Problem Statement

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 is 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.

A. **Below are the dataset and their data types which are important for a data base administrator**

Number of Rows: 1581 entries, 0 to 1580

Data columns (total 14 columns): Number of Variables

```
 #   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 dataset has 1581 rows and 14 columns. There are 8 object data types, 5 integer data types, and 1 float data type in the dataset.

Data types of above variables

Categorical Variables:

Binary:	Multilevel:	Continuous Variable:	Discrete Variable:
Gender	Make	Price	Age
Marital_ status	Education	Salary	No _of _Dependents
Personal_ loan	Profession	Partner_ salary	
House_ loan		Total_ salary	
Partner_ working			

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.**

- **Yes, there are discrepancies in the Partner_ salary and Gender data, there are null data available in both of these attributes and values need to imputed for these null values.**

```
<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
```

- **There are missing values in 2 columns of the data. There are 53 missing values in gender column and 106 missing values in partner salary column.**

```
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
```

- **We have treated these missing values for gender and partner salary column after understanding the distributions of features in the data, the relationships that exist in the data. This will help us impute these values more effectively.**

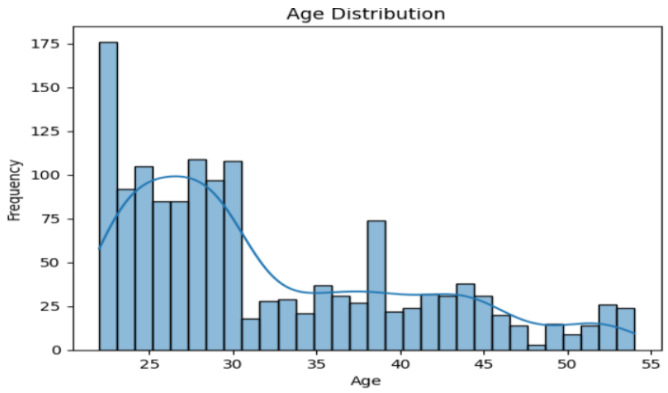
```
Age                    0
Gender                 0
Profession             0
Marital_status         0
Education              0
No_of_Dependents       0
Personal_loan          0
House_loan             0
Partner_working        0
Salary                 0
Partner_salary         0
Total_salary         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

C. Univariate Analysis

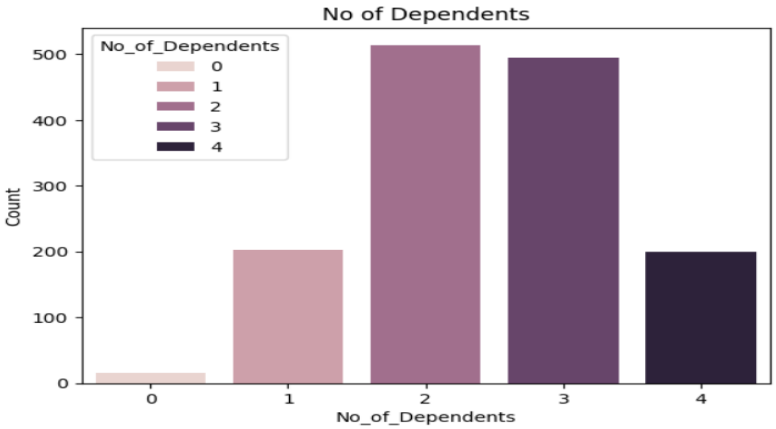
To explore all the variables (categorical and numerical) of the data separately by using appropriate visualizations and draw insights that can be utilized by the business.

