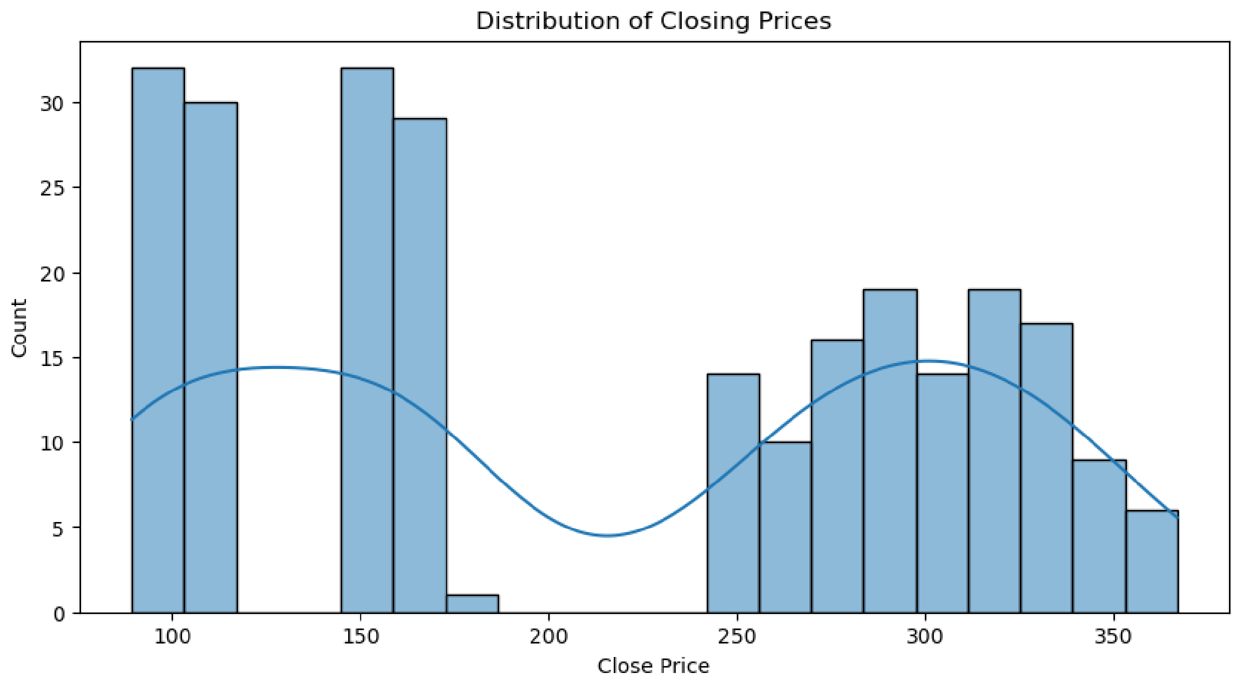
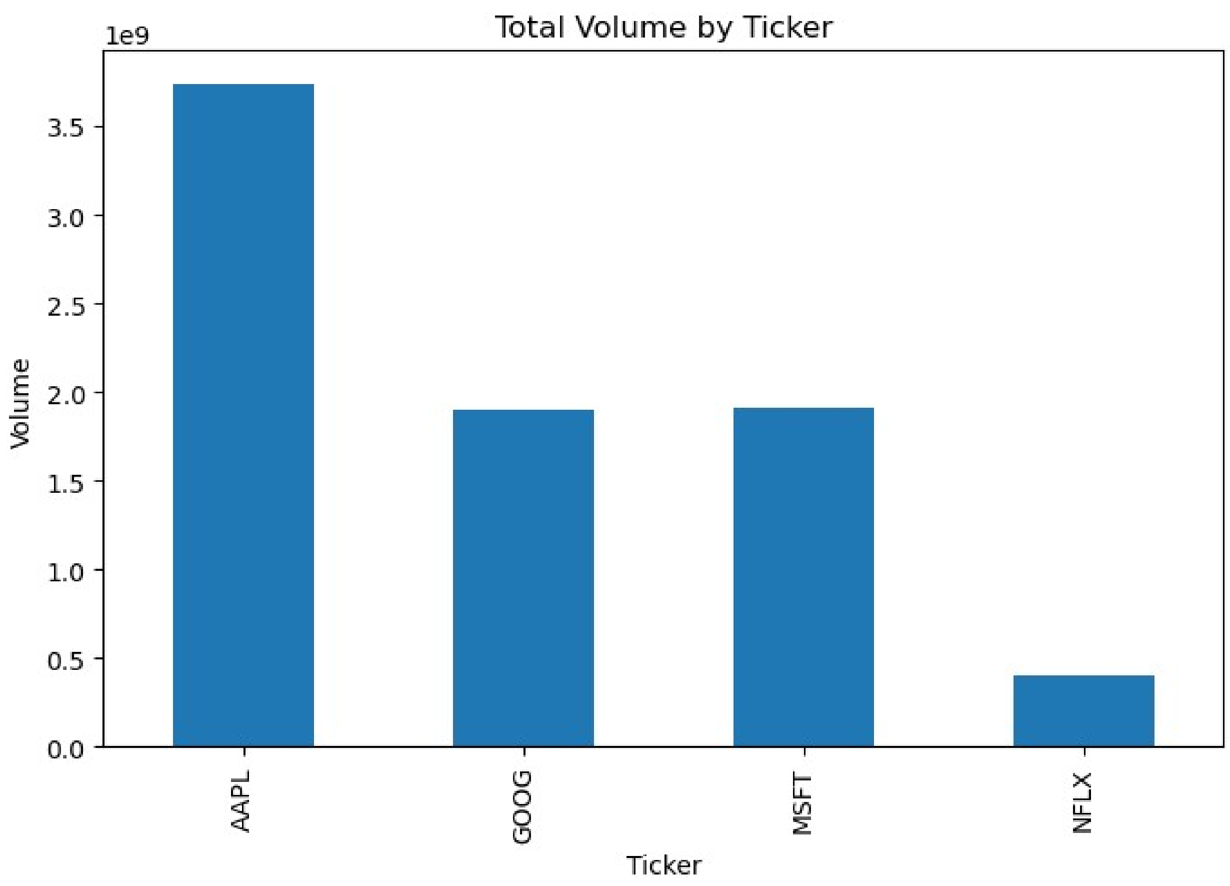
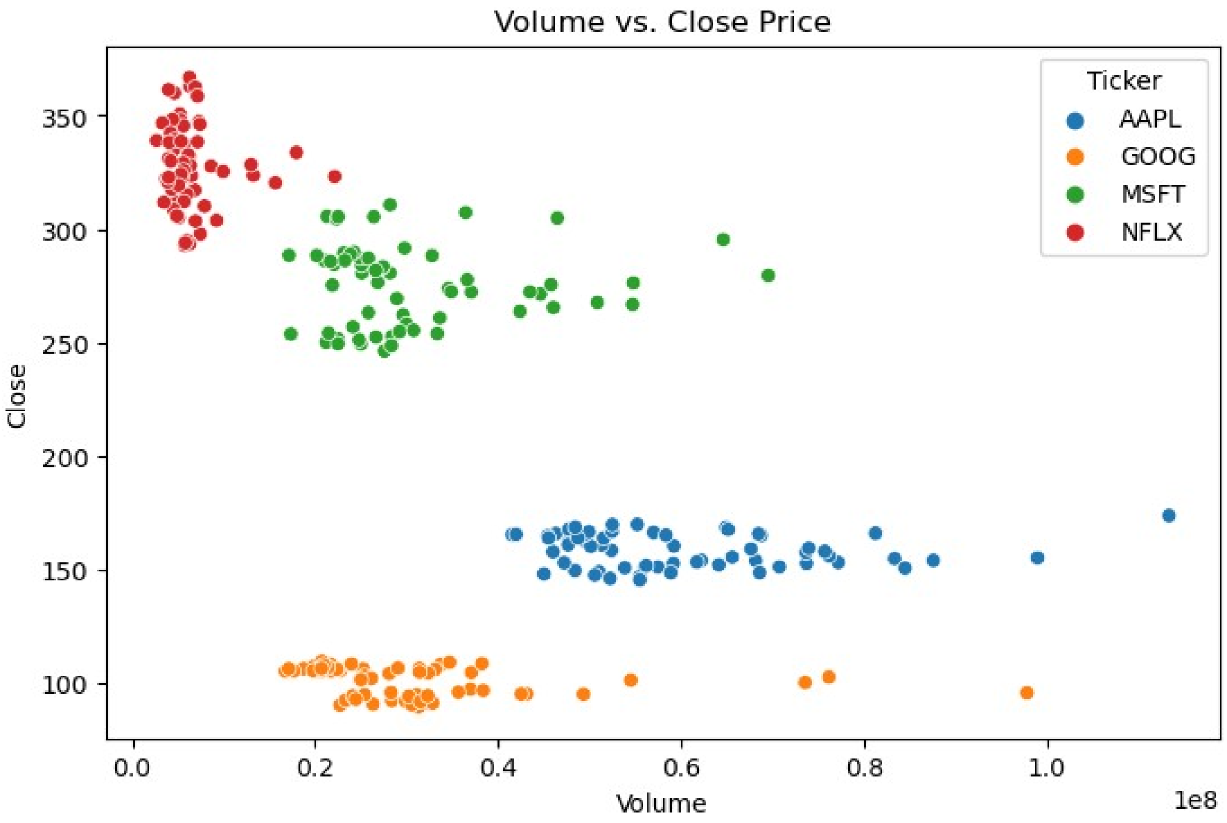
```python
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.linear_model import LinearRegression
from sklearn.model_selection import train_test_split
from sklearn.metrics import mean_squared_error, r2_score

# Load Data
file_path = r"C:\Users\pc\Downloads\stocks.csv"
data = pd.read_csv(file_path)

# Clean Data
data.columns = [col.strip().replace(' ', '_') for col in data.columns]
data['Date'] = pd.to_datetime(data['Date'])
data.sort_values(by=['Ticker', 'Date'], inplace=True)
assert data.isnull().sum().sum() == 0, "Missing values found in data"

