

**Problem:**

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.

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: Dataset: austo\_automobile**

Austo Motor Company is a leading car manufacturer specializing in SUV, Sedan, and Hatchback models.

- The dataset contains 1581 rows and 14 columns.
- Dataset contains columns having data type int, float, and object as follows:

S.NO.	COLUMN NAME	DATA TYPE
1	Age	INT64
2	No_of_Dependents	INT64
3	Salary	INT64
4	Total_salary	INT64
5	Price	INT64
6	Partner_salary	FLOAT64
7	Gender	OBJECT
8	Profession	OBJECT
9	Marital_status	OBJECT
10	Education	OBJECT
11	House_loan	OBJECT
12	Partner_working	OBJECT
13	Make	OBJECT
14	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? If yes, perform preliminary treatment of data

**Solution B:** After taking a look on the dataset:

- The columns having datatype “object” has unique values:

S.NO.	UNIQUE VALUES	COLUMN NAME	DATA TYPE
1	'Business', 'Salaried'	Profession	OBJECT
2	'Married', 'Single'	Marital_status	OBJECT

3	'Post Graduate', 'Graduate'	Education	OBJECT
4	'No', 'Yes'	Personal loan	OBJECT
5	'No', 'Yes'	House_loan	OBJECT
6	'No', 'Yes'	Partner_working	OBJECT
7	'SUV', 'Sedan', 'Hatchback'	Make	OBJECT
8	'Male', 'Femal', 'Female', nan, 'Femle'	Gender	OBJECT

- Two columns has missing values.

S.NO.	COLUMN NAME	NO.OF MISSING VALUES
1	Gender	53
2	Partner salary	106

- Two discrepancies(spelling mistakes) was seen in Gender column:  
Femle, Femal

```
### Replace the wrong words with correct words
df.Gender=df.Gender.replace(["Femal","Femle"],["Female","Female"])
```

```
df.Gender.unique()
```

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

- Total no. of males =1199, females=329, and null values were 53

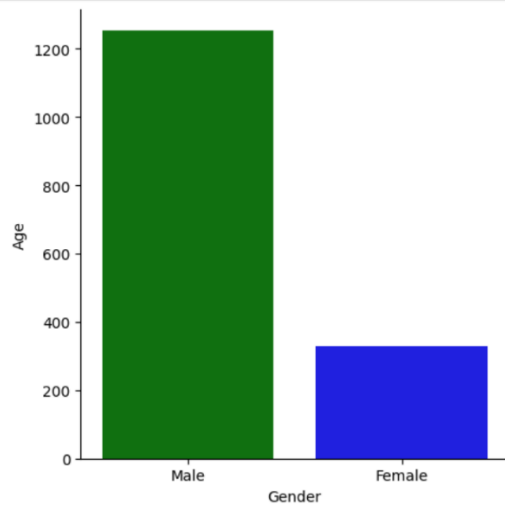
The whole description of the dataset :

Columns	count	mean	Std.	Min	max	25%	50%	75%
Age	1581	31.92	8.42	22	54	25	29	38
Salary	1581	60392.22	14674.82	30,000	99,300	51,900	59,500	71,800
Partner_salary	1475	20225.55	19573.14	0	80,500	0	25,600	38,300
Total_salary	1581	79625.99	25545.85	30,000	1,71,000	60,500	78,000	95,900
Price	1581	35597.72	13633.63	18,000	70,000	25,000	31,000	47,000

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

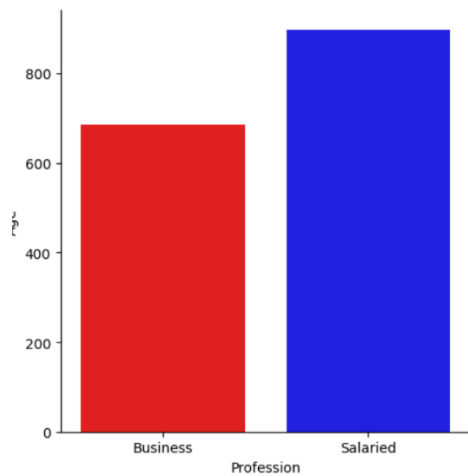
### **Solution C:**

The visuals & observations are as:



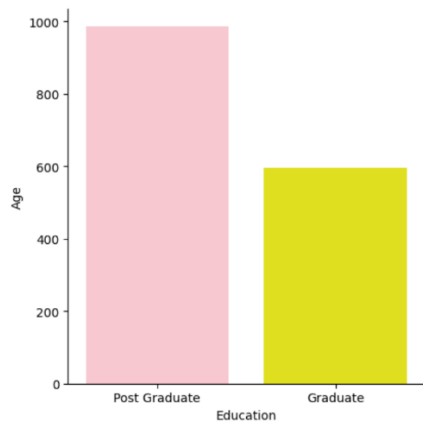
BAR GRAPH OF CATEGORY "GENDER"

