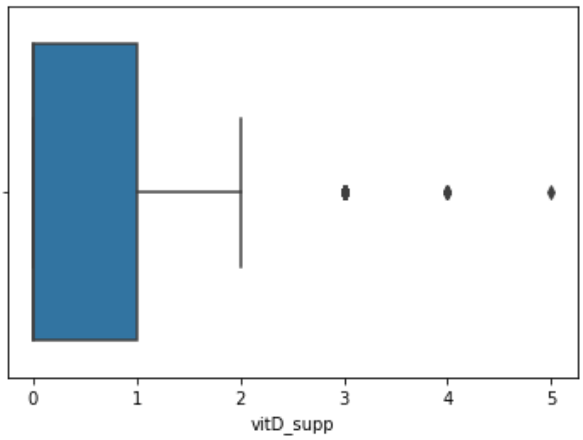
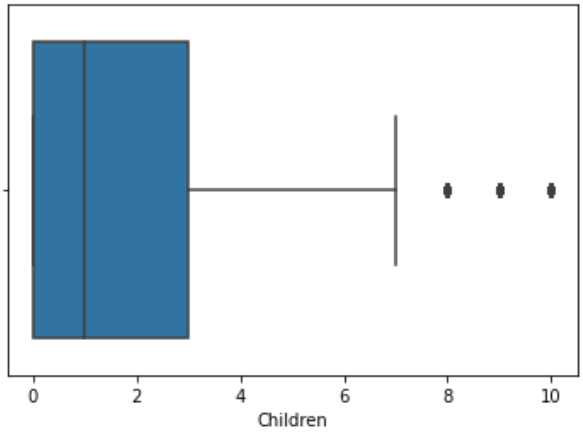
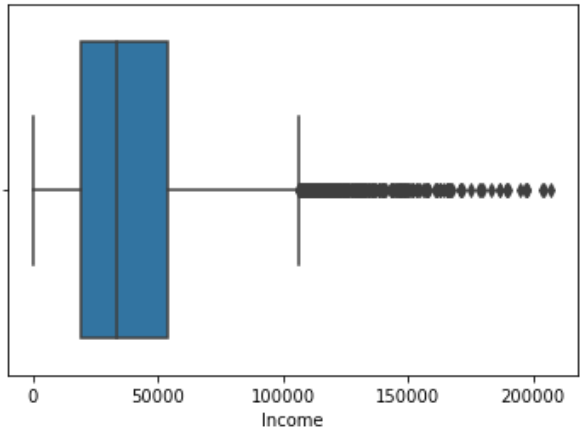
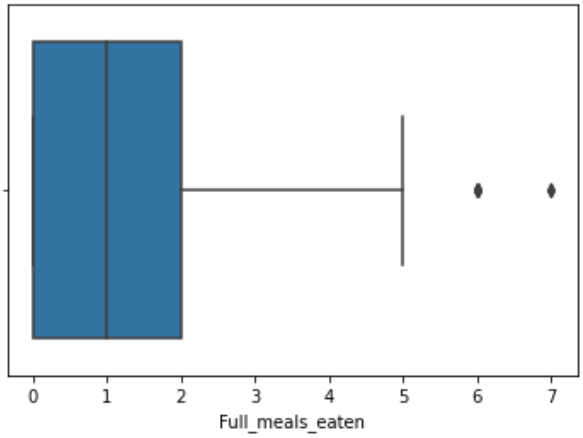
In [2]:
```
import pandas as pd

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 ols

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

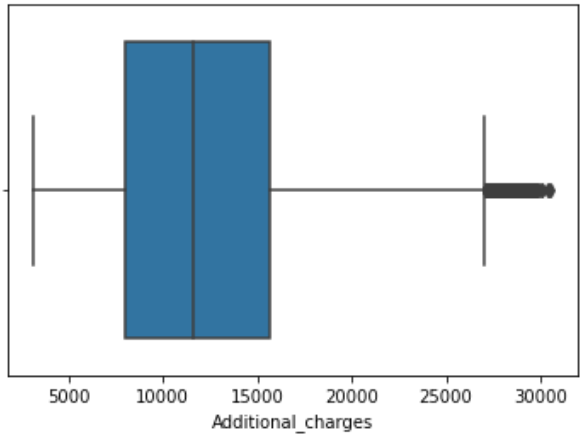
In [3]:
```
#Detection of outliers
boxplot=sns.boxplot(x='vitD_supp',data=df)
plt.show()
boxplot=sns.boxplot(x='Children',data=df)
plt.show()
boxplot=sns.boxplot(x='Income',data=df)
plt.show()
boxplot=sns.boxplot(x='Full_meals_eaten',data=df)
plt.show()
boxplot=sns.boxplot(x='Additional_charges',data=df)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

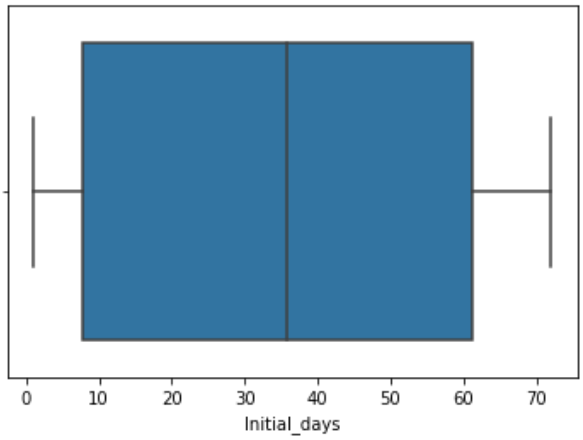


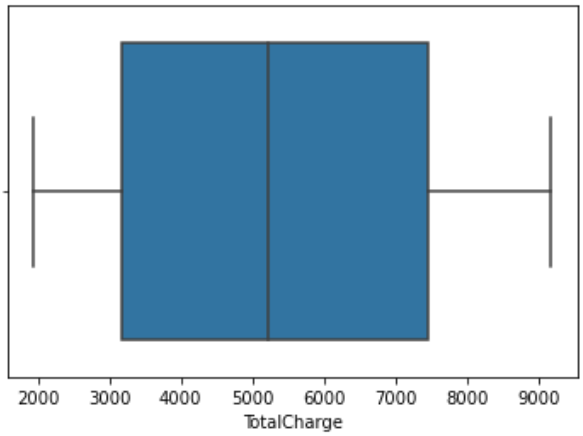

0
2
4
6
8
10
Children

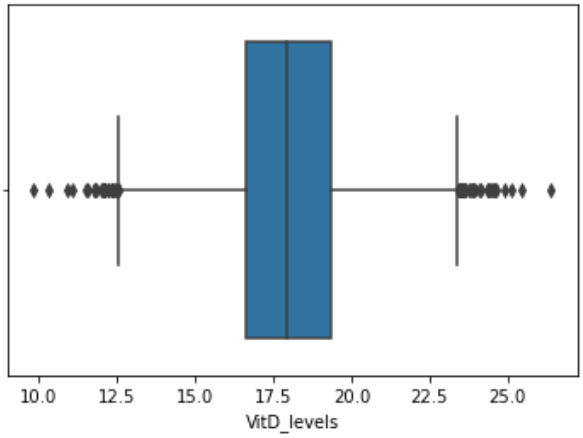

0
50000
100000
150000
200000
Income

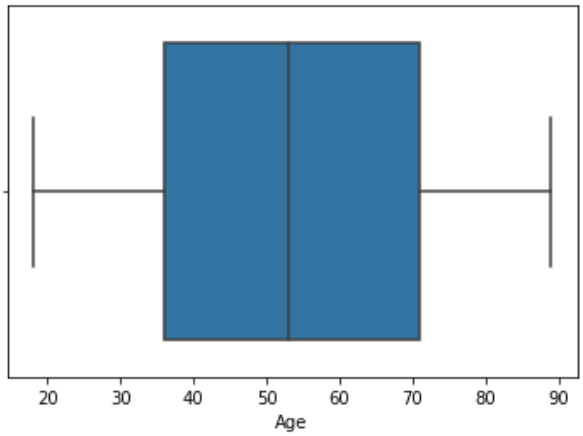

0
1
2
3
4
5
6
7
Full_meals_eaten

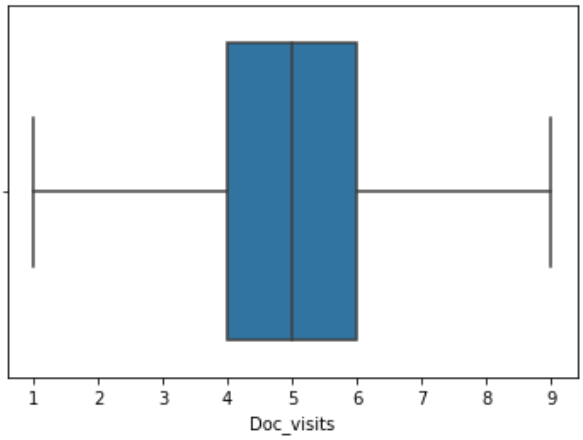

5000
10000
15000
20000
25000
30000
Additional_charges


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 = 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 74 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  Marital_Divorced                    10000 non-null  uint8  
 57  Marital_Married                     10000 non-null  uint8  
 58  Marital_Never Married               10000 non-null  uint8  
 59  Marital_Separated                   10000 non-null  uint8  
 60  Marital_Widowed                     10000 non-null  uint8  
 61  Services_Blood Work                 10000 non-null  uint8  
 62  Services_CT Scan                    10000 non-null  uint8  
 63  Services_Intravenous                10000 non-null  uint8  
 64  Services_MRI                        10000 non-null  uint8  
 65  Gender_Female                       10000 non-null  uint8  
```

