Project Repost: EDA on Stocks Market

Made by: Sumedh Bhamre(BT22CSD047)

EDA:

The title of this project encapsulates the essence of the analysis, which is a comprehensive exploration of stock market data. Through Exploratory Data Analysis (EDA), this project delves into the intricate world of financial markets, seeking to unveil patterns, trends, and insights hidden within the vast and dynamic data of stock prices, trading volumes, and related metrics. By conducting a meticulous examination of this data, we aim to gain a deeper understanding of the behaviors and factors influencing the performance of various companies and banks, ultimately shedding light on the intricate and interconnected dynamics of the stock market.

Outcomes:

1. Understanding the working of the YahooFinance API:
   Yahoo Finance is a popular and widely-used online platform that provides comprehensive financial information, including stock quotes, news, investment tools, and market insights. It offers users access to real-time and historical stock market data, enabling them to track stock prices, analyze financial trends, and make informed investment decisions. Yahoo Finance is a valuable resource for investors, traders, and anyone interested in staying informed about the world of finance and stock markets.

2. Verifying is the volume of shares of all companies/banks sold across the time frame are correlated or not?
   The pair plot clearly shows that there is not much correlation between the volume of shares sold.
3. Analyzing how the closing price of all the stocks was affected by the fall due to COVID-19 recession?

A clearly noticeable drop is seen from February to April for all stocks. The price first rose rapidly for all stocks due to panic selling and then plummeted to an all time low.

4. Observing how Tech companies got a boost post recession?
   The graphs clearly show tech companies rising rapidly post recession and also overtaking the prices of the banks as they struggled to go back to their original values.
5. Understanding returns and relationship between different industries.
   The graphs show that the tech companies are correlated to each others prices while the banks are correlated to other banks but we see that there isn't any correlation between the market situation of banking sector and tech sector.
6. Understanding Volatility,Moving Averages and other aspects using Technical graphs.
   Various plots like Bollinger plot, Moving Averages plot and Candlestick plots help us visualize the market situation better for a single company over time.
7. Understanding relevant ML models and which to use.
   Models like Linear Regression and ARIMA model are implemented to predict future stock closing prices. We see that the ARIMA model fails here but Linear Regression model performs quite well.

Dataset Used:

The data is directly taken from the Yahoo Finance APIi. This includes data of the following:
Banks:
* Bank of America
* CitiGroup
* Goldman Sachs
* JPMorgan Chase

Tech companies:
* Google
* Apple
* Microsoft
* Nvidia

Technologies/Libraries/ML algorithms used:

1. numpy (np):NumPy is a fundamental library for numerical computing in Python. It provides support for arrays, matrices, and mathematical functions, making it essential for tasks involving numerical data and computations.

2. pandas (pd):Pandas is a powerful data manipulation and analysis library in Python. It offers data structures like DataFrames and Series, making it easy to work with structured data, clean and preprocess data, and perform data analysis.

3. matplotlib.pyplot (plt):Matplotlib is a widely-used data visualization library in Python. Matplotlib.pyplot is a submodule that provides an extensive range of functions and methods to create static, interactive, and publication-quality visualizations, including line plots, bar charts, scatter plots, and more.

4. seaborn (sns):Seaborn is a data visualization library that is built on top of Matplotlib. It specializes in creating attractive and informative statistical graphics, including heatmaps, pair plots, and violin plots. Seaborn simplifies the process of creating complex visualizations with minimal code.

5. plotly.express (px):Plotly is a library for interactive data visualization. Plotly Express is a high-level interface for creating a wide range of interactive charts and plots with ease. It is often used for creating web-based and interactive visualizations.

6. Plotly Express is a high-level data visualization library built on top of Plotly, which is a versatile library for creating interactive and web-based visualizations. Plotly Express is designed to simplify the creation of a wide range of interactive charts and plots. It provides an intuitive and concise API for generating complex visualizations with minimal code.

7. yfinance (yf):yfinance is a Python library that provides an easy way to download historical stock market data from Yahoo Finance. It allows

you to access a wide range of financial data, including historical prices, volumes, and related information for various securities.

8. datetime (dt):The datetime module is a part of Python's standard library and is used for working with dates and times. It provides classes and functions to manipulate, format, and perform calculations with date and time values.

ML algorithms:

9. ARIMA (AutoRegressive Integrated Moving Average): ARIMA is a time series forecasting model that combines autoregressive, differencing, and moving average components to analyze and predict time-dependent data.

10. Linear Regression: Linear Regression is a statistical method used to model and analyze the relationship between a dependent variable and one or more independent variables. It's a versatile tool for prediction and understanding the associations between variables.