**SMDM PROJECT**

**Business Report**

Pavithra Devi

DSBA  
Great Learning

Contents

Problem 1:

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: **Come up with insights to improve the marketing campaign.**

Problem 2:

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.)

Objective: **Framing an analytics problem:- list down the top 5 important variables, along with the business justifications.**

Problem:1

1. **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)**

Before we start analyzing the data, it is important to know the basic information of the data i.e., number of columns and rows, data types available, presence of null values or duplicity etc.

Following facts can be inferred through the data provided.

The data has 1581 rows and 14 columns.

There is no duplicity of the data provided.

Datatypes of the variables:

|  |  |  |
| --- | --- | --- |
| Object | Float | Int |
| Gender | Partner\_salary | Age |
| Profession |  | Salary |
| Marital\_status |  | Price |
| Education |  | No\_of Dependents |
| Personal\_loan |  | Total\_salary |
| House\_loan |  |  |
| Partner\_working |  |  |
| Make |  |  |

By looking at the data set it is evident that the source of data collection and the information is genuine. There are few null values under Gender and

Partner\_salary column, rest all the data looks good.

2. **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.**

Data Pre-Processing:

* By describing the data, we found out that the gender column has 4 unique values under it.

Reference: ['Male', 'Femal', 'Female', nan, 'Femle']

* Gender and Partner\_salary column has some missing and null values which might need imputation.

Post correction of the bad data under gender column, we found out that

there’s 53 null values.

Since, it’s a categorical variable we can impute it with the mode value instead of dropping the null values. As we might lose out valuable inputs of the other columns.

**76% of the data is of males. Imputing the null values with the same.**

Under Partner\_salary column, we have 106 null values.

As we know, Total\_salary= Salary+Partner\_salary. And, the latter columns has no missing values. We would be able to compute Partner\_salary value using the formula: Partner\_salary= Total\_salary-Salary.

To make sure the data isn’t bad, did a minor cross check if there’s any field with

a.)Marital\_status= Single, Partner\_working=Yes, Partner\_salary=numeric value.

2.)Marital\_status= Married, Partner\_working=No, Partner\_salary=Numeric value

3.)Marital\_status= Married, Partner\_working=Yes, Partner\_salary=0.

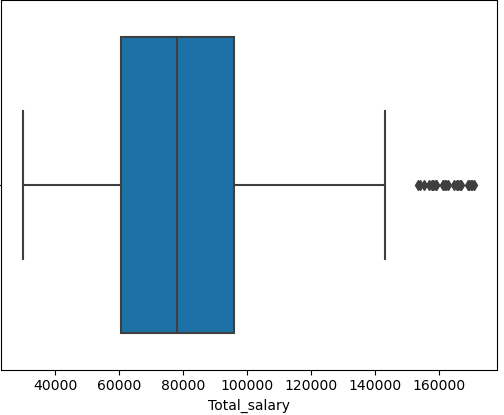
As these fields are interlinked. Bad data and anomalies will ruin our interpretation. Luckily, we do not see any such records.

**Thus, Partner\_salary column now has no null values.**

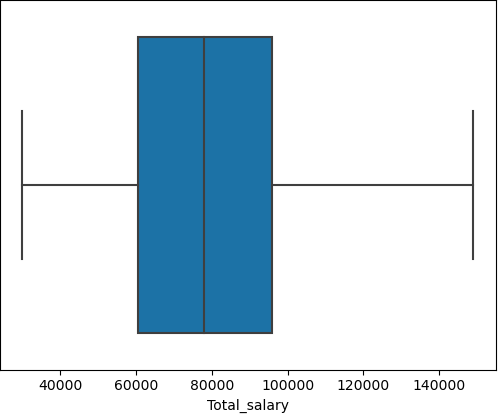
Treating Outliers:

* Outliers are important because they can have a large influence on statistics derived from the dataset.

Here, Total\_salary column has outliers.



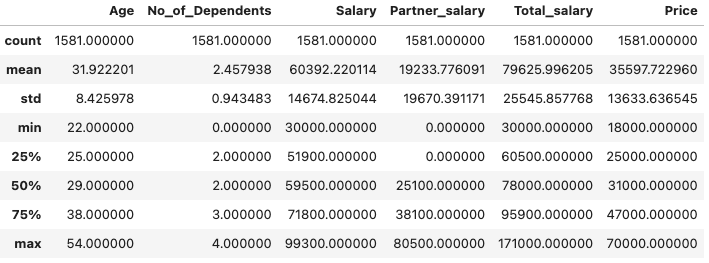
We can see that the there are extreme values in the boxplot which goes beyond the maximum line.



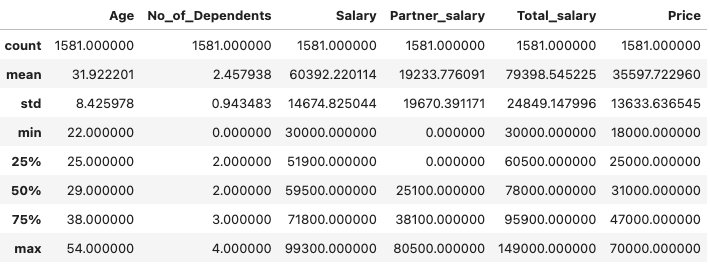
Treatment of outliers is done using the IQR method.

**Summary:**

Data Before Outlier Treatment

****

Data After Outlier Treatment

****

Focusing on the Total\_salary column of the data before and after outlier treatment :

The minimum of Total\_salary remains the same whereas the maximum value has come down to 149,000 from 171,000.

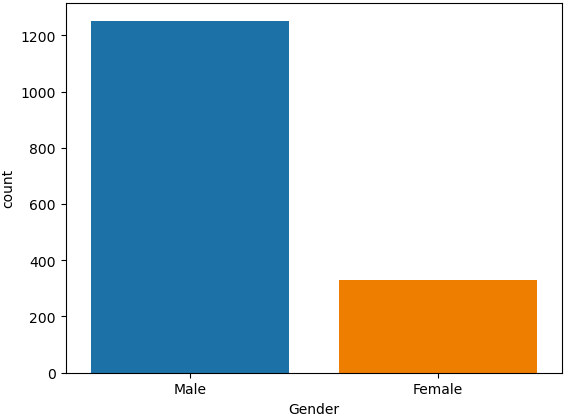
3. **Explore all the features of the data separately by using appropriate visualizations and draw insights that can be utilized by the business.**

Initially let us visualize each and every variable individually and understand their behaviour in the data to derive insights for the business.

