

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.

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)

Solution A

Data set- austo_automobile (2) (1)-1.csv

The data is about the Austo Motor Company is a leading car manufacturer specializing in SUV, Sedan, and Hatchback models. Below are some important technical information about the dataset:

- The data set contains of 1581 rows & 14 columns.
- The data contains 5 Integer, 1 float & 8 Object data type columns.

S.No	Columns	Data Type
1	Age	int64
2	Gender	object
3	Profession	object
4	Marital_status	object
5	Education	object
6	No_of_Dependents	int64
7	Personal_loan	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?

Solution B

After a critical look at quality check the following discrepancies are noticed.

- Unique values in columns with object data type.

Columns	Data Type	Unique values
Gender	object	Male, Femal, Female, nan, Femle
Profession	object	Business, Salaried
Marital_status	object	Married, Single
Education	object	Post Graduate, Graduate
Personal_loan	object	No, Yes
House_loan	object	No, Yes
Partner_working	object	No, Yes
Make	object	SUV, Sedan, Hatchback

- There are missing values in "Gender" & "Partner_salary" Columns.

Columns	Data Type	No of Missing Values
Gender	object	53
Partner_salary	float64	106

Columns	Data type	NO of missing values
Gender	object	53
Partner_salary	float64	106

- It is also noticed that in column "Gender" the spelling of Female is incorrectly mentioned adding to 2 more unique values i.e. "Femal" & "femle".

- Below is the screenshot of the spelling errors being fixed.

```
In [50]: # fixing spelling errors in 'Gender' column
df['Gender'] = df['Gender'].replace('femal','Female')
df['Gender'] = df['Gender'].replace('femle','Female')

In [51]: #updated Unique values in Gender column.
df['Gender'].unique()

Out[51]: array(['Male', 'Female', nan], dtype=object)
```

- Total No. of females = 329 & Males= 1199 and there are 53 Nan (null values).
- The quantitative characteristics of the data is as follows, this data seems to be fine with no abnormalities.

columns	count	mean	std	min	25%	50%	75%	max
Age	1581	31.9222	8.425978	22	25	29	38	54

Salary	1581	60392.22	14674.83	30000	51900	59500	71800	99300
Partner_salary	1475	20225.56	19573.15	0	0	25600	38300	80500
Total_salary	1581	79626	25545.86	30000	60500	78000	95900	171000
Price	1581	35597.72	13633.64	18000	25000	31000	47000	70000

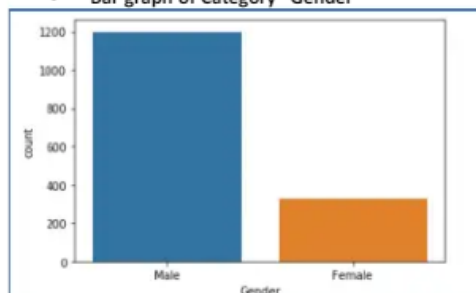
- C. Explore all the features of the data separately by using appropriate visualizations and draw insights that can be utilized by the business.

Solution C

Different Visualisations and its observations are as follows:

Uni-Variate Analysis

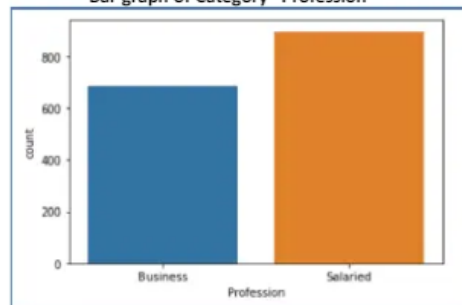
- Bar graph of Category "Gender"



Gender
Male
Female

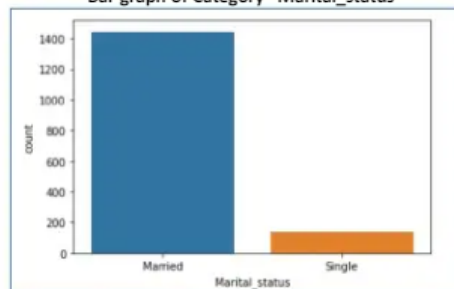
Observation: The no of males is more than the no of females in the data set provided. Which means that there are more male Buys for cars than Females.

