**Car Customers Demographics Statistical Analysis**

**Report**

Submission By: **Raj Keswani**

Contents

**Problem Statement**

**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.**

**1.A.** What is the important technical information about the dataset that a database administrator would be interested in? ......................................................................................... 3

**1.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? ....................................................................................................................................... 4

**1.C.** Explore all the features of the data separately by using appropriate visualizations and draw insights that can be utilized by the business. ................................................................................ 6

**1.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. ......................................................................................................... 10

**1.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. ................................................ 15

**1.E.1.** Steve Roger says “Men prefer SUV by a large margin, compared to the women”

**1.E.2.** Ned Stark believes that a salaried person is more likely to buy a Sedan.

**1.E.3.** 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.

**1.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. ........... 17

**1.F.1.** Gender

**1.F.2.** Personal\_loan

**1.G.** From the current data set comment if having a working partner leads to the purchase of a higher-priced car…………………………………………………………………………………………………………………… 19

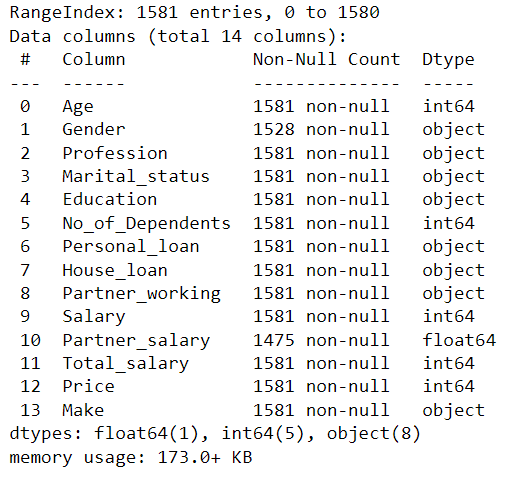
**1.H.** The main objective of this analysis is to devise an improved marketing strategy to send targeted information to different groups of potential buyers present in the data. For the current analysis use the Gender and Marital\_status - fields to arrive at groups with similar purchase history. ……………………………………………………………………………………………………………………………….… 21

Solution

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

The following technical information can be retrieved from the dataset which would be of significance to a database administrator.

* There are *1581* entries and *14* features in the dataset.
* *Age*, *No of Dependents*, *Salary*, *Total salary* and *Price* are numerical features and are of integer datatype.
* *Gender*, *Profession*, *Marital status*, *Education*, *Personal loan*, *House loan*, *Partner working* and *Make* are categorical features with object datatype.
* *Partner salary* is the only feature that is numerical in nature and has a float datatype.



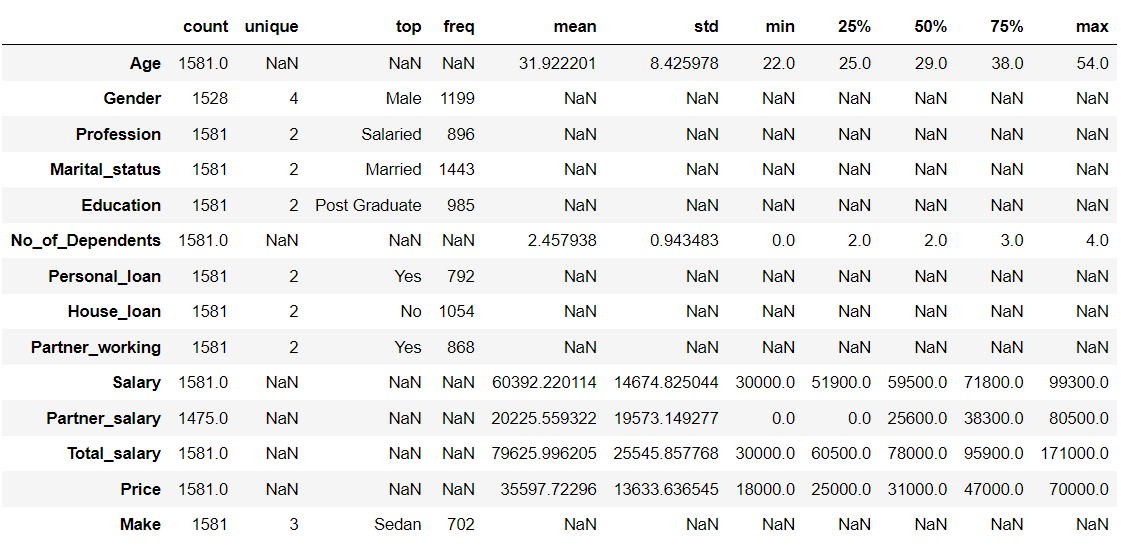
**Table 1.A.1** *Dataset Information*

1.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?

