ABSTRACT

Electric vehicles (EVs) utilize electric motors powered by rechargeable batteries instead of internal combustion engines fueled by fossil fuels. They represent a greener alternative to conventional gas-powered vehicles, emitting fewer greenhouse gases and pollutants. EVs come in different forms, including Battery Electric Vehicles (BEVs), which rely solely on electric power; Hybrid Electric Vehicles (HEVs), incorporating both electric motors and gasoline engines; and Plug-in Hybrid Electric Vehicles (PHEVs), which possess larger batteries rechargeable via plugging in. Recent years have witnessed a surge in EV adoption driven by heightened environmental awareness, government incentives, and advancements in battery technology. However, challenges such as high upfront costs, limited range, and inadequate charging infrastructure persist, hindering widespread acceptance. This study proposes an analytical approach to delineate the various segments of India's EV market, scrutinizing influential factors such as environmental consciousness, fuel expenses, and governmental support. By categorizing potential buyers into distinct groups such as urban commuters, fleet operators, and luxury consumers, the research aims to assess the market potential and identify barriers to EV adoption in India. Insights gleaned from this segmentation analysis offer valuable guidance for policymakers and industry stakeholders striving to promote EV adoption in the Indian market

INTRODUCTION

Electric vehicles (EVs) have emerged as a pivotal component of the global automotive industry's transition towards sustainability. With their potential to reduce greenhouse gas emissions and dependence on fossil fuels, EVs are increasingly being considered as a viable alternative to traditional internal combustion engine vehicles. The adoption of EVs is particularly significant in countries like India, where urbanization, rising pollution levels, and energy security concerns underscore the need for clean transportation solutions.

Our project focuses on examining the electric vehicle market dynamics in India, leveraging insights derived from a comprehensive dataset obtained from [source]. This dataset comprises essential attributes related to electric vehicles, including selling price, kilometers driven, fuel type, seller type, transmission type, and owner details. By conducting exploratory data analysis (EDA) on this dataset, our objective is to gain valuable insights into the Indian EV market's landscape and identify key trends, challenges, and opportunities

Specifically, our project aims to achieve the following objectives:

1. Analyze the distribution and characteristics of electric vehicles in the Indian market based on the dataset's attributes.
2. Identify influential factors driving the adoption of EVs among different consumer segments in India.
3. Assess the impact of environmental concerns, fuel costs, and government incentives on EV purchase decisions, as reflected in the dataset.
4. Explore the potential for market segmentation within the Indian EV market based on demographic and psychographic factors present in the dataset.
5. Highlight actionable insights that can inform stakeholders, policymakers, and industry players about strategies to promote EV adoption in India, leveraging findings from the dataset analysis.

\

PROBLEM STATEMENT

Despite the growing momentum towards electric vehicle (EV) adoption globally, the Indian automotive market still faces several challenges hindering the widespread acceptance of EVs. High upfront costs, limited charging infrastructure, and range anxiety are among the prominent barriers that need to be addressed to accelerate the transition to electric mobility in India. To effectively navigate these challenges and promote the uptake of EVs, it is imperative to gain deeper insights into the dynamics of the Indian EV market.

Therefore, the problem statement of our project is to investigate the factors influencing the adoption of electric vehicles in the Indian market and identify strategies to overcome existing barriers. Specifically, we aim to address the following key questions:

1. What are the primary drivers and deterrents affecting consumers' decisions to purchase electric vehicles in India?
2. How do factors such as environmental concerns, fuel costs, government incentives, and technological advancements impact the adoption of EVs among different consumer segments?
3. What are the prevailing trends and patterns in the Indian EV market, and how can stakeholders leverage this information to enhance market penetration?
4. What are the challenges and opportunities associated with EV adoption in India, and what measures can be implemented to mitigate challenges and capitalize on opportunities?
5. How can market segmentation strategies based on demographic, psychographic, and behavioral factors help in effectively targeting and engaging diverse consumer segments within the Indian EV market?

By addressing these questions through comprehensive data analysis and research, our project aims to provide actionable insights that can inform policymakers, industry stakeholders, and EV manufacturers about strategies to foster the sustainable growth of the electric vehicle sector in India.

