Business Report

By – Ayush Kumar

This Business report contains 2 data sets.

Case 1 – Automobile - Analyse and identify the purchasing patterns of various customers that influence the price of different segment of cars.

- The dataset provide to us has 1581 customers and their information such as age, salary, gender, profession etc.
- Our objective is to find which factors influence the total revenue and who are our target customers, what steps can be taken to increase the revenue.

Case 2 – Godigt Bank Data – Customer details like income, card type, hot list flag, net-worth category etc. Find relevant patterns of all variables.

- The data set contains 8448 rows and 28 columns.
- There are many categorical data, we have to find all the relevant patterns when compared with annual income, average spending and credit card limit.
- Identify defaulter customers and reduce the card limit and increase card limit of high earning customers.

Contents

Automobile Data Analysis

Objective:

Try to generate relevant insights to understand the purchase pattern of customers belonging to diverse field and category based on various attributes. The dataset given to us is Automobile data which contains 14 columns and 1581 rows.

Data dictionary:

Variable	Data Description	Variable Data Type
Age	The age of the individual in years.	Continuous
Gender	The gender of the individual, categorized as male or female.	Categorical
Profession	The occupation or profession of the individual. (Business or Salaried)	Categorical
Marital_status	The marital status of the individual, such as married &, single	Categorical
Education	The educational qualification of the individual Graduate and Post Graduate	Categorical
No_of_Dependents	The number of dependents (e.g., children, elderly parents) that the individual supports financially. (0,1,2,3,4)	Categorical
Personal_loan	A binary variable indicating whether the individual has taken a personal loan "Yes" or "No"	Categorical
House_loan	A binary variable indicating whether the individual has taken a housing loan "Yes" or "No"	Categorical

Partner_working	A binary variable indicating whether the individual's partner is employed "Yes" or "No"	Categorical
Salary	The individual's salary or income.	Continuous
Partner_salary	The salary or income of the individual's partner, if applicable.	Continuous
Total_salary	The total combined salary of the individual and their partner (if applicable).	Continuous
Price	The price of a product or service	Continuous
Make	The type of automobile (SUV, Sedan, Hatchback)	Categorical

The top variables are as follows:

1. Age
2. Marital Status
3. House Loan
4. Total Salary
5. Price
6. Make

Top Questions that can be asked initially:

- ✓ Which Make (Car type) is the highest selling?
- ✓ What is the total revenue and individual Make (Car type) is generating how much revenue?
- ✓ Explore variables like No of dependents and partner working?
- ✓ Which car type price is highest and who is the customer? (Gender, Age)
- ✓ Which age group customers are purchasing most cars?
- ✓ Does marital status influence the purchasing pattern of cars?

- ✓ Does Education of customers have any relevance in the purchase pattern?
- ✓ Does House loans and personal loans customers are buying more cars?
- ✓ Does partner working has any impact on purchasing cars?
- ✓ Which profession of customers are mostly purchasing cars and what type of car is preferred?
- ✓ What is the relevance of Total salary with other variables?
- ✓ What is the output of customers who has taken both loans? (House and personal)
- ✓ What is the output of customers who has not taken any loan?
- ✓ What can be done to increase the revenue of the company?

Note:

- **We created a new column Age Category and divided age in 4 groups. Age 20-30, 30-40,40-50 and 50-60. This will increase readability and make our analysis more precise.**
- **We have combined House loan and personal loan to check how many customers have taken both the loans. Does this effect and influence the purchasing pattern.**

Null Values and Irregularities treatment

There are some irregularities and null values in the data set which is highlighted below:

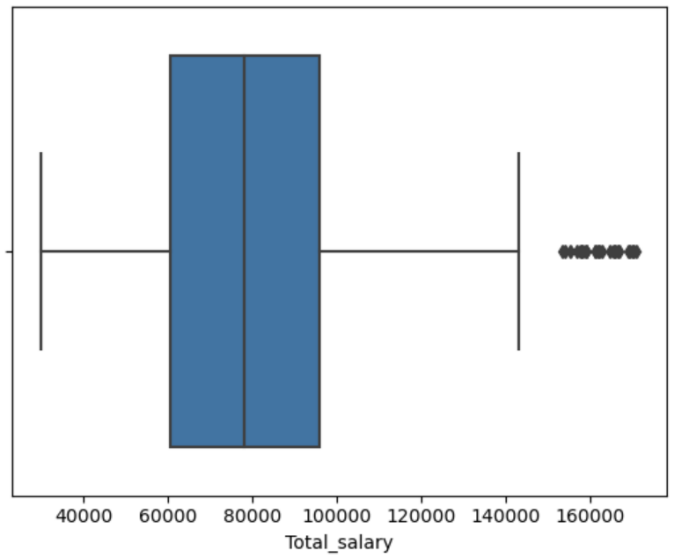
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1581 entries, 0 to 1580
Data columns (total 19 columns):
 #   Column              Non-Null Count  Dtype
---  ------              --------------  -----
 0   Age                 1581 non-null   int64
 1   Gender              1528 non-null   object
 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
 14  House               1581 non-null   int32
 15  Personal            1581 non-null   int32
 16  House_and_personal  1581 non-null   int32
 17  Both_loans          1581 non-null   object
 18  No_loans            1581 non-null   object
dtypes: float64(1), int32(3), int64(5), object(10)
memory usage: 216.3+ KB
```

Gender is a categorical variable and Partner salary is a Continuous variable. Below is the screenshot after resolving errors in both the columns

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1581 entries, 0 to 1580
Data columns (total 17 columns):
 #   Column            Non-Null Count  Dtype
---  ------            --------------  -----
 0   Age               1581 non-null   int64
 1   Gender            1581 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    1581 non-null   float64
 11  Total_salary      1581 non-null   int64
 12  Price             1581 non-null   int64
 13  Make              1581 non-null   object
 14  Both_loans        1581 non-null   object
 15  No_loans          1581 non-null   object
 16  Age_category      1581 non-null   object
dtypes: float64(1), int64(5), object(11)
memory usage: 210.1+ KB
```

Outlier treatment

After checking all the numerical variables, in column **Total Salary** there are some outliers. We can see there is a huge difference between min and max value which is resulting in the outlier, below is the boxplot graph and Total Salary summary:



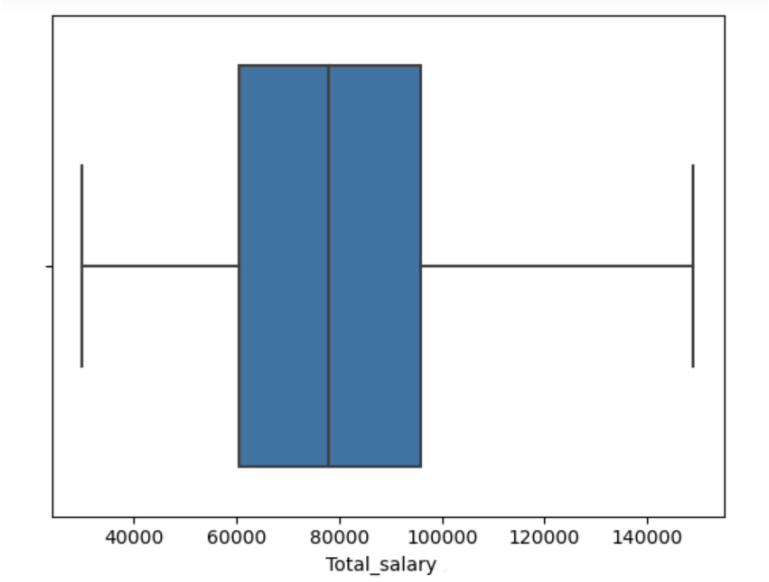
count	1581.000000
mean	79625.996205
std	25545.857768
min	30000.000000
25%	60500.000000
50%	78000.000000
75%	95900.000000
max	171000.000000

Below is the data summary of all the numerical variables **before treating the outlier**:

Data description before removing outlier:

	Age	No_of_Dependents	Salary	Partner_salary	Total_salary	Price
count	1581.000000	1581.000000	1581.000000	1475.000000	1581.000000	1581.000000
mean	31.922201	2.457938	60392.220114	20225.559322	79625.996205	35597.722960
std	8.425978	0.943483	14674.825044	19573.149277	25545.857768	13633.636545
min	22.000000	0.000000	30000.000000	0.000000	30000.000000	18000.000000
25%	25.000000	2.000000	51900.000000	0.000000	60500.000000	25000.000000
50%	29.000000	2.000000	59500.000000	25600.000000	78000.000000	31000.000000
75%	38.000000	3.000000	71800.000000	38300.000000	95900.000000	47000.000000
max	54.000000	4.000000	99300.000000	80500.000000	171000.000000	70000.000000

We will use the **Box-Plot method** to resolve the outliers, now we can see that the difference between the min and max has reduced however the median remains the same as shown in the graph and summary below:



count	1581.000000
mean	79398.545225
std	24849.147996
min	30000.000000
25%	60500.000000
50%	78000.000000
75%	95900.000000
max	149000.000000

