EDA Visulaization (Austo Motor Company)

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 ink (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

- 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.

```
[14]: # fixing spelling errors in "Gender" column
      df.Gender = df.Gender.replace(["Femal","Femle"],["Female","Female"])

[15]: # updated Unique values in Gender column.
      df.Gender.unique()

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

- Total No. of females = 329 & Males= 1199 and there are 53 Nan (null values).
- Most frequent value is Male in gender column so we can fill all the null values with Male

```
[ ]: df.Gender = df.Gender.fillna("Male")

18]: df.isnull().sum()

18]: Age          0
     Gender       0
     Profession   0
```

- Partner Salary column have 106 null values, below is the screenshot of the null values being fixed

```
[20]: # fetch the required col to check the link among them
      new=df[df.Partner_salary.notnull()][["Salary","Partner_salary","Total_salary"]]

[22]: # to check the formula
      (new.Total_salary == new.Salary + new.Partner_salary).unique()

[22]: array([ True])

[23]: # fill the null values of partner salary using formula
      df.Partner_salary = df.apply(lambda row: row.Total_salary-row.Salary
                                   if np.isnan(row.Partner_salary) else row.Partner_salary,axis=1)
```

- Now , there is no null values in our dataset
- 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	31.922201	8.425978	22	25	29	38	54
No_of_Dependents	1581	2.457938	0.943483	0	2	2	3	4
Salary	1581	60392.22	14674.825	30000	51900	59500	71800	99300
Partner_salary	1581	19233.776	19670.391	0	0	25100	38100	80500
Total_salary	1581	79625.996	25545.858	30000	60500	78000	95900	171000
Price	1581	35597.723	13633.637	18000	25000	31000	47000	70000

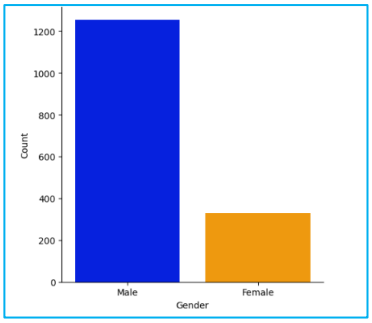
C. Explore all the features of the data separately by using appropriate visualizations and draw insights that can be utilized by the business.

Solution C

Different Visualizations and its observations are as follows

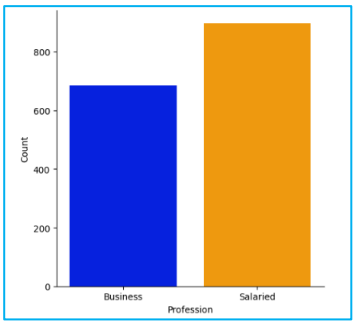
Uni-Variate Analysis

- **Bar graph of Category "Gender"**



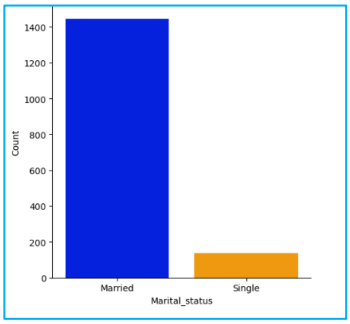
Observation: The no of males is more than the no of females in the data set provided. Which means that there are more male Buys for cars than Females.

- **Bar graph of Category "Profession"**



