Analyzing Apple's Stock Market Performance: A Data-Driven Approach to Trends, Patterns, and Forecasting

Jonathan Siegel

San Diego State University

CS577: *Principles and Techniques of Data Science*

Sean Kang

December 8, 2024

Abstract

This project analyzes Apple’s stock price and performance from 1980 to 2019 using historical stock market data. The dataset includes key financial information such as daily openings, closings, high and low prices, adjusting closing prices, and trading volume. This study will explore the long-term trends, stock volatility, and significant market events that influenced Apple’s stock price. These insights can inform investors and analysts seeking to understand Apple’s historical market behavior and make future investment decisions.

Introduction

The stock market has long been an indicator of a company's success when looking through the lens of financial health and growth potential. As one of the world's most outstanding and influential companies, Apple Inc.'s stock (https://www.kaggle.com/datasets/varpit94/apple-stock-data-updated-till-22jun2021) performance has attracted significant attention from investors and analysts. Since 1980, Apple's stock has experienced remarkable growth mirroring the company's technological advancements from their desktop and laptop series, to the introduction of iPods and iPads, and most popularly the iPhone. Understanding Apple stock's historical patterns, trends, and volatility can provide valuable insights into the factors that influence its price fluctuations and overall market performance. In this report, not only will I analyze the success of Apple in the stock market, but I will also simulate a hypothetical scenario.

More specifically, a "what if" case in which we purchase 25 shares of Apple stock on day one. We will then determine how much profit would have been made over time, explore Apple's all-time high stock price according to the dataset, predict future profit, and examine the outcome at its all-time low.

Approach

Data Collection

The dataset was derived from Kaggle, a CSV containing daily stock prices and volume for Apple Inc. from 1980 to 2019. I utilized the head() function from the pandas library to ensure I had imported the CSV correctly. Missing data was identified and removed to ensure consistency for the analysis.

```
data.head()
```
[4] ✓ 0.0s

	Date	Open	High	Low	Close	Adj Close	Volume
0	1980-12-12	0.513393	0.515625	0.513393	0.513393	0.407747	117258400.0
1	1980-12-15	0.488839	0.488839	0.486607	0.486607	0.386473	43971200.0
2	1980-12-16	0.453125	0.453125	0.450893	0.450893	0.358108	26432000.0
3	1980-12-17	0.462054	0.464286	0.462054	0.462054	0.366972	21610400.0
4	1980-12-18	0.475446	0.477679	0.475446	0.475446	0.377609	18362400.0

```
# Convert the Date column to datetime for proper analysis
data['Date'] = pd.to_datetime(data['Date'])
# Check for any missing or null values in the dataset
missing = data.isnull().sum()
print(missing)
# Display the missing values to identify any issues
data_cleaned = data.dropna()
data_cleaned.isnull().sum()
data = data_cleaned
```
[3] ✓ 0.0s

```
Date         0
Open         1
High         1
Low          1
Close        1
Adj Close    1
Volume       1
dtype: int64
```

Running data.isnull().sum() a second time returned a similar statistic above, but now consisting of a column with all zeros. The Date column was then converted into a DateTime format to allow time-series manipulation.

Exploratory Data Analysis

On the first day of the stock market, according to the dataset with 25 shares, we would have money around the range of $10 to $12. This was determined by the line below

```
#determine how much we would have at day1 with 25 shares
day1 = 25 * data.iloc[0]['Open']
print('Apple\'s day one stock price falls in the range of $', (25 * data.iloc[0]['Adj Close']), 'to $', 25 * data.iloc[0]['Open'])
```

According to Apple's website for investors, since going public there have been 5 splits in the stock market. A stock split is an action in which a company divides its existing shares into multiple new shares to boost the stock's liquidity and make it more affordable for retail investors without affecting the overall value of the company. In my Python notebook, I calculated what my stock price would be with our hypothetical scenario starting with 25 shares, and here's what happened.

```python
# Tracking the splits over time
shares1987 = day1share * 2
print('1987 share number', shares1987)
shares2000 = shares1987 * 2
print('2000 share number', shares2000)
shares2005 = shares2000 * 2
print('2005 share number', shares2005)
shares2014 = shares2005 * 7
print('2014 share number', shares2014)
shares2020 = shares2014 * 4
print('2020 share number', shares2020)

# Final number of shares after all splits
finalshares = shares2020

# Get the stock price on the last available date in the dataset (November 2019)
latest_price_2019 = data.loc[data['Date'] == data['Date'].max(), 'Adj Close'].values[0]

# Final investment value based on the latest price in 2019
finalinvestmentvalue = finalshares * latest_price_2019

# Display the results
print("Final number of shares after splits:", finalshares)
print(f"Value of investment using the latest available price in 2019: ${finalinvestmentvalue:,.2f}")
```

✓ 0.0s

```
1987 share number 50
2000 share number 100
2005 share number 200
2014 share number 1400
2020 share number 5600
Final number of shares after splits: 5600
Value of investment using the latest available price in 2019: $1,491,671.97
```

The next metric I would like to touch base on is Apple's daily stock return, which is referred to as the day-to-day percentage changes in the stock's adjusted close price. Analyzing the returns gives insights into the stock's volatility, or how much it fluctuates. This helps in asserting risk and reward potential over short timeframes. With the following code, I was able to calculate the daily return, its mean, and volatility.

```
data['Daily Return'] = data['Adj Close'].pct_change() * 100
mean_return = data['Daily Return'].mean()
volatility = data['Daily Return'].std()
```