```
 66  Gender_Male                          10000 non-null  uint8
 67  Gender_Nonbinary                     10000 non-null  uint8
 68  Initial_admin_Elective Admission     10000 non-null  uint8
 69  Initial_admin_Emergency Admission    10000 non-null  uint8
 70  Initial_admin_Observation Admission  10000 non-null  uint8
 71  Complication_risk_High               10000 non-null  uint8
 72  Complication_risk_Low                10000 non-null  uint8
 73  Complication_risk_Medium             10000 non-null  uint8
dtypes: float64(7), int64(27), object(22), uint8(18)
memory usage: 4.4+ MB
```

In [5]:

```
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 74 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  Marital_Divorced                10000 non-null  uint8  
 57  Marital_Married                 10000 non-null  uint8  
 58  Marital_Never_Married           10000 non-null  uint8  
 59  Marital_Separated               10000 non-null  uint8  
 60  Marital_Widowed                 10000 non-null  uint8  
 61  Services_Blood_Work             10000 non-null  uint8  
 62  Services_CT_Scan                10000 non-null  uint8  
 63  Services_Intravenous            10000 non-null  uint8  
 64  Services_MRI                    10000 non-null  uint8  
 65  Gender_Female                   10000 non-null  uint8  
```

```
 66  Gender_Male                          10000 non-null  uint8
 67  Gender_Nonbinary                     10000 non-null  uint8
 68  Initial_admin_Elective_Admission     10000 non-null  uint8
 69  Initial_admin_Emergency_Admission    10000 non-null  uint8
 70  Initial_admin_Observation_Admission  10000 non-null  uint8
 71  Complication_risk_High               10000 non-null  uint8
 72  Complication_risk_Low                10000 non-null  uint8
 73  Complication_risk_Medium             10000 non-null  uint8
dtypes: float64(7), int64(27), object(22), uint8(18)
memory usage: 4.4+ MB
```

In [6]:

```
##Univariate Stats Dataframe
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_Divorced	10000	0	2	uint8	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
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
Gender_Female	10000	0	2	uint8	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
City	10000	0	6072	object	False
Area	10000	0	3	object	False
ReAdmis	10000	0	2	object	False
Soft_drink	10000	0	2	object	False
HighBlood	10000	0	2	object	False
Stroke	10000	0	2	object	False
Overweight	10000	0	2	object	False
Arthritis	10000	0	2	object	False
Diabetes	10000	0	2	object	False
Hyperlipidemia	10000	0	2	object	False
BackPain	10000	0	2	object	False
Anxiety	10000	0	2	object	False
Allergic_rhinitis	10000	0	2	object	False
Reflux_esophagitis	10000	0	2	object	False
Asthma	10000	0	2	object	False

	Mean	Mode	Min
Gender_Nonbinary	0.0214	0	0
Services_MRI	0.038	0	0
Services_CT_Scan	0.1225	0	0
Population	9965.2538	0	0
vitD_supp	0.3989	0	0
Marital_Divorced	0.1961	0	0
Marital_Never_Married	0.1984	0	0
Marital_Separated	0.1987	0	0
Stroke_numeric	0.1993	0	0
Marital_Married	0.2023	0	0
Marital_Widowed	0.2045	0	0
Children	2.0972	0	0
Income	40490.49516	14572.4	154.08

Complication_risk_Low	0.2125	0	0
Initial_admin_Observation_Admission	0.2436	0	0
Initial_admin_Elective_Admission	0.2504	0	0
Soft_drink_numeric	0.2575	0	0
Diabetes_numeric	0.2738	0	0
Full_meals_eaten	1.0014	0	0
Asthma_numeric	0.2893	0	0
Additional_charges	12934.528587	3883.66416	3125.703
Services_Intravenous	0.313	0	0
Complication_risk_High	0.3358	0	0
Hyperlipidemia_numeric	0.3372	0	0
Arthritis_numeric	0.3574	0	0
ReAdmis_numeric	0.3669	0	0
Allergic_rhinitis_numeric	0.3941	0	0
HighBlood_numeric	0.409	0	0
BackPain_numeric	0.4114	0	0
Reflux_esophagitis_numeric	0.4135	0	0
Complication_risk_Medium	0.4517	0	0
Gender_Male	0.4768	0	0
Initial_days	34.455299	63.54432	1.001981
TotalCharge	5312.172769	7555.452	1938.312067
VitD_levels	17.964262	15.26009	9.806483
Age	53.5117	47	18
Gender_Female	0.5018	1	0
Doc_visits	5.0122	5	1
Initial_admin_Emergency_Admission	0.506	1	0
Services_Blood_Work	0.5265	1	0
City	-	-	-
Area	-	-	-
ReAdmis	-	-	-
Soft_drink	-	-	-
HighBlood	-	-	-
Stroke	-	-	-
Overweight	-	-	-
Arthritis	-	-	-
Diabetes	-	-	-
Hyperlipidemia	-	-	-
BackPain	-	-	-
Anxiety	-	-	-
Allergic_rhinitis	-	-	-
Reflux_esophagitis	-	-	-
Asthma	-	-	-

	Median	Max	Std	\
Gender_Nonbinary	0.0	1	0.144721	
Services_MRI	0.0	1	0.191206	
Services_CT_Scan	0.0	1	0.327879	
Population	2769.0	122814	14824.758614	
vitD_supp	0.0	5	0.628505	
Marital_Divorced	0.0	1	0.397065	
Marital_Never_Married	0.0	1	0.398815	
Marital_Separated	0.0	1	0.399042	
Stroke_numeric	0.0	1	0.399494	
Marital_Married	0.0	1	0.401735	
Marital_Widowed	0.0	1	0.403356	
Children	1.0	10	2.163659	
Income	33768.42	207249.1	28521.153293	
Complication_risk_Low	0.0	1	0.409097	
Initial_admin_Observation_Admission	0.0	1	0.429276	
Initial_admin_Elective_Admission	0.0	1	0.433265	
Soft_drink_numeric	0.0	1	0.437279	
Diabetes_numeric	0.0	1	0.44593	
Full_meals_eaten	1.0	7	1.008117	
Asthma_numeric	0.0	1	0.45346	
Additional_charges	11573.977735	30566.07	6542.601544	
Services_Intravenous	0.0	1	0.463738	
Complication_risk_High	0.0	1	0.472293	
Hyperlipidemia_numeric	0.0	1	0.472777	
Arthritis_numeric	0.0	1	0.479258	
ReAdmis_numeric	0.0	1	0.481983	
Allergic_rhinitis_numeric	0.0	1	0.488681	

HighBlood_numeric	0.0	1	0.491674
BackPain_numeric	0.0	1	0.492112
Reflux_esophagitis_numeric	0.0	1	0.492486
Complication_risk_Medium	0.0	1	0.497687
Gender_Male	0.0	1	0.499486
Initial_days	35.836244	71.98149	26.309341
TotalCharge	5213.952	9180.728	2180.393838
VitD_levels	17.951122	26.394449	2.017231
Age	53.0	89	20.638538
Gender_Female	1.0	1	0.500022
Doc_visits	5.0	9	1.045734
Initial_admin_Emergency_Admission	1.0	1	0.499989
Services_Blood_Work	1.0	1	0.499322
City	-	-	-
Area	-	-	-
ReAdmis	-	-	-
Soft_drink	-	-	-
HighBlood	-	-	-
Stroke	-	-	-
Overweight	-	-	-
Arthritis	-	-	-
Diabetes	-	-	-
Hyperlipidemia	-	-	-
BackPain	-	-	-
Anxiety	-	-	-
Allergic_rhinitis	-	-	-
Reflux_esophagitis	-	-	-
Asthma	-	-	-

	Skew	Kurt
Gender_Nonbinary	6.615434	41.772323
Services_MRI	4.833456	21.366572
Services_CT_Scan	2.303141	3.305119
Population	2.229959	5.880913
vitD_supp	1.550205	2.330763
Marital_Divorced	1.531038	0.344147
Marital_Never_Married	1.512784	0.288572
Marital_Separated	1.51042	0.281425
Stroke_numeric	1.505705	0.267202
Marital_Married	1.482369	0.197456
Marital_Widowed	1.4655	0.14772
Children	1.448013	2.076321
Income	1.405899	2.74569
Complication_risk_Low	1.405815	-0.023688
Initial_admin_Observation_Admission	1.19481	-0.572544
Initial_admin_Elective_Admission	1.152412	-0.672081
Soft_drink_numeric	1.109354	-0.769488
Diabetes_numeric	1.014712	-0.970553
Full_meals_eaten	1.009461	1.042727
Asthma_numeric	0.929485	-1.136285
Additional_charges	0.831842	-0.142684
Services_Intravenous	0.806652	-1.349583
Complication_risk_High	0.69547	-1.516625
Hyperlipidemia_numeric	0.688834	-1.525813
Arthritis_numeric	0.595206	-1.646059
ReAdmis_numeric	0.552412	-1.69518
Allergic_rhinitis_numeric	0.433498	-1.812442
HighBlood_numeric	0.370238	-1.863296
BackPain_numeric	0.360153	-1.870664
Reflux_esophagitis_numeric	0.35135	-1.876929
Complication_risk_Medium	0.194137	-1.962703
Gender_Male	0.092914	-1.991765
Initial_days	0.070286	-1.754525
TotalCharge	0.069661	-1.668267
VitD_levels	0.032435	-0.022112
Age	0.005117	-1.189527
Gender_Female	-0.007201	-2.000348
Doc_visits	-0.018563	0.025999
Initial_admin_Emergency_Admission	-0.024005	-1.999824
Services_Blood_Work	-0.106165	-1.989127
City	-	-

```
Area                    -       -
ReAdmis                 -       -
Soft_drink              -       -
HighBlood               -       -
Stroke                  -       -
Overweight              -       -
Arthritis               -       -
Diabetes                -       -
Hyperlipidemia          -       -
BackPain                -       -
Anxiety                 -       -
Allergic_rhinitis       -       -
Reflux_esophagitis      -       -
Asthma                  -       -
```