Below is the data summary of all the numerical variables **after treating the outlier**:

Data description after removing outlier:

	Age	No_of_Dependents	Salary	Partner_salary	Price	Total_salary
count	1581.000000	1581.000000	1581.000000	1581.000000	1581.000000	1581.000000
mean	31.922201	2.457938	60392.220114	20585.895003	35597.722960	79398.545225
std	8.425978	0.943483	14674.825044	18952.938643	13633.636545	24849.147996
min	22.000000	0.000000	30000.000000	0.000000	18000.000000	30000.000000
25%	25.000000	2.000000	51900.000000	0.000000	25000.000000	60500.000000
50%	29.000000	2.000000	59500.000000	25600.000000	31000.000000	78000.000000
75%	38.000000	3.000000	71800.000000	38000.000000	47000.000000	95900.000000
max	54.000000	4.000000	99300.000000	80500.000000	70000.000000	149000.000000

Q.1 Which Make (Car type) is the highest selling? (Uni-Variate)

```
Sedan        0.444023
Hatchback    0.368121
SUV          0.187856
Name: Make, dtype: float64
```

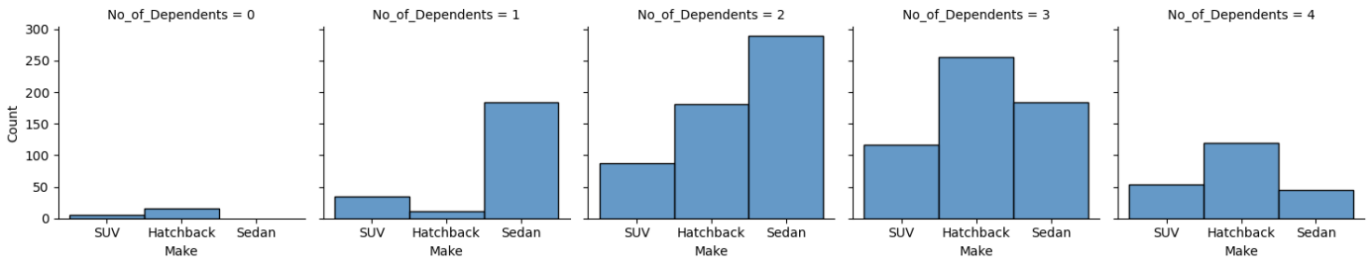
In terms of count, Sedan is the highest and contributes at 44% of entire sales of car.

Q.2 What is the total revenue and individual Make (Car type) is generating how much revenue? (Uni-Variate)

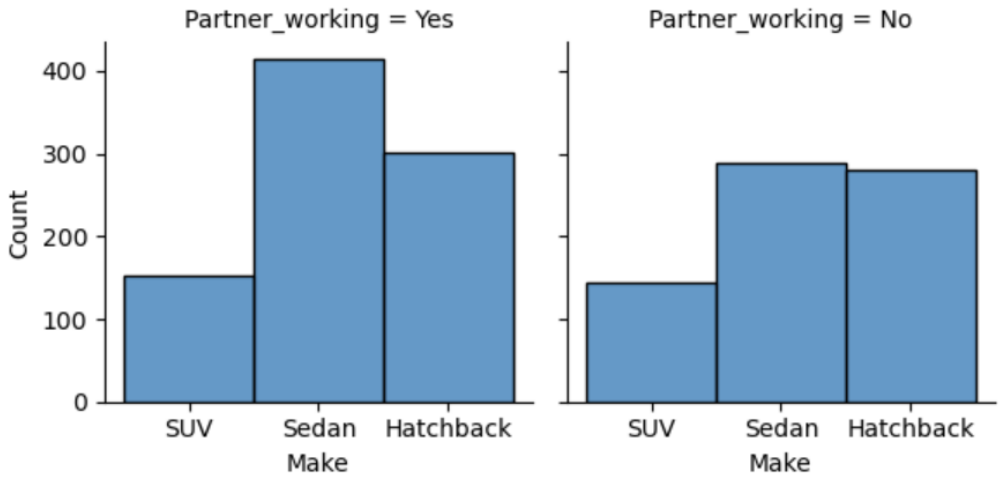
```
Make
Sedan        55868500
Hatchback    41332500
SUV          28687700
Name: Total_salary, dtype: int64
```

Sedan generates highest revenue with 43% of the entire revenue.

Q.3 Explore variables like No of dependents and partner working? (Bi-Variate)

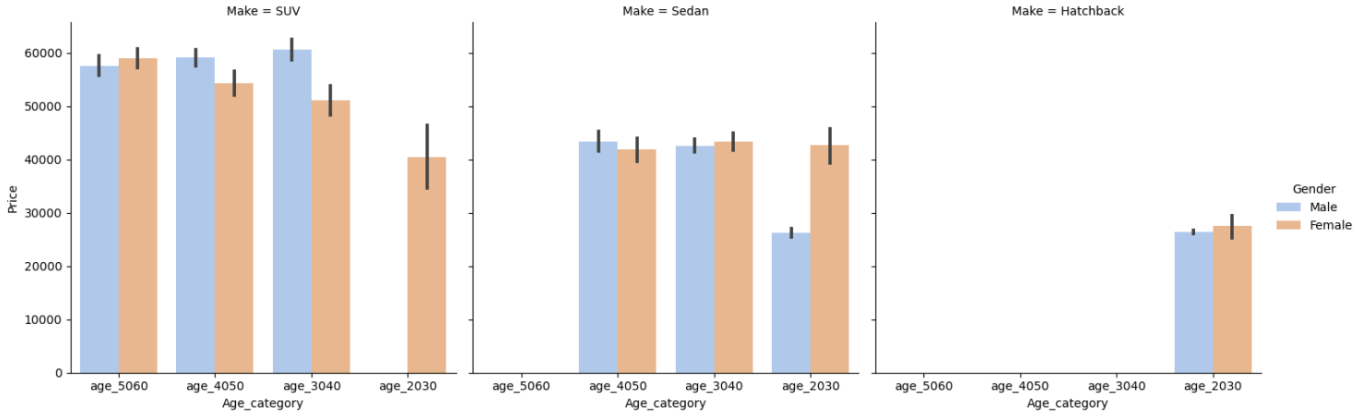


Customers who have dependents equal to 2 or 3 are purchasing more cars and sedan is the most preferred choice amongst all.



Working partners purchase more cars and in that also sedans are mostly preferred by all.

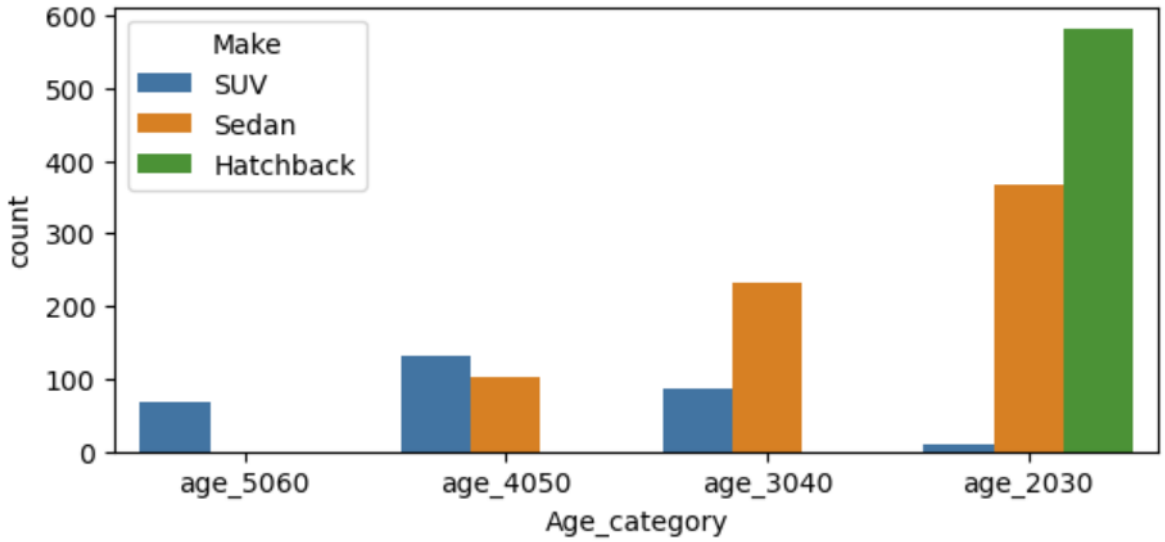
Q.4 Which car type price is highest and who is the customer? (Gender, Age) (Multi-Variate)



Below are some observations:

- SUVs are mostly preferred by all age categories.
- Females of all age category are very active in purchasing cars, SUVs is the most diverse amongst all age groups.
- Sedans are preferred by all age group except age group between 50-60 – It mean old customers are mostly purchasing SUVs.
- Hatchback are least preferred and only young customers of age between 20-30 are going after Hatchback.

