PDS Coded Project Report

Prepared By: Parthasarathi Behura

CONTENTS:

Austo Motor Company Page

LIST OF FIGURES:

LIST OF TABLES:

DSBA Project

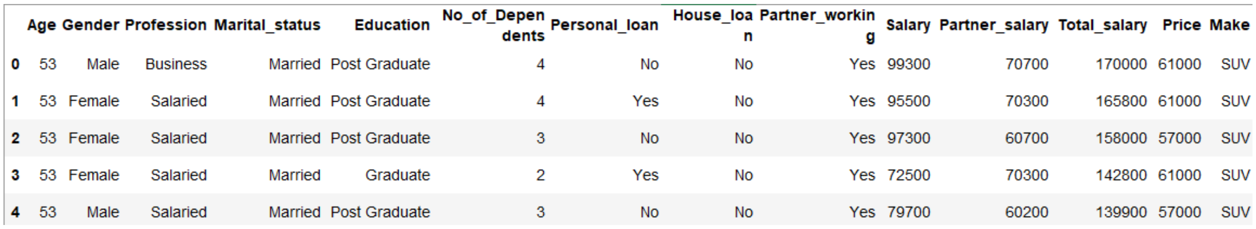
Austo Motor Company is a leading car manufacturer specializing in SUV, Sedan, and Hatchback models. In its recent board meeting, concerns were raised by the members on the efficiency of the marketing campaign currently being used. The board decides to rope in analytics professional to improve the existing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

Imported the libraries for the Data are

- Numpy
- Pandas
- Matplotlib
- Seaborn

1. There are some information about the dataset, decision makers should have a look.

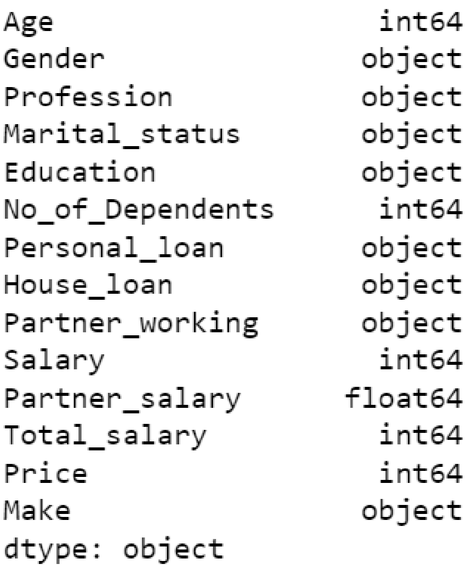
- The dataset is having 1581 rows and 14 columns.
- There is a look on the 5 sample rows to check the data type.



	Age	Gender	Profession	Marital_status	Education	No_of_Dependents	Personal_loan	House_loan	Partner_working	Salary	Partner_salary	Total_salary	Price	Make
0	53	Male	Business	Married	Post Graduate	4	No	No	Yes	99300	70700	170000	61000	SUV
1	53	Female	Salaried	Married	Post Graduate	4	Yes	No	Yes	95500	70300	165800	61000	SUV
2	53	Female	Salaried	Married	Post Graduate	3	No	No	Yes	97300	60700	158000	57000	SUV
3	53	Female	Salaried	Married	Graduate	2	Yes	No	Yes	72500	70300	142800	61000	SUV
4	53	Male	Salaried	Married	Post Graduate	3	No	No	Yes	79700	60200	139900	57000	SUV

Table 1: Top five rows of the dataset

2. While having a look on the data set information, it is found that there are 6 numerical and 8 categorial variables. The below table contains the same information.



```
Age                  int64
Gender              object
Profession          object
Marital_status      object
Education           object
No_of_Dependents     int64
Personal_loan       object
House_loan          object
Partner_working     object
Salary               int64
Partner_salary     float64
Total_salary         int64
Price                int64
Make                object
dtype: object
```

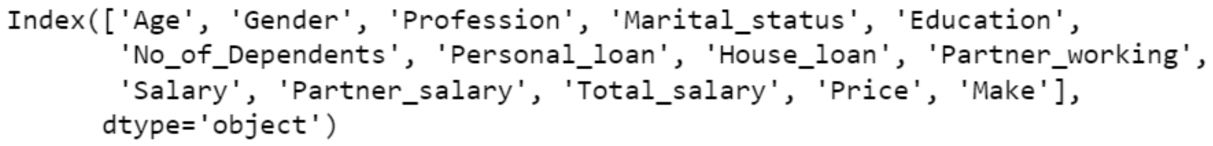
Table 2: Basic information of the data type

3. Checking the data information.

```
<bound method DataFrame.info of       Age  Gender Profession Marital_status      Educ
ation  No_of_Dependents  \
0      53    Male   Business        Married  Post Graduate                 4
1      53  Female   Salaried        Married  Post Graduate                 4
2      53  Female   Salaried        Married  Post Graduate                 3
3      53  Female   Salaried        Married       Graduate                 2
4      53    Male   Salaried        Married  Post Graduate                 3
...   ...     ...        ...            ...            ...               ...
1576   22    Male   Salaried         Single       Graduate                 2
1577   22    Male   Business        Married       Graduate                 4
1578   22    Male   Business         Single       Graduate                 2
1579   22    Male   Business        Married       Graduate                 3
1580   22    Male   Salaried        Married       Graduate                 4

     Personal_loan House_loan Partner_working  Salary  Partner_salary  \
0               No         No             Yes   99300         70700.0
1              Yes         No             Yes   95500         70300.0
2               No         No             Yes   97300         60700.0
3              Yes         No             Yes   72500         70300.0
4               No         No             Yes   79700         60200.0
...            ...        ...             ...     ...             ...
1576            No        Yes              No   33300             0.0
1577            No         No              No   32000             0.0
1578            No        Yes              No   32900             0.0
1579           Yes        Yes              No   32200             0.0
1580            No         No              No   31600             0.0

      Total_salary  Price       Make
0           170000  61000        SUV
1           165800  61000        SUV
2           158000  57000        SUV
3           142800  61000        SUV
4           139900  57000        SUV
...            ...    ...        ...
1576         33300  27000  Hatchback
1577         32000  31000  Hatchback
1578         32900  30000  Hatchback
1579         32200  24000  Hatchback
1580         31600  31000  Hatchback

[1581 rows x 14 columns]>
```

Table 3: Basic information of the data

4. Checking the columns of the dataset, to get the name of the variables.



```
Index(['Age', 'Gender', 'Profession', 'Marital_status', 'Education',
       'No_of_Dependents', 'Personal_loan', 'House_loan', 'Partner_working',
       'Salary', 'Partner_salary', 'Total_salary', 'Price', 'Make'],
      dtype='object')
```

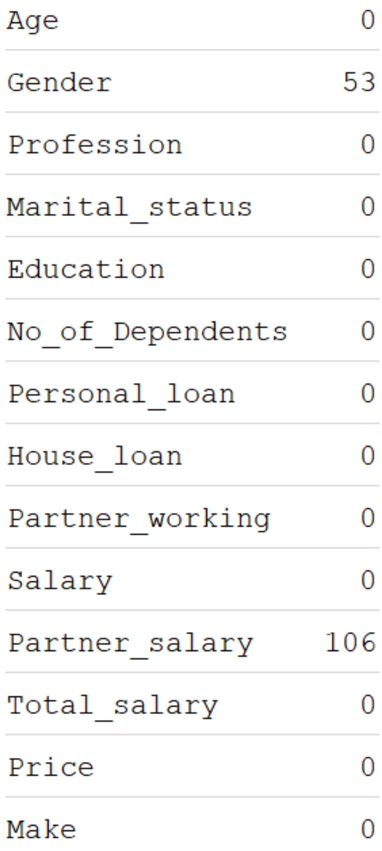
Table 4: Name of the columns present in the dataset

5. Checking the null values:

There are nulls in ‘Gender’ and ‘Partner_salary’ variables.

In ‘Gender’ it is found that there are total 53 null values.

In ‘Partner_salary’ it is found that there are total 106 null 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

Table 5: Inspecting null values in the dataset

6. In order to treat the nulls in the ‘Partner_salary’, we have checked where the ‘Total_salary’ is greater than ‘Salary’.

Then we applied condition that,

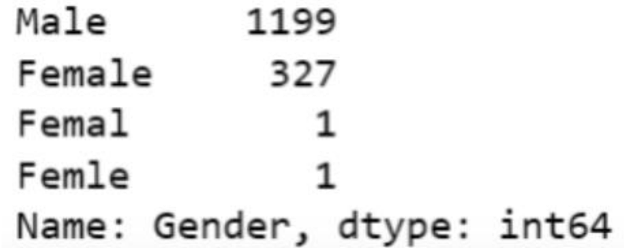
1. If , ‘Total_salary’ > ‘Salary’

then, ‘Partner_salary’ = ‘Total_salary’ - ‘Salary’

2. If, ‘Total_salary ≯ ‘Salary’

Then, ‘Partner_salary’ = 0

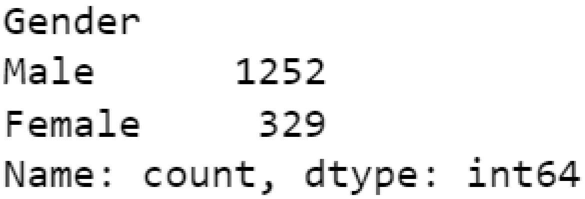
7. Checking the values count of ‘Gender’, and found that



```
Male      1199
Female     327
Femal        1
Femle        1
Name: Gender, dtype: int64
```

Table 6: Inspecting subcategories of Gender

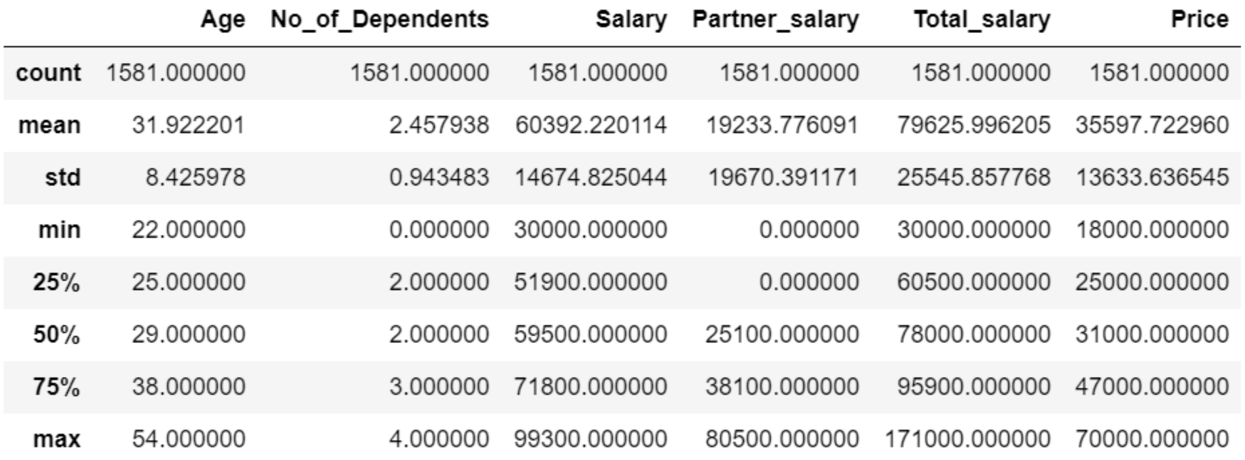
After treatment of the miss spellings in the ‘Gender’ and treating the null values, we got that



```
Gender
Male      1252
Female     329
Name: count, dtype: int64
```

Table 7: After imputing the Gender

8. Now we are to have a look on statistical summery of the numeric variables of the dataset.



	Age	No_of_Dependents	Salary	Partner_salary	Total_salary	Price
count	1581.000000	1581.000000	1581.000000	1581.000000	1581.000000	1581.000000
mean	31.922201	2.457938	60392.220114	19233.776091	79625.996205	35597.722960
std	8.425978	0.943483	14674.825044	19670.391171	25545.857768	13633.636545
min	22.000000	0.000000	30000.000000	0.000000	30000.000000	18000.000000
25%	25.000000	2.000000	51900.000000	0.000000	60500.000000	25000.000000
50%	29.000000	2.000000	59500.000000	25100.000000	78000.000000	31000.000000
75%	38.000000	3.000000	71800.000000	38100.000000	95900.000000	47000.000000
max	54.000000	4.000000	99300.000000	80500.000000	171000.000000	70000.000000

Table 8: Statistical summary of numeric variables

9. Observations:

1. The customers are between 22 and 54 years old. They could be considered to be in the working age group. The median age is 29, and the mean age is 31.92.
2. The clients' salaries vary from 30K to 99.3K. There is a very small difference between the mean and median values.
3. The minimum purchase value of the automobile is 18k and maximum value is 70k.

10. Checking the value counts of categorial variables.

```
Value counts for Gender:
Gender
Male      1252
Female     329
Name: count, dtype: int64

Value counts for Profession:
Profession
Salaried    896
Business    685
Name: count, dtype: int64

Value counts for Marital_status:
Marital_status
Married    1443
Single      138
Name: count, dtype: int64

Value counts for Education:
Education
Post Graduate    985
Graduate         596
Name: count, dtype: int64

Value counts for Personal_loan:
Personal_loan
Yes    792
No     789
Name: count, dtype: int64

Value counts for House_loan:
House_loan
No     1054
Yes     527
Name: count, dtype: int64

Value counts for Partner_working:
Partner_working
Yes    868
No     713
Name: count, dtype: int64

Value counts for Make:
Make
Sedan        702
Hatchback    582
SUV          297
```

```
Name: count, dtype: int64
```

Table 9: Value counts of categorial variables

11. Checking for the outliers or extreme values.

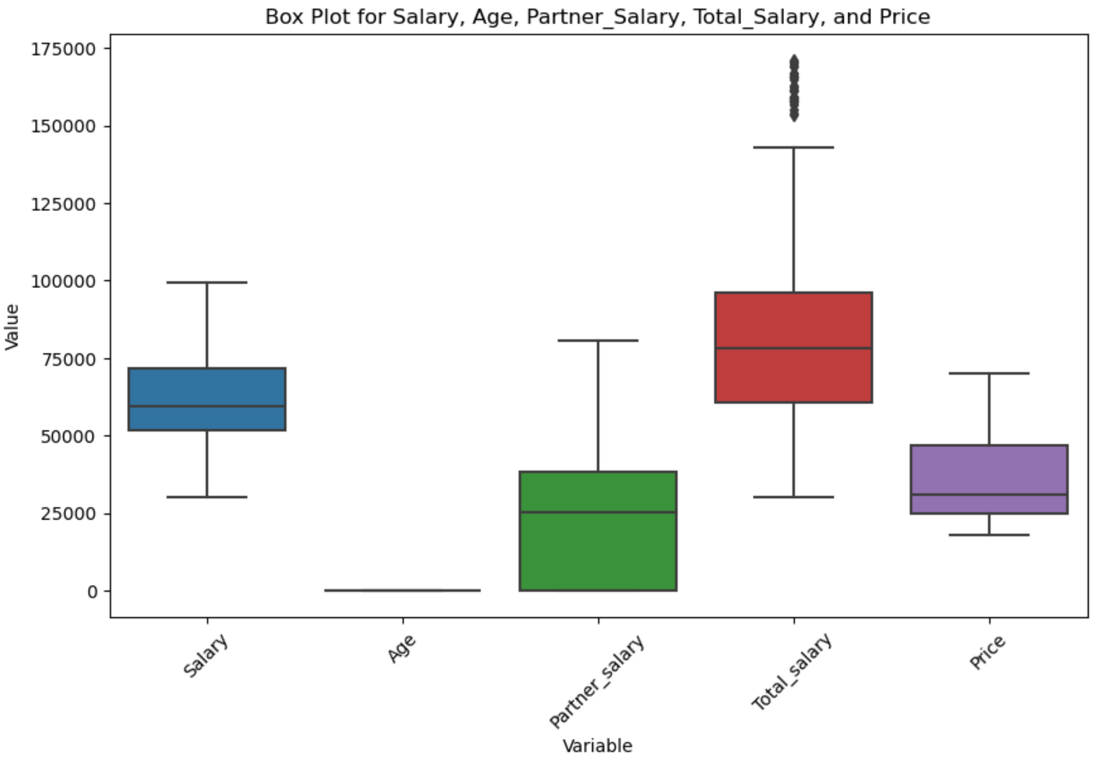
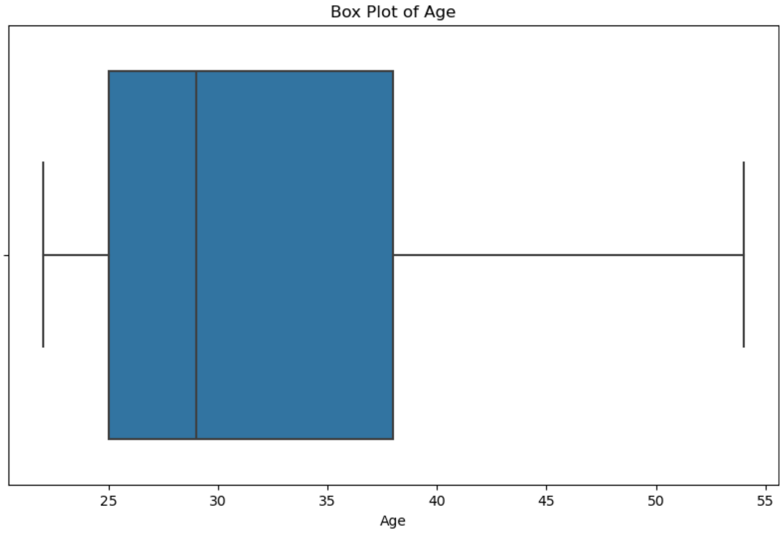
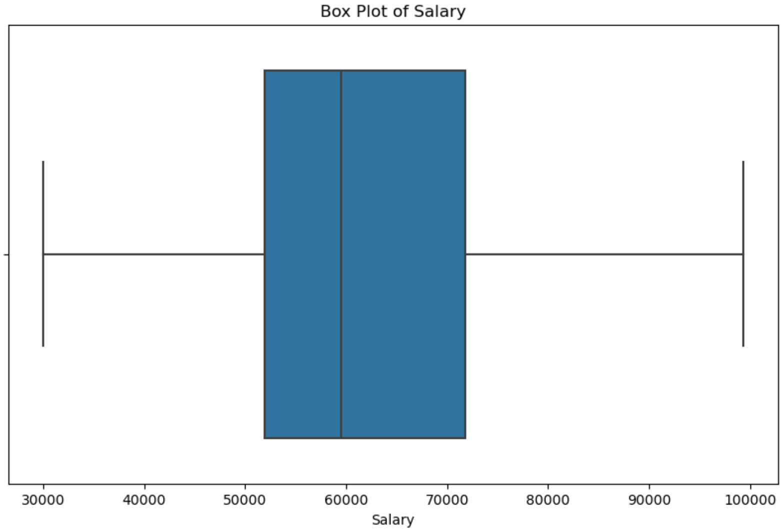


Figure-1 : Box plots of numerical variables

Analyzing box plots of every numerical variables separately:





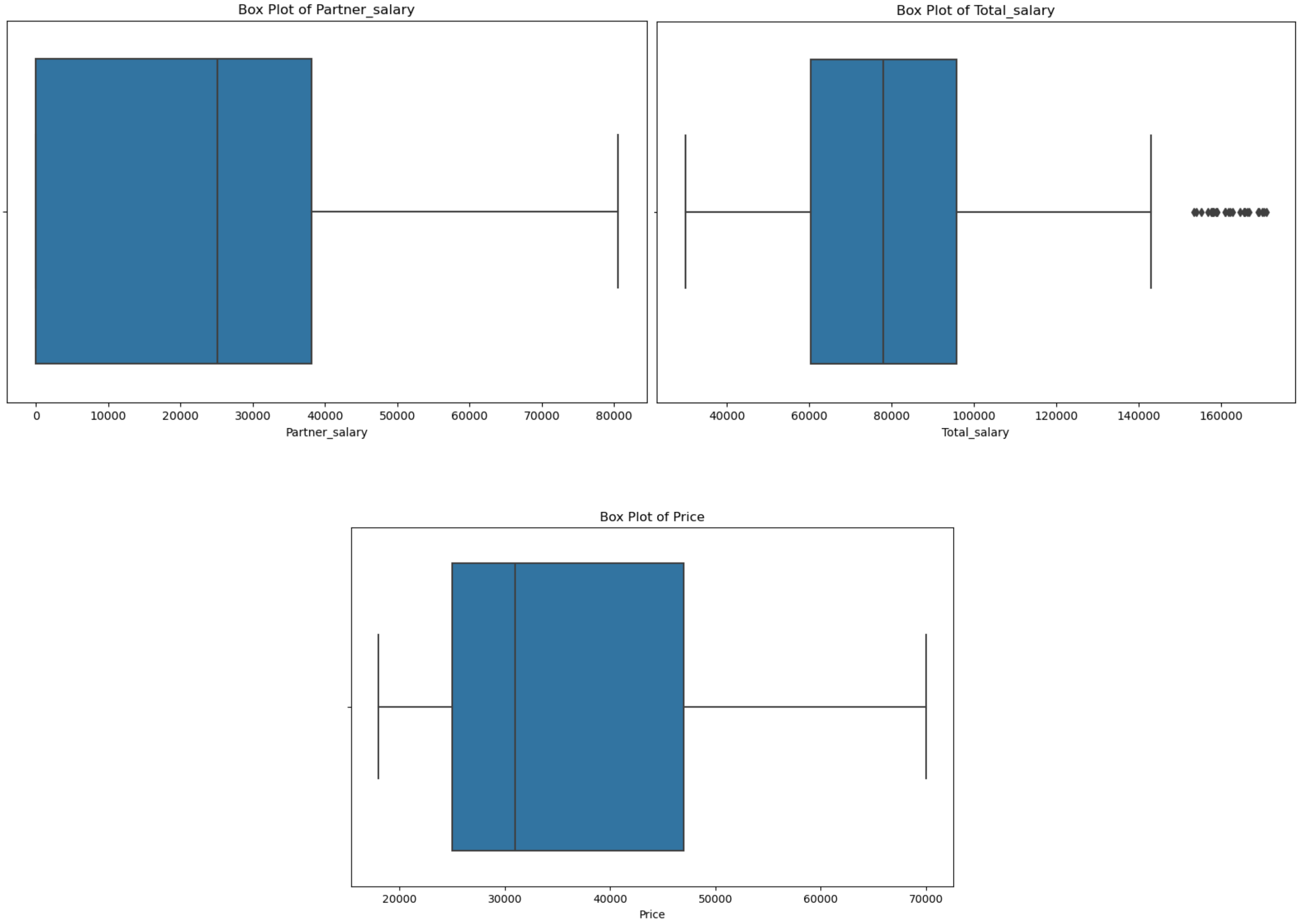
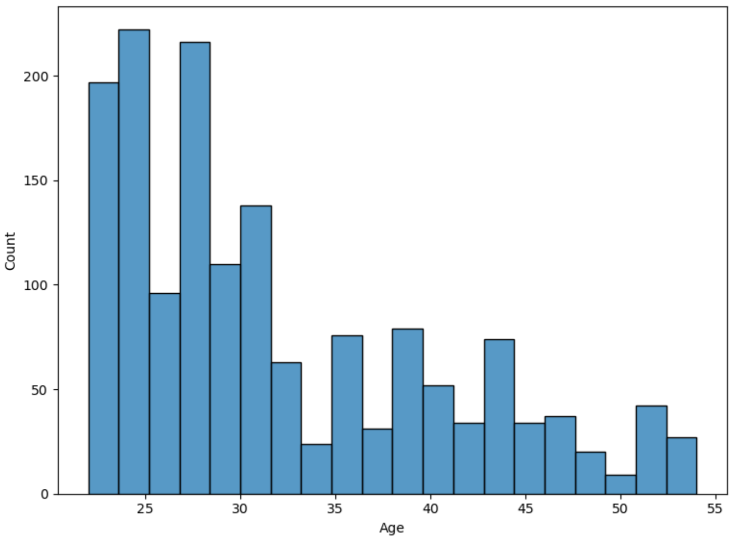
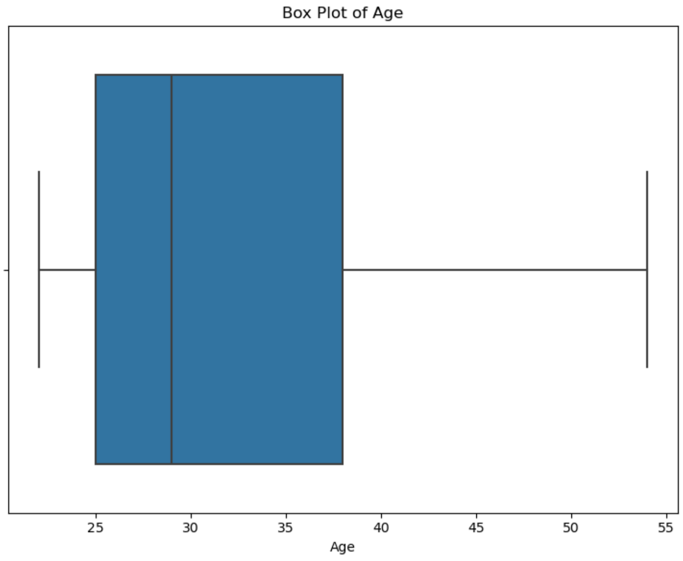


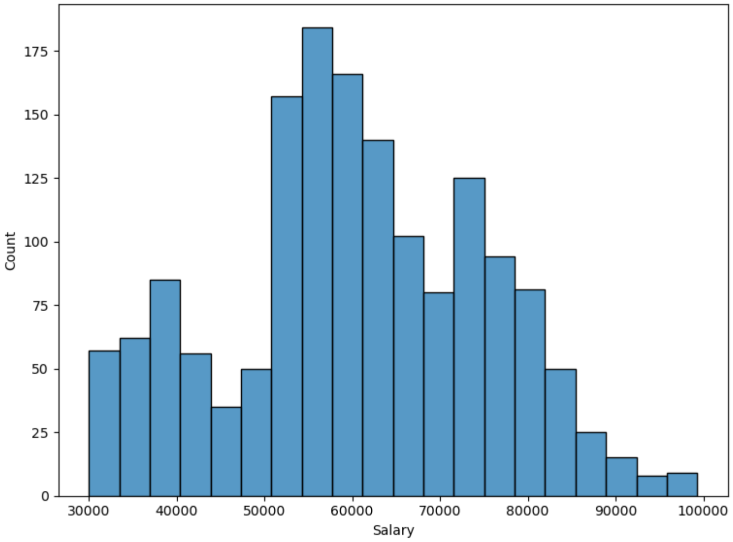
Figure-2 : Box plots of numerical variables individually

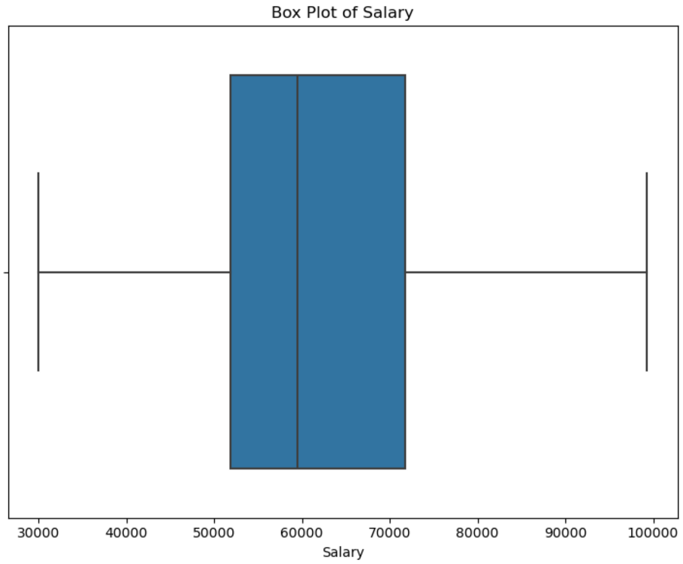
1. We can see that there are no negative values present in any numerical category.
2. The ‘Total_salary’ is having outlier.

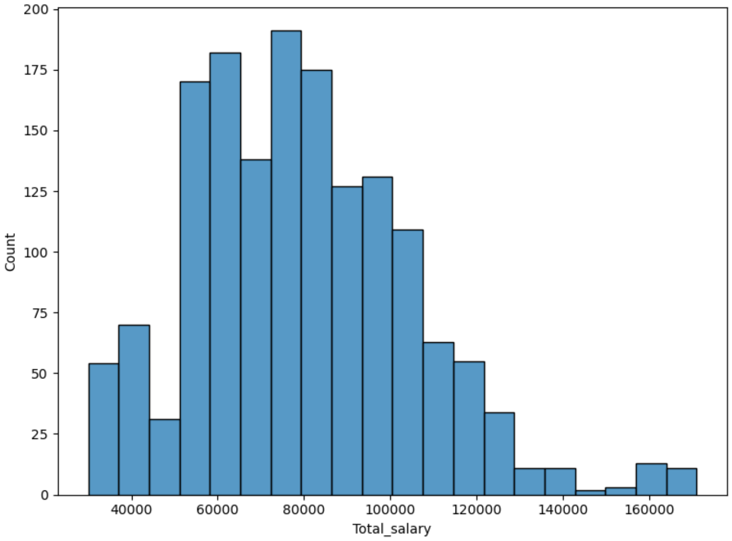
12. Univariate analysis of numerical variables.

