greatlearning
Learning for Life

Business report

Austo automobile case study

DATA SCIENCE AND BUSINESS ANALYTICS

greatlearning
Learning for Life

Content:

greatlearning
Learning for Life

1. Data overview

1.1.Check the data structure

	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

greatlearning
Learning for Life

Understanding the data structure:

- Above are first 5 columns of the dataset
- The *Age* column contains the age of customers who bought car.
- The *gender* column and the *marital status* column include the gender and Marital status of the buyer.
- The *profession* and *education* column includes profession and education of the buyer.
- *No_of_Dependents* column includes number of persons who are dependent on buyer.
- *Personal_loan* and *House_loan* includes loans taken by the buyers.
- *Partner_working* column includes whether the partner of buyer is working or not.
- *salary , partner_salary* and *total_salary* column includes salary of buyer, salary of partner and the total salary of both the partners.

greatlearning
Learning for Life

1.2. Check the data types of the column for the dataset

Range Index: 1581 entries, 0 to 1580, Data columns (total 14 columns):

- # Column Non-Null Count D type
- 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

Understanding the data type:

- All the columns have 1581 observations except Gender and Partner_salary which 1528 and 1475 observation indicating that there are some missing values in them
- There are 6 numeric columns in the dataset.
- And there are 8 object type columns in the dataset
- Data types of the columns are correct.

Understanding the shape of the dataset

- There are 1581 rows and 14 column

greatlearning
Learning for Life

1.3. Checking the statistical summary of the data

	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

Understanding the statistical summary of the data:

- **Age** : On an average there is 29 years , whereas ,75% age of the people is 38 years and the maximum age of the clients is less than or equal to 54 years.
- **No of dependents** : on an average there are 2 people who are dependents , 75% of the people having 3 members, whereas maximum number of dependents are 4 members
- Distribution of salary, partner_salary and total salary are fine.
- On an average the price of the cars is 31000 dollars and 75% is 47000 dollars whereas maximum price range of the cars is 70000 dollars.

Checking for duplicate entries in the dataset

- There are no duplicate entries in the dataset.

1.4 Checking for missing values in the dataset.

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

- There are missing values in 2 columns of the data.
- The variable, Gender have 53 missing values in them.
- Partner's salary has 106 missing values in them.
- We have used mode to fill missing values of Gender column
- We have used formula (Total salary = partners salary + salary) to fill partners salary column.

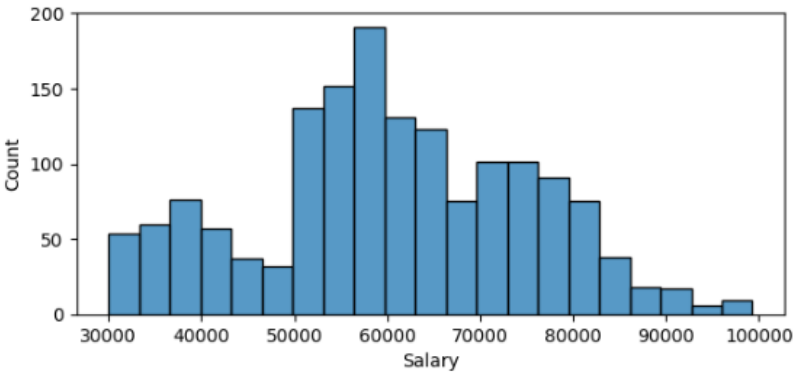
greatlearning
Learning for Life

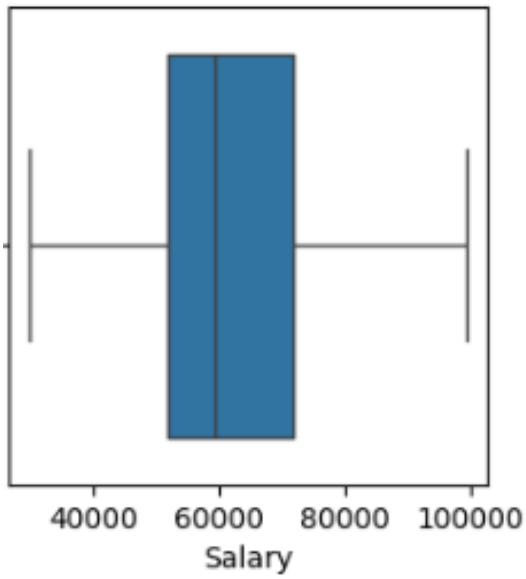
2. Univariate analysis

Univariate data visualization plots help us comprehend the descriptive summary of the particular data variable. These plots help in understanding the location/position of observations in the data variable, its distribution, and dispersion.

