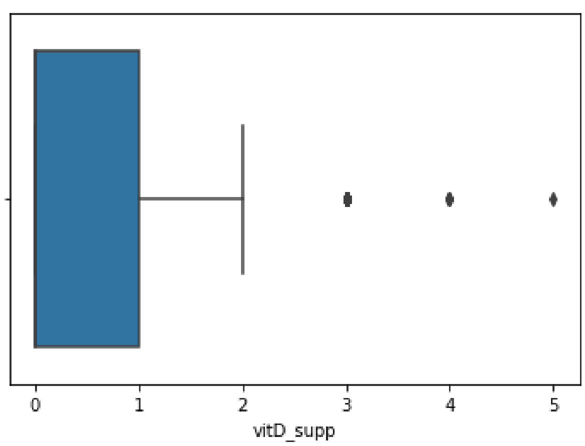
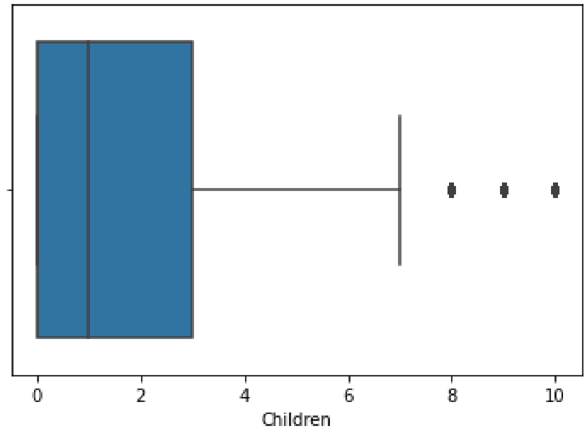
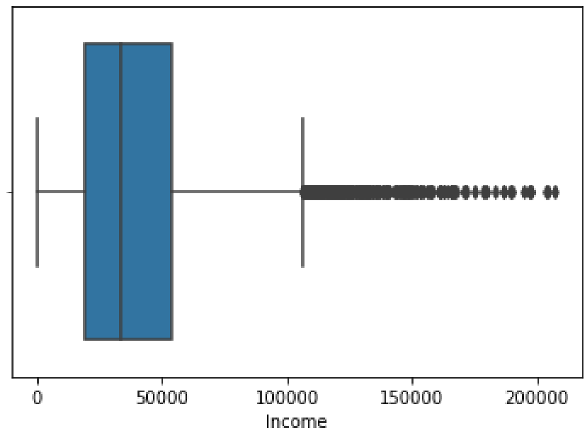
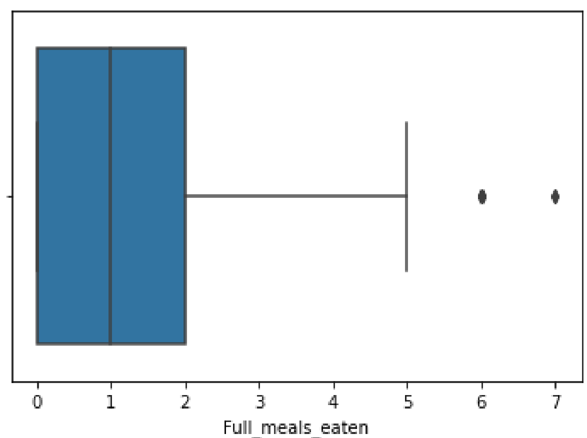
In [2]:
```
import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

import scipy.stats as stats

from scipy.stats import skew, kurtosis

import seaborn as sns

import statistics as stat

from statsmodels.formula.api import logit

#Loading the CSV of the default dataset
df = pd.read_csv(r'C:\Users\mmorg\Desktop\D208 Assessment Files\medical_clean.csv')
```

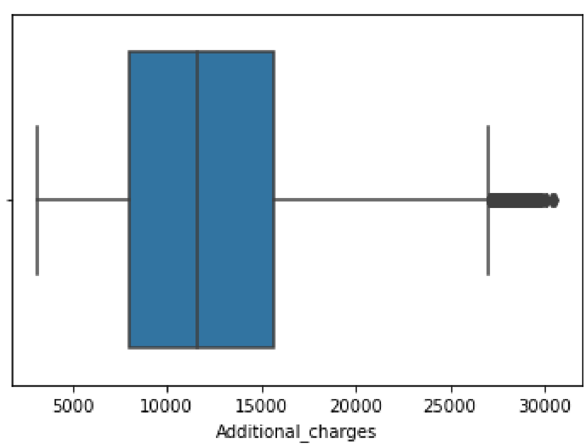
In [3]:
```
#Detection of outliers
boxplot=sns.boxplot(x='vitD_supp',data=df)
plt.show()
boxplot=sns.boxplot(x='Children',data=df)
plt.show()
boxplot=sns.boxplot(x='Income',data=df)
plt.show()
boxplot=sns.boxplot(x='Full_meals_eaten',data=df)
plt.show()
boxplot=sns.boxplot(x='Additional_charges',data=df)
plt.show()
boxplot=sns.boxplot(x='Initial_days',data=df)
plt.show()
boxplot=sns.boxplot(x='TotalCharge',data=df)
plt.show()
boxplot=sns.boxplot(x='VitD_levels',data=df)
plt.show()
boxplot=sns.boxplot(x='Age',data=df)
plt.show()
boxplot=sns.boxplot(x='Doc_visits',data=df)
plt.show()
```

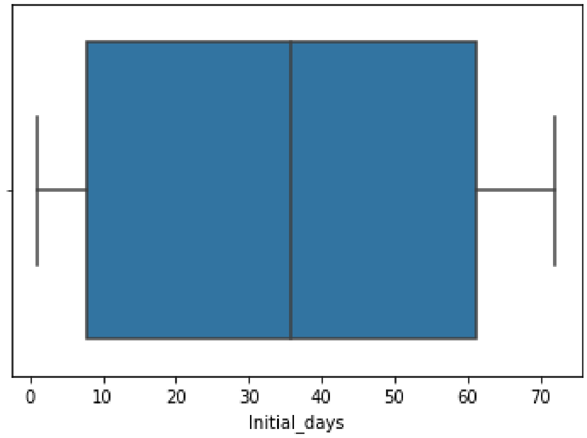


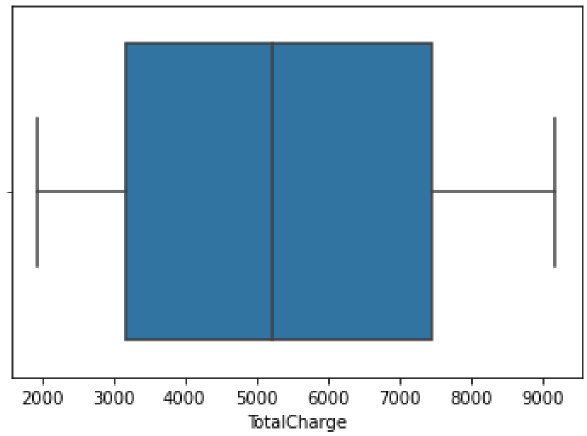

0
2
4
6
8
10
Children

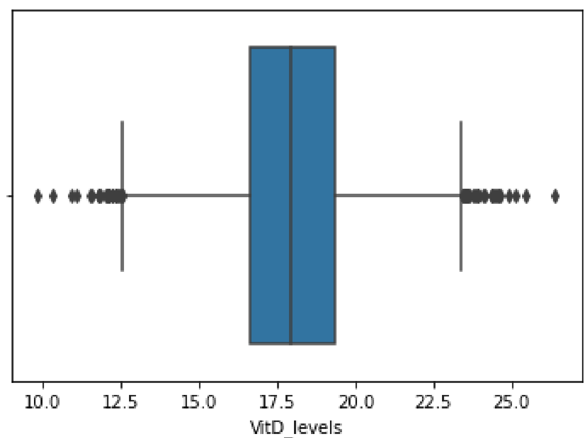

0
50000
100000
150000
200000
Income

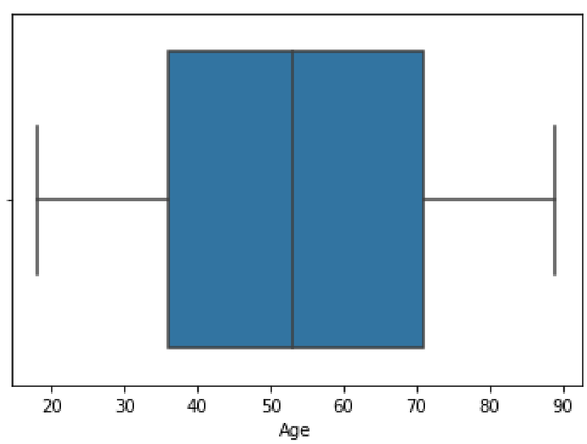

0
1
2
3
4
5
6
7
Full_meals_eaten

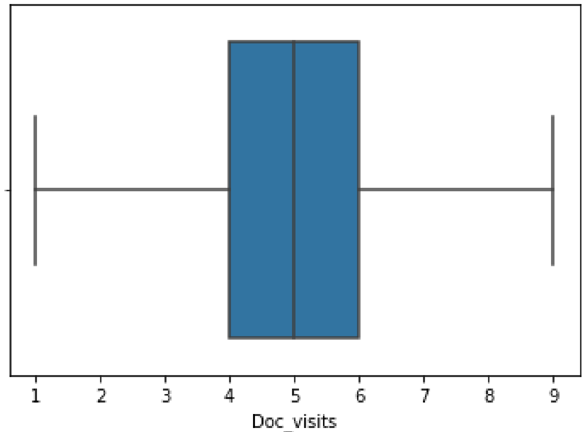

5000
10000
15000
20000
25000
30000
Additional_charges


0
10
20
30
40
50
60
70
Initial_days


2000
3000
4000
5000
6000
7000
8000
9000
TotalCharge


10.0
12.5
15.0
17.5
20.0
22.5
25.0
VitD_levels


20
30
40
50
60
70
80
90
Age



In [4]:
```
#Data Wrangling; turn categorical values into quantitative data
df['ReAdmis_numeric'] = df['ReAdmis']
dict_ReAdmis = {"ReAdmis_numeric": {"No": 0, "Yes": 1}}
df.replace(dict_ReAdmis, inplace=True)

df['Soft_drink_numeric'] = df['Soft_drink']
dict_Soft_drink = {"Soft_drink_numeric": {"No": 0, "Yes": 1}}
df.replace(dict_Soft_drink, inplace=True)

df['HighBlood_numeric'] = df['HighBlood']
dict_HighBlood = {"HighBlood_numeric": {"No": 0, "Yes": 1}}
df.replace(dict_HighBlood, inplace=True)

df['Stroke_numeric'] = df['Stroke']
dict_stroke = {"Stroke_numeric": {"No": 0, "Yes": 1}}
df.replace(dict_stroke, inplace=True)

df['Arthritis_numeric'] = df['Arthritis']
dict_arthritis = {"Arthritis_numeric": {"No": 0, "Yes": 1}}
df.replace(dict_arthritis, inplace=True)

df['Diabetes_numeric'] = df['Diabetes']
dict_diabetes = {"Diabetes_numeric": {"No": 0, "Yes": 1}}
df.replace(dict_diabetes, inplace=True)

df['Hyperlipidemia_numeric'] = df['Hyperlipidemia']
dict_hyperlipidemia = {"Hyperlipidemia_numeric": {"No": 0, "Yes": 1}}
df.replace(dict_hyperlipidemia, inplace=True)

df['BackPain_numeric'] = df['BackPain']
dict_backpain = {"BackPain_numeric": {"No": 0, "Yes": 1}}
df.replace(dict_backpain, inplace=True)

df['Allergic_rhinitis_numeric'] = df['Allergic_rhinitis']
dict_allergies = {"Allergic_rhinitis_numeric": {"No": 0, "Yes": 1}}
df.replace(dict_allergies, inplace=True)

df['Reflux_esophagitis_numeric'] = df['Reflux_esophagitis']
dict_reflux = {"Reflux_esophagitis_numeric": {"No": 0, "Yes": 1}}
df.replace(dict_reflux, inplace=True)

df['Asthma_numeric'] = df['Asthma']
dict_asthma = {"Asthma_numeric": {"No": 0, "Yes": 1}}
df.replace(dict_asthma, inplace=True)

df['Overweight_numeric'] = df['Overweight']
dict_Overweight = {"Overweight_numeric": {"No": 0, "Yes": 1}}
df.replace(dict_Overweight, inplace=True)

df['Anxiety_numeric'] = df['Anxiety']
dict_Anxiety = {"Anxiety_numeric": {"No": 0, "Yes": 1}}
df.replace(dict_Anxiety, inplace=True)
```

```
df = pd.get_dummies(df, columns=["Marital", "Services", "Gender", "Initial_admin", "Complication_risk"])

df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10000 entries, 0 to 9999
Data columns (total 76 columns):
 #   Column                      Non-Null Count  Dtype  
---  ------                      --------------  -----  
 0   CaseOrder                   10000 non-null  int64  
 1   Customer_id                 10000 non-null  object 
 2   Interaction                 10000 non-null  object 
 3   UID                         10000 non-null  object 
 4   City                        10000 non-null  object 
 5   State                       10000 non-null  object 
 6   County                      10000 non-null  object 
 7   Zip                         10000 non-null  int64  
 8   Lat                         10000 non-null  float64
 9   Lng                         10000 non-null  float64
 10  Population                  10000 non-null  int64  
 11  Area                        10000 non-null  object 
 12  TimeZone                    10000 non-null  object 
 13  Job                         10000 non-null  object 
 14  Children                    10000 non-null  int64  
 15  Age                         10000 non-null  int64  
 16  Income                      10000 non-null  float64
 17  ReAdmis                     10000 non-null  object 
 18  VitD_levels                 10000 non-null  float64
 19  Doc_visits                  10000 non-null  int64  
 20  Full_meals_eaten            10000 non-null  int64  
 21  vitD_supp                   10000 non-null  int64  
 22  Soft_drink                  10000 non-null  object 
 23  HighBlood                   10000 non-null  object 
 24  Stroke                      10000 non-null  object 
 25  Overweight                  10000 non-null  object 
 26  Arthritis                   10000 non-null  object 
 27  Diabetes                    10000 non-null  object 
 28  Hyperlipidemia              10000 non-null  object 
 29  BackPain                    10000 non-null  object 
 30  Anxiety                     10000 non-null  object 
 31  Allergic_rhinitis           10000 non-null  object 
 32  Reflux_esophagitis          10000 non-null  object 
 33  Asthma                      10000 non-null  object 
 34  Initial_days                10000 non-null  float64
 35  TotalCharge                 10000 non-null  float64
 36  Additional_charges          10000 non-null  float64
 37  Item1                       10000 non-null  int64  
 38  Item2                       10000 non-null  int64  
 39  Item3                       10000 non-null  int64  
 40  Item4                       10000 non-null  int64  
 41  Item5                       10000 non-null  int64  
 42  Item6                       10000 non-null  int64  
 43  Item7                       10000 non-null  int64  
 44  Item8                       10000 non-null  int64  
 45  ReAdmis_numeric             10000 non-null  int64  
 46  Soft_drink_numeric          10000 non-null  int64  
 47  HighBlood_numeric           10000 non-null  int64  
 48  Stroke_numeric              10000 non-null  int64  
 49  Arthritis_numeric           10000 non-null  int64  
 50  Diabetes_numeric            10000 non-null  int64  
 51  Hyperlipidemia_numeric      10000 non-null  int64  
 52  BackPain_numeric            10000 non-null  int64  
 53  Allergic_rhinitis_numeric   10000 non-null  int64  
 54  Reflux_esophagitis_numeric  10000 non-null  int64  
 55  Asthma_numeric              10000 non-null  int64  
 56  Overweight_numeric          10000 non-null  int64  
 57  Anxiety_numeric             10000 non-null  int64  
 58  Marital_Divorced            10000 non-null  uint8  
 59  Marital_Married             10000 non-null  uint8  
 60  Marital_Never Married       10000 non-null  uint8  
 61  Marital_Separated           10000 non-null  uint8  
 62  Marital_Widowed             10000 non-null  uint8  
 63  Services_Blood Work         10000 non-null  uint8  
 64  Services_CT Scan            10000 non-null  uint8  
 65  Services_Intravenous        10000 non-null  uint8  
```

```
 66  Services_MRI                         10000 non-null  uint8
 67  Gender_Female                        10000 non-null  uint8
 68  Gender_Male                          10000 non-null  uint8
 69  Gender_Nonbinary                     10000 non-null  uint8
 70  Initial_admin_Elective Admission     10000 non-null  uint8
 71  Initial_admin_Emergency Admission    10000 non-null  uint8
 72  Initial_admin_Observation Admission  10000 non-null  uint8
 73  Complication_risk_High               10000 non-null  uint8
 74  Complication_risk_Low                10000 non-null  uint8
 75  Complication_risk_Medium             10000 non-null  uint8
dtypes: float64(7), int64(29), object(22), uint8(18)
memory usage: 4.6+ MB
```