# Summary
print("Data Overview:")
print(data.head())
print("\nTickers:", data['Ticker'].unique())

# Basic EDA
plt.figure(figsize=(10, 5))
sns.histplot(data['Close'], bins=20, kde=True)
plt.title('Distribution of Closing Prices')
plt.xlabel('Close Price')
plt.show()

# Total Volume by Ticker
ticker_volume = data.groupby('Ticker')['Volume'].sum()
ticker_volume.plot(kind='bar', title='Total Volume by Ticker',
figsize=(8, 5))
plt.ylabel('Volume')
plt.show()

# Volume vs Closing Price
plt.figure(figsize=(8, 5))
sns.scatterplot(data=data, x='Volume', y='Close', hue='Ticker')
plt.title('Volume vs. Close Price')
plt.show()

# Boxplot of Close Prices
plt.figure(figsize=(8, 5))
sns.boxplot(data=data, x='Ticker', y='Close')
plt.title('Closing Price Distribution by Ticker')
plt.show()

# Moving Averages & Volatility
```

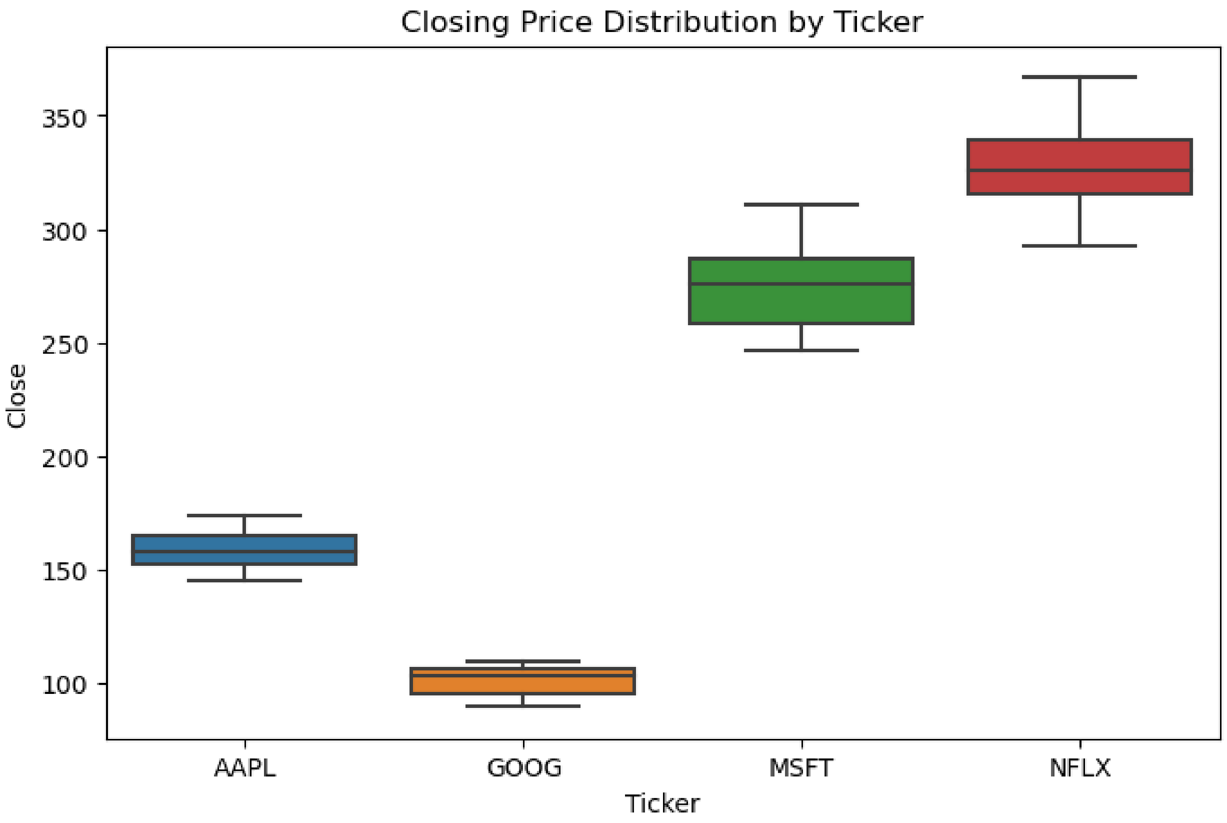
```
for ticker in data['Ticker'].unique():
    company_data = data[data['Ticker'] == ticker].copy()
    company_data.set_index('Date', inplace=True)
    company_data['MA10'] =
company_data['Close'].rolling(window=10).mean()
    company_data['Volatility'] =
company_data['Close'].rolling(window=10).std()

    plt.figure(figsize=(10, 5))
    plt.plot(company_data['Close'], label='Close Price')
    plt.plot(company_data['MA10'], label='10-Day MA')
