Contents

|  |  |  |
| --- | --- | --- |
| **Question No.** | **Problems** | **Page No.** |
|  | Austo Automobile Company Data format (Variables) | 2 |
| A | *What is the important technical information about the dataset that a database administrator would be interested in? (Hint: Information aNbout the size of the dataset and the nature of the variables)* | 3 |
| 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 |
| C | *C. Explore all the features of the data separately by using appropriate visualizations and draw insights that can be utilized by the business.* | 5 to 9 |
| 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 |
| 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.* | 11 to 13 |
| 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.* | 13 |
| G | *From the current data set comment if having a working partner leads to the purchase of a higher-priced car.* | 14 |
| 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.* | 15 to 16 |

**Austo Automobile Company Data format (Variables)**

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Gender | Profession | Marital  \_status | Education | No\_of\_  Dependents | Personal  \_loan | House  \_loan | Partner\_  working | Salary | Partner  \_salary | Total\_  salary | Price | Make |
| Male | Business | Married | Post Graduate | 4 | No | No | Yes | 99300 | 70700 | 170000 | 61000 | SUV |
| Female | Salaried | Married | Post Graduate | 4 | Yes | No | Yes | 95500 | 70300 | 165800 | 61000 | SUV |
| Female | Salaried | Married | Post Graduate | 3 | No | No | Yes | 97300 | 60700 | 158000 | 57000 | SUV |
| Female | Salaried | Married | Graduate | 2 | Yes | No | Yes | 72500 | 70300 | 142800 | 61000 | SUV |
| Male | Salaried | Married | Post Graduate | 3 | No | No | Yes | 79700 | 60200 | 139900 | 57000 | SUV |

**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)**

**The size of the data is**

1. **Rows-1581**
2. **Columns/Features 14**

***Nature of the Variables dtypes:***

1. ***float64 (1),***
2. ***int64 (5),***
3. ***object (8)***

|  |
| --- |
| # 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 |

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

There are 53 Gender & 106 Partner Salary Discrepancies in the given dataset

Age 0

Gender 53

Profession 0

Marital\_status 0

Education 0

No\_of\_Dependents 0

Personal\_loan 0

House\_loan 0

Partner\_working 0

Salary 0

Partner\_salary 106

Total\_salary 0

Price 0

Make 0

dtype: int64

**Question :Are there any discrepancies present in the data**

## *Ans: Yes, There are two values in Gender of (female) category that has spelling mistake*

Male 1199

Female 327

Femal 1

Femle 1

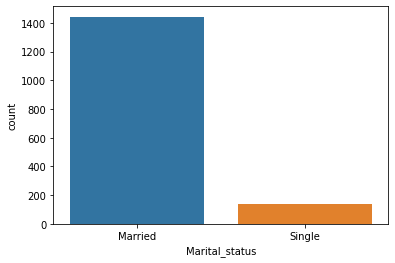
# C. Explore all the features of the data separately by using appropriate visualizations and draw insights that can be utilized by the business

# We used boxplot in below to check the outliers in our dataset visually & we found out that "Total\_salary" had many outliers. So we need to remove them.

# 

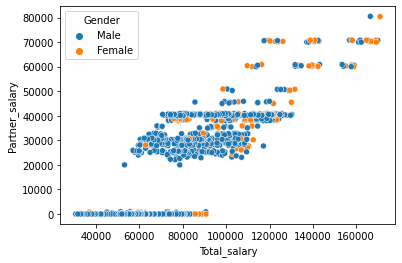
# So we can infer from the below countplots that our Gender, Marital\_status and House\_loan feature has heavy class imbalance problem which may lower accuracy of our model.

