

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)

```
[8]: file.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1581 entries, 0 to 1580
Data columns (total 14 columns):
 #   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
dtypes: float64(1), int64(5), object(8)
memory usage: 173.1+ KB
```

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.

1. Unique value in columns

```
{'Gender': array(['Male', 'Femal', 'Female', nan, 'Femle'], dtype=object),
 'Profession': array(['Business', 'Salaried'], dtype=object),
 'Marital_status': array(['Married', 'Single'], dtype=object),
 'Education': array(['Post Graduate', 'Graduate'], dtype=object),
 'No_of_Dependents': array([4, 3, 2, 1, 0], dtype=int64),
 'Personal_loan': array(['No', 'Yes'], dtype=object),
 'House_loan': array(['No', 'Yes'], dtype=object),
 'Partner_working': array(['Yes', 'No'], dtype=object),
 'Make': array(['SUV', 'Sedan', 'Hatchback'], dtype=object)}
```

2. In “Gender” column Female spelling is incorrect.

```
file.Gender.unique()
```

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

```
file.Gender = file.Gender.replace(["Femal","Femle"],["Female","Female"])
```

```
file .Gender.unique()
```

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

3. Missing values in columns

```
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
```

4. To handle missing values in “Gender” column.

```
file.groupby("Gender")["Gender"].count()
```

```
Gender
Female    329
Male     1199
```

4.1. In “Gender” column number of Males is higher than Females, So we can assume that in the missing value number of Males is higher than Females, that’s why we can fill the missing value with Male.

```
file.Gender = file.Gender.fillna("Male")
```

```
file.Gender.unique()
```

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

5. To handle missing values in “Partner_salary” column.

5.1. First find the relationship between “Salary”, “Partner_salary” and “Total_salary”

```
sel = file[file.Partner_salary.notnull()][["Salary", "Partner_salary", "Total_salary"]]
```

```
sel
```

	Salary	Partner_salary	Total_salary
0	99300	70700.0	170000
1	95500	70300.0	165800
2	97300	60700.0	158000
3	72500	70300.0	142800
4	79700	60200.0	139900
...
1574	34700	0.0	34700
1575	34000	0.0	34000
1576	33300	0.0	33300
1578	32900	0.0	32900
1580	31600	0.0	31600

1475 rows × 3 columns

```
sel["isTotal"] = sel.Salary + sel.Partner_salary == sel.Total_salary
```

```
sel.isTotal.unique()
```

```
array([ True])
```

5.2. Filling missing values in “Partner_salary”

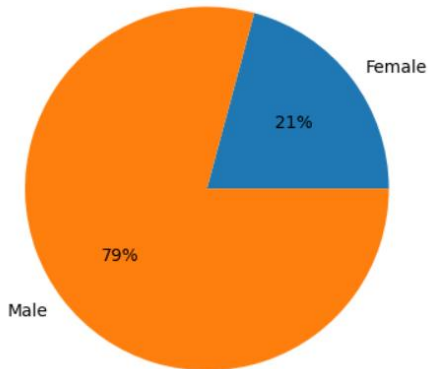
```
file.Partner_salary = file.apply(lambda row:(row["Total_salary"]-row["Salary"]) if np.isnan(row["Partner_salary"]) else row["Partner_salary"], axis=1)

file.Partner_salary.isnull().sum()

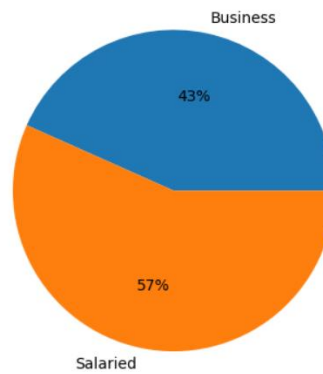
0
```

