversified portfolio. Investors can combine lower-risk assets with steady returns and higher-risk assets with potential for significant growth to develop a strategic investment plan that fits their overall financial goals and risk tolerance.

****

**Analysis of Variance (ANOVA):**

We investigated to determine if there were any differences in the closing prices of several equities such as TSLA, BTC-USD, SNAP, AMZN, GOOG, NFLX, and META. We used a statistical method called Analysis of Variance (ANOVA) to compare the means across different groups and test our hypotheses.

The Formula we used to represent the ANOVA model:

**Yij​=μ+τi​+εij​**

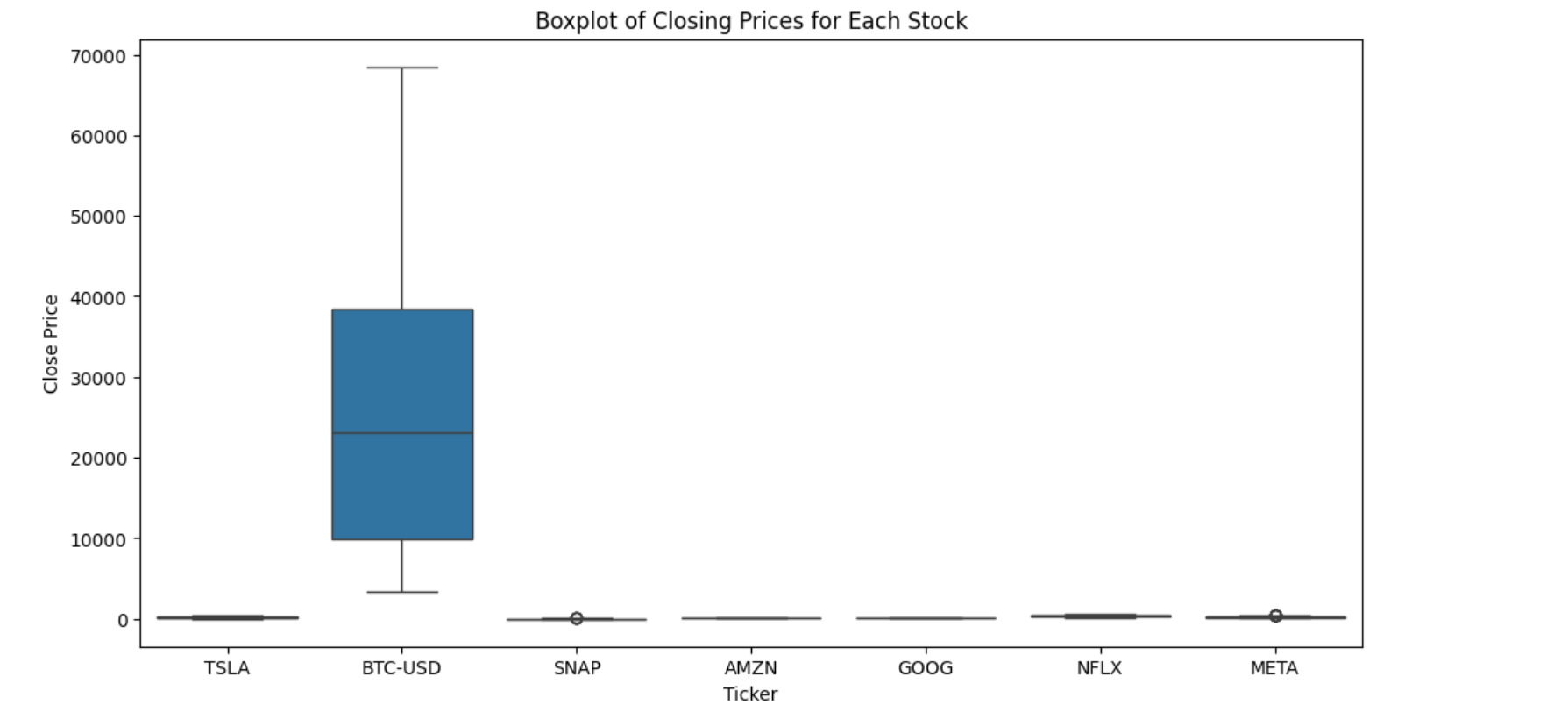
**Where:**

* **Yij:** This is the quantity that has been observed, such the closing price of a stock for group I on day j (in your example, the group is the ticker).
* **μ:** This is known as the grand mean, or the average closing price of all stocks combined.
* **τi**: This is associated with a certain stock group, often known as the "ith" group or ticker. Each stock in this group has an average closing price that is different from the average closing price of all stocks, which is also referred to as the overall mean (μ) or grand mean. This basically shows the deviation between the group's average closing price and the average closing price of all stocks.
* **εij**: Also known as the residual or error, this word denotes the discrepancy between the closing price that was observed and the value that was predicted using the group mean. It is just random variation or noise in the data.

