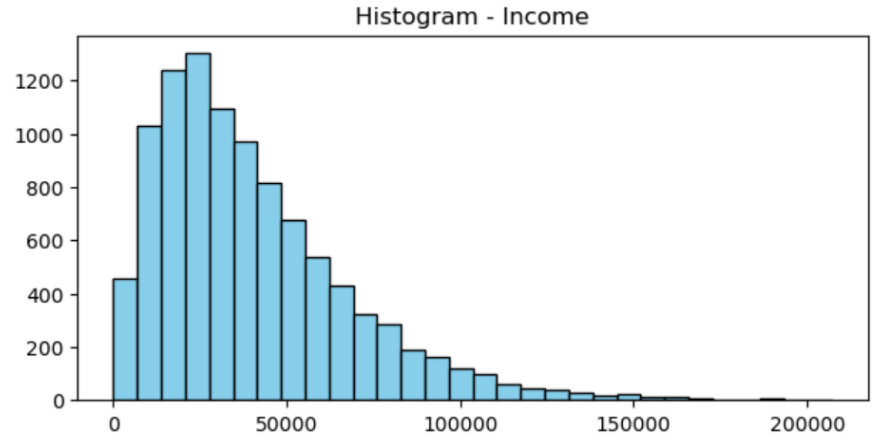
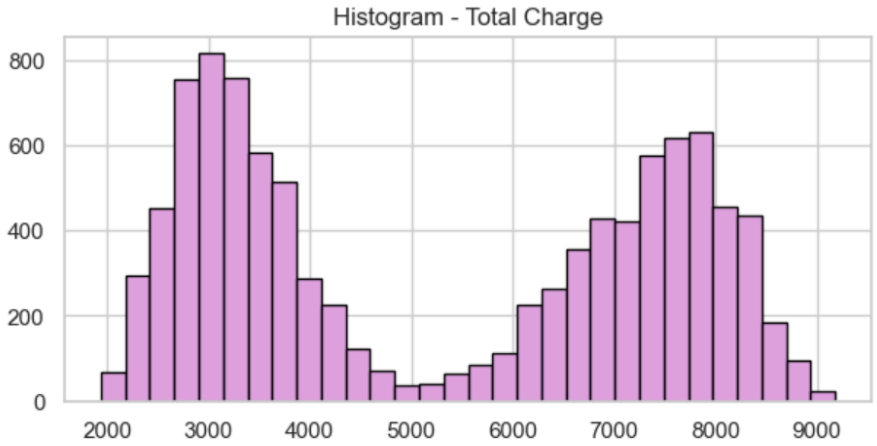
Univariate Statistics (2 continuous and 2 categorical)

C1. Visual



The income histogram shows a right-skewed distribution. This suggests that most patients have incomes below the average, while a few have very high incomes.

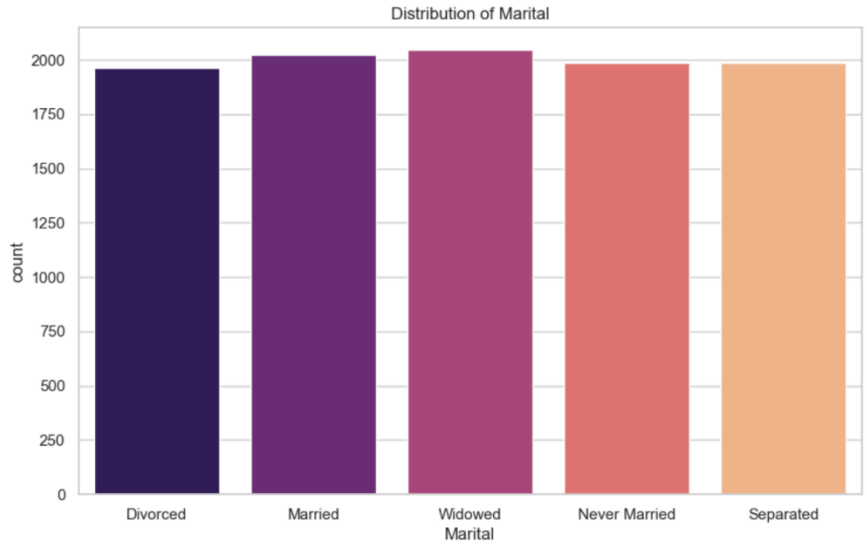
```
Mean Income: 40490.49516
Median Income: 33768.42
Variance Income: 813456185.1732982
Standard Deviation Income: 28521.15329318396
```



The total charge histogram shows a bimodal distribution. This suggests that there are two type of patients – those who are charged a lot and those who were charged a little.

