SMDM PROJECT
BUSINESS REPORT

NAME: JAYA PREETHI R M

CONTENT

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.

You as an analyst have been tasked with performing a thorough analysis of the data and coming up with insights to improve the marketing campaign.

The instructions below are given to help you complete the project –

A.What is the important technical information about the dataset that a database administrator would be interested in? (Hint: Information about the size of the dataset and the nature of the variables)

A database administrator would be interested in the following technical information about the dataset:

(a) The Total number of entries in the dataset. This information will be important because it helps us understand the size and nature of the sample and ascertain whether the statistical analysis would provide an empirical answer to our business problem. It can also. With the given dataset I learned that there are 1581 entries in total.
(b) The variables and the nature of such variables. The variables give us an essential insight into how we can use the available information to solve the business problem. Similarly, It can also help us create a framework on how we can establish relationships during Exploratory data Analysis. Another important information pertaining to variables is distinguishing the dependent variables from the independent variables.

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.

To have a primary understanding of the dataset, I loaded the shape of the dataset and the first 5 rows of the dataset.

```
(1581, 14)
```

From the following output, we can see that the dataset has 1581 rows and 14 columns. Therefore, there are a total of 1581 entries and 14 different variables within the 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

From the given table, we can take a look into the data, see how each variable has been coded and what they actually mean. For example, we can see that for the number of dependents, they have taken the total number of dependents but not any further details that are any more descriptive.

Next step is to look into the nature of the variables and the total number of variables under each one of them.

cont.

```
<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
```

From here, we can see that there are 6 numerical variables and 8 categorical variables. There were some inconsistencies or errors with the data that should be dealt with.

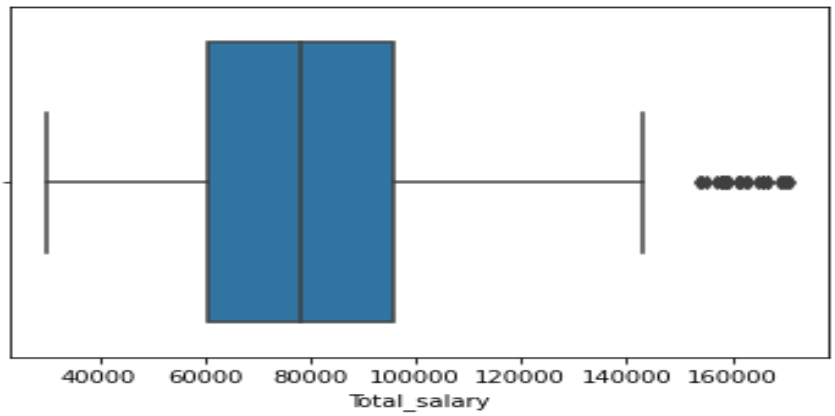
a) The total entries in the Gender variable(1528 entries) is lesser than the other variables. Thus,we can see there are some null variables.

```
array(['Male', 'Femal', 'Female', nan, 'Femle'], dtype=object)
```

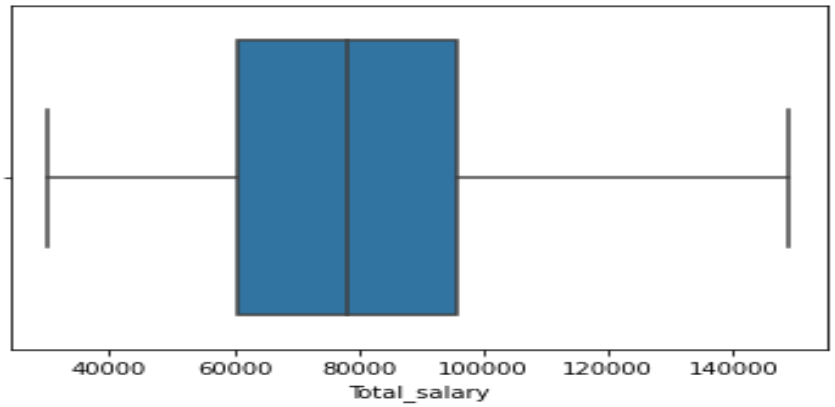
Here we can see that there are 5 unique inputs for the variable. Out of which three of them are to be dealt with, due to incorrect input. I was able to change the spelling errors of 'Femal' and 'Femle' into the correct spellings using replace function. Similarly, I had also imputed the mode of the Gender variable, i.e., Male into the null variable 'nan'.

b) The total entries of Partner's salary variable (1475 entries) is lesser than other variables. This inconsistency was due to the fact that some consumers were either single or had partner's who were not working. Thus, they have entered 0 in the Partner's Salary variable, which was coded as a null variable. I decided that I should not impute this variable because this was not misinformation and imputing this variable can warp the results of my statistical output.

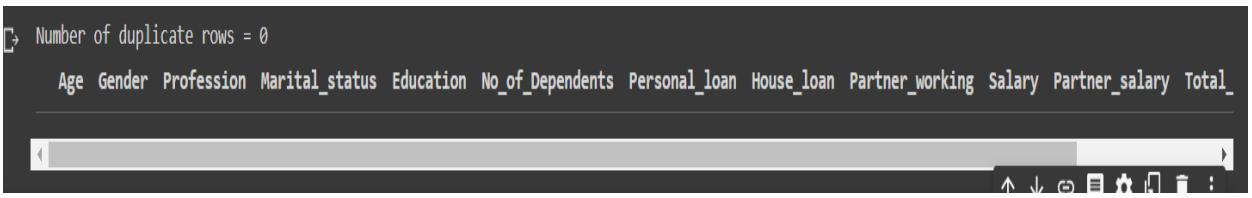
c) While checking outliers, the total salary variable had some outliers beyond the upper limit.



To give the variable an outlier treatment, I imputed the upper range value, i.e. $1,49,000 to the outlier values, i.e. inputs that are larger than the upper limit. Resultantly, I got the following boxplot output of the variable.



I had also checked for the number of duplicated rows in the dataset, but the result showed that there were no duplicates.



---

	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

From the statistical summary given above, we can see that the minimum for the Number of Dependents variable is 0. This cannot be considered an inconsistency, because some consumers can have no dependents. Thus, treating this variable can result in warped statistical analysis.

Along with this, As mentioned earlier, Partner's Salary variable has $0 as its minimum value, which also cannot be taken as an inconsistency in this specific case. Doing so can change our analytical output.

As a final stage of Pre-processing the dataset I had encoded four categorical variables - Marital Status, Personal Loan, House Loan, Partner's Working. Since all these variables were yes or no variables, I converted them to categorical variables from object variables. Then, I was able to encode these variables using the .cat.codes function. Following table was printed to check if the variables had been successfully encoded.

cont.

