Analyzing the Impact of Car Features on Price and Profitability

Description:

The automotive industry has been rapidly evolving over the past few decades, with a growing focus on fuel efficiency, environmental sustainability, and technological innovation. With increasing competition among manufacturers and a changing consumer landscape, it has become more important than ever to understand the factors that drive consumer demand for cars.
In recent years, there has been a growing trend towards electric and hybrid vehicles and increased interest in alternative fuel sources such as hydrogen and natural gas. At the same time, traditional gasoline-powered cars remain dominant in the market, with varying fuel types and grades available to consumers.
For the given dataset, as a Data Analyst, the client has asked How can a car manufacturer optimize pricing and product development decisions to maximize profitability while meeting consumer demand?
This problem could be approached by analyzing the relationship between a car's features, market category, and pricing, and identifying which features and categories are most popular among consumers and most profitable for the manufacturer. By using data analysis techniques such as regression analysis and market segmentation, the manufacturer could develop a pricing strategy that balances consumer demand with profitability, and identify which product features to focus on in future product development efforts. This could help the manufacturer improve its competitiveness in the market and increase its profitability over time.

Tech stack used:

1. Microsoft Excel - Used for data preprocessing
2. Python - EDA

Data Description:

The dataset contains information on various car models and their specifications, and is titled "Car Features and MSRP". It was collected and made available on Kaggle by Cooper Union, a private college located in New York City.

Here is a brief overview of the dataset:

- **Number of observations:** 11,159
- **Number of variables:** 16
- **File type:** CSV (Comma Separated Values)

The variables in the dataset are:

- **Make:** the make or brand of the car
- **Model:** the specific model of the car
- **Year:** the year the car was released
- **Engine Fuel Type**: the type of fuel used by the car (gasoline, diesel, etc.)
- **Engine HP:** the horsepower of the car's engine
- **Engine Cylinders:** the number of cylinders in the car's engine
- **Transmission Type**: the type of transmission (automatic or manual)
- **Driven_Wheels:** the type of wheels driven by the car (front, rear, all)
- **Number of Doors:** the number of doors the car has
- **Market Category:** the market category the car belongs to (Luxury, Performance, etc.)
- **Vehicle Size:** the size of the car
- **Vehicle Style:** the style of the car (Sedan, Coupe, etc.)
- **Highway MPG:** the estimated miles per gallon the car gets on the highway
- **City MPG:** the estimated miles per gallon the car gets in the city
- **Popularity:** a ranking of the popularity of the car (based on the number of times it has been viewed on Edmunds.com)
- **MSRP:** the manufacturer's suggested retail price of the car

This dataset could be useful for a variety of data analysis tasks, such as:

- Exploring trends in car features and pricing over time
- Comparing the fuel efficiency of different types of cars
- Investigating the relationship between a car's features and its popularity
- Predicting the price of a car based on its features and market category

However, it's important to note that the dataset was last updated in 2017, so it may not reflect current trends or prices in the automotive industry.

Understanding the Dataset:

The dataset contains information on over 11,000 car models and their specifications, including details on the car's make, model, year, fuel type, engine power, transmission, wheels, number of doors, market category, size, style, estimated miles per gallon, popularity, and manufacturer's suggested retail price (MSRP).

A data analyst could use this dataset to gain insights into various aspects of the automotive industry, such as:

Analyzing trends in car features and pricing over time: By examining the variables in the dataset, a data analyst could identify how car features and prices have changed over time, which could help manufacturers make informed decisions about product development and pricing.

Comparing the fuel efficiency of different types of cars: By looking at the MPG variables in the dataset, a data analyst could compare the fuel efficiency of different types of cars and identify which types are the most efficient. This could help consumers make informed decisions about which car to purchase.

Investigating the relationship between a car's features and its popularity: By examining the popularity variable in the dataset, a data analyst could identify which features are most popular among consumers and how they affect a car's popularity. This could help manufacturers make informed decisions about product development and marketing.

Predicting the price of a car based on its features and market category: By using the various features and market category variables in the dataset, a data analyst could develop a model to predict the price of a car. This could help manufacturers and consumers understand how different features affect the price of a car and make informed decisions about pricing and purchasing.

Overall, this dataset could be a valuable resource for data analysts interested in exploring various aspects of the automotive industry and could provide insights that could inform decisions related to product development, marketing, and pricing.

Data Preprocessing: ***(Tech Stack Used: Ms-Excel)***

- **Handling duplicate values**:

  Found duplicate rows in the dataset during analysis. We have dropped all the duplicate rows except the first instance.

  After removal of all the duplicates from the dataset, these are total = 11200 records present in the dataset with 16 columns.

- **Handling Null values:**

  There are a total 108 null values present in the given dataset. We have calculated this by using Excel function =countblank() and applied this =countblank() to each column and calculated its sum and total = 108 as nulls from the dataset.

		Lincol	Zephyr	2006	regular unleaded	221	6	AUTOMATIC	front wheel drive	4	Luxury	Midsize	Sedan	26	17	61	28995	
		Make	Model	Year	Engine Fuel Type	Engine HP	Engine Cylinders	Transmission Type	Driven_Wheels	Number of Doors	Market Category	Vehicle Size	Vehicle Style	highway MPG	city mpg	Popularity	MSRP	Total Nulls
	No.of Null	0	0	0	3	69	30	0	0	6	0	0	0	0	0	0	0	108

1. As we can see there are 3 nulls present in the Engine fuel type column, which has details like Make = Suzuki, Model = Verona and Year = 2004. I have searched for the details on the internet and replaced it with correct values.
2. Ther are 69 nulls present in the column Engine HP. For null values from the column Engine HP we have searched for the value of the Engine HP by applying filter on Make, Model and Year, and replacing the nulls with the mode values of Engine HP. For example, for electric cars, we have searched with the car model, make and year and replaced the Engine HP with the correct values.
3. There are 30 null values in the column Engine Cylinders. For Engine Cylinders column where Engine Fuel type is electric we have replaced those places with 0 as electric cars dont have engine cylinders as we know.
4. Aprt from electric fuel type, there are only 2 car types. For which we have searched on the internet and updated those null fields with corrected values, where Mazda with regular unleaded engine fuel type has 2 cylinders whereas , Mazda with engine fuel type premium unleaded has engine cylinders 4.
5. There are 6 nulls in Number of doors column. And as we can see there are only 2 cars for which nulls are for this column, we have searched online about the information and corrected the fields.

Here we can see we have removed all the null values from all the columns, by using Ms-Excel.

	Acura	ZDX	2013	premium unleac	300	6	AUTOMATIC	all wheel drive	4	Crossover,Hatcht	Midsize	4dr Hatchbac	23	16	204	50920	
	Lincoln	Zephyr	2006	regular unleade	221	6	AUTOMATIC	front wheel dri	4	Luxury	Midsize	Sedan	26	17	61	28995	
	Make	Model	Year	Engine Fuel Type	Engine HP	Engine Cylinders	Transmission Type	Driven_Wheels	Number of Doors	Market Category	Vehicle Size	Vehicle Style	highway MPG	city mpg	Popularity	MSRP	Total Nulls
No.of Null	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

- **Handling Errors:**

1. In the column ‘Model’ with Make = Saad there are some errors with the column. Whereas Saad has model type **Saad 9 - 3** except this we have entry as **9 March 2023** and with model type **Saad 9 - 5** we have entry **9 May 2023.**
2. Column Transmission type has some entries with UNKNOWN. We have replaced these entries by searching the details on the internet with corrected details.
3. For Null values in Market Category column,
    - — Separated the Categories into different columns.
    - — Searched for Market Categories of all rows of the same Car Make and Model in non null values dataframe.
    - — Found mode of the Market Categories and replaced null value with the mode.

Note: We have used Ms-Excel up to here for Data Preprocessing.

- **<u>Handling Outliers</u>**: ***(Tech Stack Used: Python)***

  From here, we will be using a Python tool for removal of outliers. By importing above processed data into the Jupyter notebook by using Python library Pandas we have started our further process.