In [5]:
```
#Renaming columns from pd.get_dummies
df = df.rename({'Initial_admin_Elective Admission': 'Initial_admin_Elective_Admission',
                'Initial_admin_Emergency Admission': 'Initial_admin_Emergency_Admission',
                'Initial_admin_Observation Admission': 'Initial_admin_Observation_Admission',
                'Marital_Never Married': 'Marital_Never_Married',
                'Services_Blood Work': 'Services_Blood_Work',
                'Services_CT Scan': 'Services_CT_Scan'}, axis ='columns')
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10000 entries, 0 to 9999
Data columns (total 76 columns):
 #   Column                      Non-Null Count  Dtype  
---  ------                      --------------  -----  
 0   CaseOrder                   10000 non-null  int64  
 1   Customer_id                 10000 non-null  object 
 2   Interaction                 10000 non-null  object 
 3   UID                         10000 non-null  object 
 4   City                        10000 non-null  object 
 5   State                       10000 non-null  object 
 6   County                      10000 non-null  object 
 7   Zip                         10000 non-null  int64  
 8   Lat                         10000 non-null  float64
 9   Lng                         10000 non-null  float64
 10  Population                  10000 non-null  int64  
 11  Area                        10000 non-null  object 
 12  TimeZone                    10000 non-null  object 
 13  Job                         10000 non-null  object 
 14  Children                    10000 non-null  int64  
 15  Age                         10000 non-null  int64  
 16  Income                      10000 non-null  float64
 17  ReAdmis                     10000 non-null  object 
 18  VitD_levels                 10000 non-null  float64
 19  Doc_visits                  10000 non-null  int64  
 20  Full_meals_eaten            10000 non-null  int64  
 21  vitD_supp                   10000 non-null  int64  
 22  Soft_drink                  10000 non-null  object 
 23  HighBlood                   10000 non-null  object 
 24  Stroke                      10000 non-null  object 
 25  Overweight                  10000 non-null  object 
 26  Arthritis                   10000 non-null  object 
 27  Diabetes                    10000 non-null  object 
 28  Hyperlipidemia              10000 non-null  object 
 29  BackPain                    10000 non-null  object 
 30  Anxiety                     10000 non-null  object 
 31  Allergic_rhinitis           10000 non-null  object 
 32  Reflux_esophagitis          10000 non-null  object 
 33  Asthma                      10000 non-null  object 
 34  Initial_days                10000 non-null  float64
 35  TotalCharge                 10000 non-null  float64
 36  Additional_charges          10000 non-null  float64
 37  Item1                       10000 non-null  int64  
 38  Item2                       10000 non-null  int64  
 39  Item3                       10000 non-null  int64  
 40  Item4                       10000 non-null  int64  
 41  Item5                       10000 non-null  int64  
 42  Item6                       10000 non-null  int64  
 43  Item7                       10000 non-null  int64  
 44  Item8                       10000 non-null  int64  
 45  ReAdmis_numeric             10000 non-null  int64  
 46  Soft_drink_numeric          10000 non-null  int64  
 47  HighBlood_numeric           10000 non-null  int64  
 48  Stroke_numeric              10000 non-null  int64  
 49  Arthritis_numeric           10000 non-null  int64  
 50  Diabetes_numeric            10000 non-null  int64  
 51  Hyperlipidemia_numeric      10000 non-null  int64  
 52  BackPain_numeric            10000 non-null  int64  
 53  Allergic_rhinitis_numeric   10000 non-null  int64  
 54  Reflux_esophagitis_numeric  10000 non-null  int64  
 55  Asthma_numeric              10000 non-null  int64  
 56  Overweight_numeric          10000 non-null  int64  
 57  Anxiety_numeric             10000 non-null  int64  
 58  Marital_Divorced            10000 non-null  uint8  
 59  Marital_Married             10000 non-null  uint8  
 60  Marital_Never_Married       10000 non-null  uint8  
 61  Marital_Separated           10000 non-null  uint8  
 62  Marital_Widowed             10000 non-null  uint8  
 63  Services_Blood_Work         10000 non-null  uint8  
 64  Services_CT_Scan            10000 non-null  uint8  
 65  Services_Intravenous        10000 non-null  uint8  
```

```
 66  Services_MRI                          10000 non-null  uint8
 67  Gender_Female                         10000 non-null  uint8
 68  Gender_Male                           10000 non-null  uint8
 69  Gender_Nonbinary                      10000 non-null  uint8
 70  Initial_admin_Elective_Admission      10000 non-null  uint8
 71  Initial_admin_Emergency_Admission     10000 non-null  uint8
 72  Initial_admin_Observation_Admission   10000 non-null  uint8
 73  Complication_risk_High                10000 non-null  uint8
 74  Complication_risk_Low                 10000 non-null  uint8
 75  Complication_risk_Medium              10000 non-null  uint8
dtypes: float64(7), int64(29), object(22), uint8(18)
memory usage: 4.6+ MB
```

In [6]:
```
##Univariate Stats Dataframe, and dropping unneeded and redundant columns
def unistats(df):
    output_df = pd.DataFrame(columns=['Count', 'Missing', 'Unique', 'Dtype', 'Numeric', 'Mean', 'Mode',

    for col in df:
        if pd.api.types.is_numeric_dtype(df[col]):
            output_df.loc[col] = [df[col].count(), df[col].isnull().sum(), df[col].nunique(), df[col].dt
        else:
            output_df.loc[col] = [df[col].count(), df[col].isnull().sum(), df[col].nunique(), df[col].dt
    return output_df.sort_values(by=['Numeric', 'Skew', 'Unique'], ascending=False)

df.drop(columns=['CaseOrder', 'Customer_id', 'Interaction', 'UID', 'State', 'County', 'Job', 'Zip', 'Tim
print(unistats(df))
```

```
                                   Count  Missing  Unique    Dtype  Numeric  \
Gender_Nonbinary                   10000        0       2    uint8     True
Services_MRI                       10000        0       2    uint8     True
Services_CT_Scan                   10000        0       2    uint8     True
Population                         10000        0    5951    int64     True
vitD_supp                          10000        0       6    int64     True
Marital_Never_Married              10000        0       2    uint8     True
Marital_Separated                  10000        0       2    uint8     True
Stroke_numeric                     10000        0       2    int64     True
Marital_Married                    10000        0       2    uint8     True
Marital_Widowed                    10000        0       2    uint8     True
Children                           10000        0      11    int64     True
Income                             10000        0    9993  float64     True
Complication_risk_Low              10000        0       2    uint8     True
Initial_admin_Observation_Admission 10000       0       2    uint8     True
Initial_admin_Elective_Admission   10000        0       2    uint8     True
Soft_drink_numeric                 10000        0       2    int64     True
Diabetes_numeric                   10000        0       2    int64     True
Full_meals_eaten                   10000        0       8    int64     True
Asthma_numeric                     10000        0       2    int64     True
Additional_charges                 10000        0    9418  float64     True
Services_Intravenous               10000        0       2    uint8     True
Anxiety_numeric                    10000        0       2    int64     True
Complication_risk_High             10000        0       2    uint8     True
Hyperlipidemia_numeric             10000        0       2    int64     True
Arthritis_numeric                  10000        0       2    int64     True
ReAdmis_numeric                    10000        0       2    int64     True
Allergic_rhinitis_numeric          10000        0       2    int64     True
HighBlood_numeric                  10000        0       2    int64     True
BackPain_numeric                   10000        0       2    int64     True
Reflux_esophagitis_numeric         10000        0       2    int64     True
Complication_risk_Medium           10000        0       2    uint8     True
Gender_Male                        10000        0       2    uint8     True
Initial_days                       10000        0    9997  float64     True
TotalCharge                        10000        0    9997  float64     True
VitD_levels                        10000        0    9976  float64     True
Age                                10000        0      72    int64     True
Doc_visits                         10000        0       9    int64     True
Initial_admin_Emergency_Admission  10000        0       2    uint8     True
Services_Blood_Work                10000        0       2    uint8     True
Overweight_numeric                 10000        0       2    int64     True

                                           Mean         Mode          Min  \
Gender_Nonbinary                       0.021400      0.00000     0.000000
Services_MRI                           0.038000      0.00000     0.000000
Services_CT_Scan                       0.122500      0.00000     0.000000
Population                          9965.253800      0.00000     0.000000
vitD_supp                              0.398900      0.00000     0.000000
Marital_Never_Married                  0.198400      0.00000     0.000000
Marital_Separated                      0.198700      0.00000     0.000000
Stroke_numeric                         0.199300      0.00000     0.000000
Marital_Married                        0.202300      0.00000     0.000000
Marital_Widowed                        0.204500      0.00000     0.000000
Children                               2.097200      0.00000     0.000000
Income                             40490.495160  14572.40000   154.080000
Complication_risk_Low                  0.212500      0.00000     0.000000
Initial_admin_Observation_Admission    0.243600      0.00000     0.000000
Initial_admin_Elective_Admission       0.250400      0.00000     0.000000
Soft_drink_numeric                     0.257500      0.00000     0.000000
Diabetes_numeric                       0.273800      0.00000     0.000000
Full_meals_eaten                       1.001400      0.00000     0.000000
Asthma_numeric                         0.289300      0.00000     0.000000
Additional_charges                 12934.528587   3883.66416  3125.703000
Services_Intravenous                   0.313000      0.00000     0.000000
Anxiety_numeric                        0.321500      0.00000     0.000000
Complication_risk_High                 0.335800      0.00000     0.000000
Hyperlipidemia_numeric                 0.337200      0.00000     0.000000
Arthritis_numeric                      0.357400      0.00000     0.000000
ReAdmis_numeric                        0.366900      0.00000     0.000000
Allergic_rhinitis_numeric              0.394100      0.00000     0.000000
HighBlood_numeric                      0.409000      0.00000     0.000000
```

```
BackPain_numeric                        0.411400     0.00000     0.000000
Reflux_esophagitis_numeric              0.413500     0.00000     0.000000
Complication_risk_Medium                0.451700     0.00000     0.000000
Gender_Male                             0.476800     0.00000     0.000000
Initial_days                           34.455299    63.54432     1.001981
TotalCharge                          5312.172769  7555.45200  1938.312067
VitD_levels                            17.964262    15.26009     9.806483
Age                                    53.511700    47.00000    18.000000
Doc_visits                              5.012200     5.00000     1.000000
Initial_admin_Emergency_Admission       0.506000     1.00000     0.000000
Services_Blood_Work                     0.526500     1.00000     0.000000
Overweight_numeric                      0.709400     1.00000     0.000000

                                          Median           Max  \
Gender_Nonbinary                        0.000000      1.000000
Services_MRI                            0.000000      1.000000
Services_CT_Scan                        0.000000      1.000000
Population                           2769.000000  122814.000000
vitD_supp                               0.000000      5.000000
Marital_Never_Married                   0.000000      1.000000
Marital_Separated                       0.000000      1.000000
Stroke_numeric                          0.000000      1.000000
Marital_Married                         0.000000      1.000000
Marital_Widowed                         0.000000      1.000000
Children                                1.000000     10.000000
Income                              33768.420000  207249.100000
Complication_risk_Low                   0.000000      1.000000
Initial_admin_Observation_Admission     0.000000      1.000000
Initial_admin_Elective_Admission        0.000000      1.000000
Soft_drink_numeric                      0.000000      1.000000
Diabetes_numeric                        0.000000      1.000000
Full_meals_eaten                        1.000000      7.000000
Asthma_numeric                          0.000000      1.000000
Additional_charges                  11573.977735   30566.070000
Services_Intravenous                    0.000000      1.000000
Anxiety_numeric                         0.000000      1.000000
Complication_risk_High                  0.000000      1.000000
Hyperlipidemia_numeric                  0.000000      1.000000
Arthritis_numeric                       0.000000      1.000000
ReAdmis_numeric                         0.000000      1.000000
Allergic_rhinitis_numeric               0.000000      1.000000
HighBlood_numeric                       0.000000      1.000000
BackPain_numeric                        0.000000      1.000000
Reflux_esophagitis_numeric              0.000000      1.000000
Complication_risk_Medium                0.000000      1.000000
Gender_Male                             0.000000      1.000000
Initial_days                           35.836244     71.981490
TotalCharge                          5213.952000   9180.728000
VitD_levels                            17.951122     26.394449
Age                                    53.000000     89.000000
Doc_visits                              5.000000      9.000000
Initial_admin_Emergency_Admission       1.000000      1.000000
Services_Blood_Work                     1.000000      1.000000
Overweight_numeric                      1.000000      1.000000

                                             Std      Skew       Kurt
Gender_Nonbinary                        0.144721  6.615434  41.772323
Services_MRI                            0.191206  4.833456  21.366572
Services_CT_Scan                        0.327879  2.303141   3.305119
Population                          14824.758614  2.229959   5.880913
vitD_supp                               0.628505  1.550205   2.330763
Marital_Never_Married                   0.398815  1.512784   0.288572
Marital_Separated                       0.399042  1.510420   0.281425
Stroke_numeric                          0.399494  1.505705   0.267202
Marital_Married                         0.401735  1.482369   0.197456
Marital_Widowed                         0.403356  1.465500   0.147720
Children                                2.163659  1.448013   2.076321
Income                              28521.153293  1.405899   2.745690
Complication_risk_Low                   0.409097  1.405815  -0.023688
Initial_admin_Observation_Admission     0.429276  1.194810  -0.572544
Initial_admin_Elective_Admission        0.433265  1.152412  -0.672081
```

```
Soft_drink_numeric                    0.437279  1.109354  -0.769488
Diabetes_numeric                      0.445930  1.014712  -0.970553
Full_meals_eaten                      1.008117  1.009461   1.042727
Asthma_numeric                        0.453460  0.929485  -1.136285
Additional_charges                 6542.601544  0.831842  -0.142684
Services_Intravenous                  0.463738  0.806652  -1.349583
Anxiety_numeric                       0.467076  0.764483  -1.415849
Complication_risk_High                0.472293  0.695470  -1.516625
Hyperlipidemia_numeric                0.472777  0.688834  -1.525813
Arthritis_numeric                     0.479258  0.595206  -1.646059
ReAdmis_numeric                       0.481983  0.552412  -1.695180
Allergic_rhinitis_numeric             0.488681  0.433498  -1.812442
HighBlood_numeric                     0.491674  0.370238  -1.863296
BackPain_numeric                      0.492112  0.360153  -1.870664
Reflux_esophagitis_numeric            0.492486  0.351350  -1.876929
Complication_risk_Medium              0.497687  0.194137  -1.962703
Gender_Male                           0.499486  0.092914  -1.991765
Initial_days                         26.309341  0.070286  -1.754525
TotalCharge                        2180.393838  0.069661  -1.668267
VitD_levels                           2.017231  0.032435  -0.022112
Age                                  20.638538  0.005117  -1.189527
Doc_visits                            1.045734 -0.018563   0.025999
Initial_admin_Emergency_Admission     0.499989 -0.024005  -1.999824
Services_Blood_Work                   0.499322 -0.106165  -1.989127
Overweight_numeric                    0.454062 -0.922526  -1.149176
```

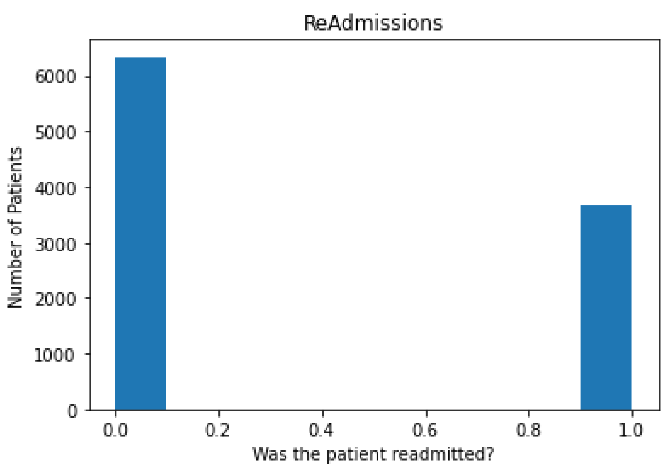
In [8]:

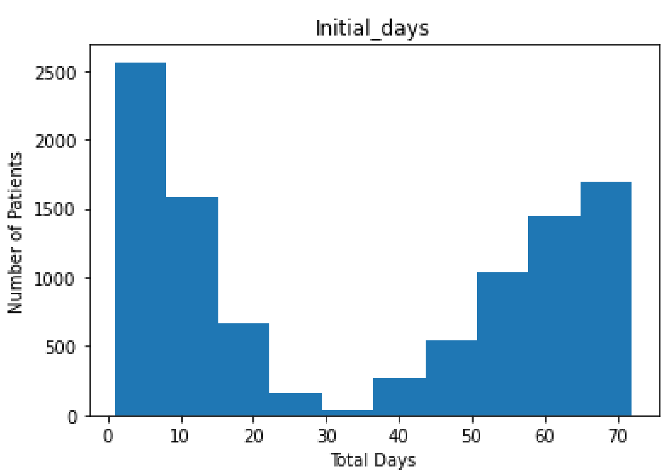
```
#Univariate Visualizations
plt.hist(df.ReAdmis_numeric)
plt.xlabel('Was the patient readmitted?')
plt.ylabel('Number of Patients')
plt.title('ReAdmissions')
plt.show()

plt.hist(df.Initial_days)
plt.xlabel('Total Days')
plt.ylabel('Number of Patients')
plt.title('Initial_days')
plt.show()

plt.hist(df.vitD_supp)
plt.xlabel('# of Vit D Administered')
plt.ylabel('Number of Patients')
plt.title('vitd_supp')
plt.show()

plt.hist(df.Children)
plt.xlabel('# of Children')
plt.ylabel('Number of Patients')
plt.title('# of children')
plt.show()

plt.hist(df.Income)
plt.xlabel('Yearly Income')
plt.ylabel('Number of Patients')
plt.title('Yearly Income')
plt.show()

plt.hist(df.Full_meals_eaten)
plt.xlabel('Full_meals_eaten')
plt.ylabel('Number of Patients')
plt.title('Full Meals Eaten')
plt.show()

plt.hist(df.Additional_charges)
plt.xlabel('Additional Charges')
plt.ylabel('Number of Patients')
plt.title('Additional Charges')
plt.show()

plt.hist(df.TotalCharge)
plt.xlabel('Total Charges')
```