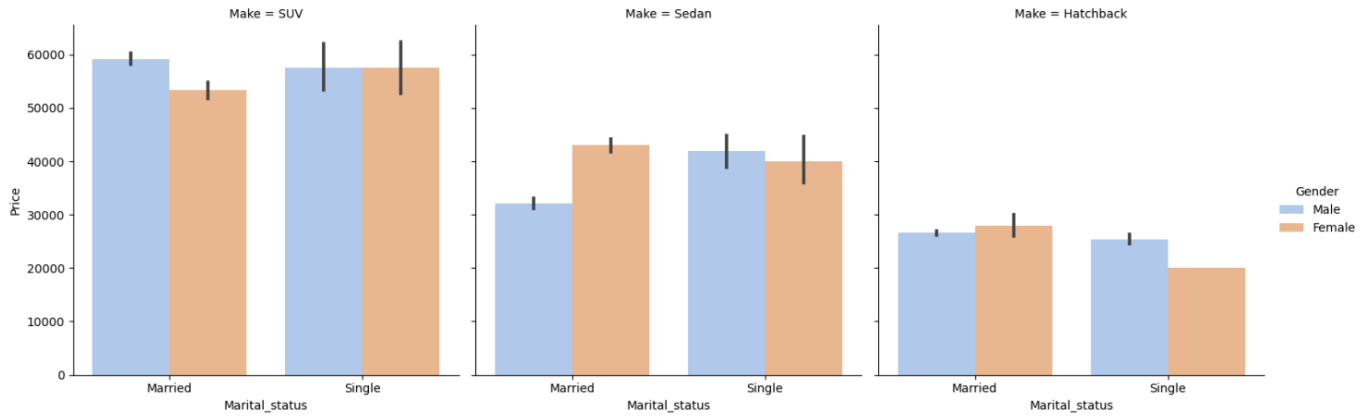
Q.5 Which age group customers are purchasing most cars? (Bi-Variate)




```
age_2030    0.605946
age_3040    0.201771
age_4050    0.148640
age_5060    0.043643
Name: Age_category, dtype: float64
```

Age group between 20-30 are purchasing 60% cars and the count of sedans is highest amongst all.

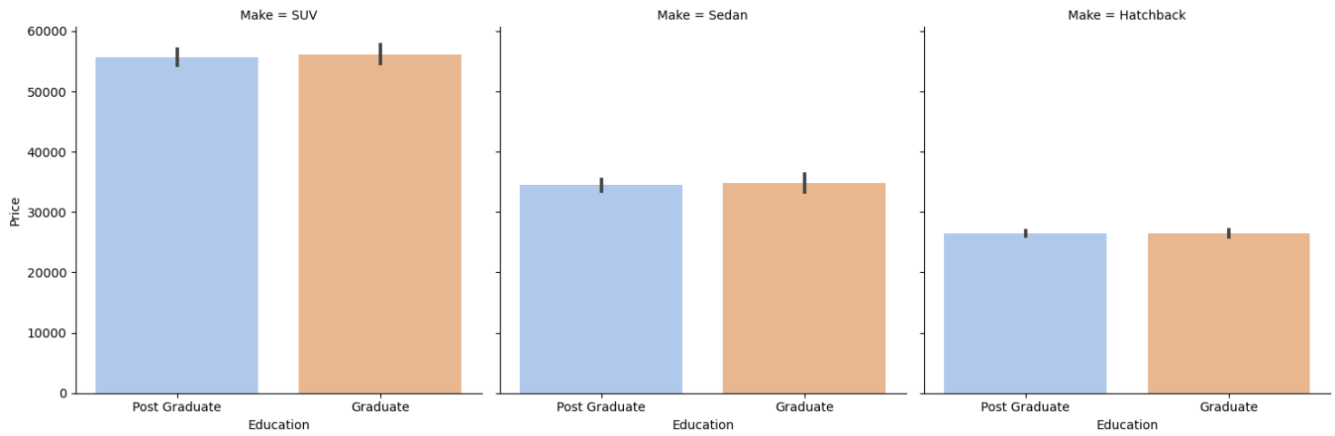
Q.6 Does marital status influence the purchasing pattern of cars? (Multi-Variate)



			sum	count
Make	Marital_status	Gender		
Sedan	Married	Male	17253000	537
Hatchback	Married	Male	12886000	484
SUV	Married	Female	8849000	166
Sedan	Married	Female	5470000	127
SUV	Married	Male	6810000	115
Hatchback	Single	Male	2110000	83
Sedan	Single	Male	1008000	24
Hatchback	Married	Female	392000	14
Sedan	Single	Female	561000	14
SUV	Single	Male	518000	9
		Female	403000	7
Hatchback	Single	Female	20000	1

91% of married customers are purchasing cars. In married customers 34% married males are purchasing Sedans, 31% married males are purchasing Hatchbacks and 10% married females are purchasing SUVs. – So, we can say females are mostly interested in purchasing SUVs rather than males.

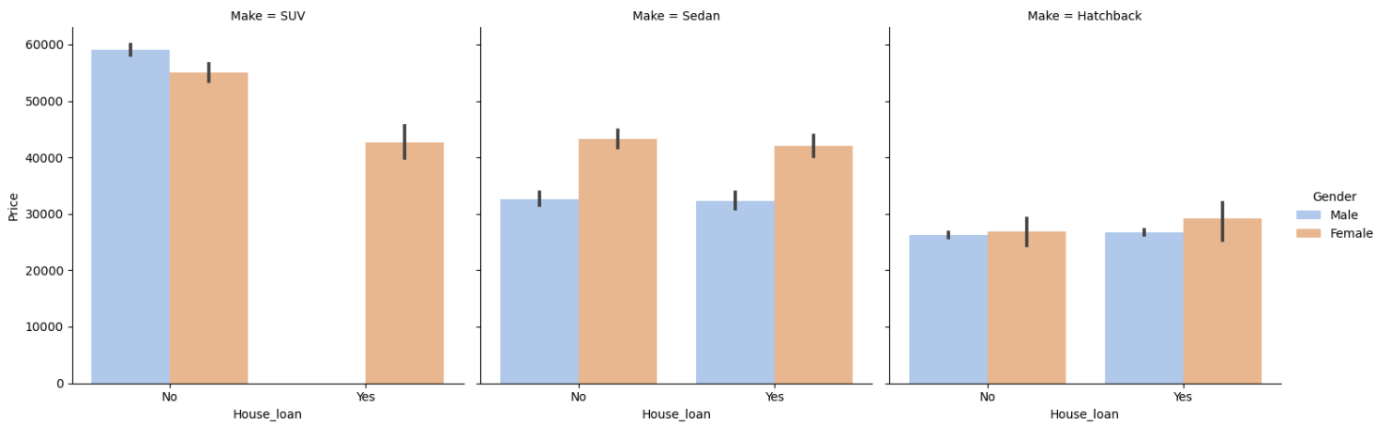
Q.7 Does Education of customers have any relevance in the purchase pattern? (Multi-Variate)



		sum	count
Make	**Education**		
Sedan	**Post Graduate**	15278000	443
Hatchback	**Post Graduate**	9594000	362
Sedan	**Graduate**	9014000	259
Hatchback	**Graduate**	5814000	220
SUV	**Post Graduate**	10007000	180
	Graduate	6573000	117

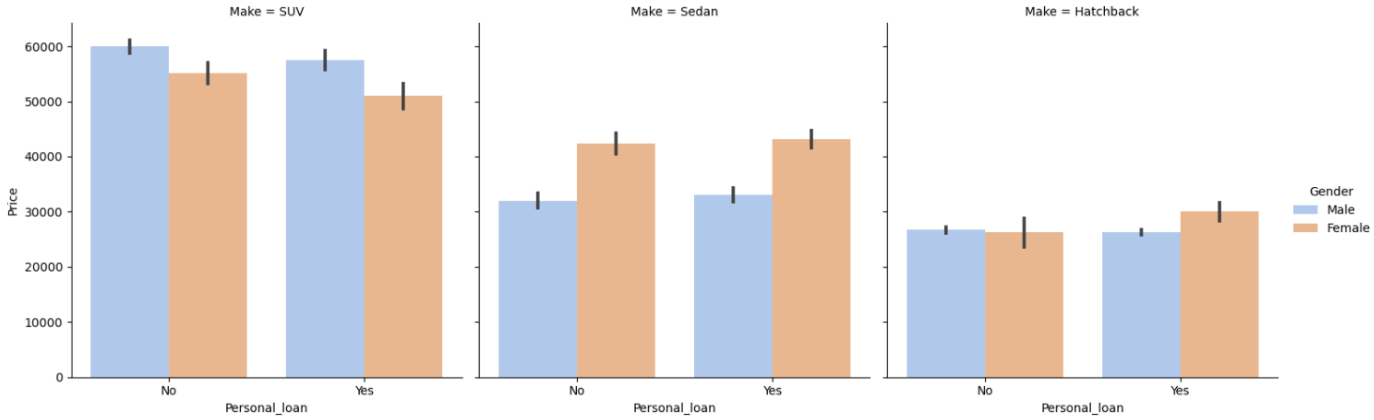
By looking at the graph we can say that most of the Education doesn't show any significant pattern but when we look at the table the numbers shows that sedans and hatchbacks are mostly purchased by post graduate customers with 28% and 23% respectively.