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

```
df= pd.read_excel(r'C:\Users\Fahim\Downloads\Dataset.xlsx')
```

```
df.head()
```

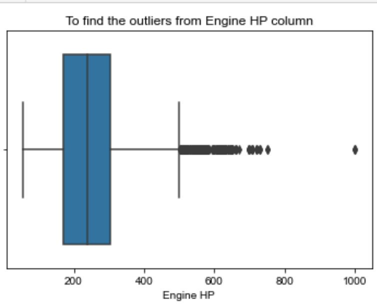
]:

	Make	Model	Year	Engine Fuel Type	Engine HP	Engine Cylinders	Transmission Type	Driven_Wheels	Number of Doors	Market Category	Vehicle Size	Vehicle Style	highway MPG	city mpg	Popul
0	BMW	1 Series M	2011	premium unleaded (required)	335	6	MANUAL	rear wheel drive	2	Factory Tuner,Luxury,High-Performance	Compact	Coupe	26	19	
1	BMW	1 Series	2011	premium unleaded (required)	300	6	MANUAL	rear wheel drive	2	Luxury,Performance	Compact	Convertible	28	19	
2	BMW	1 Series	2011	premium unleaded (required)	300	6	MANUAL	rear wheel drive	2	Luxury,High-Performance	Compact	Coupe	28	20	
3	BMW	1 Series	2011	premium unleaded (required)	230	6	MANUAL	rear wheel drive	2	Luxury,Performance	Compact	Coupe	28	18	
4	BMW	1 Series	2011	premium unleaded (required)	230	6	MANUAL	rear wheel drive	2	Luxury	Compact	Convertible	28	18	

1. For the outliers in the Engine HP column, we have verified it by plotting the box plot and we found some outliers in this column.

In [31]:

```
plt.title("To find the outliers from Engine HP column")
sns.boxplot(x=df['Engine HP'])
sns.set_style('whitegrid')
plt.figure(figsize=(12,20))
plt.show()
```



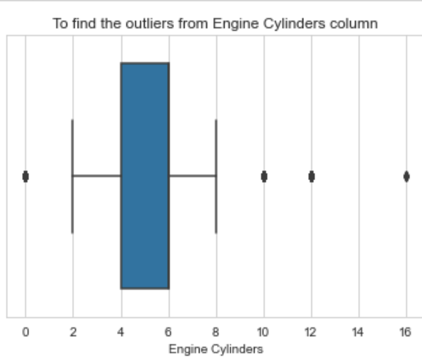
But, as we can see there are some cars which are Exotic and with High Performance. So, we did not change anything in this column.

```
# Check for the outliers:

df[df['Engine HP']>520]['Market Category'].value_counts()
```

```
7]: Exotic,High-Performance                                   109
    Factory Tuner,Luxury,High-Performance                      73
    Factory Tuner,High-Performance                             41
    Luxury,High-Performance                                    33
    Exotic,Factory Tuner,Luxury,High-Performance               29
    Exotic,Luxury,High-Performance                             21
    Crossover,Factory Tuner,Luxury,High-Performance            16
    Exotic,Flex Fuel,Factory Tuner,Luxury,High-Performance     13
    Exotic,Luxury,Performance                                  13
    Exotic,Flex Fuel,Luxury,High-Performance                   11
    Exotic,Factory Tuner,High-Performance                      11
    Exotic,Luxury                                               6
    High-Performance                                            6
    Crossover,Luxury,High-Performance                           2
    Factory Tuner,Luxury                                        2
    Flex Fuel,Factory Tuner,Luxury,High-Performance             1
    Exotic,Luxury,High-Performance,Hybrid                       1
    Name: Market Category, dtype: int64
```

2. For the column Engine Cylinders, we have applied the same logic. We have verified the outliers by plotting the box plot for the column Engine Cylinders.

```
plt.title("To find the outliers from Engine Cylinders column")
sns.boxplot(x=df['Engine Cylinders'])
sns.set_style('whitegrid')
plt.figure(figsize=(12,20))
plt.show()
```



We can see that there are outliers in this column which have Engine Cylinder values below 0 and above 8.

```
In [37]:  1 df[df['Engine Cylinders']<1]['Engine Fuel Type'].value_counts()
Out[37]: electric    66
         Name: Engine Fuel Type, dtype: int64
```

We have cross checked the values with the Engine Fuel type column and Market Category column, where cars having Engine Cylinder values below 0 are all electric cars and above 8 are all High Performance and Exotic cars

```
  1 df[df['Engine Cylinders']>10]['Engine Fuel Type'].value_counts()
35]: premium unleaded (required)                    195
     flex-fuel (premium unleaded required/E85)      24
     regular unleaded                               8
     premium unleaded (recommended)                 4
     Name: Engine Fuel Type, dtype: int64
```

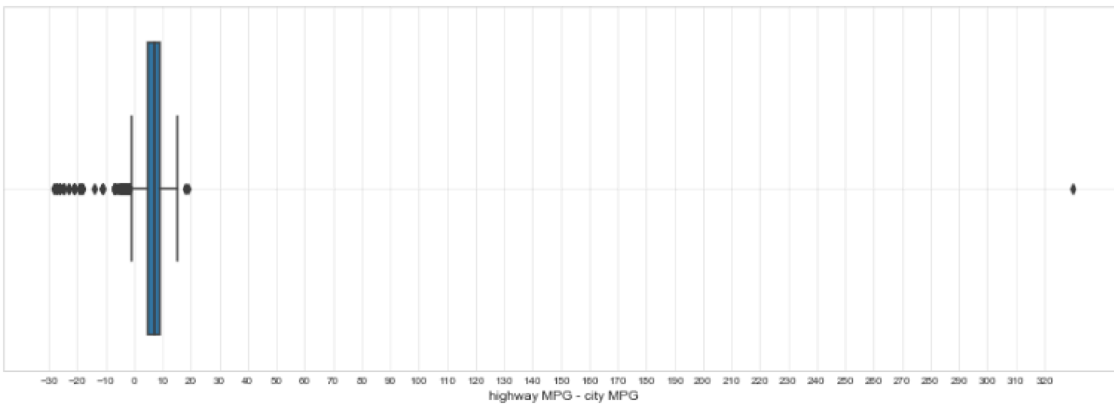
```
  1 df[df['Engine Cylinders']>8]['Market Category'].value_counts()
: Exotic,High-Performance                                   141
  Luxury,High-Performance                                    25
  Exotic,Luxury,Performance                                  25
  Exotic,Luxury,High-Performance                             19
  Exotic,Factory Tuner,High-Performance                      19
  Exotic,Flex Fuel,Factory Tuner,Luxury,High-Performance     13
  Factory Tuner,Luxury,High-Performance                      13
  Exotic,Flex Fuel,Luxury,High-Performance                   11
  Exotic,Luxury                                              10
  Luxury,Performance                                         10
  Exotic,Factory Tuner,Luxury,High-Performance                5
  Luxury                                                      3
  Factory Tuner,Luxury,Performance                            1
  Crossover,Luxury,Diesel                                     1
  Name: Market Category, dtype: int64
```

So, we didn’t take any action on these outliers from Engine Cylinders.

