**Phase-2**

**Student Name:** E.Ajay

**Register Number:** 732123104002

**Institution:** Nandha College of Technology

**Department:** BE. Computer Science And Engineering

**Date of Submission:** 01/05/2025

**Github Repository Link:**https://github.com/Balasakthi-31/Naan-Mudhalvan-phase-2

### **Problem Statement**

*Predict the future prices of stocks using historical data and artificial intelligence. This project focuses on analyzing historical stock price data of Visa Inc. to uncover meaningful patterns and trends that could support making informed investment decisions*.



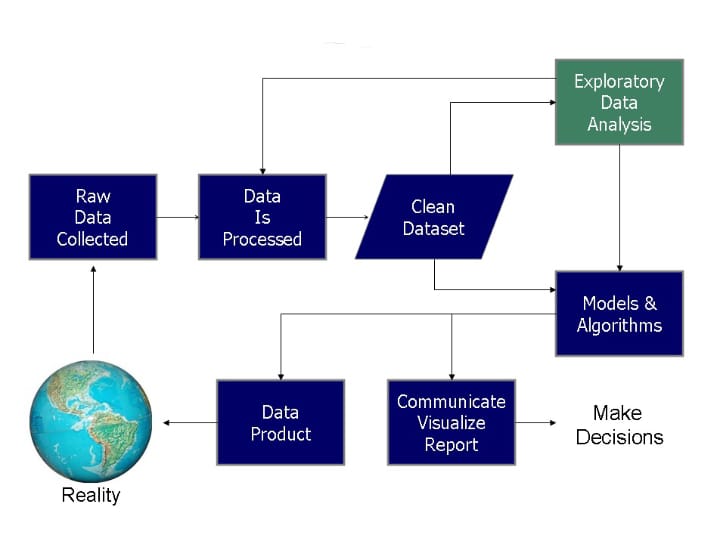
**2.Project Objectives**

**.** *Calculate and visualize short-term and long-term moving averages.*

**.** *Understand correlations between different numerical features.*

**.** *Prepare the data for potential future forecasting or modeling*

**3.Flowchart of the Project Workflow**

****

### **4. Data Description**

* *Dataset name:* *visa\_stock\_data*
* *Source:Kaggle*
* *Type of data:* *Structured, Time-Series*
* *Number of records: 996 rows*

* *Number of Features: 8 columns*
* *Dataset :* *Dynamic*
* *Target variable:* *Close, Adj Close*
* *Learning Type:Supervised Learning*

### **5. Data Preprocessing**

* *Missing values :* *No missing values found*
* *Duplicate records : No duplicate entries in the dataset*
* *Detect and treat outliers:* *Used IQR method to detect.Outliers retained due to financial data's real-world nature.*
* *Datatype Convertion :* *Date converted from string to datetime format.*
* *Categorical Encoding :* *Not applicable as dataset contains only numerical and temporal features.*
* *Normalization :* *StandardScaler applied to numerical columns to normalize them for modeling.*

### **6. Exploratory Data Analysis (EDA)**

* *Univariate Analysis:*
  + *Histogram and boxplot visualizations for Open, Close, and Volume*
* *Bivariate/Multivariate Analysis:*
  + *Correlation matrix shows high correlation between Open, High, Low, and Close*
  + *Scatter plots highlight linear trends.*
* *Insights Summary:*
  + *Strong positive correlation between Open and Close*
  + *Volume does not strongly influence price movement*