Looking at the above ANOVA output, the closing prices differed significantly among groups. A high F-statistic, along with a p-value of less than the conventional alpha of 0.05, are all supportive of this conclusion. Those numbers clearly indicate that at least the average closing price of one share greatly differs from the rest.

The closing prices boxplot for all stocks shows that the closing prices for Bitcoin (BTC-USD) are way above the others. This is made obvious from a bigger interquartile range, signaling a greater volatility of the closing price of Bitcoin. On the other hand, equities with less closing price fluctuation and low closing prices are META, NFLX, SNAP.

This, paired with the boxplot visualization, presents strong evidence that the difference in means is a critical variable with a big effect on the possible return and risk characteristics of the investment portfolio. Therefore, the expected returns by the strategy to be used and the risk tolerance level of the investing strategy must be taken into account.



**CAPM Analysis:**

The Capital Asset Pricing Model (CAPM) is a key financial model that connects an investment's expected return to its risk compared to the overall market. The idea assumes that investors should receive compensation for two things: the time value of money and the level of risk involved.

The Formula used for Calculating the CAPM results is:

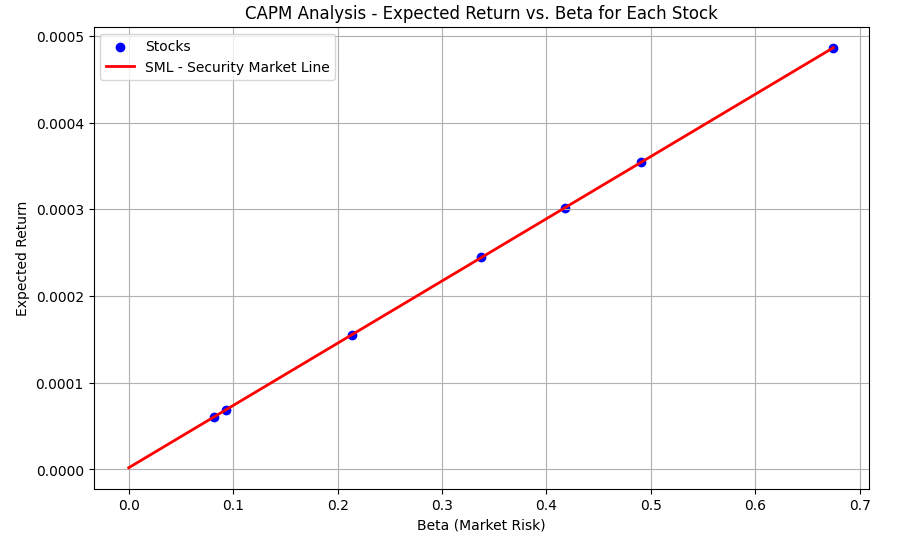
**E(Ri​)=Rf​+βi​⋅(E(Rm​)−Rf​)**

**Where:**

* **E(Ri​):** This is the estimated return of Investment.
* **Rf​:** It is the risk-free rate, representing the time value of money.
* **βi​:** It is the beta of the investment “i”
* **E(Rm​):** Expected return of the market
* **​E(Rm​)−Rf​:** Market risk Premium

We analyzed the daily returns of individual stocks as part of our study and contrasted them with the average daily return of the market, which we calculated by modelling the average daily return of all the stocks that were included. We evaluated each company's volatility in respect to the market to determine its beta. First, we converted the annual yield of a standard US Treasury bill into a daily rate to determine the risk-free rate.

The risk-return profiles of specific stocks are shown in the graph "CAPM Analysis - Expected Return vs. Beta for Each Stock". The CAPM algorithm is used to determine the beta and expected return of each blue dot on the graph, which stands for a stock. The Security Market Line (SML), a theoretical depiction of the risk-return tradeoff in the market, is represented by the red line. The degree to which the stock dots' performance aligns with the CAPM projection is indicated by how close they are to the SML. This graphic tool helps investors assess if the expected return of a stock, as well as its risk (as indicated by its beta) are well balanced.



**KEY FINDINGS:**

We start the journey to analyze from the data of the stock market, where we shall specifically look at closing prices, volumes of trade, and overall volatility of a few popular stocks and cryptocurrencies. We estimated expected returns relative to the overall market and measured the risk specific equities using statistical tools, which include the Capital Asset Pricing Model (CAPM) and Analysis of Variance (ANOVA). The results of our paper are important, as they add to the understanding of the investment risks and market behavior.

We interpret these results against the lens of contemporary financial theory and provide practical insights into optimizing portfolio management practices and rebalancing investment strategies. We present these important discoveries within the context of an analytic framework that provides a road map across the often-stormy waters of the financial markets.