The preliminary analysis and in-depth quality check to ensure consistency along with the steps taken to eliminate the discrepancies is shared as follows.

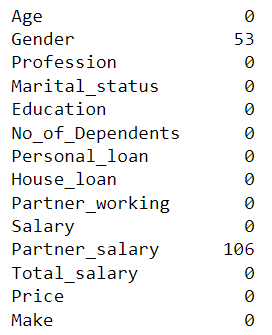
**Preliminary Analysis**

* All the numerical features seemed to be in expected range.
* Null values were there for column *Gender* and *Partner salary* that further needed to be investigated.
* Four unique values for *Gender* pointed to some possible junk values.



**Table 1.B.1** *Dataset Description*

* Upon further investigation, 53 and 106 null values were found in column *Gender* and *Partner salary* respectively.
* There were no duplicate row entries found in the dataset.



**Table 1.B.2** Sum of *Null Values column wise.*

**In Depth Quality Check** (Feature-wise)

Upon analysis, no discrepancies were found with the columns *Age*, *Marital status*, *Education*, *No of Dependents*, *Personal Loan*, *House Loan*, *Partner working*, *Salary*, *Price* and *Make*. The discrepancies and steps taken to resolve them are given feature wise as follows.

1. **Gender:**
   * *Null values present*: Substituted the Null Vales with Mode*.*
   * *Typos present*: Corrected the typos for consistency*.*
2. **Profession:**

* *Junk Values present*: Dropped the rows with junk values as the feature

categorical in nature and the difference between the

mode and the second highest occurring value is not very significant.

1. **Partner salary:**
   * *Null values present:* Substituted values based on the field *Partner working*

If partner is not working, substituted a value of 0.

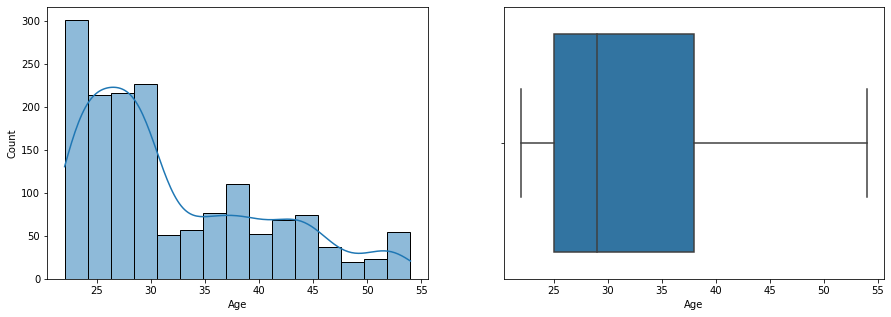
Else substituted Total\_salary - Salary for working partners.

1. **Changed the data type of columns *Salary*, *Partner salary*, *Total salary* and *Price* as float for consistency.**

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

Features from which some important business insights could be fetched are analyzed as follows.

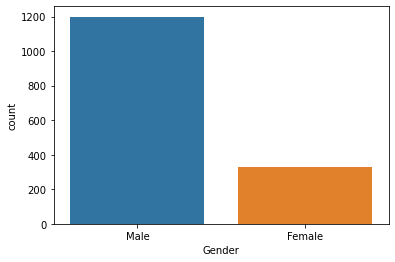
1. **Age**



**Figure 1.C.1** Age distribution of people who bought cars*.*

**Insight**: Younger people (22-30 Years) seem to buy significantly more cars than others. More than 75% of the cars have been sold to people who are <=38 years of age where median age seems to be lesser than 30.

