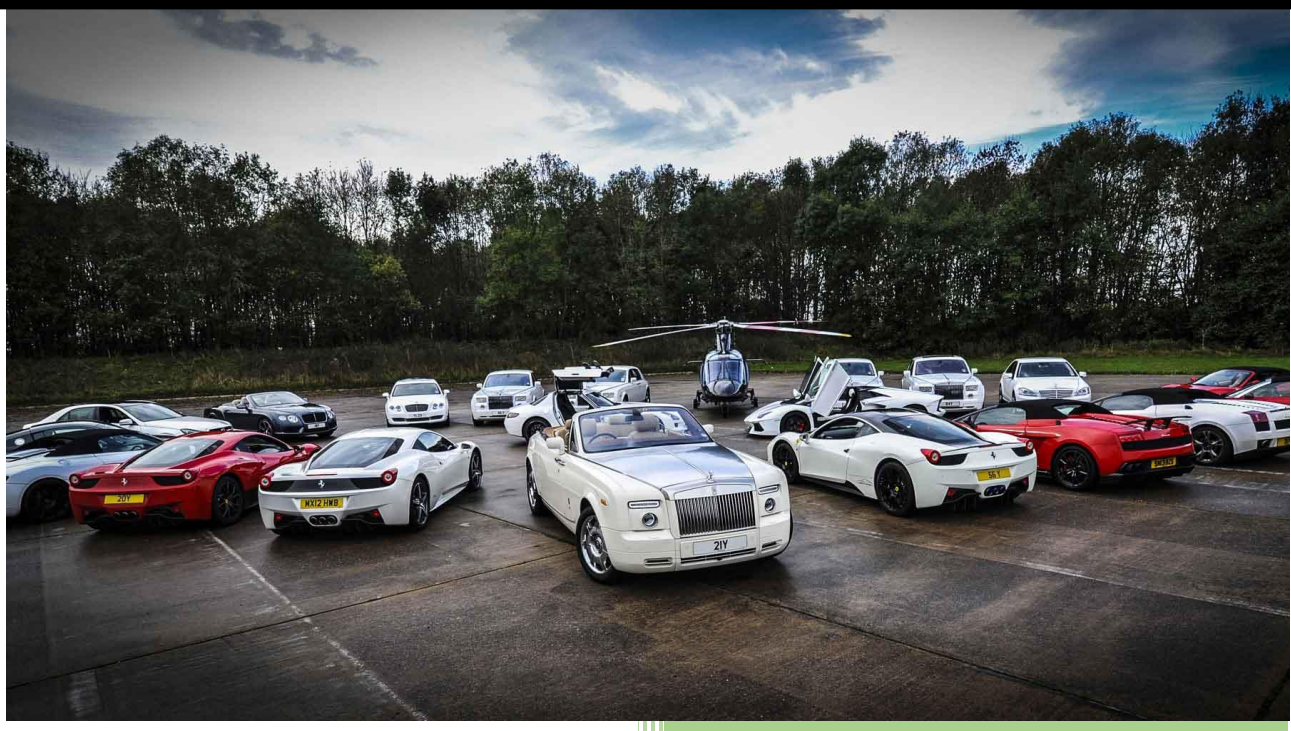
2021

Business Plan: Used Cars Dealership



VISHAL KAATAL

11/14/2021

TABLE OF CONTENTS

TABLE OF FIGURES

INTRODUCTION

The used car business is booming throughout the world because people are investing their money in properties, stocks, businesses instead of brand-new cars. (DriveNation, 2019) The brand-new car, losses over 10% of its value during the first month of purchase and up to 20% within the first year, therefore, to save these losses people are leaning towards used cars. A market survey has been conducted by Miroslav Zoricak and published the dataset called cars.csv on the Kaggle website under the name as classified ads for cars (Zoricak, 2017) which is the basis of this analysis report for setting up a used car dealership in the Czech Republic and Germany.

This report also covers the part of cleaning the datasets based on the requirement of the Business Plan. The visualization graphs have been provided to understand the data to make an investment decision. Some of the analysis questions that have been discussed in the report are as follows:

1. What is the relationship between car maker, model, and price based on car manufacturing year?
2. What is the relationship between car maker, mileage, and price?
3. What is the relationship between door count, seat count and price based on car model?
4. What kind of transmission and fuel type, do people prefer in the Czech Republic and Germany?
5. What is the top 5 vehicle manufactures recommended to invest in? Why?

DATA CLEANING

(Zoricak, 2017) The collected data has been scraped over one year and sources are completely unstructured, so as a result the data is missing values, and some values are wrong (E.g., phone numbers scraped as mileage or price). Therefore, the dataset must be cleaned so that it can answer the analysis questions described earlier. The dataset consists of roughly 3.5 million rows and 16 columns. The 16 columns consist of entities such as maker, model, mileage, manufacture year, engine displacement, engine power, body type, colour type, last emission check year, transmission, door count, seat count, fuel type, data collection date, car last seen on the website, and price in euro. Out of 16 columns, only 10 are used in the cleaning and analysis to answer the business plan.

To start a used car business, the selection of the popular car manufacturers (makers) is important such that the customers are attracted towards your dealership and ultimately make their purchase, and it gives an idea of how many cars are available in the inventory for top 10 car manufacturers in between the manufacturing year 2011 - 2015 as shown in Figure 1 and Figure 2. The manufacturing year of the car is important as the car should not be too old that it can't be operated properly or pass an inspection test. Therefore, from Figure 2, it can be noted that used car inventory is highest for the year 2015, therefore the range for the analysis can be selected just for 5 years between 2011 – 2015 when most of the data is collected to make the investment decision. Furthermore, the dataset must be cleaned to remove the unwanted values for data analysis using the hive queries mentioned in Appendix A, such as mileage should be less than 150,000 km, price should be between € 100 – € 100,000, the passenger capacity can be maximum of 7 and number of car doors can be a maximum of 5. Using these criteria, customers will be happy to select their dream car based on their budget, requirement, and purpose of their vehicle like a daily commute or high-end luxury cars for recreational purposes, etc.