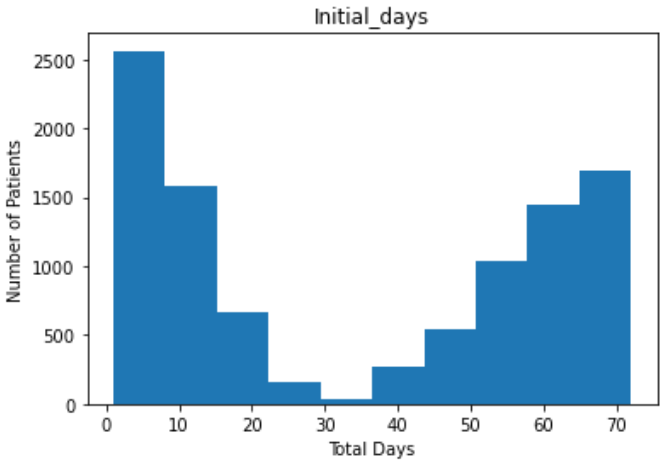
In [7]:

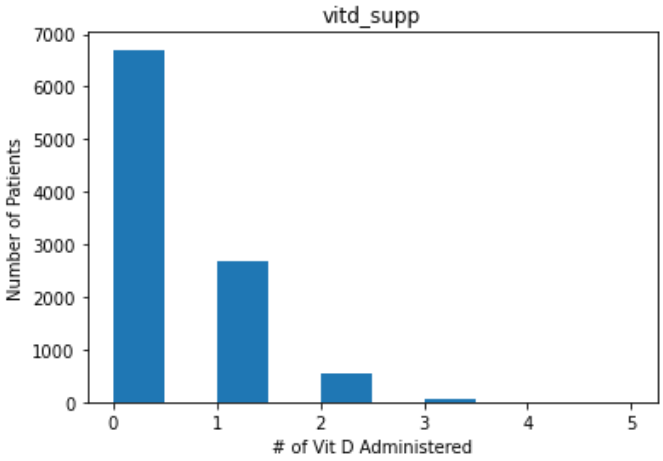
```
#Univariate Visualization
plt.hist(df.Initial_days)
plt.xlabel('Total Days')
plt.ylabel('Number of Patients')
plt.title('Initial_days')
plt.show()

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)
```

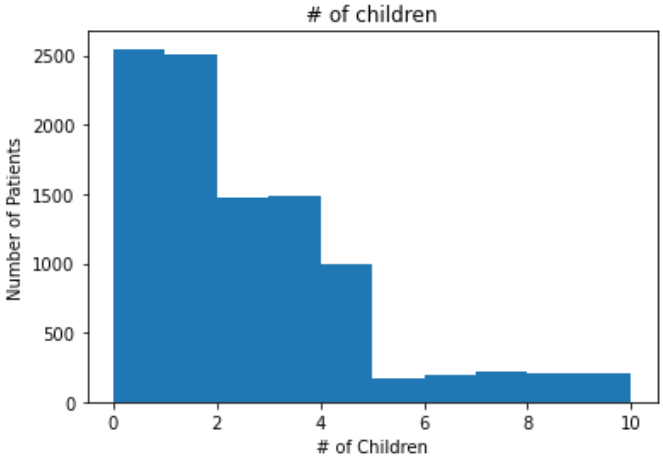
```
plt.xlabel('Doctor Visits')
plt.ylabel('Number of Patients')
plt.title('Doctor Visits')
plt.show()

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_Divorced)
plt.xlabel('Patients Marital Status')
plt.ylabel('Number of Patients')
plt.title('Marital_Divorced')
plt.show()

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

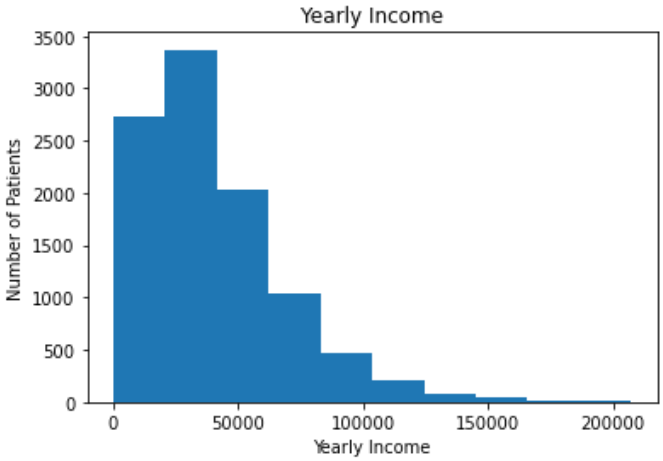
```
plt.hist(df.Marital_Never_Married)
plt.xlabel('Patients Marital Status')
plt.ylabel('Number of Patients')
plt.title('Marital_Never_Married')
plt.show()

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_Female)
plt.xlabel('What gender does the patient identify as?')
plt.ylabel('Number of Patients')
plt.title('Gender_Female')
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()
```

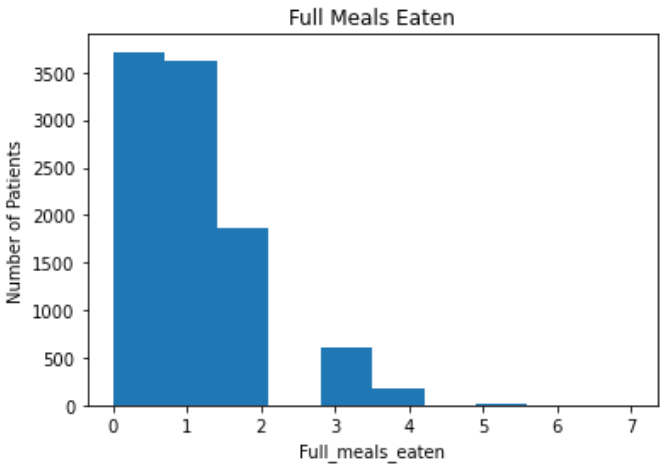
```
plt.hist(df.Initial_admin_Observation_Admission)
plt.xlabel('What brought the patient into the hospital?')
plt.ylabel('Number of Patients')
plt.title('Initial_admin_Observation_Admission')
plt.show()

