



Analysis on Indian used car market



About me :
Golla.Tharun
B.Tech(CSE)-2023

Project Overview :

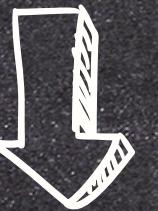
- Used cars sales are on global increase.
- Don't make a decision on buying a car based only on its price.
- Line up financing before you shop for your car to get better rates.

Problem Solution :

- A model can be developed to determine the worthiness of used car by studying variety of features like the number of used cars available for sale, the average price, and the most popular brands and models of used cars

Data Collection

- For real time analysis , data is scrapped from 'CarTrade.com website
- CarTrade website for used cars.



- Scrap the data using BeautifulSoup.



- Download url data.



- Save it to .csv file.

- Extract the features.



Data Set

- Obtained 25328 datapoints after scraping.
- It contains 8 columns and 3166 rows.

```
In [14]: df=pd.DataFrame(Data)
```

```
In [15]: df
```

```
Out[15]:
```

	Company	Model	Price	Variant	Km	Engine_type	Year	Location
0	Mercedes-Benz	E-Class	₹52,00,000	E	20,000 KMs	Diesel	2019	Pune
1	Mahindra	XUV500	₹8,95,000	W10	52,000 KMs	Diesel	2015	Delhi
2	Hyundai	i20	₹6,50,000	Active	49,000 KMs	Petrol	2015	Bangalore
3	Maruti	Suzuki	₹5,70,000	Swift	58,006 KMs	Petrol	2018	Delhi
4	Kia	Seltos	₹12,75,000	2022	14,937 KMs	Petrol	2021	Mumbai
...
3156	Tata	Altroz	₹8,19,000	XZ	23,360 KMs	Petrol	2021	Bangalore
3157	Maruti	Suzuki	₹6,79,000	Wagon	22,324 KMs	Petrol	2020	Bangalore
3158	Maruti	Suzuki	₹5,31,000	Wagon	15,363 KMs	Petrol	2020	Kolkata
3159	Maruti	Suzuki	₹9,79,000	Vitara	22,500 KMs	Petrol	2020	Mysore
3160	Hyundai	Grand	₹7,49,000	i10	17,242 KMs	Petrol	2020	Mysore

3161 rows × 8 columns

Data Cleaning

- Inside columns , details like currency symbol , kilo-meters are removed for each row.

After cleaning data and applying some important codes , Numerical and Categorical data is ready for Visualization

```
In [14]: df=pd.DataFrame(Data)
```

```
In [15]: df
```

Out[15]:

	Company	Model	Price	Variant	Km	Engine_type	Year	Location
0	Mercedes-Benz	E-Class	₹52,00,000	E	20,000 KMs	Diesel	2019	Pune
1	Mahindra	XUV500	₹8,95,000	W10	52,000 KMs	Diesel	2015	Delhi
2	Hyundai	i20	₹6,50,000	Active	49,000 KMs	Petrol	2015	Bangalore
3	Maruti	Suzuki	₹5,70,000	Swift	58,006 KMs	Petrol	2018	Delhi
4	Kia	Seltos	₹12,75,000	2022	14,937 KMs	Petrol	2021	Mumbai
...
3156	Tata	Altroz	₹8,19,000	XZ	23,360 KMs	Petrol	2021	Bangalore
3157	Maruti	Suzuki	₹6,79,000	Wagon	22,324 KMs	Petrol	2020	Bangalore
3158	Maruti	Suzuki	₹5,31,000	Wagon	15,363 KMs	Petrol	2020	Kolkata
3159	Maruti	Suzuki	₹9,79,000	Vitara	22,500 KMs	Petrol	2020	Mysore
3160	Hyundai	Grand i10	₹7,49,000	i10	17,242 KMs	Petrol	2020	Mysore

3161 rows × 8 columns

Data Cleaning

In [24]: df

Out[24]:

