**INT 375: DATA SCIENCE TOOLBOX: PYTHON PROGRAMING**

**PROJECT REPORT**

(Project Semester January-April 2025)

***Google Stock Data Analysis & Visualization***

Submitted by

KASHISH KUMARI

Registration No :

**12311829**

B. TECH CSE

SECTION: K23SG (Group 1)

Course Code: INT 217

Under the Guidance of

DR. MANPREET SINGH SEHGAL (32354)

**Discipline of CSE/IT**

**Lovely School of Computer Science & Engineering**

**Lovely Professional University, Phagwara**

**CERTIFICATE**

This is to certify that KASHISH KUMARI bearing Registration no.12311829 has completed

INT 375 project titled, **“**Google Stock Data Analysis & Visualization**”** under my guidance and supervision. To the best of my knowledge, the present work is the result of her original development, effort and study.

DR.MANPREET SINGH SEHGAL (32354)

**Signature and Name of the Supervisor**

**Associate professor**

**School of Computer Science and Engineering**

Lovely Professional University

Phagwara, Punjab.

Date: 12 APRIL, 2025

**DECLARATION**

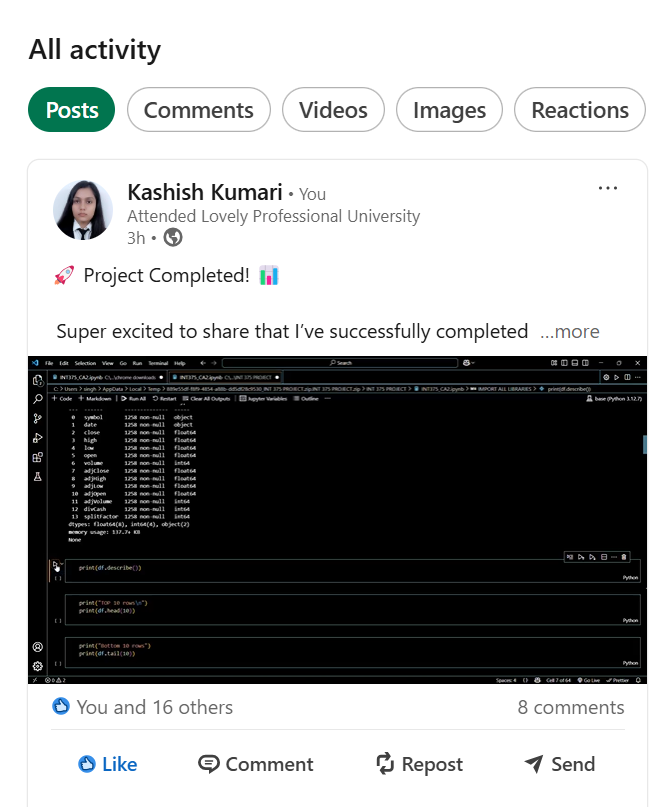
I, KASHISH KUAMRI student of B.TECH CSE under CSE/IT Discipline at, Lovely Professional University, Punjab, hereby declare that all the information furnished in this project report is based on my own intensive work and is genuine.

Date: 12 APRIL, 2025:

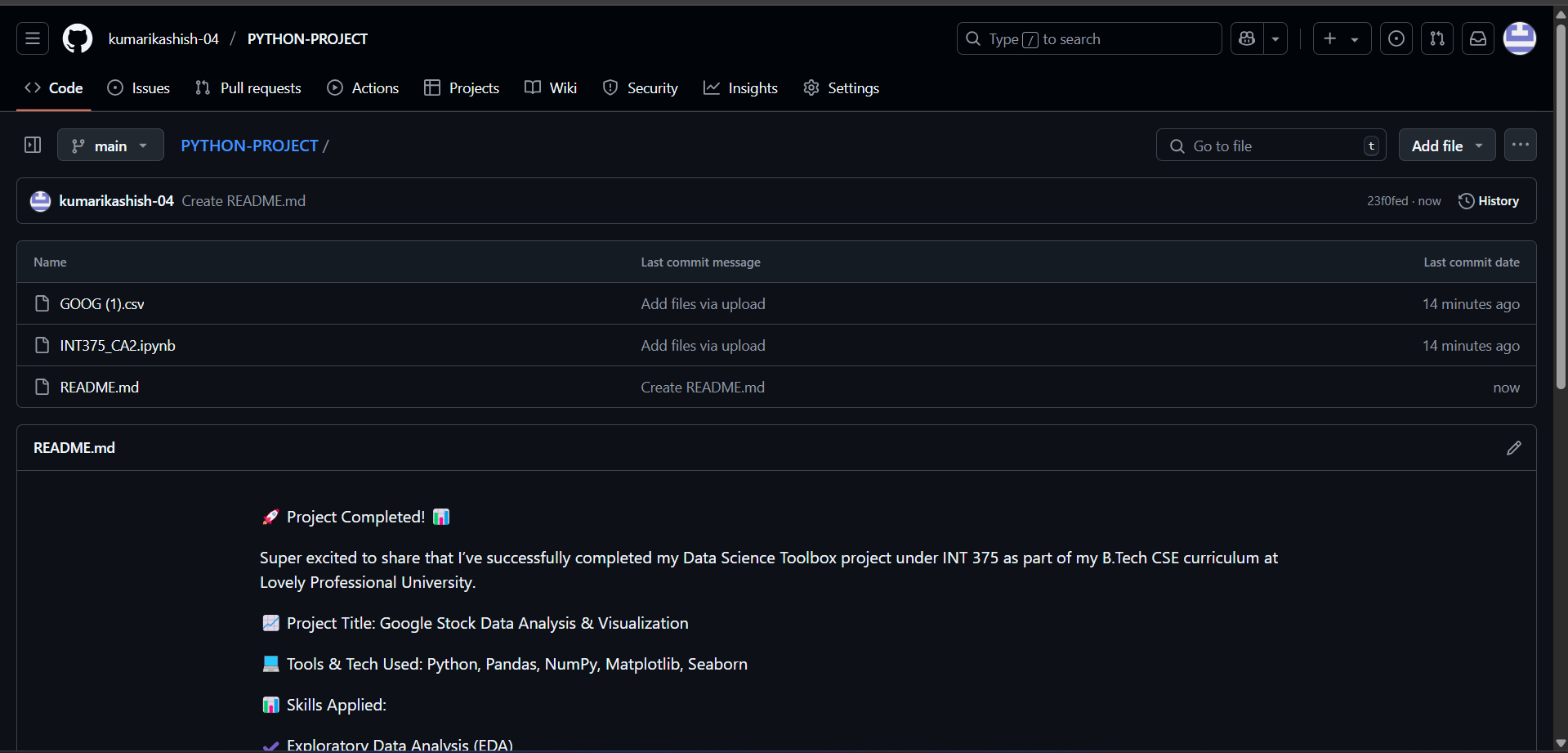
Registration No.12311829

Name of the student : KASHISH KUMARI

LINKEDIN: <https://www.linkedin.com/feed/update/urn:li:activity:7316753318871998464/>



GITHUB:

<https://github.com/kumarikashish-04/PYTHON-PROJECT:>

Acknowledgement

I would like to express my heartfelt gratitude to my respected mentor, Dr. Manpreet Singh Sehgal, for his constant encouragement, support, and expert guidance throughout the duration of this project. His valuable insights and consistent motivation played a vital role in shaping the structure and execution of this report.

I am also extremely thankful to my faculty members for creating a positive learning environment. A special thanks to my friends and classmates for their suggestions and collaborative spirit that kept me going.

Finally, I am deeply grateful to my family for their unwavering love, support, and belief in me. Their presence and encouragement helped me complete this project with dedication and focus.

This project enabled me to enhance my practical knowledge of Python programming, data analysis, and visualization techniques by applying them on real-world financial data.

Table of Content

|  |  |  |
| --- | --- | --- |
| **SNO.** | **TITLE** | **Page No.** |
| 1 | Introduction | 6 |
| 2 | Source of dataset | 7 |
| 3 | EDA Process | 8 |
| 4 | Analysis on dataset(for each analysis) | 9 |
| 5 | Conclusion | 10 |
| 6 | Future scope | 12 |

Introduction

This project focuses on a comprehensive analysis of Google's historical stock market data. The primary objective is to uncover significant trends, patterns, and inter-variable relationships that can assist in understanding the stock's behavior over time. By utilizing the capabilities of popular Python libraries such as pandas, matplotlib, and seaborn, the dataset was explored to derive insights from various angles including opening and closing price patterns, trading volume fluctuations, quarterly performance, and overall volatility.

The dataset spans multiple years and contains key financial indicators like open, high, low, close, adjusted close prices, and volume traded. These metrics are crucial in performing time series analysis and technical evaluation of stock trends. The analysis includes not just raw comparisons, but also derived fields like year and quarter to add depth to the insights.

Through a mix of statistical summaries, grouping, and visualization techniques such as line graphs, bar charts, scatter plots, and donut charts, the project transforms numeric data into visual narratives that are easier to interpret and present. The overall aim of this report is to support informed investment decision-making and strengthen the understanding of real-world financial data through Python programming.

Source of Dataset

Dataset Title: GOOG.csv

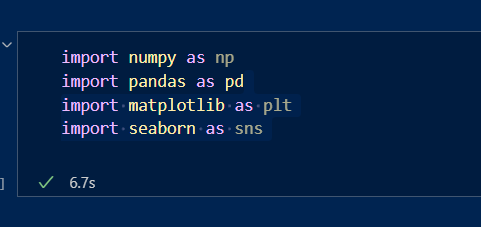
Dataset Source: Kaggle / Yahoo Finance historical stock data

Features include: open, high, low, close, adjClose, volume, adjVolume, date, etc.

Link: <https://www.kaggle.com/code/leilahasan/google-stock-prediction>

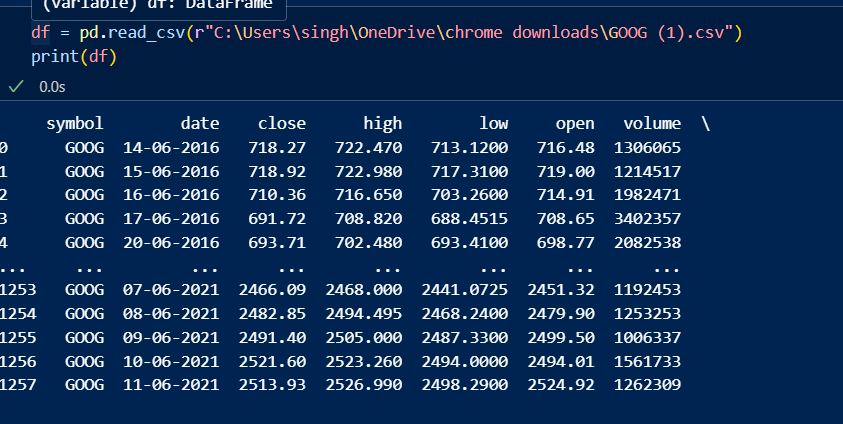
EDA Process

Exploratory Data Analysis (EDA) is a key phase that helps understand the structure, trends, and characteristics of any dataset before performing complex analyses or modeling. For this project, we analyzed the Google historical stock price data using Python libraries like pandas, matplotlib, and seaborn. Below are the main steps followed during EDA:

IMPORT LIBRARIES

* numpy – Numerical Python **(** Handles arrays, mathematical operations.)
* pandas – **Data Handling(**Load, clean, and manipulate tabular data.)
* matplotlib- Basic Plotting(Create basic plots and charts.)
* seaborn- Beautiful Statistical Plots(Make charts prettier and more informative.)

LOAD DATASET

****

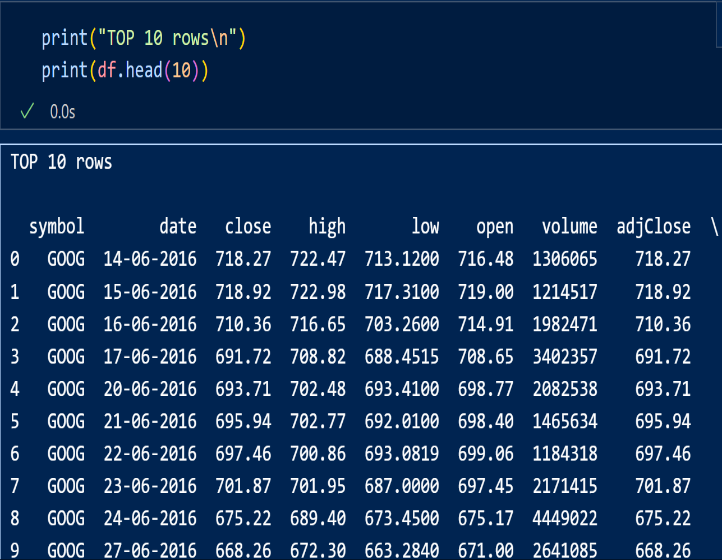
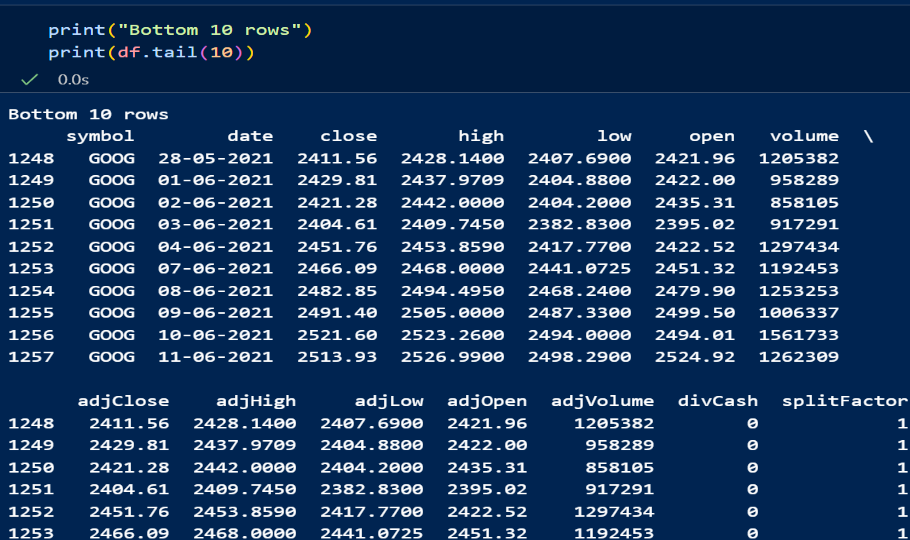
DATA SELECTION AND FILTERING

**Data Loading:**  
Read using pandas, checked data using .info(), .describe(), .head()

* Checked basic info and data types
* Confirmed that columns like open, close, high, low, adjClose, and volume were numerical.
* Verified that date column was of type object (to be converted).
* A screenshot of a computer

  AI-generated content may be incorrect.A screen shot of a computer

  AI-generated content may be incorrect.Viewed basic statistics using (df.describe())



DATA CLEANING:

* **Missing Values:**

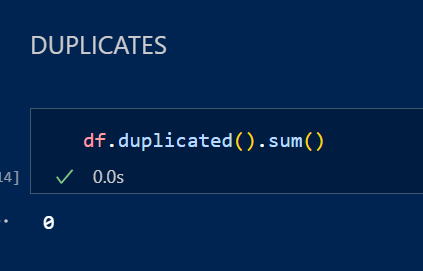
**Detect missing values with df.isnull().sum()**

**Decide how to handle missing age, height, weight, or medal fields (e.g., fill, drop, or impute)**

* **Duplicates:**

**Use df.duplicated().sum() and remove if necessary**

A screenshot of a computer program

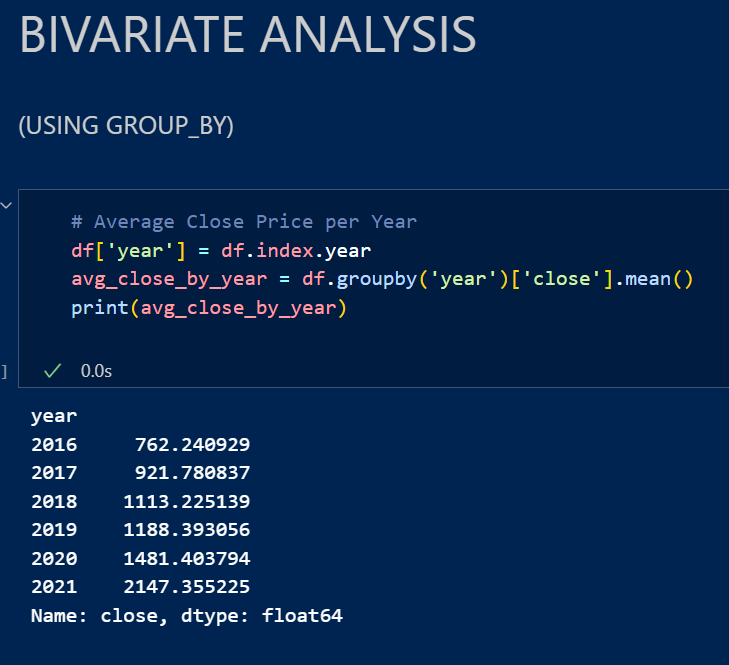
AI-generated content may be incorrect. 

**Bivariate and Multivariate Analysis**

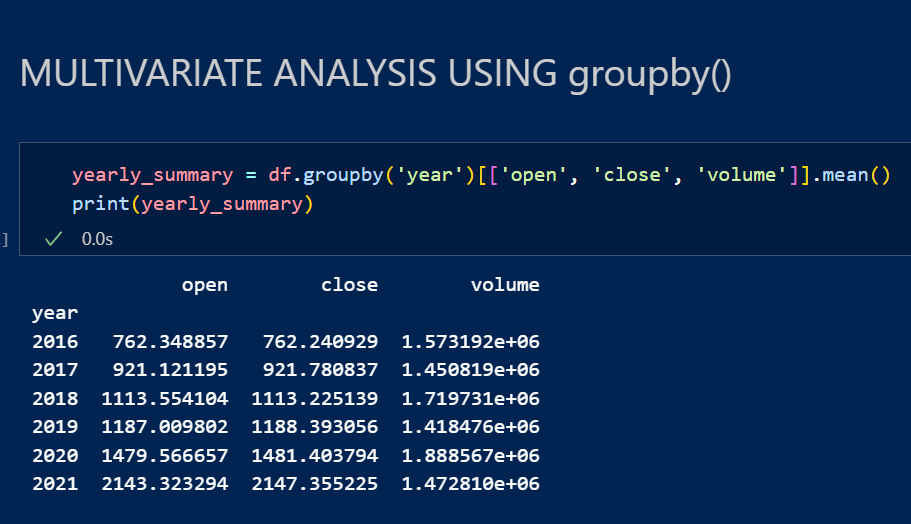
* Analyze relationships between variables:
* Average Close Price by Year
* Total Volume Traded by Year
* . Max & Min Adjusted Close Price by Year
* Average Open, Close, and Volume by Year
* Quarterly Average Adjusted Close
* Use groupby(), pivot\_table() to find correlations

A screenshot of a computer

AI-generated content may be incorrect.



A screen shot of a computer

AI-generated content may be incorrect.

Visualization of Key Insights

Create visual dashboards using:

* matplotlib and seaborn for static charts
* plotly or dash for interactive visuals
* Maps using plotly.express to show medals by country.

**Analysis on Dataset (for each objective)**

Each analysis includes:

1. Introduction
2. General Description
3. Analysis Results
4. Visualization

**1.Trend of Closing Prices Over Time**

**i. Introduction:**  
The objective is to observe how Google’s stock has performed over the available time period using the close price column.

**ii. General Description:**  
The close column represents the final trading price of the stock for each day. By plotting this over time, we can identify overall upward or downward trends.

**iii. Analysis Results:**  
From the line graph, it’s clear that Google’s stock shows a generally upward trend over the years, with some fluctuations during market corrections and global events.



2. Year-wise Total Volume Traded

**i. Introduction:**

This analysis aims to find which years had the highest trading activity.

**ii. General Description:**

The volume column indicates the number of shares traded. Summing it year-wise reveals investor activity levels.

**iii. Analysis Results:**

Years with higher trading volumes may align with market events, stock splits, or major news affecting the company.

A graph of blue bars

AI-generated content may be incorrect.

Open vs Close Price Relationship

i. Introduction:

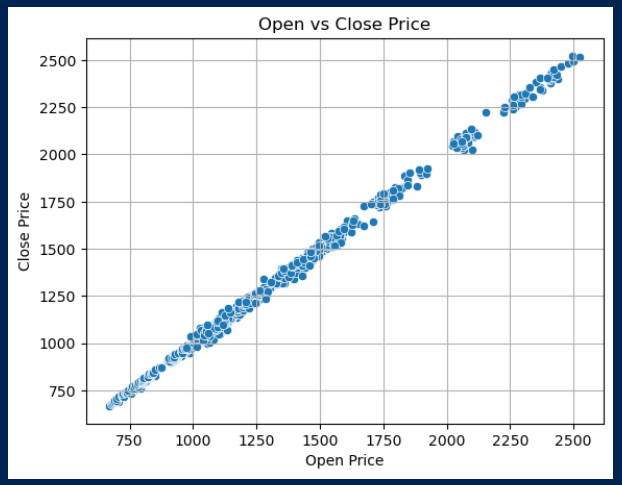
This objective is to examine how closely the opening and closing prices relate on a daily basis.

ii. General Description:

Stocks often open at a different price than they close. A scatter plot helps visualize the correlation between these two prices.

iii. Analysis Results:

There is a strong linear trend, showing that most days have small variations between open and close prices.



3.Volatility via High vs Low Prices

**i. Introduction:**

The goal here is to measure the daily volatility using the high and low prices.

**ii. General Description:**

The difference between high and low shows how much the stock fluctuated within a single day.

**iii. Analysis Results:**

On some days, the gap between high and low prices is significantly larger, indicating volatile market conditions.

A screenshot of a graph

AI-generated content may be incorrect.

Quarterly Adjusted Close Price Trend

**i. Introduction:**

This analysis focuses on identifying quarterly average adjusted closing prices.

**ii. General Description:**

The adjClose column reflects stock value after accounting for splits/dividends. Grouping by quarter gives insight into seasonal performance.

**iii. Analysis Results:**

We can observe which quarters typically yield higher adjusted values and identify any cyclic behavior.

A graph showing the growth of a company

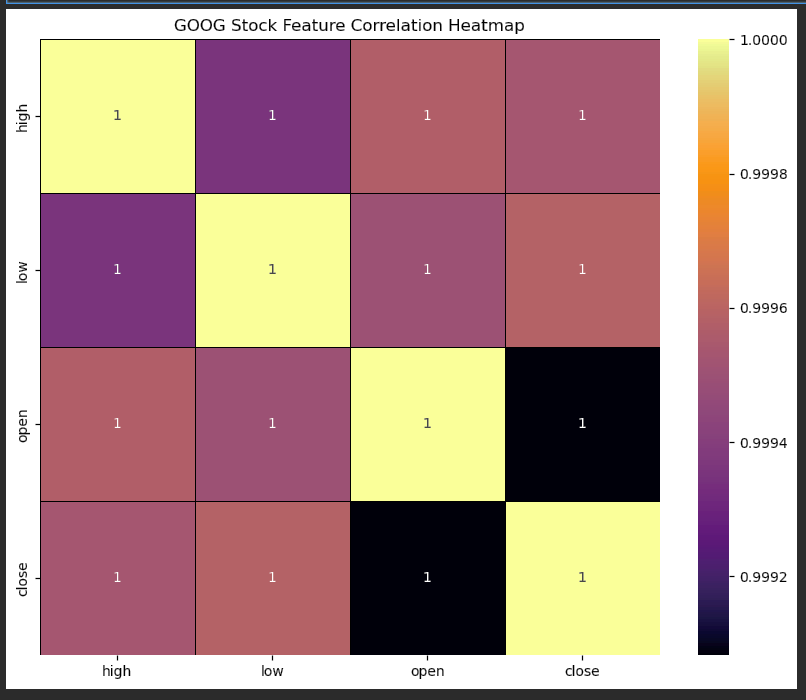
AI-generated content may be incorrect.