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()
```

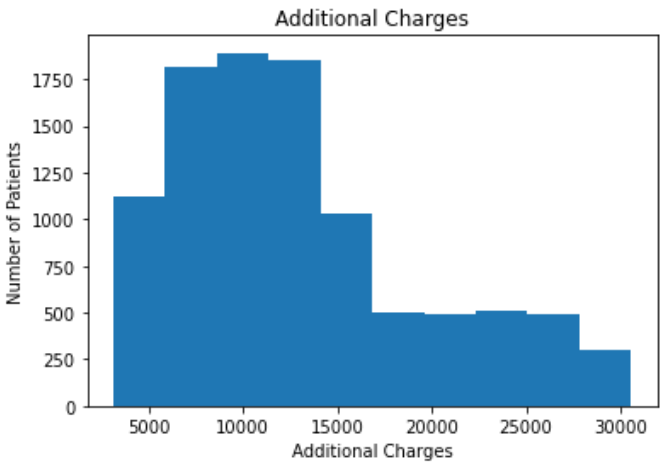


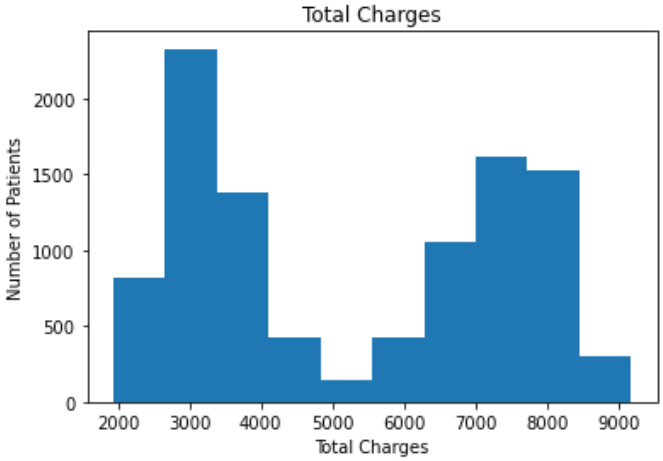


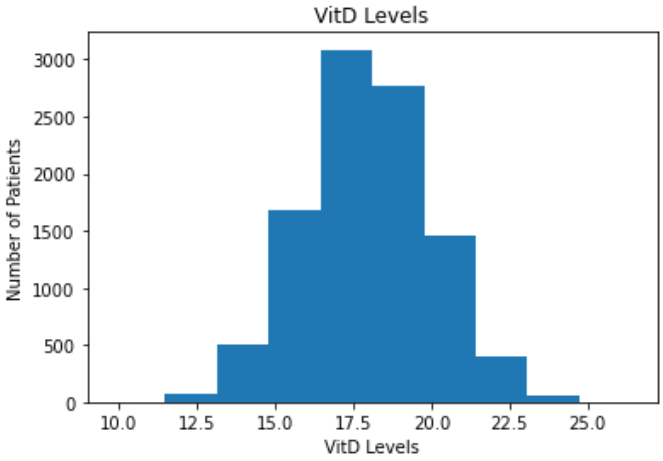

# of children
Number of Patients
# of Children

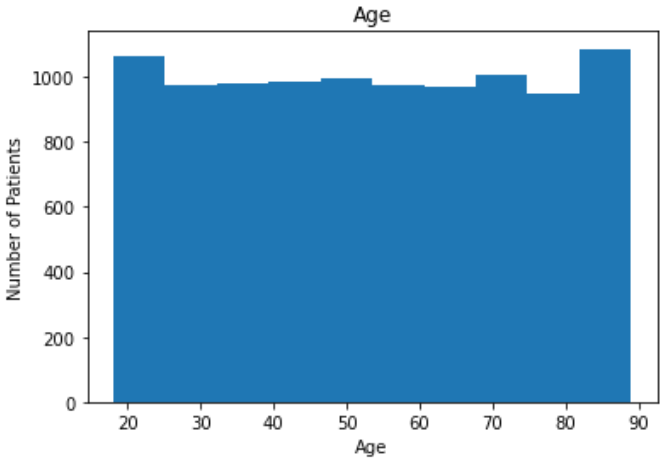

Yearly Income
Number of Patients
Yearly Income

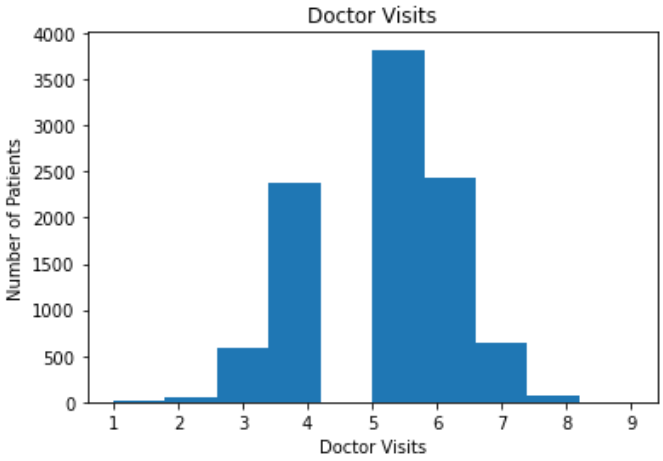

Full Meals Eaten
Number of Patients
Full_meals_eaten

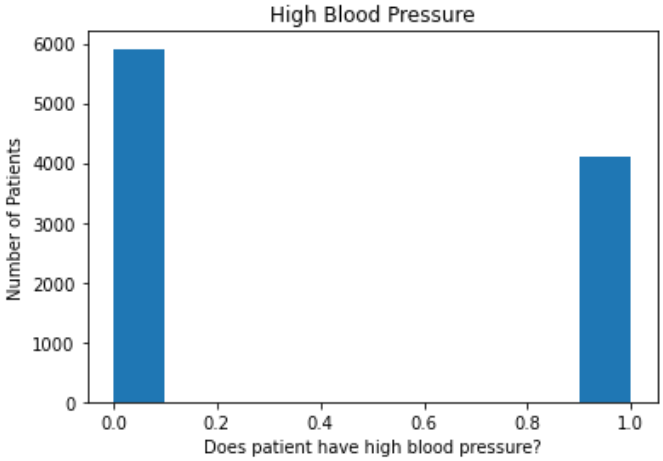

Additional Charges
Number of Patients
Additional Charges

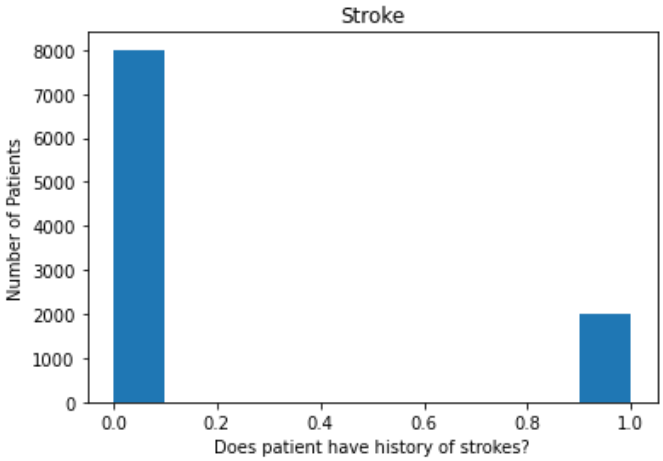

Total Charges
Number of Patients
Total Charges
0
500
1000
1500
2000
2000
3000
4000
5000
6000
7000
8000
9000

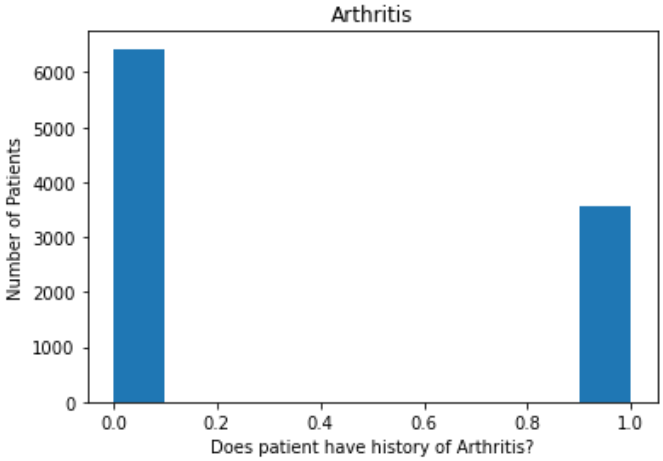

VitD Levels
Number of Patients
VitD Levels
0
500
1000
1500
2000
2500
3000
10.0
12.5
15.0
17.5
20.0
22.5
25.0

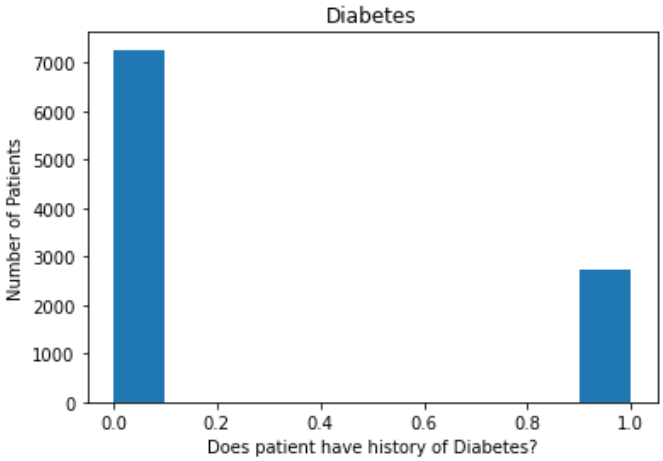

Age
Number of Patients
Age
0
200
400
600
800
1000
20
30
40
50
60
70
80
90

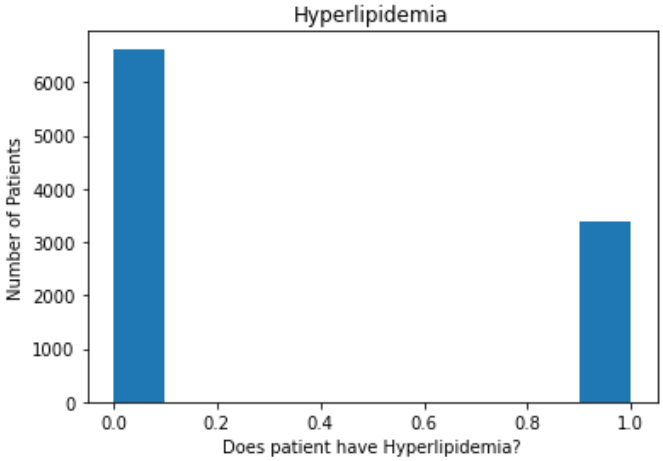

Doctor Visits
Number of Patients
Doctor Visits
0
500
1000
1500
2000
2500
3000
3500
4000
1
2
3
4
5
6
7
8
9

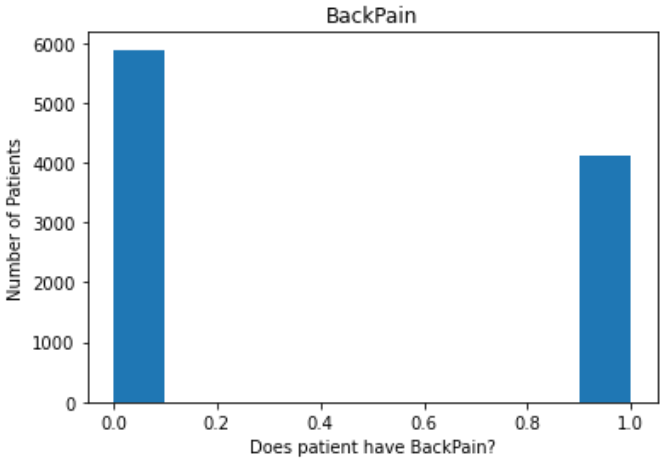

High Blood Pressure
Number of Patients
Does patient have high blood pressure?

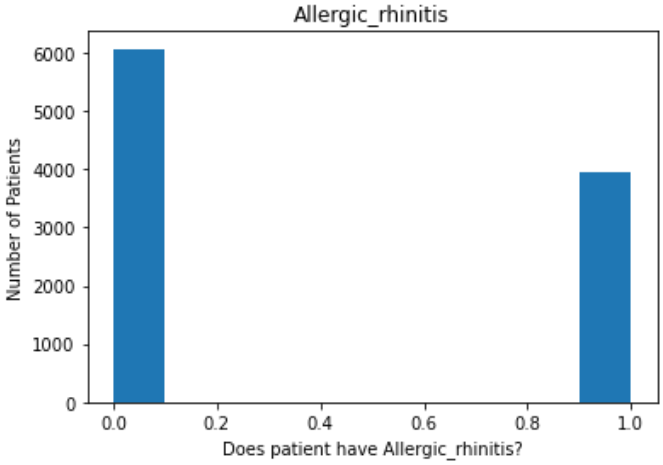

Stroke
Number of Patients
Does patient have history of strokes?

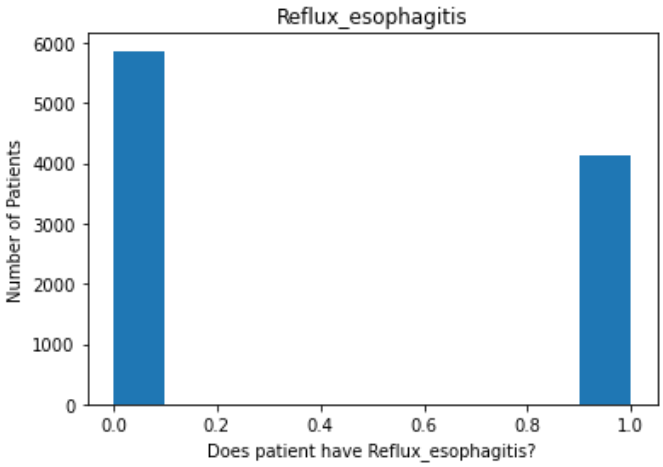

Arthritis
Number of Patients
Does patient have history of Arthritis?

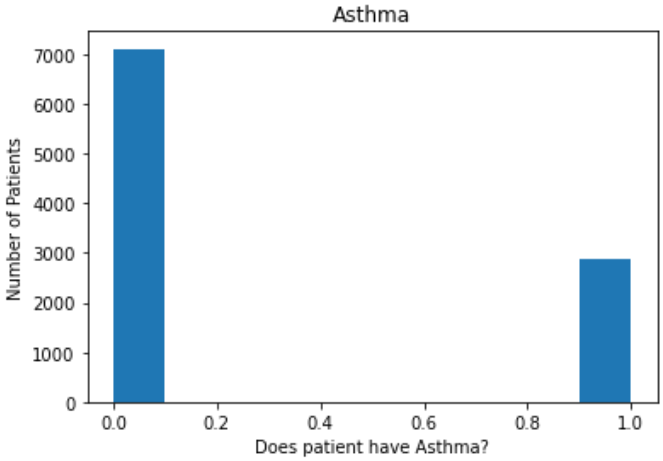

Diabetes
