Statistical Methods for Decision Making

BUSINESS REPORT

BY: - ABHISHEK K HIREMATH

Problem-1: Austo Motors

Austo Motor Company is a leading car manufacturer specializing in SUV, Sedan, and Hatchback models. In its recent board meeting, concerns were raised by the members on the efficiency of the marketing campaign currently being used. The board decides to rope in an analytic professional to improve the existing campaign.

Here is the dataset - Link

A.What is the important technical information about the dataset that the database administrator would be interested in? (Hint: Information about size of the dataset and the nature of the variables

- Size of Dataset: Dataset has 1581 rows and 14 columns.
- Data head & Data Tail: the dataset for quick reference.

	Age	Gender	Profession	Marital_status	Education	No_of_Dependents	Personal_loan	House_loan	Partner_working	Salary	Partner_salary	Total_salary	Price	Make
0	53	Male	Business	Married	Post Graduate	4	No	No	Yes	99300	70700.0	170000	61000	SUV
1	53	Femal	Salaried	Married	Post Graduate	4	Yes	No	Yes	95500	70300.0	165800	61000	SUV
2	53	Female	Salaried	Married	Post Graduate	3	No	No	Yes	97300	60700.0	158000	57000	SUV
3	53	Female	Salaried	Married	Graduate	2	Yes	No	Yes	72500	70300.0	142800	61000	SUV
4	53	Male	Salaried	Married	Post Graduate	3	No	No	Yes	79700	60200.0	139900	57000	SUV

Fig: -1

	Age	Gender	Profession	Marital_status	Education	No_of_Dependents	Personal_loan	House_loan	Partner_working	Salary	Partner_salary	Total_salary	Price	Make
1576	22	Male	Salaried	Single	Graduate	2	No	Yes	No	33300	0.0	33300	27000	Hatchback
1577	22	Male	Business	Married	Graduate	4	No	No	No	32000	NaN	32000	31000	Hatchback
1578	22	Male	Business	Single	Graduate	2	No	Yes	No	32900	0.0	32900	30000	Hatchback
1579	22	Male	Business	Married	Graduate	3	Yes	Yes	No	32200	NaN	32200	24000	Hatchback
1580	22	Male	Salaried	Married	Graduate	4	No	No	No	31600	0.0	31600	31000	Hatchback

Fig: -2

here are the some refrence of dataset.

- Dataset Information**:** There are 6 numerical and 8 categorical variables. The details of each:

```
<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
```

fig: -3

B.Take a critical look at the data and do a preliminary analysis of the variables. Do a quality check of the data so that the variables are consistent. Are there any discrepancies present in the data? If yes, perform preliminary treatment of data.

Inspecting Missing Values: There are Null records present in two variables: Gender and Partner salary.

Gender – total 53 Nulls, Partner salary – Total 106 Nulls

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	

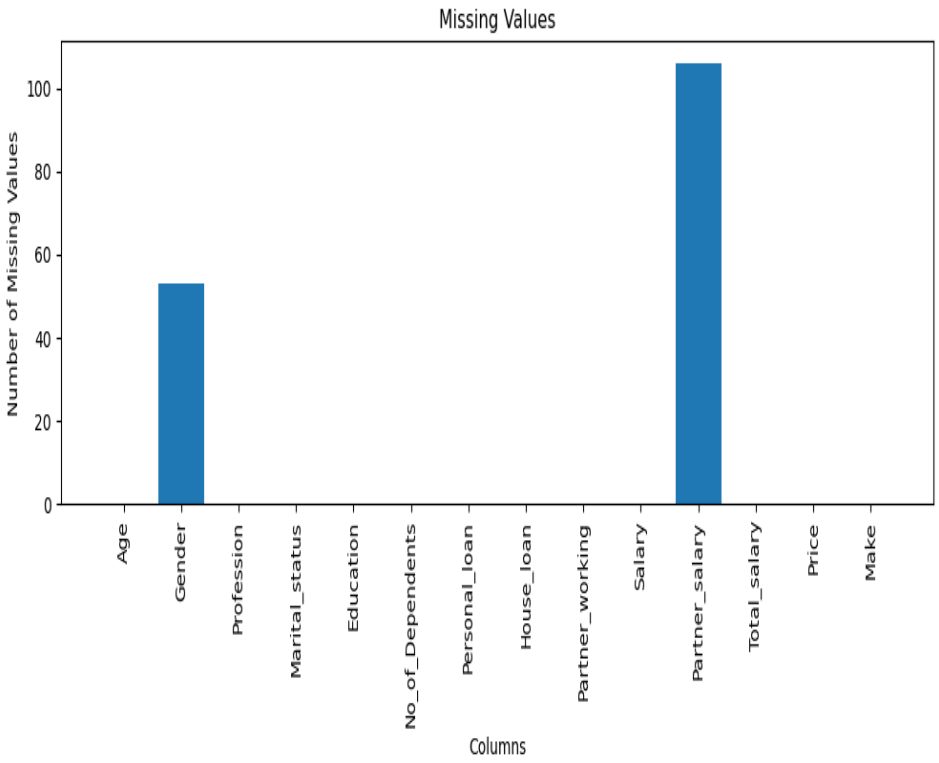


Fig: -4

fig: -5

Treating the Null values: -

a. **Gender:** Null values in Gender field can be imputed with '**Male'** having as the **mode** (maximum value in the dataset)
b. **Partner salary: fillna():** represent missing data, This is a method that replaces NaN (Not a Number) values with the value you provide. NaN values are often used in pandas to represent the missing data.

   The Data Frame df, replace all the NaN values in the Partner salary column with the mean salary of the partners, and make this change in the original Data Frame". This is a common way to handle missing data in datasets.

Duplicate Values: There are **no duplicate** records in the dataset.

Bad Values: Bad values are present in Gender as **Femal** , **Femle** and **nan** Rest of the categorical fields seem to be fine.

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

We will be treating the above by **replacing** the values Femal or Femle with **Female** and **nan** values with **male**.

Array (['Male', 'Female'], dtype=object)

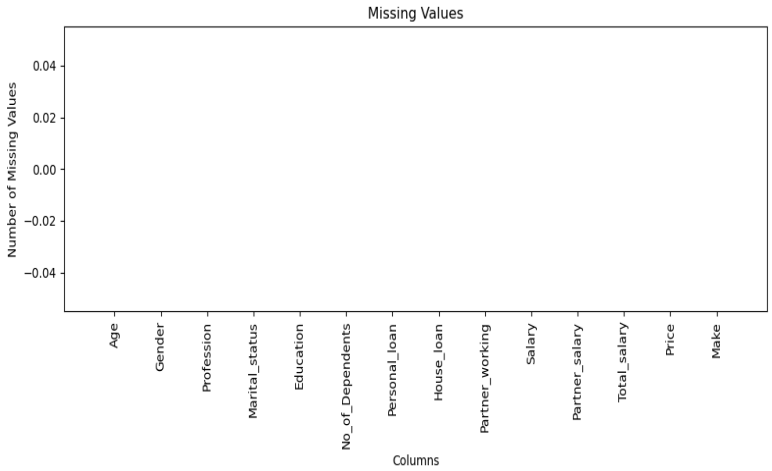


Fig: -5

Inspecting the Summary Statistics of the Dataset

	count	unique	top	freq
Gender	**1581**	**2**	**Male**	**1252**
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**
Gender Marital status	**1581**	**4**	**Male, Married**	**1136**

Fig: -6

	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

Fig: -7

1) The customer's age is between **22** and **54** years old i.e. majority might belong to **working age** group. **Mean age is 31.92** while **median age is 29** years, indicating age distribution is **positively skewed.**
2) The **Salary** of the customers **ranges between 30K and 99.3K** and the **distribution is symmetric**. The close mean and the median shows **skewness** is near to 0.
3) **Total salary** ranges between **30K and 171K** and does not show a high degree of skewness.
4) The **minimum price** of the **purchased automobile is 18K**, whereas **max is 70K.** Skewness indicates a small number of high priced purchases were made.