- Bar graph of Category "Profession"



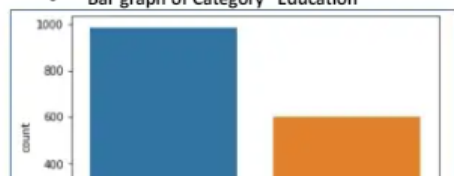
Observation: The no. of Salaried professionals is more than the no of Business professionals.

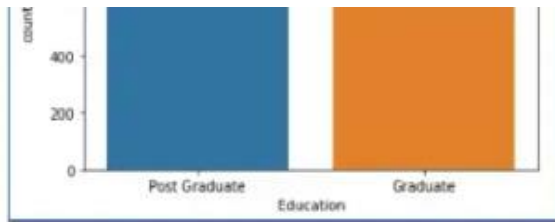
- Bar graph of Category "Marital_status"



Observation: The data set contains more than 1400 married buyers and less than 200 buyers who are single.

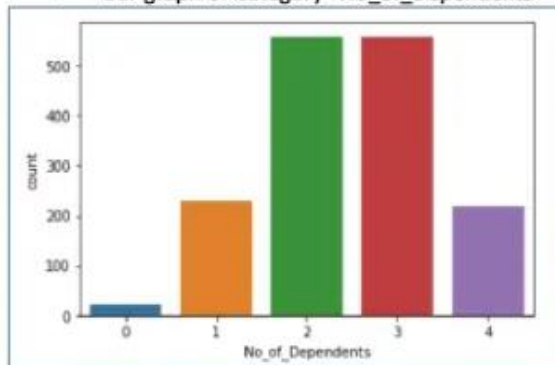
- Bar graph of Category "Education"





Observation: There are more post graduates and graduates in the data set.

- Bar graph of Category "No_of_Dependents"



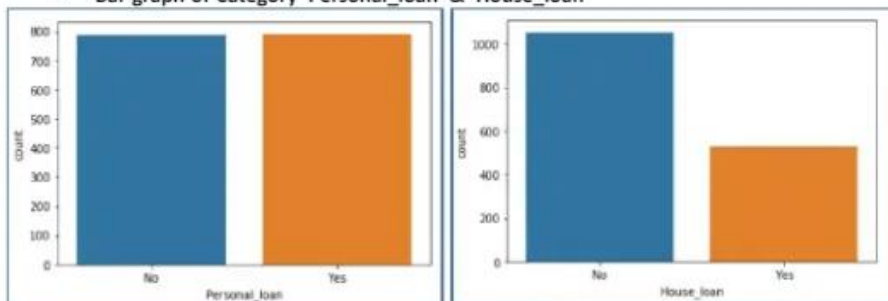
Observation:

1. There are more 500 people with no. of dependents = 2 & 3

Observation:

1. There are more 500 people with no. of dependents = 2 & 3
2. There are almost 200 people with no. of dependent = 1 & 4
3. There are less than 30 people with 0 dependents.

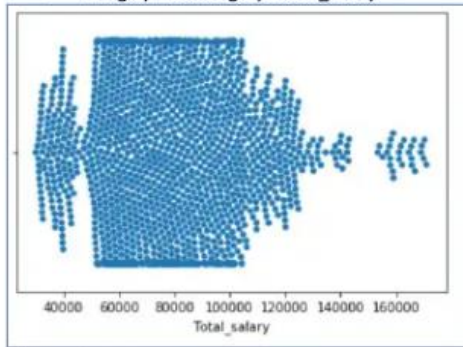
- Bar graph of Category 'Personal_loan' & 'House_loan'



Observation:

1. The no of people having & not having personal loan is same.
2. The no of people having is almost double as compared to people not having house loan is .

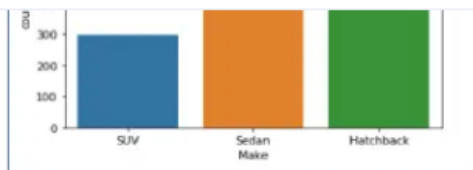
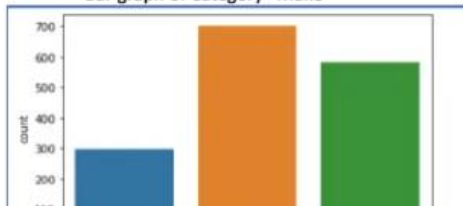
- Bar graph of Category "Total_salary"



Observation:

1. Most of the Buys have Total salary between the range of 50000-100000.
2. There are very less buys with total salary greater than 130000.