- **70% of the cars are brought by the people below 40 Years age.**

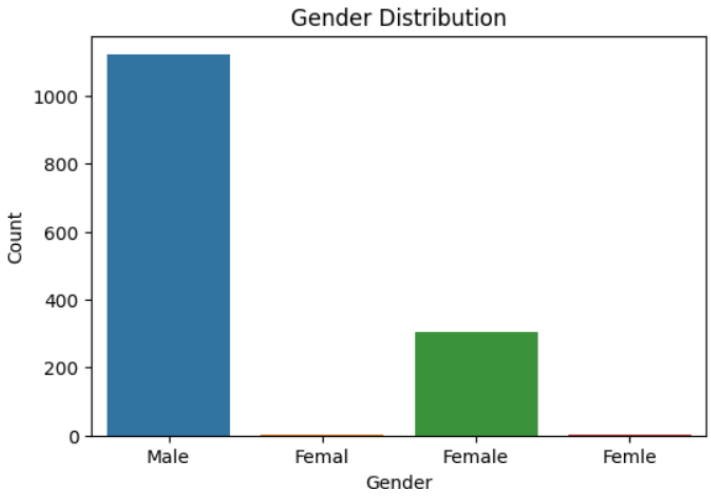
	count	mean	std	min	25%	50%	75%	max
Age	1425.0	32.037895	8.475094	22.0	25.0	29.0	38.0	54.0
No_of_Dependents	1425.0	2.462456	0.936255	0.0	2.0	2.0	3.0	4.0
Salary	1425.0	60629.052632	14768.045456	30000.0	52000.0	59600.0	72400.0	99300.0
Partner_salary	1425.0	20522.877193	19656.146189	0.0	0.0	25800.0	38400.0	80500.0
Total_salary	1425.0	81151.929825	25401.025387	30600.0	62700.0	79500.0	97200.0	171000.0
Price	1425.0	36024.561404	13492.927307	18000.0	25000.0	32000.0	47000.0	70000.0



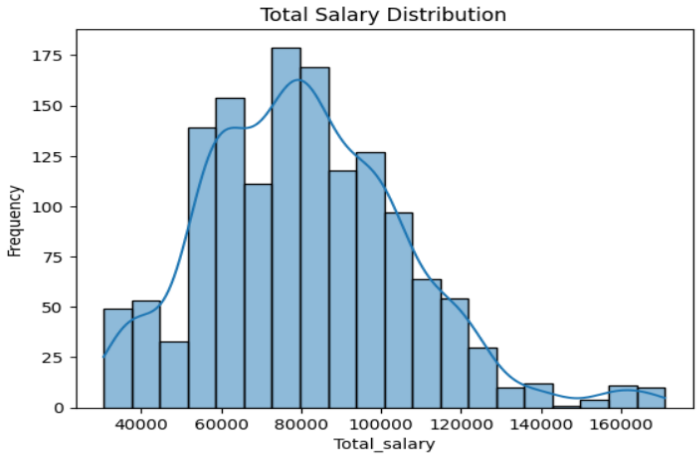
- **Maximum cars are brought by people where No_of_Dependents are 2 & 3.**



- **Maximum cars are brought by the people with male gender.**

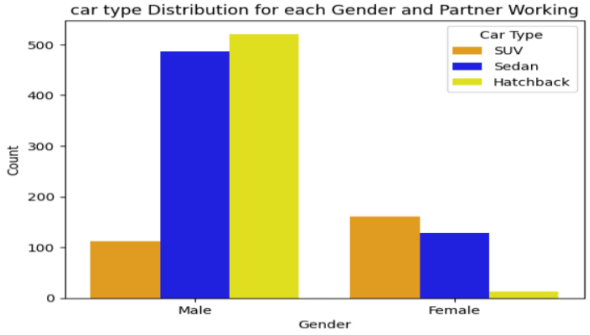


- **Total_salary is double peaked and lightly rightly skewed.**

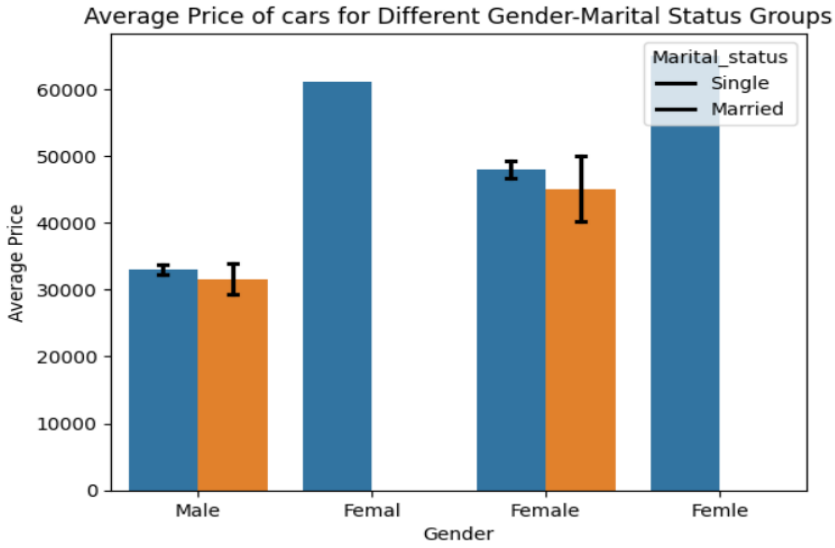


D.Understanding the relationships among the variablesin the dataset is crucial for every analytical project.Perform analysis on the data fields to gain deeper insights.Comment on your understanding of the data.

