DATA ANALYSIS FOR AUSTO MOTOR COMPANY

- **BY PAPPOPPULA AKHILESH**

- **20-04-202**

Table of Contents

LIST OF FIGURES

PROBLEM STATEMENT

Austo Motor Company is a leading car manufacturer that specializes in **SUV, sedan, and hatchback** models. During a recent board meeting, the members expressed concerns about the effectiveness of the current marketing campaign.

OBJECTIVE

Analysing the data to get a fair Idea about the demand of customers to enhance their customer experience. Performing the data analysis to answer the key questions, which will help the company improve its business.

KEY QUESTIONS

1. Do men tend to prefer SUVs more compared to women?
2. What is the likelihood of a salaried person buying a Sedan?
3. What evidence or data supports Sheldon Cooper's claim that a salaried male is an easier target for an SUV sale over a Sedan sale?
4. How does the amount spent on purchasing automobiles vary by gender?
5. How much money was spent on purchasing automobiles by individuals who took a personal loan?
6. How does having a working partner influence the purchase of higher-priced cars?

DATA DESCRIPTION

1. **Age**: The age of the individual in years.
2. **Gender**: The individual's gender, categorized as male or female.
3. **Profession**: The occupation or profession of the individual.
4. **Marital status**: The marital status of the individual, such as married &, single
5. **Education**: The educational qualification of the individual is a Graduate and postgraduate
6. **No of Dependents**: The number of dependents (e.g., children, elderly parents) that the individual supports financially.
7. **Personal loan**: A binary variable indicating whether the individual has taken a personal loan (Yes or No)
8. **House loan**: A binary variable indicating whether the individual has taken a housing loan (Yes or No)
9. **Partner working**: A binary variable indicating whether the individual's partner is employed (Yes or No)
10. **Salary**: The individual's salary or income.
11. **Partner salary**: The salary or income of the individual's partner, if applicable.

12. **Total salary**: The total combined salary of the individual and their partner (if applicable).
13. **Price**: The price of a product or service.
14. **Make**: The type of automobile

UNDERSTANDING THE STRUCTURE OF THE DATA

```
<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
```

Figure 1 (Structure of data)

- The data set contains 1581 rows and 14 columns
- There are no duplicate values in the dataset
- Every column has 1581 Values except the Gender and Partner working columns, which indicates that there are a few missing values in both columns
- The Partner salary column is read as a float64 data type it can be changed to int64 data type
- The personal loan, House loan, & Partner working column reads as objects. It is better to turn them into 0 and 1 for better statistical analysis.
- The object type columns contain categories in them
- The int64 type columns contain numerical values.

IDENTIFYING AND TREATING MISSING VALUES

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

Figure 2 (Data with missing values)

- There are missing values only in the Gender and Partner working columns, with 53 and 106 values, respectively.
- There are entries with (Femal, Femle) in the Gender column which are likely typo errors for Female, which are corrected.
- The missing values in the Gender column were addressed by imputing the mode of that column.
- Replaced the missing values in the Partner salary column for non-working partners with the value zero.
- Replaced the missing values in the partner salary column for working partners by subtracting the individual's salary from the total salary.
- All the missing values and irregularities in the data have been addressed.

STATISTICAL SUMMARY OF THE DATA

	count	mean	std	min	25%	50%	75%	max
Age	1581.0	31.92	8.43	22.0	25.0	29.0	38.0	54.0
No_of_Dependents	1581.0	2.46	0.94	0.0	2.0	2.0	3.0	4.0
Personal_loan	1581.0	0.50	0.50	0.0	0.0	1.0	1.0	1.0
House_loan	1581.0	0.33	0.47	0.0	0.0	0.0	1.0	1.0
Partner_working	1581.0	0.55	0.50	0.0	0.0	1.0	1.0	1.0
Salary	1581.0	60392.22	14674.83	30000.0	51900.0	59500.0	71800.0	99300.0
Partner_salary	1581.0	19233.78	19670.39	0.0	0.0	25100.0	38100.0	80500.0
Total_salary	1581.0	79626.00	25545.86	30000.0	60500.0	78000.0	95900.0	171000.0
Price	1581.0	35597.72	13633.64	18000.0	25000.0	31000.0	47000.0	70000.0

Figure 3 (Statistical summary of the data)

1. The average age of an individual is approximately 31.9 years, with a minimum age of 22 and a maximum age of 54.
2. On average, individuals have about 2.5 dependents with a maximum of 4.
3. 50 % of Individuals have personal Loans, while 33% of individuals have a house loan.
4. 55% of individuals have a working partner.
5. The average salary is approximately 60,392, ranging from 30,000 to 99,300.
6. The average Partner salary is approximately 19,234, ranging up to 80,500.
7. The average total salary (Individual + Partner) is approximately 35,598, ranging from 18,000 to 17,000.
8. The average price of the product is approximately 35,598, ranging from 18,000 to 70,000.

