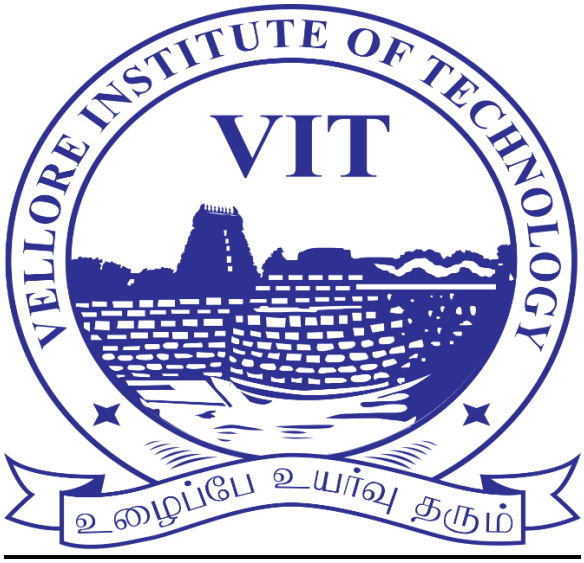


Artificial Intelligence (CSE3088)

EDA Review

Topic: Tesla and Ferrari Stock Prediction

Team:

- Shashank Singh 21MIA1110
- Dazzle A J 21MIA1119

Introduction

The dataset focuses on two prominent companies in the automotive industry: Ferrari and Tesla. Ferrari is a luxury sports car manufacturer based in Italy, known for its high-performance vehicles and prestigious brand. Tesla, on the other hand, is an innovative electric vehicle (EV) and clean energy company headquartered in the United States, renowned for its electric cars, energy storage solutions, and renewable energy products.

The dataset contains information about the daily closing prices of Ferrari and Tesla shares over an eight-year period, spanning from 2015 to 2023.

It can be used to investigate the impact of various events, market conditions, and economic factors on the share prices of these two companies.

Five Operations Performed on the dataset

- Data Summary
- Data Visualization
- Feature Scaling (Normalization)
- Feature Selection
- Dimensionality Reduction

Data Summary

Purpose: Provide an overview of the dataset by examining basic statistics and properties.

Results: Obtain descriptive statistics (mean, median, standard deviation, etc.), data types, missing values, and distribution of variables

- **Loading Up the Tesla Data:**

	Date	Open	High	Low	Close	Adj Close	Volume
0	2015-10-22	14.104000	14.383333	13.960000	14.114667	14.114667	42378000
1	2015-10-23	14.333333	14.356667	13.846000	13.939333	13.939333	63532500
2	2015-10-26	14.092000	14.392000	14.000000	14.350667	14.350667	50871000
3	2015-10-27	14.322667	14.473333	13.834000	14.023333	14.023333	52791000
4	2015-10-28	14.087333	14.230000	13.886667	14.197333	14.197333	40929000

- **Loading Up the Ferrari Data:**

	Date	Open	High	Low	Close	Adj Close	Volume
0	2015-10-22	57.070000	58.200001	55.700001	56.750000	53.604126	4545100
1	2015-10-23	57.770000	58.000000	56.270000	56.380001	53.254635	1967600
2	2015-10-26	57.000000	57.000000	54.540001	55.020000	51.970028	1466300
3	2015-10-27	54.799999	54.990002	49.360001	53.849998	50.864880	5949200
4	2015-10-28	54.020000	54.160000	50.099998	51.869999	48.994644	2411300

- **Number of Rows & Columns in Both the Dataset:**

```
Number of rows: 1885
Number of columns: 7
```

- **Data Type of Both the Dataset:**

```
Date          object
Open         float64
High         float64
Low          float64
Close        float64
Adj Close    float64
Volume         int64
dtype: object
```

- **Missing Values in Both the Dataset:**

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

- **Description of Tesla Dataset:**

	Open	High	Low	Close	Adj Close	Volume
count	1885.000000	1885.000000	1885.000000	1885.000000	1885.000000	1.885000e+03
mean	100.049049	102.330085	97.545756	99.989301	99.989301	1.181818e+08
std	110.688467	113.241143	107.806630	110.544977	110.544977	8.181768e+07
min	9.488000	10.331333	9.403333	9.578000	9.578000	1.062000e+07
25%	17.270666	17.557333	16.941334	17.243999	17.243999	6.745200e+07
50%	23.309999	23.650000	22.916000	23.290001	23.290001	9.321150e+07
75%	202.029999	207.696671	197.833328	202.070007	202.070007	1.394250e+08
max	411.470001	414.496674	405.666656	409.970001	409.970001	9.140820e+08

