

Project 1

- Austo Automobile Analysis -

NO	Name of Figure	Page no
1	Age of Customers	9
2	NO of Dependants of customers	10
3	Salary of Customers	11
4	Salary of customers partner	12
5	Total salary of Customers and their partners	13
6	Price of the Product	14
7	Customers Gender	15
8	Customers Profession	16
9	Customers Marital status	17
10	Customers education qualification	18
11	Customers with personal loan	19
12	Customers with house loan	20
13	Customers with Partner working	21
14	Type of automobile	22
15	Relationship between all numeric variables	23
16	Correlation of all numerical variables	25
17	Relationship between make and customers age	26
18	Relationship between make and customers No of Dependents	27
19	Relationship between make and Price of the product or service	28
20	Relationship between Salary and Education of the Customer	29
21	Relationship between House loan and Price of the Customers	30
22	Relationship between Make and Total Salary of the Customer	31
23	Relationship between make and Customers Gender	32
24	Relationship between make and Customers Profession	33
25	Relationship between the Customers Profession, Gender and Make	34
26	Amount spent on purchasing automobile according to Gender	35
27	Money spent on purchasing automobile by customers with personal loan	36
28	working partner and price of their Automobile	37

S.no	Topics	Page no
1	Problem- Austo Automobile Analysis	4
1.1	Problem Definition	4
1.2	Data Overview	6
1.3	Univariate Analysis	9
1.4	Bivariate Analysis	23
1.5	Answer Key Questions	32
1.6	Conclusion and Recommendations	38

Problem- Austo Automobile Analysis

1.1.Problem Definition

Context

Introduction:

Cars are motor vehicles designed primarily for transporting people and goods on roads. Since their invention in the late 19th century, cars have revolutionized personal mobility and reshaped modern society. Today, they come in various forms, from fuel-efficient sedans to electric vehicles (EVs) and autonomous driving models.

Austo Motor Company is a leading manufacturer in the automotive industry, known for its SUV, Sedan, and Hatchback models. The company prides itself on delivering vehicles that align with customer preferences and market trends. However, during a recent board meeting, concerns were raised about the effectiveness of the current marketing campaign. The board expressed uncertainty about whether the campaign is resonating with customers and driving the desired sales outcomes.

Current Challenge:

The concerns highlight a critical need to reassess the company's marketing strategies. The primary issue is to improve the efficiency of the campaign in reaching potential customers and influencing their purchasing decisions. To address this challenge, the company has decided to engage a analytics professional to conduct a thorough analysis of the campaign's performance.

Objective of the Analysis:

The overarching goal is to better understand customer demand and preferences through data analysis. By gaining these insights, Austo Motor Company aims to enhance its marketing efforts and improve the overall customer experience. The Data Scientist at the company has to analyze the provided data to answer key questions that will help refine the marketing campaign and ultimately boost business performance.

Data Description:

austo_automobile.csv: The dataset contains buyer's data corresponding to different types of Car products.

Data Dictionary:

- **Age:** The age of the customer in years.
- **Gender:** The gender of the customer, categorized as male or female.
- **Profession:** The occupation or profession of the customer.
- **Marital_status:** The marital status of the customer.
- **Education:** The educational qualification of the customer, Graduate or Post Graduate.
- **No_of_Dependents:** The number of dependents (e.g., children, elderly parents) that the customer supports financially.
- **Personal_loan:** A binary variable indicating whether the customer has taken a personal loan or not.
- **House_loan:** A binary variable indicating whether the customer has taken a housing loan or not.
- **Partner_working:** A binary variable indicating whether the customer's partner is employed or not.
- **Salary:** The salary or income of the customer.
- **Partner_salary:** The salary or income of the customer's partner.
- **Total_salary:** The total combined salary of the customer and their partner.
- **Price:** The price of a product or service.
- **Make:** The type of automobile.

1.2.Data Overview:

Import all the required python libraries followed by loading the CSV files containing the customer information of the austro automobile.

The Dataset contains 1581 rows and 14 columns. We are taking the sample of first 5 rows, which helps us to gain better intel about the dataset.

The dataset contains 5 columns of integer (int64), 1 column of float (float64), 8 columns of object as datatype.

Treatment of missing values:

There are 2 columns in the dataset that contains missing values.

- The column “Gender” has 53 missing values which sums up to 3.35% of the column.
- The column “Partner_salary” has 106 missing values which sums up to 6.71% of the column.

Column “Gender” consists of categorical values (Object). In order, to fill the missing values of the categorical columns “Gender”, the mode of the column - “Male” has been used.

Column “Partner_salary” consists of Numerical Values (Float). In order, to fill the missing values of the Numerical columns “Partner_salary”, the mean of the column - “20226.0” has been used.

Statistical summary:

The statistical summary of the numeric columns have been checked by using describe (), function which helps us to obtain count, mean, std (Standard deviation), min (minimum value), 25% (25th percentile), 50% (50th percentile), 75% (75th percentile) and Max (maximum value) for each numeric columns.

Data irregularities:

Inconsistent Data:

The inconsistent data have been found in the Gender column where “Female” have been misspelled as “Femal” and “Femle”. The misspelled words have been replaced by the word (string) “Female”.

Outliers:

The outliers were found in 2 numeric columns, namely “No_of_Dependents” and “Total_salary”. The column “No_of_Dependents” consists of 20 outliers which sums up to 1.265% and The column “Total_salary” consists of 27 outliers which sums up to 1.708%.

The outliers in the column “No_of_Dependents” is of the value 0, which we know that this value is accurate, removing or Capping would compromise the integrity of the data.

The outliers in the column “Total_salary” are the values which lies above 149000 amount of salary. We may cap the outliers which lies above the 149000, where the outliers are restricted to not exceed “upper_whisker”. So the values of outliers are replaced with the values of “upper_whiskers”.

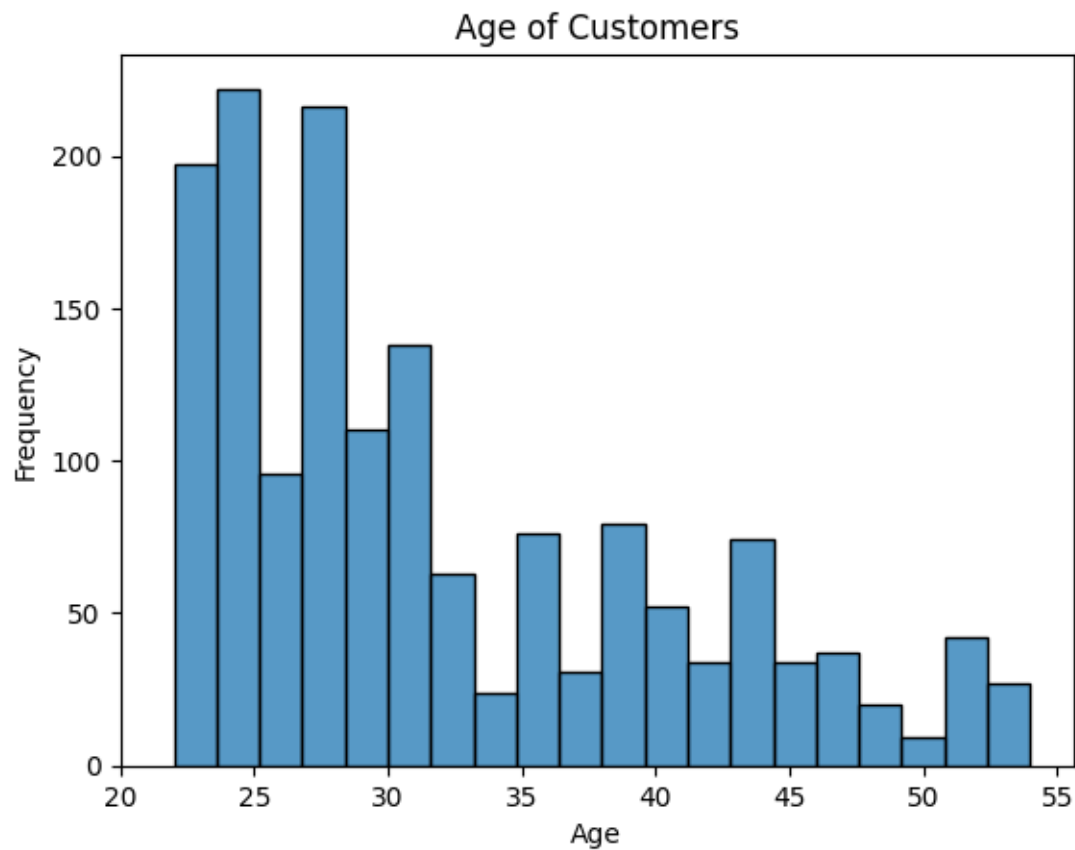
Observations and Insights:

- The dataset of austo_automobile consists of 1581 rows and 14 columns of which 5 columns of integer, 1 column of float, 8 columns of object.
- In the initial data analysis, we identified missing values in the “Gender” and “Partner_salary” columns, with “Gender” having 3.35% missing data and “Partner_salary” having 6.71%. To ensure the integrity of the analysis, we treated the missing Gender values by filling them with the mode gender and filled the missing Partner salary values with the mean Partner salary. This approach helped to retain the majority of the dataset without introducing significant bias.
- The average Total salary in the dataset is 79398.545, with a standard deviation of 24849.148. The most common product type is “Sedan”, which represents 44% of all purchases.

- Inconsistent data was identified in the Gender column, where "Female" was misspelled as "Femal" and "Femle". These misspellings were corrected by standardizing all instances to "Female", ensuring consistency and accuracy in the gender data.
- Outliers were identified in the No_of_Dependents and Total_salary columns. For the Total_salary column, outliers were treated by capping values at the lower and upper whiskers, which helps in minimizing their impact on the analysis and provides a more accurate representation of salary distributions. The outliers in the No_of_Dependents column were left untreated to retain the integrity of the data's diversity and reflect the true range of dependents. This balanced approach ensures robustness in statistical analysis while preserving the dataset's comprehensive view.

1.3.Univariate Analysis:

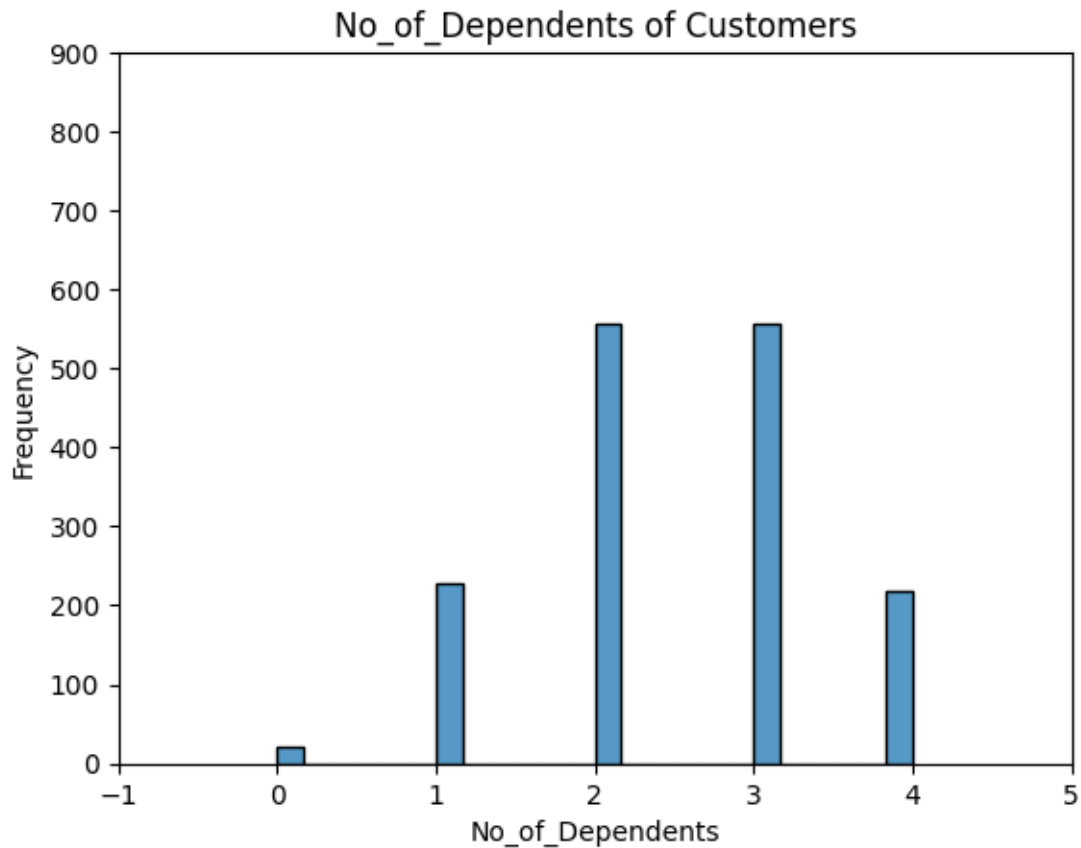
1. Age of the Customers



Interpretation:

The graph illustrates the demographic distribution of customers by age. The highest number of customers falls within the age range of 22 to 32, while the fewest customers are in the 45 to 55 age range. Additionally, the graph shows a decreasing trend: as age increases, the number of customers tends to decrease.

2. No of Dependents of Customers



Interpretation:

The graph titled illustrates the distribution of the number of dependents among customers. The majority of customers have between 2 to 3 dependents. This range represents the highest frequency of data points on the graph. A relatively small number of customers have no dependents at all, as indicated by the lower frequency of data points in this category.

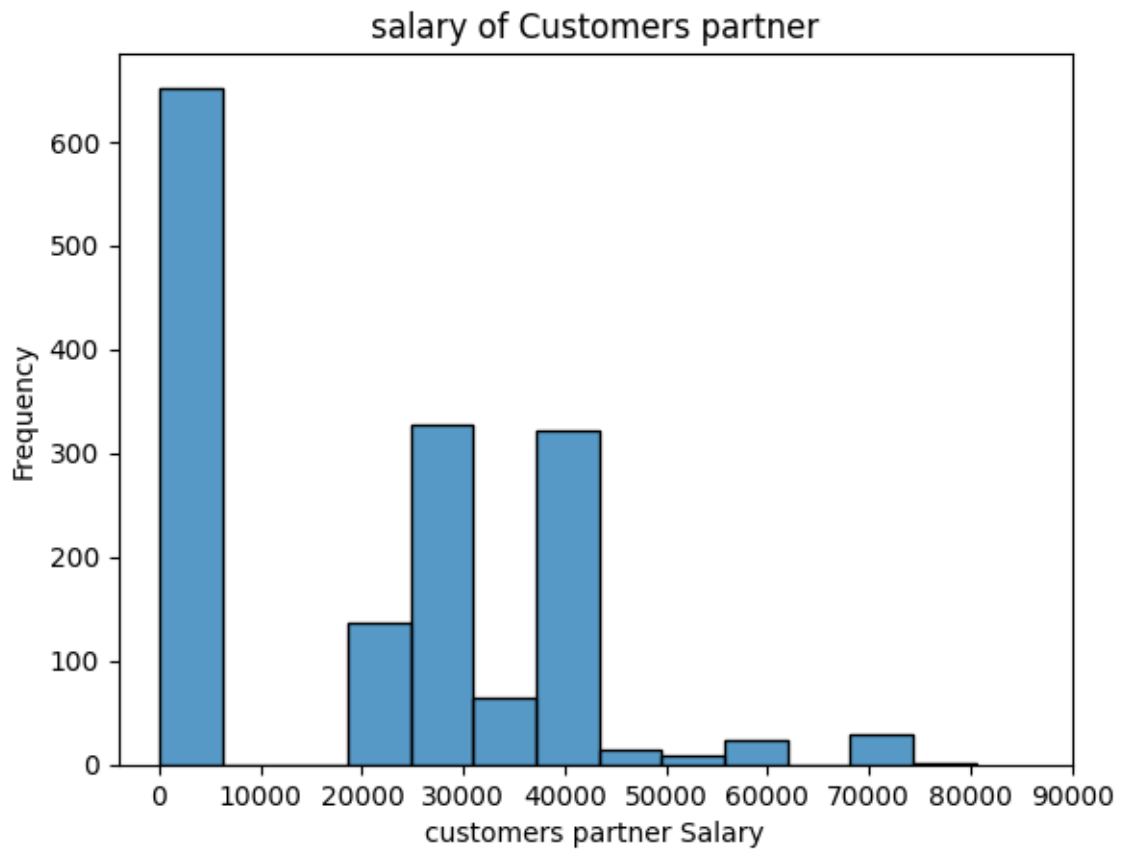
3. Salary of the Customers



Interpretation:

The graph shows that the majority of customers have salaries in the range of 50,000 to 70,000. A moderate number of customers earn between 70,000 and 90,000, while fewer customers fall within the 30,000 to 50,000 salary range.

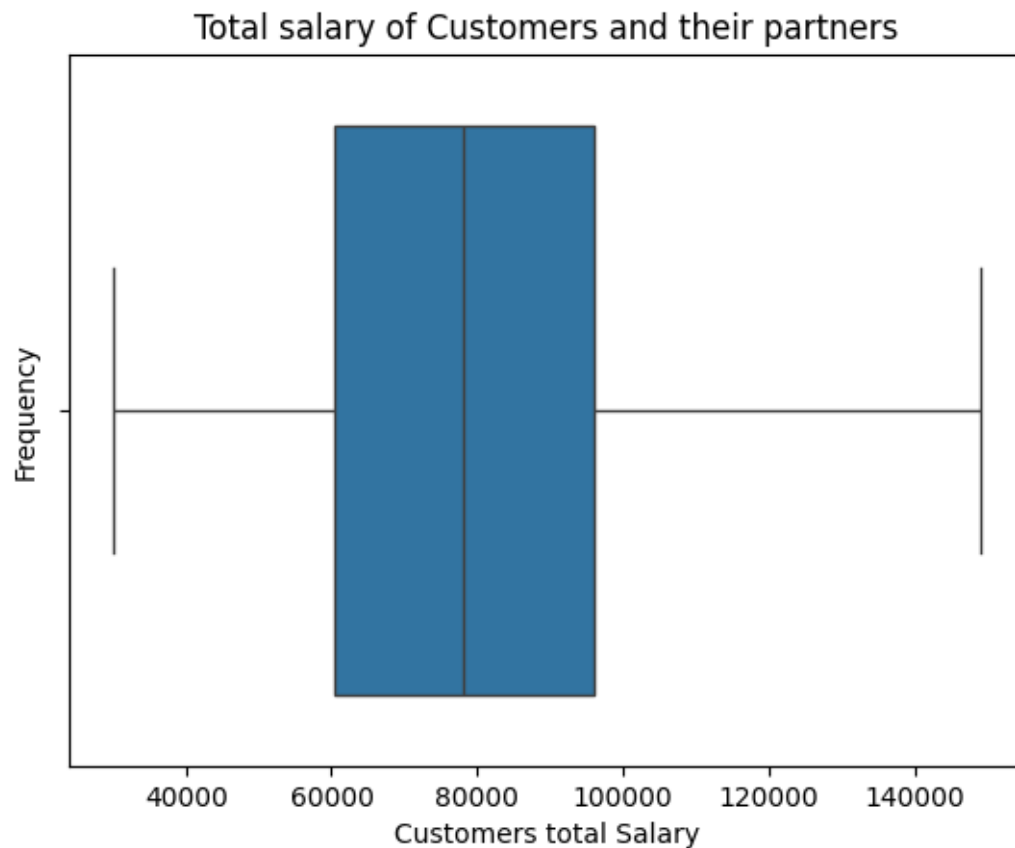
4. Salary of Customers partner



Interpretation:

The graph indicates that the majority of customers have a non-salaried partner. Among those with salaried partners, the highest number fall within the salary range of 18,000 to 44,000. Conversely, customers whose partners earn between 75,000 and 81,000 are the least common.

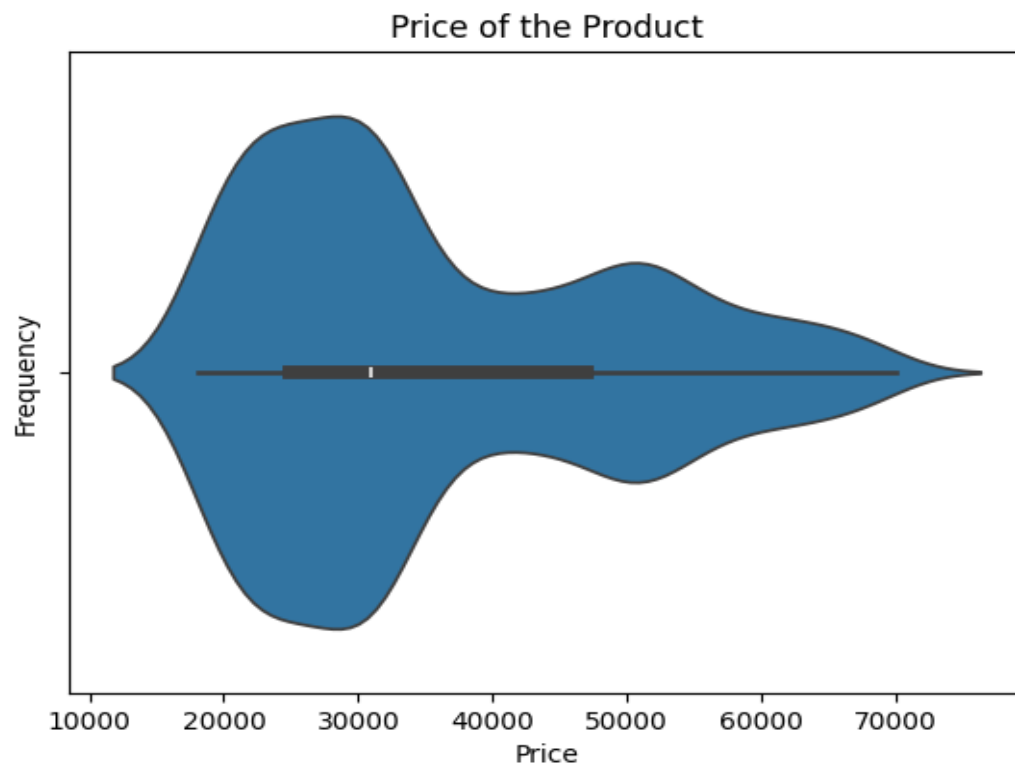
5. Total salary of Customers and their partners



Interpretation:

The boxplot shows the distribution of total salaries, including both the customer's and their partner's earnings. Most total salaries fall between 60,000 and 95,000, with a median of 78,000. The range extends from 30,000 to 1,50,000, indicating significant variability in combined earnings among customers and their partners.

6. Price of the Product

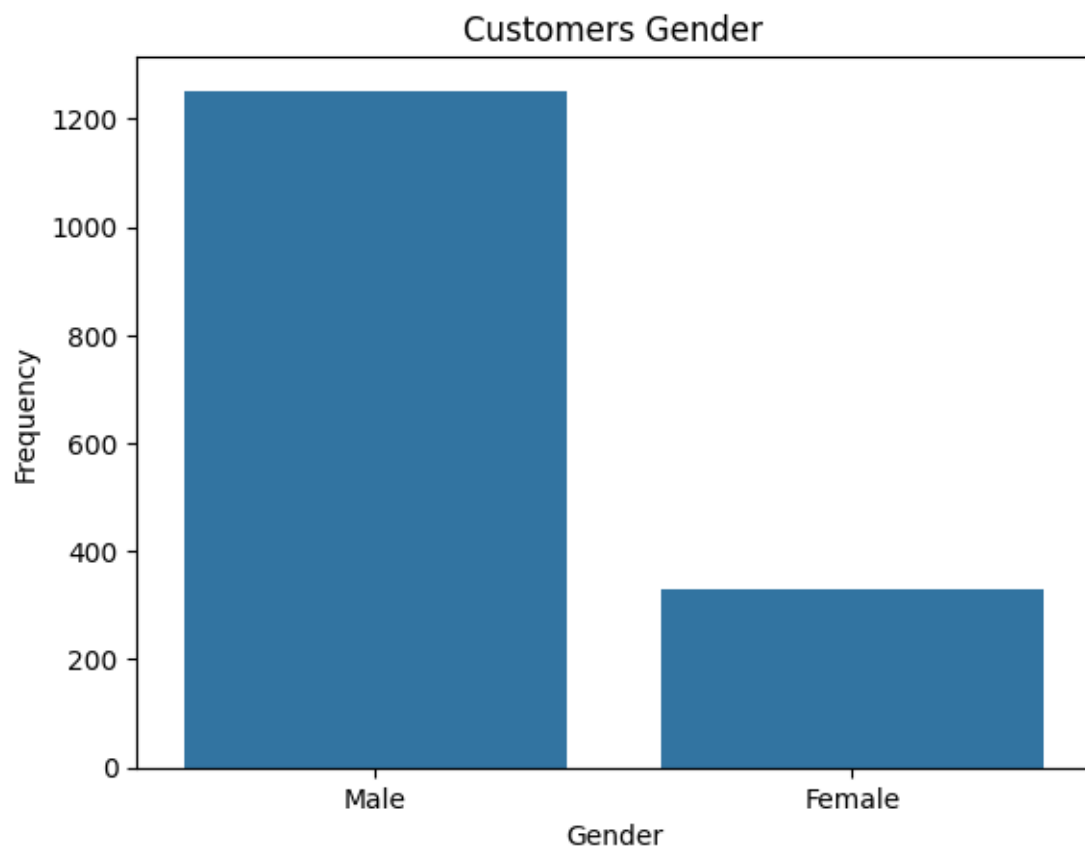


Interpretation:

The graph shows that the majority of products are priced between 20,000 and 30,000. The distribution of product prices is right-skewed, meaning that higher prices are less common. As the price exceeds 30,000, the number of products decreases significantly.

Categorical Variable:

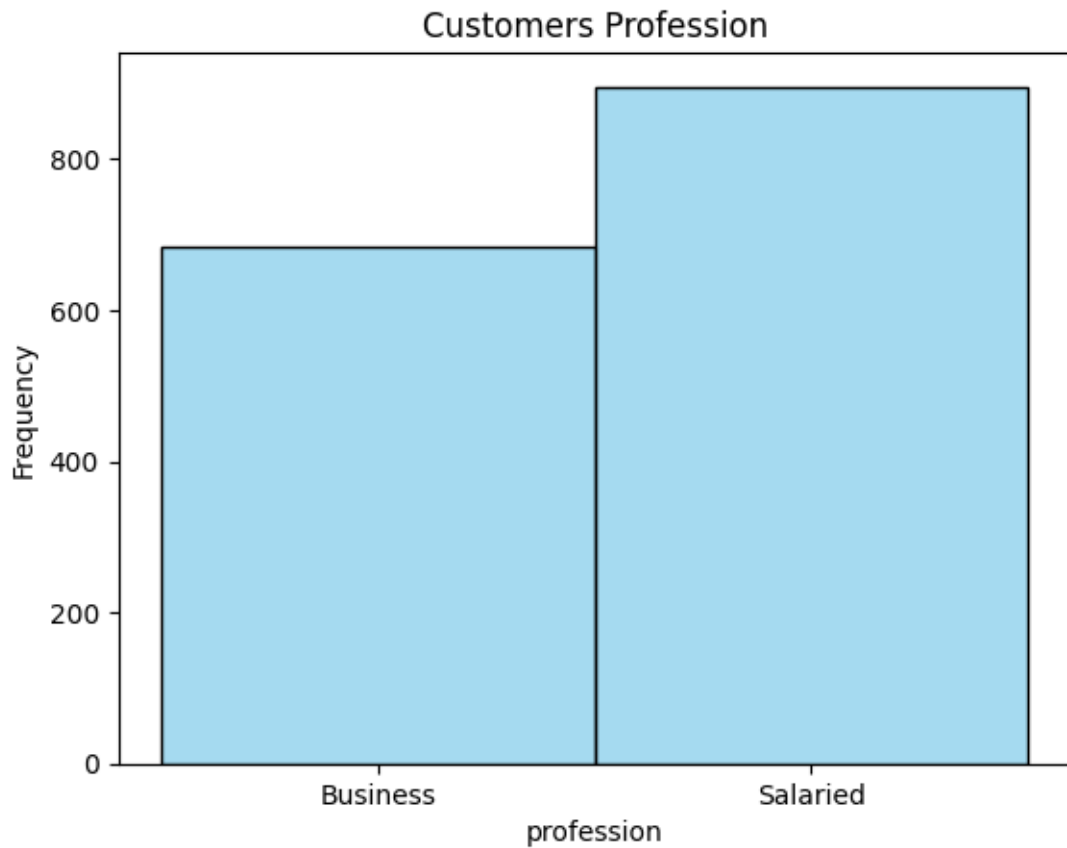
7. Customers Gender



Interpretation:

The graph shows the gender distribution of customers, a key demographic factor. It reveals that the majority of customers are male, with over 1,200 male customers represented. In contrast, the number of female customers is significantly lower, at fewer than 400. This indicates a pronounced gender disparity in the customer base.

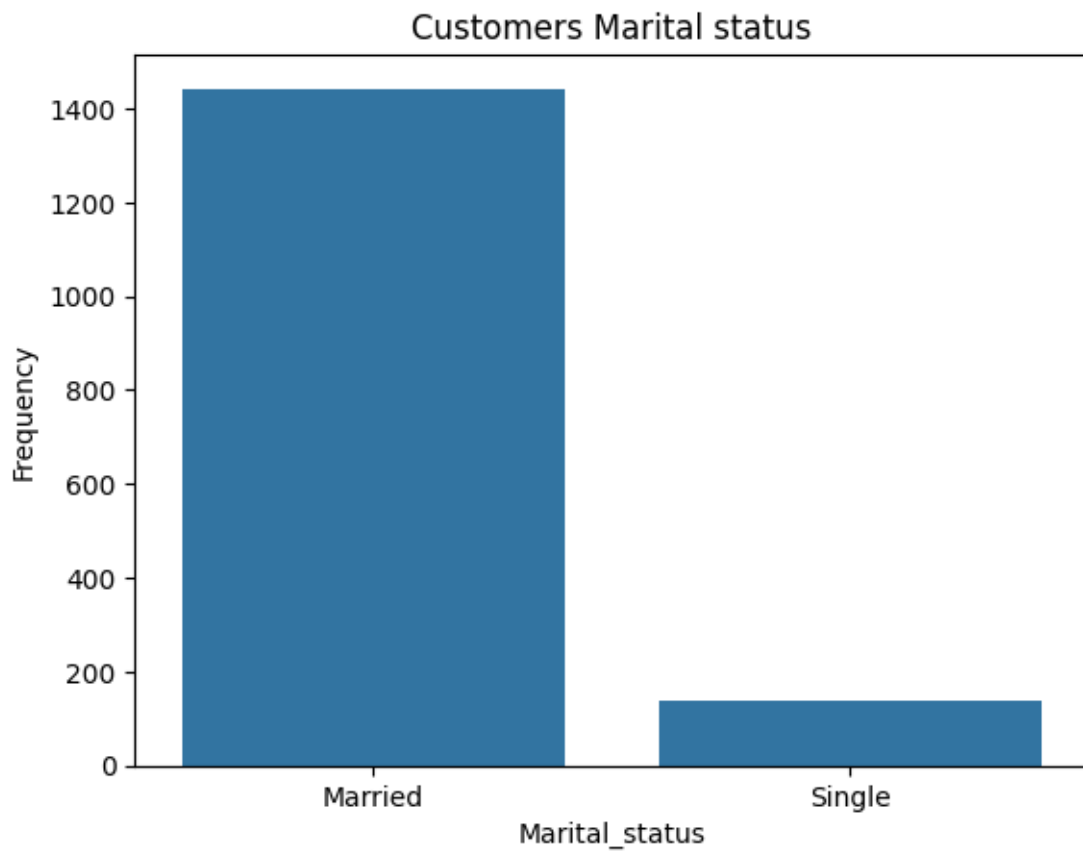
8. Customers Profession



Interpretation:

The graph shows the distribution of customers by profession. It indicates that the number of salaried customers is significantly higher compared to those who are in business. This suggests a slight predominance of salaried individuals in the customer base.

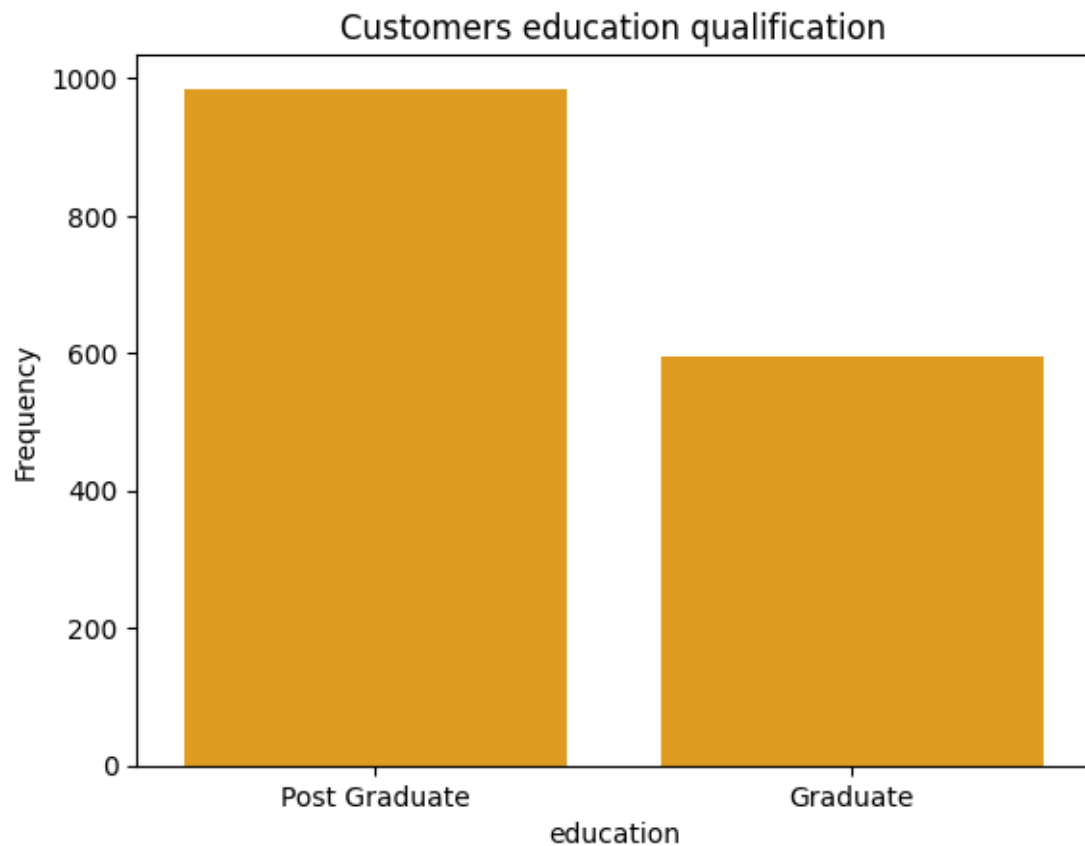
9. Customers Marital status



Interpretation:

The graph shows the marital status of customers. It reveals that there are more married customers compared to single customers. This indicates that married individuals form a larger portion of the customer base.

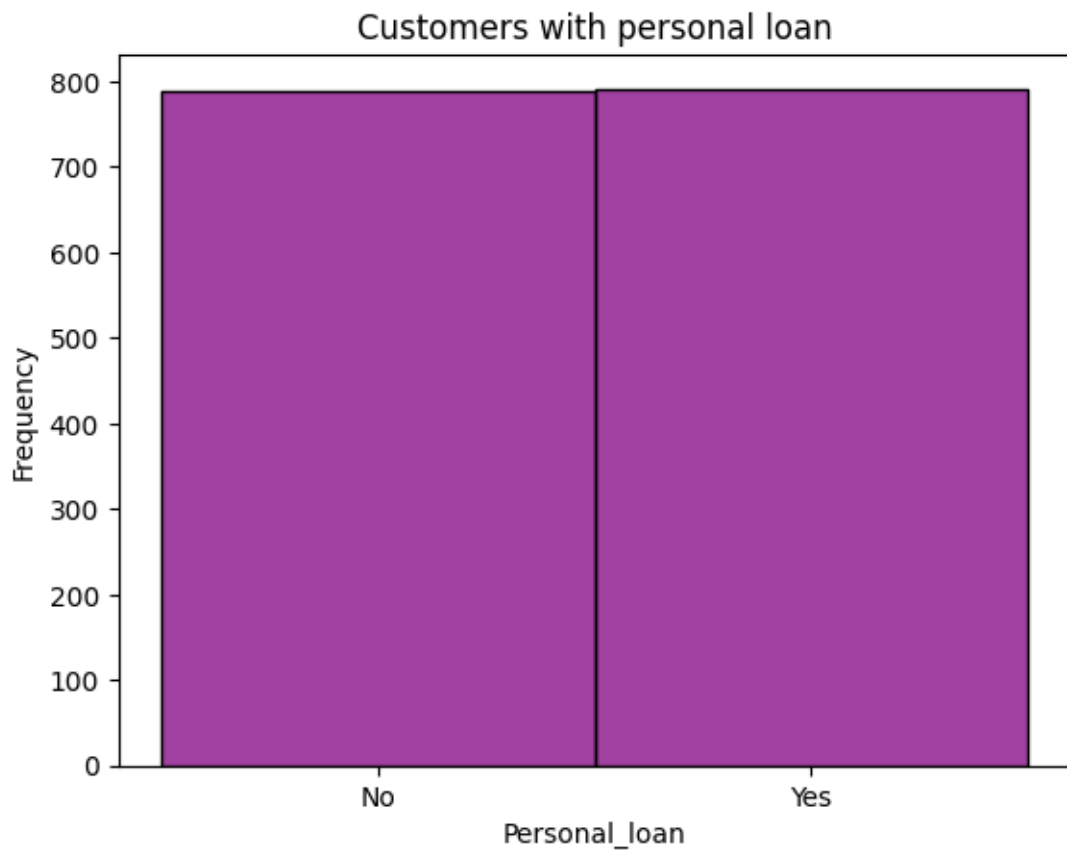
10. Customers Education qualification



Interpretation:

The graph shows the educational qualifications of customers. It indicates that the number of customers with postgraduate degrees is higher compared to those with only a graduate degree. This indicates that the customer base has a higher proportion of individuals with postgraduate qualifications.