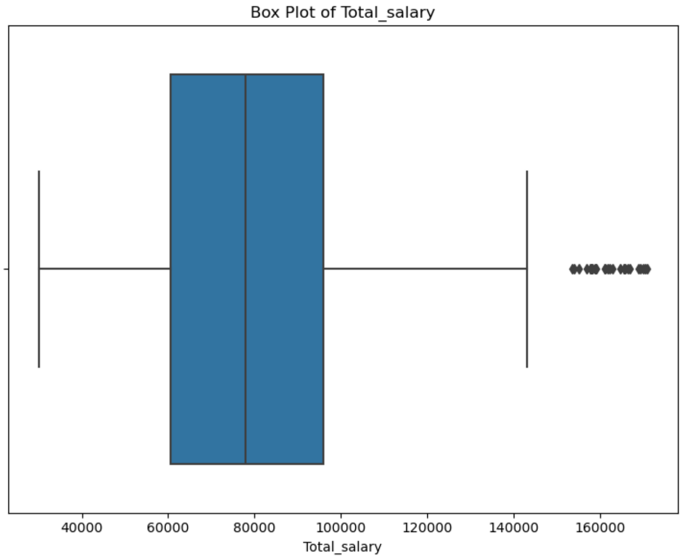

Count
0
50
100
150
200
25
30
35
40
45
50
55
Age


Box Plot of Age
25
30
35
40
45
50
55
Age


Count
0
25
50
75
100
125
150
175
30000
40000
50000
60000
70000
80000
90000
100000
Salary


Box Plot of Salary
30000
40000
50000
60000
70000
80000
90000
100000
Salary


Count
0
25
50
75
100
125
150
175
200
40000
60000
80000
100000
120000
140000
160000
Total_salary


Box Plot of Total_salary
40000
60000
80000
100000
120000
140000
160000
Total_salary

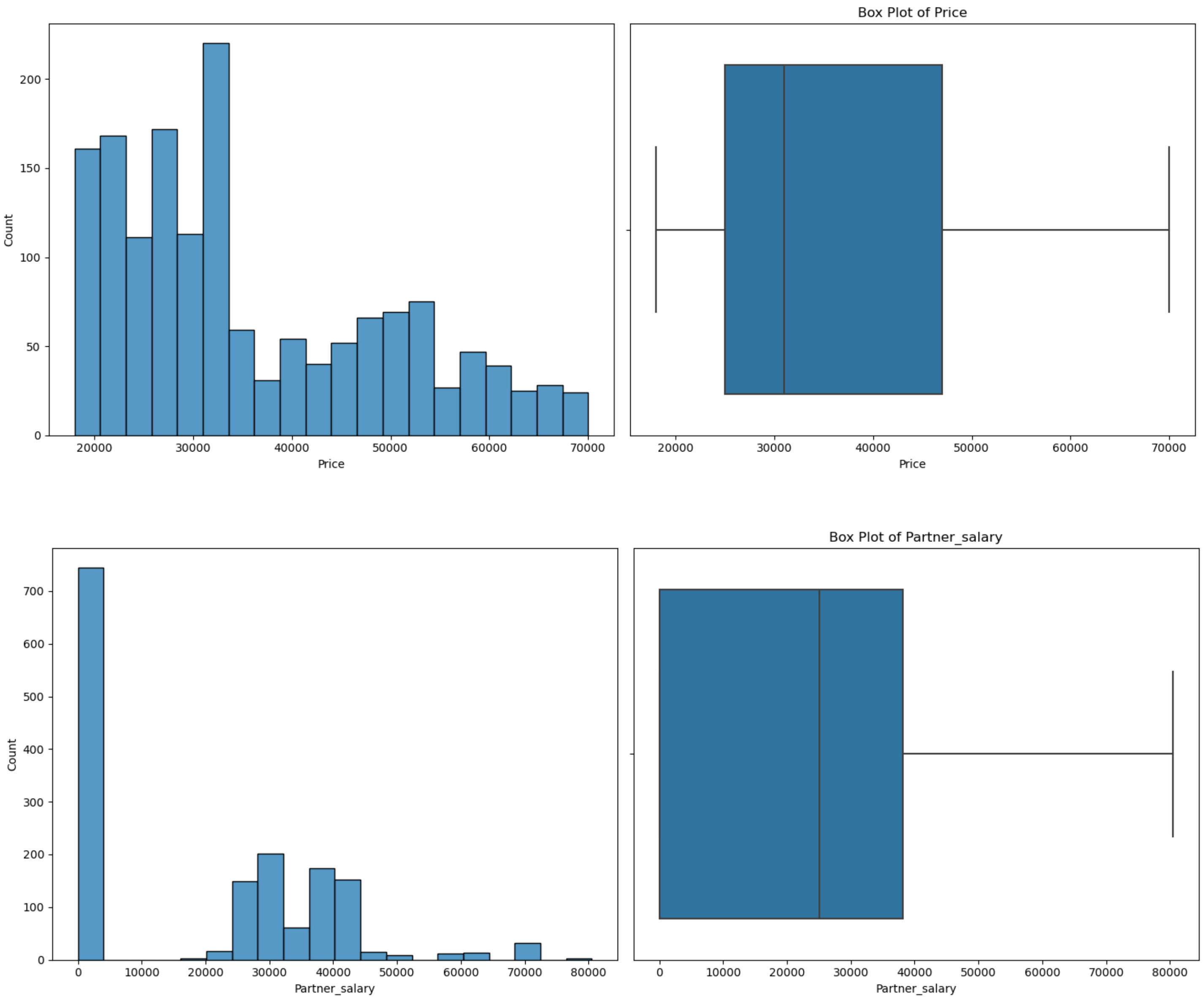
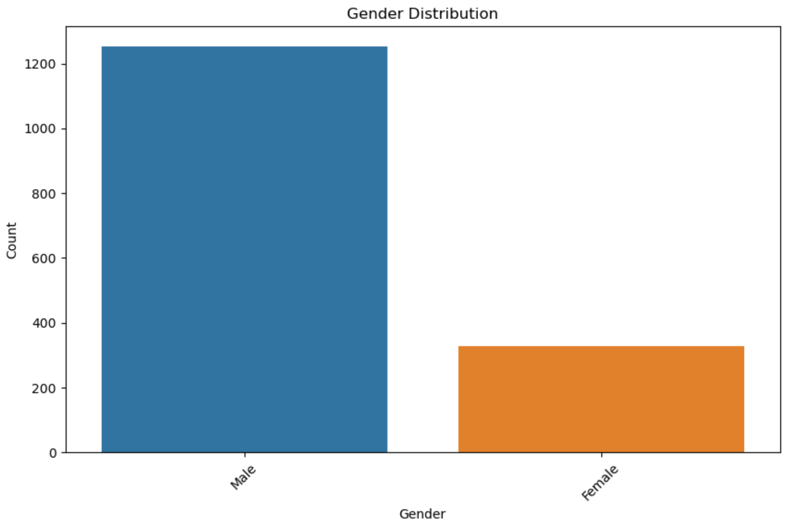


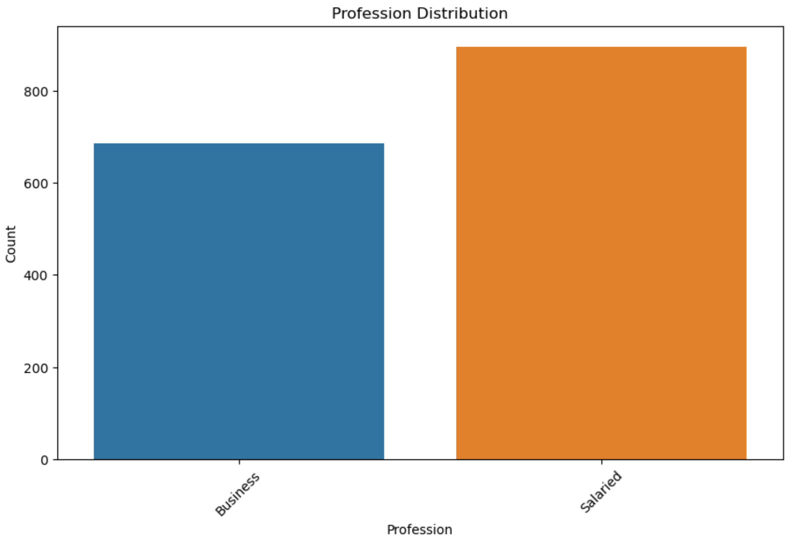
Figure-3 : Univariate analysis of numerical variables

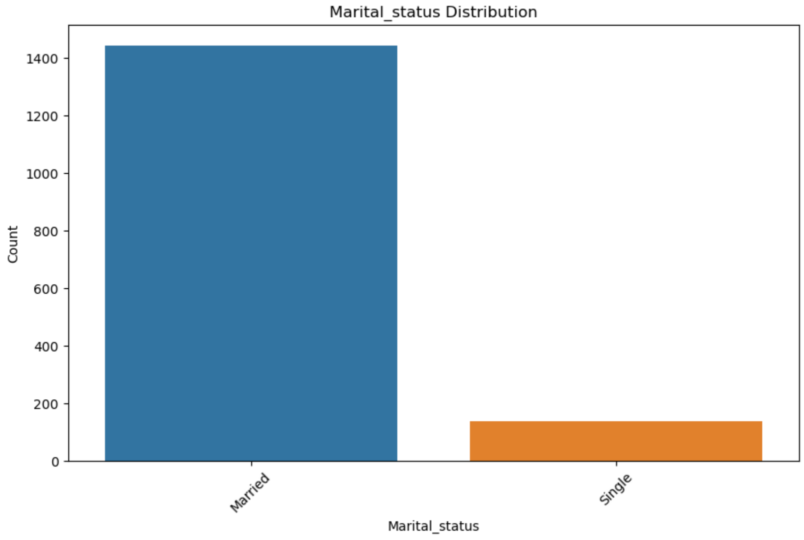
Inferences:

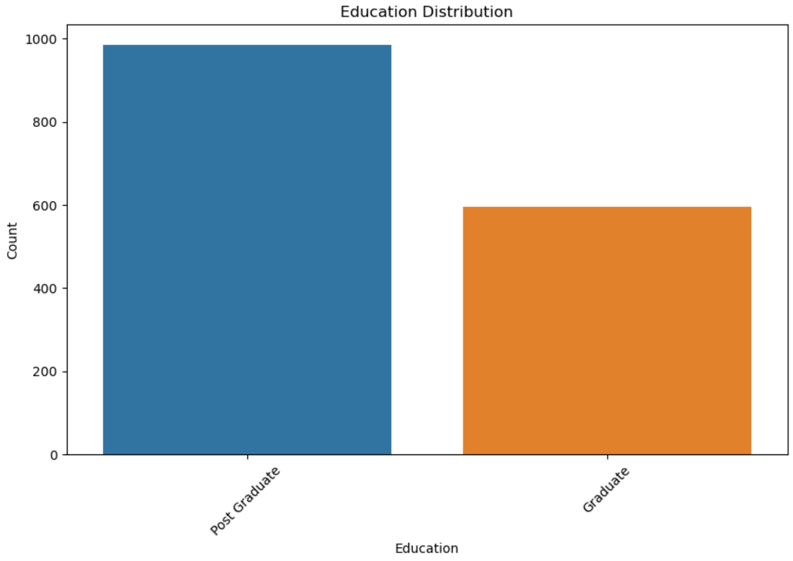
1. Salary has a range between 50k to 70k.
2. Total salary has a range between 60k to 100k.

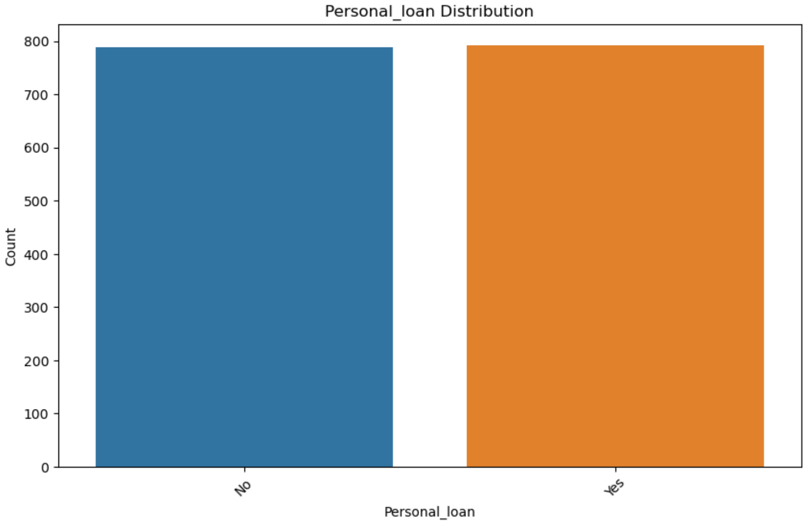
13. Univariate analysis of categorial

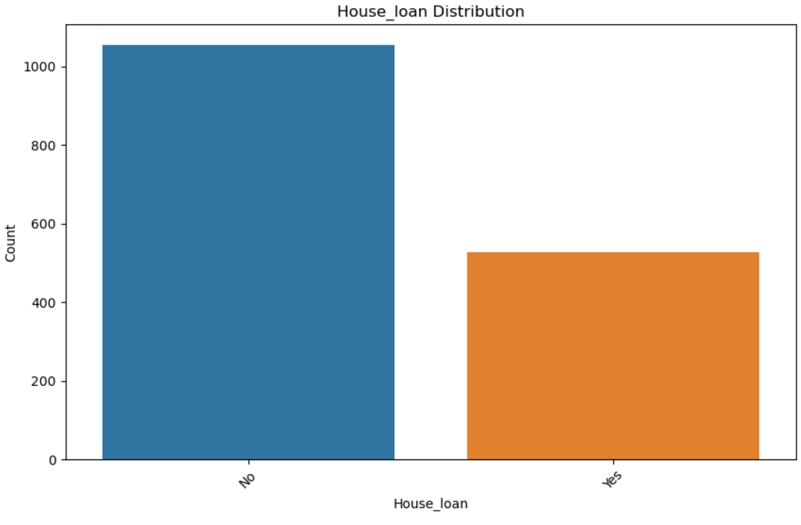

Gender Distribution


Profession Distribution


Marital_status Distribution


Education Distribution


Personal_loan Distribution


House_loan Distribution

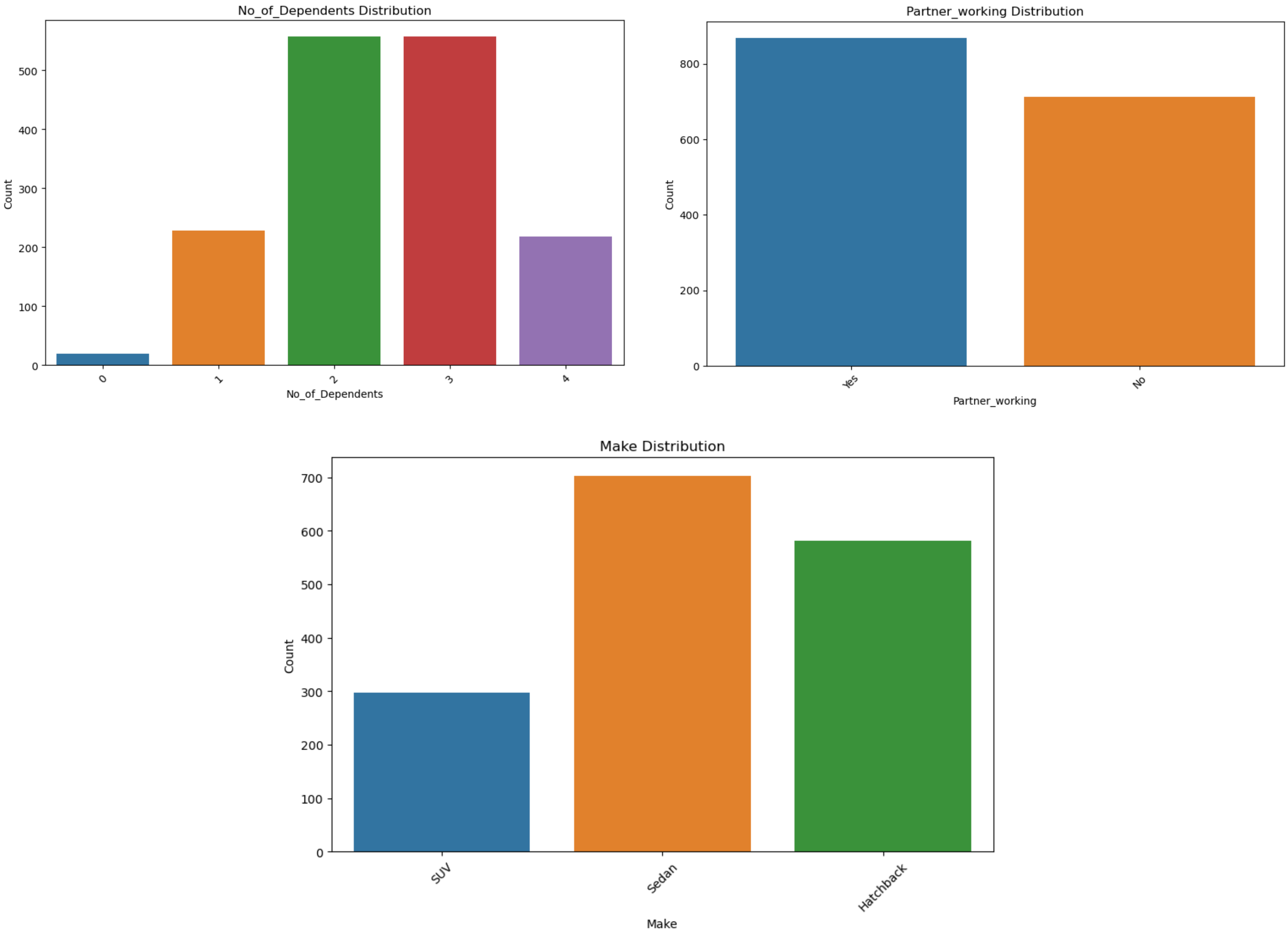


Figure-4 : Univariate analysis of categorial variables

Inferences:

1. Sedan is most preferred, after Hatchback and SUV respectively.
2. The buyers with working partner are higher than the buyers with non-working partners or single status.
3. The married buyers are very higher than the single status.
4. Major of the buyers are having postgraduate.
5. Buyers having business are little less than the number of buyers being salaried.
6. The buyers with having 2-3 dependents are higher in the dataset. Then comes the buyers with 1 & 4 dependents and the buyers having 0 dependents are very less.

14. Bivariate analysis of all the numerical variables.

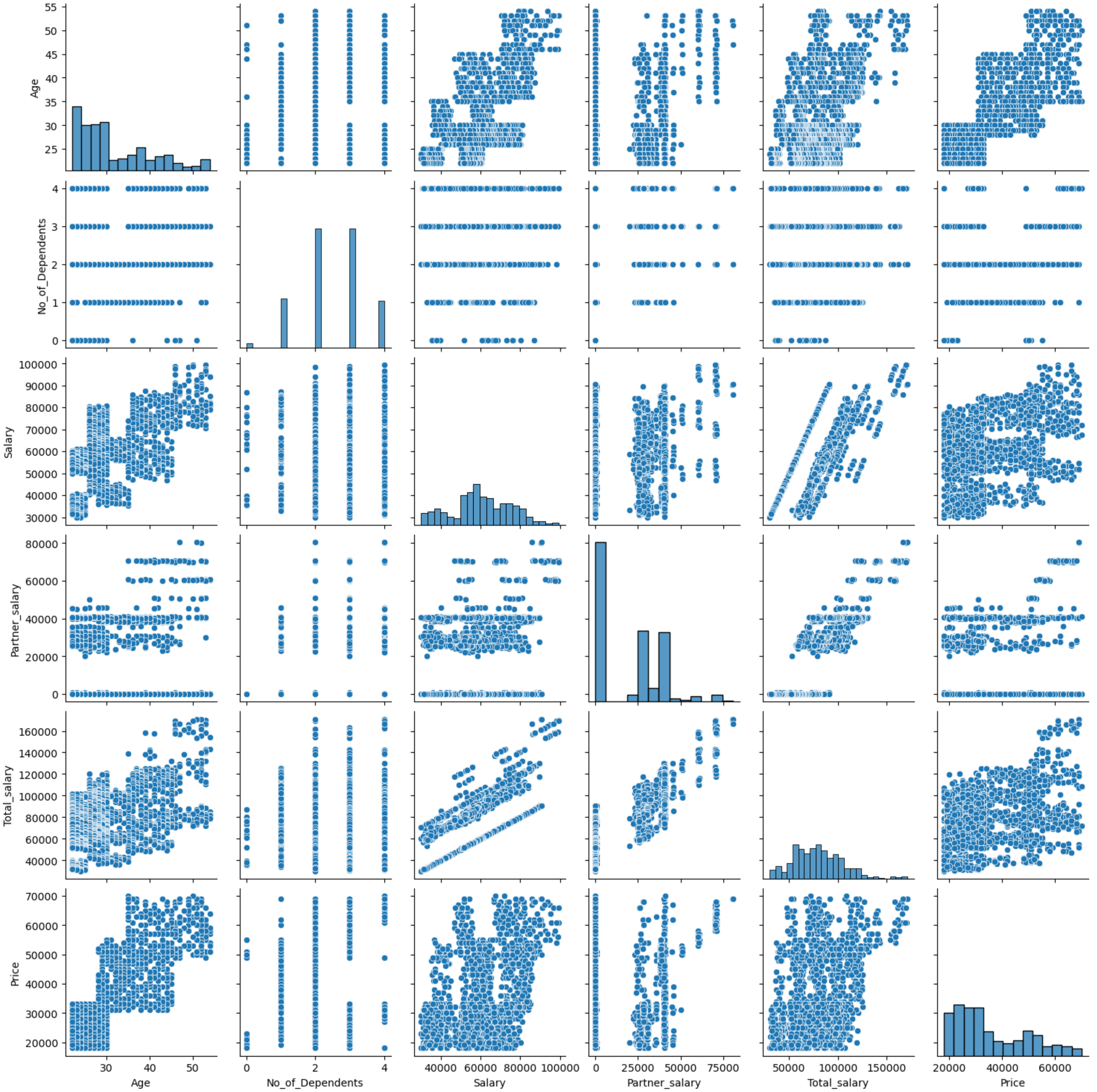


Figure-5 : Pair plot of the dataset numerical variables

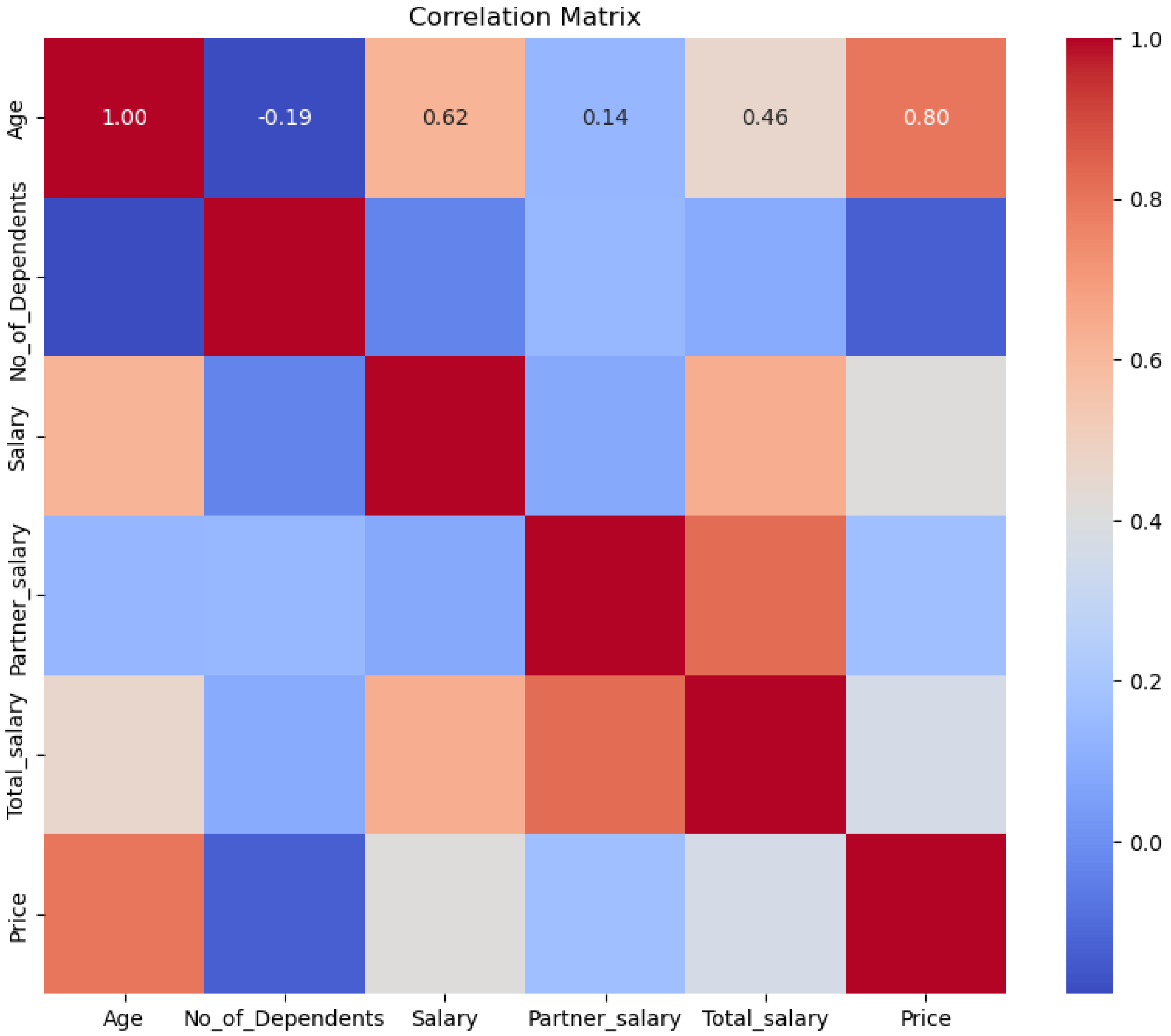
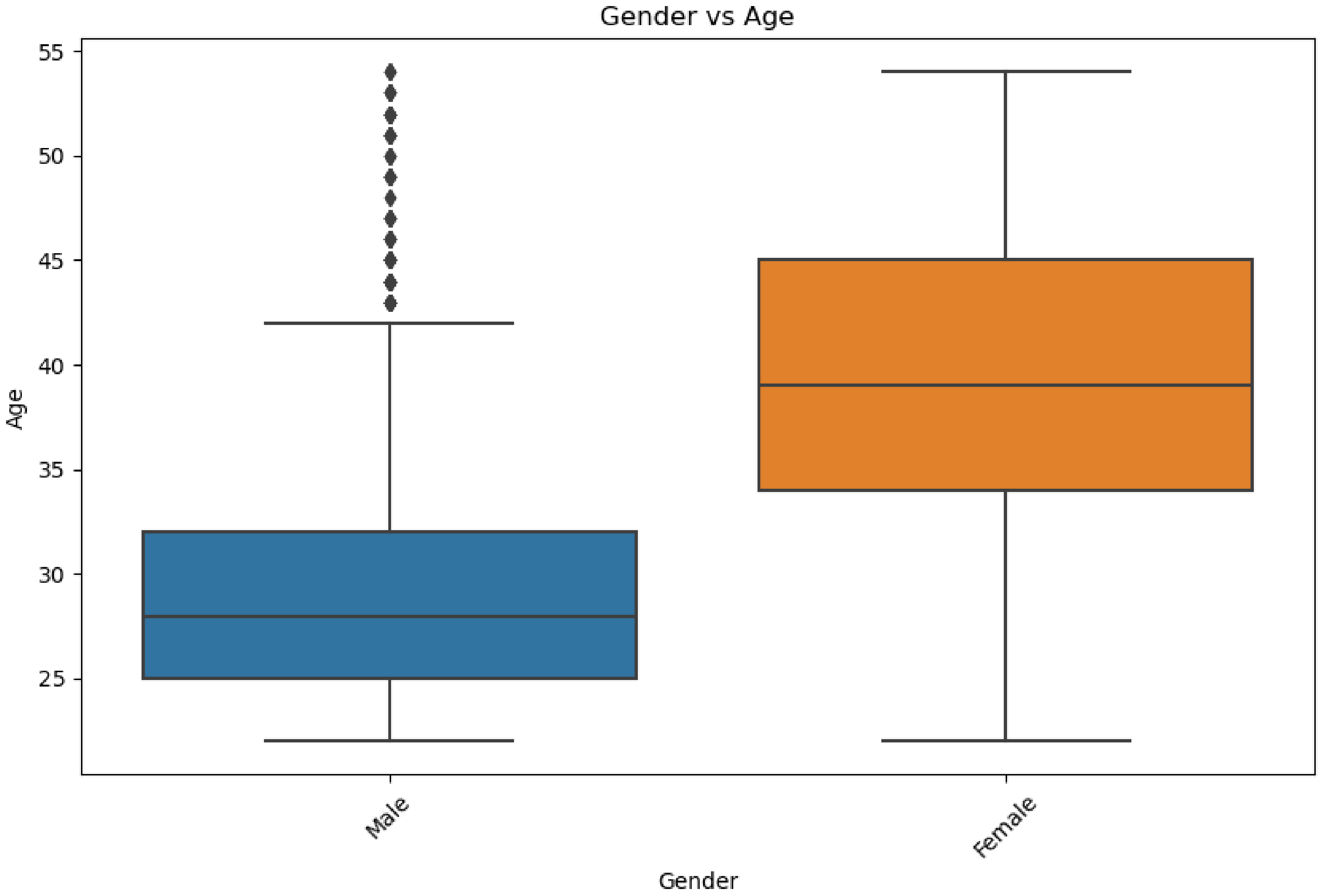
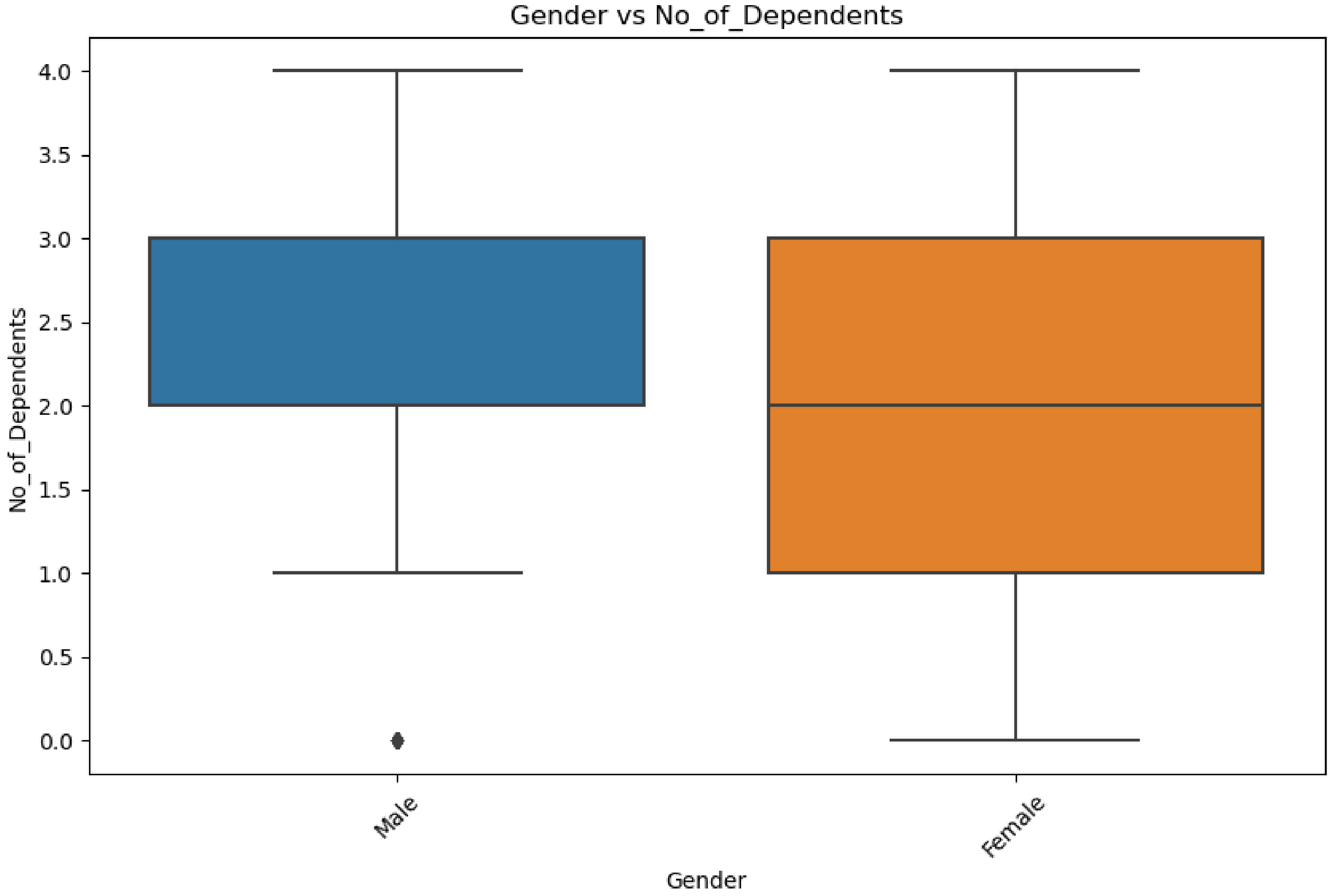
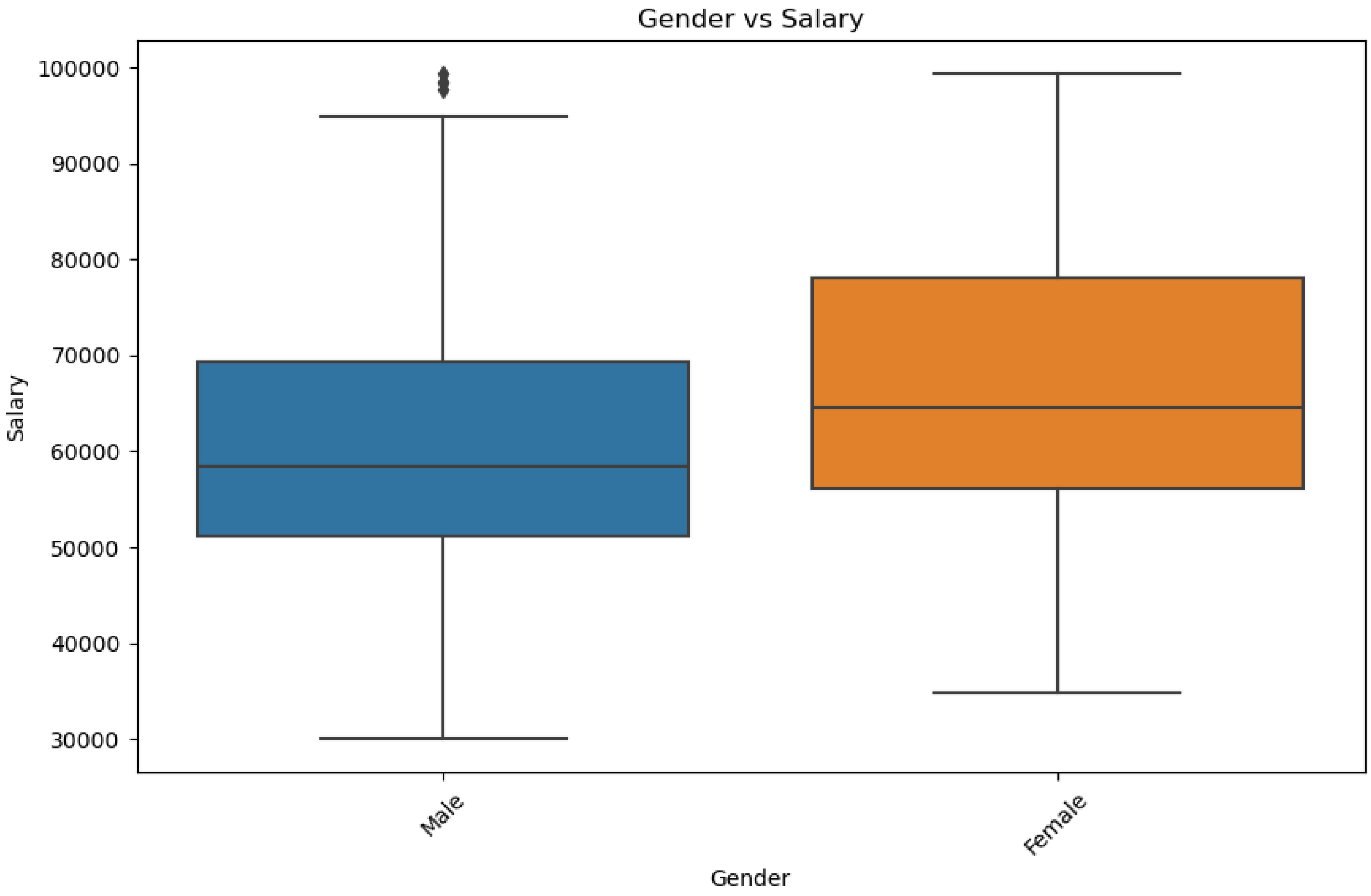