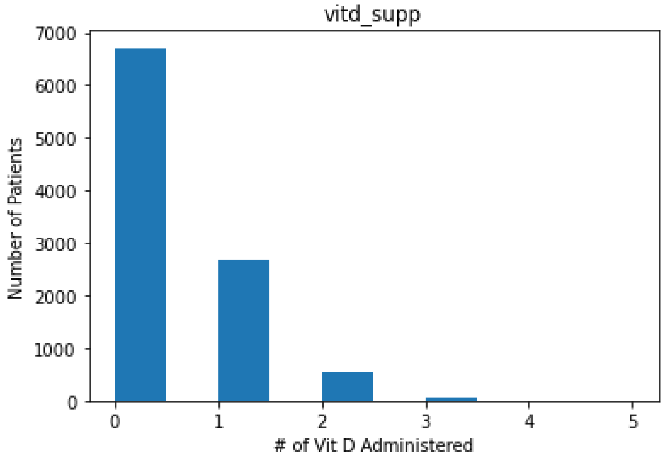
```
plt.ylabel('Number of Patients')
plt.title('Total Charges')
plt.show()

plt.hist(df.VitD_levels)
plt.xlabel('VitD Levels')
plt.ylabel('Number of Patients')
plt.title('VitD Levels')
plt.show()

plt.hist(df.Age)
plt.xlabel('Age')
plt.ylabel('Number of Patients')
plt.title('Age')
plt.show()

plt.hist(df.Doc_visits)
plt.xlabel('Doctor Visits')
plt.ylabel('Number of Patients')
plt.title('Doctor Visits')
plt.show()

plt.hist(df.HighBlood_numeric)
plt.xlabel('Does patient have high blood pressure?')
plt.ylabel('Number of Patients')
plt.title('High Blood Pressure')
plt.show()

plt.hist(df.Stroke_numeric)
plt.xlabel('Does patient have history of strokes?')
plt.ylabel('Number of Patients')
plt.title('Stroke')
plt.show()

plt.hist(df.Arthritis_numeric)
plt.xlabel('Does patient have history of Arthritis?')
plt.ylabel('Number of Patients')
plt.title('Arthritis')
plt.show()

plt.hist(df.Diabetes_numeric)
plt.xlabel('Does patient have history of Diabetes?')
plt.ylabel('Number of Patients')
plt.title('Diabetes')
plt.show()

plt.hist(df.Hyperlipidemia_numeric)
plt.xlabel('Does patient have Hyperlipidemia?')
plt.ylabel('Number of Patients')
plt.title('Hyperlipidemia')
plt.show()

plt.hist(df.BackPain_numeric)
plt.xlabel('Does patient have BackPain?')
plt.ylabel('Number of Patients')
plt.title('BackPain')
plt.show()

plt.hist(df.Allergic_rhinitis_numeric)
plt.xlabel('Does patient have Allergic_rhinitis?')
plt.ylabel('Number of Patients')
plt.title('Allergic_rhinitis')
plt.show()

plt.hist(df.Reflux_esophagitis_numeric)
plt.xlabel('Does patient have Reflux_esophagitis?')
plt.ylabel('Number of Patients')
plt.title('Reflux_esophagitis')
plt.show()

plt.hist(df.Asthma_numeric)
```

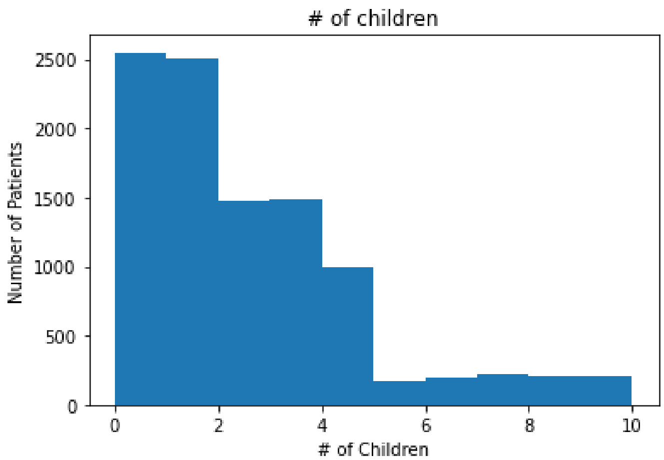
```
plt.xlabel('Does patient have Asthma?')
plt.ylabel('Number of Patients')
plt.title('Asthma')
plt.show()

plt.hist(df.Marital_Married)
plt.xlabel('Patients Marital Status')
plt.ylabel('Number of Patients')
plt.title('Marital_Married')
plt.show()

plt.hist(df.Marital_Never_Married)
plt.xlabel('Patients Marital Status')
plt.ylabel('Number of Patients')
plt.title('Marital_Never_Married')
plt.show()

plt.hist(df.Marital_Separated)
plt.xlabel('Patients Marital Status')
plt.ylabel('Number of Patients')
plt.title('Marital_Separated')
plt.show()

plt.hist(df.Marital_Widowed)
plt.xlabel('Patients Marital Status')
plt.ylabel('Number of Patients')
plt.title('Marital_Widowed')
plt.show()

plt.hist(df.Services_Blood_Work)
plt.xlabel('What services did the patient receive?')
plt.ylabel('Number of Patients')
plt.title('Services_Blood_Work')
plt.show()

plt.hist(df.Services_CT_Scan)
plt.xlabel('What services did the patient receive?')
plt.ylabel('Number of Patients')
plt.title('Services_CT_Scan')
plt.show()

plt.hist(df.Services_Intravenous)
plt.xlabel('What services did the patient receive?')
plt.ylabel('Number of Patients')
plt.title('Services_Intravenous')
plt.show()

plt.hist(df.Services_MRI)
plt.xlabel('What services did the patient receive?')
plt.ylabel('Number of Patients')
plt.title('Services_MRI')
plt.show()

plt.hist(df.Gender_Male)
plt.xlabel('What gender does the patient identify as?')
plt.ylabel('Number of Patients')
plt.title('Gender_Male')
plt.show()

plt.hist(df.Gender_Nonbinary)
plt.xlabel('What gender does the patient identify as?')
plt.ylabel('Number of Patients')
plt.title('Gender_Nonbinary')
plt.show()

plt.hist(df.Initial_admin_Elective_Admission)
plt.xlabel('What brought the patient into the hospital?')
plt.ylabel('Number of Patients')
plt.title('Initial_admin_Elective_Admission')
plt.show()
```

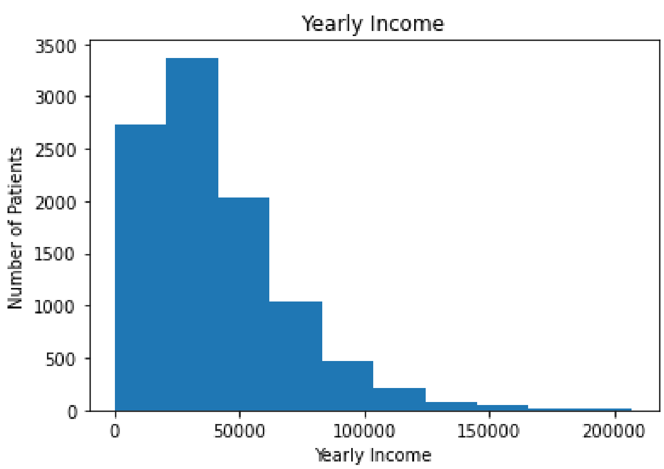
```
plt.hist(df.Initial_admin_Emergency_Admission)
plt.xlabel('What brought the patient into the hospital?')
plt.ylabel('Number of Patients')
plt.title('Initial_admin_Emergency_Admission')
plt.show()

plt.hist(df.Initial_admin_Observation_Admission)
plt.xlabel('What brought the patient into the hospital?')
plt.ylabel('Number of Patients')
plt.title('Initial_admin_Observation_Admission')
plt.show()

plt.hist(df.Complication_risk_High)
plt.xlabel('What is the patients complication risk?')
plt.ylabel('Number of Patients')
plt.title('Complication_risk_High')
plt.show()

plt.hist(df.Complication_risk_Low)
plt.xlabel('What is the patients complication risk?')
plt.ylabel('Number of Patients')
plt.title('Complication_risk_Low')
plt.show()

plt.hist(df.Complication_risk_Medium)
plt.xlabel('What is the patients complication risk?')
plt.ylabel('Number of Patients')
plt.title('Complication_risk_Medium')
plt.show()
```

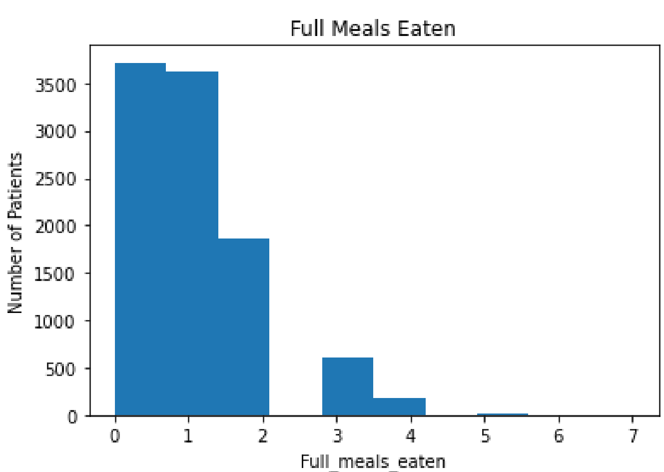


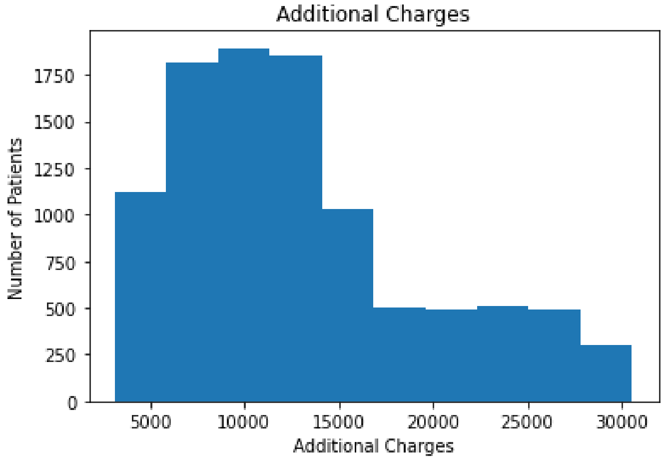


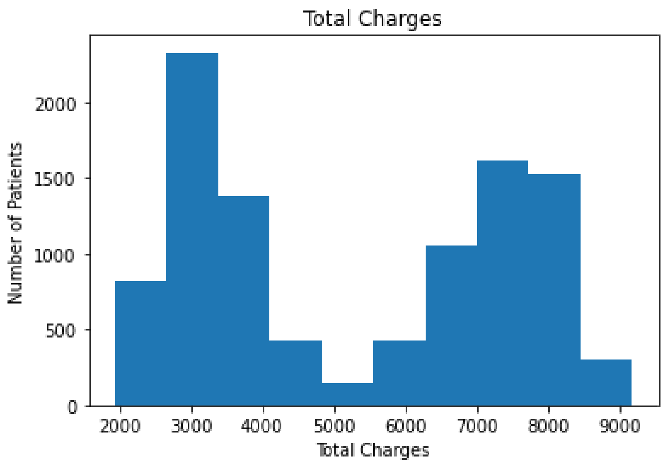

vitd_supp
Number of Patients
0
1000
2000
3000
4000
5000
6000
7000
0
1
2
3
4
5
# of Vit D Administered

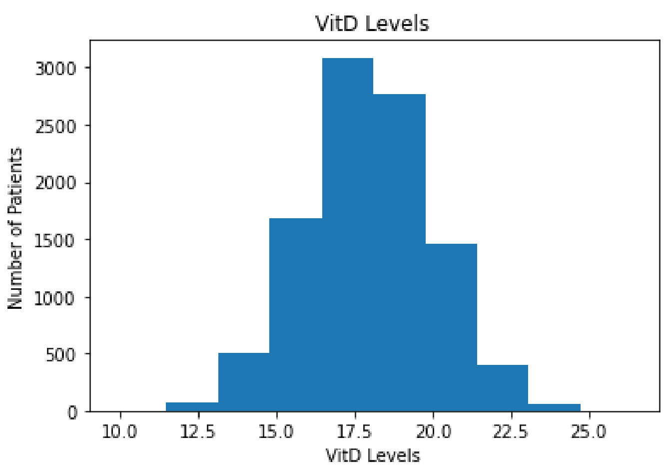

# of children
Number of Patients
0
500
1000
1500
2000
2500
0
2
4
6
8
10
# of Children

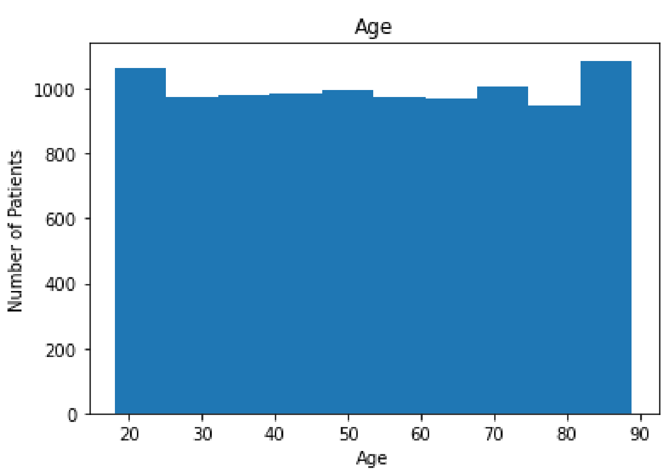

Yearly Income
Number of Patients
0
500
1000
1500
2000
2500
3000
3500
0
50000
100000
150000
200000
Yearly Income

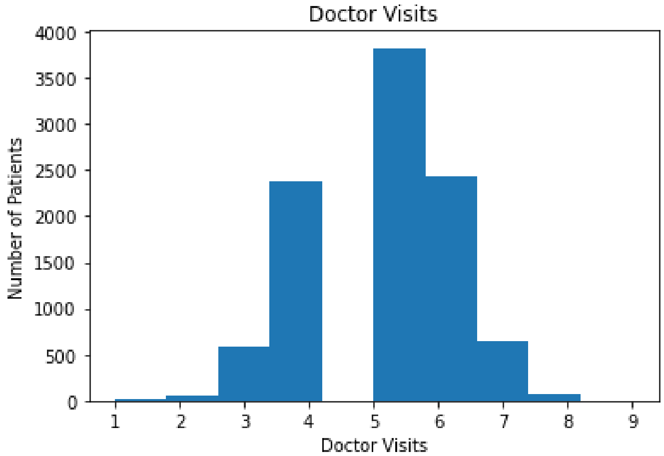

Full Meals Eaten
Number of Patients
0
500
1000
1500
2000
2500
3000
3500
0
1
2
3
4
5
6
7
Full_meals_eaten

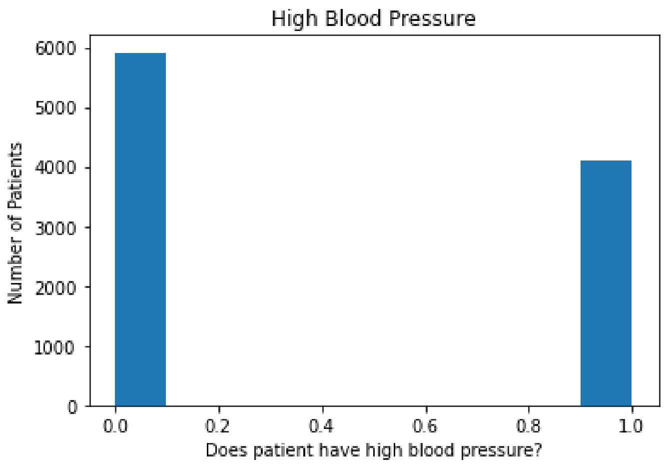

Additional Charges
Number of Patients
0
250
500
750
1000
1250
1500
1750
5000
10000
15000
20000
25000
30000
Additional Charges

