Project 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.

Dataset 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

Business aspects:

Analyzing trends in car features and pricing over time: By examining the variables in thedataset, 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.

Tech stack used: I have used google colab to run my code and display the charts. Also I have used PowerBI tool for visualisations and building the charts and Dashboards.

Solutions:

Data cleaning Approach: Firstly, we used info to get an idea about the datatypes and describe to get a statistical summary. We also standardise the column names too. Now We are going to perform some steps in data cleaning to obtain a clean,accurate, complete and relevant dataset.

Step-1:Handling null and duplicate values: We calculated the percentage of null values in the data set using a function. Our approach was to impute the most frequent category or mode for categorical variables and median for the numerical variables. Removed the duplicates values using drop_duplicates method.

Step-2: Outlier treatment : Here we will be taking the numerical columns and using iqr method we set the boundary and remove those values that exceeds the upper and lower limits. we have removed 1746 rows using outlier treatment.

Step-3:Variable Transformation: Here we check for any non-linear or skewed variables that may affect the regression model. Consider transforming variables such as logarithmic or exponential transformations to improve the model fit.We can plot the histogram or density plot of the variables to visualize their distribution. If the distribution is skewed, you may need to transform the variable. Since we have identified that the target variable "MSRP" is highly positively skewed, we need to apply a transformation to reduce the skewness. One commonly used transformation for reducing positive skewness is the log transformation.So we applied log transformation on the MSRP values.

Step-4 Feature Creation: The process of creating and deriving new relevant features from the existing features. We created new features called ‘Luxury_Status’ indicating whether the car is luxury or not, ‘HP_per_cylinder' by dividing Engine_HP by Engine_Cylinders. We also performed binning on Horsepower variable to get a new column called Engine_HP_Binned which gives the categories of low, medium, high for hp values. We have more other task to do which we will do on the regression task like checking multi collinearity, feature selection, encoding categorical variables etc.

Here i have provided the link to acess the google colab python code and powerbi file https://drive.google.com/drive/folders/1luJEf44yMe_vq0dfWZpOL4hjDN3JBjaC?usp=sharing

Task 1.A: Create a pivot table that shows the number of car models in each market category and their corresponding popularity scores.

Approach: We have created a pivot table with Market category as index, Model as column and popularity as values parameter with count as aggregate function. To find the most popular market category overall, we sum the counts across all columns using sum(axis=1), which returns a Series with the same index as the pivot table. We then use idxmax() to find the index with the highest value in the resulting Series. To find the most popular market category for each car make, we loop through the unique values in the "Make" column of the original DataFrame, and create a separate pivot table for each make. We then find the most popular market category for each pivot table using the same method as before.

```
# Create the pivot table
pivot_table = pd.pivot_table(cars_df, values='Popularity', index=['Market_Category'],columns=['Model'], aggfunc='count',fill_value=0)

pivot_table
```

```
[ ]
# Find the most popular market category overall
# Remove the "All" column from the pivot table
most_popular = pivot_table.sum(axis=1).idxmax()
print(f"The most popular market category overall is {most_popular}")
```

```
The most popular market category overall is Crossover
```

```
# Find the most popular market category for each car make
for make in cars_df['Make'].unique():
    make_pivot_table = pd.pivot_table(cars_df[cars_df['Make'] == make], values='Popularity', index='Market_Category', columns='Model', aggfunc='count', fill_value=0)
    most_popular = make_pivot_table.sum(axis=1).idxmax()
    print(f"The most popular market category for {make} is {most_popular}")
```

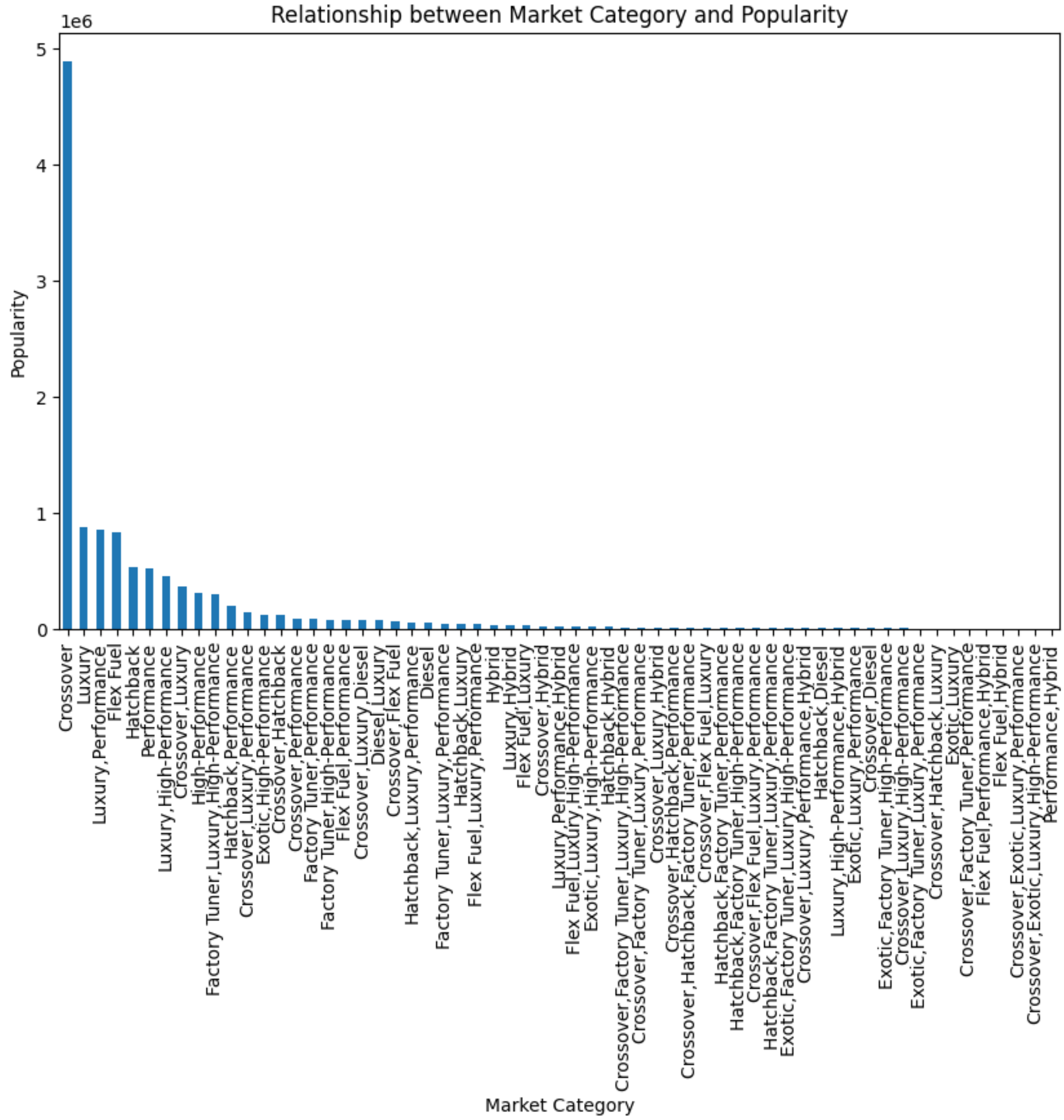
```
The most popular market category for BMW is Luxury,Performance
The most popular market category for Audi is Luxury
The most popular market category for FIAT is Crossover
The most popular market category for Mercedes-Benz is Luxury
The most popular market category for Chrysler is Crossover
The most popular market category for Nissan is Crossover
The most popular market category for Volvo is Luxury
The most popular market category for Mazda is Crossover
The most popular market category for Mitsubishi is Crossover
The most popular market category for Ferrari is Exotic,High-Performance
The most popular market category for Alfa Romeo is Luxury,High-Performance
The most popular market category for Toyota is Crossover
The most popular market category for Pontiac is Crossover
The most popular market category for Porsche is Luxury,High-Performance
The most popular market category for Saab is Luxury
The most popular market category for GMC is Crossover
The most popular market category for Hyundai is Crossover
The most popular market category for Plymouth is Crossover
The most popular market category for Honda is Crossover
The most popular market category for Oldsmobile is Crossover
The most popular market category for Suzuki is Crossover
The most popular market category for Cadillac is Luxury,Performance
The most popular market category for Kia is Crossover
The most popular market category for Bentley is Exotic,Luxury,Performance
```

Task1.B Create a stacked column chart that visualizes the relationship between Market category and popularity.

Approach: We used group by to group by market category and get the sum of popularity values and also sort in descending order using sort_values(ascending=false). Create a new fig and axes object Then we call the plot method with kind=bar , stacked=true to stack the bars on top of each other and the **ax** parameter is set to **ax** to plot the bars on the previously created axes. Then label the axes and give a title for the figure.

```python
category_popularity = cars_df.groupby('Market_Category')['Popularity'].sum().sort_values(ascending=False)

fig, ax = plt.subplots(figsize=(10, 6))
category_popularity.plot(kind='bar', stacked=True, ax=ax)
ax.set_xlabel('Market Category')
ax.set_ylabel('Popularity')
ax.set_title('Relationship between Market Category and Popularity')
plt.show()
```



Insights:

1.Cross over is the most popular category overall among all models followed closely by luxury. This suggests that consumers are prioritizing practicality and functionality when it comes to their car purchases.

2.We can see that there is a clear divide between luxury and non-luxury brands. Luxury brands such as BMW, Mercedes-Benz, Porsche, and Cadillac have the highest popularity in the Luxury, Performance, and Exotic/Luxury/High-Performance categories. Non-luxury brands such as Nissan, Toyota, Honda, and Chevrolet have the highest popularity in the Crossover category.

3.These insights could be useful for car manufacturers and marketers to understand consumer preferences and to tailor their products and marketing efforts accordingly. For example, non-luxury brands may want to focus on highlighting the practicality and functionality of their crossovers, while luxury brands may want to emphasize the high-performance and luxurious features of their vehicles.