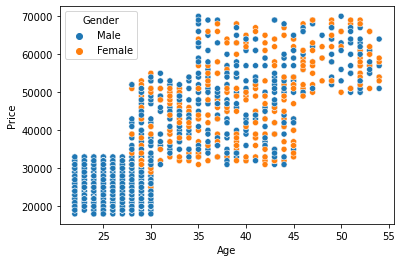
# 



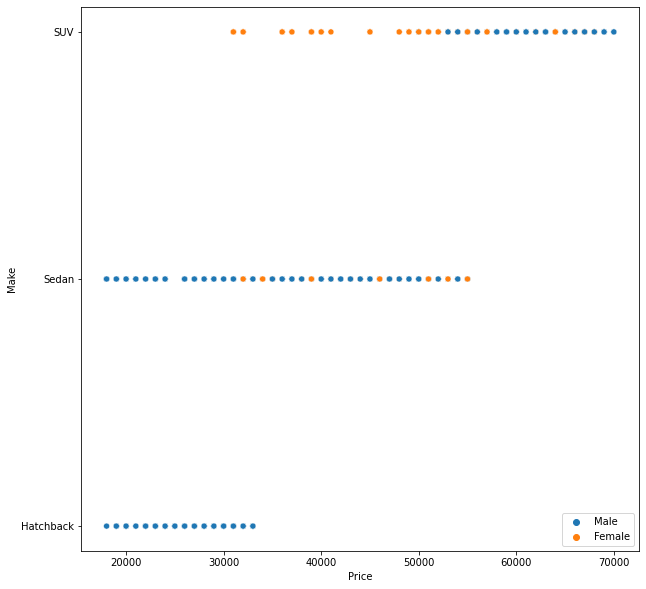


Draw insights that can be utilized by the business

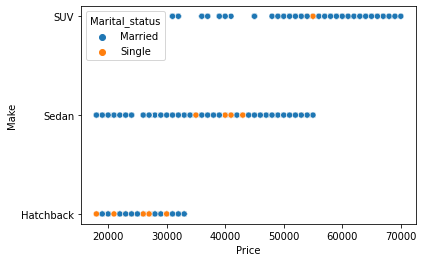


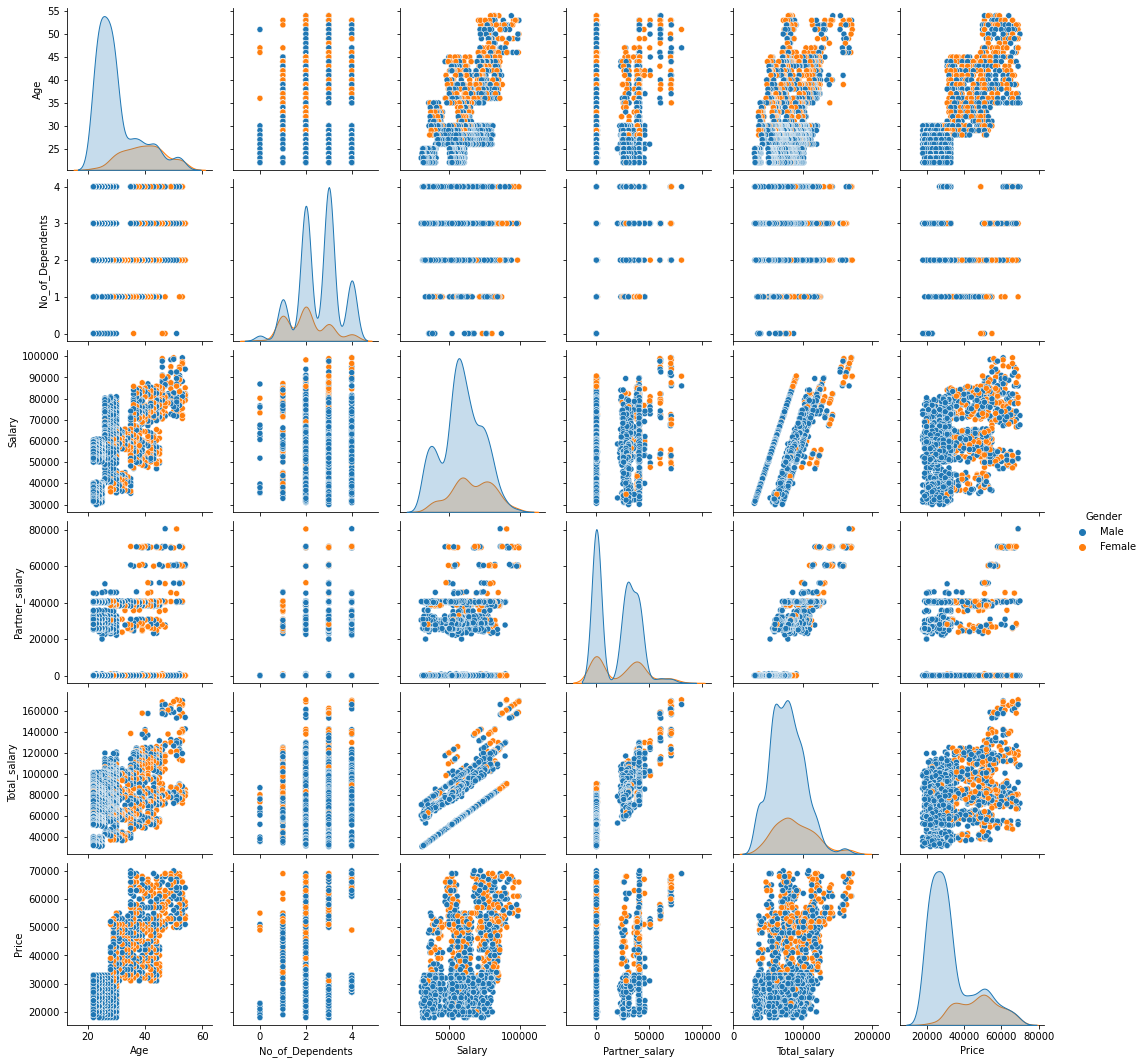


# So from the scatter plot below we can say that there were suv's considered by both male aswell as female buyers at larger price range than hatchback and sedan as suv's are much comfortable, safer and easy to drive

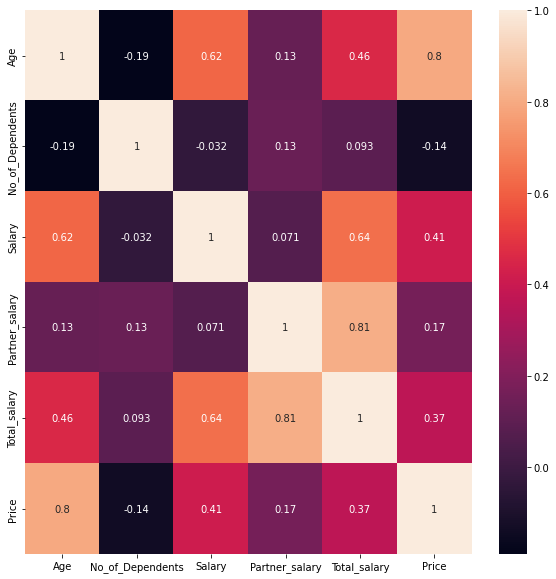