	Company	Model	Price	Variant	Km	Engine_type	Year	Location
0	Mercedes-Benz	E-Class	5200000	E	20000	Diesel	2019	Pune
1	Mahindra	XUV500	895000	W10	52000	Diesel	2015	Delhi
2	Hyundai	i20	650000	Active	49000	Petrol	2015	Bangalore
3	Maruti	Suzuki	570000	Swift	58006	Petrol	2018	Delhi
4	Kia	Seltos	1275000	2022	14937	Petrol	2021	Mumbai
...
3099	Volkswagen	Polo	285000	Comfortline	51609	Petrol	2010	Mumbai
3131	Kia	Sonet	1390000	GTX	45000	Diesel	2020	Madurai
3132	Tata	Safari	2075000	XZ	27000	Diesel	2021	Mumbai
3133	Honda	City	995000	ZX	34155	Petrol	2018	Mumbai
3134	Aston	Martin	8500000	Rapide	30000	Petrol	2011	Chennai

535 rows × 8 columns

1 df.info()

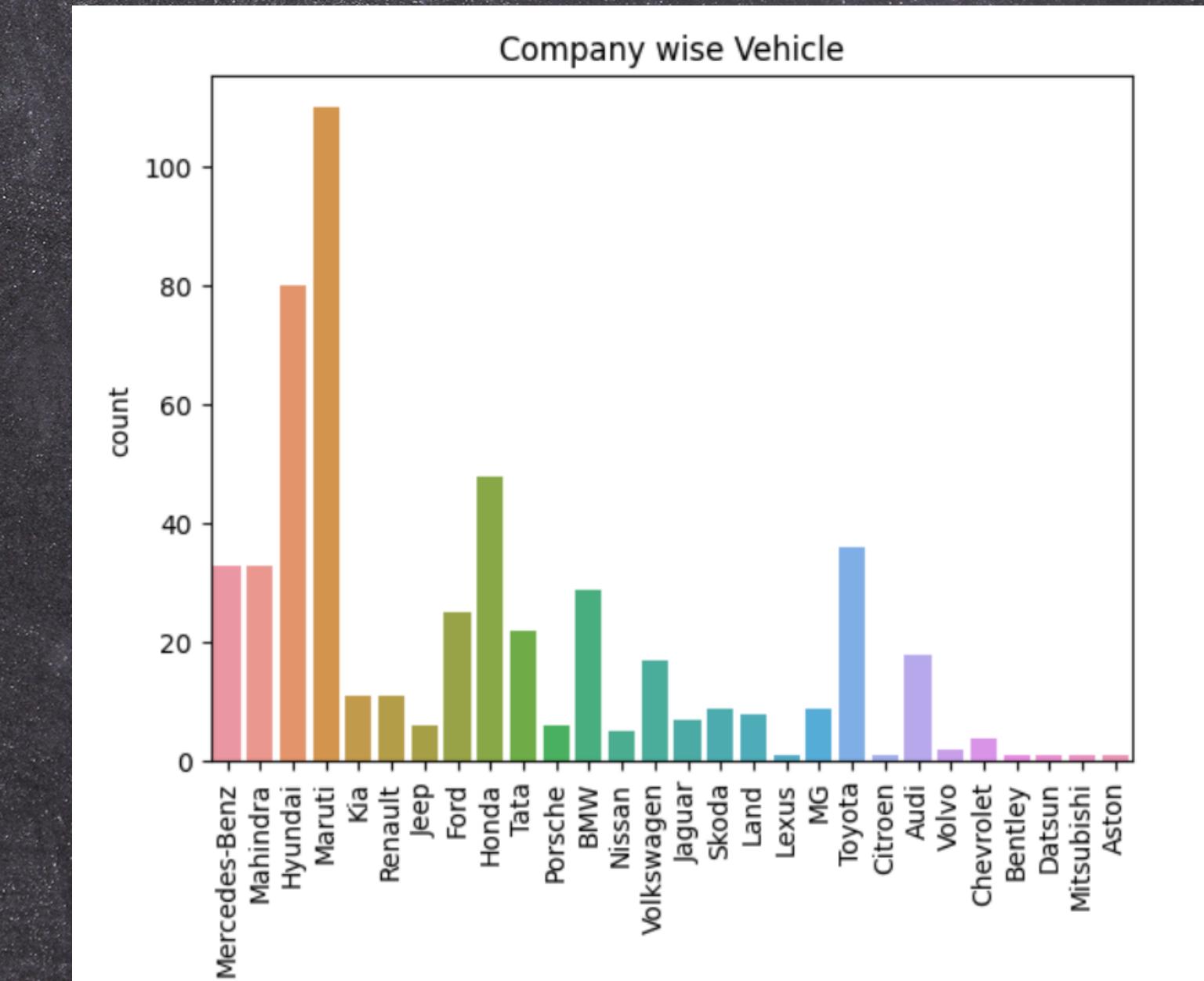
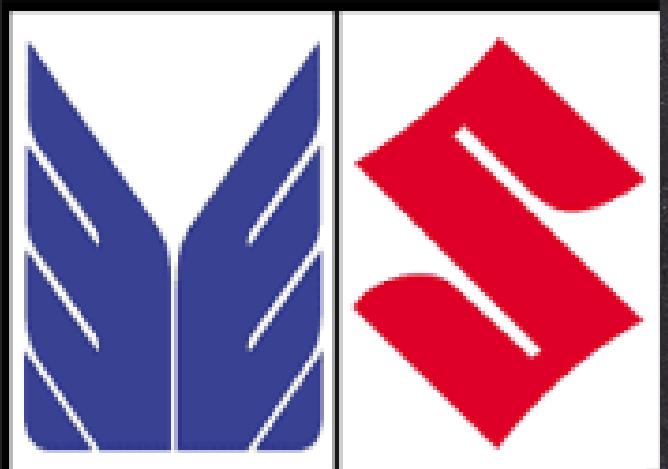
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3166 entries, 0 to 3165
Data columns (total 8 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   Company          3166 non-null   object 