**OBSERVATION:** Males are large in number as compared with the females, means more male buyers than females.



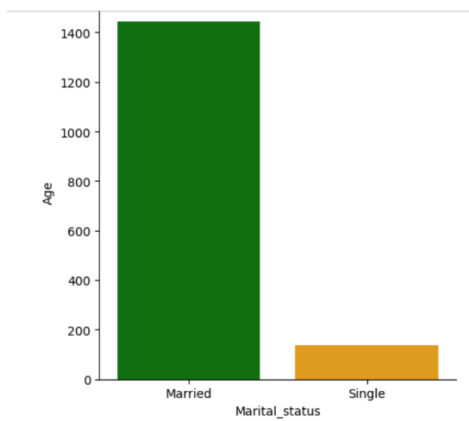
BAR GRAPH OF CATEGORY "PROFESSION"

**OBSERVATION:** The no. of salaried people are more than business class in the dataset.



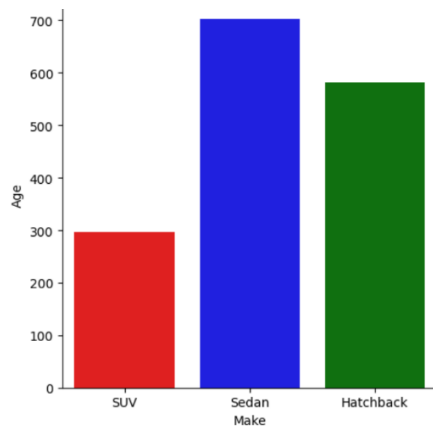
BAR GRAPH CATEGORY "EDUCATION"

**OBSERVATION:** The postgraduates are large in no. than the graduates.



BAR GRAPH CATEGORY  
"MARTIAL\_STATIUS"

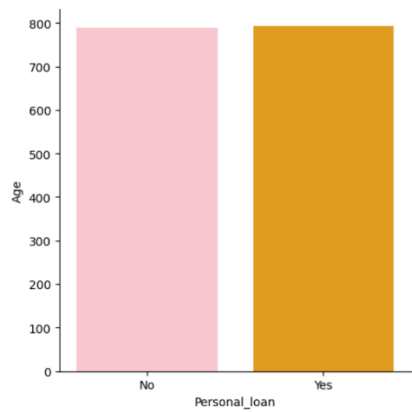
**OBSERVATION:** The no. of males =1400 and females no.is less than 200.



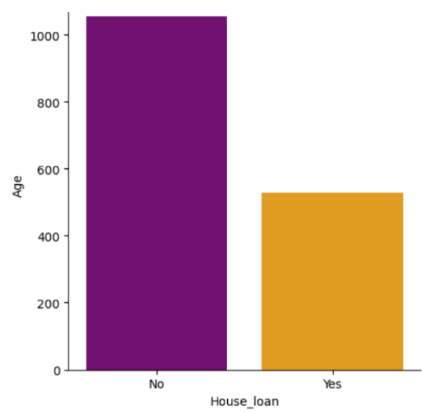
BAR GRAPH CATEGOY “ MAKE”

**OBSERVATION:**

- 1. Highest selling car is SEDAN.
- 2. Lowest selling car is SUV.



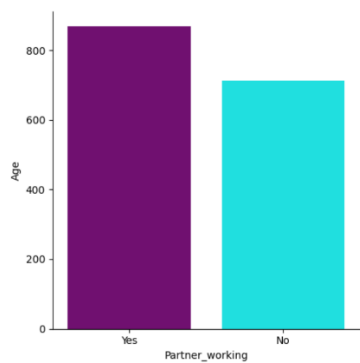
BAR GRAPH CATEGORY “PERSONAL\_LOAN”



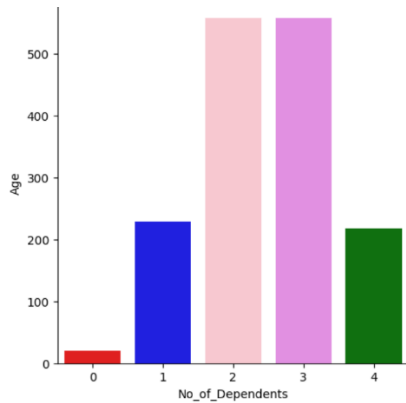
BAR GRAPH CATEGORY “HOUSE\_LOAN”

**OBSERVATION:**

- 1. The no. of people having “personal loan” and not having “personal loan” is almost equal.
- 2. The no. of people had “house loan” is almost half of people not having “house loan”.