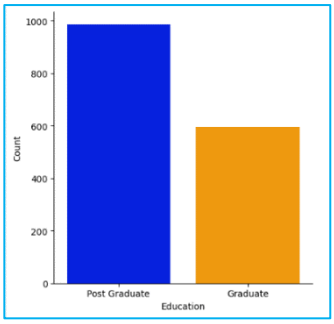
Observation : The no of Salaried professionals is more buyer than the no of Business professionals.

- **Bar graph of Category "Marital_status"**



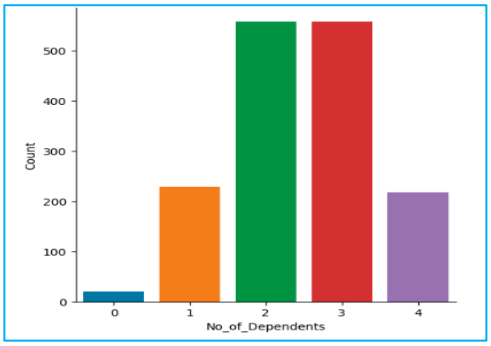
Observation : The dataset contains more than 1400 married buyers and less than 200 buyers who are single

- **Bar graph of Category "Education"**



Observation : There are more post graduates as compared to graduates in the dataset.

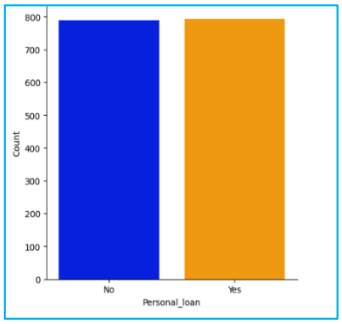
- **Bar graph of Category "No_of_Dependents"**

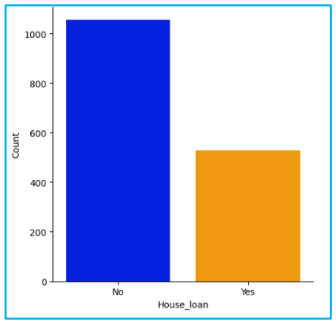


Observation :

1. There are more than 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"**

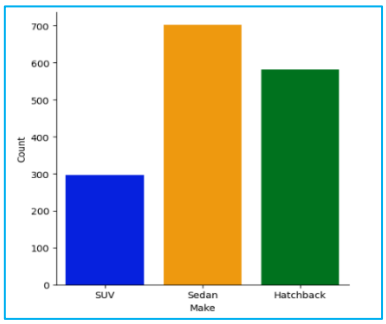




Observation :

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 "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

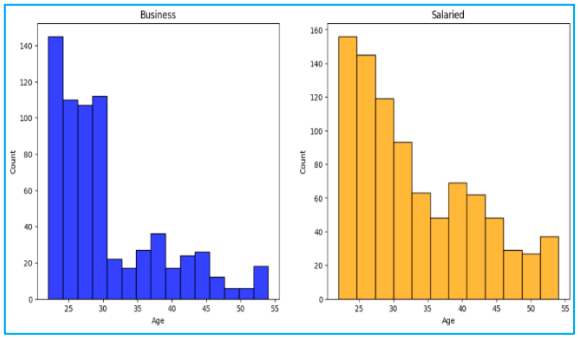
- **Pivot Table for Gender & Type of profession they are working**

Profession	Business	Salaried
Gender		
Female	105	224
Male	580	672

Observation:

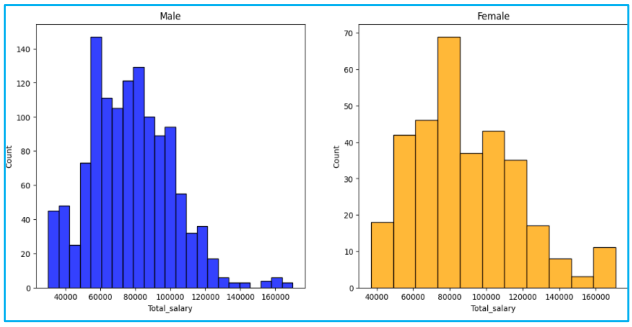
1. There are more male in the data set as compared to Females.
2. The no of salaried females is more than double the no of businesswomen.
3. The no of salaried males is more in comparison to businessmen in the data set provided

- **Histogram of Professional along with their age**



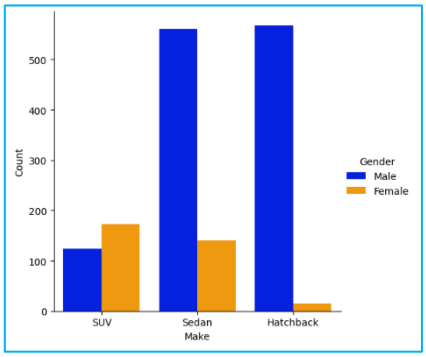
Observation : The data set contains more professional of age below 30 in both Business & Salaried category.

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



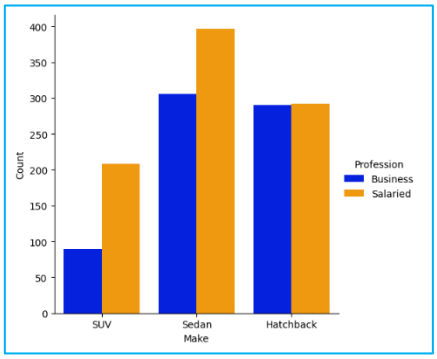
Observation : The data set contains more professional in total salary slab b/w 50000-100000 both males and females.

- **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:

1. Salaried professionals have bought more ‘SUVs’ than business professionals.
2. Salaried professionals have bought more ’Sedans’ than Business professionals
3. Both salaried & business professionals have bought almost equal no. of ’Hatchbacks’

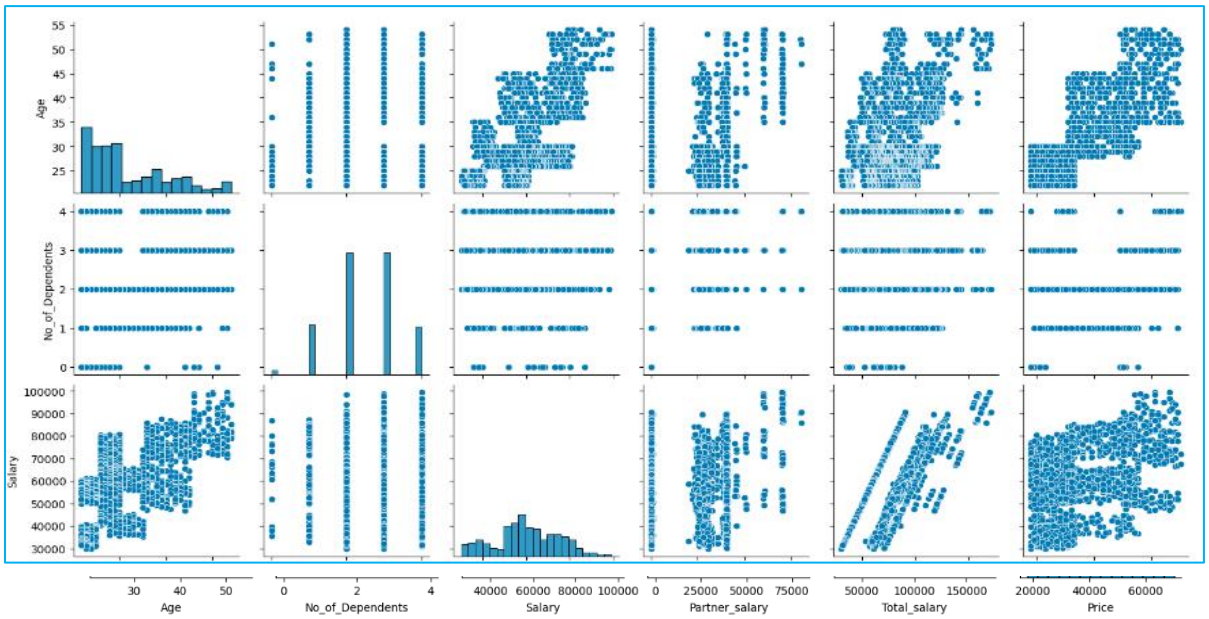
Mult i-Variate Analysis

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

No_of_Dependents	Marital_status	Make Hatchback	SUV	Sedan
0	Single	15.0	5.0	NaN
1	Married	10.0	34.0	171.0
	Single	1.0	NaN	13.0
2	Married	113.0	76.0	264.0
	Single	68.0	11.0	25.0
3	Married	256.0	117.0	184.0
4	Married	119.0	54.0	45.0