# So from the scatter plot below we can say that there were more married buyers than single one's for all type of make





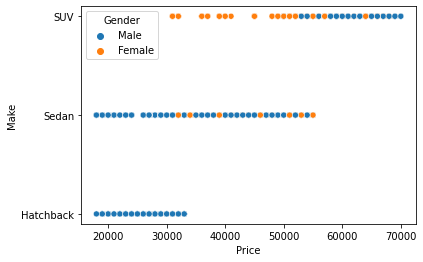
D. **Understanding the relationships among the variables in the dataset is crucial for every analytical project. Perform analysis on the data fields to gain deeper insights. Comment on your understanding of the data.**



# 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.

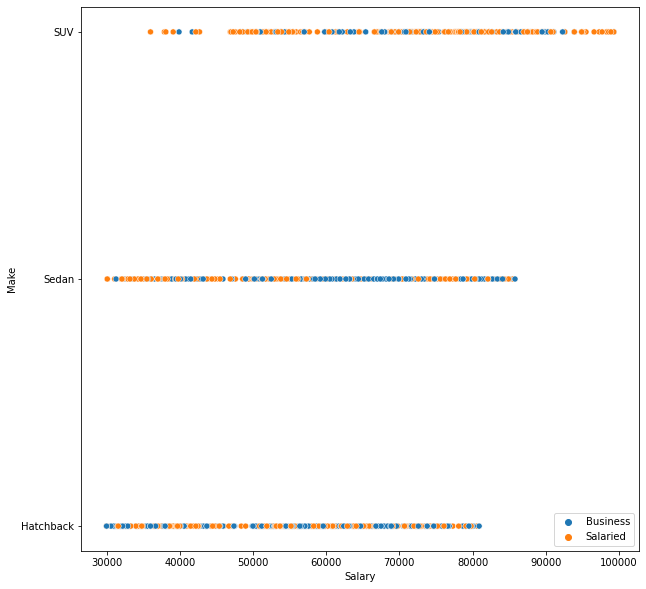
# E1) Steve Roger says “Men prefer SUV by a large margin, compared to the women”

