DSBA

PDS Project Coded - Python for Data Science

Austo Motor Company Data Analysis

Business Report

By: Sagar R Naivaruni

Contents

Description

Context

Analysts are required to explore data and reflect on the insights. Clear writing skill is an integral part of a good report. Note that the explanations must be such that readers with minimum knowledge of analytics are able to grasp the insight.

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.

Objective

They want to 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 description

- **Age:** The age of the individual in years.
- **Gender:** The gender of the individual, categorized as male or female.
- **Profession:** The occupation or profession of the individual.
- **Marital_status:** The marital status of the individual, such as married &, single
- **Education:** The educational qualification of the individual Graduate and Post Graduate
- **No_of_Dependents:** The number of dependents (e.g., children, elderly parents) that the individual supports financially.
- **Personal_loan:** A binary variable indicating whether the individual has taken a personal loan "Yes" or "No"

- **House_loan:** A binary variable indicating whether the individual has taken a housing loan "Yes" or "No"
- **Partner_working:** A binary variable indicating whether the individual's partner is employed "Yes" or "No"
- **Salary:** The individual's salary or income.
- **Partner_salary:** The salary or income of the individual's partner, if applicable.
- **Total_salary:** The total combined salary of the individual and their partner (if applicable).
- **Price:** The price of a product or service.
- **Make:** The type of automobile

Data overview

Data structure of the Austo Motor Company dataset.

	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

- Dataset contains 1581 rows and 14 columns.

Datatypes in the dataset

```
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
```

- There are a few genders missing in the gender column.
- Values are also missing in the partner salary column.
- Both need investigation to find out why they are missing.

Missing values in the dataset

```
Missing Values:
Age                   0
Gender               53
Profession            0
Marital_status        0
Education             0
No_of_Dependents      0
Personal_loan         0
House_loan            0
Partner_working       0
Salary                0
Partner_salary      106
Total_salary          0
Price                 0
Make                  0
dtype: int64
```

- There are 53 missing values in Gender and 106 missing values in Partner_salary columns
- There are no duplicate values.

Wrong entries in gender

```
Gender
Male      1199
Female     327
Femal        1
Femle        1
Name: count, dtype: int64
```

Corrected entries in gender

```
Gender
Male      1199
Female     329
Name: count, dtype: int64
```

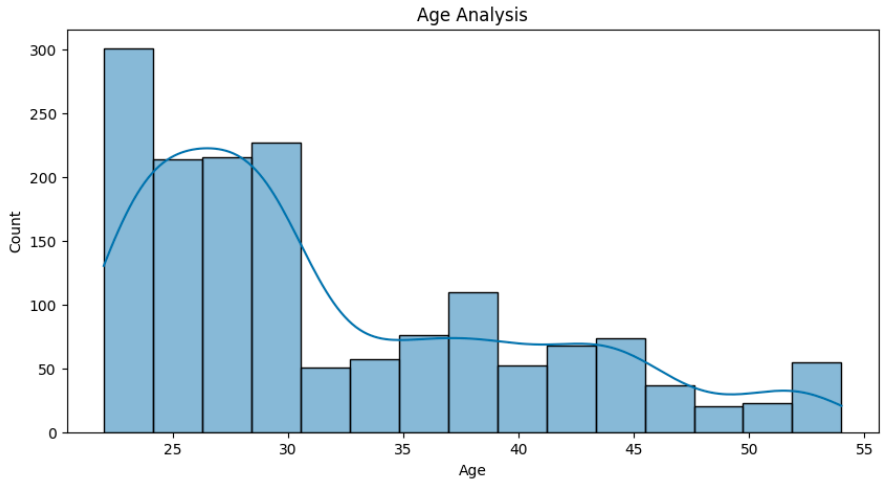
Statistical summary of the data

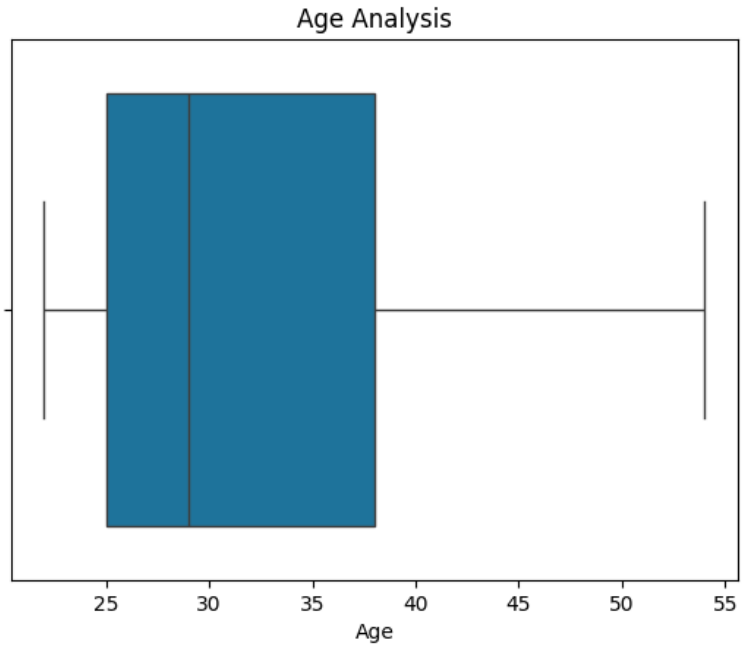
	count	unique	top	freq	mean	std	min	25%	50%	75%	max
Age	1581.0	NaN	NaN	NaN	31.922201	8.425978	22.0	25.0	29.0	38.0	54.0
Gender	1528	4	Male	1199	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Profession	1581	2	Salaried	896	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Marital_status	1581	2	Married	1443	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Education	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

- Here the mean age is 31.9 years and the maximum age is 54 years while the lowest age is 22 years.
- On an average there are 2.45 dependents.
- Average salary is $60,000 while the minimum salary is $30,000 and the maximum salary is $99,300.
- Average partner salary is $20,000 and the maximum salary is $80,500.
- Average total salary is around $80,000 and the maximum total salary is $171,000.
- Average price at which the vehicle is purchased is around $35,000 and the maximum price at which the vehicle is purchased is $70,000.

