Project 7

Impact of Car Features on Price and Profitability

Project 7: Task Analysis
Project 7: Creating dashboard

Project description:

Project deals with how a manufacturer can increase their profitability in the market by meeting the consumer demand. These days consumers of car purchasers are young who look for all the features if they are going to buy a brand new car. Not only brand matters to them, but they are also concerned about comfortability and environmental safety. Hence, besides environmental stability, manufacturers have to focus on the factors such as fuel efficiency and technological innovations, which attracts their consumer, including the way they advertise.

Description data used in the project:

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

Before performing any work with the data, it has been cleaned by removing the duplicates, so that results obtained will be fair.

Assumption made before starting any work: Honda is the popular brand among buyers.

Approach:

Pivot tables have been utilized for each question, allowing for a structured analysis of the raw data. Excel functions, along with add-ins like "Data Analysis," have been employed for regression analysis. Furthermore, filters and slicers have been implemented to refine the data visualization process. Charts have been meticulously crafted for every dataset to enhance clarity and aesthetics.
Challenges poses:
1) Editing cannot be made in the pivot table, hence for every chart, the pivot data has been rearranged as normal data.
2) Sometimes due to mistakes in the data, charts did not show the desired data.
3) Small and large quantity data cannot be put on the same chart.

Tech-Stack Used:

1) A complete excel based project, made on office 365.
2) This tech stack is used due to its user friendly nature. It was easy to access the raw data in excel because data itself is provided in the .xlsx format, hence cleaning, analysis, and giving output using excel was easier.
3) Trainity itself has provided a full lecture on using excel, which helped a lot.

Tasks Assigned: Analysis

(1) 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.

Approach:
1.A: a pivot table including Car models and its popularity for each market category has been shown in sheet **Task 1.A**. before turning it into a pivot table, i converted the data given to a table, then selected the whole table(A1:P11200) and clicked in “insert” and created a pivot table for it.
Then mentioned market category and car models in rows area and their popularity in values area. In this way we figured out what type of category of car is most populous.
From task 1.B,, we figured out that **the maximum popular market category has type Crossover, exotic, Luxury, Performance (32%). (Task 1.B)** Count of car models with respect to market category has also been show in **Car Data** Sheet.

(2) 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.

Approach: In the **sheet task 2,** first of all from the parent table the **cars MRP** and **Car HP** have been separated and listed out. After selecting them both I chose the scatter plot chart to plot them on the table. The x-axis is showing the car's power and the y-axis is showing cars' price. We see here, with the increasing cars HP, prices have increased. **Car’s Power and Car’s Price are proportionally related.**

(3)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.

Approach: Following steps are followed:
1) Work done in **sheet Task 3.**
2) Collected the numeric variables which affect the car’s price positively and negatively.
3) Data is the car's power, no. of cylinder used, highway and city mileage.
4) Search for any non numeric data present in the data.
5) Search for blank spaces.
6) Deleted the blank spaces, and prepared the final data to work on.
7) **Installed “Data Analysis” from File>Options>Add-ins>Manage>Excel Add-ins> Check box to Data Analysis> click OK.**

8) In Data, in the help bar, click Data Analysis. I chose **“Regression”** as the type of analysis to be performed.
9) Input the Y range (dependent variable), i.e., Car’s Price and the X range (independent variation), i.e., car's power, no. of cylinder used, highway and city mileage, one at a time.
10) Put Output range (the cell where you want your output).
11) Marked the labels included, and pressed OK.
12) A chart consisting of all the variations with Y range due to X range appeared.
13) Separated the Regression line, Correlation coefficients, y intercept, R square and Total Variation.
14) Correlation coefficients vary from -1 to +1. Coefficients near +1 show a strong relation of dependent variable (Car’s Price) with independent variables (stated above). Coefficients near -1 show a strong negative relation of them. Coefficients near 0 show least variations.
15) Collected all the coefficients and showed them in the **Bar chart**. It shows, **Car’s Price are significantly dependent on Car’s Power and negatively relates with car’s mileage in Highway and City.**

(4)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.

Approach: Following steps were followed:
1) Work done in sheet Task 4
2) Pivot table was created which includes **Car’s model in the Rows area, and MRP in Values area. Accordingly, the year of making was shown in the rows area.**
3) With the help of field setting, the sum of MRP (for each year, which is default) is changed to Average MRP.
4) Now the table is showing Car’s model, with each year’s average car price.
5) Car’s price with year is copied as normal range, and with the help of it, **Scatter Plot and Bar chart** was created, which shows how the maker company has made their profit each year, hence increased price.

