## NASDAQ Companies from 2010 to 2020

### **Brief Details About NASDAQ**

NASDAQ (National Association of Securities Dealers Automated Quotations) is one of the largest stock exchanges 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.

Content This dataset contains daily OPEN, CLOSE, VOLUME, HIGH, and LOW values of Amazon, Apple, Google, Microsoft, and Tesla companies as tagged by dates. and we would be making a predictive Analysis leveraging on the previous stock market data.

Acknowledgements Values are fetched from the official Kaggle website.

**Objectives**

 **Analyze stock market trends**

* Identify patterns and insights using historical stock data.

 **Handle data preprocessing efficiently**

* Clean, transform, and manage missing values and outliers in the dataset.

 **Build a predictive model using Logistic Regression**

* Train and evaluate a logistic regression model to classify or predict stock market behavior.

 **Optimize model performance**

* handle feature selection, and improve accuracy.

 **Visualize and interpret results**

* Use Seaborn and Matplotlib to generate meaningful visualizations of stock trends and model performance.
* Adding Extra Columns

 **Document key findings and recommendations**

* Provide actionable insights based on the analysis.

## 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.

This dataset helps identify patterns, volatility, and key market behaviors from mid-2010 to mid-2020. By analyzing stock price changes, volume fluctuations, and historical trends, users can make informed investment decisions, detect market opportunities, and assess potential risks.

## 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.

So I imported my data by reading in the pandas feature

import pandas as pd

Company\_data = pd.read\_csv('CompanyValues.csv')

Company\_name = pd.read\_csv('Company.csv')

# **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, dropping irrelevant values and changing data types if not correct.

In this Project I cleaned my data by;

1. Separating the dates to ensure they remain independent and placing them in different columns.
2. Dropping the date values, company name because they weren’t relevant for my analysis
3. Concatenated the two data set I imported Data = pd.concat([Company\_data, Company\_name],axis=1)
4. Filled in the missing values: i filled it up using the mapping value to enable it correspond

**Data Visualization and Identifying Outliers**

**Identifying Outliers:** I used **Seaborn's** boxplot to identify outliers in the dataset

import seaborn as sns

sns.boxplot(data=Data) After analysis, I discovered **410 rows containing outliers** across **10 columns**.

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 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

**Adding Extra Column**

I calculated the **Daily Return %** to gain insights into the stock market's performance and overall returns. Data['daily\_return'] = Data['close\_value'].pct\_change() \* 100

**Data Modeling And Encoding**

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

Standardization: I applied standardization to Daily\_return using the StandardScaler

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

## Next Steps

* Train a regression model to predict stock prices based on historical data.
* Evaluate model performance using RMSE and R² scores.
* Optimize hyperparameters for better accuracy.
* Established my Model for Prediction using Streamlit

## Dependencies

* Python 3.x
* Pandas
* Numpy
* Seaborn
* Scikit-learn
* Streamlit

## Author

Merenini Princess Better