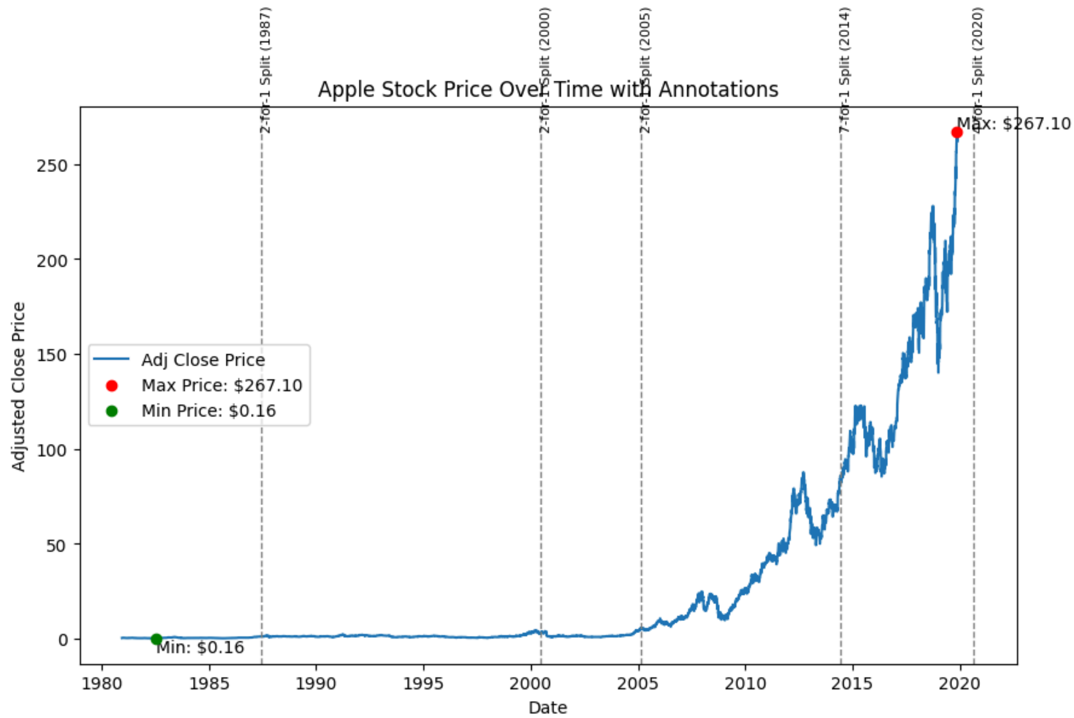
Mean Daily Return: 0.11%

Volatility (Std Dev): 2.87%

The following output tells us the following information about Apple's stock market: there is a 0.11% increase per day over our timeline. To get a sense of the annual growth, we can consider the daily return by the number of trading days in a year which comes down to 252. Annualized Return = 0.11% * 252 = 27.72%. This estimate shows that Apple’s stock price has grown 27.72% annually based on the provided dataset which aligns with Apple’s strong long-term performance. Next, our Volatility measures how much of Apple’s stock tends to fluctuate from the mean return. A 2.87% volatility means that on most days, Apple’s stock price will fluctuate +/- 2.87% above or below the mean daily return. The significance behind this is that higher volatility indicates greater risk because the stock price experiences larger fluctuations. However, it also means the potential for higher returns or losses on any given day.

Data Visualization

Below you can see a chart of Apple’s stock prices from 1980 to 2021. There are some annotations in regards to the max and mins, furthermore, the 5 splits that occurred.