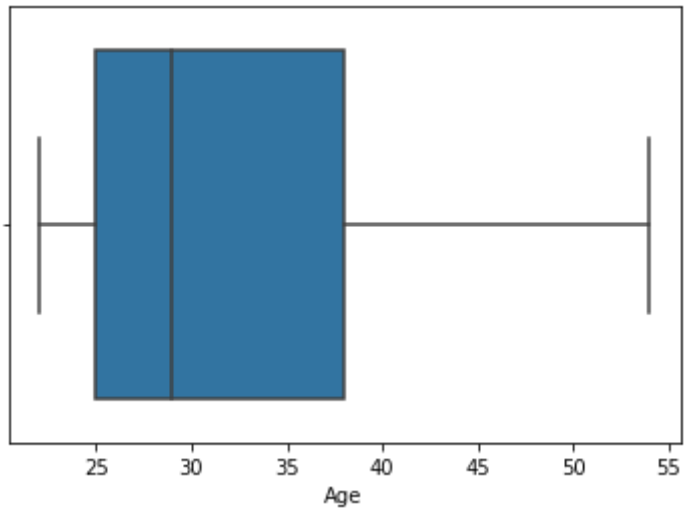
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1581 entries, 0 to 1580
Data columns (total 14 columns):
 #   Column            Non-Null Count  Dtype
---  ------            --------------  -----
 0   Age               1581 non-null   int64
 1   Gender            1528 non-null   object
 2   Profession        1581 non-null   object
 3   Marital_status    1581 non-null   int8
 4   Education         1581 non-null   object
 5   No_of_Dependents  1581 non-null   int64
 6   Personal_loan     1581 non-null   int8
 7   House_loan        1581 non-null   int8
 8   Partner_working   1581 non-null   int8
 9   Salary            1581 non-null   int64
 10  Partner_salary    1475 non-null   float64
 11  Total_salary      1581 non-null   float64
 12  Price             1581 non-null   int64
 13  Make              1581 non-null   object
dtypes: float64(2), int64(4), int8(4), object(4)
memory usage: 129.8+ KB
```

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

AGE

I have used boxplot to analyse the 'AGE' variable. This is because the boxplot could help us visualise the five point summary of a numerical variable.

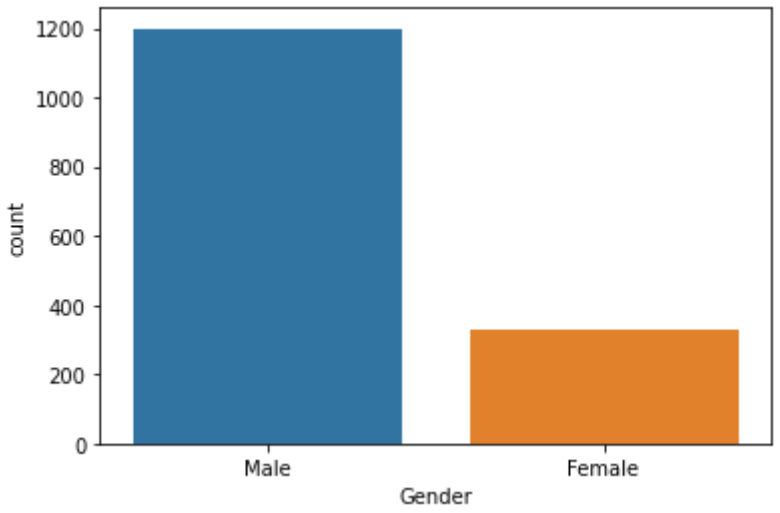
cont.



From the boxplot, we can see that most of our consumers are of age limit 22 to 54 years of age. This might be the result of purchasing an automobile being an investment for the majority of consumers and making such purchase at such a young age is not very much prevalent. As stated by *Thakuriah, Piyushmita, 2010*, "Evidence was found of an age effect, indicating that irrespective of the generation and sociodemographic grouping considered, the predicted probabilities of car ownership increase with age within young adulthood." We can also see that the variable is skewed more towards the right side and majority of the consumers fall in between the age range of 25 to 38 years.

GENDER

I have used a countplot / bar graph to visualise this variable, because this is a categorical variable and it can be better visualised this way for each comparison between the categories.

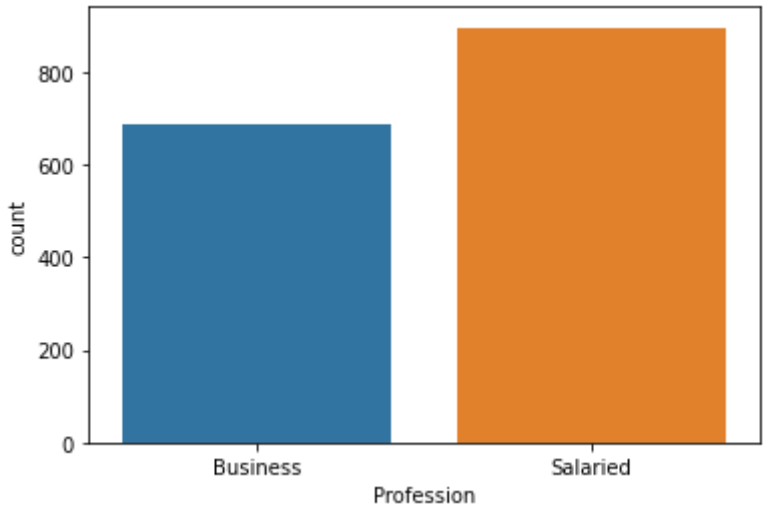


Male : 1199, Female : 329

Thus, from the bar graph, we can see that there are a significant majority of male customers than female customers, with 1199 Male customers and 329 Female customers in total. Thus, we can say that the majority of the automobile customers are of male demographic.

PROFESSION

I have used a countplot / bar graph to visualise this variable, because this is a categorical variable and it can be better visualised this way for each comparison between the categories.

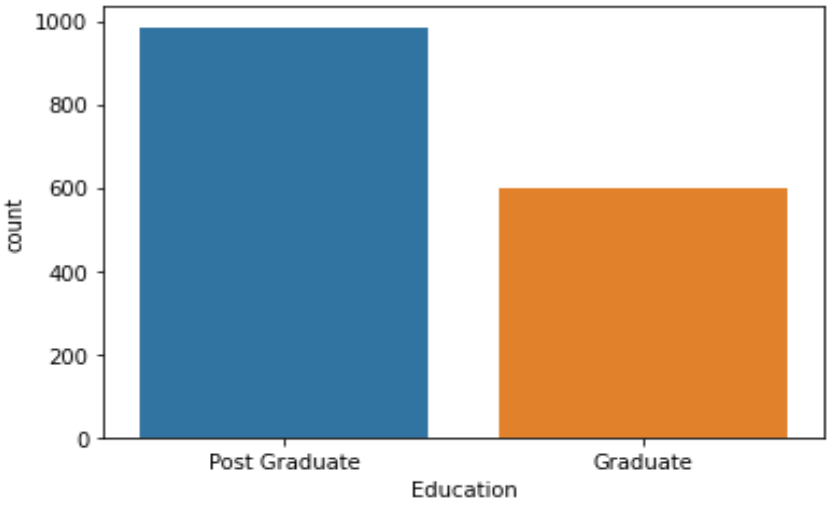


Salaried : 896, Business : 685

From the given graph, we can infer that Majority of the customers are salaried, i.e. they have been employed by a company by 896 in number. On the other hand, the minority of the customers are business entrepreneurs by 685 in number. We can also see that there are no major differences between the categories (only by 200 in numbers).

EDUCATION

I have used a countplot / bar graph to visualise this variable, because this is a categorical variable and it can be better visualised this way for each comparison between the categories.

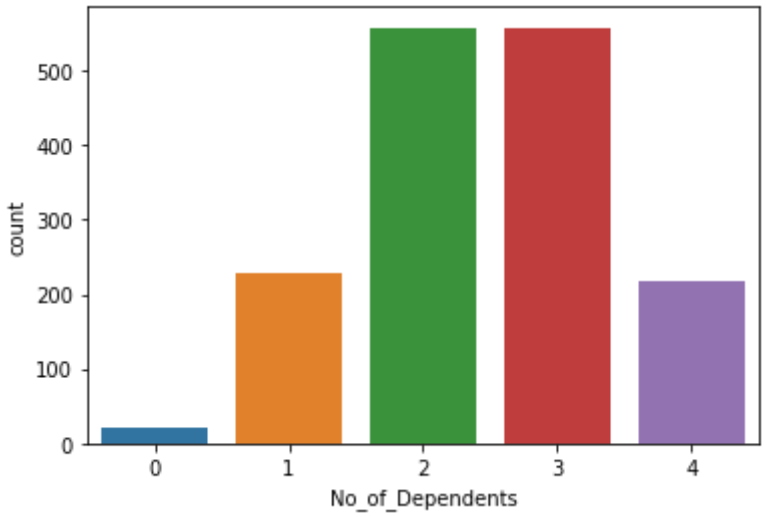


Post Graduate : 985, Graduate : 596

From the bar graph, we can infer that most of the customers are Post Graduate degree holders with 985 within the sample population. 596 customers of the sample population are Graduate degree holders, therefore keeping them in the minority.

NUMBER OF DEPENDENTS

I have used a countplot / bar graph to visualise this variable, because even though this is a numerical variable, I believe that it would be better to use bar graph in this situation since it would give us a better representation of the distribution than boxplot..