### **7. Feature Engineering**

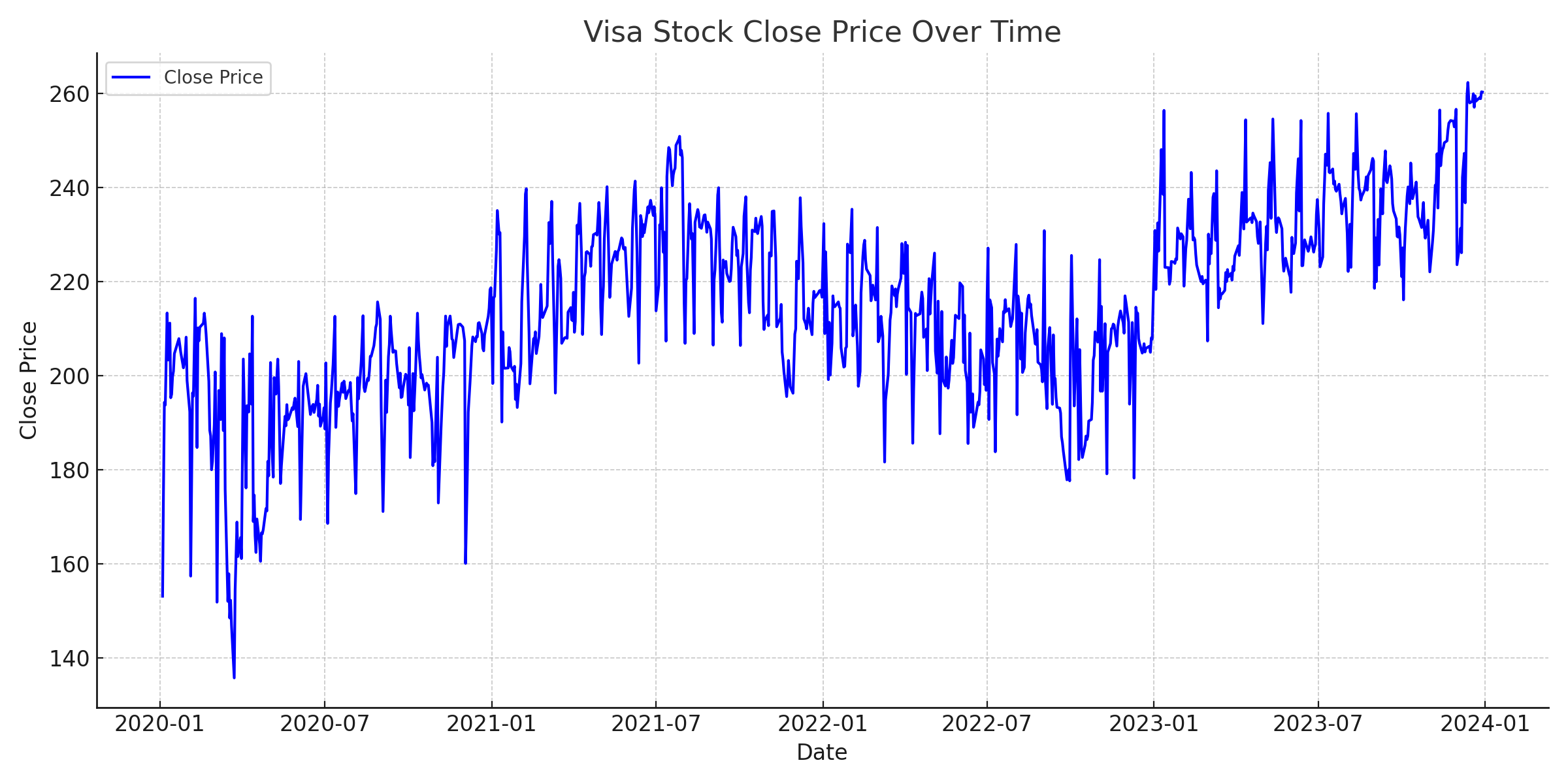
* *Created Price\_Up as a classification target for future prediction tasks.*
* *Extracted date parts from Date: Year, Month, Day for potential use.*
* *Standardized Open, High, Low, Close, Adj Close, Volume.*