A screenshot of a chart

AI-generated content may be incorrect.

The relationship between different stock price features

To visually analyze the correlation between different stock price features (high, low, open, close) of Google (GOOG) and identify how strongly they are related to each other.



**Conclusion**

The analysis of Google's historical stock data using Python programming and data visualization tools provided several valuable insights into the behavior of the stock over time. By performing exploratory data analysis, missing value treatment, and grouping through groupby() and pivot\_table(), we uncovered meaningful trends that would aid investors and analysts.

The stock exhibited a consistent long-term upward movement, with periods of high volatility captured through high vs low comparisons. Close prices were found to be strongly correlated with open prices, confirming predictable intraday price behavior. Additionally, volume-based analysis helped identify active trading periods, which often coincided with major events or market fluctuations. Seasonal (quarterly) analysis of adjusted close prices also demonstrated performance cycles that could support portfolio planning.

Overall, this project not only showcased the power of Python in financial data analysis but also emphasized how raw data can be transformed into impactful business insights through proper statistical and visual techniques.

**Future Scope**

This project provides a solid foundation for several potential extensions that can greatly enhance its depth, applicability, and usability in real-world financial analysis. The following are some well-defined and impactful directions for future work:

* **Integration of Technical Indicators:** Incorporate technical tools such as Moving Averages (MA), Exponential Moving Averages (EMA), Bollinger Bands, Relative Strength Index (RSI), and MACD. These are crucial in financial analytics and would provide traders better insight into price momentum, overbought/oversold conditions, and breakout signals.
* **Stock Price Forecasting:** Employ time-series forecasting techniques such as ARIMA, Prophet, or deep learning models like LSTM (Long Short-Term Memory networks) to predict future stock prices. This adds a predictive dimension to the current analysis, transforming it into a decision-support tool.
* **Sentiment-Driven Insights:** Merge quantitative price data with qualitative sentiment from financial news articles, Reddit posts, or tweets. Using Natural Language Processing (NLP), this approach could explain sudden price movements driven by public opinion.
* **Cross-Sector Comparison:** Extend the project by comparing Google's performance with other major players in the tech industry such as Apple, Amazon, Meta, or Microsoft. This benchmarking approach would help in identifying sector-specific trends or deviations.
* **Development of Interactive Dashboards:** Transform this static project into a dynamic and interactive dashboard using frameworks like Plotly Dash or Streamlit. This would allow users to explore charts, filter time ranges, and compare multiple metrics in real time.
* **Automated Data Pipelines:** Create scripts that fetch fresh daily stock data using APIs such as Alpha Vantage or Yahoo Finance so that the analysis stays current and can adapt to changing market scenarios.

Implementing these enhancements would not only add greater value to this project academically but also turn it into a practical tool for use in financial research, trading strategy development, or business intelligence.

**References**

1. Yahoo Finance - Google Historical Stock Data (<https://finance.yahoo.com/quote/GOOG/history>)
2. Kaggle Datasets - Google Stock Prices
3. Python Pandas Documentation - <https://pandas.pydata.org/docs/>
4. Matplotlib Documentation - <https://matplotlib.org>
5. Seaborn Documentation - <https://seaborn.pydata.org>
6. NumPy Documentation - <https://numpy.org/doc/>