

SMDM PROJECT

THANUSRI A

CONTENTS

Problem 1

A. Data overview

B. Univariate analysis

C. Bivariate analysis

D. Key questions.

1. Do men tend to prefer SUVs more compared to women?
2. What is the likelihood of a salaried person buying a Sedan?
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?
4. How does the the amount spent on purchasing automobiles vary by gender?
5. How much money was spent on purchasing automobiles by individuals who took a personal loan?
6. How does having a working partner influence the purchase of higher-priced cars?

E. Actionable Insights & Recommendations

Actionable Insights - Business Recommendations

Problem 2

Framing an analytics problem-

Analyse the dataset and list down the top 5 important variables,along with the business justifications.

Problem 1

A)

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

- Data Frame Provided - Austo Motor Company
- Number of Rows = 1581
- Number of Column= 14
- Float Type Data = 1
- Integer Type Data = 5
- Object Type Data = 8

```
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
```

- Column 'Gender' have 53 null values whereas column 'Partner_salary' have 106 null values

```
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
```

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

Statistical analysis of data as below

	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	14674.826044	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

- Column 'Gender' have 53 null values whereas column 'Partner_salary' have 106 null values
- There are no duplicate rows has been found

```

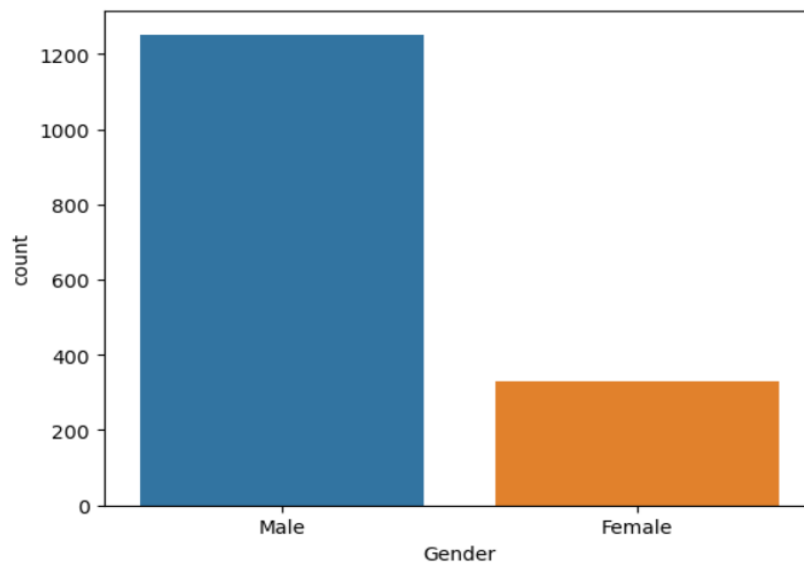
RangeIndex: 1581 entries, 0 to 1580
Data columns (total 14 columns):
#   Column                Non-Null Count  Dtype
---  -
0   Age                    1581 non-null   int64
1   Gender                 1528 non-null   object
2   Profession              1581 non-null   object
3   Marital_status         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

```

- Below unique number has been identified in column "Gender"
- ```
array(['Male', 'Femal', 'Femalè', nan, 'Femle'], dtype=object)
```
- There were two spelling error found in the column "Gender"- 'Femal & Femle'
  - This spelling error for both the error has been corrected and replaced , below is the result of the correction in the data value.

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

- After checking the mode of the 'Gender' data is 'Male' as per below plot.



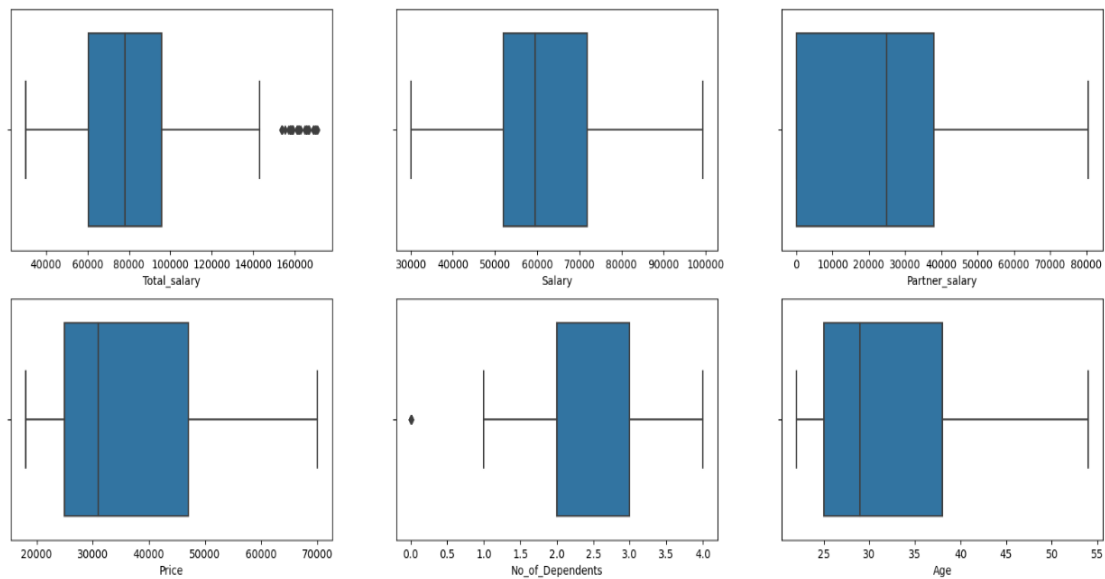
- Null value (Total=53 ) is replaced with **Mode** of the 'Gender' column.
- Replaced NaN value with Partner salary mean value in the 'Partner\_salary' column.

```
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
```

- Now there are no null values in the data set after treatment of 'Gender' column and 'Partner\_salary'
- column as per above table.

B.

### Checking of outliers in the data



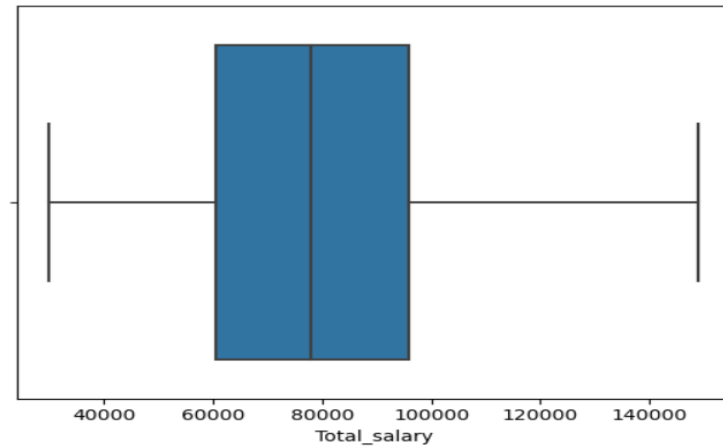
- There are outliers in the 'No\_of\_Dependents' column as well as 'Total\_salary' as per above boxplot.
- I will proceed to treat the outliers for the "Total\_Salary" only because there is probability of having '0' dependent value and treating dependent could lead to mislead the analysis. Kindly refer to below table, there are 20 values with zero Dependents, so I will continue treating only 'Total\_salary'.

```

3 557
2 557
1 229
4 218
0 20
Name: No_of_Dependents, dtype: int64

```

- Also taking mean for the Total salary in order to avoid creating any manipulative analysis and mean will provide us overall correct representation of data.
- Mean of the Total\_salary is 79625.99620493359
- Treating outliers (Total\_salary)
  - Upper range = 149000
  - Lower Range = 7400
  - Q1 = 25%
  - Q3 = 75%
  - Formula to be used = IQR = Q3 - Q1
  - Lower range = Q1 - (1.5 \* IQR)
  - Upper range = Q3 + (1.5 \* IQR)



