# SMDM PROJECT (CODED)

SRINIVASAN T

## Content:

# List of Figures:

# List of Tables:

# Data Dictionaries for the problem statements

**Austo_automobile.csv**

Data Description

- age: The age of the individual in years.
- **gender**: The gender of the individual, categorised as male or female.
- **profession**: The occupation or profession of the individual.
- **marital_status**: The marital status of the individual, such as married &, single
- **education**: The educational qualification of the individual Graduate and Post Graduate
- **no_of_dependents**: The number of dependents (e.g., children, elderly parents) that the individual supports financially.
- **personal_loan**: A binary variable indicating whether the individual has taken a personal loan "Yes" or "No"
- **house_loan**: A binary variable indicating whether the individual has taken a housing loan "Yes" or "No"
- **partner_working**: A binary variable indicating whether the individual's partner is employed "Yes" or "No"
- **salary**: The individual's salary or income.
- **partner_salary**: The salary or income of the individual's partner, if applicable.
- **Total_salary**: The total combined salary of the individual and their partner (if applicable).
- **price**: The price of a product or service.
- **make**: The type of automobile

**Godigt_cc_data.xlsx**

Data Description

**userid** - Unique bank customer-id
**card_no** - Masked credit card number
**card_bin_no** - Credit card IIN number
**Issuer** - Card network issuer
**card_type** - Credit card type
**card_source_data** - Credit card sourcing date
**high_networth** - Customer category based on their net-worth value (A: High to E: Low)
**active_30** - Savings/Current/Salary etc. account activity in last 30 days
**active_60** - Savings/Current/Salary etc. account activity in last 60 days
**active_90** - Savings/Current/Salary etc. account activity in last 90 days
**cc_active30** - Credit Card activity in the last 30 days
**cc_active60** - Credit Card activity in the last 60 days
**cc_active90** - Credit Card activity in the last 90 days
**hotlist_flag** - Whether card is hot-listed(Any problem noted on the card)
**widget_products** - Number of convenience products customer holds (dc, cc, net-banking active, mobile banking active, wallet active, etc.)
**engagement_products** - Number of investment/loan products the customer holds (FD, RD, Personal loan, auto loan)
**annual_income_at_source** - Annual income recorded in the credit card application
**other_bank_cc_holding** - Whether the customer holds another bank credit card
**bank_vintage** - Vintage with the bank (in months) as on Tth month
**T+1_month_activity** - Whether customer uses credit card in T+1 month (future)
**T+2_month_activity** - Whether customer uses credit card in T+2 month (future)
**T+3_month_activity** - Whether customer uses credit card in T+3 month (future)
**T+6_month_activity** - Whether customer uses credit card in T+6 month (future)
**T+12_month_activity** - Whether customer uses credit card in T+12 month (future)
**Transactor_revolver** - Revolver: Customer who carries balances over from one month to the next. Transactor: Customer who pays off their balances in full every month.
**avg_spends_l3m** - Average credit card spends in last 3 months
**Occupation_at_source** - Occupation recorded at the time of credit card application
**cc_limit** - Current credit card limit

# 1. Problem1 - Data Overview

## 1.1 Importing the necessary libraries:

Necessary libraries are imported using import functions.

## 1.2 Loading the dataset:

The given dataset is loaded using the **pd.read_filetype()** function.

## 1.3 Check the types of the data:

- Checking whether the data is loaded properly with head() & tail() functions

- Checking the no. of rows and no. of columns using shape() function
    - The number of rows = 1581
    - The number of columns = 14

- Checking the data types of each columns using the info() function
    - There are 1 float, 5 int & 8 objects variables are present
    - Variable: Gender & Partner_salary has null values
    - All the columns variables having respective data types(like categorical column having object data type)

**Table 1**

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1581 entries, 0 to 1580
Data columns (total 14 columns):
 #   Column            Non-Null Count  Dtype
---  ------            --------------  -----
 0   Age               1581 non-null   int64
 1   Gender            1528 non-null   object
 2   Profession        1581 non-null   object
 3   Marital_status    1581 non-null   object
 4   Education         1581 non-null   object
 5   No_of_Dependents  1581 non-null   int64
 6   Personal_loan     1581 non-null   object
 7   House_loan        1581 non-null   object
 8   Partner_working   1581 non-null   object
 9   Salary            1581 non-null   int64
 10  Partner_salary    1475 non-null   float64
 11  Total_salary      1581 non-null   int64
 12  Price             1581 non-null   int64
 13  Make              1581 non-null   object
dtypes: float64(1), int64(5), object(8)
memory usage: 173.1+ KB
```

## 1.4 Check the statistical summary:

The statistical summary can be done using describe() function. This shows us the 5 point summary ( Q1, Q2, Q3, LL, UL), Mean and Standard Deviation.

**Table 2**

| | count | mean | std | min | 25% | 50% | 75% | max |
|---|---|---|---|---|---|---|---|---|
| **Age** | 1581.0 | 31.922201 | 8.425978 | 22.0 | 25.0 | 29.0 | 38.0 | 54.0 |
| **No_of_Dependents** | 1581.0 | 2.457938 | 0.943483 | 0.0 | 2.0 | 2.0 | 3.0 | 4.0 |
| **Salary** | 1581.0 | 60392.220114 | 14674.825044 | 30000.0 | 51900.0 | 59500.0 | 71800.0 | 99300.0 |
| **Partner_salary** | 1475.0 | 20225.559322 | 19573.149277 | 0.0 | 0.0 | 25600.0 | 38300.0 | 80500.0 |
| **Total_salary** | 1581.0 | 79625.996205 | 25545.857768 | 30000.0 | 60500.0 | 78000.0 | 95900.0 | 171000.0 |
| **Price** | 1581.0 | 35597.722960 | 13633.636545 | 18000.0 | 25000.0 | 31000.0 | 47000.0 | 70000.0 |

| | count | unique | top | freq |
|---|---|---|---|---|
| **Gender** | 1528 | 4 | Male | 1199 |
| **Profession** | 1581 | 2 | Salaried | 896 |
| **Marital_status** | 1581 | 2 | Married | 1443 |
| **Education** | 1581 | 2 | Post Graduate | 985 |
| **Personal_loan** | 1581 | 2 | Yes | 792 |
| **House_loan** | 1581 | 2 | No | 1054 |
| **Partner_working** | 1581 | 2 | Yes | 868 |
| **Make** | 1581 | 3 | Sedan | 702 |

**Table 3**

The statistical summary of categorical variables shows the unique values, mode and frequency of each categorical column. The Gender variable has 4 unique values.

## 1.5 Data Preprocessing

### 1.5.1 Check for and treat bad data:

The Gender categorical variable has 4 unique values. The list has misspell words for the gender female.

```
array(['Male', 'Femal', 'Female', nan, 'Femle'], dtype=object)
```

The list has misspelt words for the gender female.