Observation: The Married people with No_of_Dependents 2 or more than 2 have mostly bought either "hatchback" or "Sedan.

- **Pair Plot for Numeric data**



Observation: The Price of cars & Age of the buyers is positively related.

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"

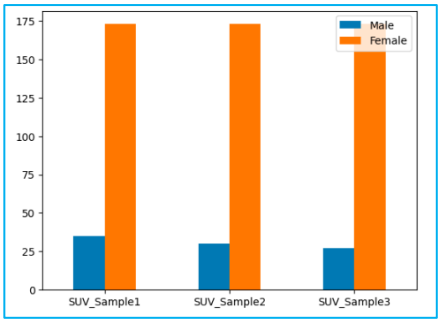
E2) Ned Stark believes that a salaried person is more likely to buy a Sedan.

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

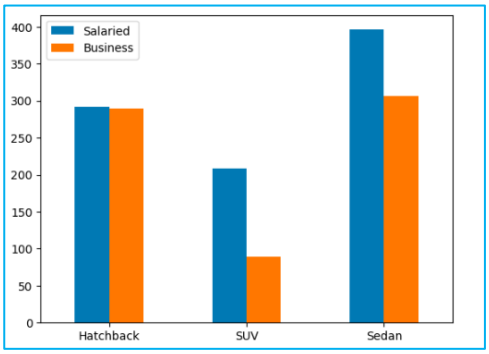
Steve Roger says “Men prefer SUV by a large margin, compared to the women:

Disagree, it is clearly visible from the bar graph that Females Prefer "SUV” ay a la e 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 claimed that a salaried male is easier target for SUV sale over Sedan sale: **Disagree,** It is clearly visible from the chart below that salaried males prefer "Sedan" much more than "the SUVs".

Profession	Gender	Make	
Business	Female	SUV	55
		Sedan	50
	Male	Hatchback	290
		SUV	34
		Sedan	256
Salaried	Female	Hatchback	15
		SUV	118
		Sedan	91
	Male	Hatchback	277
		SUV	90
		Sedan	305

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

F1) Gender

Solution F1

Gender	Make	
Female	Hatchback	412000
	SUV	9252000
	Sedan	6031000
Male	Hatchback	14996000
	SUV	7328000
	Sedan	18261000

Observation!

1. The amount spent on 'SUV' by females is more than that spend by 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

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' is more buyers having no personal loan.
2. The amount spent on 'Sedan' is more from buyers having personal loan.
3. The amount spent on 'SUV' & 'Hatchback' 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	
No	Hatchback	18000	2
		19000	9
		20000	23
		21000	2
		22000	23
			..
Yes	Sedan	51000	5
		52000	10
		53000	10
		54000	9
		55000	5

Partner_working	Make	Price	
No	SUV	39000	1
		48000	1
		32000	1
		70000	1
	Sedan	22000	1
			..
Yes	Hatchback	28000	28
		22000	30
		30000	32
		20000	35
		32000	38

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.

Soution H

Marital_status	Gender	Make	
Married	Female	Hatchback	14
		SUV	166
		Sedan	127
	Male	Hatchback	484
		SUV	115
		Sedan	537
Single	Female	Hatchback	1
		SUV	7
		Sedan	14
	Male	Hatchback	83
		SUV	9
		Sedan	24

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 '5UVs' . 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 ted 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'.