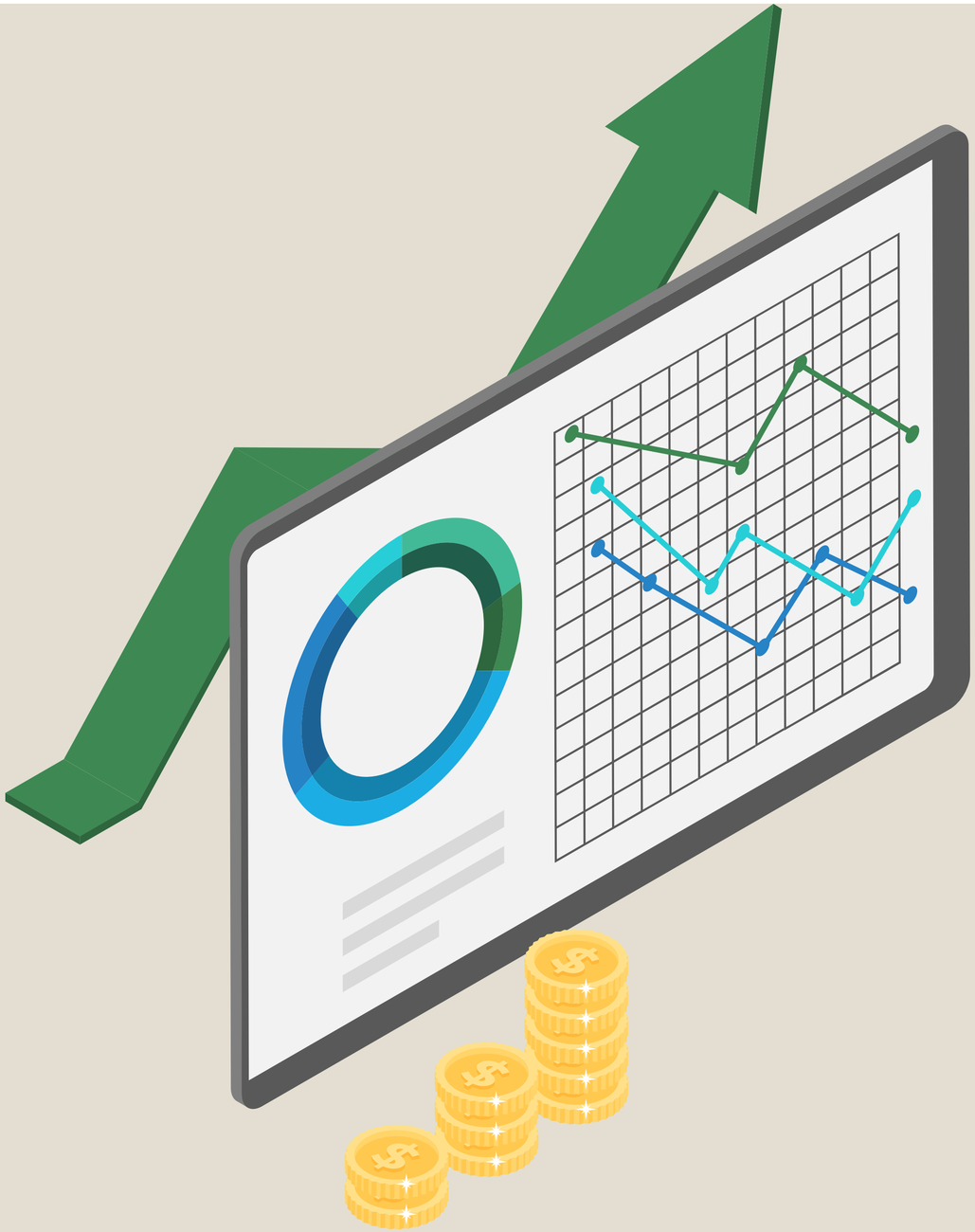
Data Analysis

Austo Motor Company

Amol Phadke

PGPDSBA.O.JULY23.A



The Problem Statement:

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.

The Objectives:

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 and features information

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

Problem 1: Data Overview

1. Data is given in the CSV format as austo_automobile.csv
2. Given data was 1581 sample size and 14 features i.e. 1581x14
3. The Data types of the features are : 1 float, 5 integers and 8 Objects as follows:

4. Description About the features:
    1. Age: Min Age is 22 years and Maximum Age is 54 Years, Median is 29 Years
    2. No of Dependents: Maximum Dependents are 4, Median is 2 , we have data 0 dependents as well
    3. Salary: Min is 30000, Maximum is 99300 and Median is : 30000, 60392
    4. Partner Salary: Min is 0, Max is 80500 and Median is 25600, Mean is 20225
    5. Total Salary : Min 30000, Max is 171000, Median is 78000, Mean is 79625
    6. Price : Min is 18000, Maximum is 70000 and Median is 31000, Mean is 35597
    7. In Gender Data there are couple of Female are mis typed which needs correction
    8. Missing Values and Null values:
        1. Partner Salary and Gender have null values i.e. 106 null values of Partner Salary and 53 for Gender
        2. There are no duplicate rows of data

```
#   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
```

- The data has 1581 instances with 14 attributes. 5 integer type, 1 float type and 8 object type(Strings in the column)

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

Problem 1: Data Overview…. contd

Bad Data Treatment and Anomalies Corrections:

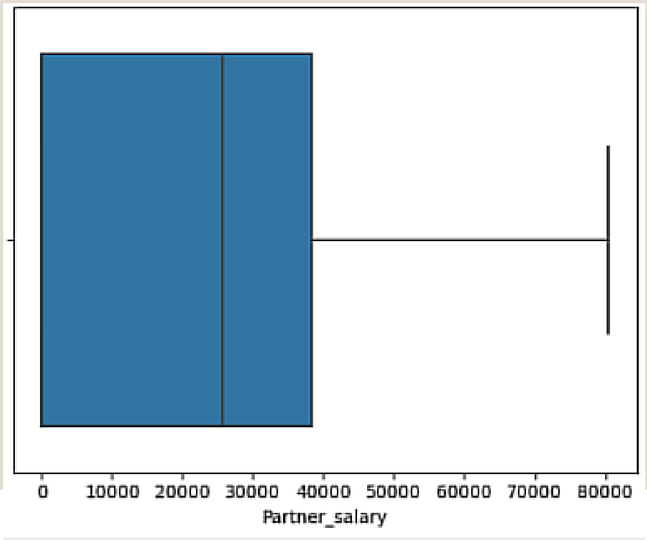
- In Gender we have 53 na values and it needs treatment
  - As we have majority of the data was Male and there are 53 missing values hence I replaced the na with Male to make the Gender data as follows:
  - Also couple Female data has typo error and it needed replacement with Female

```
Male      1252
Female     327
Femal        1
Femle        1
Name: Gender, dtype: int64
```