```
array(['Male', 'Female', nan], dtype=object)
```

The misspelt words are replaced with the actual one.

### 1.5.2 Check for and treat anomalies:

No anomalies are present/detected.

### 1.5.3 Check for and treat missing values

**Table 4**

| | |
|---|---|
| Age | 0 |
| Gender | 53 |
| Profession | 0 |
| Marital_status | 0 |
| Education | 0 |
| No_of_Dependents | 0 |
| Personal_loan | 0 |
| House_loan | 0 |
| Partner_working | 0 |
| Salary | 0 |
| Partner_salary | 106 |
| Total_salary | 0 |
| Price | 0 |
| Make | 0 |
| dtype: int64 | |

- The data set has missing values/NaN in two variables Gender and Partner salary
- The NaN values in Gender is filled with the mode value which is Male(which is 3.3%)
- The NaN value in Partner salary is filled with 0 if the partner working is ‘No’ and with the median of Partner salary if the partner working is ‘Yes’. Logically the Partner salary need to be 0 if they don’t work and can’t be filled with median for all

- Dataset after treating the null values.

**Table 5**

| | |
|---|---|
| Age | 0 |
| Gender | 0 |
| Profession | 0 |
| Marital_status | 0 |
| Education | 0 |
| No_of_Dependents | 0 |
| Personal_loan | 0 |
| House_loan | 0 |
| Partner_working | 0 |
| Salary | 0 |
| Partner_salary | 0 |
| Total_salary | 0 |
| Price | 0 |
| Make | 0 |
| dtype: int64 | |

### 1.5.4 Check for and treat duplicates:

The No. of duplicate rows in the dataset = 0

There are no duplicate rows in the dataset.

## 1.6 Observations and Insights

- The data set is loaded properly
- The data set has NaN values
- The data set has misspell objects/bad values in a column
- There are 1 float, 5 int & 8 objects variables columns
- There are NaN values in two columns
- There are no duplicate values and no anomalies
- All the columns represent the respective data types
- The statistical summary give the 5 point summary
- All the mean values are slightly higher than the 50% (median) values and we can infer that this is right skewed
- The statistical summary of categorical variables shows mode and top variable accrued.
- 1054 customers out of 1581 do not have a house loan.
- 702 customers out of 1581 prefer sedan over other car makers.

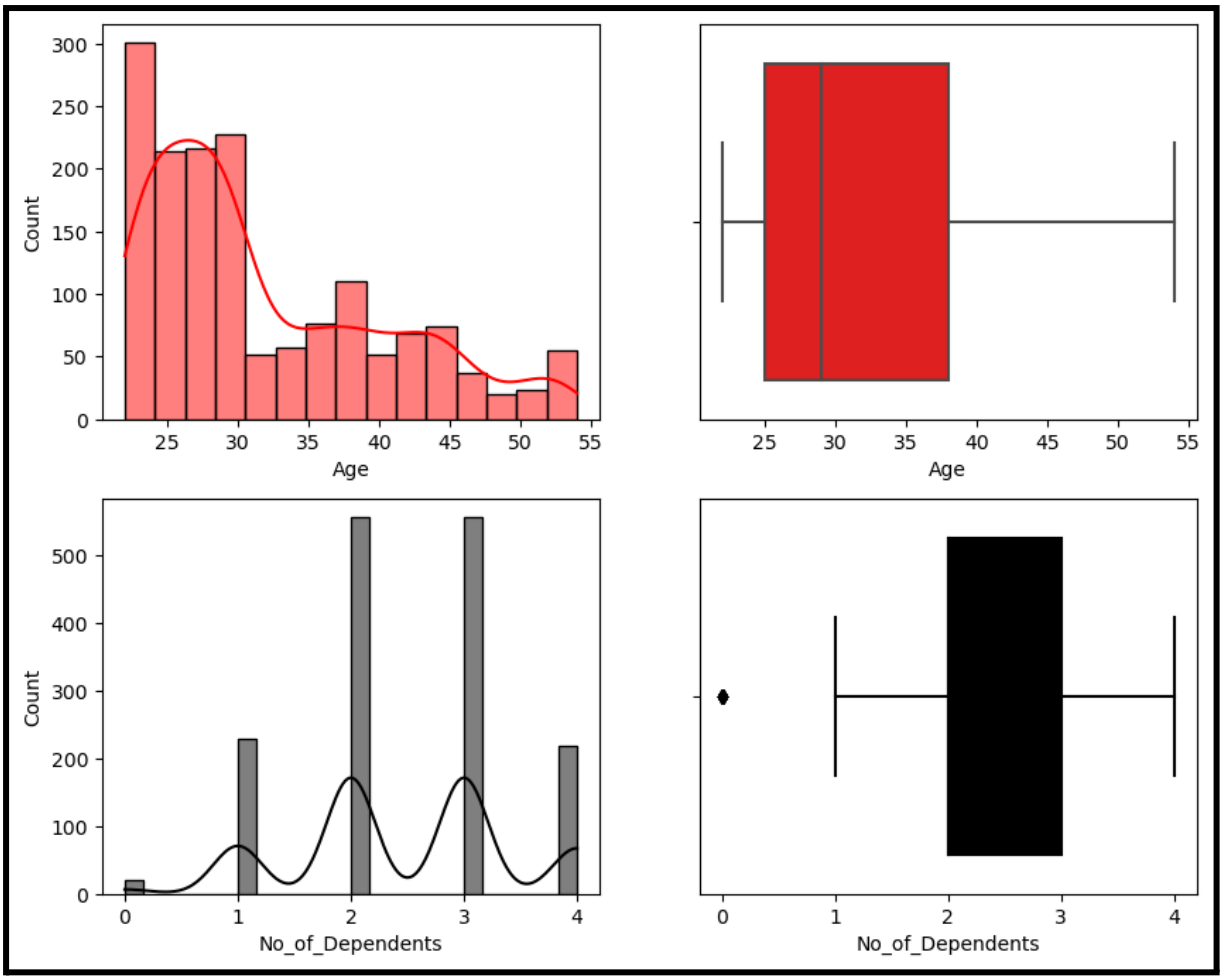
# 2. Problem1 - Data Visualization

## 2.1 Univariate Analysis

### 2.1.1 Numerical Variable Analysis

**<u>Variables:</u> Age and No. of Dependents**

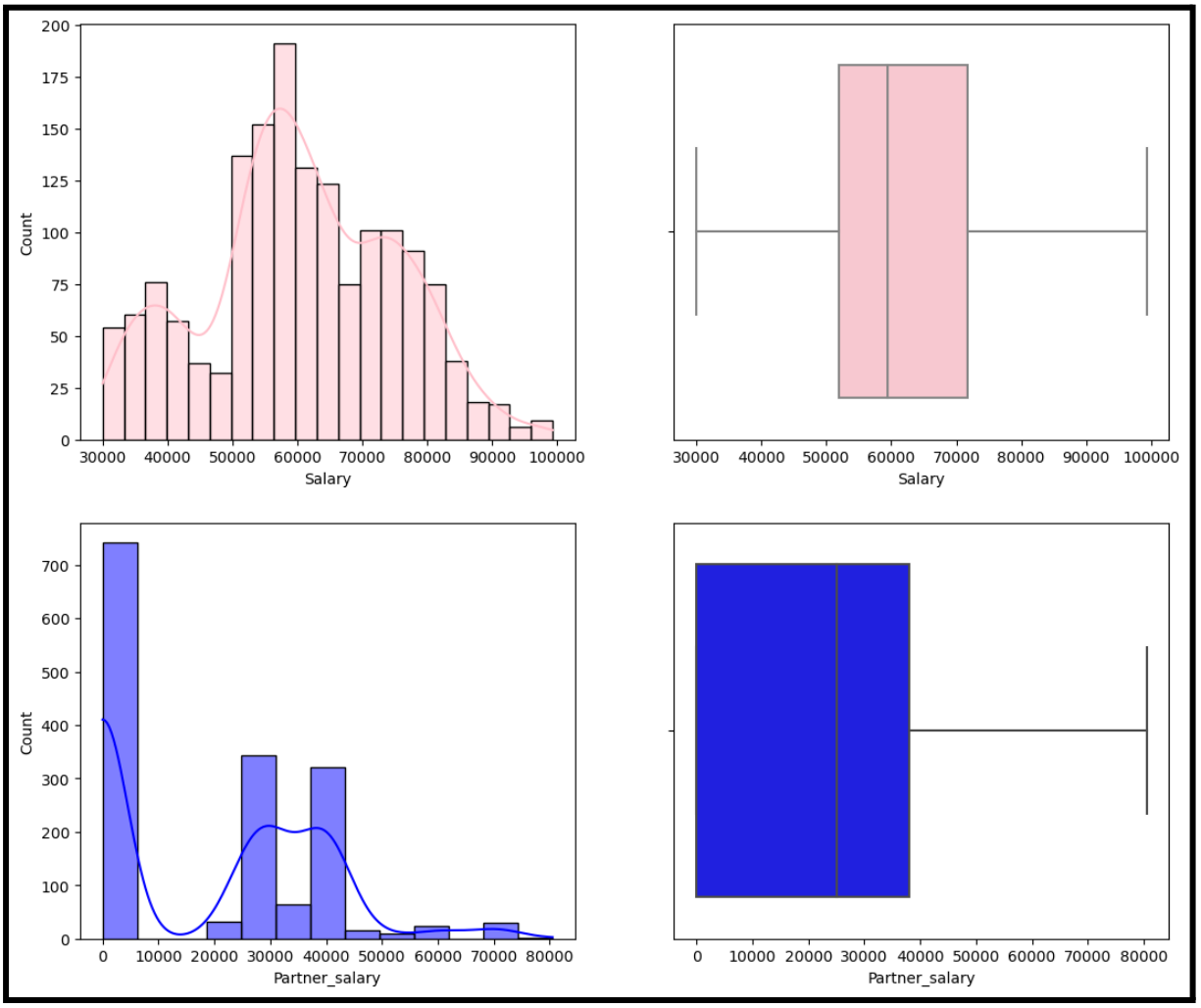
**Fig: 1.1**




- Age variable is right skewed and does not have outliers. 75% of the customers age lies between 25 - 38
- No_of_Dependents variable having more than one normal distribution and has outlier in lower limit

**<u>Variables:</u> Salary and Partner Salary**

**Fig: 1.2**

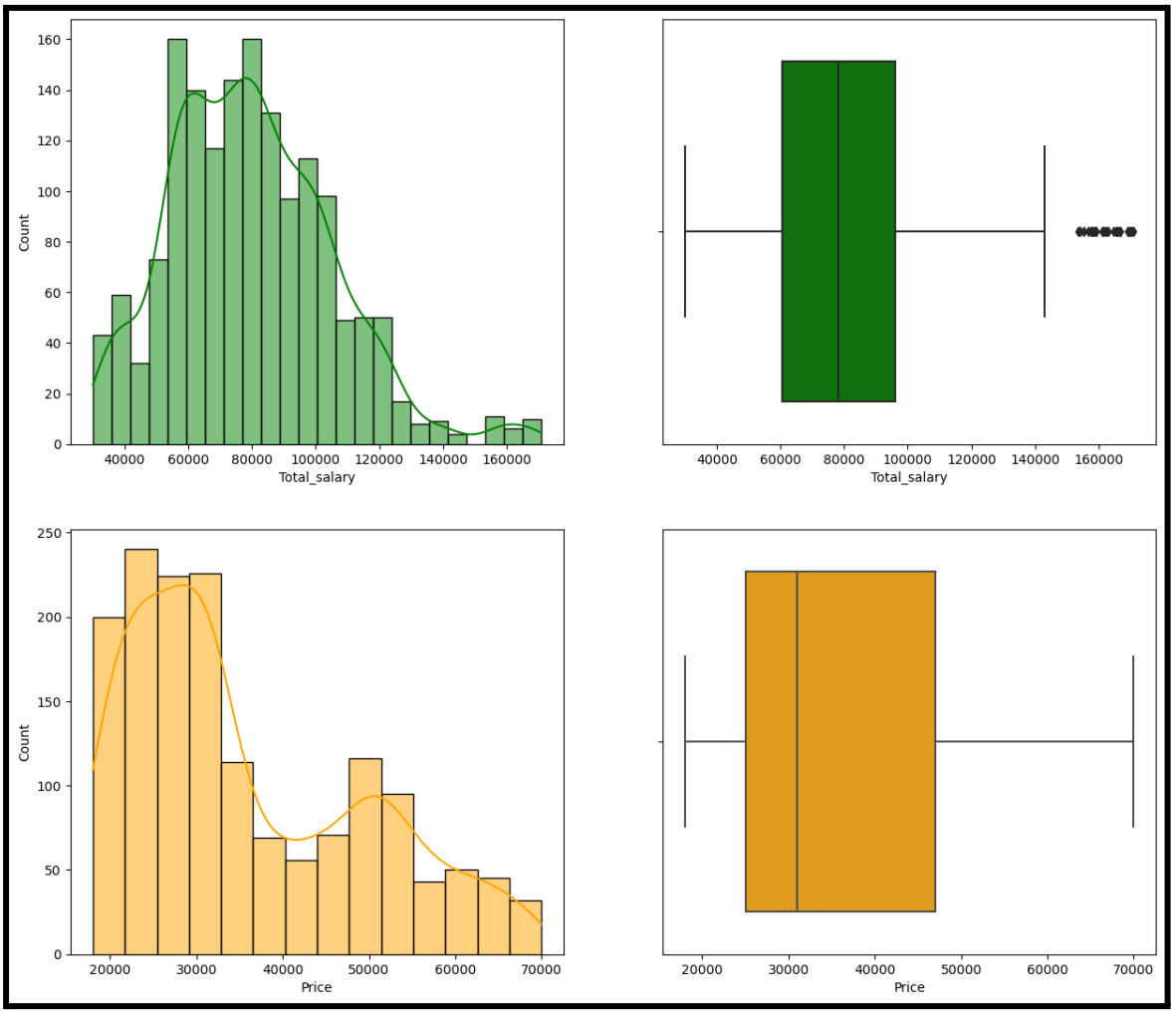



- Salary variable is right skewed and does not have outliers. The median salary is 60K.
- Partner_salary variable is left skewed, UL & Q1 = 0 (as there is no Whisker) and does not have outliers.

## **Variables:** Total Salary and Price

- Total_salary variable is almost a normal distribution or slightly right skewed and has outliers in the upper limit.
- Price variable is right skewed and does not have outliers. 50% car prices lie above 32K~.

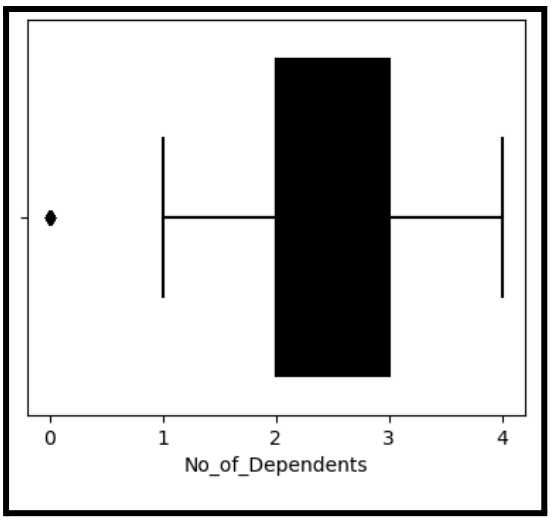
**Fig: 1.3**




## 2.1.2 Outlier Treatment

Before:

**Fig: 1.4**




After:

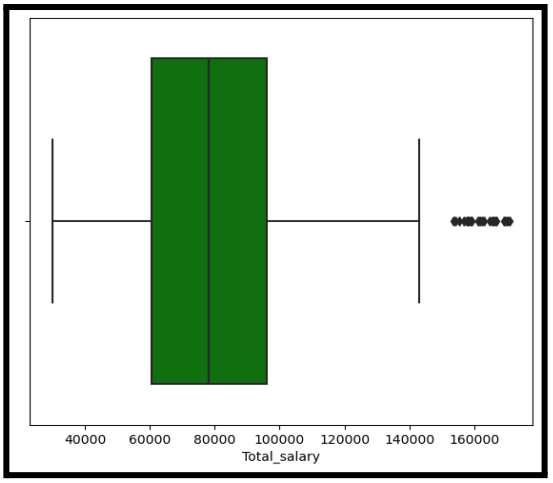
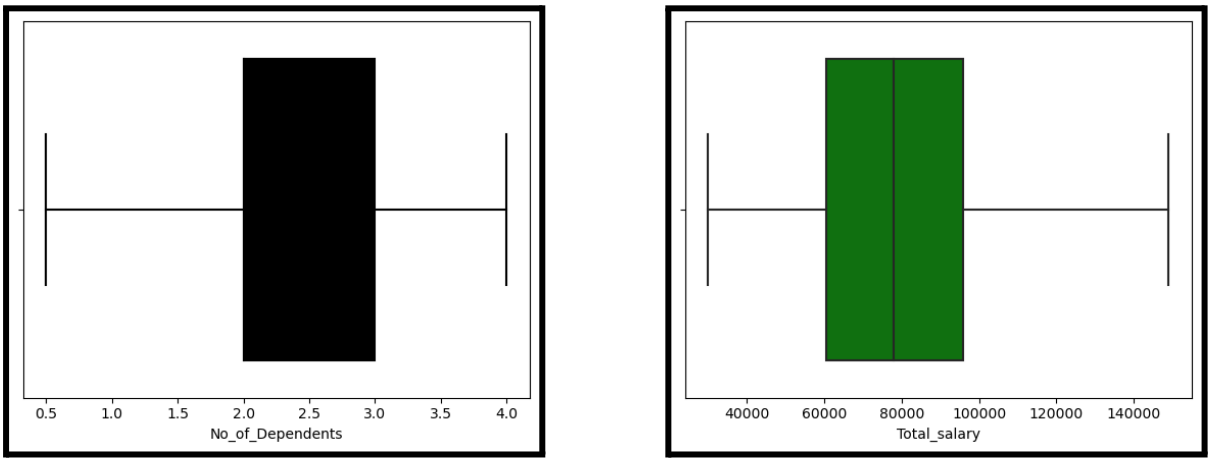


**Fig: 1.5**

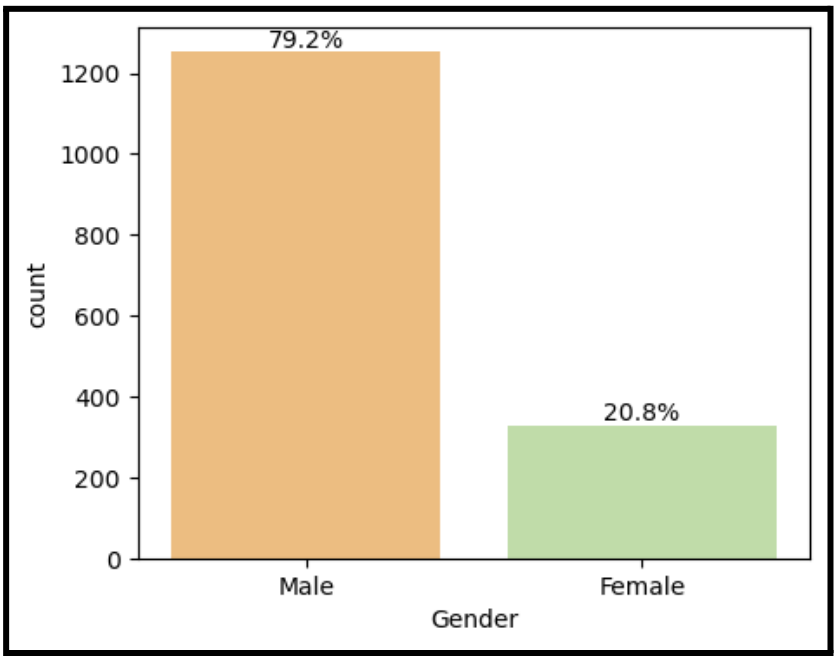



The outliers are treated and capped to the lower and upper limit of the boxplot.

### 2.1.3 Categorical Variable Analysis

**Gender:** 79.2% of the customers are male compared to overall customer base and an easier target for sales.

**Fig: 1.6**



**Profession:** 56.7% of the customers are salaried while 43.3% are business people. Targeting based on profession has somewhat equal chance or profession does not play a major role in selling cars.

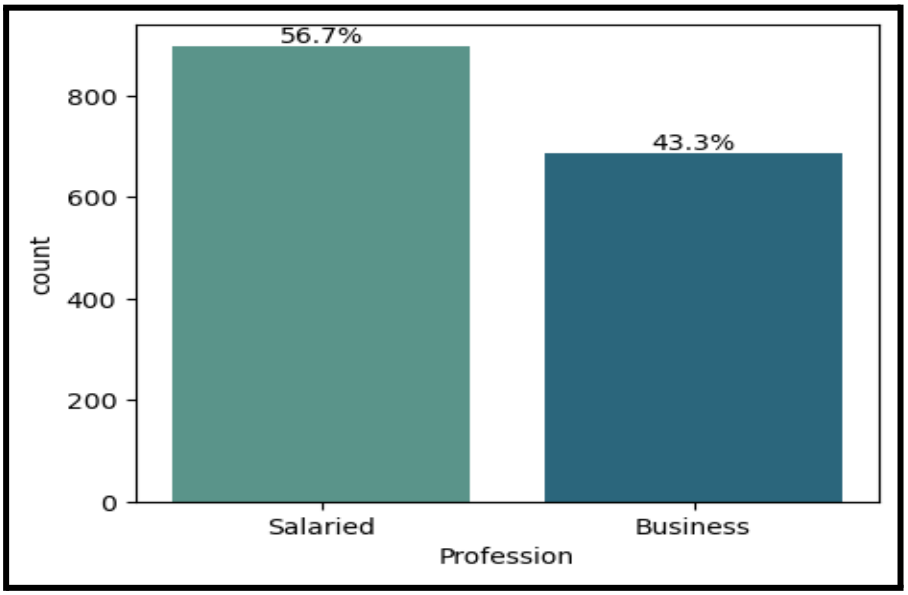



**Fig: 1.7**

**Marital Status:** 91.3% of the customers are married. Probability of selling a car to a married person is higher.

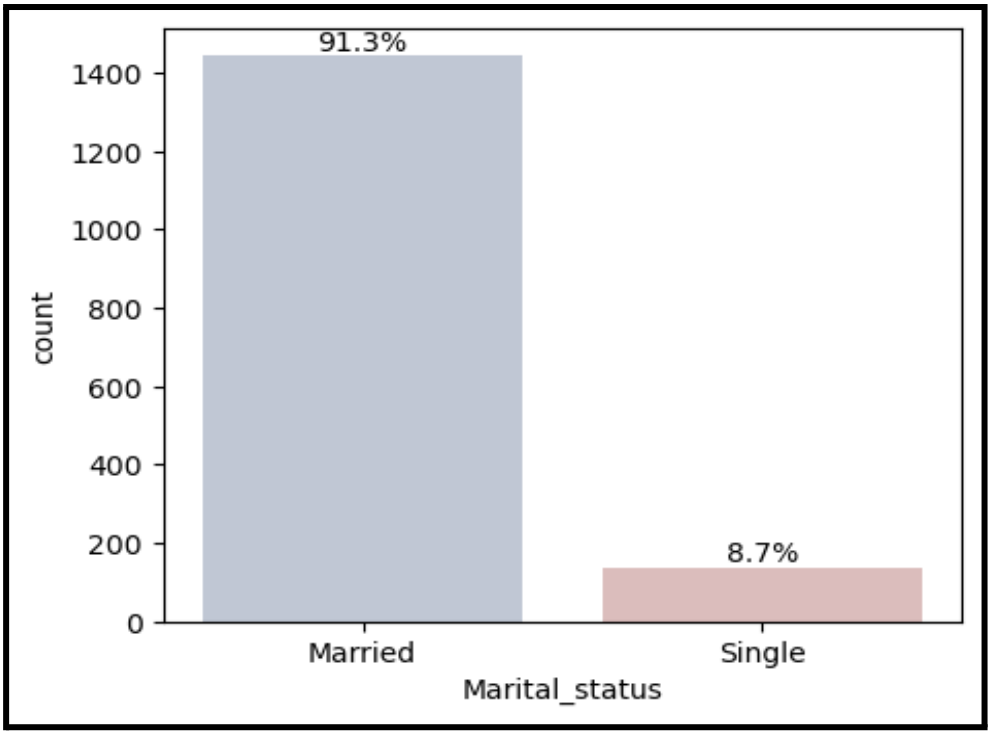



