

AUSTO MOTOR COMPANY

CASE STUDY

BUSINESS REPORT
PGPDS.O.JUL25.A

MUKESH KUMAR RAM
(17-08-2025)

CONTENTS :

List of Figures	3
List of Tables	4
1) Understanding the data and its Structure:	4
2) Data checks and treating irregularities:	5
Checking missing data fields	5
Checking data consistency	6
Checking statistical summary of the data	6
3) Univariate Analysis	8
4) Bivariate Analysis	15
5) Key Questions	18
6) Recommendations	25

LIST OF FIGURES:

- Fig 1.1 – Missing data fields
- Fig 1.2 – Unique column values
- g 1.3 – Statistical Summary
- Fig 1.4 – Bar Chart - Gender
- Fig 1.5 – Bar Chart - Profession
- Fig 1.6 – Bar Chart – Marital Status
- Fig 1.7 – Bar Chart – Education
- Fig 1.8 – Bar Chart – Personal Loan
- Fig 1.9 – Bar Chart – House Loan
- Fig 1.10 – Bar Chart – Partner Working
- Fig 1.11 – Bar Chart – Vehicle Type
- Fig 1.12 – Histogram – Age
- Fig 1.13 – Histogram – Salary
- Fig 1.14 – Histogram – Partner Salary
- Fig 1.15 – Histogram –Total Salary
- Fig 1.16 – Histogram –Price
- Fig 1.17 – Pairplot
- Fig 1.18 – Correlation Table
- Fig 1.19 – Correlation Matrix
- Fig 1.20 – Vehicle Make preference by Gender
- Fig 1.21 – Vehicle Make preference for Salaried professional
- Fig 1.22 – Vehicle Make preference for Salaried Male
- Fig 1.23 – Vehicle Make preference for Salaried Female
- Fig 1.24-Price distribution by Gender
- Fig 1.25- Average price by gender
- Fig 1.26 – Total amount spent on vehicles- By personal loan
- Fig 1.27 – Average price by partner working status

LIST OF TABLES:

Table 1.1- Data structure

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 analytics professional to improve the existing campaign.

Data 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

1) Understanding the data and its Structure:

- Data contains 1581 rows and 14 columns
- Data Type

Column Name	Data Type
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

Table 1.1- Data structure

2) Data checks and treating irregularities:

- There are no duplicates in the data

Missing Data Identification:

We've conducted a thorough check for missing data across all fields in the dataset.

Our analysis identified missing values in the following columns:

- **Gender**
- **Partner_salary**

This information is critical for understanding data completeness and informs subsequent data handling strategies.

Here is the list of columns with Null Values:

Column Name	Null 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

Fig1.1 – Missing data fields

Gender: 53 records have missing gender information. We recommend retaining these records. Removing them would lead to a loss of data for other important variables in those cases, and the missing data represents only 3% of the total dataset. Creating a new category for these missing values is not recommended due to the small proportion of affected records. We have replaced the missing values with male.

Partner Salary: 106 records have missing partner salary data. This variable is linked to marital status and whether a partner is working. We've ensured that when a partner is working, the marital status is consistently recorded as "married."

- To address the missing partner salary data, we've calculated it using the following formula: **Partner Salary = Total Salary - Salary**.
- We've confirmed that when the partner is *not* working, the partner salary value is consistently recorded as zero.

In essence: We've taken steps to preserve and improve the completeness of our data. We've addressed missing values in a way that minimizes data loss and maintains the integrity of related variables.

Data Consistency - Categorical Variables:

- We've reviewed the unique values within our categorical data to identify any inconsistencies.
- Our analysis revealed inconsistencies within the "Gender" column:
 - The original data contained variations of "Female" (e.g., "Femal", "Femle") and "Male".
 - There were also some missing values (represented as "nan").
- **Corrective Action:** To ensure data accuracy and consistency, we corrected the gender entries:
 - We standardized the variations by replacing "Femal" and "Femle" with "Female".
 - The "nan" values were replaced with Male.

Checking statistical summary of the data

Columns	count	mean	std	min	25%	50%	75%	max
Age	1581	31.922201	8.425978	22	25	29	38	54
No_of_Dependents	1581	2.457938	0.943483	0	2	2	3	4
Salary	1581	60392.2201	14674.825	30000	51900	59500	71800	99300
Partner_salary	1581	20585.895	18952.9386	0	0	25600	38000	80500
Total_salary	1581	79625.9962	25545.8578	30000	60500	78000	95900	171000
Price	1581	35597.723	13633.6365	18000	25000	31000	47000	70000

Observations

- Mean age is 31.9 years and max is 54 years
- Average number of dependents 2.45 ~ 2-3 dependents
- Average salary is ~60,000
- Average partner salary is ~ 19000
- Average total salary is ~80000 with maximum as 171000
- Average price at which the vehicle is bought is ~36000 and highest price is 70000

3) UNIVARIATE ANALYSIS

a) There are more males (1252) in the data than female (329)

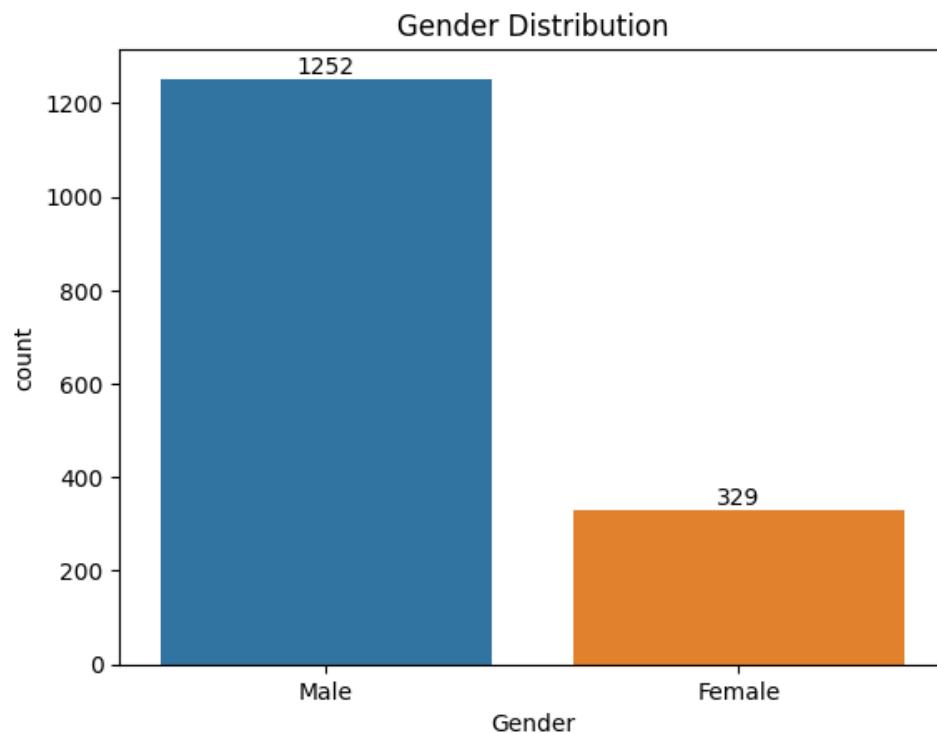


Fig 1.4 – Bar Chart – Gender

b) There are more salaried people in the data set

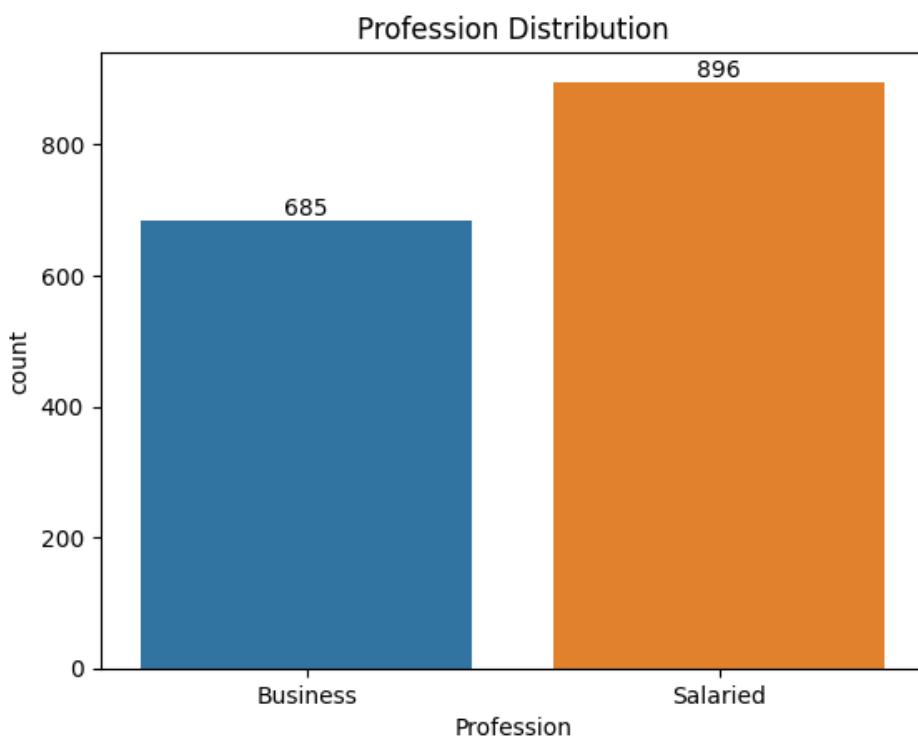


Fig 1.5 – Bar Chart - Profession

c) Marital status of most of the people is 'Married'

