Project report

```
import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

import numpy as np

import matplotlib.pyplot as plt

import seaborn as sns

import numpy as np

file_path = r"C:\Users\91931\Documents\tech_project\5-bright_automotive_company_New.csv"

df = pd.read_csv(file_path)

print(df)
```

```
      Age  Gender Profession Marital_status      Education No_of_Dependents  \
0      53    Male   Business         Married  Post Graduate                 4
1      53   Femal   Salaried         Married  Post Graduate                 4
2      53  Female   Salaried         Married  Post Graduate                 3
3      53  Female   Salaried         Married       Graduate                 ?
4      53    Male        NaN         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  99300.0         70700.0
1              Yes         No             Yes  95500.0         70300.0
2               No         No             Yes  97300.0         60700.0
3              Yes         No             Yes  72500.0         70300.0
4               No         No             Yes  79700.0         60200.0
...            ...        ...             ...      ...             ...
1576            No        Yes              No  33300.0             0.0
1577            No         No              No  32000.0             NaN
1578            No        Yes              No  32900.0             0.0
1579           Yes        Yes              No  32200.0             NaN
1580            No         No              No  31600.0             0.0

      Total_salary  Price       Make
0           170000  61000        SUV
1           165800  61000        SUV
2           158000  57000        SUV
3           142800  61000          ?
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]
```

1-Display the top 5 rows.

Age	Gender	Profession	Marital_status	Education	No_of_Dependents	Personal_loan	House_loan	Partner_working	Salary	Partner_salary	Total_salary	Price	Make
53	Male	Business	Married	Post Graduate	4	No	No	Yes	99300.0	70700.0	170000	61000	SUV
53	Femal	Salaried	Married	Post Graduate	4	Yes	No	Yes	95500.0	70300.0	165800	61000	SUV
53	Female	Salaried	Married	Post Graduate	3	No	No	Yes	97300.0	60700.0	158000	57000	SUV
53	Female	Salaried	Married	Graduate	?	Yes	No	Yes	72500.0	70300.0	142800	61000	?
53	Male	NaN	Married	Post Graduate	3	No	No	Yes	79700.0	60200.0	139900	57000	SUV

2-Display the last 5 rows

	Age	Gender	Profession	Marital_status	Education	No_of_Dependents	Personal_loan	House_loan	Partner_working	Salary	Partner_salary	Total_salary	Price
1576	22	Male	Salaried	Single	Graduate	2	No	Yes	No	33300.0	0.0	33300	27000
1577	22	Male	Business	Married	Graduate	4	No	No	No	32000.0	NaN	32000	31000
1578	22	Male	Business	Single	Graduate	2	No	Yes	No	32900.0	0.0	32900	30000
1579	22	Male	Business	Married	Graduate	3	Yes	Yes	No	32200.0	NaN	32200	24000
1580	22	Male	Salaried	Married	Graduate	4	No	No	No	31600.0	0.0	31600	31000

3-Check the shap of the dataset.

```
shap of the dataset : (1581, 14)
```

4-Check the Statistical summary

```
Statistical summary of numerical features:
                Age        Salary  Partner_salary   Total_salary          Price
count  1581.000000   1568.000000     1475.000000    1581.000000    1581.000000
mean     31.952562  60276.913265    20225.559322   79625.996205   35948.170778
std       8.712549  14636.200199    19573.149277   25545.857768   21175.212108
min      14.000000  30000.000000        0.000000   30000.000000      58.000000
25%      25.000000  51900.000000        0.000000   60500.000000   25000.000000
50%      29.000000  59450.000000    25600.000000   78000.000000   31000.000000
75%      38.000000  71700.000000    38300.000000   95900.000000   47000.000000
max     120.000000  99300.000000    80500.000000  171000.000000  680000.000000
```