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

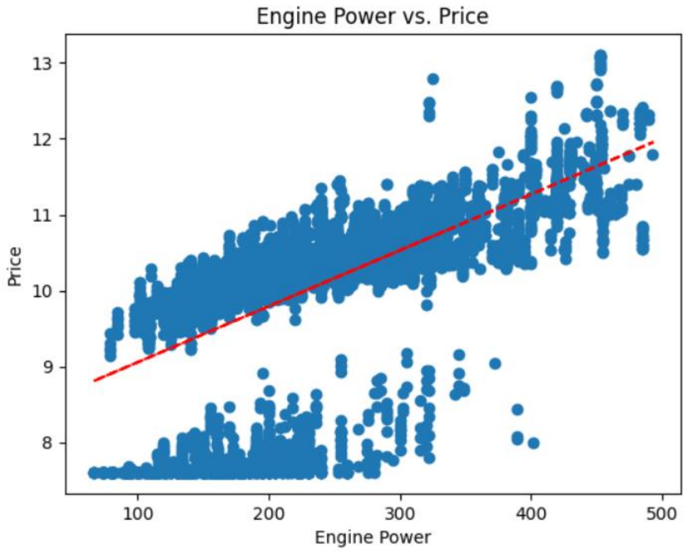
Approach: We create a scatter chart by calling plt.scatter() and passing in the x and y variables. To add a trendline, we use the np.polyfit() function to calculate the slope and y-intercept of the line of best fit, and then use the np.poly1d() function to create a function that can be used to plot the line. We then call plt.plot() and pass in the x and y values for the line, along with the "r--" argument to specify a red dashed line.

```
# create a scatter chart
plt.scatter(cars_df['Engine_HP'], cars_df['MSRP'])

# add a trendline
z = np.polyfit(cars_df['Engine_HP'],cars_df['MSRP'], 1)
p = np.poly1d(z)
plt.plot(cars_df['Engine_HP'], p(cars_df['Engine_HP']), "r--")

# set the chart title and axis labels
plt.title("Engine Power vs. Price")
plt.xlabel("Engine Power")
plt.ylabel("Price")

# show the chart
plt.show()
```



Insights:

1.Overall There is a positive linear relationship between engine power and price, which means that as the engine power of a car increases, the price tends to increase as well.

2.The trendline shows the overall trend in the data, indicating that for a given increase in engine power, there is a corresponding increase in price.It indicates that there is a significant increase in price as engine power increases, but the rate of increase slows down as engine power becomes larger.After a point, the increase in price for additional engine power becomes less significant as the engine power becomes larger. This insight could be valuable for car manufacturers, as it suggests that there may be an optimal engine power level for balancing performance and cost.

3.When the trend line does not pass through some values at the bottom where the price is below 9 but the engine power increases from 100 to 400, it means that there may be other factors affecting the price of cars besides engine power, such as the car's brand, model, market category, and features. These factors may have a more significant impact on the price of lower-priced cars, which could lead to a weaker relationship between engine power and 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.

Approach: We first check multi collinearity using Vif which determines the strength of the correlation between independent variables. The higher the value of VIF and the higher the multicollinearity with the particular independent variable.Multi collinearity adversely affect the regression results ie the estimated regression coefficients may become large and unpredictable, leading to unreliable inferences about the effects of the predictor variables on the response variable. Based on the

VIF values it appears that there is a high degree of multicollinearity between the "Year" column/Mpg values and the other columns in the dataset. Next is we do feature selection for numerical columns that is selecting only relevant features and removing the noise using correlation method which selects the features with the highest correlation to the target variable, select kbest which involves selecting the top K features with the highest predictive power and recursive feature elimination method which recursively removes the least important feature and selects the remaining features that give the best performance.

Correlation method:

```
# calculate correlation matrix
corr_matrix = cars_df_reg.corr()

# select the features with high correlation to the target variable
relevant_features = corr_matrix['MSRP'][abs(corr_matrix['MSRP']) > 0.5].index.tolist()
relevant_features
```

Select K best:

```
from sklearn.feature_selection import SelectKBest,  f_regression
X_reg=X[num_cols]
target = cars_df_reg['MSRP']
# initialize the selector
selector = SelectKBest(score_func=f_regression, k=5)

# fit the selector to the data
selector.fit(X_reg, target)

# get the selected features
relevant_features_Kbest = X_reg.columns[selector.get_support()].tolist()
relevant_features_Kbest
```

RFE:

```
from sklearn.feature_selection import RFE
from sklearn.linear_model import LinearRegression

# initialize the linear regression model
model = LinearRegression()

# initialize the RFE selector
rfe = RFE(model, n_features_to_select=5)

# fit the RFE selector to the data
rfe.fit(X_reg, target)

# get the selected features
relevant_features_rfe = X_reg.columns[rfe.support_].tolist()
relevant_features_rfe
```

Selected features from Kbest are 'Year', 'Engine_HP', 'Engine_Cylinders', 'Number_of_Doors', 'HP_per_cylinder'.

Selected features from RFE are 'Year', 'Engine_Cylinders', 'Number_of_Doors', 'city_mpg', 'HP_per_cylinder'.

Feature Selection for categorical cols: Chi-square Test. In feature selection, the chi-square test can be used to determine the importance of each feature in predicting the target variable. It works by calculating the chi-square statistic for each feature and selecting the top features based on their chi-square values. The higher the chi-square value, the more significant the feature is in predicting the target variable.

```
from sklearn.feature_selection import chi2
from sklearn.preprocessing import LabelEncoder
from sklearn.preprocessing import KBinsDiscretizer
# Perform feature selection using chi-squared test
# convert target variable to categorical variable using binning
est = KBinsDiscretizer(n_bins=3, encode='ordinal', strategy='quantile')
target_binned = est.fit_transform(target.values.reshape(-1, 1))
X_cat=cars_df_reg[cat_cols]
selector = SelectKBest(chi2, k=5) # Select top 5 features
selector.fit(X_cat, target_binned.ravel())
# get the selected features
relevant_features_cat = X_cat.columns[selector.get_support()].tolist()
relevant_features_cat
```

So the five relevant categorical features are the 'Make', 'Model', 'Market_Category', 'Vehicle_Size', 'luxury_status'

Next step is Normalization. Here Check for any variables that are on different scales or units of measurement. Normalize the variables to ensure that they are on a similar scale.Normalization helps in scaling the input features to a fixed range, typically [0, 1], to ensure that no single feature disproportionately impacts the results. It preserves the relationship between the minimum and maximum values of each features.

```
# Select numeric columns
num_cols_scale = cars_df_reg[['Year','Engine_HP','highway_MPG','city_mpg','Popularity','HP_per_cylinder']].columns
num_cols_scale
```

```
Index(['Year', 'Engine_HP', 'highway_MPG', 'city_mpg', 'Popularity', 'HP_per_cylinder'], dtype='object')
```

```
from sklearn.preprocessing import MinMaxScaler
# Perform MinMax scaling on the numerical columns

scaler = MinMaxScaler()

cars_df_reg[num_cols_scale] = scaler.fit_transform(cars_df_reg[num_cols_scale])

# Check the scale of values again after scaling
print(cars_df_reg.describe())
```

Now we use those features only selected from the feature selection. We split target variable as Y and all independent selected variables as X. Now carry out regression analysis using sm.ols model from stats library.For intercept term add constant. Get the coefficient value for each feature and print the results using summary function which gives many statistics values. Also we plot a bar chart showing the feature coefficient values which sets red color for negative values and green for positive values.

```python
# Select the features to use for prediction
X = cars_df_reg[['Year', 'Engine_HP', 'Engine_Cylinders', 'Number_of_Doors',
      'HP_per_cylinder', 'Make','Model','Market_Category', 'Vehicle_Size', 'Luxury_Status']]
# Define the target variable
y = cars_df_reg['MSRP']
```

```python
import statsmodels.api as sm
# Create a linear regression model using OLS
model = sm.OLS(y, sm.add_constant(X))
# Fit the model to the data
results = model.fit()
# Get the coefficient values for each feature
coefficients = pd.DataFrame({'Feature': X.columns, 'Coefficient': results.params[1:]})
```

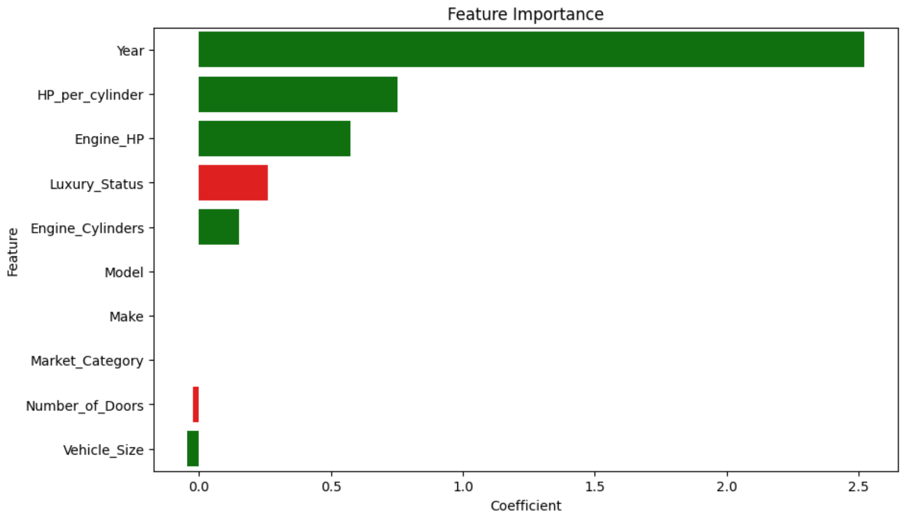
```python
# Create a bar chart of the coefficient values
plt.figure(figsize=(10, 6))
# Set the bar colors based on the sign of the coefficients
colors = np.where(coefficients['Coefficient']>=0, 'g', 'r')
sns.barplot(x='Coefficient', y='Feature', data=coefficients.sort_values(by='Coefficient', ascending=False),palette=colors)
plt.title('Feature Importance')
plt.xlabel('Coefficient')
plt.ylabel('Feature')
plt.show()
```