- Bar graph of Category "Make"



Observation:

1. Highest no of cars sold in respect to "make" category is: Sedan
 2. Lowest no of cars sold in respect to "make" category is: SUV
- 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.

Solution D

Bi-Variate Analysis

- Pivot Table for Gender & Type of profession they are working.

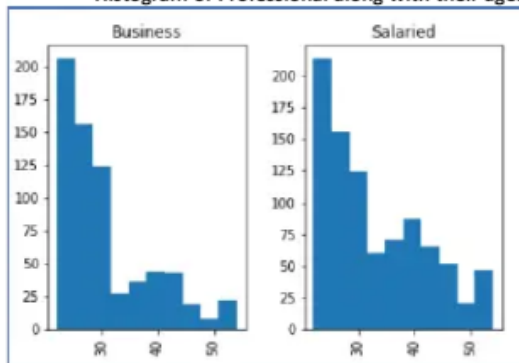
Profession	Business	Salaried
Gender		
Female	105	224
Male	559	640

Observation:

1. There are more male in the data set as compared to Females.
2. The no of salaried females is more than double the no of businesswomen.

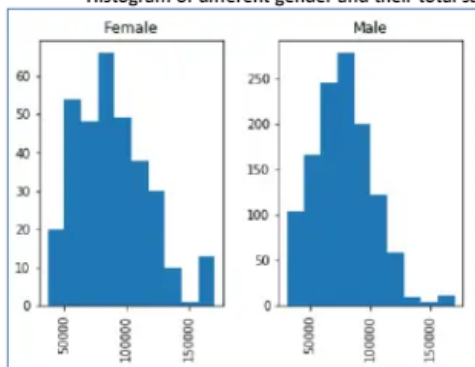
Observation:

1. There are more male in the data set as compared to Females.
 2. The no of salaried females is more than double the no of businesswomen.
 3. The no of salaried males and more in comparison to businessmen in the data set provided.
- **Histogram of Professional along with their age.**



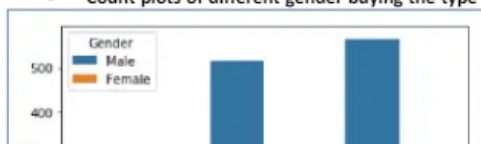
Observation: The data set contains more professional of age below 30 in both Business & salaried category.

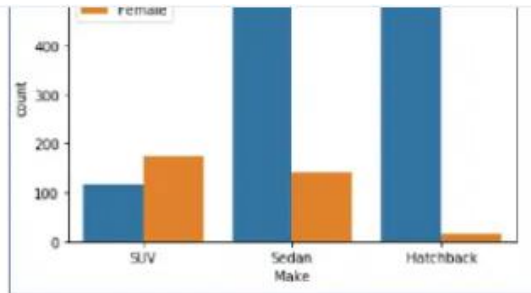
- **Histogram of different gender and their total salary.**



Observation: The data set contains more professional in total salary slab between 50000-100000 both males and females.

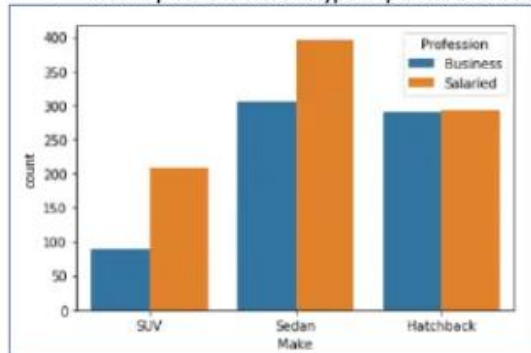
- **Count plots of different gender buying the type of car.**





Observation: There are more male buyers for 'Sedan' & 'hatchback' as opposed to 'SUVs' i.e. visa versa.

- Count plots of different type of professionals buying the type of car.



Observation:

- Salaried professionals have bought more 'SUVs' than Business professionals.