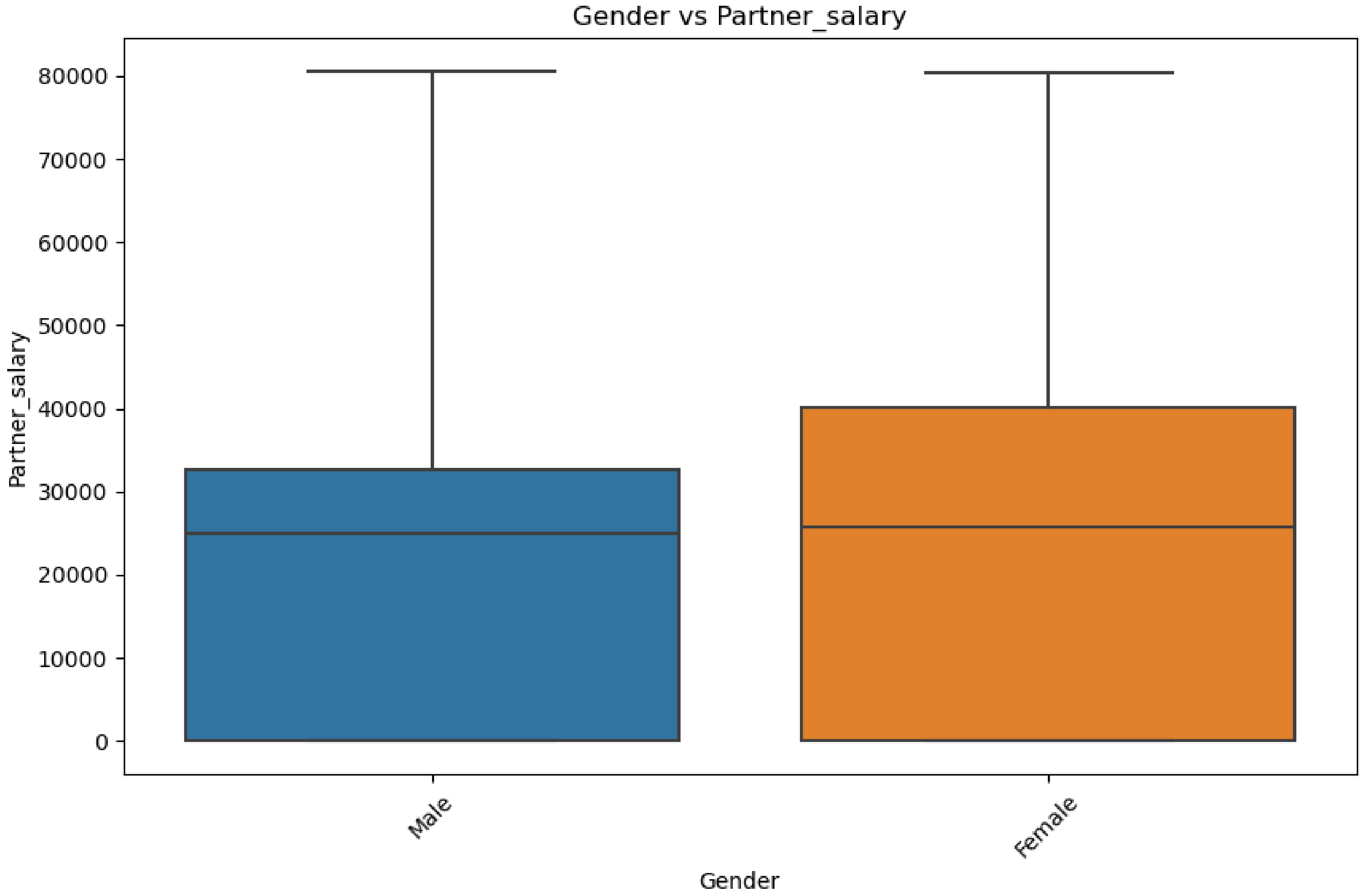
Figure-6 : Correlation heatmap of numerical variables

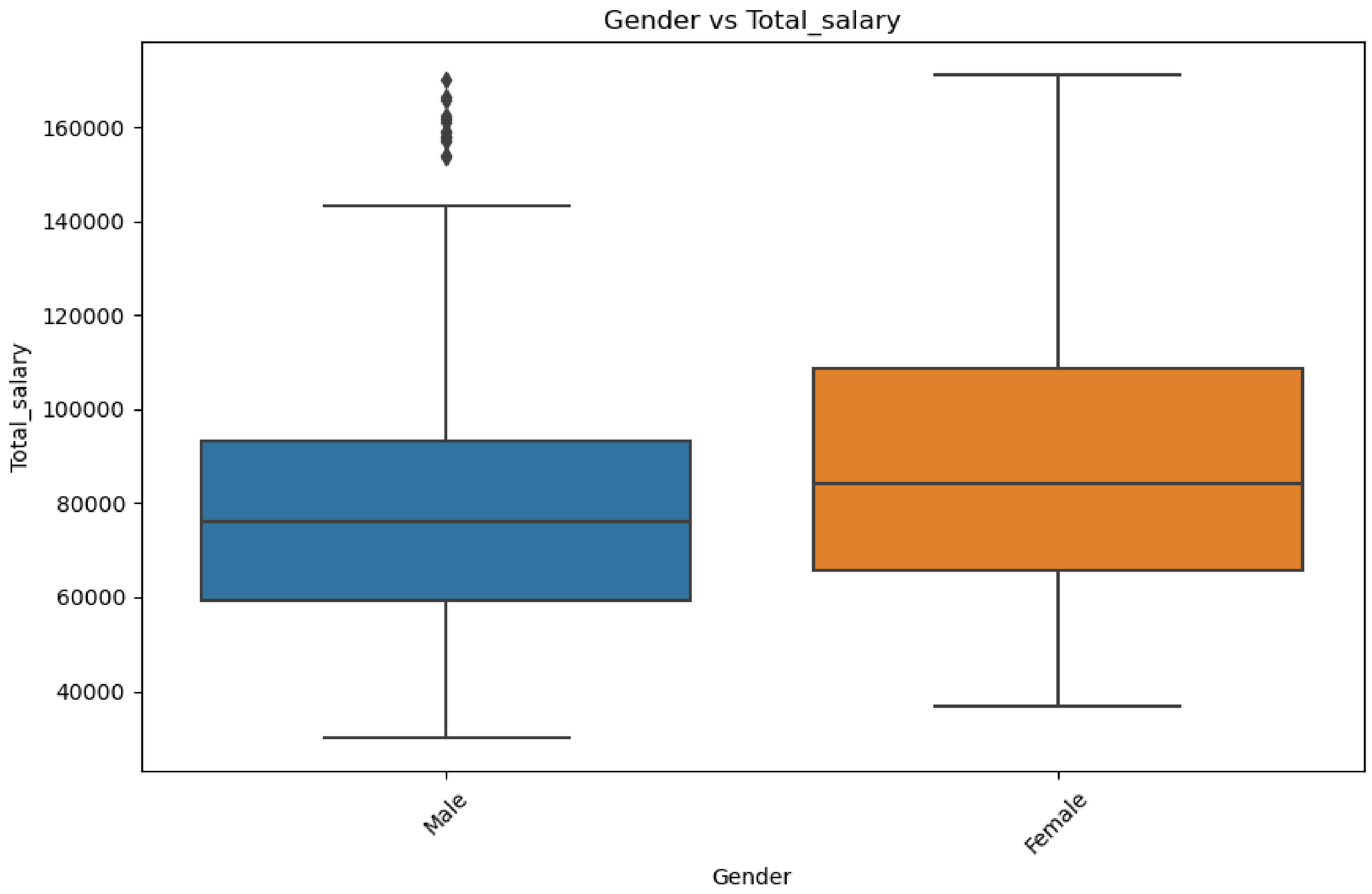
14. Bivariate analysis of all the categorial vs numerical variables.

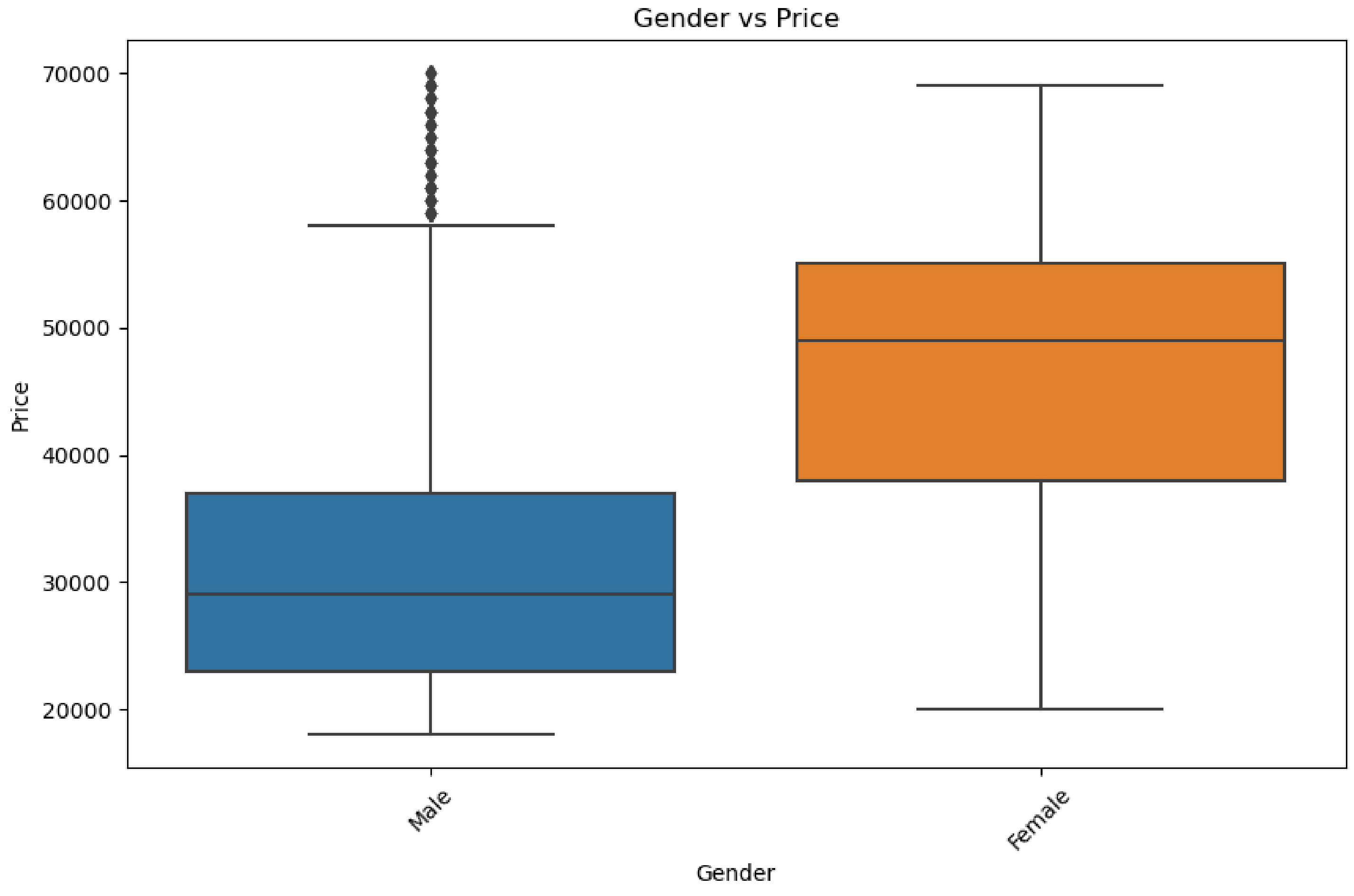


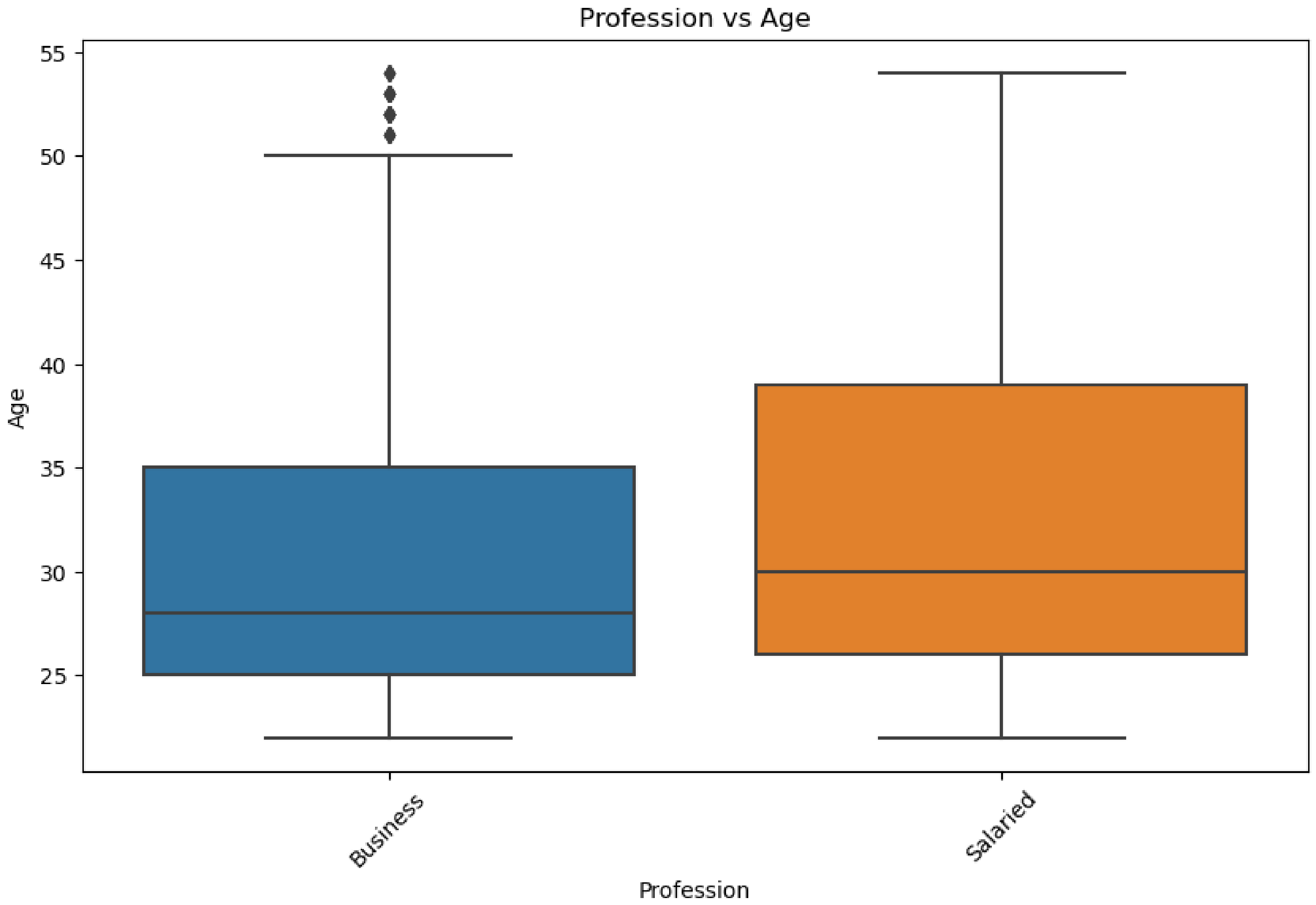


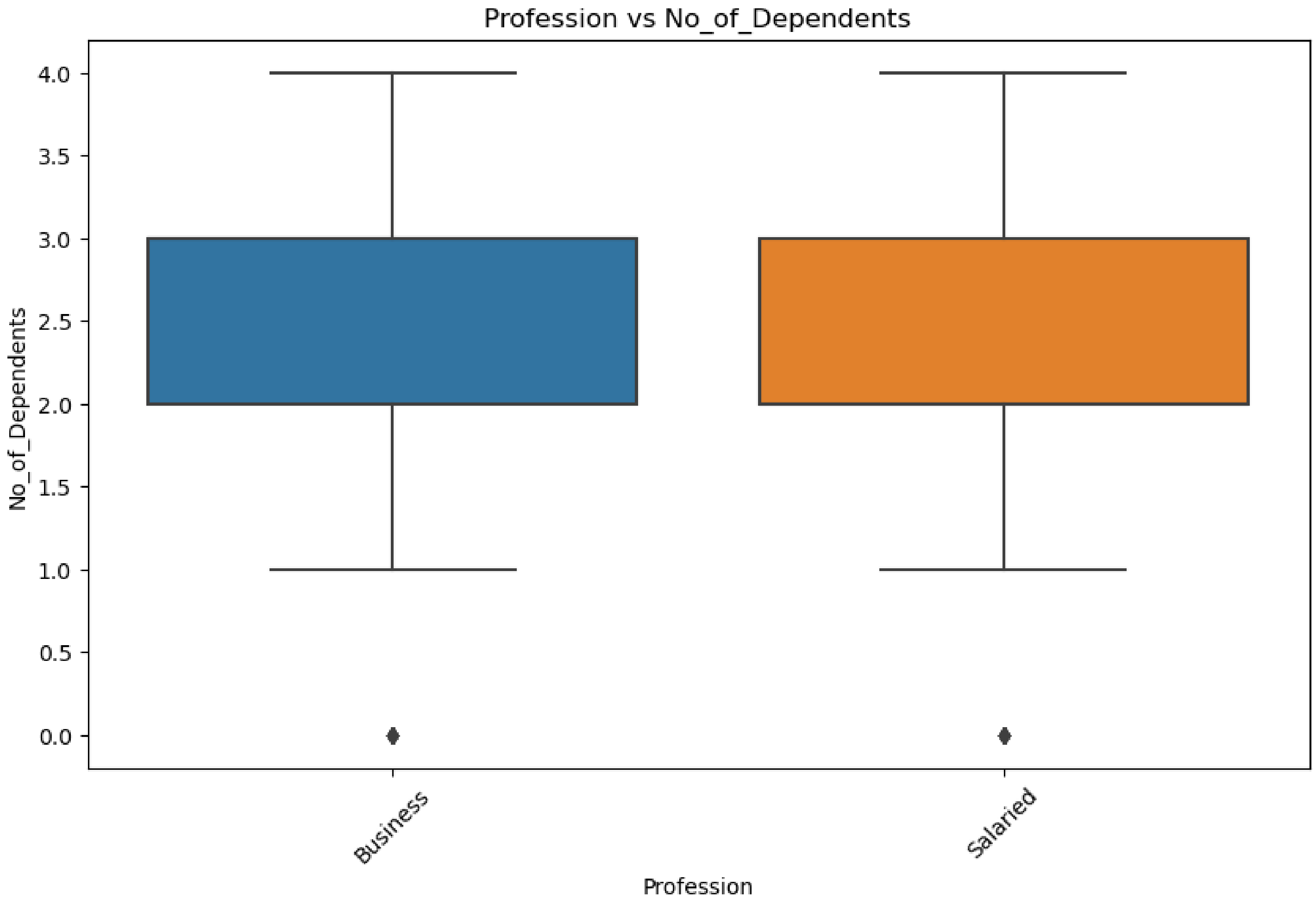

Gender vs Salary
Salary
100000
90000
80000
70000
60000
50000
40000
30000
Male
Female
Gender

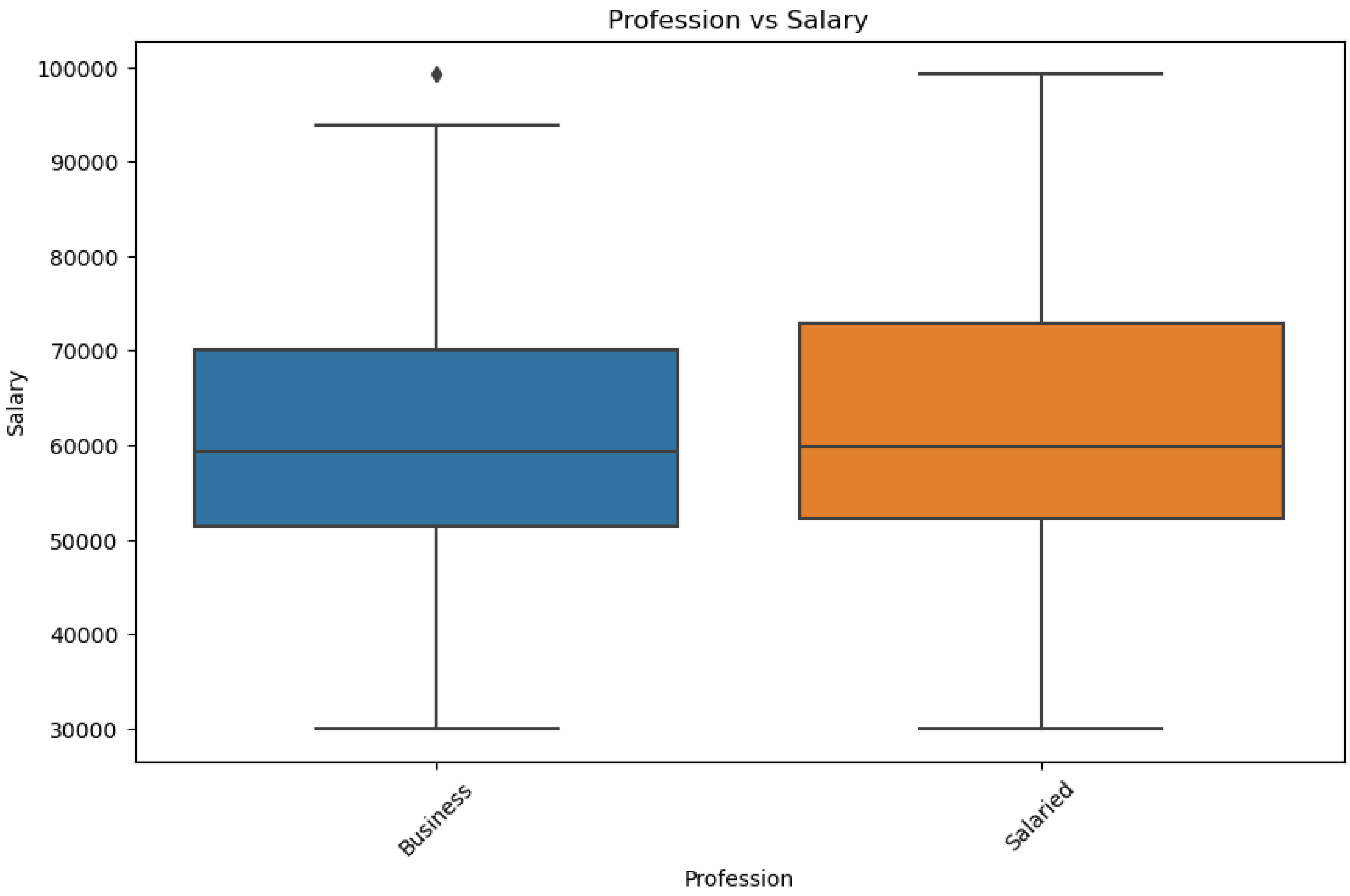

Gender vs Partner_salary
Partner_salary
80000
70000
60000
50000
40000
30000
20000
10000
0
Male
Female
Gender

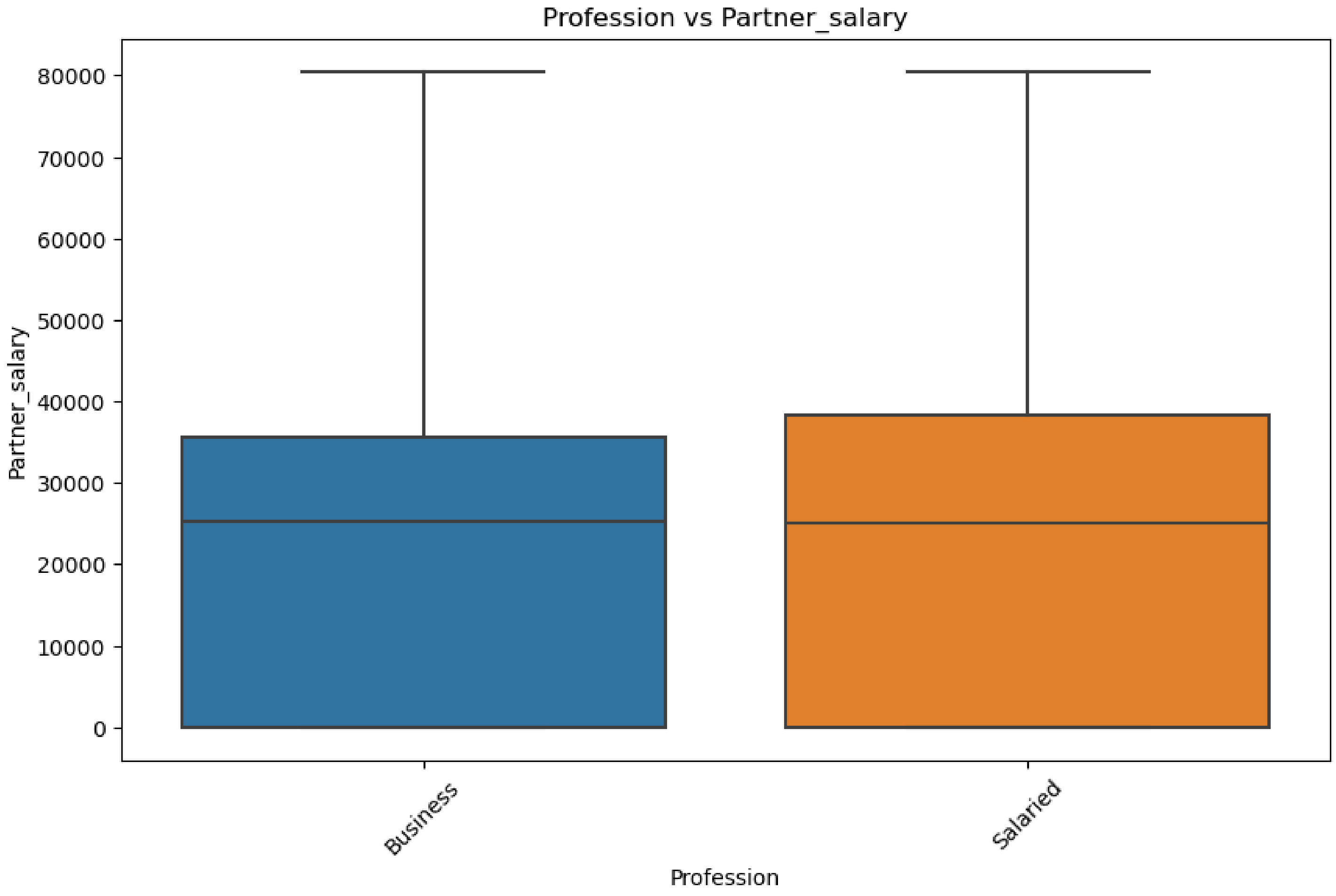

Gender vs Total_salary
Total_salary
160000
140000
120000
100000
80000
60000
40000
Male
Female
Gender