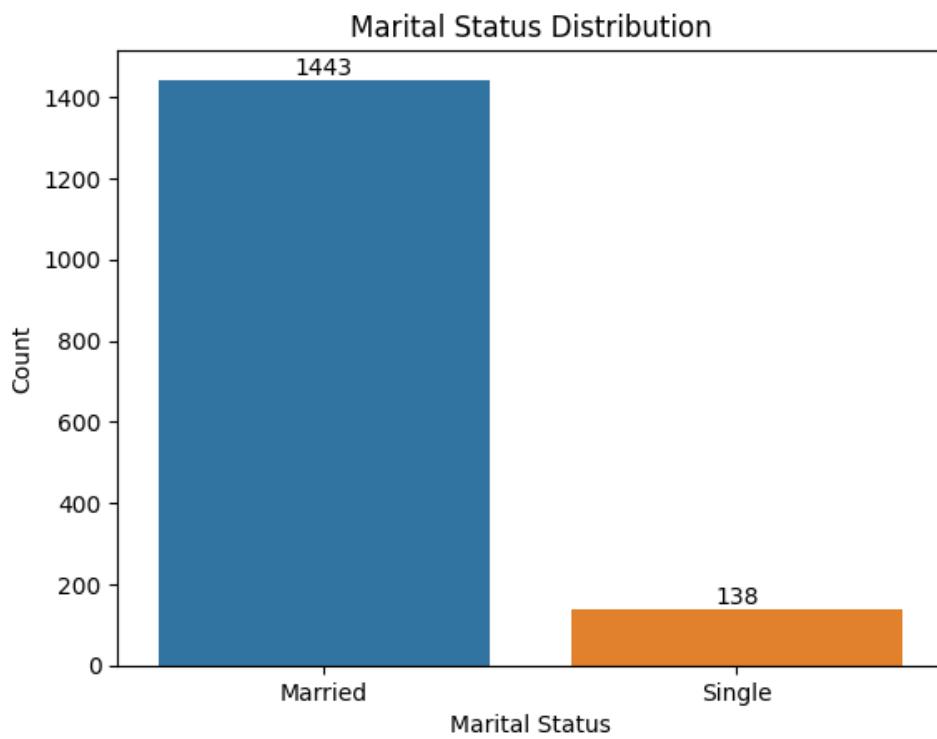


Fig 1.6 – Bar Chart – Marital Status

- d) In the data set, more people are postgraduate degree holders

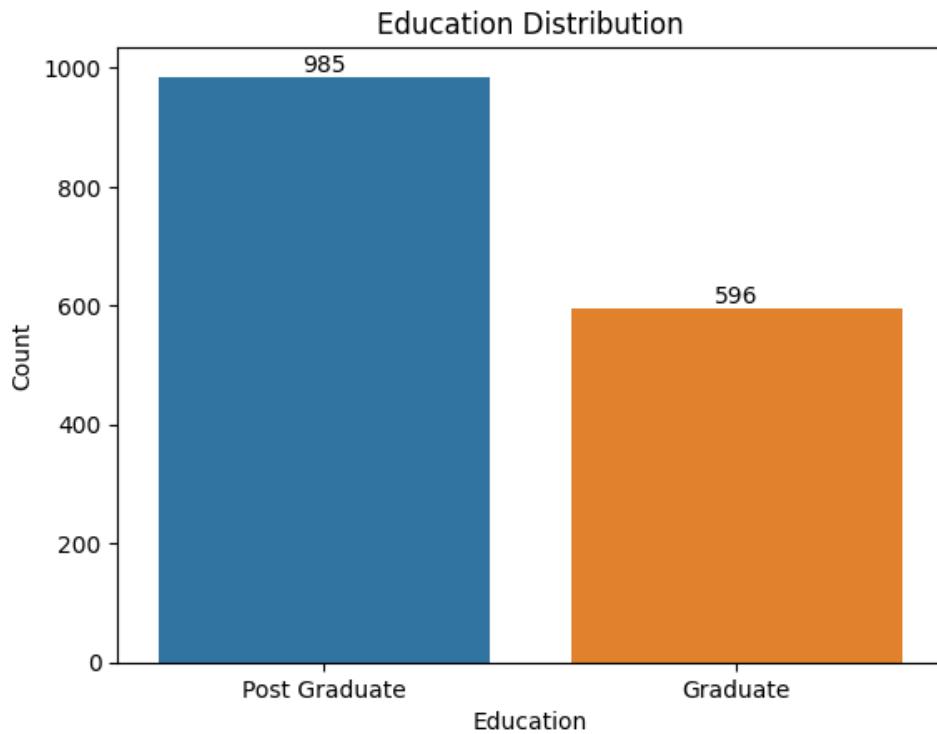


Fig 1.7 – Bar Chart – Education

- e) 50% of the people have a personal loan

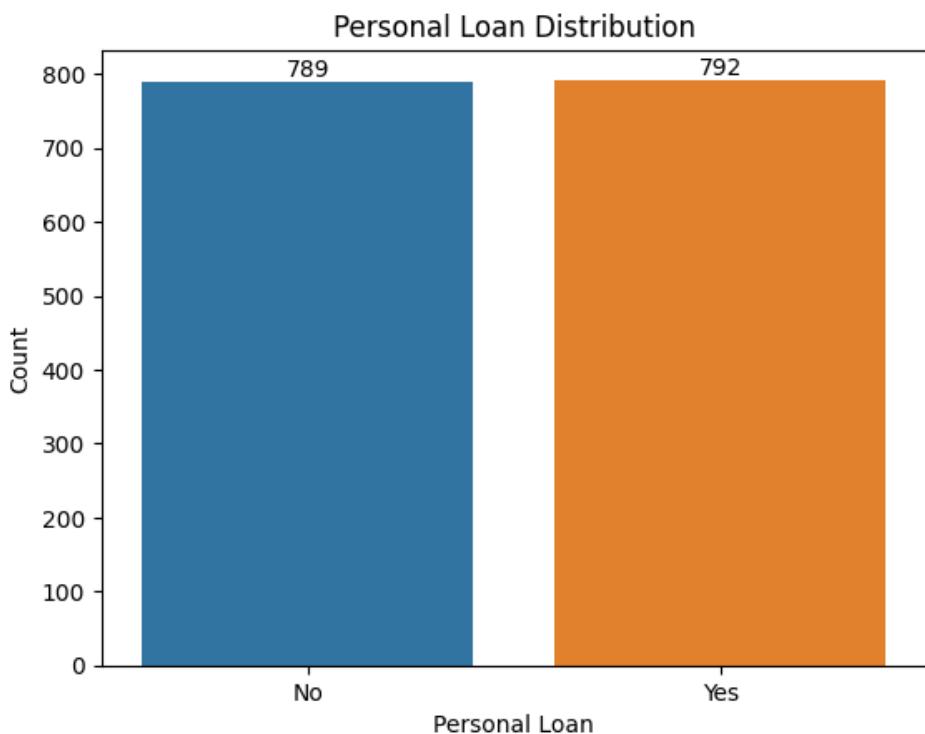


Fig 1.8 – Bar Chart – Personal Loan

f) 33% of the people have taken a house loan

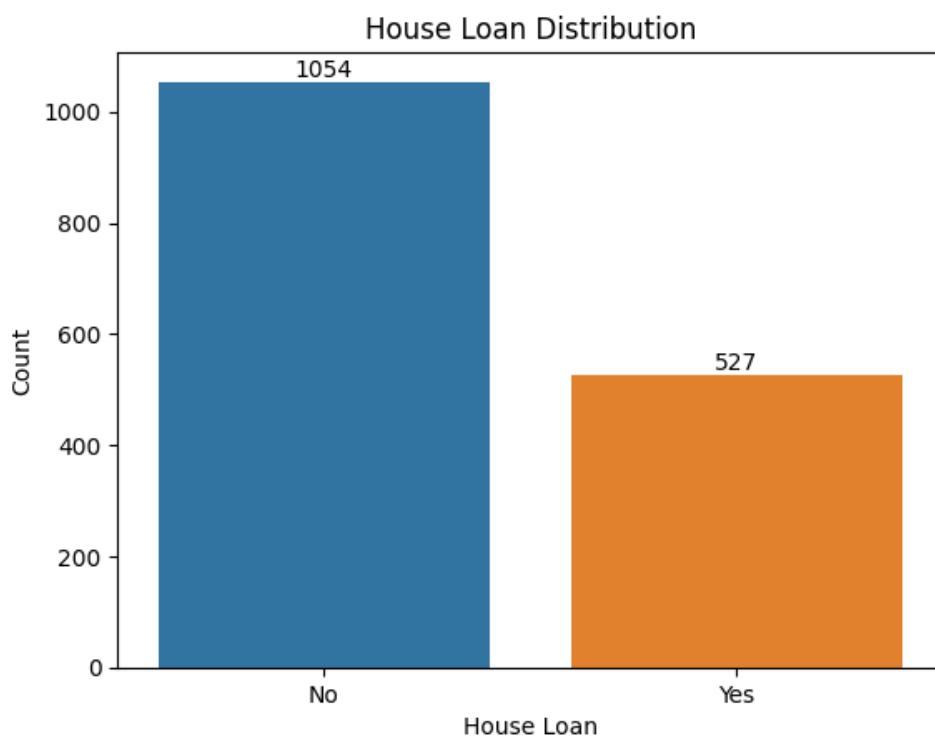


Fig 1.9 – Bar Chart – House Loan

g) More than 800 people have a working partner

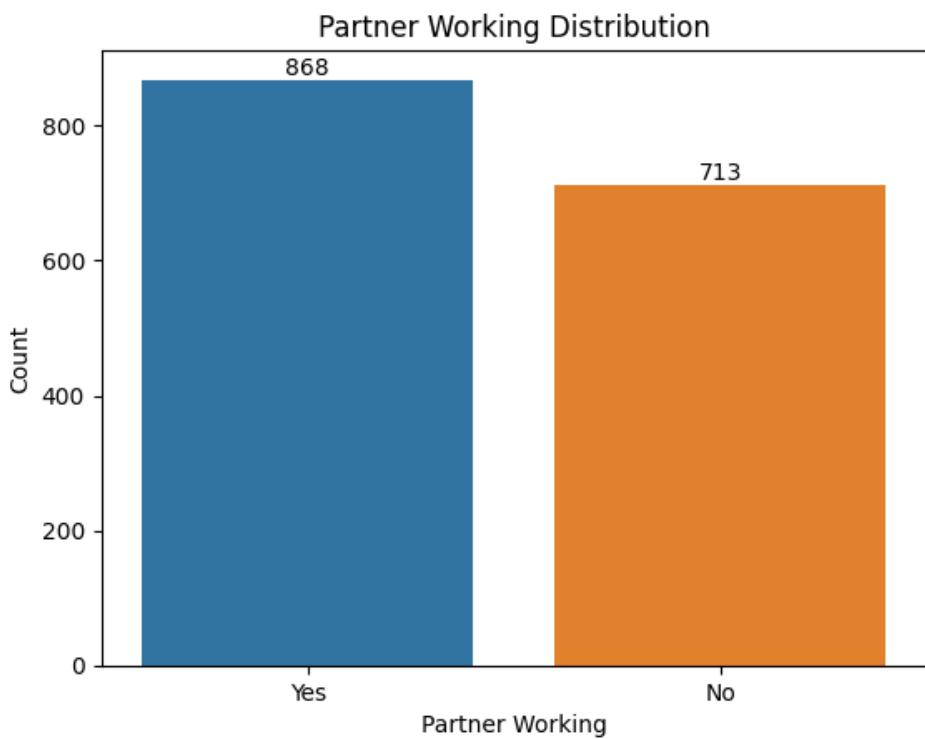


Fig 1.10 – Bar Chart – Partner Working

h) Most of the people have Sedan followed by Hatchback

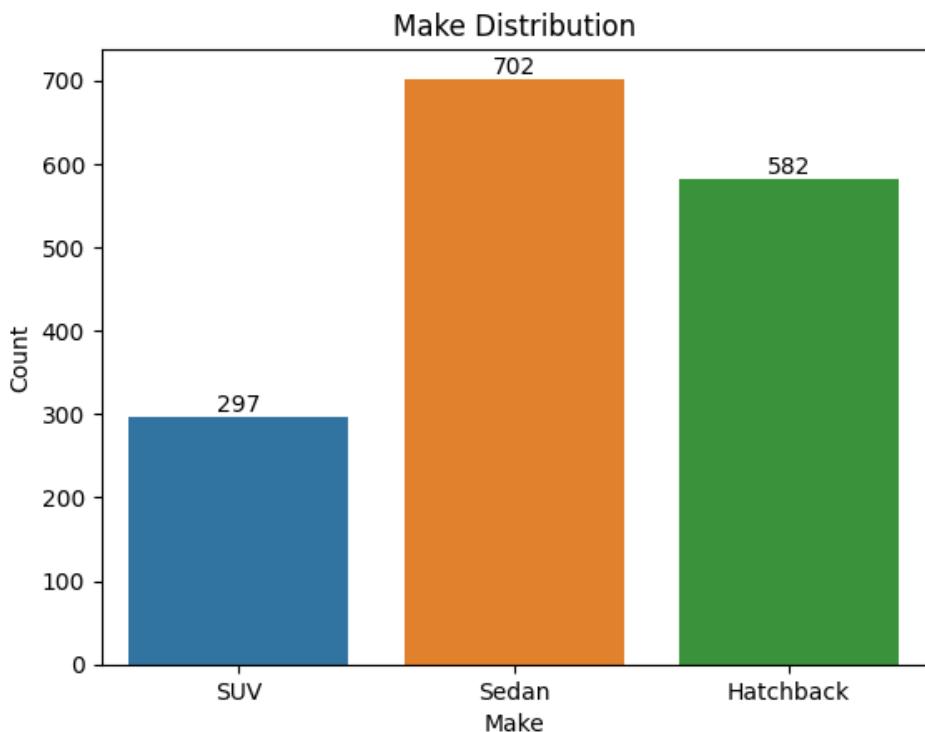


Fig 1.11 – Bar Chart – Vehicle Type

h) 60% of the sample has age less than 30

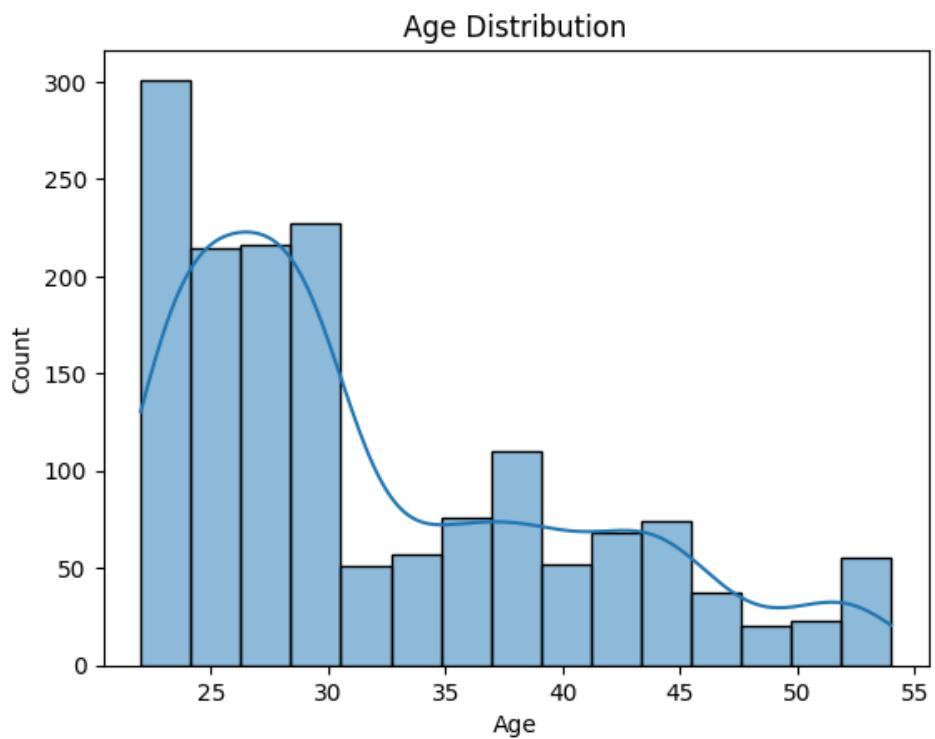


Fig 1.12 – Histogram – Age

j) Salary has a normal distribution curve

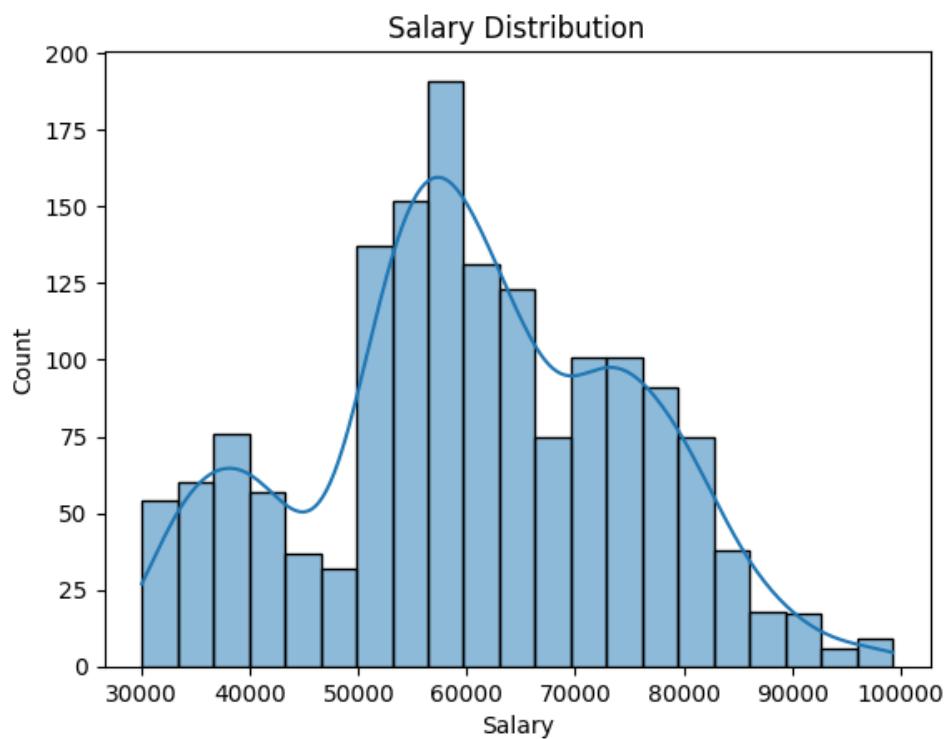


Fig 1.13 – Histogram – Salary

n) Partner salary is right skewed

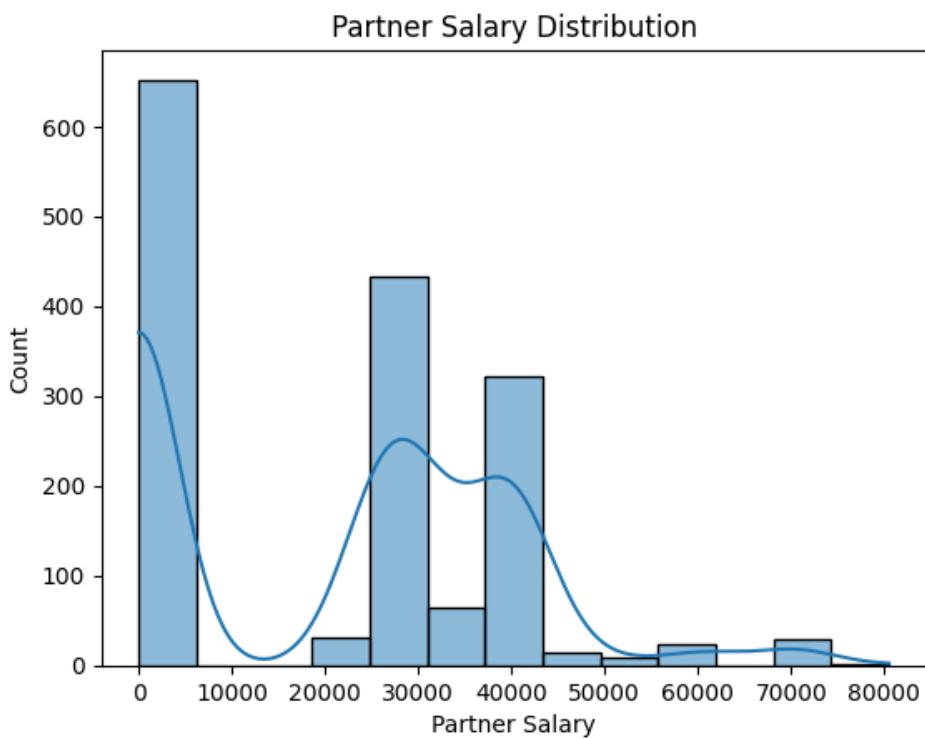


Fig 1.14 – Histogram – Partner Salary

o) Total salary is right skewed

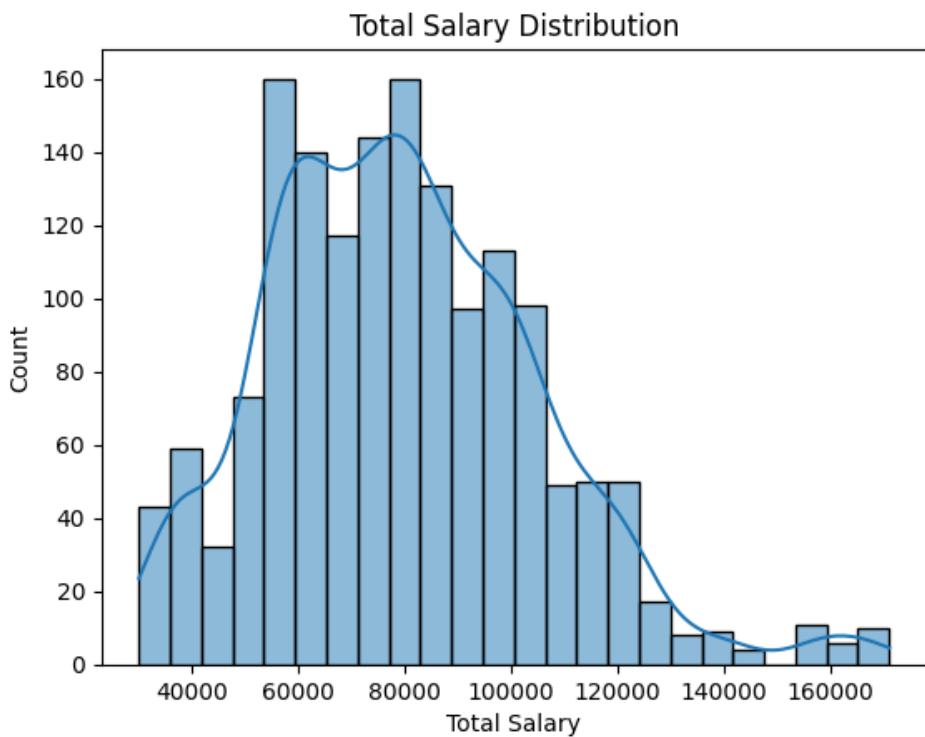


Fig 1.15 – Histogram – Total Salary

p) Price is also right skewed

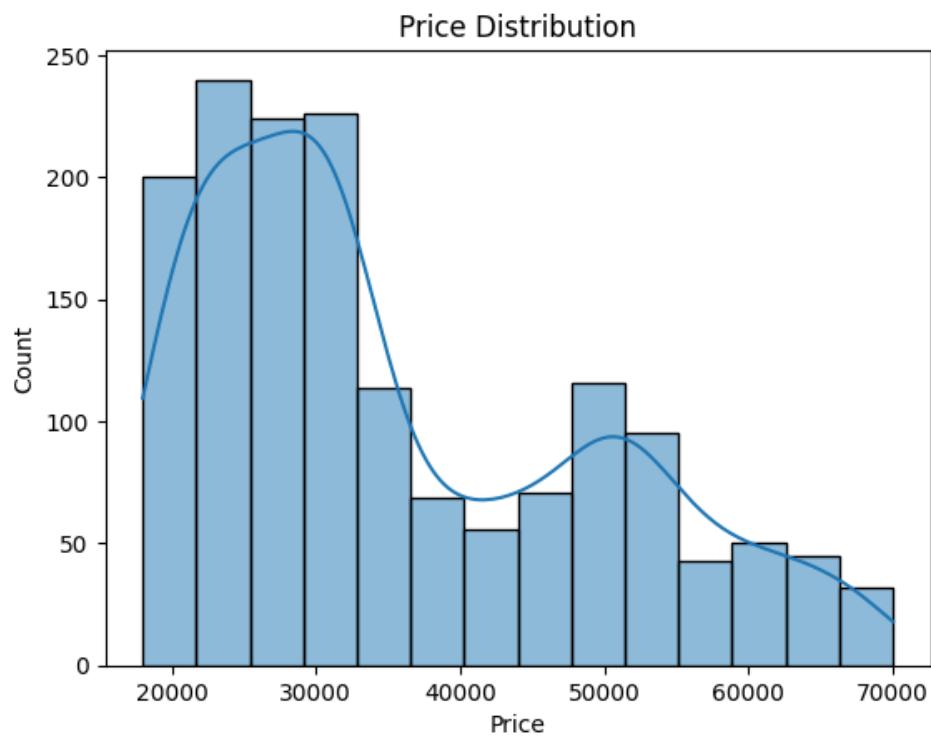


Fig 1.16 – Histogram –Price

4) BIVARIATE ANALYSIS

- Pair plot and correlation of continuous variables

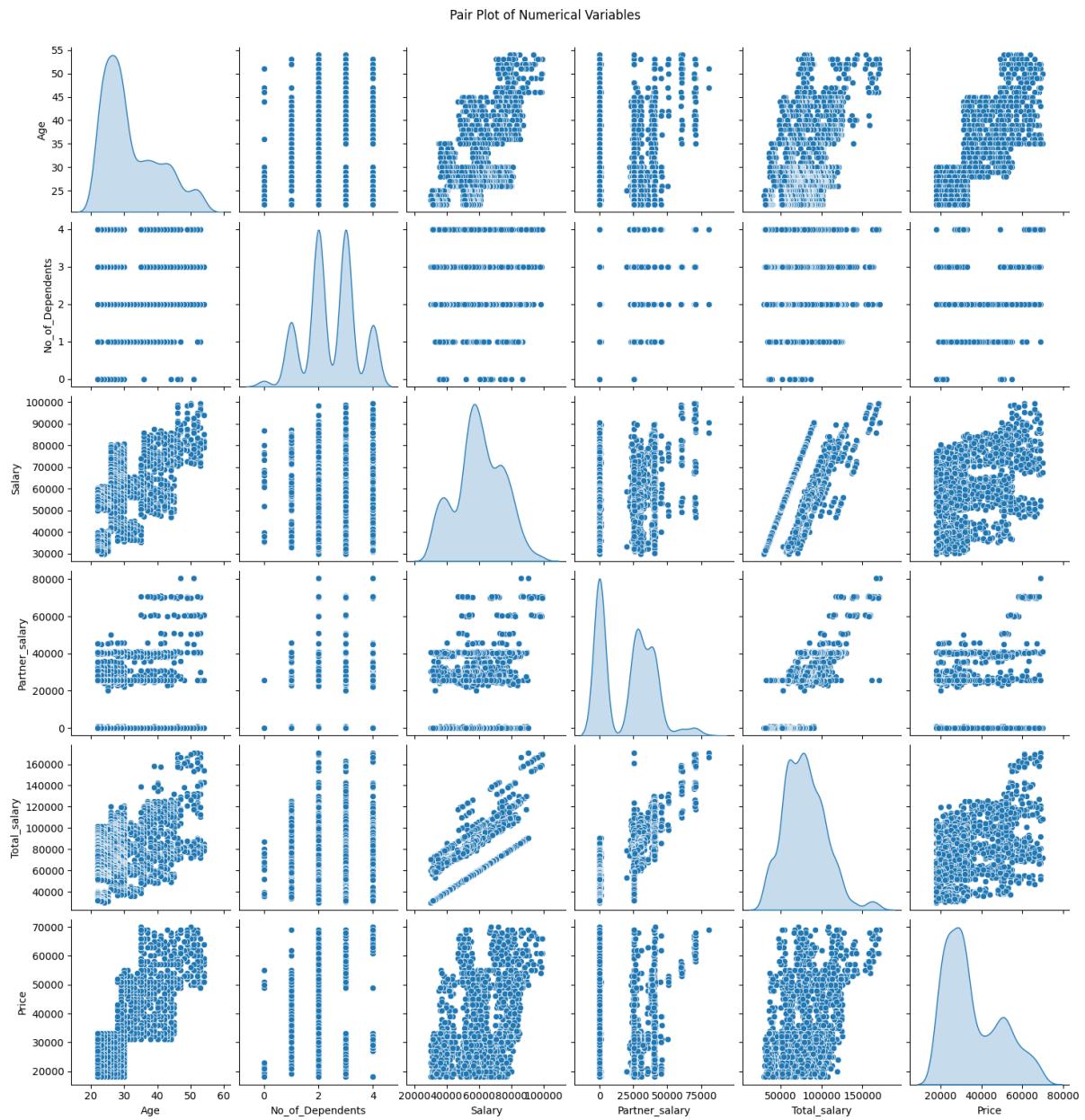


Fig 1.17 – Pairplot

There is a linear relationship between:

- Age and Salary
- Age and Total Salary
- Age and Price
- Salary and Total Salary
- Price and Salary
- Partner_Salary and Total_Salary
- Total_Salary and Price

Correlation Matrix:

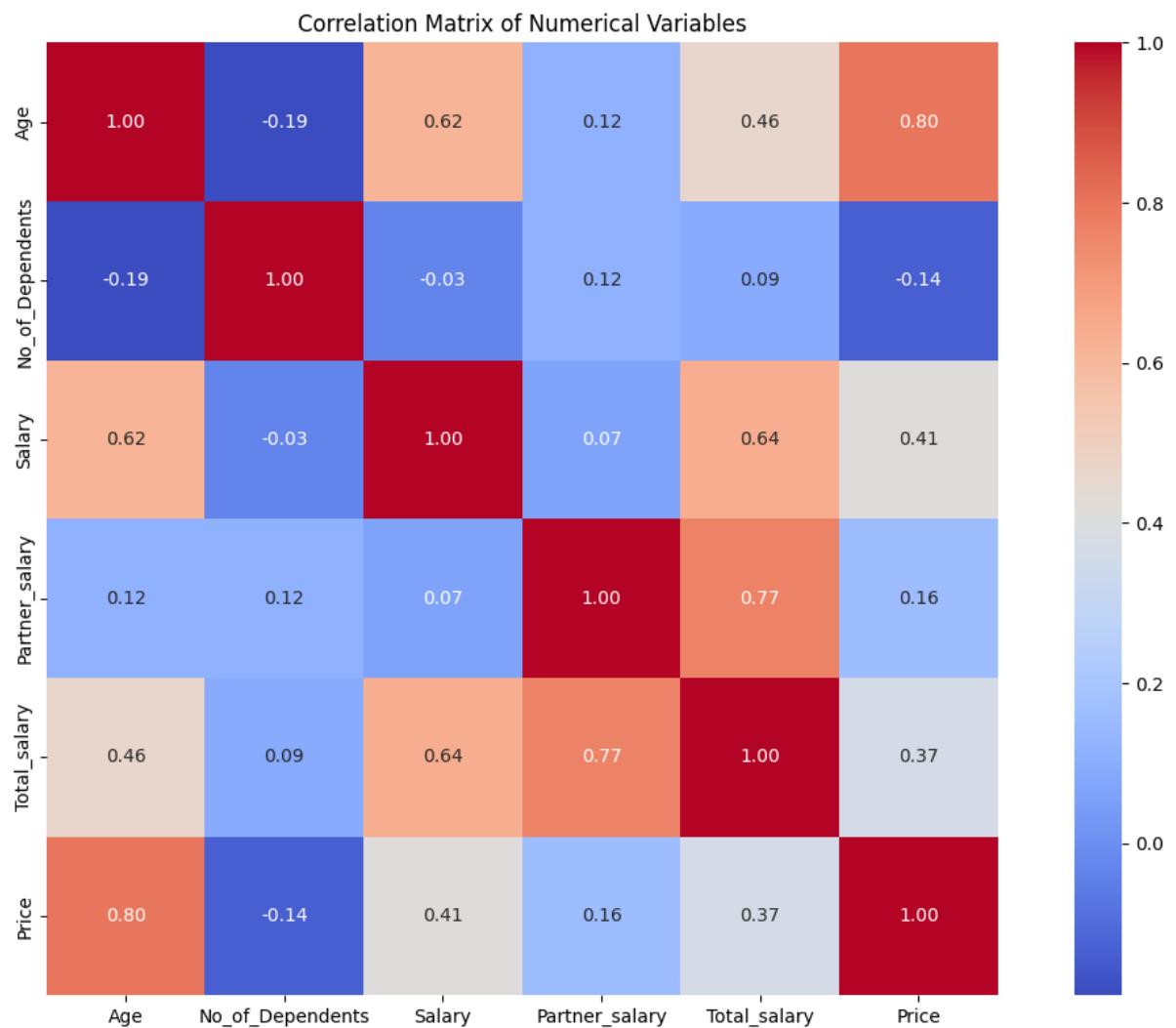


Fig 1.19 – Correlation Matrix

Below variables are highly correlated:

- Age and Price
- Age and Salary
- Salary and Total Salary
- Partner Salary and Total Salary

5) KEY QUESTIONS

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

Below are the numbers with Make by Gender:

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

Below are the numbers with Make by Gender in percentages:

Make	Hatchback	SUV	Sedan
Gender			
Female	4.56	52.58	42.86
Male	45.29	9.90	44.81

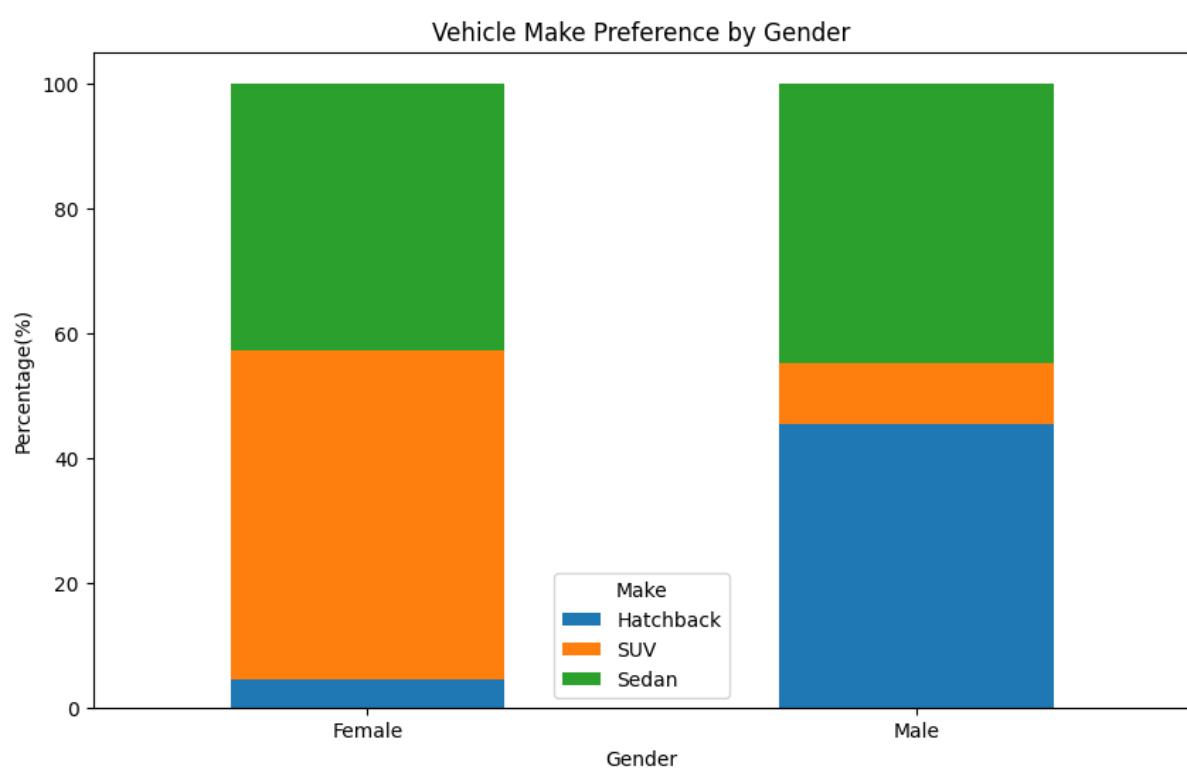


Fig 1.20 – Vehicle Make preference by Gender

Observation:

The stacked bar chart and the data table above provide a clear answer to the question of whether men prefer SUVs more than women:

Male Customers: The most popular choice among male customers is the **Hatchback**, at **45.3%**. Sedans are a close second at 44.8%, and SUVs are the least preferred option at only 9.9%.

Female Customers: In contrast, the most popular choice among female customers is the **SUV**, at **52.6%**. Sedans are the second choice at 42.9%, and Hatchbacks are the least preferred option, chosen by only 4.6% of female buyers.

Conclusion: The data shows that women in this dataset have a much stronger preference for SUVs compared to men, who prefer Hatchbacks and Sedans. This is a key insight for targeted marketing.

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

Distribution of Vehicle Makes for Salaried Professionals (%):

Make	Salaried Professionals (%)
Sedan	44.2
Hatchback	32.59
SUV	23.21

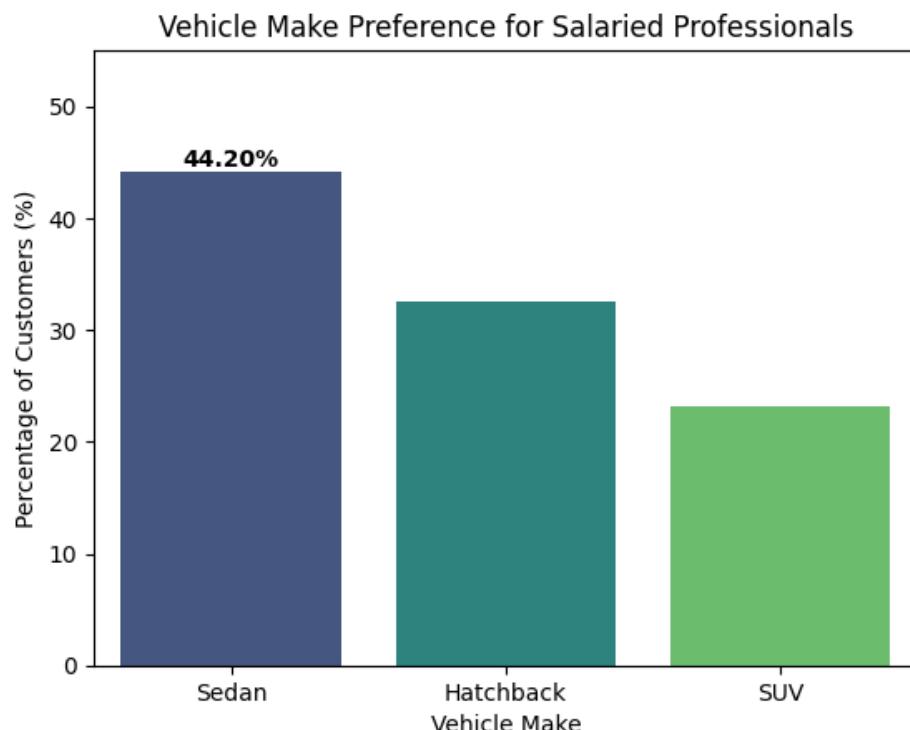


Fig 1.21 – Vehicle Make preference for Salaried professional

Observation:

The likelihood of a salaried person buying a Sedan is: 44.20%

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?

The sample size of the salaried males is: 672

Make distribution of salaried Male (%):

Make	salaried Male(%)
Sedan	45.39
Hatchback	41.22
SUV	13.39

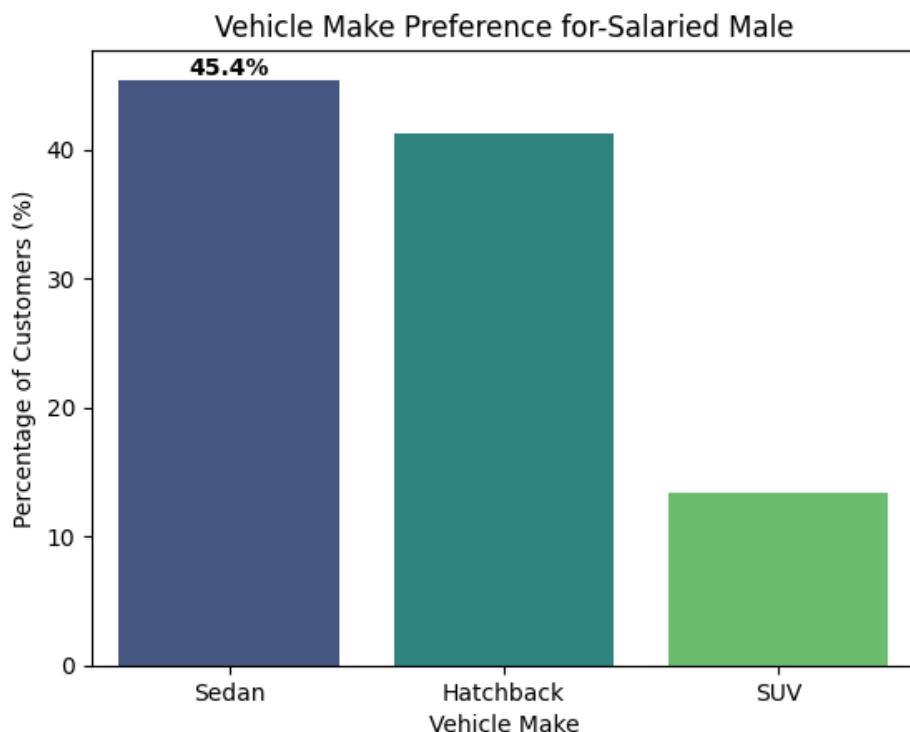


Fig 1.22 – Vehicle Make preference for Salaried Male

The sample size of the salaried females is: 224

Make distribution of salaried Female (%):

Make	salaried Female (%)
SUV	52.68
Sedan	40.62
Hatchback	6.7

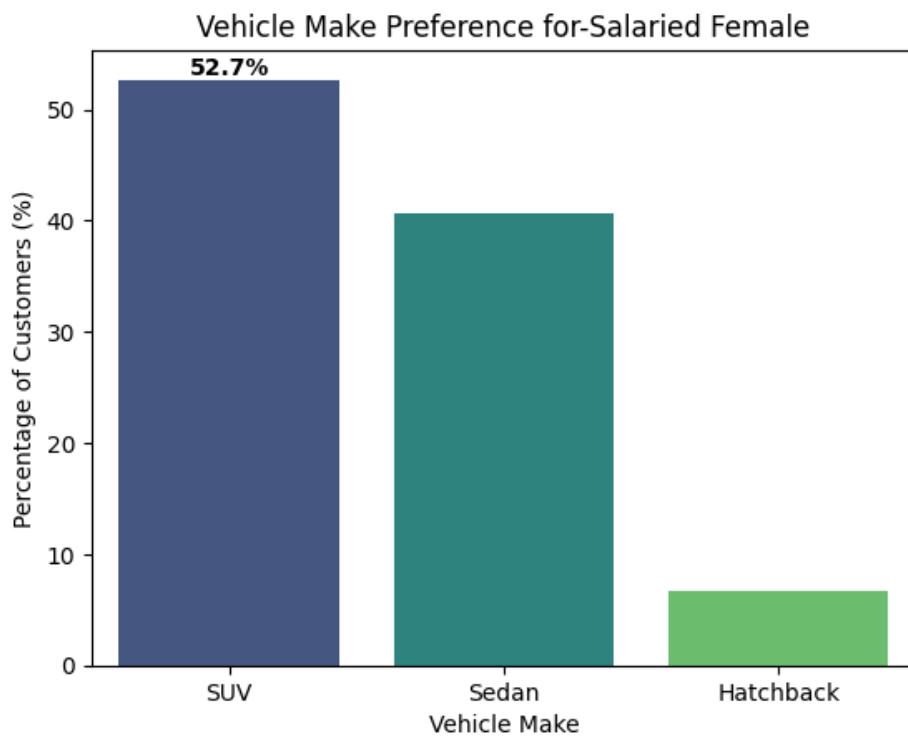


Fig 1.23 – Vehicle Make preference for Salaried Female

Observation:

The statement made by Sheldon Cooper is false as -

The percentage of salaried males who prefer SUVs is: 13.39%

The percentage of salaried males who prefer Sedans is: 45.39%

The percentage of salaried males who prefer SUVs over Sedans is: -32.00%

The percentage of salaried females who prefer SUVs is: 52.68%

The percentage of salaried females who prefer Sedans is: 40.62%

The percentage of salaried females who prefer SUVs over Sedans is: 12.06%

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

Summary Statistics by Gender:

Gender	Female	Male
count	329	1252
mean	47705.17	32416.13
std	11244.84	12366.25
min	20000	18000
25%	38000	23000
50%	49000	29000
75%	55000	37000
max	69000	70000

Relationship between Gender and Price

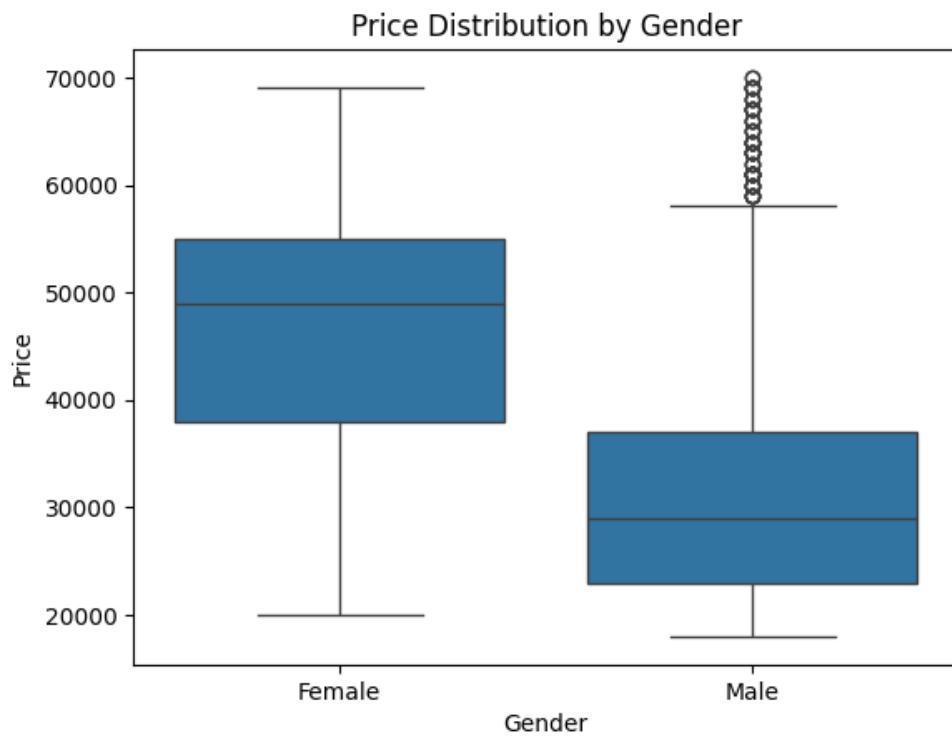


Fig 1.24-Price distribution by Gender

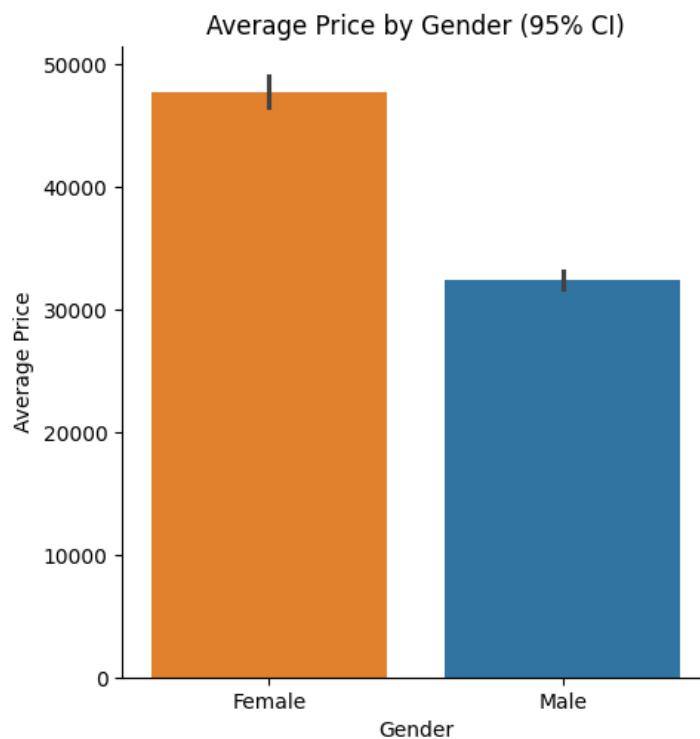


Fig 1.25- Average price by gender

Observation:

On an average, females spend more than males on cars

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

Impact of Personal Loans on Customer Spending:

Customers with a Personal Loan:

Number of Customers: 792

Total Spending: \\$27.29 million

Average Spending: \\$34,457

Customers without a Personal Loan:

Number of Customers: 789

Total Spending: \\$28.99 million

Average Spending: \\$36,743

Observations:

Customers without personal loans have a slightly higher average spending compared to those with personal loans.

The overall spending is also higher in the group without personal loans.