Univariate analysis

Age analysis

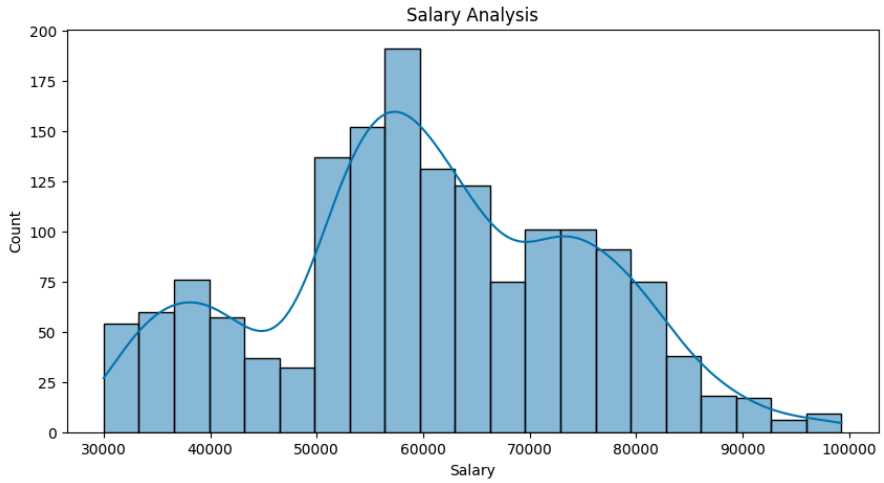


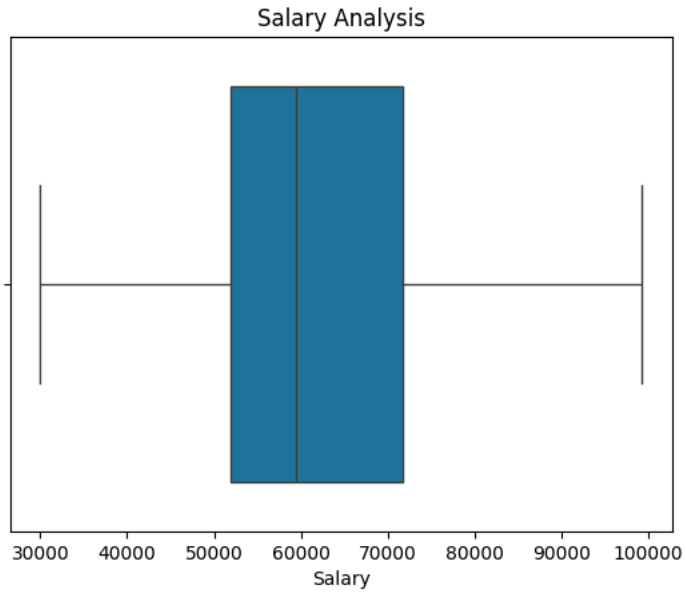


- The distribution of age is right skewed which means there are more young people than the old people.
- The median age is 29 years, meaning half of the people in the dataset are below 29 years of age and half are above 29 years of age.

- The age range is not mentioned but according to the boxplot we can estimate the age range between 20 years to 50 years of age.
- There are no outliers.

Salary analysis

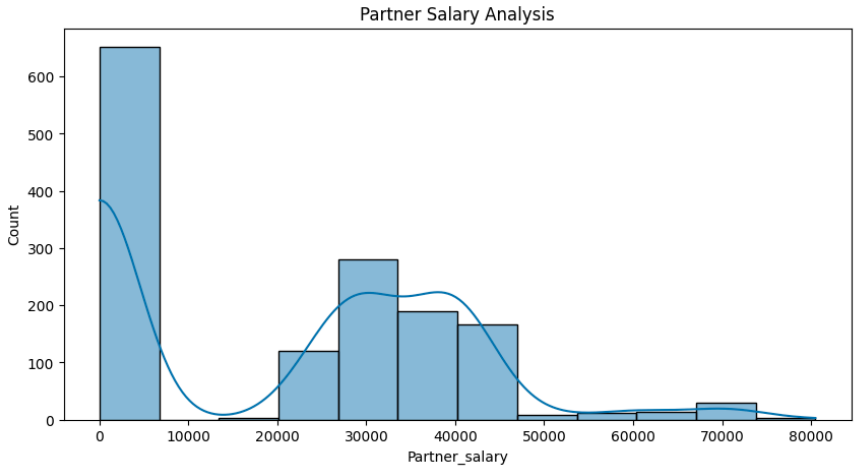


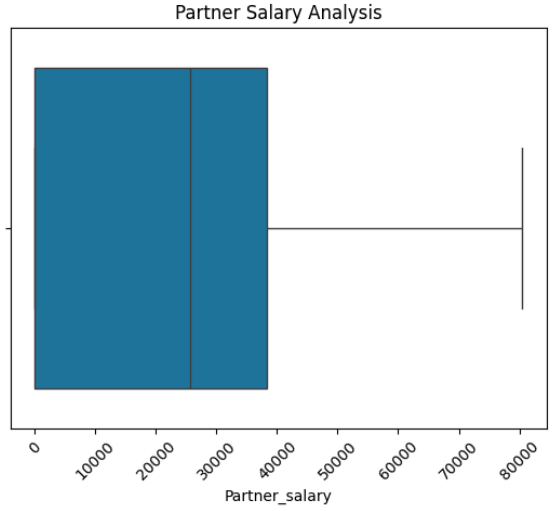


- The distribution of salary is right skewed meaning there are more individuals with lower salary compared to high salary individuals.

- The median salary is around 59000.
- There are no outliers.