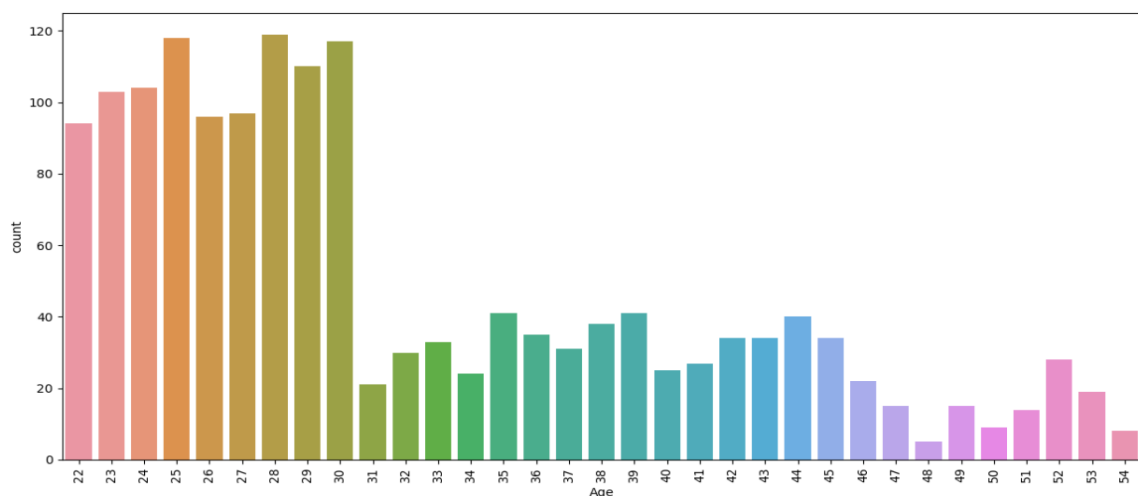
- As we can see from the above plot that outliers has been treated, now there is no outliers for the Total salary.

Exploring all the features of the data separately by using appropriate visualizations and draw insights that can be utilized by the business.

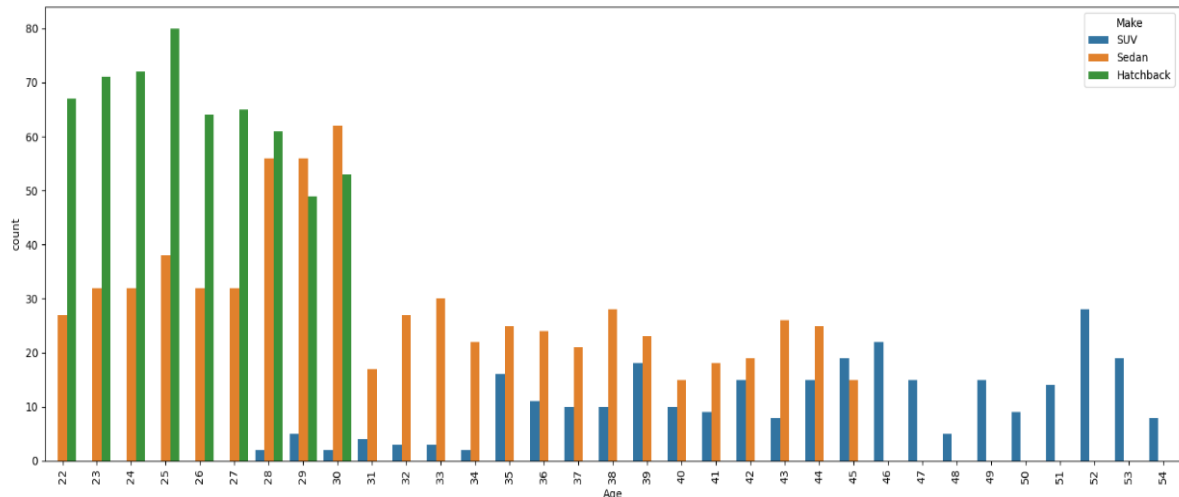
- Statistical analysis of the data which helps to summarize the data.

|                         | count  | mean         | std          | min     | 25%     | 50%     | 75%     | max      |
|-------------------------|--------|--------------|--------------|---------|---------|---------|---------|----------|
| <b>Age</b>              | 1581.0 | 31.922201    | 8.425978     | 22.0    | 25.0    | 29.0    | 38.0    | 54.0     |
| <b>No_of_Dependents</b> | 1581.0 | 2.457938     | 0.943483     | 0.0     | 2.0     | 2.0     | 3.0     | 4.0      |
| <b>Salary</b>           | 1581.0 | 60392.220114 | 14674.826044 | 30000.0 | 51900.0 | 59500.0 | 71800.0 | 99300.0  |
| <b>Partner_salary</b>   | 1475.0 | 20225.559322 | 19573.149277 | 0.0     | 0.0     | 25600.0 | 38300.0 | 80500.0  |
| <b>Total_salary</b>     | 1581.0 | 79625.996205 | 25545.857768 | 30000.0 | 60500.0 | 78000.0 | 95900.0 | 171000.0 |
| <b>Price</b>            | 1581.0 | 35597.722960 | 13633.636545 | 18000.0 | 25000.0 | 31000.0 | 47000.0 | 70000.0  |

### Analyzing the Age Variable



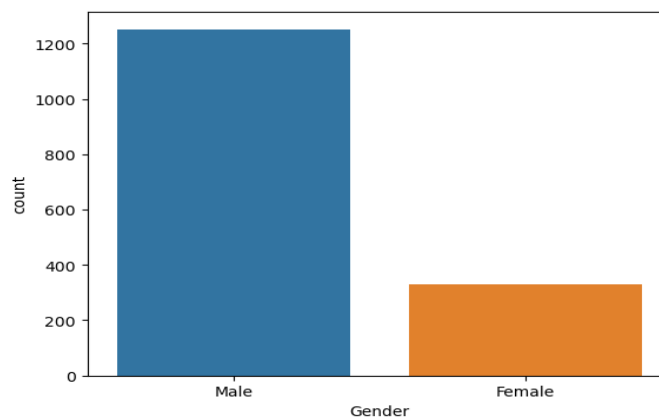
- With reference to above countplot, we can say that buying pattern for the car is w.r.t age group Younger age group (Range 20- 30) tends to buy more cars as compared to the middle aged (Range 31-45) and older age group (range from 46-55). Also there is fluctuation in buying pattern for the age group between 35-40 , sales for the cars between this age group is slightly better after young age group and compared to rest of the age group.



With reference to above ‘Age’ vs ‘Make’ graph we can conclude that

- Younger age group (20-30 ) they mostly buy Hatchback and Sedan as compared to SUV
- Mid age group (31-45) they mostly buy Sedan and SUV (few sales) with no sales of hatchback to this age group.
- Older Age group (46-55) only buy SUV with no sales of Sedan or Hatchback.
- There is zero sales of Hatchback from the age group 31 to 46.

