



BUSINESS REPORT

Of

Austo Motor Company

PDS Coded Project

PGP – DSBA

Major Analysis Covers : Demand of Customer

Submitted by : Yash Juneja

INDEX

1. About Company.....	3
2. Scope of Work/Objective.....	3

Data Overview

3. Import the Libraries.....	3
4. Load the Data.....	3
5. Structure of data.....	3
6. Types of data.....	4
7. Treating Missing Values.....	4
8. Statistical Summary.....	5
9. Check and treat data irregularities.....	5
10. Observation & Insights.....	6

Univariate Analysis

11. Explore all the Variables (categorical and Numerical) in the Data.....	6-13
12. Check and treat outliers.....	14
13. Observations and Insights.....	15

Bivariate Analysis

14. Explore the relationship between all numerical variables.....	15
15. Explore the correlation between all numerical variables.....	16
16. Explore the relationship between categorical vs numerical variables.....	16-21

Key Questions

17. Do Men tend to prefer SUVs more as compared to women ?.....	21
18. What is the Likelihood of a salaried person buying a Sedan ?.....	22
19. 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 ?.....	22
20. How does the amount spend on purchasing automobiles vary by gender ?.....	23
21. How much money was spend on purchasing automobiles by individuals who took a personal loan ?.....	23
22. How does having a working partner influence the purchase of higher-priced cars ?.....	24

Actionable Insights & Recommendations

23. Actionable Insights – Business Recommendation.....	24-26
--	-------

1.) About Company :

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 analytics professional to improve the existing campaign.

2.) Scope of Work/Objective :

Analyze the data to get a fair idea about the demand of customers which will help them in enhancing their customer experience. Suppose you are a Data Scientist at the company and the Data Science team has shared some of the key questions that need to be answered. Perform the data analysis to find answers to these questions that will help the company to improve the business.

Data Overview

3.) Import the Libraries

For Analyzing the Data set using Python, we have imported all the required libraries as NumPy, Pandas, Metplotlib and seaborn

4.) Loading the Data

Loading the data is the initial step to analysis in any Project.

5.) Checking the Structure of Data

Checking the Structure of data is important step for understanding of format and features using `df.head()`, `df.info()` and `df.shape`

This shows top 5 records and this Data has 1581 rows and 14 columns

	Age	Gender	Profession	Marital_status	Education	No_of_Dependents	Personal_loan	House_loan	Partner_working	Salary	Partner_salary
0	53	Male	Business	Married	Post Graduate	4	No	No	Yes	99300	70700.0
1	53	Femal	Salaried	Married	Post Graduate	4	Yes	No	Yes	95500	70300.0
2	53	Female	Salaried	Married	Post Graduate	3	No	No	Yes	97300	60700.0
3	53	Female	Salaried	Married	Graduate	2	Yes	No	Yes	72500	70300.0
4	53	Male	Salaried	Married	Post Graduate	3	No	No	Yes	79700	60200.0

6.) Check the Types of Data

```
<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: 173.0+ KB
```

Data types of Columns are

float64 = 1 , int64 = 5 , object = 8

Numerical = 6

Object = 8

7.) Treating Missing Values

	0
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

Missing Values affects the analysis, So identify and treating them is very important,

In this Data Set we have Missing values mentioned below :

Gender = 53

Partner Salary = 106

Right way of treating the Missing values is done by Mean, Mode and Median

Gender is replaced with Mode and Partner salary with Mean.

8.) Check the Statistical Summary

	count	unique	top	freq	mean	std	min	25%	50%	75%	max
Age	1581.0	NaN	NaN	NaN	31.922201	8.425978	22.0	25.0	29.0	38.0	54.0
Gender	1528	4	Male	1199	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Profession	1581	2	Salaried	896	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Marital_status	1581	2	Married	1443	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Education	1581	2	Post Graduate	985	NaN	NaN	NaN	NaN	NaN	NaN	NaN
No_of_Dependents	1581.0	NaN	NaN	NaN	2.457938	0.943483	0.0	2.0	2.0	3.0	4.0
Personal_loan	1581	2	Yes	792	NaN	NaN	NaN	NaN	NaN	NaN	NaN
House_loan	1581	2	No	1054	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Partner_working	1581	2	Yes	868	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Salary	1581.0	NaN	NaN	NaN	60392.220114	14674.825044	30000.0	51900.0	59500.0	71800.0	99300.0
Partner_salary	1475.0	NaN	NaN	NaN	20225.559322	19573.149277	0.0	0.0	25600.0	38300.0	80500.0
Total_salary	1581.0	NaN	NaN	NaN	79625.996205	25545.857768	30000.0	60500.0	78000.0	95900.0	171000.0
Price	1581.0	NaN	NaN	NaN	35597.72296	13633.636545	18000.0	25000.0	31000.0	47000.0	70000.0
Make	1581	3	Sedan	702	NaN	NaN	NaN	NaN	NaN	NaN	NaN

By using describe() Function in Pandas, we have generated statistical summary of this Dataset.

9.) Check for and treat(if needed) data irregularities

Upon Reviewing the data, it has been observed that there are some variables that has multiple and invalid names in Gender value, which has been corrected by using replace function

```
string = AutoMotor
string.replace('Femal','Female')
```

	Age	Gender	Profession	Marital_status	Education	No_of_Dependents	Personal_loan	House_loan	Partner_working	Salary	Partner_salary	Total_salary	Price
0	53	Male	Business	Married	Post Graduate	4	No	No	Yes	99300	70700.0	170000	61000
1	53	Female	Salaried	Married	Post Graduate	4	Yes	No	Yes	95500	70300.0	165800	61000
2	53	Female	Salaried	Married	Post Graduate	3	No	No	Yes	97300	60700.0	158000	57000
3	53	Female	Salaried	Married	Graduate	2	Yes	No	Yes	72500	70300.0	142800	61000
4	53	Male	Salaried	Married	Post Graduate	3	No	No	Yes	79700	60200.0	139900	57000
...
1576	22	Male	Salaried	Single	Graduate	2	No	Yes	No	33300	0.0	33300	27000
1577	22	Male	Business	Married	Graduate	4	No	No	No	32000	NaN	32000	31000
1578	22	Male	Business	Single	Graduate	2	No	Yes	No	32900	0.0	32900	30000
1579	22	Male	Business	Married	Graduate	3	Yes	Yes	No	32200	NaN	32200	24000
1580	22	Male	Salaried	Married	Graduate	4	No	No	No	31600	0.0	31600	31000

1581 rows x 14 columns

10.) Observation and Insights

Now, the Data Looks Perfect and Fine for analyzing the demand of Customers according to requirements, Budget and Need.

By using this Data we can Provide Information for Increasing sales and Individual experience.

Univariate Analysis

11.) Exploring Categorical and Numerical variables in the data

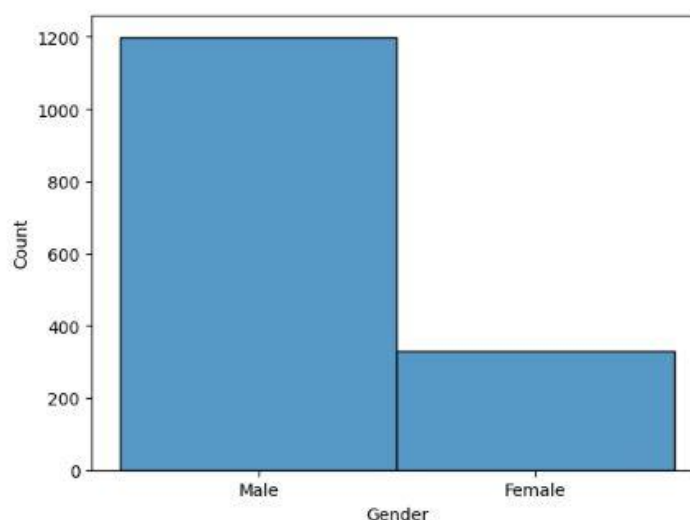
We have few categorical columns mentioned below

1. Gender
2. Profession
3. Marital status
4. Education
5. Personal loan
6. House loan
7. Partner working
8. Make

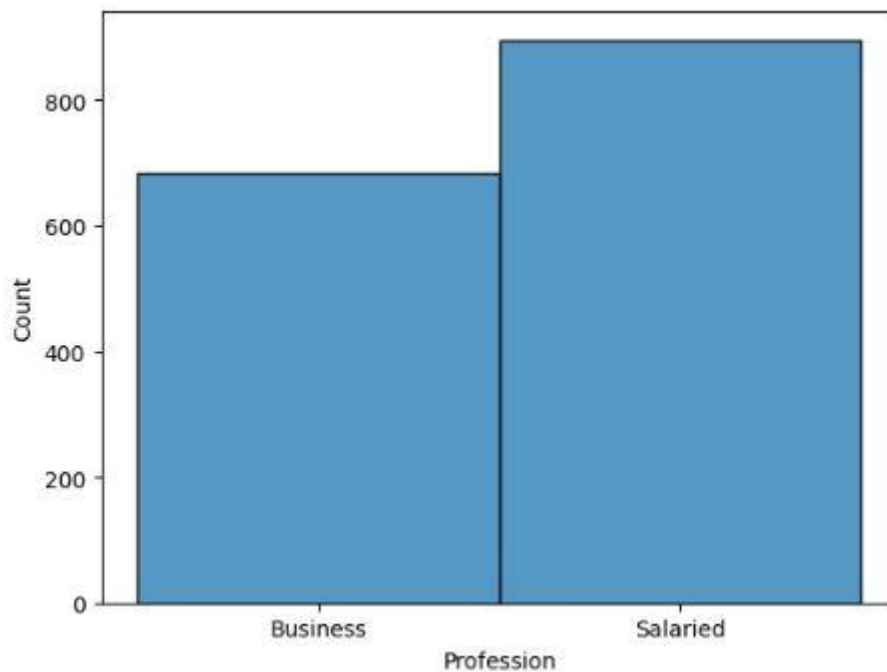
By using Plotting techniques, we will show the categorical distribution of data by using Bar Chart

Gender : Graph clearly shows Dominance of Male car owners as compared to Female.