OLS Regression Results

Dep. Variable:	MSRP	**R-squared:**	0.753
Model:	OLS	**Adj. R-squared:**	0.753
Method:	Least Squares	**F-statistic:**	2914.
Date:	Wed, 19 Apr 2023	**Prob (F-statistic):**	0.00
Time:	10:47:40	**Log-Likelihood:**	-6940.7
No. Observations:	9578	**AIC:**	1.390e+04
Df Residuals:	9567	**BIC:**	1.398e+04
Df Model:	10		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P>|t|**	**[0.025**	**0.975]**
const	6.9068	0.080	85.889	0.000	6.749	7.064
Year	2.5221	0.026	96.970	0.000	2.471	2.573
Engine_HP	0.5726	0.149	3.843	0.000	0.281	0.865
Engine_Cylinders	0.1525	0.016	9.799	0.000	0.122	0.183
Number_of_Doors	-0.0227	0.007	-3.346	0.001	-0.036	-0.009
HP_per_cylinder	0.7512	0.146	5.144	0.000	0.465	1.038
Make	-0.0002	0.000	-0.510	0.610	-0.001	0.001
Model	2.615e-05	2.34e-05	1.118	0.264	-1.97e-05	7.2e-05
Market_Category	-0.0014	0.000	-5.201	0.000	-0.002	-0.001
Vehicle_Size	-0.0439	0.007	-6.708	0.000	-0.057	-0.031
Luxury_Status	0.2606	0.015	17.804	0.000	0.232	0.289

Omnibus:	501.230	**Durbin-Watson:**	0.367
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	957.028
Skew:	0.389	**Prob(JB):**	1.53e-208



Insights:

1.The regression model has an R-squared value of 0.753, which means that 75.3% of the variance in the target variable (MSRP) can be explained by the independent variables included in the model. This is a relatively high value and suggests that the model is a good fit for the data.

2.The intercept value of 6.90 suggests that a car with no features would have an estimated price of around $6,900.

3.The car's age, horsepower, number of cylinders, horsepower per cylinder, and luxury status are the most important features in determining the car's price. The number of doors and vehicle size also have a significant impact on the car's price, but in the opposite direction.

4.The "P-value" column indicates the statistical significance of each feature. A p-value less than 0.05 is considered statistically significant, meaning that the feature is likely to have a real impact on the car's price. Here Make and model columns have a p value above 0.05 indicating they do not have much influence on the price of the car. Also the coefficients for "Make", "Model", and "Market_Category" are relatively small, indicating that they have less influence on the car's price compared to other features in the model.

5.The "Number_of_Doors" variable and 'Vehicle Size'has a negative coefficient indicating that larger cars and cars with more doors tend to have a slightly lower prices.

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

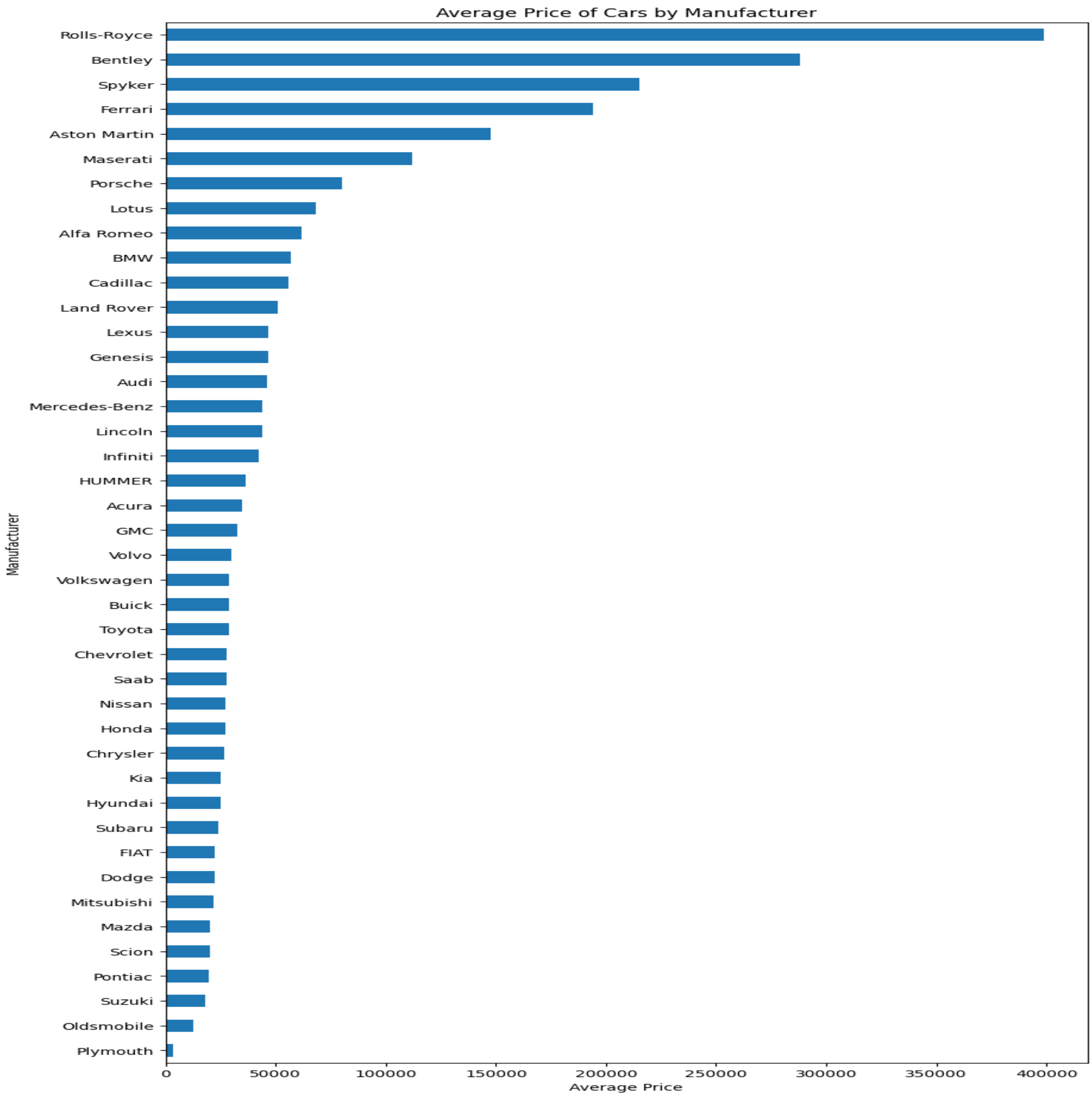
Task 4.B: Create a bar chart or a horizontal stacked bar chart that visualizes therelationship between manufacturer and average price.

Approach: We created a pivot table with make as index and Msrp as values with aggregate function is mean. For plotting purpose we used sort_values to sort the pivot table by msrp values in descending order. Then use kind=barh to plot a horizontal bar chart and customized the chart with title and axes labels etc.

```
manufacturer_avg_price = pd.pivot_table(cars_df, values='MSRP', index='Make', aggfunc='mean')
manufacturer_avg_price
```

	MSRP
Make	
Acura	34617.374486
Alfa Romeo	61600.000000
Aston Martin	147543.636364
Audi	46279.003472
BMW	56835.982394
Bentley	287832.666667
Buick	29034.189474
Cadillac	56076.222222
Chevrolet	28050.017225
Chrysler	26722.962567
Dodge	22144.196850
FIAT	22206.016949
Ferrari	193918.621622
GMC	32444.085062
Genesis	46616.666667
HUMMER	36464.411765

Genesis	46616.666667
HUMMER	36464.411765
Honda	27112.066313
Hyundai	24871.238095
Infiniti	42640.271341
Kia	25185.625000
Land Rover	51006.153846
Lexus	46656.076503
Lincoln	44123.158940
Lotus	68377.142857
Maserati	112022.057692
Mazda	20336.517157
Mercedes-Benz	44159.767176
Mitsubishi	21920.675532
Nissan	27507.273585
Oldsmobile	12843.795455



Insights:

1.There is a wide range of average prices among car manufacturers starting from 3296.87 to 398605.88 $. So other factors like as brand perception, car features, target market, and production costs also influences the price. 2.Rolls-Royce is having highest average price and plymouth having lowest price.

2.Luxury car brands such as Bentley, spyker,Aston Martin have higher average prices compared to other car manufacturers.

3.Japanese car brands such as Honda, Toyota, and Nissan have relatively lower average prices compared to other manufacturers.

4.American brands like Chevrolet and Ford have relatively lower average prices compared to other brands, which may indicate that they may target a mass market and value-conscious customers.

5.Some brands like Ferrari, Maserati, and Lotus have very high average prices, which suggests that they target a market of high-end sports car enthusiasts.

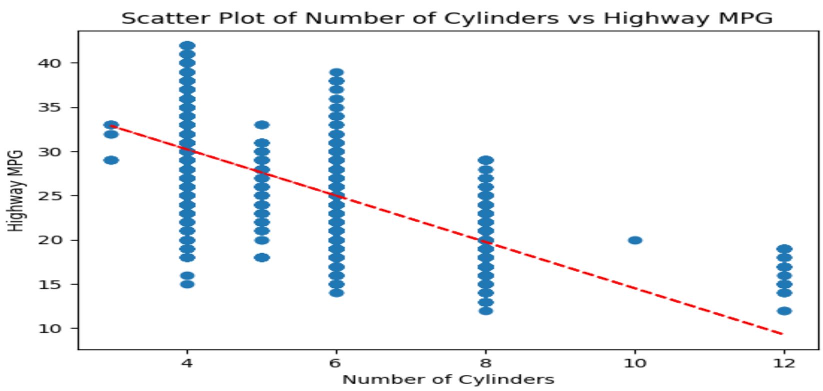
6.Brands like Kia, Hyundai, and Mitsubishi have relatively lower average prices, which may indicate that they target a younger and more budget-conscious customer base.

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.

Approach: We create a scatter chart by calling plt.scatter() and passing in the x and y variables. To add a trendline, we use the np.polyfit() function to calculate the slope and y-intercept of the line of best fit, and then use the np.poly1d() function to create a function that can be used to plot the line. We then call plt.plot() and pass in the x and y values for the line, along with the "r--" argument to specify a red dashed line.Then use df.corr() to calculate the correlation coefficient between the engine_cylinders and highway_mpg.

```
# create a scatter plot with number of cylinders on x-axis and highway MPG on y-axis
plt.scatter(cars_df['Engine_Cylinders'], cars_df['highway_MPG'])
# add a trendline
z = np.polyfit(cars_df['Engine_Cylinders'], cars_df['highway_MPG'], 1)
p = np.poly1d(z)
plt.plot(cars_df['Engine_Cylinders'],p(cars_df['Engine_Cylinders']),"r--")
# add labels and title
plt.xlabel('Number of Cylinders')
plt.ylabel('Highway MPG')
plt.title('Scatter Plot of Number of Cylinders vs Highway MPG')
# display the plot
plt.show()
```



```
# Select the relevant columns
cylinders = cars_df["Engine_Cylinders"]
highway_mpg = cars_df["highway_MPG"]
# Calculate the correlation coefficient
corr_coef = cylinders.corr(highway_mpg)
print("Correlation coefficient between number of cylinders and highway MPG:", corr_coef)
```

```
Correlation coefficient between number of cylinders and highway MPG: -0.6918908916275935
```

Insights:

1.Generally, cars with 4 cylinders have a higher highway MPG compared to cars with 6, 8, or 12 cylinders.

2.Cars with 8 and 12 cylinders have a lower range of highway MPG compared to cars with fewer (4/6)cylinders.

3.A valuable insight is such as a car with 10 cylinders that has a highway MPG of 20, which is higher than most cars with 8 or 12 cylinders.

4.The decreasing trend and the negative correlation coefficient indicates there is a negative correlation between the two variables -Engine_Cylinders and highway_MPG. This indicates that as the number of engine cylinders increases, the fuel efficiency (measured in highway miles per gallon) tends to decrease. So one possible insight from this scatter plot and trend line is that cars with fewer engine cylinders tend to have better fuel efficiency on the highway compared to cars with more engine cylinders.

Building the Dashboard:

Now for the Next portion of the Project, you need to create the Interactive Dashboard.

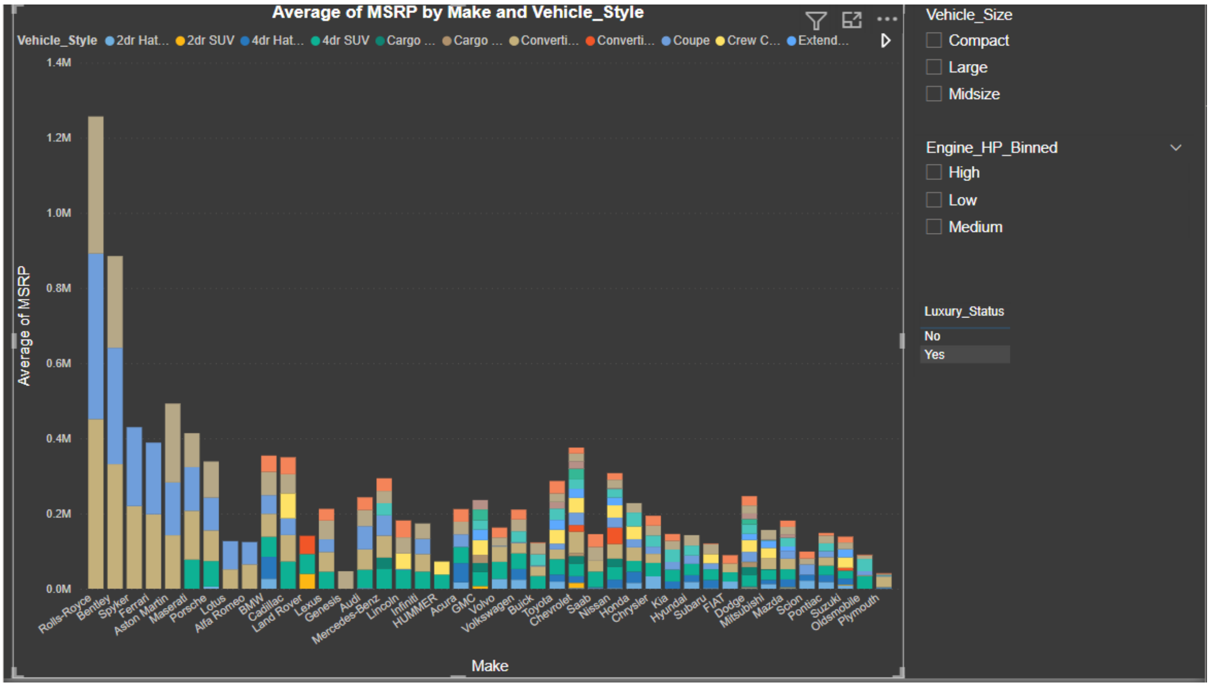
Use filters and slicers to make the chart interactive. The client has requested these

questions given below:

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

Stacked column chart to show the distribution of car prices by brand and body style. Use filters and slicers to make the chart interactive. Calculate the total MSRP for each brand and body style using SUMIF or Pivot Tables.

Approach: Stacked column charts is used to compare one category over the other .Open Power BI Desktop and click on "Get Data" from the Home tab.Select the car dataset file and import the data into Power BI. Go to the "Visualizations" pane and select "Stacked Column Chart" from the list of chart types. Drag and drop the "Make" column into the X-Axis section of the chart. Drag and drop the "Vehicle Style" column into the Legend section of the chart.Drag and drop the "MSRP" column into the Y-axis section of the chart. In X-axis grouping column or category is displayed and values are displayed in Y-axis. In Legend, we refer the column that will be displayed as different color bars. Now customize the chart using format visual giving title, changing font size etc. We also add slicers to vehicle size, hp binned, luxury status columns to make the chart interactive and helps us to filter a portion of the chart as required. Also we create a pivot table for TaskB. Go to Go to the "Visualizations" pane and select "Matrix" from the list of chart types. Drag and drop the "Make" column into the Rows section, the "Vehicle Style" column into the Column section of the chart, the "MSRP" column into values section with aggregate as sum to get the total MSRP. Rename the field as Total MSRP.



Make	2dr Hatchback	2dr SUV	4dr Hatchback	4dr SUV	Cargo Minivan	Cargo Van	Convertible	Convertible SUV	Coupe	Crew Cab Pickup	Extended Cab Pickup	Pas
Acura	4,80,917.00		3,57,440.00	26,63,505.00					6,37,748.00			
Alfa Romeo							1,29,800.00		1,78,200.00			
Aston Martin							25,63,655.00		30,87,775.00			
Audi	4,000.00			26,74,900.00			22,21,505.00		16,69,100.00			
Bentley							9,94,970.00		18,57,580.00			
BMW	80,097.00		9,32,300.00	25,58,150.00			40,45,271.00		29,65,051.00			
Buick				21,41,770.00			1,79,325.00		18,534.00			
Cadillac				71,82,555.00			9,85,607.00		27,56,859.00	5,99,150.00		
Chevrolet	2,000.00	1,93,310.00	10,54,680.00	65,09,468.00	4,20,150.00	74,688.00	20,38,910.00	1,06,300.00	25,65,085.00	59,27,617.00	31,17,951.00	
Chrysler	98,805.00			2,50,545.00			6,30,105.00		1,14,510.00			
Dodge	38,000.00	12,000.00	16,000.00	24,62,875.00	60,520.00	3,38,497.00	6,000.00		12,67,902.00	20,72,780.00	6,84,682.00	
Ferrari							31,71,211.00		40,03,778.00			
FIAT	3,25,315.00			3,69,305.00			3,27,965.00					
Genesis												
GMC		1,28,319.00		66,33,919.00	1,42,750.00	4,60,085.00				40,62,482.00	21,75,866.00	
Honda	2,05,675.00		13,62,120.00	38,00,589.00			2,52,135.00		15,88,705.00	7,50,215.00		
HUMMER				3,77,490.00						2,42,405.00		
Hyundai	7,89,650.00		5,28,880.00	19,94,390.00					6,85,920.00			
Infiniti				43,40,200.00			9,80,050.00		21,75,750.00			
Kia			4,06,960.00	20,49,645.00					1,42,630.00			
Land Rover		4,76,394.00		46,82,515.00				1,45,731.00				
Lexus				29,03,574.00			4,72,065.00		6,41,472.00			
Lincoln				34,22,570.00					17,342.00	4,53,260.00		
Lotus							4,13,260.00		15,01,300.00			
Maserati				1,55,000.00			23,42,963.00		19,72,284.00			
Mazda	18,000.00	12,000.00	8,53,180.00	30,61,165.00			8,70,505.00		5,41,879.00		5,80,033.00	
Mercedes-Benz				30,69,270.00	28,950.00		24,01,234.00		18,83,027.00			
Mitsubishi	3,70,169.00		1,32,270.00	20,09,807.00	2,000.00		2,09,893.00			2,40,210.00	1,34,360.00	
Nissan	14,683.00		10,23,090.00	41,49,630.00	1,28,620.00		14,06,552.00	1,31,075.00	19,85,392.00	24,22,300.00	10,26,379.00	
Oldsmobile				2,38,150.00			2,000.00		2,76,015.00			
Plymouth	40,000.00		14,000.00				85,631.00		8,000.00			
Pontiac	1,63,505.00		1,62,975.00	4,01,550.00			4,73,481.00		6,63,715.00			
Porsche	28,827.00			11,53,500.00			25,37,486.00		35,67,733.00			
Rolls-Royce							18,06,365.00		13,21,400.00			
Total	62,65,629.00	8,34,023.00	1,13,79,368.00	8,66,57,815.00	7,82,990.00	8,73,270.00	3,52,05,746.00	5,03,300.00	4,18,13,357.00	2,05,72,420.00	1,14,70,354.00	

Insights:

1.Luxury brands such as Audi, BMW, and Mercedes-Benz tend to have higher car prices compared to non-luxury brands like Toyota and Honda.