Q.8 Does House loans and personal loans customers are buying more cars? (Multi-Variate)



		sum	count
Make	**House_loan**		
SUV	**No**	15643000	275
Sedan	**No**	14982000	432
	Yes	9310000	270
Hatchback	**No**	9119000	347
	Yes	6289000	235
SUV	**Yes**	937000	22

Males who have taken house loan are not purchasing more SUVs. For sedans and hatchbacks females who have taken house loan show more interest than males.

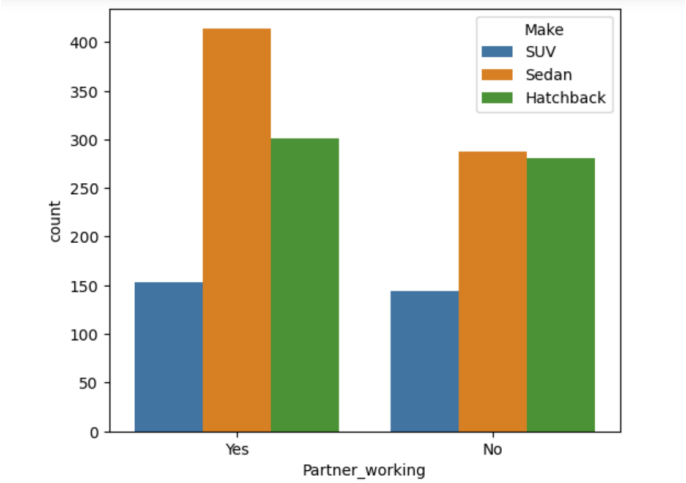


Personal loans don't show much impact when compared with Home loan. Males who have taken personal loans are showing more interest in SUVs but when compared with home loans males are not purchasing SUVs. Sedan and hatchback pattern is similar for both House and personal loan.

		sum	count
Make	**Personal_loan**		
Sedan	**Yes**	13440000	385
	No	10852000	317
Hatchback	**No**	7765000	291
	Yes	7643000	291
SUV	**No**	10373000	181
	Yes	6207000	116

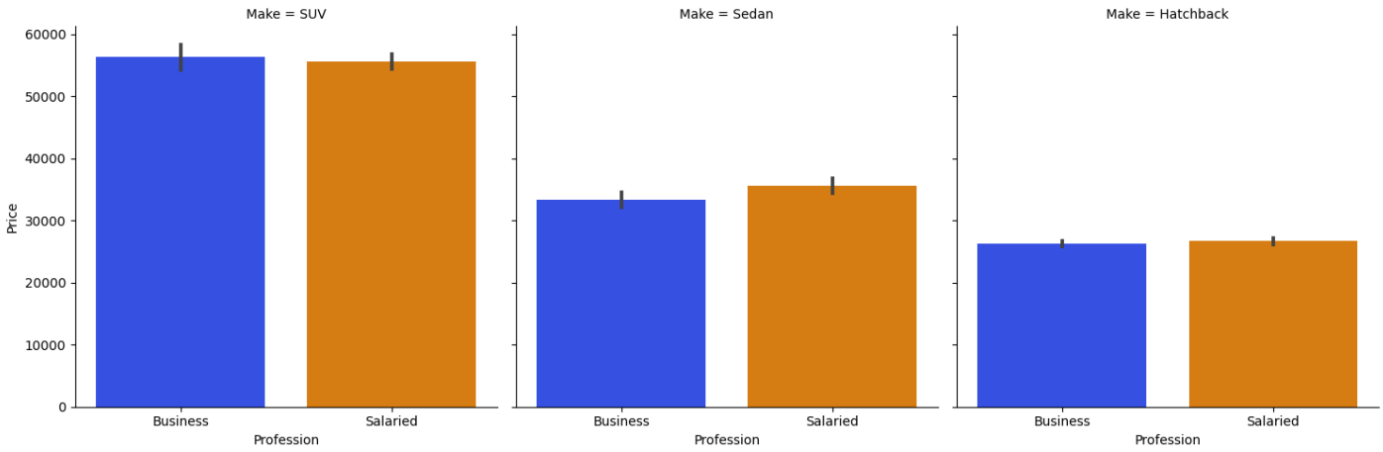
Customers who have taken personal loans are going after Sedans and in terms of money and quantity sedan is the highest. **It means that most of the customers are purchasing sedans on personal loans.**

Q9. Does partner working has any impact on purchasing cars? (Bi-Variate)



Clearly working partners purchase more cars because they contribute more in terms of revenue, also they are purchasing more sedans. But for SUVs partner working doesn't show much impact.

Q10. Which profession of customers are mostly purchasing cars and what type of car is preferred?

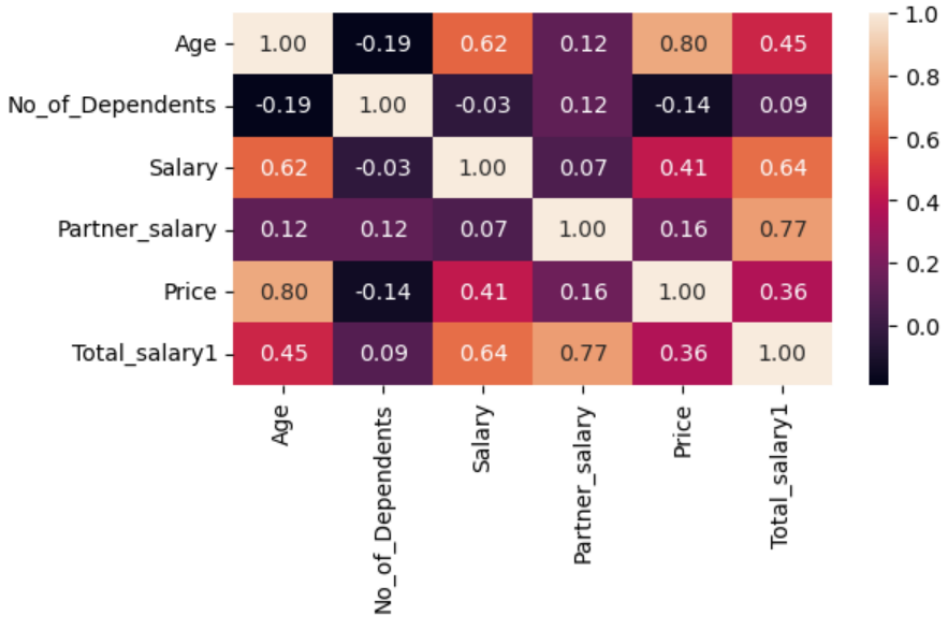


Profession weather it is business or salaried doesn't show much impact on the purchase pattern.

		sum	count
Make	**Profession**		
Sedan	**Salaried**	14092000	396
	Business	10200000	306
Hatchback	**Salaried**	7779000	292
	Business	7629000	290
SUV	**Salaried**	11562000	208
	Business	5018000	89