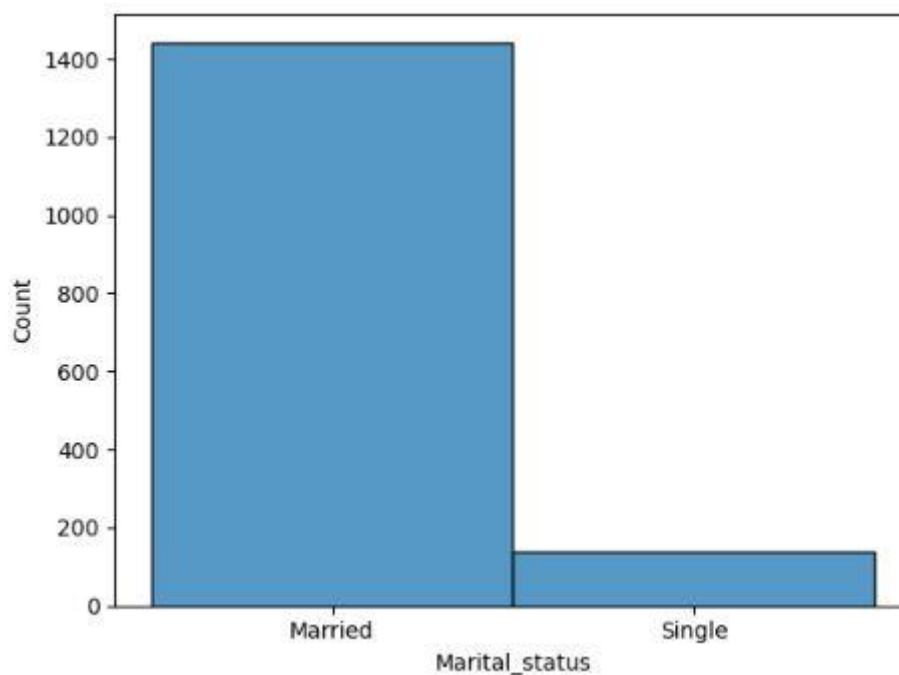
Bar Chart of Gender : According to this male Prefer buying cars more than Females



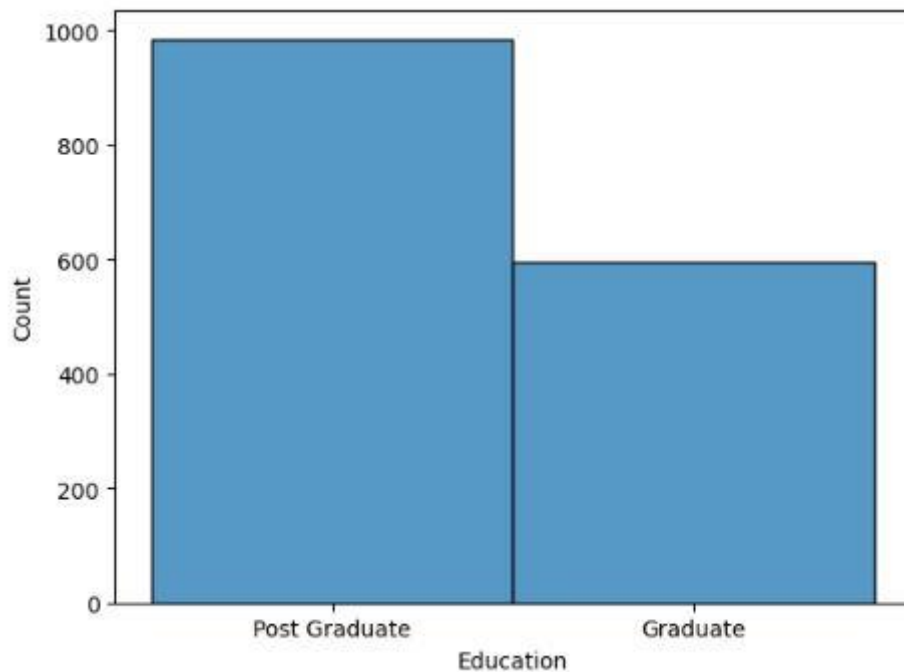
Profession : Data Suggests cars are more preferred among salaried person as compared to Business



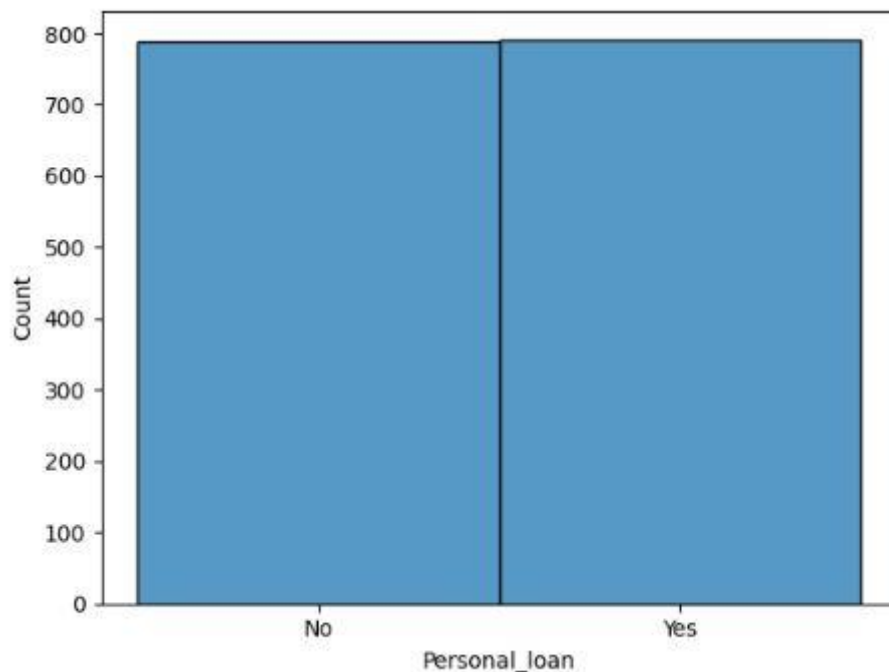
Marital Status : Married People are considering more into buying cars as compared to Single.



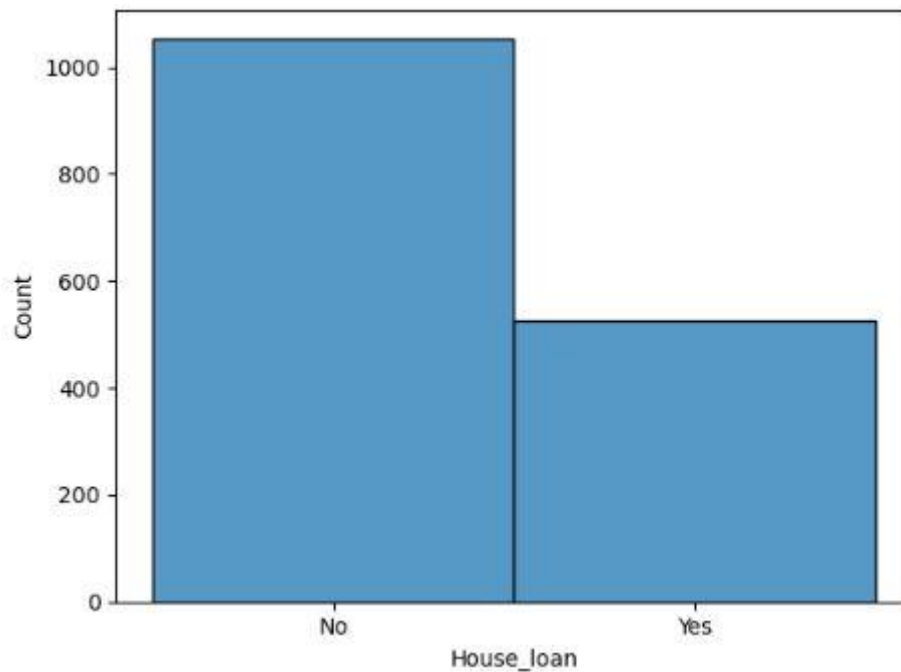
Education: Data Shows Post Graduates preferring buying Cars with comparison to Graduate.