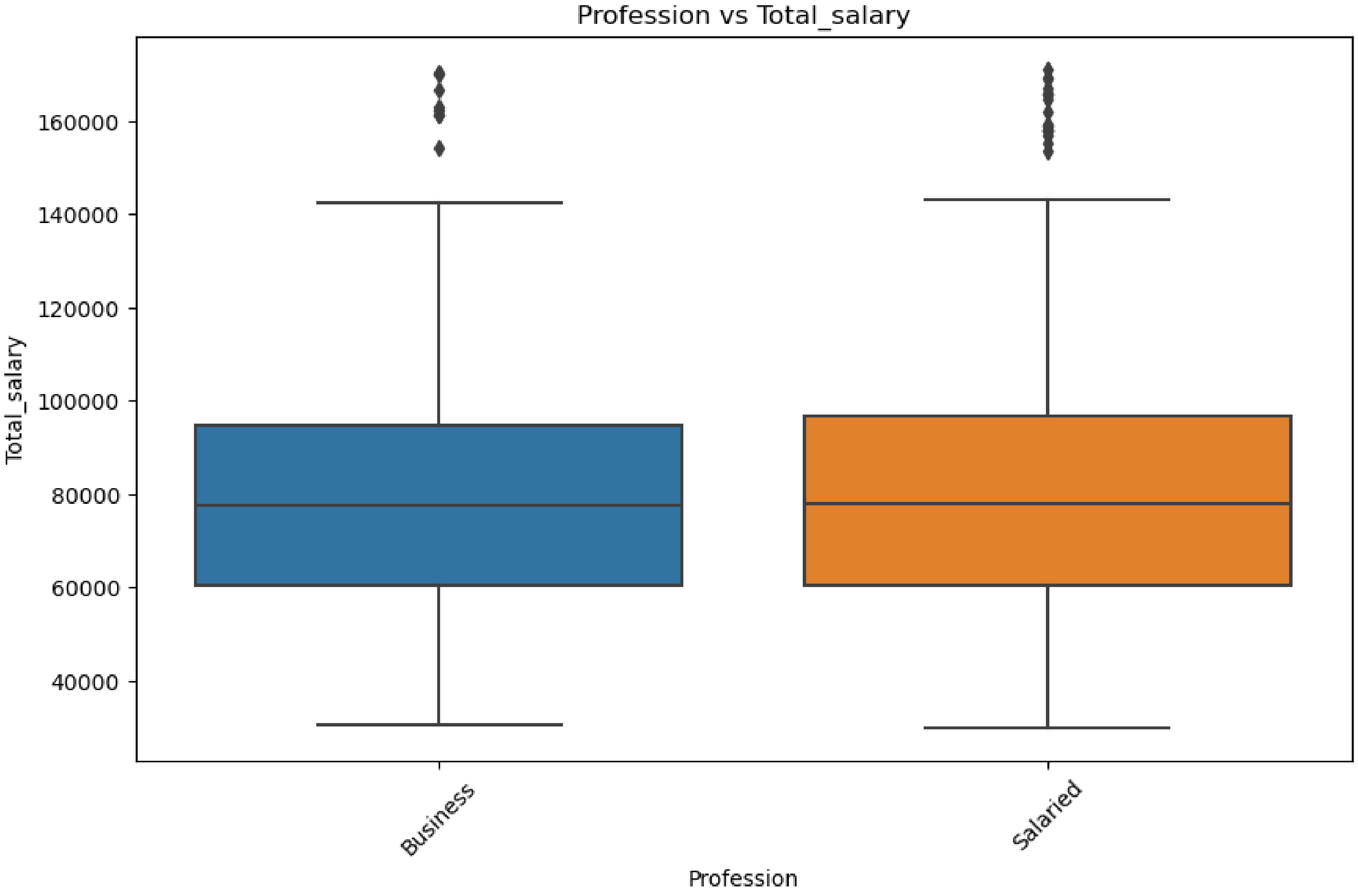

Gender vs Price
Price
70000
60000
50000
40000
30000
20000
Male
Female
Gender

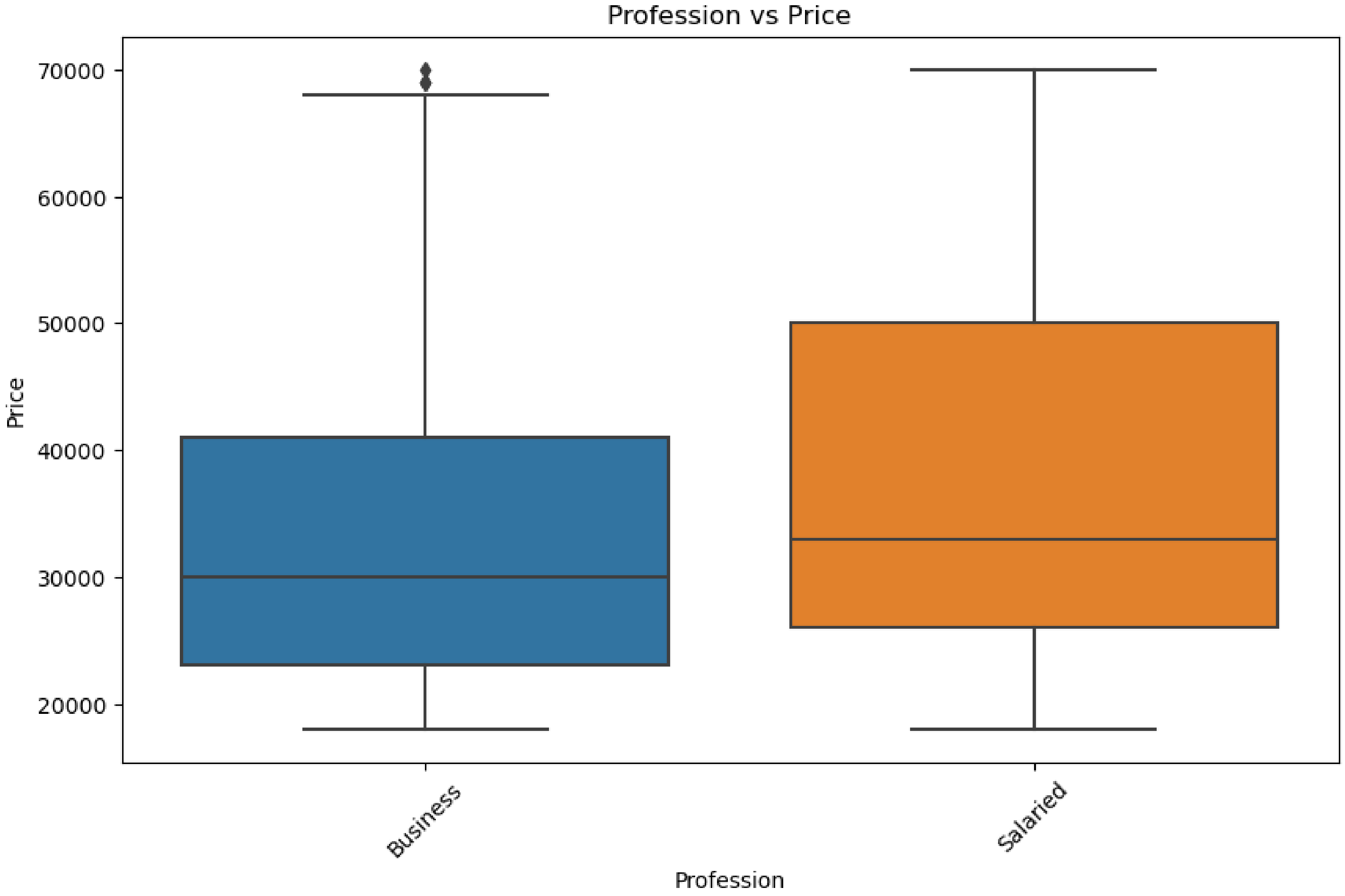

Profession vs Age
Age
55
50
45
40
35
30
25
Business
Salaried
Profession

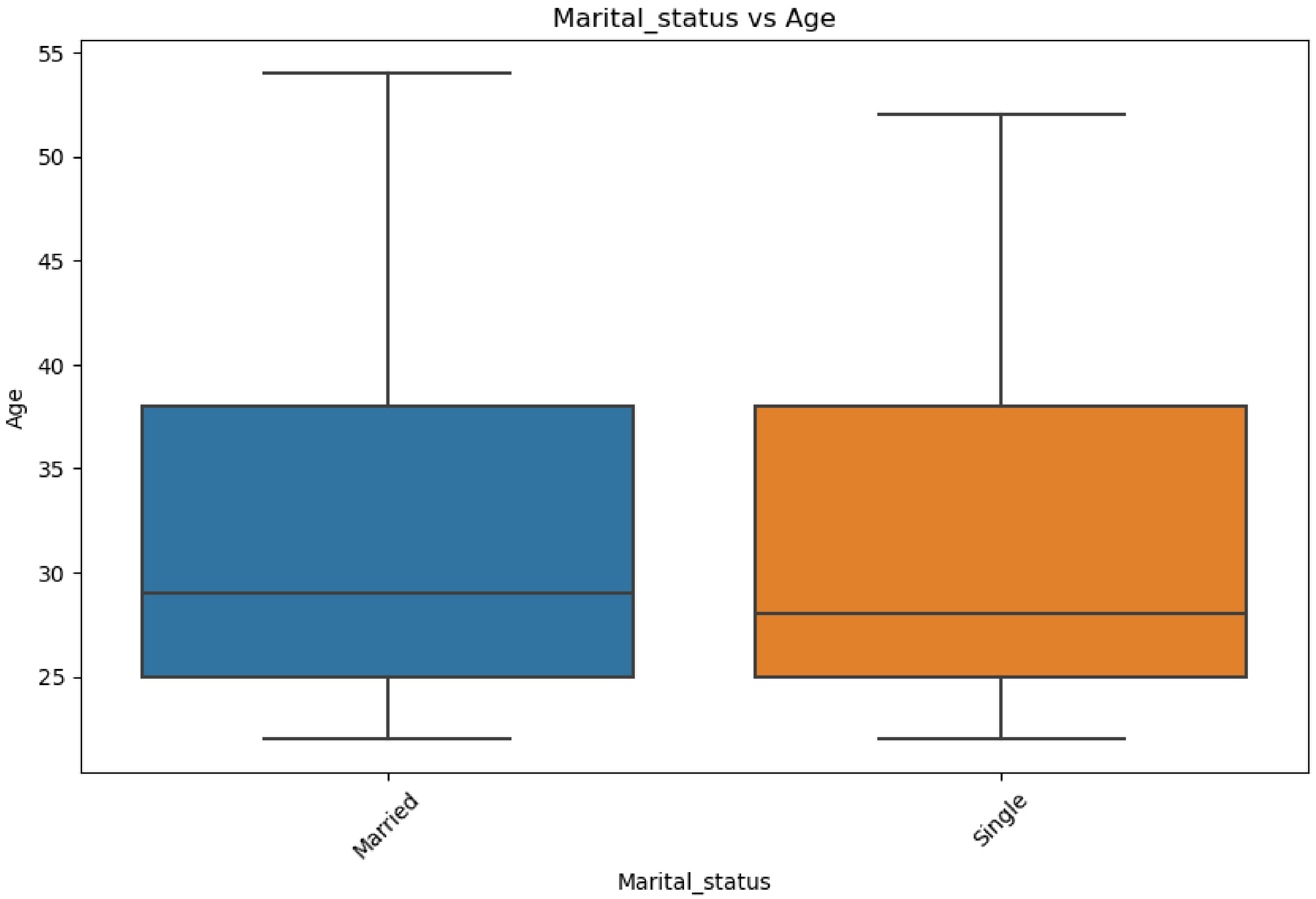

Profession vs No_of_Dependents
No_of_Dependents
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
Business
Salaried
Profession

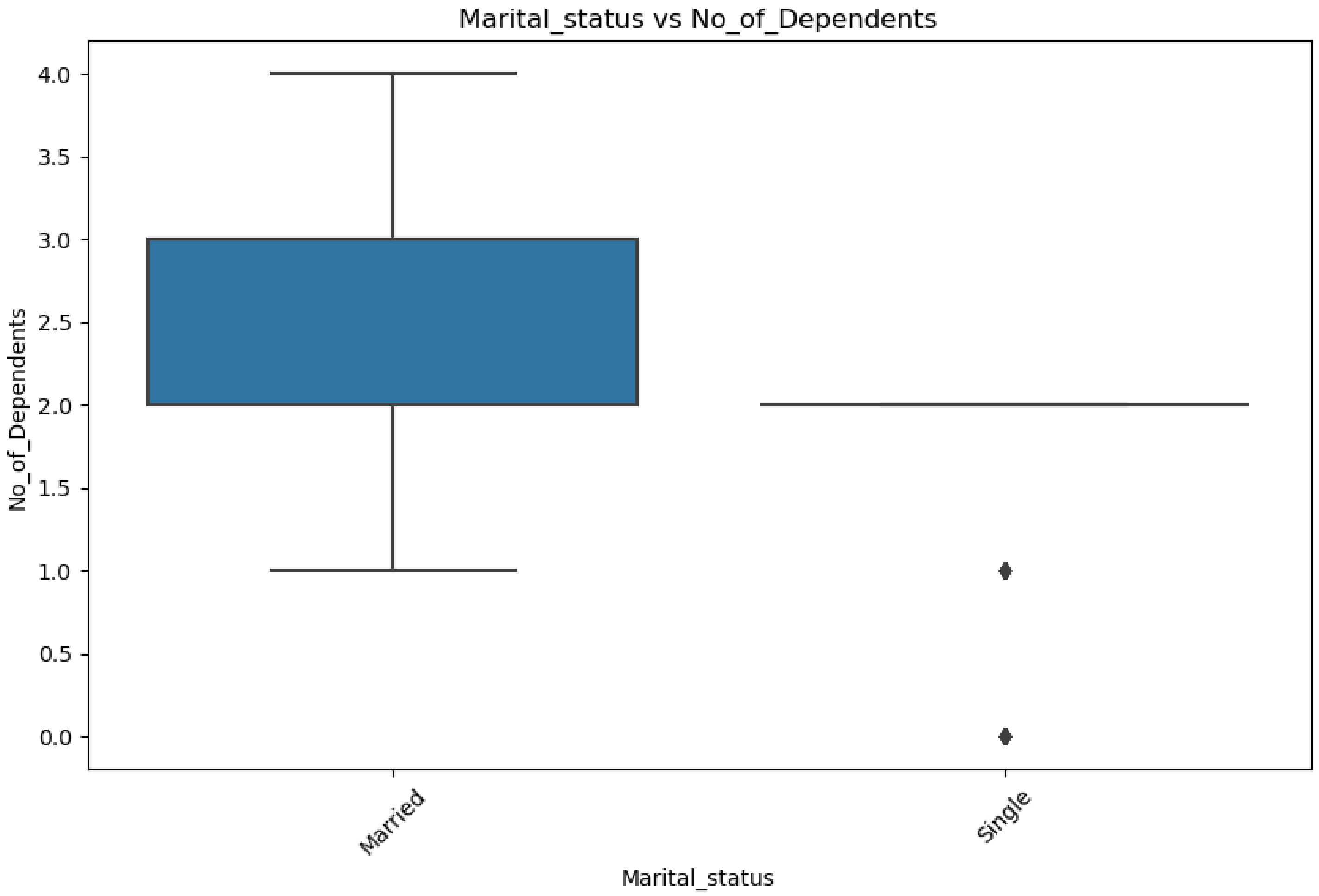

Profession vs Salary
100000
90000
80000
70000
60000
50000
40000
30000
Salary
Business
Salaried
Profession

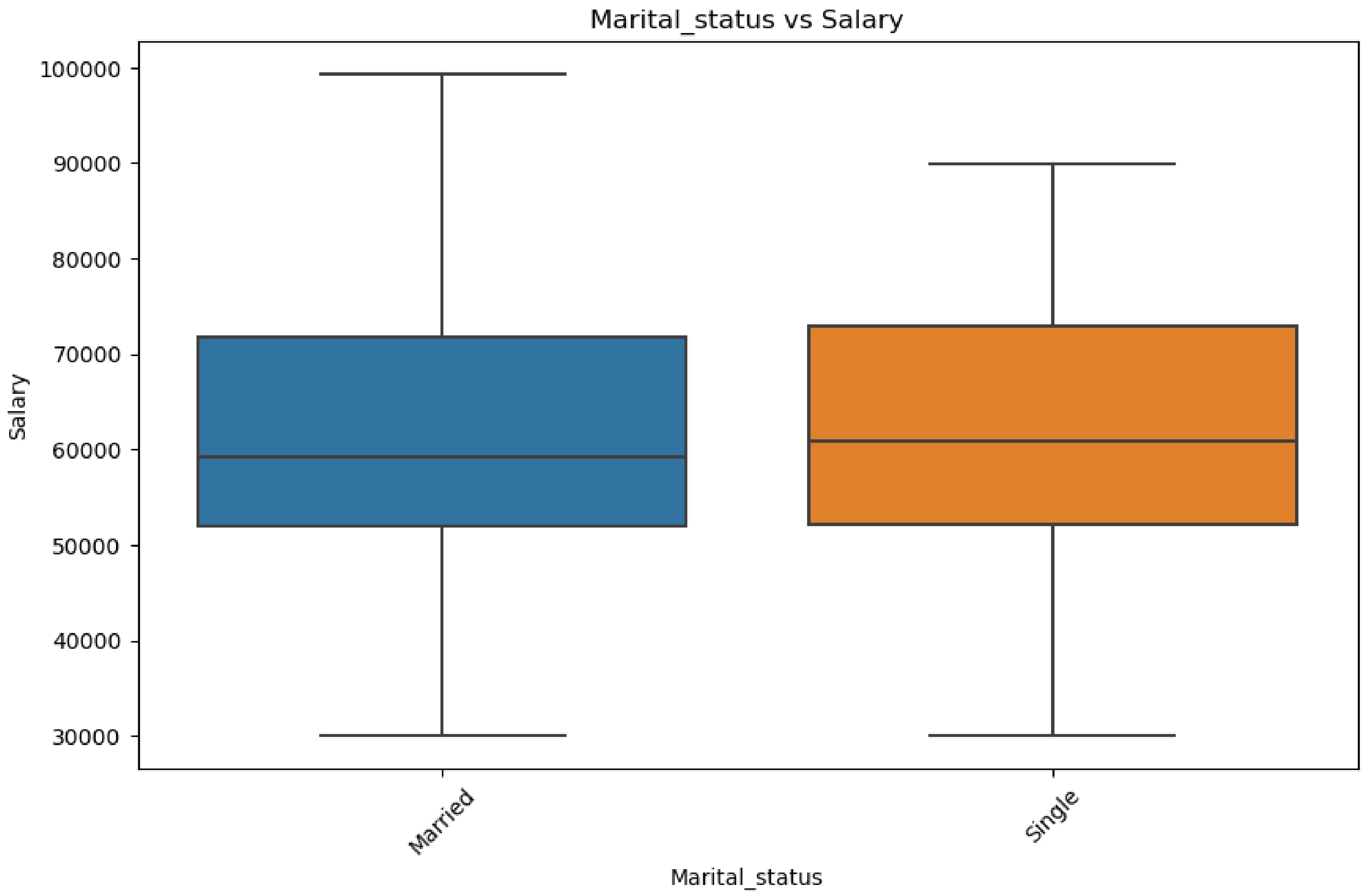

Profession vs Partner_salary
80000
70000
60000
50000
40000
30000
20000
10000
0
Partner_salary
Business
Salaried
Profession

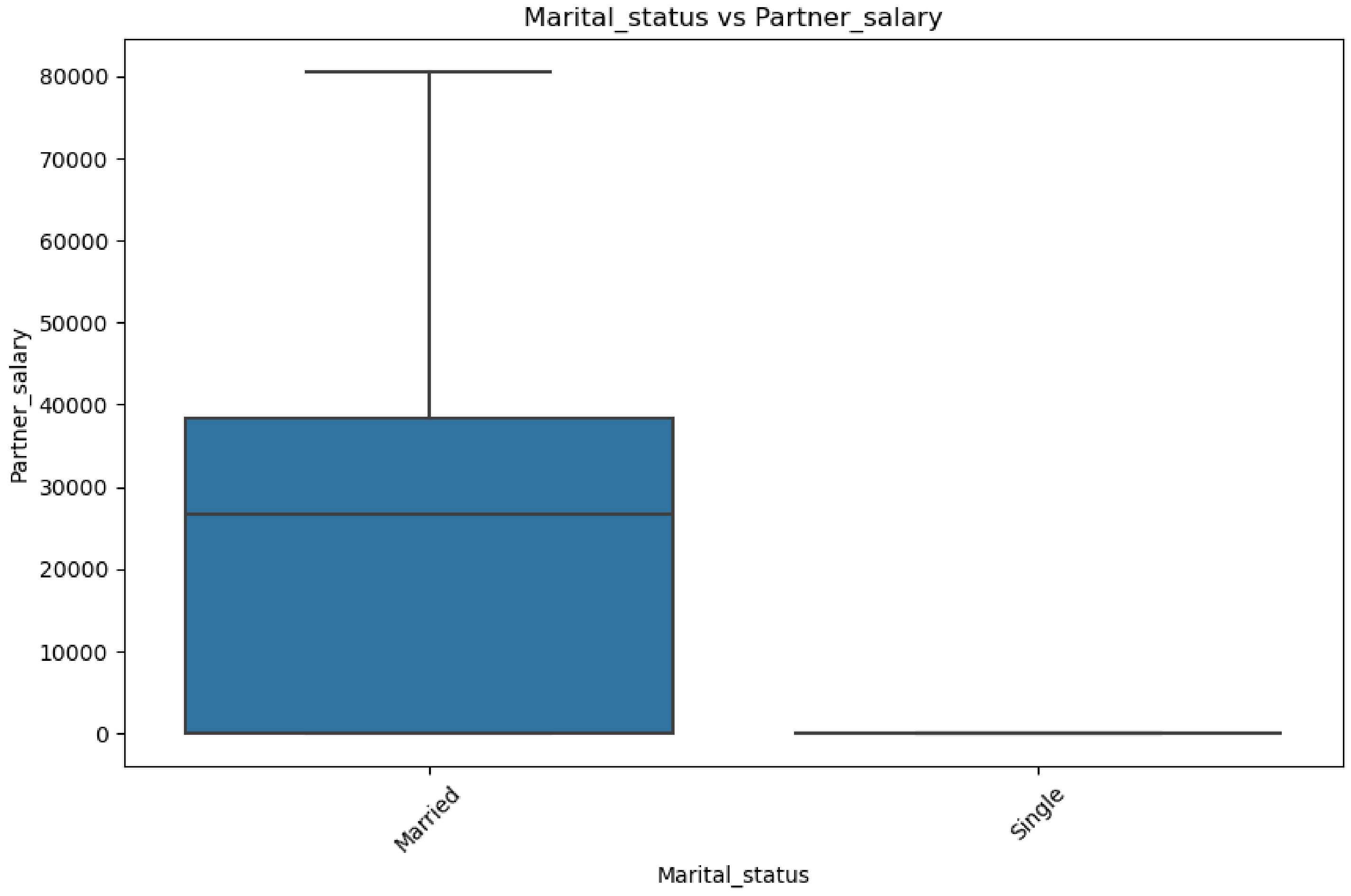

Profession vs Total_salary
160000
140000
120000
100000
80000
60000
40000
Total_salary
Business
Salaried
Profession

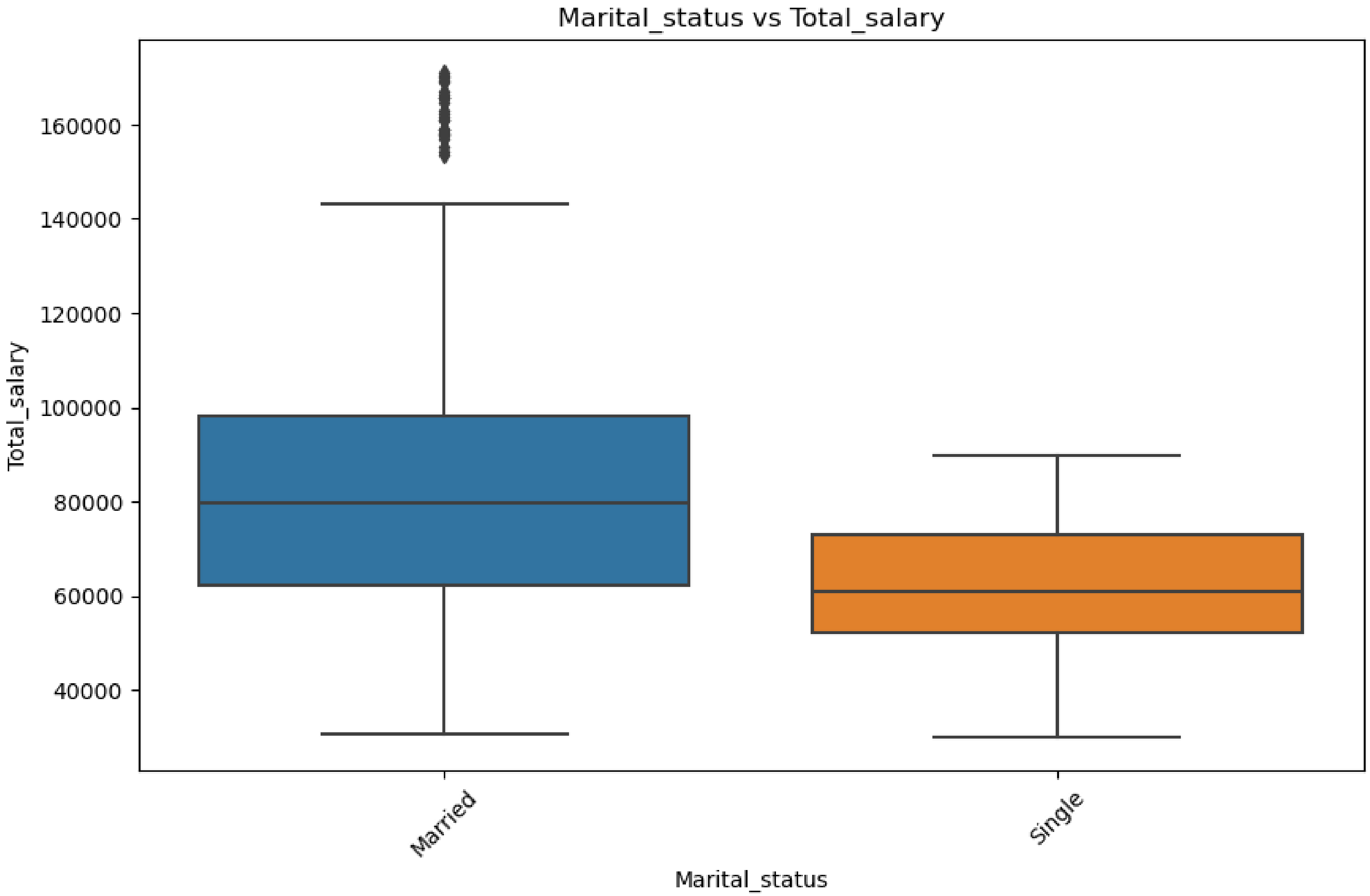

Profession vs Price
70000
60000
50000
40000
30000
20000
Price
Business
Salaried
Profession

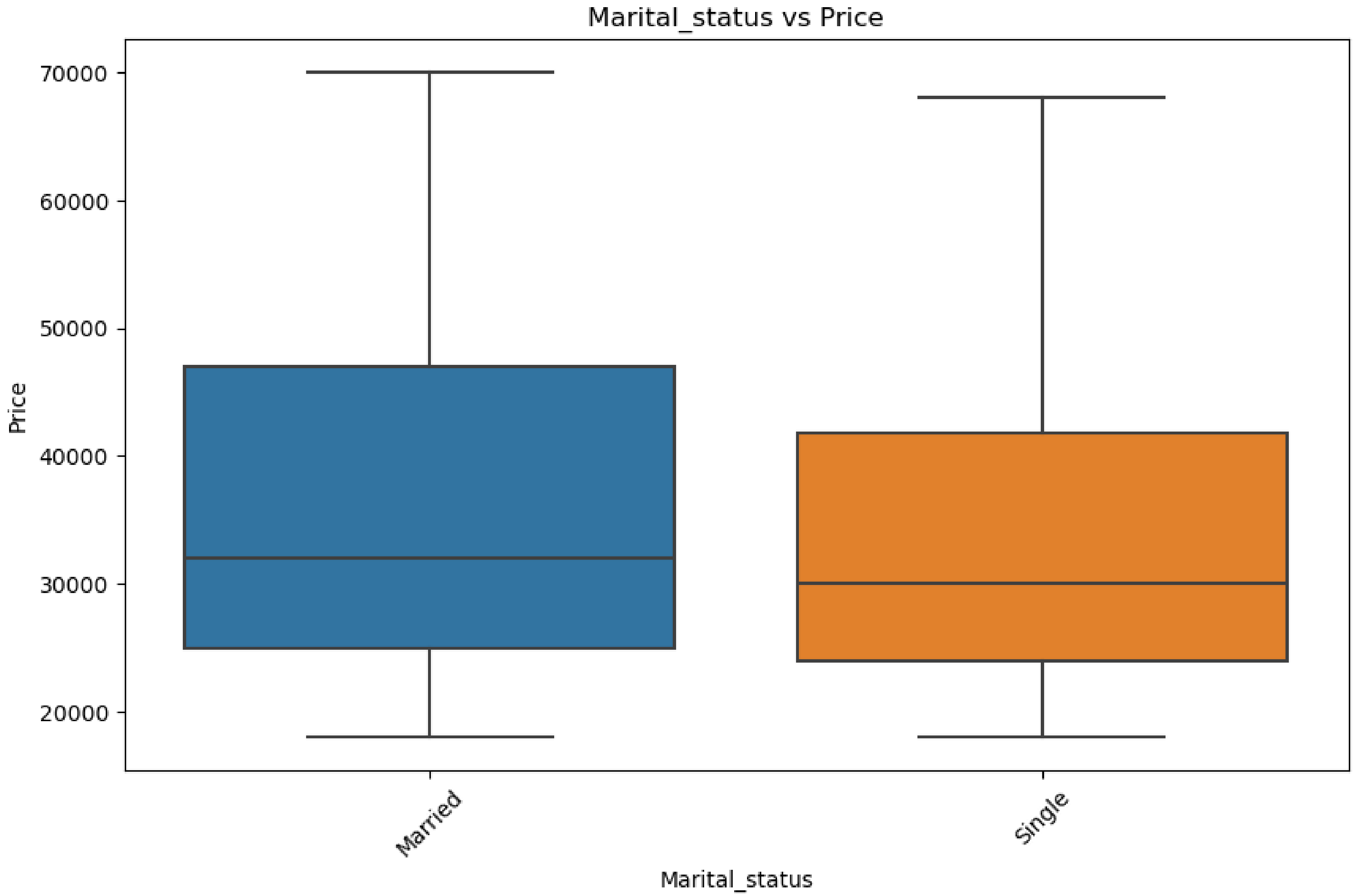

Marital_status vs Age
55
50
45
40
35
30
25
Age
Married
Single
Marital_status

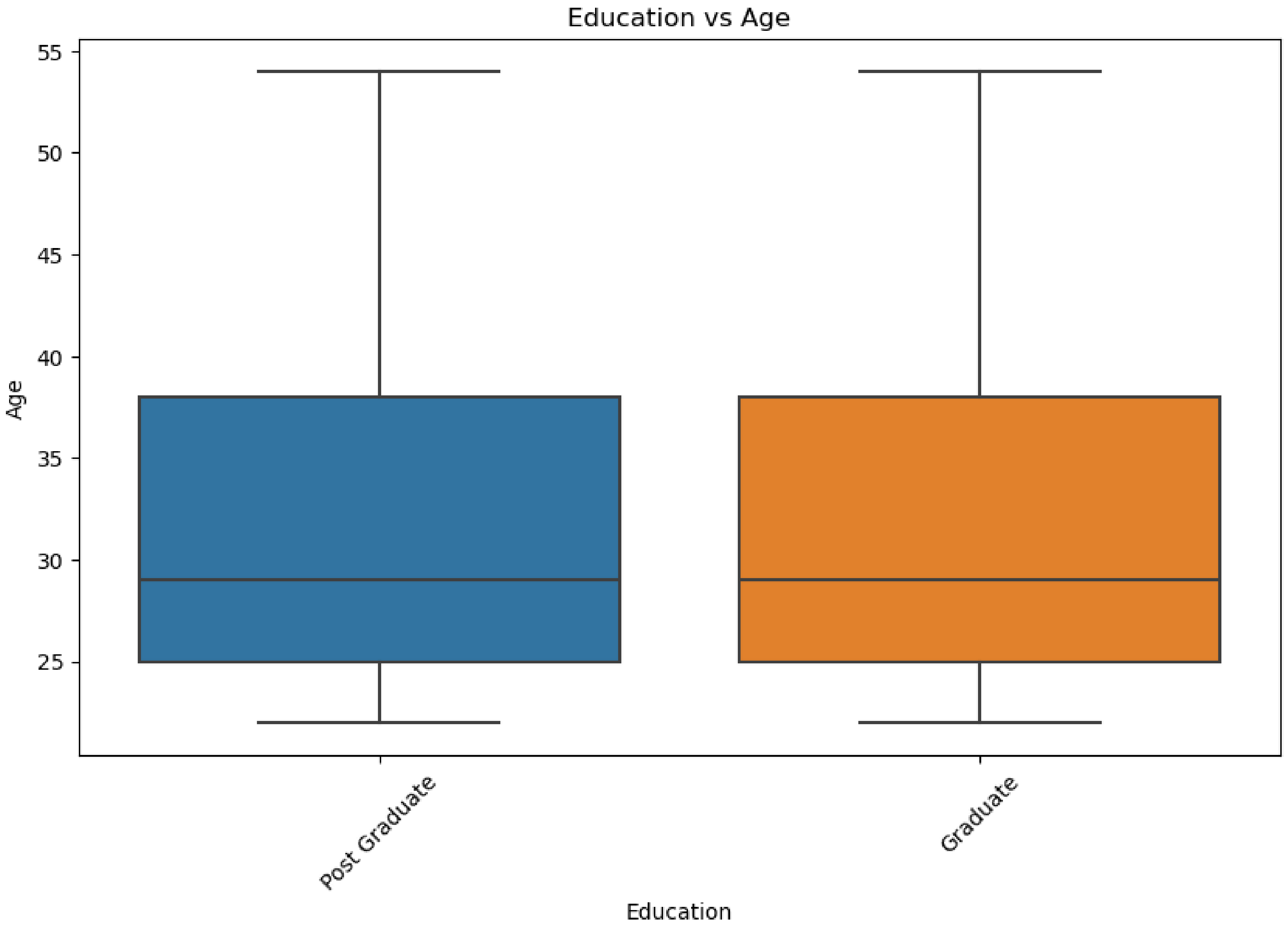

Marital_status vs No_of_Dependents
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
No_of_Dependents
Married
Single
Marital_status