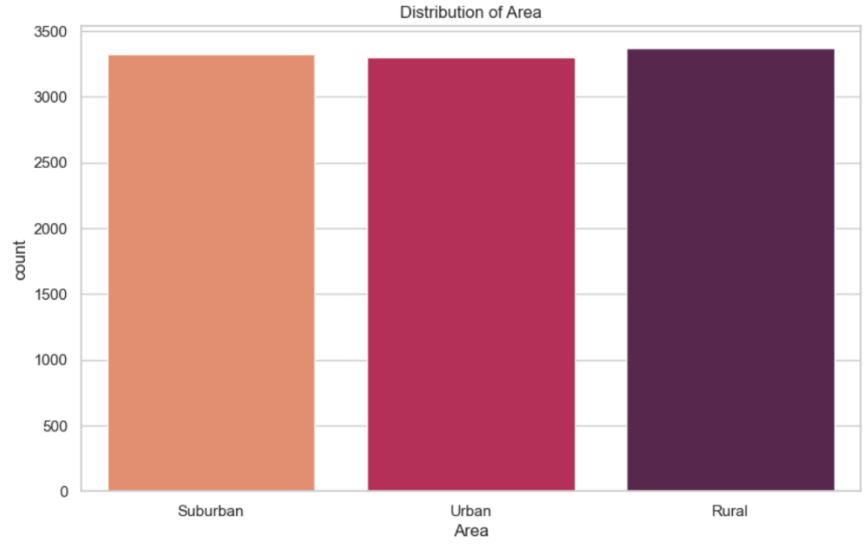
```
Mean Total Charge: 5312.1727687502
Median Total Charge: 5213.952
Variance Total Charge: 4754117.287963928
Standard Deviation Total Charge: 2180.393837810942
```

```
Frequency Distribution of Marital:
Marital
Widowed          2045
Married          2023
Separated        1987
Never Married    1984
Divorced         1961
Name: count, dtype: int64
```



The total charge histogram shows a uniform distribution. This suggests that the patient's marital status is evenly distributed across the five different categories.

```
Frequency Distribution of Area:
Area
Rural       3369
Suburban    3328
Urban       3303
Name: count, dtype: int64
```

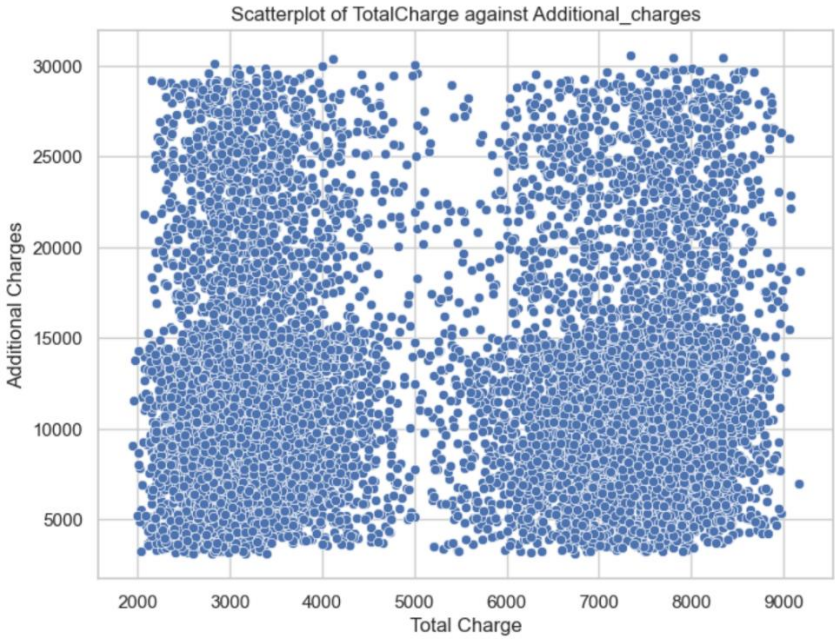


The area histogram shows a uniform distribution. This suggests that the patient's area classification is evenly distributed across the three categories.

A. Bivariate Statistics (2 continuous and 2 categorical)

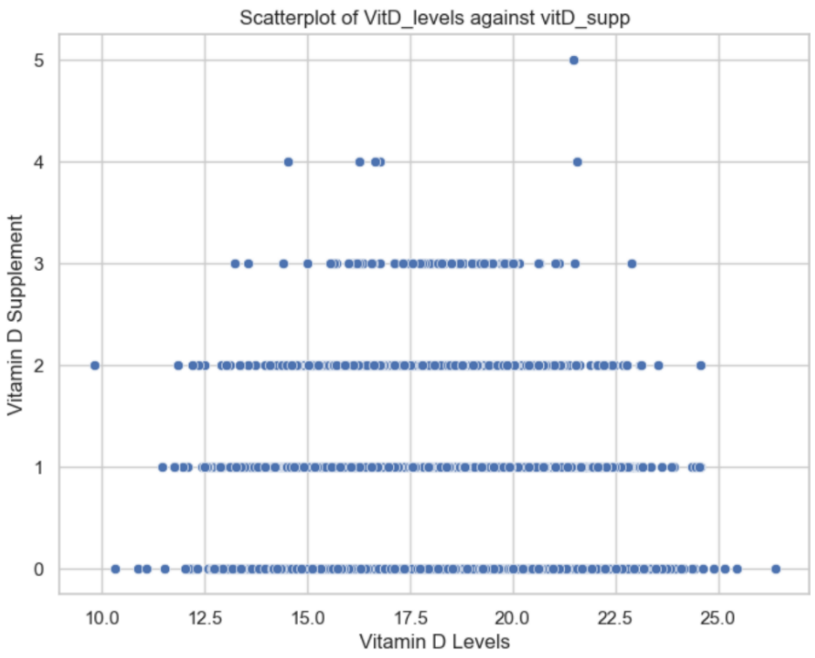
D1. Visual

```
Correlation coefficient between TotalCharge and Additional_charges: 0.02925582402378014
```