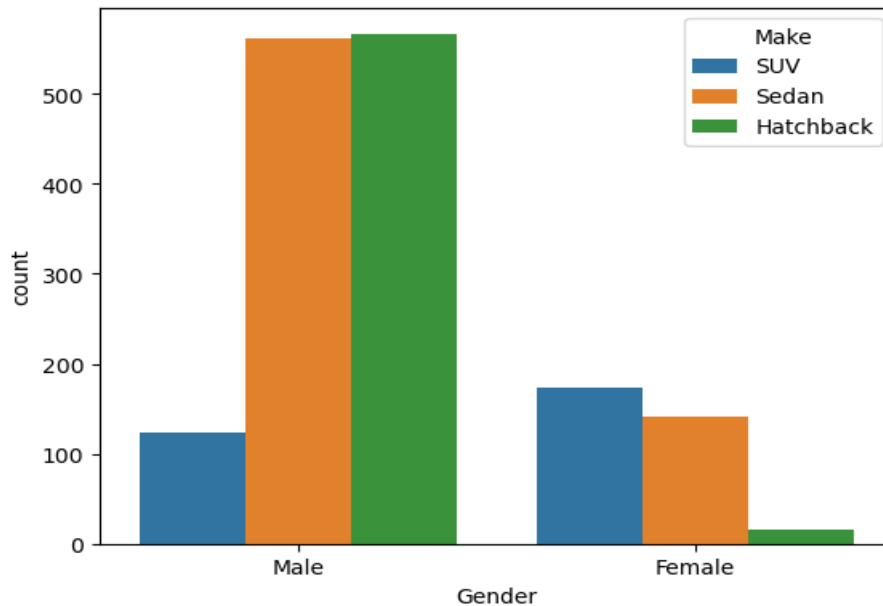
Analyzing ‘Gender’ variables





From the above graph we can say that Male buy more cars than Females.

### Analyzing Gender vs Make variables

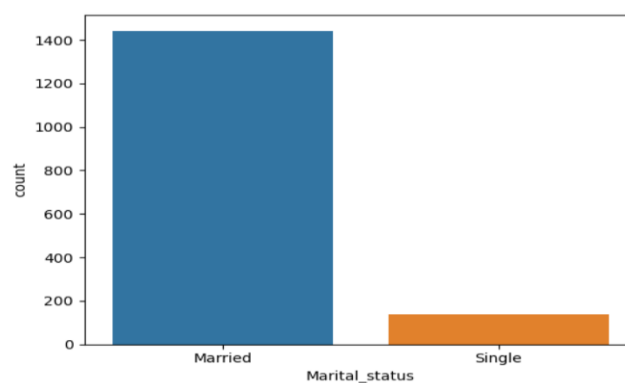


| Make   | Hatchback | SUV      | Sedan    | All      |
|--------|-----------|----------|----------|----------|
| Gender |           |          |          |          |
| Female | 0.009488  | 0.109424 | 0.089184 | 0.208096 |
| Male   | 0.358634  | 0.078431 | 0.354839 | 0.791904 |
| All    | 0.368121  | 0.187856 | 0.444023 | 1.000000 |

With the above data we can conclude that

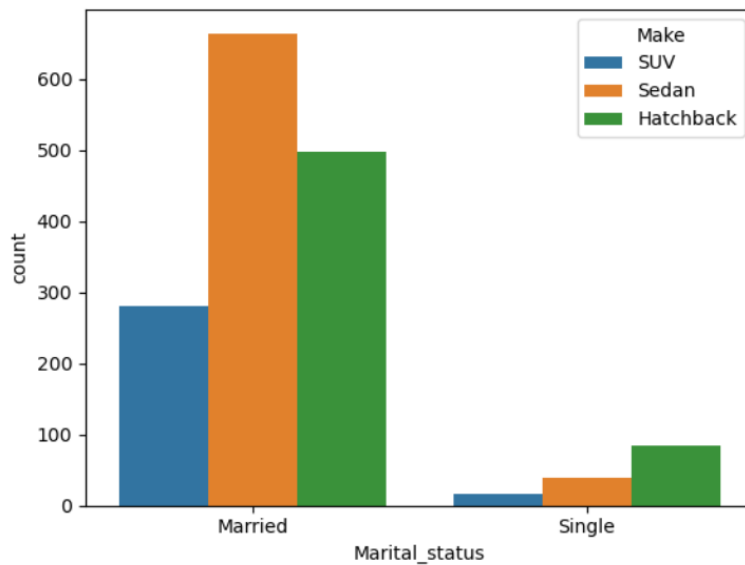
- Females preferred to buy SUV and Sedan and very few Hatchback
- Males preferred to buy Sedan and Hatchback as compared to SUV
- 44 percent of the customers tend to buy Sedan

### Analyzing Marital\_status variable



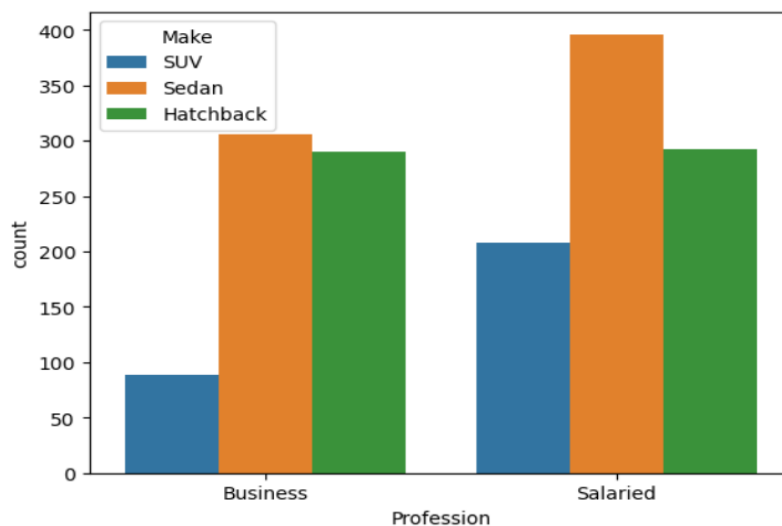
- As per above graph we can say married couple tends to buy more cars as compared to the Singles

#### Analyzing Marital status and Make variables



- Married couples preferred to buy Sedan and Hatchback as compared to SUV's.
- Single preferred to buy Hatchback as well as very little preference to Sedan and SUV.

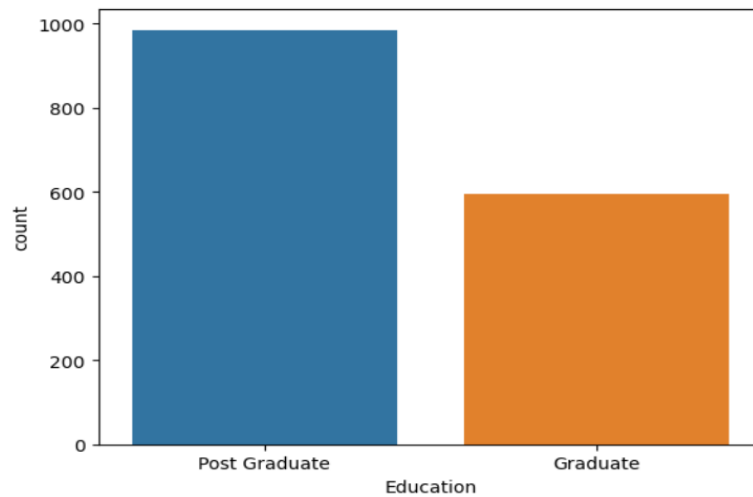
#### Analyzing Profession vs Make variables



- From the above graphs we can conclude that salaried people buy more cars as compared to the business profession group.
- Salaried prefers to buy Sedan as compared to Hatchback with little preference given to SUV.