Univariate Analysis:

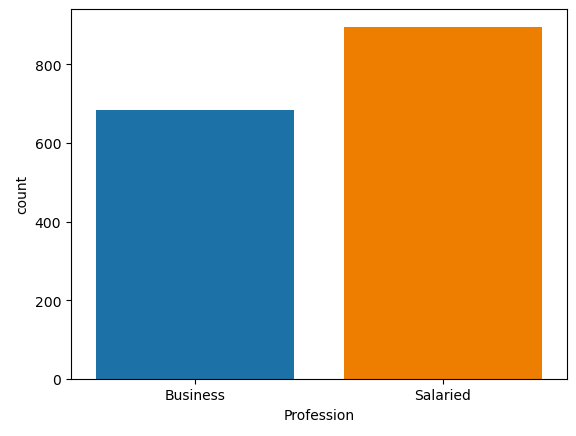
* Gender:

Given the clear indication from the graph that approximately 70% of purchases are made by males, it would be prudent to prioritize the male demographic as our primary target audience and direct our marketing strategies towards them.



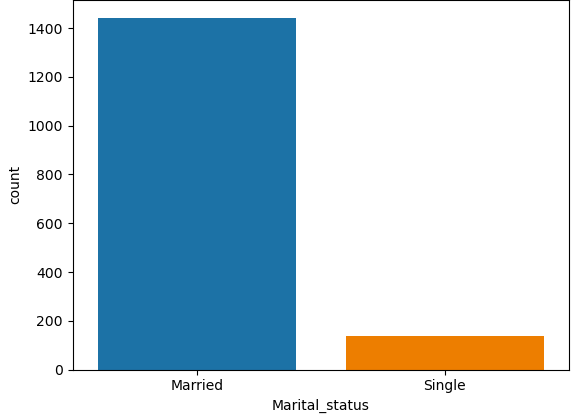
* Profession:

Salaried professionals account for over 55% of purchases, surpassing the proportion of purchases made by business professionals.



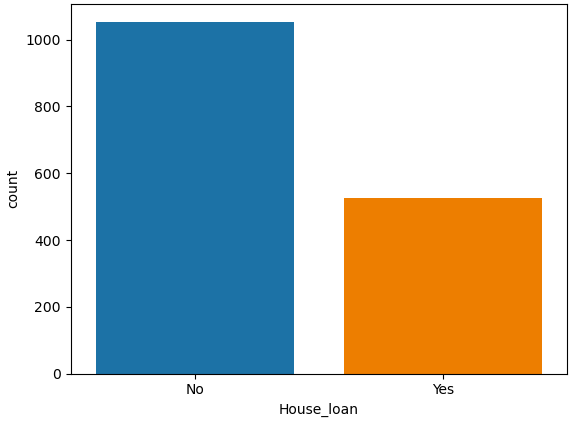
* Marital Status:

Considering the greater proportion of married individuals in comparison to singles, it can be deduced that married people exhibit a higher propensity for car purchases. As a result, focusing marketing efforts on married individuals would be advantageous to drive increased sales and profitability.



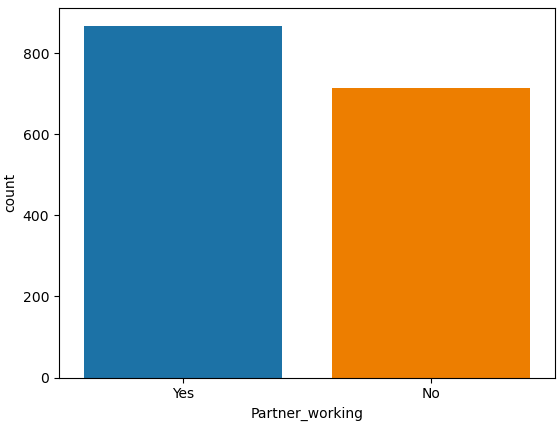
* House Loan:

The plot suggests that individuals who do not have house loans are more likely to make car purchases compared to those with existing house loans. As a result, it is recommended that businesses focus their marketing strategies towards the target audience without house loans to maximize sales and drive growth.



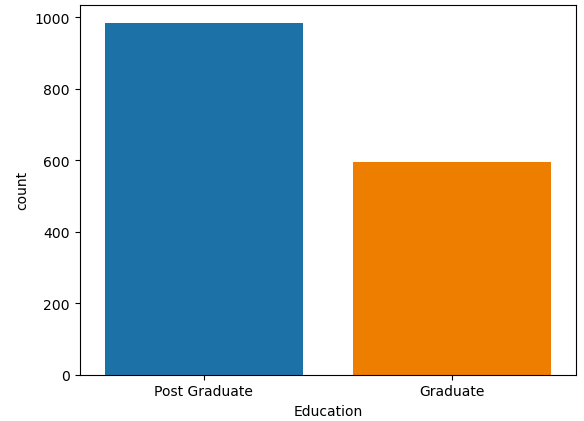
* Partner Working Status:

The data suggests a higher likelihood of car purchases among individuals who have a working partner. This indicates that targeting marketing efforts towards individuals with a working partner may lead to increased sales in the automotive industry.



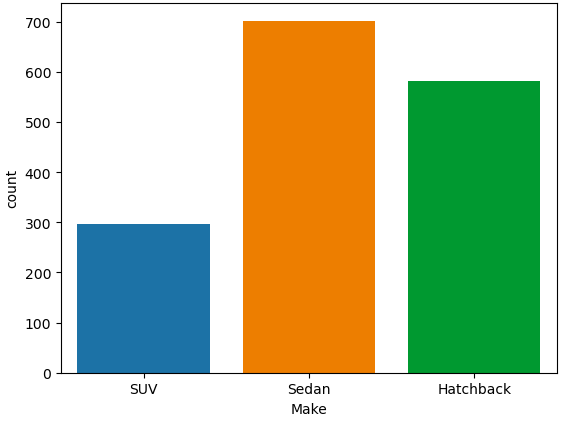
* Education:

Compared to individuals with a graduate degree, there is a higher inclination among postgraduate individuals to make car purchases. Thus, directing marketing strategies towards the postgraduate segment may be more effective in driving sales and capturing a larger market share in the automotive industry.



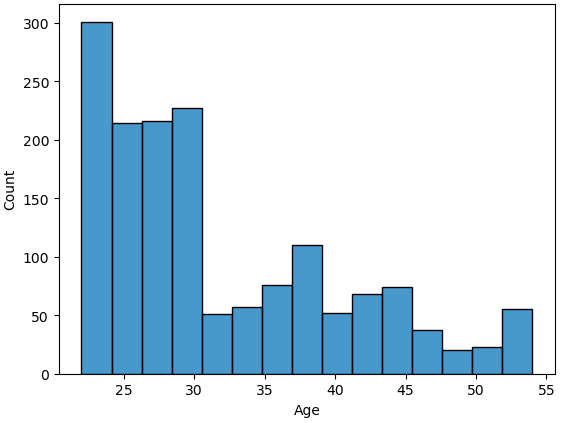
* Make:

There is a higher tendency among consumers to purchase sedans or hatchbacks compared to SUVs. Therefore, focusing marketing efforts on sedans and hatchbacks may be more fruitful in terms of increasing sales and meeting customer preferences in the automotive market.