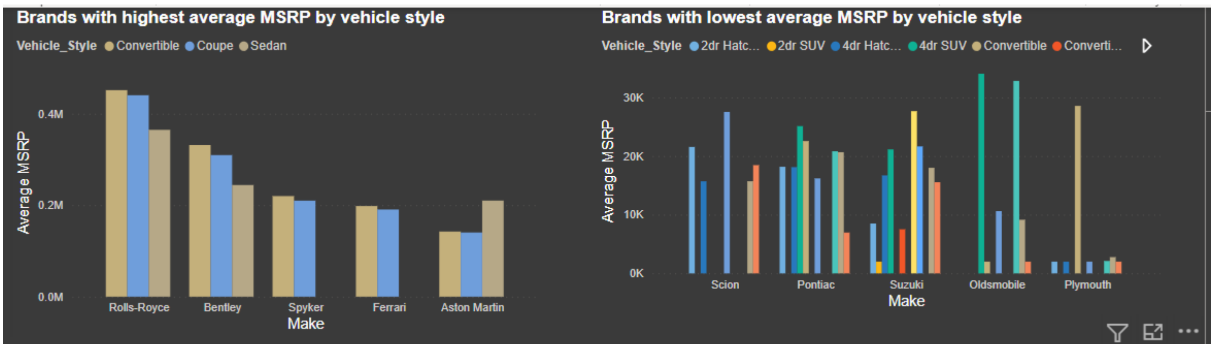
2. Within each brand, the body style can greatly affect the price. For example, BMW sedans tend to be more expensive than BMW SUVs.

3. Body style also plays a role in the price of cars. Generally, SUVs tend to be more expensive than sedans and coupes.

4. From the pivot table we can infer that Cadillac has the highest total MSRP among all the makes with a sum of 25,667,853.

Task 2: Which car brands have the highest and lowest average MSRPs, and how does this vary by body style? Clustered column chart to compare the average MSRPs across different car brands and body styles. Calculate the average MSRP for each brand and body style using AVERAGEIF or Pivot Tables.

Approach: Go to the "Visualizations" pane and select "Clustered Column Chart" from the list of chart types. Drag and drop the "Make" column into the X-Axis section of the chart. Drag and drop the "Vehicle Style" column into the Legend section of the chart.Drag and drop the "MSRP" column into the Y-axis section of the chart and change the aggregation to average. To filter the chart by highest and lowest average MSRPs, you can use the "Top N" and "Bottom N" filters in Power BI. Click on the "Filters" pane in the "Visualizations" tab of the ribbon. go to make column in filters, change the basic filter to top 10 in filter type give the number ie the number of items to display, such as "5" for the top or bottom 5 items.give average msrp is the value to sort by Click "Apply filter" to update the chart.This will filter the chart to show only the top or bottom N combinations of brand and body style based on the average MSRP. Now customize the chart using format visual giving title, changing font size etc.Also we need create a pivot table. Go to the "Visualizations" pane and select "Matrix" from the list of chart types. Drag and drop the "Make" column into the Rows section, the "Vehicle Style" column into the Column section of the chart, the "MSRP" column into values section with aggregate as Average to get the total MSRP. Rename the field as Average MSRP.



Make	2dr Hatchback	2dr SUV	4dr Hatchback	4dr SUV	Cargo Minivan	Cargo Van	Convertible	Convertible SUV	Coupe	Crew Cab Pickup	Extended Cab Pickup	Passenger Minivan
Acura	17,175.61		51,062.86	42,959.76					33,565.68			
Alfa Romeo							64,900.00		59,400.00			
Aston Martin							1,42,425.28		1,40,353.41			
Audi	2,000.00			48,634.55			54,183.05		61,818.52			
Bentley							3,31,656.67		3,09,596.67			
BMW	26,699.00		58,268.75	53,294.79			61,291.98		49,417.52			
Buick				33,996.35			25,617.86		2,059.33			30,005.91
Cadillac				72,551.06			70,400.50		44,465.47	66,572.22		
Chevrolet	2,000.00	13,807.86	17,578.00	33,553.96	20,007.14	8,298.67	56,636.39	17,716.67	32,885.71	39,255.74	24,170.16	24,934.29
Chrysler	32,935.00			35,792.14			24,234.81		19,085.00			29,751.45
Dodge	2,000.00	2,000.00	2,000.00	31,175.63	20,173.33	12,536.93	2,000.00		26,414.63	31,405.76	16,301.95	25,337.50
Ferrari							1,98,200.69		1,90,656.10			
FIAT	19,136.18			24,620.33			23,426.07					
Genesis												
GMC		7,128.83		37,479.77	23,791.67	21,908.81				39,062.33	27,895.72	25,105.00
Honda	15,821.15		29,611.30	28,575.86			36,019.29		21,763.08	34,100.68		36,879.00
Total	**16,575.74**	**13,451.98**	**21,925.56**	**38,583.18**	**20,605.00**	**15,320.53**	**57,904.19**	**17,975.00**	**44,768.05**	**36,475.92**	**22,894.92**	**26,637.27**

Insights:

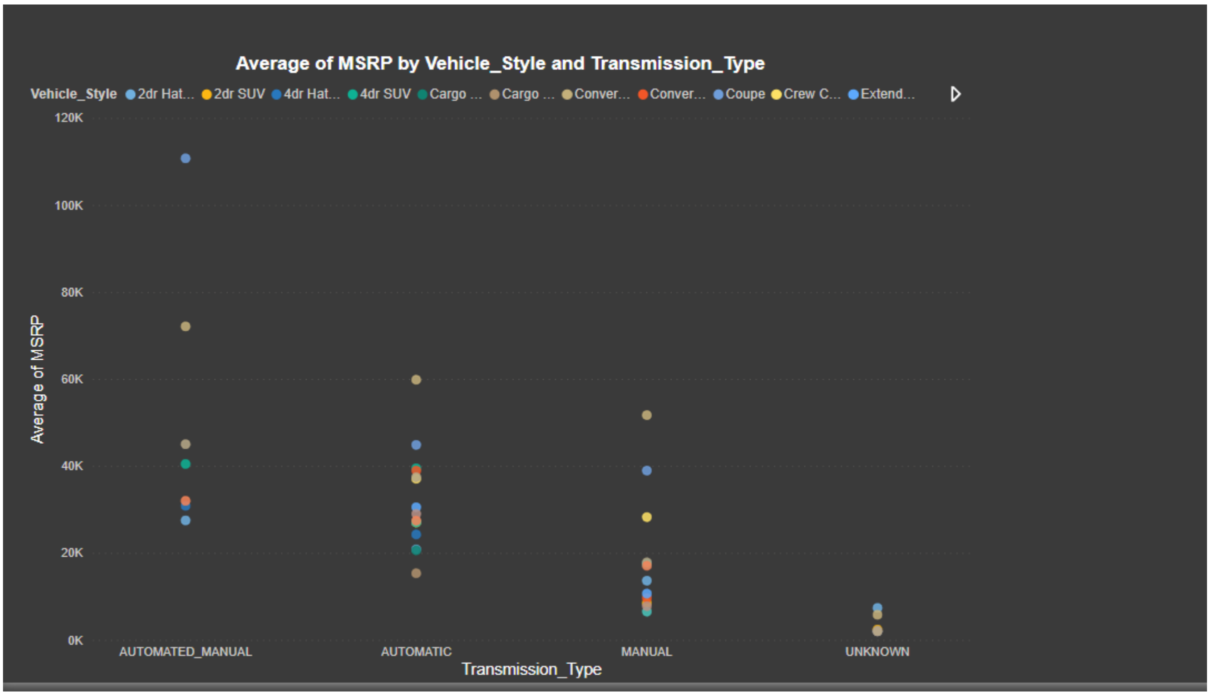
1. Roll Royce,Bentley, Spyker, Ferrari and Aston Martin are the cars having highest Average MSRPs. Here except Aston Martin, in all other brands Convertible and Coupe body styles tend to be expensive than sedan body style. Only in Aston Martin brand, Sedan tends to be expensive than Convertible and Coupe body styles

2.Scion,Pontiac,Suzuki,Oldsmobile and Plymoutha are the cars having lowest Average MSRPS. Here body style has no clear trend. In each of the brands, cars with different body style is expensive. Coupe tends to be expensive for Scion, Suv tends to be expensive for Pontiac etc.

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

Scatter plot chart to visualize the relationship between MSRP and transmission type, with different symbols for each body style. Calculate the averageMSRP for each combination of transmission type and body style using AVERAGEIFS or Pivot Tables.

Approach: Go to the "Visualizations" pane and select "Scatter Chart" from the list of chart types. Drag and drop the "Transmission_Type" column into the X-Axis section of the chart. Drag and drop the "Vehicle Style" column into the Legend section of the chart. Drag and drop the "MSRP" column into the Y-axis section of the chart and change the aggregation to average. Now customize the chart using format visual giving title, changing font size etc. Also we need create a pivot table. Go to the "Visualizations" pane and select "Matrix" from the list of chart types. Drag and drop the " Transmission_Type " column into the Columns section, the "Vehicle Style" column into the Rows section of the chart, the "MSRP" column into values section with aggregate as Average to get the total MSRP. Rename the field as Average MSRP.