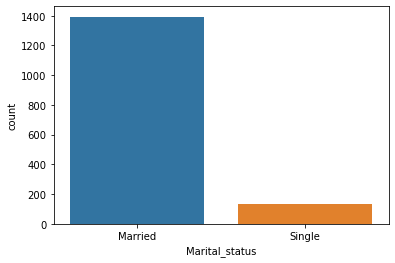
1. **Gender**



**Figure 1.C.2** Age distribution of people who bought cars*.*

**Insight**: Men seem to buy significantly more cars as compared to women.

1. **Marital Status**



**Figure 1.C.3** Marital Status of people who bought cars*.*

**Insight**: Married people seem to buy significantly more cars as compared to single people.

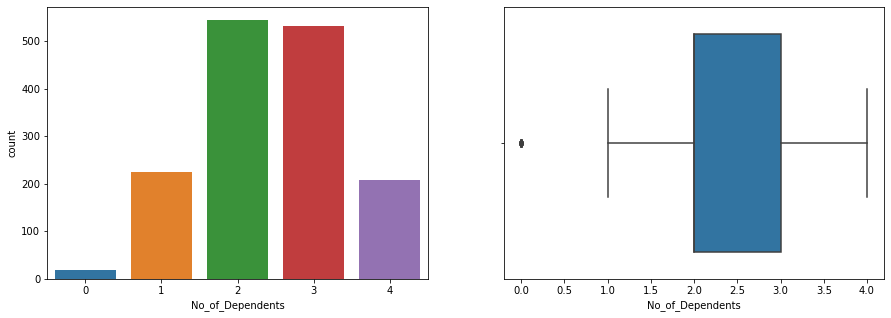
1. **Education**



**Figure 1.C.4** Education Background of people who bought cars*.*

**Insight**: Post Graduate people seem to buy more cars as compared to people who are Graduates.

1. **No of dependents**



**Figure 1.C.5** Education Background of people who bought cars*.*

## Insight: Most people have 2 to 3 dependents. Followed by 1 and 4. It is safe to assume people with 0 dependents don't buy cars frequently.

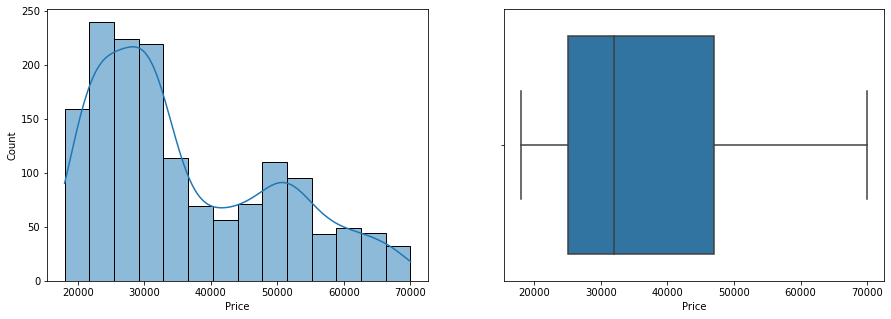
1. **House Loan**



**Figure 1.C.6** House loan status of people who bought cars*.*

## Insight: People prefer to buy a car if they do not have an outstanding House loan.

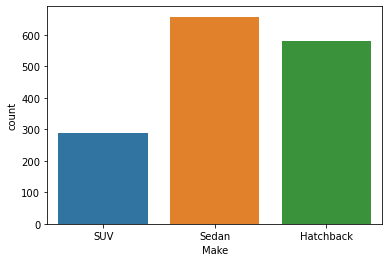
1. **Price**



**Figure 1.C.7** Price distribution of the Cars bought*.*

**Insight:** There is an inclination of people towards economical cars.

1. **Make**



**Figure 1.C.8** Make distribution of the Cars bought*.*

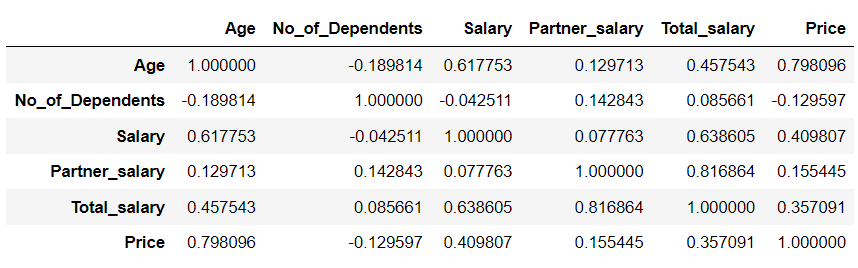
## Insight: Sedans are most popular. Followed by Hatchbacks and then by SUVs.