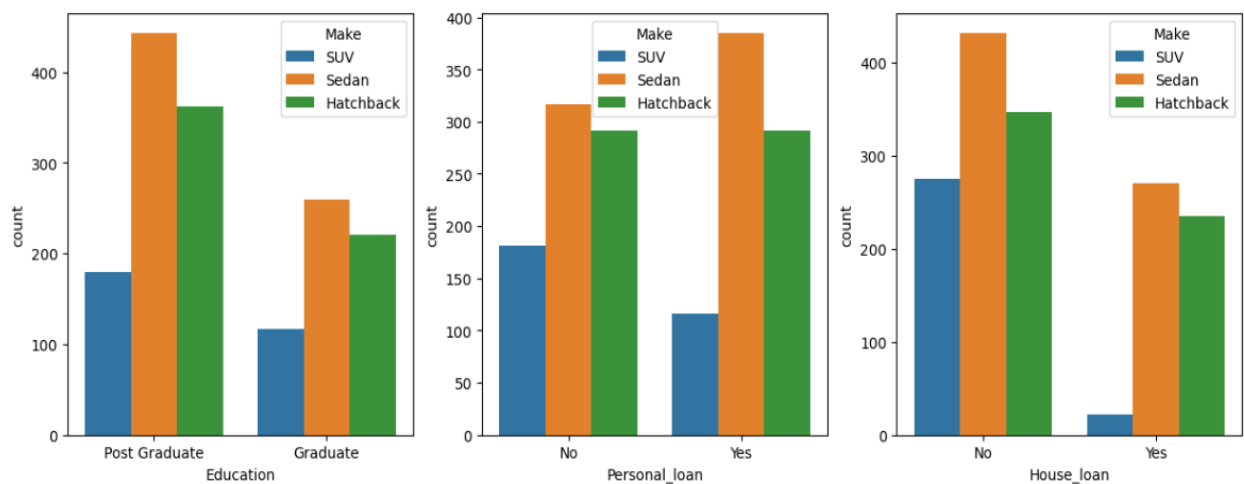
- Whereas Business professionals prefer to buy Sedan and Hatchback as compared to SUV. SUV is not much preferred by the Business people.

#### Analyzing Education variables



As per the above graph we can say that Post Graduate buys more Cars than Graduates.

#### Analyzing the Education, Personal loan and House loan variables

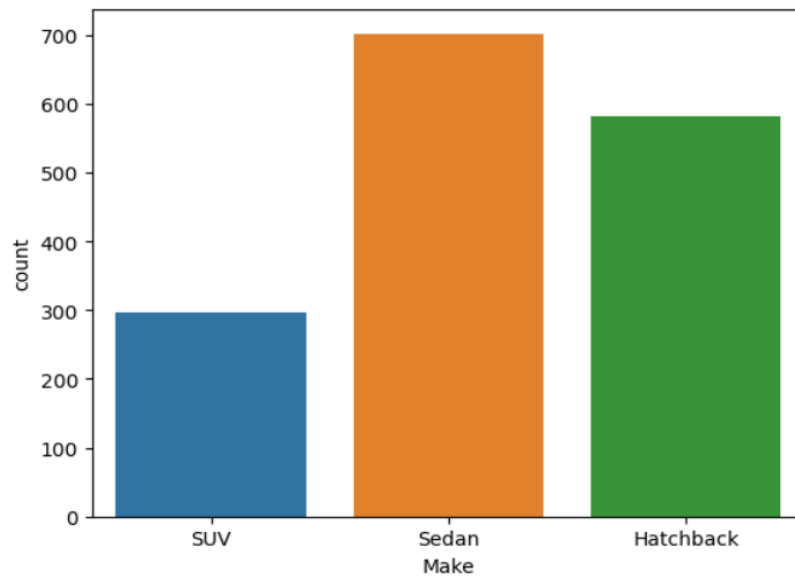


As per above data and graph we can conclude that

- Post graduates student mostly buy Sedan and Hatchback with lesser preference to SUV as compared to other make.
- Graduates also prefer to buy Sedan and hatchback as compared to SUV.
- Customers with and without Personal\_loan buy more Sedan and Hatchback as compared to SUV
- Customers with and without House\_loan buy more Sedan and Hatchback with lesser preference given to SUV.

- Customers without house loan buys more SUV as compared to customer with House\_ loan

Analtzing Make variable



- With above plot we can say Company is making more Sedan and hatchback as compared to SUV , Hatch back manufacturing takes second place after Sedan.

With above all the data we can concludes that

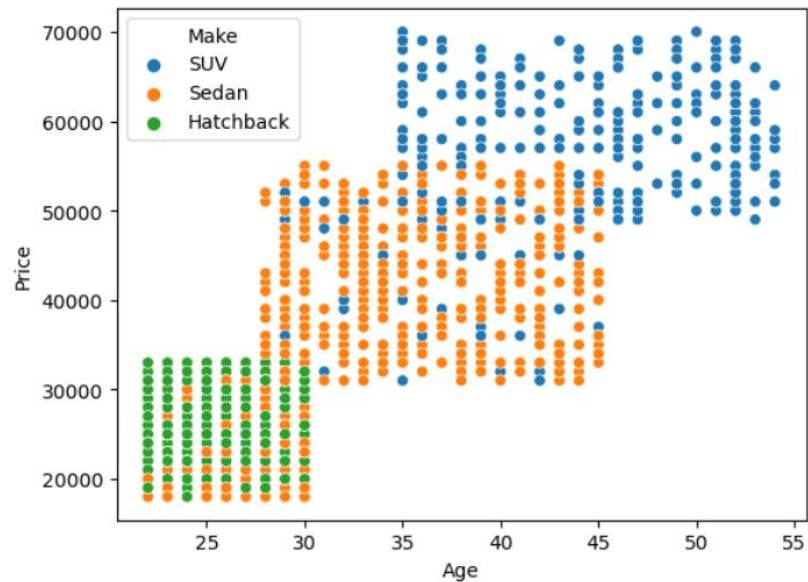
- Average age for buying cars is 29.
- Average price to buy car is 35597.722960151805
- Average price for the Make is as below

```
Make
Hatchback 26474.226804
SUV 55824.915825
Sedan 34603.988604
Name: Price, dtype: float64
```

C.

Understanding the relationships among the variables in the dataset to gain deeper insights.

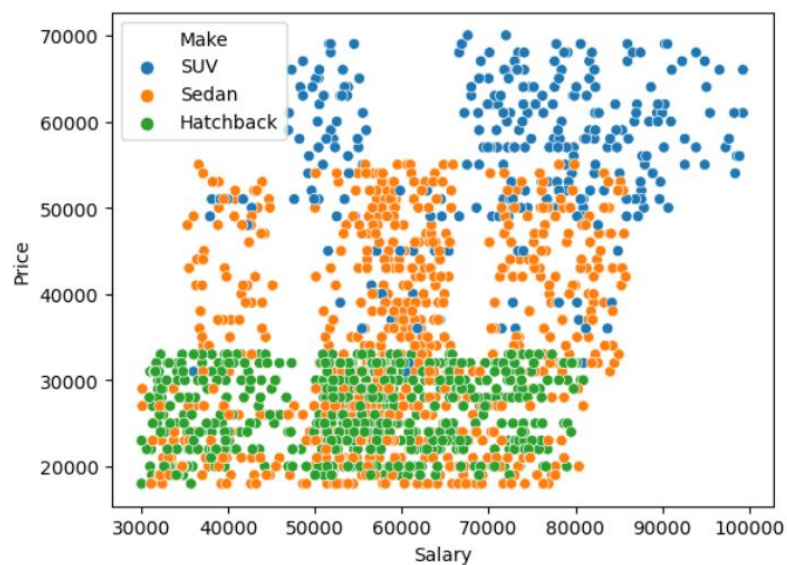
Establishing co-relation between Age & Price



With above figure as reference, there is positive correlation between age of customer and amount of moneyspent on the buying cars and as the customer age increases they tends to buy more expensive cars , this clearly shown in the above scatterplot.