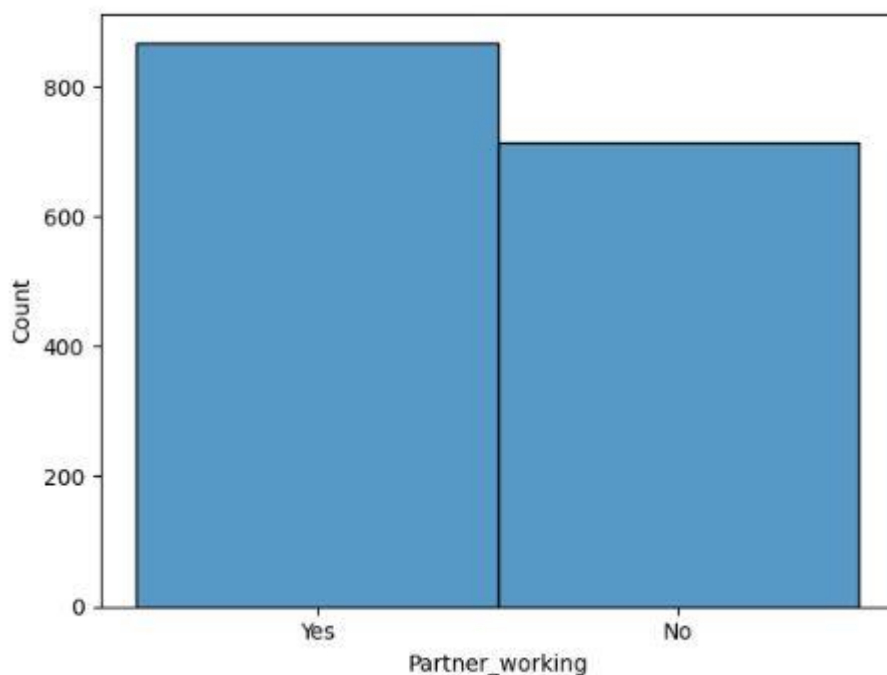
Personal Loan : There is no impact of Personal Loan in buying of Cars



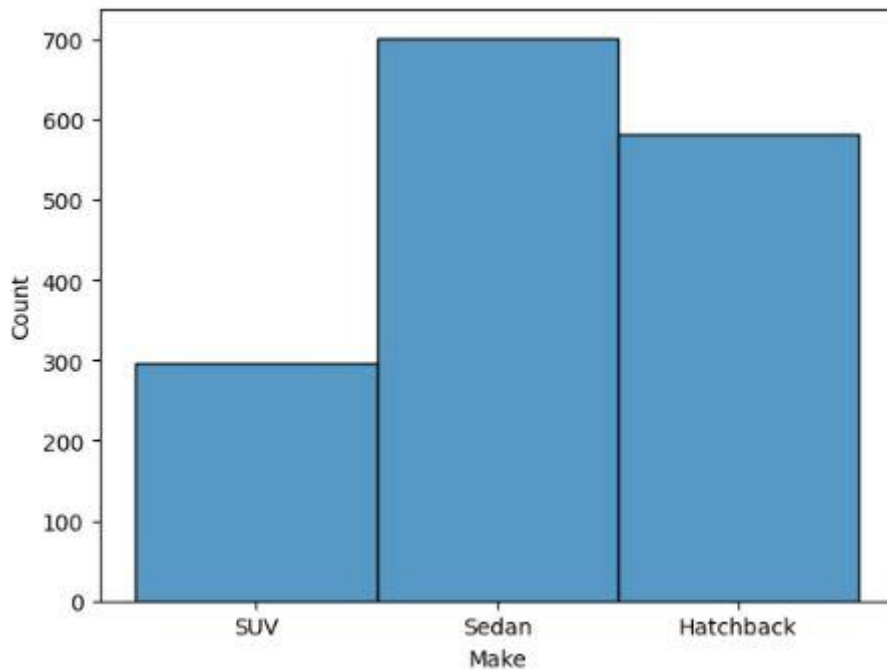
House Loan : Data Shows that People with No House loan are buying more cars, it looks like House loan disbalances the budget.



Partner Working : This data shows that, whose Partner is working preferring buying more cars as compared to Partner not working, that means having money in hand is easy decision of buying cars compared to the person with Non-working partners.



Make : This Graph shows that People love to buy and travel from Sedan, Their may be a reason of Budget, as Sedan comes in the middle class budget and SUV is costly and only preferred by those whose salary and income is high. Hatchback comes in Between both.



Exploring Numerical Variables mentioned below, It can be both int and float

- Age
- No of Dependents
- Salary
- Partner Salary
- Total Salary
- Price

For Numerical variable and distribution, box plot will be best for understanding

Box Plot is a graphical representation that displays the distribution of datasets and provides summary. It consists of 5 major key values that represent important characteristics of the data.

Minimum(min) = Smallest value in the dataset

First Quartile(Q1) = The Median of the lower half of Data set

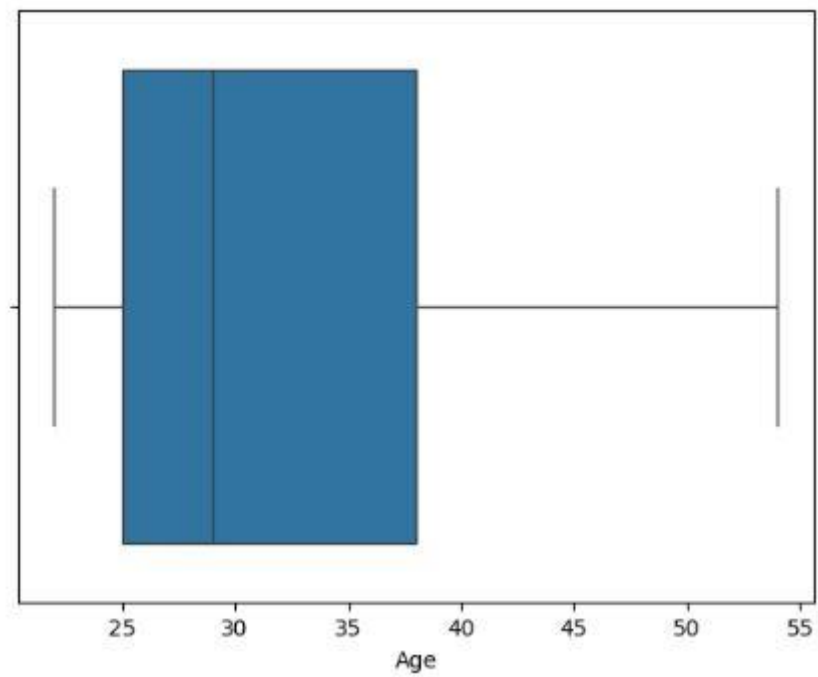
Median = The Middle value of the Dataset.