1.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.

**Preliminary Analysis**

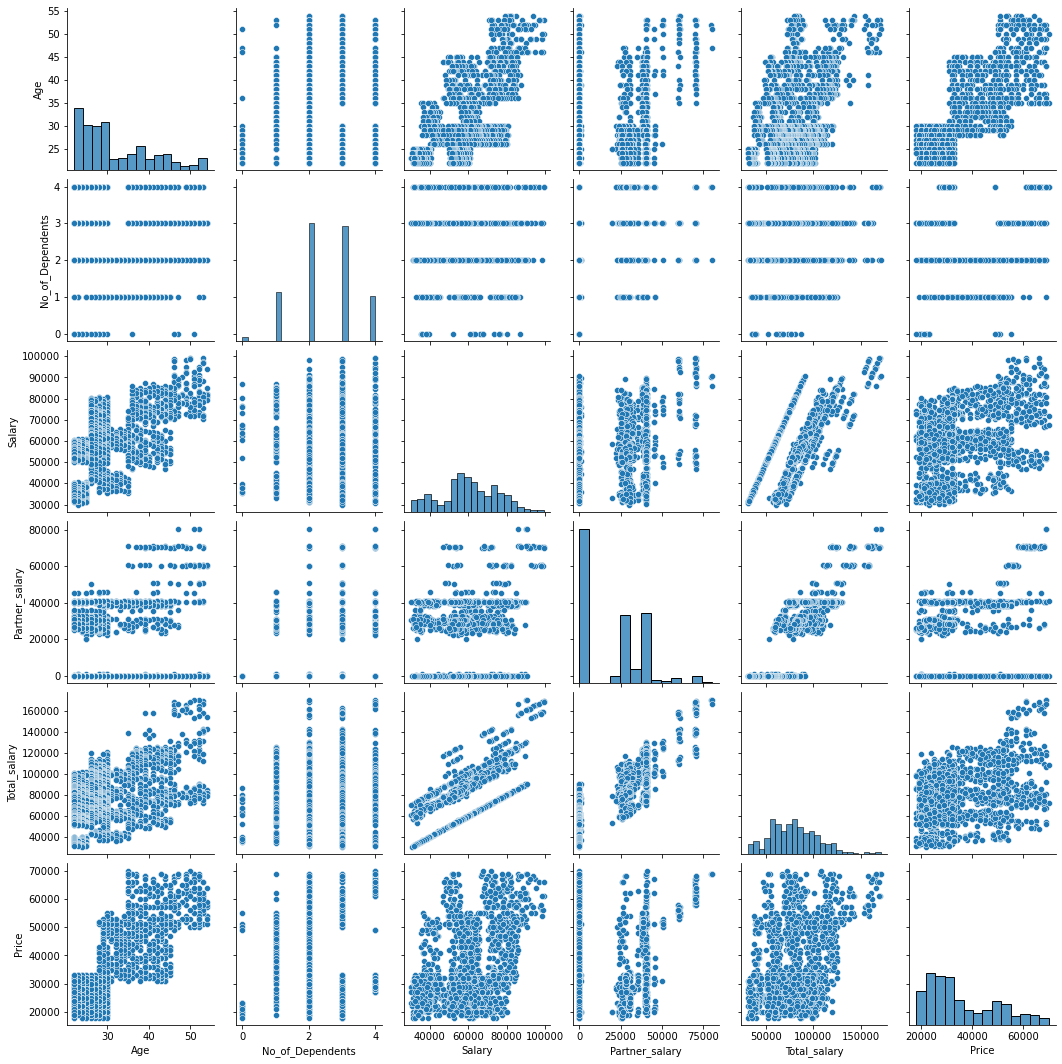
Preliminary analysis to study the relationships between variables was performed and observations are shared below.

