

Business Report - 1

PG Program in Data Science and Business Analytics

submitted by

Sangram Keshari Patro

BATCH:PGPDSBA.O.AUG24.B



Contents

1	Context	2
2	Data Description	2
3	Data Overview	2
3.1	Importing necessary libraries and the dataset	2
3.2	Structure and type of data	3
3.3	Statistical summary	4
3.4	Missing value treatment	4
3.4.1	Treating 'Gender' column	4
3.4.2	Treating 'Partner_salary' column	4
4	Exploratory Data Analysis	4
4.1	Univariate Analysis	4
4.1.1	Numerical columns	4
4.1.2	Categorical columns	8
4.2	Bivariate Analysis	13
4.2.1	Numerical variables	13
4.2.2	Categorical vs numerical variables	16
4.2.3	Categorical vs categorical variables	19
4.3	Key questions to ponder	25
4.3.1	Do men tend to prefer SUVs more compared to women?	25
4.3.2	What is the likelihood of a salaried person buying a Sedan?	25
4.3.3	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?	25
4.3.4	How does the the amount spent on purchasing automobiles vary by gender?	26
4.3.5	How much money was spent on purchasing automobiles by individuals who took a personal loan?	27
4.3.6	How does having a working partner influence the purchase of higher-priced cars?	28
4.4	Some extra intersting questions to ponder	29
4.4.1	How does the level of Education impact price of cars purchased?	29
4.4.2	How does the total salary of customers influence the price range of different car models (SUV, Sedan, Hatchback)?	32

1 Context

They want to analyze the data to get a fair idea about the demand of customers which will help them in enhancing their customer experience. Suppose you are a Data Scientist at the company and the Data Science team has shared some of the key questions that need to be answered. Perform the data analysis to find answers to these questions that will help the company to improve the business.

2 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

3 Data Overview

3.1 Importing necessary libraries and the dataset

The dataframe is printed. It has 1581 rows & 14 columns.

Age	Gender	Profession	Marital_status	Education	No_of_Dependents	Personal_loan	House_loan	Partner_working	Salary	Partner_salary	Total_salary	Price	Make
53	Male	Business	Married	Post Graduate	4	No	No	Yes	99300	70700.0	170000	61000	SUV
53	Female	Salaried	Married	Post Graduate	4	Yes	No	Yes	95500	70300.0	165800	61000	SUV
53	Female	Salaried	Married	Post Graduate	3	No	No	Yes	97300	60700.0	158000	57000	SUV
53	Female	Salaried	Married	Graduate	2	Yes	No	Yes	72500	70300.0	142800	61000	SUV
53	Male	Salaried	Married	Post Graduate	3	No	No	Yes	79700	60200.0	139900	57000	SUV
...
22	Male	Salaried	Single	Graduate	2	No	Yes	No	33300	0.0	33300	27000	Hatchback
22	Male	Business	Married	Graduate	4	No	No	No	32000	NaN	32000	31000	Hatchback
22	Male	Business	Single	Graduate	2	No	Yes	No	32900	0.0	32900	30000	Hatchback
22	Male	Business	Married	Graduate	3	Yes	Yes	No	32200	NaN	32200	24000	Hatchback
22	Male	Salaried	Married	Graduate	4	No	No	No	31600	0.0	31600	31000	Hatchback

Figure 1: Dataframe

3.2 Structure and type of data

Data is explored further. Data doesn't have any duplicate rows.

```

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

```

Figure 2: Table depicting the datatype and Non-Null values in each column. This also denotes the columns having 'NaN' values.

3.3 Statistical summary

	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

Figure 3: Statistical summary of the data

From this table we can observe that there are outliers in the 'Total_salary' column

3.4 Missing value treatment

3.4.1 Treating 'Gender' column

Typos in the 'Gender' column were corrected. For instance, entries like 'Femal' and 'Female' were standardized to 'Female', while missing or NaN values were replaced with 'Unknown.'

3.4.2 Treating 'Partner_salary' column

The boxplot and histogram for the 'Partner_salary' column indicate no extreme outliers. However, despite the absence of outliers, it's preferable to use the median to replace NaN values rather than the mean because the data distribution is not Gaussian (i.e., it may be skewed or non-symmetric). The median is a more robust measure of central tendency in such cases, as it is less influenced by any skewness or irregularities in the data distribution.

4 Exploratory Data Analysis

4.1 Univariate Analysis

4.1.1 Numerical columns

- Age

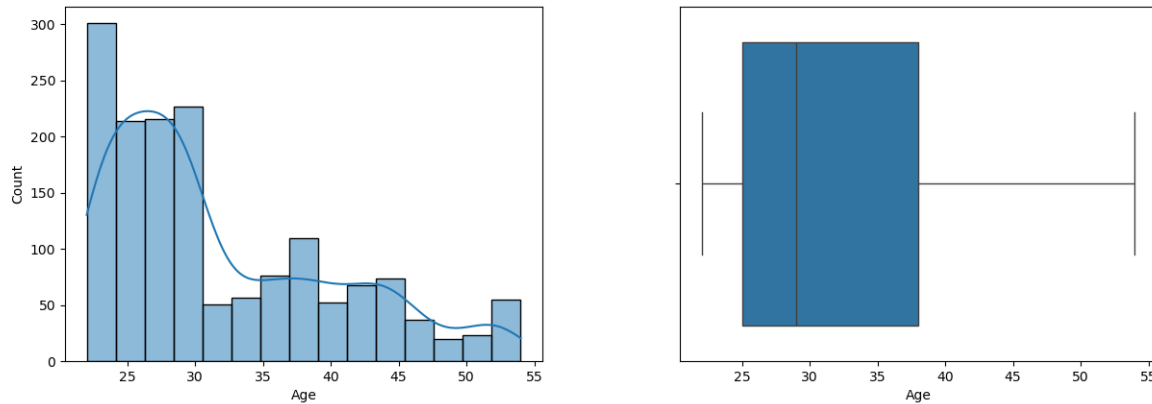


Figure 4: Histogram and boxplot of 'Age' column

1. The histogram and kernel density estimate depict the age distribution of customers. Notably, this distribution is right-skewed, indicating that there are more younger customers and relatively fewer older ones.
2. Most customers appear to be in their younger years (between 20 and 30). This insight suggests that targeting marketing efforts toward this younger demographic could yield better results.
3. The box plot reveals the spread of ages within the sample population. The interquartile range (IQR) spans from approximately 25 to 38, covering the middle 50% of ages. The median age is 29. Targeting this age group might yield the highest engagement, as this age group constitutes the bulk of the customer base.

- No_of_Dependents

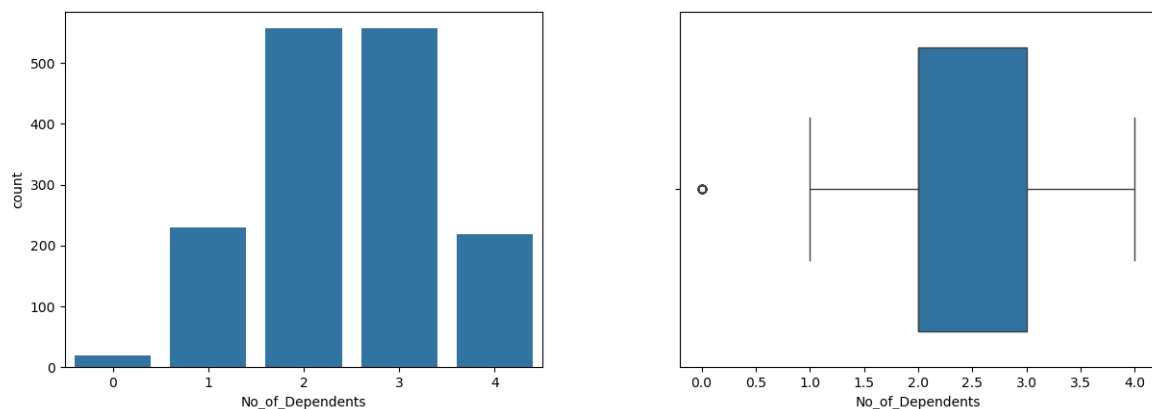


Figure 5: Histogram and boxplot of 'No_of_Dependents' column

1. The histogram displays the count of individuals based on the number of dependents they have. The tallest bars correspond to individuals with 2 or 3 dependents. Austo Motor Company can leverage this information to tailor their marketing strategies effectively related to family-oriented campaigns or financial planning.
2. In the box plot, we observe an outlier—an individual with no dependents. The interquartile range (IQR), which represents the middle 50% of the data, spans from approximately 2 to 3 dependents. We also can observe that the median line is coinciding with the Q1. The highest reported number of dependents is 4.

- **Salary**

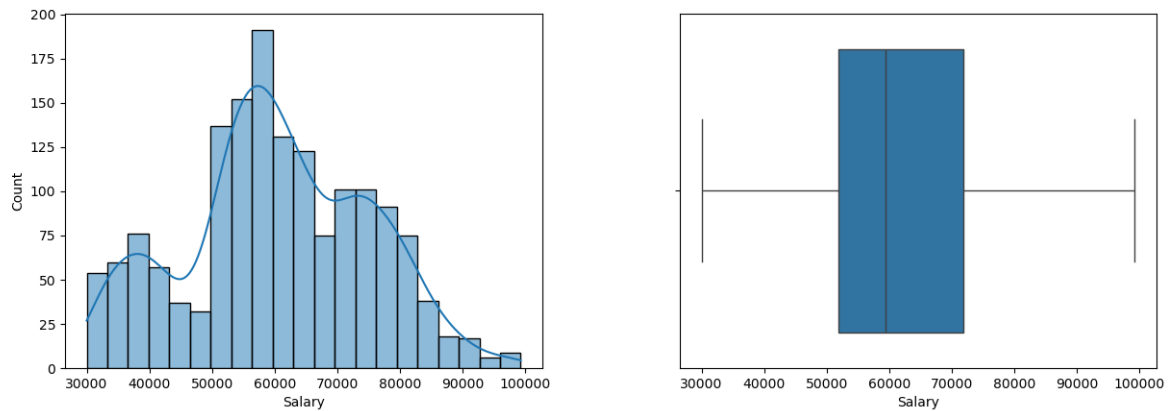


Figure 6: Histogram and boxplot of "Salary" column

1. The histogram shows the distribution of salaries. Most employees fall within the salary range around 60,000. The shape of the histogram is somewhat bell-shaped but skewed to the right, indicating fewer individuals with higher salaries and more with salaries below the peak.
2. If Austo's vehicles are considered luxury items, targeting higher salary brackets might be more effective. If affordability is key, focusing on the middle-income segment could improve campaign efficiency.
3. The interquartile range (IQR) spans from 60,500 Rs. to 95,900 Rs. Allocate resources effectively based on customer segments. Tailor messaging to resonate with different income groups.

- **Partner_salary**

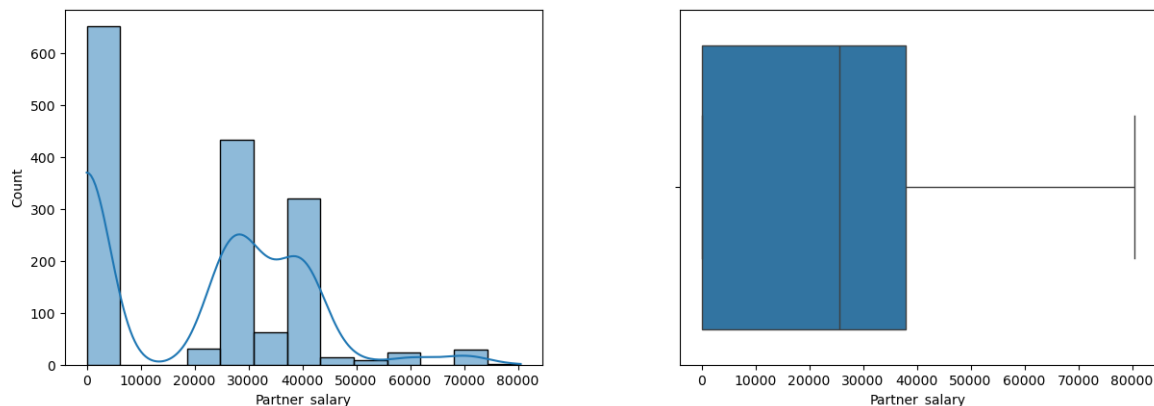


Figure 7: Histogram and boxplot of 'Partner_salary' column

1. The histogram shows that most partners have salaries in the lower range (around 0 to 10,000 Rs). Fewer partners fall into higher salary brackets. Customized marketing messages can be crafted for each segment. Low-income partners might respond better to cost-saving benefits, while high-income partners may prioritize performance or prestige.
2. The interquartile range (IQR) for the data lies between 0 and 38,300 Rs, with a median value of 25,600 Rs.

- **Total_salary**

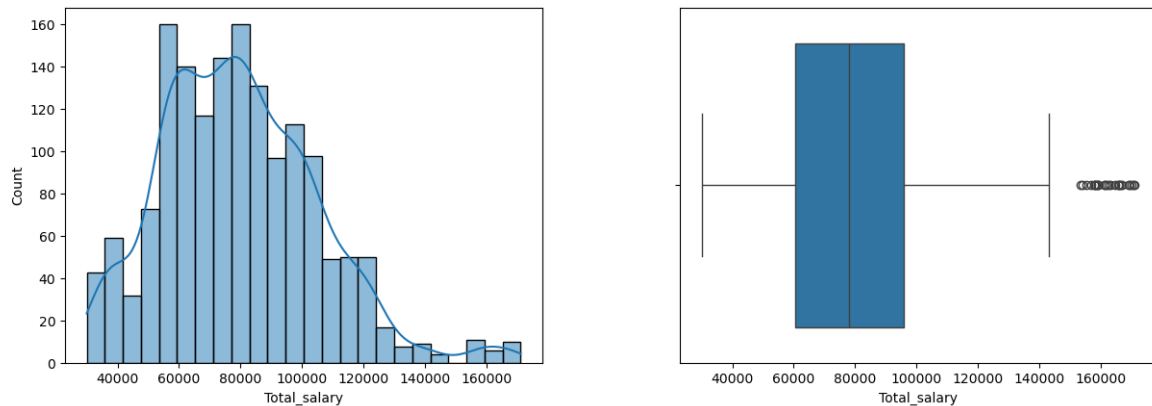


Figure 8: Histogram and boxplot of 'Total_salary' column

1. The histogram shows the distribution of total salaries among customers. There are two prominent peaks in salary frequency: around 60,000 Rs. and 80,000 Rs.
2. The median salary 78,000 Rs. provides insight into the typical salary level. The interquartile range (IQR) spans from 51,900 Rs. to 71,800 Rs. We can observe the outliers in this dataset having high total salary. The mean total salary is 79,626 Rs. which is close to the median i.e. the data is not skewed.

Treating outliers in Total_salary column

All the values smaller than lower_whisker will be assigned the value of lower_whisker & all the values greater than upper_whisker will be assigned the value of upper_whisker. The plots after treatment is attached below.

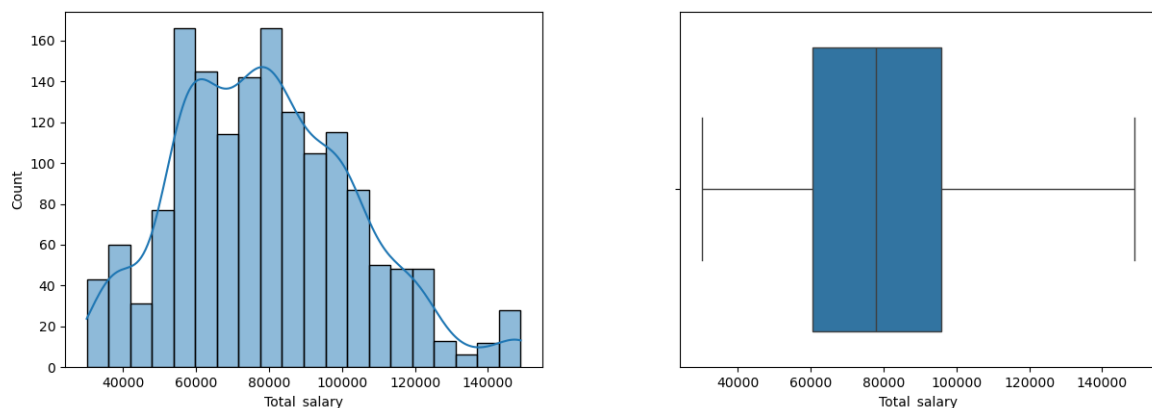


Figure 9: Histogram and boxplot of 'Total_salary' column after treatment of outliers

Although the overall shape of the curve remains the same, the boxplot no longer includes outliers.

- Price

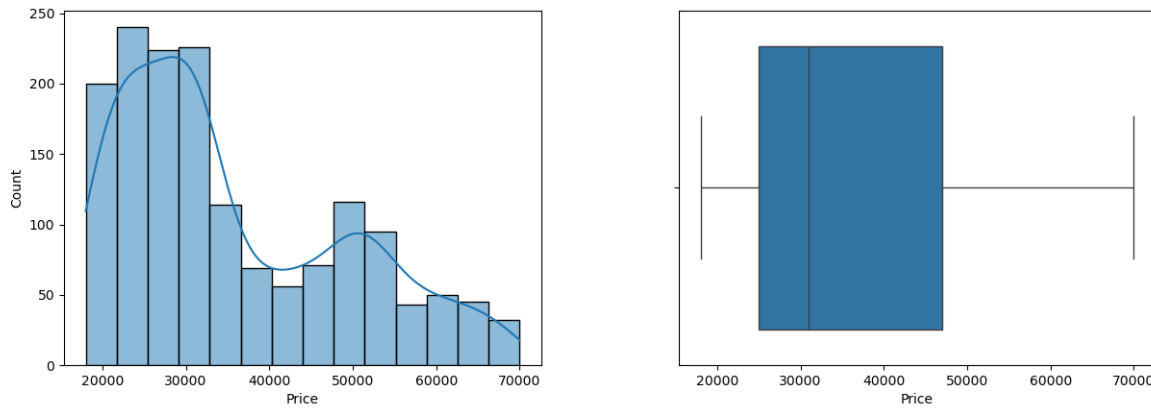


Figure 10: Histogram and boxplot of 'Price' column

1. The histogram shows that the majority of customers prefer cars priced between approximately 20,000 and 30,000 units. This price range has the highest concentration of data points. We can consider focusing marketing efforts on this segment, as it represents the largest customer base.
2. Most prices cluster around the median, suggesting a common price preference among customers.

4.1.2 Categorical columns

- Gender

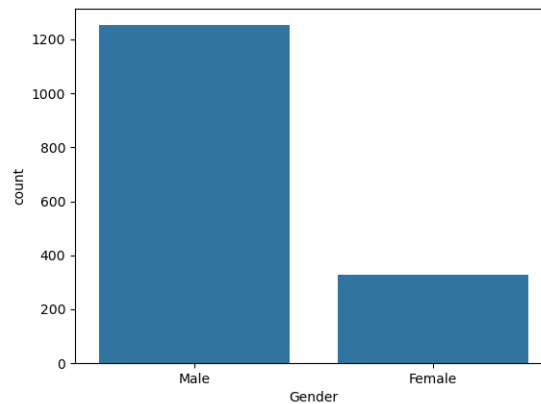


Figure 11: Barchart of 'Gender' column

Observations

Observations: The provided bar chart indicates a higher demand among **male** customers compared to **female** customers.

Insights

Insights: To enhance customer experience and improve marketing efficiency, Austo Motor Company should consider tailoring campaigns to attract more **female** customers or further capitalize on the high **male** interest.

- **Profession**

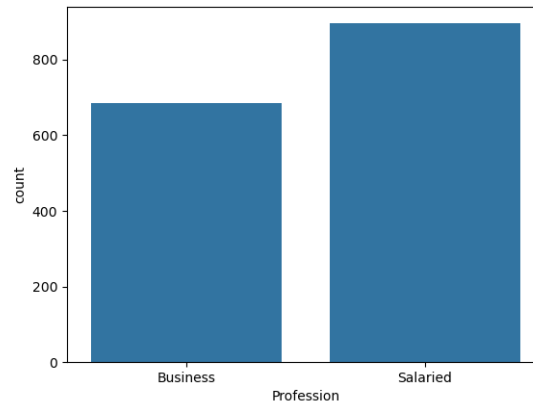


Figure 12: Barchart of 'Profession' column

Observations

Observations:

- Notably, the number of **Salaried** individuals is significantly higher, indicating a larger representation in the dataset.

Insights

Insights:

- Given the higher proportion of **Salaried** professionals, it appears that this group could be a promising target market for marketing campaigns.
- To improve campaign efficiency, consider directing marketing efforts toward **Salaried** customers.
- **Marital_status**

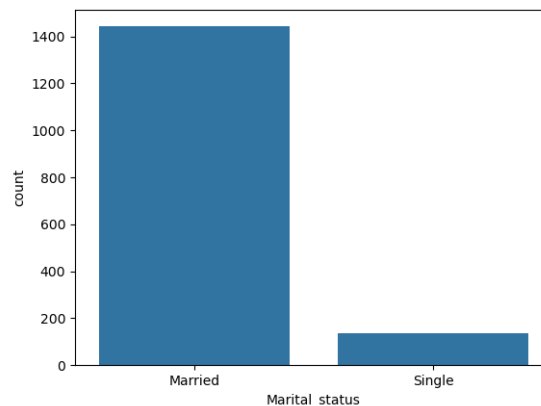


Figure 13: Barchart of 'Marital_status' column

Observations

1. The majority of customers are **Married**, with a count exceeding 1400.

Insights

1. **Marketing Strategy:** Focus **Sedan/Hatchback** ads on **Married** individuals; tailor **Hatchback** campaigns for **Singles**.

- **Education**

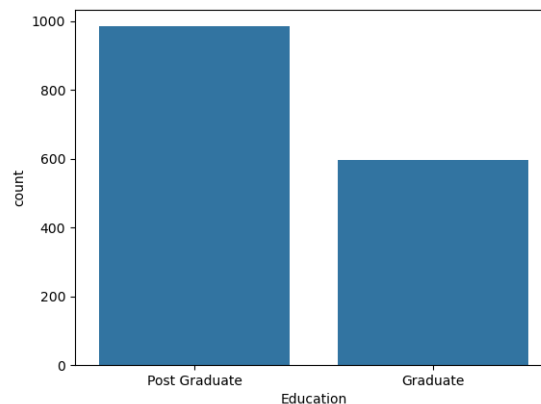


Figure 14: Barchart of '**Education**' column

Observations

1. The **bar chart** analysis indicates higher counts for **Post Graduates** compared to **Graduates**.

Insights

1. This insight can guide Austo Motor Company in enhancing marketing strategies towards a more educated customer base, potentially improving customer engagement and sales.

- **Personal_loan**

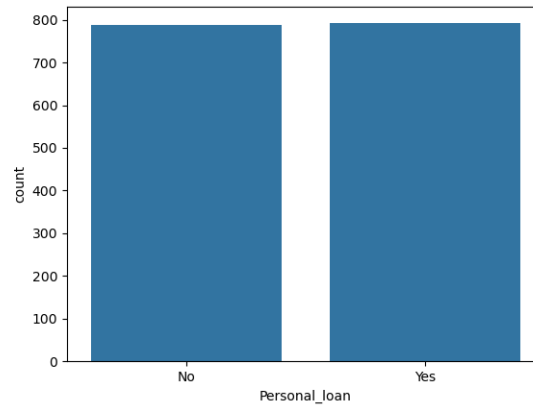


Figure 15: Barchart of 'Personal_loan' column

Observations

The bar chart shows a nearly equal distribution of customers who have taken personal loans ("Yes") and those who haven't ("No").

Insights

Given the balanced distribution, the marketing campaign should target both loan and non-loan customers. Further analysis could explore campaign effectiveness across different customer segments (e.g., SUV, Sedan, Hatchback).

- **House_loan**

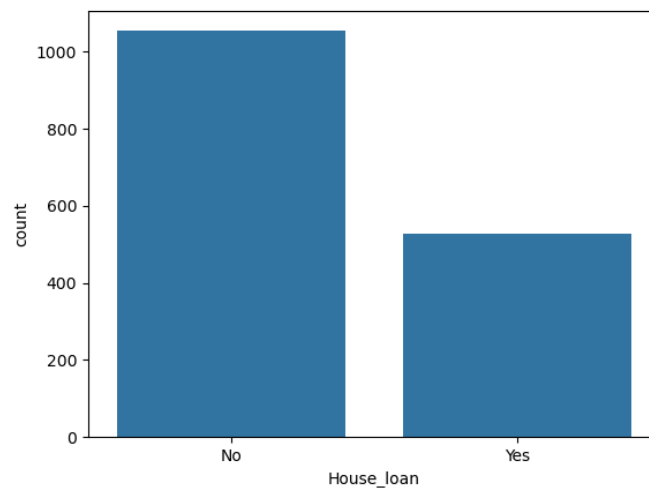


Figure 16: Barchart of 'House_loan' column

Observations

The majority of customers do not have house loans.

Insights

This information can guide targeted marketing efforts based on customer loan status.

- **Partner_working**

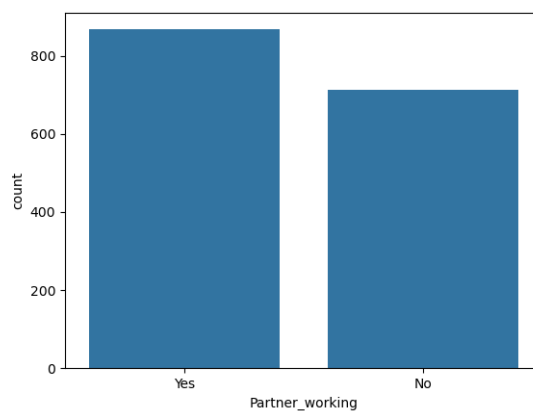


Figure 17: Barchart of 'Partner_working' column

Observations

Customers with working partners have a higher count.

Insights

Consider partner employment status in marketing strategies..

- **Make**

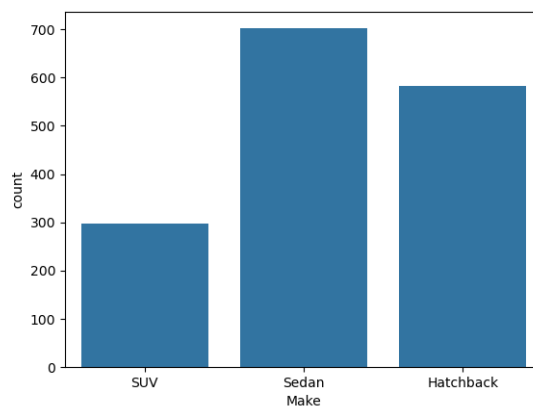


Figure 18: Barchart of 'Make' column

Observations

The provided image shows a bar chart with counts for three car models: **SUV**, **Sedan**, and **Hatchback**. The **Sedan** model exhibits the highest demand with around 700 units, followed by **Hatchback** at approximately 580 units, and **SUV** at around 300 units. This data is crucial for Austo Motor Company's marketing campaign efficiency analysis.

Insights

Consider partner employment status in marketing strategies..

4.2 Bivariate Analysis

4.2.1 Numerical variables

- **Heatmap**

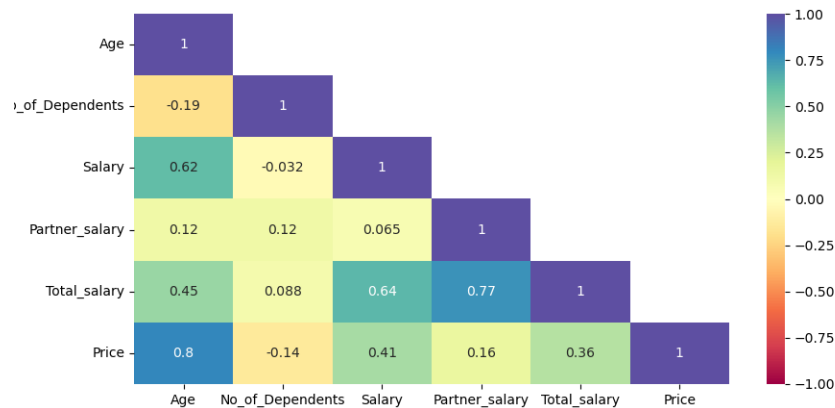


Figure 19: Heatmap of all numerical variables

Key Insights:

- **Age** has a strong positive correlation with **Price** (0.8), indicating older customers prefer expensive cars.
- **Salary** and **Price** have a moderate positive correlation (0.41), showing higher salaries lead to more expensive purchases.
- **Total salary** shows a moderate correlation with **Price** (0.36), highlighting household income's influence on car prices.
- **No. of Dependents** has a weak negative correlation with **Price** (-0.14), indicating a slight decrease in expensive car purchases with more dependents.

- **Pairplot**

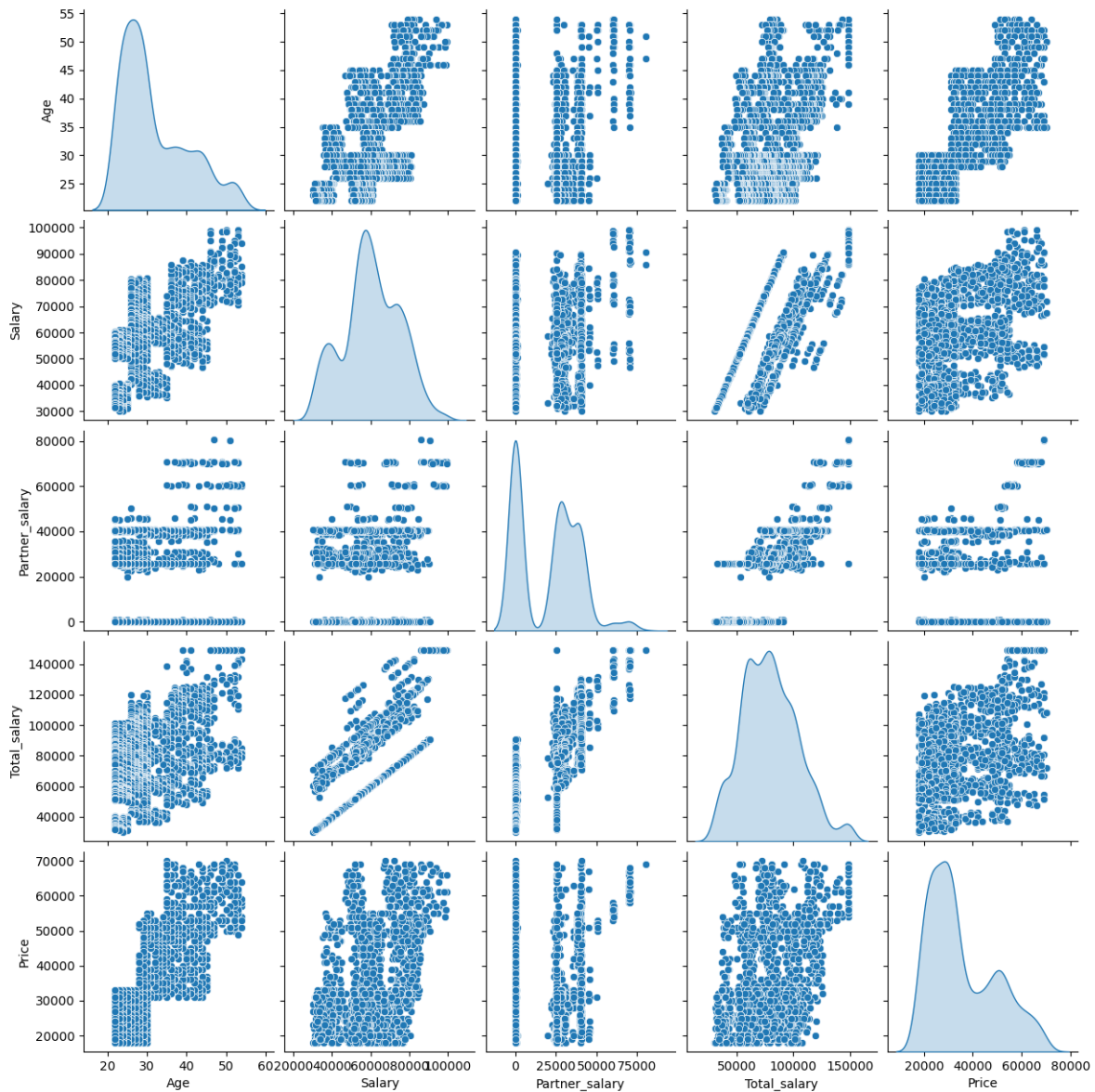


Figure 20: Pairplot of all numerical variables

Observations

Diagonal (Density Plots):

- **Age:** Skewed left, suggesting most customers are younger (around 25-35 years old).
- **Salary:** Mostly centered around 40,000-60,000, indicating a middle-income segment.
- **Partner_salary:** Bimodal distribution with peaks at 0 (no partner salary) and around 25,000.
- **Total_salary:** Peaks around 50,000-100,000, indicating most households earn in this range.
- **Price:** Slightly skewed to the right, with many cars priced between 20,000 and 50,000.

Scatter Plots:

- **Age vs Salary:** Positive correlation. Younger customers tend to have lower salaries, while older customers have higher salaries.

- **Age vs Partner_salary:** No clear pattern, suggesting partner salary does not vary significantly with age.
- **Age vs Total_salary:** Positive correlation. Older customers generally have higher combined household incomes.
- **Age vs Price:** Strong positive correlation. Older customers tend to buy more expensive cars.
- **Salary vs Partner_salary:** Weak relationship. Many customers with similar salaries have a wide range of partner salaries, including zero.
- **Salary vs Total_salary:** Strong linear correlation. Higher individual salaries lead to higher combined household incomes.
- **Salary vs Price:** Moderate positive correlation. Higher salaries lead to purchases of higher-priced cars.
- **Partner_salary vs Total_salary:** Moderate correlation, but many customers report a partner salary of zero, reducing the strength of this relationship.
- **Partner_salary vs Price:** Weak positive correlation, but no strong trend indicating a direct influence of partner salary on car price.
- **Total_salary vs Price:** Positive correlation. Higher household income is associated with higher-priced car purchases.

Key Insights

- **Age** strongly influences **Price** and **Total_salary**, showing that older customers tend to buy more expensive cars and have higher combined incomes.
- **Salary** is a key factor for both **Total_salary** and **Price**, suggesting that customers with higher individual incomes are likely to buy higher-end models.
- **Partner_salary** has a limited direct impact on **Price**, indicating individual income might be more important in decision-making.

4.2.2 Categorical vs numerical variables

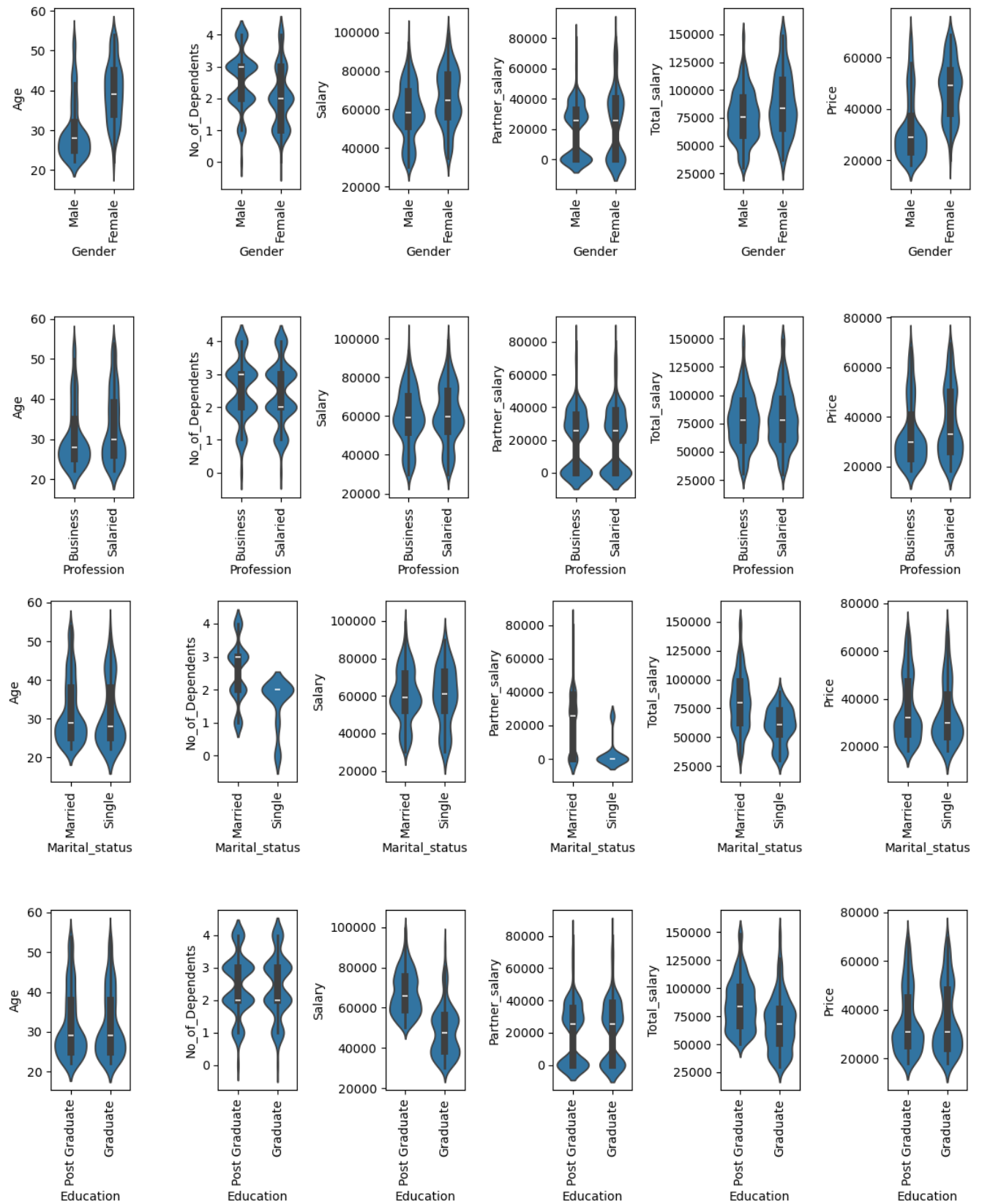


Figure 21: Violinplot of categorical vs numerical variables

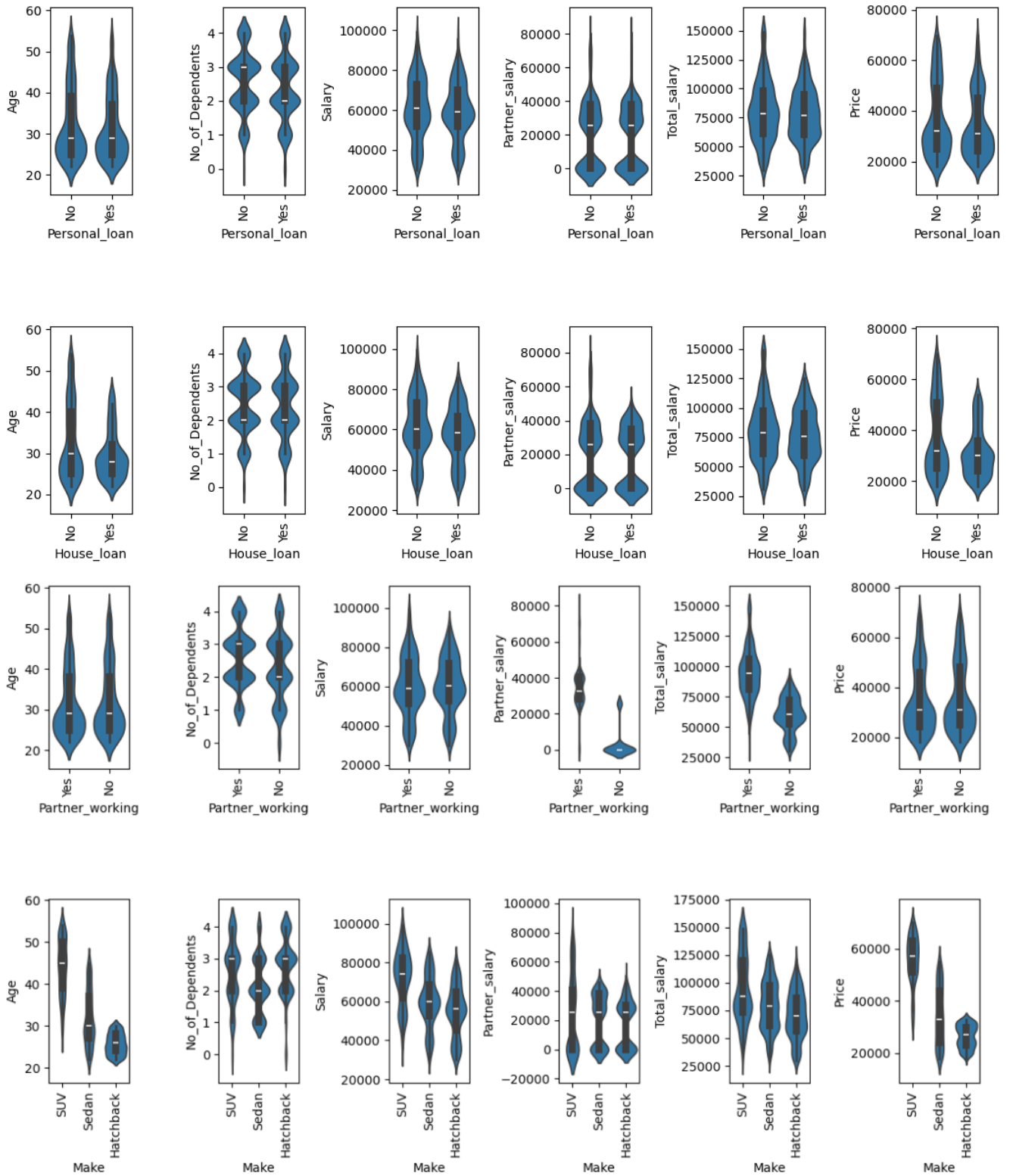


Figure 22: Violinplot of categorical vs numerical variables

Observations

Diagonal Plots (Violin Plots of Numerical Variables by Categorical Variables)

Gender:

- **Age:** Females slightly older on average.
- **No_of_Dependents:** Females have fewer dependents, mostly 1-2.
- **Salary:** Males have more varied salaries but median salary is higher for female as older females are high in number in our dataset.
- **Partner_salary:** Similar distributions across genders, many with no partner salary.
- **Total_salary:** Females have higher median total salaries, broader distribution.
- **Price:** Females buy more expensive cars.

Profession:

- **Age:** A larger proportion of business professionals fall within the 20-30 age range compared to salaried customers.
- **No_of_Dependents:** Slightly more dependents in business.
- **Salary:** Median salary of both professionals are same.
- **Partner_salary:** No notable difference.
- **Total_salary:** Median salary of both professionals are same.
- **Price:** Business professionals buy more number of less-priced cars than salaried professionals.

Marital Status:

- **Age:** Married people are generally older.
- **No_of_Dependents:** Married individuals have more dependents.
- **Salary:** Median salary is slightly higher for single people.
- **Partner_salary:** Broader range for married individuals, many without partner salary. This indicates that a salary of 25,600 Rs was incorrectly entered by approximately 11.6 percentage customers, so we replaced these values with 0.
- **Total_salary:** Married individuals have higher incomes.
- **Price:** Married people buy more expensive cars.

Education:

- **Age:** Postgraduates tend to be slightly older.
- **No_of_Dependents:** No major difference.
- **Salary:** Post graduates have higher salaries.
- **Partner_salary:** Similar distributions by education level. A higher number of postgraduates have partners with no salary compared to graduates.
- **Total_salary:** Postgraduates have higher household incomes.
- **Price:** Surprisingly graduates buy more expensive cars.

Personal Loan:

- **Age:** Loan holders are younger.

- **No_of_Dependents:** Number of dependents are less with customers with a personal loan considering the median.
- **Salary:** Slightly lower for loan holders.
- **Partner_salary:** No major differences.
- **Total_salary:** Slightly lower total incomes for loan holders.
- **Price:** Loan holders buy less expensive cars.

House Loan:

- **Age:** Older individuals doesn't hold house loans.
- **No_of_Dependents:** No major differences.
- **Salary:** Lower salaries for loan holders.
- **Partner_salary:** Customers with higher partner salaries doesn't hold loan.
- **Total_salary:** Lower total incomes for house loan holders.
- **Price:** Loan holders buy cheaper cars.

Partner Working:

- **Age:** No significant difference.
- **No_of_Dependents:** Customers with working partners have zero dependents.
- **Salary:** Slightly lower for individuals with working partners.
- **Partner_salary:** This indicates that a salary of 25,600 Rs was incorrectly entered by approximately 12.6 percentage customers, so we replaced these values with 0.
- **Total_salary:** Higher total incomes for working partners.
- **Price:** slightly more customers buy expensive cars without working partners .

Make (Car Type - SUV, Sedan, Hatchback):

- **Age:** SUV buyers are older, Hatchback buyers younger.
- **No_of_Dependents:** Sedan buyers have less dependents.
- **Salary:** Higher for SUV buyers, lower for Hatchback buyers.
- **Partner_salary:** Customers with high partner salaries buy SUVs. Customers with zero partner salaries prefer Hatchback
- **Total_salary:** Higher for SUV buyers.
- **Price:** SUVs are pricier, Hatchbacks are cheaper.

Key Insights

- **Age:** Older buyers favor expensive cars (SUVs), with higher incomes.
- **Gender:** Females earn more and buy pricier cars as the male customers in our dataset are more younger their median salaries are less as compared to female. We need to add more number of older males to get a better insight.
- **Profession:** Salaried professionals earn more and buy pricier cars.
- **Marital Status:** Married individuals have lower incomes , but buy more expensive cars.
- **Education:** Postgraduates have higher incomes, but graduates buy more expensive cars.
- **Loans:** Loan holders (personal/house) buy lower-priced cars.
- **Partner Working:** Working partners boost household incomes and car price.
- **Make:** SUVs are bought by wealthier, older individuals; Hatchbacks by younger, lower-income buyers.

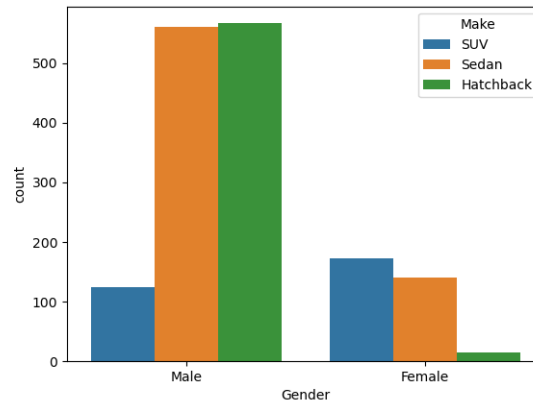


Figure 23: Barchart of 'Gender' column

Observations

1. **Males** prefer **Sedans** and **Hatchbacks** equally, with fewer choosing **SUVs**.
2. **Females** slightly favor **SUVs** over **Sedans**, with low interest in **Hatchbacks**.
3. Overall, **males** dominate the **Sedan** and **Hatchback** categories, while **SUVs** are more balanced across genders.

Insights

1. **Gender impacts car preferences:** Males favor **Sedans/Hatchbacks**, females prefer **SUVs**.
2. **Marketing Strategy:** Target **males** for **Sedans/Hatchbacks** and **females** for **SUVs**.
3. **Product Focus:** Tailor promotions to each gender's preferences. For instance, offering promotions, features, and advertisements that cater to males for Sedans and Hatchbacks, and to females for SUVs.
4. **Supply Balance:** Align inventory with gender-based demand to maximize efficiency.

This approach will help **Austo Motor Company** optimize customer engagement and product offerings.

* Profession vs Make

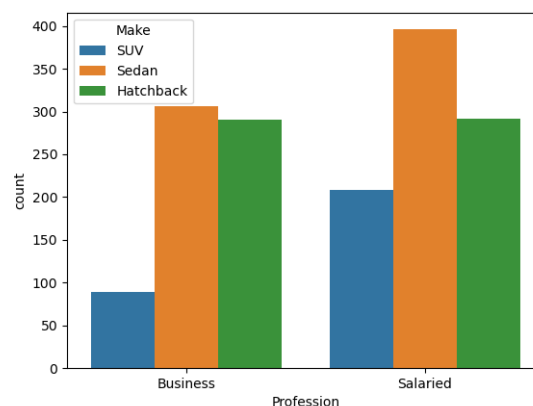


Figure 24: Barchart of 'Profession' column

Observations

1. **Business owners** favor **Sedans** and **Hatchbacks** equally, with less preference for **SUVs**.
2. **Salaried professionals** prefer **Sedans**, followed by **Hatchbacks**, with moderate interest in **SUVs**.

Salaried professionals have more interest in products of Austo Motor Company than business professionals.

Insights

1. **Profession impacts car choice:** **Business** prefers **Sedans/Hatchbacks**, while **Salaried** professionals lean toward **Sedans**.
2. **Marketing Strategy:** Focus **SUV** ads on **Salaried** group and **Sedan/Hatchback** ads for both groups.

* **Marital_status vs Make**

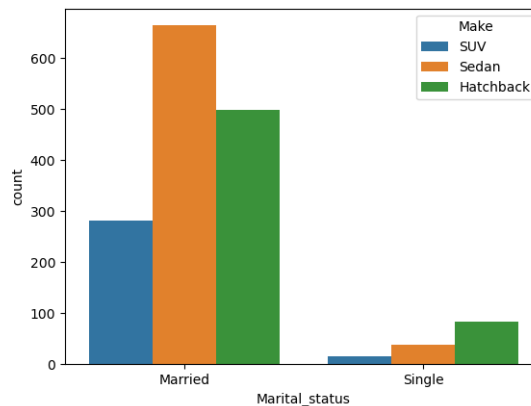


Figure 25: Barchart of 'Marital_status' column

Observations

1. **Married individuals** prefer **Sedans** most, followed by **Hatchbacks** and **SUVs**.
2. **Single individuals** show little interest in **Sedans** and **SUVs**, preferring **Hatchbacks** slightly more.

Insights

1. **Marital status affects car preference:** **Married** people prefer **Sedans/Hatchbacks**, while **Singles** favor **Hatchbacks**.
2. **Marketing Strategy:** Focus **Sedan/Hatchback** ads on **Married** individuals; tailor **Hatchback** campaigns for **Singles**.

* **Education vs Make**

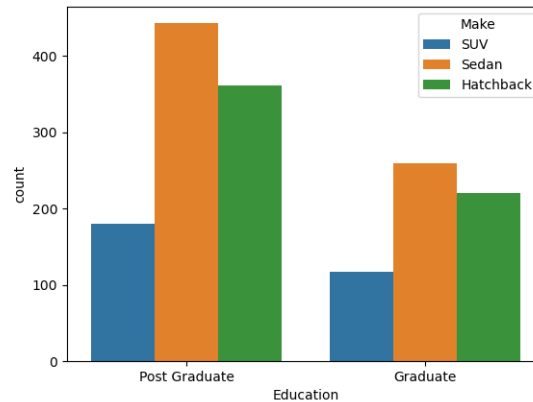


Figure 26: Barchart of 'Price' column

Observations and Insights

- * **SUVs** are the least popular among both graduates and postgraduates.
- * **Sedans** are the most popular among postgraduates, followed by hatchbacks.
- * **Hatchbacks** are the second most preferred car among both education levels, with higher preference among postgraduates.
- * Graduates generally prefer sedans more than hatchbacks and SUVs.

In the data we have, the number of **postgraduates** is more as compared to graduates. Given the higher count of postgraduates, the marketing team could tailor campaigns to appeal to this educated segment.

Impact on Business

- * The marketing strategy should focus on promoting **SUVs**, as they are less popular.
- * For graduates, emphasize **sedans** and **hatchbacks**. For postgraduates, sedans can drive higher engagement.
- * **Personal_loan vs Make**

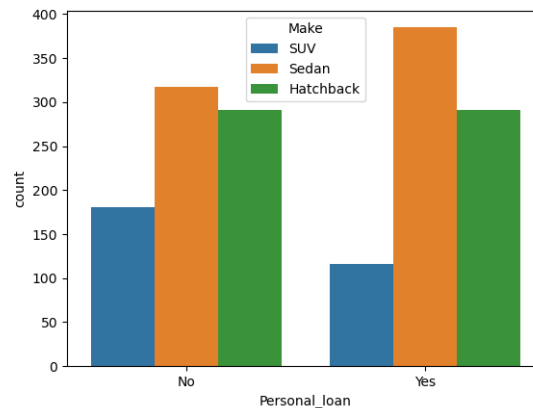


Figure 27: Barchart of 'Personal_loan' column

Observations and Insights

- * **Sedans** are the most popular among customers with personal loans, followed by **hatchbacks**.
- * **SUVs** are the least preferred for both personal loan and non-loan customers.
- * Among non-loan customers, **sedans** and **hatchbacks** show similar demand, both higher than SUVs.

Impact on Business

- * Increase **SUV** promotions across both customer segments.
- * Focus on **sedan** promotion for customers with personal loans.
- * Continue a balanced **hatchback** promotion strategy, emphasizing loans for increased engagement.
- * **House_loan vs Make**

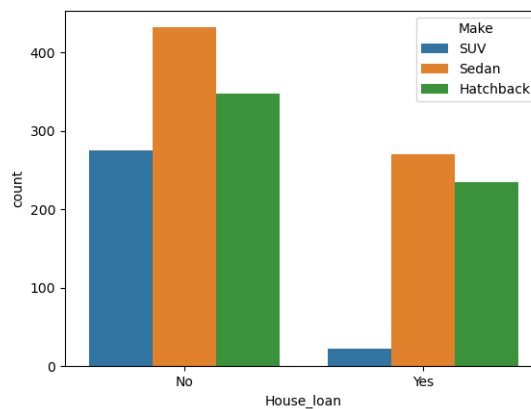


Figure 28: Barchart of 'House_loan' column

Observations and Insights

- * **Sedans** are the most popular among customers without house loans, followed by **hatchbacks**.
- * **SUVs** are the least preferred among both house loan and non-house loan customers.
- * Customers with house loans show lower overall demand for cars, with **sedans** and **hatchbacks** preferred.

The dataset shows that the number of customers with house loans is significantly lower.

Impact on Business

- * Increase **SUV** promotions across both customer segments, especially for house loan customers.
- * Focus on **sedans** and **hatchbacks** for non-house loan customers.
- * Target house loan customers with a focused approach to boost **sedan** and **hatchback** sales.
- * **Partner_working vs Make**

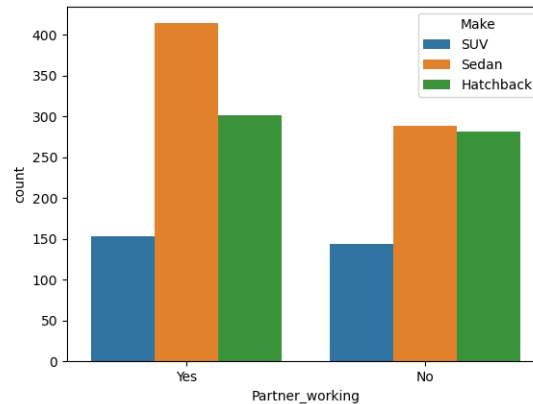


Figure 29: Barchart of 'Partner_working' column

Observations:

- * The chart displays preferences for **SUVs**, **Sedans**, and **Hatchbacks** based on the partner's working status.
- * **Sedans** are highly favored by customers with working partners, followed by **Hatchbacks**.
- * Customers without working partners show almost equal preference for **Sedans** and **Hatchbacks**.
- * **SUVs** are the least popular in both cases.

Insights:

- * **Sedans** are in high demand among customers with working partners, possibly for their comfort and space.
- * Low **SUV** demand suggests a need to reconsider its promotion in marketing strategies.
- * **Hatchbacks** are consistently popular across both groups, indicating stable demand.
- * **Make vs Price**

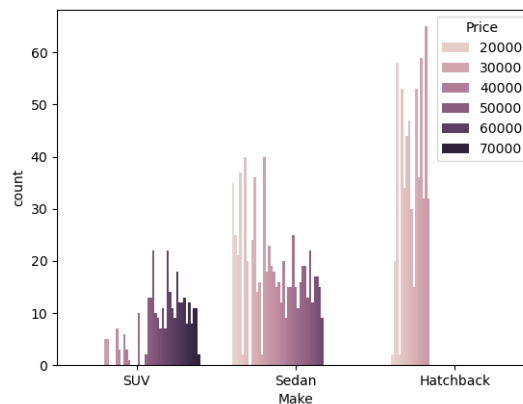


Figure 30: Barchart of 'Make' column

Observations:

- * The chart shows the distribution of car types (**SUV**, **Sedan**, **Hatchback**) based on **Price**.
- * **SUVs** are primarily priced between 20,000 and 70,000, with many high-priced models.
- * **Sedans** have a broader price range from 20,000 to 60,000.
- * **Hatchbacks** are mainly concentrated in the 20,000 to 40,000 price range.

Insights:

- * **SUVs** are concentrated in the mid-price range, implying limited demand for high-end models.
- * **Sedans** appeal to a wide customer base across various price ranges.
- * **Hatchbacks** are popular in the lower price segment, appealing to budget-conscious buyers.

4.3 Key questions to ponder

4.3.1 Do men tend to prefer SUVs more compared to women?

No Women prefer SUVs more as compared to men.

4.3.2 What is the likelihood of a salaried person buying a Sedan?

Likelihood of a salaried person buying a Sedan is **44.196%**.

4.3.3 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?

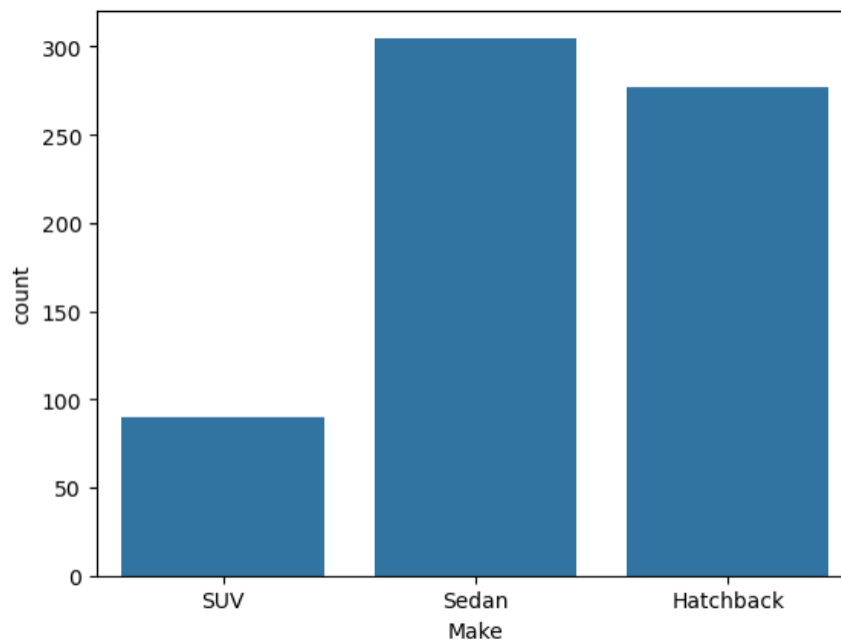
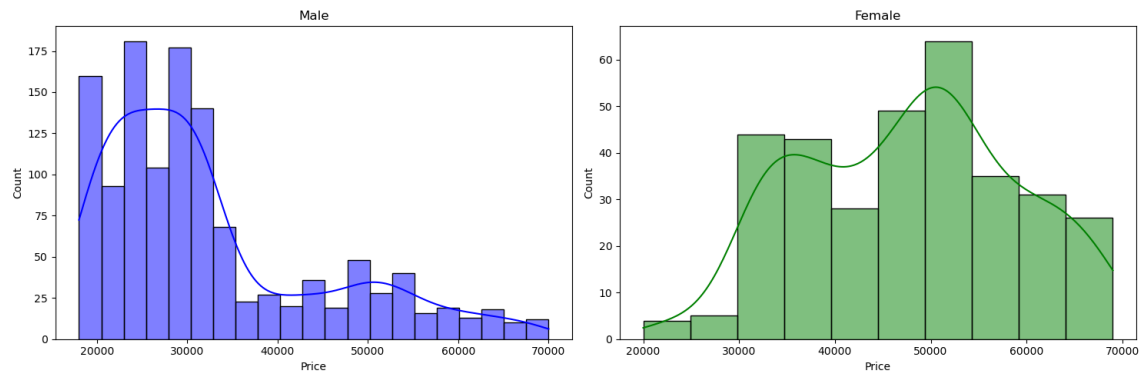


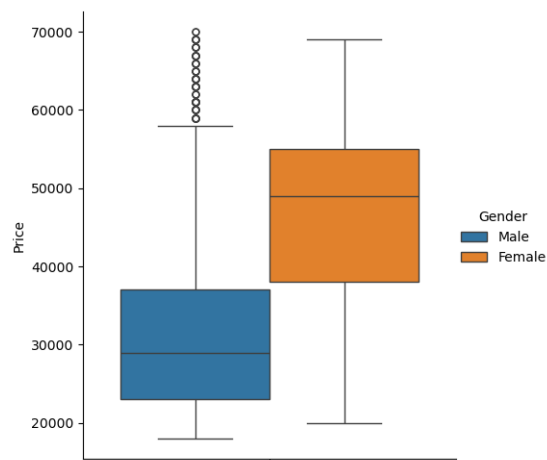
Figure 31: Barchart of 'Make' column for salaried male

According to the data, salaried males show a stronger preference for purchasing sedans over SUVs. Sheldon Cooper's assertion that salaried males are more susceptible to SUV sales doesn't hold true based on this evidence.

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



(a) Countplot



(b) Boxplot

Figure 32: 'Gender' vs 'Price'

Observation and Insights

Male vs Female Spending Patterns:

- * **Males** predominantly purchase vehicles priced between 20,000 and 40,000 units, with a noticeable decline in higher price ranges.
- * **Females** tend to spend more, as most purchases are concentrated between 40,000 and 60,000 units.

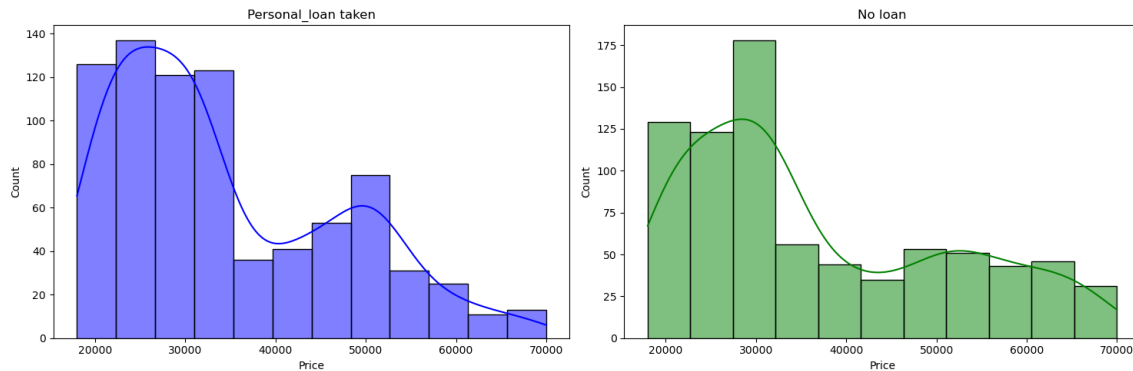
Gender-wise Spending Distribution:

- * The boxplot shows that **females** typically spend more on vehicles, with a higher median and a broader range of prices.
- * **Males** exhibit lower spending overall, with several outliers on the higher side, indicating a few instances of luxury purchases.

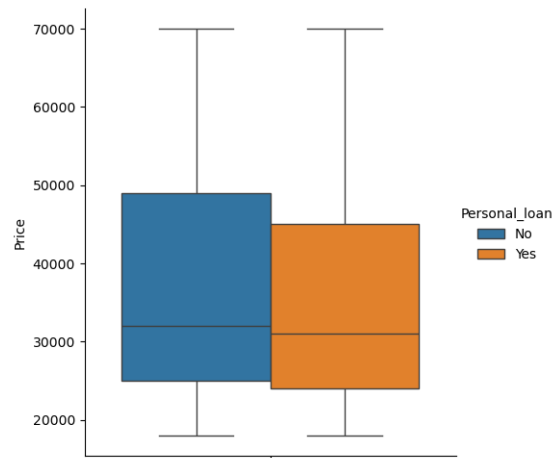
Business Insight:

- * **Males** are price-conscious and purchase lower-cost models more frequently.
- * **Females** are willing to invest in higher-priced models, suggesting the potential for marketing premium models to the female demographic.

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



(a) Countplot



(b) Boxplot

Figure 33: 'Personal_loan' vs 'Price'

Observation and Insights

Individuals with Personal Loans:

- * Most purchases fall between **20,000 and 36,000** units, with fewer purchases in higher price brackets.
- * The distribution skews towards lower-priced vehicles, suggesting loan-takers buy less expensive models.

Individuals without Personal Loans:

- * Purchases are concentrated between **20,000 and 35,000** units.
- * This group shows a broader range of spending, indicating flexibility in purchasing higher-priced models.

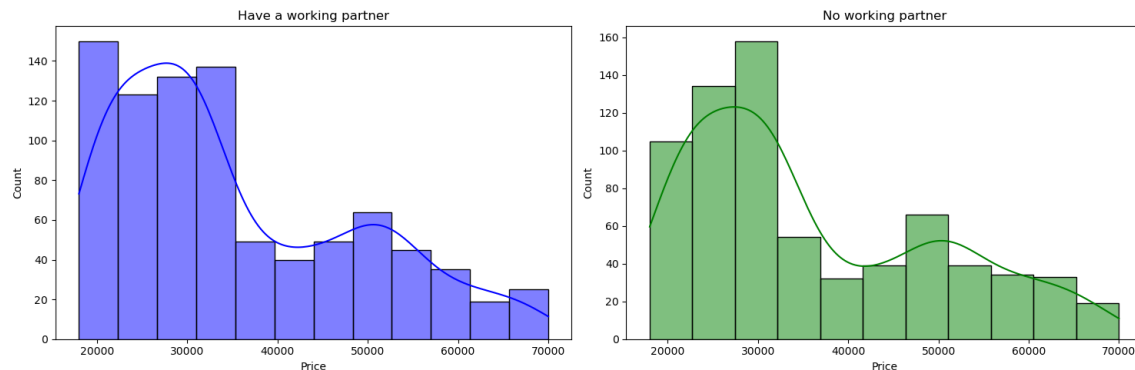
Boxplot Comparison:

- * The **median spending** of individuals without loans is **higher** than that of loan-takers.
- * Loan-takers exhibit a more constrained spending range, while non-loan takers show a wider variation, purchasing higher-end models more frequently.

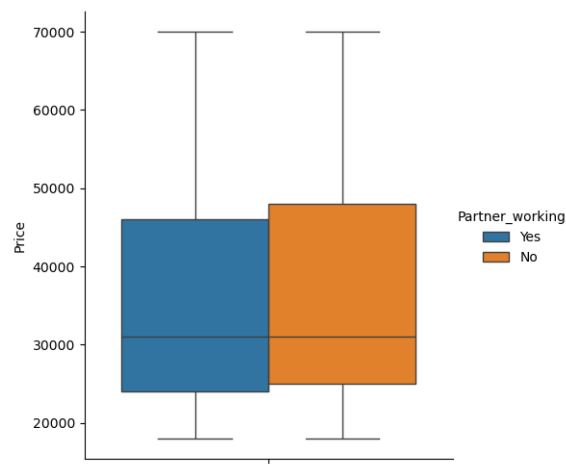
Business Insight:

- * Non-loan customers tend to spend more, suggesting premium models can be marketed to them.
- * Loan-takers prefer budget vehicles, highlighting the importance of financing options for lower-cost models.