3. For the Outliers in highway MPG and city mpg columns, we plotted a box plot of difference values between the two columns.

```
fig, axs = plt.subplots(1, 1,figsize=(15, 5))
sns.boxplot(df['highway MPG']-df['city mpg'], orient='h', ax=axs)
axs.set_xticks(np.arange(-30,330,10))
axs.tick_params(axis='x',pad=5,length=0, labelsize=8)
axs.set_xlabel('highway MPG - city MPG')
axs.grid(axis='y',alpha=0.4)
axs.grid(axis='x',alpha=0.4)
axs.set_axisbelow(True)
plt.show()
```

```
C:\Users\Fahim\anaconda3\lib\site-packages\seaborn\_decorators.py:36: FutureWarning: Pass the following variable as a ke
d arg: x. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an e
cit keyword will result in an error or misinterpretation.
  warnings.warn(
```



We observed that there are few negative values of differences. Generally Mileage in Highway is more than that of City.

```
d = df['highway MPG']-df['city mpg']
d
```

So we replaced the highway MPG of such rows with the corresponding city mpg. Also, there were few rows where highway MPG is a lot more than city mpg. On further investigation, we found that these are not correct values. So we searched online for the correct Highway and City mileages and replaced them with the correct values.

```
df['highway MPG'].loc[df[d<0].index] = df['city mpg'].loc[df[d<0].index]
```

```
C:\Users\Fahim\anaconda3\lib\site-packages\pandas\core\indexing.py:670: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  iloc._setitem_with_indexer(indexer, value)
```

```
df[d>15]
```

]:

	Make	Model	Year	Engine Fuel Type	Engine HP	Engine Cylinders	Transmission Type	Driven_Wheels	Number of Doors	Market Category	Vehicle Size	Vehicle Style	highway MPG
1083	Audi	A6	2017	premium unleaded (recommended)	252	4	AUTOMATED_MANUAL	front wheel drive	4	Luxury	Midsize	Sedan	354
2990	Chevrolet	Cruze	2015	diesel	148	4	AUTOMATIC	front wheel drive	4	Diesel	Midsize	Sedan	46
4030	Lotus	Evora 400	2017	premium unleaded (required)	400	6	MANUAL	rear wheel drive	2	Exotic,High-Performance	Compact	Coupe	39

So we searched online for the correct Highway and City mileages and replaced them with the correct values.

```
df['highway MPG'].iloc[1083]=35
df['highway MPG'].iloc[2990]=40
df['highway MPG'].iloc[4030]=24
```

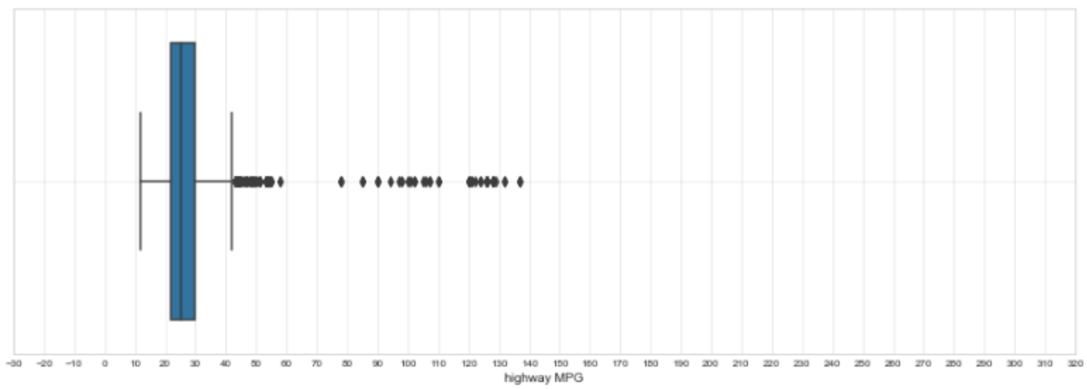
```
C:\Users\Fahim\anaconda3\lib\site-packages\pandas\core\indexing.py:670: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  iloc._setitem_with_indexer(indexer, value)
```

4 .For the outliers from the highway mpg column, we have plotted box plot distplot. If we consider 40 as threshold value, then there is no need to change anything in this column. Because electric cars have a high percentage of mileage.

Checks on Boxplot:

```python
fig, axs = plt.subplots(1, 1,figsize=(15, 5))
sns.boxplot(df['highway MPG'], orient='h', ax=axs)
axs.set_xticks(np.arange(-30,330,10))
axs.tick_params(axis='x',pad=5,length=0, labelsize=8)
axs.set_xlabel('highway MPG')
axs.grid(axis='y',alpha=0.4)
axs.grid(axis='x',alpha=0.4)
axs.set_axisbelow(True)
plt.show()
```

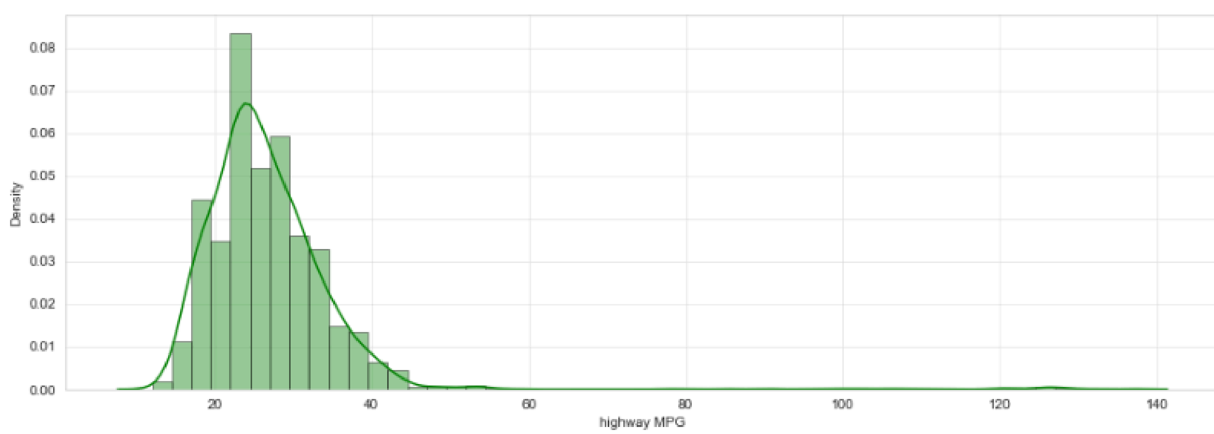
```
C:\Users\Fahim\anaconda3\lib\site-packages\seaborn\_decorators.py:36: FutureWarning: Pass the following variable as a keywor
d arg: x. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an expli
cit keyword will result in an error or misinterpretation.
  warnings.warn(
```



Checks of distplot:

```python
fig, axs = plt.subplots(1, 1,figsize=(15, 5))
sns.distplot(a=df['highway MPG'], color='green',
             hist_kws={"edgecolor": 'black'})
axs.grid(axis='y',alpha=0.4)
axs.grid(axis='x',alpha=0.4)
axs.set_axisbelow(True)
plt.show()
```

```
C:\Users\Fahim\anaconda3\lib\site-packages\seaborn\distributions.py:2551: FutureWarning: `distplot` is a deprecated 
and will be removed in a future version. Please adapt your code to use either `displot` (a figure-level function wit
r flexibility) or `histplot` (an axes-level function for histograms).
  warnings.warn(msg, FutureWarning)
```

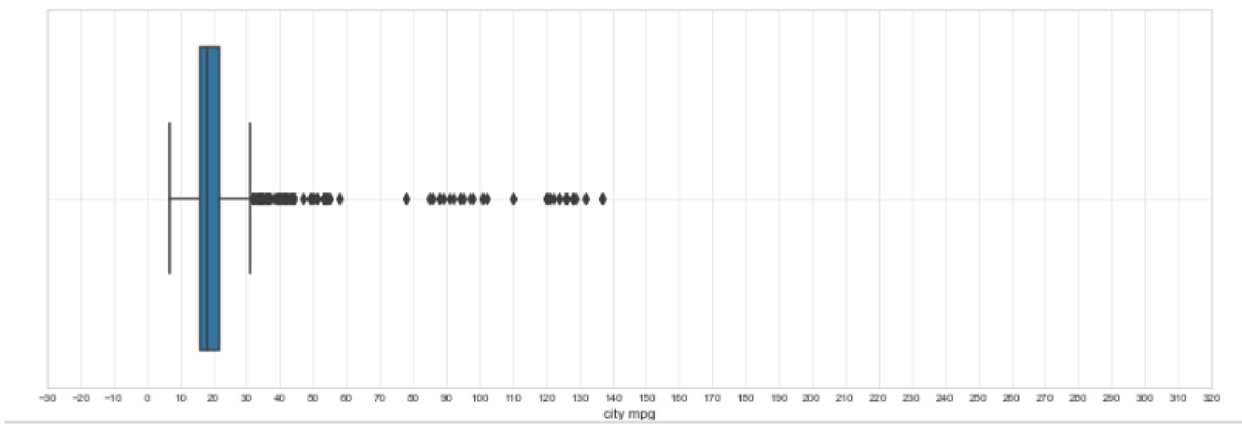


5. Same for the city mpg column. We verified outliers by plotting boxplot and distplot for this column as well. If we consider 28 as threshold value, we observed that a large number of vehicles have high city mileage as well and those vehicles must be electric vehicles.