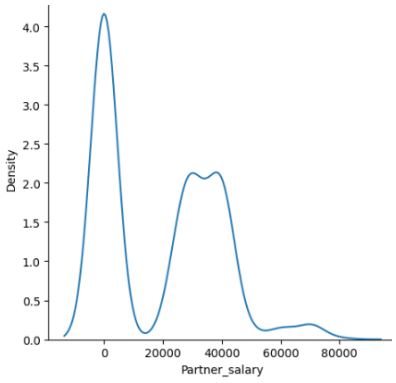
2.1. Exploring all variables of the data set:

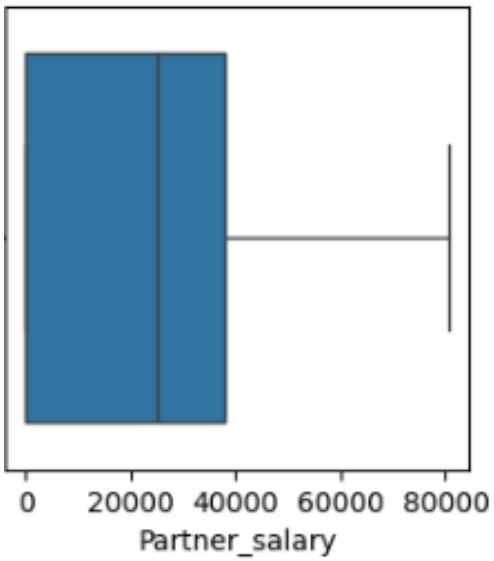
a) Salary





b) Partner's salary



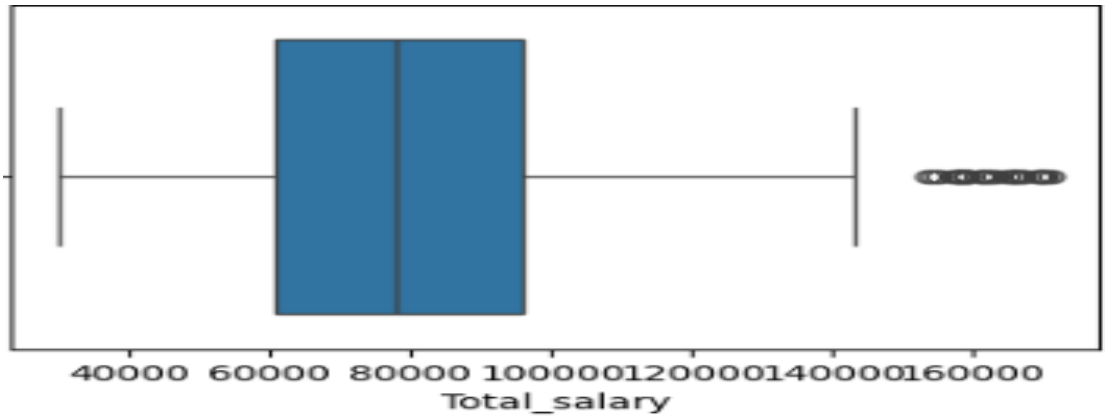


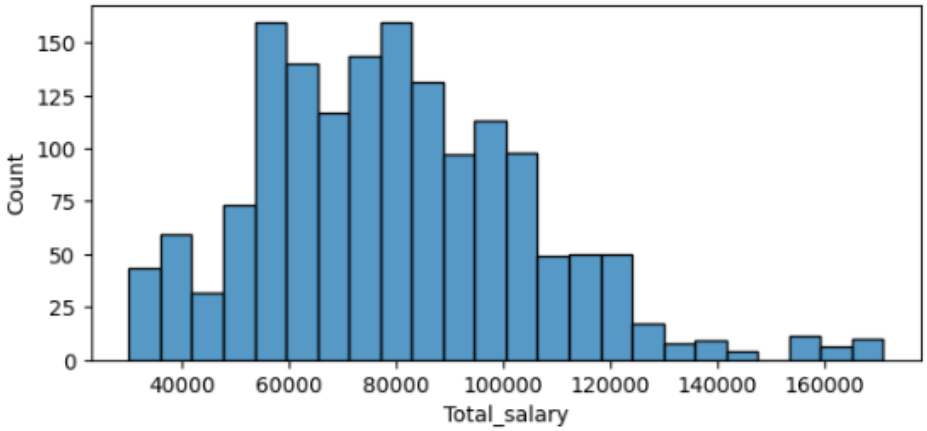
Observations:

- More number of car buyers is receiving salary between 55000 to 60000 dollars.
- The people who are having a salary range between 95000 to 100000 dollars are not much spending their money on cars.
- There is no outliers
- And the median is at 60000 dollars
- Partners of the buyers are getting below 10000 dollars.

In this data, there are variables like **total salary, price, number of dependents, make** etc that affects the number of selling.
Let us explore some of the variables and how they are distributed.

c) Total salary:

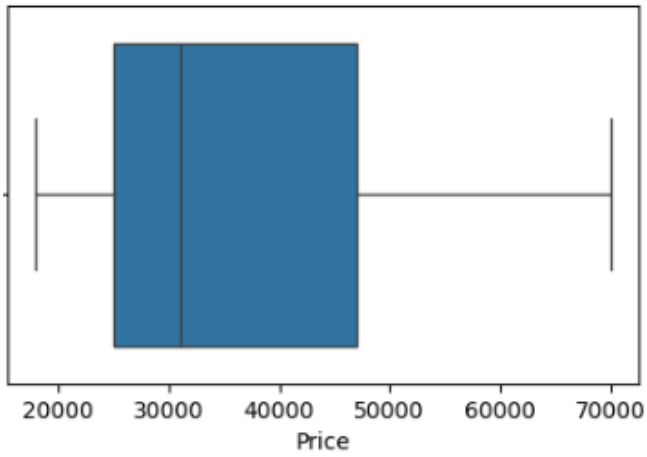


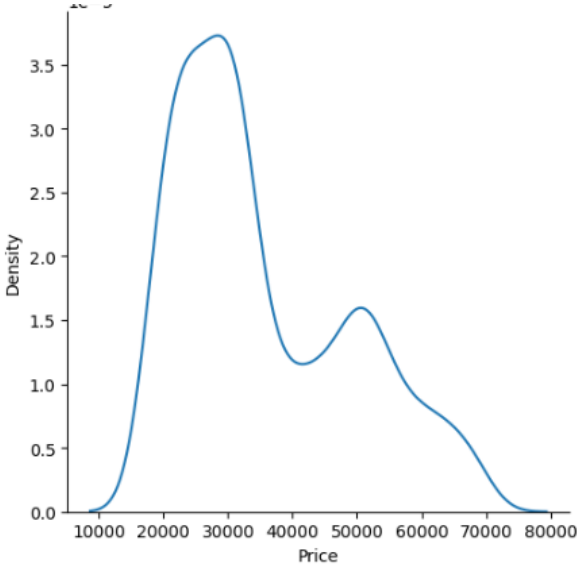


Observations:

- Most of the cars are bought by those who are getting salary around 60000 and 80000
- Outliers are there in the right
- Here, Salary above 140000 dollars is considered as outliers.

d) Price:

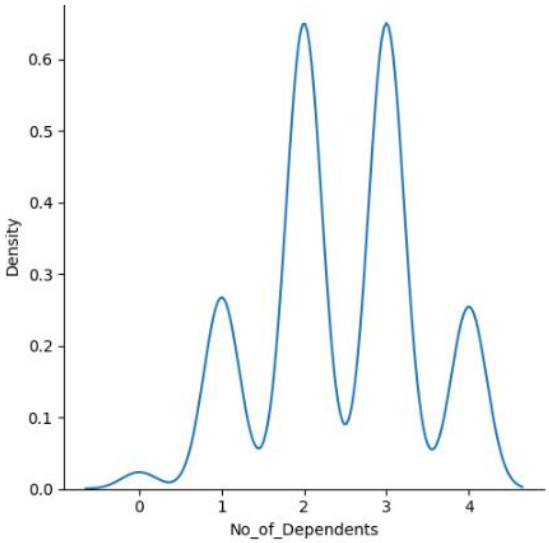


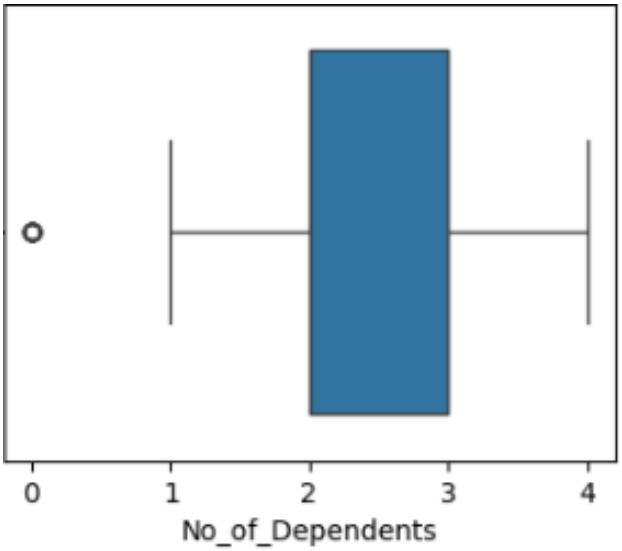


Observations:

- Minimum price of the cars is 20000 dollars which bought more as compared to other cars.
- Maximum price of the cars is 70000 dollars which bought lower. Maximum sales are between 25000 to 35000 dollars price.

e) Number of dependents:

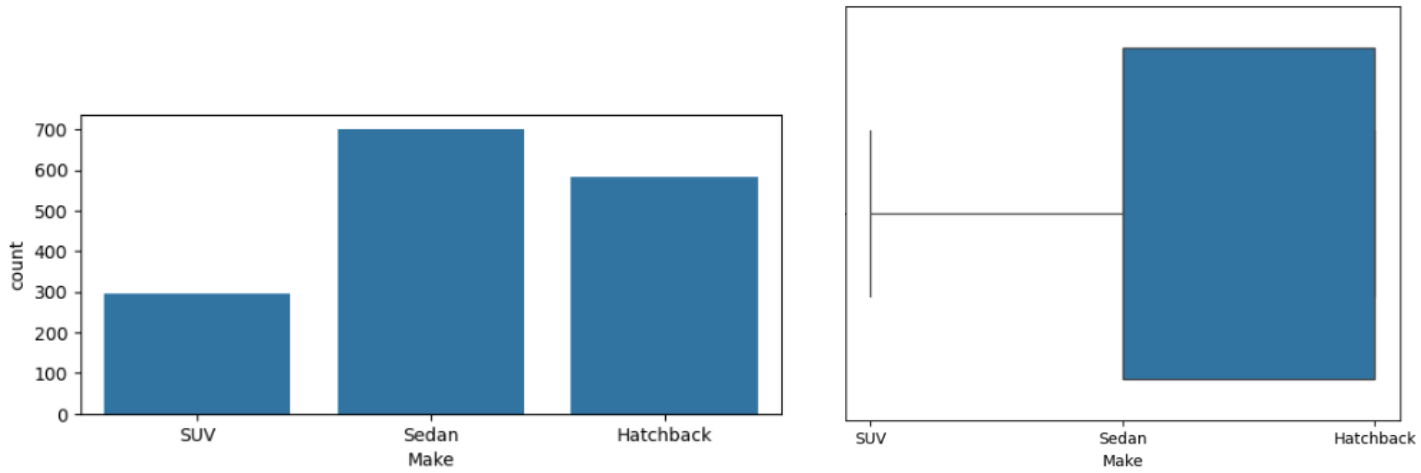




Observation:

- Some of the clients have zero dependence on them.
- Those who are buying the cars more having 2 and 3 dependents on them.
- Maximum numbers of dependence on the car purchaser are 4 members

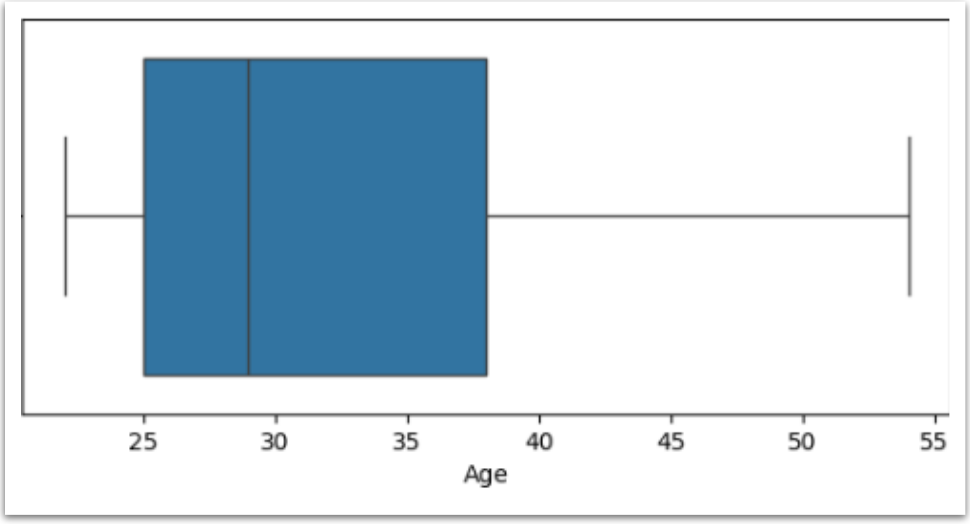
f) Make:

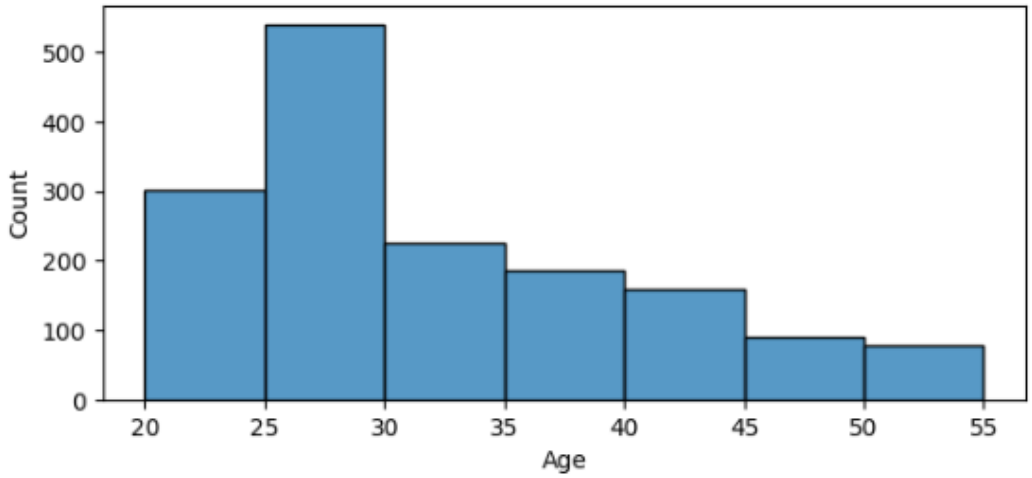


Observations:

- Sedan cars are most preferable than the SUV and the hatchback.
- Sedan car selling is around 700 whereas SUV and hatchback are around 300 and 600 counts.

g) Age:



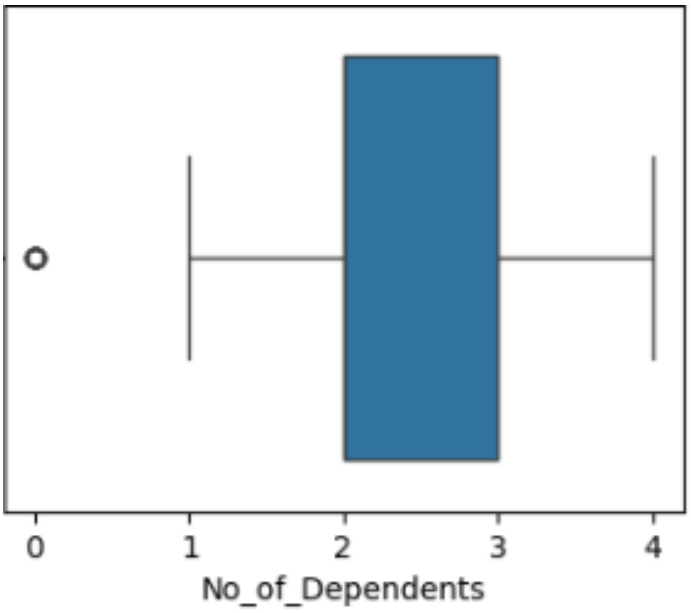


Observations:

- The distribution is skewed towards right
- The cars are bought by the people who are between the age 25 to 30 years
- Less number of cars bought by the people at the age of 50.
- There no outliers
- Median is between 25 years to 30 years range

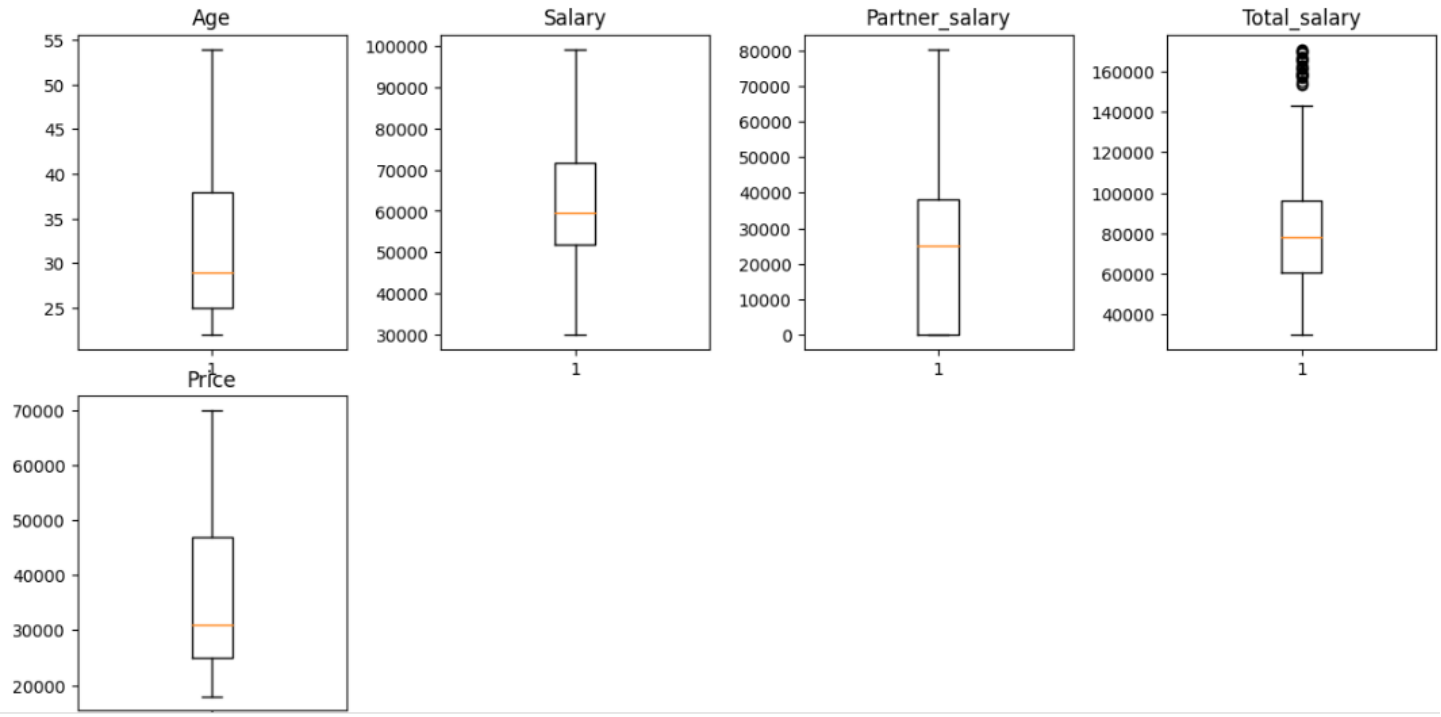
2.2 checking for and treating Outliers: - observations and insights:

a) Number of dependents:



- We are not going to treat outliers because it seems to be genuine and we can miss out any of important information by treating genuine information.
- Many people might not have any dependents on them, that is why we are not going to treat outliers.

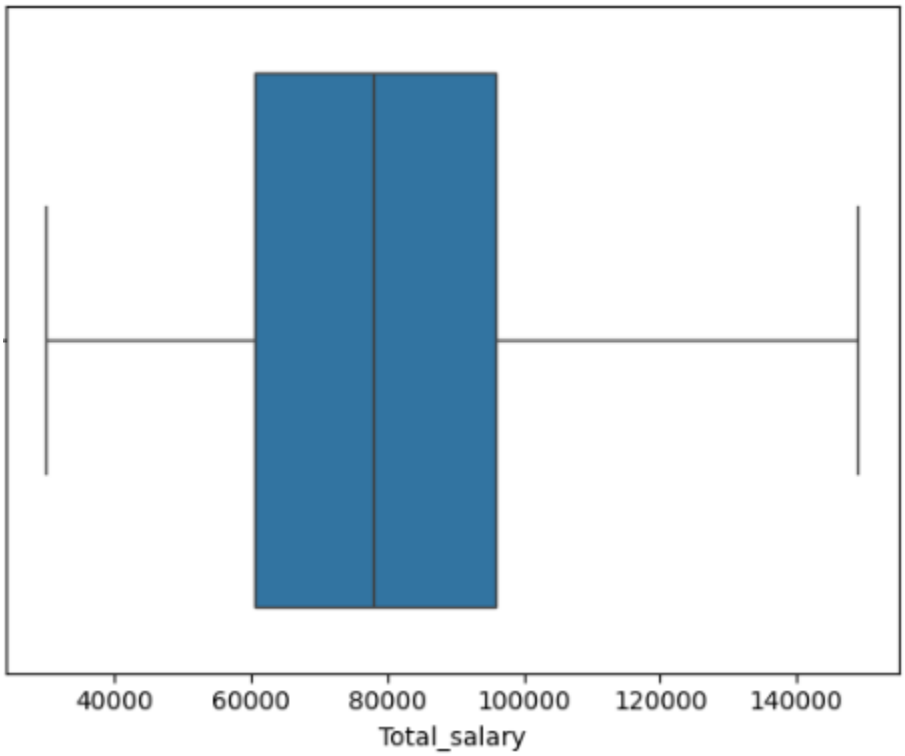
b) Checking outliers of numerical variables:



Observations:

- Total salary column has a wide range of values with lots of outliers. However we should treat these outliers.
- Except total salary there are no outliers in any other variable.

Outliers of total salary is treated, after the treatment we got the below plot

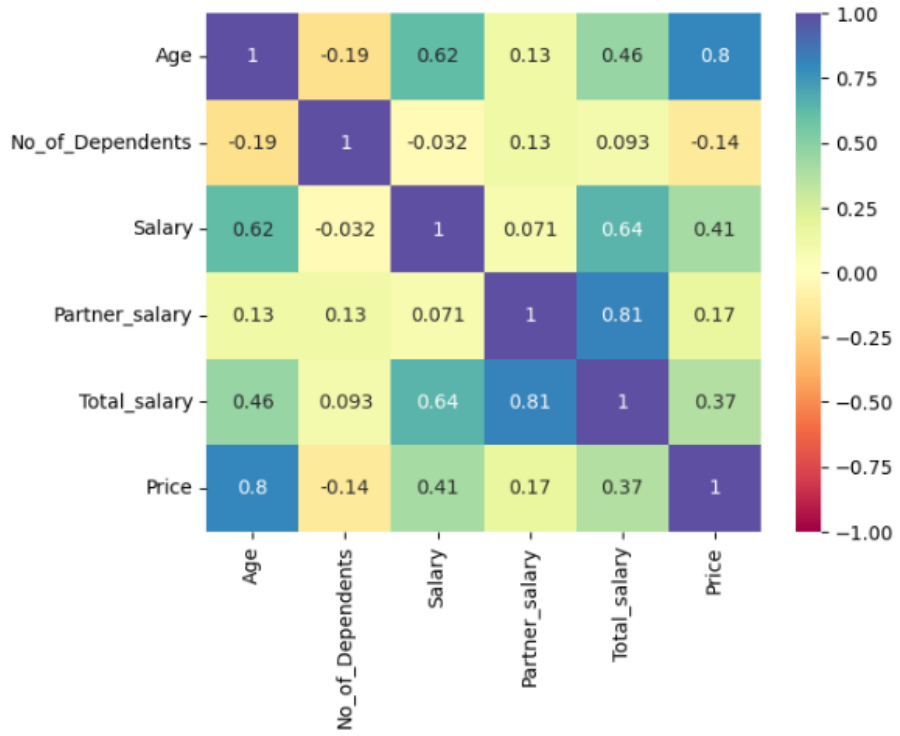


- Successfully removed the outliers of total salary.

greatlearning
Learning for Life

3) BIVARIATE ANALYSIS –

Correlation matrix:

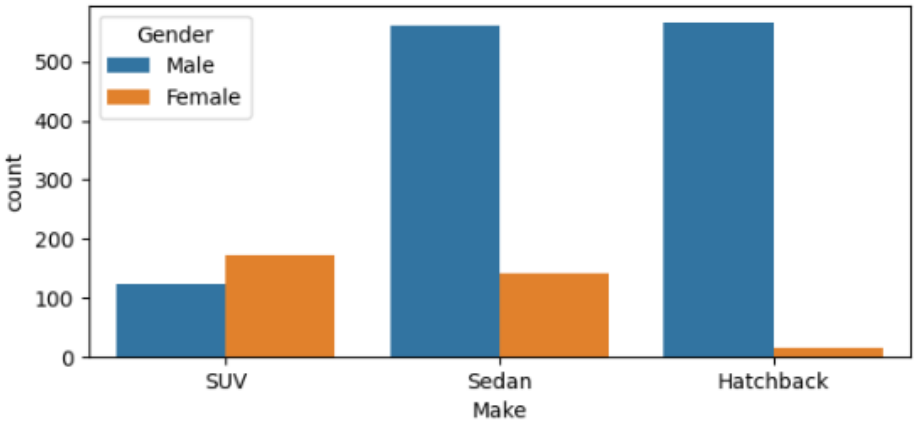


- Partner's salary has high correlation with total salary.
- Age negatively correlated with number of dependents which means when the age is high number of dependents is low.
- Price is also negatively correlated with number of dependents, which shows when numbers of dependents are less buyer purchase higher price cars which make sense too.
- Salary is negatively correlated with number of dependents.
- Age is positively correlated with price which shows when age increases capacity of buying higher price cars is also increases.
- Salary is positively correlated with total salary, when salary increases total salary is also an increase which is expected.