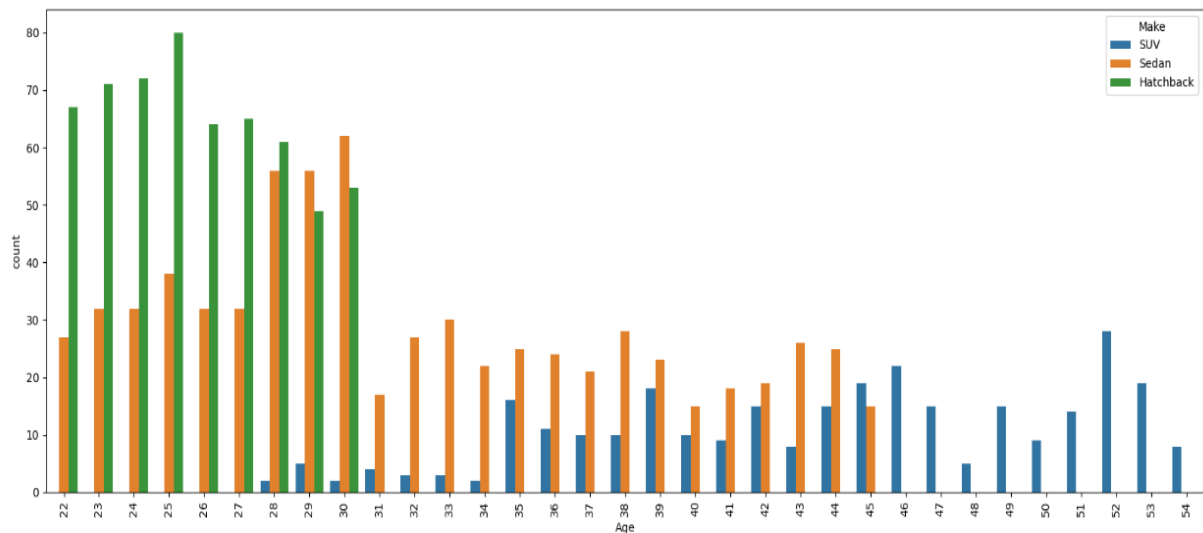
As the age of the Customer increases the amount of money spent on this automotive sector also increases. Age and Price are positively correlated.

Establishing co-relation between Salary & Price



Insights for the above scatter plot reveals that as the salary of individual increases then price of the cars is also increasing. Hence, Price and Salary are positively correlated.

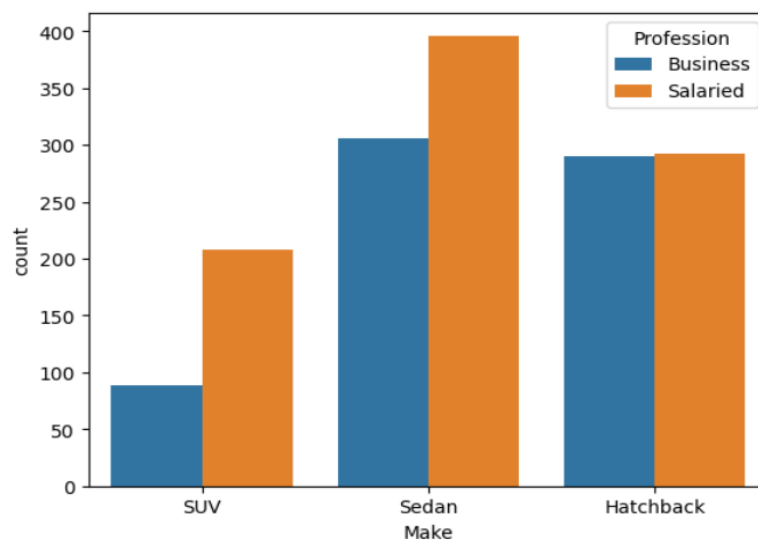
### Establishing co-relation between Age & Make



With reference to above 'Age' vs 'Make' graph we can conclude that

- Younger age group (20-30 ) they mostly buy Hatchback and Sedan as compared to SUV
- Mid age group (31-45) they mostly buy Sedan and SUV (few sales only) with no sales of hatchback to this age group.
- Older Age group (46-55) only buy SUV with no sales of Sedan or Hatchback.
- There is zero sales of Hatchback from the age group 31 to 46.

### Establishing co-relation between Professional & Gender



- Male Business professionals first choice is Hatch back and second is Sedan and comparatively less preferred is SUV.
- Whereas female Business Professions prefer to buy Sedan as well as SUV with similar interest in Make.
- Salaried Female customer first choice is SUV whereas second choice is Sedan and with fewer sales of hatchback amongst them.
- Salaried Male Customer first choice is either Sedan or Hatchback as compared to SUV, SUV is comparatively less demanding amongst them

Establishing co-relation with heatmap with Price, Age, No\_of\_Dependants, salary, Partner salary & Total salary



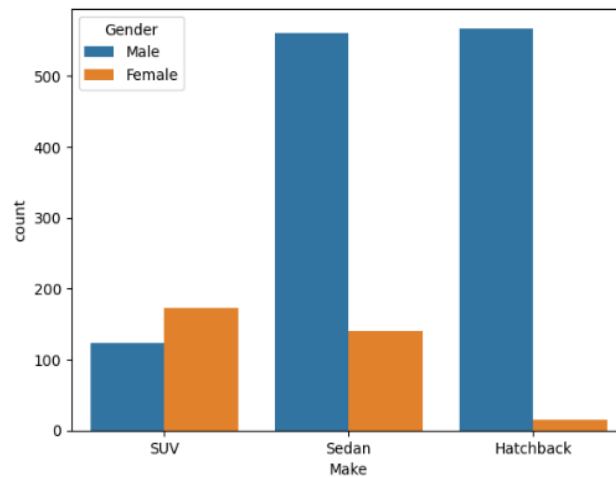
-1 to 0 = Negative correlations

0 to 1 = Positive correlation

- There is a Strong correlation between Age and Price.
- There is a strong correlation between Partner salary and Total salary
- There is medium correlation between Salary and Price
- There is medium correlation between Total salary and Price

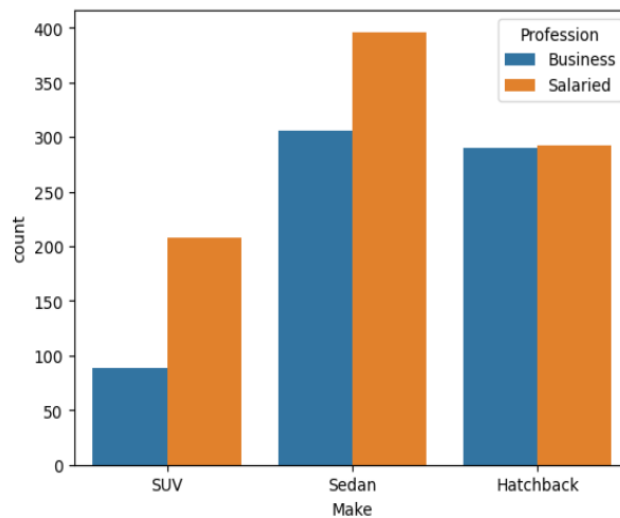
## D.Exploring the data

1. Do men tend to prefer SUVs more compared to women?



No, the following countplot shows that in reality Women prefer SUV more compared to Men.

2. What is the likelihood of a salaried person buying a Sedan?



Yes, the count plot above salaried person buys more Sedan than Business professionals.

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?