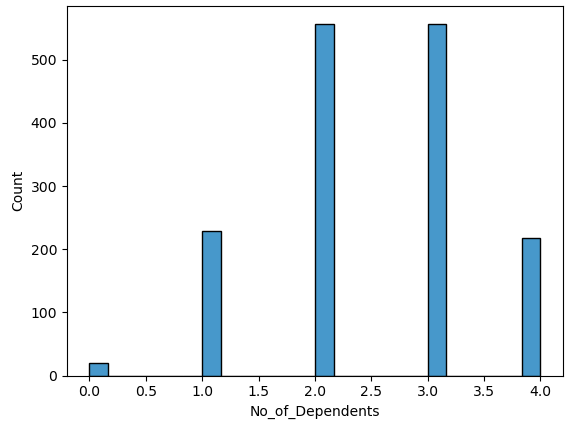
* Age:

The data reveals a higher count of car purchases within the age group of 22-30, accounting for approximately 50% of the dataset. Furthermore, approximately 75% of the data indicates that customers purchasing cars are within the age group of 40 and below. These insights suggest that targeting marketing efforts towards younger and middle-aged demographics would be advantageous for maximizing sales and effectively reaching the majority of the customer base.



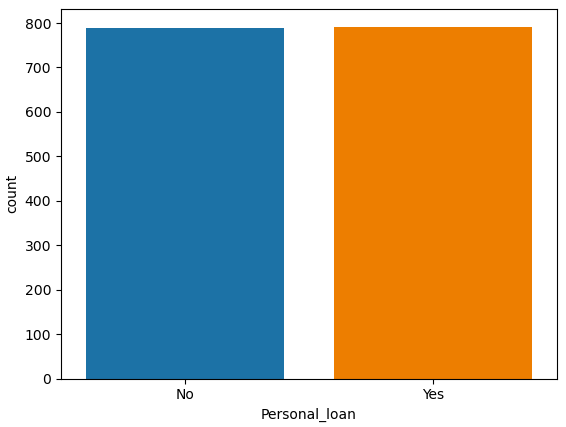
* Number of Dependents:

The graph exhibits a multimodal distribution, with the highest peak representing households with 2 or 3 dependents, and the second highest peak representing households with 1 or 4 dependents. This indicates that the majority of individuals purchasing cars have either 2 or 3 dependents, followed closely by those with either 1 or 4 dependents. Understanding these patterns can assist in tailoring marketing strategies to cater to the needs and preferences of customers in different dependent groupings, thereby increasing sales and customer satisfaction.



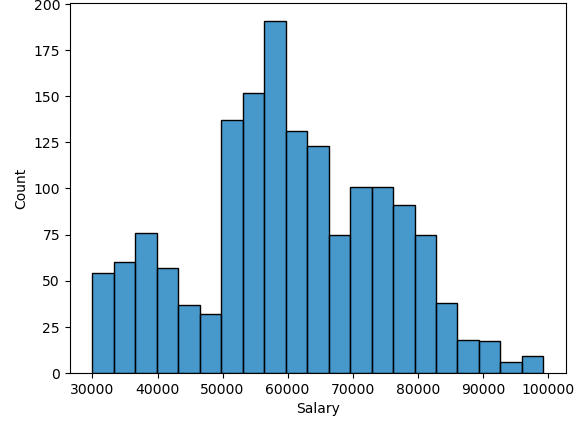
* Personal Loan:

The data suggests that the presence of a personal loan does not appear to have a substantial impact on the intention to purchase a car. Therefore, for the purpose of developing your marketing strategy, it may be reasonable to exclude the personal loan variable as a standalone factor.



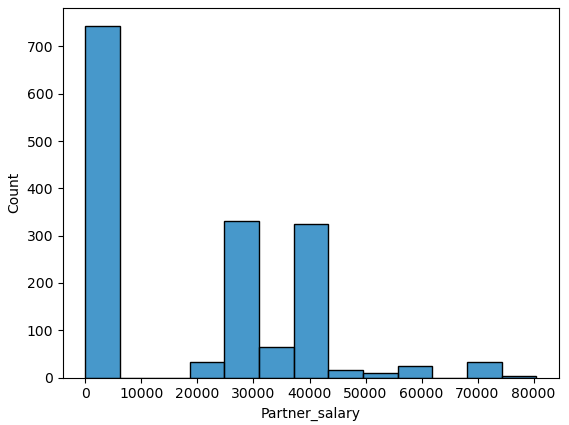
* Salary:

Individuals within the salary range of 50k-80k generally exhibit a higher purchasing power compared to those with lower incomes. This finding suggests that targeting this specific salary range in your marketing strategy could potentially yield positive results in terms of increasing sales.



* Partner Salary:

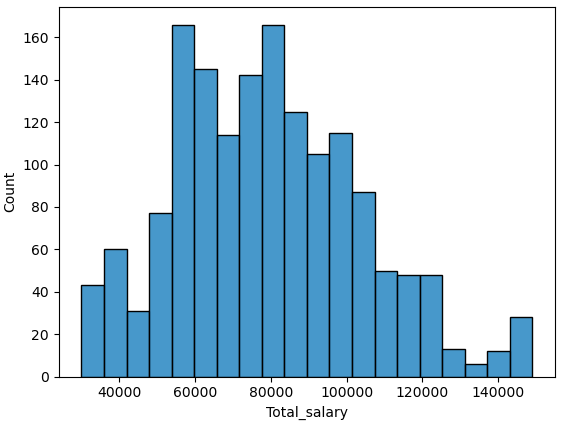
The data suggests that a significant number of individuals with partner salaries ranging from 0-5k are willing to purchase a car. This finding indicates an interesting opportunity to target this specific segment in your marketing strategy, as they show a willingness to make a car purchase despite their lower partner salary.