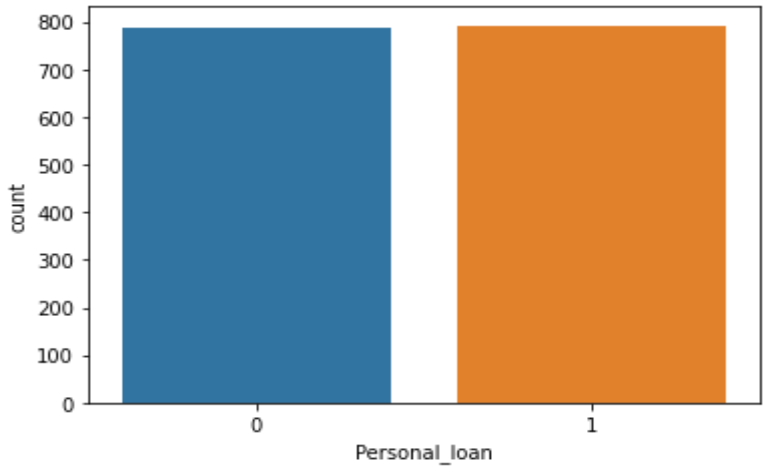


0 : 20	1 : 229	2 : 557	3 : 557	4: 218

From the bar graph, we can see that the majority of the consumers have 2 or 3 dependents with 557 consumers falling in both categories respectively. On the other hand, a small minority of 20 people from the sample population do not have any dependents. Thus, we can say that the majority of the customers have one or more than one dependents.

PERSONAL LOAN

I have used a countplot / bar graph to visualise this variable, because this is a categorical variable and it can be better visualised this way for each comparison between the categories.

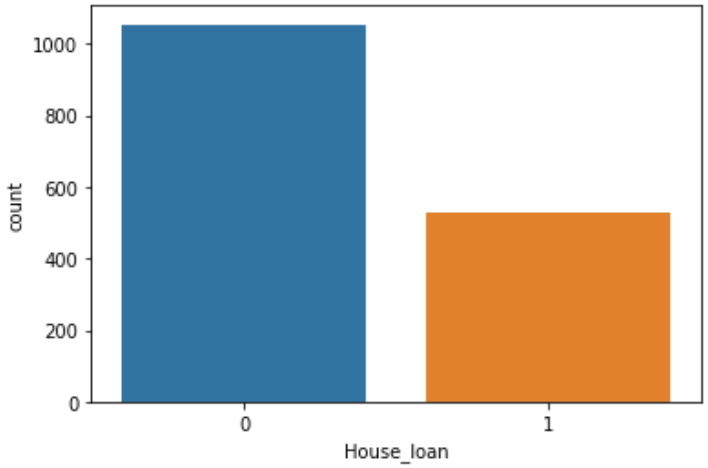


1(YES) : 792, 0(NO) : 789

From the given bar graph, we can see that the people who took personal loans (792) are not significantly larger in number than the people who did not take personal loan (789).

HOUSE LOAN

I have used a countplot / bar graph to visualise this variable, because this is a categorical variable and it can be better visualised this way for each comparison between the categories.

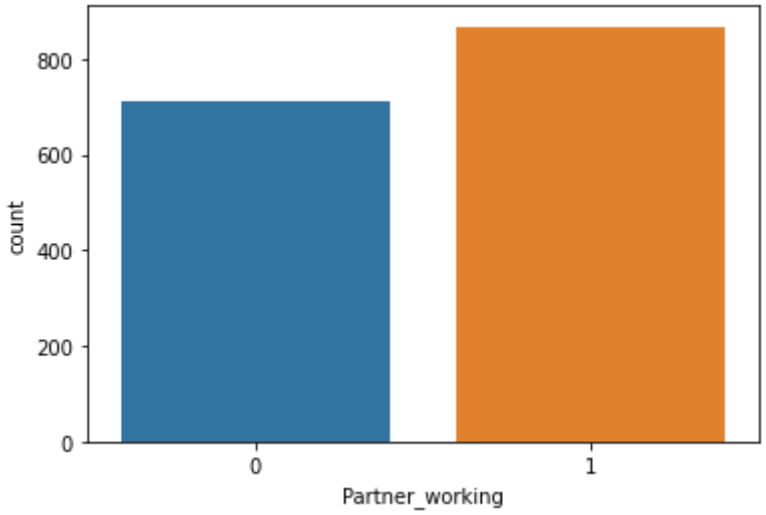


0 (No) : 1054, 1(Yes) : 527

From the following bar graph, we can see that the majority of the population has not taken a house loan with 1054 customers within the sample population and other 527 people in the sample population have taken a house loan.

PARTNER WORKING

I have used a countplot / bar graph to visualise this variable, because this is a categorical variable and it can be better visualised this way for each comparison between the categories.

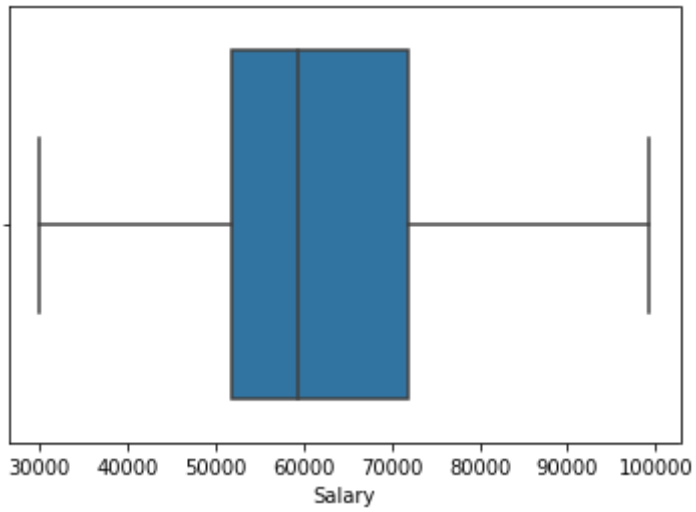


1 (Yes) : 868, 0 (No) : 713

From the bar graph, we can see that the majority of the sample population (868) have a working partner, i.e. they have an additional income within their household. On the other hand, 713 of the sample population do not have a working partner, thus they do not have an additional source of income in their household.

SALARY

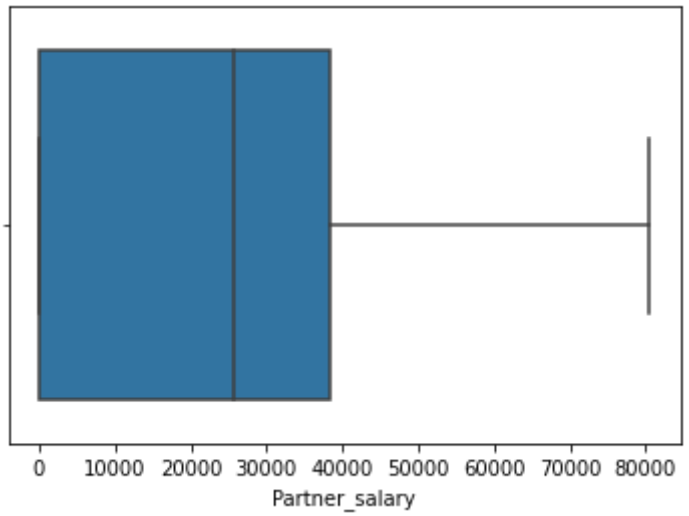
I have used boxplot to analyse the ‘AGE’ variable. This is because the boxplot could help us visualise the five point summary of a numerical variable.



The Salary variable has a lower and upper range of $30,000 and $99,300 respectively. We can also see that the data is not skewed to either side. Thus, the data is normally distributed. The majority of the income base falls in between $51,900 and $71,800. Thus we can infer that the salary of the customers are pretty much evenly dispersed and most of the customers have fallen in between the given range.

PARTNER’S SALARY

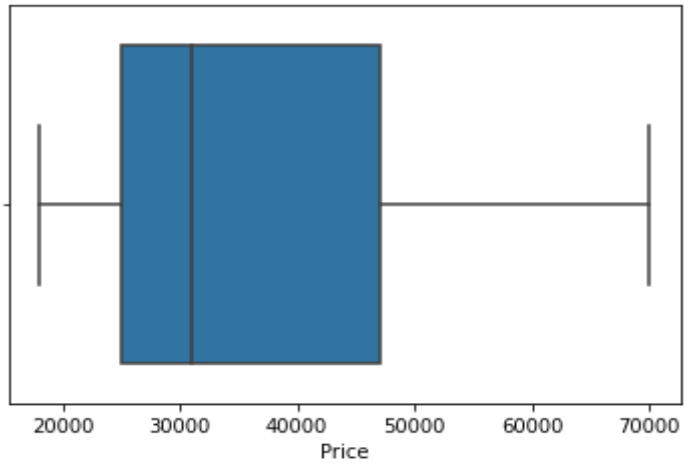
I have used boxplot to analyse the ‘PARTNER’S SALARY’ variable. This is because the boxplot could help us visualise the five point summary of a numerical variable.



We can see from the boxplot that the data has been very much skewed to the right. This is because a significant amount of consumer's have zero income generated from their partner's thus the lower limit is set at 0. Thus, the lower range and upper range is set at $0 and $80,500 respectively, with the majority of the population falling in between the salary range of $0 to $38,300.

PRICE

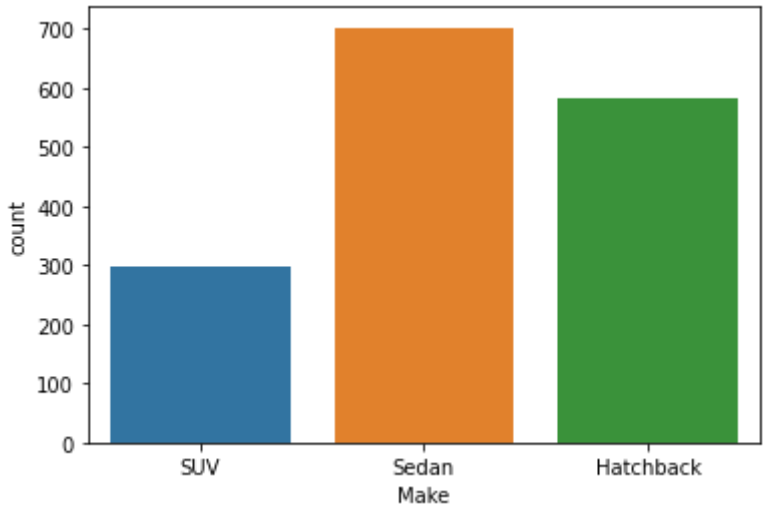
I have used boxplot to analyse the 'PRICE' variable. This is because the boxplot could help us visualise the five point summary of a numerical variable.



We can see from the boxplot that the data is skewed to the right, This is because the customers would prefer to buy cheaper automobiles than the expensive ones. The upper and the lower range of prices are $18,000 and $75,000, with the majority of consumers buying automobiles that are ranging from $25,000 to $47,000 in cost.