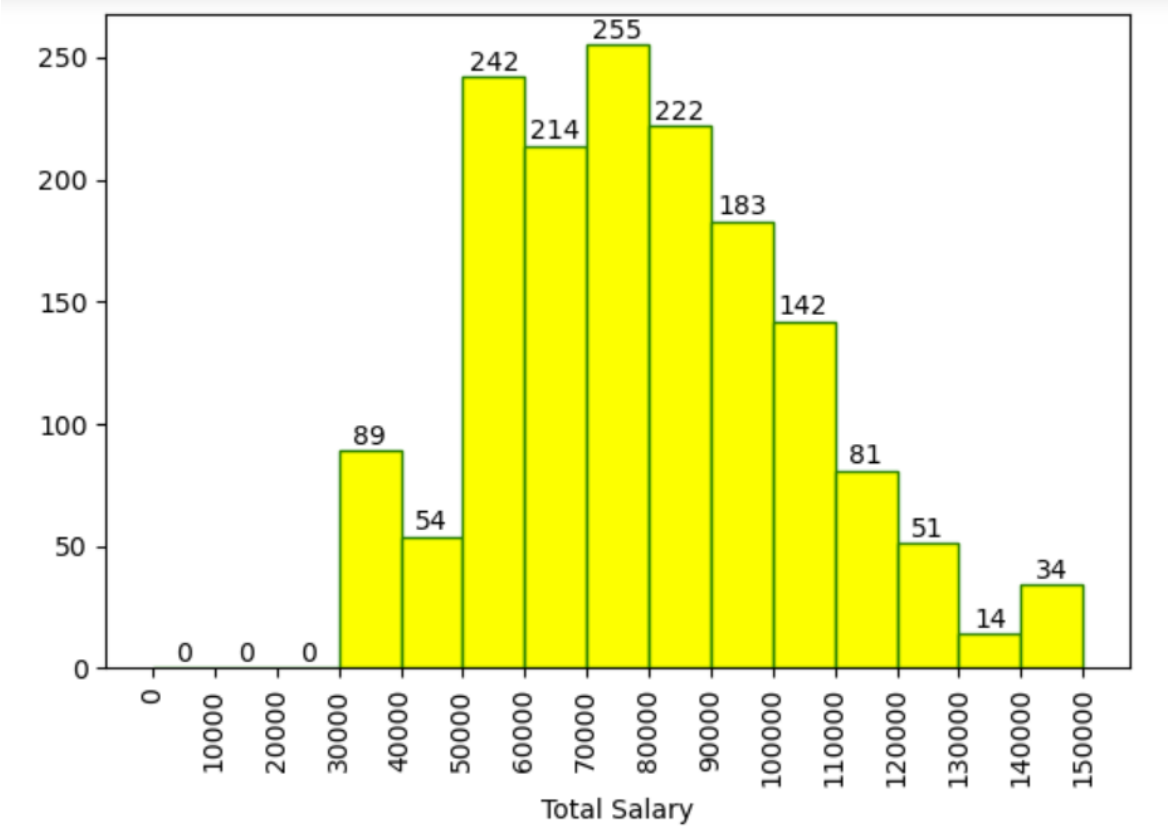
But if we dive deep then salaried customers are mostly purchasing sedans. 25% of salaried customers are purchasing sedans.

Q11. What is the relevance of Total salary with other variables? (Multi-Variate)



	Age	No_of_Dependents	Salary	Partner_salary	Price	Total_salary
Age	1.000000	-0.189614	0.616899	0.121187	0.797831	0.452844
No_of_Dependents	-0.189614	1.000000	-0.031746	0.121555	-0.135839	0.087606
Salary	0.616899	-0.031746	1.000000	0.065348	0.409920	0.638625
Partner_salary	0.121187	0.121555	0.065348	1.000000	0.161136	0.765147
Price	0.797831	-0.135839	0.409920	0.161136	1.000000	0.359651
Total_salary	0.452844	0.087606	0.638625	0.765147	0.359651	1.000000

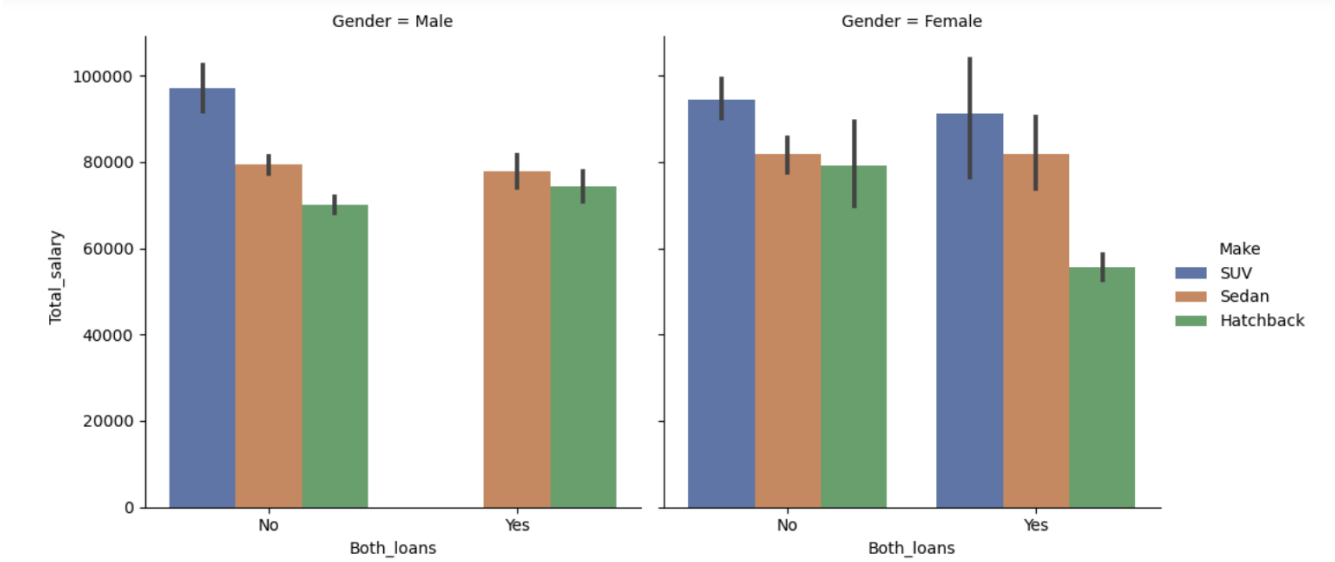
Total Salary of customer has 36% corelation with price of car. If the partner is working then partner salary contributes in the total salary and has a good corelation of 77%.



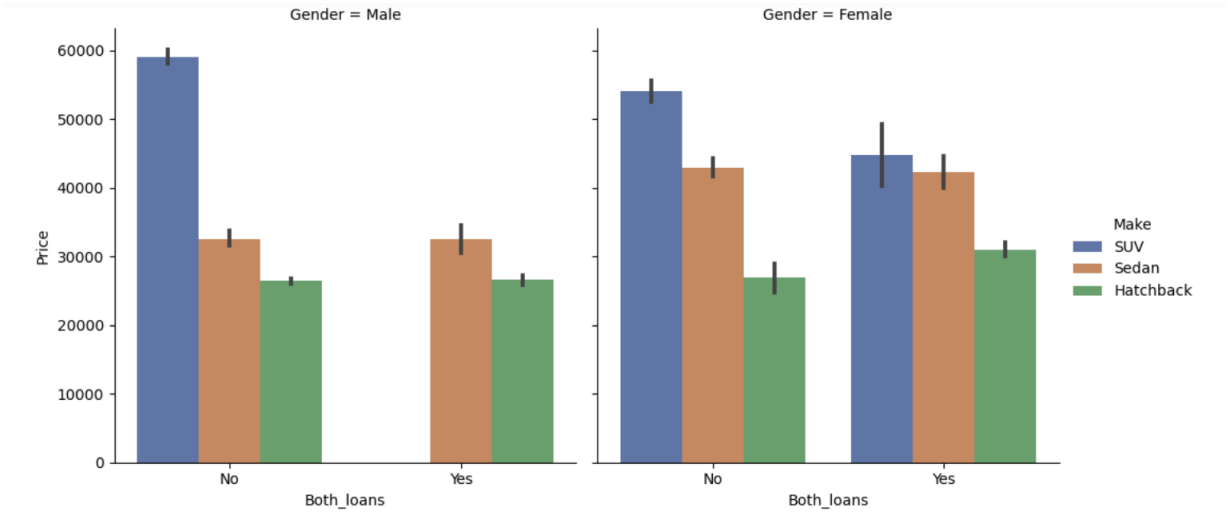
71 % employee total salary is between the 50000-100000 and 16% employee total salary falls between 70000-80000.

Q12. What is the output of customers who *has taken both loans*? (House and personal) (Multi-Variate)

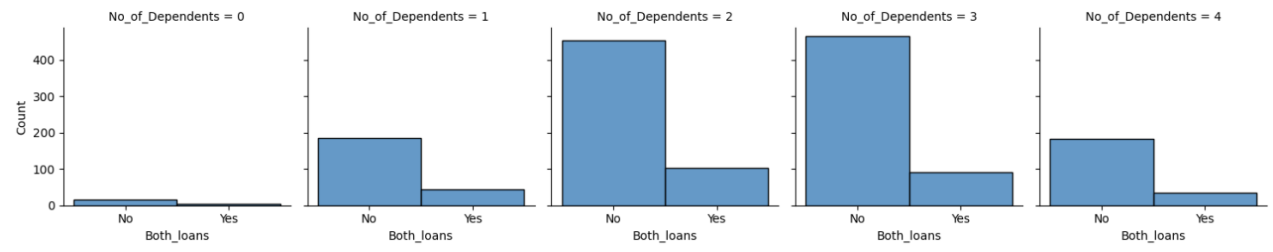
When compared with total salary, males who have taken both loans are not buying SUVs only Sedans and Hatchbacks are preferred but women of high salary with both loans are purchasing SUVs also. **(Personal loan doesn't have much impact when compared with house loan)**



In terms of car price also, same pattern as above males with both loans are not buying SUVs only Sedans and Hatchbacks but women with both loans are buying expensive SUVs.