DATA:

In order to address the challenges and gain insights I have taken an open source data from the car dekho website which contains data about the sales of the cars from three big companies such as Maruti, Honda, Chevrelot across india. The data contains 1720 rows and 9 columns. The data set looks like the below image

```
In [2]: df=pd.read_csv("Car_Details_Car_Dekho.csv")
        df
```

Out[2]:

	Company	name	year	selling_price	km_driven	fuel	seller_type	transmission	owner
0	Honda	Honda Amaze VX i-DTEC	2014	450000	141000	Diesel	Individual	Manual	Second Owner
1	Honda	Honda Amaze VX i-DTEC	2014	450000	141000	Diesel	Individual	Manual	Second Owner
2	Honda	Honda City V MT	2018	930000	14500	Petrol	Dealer	Manual	First Owner
3	Honda	Honda City i DTEC V	2014	560000	74000	Diesel	Individual	Manual	Second Owner
4	Honda	Honda City i DTEC VX	2014	675000	90000	Diesel	Dealer	Manual	First Owner
...	...	...	...	...	...	...	...	...	...
1715	Maruti	Maruti 800 AC	2014	195000	75000	Petrol	Individual	Manual	Second Owner
1716	Maruti	Maruti Alto 800 Base	2015	155000	40000	Petrol	Individual	Manual	First Owner
1717	Maruti	Maruti Alto LXi	2000	65000	90000	Petrol	Individual	Manual	Second Owner
1718	Maruti	Maruti Ritz VDi	2012	225000	90000	Diesel	Individual	Manual	Second Owner
1719	Maruti	Maruti 800 AC BSIII	2009	110000	83000	Petrol	Individual	Manual	Second Owner

1720 rows × 9 columns

The following image shows the various type of fuel type cars that are sold around these three companies.

```
In [3]: df["fuel"].value_counts()

Out[3]: fuel
        Petrol     1096
        Diesel      578
        CNG          31
        LPG          15
        Name: count, dtype: int64
```

From the above image we can see that Petrol and Diesel cars are the most sold ones and when we plan to introduce the electric vehicle we should keep in mind that it should attract these two main category customers in order to achieve better result.

Columns:

```
In [5]: df.columns
```

```
Out[5]: Index(['Company', 'name', 'year', 'selling_price', 'km_driven', 'fuel',
               'seller_type', 'transmission', 'owner'],
              dtype='object')
```

The above image shows the 9 columns present in the data.

Descriptive Statistics:

```
In [6]: df.describe()
```

Out[6]:

	year	selling_price	km_driven
count	1720.000000	1.720000e+03	1720.000000
mean	2012.659884	3.595410e+05	63336.986047
std	4.618048	2.445697e+05	44724.303652
min	1992.000000	3.000000e+04	1000.000000
25%	2010.000000	1.600000e+05	35000.000000
50%	2014.000000	3.000000e+05	60000.000000
75%	2016.000000	5.099990e+05	80825.250000
max	2020.000000	1.800000e+06	806599.000000

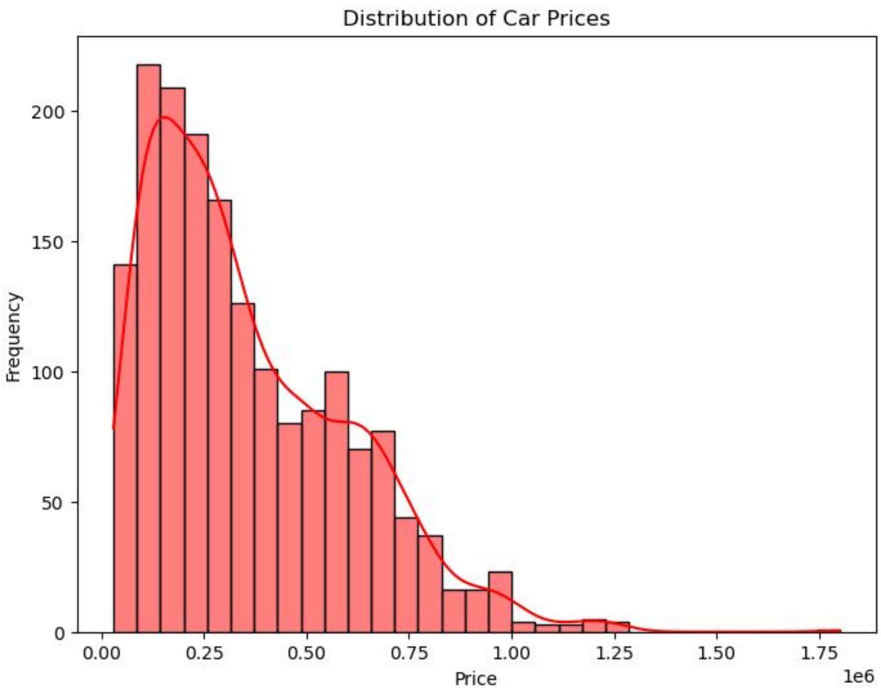
These three columns are the numerical columns present in the dataset

The above image describes the descriptive statistics which tells us about the things such as count, mean, standard deviation, minimum value, 25th percentile, 50th percentile, 75th percentile and finally the maximum value of the numerical data present in the dataset.