5-Check the null values

```
null values in each column:
 Age                    0
Gender                 53
Profession              6
Marital_status          0
Education               0
No_of_Dependents        0
Personal_loan           0
House_loan              0
Partner_working         0
Salary                 13
Partner_salary        106
Total_salary            0
Price                   0
Make                    0
dtype: int64
```

6-Check the duplicate values

```
 duplicate_rows = df.duplicated().sum()
print ("no of  duplicate values in rows :", duplicate_rows )
```

```
no of  duplicate values in rows : 0
```

```
duplicate_columns = df.duplicated().sum()
print ("no of  duplicate values in columns :", duplicate_columns )
```

```
no of  duplicate values in columns : 0
```

7-Check the anomalies or wrong entries.

```
     Age  Gender Profession Marital_status      Education No_of_Dependents  \
18   120  Female   Business        Married  Post Graduate                3
414   14    Male   Salaried        Married  Post Graduate                2

    Personal_loan House_loan Partner_working   Salary  Partner_salary  \
18             No         No             Yes  92600.0         70300.0
414           Yes        Yes             Yes  80600.0         40500.0

     Total_salary  Price   Make
18         162900  58000    SUV
414        121100  43000  Sedan
     Age  Gender Profession Marital_status      Education No_of_Dependents  \
18   120  Female   Business        Married  Post Graduate                3
414   14    Male   Salaried        Married  Post Graduate                2

    Personal_loan House_loan Partner_working   Salary  Partner_salary  \
18             No         No             Yes  92600.0         70300.0
414           Yes        Yes             Yes  80600.0         40500.0

     Total_salary  Price   Make
18         162900  58000    SUV
414        121100  43000  Sedan
```

8-Check the outliers and their authenticity.

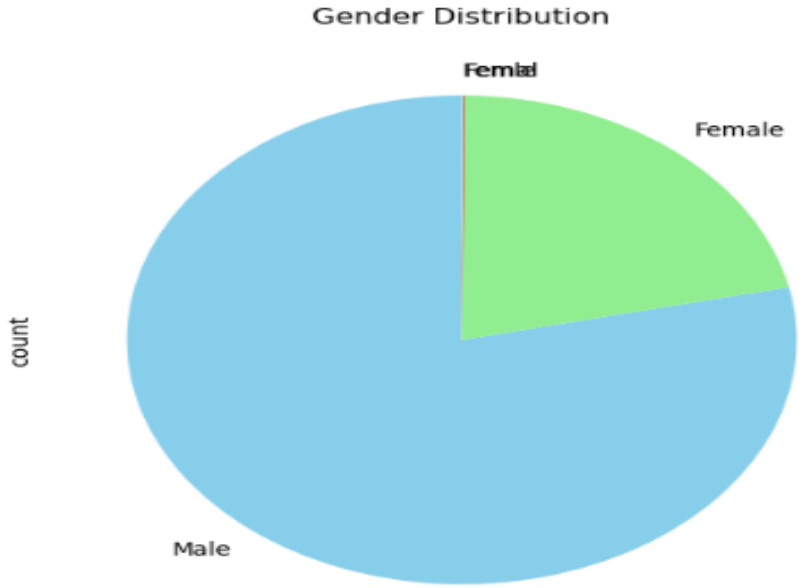
```
     Age  Gender Profession Marital_status      Education No_of_Dependents  \
18   120  Female   Business        Married  Post Graduate                3
414   14    Male   Salaried        Married  Post Graduate                2

    Personal_loan House_loan Partner_working   Salary  Partner_salary  \
18             No         No             Yes  92600.0         70300.0
414           Yes        Yes             Yes  80600.0         40500.0

     Total_salary  Price   Make
18         162900  58000    SUV
414        121100  43000  Sedan
     Age  Gender Profession Marital_status      Education No_of_Dependents  \
18   120  Female   Business        Married  Post Graduate                3
414   14    Male   Salaried        Married  Post Graduate                2

    Personal_loan House_loan Partner_working   Salary  Partner_salary  \
18             No         No             Yes  92600.0         70300.0
414           Yes        Yes             Yes  80600.0         40500.0

     Total_salary  Price   Make
18         162900  58000    SUV
414        121100  43000  Sedan
```

What are the mean, median, and standard deviation of the ages of individuals in the dataset?

```
Mean age: 31.952561669829223
Median age: 29.0
Standard deviation of age: 8.71254886208395
```

What is the distribution of gender in the dataset? Represent it using a pie chart.