- **Sedan cars are purchased more when there are working partners.**

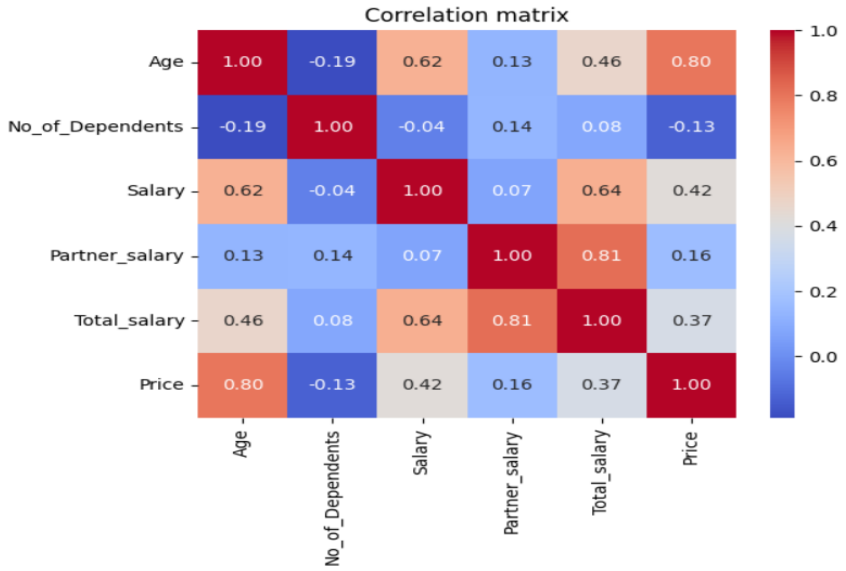


- **Married people purchased more high priced cars especially sedan**

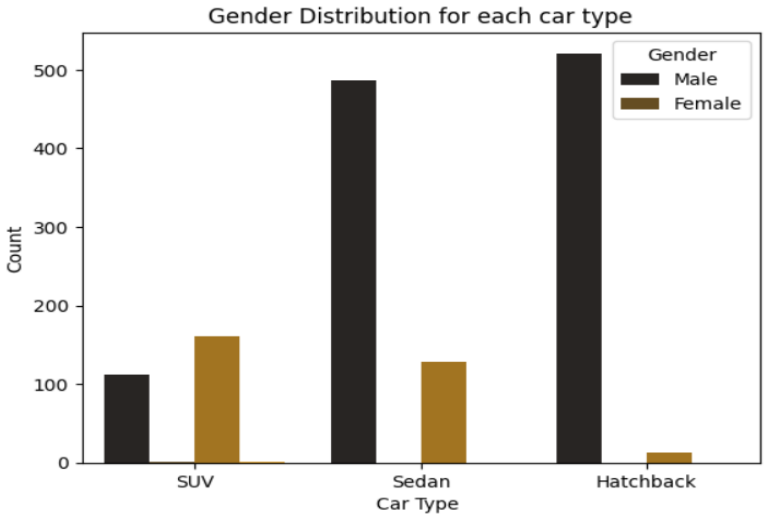


- **No much correlation between the salary and the price of car that they purchased.**

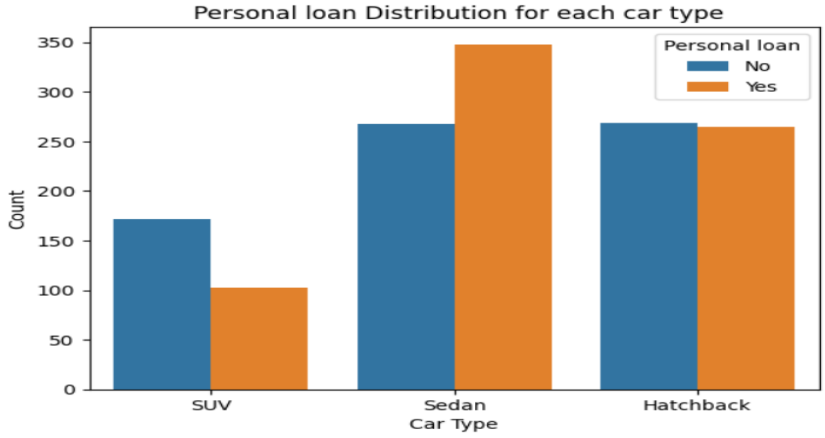
- **Based on the age group the price range of the car purchased is similar.**