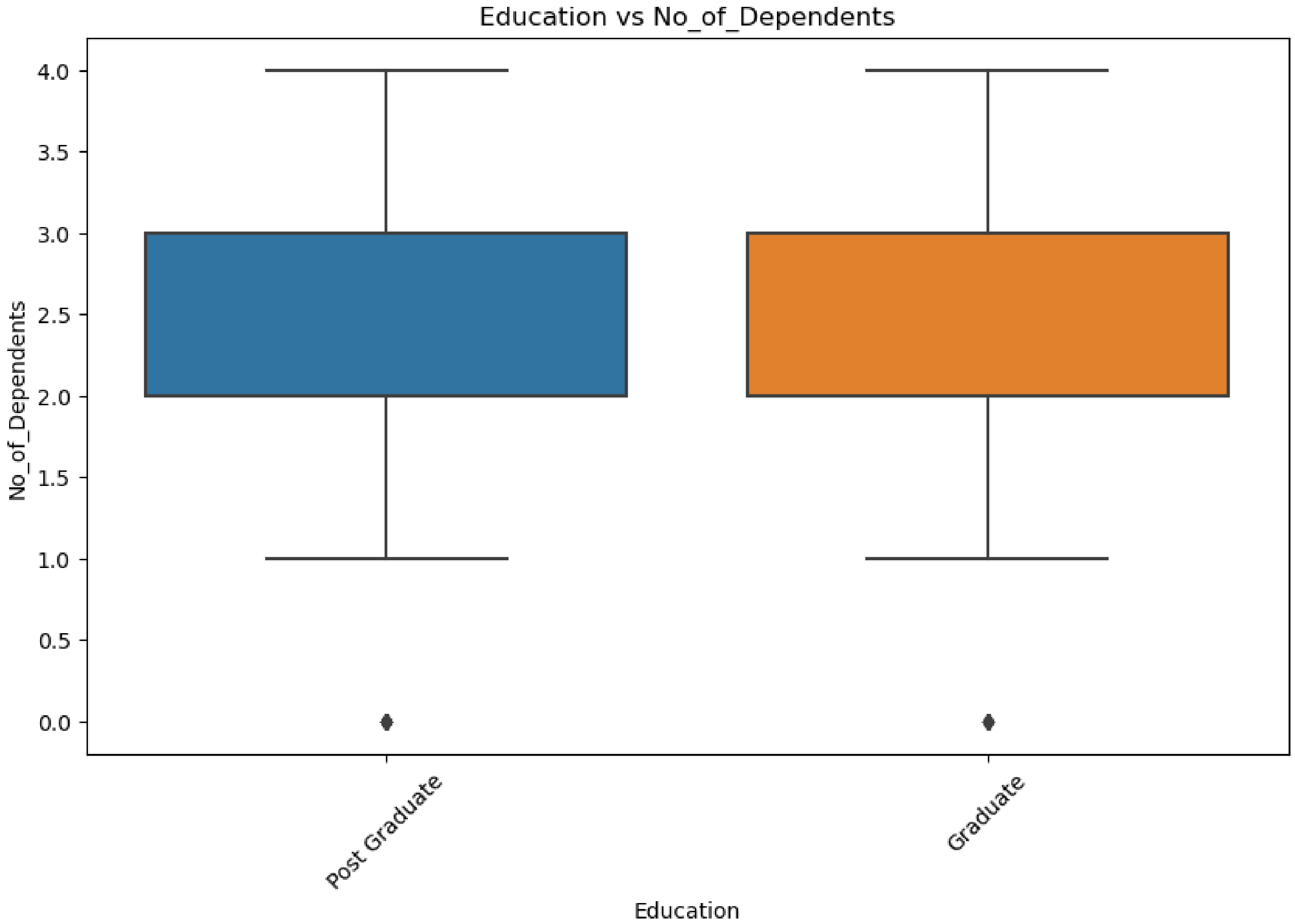

Marital_status vs Salary
Salary
100000
90000
80000
70000
60000
50000
40000
30000
Married
Single
Marital_status

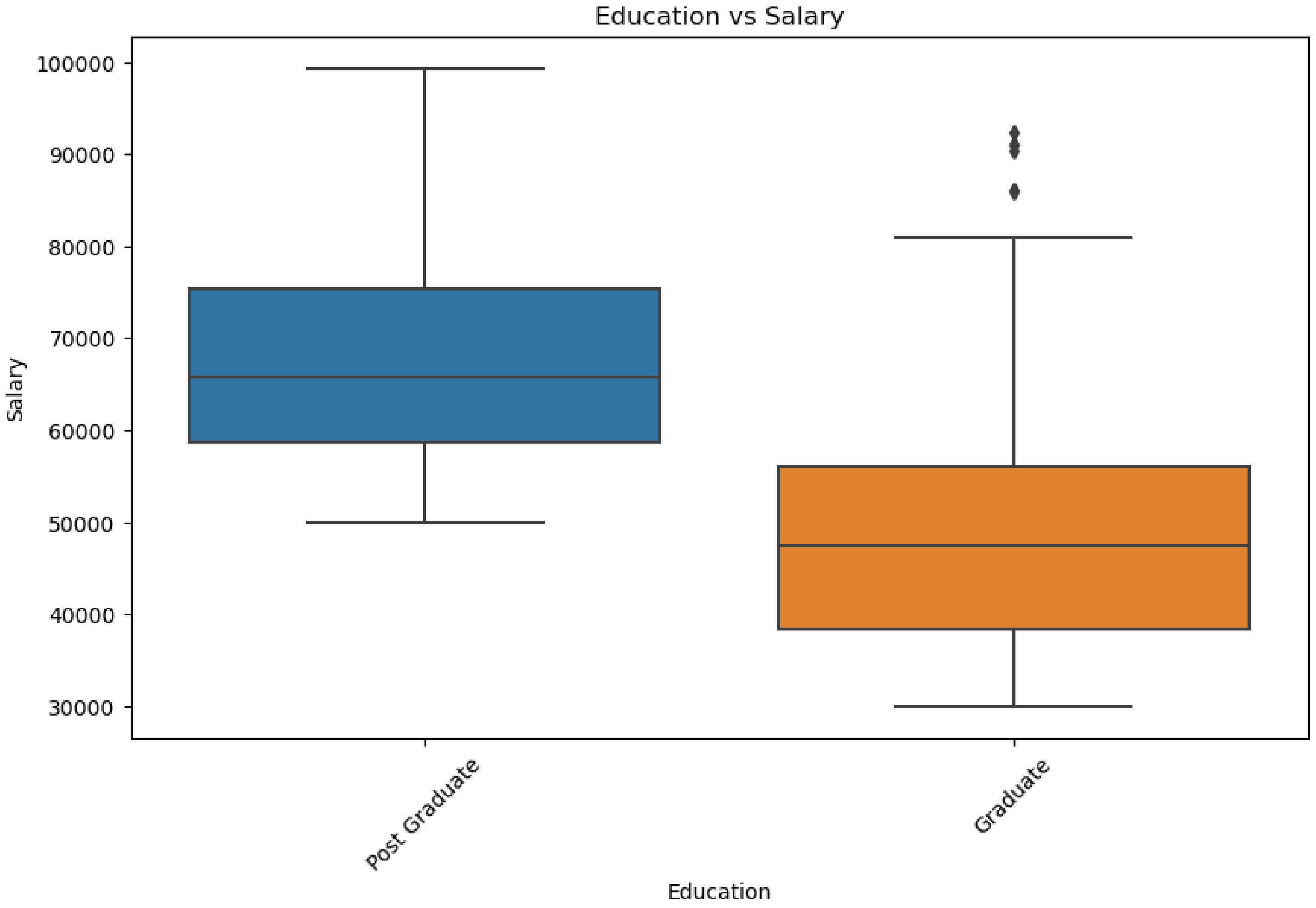

Marital_status vs Partner_salary
Partner_salary
80000
70000
60000
50000
40000
30000
20000
10000
0
Married
Single
Marital_status

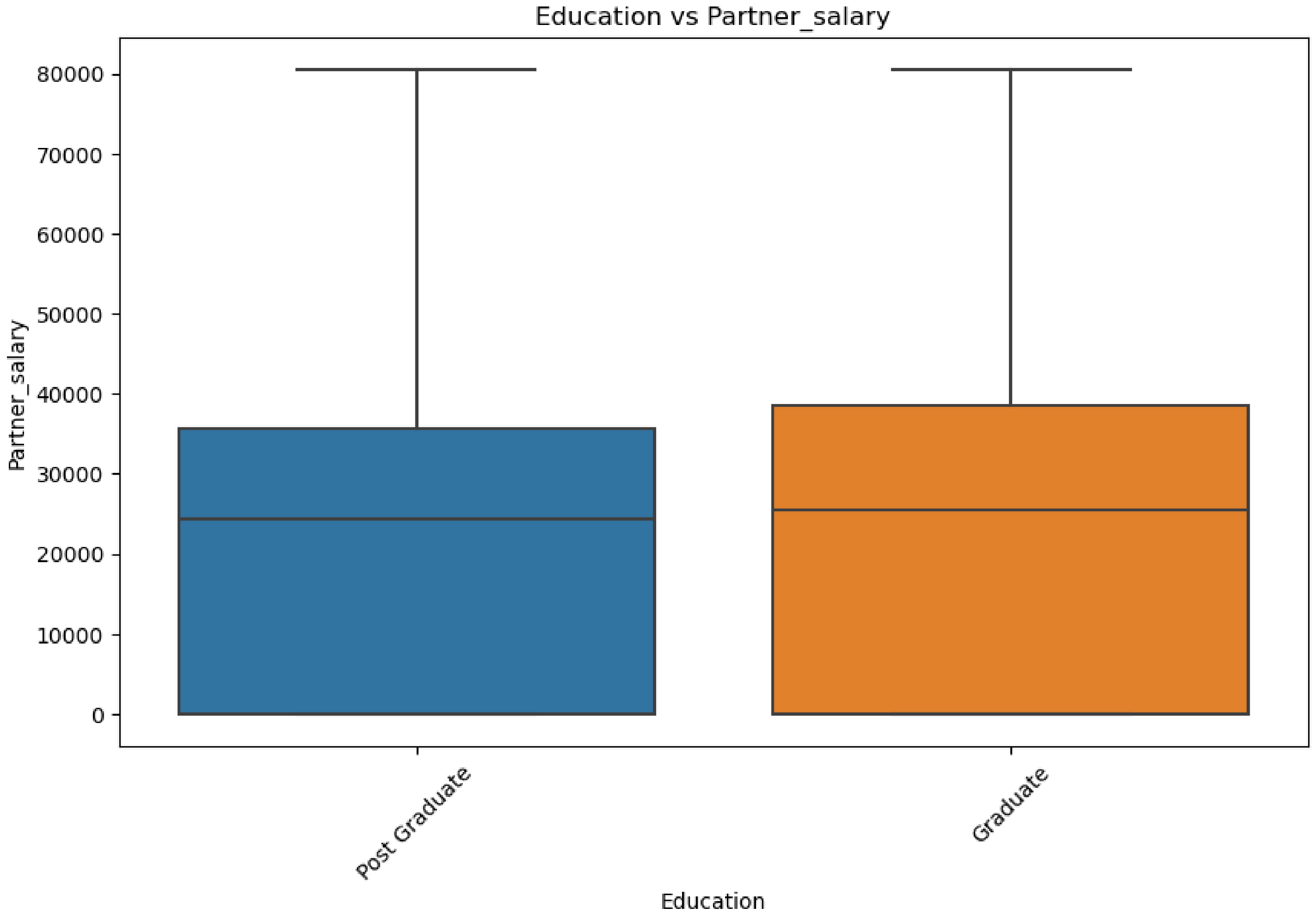

Marital_status vs Total_salary
Total_salary
160000
140000
120000
100000
80000
60000
40000
Married
Single
Marital_status

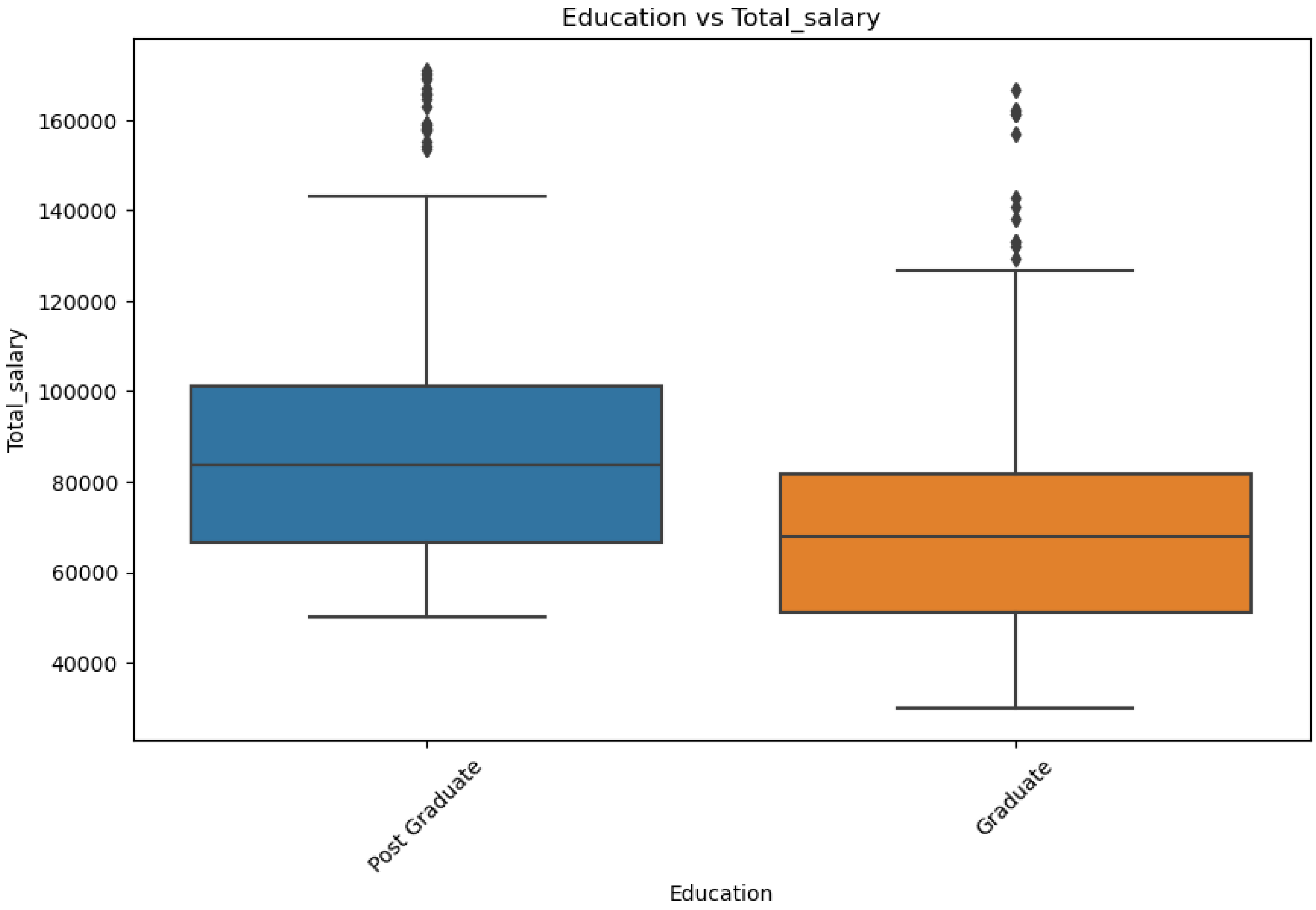

Marital_status vs Price
Price
70000
60000
50000
40000
30000
20000
Married
Single
Marital_status


Education vs Age
55
50
45
40
35
30
25
Age
Post Graduate
Graduate
Education

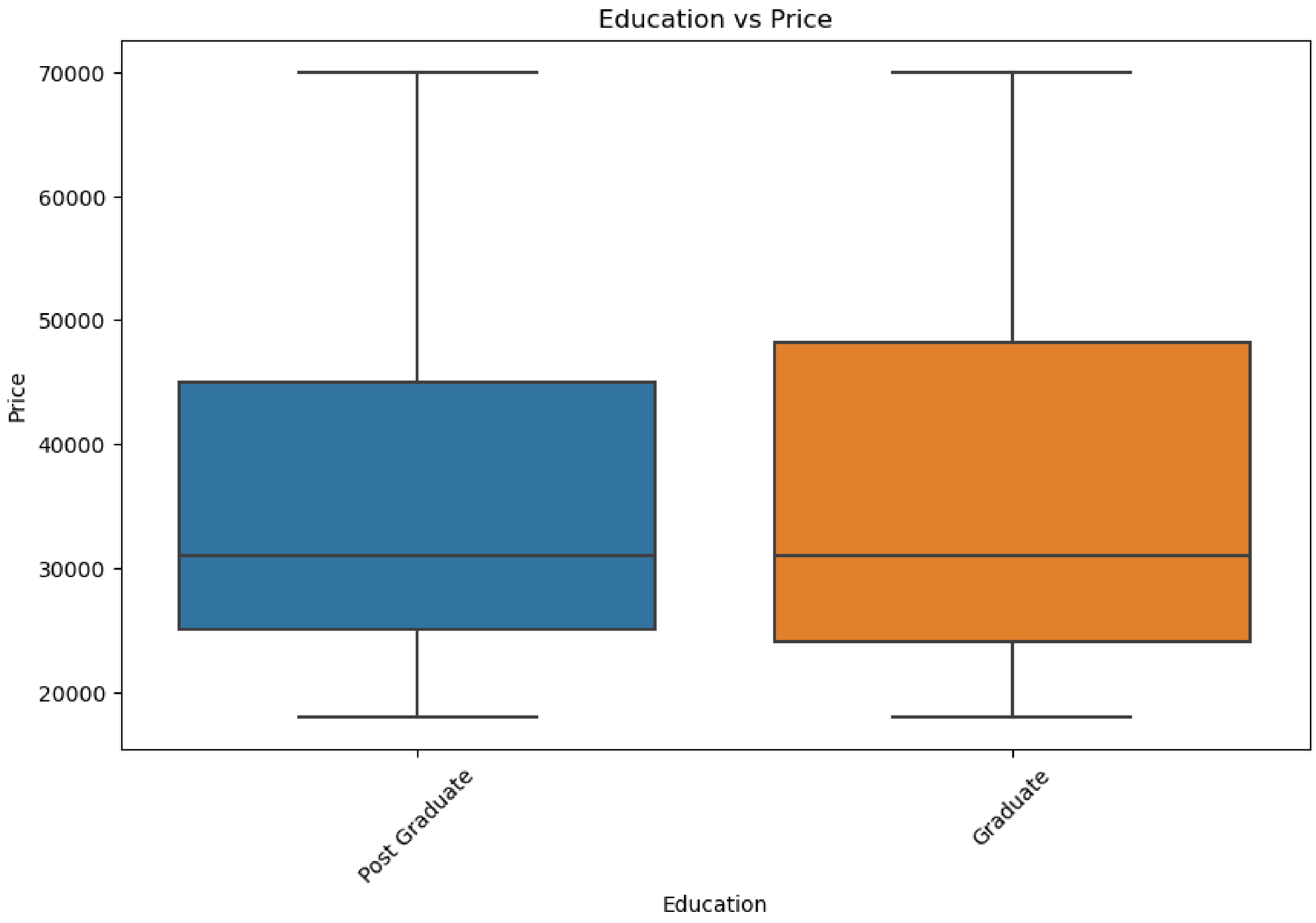

Education vs No_of_Dependents
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
No_of_Dependents
Post Graduate
Graduate
Education


Education vs Salary
Salary
100000
90000
80000
70000
60000
50000
40000
30000
Post Graduate
Graduate
Education

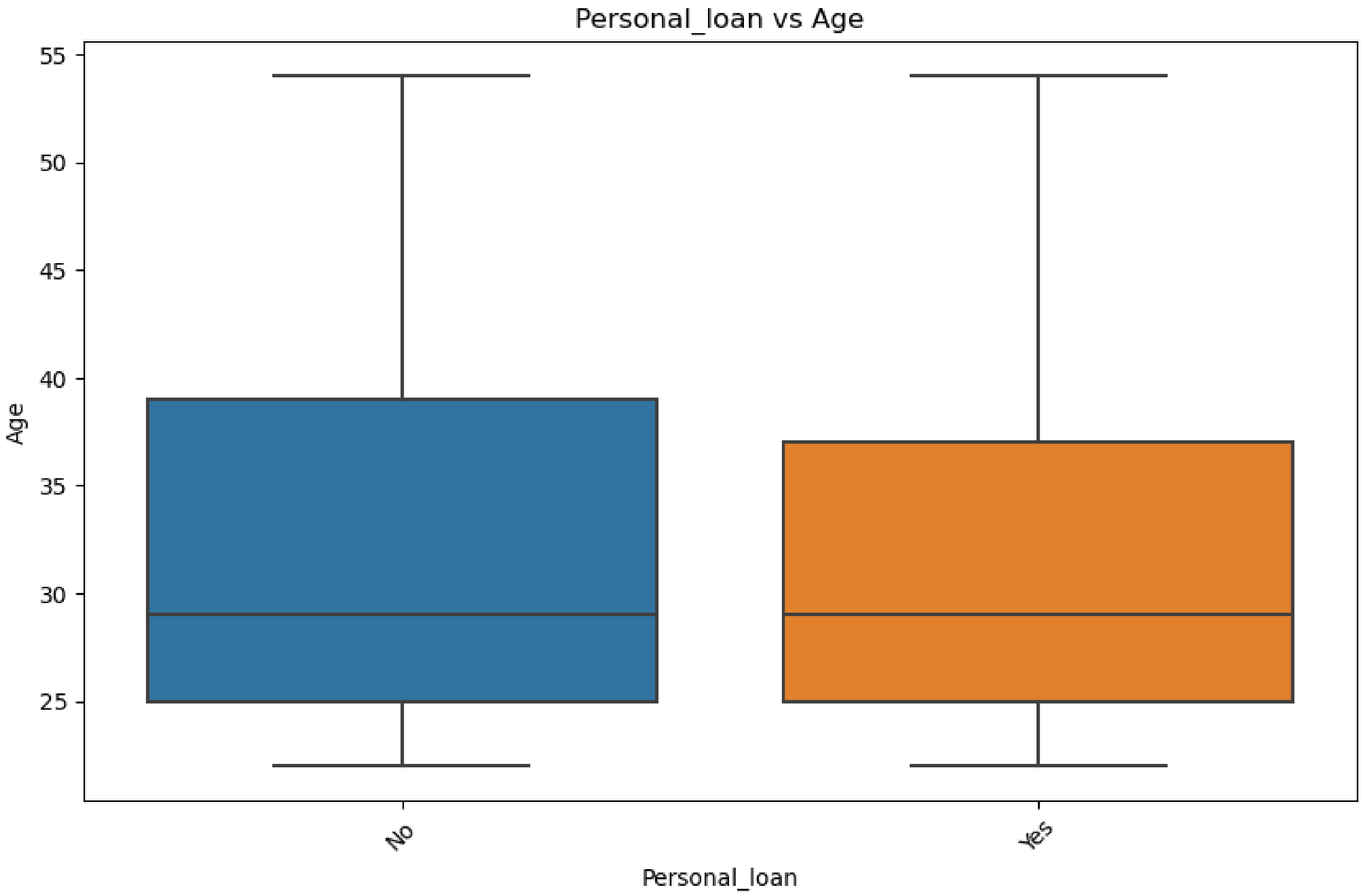

Education vs Partner_salary
Partner_salary
80000
70000
60000
50000
40000
30000
20000
10000
0
Post Graduate
Graduate
Education