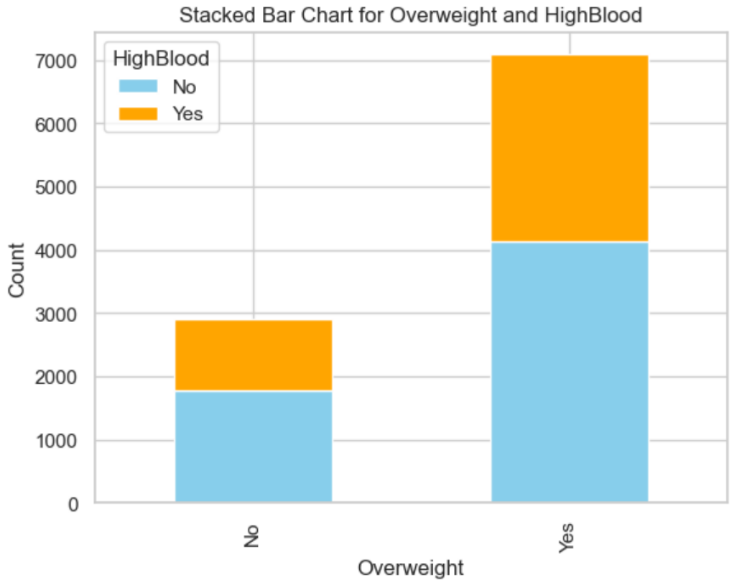
The scatter plot shows the relationship between total charge and additional charges. The correlation coefficient is approximately 0.029, which indicates a positive but negligible linear relationship between the two variables.

```
Correlation coefficient between VitD_levels and vitD_supp: -0.007203220113302815
```



The scatter plot shows the relationship between vitamin D levels and vitamin D supplements. The correlation coefficient is approximately -0.0072, which indicates a negative and negligible linear relationship between the two variables.

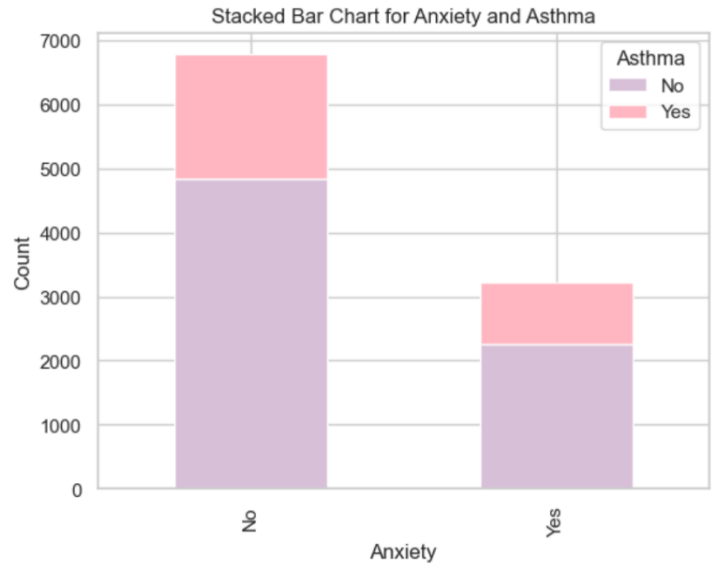
```
Chi-square test for Overweight and HighBlood:
Contingency Table:
HighBlood     No   Yes
Overweight
No          1776  1130
Yes         4134  2960
Chi-square statistic: 6.763425556265908
P-value: 0.009304497772567753
Expected frequencies:
[[1717.446 1188.554]
 [4192.554 2901.446]]
There is a significant association between the variables.
```



The stacked bar chart shows the relationship between overweight and high blood. The chart shows that patients who are not overweight have a higher count of people without high blood pressure, while among those who are overweight, the counts of patients with and without high blood pressure are closer.

The p-value is 0.009. Compared to the alpha, which is 0.05, the p-value is smaller, indicating a significant association between being overweight and having high blood pressure.

```
Chi-square test for Anxiety and Asthma:
Contingency Table:
Asthma     No   Yes
Anxiety
No       4847  1938
Yes      2260   955
Chi-square statistic: 1.3274918794894046
P-value: 0.24925190558821012
Expected frequencies:
[[4822.0995 1962.9005]
 [2284.9005  930.0995]]
There is NO significant association between the variables.
```



The stacked bar chart shows the relationship between anxiety and asthma. The chart shows that patients who have anxiety have a slightly higher count of people with asthma than those who do not have anxiety, but the difference is not too significant.

The p-value is 0.249. Compared to the alpha, which is 0.05, the p-value is larger, indicating no significant association between anxiety and asthma.