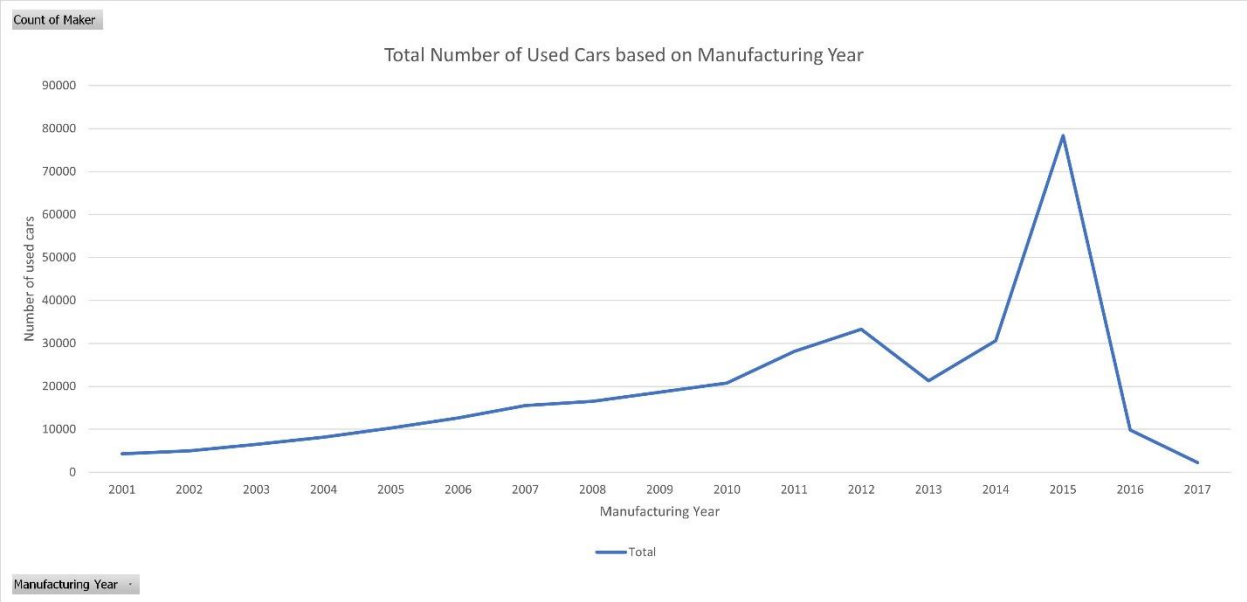


Figure 1: Total Number of used cars based on the manufacturing year

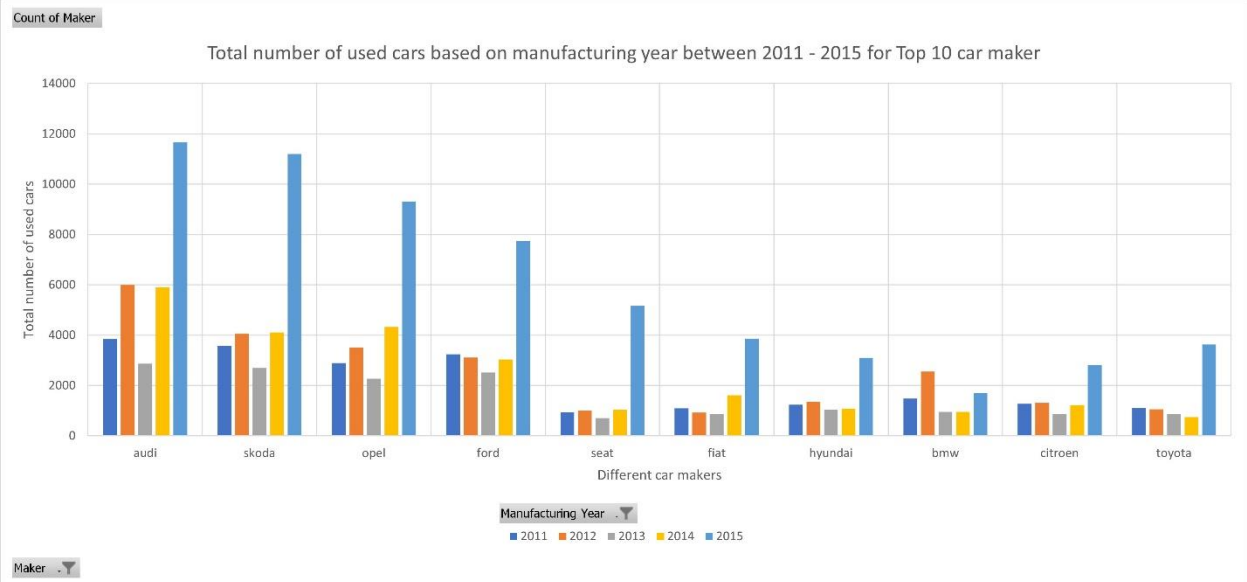


Figure 2: Total number of used cars based on manufacturing year between 2011 - 2015 for Top 10 car maker

ANALYSIS RESULT

This part of the report answers the analysis questions stated earlier to make an investment decision.

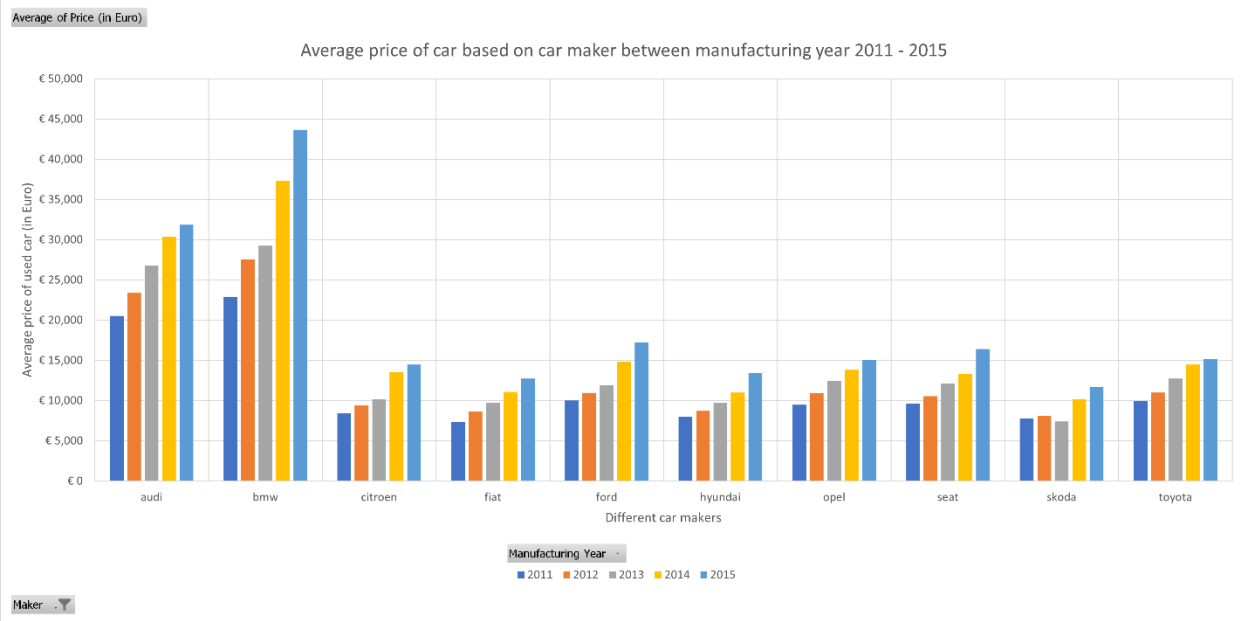


Figure 3: Price of a car based on the maker

Based on Figures 3 and 4, it can be stated that the car maker with the highest average price is BMW whereas Audi took the lead based on the highest number of car model prices as 12 out of 20 car models. Also, figure 4, shows that only 4 car makers i.e., Audi, BMW, Ford, Toyota compete for an expensive luxury car for the manufacturing year 2011 – 2015. Figure 5 represents various models of Audi, the car maker based on the average price of the car from highest to lowest. This analysis helps to understand which car maker and model to focus on to achieve higher profit and help to select the most expensive car for the dealership which, in the case of BMW is the BMW i8.