  - After Treatment Gender Data looks as follows:

```
Male      1252
Female     329
Name: Gender, dtype: int64
```

Now Treated the Partner Salary which has na values
- As we see in the box plot there are no outliers
- Hence its better to replacement the na values with it’s mean



Problem 1: Data Overview…. contd

After Doing the missing data treatment and Anomalies adjustment our data looks as follows:

	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	888	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	13833.636545	18000.0	25000.0	31000.0	47000.0	70000.0
Make	1581	3	Sedan	702	NaN	NaN	NaN	NaN	NaN	NaN	NaN

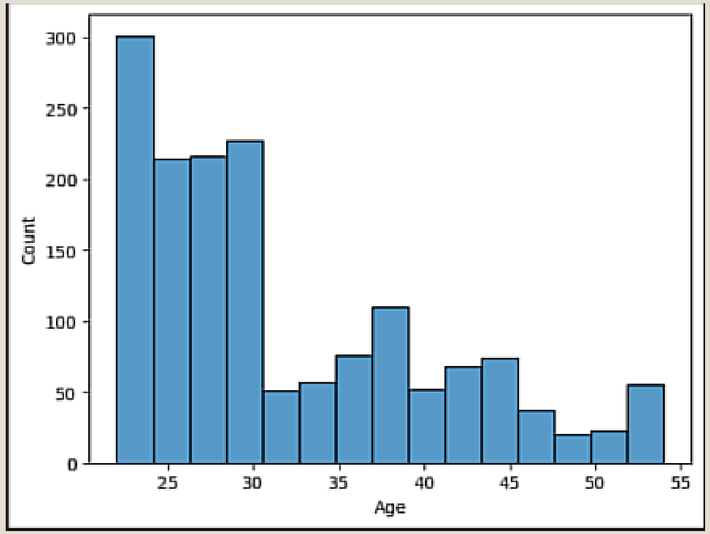
```
Age                0
Gender             0
Profession         0
Marital_status     0
Education          0
No_of_Dependents   0
Personal_loan      0
House_loan         0
Partner_working    0
Salary             0
Partner_salary     0
Total_salary       0
Price              0
Make               0
dtype: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

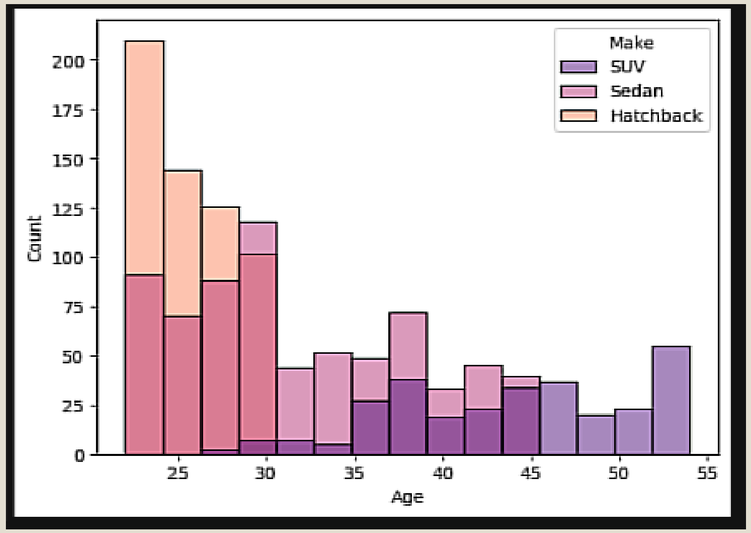
Observations and Insights:

- We have seen the data is 1581 rows and 14 features. Data types of these are mentioned in earlier slide
- In the we have also seen Gender and Partner Salary had missing data