**Fig: 1.8**

**Education:** The customer who buys car are 62.3% postgraduate completed persons

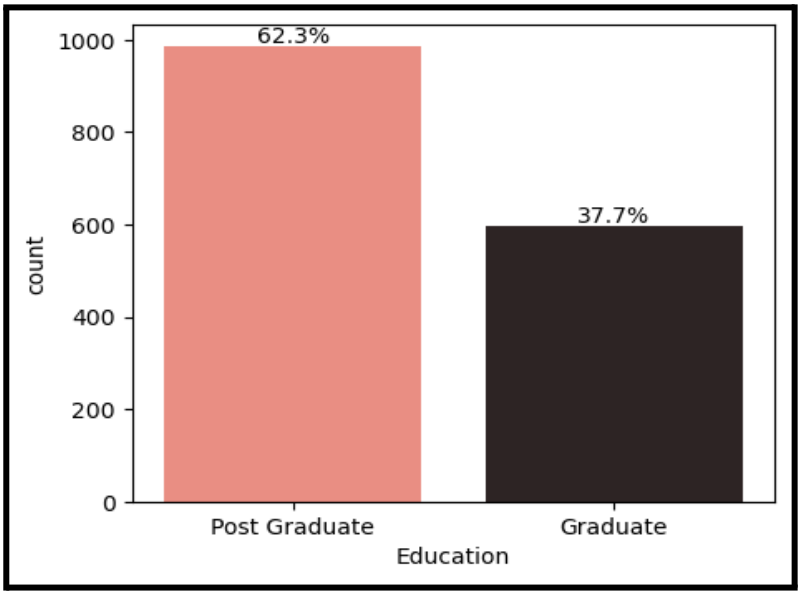



**Fig: 1.9**

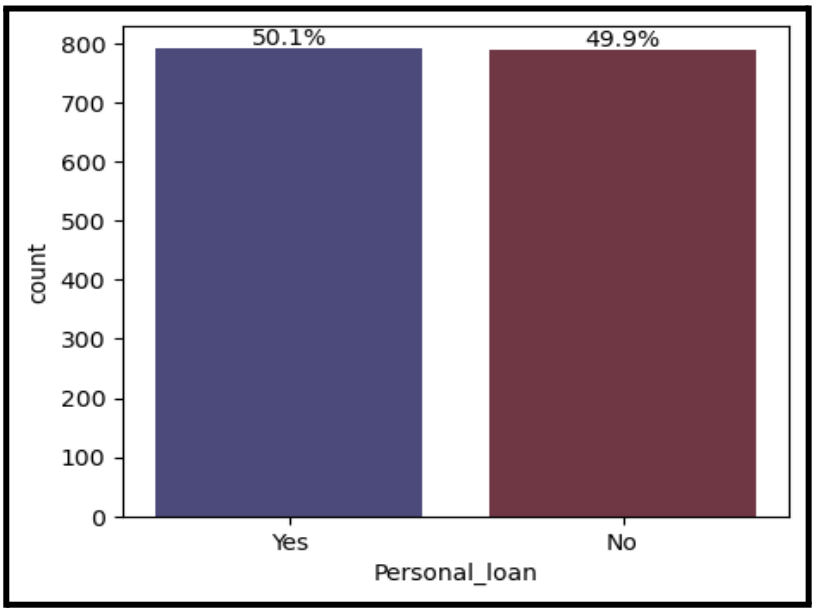



**Fig: 2.0**

**Personal Loan:** Half of the customer base has a personal loan. So a customer with a personal loan as a liability is not affecting the car buying decision.

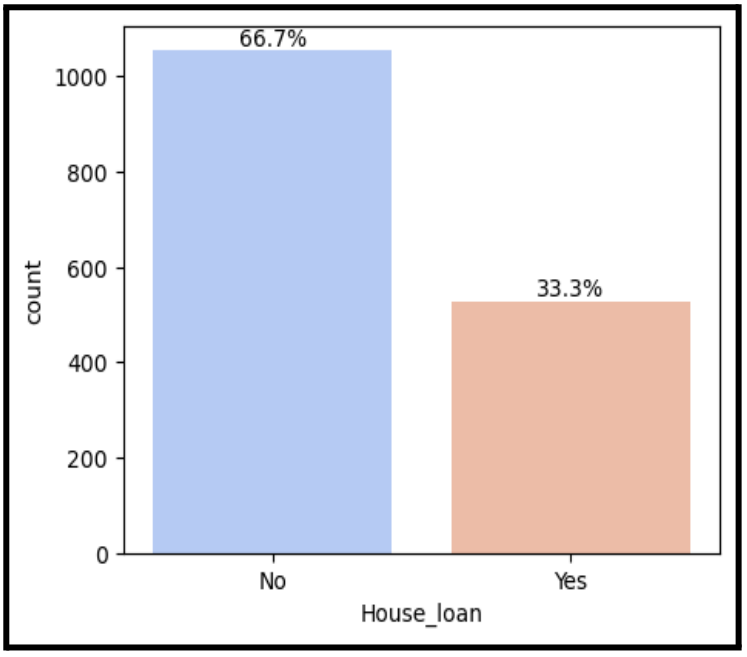
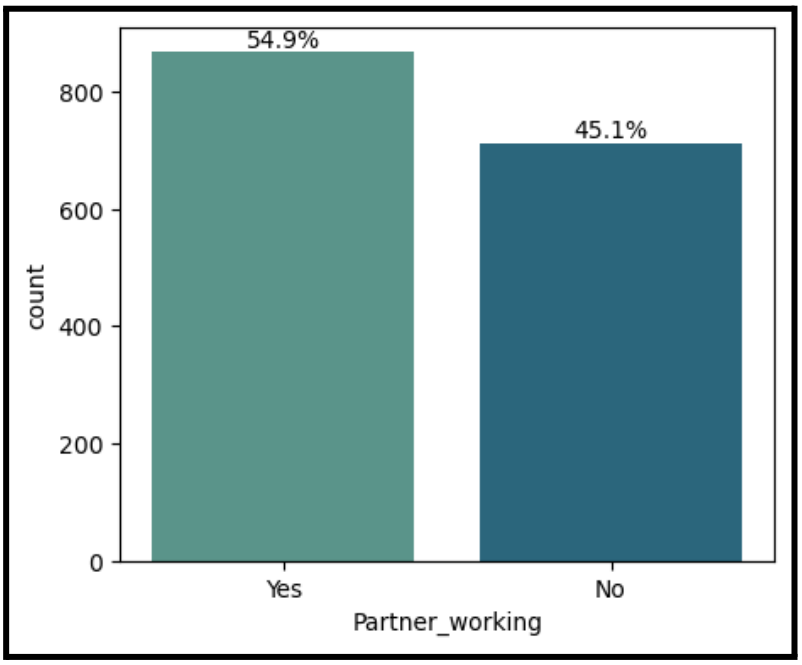



**Fig: 2.1**

**House Loan:** 66.7% of the customer does have a house loan. People without a house loan tend to buy cars as there is no liability.

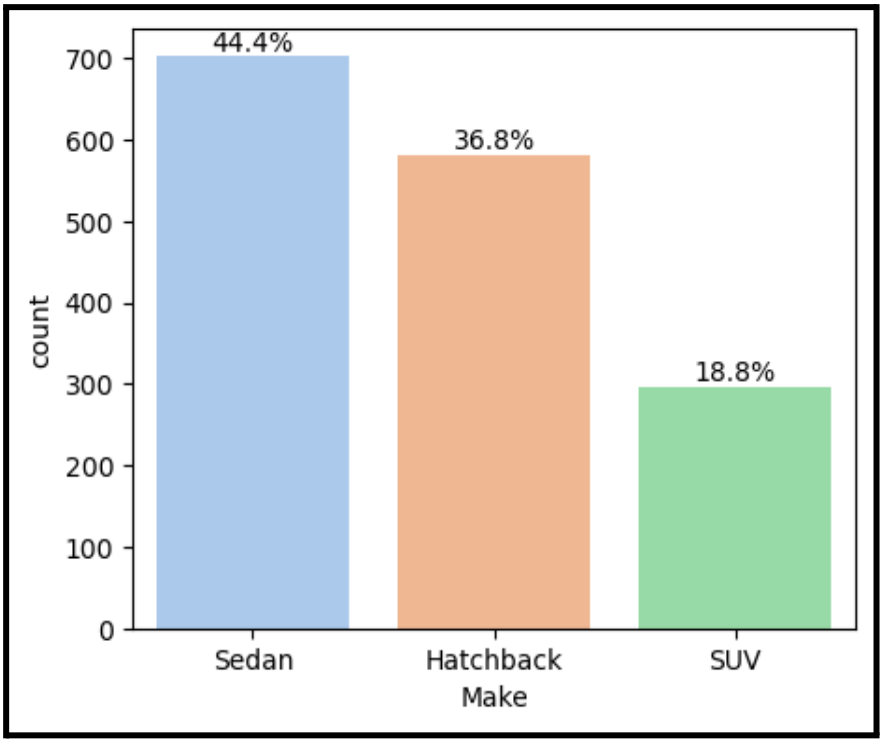
**Partner Working:** 54.9% of the customers' partners are working. There is a 5%~ chance of customers buying cars whose partner is working.

**Fig: 2.2**




**Make:** Sedan is the highest contributor with 44.4% and shows the high probability of people buying sedan car types. Next the Hatchback stands with 36.8%. So targeting customers for these two make types will 81.2% buying probability.

**Fig: 2.3**



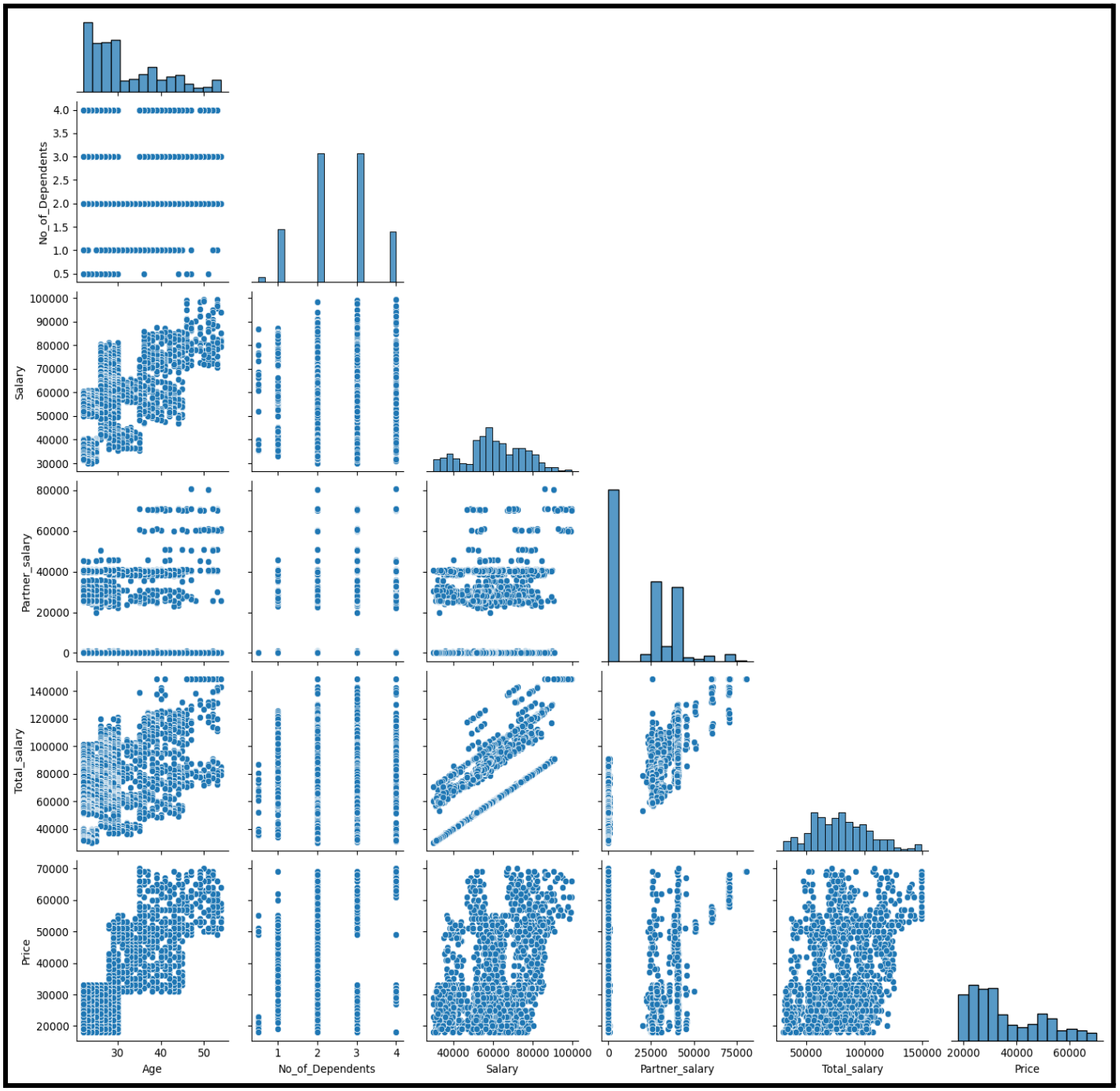
### 2.1.4 Observations and insight

- Age variable is right skewed and does not have outliers. 75% of the customers age lies between 25 – 38.
- No_of_Dependents variable having more than one normal distribution and has outlier in lower limit
- Salary variable is right skewed and does not have outliers. The median salary is 60K.
- Partner_salary variable is left skewed, UL & Q1 = 0 (as there is no Whisker) and does not have outliers.
- Total_salary variable is almost a normal distribution or slightly right skewed and has outliers in the upper limit.
- Price variable is right skewed and does not have outliers. 50% car prices lies above 32K~
- 79.2% of the customers are male compared to overall customer base and an easier target for sales.
- 56.7% of the customers are salaried while 43.3% are business people. Targeting based on profession has somewhat equal chance or profession does not play a major role in selling cars.
- 91.3% of the customers are married. Probability of selling a car to married person is higher.
- The customers who buy cars are 62.3% postgraduate completed persons.
- Half of the customer base has a personal loan. So, a customer with a personal loan as a liability is not affecting the car buying decision.
- 66.7% of the customer does have a house loan. People without a house loan tend to buy cars as there is no liability.
- 54.9% of the customers' partners are working. There is a 5%~ chance of customers buying cars whose partner is working.
- Sedan is the highest contributor with 44.4% and shows the high probability of people buying sedan car types. Next the Hatchback stands with 36.8%. So targeting customers for these two make types will 81.2% buying probability.

## 2.2 Bivariate Analysis

### 2.2.1 Numerical vs Numerical Variable Analysis

**Fig: 2.4**

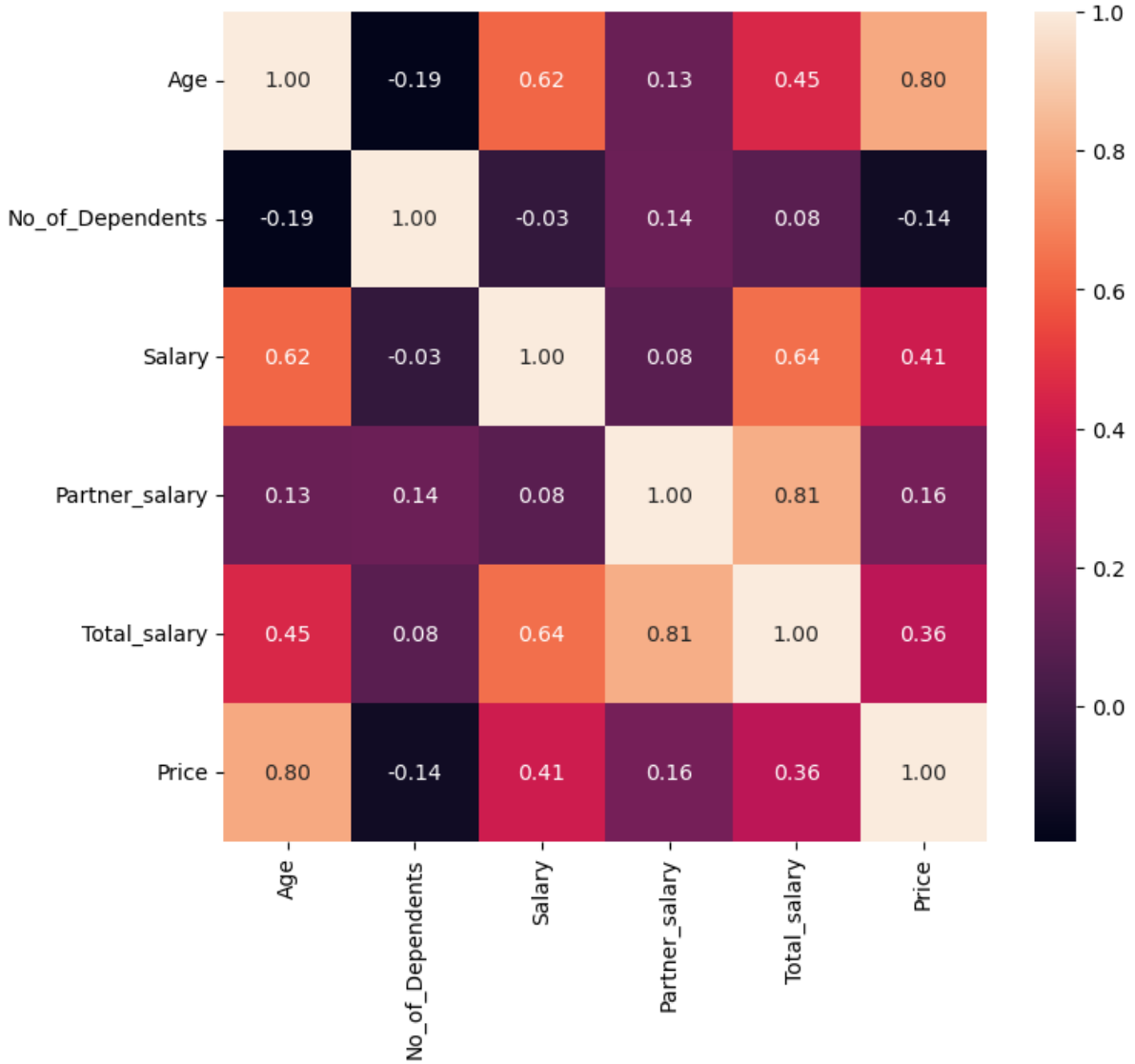



- Age vs Price, Total Salary, Partner Salary & Salary have Positive correlation.
- Salary & Total Salary have Positive correlation.

- Total Salary & Partner Salary have Positive correlation. When partner salary is added the total salary increases.
- Age vs Price: Whenever the age increases the value of the car price also increases.
- Age vs Salary: The increase in age (experience)  the salary increases.

### 2.2.2 Correlation between all numerical variables

**Fig: 2.5**

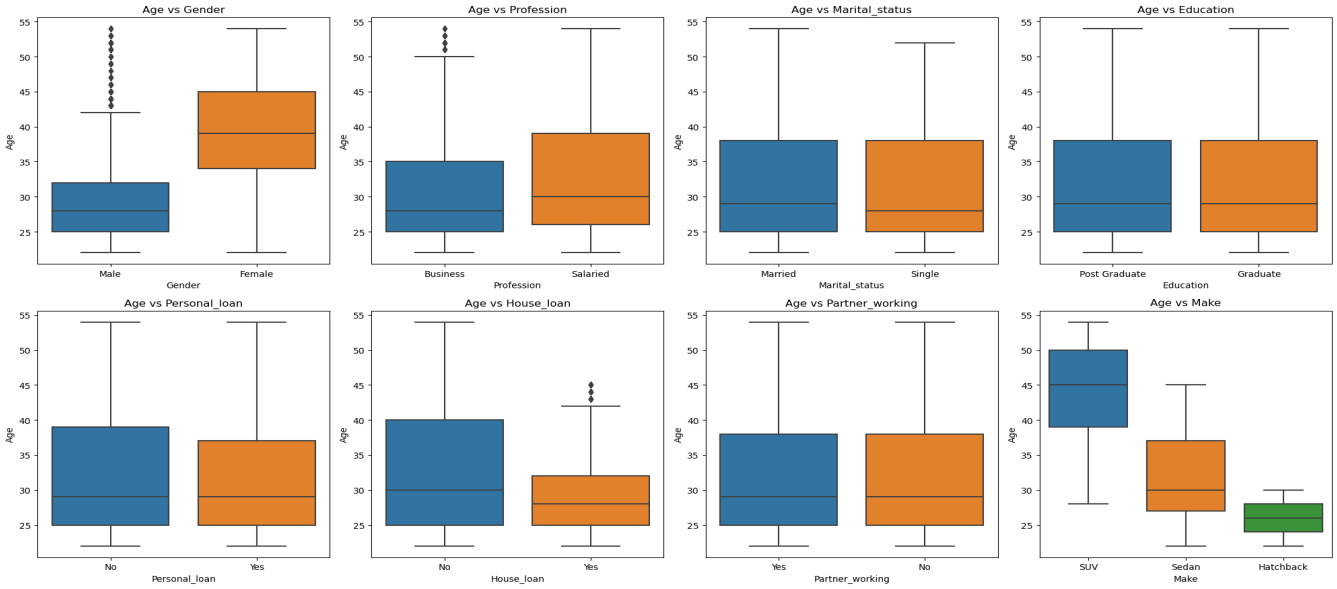



- Price & Age, Total salary & Partner salary has a high correlation. Targeting higher age customers, customers with higher total salary(above median/mean) and customers with partner salary having high probability of getting high value cars.

- Total salary & salary is having a good correlation (targeting a partner working per for SUV's)
- No. of dependents does not have correlation with any other variables