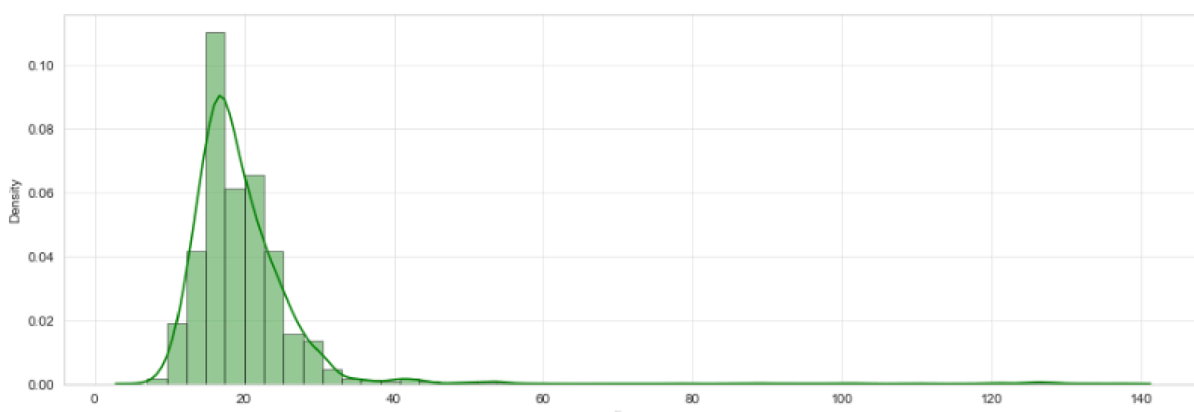
Check on boxplot:

```
fig, axs = plt.subplots(1, 1,figsize=(15, 5))
sns.boxplot(df['city mpg'], orient='h', ax=axs)
axs.set_xticks(np.arange(-30,330,10))
axs.tick_params(axis='x',pad=5,length=0, labelsize=8)
axs.set_xlabel('city mpg')
axs.grid(axis='y',alpha=0.4)
axs.grid(axis='x',alpha=0.4)
axs.set_axisbelow(True)
plt.show()
```

```
C:\Users\Fahim\anaconda3\lib\site-packages\seaborn\_decorators.py:36: FutureWarning: Pass the following variable as
d arg: x. From version 0.12, the only valid positional argument will be `data`, and passing other arguments withou
cit keyword will result in an error or misinterpretation.
  warnings.warn(
```



Checks on displot:



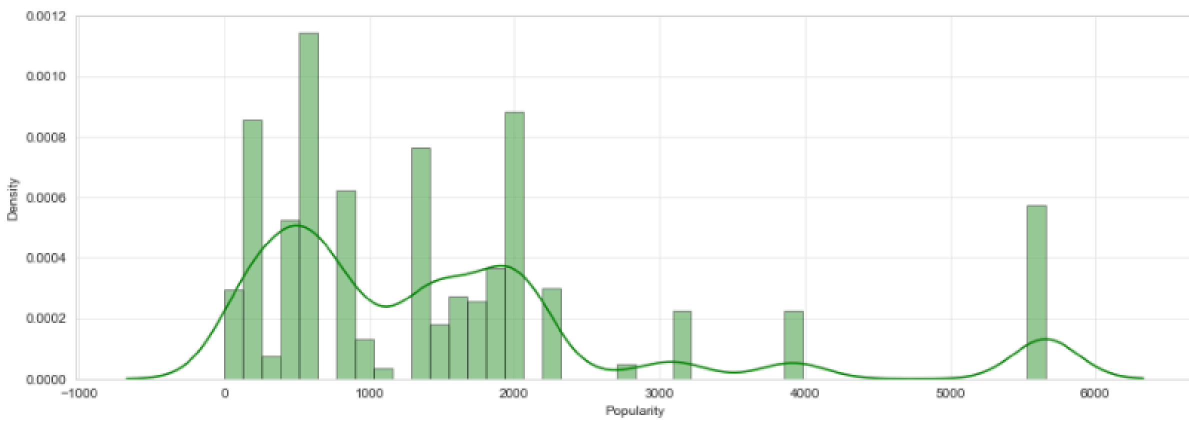
6. In the Popularity column we didn't change anything. Because if we consider 4000 as threshold value for this column then records which are greater than 4000 is Ford as per dataset, which is reality as well.

```
d = df[df['Popularity'] > 4000]
d['Make'].unique()
```

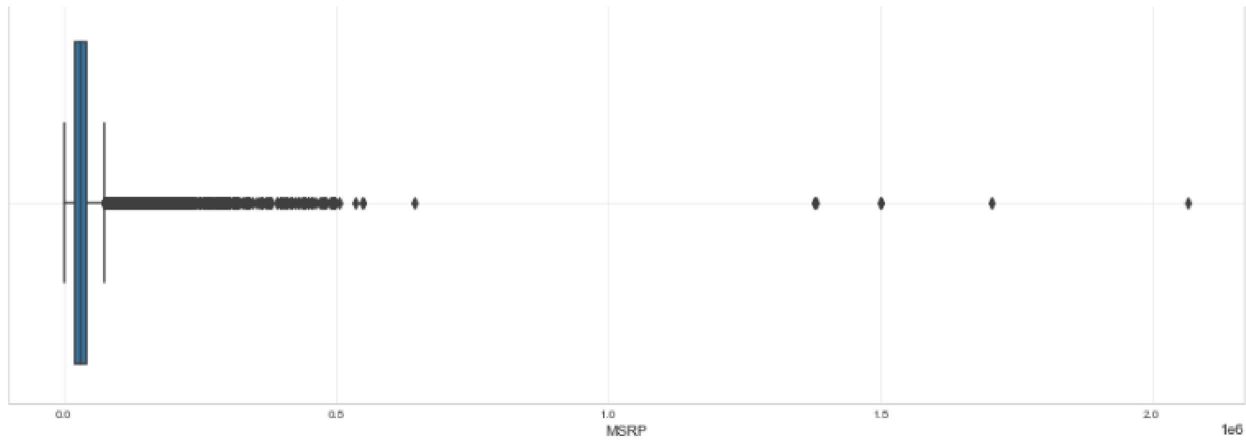
```
0]: array(['Ford'], dtype=object)
```

```
fig, axs = plt.subplots(1, 1,figsize=(15, 5))
sns.distplot(a=df['Popularity'], color='green',
             hist_kws={"edgecolor": 'black'})
axs.grid(axis='y',alpha=0.4)
axs.grid(axis='x',alpha=0.4)
axs.set_axisbelow(True)
plt.show()
```

```
C:\Users\Fahim\anaconda3\lib\site-packages\seaborn\distributions.py:2551: FutureWarning: `distplot` is a deprecated func
and will be removed in a future version. Please adapt your code to use either `displot` (a figure-level function with si
r flexibility) or `histplot` (an axes-level function for histograms).
  warnings.warn(msg, FutureWarning)
```



7. For the Outliers in MSRP column, we plotted the box plot. Considering 100000 as the threshold, we observed that the cars with price above 100000 are all Exotic or Performance or Luxury cars. So we didn't change anything.



Here, our dataset is free from outliers now. We got the clean outliers free data for analysis.

Task: Analysis: (Tech stack used - Ms- Excel)

We have downloaded the .csv format of the dataset from the Jupyter notebook to answer the questions asked by the user.