* Total Salary:

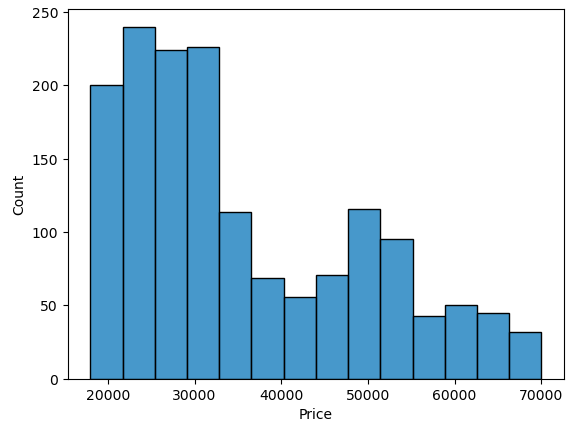
The close proximity of the mean and median suggests that the salary distribution is relatively symmetrical or has a less skewed distribution. This indicates that there is a balance between higher and lower salaries within the dataset.

Additionally, the presence of a maximum value of $149,000 indicates that there are individuals with higher salaries in the dataset. It's important to consider this maximum value in conjunction with the mean and median to understand the range and dispersion of salaries.



* Price:

A significant majority, comprising approximately 75% of customers, tend to purchase cars within the price range of 18k to 47k. Furthermore, half of the customers (50%) opt for cars falling below the 32k price range slab, indicating a strong preference for more affordable options.



4. **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.**

Bivariate Analysis:

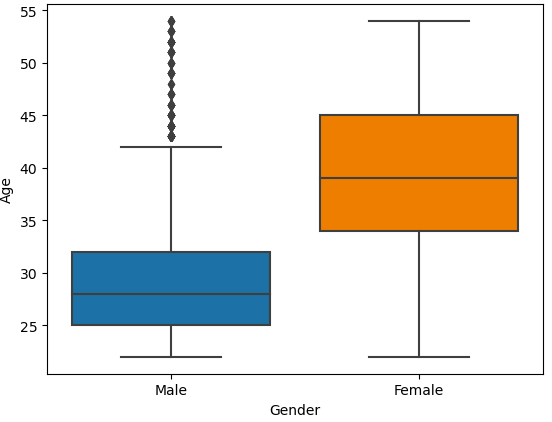
It is very important to know the customers to understand their purchase behaviour and various factors associated….

Let us begin by analysing how 2 variables are related.

* Age and Gender:

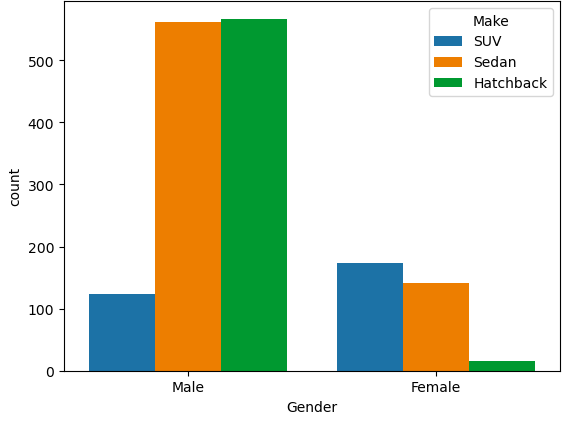
The box plots reveal that the age distribution for males is relatively younger, with a wider range and the presence of outliers up to 54 years.

In contrast, 50% women under the age category 35-44years are likely to buy cars.



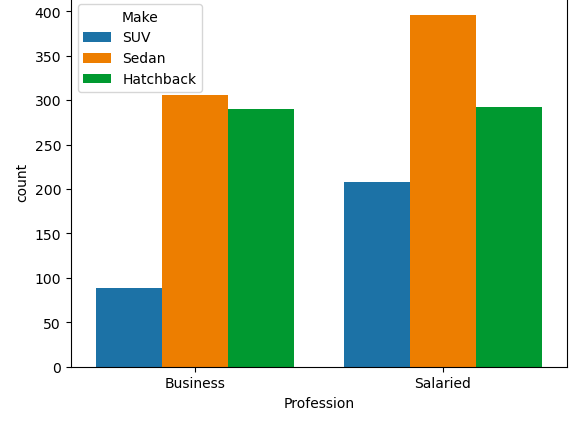
* Gender and Make:

There’s a clear disparity in the choice between male and female customers. More than 70% of the males prefer a Sedan or a Hatchback over SUVs. Whereas more than 70% females prefer SUV or a Sedan over a Hatchback. Clearly indicating there’s high demand of Sedan and Hatchback among males and SUV and Sedan among females.



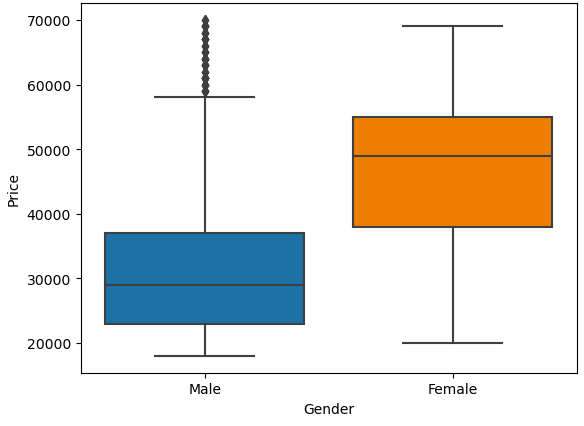
* Profession and Make:

Salaried individuals prefer a Sedan over any other models.



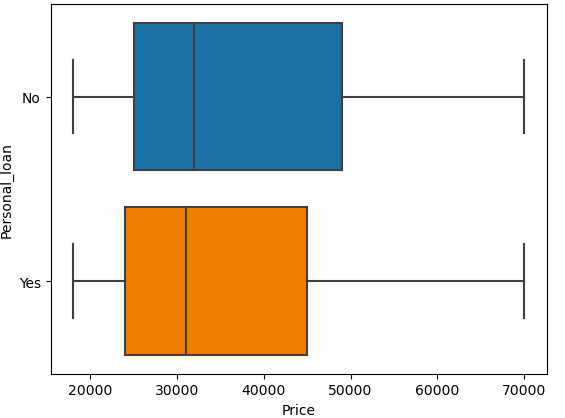
* Gender and Price:

Females generally have the tendency to purchase cars of a higher value comparatively.



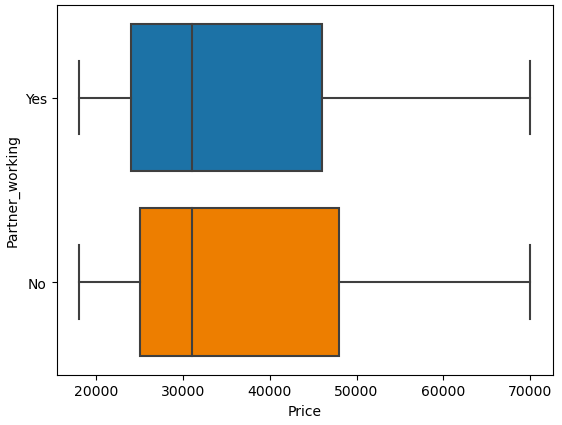
* Personal Loan and Price:

The presence of a personal loan does not have an impact on the kind/. value of the car that is being purchased. The univariate analysis on Personal Loan and Price confirms this.