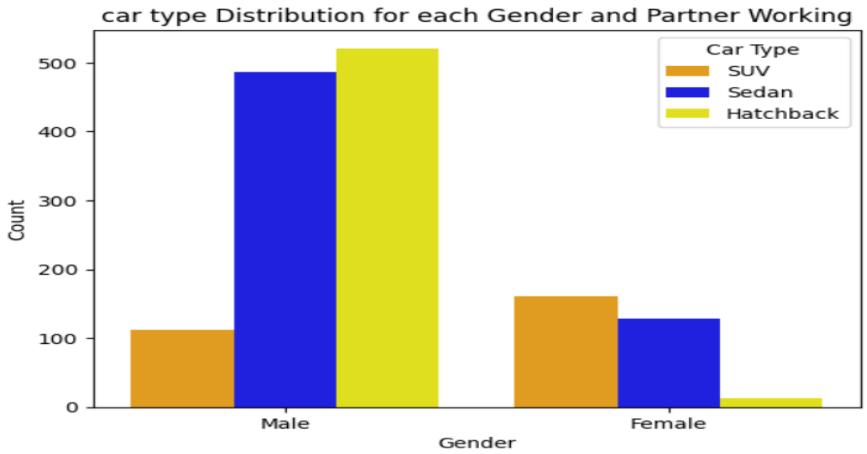
- **SUV cars are more preferred by females, followed by sedan and then hatchback.**



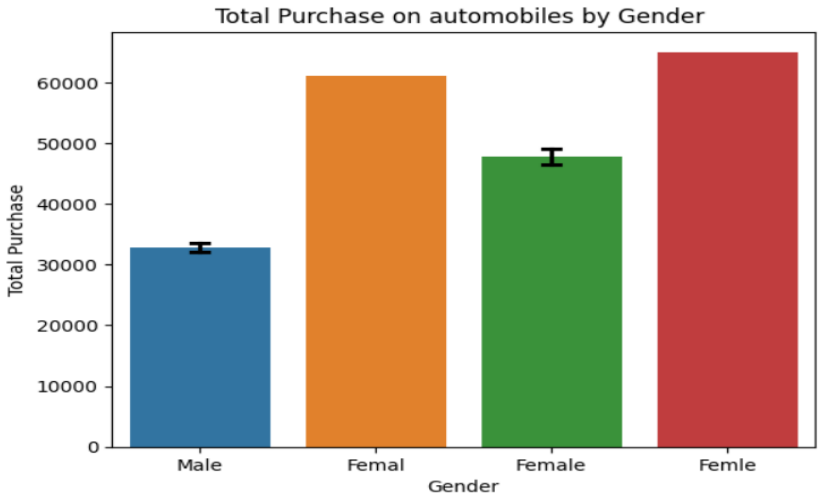
- **A salaried person (likely to have a personal loan) prefers sedan than any other car.**



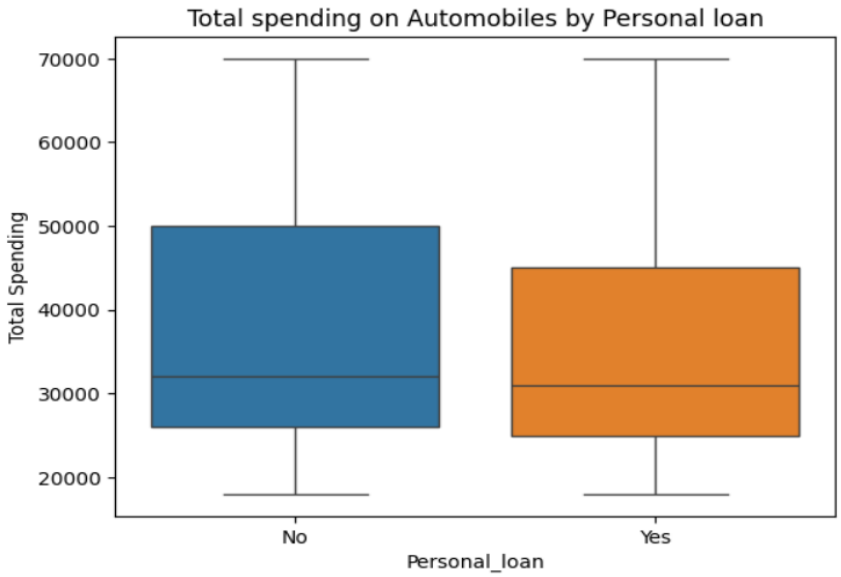
- By examining this plot, we can find if the salaried males show a higher preference for SUVs compared to Sedans. Justification:If Sheldon's claim is true , we should observe a higher number of salaried males choosing SUVs over Sedans compared to other groups(salaried females and non-salaried males and females).



