

Stock Prediction Analysis

Group 8

Data Science Capstone Project Launch Report

1/20/2025

Team Members:

Name: Robert Lignowski - rml345@drexel.edu

Name: Udit Shah - us54@drexel.edu

Name: Ahmad Javed - aj3235@drexel.edu

Name: Steven Sullivan - sas683@drexel.edu

The System/Product

System/Product Name: Stock Price Time Series Forecast

Introduction:

Given the most recent election cycle, there has been a lot of discussion and controversy surrounding upcoming policies in regards to trade and importing/exporting goods. One of the policies that has garnered a lot of questions is in regards to tariffs, specifically those placed by the United states on other countries. Along with these tariffs, the increase in political involvement of Tesla CEO Elon Musk has made the car industry an area of interest in the coming years.

The goal of this project is to learn about and implement time series forecasting to make predictions based on historical data. Our project will be focused on four automobile companies representing 4 different areas of interest: American, European, Asian, and Electric.

Electric:

Tesla - TSLA

Tesla went public on June 29, 2010. The company is led by CEO Elon Musk, who has been at the helm since 2008. As of January 18, 2025, Tesla's stock price stands at \$426.50, reflecting its strong market presence and investor confidence. Over the years, Tesla has sold more than 3 million vehicles globally, cementing its position as a leader in the electric vehicle industry.

Asian:

Toyota - TM

Founded in 1937, Toyota's headquarters and several of its manufacturing plants are located in Toyota, Aichi, Japan. Formerly named "Koromo," the close ties between the company and the city resulted in it being renamed in 1959. Today, the company is the largest automobile manufacturer in the world, producing more than 10 million cars every year.

American:

Ford - F

The Ford Motor Company was founded in 1903 and named after its founder Henry Ford. Ford is recognized as one of America's most prominent automobile companies. Based out of

Dearborn, Michigan, a suburb of Detroit, the Ford Motor company sells approximately 4 to 5 million vehicles per year.

European:

Volkswagen - VW

Founded in 1937 and headquartered in Wolfsburg, Germany, Volkswagen is one of the largest automobile manufacturers in the world. Known for its iconic models like the Beetle and Golf, Volkswagen produces over 10 million vehicles annually, catering to a wide range of consumers. The company is a cornerstone of the Volkswagen Group, which owns several other prominent brands, including Audi and Porsche.

Our analysis and predictions will be based on available stock prices and google trends over the past few years, to help us identify what are the most influential factors in determining stock prices. The final outcome of the project will be individualized predictions for each of the twelve automobile companies, as well as a general prediction on what the market will look like for each of the groups in our project.

Highlighted Features/Functionality:

The Stock Price Time Series Forecast project offers a multi-dimensional analysis by integrating diverse data sources, including stock price trends and Google Trends, to provide a comprehensive understanding of market dynamics. By focusing on industry-relevant external factors such as tariff policies, political events, and market shifts, the project ensures a holistic approach to stock price forecasting. The analysis is centered on four automobile categories: Electric (Tesla), Asian (Toyota), American (Ford), and European (Volkswagen) to represent international perspectives and regional variations in the automobile industry.

To achieve accurate predictions, the project employs advanced time series forecasting models such as ARIMA, ensuring robust trend analysis, along with Logistic Regression to interpret key drivers of stock price fluctuations. Custom data visualizations are a key feature, designed to clearly present trends and insights, including comparative trend graphs across companies and regions. Additionally, these visualizations will highlight the relationship between external factors such as tariffs and EV adoption rates, providing valuable insights for stakeholders. By leveraging these innovative features, the project differentiates itself as a comprehensive and insightful tool for stock price analysis in the automobile sector.

Sponsor or Proxy User:

At this moment we do not have an external sponsor for this project, but we are open to work with anyone who has experience with stock prices, the automobile industry, and any other related fields.