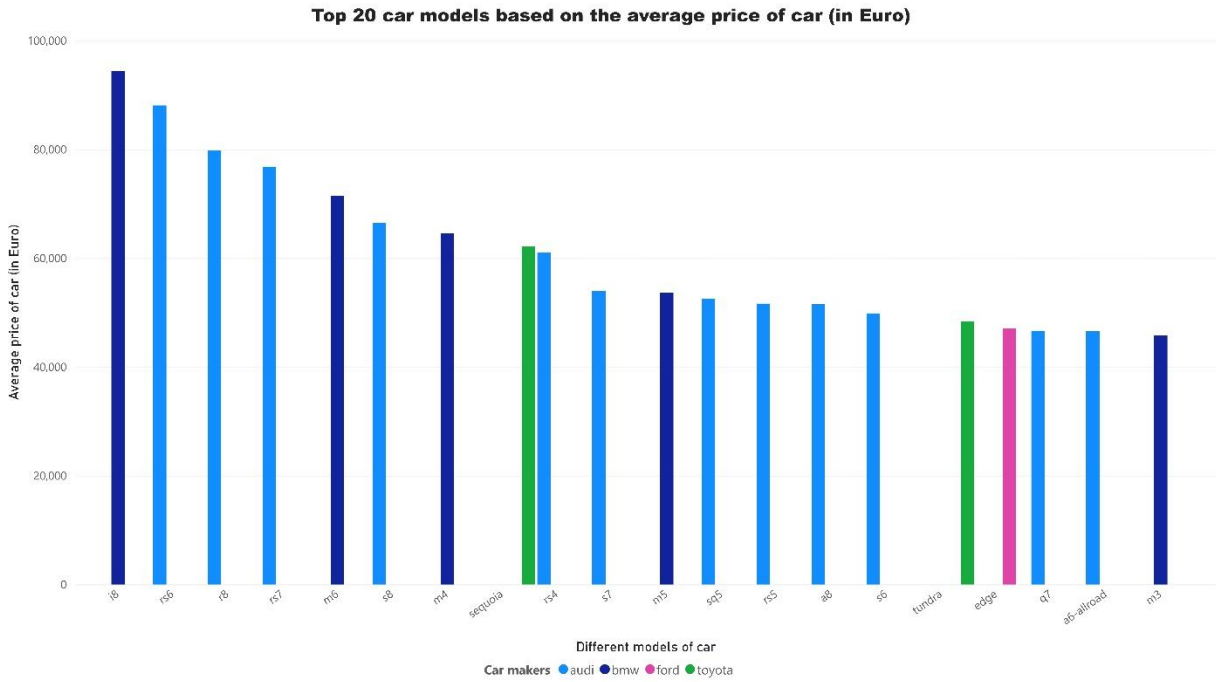


Figure 4: Top 20 car models based on price

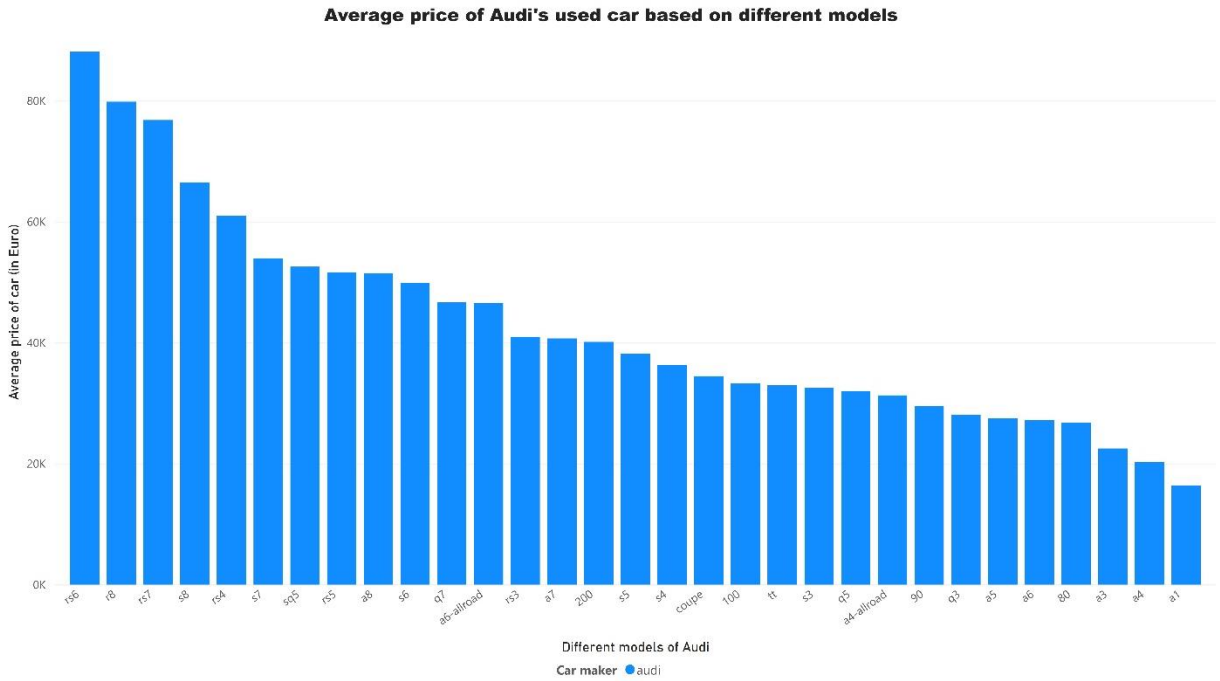


Figure 5: Price of Audi's different models

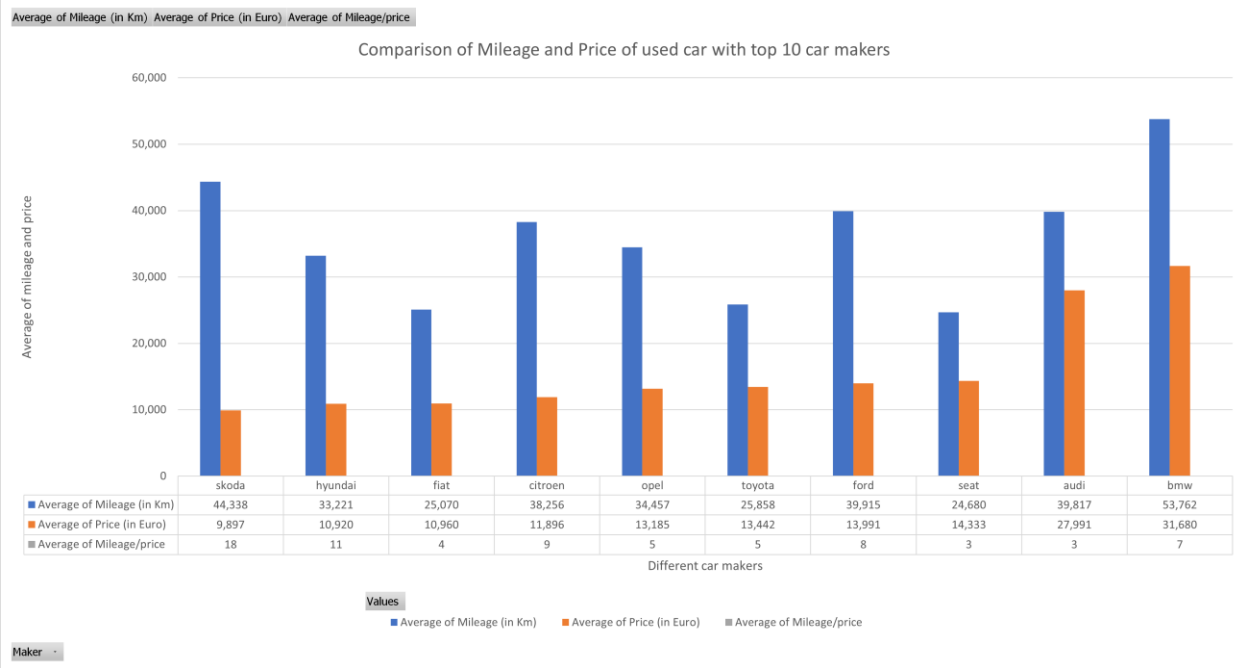


Figure 6: Comparison of mileage, price, and car maker

Figure 6 represents the relationship between mileage, price, and car maker. In general, as the mileage of cars increases, the price of cars decreases. Also, the price variation will depend on which car manufacturer is selected, such as Skoda is the least expensive car and provides the highest mileage/price ratio of 18 or 18 km/ 1€. On the other hand, car makers such as Seat and Audi has the lowest mileage/price ratio of 3 or 3 km/ 1€. So, if the ratio is low then the price will be high, even the mileage is the same in each car maker. This is one of the parameters to calculate the actual cost of the car, other parameters include the service/maintenance log, accident record, rust proofing, colour paint, the interior of the car, etc. This analysis helps us to understand how the mileage affects the price of a car based on car makers and its drastic price change.