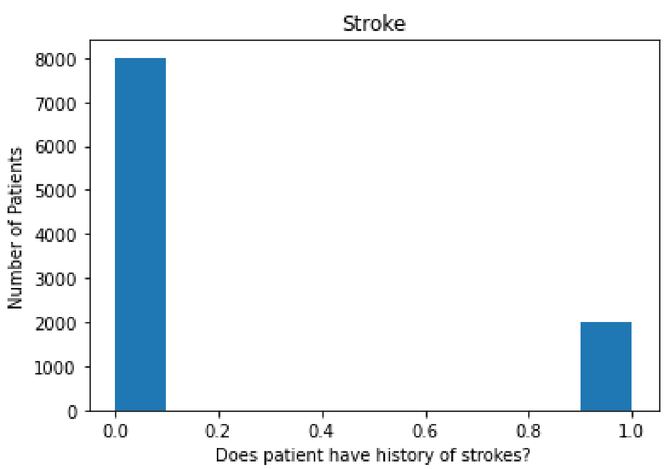

Total Charges
Number of Patients
0
500
1000
1500
2000
2000
3000
4000
5000
6000
7000
8000
9000
Total Charges

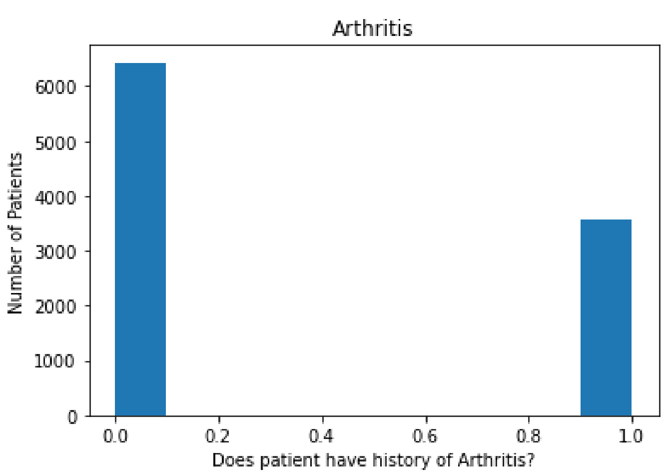

VitD Levels
Number of Patients
0
500
1000
1500
2000
2500
3000
10.0
12.5
15.0
17.5
20.0
22.5
25.0
VitD Levels

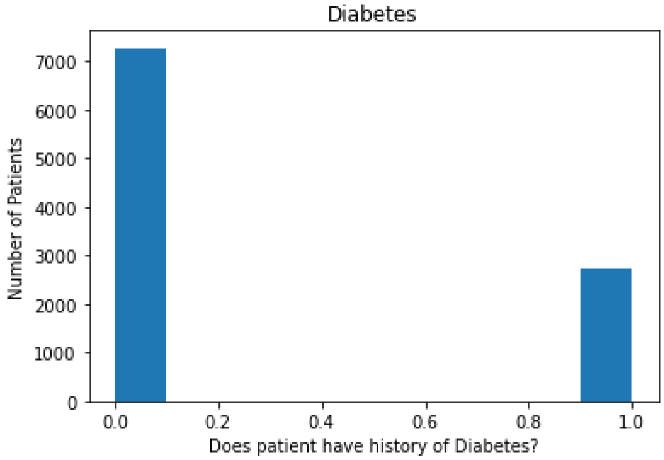

Age
Number of Patients
0
200
400
600
800
1000
20
30
40
50
60
70
80
90
Age

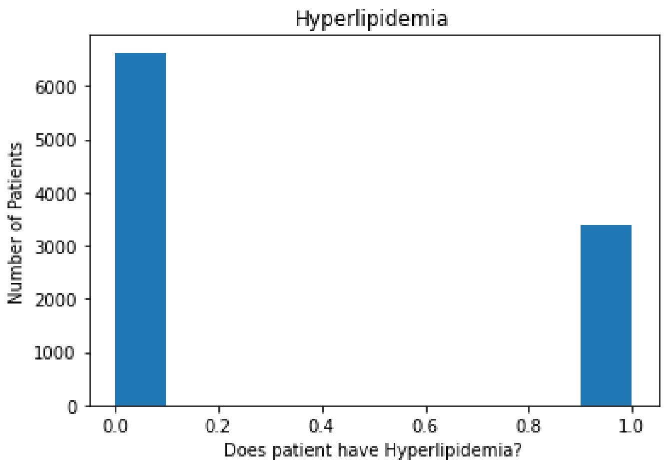

Doctor Visits
Number of Patients
0
500
1000
1500
2000
2500
3000
3500
4000
1
2
3
4
5
6
7
8
9
Doctor Visits

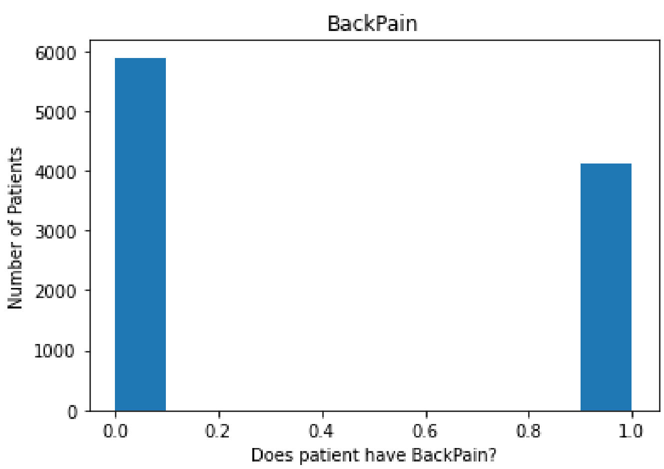

High Blood Pressure
Number of Patients
0
1000
2000
3000
4000
5000
6000
0.0
0.2
0.4
0.6
0.8
1.0
Does patient have high blood pressure?

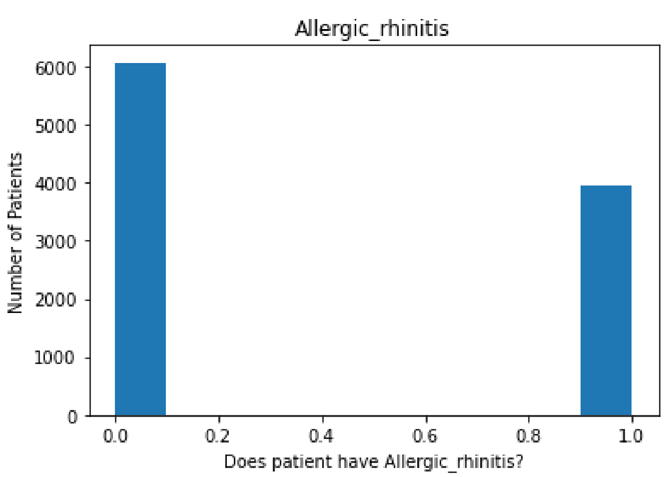

Stroke
Number of Patients
0
1000
2000
3000
4000
5000
6000
7000
8000
0.0
0.2
0.4
0.6
0.8
1.0
Does patient have history of strokes?

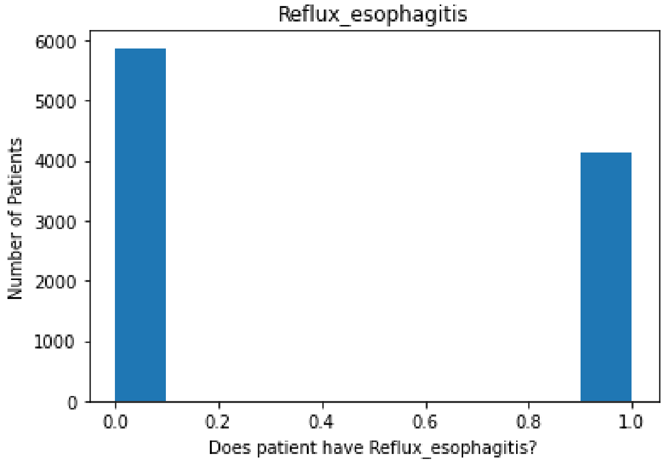

Arthritis
Number of Patients
0
1000
2000
3000
4000
5000
6000
0.0
0.2
0.4
0.6
0.8
1.0
Does patient have history of Arthritis?

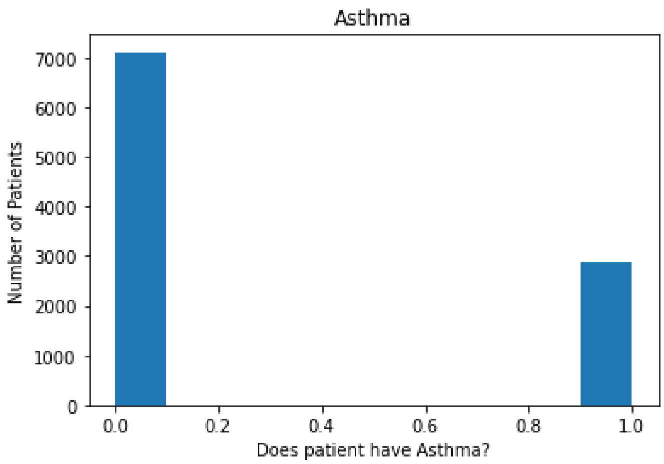

Diabetes
Number of Patients
Does patient have history of Diabetes?

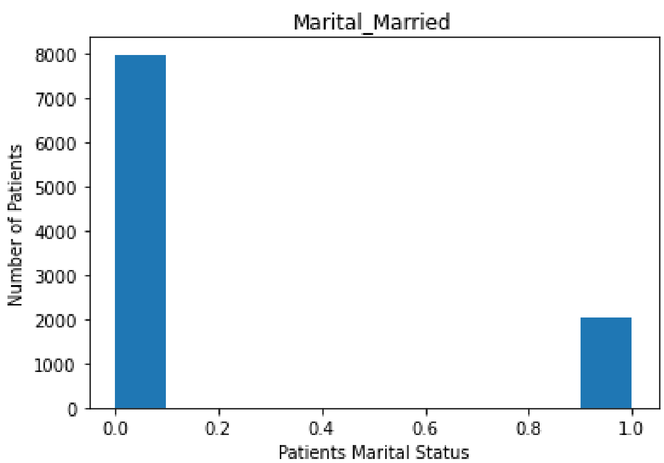

Hyperlipidemia
Number of Patients
Does patient have Hyperlipidemia?

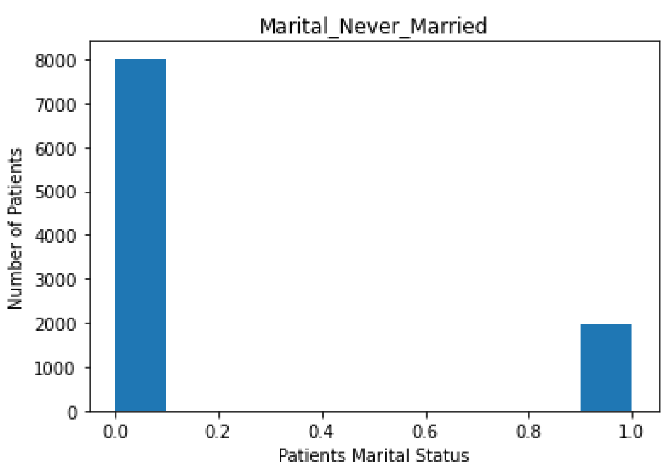

BackPain
Number of Patients
Does patient have BackPain?

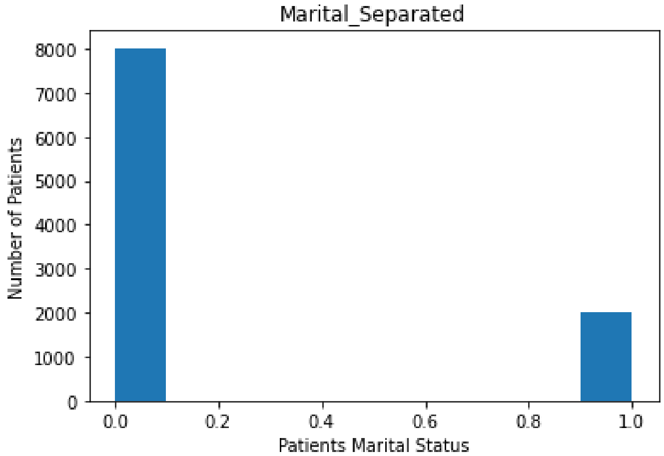

Allergic_rhinitis
Number of Patients
Does patient have Allergic_rhinitis?

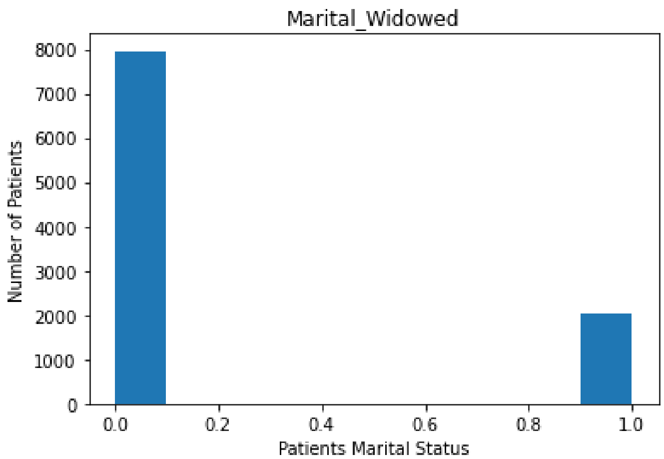

Reflux_esophagitis
Number of Patients
Does patient have Reflux_esophagitis?

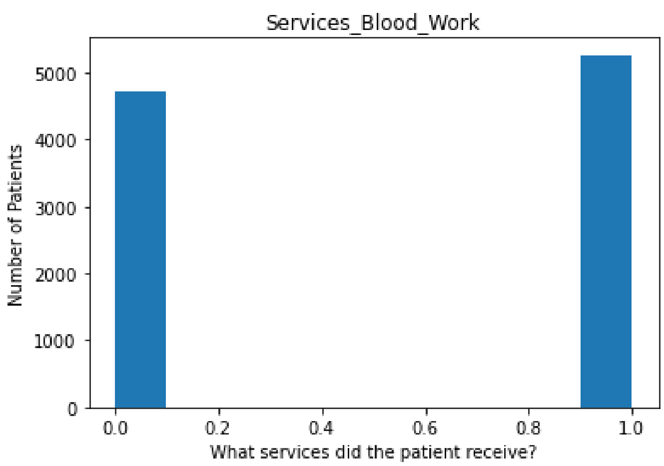

Asthma
Number of Patients
Does patient have Asthma?

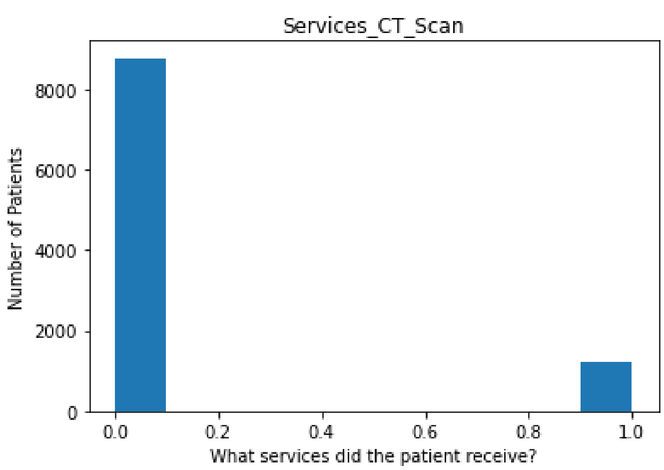

Marital_Married
Number of Patients
Patients Marital Status

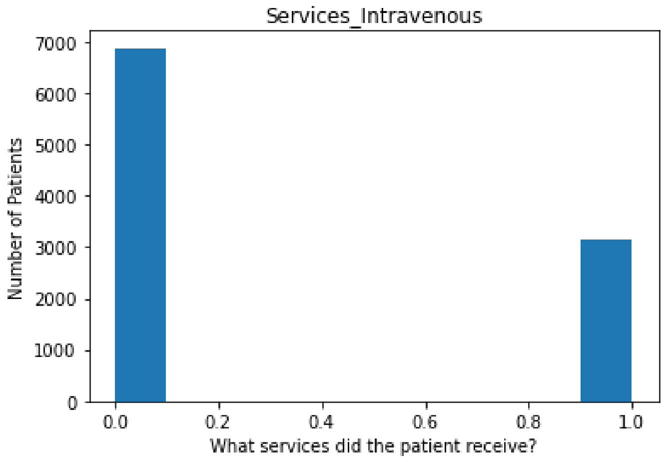

Marital_Never_Married
Number of Patients
Patients Marital Status

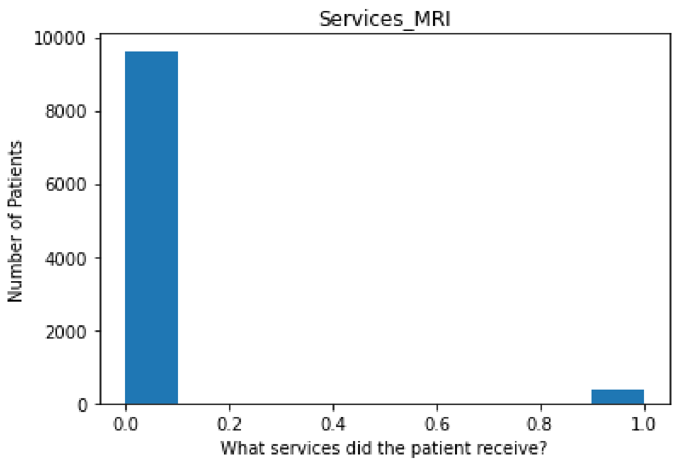

Marital_Separated
Number of Patients
Patients Marital Status
0
1000
2000
3000
4000
5000
6000
7000
8000
0.0
0.2
0.4
0.6
0.8
1.0