- I have replaced 53 rows of missing data in Gender feature by Males as Male population was significantly higher than Females
- As far as Partner Salaries are concerned it doesn’t have any outlier hence it was safe to replace them with mean of Partner Salary which was done
- In Gender we had 2 spelling mistakes for Female which was corrected by replacing them with Females

UNIVARIANT ANALYSIS

Age of Customers

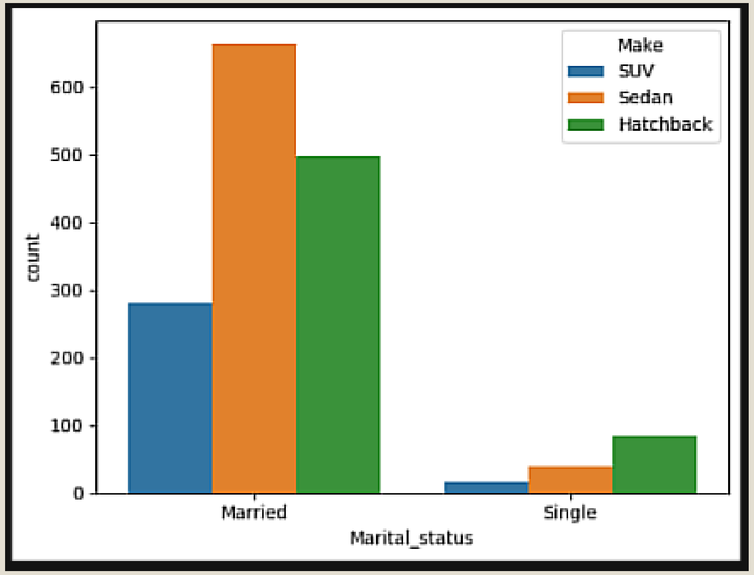


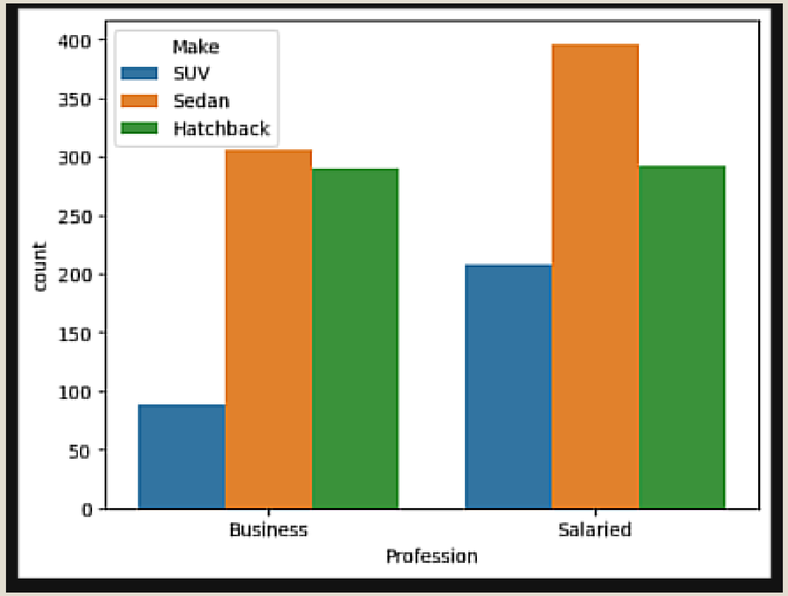


Observations and Insights:

- We see highest Population of sample is less than 30 years of age
- We see here SUV’s are preferred by people with higher ages
- Most of the young people are inclined towards Sedan or Hatchback cars

Based on Marital Status and Profession





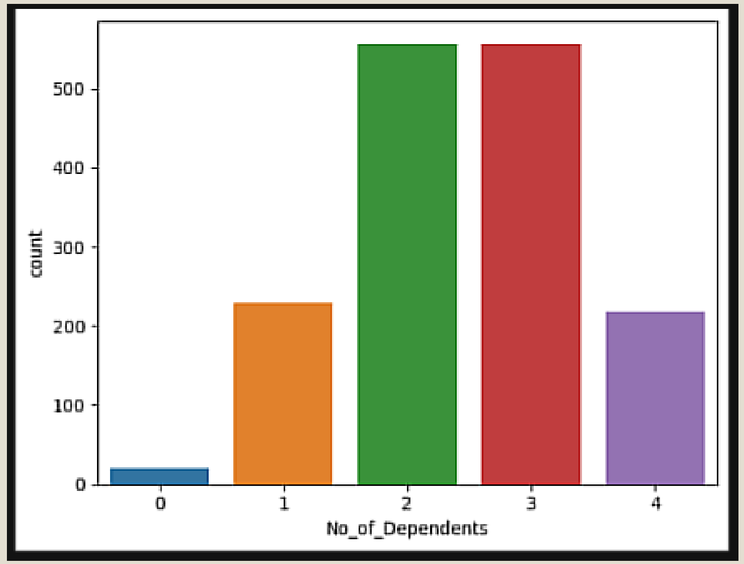
Observations and Insights:

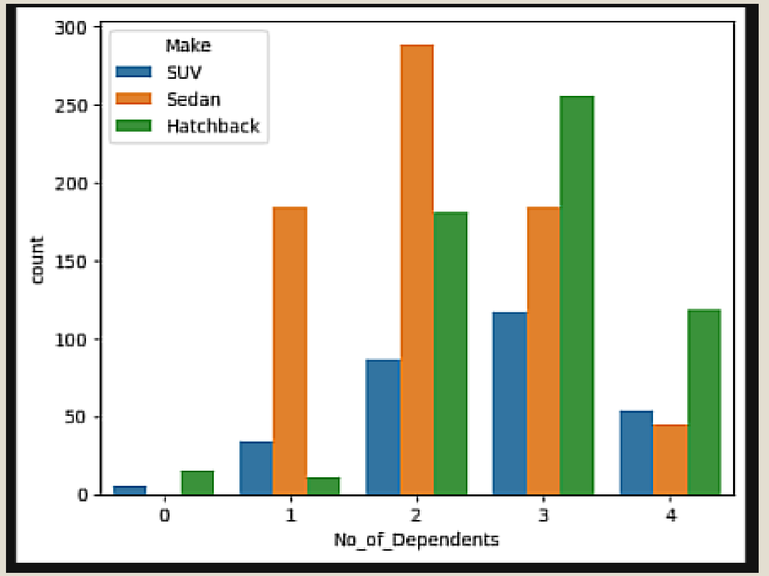
- SUV is least preferred car for both Married as well as Single People
- Married people prefers Sedan where Hatchback is choice for many Singles
- Majority of people who are buying cars are Married
- 91% people who are buying cars are married and only 9% of the singles have bought cars

Observations and Insights:

- In the data provided we see 57% of the people are salaried and 43 are business professionals who are buying cars
- Sedan is the most preferred car for both types of professionals
- Hatchback is next preferred car
- SUV is least preferred vehicle however salaried people are buying more SUV’s compare of Business professionals

Based on number of Dependents

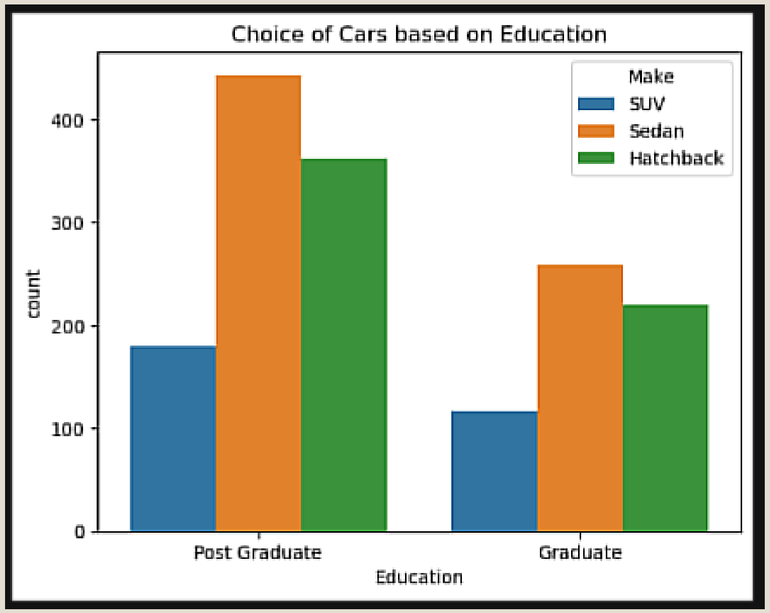




Observations and Insights:

- Majority of the population has 2 or 3 dependents
- Less sample without any dependents
- Irrespective of no of dependents Sedan is the preferred choice by 1 and 2 dependents
- Interesting things with 3 and 4 dependents we see Hatchback is the choice of cars
- SUV is the again least preferred by people with 2, 3 and 4 dependents
- Small family with 1 dependent buy mostly Sedan cars

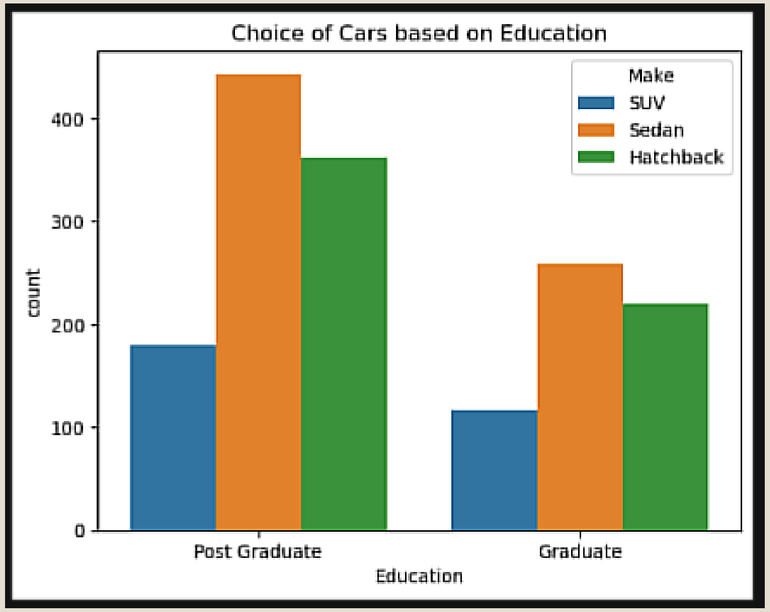
Based on Graduation



Observations and Insights:

- Sedan is the most preferred choice by both Graduates as well as Postgraduate people
- Hatchback is the next choice of cars
- SUV is preferred by Postgraduate people more than Graduates

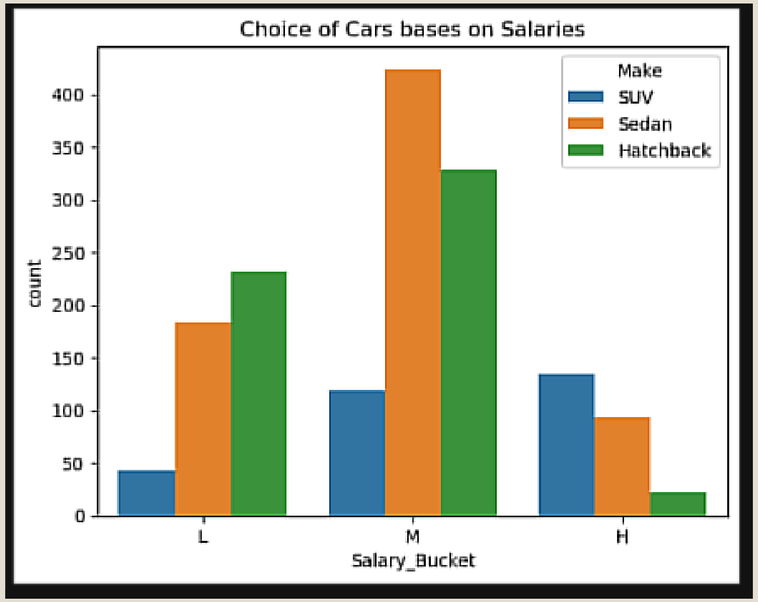
Based on Graduation

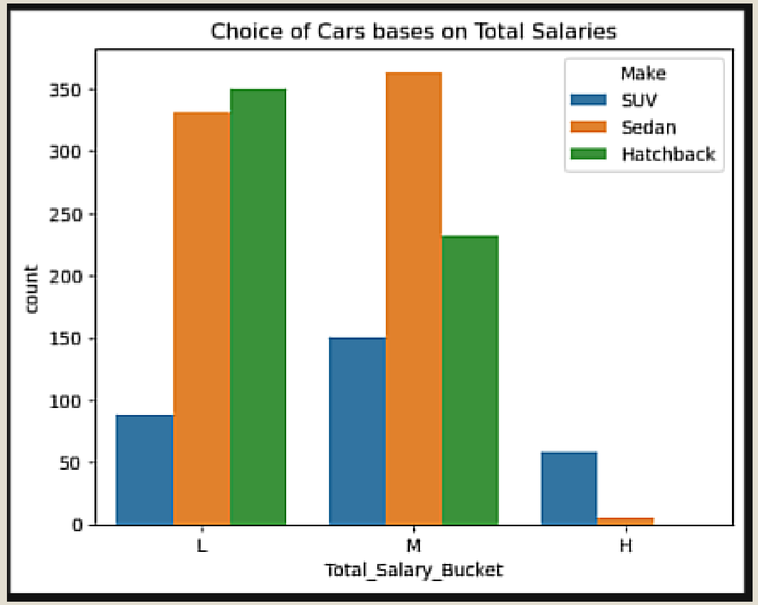


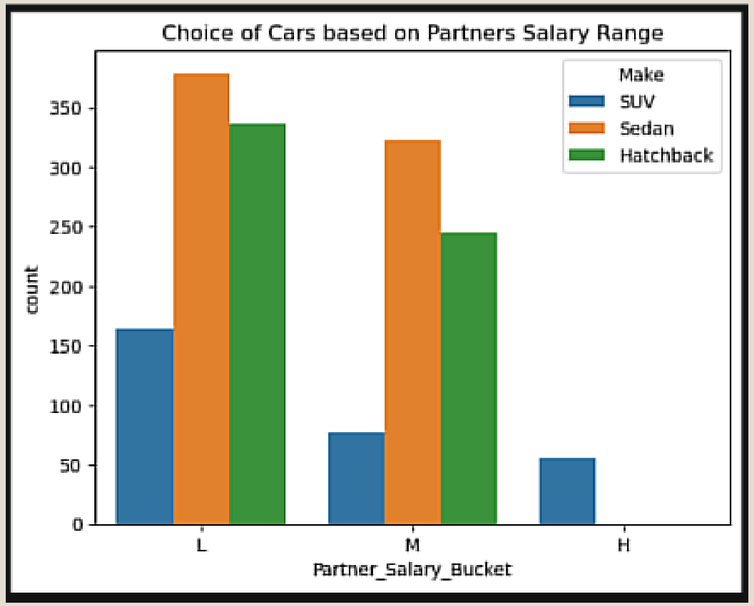
Observations and Insights:

- Sedan is the most preferred choice by both Graduates as well as Postgraduate people
- Hatchback is the next choice of cars
- SUV is preferred by Postgraduate people more than Graduates
- 62% of car buyers are Postgraduate and 38% are Graduates

Based on Salaries



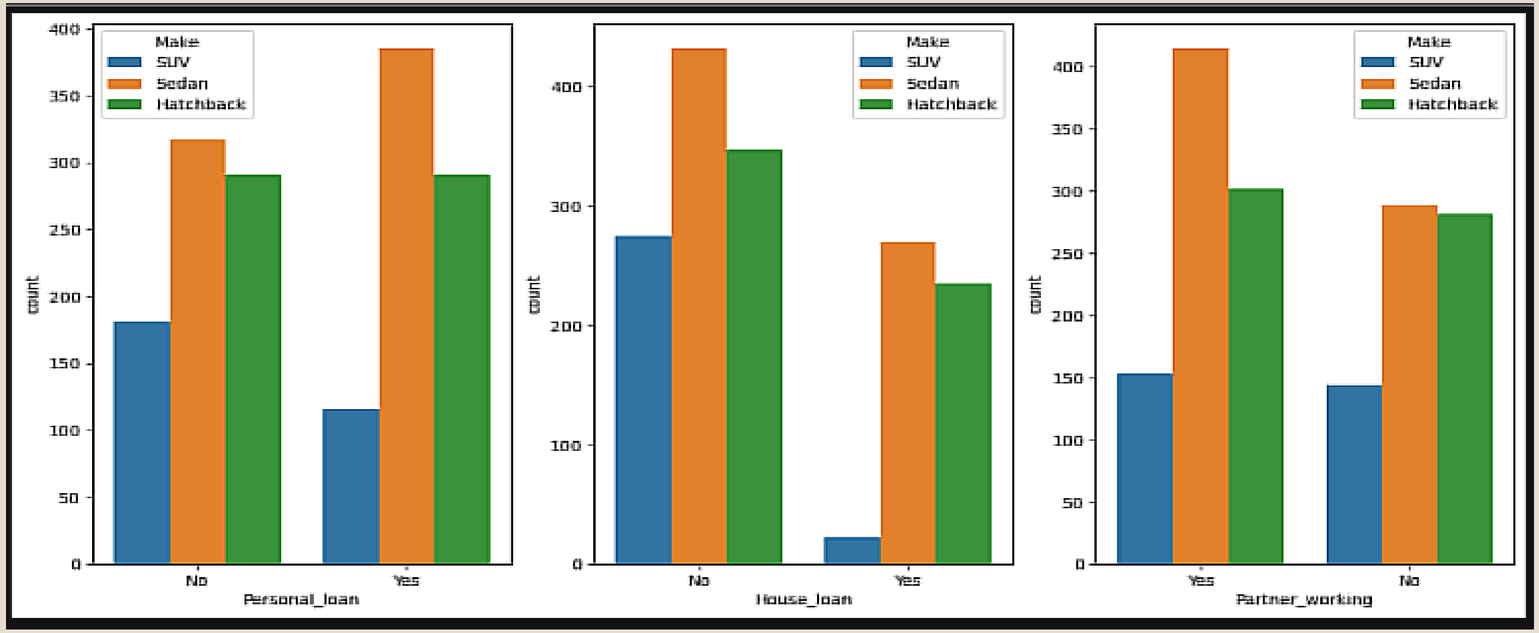




Observations and Insights:

- People with lower salaries Prefer Hatchback cars
- People with High salaries Prefer SUV
- People with Medium salary range prefer Sedan
- High Total Salary people are buying SUV’s
- Medium Total Salary People are buying Sedans
- People with Low Total Salaries are buying Hatchback and then followed by Sedan. Few are buying SUV’s too
- People whose partners are working and are earning high salaries buys SUV’s
- Rest of the others are buying Sedan more than Hatchback cars

Based on Loans and Working Partners

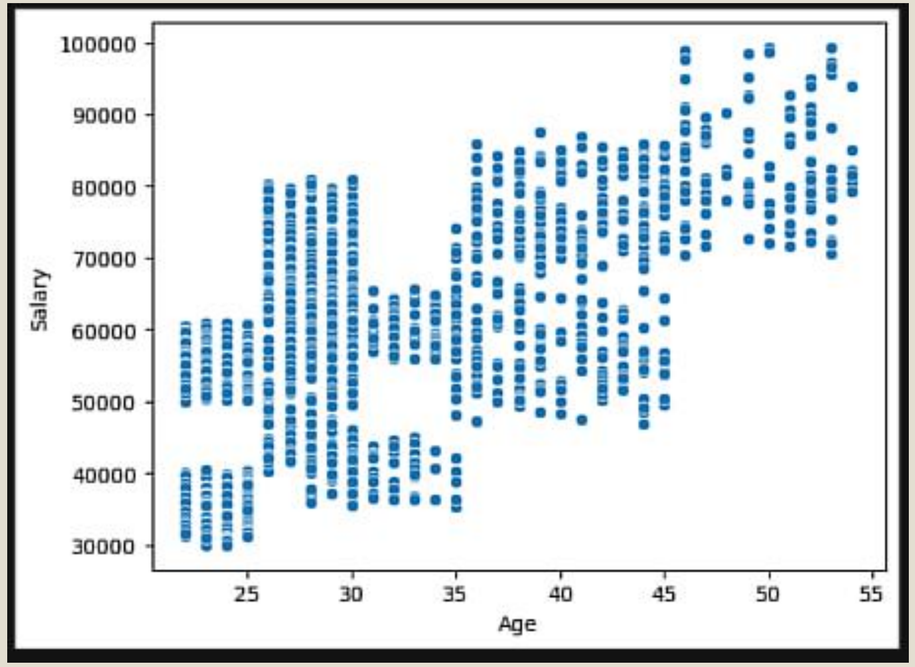


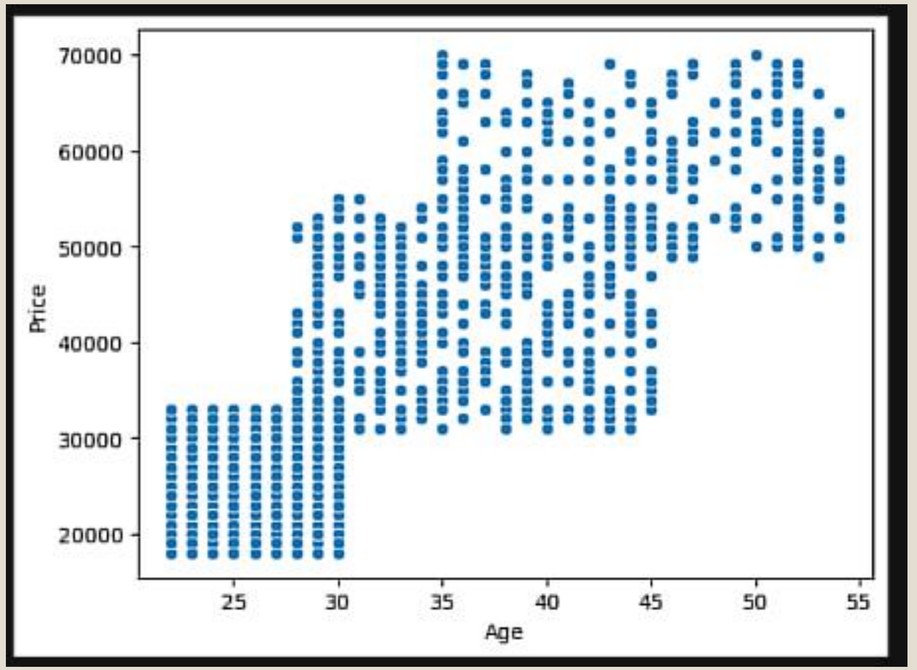
Observations and Insights:

- People with Personal loan seems to have more Sedan cars then no Personal Loan People. Are they taking PL for buying cars?
- People who don't have home loans have buys more cars than people who have home loans.
- People whose partners are working seem to slightly more Sedan cars than whose partners don't work