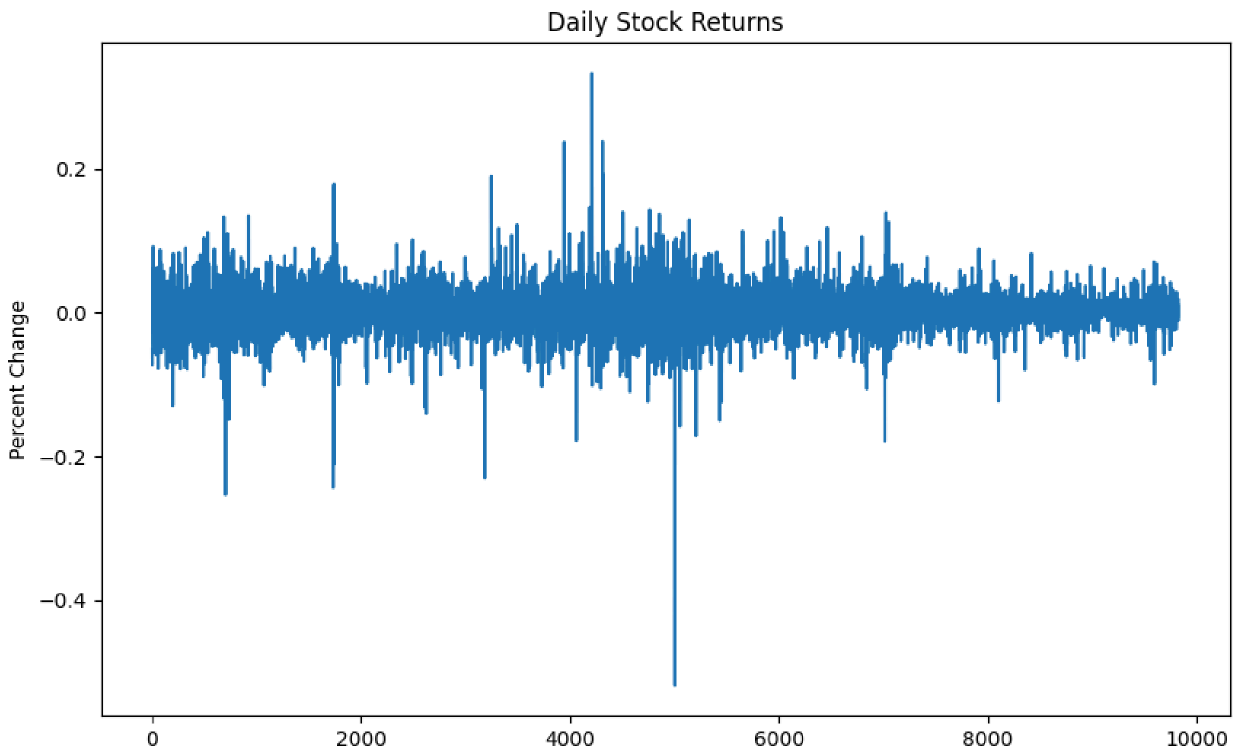


To verify this information regarding the minimum and maximum price I ran the following code blocks. The dates are accurate however, comparing it with the most up-to-date data via Yahoo the prices yet again went down. This is due to the dataset not being entirely up to date (1980 - 2024).

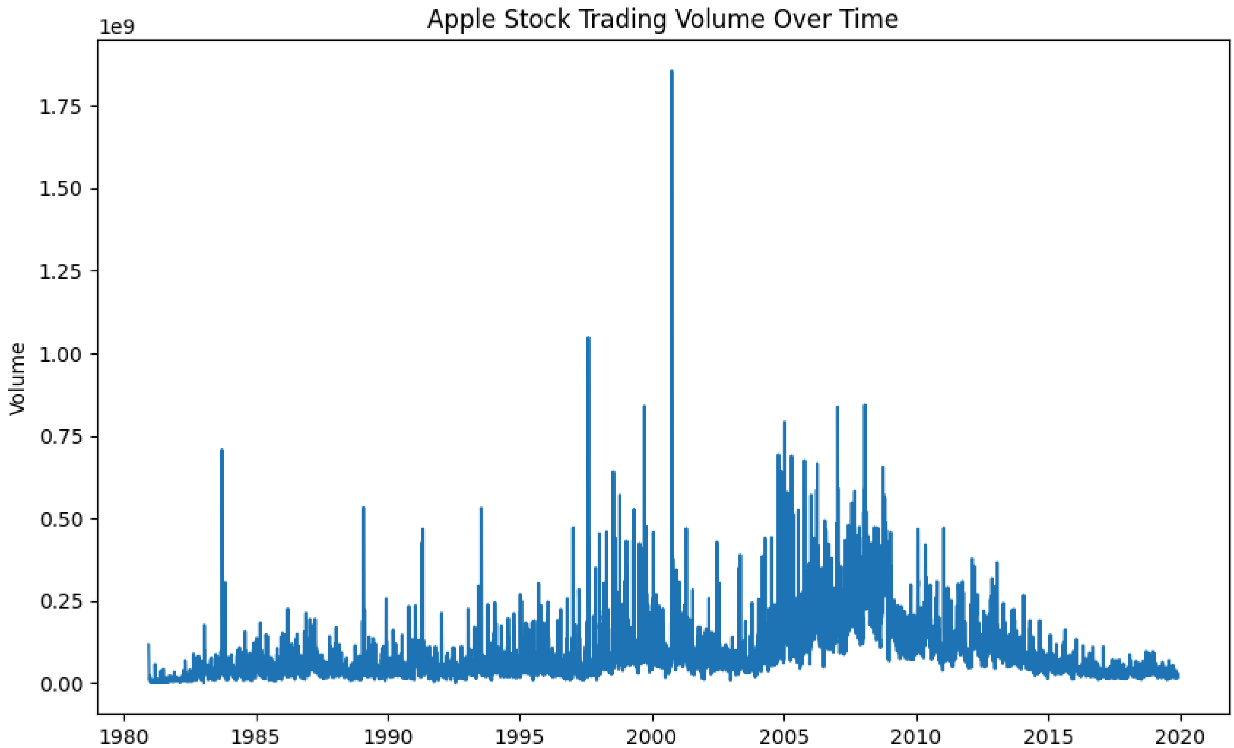
```
Minimum Adjusted Price: $0.16

Date of Minimum Price: 1982-07-08T00:00:00.000000000

Maximum Adjusted Price: $267.10

Date of Maximum Price: 2019-11-18T00:00:00.000000000
```

Some further cross-validating with Google shows that the graph still skews left showing that even after 2019, Apple’s success in the stock market continues to grow.

The dramatic increase is due to the innovations of the iPhone, iPad, and iPod touch. On June 29, 2007, the first iPhone was released. It has combined everything great about the iPod touch, the fact that it was more than a listening device but you could also surf the internet with it. Now with the iPhone, calling and texting have become more efficient, and it is easy for people to abandon their flip phones and upgrade to something more revolutionary.