Checking Outliers in the numerical variable

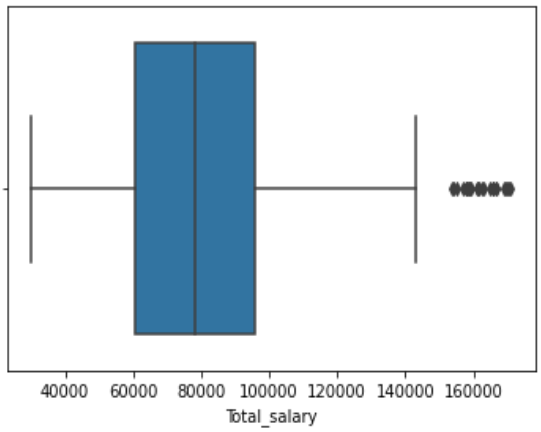


Fig: -8

- There are **no negative** values present in the numerical fields.
- From the boxplots we can observe **outlier values are present in Total salary** variables.
- Outliers are treated by using **IQR method**, i.e. bringing the larger outliers (Data points above the Q3 + 1.5 *IQR value) to the upper whisker

These calculations are part of a method called the **IQR method** for outlier detection, which is based on the statistical concept of quartiles. It's a commonly used method because it's simple to understand and implement, and it's robust to outliers. This method assumes that the data is unimodal and roughly symmetric, which is the case for many natural phenomena. However, it might not be the best choice for data that is multimodal or highly skewed. In those cases, other methods for outlier detection might be more appropriate

Boxplot after Outlier treatment

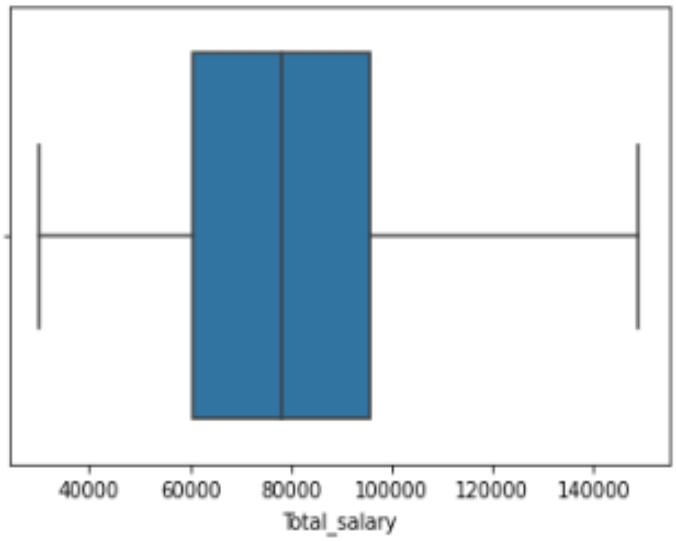


Fig: -9

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

Univariate Analysis of Numerical fields

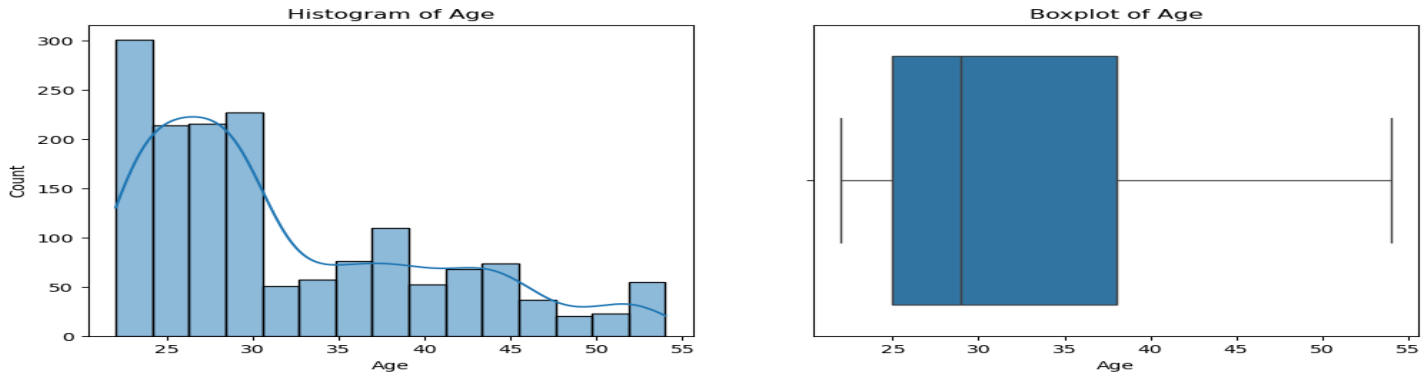


Fig: - 10

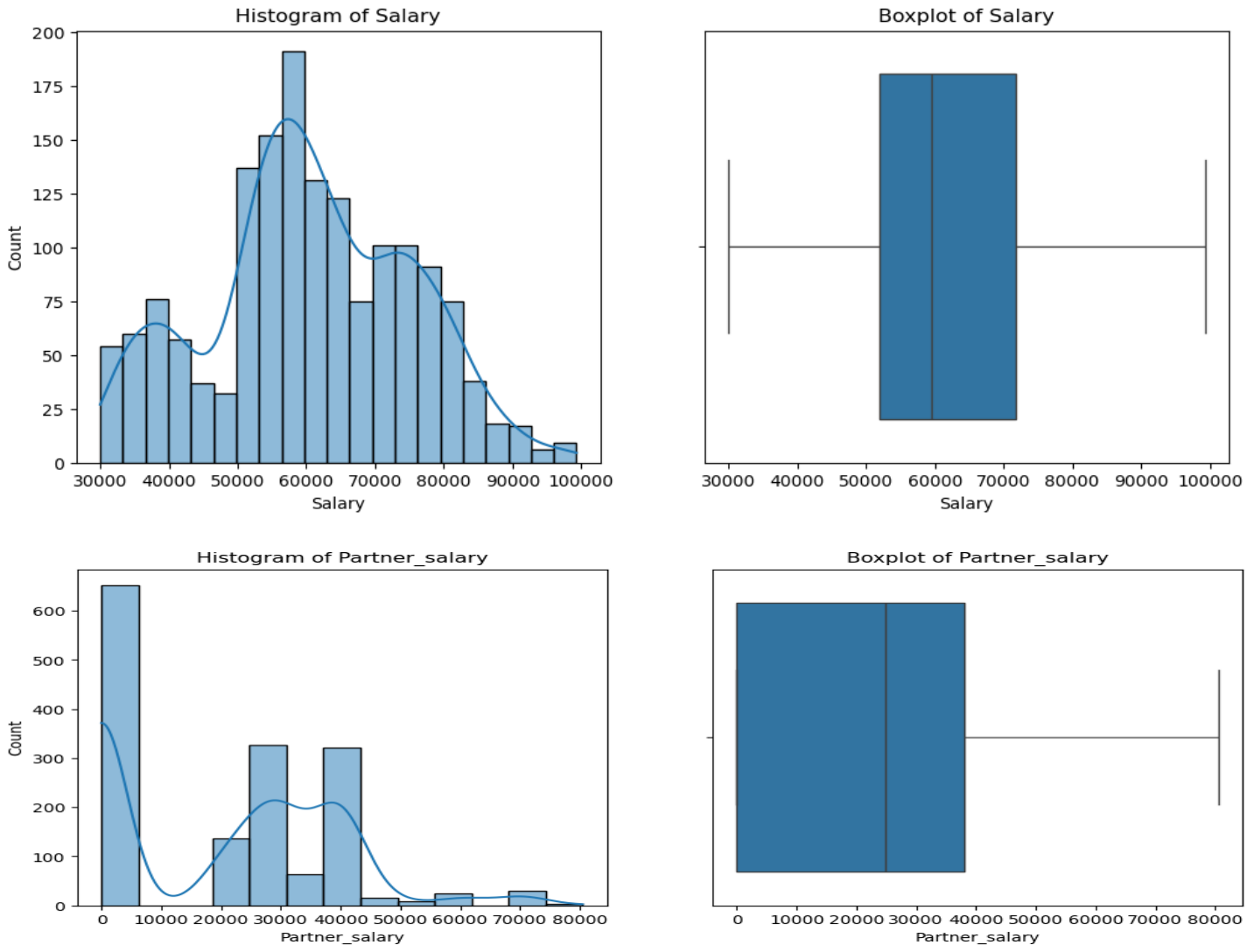

Histogram of Salary
Count
Salary
Boxplot of Salary
Salary
Histogram of Partner_salary
Count
Partner_salary
Boxplot of Partner_salary
Partner_salary