greatlearning
Learning for Life

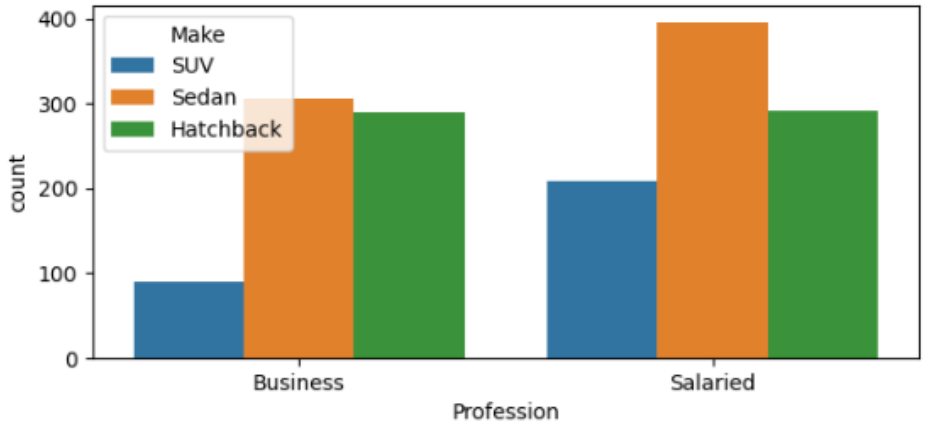
4)KEY QUESTIONS with answers through BIVARIATE ANALYSIS

1) Do men tend to prefer SUVs more than compared to women?



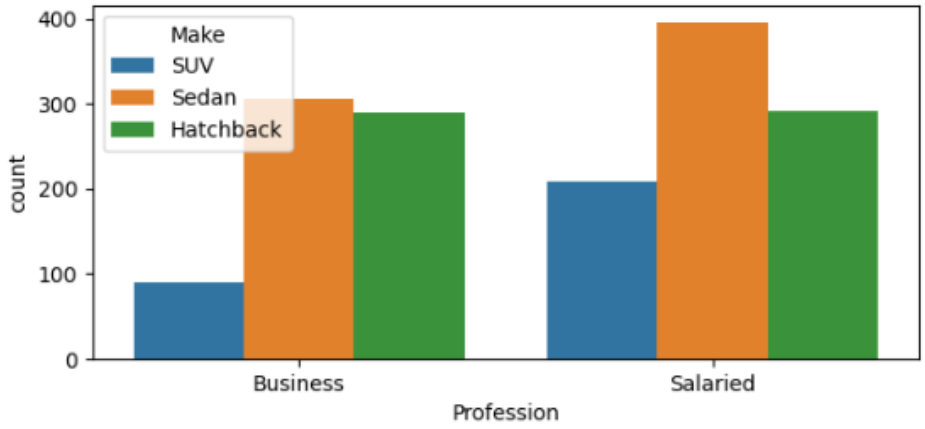
From the above graph more preferable cars for men are hatchback and Sedan than the SUV and for women SUV is more preferable than the other cars.

2) What is the likelihood of a salaried person buying a sedan?



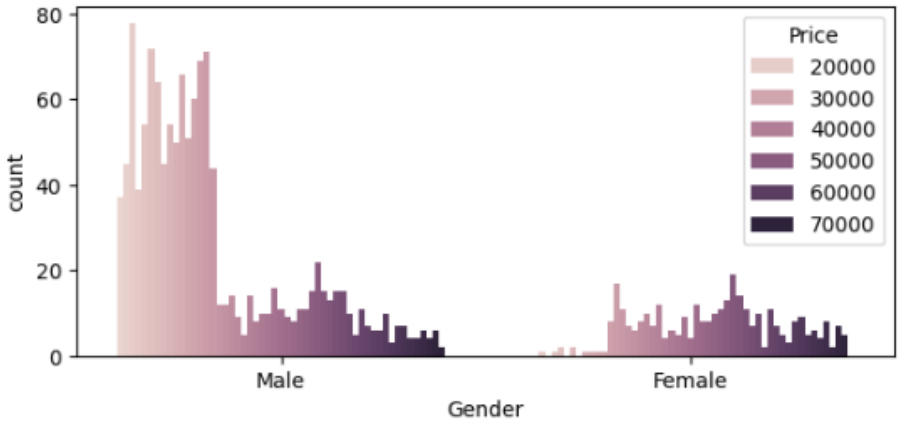
From the above graph sedan is more preferable by salaried person after sedan hatchback is preferable. Sedan is more preferable by salaried ones than the businessmen.

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?



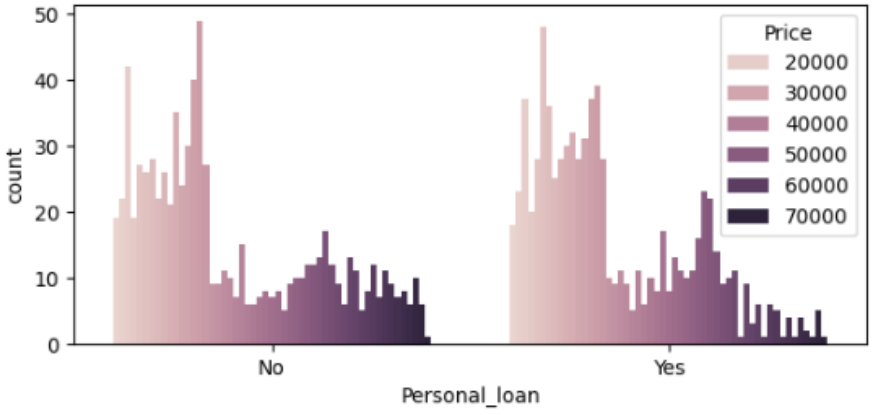
As said above , From the above graph sedan is more preferable by salaried person after sedan hatchback is preferable. Sedan is more preferable by salaried ones than the businessmen. SUV is not much preferable. So salaried male is not an easier target for a SUV sale over sedan sale.