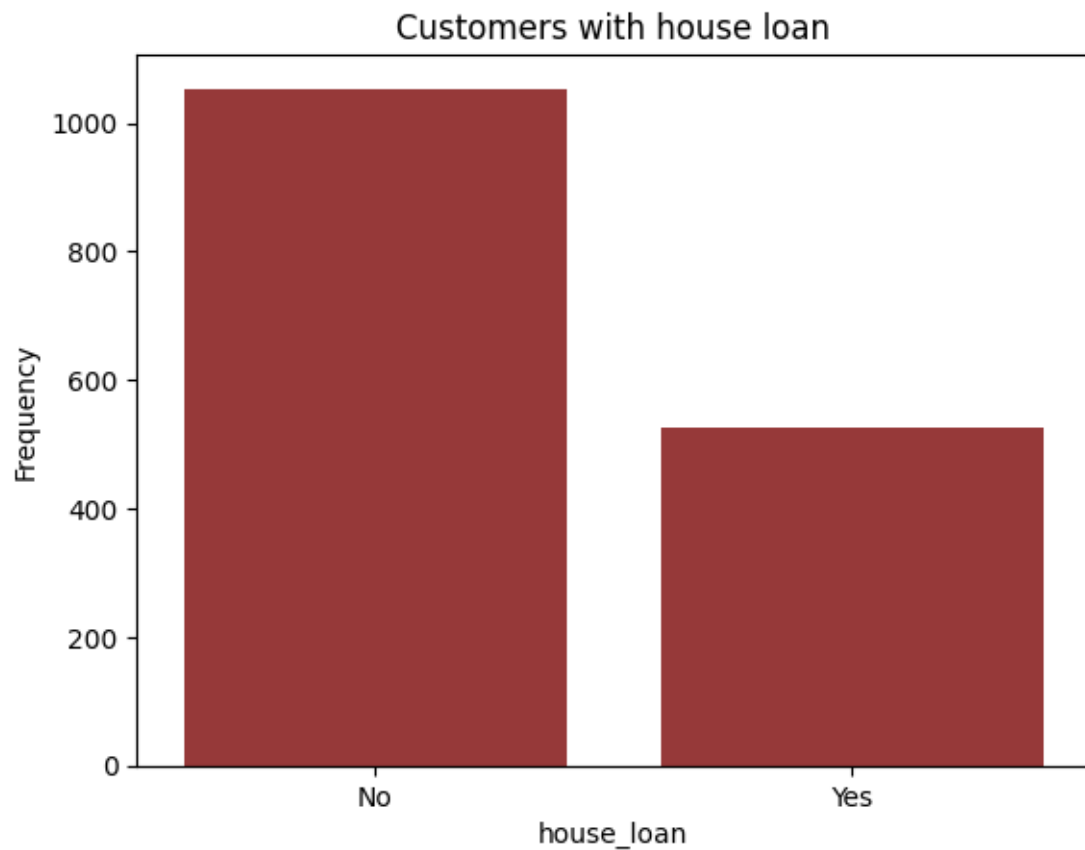
11. Customers with personal loan



Interpretation:

The graph shows the number of customers with personal loans. The number of customers with personal loans is roughly equal to the number of those without personal loans, indicating a balanced distribution between the two groups.

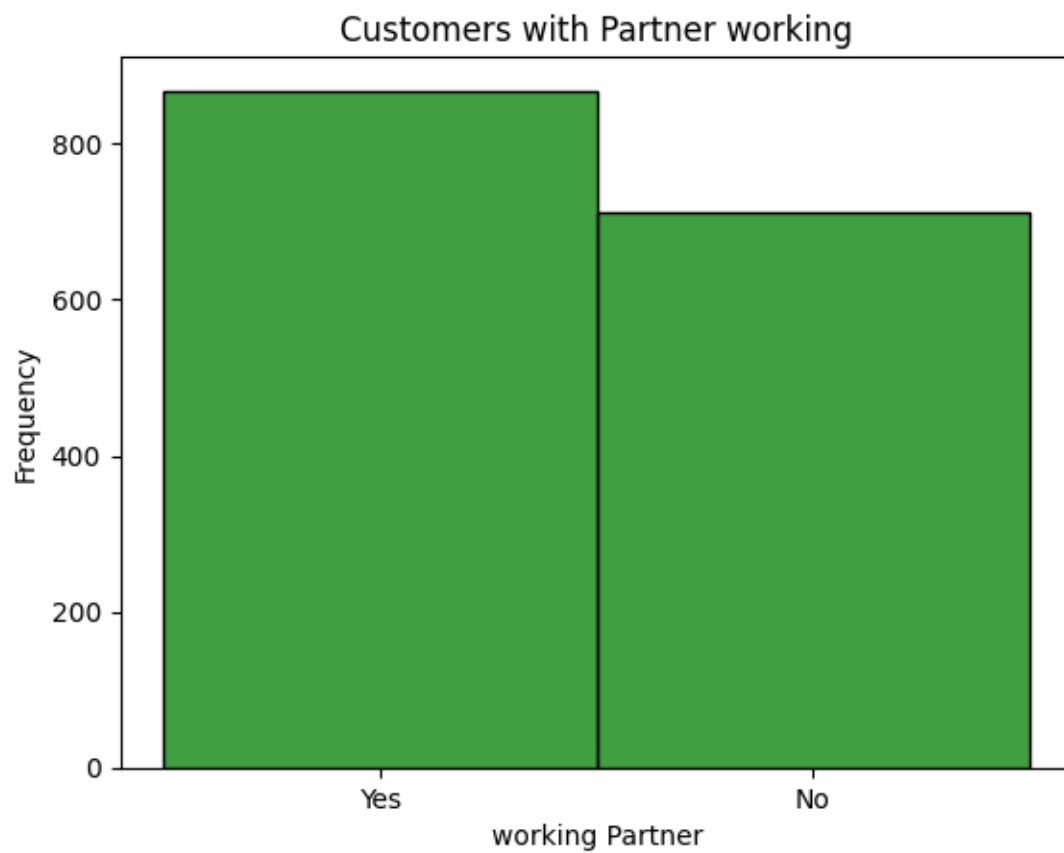
12. Customers with house loan



Interpretation:

The graph shows the house loan status of customers. It indicates that fewer customers have house loans compared to those who do not have house loans. This suggests that house loan holders are in the minority relative to non-house loan holders.

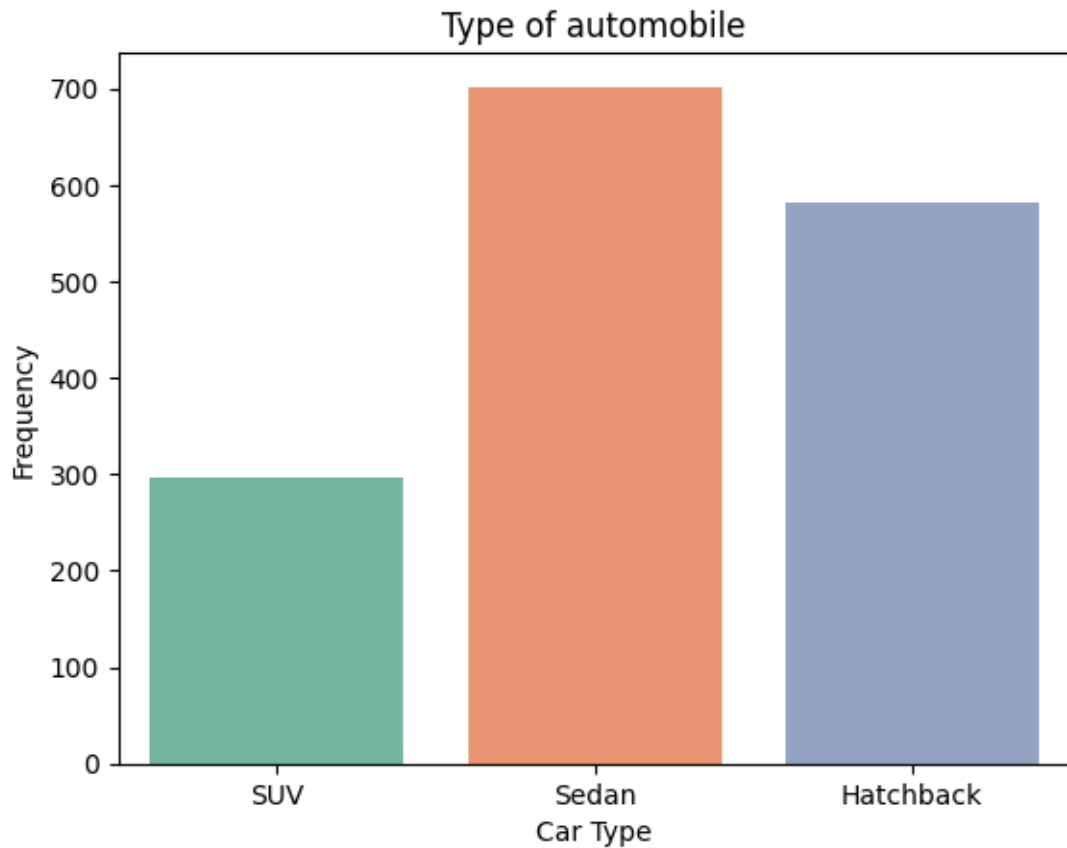
13. Customers with Partner working



Interpretation:

The graph shows whether customers' partners are working. It indicates that there are slightly more customers with working partners compared to those with non-working partners. This suggests a modest predominance of customers whose partners are Working.

14. Type of automobile



Interpretation:

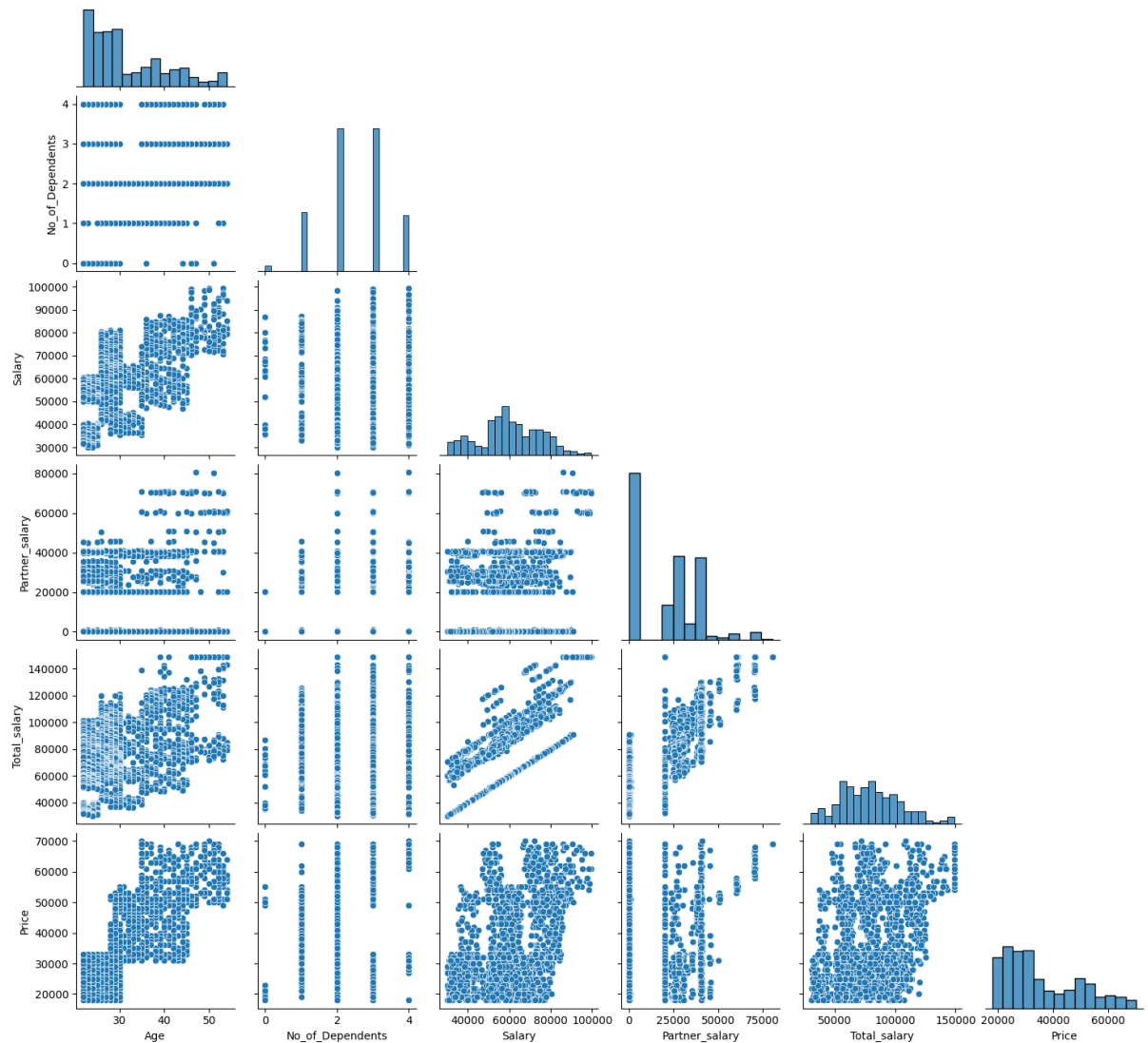
The graph displays the types of automobiles owned by customers. It shows that sedans are the most commonly owned vehicle type, with the highest number of customers having them. Hatchbacks are the next most popular, while SUVs are owned by a relatively small number of customers. This indicates a clear preference for sedans among the customer base, with SUVs being less favored.

1.4.Bivariate analysis:

Numerical variables

15. Relationship between numeric variables

Relationship between all numeric variables

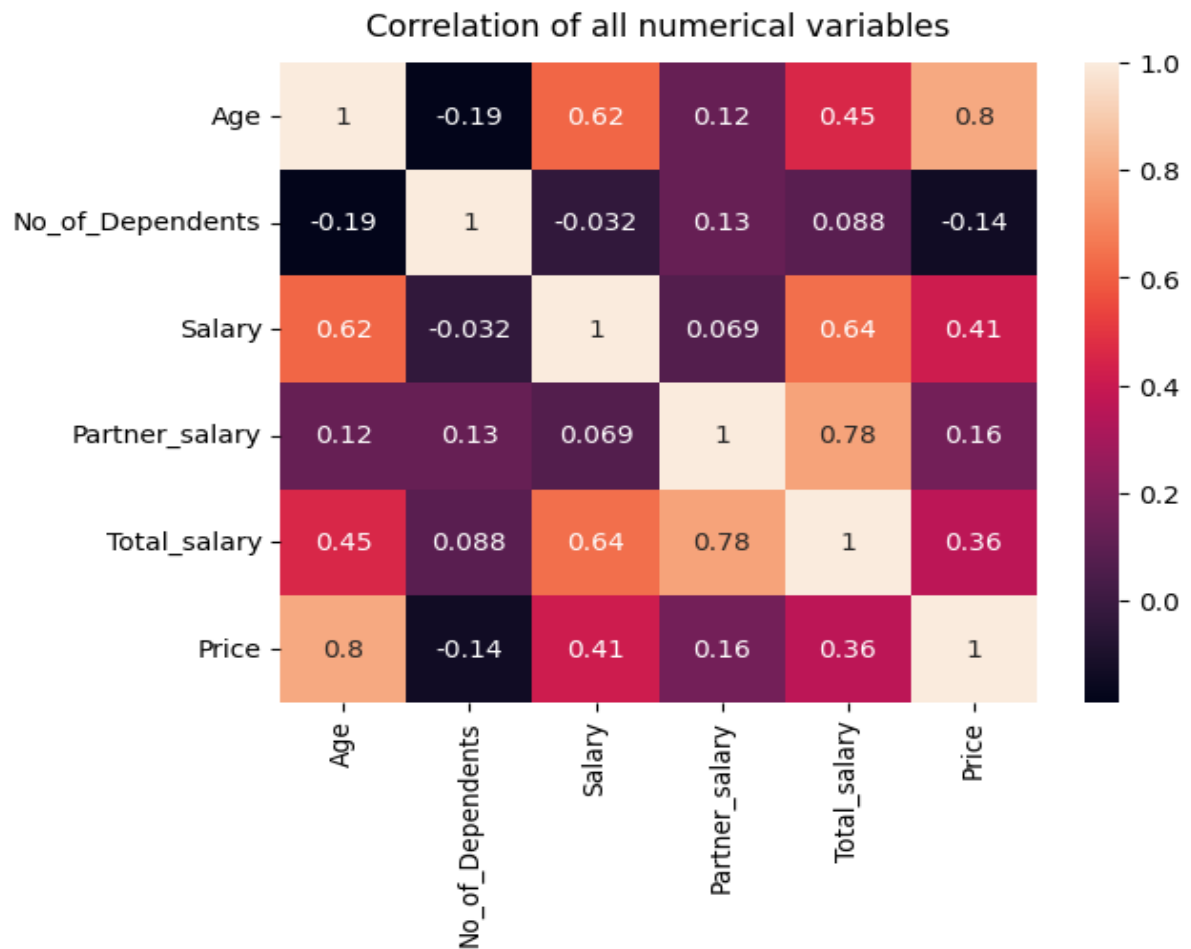


Interpretation:

- **Age and Number of Dependents:** Customers over the age of 30 tend to have more dependents, while those under 30 are more likely to have no dependents. This indicates a correlation between age and the number of dependents.

- **Salary and Age:** The graph shows an increasing trend where customer salaries rise with age. This suggests that, generally, older customers have higher salaries.
- **Salary and Number of Dependents:** Customers with salaries exceeding 85,000 are more likely to have 2 or more dependents. This indicates that higher earners tend to support more dependents.
- **Partner's Salary and Age:** Partners of customers who are under 30 years old are most likely to earn below 55,000. This suggests a correlation between the age of partners and their earnings.
- **Partner's Salary and Customer's Salary:** Customers earning below 45,000 typically have partners earning less than 44,000. This indicates that lower-income customers also tend to have lower-earning partners.
- **Total Salary and Age:** As customers age, their total salary tends to increase. This suggests a positive relationship between age and the combined income of customers and their partners.
- **Total Salary vs. Customer's Salary and Partner's Salary:** As both the customer's and their partner's salaries increase, the total salary of the customer also rises. This reflects that higher individual and combined earnings contribute to a higher total salary.
- **Price and Age:** As the customer's age increases, the price of their automobile also tends to increase. This indicates that older customers often own more expensive vehicles.
- **Price and Total Salary:** Customers with a total salary of less than 40,000 are likely to have automobiles priced under 55,000. This suggests that lower total salaries are associated with less expensive vehicles.

16. Correlation of all numerical variables

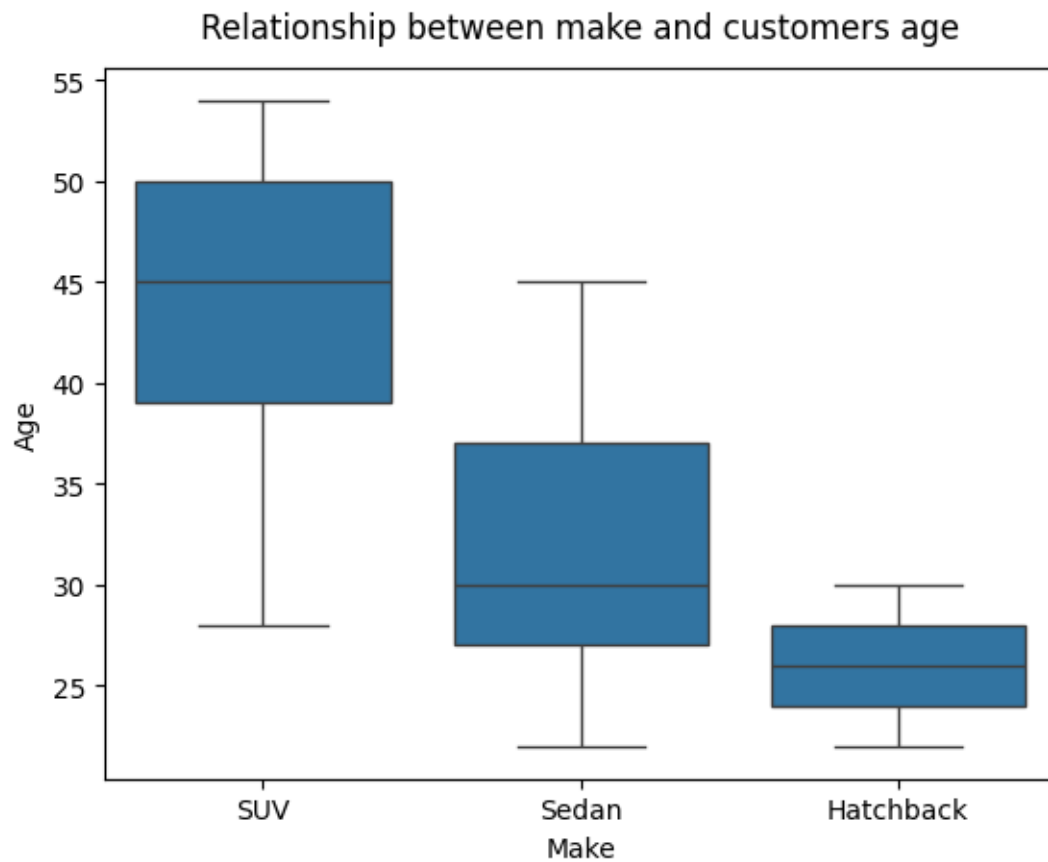


Interpretation:

- **Price and Number of Dependents:** The correlation between the price of automobiles and the number of dependents is the lowest, suggesting minimal relationship between these variables. This is followed by the correlation between age and the number of dependents, which is also relatively weak.
- **Total Salary and Partner's Salary:** The total salary shows the highest correlation with the partner's salary, indicating a strong relationship. This is followed by a strong correlation between total salary and the customer's own salary.
- **Total Salary and Age, Salary and Automobile Price:** The total salary has a moderate correlation with the customer's age, and there is also a moderate correlation between salary and the price of automobiles. This suggests a moderate relationship between these variables.

The relationship between categorical vs numerical variables

17. Relationship between make and customers age

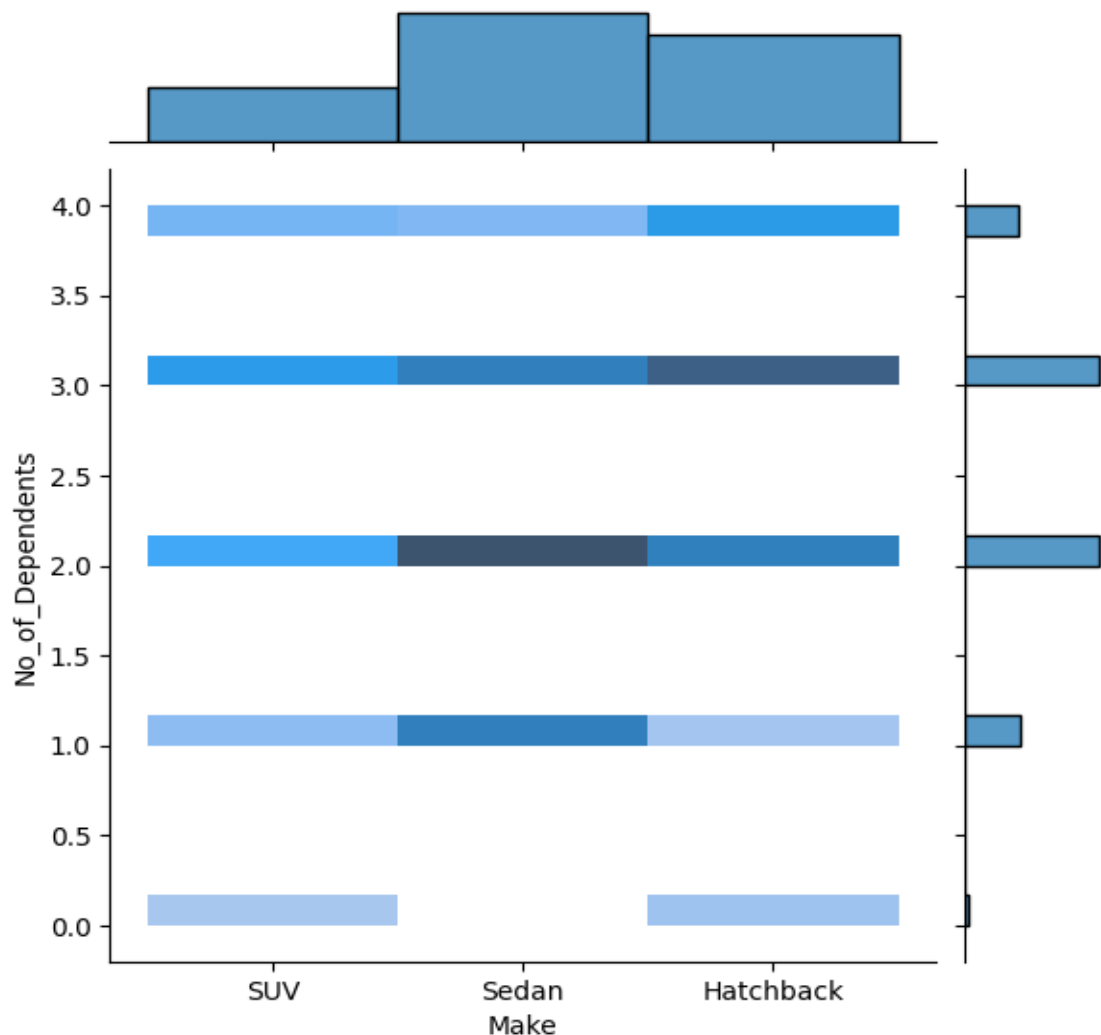


Interpretation:

The graph illustrates the relationship between age and the type of automobile preferred. Customers over 40 years old predominantly prefer SUVs. In contrast, those aged between 28 and 37 are most likely to choose sedans. Meanwhile, customers under 28 tend to favor hatchbacks. This indicates varying automobile preferences across different age groups.

18. Relationship between make and customers No of Dependents

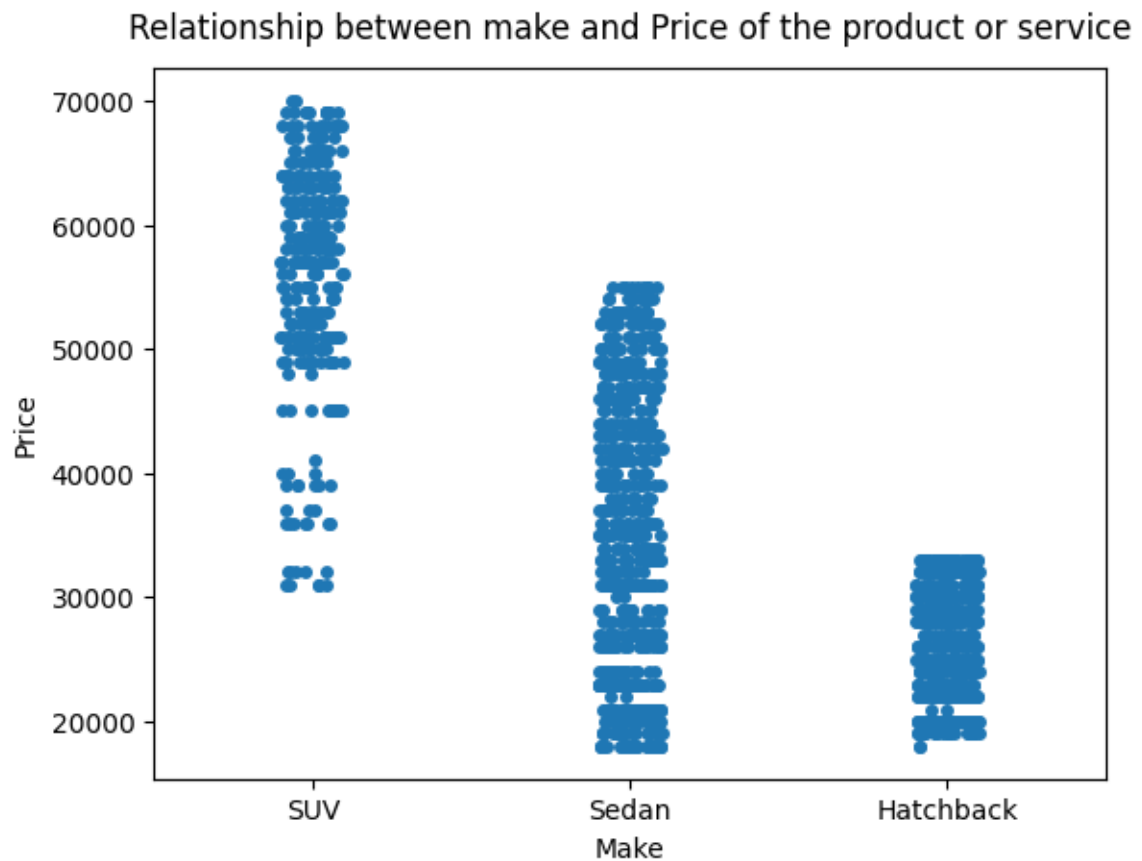
Relationship between make and customers No of Dependents



Interpretation:

The graph indicates that customers with 0 dependents show no particular preference for sedans. In contrast, customers with 2 dependents exhibit the highest preference for sedans. Those with 3 dependents tend to favor hatchbacks, and even customers with 4 dependents display a slightly increased preference for hatchbacks, highlighting a shift in vehicle choice as the number of dependents rises.

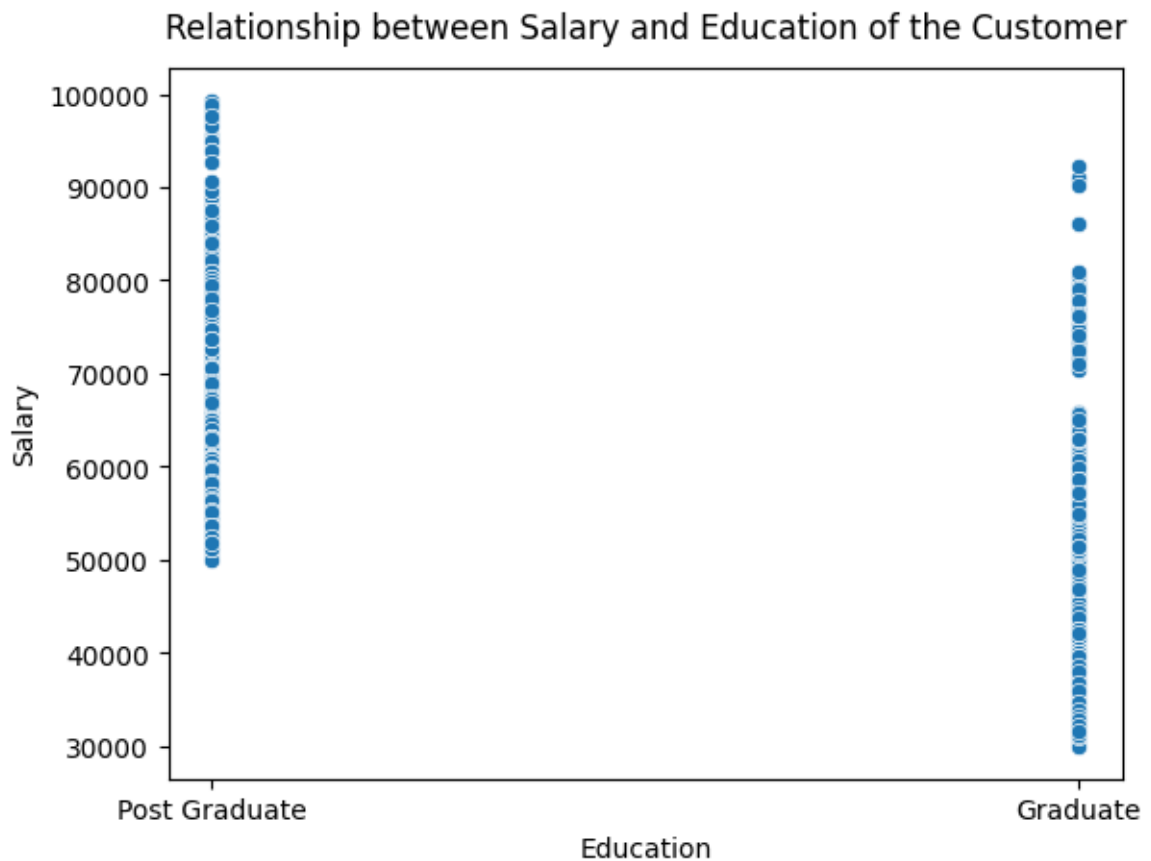
19. Relationship between make and Price of the product or service



Interpretation:

SUVs tend to be priced higher compared to other automobile makes. Sedans are priced within a range from below 58,000 to around 18,000, while hatchbacks are generally the most affordable, priced at the lowest end. This suggests that SUVs are positioned as premium options, with hatchbacks catering to more budget-conscious customers.

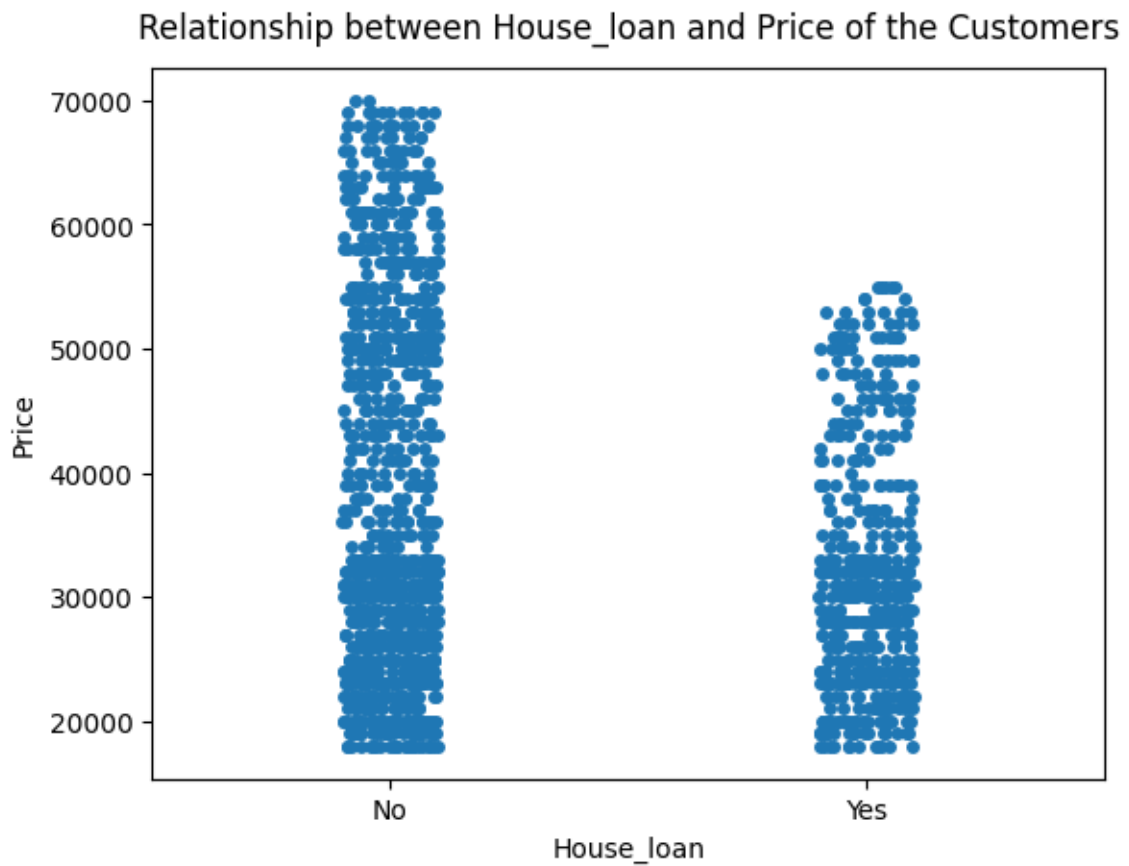
20. Relationship between Salary and Education of the Customer



Interpretation:

Most customers who have completed postgraduate studies earn a salary between 50,000 and 100,000. In contrast, the majority of customers with only a graduate degree tend to have slightly lower salary packages, ranging from less than 30,000 to 70,000. This indicates a positive correlation between higher education and increased earning potential.

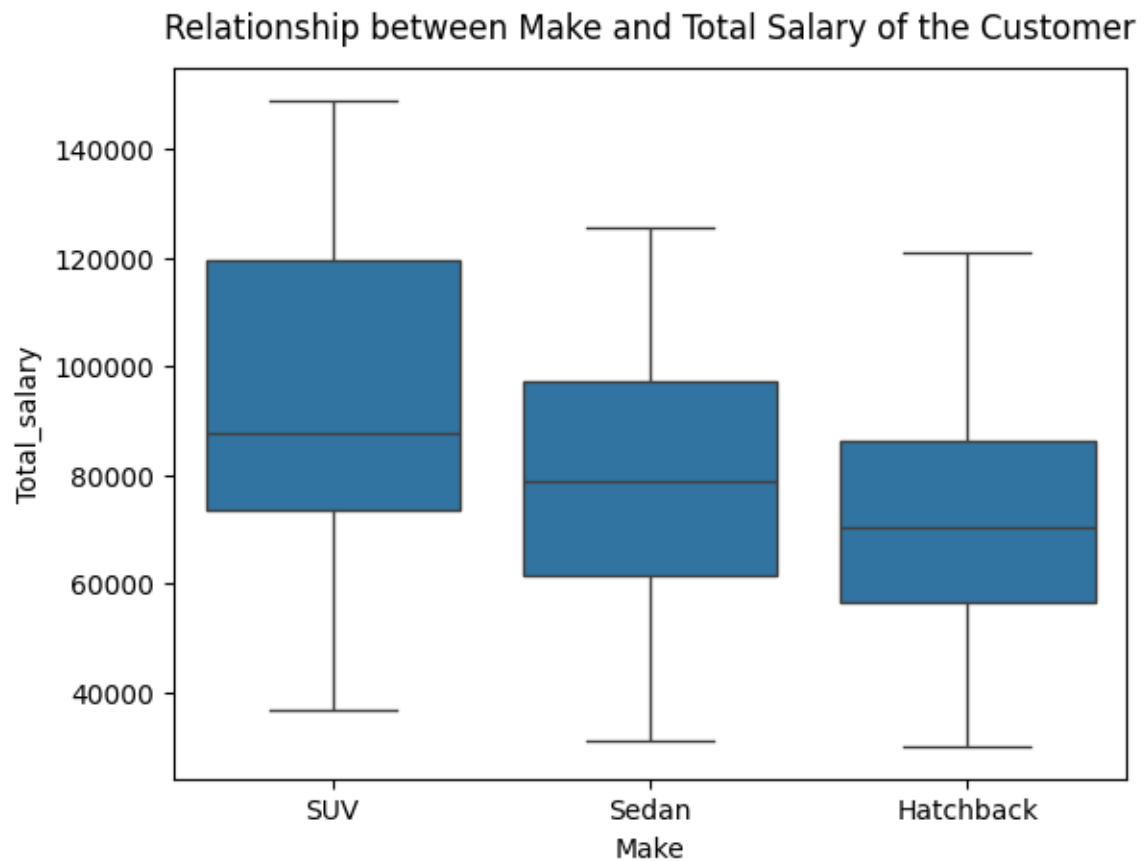
21. Relationship between House loan and Price of the Customers



Interpretation:

The graph represents the relationship between customers with house loans and the price of their automobiles. It shows that customers with house loans tend to purchase lower-priced automobiles compared to those without house loans, indicating a possible link between loan status and vehicle affordability.

22. Relationship between Make and Total Salary of the Customer

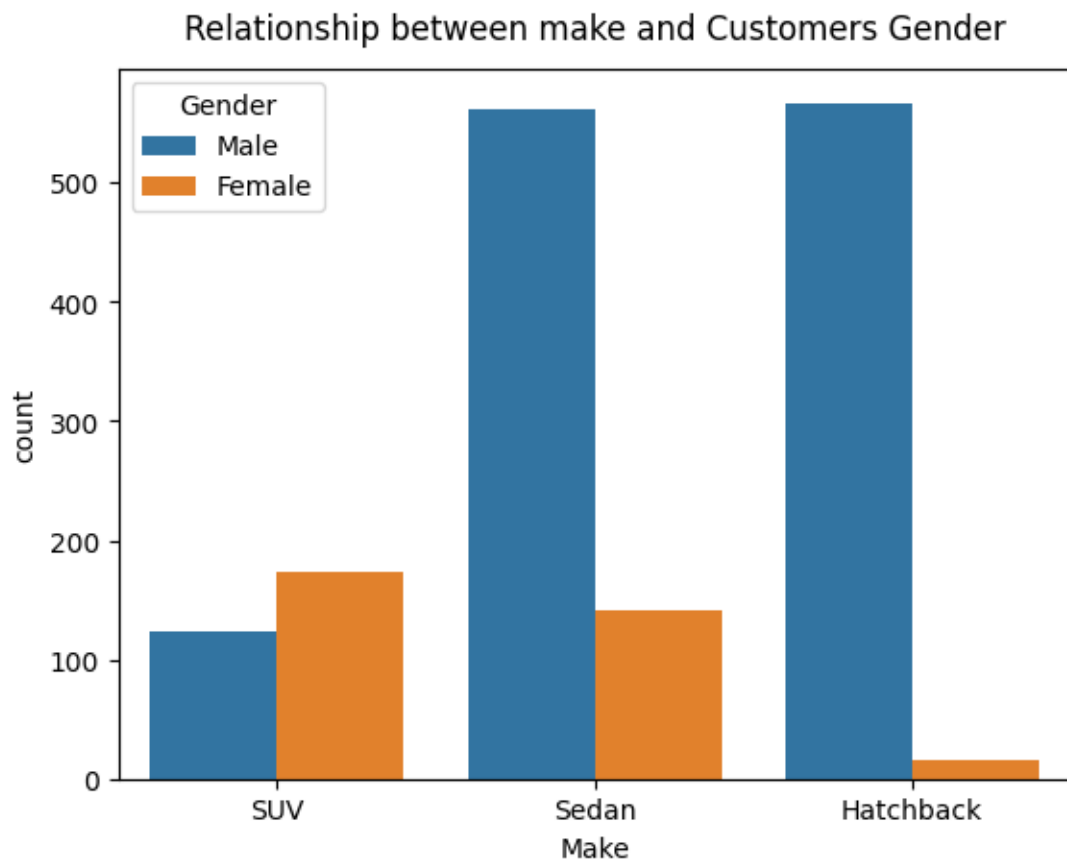


Interpretation:

The graph shows the relationship between car make and the total salary of customers. Customers with higher salaries are more likely to purchase SUVs compared to other makes. Those with moderate salaries tend to buy sedans, while individuals with comparatively lower salaries often opt for hatchbacks.

1.5.Key Answers for the key questions:

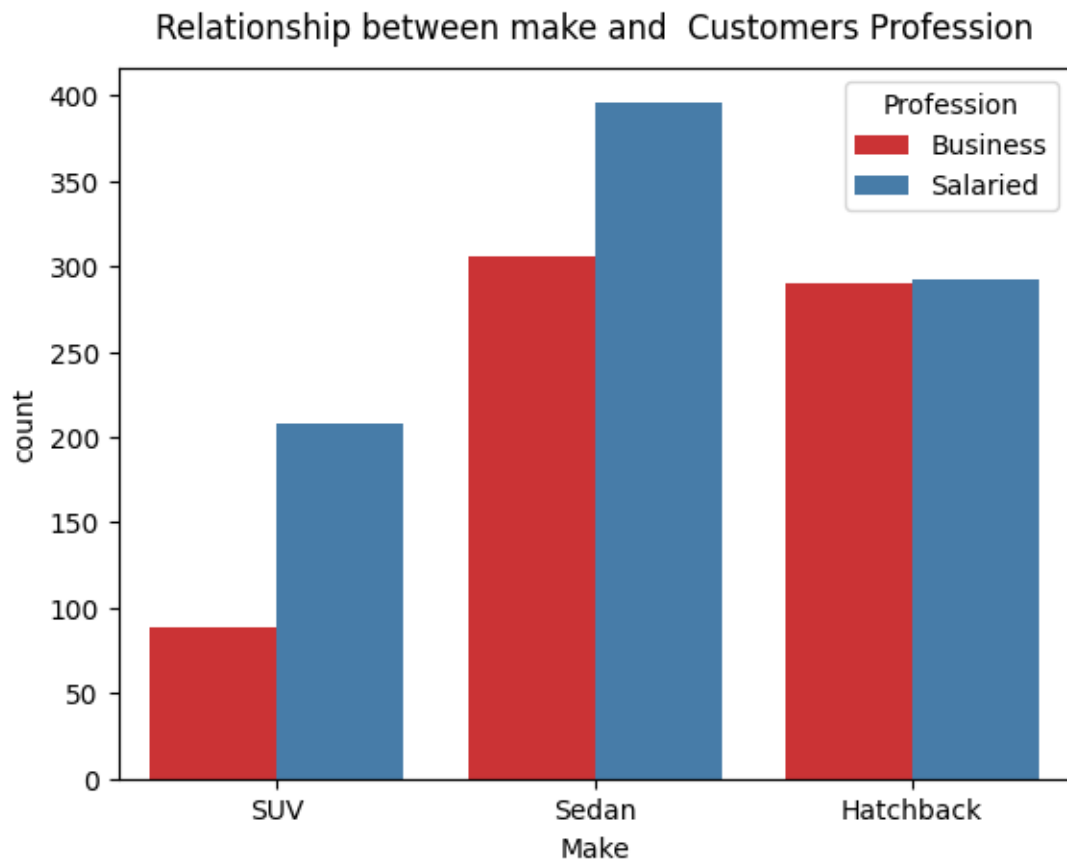
23. Relationship between make and Customers Gender



Answer 1:

Women prefer SUVs more than men. The data reveals that female customers are more inclined to choose SUVs compared to their male. This trend highlights a stronger preference for SUVs among women in the customer base. So male do not tend to prefer SUVs compared to woman.

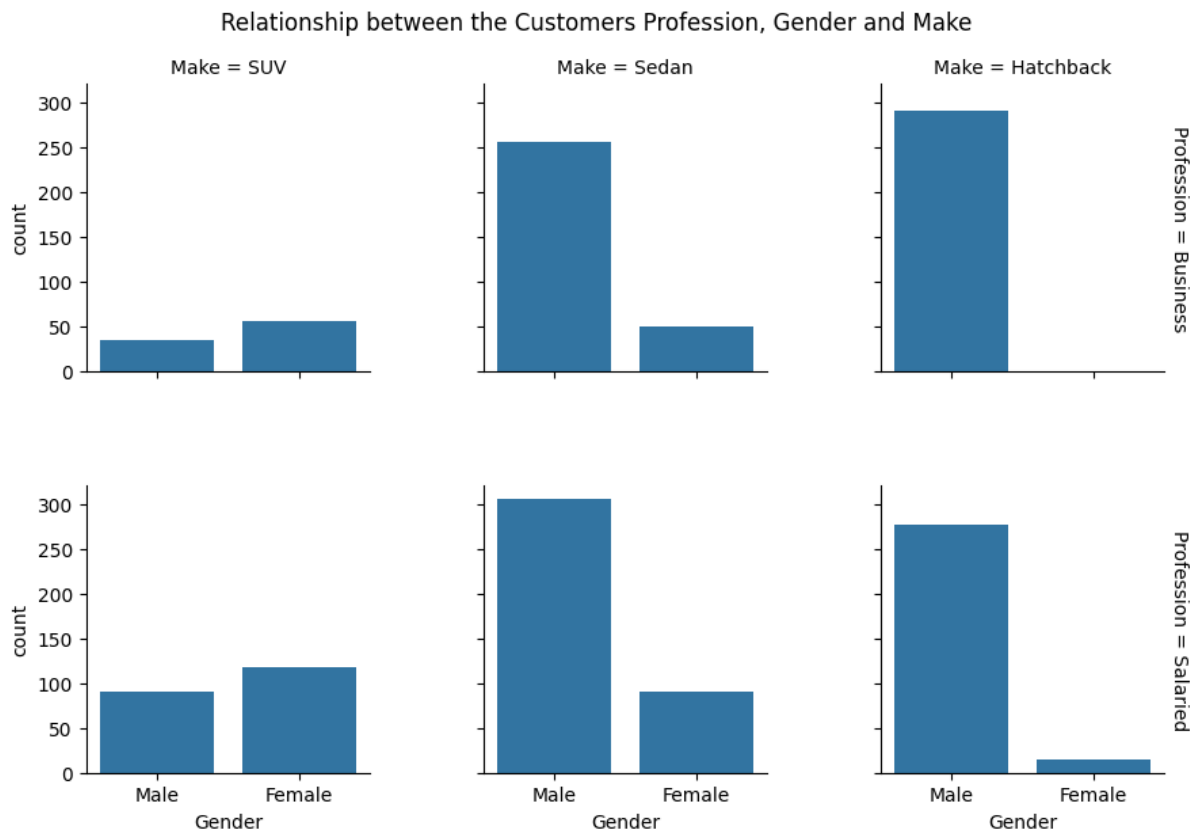
24. Relationship between make and Customers Profession



Answer 2:

A salaried person is more likely to buy a sedan compared to customers who are engaged in business. The data shows that individuals with salaried positions exhibit a higher preference for sedans, making sedans a popular choice among salaried customers.

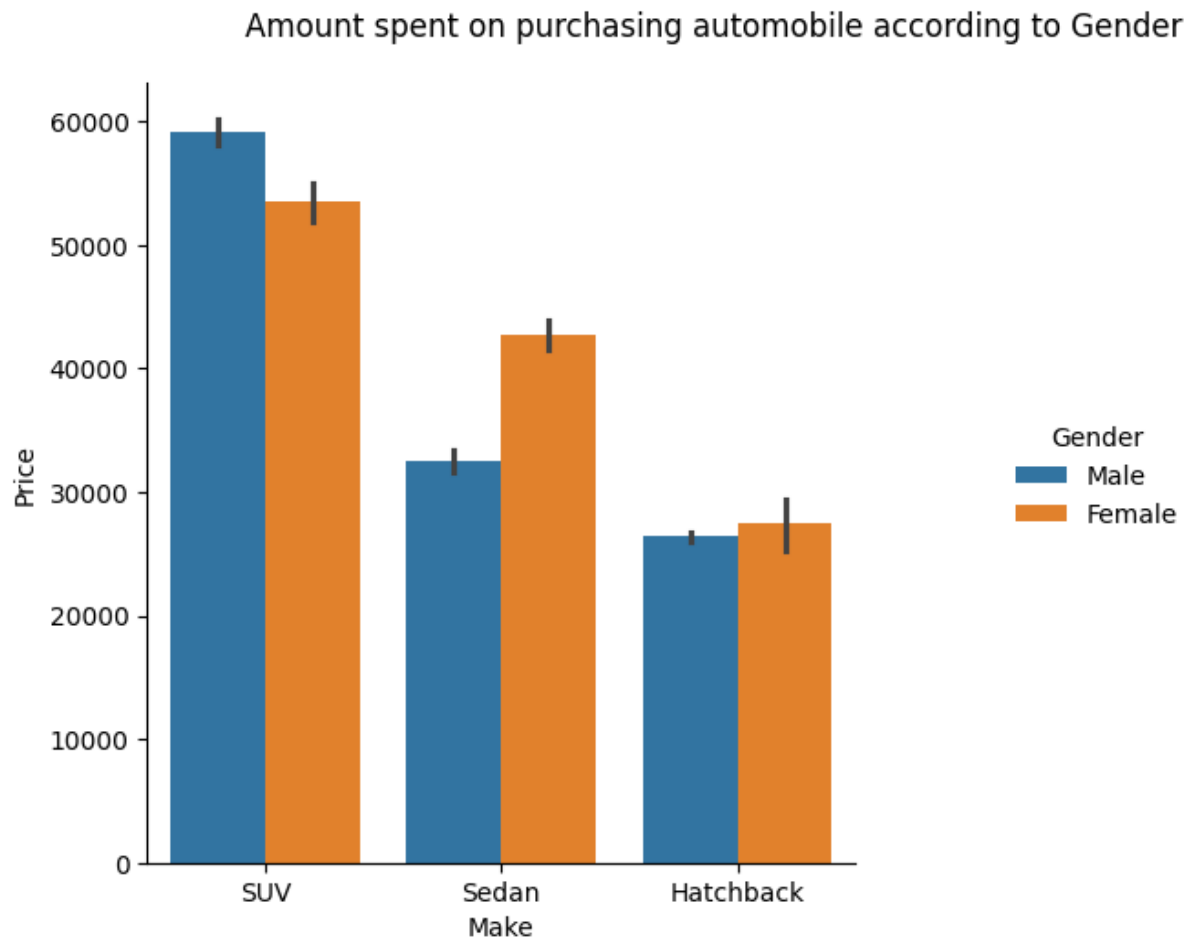
25. Relationship between the Customers Profession, Gender and Make



Answer 3:

The evidence contradicts Sheldon Cooper's claim. Data shows that salaried males actually prefer sedans over SUVs. This indicates that, contrary to the claim, salaried males are not an easier target for SUV sales compared to sedans.

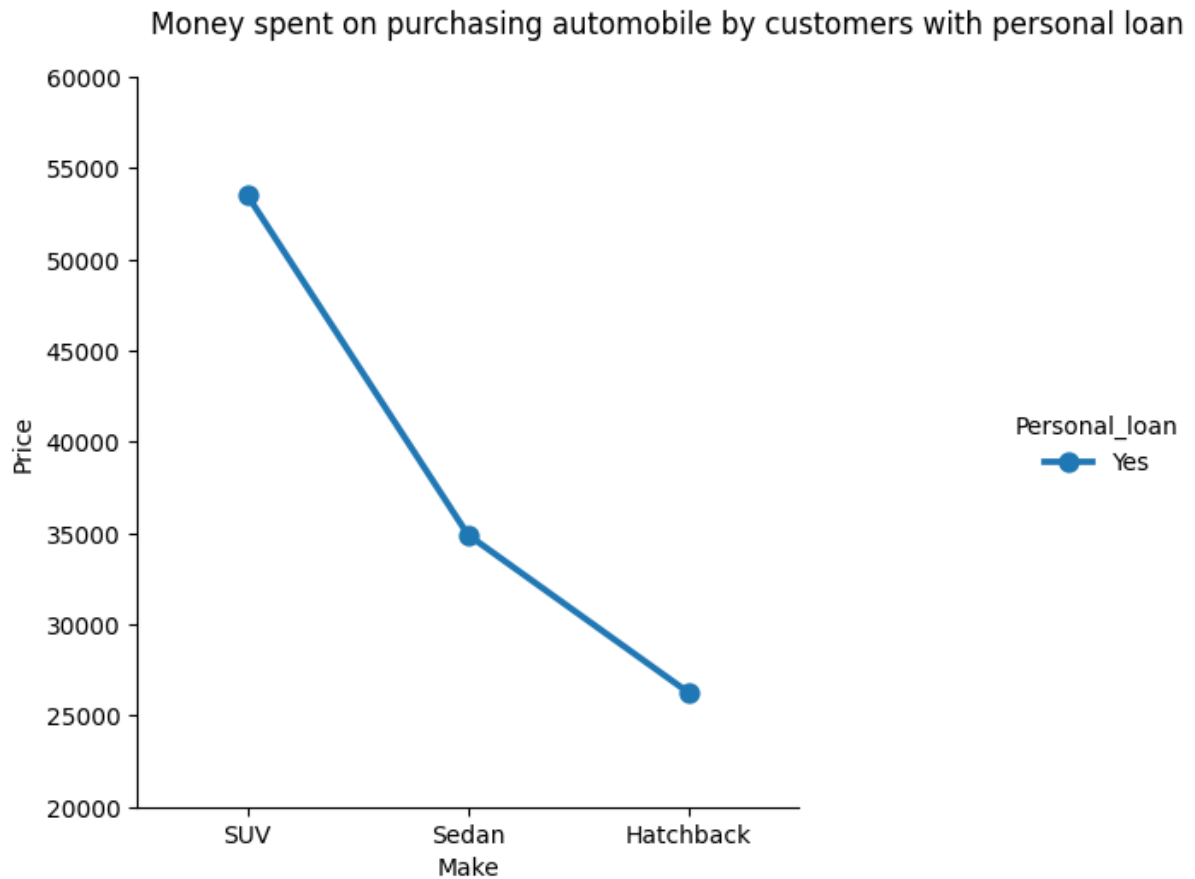
26. Amount spent on purchasing automobile according to Gender



Answer 4:

When considering SUVs, males tend to spend more money compared to females. However, in the case of sedans, females generally spend more than males. For hatchbacks, females also tend to spend slightly more than males.

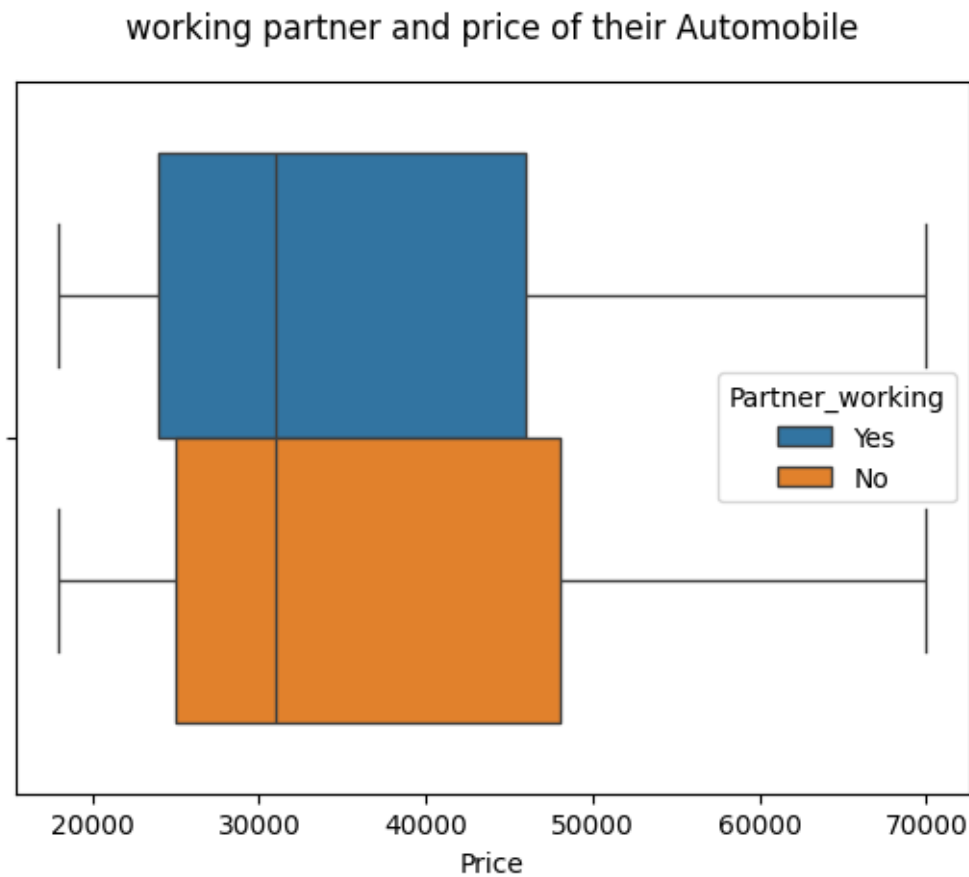
27. Money spent on purchasing automobile by customers with personal loan



Answer 5:

The graph shows the relationship between the price and make of automobiles for customers with personal loans. Customers with personal loans spend approximately 25,000 on hatchbacks, around 35,000 on sedans, and about 54,000 on SUVs. This indicates that customers with personal loans spend more on SUVs compared to hatchbacks and sedans.

28. working partner and price of their Automobile



Answer 6:

Having a working partner does not influence the purchase of higher-priced cars. In fact, customers with working partners tend to purchase cars at lower price points compared to those without working partners. This suggests that the presence of a working partner does not correlate with a preference for higher-priced vehicles.

1.6.Actionable Insights & Recommendations

Actionable Insights:

1) SUV Preference Among Females:

- a. Female customers prefer SUVs more than male customers.

2) Salaried Individuals and Sedans:

- a. Salaried individuals tend to prefer sedans over other vehicle types.

3) Spending Patterns by Vehicle Type:

- a. Customers with personal loans spend approximately 25,000 on hatchbacks, around 35,000 on sedans, and about 54,000 on SUVs.

4) Price Sensitivity Based on Vehicle Type:

- a. SUVs are priced higher compared to sedans and hatchbacks, yet customers with personal loans still spend more on SUVs.

5) Gender-Based Spending Trends:

- a. Males spend more on SUVs, while females spend more on sedans and hatchbacks.

Business Recommendations:

1) Target Marketing for SUVs:

- a. Focus marketing campaigns for SUVs on female customers and higher-income individuals. Highlight the premium features of SUVs and create targeted advertisements to attract these demographics.

2) Promote Sedans to Salaried Professionals:

- a. Design promotional activities specifically for salaried professionals emphasizing the practicality and affordability of sedans. Offer tailored deals to this segment.

3) Adjust Financing and Promotions for Personal Loan Holders:

- a. Develop special financing options and promotions to encourage customers with personal loans to consider higher-priced vehicles like SUVs. Highlight the benefits and flexible financing plans available.

4) **Refine Pricing Strategies:**

- a. Review and adjust the pricing strategy for SUVs to ensure it aligns with customer spending behavior. Consider offering value-added services or features to justify the higher price.

5) **Personalize Customer Interactions:**

- a. Utilize CRM systems to track and analyze customer preferences based on demographic and spending data. Provide personalized recommendations and improve customer service based on these insights.

6) **Monitor Market Trends:**

- a. Regularly analyze changing trends in vehicle preferences and customer demographics. Adapt marketing strategies and product offerings to align with evolving customer needs and market conditions.