MAKE

MAKE variable tells us the different types of automobiles that are being purchased by the customers. I have used a countplot / bar graph to visualise this variable, because this is a categorical variable and it can be better visualised this way for each comparison between the categories.



Sedan : 702, Hatchback : 582, SUV : 297

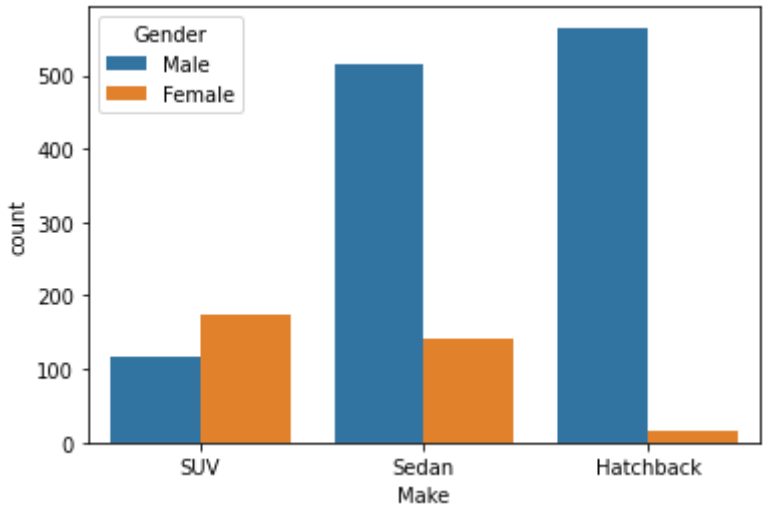
Majority of the sample population would like to purchase Sedan (702) as evidenced by the visualisation above. The second most desired make is Hatchback (297). The least purchased automobile among the sample population is SUV at 297 people buying the automobile.

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.

I have done several bivariable and multivariable analyses with the dataset. They are as follows:

a) GENDER & MAKE

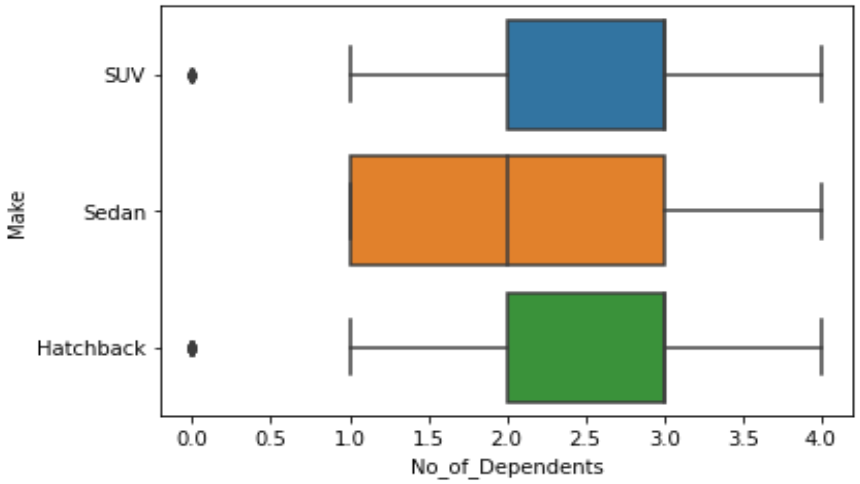
cont.



The bar graph gives us the visualisation of the distribution of preferences between the gender demographics. One major thing we can notice is that the customer base is far dominated by male demographic rather than the female demographic. We can see that among male demographic (being represented in blue bars), most men prefer Hatchback model automobiles and least prefer SUV model automobiles. Another significant part of the male demographic prefers Sedan model automobiles. On the other hand, the Female demographic contrastingly prefers SUV model automobiles the most and Least prefer Hatchback model automobiles. However, similar to the male demographic, female customers want Sedan model automobiles as their second most preferred automobiles. Thus, we can conclude that male consumers are more likely to buy Hatchback model automobiles than others, Whereas, female consumers are more likely to buy SUV model automobiles than others.

---

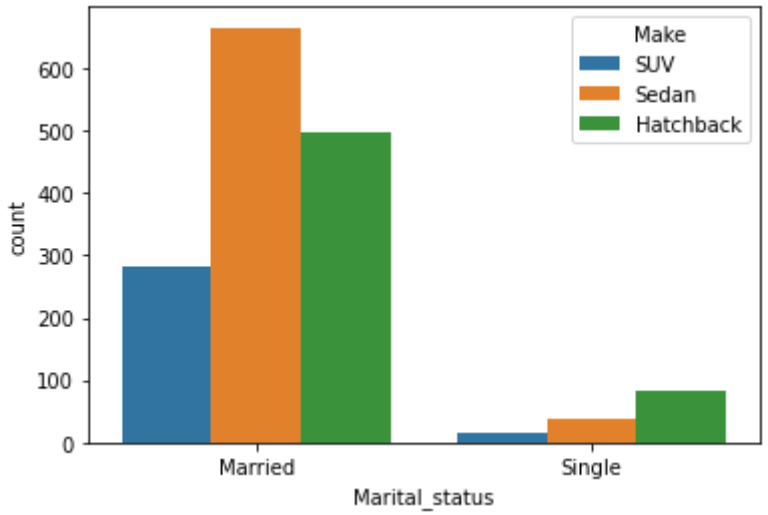
b) MAKE & NUMBER OF DEPENDENTS



From the given boxplots, we can analyse the relationship between the number of dependents of the consumer and the type of automobile they prefer. We can notice that all of the boxplots range in between 1 to 4 dependents with some outliers at no (0) dependents. While SUV and Hatchback data are not skewed either side, Sedan data are skewed towards the right hand side. Therefore, most customers with 2 to 3 dependent prefer SUV and Hatchback model automobiles. Whereas, customers with 1 to 3 dependents prefer Sedan model automobiles, Thus, having a wider range of customers than the other types. Thus, we can conclude that a consumer with any dependents is highly likely to buy a Sedan model automobile.

---

c) MARITAL STATUS & MAKE

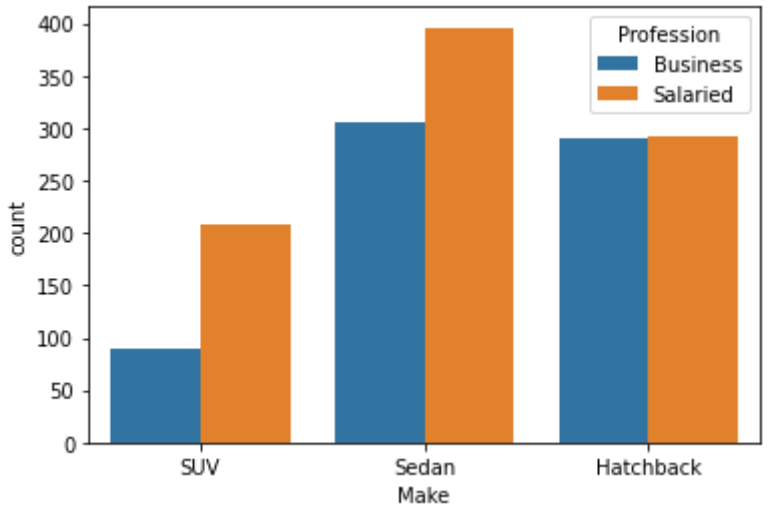


The above bar graph visualises the distribution of married and single customers among the three given makes of automobiles. First, we can notice that the majority of the customer base is married whereas the unmarried people are in the minority. Among the married population, the most preferred automobile is Sedan and the least preferred automobile is SUV. Among the unmarried population, the most preferred automobile is the Hatchback. Interestingly, In both the population, the least preferred make is SUV. Thus, we can conclude that Married consumers are most likely to purchase Sedan model automobiles and unmarried consumers are most likely to purchase Hatchback model automobiles.

---

d) MAKE & PROFESSION

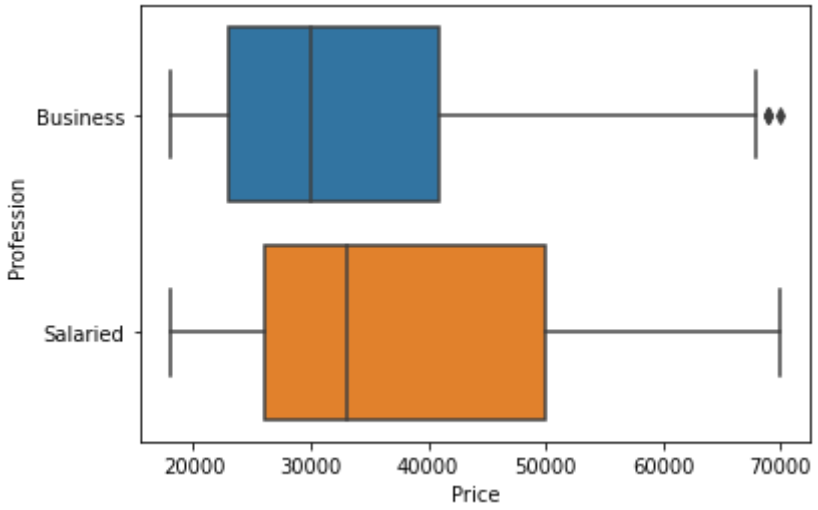
cont.



