

# PYTHON FINAL PROJECT

NASDAQ COMPANIES  
FROM 2010 TO 2020

PRESENTATION

By Merenini Princess Better



# BRIEF DETAILS ABOUT NASDAQ

NASDAQ (National Association of Securities Dealers Automated Quotations is one of the largest stock exchanges market in the world, headquartered in New York City. It was founded in 1971 and is known for being the first electronic stock exchange, allowing investors to trade securities digitally without a physical trading floor.



# OVERVIEW

This dataset provides insights into the daily market share value changes of Amazon, Apple, Google, Microsoft, and Tesla from mid-2010 to mid-2020, making it valuable for those analyzing stock trends over time.



# OBJECTIVES

1

**ANALYZE  
STOCK MARKET  
TRENDS**

2

**Handle data  
preprocessing  
efficiently**

3

**BUILD A  
PREDICTIVE  
MODEL USING  
LOGISTIC  
REGRESSION**

4

**OPTIMIZE  
MODEL  
PERFORMANCE**

5

**VISUALIZE AND  
INTERPRET  
RESULTS**

6

**DOCUMENT KEY  
FINDINGS AND  
RECOMMENDATIONS**

# PROBLEM STATEMENT

Investors and analysts need to understand stock market trends, price movements, and trading volumes for major companies like Amazon, Apple, Google, Microsoft, and Tesla.

day_date	close_value	volume	open_value	high_value	low_value
2020-05-29	317.94	38399530	319.25	321.15	316.4700
2020-05-28	318.25	33449100	316.77	323.44	315.6300
2020-05-27	318.11	28236270	316.14	318.71	313.0900
2020-05-26	316.73	31380450	323.50	324.24	316.5000
2020-05-22	318.89	20450750	315.77	319.23	315.3500
...	...	...	...	...	...
2019-12-21	405.59	14785210	410.29	413.00	400.1850
2019-12-22	405.59	14785210	410.29	413.00	400.1850
2019-12-25	425.25	8054720	418.36	425.47	412.6875
2019-12-28	430.38	9956827	435.00	435.31	426.1100
2019-12-29	430.38	9956827	435.00	435.31	426.1100



# DATA IMPORTATION

Data importation is a process of bringing in your dataset into your Notebook to enable you have a better Overview of the data you will be working with and also help you make your analysis and provide useful insights.

```
import pandas as pd  
  
company_data = pd.read_csv('CompanyValues.csv')  
company_name = pd.read_csv('Company.csv')  
  
company_data  
  
ticker_symbol      day_date    close_value      volume
```

# DATA CLEANING/ PREPROCESSING



this is a process in data analysis where you clean up your data set by removing Null values, Removing Duplicates, changing inconsistency etc

17524	TSLA	2019-12-21	405.59	99.0
17525	TSLA	2019-12-22	405.59	99.0
17526	TSLA	2019-12-25	425.25	99.0
17527	TSLA	2019-12-28	430.38	99.0
17528	TSLA	2019-12-29	430.38	99.0