BAR GRAPH CATEGORY “PARTNER\_WORKING”



BAR GRAPH CATEGORY "NO. OF DEPENDENTS"

### **OBSERVATION:**

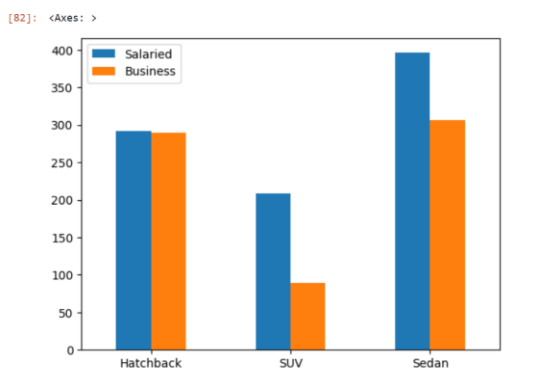
1. The bar height shows that people having dependents 2 & 3 are more than 500.
2. Dataset shows that people having dependents 1 & 4 are more than 200.
3. The bar shows that people having dependents 0 are less than 50.

**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:**

### **OBSERVATIONS:**

1. Salaried class people is more interested in SUV'S and SEDAN as compared with HATCHBACK.



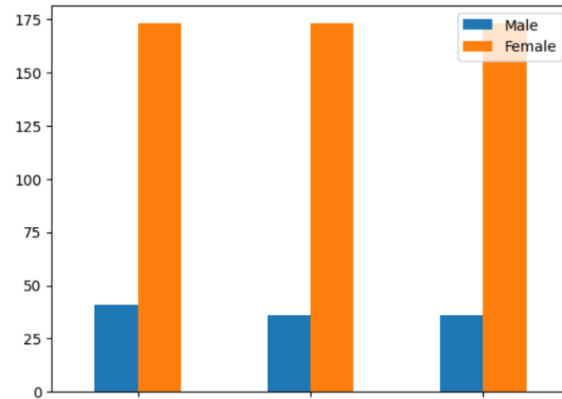
**E.** Employees working on the existing marketing campaign have made the following remarks. Based on the data and your analysis state whether you agree or disagree with their observations. Justify your answer Based on the data available.

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.

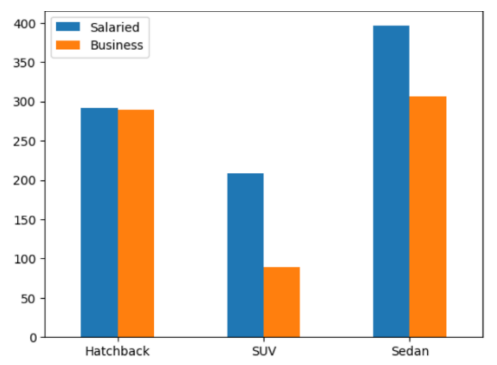
E3) Sheldon Cooper does not believe any of them; he claims that a salaried male is an easier target for a SUV sale over a Sedan Sale

### **SOLUTION E1:**



**OBSERVATION:** Female buyers are interested in SUV'S as compared with male buyers.

### **SOLUTION E2:**



**OBSERVATION:** Yes, I agree with Ned Stark. It is clearly shown by the graph.

### **SOLUTION E3:**

```
Sal_Male=df.where((df["Profession"]=="Salaried")&(df.Gender=="Male")&(df.Make=="SUV")).dropna()[["Profession","Make"]]  
Sal_Female=df.where((df["Profession"]=="Salaried")&(df.Gender=="Female")).dropna()[["Profession","Make"]]
```

```
8]: Sal_Male
```

```
8]:
```

	Profession	Make
4	Salaried	SUV
7	Salaried	SUV
8	Salaried	SUV
9	Salaried	SUV
10	Salaried	SUV
...	...	...
480	Salaried	SUV
493	Salaried	SUV
496	Salaried	SUV
509	Salaried	SUV
511	Salaried	SUV

90 rows × 2 columns

```
Sal_Male=df.where((df["Profession"]=="Salaried")&(df.Gender=="Male")&(df.Make=="Sedan")).dropna()[["Profession","Make"]]
```

```
Sal_Male
```

```
:
```

	Profession	Make
144	Salaried	Sedan
149	Salaried	Sedan
152	Salaried	Sedan
159	Salaried	Sedan
163	Salaried	Sedan
...	...	...
1539	Salaried	Sedan
1547	Salaried	Sedan
1557	Salaried	Sedan
1568	Salaried	Sedan
1574	Salaried	Sedan

305 rows × 2 columns

**OBSERVATION:** No, I disagree with Sheldon Cooper.

**E.** 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.

```
male=df.where((df["Gender"]=="Male")).dropna()[["Make","Price"]].groupby("Make").sum()
```

```
male
```

	Price
Make	
Hatchback	14996000.0
SUV	7328000.0
Sedan	18261000.0



```
female=df.where((df["Gender"]=="Female")).dropna()[["Make","Price"]].groupby("Make").sum()
```

female	
	Price
Hatchback	412000.0
SUV	9252000.0
Sedan	6031000.0

### **OBSERVATIONS:**

1. The amount spend by males on 'HATCHBACK' is greater than three times the amount Spend by females.
2. Amount spend on SUV'S is almost equal.
3. Money spend for SEDAN'S by males is almost three times higher than females.

### **F2) Personal\_loan.**

```
yes=df.where((df["Personal_loan"]=="Yes")).dropna()[["Make","Price"]].groupby("Make").sum()
```

yes	
	Price
Hatchback	7643000.0
SUV	6207000.0
Sedan	13440000.0

```
no=df.where((df["Personal_loan"]=="No")).dropna()[["Make","Price"]].groupby("Make").sum()
```

no	
	Price
Hatchback	7765000.0
SUV	10373000.0
Sedan	10852000.0

### **OBSERVATIONS:**

1. Amount spend on 'HATCHBACK' is almost equal.
2. Larger amount spend on 'SUV' is by people not having personal loan.
3. Amount spend on 'SEDAN' is largely done by people 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:**

```
]:
```

```
1]: df.groupby(['Partner_working', 'Make', 'Gender', 'Price'])['Price'].count()
```

```
1]:
```

Partner_working	Make	Gender	Price	
No	Hatchback	Female	20000	1
			25000	1
			29000	1
			30000	1
			32000	2
			..	
Yes	Sedan	Male	51000	3
			52000	5
			53000	5
			54000	6
			55000	3

Name: Price, Length: 257, dtype: int64

```
]:
```

```
[147]: Partner_working=df.where((df["Partner_working"]=="Yes")).dropna()[["Price","Make"].groupby("Price").count()
```

```
[146]: Partner_working
```

[146]:

	Make
Price	
18000.0	21
19000.0	28
20000.0	47
21000.0	23
22000.0	31
23000.0	36
24000.0	35
25000.0	20
26000.0	32
27000.0	33
28000.0	37

---

28000.0	37
---------	----

29000.0	30
---------	----

30000.0	32
---------	----

31000.0	40
---------	----

32000.0	48
---------	----

33000.0	32
---------	----

34000.0	9
---------	---

35000.0	8
---------	---

36000.0	11
---------	----

37000.0	13
---------	----

38000.0	8
---------	---

39000.0	17
---------	----

40000.0	5
---------	---

41000.0	10
---------	----

42000.0	5
---------	---

43000.0	20
---------	----

42000.0 20

43000.0 20

44000.0 6

45000.0 11

46000.0 11

47000.0 8

48000.0 13

49000.0 12

50000.0 20

51000.0 16

52000.0 16

53000.0 14

54000.0 12

55000.0 14

56000.0 5

57000.0 11

58000.0 5

---

58000.0	5
59000.0	4
60000.0	4
61000.0	11
62000.0	6
63000.0	5
64000.0	5
65000.0	3
66000.0	7
67000.0	4
68000.0	6
69000.0	7
70000.0	1

**OBSERVATION:** Yes, based on the above results it's true that working partners tends to purchase 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.

**SOLUTION H:**

```
Married_Males = df.where((df["Marital_status"]=="Married") & (df["Gender"]=="Male")).dropna()[["Marital_status","Make"]].groupby("Make").count()

Married_Males

Marital_status
Make
Hatchback    484
SUV          115
Sedan        537

Married_females = df.where((df["Marital_status"]=="Married") & (df["Gender"]=="Female")).dropna()[["Marital_status","Make"]].groupby("Make").count()

Married_females

Marital_status
Make
Hatchback    14
SUV          166
Sedan        127

7]: Single_females = df.where((df["Marital_status"]=="Single") & (df["Gender"]=="Female")).dropna()[["Marital_status","Make"]].groupby("Make").count()
8]: Single_females

8]: Marital_status
Make
Hatchback    1
SUV          7
Sedan        14

9]: Single_Males = df.where((df["Marital_status"]=="Single") & (df["Gender"]=="Male")).dropna()[["Marital_status","Make"]].groupby("Make").count()
0]: Single_Males

0]: Marital_status
Make
Hatchback    83
SUV          9
Sedan        24
```

## OBSERVATIONS:

1. Married males buys SEDAN & HATCHBACK so some discounts or offer scheme should be given on SUV's.
2. Single males are more interested in HATCHBACK. Marketing strategy should be devised for targeting single males.
3. Married females should be targeted to increase the sales of SUV's & SEDAN.