Impact of Car Features

Project Description:

Automotive industry is rapidly growing due to its increased demand which increases the importance to understand the factors that drive the consumer demand for cars. The main purpose of this project is to analyse the impact of various car features on the manufacturing price and to find the relationships between them. A dataset containing the details on various car features is provided.
The dataset is first cleaned by removing all the duplicates and null values. Appropriate values are replaced in place of missing values and further analysis is done on this dataset.

Approach:

I performed the analysis in a systematic manner to ensure better results by making use of all concepts that I have learned so far. For each of the tasks, I used pivot tables to get the values in an ordered and tabular form and I made use of different visualisation techniques including the scatter plot, column charts, line plot etc for better understanding.I have also performed regression analysis for one of the tasks to identify the variable with strongest relationship with the car price.

Tech-Stack Used:

I have used python and tableau for this project. Python is very useful because of its high flexibility and support. It is a free and open source programming language. It is extremely powerful.I used python to clean the data and to perform regression analysis. For the rest of the tasks, I used tableau for visualisation. Tableau is a leading visualisation tool for data analysis and business intelligence.Tableau makes the visualisation purpose very easy.

Insights:

I have gained several insights from doing this project. As this project mainly deals with the impact of car features, I saw how it affects the market price.I came to know that bugatti has the most average price compared to other brands , the average manufacturing price and the engine horsepower share a positive exponential curve which means that price is increasing with respect to engine horsepower. Most people prefer crossovers whereas hybrid performance cars have the least popularity.All these insights gained plays a crucial role in automobile industries as it helps them in many ways to increase their turnovers.

Result:

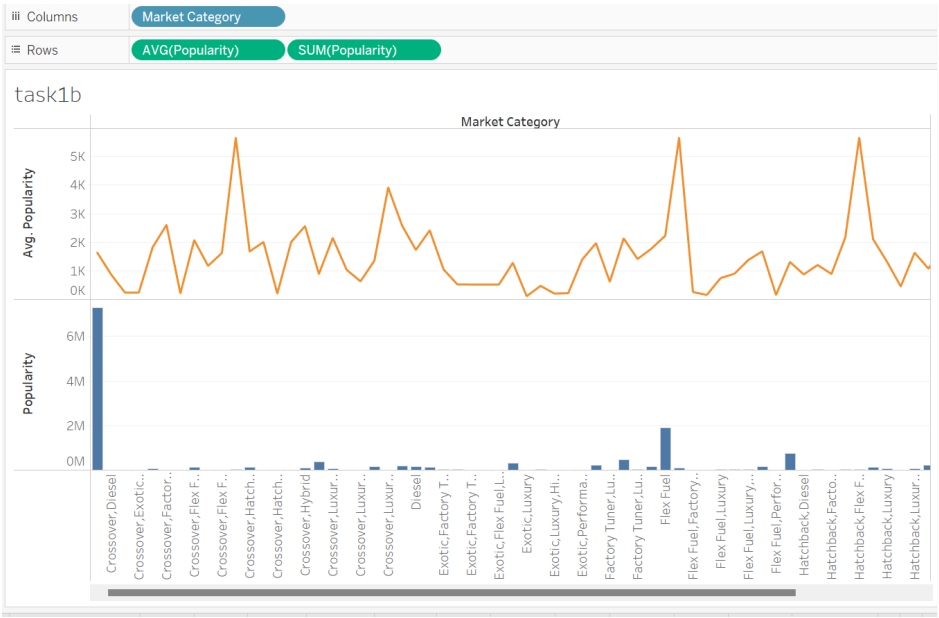
Task analysis:

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

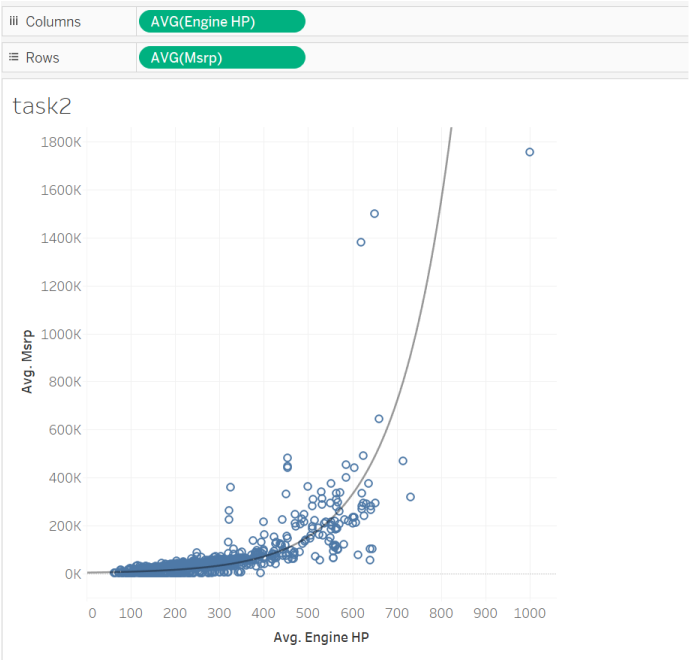
task1a

Market Cate..	Count of ..	Popularity
Flex Fuel,Hybrid	2	310
Flex Fuel,Luxury	39	29,115
Flex Fuel,Luxury,..	32	28,746
Flex Fuel,Luxury,..	28	38,642
Flex Fuel,Perfor..	87	146,201
Flex Fuel,Perfor..	2	310
Hatchback	574	751,167
Hatchback,Diesel	14	12,222
Hatchback,Facto..	13	15,667
Hatchback,Facto..	9	7,982
Hatchback,Facto..	21	45,648
Hatchback,Flex F..	7	39,599
Hatchback,Hybrid	64	135,114
Hatchback,Luxury	45	59,541
Hatchback,Luxur..	3	1,362
Hatchback,Luxur..	36	58,761
Hatchback,Perfo..	198	212,585
High-Performance	198	361,029
Hybrid	121	256,107
Luxury	819	883,877
Luxury,High-Perf..	334	557,118
Luxury,High-Perf..	12	6,826
Luxury,Hybrid	52	35,029
Luxury,Performa..	659	852,128
Luxury,Performa..	11	25,665
Performance	520	735,909
Performance,Hy..	1	155

- **Task 1.B:** Create a combo chart that visualizes the relationship between market category and popularity.



- **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.



- **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.

```
In [31]: y = df['MSRP']

In [32]: x = df[['Engine HP','Engine Cylinders','Number of Doors','highway MPG','city mpg','Popularity']]

In [33]: x = sm.add_constant(x)

In [34]: model = sm.OLS(y, x).fit()
         summary = model.summary()

In [35]: print(summary)
```