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



Is there a correlation between age and salary? Provide the correlation coefficient and interpret the result.

```
Correlation coefficient between Age and Salary: 0.602354735794213
There is a positive correlation between age and salary.
```

What is the average number of dependents for married individuals versus single individuals?

```
Average number of dependents for married individuals: 2.5385149201943094
Average number of dependents for single individuals: 1.608695652173913
```

How does the employment status of a partner affect the total combined salary?

```
Partner_working
No     60820.819672
Yes    95222.065728
Name: Total_Combined_Salary, dtype: float64
```

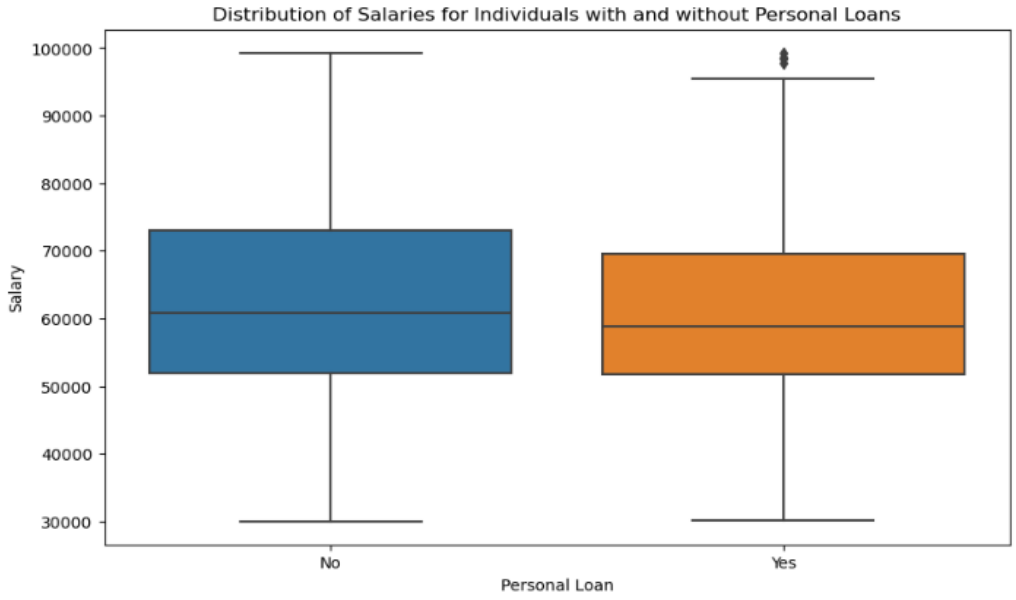
Compare the average salary of individuals whose partners are working versus those whose partners are not working.

```
Partner_Working
Not Working    60276.913265
Name: Salary, dtype: float64
```

