# SMDM\_final assessment report

Table of Content

Problem 1: Austo Motor.........................................................................................................................2

1. What is the important technical information about the dataset that a database administrator would be interested in?.............................................................................................................2
2. Take a critical look at the data and do a preliminary analysis of the variables. Do a quality check of the data so that the variables are consistent? Are there any discrepancies present in the data?....................................................................................................................................3
3. Explore all the features of the data separately by using appropriate visualizations and draw insights that can be utilized by the business.............................................................................4
4. Understanding the relationships among the variables in the dataset is crucial for every analytical project. Perform analysis on the data fields to gain deeper insights. Comment on your understanding of the data.................................................................................................7
5. Employees working on the existing marketing campaign have made the following remarks. Based on the data and your analysis state whether you agree or disagree with their observations. Justify your answer Based on the data available................................................8

E1) Steve Roger says "Men prefer SUV by a large margin, compared to the women"..............8

E2) Ned Stark believes that a salaried person is more likely to buy a Sedan.............................8

E3) Sheldon Cooper does not believe any of them; he claims that a salaried male is an easier target for a SUV sale over a Sedan Sale.....................................................................................8

1. From the given data, comment on the amount spent on purchasing automobiles across the following categories. Comment on how a Business can utilize the results from this exercise. Give justification along with presenting metrics/charts used for arriving at the conclusions.

6.1) Gender................................................................................................................................9

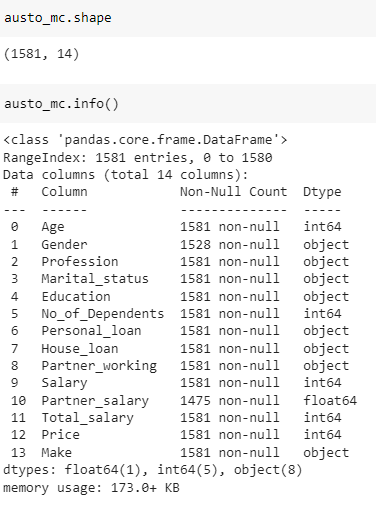
6.2) Personal\_loan.....................................................................................................................9

1. From the current data set comment if having a working partner leads to purchase of a higher priced car.................................................................................................................................10
2. The main objective of this analysis is to devise an improved marketing strategy to send targeted information to different groups of potential buyers present in the data. For the current analysis use the Gender and Marital\_status - fields to arrive at groups with similar purchase history......................................................................................................................10

Problem2: Go Digit Credit Card............................................................................................................11

# Problem 1: Austo Motor Company

## 1 - What is the important technical information about the dataset that a database administrator would be interested in?



The Dataset has a total of 1581 rows and 14 columns

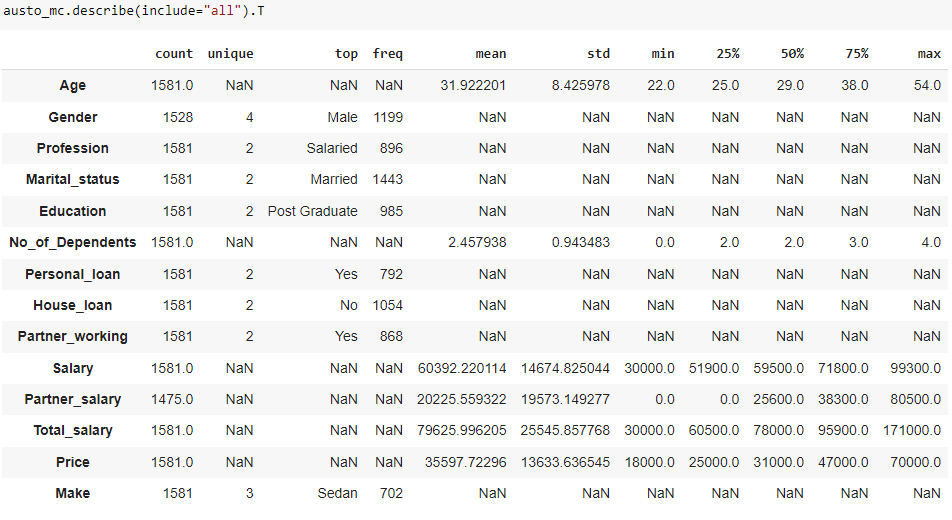
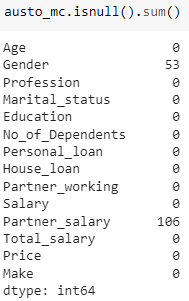
Out of the 14 variables there are 6 numerical and 8 categorical variables

There are no duplicate values in the dataset

There are a 53 null values in the Gender and 106 null values in Partner\_salary variables

Gender also shows 4 unique value, it needs to be processed





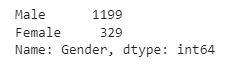
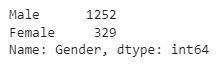
## 2 - Take a critical look at the data and do a preliminary analysis of the variables. Do a quality check of the data so that the variables are consistent? Are there any discrepancies present in the data?

There are misspelled values for the value 'Female' as 'Femal' and 'Femle' in Gender variable.

Gender has 53 missing values, dropping them might impact the analysis, so we imputed the missing values with Mode (highest frequency)

After applying Mode

Before applying Mode

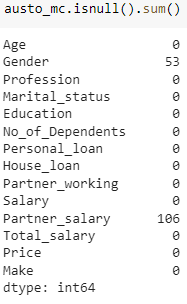
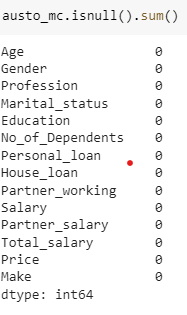


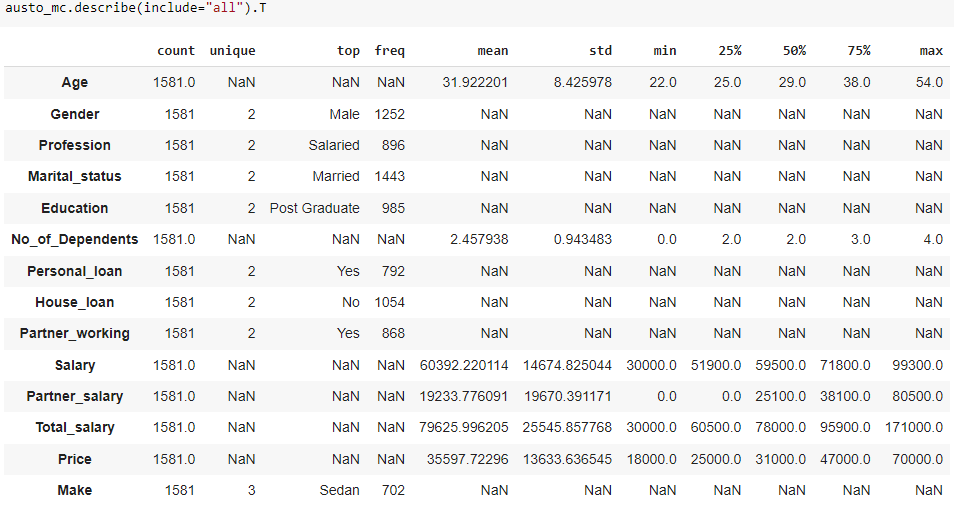
Partner\_Salary has 106 rows with missing values.

We treated this missing values considering ‘Total\_salary - Salary, where the Partner Working option is mentioned Yes

After imputing missing values Mode

Before imputing missing values Mode



3 - Explore all the features of the data separately by using appropriate visualizations and draw insights that can be utilized by the business

Minimum age of a person is 22 and maximum is 54 but the 75%ile is 38 years

There are maximum no. of 4 dependents

Salary ranges from 30000 to 99300

Range of price of cars varies from 18000 to 70000

Minimum age of a person is 22 and maximum is 54 but the 75%ile is 38 years

There are maximum no. of 4 dependents

Salary ranges from 30000 to 99300

Range of price of cars varies from 18000 to 70000

Numerical Analysis – univariate

**Age** is skewed towards right and median is around 29 showing more people below the age of 30 are purchasing the vehicle. 75% of the buyers are below 38 years of age

**Partner salary** is heavily skewed towards the right which indicating around 75% of the partner's salary is below 38000 while highest going more than 80000

**Price** is also skewed towards right and 75% of the people are buying vehicles below 47000 where maximum price goes up to 70000

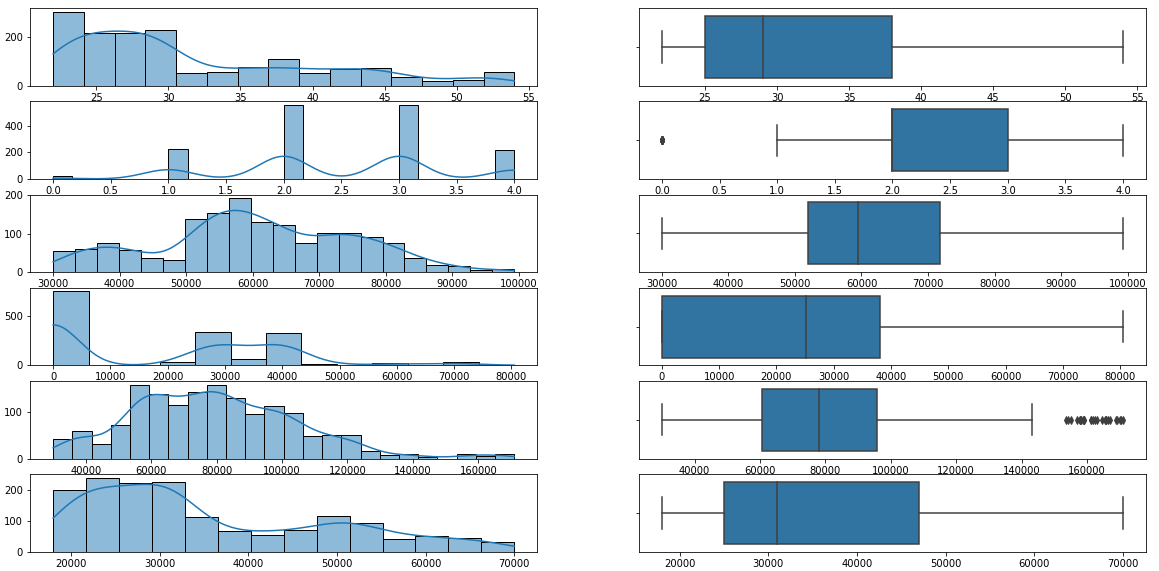
Total salary

Price

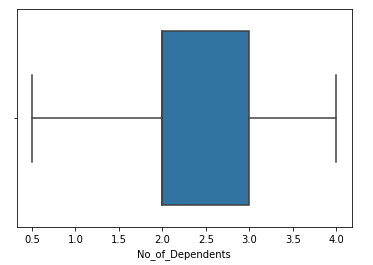
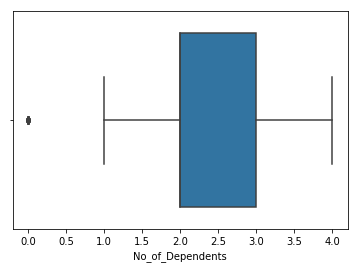
Partner salary

No of Dependents

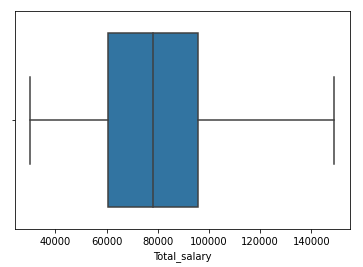
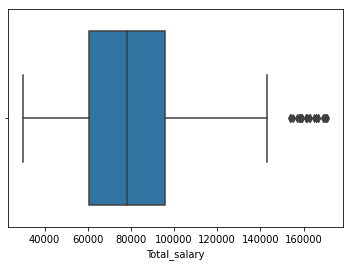
Age



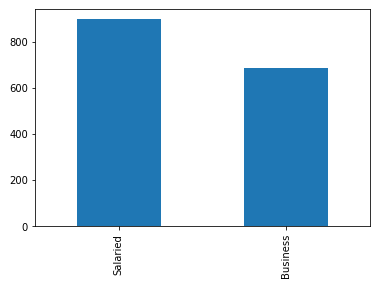
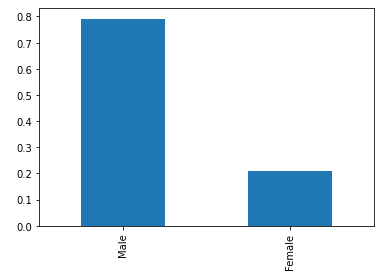
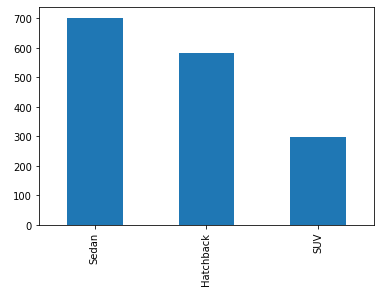
Salary



**No\_of\_Dependents** has outliers showing 0. Replaced outlier for lower range.



**Total\_salary** has outliers showing on the higher range. Replaced outlier using IQR method value for upper range.

Categorical variables analysis – univariate

**Make**

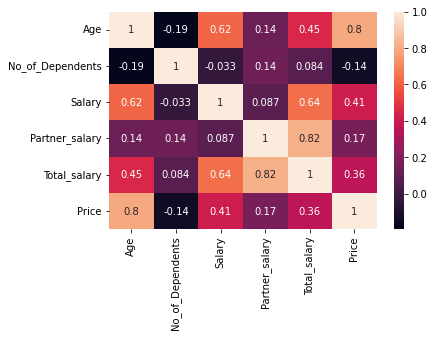
**Profession**

**Gender**

**Gender** has around 80% of Male category and 20% of Female category

There are around 56% of people who are salaried vs 43% who are doing business in **Profession** category

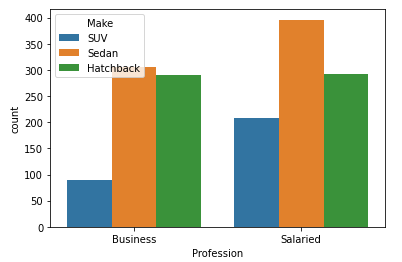
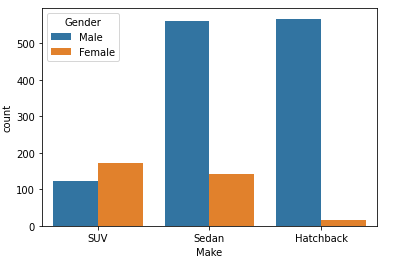
SUV’s are less preferred than hatchback or sedans in **Make** category

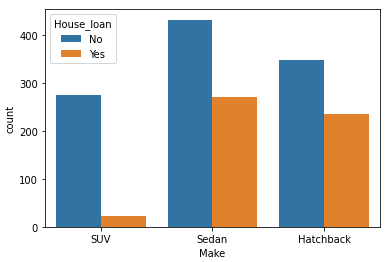
4 - Understanding the relationships among the variables in the dataset is crucial for every analytical project. Perform analysis on the data fields to gain deeper insights. Comment on your understanding of the data

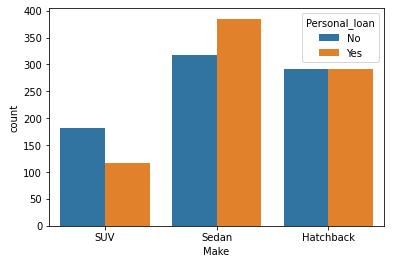
Correlation between **Age** and **Price** is the highest 0.8

Correlation between **Partner\_salary** and **Total\_salary** shows 0.82

The other variables have limited correlation among them







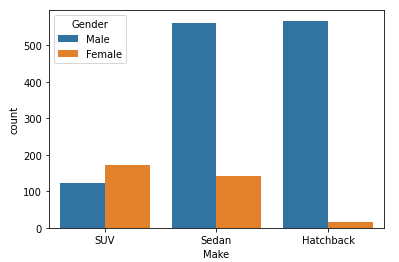
Males prefer sedan and hatchback more whereas Female prefer SUVs in **Make** category

People with House loan prefer less of SUV in **Make** category

People who have Personal loan prefer more of sedan in **Make** category

5 - Employees working on the existing marketing campaign have made the following remarks. Based on the data and your analysis state whether you agree or disagree with their observations. Justify your answer Based on the data available

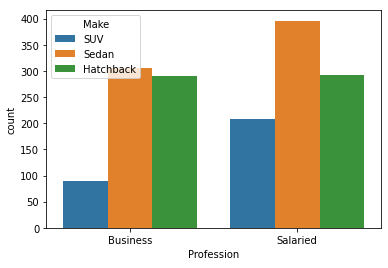
5.1 - Steve Roger says "Men prefer SUV by a large margin, compared to the women"



Statement is **False**

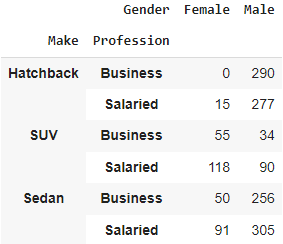
Female prefer more of SUV compared to Men who prefer more of sedan and Hatchback

5.2 - Ned Stark believes that a salaried person is more likely to buy a Sedan



Statement is **True**

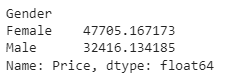
Salaried people are more likely to buy Sedan compared to business profession

5.3 - Sheldon Cooper does not believe any of them; he claims that a salaried male is an easier target for a SUV sale over a Sedan Sale

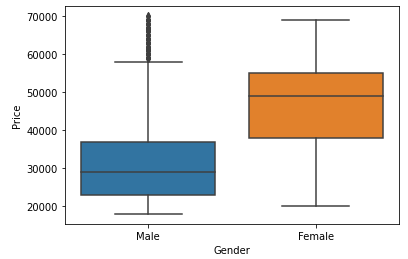
Statement is **False**

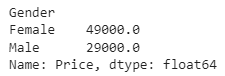
Salaried Male people are prefer to buy Sedan over SUV compared to business profession

6 - From the given data, comment on the amount spent on purchasing automobiles across the following categories. Comment on how a Business can utilize the results from this exercise. Give justification along with presenting metrics/charts used for arriving at the conclusions

6.1 – Gender

Mean



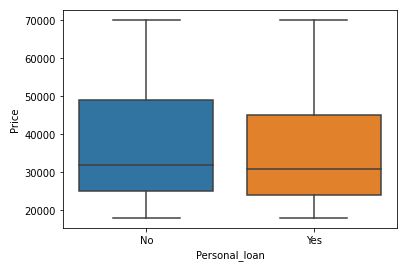


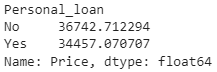
Median

Mean and Median Price for Female customers is higher than Male customers

Business should focus on Female as their spending are higher than Men

6.2 – Personal Loan



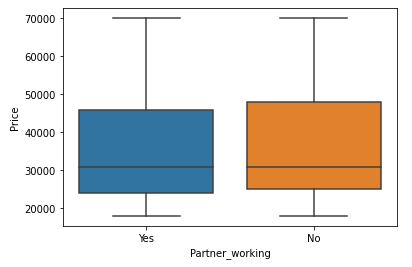


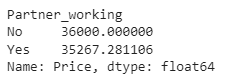
Mean

People who do not have Personal Loan spend more for the vehicles

Business should focus on people with no personal loan

## 7. From the current data set comment if having a working partner leads to purchase of a higher priced car

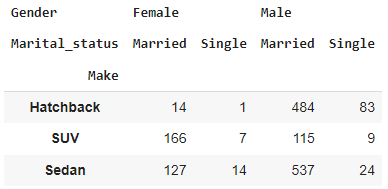


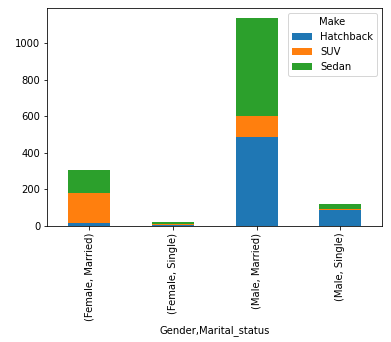


Mean

There is only slight difference in Average Price for both people who have working partner and people who do not have a working partner, so vehicles purchased by people does not depend on Partner working category

## 8. The main objective of this analysis is to devise an improved marketing strategy to send targeted information to different groups of potential buyers present in the data. For the current analysis use the Gender and Marital\_status - fields to arrive at groups with similar purchase history





Married male and female are more into buying cars than singles. Business can approach married people as target audience

SUV are majorly bought by women compared to Men. This could be marketing strategy to brand among more women

Male prefer sedan over other cars and can be branded to them

# Problem 2: Go Digit Credit Card

The Top Variables for the analysis of this dataset will be:

* Annual\_income\_at\_source – It can help to categorise customers into different income groups.
* cc\_limit – With the credit limits, business can look for more customers
* Avg\_spends\_3m
* T+1 month\_activity
* other\_bank\_cc\_holding