The above bar graph visualises the distribution of the sample population's preference on the make of the automobile on the basis of their profession. From observing the graph we can see that all the preferences are similar for both salaried and business owning individuals. Both the categories prefer Sedan model automobiles the most, Hatchback the second most and SUV the least. Therefore, we can conclude that consumers of any profession are more likely to buy Sedan model automobiles than SUV model automobiles.

---

e) PROFESSION & PRICE

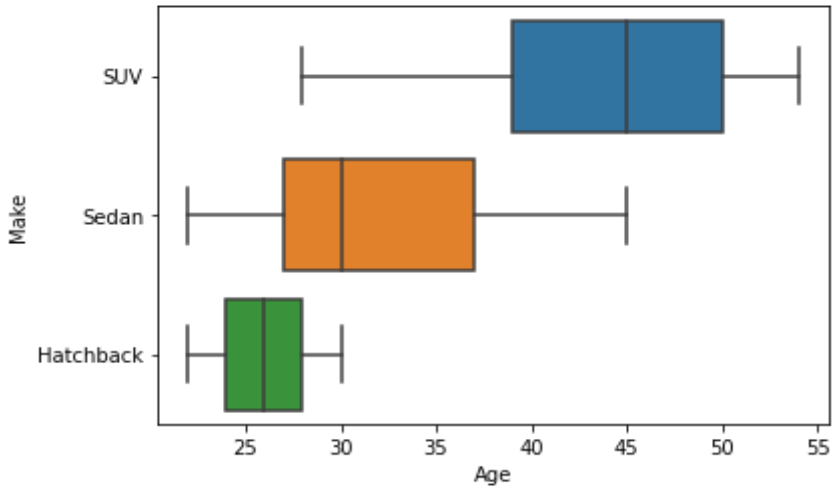


From comparing the boxplots of both business owners and salaried persons in terms of Price of Automobiles, we can compare the preference of range of price both the types of customers are willing to pay for the product. From the given diagram, we can infer from both the boxplot that the data is skewed to the right side, where the business data is more skewed than salaried. The majority of Business owners would like to own automobiles that are in between price range $20,000 to $68,000. Whereas, the majority of salaried people prefer to own

automobiles that are in between price range $18,000 and $70,000. Thus, we can conclude that salaried individuals are more likely to buy higher end automobiles than business owners.

---

f) MAKE & AGE

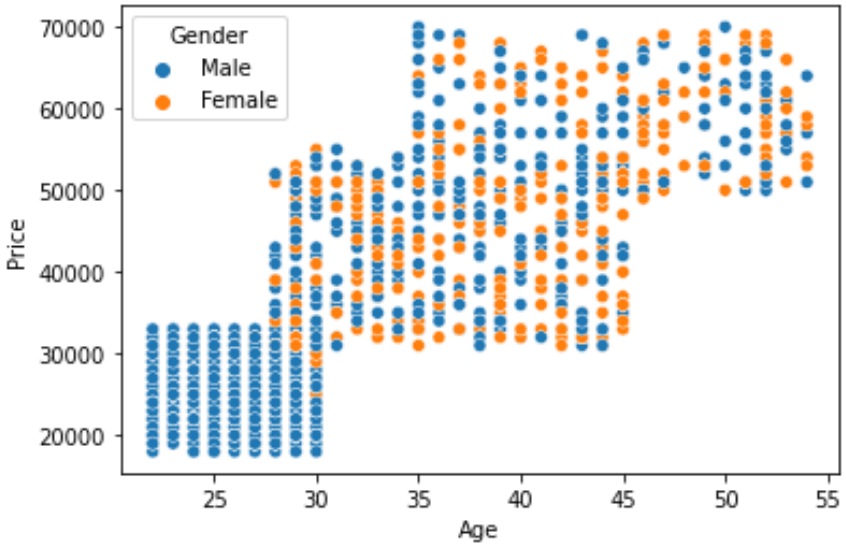


From the given diagram, we can see that the Hatchback model automobiles are most preferred among younger consumers from the age of 22 to 30 years and have the shortest range. Both SUV and Sedan have a wider range with SUV falling more into older consumers. Sedan is most preferred among 22 to 45 year olds. Whereas, SUV is most preferred amongst 27 to 54 year olds. Thus, we can conclude that the younger consumers are more likely to buy a Hatchback model automobile, middle age consumers are more likely to buy a Sedan model automobile, and consumers above 30 years of age are more likely to buy SUV model automobiles.

---

g) AGE, PRICE & GENDER

cont.



The above given scatterplot visualises the preferred price range of consumers on the basis of their age and gender. We are able to infer that most male consumers are concentrated in between the price range of $20,000 and $35,000. The distribution of male consumers tend to get more sparse as we move along the age. On the other hand, there are no female consumers in the cheaper price range with a minimum price range of $28,000. Thus, we can see that most women who buy cars are older compared to male demographic and their preferred price range is also higher compared to male consumers. Thus, we can conclude by saying that young male consumers (minimum age of 22) are more likely to buy cars at a cheaper price range than older male consumers. Female consumers are usually older than male consumers (minimum age of 27) and are more likely to buy medium to expensive cars than male consumers.

---

h) ALL NUMERICAL VARIABLE

cont.

The above given heatmap is the visualisation of correlation between all the numerical variables. There is a correlation of 0.62 between salary and age variable. Thus, showing a positive relationship between those two variables. There is a strong correlation between both Partner's salary and Total Salary(0.81), and Salary and Total Salary (0.64). This is because Total salary is the sum of Partner's salary and Salary. We can also see that there is a strong positive correlation of 0.80 between Age and Price. Thus, we can conclude that along with age, the price of products purchased also increases.

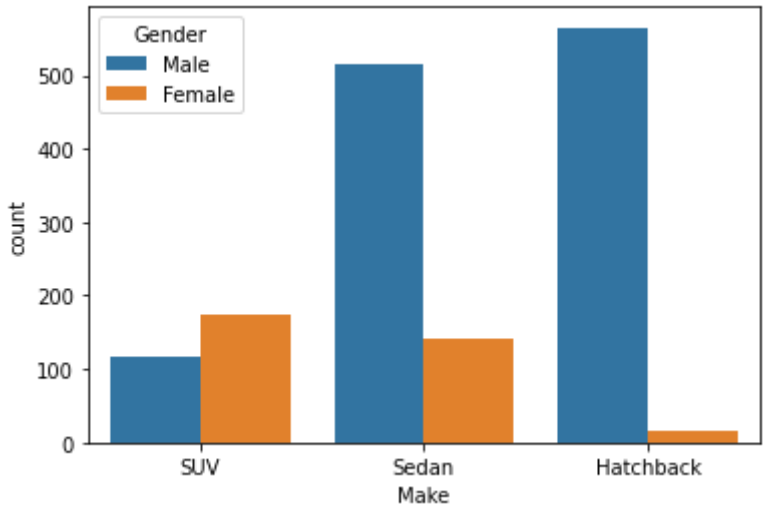
E. Employees working on the existing marketing campaign have made the following remarks. Based on the data and your analysis state whether you agree or disagree with their observations. Justify your answer Based on the data available.

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

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

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.