EXPLORATORY DATA ANALYSIS

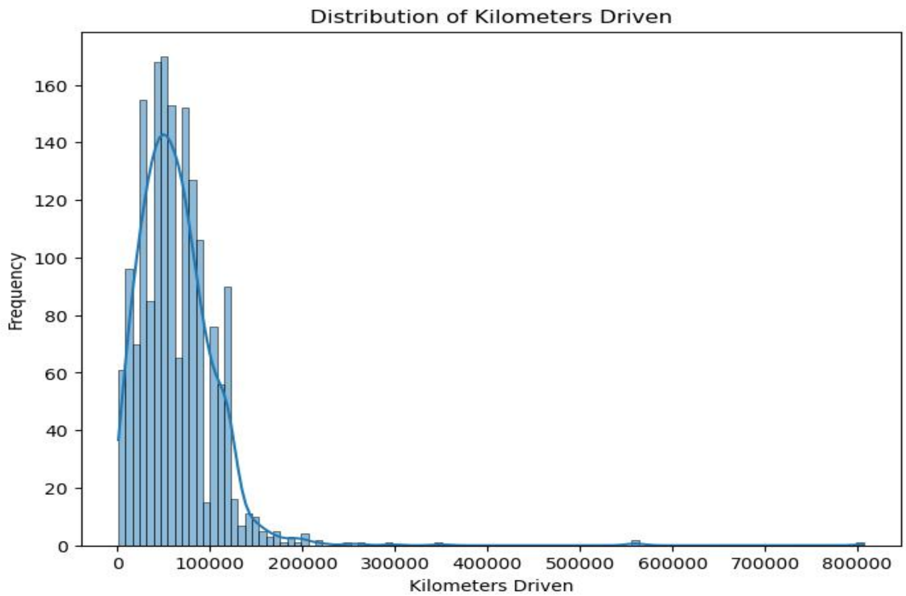
Exploratory Data Analysis (EDA) is a crucial step in data analysis that involves exploring and understanding the characteristics of a dataset. The primary goal of EDA is to uncover insights, patterns, and relationships within the data, which can inform subsequent analysis and decision-making processes. At first let us start by plotting a histogram to visualize the selling price of cars and identify their skewness using the kernel density estimation (kde).

In [7]:
```
plt.figure(figsize=(8, 6))
sns.histplot(df['selling_price'], kde=True,color="r")
plt.title('Distribution of Car Prices')
plt.xlabel('Price')
plt.ylabel('Frequency')
plt.show()
```



The histogram tells us that the selling price is left skewed as there are many cars with lower selling price.Now let us see another histogram which plots the kilometers driven by the car and also see its skewness.

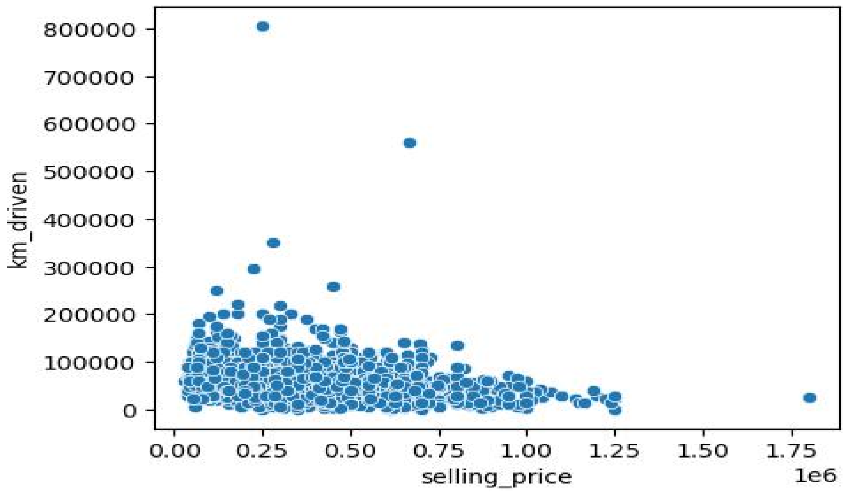
```
In [8]: plt.figure(figsize=(8, 6))
        sns.histplot(df['km_driven'], kde=True)
        plt.title('Distribution of Kilometers Driven')
        plt.xlabel('Kilometers Driven')
        plt.ylabel('Frequency')
        plt.show()
```



This histogram is also a left skewed which means the cars which have lesser kilometers driven are the ones which have been sold the most.