- **Description of Ferrari Dataset:**

	Open	High	Low	Close	Adj Close	Volume
count	1885.000000	1885.000000	1885.000000	1885.000000	1885.000000	1.885000e+03
mean	146.694973	147.990397	145.406204	146.696976	144.380032	4.850036e+05
std	64.586379	65.092660	64.076793	64.569438	65.175258	4.721486e+05
min	32.290001	32.480000	31.660000	32.000000	30.226114	5.380000e+04
25%	105.720001	106.580002	104.260002	105.190002	101.958740	2.602000e+05
50%	151.500000	153.259995	150.000000	151.039993	148.161163	3.703000e+05
75%	200.860001	202.240005	199.264999	200.729996	199.868561	5.471000e+05
max	284.940002	286.609985	284.350006	285.529999	285.529999	7.485100e+06

- **Number of Duplicate Rows in Both the Dataset:**

```
Number of duplicate rows: 0
```

- **Earliest & Latest Date of Shares of Both Tesla and Ferrari Datasets:**

```
Earliest date: 2015-10-22
Latest date: 2023-04-19
```

- **Total Trading Volume of Tesla Dataset:**

```
Total trading volume per year:
Year
2015     2419168500
2016    17435425500
2017    23850471000
2018    32424582000
2019    34620726000
2020    57158737200
2021    20708071500
2022    21821019600
2023    12334431500
Name: Volume, dtype: int64
```

- **Total Trading Volume of Ferrari Dataset:**

```
Total trading volume per year:
Year
2015     47745400
2016    173509400
2017    143317200
2018    157306800
2019     99901700
2020     86284800
2021     81671600
2022     97582000
2023     26912900
Name: Volume, dtype: int64
```

- **Average Share Price of Tesla Dataset Per Year:**

```
Tesla average share price per year:
Year
2015     14.910694
2016     13.984484
2017     20.954420
2018     21.153995
2019     18.235347
2020     96.665689
2021    259.998162
2022    263.093081
2023    176.328108
Name: Close, dtype: float64
```

- **Average Share Price of Ferrari Dataset Per Year:**

```
Ferrari average share price per year:
Year
2015     49.474694
2016     46.020278
2017     91.249322
2018    124.199203
2019    147.236786
2020    178.309683
2021    218.776944
2022    206.874063
2023    258.806081
Name: Close, dtype: float64
```

- **Average Share Price of Tesla Per Month:**

```
Tesla average price per month:
Month
1     105.003316
2     109.659089
3     105.313809
4     104.452530
5      81.433554
6      83.543694
7      91.177265
8     103.141966
9     109.648490
10     98.207217
11    102.374901
12    103.466480
Name: Close, dtype: float64
```