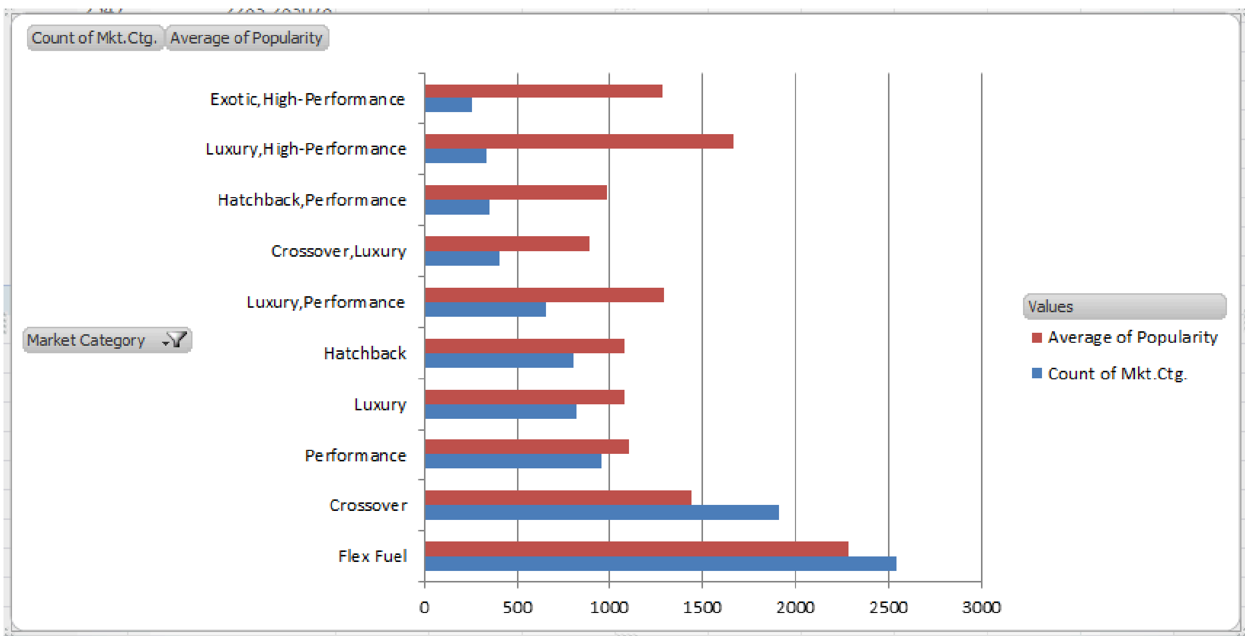
```
# After data cleaning download cleaned dataset:

df.to_csv('~/Desktop/Dataset.csv')
```

Insight Required: How does the popularity of a car model vary across different market categories?

- **Task 1.A:** Create a pivot table that shows the number of car models in each market category and their corresponding popularity scores.
- **Task 1.B:** Create a combo chart that visualizes the relationship between market category and popularity.

Solution: We can observe that the average popularity of cars based on their Market Category mainly ranges from 1200 to 1800 with the exception of Exotic cars being the least popular and Flex Fuel cars being the most popular**.**

Row Labels	Count of Mkt.Ctg.	Average of Popularity
Flex Fuel	2547	2283.283078
Crossover	1908	1439.181866
Performance	951	1102.728707
Luxury	820	1078.087805
Hatchback	806	1081.583127
Luxury,Performance	659	1293.062215
Crossover,Luxury	406	889.2142857
Hatchback,Performance	351	986.1937322
Luxury,High-Performance	334	1668.017964
Exotic,High-Performance	254	1280.047244

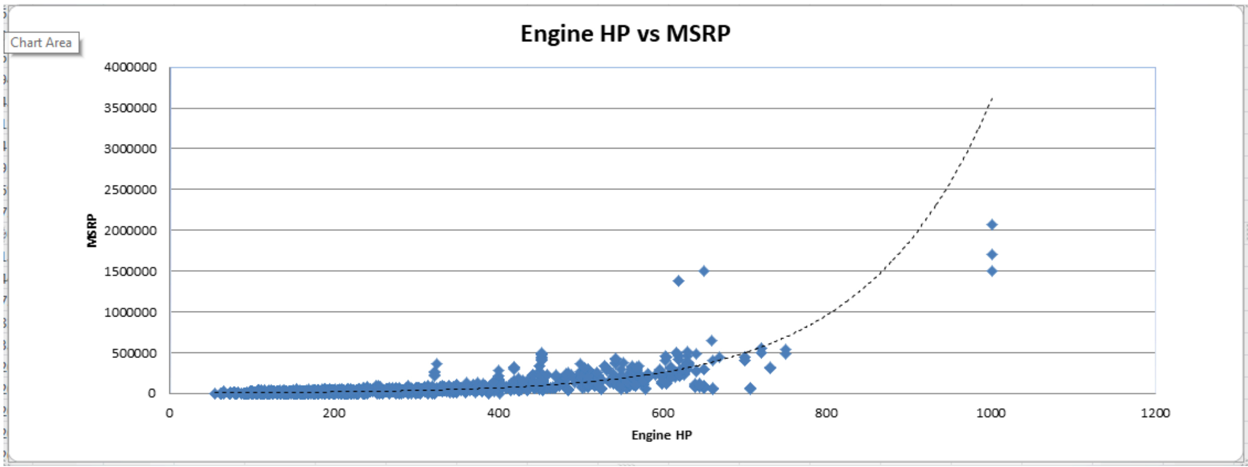


The dataset has high popularity of Flex fuel followed by Luxury with High Performance and Hatchback with Performance.

Insight Required: What is the relationship between a car's engine power and its price?

Task 2: Create a scatter chart that plots engine power on the x-axis and price on the y-axis. Add a trendline to the chart to visualize the relationship between these variables.

Solution: As we can see the relation between Engine HP and Price (MSRP) is positive, where the trendline is moving by creating a positive relation between the field.



This is logical that higher Engine HP requires more complex design and more expensive parts as well.

Insight Required: Which car features are most important in determining a car's price?

Task 3: Use regression analysis to identify the variables that have the strongest relationship with a car's price. Then create a bar chart that shows the coefficient values for each variable to visualize their relative importance.

Solution:

Regression Analysis

SUMMARY OUTPUT

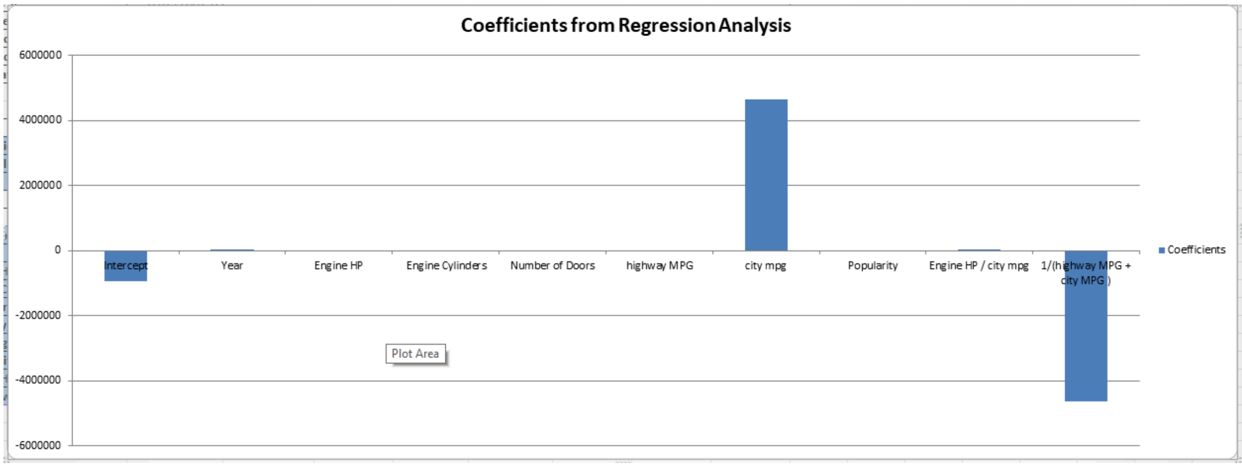
Regression Statistics	
Multiple R	0.821260237
R Square	0.674468377
Adjusted R Square	0.674206532
Standard Error	35123.17187
Observations	11199

