AUSTO AUTOMOBILES

PYTHON DATA SCIENCE PROJECT REPORT

MAY 18, 2025

ANNAPOORNA S

Table of Contents

PDS Project Report: Austo Automobiles

This report presents an evaluation of Austo Motor Company's existing marketing campaign for its automobiles. Austo Motor Company is a leading car manufacturer specializing in SUV, Sedan, and Hatchback models.

The purpose of analysing this dataset is to understand customer demand and improve marketing strategies for a car manufacturer.

The analysis will involve examining relevant datasets to identify critical key performance and formulating data-driven recommendations aimed at improving campaign reach, engagement, and conversion rates to drive sales and business growth.

Data Overview and Preparation

Data Description

The dataset encompasses customer information for Austo Motor Company, focusing on demographic details such as age, gender, and marital status, automobile preferences as well as financial information like salary and loan status. Key columns in the dataset include:

Variables	Description
Age	The age of the individual in years.
Gender	The gender of the individual, categorized as male or female.
Profession	The occupation or profession of the individual.
Marital_status	The marital status of the individual, such as married &, single
Education	The educational qualification of the individual Graduate and Post Graduate
No_of_Dependents	The number of dependents (e.g., children, elderly parents) that the individual supports financially.
Personal_loan	A binary variable indicating whether the individual has taken a personal loan "Yes" or "No"
House_loan	A binary variable indicating whether the individual has taken a housing loan "Yes" or "No"
Partner_working	A binary variable indicating whether the individual's partner is employed "Yes" or "No"
Salary	The individual's salary or income.
Partner_salary	The salary or income of the individual's partner, if applicable.
Total_salary	The total combined salary of the individual and their partner (if applicable).
Price	The price of a product or service.
Make	The type of automobile

Data Loading, Types and Structure

To understand the data, first we need to load the data and validate the structure and types of data.

- The given data set *austo_automobile.csv is* loaded and read using *read_csv()*
- The dataset contains 1,581 rows and 14 columns.
- Data types were verified to ensure numerical columns are classified as integers or floats, and categorical columns are recognized as objects.
- There are 6 numerical columns and 8 object type columns

	0
Age	int64
Gender	object
Profession	object
Marital_status	object
Education	object
No_of_Dependents	int64
Personal_loan	object
House_loan	object
Partner_working	object
Salary	int64
Partner_salary	float64
Total_salary	int64
Price	int64
Make	object

- Let's see statistical summary of numerical variables and object variables

	Age	No_of_Dependents	Salary	Partner_salary	Total_salary	Price
count	1581.000000	1581.000000	1581.000000	1475.000000	1581.000000	1581.000000
mean	31.922201	2.457938	60392.220114	20225.559322	79625.996205	35597.722960
std	8.425978	0.943483	14674.825044	19573.149277	25545.857768	13633.636545
min	22.000000	0.000000	30000.000000	0.000000	30000.000000	18000.000000
25%	25.000000	2.000000	51900.000000	0.000000	60500.000000	25000.000000
50%	29.000000	2.000000	59500.000000	25600.000000	78000.000000	31000.000000
75%	38.000000	3.000000	71800.000000	38300.000000	95900.000000	47000.000000
max	54.000000	4.000000	99300.000000	80500.000000	171000.000000	70000.000000

	Gender	Profession	Marital_status	Education	Personal_loan	House_loan	Partner_working	Make
count	1528	1581	1581	1581	1581	1581	1581	1581
unique	4	2	2	2	2	2	2	3
top	Male	Salaried	Married	Post Graduate	Yes	No	Yes	Sedan
freq	1199	896	1443	985	792	1054	868	702

  - The average age of customers is approximately 31.9 years, with most individuals between 22 and 54 years old.
  - There are few individuals who wants to buy a car, some are with the dependants with maximum of 4 number of dependants.
  - The average salary is $60392.220114, but significant variability exists, as shown by a standard deviation of $14674.825044 indicating a diverse customer base.
  - An individual earns minimum of $30,000 to $99,300, where we can notice diverse customer base.
  - Some customers partners also working with a significant salary which adds additional income and highly efficient to purchase car.
  - The mean total salary is $79,626, suggesting substantial combined incomes among customers.
  - The dataset predominantly consists of Male customers
  - A significant portion of customers are in Salaried profession than in Business
  - Most of the customers are Married, highlighting a potential focus for family-oriented marketing strategies.
- Missing data was identified as follows:
  - 'Gender': 53 missing entries
  - 'Partner_salary': 106 missing entries

Missing Value Treatment

- Partner_salary missing values were imputed using the formula: ***Total_Salary = salary + Partner_Salary*** for valid entries to ensure non-negative values.
- There were few data in Gender where there was a spelling mistakes for female ('Femal', 'Female', 'Femle'). So, if Gender starts from F, it is treated as Female
- Gender missing values were filled using the mode, which was found to be Male.

Statistical Summary with observation and insights

Below is a statistical summary for key numerical variables and object variable after imputation:

	Age	No_of_Dependents	Salary	Partner_salary	Total_salary	Price
count	1581.000000	1581.000000	1581.000000	1581.000000	1581.000000	1581.000000
mean	31.922201	2.457938	60392.220114	19233.776091	79625.996205	35597.722960
std	8.425978	0.943483	14674.825044	19670.391171	25545.857768	13633.636545
min	22.000000	0.000000	30000.000000	0.000000	30000.000000	18000.000000
25%	25.000000	2.000000	51900.000000	0.000000	60500.000000	25000.000000
50%	29.000000	2.000000	59500.000000	25100.000000	78000.000000	31000.000000
75%	38.000000	3.000000	71800.000000	38100.000000	95900.000000	47000.000000
max	54.000000	4.000000	99300.000000	80500.000000	171000.000000	70000.000000

- The mean age of customers is approximately 31.92, with a minimum of 22 years and a maximum of 54 years, indicating a relatively young customer base.
- The mean salary is about $60,392 with a high standard deviation of $14,674, suggesting varied income levels among customers.
- Partner_salary has a considerable deviation and a maximum value of $80,500, but also a significant number of entries at $0, indicating some partners may not be employed or contributing financially to the household.
- The Total_salary represents the combined income, showcasing the purchasing power within households—averaging $79,626.