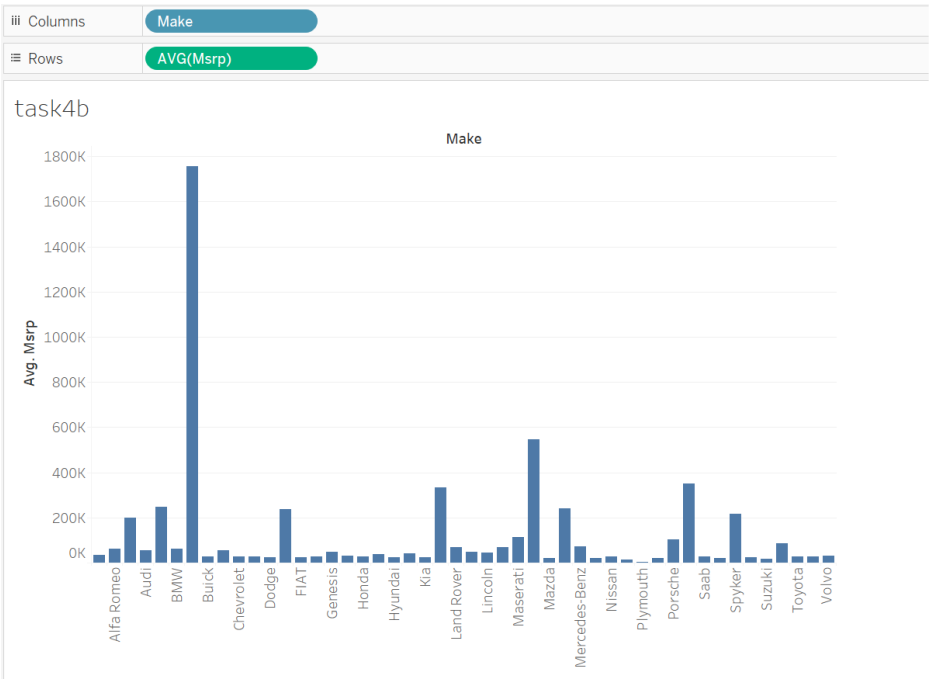
```
                            OLS Regression Results
==============================================================================
Dep. Variable:                   MSRP   R-squared:                       0.461
Model:                            OLS   Adj. R-squared:                  0.461
Method:                 Least Squares   F-statistic:                     1598.
Date:                Wed, 11 Oct 2023   Prob (F-statistic):               0.00
Time:                        15:41:18   Log-Likelihood:            -1.3592e+05
No. Observations:               11199   AIC:                         2.719e+05
Df Residuals:                   11192   BIC:                         2.719e+05
Df Model:                           6
Covariance Type:            nonrobust
==============================================================================
```

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

task4a

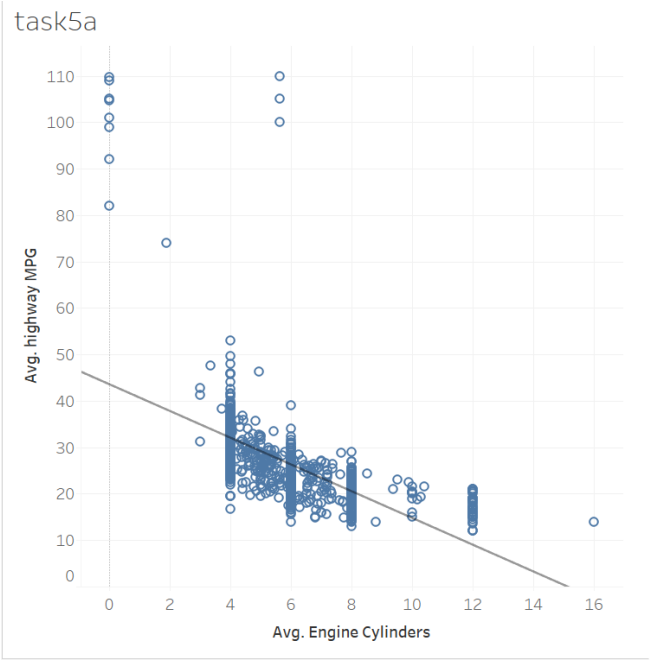
Make	
Acura	35,087
Alfa Romeo	61,600
Aston Martin	198,123
Audi	54,574
Bentley	247,169
BMW	62,163
Bugatti	1,757,224
Buick	29,034
Cadillac	56,368
Chevrolet	29,075
Chrysler	26,723
Dodge	24,857
Ferrari	238,219
FIAT	22,670
Ford	28,511
Genesis	46,617
GMC	32,444
Honda	26,655
HUMMER	36,464
Hyundai	24,926
Infiniti	42,640
Kia	25,514
Lamborghini	331,567
Land Rover	68,067
Lexus	47,549
Lincoln	43,861
Lotus	68,377

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



- **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.

task5a

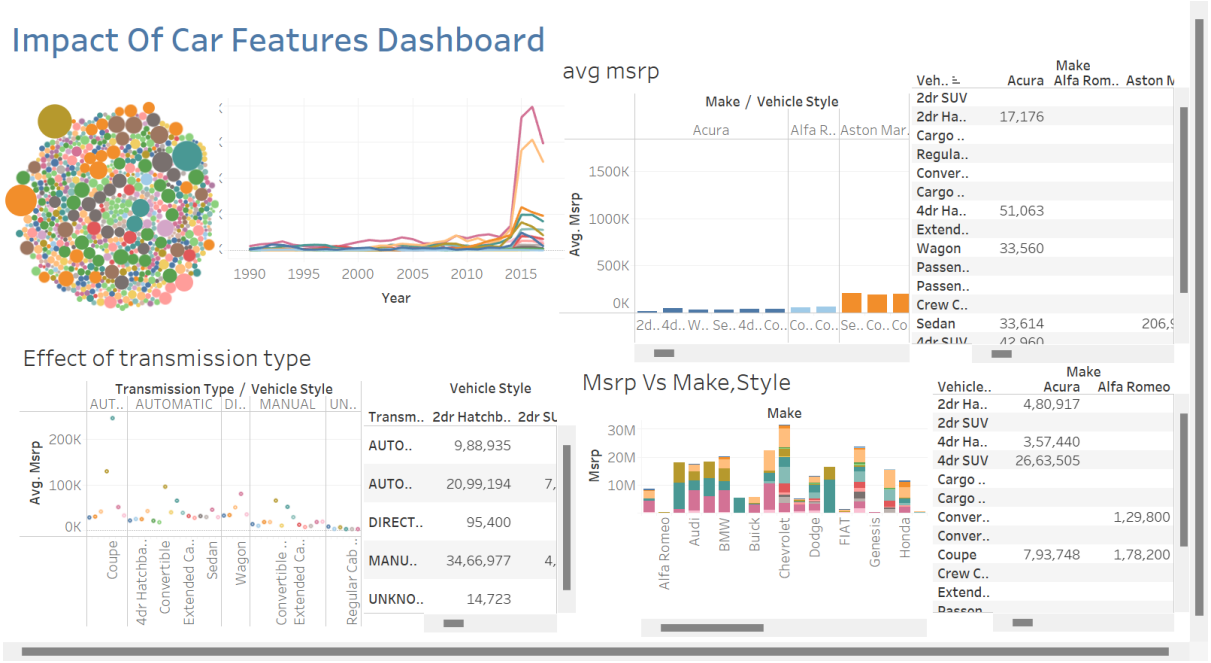


- **Task 5.B:** Calculate the correlation coefficient between the number of cylinders and highway MPG to quantify the strength and direction of the relationship.

task5b

Make	
Acura	-0.525
Alfa Romeo	
Aston Martin	-0.505
Audi	-0.224
Bentley	0.079
BMW	-0.701
Bugatti	
Buick	-0.719
Cadillac	-0.854
Chevrolet	-0.615
Chrysler	-0.595
Dodge	-0.649
Ferrari	-0.406
FIAT	-0.973
Ford	-0.739
Genesis	-0.803
GMC	-0.590
Honda	-0.674
HUMMER	-1.000
Hyundai	-0.713
Infiniti	-0.582
Kia	-0.732
Lamborghini	-0.857
Land Rover	-0.854
Lexus	-0.722
Lincoln	-0.739
Lotus	-0.317

Dashboard:



Doing this project has proven to be very beneficial for me as it got me started with tableau. Now all the visualisation process seems very easier.I have gained one more relevant skill.I learned to do regression analysis in python . Building regression models, getting coefficients , identifying the relationships and significant patterns has become very familiar to me.I also explored a lot of different charts while doing the analysis. Overall doing this project has sharpened my skills and equipped with more as a data analyst .