Selling Price vs Kilometers Driven :

```
In [27]: plt.figure(figsize=(5,4))
         sns.scatterplot(data=df,x="selling_price",y="km_driven")
         plt.show()
```

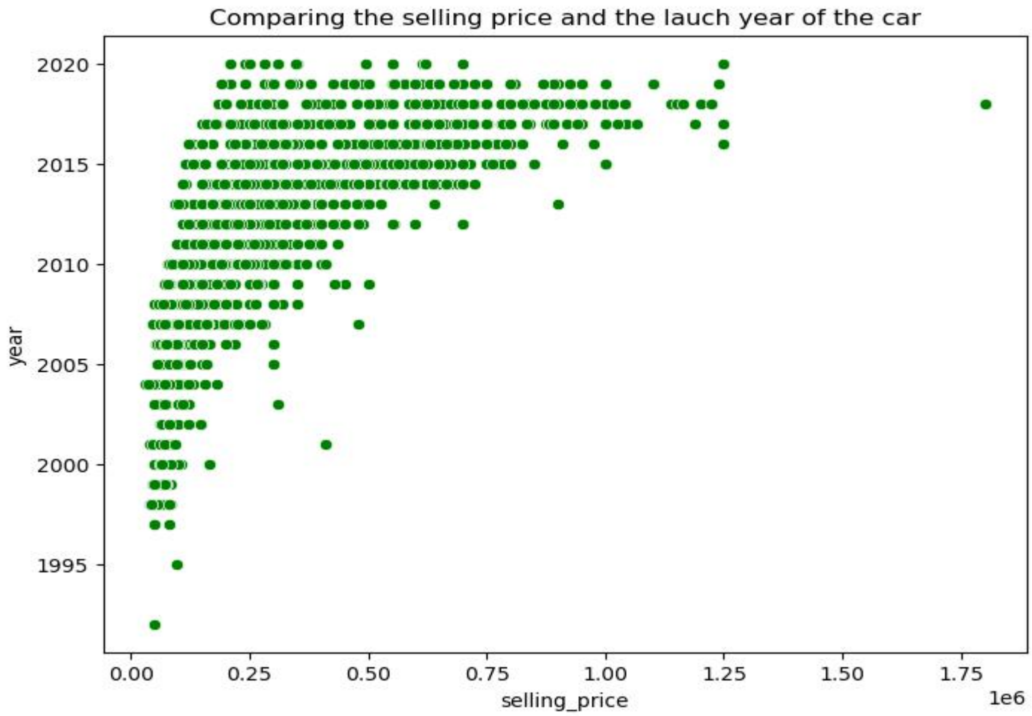


Scatterplot between the selling price and kilometers driven by the car.

.From the above scatterplot we can see that the cars that are with less km driven are easily sold and in the right bottom of the graph we can see one car with very low km has the highest selling price.

Selling Price Based on Year:

In [31]:
```
plt.figure(figsize=(8,6))
sns.scatterplot(x=df["selling_price"],y=df["year"],color="g")
plt.title("Comparing the selling price and the lauch year of the car")
plt.show()
```

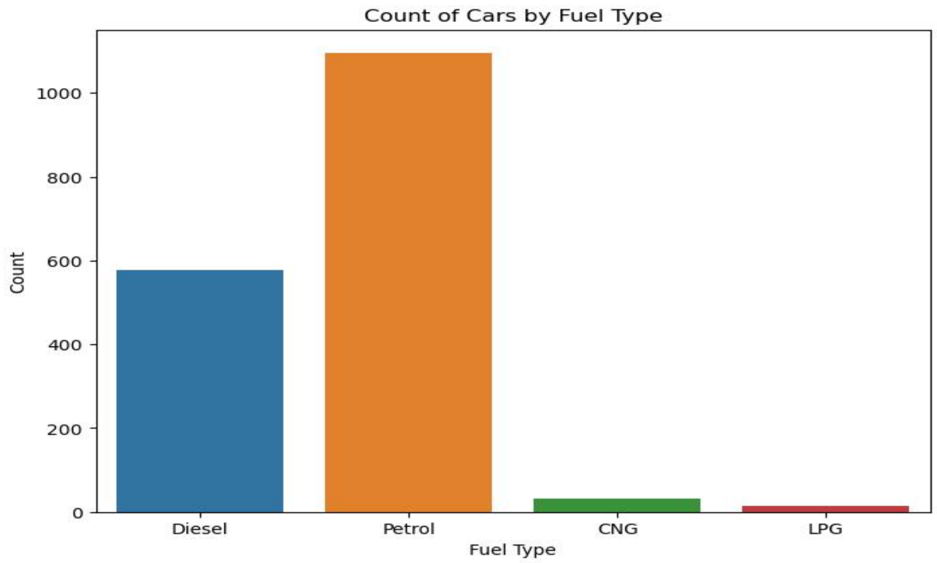


Scatterplot between the Selling price and the Year of manufacturing of the car.

From the above scatterplot we can see that cars model year after 2010 has the most number of sales and also with better price for different models and the older models have lesser selling price this maybe due to the kms driven by the cars.