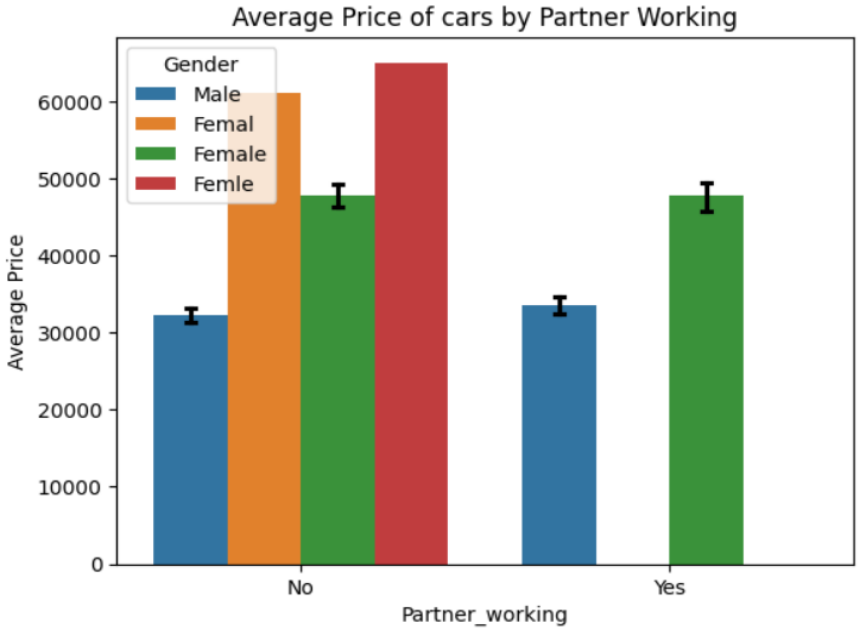
- If we observe a significant difference between males and females , the business can utilize this information to tailor marketing strategies and promotions for each gender.For example, if males tend to spend more, the business can offer premium or high end car models , while for females, they can focus on providing options that align with their preferences and budget.



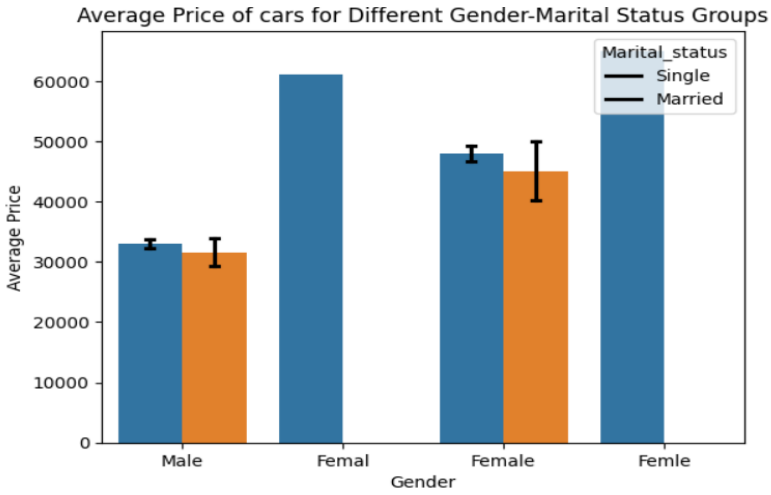
- **If we observe a difference in the distribution of spending between customers with and without personal loans, the business can utilize this information to create targeted offers.For instance, customers with personal loans may have different financial considerations , so the business can offer financing options or deals that cater to their specific needs.On the other hand , customers without personal loans may prefer one-time payment options or incentives for immediate purchase.**



- **Working partner influence the purchase of higher-priced cars.**



E.The main objective of this analysis id to devise 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.



Actionable Insights & Recommendations for marketing strategy

- **Singe males and females:If the average price of cars is higher for singles,they may prefer premium car models.Target them with campaigns showcasibg luxury features.**
- **Married males and females:If the average price of cars is higher for married customers, they may be interested in family-oriented or SUV models.Market spacious and safe vehicles to be assigned for them.**
- **Gender-specific Preferences:Compare average prices between single males and females ,as well as married males and females, to identify gender-specific preferences.Tailor marketing to address their unique needs.**