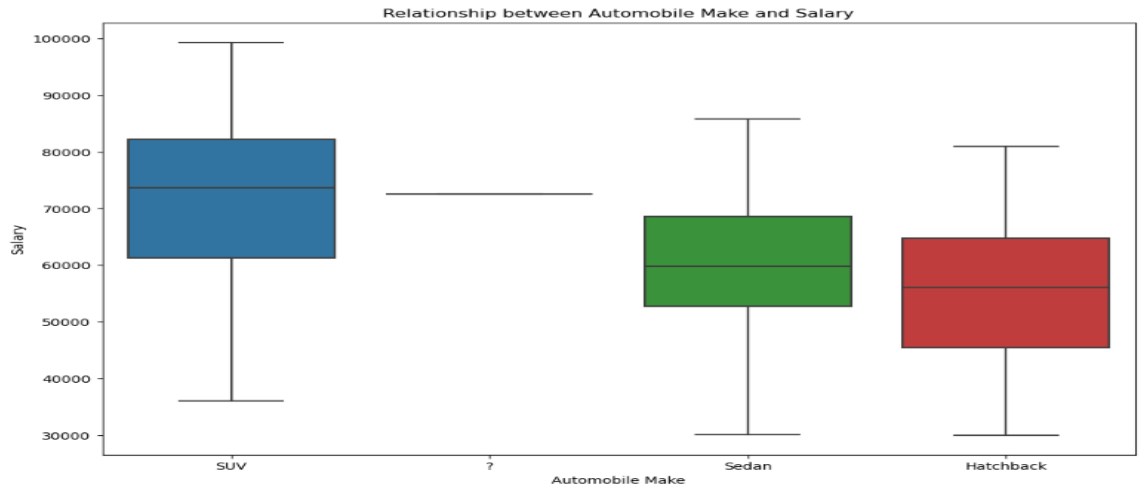
What is the proportion of individuals with house loans based on their profession?

```
House_loan  Proportion_with_Loan
Profession
Business                0.334307
Salaried                0.334831
```

What is the distribution of salaries for individuals with personal loans versus those without personal loans? Represent it using a box plot.



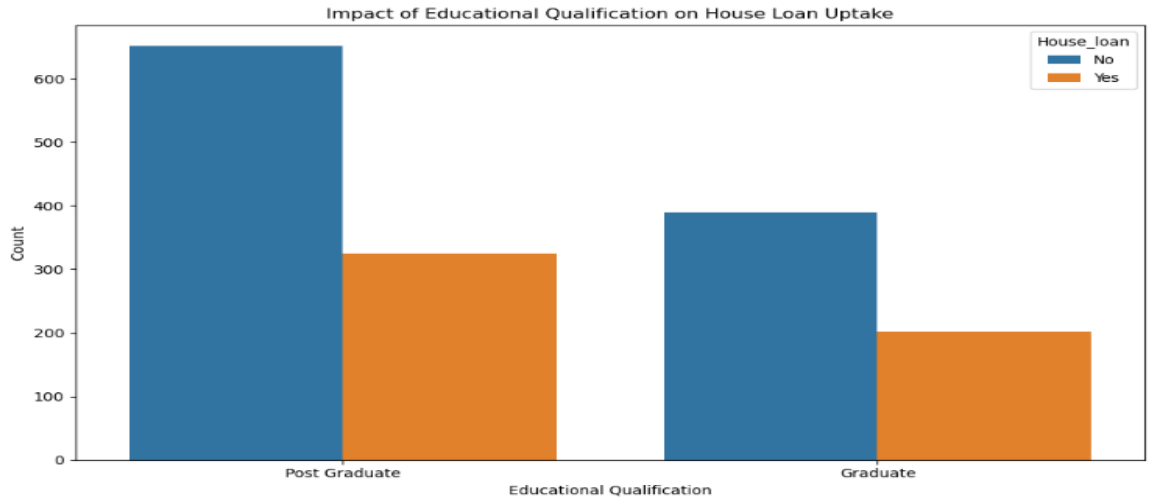
How does the type of automobile relate to the salary of the individuals? Provide insights based on the make of the automobile.