 1   Model            3166 non-null   object 
 2   Variant          3166 non-null   object 
 3   Engine Type     3166 non-null   object 
 4   Year             3166 non-null   int32  
 5   Driven           3166 non-null   int32  
 6   Price            3166 non-null   int32  
 7   Location         3166 non-null   object 
dtypes: int32(3), object(5)
memory usage: 160.9+ KB
```

Data Visualization

Compare all the companies you are considering.

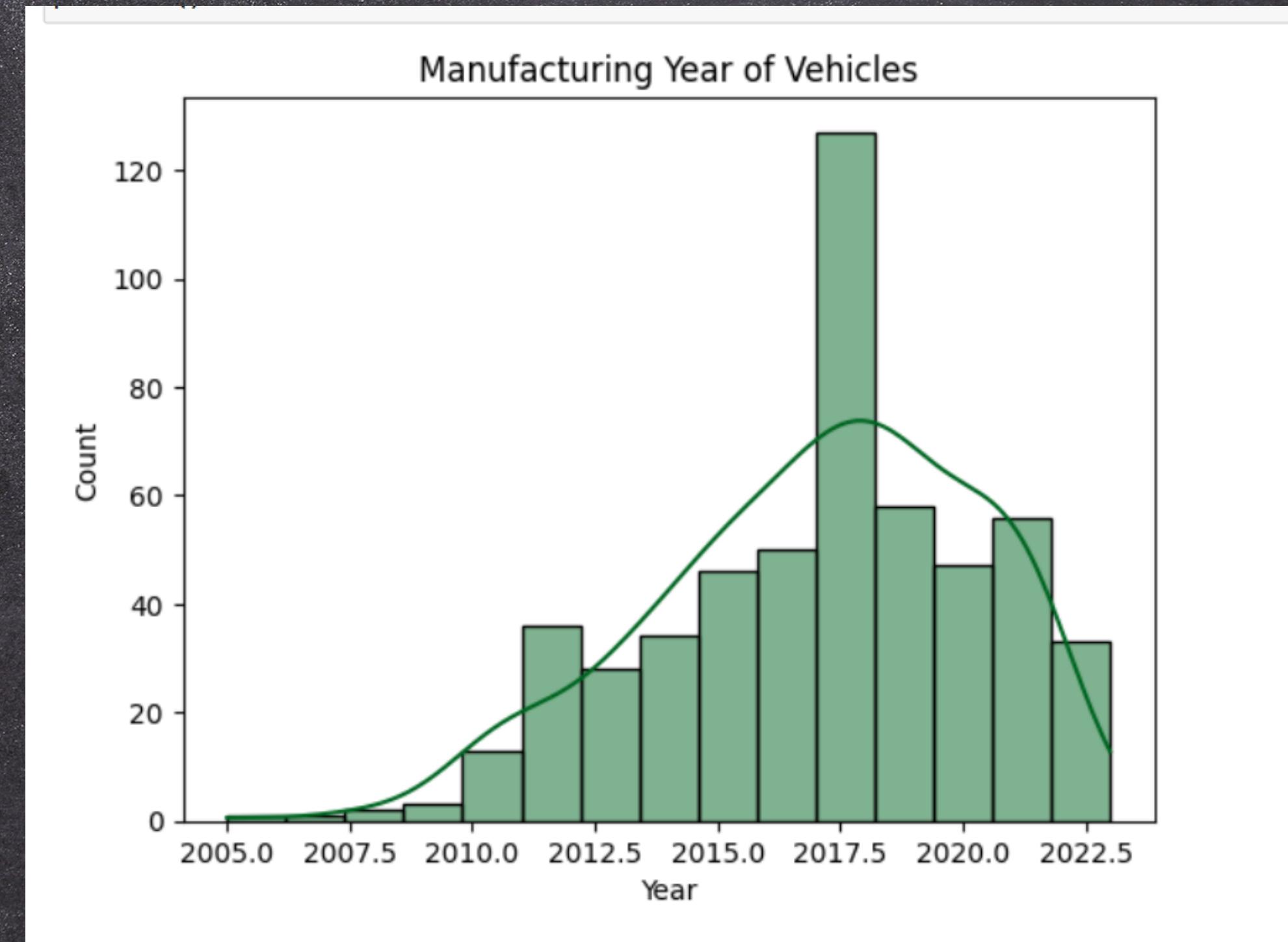
Top 2 Businesses Are :

1. Maruti
2. Hyundai



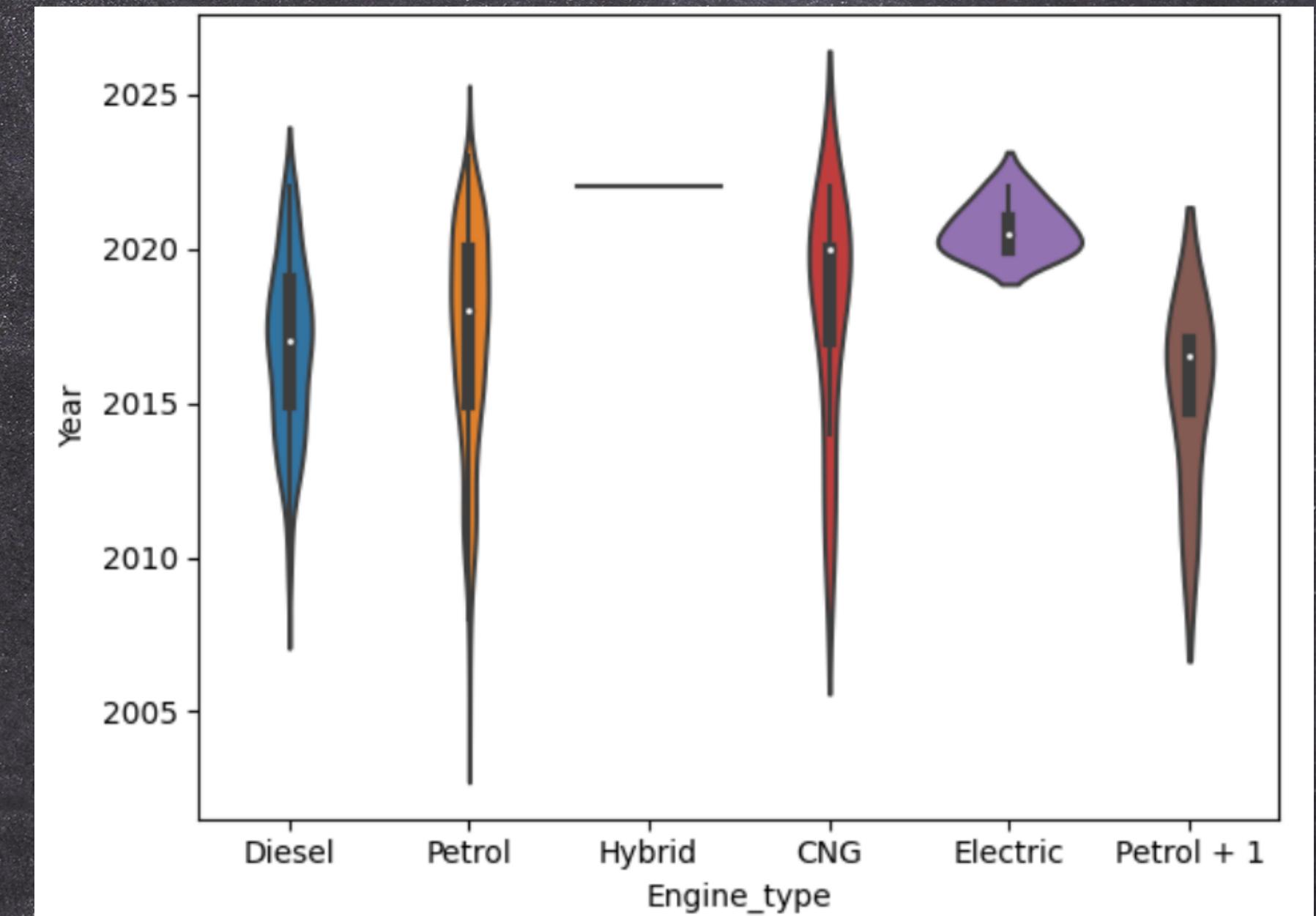
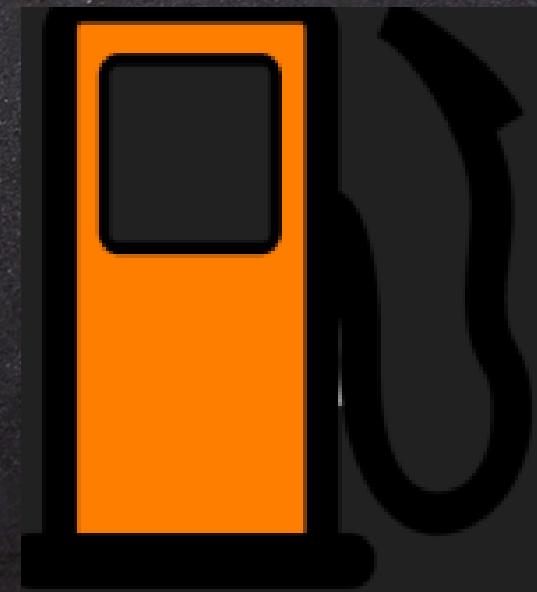
Analysis of the Most Manufacturing Years in a Row

2018 saw the most manufactured vehicles ever.



Identify differences in Engine types across time

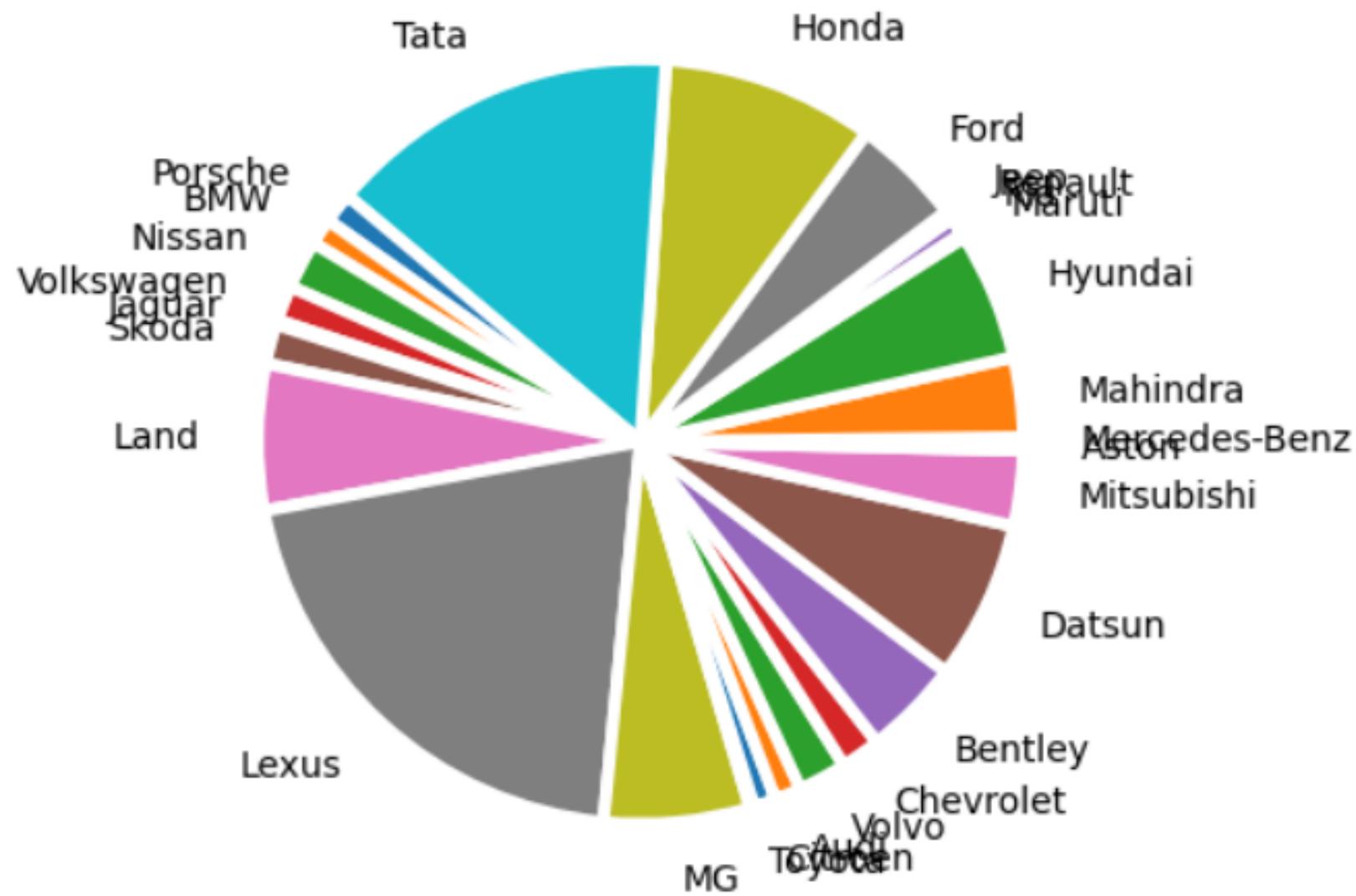
This graph compares the performance of petrol, diesel, and CNG engines.



- All Types of Companies.

In [47]:

```
# Company wise pie plot
plt.pie(dup_company, labels=df['Company'].unique(), labeldistance=1.15,
```

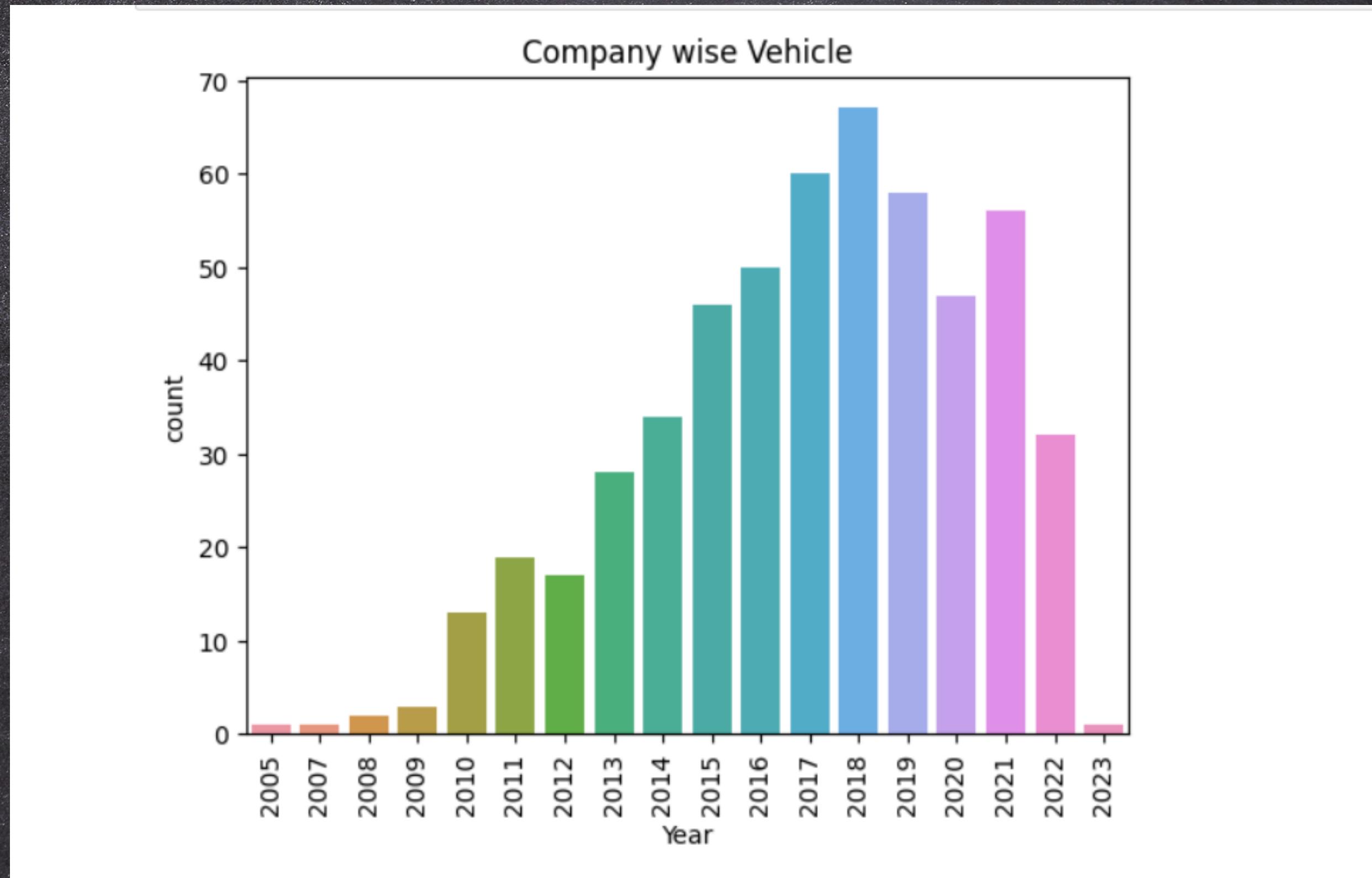


Engine Type Max count

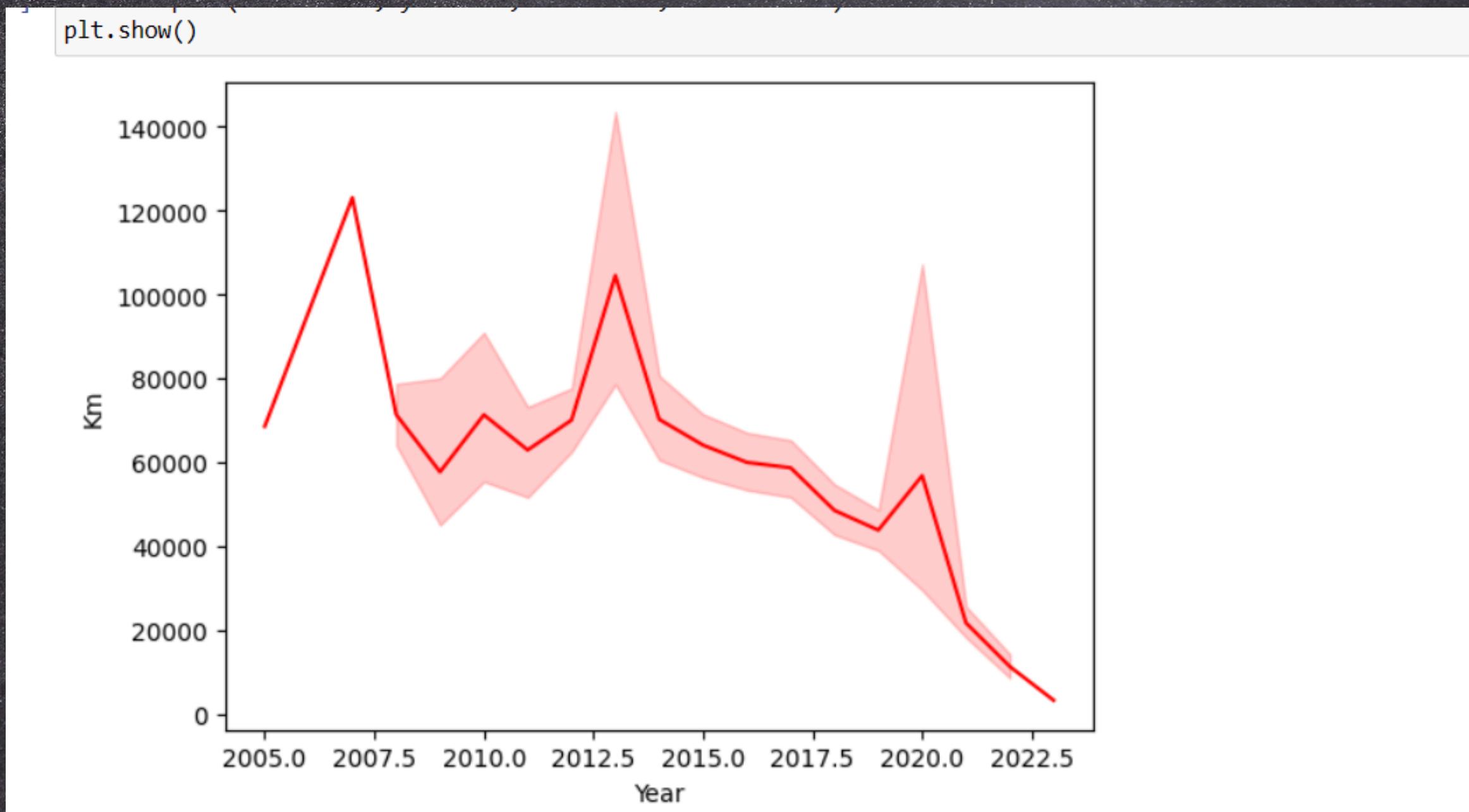
In [41]:

```
# Engine type and year wise
```

- Company wise Vehicle
- Year



- Over all Km and Year's



Libraries

- Numpy
- Pandas
- Matplot lib
- Sea born
- Beautiful Soup.



Libraries

Numpy: The name “Numpy” stands for “Numerical Python”. It is the commonly used library. It is a popular machine learning library that supports large matrices and multi-dimensional data.

Pandas : Pandas are an important library for data scientists. It is an open-source machine learning library that provides flexible high-level data structures and a variety of analysis tools. It eases data analysis, data manipulation, and cleaning of data

Matplotlib : This library is responsible for plotting numerical data. And that's why it is used in data analysis. It is also an open-source library and plots high-defined figures like pie charts, histograms, scatterplots, graphs, etc.

Libraries

Seaborn : (sns)

Seaborn is a library for making statistical graphics in Python. It builds on top of matplotlib and integrates closely with pandas data structures. Seaborn helps you explore and understand your data.

Challenges Faced

- Gathering Data
- 2.Analyzing the Data
- 3.Time management
- 4.Individual project



Conclusion

This project will provide a thorough understanding of the used car market using Python's Pandas DataFrame functionality, enabling stakeholders to decide wisely, spot market trends, and gain insightful knowledge into pricing dynamics, well liked models, and regional variations.



Thank You