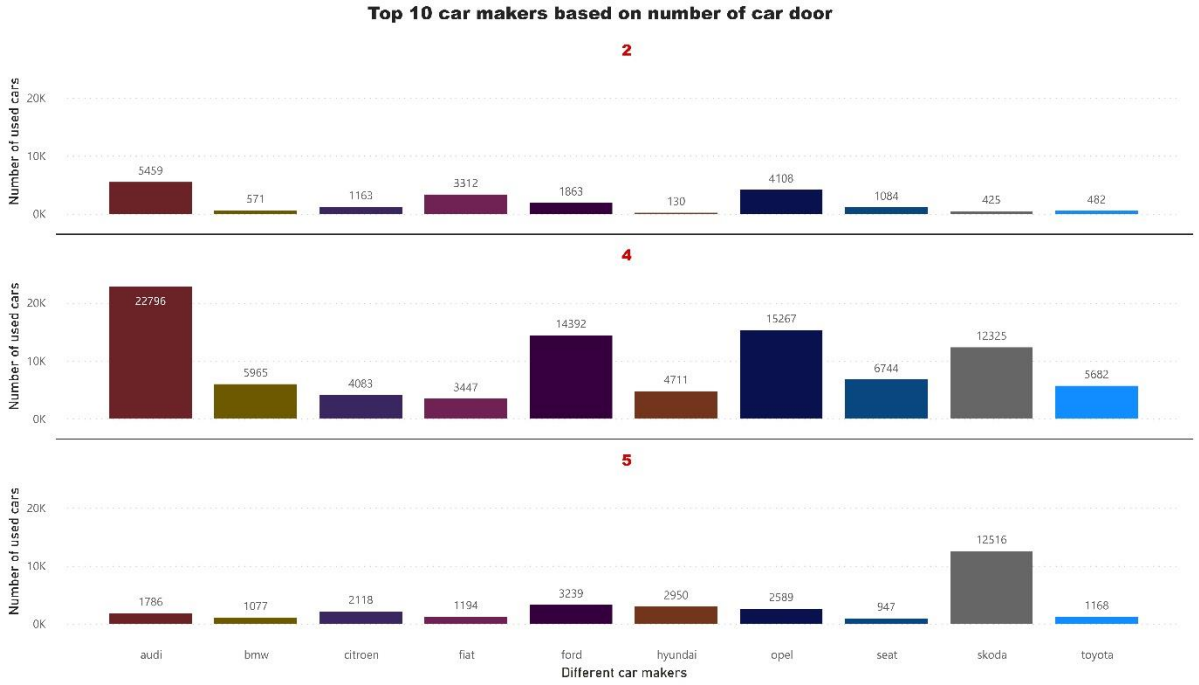


Figure 7: Number of car doors based on model

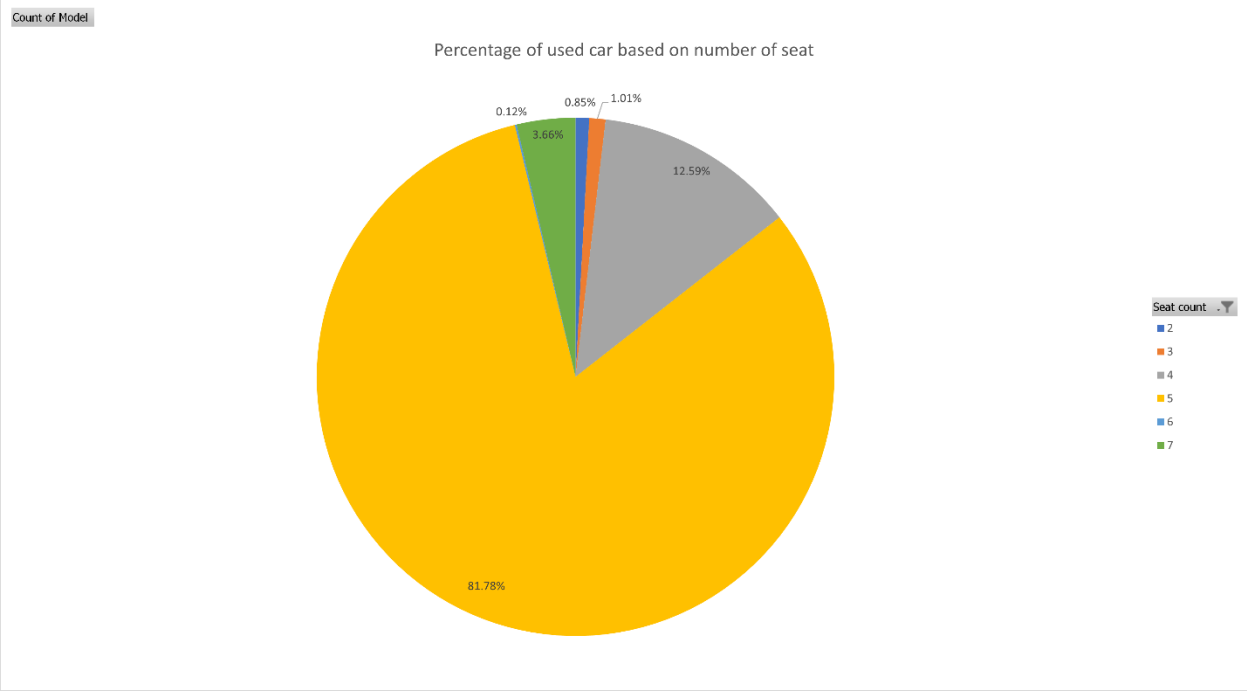


Figure 8: Number of seats in a car

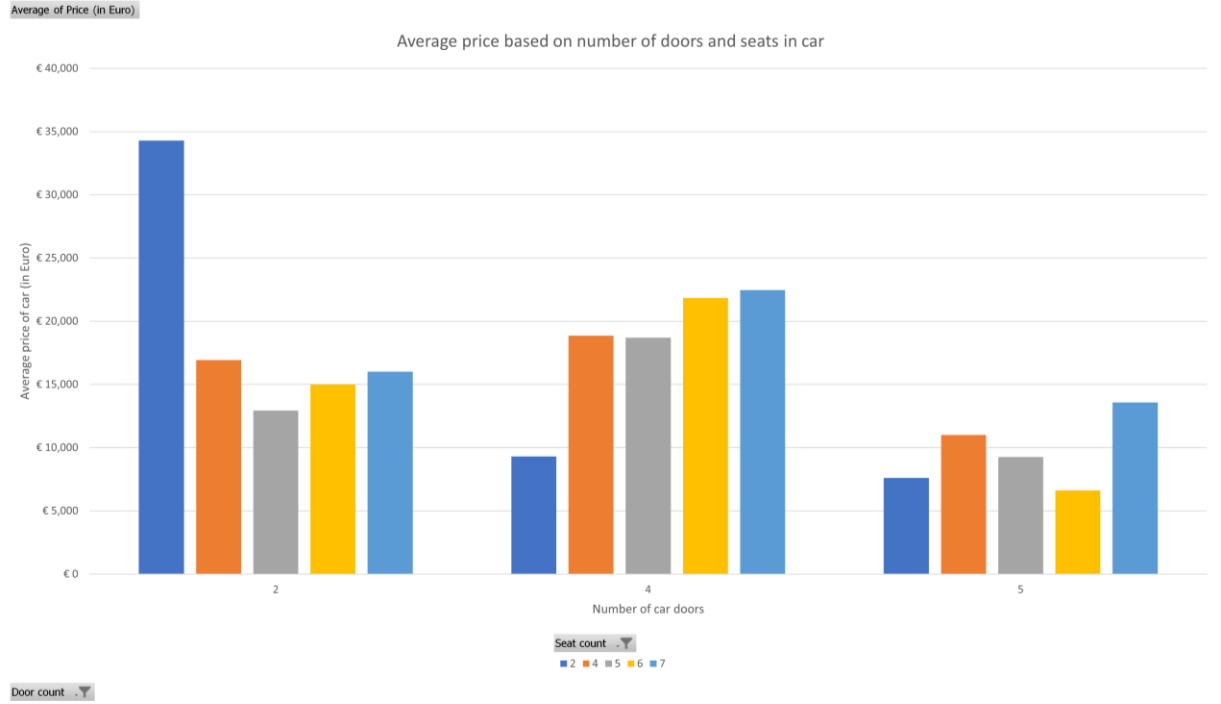


Figure 9: Comparison of average price based on number of car doors and seats

From Figures 7 and 8, the average seating capacity in a car should be 5 and the doors should be 4, which makes an ideal choice for each car maker. Figure 9 represents the comparison of the average price of a car based on the number of car doors and seats since most fast cars have 2 seats and 2 doors that lead to taking the average price of the car to maximum, whereas 4 doors and 4 seats or above have a similar price range of cars. This analysis helps to focus our business model around 2 or 4 door cars with 4 or above seats for maximizing the profit.

From Figures 10 and 11, it can be derived that many cars in the Czech Republic and Germany are manual cars instead of automatic. Also, only 2 major car makers Audi and BMW produce the majority of cars with automatic transmission, whereas others tend to produce manual cars. The reason behind manual vs automatic is the cost (price) as for all car makers, cars with automatic transmission is expensive than the ones with manual transmission.