### **8. Model Building**

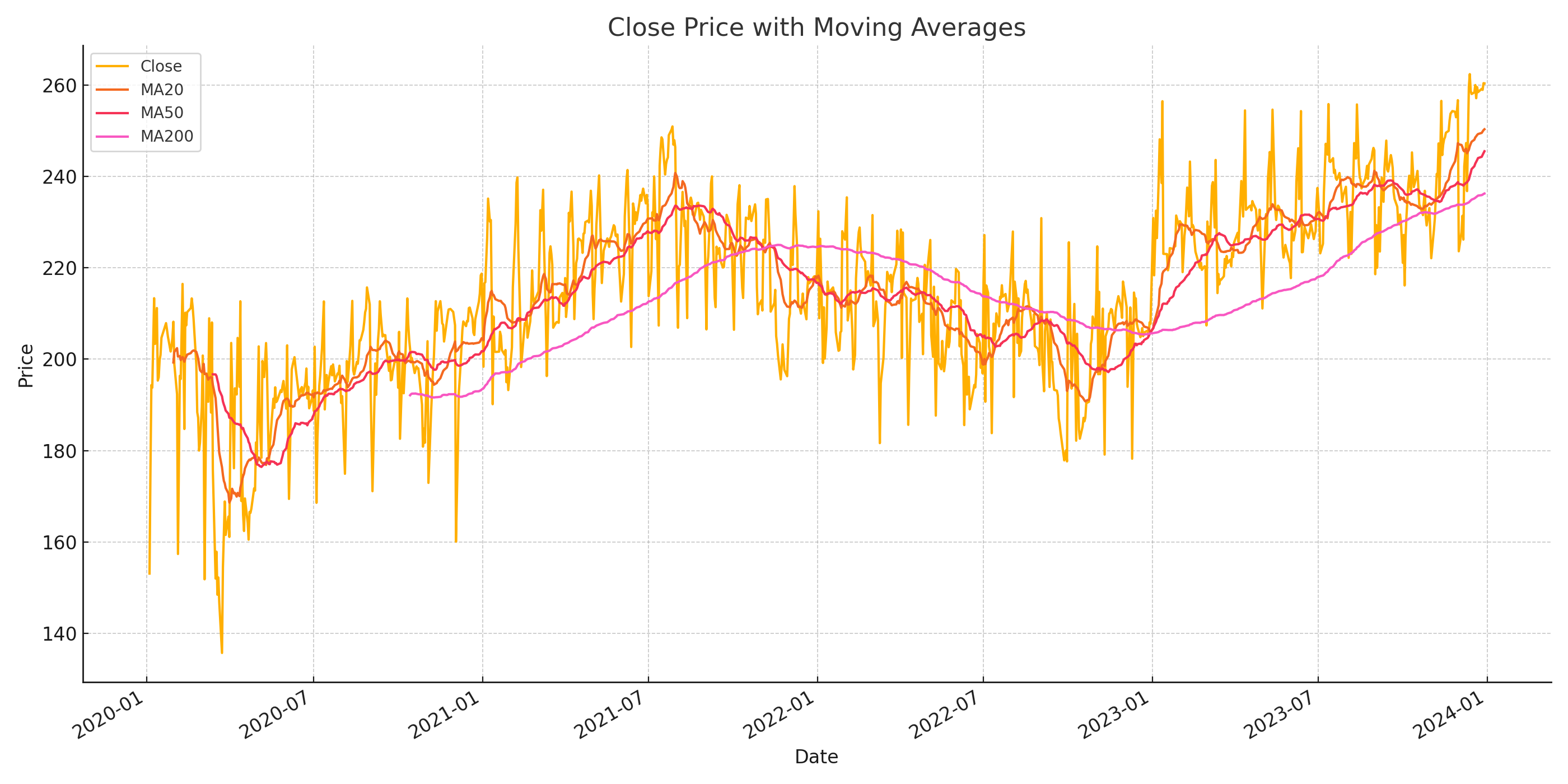
* *Models used:*
  + *Linear Regression and Decision Tree .*
* *Train-Test Split:80%-20%*
* *Evaluation Metrics: MAE, RMSE, R² Score*
* *why these models:*
  + *Linear Regression provides a baseline with interpretability*
  + *Decision Tree handles non-linear trends and is robust to outliers.*