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



(a) Countplot



(b) Boxplot

Figure 34: 'Partner_working' vs 'Price'

Observations

1. Histograms:

- * For customers **with working partners**, demand is more evenly distributed across price ranges, peaking around 30,000 Rs..
- * For customers **without working partners**, demand also peaks around 30,000 Rs., but declines sharply after 50,000 Rs..

2. Boxplot:

- * Median car prices are similar for both groups.
- * **Customers without working partners** exhibit slightly more price variability.

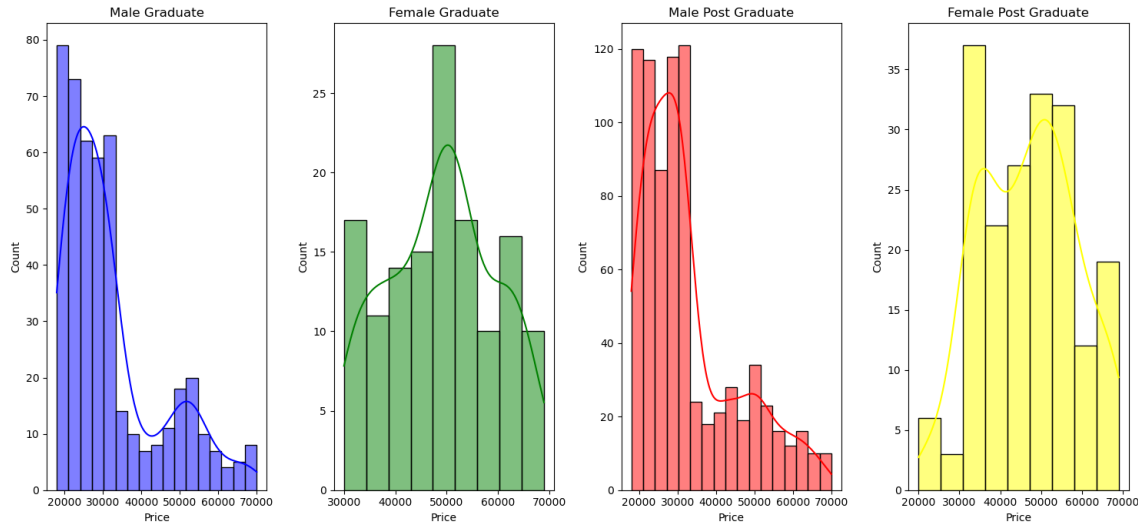
Insights

- * The price range around 30,000 Rs. has the highest demand across all customers.

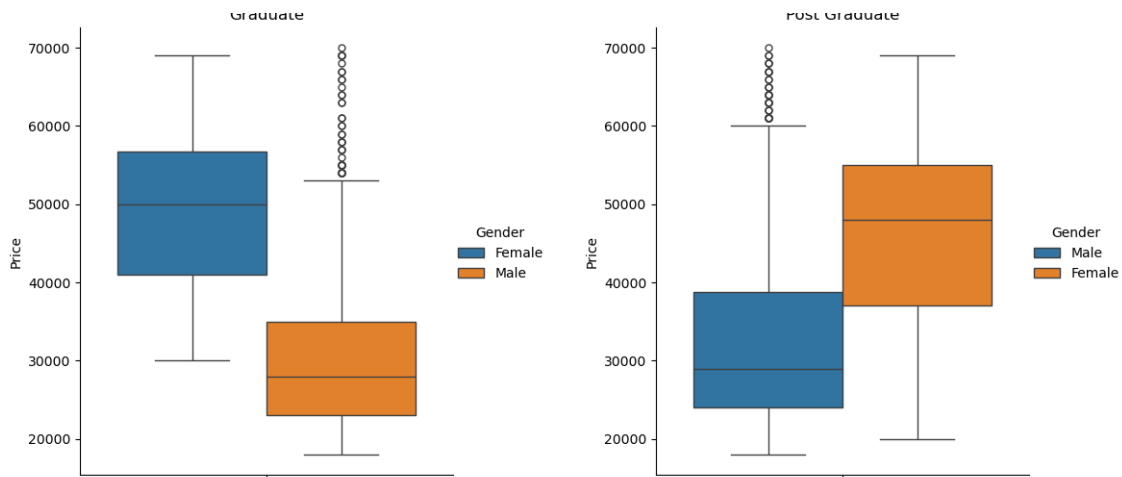
- * **Working partners** may be linked to a broader range of affordability.
- * **Marketing campaigns** should emphasize the **30,000 Rs.** price point, with tailored strategies for higher-end models aimed at customers with working partners.

4.4 Some extra interesting questions to ponder

4.4.1 How does the level of Education impact price of cars purchased?



(a) Countplot



(b) Boxplot

Figure 35: 'Education' vs 'Price'

Observations

1. Histograms:

- * **Male Graduates** show the highest car purchase demand around **Rs.30,000**, with a sharp decline afterward.
- * **Female Graduates** show the highest demand around **Rs.50,000**.
- * **Male Postgraduates** have concentrated demand around **Rs.30,000**, similar to male graduates.

- * **Female Postgraduates** show a broader demand distribution, peaking around **Rs.45,000** and **Rs.55,000**.

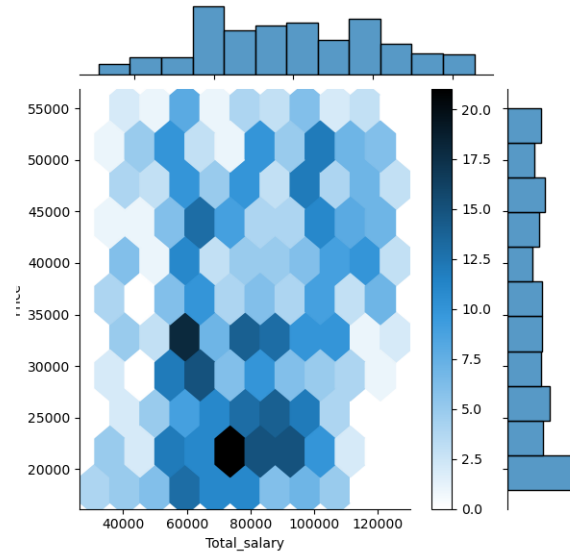
2. Boxplots:

- * **Postgraduates** (both genders) show higher variability in car prices compared to graduates.
- * The median price for **Male Graduates** is significantly lower than for **Female Graduates**.
- * **Male Postgraduates** have lower median prices compared to **Female Postgraduates**.

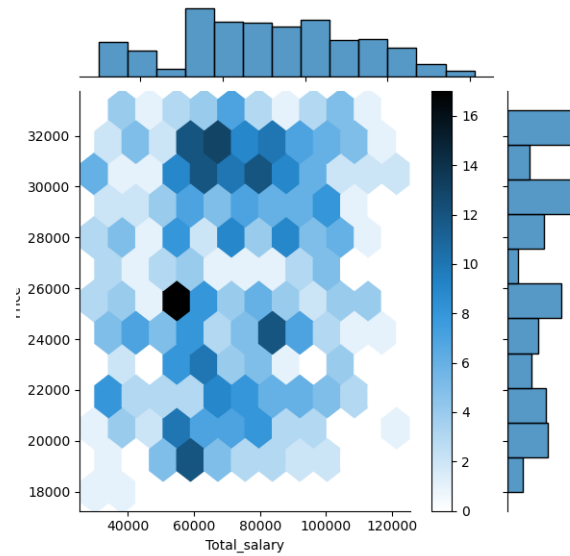
Insights

- * **Graduates** (especially males) prefer cars around **Rs.30,000**, while **Female Graduates** opt for higher-priced cars.
- * **Postgraduates** display more variability in spending, especially **Female Postgraduates**, with higher median prices.
- * **Marketing campaigns** should emphasize affordability around **Rs.30,000** for male customers, while highlighting mid-range options for females and postgraduates.

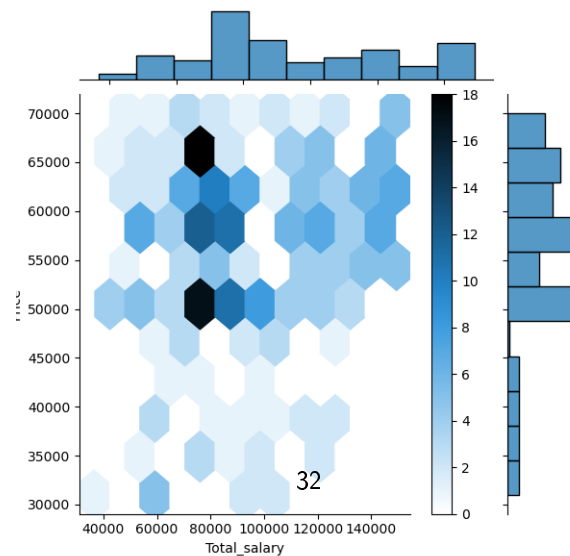
4.4.2 How does the total salary of customers influence the price range of different car models (SUV, Sedan, Hatchback)?



(a) Make - 'Sedan'



(b) Make - 'Hatchback'



(c) Make - 'SUV'

Observations and Insights

1. SUV:

- * Most buyers have **salaries Rs. 60k-80k**, purchasing at **Rs. 48k-53k**.
- * A smaller segment consists of buyers with **salaries > Rs. 100k**, buying at **Rs. 60k-70k**.

2. Sedan:

- * Clustered buyers have **salaries Rs. 60k-80k**, purchasing at **Rs. 20k-25k**.

3. Hatchback:

- * Buyers mainly have **salaries Rs. 55k-70k**, purchasing at **Rs. 24k-26k** and **Rs. 30k-32k**.

Insights for Business

- * **SUV**: Target **mid-to-high salary** customers.
- * **Sedan**: Focus on **mid-range** pricing.
- * **Hatchback**: Market to **budget-conscious** buyers.