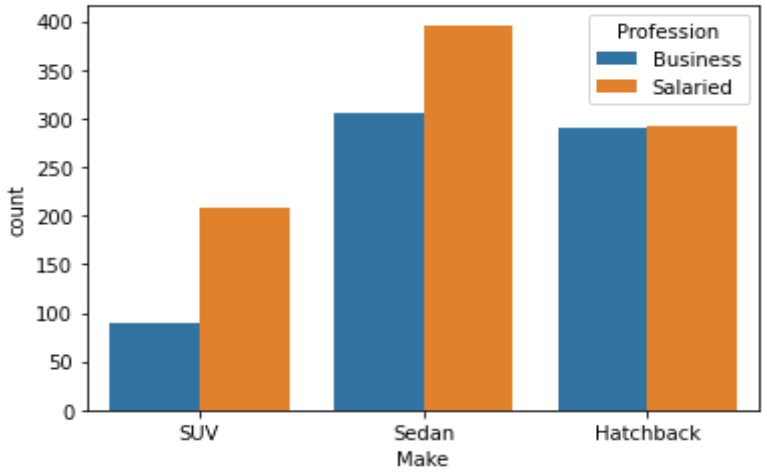
E1) Steve Roger says “Men prefer SUV by a large margin, compared to the women”



From the given diagram, we can see that the blue bar (male) is lower than the orange bar (female) for the visualisation of distribution under SUV purchasers. This shows that female consumers prefer SUV model cars more than male consumers. Therefore, I disagree with Steve Rogers.

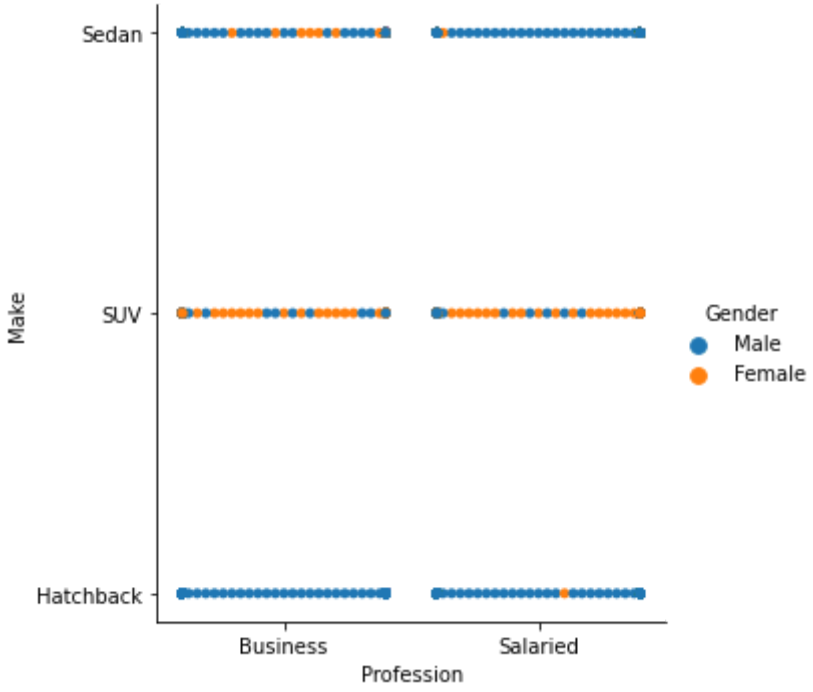
E2) Ned Stark believes that a salaried person is more likely to buy a Sedan.

cont.



In the given bar graph, Salaried persons have been picturised using orange bars and business owners are being picturised using blue bars. From the diagram, we can see that most salaried persons prefer Sedan model cars. Conclusively, we can say that salaried people are more likely to buy Sedan model cars. Therefore, I agree with Ned Stark.

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.



From the above given swarm plot, we can see the make preferences of the consumer base on the basis of their gender and Profession. With orange dots representing female consumers and blue dots representing male consumers, we can see that the SUV sale by Salaried people seems to be dominated by female consumers rather than male consumers. Whereas, Sedan sales by Salaried people seem to be dominated by male consumers. Thus, we can infer that

salaried male consumers prefer Sedan and Hatchback over SUVs. Therefore, I disagree with Sheldon Cooper.

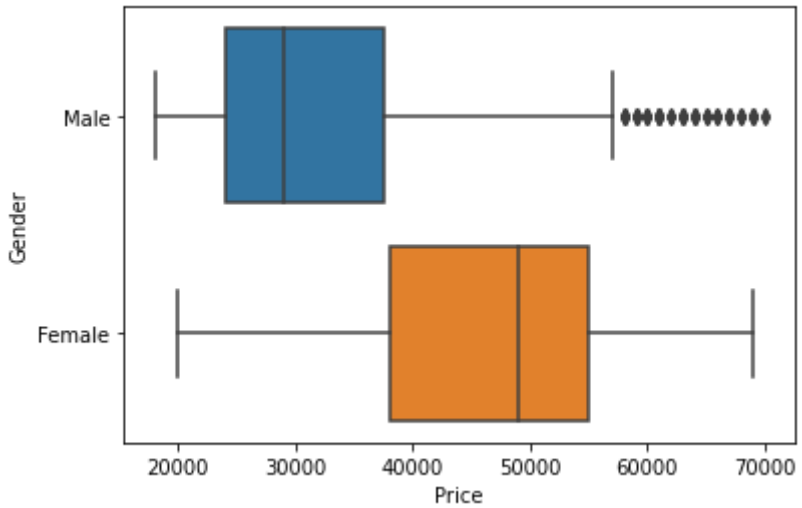
F. From the given data, comment on the amount spent on purchasing automobiles across the following categories. Comment on how a Business can utilize the results from this exercise. Give justification along with presenting metrics/charts used for arriving at the conclusions.

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

F1) Gender

F2) Personal_loan

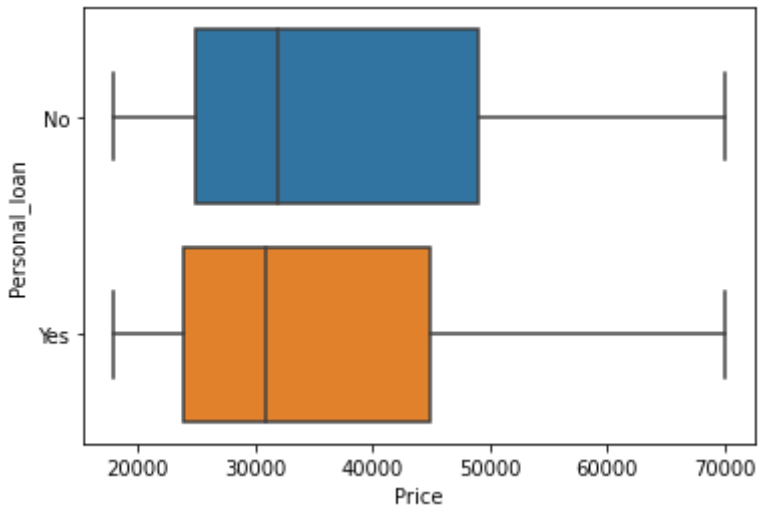
F1) Gender



From the given boxplot, we can see the price range in which the consumers are spending for the purchase of an automobile based on their gender. While looking at blue boxplot (male), we can see that the maximum a male consumer would spend on their car is approximately $58,000, with several outliers lying above the maximum range. The majority of male consumers spend in the average range of $18,000 to $58,000. On the other hand, Female consumers can spend a maximum of $69,000 and the majority of female consumers spend on average $20,000 to $69,000. Therefore, we can infer that Male consumers are drawn towards cheaper cars and if given discounts or EMI offers on expensive cars, they can be persuaded to buy higher end cars. As stated by *Rastogi (2021)*, 'Providing the Buy Now Pay Later option instantly for your customers can ensure quick transactions without any hold-ups, thus, greatly increasing the chances of high-value sales. The time-saving option of EMI at POS (point of sale) can boost the overall chances of transactions.' On the other hand, Although female

consumers are motivated to buy higher end vehicles, we should also take into consideration that there are far less female consumers than male consumers. Therefore, the company must make an effort to persuade women to buy their automobile by creating marketing campaigns targeting female consumers.

F2) Personal_loan

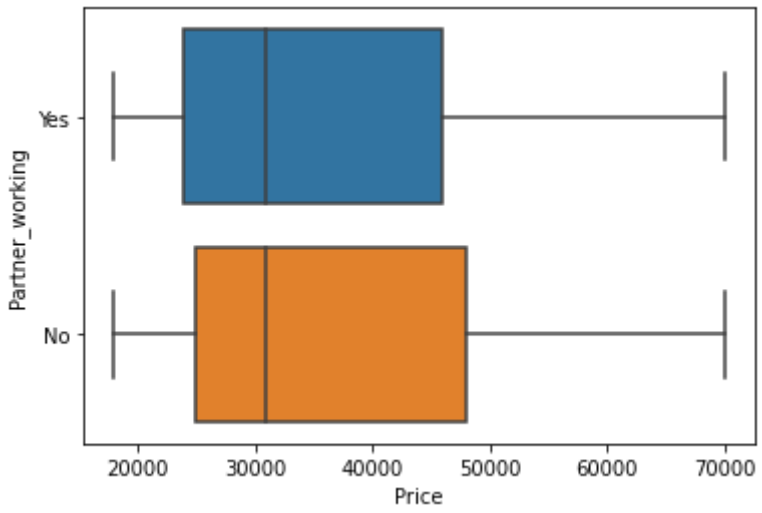