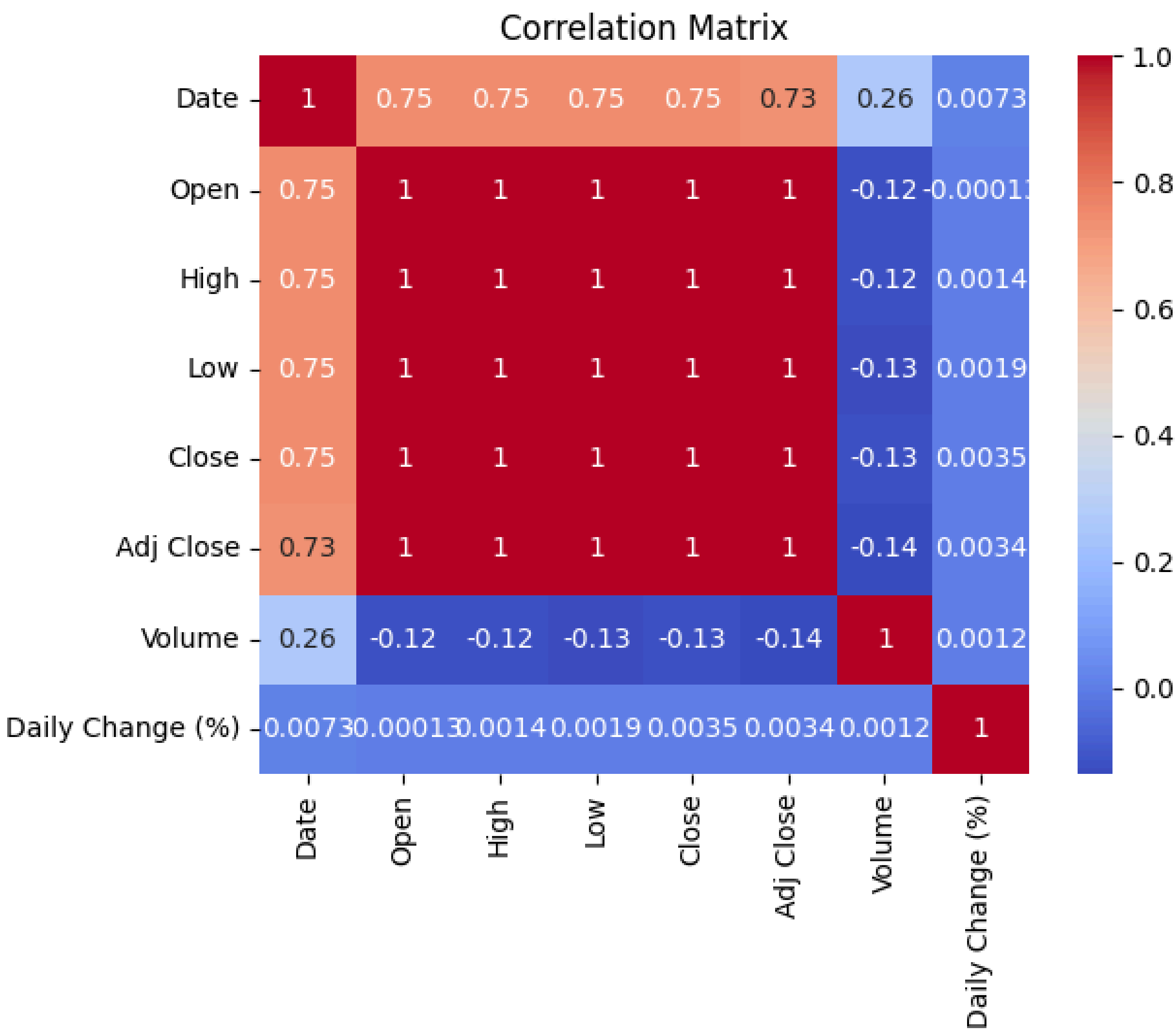


The graph above is titled "Daily Stock Returns" which shows the percentage changes in Apple's stock price daily over time. The first instance the graph shows is volatility, these are presented amongst themselves in the large spikes and dips the graph. These spikes and dips are most common during major market events, economic downturns, and even company-specific announcements. There are some notable outliers too, going back to the large negative spikes, this shows market crashes or other downturns which may include the 2008 financial crisis, the dot-com bubble, and early instances of COVID-19. In contrast, as I had mentioned earlier the upward spikes can be due to an evolutionary release from Apple like another iPhone or MacBook laptop which may introduce a new chip. In sum, The slightly positive average daily return suggests long-term upward momentum while the low frequency could mean there is controlled volatility for a stock of its size and market influence.



This graph shows the relationship between trading volume over time in Apple's stock history from the 1980s to the 2020s. In the earlier periods (1980s - 1990s) this was Apple's early years as a public company which is why the trading volume is relatively low. This reflects Apple's smaller market presence at launch which also meant less investors and attention at the time. From the 1990s to the 2000s the volume increased, this is due to the dot-com boom and the growing recognition of Apple as a major competitor in the tech industry. Now we have the 2000 spike, outside of the dot-com stock boost, Steve Jobs returned to Apple in 1997 and the company started a turnaround. This turnaround started with the release of the iMac G3 in 1998 which allowed the "fashion" aspect of computers in appearance. We also introduced Mac OS X in 2001, arguably Apple's first interactive operating system that continues to be updated to this day. Lastly, we had the release of the iPod in 2001, this release redefined Apple as a consumer electronic leader and expanded their market appeal. Post 2005 - 2015 we see a small increase in trading volume then shortly after, trade starts to decline. This may indicate that there is market stabilization or a shift to long-term investors who trade less frequently.