Observation:

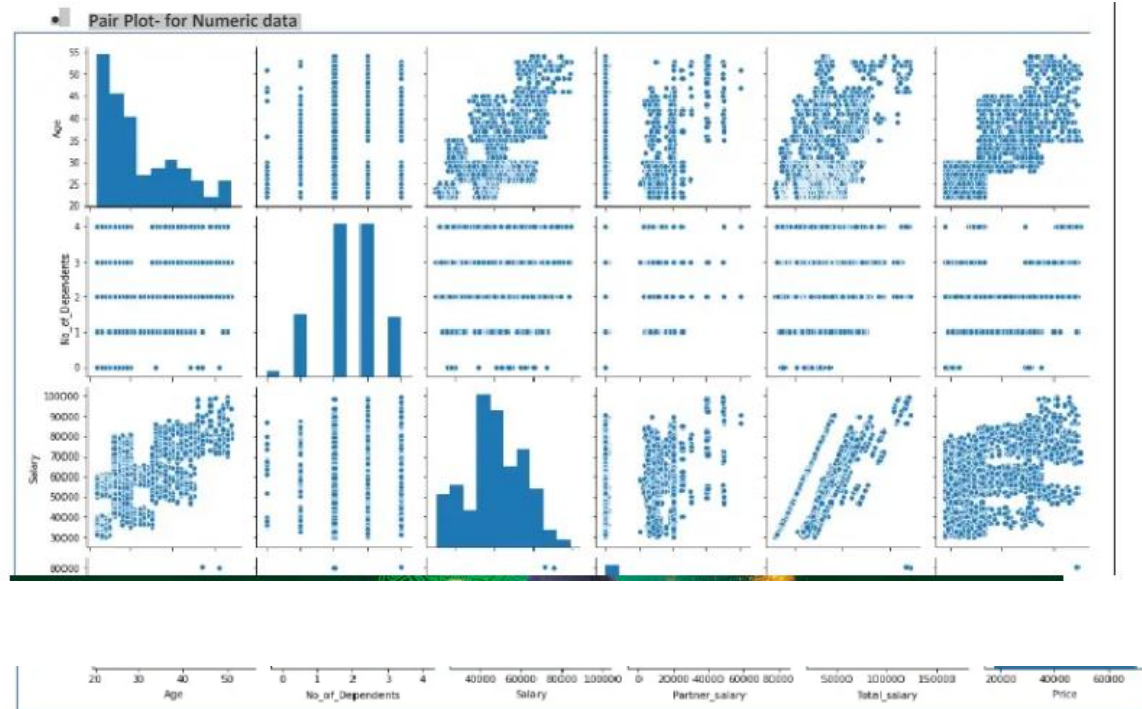
- Salaried professionals have bought more 'SUVs' than Business professionals.
- Salaried professionals have bought more 'Sedans' than Business professionals
- Both salaried & business professionals have bought almost equal no. of 'Hatchbacks'

Multi-Variate Analysis

- Pivot table depicting marital status along with No. of dependents and the type of Car bought.

No_of_Dependents	Marital_status	Age		
		Make	Hatchback	Sedan
0	Single		15.0	5.0
1	Married		10.0	34.0
	Single		1.0	13.0
2	Married		113.0	76.0
	Single		68.0	25.0
3	Married		256.0	117.0
4	Married		119.0	54.0

Observation: The Married people with children more than 2 have mostly bought either "hatchback" or "Sedan". The reason behind this can be that they have also have personal loans & house loans and there is no budget to buy SUV's.



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.

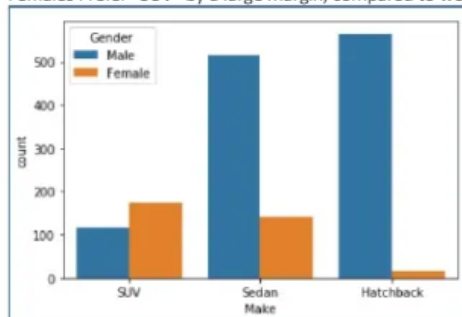
E1) Steve Roger says "Men prefer SUV by a large margin, compared to the women"

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.