UNIVARIATE ANALYSIS

a) Observation on Age

- The histogram shows a high count of younger individuals in their early 20s with a peak around the 20-25 age range.
- The boxplot indicates that the distribution is skewed towards the right, indicating the large number of younger individuals.

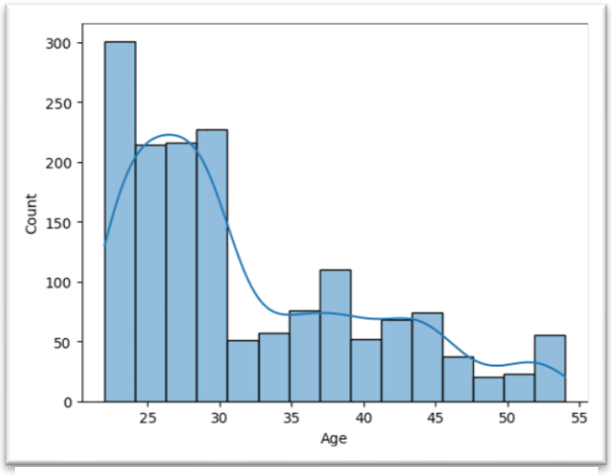

Figure 4 (Histogram of Age)

b) Observation on No of Dependents

- The histogram suggests that there is a significant portion of individuals who have 2 or 3 dependents, with fewer individuals having either zero or a single dependent
- The highest counts are observed for 2 and 3 dependents, with over 500 individuals in each category.

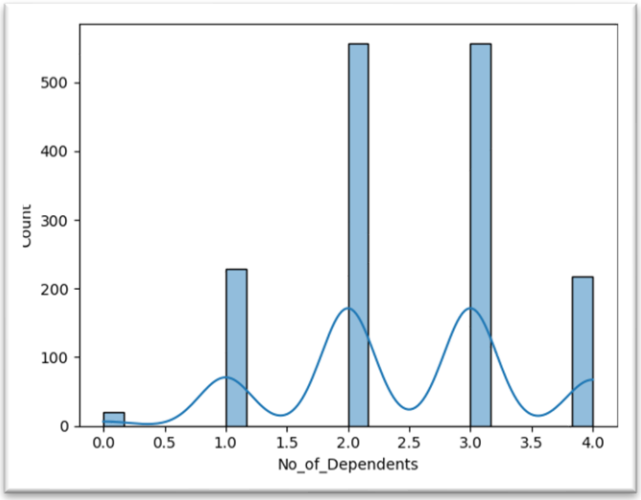

Figure 5 (Histogram of Number of Dependents)

c) Observation on Salary

- The distribution is symmetrical.
- The median salary appears to be around 60,000.
- The minimum salary is around 30,000, and the maximum is around 100000.
- There are no outliers in this column.

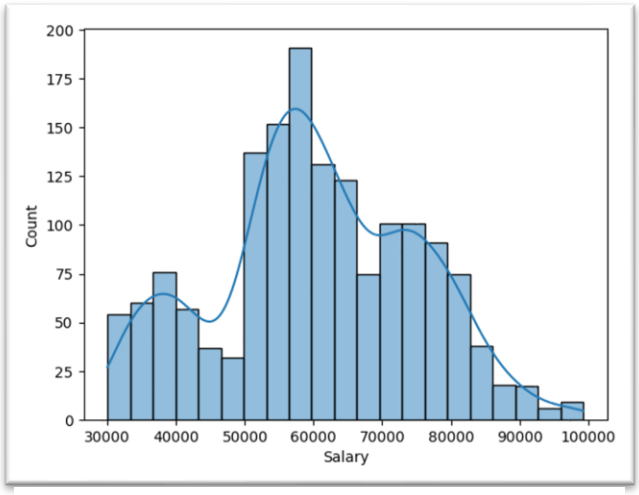

Figure 6 (Histogram of Salary)

d) Observation on Partner Salary

- The distribution appears to be right-skewed.
- The whisker extends to approximately 80,000, indicating the maximum salary in the data set, and the minimum salary is 0.
- There are no outliers in this column.

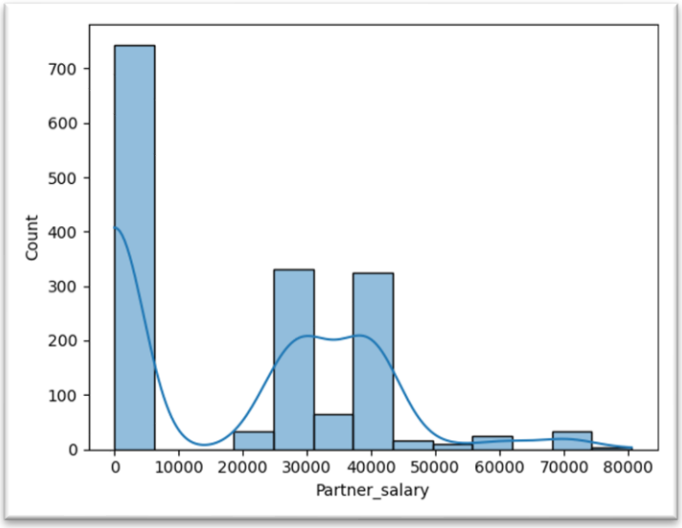

Figure 7 (Histogram of Partner Salary)

e) Observation on Total Salary

- The distribution is skewed towards the right side.
- There are many outliers present above the upper whisker, indicating some of the Individuals total salary is significantly higher.
- Most of the Individuals total salary falls between $60,000 to $90,000