Vehicle_Style	AUTOMATED_MANUAL	AUTOMATIC	MANUAL	UNKNOWN	Total
2dr Hatchback	27,470.42	20,806.75	13,614.99	7,361.50	16,575.74
2dr SUV		27,004.11	8,461.63	2,371.00	13,451.98
4dr Hatchback	30,828.71	24,250.94	17,504.03		21,925.56
4dr SUV	40,451.15	39,464.85	17,422.09		38,583.18
Cargo Minivan		20,605.00			20,605.00
Cargo Van		15,320.53			15,320.53
Convertible	72,088.70	59,818.62	51,682.98	5,783.50	57,904.19
Convertible SUV		38,925.50	9,594.80		17,975.00
Coupe	1,10,733.71	44,842.74	38,927.31	2,000.00	44,768.05
Crew Cab Pickup		37,054.64	28,233.11		36,475.92
Extended Cab Pickup		30,501.46	10,653.14		22,894.92
Passenger Minivan		27,125.21	6,510.00		26,637.27
Passenger Van		27,022.30			27,022.30
Regular Cab Pickup		28,960.56	7,757.03	2,000.00	17,869.90
Sedan	45,007.19	37,415.10	17,823.44	2,000.00	34,131.17
Wagon	31,985.28	27,419.85	17,066.01		24,969.66
Total	**54,403.13**	**36,770.06**	**23,957.47**	**3,647.83**	**34,348.73**

Insights:

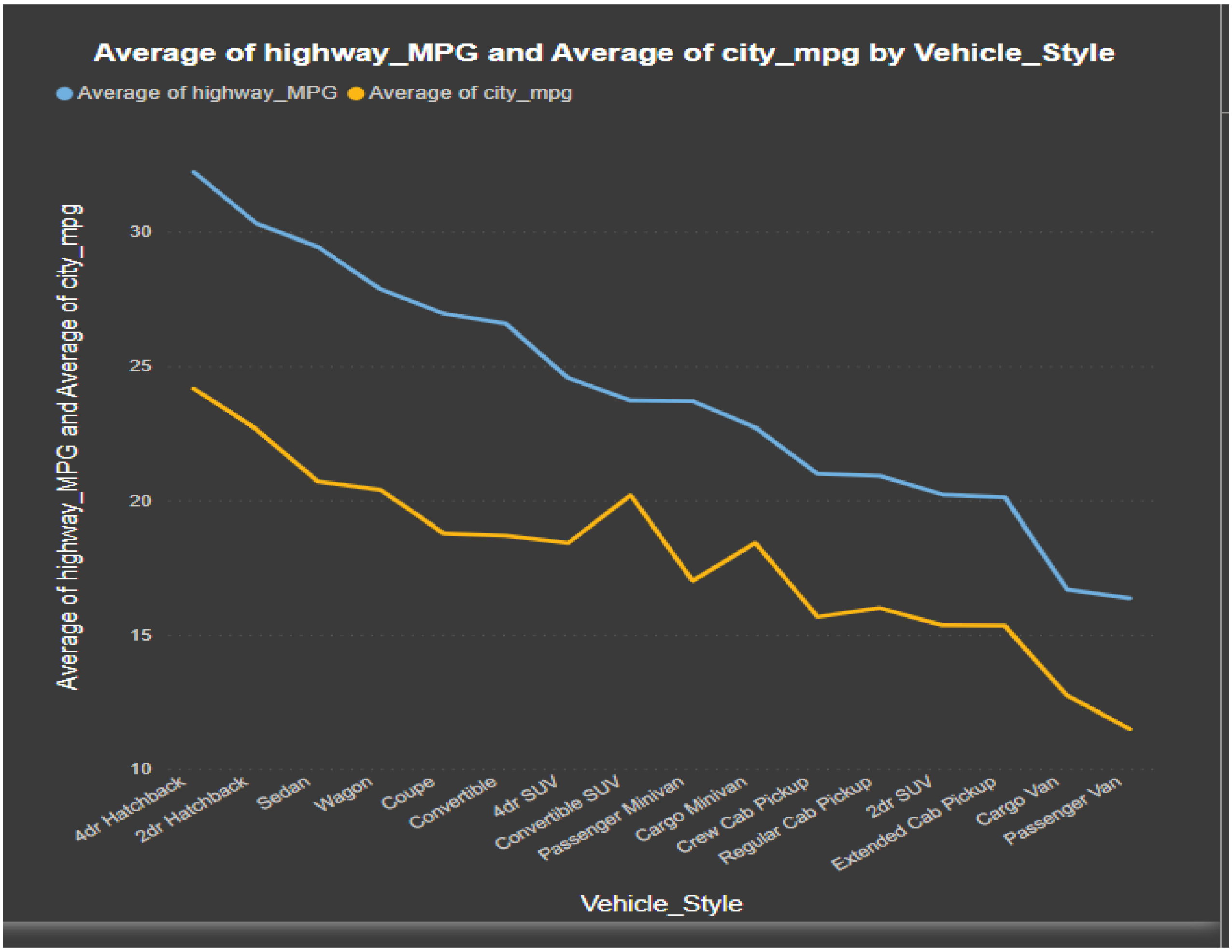
1. Cars with Automated_Manual transmission type and Coupe vehicle style has higher average MSRP values in general. Coupe cars of automated manual transmission type tends to be the most expensive with average MSRP of 110733.71

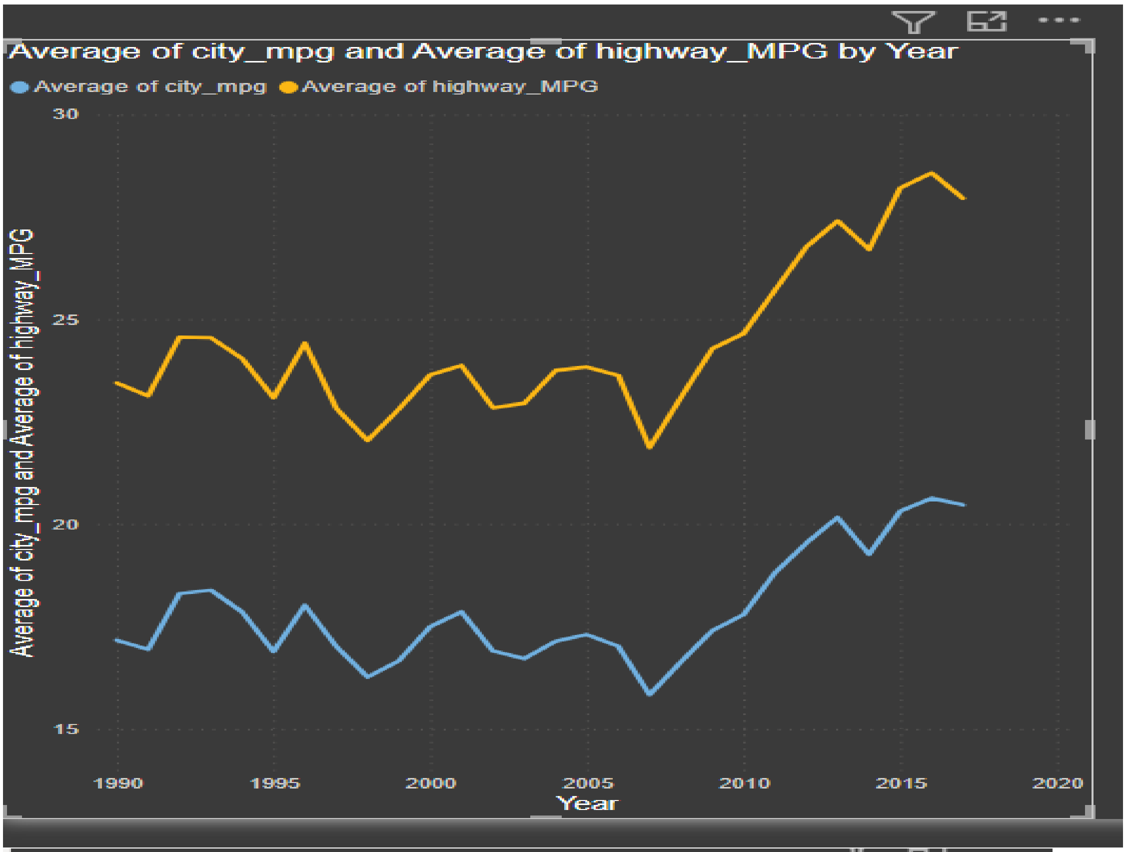
2. Cars with Unknown Transmission Type and sedan vehicle style has lowest average MSRP values. Here irrespective of the body style all the Cars with Unknown Transmission Type all are less expensive.
3. Cars with Convertible body style are very expensive for all categories of Transmission type except unknown.

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

Line chart to show the trend of fuel efficiency (MPG) over time for each body style. Calculate the average MPG for each combination of body style and model year using AVERAGEIFS or Pivot Tables.

Approach: Go to the "Visualizations" pane and select "Line Chart" from the list of chart types. Drag and drop the "Year " and “Vehicle style” columns into the X-Axis section of the chart. Drag and drop the "Highway mpg" and “city mpg” columns into Y-axis section of the chart. change the aggregation of highway mpg and city mpg columns to average. Now customize the chart using format visual giving title, changing font size etc. Also we need create a pivot table. Go to the "Visualizations" pane and select "Matrix” from the list of chart types. Drag and drop the " Vehicle style " column into the Columns section, the " Year " column into the Rows section of the chart, the "Highway mpg" and “city mpg” column into values section with aggregate as Average to get the average mpg values.


Average of highway_MPG and Average of city_mpg by Vehicle_Style
Average of highway_MPG
Average of city_mpg
Average of highway_MPG and Average of city_mpg
30
25
20
15
10
4dr Hatchback
2dr Hatchback
Sedan
Wagon
Coupe
Convertible
4dr SUV
Convertible SUV
Passenger Minivan
Cargo Minivan
Crew Cab Pickup
Regular Cab Pickup
2dr SUV
Extended Cab Pickup
Cargo Van
Passenger Van
Vehicle_Style



Vehicle_Style Year	2dr Hatchback Average of highway_MPG	Average of city_mpg	2dr SUV Average of highway_MF
1990	27.00	20.50	20.
1991	27.60	20.00	16.
1992	29.11	22.11	18.
1993	27.88	21.15	18.
1994	27.05	20.42	18.
1995	25.50	18.00	15.
1996	26.00	18.00	26.
1997	23.60	17.20	22.
1998	23.20	17.20	26.
1999	25.50	19.33	18.
2000	28.60	21.40	18.
2001	29.00	22.29	18.
2002	25.25	17.00	
2003	29.75	22.00	19.
2004	29.71	22.29	18.
2005	30.33	22.56	18.
2006	27.25	19.67	
2007	25.09	17.73	
2008	26.43	18.86	
2009	29.00	20.25	
2010	27.13	19.00	
2011	27.83	19.83	
2012	30.21	21.36	
2013	31.40	22.20	
2014	33.60	25.67	
2015	34.81	26.44	30.
2016	34.60	26.33	30.
2017	33.07	25.80	29.
Total	30.32	22.66	20.