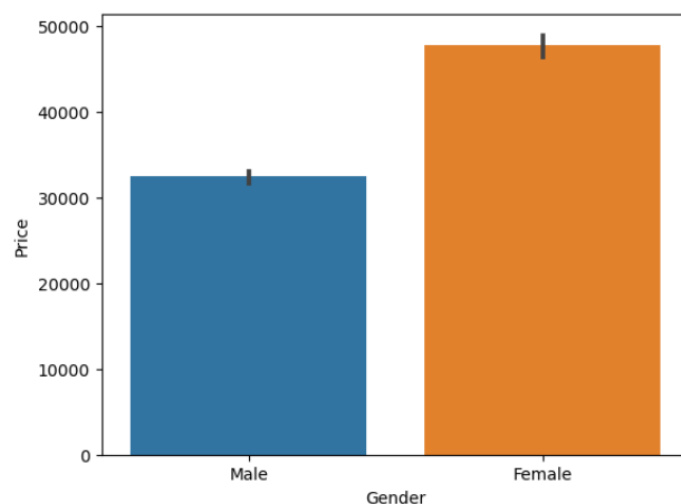


| Gender | Profession | Make      |     |
|--------|------------|-----------|-----|
| Female | Business   | SUV       | 55  |
|        |            | Sedan     | 50  |
|        | Salaried   | SUV       | 118 |
|        |            | Sedan     | 91  |
| Male   | Business   | Hatchback | 15  |
|        |            | Hatchback | 290 |
|        |            | Sedan     | 256 |
|        | Salaried   | SUV       | 34  |
|        |            | Sedan     | 305 |
|        |            | Hatchback | 277 |
|        |            | SUV       | 90  |

Name: Make, dtype: int64

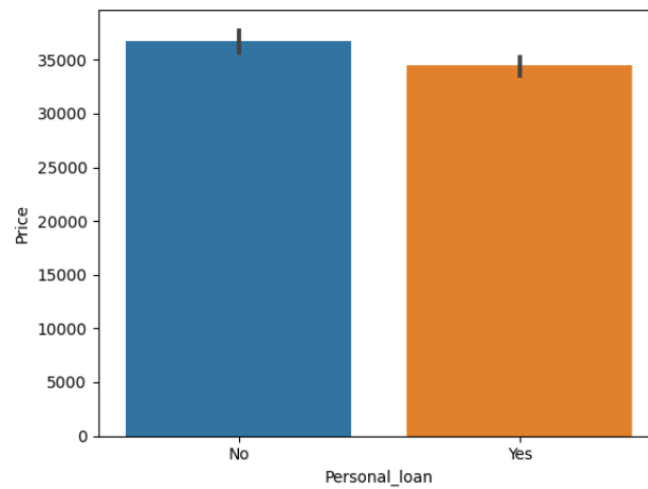
No, Sheldon Cooper is wrong according to data above, their preference is to buy Hatchback and Sedan ratherthan SUV. SUV is less preferred amongst salaried male.

4. How does the the amount spent on purchasing automobiles vary by gender?



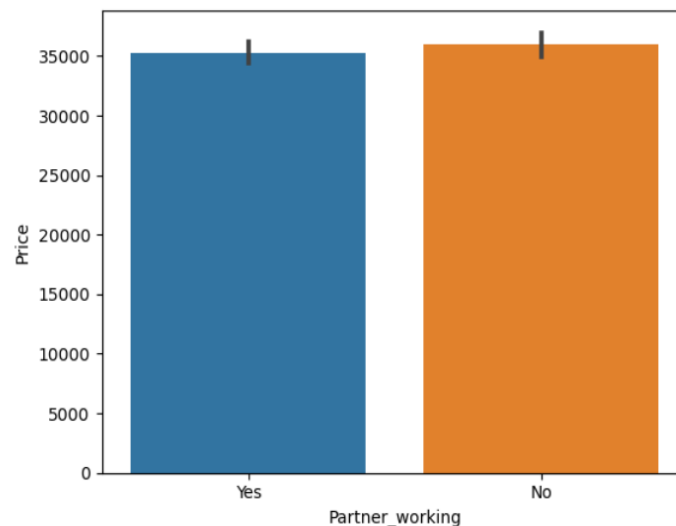
As per above calculation Gender vs Price Mean/Average, we can say Female has bought altogether moreexpensive car than Male .  
 From the above plot, customer who don't take loan buy more expensive cars than who avails Personal loan.

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



From the above plot, customer who don't take loan buy more expensive cars than who avails Personal loan.

6. How does having a working partner influence the purchase of higher-priced cars?



With the above graph we can conclude that it doesn't matter if partner is working or not working Customer will buy their preferred car, there is although a marginal difference between them which slightly shows that customers whose partner is not working buy more expensive cars.

E.

### Actionable Insights & Recommendations

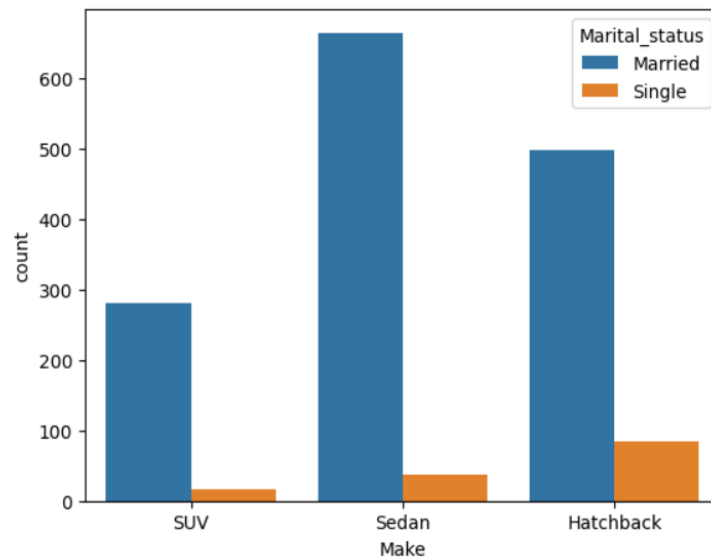
As per below crosstab information with Gender aspect , we can conclude that total Male customer buys more cars with highest number of Hatchback followed by Sedan on second number and SUV takes third position in buying preference.

For Female customer, they buy more SUV as compared to Sedan, Hatchback takes last position in the buying preference list for Females.

| Make   | Hatchback | SUV | Sedan | All  |
|--------|-----------|-----|-------|------|
| Gender |           |     |       |      |
| Female | 15        | 173 | 141   | 329  |
| Male   | 567       | 124 | 561   | 1252 |
| All    | 582       | 297 | 702   | 1581 |

Insights for the below data values

- Zero sales amongst female with business professionals for the Hatchback, Their first preference is to buy SUV and second choice is Sedan.
- Where as Salaried females first choice is also SUV followed by Sedan with few females prefer to buy Hatchback.
- Male Business professionals prefers to buy Hatchback followed by Sedan and fewer choices of SUV.
- Salaried Male prefer to buy Sedan followed by Hatchback and SUV takes third position for the choice.
- Married customers buys more Sedan as compared to Hatchback and SUV becomes the last choice.
- Single customer buys more Hatchback as compared to Sedan and SUV takes last position for the choice



| Make           | Hatchback | SUV | Sedan | All  |
|----------------|-----------|-----|-------|------|
| Marital_status |           |     |       |      |
| Married        | 498       | 281 | 664   | 1443 |
| Single         | 84        | 16  | 38    | 138  |
| All            | 582       | 297 | 702   | 1581 |

As per below data and Graph below insights can be derived.