Third Quartile(Q3) = Median of Upper half in Dataset

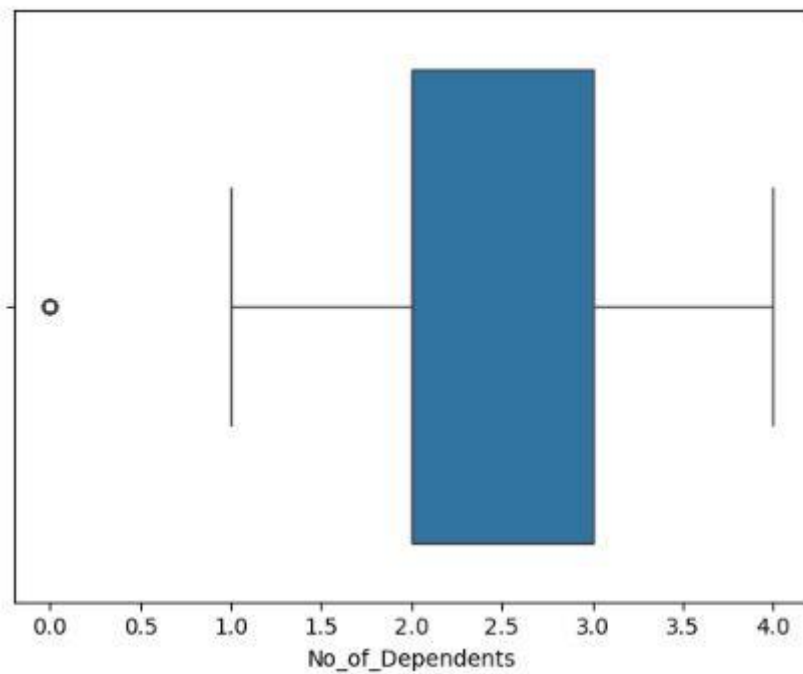
Maximum(Max) = Largest Value in Dataset

Here Shows Graphical Representation

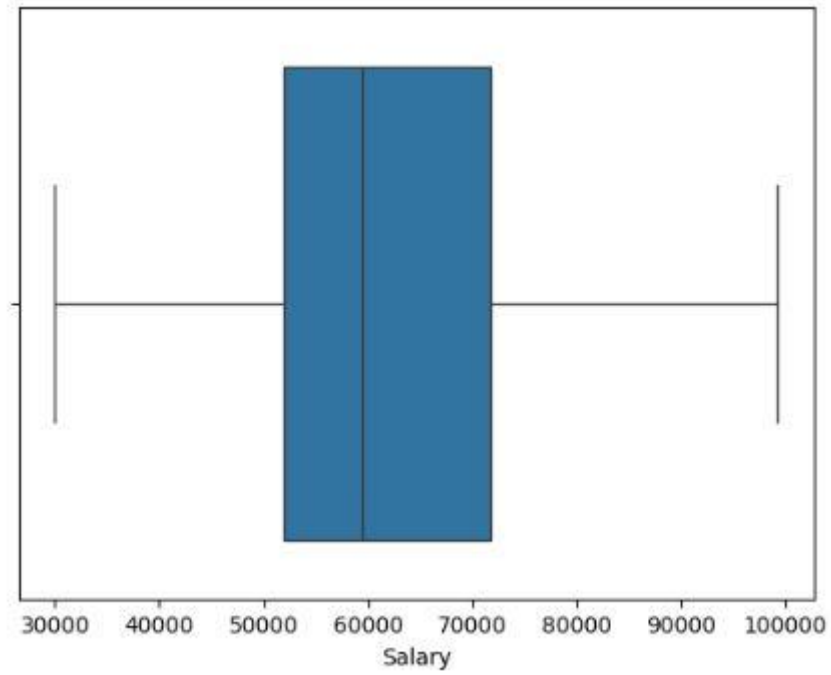
Box Plot of Age



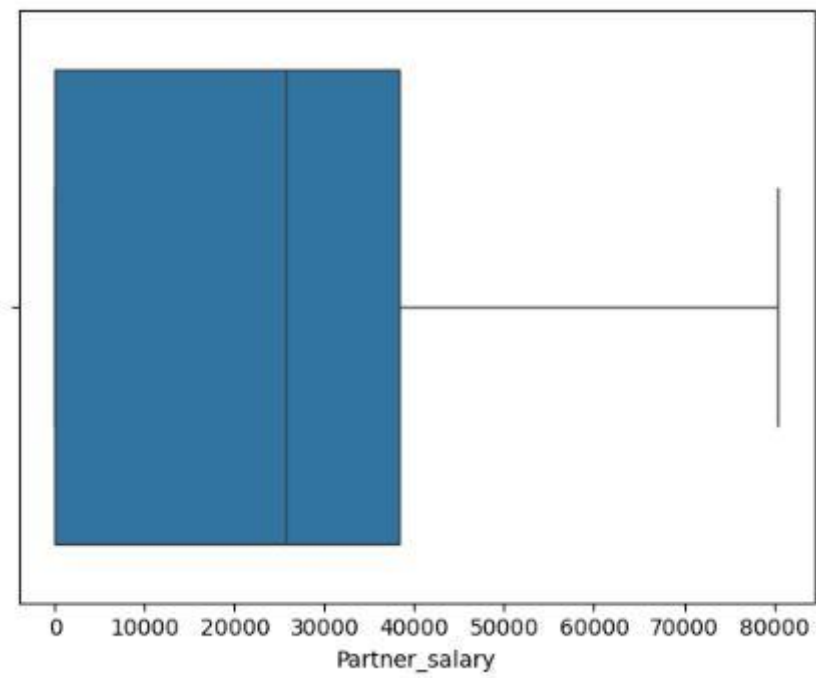
Box Plot of No of Dependents



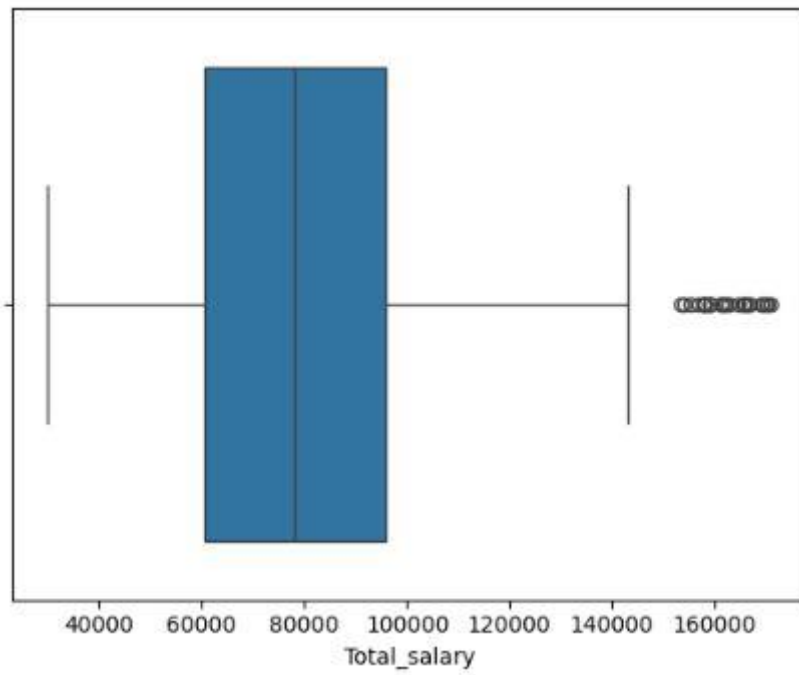
Box Plot of Salary



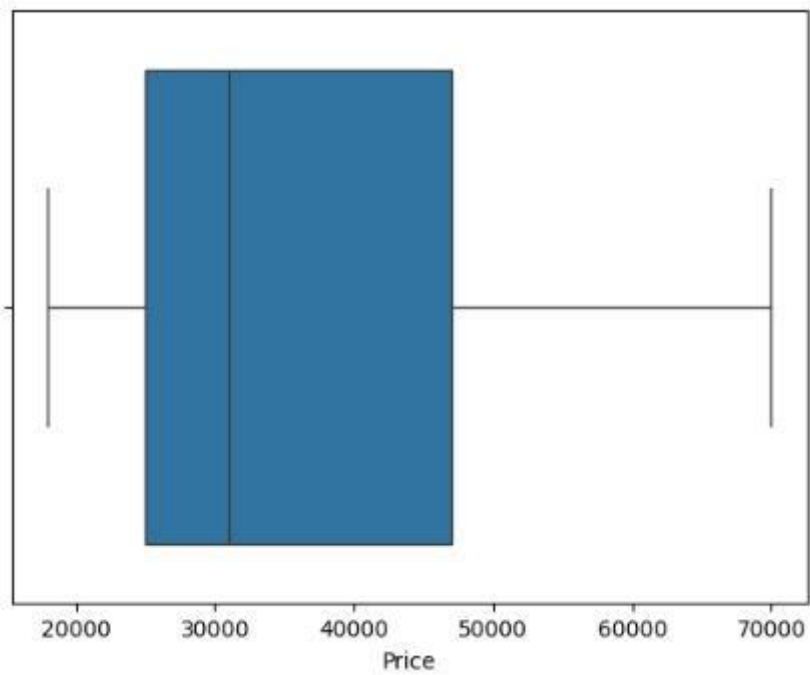
Box Plot of Partner Salary



Box Plot of Total Salary



Box Plot of Price

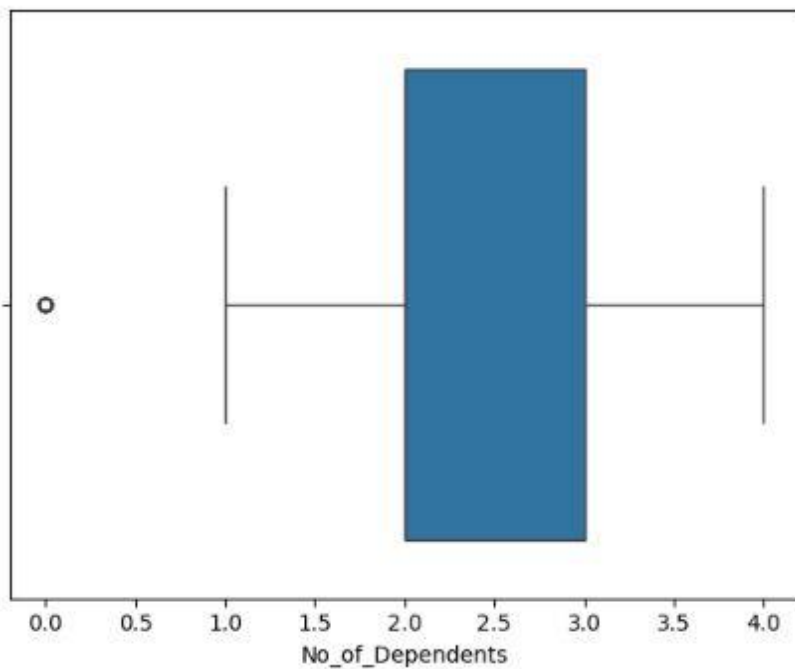


12.) Identification and treatment of outliers

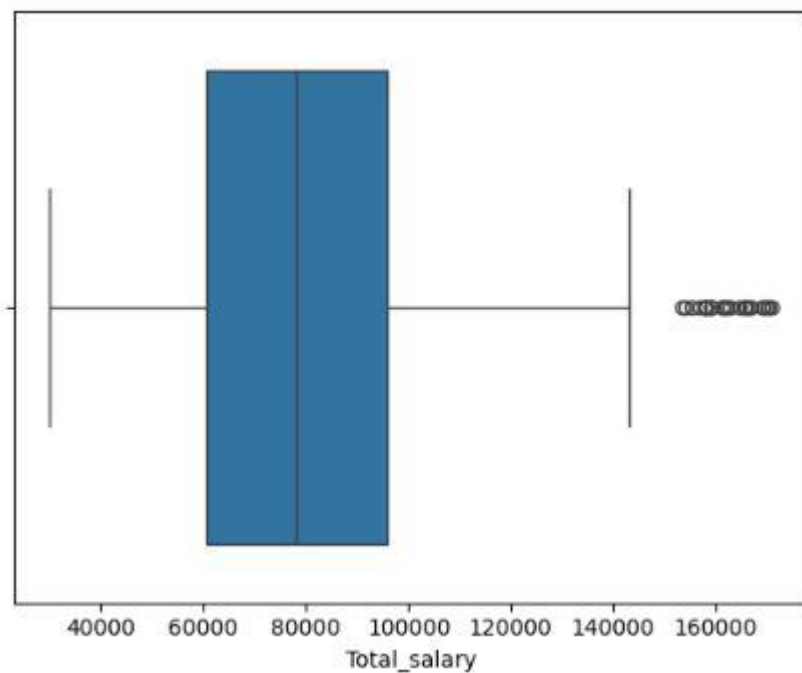
We can see that there are some outliers present in few variables of this Data, which can impact the analysis of Data, This leads to have bad effects on data analysis.