### 2.2.3 Numerical vs Categorical Variable Analysis

**Fig: 2.6**




Age vs Gender: The median age of females is almost double the male. The male has outliers and it is right skewed.

Age vs Profession: The salaried people's age are right skewed. Business median age is around 27 and have outliers of 50+ age.

Age vs Marital Status: The distribution of marital status age looks similar and it is right skewed.

Age vs Education: The distribution of education status age looks similar and it is right skewed.

Age vs Personal Loan: The distribution of personal loans of age is normal and it is right skewed.

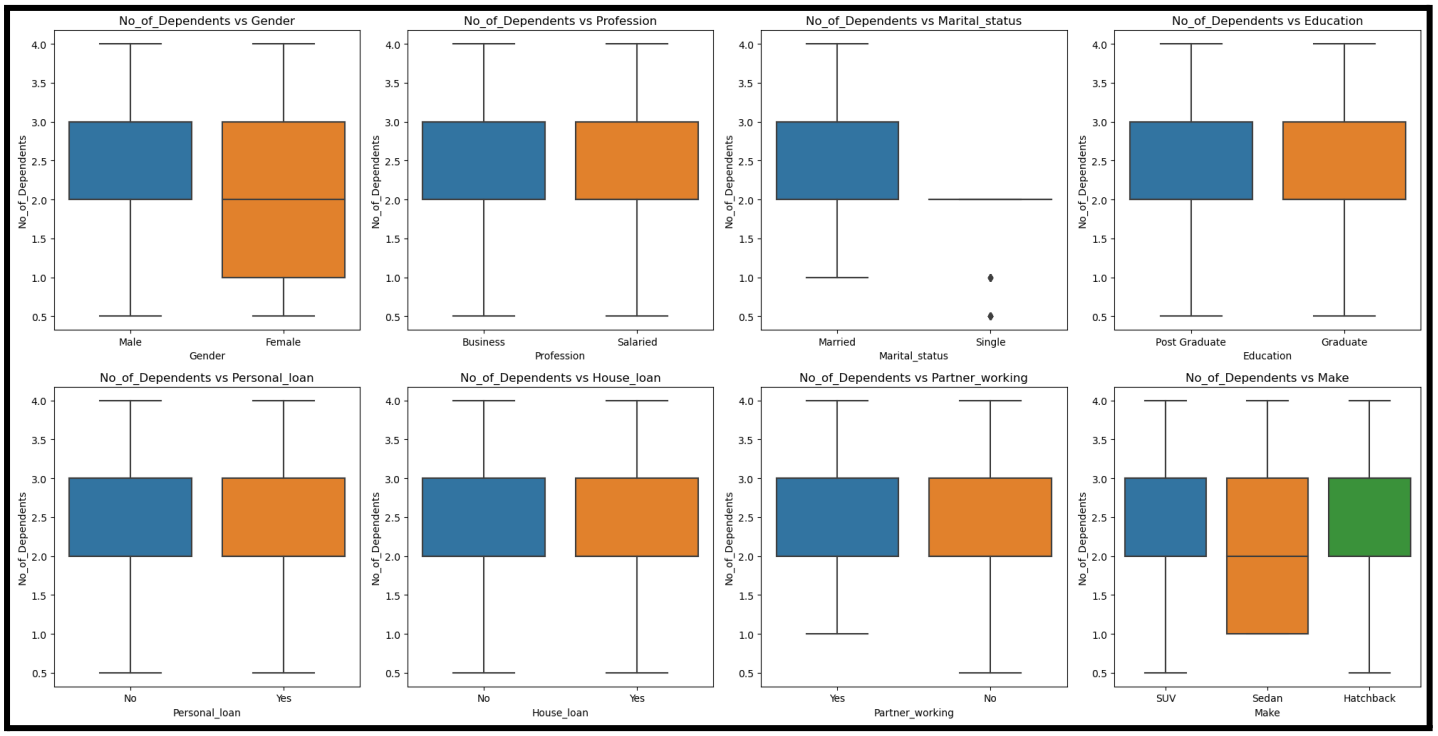
Age vs House Loan: People with house loan have their median age around 25-30 with a max of 40-45.

Age vs Partner Working: The distribution of partner working age looks similar and it is right skewed.

Age vs Make: 75% owning a SUV are above 40 age, Hatchback are most preferred by less 30 age people and 75% people with age less than 37 prefer Sedan

No. of Dependents vs Gender: The depth of females is higher than the male.

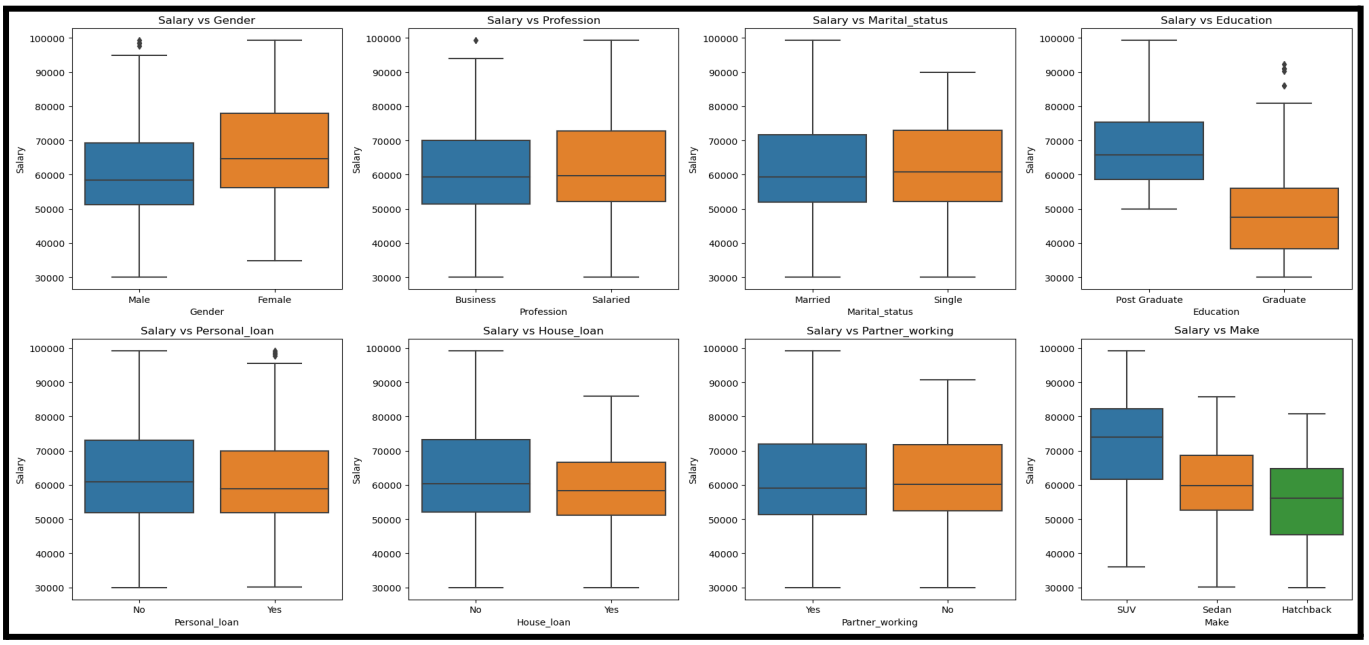
**Fig: 2.7**




No. of Dependents vs Marital Status: The distribution of marital status single looks as it does not have values.

No. of Dependents vs Make: Sedan cars are owned by customers whose median is 2.

**Fig: 2.8**




Salary vs Gender: The median salary of females is higher than male. The male has outliers and both male & female are right skewed.

Salary vs Profession: The depth of salaried customers are higher. The median salary looks similar. Business customers having outliers in salary.

Salary vs Marital Status: The median salary of single customers are slightly higher than married but the highest/maximum salary is from married customer

Salary vs Education: The 25% of post graduates is higher than the 75% of the graduate customers. Both the distributions with respect to salary are right skewed and graduate customer have outliers

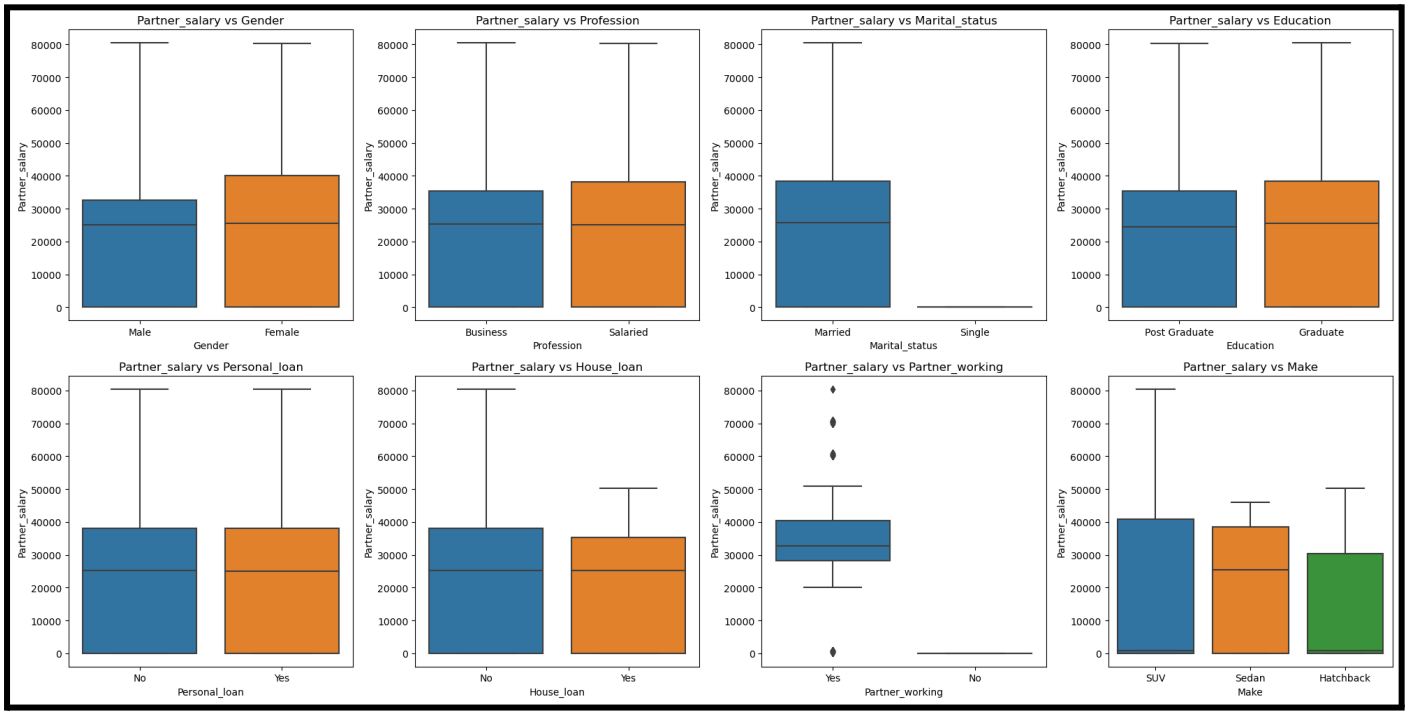
Salary vs Personal Loan: The customers who have not opted for personal loan depth are higher. The median salary is lesser for customers who opted for personal loans with outliers.

Salary vs House Loan: The customers who have not opted for house loan depth are higher. The median salary is lesser for customers who opted for house loans.

Salary vs Partner Working: The customers having partner working having maximum salary

Salary vs Make: Median salary of SUV customers > Sedan > Hatchback

**Fig: 2.9**




Partner Salary vs Gender: The partner salary of female customers are higher with depth and both are right skewed.

Partner Salary vs Profession & Education: The profession and education looks similar distribution pattern.

Partner Salary vs Marital Status: Obviously single status customers does not have partner salary.

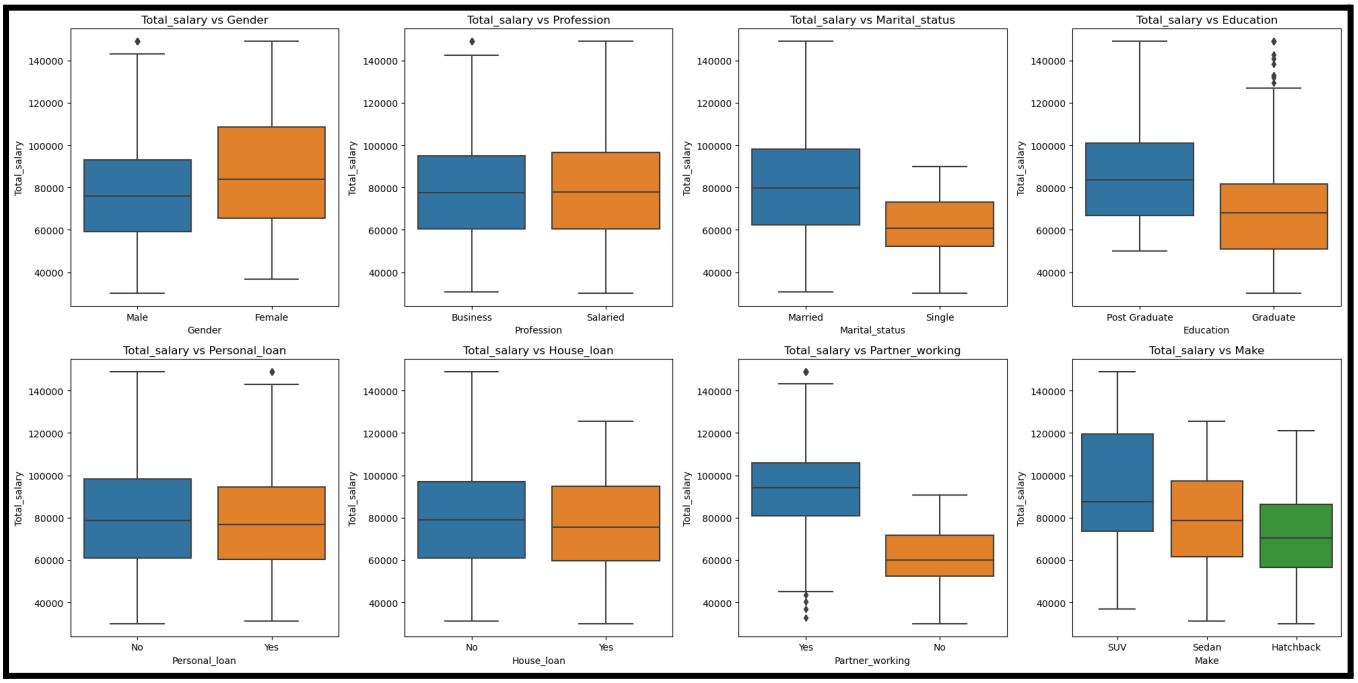
Partner Salary vs Personal Loan: The distribution looks similar for personal loan.

Partner Salary vs House Loan: The distribution looks similar for personal loan with the partner salary is maximum for customers who have not opted for house loan.

Partner Salary vs Partner Working: The customers do not have a partner salary who do not have a partner . The media partner salary varies between 30k-40k for whose having partner and has outliers below and above the minimum and maximum values.

Partner Salary vs Make: Higher customers with partner salary higher the purchase of SUV then Sedan and Hatchback.

**Fig: 3.0**




Total Salary vs Gender: The total salary of female customers is always higher since most of the male partners are working. Both distributions are right skewed.

Total Salary vs Profession: The median value look similar for business(with outliers) and salaried

Total Salary vs Marital Status: The married people have higher median total salary as that the single status customer does not have a partner to add up the total salary

Total Salary vs Education: The total salary of post graduate is higher with higher median and maximum. The graduate has outliers in total salary

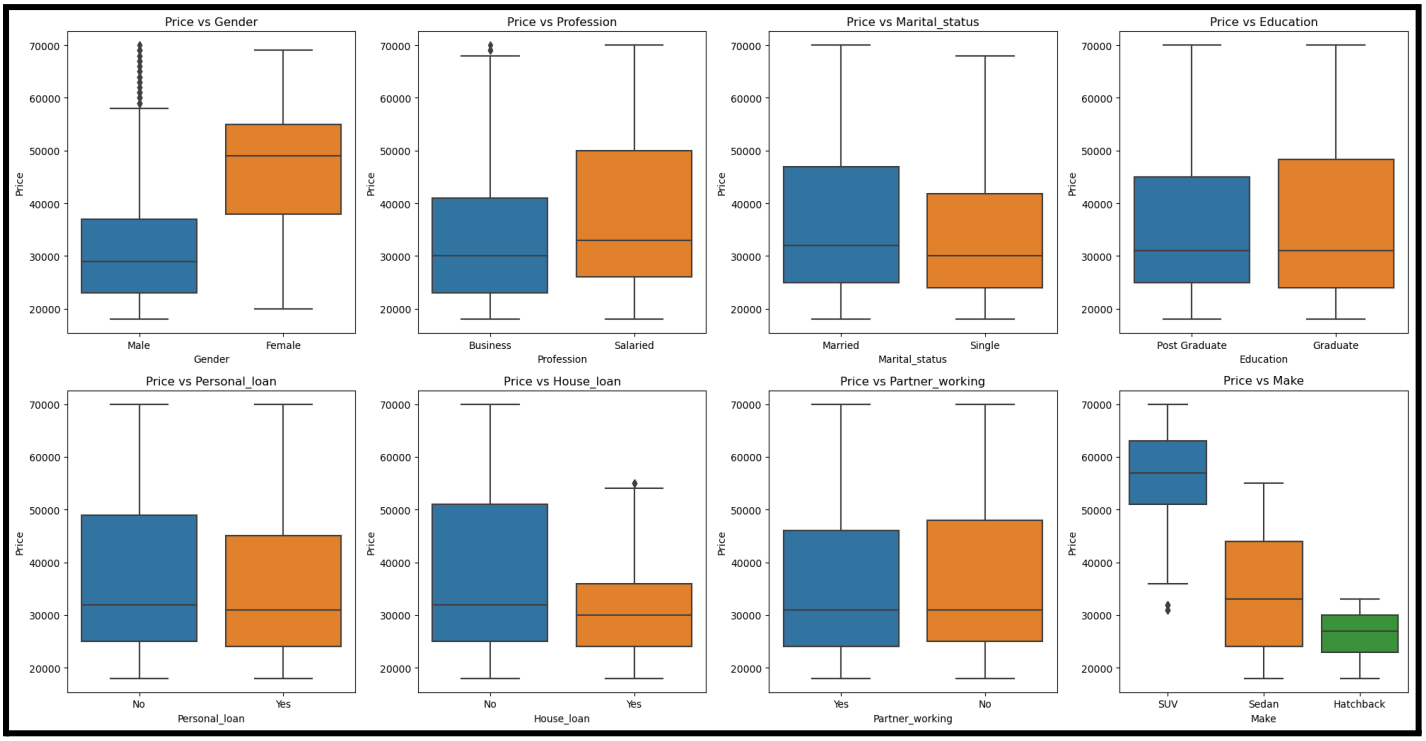