* **Correlation matrix**



**Table 1.D.1** Make distribution of the Cars bought*.*

* As *Total Salary* is a derived column that depends upon columns *Salary* and *Partner Salary,* its positive correlation is expected and is of no significance here.
* The following parameters are positively correlated
  + - Age and Price
    - Age and Salary
    - Salary and Price
* **Pair Plot**



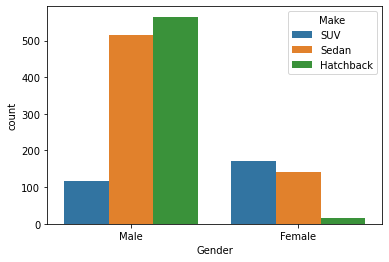
**Figure 1.D.1** pair plot for all the variables in the dataset.

* The relationship between Age and Price required further analysis as there seems to be another factor playing a role here.

**In Depth Analysis**

The insights fetched after analyzing the relationships between the features in detail are shared as follows.

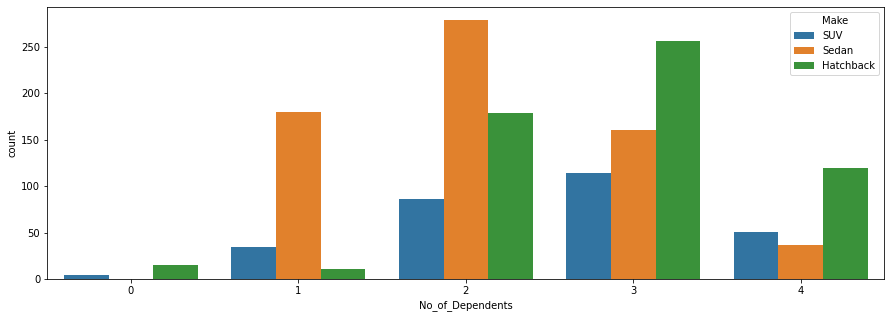
1. **Relationship between Gender and Make.**



**Figure 1.D.2** Distribution of Make the Cars across Customer gender*.*

* The most popular make across Males is Hatchback whereas Females seem to prefer SUV.
* Sedans are somewhat liked by both since it has highest sales.
* Males least prefer the SUV whereas Females least prefer Hatchbacks.

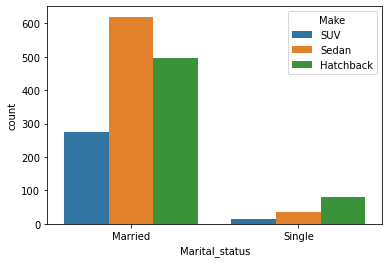
1. **Relationship between No. of Dependents and Make.**



**Figure 1.D.3** Distribution of Make the Cars across No. of Dependents customer has.

* As the number of People with high number of dependents (3 or 4) seem to prefer hatchbacks over other makes.
* For people with 1 to 2 dependents, the choice is a Sedan.
* As number of dependents increase, the chance of people buying an SUV also increases.

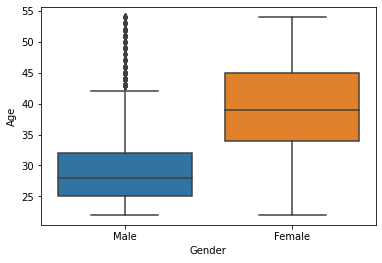
1. **Relationship between Marital Status and Make.**



**Figure 1.D.4** Distribution of Make the Cars across Marital Status of customers

* Married people seem to prefer a Sedan
* Single people prefer Hatchback

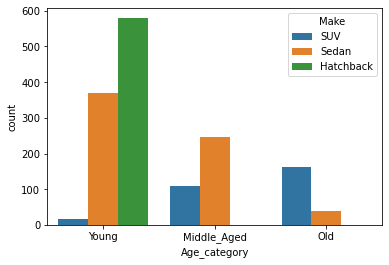
1. **Relationship between Gender and Age.**

.