Number of Patients
Does patient have history of Diabetes?

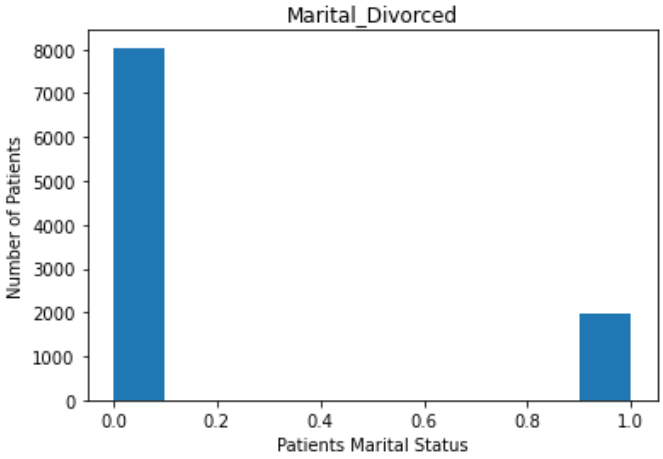

Hyperlipidemia
Number of Patients
0
1000
2000
3000
4000
5000
6000
0.0
0.2
0.4
0.6
0.8
1.0
Does patient have Hyperlipidemia?

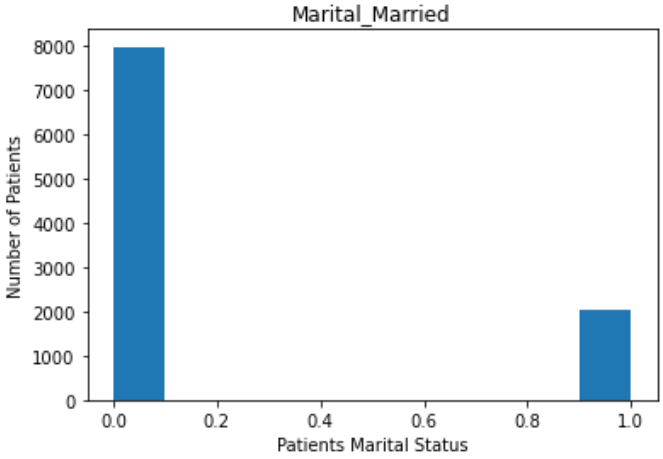

BackPain
Number of Patients
0
1000
2000
3000
4000
5000
6000
0.0
0.2
0.4
0.6
0.8
1.0
Does patient have BackPain?

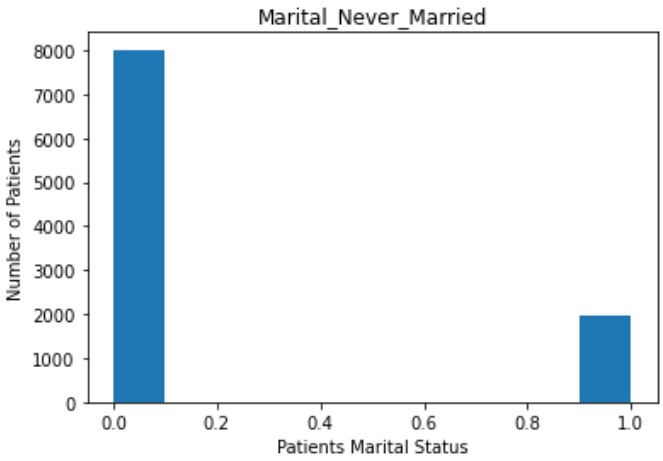

Allergic_rhinitis
Number of Patients
0
1000
2000
3000
4000
5000
6000
0.0
0.2
0.4
0.6
0.8
1.0
Does patient have Allergic_rhinitis?

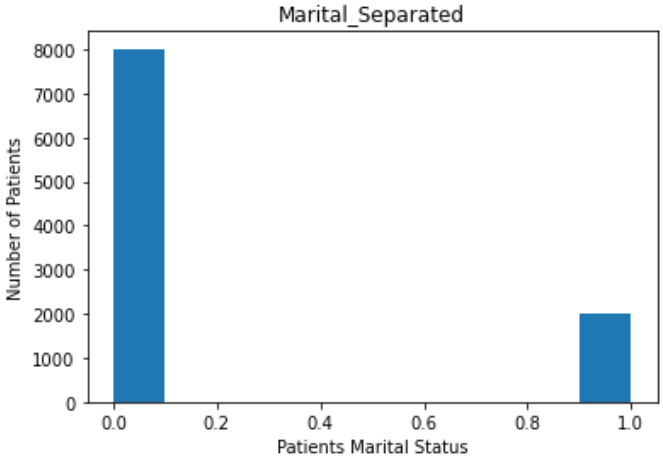

Reflux_esophagitis
Number of Patients
0
1000
2000
3000
4000
5000
6000
0.0
0.2
0.4
0.6
0.8
1.0
Does patient have Reflux_esophagitis?

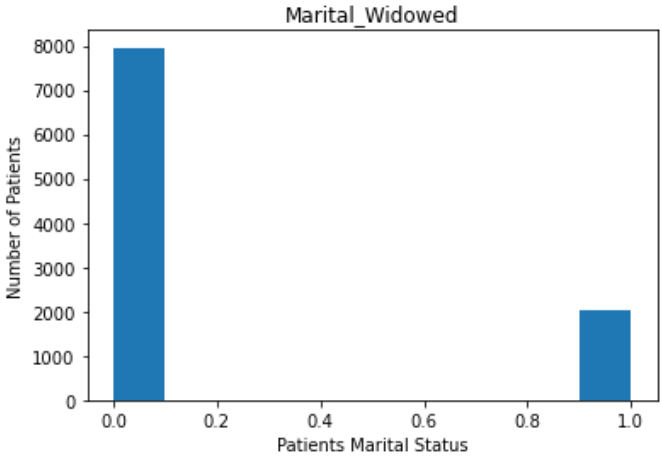

Asthma
Number of Patients
Does patient have Asthma?

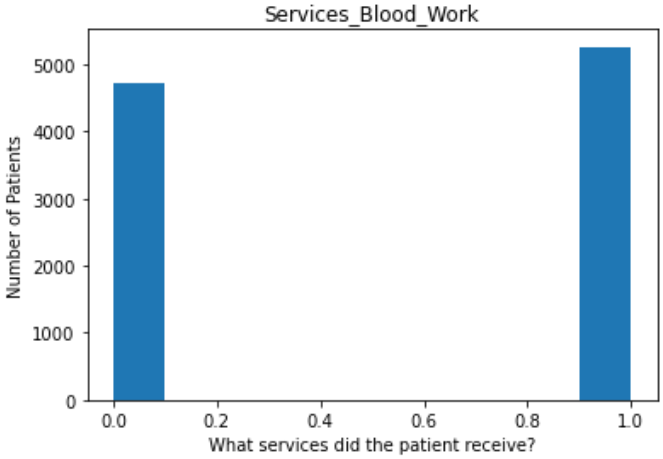

Marital_Divorced
Number of Patients
Patients Marital Status

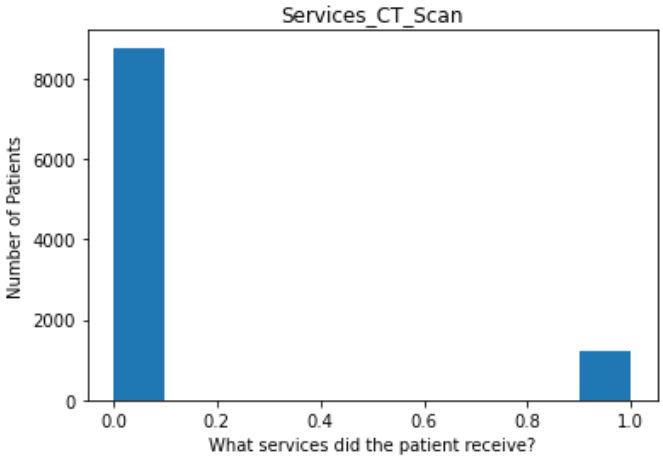

Marital_Married
Number of Patients
Patients Marital Status

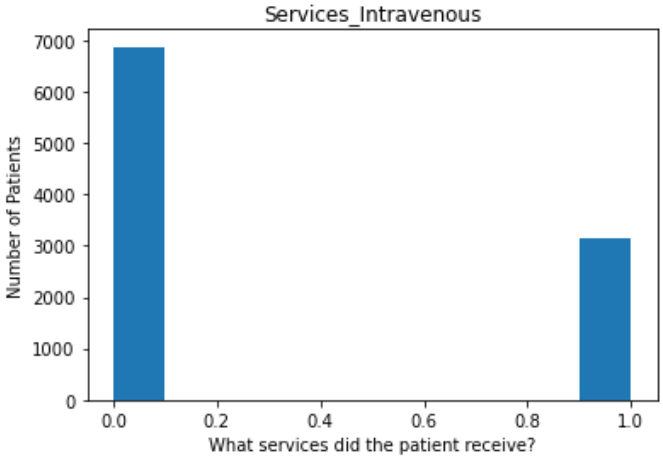

Marital_Never_Married
Number of Patients
Patients Marital Status

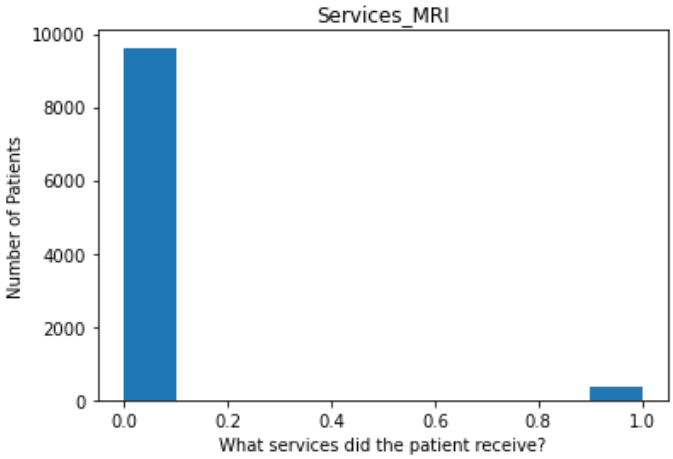

Marital_Separated
Number of Patients
Patients Marital Status
0
1000
2000
3000
4000
5000
6000
7000
8000
0.0
0.2
0.4
0.6
0.8
1.0

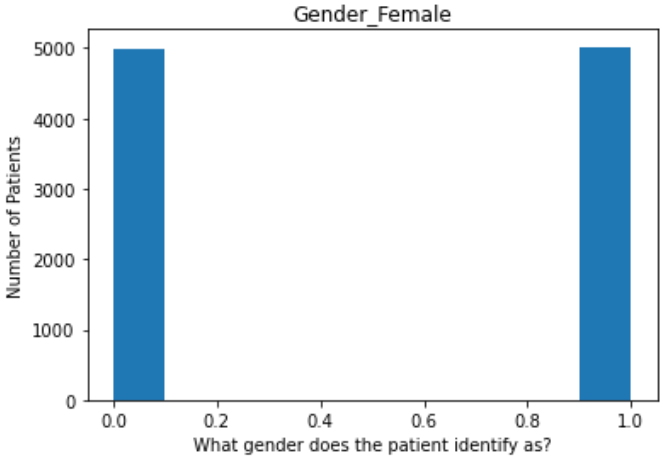

Marital_Widowed
Number of Patients
Patients Marital Status
0
1000
2000
3000
4000
5000
6000
7000
8000
0.0
0.2
0.4
0.6
0.8
1.0