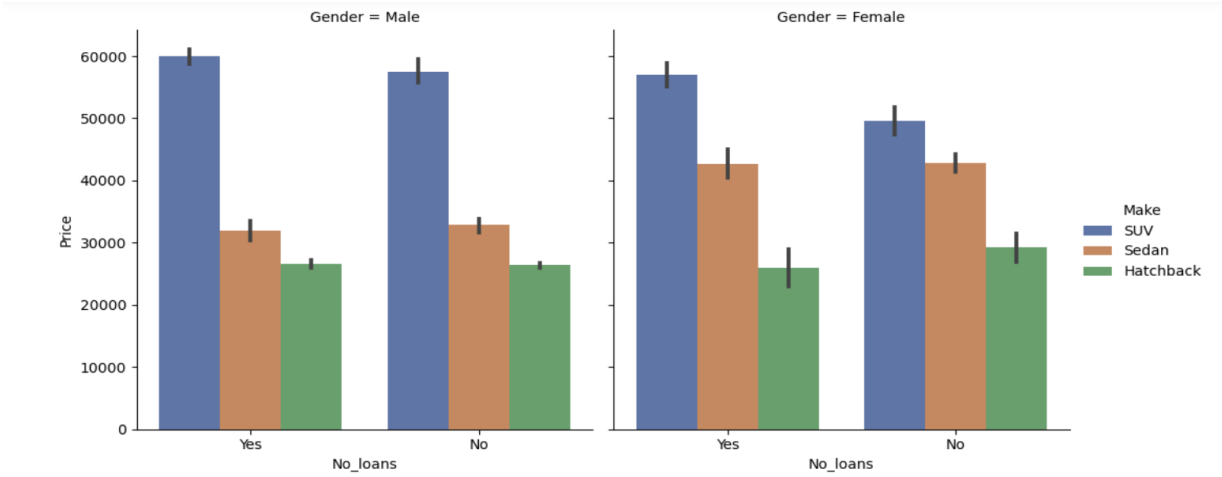


Customers with 2 and 3 dependents are purchasing more cars and have also taken both loans.

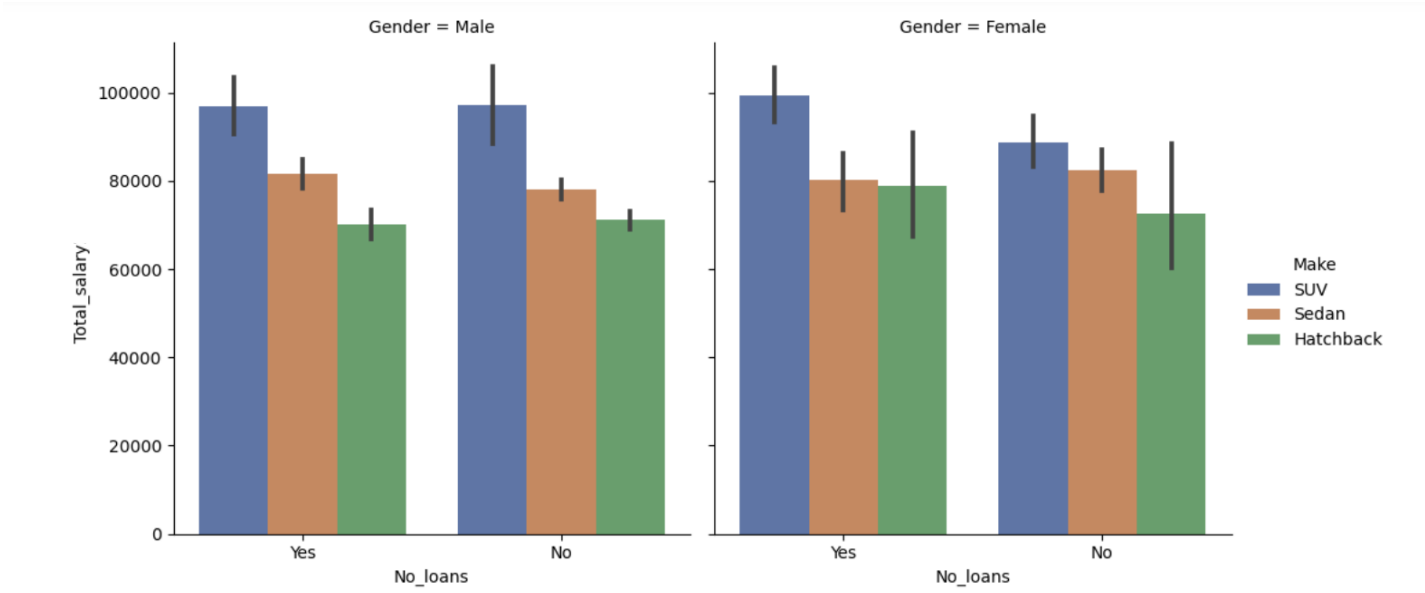


Q13. What is the output of customers who has *not taken any loan*? (House and personal) (Multi-Variate)

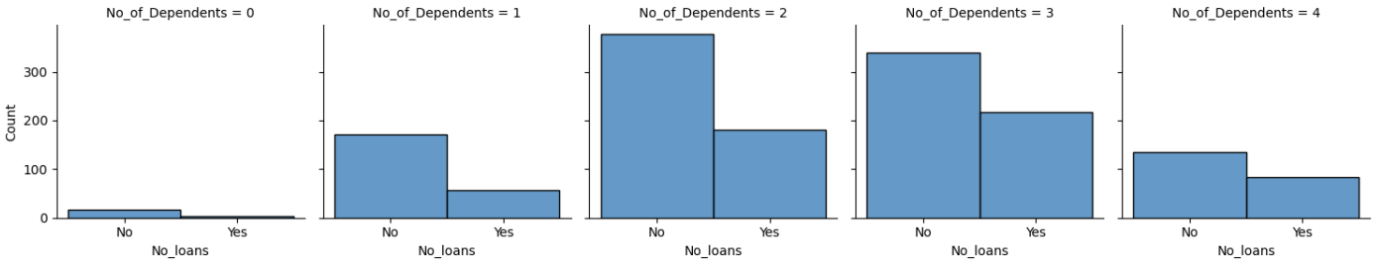
Below is the graph for customers who has not taken any kind of loan in terms of car price.



Below is the graph for customers who has not taken any kind of loan in terms of Total Salary.



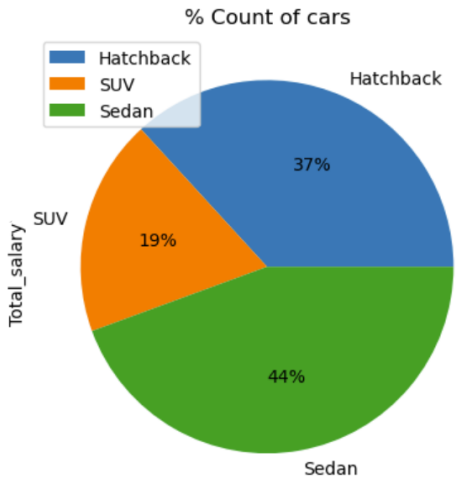
Customers with 2 and 3 dependents are purchasing more cars who has not taken any loans.



Q.14 What can be done to increase the revenue of the company?

Before commenting first let's analyse the different data points below:

1. The Count of different car type. (Sedan, SUV, Hatchback)

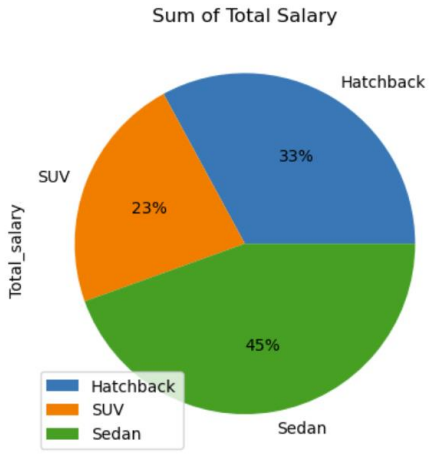



```
Make
Sedan        702
Hatchback    582
SUV          297
Name: Price, dtype: int64
```

We can say that 44% purchase is of sedans which is the highest in the category.

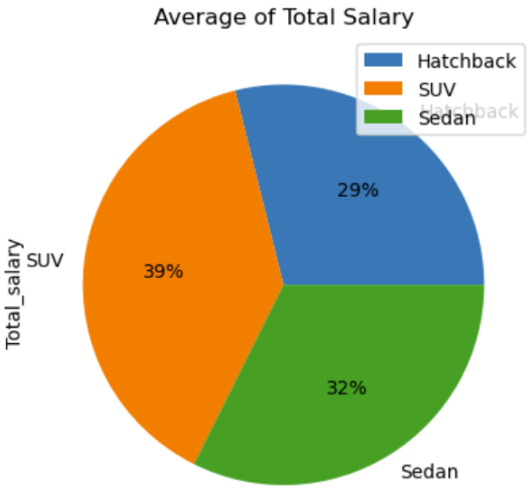
2. Sum and average Total salary of customers.