**Figure 1.D.5** Distribution of Age of Customers across Gender of customers.

* Most Males seem to by a car at an earlier point in life centered at 27-28 Years of age apart for a few exceptions.
* Females by a car at a later point in time, centered at about 40 Years of age.

1. **Relationship between Age category and Make**

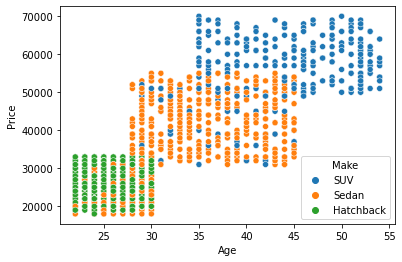


**Figure 1.D.6** Distribution of Age of Customers across

different Makes of cars sold.

* Young people prefer Hatchback.
* Middle aged people prefer Sedan.
* People older in age prefer SUVs.

1. **Relationship between Age and Price in terms of Make**



**Figure 1.D.7** Scatterplot of Price of Cars and Age of Customers across Makes.

* Most of the hatchbacks cost around 18k-33k ad are bought by people in the age of 22-30.
* Most of the sedans cost around 30k-55k and are bought by people in the age of 28-45.
* Most of the SUVs cost around 50k-70k ad are bought by people in the age of 35-54.

1.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.

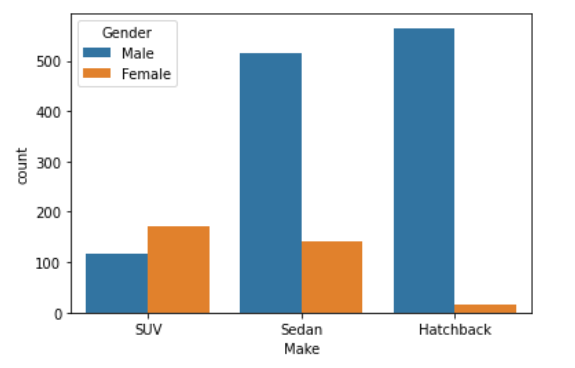
1.E.1. Steve Roger says “Men prefer SUV by a large margin, compared to the women”

1.E.2. Ned Stark believes that a salaried person is more likely to buy a Sedan.

1.E.3. 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.

After analyzing the available data, the following observations are made.

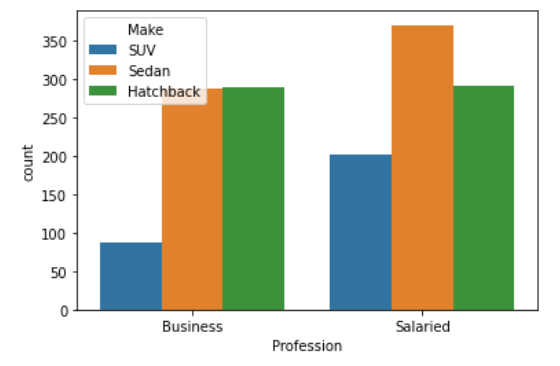
1. ***Steve Roger’s statement seems to be incorrect as Females seem to prefer SUVs more when compared to Males.***



**Figure 1.E.1** Number of people buying different Makes

across Genders*.*

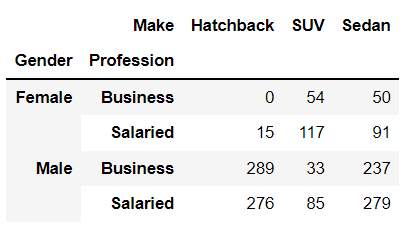
1. ***Ned Stark’s statement is correct. A salaried person is more likely to buy a Sedan.***



**Figure 1.E.2** Number of people buying cars

from different Professions across Makes*.*

1. ***Sheldon Cooper’s statement is also incorrect as Salaried Males prefer Sedan or Hatchback over SUV.***



**Table 1.E.1** Number of people buying various Makes

Across Genders and profession.