The above given boxplot visualises the price range a consumer has spent on automobiles based on whether they have taken a personal loan or not. From the diagram, we can see that the price range for both types of consumer is more or less similar, ranging from $18,000 to $70,000. However, the data for the consumers who have taken personal loans is slightly skewed more towards right than the data for the consumers who have not taken any loan. Therefore, we can infer that the median spending of a consumer who takes personal loan is slightly lesser than the median spending of a consumer who does not take personal loan. Thus, the company can make efforts to give out promotional offers to loan takers on expensive models to stimulate the rise in median spending of a consumer who takes personal loan.

---

cont.

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

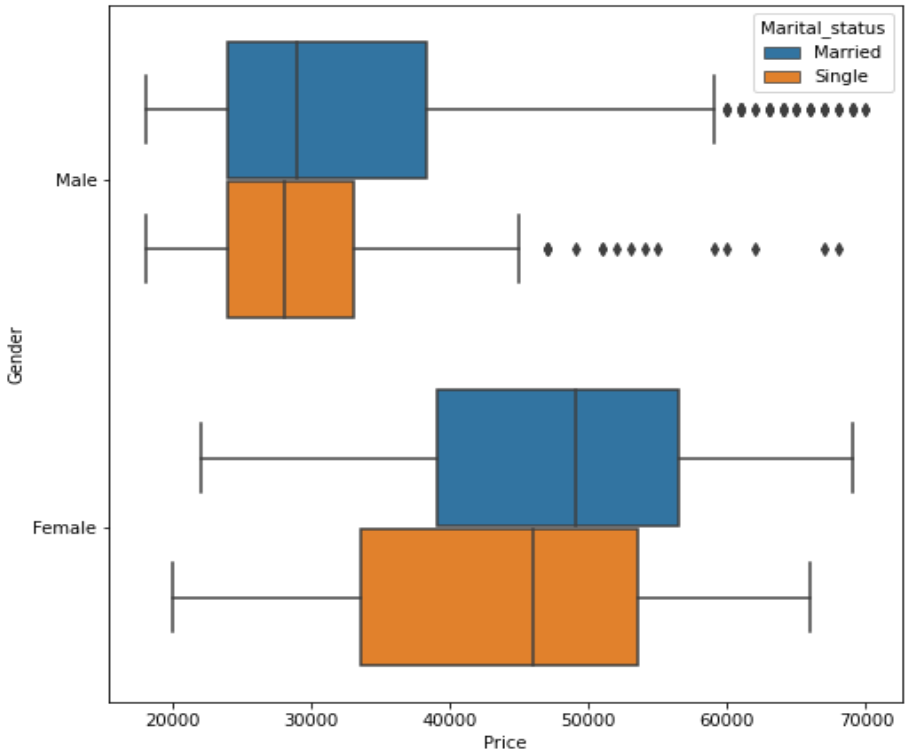


The above given diagram is the picturisation of the price preferences of a consumer based on whether they have a partner who is currently working or not. From the diagram, we can see that the preferred price range for both types of consumers is the same. They both prefer cars that range from $18,000 to $70,000. Similarly, the median spending of a consumer with a working partner is the same as the consumer with a non-working partner. Therefore, we can positively say that it is not certain that a consumer with a working partner will purchase a higher priced car than the other.

---

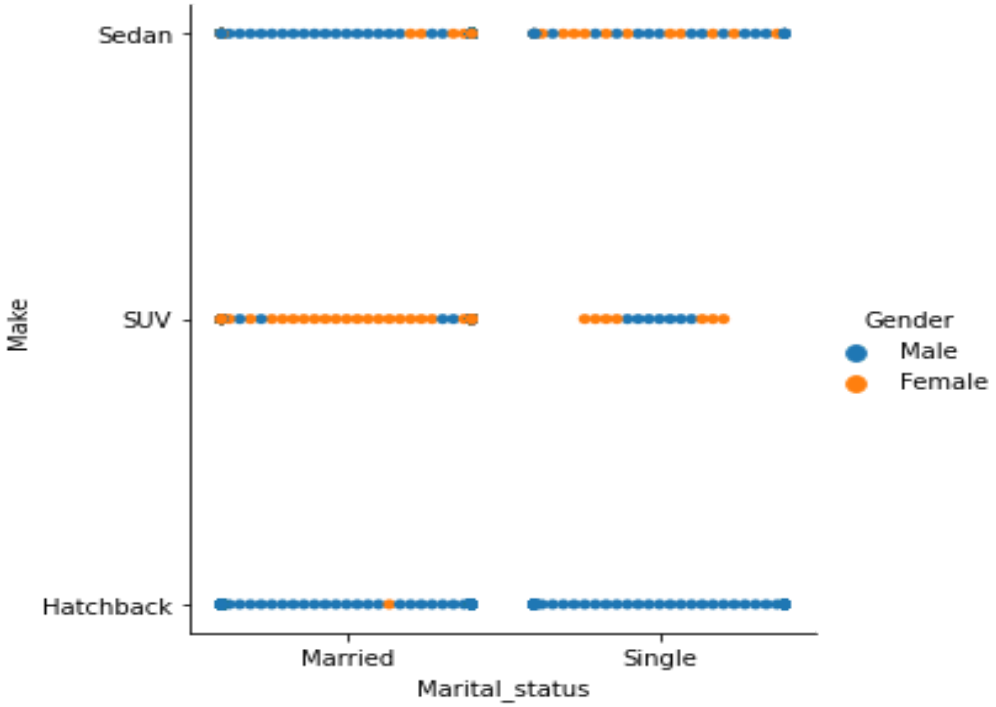
cont.

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.



Above box plot diagram shows the visualisation of the price preferences of consumers based on their gender and marital status. From the diagram we can see that men generally prefer cheaper cars than women and the price preference range is more expansive for women than men, with single men having the shortest price preference range in comparison. Married consumers are more willing to buy expensive cars than single individuals. This can be evidenced by the above given diagram where the median price range for married individuals is higher than the single individuals. Another noticeable trend is that the data from male consumers are generally skewed to the right. Whereas, the data from female consumers are generally skewed to the left.

cont.



The above given swarm plot is the visualisation of the model preferences of consumers based on their gender and marital status. From the given diagram, we can see that the Hatchback category predominantly consists of blue dots (male). Thus, we can say that Hatchback modelled cars are mostly preferred by men. Whereas, the SUV category mostly consists of Female consumers. This shows that most women prefer SUVs over other models. When talking about the Sedan category, Most married men prefer Sedan and both unmarried male and female consumers prefer Sedan cars.

II. 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.)

Problem 2 Question: (Analyze the dataset and list down the top 5 important variables, along with the business justifications. (10 Points) Data Dictionary - Link)

- **Transactor/Revolver**

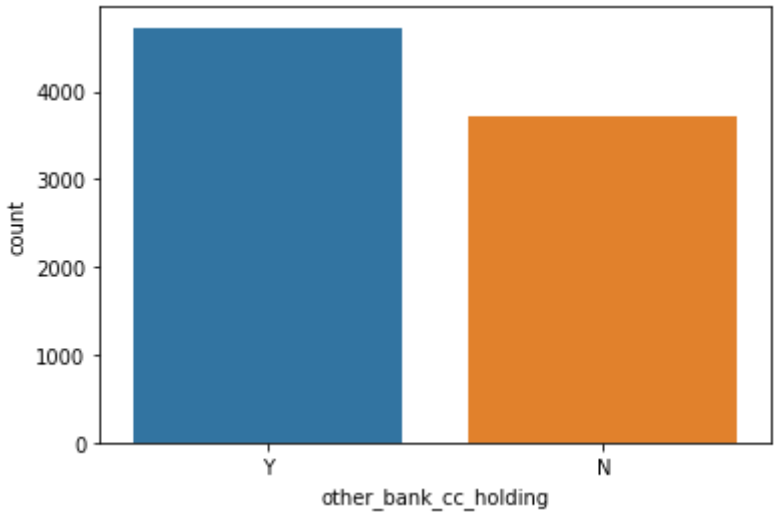
The first variable that is essential to the problem statement is the Transactor/Revolver variable. This variable essentially tells us if the consumer is a Transactor or an Revolver. Where, the transactor is a customer who pays off their balances in full and Revolver is a customer who carries balance from one month to next. This variable gives us essential details on how often a customer can default on their credit card payments or loan payments. Usually, a revolver customer is highly likely to default than transactors.

Transactor : 7115, Revolver : 1295

The above given diagram visualises the variable. We can see that there are significantly more transactors than revolvers. With 7115 customers being transactors and 1295 being revolvers. This shows that the company must make an effort to reduce the number of revolvers to avoid attrition rates.

- **Other bank cc holding**