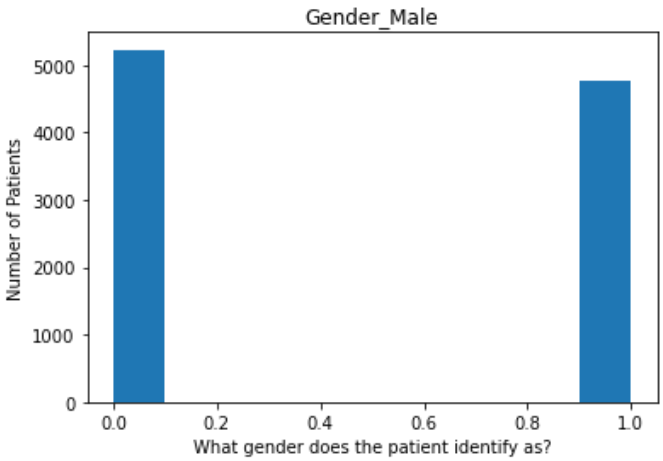

Services_Blood_Work
Number of Patients
What services did the patient receive?
0
1000
2000
3000
4000
5000
0.0
0.2
0.4
0.6
0.8
1.0

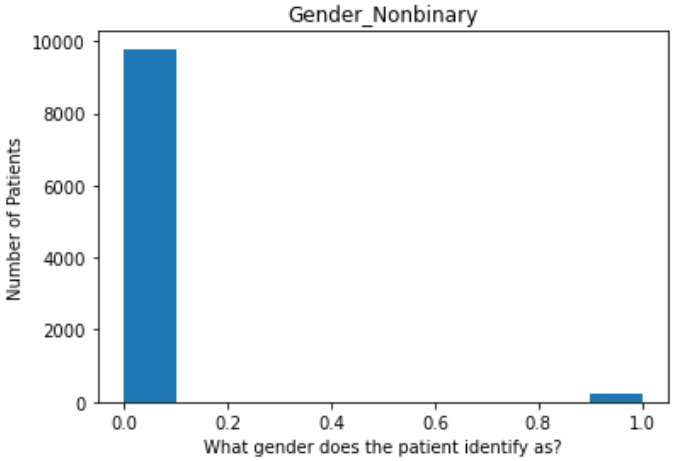

Services_CT_Scan
Number of Patients
What services did the patient receive?
0
2000
4000
6000
8000
0.0
0.2
0.4
0.6
0.8
1.0

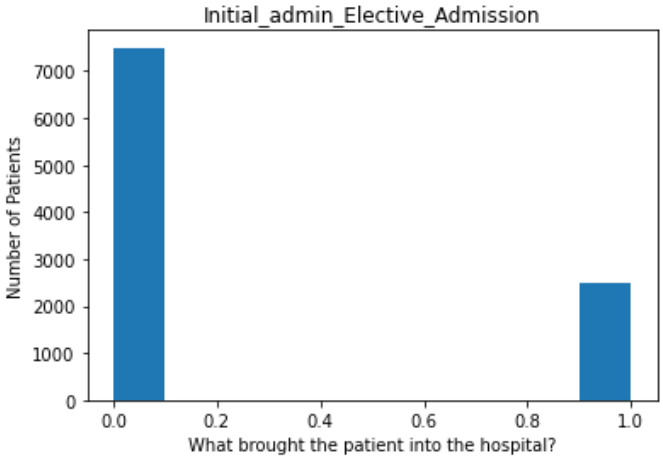

Services_Intravenous
Number of Patients
What services did the patient receive?

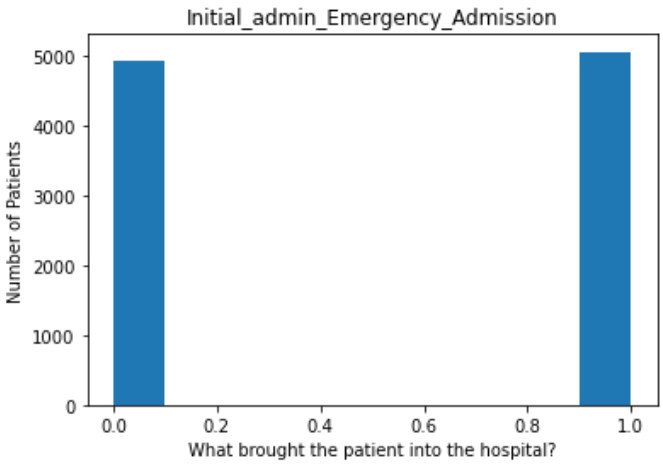

Services_MRI
Number of Patients
What services did the patient receive?

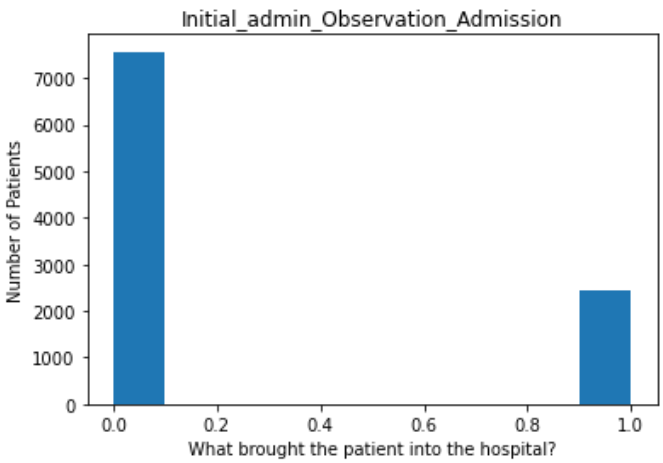

Gender_Female
Number of Patients
What gender does the patient identify as?

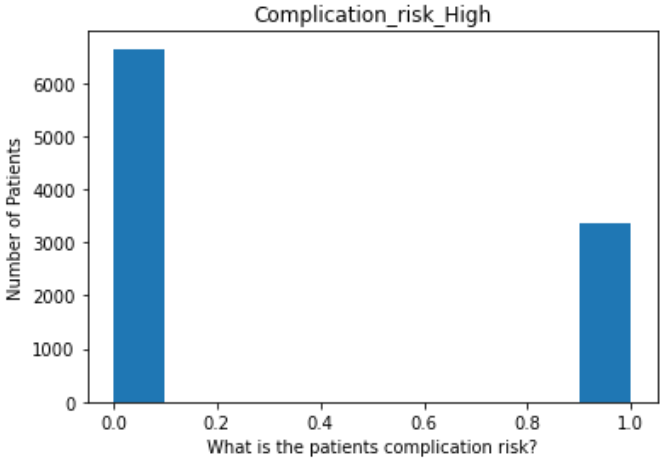

Gender_Male
Number of Patients
What gender does the patient identify as?

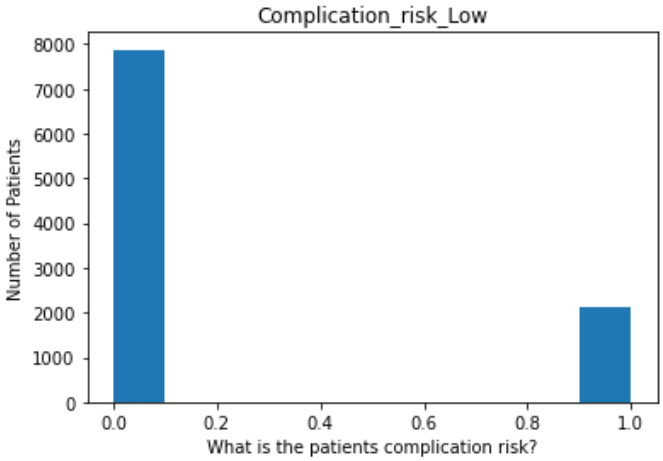

Gender_Nonbinary
Number of Patients
What gender does the patient identify as?

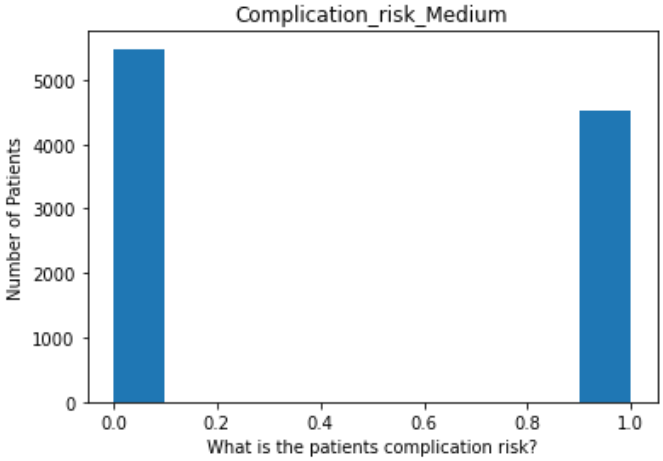

Initial_admin_Elective_Admission
Number of Patients
What brought the patient into the hospital?


Initial_admin_Emergency_Admission
Number of Patients
What brought the patient into the hospital?


Initial_admin_Observation_Admission
Number of Patients
What brought the patient into the hospital?



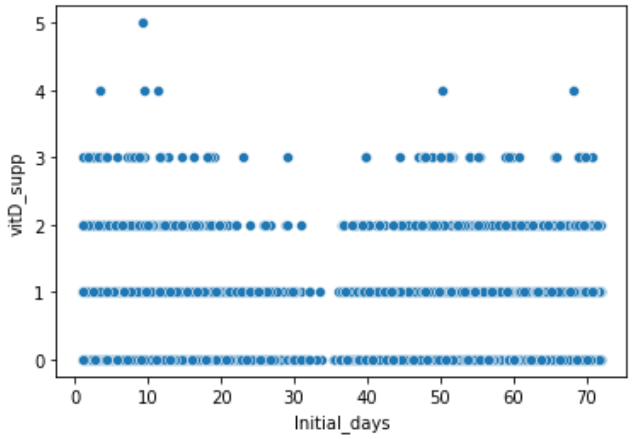


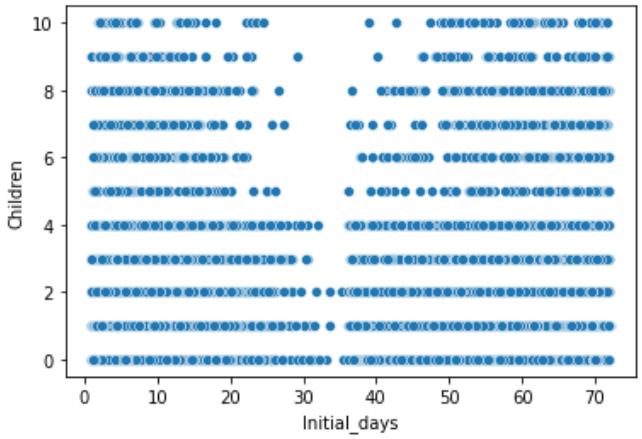


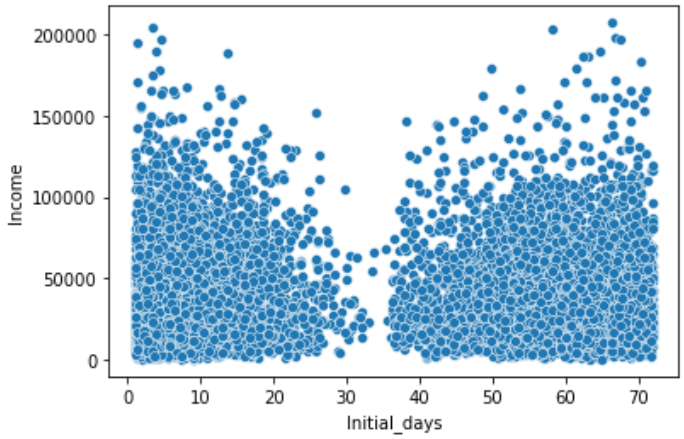
In [8]:
```
#Bivariate Visualization
sns.scatterplot(data=df, y="vitD_supp", x="Initial_days")
plt.show()