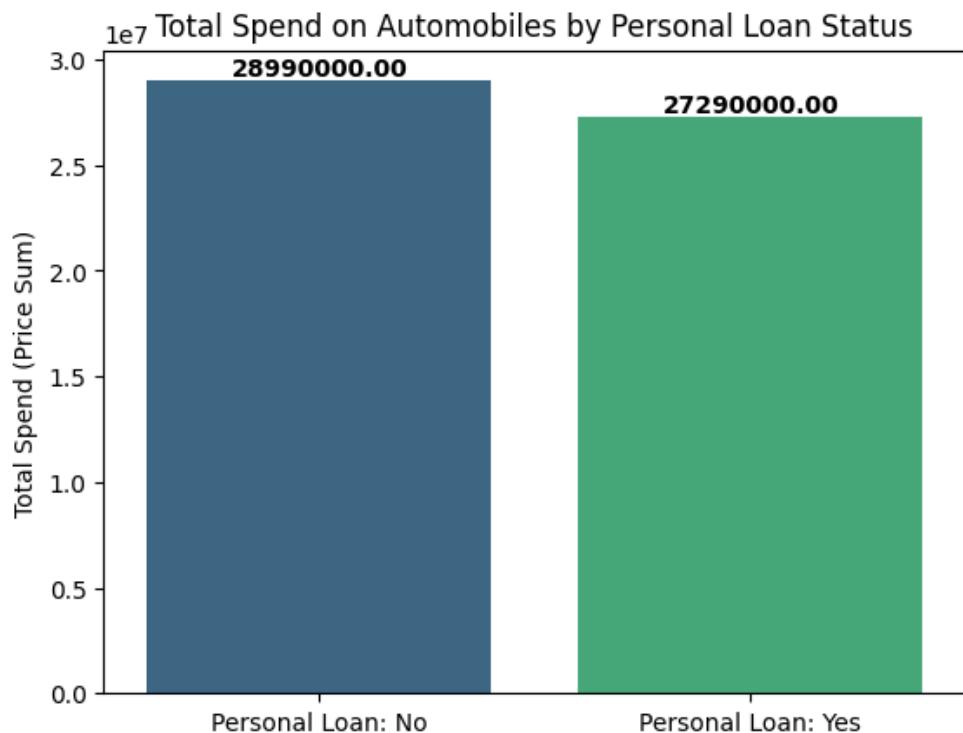


Fig 1.26 – Total amount spent on vehicles- By personal loan

Observation:

People who do not have a personal loan spend slightly more on an average as compared to people who do not

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

Typical Price: The most common price point for our products is \$31,000.

High-End Purchases: About 25% of our sales are for items priced at \$47,000 or more.

Impact of Partner Employment on Purchase Price:

We've analysed how having a working partner affects the price of purchases:

Purchases Above \$31,000: Customers without a working partner are slightly more likely to buy items above the \$31,000 price point.

Purchases Above \$47,000: Similarly, customers without a working partner are a bit more likely to buy items at or above the \$47,000 price point.

Partner employment seems to have a small influence on purchase price, with customers without a working partner slightly more inclined towards higher-priced items.

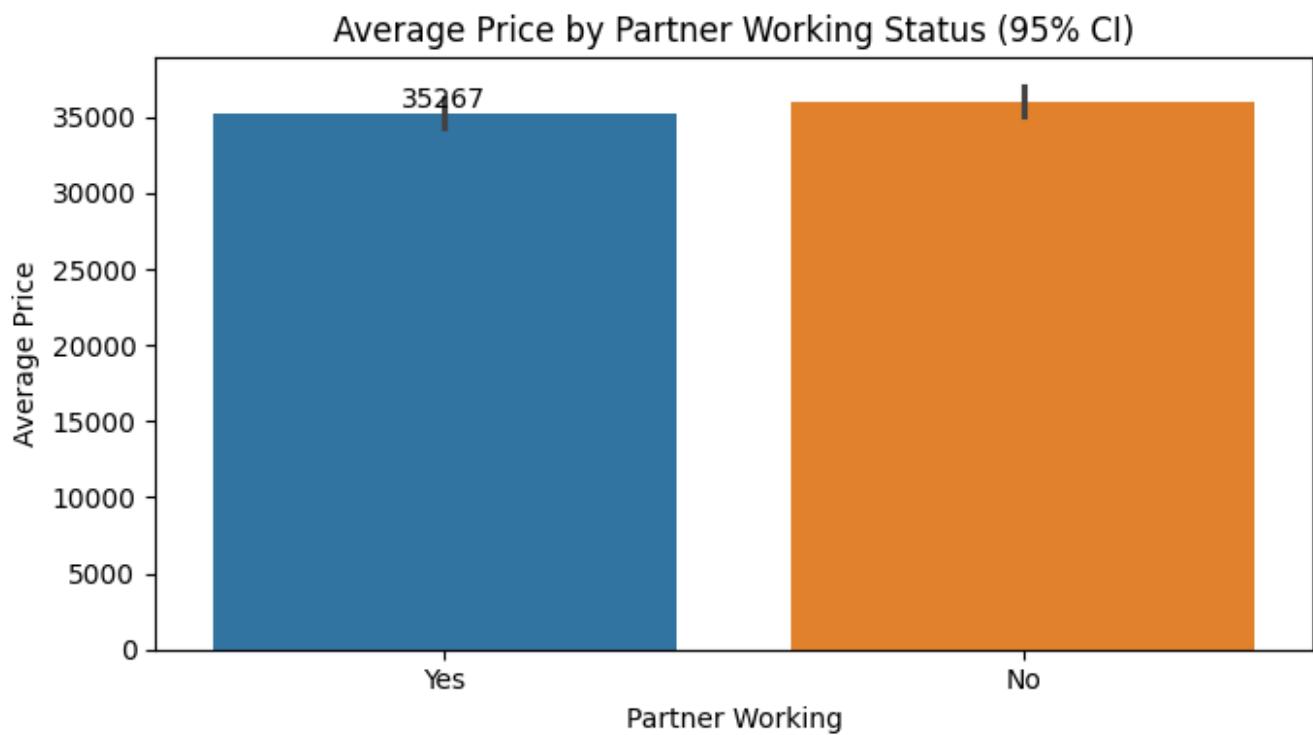


Fig 1.27 – Average price by partner working status

RECOMMENDATIONS:

⌚ KEY BUSINESS INSIGHTS & ACTIONABLE RECOMMENDATIONS

📋 DATA QUALITY INSIGHTS:

- Dataset contains 1581 customer records
- Missing data: Partner_salary (6.7%), Gender (3.4%)
- Data entry errors found in Gender field (2 misspellings)
- No duplicate records found

👤 CUSTOMER DEMOGRAPHICS:

- Gender: 75.8% Male, 24.2% Female
- Marital Status: 91.3% Married
- Profession: 56.7% Salaried, 43.3% Business
- Education: 62.3% Post Graduate

💰 FINANCIAL PROFILE:

- Average Total Salary: \$79,626
- Average Car Price: \$35,598
- Personal Loan Holders: 50.1%

🚗 CAR PREFERENCES:

- Sedan: 44.4% (Most Popular)
- Hatchback: 36.8%
- SUV: 18.8% (Premium Segment)

🔗 KEY CORRELATIONS:

- Age strongly correlates with Price (0.80) - Older customers buy expensive cars
- Partner_salary correlates with Total_salary (0.81) - Dual income families
- Total_salary correlates with Price (0.37) - Income influences car choice

STRATEGIC BUSINESS RECOMMENDATIONS

1. TARGET MARKET SEGMENTATION:

- PRIMARY TARGET: Males aged 30-45 with post-graduate education
- SECONDARY TARGET: Female customers (higher spending power identified)
- Focus on married customers (91% of customer base)

2. PRODUCT STRATEGY:

- SEDAN FOCUS: Maintain strong sedan inventory (44% market share)
- SUV OPPORTUNITY: Premium SUV segment shows highest prices (\$55,825 avg)
- HATCHBACK VALUE: Entry-level segment with consistent demand (37%)

3. PRICING STRATEGY:

- FEMALE PREMIUM: Female customers pay \$47,612 vs males \$32,817
- AGE-BASED PRICING: Older customers (40+) willing to pay premium
- DUAL INCOME TARGETING: Couples with working partners show higher spending

4. FINANCING RECOMMENDATIONS:

- LOAN PARTNERSHIPS: 792 customers have personal loans
- MORTGAGE INTEGRATION: 527 customers have house loans
- Customers with loans tend to buy slightly cheaper cars (financing-conscious)

...

6. MARKETING CAMPAIGN FOCUS:

- SUV CAMPAIGNS: Target age 44+ with salary \$96,592+
- FEMALE OUTREACH: Untapped high-value segment (higher spending per purchase)
- PROFESSIONAL TARGETING: Focus on salaried professionals for premium vehicles

ROI PROJECTIONS & IMPLEMENTATION ROADMAP

REVENUE OPPORTUNITIES:

- Female customers spend \$14,794 more per purchase
- If 10% of male customers matched female spending:
- Potential additional revenue: \$1,773,833
- SUV upselling opportunity: \$21,221 premium per vehicle
- Converting 5% sedan buyers to SUV: \$744,855 additional revenue

IMPLEMENTATION TIMELINE:

PHASE 1 (0-3 months): Data Quality & System Setup

- Fix data entry validation
- Implement customer segmentation system
- Develop targeted marketing materials

PHASE 2 (3-6 months): Market Testing & Campaigns

- Launch female-focused marketing campaign
- Test age-based pricing strategies
- Develop SUV upselling program

PHASE 3 (6-12 months): Scale & Optimize

- Measure campaign effectiveness

 EDA ANALYSIS COMPLETE - Ready for Business Action! 