There are total 1443 married and 138 Singles , henceforth there are more married customers in the company record

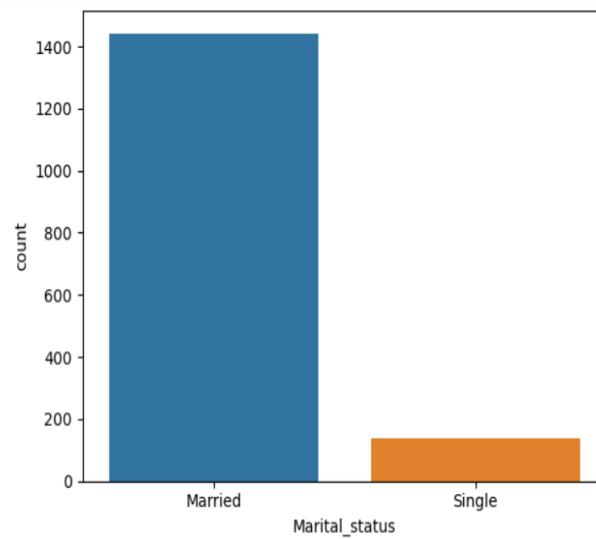
Married business professional they prefer to buy Sedan followed by Hatchback and SUV became last choice for them.

Single business professional tends to buy more Hatchback than Sedan with very few prefers to buy SUV.

Salaried and Married prefers to buy more Sedan than Hatchback and SUV is the last choice for them.

Salaried and single prefers to buy Hatchback followed by Sedan with fewer choice for SUV.

```
Married 1443
Single 138
Name: Marital_status, dtype: int64
```



```
Profession Make Gender
Business Hatchback Male 290
 SUV Female 55
 Male 34
 Sedan Male 256
 Female 50
Salaried Hatchback Male 277
 Female 15
 SUV Female 118
 Male 90
 Sedan Male 305
 Female 91
Name: Gender, dtype: int64
```

## Problem 2 - Framing Analytics Problem

Analyse the dataset and list down the top 5 important variables, along with the business justifications.

- Data Frame Provided ----- godigt\_cc\_data
- There are 8 object type, 19 integer type and datetime type variable in the data set.
- There are no duplicates in the data
- There are Data Rows : 8448 ; Data Columns : 28
- There are total 38 null values in 'Transactor\_revolve'

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8448 entries, 0 to 8447
Data columns (total 28 columns):
Column Non-Null Count Dtype
--- -
0 userid 8448 non-null int64
1 card_no 8448 non-null object
2 card_bin_no 8448 non-null int64
3 Issuer 8448 non-null object
4 card_type 8448 non-null object
5 card_source_date 8448 non-null datetime64[ns]
6 high_networth 8448 non-null object
7 active_30 8448 non-null int64
8 active_60 8448 non-null int64
9 active_90 8448 non-null int64
10 cc_active30 8448 non-null int64
11 cc_active60 8448 non-null int64
12 cc_active90 8448 non-null int64
13 hotlist_flag 8448 non-null object
14 widget_products 8448 non-null int64
15 engagement_products 8448 non-null int64
16 annual_income_at_source 8448 non-null int64
17 other_bank_cc_holding 8448 non-null object
18 bank_vintage 8448 non-null int64
19 T+1_month_activity 8448 non-null int64
20 T+2_month_activity 8448 non-null int64
21 T+3_month_activity 8448 non-null int64
22 T+6_month_activity 8448 non-null int64
23 T+12_month_activity 8448 non-null int64
24 Transactor_revolver 8410 non-null object
25 avg_spends_l3m 8448 non-null int64
26 Occupation_at_source 8448 non-null object
27 cc_limit 8448 non-null int64
dtypes: datetime64[ns](1), int64(19), object(8)
memory usage: 1.8+ MB
```

- There are total 38 null values in 'Transactor\_revolve'.

```
userid 0
card_no 0
card_bin_no 0
Issuer 0
card_type 0
card_source_date 0
high_networth 0
active_30 0
active_60 0
active_90 0
cc_active30 0
cc_active60 0
cc_active90 0
hotlist_flag 0
widget_products 0
engagement_products 0
annual_income_at_source 0
other_bank_cc_holding 0
bank_vintage 0
T+1_month_activity 0
T+2_month_activity 0
T+3_month_activity 0
T+6_month_activity 0
T+12_month_activity 0
Transactor_revolver 38
avg_spends_13m 0
Occupation_at_source 0
cc_limit 0
dtype: int64
```

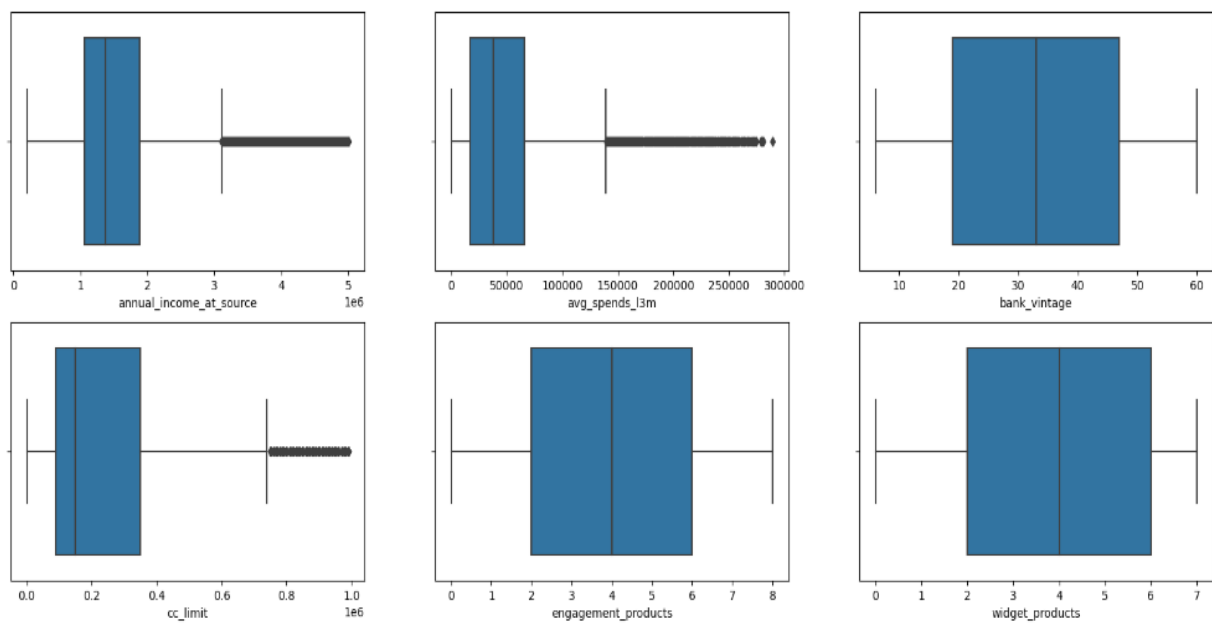
- Treating Transactor\_revolver, this will be done by replacing the missing values with the mode of 'Transactor\_revolver'.
- The mode of 'Transactor\_revolver' is 'T', replacing the missing values with T.
- After treatment there is no null values