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

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

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

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

sns.scatterplot(data=df, y="TotalCharge", x="Initial_days")
```

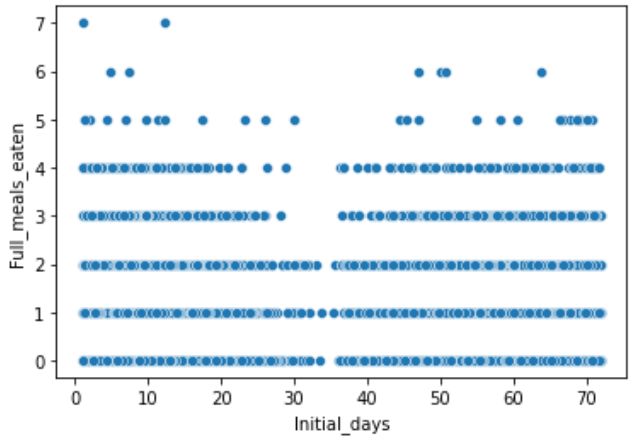
```
plt.show()

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

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

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

sns.scatterplot(data=df, y="HighBlood", x="Initial_days")
plt.show()

sns.scatterplot(data=df, y="Stroke", x="Initial_days")
plt.show()

sns.scatterplot(data=df, y="Arthritis", x="Initial_days")
plt.show()

sns.scatterplot(data=df, y="Diabetes", x="Initial_days")
plt.show()

sns.scatterplot(data=df, y="Hyperlipidemia", x="Initial_days")
plt.show()

sns.scatterplot(data=df, y="BackPain", x="Initial_days")
plt.show()

sns.scatterplot(data=df, y="Allergic_rhinitis", x="Initial_days")
plt.show()

sns.scatterplot(data=df, y="Reflux_esophagitis", x="Initial_days")
plt.show()

sns.scatterplot(data=df, y="Asthma", x="Initial_days")
plt.show()

sns.scatterplot(data=df, y="Marital_Divorced", x="Initial_days")
plt.show()