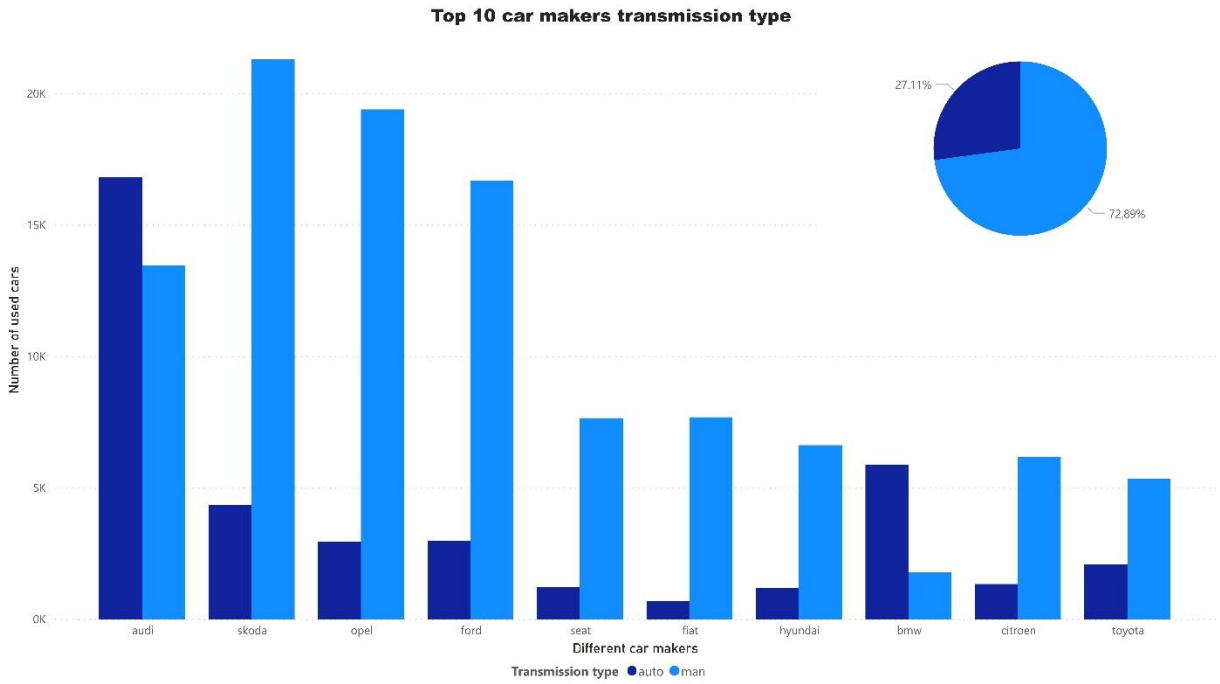


Figure 10: Top 10 car makers transmission type

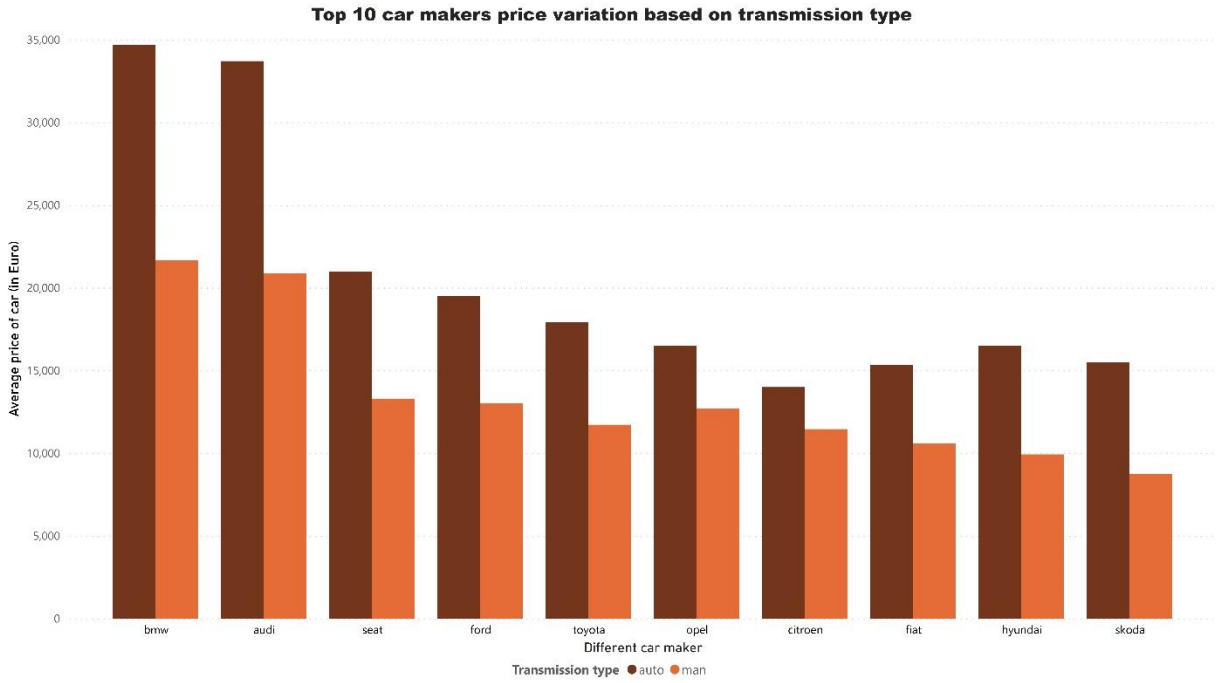


Figure 11: Price of a car based on the transmission type

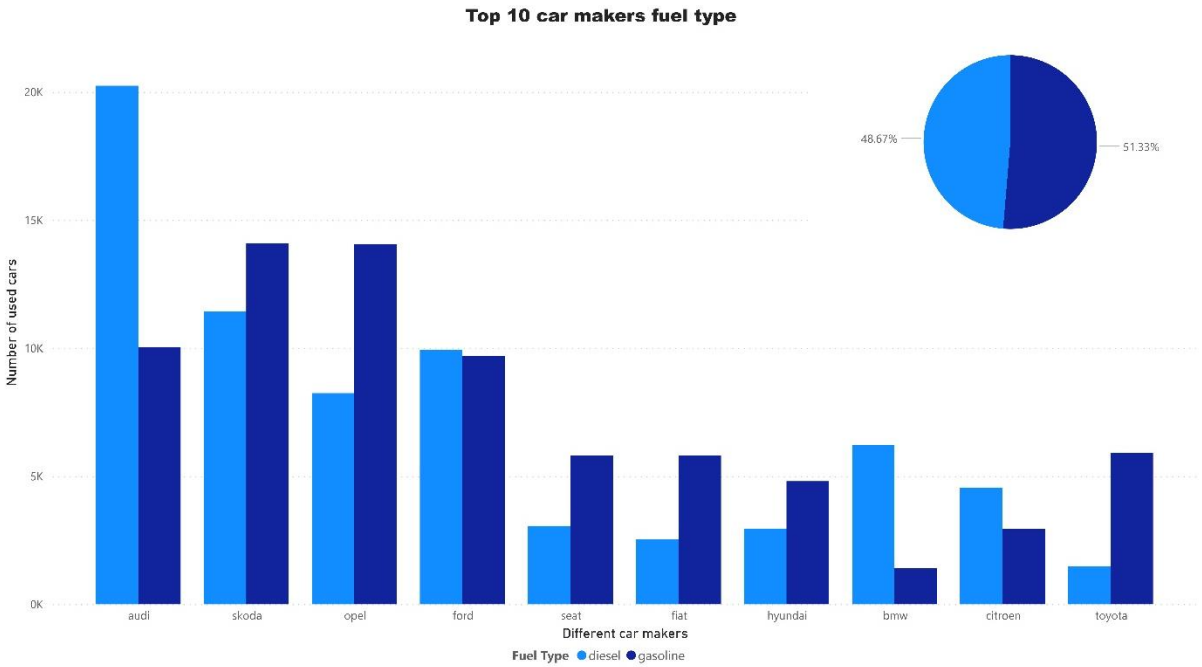


Figure 12: Top 10 car makers fuel type

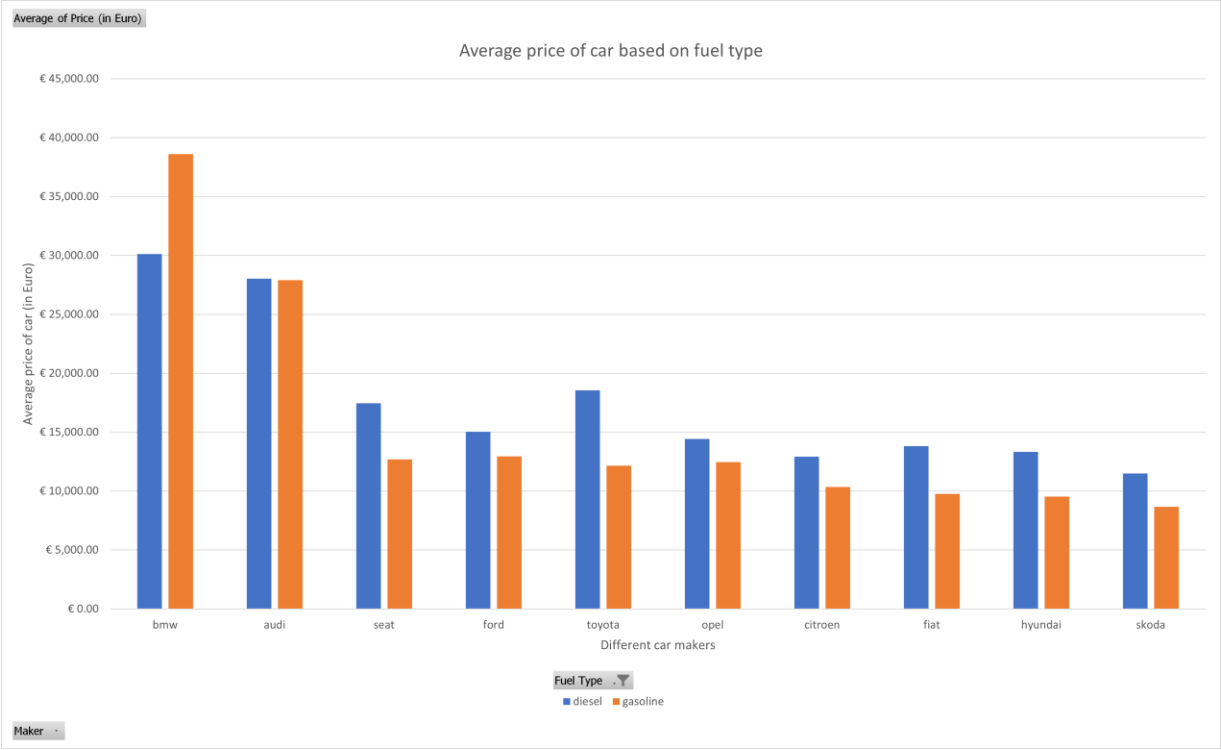


Figure 13: Price of a car based on fuel type

From figures 12 and 13, it can be stated that cars in the Czech Republic and Germany are both gasoline and diesel in approximate equal ratio. It is interesting to note that Audi, BMW, Citroen car manufactures have most of their cars running on diesel as compared to gasoline, whereas other car makers manufacture has more cars running on gasoline. Also, the price of a diesel car is more than the gasoline one except for BMW manufacturer, the reason being the cheaper price of diesel as compared to gasoline (autotraveler, n.d.).

Therefore, from Figures 10 – 13, it can be analyzed that the top 5 average prices of the cars based on car makers are BMW, Audi, Seat, Ford, and Toyota irrespective of transmission or fuel type.

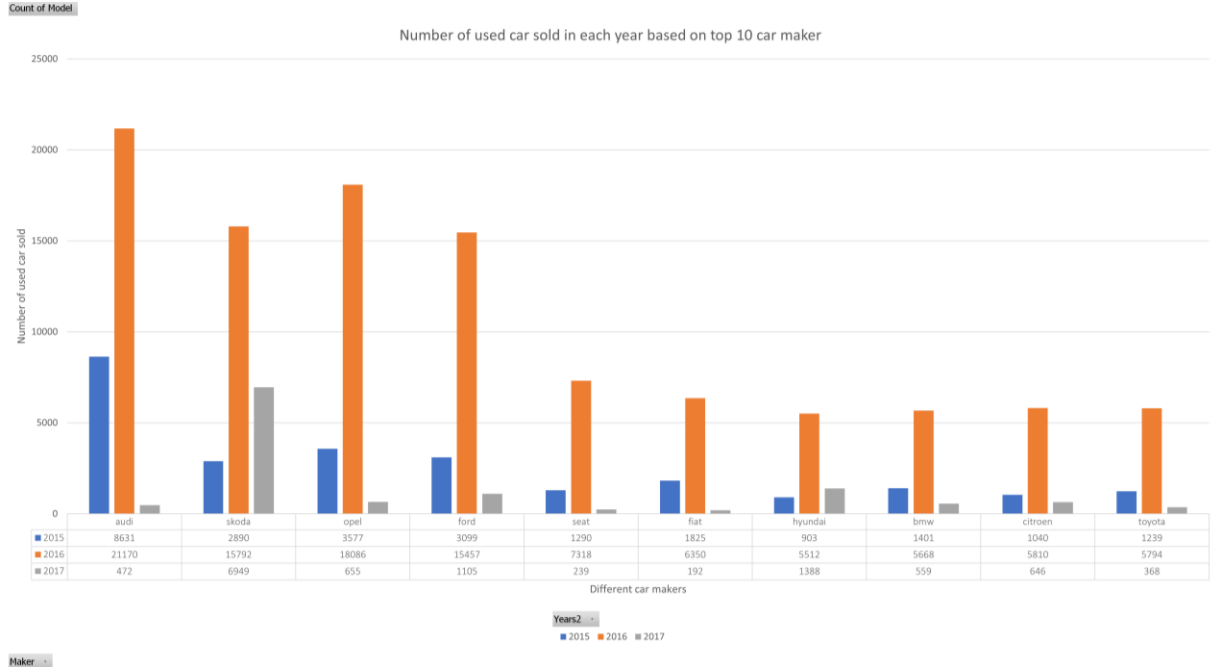


Figure 14: Sales figure of top 10 car makers

Figure 14 represents the sales of used cars in 2015, 2016 and 2017. Since the data is collected mostly in 2016, that's the reason to have a maximum number of sales. This analysis clears that the first 5 car makers have the highest number of sales. Therefore, the recommended 5 vehicle manufacturers that should be considered to make the business successful are Audi, Skoda, Seat, Ford, and Opel which will be sold easily with variations of models, price, transmission, and fuel type.