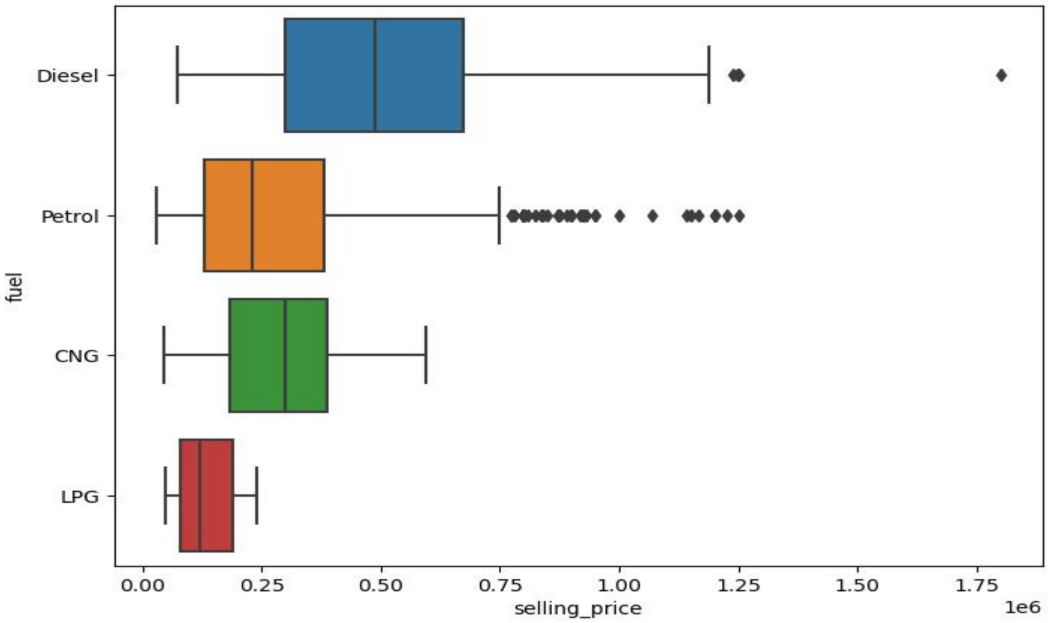
Fuel Type :

```
In [11]: plt.figure(figsize=(8, 6))
         sns.countplot(x='fuel', data=df)
         plt.title('Count of Cars by Fuel Type')
         plt.xlabel('Fuel Type')
         plt.ylabel('Count')
         plt.show()
```



Countplot for Fuel type of the cars sold.

```
In [33]: plt.figure(figsize=(8,6))
         sns.boxplot(data=df,x="selling_price",y="fuel")
         plt.show()
```

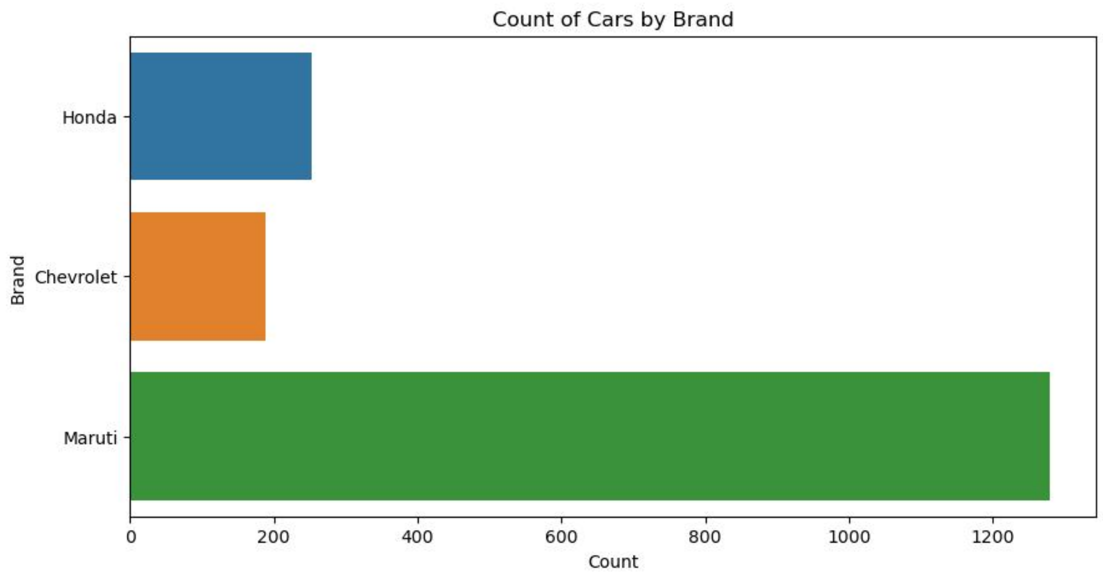


Boxplot showing the Selling price based on the Fuel type.

From the above boxplots we can see that there are more sales with the petrol cars and the diesel cars have the high selling price compared to others and is closely followed by petrol cars.

