# Graded Project SMDM

## Problem 1

**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.**

**A. What is the important technical information about the dataset that a database administrator would be interested in?**

The 1581 records in the given data set include details on the purchaser's total income, gender, marital status, loan status, number of dependents, and the type and cost of the car they bought.

The size of the data and the data types for each variable would be of interest to the database administrator. In addition, the administrator would be concerned with **replacing any null values (Gender & Partner Salary) and deleting duplicates, useless data, and bad values** from the data.

```
df.info()
```

```
<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
```

```
df.shape
```

```
(1581, 14)
```

**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?**

Yes, there are discrepancies in the given data. It contains null values and unwanted information.

Under Partner salary and Gender, there are numerous null values.

```
In [14]: df=pd.read_csv("austo_automobile+(2)+(1).csv")
         df.isnull().sum()

Out[14]: 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
```

So, we can replace the missing values by using below mechanism.

**Partner_salary:**

df['Partner_salary'] = impute.fit_transform(df[['Partner_salary']])

I've used KNNImputer algorithm to get nearest value to replace the missing value of partner salary.

**Gender:**

Simple new values like "Others" can be created to replace null values with gender (categorical data).

We are aware that "Femal" and "Femle" should be replaced with "Female." I therefore changed the same.

```
df["Gender"].fillna("Others", inplace=True)
df['Gender']=np.where(df['Gender']=="Femal", "Female", df['Gender'])
df['Gender']=np.where(df['Gender']=="Femle", "Female", df['Gender'])
```

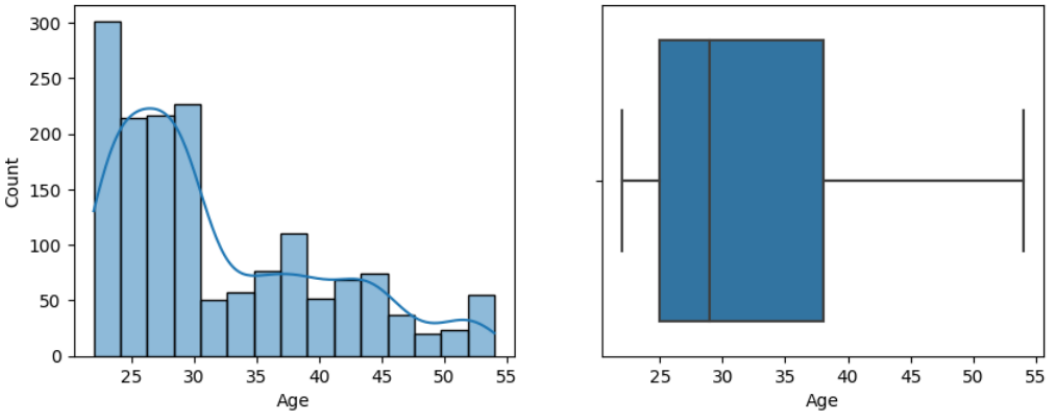
**Education:**

We can omit the "Education" section as we already have a salary. Thus, whoever has a larger salary will buy the car. We therefore do not need an Education column here.

df.drop(['Education'], axis=1, inplace=True)

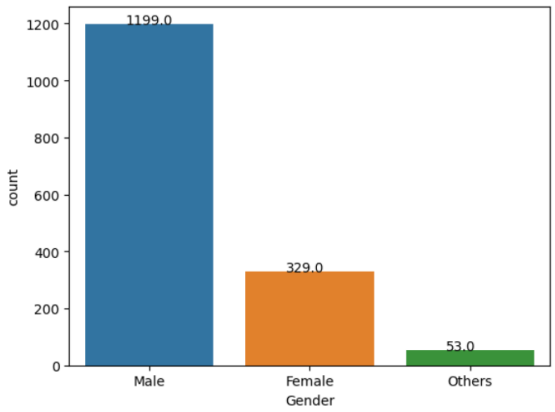
C. **Explore all the features of the data separately by using appropriate visualizations and draw insights that can be utilized by the business.**

**Age:**



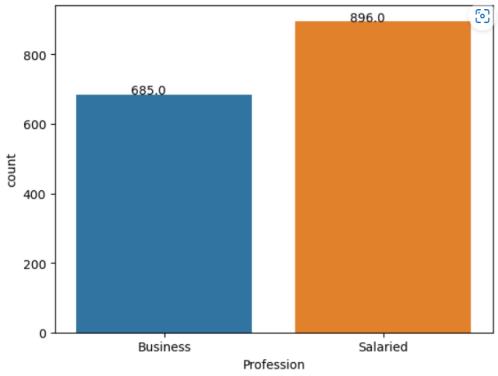

The data appears to be right-skewed, and the average value is 29. And the majority of the buyers were between the ages of 25 and 38.

**Gender:**



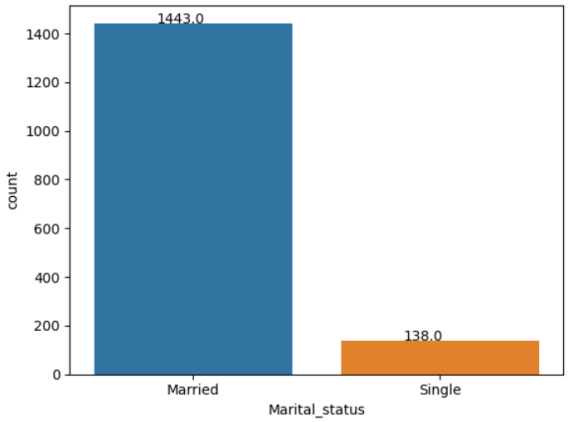

- More cars are purchased by men than by women.

**Profession:**



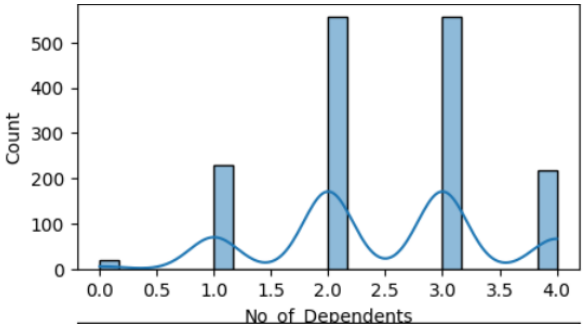

- 57 percent of consumers are salaried.
- Business owners make up 43% of purchases.