**ANOVA Analysis for the Dataset:**

An ANOVA test of if the closing prices are significantly different or not for the stocks represented by each symbol was conducted. The summary has been tabled below:

**Sum of Squares (sum\_sq)** displays both the variance because of group interactions and the overall variance within the data. There is, therefore, a significant variation of closing prices among stocks in the 'Ticker' group, with an approximate sum of squares of 9.78997 × 10¹¹.

The 'Ticker' group has 6 **degrees of freedom (df)**, precisely the number of levels (stocks) minus one. The total number of observations, 10324, less the number of levels, 607, this difference of 9718 is the number of residual degrees of freedom.

The F-value, which is **3073.697472**, measures the deviation of the means of several groups from the general mean. It is adjusted for group variance. The higher the F-value, the meaning of at least one group is substantially different from the meanings of the others. The rounded to **0.0 P-value** is signaling that the observed F-value is very much significant, thus rejecting the null hypothesis of equal variances. By doing this, it, therefore, concludes that the closing prices of different stocks do show a significant difference.

Anova results

From the ANOVA results, it is shown that the closing price performance definitely has differences among different stocks at a 95% confidence level.

This further goes ahead to emphasize the importance of diversifying holdings. Investors should buy a lot of different stocks to distribute the risk brought about by price swings in any one stock. Differences in the closing prices can also signal to investors which businesses may be more in line with their investment objectives, whether these are businesses that value stability or are ready to take on more risk to get potentially bigger returns.

**CAPM Analysis for the Dataset:**

Current research has been conducted to test the risk-return nature of stocks that include renowned tech firms and cryptos under the Capital Asset Pricing Model (CAPM). One of the important components of the CAPM study is the figure for beta. It is a measure of how sensitive the returns to stock are with respect to market fluctuations. Where β > 1, it means the stock is more volatile than the market, and where β < 1, it means it has less volatility.

Here are the beta levels and predicted returns for each stock according to the CAPM analysis:

With a beta of 0.893255, the stock price of Amazon is more volatile than the market. This thus justifies the risk associated with a higher return than average and that too by a spread of 0.00069. But then there is more risk attached to this higher return. On the other hand, the beta for Bitcoin (BTC-USD) is less than one, standing at 0.417975. This value points out that the price of Bitcoin is less erratic compared to the behavior in the general market. However, its estimated return is just 0.000302. This is not very high, considering the level of risk involved.

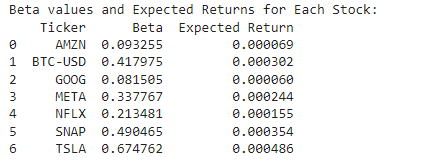
This might be because of the special risks associated with cryptocurrencies. Google (GOOG) has volatility like the average market as revealed by its beta of 0.881505. It has a fair predicted return of 0.000660.

Meta Platforms (META) has a beta of 0.337767, indicating that it is much less volatile compared to the general market. With a degree of risk involved, the estimated return of 0.000244 seems sensibly entailed.

Netflix (NFLX) has a beta of 0.213481, indicating considerable volatility. It also shows an expected return of 0.000155, and hence, it is a less risky way of investing.

On the other hand, the beta of Snap Inc. (SNAP) is 0.490465, which is lower and less erratic than the market. Moreover, it has a forecast return of 0.000354.

Of all selected equities, Tesla (TSLA) has the highest beta value, 0.674762, which means higher volatility for sure. It also gives the highest expected return out of all equal to 0.000486 by its value. It means that higher risk may be expected to offer higher rewards.



In this context, the study related to Capital Asset Pricing Model (CAPM) shows that the risk profile of each stock is much different. In other words, investors who look for some level of high returns on their investment will be somewhat attracted to companies like Amazon and Tesla, with higher beta values and expected returns indicating a potentially bigger payoff for the accepted risk. On the other hand, investors who are less risk-tolerant would probably view it otherwise and think of investment in such companies as META and NFLX, with low expected returns and beta values indicating somewhat stability of the investment. It is highly important to recollect the fact that the beta value of Bitcoin is marked in its market conditions, which are less volatile than expected. And these are related to its unique qualities as a non-stock asset class. The data further underscores the importance of both the beta and projected returns while charting down a diversified portfolio.

They inform the retail investor about the risk and potential rewards of the various investing options. This will result in a well-educated, well-balanced portfolio due to the incorporation of these insights into an investment plan and will meet the expected return of the investor as well as risk tolerance.

**INVESTMENT RECOMMENDATIONS:**

From the CAPM research, Tesla (TSLA) and Amazon (AMZN) should have higher returns than other equities since they have larger beta values than other stocks. "These equities offer an attractive risk-return trade-off to investors willing to take on added market risk in their quest for growth." And further, Tesla's dominant share in the electric vehicle space shows that there is a whole lot of growth ahead for it, besides having a gigantic platform in e-commerce and cloud computing. One should critically assess one’s investment horizon and risk appetite to bring the high volatility, which is an innate factor for these stocks, to a lower level.