Treating outliers appropriately is most important for Stable results in Data Analysis

Below graph shows the variable of No of Dependents



Below graph shows the variable of Total Salary



Treatment of outliers:

By using Median technique, treatment of outliers can be done, In this Outliers has been replaced by Median Value.

13.) Observations & Insights

Data had outliers which has been treated now data has been fully ready for analysis.

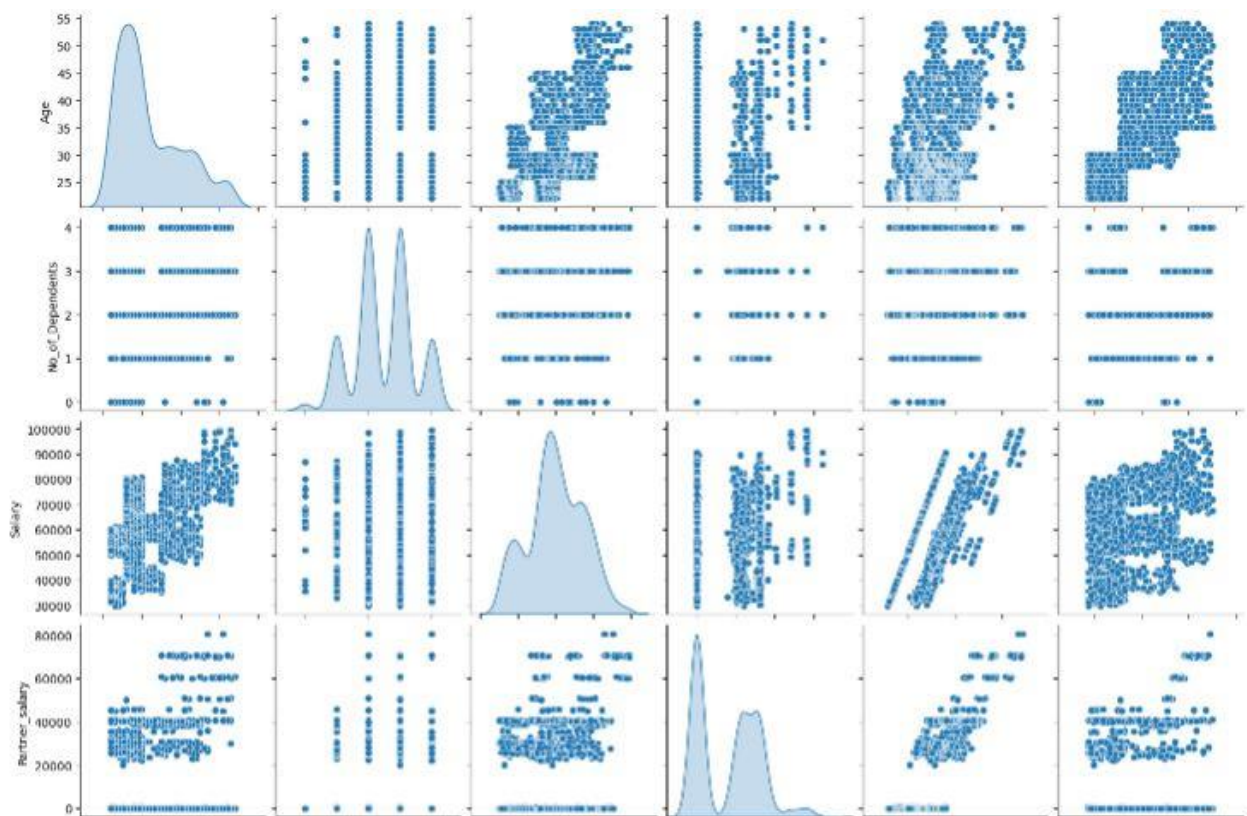
- Analysis shows dominance of Salaried male who is married.
- Personal loan doesn't really make any difference in buying a car as per data.
- Sedan is the most preferred choice as per car.

Bivariate Analysis

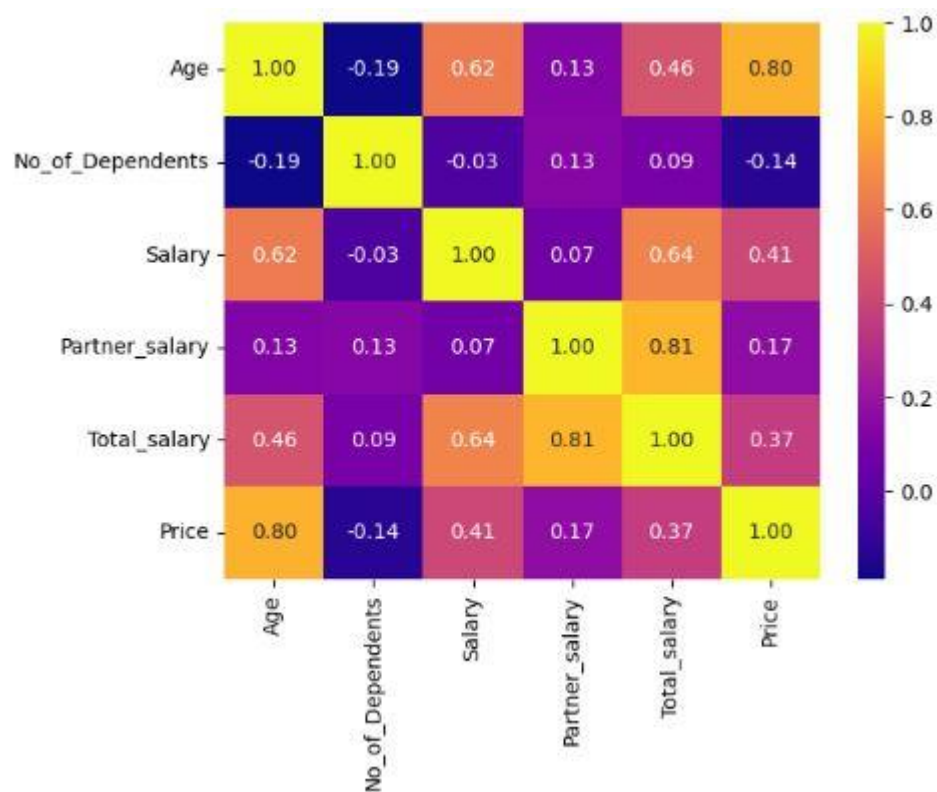
14.) Explore the relationship between all numerical variables

By using Pair Plot to draw relationship between all the numerical values. This chart shows and represents the relationship between all numerical variables.

This Chart helps in identifying the patterns, relationships, analysis.



15.) Explore the correlation between all numerical variables

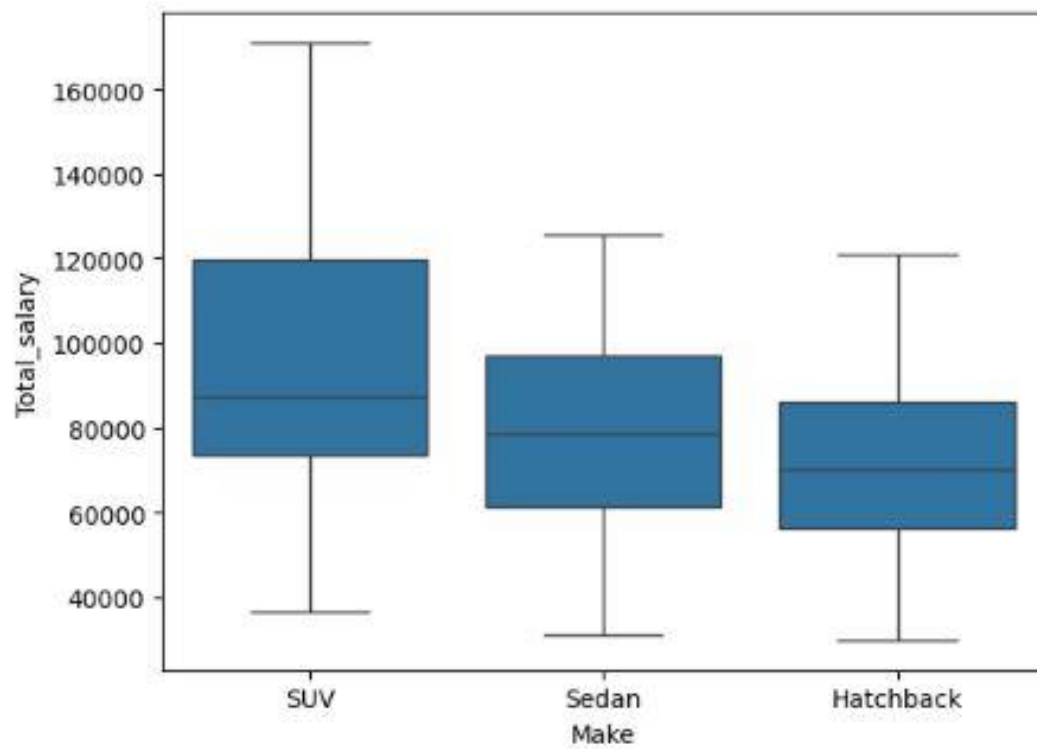


16.) Explore the relationship between categorical vs numerical variables

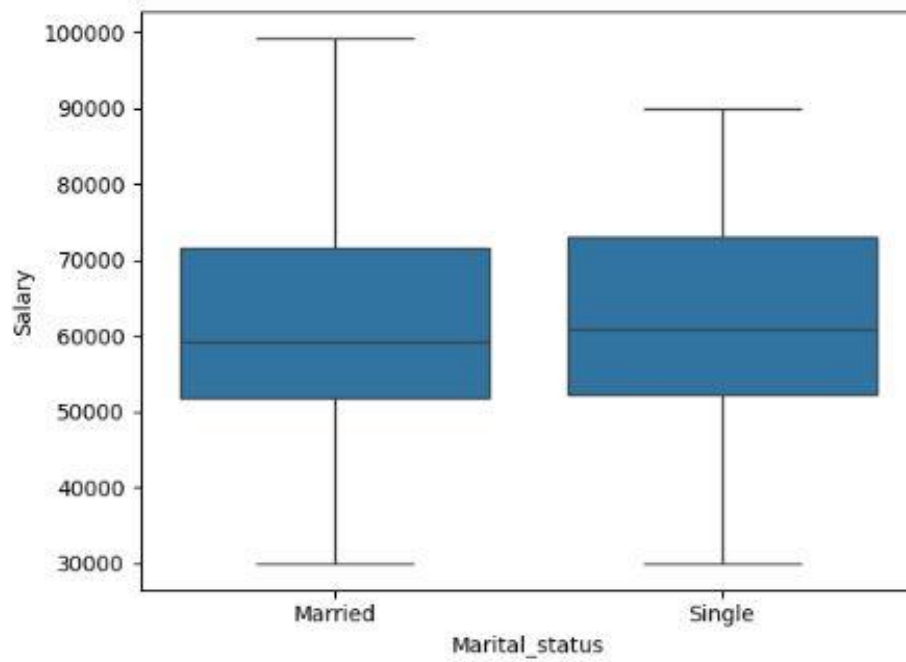
Co-related a few important Categorical and numerical variables by using Box Plot to explore these relationships

- Total Salary Vs Make
- Marital Status Vs Salary
- Education Vs Salary
- Personal Loan Vs Total Salary
- House Loan Vs Price
- Gender Vs Age
- Car Choice Across Age Groups

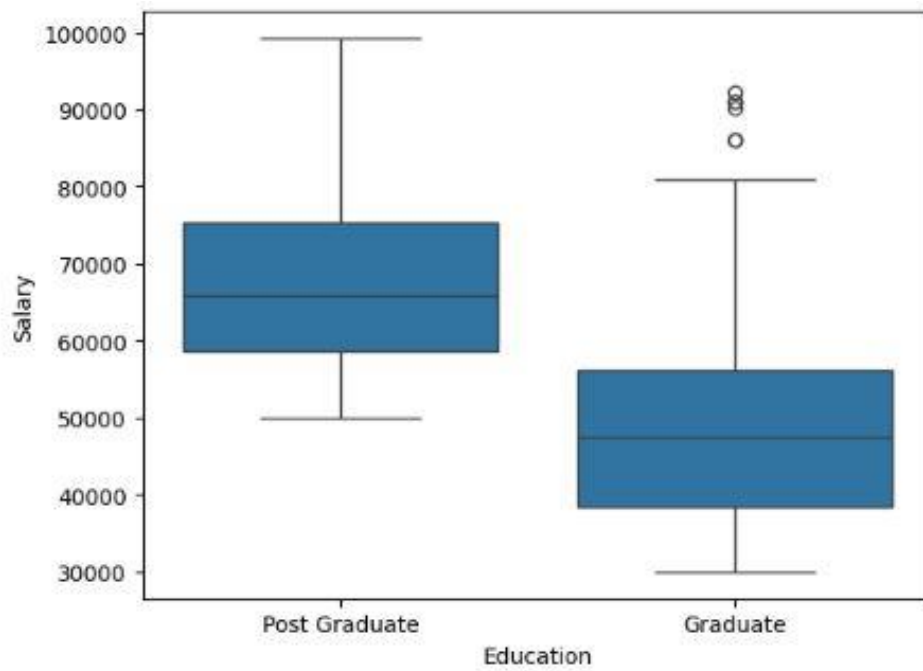
Box Plot of Total Salary by Make



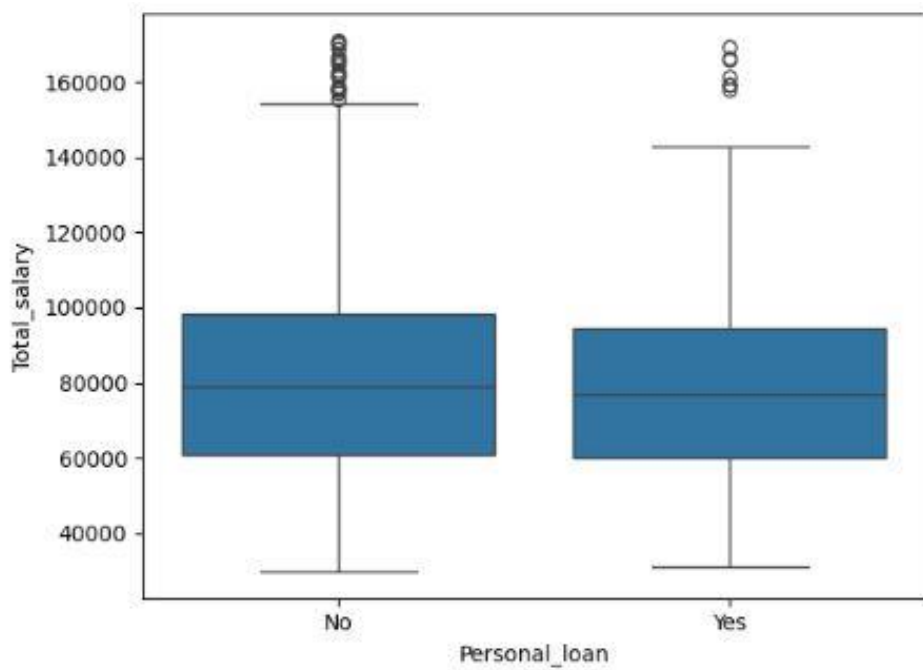
Box Plot of Salary by Marital Status



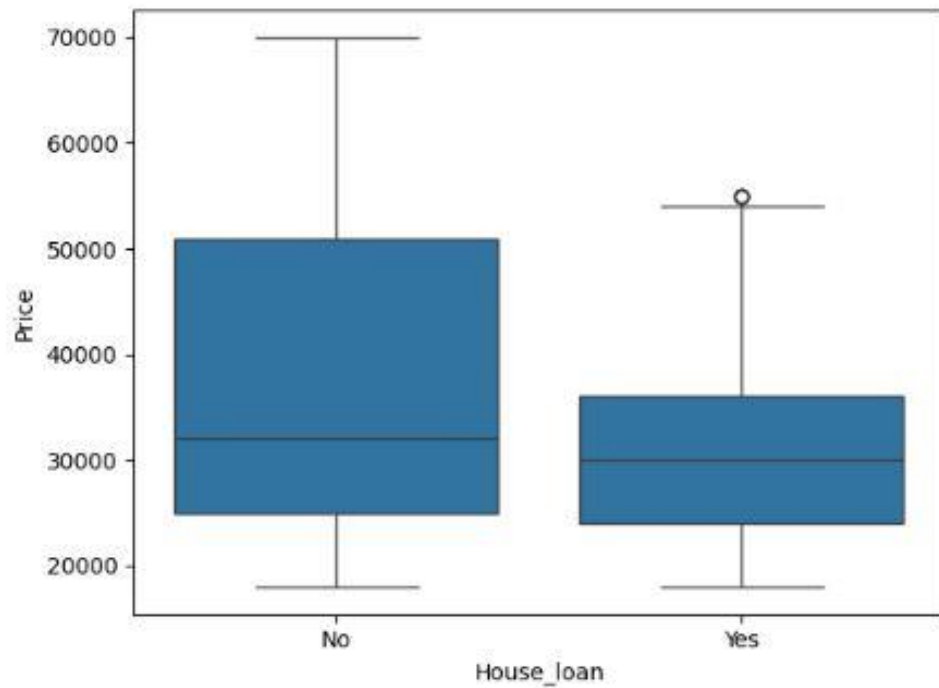
Box Plot of Salary by Education



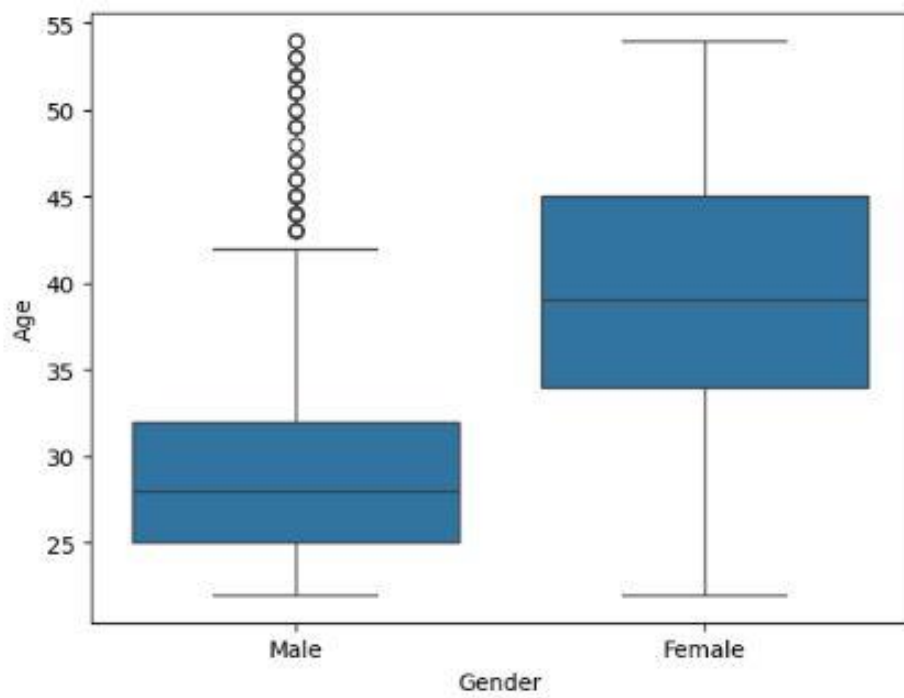
Box Plot of Total Salary by Personal Loan



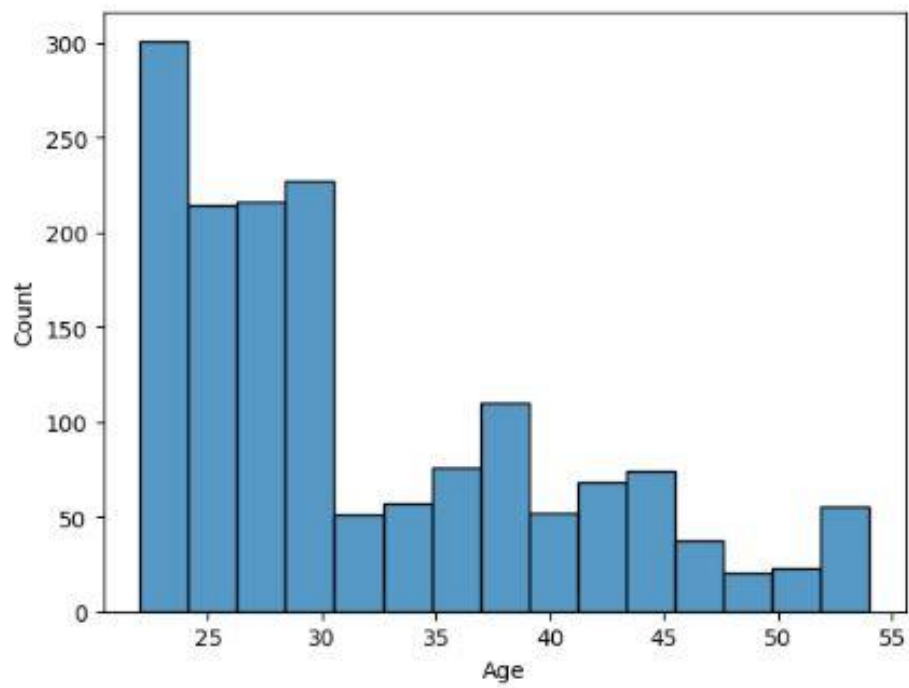
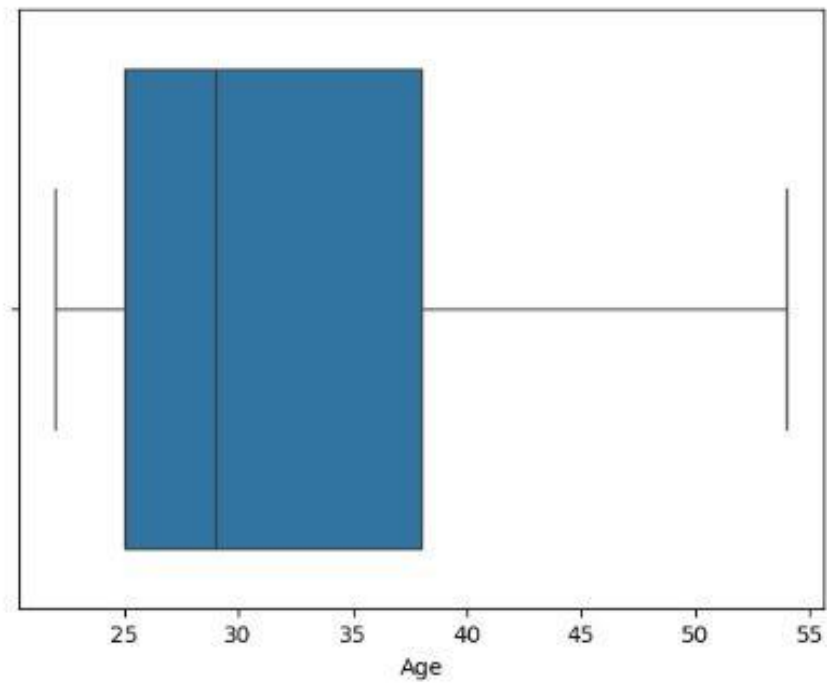
Box Plot of Price by House Loan

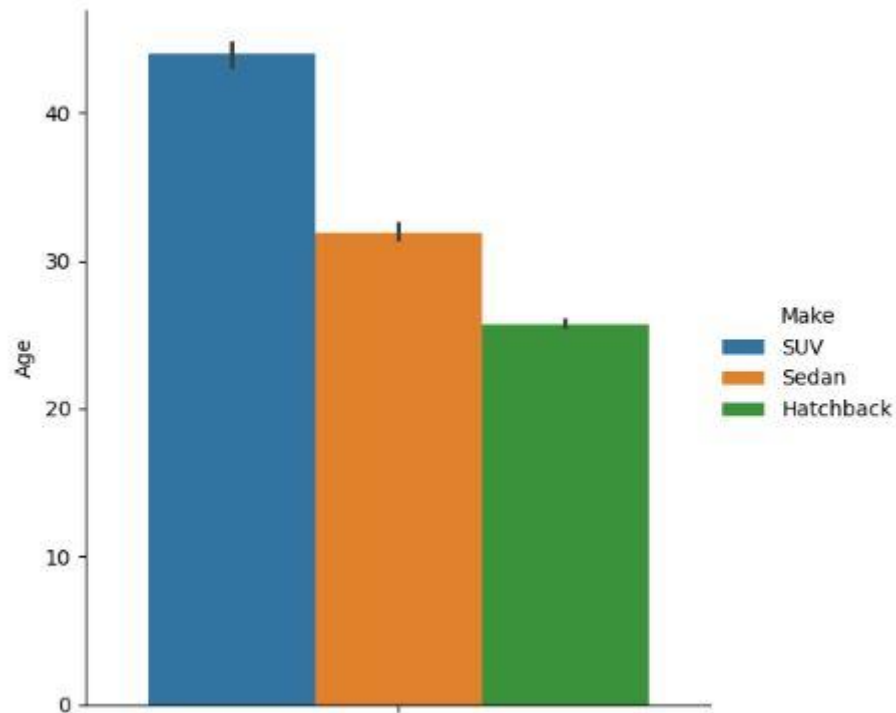


Box Plot of Age by Gender



Car Choices across Age Groups

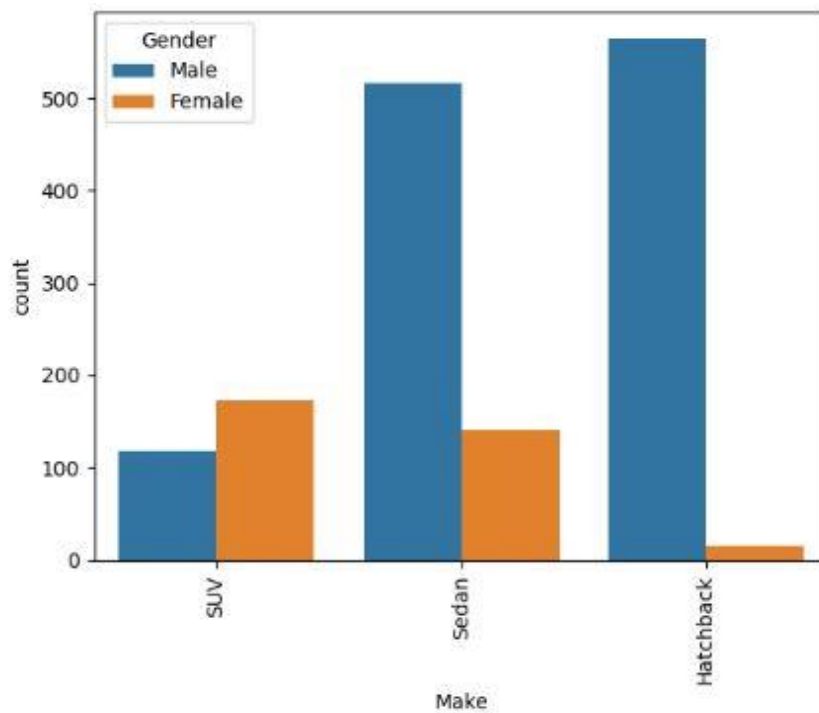




Key Questions

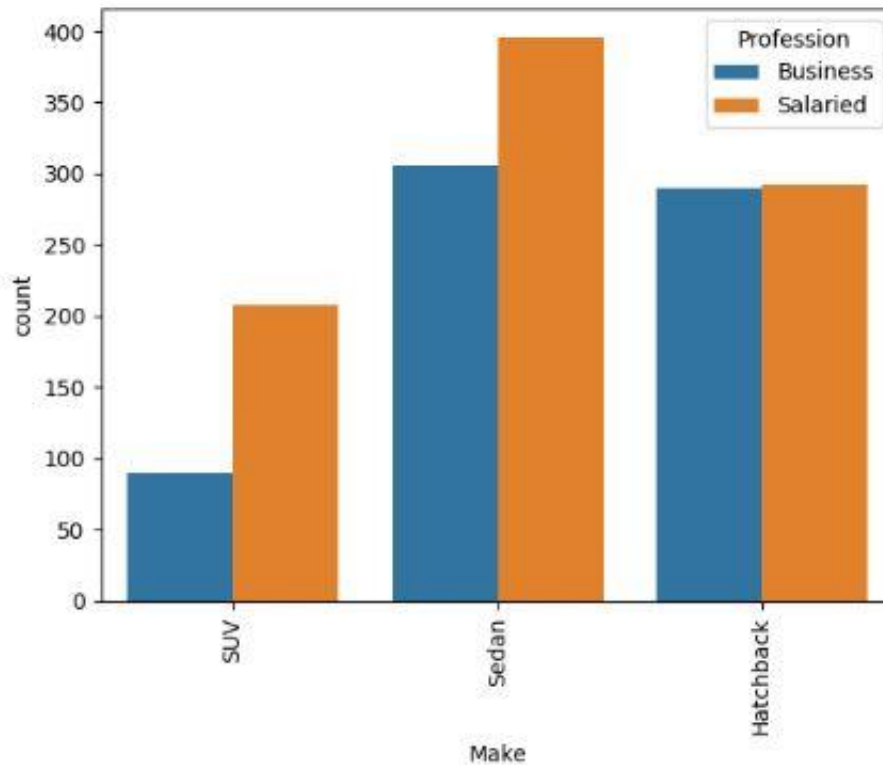
17.) Do men tend to prefer SUVs more compared to women ?

No, As per the Graph shown below, Female tend to prefer SUVs more as compared to men



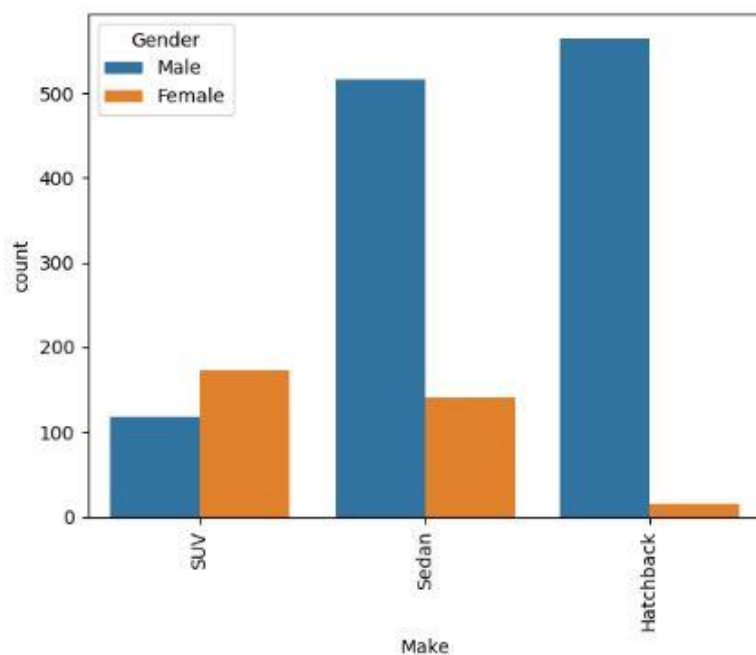
18.) What is the likelihood of a salaried person buying a Sedan ?

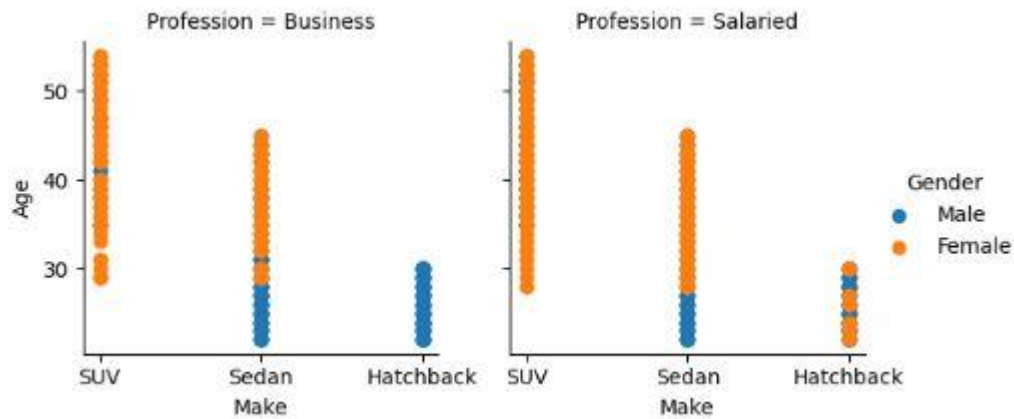
As per the graph shown Below, Salaried person is more likely to buy Sedan.



19.) 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 ?

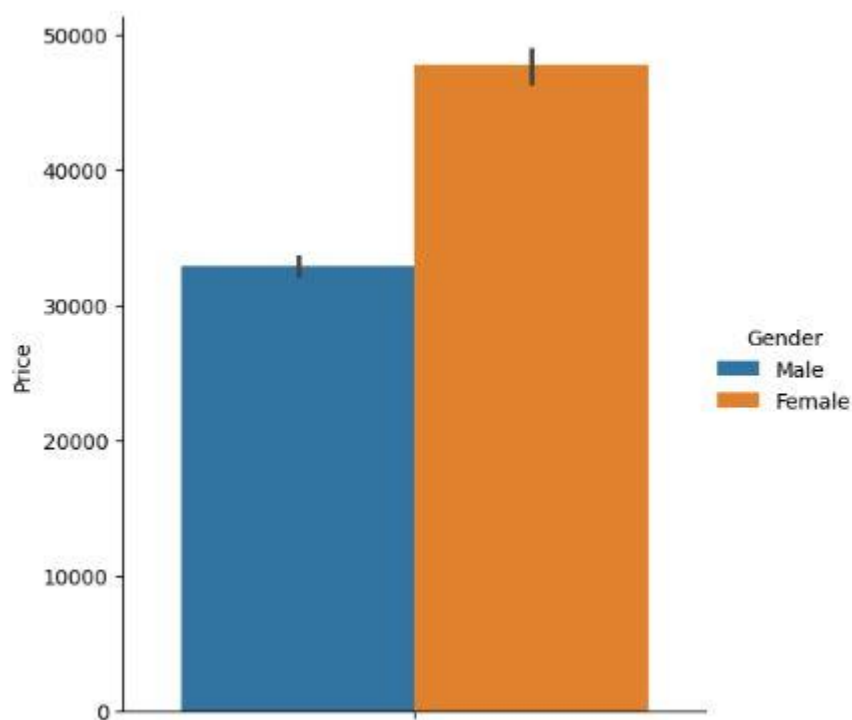
As per the graph shown below, SUV's are more preferred by Salaried women as compared to Salaried Men.





20.) How does the amount spend on purchasing automobiles vary by gender ?

As per below mentioned graph, it suggests that Female are spending more money on purchasing Automobiles as compared to Male



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

Below attached data shows the average amount of money spent by individuals who took Personal loan.

Price	
Personal_loan	
No	36742.712294
Yes	34457.070707

dtype: float64

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

Data Shows that there is a moderate influence for purchase of higher-priced cars

Price	
Partner_working	
No	36000.000000
Yes	35267.281106

dtype: float64

Actionable Insights & Recommendations

23.) **Actionable Insights & Business Recommendations**

- Demographic Analysis

Insight : This Dataset is Majorly in favour as a Male Salaried dominant and married people, Most of them have Post Graduate Degree.

Recommendation : Need to make a plan to reach out to more single people for buying a car and Hatchback can be promoted for salary Professionals and Sedan can be promoted more to increase more revenue.

- Income and Affordability

Insight : Most people have working partners with partner earning less. Data shows people have taken Personal loans more than House Loans

Recommendation : We have seen from the data that people with working partners are not spending much on buying higher prices cars. Hatchbacks with higher prices can be recommended to them considering their high income, Dependents and marital status.

- Loan Analysis

Insight : Many of the people in this data set have taken personal loans. Data shows that people have taken Personal loans more as compared to House loans

Recommendation : collaboration with financial organizations to offer attractive loan packages with less interest or any other facilities for car buyers with lesser loan liabilities.

- Gender Analysis

Insight : It has been observed that Female prefers less buying cars as compared to male

Recommendation : We can target Salaried male or female, depend upon the place of their organization, if they can communicate by car.

- Make Preference Analysis

Insight : We have seen that sedan is the most preferred choice followed by hatchback and SUV is the least Preferred.

Recommendation : Need to target audience for SUV's Sales by creating Digital marketing strategies by customized offers and other Discounted Liabilities.

- Professional Analysis

Insight : It has been observed that Business person are Less preferring buying any of the cars including SUV and Sedan as compared to Salaried person, But in terms of Hatchback its almost same.

Recommendation : Need to target the Business owners with Affordable Hatchback in the budget or according to their need and preference including some business benefits, if they buy with their company name.

- Choice Analysis

Insight : According to Analysis, it has been observed that Male prefers buying Sedan and Hatchback more & Female prefers buying SUV more as compared to Men.

Recommendation :When we target the audience we should keep in mind while doing online campaigns, We can target more SUV to female, and target Sedan, Hatchback to male, in order to get more revenue and sales.

THANK YOU