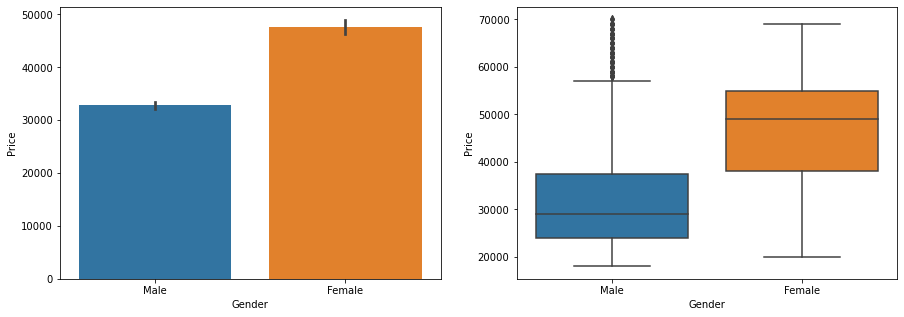
1.F. From the given data, comment on the amount spent on purchasing automobiles across the following categories. Comment on how a business can utilize the results from this exercise. Give justification along with presenting metrics/charts used for arriving at the conclusions.

1.F.1. Gender

1.F.2. Personal loan

The distribution of price is analyzed across gender and personal loan and the observations and recommended target strategies are shared as follows.

1. **Gender**

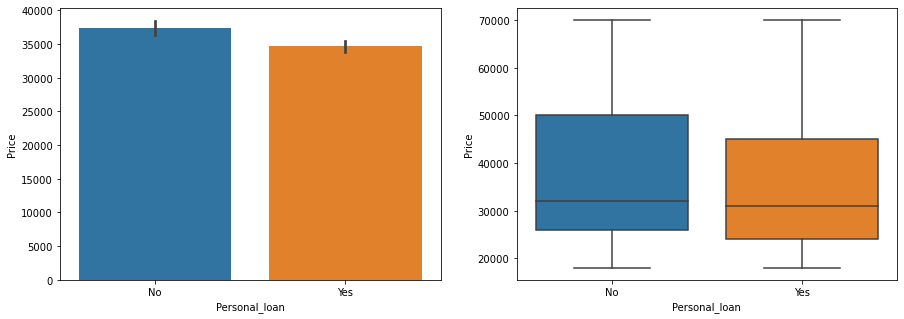


**Figure 1.F.1** *Bar plot and Box plot of Price of cars across Genders.*

*Observation:* Females spend extravagantly while buying a car while Males tend to be more economical apart from a few exceptions.

*Target Strategy:* Economical makes of cars and economical variants of the same make can be targeted towards Males while the Premium makes of cars and premium variants of the same make can be targeted towards Female customers.

1. **Personal loan**



**Figure 1.F.2** *Bar plot and Box plot of Price of cars across outstanding Personal Loan.*

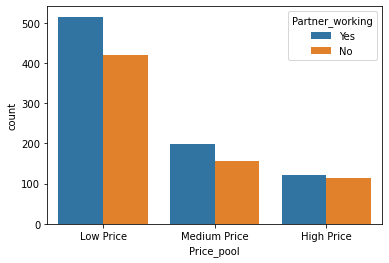
*Observation:* People with a personal loan tend to spend less while buying a car as compared to people with no outstanding personal loan. Although the difference doesn’t seem to be very much.

*Target Strategy:* Variants with less features can be targeted towards the people with an outstanding personal loan while variants with more features can be targeted towards the people without any personal loan.

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

To study the impact of having a working partner while buying a high-priced car, the following observations have been made. The price range of the cars is divided in three price pools as Low priced, Medium priced and High priced.

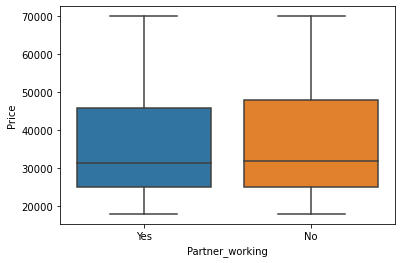
1. As the price of car increases, the difference between the number of people with and without a working partner decreases.



**Figure 1.G.1** *Number of people with and without a Working Partner*

*against the Price Pool of the car purchased.*

1. The distribution of Price of car purchased by the people who do not have a working partner seems to be slightly higher than that of purchased by the people who have a working partner.