Below is the **Sum of total salary** of customers and sedan is contributing the most about 45%.




```
Make
Sedan          55868500.0
Hatchback      41332500.0
SUV            28328100.0
Name: Total_salary , dtype: float64
```

Below is the **average (mean) total salary** of customers and here SUVs is contributing more because SUVs are expensive.

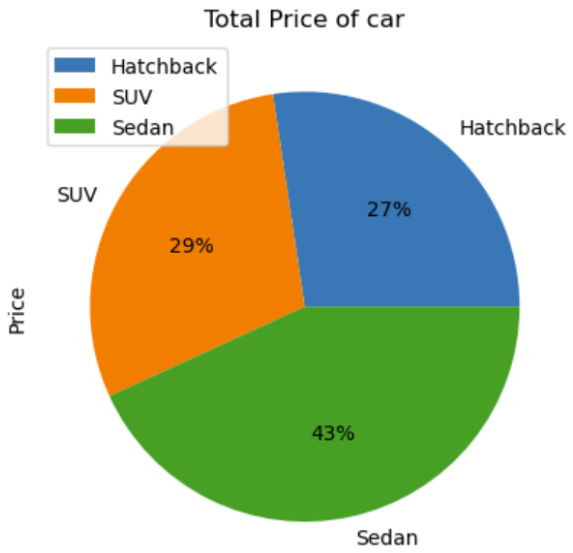



```
Make
SUV            95380.808081
Sedan          79584.757835
Hatchback      71018.041237
Name: Total_salary , dtype: float64
```

3. Sum and average price of different car Make. (Sedan, SUV, Hatchback)

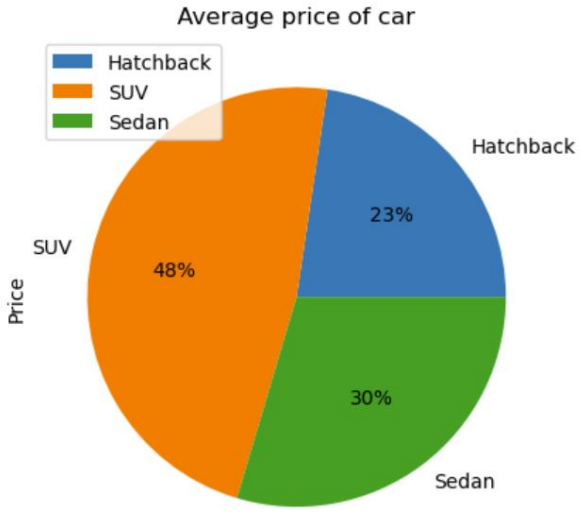
Below is the pie chart of **car total price** and sedan is contributing the most around 43% of the total revenue.

```
Make
Sedan        24292000
SUV          16580000
Hatchback    15408000
Name: Price, dtype: int64
```



Below pie chart shows **average car price** and here SUVs is the highest amongst all and contribute around 48%

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



The above cohorts say that, some customers who are buying sedans and hatchbacks have high salary and can be pitched to buy SUVs which will increase the revenue. Now we will identify which customers to pitch.

Target customers to pitch:

1. **Let's consider customers of salary 1,00,000 and above to be rich customers, the count of these customers is 322 out of 1581 which is around 20%.**

2. **Out of those 20% rich customers 9% customers are buying sedans and 4% are buying hatchbacks. We should target women because mostly women show more interest in purchasing SUVs.**

3. **If we target these rich customers to buy SUVs then our revenue will increase.**

What Steps can we take to target the rich customers to increase the SUVs sales:

1. **Give customers spot offers discount for first time visitors.**

2. **Give customers first 2 car servicing free if they go for SUVs.**

3. **Give them some offers in bank loans if applicable.**

4. **Organise an event regarding SUVs purchase and put some offers like spin-the-wheel and give them some gift vouchers to increase sales.**

Summary & recommendations:

1. Sedan contributes 44% of the entire sales and 43% of the total revenue.
2. Sedans are preferred by all age group except age group between 50-60, old customers are mostly purchasing SUVs.
3. 20-30 age group customers are purchasing 60% cars.
4. In terms of marital status as married 34% Males are purchasing Sedans, 31% males are purchasing Hatchbacks. 10% females are purchasing SUVs and also females are mostly interested in purchasing SUVs.
5. Males who have taken house loan are not purchasing SUVs. For sedans and hatchbacks females who have taken house loan show more interest than males.
6. Customers who have taken personal loans are going after Sedans and in terms of money and quantity sedan is the highest. It means that most of the customers are purchasing sedans on personal loans.
7. Working partners purchase more cars because they contribute more in terms of revenue, also they are purchasing more sedans. But for SUVs partner working doesn't show much impact.
8. Profession weather it is business or salaried doesn't show much impact on the purchase pattern.
9. 57% employee total salary is between the 50000-100000 and 16% employee total salary falls between 70000-80000.

GODIGT Bank Data Analysis

Objective:

In Godigt bank data find out relevant customer pattern in all categorical variables, how to increase the bank revenue, which customers to target, which customer to avoid and reduce their cc limit also what steps can be taken to avoid bad customers in future.

Data dictionary:

Variable	Data Description	Variable Data Type
Userid	Unique bank customer-id	Continuous
Card_no	Masked credit card number	Continuous
Card_bin_no	Credit card IIN number	Continuous
Issuer	Card network issuer	Categorical
Card_type	Credit card type	Categorical
Card_source_date	Credit card sourcing date	Date Time
High_networth	Customer category based on their net-worth value (A: High to E: Low)	Categorical
Active_30	Savings/Current/Salary etc. account activity in last 30 days	Categorical
Active_60	Savings/Current/Salary etc. account activity in last 60 days	Categorical
Active_90	Savings/Current/Salary etc. account activity in last 90 days	Categorical
Cc_active_30	Credit Card activity in the last 30 days	Categorical
Cc_active_60	Credit Card activity in the last 60 days	Categorical
Cc_active_90	Credit Card activity in the last 90 days	Categorical

Hotlist_flag	Whether card is hot-listed (Any problem noted on the card)	Categorical
Widget_products	Number of convenience products customer holds (dc, cc, net-banking active, mobile banking active, wallet active, etc.)	Categorical
Engagement_products	Number of investment/loan products the customer holds (FD, RD, Personal loan, auto loan)	Categorical
Annual_income_at_source	Annual income recorded in the credit card application	Continuous
Other_bank_cc_holding	Whether the customer holds another bank credit card	Categorical
Bank_vintage	Vintage with the bank (in months) as on Tthmonth	Continuous
T+1_month_activity	Whether customer uses credit card in T+1 month (future)	Categorical
T+2_month_activity	Whether customer uses credit card in T+2 month (future)	Categorical
T+3_month_activity	Whether customer uses credit card in T+3 month (future)	Categorical
T+6_month_activity	Whether customer uses credit card in T+6 month (future)	Categorical
T+12_month_activity	Whether customer uses credit card in T+12 month (future)	Categorical
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.	Categorical
Avg_spends_l3m	Average credit card spends in last 3 months	Continuous
Occupation_at_source	Occupation recorded at the time of credit card application	Categorical
Cc_limit	Current credit card limit	Continuous

The top variables are as follows:

1. High Net Worth
2. Hot List Flag
3. Annual Income
4. Average Credit card spend in 3 months
5. CC limit
6. Occupation
7. Card Source Date

Data Description:

There are 8448 rows and 28 columns, below is the data description of all the variables.

```
<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 3 numerical variables which are **“annual_income_at_source”, “avg_spends_l3m” and “cc_limit”**. Below is the summary of the numerical data and corelation between them.

Data Summary:

	annual_income_at_source	avg_spends_l3m	cc_limit
count	8.448000e+03	8448.000000	8448.000000
mean	1.674595e+06	49527.365530	251706.912879
std	1.064307e+06	46244.954836	229114.856385
min	2.000950e+05	0.000000	0.000000
25%	1.061104e+06	17110.000000	90000.000000
50%	1.372134e+06	37943.000000	150000.000000
75%	1.881734e+06	66095.750000	350000.000000
max	4.999508e+06	289292.000000	990000.000000

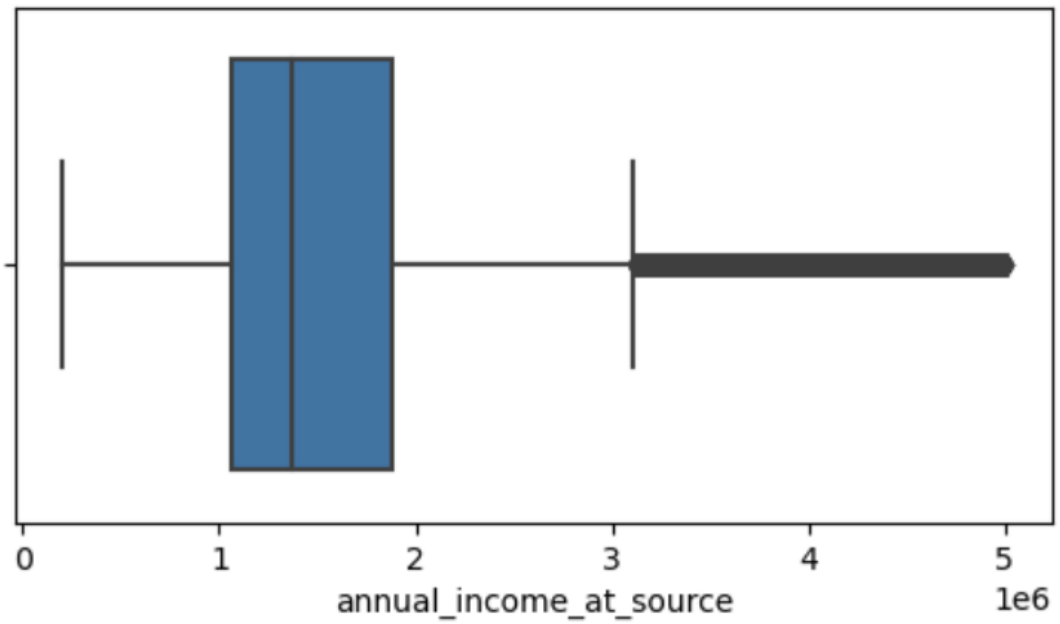
Data Corelation:

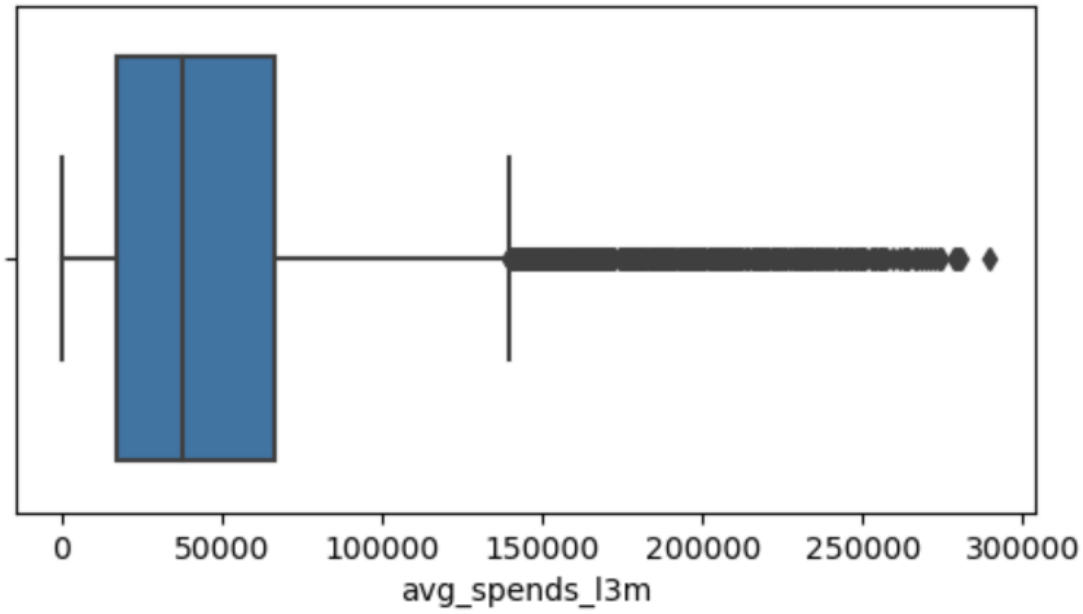
	annual_income_at_source	avg_spends_l3m	cc_limit
annual_income_at_source	1.000000	0.684695	0.771886
avg_spends_l3m	0.684695	1.000000	0.527270
cc_limit	0.771886	0.527270	1.000000

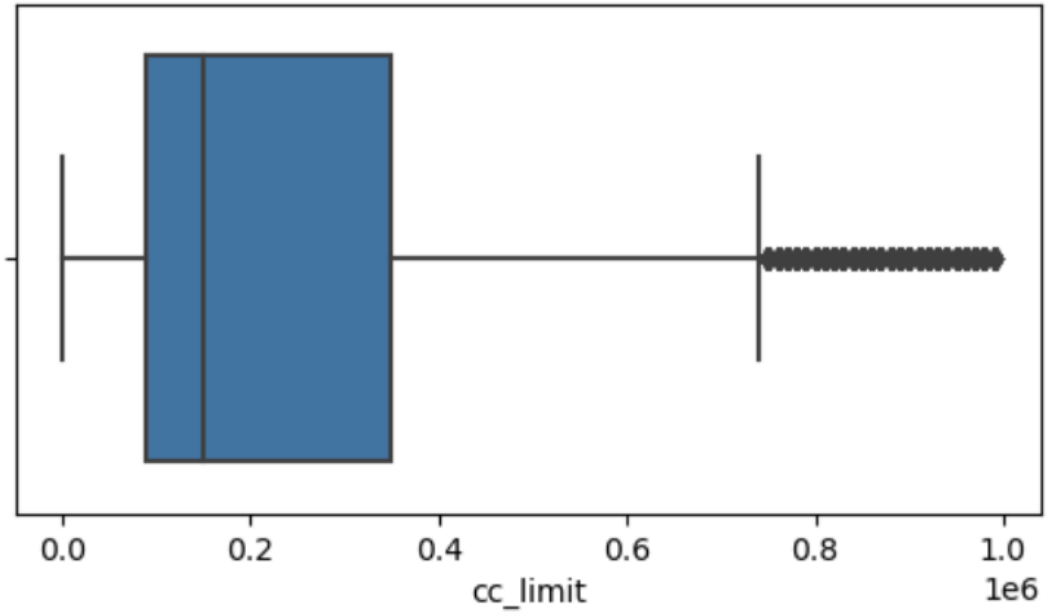
Outlier and null values identification

All 3 numerical columns **“annual_income_at_source”, “avg_spends_l3m” and “cc_limit”** contain outliers. We can use the boxplot method to treat the outliers. Below is the screenshot attached:

Note: Here we are only identifying the null values and outliers and not treating them.







In “Transactor_revolver” variable there are 38 null values, the count is 8410 rows. We can consider the mode to treat the null values. In the below screenshot the mode is “**Transactor**”.

Data summary of column “Transactor_revolver”

```
count     8410
unique       2
top          T
freq      7115
Name: Transactor_revolver, dtype: object
```

Value counts of column “Transactor_revolver”

```
T    7115
R    1295
Name: Transactor_revolver, dtype: int64
```

Note:

We can break the *date variable column into three columns like Day, month and year*. This will add more value in our analysis and we can dive deeper into the data.

Top Questions that can be asked initially:

- ✓ Which issuer does most transaction and what is the total annual income?
- ✓ Which card type is widely used and what is the total annual income?
- ✓ Which year the count and annual income are highest?
- ✓ Which year contributes the highest count of customer’s hotlist flag?
- ✓ How many customer’s salary accounts are active within 30 days, 60 days and 90 days?
- ✓ How many widget products are popular when compared with card type?
- ✓ Which card type has different engagement products?
- ✓ How many customers have accounts in other bank, are they high net worth customers?
- ✓ How many customer’s credit card are active within 30 days, 60 days and 90 days?
- ✓ Show count and annual income of T+1 month activity, T+2 month activity, T+3 month activity, T+6 month activity, T+12 month activity and compare which one is highest?
- ✓ Which occupation is highest in terms of count and annual income?

- ✓ Which card type and has more CC limit?
- ✓ Which customer comes most under hotlist flag?
- ✓ In which month the count and annual income is the highest?
- ✓ What is the count and annual income of high net-worth customers?
- ✓ How many high net-worth customers are under hotlist flag?
- ✓ What is the corelation between all variables?
- ✓ How many customers have not activated their savings account in past 90 days?
- ✓ How many customers have not done any transactions by their credit card in past 90 days?
- ✓ What is the relevance of bank vintage with other columns?
- ✓ What is the relevance of high net-worth customers with widget products?
- ✓ What is the relevance of high net-worth customers with engagement products?
- ✓ What is the average spending of high net-worth customers, in which year and month it is highest?
- ✓ What steps can be taken to increase the average spending of customers?
- ✓ How can we encourage and convert inactive customers to active customers?
- ✓ How many customers are transactor and how many are revolvers?
- ✓ What is the relevance transactor revolver column with other variables like hotlist flag, annual income, CC limit and high net-worth?
- ✓ What is the count of transactor and revolver column year and month wise?
- ✓ What is the average spending of customers based on Occupation and transactor and revolver variable?
- ✓ Can we say that customers having other bank accounts are less active and do less transaction?
- ✓ What is the relevance between occupation of customer and card type?

The End