BIVARIANT ANALYSIS

Analysis between Age and Salaries/Price



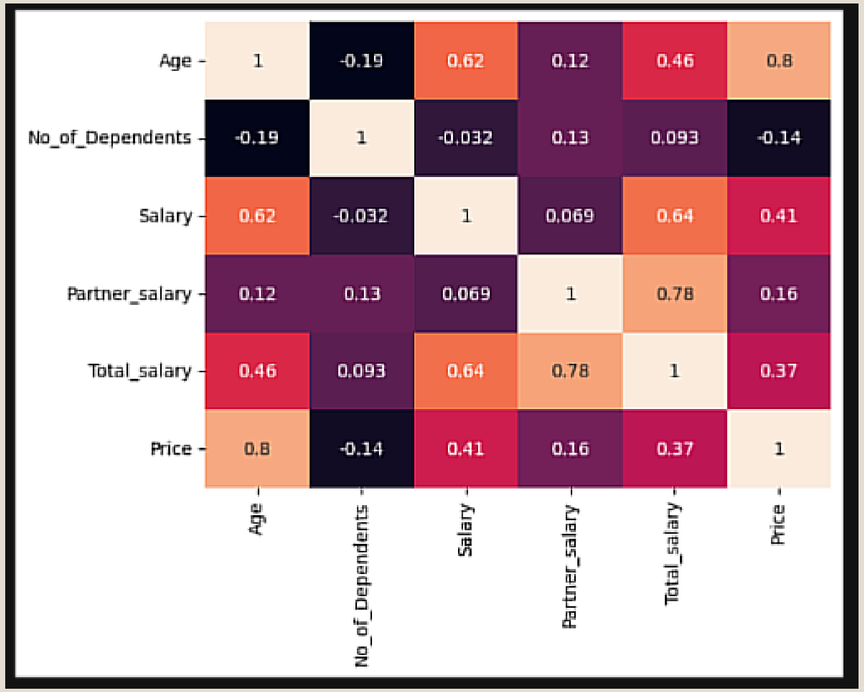


Observations and Insights:

- As we Age and Salaries have positive correlation
- This also proves our earlier hypothesis that more aged people have SUV's and they earn more than Young People
- Young People don't prefer SUVs as their salaries are relatively lower
- Age and Price also has positive correlation i.e. People with Higher Age spends on higher priced cars

- This hypothesis can be made i.e. Older People earns more and buys more SUV's as they are expensive.

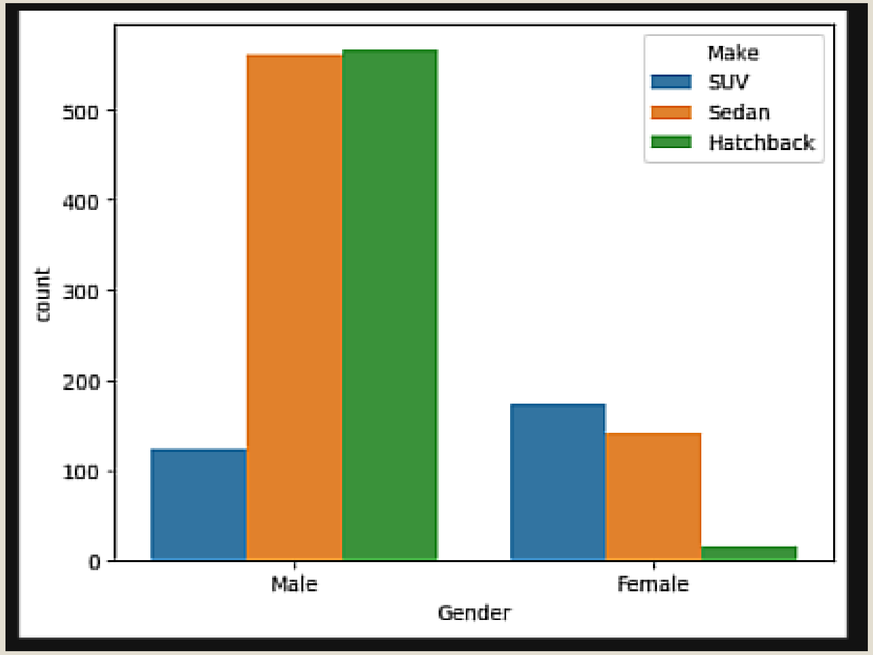
Explore the correlation between all numerical variables



Observations and Insights:

- Age and Salary has positive correlation
- Age and Price also has positive correlation
- Partner Salary and Price have a positive correlation but a very small one

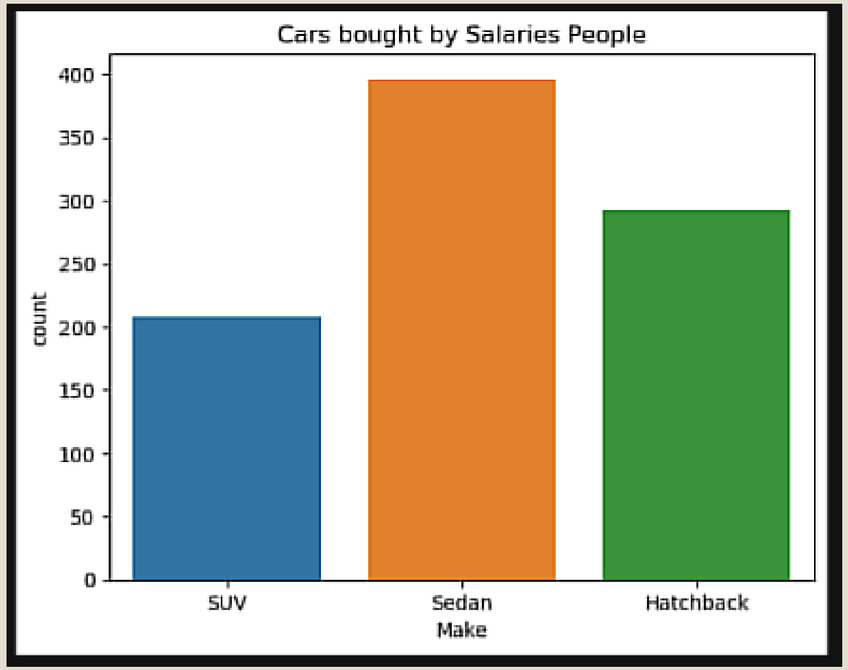
Question: Do men tend to prefer SUVs more compared to women?



Observations and Insights:

- Answer to this key question is NO
- Though we have few women's in the data but they are buying more SUV’s than Men

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

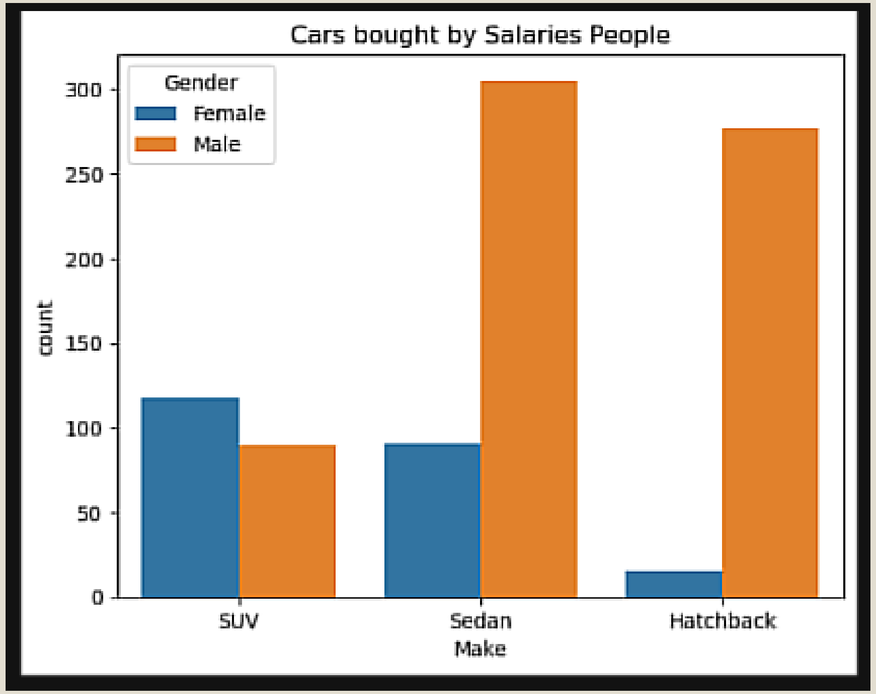