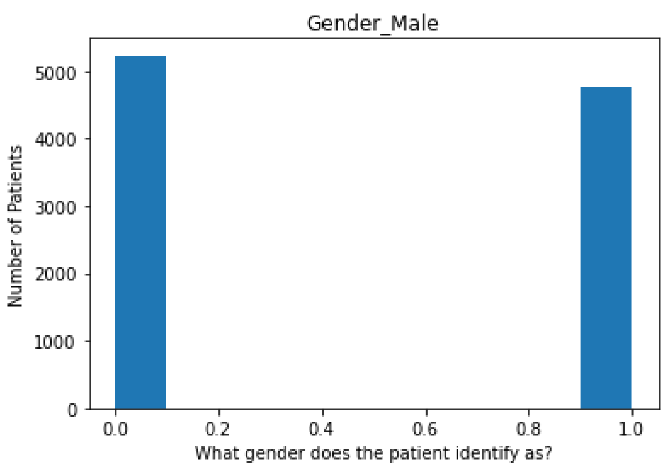

Marital_Widowed
Number of Patients
Patients Marital Status
0
1000
2000
3000
4000
5000
6000
7000
8000
0.0
0.2
0.4
0.6
0.8
1.0

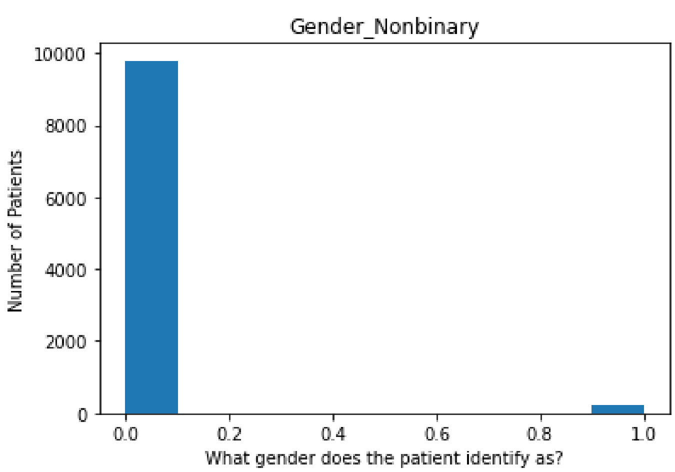

Services_Blood_Work
Number of Patients
What services did the patient receive?
0
1000
2000
3000
4000
5000
0.0
0.2
0.4
0.6
0.8
1.0

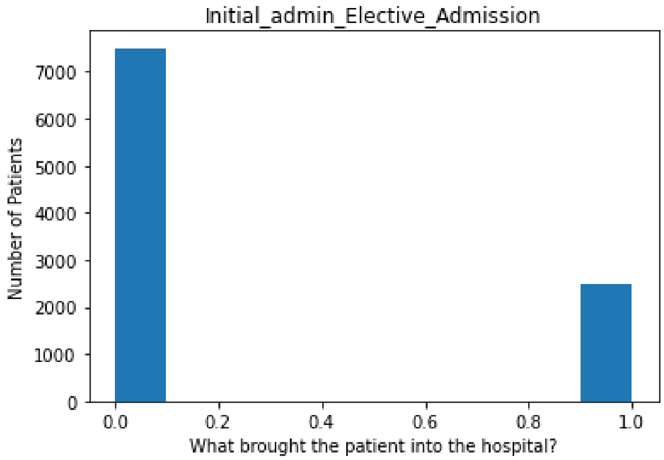

Services_CT_Scan
Number of Patients
What services did the patient receive?
0
2000
4000
6000
8000
0.0
0.2
0.4
0.6
0.8
1.0

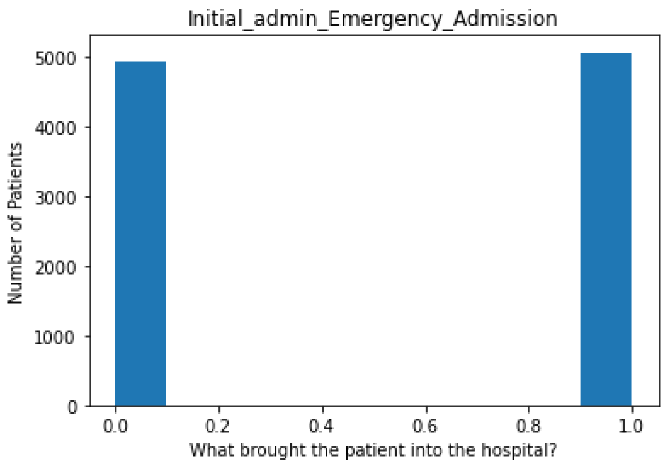

Services_Intravenous
Number of Patients
What services did the patient receive?

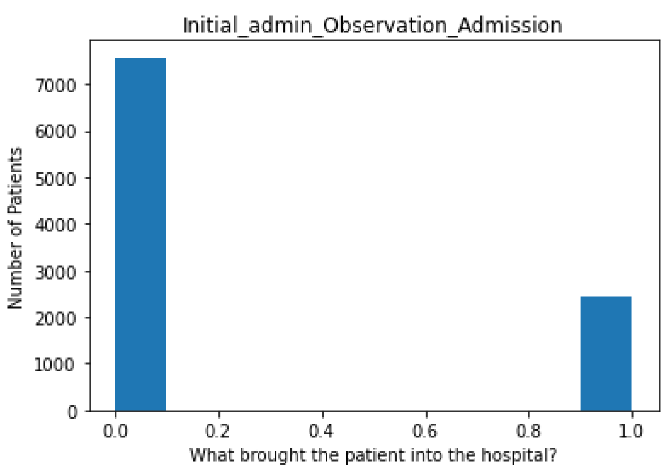

Services_MRI
Number of Patients
What services did the patient receive?

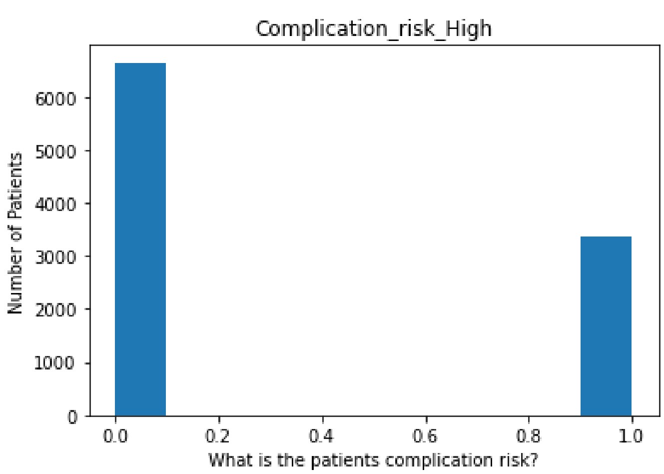

Gender_Male
Number of Patients
What gender does the patient identify as?

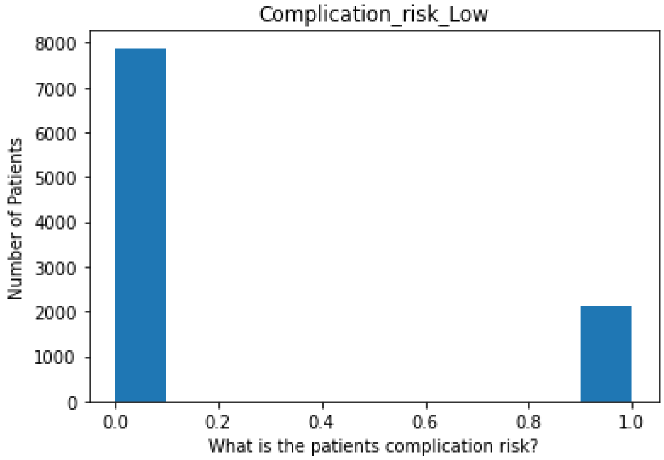

Gender_Nonbinary
Number of Patients
What gender does the patient identify as?

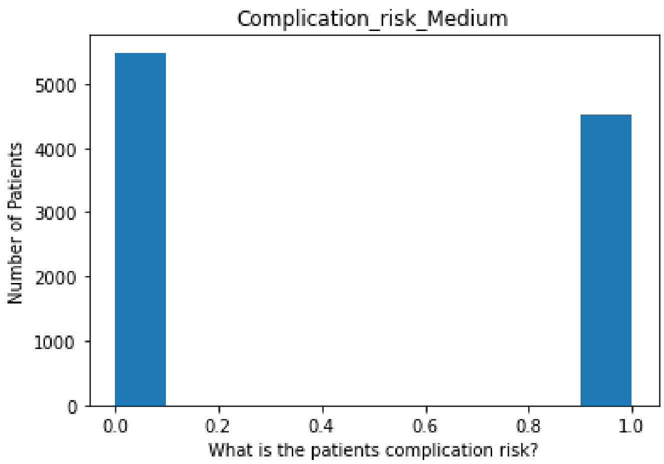

Initial_admin_Elective_Admission
Number of Patients
0
1000
2000
3000
4000
5000
6000
7000
0.0
0.2
0.4
0.6
0.8
1.0
What brought the patient into the hospital?


Initial_admin_Emergency_Admission
Number of Patients
0
1000
2000
3000
4000
5000
0.0
0.2
0.4
0.6
0.8
1.0
What brought the patient into the hospital?


Initial_admin_Observation_Admission
Number of Patients
0
1000
2000
3000
4000
5000
6000
7000
0.0
0.2
0.4
0.6
0.8
1.0
What brought the patient into the hospital?


Complication_risk_High
Number of Patients
0
1000
2000
3000
4000
5000
6000
0.0
0.2
0.4
0.6
0.8
1.0
What is the patients complication risk?

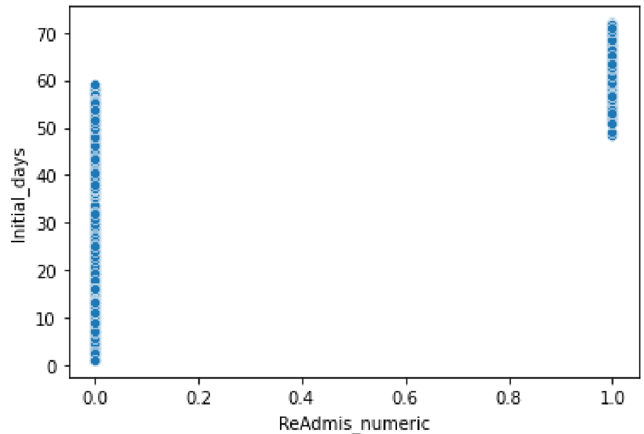


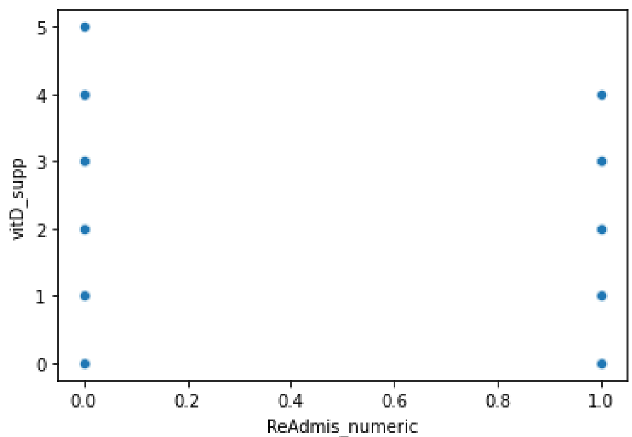


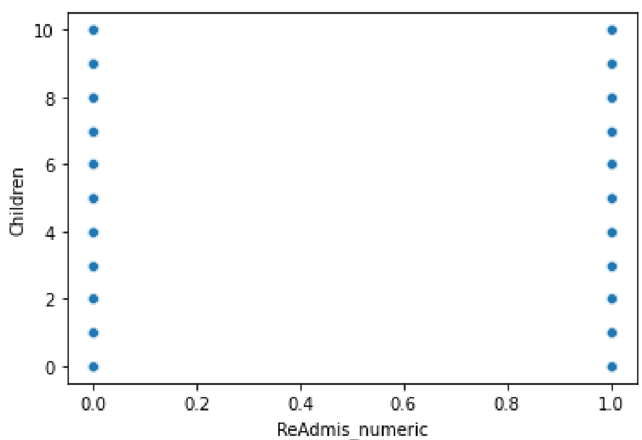
In [9]:
```
#Bivariate Visualizations
sns.scatterplot(data=df, y="Initial_days", x="ReAdmis_numeric")
plt.show()

sns.scatterplot(data=df, y="vitD_supp", x="ReAdmis_numeric")
plt.show()

sns.scatterplot(data=df, y="Children", x="ReAdmis_numeric")
plt.show()

sns.scatterplot(data=df, y="Income", x="ReAdmis_numeric")
plt.show()

sns.scatterplot(data=df, y="Full_meals_eaten", x="ReAdmis_numeric")
plt.show()

sns.scatterplot(data=df, y="Additional_charges", x="ReAdmis_numeric")
plt.show()

sns.scatterplot(data=df, y="TotalCharge", x="ReAdmis_numeric")
plt.show()

sns.scatterplot(data=df, y="VitD_levels", x="ReAdmis_numeric")
plt.show()

sns.scatterplot(data=df, y="Age", x="ReAdmis_numeric")
plt.show()

sns.scatterplot(data=df, y="Doc_visits", x="ReAdmis_numeric")
plt.show()

sns.scatterplot(data=df, y="HighBlood_numeric", x="ReAdmis_numeric")
plt.show()

sns.scatterplot(data=df, y="Stroke_numeric", x="ReAdmis_numeric")
```

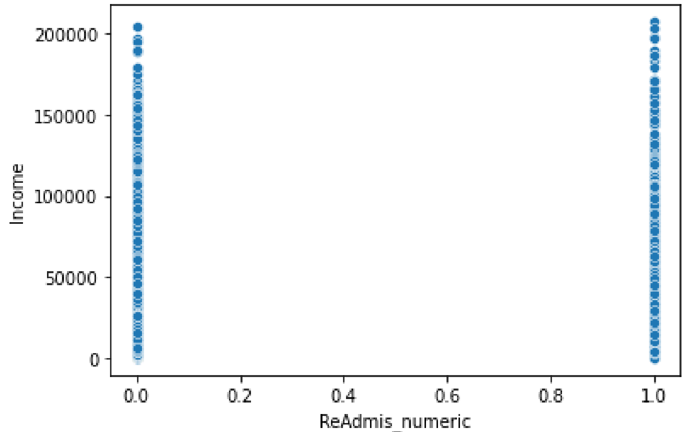
```
plt.show()

sns.scatterplot(data=df, y="Arthritis_numeric", x="ReAdmis_numeric")
plt.show()

sns.scatterplot(data=df, y="Diabetes_numeric", x="ReAdmis_numeric")
plt.show()

sns.scatterplot(data=df, y="Hyperlipidemia_numeric", x="ReAdmis_numeric")
plt.show()

sns.scatterplot(data=df, y="BackPain_numeric", x="ReAdmis_numeric")
plt.show()

sns.scatterplot(data=df, y="Allergic_rhinitis_numeric", x="ReAdmis_numeric")
plt.show()

sns.scatterplot(data=df, y="Reflux_esophagitis_numeric", x="ReAdmis_numeric")
plt.show()

sns.scatterplot(data=df, y="Asthma_numeric", x="ReAdmis_numeric")
plt.show()

sns.scatterplot(data=df, y="Overweight_numeric", x="ReAdmis_numeric")
plt.show()

sns.scatterplot(data=df, y="Anxiety_numeric", x="ReAdmis_numeric")
plt.show()

sns.scatterplot(data=df, y="Marital_Married", x="ReAdmis_numeric")
plt.show()

sns.scatterplot(data=df, y="Marital_Never_Married", x="ReAdmis_numeric")
plt.show()

sns.scatterplot(data=df, y="Marital_Separated", x="ReAdmis_numeric")
plt.show()

sns.scatterplot(data=df, y="Marital_Widowed", x="ReAdmis_numeric")
plt.show()

sns.scatterplot(data=df, y="Services_Blood_Work", x="ReAdmis_numeric")
plt.show()

sns.scatterplot(data=df, y="Services_CT_Scan", x="ReAdmis_numeric")
plt.show()

sns.scatterplot(data=df, y="Services_Intravenous", x="ReAdmis_numeric")
plt.show()

sns.scatterplot(data=df, y="Services_MRI", x="ReAdmis_numeric")
plt.show()

sns.scatterplot(data=df, y="Gender_Male", x="ReAdmis_numeric")
plt.show()

sns.scatterplot(data=df, y="Gender_Nonbinary", x="ReAdmis_numeric")
plt.show()

sns.scatterplot(data=df, y="Initial_admin_Elective_Admission", x="ReAdmis_numeric")
plt.show()

sns.scatterplot(data=df, y="Initial_admin_Emergency_Admission", x="ReAdmis_numeric")
plt.show()

sns.scatterplot(data=df, y="Initial_admin_Observation_Admission", x="ReAdmis_numeric")
plt.show()

sns.scatterplot(data=df, y="Complication_risk_High", x="ReAdmis_numeric")
plt.show()
```