Other bank cc holding is another essential variable to the problem statement, where it gives us an understanding if customers have several credit cards from other banks. This helps us to assess if we must stimulate more consumer activity.

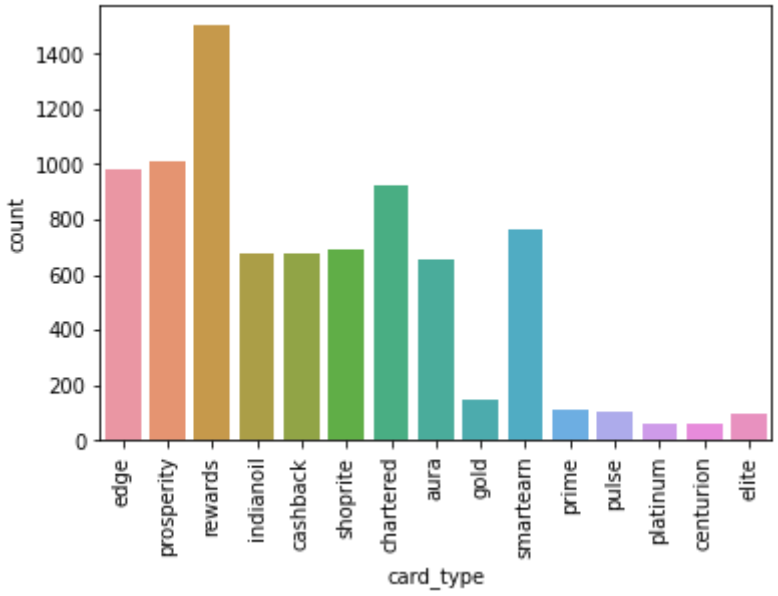


Y : 4728, N : 3720

The above given diagram shows the total number of customers with credit cards from other banks. We can see that the majority of the consumers have other credit cards. 4728 consumers have credit cards and 3720 consumers don't have credit cards from other banks. We can learn that the bank must make more efforts to review its service and credit policy to compete with other companies.

- **Card type**

Card type variable helps us analyse the distribution of the type of cards amongst the consumers. This variable can help us analyse the frequency of the type of cards and compare it with the consumer demographic to assess if they are fruitful to both consumers and the business. We can also learn the motivation of the consumers to get credit card from the bank.



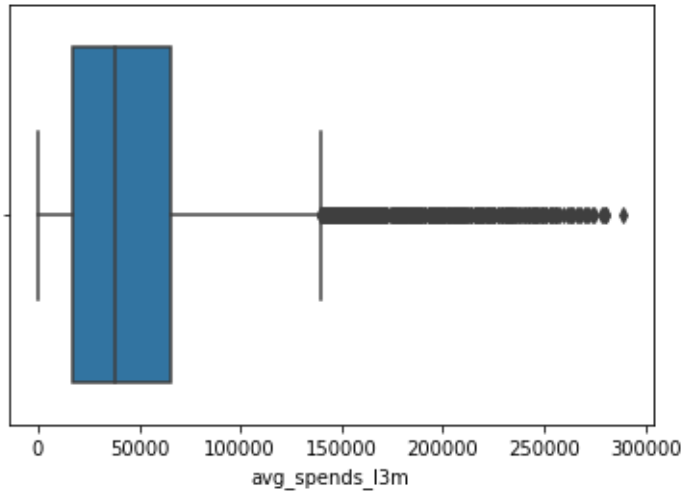
rewards	1502
prosperity	1007

edge	980
chartered	923
smartearn	765
shoprite	688
indianoil	680
cashback	676
aura	652
gold	145
prime	112
pulse	101
elite	96
centurion	62
platinum	59

From the above diagram about the type of credit cards, we can see that most consumers opt for cards with benefits like rewards cards or prosperity cards. These cards tend to provide high benefits like cashback or reward points for further purchases. Whereas, cards with added responsibility like centurion cards and platinum cards are not preferred by most customers.

- **Average Credit Card spend**

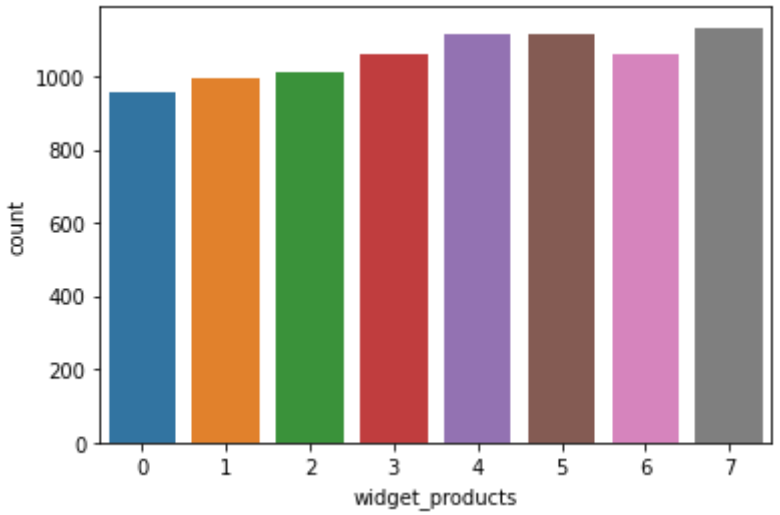
The Average credit card spending variable shows us how often a consumer utilises their card in everyday practices. Thus, we can understand the consumption activity of the consumer and ascertain if they are an active consumer or not.



From the given boxplot, we can see that most of the consumers spend between Re.1 to Rs. 1,50,000. We must also notice that there are some significant number of outliers above the maximum limit.

- **Widget product**

Widget product shows us the total number of convenience products the customer holds, like net banking active, mobile banking active, wallet active. I believe that this variable is more important than engagement products because Widget products show us how often consumers use the banking services and how often they use it. Whereas, engagement products show us the consumer's loan activity. Although they are both important, the Widget product shows us how many consumers are willing to use the bank's services on an everyday basis.



7	1132
6	1117
5	1115
4	1062
3	1060
2	1010
1	997
0	955

We can see that the distribution of data in each category is pretty similar. With an average consumer's usage of the widget products ranging from 0 to 7. Therefore, the bank must make efforts to increase the consumer interaction with the widget product to increase brand loyalty.

REFERENCES

Thakuriah, P. (2010). Car Ownership Among Young Adults - Generational and Period-Specific Perspective. *Journal of the Transportation Research Board*. 10.3141

Malik, Sakshi and Kaur, Muskan and Kapoor, Anuj Pal, Purchase Now and Pay Later: Consumer Preferences Towards No Cost EMI's in India (December 5, 2020). e-journal - First Pan IIT International Management Conference – 2018, Available at SSRN: https://ssrn.com/abstract=3743415 or http://dx.doi.org/10.2139/ssrn.3743415

Rastogi, A. (2021, October 22). *How Can EMI Payments Help You Maximize Your Sales In Festive Season*. Ezetap. Retrieved March 5, 2023, from https://corp.ezetap.com/blogs/emi-payments-maximize-sales-festive-season