Total Salary vs Personal Loan: The distribution looks similar with outlier in personal loan opted customers

Total Salary vs House Loan: The distribution looks similar with maximum total salary in house loan opted customers

Total Salary vs Partner Working: If partner working is yes then the obviously the total salary is higher compared to partner not working

Total Salary vs Make: Median for SUV> Sedan > Hatchback

**Fig: 3.1**




Price vs Age: 25% of female customers car price is higher than 75% of the male customers car price

Price vs Profession: The salaried person's value of buying cars is higher and it is right skewed.

Price vs Marital Status: Married customers bought cars of high values

Price vs Education: Graduate's car buying value is higher compared to postgraduate

Price vs Personal Loan: The distribution looks similar and right skewed

Price vs House Loan: People without house loan have higher depth in terms of car buying price

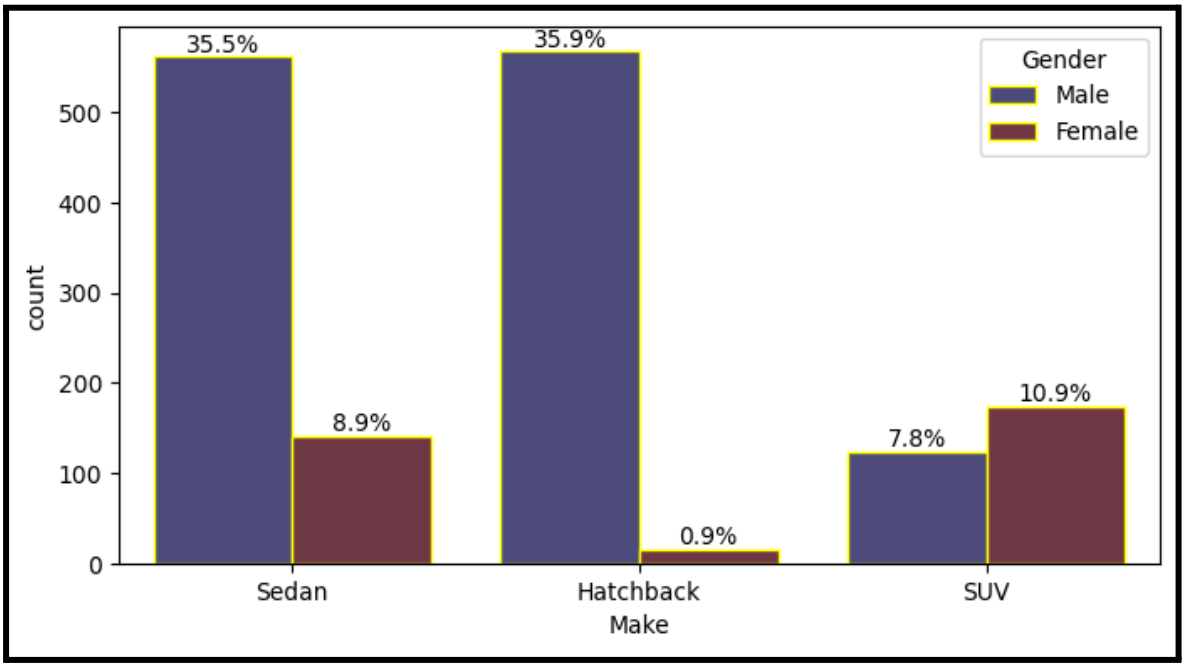
Price vs Partner Working: The distribution looks similar and it is right skewed.

Price vs Make: SUV has higher value compared to other make's

# 3. Problem1 - Key Questions

## 3.1 Do men tend to prefer SUVs more compared to women?

**Fig: 3.2**

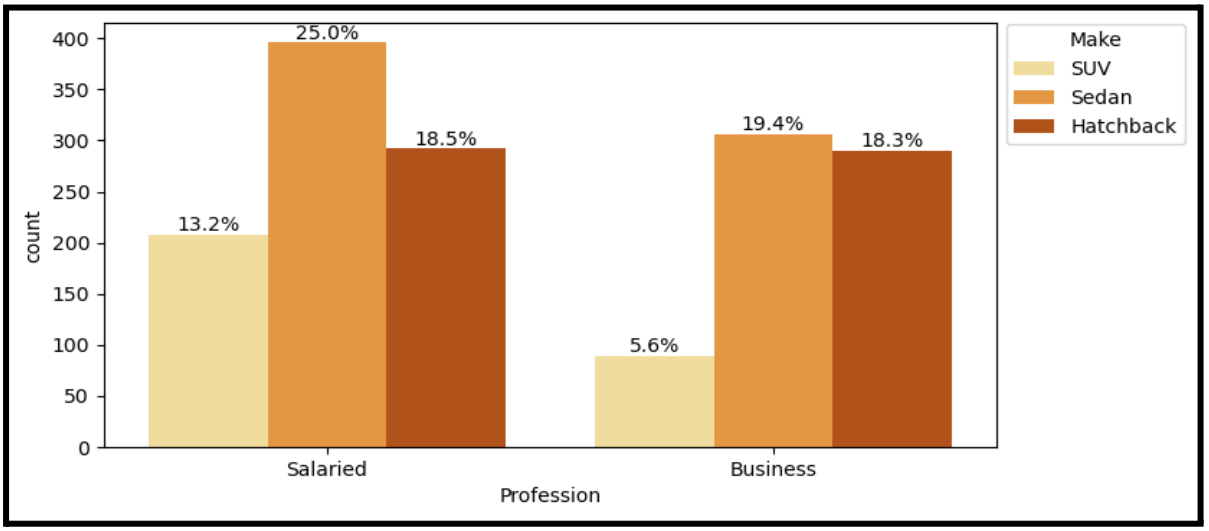



The data clearly shows that the female customers contribute more to SUV cars than male customers.

## 3.2 What is the likelihood of a salaried person buying a Sedan?

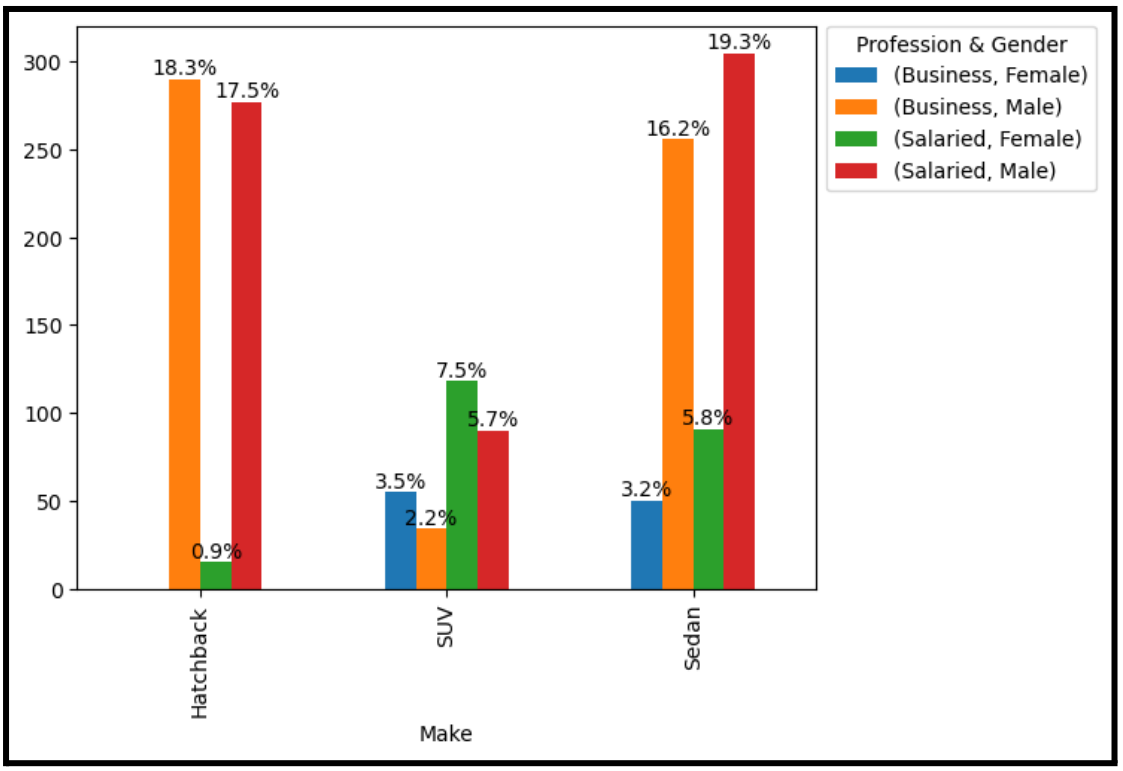
There is a 25% probability of a salaried person buying a sedan.

**Fig: 3.3**



### 3.3 What evidence or data supports Sheldon Cooper's claim that a salaried male is an easier target for a SUV sale over a Sedan sale?

**Fig: 3.4**




There is no evidence supporting Sheldon Cooper's claim that a salaried male is an easier target for a SUV sale over a Sedan sale. But salaried females can be the right target for an SUV.

### 3.4 How does the amount spent on purchasing automobiles vary by gender?

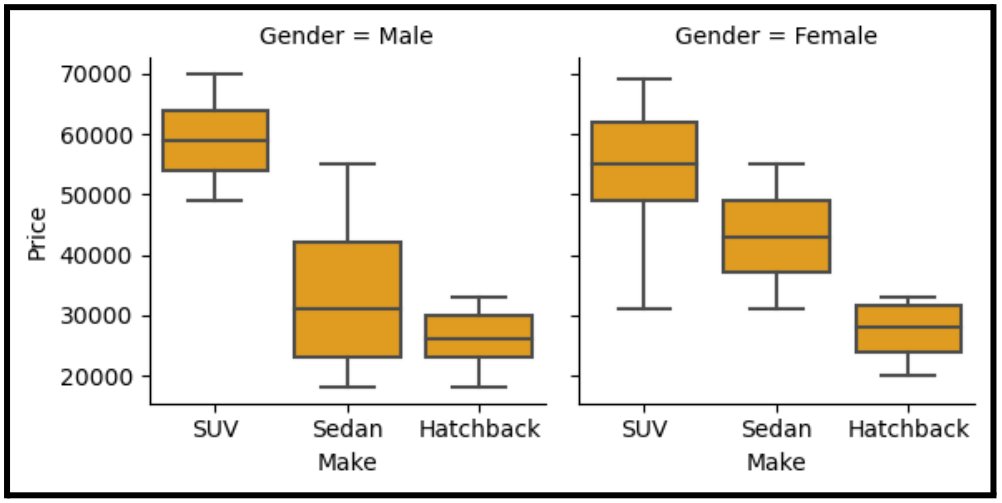



**Fig: 3.5**

SUV: The total amount spent on SUV by females is higher than male. But the female has the minimum value for an SUV. The median value of a male buying a SUV is higher than a female customer.

Sedan: Male contribute more to Sedan cars. The median value of a female buying a Sedan is higher than a male customer. High value sedan cars can be targeted to females rather than male.

Hatchback: The amount spent on Hatchback by both male and female does not show much difference.

### 3.5 How much money was spent on purchasing automobiles by individuals who took a personal loan?

**Table 6**

| Personal_loan | Price |
|---|---|
| No | 28990000 |
| Yes | 27290000 |

The total money spent on purchasing automobiles by individuals who took a personal loan is 2,72,90,000.

## 3.6 How does having a working partner influence the purchase of higher-priced cars?

**Fig: 3.6**

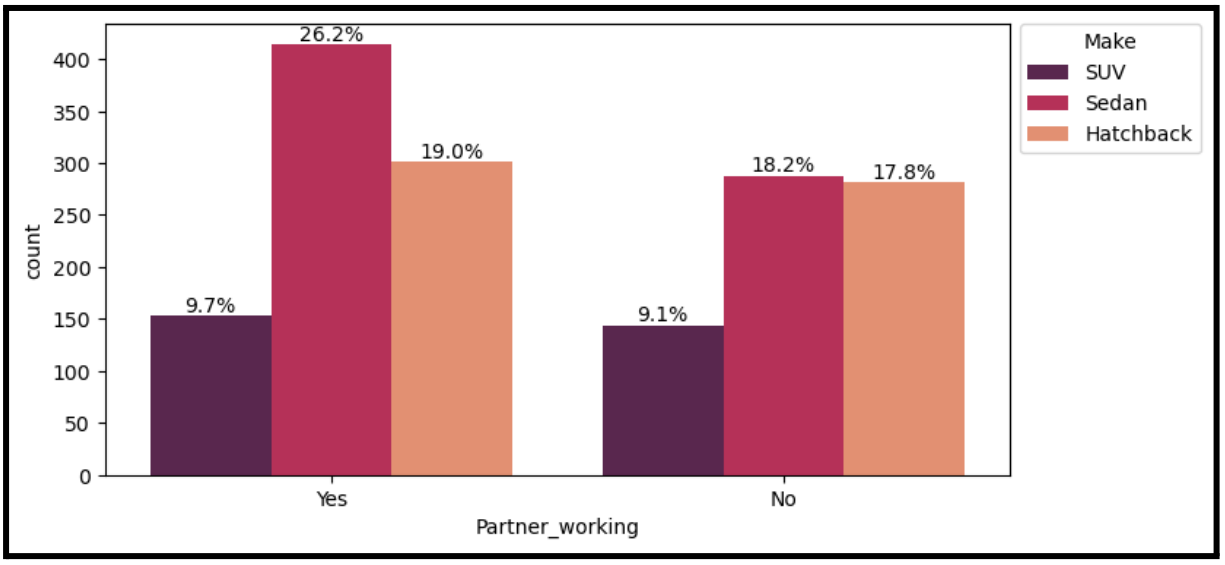



**Table 7**

| ender | Profession | Marital_status | Education | No_of_Dependents | Personal_loan | House_loan | Partner_working | Salary | Partner_salary | Total_salary | Price | Make |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Male | Business | Married | Post Graduate | 4.0 | Yes | No | Yes | 67600 | 40900.0 | 108500.0 | 70000 | SUV |
| Male | Salaried | Married | Graduate | 4.0 | No | No | No | 72000 | 0.0 | 72000.0 | 70000 | SUV |
| emale | Business | Married | Post Graduate | 4.0 | No | No | Yes | 90300 | 25600.0 | 149000.0 | 69000 | SUV |
| emale | Business | Married | Post Graduate | 3.0 | Yes | No | No | 80600 | 0.0 | 80600.0 | 69000 | SUV |
| emale | Salaried | Married | Graduate | 1.0 | Yes | No | Yes | 77800 | 40300.0 | 118100.0 | 69000 | SUV |
| emale | Salaried | Married | Post Graduate | 2.0 | No | No | Yes | 90600 | 80400.0 | 149000.0 | 69000 | SUV |
| Male | Salaried | Married | Post Graduate | 2.0 | No | No | No | 67000 | 0.0 | 67000.0 | 69000 | SUV |
| Male | Business | Married | Graduate | 4.0 | Yes | No | Yes | 86000 | 80500.0 | 149000.0 | 69000 | SUV |