**Figure 1.G.2** Distribution of *Price of Cars for people*

*with and without a Working Partner*

**Based on above observations, the following information is inferred.**

* There seems to be a positive impact of having a working partner while purchasing a car.
* The impact seems to diminish as the price of the car increases.
* For high priced cars, even though the impact of having a working partner is still positive but is statistically insignificant.
* If there was a strong impact of having a working partner on purchase of higher priced cars, the distribution of Price for people with a working partner should have been slightly higher than the people without a working partner.

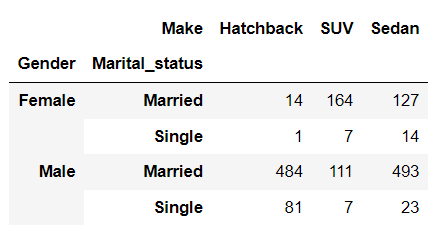
1.H. The main objective of this analysis is to devise an improved marketing strategy to send targeted information to different groups of potential buyers present in the data. For the current analysis use the Gender and Marital status - fields to arrive at groups with similar purchase history.

The analysis has been broken down into two sections. In the first section, analysis is directed towards identifying the preferred Make of various subgroups in terms of Gender and Marital Status. In the second section, the focus is transferred to the study the spending patterns of the subgroups. At the end of the two sections, the insights are clubbed, and target strategies are recommended for each subgroup.

**Section 1.H.a**

The following insights can be derived while analyzing the purchase history of various genders across marital statuses.

* Most of the Married Females prefer SUV.
* Most of the Unmarried Females prefer Sedan.
* In general, Females don’t seem to prefer Hatchback.
* Most of the Married Males either prefer Sedan or a Hatchback.
* Most of the single men prefer Hatchback.



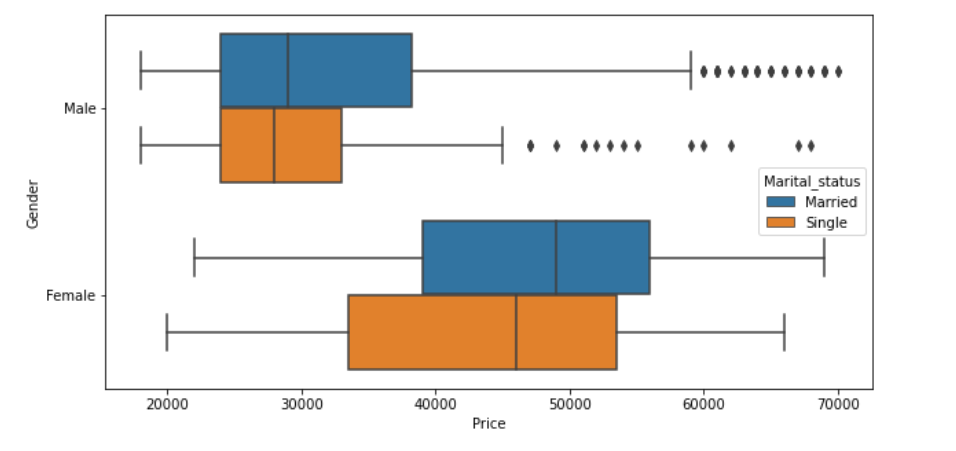
**Table 1.H.a.1** *Purchase history of different Car Makes across*

*Genders and Marital Statuses*

**Section 1.H.b**

The following insights can be derived while trying to analyze the spending history of various genders across marital statuses.

* Most males seem to be more economical while buying a car with a few exceptions whereas Females seem o be extravagant while buying a car.
* Married people seem to spend more while buying a car in comparison with Unmarried people.



**Figure 1.H.b.1** *Spending history of various Genders across Marital Statuses*

*while buying a car.*

**Recommended Target Strategy for potential buyers**

1. *Single Males:*Economical Hatchbacks
2. *Married Males:*Premium Hatchbacks or Economical Sedans
3. *Single Females:*Economical SUVs and Premium Sedans
4. *Married Females:* Premium SUVs and Sedans