The last basis of visual description I would like to touch upon is the correlation matrix.



The correlation matrix for Apple’s stock data reveals strong and positive correlations between the Open, High, Low, Close, and Adjusted close pieces. This indicates that these price metrics move almost identically. This is expected for a stock market dataset because the daily price fluctuates through consistent market trends and investor behavior. For instance, when the opening price of Apple stock is high, likely, the other price features (high. Low, close etc) will likely also reflect this trend during that trading day. On a more interesting note, the volume shows a weak negative correlation, being -0.12 to -0.14 with the price metrics. With this information, I can conclude that high trading volume does not lead to higher prices and in some cases, it could align with slight downward pressure. Buying and selling are not always tied directly to price increases but may reflect broader market sentiment or volatility.

Overall, the matrix emphasizes that Apple stock prices are tightly interlinked but only weakly associated with trading volume and daily changes.

Statistical Analysis

```
data.describe()
```

	Date	Open	High	Low	Close	Adj Close	Volume
count	9822	9822.000000	9822.000000	9822.000000	9822.000000	9822.000000	9.822000e+03
mean	2000-05-28 06:28:48.527794688	30.350057	30.640633	30.048298	30.352940	28.364033	8.620720e+07
min	1980-12-12 00:00:00	0.198661	0.198661	0.196429	0.196429	0.156008	3.472000e+05
25%	1990-08-30 06:00:00	1.062500	1.085357	1.043571	1.066964	0.914089	3.318138e+07
50%	2000-05-18 12:00:00	1.709286	1.742366	1.671429	1.712857	1.452872	5.798940e+07
75%	2010-02-24 18:00:00	30.393928	30.555714	29.865358	30.233214	26.305946	1.076320e+08
max	2019-11-25 00:00:00	267.899994	268.000000	265.390015	267.100006	267.100006	1.855410e+09
std	NaN	53.443016	53.908364	52.983661	53.464270	51.603590	8.623138e+07

The output above is from a function within the data frame library called “describe” This outputs information like count, mean, min, and other meaningful metrics, and here is what we could derive from it. The mean is the average value of each column over the dataset’s timespan. From this, we know that the average opening price is $30.35 and the average closing price is $28.36. The average trading volume per day is approximately 86.2 million shares. Our data shows that

25% of our data shows that opening prices were at one point $1.06 which could either mean that this is reflecting how prices were around a day or at some point apple might have taken some hits which could have been caused by a lot of factors like bad product. However, when Apple performed in the stock market, it never disappointed. The 75% tile shows that the opening prices were just a little under $30.39 and the adjusted closing prices were $26.31. The trading volumes were just below 107.63 million shares. A lot more money and trading activity to be seen since day 1, or at least compared to our first 25%.

Results

Given all the research and deep diving that has been discovered throughout this study, we can safely assume that Apple is successful in the markets. Although our dataset only covers the last few years, through personal anecdotes it is no secret that Apple continues to succeed.

Outcome 1 - Training our Model to Evaluate Accuracies

```
data['Date'] = pd.to_datetime(data['Date'])
data['Daily Change (%)'] = (data['Close'] - data['Open']) / data['Open'] * 100
data.dropna(inplace=True)

# Select features (X) and target (y)
X = data[['Open', 'High', 'Low', 'Volume']]
y = data['Close']
```

This code preprocesses a stock price dataset to prepare it for predictive modeling. We ensure this process happens by first converting the Date column from a string format to DateTime object, this makes it easier for manipulation and time-based analysis of the data. Next, we calculate a new column called Daily Change, which represents the percentage change in stock price during a trading day. This is done by taking the difference between the Close and

Open prices, dividing it by the Open, and multiplying by 100. The code then removes any missing values to ensure that the dataset is clean and ready for analysis.

```
# Split the data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Train and evaluate regression algorithms
models = {
    'Linear Regression': LinearRegression(),
    'Random Forest': RandomForestRegressor(random_state=42, n_estimators=100)
}

results = {}
predictions = {}

# Train and evaluate each model
for name, model in models.items():
    model.fit(X_train, y_train)
    y_pred = model.predict(X_test)
    predictions[name] = y_pred
    mae = mean_absolute_error(y_test, y_pred)
    mse = mean_squared_error(y_test, y_pred)
    results[name] = {'MAE': mae, 'MSE': mse}
```