Insights:

1. Cars with 4dr hatch back as body style is having the highest fuel efficiency as it has largest average mpg values in highway(32.23 miles per galon) and city(24.16 mpg), while those with passenger van is having the lowest mpg values in city (11.59)and highway(16.34).

2. The average highway and city mpg has increased over the years. For example, in 1990 the average highway mpg was 24.63 and the average city mpg was 17.05, while in 2016 it reaches the maximum where the average highway mpg increased to 28.57 and the average city mpg increased to 20.63.

3. From 1992 to 1993 there was no much significant change that is the city and highway mpg values remain constant. Same constant behaviour is seen from 2004 to 2005 to 2006 years where the city mpg values around 17.1 and highway mpg values around 23.75. Years of significant increase in mpg values are from 1995-1996, 2014-2015. There is a steep continuous increase in mpg values from 2007(city=15.84, highway=21.87) to 2013 (city=20.16, highway=27.40) after that it drops in 2014.

4. Across all the years, Sedan is the vehicle style that appears in the highest number of times with the highest average highway mpg, with a maximum average highway mpg of 25.83 in 1996.The Convertible SUV is the vehicle style that appears in the highest number of times with the highest average city mpg, with a maximum average city mpg of 24 in 1993. The SUV and Pickup styles have the lowest average highway and city mpg compared to other vehicle styles in most of the years.

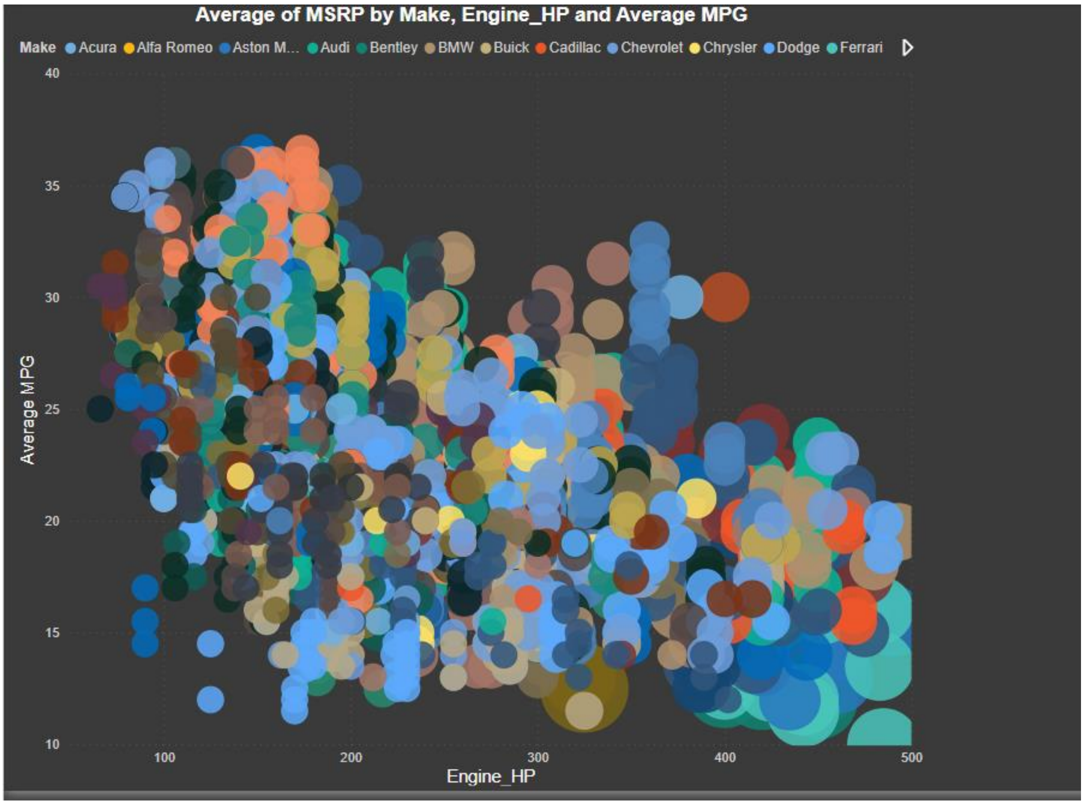
5. Overall, we can see that the trend of average mpg has increased over the years, with Sedan being the most fuel-efficient vehicle style in terms of highway mpg and Convertible SUV being the most fuel-efficient vehicle style in terms of city mpg.

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

Bubble chart to visualize the relationship between horsepower, MPG, and price across different car brands. Assign different colors to each brand and label the bubbles with the car model name. Calculate the average horsepower, MPG, and MSRP for each car brand using AVERAGEIFS or Pivot Tables.

Approach: A bubble graph is used to visualize data set with three dimensions. It is quite related to the Scatter chart. A scatter chart has two value axes. One numerical data long horizontal axis and one along the vertical axis. A scatter chart shows the relationship between the two numerical values. And Bubble chart replaces the data points with bubble size representing the third data dimension. Basically, the Bubble chart represents three sets of data in a graph. One is X-axis coordinate, second is Y-axis coordinate and the final is the Bubble size data set. First we have created a calculated column taking the average of the city mpg and highway values and named this column as Average MPG. that is Average MPG =(city mpg + highway mpg)/2. We drag engine hp to x axis , Average MPG the calculated column to y axis, make to legend , MSRP to size section of the chart. Change the aggregation of MSRP to average.By adding the size parameter we can create a bubble chart that calculates the relationship between the 3 variables. . Now customize the chart using format visual giving title, changing font size etc. Also we need create a pivot table. Go to the "Visualizations" pane and select "Matrix" from the list of chart types. Drag and drop the " Make " column into the Rows section of the chart, the "MSRP" and "ENGINE HP","Average MPG" columns into values section with aggregate as Average to get the average values of all 3 columns.

Make	Average of MSRP	Average of Engine_HP	Average of Average MPG
Acura	34,617.37	244.72	24.00
Alfa Romeo	61,600.00	237.00	29.00
Aston Martin	1,47,543.64	436.25	16.18
Audi	46,279.00	260.17	23.96
Bentley	2,87,832.67	436.67	12.30
BMW	56,835.98	309.62	24.06
Buick	29,034.19	220.01	22.90
Cadillac	56,076.22	327.17	21.41
Chevrolet	28,050.02	242.46	21.77
Chrysler	26,722.96	229.14	22.06
Dodge	22,144.20	237.51	19.86
Ferrari	1,93,918.62	447.16	12.88
FIAT	22,206.02	143.56	29.97
Genesis	46,616.67	347.33	20.83
GMC	32,444.09	267.65	18.62
Honda	27,112.07	205.47	26.64
HUMMER	36,464.41	261.24	15.41
Hyundai	24,871.24	205.75	25.44
Infiniti	42,640.27	310.68	21.32
Kia	25,185.63	207.94	25.03
Land Rover	51,006.15	258.27	20.01
Lexus	46,656.08	284.68	21.92
Lincoln	44,123.16	287.79	20.15
Lotus	68,377.14	271.54	22.39
Maserati	1,12,022.06	413.58	16.60
Mazda	20,336.52	172.72	24.55
Mercedes-Benz	44,159.77	293.16	21.86
Mitsubishi	21,920.68	183.30	22.16
Nissan	27,507.27	236.87	23.19
Oldsmobile	12,843.80	179.73	21.92
Plymouth	3,296.87	133.75	23.90
Pontiac	19,800.04	192.34	22.79
Porsche	79,950.43	352.69	21.92
Rolls-Royce	3,98,605.88	414.65	14.38
Saab	27,879.81	221.17	22.07
Scion	20,263.18	159.71	28.09
Spyker	2,14,000.00	400.00	15.50
Total	34,348.73	240.28	22.71



Insights:

1. A clear trend is visible in the data that shows a positive relationship between MSRP (price) and Engine_HP and a negative relationship between Engine-Hp and average MPG In general, the higher the price of a car, the more horsepower it tends to have.

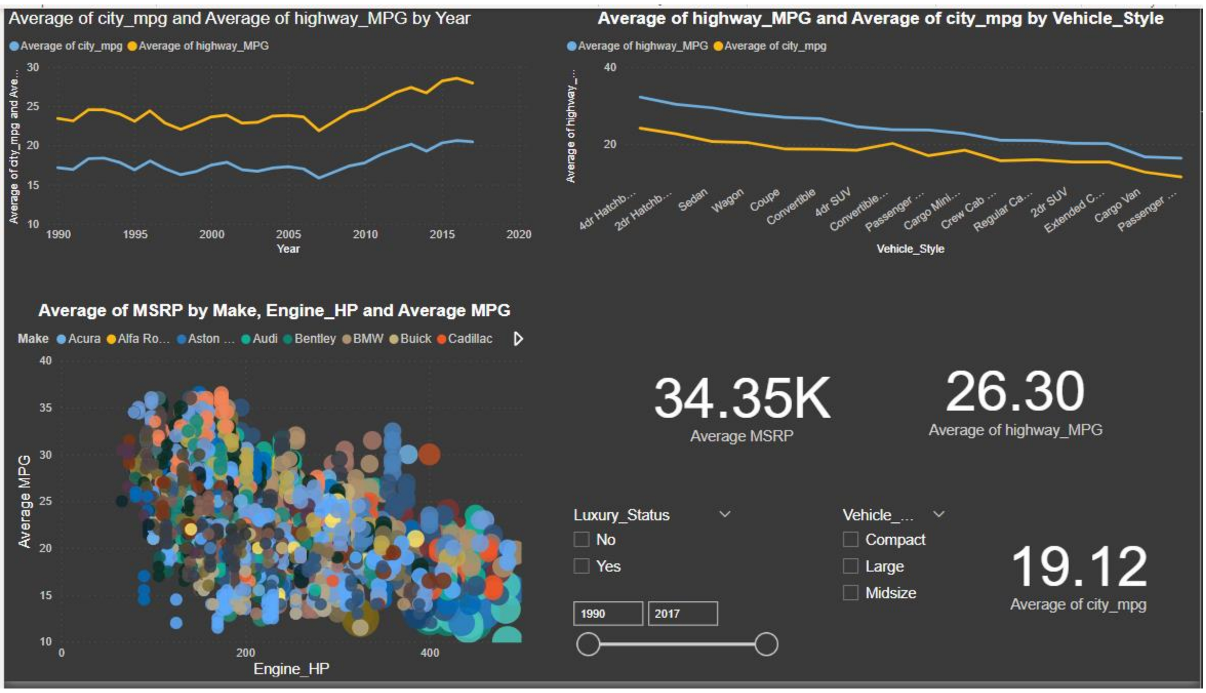
2. The most expensive cars, on average, are made by Rolls-Royce with an average MSRP of $398,605.88. Ferrari has the highest average engine horsepower at 447.16.The most fuel-efficient cars, on average, are made by FIAT with an average MPG of 29.97.