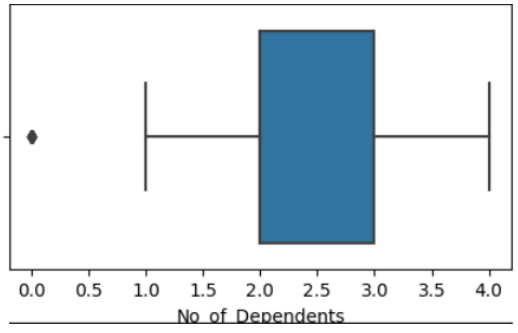
**Marital_status:**




- Compared to single people, married people made up the majority of car buyers.

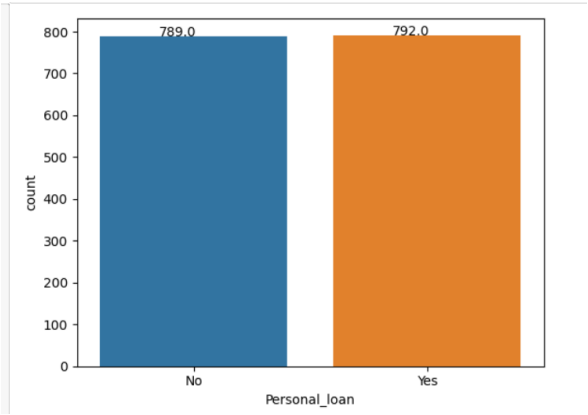
**Number of Dependents:**



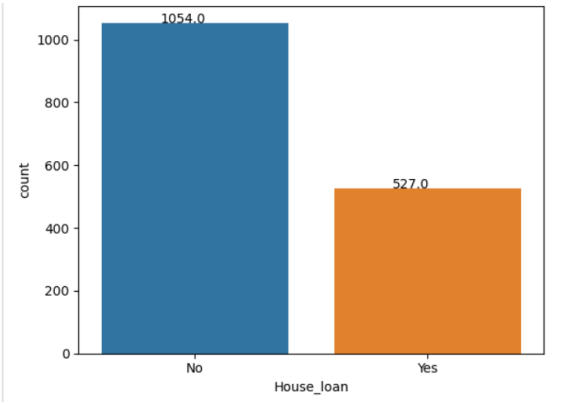

- Those with dependents between two and three bought more cars.
- Minimum dependent value is 1
- Maximum dependents value is 4

**Personal Loan**



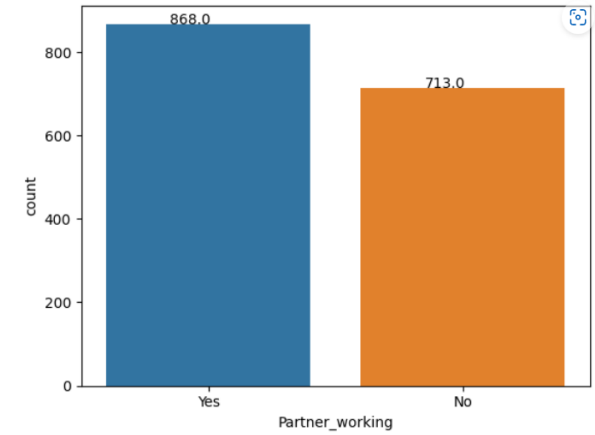

- The half of the people who purchase cars do so with personal loans.

**House Loan**



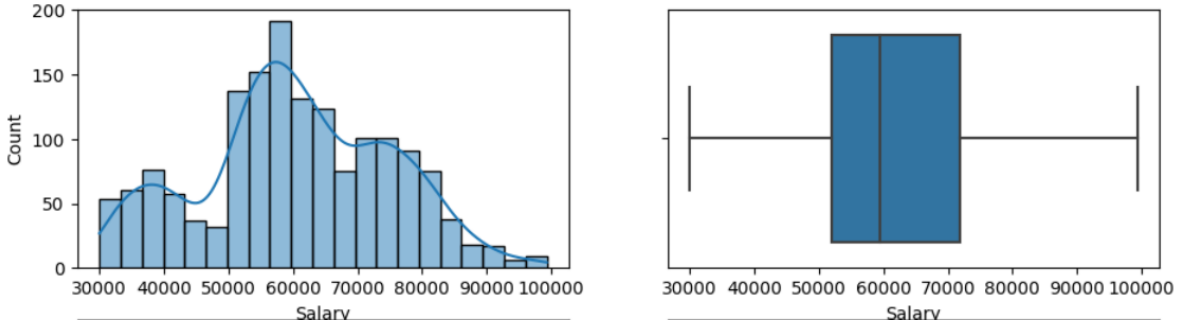

- The percentage of persons who bought cars while having house loan is only 30%.

**Partner working**



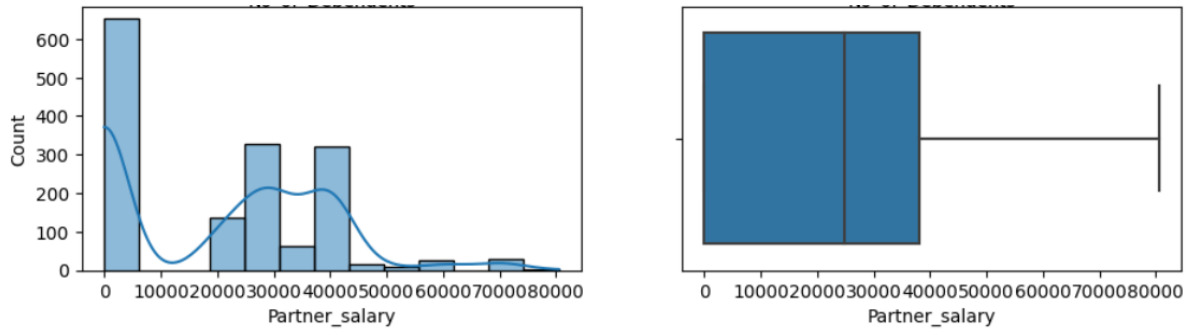

- 54% of buyers of the cars whose partners are employed.