ANOVA

	df	SS	MS	F	Significance F
Regression	9	2.85988E+13	3.1776E+12	2575.82986	0
Residual	11189	1.38032E+13	1233637202		
Total	11198	4.24019E+13			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	-950042.2055	129643.4352	-7.32811657	2.4948E-13	-1204166.16	-695918.252	-1204166.16	-695918.252
Year	589.7467467	64.34267293	9.16571724	5.7795E-20	463.6237819	715.869712	463.623782	715.869712
Engine HP	-541.3658636	11.55996891	-46.831083	0	-564.025438	-518.70629	-564.025438	-518.70629
Engine Cylinders	-2432.89896	417.7229157	-5.82419319	5.8969E-09	-3251.7094	-1614.08852	-3251.7094	-1614.08852
Number of Doors	-1259.939581	410.6334238	-3.06828307	0.00215807	-2064.85337	-455.025789	-2064.85337	-455.025789
highway MPG	-4838.819455	235.9277614	-20.5097502	8.495E-92	-5301.2794	-4376.35951	-5301.2794	-4376.35951
city mpg	4636258.992	104874.255	44.2077895	0	4430686.991	4841830.99	4430686.99	4841830.99
Popularity	-1.881250677	0.234163561	-8.03391726	1.0388E-15	-2.34025248	-1.42224888	-2.34025248	-1.42224888
Engine HP / city mpg	12969.9692	152.7394146	84.9156666	0	12670.57306	13269.3653	12670.5731	13269.3653
1/(highway MPG + city MPG)	-4631984.324	104747.3999	-44.2205184	0	-4837307.67	-4426660.98	-4837307.67	-4426660.98

- Using Regression analysis from Ms-excel we have calculated the above report.Here we have calculated two more columns as well (Engine HP / City MPG) and (1/(highway mpg + city mpg).
- We can see that R-square is calculated as 0.67 which is a good score value.

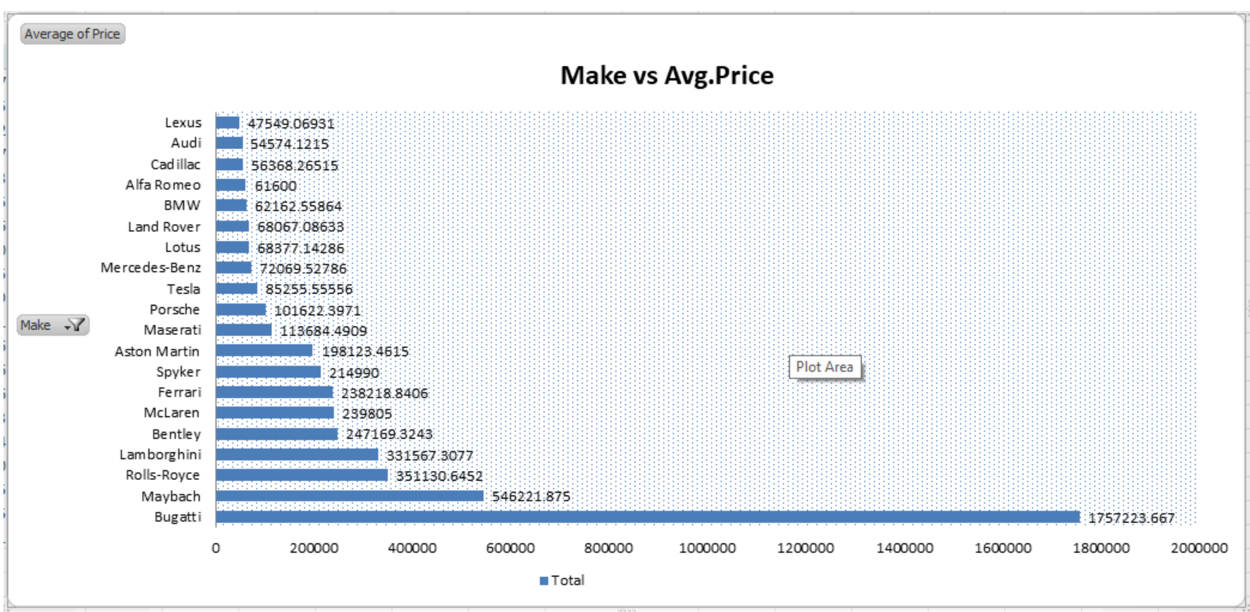


The Bar chart on coefficient is as above.

Insight Required: How does the average price of a car vary across different manufacturers?

Task 4.A: Create a pivot table that shows the average price of cars for each manufacturer.

Task 4.B: Create a bar chart or a horizontal stacked bar chart that visualizes the relationship between manufacturer and average price.



Row Labels	Average of Price
Bugatti	1757223.66
Maybach	546221.87
Rolls-Royce	351130.645
Lamborghini	331567.307
Bentley	247169.324
McLaren	23980
Ferrari	238218.840
Spyker	21499
Aston Martin	198123.461
Maserati	113684.490
Porsche	101622.397
Tesla	85255.5555
Mercedes-Benz	72069.5278
Lotus	68377.1428
Land Rover	68067.0863
BMW	62162.5586
Alfa Romeo	6160
Cadillac	56368.2651
Audi	54574.121
Lexus	47549.0693

- As we can see, the most expensive car is Bugati, followed by Maybach and Rolls-Royse.
- These cars are so expensive.

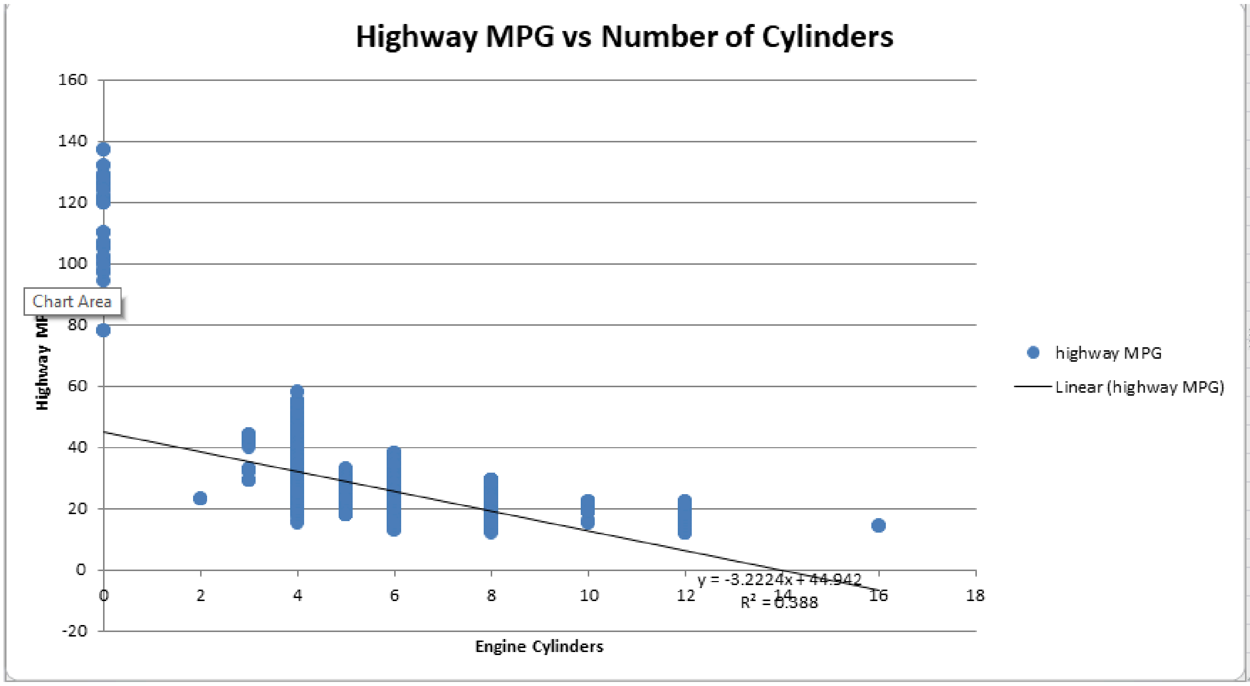
Insight Required: What is the relationship between fuel efficiency and the number of cylinders in a car's engine?