What is the average price of the product/service in the dataset? How does this price vary based on the individual's total salary?

```
The average price of the product/service is: $35801.02

Price variation based on individual's total salary:
For a total salary of 30000.00, the average price is 18000.00
For a total salary of 30600.00, the average price is 22000.00
For a total salary of 30900.00, the average price is 28000.00
For a total salary of 31100.00, the average price is 20000.00
For a total salary of 31200.00, the average price is 18000.00
For a total salary of 31300.00, the average price is 26000.00
For a total salary of 31500.00, the average price is 31000.00
For a total salary of 31600.00, the average price is 31000.00
For a total salary of 31800.00, the average price is 28000.00
For a total salary of 31900.00, the average price is 28000.00
For a total salary of 32000.00, the average price is 25000.00
For a total salary of 32200.00, the average price is 24333.33
For a total salary of 32300.00, the average price is 28500.00
For a total salary of 32500.00, the average price is 26000.00
For a total salary of 32900.00, the average price is 25000.00
```

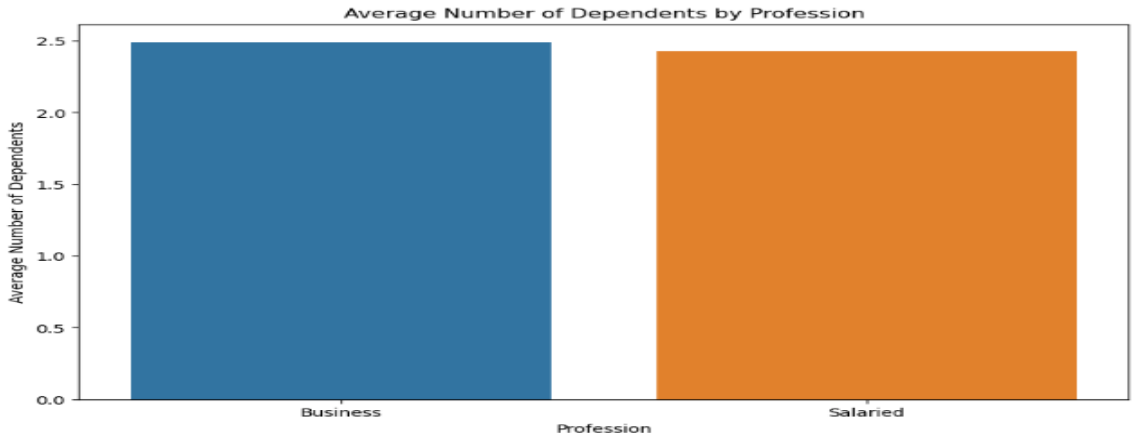
How does educational qualification impact the likelihood of taking a house loan?




```
Percentage of individuals taking house loans by educational qualification:
       Education House_loan  Percentage
0       Graduate         No   65.878378
1       Graduate        Yes   34.121622
2  Post Graduate         No   66.700820
3  Post Graduate        Yes   33.299180
```

Analyze the number of dependents based on the profession of the individual. Which profession has the highest average number of dependents?

```
Profession with the highest average number of dependents: Business
Average number of dependents: 2.491202346041057
```



Is there a significant difference in salaries between males and females? Provide statistical evidence.

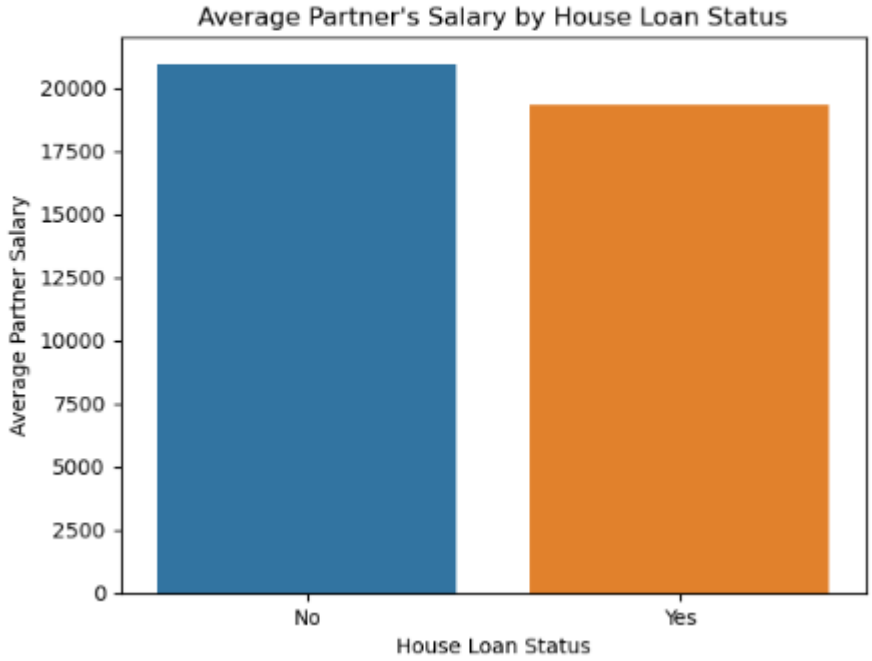
```
T-statistic: -7.999645190591819
P-value: 2.462049251138918e-15
There is a significant difference in salaries between males and females.
```

How does having a personal loan affect the total combined salary of the individual and their partner?

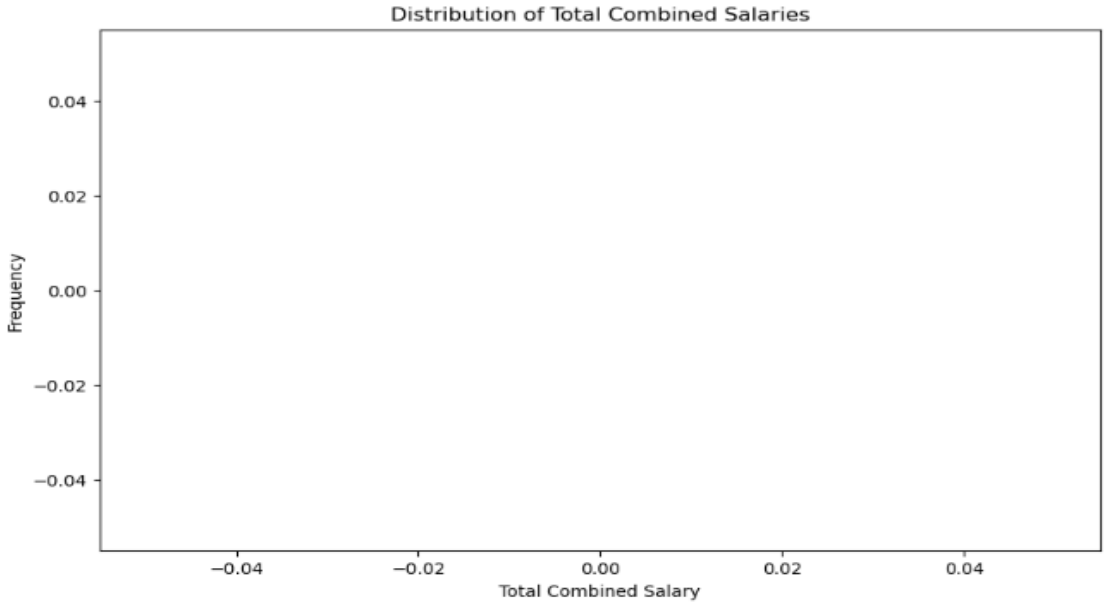
```
Average Total Combined Salary by Personal Loan Status:
Personal_loan
No     82410.807114
Yes    79326.402189
Name: Total_Combined_Salary, dtype: float64
```

What is the average partner's salary for individuals with and without house loans?

```
Average Partner's Salary by House Loan Status:
House_loan
No     20956.521739
Yes    19332.056452
Name: Partner_salary, dtype: float64
```



Create a histogram showing the distribution of total combined salaries. Identify and discuss any skewness or outliers in the data.




```
Skewness: nan
Number of outliers: 0
Outliers:
Empty DataFrame
Columns: [Salary, Partner_salary, Total_Combined_Salary]
Index: []
```