Conservative investors or those whose objective is portfolio stability may prefer equities with lower beta levels. Netflix (NFLX) and Meta Platforms (META) are some of them. Both belong to the growth-oriented technology space but have more moderate return profiles since they are less exposed to market dynamics. Worth noting, the Barclays ETNs are designed for those investors who seek to reduce risk without forgoing their access to opportunities in popular market sectors. Besides, investors who seek further diversification from the ordinary stock and take advantage of the explosion in the digital asset market have another alternative flowing from the unique positioning of Bitcoin (BTC-USD), courtesy of its low beta and high trading volume.

In conclusion, a strategy with a combination of low-beta stocks providing stability and high-beta stocks that have room to run may. Thus, investors must make sure they choose their stocks based on sound analytical statistics and their specific level of risk tolerance market outlook.

**CONCLUSION:**

Here, we investigated lots of equities, efficiently, from diverse industries, including renowned tech firms and cryptocurrency. In this research, we used sophisticated statistical methods and found different patterns in volatility, trade volume distribution, and market liquidity. Bitcoin's trading volume vastly exceeded what typical stocks would do. This is an illustration that different asset classes indeed have very different market behaviors. This result also shows the necessity of well-informed decisions that are based on the planned risk-return profile.

To calculate the expected returns from beta values of each stock, which indicate the various degrees of risk entailed in every stock, we undertook CAPM analysis. Those investors who are more inclined to the growth prospects of a stock and are willing to take larger levels of risk will be most attracted to stocks like TSLA and AMZN, both noted for their higher volatility and potential for bigger returns. Nevertheless, less daring investors could pay attention to less volatile stocks like NFLX and META, as they pose steady investment opportunities. The ANOVA result has shown that there is some difference in closing prices between these companies.

This goes on to emphasize just how essential it is to make intelligent choices of your assets in a bid to gain maximum results. The different performance of the stocks clearly shows that a personalized investment can be made tailored to the financial goals and the willingness of an investor to take risks. Research indicates that for any financial decision, data-driven decisions can be very helpful.

In a nutshell, it is only through going largely in the market data analysis that the investors will come up with valuable insights for possible investment and tailor a portfolio that matches their unique goals. A flexible investment strategy that can capitalize on emerging trends, react to changing market circumstances, and that can strike a balance between risk and reward is laid by this project. Thus, the analytical results that emanate from financial market research can be applied to strategists and investors. These findings provide quantifiable ways by which the complex behavior of financial markets may be understood. Investment decisions can get to be more than gut or take a stronger, analytically grounded approach based on the evidence. In the end, this becomes the most profitable and informed investing decision.

**BIBILOGRAPHY:**

[1] E. F. Fama and K. R. French, "The Capital Asset Pricing Model: Theory and Evidence," SSRN Electronic Journal, 2003. [Online]. Available: <https://doi.org/10.2139/ssrn.440920>

[2] R. Gueorguieva and J. H. Krystal, "Move over ANOVA: progress in analyzing repeated-measures data and its reflection in papers published in the Archives of General Psychiatry," Archives of General Psychiatry, vol. 61, no. 3, pp. 310–317, 2004. [Online]. Available: <https://doi.org/10.1001/archpsyc.61.3.310>

[3] T. Rishi, "Stock Market Analysis Using Linear Regression," Proceedings of the Jepson Undergraduate Conference on International Economics, vol. 4, no. 1, Article 4, 2022. [Online]. Available: <https://scholarworks.uni.edu/jucie/vol4/iss1/4>

[4] Yahoo! Inc., "Yahoo Finance," 2024. [Online]. Available: <https://finance.yahoo.com/>

[5] F. Zhou, Z. Qun, D. Sornette, and L. Jiang, "Cascading Logistic Regression onto Gradient Boosted Decision Trees to Predict Stock Market Changes Using Technical Analysis," Swiss Finance Institute Research Paper No. 18-50, July 24, 2018. [Online]. Available: <https://doi.org/10.2139/ssrn.3218941>

[6] N. Adlakha and A. Katal, "Real Time Stock Market Analysis," in 2021 International Conference on System, Computation, Automation and Networking (ICSCAN), Puducherry, India, 2021, pp. 1-5. [Online]. Available: <https://doi.org/10.1109/ICSCAN53069.2021.9526506>

[7] S. George and M. Changat, "Network approach for stock market data mining and portfolio analysis," in 2017 International Conference on Networks & Advances in Computational Technologies (NetACT), Thiruvananthapuram, India, 2017, pp. 251-256. [Online]. Available: <https://doi.org/10.1109/NETACT.2017.8076775>