```
Sedan          396
Hatchback      292
SUV            208
Name: Make, dtype: int64
```

Observations and Insights:

- It is very likely that Salaried people buys Sedan
- As shows in the graph and data we see 44% of salaried people buys Sedan, 33% people buy Hatchback and only 22% salaried people buys SUV

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?

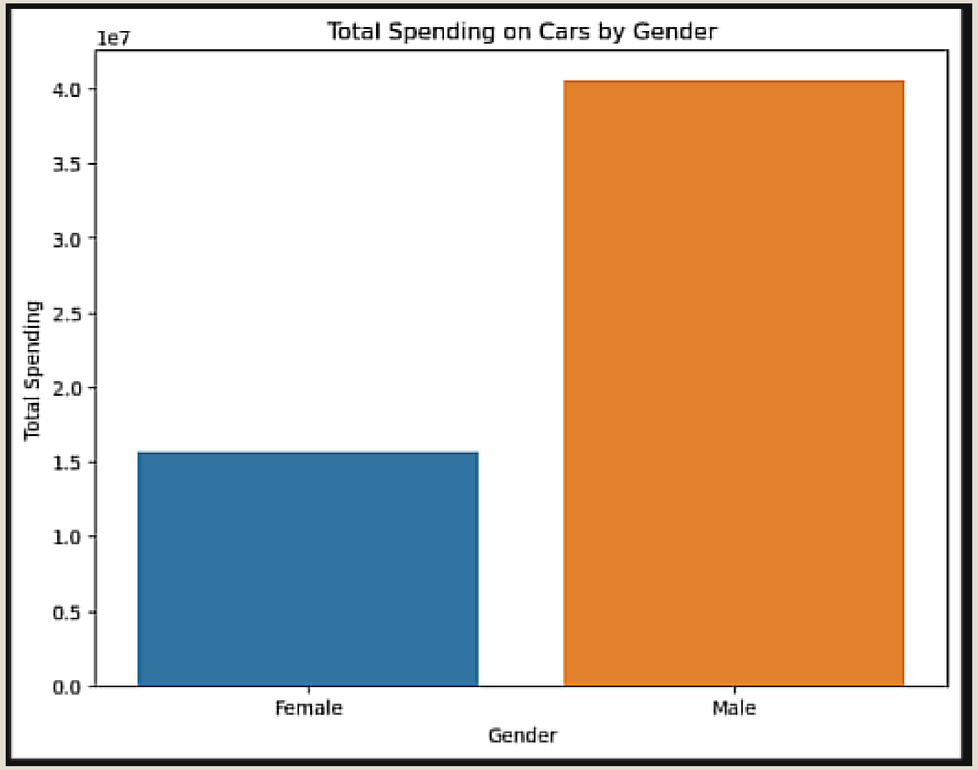


Observations and Insights:

- It’s not true that Salaried Male are easier Targets for SUV sale over Sedan
- If you see here Salaried Male are buying More Sedan and then followed by Hatchback
- In fact in SUV sale salaried females are buying more SUV’s
- As far as Salaried Males are concerned they don’t buy too many SUV’s

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