Fig: -11

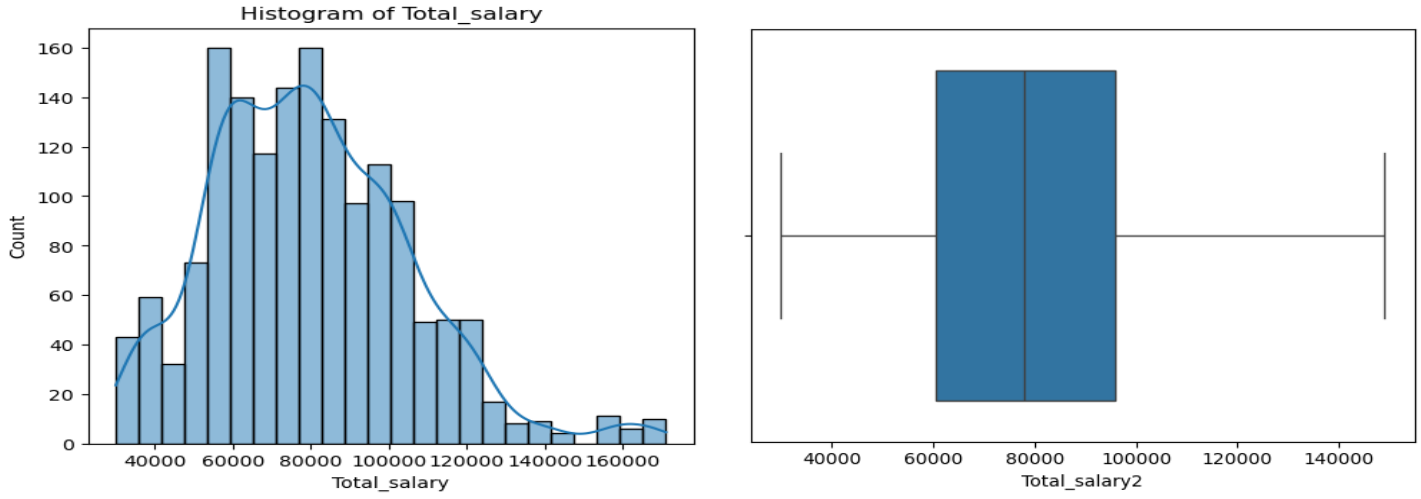

Histogram of Total_salary
Count
Total_salary
Total_salary2

Fig: -12

Univariate Analysis of Categorical fields

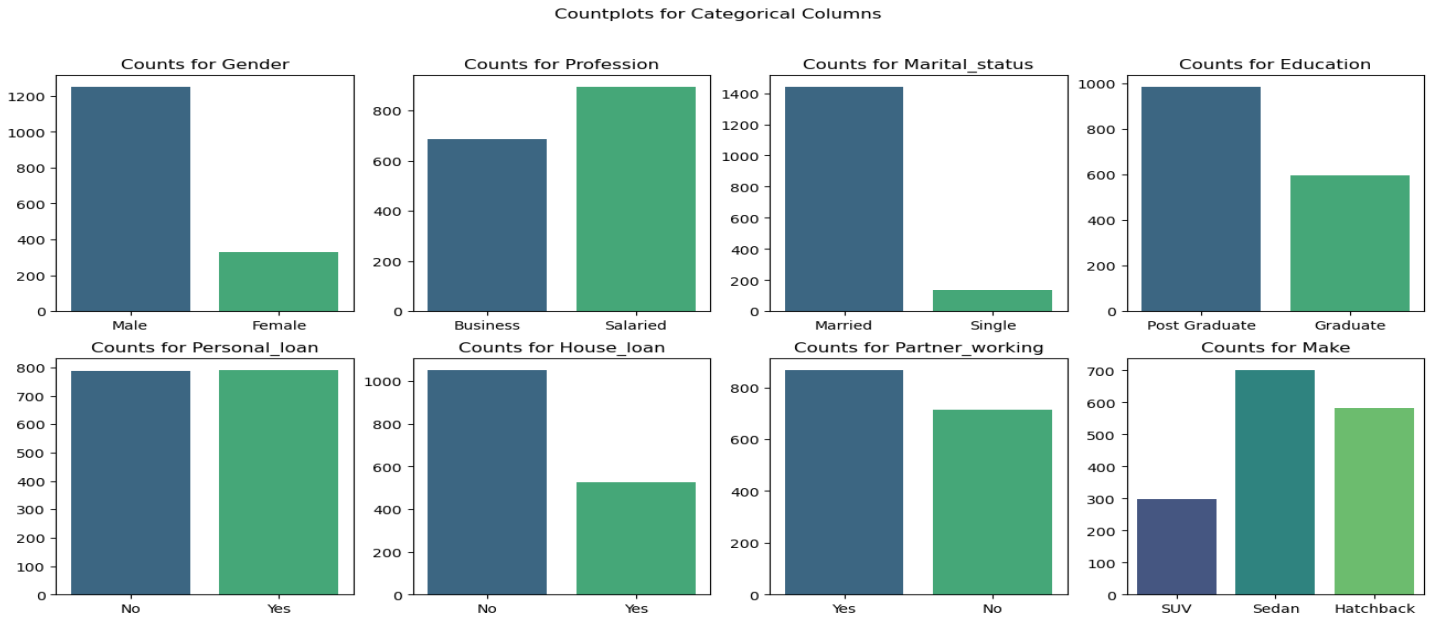


Fig: -13

Inferences

- **Majority** of the customers in the dataset are **Post Graduate**.
- **Sedan** is the **most preferred purchase**, followed by **Hatchback** and **SUV**
- **Salaried customer count** is **slightly higher** than that of **Business customers**.
- The number of **customers** having a **working partner** are **slightly higher** than customers with **non-working partner or singles**
- Most individuals do not have a **house loan.** And even **Personal loan**

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

Bivariate analysis of Numerical variables

Pair plot on the Data set: -

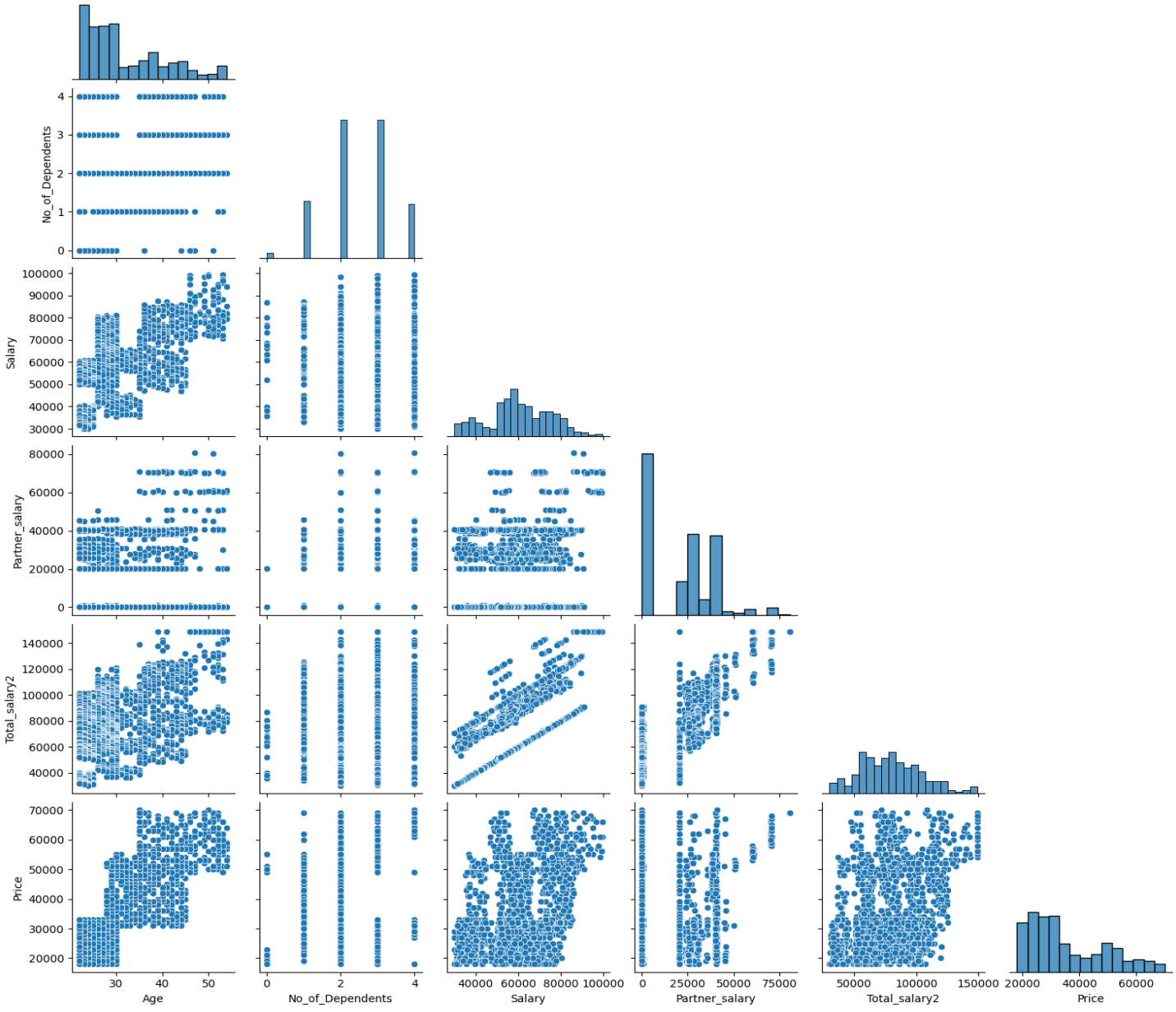
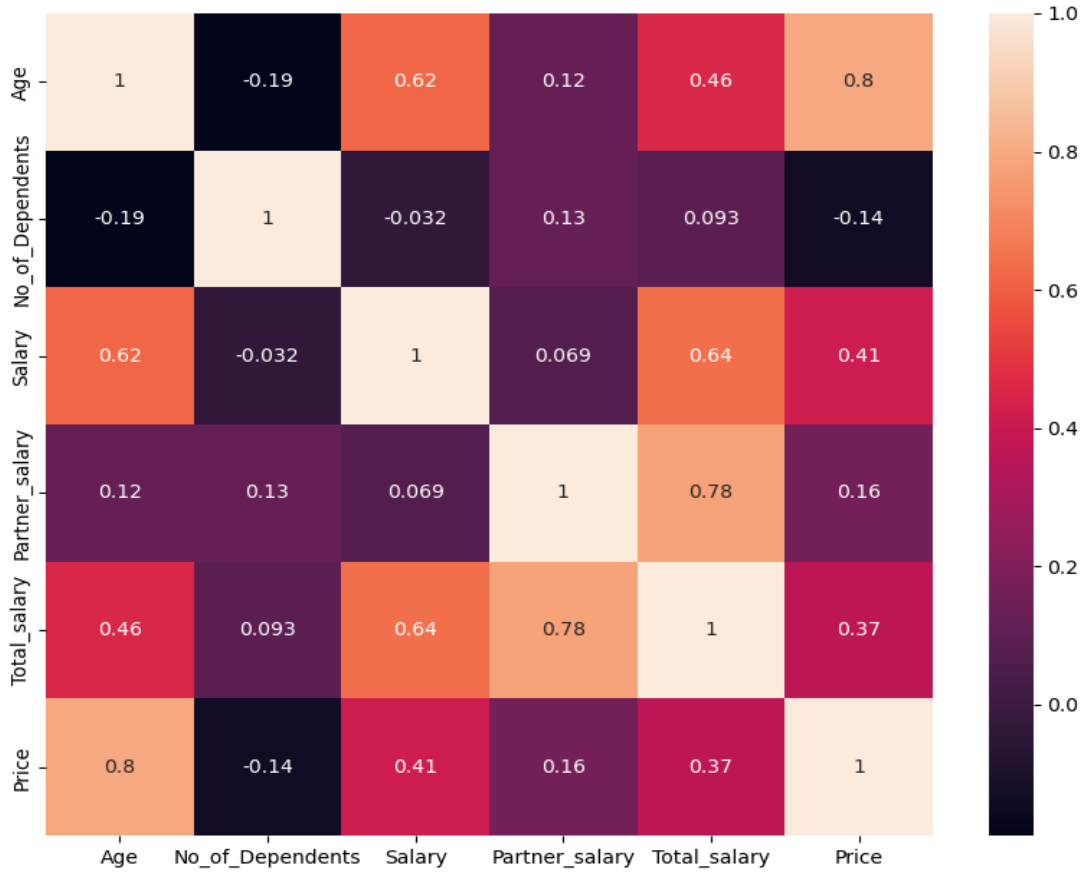