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

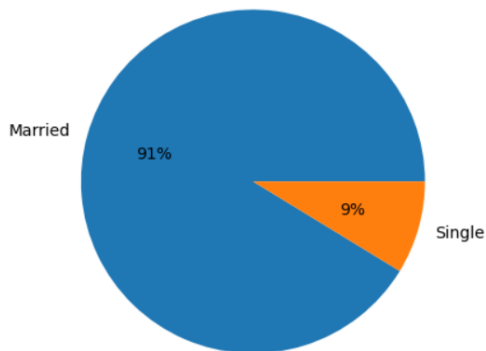
Gender



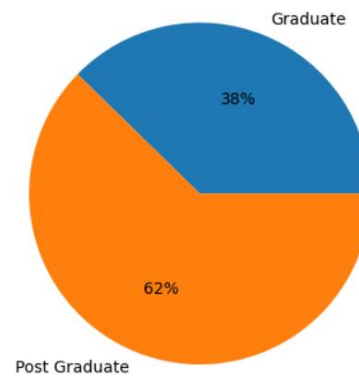
Profession



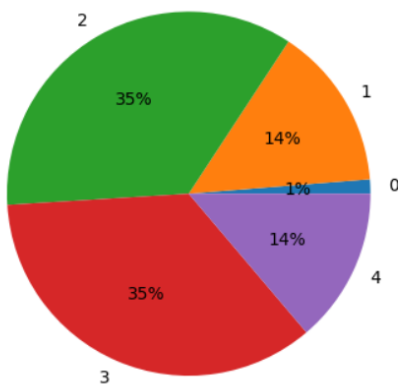
Marital Status



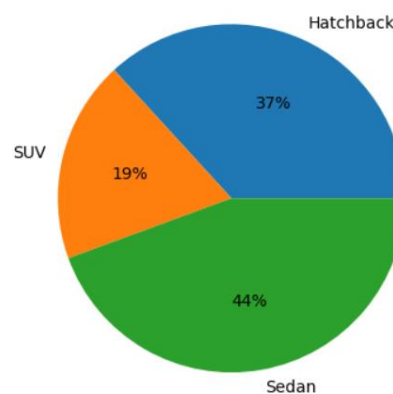
Education



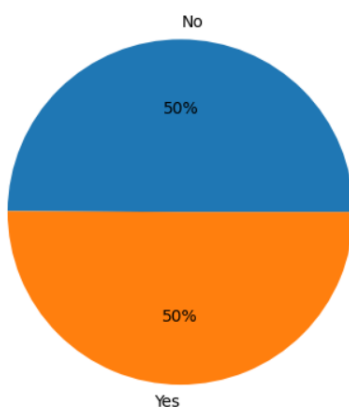
Number of Dependents



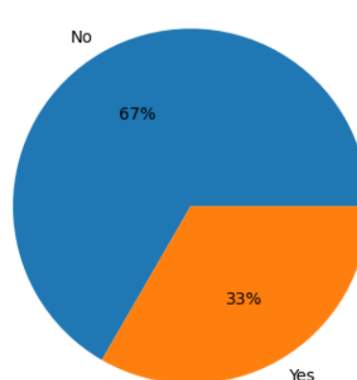
Make



Personal Loan



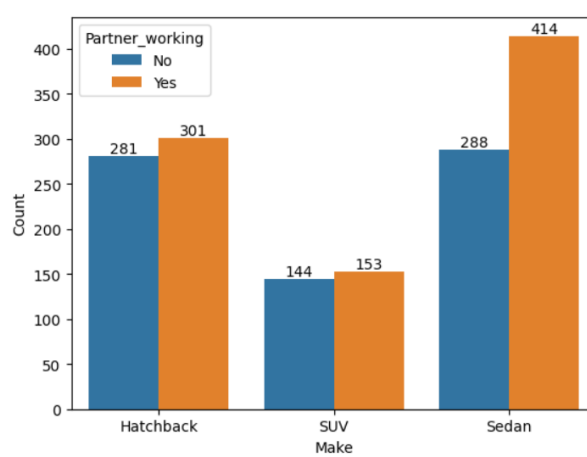
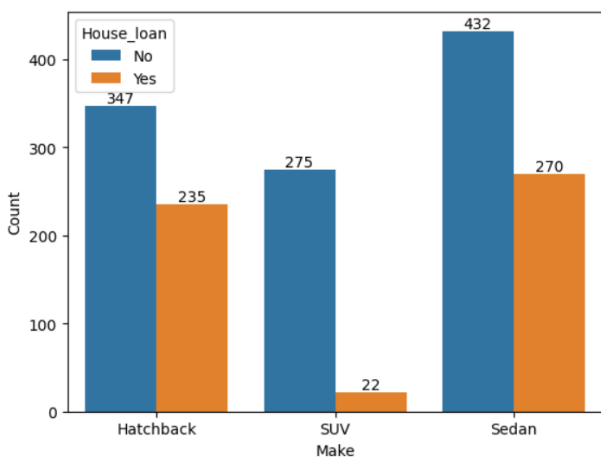
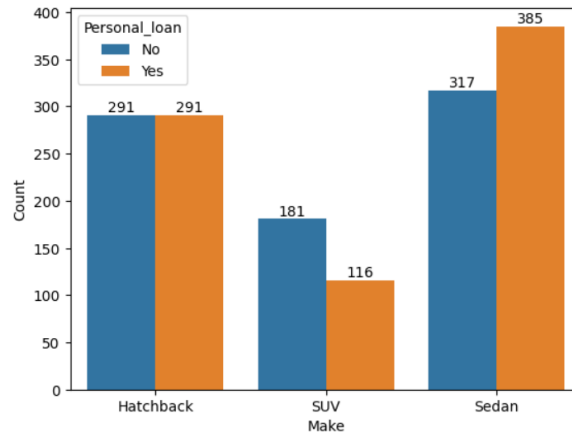
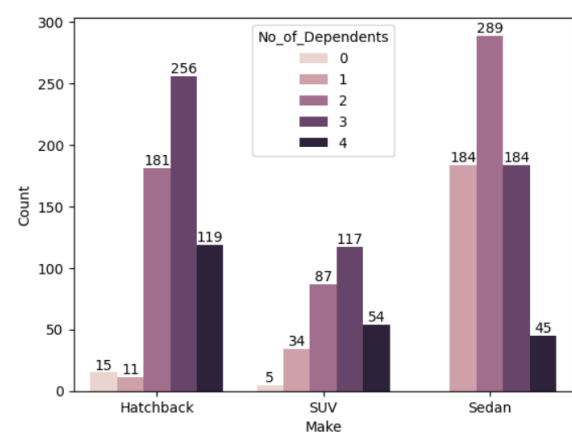
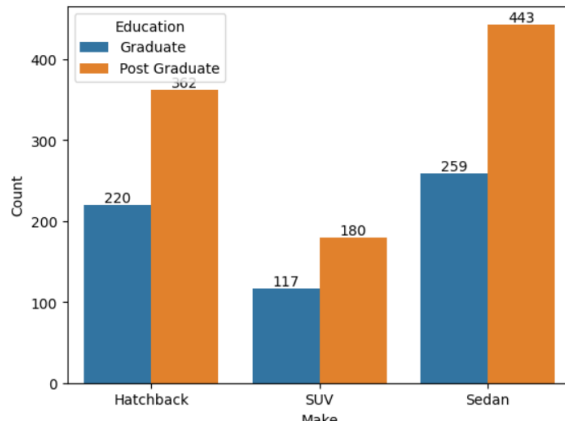
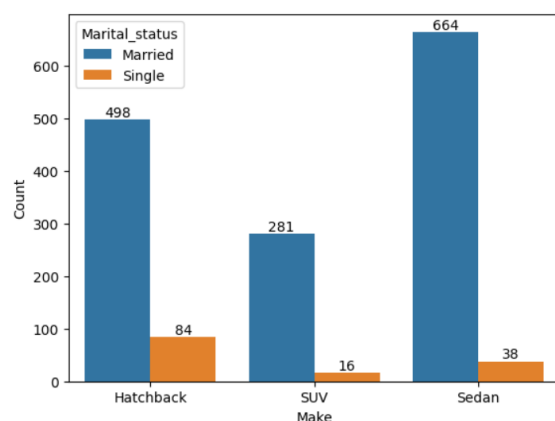
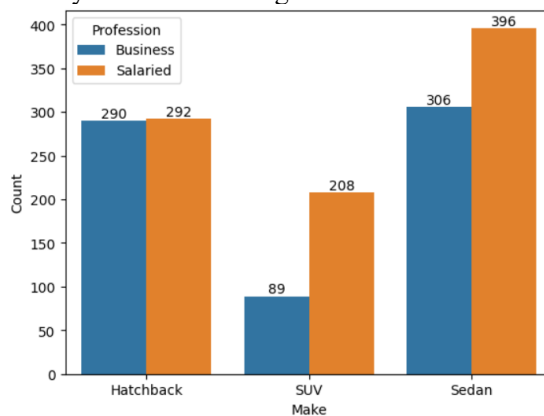
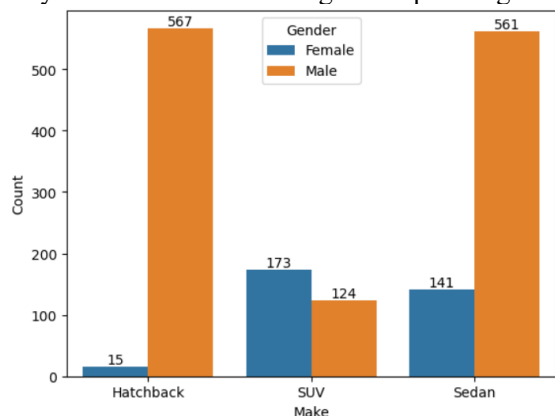
House Loan



From the above visualization we can say that –

- 1 Male used to buy more cars than Female.
- 2 Salaried people used to buy more cars than Business person.
- 3 Married people used to buy more cars than Single.
- 4 Post-graduates people buy more cars than Graduates.
- 5 Most people like Sedan car.
- 6 Most people don't want to buy a car using House Lone.

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.

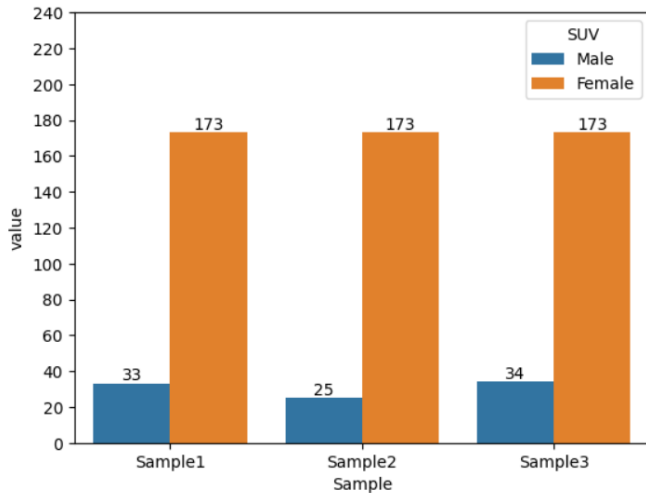


From the above visualization we can say that –

- 1 Male used to buy Hatchback while Female prefer SUV.
- 2 Salaried and Business people used to buy Sedan.
- 3 Married people used to buy Sedan while Single people prefer Hatchback.
- 4 Post Graduate and Graduate used to buy Sedan.

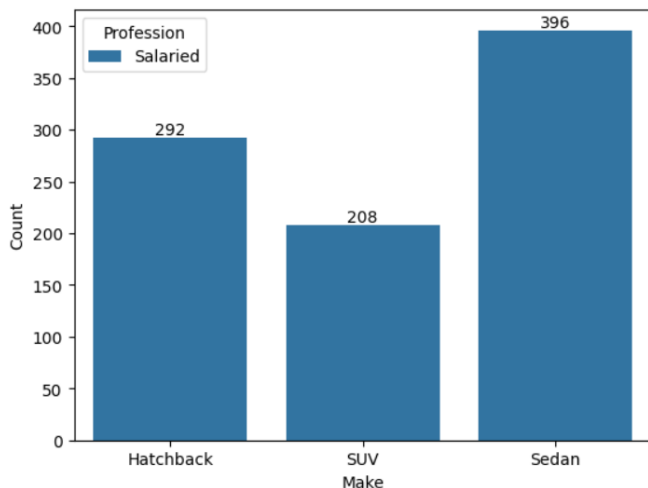
E. Employees working on the existing marketing campaign have made the following remarks. Based on the data and your analysis state whether you agree or disagree with their observations. Justify your answer Based on the data available.

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



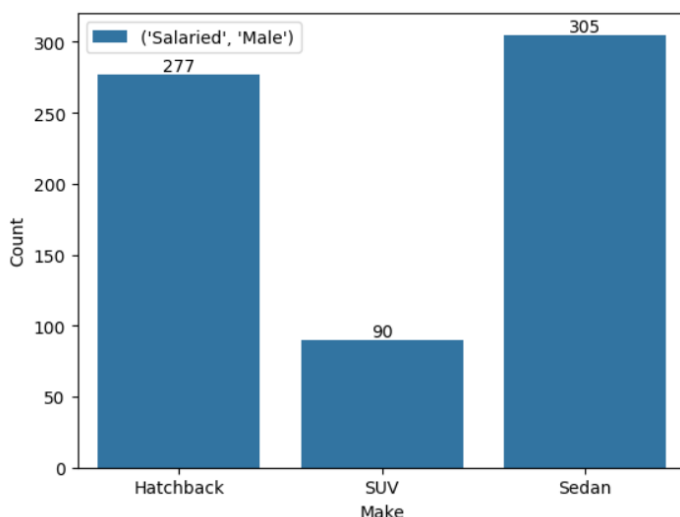
Disagree, From the above visualization we can say that Female prefer SUV by large margin as compared to mail.

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



Agree, we can clearly see that in the above visualization, 396 Salaried persons buy Sedan.

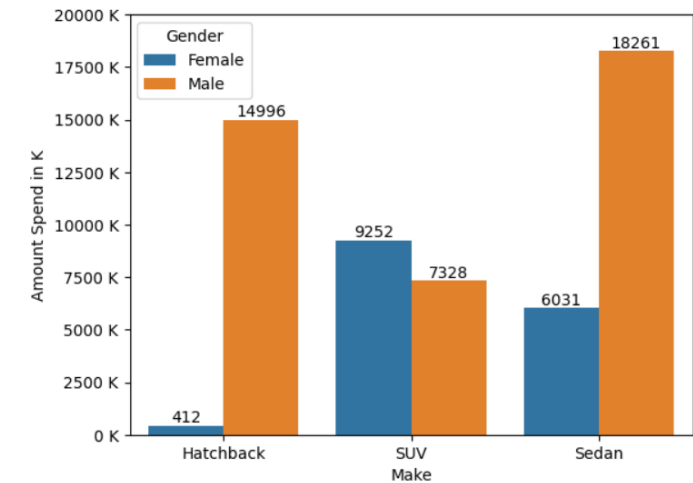
E3) Sheldon Cooper does not believe any of them; he claims that a salaried male is an easier target for a SUV sale over a Sedan Sale.



Disagree, we can clearly see that in the above visualization, 305 Salaried Male persons buy Sedan over SUV.

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.

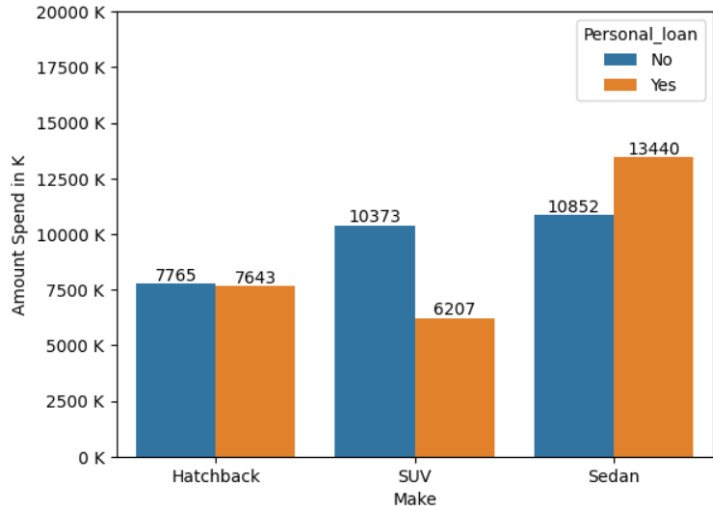
F1) Gender



Observation:

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

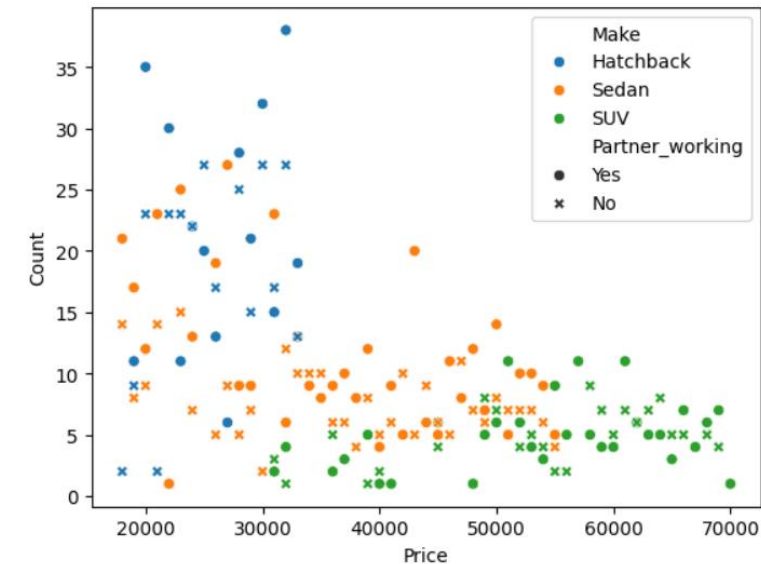
F2) Personal_loan



Observation:

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



Partner_working		Make	Price	Count
94	Yes	Hatchback	32000	38
83	Yes	Hatchback	20000	35
92	Yes	Hatchback	30000	32
84	Yes	Hatchback	22000	30
90	Yes	Hatchback	28000	28
12	No	Hatchback	30000	27
14	No	Hatchback	32000	27
135	Yes	Sedan	27000	27
7	No	Hatchback	25000	27
10	No	Hatchback	28000	25

Partner_working		Make	Price	Count
44	No	SUV	70000	1
49	No	Sedan	22000	1
131	Yes	Sedan	22000	1
22	No	SUV	48000	1
101	Yes	SUV	40000	1
19	No	SUV	39000	1
17	No	SUV	32000	1
102	Yes	SUV	41000	1
104	Yes	SUV	48000	1
126	Yes	SUV	70000	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.

	Gender	Make	Min Age	Max Age	Avg Age	count
0	Female	Hatchback	22	30	26.200000	15
1	Female	SUV	28	54	43.023121	173
2	Female	Sedan	28	45	36.652482	141
3	Male	Hatchback	22	30	25.714286	567
4	Male	SUV	35	54	45.338710	124
5	Male	Sedan	22	45	30.771836	561

Observation:

I. Most 43-year-old female wants to buy SUV.

2. Most 25-year-old male wants to buy Hatchback.

3. Most 30-year-old male wants to buy Sedan.