(5)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.

Approach: Following steps have been followed:

1) Pivot table showing Engine fuel type and Highway Mileage on the rows area, and count of cylinder on Values area.
2) Selected the column of Highway Mileage and count of cylinders.
3) Made a scatter plot where No. of cylinder are on X axis and Highway mileage on Y axis.
4) Their slopes were taken.
5) Correlations were done on Highway Mileage and count of cylinders.
6) No or least negative relation between Highway Mileage and count of cylinders in diesel, electric, flex fuel premium unleaded recommended, flex fuel premium unleaded required and premium unleaded type fuels.
7) Strong negative relation between Highway Mileage and count of cylinders in Regular Unleaded type fuel.

Tasks Assigned: Building a Dashboard

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

Approach: Steps followed:

1) Pivot table was made which includes Car brand in rows area, car style in column area, and Car MSRP in values area.
2) We use the SUMIF function or Pivot Tables for this. We use a formula like =SUMIF(A:A, "Brand Name", C:C) to calculate the total MSRP for each brand. Do the same for each car style.
3) Select the pivot table, go to insert.
4) Select column chart, and it will give you a chart on Car Brand and Sum of MSRP of Car's style.
5) Next we made a slicer on car MSRP, car brand and car style; Click on each to get filtered data.

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

Approach: Following steps were followed:

Unique Brands	Increasing order of AVG MSRP
2dr SUV	14306.54945
2dr Hatchback	16063.15159
Cargo Van	17019.29762
Regular Cab Pickup	17854.64928
Convertible SUV	17975
Cargo Minivan	20292.93103
4dr Hatchback	22061.31925
Extended Cab Pickup	23041.77219
Wagon	26084.29084
Passenger Minivan	26152.10417
Passenger Van	30578.06612
Crew Cab Pickup	37183.11145
Sedan	40151.95589
4dr SUV	40736.35037
Coupe	78480.15931
Convertible	88216.79217

1) Data like car maker, car style and their MSRP were chosen.
2) Made a unique table of car styles, and accordingly their average MSRP were taken with the help of AVERAGEIF Function.
3) Sorting was done of the average MSRP data of cars to know what is the maximum and minimum average MSRP of each car style.
4) Maximum car models are Sedan and minimum car models are Convertible SUV.
5) Next, the pivot table was made, where car brand and vehicle style are in the rows area and average of MSRP in values area.
6) Here we could see the increasing or decreasing MSRP of each car.
7) Highlighted the **maximum Average price with Green box and Lowest average price with Red box. Maximum is for coupe models from bugatti and minimum from 2dr Hatchback.**
8) Column charts were created showing the decreasing average price of some of the car brands.

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

Approach: Following are the steps followed:

1) Pivot table was made using the car data. The transmission type and car style are on the row area while the average MSRP was on values area.

2) Found the highest and lowest average MSRP by MAX & MIN function respectively.
3) For each transmission type, average price was recorded according to each brand.
4) We found the **maximum average price in Automated_manual type transmission.** I.e., 245977.42 and the **minimum of average MSRP 2000 in unknown transmission type.**
5) For each type of transmission type, Column charts and scatter charts are shown.
6) Each column chart shows their maximum and minimum average MSRP.

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

Approach: following steps were followed:

1) Pivot table was made using the car data. The year and car style are on the row area while the average car MPG (highway and city) are on values area.
2) Line chart showing average car MPG (highway and city) are on values area for each car style is shown.
3) Simply “change field” is applied to convert default sum() into avg() of Car MPG.
4) Highest average MPG for each year have been highlighted by conditional formatting. **We can see that maximum MPG is given by car style Sudan.**

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

Approach: Following steps were followed:

1) Pivot table was made using the car data. The car brand is on the row area while the average car HP, Average highway MPG, Average Car MSRP are on values area.
2) Data was copied to be modified.
3) Bubble chart was created to compare between 3 data.

Insights:

Working on critical data analysis for a car company involves diving deep into various aspects such as car brand, 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). It's a fascinating yet demanding task that requires extensive use of Excel functions, pivot tables, and careful examination of the data. This project has been both challenging and rewarding, pushing me to explore and experiment with multiple functions to extract meaningful insights. It's undoubtedly going to be a standout addition to my resume, showcasing my analytical skills and proficiency in handling complex datasets.

Result:

I Learned about excel functions and also excel add-ins like Data Analysis, which helped me in climbing one more step toward becoming a data analyst.