Fig: -14

Fig: -15



Inferences –

1) Hardly any linear relationships present among the fields.
2) **Positive correlation** between **Price and Age** is **0.8**, and **Total salary and Partner salary** has **0.78**

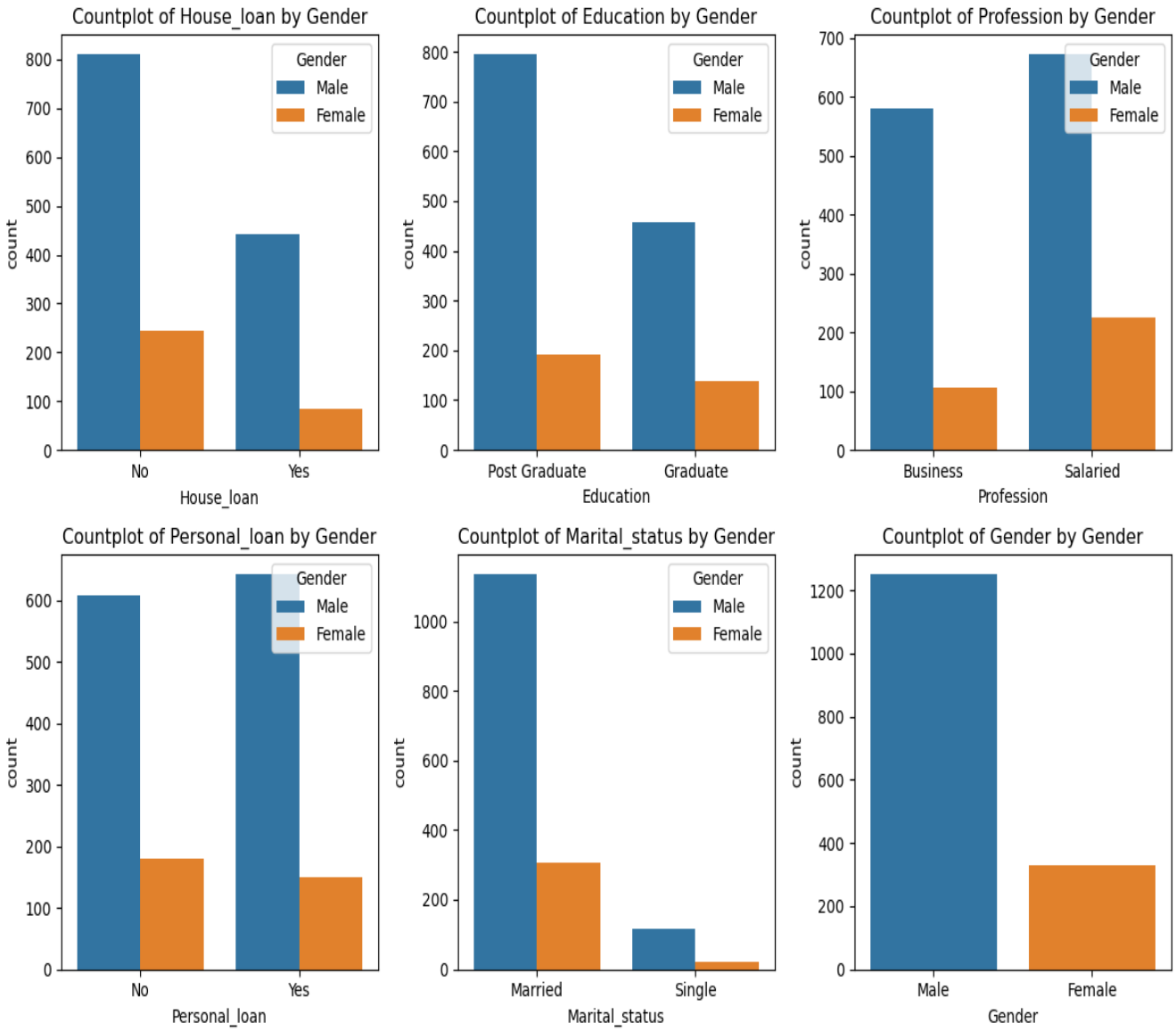
Bi- Variate analysis of Categorical vs Categorical variables –

Inferences –

1. The distribution of **personal loans** by gender. It shows that the majority of both **males** and **females** in the dataset do not have a personal loan. This could suggest that most individuals in the dataset either do not need or do not qualify for personal loans.
2. As we can see in the **Education** field most of the **Male** & **Female** are **Post Graduate** than the **Graduate.**
3. In the **Profession** most of the **Male** and **Female** are **Salaried** who works in IT sector, construction and automobiles sector all works in the company. As there are some **Male** and **Female** who has the **Business** field.

4. Most of the **Male** and less **Female** are having the **Personal Loan** may be some of them are not eligible to the requirement of the **Personal Loan** are may be not needed.
5. In the dataset most of them are married and there are less people who are still single.