CONCLUSION

In general, any business plan has risks whether it's business model will be successful or not. During the analysis, its clear that if the desired selection of used car maker, model, manufacturing year is selected then the business shall not fail. All the parameters have been discussed in previous chapters and appropriate graphs has been generated such that visualization should be giving us better understanding. Most important part of data cleaning has been done using hive such that the decisions can be data-driven for the investment in used car dealership business.

It can be concluded that, if the investment firm is investing in this business plan, then it can generate a huge profit in a very short time. Therefore, to make it successful the car manufacturers should be selected are Audi, Skoda, Ford, Seat, BMW, and Opel which are manufactured between 2011 – 2015 with car door as 2 or 4 and number of seats above 4 with any transmission type or fuel type.

BIBLIOGRAPHY

autotraveler. (n.d.). *Trend in gasoline prices in Germany. Statistics for the last few years. Gasoline prices in Germany.* Retrieved 11 12, 2021, from autotraveler.ru: https://autotraveler.ru/en/germany/trend-price-fuel-germany.html

DriveNation. (2019, 05 03). Retrieved from drivenation.ca/blog: https://www.drivenation.ca/blog/depreciation-vehicle-much-money-will-lose-buying-new-car/

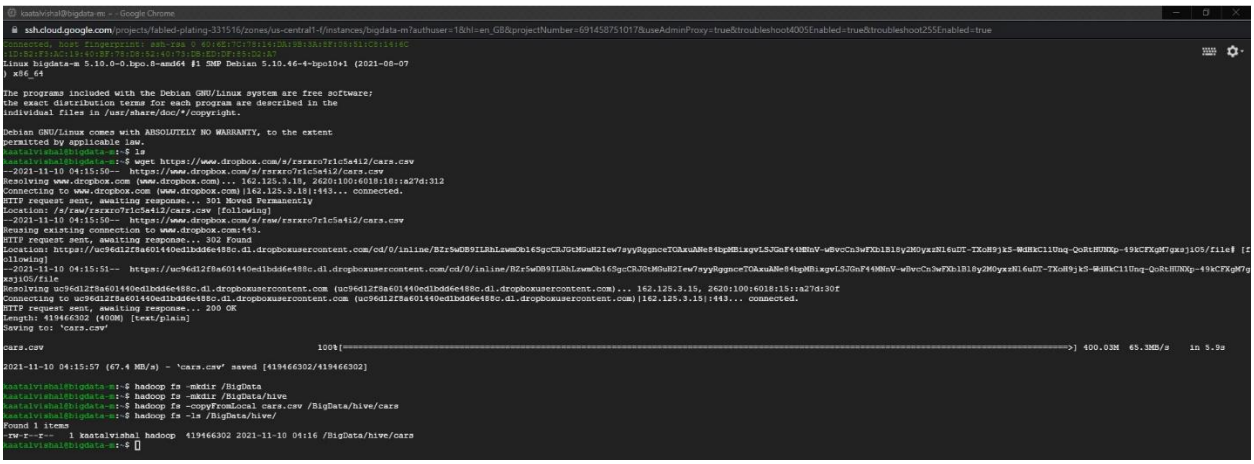
Zoricak, M. (2017, 03 16). *Classified Ads for Cars*. Retrieved from Kaggle: https://www.kaggle.com/mirosval/personal-cars-classifieds

APPENDIX

////Loading of Data in GCP///

```
wget https://www.dropbox.com/s/rsrxro7r1c5a4i2/cars.csv
```

////Creating hive Directory inside BigData folder and move the cars.csv to hive////



////Start hive and create a database & table of cars////

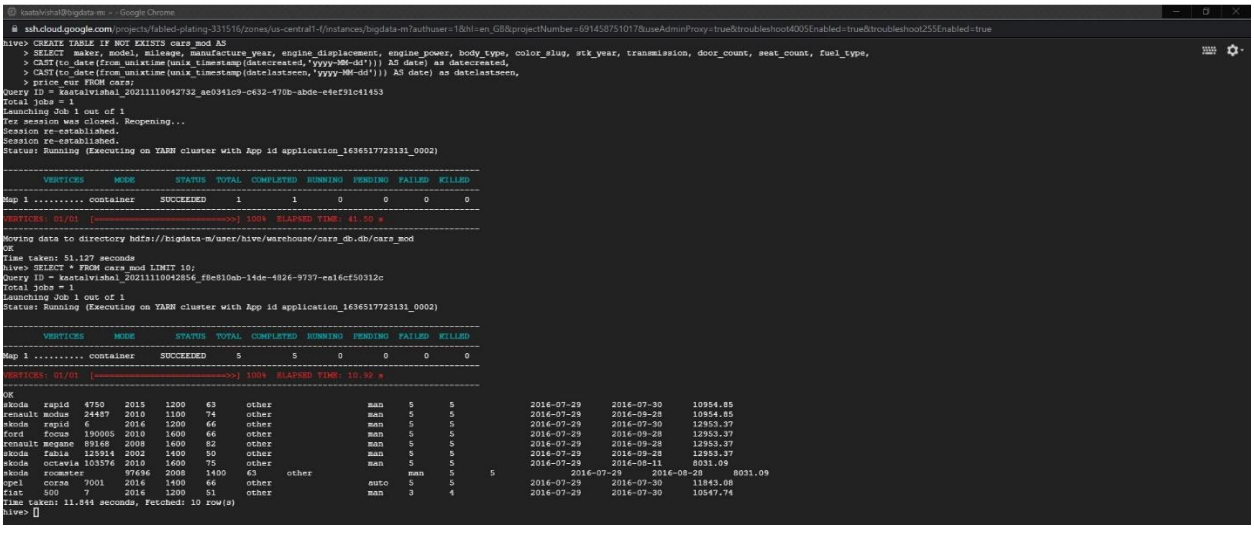
```
hive> CREATE DATABASE cars_db;

hive> USE cars_db;

hive> CREATE EXTERNAL TABLE IF NOT EXISTS cars (
maker STRING,
model STRING,
mileage INT,
manufacture_year INT,
engine_displacement INT,
engine_power INT,
body_type STRING,
color_slug STRING,
stk_year STRING,
transmission STRING,
door_count INT,
seat_count INT,
fuel_type STRING,
datecreated STRING,
datelastseen STRING,
price_eur FLOAT)
ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
LOCATION '/BigData/hive/'
TBLPROPERTIES("skip.header.line.count"="1");
```

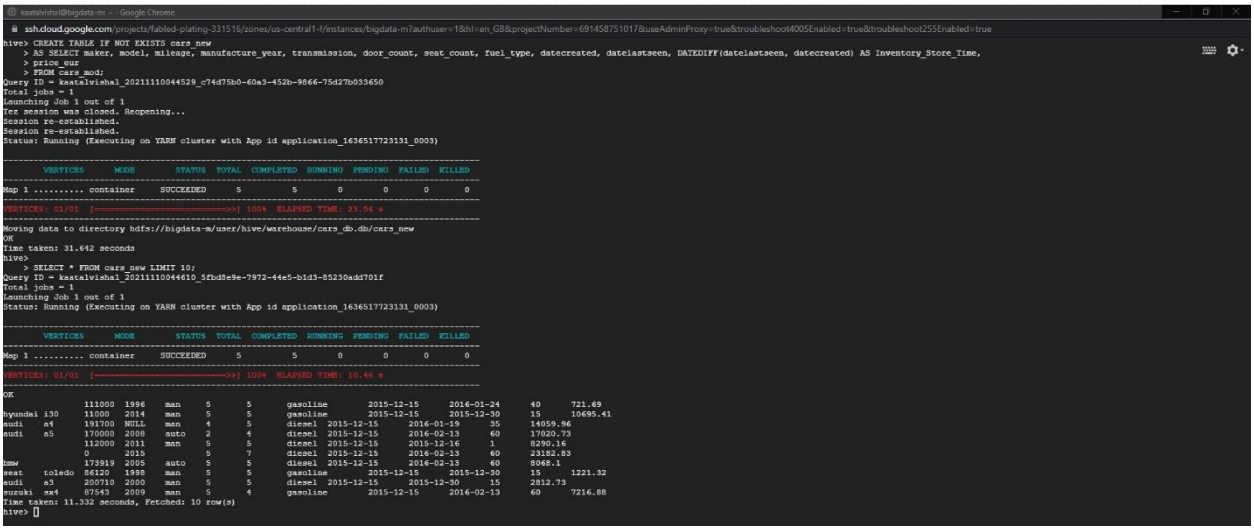
////Converting STRING Date column to actual date values////

```
hive> CREATE TABLE IF NOT EXISTS cars_mod AS
SELECT  maker, model, mileage, manufacture_year, engine_displacement,
engine_power, body_type, color_slug, stk_year, transmission, door_count,
seat_count, fuel_type,
CAST(to_date(from_unixtime(unix_timestamp(datecreated,'yyyy-MM-dd')))
AS date) as datecreated,
CAST(to_date(from_unixtime(unix_timestamp(datelastseen,'yyyy-MM-dd')))
AS date) as datelastseen,
price_eur FROM cars;
```



////Creation of new table based on required columns for the analysis using the last table////

```
hive> CREATE TABLE IF NOT EXISTS cars_new
AS SELECT maker, model, mileage, manufacture_year, transmission,
door_count, seat_count, fuel_type, datecreated, datelastseen,
DATEDIFF(datelastseen, datecreated) AS Inventory_Store_Time,
price_eur
FROM cars_mod;
```



////Cleaning of the dataset which includes blank space, null values, outside scope values such mileage 1 billion, price of the car as 2706 billion where each change has been stored as separate table////

STEP 1: Sorting the dataset for manufacturing year 2011 – 2015 and collecting only manual or automatic transmission type

```
hive> CREATE TABLE IF NOT EXISTS cars_clean1
AS SELECT maker, model, mileage, manufacture_year, transmission,
door_count, seat_count, fuel_type, price_eur, datecreated, datelastseen,
Inventory_Store_Time
FROM cars_new
WHERE manufacture_year > 2010 AND manufacture_year < 2016 AND
transmission in ('man','auto');
```

STEP 2: Removing the empty spaces from model and maker column

```
hive> CREATE TABLE IF NOT EXISTS cars_clean2
AS SELECT maker, model, mileage, manufacture_year, transmission,
door_count, seat_count, fuel_type, price_eur, datecreated, datelastseen,
Inventory_Store_Time
FROM cars_clean1
WHERE model != '' AND maker != '';
```

STEP 3: Selecting the fuel type as diesel or gasoline and mileage less than 150,000 kms

```
hive> CREATE TABLE IF NOT EXISTS cars_clean3
AS SELECT maker, model, mileage, manufacture_year, transmission,
door_count, seat_count, fuel_type, price_eur, datecreated, datelastseen,
Inventory_Store_Time
FROM cars_clean2
WHERE mileage < 150000 AND fuel_type in ('diesel','gasoline');
```

STEP 4: Selection of the range of Price of car between €100 – €100,000

```
hive> CREATE TABLE IF NOT EXISTS cars_clean4
AS SELECT maker, model, mileage, manufacture_year, transmission,
door_count, seat_count, fuel_type, price_eur, datecreated, datelastseen,
Inventory_Store_Time
FROM cars_clean3
WHERE price_eur > 100 AND price_eur < 100000;
```

STEP 5: Selecting the range of number of car seats between 1 - 8

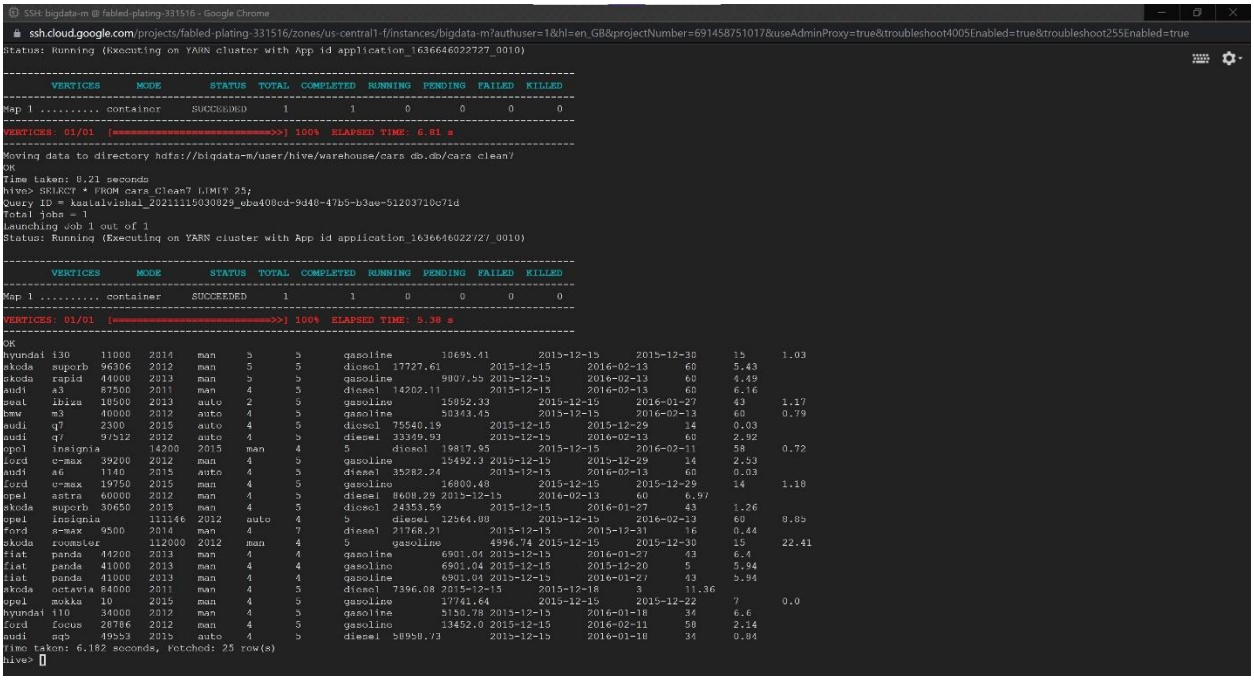
```
hive> CREATE TABLE IF NOT EXISTS cars_clean5
AS SELECT maker, model, mileage, manufacture_year, transmission,
door_count, seat_count, fuel_type, price_eur, datecreated, datelastseen,
Inventory_Store_Time
FROM cars_clean4
WHERE seat_count > 1 AND seat_count < 8;
```

STEP 6: Selecting the range of number of car doors between 1 - 6

```
hive> CREATE TABLE IF NOT EXISTS cars_clean6
AS SELECT maker, model, mileage, manufacture_year, transmission,
door_count, seat_count, fuel_type, price_eur, datecreated, datelastseen,
Inventory_Store_Time
FROM cars_clean5
WHERE door_count > 1 AND door_count < 6;
```

STEP 7: Selection of top 10 car manufacturers

```
hive> CREATE TABLE IF NOT EXISTS cars_clean7
AS SELECT maker, model, mileage, manufacture_year, transmission,
door_count, seat_count, fuel_type, price_eur, datecreated, datelastseen,
Inventory_Store_Time, ROUND(CAST(mileage/price_eur AS FLOAT),2) AS MPP
FROM cars_clean6
WHERE maker in
('audi','skoda','opel','ford','seat','fiat','hyundai','bmw','citroen',
'toyota');
```



////To download the file from GCP, copy the file to local and download it////

```
hive> INSERT OVERWRITE LOCAL DIRECTORY '/home/kaatalvishal'
ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
SELECT * FROM cars_clean7;
```

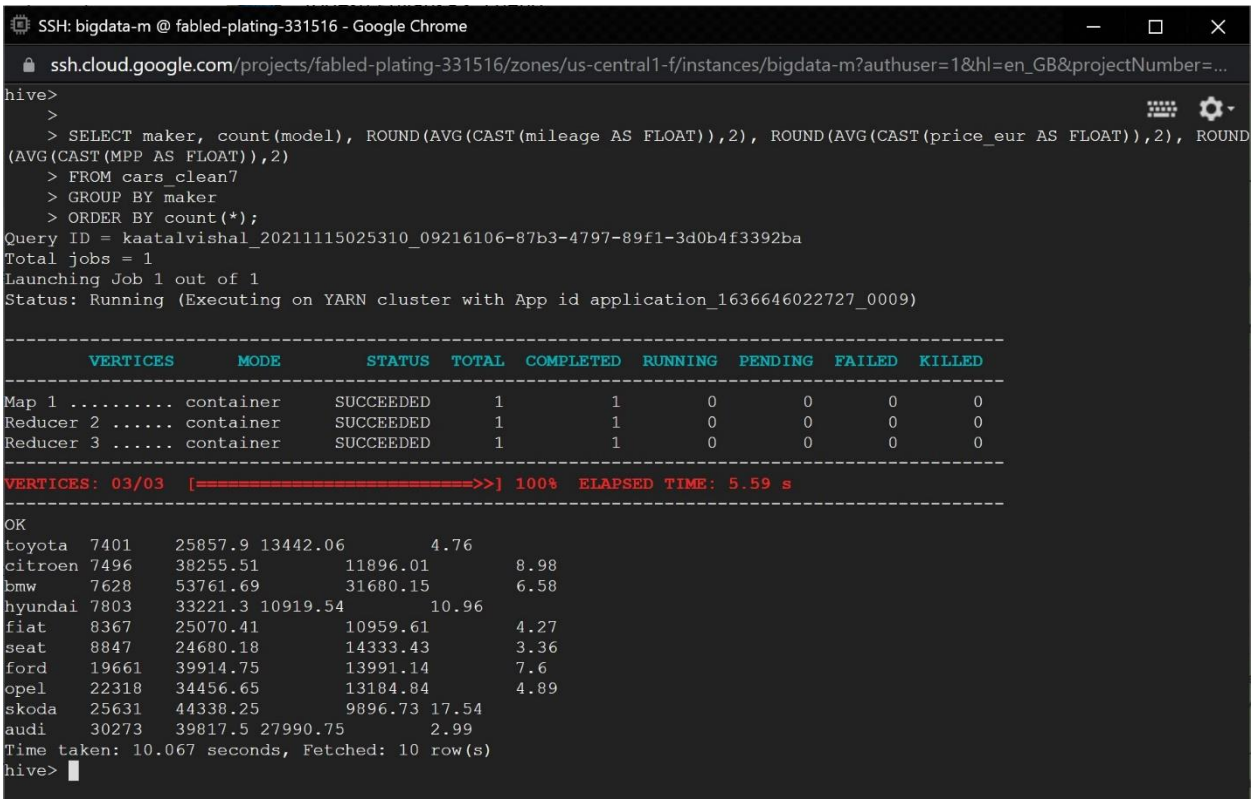
////Analysis Graphs////

```
hive> SELECT maker, count(maker)
FROM cars_clean7
GROUP BY maker
ORDER BY count(*) DESC;
```

```
hive> SELECT manufacture_year, count(maker)
FROM cars_clean7
GROUP BY manufacture_year
ORDER BY count(*) DESC;
```

```
hive> SELECT maker, ROUND(AVG(CAST(price_eur AS FLOAT)),2)
FROM cars_clean7
GROUP BY maker
ORDER BY count(*) DESC;
```

```
hive> SELECT maker, ROUND(AVG(CAST(mileage AS FLOAT)),2),
ROUND(AVG(CAST(price_eur AS FLOAT)),2) , ROUND(AVG(CAST(MPP AS
FLOAT)),2)
FROM cars_clean7
GROUP BY maker
ORDER BY count(*) DESC;
```




```
hive> SELECT seat_count, count(seat_count)
FROM cars_clean7
GROUP BY seat_count;

hive> SELECT transmission, count(transmission)
FROM cars_clean7
GROUP BY transmission;

hive> SELECT maker, count(transmission)
FROM cars_clean7
GROUP BY maker;

hive> SELECT door_count, count(door_count)
FROM cars_clean7
GROUP BY door_count;

hive> SELECT fuel_type, count(fuel_type)
FROM cars_clean7
GROUP BY fuel_type;

hive> SELECT maker, count(fuel_type)
FROM cars_clean7
GROUP BY maker;
```