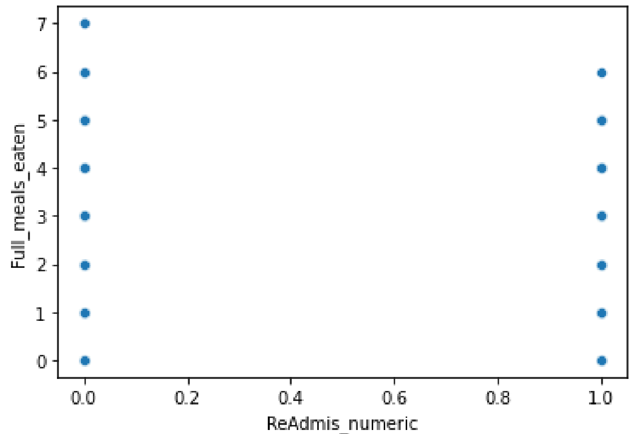
```
sns.scatterplot(data=df, y="Complication_risk_Medium", x="ReAdmis_numeric")
plt.show()

sns.scatterplot(data=df, y="Complication_risk_Low", x="ReAdmis_numeric")
plt.show()
```

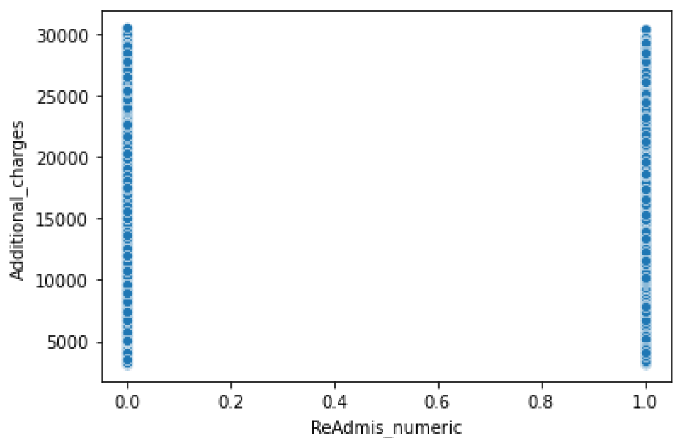


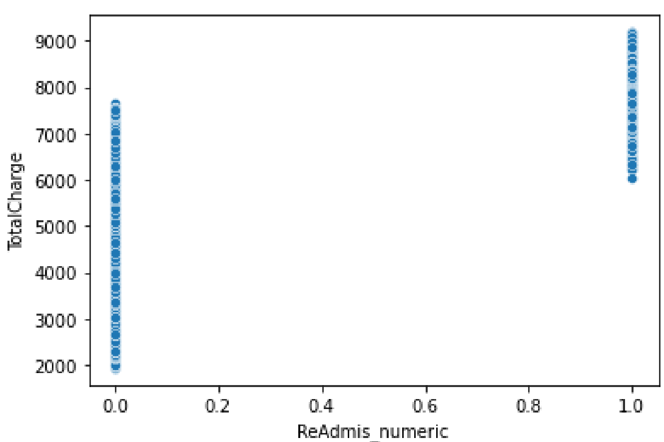


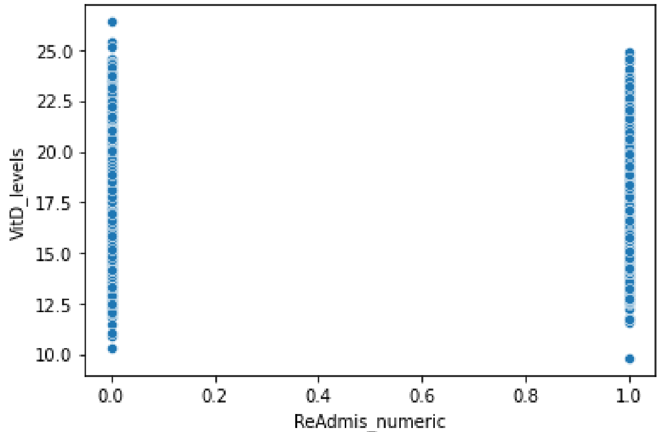


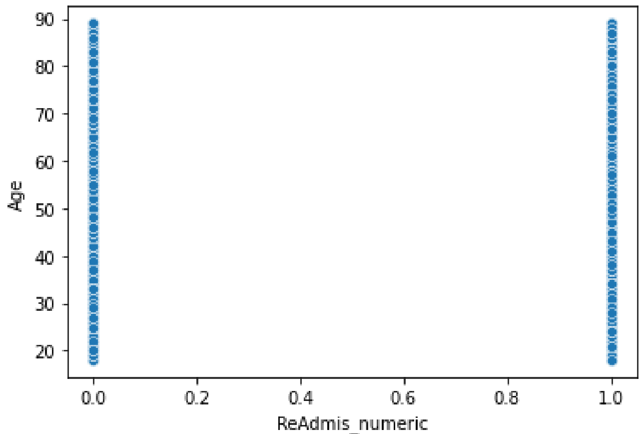

Income
ReAdmis_numeric
0
50000
100000
150000
200000
0.0
0.2
0.4
0.6
0.8
1.0

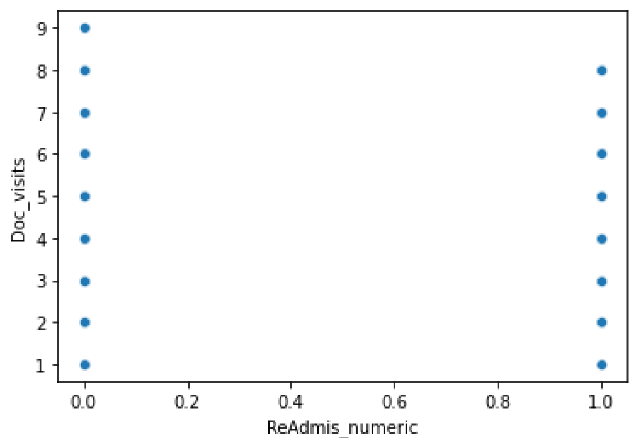

Full_meals_eaten
ReAdmis_numeric
0
1
2
3
4
5
6
7
0.0
0.2
0.4
0.6
0.8
1.0

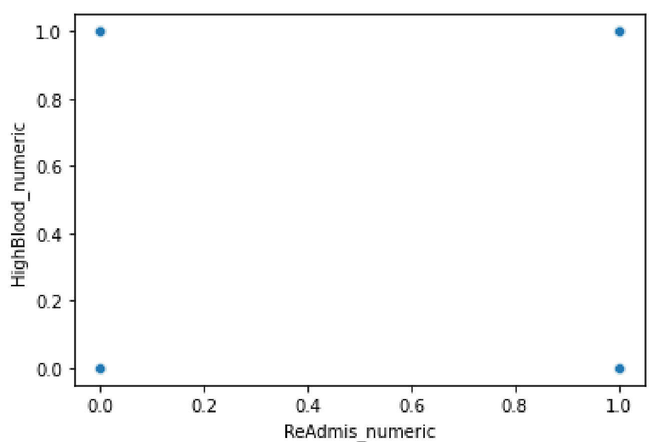

Additional_charges
ReAdmis_numeric
5000
10000
15000
20000
25000
30000
0.0
0.2
0.4
0.6
0.8
1.0

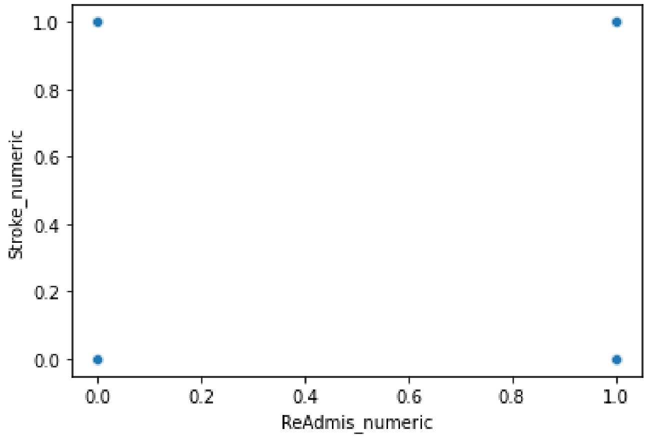

TotalCharge
ReAdmis_numeric
2000
3000
4000
5000
6000
7000
8000
9000
0.0
0.2
0.4
0.6
0.8
1.0

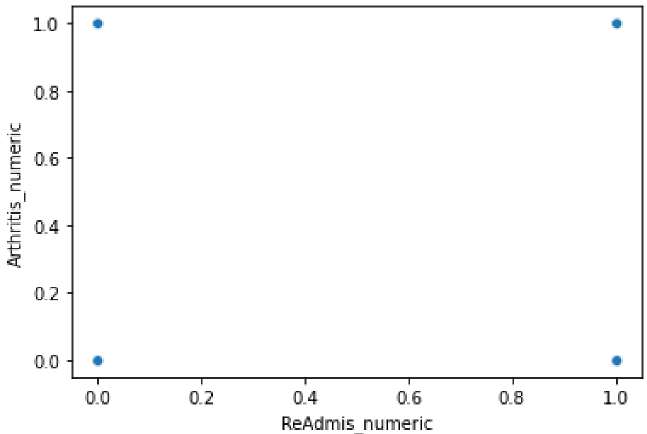

VitD_levels
25.0
22.5
20.0
17.5
15.0
12.5
10.0
0.0
0.2
0.4
0.6
0.8
1.0
ReAdmis_numeric

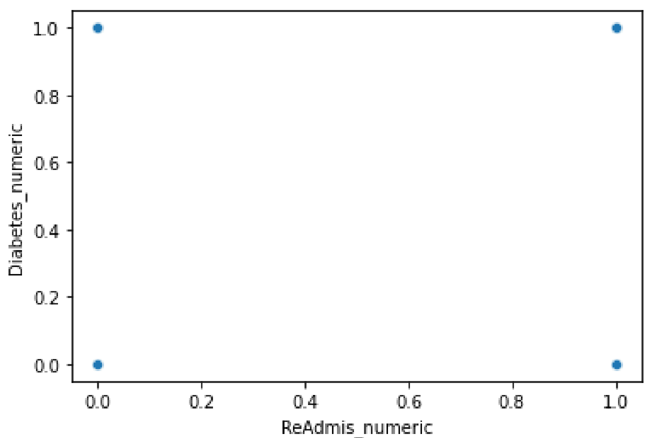

Age
90
80
70
60
50
40
30
20
0.0
0.2
0.4
0.6
0.8
1.0
ReAdmis_numeric

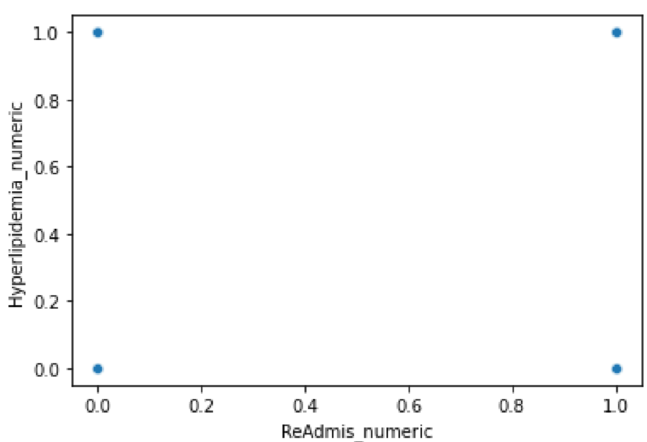

Doc_visits
9
8
7
6
5
4
3
2
1
0.0
0.2
0.4
0.6
0.8
1.0
ReAdmis_numeric

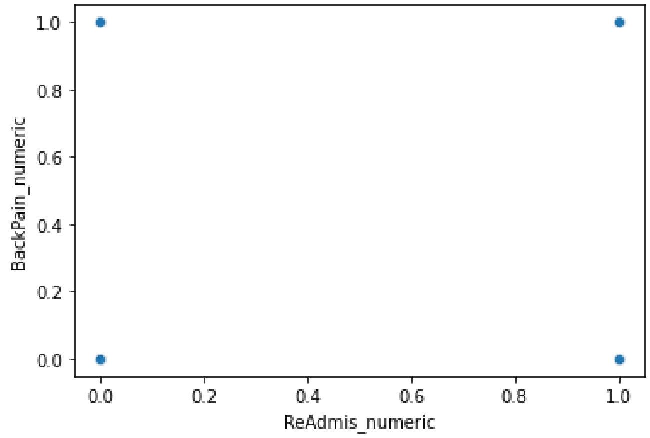

HighBlood_numeric
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0
ReAdmis_numeric