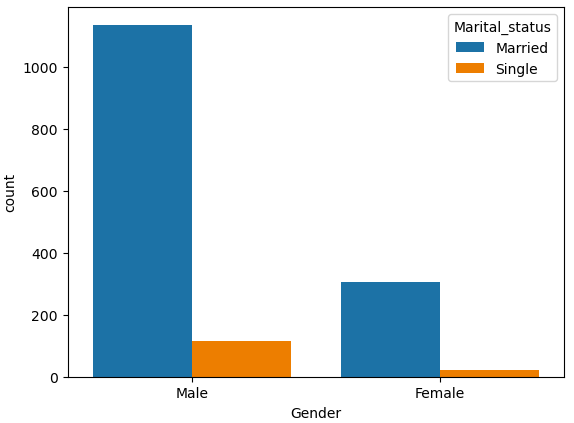
* Partner Working Status and Price:

Irrespective of working partner or not, the median of the both the boxplots is under the same range, showing there isn’t much difference in the behaviour.



* Gender and Marital Status:

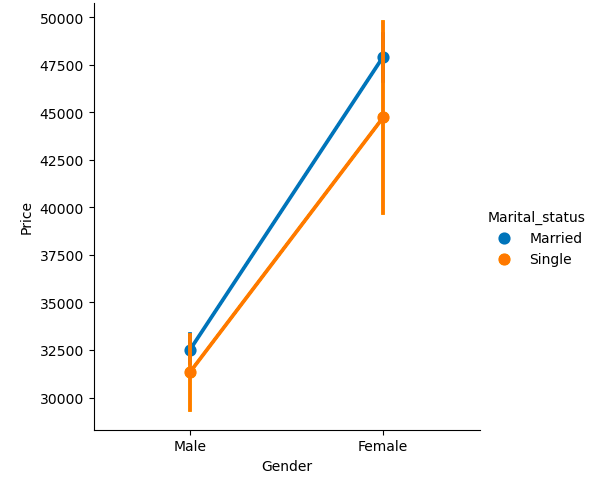
Observations tells us that married male and female have the tendency to purchase cars overall.



Multivariate Analysis:

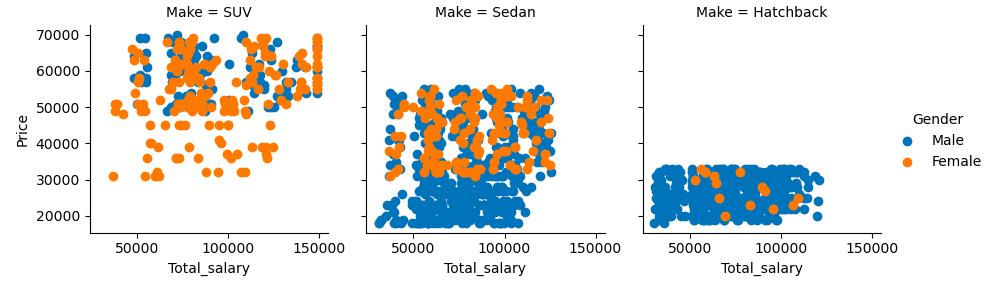
* Gender, Marital Status, Price:

In the multivariate analysis considering Gender, Marital Status, and Price, it is evident that both married males and females tend to purchase cars with higher values. However, females emerge as the primary purchasers of high-value cars.

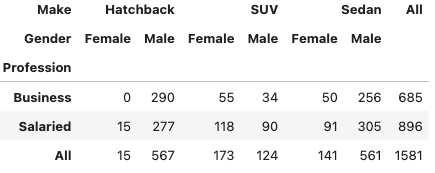


* Make, Total Salary, Price, Gender:

SUVs are predominantly purchased by females and are associated with a higher overall salary, often reaching up to 150k. However, sedans and hatchbacks tend to have a salary limit that typically does not exceed 130k. So, people having salary below S125000 approx. mostly prefer Sedan and Hatchback.

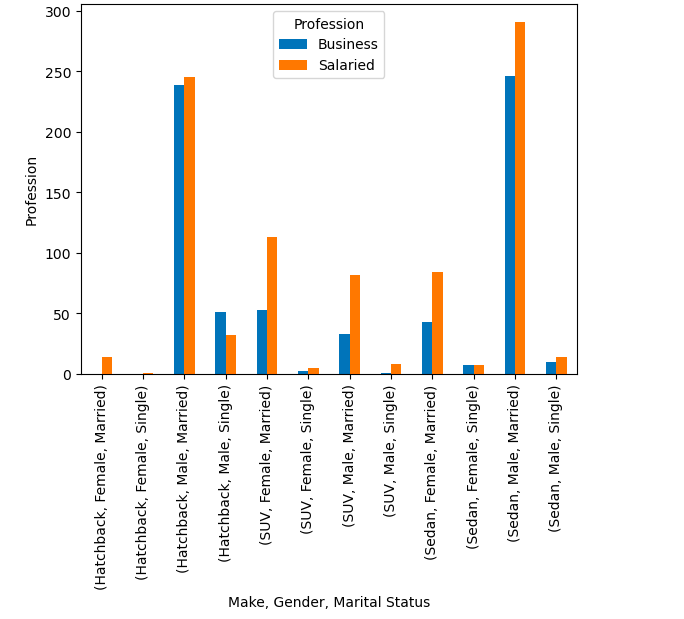


* Gender, Profession, Make:



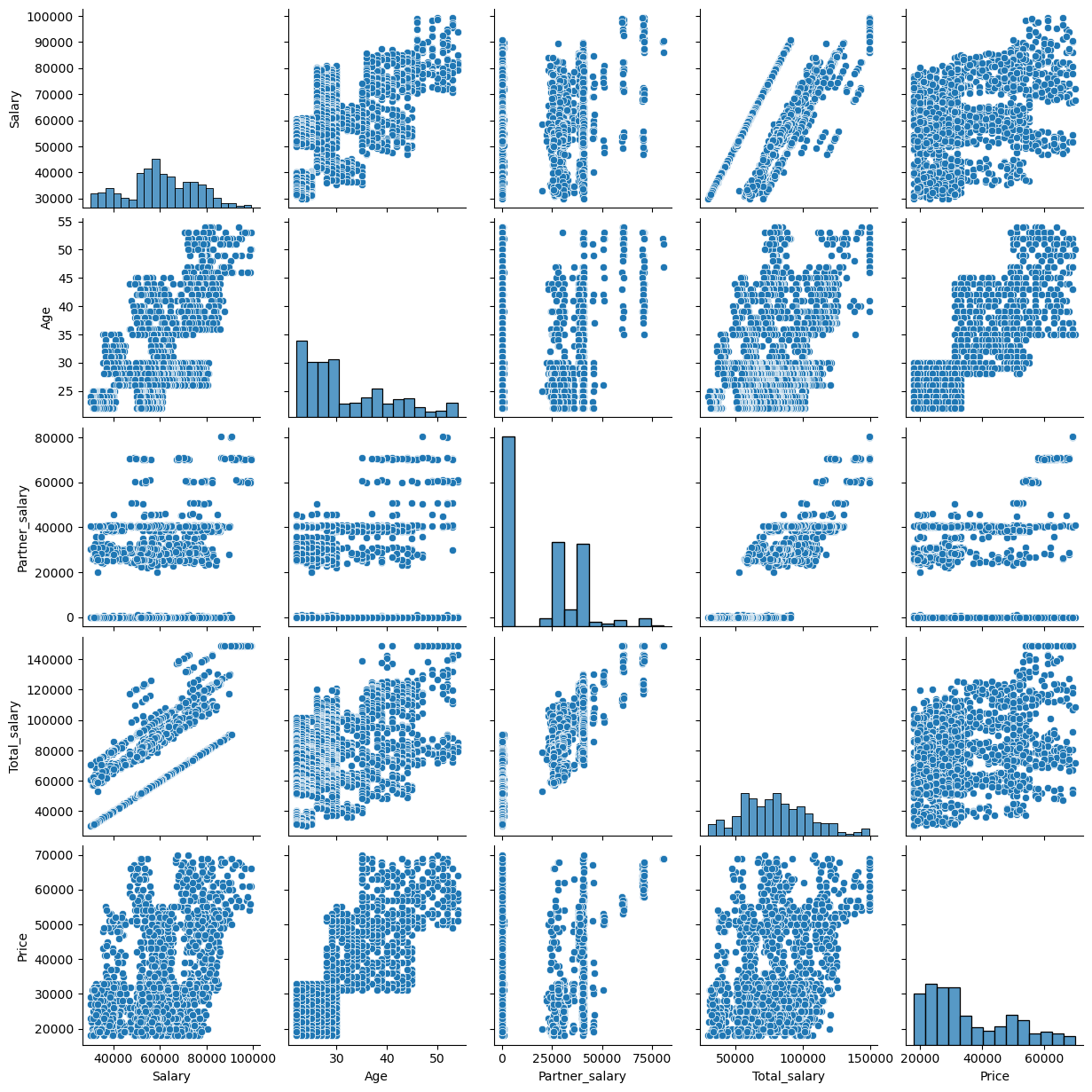
Observations shows that business Females don't prefer Hatchback at all. They prefer either a SUV or a Sedan. Salaried males prefer either a Hatchback or a Sedan. Hatchbacks are preferred Business males.

To bifurcate furthermore, let us include the marital status of the individuals in the data presented.



Based on the data, it can be inferred that married males, whether employed in a salaried or business profession, tend to prioritize Sedans as their first choice. Hatchback models follow as their second preference. Conversely, for married females employed in business or salaried positions, SUVs are their primary preference, followed by Sedans.

Let's explore the correlation between the continuous variables in the dataset to gain a better understanding of their relationship.



By analyzing the pairplot, it becomes evident that the following pairs of variables exhibit a linear relationship:

1. Age and Salary

2. Age and Total Salary

3. Age and Price

4. Salary and Total Salary

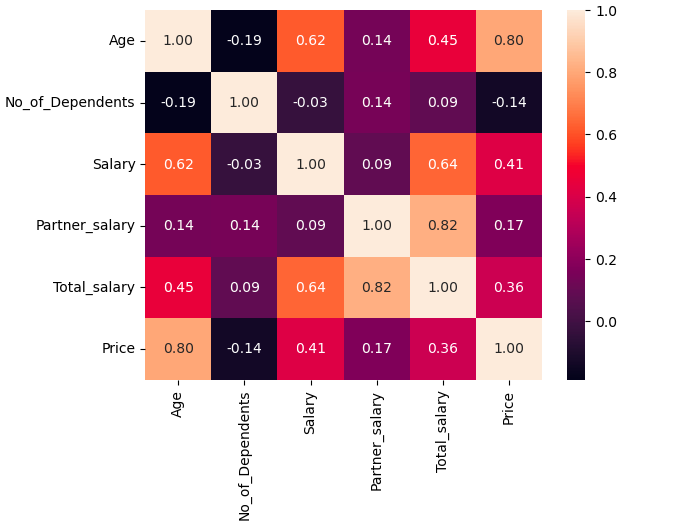
5. Price and Salary

6. Partner Salary and Total Salary

7. Total Salary and Price

These pairs of variables with lighter color, indicating that changes in one variable are associated with proportional changes in the other variable.

Upon analysis, it is observed that the number of dependents does not exhibit any significant correlation with the other variables in the dataset.

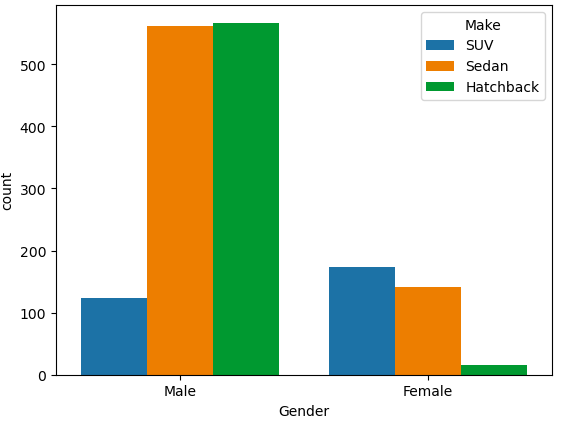


Summary

1. More than 75% of the cars are purchased by males. It would be prudent to prioritize the male demographic as our primary target audience and direct our marketing strategies towards them.
2. However, females emerge as the primary purchasers of high-value cars.
3. Observations tells us that married male and female have the tendency to purchase cars overall.
4. Females don't prefer Hatchback at all.
5. The majority of individuals purchasing cars have either 2 or 3 dependents.
6. Compared to individuals with a graduate degree, there is a higher inclination among postgraduate individuals to make car purchases.
7. Salaried professionals account for over 55% of purchases, surpassing the proportion of purchases made by business professionals.
8. Individuals who do not have house loans are more likely to make car purchases compared to those with existing house loans.
9. The data reveals a higher count of car purchases within the age group of 22-30, accounting for approximately 50% of the dataset. Furthermore, approximately 75% of the data indicates that customers purchasing cars are within the age group of 40 and below.
10. Individuals within the salary range of 50k-80k generally exhibit a higher purchasing power compared to those with lower incomes.
11. A significant majority, comprising approximately 75% of customers, tend to purchase cars within the price range of $18k to $47k. Furthermore, half of the customers (50%) opt for cars falling below the $32k price range slab, indicating a strong preference for more affordable options.
12. Individuals having salary below S125000 approx. mostly prefer Sedan and Hatchback.

