Chase Rawson

DATA 520.1OL - Data Mining/Cleaning/Visualization

March 22, 2023

New Cars Price 2019

Finance and Travel Industries

The following data set I chose contains 32,317 observations of car models manufactured in the United States between 1990 and 2019. Each row of data includes its original MSRP, MPG for cars that are not electric, engine type, weight, horsepower, torque, year manufactured, manufacturer, and whether it is a car or an SUV. This dataset will be used in the scenario of a used car dealership, analyzing the differences between the car make and each characteristic listed above. As part of the used car dealerships’ operations, they acquire cars and SUVs from auctions at low prices and sell them at higher prices to make a profit. Those who consider themselves car enthusiasts look for cars to buy based on those above characteristics and an affordable price. Used car dealerships look to keep their clientele years down the road. However, they need a data-driven way to purchase vehicles at auction that would keep these enthusiasts impressed, which most times can be challenging as the Auto Remarketing magazine described, “dealers still scrambling for inventory to others experiencing a softening of demand and the beginning return of margin compression.” (Rubenoff, 2023). As a member of an investment group looking to start a chain of used car dealerships, our mission is to ensure transparency between our customers and management, providing them with options of the best vehicles on our lots, ensuring a positive relationship while adding profit to build the dealership of the future.

I chose this data set to discover which cars are the most affordable to acquire at an auction by the MSRP. I want to compare the MSRP of different car models to determine which carmaker will fulfill this goal. I am also analyzing whether I should acquire cars or SUVs, and determine which affordable models have a higher MPG rating, year manufactured, horsepower, torque, and engine type. With this dataset, I believe that all the answers I am looking for will be provided with the following KPIs I will be using.

**Manufacturer by MSRP**

* KPI from a customer perspective.
* I chose this KPI to determine which car manufacturers have the lowest average MSRP rating within the dataset.
* This KPI is being measured to determine which manufacturers my investors should focus on acquiring to run successful dealerships.
* This measure is based on the average MSRP of all car manufacturers listed in the dataset.
* The target I hope to achieve with my data is to find the car manufacturers with the lowest average MSRP and the highest average MSRP to determine the cars we should avoid acquiring.
* All raw data will come from Kaggle from the “New Cars Price 2019” dataset. I will use the columns MSRP and manufacturer in the dataset to determine the lowest and highest average MSRP. The results will be shown on a bar chart, with each bar showing each manufacturer’s average MSRP.
* This KPI will relate to the firm’s goal by providing our potential customers with the most affordable cars they can purchase. Acquiring cars at auction without regard to their average price would not make our potential business profitable.
* This KPI shows that sportscars like Lamborghini, Maserati, Tesla, and Lotus are likely to cost more than a vehicle that is more likely to be seen on the road.

**Category by MSRP compared with MPG**

* KPI from a customer perspective.
* I chose this KPI because I want to determine which category of vehicles, either cars, pickups, SUVs, or vans, have the lowest average MSRP rating within the dataset.
* This KPI is being measured to determine which category of vehicles my investors should focus on acquiring to run successful dealerships.
* This measure is based on the average MSRP of all vehicles listed in the dataset and compares the average MSRPs of each vehicle category.
* The target I hope to achieve with my data is to find the vehicle category with the lowest average MSRP and the highest average MSRP to determine which category we should focus on acquiring at auction.
* All raw data will come from Kaggle from the “New Cars Price 2019” dataset. I will use the columns titled MSRP, category, and MPG in the dataset to determine the lowest and highest average MSRP. The results will be shown on a bar chart, with each bar showing each vehicle category’s average MSRP.
* This KPI will relate to the firm’s goal by providing our potential customers with the most affordable cars they can purchase. Each vehicle category is manufactured differently, affecting its average MSRP depending on labor and materials.
* In this KPI, pickups are most likely to cost less than a car and have lower than average MPG ratings.