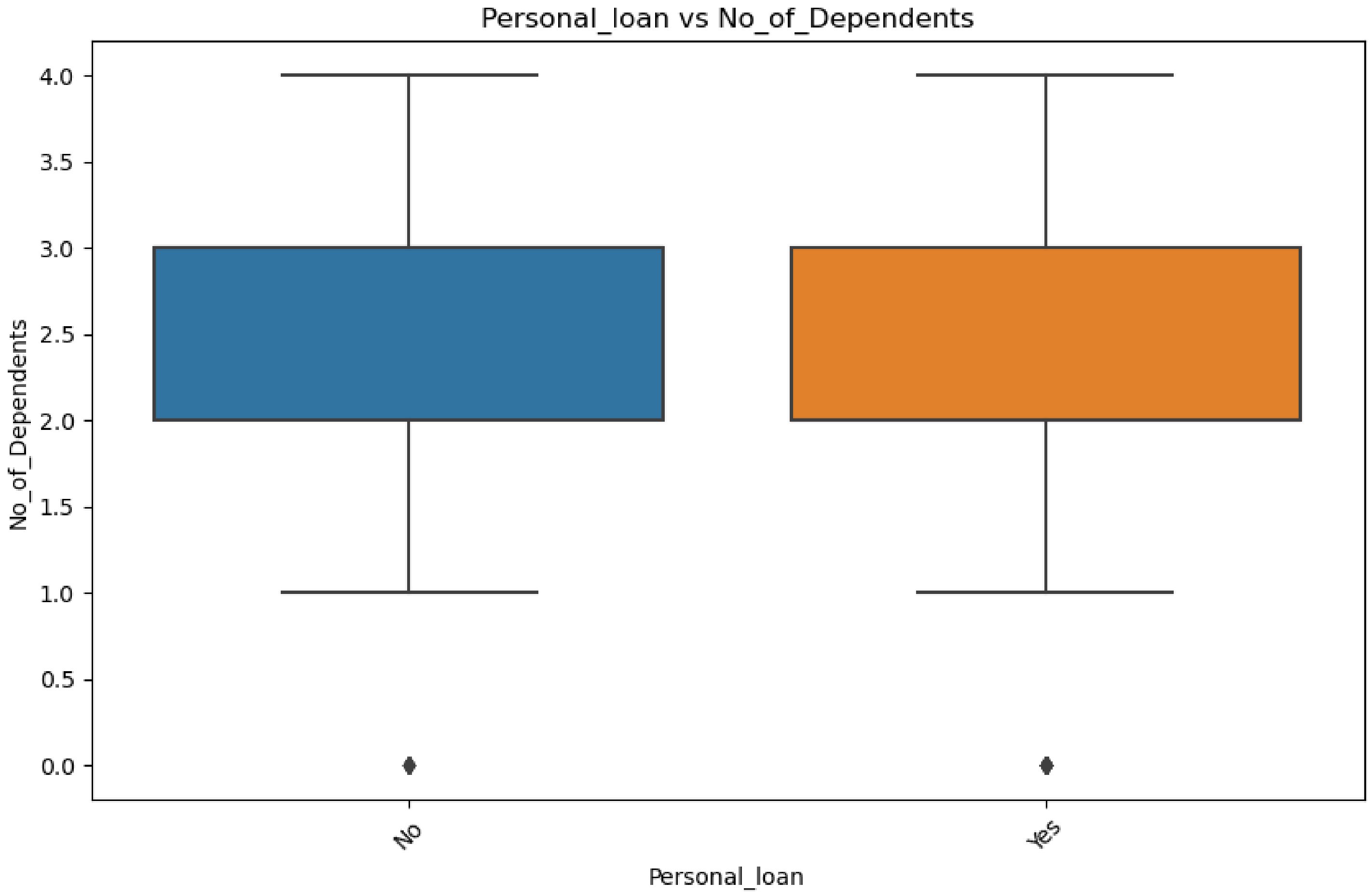


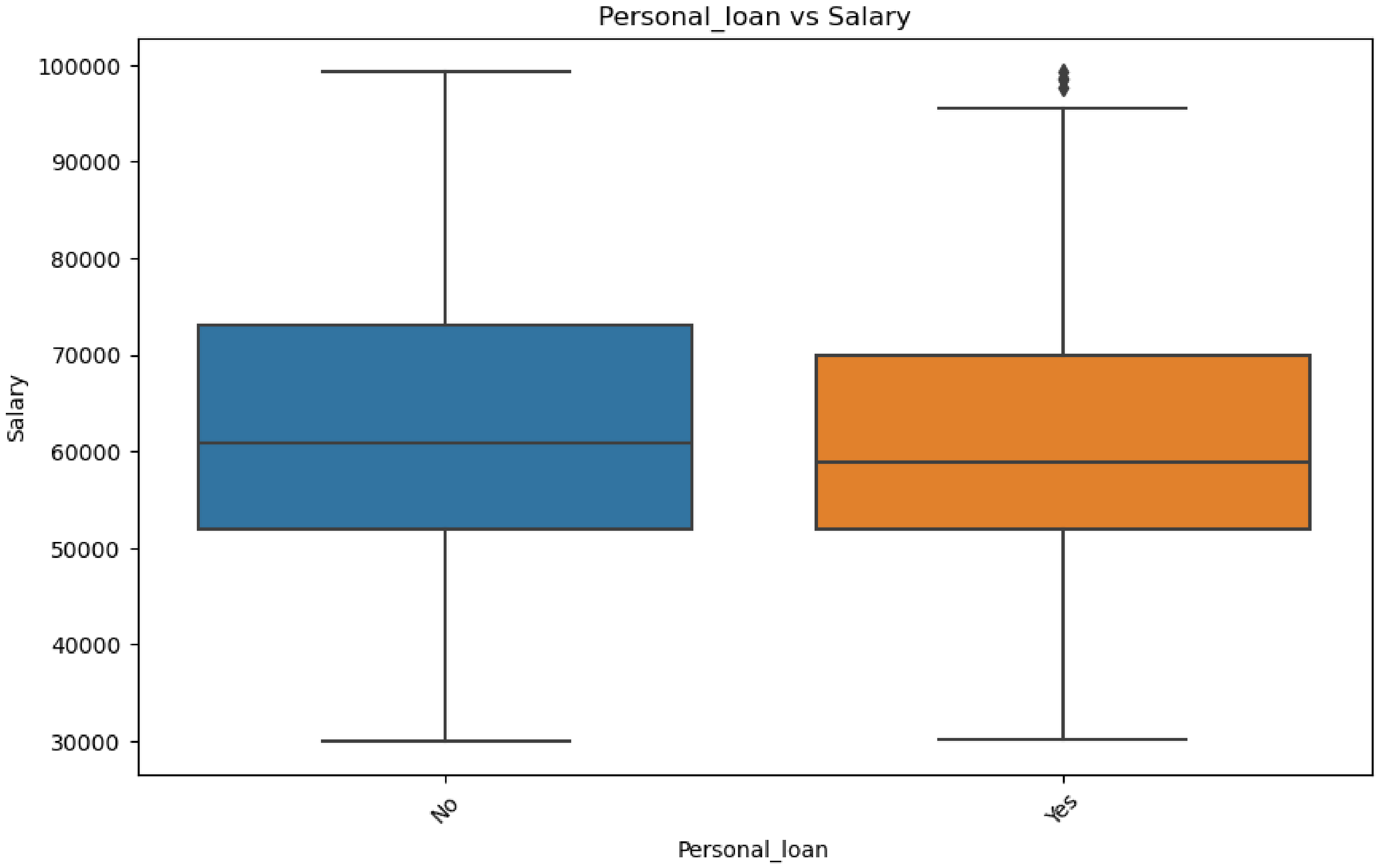
Education vs Total_salary

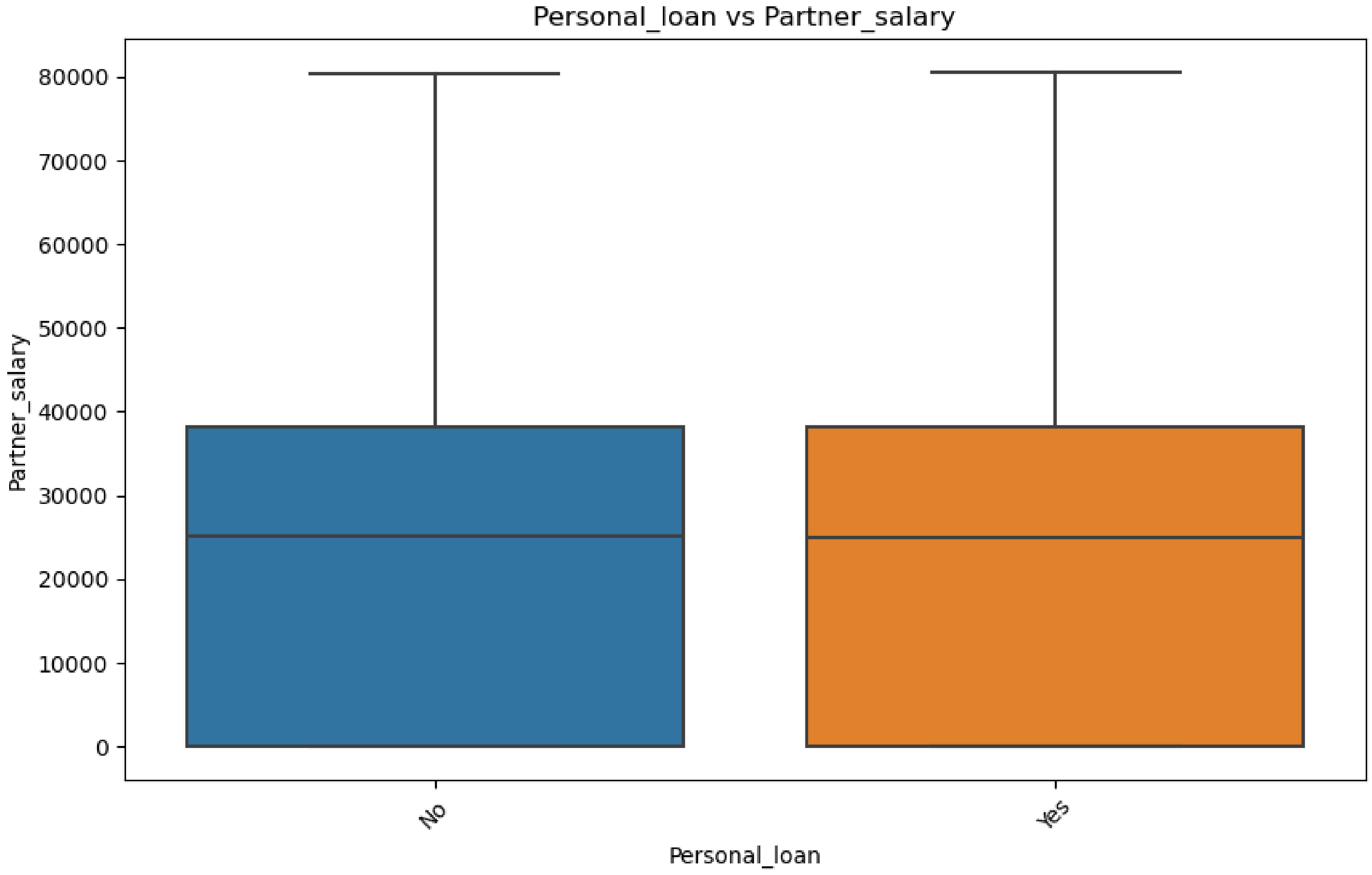


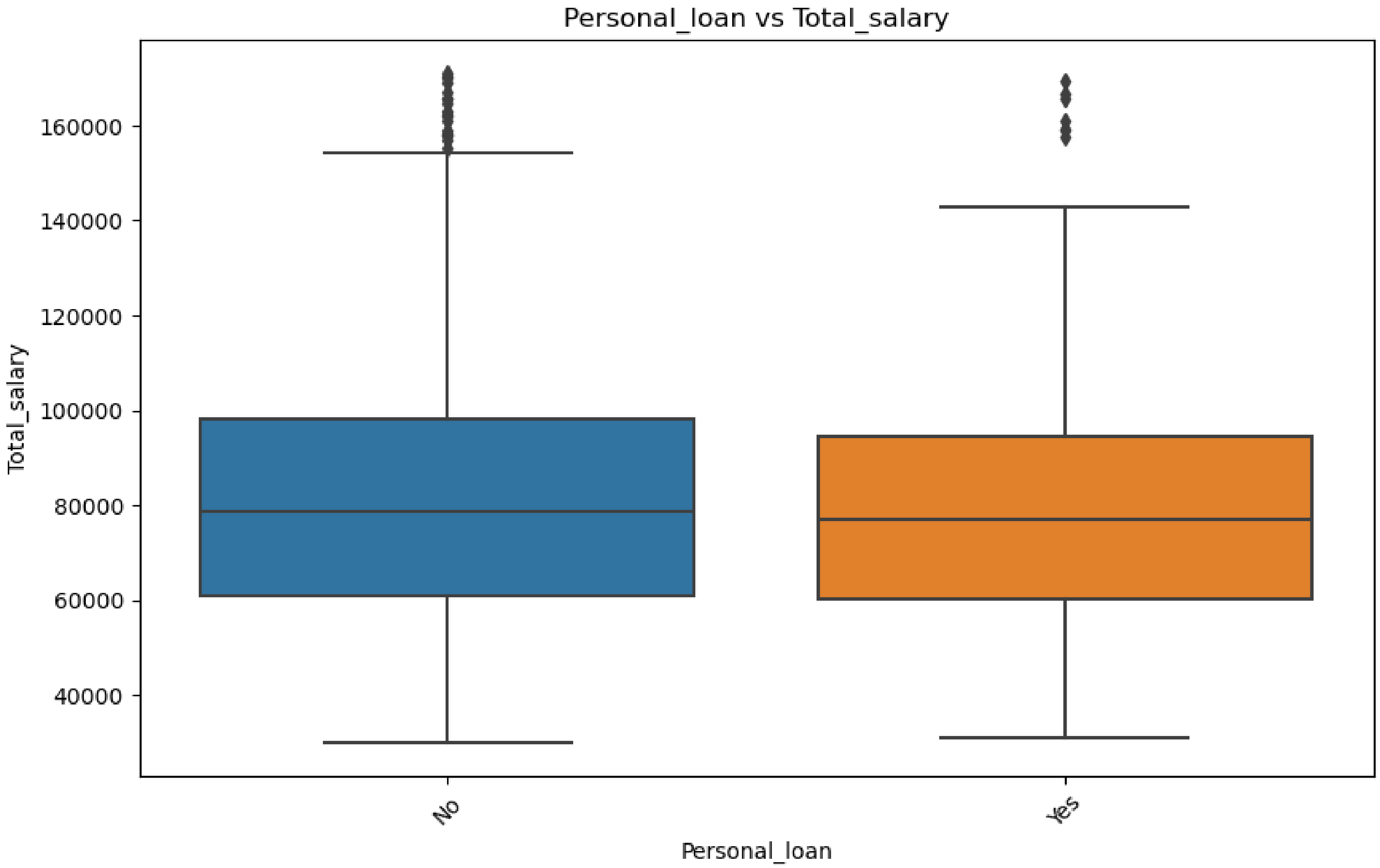
Education vs Price

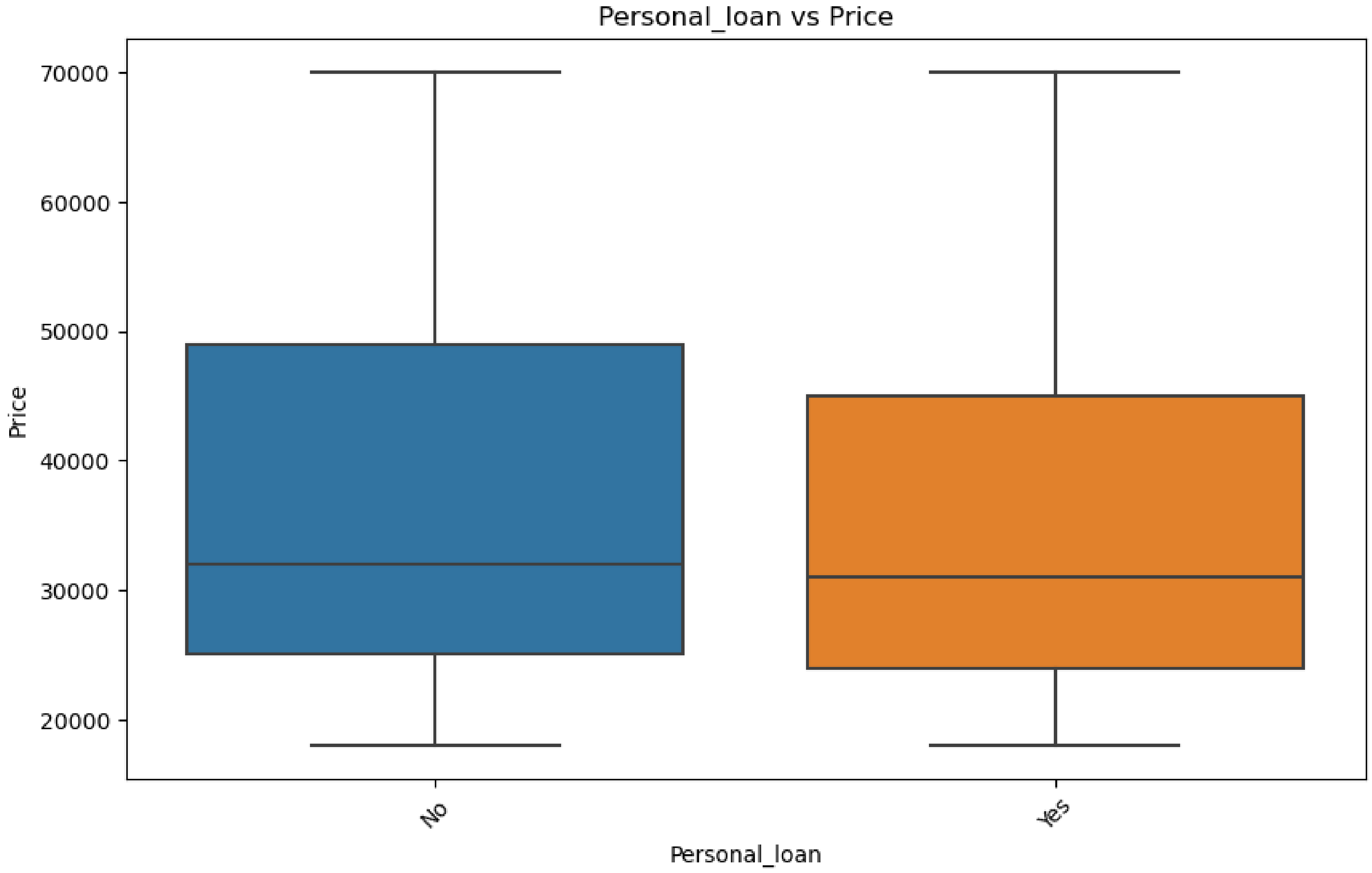

Personal_loan vs Age
55
50
45
40
35
30
25
Age
No
Yes
Personal_loan

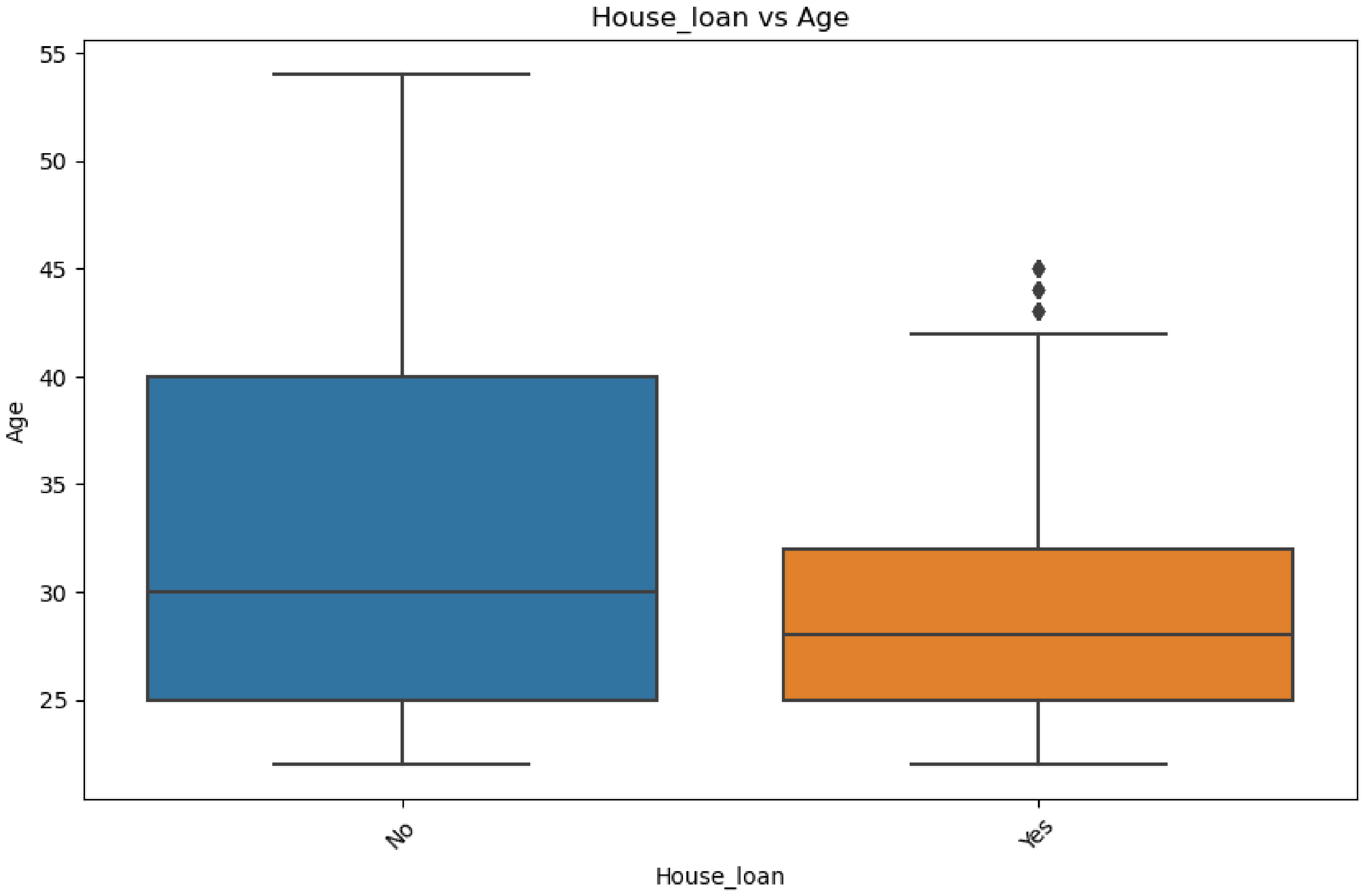

Personal_loan vs No_of_Dependents
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
No_of_Dependents
No
Yes
Personal_loan

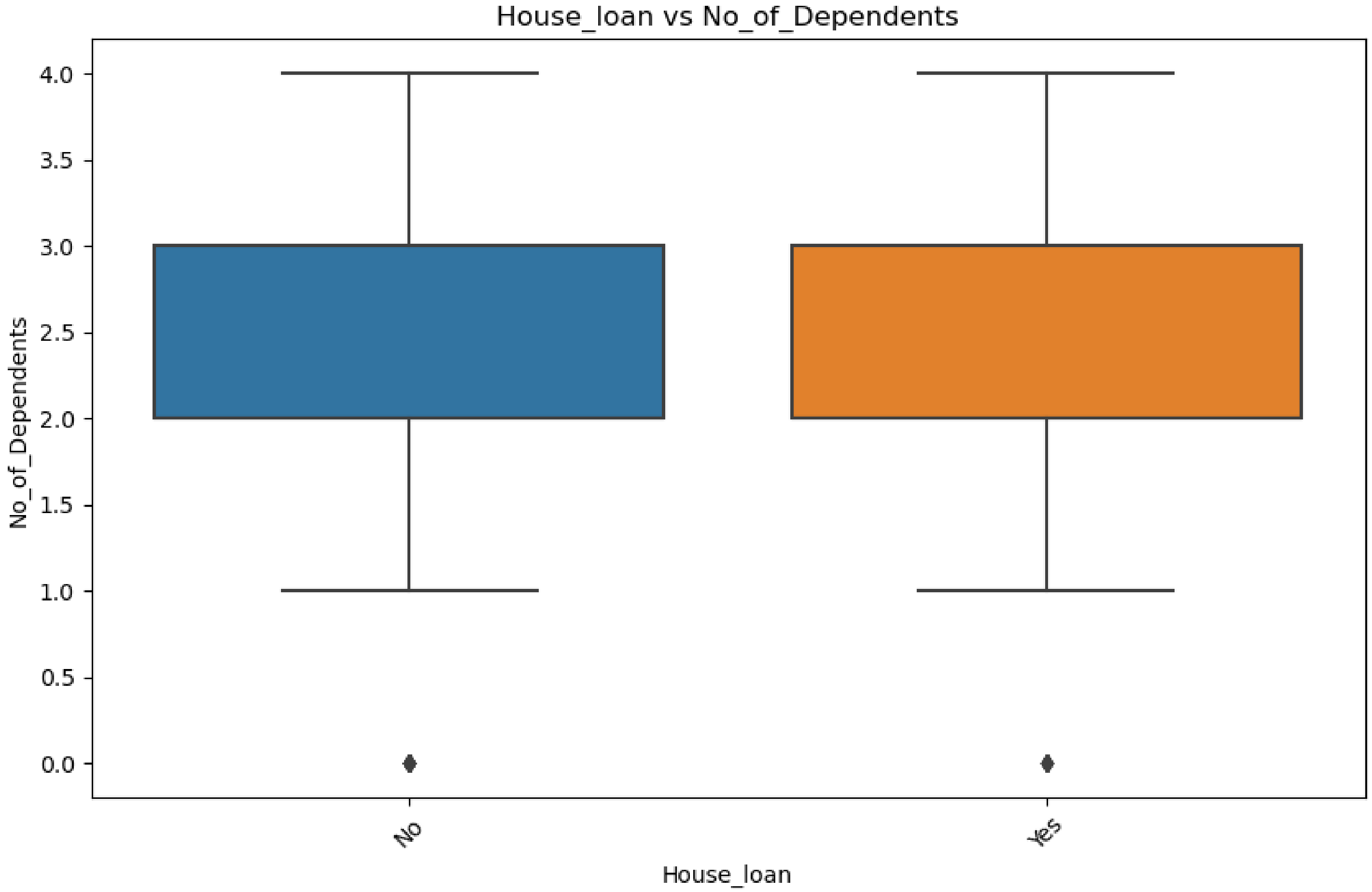

Personal_loan vs Salary
100000
90000
80000
70000
60000
50000
40000
30000
Salary
No
Yes
Personal_loan

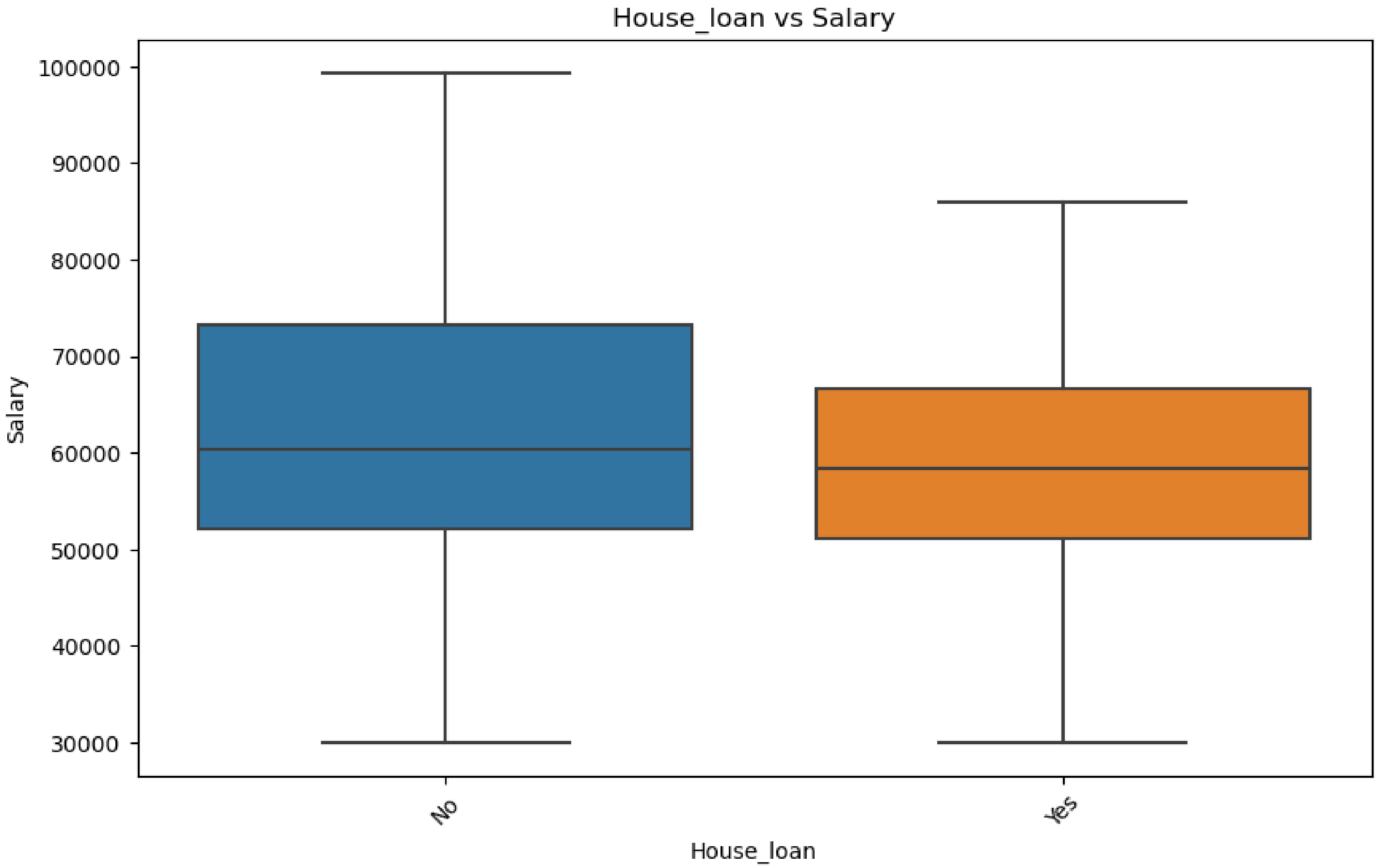

Personal_loan vs Partner_salary
80000
70000
60000
50000
40000
30000
20000
10000
0
Partner_salary
No
Yes
Personal_loan

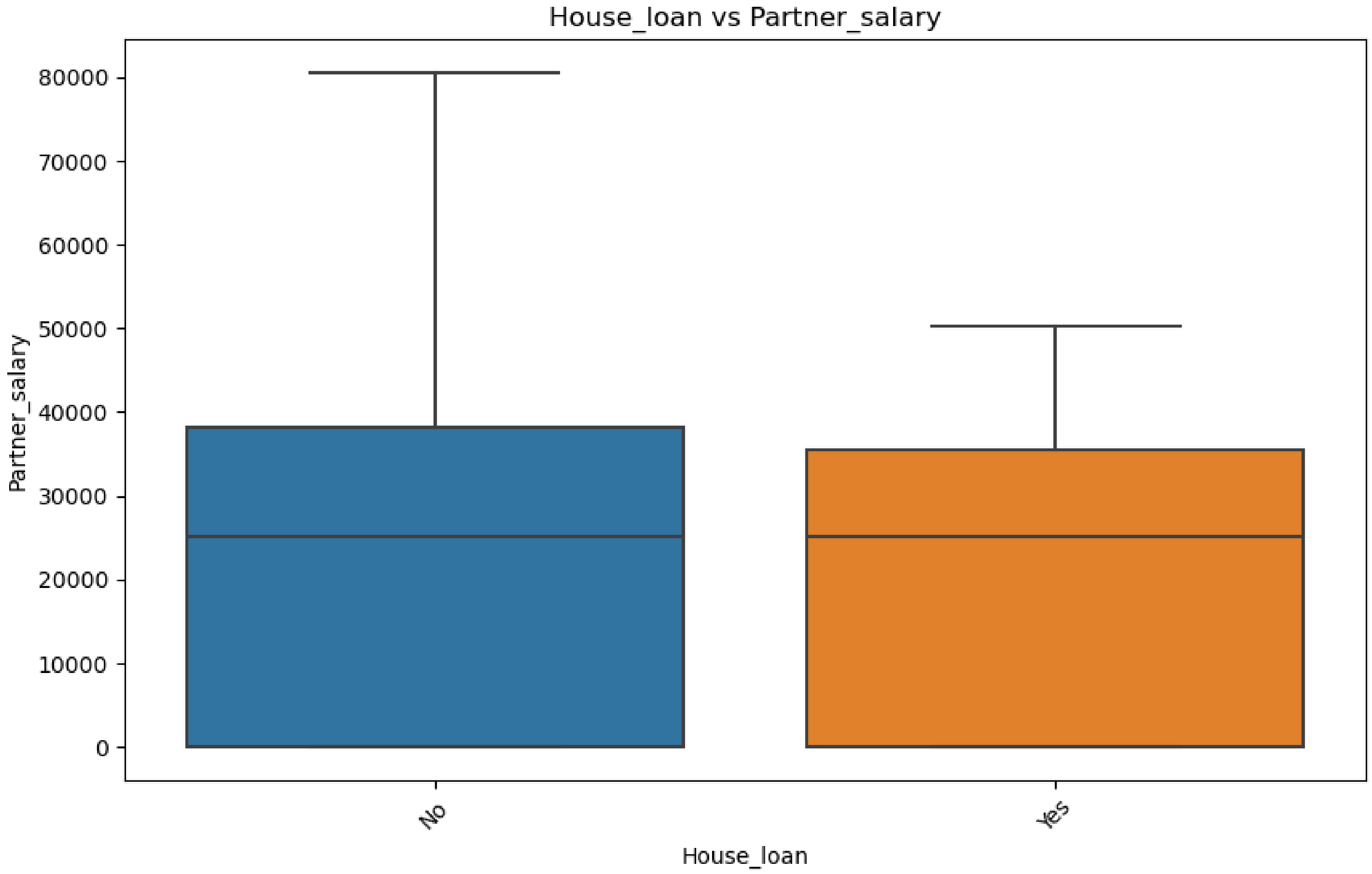

Personal_loan vs Total_salary
160000
140000
120000
100000
80000
60000
40000
Total_salary
No
Yes
Personal_loan

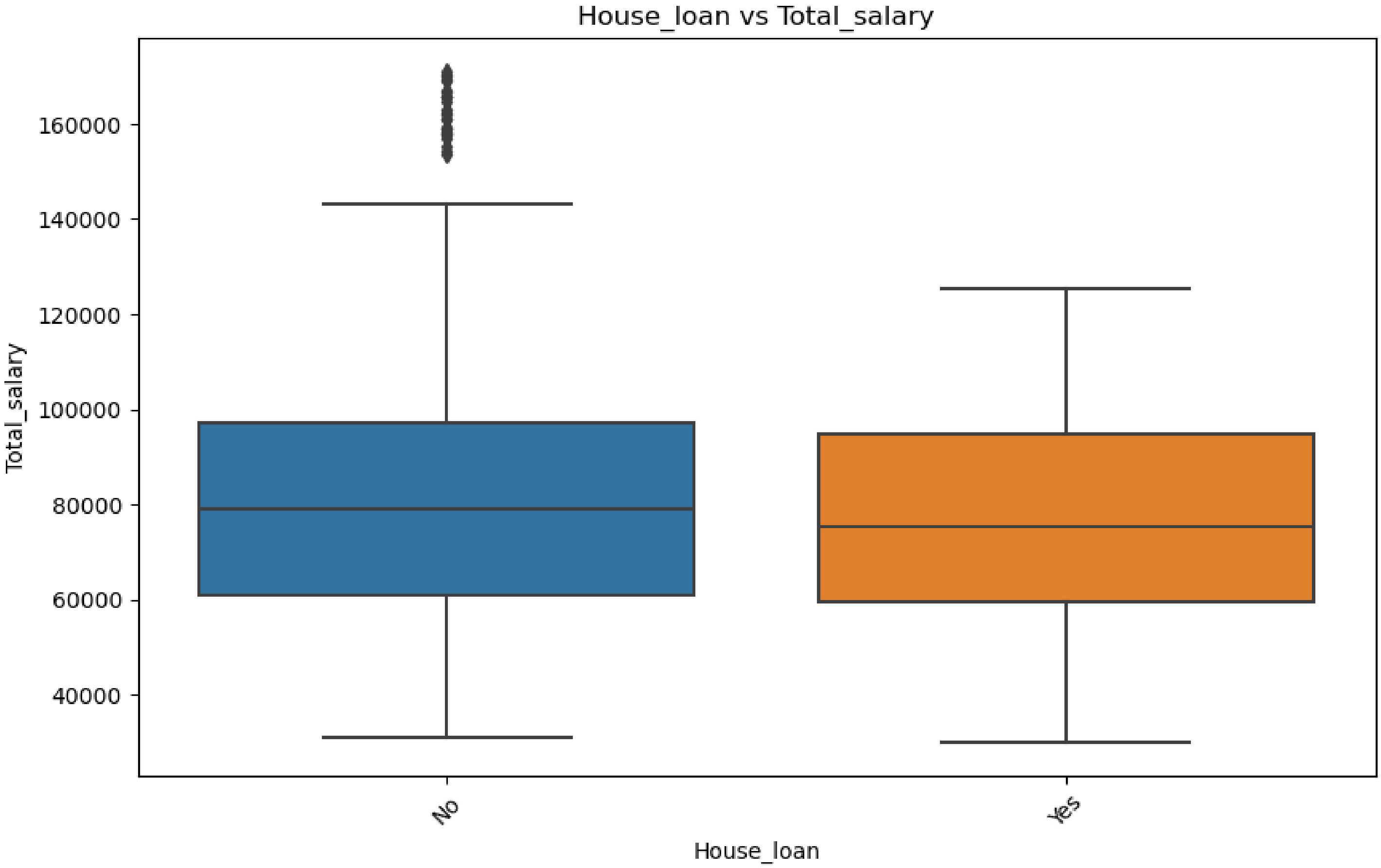

Personal_loan vs Price
70000
60000
50000
40000
30000
20000
Price
No
Yes
Personal_loan

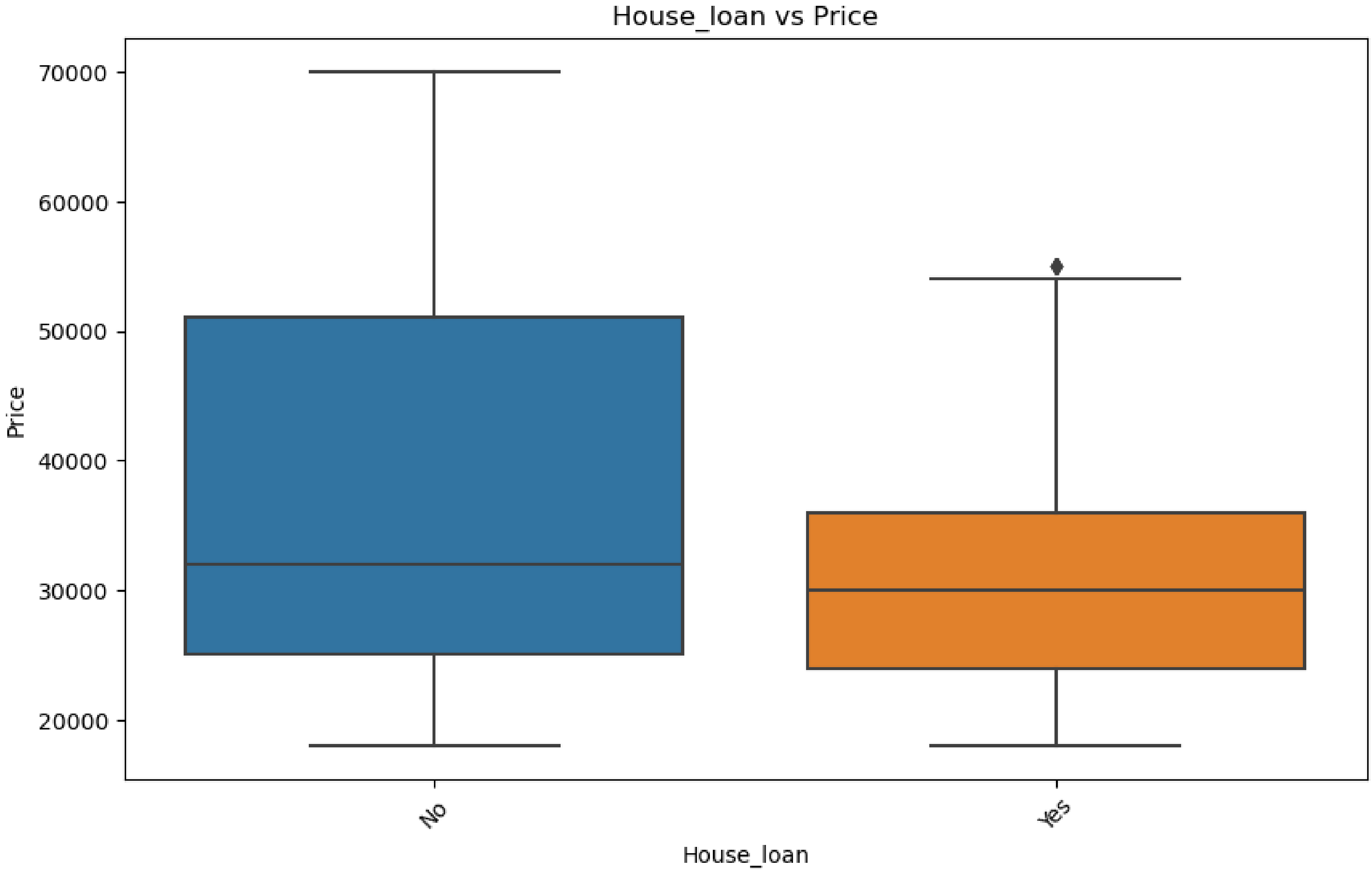

House_loan vs Age
55
50
45
40
35
30
25
Age
No
Yes
House_loan

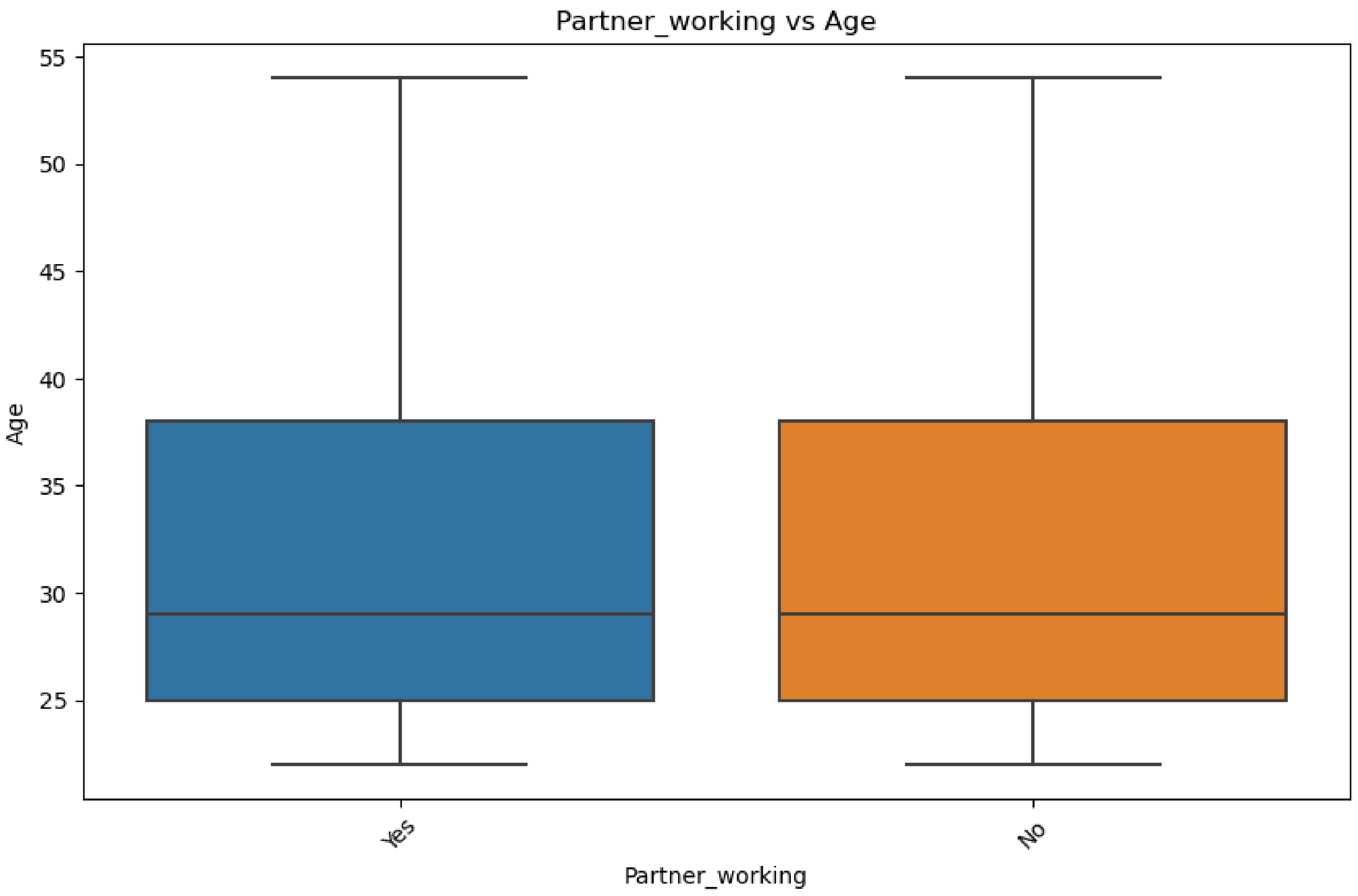

House_loan vs No_of_Dependents
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
No_of_Dependents
No
Yes
House_loan