	Gender	Profession	Marital_status	Education	Personal_loan	House_loan	Partner_working	Make
count	1581	1581	1581	1581	1581	1581	1581	1581
unique	2	2	2	2	2	2	2	3
top	Male	Salaried	Married	Post Graduate	Yes	No	Yes	Sedan
freq	1252	896	1443	985	792	1054	868	702

- Even after imputation, Gender distribution shows a predominant number of males (1,252), indicating a potential target market trend.
- Predominantly 896 customers are Salaried and 685 do the business
- Marital status indicates that most customers are married (1,443), pointing to household-oriented marketing strategies where partner also working and adds on to the partner salary.
- Education levels reflect a higher number of customers with a Post Graduate degree (985), possibly correlating with higher income levels.
- 792 customers have their personal loan but most of them don't have house loan.
- The preference for Sedans (702) as the most popular car type, as observed in the Make category, suggests that marketing campaigns should include a focus on this type of vehicle.

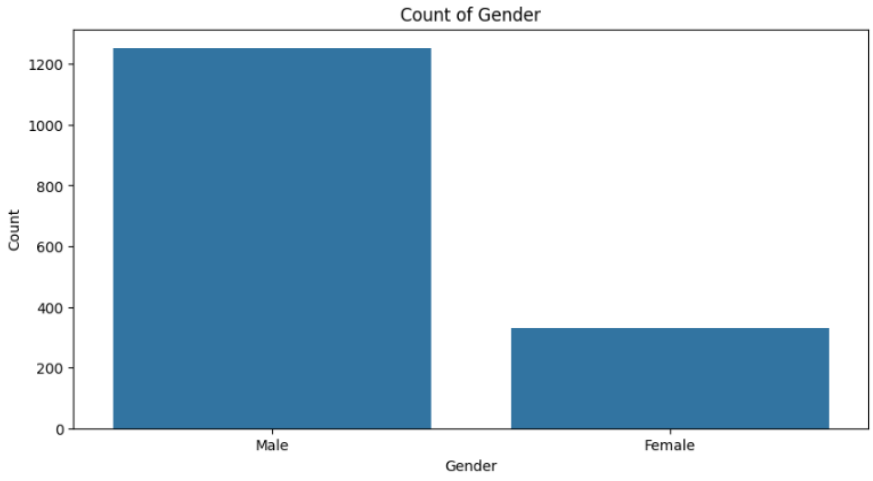
Univariate Analysis

Summarize the univariate analysis performed on the Austo Motor Company dataset, exploring both categorical and continuous variables and identify the outliers.

Categorical Variables

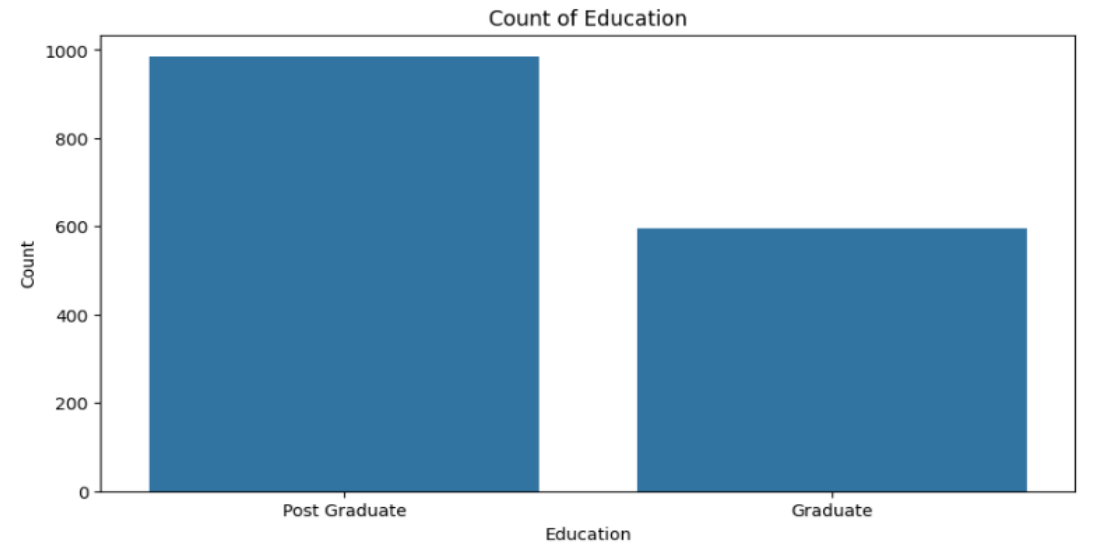
Analysis of categorical variables was carried out with count plots illustrated below for each category.

- **Gender Distribution**:



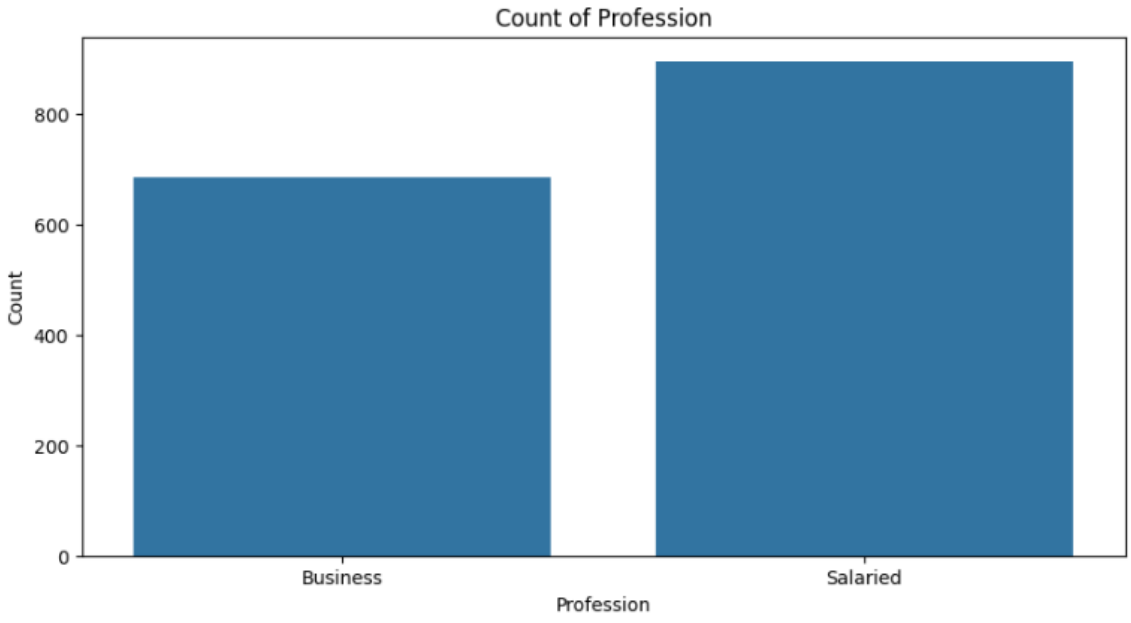
The dataset shows a 20.8% Female and 79.2% Male demographic representation which is predominantly Male purchasing car.

- **Education:**



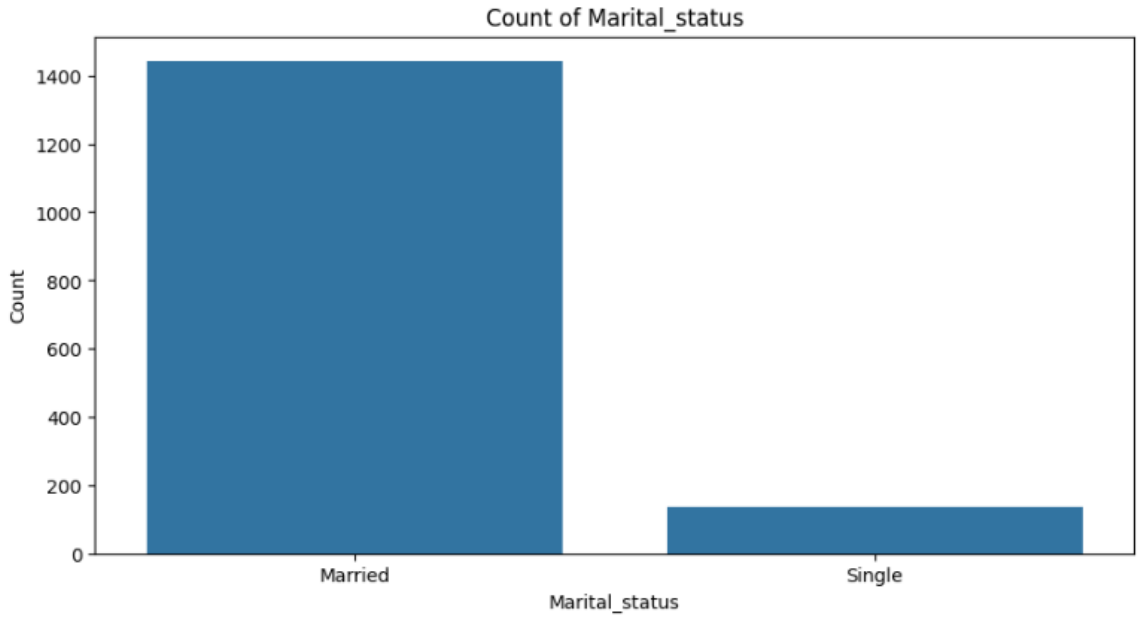
Most of the customers are Post graduated comparatively with high qualifications.

- **Profession Distribution**:



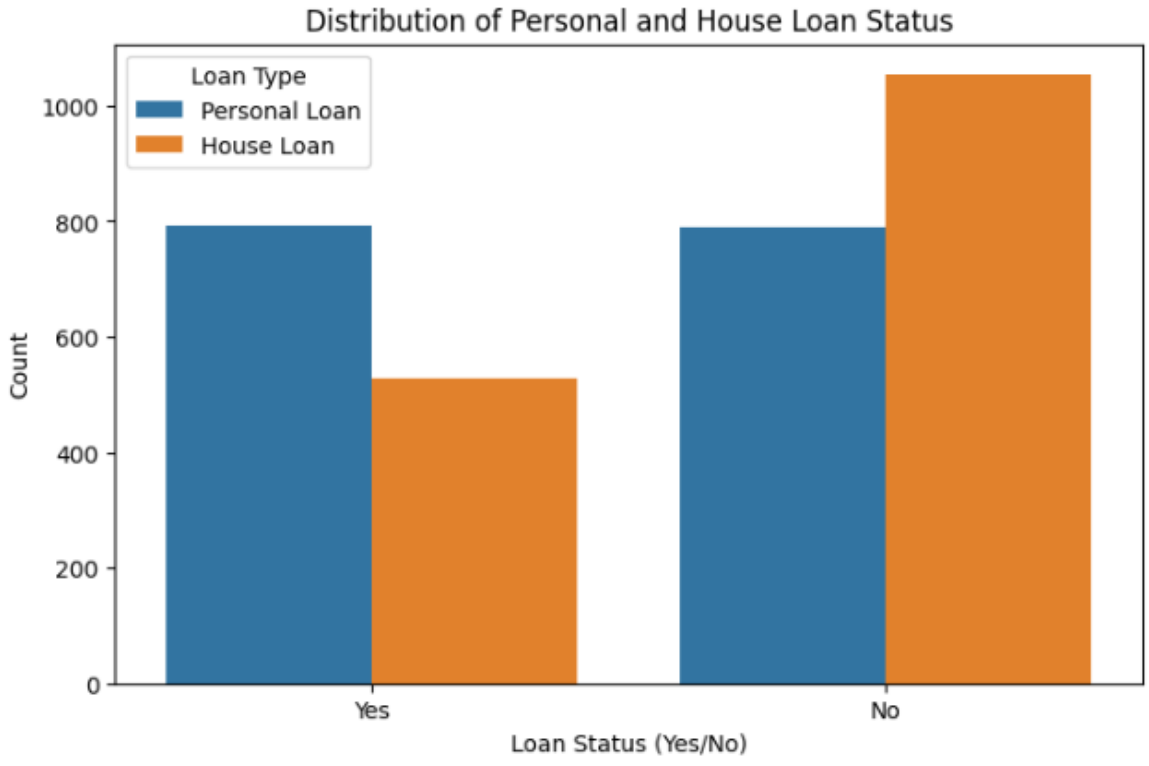
- Majority are engaged in Salaried profession; few are doing their business.

- **Marital Status Distribution:**



  - The dataset contains 1443 are Married individuals out of 1881, indicating potential family-oriented marketing strategies.