**Model Year by MSRP compared with MPG**

* KPI from a customer perspective.
* I chose this KPI to determine the average MSRP rating within the dataset by the year the car was manufactured.
* This KPI is being measured to determine which manufacturers my investors should focus on acquiring to run successful dealerships.
* This measure is based on the average MSRP by the year manufactured in the dataset. I am also measuring which year has the highest and lowest average MPG ratings.
* The target I hope to achieve with my data is to find vehicles with the lowest average MSRP and the highest average MSRP by the year it was made to determine the cars we should avoid acquiring.
* All raw data will come from Kaggle from the “New Cars Price 2019” dataset. In the dataset, I will use the column titled MSRP and the column titled Model Year to determine the lowest and highest average MSRP. The results will be shown on a bar chart, with each bar showing the average MSRP by the year it was manufactured.
* This KPI will relate to the firm’s goal by providing our potential customers with the most affordable cars they can purchase. Acquiring cars at auction without regard to their average price would not make our potential business profitable.
* In this KPI, cars before 1995 do not have a recorded MPG rating. That year, as well as in 2008, was the lowest MPG rating for vehicles sold in those two years, while cars sold more recently tend to have higher MPG ratings.

**Engine by MSRP compared with MPG**

* KPI from a customer perspective.
* I chose this KPI because I want to determine which type of engine has the lowest average MSRP rating within the dataset.
* This KPI is being measured to determine which engine types my investors should focus on acquiring to run successful dealerships.
* This measure is based on the average MSRP of all vehicles listed in the dataset by each engine type listed in the dataset.
* The target I hope to achieve with my data is to find the engine type with the lowest average MSRP and the highest average MSRP to determine which category we should focus on acquiring at auction; as well as comparing each engine type with the average MPG rating.
* All raw data will come from Kaggle from the “New Cars Price 2019” dataset. In the dataset, I will use the columns titled MSRP and MPG and the column titled engine to determine the lowest and highest average MSRP. The results will be shown on a bar chart, each bar showing each engine type’s average MSRP.
* This KPI will relate to the firm’s goal by providing our potential customers with the most affordable yet thrilling cars they can purchase. Each vehicle engine is manufactured differently, affecting its average MSRP depending on labor and materials.
* In this KPI, we notice that engine types that are least likely to be found in a sportscar tend to have a higher MPG rating and lower than-average MSRP.

**Manufacturer by MPG**

* KPI from a customer perspective.
* I chose this KPI to compare each vehicle manufacturer by the average MPG.
* This KPI is being measured to determine which vehicle manufacturers my investors should focus on acquiring to run successful dealerships.
* This measure is based on the average MPG of all vehicle manufacturers listed in the dataset.
* The target I hope to achieve with my data is to find the car manufacturers with the highest average MPG and the lowest average MPG to determine the cars we should avoid acquiring.
* All raw data will come from Kaggle from the “New Cars Price 2019” dataset. In the dataset, I will use the column titled EPA Fuel Economy Est - City (MPG) and the column titled Manufacturer to determine the lowest and highest average horsepower a specific vehicle has. The results will be shown on a bar chart, with each bar showing each manufacturer’smanufacturer’s average horsepower.
* This KPI will relate to the firm’s goal by providing our potential customers with the most thrilling cars they can purchase. In order to keep a positive relationship between a customer and management, we need to find vehicles that would suit the customer’s needs. Vehicles with higher MPG ratings will provide a more thrilling experience to our customers; longer rides without filling up the gas tank.
* For this KPI, as seen on the dashboard, we notice that Tesla does not have an MPG rating because the car manufacturer only sells all-electric vehicles. In contrast, sportscars like Maserati and Lamborghini have lower-than-average MPG ratings.

**Manufactuer by Torque and Horsepower**

* KPI from a customer perspective.
* I chose this KPI because I want to compare each vehicle manufacturer by the amount of torque and horsepower it has.
* This KPI is being measured to determine which vehicle manufacturers my investors should focus on acquiring to run successful dealerships.
* This measure is based on the average torque horsepower of all vehicle manufacturers listed in the dataset.
* The target I hope to achieve with my data is to find the car manufacturers with the highest average torque and horsepower and the lowest average torque and horsepower to determine the cars we should avoid acquiring.
* All raw data will come from Kaggle from the “New Cars Price 2019” dataset. In the dataset, I will use the columns titled SAE Net Torque @ RPM and SAE Net Horsepower @ RPM and the column titled manufacturer to determine the lowest and highest average torque and horsepower a specific vehicle has. The results will be shown on a bar chart, with each bar showing each manufacturer’smanufacturer’s average horsepower.
* This KPI will relate to the firm’s goal by providing our potential customers with the most thrilling cars they can purchase. In order to keep a positive relationship between a customer and management, we need to find vehicles that would suit the customer’s needs.
* For this KPI, as seen on the dashboard, we can view a trend where car models with more torque will have more horsepower. Another trend is that with more torque and horsepower, these car models will likely have higher MSRPs.