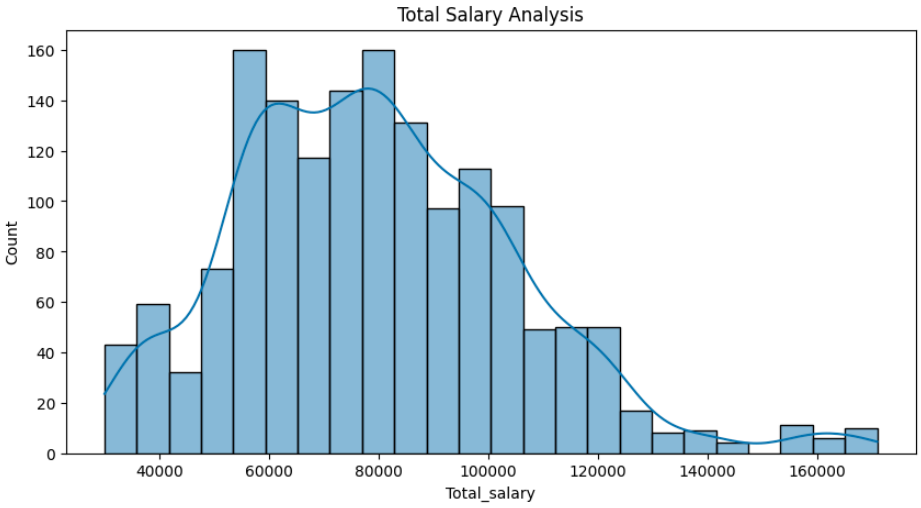
Partner salary analysis

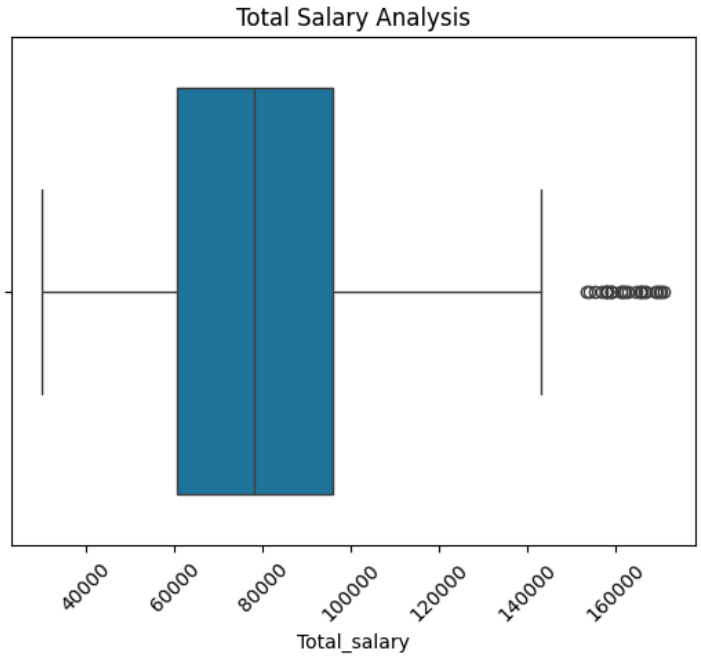




- The distribution of partner salary is right skewed.
- The median partner salary is around 26000
- There are no outliers.

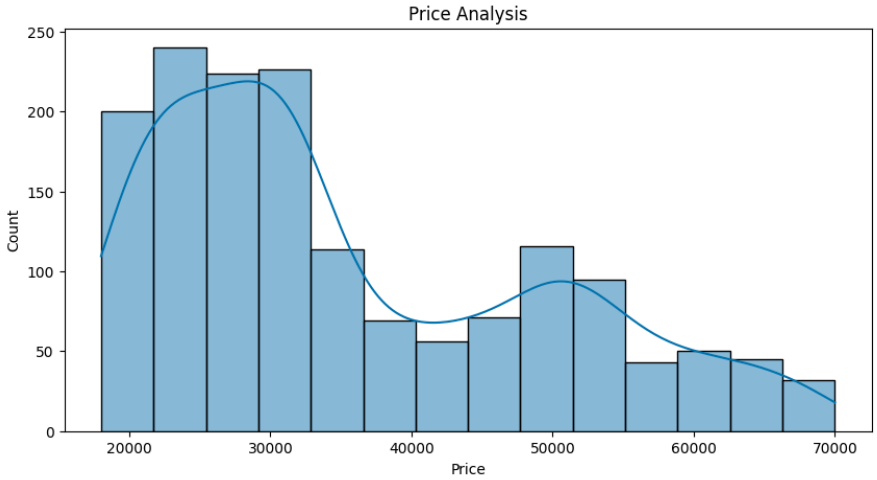
Total salary analysis

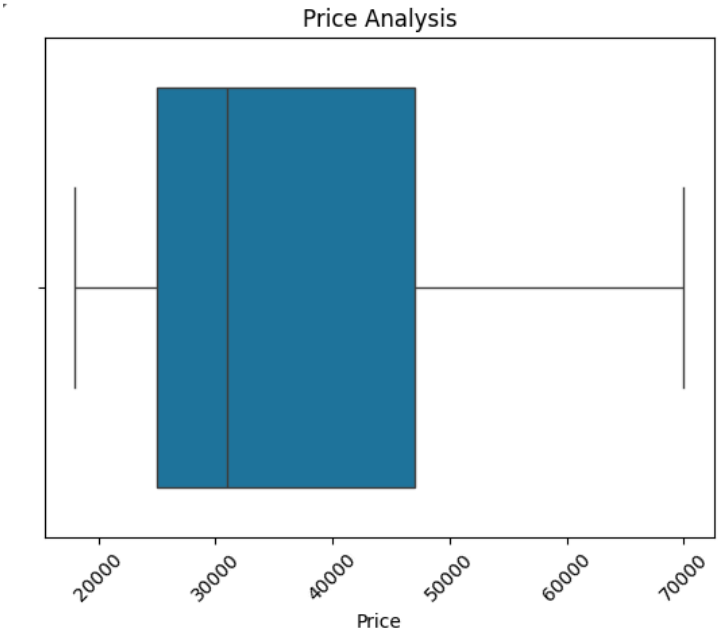




- The distribution of total salary is right skewed.
- The median total salary is around 78000.
- There are a few outliers indicating people with higher total salary compared to the majority.