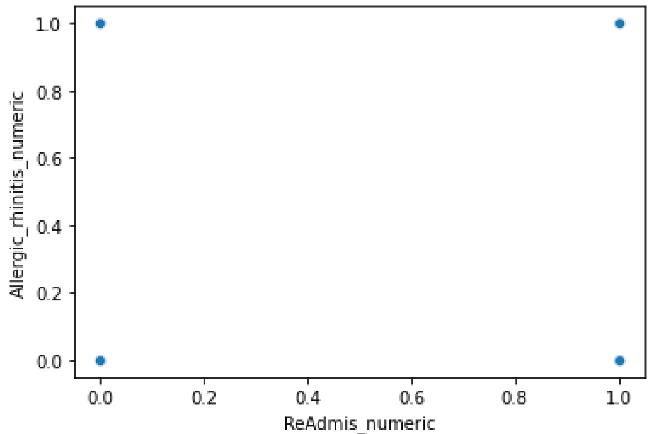

Stroke_numeric
ReAdmis_numeric

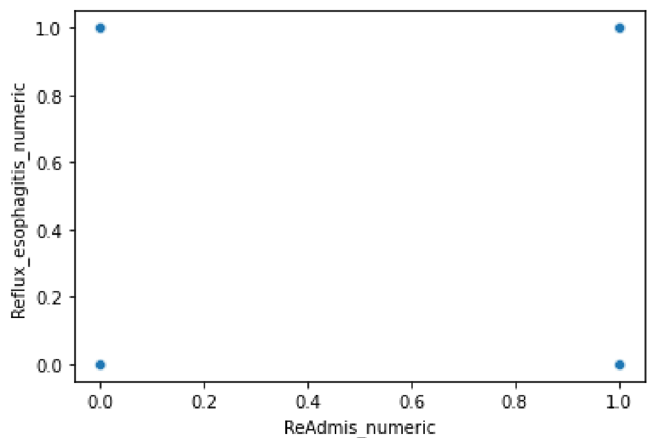

Arthritis_numeric
ReAdmis_numeric

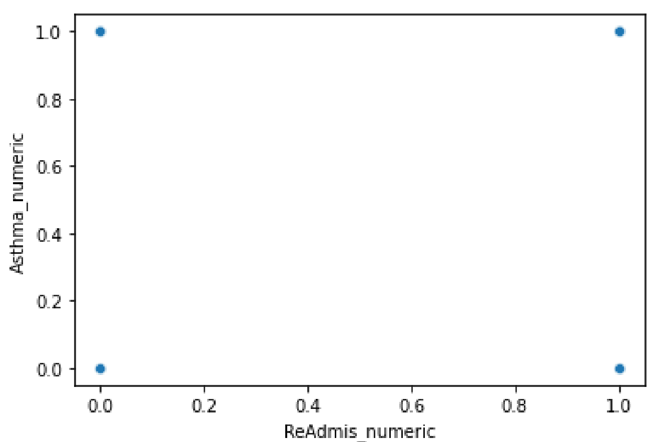

Diabetes_numeric
ReAdmis_numeric

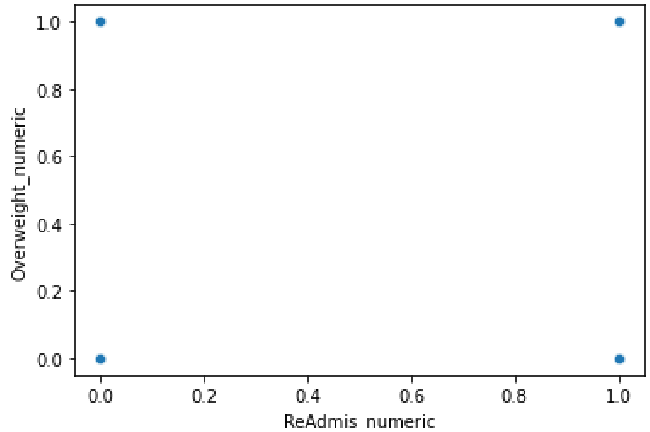

Hyperlipidemia_numeric
ReAdmis_numeric

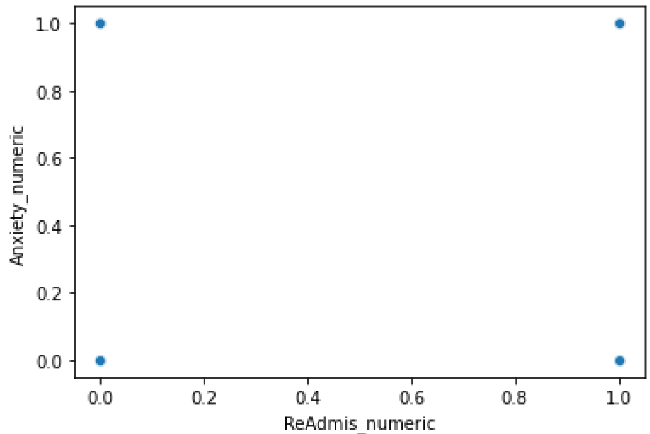

BackPain_numeric
ReAdmis_numeric

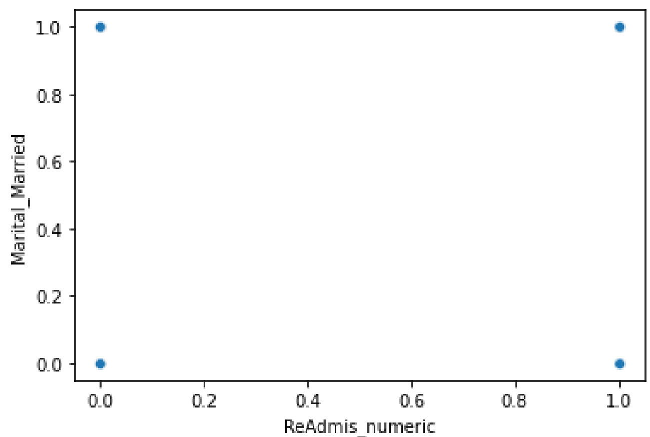

Allergic_rhinitis_numeric
ReAdmis_numeric

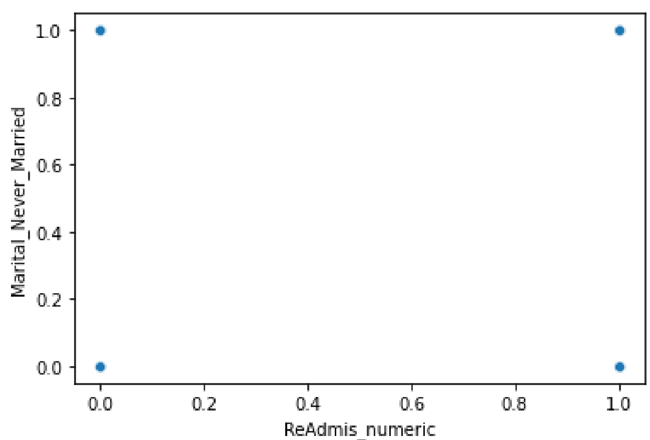

Reflux_esophagitis_numeric
ReAdmis_numeric

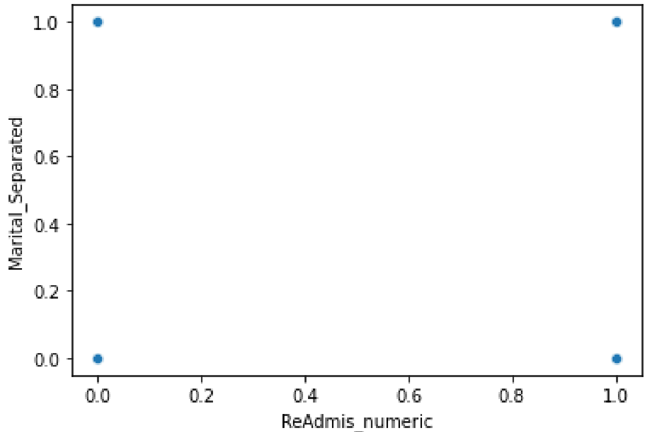

Asthma_numeric
ReAdmis_numeric

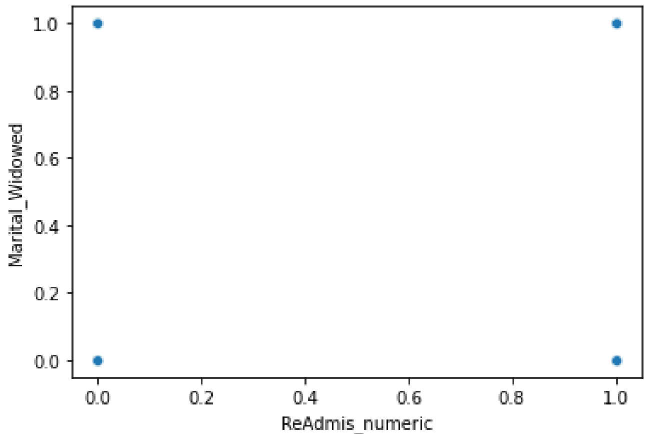

Overweight_numeric
ReAdmis_numeric

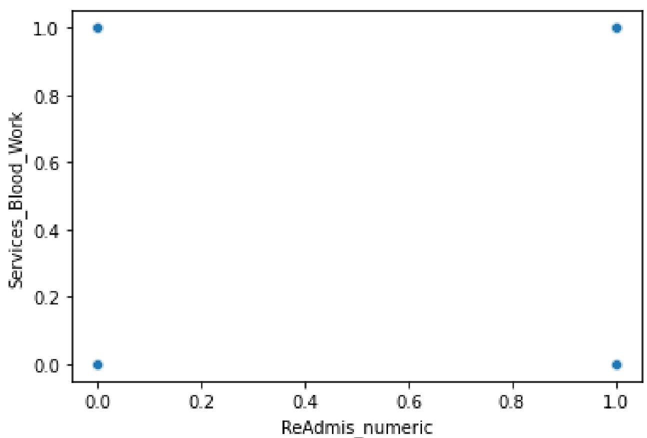

Anxiety_numeric
ReAdmis_numeric

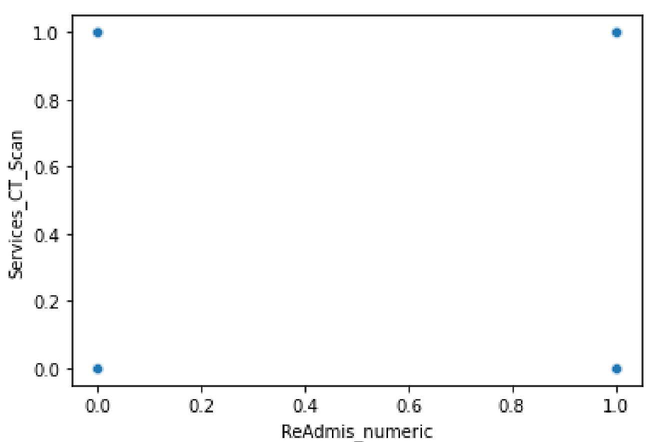

Marital_Married
ReAdmis_numeric

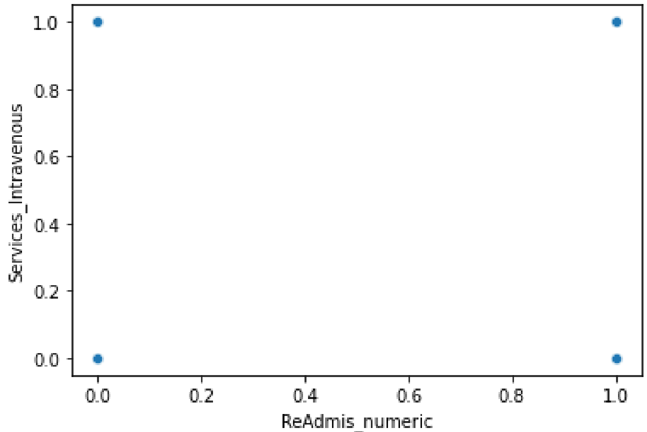

Marital_Never_Married
ReAdmis_numeric

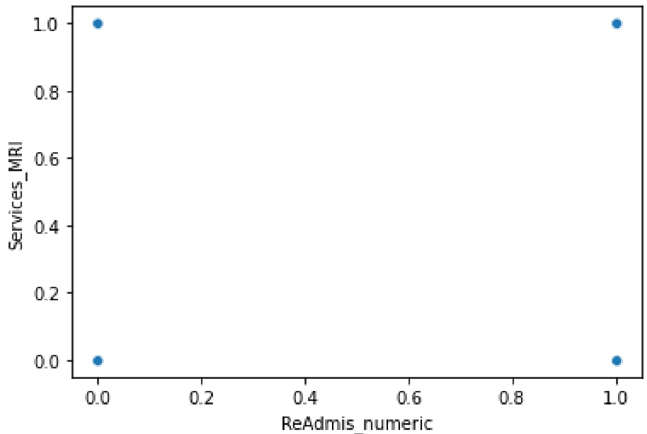

Marital_Separated
ReAdmis_numeric

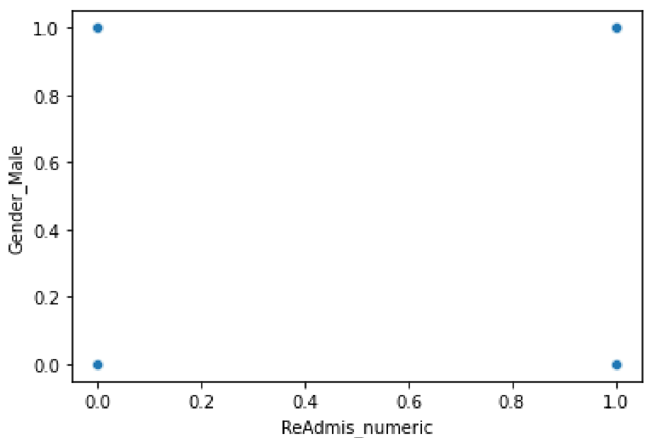

Marital_Widowed
ReAdmis_numeric

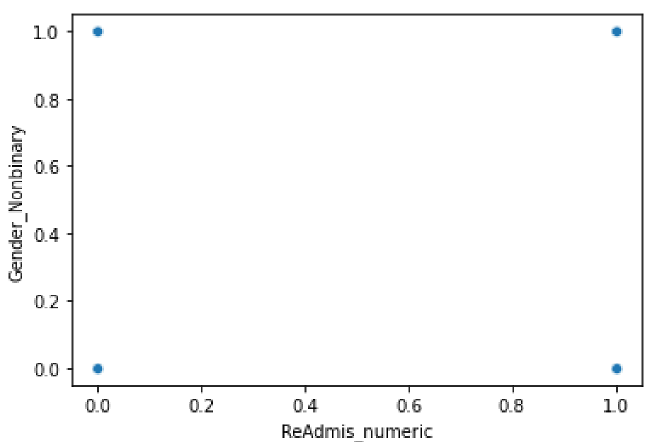

Services_Blood_Work
ReAdmis_numeric

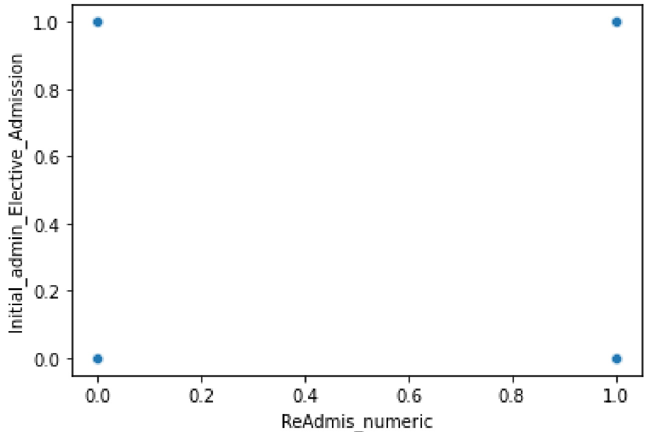

Services_CT_Scan
ReAdmis_numeric

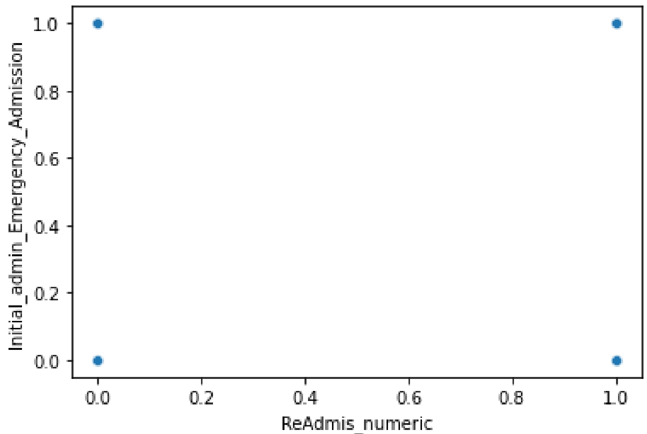

Services_Intravenous
ReAdmis_numeric
0.0
0.2
0.4
0.6
0.8
1.0

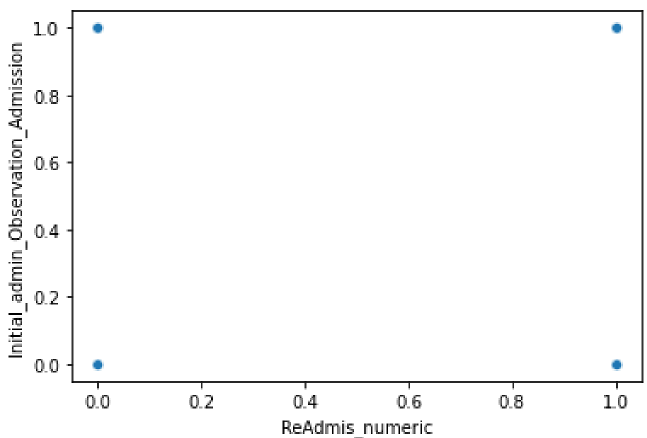

Services_MRI
ReAdmis_numeric
0.0
0.2
0.4
0.6
0.8
1.0

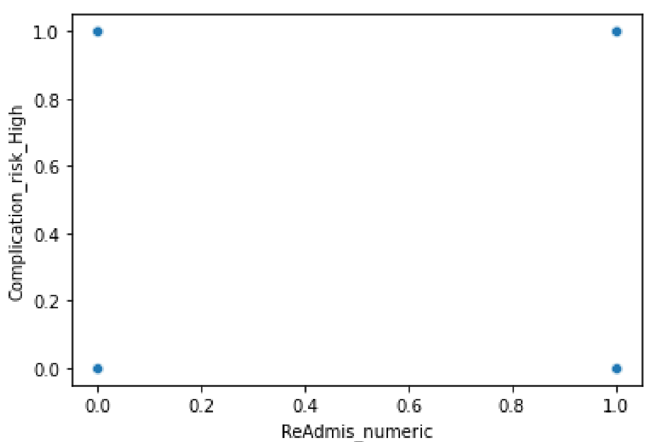

Gender_Male
ReAdmis_numeric
0.0
0.2
0.4
0.6
0.8
1.0

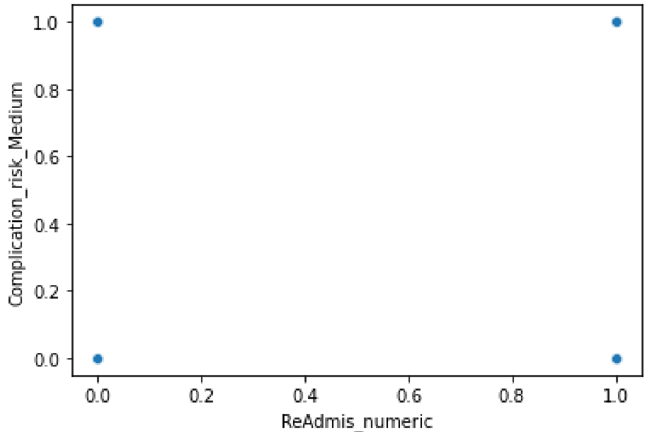

Gender_Nonbinary
ReAdmis_numeric
0.0
0.2
0.4
0.6
0.8
1.0

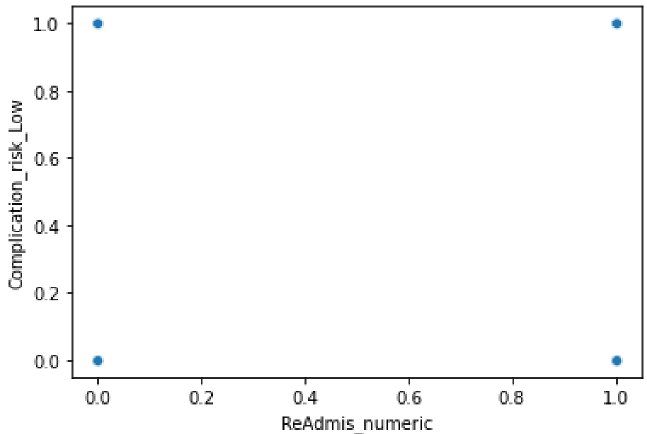

Initial_admin_Elective_Admission
ReAdmis_numeric
0.0
0.2
0.4
0.6
0.8
1.0


Initial_admin_Emergency_Admission
ReAdmis_numeric
0.0
0.2
0.4
0.6
0.8
1.0


Initial_admin_Observation_Admission
ReAdmis_numeric
0.0
0.2
0.4
0.6
0.8
1.0


Complication_risk_High
ReAdmis_numeric
0.0
0.2
0.4
0.6
0.8
1.0





In [10]:
```
df.to_csv(r'C:\Users\mmorg\Desktop\D208 Assessment Files\Cleaned208data.csv')
```