5. **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.**

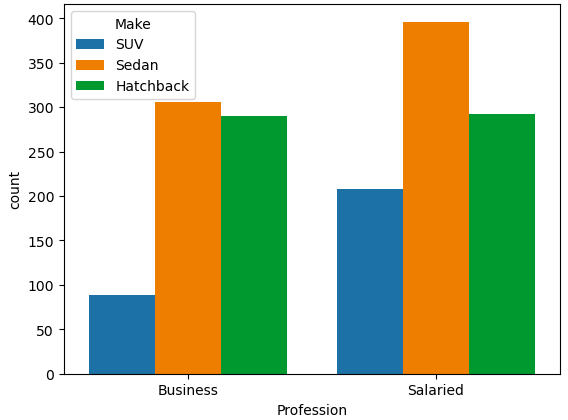
E1: **Steve Roger says “Men prefer SUV by a large margin, compared to the women”.**



The data clearly indicates there’s a high demand for Sedan and Hatchback among males and SUV and Sedan among females. SUV is the least preferred model among males. But highly preferred by females.

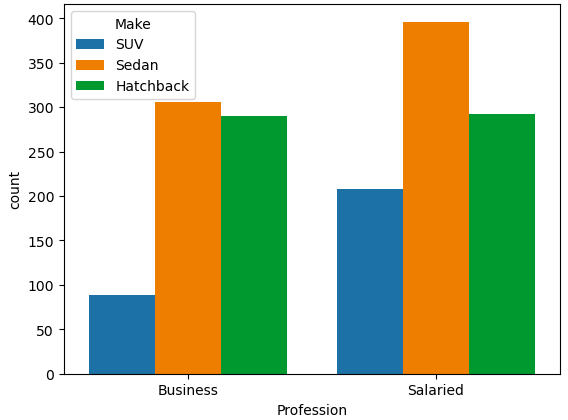
So, the statement made by Steve Roger is wrong.

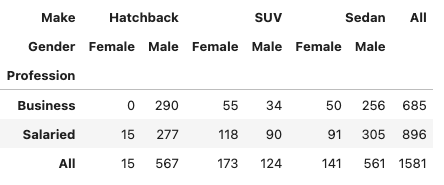
E2: **Ned Stark believes that a salaried person is more likely to buy a Sedan.**



Salaried professionals exhibit a distinctively high preference for Sedan cars. So, the statement made by Ned Stark is correct.

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.**



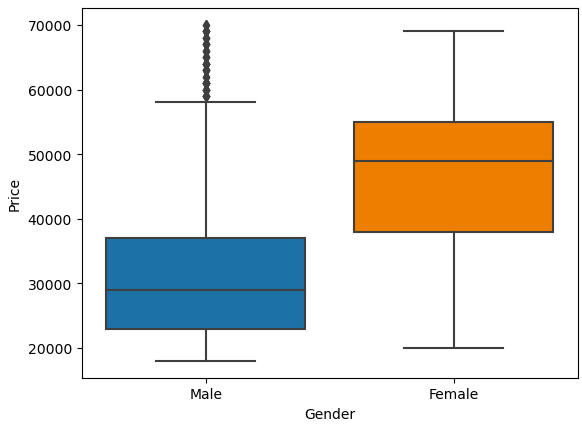


Among 896 salaried individuals, 672 are males and about 305(45%) males prefer Sedans over any other cars. The second preference is relatively higher for hatchback. The least is for SUV’s.

So, the statement made by Sheldon Cooper is wrong.

6. **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. \*\*\*F1) Gender \*\*\*F2) Personal\_loan**

F1: Price and Gender:

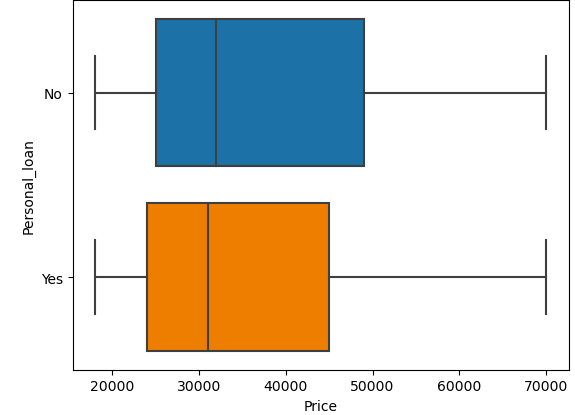


Irrespective of the fact that the higher purchases are done by males. The higher value cars are purchased by females. The average amount spent by a female is $47,705 whereas for a male it is about $32,416. The revenue generated from each source i.e., Male: $40,585,000, Female: $15,695,000. Females owning just 25% of the overall purchases are able to contribute 39% of the male purchase revenue. Which clearly shows that there’s high potential of revenue that can be unlocked from female category through effective marketing strategies.

Now, that we know females prefer sedans and SUVs over hatchback models, we can have a special focus on the same.

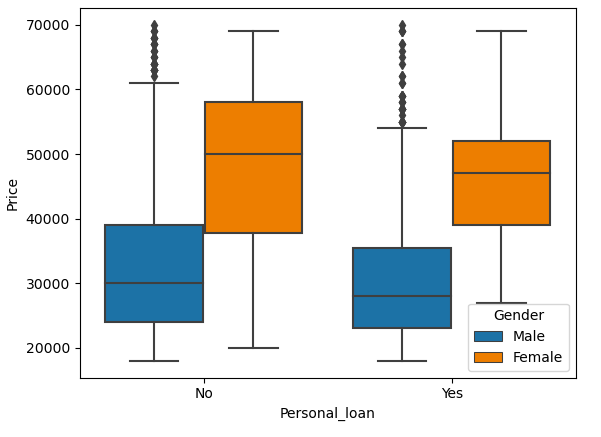
F2: Price and Personal Loan:

In the bivariate analysis of Price and Personal loan variable, it was found out that customers irrespective of having the personal loan had similar purchasing behavior as the Non personal loan holders. Here, it confirms the pattern. The median of both the plots are almost the same i.e., YES: $31,000, NO: $32,000. The mean is also almost similar i.e., YES: $36,742, NO: $34,457. The total sum of purchases done by a personal loan holder is $27290000 and No-Personal loan holder is $28990000.