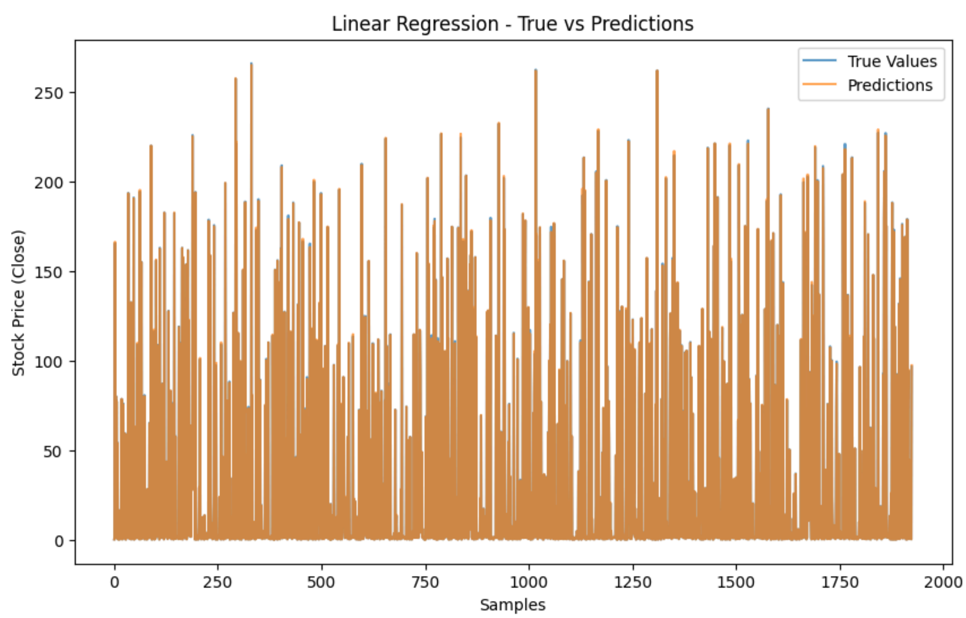
The code above is designed to train and evaluate two different types of regression models, Linear Regression and Random Forest, to predict a target variable. The dataset is split into training and testing sets using the train_test_split function where 80% of the data is used for training and 20% is reserved for testing. The random state ensures that the split is reproducible. The linear regression model captures simple linear relationships whereas the Random Forest ensembles a model of decision trees, capturing more complex, non-linear patterns. Each model is trained on the training data (x_train and y train) which is then used to make a prediction. To evaluate the models, the Mean Absolute Error and the Mean Squared error are calculated.

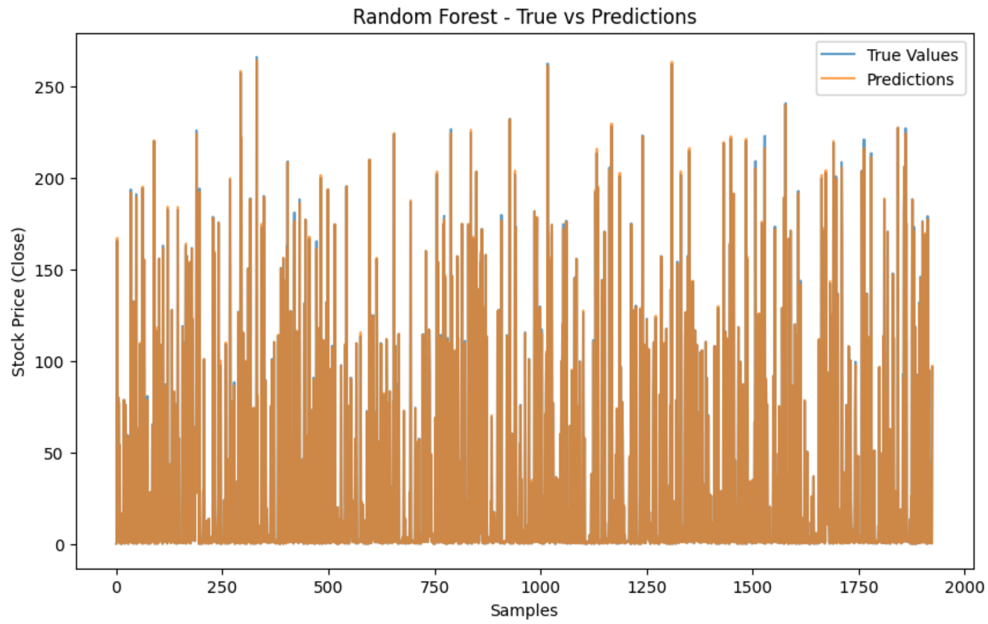
Model Evaluation Metrics:

	MAE	MSE
Linear Regression	0.128536	0.102544

Random Forest 0.190390 0.246716

These metrics measure how close the predicted values are to the actual ones, with lower values representing better performance. The Linear Regression outperforms the Random Forest in this scenario, with lower MAE and MSE values. This implies that the relationship between the features and the target variables is linear, making Linear Regression a more suitable model for this dataset. However, Random Forest's performance, while reasonable, does not justify the added complexity of our task. Below you can see how our model trained in terms of True Values vs Predictions.