# Ans 1. No, as there are more female data points for suv than male according to our scatter plot.[¶](http://localhost:8888/notebooks/Downloads/Python%20Downloads/final_final.ipynb#Ans-1.-No,-as-there-are-more-female-data-points-for-suv-than-male-according-to-our-scatter-plot.)



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

## Ans 2. Yes, 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

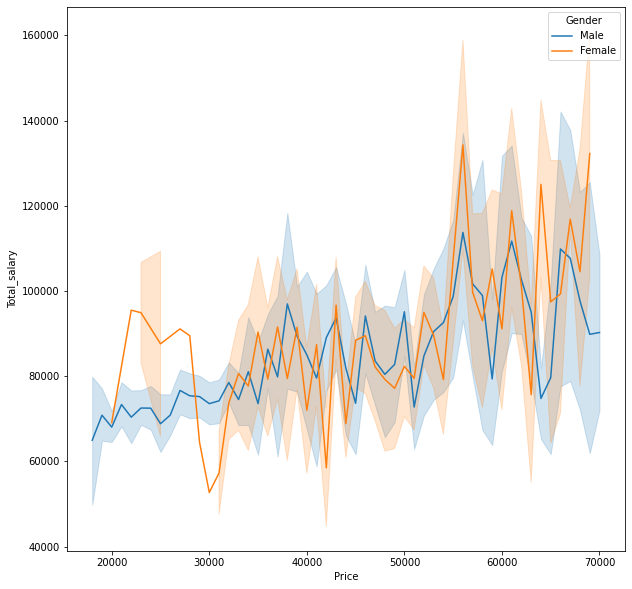
## Ans 3. Yes its true that a salaried male is an easier target for a SUV sale over a Sedan Sale but he is wrong about the statements of Steve rogers and ned stark



# 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.

**So from the below plots firstly we can infer that female buyers are more likely to buy expensive cars than male buyer.**



# Secondly people who have higher total salary and can also take personal loan are likely to buy expensive cars with respect to people who have lower total salary and have not opted for personal loan.[¶](http://localhost:8888/notebooks/Downloads/Python%20Downloads/final_final.ipynb#Secondly-people-who-have-higher-total-salary-and-can-also-take-personal-loan-are-likely-to-buy-expensive-cars-with-respect-to-people-who-have-lower-total-salary-and-have-not-opted-for-personal-loan.)

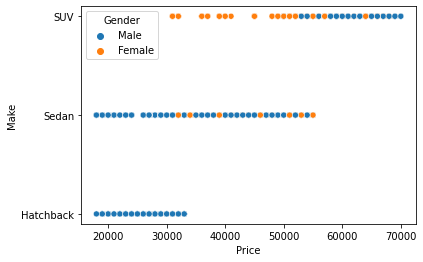
# 

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

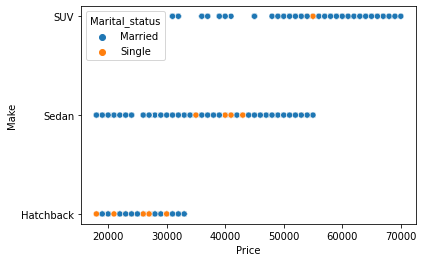
# 

**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.**

# *So from the scatter plot below we can say that there were suv's considered by both male aswell as female buyers at larger price range than hatchback and sedan as suv's are much comfortable, safer and easy to drive.*

****

# So from the scatter plot below we can say that there were more married buyers than single one's for all type of make.[¶](http://localhost:8888/notebooks/Downloads/Python%20Downloads/final_final.ipynb#So-from-the-scatter-plot-below-we-can-say-that-there-were-more-married-buyers-than-single-one's-for-all-type-of-make.)

****