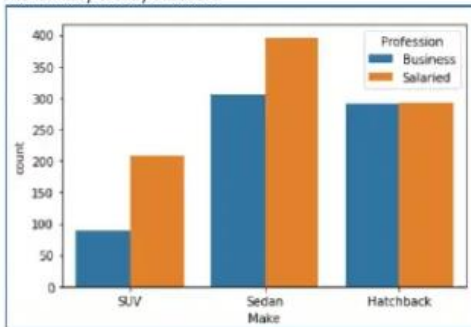
Solution E1

Steve Roger says "Men prefer SUV by a large margin, compared to the women" : **Disagree**, it is clearly visible from the bar graph that Females Prefer "SUV" by a large margin, compared to women



Solution E2

Ned Stark believes that a salaried person is more likely to buy a Sedan: **Agree**, it is clearly visible from the bar graph that salaried person is more likely to buy a Sedan.



Solution 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: **Disagree**, it is clearly visible from the chart below that salaried males prefer 'Sedan' much more than the 'SUVs'.

Profession	Gender	Make	count
Business	Female	SUV	55
		Sedan	50
	Male	Hatchback	289
		SUV	33
		Sedan	237
Salaried	Female	Hatchback	15
		SUV	118
	Male	Sedan	91
		Hatchback	276
		SUV	85

Salaried	Female	Sedan	237
		Hatchback	15
	Male	SUV	118
		Sedan	91
		Hatchback	276

- 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

Solution F1

Gender	Make	Amount Spent
Female	Hatchback	412000
	SUV	9252000
	Sedan	6031000
Male	Hatchback	14959000
	SUV	7031000
	Sedan	17358000

Observation:

1. The amount spent on 'SUV' by females is more than that spend my males.
2. The amount spent on 'Hatchback' is more than triple, compared to amount spend by females.
3. The amount spent on 'Sedan' is more almost triple, compared to amount spend by females.

F2) Personal_loan

2. The amount spent on 'Hatchback' is more than triple, compared to amount spend by females.
3. The amount spent on 'Sedan' is more almost triple, compared to amount spend by females.

F2) Personal_loan

Solution F2

Personal_loan	Make	
No	Hatchback	7765000
	SUV	10373000
	Sedan	10852000
Yes	Hatchback	7643000
	SUV	6207000
	Sedan	13440000

Observation:

1. The amount spent on 'SUV' & 'Sedan' is more from buyers having no personal loan.
2. The amount spent on 'Sedan' is more from buyers having personal loan.
3. The amount spent on 'SUV' & 'Hachback' is less from buyers having personal loan.

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

Solution G

Partner_working	Make	Price	
Yes	Hatchback	32000	38
		20000	35
		30000	32
		22000	30
		28000	28
No	Sedan	27000	27
		32000	27
		30000	27
		25000	27
		23000	25

Partner_working	Make	Price	
Yes	Sedan	22000	1
		32000	1
		39000	1
		48000	1
		70000	1
No	SUV	40000	1
		41000	1
		48000	1
		70000	1
		22000	1

Observation: Yes, From the current data set provided the buyers having working partner tend to purchase car with Higher price.

- 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.

Solution H

Marital_status	Gender	Make	
Married	Female	Hatchback	14

Marital_status	Gender	Make	
Married	Female	Hatchback	14
		SUV	166
		Sedan	127
	Male	Hatchback	484
		SUV	111
		Sedan	493
Single	Female	Hatchback	1
		SUV	7
		Sedan	14
	Male	Hatchback	81
		SUV	7
		Sedan	23

Observations: From the above chart the marketing strategy can be devised to send targeted information to different groups of potential buyers present in the data i.e.

1. As Married Males tend to buy more of 'Sedans' & 'Hatchback' may be due to budget constraints and other liabilities like 'Personal loans' & 'house loans. Multiple schemes can be provided to them for discounts on 'SUVs' . Which can come under their budget and is more spacious as per the family requirements.
2. Single Males tend to buy more of 'Hatchback' cars as compared to other two categories. Marketing strategy can be devised in a way to target only single males for increasing sales of 'hatchback' cars.
3. Married females tend to buy more of 'SUVs' & 'Sedans' as cars as compared to 'hatchback' . Marketing strategy can be devised in a way to target only married females for increasing sales of 'SUVs' & 'Sedans'.