Price analysis

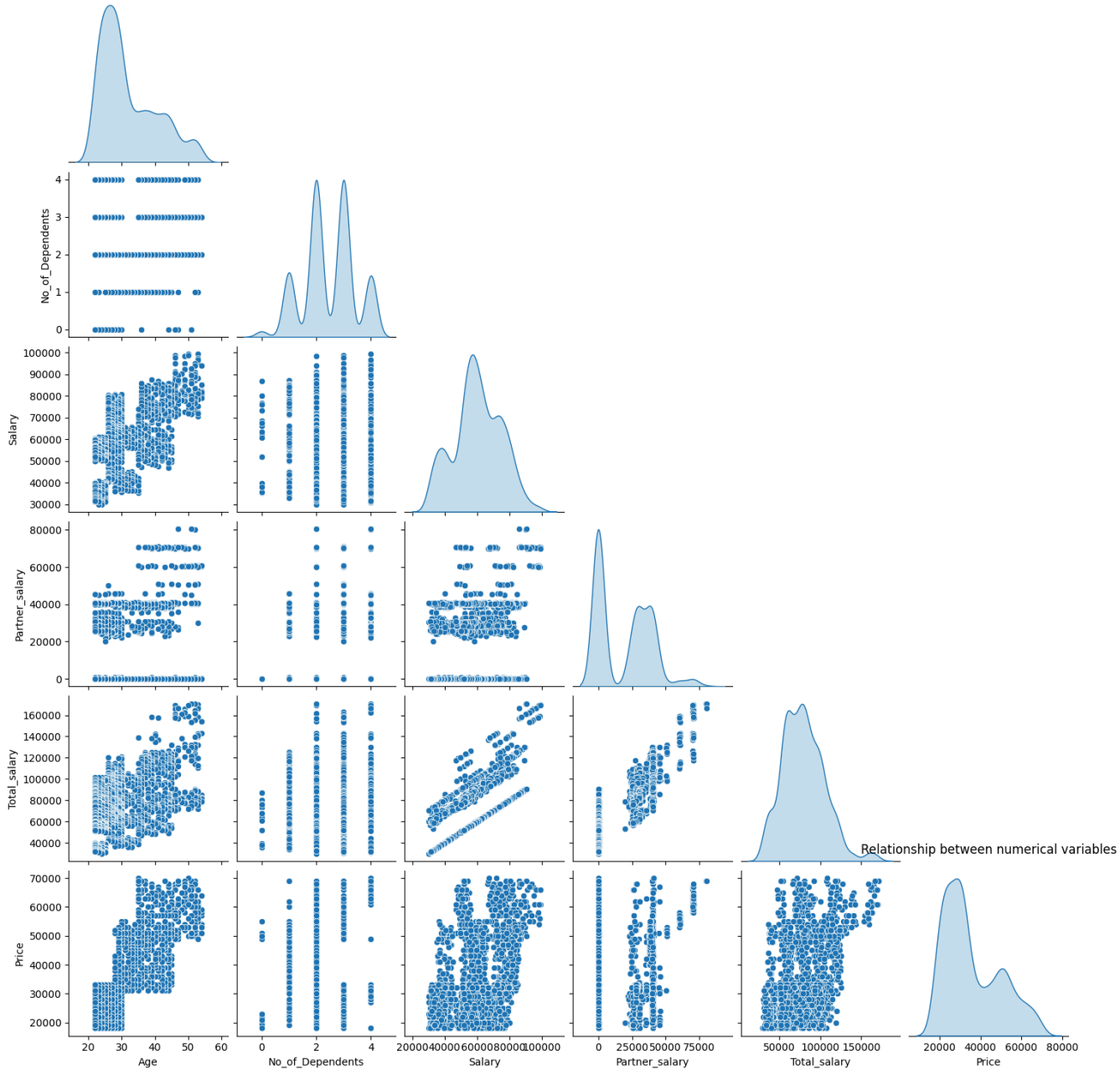




- The price distribution is right skewed meaning there are more cars with lower prices than the cars with higher prices.
- The median price of the cars is 31000.
- The price range of the car is not mentioned but from the histplot we can estimate the price to range from 25000 to 70000

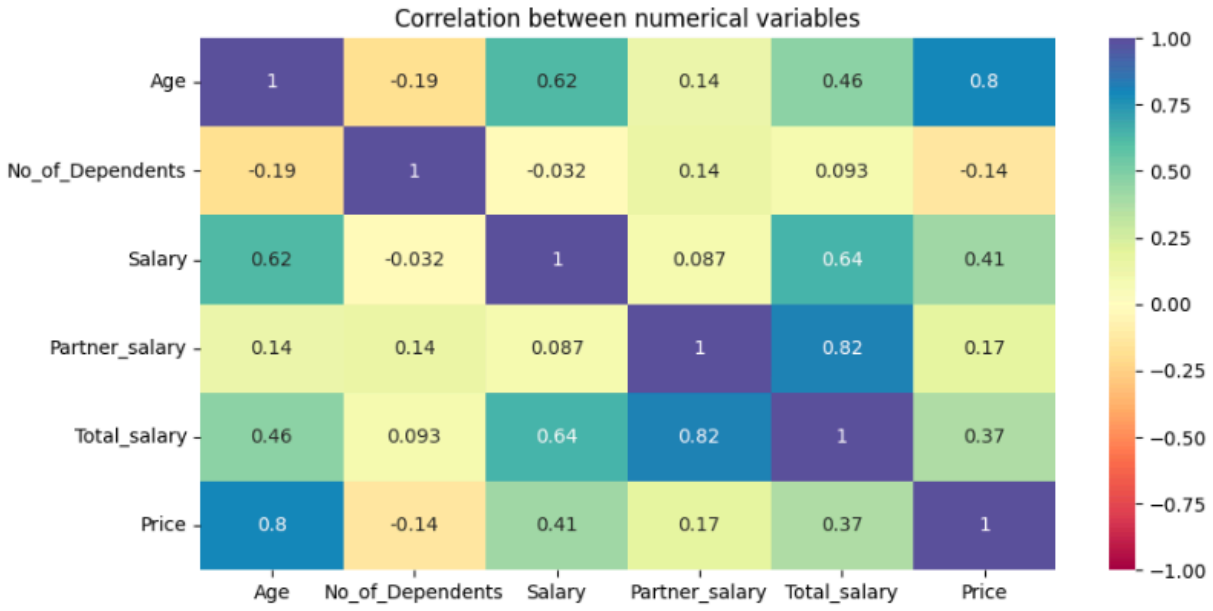
Bivariate analysis

Relationship between numerical variables in the dataset



- There is a high positive correlation between price and age.
- A linear relationship can be hardly seen among the variables.

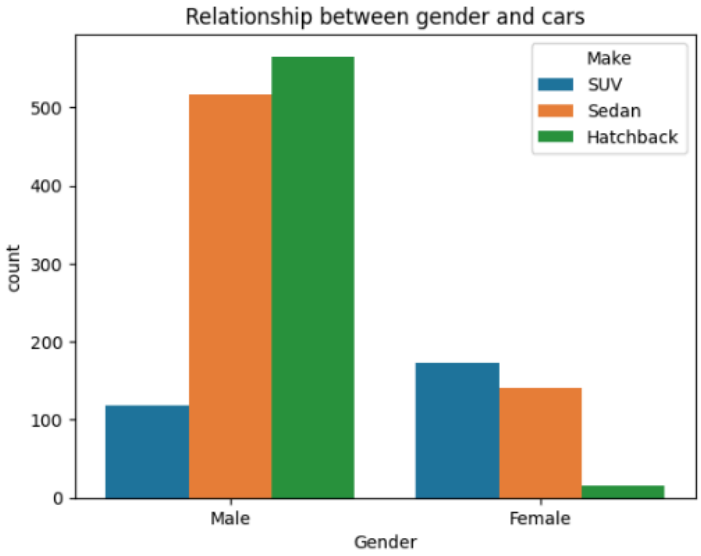
Correlation between numerical variables



- Age & Price has a strong positive correlation of 0.8 indicating a strong positive relationship suggesting older people buy cars with higher prices.
- Number of dependents & has a correlation of -0.14, indicating a weak negative relationship meaning young people tend to have fewer dependents.
- Salary & Total salary has a correlation of 0.64, indicating a moderate positive relationship meaning that when individual salary increases, the total salary also tends to increase, but not as strongly as the relationship between partner salary and total salary.
- Partner salary & Total salary has a correlation of 0.81, indicating a very strong positive relationship meaning that when partner salary increases, the total salary also tends to increase significantly.

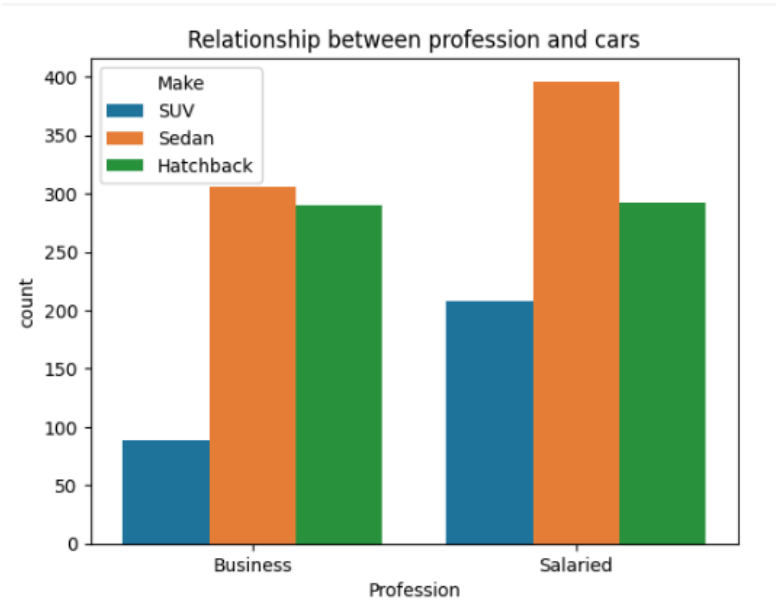
Relationship between categorical and numerical variables

Relationship between gender and cars



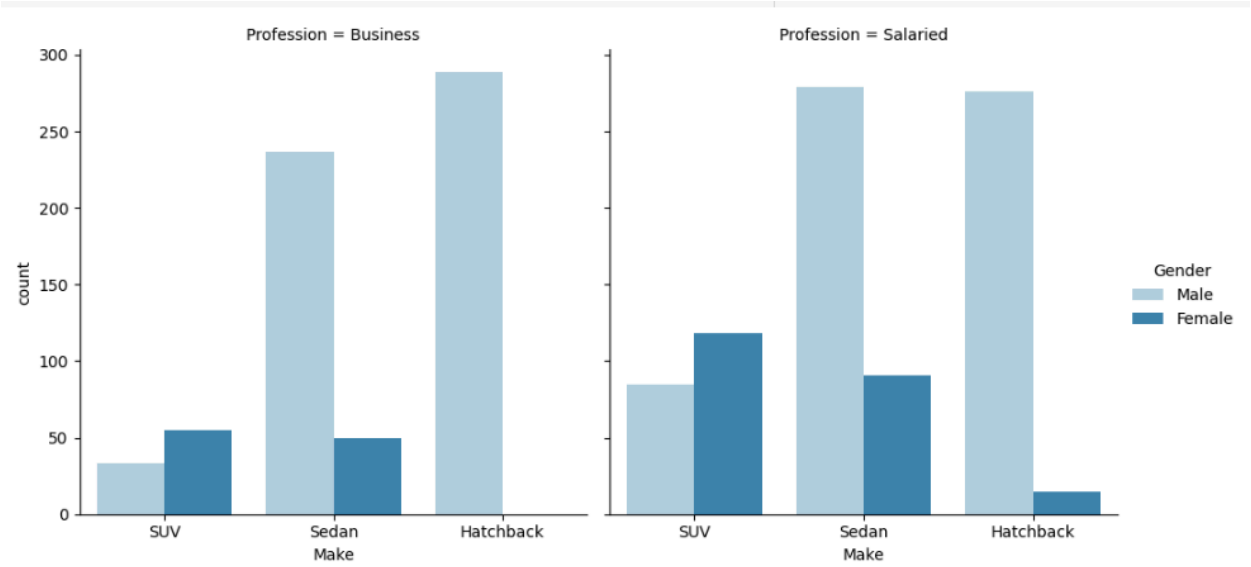
- Hatchbacks are more preferred by men followed by Sedan and SUV.
- SUVs are more preferred by women followed by Sedan and a very little prefer hatchbacks.

Relationship between profession and cars



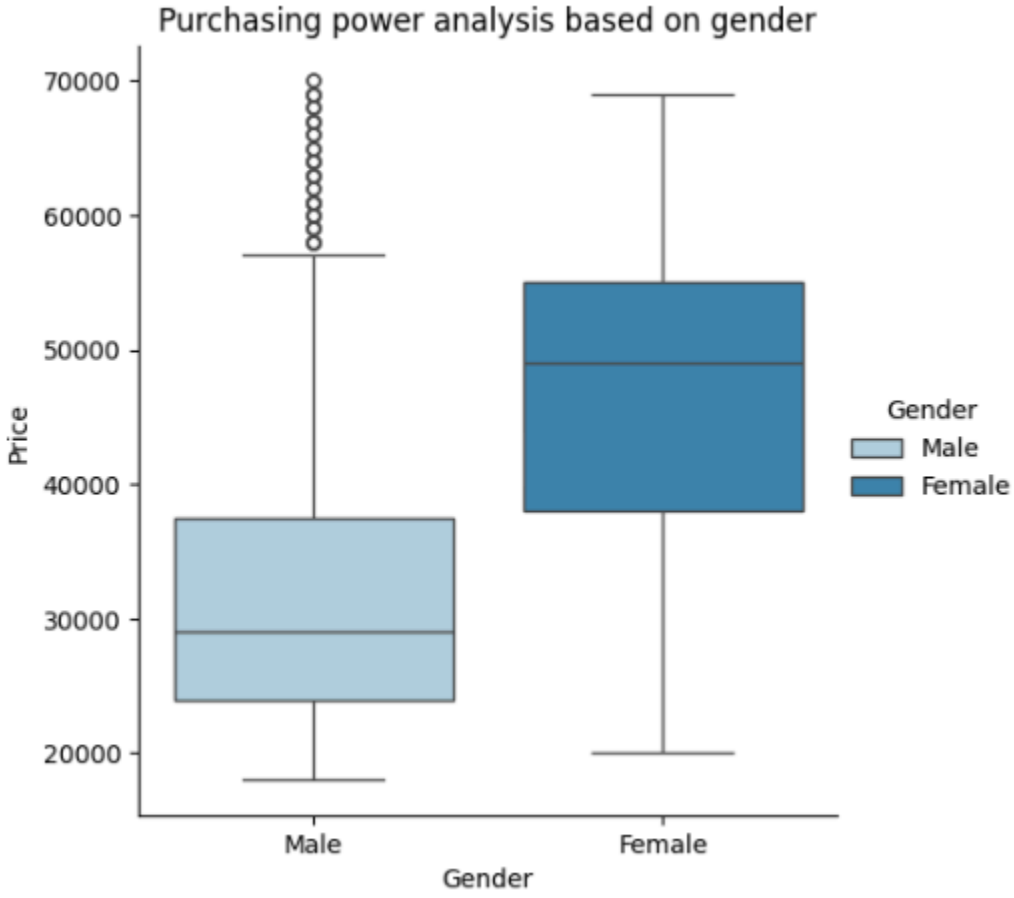
- Individuals who have their own business prefer Sedans which are followed by Hatchbacks and SUVs.
- Salaried individuals prefer Sedans followed by Hatchbacks and Suvs.

Car preference male vs female



- Men who have their own business prefer Hatchbacks the most followed by Sedans and SUVs.
- Women who have their own business prefer SUVs followed by Sedans.
- Men who are salaried prefer Hatchback and Sedan almost the same followed by SUV's.
- Women who are salaried prefer SUVs the most followed by Sedans and very little prefer hatchbacks.

Purchasing power analysis based on gender

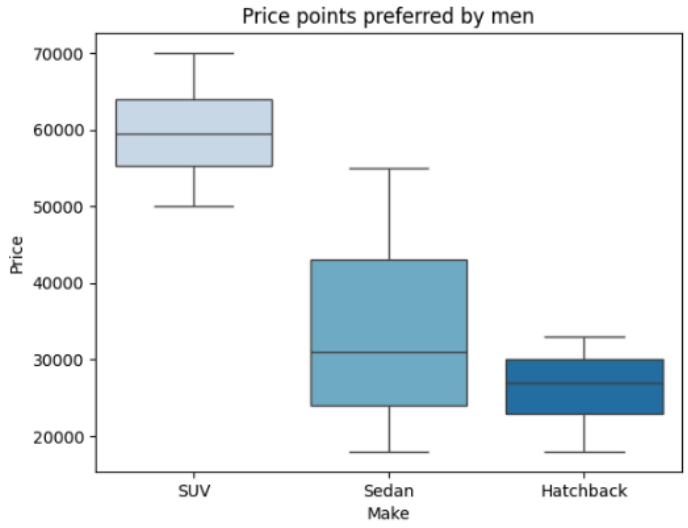


- Women tend to purchase automobiles more than men and also tend to spend more.
- 75% of the men limit their spending on automobiles to 38000 while women spending on automobiles start at 38000.
- There are outliers observed in Men.

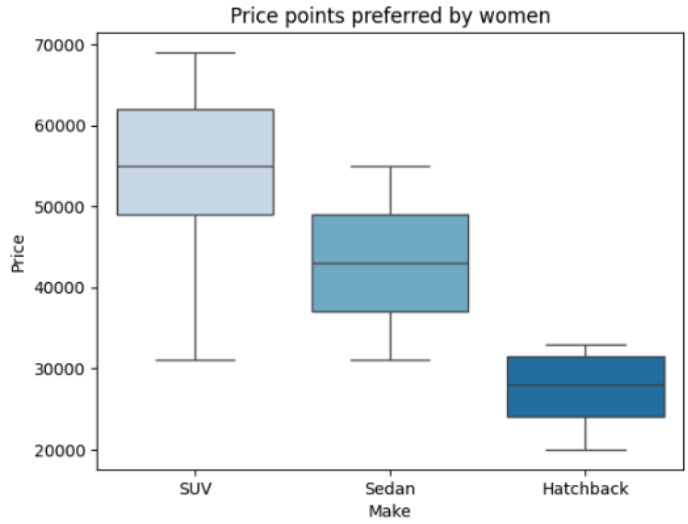
Price Points of different car makes

- SUVs cost more than Sedans and hatchbacks i.e SUV>Sedan>Hatchback

Price Points preferred by men

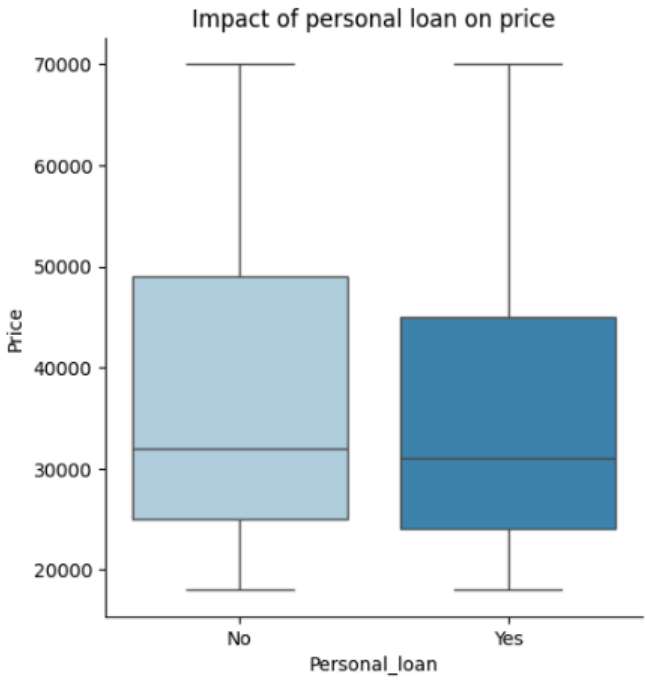


Price Points preferred by women



- Men prefer Hatchback however they also buy Sedans equally but men purchased lower segment sedans with the mean price of 32000
- Women prefer SUVs which are priced high followed by Sedans, where the mean price of SUVs are 55000 and Sedans are 43000.

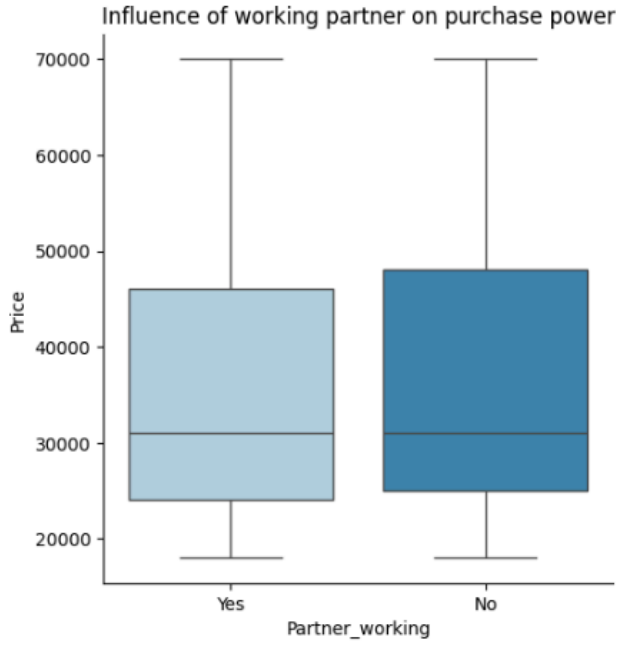
Impact of personal loan on price



Personal_loan	count	mean	std	min	25%	50%	75%	max
No	789.0	36742.712294	14534.344526	18000.0	25000.0	32000.0	49000.0	70000.0
Yes	792.0	34457.070707	12578.780338	18000.0	24000.0	31000.0	45000.0	70000.0

- The median purchase price with people taking personal loan and not taking personal loans is in between 30000 to 32000.
- There is not much difference in purchasing price whether or not a personal loan is taken or not.

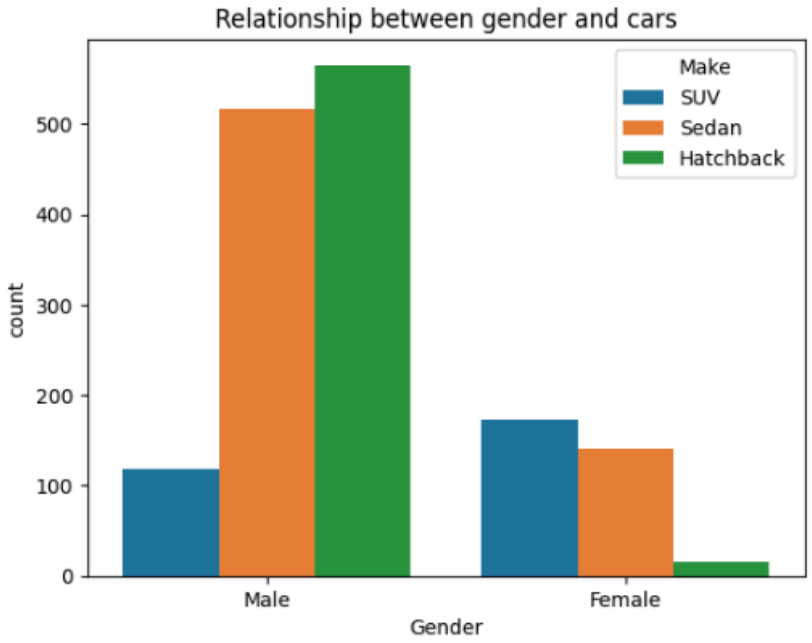
Influence of working partner on purchasing power



- No significant difference is seen on the purchasing power whether or not the partner is working or not.

Questions

Question 1: Do men tend to prefer SUVs more compared to women?

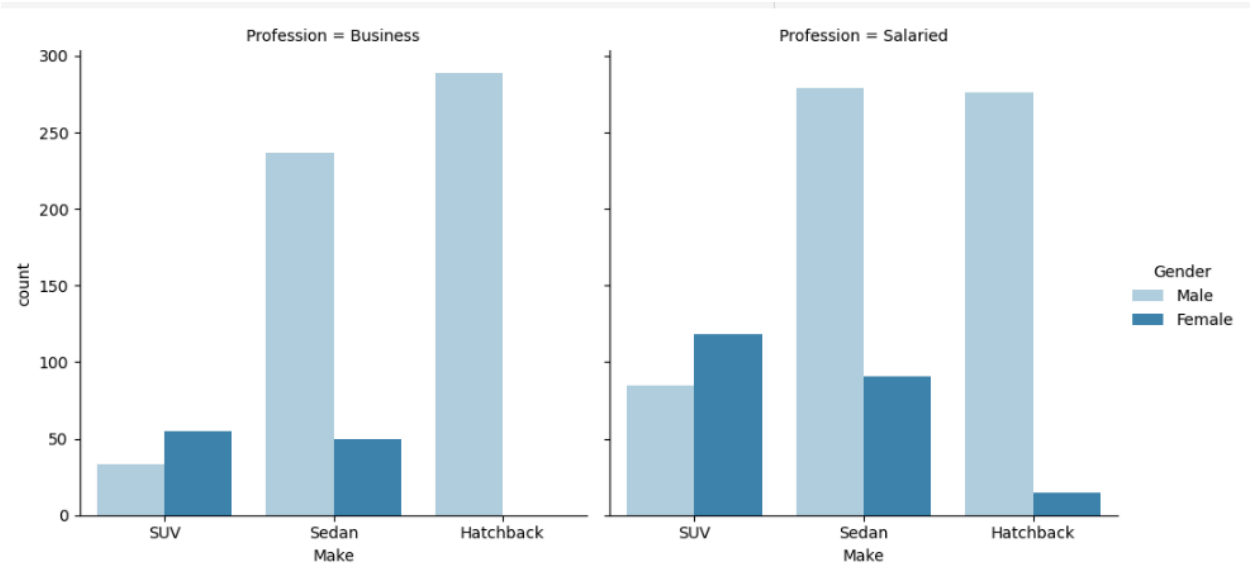


- No men tend to prefer Hatchbacks which are followed by Sedan and the least preferred are SUVs.
- SUVs are more preferred by women.

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

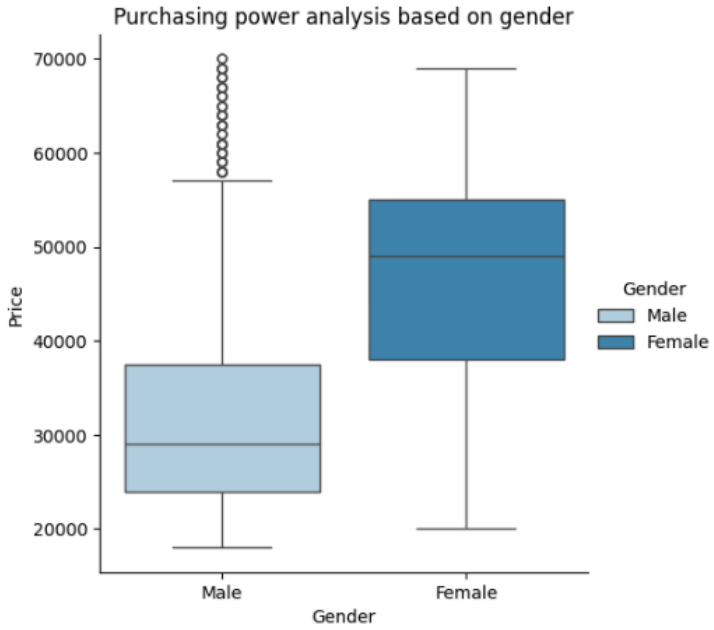
- The likelihood of a salaried person buying a Sedan is 44.19%.
  i.e [salaried_sedan/Total_salaried_people] x 100

Question 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?



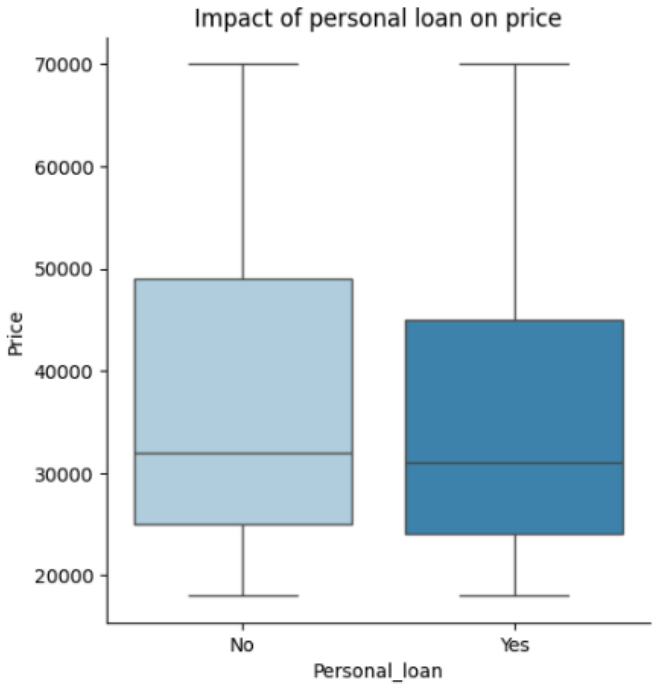
- Sheldon Cooper's claim that a salaried male is easier target for a SUV over a Sedan sale is completely wrong as the above chart clearly says that salaried men would prefer a Sedan or a Hatchback than an SUV.

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



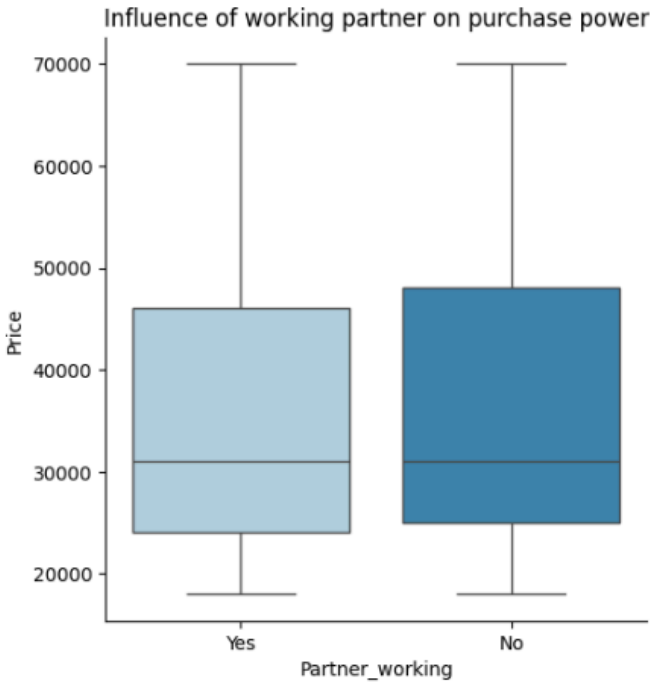
- Women tend to purchase automobiles more than men and also tend to spend more.
- 75% of the men limit their spending on automobiles to 38000 while women spending on automobiles start at 38000.

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



- People who took a personal loan spent upto $45,000 on automobile purchases.

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



- There is no significant difference in purchasing power whether or not the partner is working or not.

Actionable insights and business recommendation

Actionable insights

- Men prefer hatchbacks and sedans over SUVs.
- Women prefer SUVs over other vehicle types.
- Of the salaried population, about 44.19% have bought sedan vehicles.
- Compared to men, women typically spend more money on cars.
- Those who purchased cars with personal loans typically spent up to INR 45,000.
- The presence or absence of a working partner does not significantly alter the purchasing habits of the couple.

Business recommendation

- Women should be the primary target of SUV marketing campaigns, and advertising tactics should be modified to suit their tastes.
- Target Sedans' marketing campaigns at salaried people, emphasizing amenities or perks that appeal to them.
- To increase SUV sales, think about focusing on salaried men with incentives or campaigns designed especially for them.
- To capitalize on women's desire to spend, businesses can provide a wide variety of car options, such as SUVs and upper category sedans.
- Provide loan or financing options that are suited to consumers' spending patterns, sometimes with discounts or incentives for loan applicants.
- Maintain your marketing of more expensive cars to both client groupings, but think about modifying your approach depending on additional variables like gender or occupation.