```
userid 0
card_no 0
card_bin_no 0
Issuer 0
card_type 0
card_source_date 0
high_networth 0
active_30 0
active_60 0
active_90 0
cc_active30 0
cc_active60 0
cc_active90 0
hotlist_flag 0
widget_products 0
engagement_products 0
annual_income_at_source 0
other_bank_cc_holding 0
bank_vintage 0
T+1_month_activity 0
T+2_month_activity 0
T+3_month_activity 0
T+6_month_activity 0
T+12_month_activity 0
Transactor_revolver 0
avg_spends_13m 0
Occupation_at_source 0
cc_limit 0
dtype: int64
```

## Statistical analysis of the data

|                         | count  | mean         | std          | min      | 25%        | 50%       | 75%        | max       |
|-------------------------|--------|--------------|--------------|----------|------------|-----------|------------|-----------|
| userid                  | 8448.0 | 4.224500e+03 | 2.438872e+03 | 1.0      | 2112.75    | 4224.5    | 6336.25    | 8448.0    |
| card_bin_no             | 8448.0 | 4.367470e+05 | 3.048975e+04 | 376916.0 | 426241.00  | 437551.0  | 438439.00  | 524178.0  |
| active_30               | 8448.0 | 2.923769e-01 | 4.548815e-01 | 0.0      | 0.00       | 0.0       | 1.00       | 1.0       |
| active_60               | 8448.0 | 4.947917e-01 | 5.000025e-01 | 0.0      | 0.00       | 0.0       | 1.00       | 1.0       |
| active_90               | 8448.0 | 6.420455e-01 | 4.794271e-01 | 0.0      | 0.00       | 1.0       | 1.00       | 1.0       |
| cc_active30             | 8448.0 | 2.840909e-01 | 4.510070e-01 | 0.0      | 0.00       | 0.0       | 1.00       | 1.0       |
| cc_active60             | 8448.0 | 4.844934e-01 | 4.997891e-01 | 0.0      | 0.00       | 0.0       | 1.00       | 1.0       |
| cc_active90             | 8448.0 | 6.323390e-01 | 4.821970e-01 | 0.0      | 0.00       | 1.0       | 1.00       | 1.0       |
| widget_products         | 8448.0 | 3.614583e+00 | 2.273193e+00 | 0.0      | 2.00       | 4.0       | 6.00       | 7.0       |
| engagement_products     | 8448.0 | 3.991122e+00 | 2.572135e+00 | 0.0      | 2.00       | 4.0       | 6.00       | 8.0       |
| annual_income_at_source | 8448.0 | 1.674595e+06 | 1.064307e+06 | 200095.0 | 1061104.00 | 1372133.5 | 1881734.25 | 4999508.0 |
| bank_vintage            | 8448.0 | 3.316418e+01 | 1.586834e+01 | 6.0      | 19.00      | 33.0      | 47.00      | 60.0      |
| T+1_month_activity      | 8448.0 | 1.112689e-01 | 3.144835e-01 | 0.0      | 0.00       | 0.0       | 0.00       | 1.0       |
| T+2_month_activity      | 8448.0 | 4.794034e-02 | 2.136527e-01 | 0.0      | 0.00       | 0.0       | 0.00       | 1.0       |
| T+3_month_activity      | 8448.0 | 8.037405e-02 | 2.718875e-01 | 0.0      | 0.00       | 0.0       | 0.00       | 1.0       |
| T+6_month_activity      | 8448.0 | 8.877841e-03 | 9.380867e-02 | 0.0      | 0.00       | 0.0       | 0.00       | 1.0       |
| T+12_month_activity     | 8448.0 | 9.469697e-03 | 9.685625e-02 | 0.0      | 0.00       | 0.0       | 0.00       | 1.0       |
| avg_spends_l3m          | 8448.0 | 4.952737e+04 | 4.624495e+04 | 0.0      | 17110.00   | 37943.0   | 66095.75   | 289292.0  |
| cc_limit                | 8448.0 | 2.517069e+05 | 2.291149e+05 | 0.0      | 90000.00   | 150000.0  | 350000.00  | 990000.0  |

## Checking outliers in the dataset





- There are outliers in 'annual income at source', 'avg\_spends\_13m', 'cc\_limit'.
- Outliers will not be treated as it will impact the analysis.

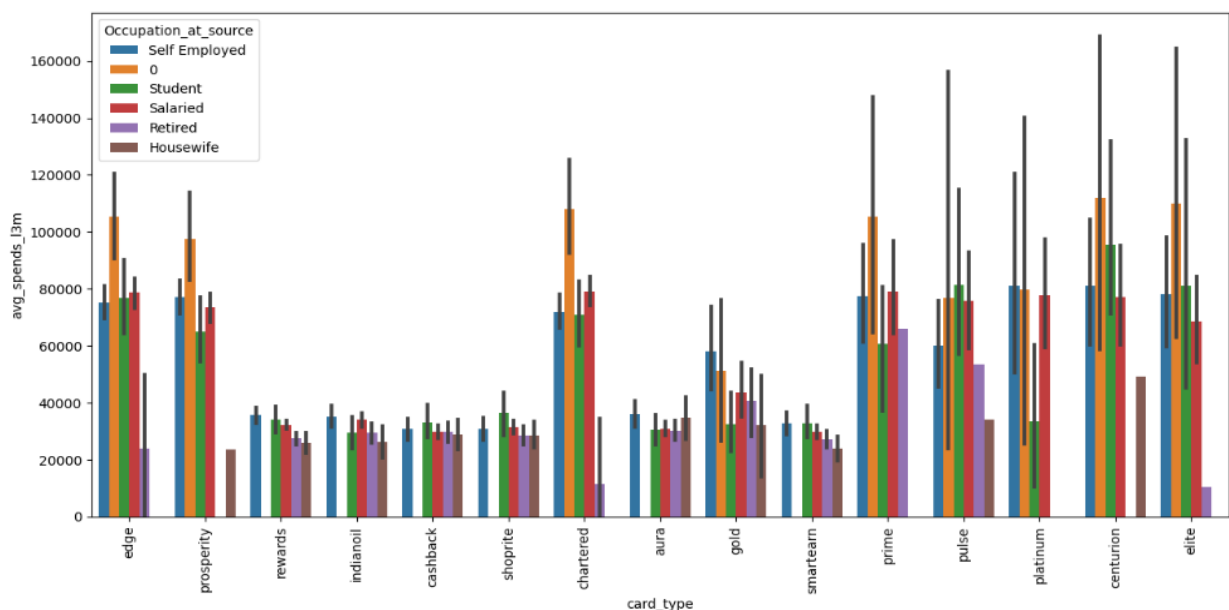
Framing few analytics problem which bank can include in their strategies.

- What is the relation between active user ID and total spend amount?
- Customer holding multiple bank credit card and their spending analysis to understand which bank cards are more preferred by the customers and what might be the reason?
- Does credit card limit restrict customer on spending and do they have any correlations.?
- What is the relationship between customers Annual income and Average spending?
- Does spending increase with the increase in Annual income?
- Spending pattern for different customers with different occupations?
- Which card types should be incorporated as per the type of customers? and it can be done by analysing and evaluating past records.

Most important variables from the data

### 1.Card\_type

Card types are the most important variable. Based on above data Salaried customer and high profile customer prefers to have more reward cards, Card\_types helps to increase the avg spent.



Card usage based on occupations

Housewife – 384

Retired-1089

Salaried -417

Self- employed – 2175

Students– 621

## 2.Occupation\_at\_source

Bank should strategize to sell the correct card\_type as per the occupations which will increase the avg spends and will eventually increase the profits.

- Highest CC Customers are self Employed and salaried and also their avg spent is high.
- Students prefer to have Reward CC

## 3.High\_networth

Banks accumulate profits thru interest earned on Average spent so this variable is very important.

High networth customers spent significantly which reciprocate as high profits to the bank.

## 4.Transactor Revolver

High transaction customers needs to be targeted for the revolver payments as customer can easily pay the amount and not being the defaulters. From the data given we can say Self-employed, salaried and students opt for revolver payment.

This variable is important as high transactor ,avg spending is more and uses revolver to do the payment which is beneficial for the bank to generate the profits.

## 5.Avg\_spends\_13

Avg spends is very important variable as this will give window to the bank for generating profits. These are directly proportional to the card\_type, high networth, occupation\_at\_source and Transactor\_revolver.

This is also proportional to the interest received, more spending means more interest and thus becomes the mode for revenue generation to the bank.