

A decorative graphic on the left side of the slide featuring a blue parallelogram and a light green parallelogram, both tilted at an angle, set against a dark blue background with diagonal stripes.

Stock Price Time Series Forecast: Data Challenges and Preparation

Group 8: Ahmad Javed, Steven Sullivan,
Robert Lignowski, Udit Shah



Introduction

- Recent discussions and controversies have emerged regarding trade policies, particularly U.S. tariffs on other countries.
- The increasing political involvement of Tesla CEO Elon Musk has drawn attention to the car industry.
- The project aims to use time series forecasting to predict trends for four automobile companies representing American, European, Asian, and Electric markets.



Features

- The project integrates diverse data sources, including stock price trends and Google Trends, to analyze market dynamics with a focus on external factors like tariffs and political events.
- It examines four automobile categories—Electric (Tesla), Asian (Toyota), American (Ford), and European (Volkswagen)—to capture international and regional industry variations.
- Advanced time series forecasting models like ARIMA and Logistic Regression are used for accurate predictions, complemented by custom visualizations to highlight trends and external influences on stock prices.

Company Stocks

TESLA (TSLA)



TOYOTA (TM)



VOLKSWAGEN (VWAGY)



FORD (F)





Data Sources

- Main source:
 - Historical stock data from Investing.com
 - Downloaded from “General - Historical Data” section
 - Reliable and up-to-date
- Data covers a period from 2020-2025 for the four main companies
- These companies were chosen because:
 - Significant presence in global automobile industry
 - Mixture of traditional and electric vehicle manufacturers
- Insights:
 - Market behavior
 - Investor sentiment
 - Financial performance over time



Data Acquisition Process

- All data was downloaded as a CSV file
 - Publicly available, so no need for API
- Preprocessing:
 - Standardize date formats
 - Handling Missing Values
 - Ensuring uniformity by renaming headers
- Data Cleaning
 - Removing unnecessary columns
 - Renaming headers



Data

	Unnamed: 0	Date	Price	Open	High	Low	Vol.	Change %
0	1279	2020-01-02	142.24	142.00	142.35	141.53	118090.0	1.21
1	1278	2020-01-03	140.75	141.23	141.42	140.30	174890.0	-1.05
2	1277	2020-01-06	140.77	139.46	140.89	139.46	150390.0	0.01
3	1276	2020-01-07	141.51	142.13	142.43	141.36	123670.0	0.53
4	1275	2020-01-08	141.16	140.91	141.58	140.60	117710.0	-0.25

Data Acquisition Issues



Tesla: Missing values, inconsistent volume formatting, outliers.



Volkswagen: Volume data inconsistencies ('K' vs. 'M'), missing values, market volatility.



Toyota: Similar volume issues, missing values filled, NYSE data used.



Ford: Formatting adjustments, but no missing values.

TOYOTA



Data Preprocessing Steps

- 1) **Handling Missing Data:** Filled missing values with median.
- 2) **Standardizing Volume Formatting:** Converted 'K' to 1,000 and 'M' to 1,000,000.
- 3) **Cleaning Non-Numeric Characters:** Removed '%' from "Change %" column.
- 4) **Date Standardization:** Converted "Date" column to datetime format.
- 5) **Ensuring Consistency Across Datasets:** Standardized column names and aligned time frames.

Example Data Transformation

	A	B	C	D	E	F	G
1	Date	Price	Open	High	Low	Vol.	Change %
2	1/2/2025	9.08	9.22	9.23	9.03	307.21K	-2.47%
3	#####	9.31	9.36	9.41	9.27	451.22K	-0.53%
4	#####	9.36	9.42	9.48	9.36	759.57K	-0.74%
5	#####	9.43	9.44	9.49	9.4	427.29K	1.51%
6	#####	9.29	9.3	9.37	9.2	436.21K	-0.21%
7	#####	9.31	9.26	9.31	9.25	171.74K	0.22%
8	#####	9.29	9.13	9.3	9.09	415.86K	-1.28%
9	#####	9.41	9.4	9.71	9.16	465.19K	2.84%
10	#####	9.15	9.37	9.37	9.15	489.04K	-0.76%
11	#####	9.22	9.39	9.44	9.16	466.01K	-1.07%
12	#####	9.32	9.23	9.35	9.22	469.82K	0.43%
13	#####	9.28	9.23	9.31	9.14	622.17K	-1.38%
14	#####	9.41	9.51	9.53	9.26	551.74K	0.86%
15	#####	9.33	9.29	9.38	9.25	296.01K	0.27%
16	#####	9.31	9.34	9.36	9.23	449.15K	0.59%
17	#####	9.25	9.29	9.31	9.2	531.69K	1.31%
18	12/9/2024	9.13	9.09	9.19	9.08	707.63K	1.11%
19	12/6/2024	9.03	9.1	9.11	9	488.83K	1.23%
20	12/5/2024	8.92	8.91	8.97	8.87	561.73K	0.60%
21	12/4/2024	8.87	8.89	8.9	8.82	813.16K	1.91%



