**Data Analytics Capstone Topic Approval Form**

**Student Name:** Jonathon Fryman

**Student ID:** 000974544

**Capstone Project Name:** Analyzing Tech Stock – Using Time Series to forecast future stock performance

**Project Topic**: Prices within the stock market are known to be volatile. This can be caused by circumstances specific to a certain company or through general market volatility caused by external factors such as government policy, industry regulation, or general economic performance. The focus of this project will be to create a forecast of future stock performance for some of the biggest technology companies.

**This project does not involve human subjects research and is exempt from WGU IRB review.**

**Research Question:** To what extent can a company’s future daily per-share closing stock dollar value be accurately forecast?

**Hypothesis**: The hypothesis is that the associated Mean Absolute Percentage Error(MAPE) score for predicted stock closing prices will be below 20%. A MAPE score below 20% indicates the difference between the dollar value for the predicted and actual stock closing price is smaller than 20% which is a generally considered good MAPE score.   
  
**Null hypothesis**- The null hypothesis is that the MAPE score is 20% or greater. This would indicate a less than 80% similarity on the predicted dollar value of the stock closing price.  
  
**Alternate Hypothesis**- The alternate hypothesis is that the MAPE score is less than 20%. This would indicate a greater than 80% similarity on the predicted dollar value of the stock closing price.

**Context:** Due to market fluctuation and the need for prudent investing, it would be of great benefit to current and future investors to have the capability of utilizing historical trends of a specific tech stock’s performance to have confidence in the best time to make an investment. Using time-series analysis, the objective is to extract meaningful statistics and other characteristics of the data.

**Data:** The dataset to be used for this analysis is publicly available at <https://www.kaggle.com/datasets/evangower/big-tech-stock-prices>.

The full dataset consists of 14 unique csv files. Each file is associated with a specific tech companies stock performance for trading days beginning in 2010. These files contain the trading date, open and closing price, the high and low value for that trading day, the adjusted close value, and overall volume.   
  
The tech companies included within this dataset are Apple, Adobe, Amazon, Salesforce, Cisco, Google, IBM, Intel, Meta, Microsoft, Netflix, Oracle, and Tesla.

The data being used for this analysis is provided under a [CC0 1.0 Universal Public Domain Dedication](https://creativecommons.org/publicdomain/zero/1.0/" \t "_blank) license that allows users to share and adapt the data with proper credit given to the original data provider

**Data Gathering:** The data-gathering methodology to be used for this analysis is documents and records. This methodology consists of examining existing data. For this specific analysis, this includes examining existing records related to historical stock prices for 14 tech companies over a period beginning in 2010. This includes historical open and close prices, and overall volume.

**Data Analytics Tools and Techniques**: For this report, the primary data analytic technique to be used is a time series analysis. Time series analysis is a method of analyzing a sequence of data points collected over consistent internals of time. This analysis technique allows insight into how specific variables change over time.

Time Series analysis requires many data points to ensure consistency and reliability. Ensuring that there is sufficient data helps ensure that any trends or patterns identified are not outliers and can account for seasonal variance. Additionally, time series data can be used for forecasting/predicting future data based on historical data.

**Justification of Tools/Techniques:**

The planned tools to be used for this analysis include the Python programming language, Jupyter Notebooks, and Tableau.

Jupyter Notebooks will be utilized for any required data cleaning, data analysis, data visualization, and the development of a time series forecast to predict future fluctuations in the stock price for the associated tech companies. This will be achieved using the Python programming language and its many available packages and libraries that facilitate data analysis and model development.

Additionally, Tableau is planned to be utilized to create additional data visualizations for use within the written report associated with this project.

**Project Outcomes**: The primary project outcomes associated with this analysis is to identify any trends or seasonality within the fluctuation of daily closing prices associated with 14 large tech companies. This will allow investors to be better informed regarding the most prudent time for investment and provide a general forecast for an expected value of stock price at the end of a given trading day.   
  
This will also facilitate a prediction of future stock price performance to give the investor a baseline gauge of what type of return on investment may be expected presuming that unforeseen circumstances within the company or caused by external factors are not encountered during the forecasted period.

**Projected Project End Date**: 4/15/2023

**Sources**:

* *Creative Commons License Deed*. Creative Commons - Attribution-NonCommercial-ShareAlike 4.0 International - CC BY-NC-SA 4.0. (n.d.). Retrieved February 23, 2023, fromhttps://creativecommons.org/licenses/by-nc-sa/4.0/
* Gower, E. (2023, January 30). *Big Tech Stock prices*. Kaggle. Retrieved February 23, 2023, fromhttps://www.kaggle.com/datasets/evangower/big-tech-stock-prices
* *Time series analysis: Definition, types, techniques, and when it's used*. Tableau. (n.d.). Retrieved February 23, 2023, from https://www.tableau.com/learn/articles/time-series-analysis#definition
* Li, S. (2018, September 5). *An end-to-end project on time series analysis and forecasting with python*. Medium. Retrieved February 23, 2023, from <https://towardsdatascience.com/an-end-to-end-project-on-time-series-analysis-and-forecasting-with-python-4835e6bf050b>
* Stephen Allwright. (2022, December 6). *What is a good MAPE score? (simply explained)*. Stephen Allwright. Retrieved February 24, 2023, from https://stephenallwright.com/good-mape-score/

**Course Instructor Signature/Date:**

The research is exempt from an IRB Review.

An IRB approval is in place (provide proof in appendix B).

Course Instructor’s Approval Status: Approved

Date: Click here to enter a date.

Reviewed by:

Comments: Click here to enter text.