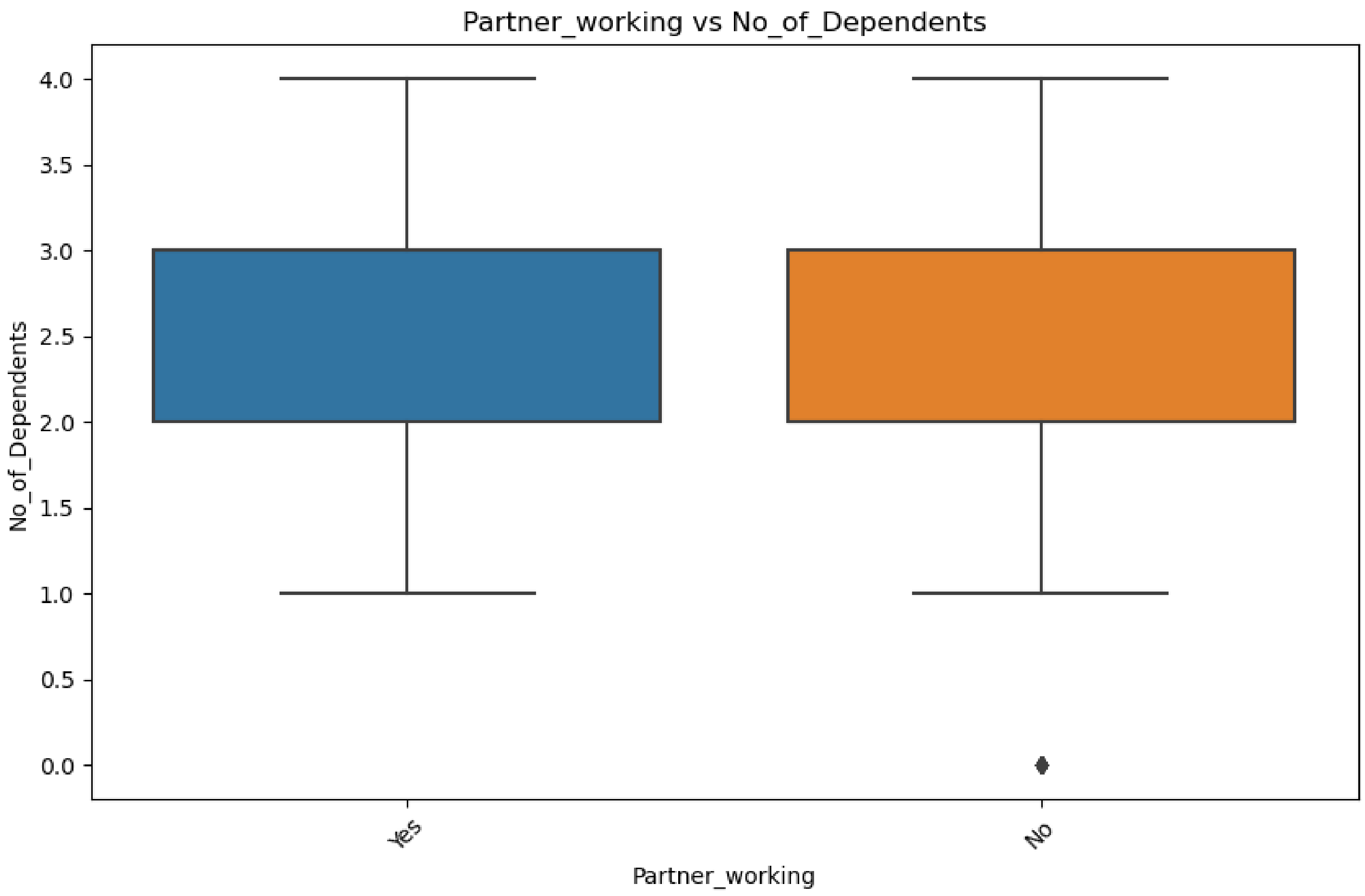

House_loan vs Salary
100000
90000
80000
70000
60000
50000
40000
30000
Salary
No
Yes
House_loan

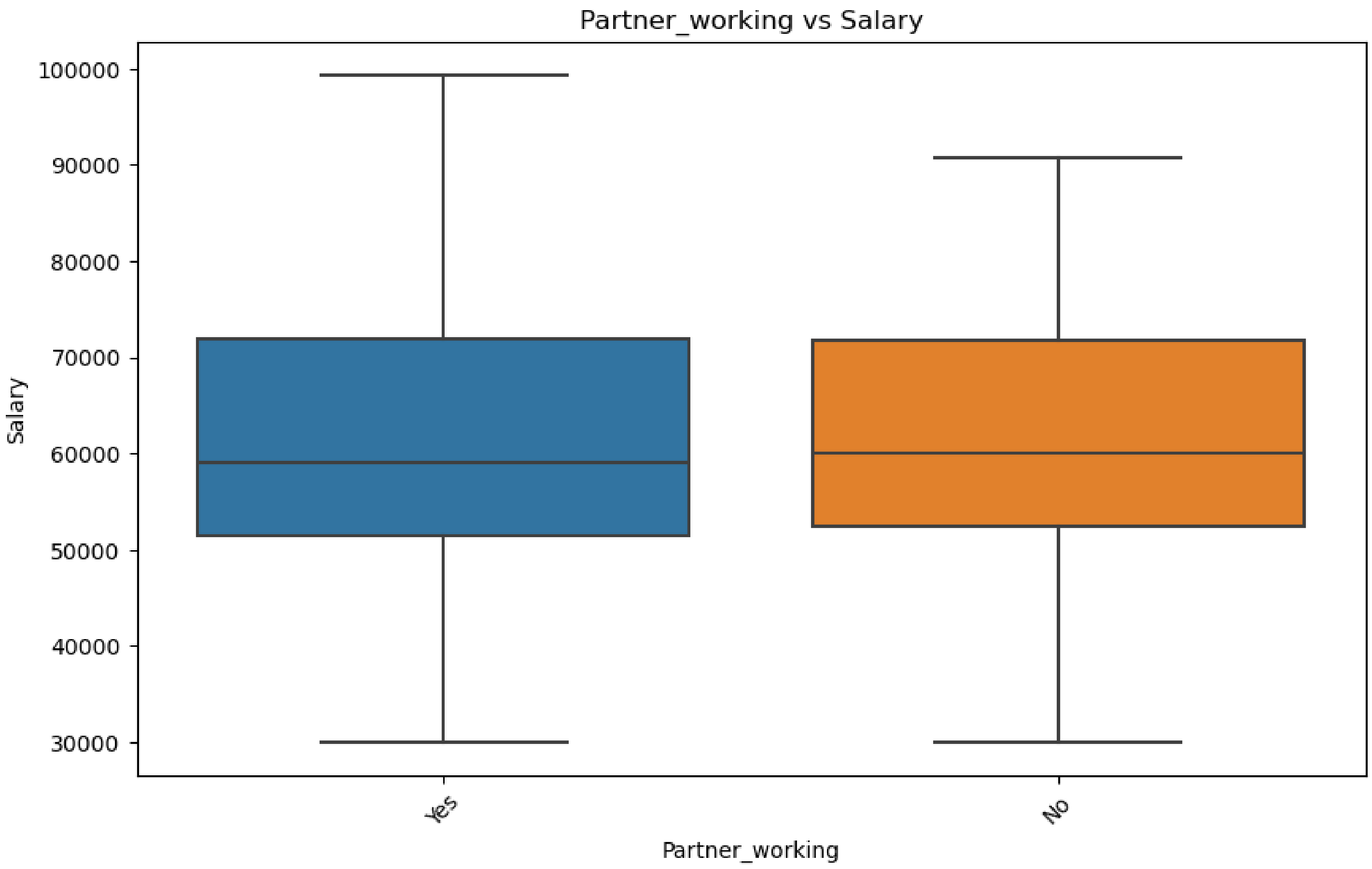

House_loan vs Partner_salary
80000
70000
60000
50000
40000
30000
20000
10000
0
Partner_salary
No
Yes
House_loan

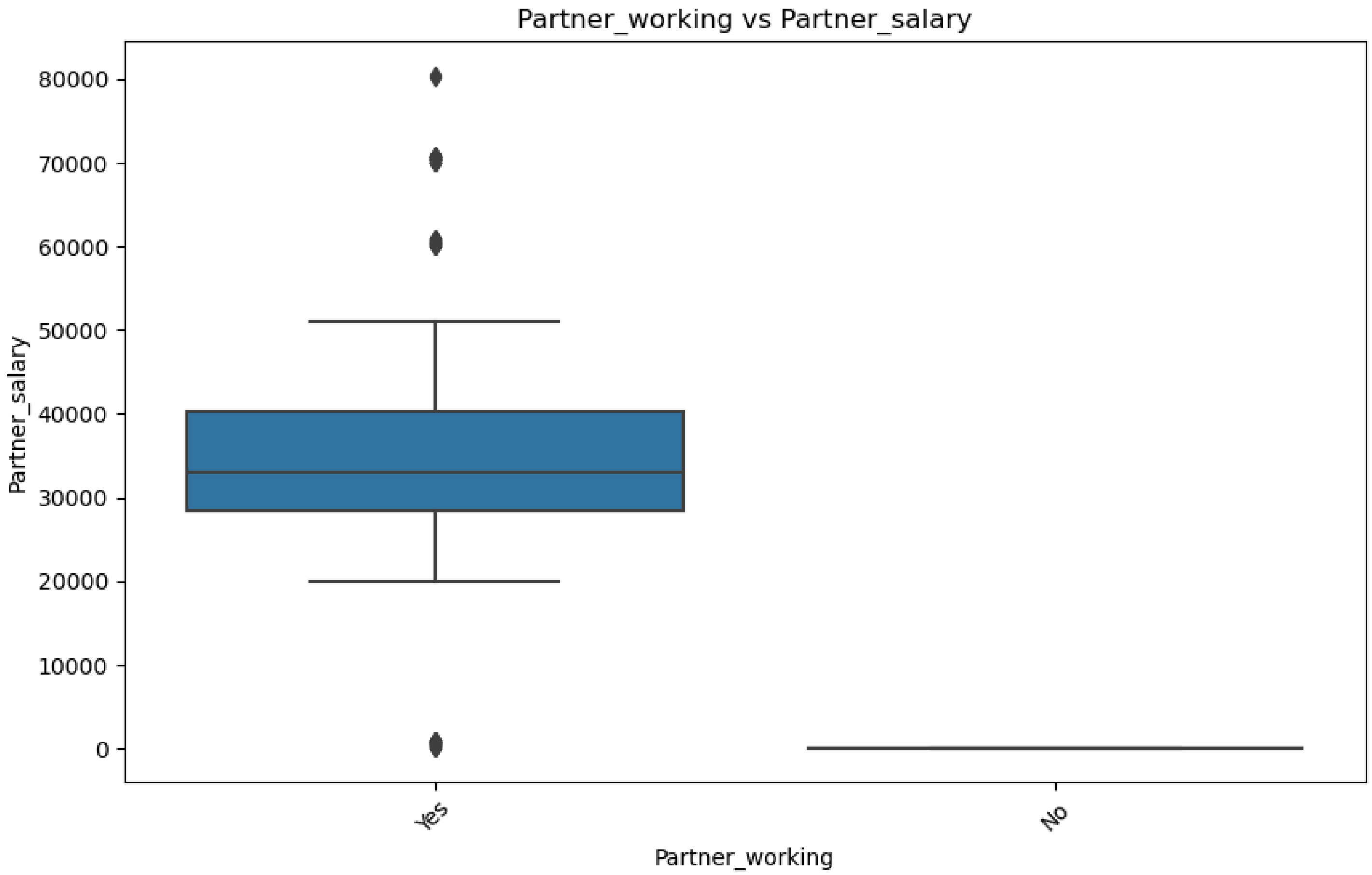

House_loan vs Total_salary
Total_salary
160000
140000
120000
100000
80000
60000
40000
No
Yes
House_loan

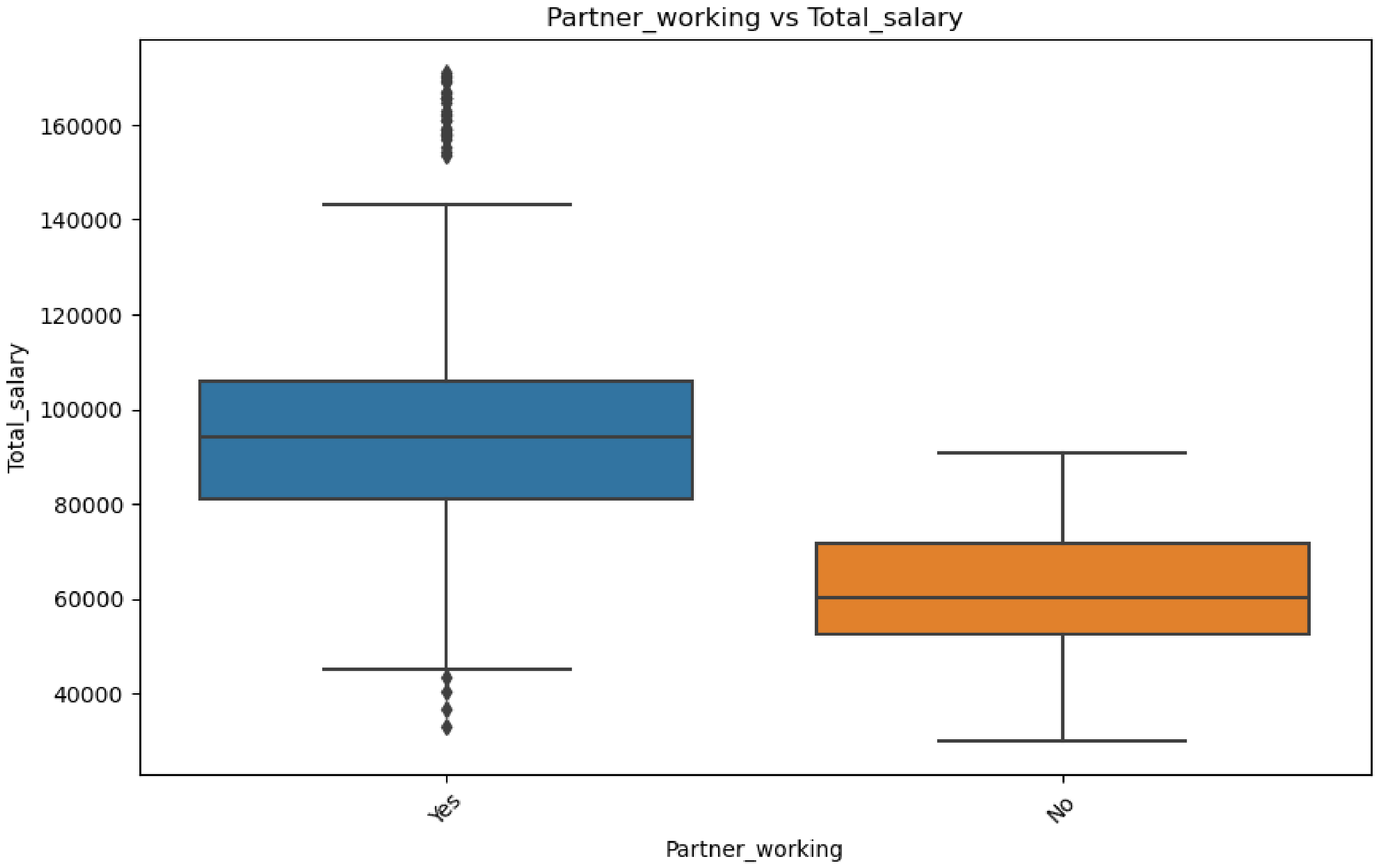

House_loan vs Price
Price
70000
60000
50000
40000
30000
20000
No
Yes
House_loan

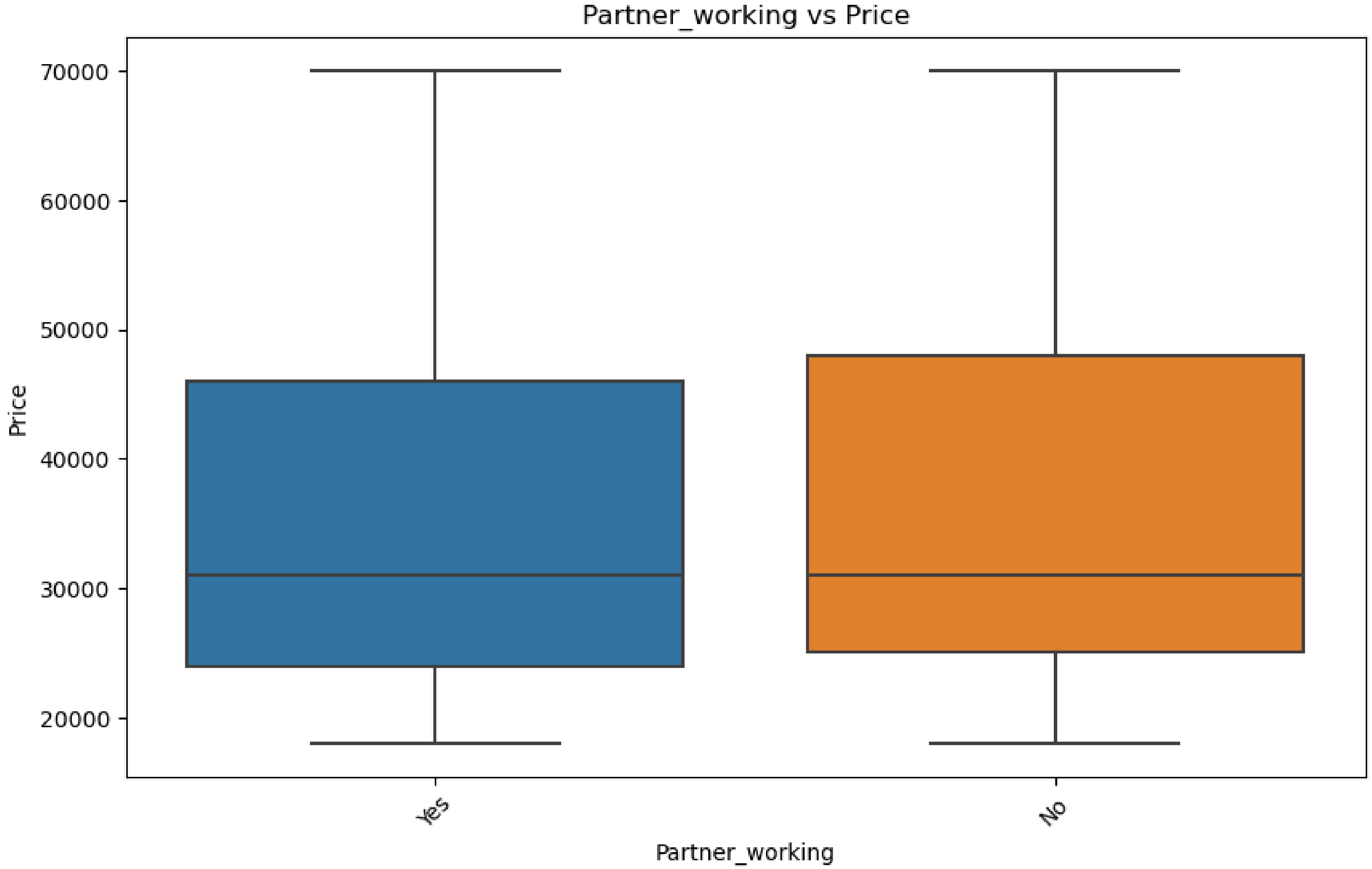

Partner_working vs Age
55
50
45
40
35
30
25
Age
Yes
No
Partner_working

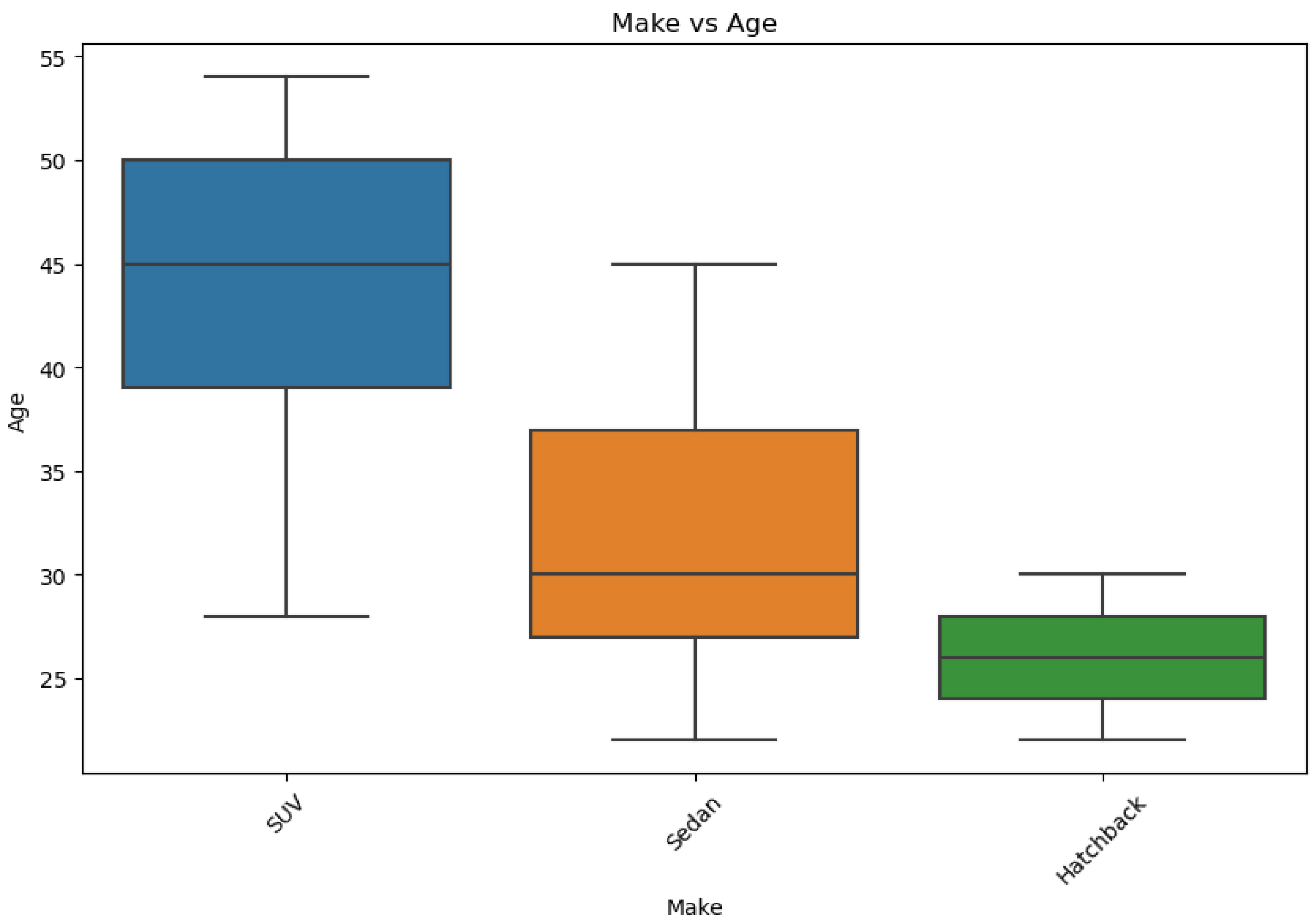

Partner_working vs No_of_Dependents
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
No_of_Dependents
Yes
No
Partner_working

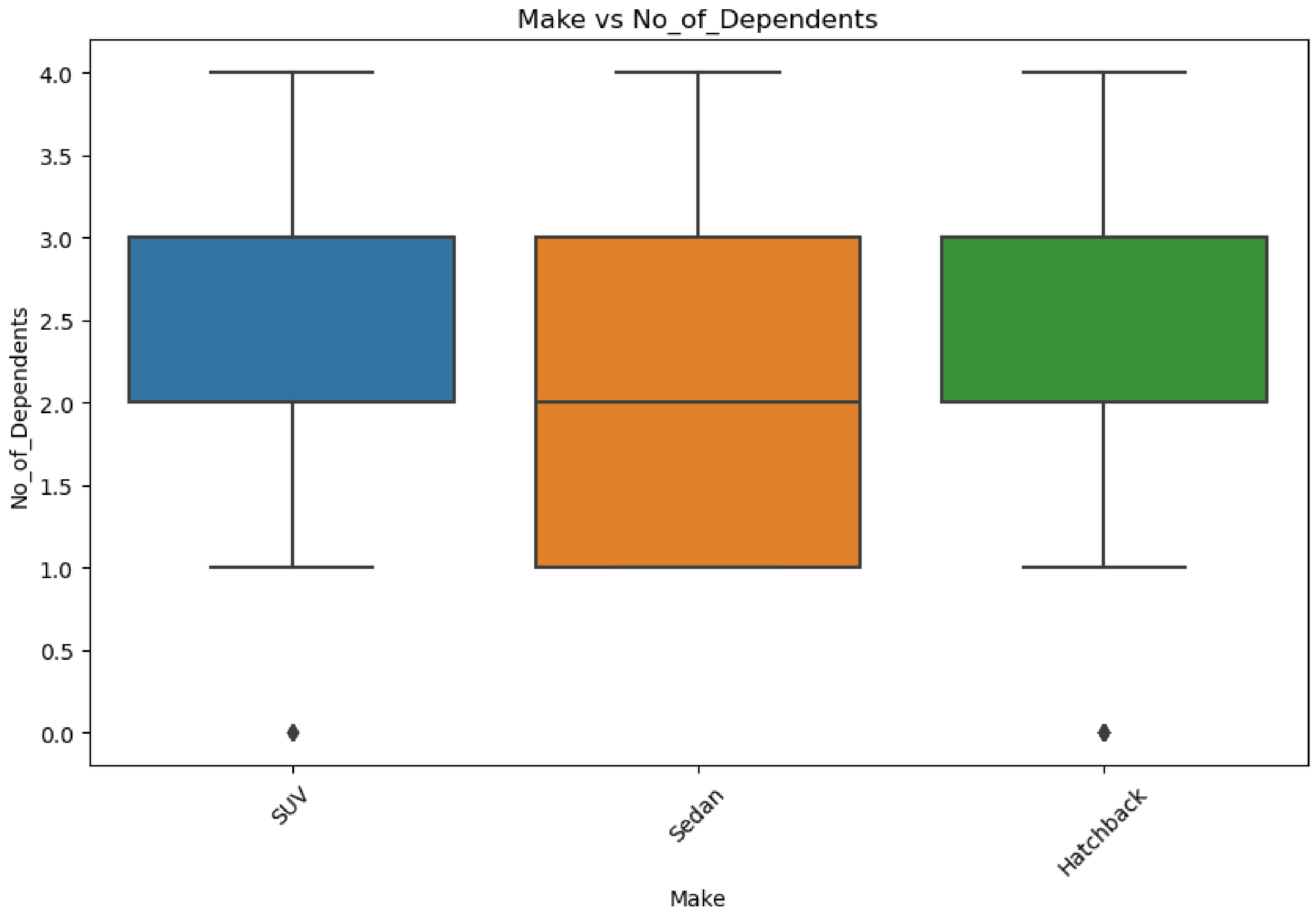

Partner_working vs Salary
Salary
100000
90000
80000
70000
60000
50000
40000
30000
Yes
No
Partner_working

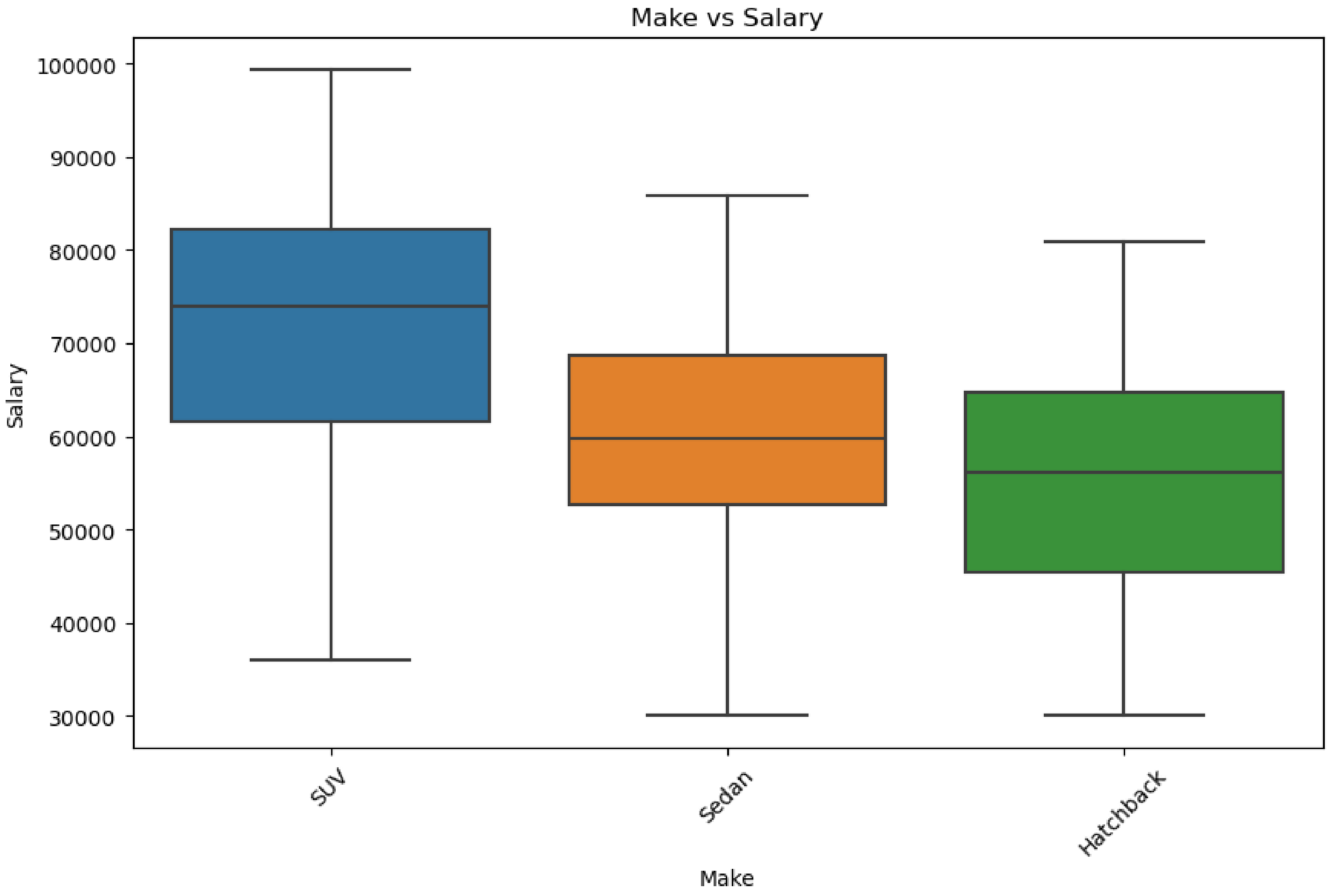

Partner_working vs Partner_salary
Partner_salary
80000
70000
60000
50000
40000
30000
20000
10000
0
Yes
No
Partner_working