    plt.title(f'{ticker} - Close Price & Moving Average')
    plt.legend()
    plt.show()

# Correlation Matrix
corr = data[['Open', 'High', 'Low', 'Close', 'Adj_Close',
'Volume']].corr()
plt.figure(figsize=(8, 6))
sns.heatmap(corr, annot=True, cmap='coolwarm')
plt.title('Correlation Matrix')
plt.show()

# Feature Engineering
data['Day'] = data['Date'].dt.day
data['Month'] = data['Date'].dt.month
data['Year'] = data['Date'].dt.year

# Predicting Close Price using Linear Regression
model_data = data.copy()
model_data = pd.get_dummies(model_data, columns=['Ticker'],
drop_first=True)
features = ['Open', 'High', 'Low', 'Volume', 'Day', 'Month', 'Year'] +
[col for col in model_data.columns if col.startswith('Ticker_')]
X = model_data[features]
y = model_data['Close']

X_train, X_test, y_train, y_test = train_test_split(X, y,
test_size=0.2, random_state=42)

lr = LinearRegression()
lr.fit(X_train, y_train)
y_pred = lr.predict(X_test)

print("\nModel Evaluation:")
print("RMSE:", np.sqrt(mean_squared_error(y_test, y_pred)))
print("R2 Score:", r2_score(y_test, y_pred))

# Actual vs Predicted
plt.figure(figsize=(8, 5))
plt.scatter(y_test, y_pred, alpha=0.5)
plt.xlabel("Actual Close Price")
plt.ylabel("Predicted Close Price")
plt.title("Actual vs Predicted Close Prices")
plt.show()

Data Overview:
  Ticker        Date        Open        High         Low       Close  \
0   AAPL 2023-02-07  150.639999  155.229996  150.639999  154.649994
1   AAPL 2023-02-08  153.880005  154.580002  151.169998  151.919998
2   AAPL 2023-02-09  153.779999  154.330002  150.419998  150.869995
3   AAPL 2023-02-10  149.460007  151.339996  149.220001  151.009995
4   AAPL 2023-02-13  150.949997  154.259995  150.919998  153.850006

    Adj_Close    Volume
0  154.414230  83322600
1  151.688400  64120100
2  150.639999  56007100
3  151.009995  57450700
4  153.850006  62199000

Tickers: ['AAPL' 'GOOG' 'MSFT' 'NFLX']
```



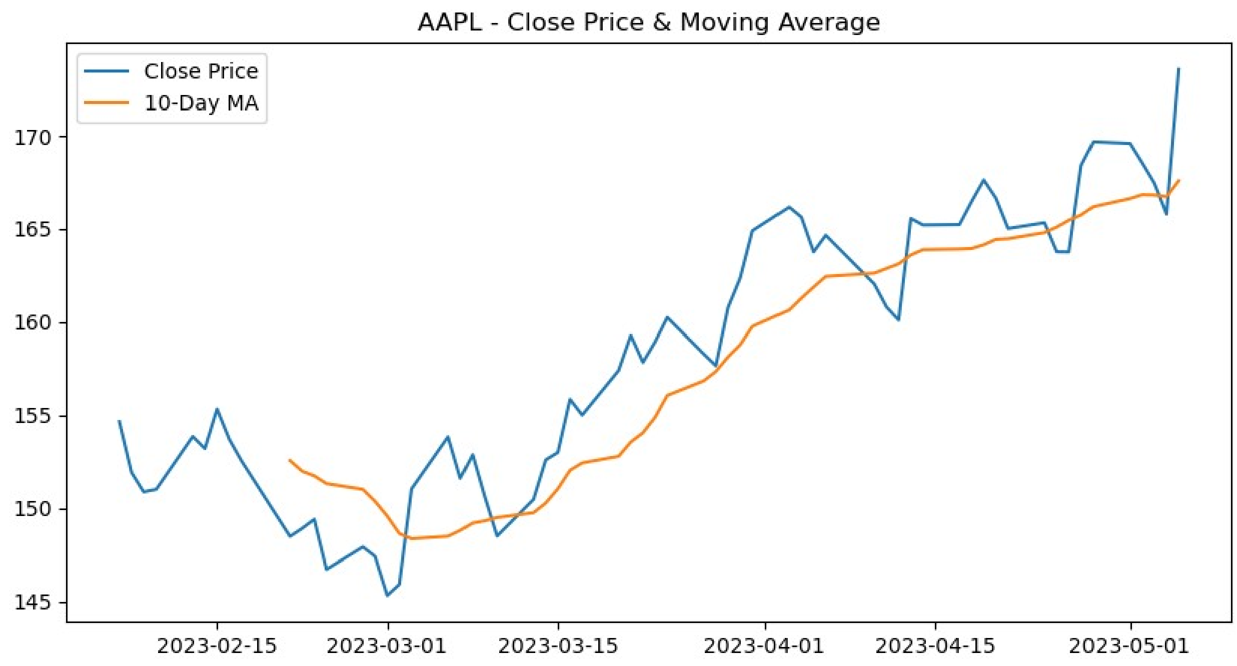

Total Volume by Ticker
1e9