7528 rows × 7 columns

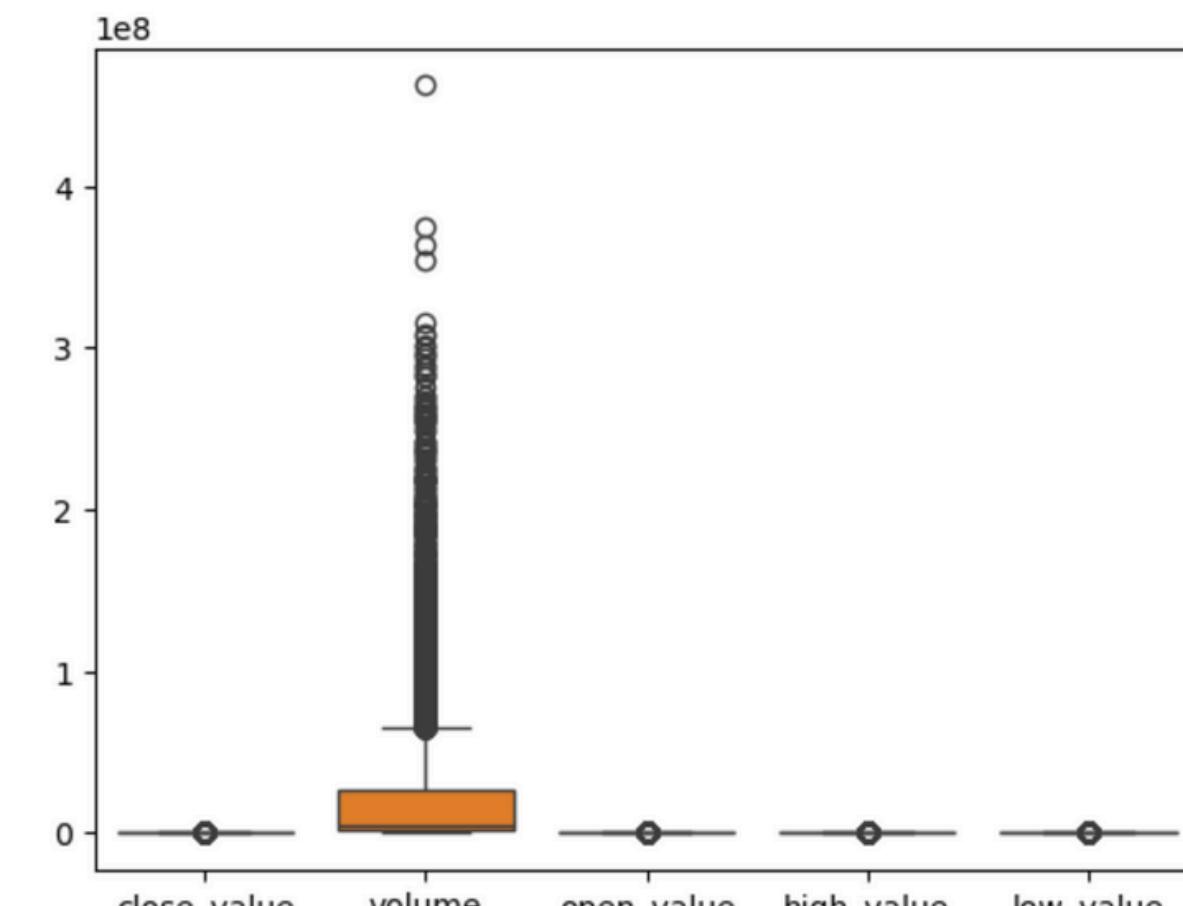
```
Company_data['day'] = Company_data['day_date'].dt.day  
Company_data['month'] = Company_data['day_date'].dt.month  
Company_data['year'] = Company_data['day_date'].dt.year
```

# DATA VISUALIZATION AND IDENTIFYING OUTLIERS

**Data Visualization:** I visualized the Stock Opening Prices Over Time, Stock Closing Prices Over Time and Stock Trading Volume Over Time to enable my predictions perfect, using the Matplotlib and scatterplot