| Male | Salaried | Married | Post Graduate | 4.0 | No | No | Yes | 66900 | 40200.0 | 107100.0 | 69000 | SUV |
| Male | Salaried | Married | Graduate | 4.0 | Yes | No | No | 51800 | 0.0 | 51800.0 | 69000 | SUV |
| emale | Salaried | Married | Post Graduate | 3.0 | No | No | Yes | 80600 | 40600.0 | 121200.0 | 69000 | SUV |
| Male | Salaried | Married | Graduate | 4.0 | Yes | No | No | 54500 | 0.0 | 54500.0 | 69000 | SUV |
| Male | Salaried | Married | Graduate | 4.0 | No | No | Yes | 51600 | 40500.0 | 92100.0 | 69000 | SUV |
| emale | Salaried | Married | Post Graduate | 3.0 | No | No | Yes | 87500 | 70700.0 | 149000.0 | 68000 | SUV |
| emale | Business | Married | Graduate | 2.0 | No | No | No | 73400 | 0.0 | 73400.0 | 68000 | SUV |
| emale | Salaried | Married | Post Graduate | 3.0 | No | No | Yes | 81900 | 27800.0 | 109700.0 | 68000 | SUV |
| emale | Salaried | Married | Graduate | 2.0 | No | No | Yes | 51800 | 28500.0 | 80300.0 | 68000 | SUV |
| Male | Salaried | Single | Post Graduate | 2.0 | No | No | No | 70100 | 0.0 | 70100.0 | 68000 | SUV |
| emale | Salaried | Married | Post Graduate | 2.0 | Yes | No | No | 66600 | 0.0 | 66600.0 | 68000 | SUV |

The first 40 records show the high-priced cars as a SUV. So, there is ~0.6% difference between having a working partner buying a high-priced car.

# 4. Problem1 - Actionable Insights & Recommendations

## 4.1 Actionable Insights

- Overall 75% of the customers age lies between 25 - 38~.
- 50% of car prices lie above 32K~. Customers with above 32k salary can be targeted for high value cars
- 79.2% of the customers are male compared to overall customer base and an easier target for sales
- 56.7% of the customers are salaried while 43.3% are business people
- 91.3% of the customers are married. Probability of selling a car to married person is higher
- 50%~ of the customer base has a personal loan. 66.7% of the customer does have a house loan.
- 54.9% of the customers' partners are working. Who are tend to buy high value cars comparatively
- Sedan is the highest contributor with 44.4% and shows the high probability of people buying sedan car types. Next the Hatchback stands with 36.8%
- Total Salary & Partner Salary have Positive correlation. When partner salary is added the total salary increases
- Whenever the age increases/higher the value of the car price also increases/higher
- Sedan car is owned by customers who no. of dependents whose median is 2
- The customers having partner working having maximum salary which in turn tend to buy high value cars
- The median salary of SUV customers > Sedan > Hatchback
- The total salary of female customers is always higher since most of the male partners are working. Both distributions are right skewed.
- 25% of female customers car price is higher than 75% of the male customers car price
- Married customers bought cars of high values and in high numbers
- SUV has higher value compared to other make’s
- The data clearly shows that the female customers contributes more to SUV cars than male customers
- There is a 25% possibility of a salaried person buying a sedan
- Salaried male is not an easier target for a SUV sale over a Sedan sale

- The total amount spent on SUV cars by females is higher than male. But the female has the minimum value for an SUV. The median value of a male buying a SUV is higher than a female customer
- Male contribute more to Sedan cars. The median value of a female buying a Sedan is higher than a male customer. High value sedan cars can be targeted to females rather than male.

## 4.2 Recommendations

- Hatchback:

    - Male customers contribute more to the car sales. Especially married male contribute to 31.7% hatchback car sales
    - Younger people (~22 to 27 of age) are the easy target for selling Hatchback cars
    - The median salary is >~57K which is lesser among other car makers
    - People buying car for the first time are target customers for Hatchback cars
    - The price ranges from min 18k to max 33k~ which is the lowest high price among other car makers