### **9. Visualization of Results & Model Insights**

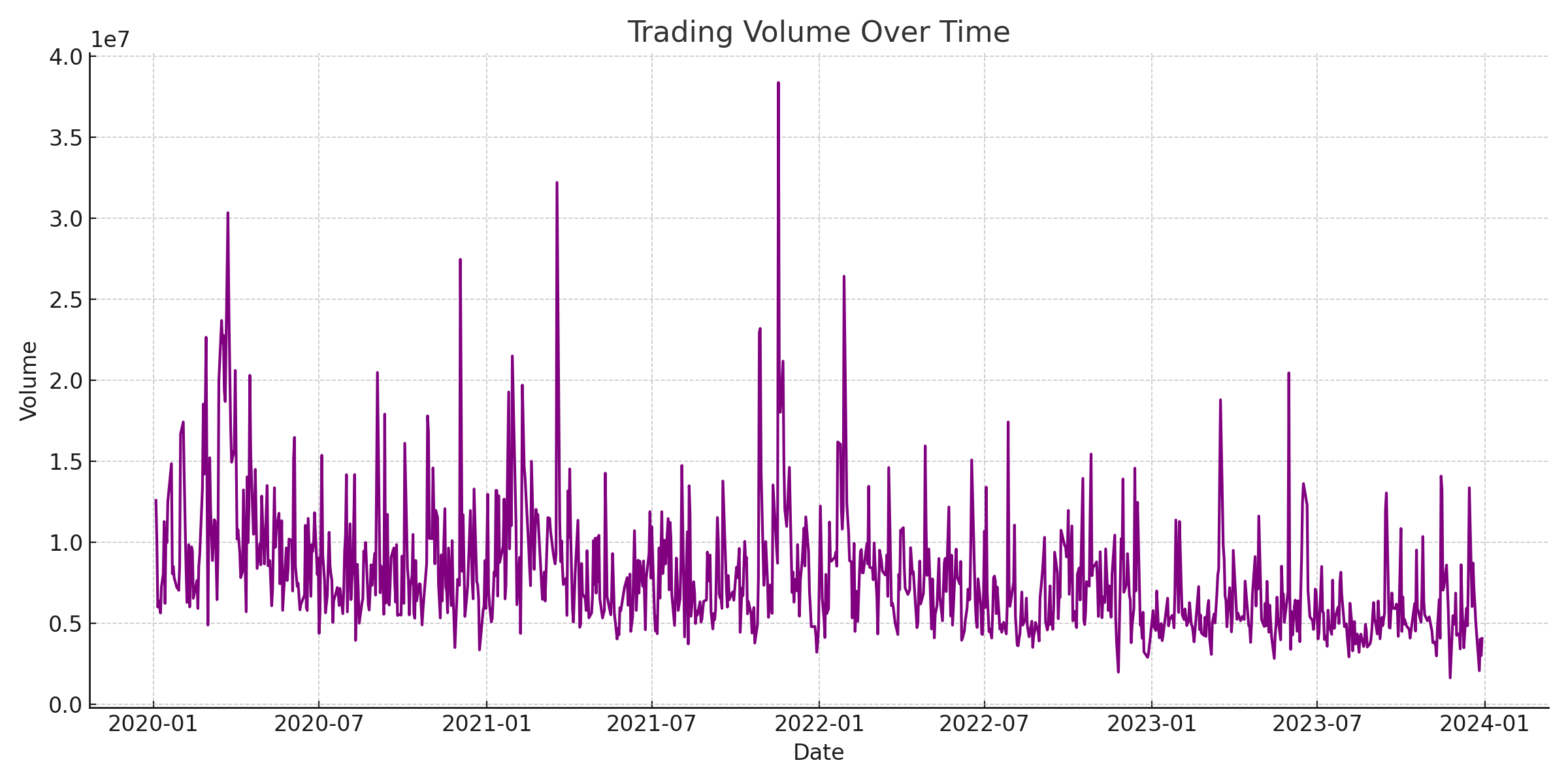
*Visa Stock Close Price Over Time*



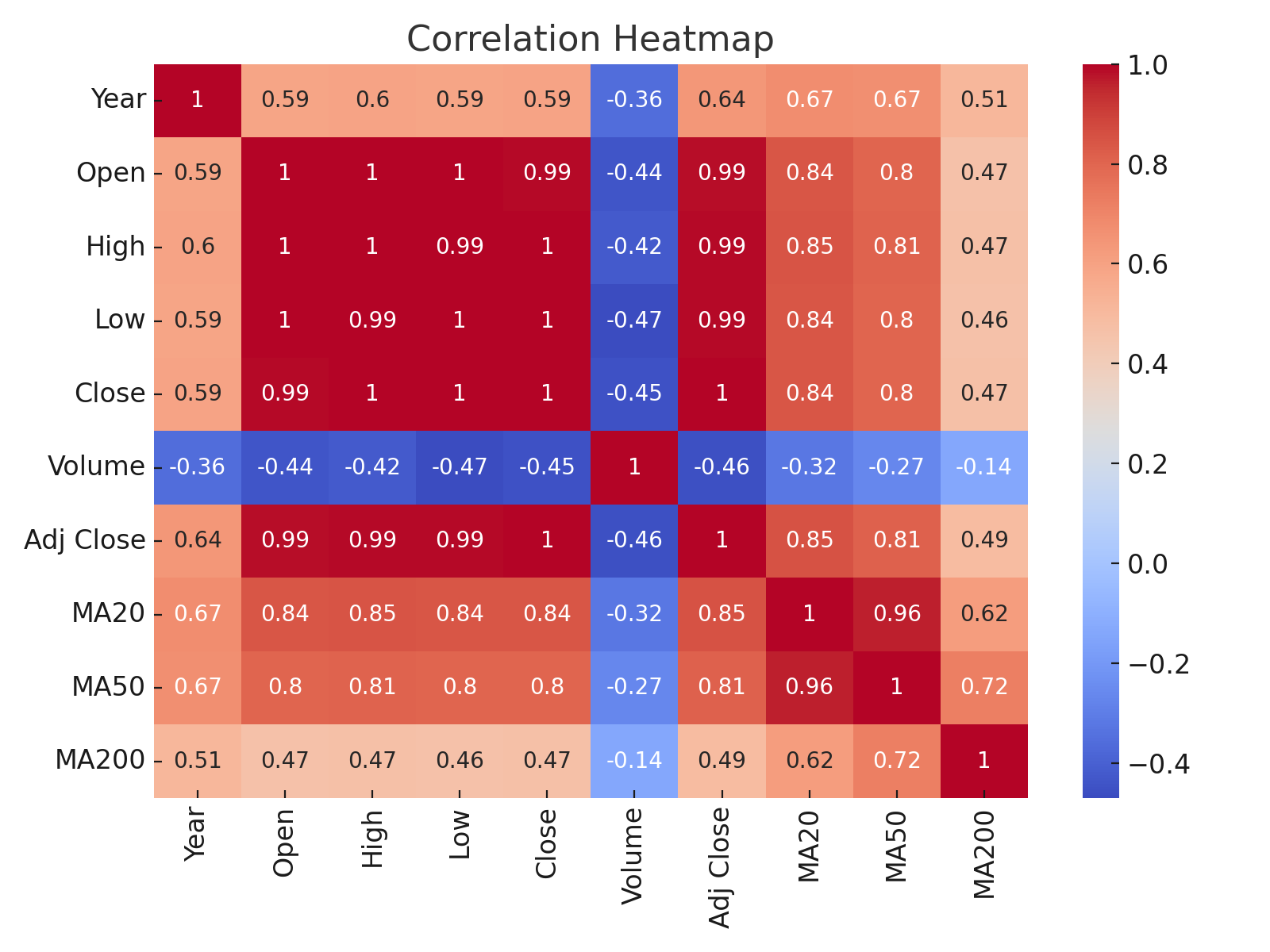
**.** *Short- and long-term Moving averages*



* *Trading Volume Over Time*



* *Correlation Heatmap*



### **10. Tools and Technologies Used**

* *Language: Python.*
* *Development Environment: Google Colab*
* *Libraries: pandas, numpy, matplotlib and seaborn.*
* *Visualization Tools: Matplotlib and seaborn*

### **11. Team Members and Contributions**

* *E.Ajay : Prepares the dataset and helps with data cleaning.*
* *R.Deepak : Explores the data analysis (EDA) and feature engineer.*
* *R. Balasakthi : Prepares the documentation & reporting and model development.*