Company vs Selling Price :

In [34]:
```
plt.figure(figsize=(10, 5))
sns.countplot(y='Company', data=df)
plt.title('Count of Cars by Brand')
plt.xlabel('Count')
plt.ylabel('Brand')
plt.show()
```

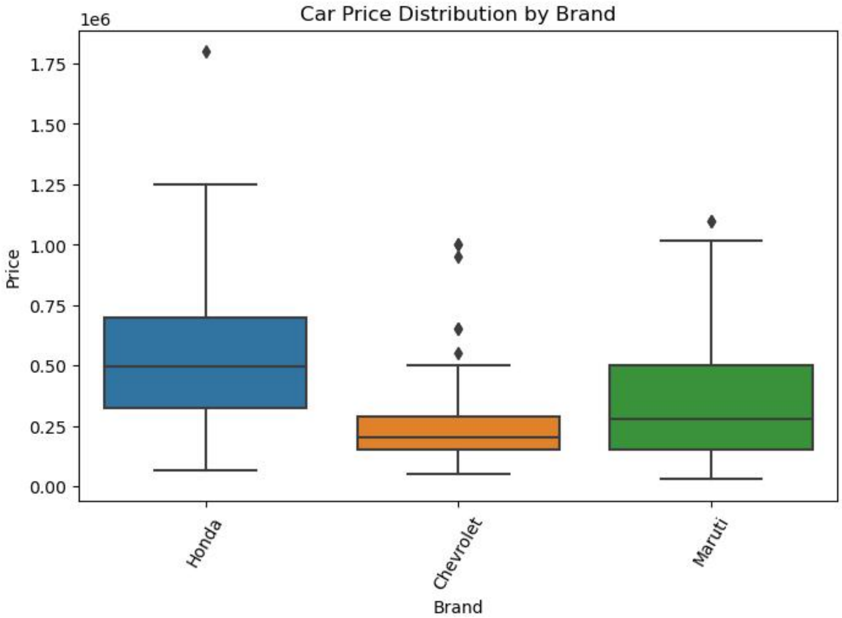


Countplot of the Companies and the number of cars sold on their brand

From this countplot we can infere that Maruti company has the highest number of car sales and Honda with the second place and chevrolet being the last.

This can be due to the price of the cars Maruti have more variety of models ranging from the most affordable to costlier ones. Hence it has the most number of customers.

```
In [35]: plt.figure(figsize=(8, 5))
sns.boxplot(x='Company', y='selling_price', data=df)
plt.title('Car Price Distribution by Brand')
plt.xlabel('Brand')
plt.ylabel('Price')
plt.xticks(rotation=60)
plt.show()
```



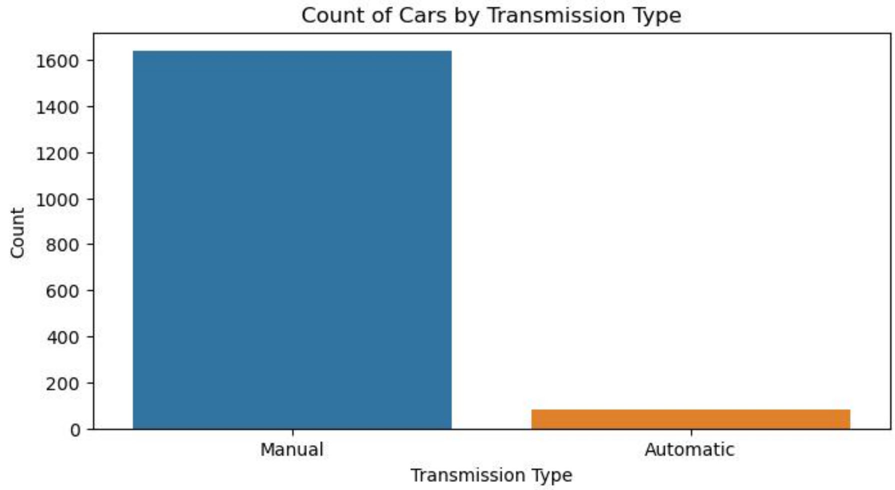
Boxplot for Company vs Selling price

In this we can see clearly why Maruti has the most sales it have wide variety of models in wide price range making it an option for more people where as Honda sitting at second might not be affordable to all because of it price. On the other hand Chevrelet is sitting at last due to its lack of variety in the models.

Transmission:

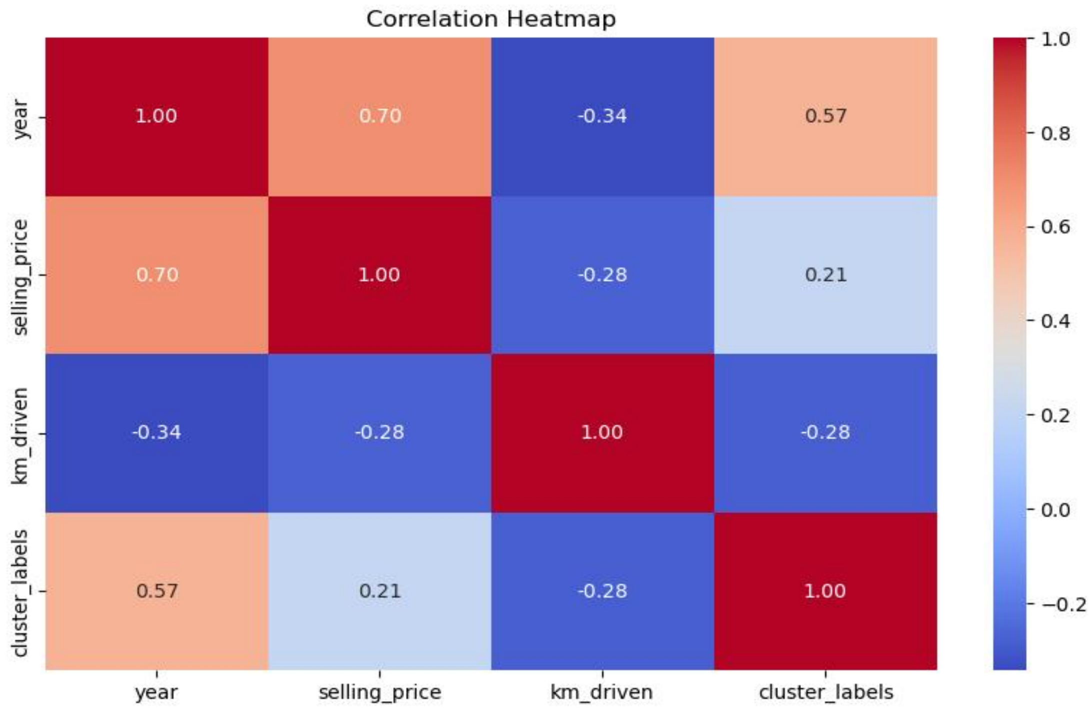
With the following countplot below we can see there are too many Manual cars and fewer automatic cars this shows how people still love the old manual transmission cars rather than having an automatic one. This is a crucial information as a EV car which is automatic may not receive as much attention as we expect in the company.

```
In [36]: plt.figure(figsize=(8, 4))
         sns.countplot(x='transmission', data=df)
         plt.title('Count of Cars by Transmission Type')
         plt.xlabel('Transmission Type')
         plt.ylabel('Count')
         plt.show()
```



Heatmap for Correlation :

```
In [37]: df_num=df.select_dtypes("int")
         plt.figure(figsize=(10, 6))
         sns.heatmap(df_num.corr(), annot=True, cmap='coolwarm', fmt=".2f")
         plt.title('Correlation Heatmap')
         plt.show()
```



Heatmap of numerical columns

Heatmaps are used to find the correlation between the numerical columns of the data to find how well are they corelated to each other. In this dataset there are four numerical data and we have plotted the heatmap to identify their correlation.

KMEANS CLUSTERING:

K-means clustering is a fundamental unsupervised machine learning algorithm designed to partition a dataset into a predefined number of clusters, aiming to group similar data points together while minimizing within-cluster variance, also known as inertia. This process begins by randomly initializing K cluster centroids, which serve as the initial cluster centers in the feature space. Subsequently, each data point is assigned to the nearest cluster centroid based on a chosen distance metric, typically Euclidean distance, thereby forming K clusters. The centroids of these clusters are then recalculated iteratively by computing the mean of all data points assigned to each cluster. This iterative process continues until convergence is achieved, characterized by minimal changes in cluster assignments or upon reaching a predetermined number of iterations. A critical consideration in employing K-means clustering is determining the appropriate number of clusters, which often involves employing techniques such as the elbow method or silhouette score. Moreover, the algorithm's sensitivity to the initial random initialization of centroids necessitates running multiple iterations with different initializations to select the solution with the lowest inertia. Widely applicable across various domains, K-means clustering finds utility in customer segmentation, image segmentation, anomaly detection, and market basket analysis, among other areas. In conclusion, K-means clustering serves as a versatile tool for data exploration, pattern recognition, and decision-making, empowering users to uncover meaningful insights from their datasets.

```
In [20]: from sklearn.cluster import KMeans
         kmeans = KMeans(n_clusters=3,random_state=32)
         kmeans.fit(df_scaled)
```

Out[20]:

```
KMeans
KMeans(n_clusters=3, random_state=32)
```

```
In [21]: cluster_labels = kmeans.labels_
```

```
In [39]: df["cluster_labels"] = cluster_labels
         df
```

Out[39]:

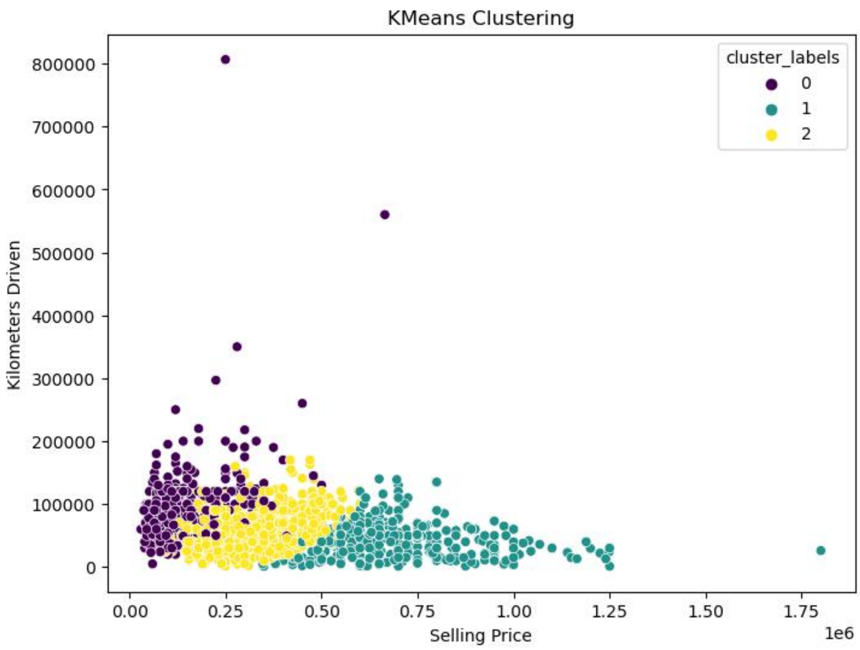
	Company	name	year	selling_price	km_driven	fuel	seller_type	transmission	owner	cluster_labels
0	Honda	Honda Amaze VX i-DTEC	2014	450000	141000	Diesel	Individual	Manual	Second Owner	2
1	Honda	Honda Amaze VX i-DTEC	2014	450000	141000	Diesel	Individual	Manual	Second Owner	2
2	Honda	Honda City V MT	2018	930000	14500	Petrol	Dealer	Manual	First Owner	1
3	Honda	Honda City i DTEC V	2014	560000	74000	Diesel	Individual	Manual	Second Owner	2
4	Honda	Honda City i DTEC VX	2014	675000	90000	Diesel	Dealer	Manual	First Owner	1
...	...	...	...	...	...	...	...	...	...	...
1715	Maruti	Maruti 800 AC	2014	195000	75000	Petrol	Individual	Manual	Second Owner	2
1716	Maruti	Maruti Alto 800 Base	2015	155000	40000	Petrol	Individual	Manual	First Owner	2
1717	Maruti	Maruti Alto LXi	2000	65000	90000	Petrol	Individual	Manual	Second Owner	0
1718	Maruti	Maruti Ritz VDi	2012	225000	90000	Diesel	Individual	Manual	Second Owner	2
1719	Maruti	Maruti 800 AC BSIII	2009	110000	83000	Petrol	Individual	Manual	Second Owner	0

1720 rows × 10 columns

Here I have made 3 clusters using the KMeans Clustering algorithm to partition the numerical data.

Visualising the clusters ¶

```
In [24]: plt.figure(figsize=(8, 6))
         sns.scatterplot(x='selling_price', y='km_driven', hue='cluster_labels', data=df, palette='viridis')
         plt.title('KMeans Clustering')
         plt.xlabel('Selling Price')
         plt.ylabel('Kilometers Driven')
         plt.show()
```



CONCLUSION

In conclusion, the integration of exploratory data analysis (EDA) and K-means clustering techniques has yielded a wealth of valuable insights into the intricate landscape of the car market. Through meticulous examination of the Car Dekho dataset, we have delved deep into the diverse facets of consumer behavior, preferences, and underlying market dynamics. By leveraging the power of data-driven analysis, we have unraveled distinct patterns and trends within the dataset, shedding light on the nuanced nuances of consumer decision-making processes.The segmentation of the market into three discernible clusters has not only provided clarity but has also offered a framework for understanding the heterogeneous preferences of consumers. Each cluster encapsulates a unique set of characteristics, reflecting the varied needs, aspirations, and priorities of different segments of the consumer base. From budget-conscious buyers seeking practicality and affordability to luxury enthusiasts inclined towards premium features and status symbols, the clusters paint a vivid picture of the diverse spectrum of consumer preferences within the automotive landscape.Moreover, the insights gleaned from the EDA and clustering analysis extend far beyond mere observation, offering actionable guidance for strategic planning and decision-making. Armed with a comprehensive understanding of consumer behavior and market dynamics, businesses are empowered to craft targeted marketing strategies, develop tailored product offerings, and optimize resource allocation to capitalize on emerging opportunities and mitigate potential risks.By leveraging these insights as a strategic compass, organizations can chart a course towards sustainable growth, enhanced customer satisfaction, and enduring success in an ever-evolving marketplace.