Task 5.A: Create a scatter plot with the number of cylinders on the x-axis and highway MPG on the y-axis. Then create a trendline on the scatter plot to visually estimate the slope of the relationship and assess its significance.

Task 5.B: Calculate the correlation coefficient between the number of cylinders and highway MPG to quantify the strength and direction of the relationship.

Solution: Here we can see correlation between Highway MPG and Number of cylinders is -0.62293632.

Correlation between Highway MPG and Number of Cyliners	-0.62293632

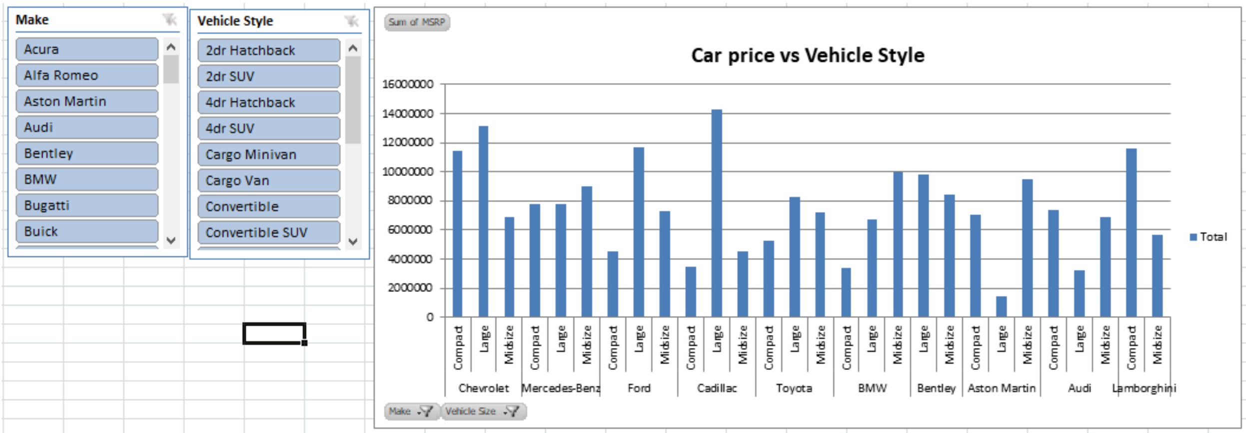


- We can observe that the plot between highway MPG and Engine Cylinders has a negative slope with a value of -3.2224.
- The correlation coefficient is also Negative with a value of -0.62293634
- This is logical because as the number of Engine Cylinders increases, the amount of fuel to be burnt also increases, thus decreasing the mileage (highway MPG).

Building the Dashboard: (Tech stack used - Ms-excel)

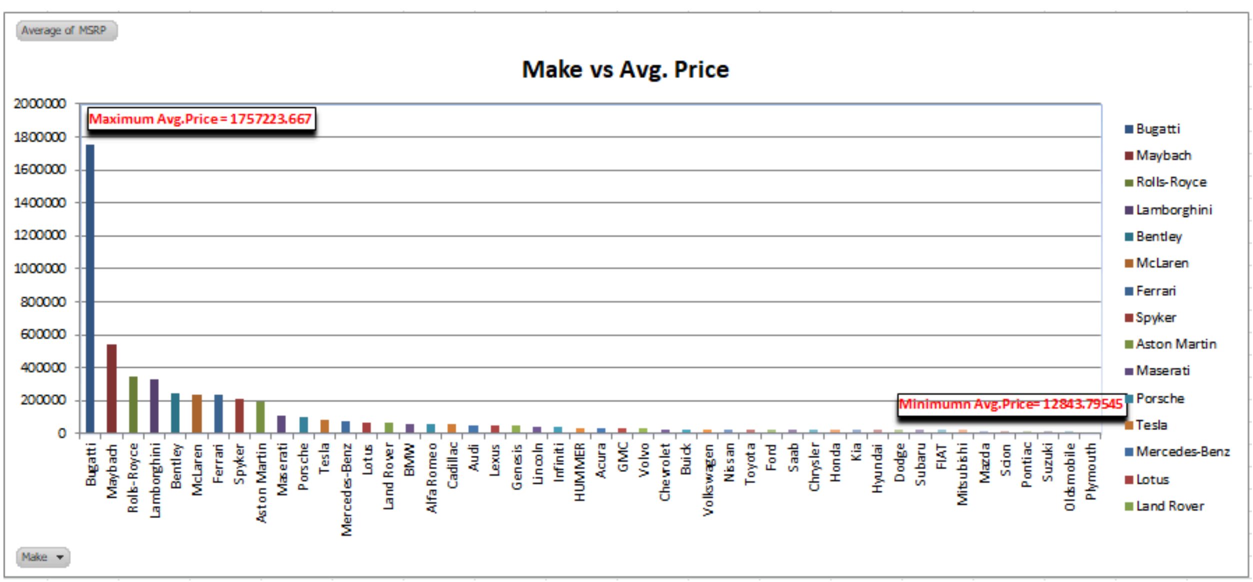
Task 1: How does the distribution of car prices vary by brand and body style?

Solution:



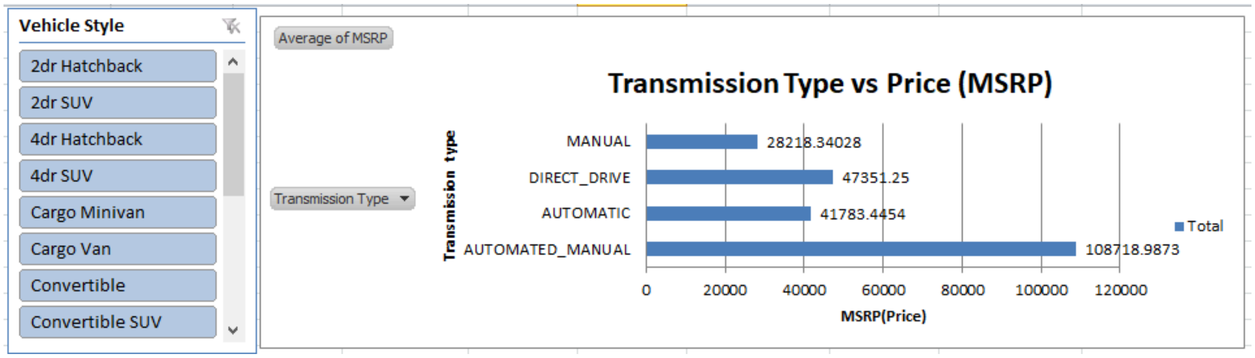
Task 2: Which car brands have the highest and lowest average MSRPs, and how does this vary by body style?