- Sedan:

    - Male customers contribute 33.8% to the sedan car sales but females tend to buy higher value cars than male as the median for sedan cars for females is higher than male. Targeting female customers for high value sedan cars is better choice
    - Married customers tend to buy more and higher value sedan cars.
    - Salaried Male & Business Male contribute to 33.8% of the car sales which is sedan
    - Customers with no. of Dependents between 1-3 has high chance of buying sedan
    - Sedan is preferred by salaried person with 25% contribution
    - Customer whose partner is working having impact of sales in sedan car types with 26.2%

- SUV:
    - Female customers contribute more to SUV cars than male or high value cars than male - 11.3%
    - Salaried  customers shows more intention towards buying an SUV especially salaried female has contribution of 7.7%
    - Target married customers who always have a higher chance of buying 18.4% compared to 1% of unmarried customers. Even the median value of SUV's for married people are higher
    - Customers who do not have a house tend to buy high value cars (18%) and who contribute a high number of car purchases overall (69%). Target a customers without a house loan have higher probability for high value cars
    - Customers with higher total salary where their age is ~37+, gender as female and partner working status as yes having a high probability of getting SUV cars with high value
    - Since age has a positive with the car price the higher age customers will tend to buy high value cars
    - Salaried customers could be a better target for SUV since their total salary increase based on partner salary especially female customers.

# 5. Problem2 - Farming Analytical Problem

## 5.1 Analyse the dataset

## 5.2 Framing initial questions

Q1: Number of customers using different type of cards

Q2: Customer who carries balances over from one month to the next?

Q3: Customer category based on their net-worth value?

Q4: What is the number of widget products & number of engagement products used by customers?

Q5: What is the future use of credit cards by the customers from T+1,2,3,6 & 12 months ?

Q6: What is the Average credit card spent in the last 3 months by customers?

Q7: What is the effect of change in CC limit based on their income levels?

Q8: What is the relation between L3M avg spend and their income levels?

Q9(1,2,3): What is the number of customers who are using their cc for last 30,60,90 days and their future usage of the cc

Q10: What is the relation between L3M avg spend and their cc limit?

Q11: What is the networth of customers based on the cardtype they use?

Q12: What is the networth of customers who carry forward the balance to next month?

Q13: Which occupation of customers contributes more to the transactors category?

Q14: How many customers based on their occupation hold other banks?

Q15: How many customers carry forward the balance to next month, having cc in other banks?

Q16: What is the cc limit provided to customers and the cc card type they use?

Q17: What is the cc active status of the customers in 30,60 and 90 days who use other bank credit cards?

Q18: How many customers with different occupation status hold other bank credit cards?

Q19: What is the cc active status for 30, 60 & 90 days for customers who carry forward their balance?

Q20: What is the cc active status for T+1,2,3,6 & 12 months for customers who carry forward their balance?

Q21: what is the last 3 months avg spend of credit card of customers with different card types?

Q22: what is the income recorded at source for cc application of customers with different card types?

Q23: What is the correlation between the products usage, income, bank vintage, avg spend and cc limit?

# 5.3 5 Important variables & Business Justification

## 5.3.1 Card Type

**Fig: 3.7**

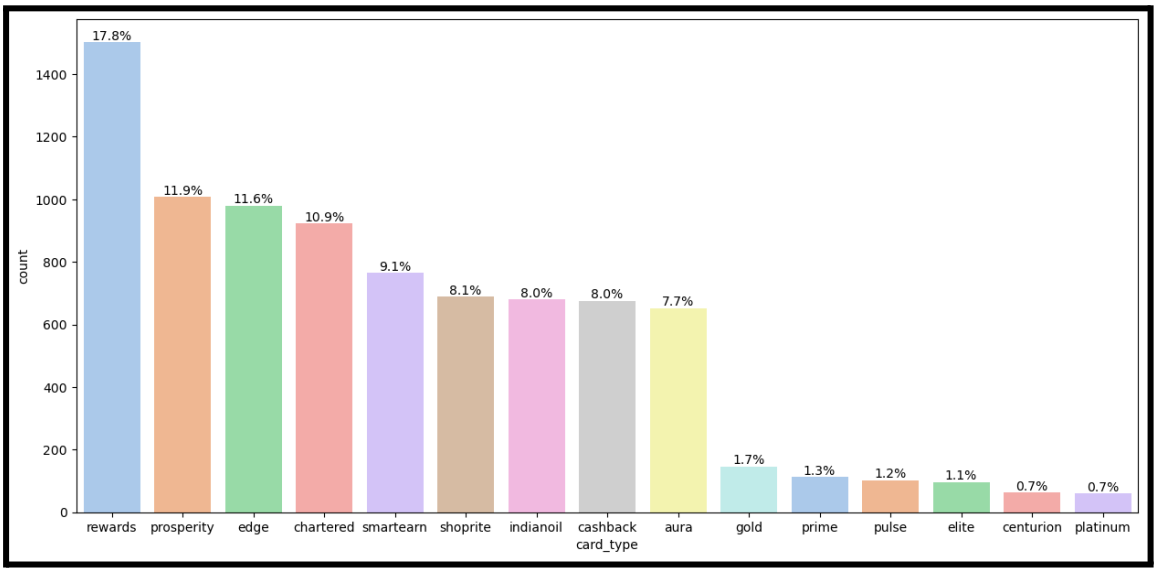



**Fig: 3.8**

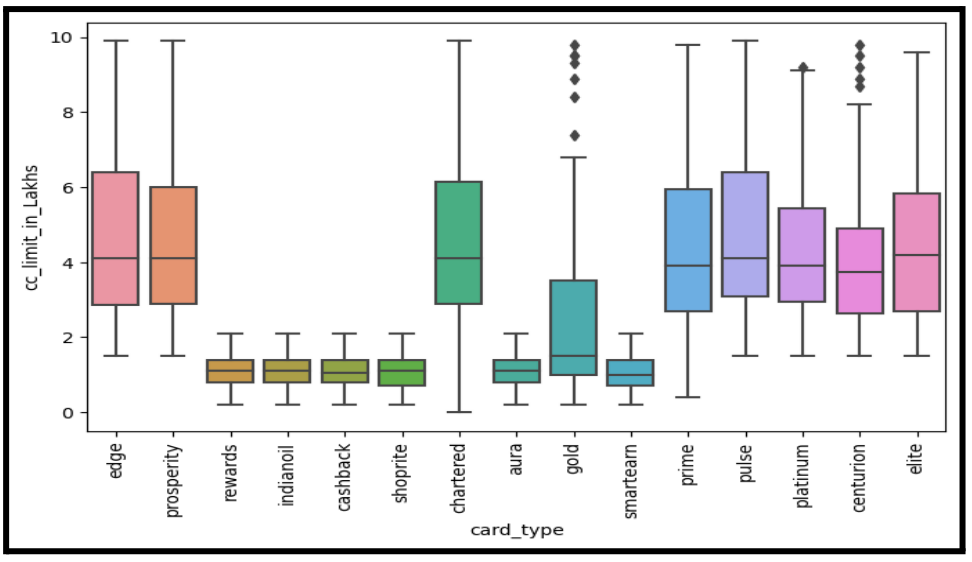


**Fig: 3.9**

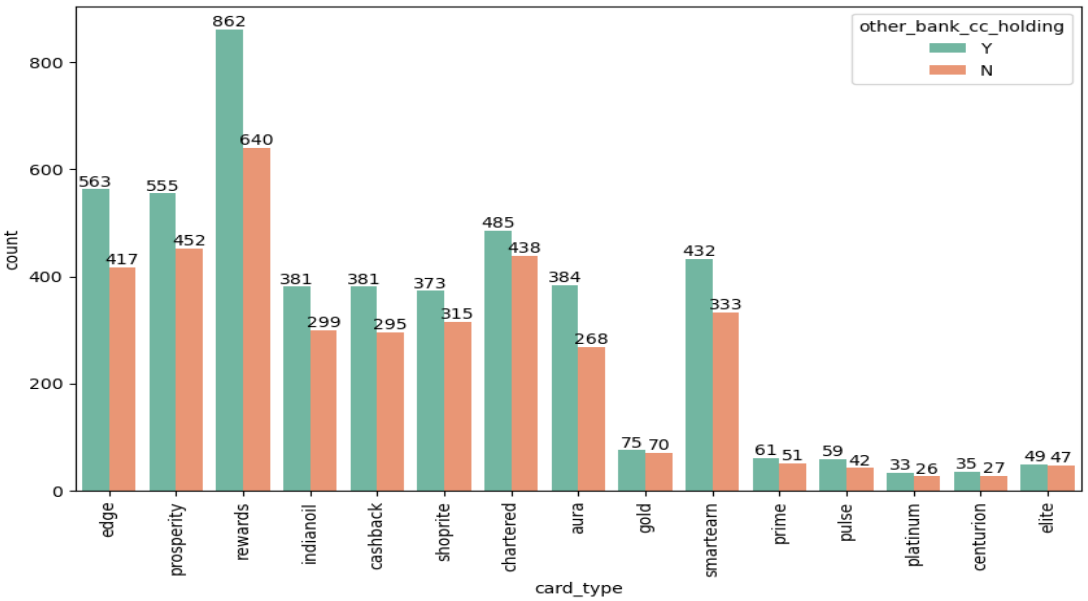