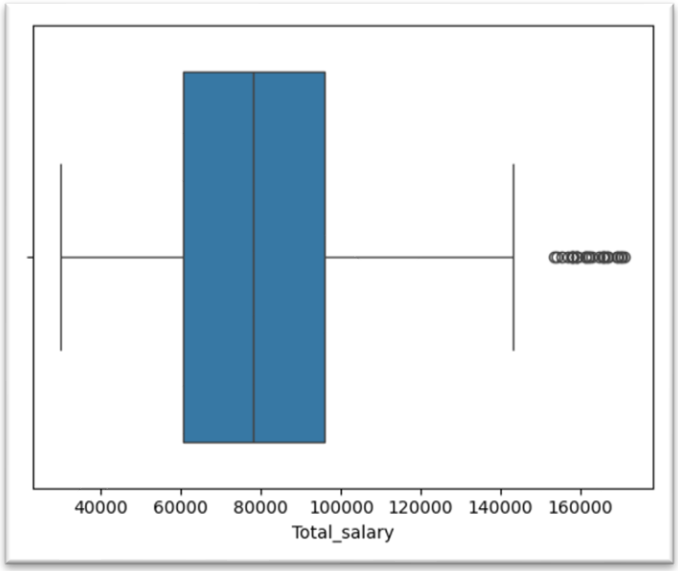

Figure 8 (Boxplot of Total salary)

f) Observation on the price

- The distribution is skewed towards the right side.
- The minimum price of a product is 18000, and the maximum price is 70000.
- There are no outliers present in this column.

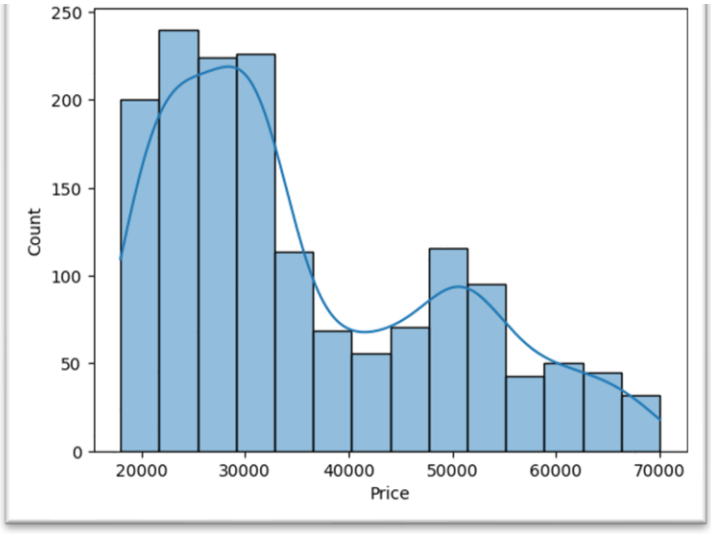

Figure 9 (Histogram of Price)

g) Observation on Gender

- The dataset contains significantly more males than females, with 1,252 males and 329 females.

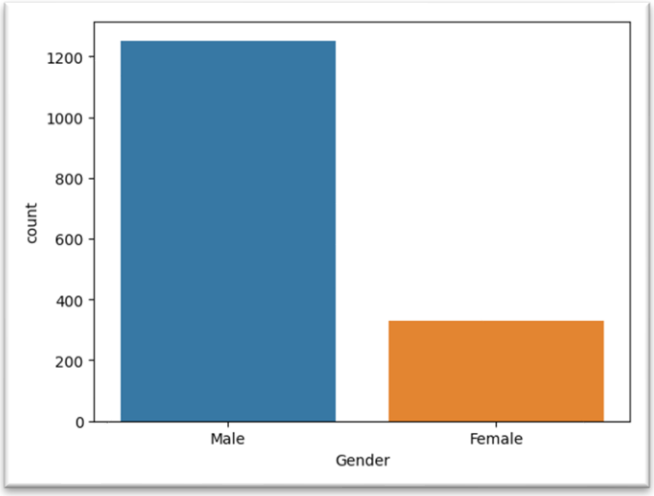

Figure 10 (Count plot of Gender)

h) Observation on Profession

- The chart indicates that individuals in salaried professions outnumber those in business professions.
- Number of Business profession Individuals: 685
- Number of Salaried professional Individuals: 896

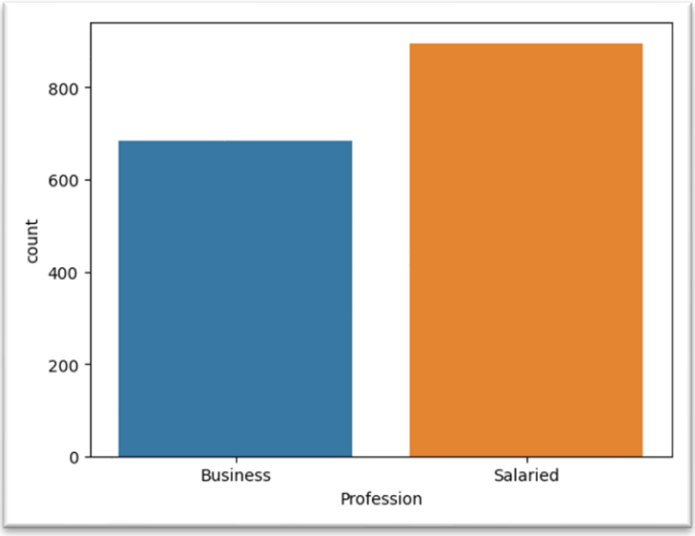

Figure 11 (Count plot on Profession)

BIVARIATE ANALYSIS

a) Correlation by Heat map for all numerical variable columns

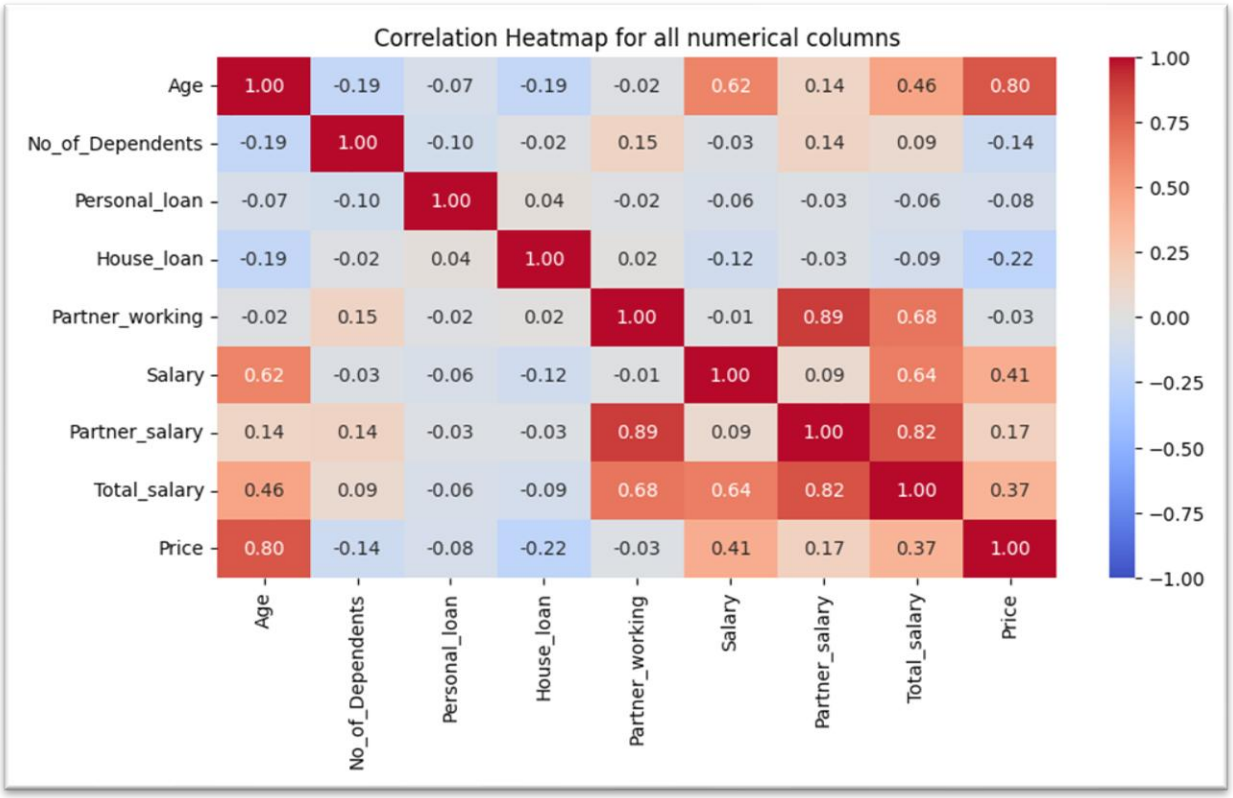

Figure 12 (Correlation Heat map)

- There is a strong positive correlation between Age and Price (0.80), which means older Individuals tend to have high-priced products.
- The Age has a moderate to strong correlation with Salary and Total Salary, which means older Individuals tend to have higher salaries and higher total salaries.
- The Partner salary has a strong correlation with partner working and Total salary, which means the partner salaries tend to contribute to higher total salaries

b) Relationship between Age and the Purchase Price of the Product

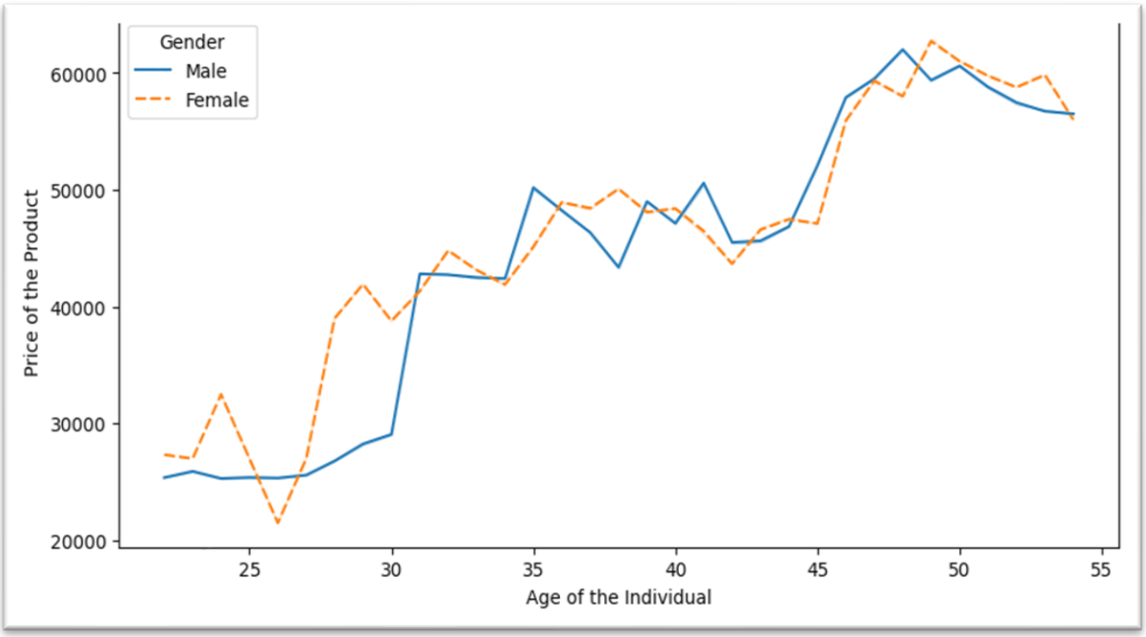

Figure 13 (Relationship between Age and Purchase price of the Product)

- We can see that an individual's age is strongly linked to the purchase of high-priced cars.
- Young individuals tend to purchase lower-priced cars, while older individuals are likely to buy higher-priced cars.

c) Price Distribution for Different Car Models

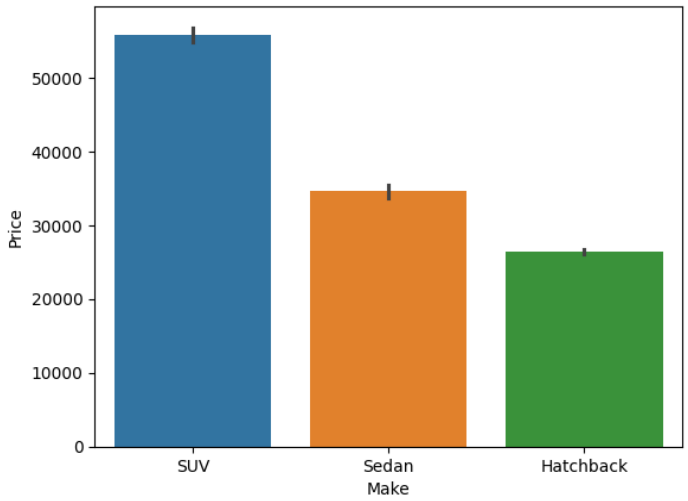

Figure 14 (Bar plot on Price Distribution for Different Car Models)

The above plot shows that SUVs and sedans have a higher price structure than Hatchback model cars.

d) Purchase Price Distribution by Home Loan Status

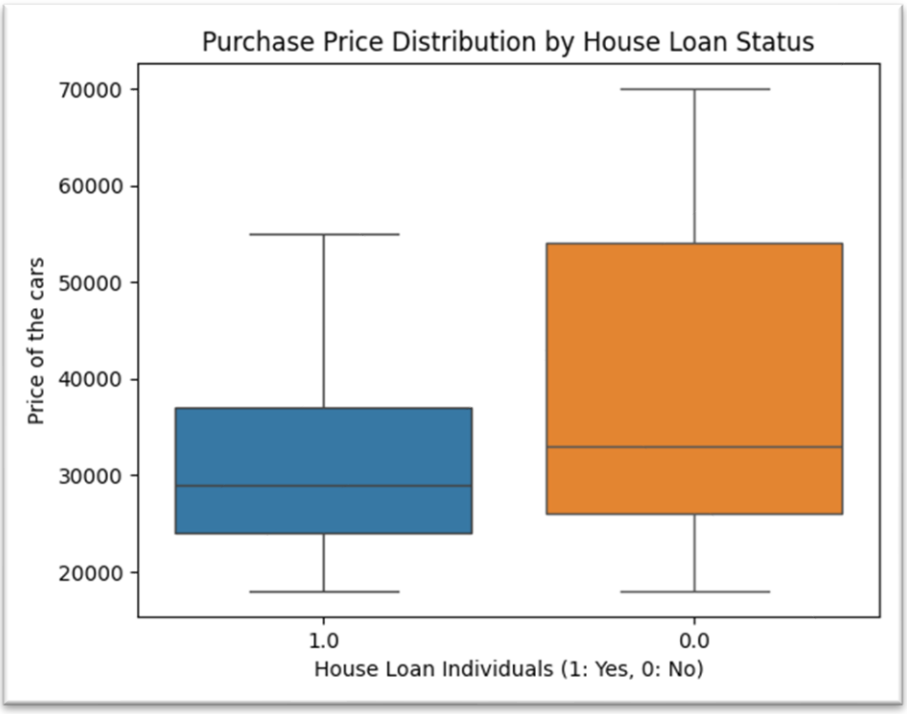

Figure 15 (Box plot on Purchase Price Distribution by Home Loan Status)

- A total of 527 individuals took out a House loan, and a total of 1054 individuals did not take out a House loan.
- The median purchase price of individuals tends to be higher than that of individuals with house loans.
- The individuals who do not have a home loan tend to spend a higher amount on purchasing automobiles.
- The data suggests that an individual's home loan status impacts the purchase price of automobiles

KEY QUESTIONS

Q1. Do men tend to prefer SUVs more compared to women?

- Females tend to prefer SUVs more than males, although their overall count is lower. Specifically, 173 females prefer SUVs, while only 124 males choose SUVs.
- Males show a strong preference for Hatchbacks and Sedans, with counts of 567 and 561.
- Hatchbacks are less popular compared to SUVs and Sedans among females.

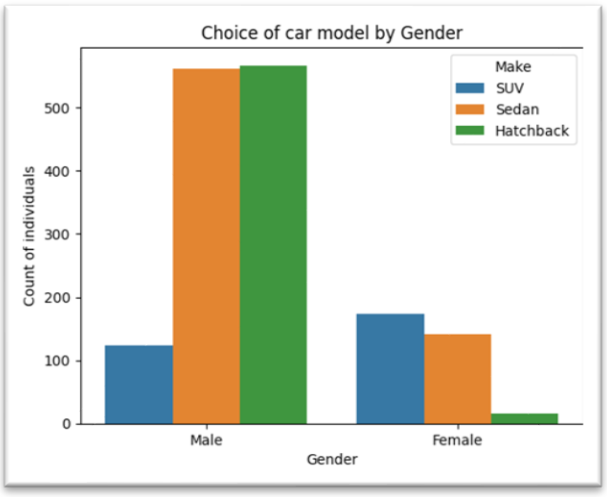

Figure 16 (Choice of car model by Gender)

Q2. What is the likelihood of a salaried person buying a Sedan?

- Sedans are the most popular type of car among both professions, but salaried individuals show a particularly high preference for sedans.
- so salaried customer is most likely to purchase a sedan.
- Hatchbacks have similar popularity among both professions.
- Additionally, salaried individuals tend to own SUVs more often than business individuals do

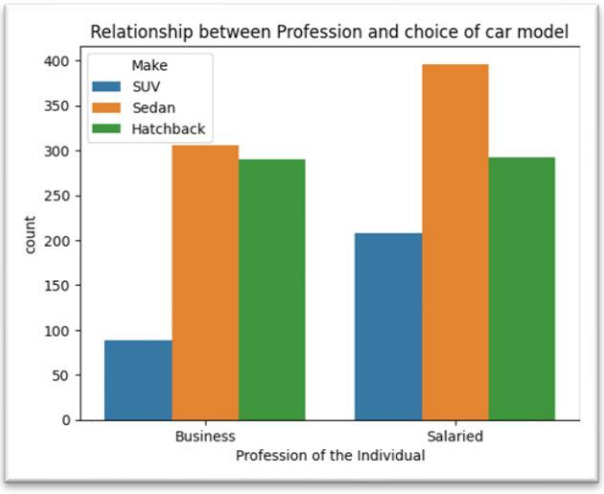

Figure 17(Relationship between Profession and choice of car model)

Q3. What evidence or data supports Sheldon Cooper's claim that a salaried male is an easier target for an SUV sale over a Sedan sale?