The only inference that can be drawn from the above data is that people with a personal loan tend to slightly purchase cars of a lower value than the non-personal loan holders.

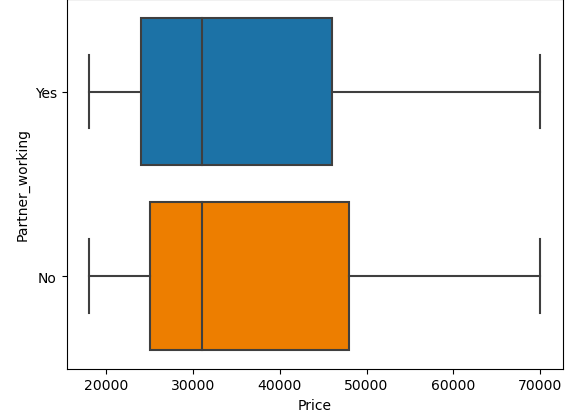
The below graph represents the detail evidently…



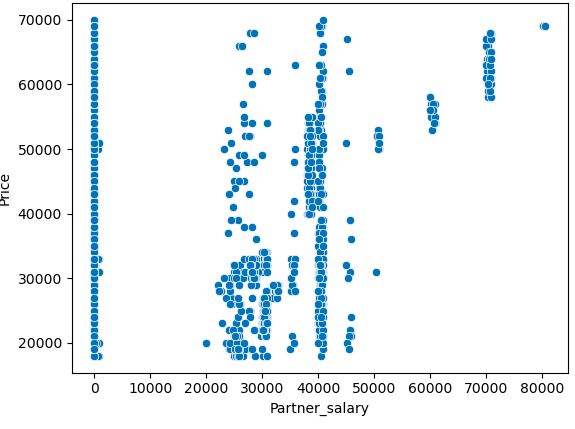
EMI’s, low interest etc can be added to the promotions or ads to help the customers with the money burden.

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

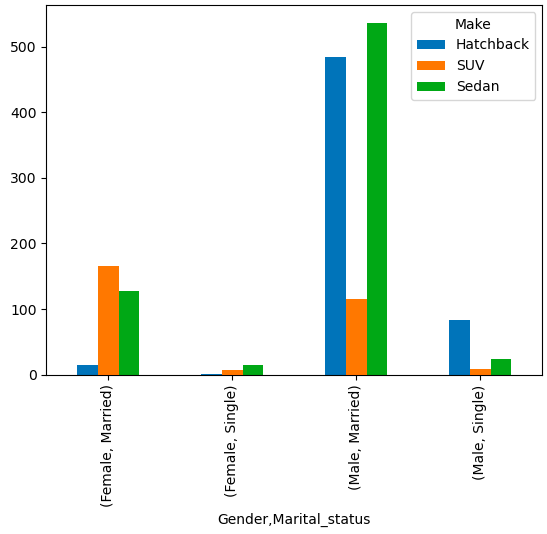
The below boxplot demonstrates that there is not much difference between the purchase behaviour of the individuals whose partners are working and not working.(The individuals having working partners might be willing to pool around $2,000 extra compared to the individuals whose partners are not working. The median of both the plots coincide i.e.; $31,000.



The below scatterplot, shows the correlation between the variable Price and Partner salary. As we can see, individuals having $0 as their partner salary is still able to afford cars ranging from [< $20k->$65k]. And as the distribution does not have any pattern in the spread of the data. It is evident that the variables are not correlated. From the inference drawn from the Heapmap values. The correlation of Partner Salary and Price is 0.17. **Correlation coefficient values below 0.3 are considered to be weak.Therefore, having a working partner does not lead to purchase of a higher priced car.**



8. **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 Gender and Marital\_status - fields to arrive at groups with similar purchase history.**



Based on the provided plot, it is evident that Married Males have a higher propensity to purchase Sedans/Hatchbacks compared to other groups. SUVs, on the other hand, are predominantly favored by Married Females. Additionally, Single Females display the lowest purchasing activity across all car types. Following Married Males, Married Females exhibit a relatively higher preference for Sedans in comparison to other groups.

A potential marketing strategy can focus on Sedans and Hatchbacks, targeting Married Males as they have shown a higher preference for these car types. Similarly, Married Females can be targeted for SUVs, as they have displayed a greater inclination towards this category. Effective marketing tactics ads can be employed to attract sales from these specific groups.

Problem: 2

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

Considering the objective of maximizing credit card usage and reducing attrition, the customer's behavior, his/her activity is of significant importance.

To list out the 5 important variables:

* Credit Card Activity for Last 30 Days, 60 Days, and 90 Days: Understanding the activity of customers over different time periods can provide insights into their credit card usage patterns. By analyzing these variables, the bank can identify customers who have consistent and active credit card usage, indicating higher intent to use the card. This information can help in targeting these customers for personalized offers, rewards, and retention strategies.
* Other Bank Credit Card Holding: There’s high competition in the market. Therefore, identifying whether customers already hold credit cards from other banks can provide insights into their existing credit card usage and preferences. Customers who already possess other credit cards may have a higher intent to use credit cards and can be targeted with specific offers and benefits to encourage them to use the bank's credit card more frequently.
* Last 3 Months Average Spends of the Customer: This variable provides insights into the customer's spending behavior over the past three months. By analyzing this data, the bank can identify customers who have shown consistent and substantial credit card usage. These customers are less likely to churn as they have demonstrated an ongoing need and interest in using their credit cards. The bank can focus on retaining these customers by providing personalized offers, rewards, and targeted communication to encourage continued credit card usage.
* Customer Spends for the current+next Months: Predicting and understanding customer spending patterns for future months can help the bank proactively engage with customers and encourage them to continue using their credit cards. By analyzing this variable, the bank can identify customers who are likely to have higher intent and predict future spending. This information can be used to design relevant marketing campaigns, offers, and promotions to drive credit card usage and reduce attrition.
* Card Type and Credit Card Limit: Right card to the right customer. That’s how we will be able to meet the customer needs and maximize credit card usage.The credit card limit assigned to a customer is an important factor in determining their spending capacity. Ensuring that customers have appropriate credit limits based on their income and creditworthiness can encourage them to utilize their credit cards more frequently. By monitoring and adjusting credit limits as necessary, the bank can empower customers to make desired purchases and increase credit card usage, reducing the likelihood of attrition.