- **Average Share Price of Ferrari Per Month:**

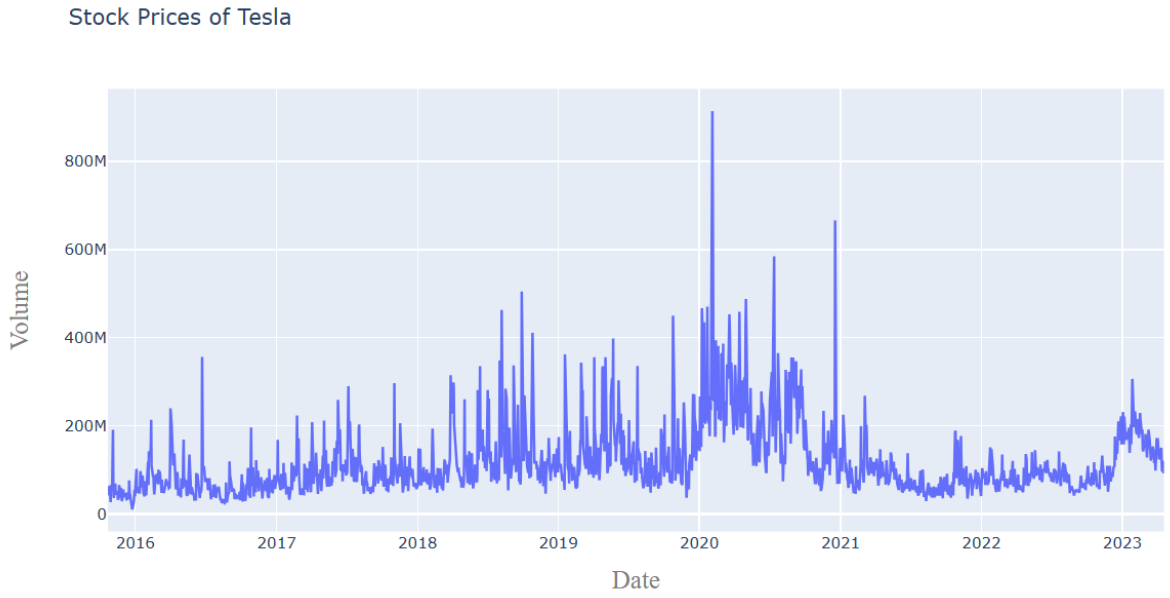
```
Ferrari average price per month:
Month
1     149.508199
2     150.777385
3     148.198258
4     147.921987
5     135.882973
6     140.782800
7     147.360479
8     150.573057
9     149.956924
10    143.956124
11    146.352121
12    148.341071
Name: Close, dtype: float64
```

Data Visualization

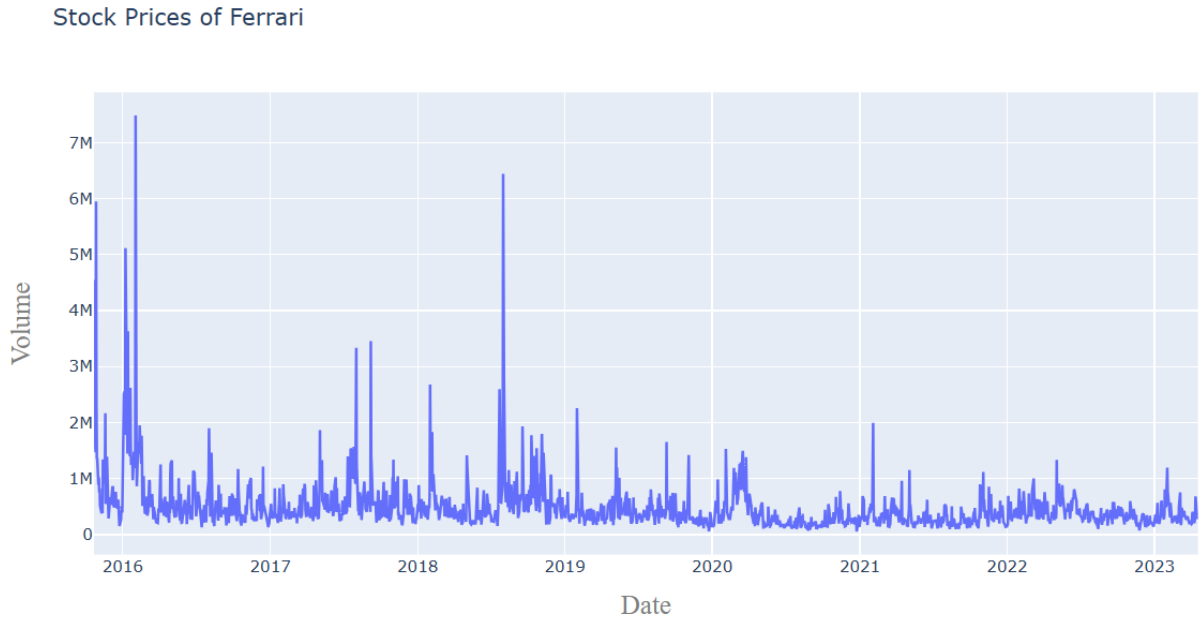
- **Graph of Price vs Year of Tesla Shares:**