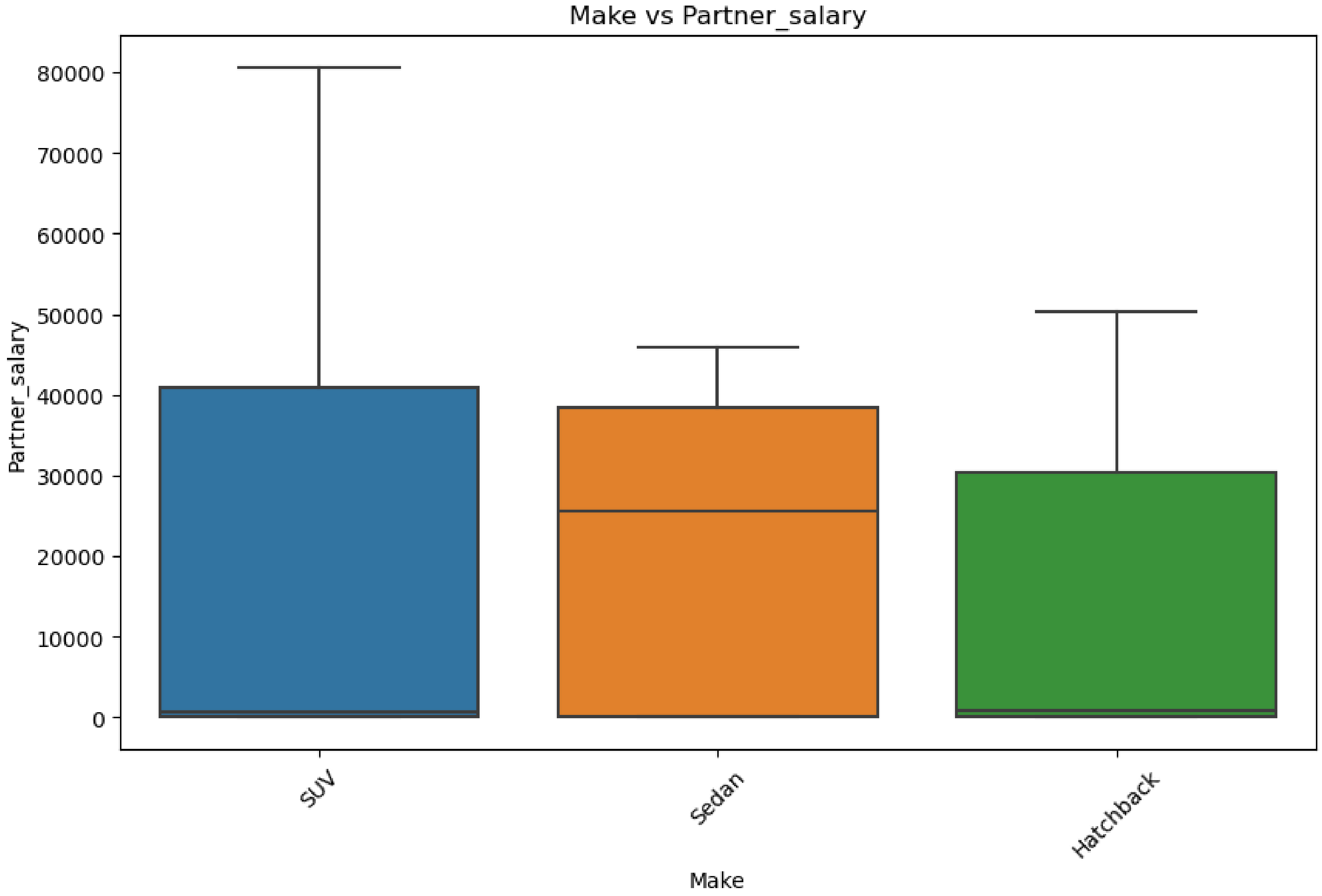

Partner_working vs Total_salary
Total_salary
160000
140000
120000
100000
80000
60000
40000
Yes
No
Partner_working


Partner_working vs Price
Price
70000
60000
50000
40000
30000
20000
Yes
No
Partner_working


Make vs Age
55
50
45
40
35
30
25
Age
SUV
Sedan
Hatchback
Make


Make vs No_of_Dependents
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
No_of_Dependents
SUV
Sedan
Hatchback
Make


Make vs Salary
Salary
100000
90000
80000
70000
60000
50000
40000
30000
SUV
Sedan
Hatchback
Make


Make vs Partner_salary
Partner_salary
80000
70000
60000
50000
40000
30000
20000
10000
0
SUV
Sedan
Hatchback
Make

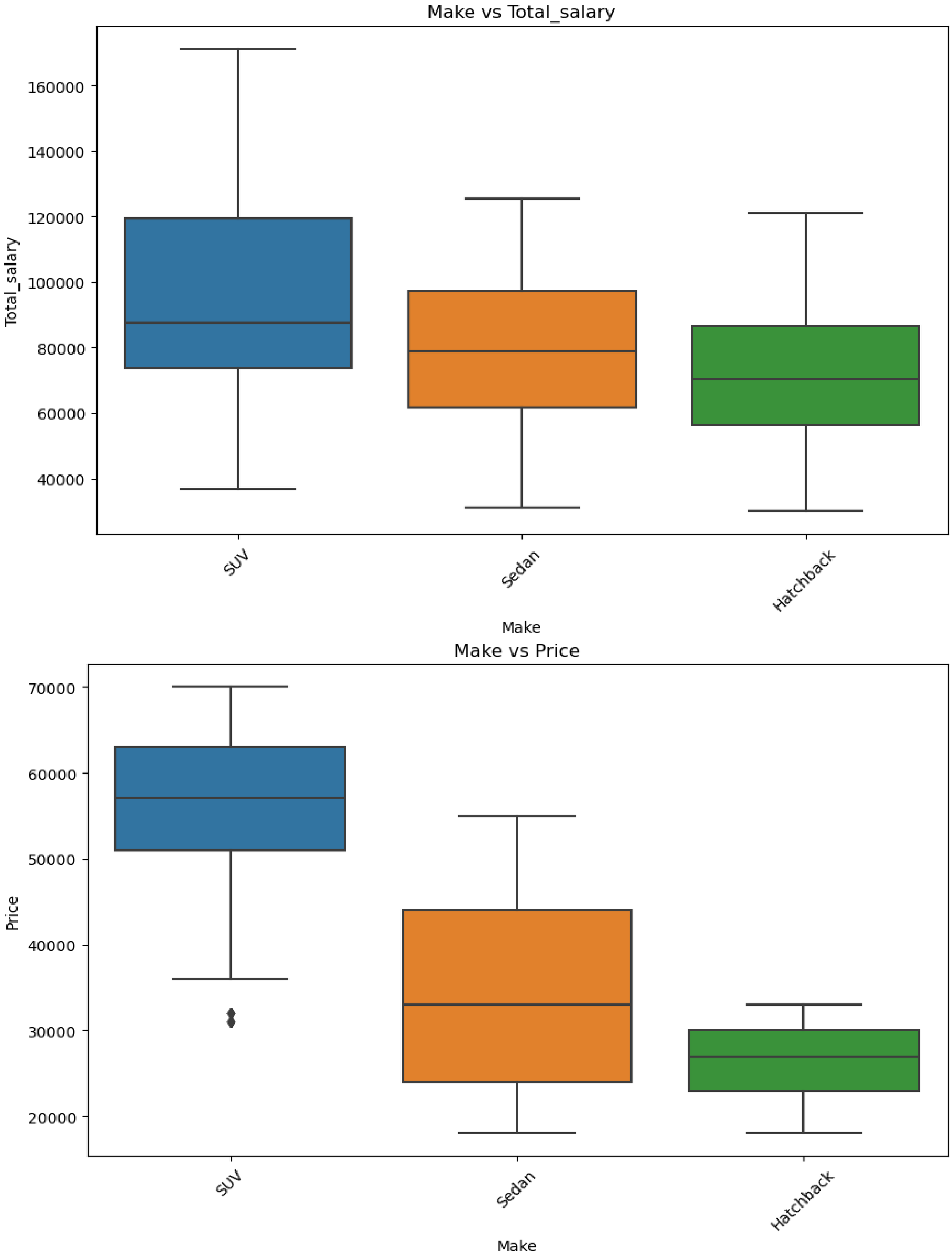


Figure-7 : Bivariate relationship of categorial vs numerical variables

Inferences:

1. Females are in higher age or older than males.
2. Females are having dependents 1-3, where males are having dependents 2-3.
3. Females are having salary ranging higher than males. While females are having salary range 34800-99300. While males are having salary range from 30000-99300.
4. Partner's salary of some females are little bit higher than partner's of males.
5. Females are having higher total salary than males.
6. Females have bought higher priced automobiles than males.
7. Some salaried buyers are older than the buyers having business.
8. Both salaried and business buyers are having 2-3 dependents.
9. Salaried buyers are little more earning than the buyers having business.
10. Salaried buyer's partner's salary is little higher, not so significant than the partners of buyers having business.
11. Some salaried buyer's total salary is having more outliers than the total salary of buyers having business.
12. Salaried buyers have spent more in automobiles than business buyers.
13. The buyers who are single are in near about within the same age group who are married.
14. Some single buyers are having 1 dependent.
15. Salary of both single and married buyers are nearly same.
16. Married buyers are having more total salary than the buyers who are single.
17. Married buyers have spent more on buying automobiles than the buyers who are single.
18. The buyers who are graduate are almost in the same age group of buyers who are having post-graduate.
19. Both the graduate and post-graduate buyers are having same number of dependents, i.e 2-3.
20. Post-graduate buyers are having more salary than the graduate buyers.
21. The graduate buyers are having a little more partner salary than the post-graduate buyers.
22. The post-graduate buyers are having more total salary than the graduate buyers.
23. The graduate buyers had spent a little more on buying the cars than post-graduate buyers.
24. The buyers whose salary is more then 70k have not taken personal loan.
25. Some buyers having 2-3 dependents have taken house loans.
26. Mostly the buyers who are having a working partner, they have larger total salary than the buyers who are not having working partner.

27. The buyers of age group 38-50 have preferred to buy SUV, where as the age group of 27-37 have preferred to buy sedan and the buyers who have preferred to buy Hatchback are 25-28 years.
28. The buyers having 2-3 dependents have preferred SUV and Hatchback, where as the buyers having 1-3 dependents have purchased Sedan.
29. The buyers having 62k-82k salaried have purchased SUV, 52k-68k salaried have purchased Sedan and 44k-66k salaried buyers have purchased Hatchback.
30. The buyers having 66k-120k total salary have purchased SUV, 64k-88k total salary have purchased Sedan and 60k-84k total salary buyers have purchased Hatchback.
31. SUV price range purchased are 52k-64K, where as Sedan cars have been purchased in mostly price range of 26k-44k and Hatchback cars have been mostly purchased in a range of 24k-30k.

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

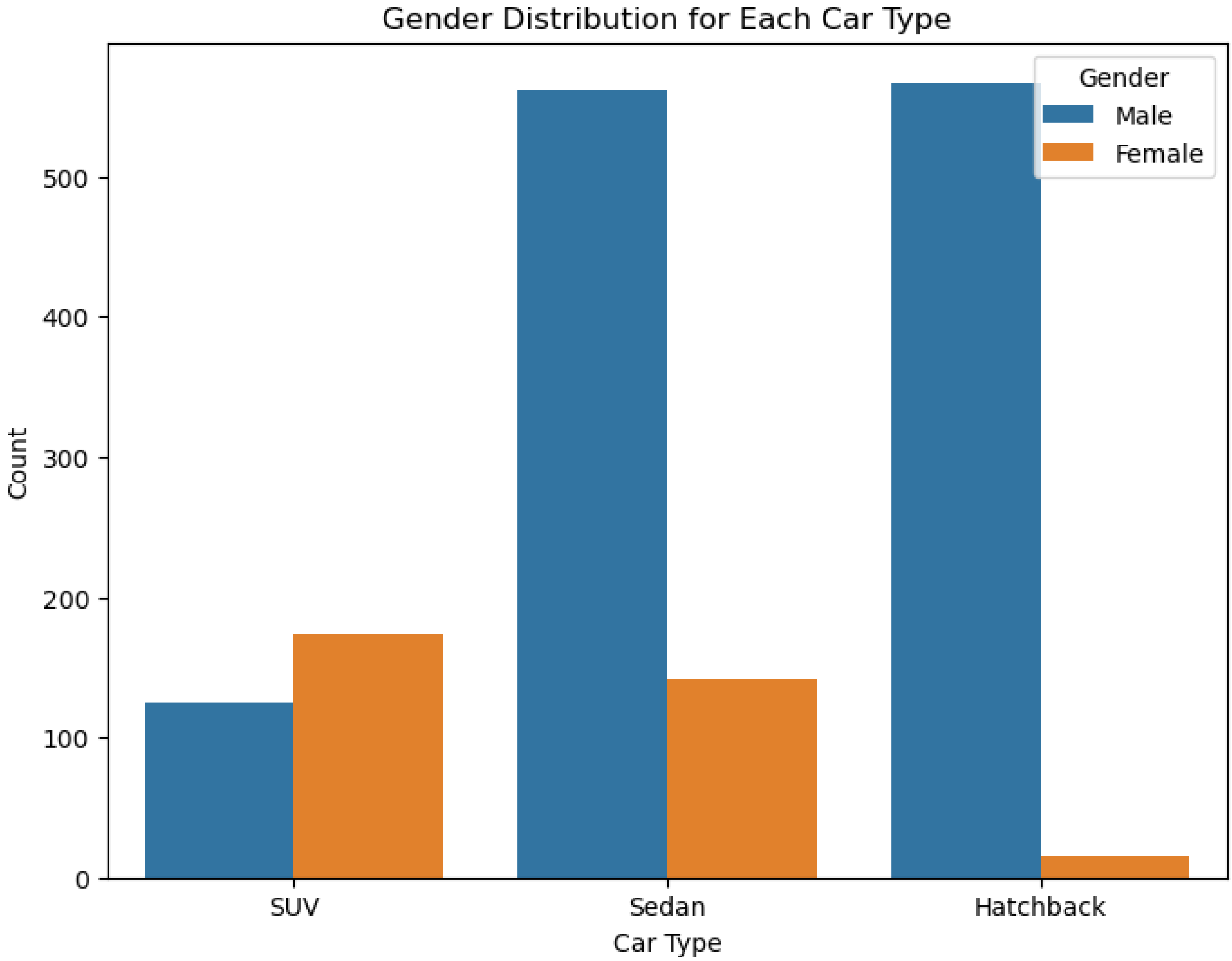


Figure-8 : Count plot of Gender vs Make

Ans:

We can properly see that the women more likely prefer SUVs compared to men.

So, the answer for the Q1 is ‘No’.

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

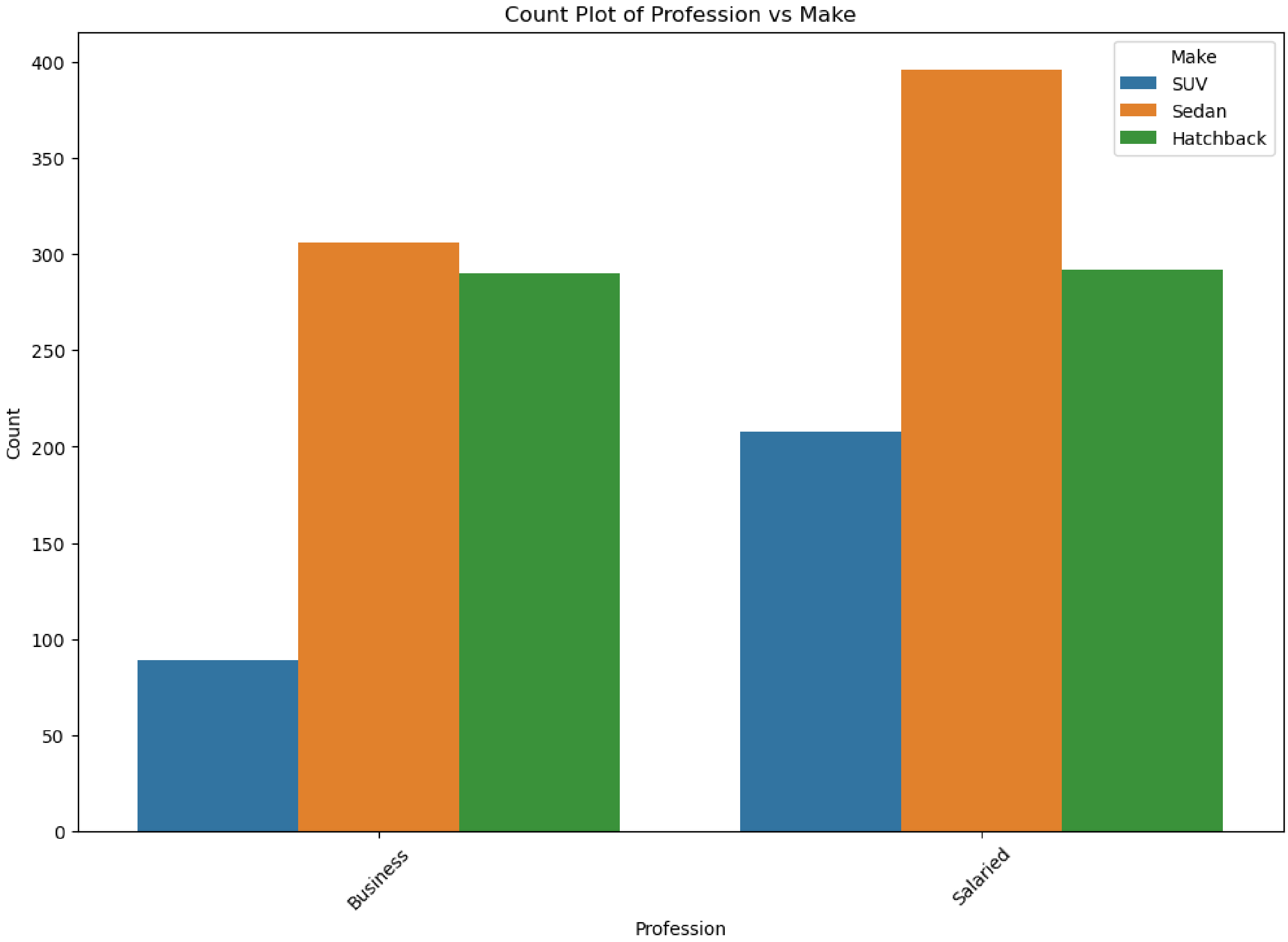


Figure-9 : Count plot of Profession vs Make

Ans:

From the above chart, it is evident that salaried person is more likely to buy a Sedan.

So, this statement is True.

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?

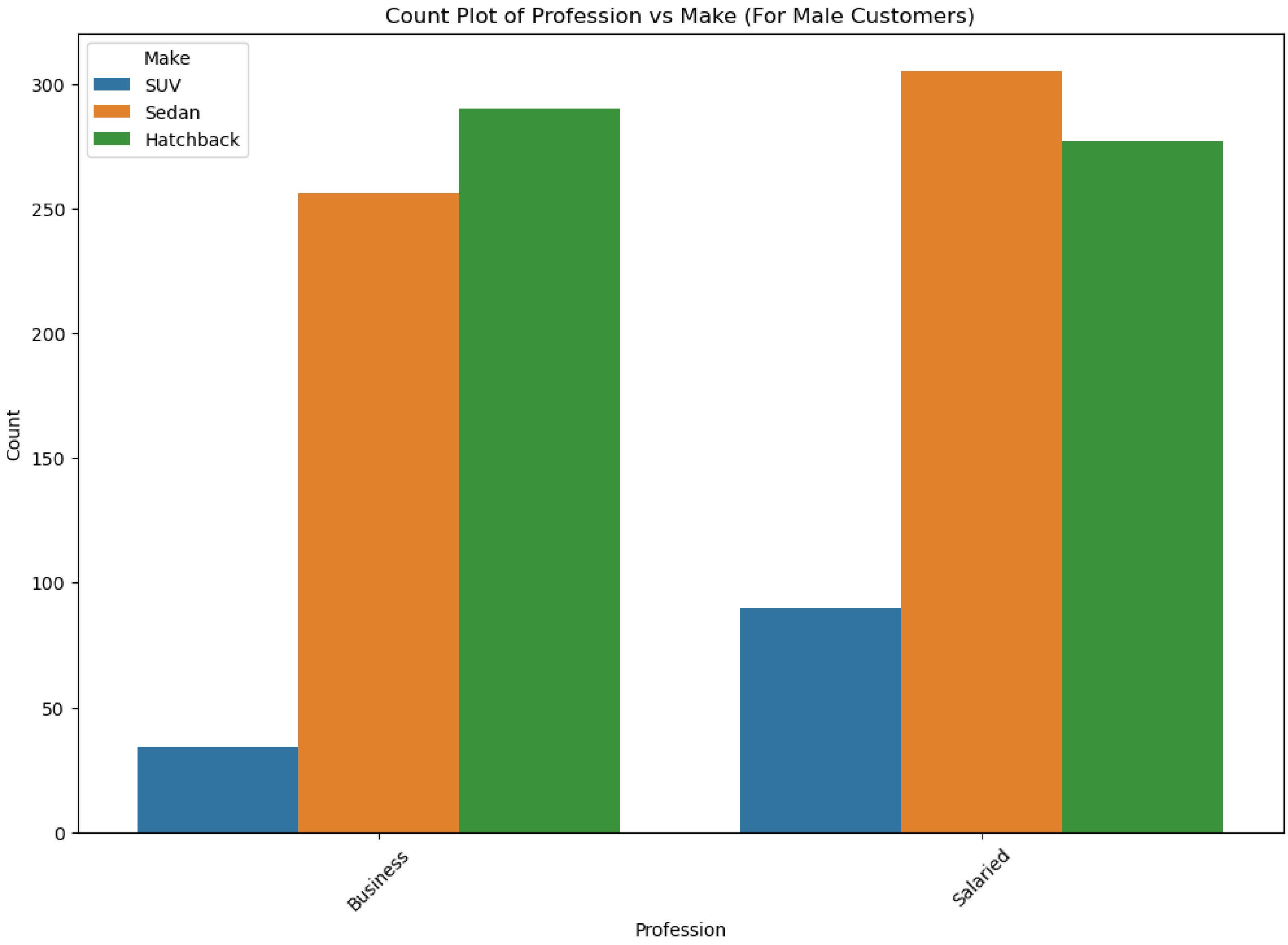


Figure-10 : Count plot of Profession vs Make for Male buyers

Ans:

From the above chart, it is evident that Salaried male prefers Sedan over SUV.

So, this statement is incorrect.

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

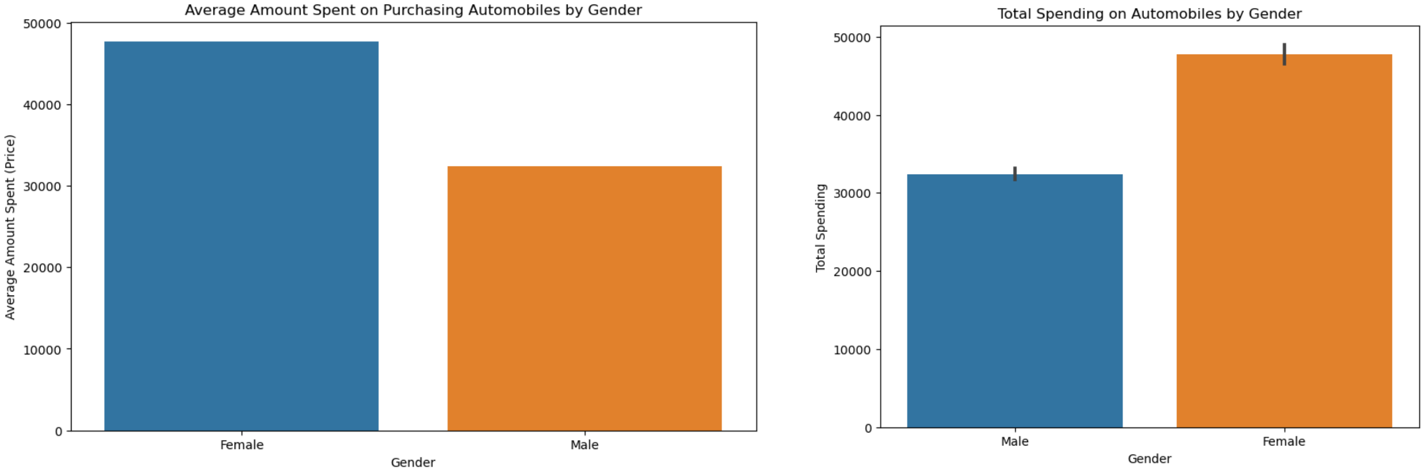


Figure-11 : Count plot of spending on automobiles by Gender

Ans:

We can clearly see that the spending on automobiles is done by females than males.

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

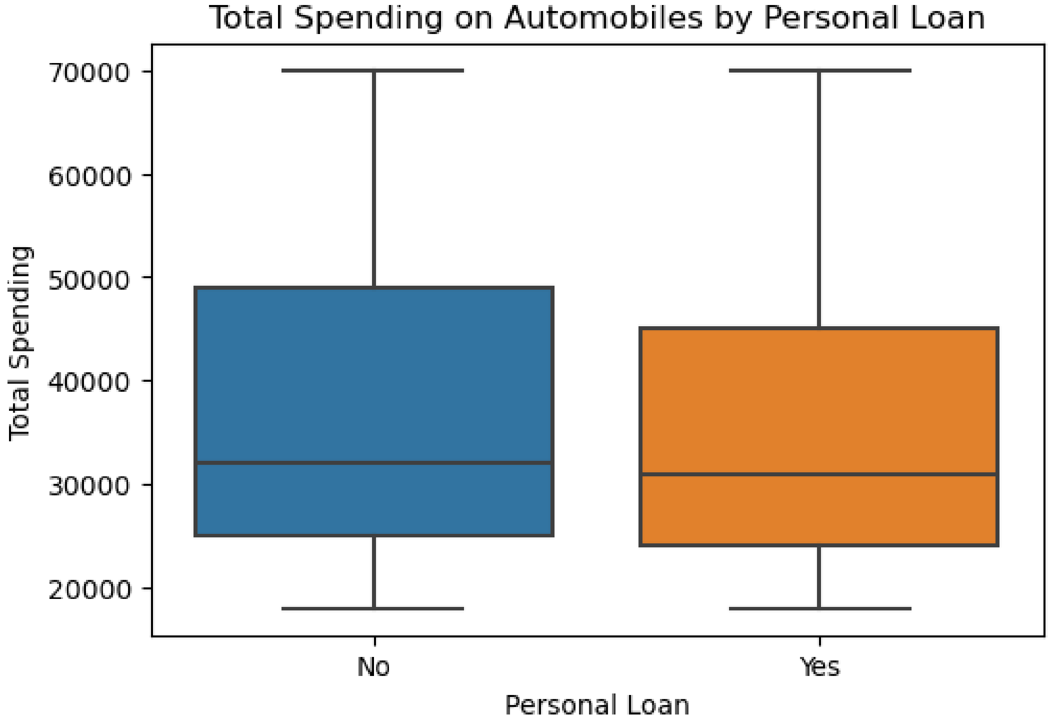
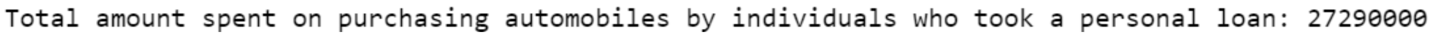


Figure-12 : Box plot of total spending on automobiles by personal loan

Ans:



```
Total amount spent on purchasing automobiles by individuals who took a personal loan: 27290000
```

Q6. How does having a working partner influence the purchase of higher-priced cars?

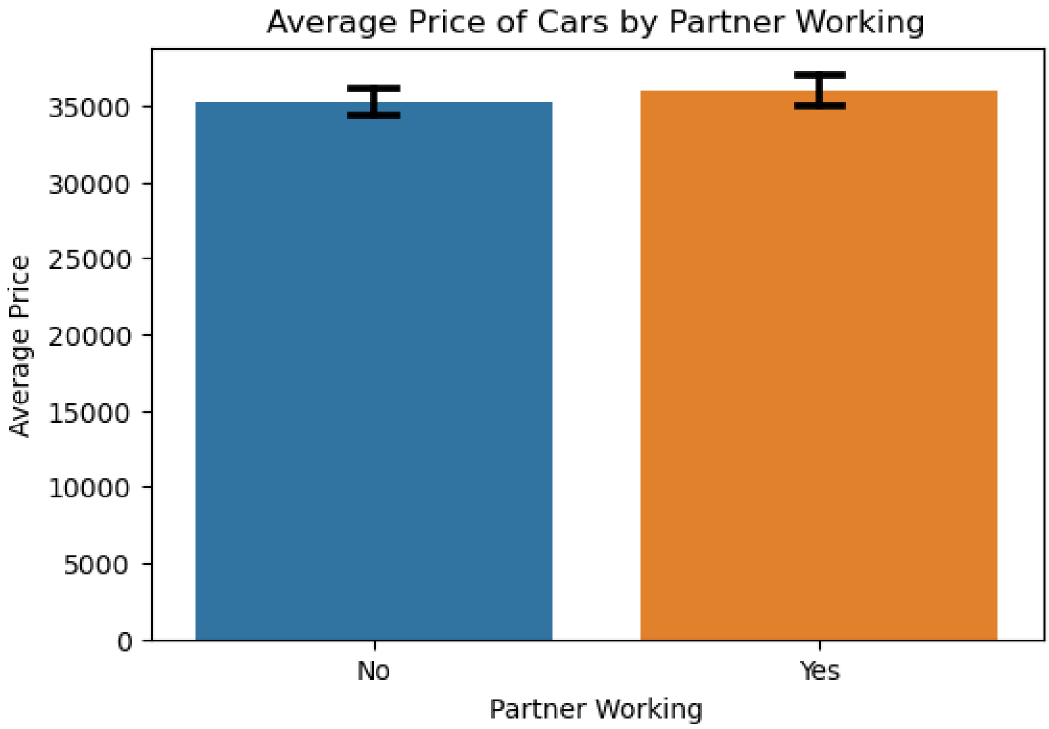
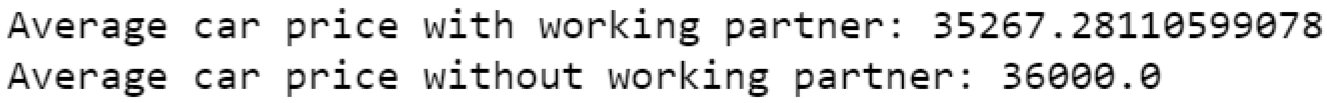


Figure-13 : Box plot of average price of cars by partner working

Ans:



```
Average car price with working partner: 35267.28110599078
Average car price without working partner: 36000.0
```