Solution:



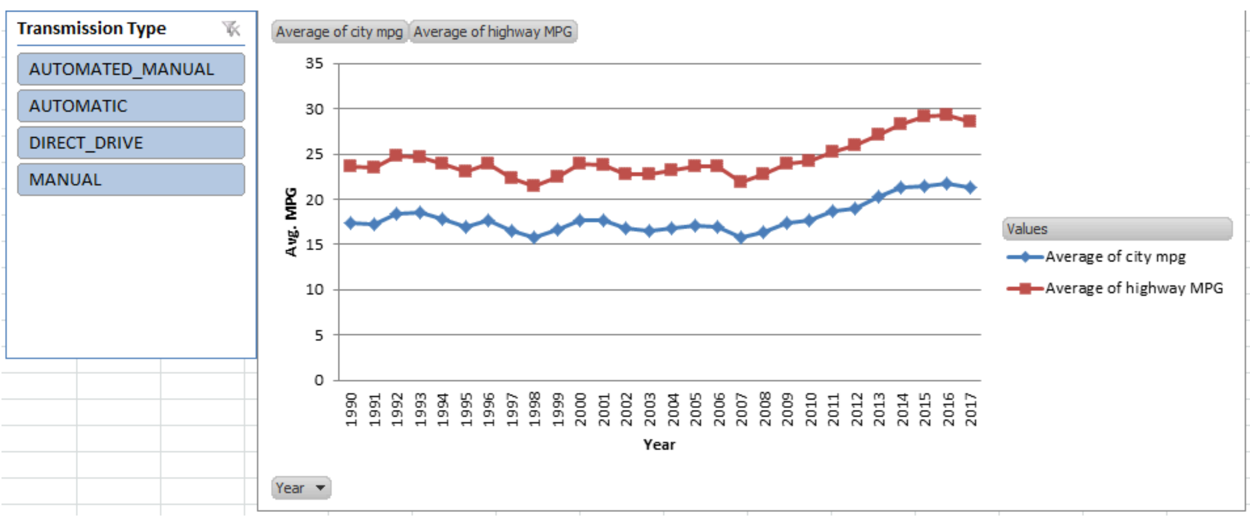
Task 3: How do the different features such as transmission type affect the MSRP, and how does this vary by body style?

Solution:



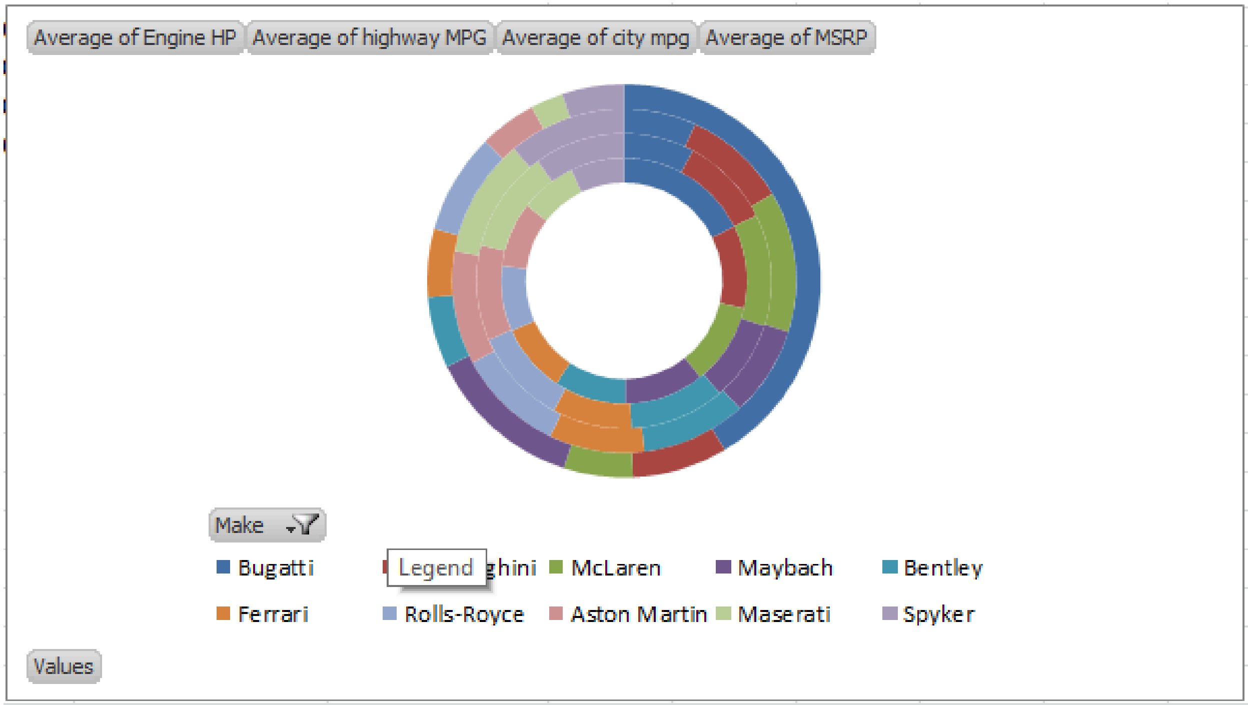
Task 4: How does the fuel efficiency of cars vary across different body styles and model years?

Solution:

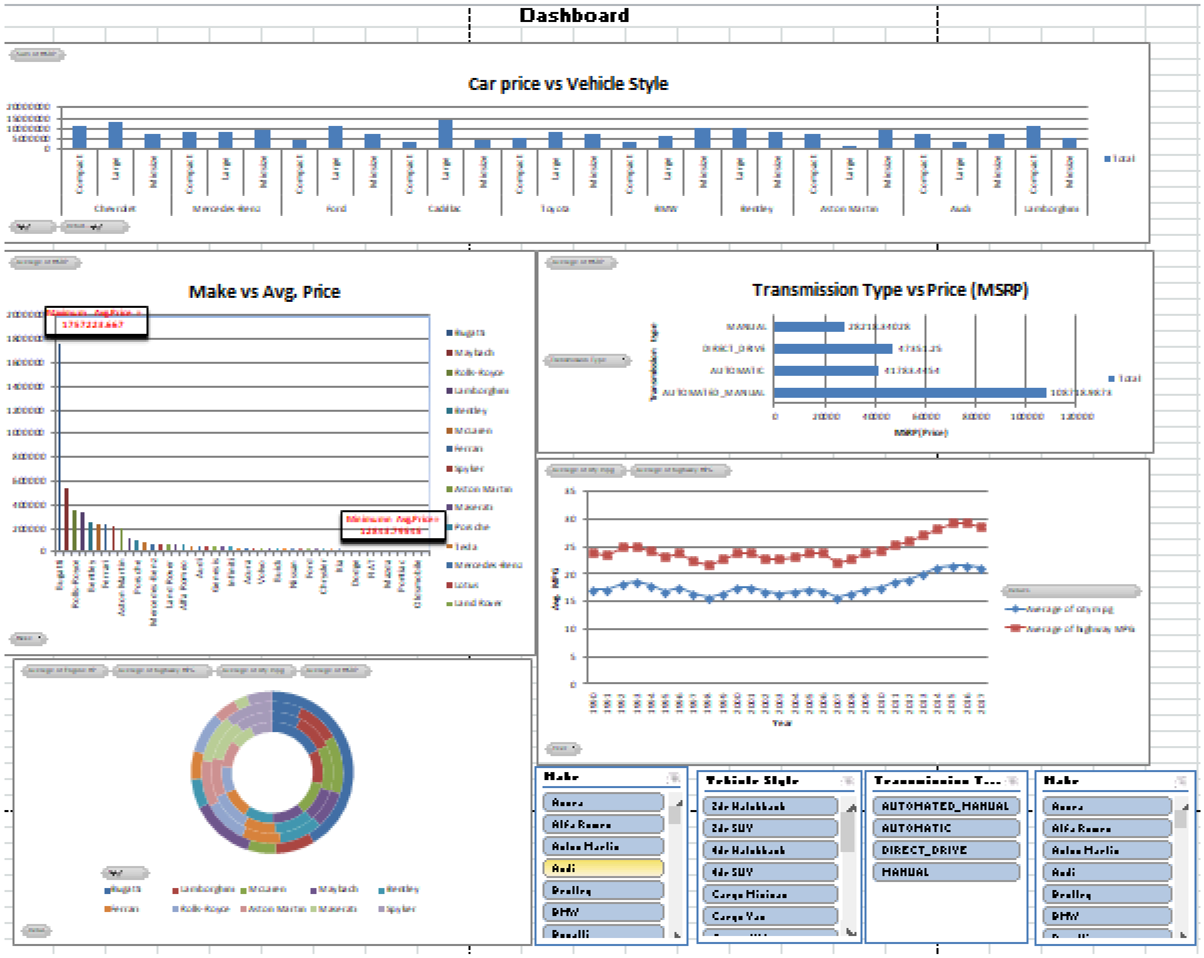


Task 5: How does the car's horsepower, MPG, and price vary across different Brands?

Solution:



Final Dashboard:



Thank You: