SMDM BUISNESS REPORT

BY

Raghavendra Kumar

CONTENTS:

Problem -1

A.What is the important technical information about the dataset that a database administrator would be interested in? (Hint: Information about the size of the dataset and the nature of the variables)

B. Take a critical look at the data and do a preliminary analysis of the variables. Do a quality check of the data so that the variables are consistent? Are there any discrepancies present in the data? If yes, perform preliminary treatment of data.

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

D. Understanding the relationships among the variables in the dataset is crucial for every analytical project. Perform analysis on the data fields to gain deeper insights. Comment on your understanding of the data.

E. Employees working on the existing marketing campaign have made the following remarks. Based on the data and your analysis state whether you agree or disagree with their observations. Justify your answer Based on the data available.

E1) Steve Roger says “Men prefer SUV by a large margin, compared to the women”

E2) Ned Stark believes that a salaried person is more likely to buy a Sedan.

E3) Sheldon Cooper does not believe any of them; he claims that a salaried male is an easier target for a SUV sale over a Sedan Sale.

F. From the given data, comment on the amount spent on purchasing automobiles across the following categories. Comment on how a Business can utilize the results from this exercise. Give justification along with presenting metrics/charts used for arriving at the conclusions.

Give justification along with presenting metrics/charts used for arriving at the conclusions.

F1) Gender

F2) Personal loan

G. From the current data set comment if having a working partner leads to the purchase of a higher-priced car.

H. The main objective of this analysis is to devise an improved marketing strategy to send targeted information to different groups of potential buyers present in the data. For the current analysis use the Gender and Marital status - fields to arrive at groups with similar purchase history.

A.What is the important technical information about the dataset that a database administrator would be interested in? (Hint: Information about the size of the dataset and the nature of the variables)

As a database administrator, you would be interested in several technical aspects of a dataset, by gathering and analysing this information, a database administrator can make informed decisions regarding database design, storage allocation, performance optimization, and data management strategies.

1) Size of the dataset (df) example basic checks like checking how many rows and columns are there in dataset – **df.shape**
2) Data Types: In provided dataset identifying how many integer, float and objects.
   df.info ( ) / df.dtypes
3) Data structure: The schema defines the structure of the dataset, including the names of tables, columns, and their relationships.
4) Data distribution: It is essential to understand the distribution of data within the dataset.

B. Take a critical look at the data and do a preliminary analysis of the variables. Do a quality check of the data so that the variables are consistent? Are there any discrepancies present in the data? If yes, perform preliminary treatment of data.

Basic scenario: -To perform a preliminary analysis of the variables and conduct a quality check of the data, you can follow these steps:

We first load the dataset using **pd.read_csv ( )** from the Pandas library. Then, we perform a preliminary analysis by checking the shape of the dataset and using **df.info ( )** to get summary information about the variables. To identify discrepancies, we can check for missing values using **df.isnull ( ).sum ( )**. Depending on the situation, you can either drop rows with missing values using **df.dropna ( )** or fill the missing values with appropriate methods using **df.fillna ( ).** For handling inconsistent data, finally, after the treatment, we recheck the data using **df.shape ( )** and **df.info ( ),** to ensure consistency and verify the changes made. We can customize the treatment steps based on your specific dataset and requirements.

Dups = dataset. duplicated ( ).sum ( ) - remove the duplicate numbers from all columns.

dataset.describe ( ) – Provides statistical summary about dataset.

df.isnull ( ).sum ( ). – shows the dataset missing values / null values.

Num_col= dataset ['numerical_column'].median () – replacing null values with median values.

Data profiling: Examine the dataset to understand its structure, variables, and general characteristics. This includes reviewing the number of rows and columns, variable names, data types, and initial sample values.

Variable analysis: Analyse each variable individually to identify any discrepancies or inconsistencies. Consider the following aspects:

a. Data types: Check if the data types assigned to each variable are appropriate and consistent with their expected values. For example, numeric variables should not contain string values.

b. Missing values: Identify if there are any missing values and determine their extent. Consider the reasons for missingness and decide how to handle them, either through imputation or exclusion.

c. Unique values: Identify variables with categorical or discrete values and check for the presence of unexpected or inconsistent values. Look for any misspellings, typographical errors, or variations of the same value.

d. Cleaning: Correcting or removing erroneous values, inconsistencies, or outliers based on domain knowledge or predefined rules.

e. Transformation: Converting variables to appropriate formats or units, standardizing values, or normalizing data where necessary.

f. Imputation: Filling in missing values using suitable imputation techniques, such as mean, median, or predictive models.

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

Here are some of the commonly used plots in data visualization. The choice of plot depends on the type of data, the variables of interest, and the insights want to convey. Additionally, there are various libraries in different programming languages (such as Matplotlib, Seaborn, Plotly, in Python) that provide tools and functions for creating these plots.

Identify the variables: Understand the variables in your dataset and their nature (numeric, categorical, etc.). This will help you choose appropriate visualization techniques for each variable.

Univariate analysis: Perform a separate analysis for each variable to understand its distribution and characteristics. Here are some visualization techniques you can use: Histograms, Bar charts, Pie charts and Box plots.

In These formulas and concepts provide a general understanding of the main components in each plot.

Histogram:
Bin width: Determines the width of each bin in the histogram.
Frequency: Number of data points falling within each bin.
Bin count: Total number of bins in the histogram.

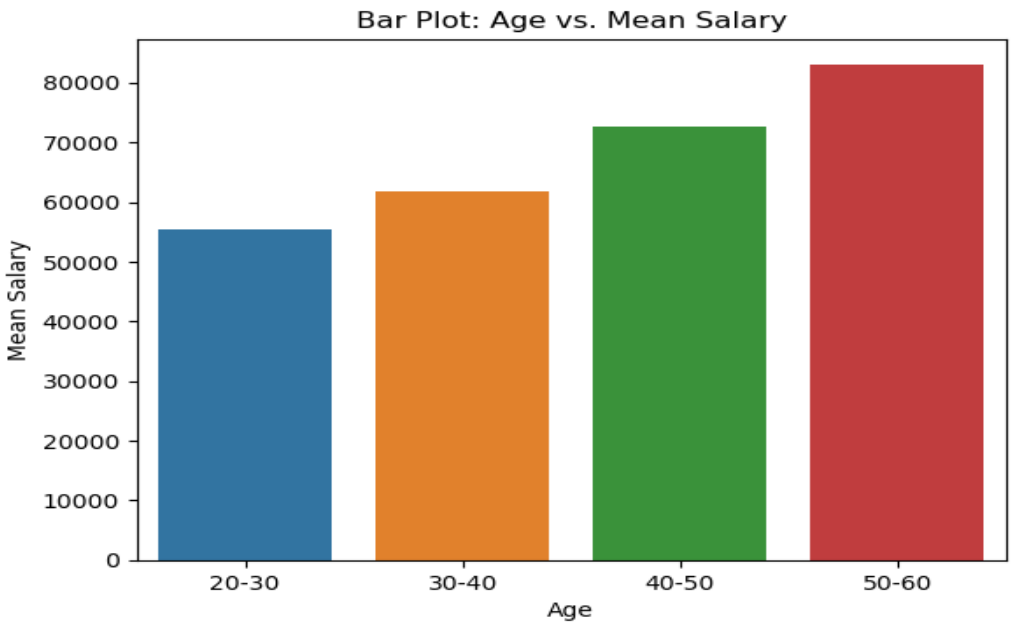
Bar chart:
Bar height: Represents the value or frequency of each category.
Bar width: Determines the width of each bar in the chart.
Category: Represents the distinct categories or groups.

Pie chart:
Sector angle: Represents the proportion of each category as a fraction of the total.
Category: Represents the distinct categories or groups.
Percentage: Shows the percentage of each category in relation to the whole.

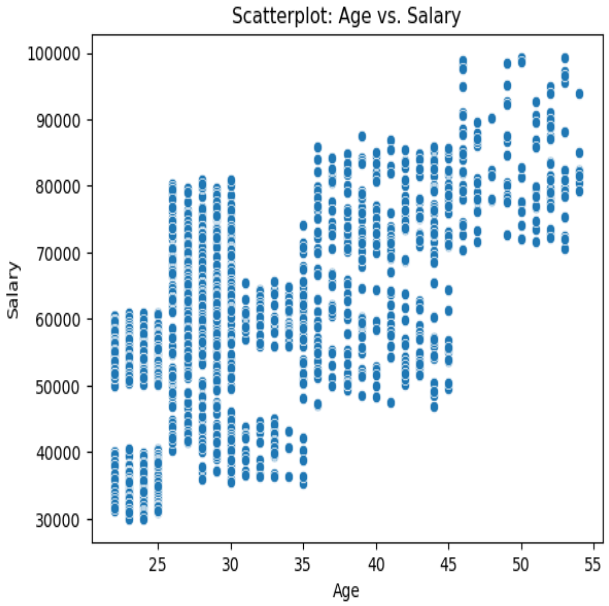
Box plot:
Median: Middle value of the dataset.
Quartiles: Divides the dataset into four equal parts.
Q1 (lower quartile): Median of the lower half of the data.
Q2 (median): Middle value of the dataset.Q3 (upper quartile): Median of the upper half of the data.
Interquartile Range (IQR): Difference between the upper and lower quartiles.
Whiskers: Lines extending from the box that represent the range of the data, excluding outliers.
Outliers: Data points that lie significantly outside the range of the whiskers.

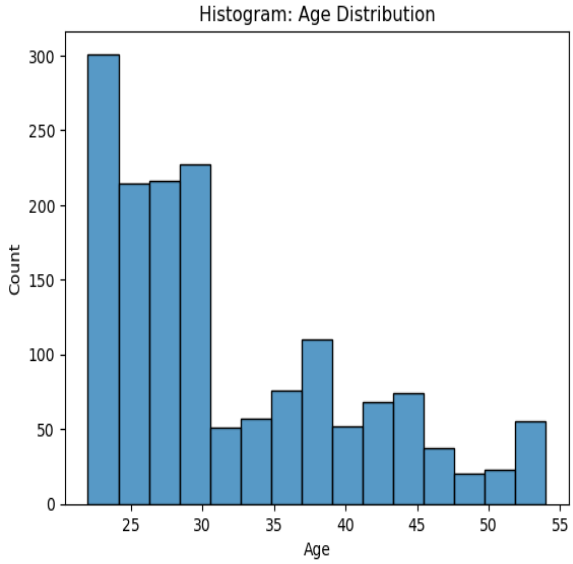
Scatter plot:
Data points: Represent individual data pairs (x, y).
X-axis values: Independent variable values.
Y-axis values: Dependent variable values.
Correlation: Degree of linear relationship between the x and y variables

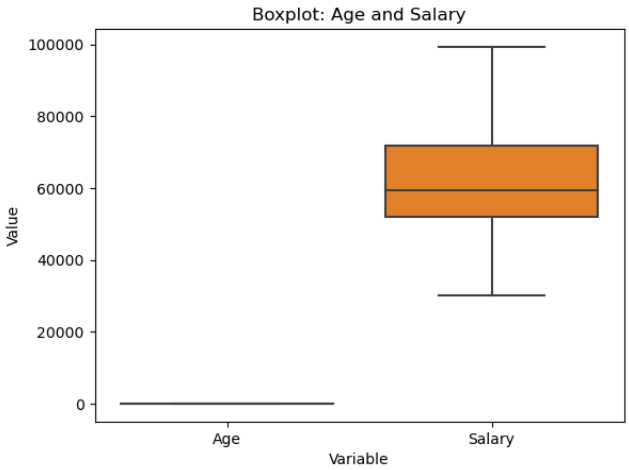
Numerical to Numerical variables graphs:



Bar plot that visualizes the relationship between age groups and the mean salary. The dataset is first divided into age groups using the **pd.cut ( )** function and then grouped by these age groups to calculate the mean salary. The resulting mean salary values are then plotted using **sns.barplot ( ).** The plot shows the mean salary for each age group, allowing for a comparison of salary levels across different age ranges.

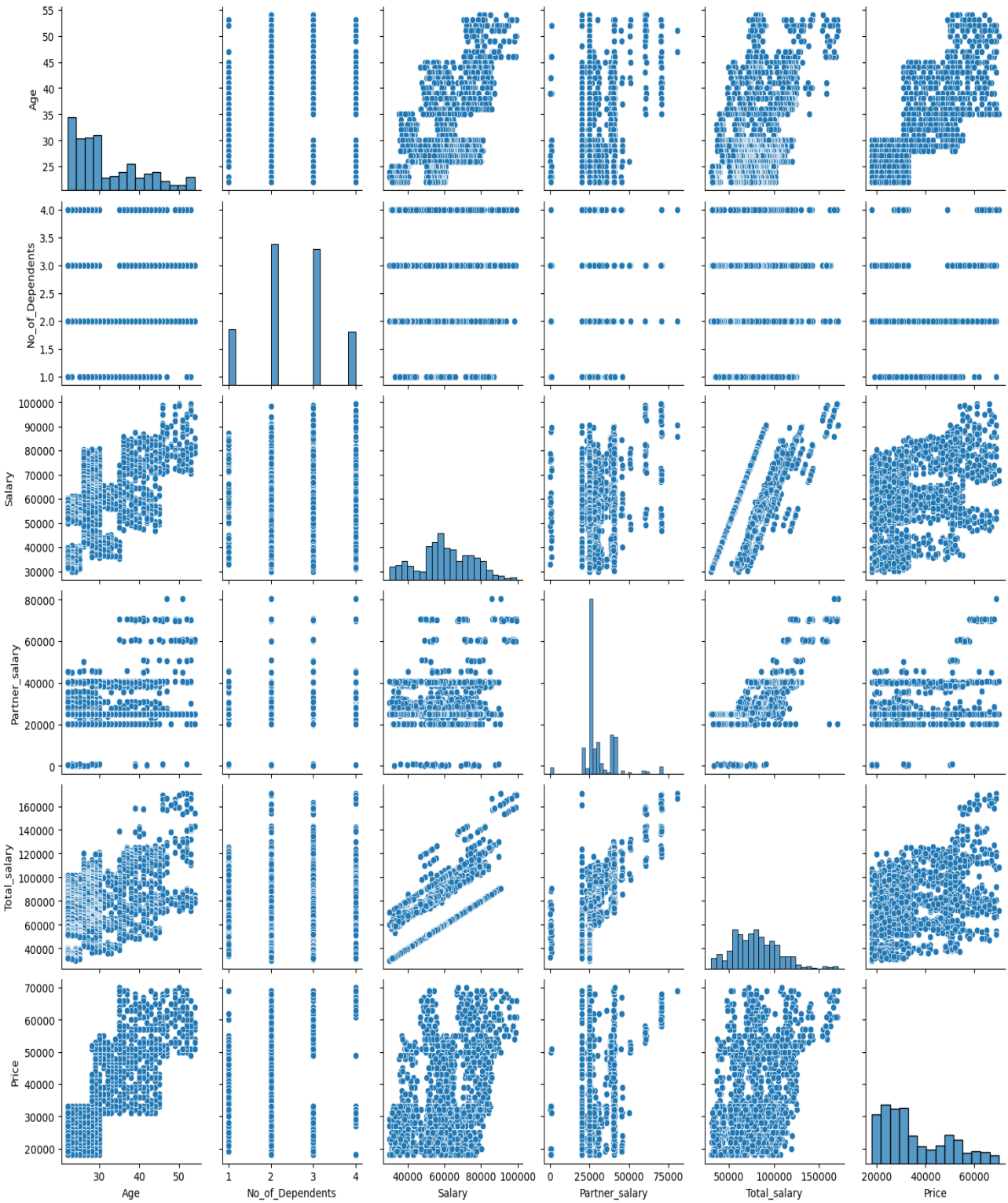






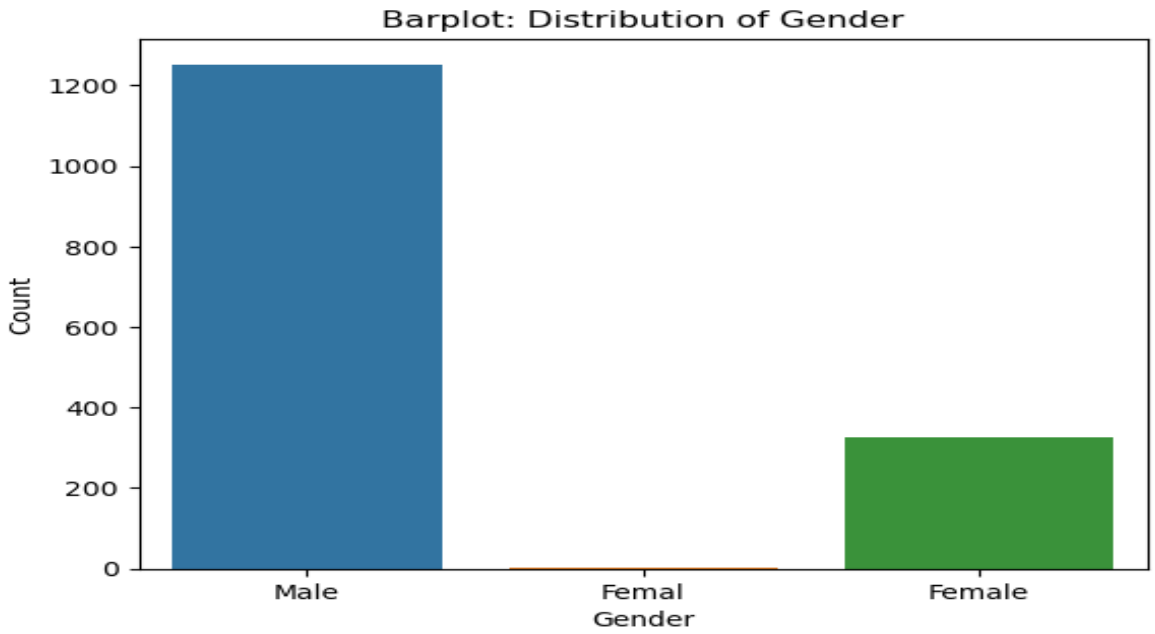
This above graphs are examples of Scatter, boxplot, bar plot and line plots that visualizes the selected variables in the dataset.

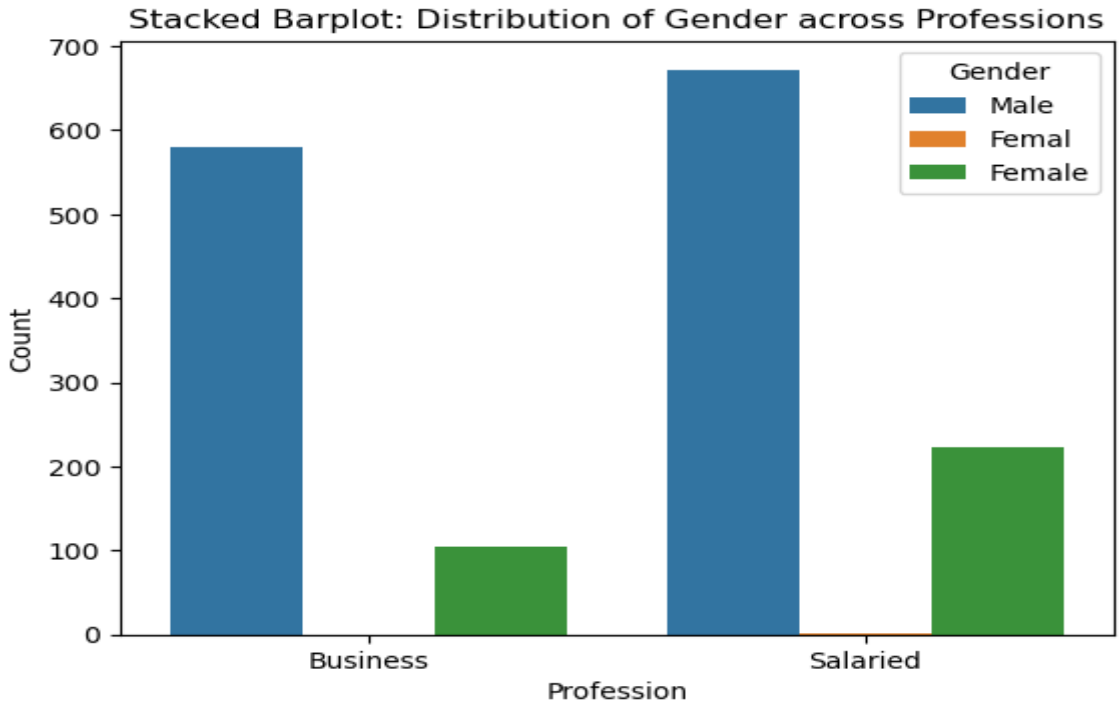
Numerical columns pair plots

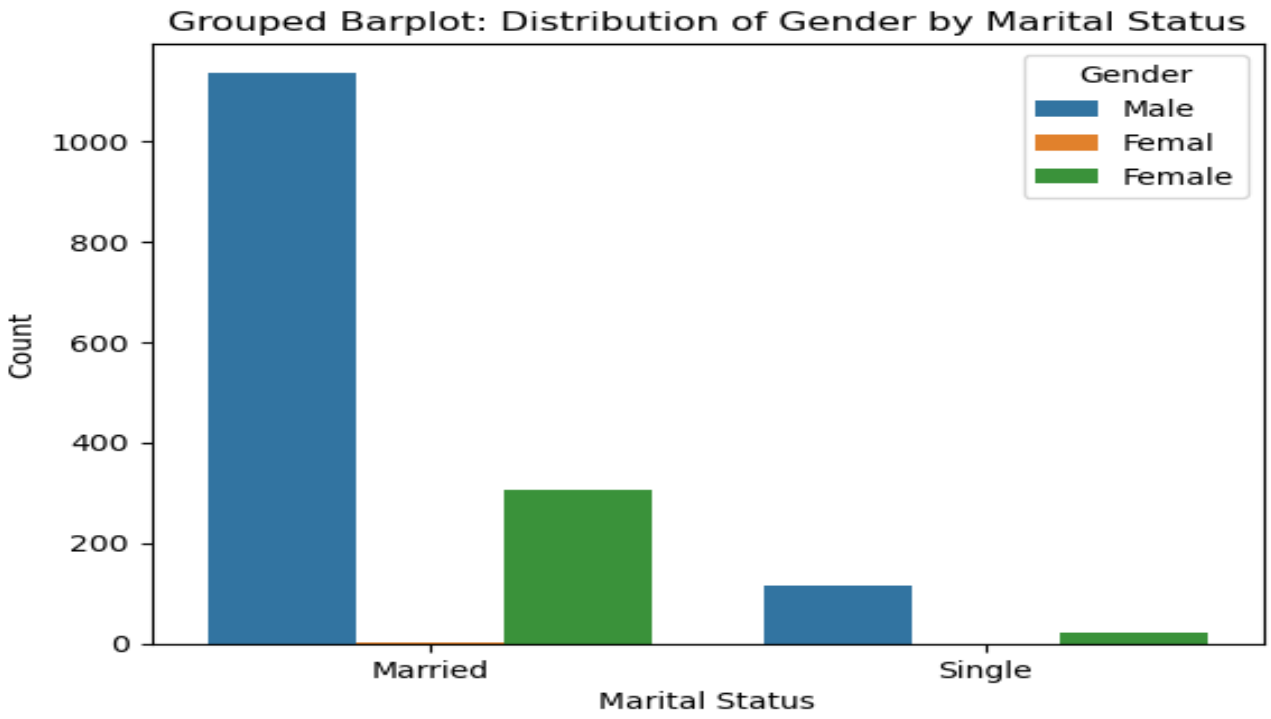


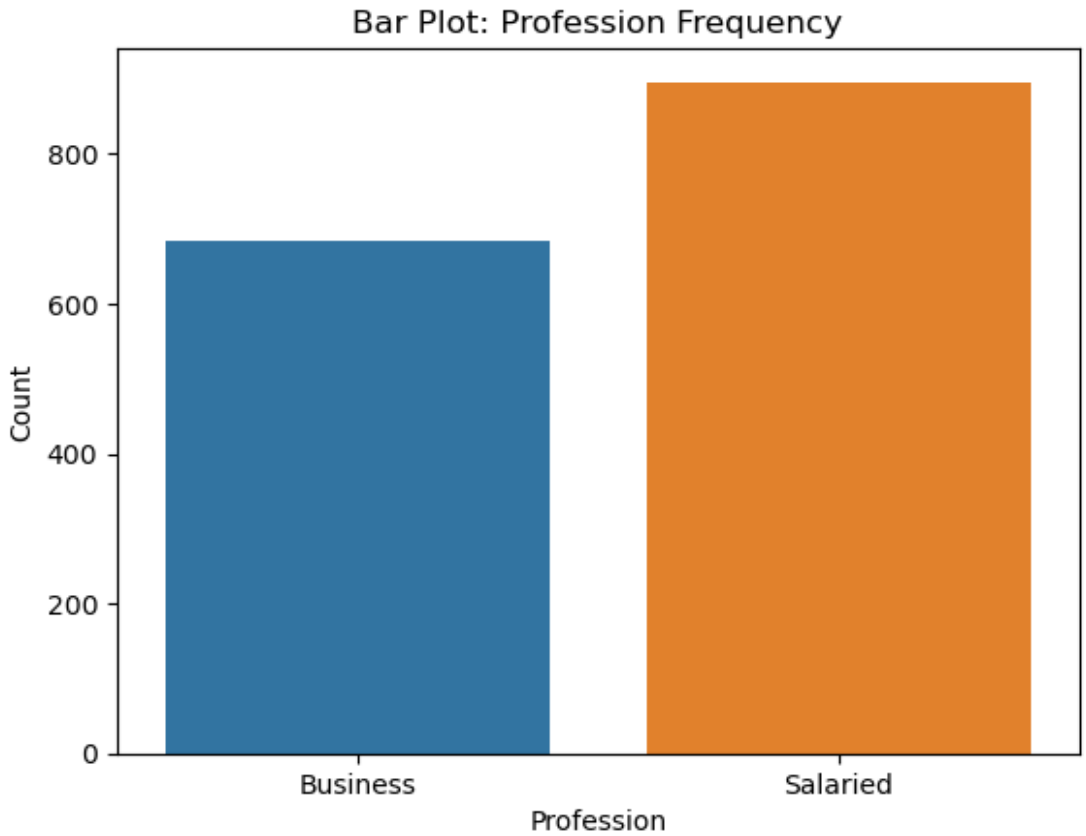
Numerical variables: `['Age', 'No_of_Dependents', 'Salary', 'partner salary ', 'Total salary', 'Price']`

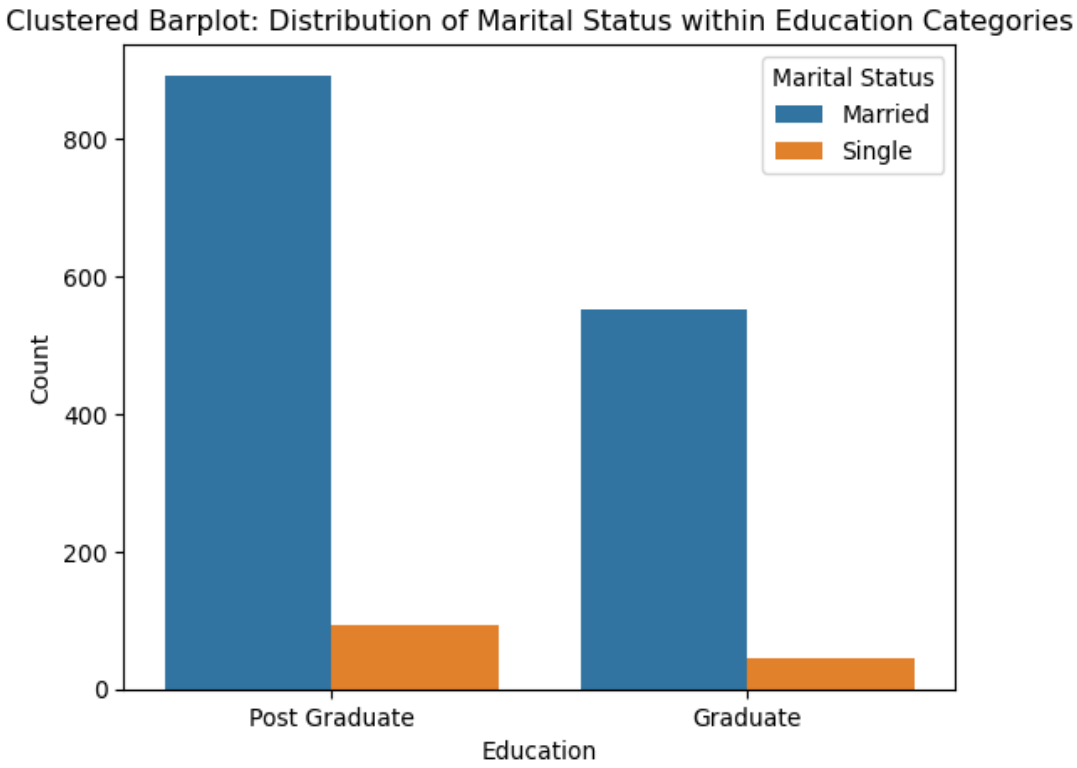
Categorical to Categorical variables





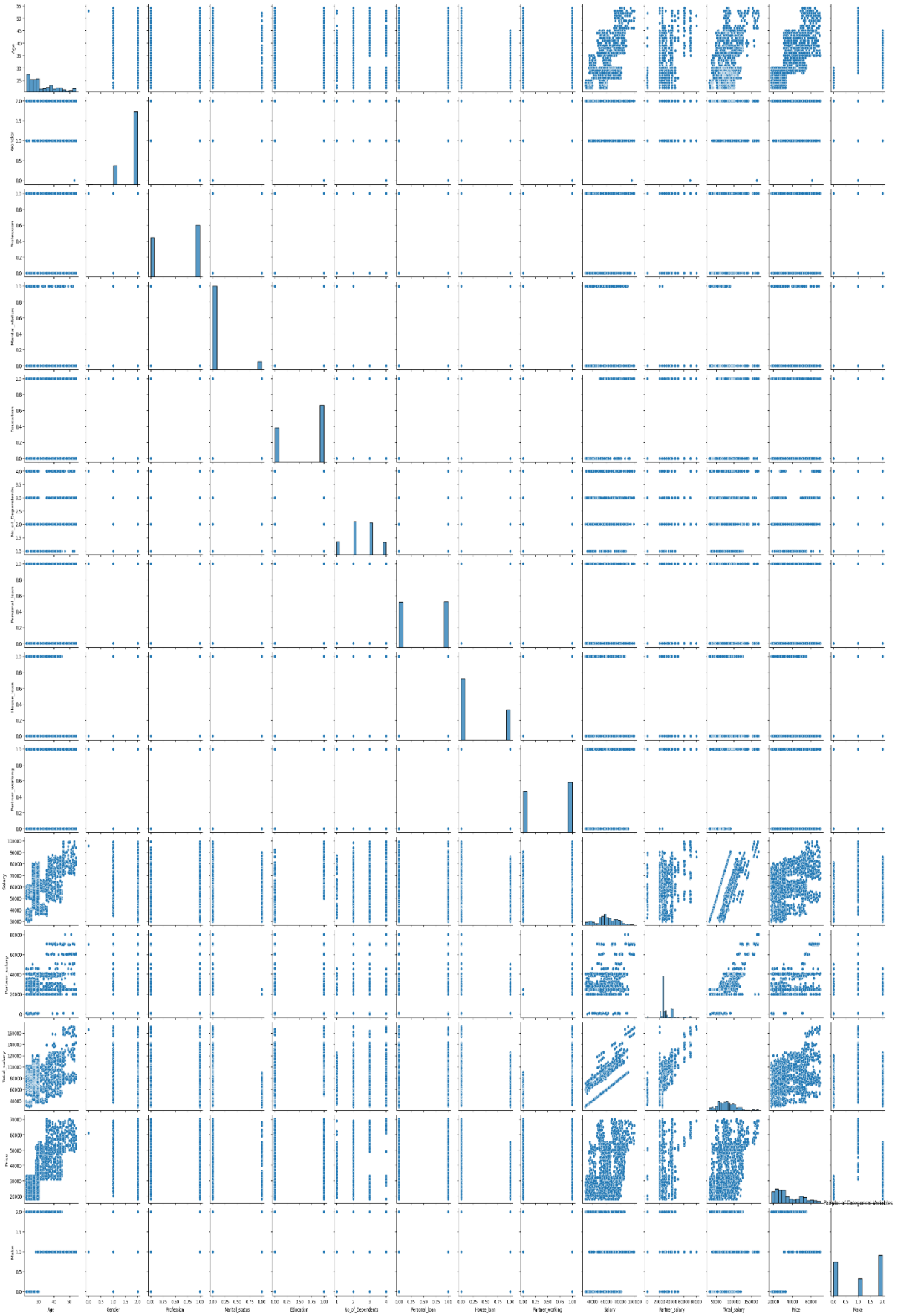

Grouped Barplot: Distribution of Gender by Marital Status
Gender
Male
Femal
Female
1000
800
600
400
200
0
Count
Married
Single
Marital Status


Bar Plot: Profession Frequency
800
600
400
200
0
Count
Business
Salaried
Profession

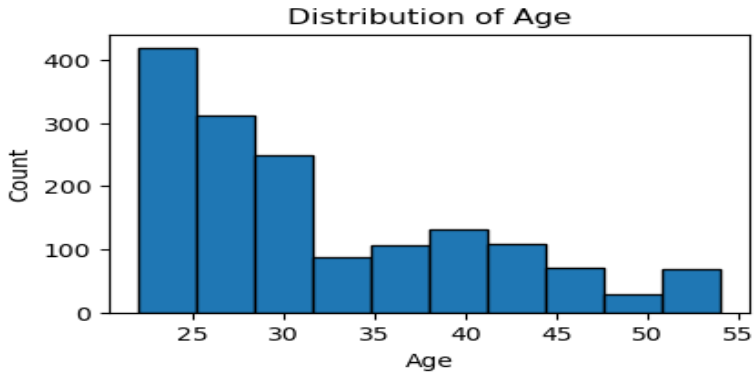


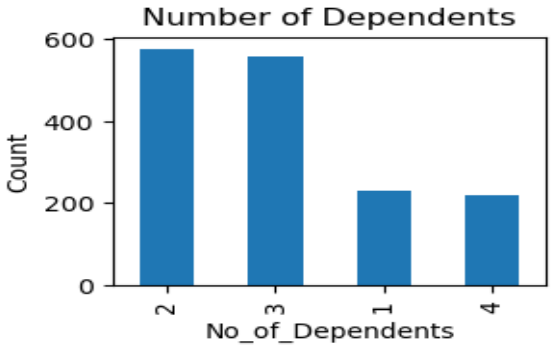
Categorical variables: `['Gender', 'Profession', 'Marital status', 'Education', 'Personal loan', 'House loan', 'Partner working', 'Make']`

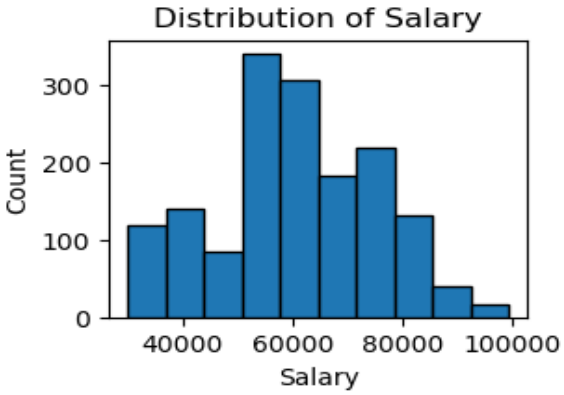
Categorical column Pair plots

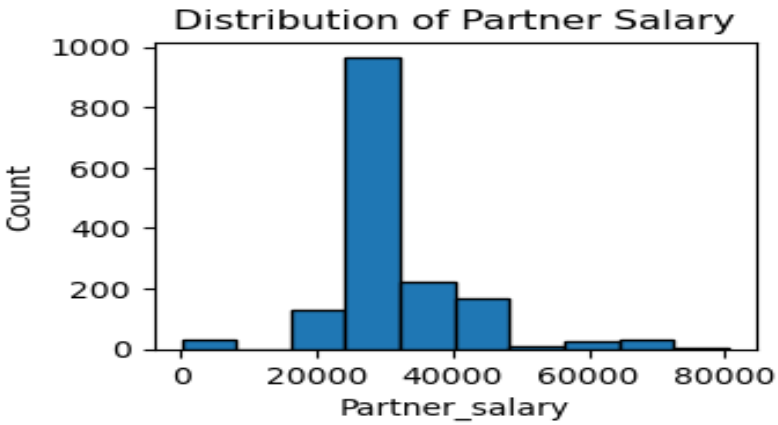


Univariate analysis

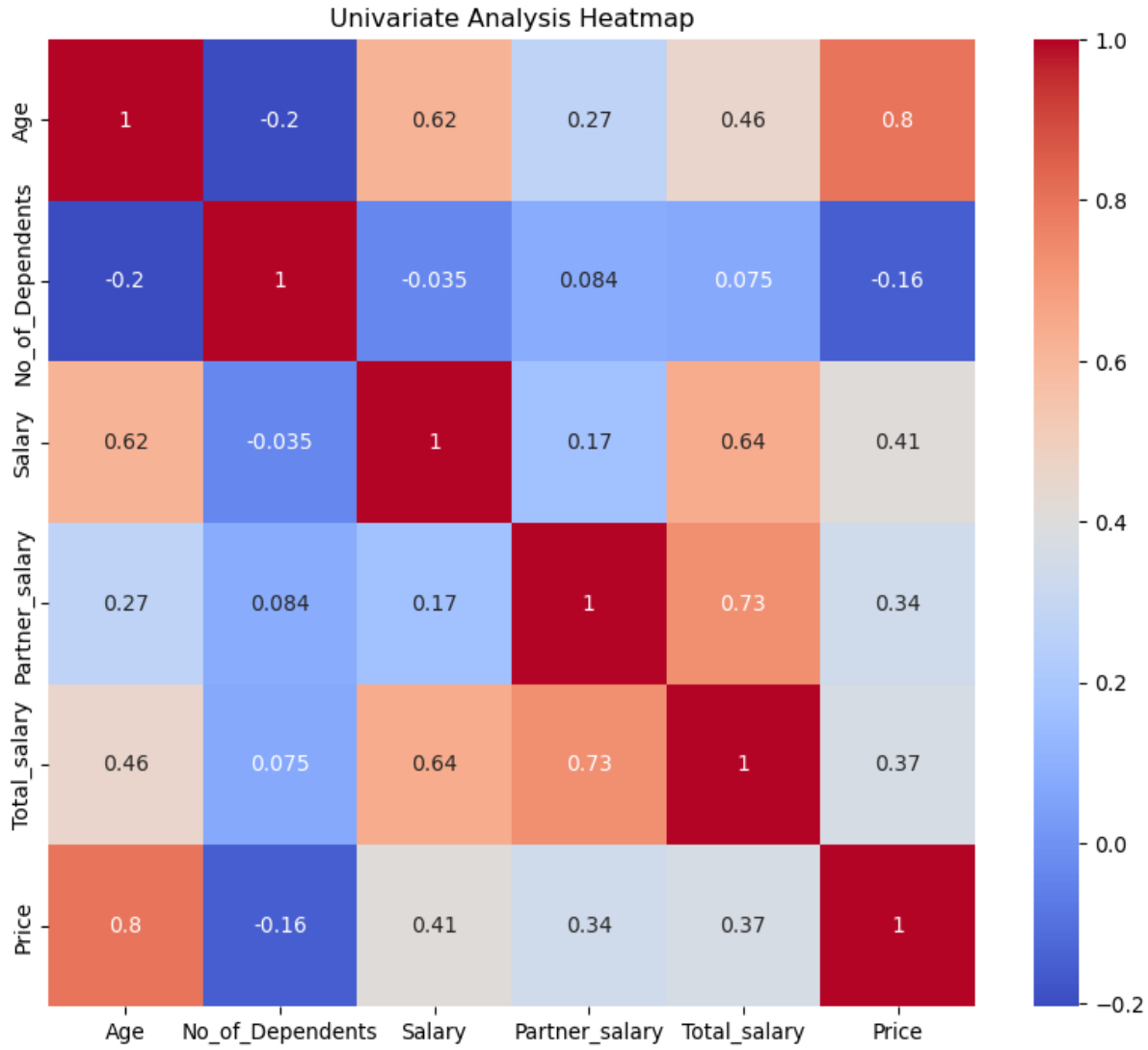








Univariate Analysis Heat map



D. Understanding the relationships among the variables in the dataset is crucial for every analytical project. Perform analysis on the data fields to gain deeper insights. Comment on your understanding of the data.

To perform analysis on a dataset and gain deeper insights, we can follow these steps:

Data Exploration: Start by understanding the structure and content of the dataset. Look at the number of variables, their types (numeric, categorical, etc.), and the range of values they take. Identify any missing or incomplete data,

Descriptive Statistics: Calculate summary statistics for numerical variables such as mean, median, standard deviation, minimum, and maximum. This will give you a sense of the distribution and variability of the data.

Bivariate analysis: Explore relationships between pairs of variables to uncover insights and patterns. Here are some techniques for visualizing relationships: Scatter plots, bar plots. Line plots and Heatmaps.

Multivariate analysis: Explore relationships among multiple variables simultaneously. Techniques such as parallel coordinates, scatterplot matrices, or heat maps with hierarchical clustering can help visualize complex interactions and identify patterns.

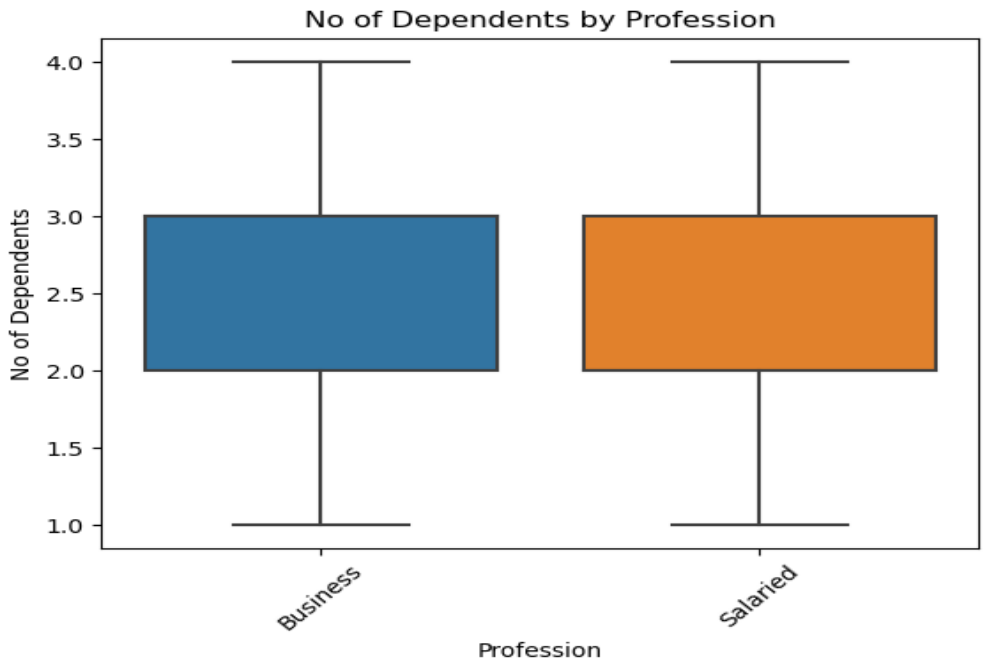
Extract insights: Based on your visualizations, extract meaningful insights that can be utilized by the business. Look for trends, outliers, correlations, or patterns that can inform decision-making, identify potential business opportunities, or address challenges.

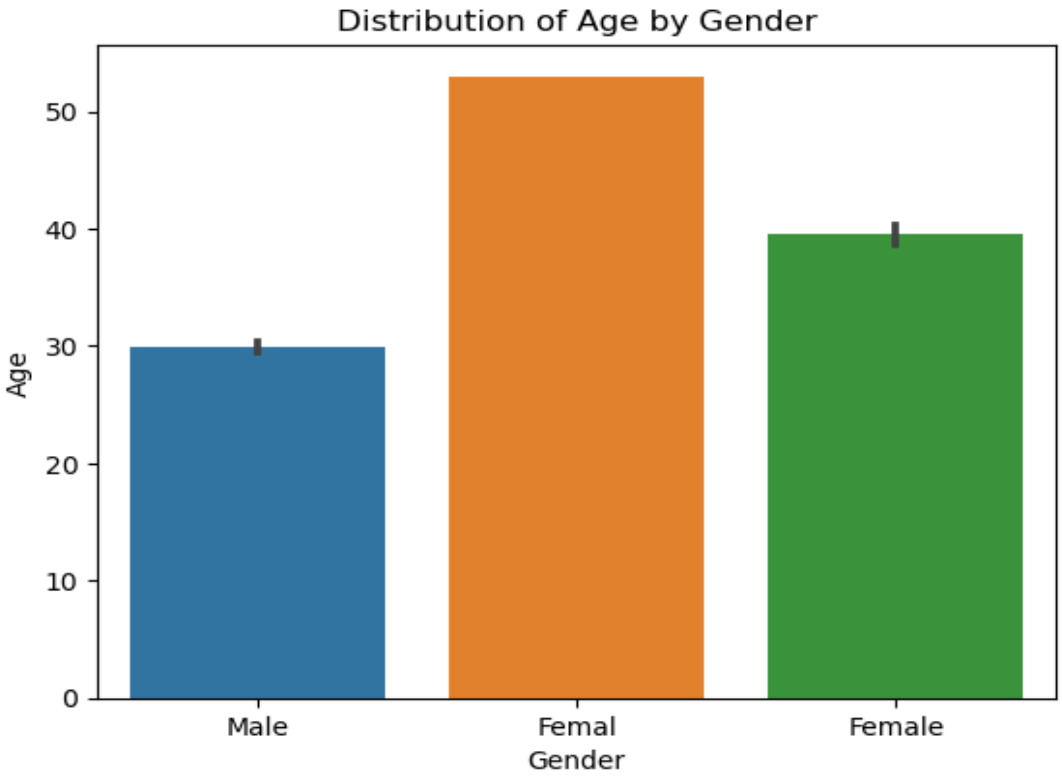
Data Visualization: Create visualizations such as histograms, scatter plots, or box plots to understand the distribution and relationships between variables. Visualizations can help identify patterns, outliers, and potential correlations.

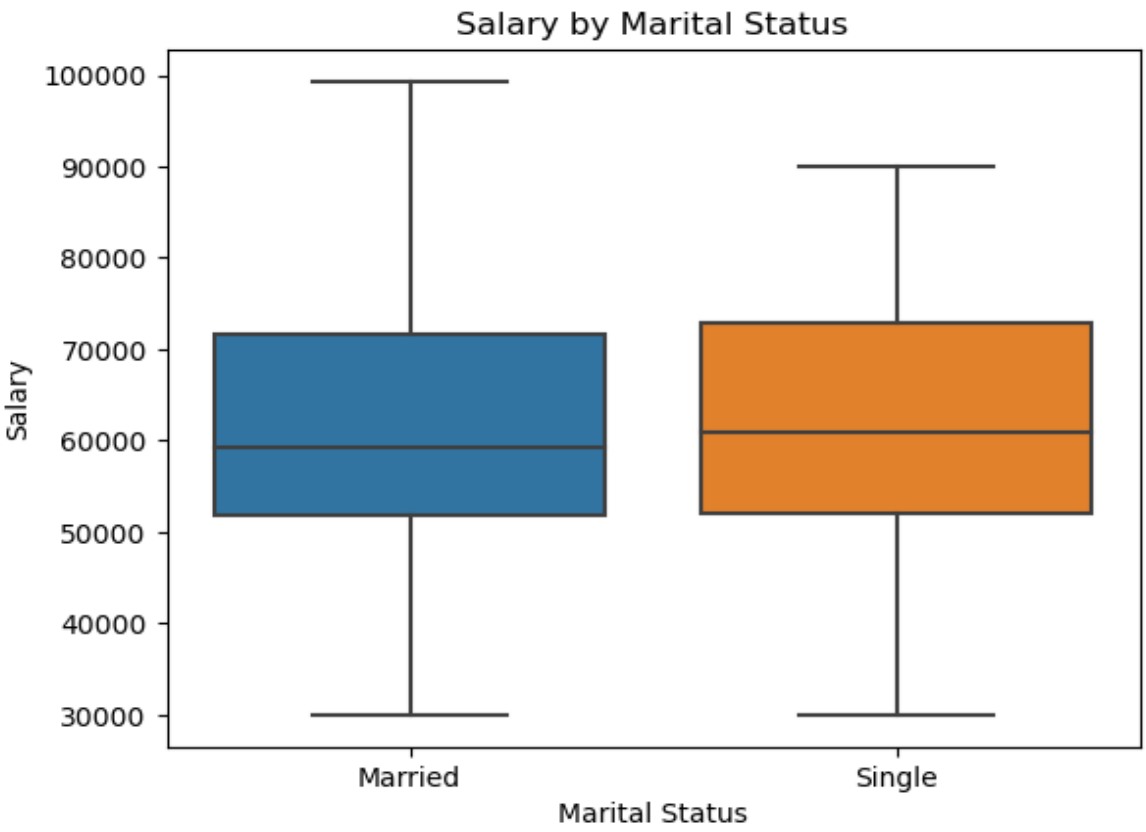
Correlation Analysis: Compute correlation coefficients (e.g., Pearson correlation) to measure the strength and direction of the linear relationship between pairs of numerical variables. This will help you identify variables that are highly correlated and potentially redundant.

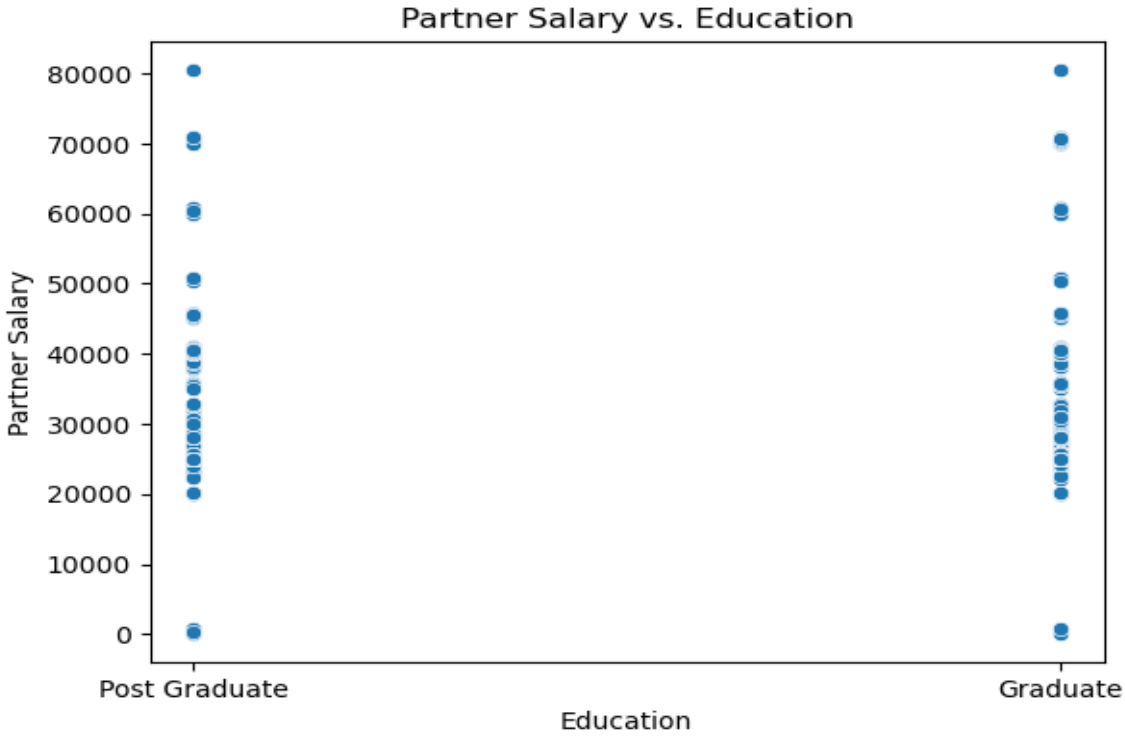
Throughout the analysis, document the findings and insights. Look for interesting patterns, outliers, relationships, or any unexpected observations. Commenting the strengths and limitations of the data, potential biases, and any additional data that might be useful for a more comprehensive analysis.

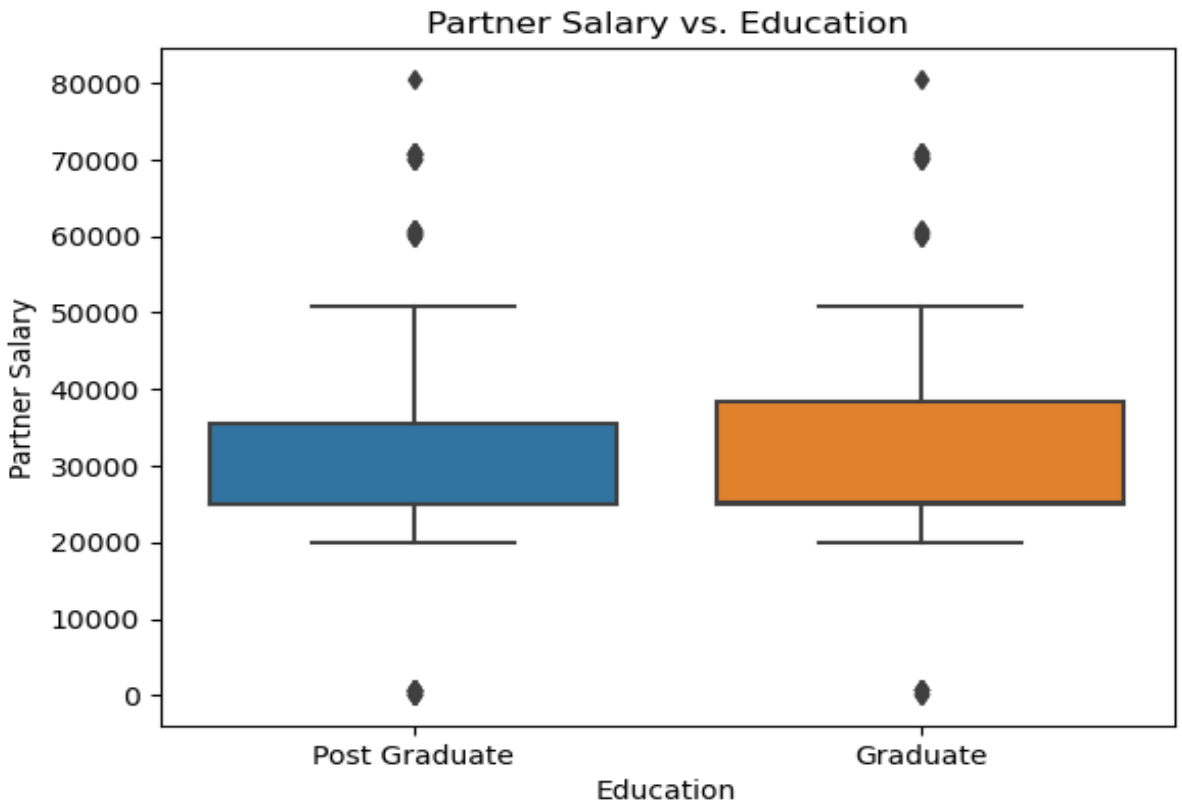
Two variable plots –includes Bivariate and Multivariate plots

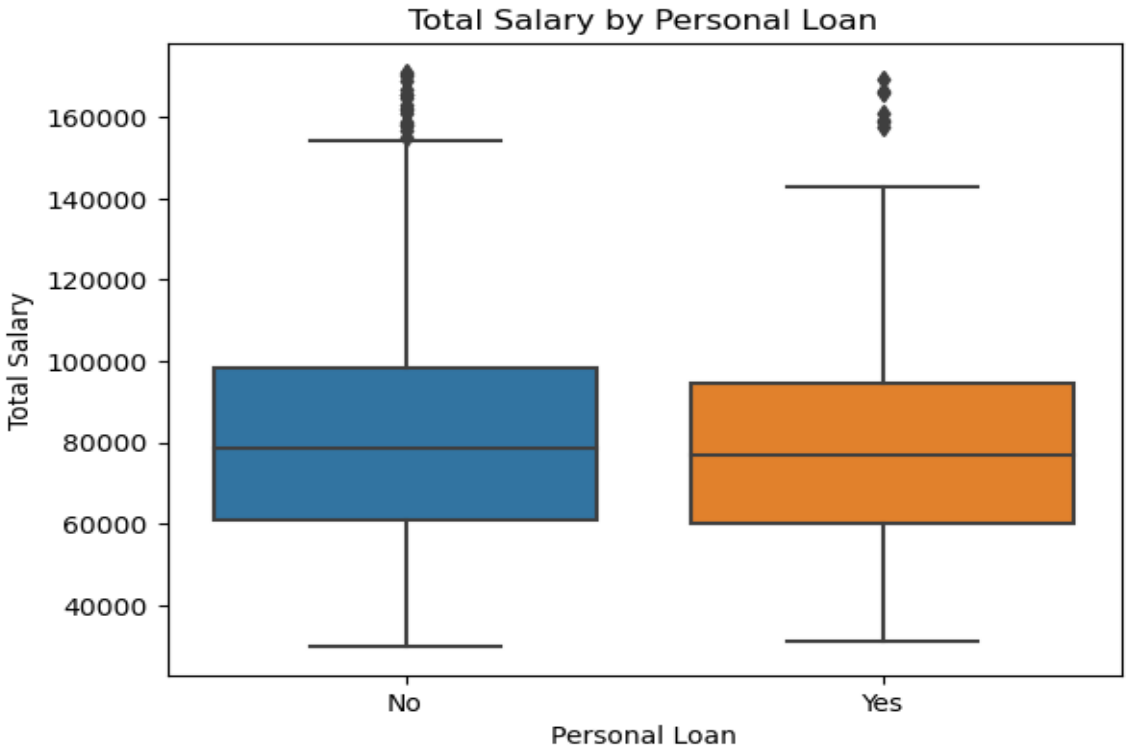


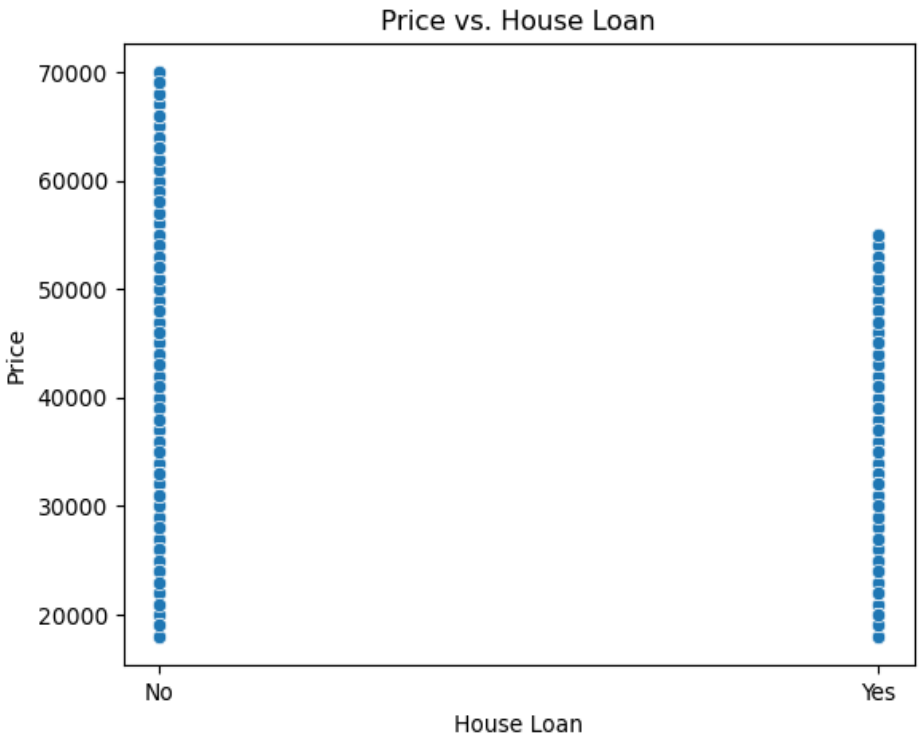


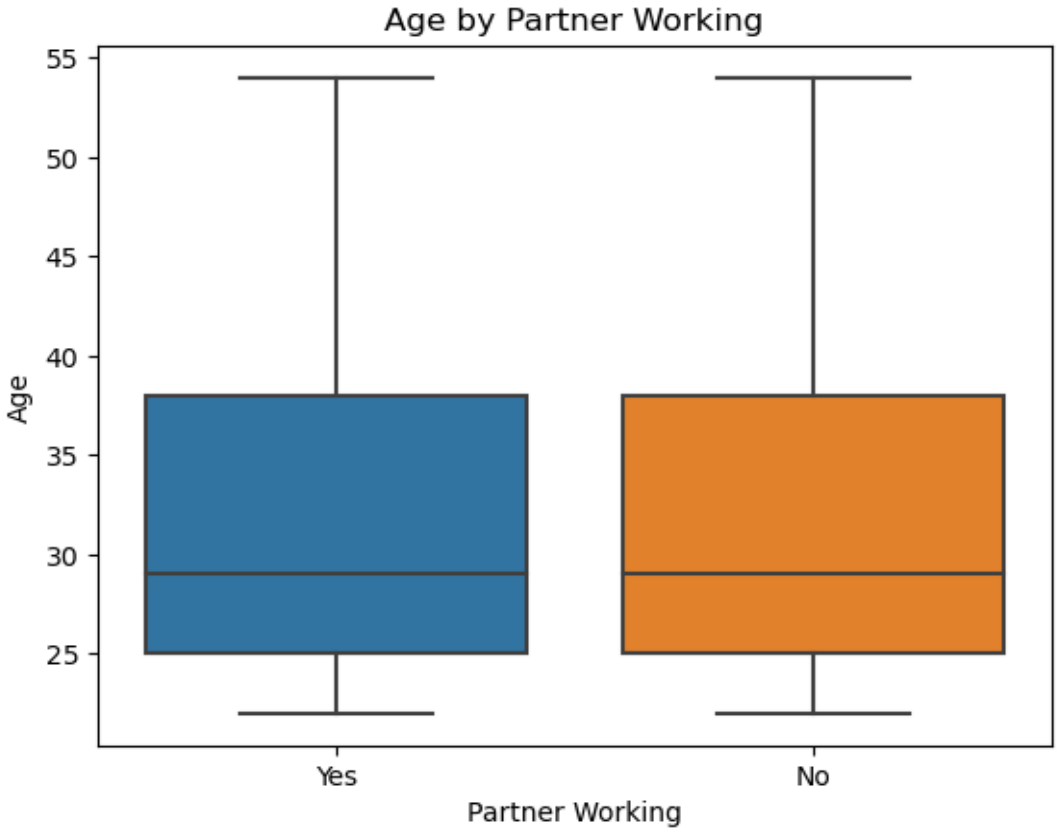

Salary by Marital Status
100000
90000
80000
70000
60000
50000
40000
30000
Salary
Married
Single
Marital Status

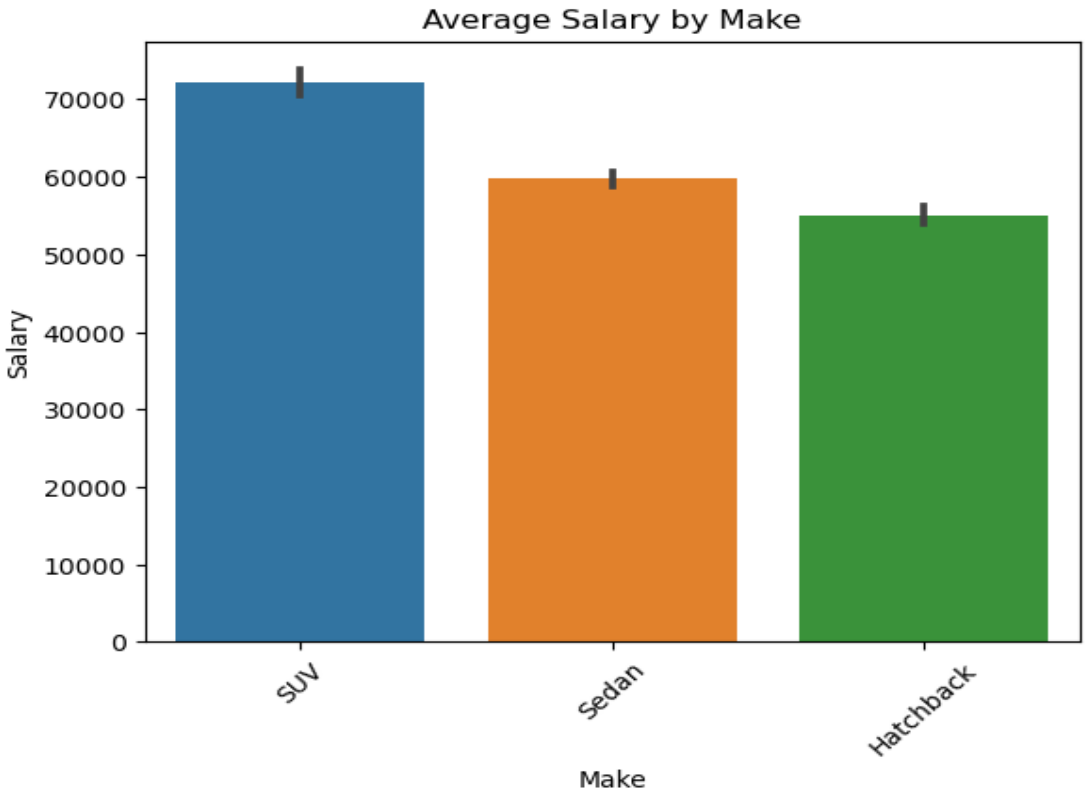

Partner Salary vs. Education
80000
70000
60000
50000
40000
30000
20000
10000
0
Partner Salary
Post Graduate
Graduate
Education

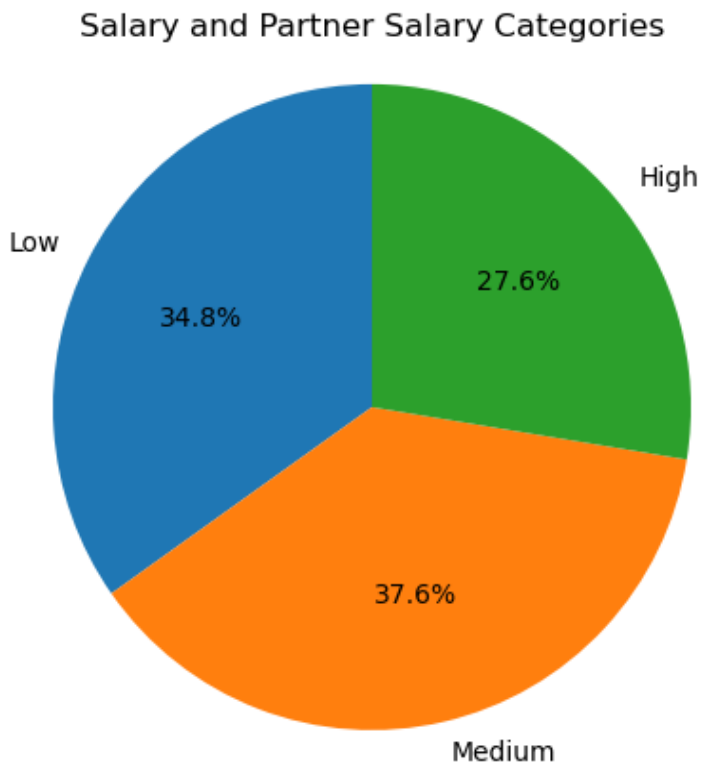

Partner Salary vs. Education
Partner Salary
80000
70000
60000
50000
40000
30000
20000
10000
0
Post Graduate
Graduate
Education


Total Salary by Personal Loan
Total Salary
160000
140000
120000
100000
80000
60000
40000
No
Yes
Personal Loan

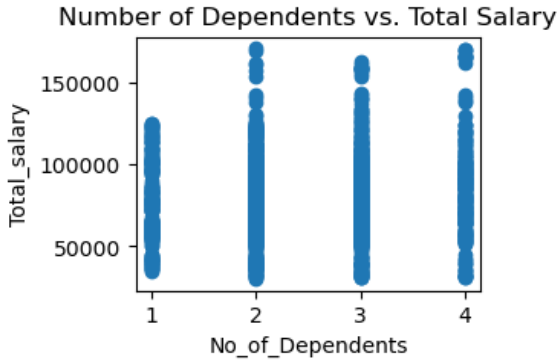

Price vs. House Loan
70000
60000
50000
40000
30000
20000
Price
No
Yes
House Loan

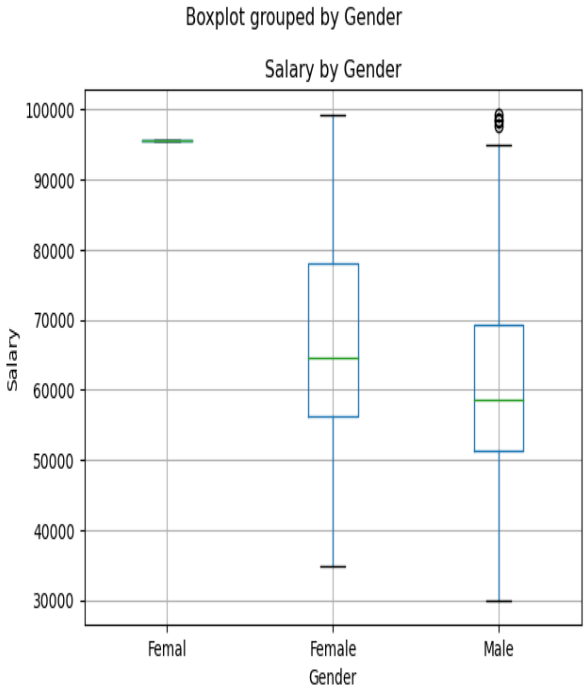

Age by Partner Working
55
50
45
40
35
30
25
Age
Yes
No
Partner Working

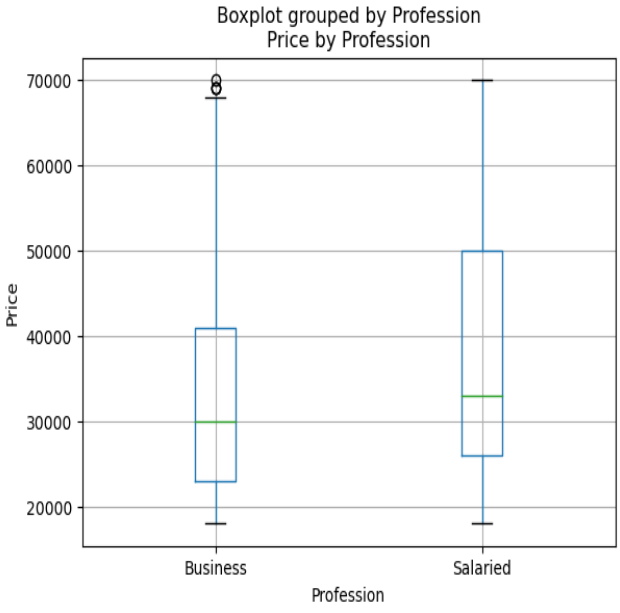

Average Salary by Make
Salary
70000
60000
50000
40000
30000
20000
10000
0
SUV
Sedan
Hatchback
Make


Salary and Partner Salary Categories
Low
34.8%
High
27.6%
37.6%
Medium

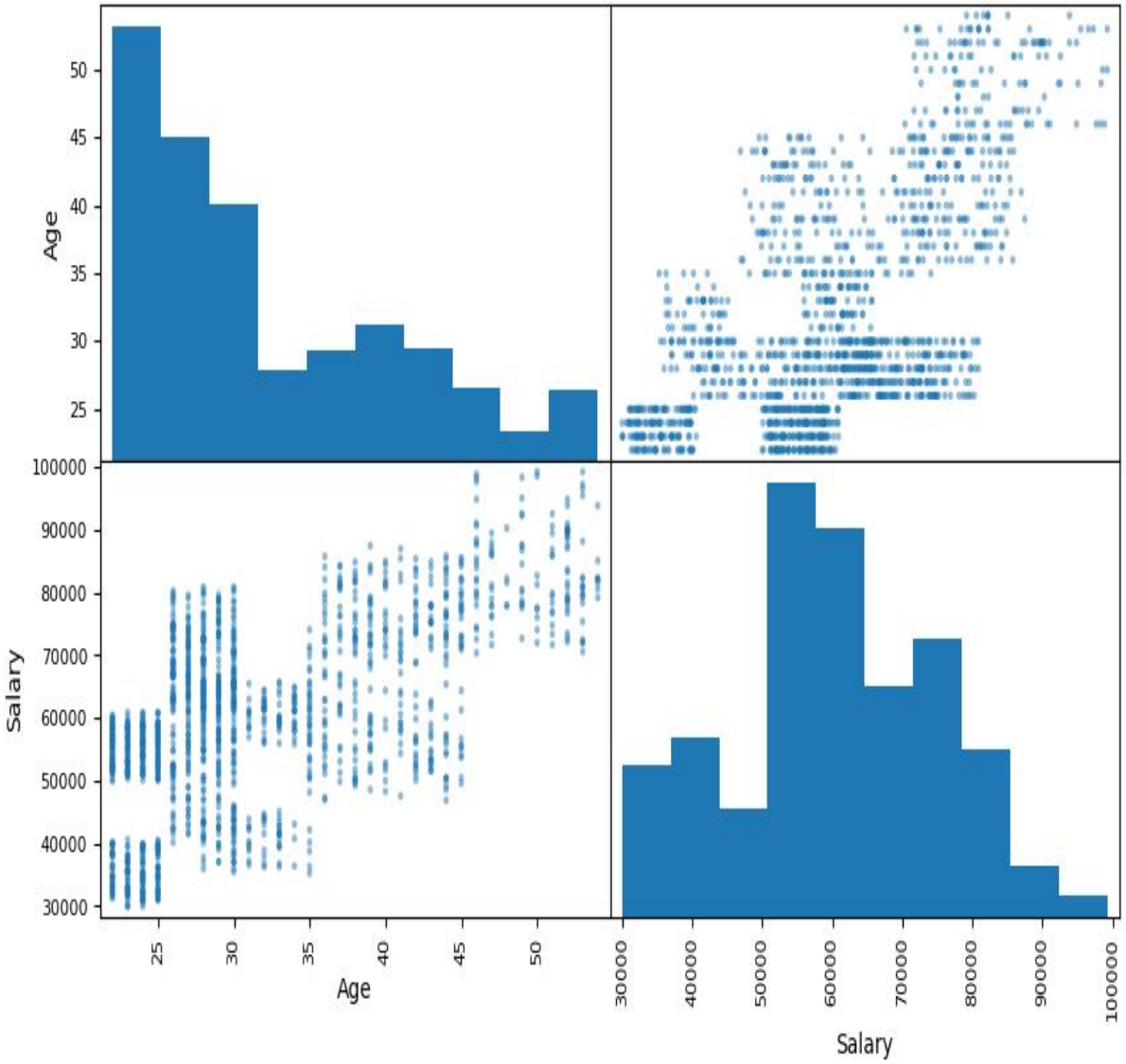
Bivariate analysis

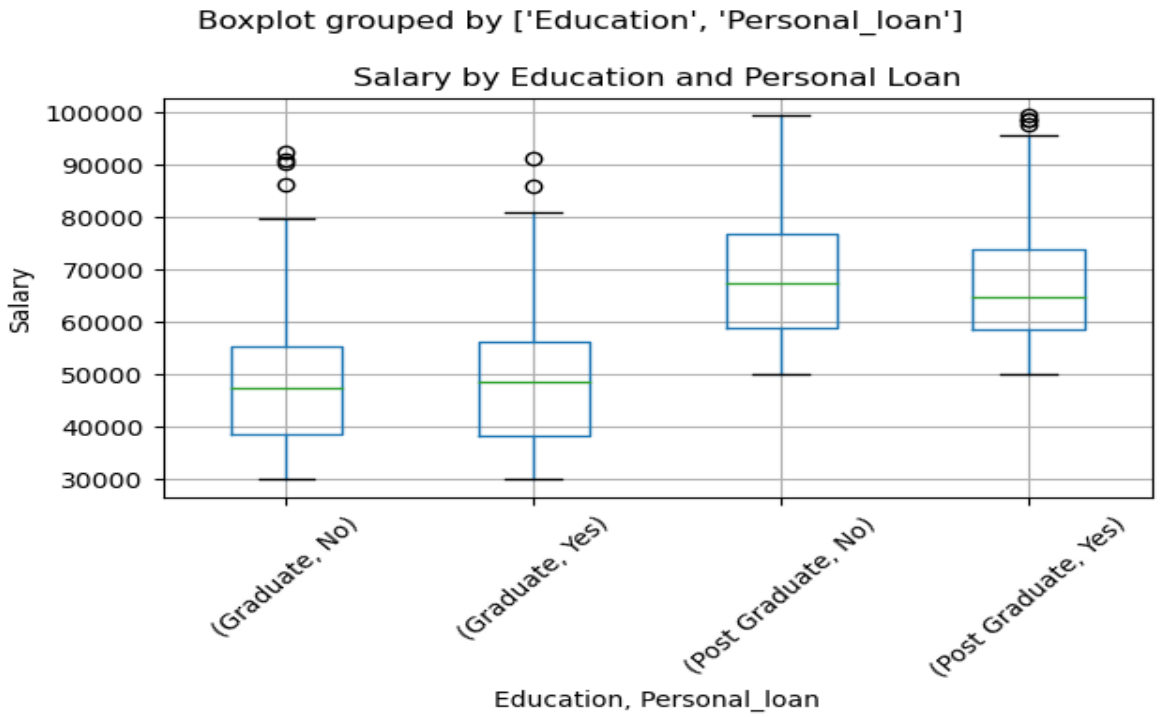


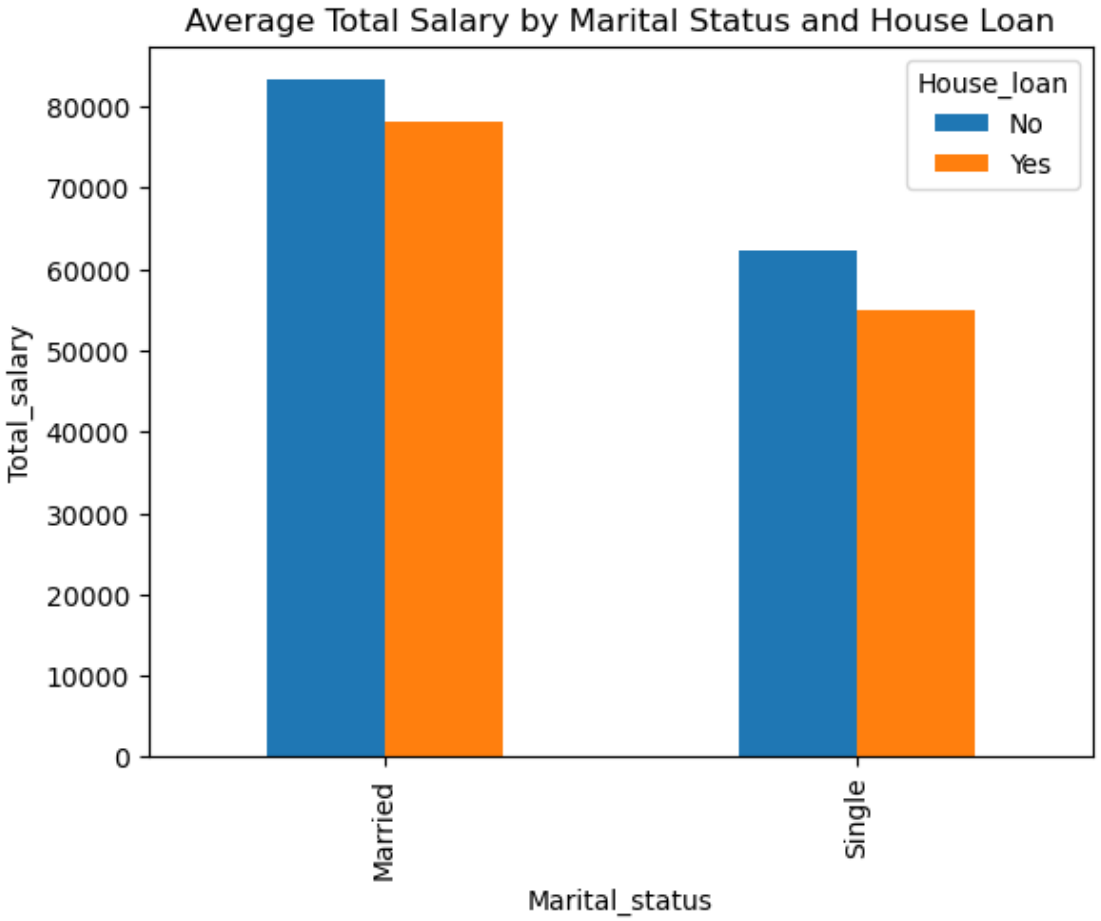


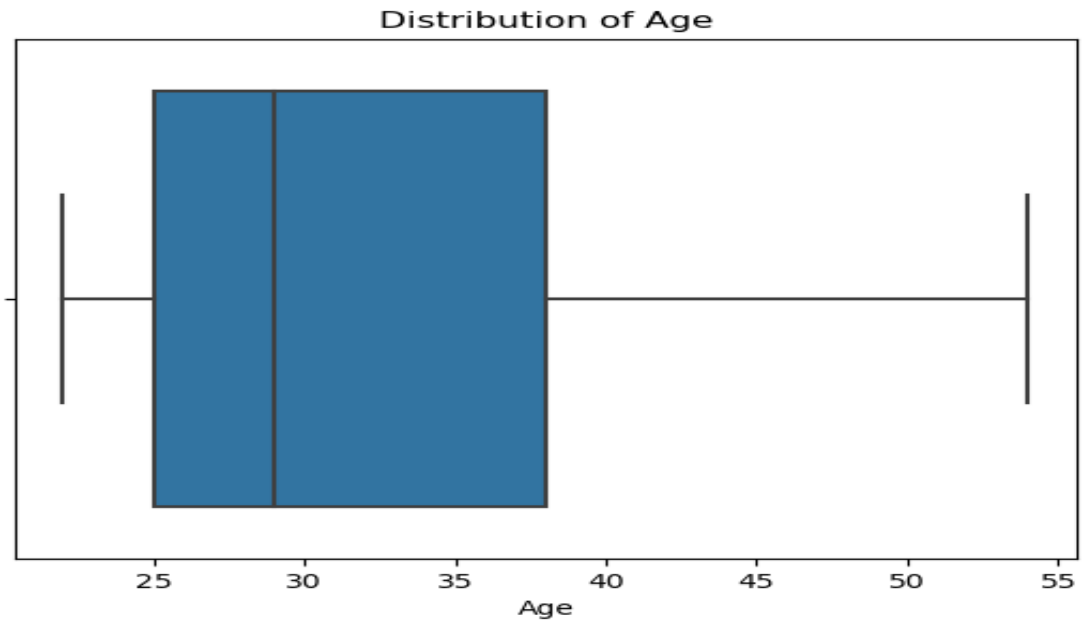


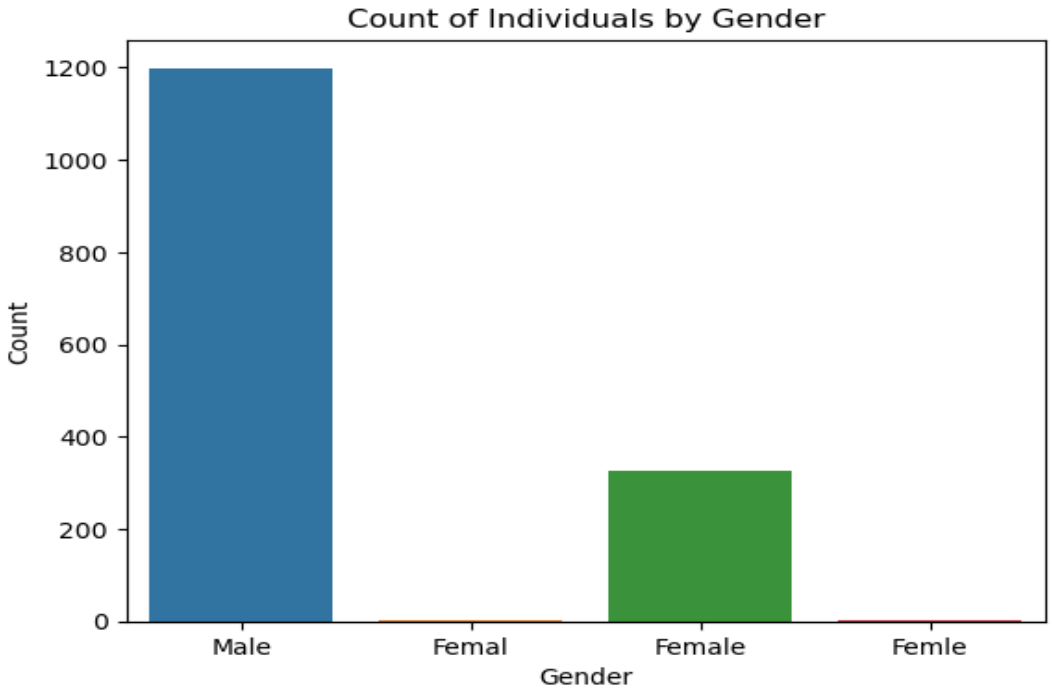
Multivariate analysis

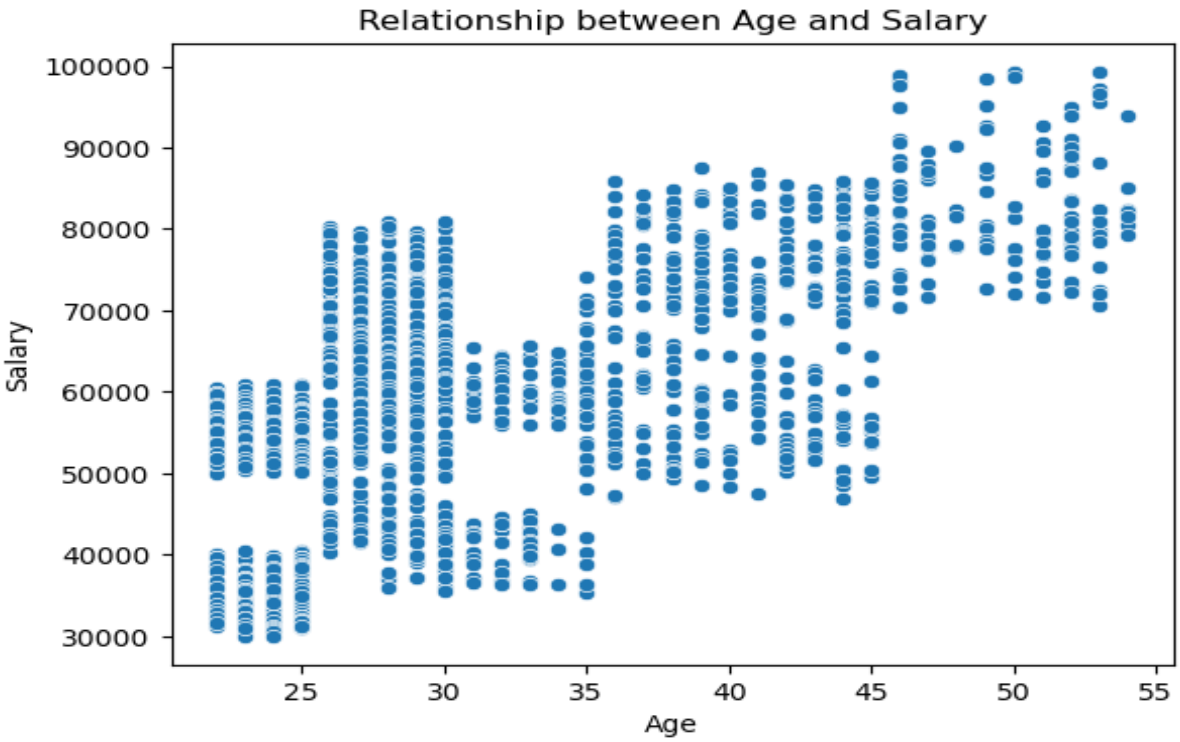


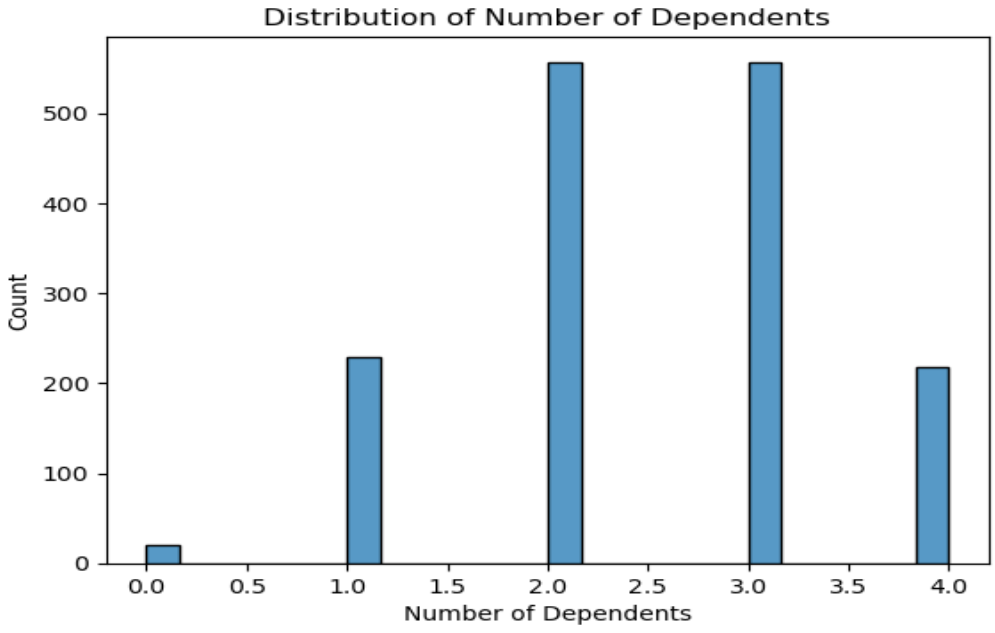

Boxplot grouped by ['Education', 'Personal_loan']
Salary by Education and Personal Loan
Salary
100000
90000
80000
70000
60000
50000
40000
30000
(Graduate, No)
(Graduate, Yes)
(Post Graduate, No)
(Post Graduate, Yes)
Education, Personal_loan

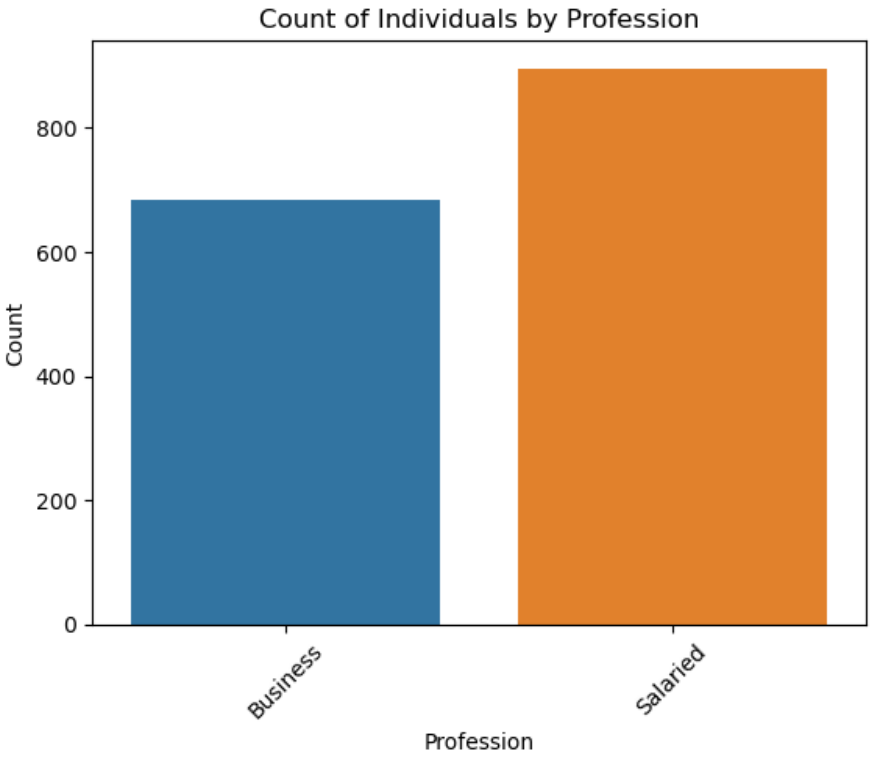

Average Total Salary by Marital Status and House Loan
House_loan
No
Yes
Total_salary
80000
70000
60000
50000
40000
30000
20000
10000
0
Married
Single
Marital_status

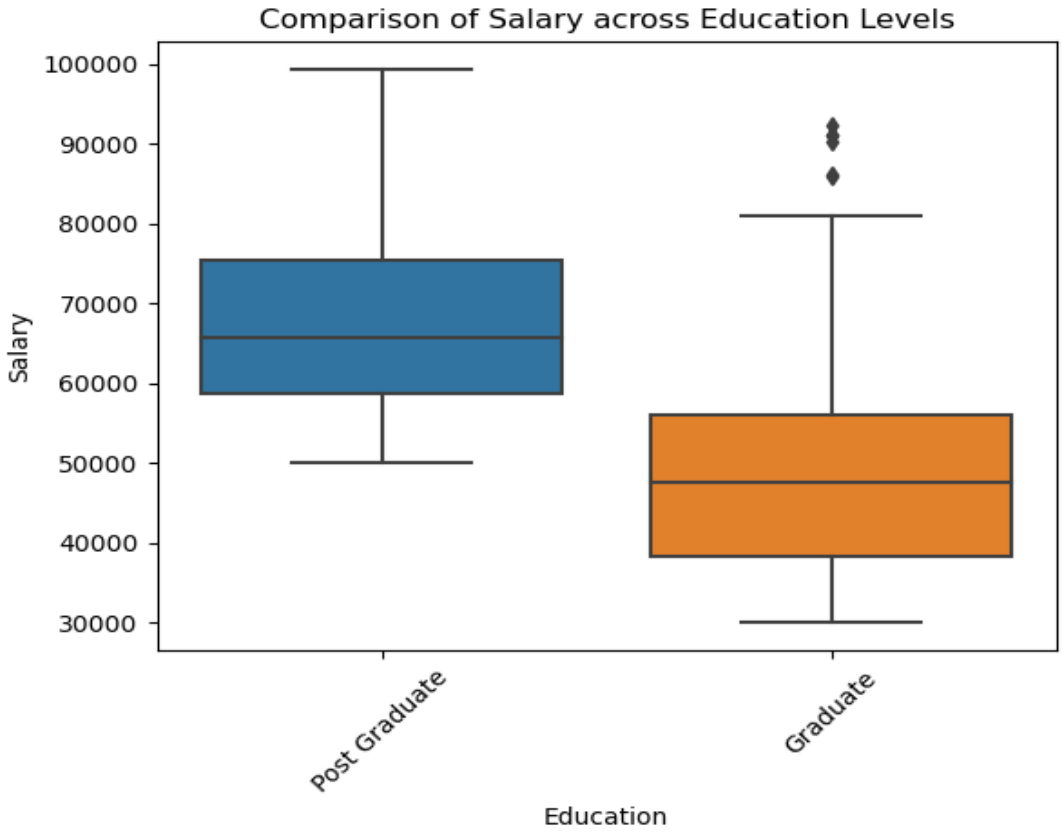

Distribution of Age
25
30
35
40
45
50
55
Age

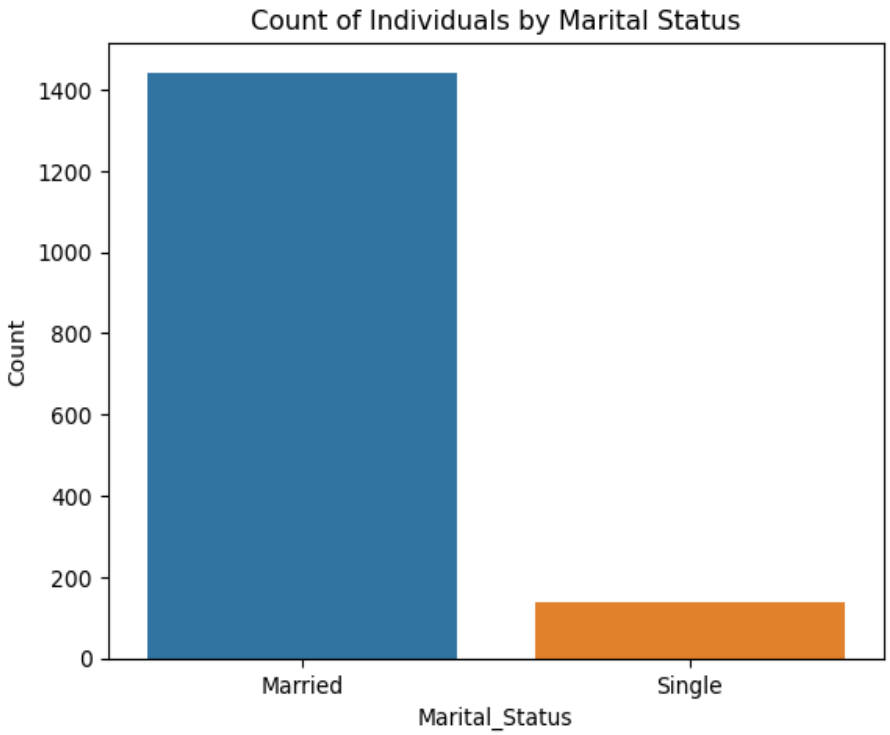

Count of Individuals by Gender
1200
1000
800
600
400
200
0
Count
Male
Femal
Female
Femle
Gender

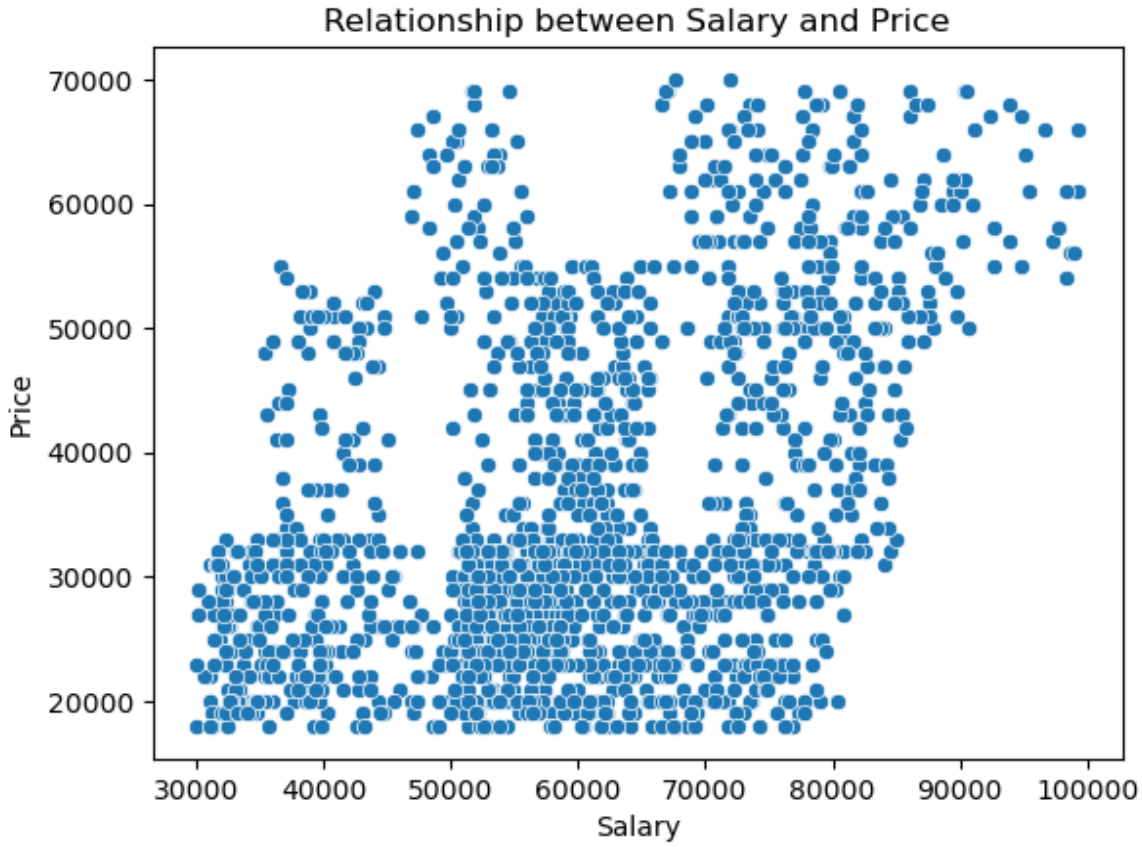

Relationship between Age and Salary
Salary
Age
100000
90000
80000
70000
60000
50000
40000
30000
25
30
35
40
45
50
55

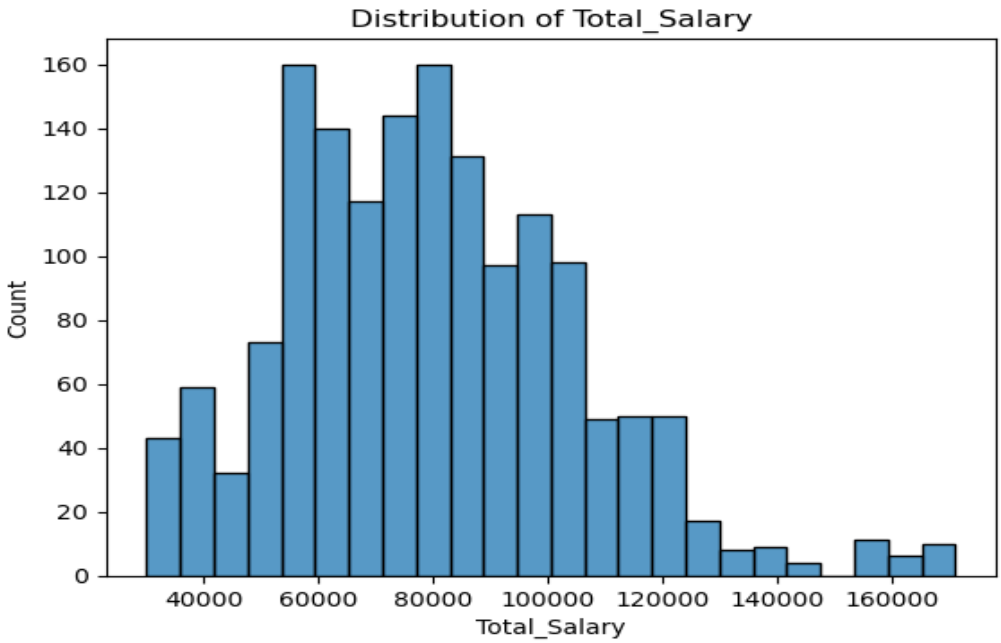

Distribution of Number of Dependents
Count
Number of Dependents
500
400
300
200
100
0
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0

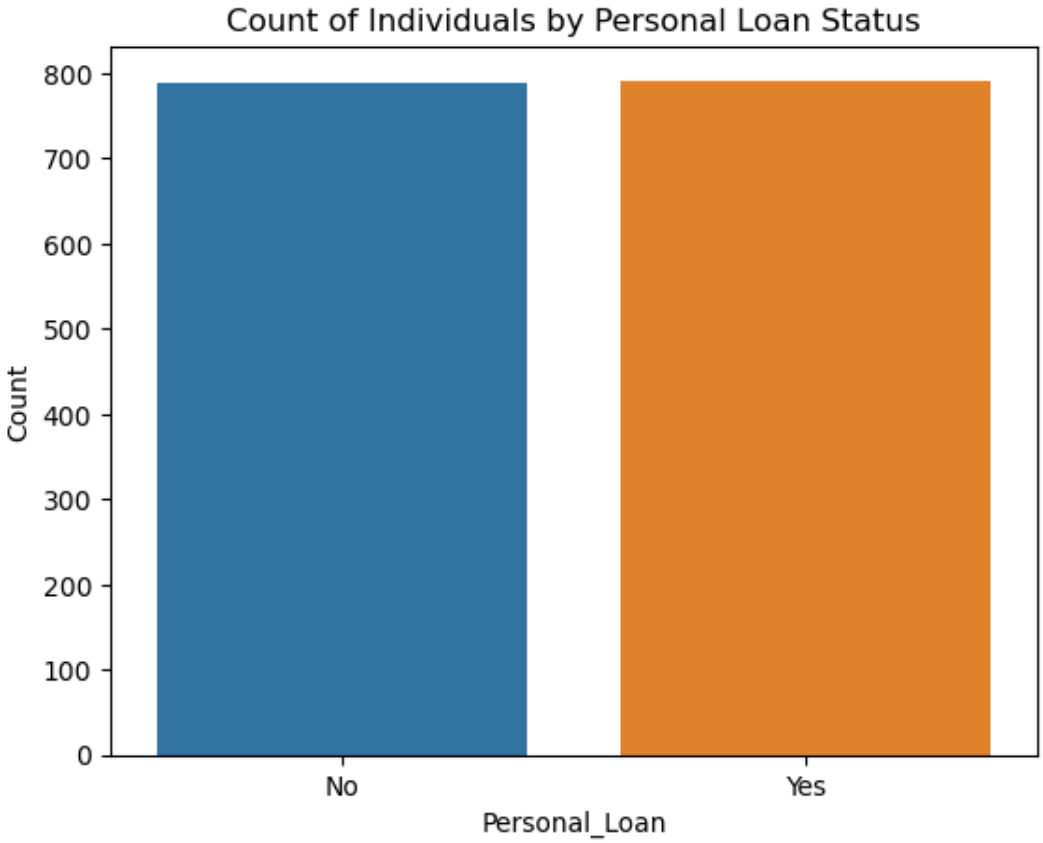

Count of Individuals by Profession
800
600
400
200
0
Count
Business
Salaried
Profession

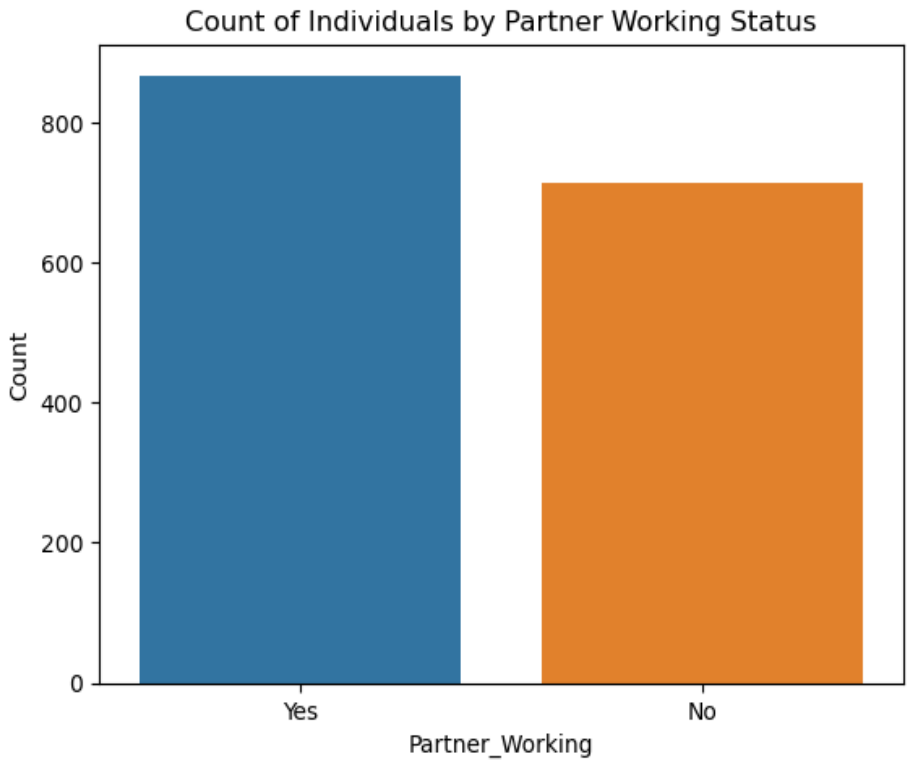

Comparison of Salary across Education Levels
100000
90000
80000
70000
60000
50000
40000
30000
Salary
Post Graduate
Graduate
Education

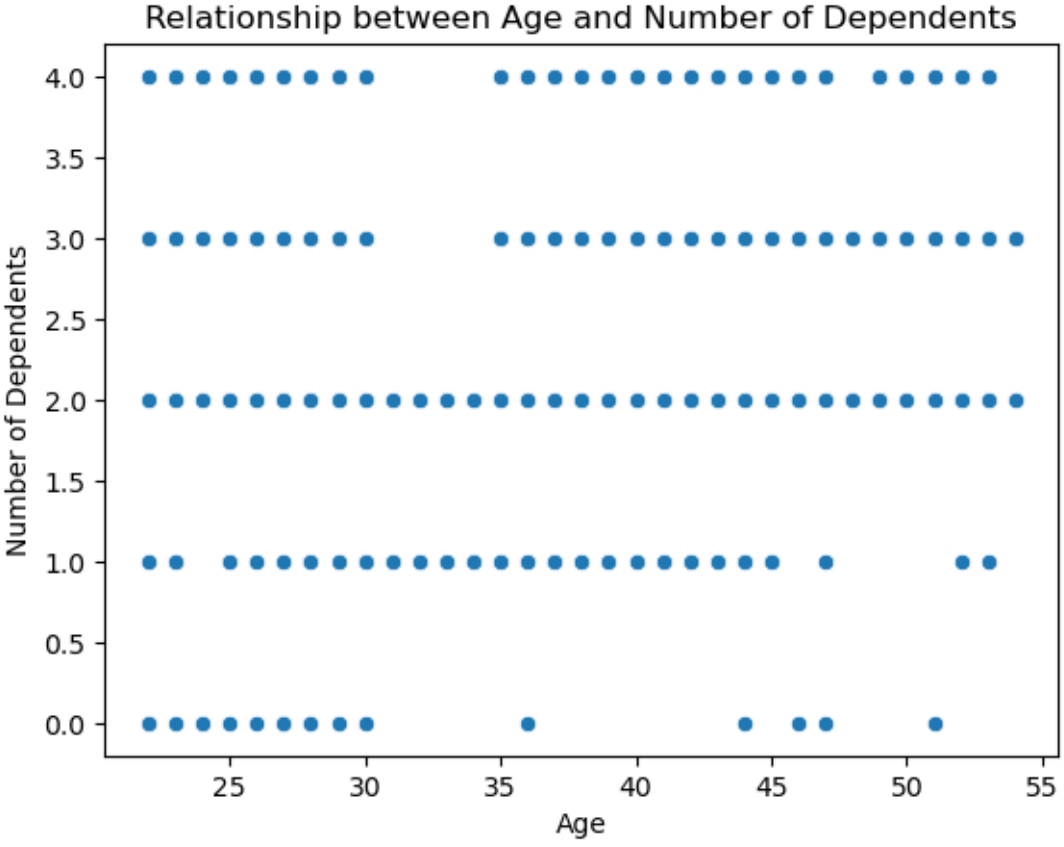

Count of Individuals by Marital Status
1400
1200
1000
800
600
400
200
0
Count
Married
Single
Marital_Status

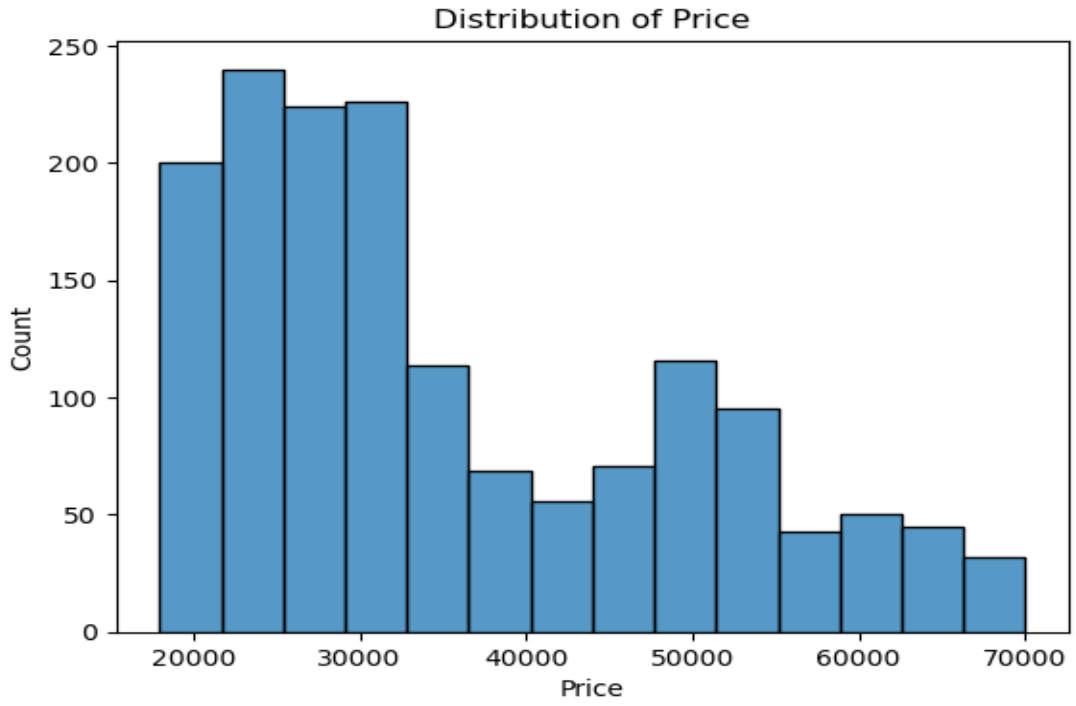

Relationship between Salary and Price
70000
60000
50000
40000
30000
20000
Price
30000
40000
50000
60000
70000
80000
90000
100000
Salary

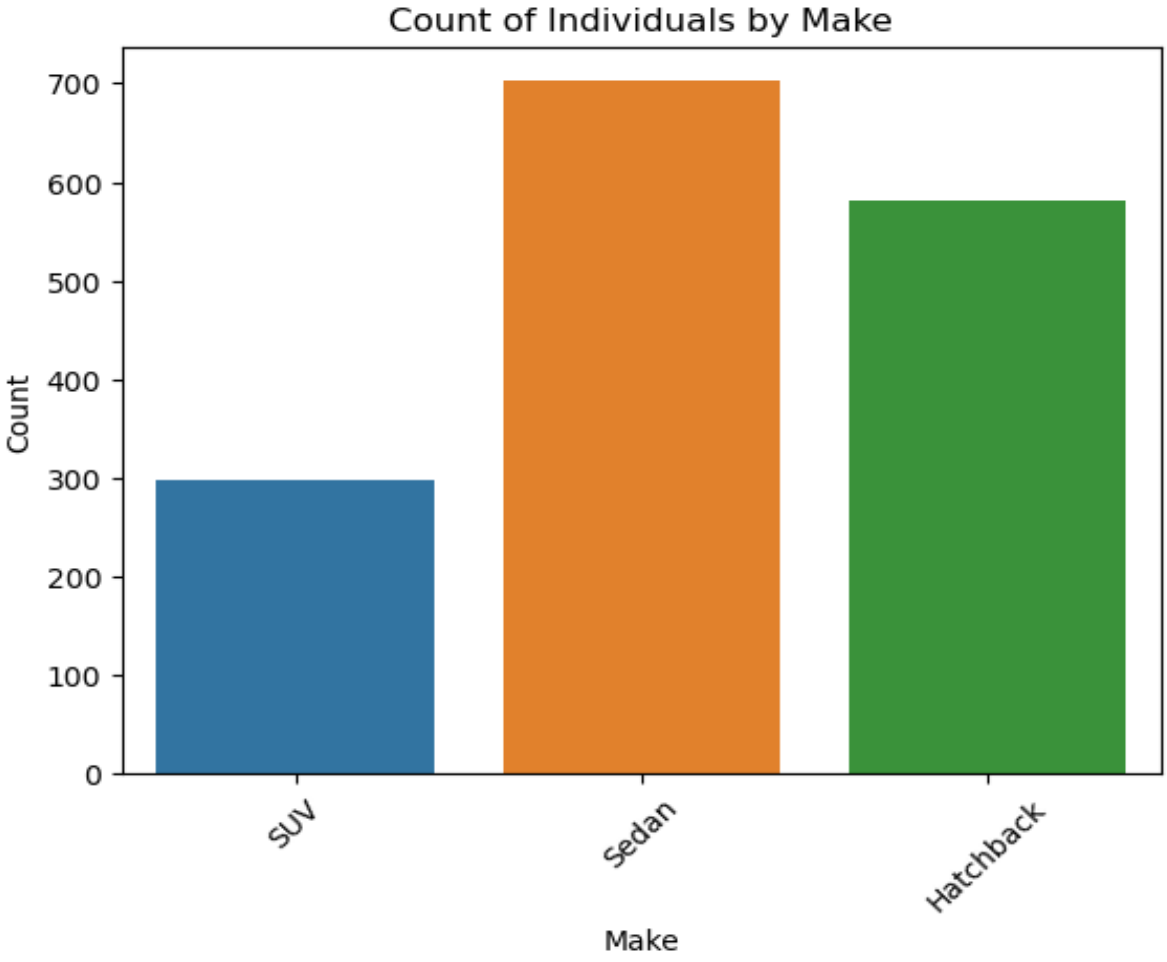

Distribution of Total_Salary
Count
Total_Salary
0
20
40
60
80
100
120
140
160
40000
60000
80000
100000
120000
140000
160000

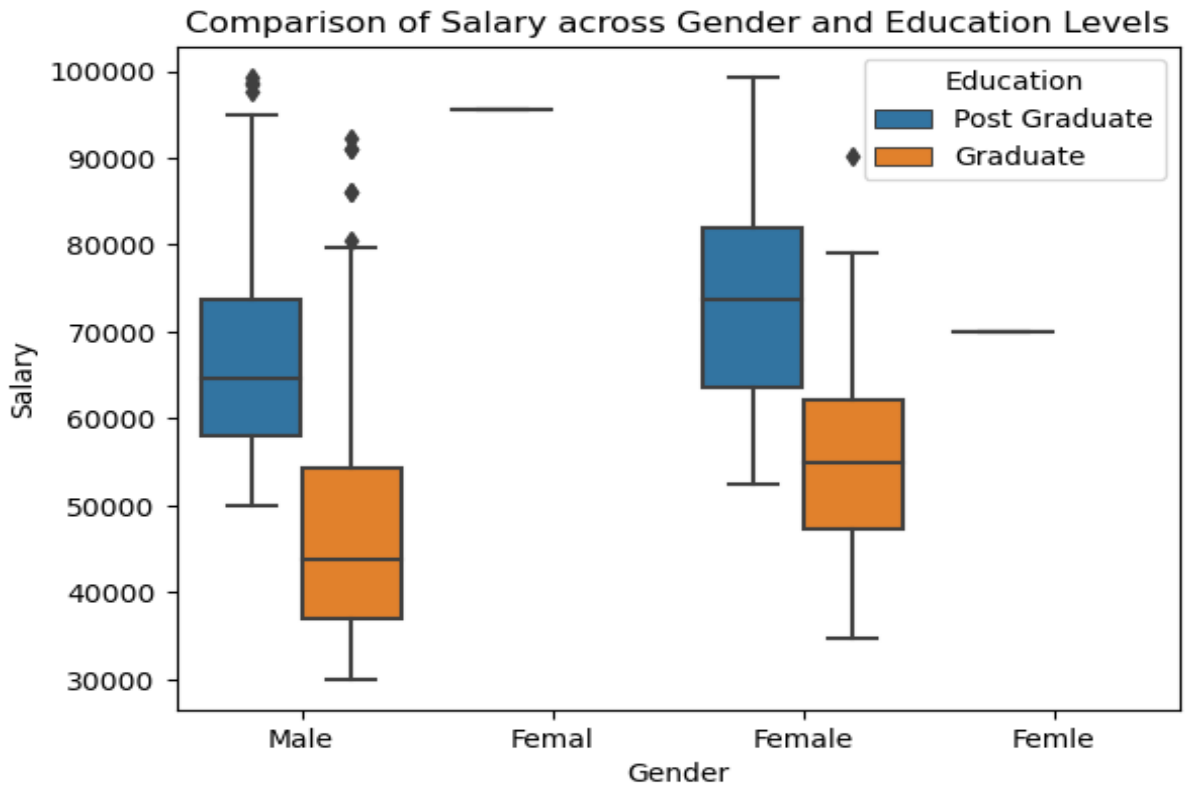

Count of Individuals by Personal Loan Status
Count
Personal_Loan
0
100
200
300
400
500
600
700
800
No
Yes

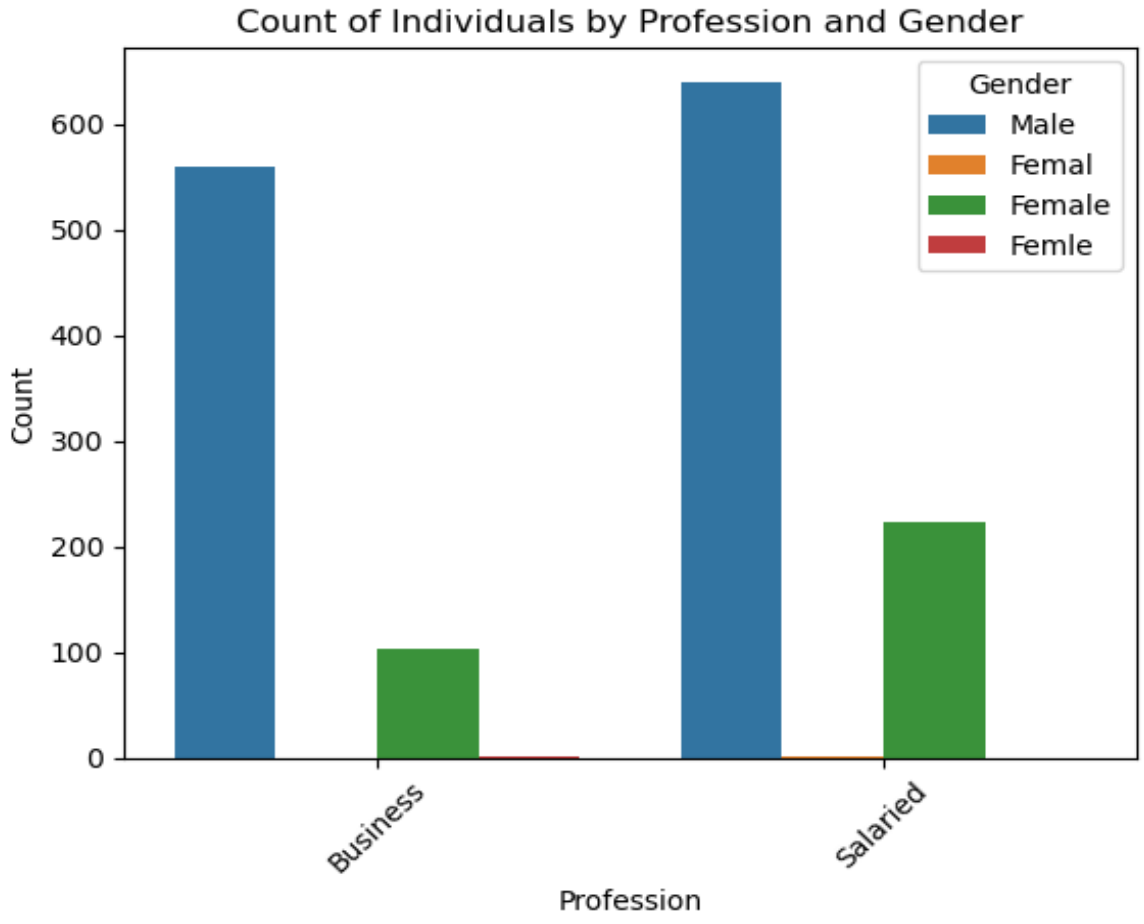

Count of Individuals by Partner Working Status
Count
800
600
400
200
0
Yes
No
Partner_Working

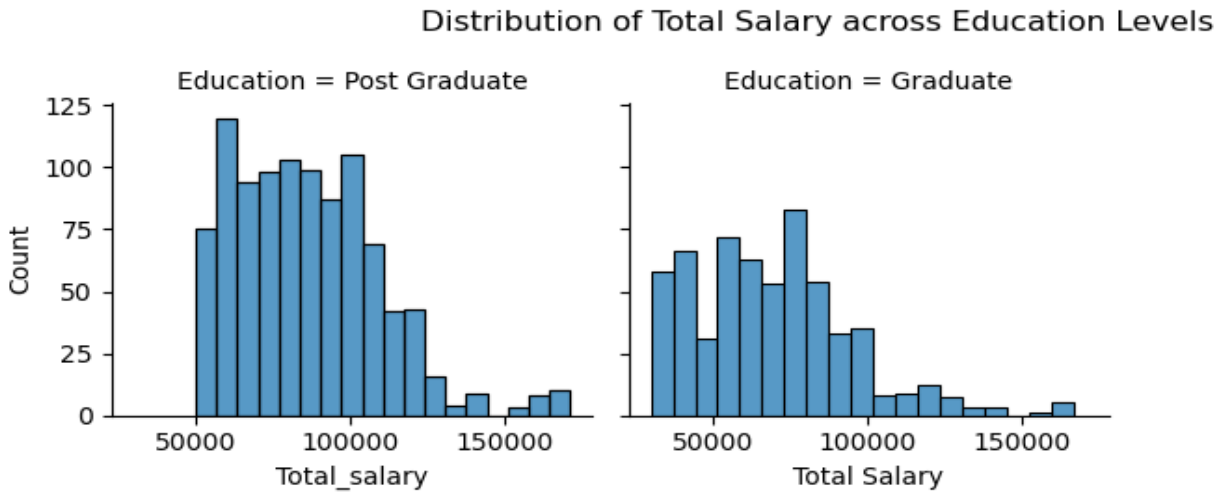

Relationship between Age and Number of Dependents
Number of Dependents
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
25
30
35
40
45
50
55
Age

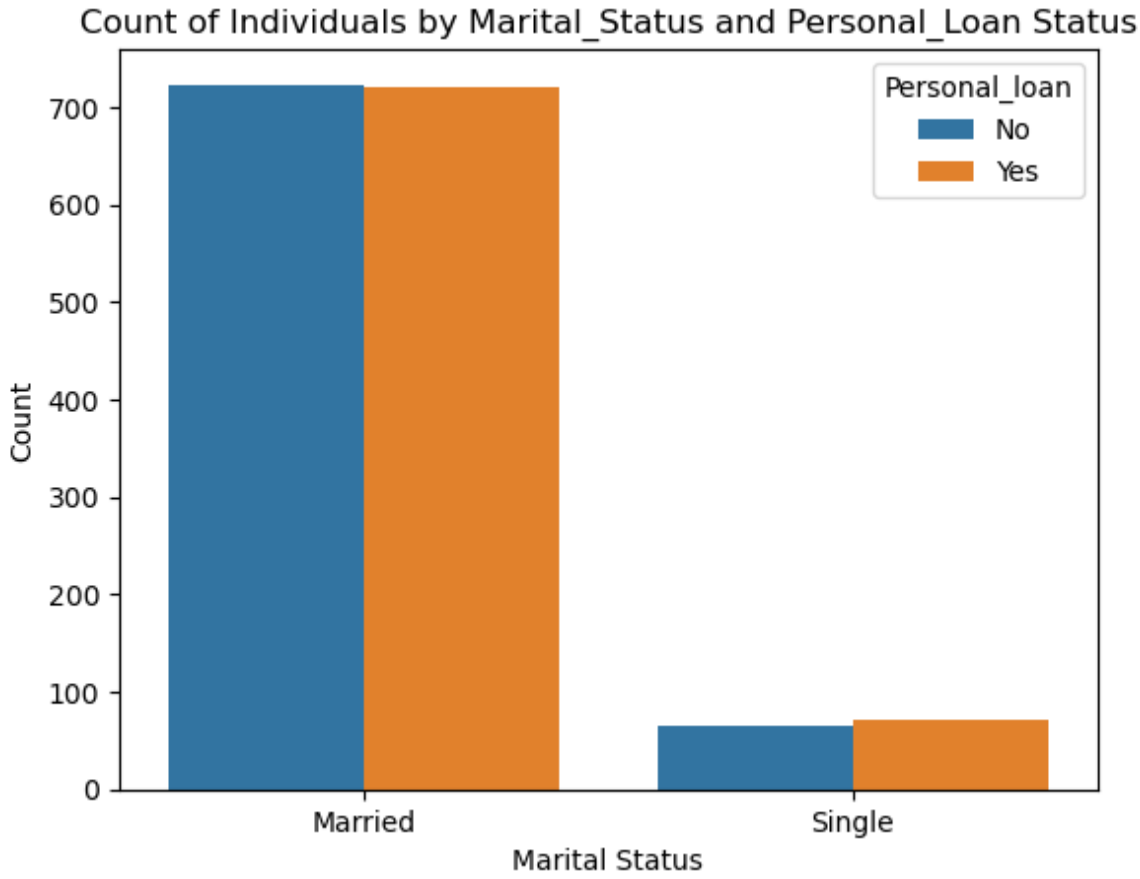

Distribution of Price
250
200
150
100
50
0
Count
20000
30000
40000
50000
60000
70000
Price

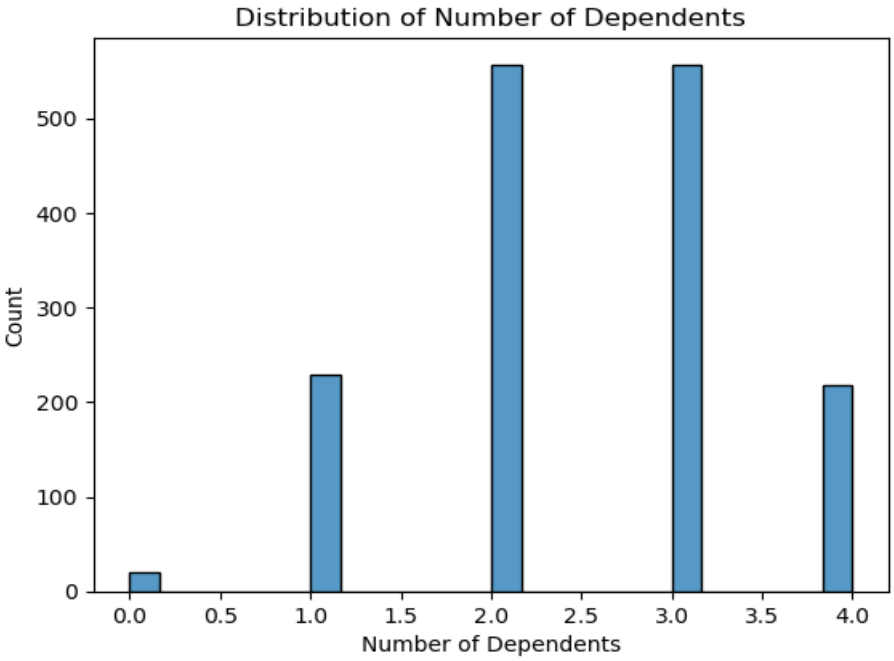

Count of Individuals by Make
700
600
500
400
300
200
100
0
Count
SUV
Sedan
Hatchback
Make

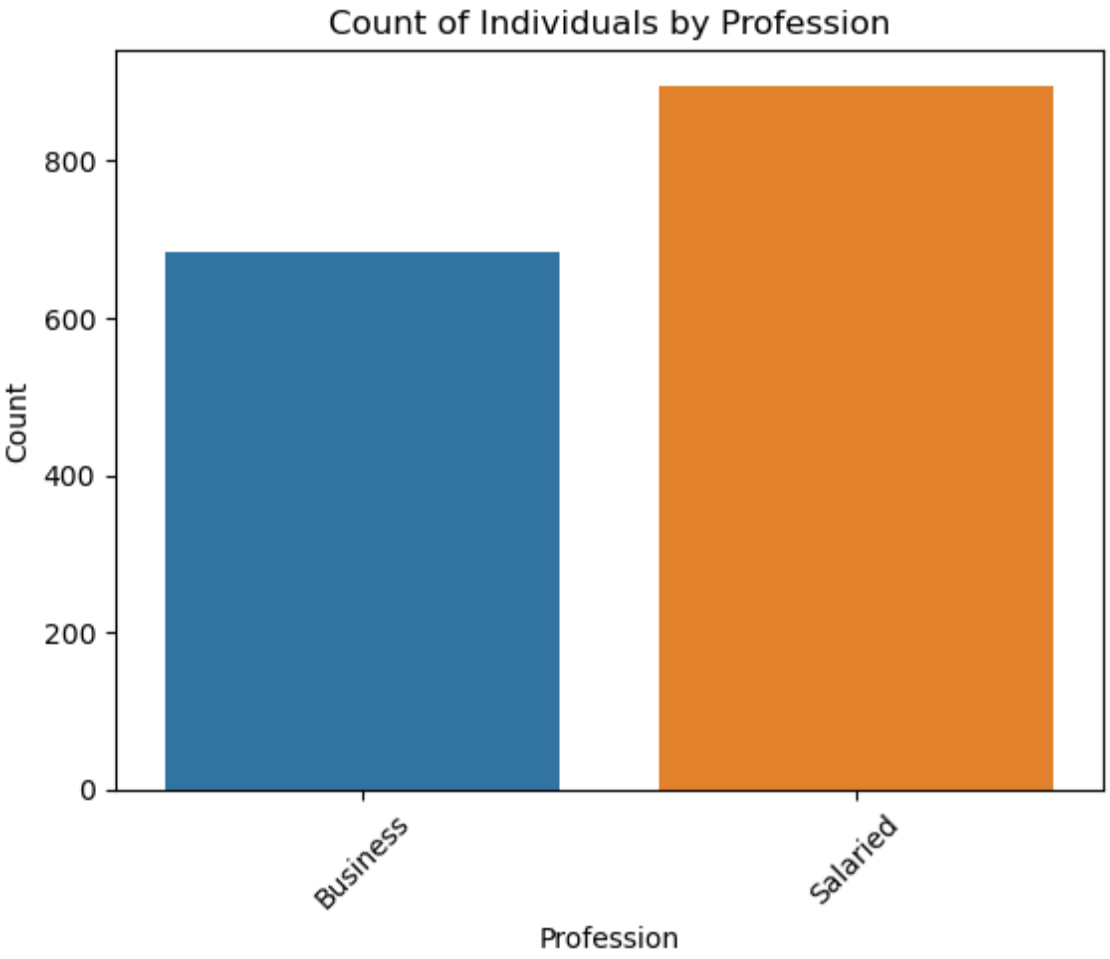

Comparison of Salary across Gender and Education Levels
Education
Post Graduate
Graduate
Salary
100000
90000
80000
70000
60000
50000
40000
30000
Male
Femal
Female
Femle
Gender

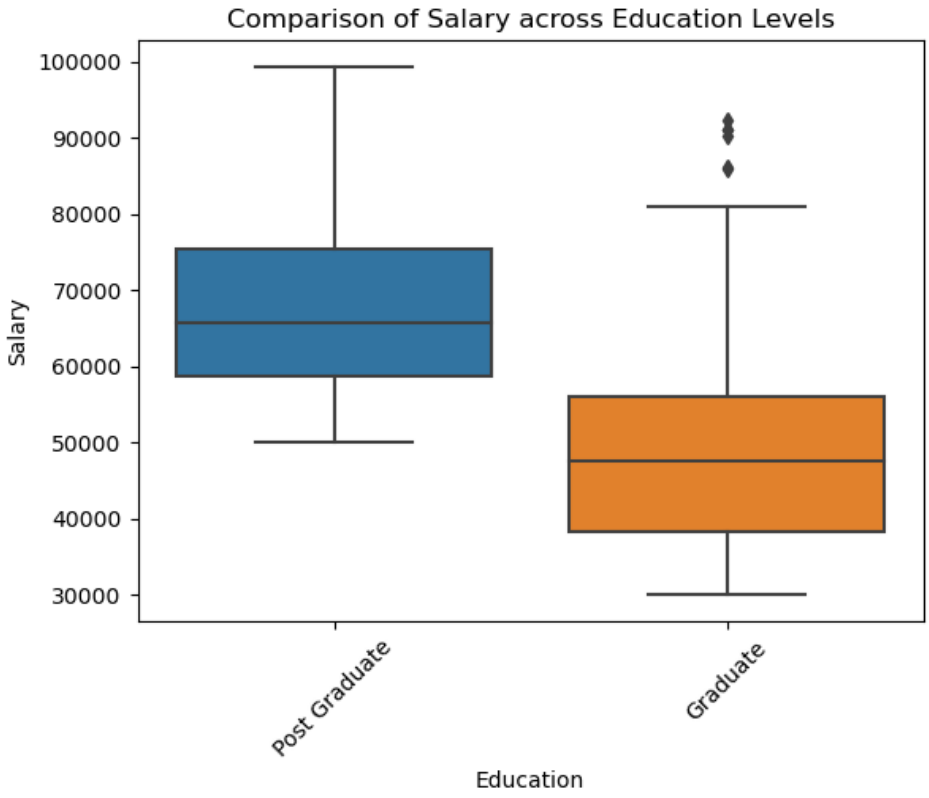

Count of Individuals by Profession and Gender
Gender
Male
Femal
Female
Femle
Count
600
500
400
300
200
100
0
Business
Salaried
Profession

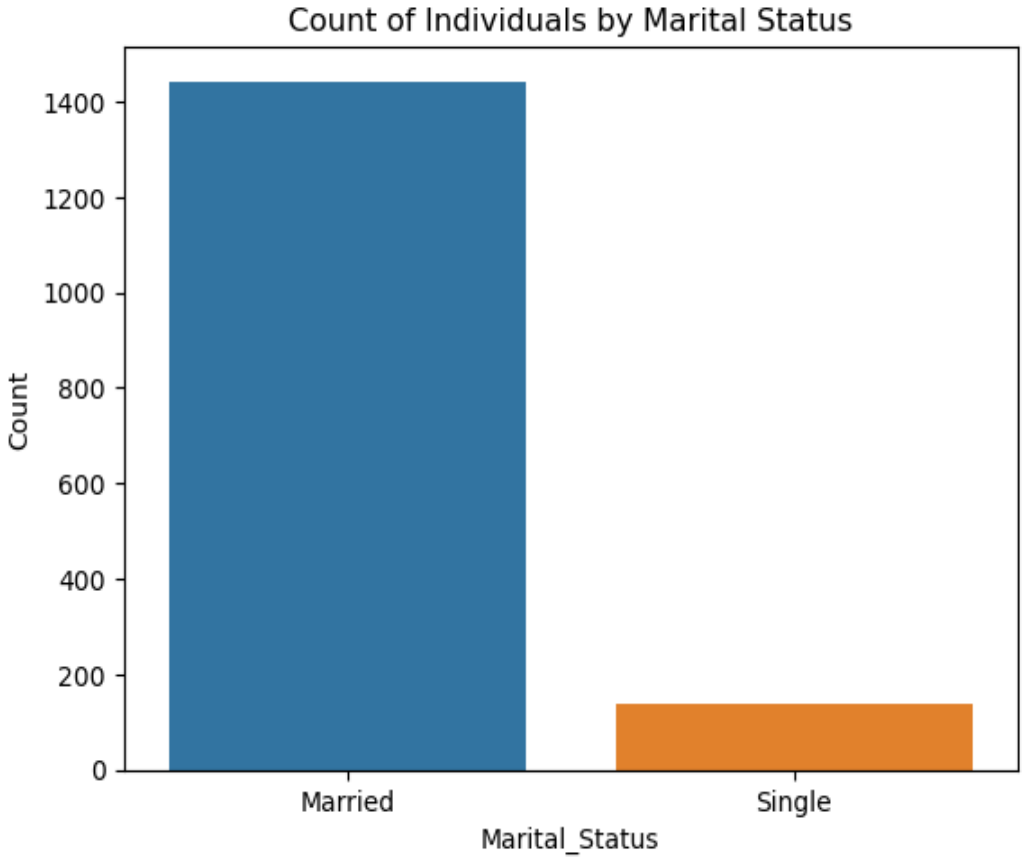

Distribution of Total Salary across Education Levels
Education = Post Graduate
Education = Graduate
Count
125
100
75
50
25
0
50000
100000
150000
Total_salary
Total Salary

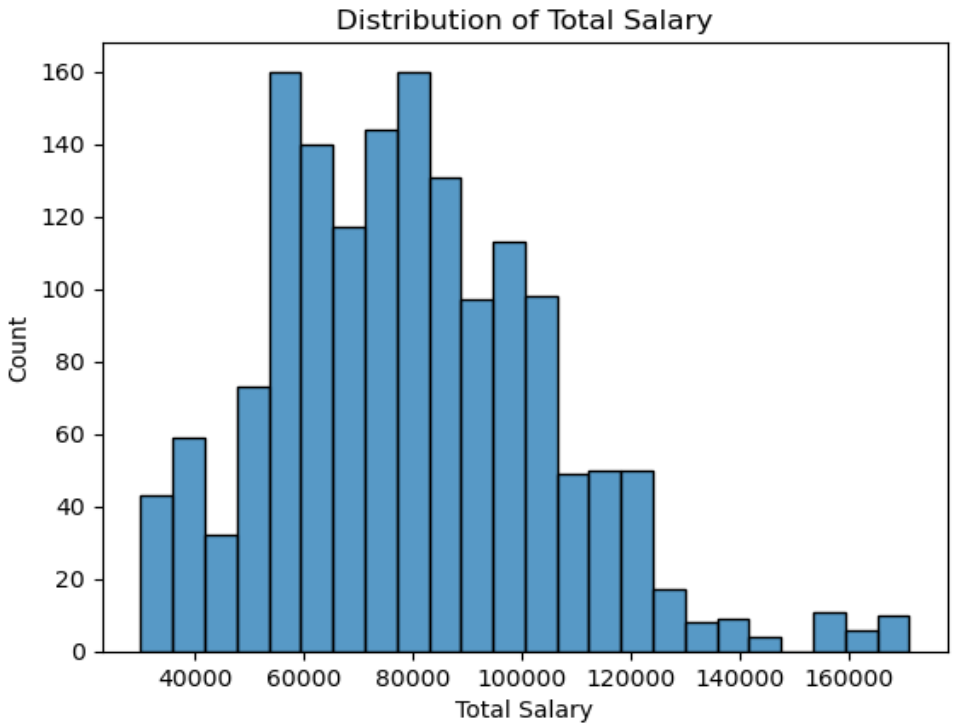

Count of Individuals by Marital_Status and Personal_Loan Status
Personal_loan
No
Yes
Count
700
600
500
400
300
200
100
0
Married
Single
Marital Status

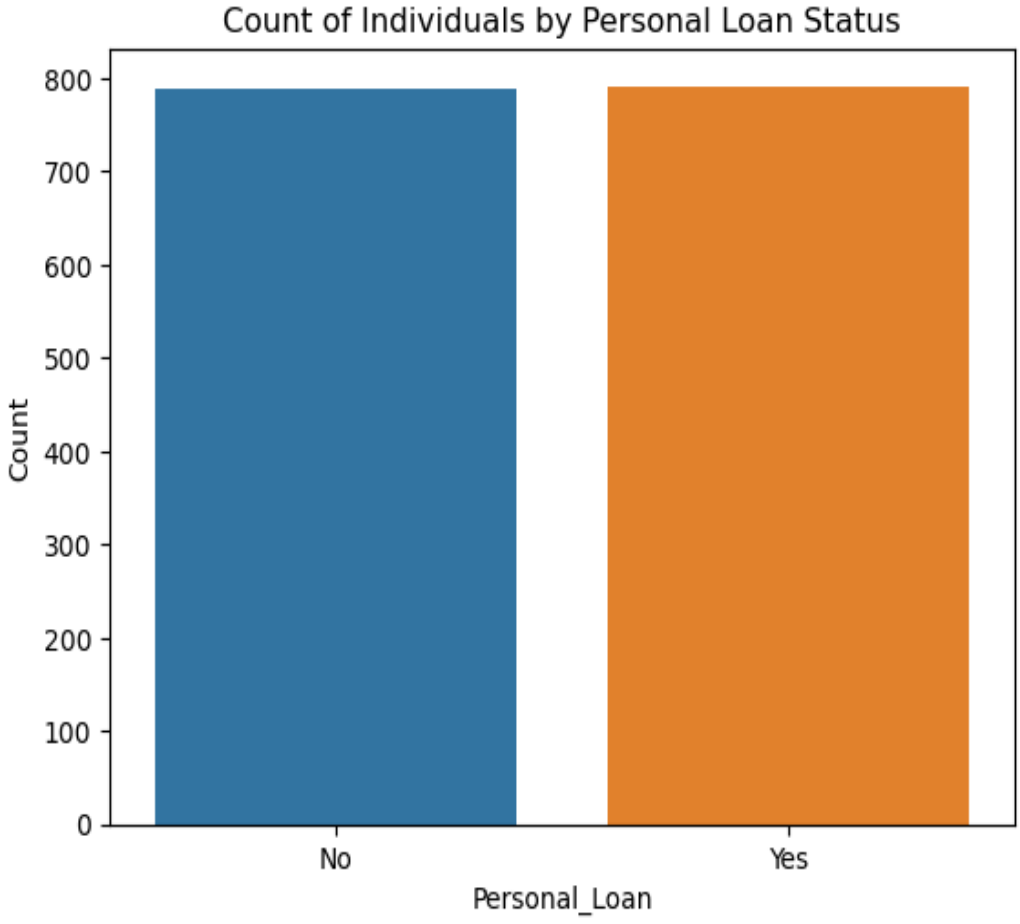

Distribution of Number of Dependents
Count
0
100
200
300
400
500
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
Number of Dependents

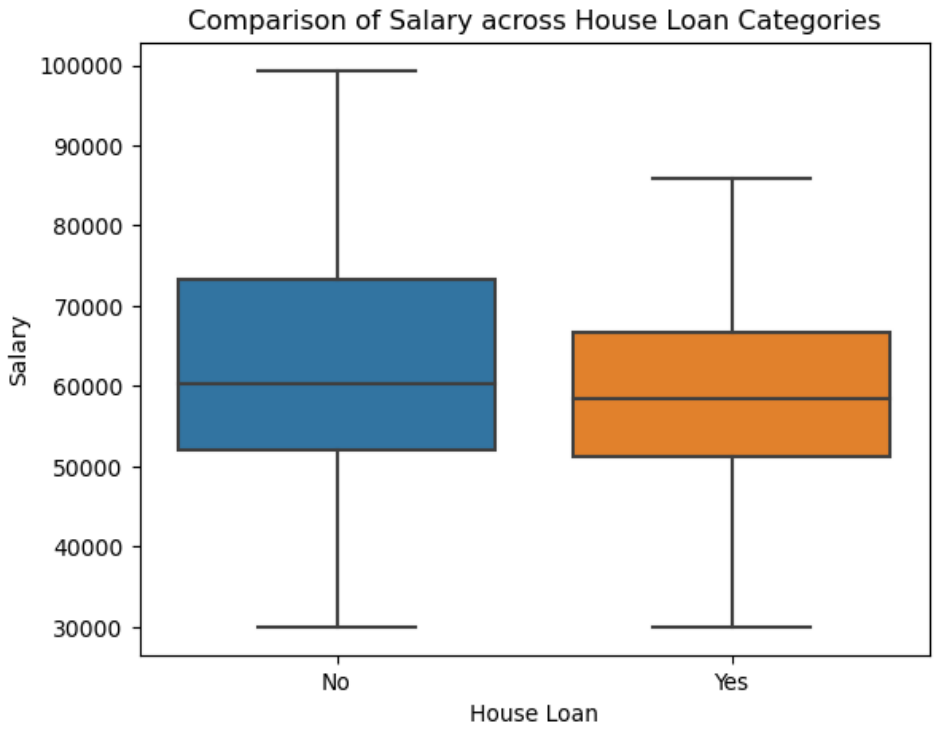

Count of Individuals by Profession
Count
0
200
400
600
800
Business
Salaried
Profession

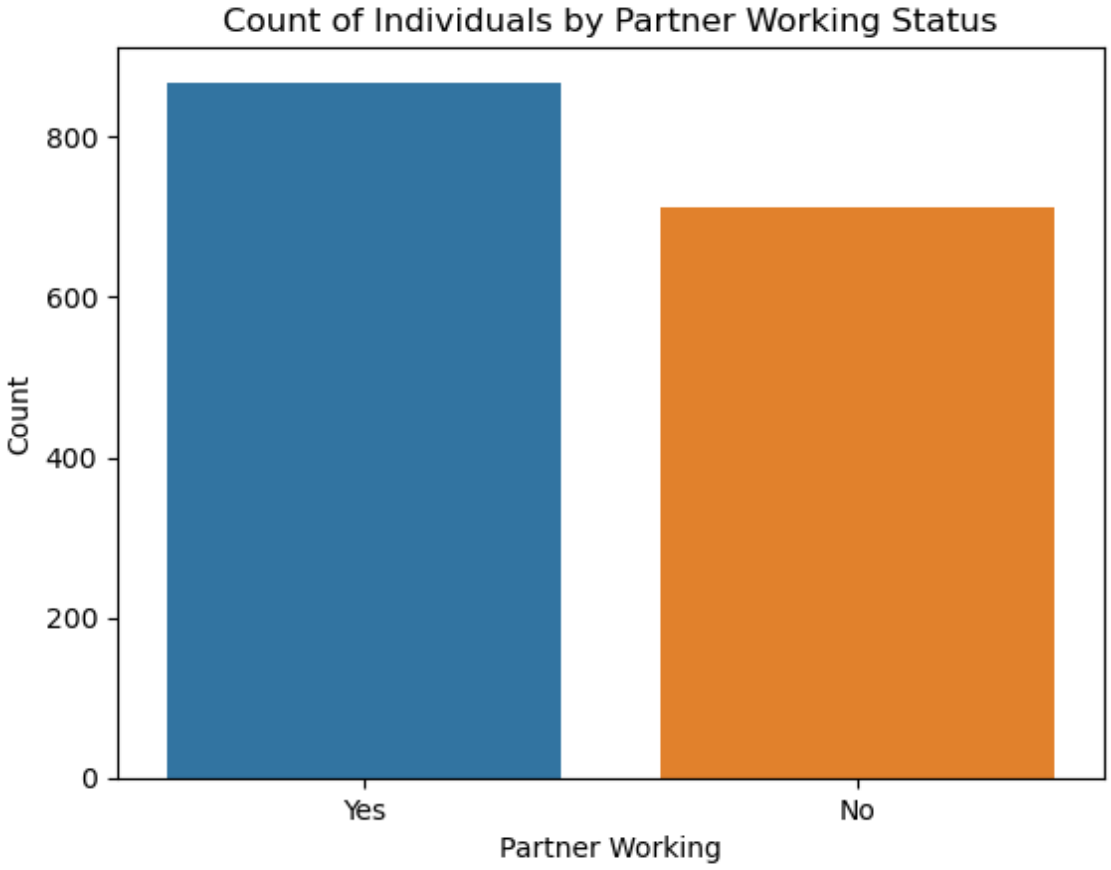

Comparison of Salary across Education Levels
Salary
100000
90000
80000
70000
60000
50000
40000
30000
Post Graduate
Graduate
Education

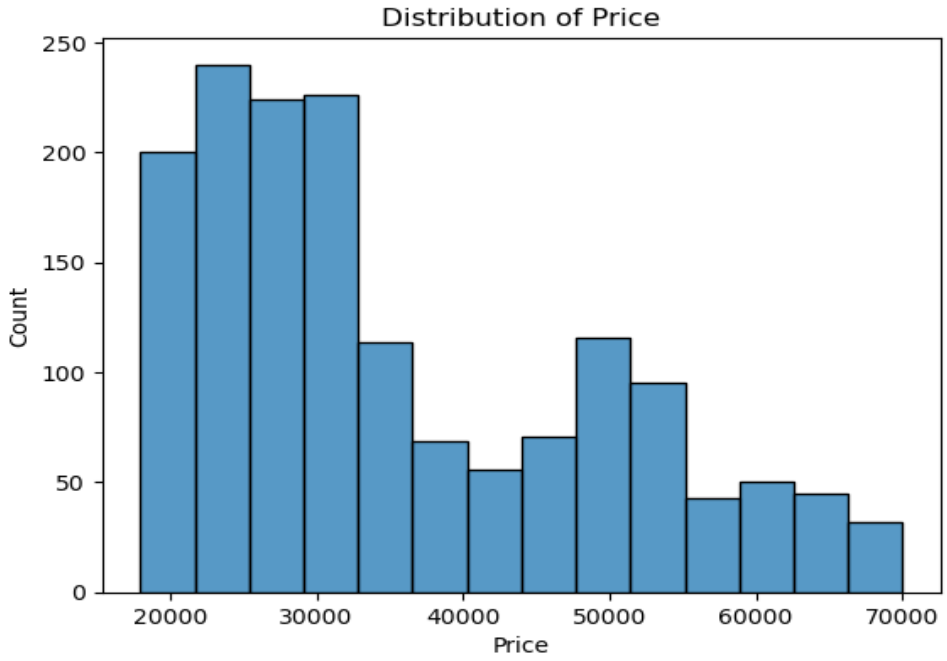

Count of Individuals by Marital Status
Count
1400
1200
1000
800
600
400
200
0
Married
Single
Marital_Status

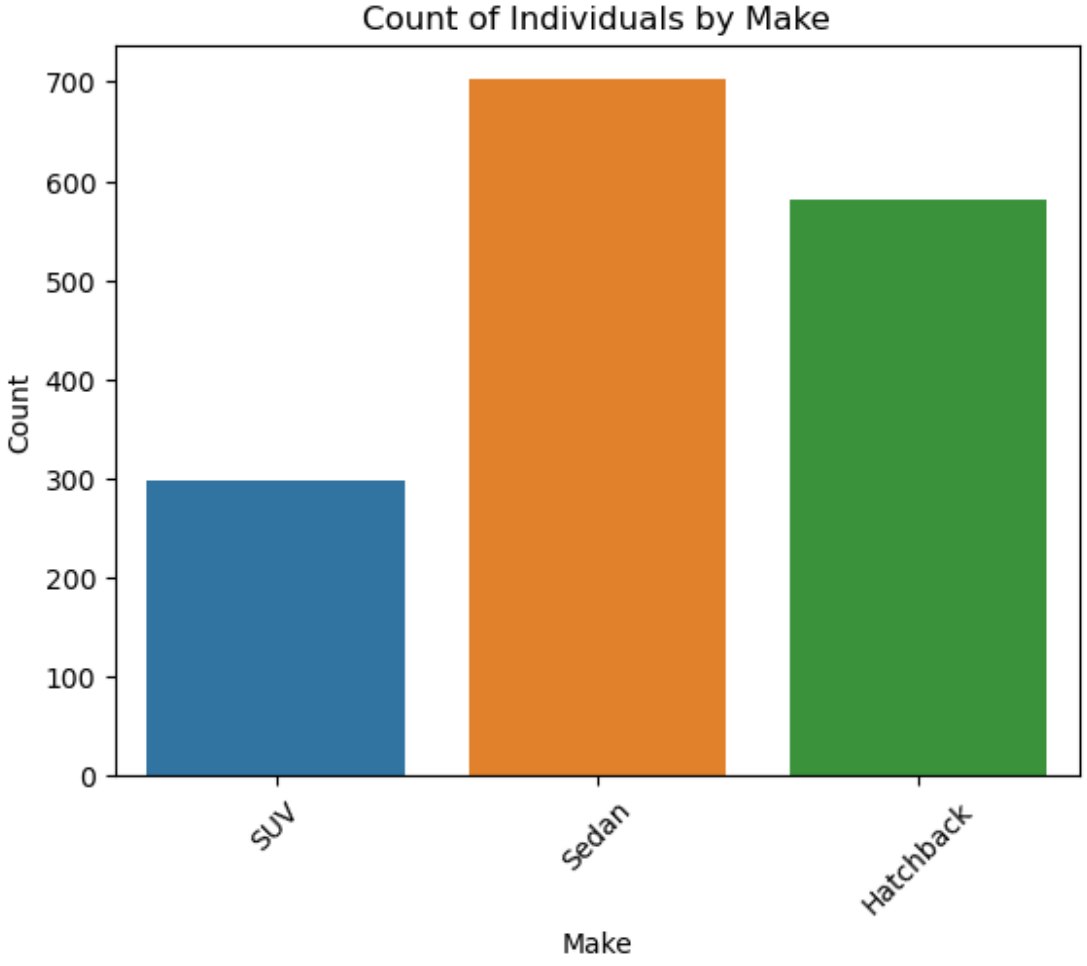

Distribution of Total Salary
Count
Total Salary
0
20
40
60
80
100
120
140
160
40000
60000
80000
100000
120000
140000
160000

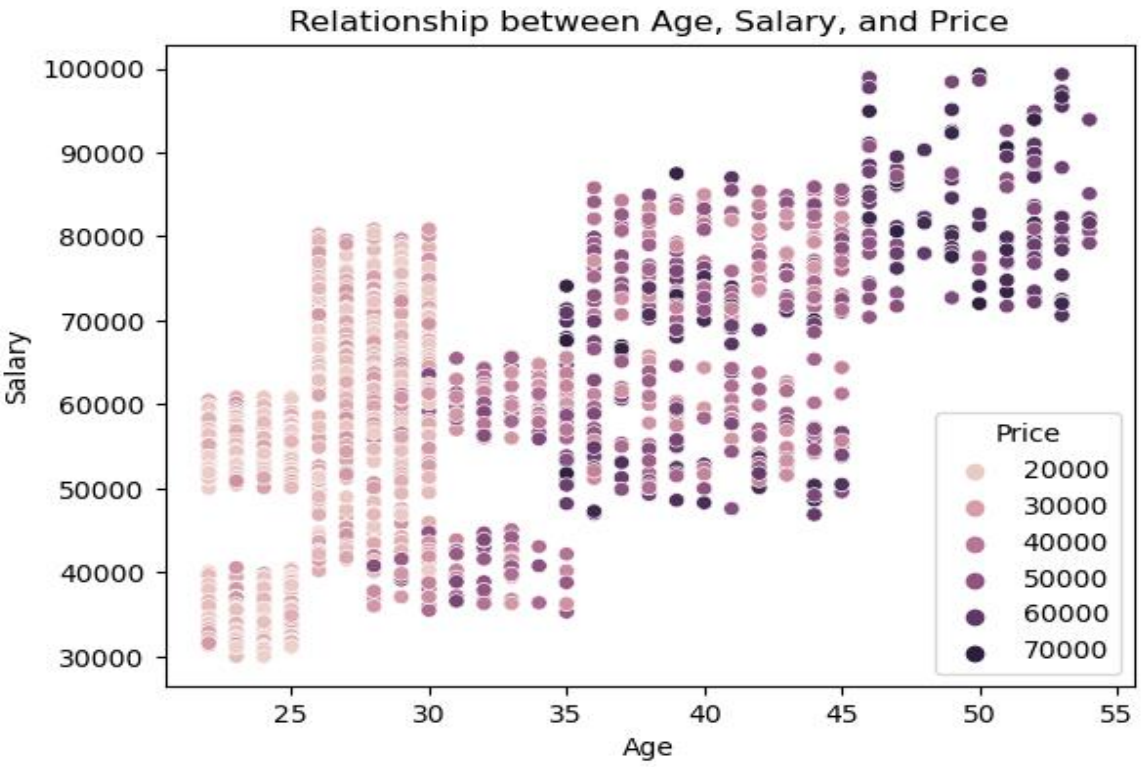

Count of Individuals by Personal Loan Status
Count
Personal_Loan
0
100
200
300
400
500
600
700
800
No
Yes

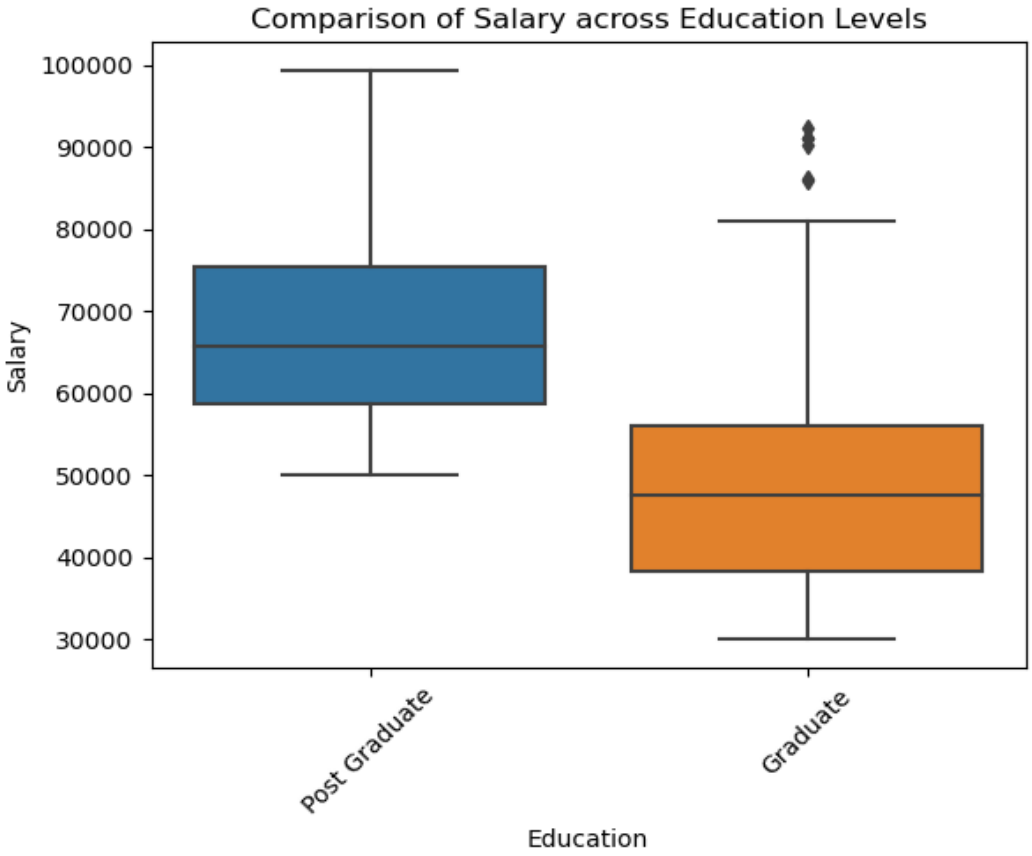

Comparison of Salary across House Loan Categories
100000
90000
80000
70000
60000
50000
40000
30000
Salary
No
Yes
House Loan

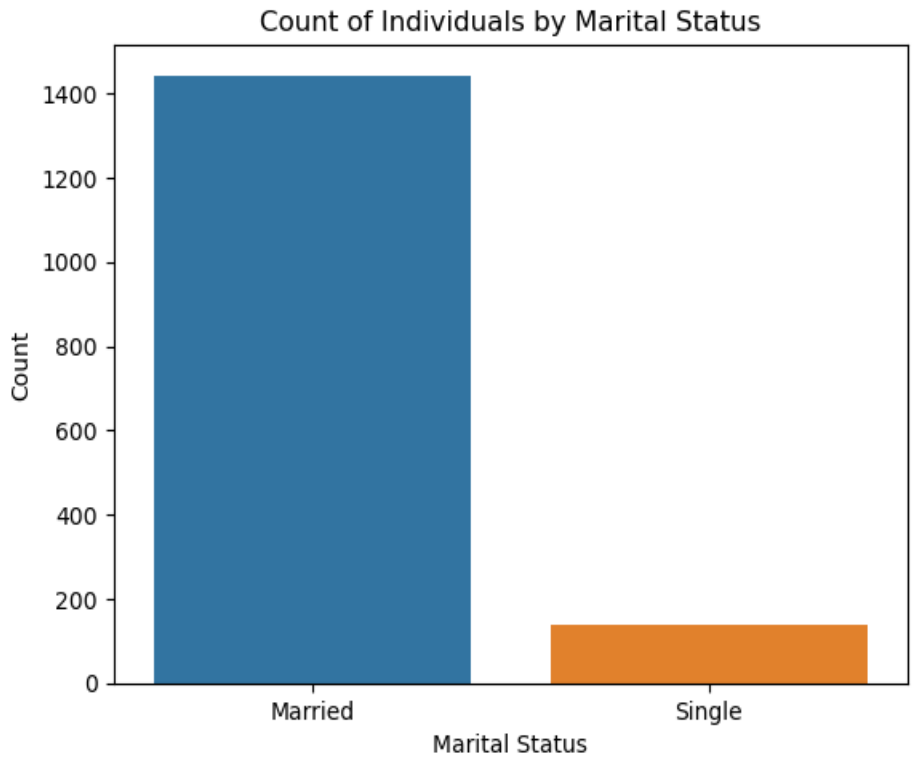

Count of Individuals by Partner Working Status
800
600
400
200
0
Count
Yes
No
Partner Working

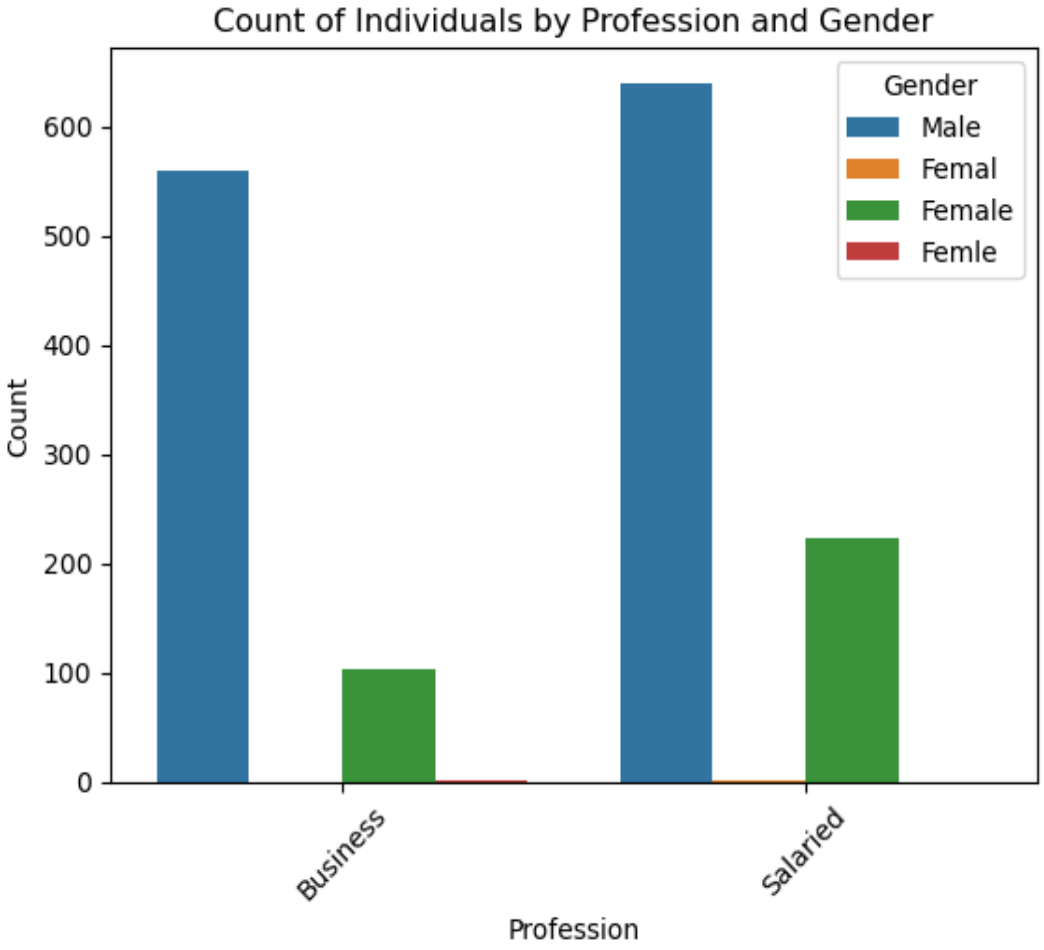

Distribution of Price
250
200
150
100
50
0
Count
20000
30000
40000
50000
60000
70000
Price

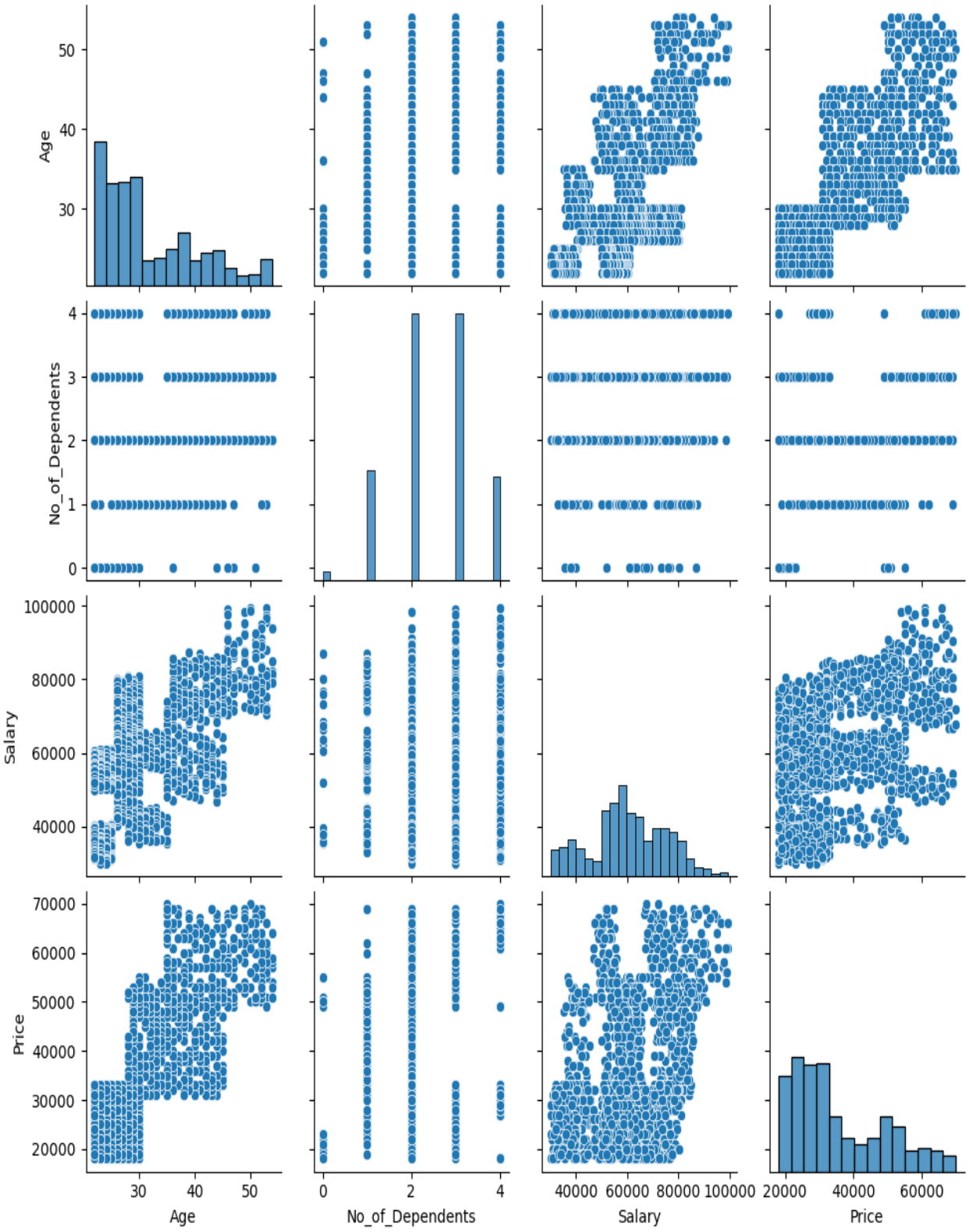

Count of Individuals by Make
700
600
500
400
300
200
100
0
Count
SUV
Sedan
Hatchback
Make

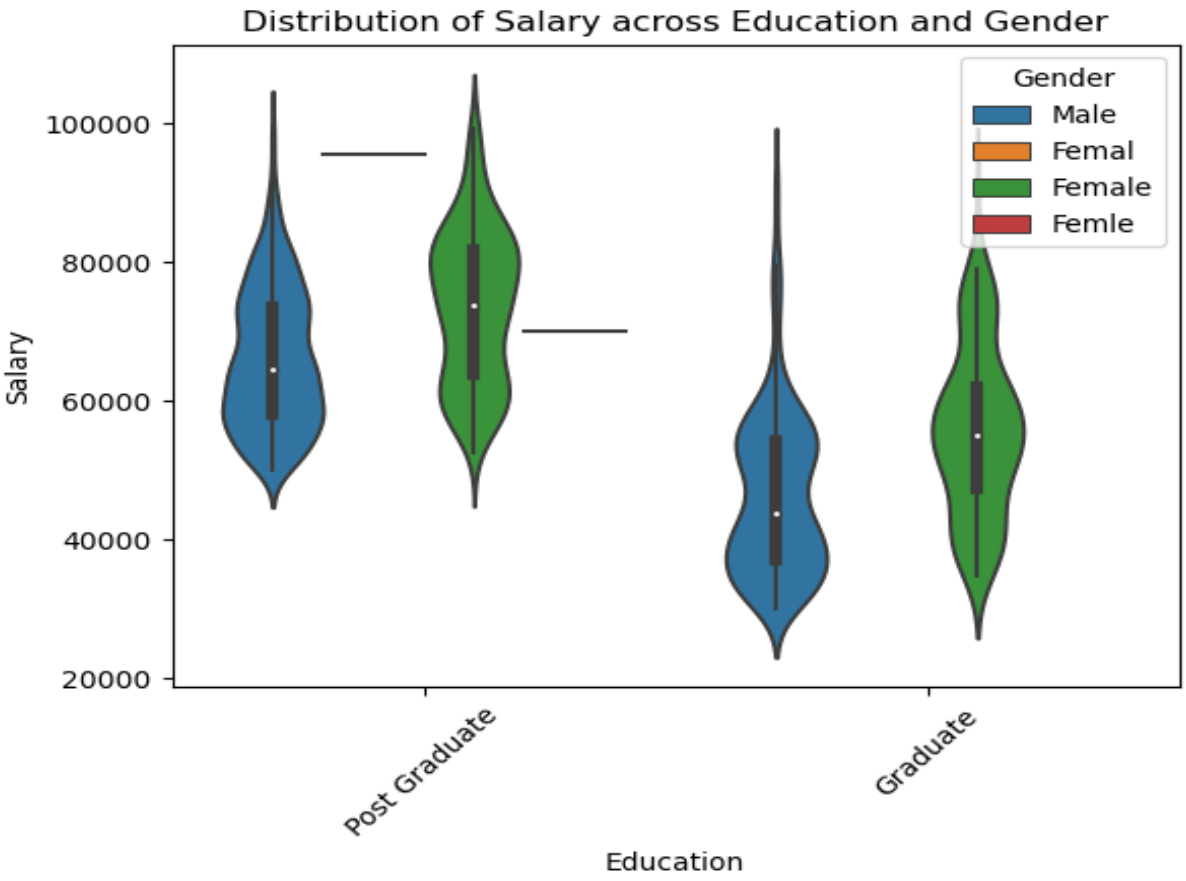

Relationship between Age, Salary, and Price
Salary
Age
Price
20000
30000
40000
50000
60000
70000
100000
90000
80000
70000
60000
50000
40000
30000
25
30
35
40
45
50
55

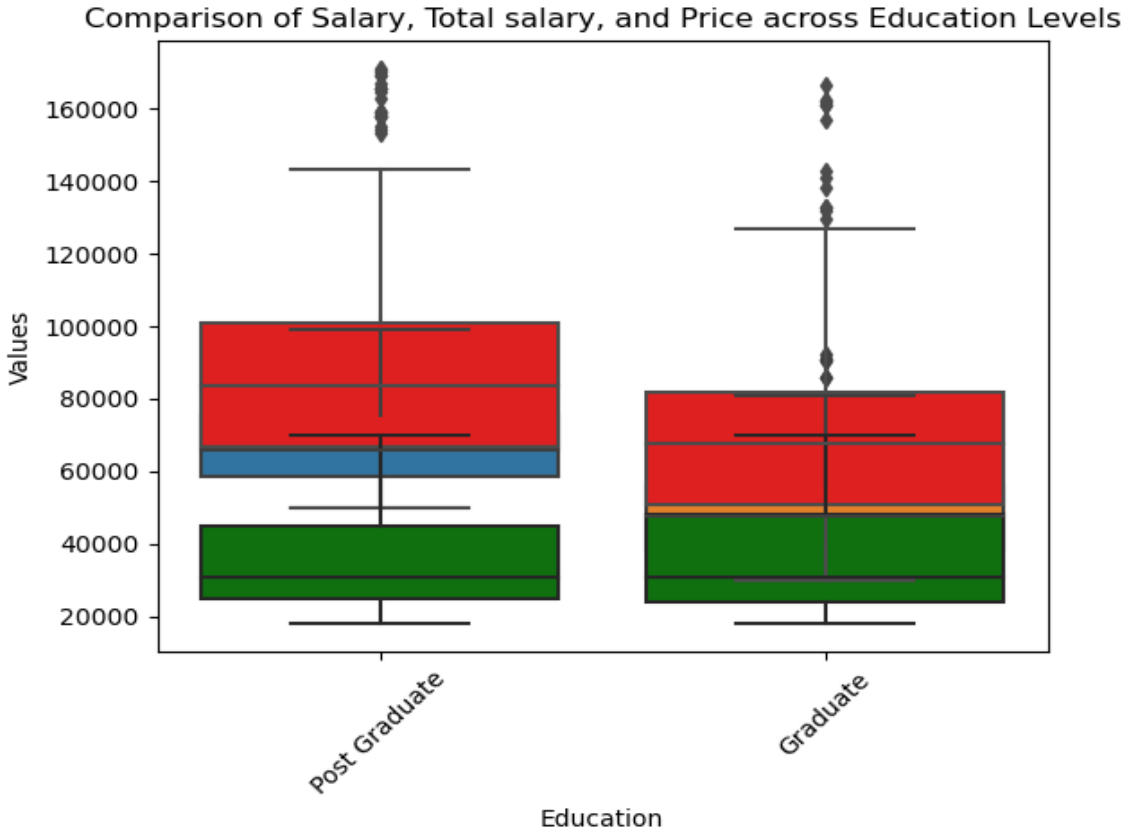

Comparison of Salary across Education Levels
Salary
Education
Post Graduate
Graduate
100000
90000
80000
70000
60000
50000
40000
30000

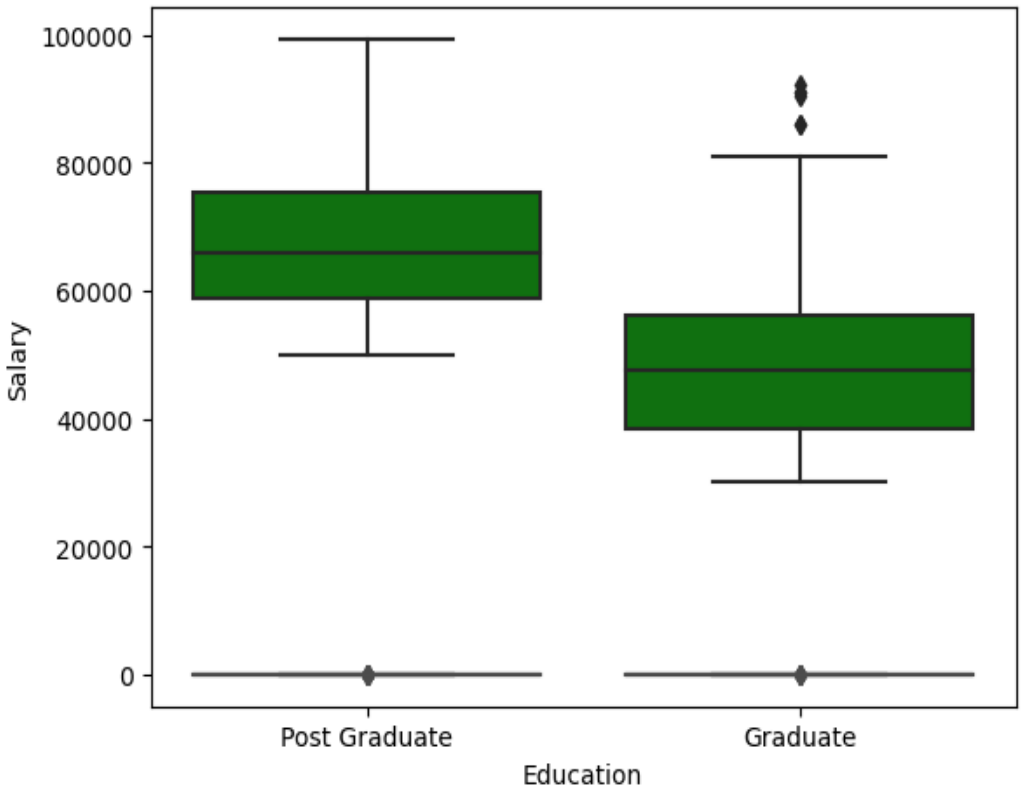

Count of Individuals by Marital Status
1400
1200
1000
800
600
400
200
0
Count
Married
Single
Marital Status

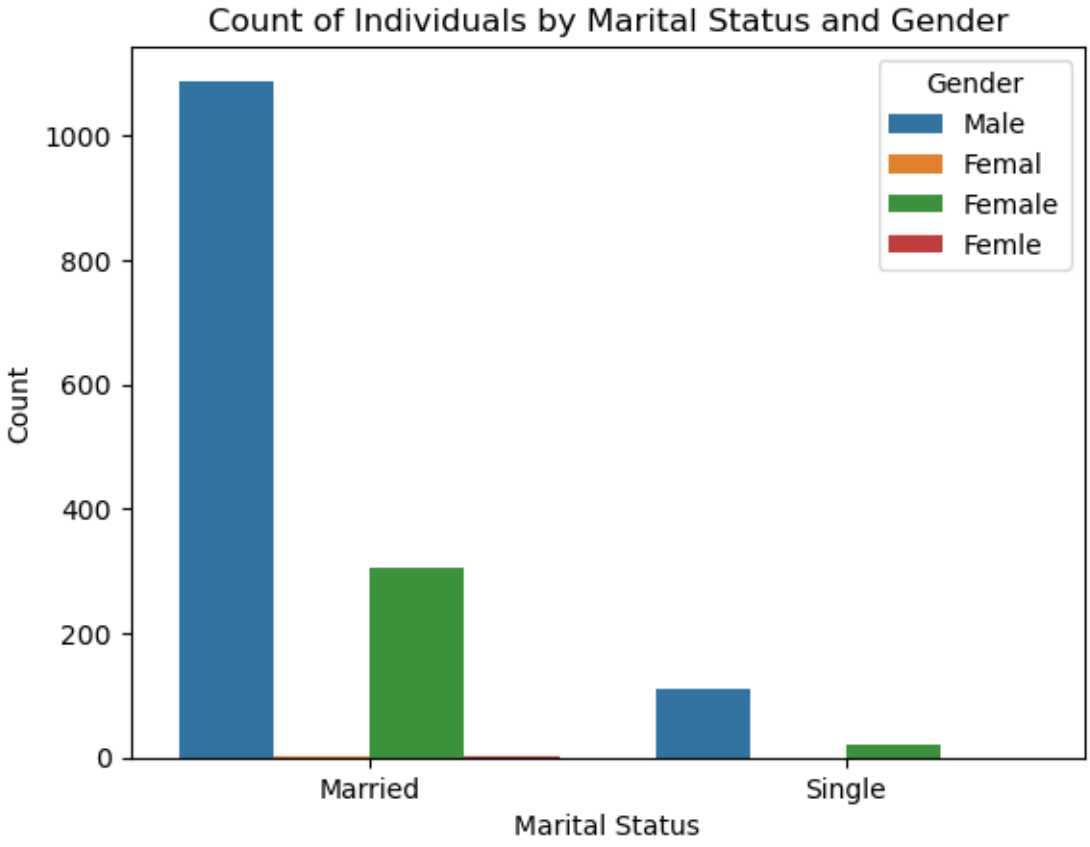

Count of Individuals by Profession and Gender
Gender
Male
Femal
Female
Femle
600
500
400
300
200
100
0
Count
Business
Salaried
Profession

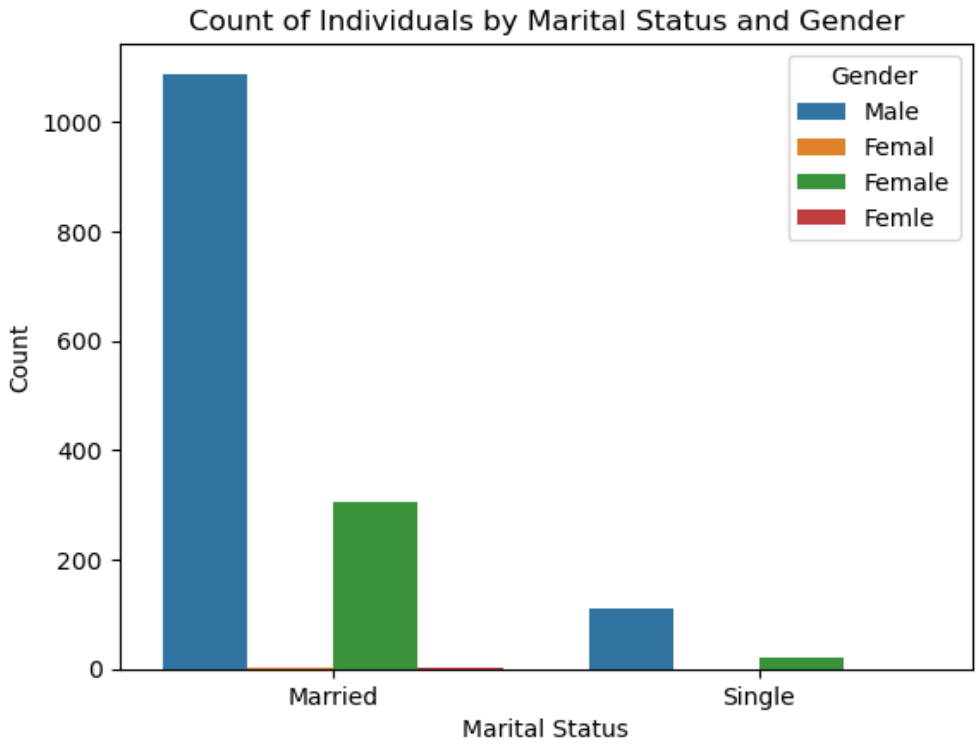

Age
No_of_Dependents
Salary
Price
50
40
30
4
3
2
1
0
100000
80000
60000
40000
70000
60000
50000
40000
30000
20000
30
40
50
0
2
4
40000
60000
80000
100000
20000
40000
60000

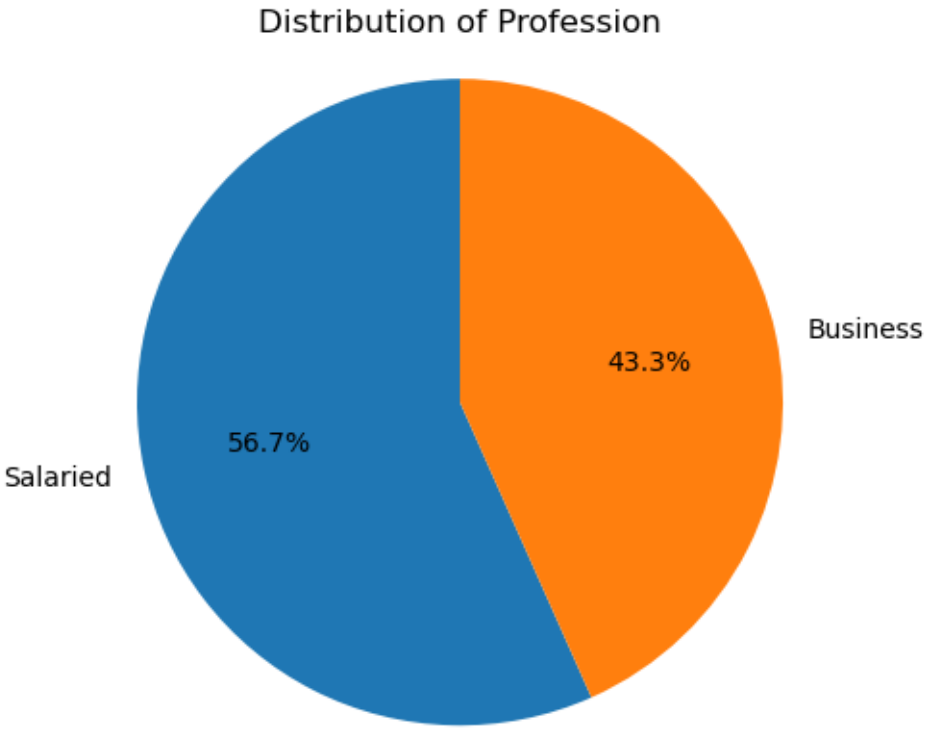

Distribution of Salary across Education and Gender
Gender
Male
Femal
Female
Femle
Salary
100000
80000
60000
40000
20000
Post Graduate
Graduate
Education

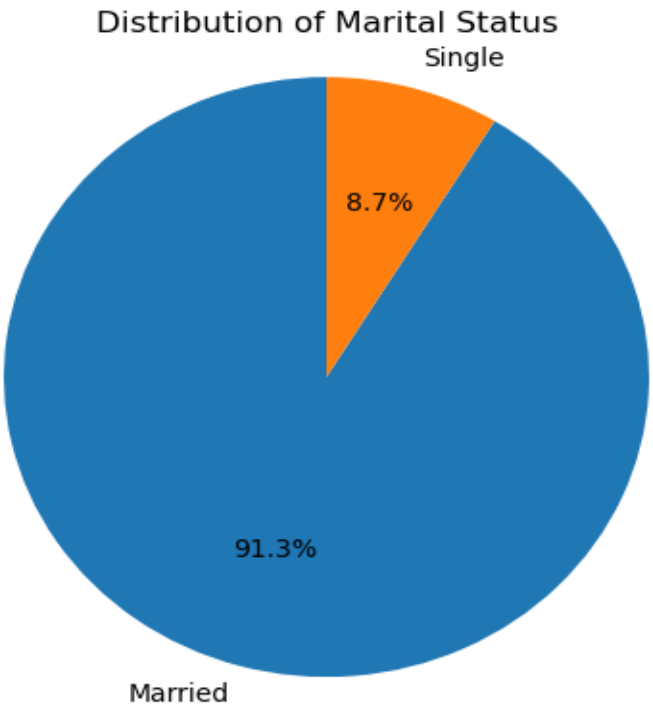

Comparison of Salary, Total salary, and Price across Education Levels
Values
160000
140000
120000
100000
80000
60000
40000
20000
Post Graduate
Graduate
Education

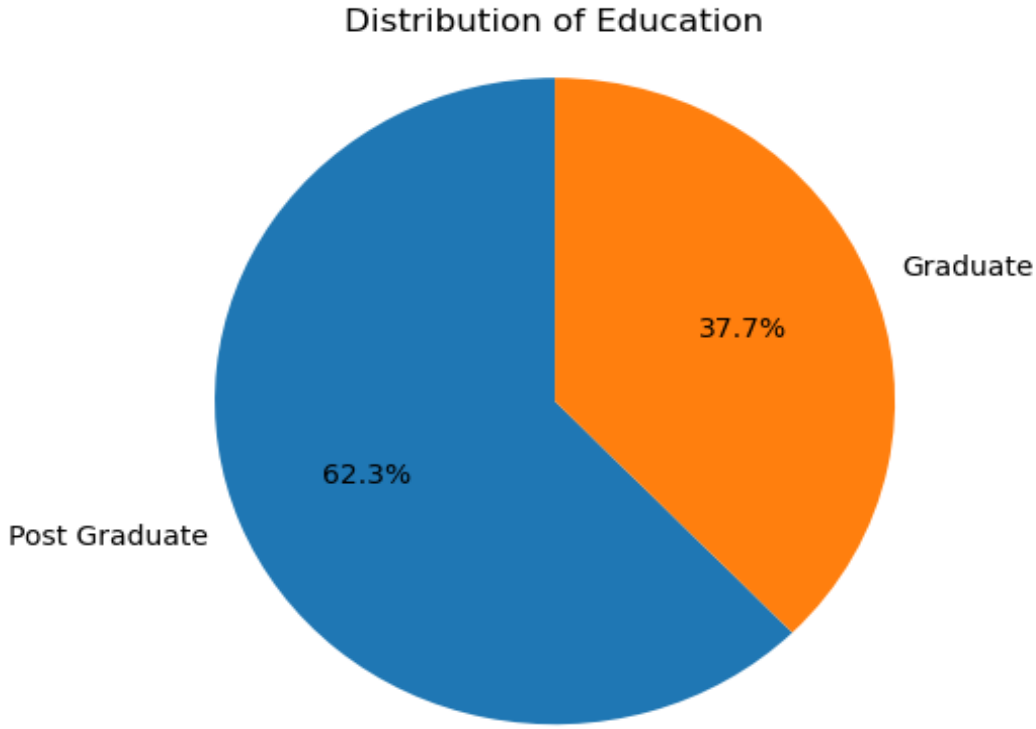

100000
80000
60000
40000
20000
0
Salary
Post Graduate
Graduate
Education

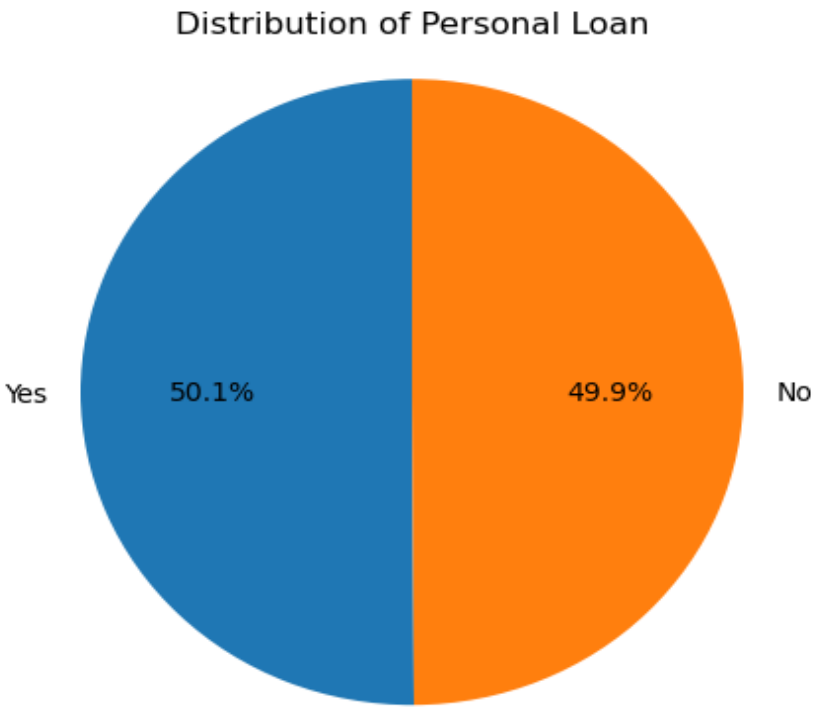

Count of Individuals by Marital Status and Gender
Gender
Male
Femal
Female
Femle
1000
800
600
400
200
0
Count
Married
Single
Marital Status

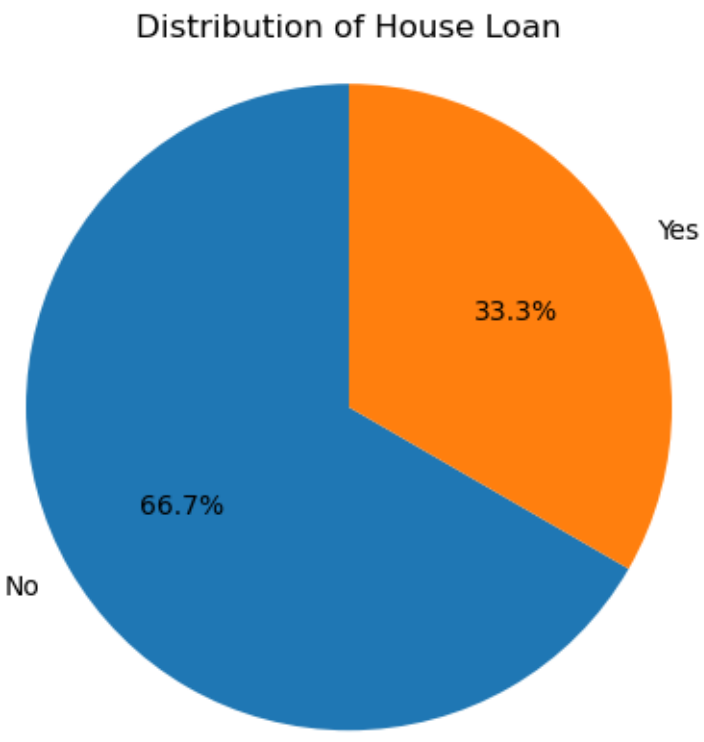

Count of Individuals by Marital Status and Gender
Gender
Male
Femal
Female
Femle
1000
800
600
400
200
0
Count
Married
Single
Marital Status

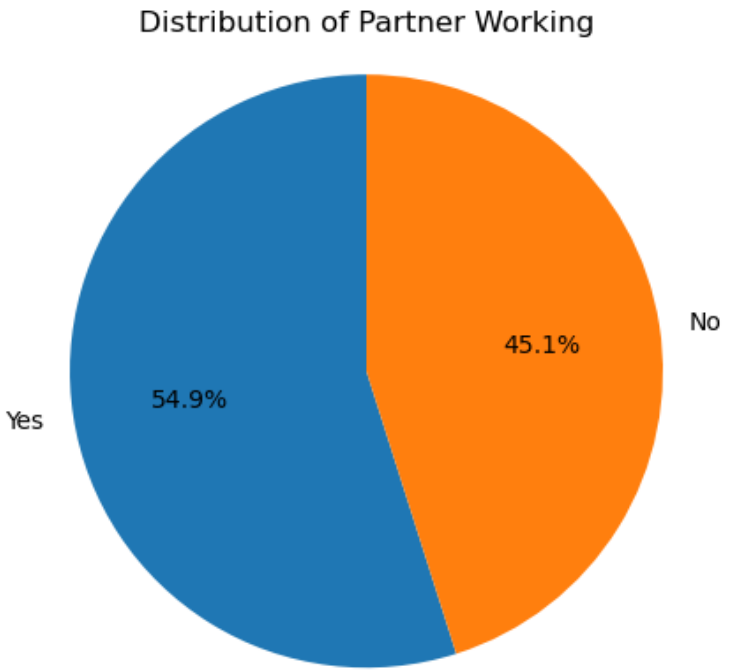

Distribution of Profession
Business
43.3%
56.7%
Salaried

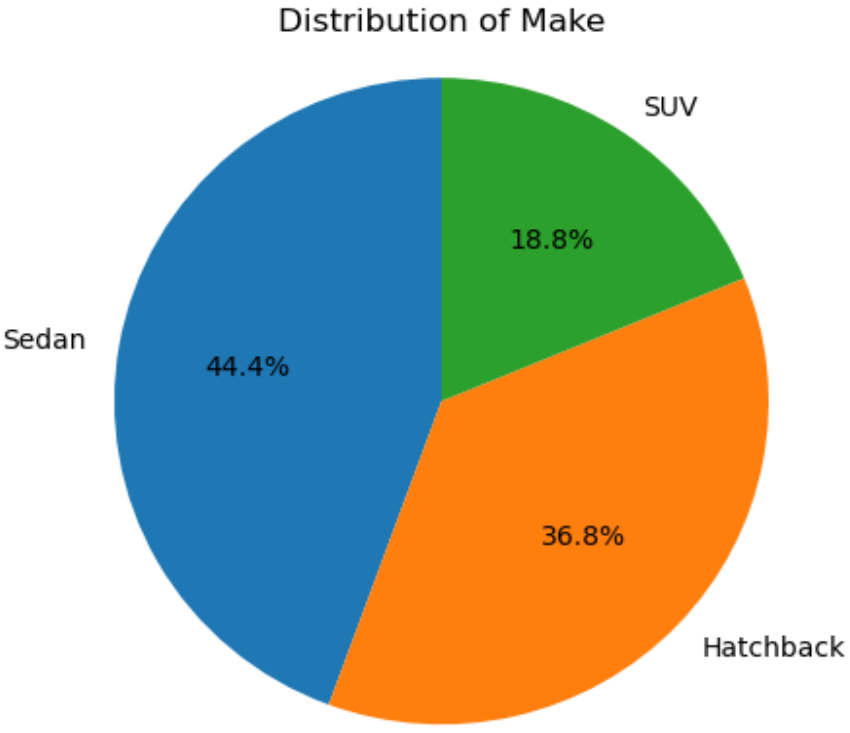

Distribution of Marital Status
Single
8.7%
91.3%
Married

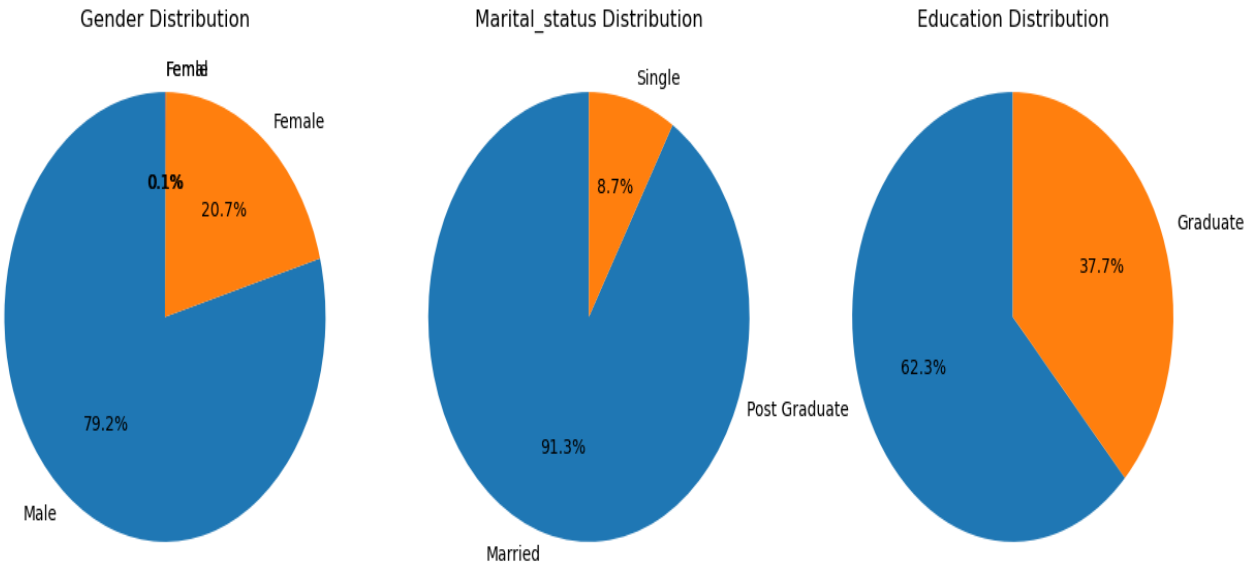

Distribution of Education
Graduate
37.7%
62.3%
Post Graduate

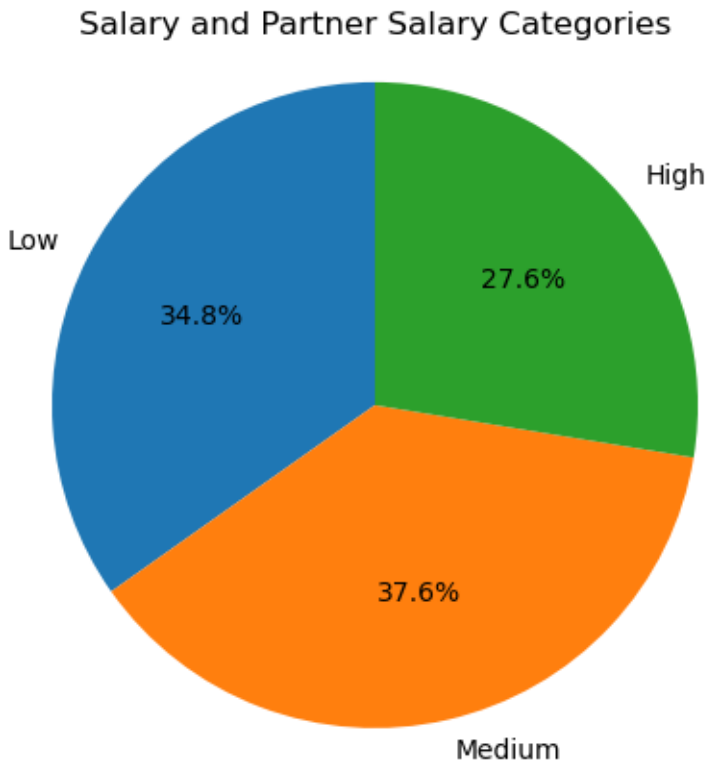

Distribution of Personal Loan
Yes
50.1%
49.9%
No


Distribution of House Loan
Yes
33.3%
66.7%
No

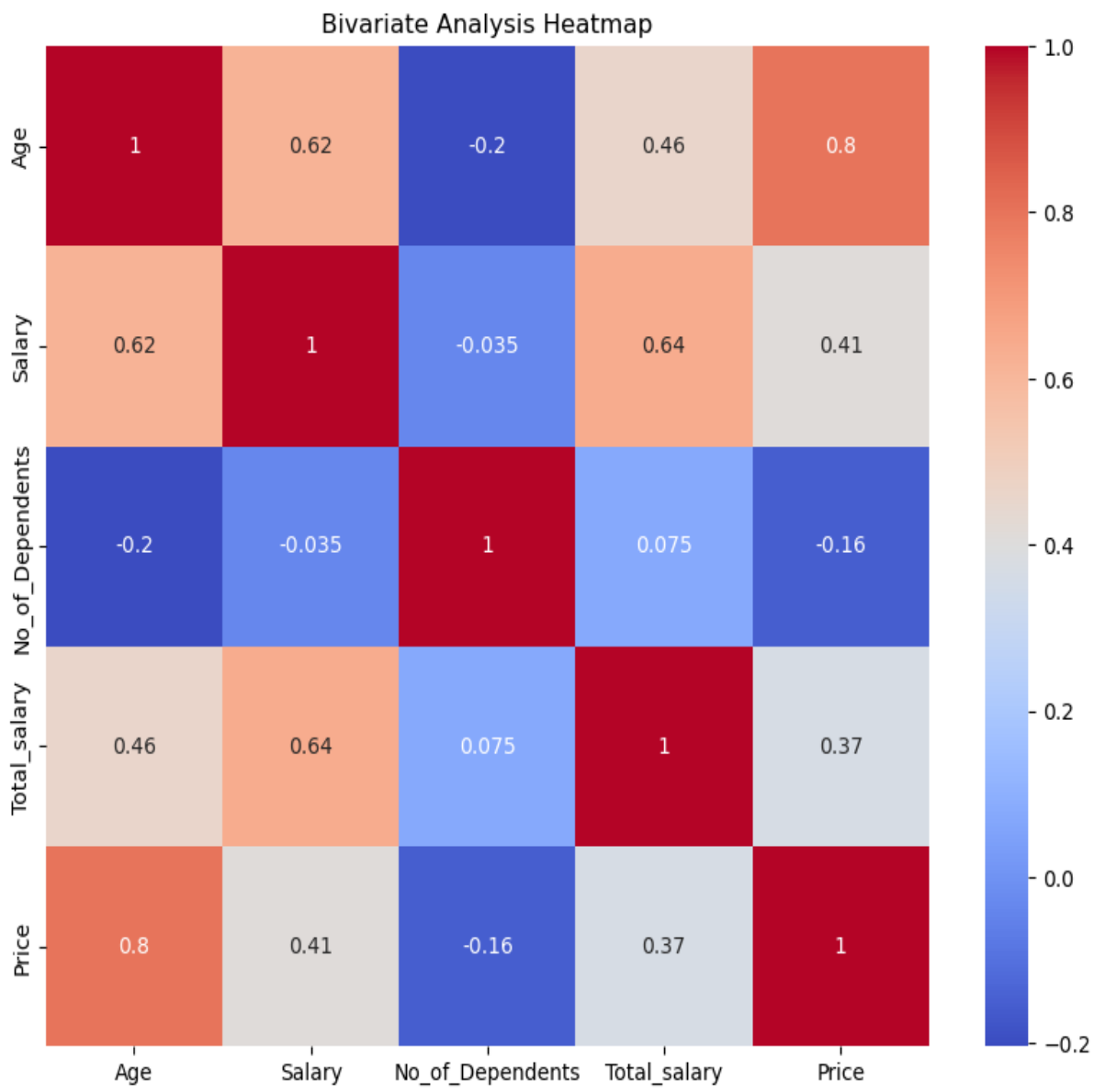

Distribution of Partner Working
No
45.1%
54.9%
Yes


Distribution of Make
SUV
18.8%
Sedan
44.4%
36.8%
Hatchback

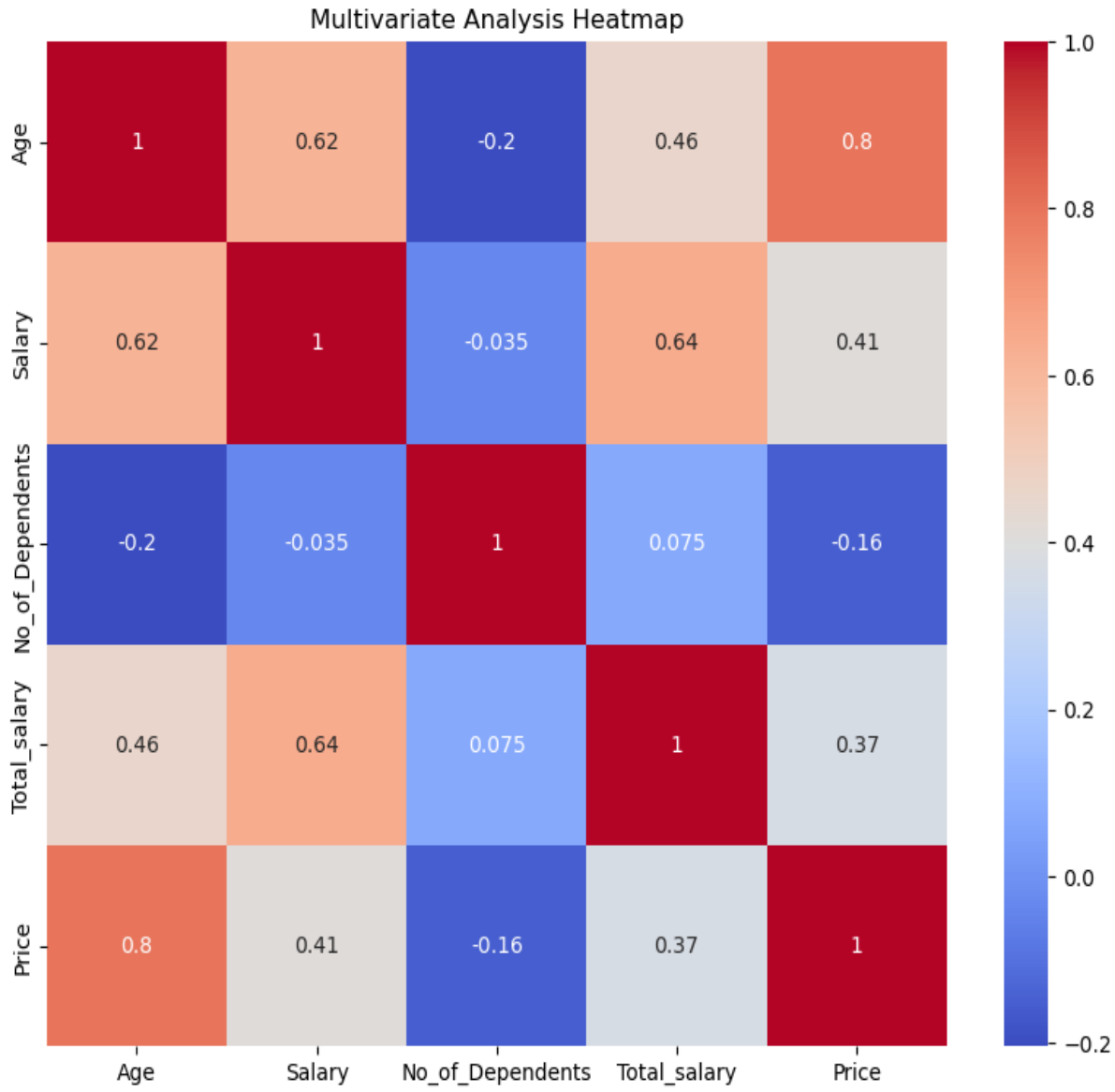

Gender Distribution
Female
Female
0.1%
20.7%
79.2%
Male
Marital_status Distribution
Single
8.7%
91.3%
Married
Post Graduate
Education Distribution
Graduate
37.7%
62.3%


Salary and Partner Salary Categories
High
Low
27.6%
34.8%
37.6%
Medium

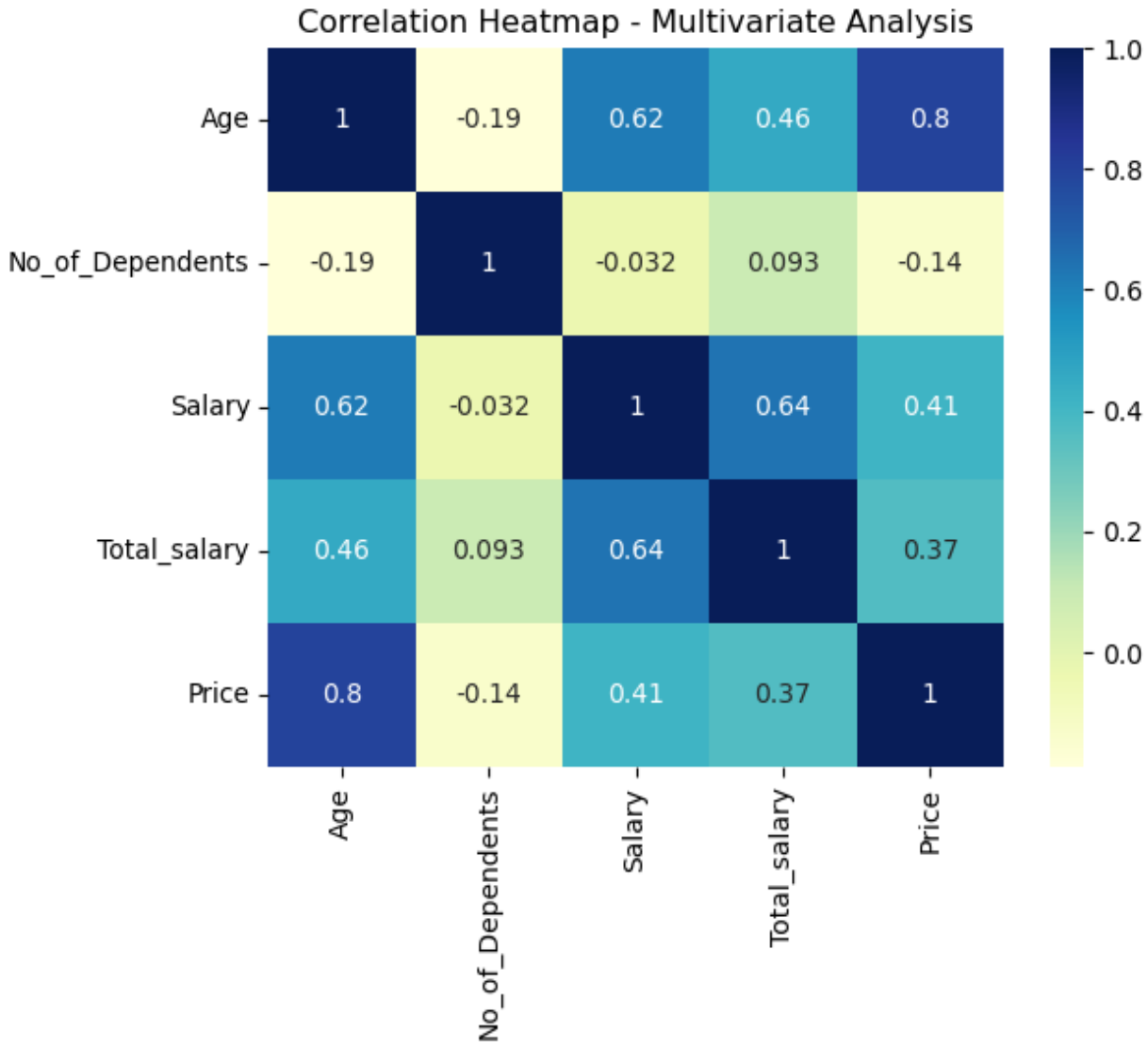
Bivariate Analysis Heat map



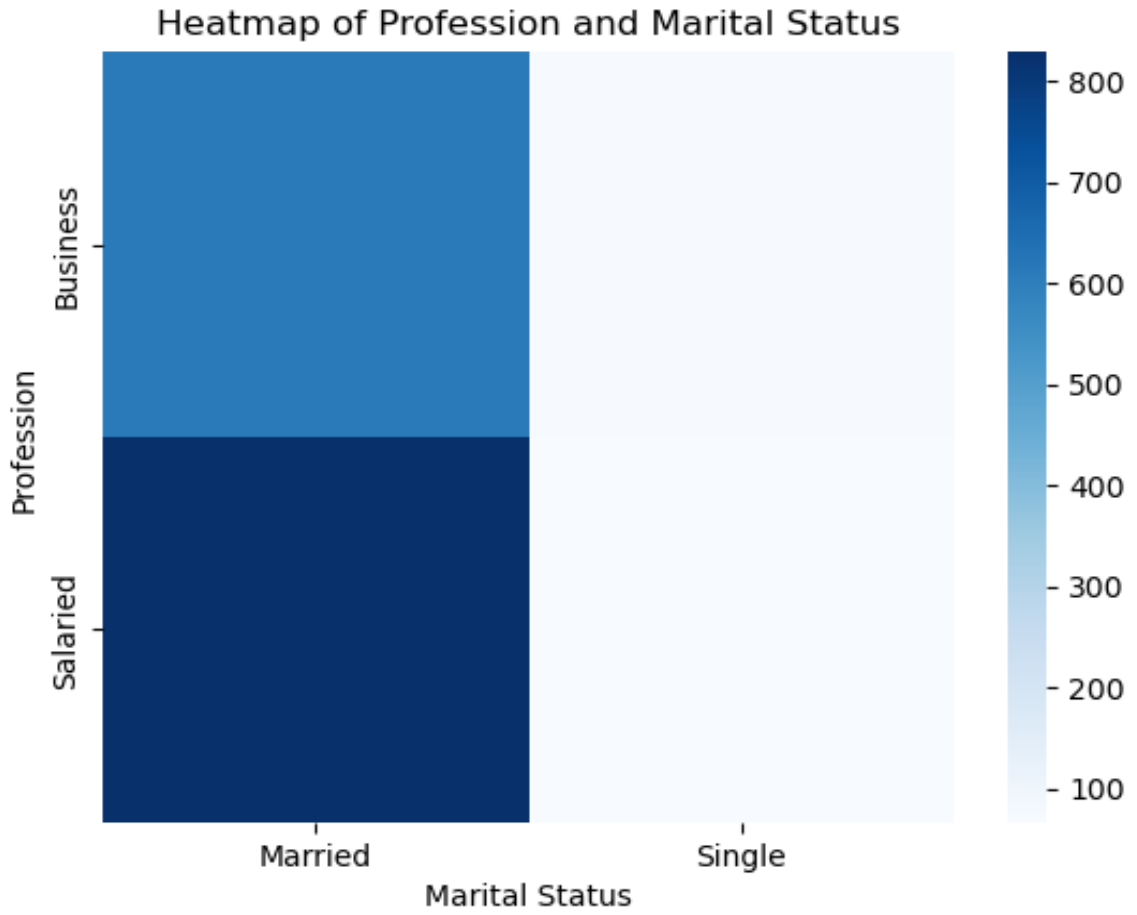
Multivariate Analysis Heat map



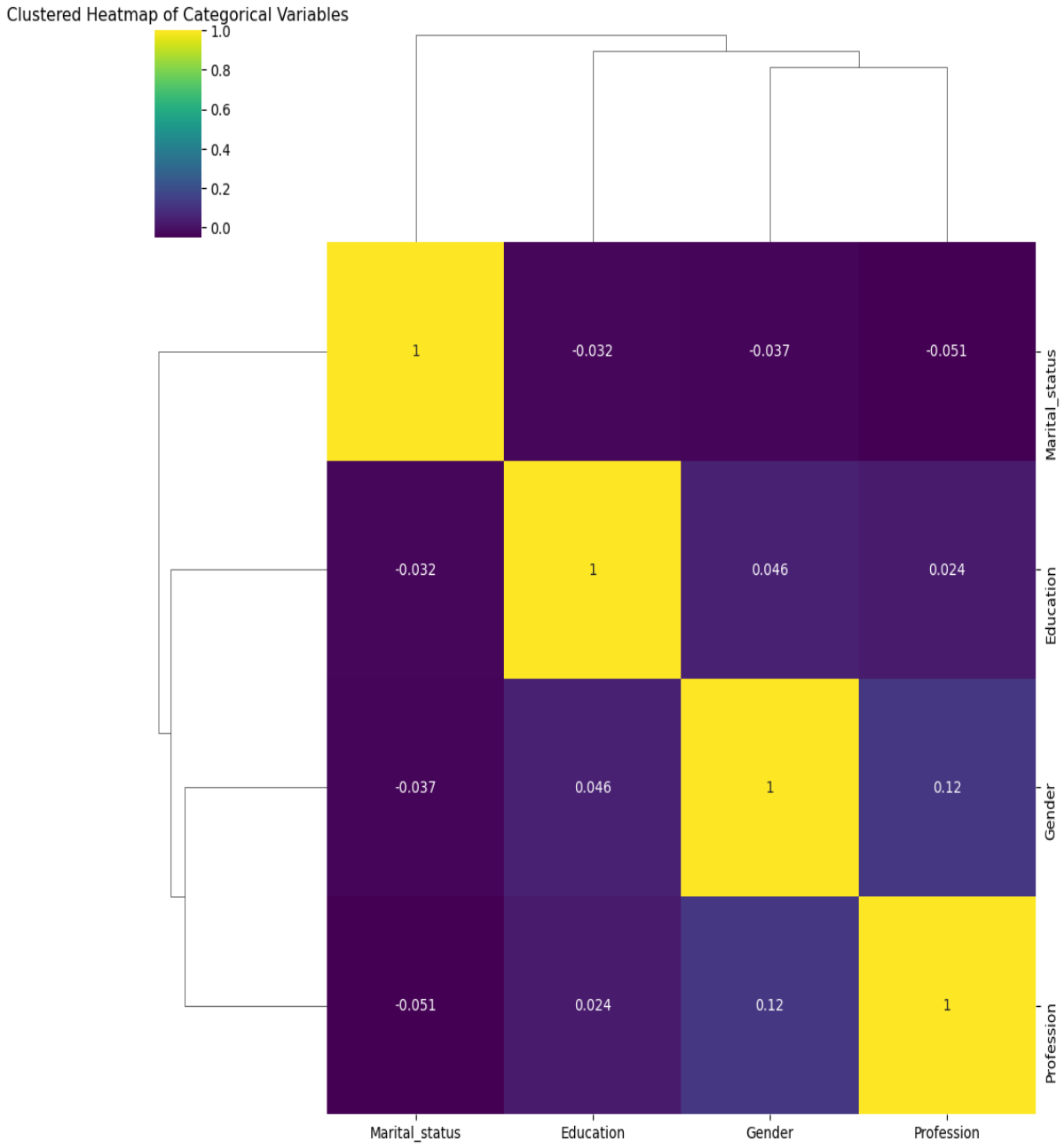
 # Correlation heat map of multivariate analysis



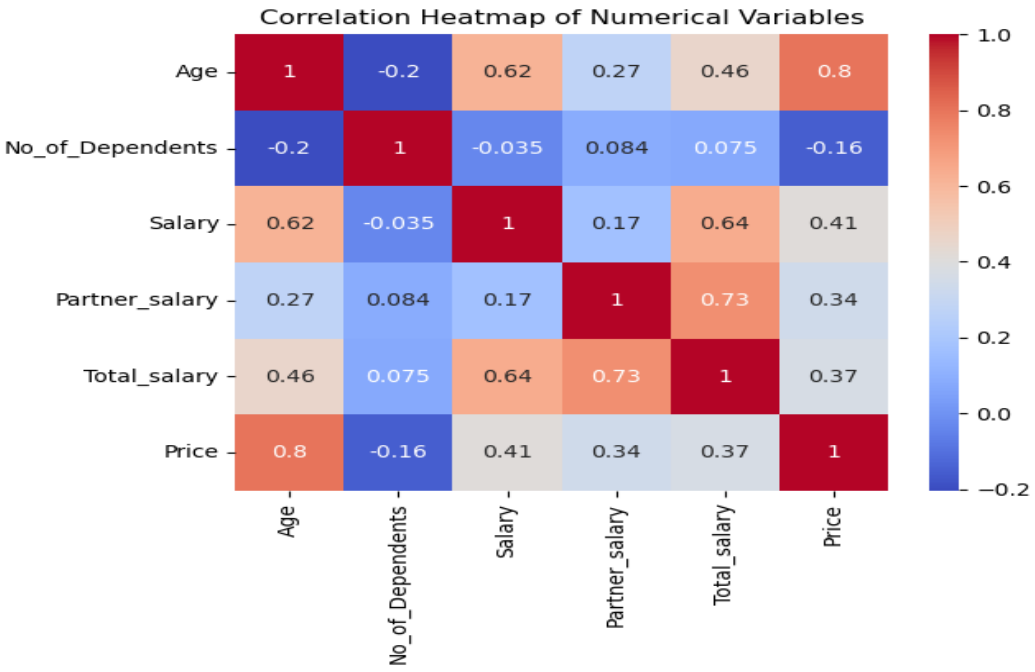
Visualize the relationship between two categorical variables using a color-coded heat map.



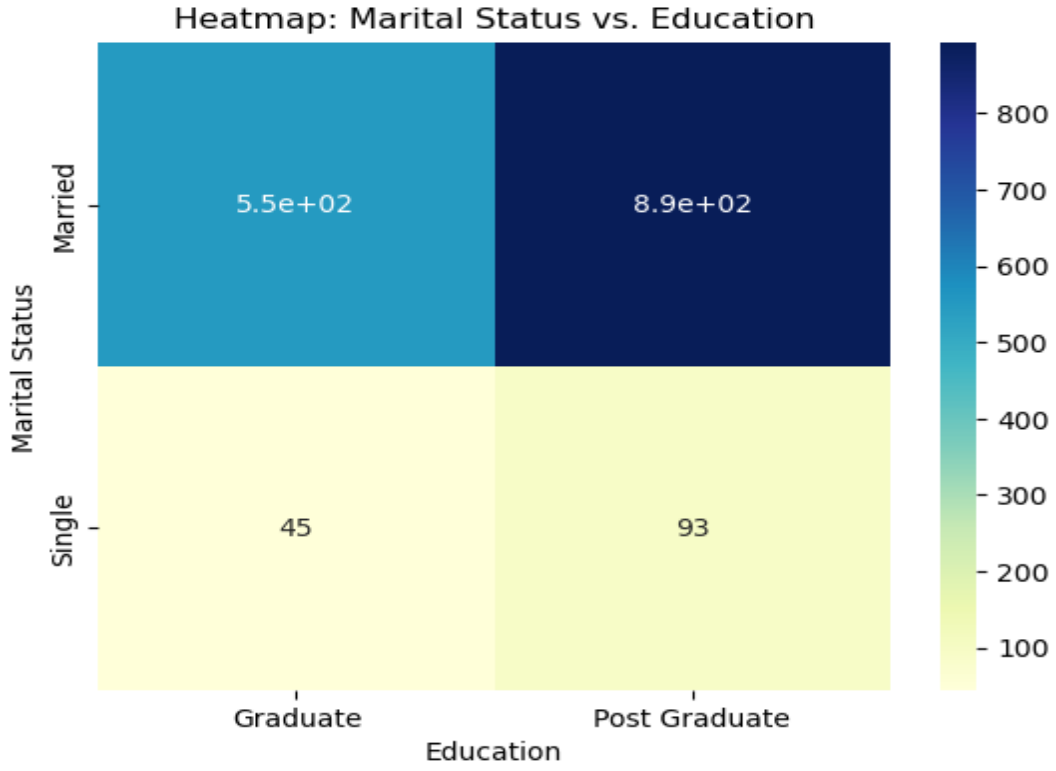
Correlation Heat map: Display the correlation matrix between multiple numerical variables.



Correlation Heat map of Numerical Variables



Contingency Table:



Above plots information:

 # Box Plot: Visualize the distribution of 'Age'

 # Bar Plot: Count the number of individuals for each 'Gender'

 # Scatter Plot: Analyse the relationship between 'Age' and 'Salary'

 # Histogram: Observe the distribution of 'No_of_Dependents'

 # Count Plot: Count the occurrences of each 'Profession'

 # Box Plot: Compare 'Salary' across different 'Education' levels

 # Bar Plot: Display the count of individuals based on 'Marital status'

 # Scatter Plot: Explore the relationship between 'Salary' and 'Price'

 # Histogram: Examine the distribution of 'Total salary'

 # Count Plot: Count the occurrences of each 'Personal loan' status

 # Box Plot: Compare 'Salary' across different 'House loan' categories

 # Bar Plot: Count the number of individuals with 'Partner working' or not

 # Scatter Plot: Analyse the relationship between 'Age' and 'No_of_Dependents'

 # Histogram: Observe the distribution of 'Price'

 # Count Plot: Count the occurrences of each 'Make'

 # Box Plot: Compare 'Salary' across different 'Gender' and 'Education' levels

 # Bar Plot: Display the count of individuals based on 'Profession' and 'Gender'

 # Scatter Plot: Explore the relationship between 'Salary', 'Price', and 'Age'

 # Histogram: Examine the distribution of 'Total salary' across 'Education' levels

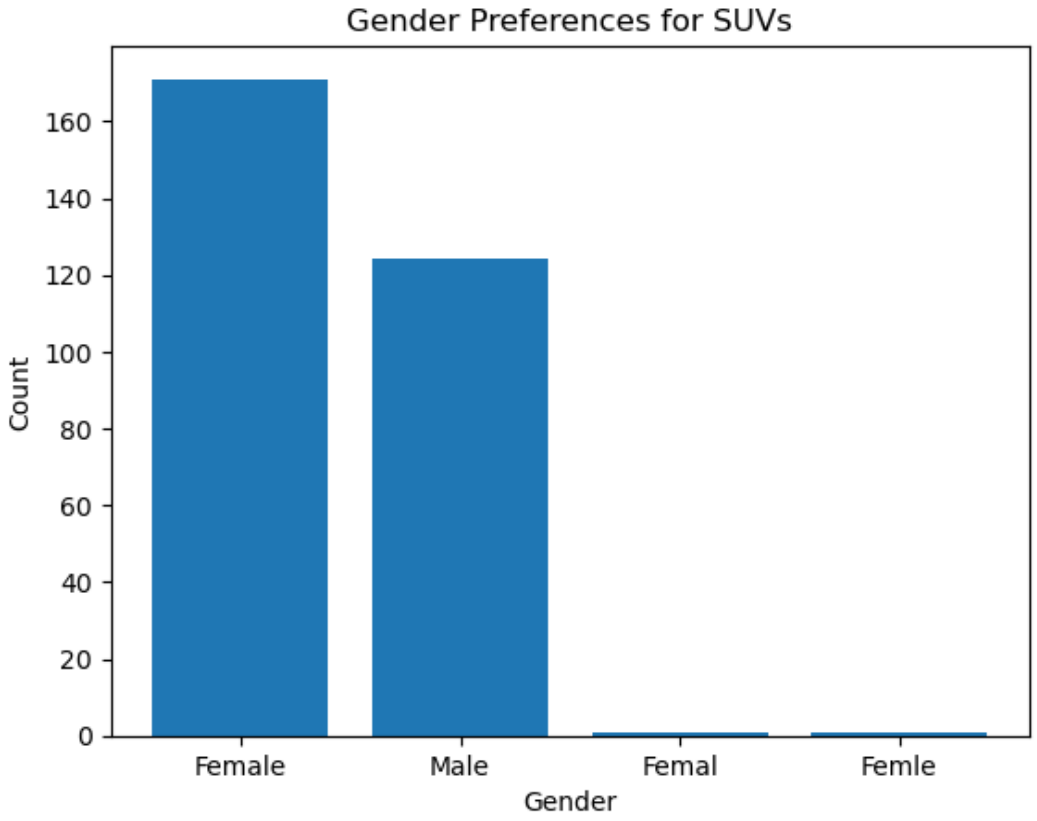
 # Count Plot: Count the occurrences of each 'Personal loan' status based on 'Marital status'

E. Employees working on the existing marketing campaign have made the following remarks. Based on the data and your analysis state whether you agree or disagree with their observations. Justify your answer Based on the data available.

E1) Steve Roger says "Men prefer SUV by a large margin, compared to the women"

```
suv_preferences = dataset [dataset ['Make'] == 'SUV'] ['Gender'].value_counts ( )
```

Output: Female 172
Male 124



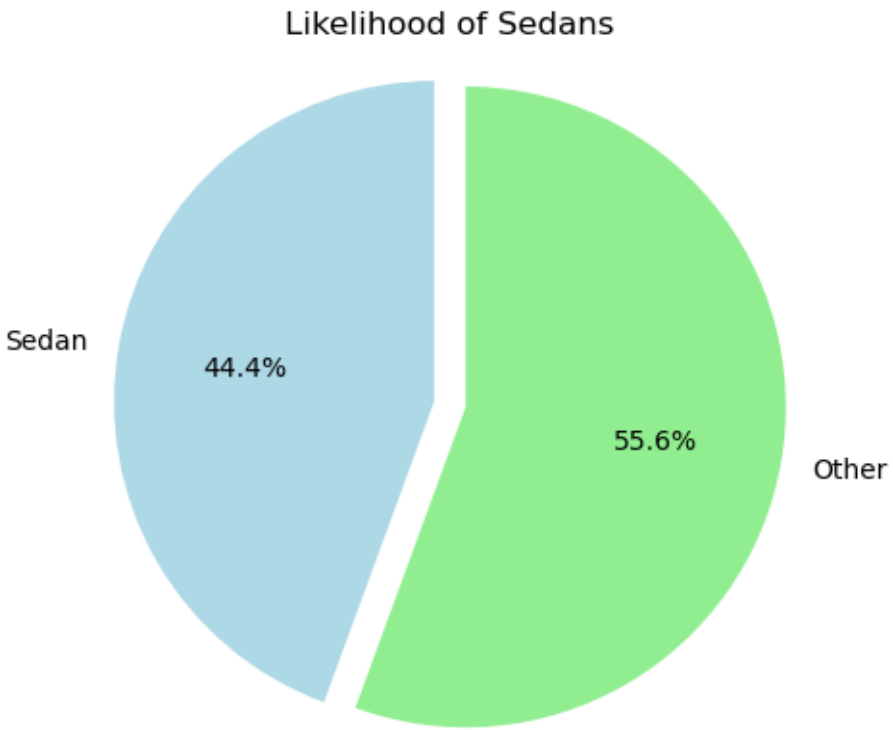
Short summary:

According to Steve Roger's study of the dataset, it is clear that men and women have very differe nt preferences for SUVs. The data finds that there are 172 cases when females are the gender ide ntified among the items particularly related to SUVs, while males account for 124 occurrences. Th us, based on the data, it can be inferred that, contrary to Steve Roger's initial assertion, women ex hibit a larger desire for SUVs than do males.

E2) Ned Stark believes that a salaried person is more likely to buy a Sedan.

```
sedan_likelihood = dataset [dataset ['Make'] == 'Sedan'] ['Salary'].count ( ) / dataset ['Salary'].count ( )
```

Output: 0.444022770398482



Short summary:

According to Ned Stark, people who are paid a salary are more likely to buy sedans. The methodology involved determining the ratio of sedan purchases to all entries in the dataset, which represents the likelihood of purchasing a sedan, in order to assess this assertion. The output value of 0.444022770398482 shows that relative to the total number of entries, sedan purchases make up about 44.4% of the dataset. This statistic lends some weight to Ned Stark's hypothesis that a salaried person is somewhat likely to purchase a sedan.

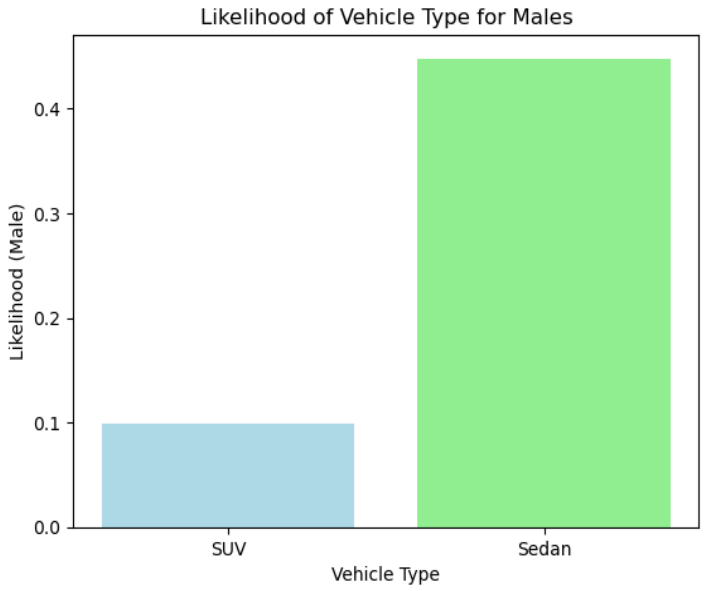
E3) Sheldon Cooper does not believe any of them; he claims that a salaried male is an easier target for a SUV sale over a Sedan Sale.

```
suv_likelihood_male = (dataset [(dataset ['Make'] == 'SUV') & (dataset ['Gender'] == 'Male') & (dataset ['Salary'] > 0)].shape [0]) / dataset [(dataset ['Gender'] == 'Male') & (dataset ['Salary'] > 0)].shape [0]

sedan_likelihood_male = (dataset [(dataset ['Make'] == 'Sedan') & (dataset ['Gender'] == 'Male') & (dataset ['Salary'] > 0)].shape [0]) / dataset [(dataset ['Gender'] == 'Male') & (dataset ['Salary'] > 0)].shape [0]

suv_likelihood_male, sedan_likelihood_male
```

Output :(0.09904153354632587, 0.4480830670926518)



Short summary:

Based on their gender and financial situation, men are more likely to buy SUVs and sedans, according to the study done on the dataset.

According to data on suv_likelihood_male, which was found to be roughly 0.09904153354632587, 9.9% of men with positive salaries are likely to buy SUVs.

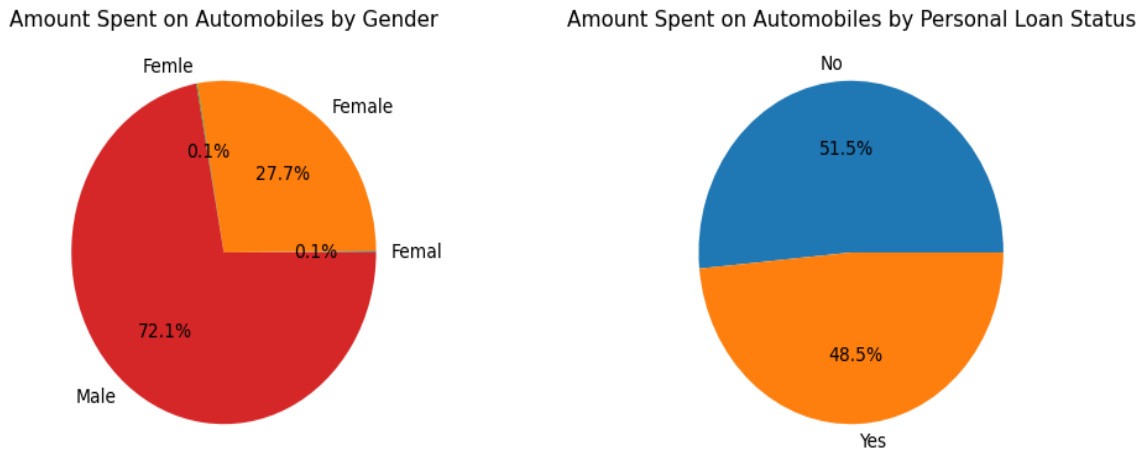
The sedan_likelihood_male, on the other hand, is almost 0.4480830670926518, which means that among males with a positive wage, about 44.8% are likely to buy sedans.

According to the survey, men with positive salaries are substantially more likely to buy sedans (44.8%) than SUVs (9.9%), according to the analysis's findings. When taking into account their gender, financial situation, and vehicle preference, this suggests that men have a stronger preference for sedans.

F. From the given data, comment on the amount spent on purchasing automobiles across the following categories. Comment on how a Business can utilize the results from this exercise. Give justification along with presenting metrics/charts used for arriving at the conclusions. Give justification along with presenting metrics/charts used for arriving at the conclusions.

*****F1) Gender ***F2) Personal loan**

```
gender_spending = dataset.groupby ('Gender') ['Price'].sum ( ) loan_spending =
dataset.groupby ('Personal_loan') ['Price'].sum ( )
Fig, axes = plt.subplots (1, 2, figsize= (12, 4))
Axes [0].pie gender_spending, labels=gender_spending.index, autopct='%1.1f %%')
Axes [0].set_title ('Amount Spent on Automobiles by Gender')
Axes [1].pie (loan_spending, labels=loan_spending.index, autopct='%1.1f %%')
Axes [1].set_title ('Amount Spent on Automobiles by Personal Loan Status')
plt.show ( )
```



Short summary:

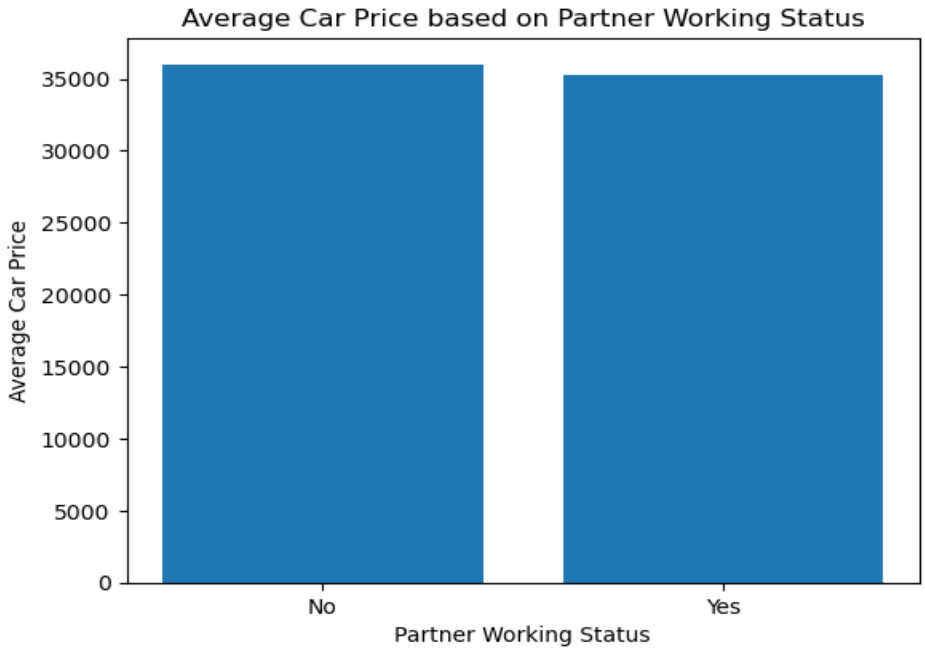
Based on gender, the data displays the overall amount spent on autos. The pie chart shows how money is distributed, with each slice denoting a particular gender category (such as male and female), and the accompanying percentage showing how much money is spent by each gender. In contrast to women, the chart demonstrates that men spend more money on cars.

Based on personal loan status, the study looks into the amount spent on cars. The pie chart shows how spending is distributed, with each slice representing a category of personal loan status (such as having a personal loan and not having one), and the accompanying percentage showing how much of spending falls into each group. According to the graph, people who do not have personal loans spend more money on cars than people who do.

In conclusion, businesses can gain insights into client behaviour and preferences by analysing the amount spent on cars by gender and personal loan status. Businesses can maximise customer happiness and boost sales by utilising these data to effectively allocate resources, create focused marketing tactics, and customise their product offerings.

G. From the current data set comment if having a working partner leads to purchase of a higher priced car.

```
average_prices = dataset.groupby ('Partner_working') ['Price'].mean ( )
plt.bar (average_prices.index, average_prices.values)
plt.xlabel ('Partner Working Status')
plt.ylabel ('Average Car Price')
plt.title ('Average Car Price based on Partner Working Status')
plt.show ( )
```



Short summary:

The analysis of the current dataset looks at the connection between having a co-worker and buying more expensive autos. The partner's employment status—whether the partner is employed or not—is used to determine the average car prices. The average car price for each partner working status category is displayed using a bar chart.

The average cost of a car for those who work with a partner and those who do not is shown in the bar chart. According to the data, those who work with a partner typically buy more expensive cars than people who don't.

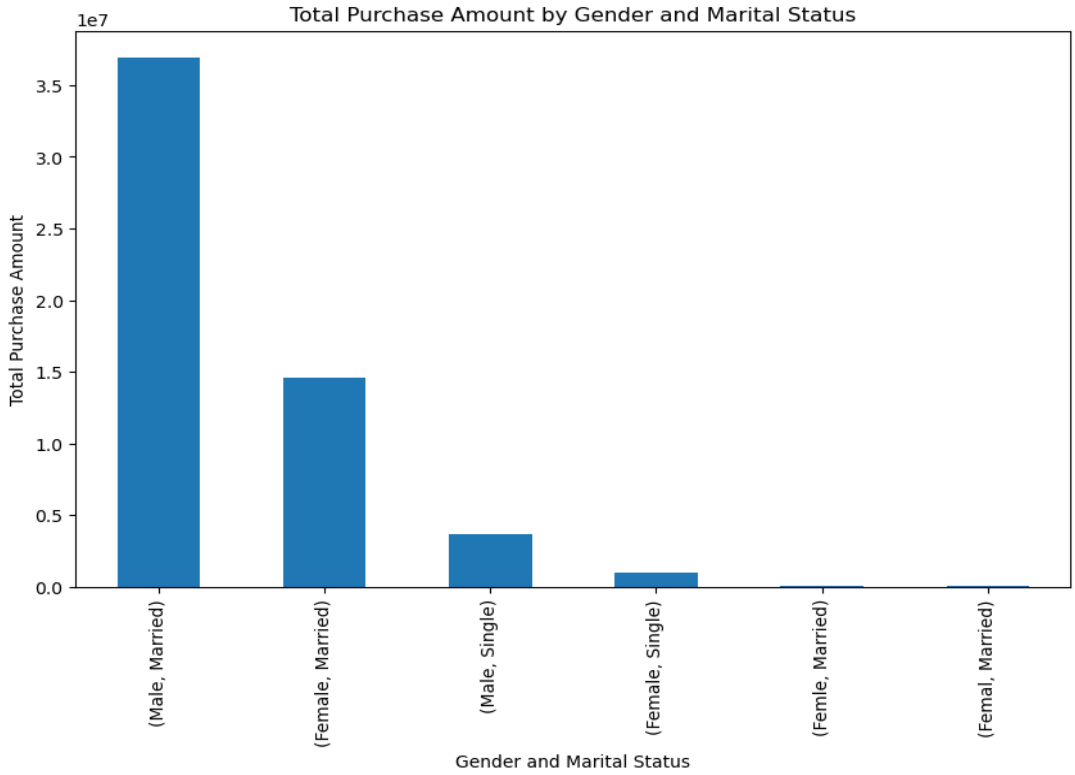
These insights can help businesses comprehend how partner employment status affects car purchase choices. Businesses can modify their marketing strategy and product offers to target this particular demographic by realising that having a working partner is linked to more expensive car purchases. This can entail highlighting attributes or financing choices that appeal to families with two incomes and greater spending power. Businesses can increase their sales and better serve potential clients by efficiently targeting and catering to this customer category.

H. The main objective of this analysis is to devise an improved marketing strategy to send targeted information to different groups of potential buyers present in the data. For the current analysis use Gender and Marital status - fields to arrive at groups with similar purchase history.

```
grouped_purchase_history = dataset.groupby (['Gender', 'Marital_status']) ['Price'].sum ( )
sorted_purchase_history = grouped_purchase_history.sort_values (ascending=False)
Print (sorted_purchase_history)
```

Gender Marital_status		
Male	Married	36949000
Female	Married	14585000
Male	Single	3636000
Female	Single	984000
Female	Married	65000
Female	Married	61000

```
grouped_purchase_history = dataset.groupby (['Gender', 'Marital_status']) ['Price'].sum ( )
sorted_purchase_history = grouped_purchase_history.sort_values (ascending=False)
plt.figure (figsize= (10, 6))
sorted_purchase_history.plot (kind='bar', stacked=True)
plt.xlabel ('Gender and Marital Status')
plt.ylabel ('Total Purchase Amount')
plt.title ('Total Purchase Amount by Gender and Marital Status')
plt.show ( )
```



Short Summary:

Based on gender and marital status, the analysis sought to classify distinct groups of potential customers in order to develop a more successful marketing approach to reach these groups. The following groupings, along with the associated total purchase amounts, were determined by the analysis:

Male, Married: With a total purchase of 36,949,000, this group showed the highest average price. They represent a sizable group of prospective customers who may be catered to with special marketing messages and promotions.
Female, Married: This category also exhibits significant purchasing power, with a total purchase amount of $14,585,000. To increase sales, targeted marketing campaigns can be created to meet the interests and demands of married women.

Single: This demographic spent a total of $3,636,000, which suggests that there may be opportunities to attract single males with targeted marketing tactics and product selections that suit their tastes and way of life.
Female, Single: With $984,000 in total spending, this category gives businesses the chance to entice and retain single women as customers by getting to know their specific preferences and providing for them.
Female, Married: With only 65,000 spent overall, this category had a significantly lower average.
Female, Married: In comparison, just 61,000 worth of purchases were made by this category.
Correcting the error in the gender designation is crucial.

Businesses can create tailored marketing campaigns, improve product offers, and create specialised promotional methods to efficiently engage and convert potential customers by identifying these unique groups and their purchasing behaviours. This analysis helps to maximise marketing initiatives and allocate resources to the most potential buyer segments, ultimately boosting customer happiness and driving sales.