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
Volume
AAPL
GOOG
MSFT
NFLX
Ticker

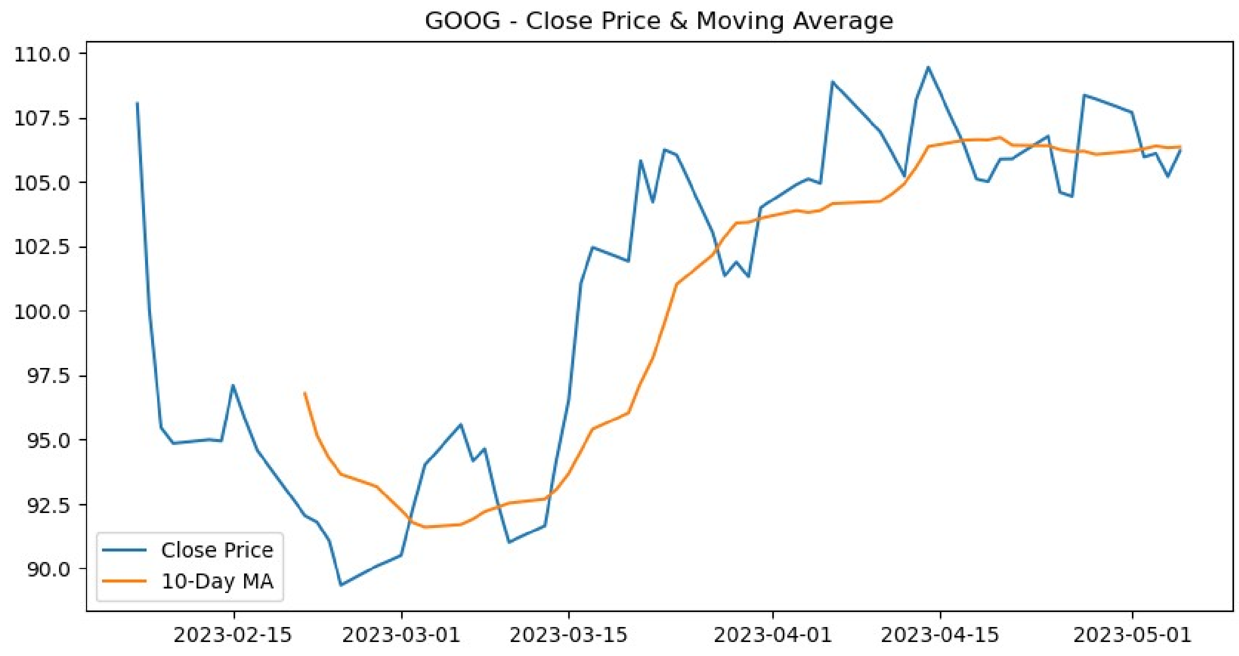


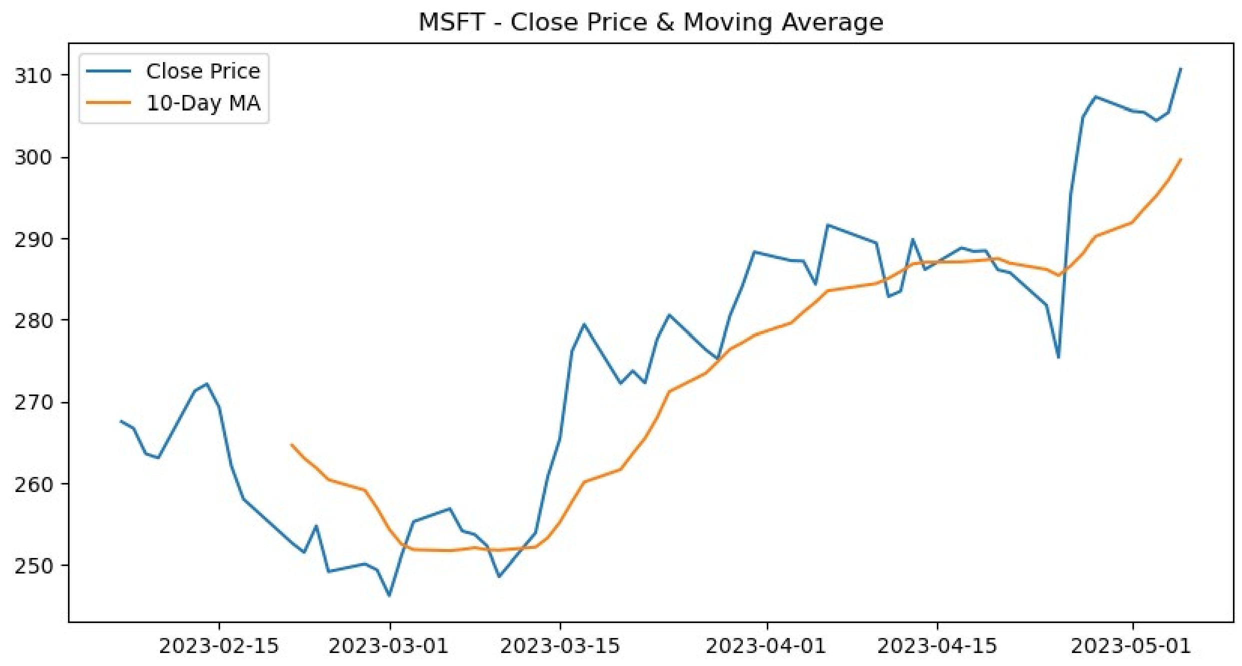
Volume vs. Close Price
Ticker
AAPL
GOOG
MSFT
NFLX
Close
350
300
250
200
150
100
0.0
0.2
0.4
0.6
0.8
1.0
Volume
1e8

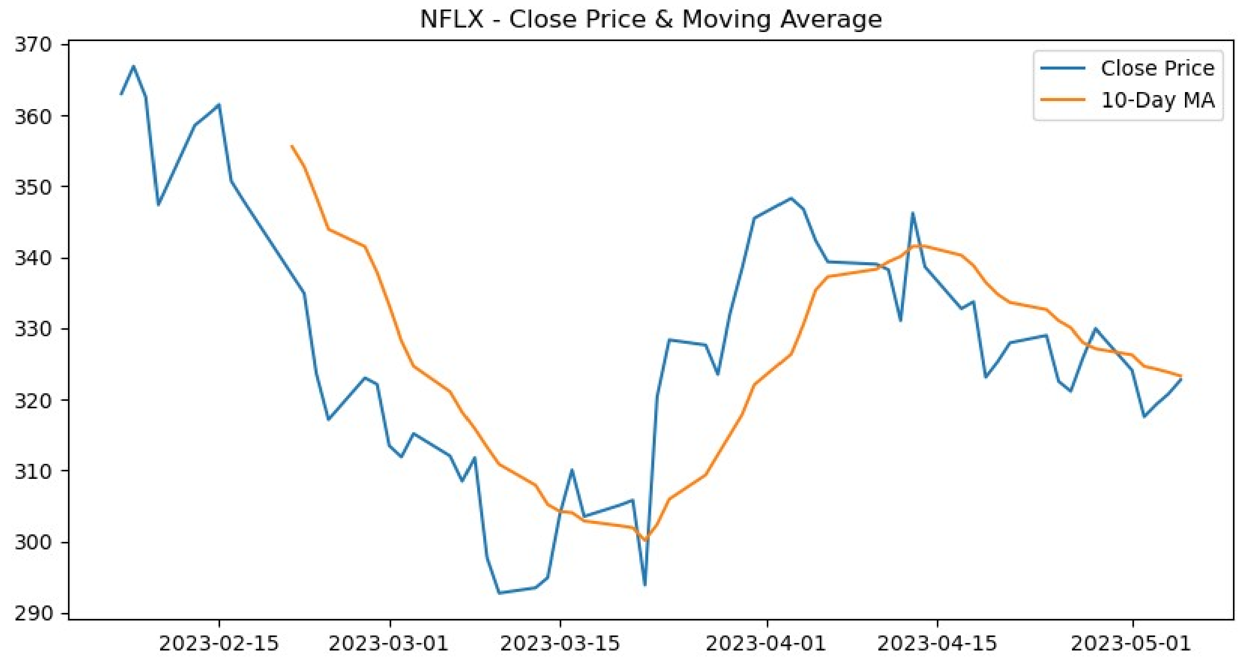


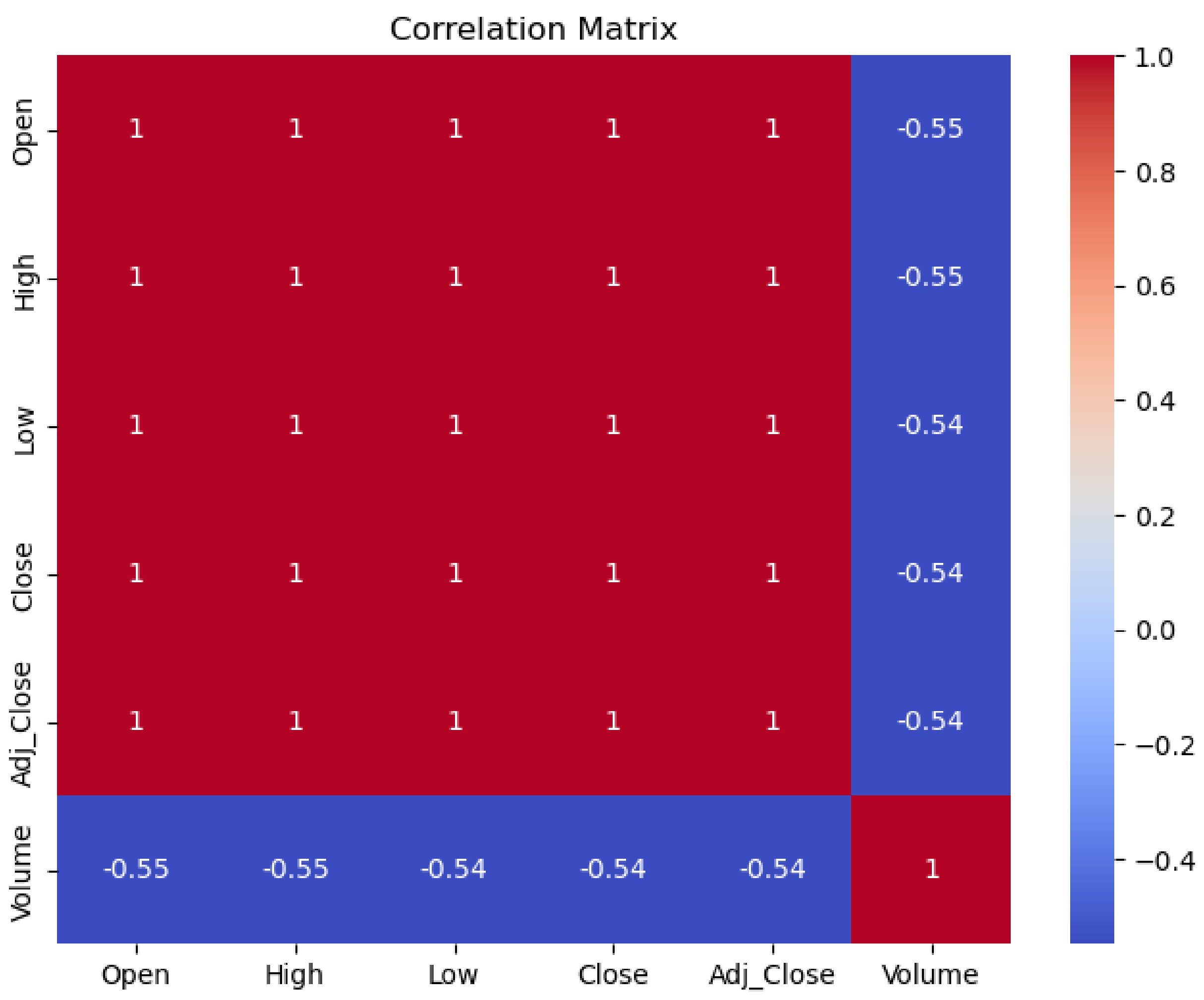
Closing Price Distribution by Ticker
Close
350
300
250
200
150
100
AAPL
GOOG
MSFT
NFLX
Ticker


AAPL - Close Price & Moving Average
Close Price
10-Day MA
170
165
160
155
150
145
2023-02-15
2023-03-01
2023-03-15
2023-04-01
2023-04-15
2023-05-01

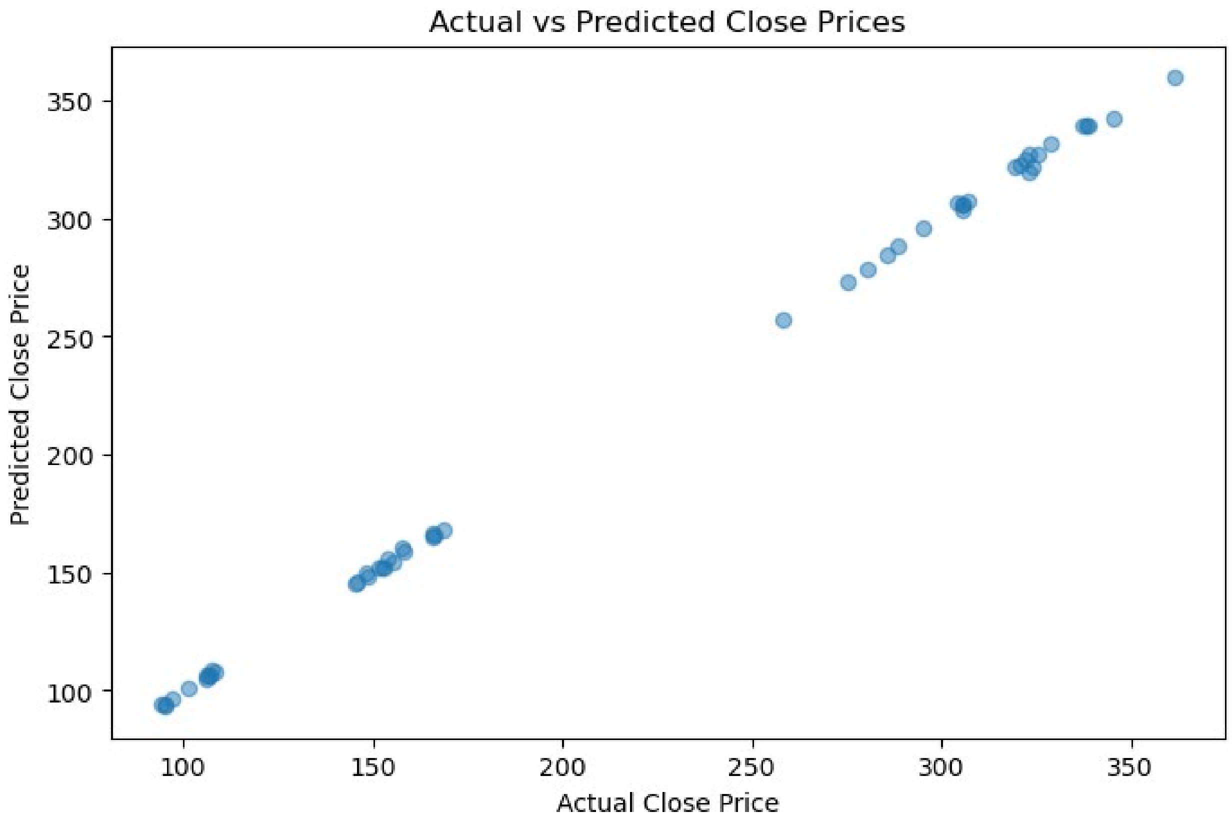

GOOG - Close Price & Moving Average
110.0
107.5
105.0
102.5
100.0
97.5
95.0
92.5
90.0
Close Price
10-Day MA
2023-02-15
2023-03-01
2023-03-15
2023-04-01
2023-04-15
2023-05-01


MSFT - Close Price & Moving Average
Close Price
10-Day MA
310
300
290
280
270
260
250
2023-02-15
2023-03-01
2023-03-15
2023-04-01
2023-04-15
2023-05-01


NFLX - Close Price & Moving Average
Close Price
10-Day MA
370
360
350
340
330
320
310
300
290
2023-02-15
2023-03-01
2023-03-15
2023-04-01
2023-04-15
2023-05-01



```
Model Evaluation:
RMSE: 1.6088819481286896
R2 Score: 0.9997047590407421
```


Actual vs Predicted Close Prices
Predicted Close Price
Actual Close Price
100
150
200
250
300
350