3.Land Rover has the lowest average MPG at 20.01, which means that Land Rover cars are less fuel-efficient compared to other car brands. Alfa Romeo has the highest average MPG among all brands with 29, which indicates that Alfa Romeo cars are more fuel-efficient compared to other car brands.

4. Luxury car brands such as Aston Martin, Bentley, Ferrari, Maserati, and Rolls-Royce have some of the highest average MSRP values in the dataset, indicating that they are typically associated with high prices and exclusivity.

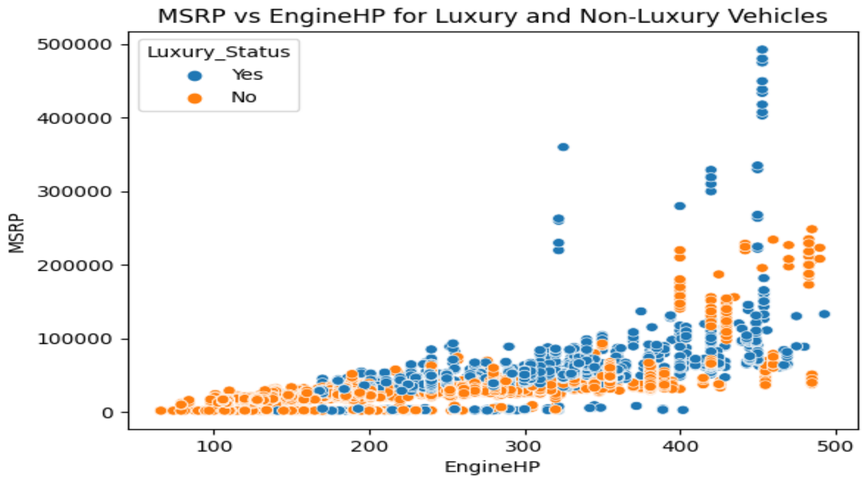
5.Popular Brands such as Chevrolet, Honda, Nissan, and Toyota have relatively lower average MSRP values compared to luxury brands, but they are more popular among consumers and have a wider range of models available.

Building Dashboard: I have used all the charts in tasks to create a dashboard. Also to make it interactive , added slicers for year, Market category, Vehicle size, luxury status etc . Also added card visuals displaying the average msrp , average city mpg and average highway mpg of the current selection.



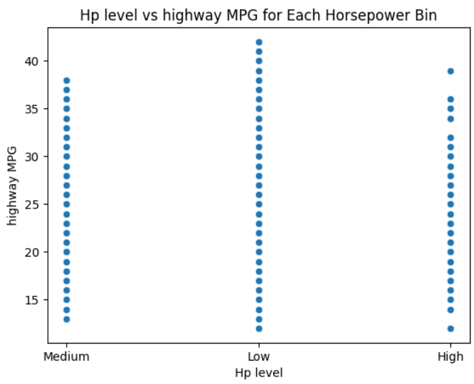
Further Analysis: I have done some further analysis on my derived columns in the dataset like Engine hp binned , luxury status etc. I will be sharing those insights also here.

```
# Scatter plot for MSRP vs EngineHP for Luxury and Non-Luxury Vehicles separately
sns.scatterplot(x='Engine_HP', y='MSRP', hue='Luxury_Status', data=cars_df)
plt.title('MSRP vs EngineHP for Luxury and Non-Luxury Vehicles')
plt.xlabel('EngineHP')
plt.ylabel('MSRP')
plt.show()
```



So for luxury and non luxury vehicles, engine hp and msrp holds a positive relationship. Also, it seems that there is a difference in the relationship between Engine HP and MSRP for luxury and nonluxury vehicles, especially for Engine HP values between 400 and 500. Specifically, for Engine HP values in this range, luxury vehicles have higher MSRP values compared to nonluxury vehicles.possible explanation could be that luxury vehicles are generally priced higher than non-luxury vehicles due to various factors such as brand image, quality of materials used, and additional features, and the difference in MSRP between luxury and non-luxury vehicles for the same Engine HP value could be attributed to these factors.

```
# Scatter plot for MSRP vs City MPG for each horsepower bin
sns.scatterplot(x='Engine_HP_Binned', y='highway_MPG', data=cars_df)
plt.title('Hp level vs highway MPG for Each Horsepower Bin')
plt.xlabel('Hp level')
plt.ylabel('highway MPG')
plt.show()
```



It seems that there is a variation in the relationship between Engine HP level and Highway MPG. Specifically, for the low, medium, and high Engine HP levels, the range of Highway MPG values seems to overlap, with medium and high levels having similar ranges. It's also worth noting that the range of Highway MPG values for each Engine HP level is quite wide, which suggests that there may be considerable variation in fuel efficiency even for vehicles with similar Engine HP levels.

Another possible insight is that there may be a trade-off between Engine HP and fuel efficiency, as higher Engine HP levels may come at the cost of lower fuel efficiency. This trade-off may be more pronounced for low Engine HP levels, where the range of Highway MPG values is wider compared to the medium and high Engine HP levels.the relationship between Engine HP and Highway MPG may not be straightforward, and other factors such as transmission type, weight of the vehicle, and driving conditions may also play a role.

Other insights:

```
# Create a pivot table with Engine_HP_Binned and Make as rows and count of popularity as values
Make_Hp_pivot_table = pd.pivot_table(cars_df, values='Popularity', index='Engine_HP_Binned', columns='Make', aggfunc='count')

most_popular_brands = Make_Hp_pivot_table.idxmax(axis=1)

print(most_popular_brands)
```

```
Engine_HP_Binned
High      Chevrolet
Low       Chevrolet
Medium    Chevrolet
dtype: object
```

```
# Create a pivot table with Luxury status and Make as rows and count of popularity as values
Make_Luxury_pivot_table = pd.pivot_table(cars_df, values='Popularity', index='Luxury_Status', columns='Make', aggfunc='count')

most_popular_brands_luxury = Make_Luxury_pivot_table.idxmax(axis=1)

print(most_popular_brands_luxury)
```

```
Luxury_Status
No     Chevrolet
Yes     Cadillac
dtype: object
```

```
# Create a pivot table with  Engine_HP_Binned as rows,make as columns  and mean of price as values
Make_Price_pivot_table = pd.pivot_table(cars_df, values='MSRP', index='Engine_HP_Binned', columns='Make', aggfunc='mean')

brands_price = Make_Price_pivot_table.idxmax(axis=1)

print(brands_price)
```

```
Engine_HP_Binned
High      Rolls-Royce
Low             Lotus
Medium    Rolls-Royce
dtype: object
```

1.Chevrolet is the most popular brand in all three levels of horsepower and in the non-luxury category, while Cadillac is the most popular in the luxury category, suggests that these brands have established a strong market presence across different vehicle types and horsepower levels.

2.Rolls royce is having highest average price in high and medium levels of horsepower while Lotus tends to be expensive in low level of horsepower.
With the highest average price in the high and medium levels of horsepower, Rolls Royce is likely targeting customers who are willing to pay a premium for both performance and luxury.

3.On the other hand, Lotus appears to have a different pricing strategy for their vehicles. While they are still a luxury brand, they tend to be more expensive in the low level of horsepower range. This could be because they are focusing on producing lightweight and agile sports cars that prioritize handling over raw power.

Final Conclusions:

1. Consumers prioritize practicality and functionality when purchasing cars, with crossovers being the most popular category overall.
2. There is a positive linear relationship between engine power and price, but the rate of increase slows down as engine power becomes larger, suggesting an optimal engine power level for balancing performance and cost.
3. Car age, horsepower, number of cylinders, horsepower per cylinder, and luxury status are the most important features in determining the car's price, and luxury brands tend to have higher prices compared to non-luxury brands.
4. Cars with fewer cylinders tend to have better fuel efficiency on the highway compared to cars with more engine cylinders.
5. Body style plays a significant role in the price of cars, with SUVs generally being more expensive than sedans and coupes.
6. Cars with 4dr hatchback as a body style have the highest fuel efficiency, and the average highway and city mpg values have increased over the years.

Recommendations:

Based on the above insights, here are some recommendations for car manufacturers:

1. **Electric and hybrid vehicles are gaining popularity, and manufacturers should focus on developing more efficient and affordable models to meet the increasing demand.**
2. **Alternate fuels, such as hydrogen fuel cells and biofuels, are also gaining attention, and manufacturers should consider incorporating these technologies into their product lines.**
3. **Fuel efficiency is becoming increasingly important to consumers, and manufacturers should focus on developing cars with better mpg ratings to remain competitive in the market.**
4. **Horsepower is still important to some consumers, but manufacturers should balance this with fuel efficiency to offer a more well-rounded product.**

5. **Luxury features, such as advanced technology and comfort features, are becoming more important to consumers, and manufacturers should continue to invest in these areas.**
6. **To optimize pricing decisions for profitability, manufacturers should conduct market research to understand the price sensitivity of their target market and adjust pricing accordingly.**

Results: I have learnt what are the important features in predicting the price of a car and the factors that drive consumer demand, how the features and prices of car changes over time, which all brands are most fuel efficient etc. Overall, this dataset was a valuable resource in exploring various aspects of the automotive industry and could provide insights that could inform decisions related to product development, marketing, and pricing.

THANK YOU