Business Report

Automobile Market Analysis Using Python-Based Exploratory Data Analysis

Prepared by

Arkaprava Mazumder

May 18, 2025

Contents

INTRODUCTION 4

Overview of the Project 4

Objectives of the Analysis 4

TOOLS & TECHNOLOGIES 4

DATA OVERVIEW 5

Description of the data 5

Data Loading and Initial Exploration 6

Statistical Summary of the Data 7

Missing Values & Treatment 8

UNIVARIATE ANALYSIS 11

Relationship between numerical variables 11

1. Analysis Of Age 11
2. Salary Analysis 12
3. Car Price Distribution 14

Relationship between categorical variables 14

4. Car Type Distribution 14
5. Gender Distribution 15
6. Marital Status Analysis 15
7. Loan Status Analysis 16

Outliers in Total Salary 16

BIVARIATE AND MULTIVARIATE ANALYSIS 18

Relationship between all the numerical variables by Car Type 18

Correlation between all numerical variables 19

Relationship between categorical vs numerical variables 21

1. Understanding of spending by Gender 21
2. Car Type Distribution by Gender 22
3. Relationship between Personal Loans and Price 23
4. Car Prices by Partner Working Status 24
5. Income Vs. Car type distribution 25

6. Salary distribution by Gender 26
7. Car Type by Profession 27
8. Age Distribution Vs Personal Loan status 28
KEY QUESTIONS & ANSWERS 29
1. Do men tend to prefer SUVs more compared to women? 29
2. What is the likelihood of a salaried person buying a Sedan? 30
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? 30
4. How does the amount spent on purchasing automobiles vary by gender? 32
5. How much money was spent on purchasing automobiles by individuals who took a personal loan? 33
6. How does having a working partner influence the purchase of higher-priced cars? 34
ACTIONABLE INSIGHTS & RECOMMENDATIONS 35
Insight 1: High Income Drives SUV Purchases 35
Insight 2: Female Customers Show Preference for SUVs 35
Insight 3: Loan Eligibility Influences Purchase Behaviour 36
Insight 4: Car Type Varies by Income Level 36
Insight 5: Partner's Employment Status Affects Spending Power 36
Insight 6: Young Customers & Budget Sensitivity 37
CONCLUSION 37

INTRODUCTION

Maintaining a competitive edge in the rapidly expanding automotive industry requires an understanding of consumer trends and preferences. Austo Motor Company, a leader in producing SUV, sedan, and hatchback models, faces a unique challenge: improving the efficiency of its marketing campaigns to better align with customer demand. In response to this challenge, the company has turned to data analysis to gain insights into customer behaviour and preferences.

Overview of the Project

This project involves analyzing customer-related data to help Austo Motor Company enhance its marketing strategies and improve customer experience. By leveraging data analytics, the company aims to better understand the factors influencing vehicle purchase decisions, customer preferences across different demographics, and the overall demand for their vehicle models.

Objectives of the Analysis

The primary objectives of this analysis are as follows:

1) To identify which vehicle types are preferred by different customer segments, with a focus on gender-specific preferences.

2) For analyzing how different customer preferences for specific vehicle models such as sedans or SUVs.

3) To explore the relationship between financial factors, such as salary and loans, and the purchasing decisions of customers.

4) For obtaining actionable insights for marketing and sales strategies, ultimately aligning them with customer needs and improving campaign efficiency.

TOOLS & TECHNOLOGIES

- **Programming Language:** Python
- **Libraries Used:** Pandas, NumPy, Matplotlib, Seaborn
- **Data Source:** Austo Automobiles Dataset (austo_automobile .csv)

DATA OVERVIEW

In the first step of the analysis, the austo_automobile dataset was imported and preliminary examinations were carried out to understand its structure and contents.

Description of the data

The austo_automobile dataset comprises a range of variables that describe customer demographics, financial attributes, and vehicle preferences. Each of these variables plays a crucial role in understanding customer behaviour and purchase decisions. Below is a detailed description of the key variables in the dataset:

- **Age:** The person's age in years.
- **Gender:** Whether the person is male or female.
- **Profession:** The person's job or occupation (e.g., salaried or business).
- **Marital Status:** Whether the person is married or single.
- **Education:** The person's highest level of education (e.g., Graduate or Post Graduate).
- **Number of Dependents:** How many people (like children or elderly parents) the person supports financially.
- **Personal Loan:** Shows if the person has taken a personal loan (Yes or No).
- **House Loan:** Shows if the person has taken a housing loan (Yes or No).
- **Partner Working:** Indicates if the person's spouse or partner is employed (Yes or No).
- **Salary:** The income or salary of the individual.
- **Partner Salary:** The income of the person's partner, if they are working.
- **Total Salary:** The combined income of the person and their partner.
- **Price:** The price of the car the person is interested in or has purchased.
- **Make:** The type or category of the car (e.g., SUV, Sedan, Hatchback).

Data Loading and Initial Exploration

a) The dataset was successfully loaded from the specified directory path. Upon loading, I previewed the dataset by displaying the first five rows.

	Age	Gender	Profession	Marital_status	Education	No_of_Dependents	Personal_loan	House_loan	Partner_working	Salary	Partner_salary	Total_salary	Price	Make
0	53	Male	Business	Married	Post Graduate	4	No	No	Yes	99300	70700.0	170000	61000	SUV
1	53	Femal	Salaried	Married	Post Graduate	4	Yes	No	Yes	95500	70300.0	165800	61000	SUV
2	53	Female	Salaried	Married	Post Graduate	3	No	No	Yes	97300	60700.0	158000	57000	SUV
3	53	Female	Salaried	Married	Graduate	2	Yes	No	Yes	72500	70300.0	142800	61000	SUV
4	53	Male	Salaried	Married	Post Graduate	3	No	No	Yes	79700	60200.0	139900	57000	SUV

b) The dataset consists of 1581 rows and 14 columns, as confirmed by examining its shape.

```
(1581, 14)
```

c) To delve deeper, I checked the dataset's columns and their data types. This step is vital to ensure that all data elements are in their expected formats, which influences subsequent analysis procedures. For example, categorical and numerical variables need to be identified to apply appropriate analysis techniques.

```
<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
```

From the basic information, I also observed most of the columns are having 1581 records except 'Gender' and 'Partner_salary'. It indicates some missing values are present in these columns.

Statistical Summary of the Data

In order to get a comprehensive understanding of the dataset, a statistical summary was conducted, focusing separately on numerical and categorical variables.

	count	mean	std	min	25%	50%	75%	max
Age	1581.0	31.922201	8.425978	22.0	25.0	29.0	38.0	54.0
No_of_Dependents	1581.0	2.457938	0.943483	0.0	2.0	2.0	3.0	4.0
Salary	1581.0	60392.220114	14674.825044	30000.0	51900.0	59500.0	71800.0	99300.0
Partner_salary	1475.0	20225.559322	19573.149277	0.0	0.0	25600.0	38300.0	80500.0
Total_salary	1581.0	79625.996205	25545.857768	30000.0	60500.0	78000.0	95900.0	171000.0
Price	1581.0	35597.722960	13633.636545	18000.0	25000.0	31000.0	47000.0	70000.0

Statistic of Numerical Columns

	count	unique	top	freq	mean	std	min	25%	50%	75%	max
Age	1581.0	NaN	NaN	NaN	31.922201	8.425978	22.0	25.0	29.0	38.0	54.0
Gender	1528	4	Male	1199	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Profession	1581	2	Salaried	896	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Marital_status	1581	2	Married	1443	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Education	1581	2	Post Graduate	985	NaN	NaN	NaN	NaN	NaN	NaN	NaN
No_of_Dependents	1581.0	NaN	NaN	NaN	2.457938	0.943483	0.0	2.0	2.0	3.0	4.0
Personal_loan	1581	2	Yes	792	NaN	NaN	NaN	NaN	NaN	NaN	NaN
House_loan	1581	2	No	1054	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Partner_working	1581	2	Yes	868	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Salary	1581.0	NaN	NaN	NaN	60392.220114	14674.825044	30000.0	51900.0	59500.0	71800.0	99300.0
Partner_salary	1475.0	NaN	NaN	NaN	20225.559322	19573.149277	0.0	0.0	25600.0	38300.0	80500.0
Total_salary	1581.0	NaN	NaN	NaN	79625.996205	25545.857768	30000.0	60500.0	78000.0	95900.0	171000.0
Price	1581.0	NaN	NaN	NaN	35597.72296	13633.636545	18000.0	25000.0	31000.0	47000.0	70000.0
Make	1581	3	Sedan	702	NaN	NaN	NaN	NaN	NaN	NaN	NaN

Statistics including all the Numerical and Categorical columns

Observations:

- Most of our customers are between the ages of 23 and 40, with an average age of 32.
- On average, customers have 2 to 3 dependents, which shows that many of them likely have families.
- The typical customer earns around $60,000 per year. However, information about their partner's income is often missing or not provided.
- A majority of customers do not have a home loan, though about half have taken personal loans.

- Customer groups such as gender, profession, education level, and marital status are fairly well balanced. However, gender information is missing for 53 customers.
- Among the three types of cars offered, the Sedan is the most popular choice.

Missing Values & Treatment

As part of ensuring data quality and integrity, an analysis was conducted to identify any missing values within the dataset.

Previously we found that there are missing values in “Gender” and “Partner_salary” column.

Now, I have deep dived into it to find the count of 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
dtype: int64
```

As a result, I have found, there are 53 missing values for "Gender" column and 106 missing values in "Partner_salary".

Next, I have checked for duplicate records in the data, but dataset doesn't contain any duplicate row.

Following this, I examined how much information is missing in the "Gender" and "Partner Salary" fields. We calculated both the number of missing entries and what percentage (proportion) they represent out of the total data, to understand the extent of these gaps.

```
{'Count': Gender
Male      1199
Female     327
NaN         53
Femal        1
Femle        1
Name: count, dtype: int64, 'Percentage': Gender
Male      75.838077
Female    20.683112
NaN        3.352309
Femal      0.063251
Femle      0.063251
Name: proportion, dtype: float64}
```

- From the above operations we can say that, from these 4 categorical values, 2 of them ('Femal': 1, 'Femle': 1) are wrong entries. Maybe it's a type mistake when entering the data.
- So, we can replace these two entries with "Female" only.

After replacing the entries, I rechecked and under Female column there were 329 entries, which means the wrong entries added to 'Female' category successfully. The blank (NaN) column, we leave for now as maybe it results from a systematic entry issue or optional form field that some users skipped.

Gender	count
Male	1199
Female	329
NaN	53

dtype: int64

We further looked into the missing "Partner Salary" data by grouping it based on whether the partner is working. Out of the 106 missing entries:

- 90 are cases where the partner is not working — in these instances, we can reasonably assume the partner's income is zero.
- 16 entries are cases where the partner is working, but their income is missing.

Partner_working	0
No	90
Yes	16

dtype: int64

For these 16 records, we noticed that the total household income is greater than the individual's income. This allows us to estimate the partner's salary by subtracting the individual's salary from the total income. (As, **Total Salary = Applicant Salary + Partner Salary**)

Observations:

- Dataset has 1581 rows and 14 columns.
- No duplicate records found.

Missing Values:

- 53 missing values in Gender, and 2 typos (Femal, Femle) — corrected.
- 106 missing values in Partner_salary:
- 16 records in Partner_salary calculated as (Total_salary – Salary) (for partner working)
- After treatment, no missing values remain.

Statistical Insights:

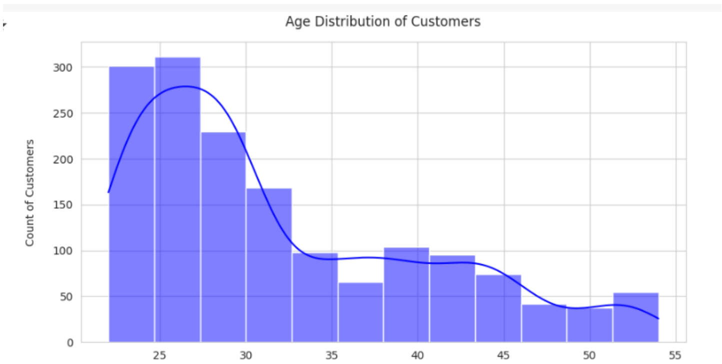
- Most customers are aged 23–40, average ~32 years.
- Typical number of dependents: 2–3, suggesting many buyers are families.
- Average Salary is around $60,000.
- House loan is rare, but many have personal loans.
- Three car types observed: SUV, Sedan, Hatchback.

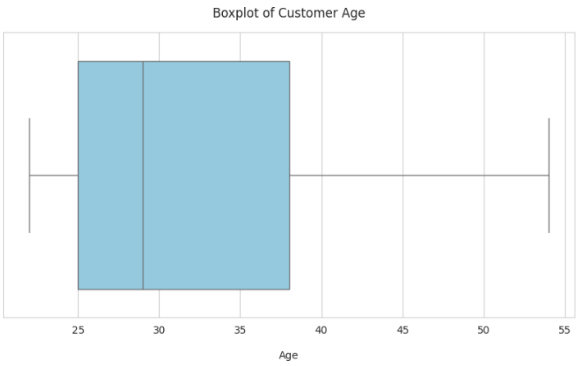
UNIVARIATE ANALYSIS

In this section, we examined each individual customer attribute on its own to better understand the overall profile of our customer base. This includes factors such as age, income, number of dependents, and types of loans. By analyzing each feature separately, we can identify patterns and key trends that help explain who our customers are and what they might need.

Relationship between numerical variables

1. Analysis Of Age





To begin the age analysis, I used both a histogram and a boxplot to visualize the distribution of customer ages.

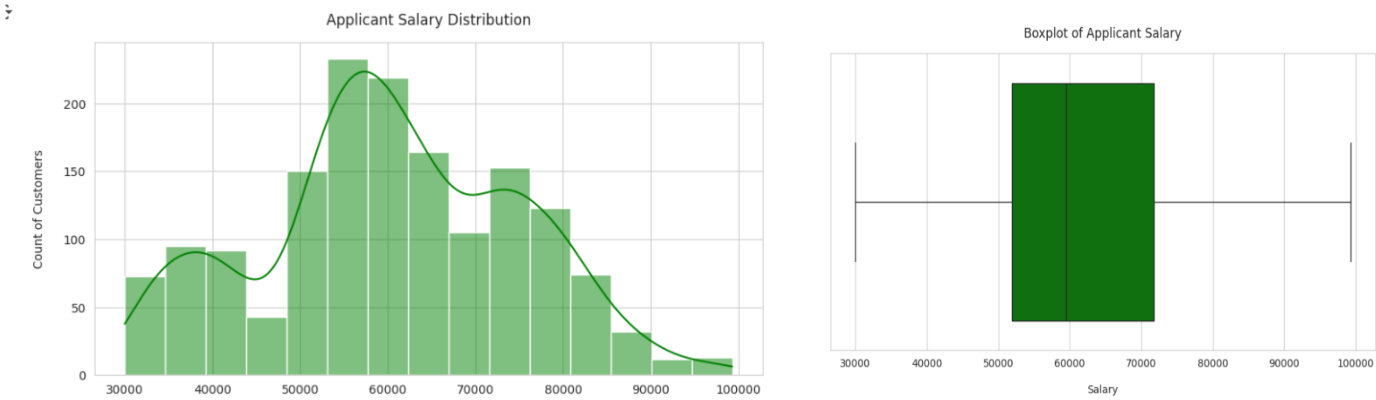
Observations:

- The majority of customers are between 23 and 27 years old.
- There is a noticeable decline in customer count after the age of 30.