```
Gender
Male      40585000
Female    15695000
Name: Price, dtype: int64
```

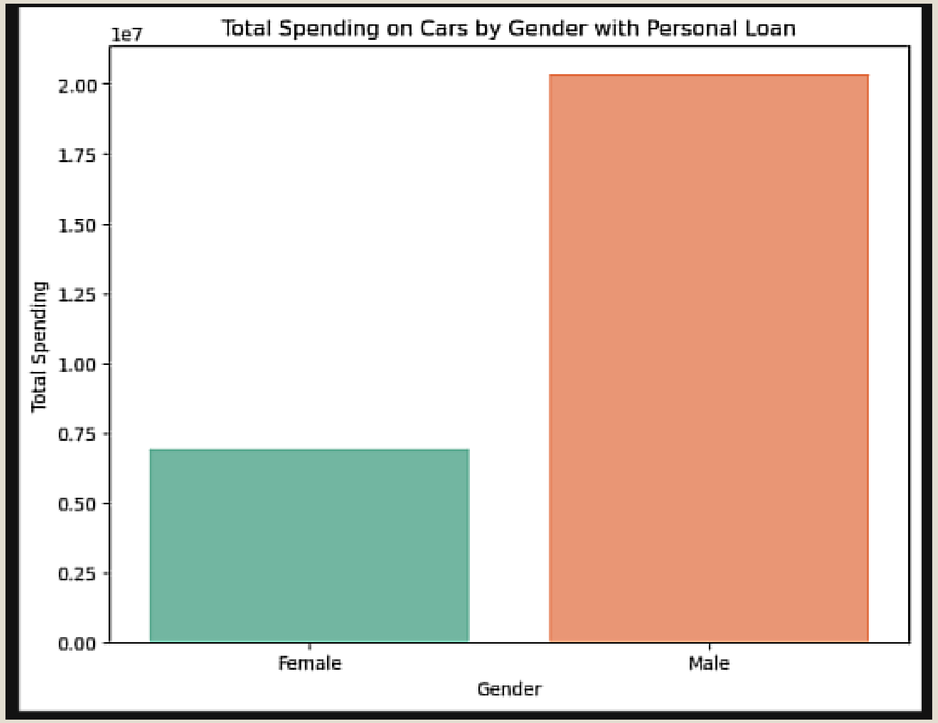


Observations and Insights:

- ~75% of the amount spent on purchasing automobiles are Men
- 25% of the amount spending on purchasing automobiles are Women

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

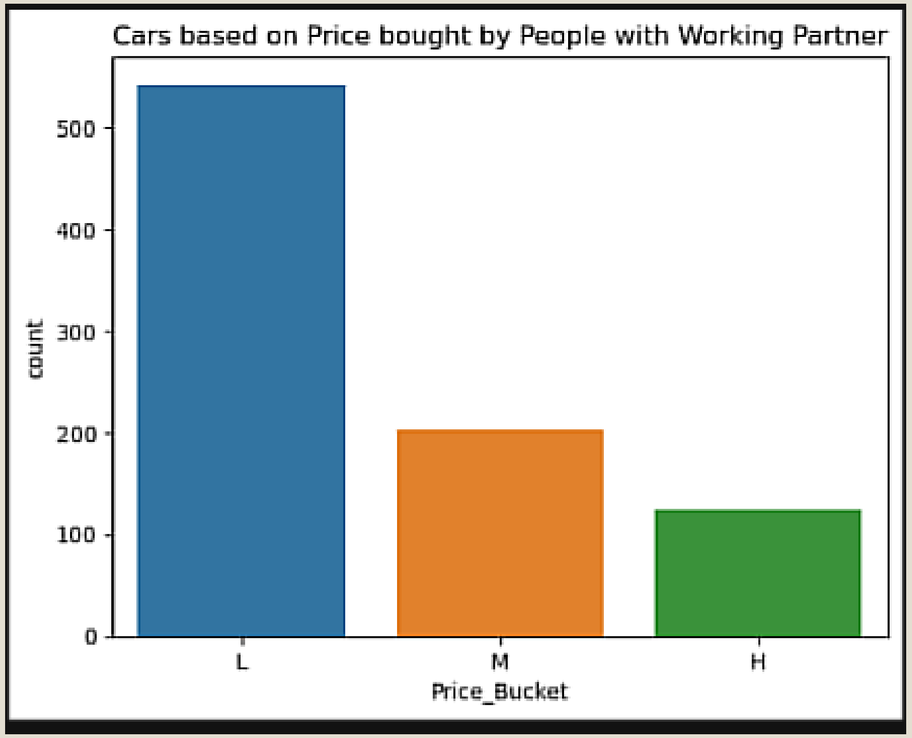
```
Gender
Male      20357000
Female     6933000
Name: Price, dtype: int64
```



Observations and Insights:

- ~72% of the amount spent on purchasing automobiles are Men
- 28% of the amount spending on purchasing automobiles are Women

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



Observations and Insights:

- Even people with working Partner seem to buying cars with low price
- 62% people are buying low priced cars, 23% buying Medium size cars and 14% people are buying High priced cars

Actionable Insights & Recommendations

Business Recommendations:

- Most of the people are buying the Sedan cars
  - Austo Motor Company should invest more in Sedan cars and launch new brands in the market
  - Company should also focus on launching new Sedan cars with Low and Competitive prices
  - Company should focus on increase Women’s customers as well
  - Low priced cars are preferred by Many and some innovations should come up in this segment with low priced cars

- Next Preferred vehicles are Hatchback
  - Company should again focus on increasing Market share on Hatchback cars by offering more options
  - Company here can look to attract more women buyers which don’t seem to many vehicles
  - Low priced cars are preferred by Many and some innovations should come up in this segment with low priced cars

- SUV’s are not bought by many
  - Austo should focus on existing SUV vehicles and can look to offer some discounts on them
  - Company shouldn’t introduce new models in the same segments instead they should focus on existing SUV’s