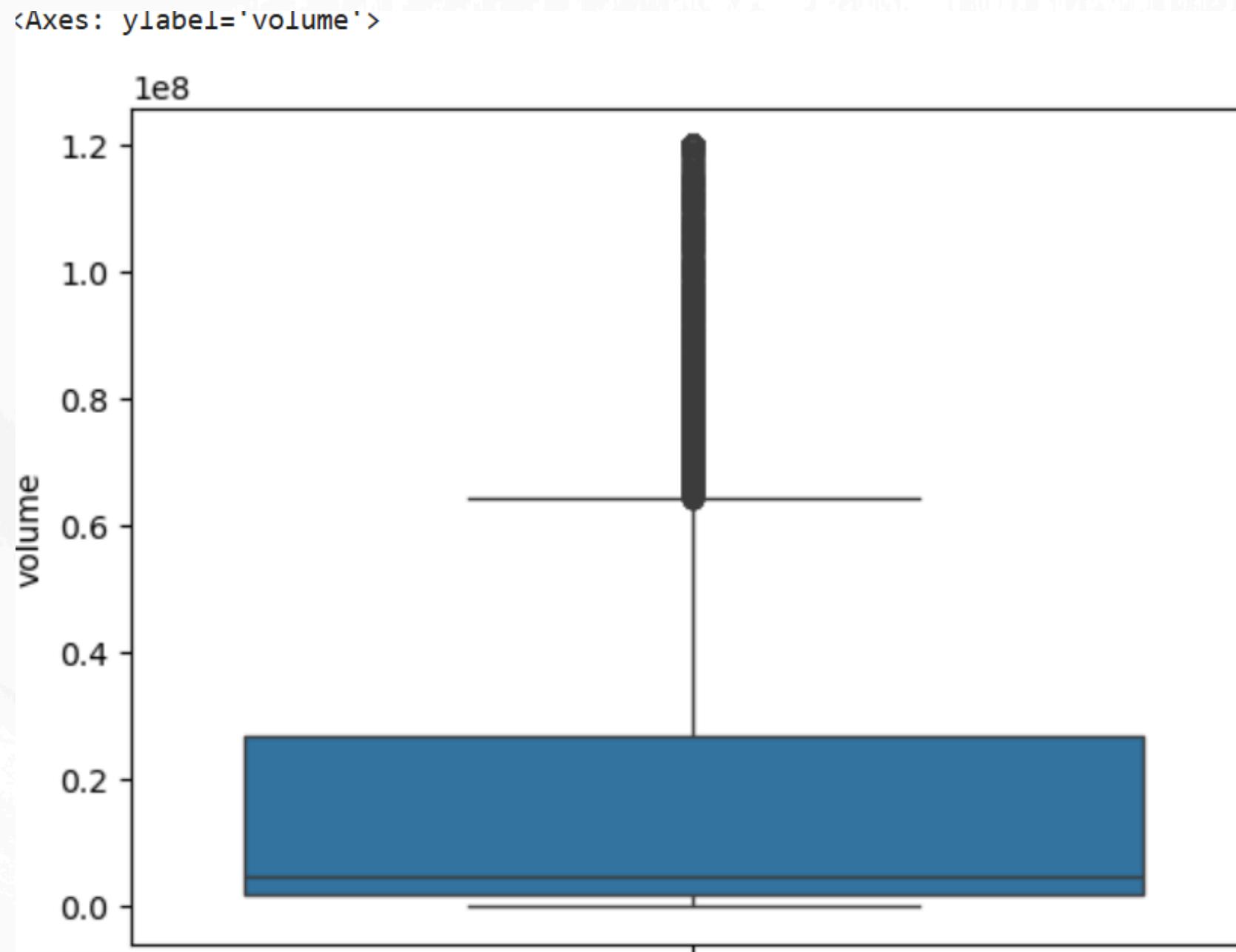


**Identifying Outliers:** I used Seaborn'sboxplot to identify outliers in the dataset  
I discovered 410 rows containing outliers across 10 columns.



# MOVING OUTLIERS

I USED THE CAPPING METHOD WHICH HELPED ME CHANGE THE OUTLIERS VALUES TO UPPER (OR) LOWER LIMIT VALUES, INSTEAD OF DELETING THEM TO AVOID ERRORS IN MY ANALYSIS.



# DATA MODELING AND ENCODING



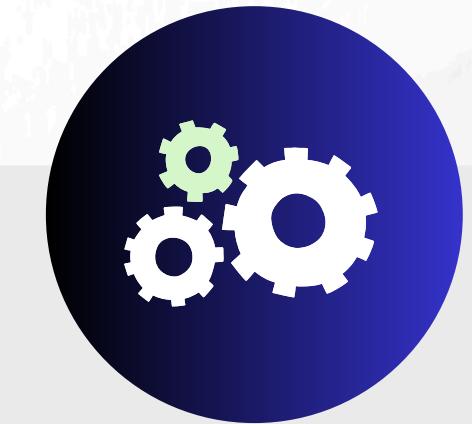
## ENCODING

I encoded my ticker\_symbol and company\_name because they are in object and machine Learning Needs Numeric Values



## FEATURE SELECTION

The target variable selected was volume, while other numerical features were used as predictors



## DATA SPLITTING

The dataset was split into training (70%) and testing (30%) sets

# KEY INSIGHTS FROM THE ANALYSIS

01

Stock  
Market  
Trends

02

Outliers  
Identified

03

Daily  
Returns

04

Feature  
Engineering

# FINAL STEPS

01

Train a regression model to predict stock prices based on historical data

02

I Evaluate model performance using RMSE and R<sup>2</sup> scores.

03

Optimize hyperparameters for better accuracy

04

Established my Model for Prediction using Streamlit

# RECOMMENDATIONS

- **Trend-Based Investment:** Stocks like Tesla, known for high volatility, may be suitable for risk-tolerant investors, while steady-growing stocks like Apple and Microsoft may be safer options.
- **Outlier Management:** Investors should watch for significant spikes in stock prices or trading volume, as these often indicate market-moving events.
- **Risk Management:** Investors should monitor daily return fluctuations to determine potential market corrections or downturns.

# DEPENDENCIES

- PYTHON 3.X
- PANDAS
- NUMPY
- SEABORN
- SCIKIT-LEARN
- STREAMLIT



# CONCLUSION

The analysis of NASDAQ stock data from 2010 to 2020 provided valuable insights into market trends, stock performance, and volatility. Identifying outliers and calculating daily returns helped in understanding market fluctuations and potential investment risks. These findings can guide investors in making data-driven decisions. Future enhancements, such as integrating sentiment analysis and advanced predictive models, could further improve forecasting accuracy and market insights.

# THANK YOU