We can see the characteristics that make up many different car models manufactured over 30 years. The dataset comes with many different comparisons, but most importantly, we are comparing the vehicle manufacturer with its MSRP, MPG, horsepower, and torque. Also, we are comparing the MSRP by its model year, vehicle category, whether it is a pickup, van, SUV, or car, and engine type. The vehicles manufacturers that are showcased are Mitsubishi, Smart, Hyundai, Fiat, Kia, Subaru, Volkswagen, Mazda, Honda, Mini, Jeep, Toyota, Nissan, Ford, Buick, Dodge, GMC, Chevrolet, Chrysler, Ram, Volvo, Audi, Alfa Romeo, Infiniti, Lincoln, Acura, Genesis, Lexus, BMW, Cadillac, Jaguar, Land Rover, Mercedes-Benz, Lotus, Tesla, Porsche, Maserati, Aston Martin, Bentley, McLaren, Ferrari, Lamborghini, and Rolls-Royce manufactured between the years 1990 and 2019.

In the strategic plan for our organization, which consists of a group of individuals looking to invest and build a collection of used car dealerships, this data will help us determine the type of vehicles we should be acquiring for our customers. We have no experience working in the automotive industry in a previous setting, and even though our clients would have more of a say regarding the vehicle they want to purchase, this data will help us start our businesses using a data-driven method.

Using the feedback from Part B, I have modified my dashboard by providing a variety of visuals as well as merging two of my KPIs. First, at least four KPIs, Category by MSRP, Model Year by MSRP, and Engine by MSRP have been compared with MPG to determine which engine and vehicle type will likely have a higher or lower average MPG. I have also used MPG to portray the average MPG over 15 years to determine which year has the highest and lowest average MPG, as depicted in yellow. This also led to the creation of bullet charts. Previously all seven of my charts were bar graphs; I am now using two bullet charts, a combination of bar chart and line chart, and four regular bar charts. I also used the feedback from Part B to add two more slicers, one slicer comparing the years the vehicles were manufactured and a slicer that compares the vehicle categories and MPG. I have also cleaned up my dataset by adding grouping to portions of the columns; for example, one of the groups I created was to merge the data on the vehicle features, and another group merged the data on roadside assistance, among other groups. Also, since Part C, one of the changes I have made to my dashboard is that I have resized all of the graphs so that the dashboard fits the page at 100% zoom.

Two of the items I have learned from this project are the ability to organize the dataset differently and add more detail to my graphs. First, because my dataset is shown with many columns and can be extended to view, it was recommended that I create groups, which are a helpful way to view specific columns at once and can hide them so my dataset is not clustered. One group I created shows statistics from each car model, showing the passenger capacity, the number of passenger doors, and the car’s weight in one group. I also created groups that showed airbag statistics, accessory statistics such as whether the car has traction control or a backup camera, roadside assistance data, and front tire statistics. I also learned about adding detail to my graphs. With a previous dashboard that I completed, I felt that there was not enough color, the graphs were only bar graphs, and I was only comparing two columns of data for each KPI. This time, four of the six KPIs show three or more columns of data being compared. For example, on one KPI, I compare the vehicle category by MSRP (bar graph) and MPG (line graph). This helped show that cars are more likely to be more expensive and have higher MPG ratings than an SUV, vans, and a pickup. I also compared the vehicle manufacturer by torque (blue graph) and horsepower (yellow graph).

**References**

Rubenoff, S. (2023, January 24). *Exploring used-car inventory acquisition in ever-changing environment*. Auto Remarketing. Retrieved January 26, 2023, from <https://www.autoremarketing.com/trends/exploring-used-car-inventory-acquisition-ever-changing-environment>

**Raw Dataset Chosen**

See attached dataset titled “New Cars Price 2019” or <https://www.kaggle.com/datasets/prassanth/new-cars-price-2019?select=New_cars_cleaned.csv>