**Salary**



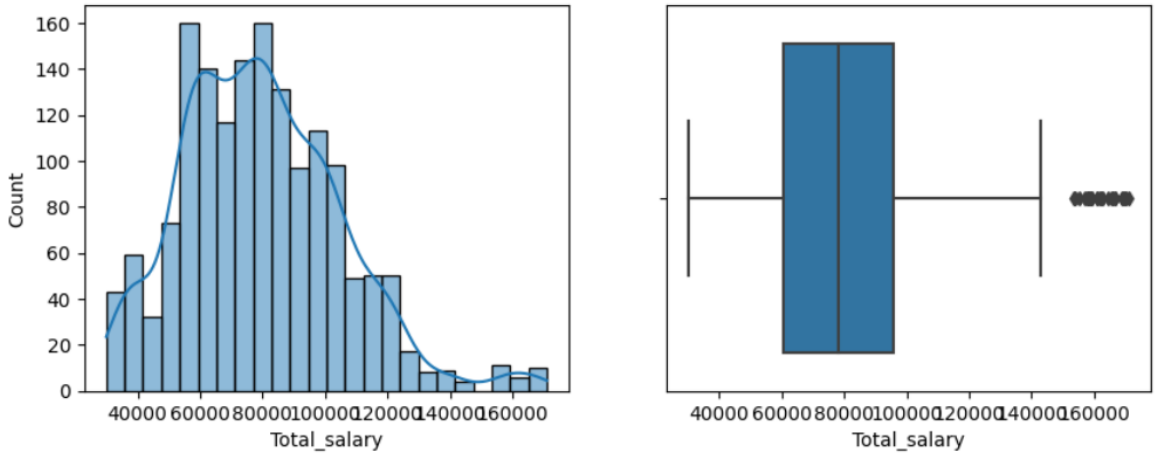

- 50% of car buyers make income between 30,000 and 59500.
- 75% of car buyers make income between 30,000 and 71,800.
- Minimum salary is 30,000.
- Maximum salary is 99300.

**Partner Salary**




- 45% of partners are unemployed, while the average partner earnings is 24,000.
- Maximum of partner salary is 80500.

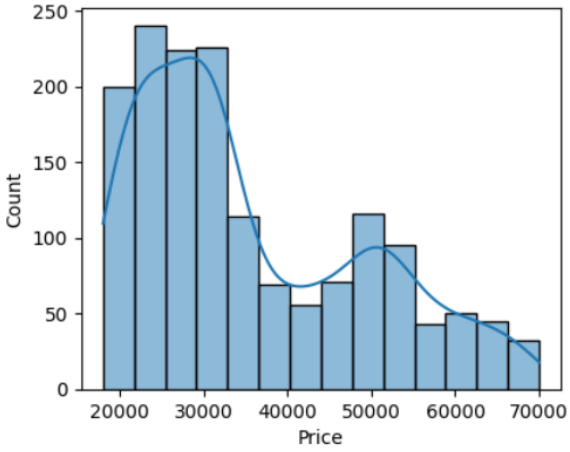
**Total Salary**

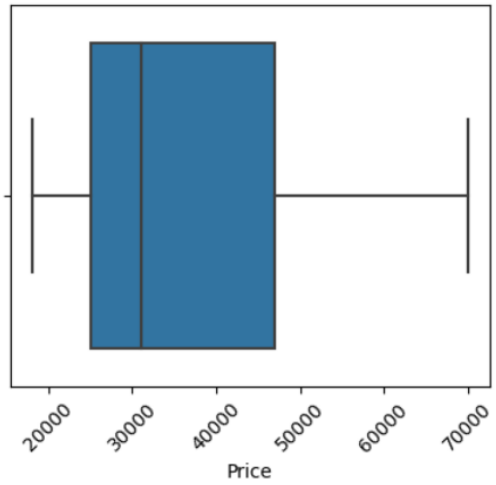



- 50% of car buyers make total income between 30,000 and 78000.
- 75% of car buyers make total income between 30,000 and 95590.

- Maximum total income value is 171000.
- Minimum total income value is 30,000.

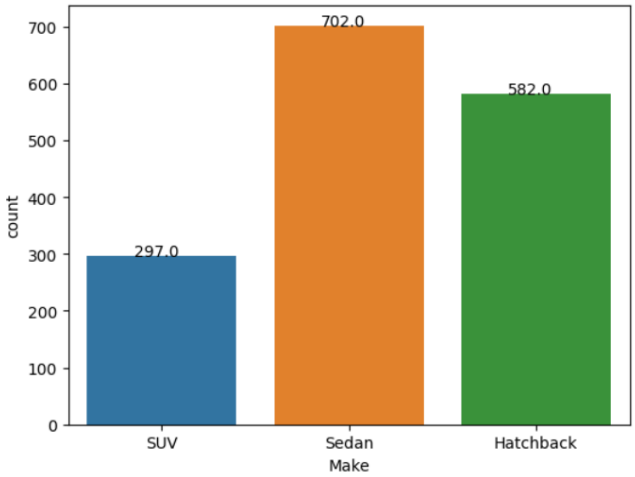
**Price**




- Average price of car is 31000.
- Minimum value of car is 18000.
- Maximum value of car is 70000.
- 50% of  the people purchased vehicles between the price 18,000 and 31,000

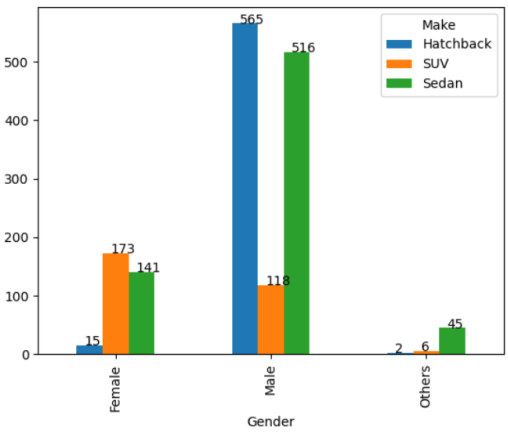
**Make:**




- 44% of purchasers selected sedan-style cars.
- 37% of purchasers chose hatchback-style cars.
- 19% of purchasers selected SUV-style cars.

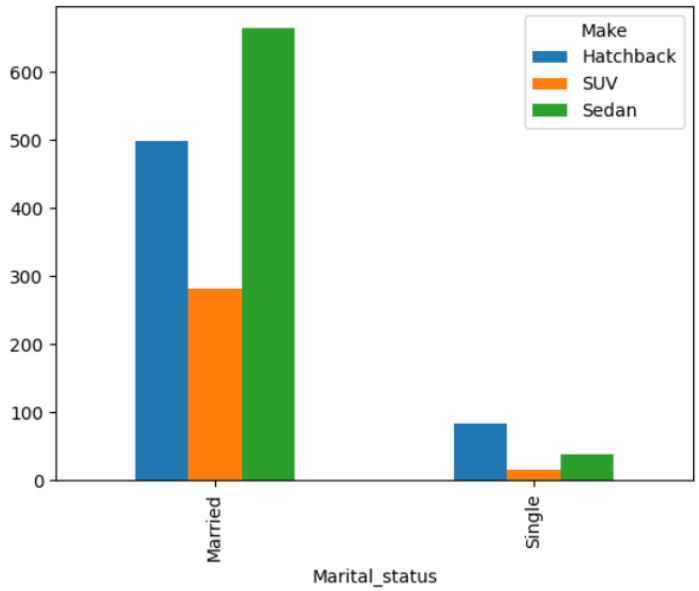
D. Understanding the relationships among the variables in the dataset is crucial for every analytical project. Perform analysis on the data fields to gain deeper insights. Comment on your understanding of the data.

**Gender vs Make**



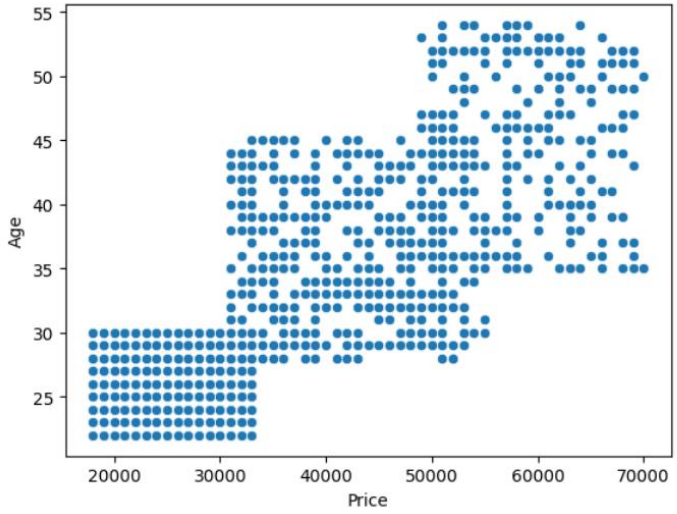

Males are more likely to purchase sedan and hatchback style vehicles than SUV style vehicles. When compared to hatchbacks, SUVs are more appealing to women buyers.

**Marital Status vs Make**



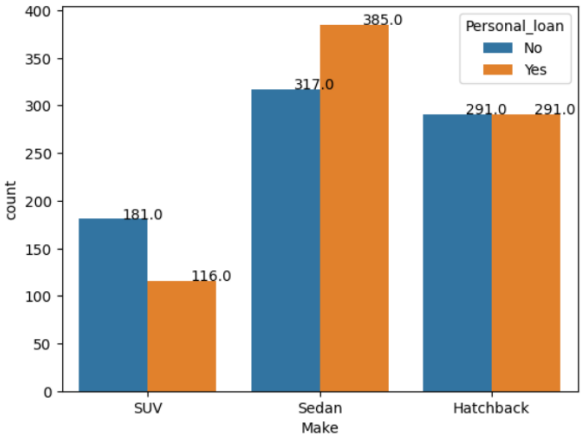

- Married folks are more drawn to sedan-style vehicles.
- The hatchback style of vehicles is more appealing to single people.

**Price vs Age**



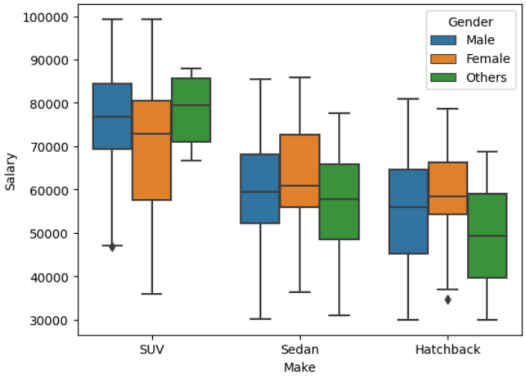

- The majority of elderly folks purchase the most expensive vehicles.
- Middle-aged people are drawn to moderately priced vehicles.

**Make vs Personan Loan**

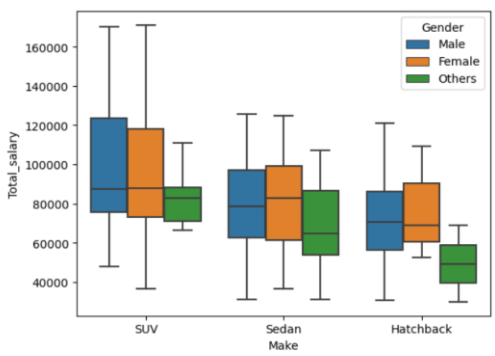



- 25% of buyers of sedan-style vehicles took out personal loans.
- 18% of those who purchased hatchback vehicles did so using a personal loan.

**Make vs Salary**

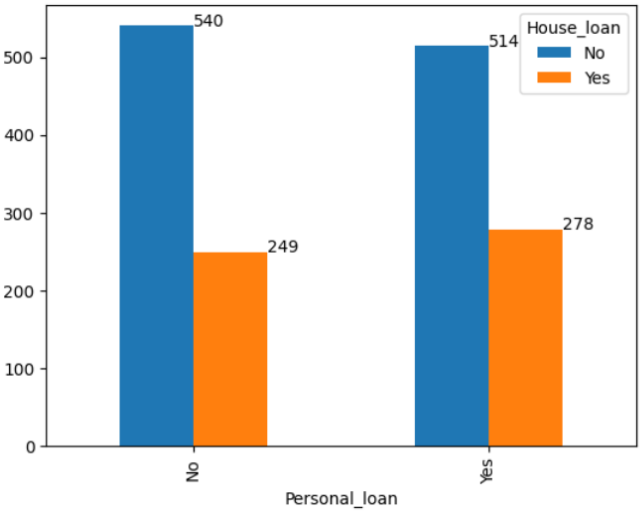



**Make vs Total Salary**



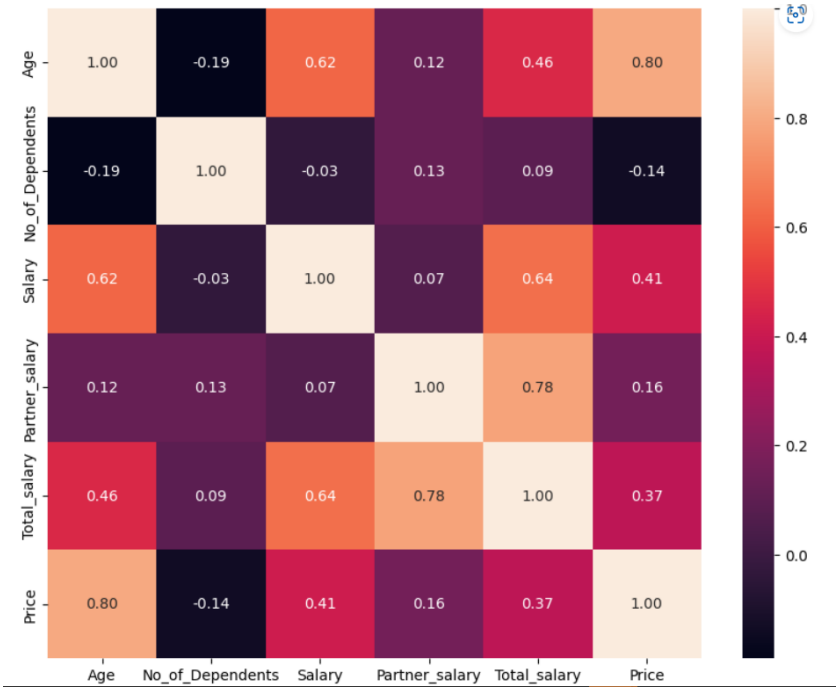

- Those who earned more money bought SUV-style vehicles.
- Vehicles of the sedan type were bought by people making income between 30,000 to 130000.

**Make vs Personal Loan vs House Loan**




- Only 18% people bought cars when they had both personal loans and house loans.
- 15% of consumers purchased cars without using personal or house loans.

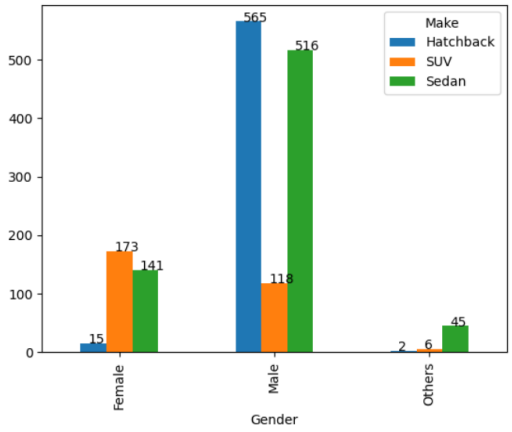
Co relation




- Price and Age is having more co relation and they are positively co related.

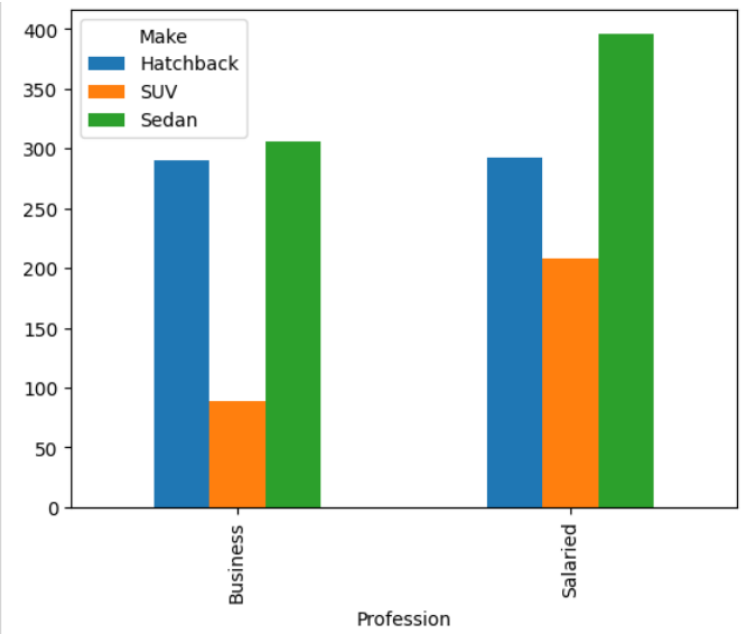
**E. Employees working on the existing marketing campaign have made the following remarks. Based on the data and your analysis state whether you agree or disagree with their observations. Justify your answer Based on the data available.**

E1) Steve Roger says “Men prefer SUV by a large margin, compared to the women”




According to the graphic portrayal, women are more likely than males to purchase an SUV.

E2) Ned Stark believes that a salaried person is more likely to buy a Sedan.



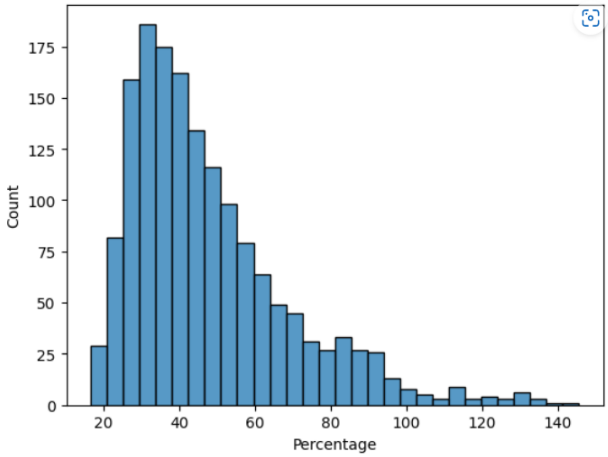

Certainly, those with salaries purchased a greater quantity of sedan cards.

E3) Sheldon Cooper does not believe any of them; he claims that a salaried male is an easier target for a SUV sale over a Sedan Sale.

No, salaried individuals are more drawn to sedans than SUVs.

**F. From the given data, comment on the amount spent on purchasing automobiles across the following categories. Comment on how a Business can utilize the results from this exercise. Give justification along with presenting metrics/charts used for arriving at the conclusions.**

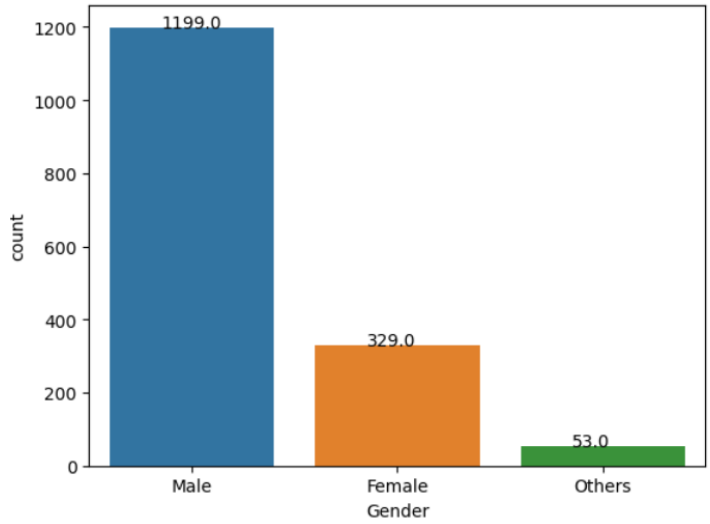
Totally 56280000 has spent to purchase the vehicles.




- Percentage of total income spent on buying cars.

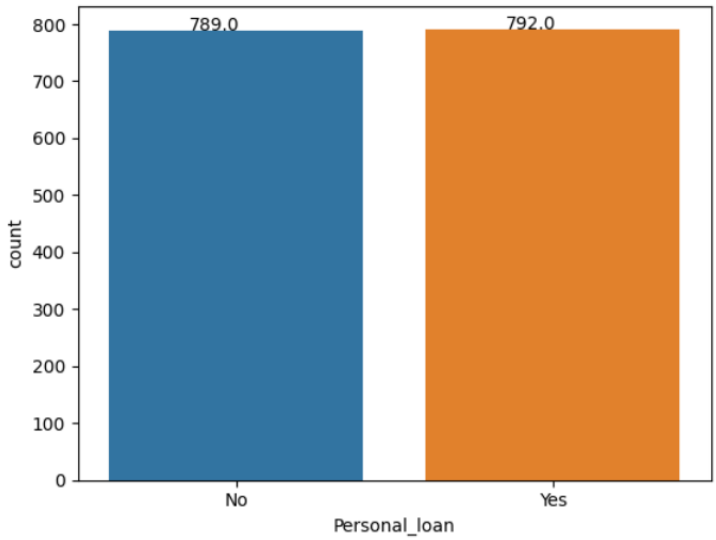
Give justification along with presenting metrics/charts used for arriving at the conclusions.

F1) Gender




- Men are more interested in buying cars than women are.
- Only 25% of women bought cars, as opposed to 75% of men.

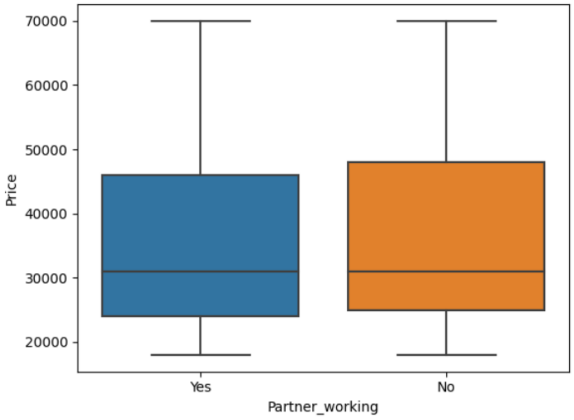
F2) Personal_loan




- The half of the people (792) who purchase cars do so with personal loans.

G. From the current data set comment if having a working partner leads to the purchase of a higher-priced car.

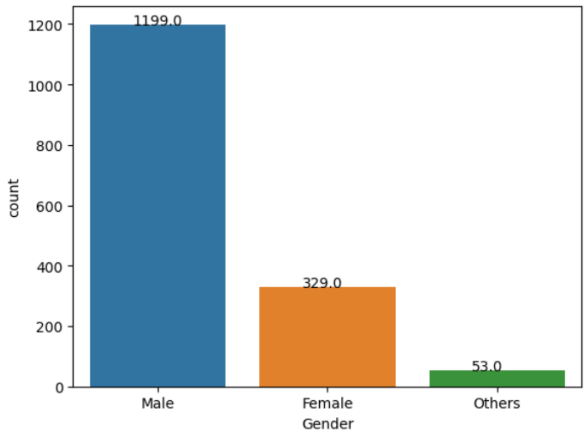
```
In [119]: sns.boxplot(x='Partner_working', y='Price', data=df)
Out[119]: <AxesSubplot: xlabel='Partner_working', ylabel='Price'>
```



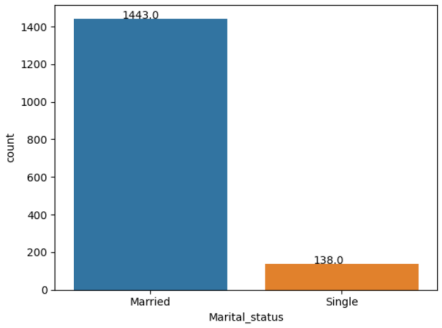

- The data appears to be slightly right-skewed when we take into account the working partner. Yet, we don't notice any leads for more expensive car purchases.
- In both instances, it is clear that the car priced at 70,000.00 is more expensive.

**H. The main objective of this analysis is 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.**

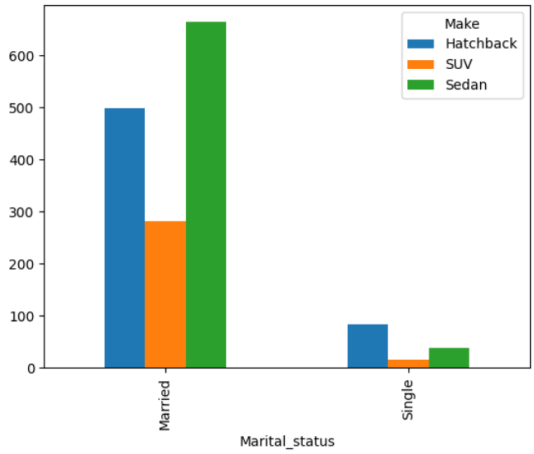
**Gender**



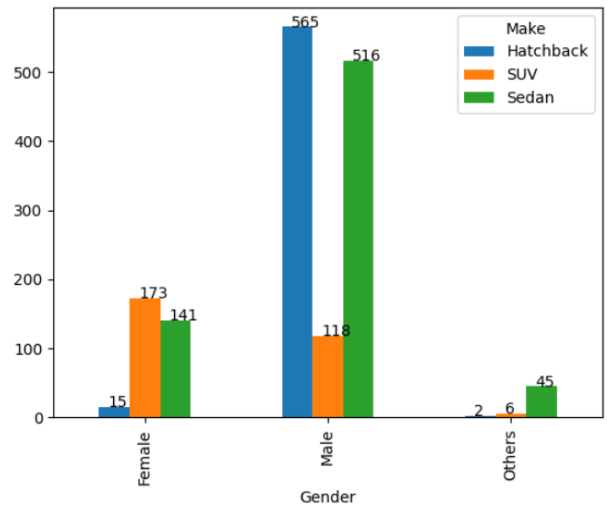
## Marital Status




## Marital Status Vs Make




## Gender Vs Make




- Compared to single people, married people made up the majority of car buyers.
- Married folks are more drawn to sedan-style vehicles.
- The hatchback style of vehicles is more appealing to single people.

## Problem 2

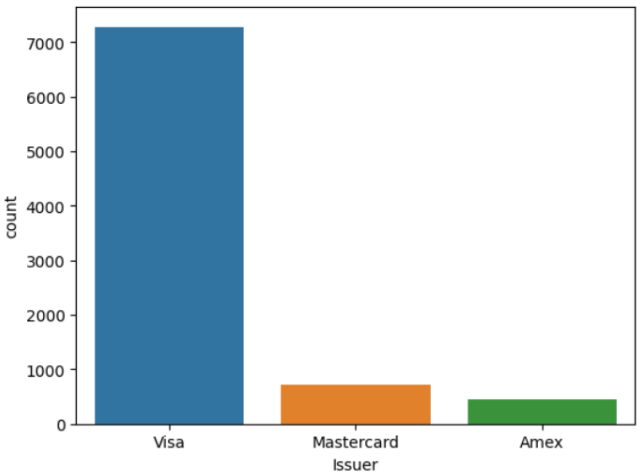
A bank can generate revenue in a variety of ways, such as charging interest, transaction fees and financial advice. Interest charged on the capital that the bank lends out to customers has historically been the most significant method of revenue generation. The bank earns profits from the difference between the interest rates it pays on deposits and other sources of funds, and the interest rates it charges on the loans it gives out.

GODIGT Bank is a mid-sized private bank that deals in all kinds of banking products, such as savings accounts, current accounts, investment products, etc. among other offerings. The bank also cross-sells asset products to its existing customers through personal loans, auto loans, business loans, etc., and to do so they use various communication methods including cold calling, e-mails, recommendations on the net banking, mobile banking, etc.

GODIGT Bank also has a set of customers who were given credit cards based on risk policy and customer category class but due to huge competition in the credit card market, the bank is observing high attrition in credit card spending. The bank makes money only if customers spend more on credit cards. Given the attrition, the Bank wants to revisit its credit card policy and make sure that the card given to the customer is the right credit card. The bank will make a profit only through the customers that show higher intent towards a recommended credit card. (Higher intent means consumers would want to use the card and hence not be attrite.)

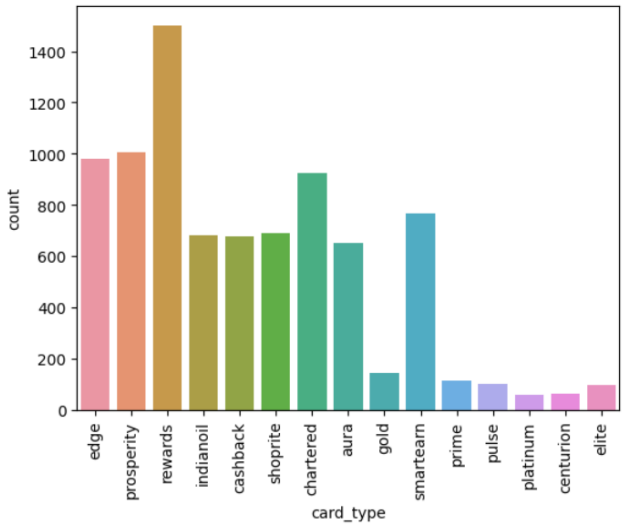
*****Framing An Analytics Problem*** Analyse the dataset and list down the top 5 important variables, along with the business justifications.**

**Issuer:**



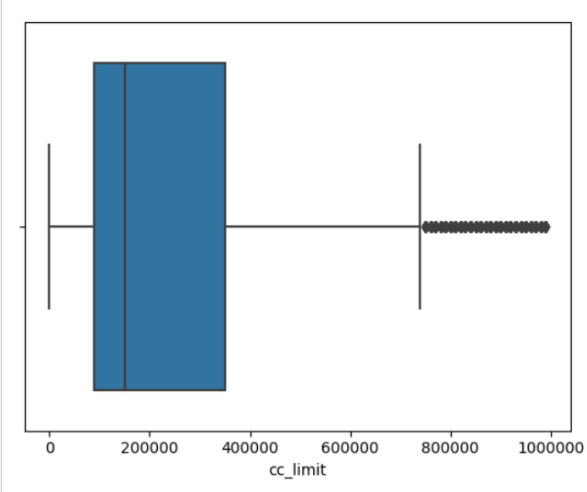

- Most of the credit card holders are using visa cards.
- To raise the number of MasterCard and Amex cards still in circulation, we must comprehend the Visa offers and customer accountability.

**Card Type**



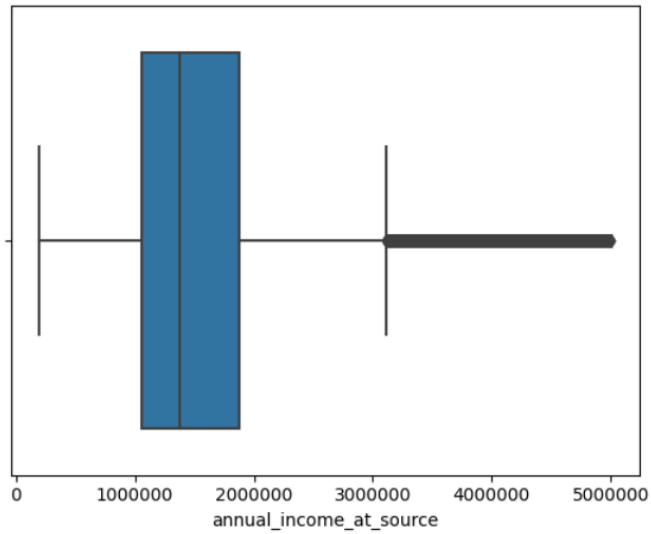

- Rewards cards are more popular among card users.
- Centurion, Elite, and Platinum have fewer customers.

**CC Limit**



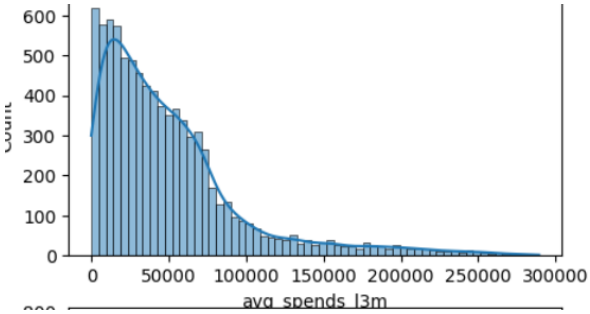

- Average limit of Credit Card limit is 150000.
- Minimum Credit Card limit is 0.
- Maximum of Credit card limit is 990000.
- 75% of people using card limit between 0 to 350000.
- There are multiple higher whiskers value presented in the Credit Card Limit.

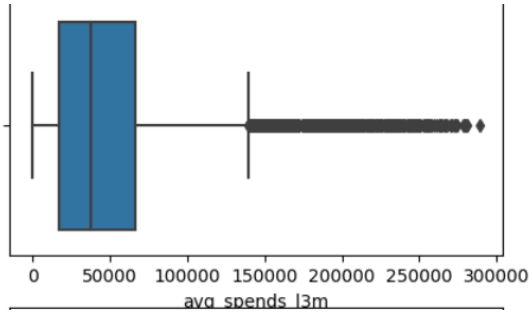
**Annual Income at source**




- 25-75% people using credit card and who is earning between 1061104 and 1881734.

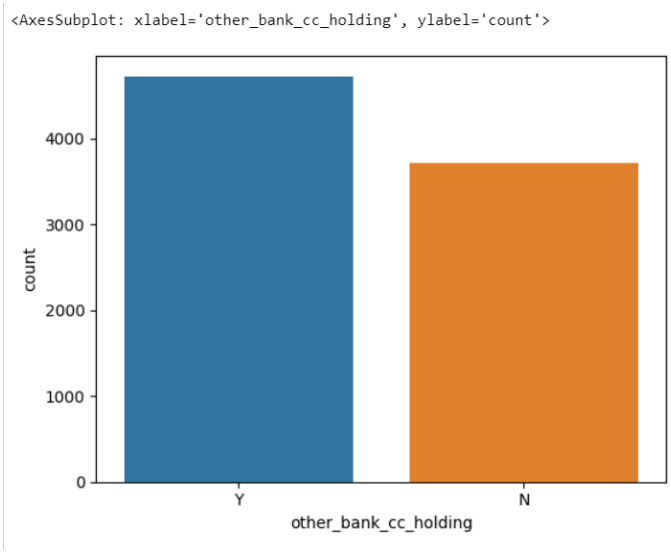
**Average Spends**




The monthly average amount spent is 37943.

**Other card holding:**




- Increasingly people are using two credit cards simultaneously.