In [11]:
```
#Initial Logistic Regression Model
mdl_readmis_vs_variables = logit("ReAdmis_numeric ~ Initial_days + vitD_supp + Children + Income + Full_

print(mdl_readmis_vs_variables.params)

mdl_readmis_vs_variables.summary()
```

```
Warning: Maximum number of iterations has been exceeded.
         Current function value: 0.032914
         Iterations: 35
Intercept                             -78.391168
Initial_days                           -1.143612
vitD_supp                              -0.105465
Children                                0.088933
Income                                  0.000002
Full_meals_eaten                        0.048269
Additional_charges                      0.000047
TotalCharge                             0.032097
VitD_levels                             0.029913
Age                                    -0.008115
Doc_visits                              0.006554
HighBlood_numeric                      -3.125420
Stroke_numeric                          1.651497
Arthritis_numeric                      -3.669106
Diabetes_numeric                       -1.908568
Hyperlipidemia_numeric                 -2.757763
BackPain_numeric                       -2.508272
Allergic_rhinitis_numeric              -2.268662
Reflux_esophagitis_numeric             -2.342286
Asthma_numeric                         -1.389130
Overweight_numeric                     -0.286489
Anxiety_numeric                        -3.817203
Marital_Married                         0.268609
Marital_Never_Married                   0.356522
Marital_Separated                      -0.127193
Marital_Widowed                         0.136171
Services_Blood_Work                   -20.676842
Services_CT_Scan                      -19.061885
Services_Intravenous                  -20.683935
Services_MRI                          -17.968342
Gender_Male                             0.169142
Gender_Nonbinary                        0.365800
Initial_admin_Elective_Admission      -21.750667
Initial_admin_Emergency_Admission     -35.680140
Initial_admin_Observation_Admission   -20.960246
Complication_risk_High                -34.251935
Complication_risk_Low                 -22.802367
Complication_risk_Medium              -21.336813
dtype: float64
```

```
C:\Users\mmorg\anaconda3\lib\site-packages\statsmodels\base\model.py:604: ConvergenceWarning: Maximum Li
kelihood optimization failed to converge. Check mle_retvals
  warnings.warn("Maximum Likelihood optimization failed to "
```

Out[11]:

Logit Regression Results

Dep. Variable:	ReAdmis_numeric	**No. Observations:**	10000
Model:	Logit	**Df Residuals:**	9965
Method:	MLE	**Df Model:**	34
Date:	Thu, 24 Nov 2022	**Pseudo R-squ.:**	0.9499
Time:	22:22:07	**Log-Likelihood:**	-329.14
converged:	False	**LL-Null:**	-6572.9
Covariance Type:	nonrobust	**LLR p-value:**	0.000

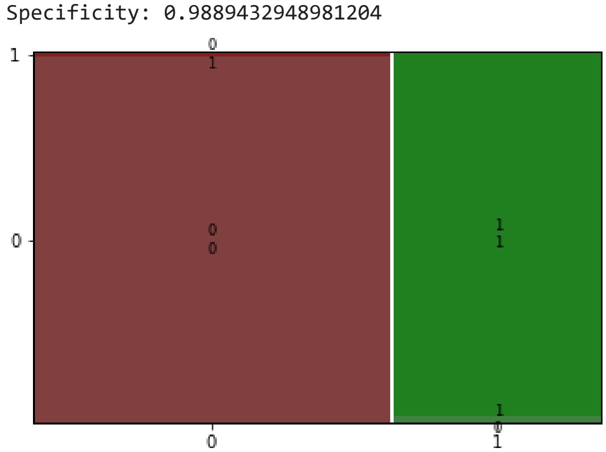
	coef	std err	z	P>|z|	[0.025	0.975]
Intercept	-78.3912	nan	nan	nan	nan	nan
Initial_days	-1.1436	1.5e+04	-7.62e-05	1.000	-2.94e+04	2.94e+04
vitD_supp	-0.1055	0.167	-0.632	0.528	-0.433	0.222
Children	0.0889	0.046	1.913	0.056	-0.002	0.180
Income	1.572e-06	3.62e-06	0.434	0.664	-5.53e-06	8.67e-06
Full_meals_eaten	0.0483	0.103	0.469	0.639	-0.153	0.250
Additional_charges	4.664e-05	6.25e-05	0.746	0.455	-7.58e-05	0.000
TotalCharge	0.0321	183.195	0.000	1.000	-359.024	359.088
VitD_levels	0.0299	0.049	0.613	0.540	-0.066	0.126
Age	-0.0081	0.015	-0.549	0.583	-0.037	0.021
Doc_visits	0.0066	0.098	0.067	0.947	-0.186	0.200
HighBlood_numeric	-3.1254	2.06e+04	-0.000	1.000	-4.03e+04	4.03e+04
Stroke_numeric	1.6515	0.272	6.069	0.000	1.118	2.185
Arthritis_numeric	-3.6691	1.32e+04	-0.000	1.000	-2.58e+04	2.58e+04
Diabetes_numeric	-1.9086	1.38e+04	-0.000	1.000	-2.7e+04	2.7e+04
Hyperlipidemia_numeric	-2.7578	1.72e+04	-0.000	1.000	-3.38e+04	3.37e+04
BackPain_numeric	-2.5083	1.56e+04	-0.000	1.000	-3.06e+04	3.06e+04
Allergic_rhinitis_numeric	-2.2687	1.11e+04	-0.000	1.000	-2.18e+04	2.17e+04
Reflux_esophagitis_numeric	-2.3423	1.09e+04	-0.000	1.000	-2.14e+04	2.14e+04
Asthma_numeric	-1.3891	0.237	-5.852	0.000	-1.854	-0.924
Overweight_numeric	-0.2865	0.229	-1.250	0.211	-0.736	0.163
Anxiety_numeric	-3.8172	1.58e+04	-0.000	1.000	-3.09e+04	3.09e+04
Marital_Married	0.2686	0.331	0.811	0.418	-0.381	0.918
Marital_Never_Married	0.3565	0.338	1.055	0.292	-0.306	1.019
Marital_Separated	-0.1272	0.344	-0.369	0.712	-0.802	0.548
Marital_Widowed	0.1362	0.333	0.409	0.682	-0.516	0.788
Services_Blood_Work	-20.6768	4.07e+06	-5.08e-06	1.000	-7.98e+06	7.98e+06
Services_CT_Scan	-19.0619	4.07e+06	-4.68e-06	1.000	-7.98e+06	7.98e+06
Services_Intravenous	-20.6839	4.07e+06	-5.08e-06	1.000	-7.98e+06	7.98e+06
Services_MRI	-17.9683	4.07e+06	-4.41e-06	1.000	-7.99e+06	7.99e+06
Gender_Male	0.1691	0.210	0.807	0.420	-0.242	0.580
Gender_Nonbinary	0.3658	0.714	0.512	0.608	-1.033	1.765

Initial_admin_Elective_Admission	-21.7507	nan	nan	nan	nan	nan
Initial_admin_Emergency_Admission	-35.6801	nan	nan	nan	nan	nan
Initial_admin_Observation_Admission	-20.9602	nan	nan	nan	nan	nan
Complication_risk_High	-34.2519	2.15e+06	-1.59e-05	1.000	-4.22e+06	4.22e+06
Complication_risk_Low	-22.8024	2.34e+06	-9.75e-06	1.000	-4.59e+06	4.59e+06
Complication_risk_Medium	-21.3368	2.34e+06	-9.12e-06	1.000	-4.59e+06	4.59e+06

Possibly complete quasi-separation: A fraction 0.82 of observations can be perfectly predicted. This might indicate that there is complete quasi-separation. In this case some parameters will not be identified.

In [12]:
```
#Confusion Matrix for Initial Model

conf_matrix = mdl_readmis_vs_variables.pred_table()

print(conf_matrix)

from statsmodels.graphics.mosaicplot import mosaic

mosaic(conf_matrix)

#Calculating accuracy: the proportion of correct predictions

TN = conf_matrix[0,0]
TP = conf_matrix[1,1]
FN = conf_matrix[1,0]
FP = conf_matrix[0,1]

acc = (TN + TP) / (TN + TP + FN + FP)
print('Accuracy:', acc)

#Sensitivity: proportion of true positives

sens = TP / (FN + TP)
print('Sensitivity:', sens)

#Specificity: proportion of true negatives

spec = TN / (TN + FP)
print('Specificity:', spec)
```

```
[[6261.   70.]
 [  60. 3609.]]
Accuracy: 0.987
Sensitivity: 0.9836467702371219
Specificity: 0.9889432948981204
```



In [13]:
```
#Variable Selection
# Checking for the VIF values of the variables.
from statsmodels.stats.outliers_influence import variance_inflation_factor

X = df[['Initial_days', 'vitD_supp', 'Children', 'Income', 'Full_meals_eaten', 'Additional_charges', 'To


# VIF dataframe
vif_data = pd.DataFrame()
vif_data["feature"] = X.columns

# calculating VIF for each feature
vif_data["VIF"] = [variance_inflation_factor(X.values, i)
                          for i in range(len(X.columns))]

print(vif_data)
```

```
                                feature          VIF
0                          Initial_days  2880.163153
1                             vitD_supp     1.003676
2                              Children     1.003506
3                                Income     1.002683
4                      Full_meals_eaten     1.004107
5                    Additional_charges    16.303881
6                           TotalCharge  2944.078834
7                           VitD_levels     1.003914
8                                   Age     9.273563
9                            Doc_visits     1.003377
10                     HighBlood_numeric     9.711378
11                        Stroke_numeric     1.010014
12                     Arthritis_numeric     1.760819
13                      Diabetes_numeric     1.696512
14                 Hyperlipidemia_numeric     2.198017
15                      BackPain_numeric     2.112077
16              Allergic_rhinitis_numeric     1.551215
17             Reflux_esophagitis_numeric     1.527188
18                        Asthma_numeric     1.003104
19                       Marital_Married     1.627325
20                 Marital_Never_Married     1.618488
21                     Marital_Separated     1.617166
22                       Marital_Widowed     1.630512
23                   Services_Blood_Work          inf
24                      Services_CT_Scan          inf
25                  Services_Intravenous          inf
26                          Services_MRI          inf
27                           Gender_Male     1.026146
28                      Gender_Nonbinary     1.023726
29      Initial_admin_Elective_Admission          inf
30     Initial_admin_Emergency_Admission          inf
31   Initial_admin_Observation_Admission          inf
32                 Complication_risk_High          inf
33                  Complication_risk_Low          inf
34               Complication_risk_Medium          inf
```

```
C:\Users\mmorg\anaconda3\lib\site-packages\statsmodels\stats\outliers_influence.py:195: RuntimeWarning:
divide by zero encountered in double_scalars
  vif = 1. / (1. - r_squared_i)
```

In [14]:
```
#Reduced Model removing complication risk, initial admin, services, demographics, and charges columns du

mdl_readmis_vs_variables1 = logit("ReAdmis_numeric ~ Initial_days + Children + Stroke_numeric + Asthma_n

print(mdl_readmis_vs_variables1.params)

mdl_readmis_vs_variables1.summary()
```

```
Optimization terminated successfully.
         Current function value: 0.045704
         Iterations 13
Intercept        -57.974445
Initial_days       1.065053
Children           0.069841
Stroke_numeric     1.273473
Asthma_numeric    -0.944019
dtype: float64
```

Out[14]:

Logit Regression Results

Dep. Variable:	ReAdmis_numeric	**No. Observations:**	10000
Model:	Logit	**Df Residuals:**	9995
Method:	MLE	**Df Model:**	4
Date:	Thu, 24 Nov 2022	**Pseudo R-squ.:**	0.9305
Time:	22:23:54	**Log-Likelihood:**	-457.04
converged:	True	**LL-Null:**	-6572.9
Covariance Type:	nonrobust	**LLR p-value:**	0.000

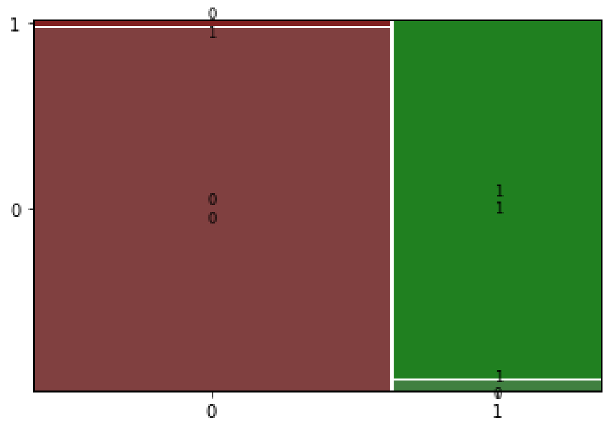
	coef	std err	z	P>|z|	[0.025	0.975]
Intercept	-57.9744	2.749	-21.087	0.000	-63.363	-52.586
Initial_days	1.0651	0.050	21.125	0.000	0.966	1.164
Children	0.0698	0.038	1.839	0.066	-0.005	0.144
Stroke_numeric	1.2735	0.223	5.706	0.000	0.836	1.711
Asthma_numeric	-0.9440	0.190	-4.975	0.000	-1.316	-0.572

Possibly complete quasi-separation: A fraction 0.75 of observations can be perfectly predicted. This might indicate that there is complete quasi-separation. In this case some parameters will not be identified.

In [15]:

```
#Confusion Matrix for Reduced Model

conf_matrix = mdl_readmis_vs_variables1.pred_table()

print(conf_matrix)

from statsmodels.graphics.mosaicplot import mosaic

mosaic(conf_matrix)

#Calculating accuracy: the proportion of correct predictions

TN = conf_matrix[0,0]
TP = conf_matrix[1,1]
FN = conf_matrix[1,0]
FP = conf_matrix[0,1]

acc = (TN + TP) / (TN + TP + FN + FP)
print('Accuracy:', acc)

#Sensitivity: proportion of true positives

sens = TP / (FN + TP)
print('Sensitivity:', sens)

#Specificity: proportion of true negatives

spec = TN / (TN + FP)
print('Specificity:', spec)
```

```
[[6228.  103.]
 [ 102. 3567.]]
Accuracy: 0.9795
Sensitivity: 0.9721995094031071
Specificity: 0.9837308482072342
```



In [16]:
```
#Calculating AIC of Initial Model
from sklearn.linear_model import LinearRegression
import statsmodels.api as sm

#define response variable
y = df['ReAdmis_numeric']

#define predictor variables
x = df[['Initial_days', 'vitD_supp', 'Children', 'Income', 'Full_meals_eaten', 'Additional_charges', 'To

#add constant to predictor variables
x = sm.add_constant(x)

#fit regression model
model = sm.OLS(y, x).fit()

#view AIC of model
print(model.aic)
```

```
897.5899676679401
```

In [17]:
```
#Calculating AIC of Reduced Model #1

#define response variable
y = df['ReAdmis_numeric']

#define predictor variables
x = df[['Initial_days', 'Children', 'Stroke_numeric', 'Asthma_numeric', 'Overweight_numeric']]

#add constant to predictor variables
x = sm.add_constant(x)

#fit regression model
model = sm.OLS(y, x).fit()

#view AIC of model
print(model.aic)
```

```
918.392404996368
```

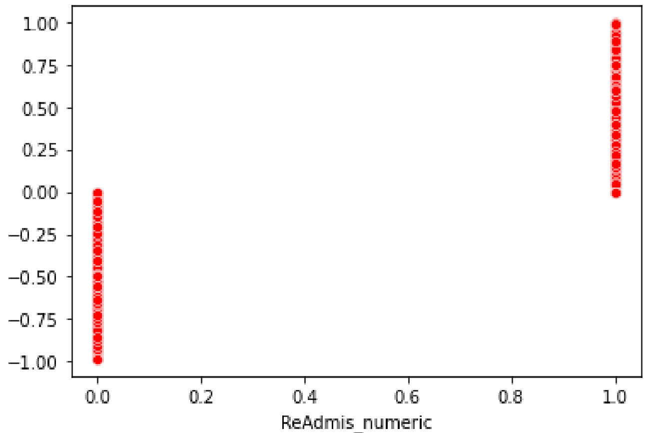
In [18]:
```
#Residual Standard Error for reduced model
np.sqrt(mdl_readmis_vs_variables1.scale)
```

Out[18]:
```
1.0
```

In [19]:
```
#Residual plot for reduced model
df['intercept'] = 1
```

```
residuals = df['ReAdmis_numeric'] - mdl_readmis_vs_variables1.predict(df[['Initial_days', 'Children', 'S
sns.scatterplot(x=df['ReAdmis_numeric'], y=residuals, color='red')
```

Out[19]: <AxesSubplot:xlabel='ReAdmis_numeric'>



In [ ]: