**AUSTO MOTOR COMPANY**

**& GODIGIT BANK**

a

A picture containing text, clipart, vector graphics

Description automatically generatedSM

**DSBA**

**SMDM BUSINESS REPORT**

Presented by -AMAYIKA

**Email:** Kashyap.amayika009@gmail.com

**4th Sept, 2022**

SMDM Module

# Abstract

**Problem -1**

**Austo Motor Company is a leading car manufacturer specializing in SUV, Sedan, and Hatchback models. In its recent board meeting, concerns were raised by the members on the efficiency of the marketing campaign currently being used.**

We will explore the Data set and perform the exploratory data analysis on the dataset. This analysis and the useful insights given by us will help The Company, its Board members and its employees to enhance its existing marketing strategy and increase its efficiency so that they can take The Company to new heights.

**Problem -2**

GODIGT Bank is a mid-sized private bank that deals in all kinds of banking products, such as savings accounts, current accounts, investment products, etc. among other offerings. Given the attrition in its credit card, the Bank wants to revisit its credit card policy and make sure that the card given to the customer is the right credit card. So, we will go through the dataset its variables and categories and will try to understand that how data is behaving and then we will FRAME ANALYTICS PROBLEMS based on our research so that customers that show higher intent towards a recommended credit card and help in increasing the profit margin of the company.

Table of Contents

**Problem-1**

[SECTION-A Important Technical Information 3](#_Toc68199677)

SECTION-B [Preliminary Analysis](#_Toc68199678) 5

SECTION-C Exploration of Data Set AND Insights based on the analysis8

SECTION-D [Relationship Among Variables and digging deeper Insights](#_Toc68199680) 18

SECTION-E [Analysis of Existing Marketing Campaigns 24](#_Toc68199681)

SECTION-F [Amount Spent On different Automobile categories 29](#_Toc68199683)

SECTION-G Analysis of Working Partner35

SECTION-H Gender & Marital status-based Analysis…………………………………………………………………………………………………………………………………………………38

**Problem -2**

Framing the Analytical questions and their Justifications……………………………………………………………………41

# SECTION A

# Technical And Preliminary Analysis

## Basic Data Exploration

## Important technical Information about the dataset that will help a database administrator in analyzing the business problem and its solutions

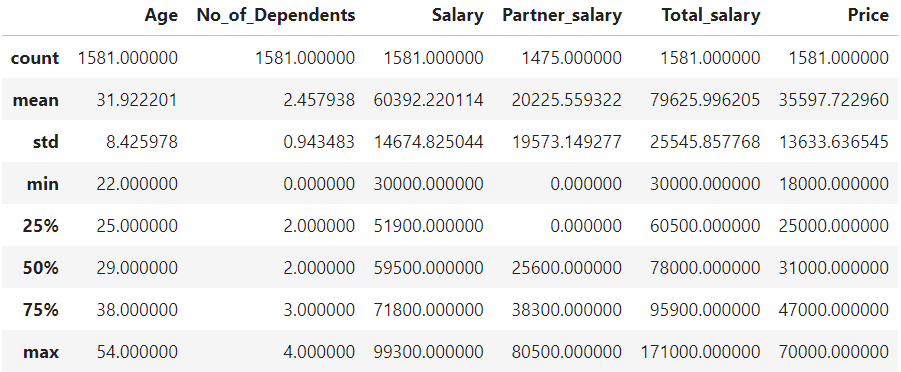
**In this step, we will perform the below operations to check what the data set comprises of. We will check the below things:**

* **head of the dataset**
* **shape of the dataset**
* **info of the dataset**
* **summary of the dataset**
* **Head of the dataset - head () function will tell us the top records in the data set. By default, it’s the top 5 rows) This will help us to know the type of data that we have and their characteristics**
* **shape of the dataset - Shape attribute tells us number of observations and variables we have in the data set. This function "shape" returns the shape of an array. It is used to check the dimension of data. The Austo Motor Company data set has 1581 observations and 14 variables in the data set.**
* **Info of the dataset** - **info () is used to check the Information about the data and the datatypes of each respective attributes. Here it shows that we have 1 float datatype, 5 int datatypes and 8 object datatypes. we can further divide the variables based on our Inference that we have. Our variables divided in following categories**-

|  |  |  |
| --- | --- | --- |
| Categorical Variables | | Continuous Variables |
| Binary | Multi-Level |
| Personal loan | Make | Age |
| House loan |  | No. Of Dependents |
| Partner working |  | Salary |
| Gender |  | Partner Salary |
| Marital status |  | Total Salary |
| Education |  | Price |
| Profession |  |  |

Here we can see that basically to increase the profit of the Austo Motor Company ‘**PRICE’ (numerical)** & ‘**MAKE’ (categorical)** are the dependent variables and rest are independent variables which will help the company to enhance its profit margin.

**Summary of the dataset - The describe method will help to see how data has been spread for the numerical values. We can clearly see the minimum value, mean values, different percentile values and maximum values of all the numerical variables in the given table.**

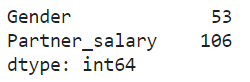
****

## SECTION -B

## Preliminary Analysis

**Preliminary Analysis of the variables and doing the quality check of the data by dealing with any discrepancies present in the data.**

**Checking null values** – **is null () function will tell us about the missing values in the data set, so that we can know if there are any discrepancies in the data set and further using the unique () function we will be able to know how many unique discrepancies are there. This will further help us to treat the wrong data and impute it with the required input. So, the data set has following null values and discrepancies.**

****

|  |  |
| --- | --- |
| Unique Variables | Counts |
| Male | 1199 |
| Female | 327 |
| Femle | 1 |
| Femal | 1 |
| Nan | 5 |

**Here we can clearly see that there are two misprinted data in the Gender column along with 53 Nan values. First, we will treat the misprinted column and then the Nan values.**

We are first replacing the misprinted words 'femle' and 'femal' as Female first using. replace () function where we have shown (in place=True) that will make the changes in the dataset as permanent values and then checking the unique values using. **unique ()** function again to know the exact Nan values now.

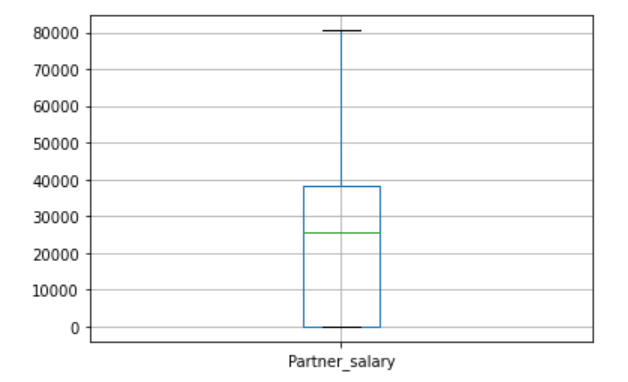
For Nan values as the **Mode** of the Gender since it’s a categorical variable so '**Mode**' will be more suitable central tendency to replace the Nan value than the 'Median’. So, first we will calculate the Mode which is a statistical tool and then replace the Nan values with it

Gender:

|  |  |
| --- | --- |
| Male | 1252 |
| Female | 329 |

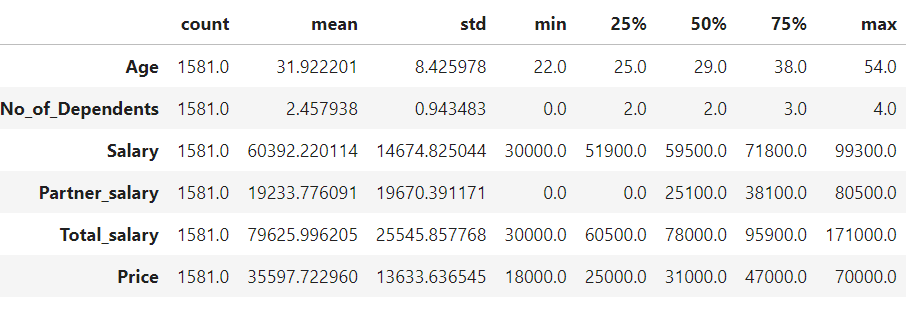
### **Similarly, we can do that in case of Partner salary but in that column two different cases needs to be taken care of and hence two conditional statements are required.**

**So here as we can observe there are 106 Nan values in the column 'Partner salary' in the Dataset, in which 16 rows are such where partners are working but still the value is Nan and, in the rest, (106-16) that is 90 rows the partners are not working hence those Nan values has to be imputed with 0 we will do this in two steps taking two different cases. In First Case we will impute the values after subtracting the Salary from the Total salary column, after checking the outliers and in second case replace Nan with float value 0.0**

****

**As the boxplot shows it has no outliers, we can easily impute the values with the value which we will get when we will subtract the Salary column from Total salary column so as to know the actual value of the working partner only in the case when He or She is working (first case)**

### Now that we have computed the missing values in the data, now let’s check the summary again and then we will observe the slight changes in the values of central tendency in the column Partner salary which shows that values have been imputed properly

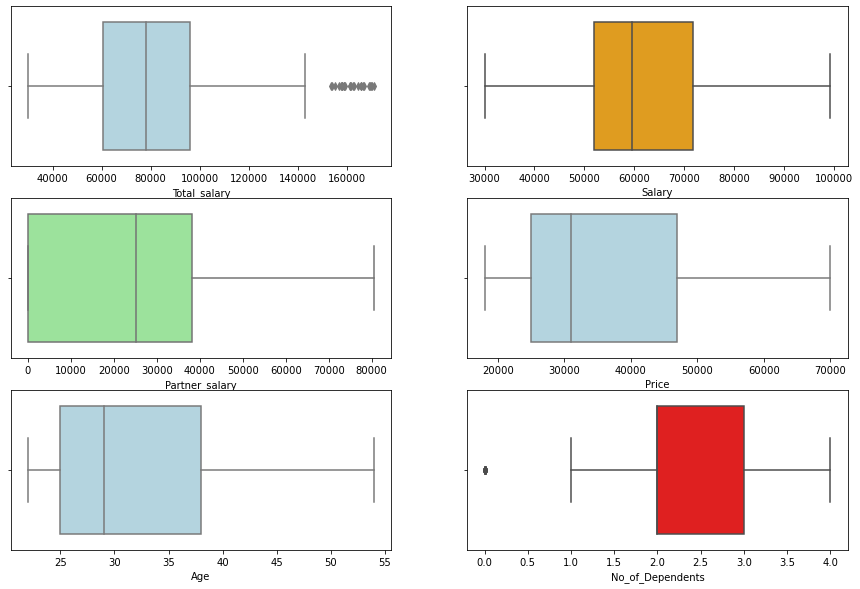


**Checking the duplicates** - Number of duplicate rows = 0

So, we will further move ahead and check for Outliers in the dataset

**Outliers Treatment**

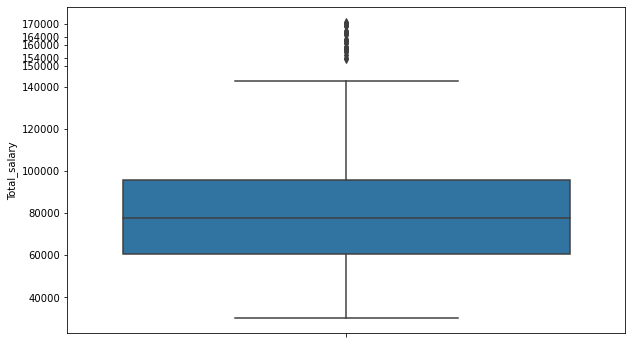
**To check for outliers in case of the continuous variables, we will be plotting the box plots, which are as follows.**



**Looking at the box plot, we can examine that the variables Total salary and no of dependents has outliers in the dataset. These outliers cab be treated before we further proceed and there are several ways of treating them:**

* **Drop the outlier value**
* **Replace the outlier value using the IQR**

#### But it is not necessary to treat them always as they are not always the bad data as here in case of Total salary. As we can see Total salary shows some outliers after the Q3 value. so first let’s find out what percentage of the given data is present in the outliers and is it really important to treat them or we can easily go on with same dataset and further do the analysis



### We see that nearly 2.1% of the data in the "Total salary" represents outliers, and in all the cases we have observed that both the partners are working and hence it is normal to have values higher than in case of both partners not working. So, this is not really required to be treated as it does not represent the bad data. Also, in case of "No of Dependents" the outlier represents the value 0 which is quite normal for those who are not having any dependents with them. So, we are good to go with the outliers in both the cases

# Section-C

## Now let’s Explore all the features of the data separately by using appropriate visualization technique and draw insights that can be utilized by the business. This includes both Univariate and Bivariate

**Let’s form some questions so that we can derive proper insights from the data based on those questions which will help the business to enhance its marketing strategy**

**1)What is the current status of the MAKE? Which car is in demand when it comes to married couple than to singles?**

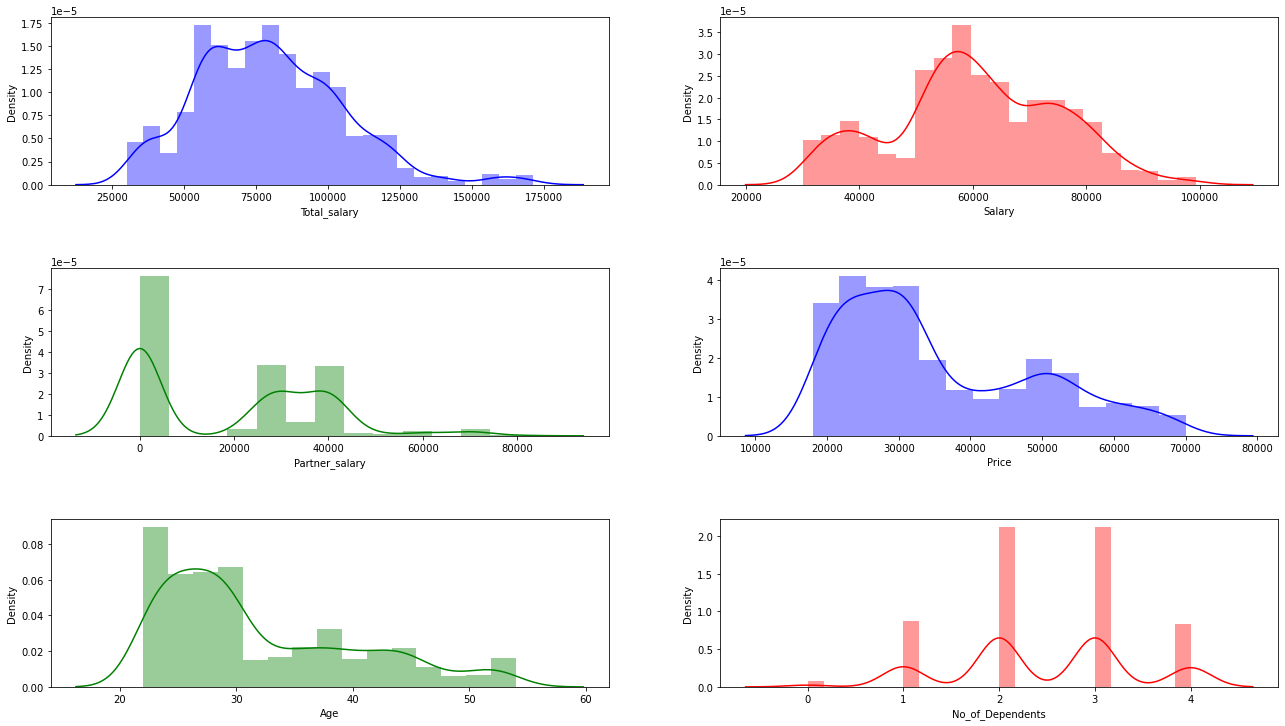
**2)How is the profession plays an important role in purchasing the car?**

**3)Which car is fav when it comes to different Genders also what are the chances when their marital status changes?**

**4)Is No of Dependents really plays an important role when it comes to purchasing capacity?**

### **Univariate Analysis**

****

****

**The skewness values can be interpreted in the following manner:**

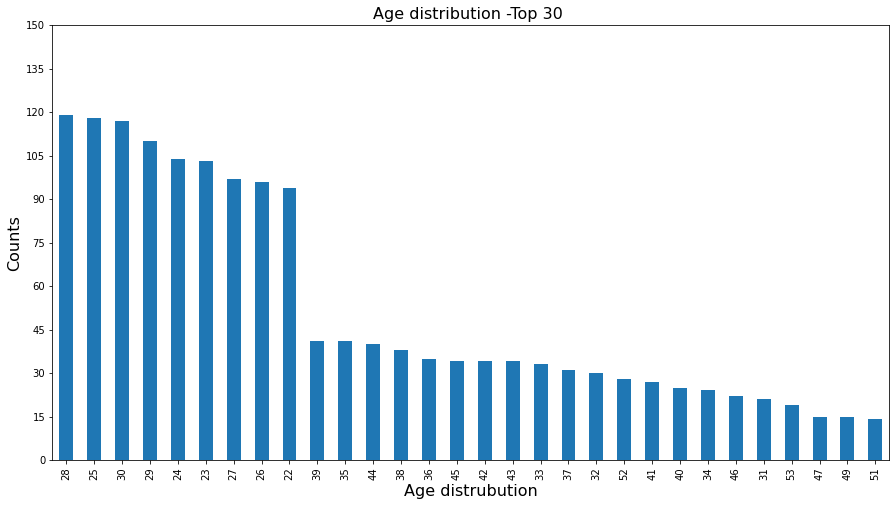
1. **Highly skewed distribution: If the skewness value is less than −1 or greater than +1.**
2. **Moderately skewed distribution: If the skewness value is between −1 and −½ or between +½ and +1.**
3. **Approximately symmetric distribution: If the skewness value is between −½ and +½.**

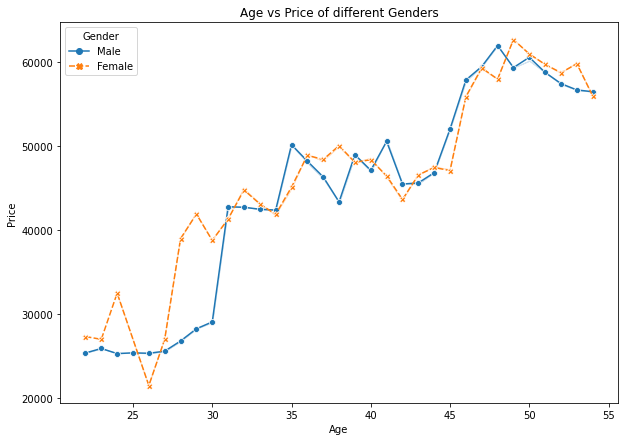
### From above figure, we can say that the Negative values for the skewness indicate data that are skewed left and positive values for the skewness indicate data that are skewed right. The skewness for a normal distribution is zero, and any symmetric data should have a skewness near zero**. So No of Dependents = -0.129685 and Salary= -0.011560** here are a bit **left skewed curves**, salary being the almost symmetric though and rest four are rightly skewed curves that too moderately.

### Now let’s do Analysis one by one.

### First, we will talk about the variables Age and No of Dependents and see their distributions and their statistics

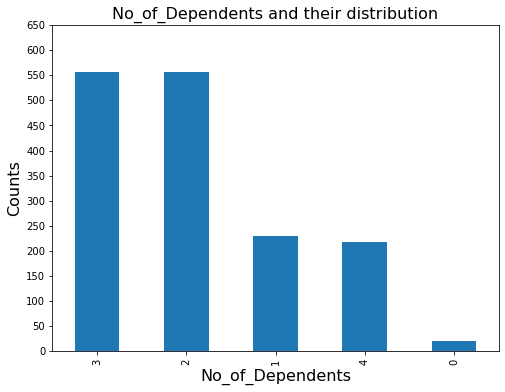
**Age Distribution**

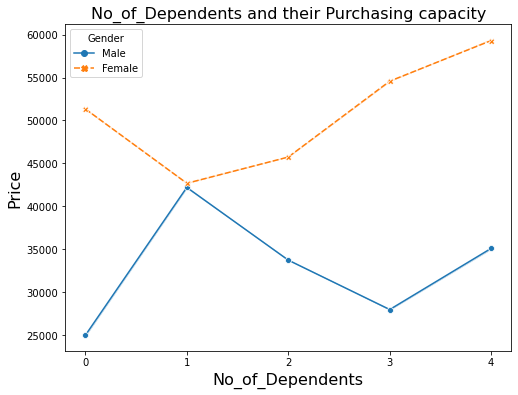
****

****

#### So, from the above data we can observe that maximum No of counts for Age distribution of the customers lies in between 22 to 35 years with the maximum value 28 and then it slowly decreases as the age increases with few exceptions here and there. So, the company should more focus on people who are in their late 20s and early 30s if they want to add new customers to their group, especially younger males. They should make strategies so that they can attract more customer base from this Age group to buy more higher priced cars. Also, they should make different strategies for customers who are much older and nearly in 50s because they are also be valuable customers if not potential customers and clearly, they are buying more valuable cars as they gain experience in their work field.

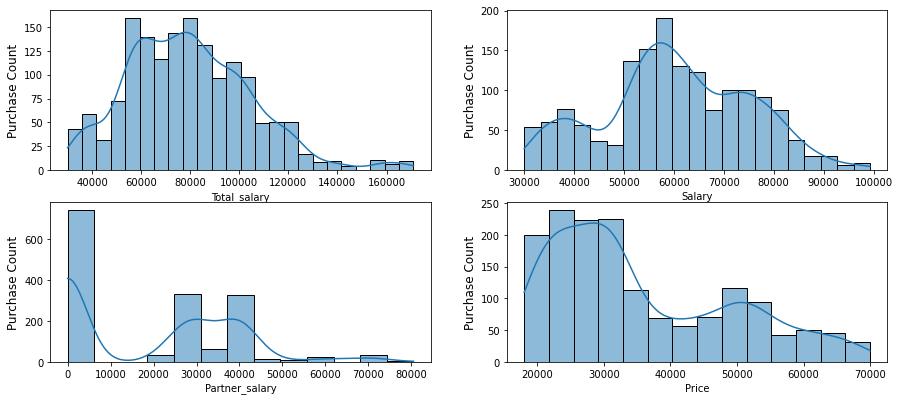
**No Of Dependents-**

****

****

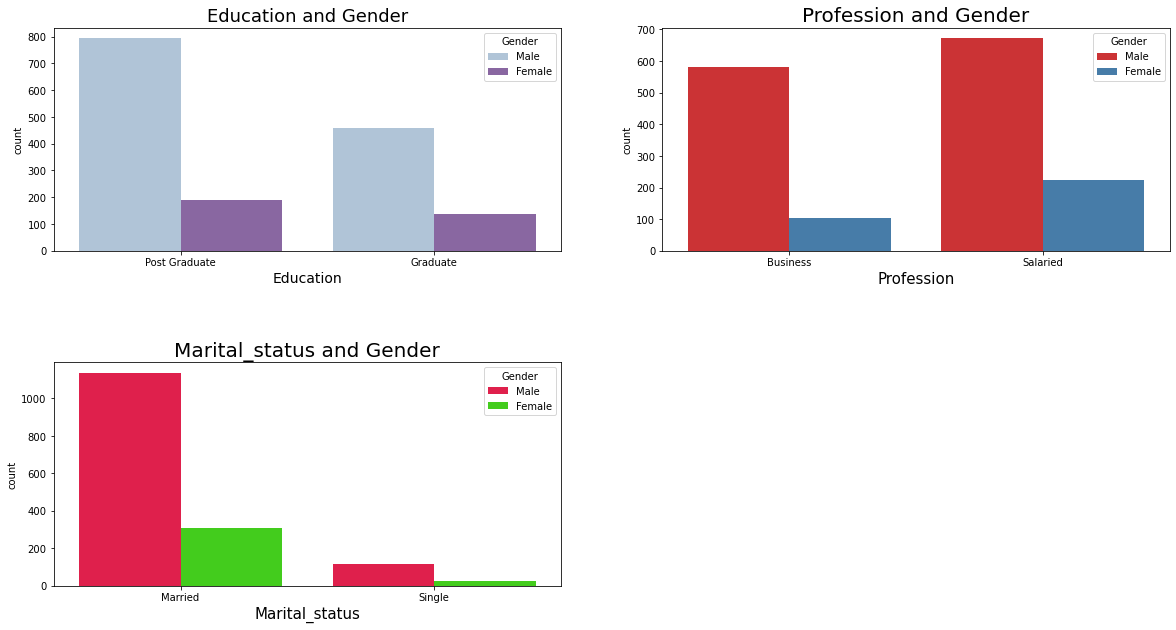
### We observe from the data that maximum number of customers have 2 or 3 no of dependents on them, but still females are much more valuable buyers here than males. So, strategy should be made by the marketing team to attract the male customers more despite their No of dependents so that they can come in completion with females and increase the sales of the company.

**Distribution of all the Numerical Variables**

****

## Bivariate Analysis

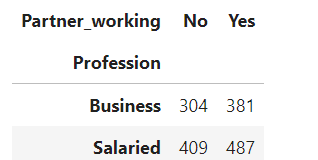
### Relationship between Education and Gender, Profession and Gender of customers and Marital status and Gender of the customers

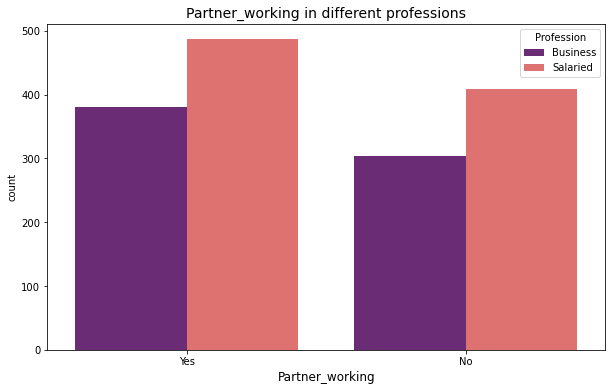
****

**Here we can infer that out of 329 females almost around 200 female customers are preferring salaried jobs when compared to Business and similarly for Males out of 1252 males around 700 are into salaried jobs. We can also interpret that most of the male and female customers in the dataset are Post graduated when it comes to educational qualification and most of these male and female customers are married, than that of singles.**

**So, it can be predicted that if the company goes for well settled customers in terms of marriage and profession, they are more likely to make profit in the coming future.**

**Similarly, now we will look in terms of Profession**

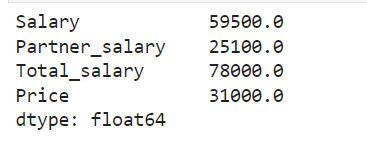
****

****

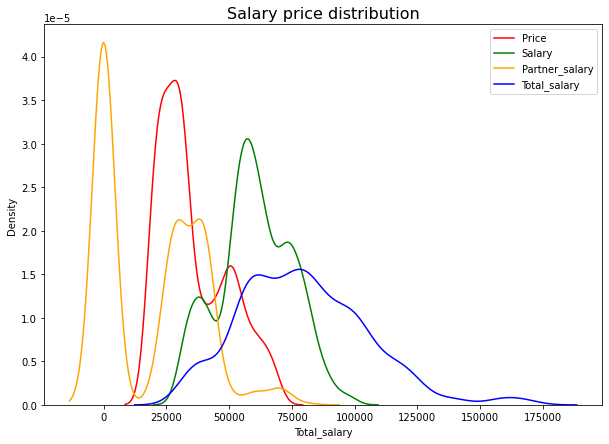
So, we can easily tell from the above graph and analysis that most of customers here are preferring for salaried jobs than going for business despite of the fact that their partners are working or not working. So Austo Motor should make a targeted move for salaried customers by making different strategies for these two categories.

**Salary, Partner salary and Total salary**

We will find out the medians of the given variables first that will help us to differentiate between various clusters based on those numbers. The medians are as follows:

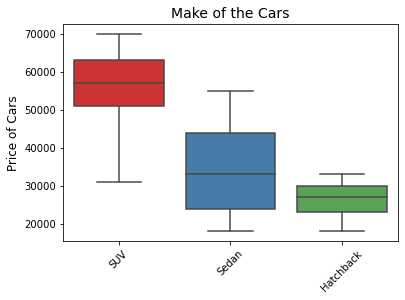


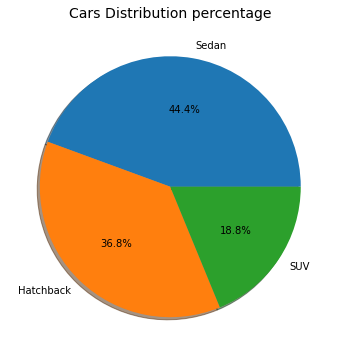
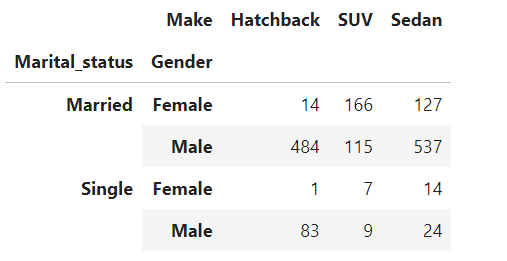
Plotting a density distribution curve of the Continuous variables to see their variations

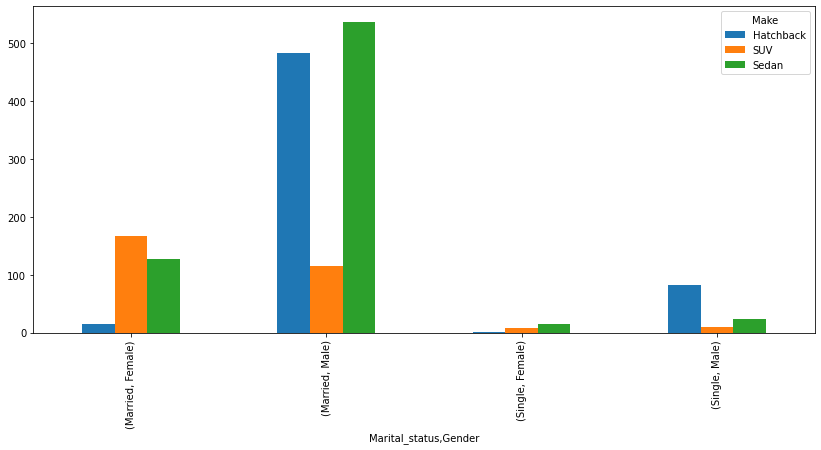


#### As we observe the Density plots of various kinds of salaries and the price of car, we can see that Partner salary has maximum frequency as 0, which means in most of the cases partners are not working, be it male or female, and when they are working, most of them are getting paid near 25000. This combination of salary and partner salary takes the graph of total salary in the range of 150000 to 175000, where 57100 represents the maximum frequency. Price of car ranges between near 17000 to around76000 and 32000 being the maximum occurring number which shows that this is the most preferred range by the valuable customers.

### **Make of Cars and their Price Distribution**



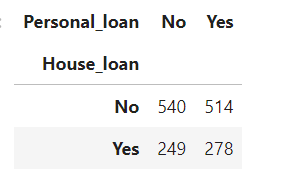
 



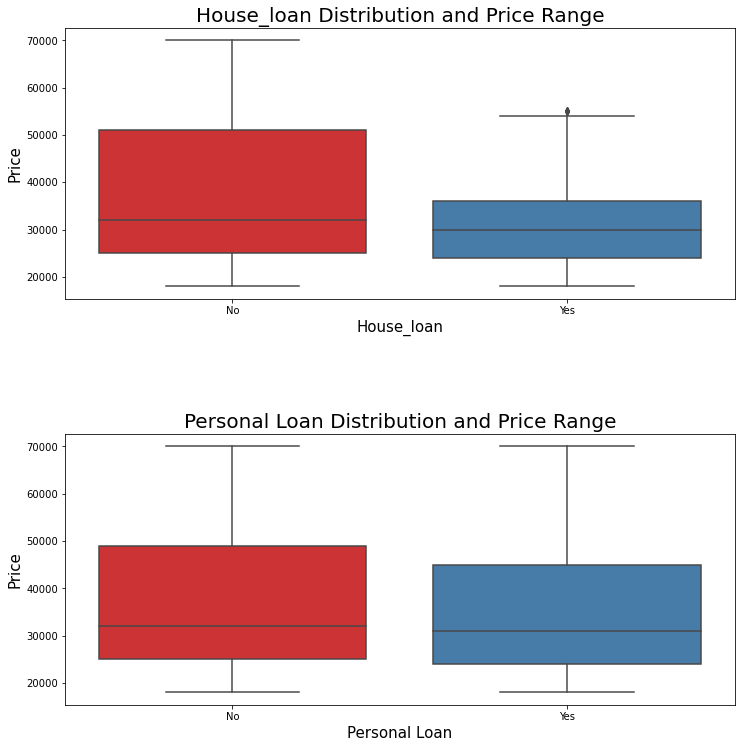
So, it’s evident that although SEDANs and HATCHBACKs are bought more in number by the customers the price distribution is highest for SUVs, and the reason being its price range which is quite higher compared to these two, especially HATCHBACKs.

#### We can also say that here, Married people are more valuable customers than that of singles. The observation says that for married males SEDAN is more preferred but for Married Females, it’s the SUVs. But when it comes to Single Males SUVs are the fav choice followed by SEDANs and then HATCHBACKs. Although the single females are not potential customers for the company, but they can be ‘focused customers’ if the company needs to increase the sales specially in SEDANs followed by HATCHBACKs for single females

### **How Personal Loan and House Loan affects the purchasing power of customers let’s get an insight from the visualization data**



So, this is the table representing the customers and their Loan distribution in the two categories.



The box plots plotted above clearly shows that people who are having personal loan are not cutting the cost for their cars much but when it comes to House Loan, it does have an impact on the price range of car people are choosing.

#### House loan does impact the purchasing power, personal loan is not much significant in this case, although it shows some variation in the median and Q3 value. But the customer who has the liability of House loan is opting for less priced cars than that of No liability. So, company can target these customers and provide them with some better deals to increase their sales in these potential areas

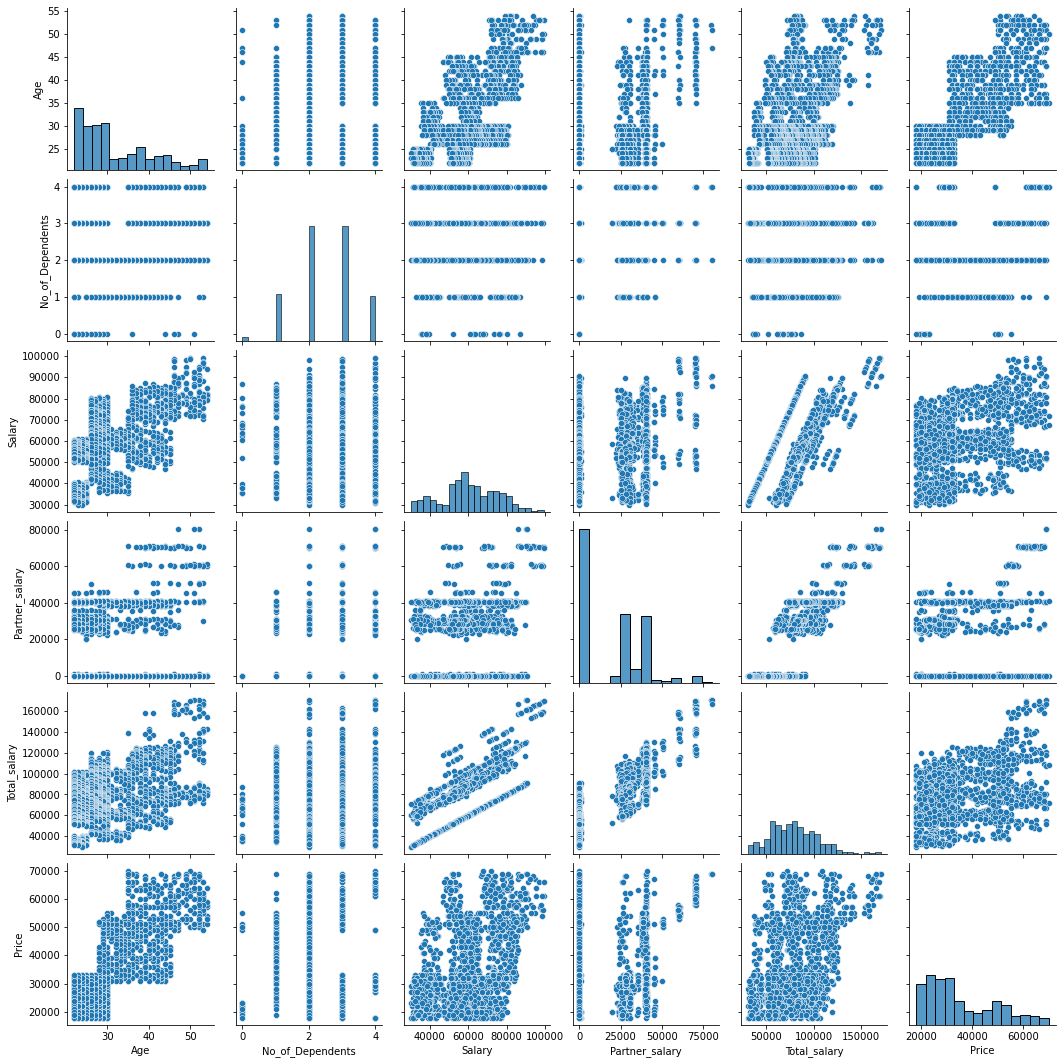
# SECTION -D

### Let’s form some questions so that we can derive proper insights from the data based on those questions which will help the business to enhance its marketing strategy

**5)How is Age related to the Car market and their Make?**

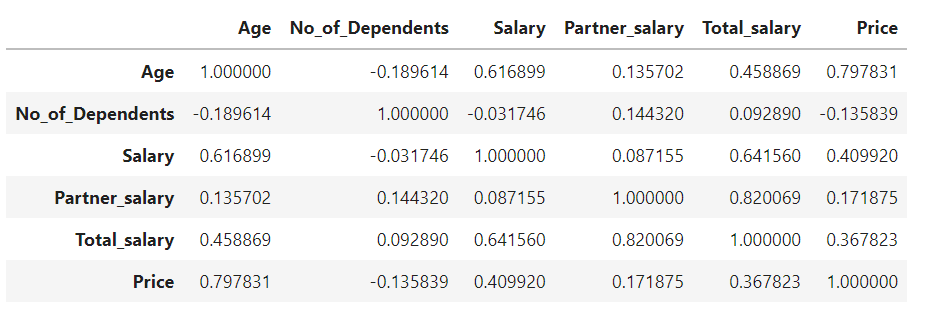
**6)How does Education and Profession affect the buying capacity (Make of car and Price of Car)**

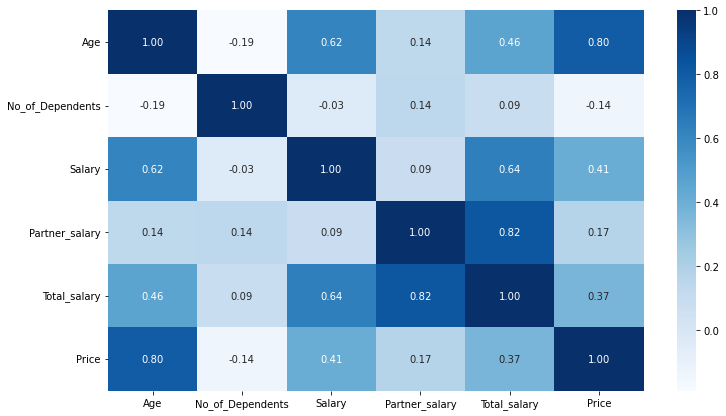
**8)Does Personal loan and House Loan Together make big impact on the buying capacity?**

****

## Correlation Heatmap

#### Before starting the relationship analysis lets plot a correlation heatmap for each variable and see how they are correlated to each other and what is their extent of correlation

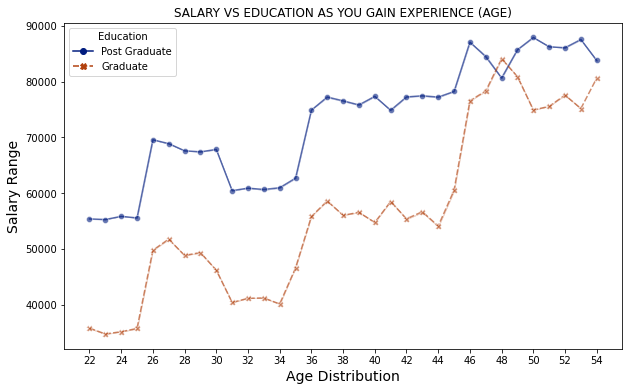




### From the heatmap we can see that Partner salary is highly correlated to Total salary, so the customers whose partners are working they together make a good Total salary. Also, Age is highly correlated to the Price of cars and for doing further analysis we will observe each of these variables together on a separate graph plot to get the useful insights.

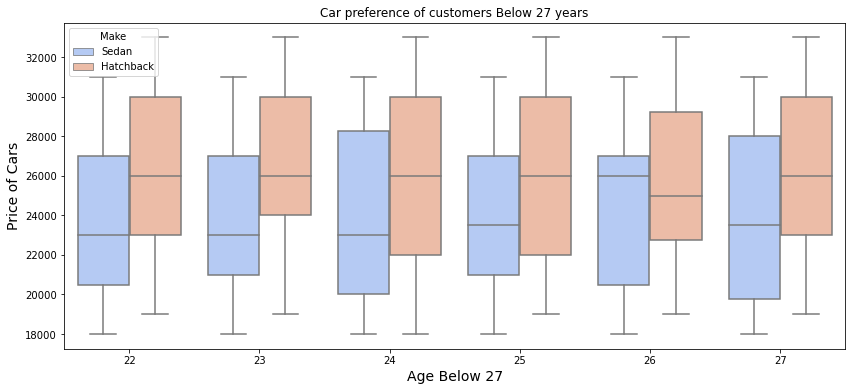
## How is Age related to the Car market and their Make

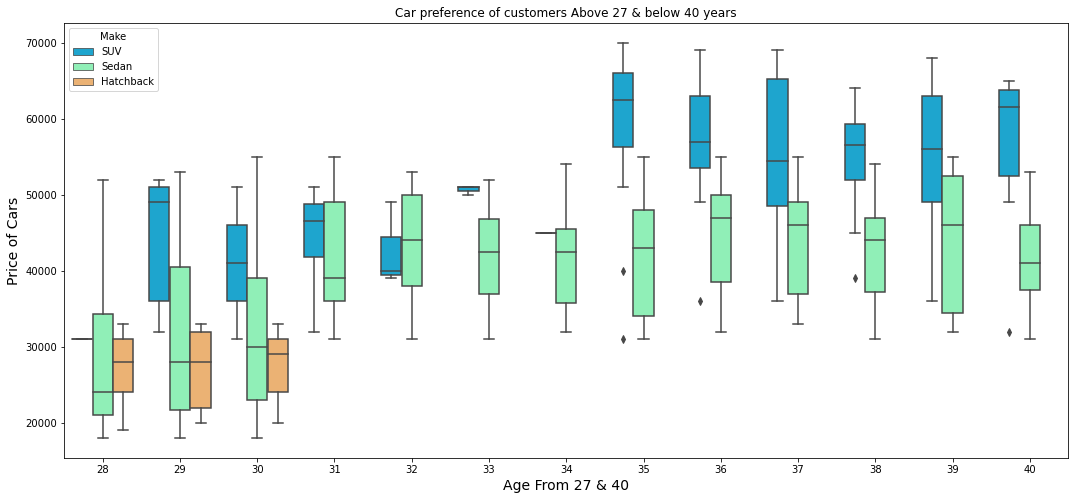
### It’s a common perception that as the Age increases our knowledge and experience increases (be it Business or Job) and so do our salaries, but Educational Qualification and skills are something which will always keep us ahead of our peers despite our young Age. So, let’s first plot a graph and observe that how our data behaves as Age increases along with Education pattern and salary obtained.

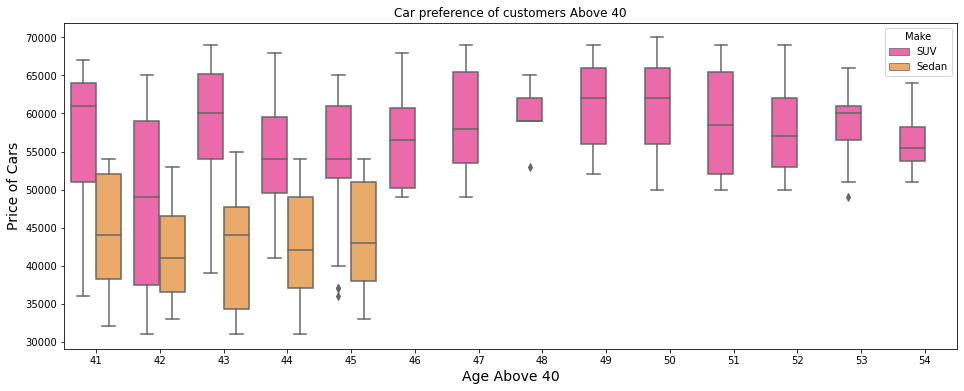


### As We can see Post graduate people tend to get more salary than graduate people from the beginning and the salary increases as the experience increases (Age), with some exceptions in the last stage, now we will differentiate the Age groups and then let’s find out about the car Makes and the purchasing power of the people.

**Age group of in the data frame and study of different clusters**

****





#### So here we have defined three categories of Age Groups and tried to find out Car preference of the customers. The three age group clusters are

* Below 27 years
* Between 27 and 40 years
* Above 40 years of age

#### From the Above three graphs its quite evident that people in the Age group below 30 are mostly preferring HATCHBACK and SEDAN where Hatchback is of slightly higher Range in that cluster, but as we move forward, we see that in the middle Age group as the salary of customers increases their preference mostly shift towards SEDAN and SUV of higher price range although some people still are Comfortable with HATCHBACK. But then in the last cluster, in the Age group above 40 we see mostly SEDANs and SUVs that to of higher price range compared to that of the middle cluster. It clearly shows that company can focus on the desired customers for targeted sale of Cars

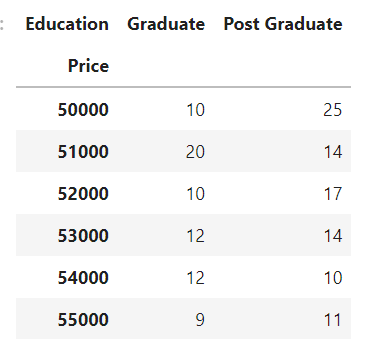
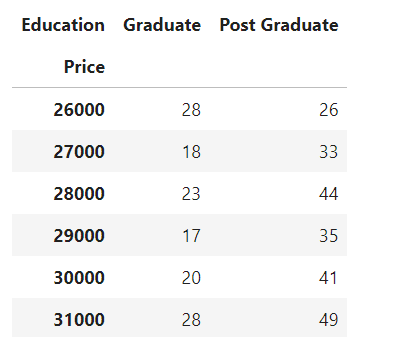
## How does Education and Profession affect the buying capacity (Make & Price of Cars)

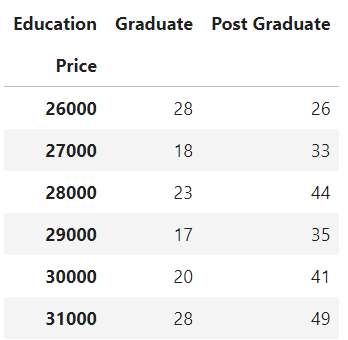
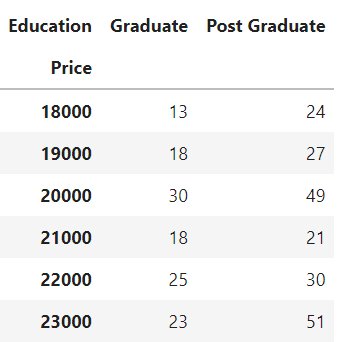
### We have already seen the salary increases when experience increases now let’s focus on the Education and Profession of the customers and find out some insights regarding the dataset.

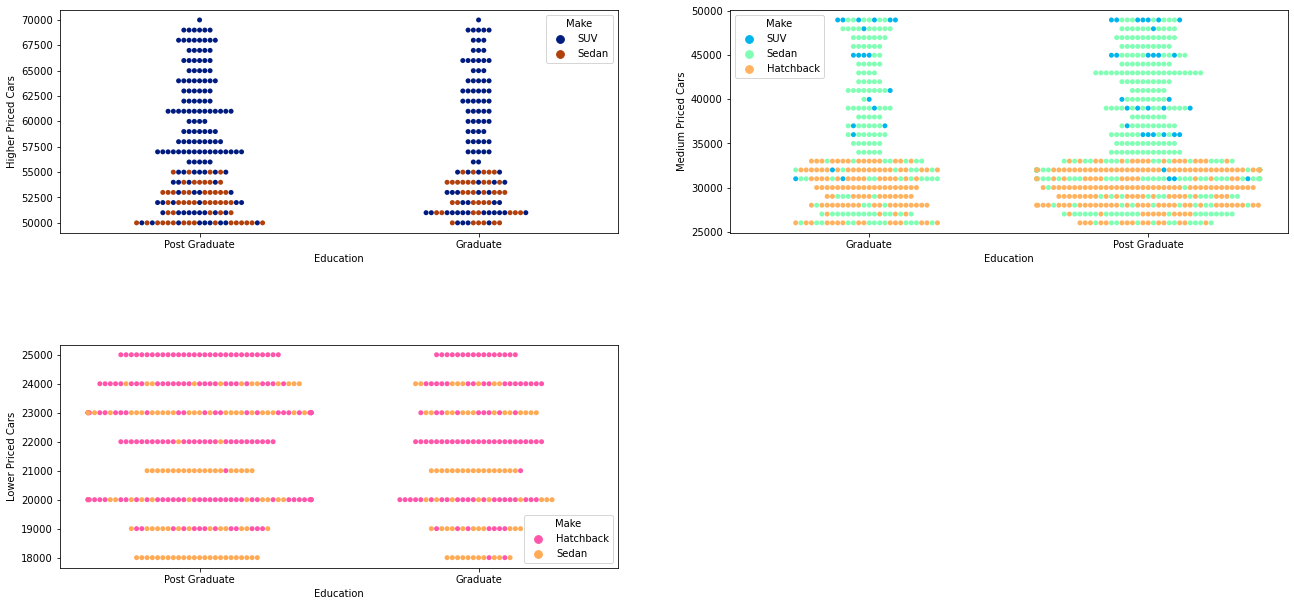
#### For this lets first set a range for Higher, Medium and Lower priced ranged cars for reference.

* **Higher priced cars = ['Price'] >= 50000**
* **Medium priced cars = ['Price'] <50000& ['Price'] >= 26000**
* **Low priced cars = ['Price'] < 26000**

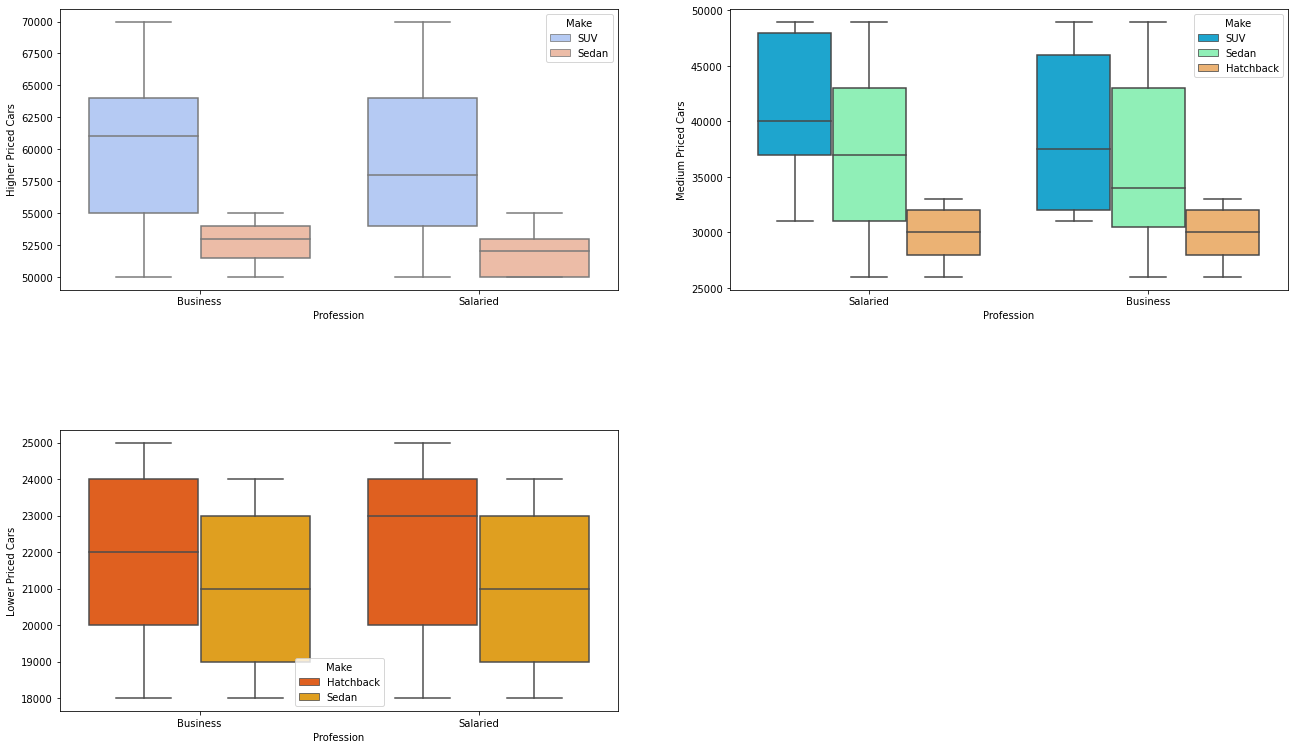
Checking the frequencies of top 6 values in each of the clusters and this is what we got as the output.



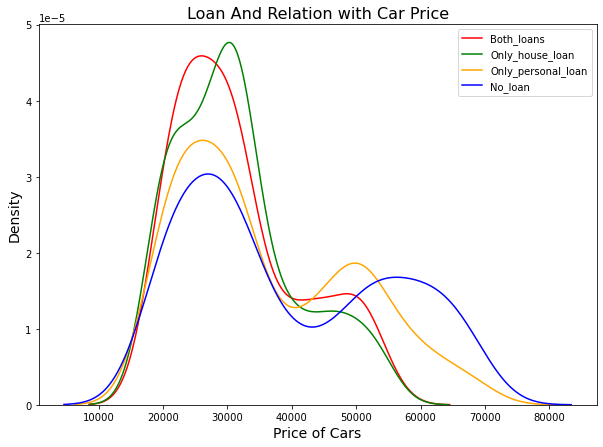
#### So, what we can infer from the above swarm plots is that as the educational qualification of customers increases their buying capacity increases and they prefer higher priced cars Where SUVs are the highest of that order**, SEDAN** being the second favorite and **HATCHBACK** being the third. Also, we can observe that medium priced range cars are denser than the other two clusters, so mostly post Graduates are preferring **SEDANs and HATCHBACKS** in the range **25000 to 35000** and then slowly shifting towards **SEDANs and SUVs** further in the range **35000** to 50000.Graduates although prefer the same trend but they are still lagging behind than that of Post Graduates when it comes to purchasing capacity. SUVs are most purchased cars in the higher range cars and their density increases as the price increases.



#### We observe that Profession doesn't play a very significant role in the Choice of cars, its rather their Price and the customers’ Capacity. Although their Medians differ in two cases their Max and Min range remains almost the same.

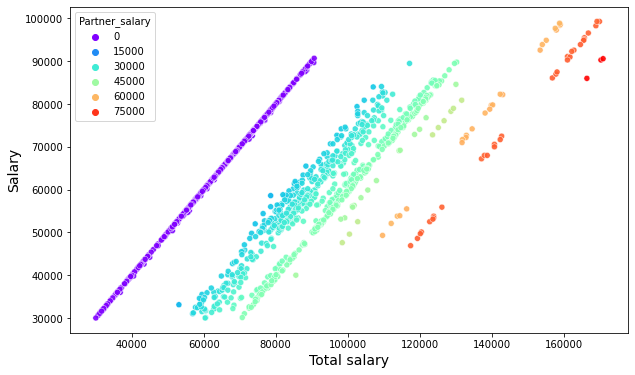
### **Does personal loan and House loan Together make big impact on the buying capacity?**

#### We have already seen the distribution of House loan and Personal loan separately in section C. Now let’s combine these two and see how it shows the variation with Car Price and type if a person is paying both the loans together and not paying any loans. For that we will have four cases so we will make four data subsets for the same.



These are very insightful density plots here which actually tells us about the Two Loans liabilities and how they impact the purchasing capacity .We have already seen that House loan does impact the purchasing power , personal loan is not much significant in this case **, but when we combine the two loans it shows a greater impact and reduces the Highest car price to around 60000 to** **65000** ,which is not the case when a customer has No Loans liability or only the personal loan liability, since the range then **increases up to 80000**

### **Now let’s see the relationship between Total salary and salary in case of working and non-working partners**

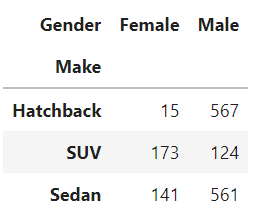
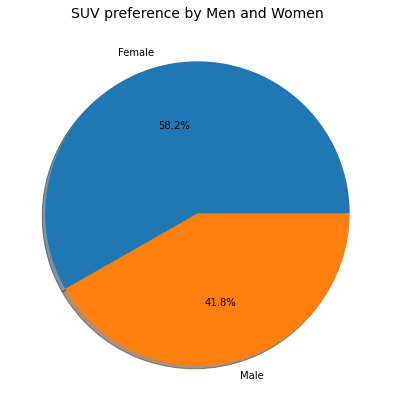


**Total salary automatically increases when the other partner is working which is clearly visible from the graph, although working partner doesn’t play a very significant role when it comes to purchasing capacity.**

# SECTION-E

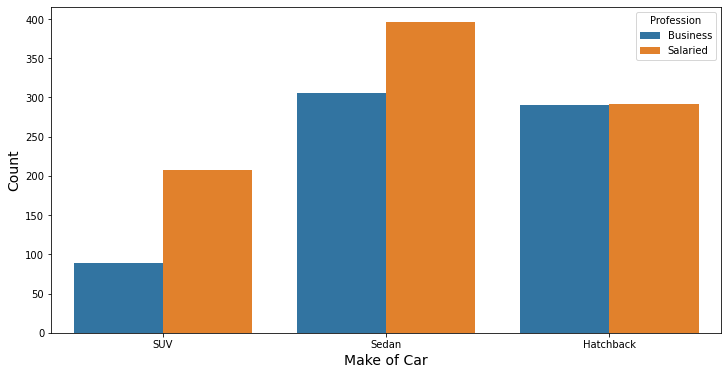
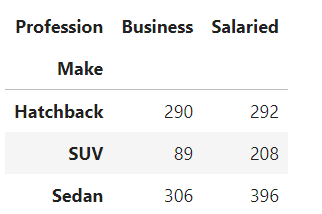
### Employees working on the existing marketing campaign have made the following remarks. Based on the data and your analysis state whether you agree or disagree with their observations. Justify your answer Based on the data available.

### **E1) Steve Roger says “Men prefer SUV by a large margin, compared to the women**”

### So, we have made a data frame named ‘df\_SUV’, where Make==SUV and then we have found out the value counts of men and women, and we can clearly see that women are preferring SUVs more clearly by 18%. So, the statement by Steve Roger is FALSE

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

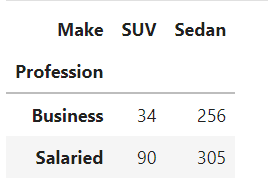
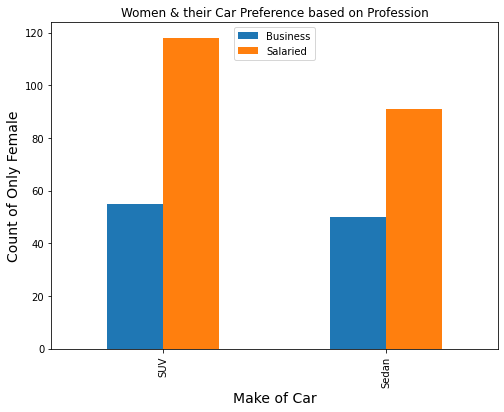
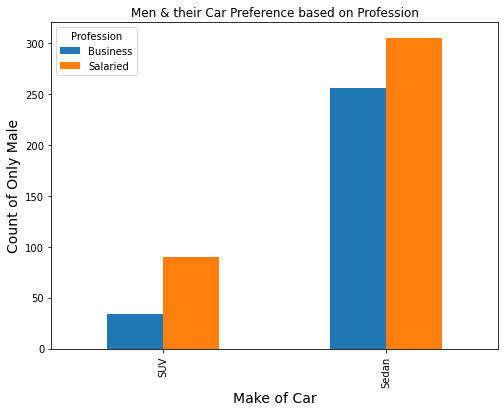
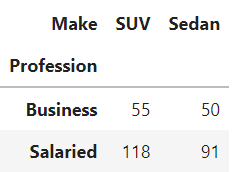


### The statement by Ned Stark is TRUE. It’s clearly visible that salaried person is more inclined towards SEDAN than that of business-oriented person. In fact, salaried person prefers both SUVs and HATCHBACKs also more than that of Business oriented person

### **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.**

#### As Sheldon Cooper is only interested in SUVs and SEDANs so we are making a subset of data frame for both Male and Female separately which does not include HATCHBACKs

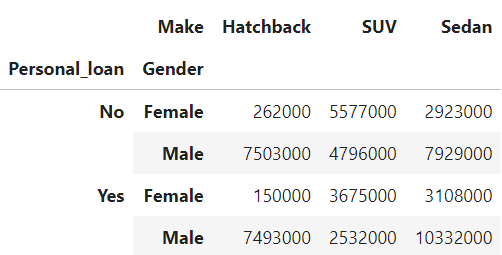
The two frequency tables show the number of cars owned by Men and Women respectively, and from the below graphs we can say that the marketing strategy of Mr. Cooper is not up to the mark and it’s a false statement.

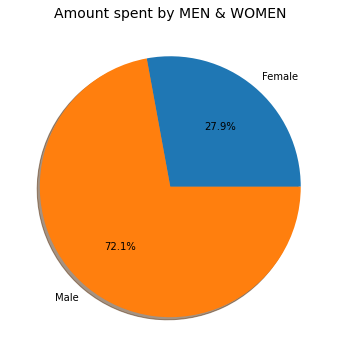
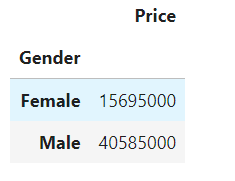
 

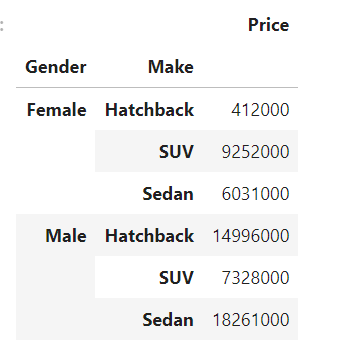
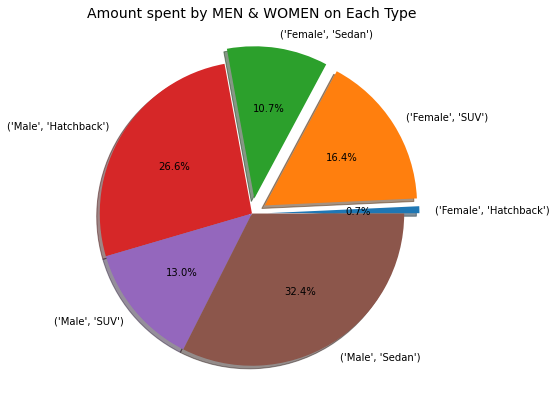
# SECTION -F

### Let’s research about the amount spent on purchasing automobiles across the following categories and see how a business can utilize the results from this exercise by arriving at the conclusions**.**

**F1) Gender**

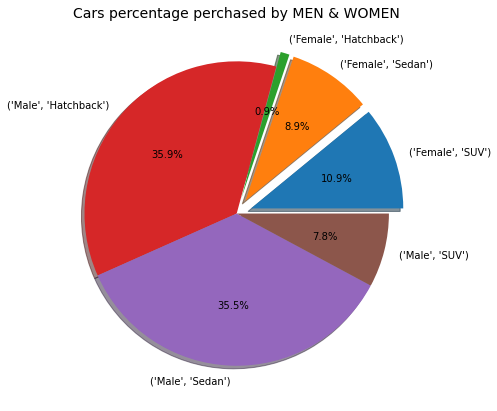
****

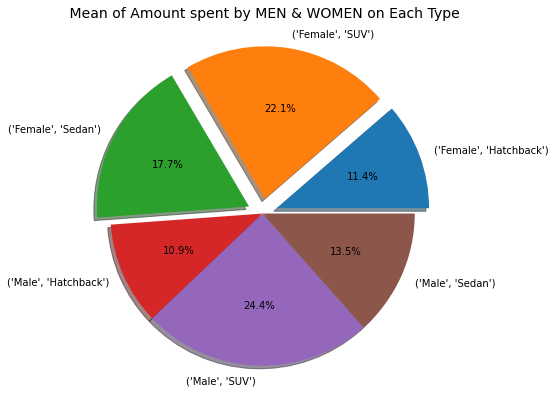
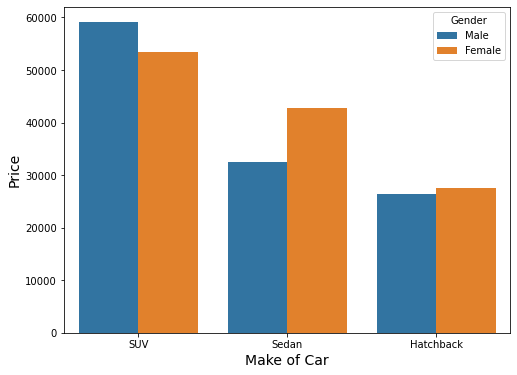
**** ****

** **

### From the above pie chart, we can observe that the from **most amount of revenue is generated** the Males who invested on SEDANs followed by Males who bought HATCHBACKs and 3rd being the Females who are going for SUVs. Males who are preferring SUVs are least as compared to males who are buying **SEDANs and HATCHBACKs** In case of women, SUVs are the most invested car, second being SEDANs and least being HATCHBACKs

#### But we also know that when it comes to pricing of the cars highest priced cars mostly contains SUVs with a few percentages of SEDANs and mostly Medium-priced cars are bought by customers which **range between 26000 to 50000**. So, if company wants to increase its revenue, it should focus more on the sale of SUVs in bulk specially by males as they are lagging behind in this type from the other two and company should overall focus on the Female customers so that they come at par with males in the purchasing capacity.

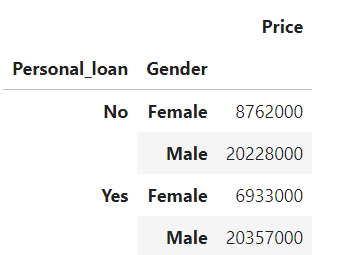
****

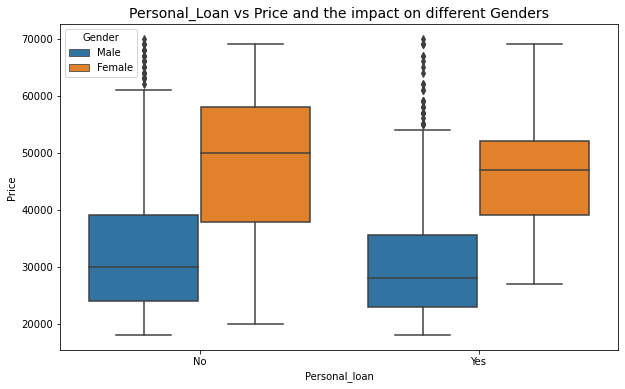
** **

### When we plotted the price of cars against their type, we can see that the mean of SUV is much higher than that of SEDAN which is second and HATCHBACK which has the least mean value ranges. So, although the Males have bought a smaller number of SUVs than females, but it is quite evident that the MEAN amount in case of SUVs for males (24.4%) are slightly higher than that of females (22.2%). In case of SEDANs the female mean amount is higher than that of male mean amount and in case of HATCHBACKs its almost similar in both the cases.

### So, the company should focus more on the sales of SUVs and SEDANs in bulk, by giving attracting deals and opting for a good marketing strategy here, and focus on the potential customers.

## F2- Personal Loan

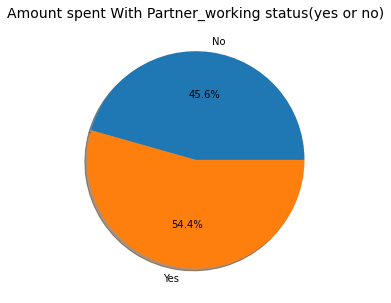


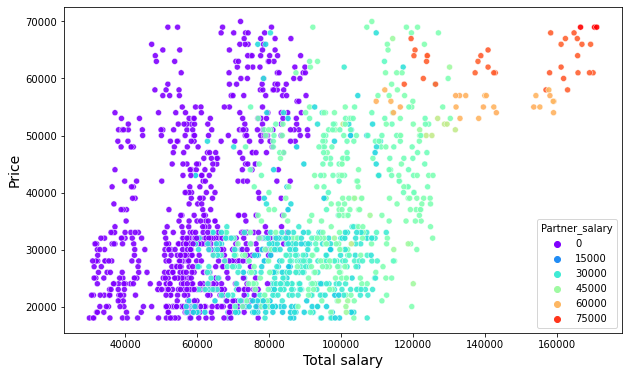


### It’s evident from the Boxplot that women are potential customers when it comes to higher priced cars whether they have taken personal loans or not, although in case of No personal loan clearly their spending capacity is slightly higher than that the case when they are paying loans. On the other hand, men are not much impacted by the loan and their purchasing power needs some improvement, which can be done by great deals and focused approach by the company

# SECTION -G.

### **Working partner and their analysis on the purchase of a higher-priced car.**

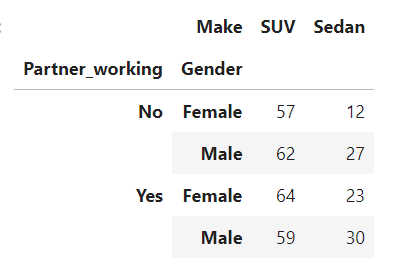




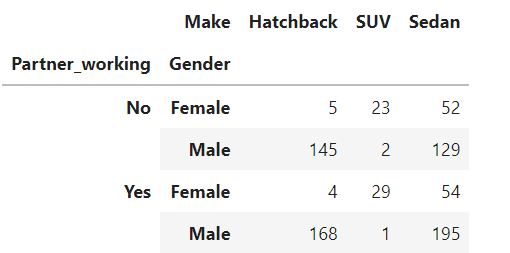
### Let’s set a range for cars which are higher priced, medium priced and Low priced. So, let’s assume the cars which range from 50000 and above are higher priced cars. For medium priced cars, we can assume that to be in the range between 25000 and 50000(since the median is around 31000) and below that is low priced cars. Now we will make the new datasets based on the following ranges and further do the analysis.

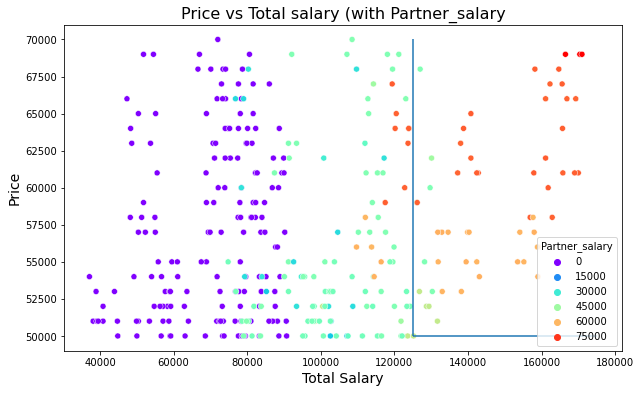
* **Higher priced cars = ['Price'] >= 50000**
* **Medium priced cars = ['Price'] <50000& ['Price'] >= 26000**
* **Low priced cars = ['Price'] < 26000**

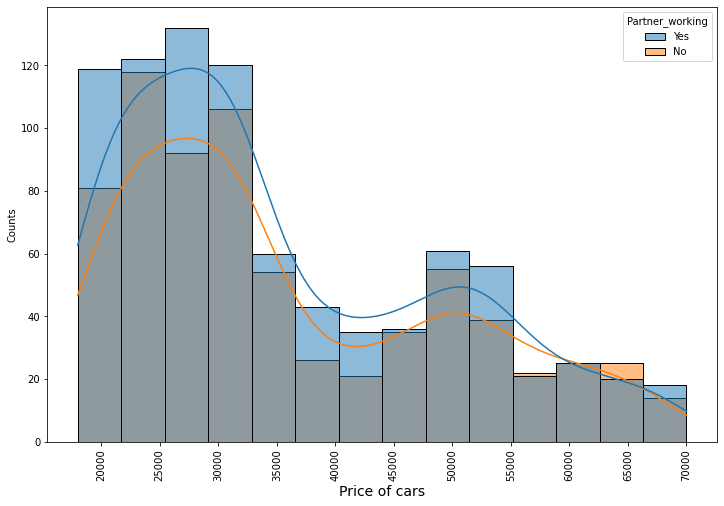
**Higher priced cars = ['Price'] >= 50000**

****

**Medium priced cars = ['Price'] <50000& ['Price'] >= 26000**







### So, from the graph analysis and data we can say that although working partners helps in increasing the purchasing capacity of customers but in case of higher priced cars the working partners doesn't play a very significant role specially when it comes to higher priced SUVs. In case of higher priced SEDANs, the numbers are noticeable though.

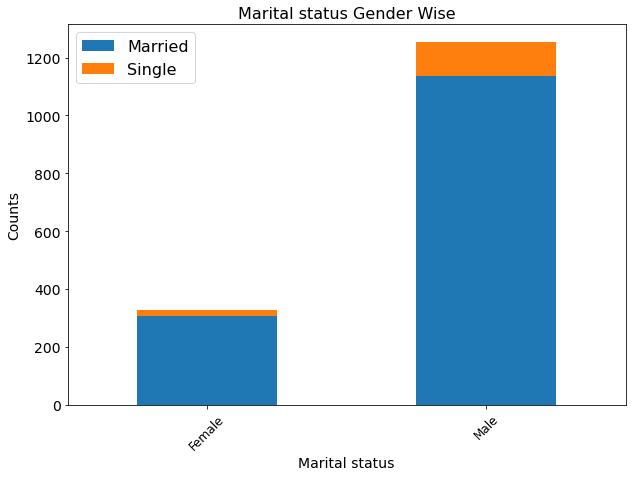
### We can see from the graph the highest priced car that is of 70000 is bought in both the cases when a partner is working and when he or she is non-working.

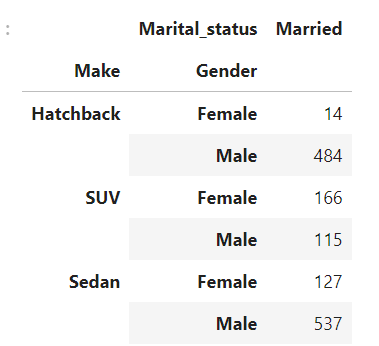
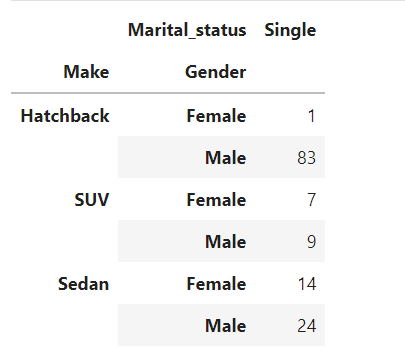
### But yes, working partners have some role to play when it comes to Medium priced cars as the numbers are quite significant in case of SEDANs and HATCHBACKs

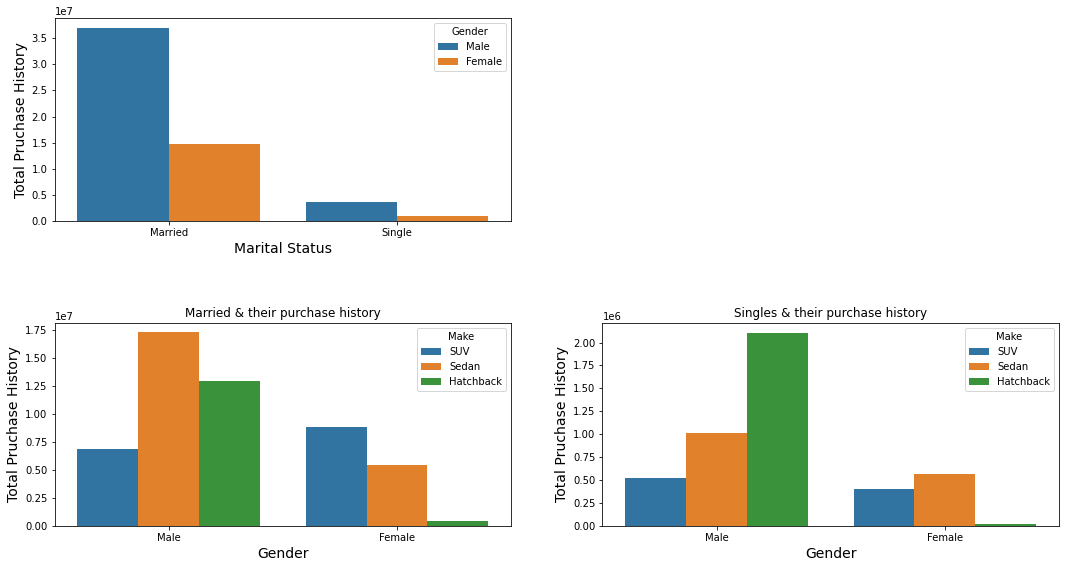
# SECTION-H

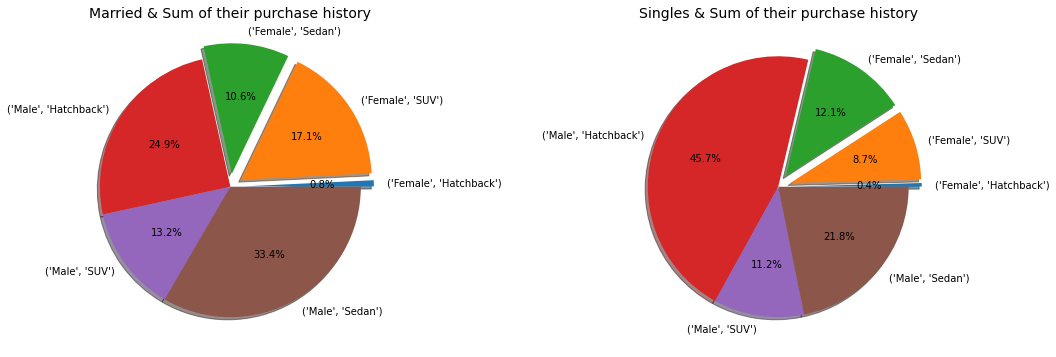
### The main objective of this analysis is to devise an improved marketing strategy to send targeted information to different groups of potential buyers present in the data.

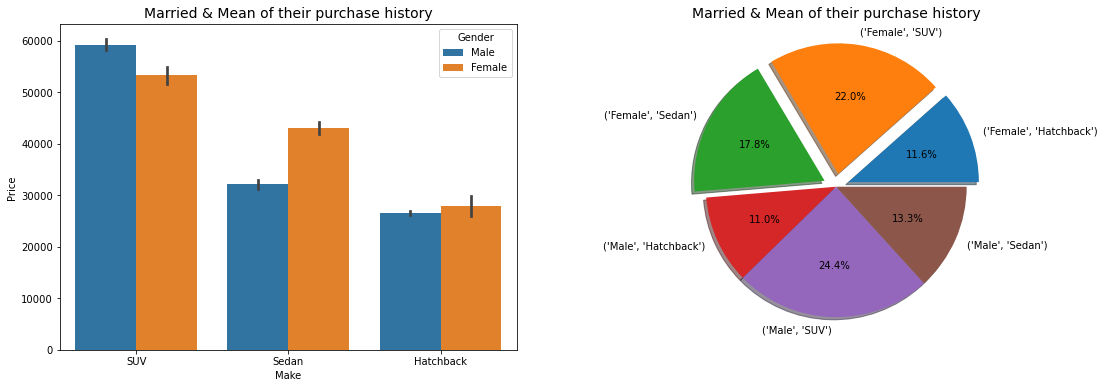
**Gender and Marital status - with similar purchase history.**

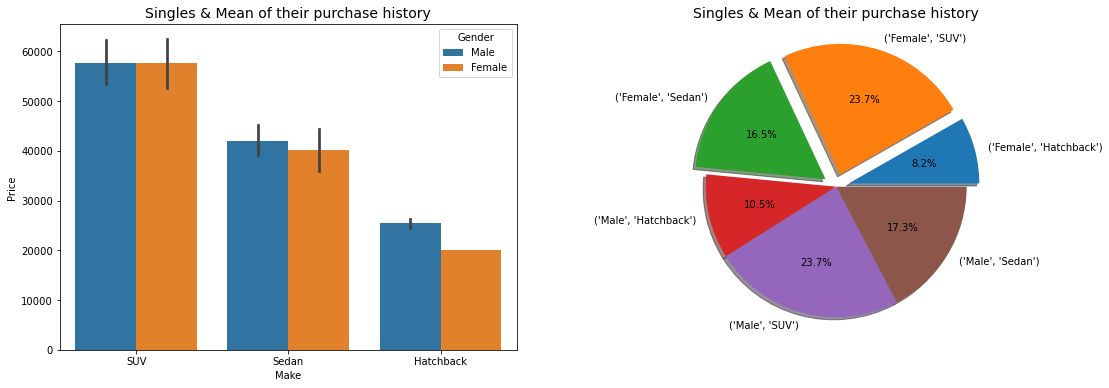
****









### **We can observe the following**-

#### So, we can see the various graphs and pie charts showing both the sum and mean price ranges of the Singles and Married customers separately. It shows that Mean values for cars are same in case of SINGLES who are having SUVs in case of both male and female & that is around 58000 and almost equal to 23.7% of total value.

#### FEMALES who are single and bought SEDAN has almost equal purchase history when it comes to FEMALES who are married and bought SEDAN (in terms of summation of Price)

**CONCLUSION-**

To lead in the market Austo Motor Company requires to modify its present marketing strategies and above analysis which consists of various useful insights including descriptive analysis of each variable can definitely help their new marketing approach, which will further help the company to increase its profit margin and establish itself in the market with the ever-increasing growing demand. These insights will be the needful for the company’s Growth as a whole

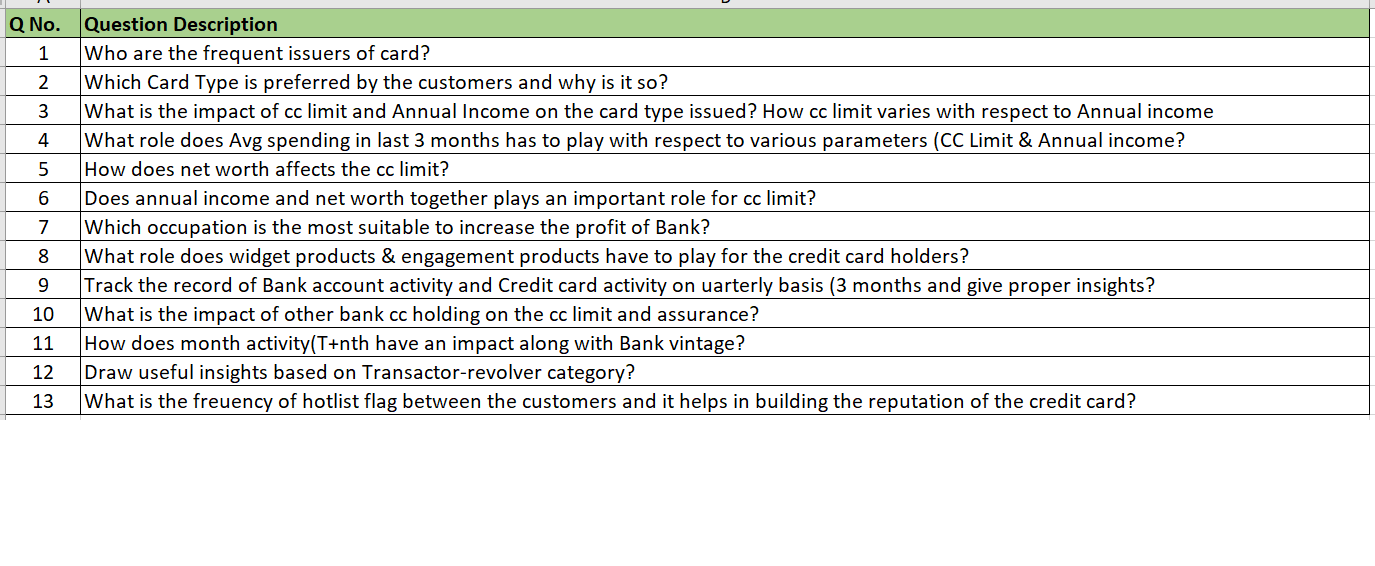
## PROBLEM-2

## GODIGT Bank credit card

### First, we will start with uploading the excel file using read excel () function and we will go through the dataset its variables and categories. We will try to understand that how data is behaving and then only we will be able to FRAME ANALYTICS PROBLEMS based on that

## Framing Analytics Problems

### Let’s frame some of the questions based on the data given and analyze which are the most important factors for the credit card company

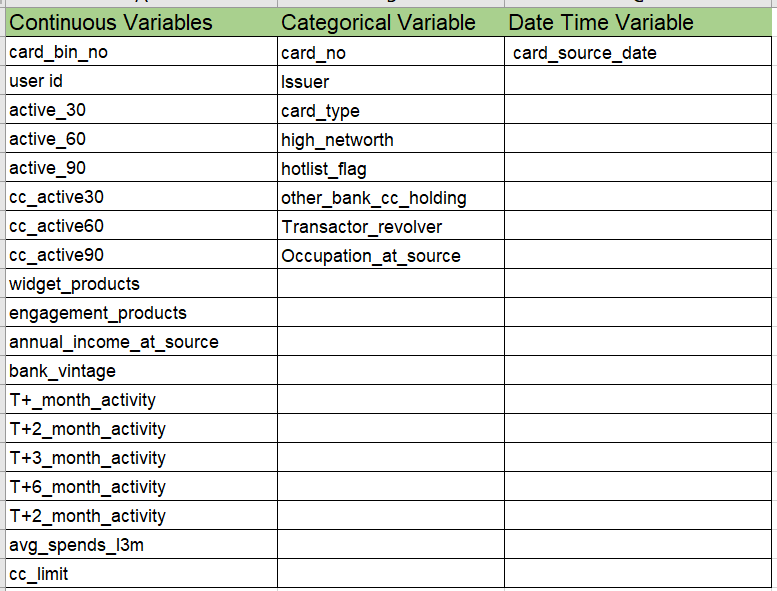
**

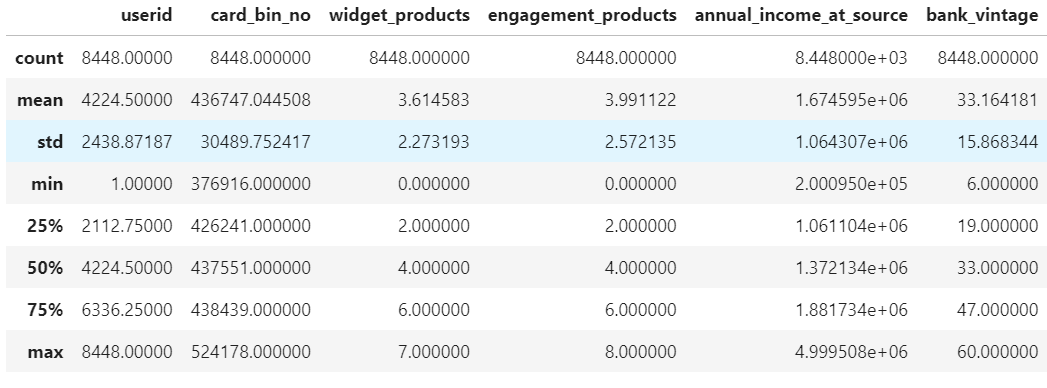
## Preliminary analysis

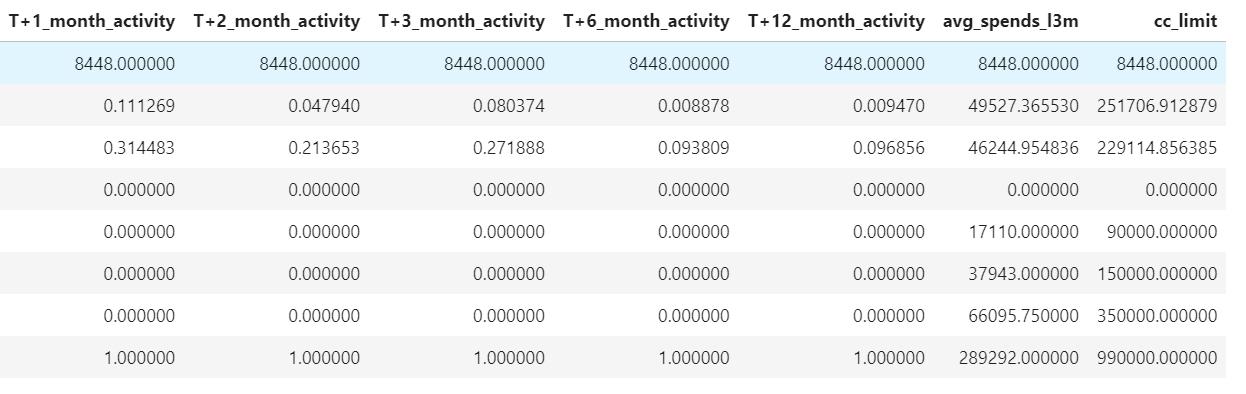
So starting with the analysis, We found out that the shape of the data set has 8448 rows and 28 columns in it with following data types -

* datetime64[ns] (1)
* int64(19)
* object (8)

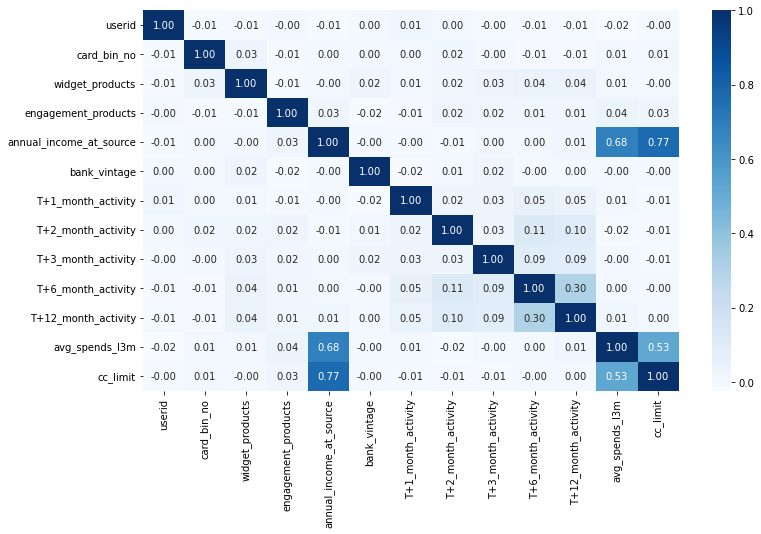
We have divided our columns into various categories so that we can do further analysis. The categories are in the below given table.



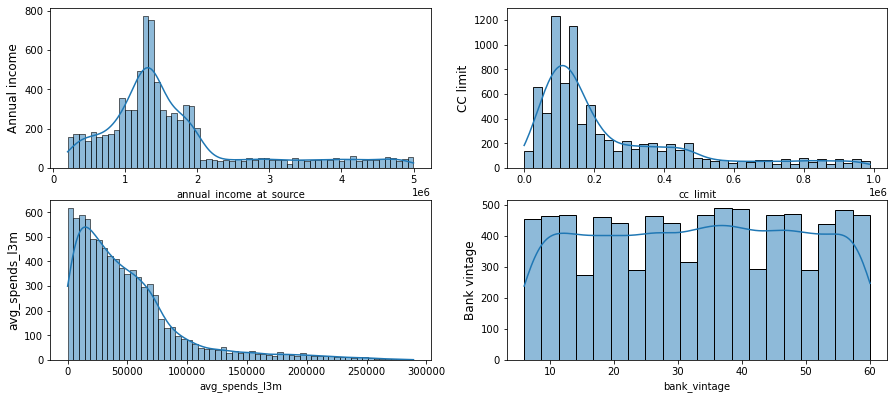


**CORRELATION HEAT MAP**

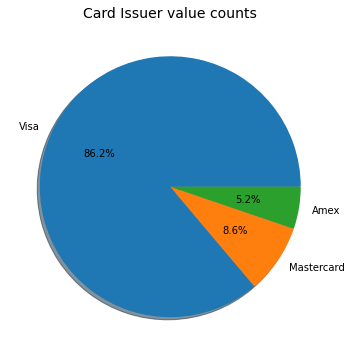


## Histograms of various numeric variables to see the distribution of the data



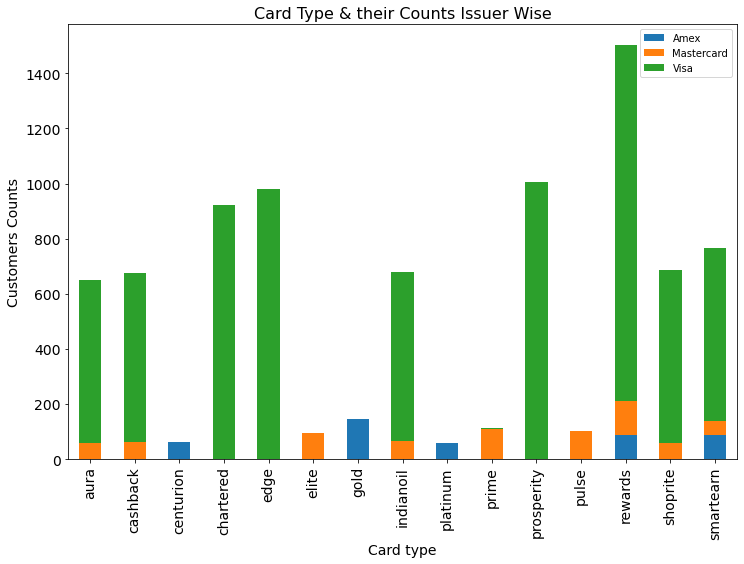
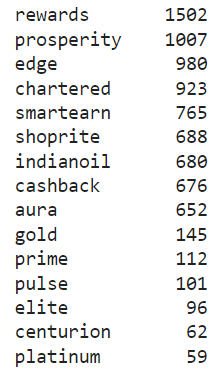
### Skewness with values annual income at source 1.446943, credit card limit 1.512063 and avg spends in last 3 months 1.888457 shows they are highly skewed data and that to rightly skewed but values such as bank vintage-0.016884, widget products -0.060988 and engagement products 0.005784 shows that they are somewhat neutral in its approach

### Q1 Card Issuers and its frequency



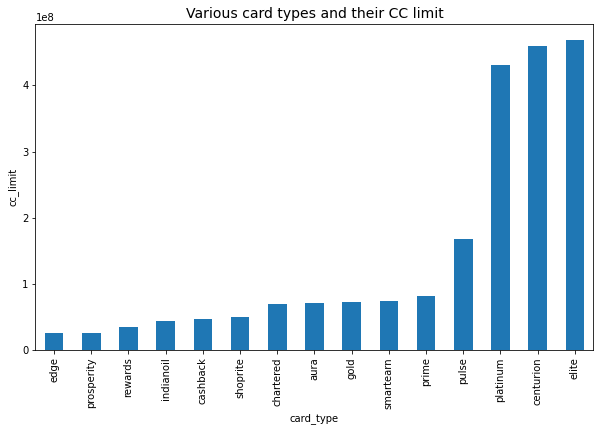
The pie chart presented here shows the maximum customers are holding the Visa type card followed by Mastercard and lastly Amex.

### Q2 Which card type has the highest frequency, plotting a graph in descending order

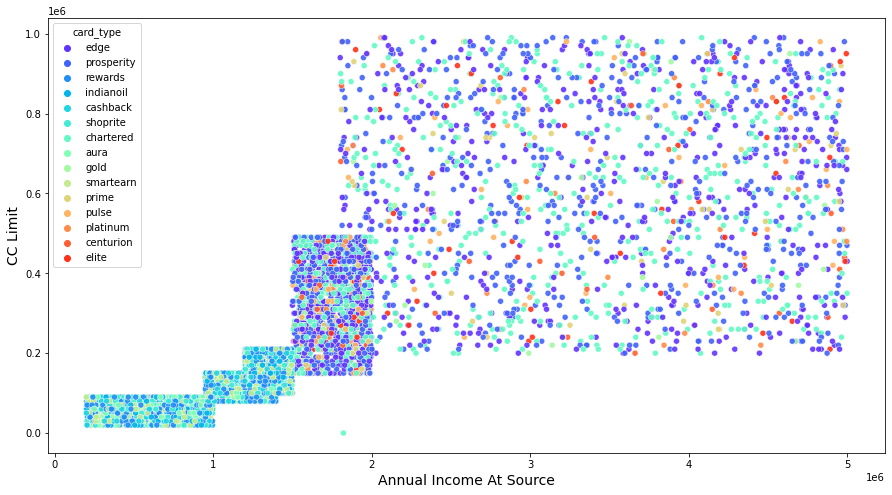
### *Almost every issuer is issuing all the cards except for GOLD & CENTURION by AMEX, PLATINUM by AMEX, ELITE and PULSE by MASTERCARD*

### **Card type and their credit card limit based on sum in ascending order**



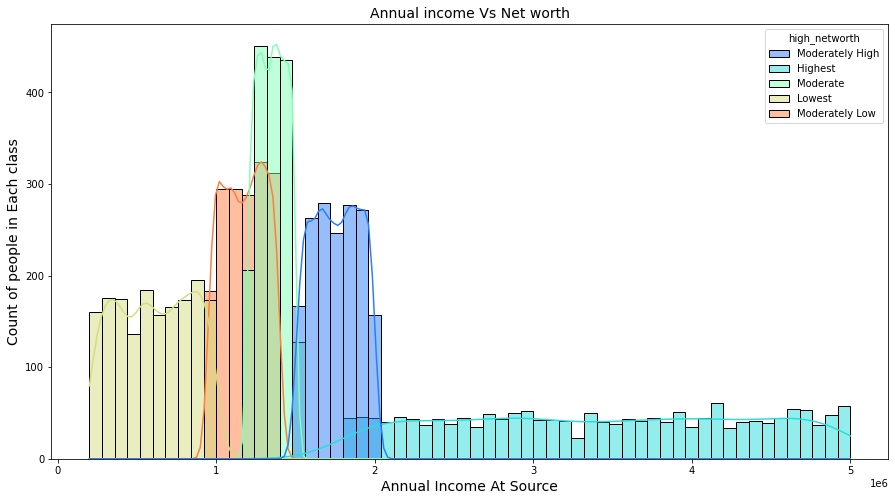
#### From these two graphs its evident that the cards like REWARS, EDGE, PROSPERITY whose cc limits are generally low are preferred by the maximum number of customers. As the cc limit increases the annual charges on the card increases and hence it can be afforded by only those whose annual income at source is really high or net worth is high. These cards include ELITE, CENTURION, PLATINUM. We will further plot graphs and try to relate it with net worth and annual income.

### Q3 Net Worth and its variation with the cc limit

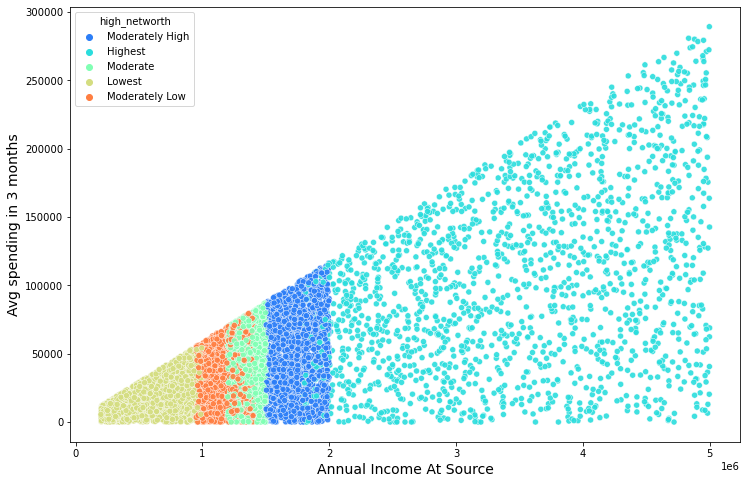
****

### the graph above shows that when the annual income increases the cc limit also increases and hence the annual charges on the card also increases. So, cards like Elite and Platinum are found in the upper range only

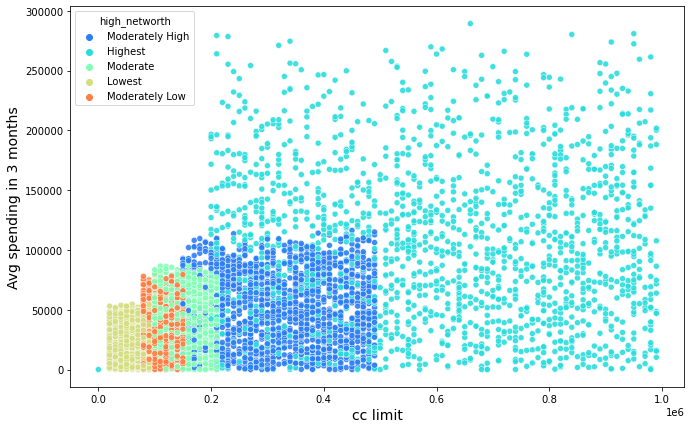
### Q3 Relationship between Avg spend limit in 3 months by the customer with respect to annual income and credit card limit

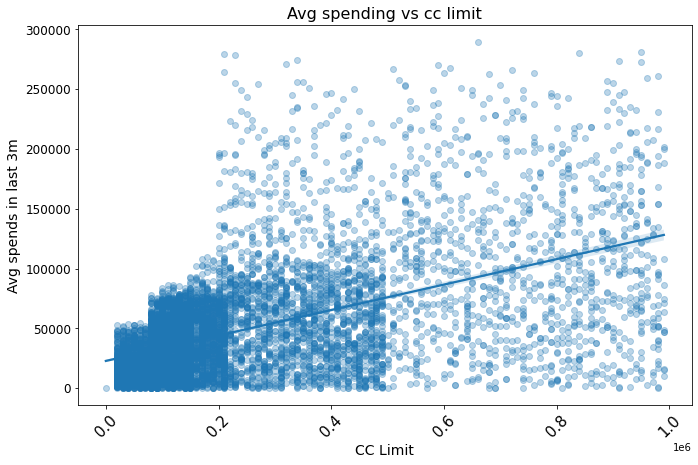


### It shows from the graphical representation that count of customers who are having the "highest net worth" is lowest and the customers who are in range of Moderate Net worth represents the maximum frequency. Now let’s analyze their cc limit and spending capacity.



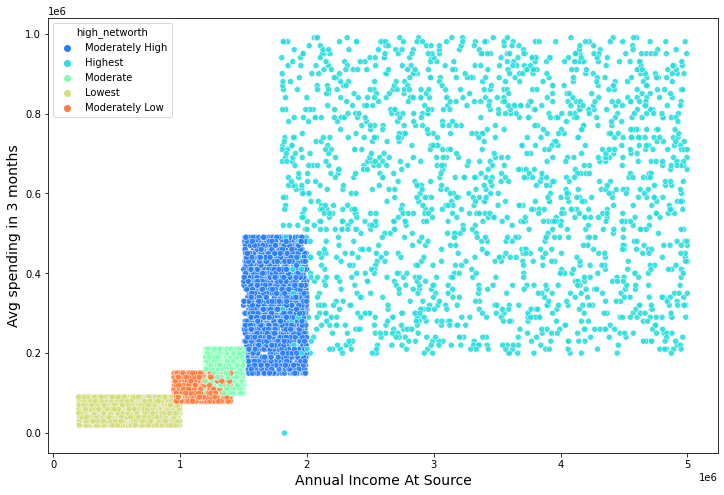
### Q4 Net worth vs cc limit and avg spending in last 3 months





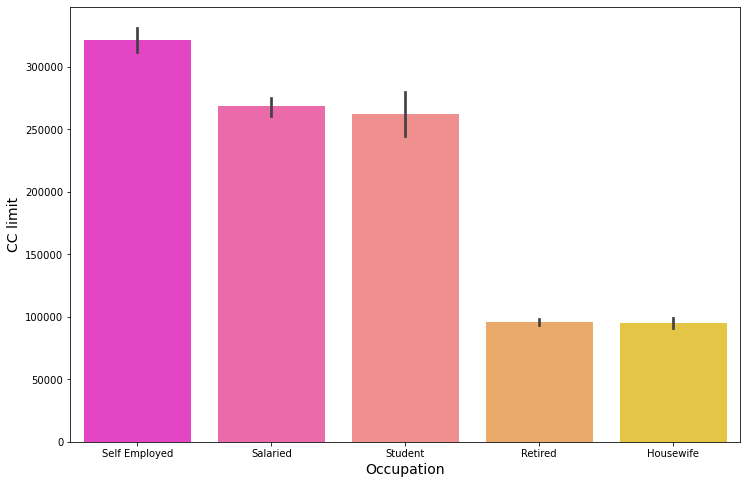
### So, we can say that although CC limit increases as the net worth of the customers increases but it’s not necessary that avg spending in last 3 months will definitely increase, although it shows an increasing trend but it’s not highly correlated, and the line tends to get constant in near future

### Q5 Does annual income and net worth together plays an important role for cc limit



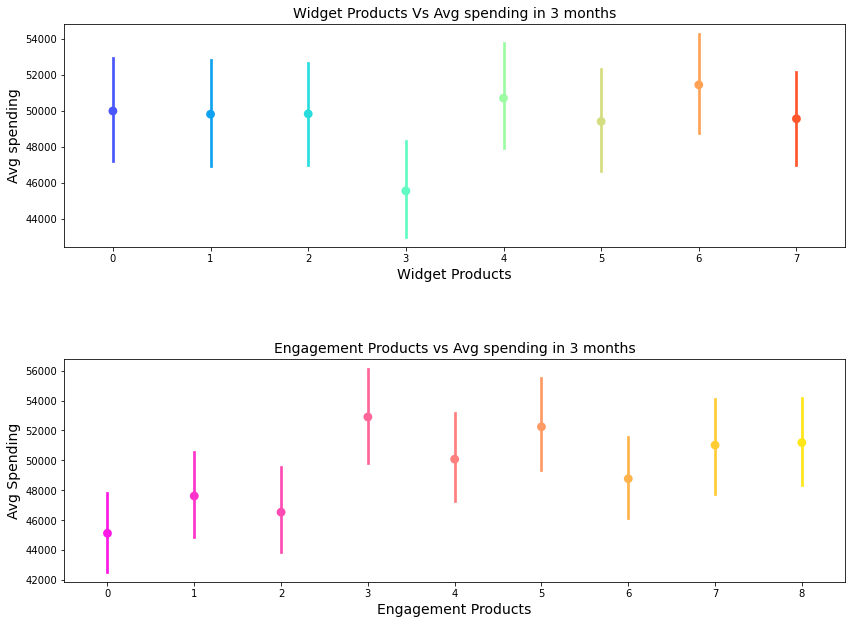
### Yes, these 3 are quite correlated to each other as we can see as the annual salary increases the net worth also increases along with the cc limit

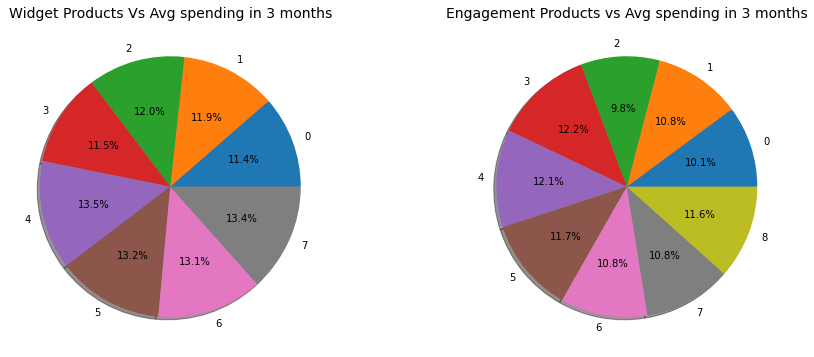
### Q6 Which occupation is the most suitable to increase the profit? occupation and cc limit along with avg spending



The occupation graph clearly indicates that business (that is self-employed) and salaried customers are the most favorable ones here in this case, and it decreases as the earning decreases, like in case of students then to retired customers and lastly to the housewives.

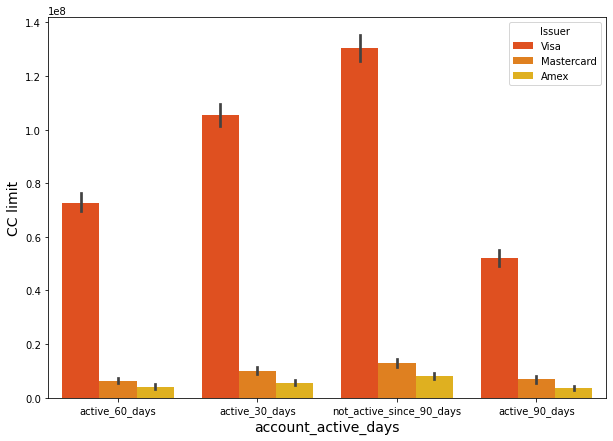
### Q7 What role does widget products & engagement products have to play for the credit card holders?

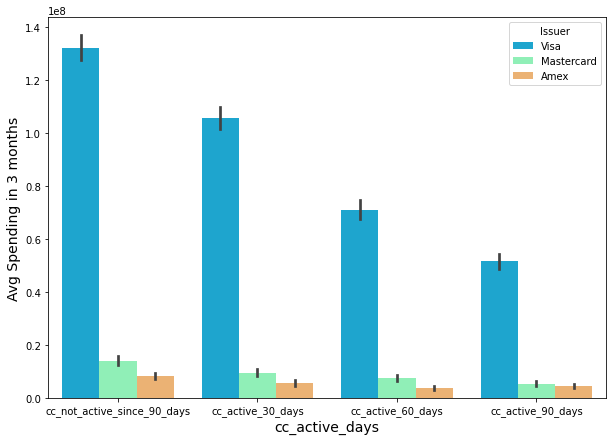




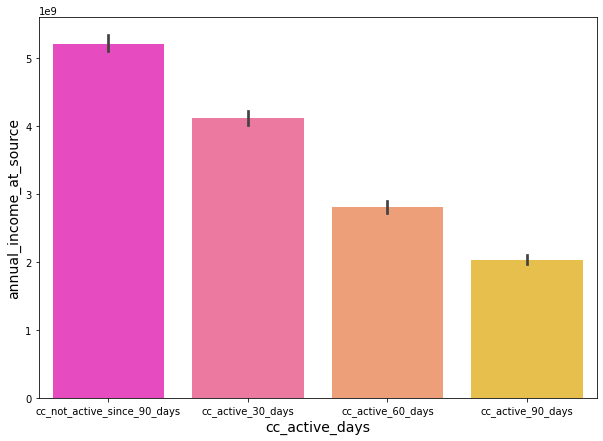
### We can interpret that here widget products and engagement products are not much related to the cc limit and avg spending. It’s mostly similar except for few exceptions and not having much impact on the credit card holders. Yes, in case of Engagement products, when it increases it somewhat curbs the spending capacity

### Q8 Track the record of Bank account activity and Credit card activity on quarterly basis (3 months)



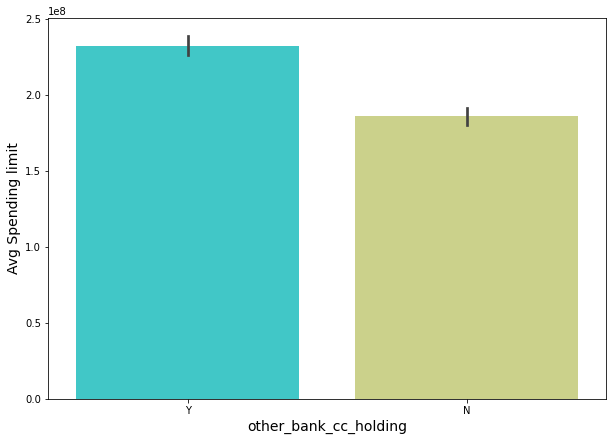


Average spending here has a decreasing trend with the days, and most of the credit card users spending shows they are not really active since last 90 days, followed by those who were active till last month and their spendings were higher to those who were active in last 60 days.



Although annual income is high still the non-active (in last 90 days) card holders are higher and the trend decreases as the active days’ increases.

### Q9 What is the impact of other bank cc holding on the cc limit and assurance?

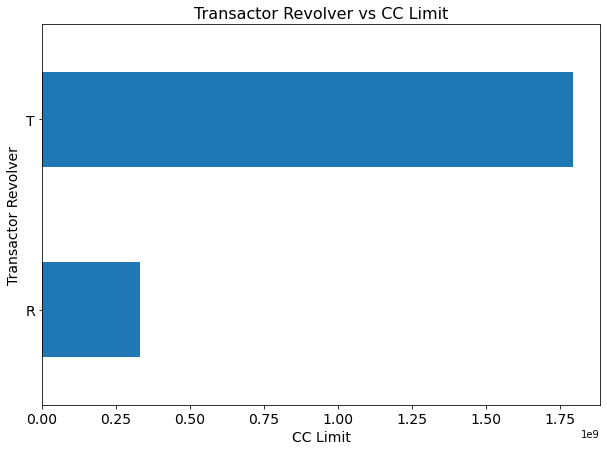
It shows that other card holdings actually is a good deal breaker for credit card companies as it shows the reliability of the customers in the market. It also helps them to gain the trust in some of the cases and in return increase in the credit card limit for the customers . if you are already a credit card holder it increases your chances to easily get another credit card in the desired limit.

### Q10 Bank vintage vs T+1\_month activity, plotting the graphs to know the variations



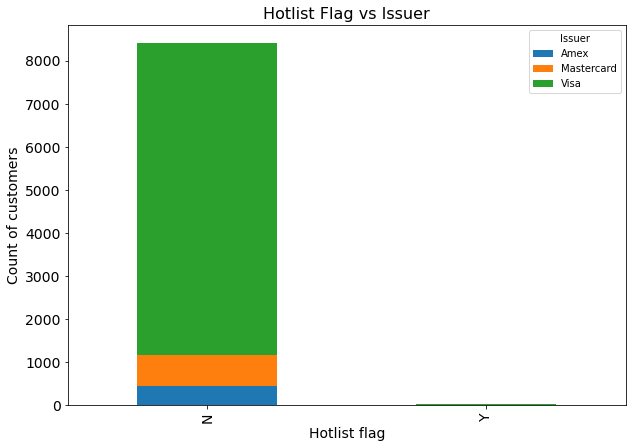
### This analysis shows that most of the people tend to spend in T+1\_month and slowly it decreases except for in case of T+3\_month which shows a little bit of variation.

#### Q11 Draw useful insights based on Transactor-revolver category and its relation to cc limit?



If the customer belongs to the Transactor category it automatically helps them to increase their credit card limit and hence increase the average spending limits. In case of Revolvers, those who pays EMIs in the next months their credit card limit is generally low.

### Q12 Bank vintage vs Hotlist Flag including the issuer of the cards



Its almost negligible here in case of the data set given to us as it shows, so it’s not very important when it comes to profit margin of the credit card company.

### So, the most required features for the company would be interested in are as follows-

**CONCLUSION-**

The above analysis done by me consists of various useful insights from the data set including descriptive analysis of each independent variable along with the necessary dependent variables, which can help the company to increase its customer base and gain popularity in the market.

\*\****References*** *– Team Great Learning Mentors*