Business Report

Statistical Methods for Decisions Making

By Gagandeep Singh

Table of Contents

Problem 1

Context

Analysts are required to explore data and reflect on the insights. Clear writing skill is an integral part of a good report. Note that the explanations must be such that readers with minimum knowledge of analytics are able to grasp the insight.

Austo Motor Company is a leading car manufacturer specializing in SUV, Sedan, and Hatchback models. In its recent board meeting, concerns were raised by the members on the efficiency of the marketing campaign currently being used. The board decides to rope in an analytics professional to improve the existing campaign.

Objective

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.

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

Data Overview

Structure of the Data:

- Number of Rows: 1581
- Number of Columns: 14
- Memory Usage: 173.1+ KB
- Range Index: 0 to 1580
- Data Types: Float, Int, and Object

Data Type:

The different datatypes in the dataset are as follows
a.There are 5 columns in the with int64 data type
b.There are 8 columns in the with object data type
c.There are 1 columns in the with float64 data type

```
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
```

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
```

Gender Column:

- Missing Values: 53

Partner Salary Column:

- Missing Values: 106

Missing values Treatment:

- For the "Gender" column, a standard method to address missing categorical values involves imputing them with the mode of the data, representing the most frequently occurring category.
- Regarding the "Partner Salary" column, it appears to be related to another column named "Salary," and the "Total Salary" is derived by summing the values of both "Partner Salary" and "Salary." To fill in the missing values in the "Partner Salary" column, they have been computed using the formula: Total Salary - Salary.

```
Age                  0
Gender               0
Profession           0
Marital_status       0
Education            0
No_of_Dependents     0
Personal_loan        0
House_loan           0
Partner_working      0
Salary               0
Partner_salary       0
Total_salary         0
Price                0
Make                 0
dtype: int64
```

All the missing values have been addressed.

Statistical Summary:

	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	1581.0	19233.776091	19670.391171	0.0	0.0	25100.0	38100.0	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

- The "Total_salary" column displays a positively skewed distribution, evident from the mean being lower than the median.
- The "Price" column demonstrates a positively skewed distribution, as the mean exceeds the median.
- Both "No of Dependents" and "Partner Salary" share the same minimum value, while "Total Salary" and "Salary" also have identical minimum values.
- Age spans from a minimum of 22 to a maximum of 54.
- Prices range from a minimum of 18000 to a maximum of 70000.

Data Irregularities and Treatment:

```
The unique values in Gender columns -
 ['Male' 'Femal' 'Female' 'Femle']
```

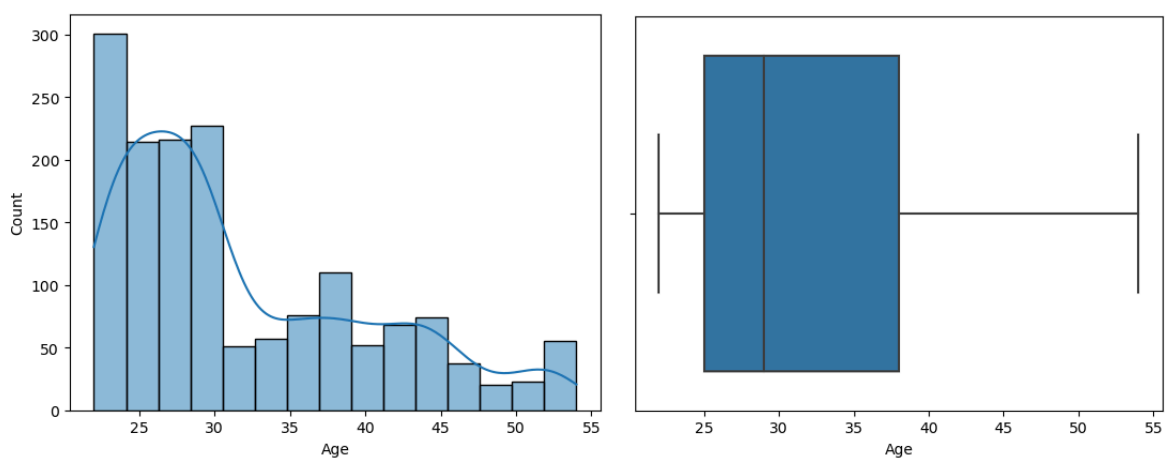
- In the "Gender" column, it was observed that there are four unique values: 'Male,' 'Femal,' 'Female,' and 'Femle.'
- To address this inconsistency, 'Femal' and 'Femle' were standardized to 'Female' for uniformity and accuracy in gender representation.

Univariate Analysis

Explore all the features of the data separately by using appropriate visualizations and draw insights that can be utilized by the business.

Analysis:

Age



Skewness:

1. The age distribution is positively skewed, indicating that there are relatively more individuals with younger ages, and the tail of the distribution extends towards higher ages.

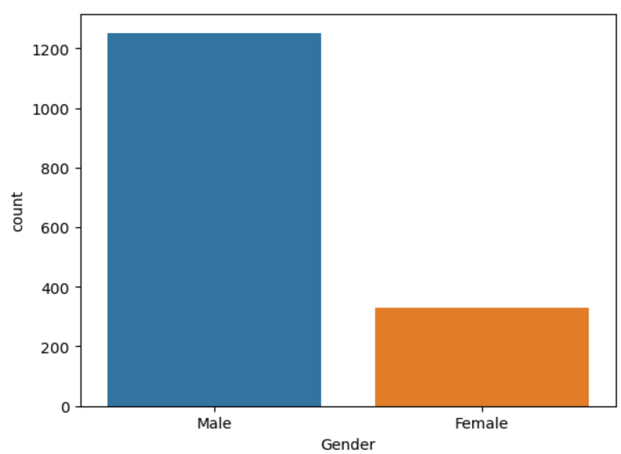
Median Age:

1. The median age is 29, suggesting that it is the middle point of the age distribution. In a positively skewed distribution, the median is typically less than the mean.

Preference by Age:

1. The histogram reveals a trend where individuals in the younger age group, specifically between 22 and 30, show a higher preference for buying cars.
2. There is a noticeable decrease in car buying preference after the age of 30.

Gender



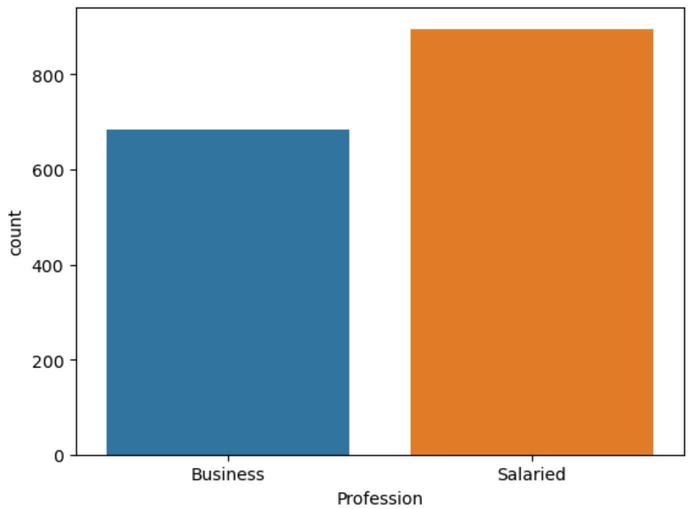
Gender Distribution:

The dataset comprises a greater proportion of males, constituting 79% of the entire dataset, while the representation of females is comparatively smaller.

Buying Preferences:

The observation suggests that, based on the available data, there is a trend indicating that males show a higher preference for buying cars compared to females.

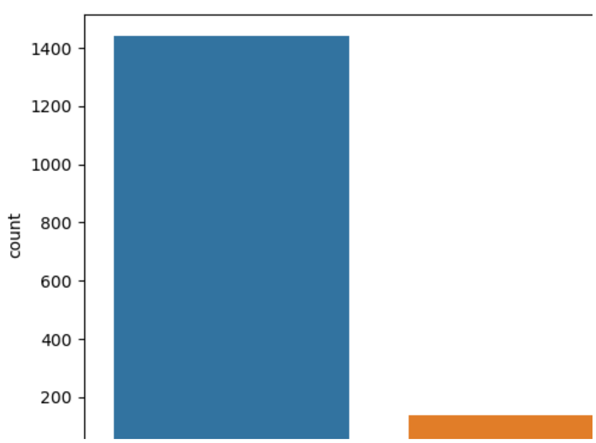
Profession



Preference by Profession:

The observation suggests that individuals categorized as business professionals demonstrate a lower preference for purchasing cars compared to those classified as salaried employees.

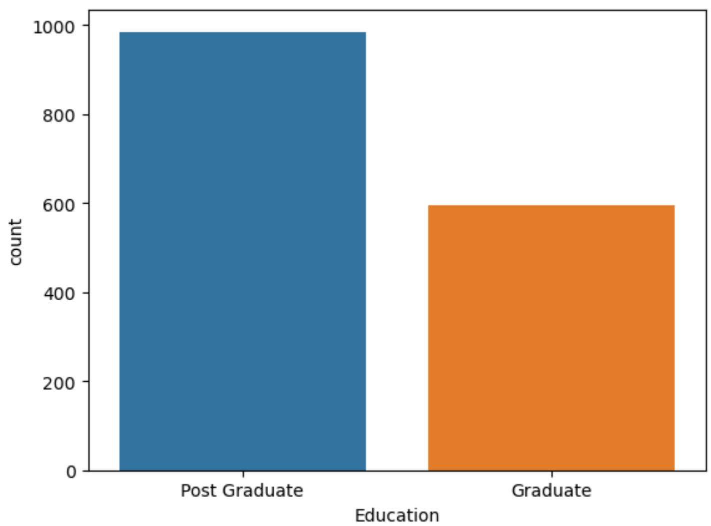
Marital_status



Ratio of Single to Married Individuals:

1. The dataset reveals a notably higher ratio of married individuals in comparison to single individuals.
2. More specifically, 92% of the dataset consists of married individuals, indicating a strong representation of this demographic in the context of car purchasing.

Education



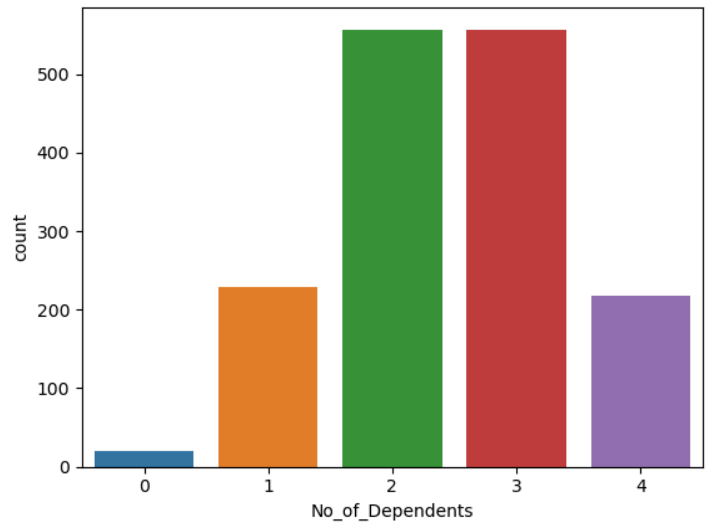
Educational Distribution:

The dataset is dominated by postgraduate individuals, constituting 62% of the data, in comparison to those with a graduate education.

Buying Trends by Education:

The observation suggests a trend where individuals with postgraduate degrees have a higher representation in the dataset, indicating a slightly higher likelihood of postgraduates buying cars.

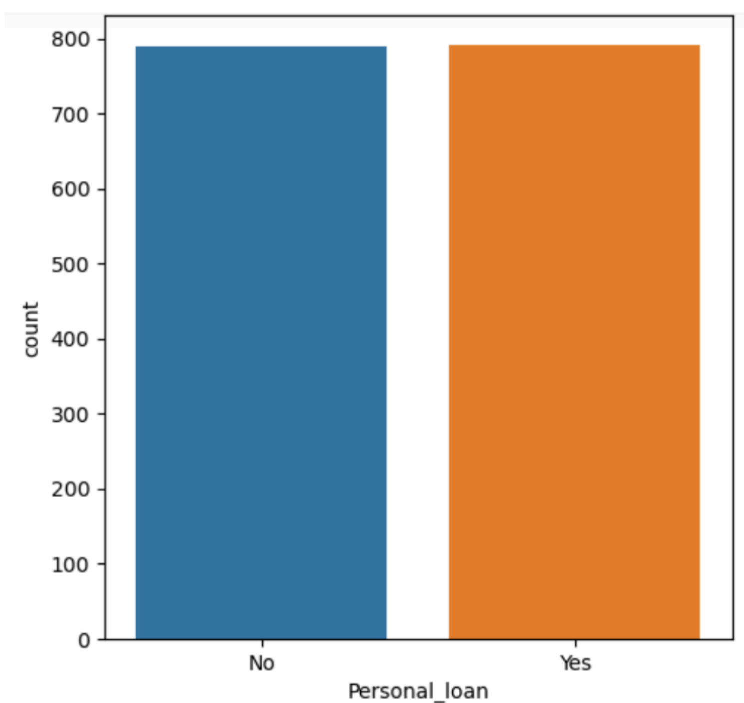
No_of_Dependents



Buying Preferences by Number of Dependents:

1. Individuals with 2 or 3 dependents show a higher preference for buying a car compared to other dependency values. Specifically, this group constitutes 70% of the entire dataset.
2. Outliers are present in the data with 0 as the number of dependents.

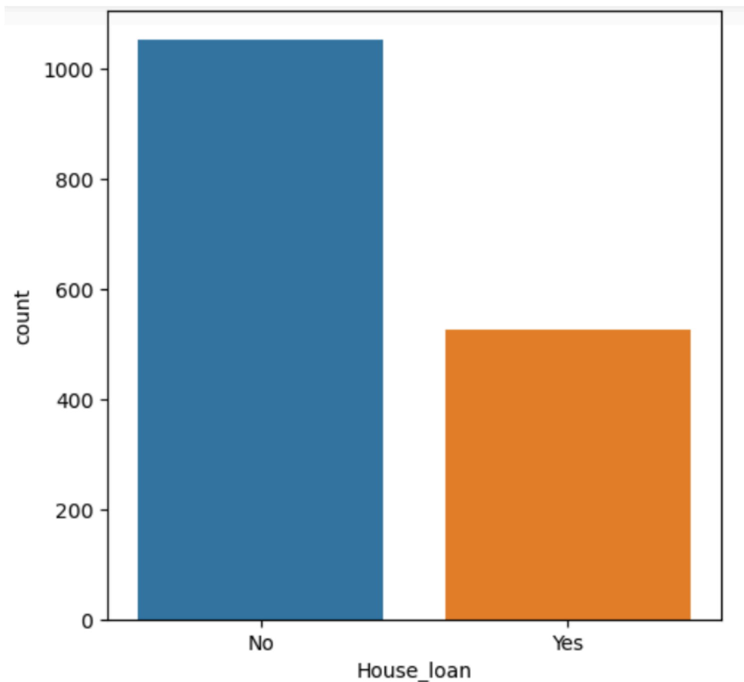
Personal_loan



Loan Status and Car Buying:

1. Individuals with personal loans and those without personal loans exhibit roughly equal ratios in the dataset.
2. This suggests that both groups show a similar preference for buying cars.

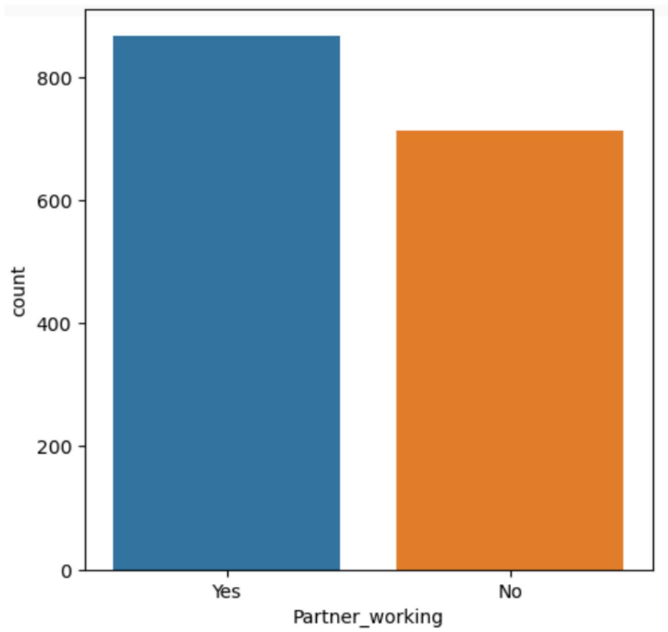
House_loan



Car Buying Preference and Home Loans:

1. The dataset reveals a ratio where individuals without house loans prefer to buy a car in contrast to those with ongoing home loans.
2. Specifically, the ratio is observed as 33 to 66, indicating a higher preference for car buying among individuals without home loans.

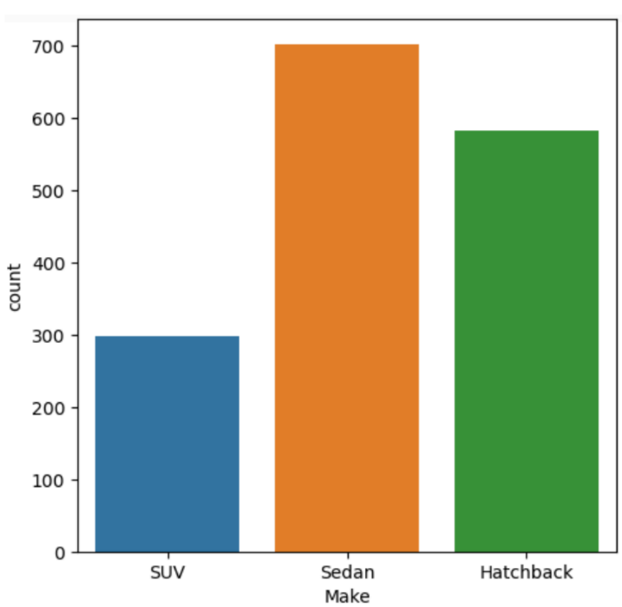
Partner_working



Partner's Employment Status and Car Buying:

1. The dataset indicates that there is not a significant difference in the number of individuals who prefer to buy a car based on whether their partner is working or not.
2. However, there is a slightly higher likelihood of individuals having a car if their partner is employed.

Make



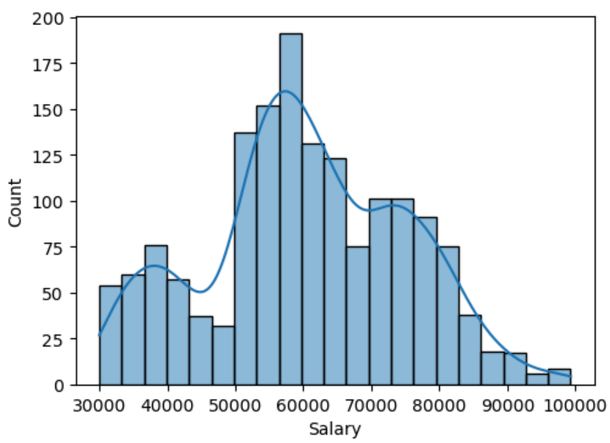
Most Preferred Car Type:

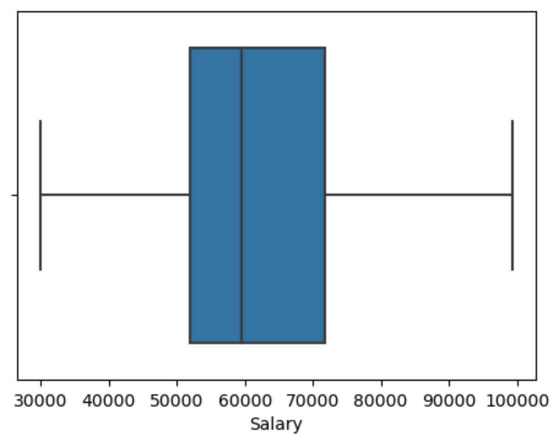
Sedan emerges as the most favored car type among all individuals in the dataset, constituting 44% of the preferences.

Least Preferred Car Type:

SUV ranks as the least preferred car type, representing only 18% of the preferences.

Salary

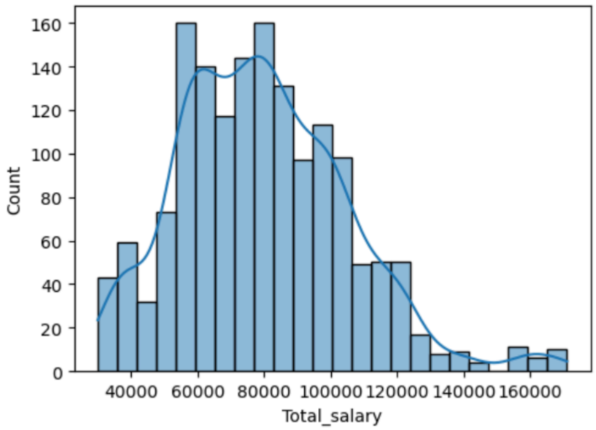


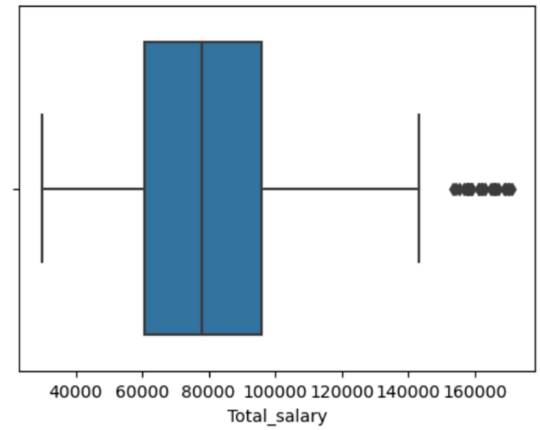


Salary Analysis -

1. The dataset indicates a higher number of individuals with salaries in the range of 50000 to 80000.
2. Conversely, there is a notably lower number of individuals with salaries exceeding 90000.

Total _Salary

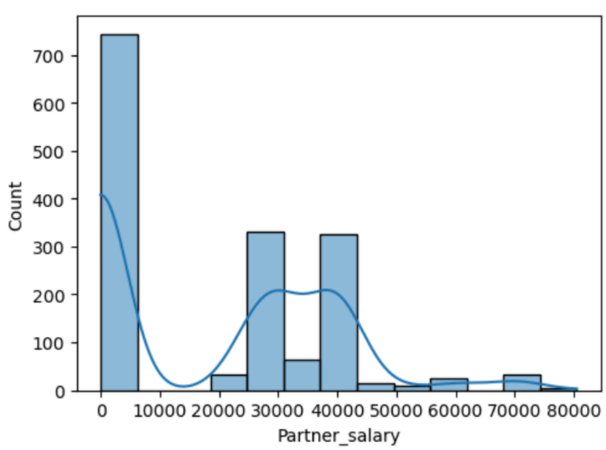


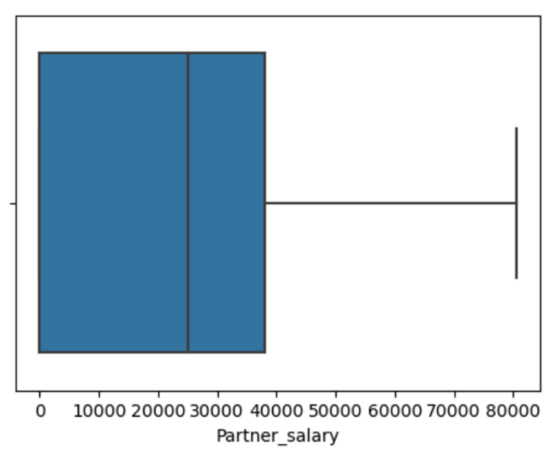


Total Salary Analysis:

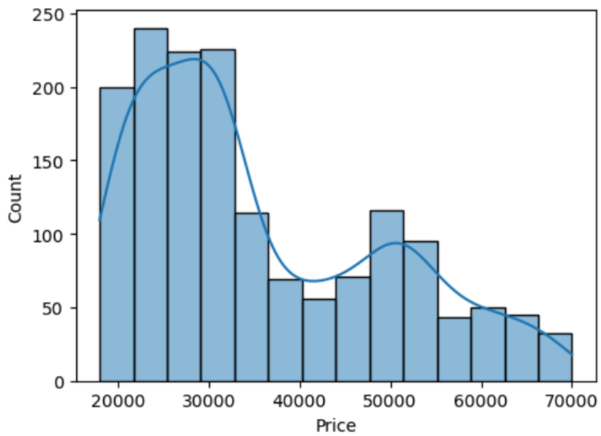
1. The distribution of total salaries is right-skewed, indicating a higher frequency of lower total salary values and a gradual decrease towards higher values.
2. An outlier has been identified in the dataset, specifically in the total salary column. This outlier reflects a scenario where the combined salary of high-income individuals and their partners contributes to an exceptionally high total salary value.
3. Notably, there is a sudden dip in the distribution after the 100000 mark.
4. This dip suggests a decrease in the number of individuals with total salaries exceeding 100000, signifying a less frequent occurrence of higher total salary values.

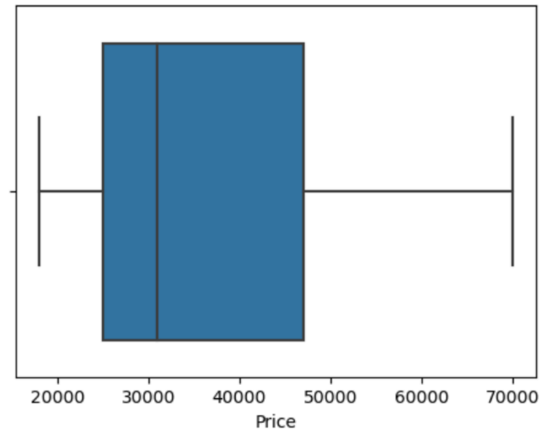
Partner_salary





Price

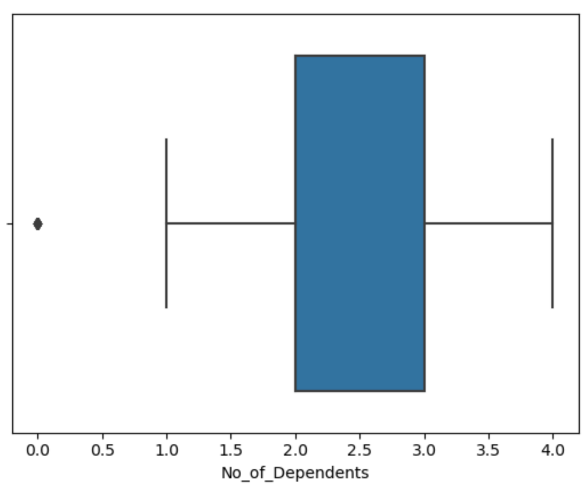


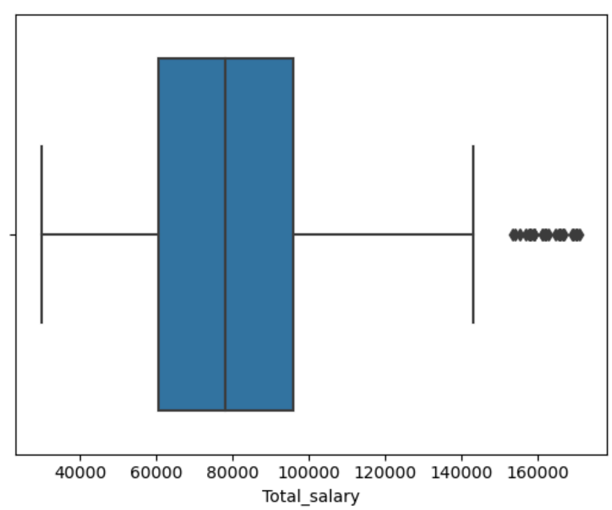


Price Analysis:

1. The distribution of car prices is right-skewed, suggesting a higher occurrence of lower-priced cars and a gradual decline in frequency as prices increase.
2. The observation indicates a higher preference for cars within the price range of 18000 to approximately 32000.

<u>Outliers and Treatment:</u>





There are two columns with Outliers -

1. No_of_Dependents
2. Total_salary

Outlier Treatment Summary:

Number of Dependents:

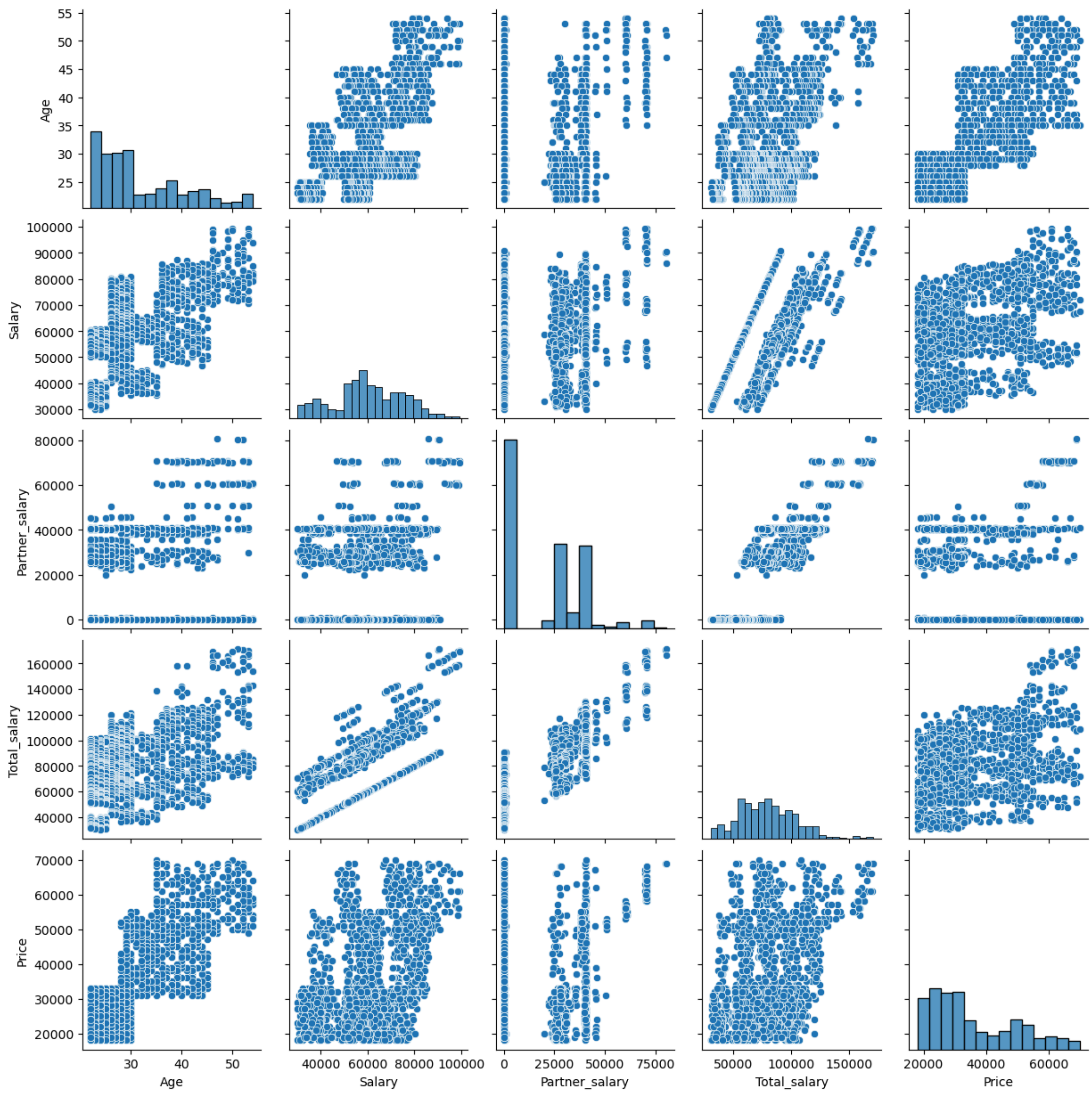
The "Number of Dependents" column has undergone a transformation in its data type to category. This change reflects the understanding that the number of dependents is more appropriately represented as a categorical variable, considering its nature as a whole number.

Total Salary:

Outliers in the "Total Salary" column have not been treated. The decision not to treat these outliers is based on the consideration that adjusting them could impact the "Partner Salary" and "Salary" columns, As "Total Salary" is derived from their sum

Bivariate Analysis

Relationship and correlation between all numerical variables-


Age
Salary
Partner_salary
Total_salary
Price

Insights-

Age and Price:

- There is a significant positive correlation between age and car price.
- As age increases, there is a preference for higher-priced cars.

Salary and Price:

- The correlation analysis indicates that the relationship between salary and car price is relatively weak.
- An increase in salary does not exhibit a strong correlation with the inclination to buy more expensive cars.

Age and Salary:

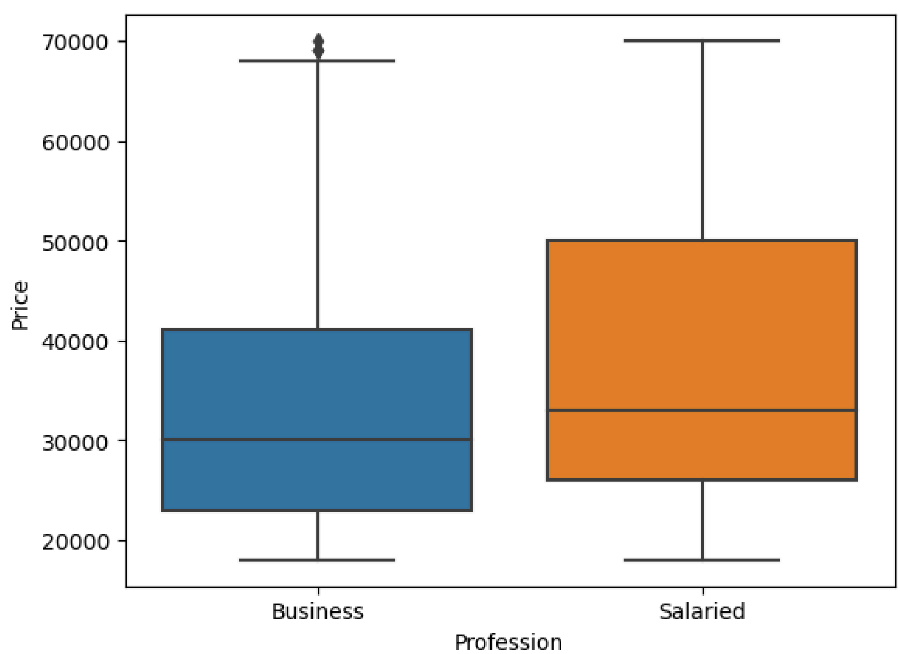
- Age and salary have a strong positive correlation.
- As age increases, salary tends to show an upward trend.

Number of Dependents and Price:

- There is no significant correlation between the number of dependents and car price.
- The presence or absence of dependents does not strongly influence the price preferences for cars.

Relationship between categorical vs numerical variables

Profession and Price



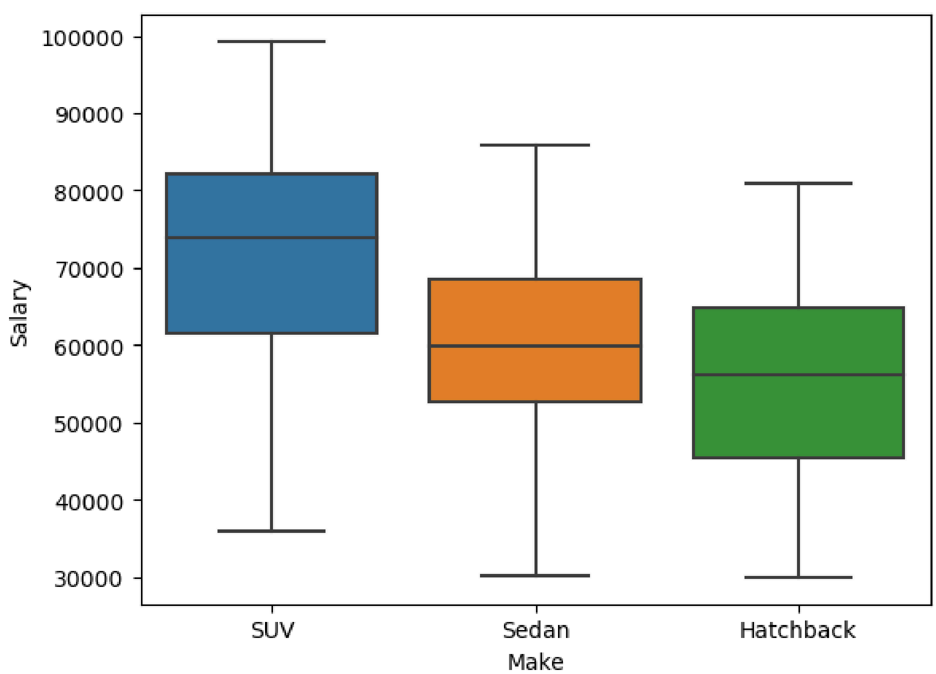
Profession and Price -

Median Price for Salaried Individuals:
The median price of cars preferred by salaried individuals is higher compared to that of business individuals.

Salaried Individual Preferences:
This suggests that, on average, salaried individuals show a greater preference for higher-priced cars.

Outliers in Business Individuals:
Outliers among business individuals indicate that there are a few individuals in the business category who exhibit a preference for higher-priced cars.

Salary and Make



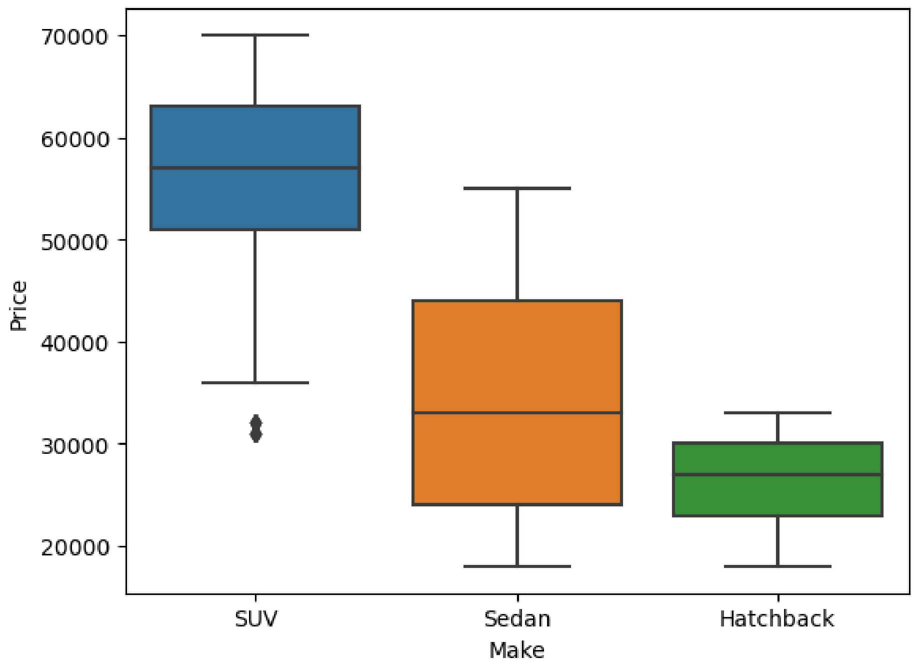
Median Value Comparison:

The median value of SUVs, in relation to individual salaries, is higher compared to hatchbacks and sedans. Conversely, the median value of hatchbacks is the lowest among the three car types in relation to individual salaries.

Salary Preferences:

Individuals with high salaries exhibit a stronger preference for SUVs over sedans and hatchbacks.

Price and Make



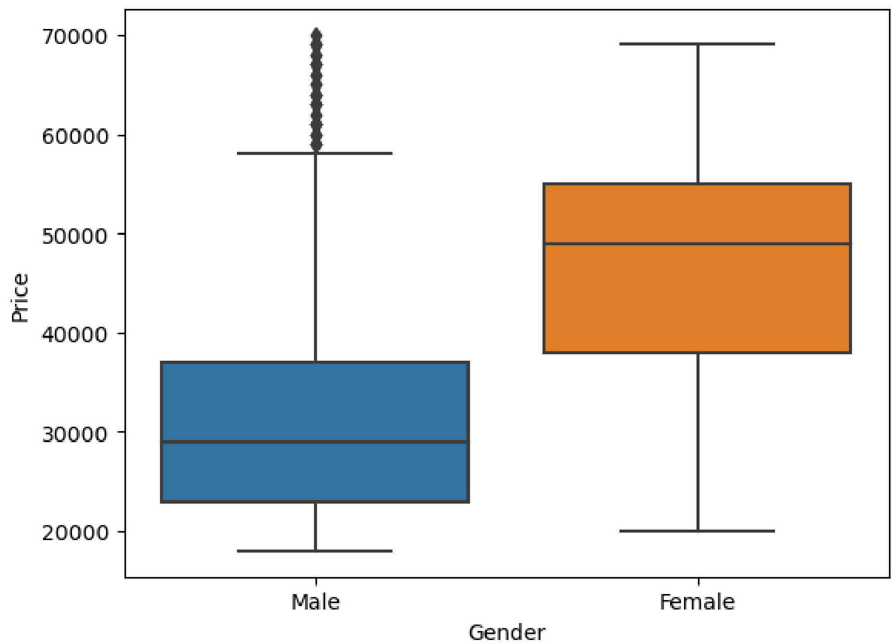
Median Price Comparison:

SUVs have the highest median price compared to sedans and hatchbacks.

Relative Expensiveness:

This observation suggests that, on average, SUVs are more expensive compared to sedans and hatchbacks.

Price and Gender

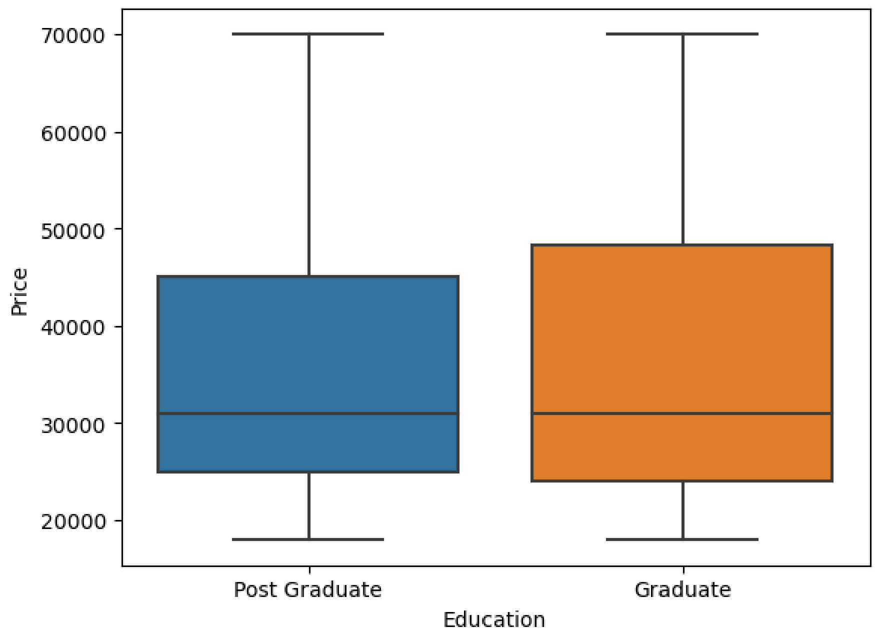


Median Comparison:
Female individuals have a higher median car price compared to males, suggesting a preference for higher-priced cars among females.

Female Car Preferences:
The observation indicates that, on average, more females prefer to buy higher-priced cars.

Male Outliers:
Outliers in the male category suggest that there are some males who exhibit a preference for high-priced cars, although this is not as common as in females.

Education and Price



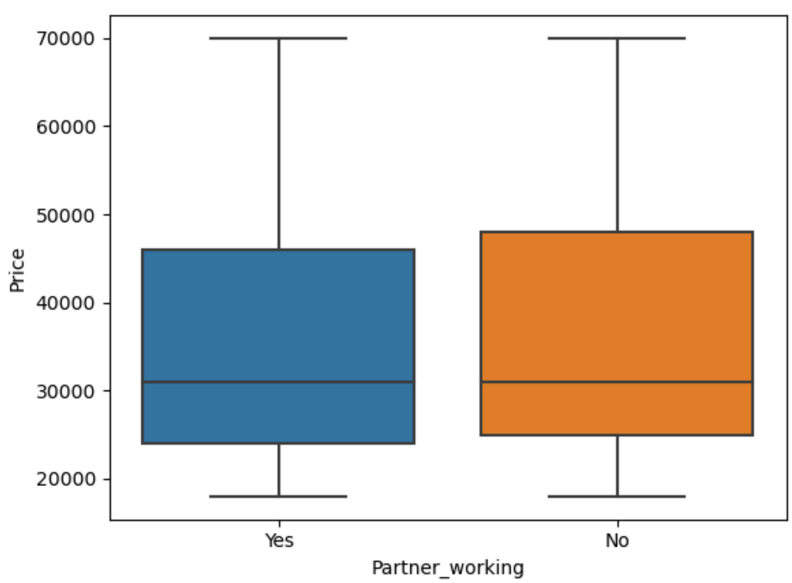
Median Comparison:

The median car prices are approximately the same across different education levels.

Education and Car Price Relationship:

This suggests that there is not a significant impact on the choice of car prices based on the education level of the individual.

Partner_working and Price



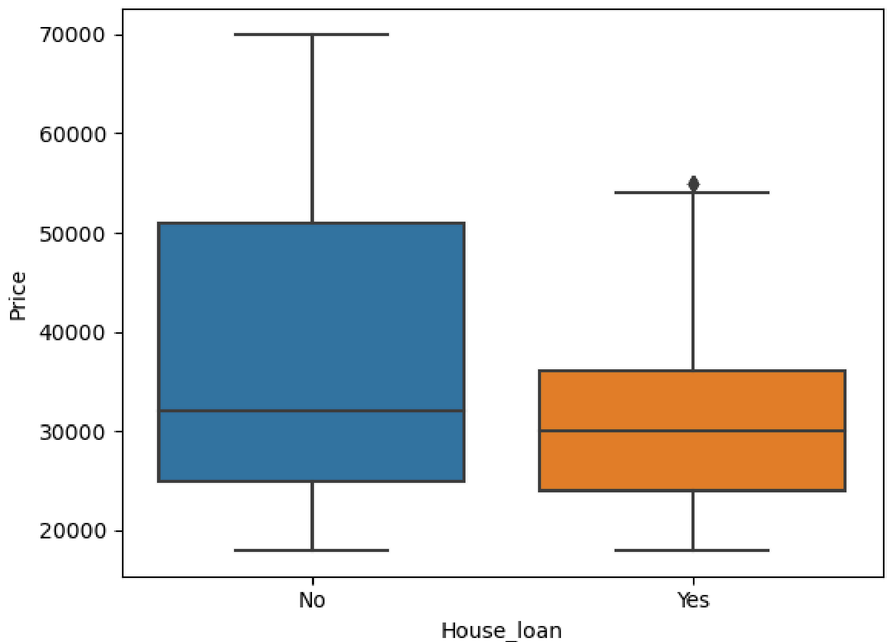
Median Value Comparison:

The median values for car prices are approximately equal for individuals whether their partner is working or not.

Partner's Employment and Car Price Relationship:

This indicates that there is not a significant impact on the preference for car prices based on whether the individual's partner is working or not.

House_loan and Price



Median Value Comparison:

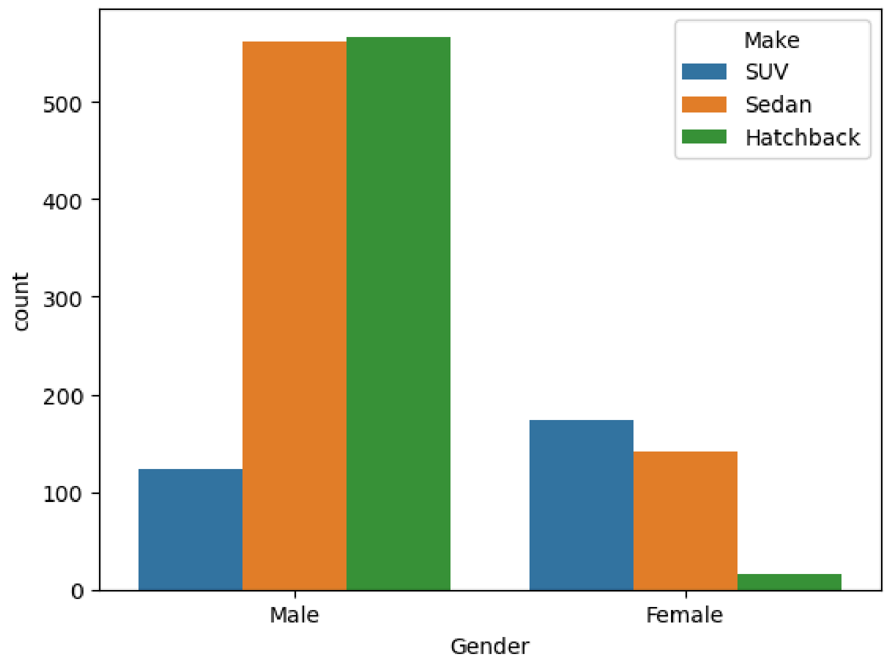
The median value for individuals without a house loan is slightly higher, indicating a potentially higher likelihood for them to buy a higher-priced car compared to individuals with a house loan.

Home Loan and Car Price Relationship:

This suggests that there may be a correlation between not having a house loan and a higher preference for higher-priced cars.

Key Questions

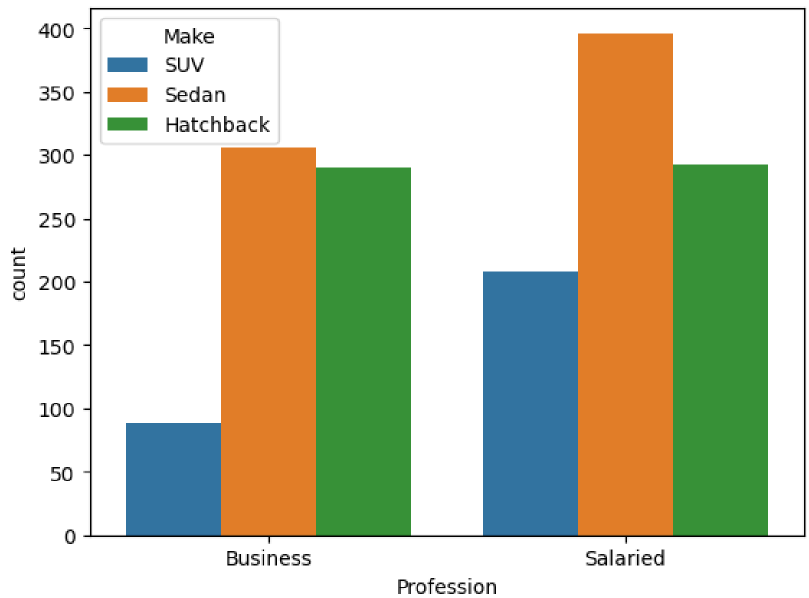
Q1. Do men tend to prefer SUVs more compared to women?



No,Females tend to prefer SUVs more compared to males, based on a higher count.

Q2. What is the likelihood of a salaried person buying a Sedan?

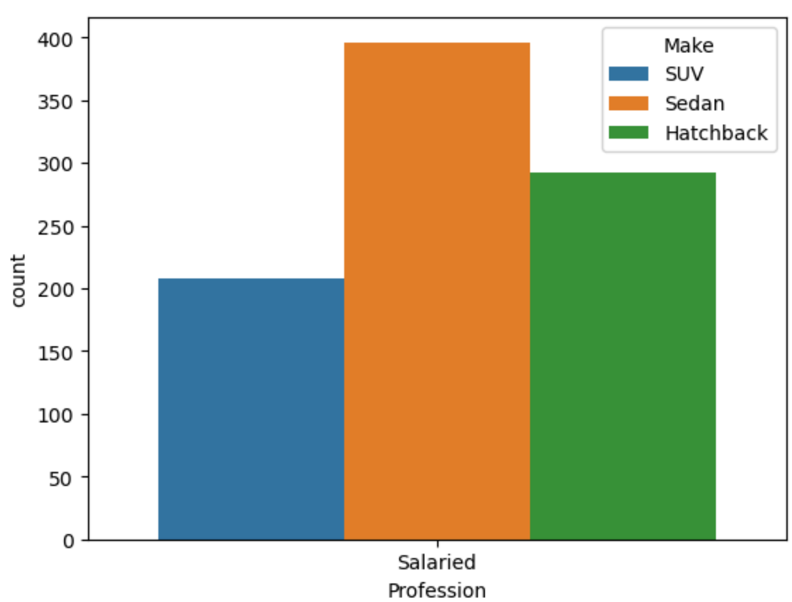
Make	Hatchback	SUV	Sedan	All
Profession				
Business	0.183428	0.056293	0.193548	0.43327
Salaried	0.184693	0.131562	0.250474	0.56673
All	0.368121	0.187856	0.444023	1.00000



There is a higher likelihood of salaried individuals buying sedans.¶

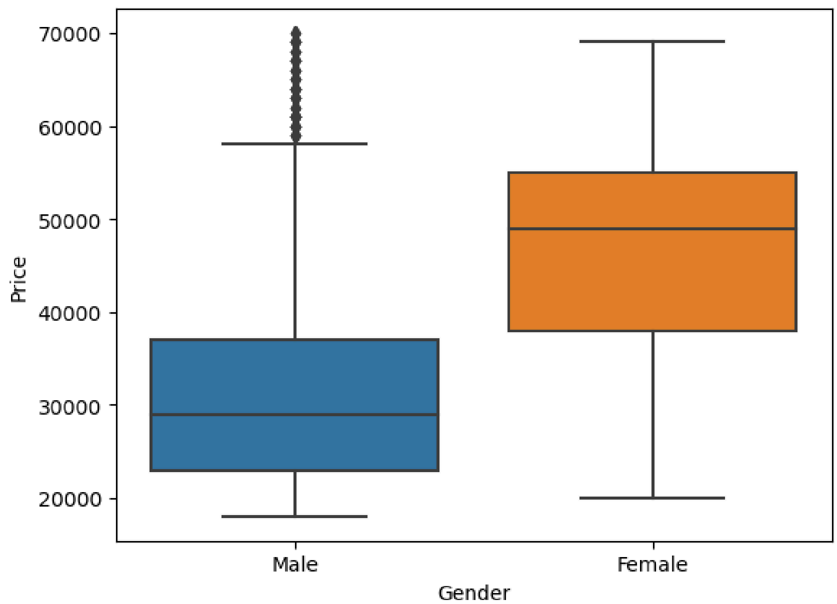
Specifically, 25% of salaried individuals prefer sedans, while the preference among business individuals is 19%.

Q3. 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?



Contrary to Sheldon Cooper's claim, the available data indicates that the salaried profession generally prefers sedans over SUVs and hatchbacks. It appears that the count data reveals a different trend, with a preference for sedans among salaried individuals.

Q4. How does the amount spent on purchasing automobiles vary by gender?

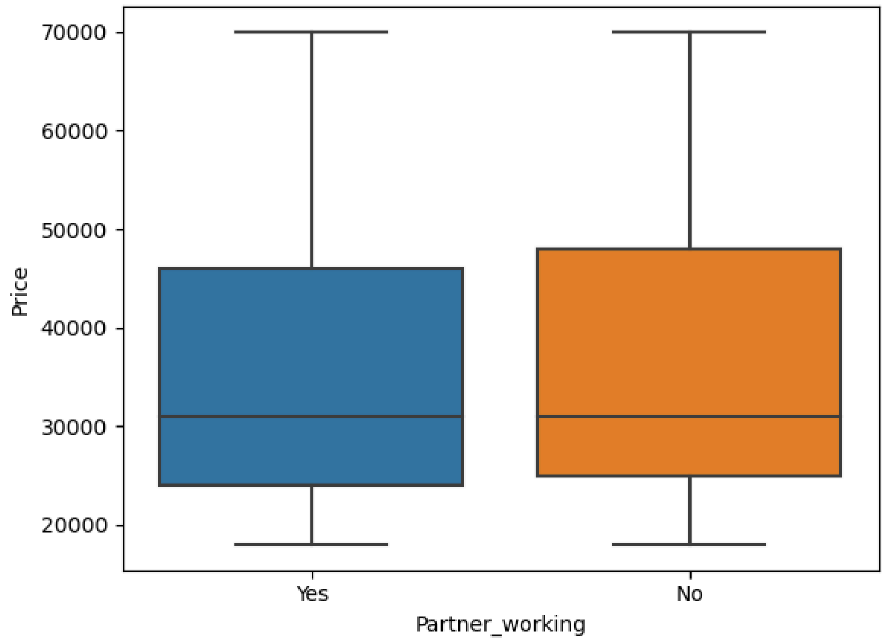


The median car prices for female individuals are higher than those for males, indicating a higher likelihood of females buying higher-priced cars. However, there are outliers among male individuals, suggesting that some males exhibit a preference for higher-priced cars.

Q5. How much money was spent on purchasing automobiles by individuals who took a personal loan?

By grouping individuals who took a personal loan into a new data frame and then summing the prices, it was found that a total of 27,290,000 was spent on purchasing automobiles by these individuals.

Q6. How much money was spent on purchasing automobiles by individuals who took a personal loan?

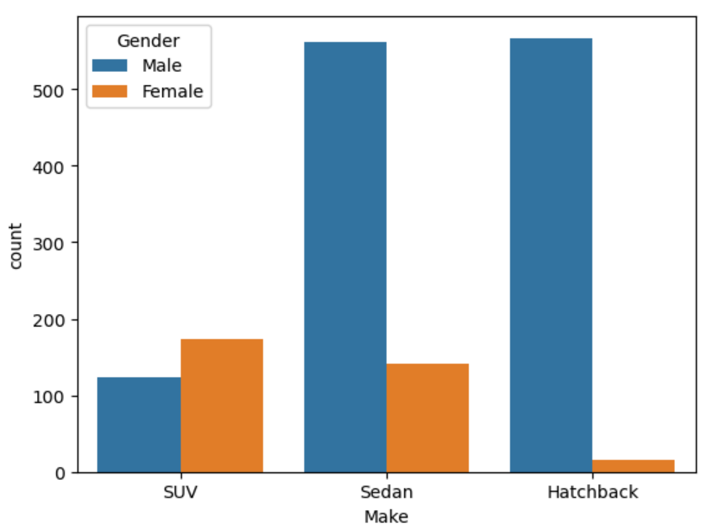


The median car prices for individuals, whether their partner is working or not, are approximately equal. This suggests that the employment status of an individual's partner does not have a significant effect on the choice of car prices within the dataset.

Actionable Insights

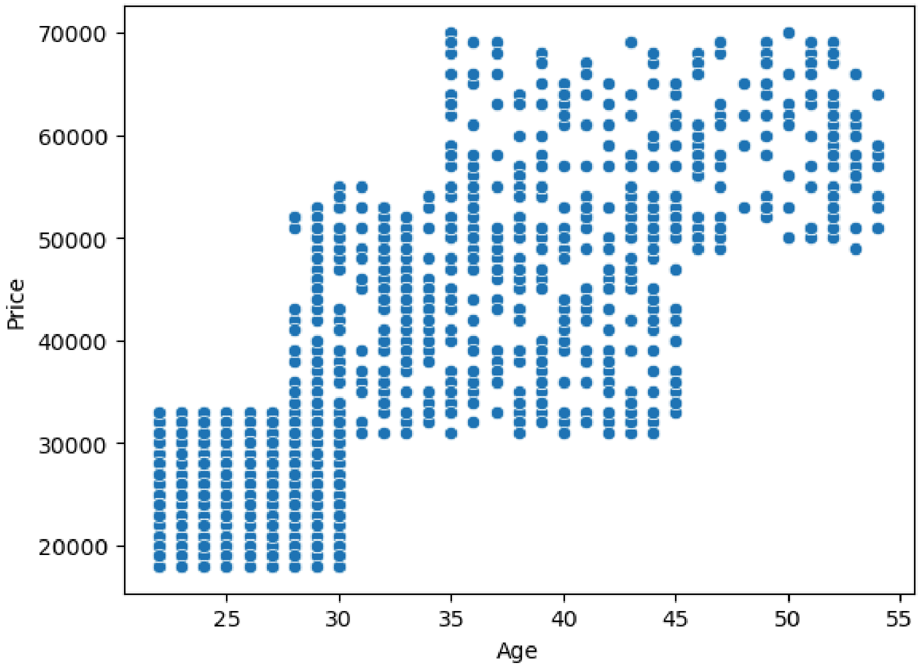
The key takeaways for action are as follows:

- **Targeting SUVs to Females and Sedan, Hatchback to Males:**



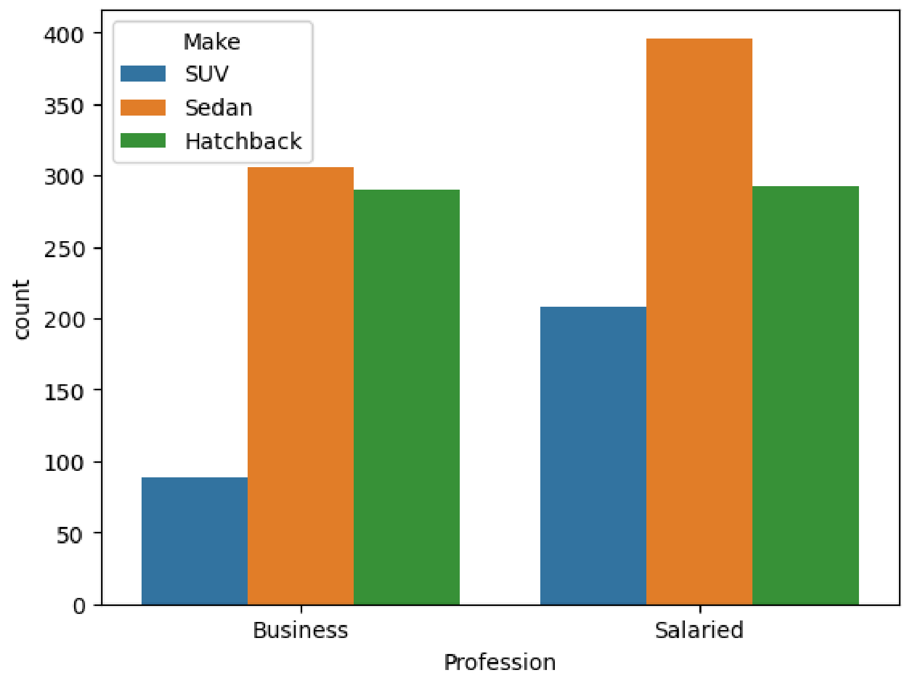
Observing the data reveals a clear preference among males for hatchback and sedan models compared to females. Leveraging this insight, the company can effectively target and tailor marketing strategies to promote these specific car models to the male demographic. This targeted approach may enhance engagement and potentially increase sales among the male customer base.

- **Age-Based Targeting for Cars:**



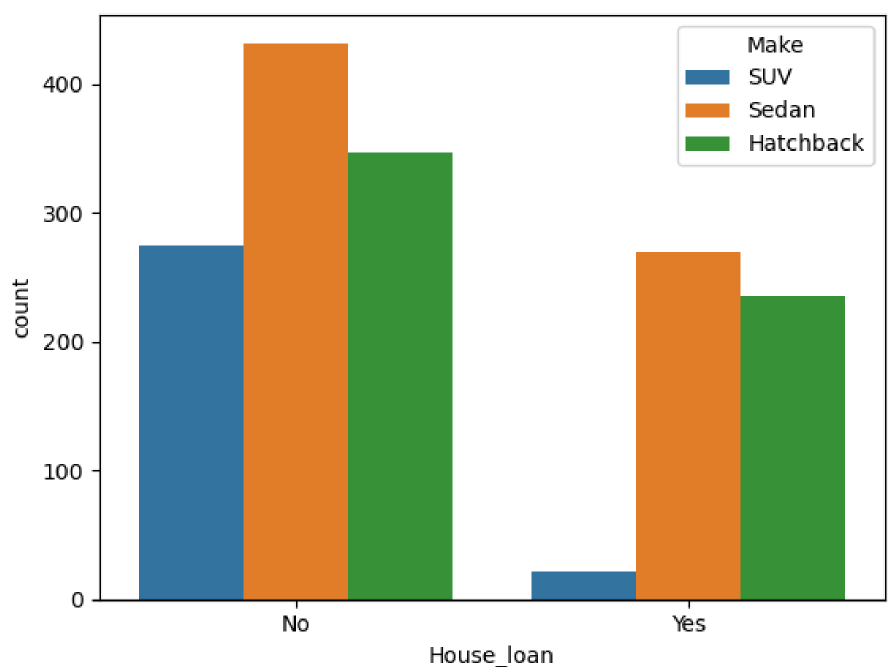
By analyzing the data, it's evident that individuals aged 22 to 27 show a preference for cars priced below 32,000, while those aged 46 and beyond tend to favor cars with prices exceeding 50,000. Leveraging this information, the company can create targeted marketing campaigns and product offerings for specific age groups

- **Occupation-Based Targeting:**



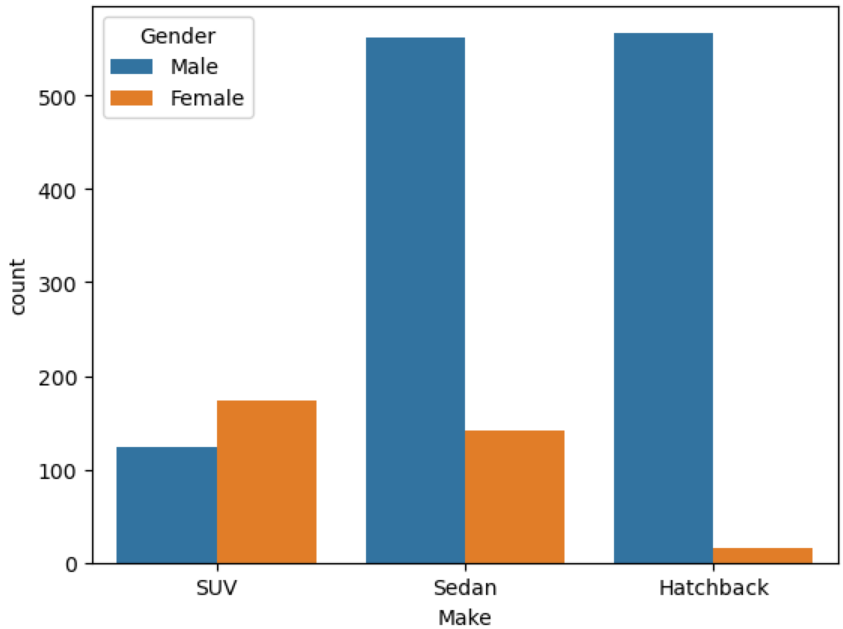
Consider a targeted approach for salaried individuals, emphasizing Sedans and SUVs, as they have shown a higher likelihood of preferring these car types compared to business individuals.

- **Home Loan Influence on Car Preferences:**



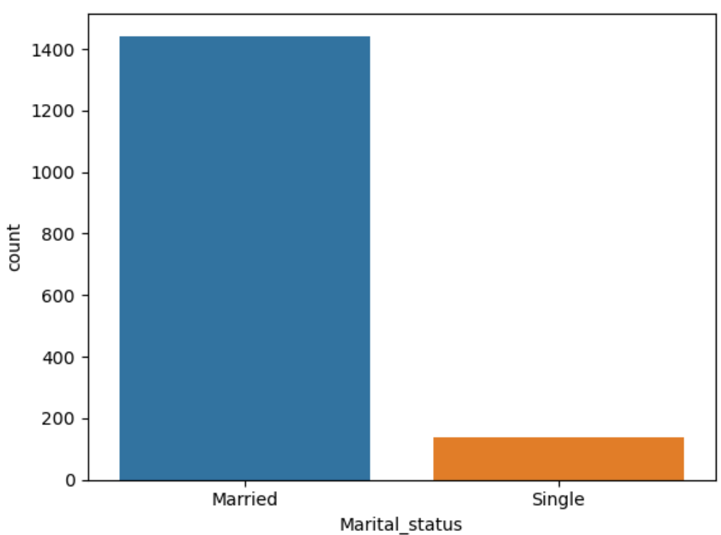
Recognize that individuals with house loans prefer cars below $50,000, while those without house loans lean towards more expensive cars. Tailor marketing strategies accordingly.

- **Market Size Consideration:**



Acknowledge that Sedans and Hatchbacks have a larger market share compared to SUVs. Align inventory and marketing efforts to cater to the broader appeal of Sedans and Hatchbacks.

- **Marital Status and Car Buying:**



Recognize the higher likelihood of car buying among married individuals. Target marketing efforts and promotions to appeal to the preferences and needs of this demographic.

Problem 2

Context

A bank generates revenue through interest, transaction fees, and financial advice, with interest charged on customer loans being a significant source of profits. GODIGT Bank, a mid-sized private bank, offers various banking products and cross-sells asset products to existing customers through different communication methods. However, the bank is facing high credit card attrition, leading them to reevaluate their credit card policy to ensure customers receive the right card for higher spending and intent, resulting in profitable relationships.

Objective

As a Data Scientist at the company, the Data Science team has shared some data. You are supposed to find the key variables that have a vital impact on the analysis which will help the company to improve the business.

Data Description

userid - Unique bank customer-id

card_no - Masked credit card number

card_bin_no - Credit card IIN number

Issuer - Card network issuer

card_type - Credit card type

card_source_data - Credit card sourcing date

high_networth - Customer category based on their net-worth value (A: High to E: Low)

active_30 - Savings/Current/Salary etc. account activity in last 30 days

active_60 - Savings/Current/Salary etc. account activity in last 60 days

active_90 - Savings/Current/Salary etc. account activity in last 90 days

cc_active30 - Credit Card activity in the last 30 days

cc_active60 - Credit Card activity in the last 60 days

cc_active90 - Credit Card activity in the last 90 days

hotlist_flag - Whether card is hot-listed(Any problem noted on the card)

widget_products - Number of convenience products customers hold (dc, cc, net-banking active, mobile banking active, wallet active, etc.)

engagement_products - Number of investment/loan products the customer holds (FD, RD, Personal loan, auto loan)

annual_income_at_source - Annual income recorded in the credit card application

other_bank_cc_holding - Whether the customer holds another bank credit card

bank_vintage - Vintage with the bank (in months) as on Tthmonth

T+1_month_activity - Whether customer uses credit card in T+1 month (future)

T+2_month_activity - Whether customer uses credit card in T+2 month (future)

T+3_month_activity - Whether customer uses credit card in T+3 month (future)

T+6_month_activity - Whether customer uses credit card in T+6 month (future)

T+12_month_activity - Whether customer uses credit card in T+12 month (future)

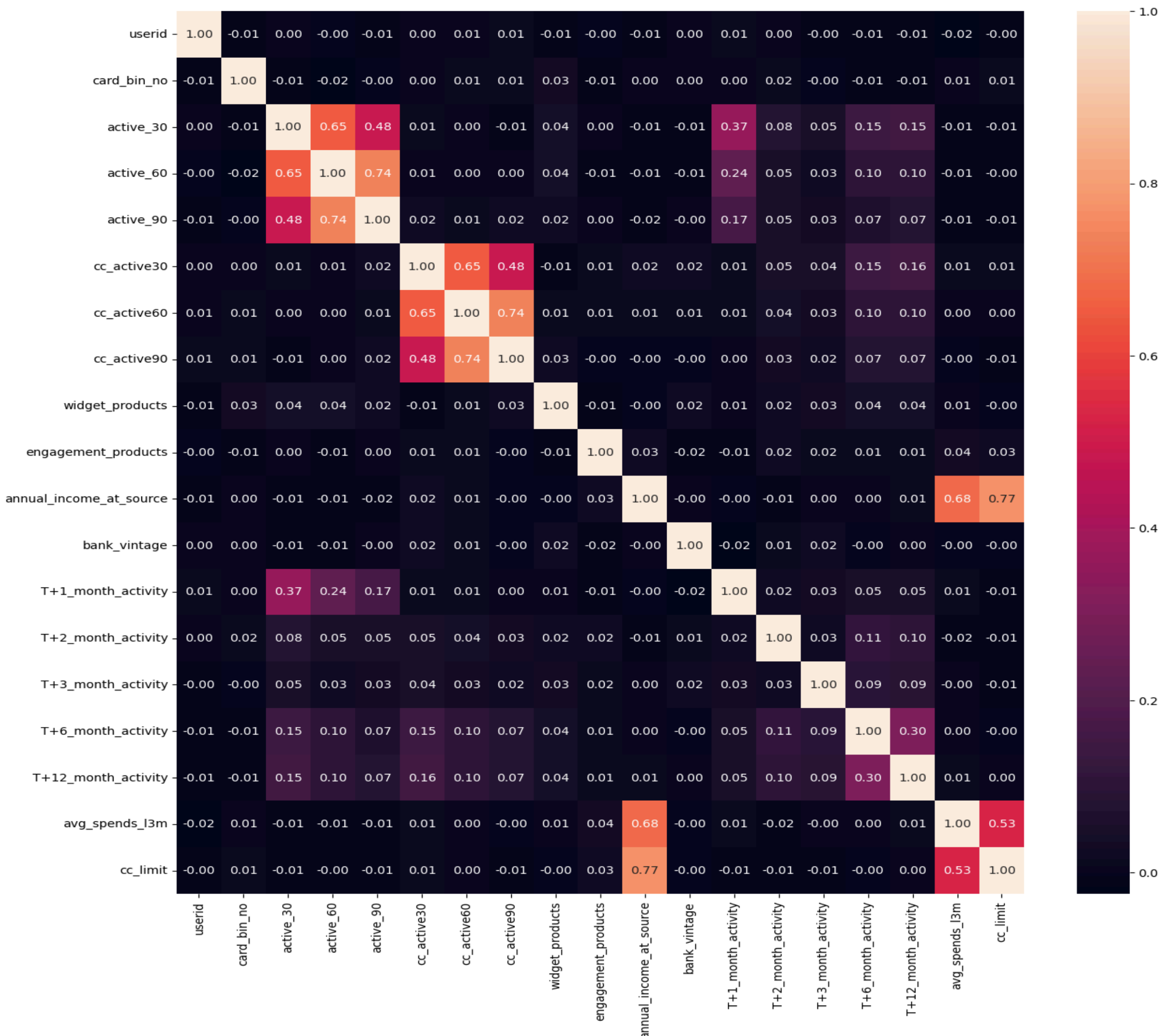
Transactor_revolver - Revolver: Customer who carries balances over from one month to the next. Transactor: Customer who pays off their balances in full every month.

avg_spends_l3m - Average credit card spends in last 3 months

Occupation_at_source - Occupation recorded at the time of credit card application

cc_limit - Current credit card limit

Analyse the dataset and list down the top 5 important variables, along with the business justifications.



- **Annual_income_at_source** - It exhibits a strong correlation with both "avg_spends_l3m" and "cc_limit." This suggests that an individual's annual income at the source plays a pivotal role in determining their spending capacity.
- **avg_spends_l3m -** It demonstrates a robust relationship with "cc_limit," providing valuable insights into the spending patterns of customers. The amount spent over the last three months is strongly linked to the credit card limit, indicating the potential influence of credit limits on customer spending behaviors.
- **Occupation_at_source -** This variable is important to identify the nature of occupation which are the biggest spenders and income earners. This would also help identify the kind of occupation which are currently generating the major

chunk of revenue and which can be the focus of future campaigns to improve market share.

- **Card_type** - The type of card and its utilization serve as indicators of the features' usability and relevance in the market. Analyzing the features offered by each card type is essential to ensure they provide value to users, ultimately driving increased spending.

- **Transactor_revolver**- Based on the preliminary analysis of the data, Number of transactions and Total Amount spent by users who generally pay off their balances in full every month is significantly higher than the users who carry the balances to next month.