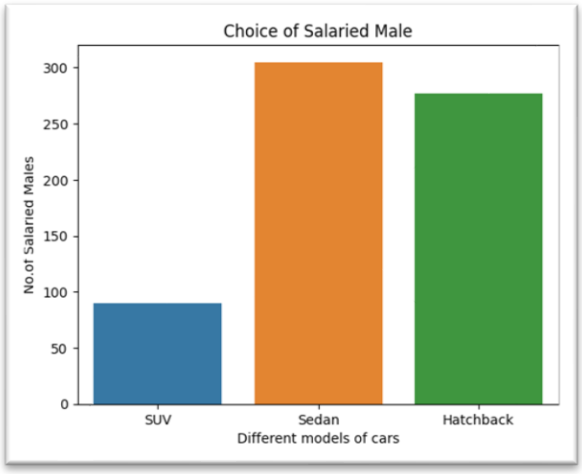

Figure 18 (Choice of Salaried Male)

- The chart shows that salaried males are less likely to buy SUVs compared to sedans, as the data indicates a significantly higher number of males targeted for sedan sales.
- So, Salaried males are not an easy target for SUV sales over a Sedan sale.

Q4. How does the amount spent on purchasing automobiles vary by gender?

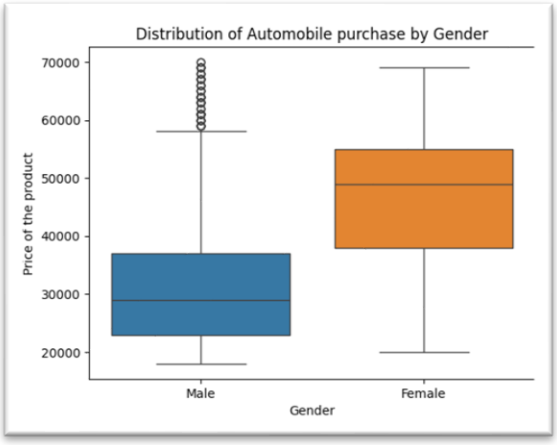

Figure 19 (Distribution of Automobile Purchases by Gender)

- The Median purchase price for females is higher than that of males.
- There are outliers in the male category, indicating some males purchased automobiles at significantly higher prices. There aren't any apparent outliers in the female category.
- The data for males is more skewed right side, while the female data is evenly distributed.
- In summary, females tend to purchase more automobiles with a de range of purchase prices, while males have a narrower range, and outliers are at higher prices.

Q5. How much money was spent on purchasing automobiles by individuals who took a personal loan?

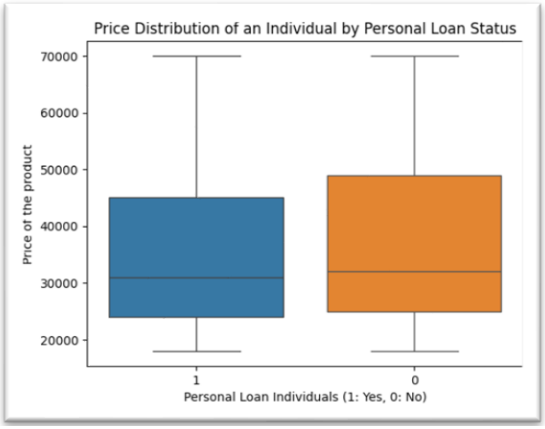

Figure 20 (Price Distribution of an Individual by Personal Loan Status)

- A total of 792 individuals took out a Personal loan, and a total of 789 individuals did not take out a personal loan.
- The median purchase price is similar for both groups, around 32000
- Both groups have similar maximum and minimum prices.
- The above data suggests that taking a personal loan by an Individual does not significantly affect the median price of an automobile purchased.

Q6. How does having a working partner influence the purchase of higher-priced cars?

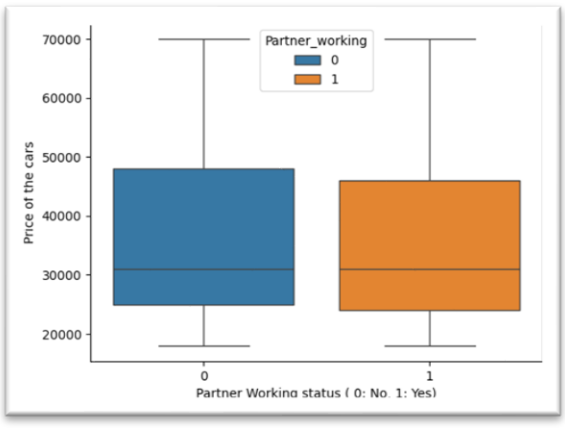

Figure 21 (Purchase Price Distribution by Working Partner Status)

- The median price of purchasing the products appears to be similar for both groups (working and non-working partners).
- Based on this visualization, the working status of a partner does not seem to have a substantial influence on the purchase of high-priced cars.

OUTLIERS TREATMENT

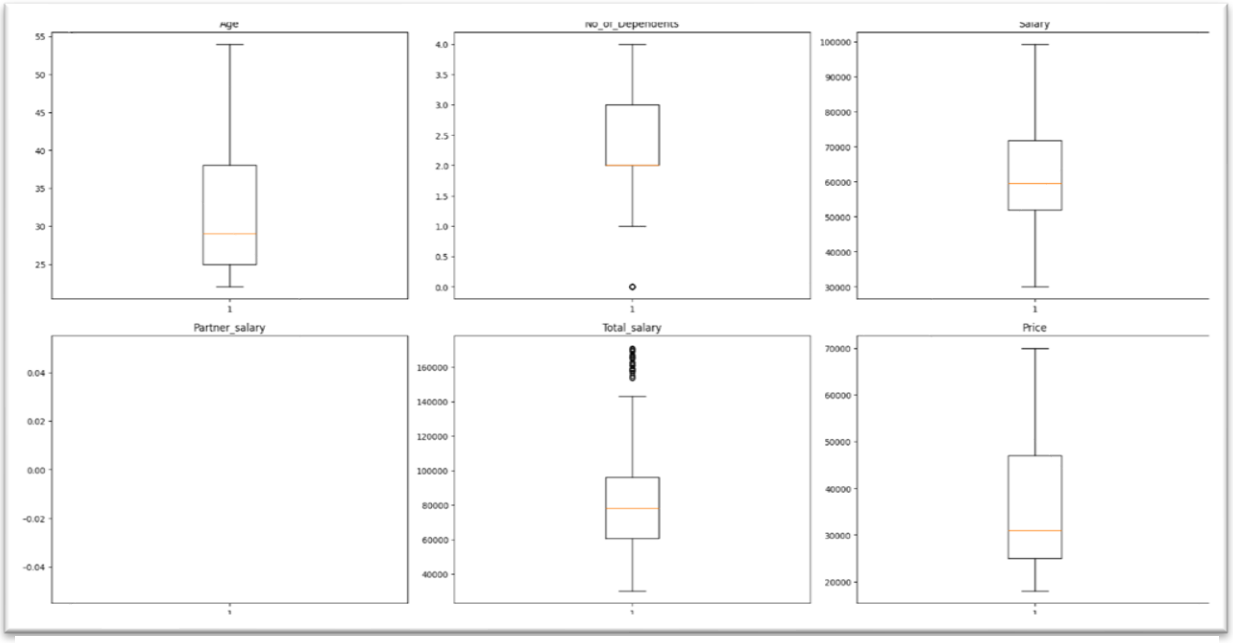

Figure 22 (Box plots of all numerical columns)

- We can observe outliers only in the Number of Dependents and Total Salary columns; the other columns do not have any outliers.
- However, we are not going to treat both columns since the Number of Dependents and Total Salary columns exhibit a wide range of values. We might miss out on genuine data if we treat the outliers

ACTIONABLE INSIGHTS

We have analysed a dataset of 1,581 Austo Automobiles company customers information to understand their demands and purchase patterns, aiming to improve the efficiency of the current marketing campaign. This analysis is key to identifying the ideal customers for different model cars.

From this analysis, we have been able to conclude that:

1. Age plays a crucial role in car purchase decisions. Younger age customers tend to purchase lower-priced cars, while older age customer is likely to buy higher-priced cars.
2. Females prefer SUV models more than males do.
3. Males have a strong preference for Sedans and Hatchbacks.
4. Sedans are the most popular choice for both Business and Salaried customers, but a salaried customer is most likely to purchase a sedan over other model cars.
5. Females tend to spend more money on purchasing automobiles compared to men, with a high median price.
6. The personal loan status of an individual does not impact the purchase price of the product, while the home loan status has a significant influence on the purchase price.
7. The working status of the Customer's partner does not have much influence on the decision to buy high-priced products.

RECOMMENDATIONS TO THE BUSINESS

1. The company should tailor their marketing campaigns based on customer's demographics and their preferences.
2. Offer affordable and entry-level cars to younger customers, while showcasing premium and high-end models to older customers.
3. Customize marketing messages and channels based on the age of the customers.
4. Highlight SUV models in marketing campaigns targeting females.
5. Emphasize sedan and hatchback models in marketing campaigns targeting male consumers.
6. Target sedans for business and salaried customers, prioritizing salaried customers.
7. Consider offering financial products for customers with or without home loans to encourage their purchase of these products.

- By implementing these recommendations, Austo Motor Company can better target their customers, improve the efficiency of their marketing campaigns, and increase their sales.