Issues:

Potential issues may arise from certain companies having more sources of data than others (i.e., a company may have more social media activity and/or more news articles written about them compared to other companies). To counter this, the team will focus on obtaining as much data as possible to ensure all four companies have the same amount of data to be used for the project.

The Team

Team Name: Group 8

Team Members and their specialties:

Each Team Member has been given a group of car companies for this project. Each group member will have their own creative freedom for data analysis and exploration. Once each member has filtered and evaluated their respective data, the team will convene and work on creating a time forecasting model and will implement it on all car companies in this project. Team members will coordinate with one another on any ideas or preferences for this project.

Uditi Shah: Masters in Data Science with skills Python, AI, Machine Learning, SQL. Recently worked as a data science intern at Dataing. In this project, I want to focus on data analysis and show ongoing trends, graphs, in the stock market.

Robert Lignowski: Masters in Data Science, with skills in Python and the OpenBB package. Recently worked as a data analyst with a professor in the Health Sciences department.

Ahmad Javed: Masters in Data Science with skills in Python, SQL, Microsoft Excel/ VBA, and Data visualization softwares such as Tableau and PowerBI. Currently working as a Data Analyst for a biotechnology company in the Northeast.

Steven Sullivan: Masters in Data Science with skills in Python, SQL, Java, and C, alongside experience in machine learning, time series forecasting, and data visualization.

Team Communication:

The team has created a discord group for immediate access to one another. The discord group is where the team will share insights, project related articles, videos and generally set up a time to discuss the project when everyone is available.

The team will also be meeting with our professor every Tuesday from 7:30-8:00 pm to get insight into our project and what we can do to enhance the results of our project.

Team Issues:

For Phase II, the team plans to use ARIMA (AutoRegressive Integrated Moving Average) which the team has limited knowledge about. The goal of weeks 7 - 8 will be to learn about ARIMA and how it can be implemented into the project for Phase II.

Plan

Week	Objective
1	Create Group <ul style="list-style-type: none"> ● Introduction to teammates ● Identify project filed and goals
2	Create Launch Report <ul style="list-style-type: none"> ● System/Product ● Team ● Plan
3-4	Obtain and Combine Data Sources <ul style="list-style-type: none"> ● Find source(s) for data ● Stock Prices ● Google Trends ● News Articles ● Social Media Posts (X, Facebook, Instagram)
5	Data Exploration and Analysis <ul style="list-style-type: none"> ● Find common trends amongst car companies ● Display initial findings with visualizations
6	Data Acquisition and Preprocessing Report <ul style="list-style-type: none"> ● Identifying Data ● Data Acquisition Process
7-8	Learn about Modeling Algorithms for Phase II: <ul style="list-style-type: none"> ● (1st Modeling Algorithm) Learn about ARIMA (AutoRegressive Integrated Moving Average) <ul style="list-style-type: none"> ○ Goal is to acquire skill of ARIMA and learn how to implement it into a project

	<ul style="list-style-type: none"> • (2nd Modeling Algorithm) Logistic Regression <ul style="list-style-type: none"> ◦ Since team is familiar with Logistic Regression, the focus will be more on learning the first algorithm while sharpening skills on Logistic Regression
9	Exploratory Data Analytics Report <ul style="list-style-type: none"> • Analysis the basic metrics of variables • Non-graphical and graphical univariate analysis • Missing value analysis and outlier analysis • Feature engineering and analysis
10	Status Presentation <ul style="list-style-type: none"> • Present Findings

Learning Outcomes:

- How to obtain data between different types of sources
 - Stocks and Trends
 - News articles
 - Social Media Activity
- How to set up data for Time series exploration
- What is ARIMA and how to implement it (New Skill)
- How to ensure all companies are represented equally so that analysis is fair and accurate

Table of Contributions

The table below identifies contributors to various sections of this document.

	Section	Writing	Editing
1	Project	Ahmad, Steven	Robert, Udit
2	Team	All Members	All Members
3	Plan	Robert, Udit	Ahmad, Steven