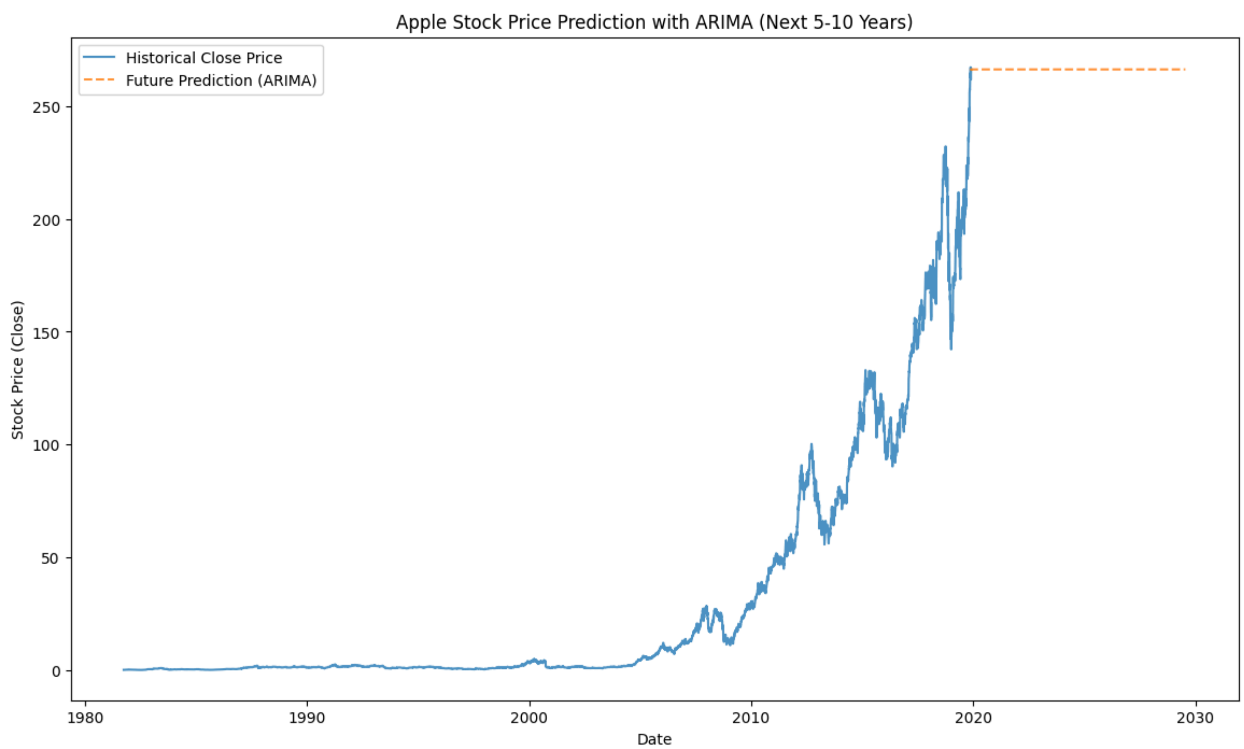
As you can see although they are very close the Linear Regression and Random Forest are really close to each other, but the Linear regression model shows less blue then the random forest model.

Outcome 2 - Predicting How our Hypothetical Scenario will look in 5-10 years

To determine how our stock will do in the next 5-10 years could be evaluated with the code below

```
warnings.filterwarnings("ignore")
arima_data = data['Close']
arima_model = ARIMA(arima_data, order=(5, 1, 0))
arima_result = arima_model.fit()
future_steps = 10 * 252  # 10 years of trading days
arima_forecast = arima_result.forecast(steps=future_steps)

future_arima_dates = pd.date_range(start=data['Date'].max(), periods=future_steps, freq='B')

# Plot historical and forecasted prices
plt.figure(figsize=(14, 8))
plt.plot(data['Date'], data['Close'], label='Historical Close Price', alpha=0.8)
plt.plot(future_arima_dates, arima_forecast, label='Future Prediction (ARIMA)', linestyle='--', alpha=0.8)
plt.title('Apple Stock Price Prediction with ARIMA (Next 5-10 Years)')
plt.xlabel('Date')
plt.ylabel('Stock Price (Close)')
plt.legend()
plt.show()
```

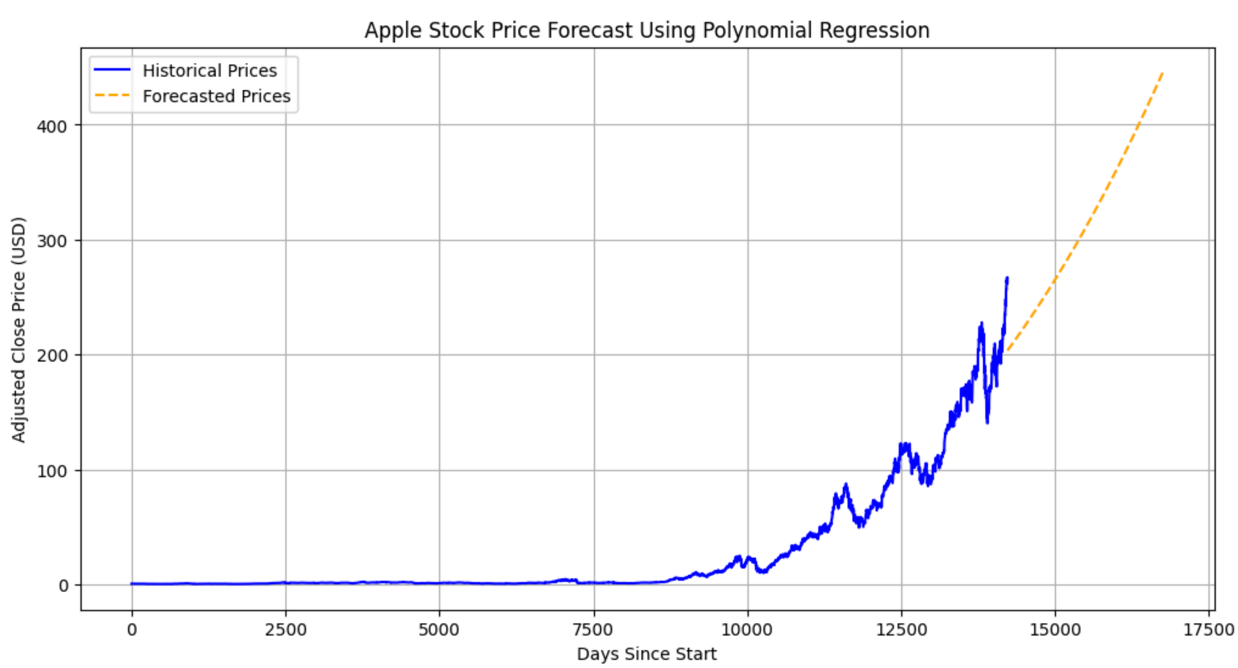
ARIMA is an auto-regression component that is similar to linear regression, but it is more tailored to time-series analysis and could include additional mechanisms like differencing and error modeling. From that, I was able to generate the following graph