- **Loan Status Distribution:**

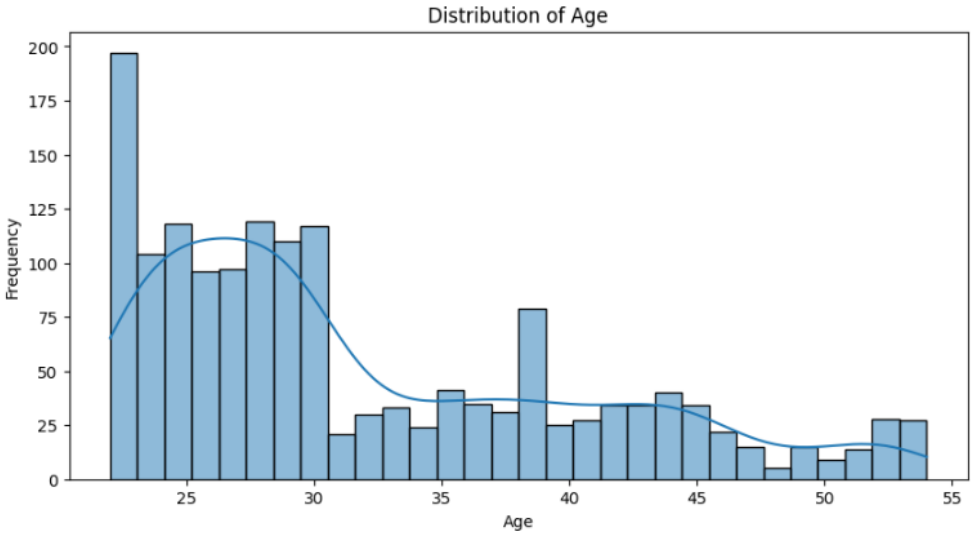


- Customers have personal loan more than the house loan, some people neither have house loan nor personal loan considered fully stable and can purchase car appropriately.

Numerical Variables

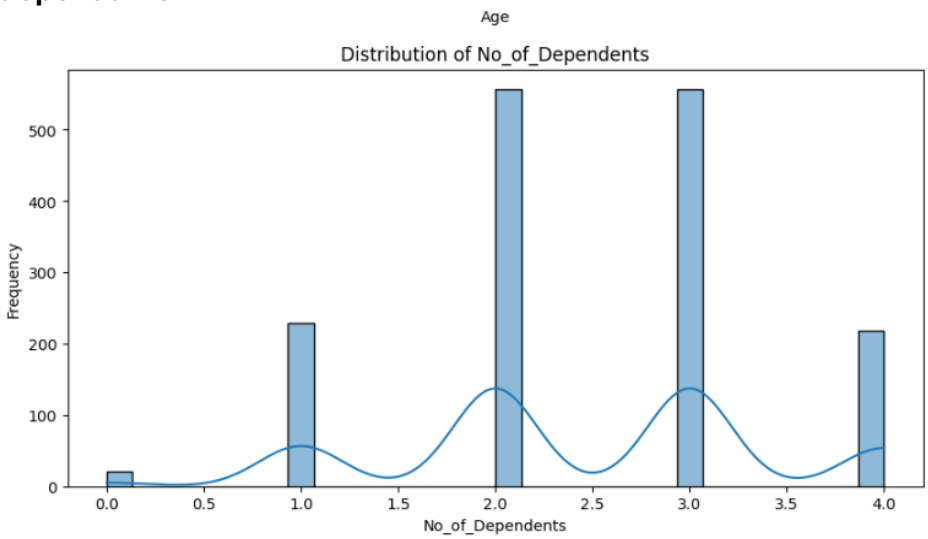
Explorations of numerical variables showcased multiple visualizations:

- **Age Distribution**:



Most customers are between ages 25-35.

- **No_of_dependants:**



557 number of customers have dependants of 3 and 2 majorly.

- **Salary Distribution**:

- Some customers have support of their partners salary which makes their total income high for a stable life.

Total salary = Customer Salary + Partner's salary

- However, we can see some **outliers** from total salary which beyond the upper bound.

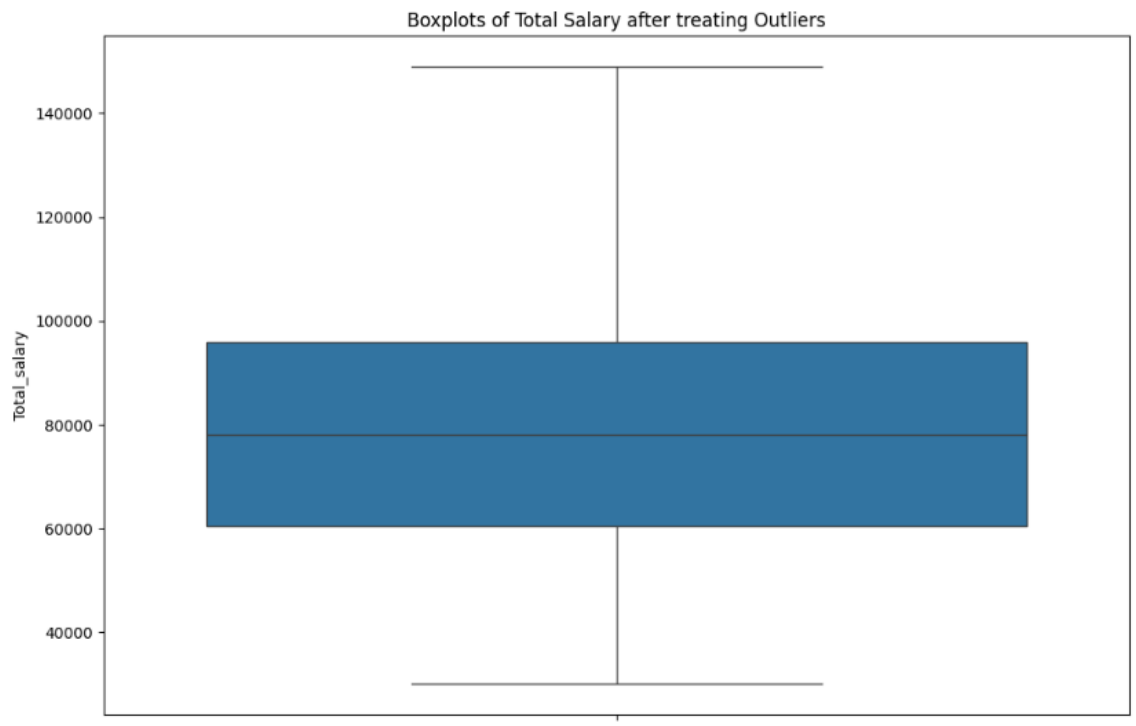
OUTLIER TREATMENT

In the above graph we see the outliers in total salary.

- We treated those outliers with IQR method (Replace the data points with the lower whisker (Q1 - 1.5 * IQR) or upper whisker (Q3 + 1.5 * IQR) value)

```
IQR = Q3 - Q1
lower_bound = Q1 - 1.5 * IQR
upper_bound = Q3 + 1.5 * IQR

where Q1 is 25th percentile of Total salary,
Q2 is 75th percentile of Total Salary
```



This is the graph after treating outliers in Total Salary.

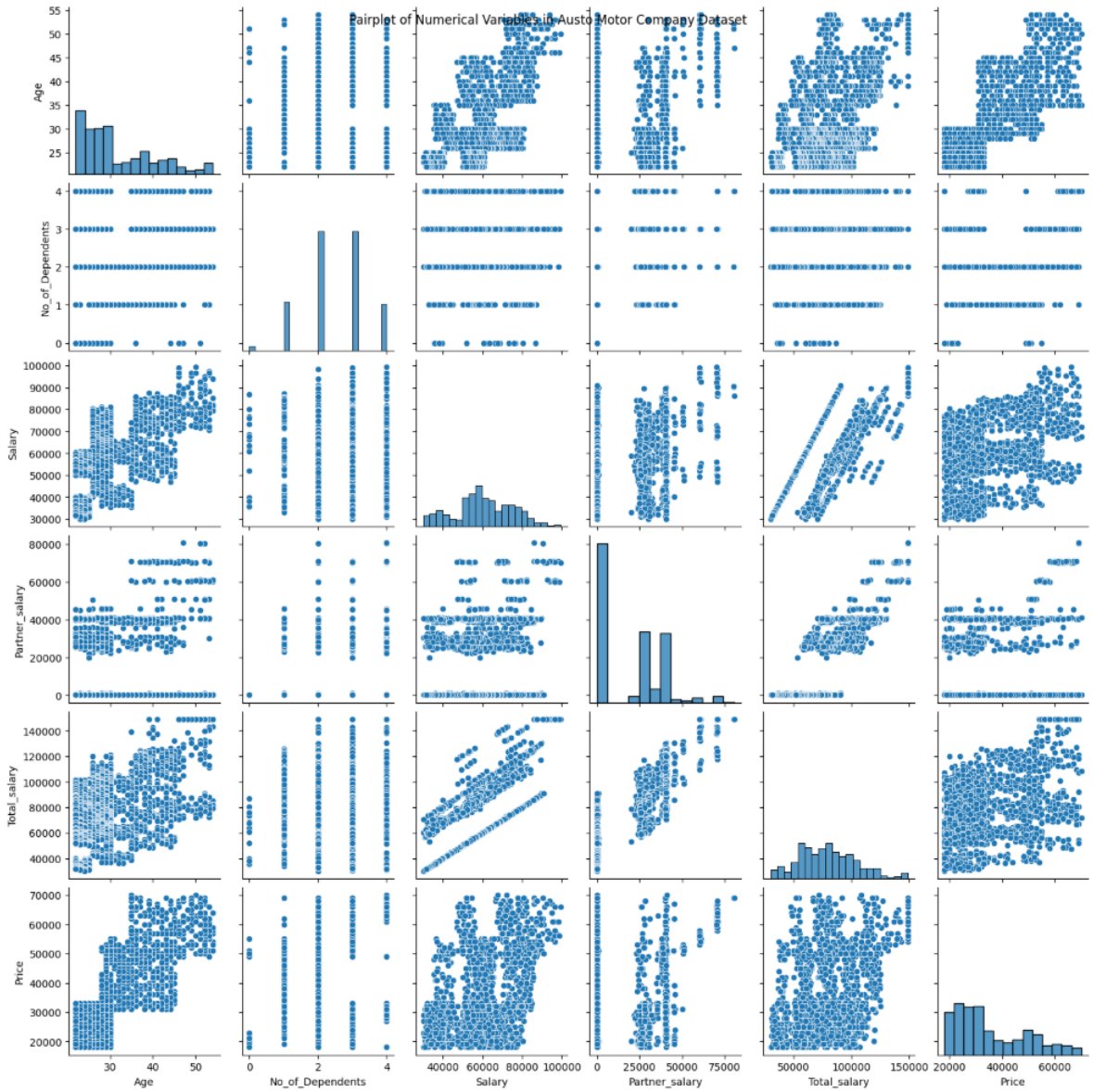
Bivariate Analysis

Analyse the data seeing relationships between two variables and their correlation to know the trends over categorical variables, numerical variables and also between them.

Explore Relationships Between Numerical Variables

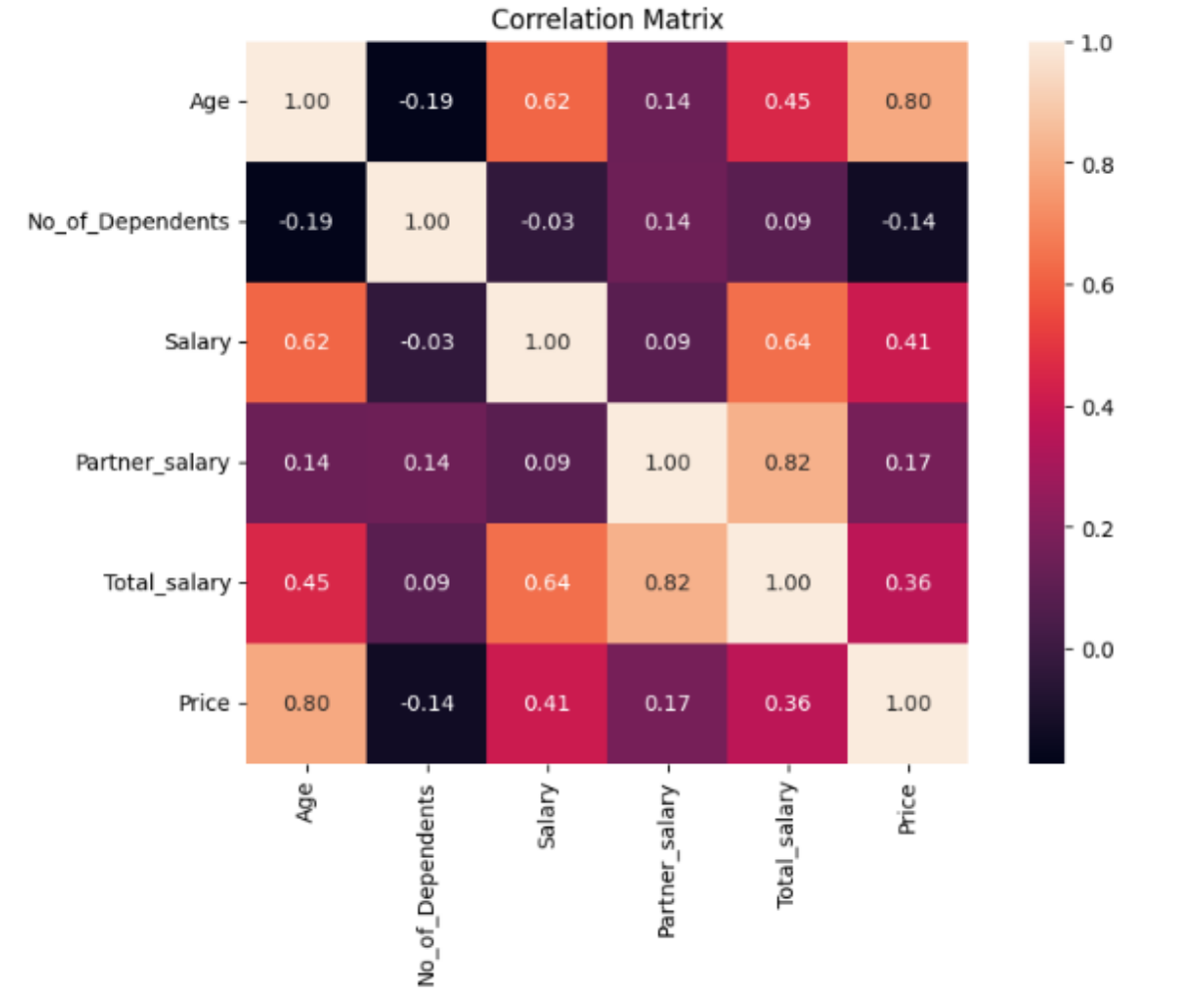
We use pairplot to analyse relationship between each numerical data.

Relationship[between each two numerical data could be seen in the below graph.



1. There is a **strong positive correlation between Salary and Total Salary**, indicating that higher individual earnings correspond to higher combined household incomes, which is crucial for understanding customer purchasing potential.
2. The graph between **Age and Salary suggests a slight positive** trend, indicating that older individuals generally tend to have higher salaries, potentially reflecting increased work experience and seniority.
3. The pairwise comparisons show that **prices are positively associated with both Salary and Total Salary,** indicating that as customers' incomes increase, so do their willingness to spend on higher-priced vehicles.

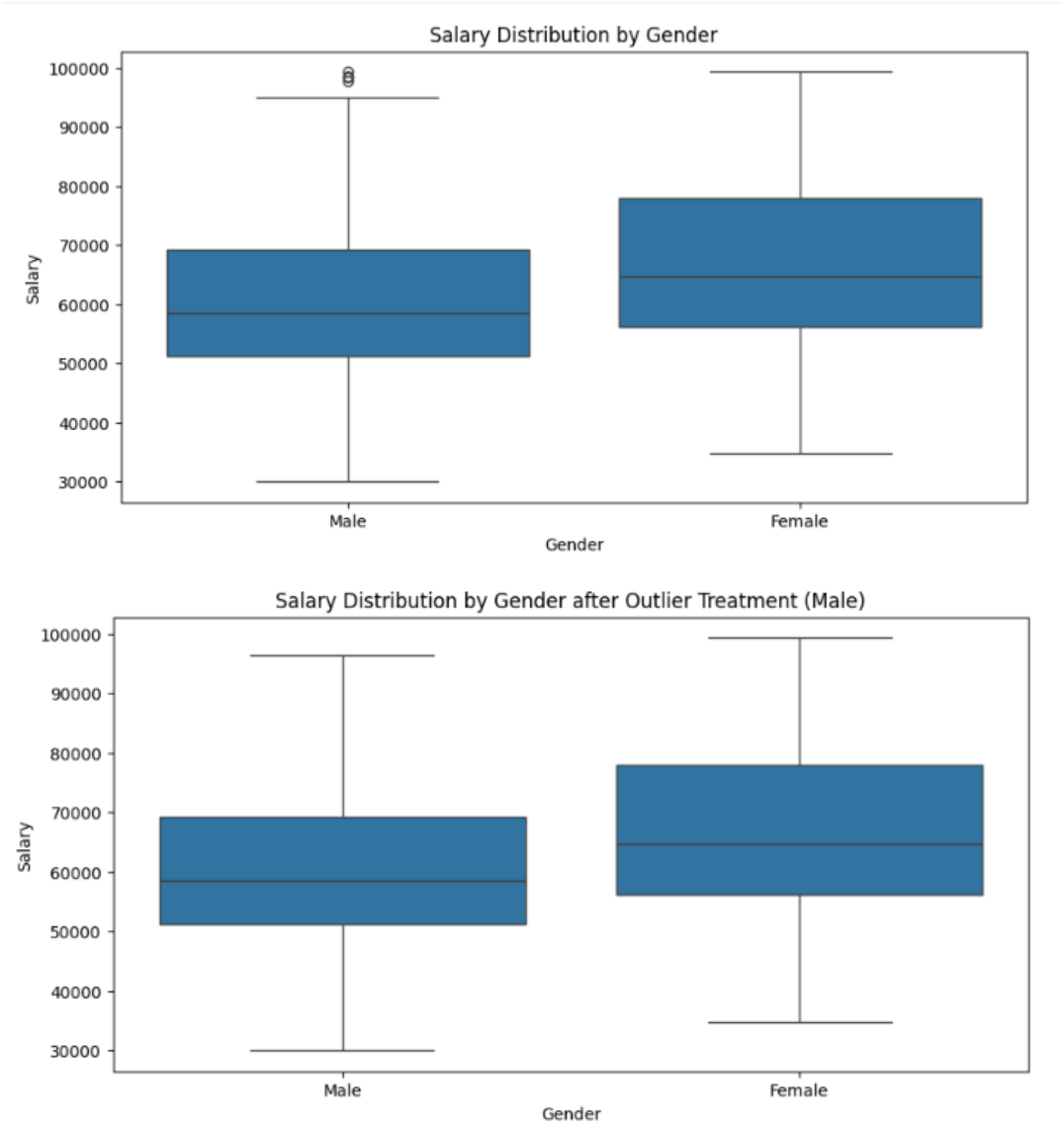
EXPLORE THE CORRELATION BETWEEN ALL NUMERICAL VARIABLES



- There is a strong positive correlation between Partner_salary and Total Salary(0.82), which meanshaving partner support with their salary increasing with his total income, which is crucial for understanding customer purchasing potential
- Additionally, Age shows a significant positive correlation with Price (0.80), suggesting that older customers tend to purchase more expensive vehicles.
- The matrix reveals a negative correlation between No_of_Dependents and Age (-0.19), indicating that younger individuals may have fewer dependents. Similarly, Price correlates negatively with No_of_Dependents (-0.14), suggesting that households with more dependents might opt for lower-priced vehicle options.

EXPLORE THE RELATIONSHIP BETWEEN CATEGORICAL VS NUMERICAL VARIABLES

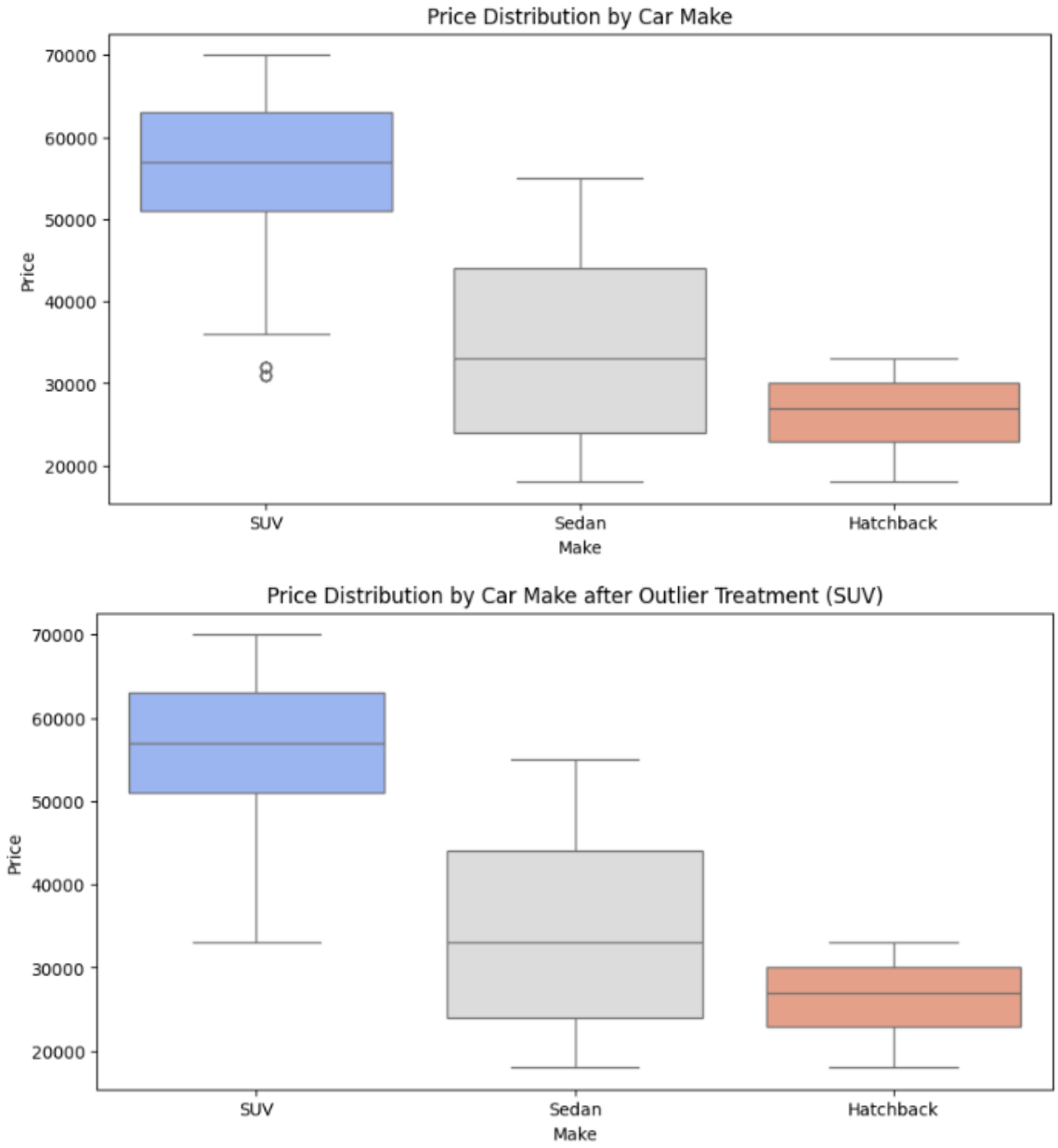
Salary distribution by Gender



There were few extreme high outliers noticed between Salary and male Gender. After outlier treatment with IQR method, the distributions appear more balanced retaining all the data inplace.

From outliers we can say there are men who earns extremely high salary, but women earns almost higher than the male.

Price distribution by Car Make



There were few customers who buys SUV at very low price. Those outliers are treated with IQR where lower bound value is considered and remove the outleirs and retaining all the values.

Considering the above plots, we can say that Hatchback is lower priced and SUV is higher priced, Seden is in between.

Relationship between salary and a price of cars

From the above graph, we can say that a person with low salary tends to buy low prised car such as Hatchback, above that with some more salary customer tends to buy Sedan which is in low to medium range of price. A person whose income is very high they mostly tends to buy SUV.

KEY QUESTIONS

Explore the data to answer the following key questions:

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

Make	Hatchback	SUV	Sedan
Gender			
Female	15	173	141
Male	567	124	561

No, Women tend to prefer SUVs compared to men in given dataset. 173 females and 124 male bought SUV.