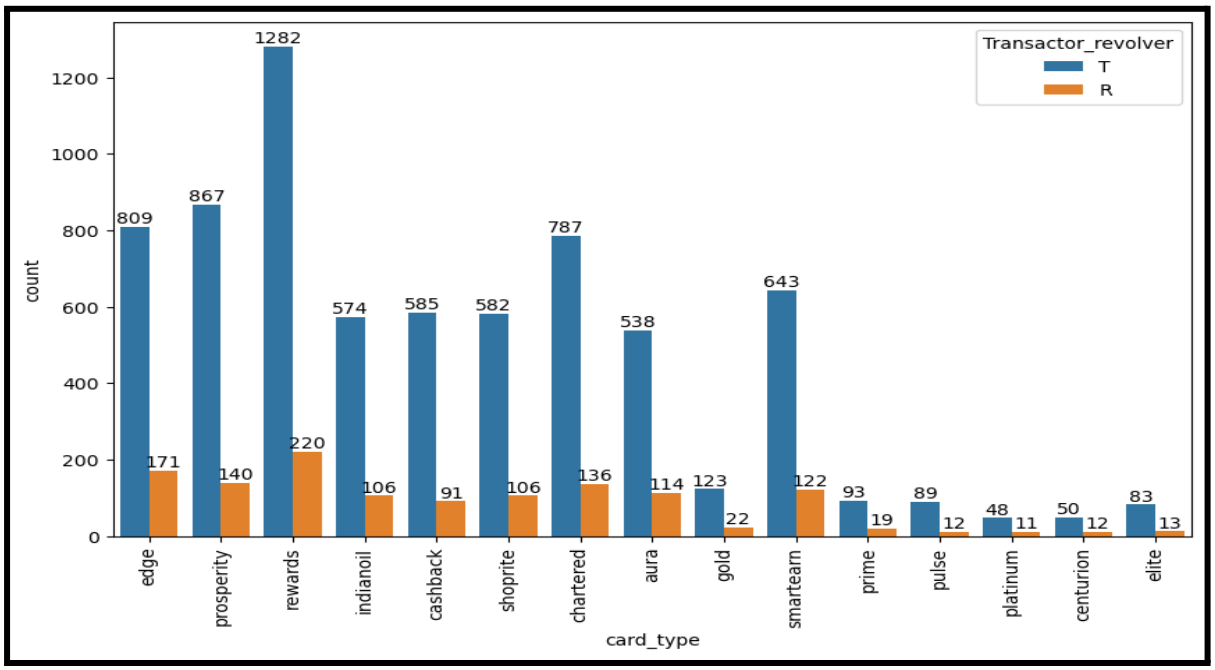

other_bank_cc_holding
Y
N
count
card_type
edge
prosperity
rewards
indianoil
cashback
shoprite
chartered
aura
gold
smarteam
prime
pulse
platinum
centurion
elite
563
417
555
452
862
640
381
299
381
295
373
315
485
438
384
268
75
70
432
333
61
51
59
42
33
26
35
27
49
47

**Fig: 4.0**


Transactor_revolver
T
R
count
card_type
edge
prosperity
rewards
indianoil
cashback
shoprite
chartered
aura
gold
smarteam
prime
pulse
platinum
centurion
elite
809
171
867
140
1282
220
574
106
585
91
582
106
787
136
538
114
123
22
643
122
93
19
89
12
48
11
50
12
83
13

## Business Justification:

- The ‘reward’ type credit card has a higher customer contribution with 17.8%. But the max. cc limit given was 2L~
- Also, the no. of customers who carry forward their balance was higher with 1.2K customers. Which means most of the customers who use the credit card to circulate the value.
- The same set of customers has the highest contribution towards holding other bank credit cards. which will clearly indicate us that the customers may stop using the card or cancel/close the credit card

## 5.3.2 Other Bank Credit Card Holding

**Fig: 4.1**

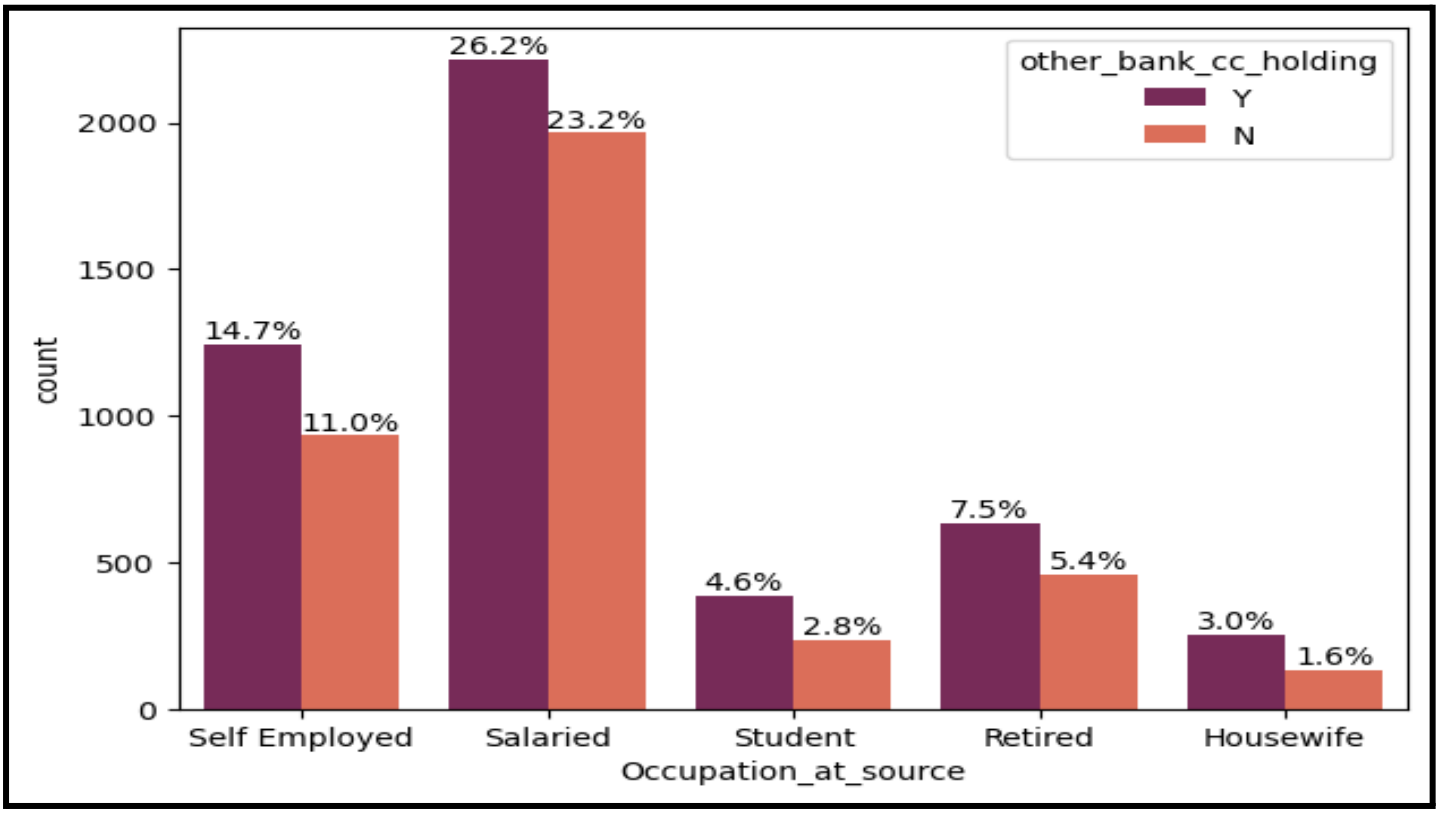
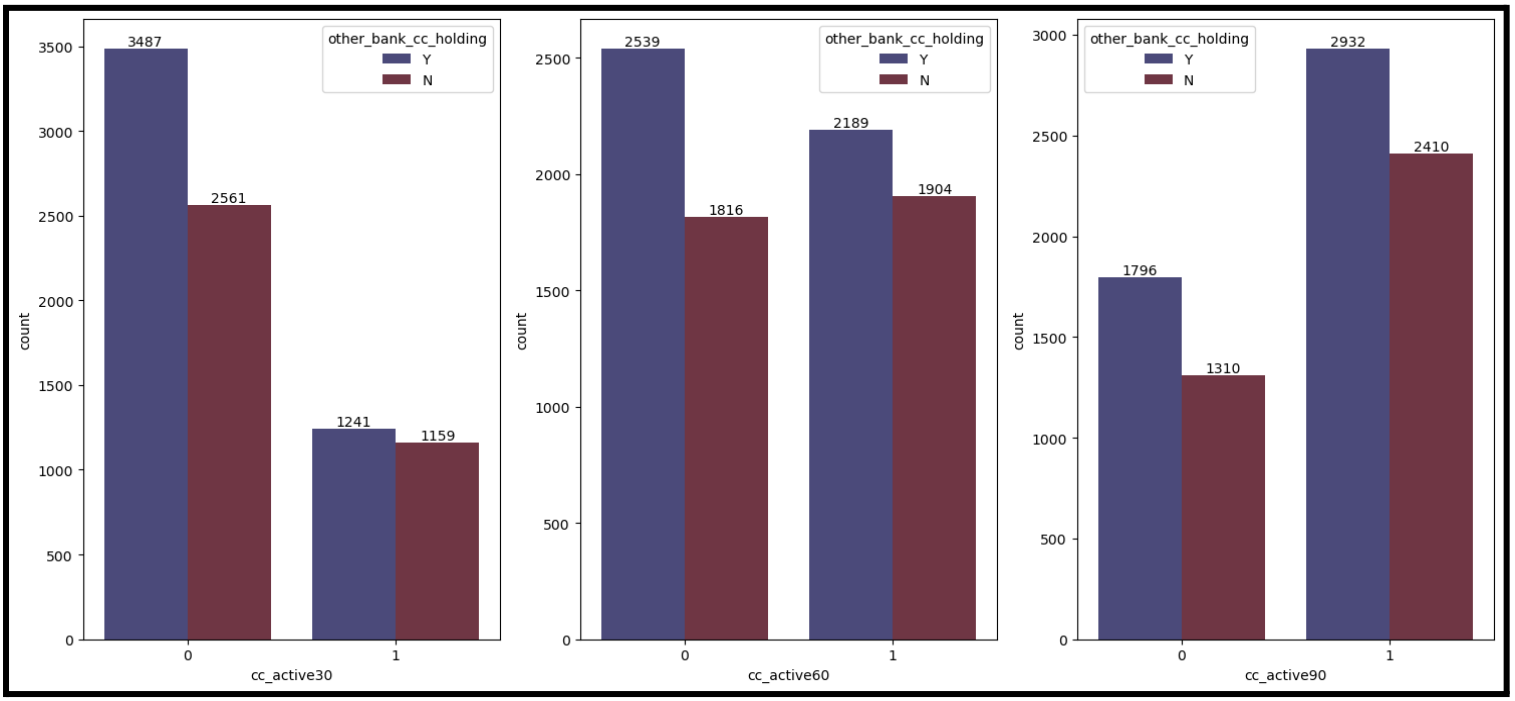


**Fig: 4.2**


other_bank_cc_holding
Y
N
count
3487
2561
1241
1159
cc_active30
2539
1816
2189
1904
cc_active60
1796
1310
2932
2410
cc_active90

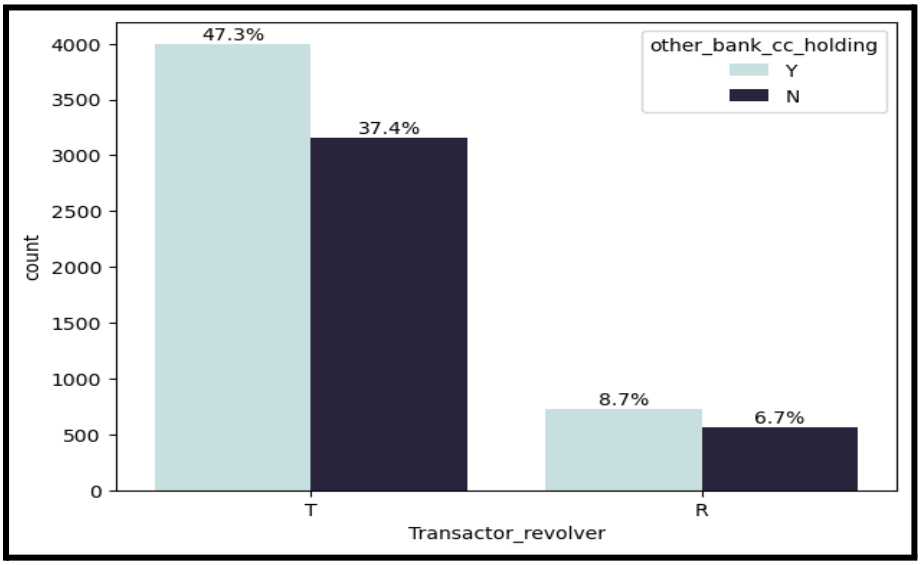
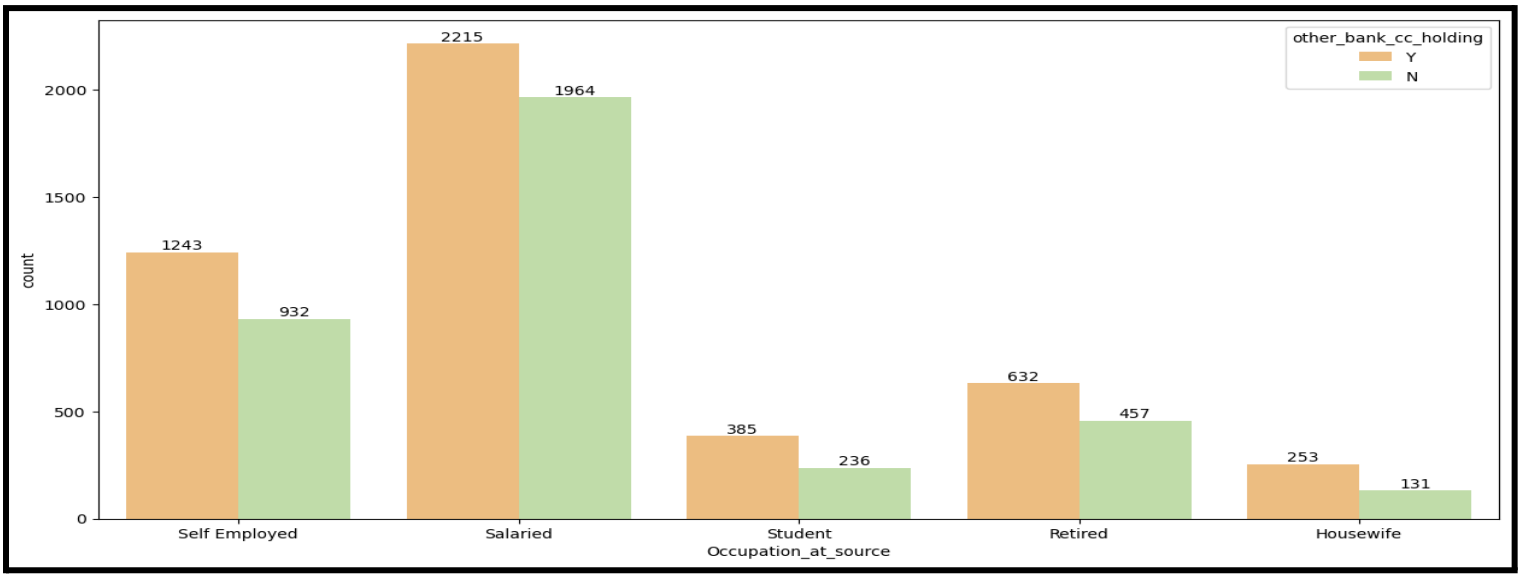

other_bank_cc_holding
Y
N
count
47.3%
37.4%
8.7%
6.7%
T
R
Transactor_revolver

**Fig: 4.3**

**Fig: 4.4**




## Business Justification:

- Holding other bank credit cards have a high chance of customer attrition
- Salaried and self employed customers have major contribution of holding other bank credit cards
- Almost 6048 customers do not have any cc activity in the last 30 days but the usage of credit cards increased over a 90 days period of time. which clearly indicates the low usage of credit cards
- Overall 84.7% of customers are holding other bank credit cards

### 5.3.3 Transactor_Revolver

**Fig: 4.5**

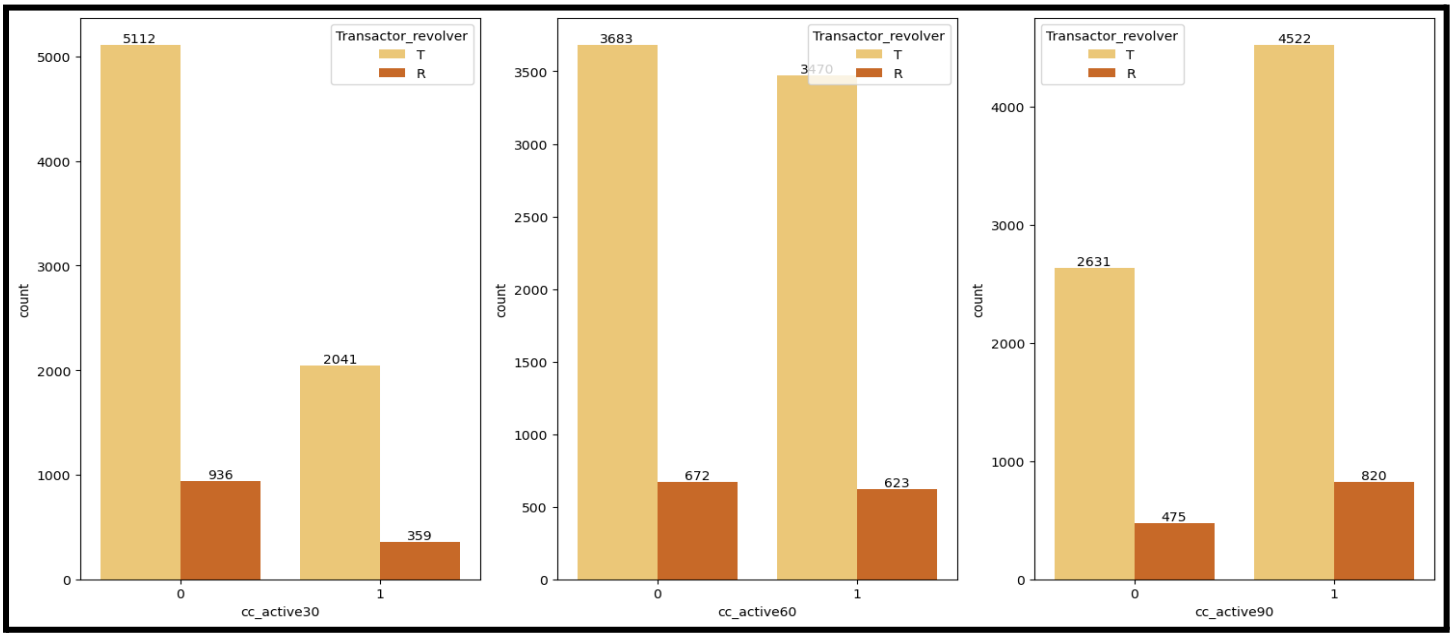



**Fig: 4.6**

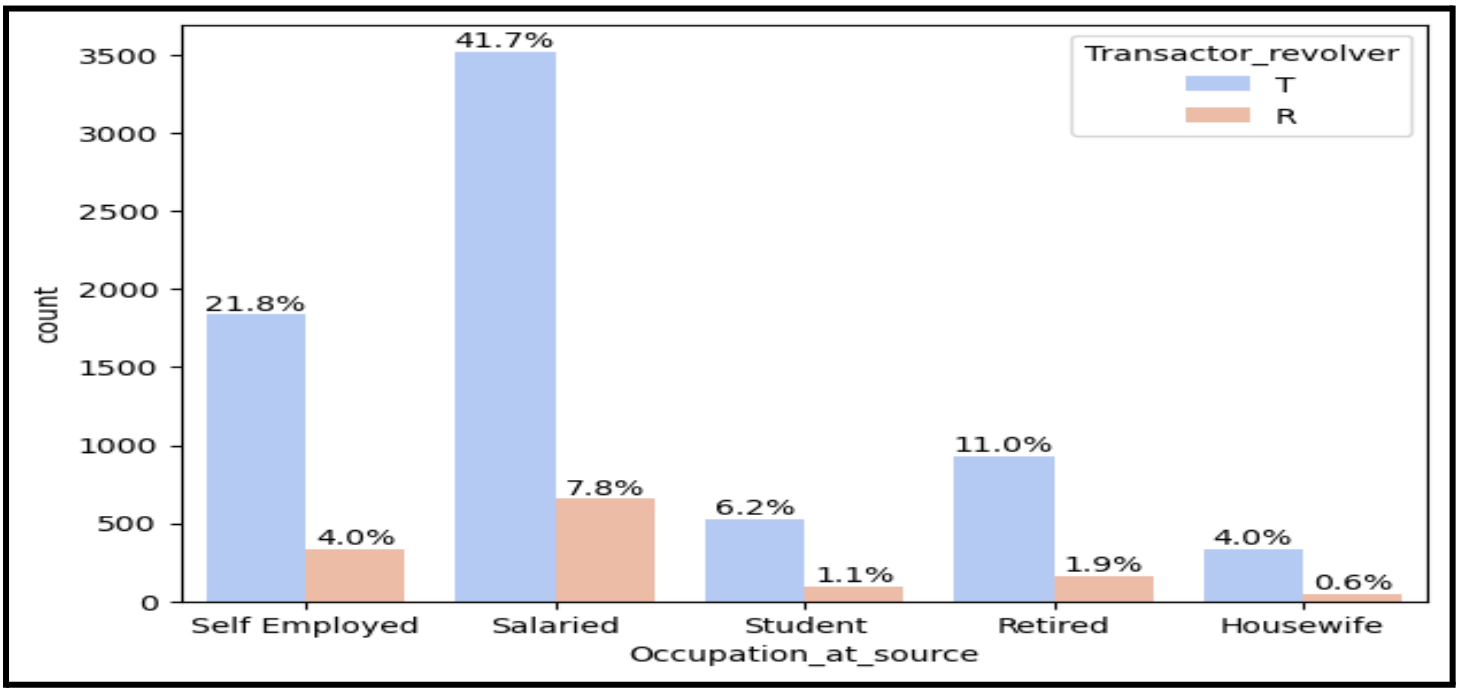
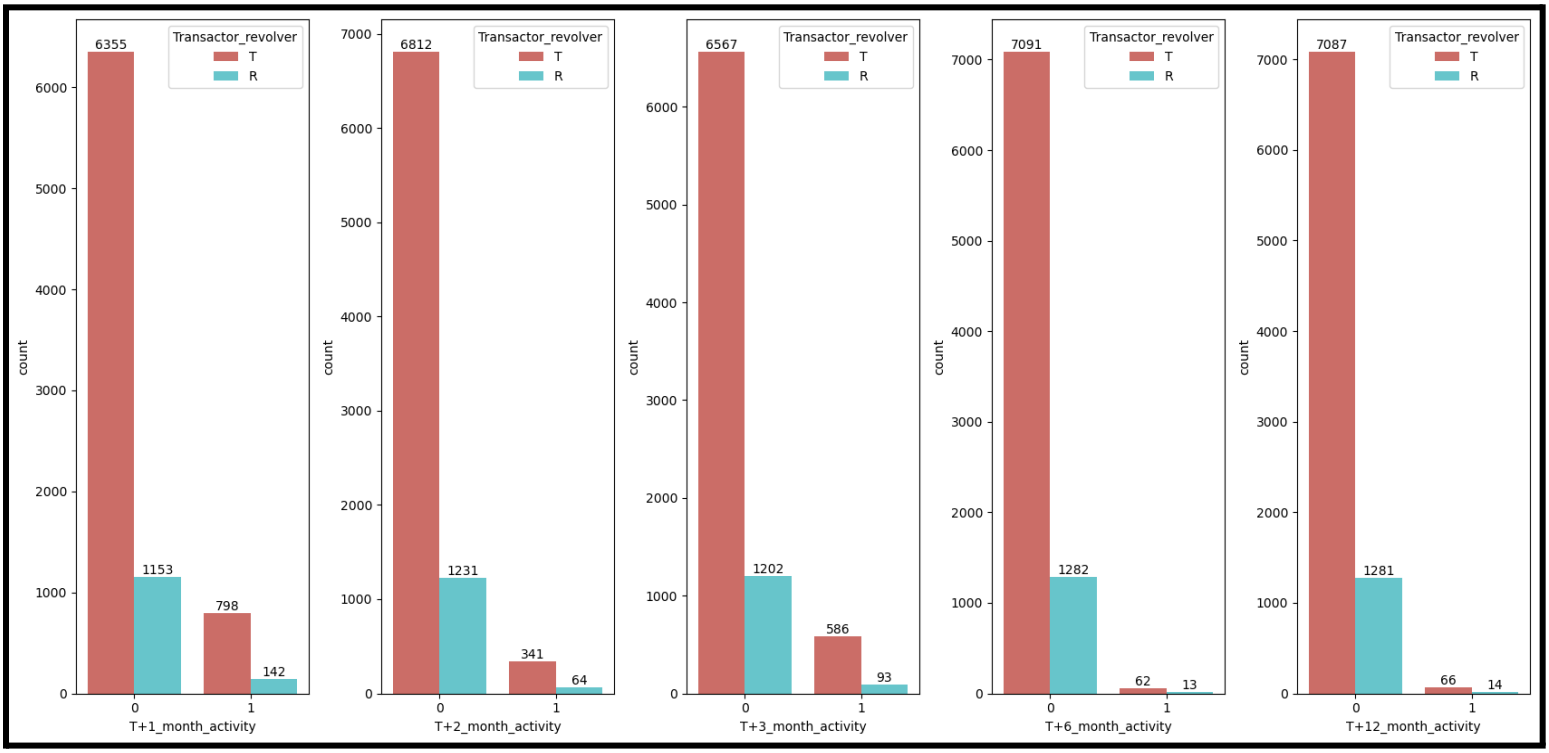


**Fig: 4.7**




## Business Justification:

- Transactor customers tend to use the card again and again. The usage was higher in 60 days and 90 days compared to the usage of 30 days.
- Salaried and self employed contribute to 63.5% compared to other occupation
- The future usage of credit card by these type of customers will be negligible
- The transactor customers  who use other bank cards and having less usage in last 30 day and future usage need to be treated properly for manage the attrition

## 5.3.4 Avg_spends_l3m

**Fig: 4.8**

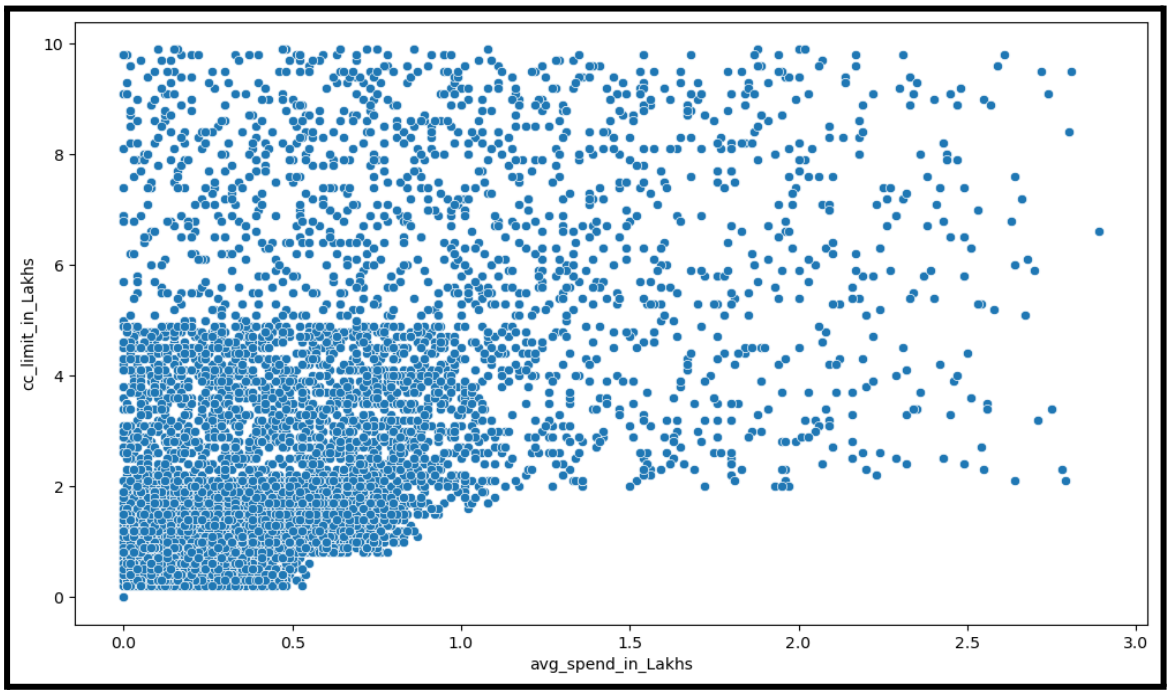
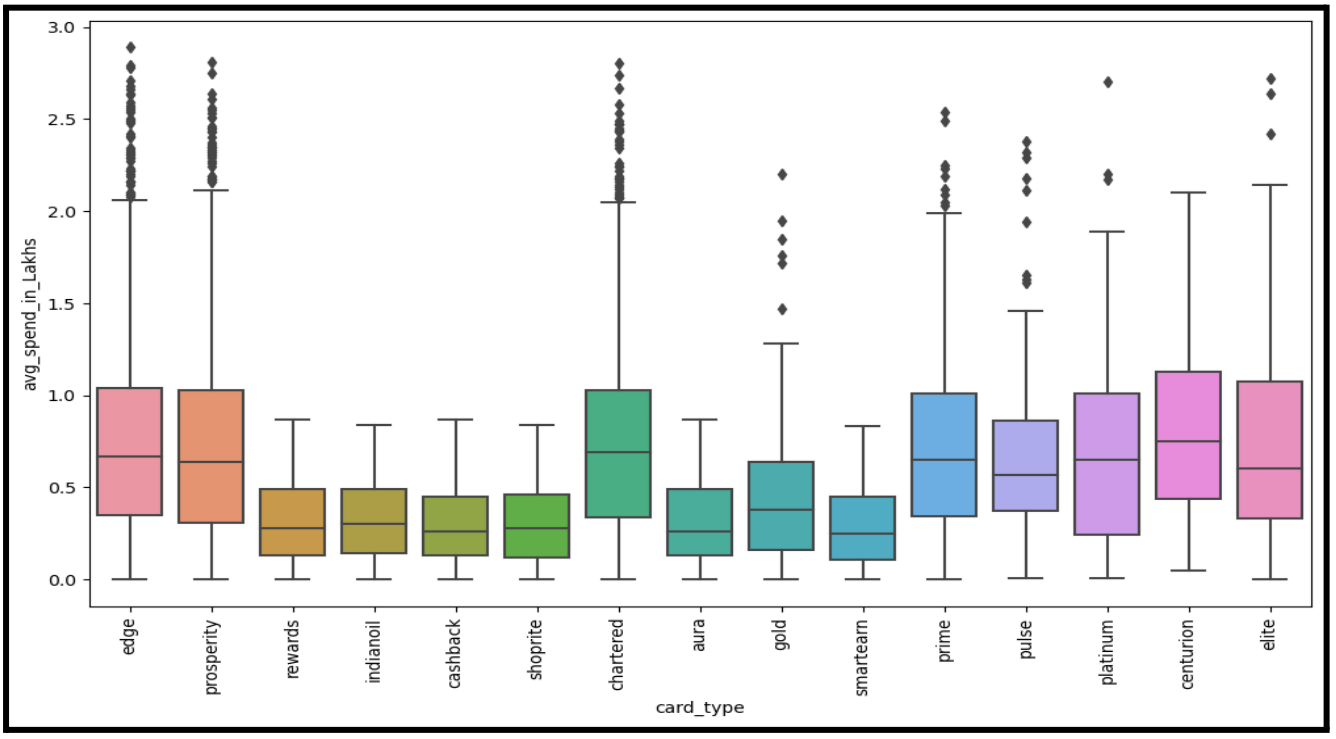



**Fig: 4.9**



## Business Justification:

- Avg amount spend on credit card decides the effective usage of the card
- Customer who has higher avg spend has higher cc limit
- The scatter plot gives the that the increase in cc limit increases the avg spend of customers
- 8 out of 15 credit card types have outliers in the avg spend. Focusing more that particular card type and providing same benefits will increase the usage of other cards

## 5.3.5 Annual_income_at_source

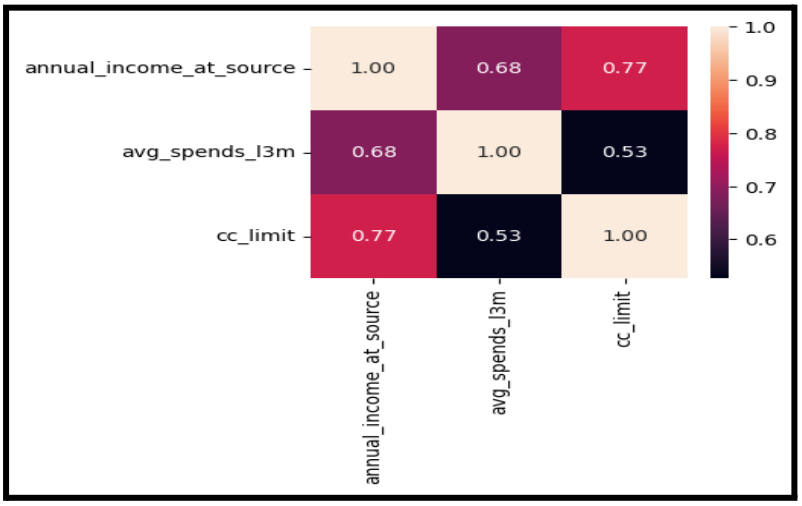



**Fig: 5.0**

**Fig: 5.1**

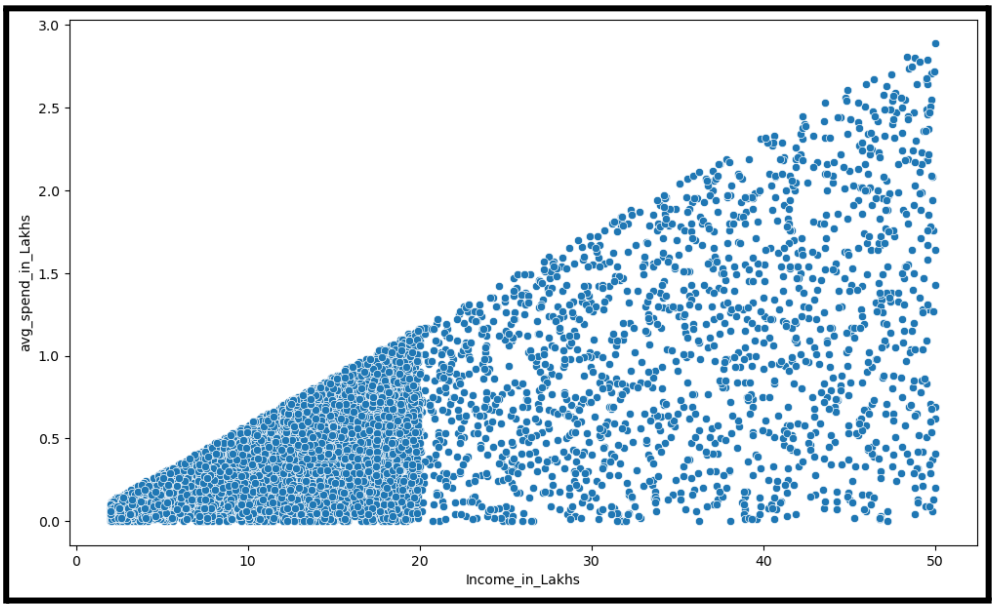
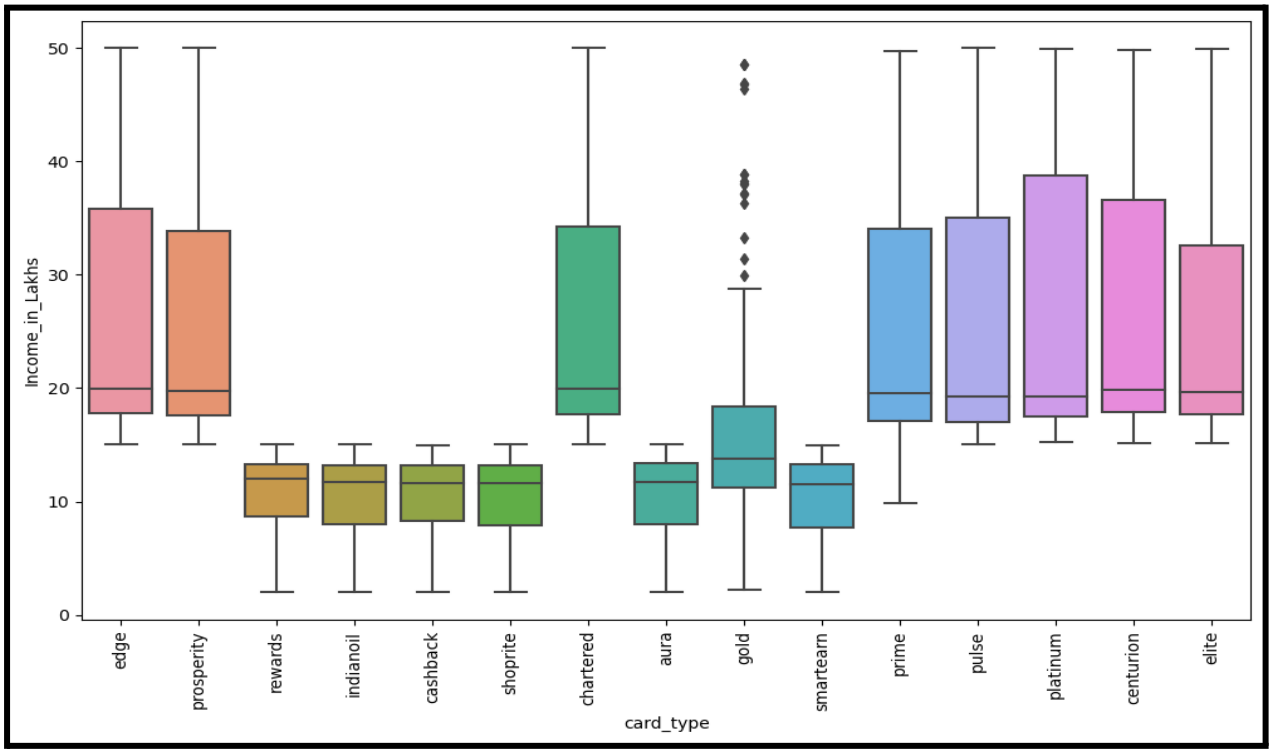


**Fig: 5.2**




## Business Justification:

- Income and cc_limit have positive correlation. Customers above 20 lakhs of income the cc limit vary from 2 lakhs to 10 lakhs~
- The income at source is the key factor for deciding the cc limit
- The income has 0.68 correlation with the avg spend of the credit card
- Based on the income the customers are classified for card types and usage of card based on the same
- Identifying the transactors and customers with min usage of cc/future usage with a customization plan may retain them