	A	B	C	D	E	F	G
1	Date	Price	Open	High	Low	Vol.	Change %
2	1/2/2025	9.08	9.22	9.23	9.03	307210	-2.47
3	#####	9.31	9.36	9.41	9.27	451220	-0.53
4	#####	9.36	9.42	9.48	9.36	759570	-0.74
5	#####	9.43	9.44	9.49	9.4	427290	1.51
6	#####	9.29	9.3	9.37	9.2	436210	-0.21
7	#####	9.31	9.26	9.31	9.25	171740	0.22
8	#####	9.29	9.13	9.3	9.09	415860	-1.28
9	#####	9.41	9.4	9.71	9.16	465190	2.84
10	#####	9.15	9.37	9.37	9.15	489040	-0.76
11	#####	9.22	9.39	9.44	9.16	466010	-1.07
12	#####	9.32	9.23	9.35	9.22	469820	0.43
13	#####	9.28	9.23	9.31	9.14	622170	-1.38
14	#####	9.41	9.51	9.53	9.26	551740	0.86
15	#####	9.33	9.29	9.38	9.25	296010	0.27
16	#####	9.31	9.34	9.36	9.23	449150	0.59
17	#####	9.25	9.29	9.31	9.2	531690	1.31
18	12/9/2024	9.13	9.09	9.19	9.08	707630	1.11
19	12/6/2024	9.03	9.1	9.11	9	488830	1.23
20	12/5/2024	8.92	8.91	8.97	8.87	561730	0.6
21	12/4/2024	8.87	8.89	8.9	8.82	813160	1.91



Data Analysis and Visualization

- For each company, we have analyzed our data and created visualizations to show how the stock prices have changed over the years
- The following graphs show the change in stock prices (at closing) for each of our companies
- Along with our graphs, we have identified over the past 5 years:
 - The highest stock price at opening
 - The highest stock price at closing
 - The highest stock price overall
 - The lowest stock price overall

Data Visualization (Tesla)



Best Open: \$475.90

Best Close: \$479.86

Highest Stock Price: \$488.54

Lowest Stock Price: \$23.37

Data Visualization (Volkswagen)



Best Open: \$42.10

Best Close: \$42.33

Highest Stock Price: \$48.72

Lowest Stock Price: \$8.57

Data Visualization (Toyota)