Fig: -16



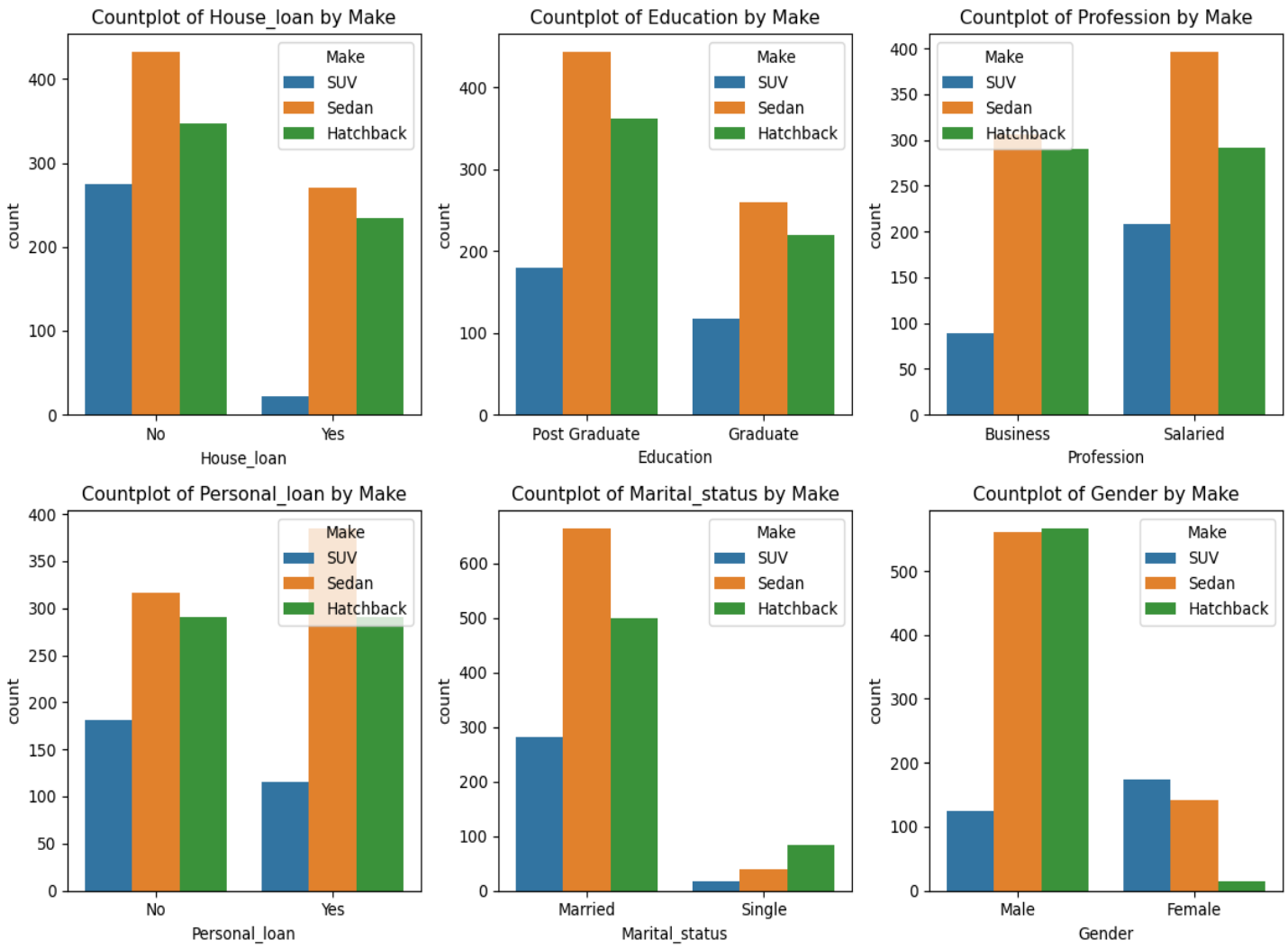


Fig: -17

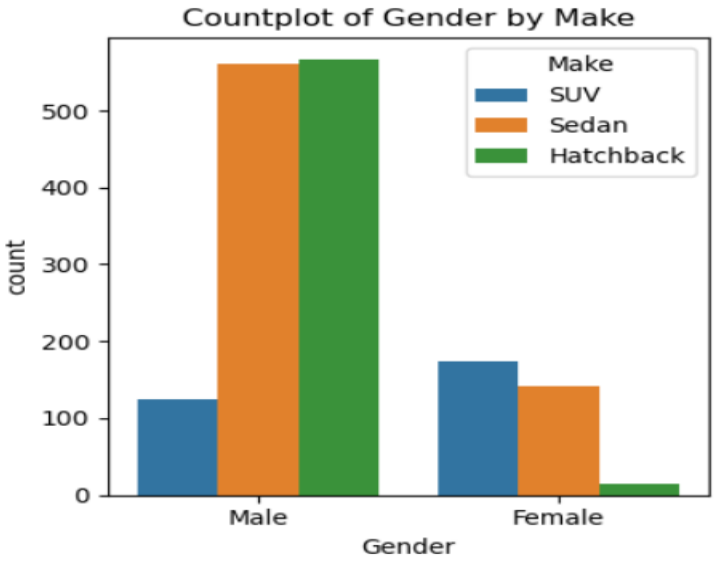
Inferences –

1. Customers who have a **house loan** are **not likely to buy an SUV** (which is the costliest make among the three).
2. **Females prefer SUV** and are **least likely to buy a Hatchback**, whereas **Male prefer Sedan or hatchback**. **SUV** is **least preferable** among **males.**
3. **Married** customers **prefer Sedan** whereas **single** customers **prefer Hatchbacks**.
4. Customers who are in both Profession (Business & Salaried) they prefer Sedan and followed by Hatchback and SUV
5. Customers who have both **Personal loan** & **House loan** they prefer to buy **Sedan** than the **SUV**
    - Over all we can say that **Male** prefer **Sedan** & **Hatchback** which is less affordable than the other **Make**.
    - **Female** customers most preferably buy the **SUV's** because as I assume they prefer style and comfort.
    - Customers who have the loans they prefer the **sedan** over **SUV.**

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

E1) "Do men tend to prefer SUVs more compared to women?**"**

Analyzing the ratio of SUV purchases for both the Genders, we get:



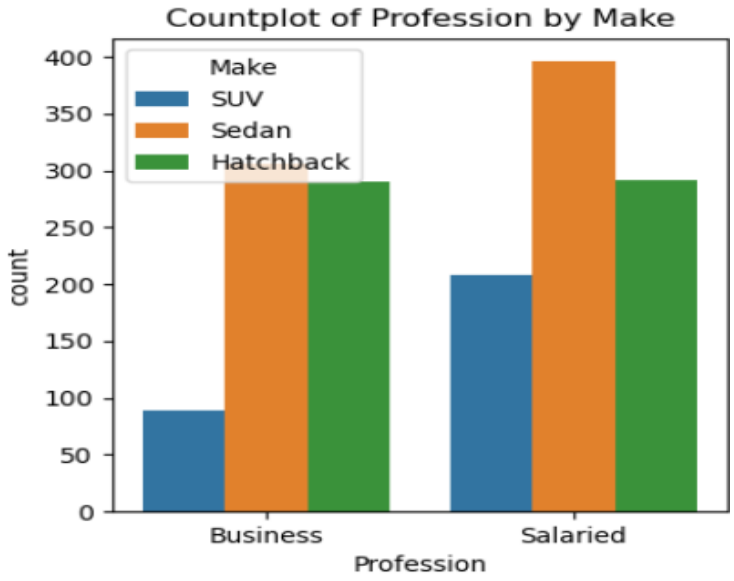
Make	Hatchback	SUV	Sedan
Gender			
Female	15	173	141
Male	567	124	561

Fig: -18

From above data, we can **conclude** that the **Men** prefer to buy Sedan and Hatchback, because basically the **Sedan** and **Hatchback** are cheaper than the **SUV**.

E2) What is the likelihood of a **salaried** person buying a **Sedan**?

Analyzing the ratio of **Sedan** purchases against **profession**, we get:



Make	Hatchback	SUV	Sedan
Profession			
Business	290	89	306
Salaried	292	208	396

Fig: -19

From above data, we can **conclude** that the **Salaried** person prefers to buy the **Sedan** more than the Business Customers. Hence we can go with the Statement

E3) 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?

Analyzing the ratio of **SUV** purchases against **profession**, we get:

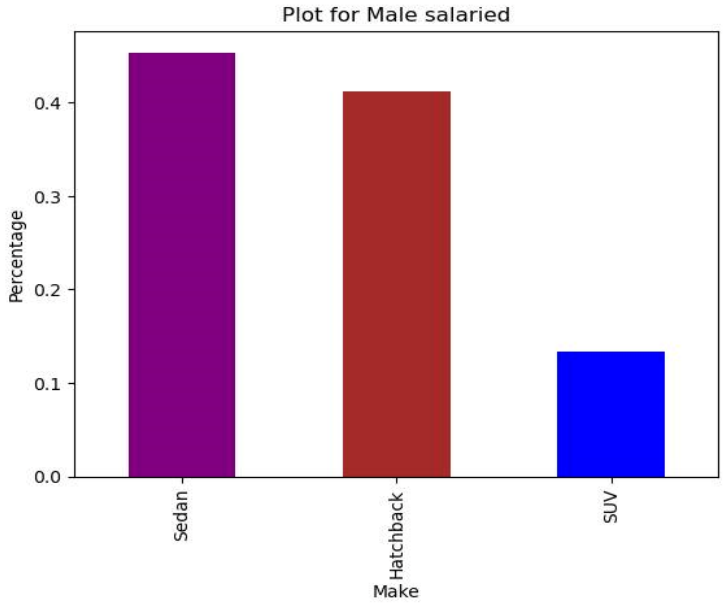


Fig:-20

From above data, we can **conclude** that the **statement** made by **Sheldon Cooper is Incorrect**

E4) How does the amount spent on purchasing automobiles vary by **Gender**?

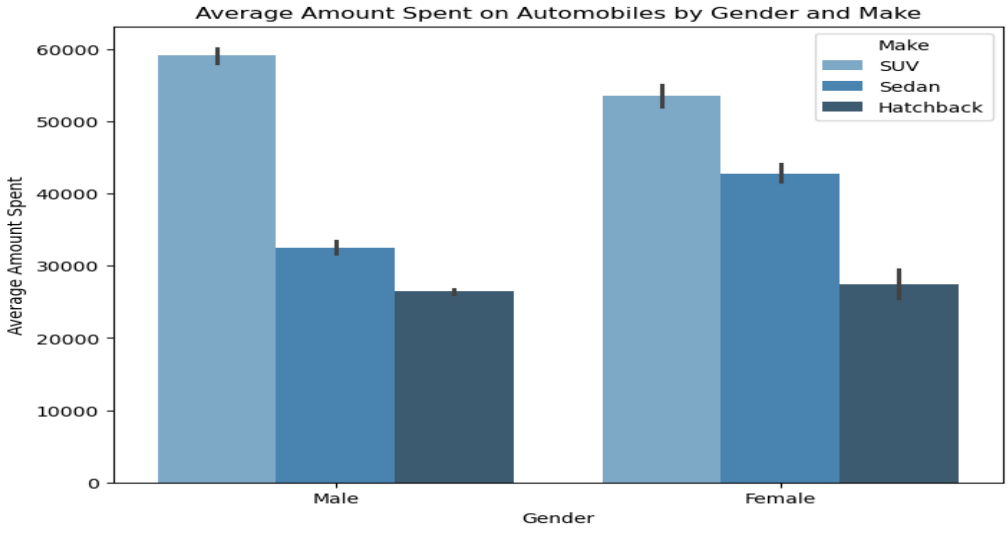


Fig: - 21

- From above data, we can **conclude** that **Male** spent **58K-59K** on the **SUV**,And followed by **31K-32K** on **Sedan** then on **Hatchback** they spent approximately **25K**
- As of **Female** Average amount spent on **SUV** is **55K**, on the **Sedan** they spent average amount of **42K** and on the **Hatchback** they spent approximately of **25K.**

E5) How much money was spent on purchasing automobiles by **individuals** who took a **personal loan**?

Total amount spent on purchasing automobiles by individuals who took a **personal loan: 34457**.

E6) How does having a **working partner** influence the purchase of higher-priced cars?

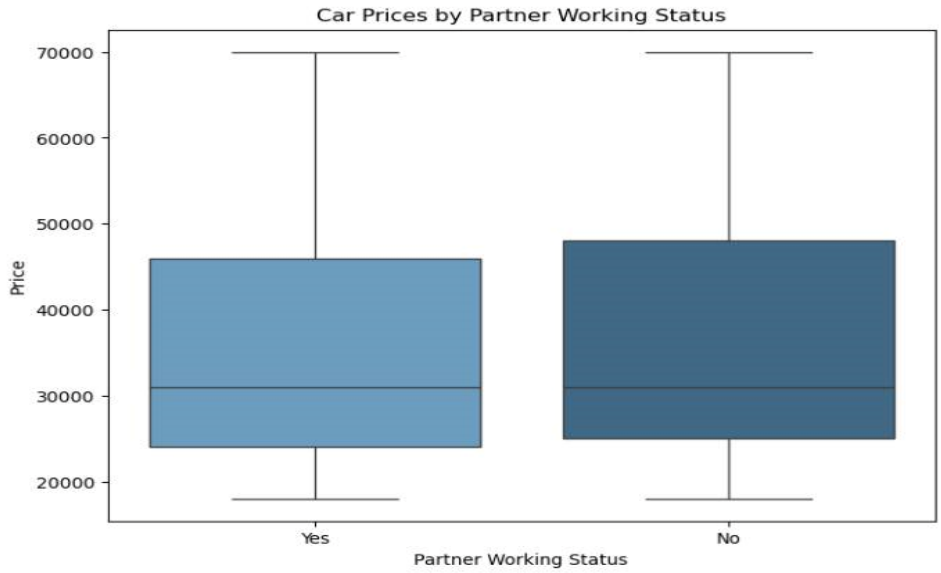


Fig: - 22

- From above data, we can **conclude** that it appears that when the partner is **working ('Yes'),** the car prices show a wider range and variability compared to when the partner is not **working ('No'**). This could suggest that individuals with **working partners** tend to purchase cars across a wider range of prices, potentially including **higher-priced cars**.
- However, the median car price (indicated by the line within each box) seems to be similar for both categories. This means that the 'typical' car price does not significantly differ whether the partner is working or not.
- **Mean of Price across Partner working:**
- Partner working: No = 36000
- Partner working: Yes = 35267

- **Median of Price across Partner working:**
- Partner working: No = 31000
- Partner working: Yes = 31000

- **The Mean and Median price** of the purchased automobile is **almost similar across the Partner working category**, thus indicating whether **partner is working or not,** it has **slight effect** on the **Purchase made by the customer.**

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

Give justification along with presenting metrics/charts used for arriving at the conclusions.

F1) Gender

Females are more likely to buy **SUV's** and on average spend more on cars than males **47705** Units against **32416** Units.

- Mean of Price across Gender:
- Female = 47705
- Male = 32416
- Median of Price across Gender:
- Female = 49000
- Male = 29000

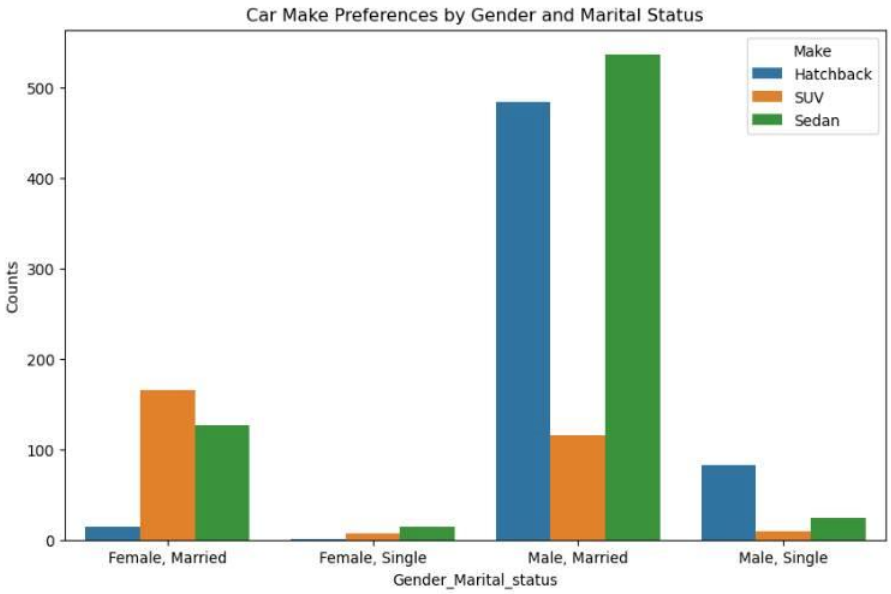
- **Mean and Median Price** for **Female** customers is **higher than Male customers.**

F2) Personal loan

- Mean of Price across Personal Loan:
- Personal Loan: No= 36742
- Personal Loan: Yes= 34457
- Median of Price across Personal Loan:
- Personal Loan: No= 32000
- Personal Loan: Yes= 31000

- **Mean and Median of Price** for purchase made by customers **without a Personal loan** is **slightly higher** than **customers who have a Personal Loan**.
- **To ensure increased spend** of customers with Personal loans, the **business can look at cheaper interest rates** (for Automobile purchase) or **easy the repayment terms**.

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



Gender	Marital_status			
Female	Married	14	166	127
	Single	1	7	14
Male	Married	484	115	537
	Single	83	9	24

Fig: -23

Most frequently purchased Car make grouped on Marital Status and Gender, we find:

- Female – Married: SUV
- Male – Married: Sedan
- Male – Single: Hatchback
- Female – Single: Sedan

Analyzing the **mean Price of purchased car** across the Marital status and Gender, we find:

- Mean Price for purchases made by Married Females = 62857
- Mean Price for purchases made by Married Males = 60692

- **Mode** of the Car make for Gender and Marital status fields shows that **both the married groups preferring SUV**.
- Similarly, the **Mean of Price** for **Male Married is approx. 60K** while it is **62K for Female Married**.
- All the **Male Married Customers** with **Total Salary greater than 149 K purchased SUV**. Whereas **Married male** with **lower Total salary preferred Sedan**

H. Actionable Insights – Business Recommendation: -

1. Recognize Consumer Behavior: Keep an eye out for evolving consumer priorities, financial healths, plans to share, and climate change. Utilize these data-driven insights to react and adjust to changing consumer and automobile industry demands.

2. Sustainability plans: Take advantage of the chance to explore sustainability comfortness of the car make and design that the customers and less price,

3. Product Strategy and Development: Direct the development, management, and strategies of your products, including portfolio optimization and the formulation of go-to-market plans for novel technologies.

4. Keep Up with Industry Trends: Keep tabs on consumer preferences and emerging technologies that could have a major influence on the automobile sector. The Global Automotive Consumer Study by Deloitte provides insights that can assist automakers in overcoming certain obstacles,

5. Adapt to Market Disruptions: The changing market environment of the automobile sector presents many automotive firms with continual challenges, ranging from sophisticated technology to supply chain disruptions and climate change policies. Be ready to respond to these challenges with innovation and adaptation.

6. Pay Attention to Electric Vehicles (EVs): As the world moves toward sustainability, there is an increasing need for EVs. Making investments in EV development could prove to be a wise strategic choice in the future.

(Note: some of the sentences are borrowed from the WEB regarding the technical things of Automobiles and its business field)

Problem-2 GODIGT Bank

A bank can generate revenue in a variety of ways, such as charging interest, transaction fees and financial advice. Interest charged on the capital that the bank lends out to customers has historically been the most significant method of revenue generation. The bank earns profits from the difference between the interest rates it pays on deposits and other sources of funds, and the interest rates it charges on the loans it gives out.

GODIGT Bank is a mid-sized private bank that deals in all kinds of banking products, such as savings accounts, current accounts, investment products, etc. among other offerings. The bank also cross-sells asset products to its existing customers through personal loans, auto loans, business loans, etc., and to do so they use various communication methods including cold calling, e-mails, recommendations on the net banking, mobile banking, etc.

GODIGT Bank also has a set of customers who were given credit cards based on risk policy and customer category class but due to huge competition in the credit card market, the bank is observing high attrition in credit card spending. The bank makes money only if customers spend more on credit cards. Given the attrition, the Bank wants to revisit its credit card policy and make sure that the card given to the customer is the right credit card. The bank will make a profit only through the customers that show higher intent towards a recommended credit card. (Higher intent means consumers would want to use the card and hence not be attrite.)

Framing An Analytics Problem - Analyze the dataset and list down the top 5 important variables, along with the business justifications.

- Size of Dataset: Dataset has 8448 rows and 28 columns.

- Data headers: PFB the data headers present in the dataset for quick reference

	userid	card_no	card_bin_no	issuer	card_type	card_source_date	high_networth	active_30	active_60	active_90	cc_active30	cc_active60	cc_active90	hotlist_flag	widget_products	engagement_products	annual_income_at_source	other_bank_cc_holding	bank_vintage	T+1_month_activity	T+2_month_activity	T+3_month_activity	T+6_month_activity	T+12_month_activity	Transactor_revolver	avg_spends_l3m	Occupation_at_source	cc_limit
0	1	4364 05XX XXXX XXXX	430459	Visa	edge	2019-09-29	B	0	1	1	0	0	0	N	1	3	152211	Y	27	0	0	0	0	0	T	27729	Self Employed	250000
1	2	4377 49XX XXXX XXXX	437748	Visa	prosperity	2002-10-30	A	1	1	1	0	0	0	N	4	1	403871	Y	52	0	0	0	0	0	R	269854	0	950000
2	3	4377 49XX XXXX XXXX	437748	Visa	rewards	2013-10-05	C	0	0	0	0	0	0	N	4	2	1345429	N	23	1	0	0	0	0	R	76597	Student	210000
3	4	4258 05XX XXXX XXXX	425806	Visa	indianoil	1998-06-01	E	0	1	1	1	1	1	N	6	0	886560	N	49	0	0	1	0	0	T	9156	Self Employed	80000
4	5	4377 49XX XXXX XXXX	437748	Visa	edge	2006-06-13	B	1	1	1	0	1	1	N	4	3	1808582	N	21	1	0	0	0	0	T	38108	Salaried	220000

Fig: -24

- Dataset Information: There are 19 numerical and 8 categorical variables. PFB the details of each: -

Five Points Summary:

```
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       8448 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
```

Fig: -25

	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

Fig: -26

Exploring the data: -

- No duplicate values found

- In **Transactor revolver** we found **38 null values**

Histograph for the numerical values

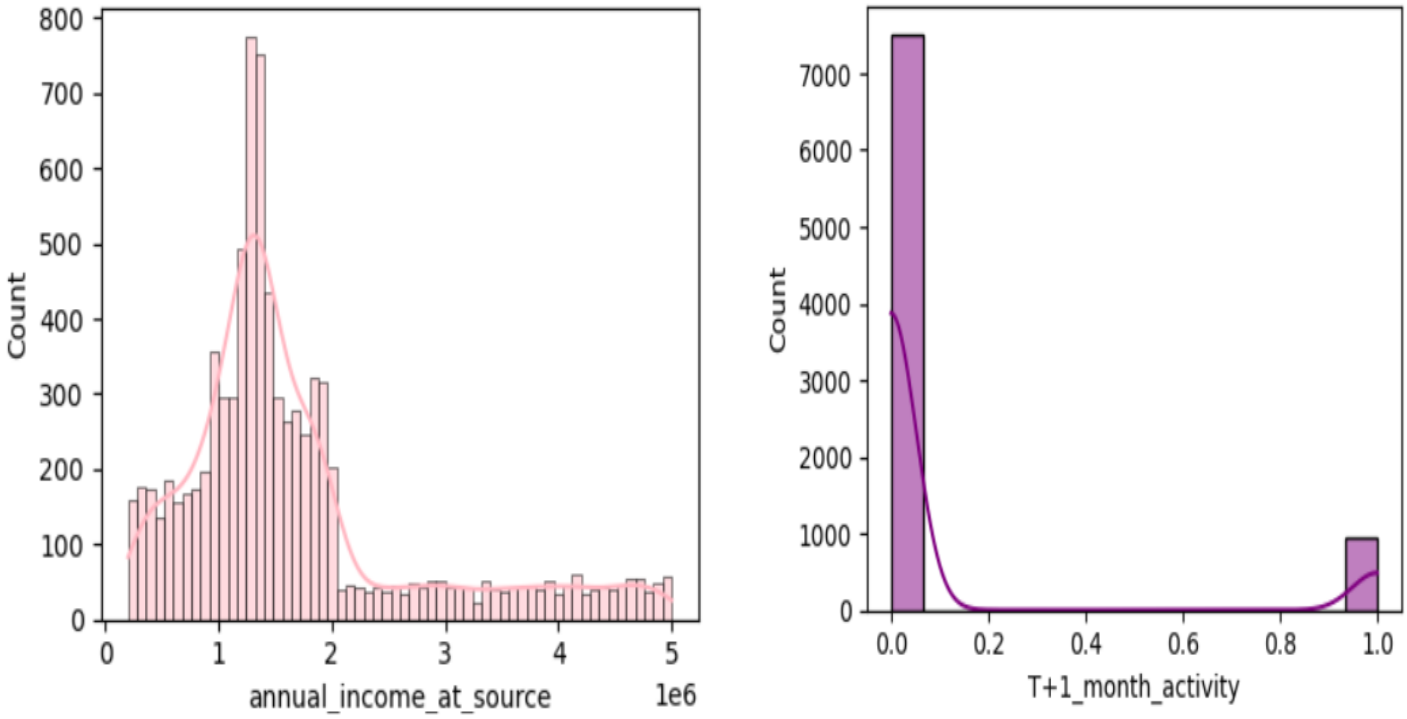


Fig: -27

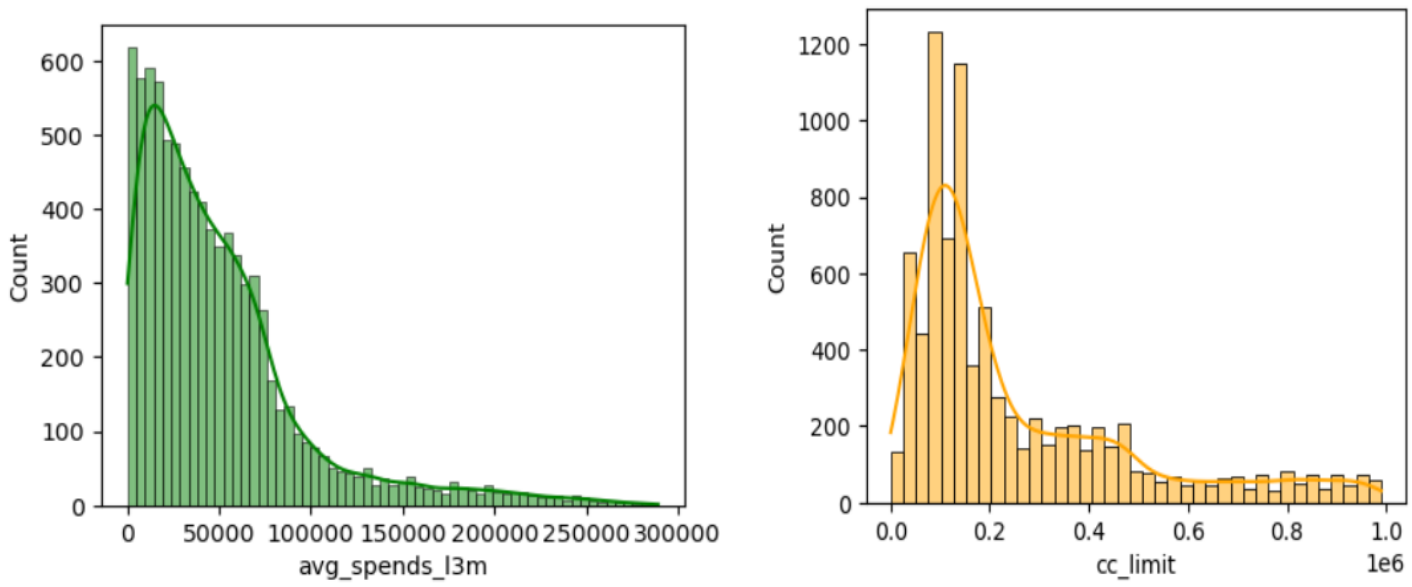
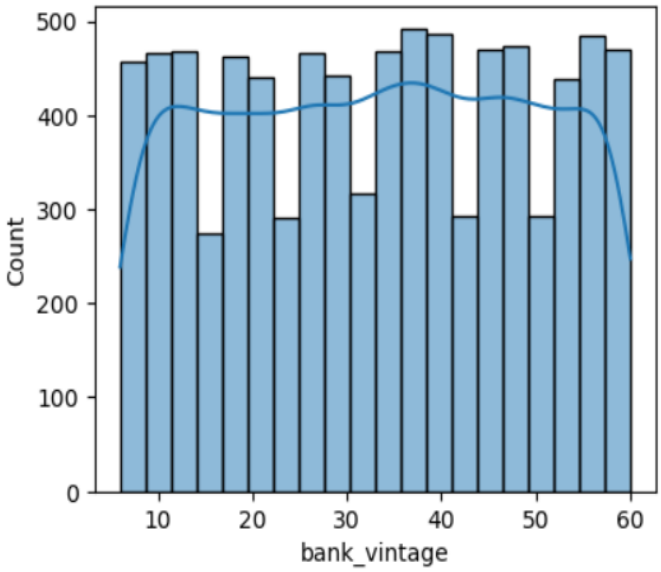


Fig: -28


Count
bank_vintage

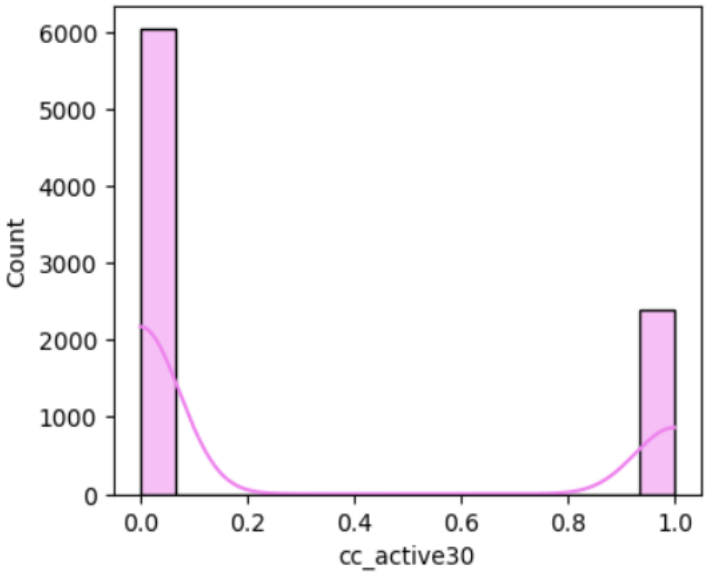
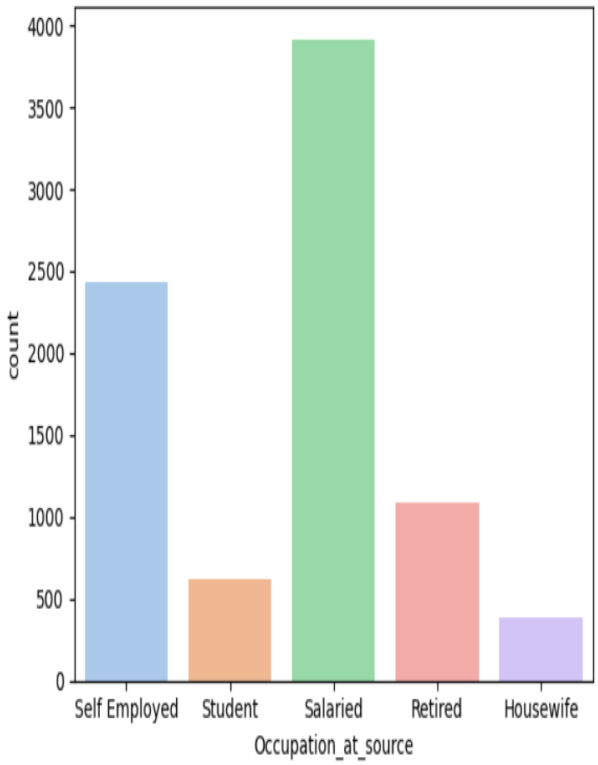

Count
cc_active30

Fig: -29


count
Self Employed
Student
Salaried
Retired
Housewife
Occupation_at_source

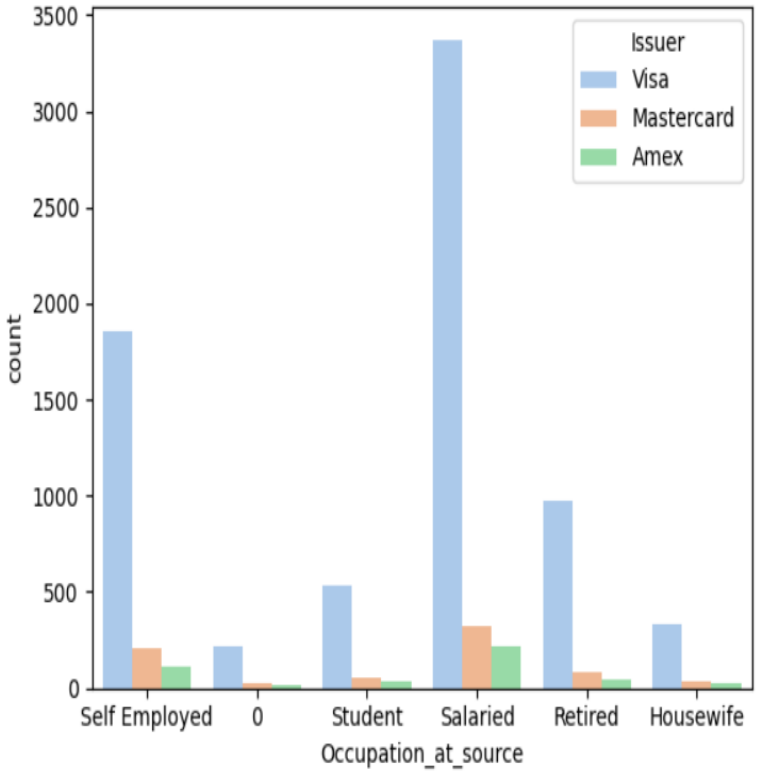

Issuer
Visa
Mastercard
Amex
count
Self Employed
0
Student
Salaried
Retired
Housewife
Occupation_at_source

Fig: -30

Below are the Top 5 important variables from the given dataset with justification.

1) Annual Income at source-

Annual income plays a big role in the purchasing power of an individual hence is a vital piece of info. Income can be used by the banks to make better decisions in areas such as risk profiling, targeted ads, campaigns, offers, loan limits etc.

2) CC limit -

Defining Credit Card limit for customers basis their attributes (such as income, CIBIL Score, etc.) is part of the Risk Management practice wherein the banks try to minimize the number of defaulters.

3) CC_active30 –

Flag variables such as cc_active30, cc_active60 can be used to get an understanding over how frequently does the customer use the credit card, if the account is dormant or if the customer is experiencing any issues leading to reduced usage of the card etc.

4) Occupation at source-

Profession is a key factor in credit risk assessment. Salaried individuals might have a steady income stream, potentially indicating a lower credit risk compared to self-employed individuals or students. This could influence decisions related to credit limits, loan approvals, and interest rates.

5) avg_spends_13m-

The avg_spends_l3m variable can give important insights on the customer spending behavior. It can be used to identify whether the credit card is primary or secondary card of customer, i.e. high spend indicates primary account whereas lower spend would mean secondary account. Campaigns can be rolled out on the basis of the customer preference, customized offers can begiven to lure customers into using the credit account more frequently.

(Note: some of the sentences are borrowed from WEB regarding the banking sector etc.)

----------End of Report----------