- Very few customers are older than 48.
- Overall, the age distribution leans toward a younger audience, highlighting that our customer base is primarily made up of young adults.

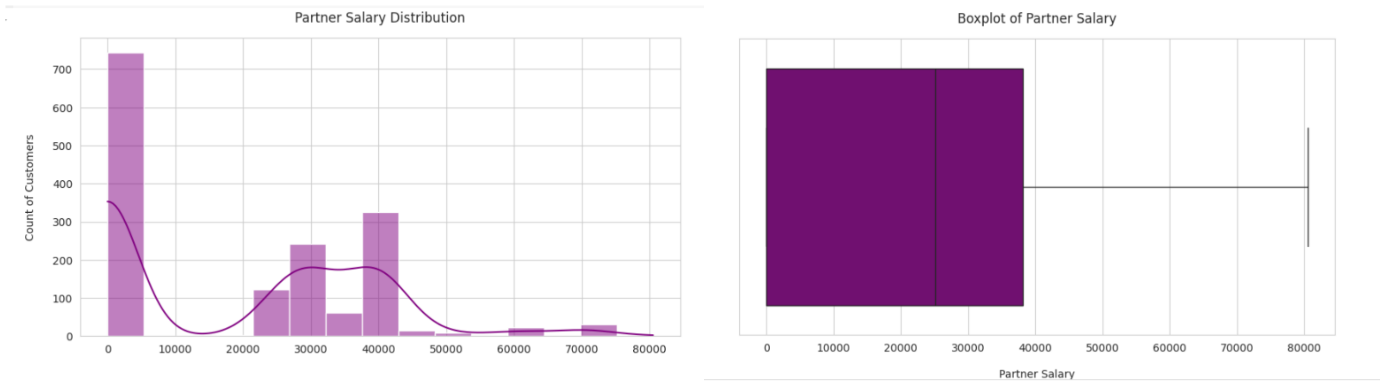
2. Salary Analysis

For the salary analysis, we examined three key income-related columns:

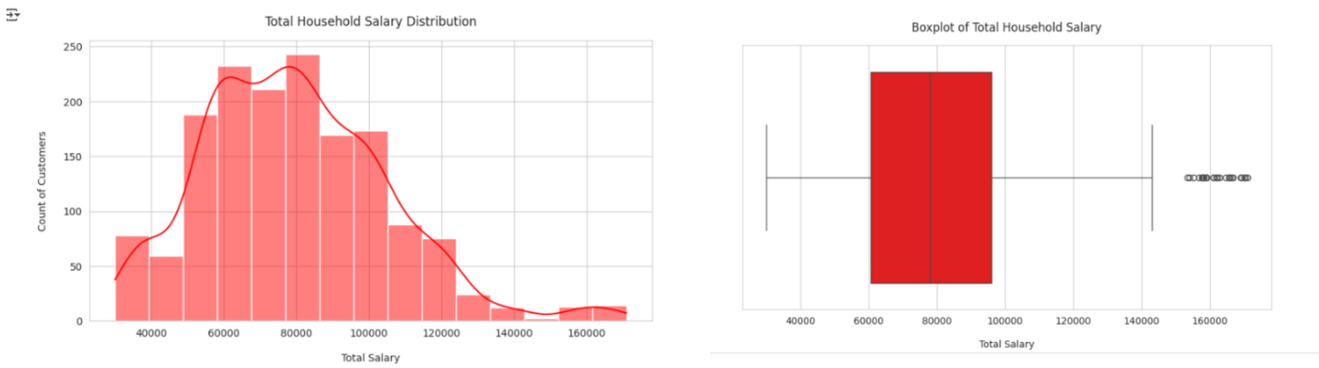
- Customer Salary – the individual income of each customer
- Partner Salary – income reported by the customer's partner (where available)
- Total Salary – combined household income



Salary Distribution of Customers



Partner's Salary Distribution



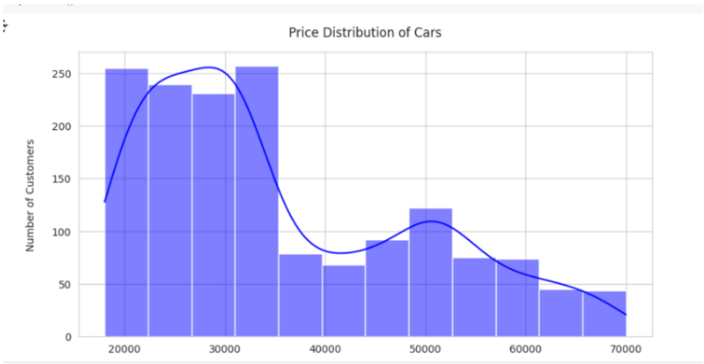
Total Household Salary Distribution

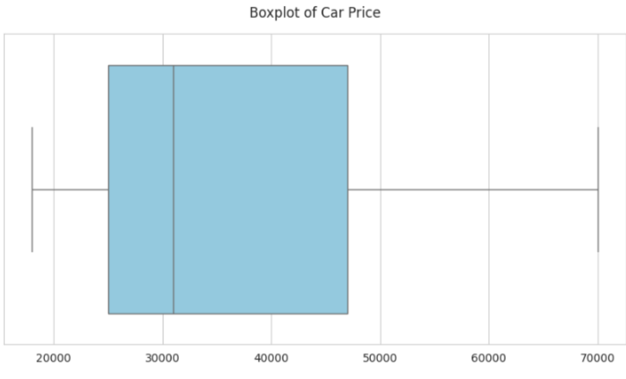
By analyzing above 3 plots we can find out some key observations and insights.

Observations and Insights:

- Most customers earn between $50,000 and $70,000 per year, with the typical (median) salary around $60,000 to $65,000. There aren't any unusual or extreme values, which means the income range for main earners is quite steady.
- When looking at partner incomes, many partners either do not earn or have very low incomes. However, for those who do earn, their salaries often fall between $25,000 and $45,000. There are no extreme values in this group either.
- Combined household incomes (customer + partner) are understandably higher, with most families earning between $60,000 and $100,000. The typical total household income is around $80,000.
- A few families earn significantly more — over $140,000 — and these high earners raise the overall average slightly.
- These insights suggest that Austo primarily serves a middle-income customer base, many of whom are dual-income households.

3. Car Price Distribution



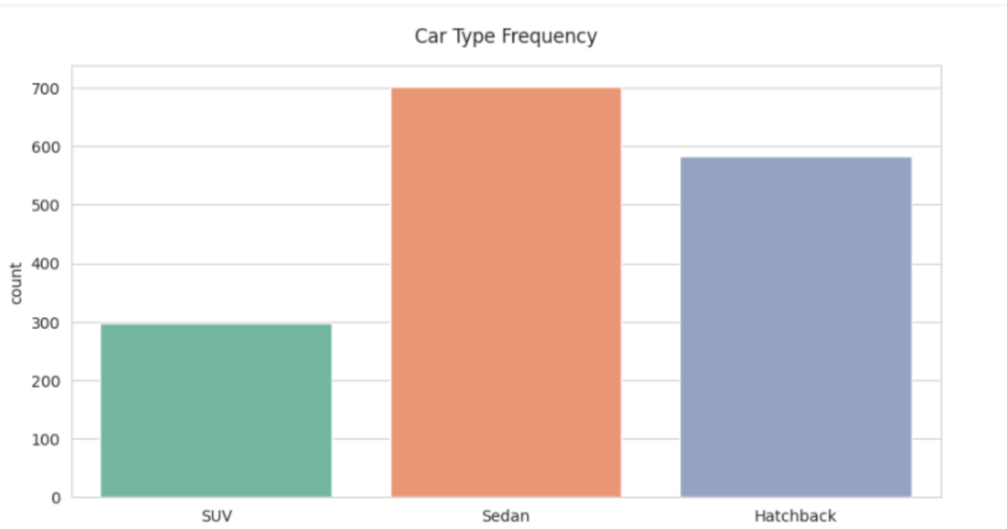


Observations and Insights:

- Car prices tend to fall into two main groups — one around $30,000 and another around $50,000.
- There are a few more expensive cars above $60,000, but not many.
- The middle price (median) is about $35,000, meaning prices are fairly balanced overall.
- Most cars are priced between $30,000 and $45,000.
- There aren't any prices that stand out as way too high or too low, so the prices seem pretty consistent.

Relationship between categorical variables

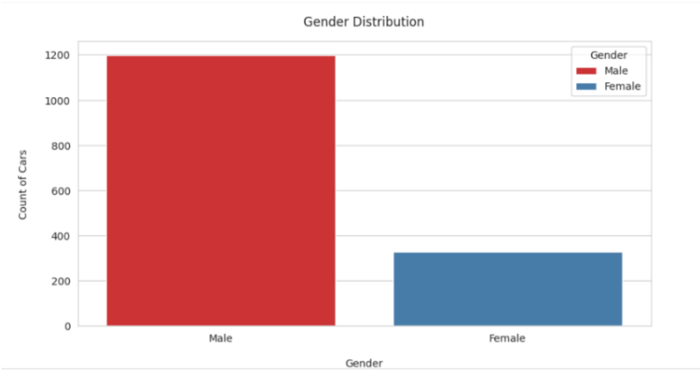
4. Car Type Distribution



Observations:

- Sedans are the most popular car type, with approximately 700 customers choosing them.
- Hatchbacks are the second most preferred, with close to 580 buyers.
- SUVs are the least chosen, with just under 300 customers.

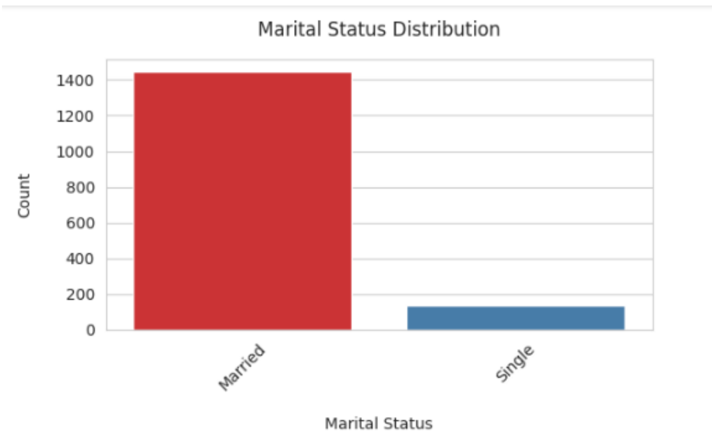
5. Gender Distribution



Observations:

- Most of our buyers are men — about 1,200 men compared to only around 350 women.
- This big difference suggests that our current marketing may be more appealing to men.
- At the same time, it shows a great chance to grow by creating marketing that speaks more to women.

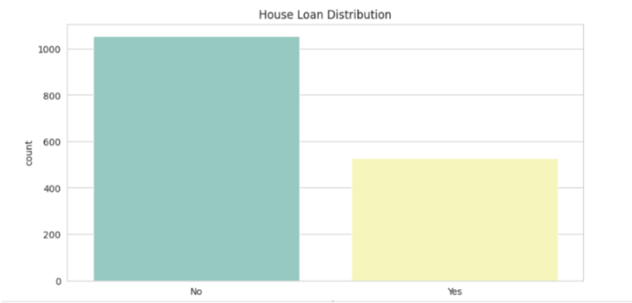
6. Marital Status Analysis



Observations and Insights:

- A large majority of customers are married, compared to a very small number of single individuals.

7. Loan Status Analysis



Observation:

- Around 800 people have taken Personal Loan which is 60% of the customers
- 30% of customers (near about 450) have house loans
- Customers without no loans will be better target for premium models.

Outliers in Total Salary

We have seen previously from the Salary Analysis, there are some higher outliers / right outliers in Total salary column (where Tot al Salary is **greater than ~14500).**

Let's see with the formula as well:

First Quartile (Q1) = 25th percentile of Total_salary

Third Quartile (Q3) = 75th percentile of Total_salary

So, **Interquartile Range (IQR) = Q3 – Q1**

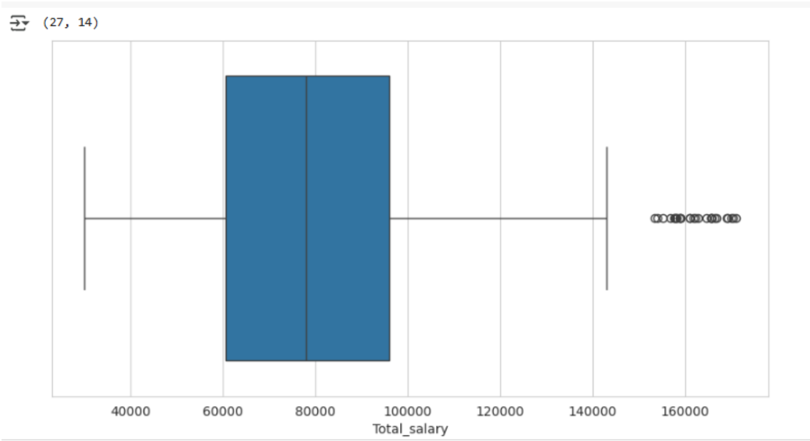
Lower Boundary = Q1 - 1.5 * IQR

Upper Boundary = Q3 +1.5 * IQR

From the Boxplot of Total_salary we can see the Outliers are present at upper boundary.

So, we can say the upper outliers can be calculated as The Salaries which are greater than upper boundary

Or, we can say **Upper Outliers = Total salary > Upper Boundary.**



From this formula we have found that there are 27 records which can be identified as outliers. We can keep them in data as high-income households and they are important for premium targeting.

Some Findings in the Outliers data,

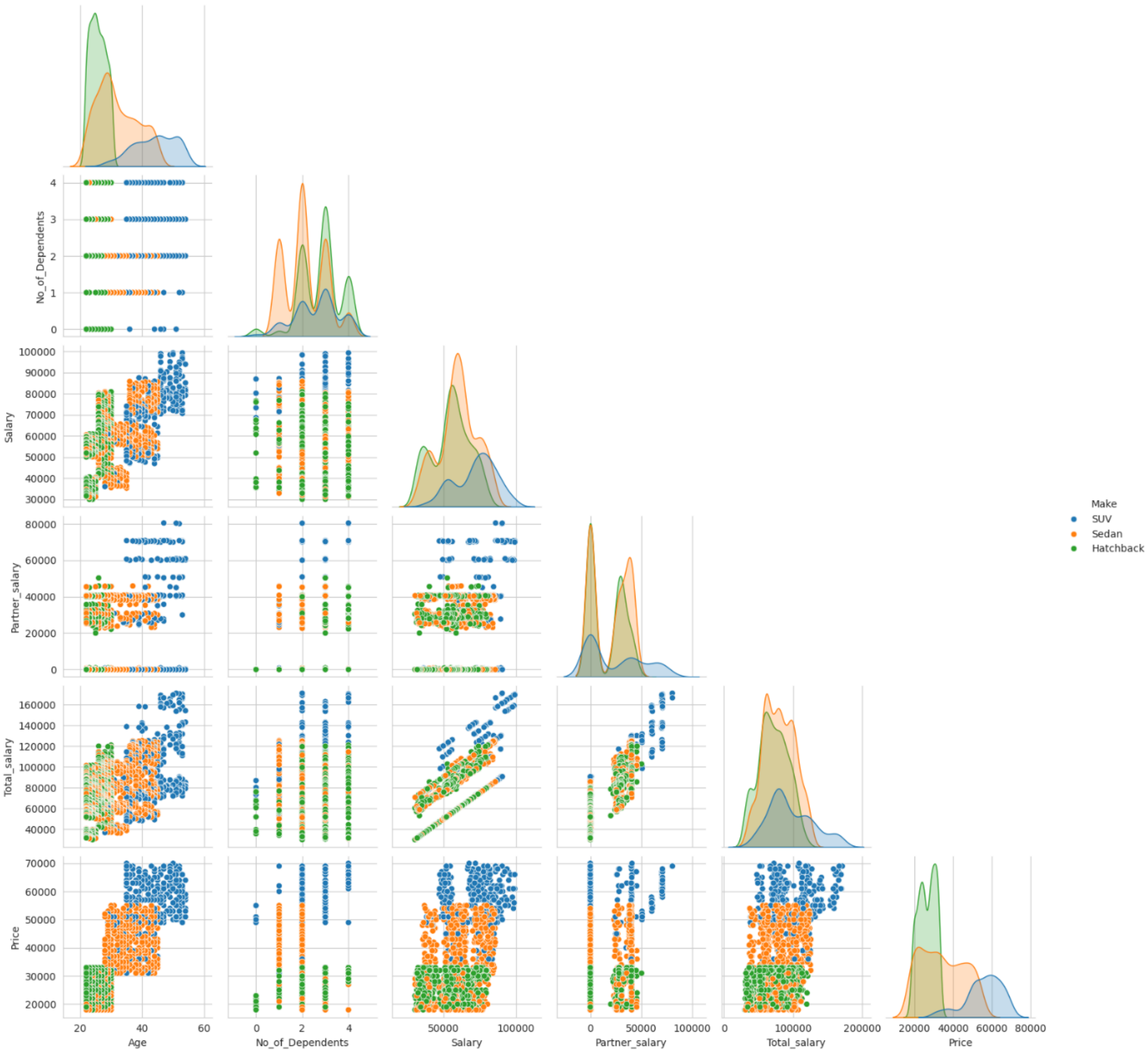
- All have a working partner, which directly contributes to higher combined income.
- Majority hold a Post Graduate degree, suggesting education is a strong income factor.
- Gender is balanced, showing no dominant gender trend among high earners.
- All own SUVs, implying affordability of higher-end vehicles in this group.
- Very few have Personal loans but no one have House loans.

BIVARIATE AND MULTIVARIATE ANALYSIS

To better understand the factors influencing customer behaviour and business outcomes, it's important to look at how different pieces of data relate to each other. In this section, we explore these relationships — first among multiple variables at once (multivariate analysis) then between pairs of variables (bivariate analysis), and— to uncover patterns that can guide smarter decisions.

Relationship between all the numerical variables by Car Type

Analysed the relationship among all numerical variables using pair plot segmented by Car Type to uncover key patterns influencing vehicle preferences.



Observations and Insights:

- SUV buyers generally have the highest total salaries and tend to spend more on cars.
- Sedan buyers cluster in the middle salary and price range, suggesting a balance of affordability and value.
- Hatchback buyers typically have lower total salaries and purchase the least expensive cars.
- There's a positive correlation between total salary and car price across all car types, suggests the more salary the more expensive car will be purchased by customers.

Correlation between all numerical variables

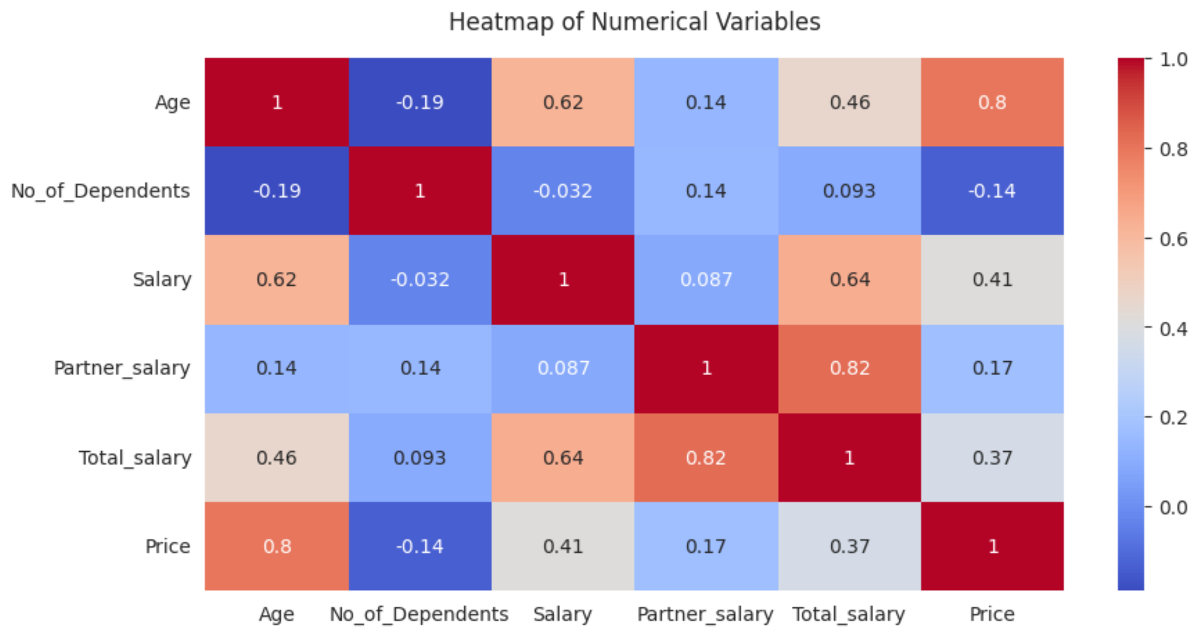
Correlation tells us whether and how strongly two things are connected.

For example, if people who earn more also tend to buy more expensive cars, then **salary and car price have a strong positive correlation**.

- A **positive correlation** means: when one goes up, the other also goes up.
- A **negative correlation** means: when one goes up, the other goes down.
- If there's **no clear pattern**, we say there's **no correlation**.

Using the heatmap, we can **quickly spot** which numbers in our dataset are related.

- Dark colors (closer to 1) mean strong positive correlation – when one value goes up, the other usually goes up too.
- Light or neutral colors (around 0) mean no clear connection – changes in one value don't affect the other much.
- Negative values (closer to -1) mean strong negative correlation – when one value goes up, the other tends to go down.



Strongest Correlations:

- Age and Price (0.80): Older buyers tend to purchase higher-priced vehicles.
- Partner_salary and Total_salary (0.82): Partner income heavily contributes to household income.
- Salary and Total Salary (0.64) when the salary increases the Total salary also increases.
- Age and Salary (0.62): Age likely reflects experience and higher pay.

Moderate Correlations:

- Salary and Price (0.41): Indicates income affects vehicle price decisions.
- Total_salary and Price (0.37): Reinforces the impact of total income on vehicle cost.

Weak or No Correlation:

- No_of_Dependants with most variables (e.g., with Price -0.14, Salary -0.032): Minimal impact on spending or earnings.
- Partner_salary and Salary (0.071): Suggests income independence between partners.

Relationship between categorical vs numerical variables

Explored the relationship between categorical and numerical variables to identify how customer segments influence key quantitative metrics.

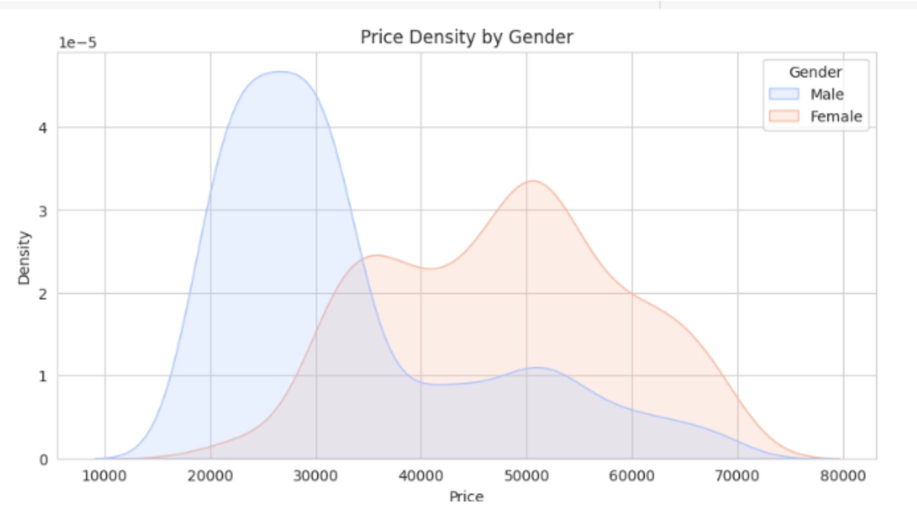
1. Understanding of spending by Gender

The purpose of this section is to explore how car purchase prices vary between male and female customers.

Firstly, I calculated the **average car price** and **price variation** (standard deviation) for male and female customers separately.

Gender	mean	std
Female	47705.167173	11244.836378
Male	32817.347790	12299.239195

Secondly, I created a **density plot** to show how car prices are distributed for each gender, helping us visually compare purchasing trends.



By analysing this relation and the plot we got some findings,

Observations and Insights:

- Female customers, on average, spend approximately (47000 – 32000) = $14,000 more on cars than male customers.
- Male buyers tend to prefer cars priced in the $20,000 to $32,000 range.

- Female buyers are more likely to purchase cars in the $32,000 to $55,000 range and highest buyers are on near $35k and $50k
- The price range among female buyers is broader, indicating more variety in their spending compared to male buyers.

2. Car Type Distribution by Gender

Analyzed Car Type Preferences by Gender:

- Calculated the **percentage distribution** of car types (Hatchback, SUV, Sedan) for each gender. (including each row)

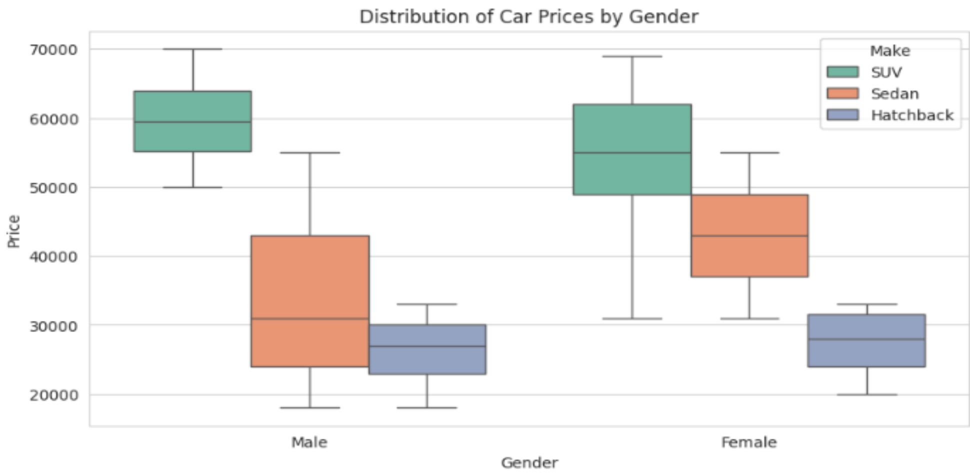
Make	Hatchback	SUV	Sedan
Gender			
Female	4.559271	52.583587	42.857143
Male	47.122602	9.841535	43.035863

- I also performed a **grouping operation** to get the count of males and females buying each car type (Hatchback, SUV, Sedan).

Make	Gender	count
Hatchback	**Male**	565
	Female	15
SUV	**Female**	173
	Male	118
Sedan	**Male**	516
	Female	141

dtype: int64

- I used a boxplot to visualize the distribution of car prices for both males and females, separated by car type (Hatchback, SUV, Sedan).



Observations and Insights:

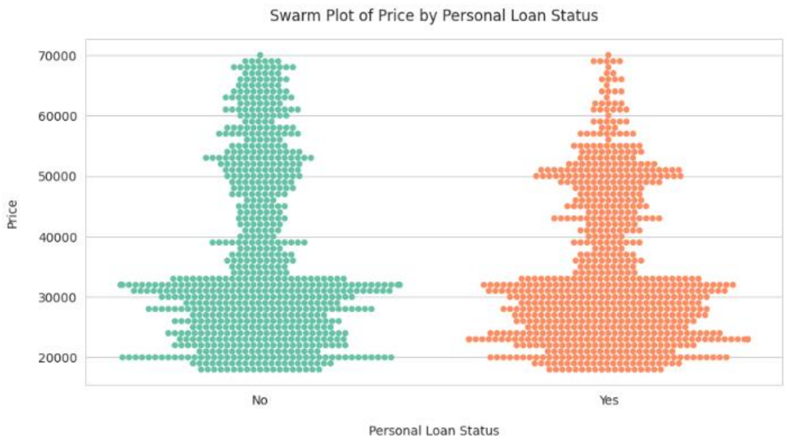
- **SUV Preferences**:
    - **Females** have a strong preference for **SUVs** (53%) and are willing to pay higher prices for them.
    - **Females** also show **greater price variability** for SUVs, with a **wider range** of prices and a **higher average price** compared to males.
- **Hatchback Preferences**:
    - **Males** prefer **Hatchbacks** (47%) and generally spend **less** on them.
    - The price distribution for Hatchbacks is more **consistent for males**, with a **lower average price** and a **narrower range**.
- **Sedan Preferences**:
    - **Both males and females** have a similar preference for **Sedans** (around 43% each).
    - However, **females tend to spend more** on **Sedans** than males, showing a slight difference in spending behaviour.
- **Price Insights for SUVs**:
    - **SUVs** are the **most expensive cars** for both genders.
    - **Females** show a **wider range** of prices, meaning they are more likely to buy both cheaper and more expensive SUVs, while **males** generally spend in a more **focused price range**.

3. Relationship between Personal Loans and Price

To understand how car prices vary based on personal loan status, you first calculated **summary statistics** for car prices, separated by whether customers had a personal loan (Yes or No), we computed Mean and Standard Deviation.

Personal_loan	mean	std
No	36742.712294	14534.344526
Yes	34457.070707	12578.780338

To visualize the distribution of car prices based on personal loan status, I created a **swarm plot**. It allows us to observe how prices are distributed.



Observations and Insights:

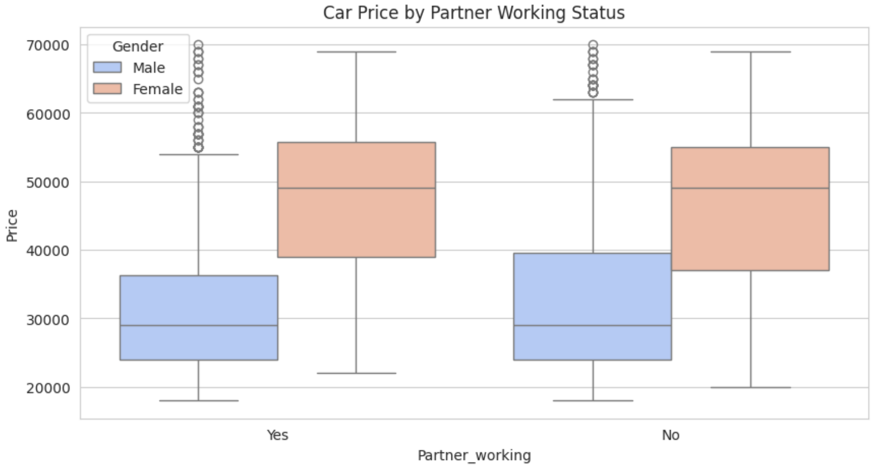
- Individuals without a personal loan tend to purchase higher-priced cars on average (36k vs. 34k)
- Both plots reveal a significant overlap in price distributions, but non-loan buyers show a slight shift toward higher car prices.

4. Car Prices by Partner Working Status

At first, calculated Summary Statistics to explore how car prices differ based on whether a customer's partner is working or not.

Partner_working	mean	std
No	36000.000000	13817.734086
Yes	35267.281106	13479.532555

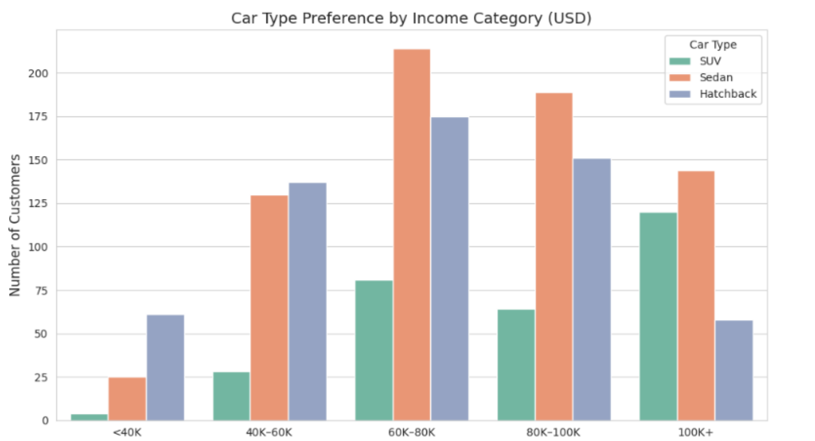
Generated the boxplot compares the price range, median, and outliers of car prices between customers with a working partner (Yes) and without a working partner (No), with additional distinction by gender.



Observations and Insights:

- Customers (Male) without a working partner spend slightly more on cars.
- The spending difference between partner status groups is small and likely not significant.
- Females consistently spend more on cars than males across both partner status groups.
- For males, spending range is slightly higher when their partner is not working.
- Male buyers generally avoid higher-priced cars, but showing fewer much expensive purchases.

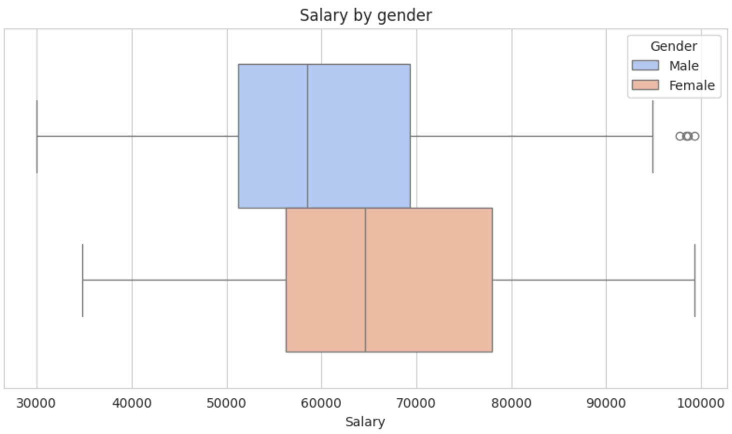
5. Income Vs. Car type distribution



Observations:

- Sedans are the most preferred car type overall, especially among customers earning between \$60K–\$100K.
- Hatchbacks dominate the <\$60K segments, indicating strong preference among lower-income customers.
- SUV popularity surges in the \$100K+ income group, nearly matching sedan sales.
- Car preference shifts with income, from hatchbacks (low) → sedans (middle) → SUVs (high).

6. Salary distribution by Gender



I have found some statistical information of Salary filtered by Gender.

Gender	max	min	mean	median
Female	99300	34800	66194.528875	64600.0
Male	99300	30000	58878.065054	58500.0

Observations and Insights:

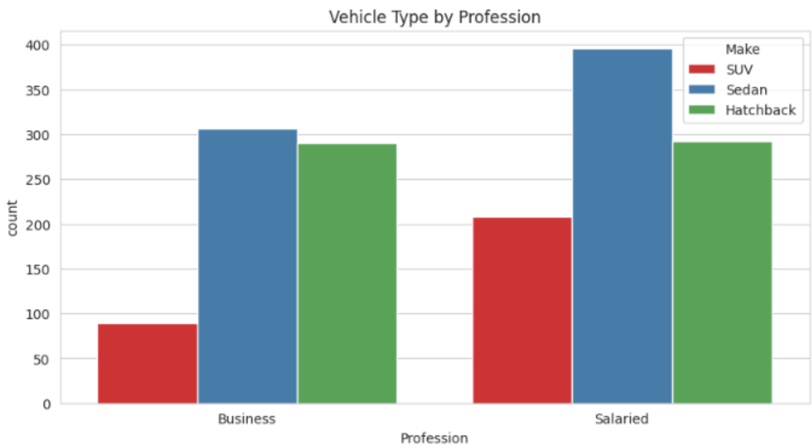
- Females have a higher median and mean salary than males.
- Males and females share the same maximum salary (99,300).
- Males have a lower minimum salary than females.
- Males show high-end salary outliers in the boxplot.

7. Car Type by Profession

I have calculated the proportion of different type of car choice by salaried people and businessmen.

Make	Hatchback	SUV	Sedan
Profession			
Business	42.335766	12.992701	44.671533
Salaried	32.589286	23.214286	44.196429

Also, visualized the count of various car types among different profession.

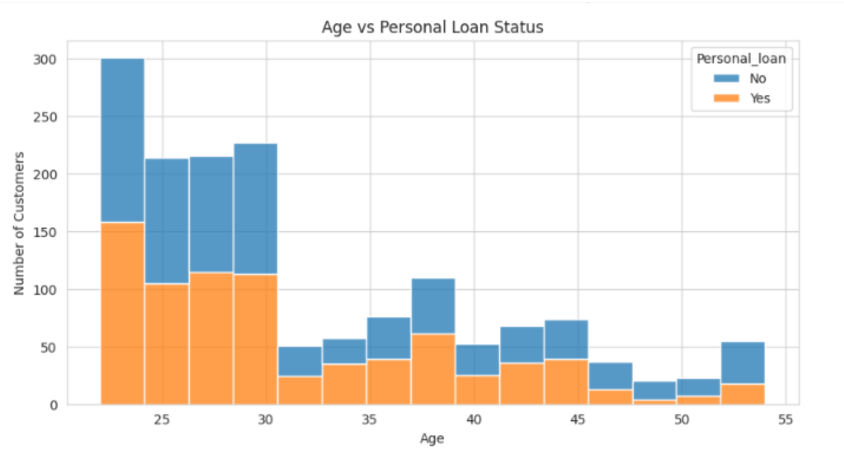


Observations & Insights:

- Sedans are the top choice overall 44%.
- Hatchbacks are a close second, also around 43%.
- SUVs are least preferred, with only 13% share.
- Salaried individuals prefer Sedans the most, while business user considers between Sedans and Hatchbacks.

8. Age Distribution Vs Personal Loan status

Based on the stacked bar chart we can get the distribution of ages taking loans.



Observations:

- Younger customers (22–30) are more likely to take personal loans, making up the bulk of loan applicants.
- Loan uptake drops significantly after age 35, indicating reduced loan dependency with age.
- A larger portion of customers above 40 years avoid personal loans altogether.
- Customers aged 25–30 show a near-equal split between loan and no-loan, indicating mixed financial behaviour in early careers.

KEY QUESTIONS & ANSWERS

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

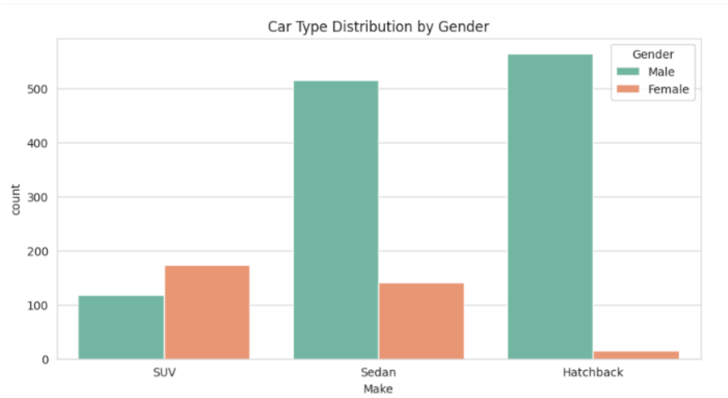
Ans:

We have seen how gender varies with different car types from the "Car Type Distribution by Gender" relationship previously.

Make	Hatchback	SUV	Sedan
Gender			
Female	4.559271	52.583587	42.857143
Male	47.122602	9.841535	43.035863

Number of vehicles purchased by men and women in the SUV, sedan, and hatchback segments:

Gender	Make	count
Female	SUV	173
	Sedan	141
	Hatchback	15
Male	Hatchback	565
	Sedan	516
	SUV	118



The computation above shows that

a) SUVs are preferred by 9.8% of men and 52.6% of women.

b) If we count the number of SUV sales, women purchase 173 of them, compared to 118 for men.

c) We can also infer the same thing from the Count plot.

Therefore, it can be said, women tend to prefer SUVs instead of men.

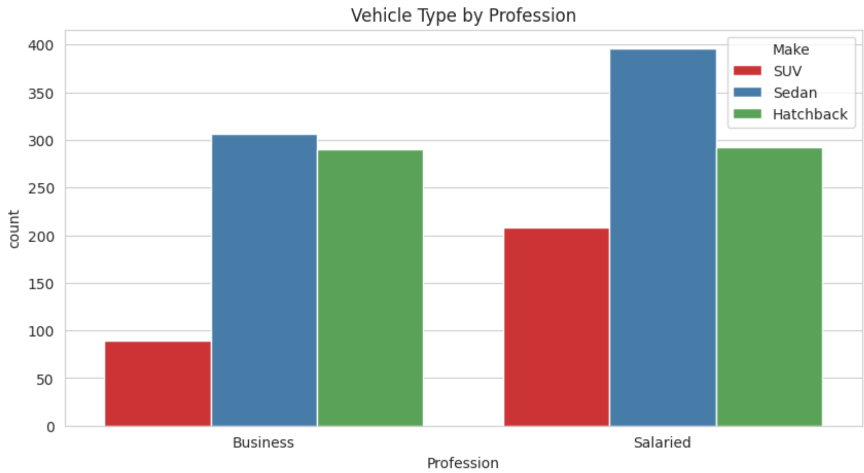
2. What is the likelihood of a salaried person buying a Sedan?

Ans:

The likelihood of purchasing various cars was previously obtained in the "Car Type by Profession" section of the bivariate analysis:

Make	Hatchback	SUV	Sedan
Profession			
Business	42.335766	12.992701	44.671533
Salaried	32.589286	23.214286	44.196429

A visual representation of the distribution of various car types by occupation:



Therefore, we can conclude that 44.19% of salaried men prefer to purchase sedans, which is roughly the same as 44.67% of businessmen. However, if we look at the numbers, we can see that salaried men have purchased more sedans than businesspeople.

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?

Ans:

I divided the data into various conditions and compared them in order to answer the question:

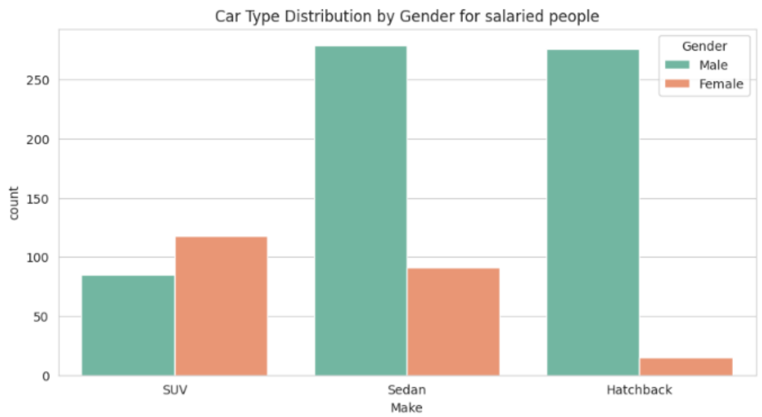
a) Calculated the percentage of distinct car preferences after filtering the data for salaried males and females.

Make	proportion
Sedan	43.59375
Hatchback	43.12500
SUV	13.28125

Male Salaried

Make	proportion
SUV	52.678571
Sedan	40.625000
Hatchback	6.696429

Female Salaried



Observations:

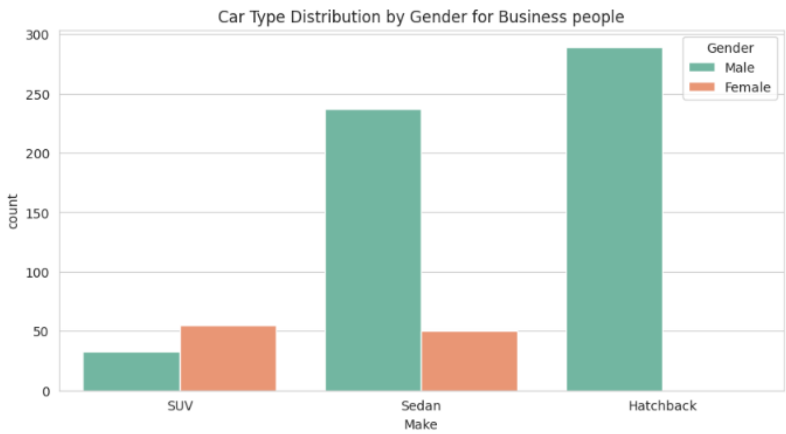
- The percentage of men (in percentage) who prefer sedans is 43.6%, which is the nearly same as those who prefer hatchbacks (43.1%).
- ·The SUV segment has a much lower percentage (13.3%).
- Women prefer SUVs 52.6% that is more than men do.
- It's about the same for women in the Sedan category. - 40.6%

b) Calculated the percentage of various car preferences after filtering the data for businesspeople (male and female).

Make	proportion
Hatchback	51.699463
Sedan	42.397138
SUV	5.903399

Make	proportion
Hatchback	51.699463
Sedan	42.397138
SUV	5.903399

Businessman **Businesswoman**



Observations:

- Of the businessmen, 51.7% prefer to purchase hatchbacks, with 42.3% preferring sedans.
- For businesswomen, 52% prefer SUVs over sedans, with 47.6% preferring them and nearly none preferring hatchbacks.

When comparing the two scenarios (businessmen and salaried individuals), we can conclude that, following the segmentation of car preferences for men and women, **women are the biggest SUV buyers, while men prefer sedans and hatchbacks**. Hence, Sheldon Cooper's claim that, "A salaried male is an easier target for a SUV sale over a Sedan sale" fails for this dataset and proven wrong.

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

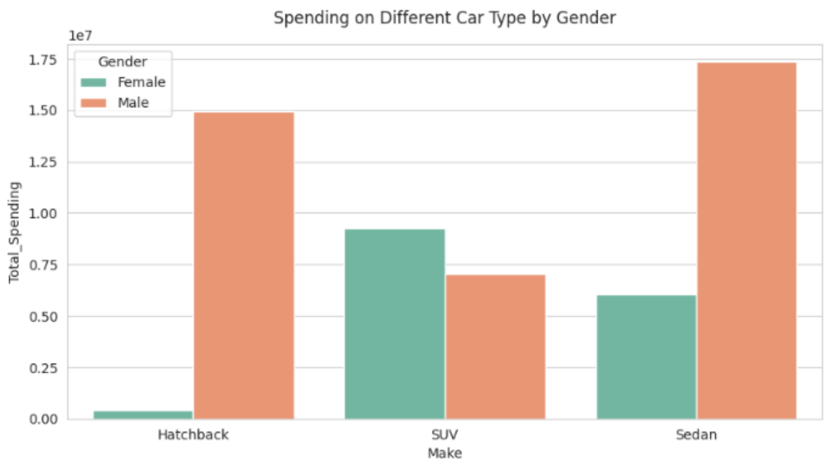
Ans:

Total amount spent by men and women on cars:

Gender	Total_Spending
Female	15695000
Male	39348000

Statistical details of spendings, filtered by gender and car type:

	Gender	Make	Total_Spending	Average_Price	Price_Std_Dev
4	Male	SUV	7031000	59584.745763	5673.941441
1	Female	SUV	9252000	53479.768786	10378.079311
2	Female	Sedan	6031000	42773.049645	7248.023173
5	Male	Sedan	17358000	33639.534884	10752.371921
0	Female	Hatchback	412000	27466.666667	4240.395310
3	Male	Hatchback	14959000	26476.106195	4273.646655



The table and bar graph provided us with some important insights:

a) Men spend the most on average in the SUV segment, while women spend the most in the sedan and hatchback segments.

b) Males spend the most on sedans and hatchbacks when it comes to total spending, while females are leading the SUV market.

c) When looking at all the data, men spend twice as much on cars ($39348,000) as women ($15695,000).

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

Ans:

Total spent by customers who took personal loans:

$27,290,000.00

The total amount spent by various professions is:

Profession	Total_Spending
Business	11711000
Salaried	15579000

Total amount spent on various car models:

Profession	Make	Total_Spending
Business	Hatchback	3952000
	SUV	1830000
	Sedan	5929000
Salaried	Hatchback	3691000
	SUV	4377000
	Sedan	7511000

Observations:

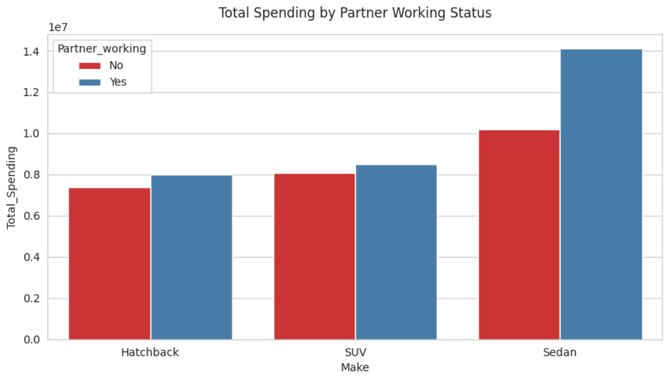
- Compared to business professionals, salaried people who took out personal loans spent more money overall.
- Among salaried individuals with personal loans, spending was highest on Sedans, followed by SUVs.
- Business professionals with personal loans predominantly purchased Sedans, followed by Hatchbacks.

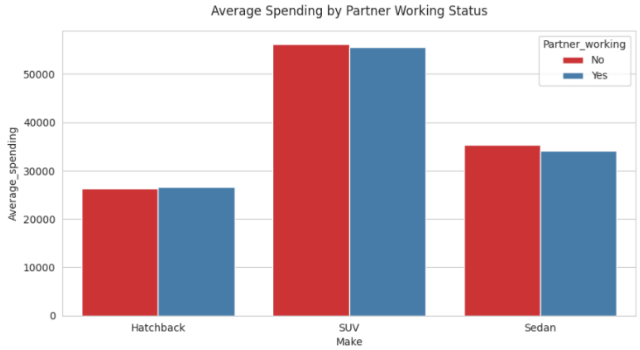
6. How does having a working partner influence the purchase of higher-priced cars?

Ans:

Average and total car expenses by partner's employment status:

Partner_working	Make	Total_Spending	Average_spending
No	Hatchback	7397000	26323.843416
	SUV	8089000	56173.611111
	Sedan	10182000	35354.166667
Yes	Hatchback	8011000	26614.617940
	SUV	8491000	55496.732026
	Sedan	14110000	34082.125604





Individuals with working partners **spent slightly more** on Sedans in terms of **total spendings**.

Individuals with working partners **spent slightly more in total across all car types, though average spending** per car was similar or even lower in some cases—**indicating a modest but not significant tendency toward higher-priced car choices in dual-income households.**

ACTIONABLE INSIGHTS & RECOMMENDATIONS

Insight 1: High Income Drives SUV Purchases

Observation: Customers with annual salaries above $60,000 and working partners are highly likely to purchase Sedans followed by SUVs.

Target Segment: Dual-income families with combined income >$100,000.

Recommendation: Launch premium SUV marketing campaigns focused on family safety, comfort, and status appeal. Use targeted online ads, especially on LinkedIn and YouTube.

Insight 2: Female Customers Show Preference for SUVs

Observation: A significant proportion of female customers prefer SUVs over other car types. Recommendation:

Target Segment: Professional women aged 30–50.

Recommendation: Design women-focused promotions highlighting SUV features like safety, advanced tech, and ease of driving. Partner with female influencers for campaigns.

Insight 3: Loan Eligibility Influences Purchase Behaviour

Observation: Customers eligible for loans are more inclined to purchase medium-priced cars like Sedans.

Target Segment: Middle-income earners not eligible for loans.

Recommendation: Offer easy financing options, pre-approved loan schemes, or zero-down payment plans to convert hesitant buyers which can expand sales of premium models like SUVs.

Insight 4: Car Type Varies by Income Level

Observation:

Low-income customers ($30–60k): Prefer hatchbacks.

Middle-income customers ($60–100k): Prefer sedans.

High-income customers ($100k+): Prefer SUVs.

Recommendation:

Develop tiered marketing strategies:

Promote hatchbacks via EMI plans to recent graduates or entry-level workers.

Advertise sedans as practical upgrades for growing families.

Market SUVs as aspirational, lifestyle-oriented vehicles for affluent buyers.

Insight 5: Partner's Employment Status Affects Spending Power

Observation: Customers with working partners are more likely to buy expensive car models. Recommendation:

Target Segment: Households with dual earners.

Recommendation: Cross-sell premium services like extended warranties and premium insurance bundles. Emphasize lifestyle branding in your campaigns.

Insight 6: Young Customers & Budget Sensitivity

Observation: Younger customers (age 25–35) tend to prefer lower-priced models and are often not loan-eligible.

Target Segment: First-time car buyers and young professionals.

Recommendation: Launch a “Young Driver Program” with affordable instalment plans, trade-in offers, and co-buyer (parent) finance options.

(**Note:** References for the insights shared in this report can be found in the "Actionable Insights & Recommendations" section of **‘.ipynb’** file)

CONCLUSION

Austo Motor Company's exploratory data analysis provides insightful information about the financial characteristics, vehicle preferences, and customer demographics. A better grasp of the company's present clientele and their purchasing habits is offered by this study, which looks at trends across a number of characteristics, including age, gender, marital status, income, and car ownership.

The majority of customers are middle-aged, married, frequently male, and have moderate to high household incomes, according to key findings. The most popular car type was found to be sedans, whereas SUVs and hatchbacks displayed clear trends associated with gender and income levels. Vehicle choices and spending capacity were also found to be influenced by factors like partner income, employment status, and the presence of children.

The analysis also pointed out areas that needed work, like typos and missing values in the data, which, when fixed, helped produce more accurate interpretations. The amount spent on vehicles is significantly influenced by gender, household income, and car type, according to multivariate relationships.

Overall, Austo Motor Company can benefit from the insights gained from this analysis by refining its product offerings, identifying prospective customer segments for expansion, and optimizing its marketing strategies. The business can more effectively match its outreach and sales initiatives to the changing tastes of its wide range of clients by using data-driven decision-making.