- **Graph of Price vs Year of Ferrari Shares:**

- **Graph of Volume vs Year of Tesla Shares:**

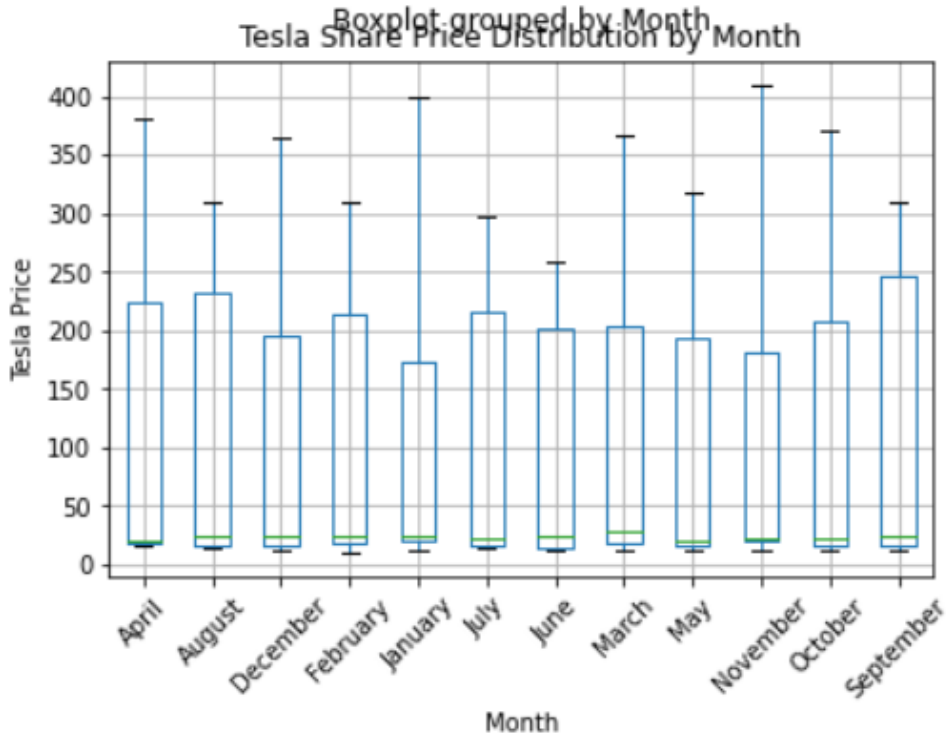


- **Graph of Volume vs Year of Ferrari Shares:**



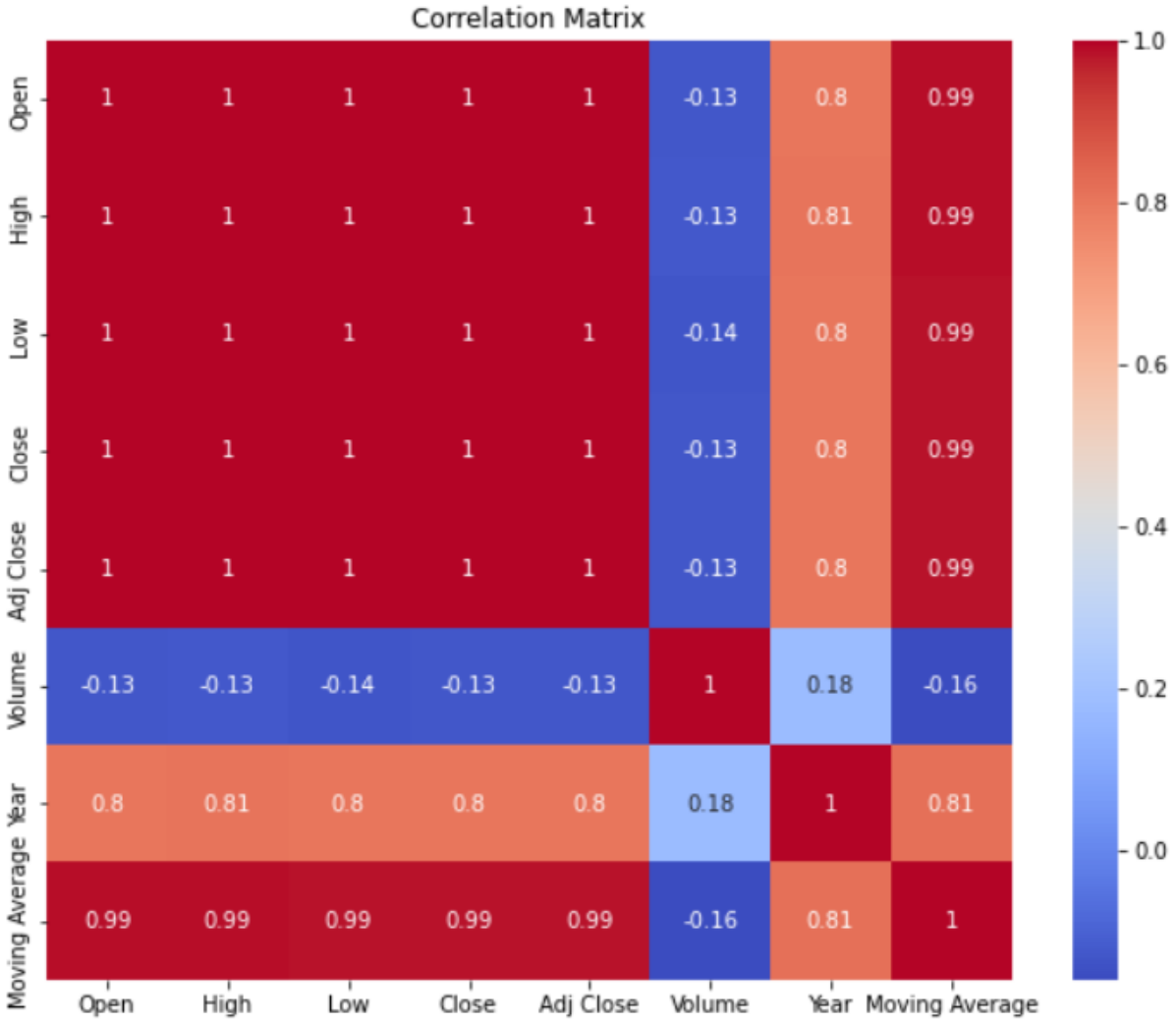
Now we have done box plot to visualize the stock price by month

- **Tesla Share Price Distributed By Month:**

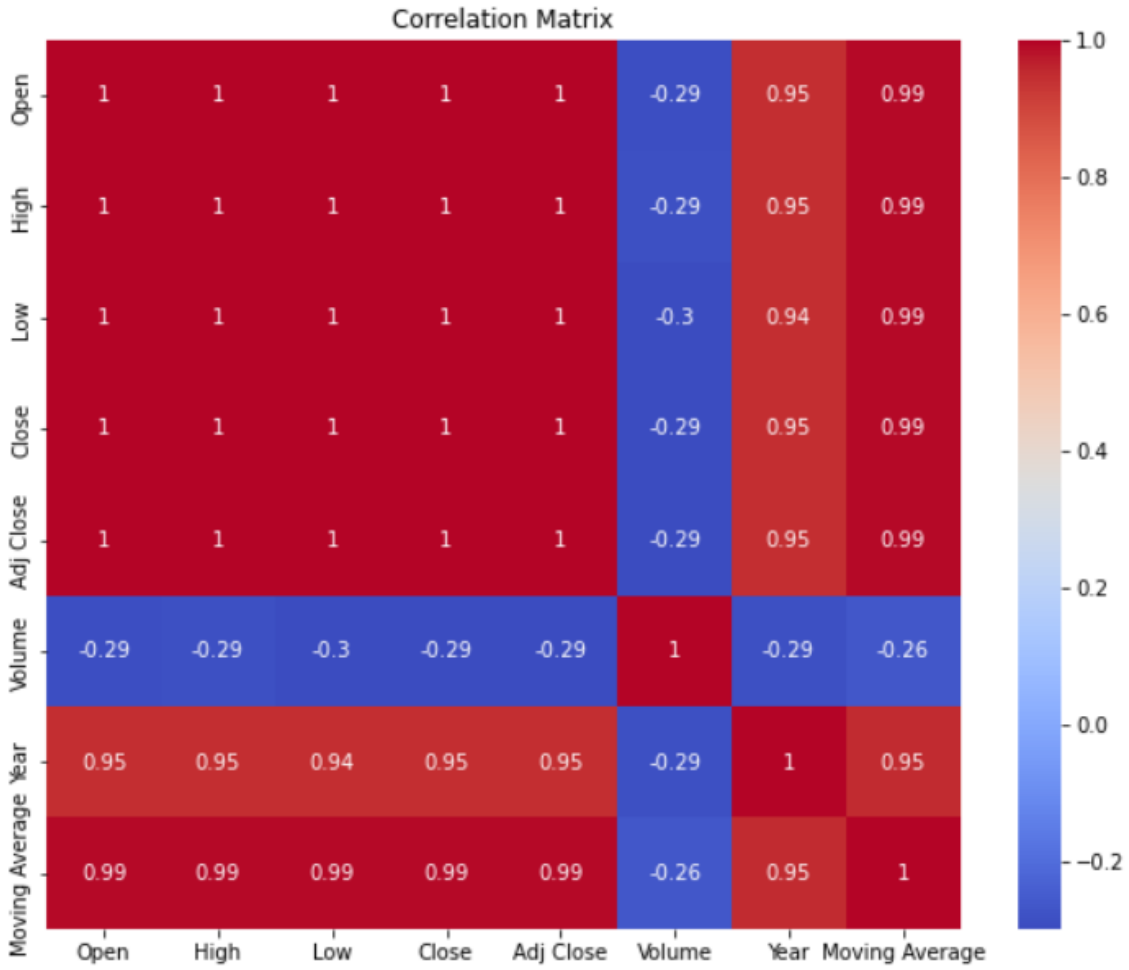


- **Tesla Correlation Matrix:**

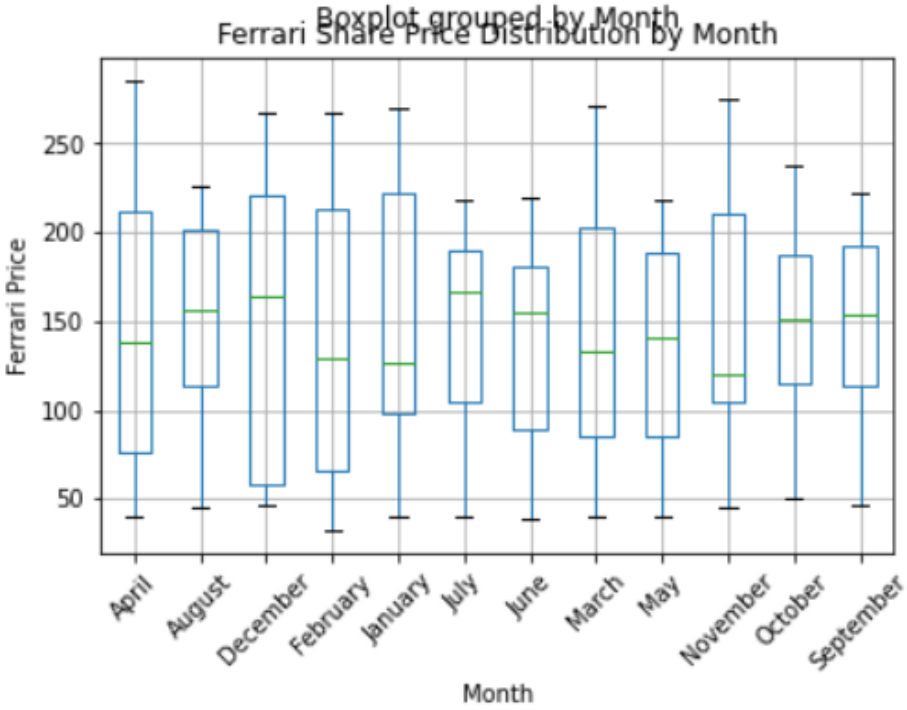
Insights: This heatmap or correlation matrix will provide insights into the correlation between numerical variables, highlighting the strength and direction of the relationships



- **Ferrari Correlation Matrix:**

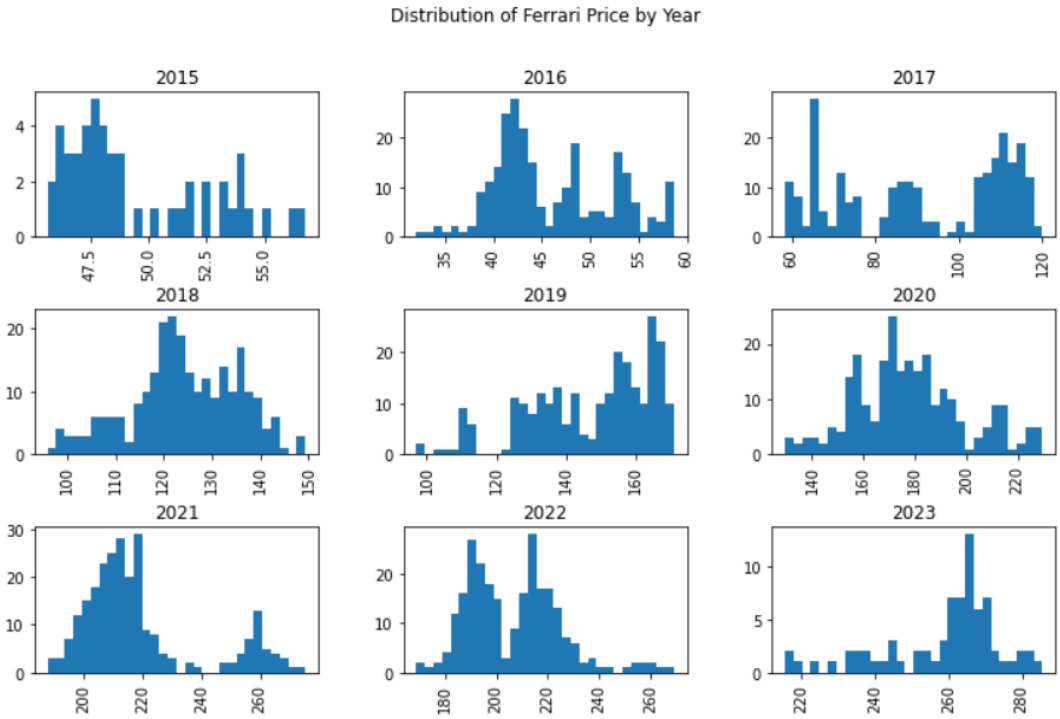


- **Ferrari Share Price Distributed by Month:**



- **Distribution of Tesla Share Price by Year:**

- **Distribution of Ferrari Share Price by Year:**



Feature Scaling

Now we have done feature engineering such as normalization

- **Normalization of Tesla Dataset:**

```
0       0.011331
1       0.010893
2       0.011920
3       0.011102
4       0.011537
          ...
1880    0.440373
1881    0.438126
1882    0.443221
1883    0.436402
1884    0.427111
Name: Close, Length: 1885, dtype: float64
```

- **Normalization of Ferrari Dataset:**

```
0       0.097622
1       0.096162
2       0.090798
3       0.086183
4       0.078373
          ...
1880    0.989784
1881    1.000000
1882    0.982290
1883    0.980555
1884    0.972784
Name: Close, Length: 1885, dtype: float64
```

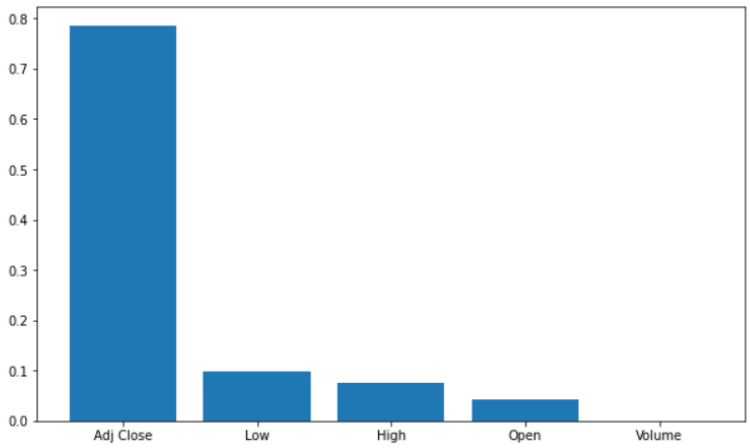
Feature Selection

Feature selection has been done to remove all the unwanted columns to get the precise columns for use.

- **Feature Selection of Ferrari Dataset:**

```
     Feature  Importance
3  Adj Close    0.785072
2        Low    0.097258
1       High    0.075381
0       Open    0.042277
4     Volume    0.000012
['Adj Close', 'Low', 'High', 'Open', 'Volume']
```

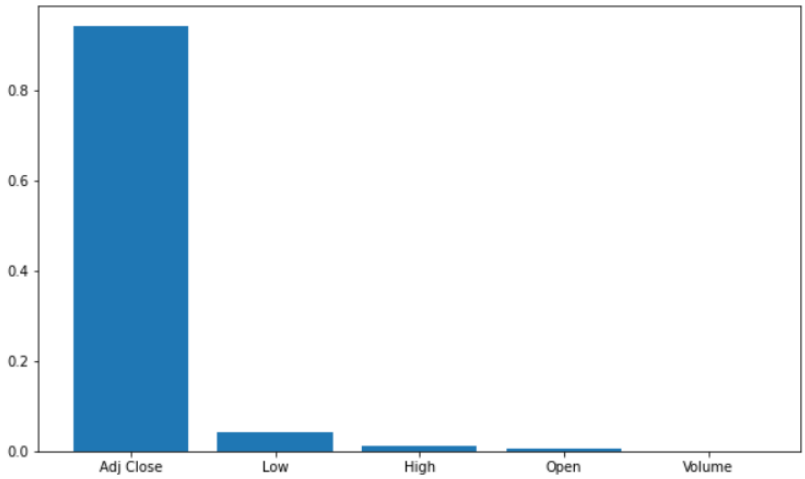
Out[140]: <BarContainer object of 5 artists>



- **Feature Selection of Tesla Dataset:**

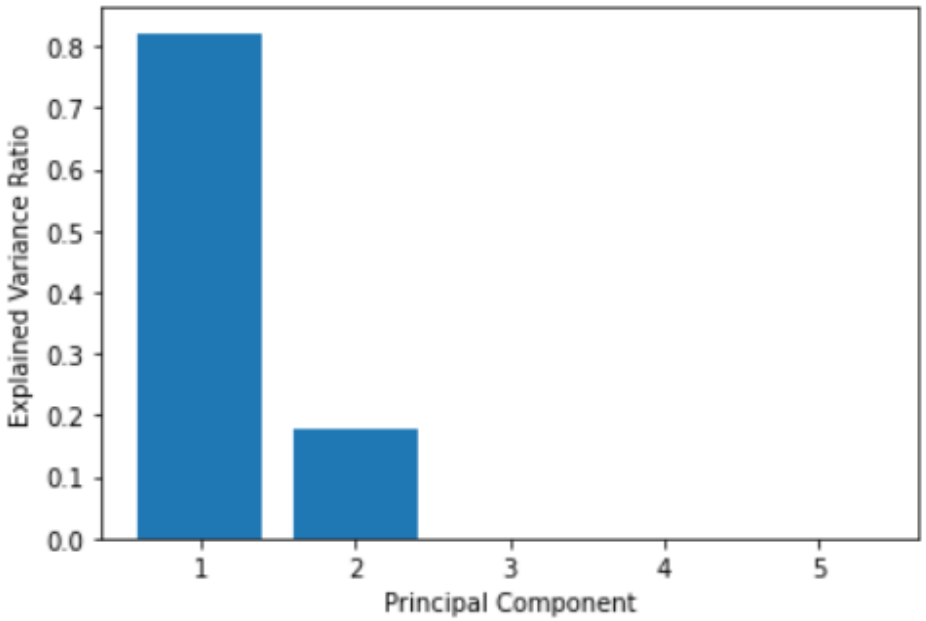
```
     Feature  Importance
3  Adj Close    0.940969
2        Low    0.041889
1       High    0.013024
0       Open    0.004114
4     Volume    0.000005
['Adj Close', 'Low', 'High', 'Open', 'Volume']
```

Out[3]: <BarContainer object of 5 artists>



Dimensionality Reduction

- **PCA of Tesla Dataset:**



- **PCA of Ferrari Dataset:**