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

number of Salaried individuals who purchased a Sedan = 396

total number of salaried individuals = 896

likelihood = Salaried individuals who purchased a Sedan / total number of salaried individuals

44.20% for salaried individuals purchasing a Sedan indicates a significant market presence for this vehicle type

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

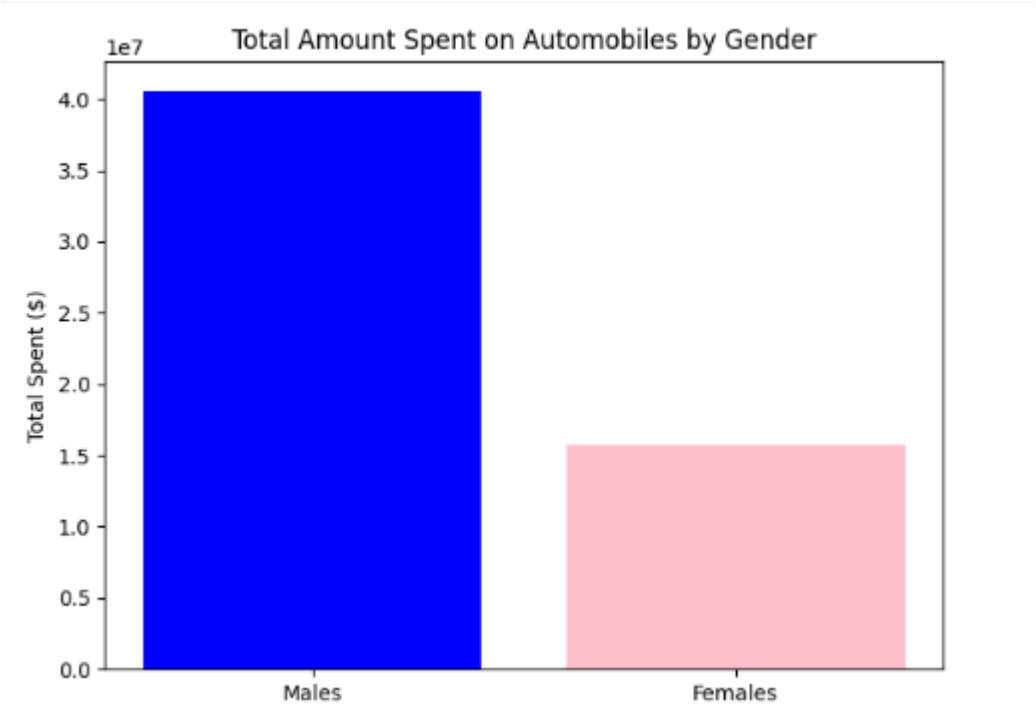
A salaried male purchases, SUV: 90, Sedan: 305

Percentage of SUV Purchases: 22.78%

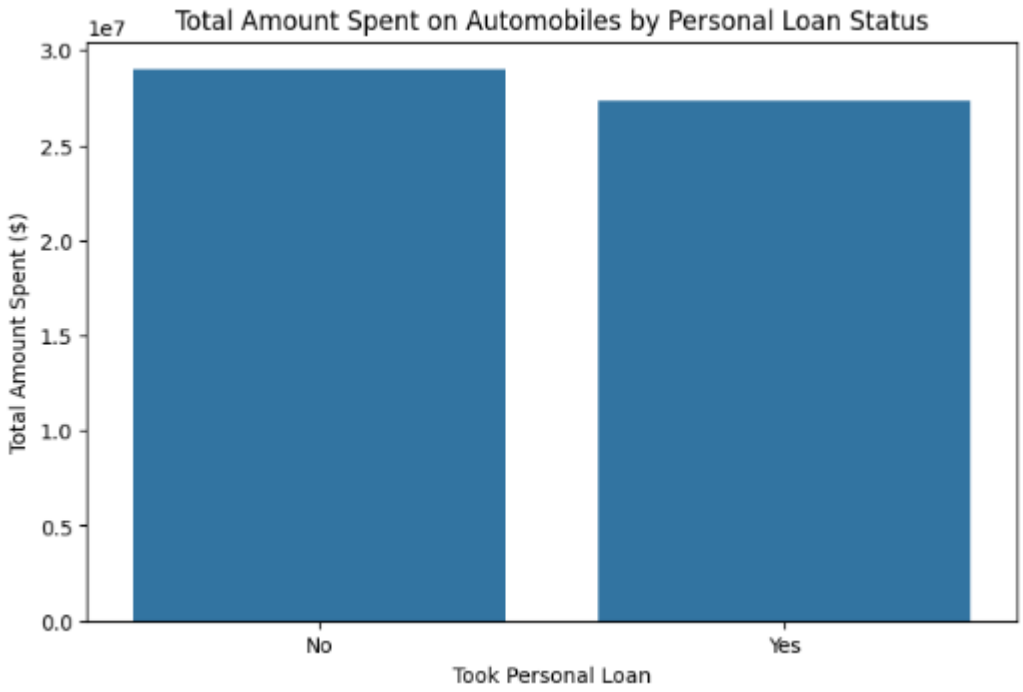
Percentage of Sedan Purchases: 77.22%

From this data, I would say Salaried males show a higher preference for Sedans compared to SUVs.

4. **How does the amount spent on purchasing automobiles vary by gender?**
   The analysis shows that males spent a total of $440585000 on automobiles, while females spent $15710000. The average spending per male was $32416.13, compared to females at $47750.75. These differences suggest that ***males tend to invest more*** in automobiles, which could inform targeted marketing strategies aimed at different gender.



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



Total money spent on automobiles by individuals with a personal loan: $27,300,000.00

6. **How does having a working partner influence the purchase of higher-priced cars?**
   Let's consider the 75th percentile as high-priced cars - $47,000.00
   Number of Higher-Priced Cars Purchased with Partner: 201
   Number of Higher-Priced Cars Purchased without Partner: 180
   Individuals with working partners tend to purchase more higher-priced cars.

ACTIONABLE INSIGHTS & RECOMMENDATION

After analysing the given data set there are few observation insights I would like to include:

- Males are predominant customers for purchasing cars compare to female
- Strategies to engage female customers also more effectively, highlighting the features that reasonate their preferences
- Couples with partners who contribute to household income tend to make more car purchases.
- Young individuals tend to purchase less expensive cars, while older, more experienced buyers invest in higher-priced cars.

- Ther are quite few individuals with no dependants purchase highly priced cars
- Customers are highly qualified with their Postgraduation or minimum with their Graduation
- If the person has no personal and home loans are more likely to buy cars compare to person who have their loans. But most of the person tends to take their personal loans to buy car.
- Individuals with lower salaries typically opt for more affordable models, such as Hatchbacks.
- As income increases, customers often transition to purchasing Sedans, which generally fall within the low to medium price range.
- Conversely, those with significantly higher incomes tend to favour SUVs, indicating a preference for larger and often more expensive vehicles.
- Ensure that inventory reflects customer preference trends, placing more emphasis on SUVs while maintaining a healthy supply of efficient Sedans for younger consumers.