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



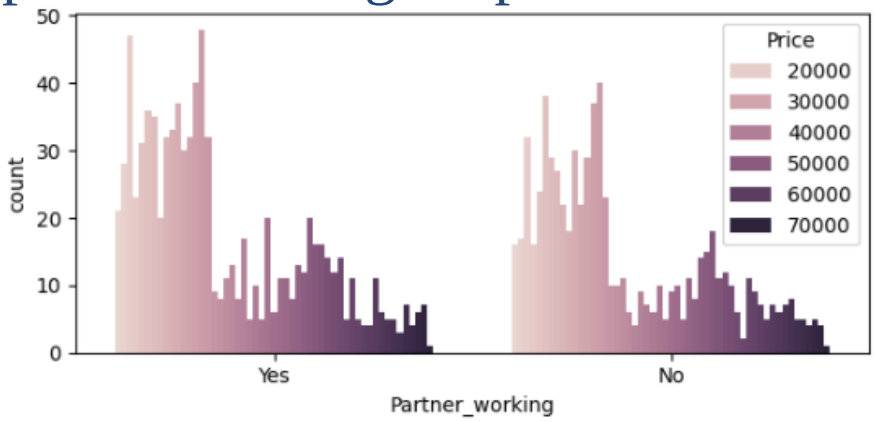
- Among number of buyers males are dominant in spending on cars rather than females
- Female buying cars are rare as compared to male.
- Males buy more cars from 20000 to 40000.
- Selling cars for male from 40000 to 70000 is as similar as female.

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



- Numbers of buyers who have personal loan are as same as the buyers who don't have any personal loan.
- Those who have loan buy most of the cars of the price 20000.
- Whereas the people who don't have any personal loan buy most of the cars of price 30000.

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



- People who have working partner tend to buy cars and also tend to buy higher priced cars, which is expected.
- Female who has a working partner buy cars of the price around 50000 to 60000.
- Male who has a working partner but does not spend amount on buying of cars around 60000 to 70000 but female spend most than the male according to above plot.

5) Actionable insights and business recommendations.

Actionable Insights:

We analyzed a dataset of Austo Motor Company which is a leading car manufacturer specializing in SUV, Sedan, and Hatchback models to get a fair idea about demand of customers which will help them in enhancing their customer experience. Both from an environmental and business perspective, having cars which has less demand or having no cars which has higher demand is inefficiency. Thus we determined the factors that affect sales and the nature of their effect.

We have been able to conclude that-

1. Demand for the lesser priced cars is higher than the high priced cars.

2. People buy more cars at the age of 25 to 30.While age increases the demand for cars decrease. Buying hatchback is less and constant between the age of 20 to 30 years and price is between 25000 to 30000 .When the age is between 20 to 30 years old the buying of sedan becomes constant which is of 25000 .However the age increases the buying capacity of the buyers is also increases and buys sedan more than the hatchback. Buying SUV begins at the age between 25 to 30 years .As said before buying capacity of buyer is increasing while the age increases.

3. Increasing in salary increases in net spending on cars which is expected.

4. When a person has 2 or 3 dependents on him or her, then he or she decides to buy a car, which makes sense. When the depends is less buyers go for sedan more than hatchback and SUV .When increase in the number

of dependents ,increase in buying of hatchback and SUV. When increase in salary and increase in number of dependents, increases in the buying of SUV and sedan.

5. Among number of buyer males are dominant in buying cars rather than female

6. Cars bought by married ones than the singles.

7. Salaried people are more than the businessmen who are the buying cars.

8. People who are post graduate buys frequently then graduates.

9. Buyers who have house loan are less than the people who do not have any house loan.

10. Buyers who have house loan are less than the people who do not have any house loan.

11. Sedan cars are most preferable than the SUV and the hatchback. Female prefer SUV cars than the sedan and the hatchback. Male prefer more sedan later the hatchback and the SUV.

12. Salaried person prefer sedan cars more than SUV and hatchback.

13. Those who have loan buy most of the cars of the price 20000.Whereas the people who don’t have any personal loan buy most of the cars of price 30000.

14. Those who have working partner buy higher priced cars as compared to those whose partners are not working.

greatlearning
Learning for Life

Business Recommendations:

1. Demand for the lesser priced cars is more than the higher priced cars ,so should focus on lesser priced cars.

2. Salaried couples who are post graduates buys car the most so should focus on them to increase sales.

3. Buying hatchback is less and constant between the age of 20 to 30 years and price is between 25000 to 30000, so to sell hatchback company should focus who are at the age of 20 to 30 years.

4. Men who doesn’t have house loan tend to buy cars more than the women so should focus on men.