Although our model appears to do well, ARIMA relies on statistical properties like mean, variance, and covariance and does not account for exponential or compound growth for a stock like Apple's. If we dig into our dataset, if recent historical stock prices have less variability or growth our model might have a hard time identifying that thus struggling to show an upward trend. Lastly, ARIMA does not incorporate factors like market growth, product launches, economic trends etc.

```python
# Prepare data for polynomial regression
X = apple_prices[['Days']].values
y = apple_prices['Adj Close'].values

# Split the data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Apply polynomial features for non-linear trends
poly = PolynomialFeatures(degree=3)
X_poly_train = poly.fit_transform(X_train)
X_poly_test = poly.transform(X_test)

# Train a linear regression model on the polynomial data
model = LinearRegression()
model.fit(X_poly_train, y_train)

# Forecast the stock price 10 years into the future (252 trading days per year)
future_days = np.arange(X[-1][0], X[-1][0] + 10 * 252).reshape(-1, 1)
future_poly = poly.transform(future_days)
forecast_prices = model.predict(future_poly)

# Calculate the predicted price for 5 and 10 years
price_5_years = forecast_prices[252 * 5 - 1]
price_10_years = forecast_prices[252 * 10 - 1]
```

The following code above implemented another type of regression called polynomial regression. From the output of our code above it is clear that apple stock does not grow on a linear trend and more of a exponential or polynomial trend and got a way more accurate output with the graph below.



With this information too, the code was able to determine how much money we would have (hypothetically) if we bought 25 shares of Apple stock on day 1.

Value of 25 shares in 5 years: $7735.09

Value of 25 shares in 10 years: $11131.95

To reiterate, of course, there are numerous amounts of factors that could either prove the accuracy of this model or hurt the accuracy of this model. Ultimately it comes down to how Apple decides to move forward as a company with its technological releases, or how it decides to move forward with financial strategy, like splitting the stock again.

Discussion

Given all the research and deep diving that has been discovered throughout this study, we can safely assume that Apple is successful in the markets. Although our dataset only covers the last few years, through personal anecdotes it is no secret that Apple continues to succeed. In recent times we have the release of the iPhone 16 series, this iPhone series introduced Artificial intelligence through the powerful A18 chip. This phone series has done tremendous numbers selling more than 9 million units. Apple has launched a new series of products called the Apple Vision Pros which takes the next level on virtual reality simulation. The Wall Street Journal had anticipated that the VR headset would only sell 500k units which is already impressive on its own, but Apple was able to sell somewhere between 700,000 and 800,000k units.

Open	243.99	Mkt cap	3.67T	CDP score	A-
High	244.54	P/E ratio	39.95	52-wk high	244.54
Low	242.13	Div yield	0.41%	52-wk low	164.08

Apple's commitment to innovation has driven its stock price higher since 2021. As the company continues to introduce groundbreaking products and services, its future in the market looks promising.

References

"Apple Inc. (AAPL) Stock Price, News, Quote & History." *Yahoo! Finance*, Yahoo!, 25 Oct. 2024, finance.yahoo.com/quote/AAPL/.

"Apple Inc. (AAPL) Stock Price, News, Quote & History." *Yahoo! Finance*, Yahoo!, 25 Oct. 2024, finance.yahoo.com/quote/AAPL/.

Brown, David. "An In-Depth Analysis of AAPL Stock and Apple Inc." *DB Investing - Broker You Can Trust*, 3 Oct. 2023, dbinvesting.com/apple-aapl/.

Christoph. "Timeline – the History of Apple (2000 – 2009) - MAC History." *Mac History*, 16 Dec. 2022, www.mac-history.net/2022/12/16/timeline-the-history-of-apple-2000-2009/.

DeJohn, Jaclyn. *History of Apple: Company and Stock*, 20 Aug. 2024, smartasset.com/investing/apple-history.

"FAQ - Apple Investor Relations." *Apple*, investor.apple.com/faq/default.aspx. Accessed 27 Oct. 2024.

Kerin, James. "How Product Releases Affect Apple's Stock Price." *Investopedia*, Investopedia, www.investopedia.com/articles/stocks/12/history-apple-stock-increases.asp. Accessed 27 Oct. 2024.

Zucchi, Kristina. "What Moves Apple's Stock Price?" *Investopedia*, Investopedia, www.investopedia.com/stock-analysis/061115/what-moves-apples-stock-price-aapl.aspx. Accessed 27 Oct. 2024.