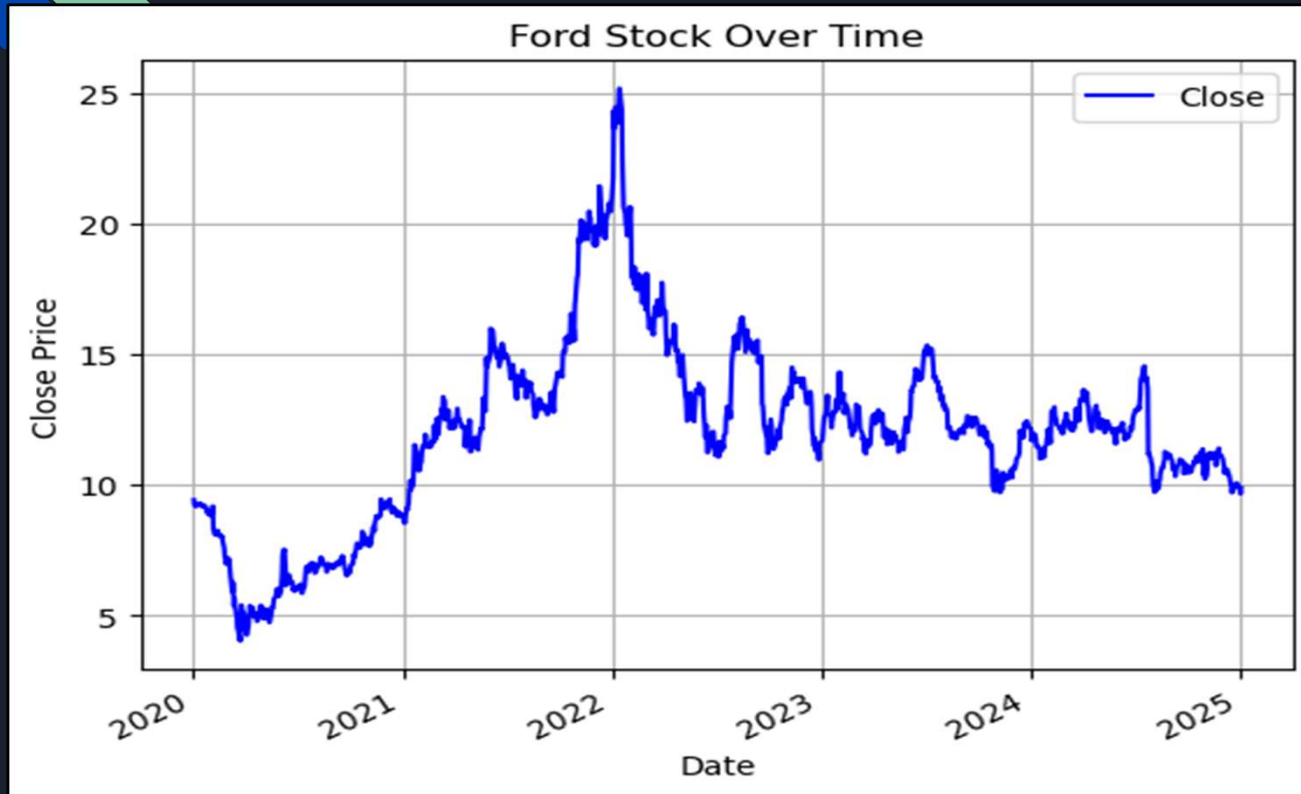
Best Open: \$255.00

Best Close: \$254.77

Highest Stock Price: \$255.23

Lowest Stock Price: \$108.01

Data Visualization (Ford)



Best Open: \$23.33

Best Close: \$23.63

Highest Stock Price: \$24.27

Lowest Stock Price: \$3.72



Pseudocode for Data Cleaning

1. Load the dataset from CSV file
 - Read the CSV file into a dataframe.
2. Clean and preprocess data
 - Convert the 'Vol.' column:
 - If it contains 'M', multiply by 1,000,000.
 - If it contains 'K', multiply by 1,000.
 - Remove non-numeric characters and convert to float.
 - Convert the 'Change %' column:
 - Remove '%' symbol.
 - Convert to float.
 - Convert financial columns ('Price', 'Open', 'High', 'Low', etc.):
 - Remove '\$' or any non-numeric characters.
 - Convert to float.
 - Convert the 'Date' column to datetime format.



Pseudocode for Data Cleaning

1. **Handle missing values**
 - Identify missing values.
 - Fill missing values with the median of the respective column.
2. **Perform Exploratory Data Analysis (EDA)**
 - a. Display basic statistics of the dataset (mean, median, standard deviation, etc.).
 - b. Check for missing values.
 - c. Plot time series of:
 - Closing Price vs. Date.
 - Trading Volume vs. Date.
 - Percentage Change vs. Date.
 - d. Identify the most common stock prices for 'High', 'Low', 'Open', and 'Close'.



Conclusion

In this report, we have detailed the process of acquiring and preprocessing historical stock data for Ford, Volkswagen, Toyota, and Tesla from Investing.com. We carefully selected these companies to analyze trends in both traditional and electric vehicle markets. Our acquisition process involved downloading structured CSV files, ensuring consistency across datasets, and standardizing formats for accurate analysis.

During preprocessing, we addressed challenges such as missing values, inconsistent volume formatting, and data noise. We employed statistical methods, such as median imputation for missing data and numerical conversions for stock volumes and percentage changes, ensuring data integrity. Additionally, date formats were standardized to facilitate time-series analysis.



Conclusion

Beyond technical preparation, it is crucial to recognize that **stock prices are influenced by a variety of external factors, including company news, leadership changes, and strategic investments.** For instance, **Tesla's stock surged to an all-time high following Donald Trump's 2024 election victory,** attributed to expectations of favorable policies and Elon Musk's influence. Toyota saw a **decline during the 2021-2022 chip shortage,** but investor confidence rebounded in 2023 due to its advancements in **solid-state battery technology.** Ford experienced **stock surges after the F-150 Lightning launch** but also faced declines after announcing **job cuts in 2023.** Volkswagen benefited from its **EV expansion post-Dieseldgate** but struggled in **2023 due to declining sales in China.**

With a **clean and well-structured dataset,** we can now proceed to **exploratory data analysis, feature engineering, and predictive modeling.** Understanding the interplay between stock movements and external factors will be essential in building **robust machine learning models** for stock trend prediction, ultimately aiding in more informed investment decisions.