Austo Automobile
Case Study

Post Graduate Program in Data Science with Generative AI

JUNITHA M

1. Introduction

- Auto Motor Corporation is an established automobile company manufacturing SUVs, sedans, and hatchbacks.
- The management identified issues in the current marketing strategy affecting customer engagement and sales performance.
- The management decided to involve a data analyst to evaluate customer behavior and provide insights that could help refine and strengthen the existing marketing approach.

2. Objective

The analysis specifically focuses on the following aspects:

1. Comparison of SUV preferences between male and female customers.
2. Probability of salaried professionals opting for sedan models.
3. Validation of Sheldon's assumption regarding SUV purchases by salaried men.
4. Analysis of automobile expenditure based on gender.
5. Impact of personal loans on vehicle spending behavior.
6. Role of a working spouse in influencing the purchase of high-value cars.

3. Dataset Overview

The dataset contains various demographic, financial, and automobile-related attributes of individuals. Each variable and its explanation are given below.

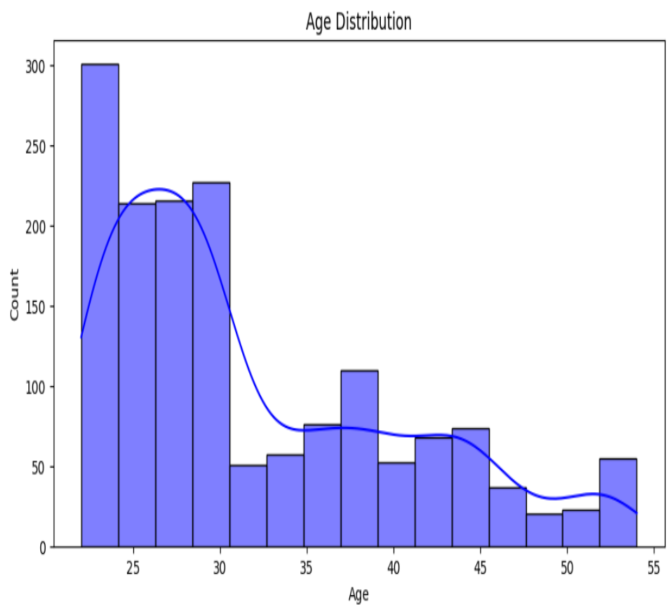
Variables	Description
Age	Represents the individual's age measured in years.
Gender	Categorized as Male or Female
Profession	Occupation of the individual.
Marital_status	Marital condition of the individual, such as married or unmarried.
Education	(Graduate or Post-Graduate).
No_of_Dependents	Number of people financially dependent on the individual, including children or elderly parents.
Personal_loan	Identifies whether the individual has availed a personal loan (Yes/No).
House_loan	Indicates whether the individual has taken a home loan (Yes/No).
Partner_working	Specifies if the individual's spouse/partner is employed (Yes/No).
Salary	The monthly or annual income earned by the individual.
Partner_salary	Income earned by the individual's partner, where applicable.
Total_salary	Combined income of the individual and their partner.

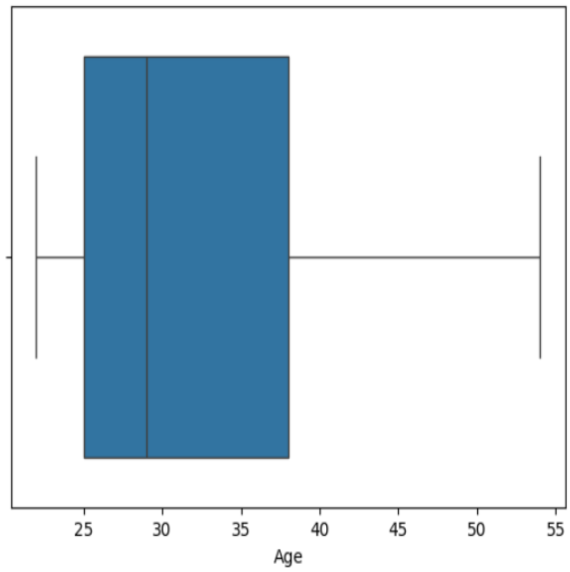
Price	Cost of the product or service considered.
Make	Category or brand of the automobile.

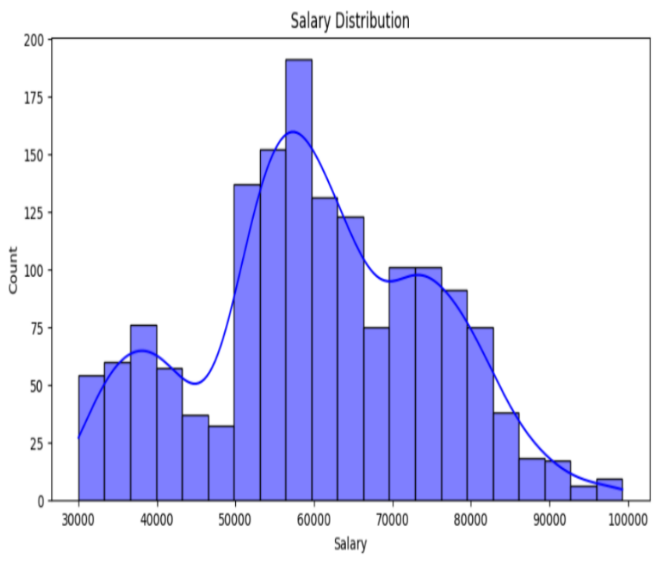
- The dataset contains **1,581 records** and **14 variables**.
- No duplicate entries were identified; however, 53 records lack gender information, and 106 records have missing partner salary values.
- Two incorrect gender values were found, namely “Femal” and “Femle.”
- These incorrect gender entries were standardized and replaced with “Female.”
- Missing values in the Gender column were retained as ‘unknown’ to avoid introducing bias.
- For cases where the partner was not employed, the partner salary was set to 0.
- In 16 records, where the partner was employed but salary data was unavailable, the partner’s income was calculated by subtracting the individual’s salary from the total salary.

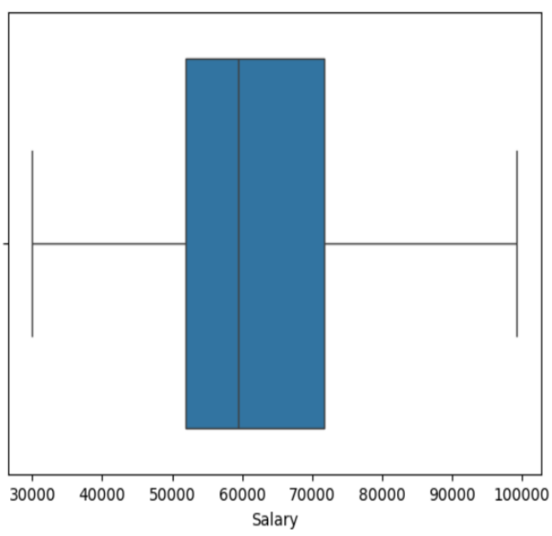
4. Univariate Analysis

4.1 . Univariate Analysis using Numerical Data









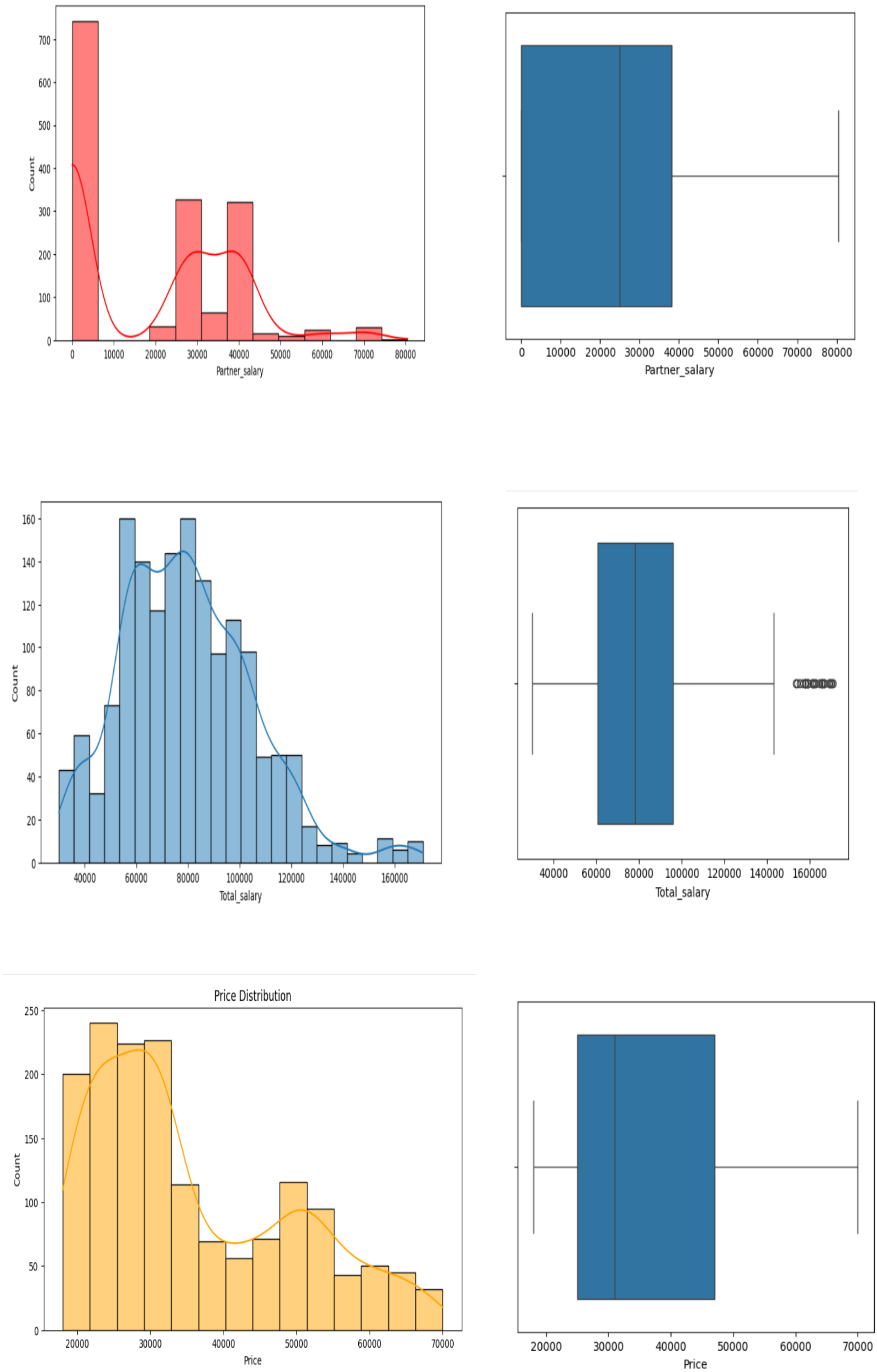


Fig : Univariate Analysis of numerical data

Observations and Insights

- **Age** values are mainly concentrated between 25–38 years and show a right-skewed distribution, representing the active working population.
- **Salary** data shows multiple peaks with an approximately normal trend, where most values lie between ₹50,000 and ₹70,000.
- **Partner salary** is highly dispersed due to the inclusion of zero values for non-working partners, affecting the overall distribution.
- **Total salary** exhibits skewness with few outliers (≈1.7%); hence, no outlier treatment was applied, and most values fall between ₹60,000–₹100,000.
- **Price** follows a bi-modal, right-skewed pattern, indicating two purchase ranges, with a major concentration between ₹25,000 and ₹38,000.
- Most numerical variables are skewed, and therefore none strictly follow a perfect normal distribution.

4.2 Univariate Analysis of Categorical Data

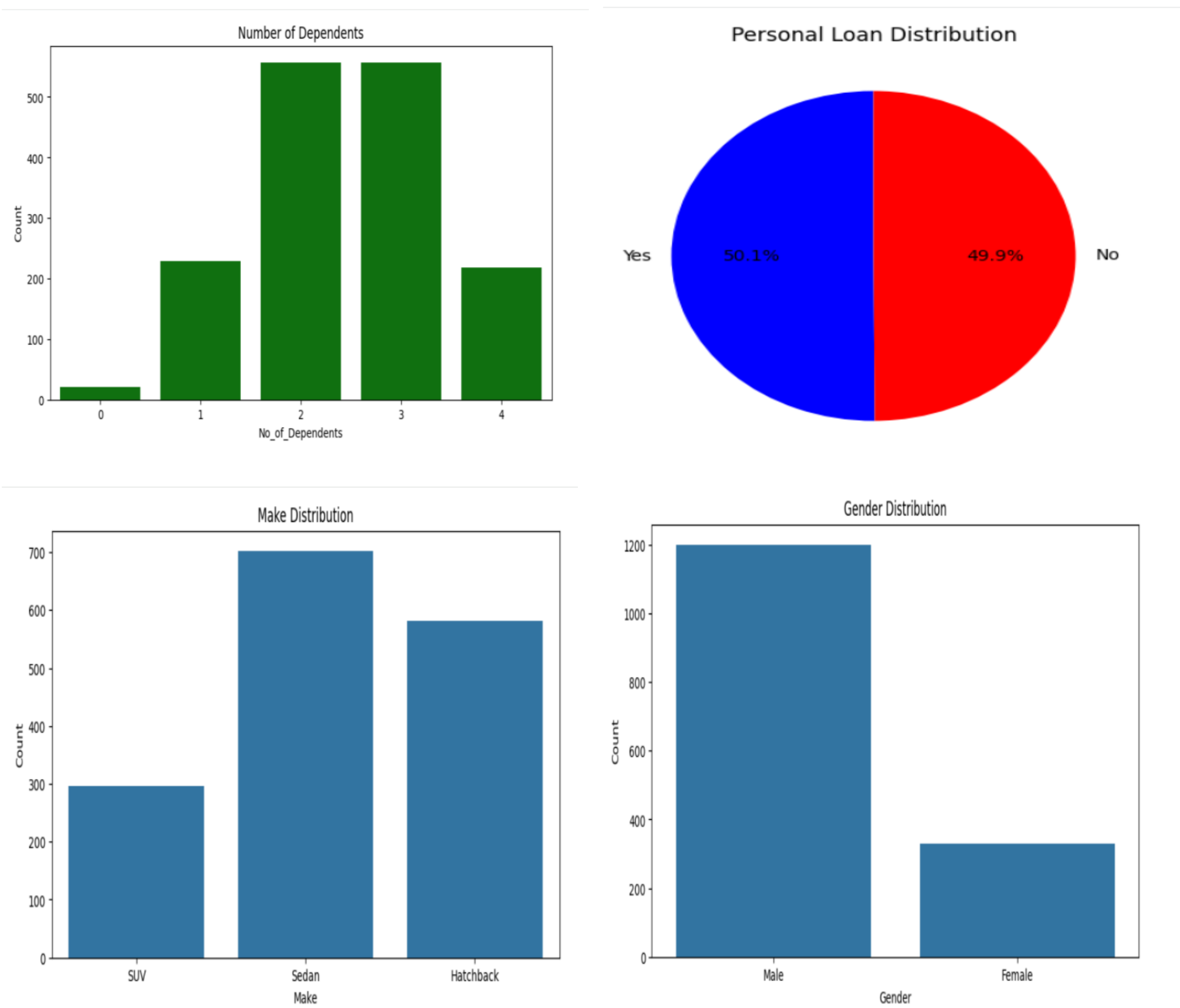


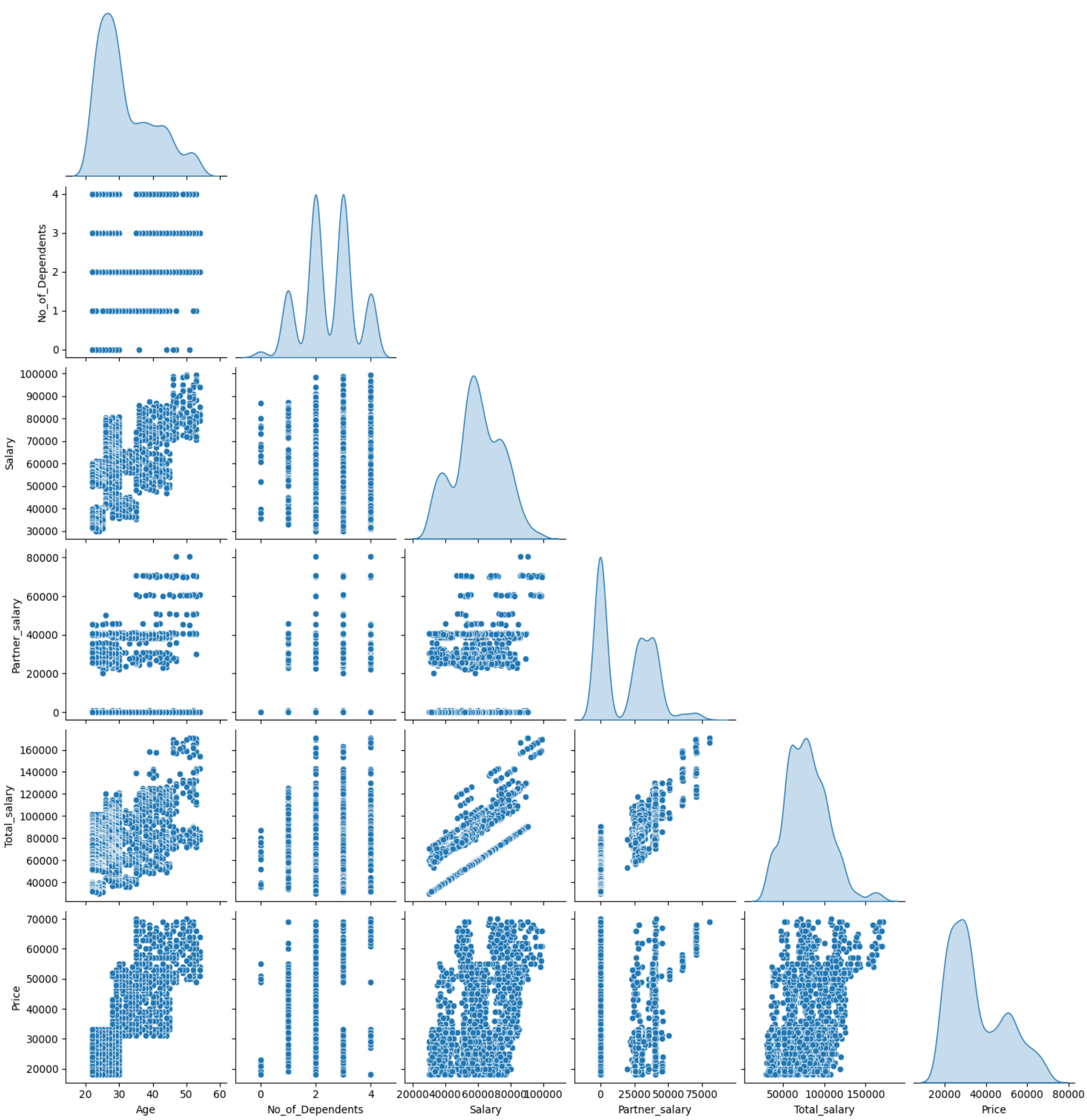
Fig : Univariate Analysis of Categorical Data

Observations and Insights

- **Analysis of dependents** reveals that most customers have two or three dependents, followed by those with one or four. Only a small number of customers have no dependents.
- The **personal loan** variable shows an almost even split between customers who have taken a loan and those who have not, approximately **50:50**.
- Analysis of **vehicle preference** indicates that sedans are the most commonly purchased automobiles, followed by hatchbacks and SUVs.
- The gender-wise distribution shows that **male customers significantly outnumber female customers** in the dataset.

5. Bivariate Analysis

5.1 . Bivariate Analysis using Numerical Data



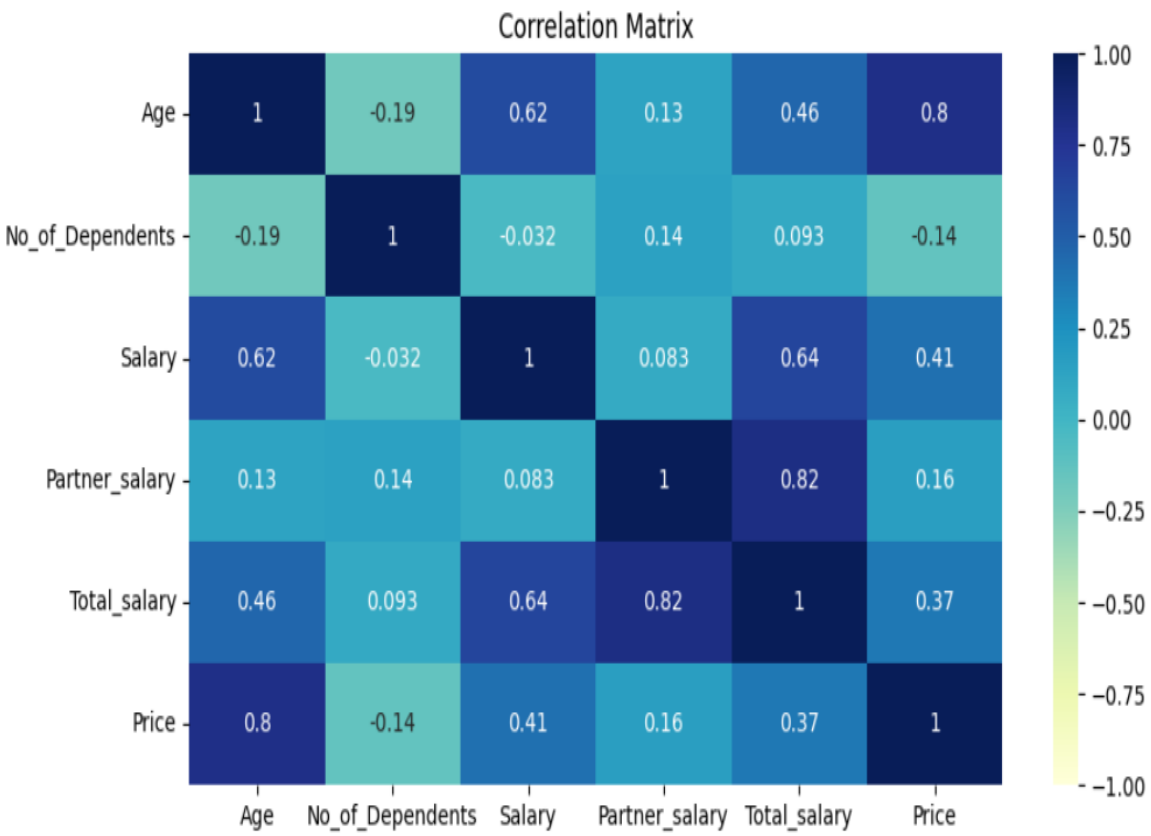
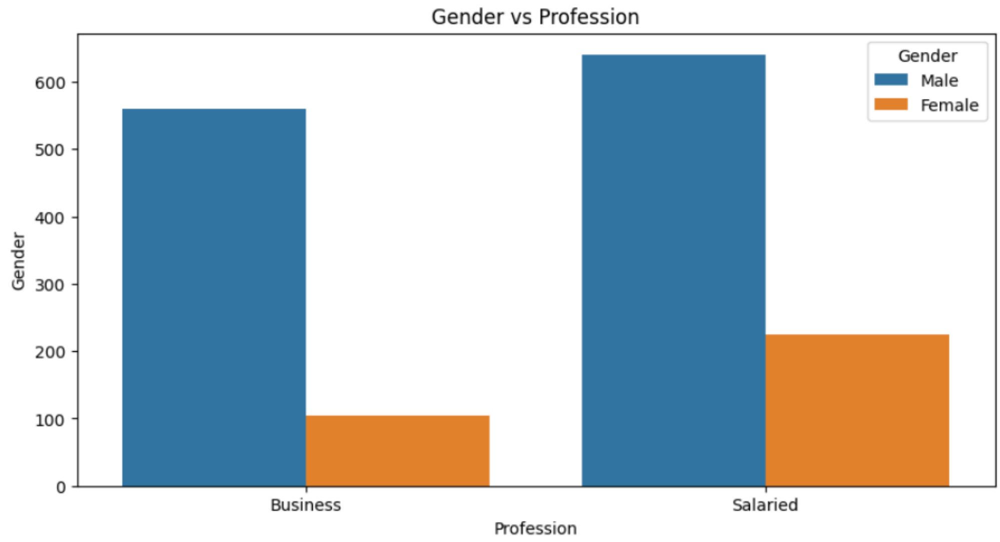


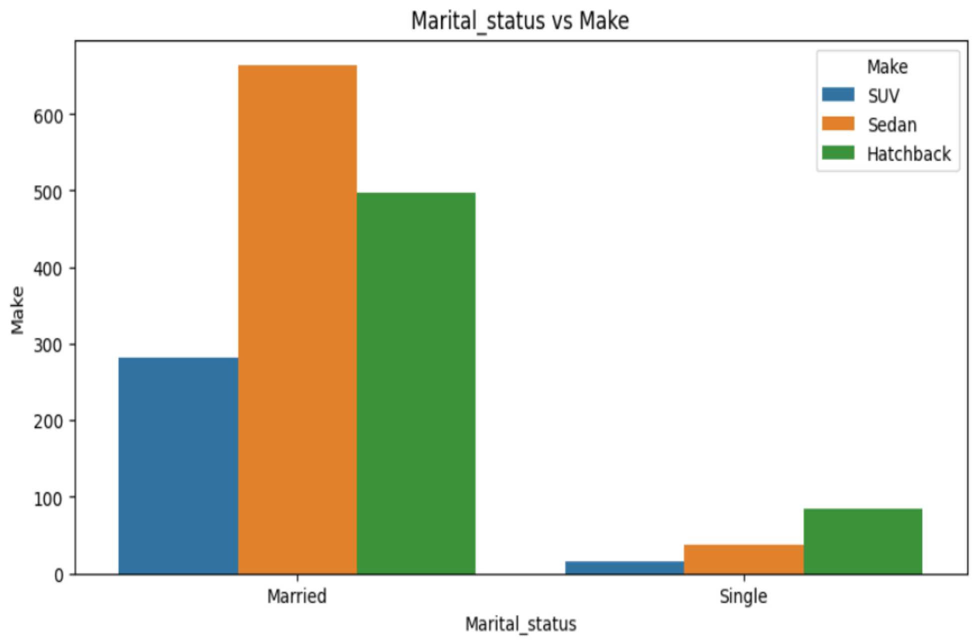
Fig : Bivariate Analysis using Numerical Data

Observations & Insights

- The pair plot indicates that most numerical features do not exhibit strong associations, except for noticeable relationships between **age & salary** and **age & price**.
- Customers in the **20–30 age group** generally show a preference for **lower-cost vehicles**, as reflected in the Age versus Price analysis.
- An inverse trend is observed between the **number of dependents and vehicle price**, where customers with more dependents tend to opt for lower-priced options; however, those with **two dependents** are seen purchasing vehicles across a wide price range.
- The correlation heatmap highlights a **strong positive relationship** between **age and price**, as well as between **partner salary and total salary**, while the association between **age and individual salary** is moderately positive.
- **Negative correlations** are evident between **price and dependents**, **age and dependents**, and **dependents and salary**, suggesting that an increase in dependents is associated with lower values of these variables.

5.2 Bivariate Analysis using Categorical Data





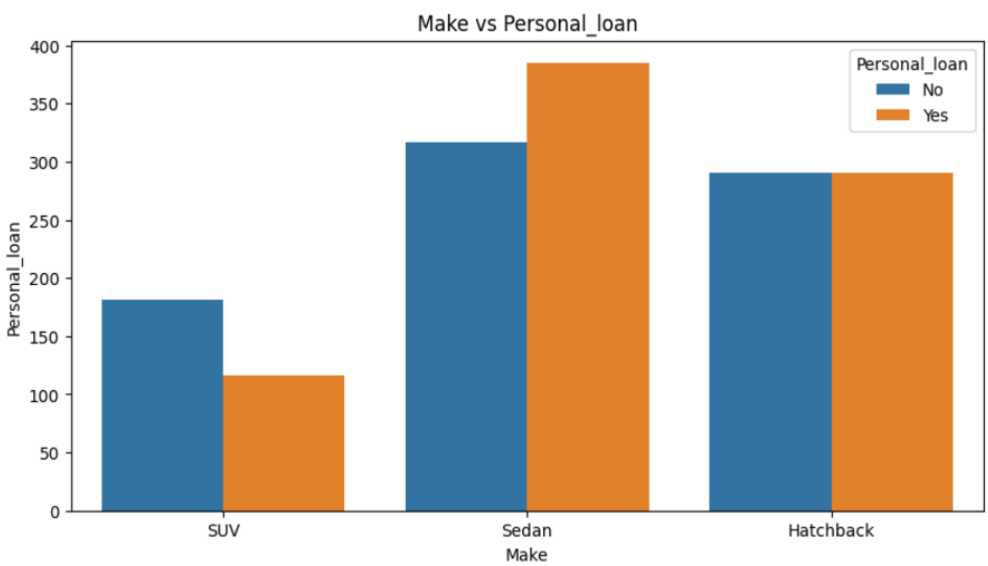
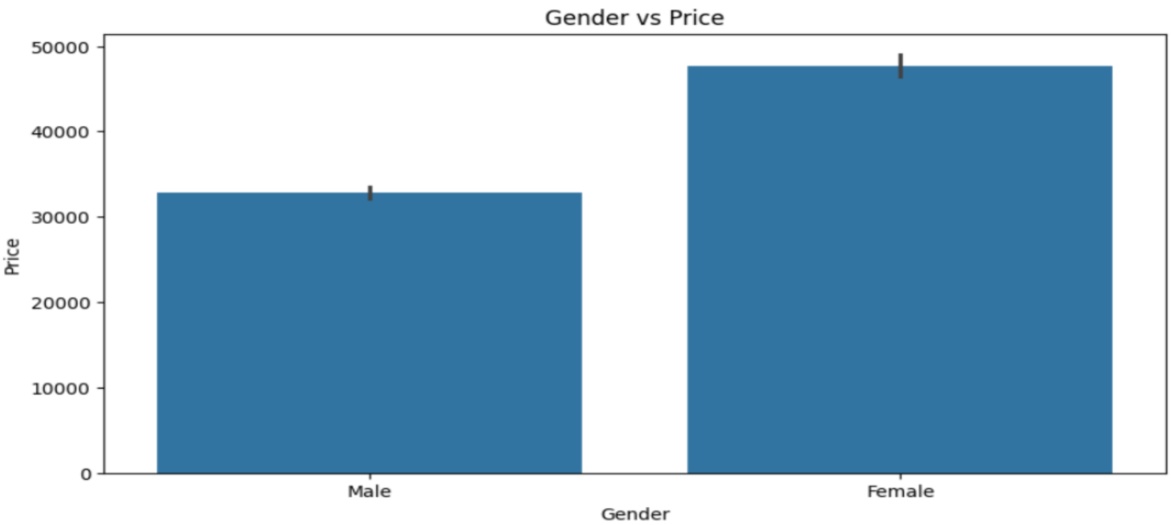


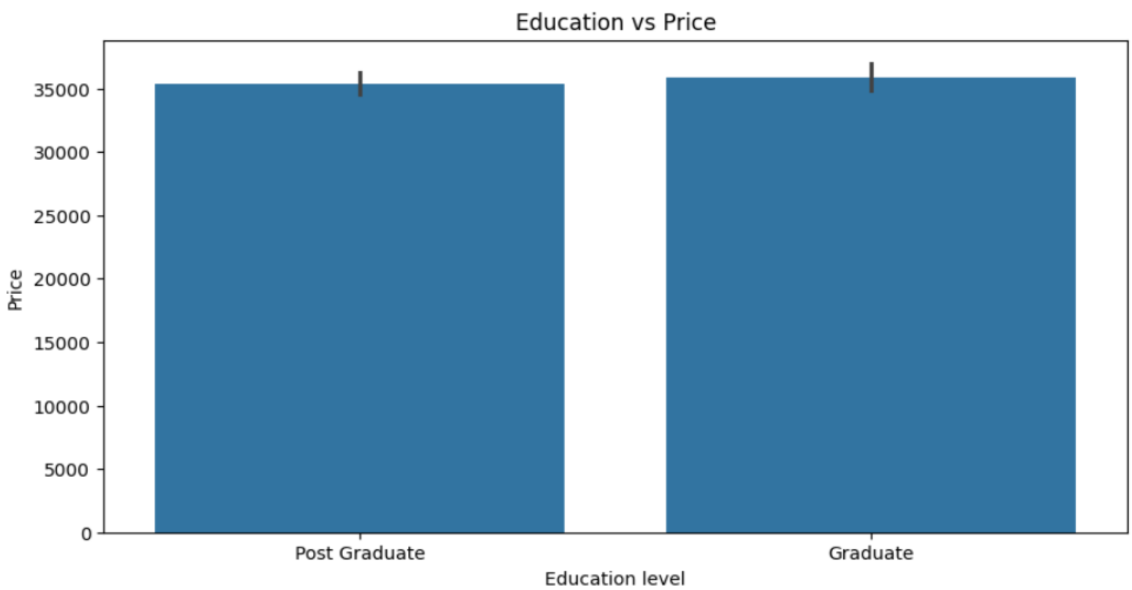
Fig : Bivariate Analysis using Categorical Data

Observations and Insights

- When comparing **profession and gender**, salaried individuals form a larger share of the dataset than business professionals across both male and female groups.
- Examination of **profession versus personal loan** status reveals no significant variation, suggesting that profession does not strongly influence personal loan uptake.
- The relationship between **marital status and vehicle type** indicates that married customers mainly prefer sedans, followed by hatchbacks and SUVs. In contrast, single customers show a stronger inclination toward hatchbacks, with sedans and SUVs chosen less frequently, likely due to budget considerations.
- In the **personal loan and vehicle type** comparison, sedans remain the most commonly selected option regardless of loan status. Hatchback purchases are largely unaffected by personal loan decisions, while customers with personal loans tend to purchase fewer SUVs compared to non-loan customers.

5.3 Relationship between Categorical Data and Numerical Data





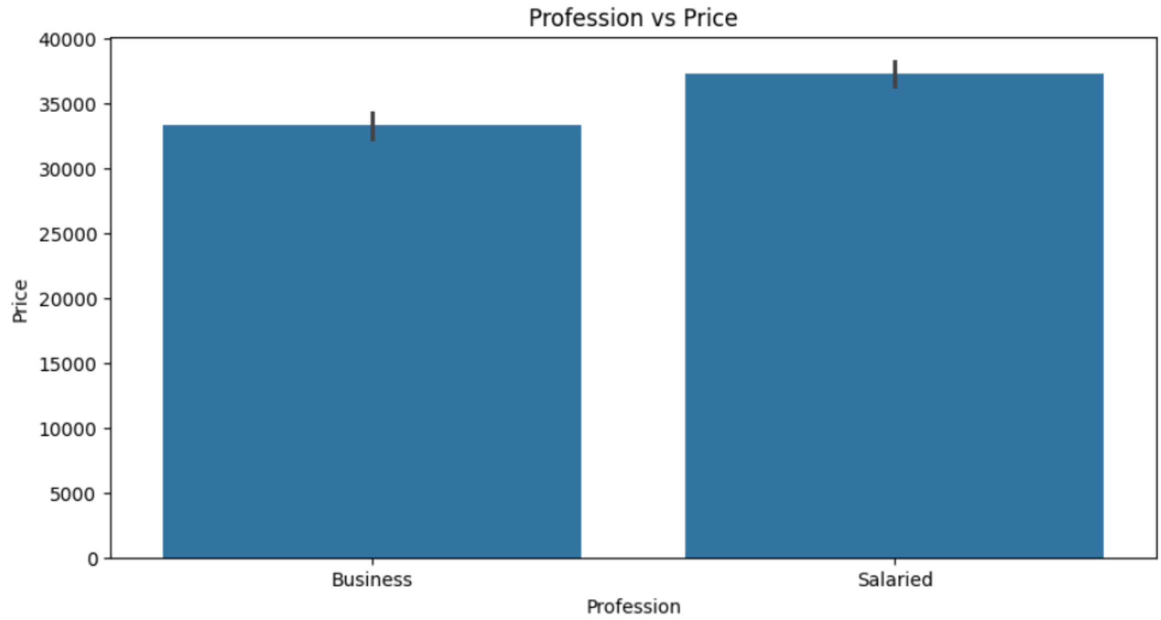


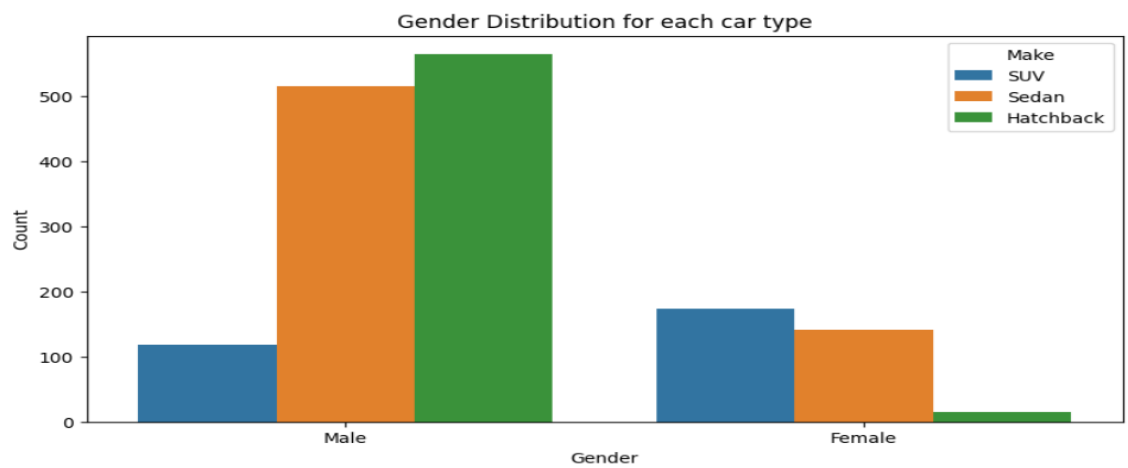
Fig: Relationship between Categorical Data and Numerical Data

Observation and Insights

- Gender vs Price: Female customers tend to spend more on average compared to male customers.
- Education vs Price: Education does not significantly impact the amount spent on vehicle purchases.
- Profession vs Price: Salaried customers show marginally higher average spending compared to business customers.

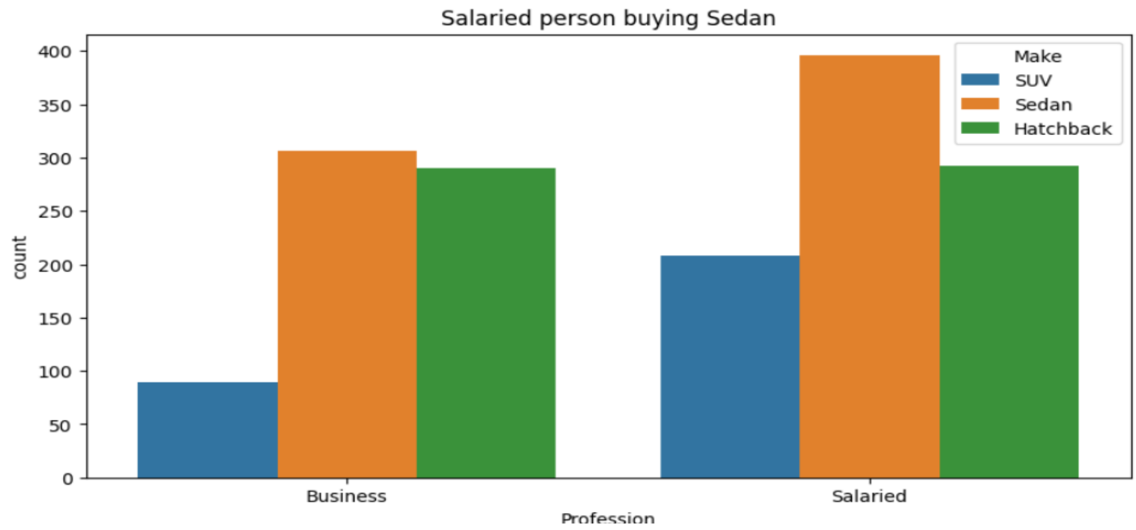
Key Questions:

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



From the count plot, female customers account for approximately 53% of SUV purchases, while male customers contribute around 47%. In comparison to their total purchases, the proportion of SUV buyers is higher among females, indicating that women show a relatively stronger preference for SUVs than men.

2. Likelihood of a Salaried person buying a Sedan?

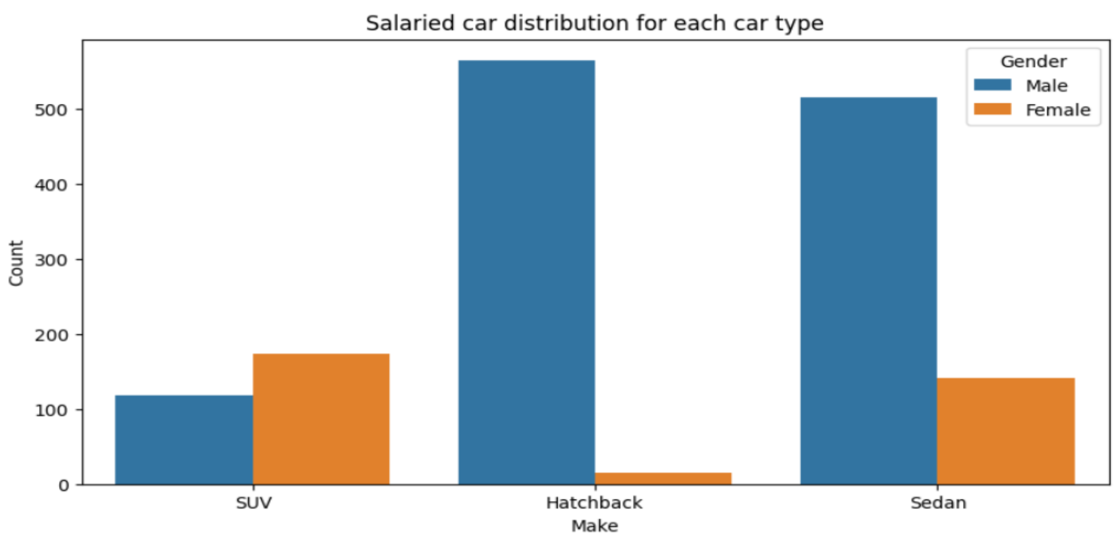



```
# Total number of Salaried individuals who brought Sedan
sedan_buyers = df.query("Profession == 'Salaried' and Make == 'Sedan'")
sedan_count=len(sedan_buyers)

# Total salaried individuals
total_salaried=df[df["Profession"] == "Salaried"].shape[0]

likelihood=(sedan_count/total_salaried) * 100

print("Likelihood of a Salaried person buying a Sedan",likelihood,'%')
```

```
Likelihood of a Salaried person buying a Sedan 44.19642857142857 %
```

Among salaried customers, Sedans account for about 44% (396 purchases), followed by Hatchbacks at nearly 33% (292 purchases), while SUVs contribute around 23% (208 purchases). This indicates that sedans are the most preferred choice among salaried individuals, followed by hatchbacks and SUVs.

3. What evidence from the count plot supports or contradicts the claim that a salaried male is more likely to purchase an SUV than a Sedan?

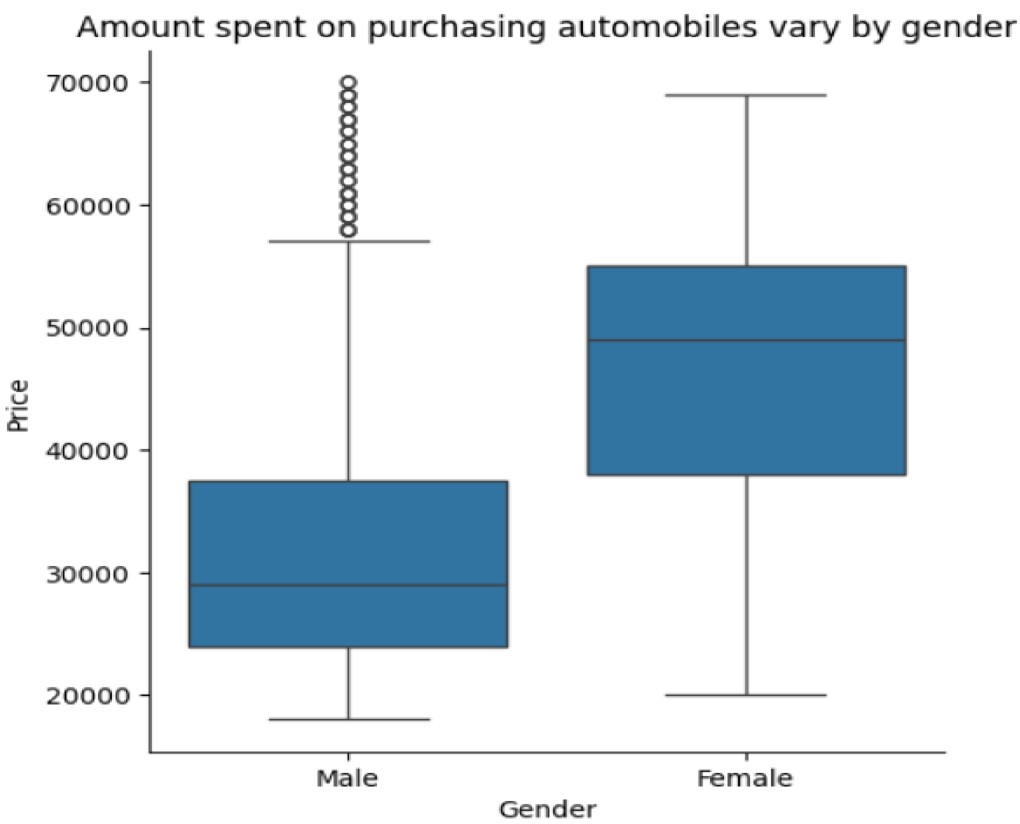


To evaluate whether a salaried male is more likely to purchase an SUV than a Sedan, a count plot was used with Make as the category and the data filtered for salaried male customers.

The count plot clearly shows that Sedans have the highest purchase count, followed by Hatchbacks, while SUVs have the lowest count among salaried males. Numerically, around 305 salaried males purchased Sedans, compared to only 90 purchasing SUVs.

This indicates that Sedans are significantly more preferred than SUVs within this customer segment. Therefore, the visual evidence from the count plot does not support the claim that a salaried male is an easier target for SUV sales compared to Sedan sales.

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

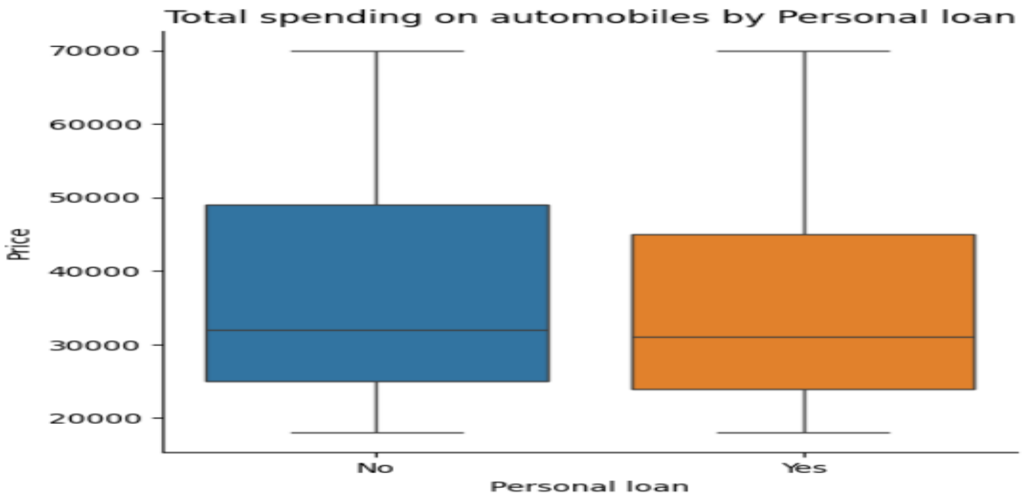


Female customers spend more on automobile purchases compared to males.

The average spending by females is around ₹48,000, while males spend approximately ₹32,000.

Median values also follow the same trend, indicating consistently higher expenditure by females, likely due to their preference for higher-priced vehicle segments such as SUVs.

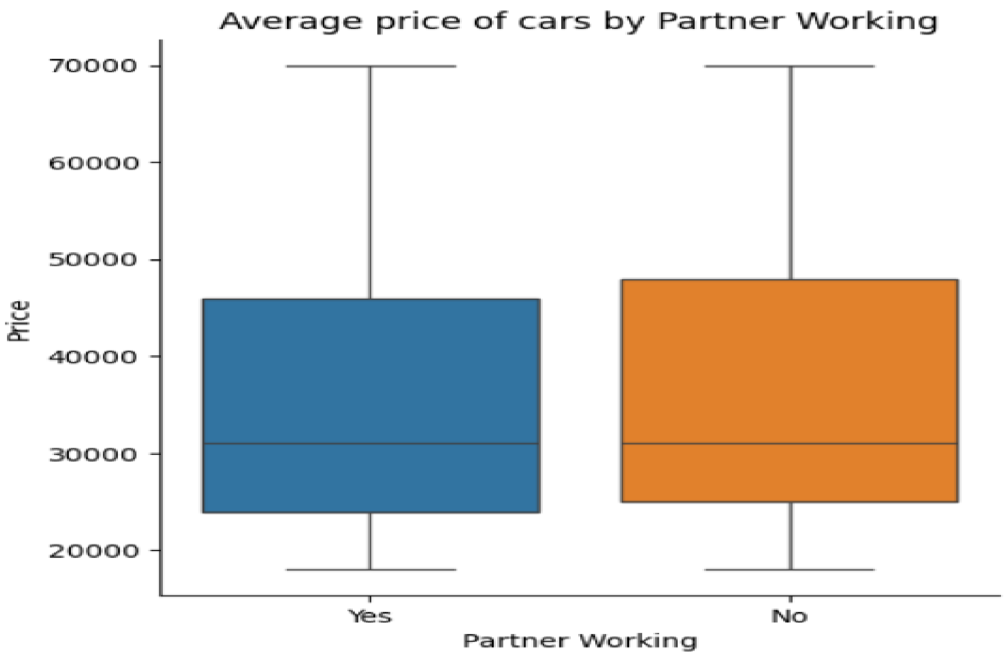
5. **How much money was spent on purchasing automobiles by individuals who took a personal loan?**



The visualization shows very little difference in spending between customers who took a personal loan and those who did not. Both groups spend around ₹34,000 on average.

Median spending is also similar, indicating that taking a personal loan does not significantly influence the vehicle purchase amount.

6. **How does having a working partner influence higher-priced purchases?**



The bar plot indicates that customers with and without a working partner spend nearly the same average amount (around ₹34,000) on automobiles. This suggests that a partner's employment status has minimal impact on purchasing higher-priced vehicles.

Actionable Insights

- Sedan models are the most chosen vehicles among female customers.
- Customers aged above 45 years generally show a preference for higher-priced cars.
- Buyers who have commitments such as personal loans or home loans usually opt for economical vehicles, particularly hatchbacks.
- Age and vehicle price demonstrate a strong relationship in car purchasing decisions, with individuals between 22 and 29 years showing a greater inclination toward hatchback cars.
- Customers with higher income levels are more likely to purchase premium vehicles, and an increase in the number of dependents often leads to choosing larger car models.

Business Recommendations

- To boost SUV sales, special discount schemes can be introduced for middle-aged and senior customers, making premium vehicles more affordable for these groups.
- Since younger buyers predominantly prefer hatchbacks while older customers lean toward SUVs, predictive analytics and machine learning techniques can be applied to better target customer segments and optimize marketing strategies.
- Offer flexible EMI and loan-linked schemes for customers with existing financial commitments to convert price-sensitive buyers toward mid-range sedans.