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

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

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

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

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

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

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

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

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

sns.scatterplot(data=df, y="Gender_Female", x="Initial_days")
plt.show()
```

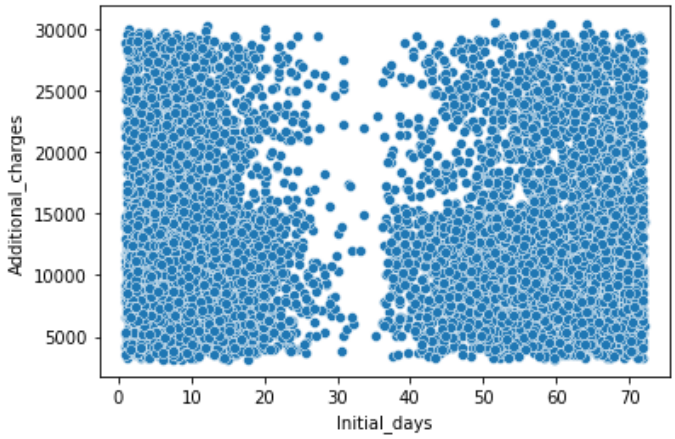
```
sns.scatterplot(data=df, y="Gender_Nonbinary", x="Initial_days")
plt.show()

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

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

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

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

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

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

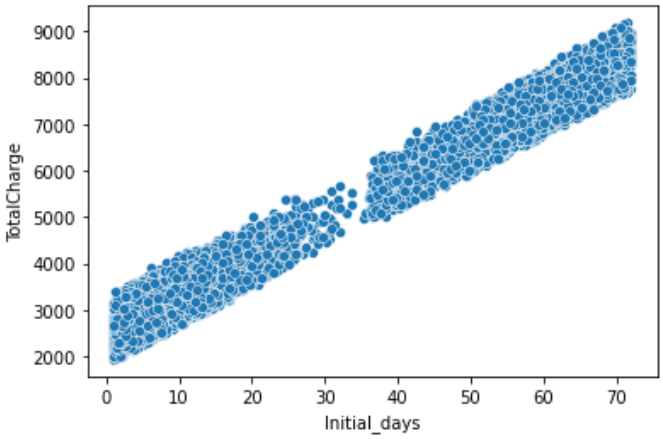


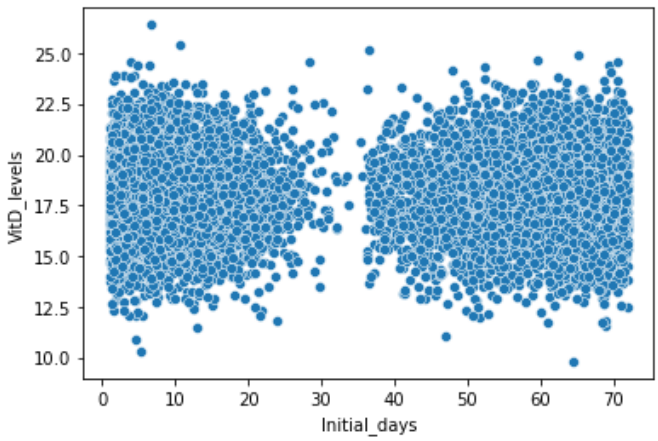


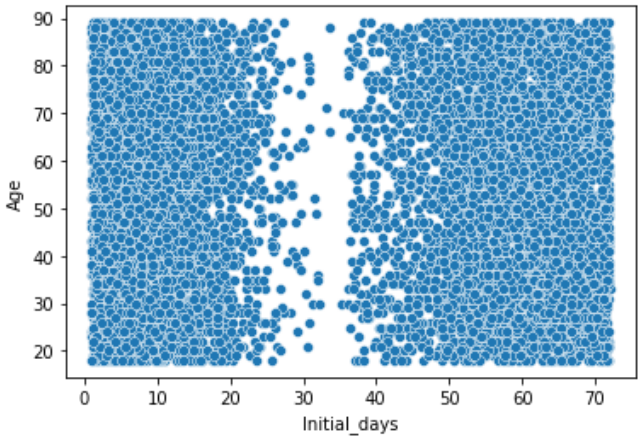


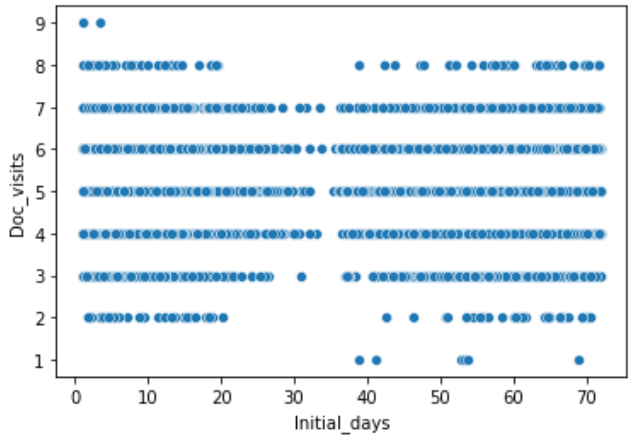

Full_meals_eaten
Initial_days

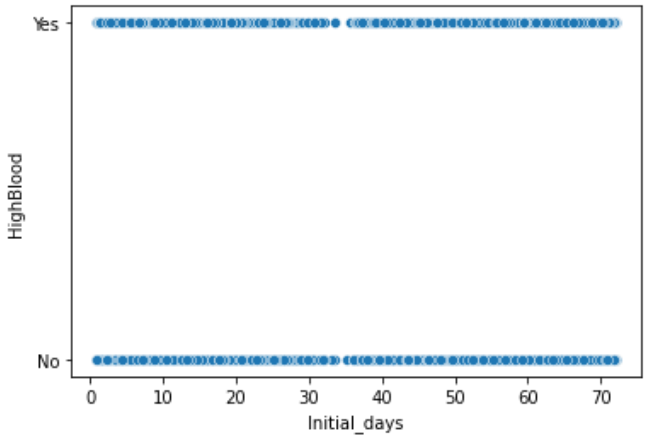

Additional_charges
Initial_days

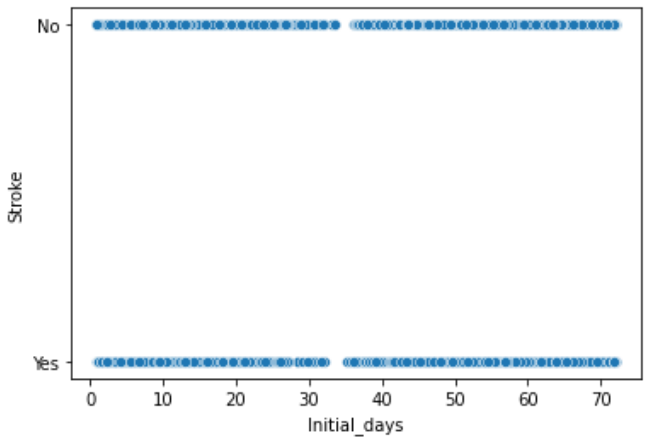

TotalCharge
Initial_days

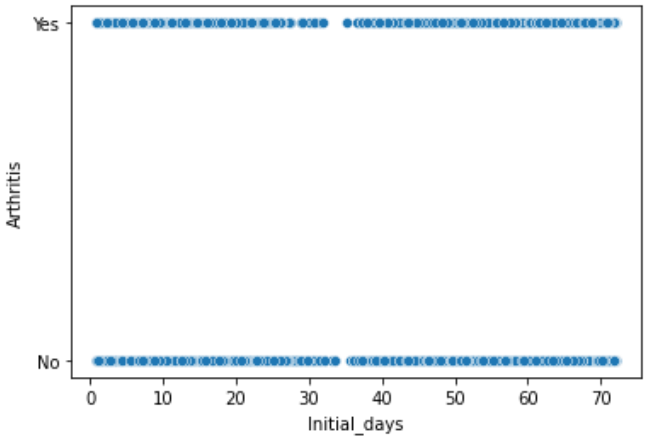

VitD_levels
Initial_days

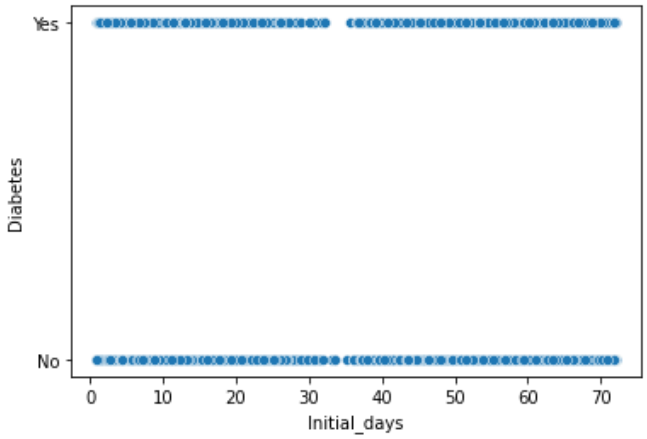

Age
Initial_days

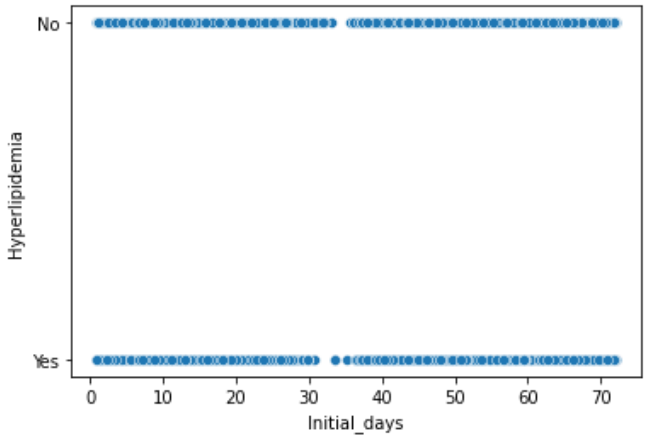

Doc_visits
Initial_days

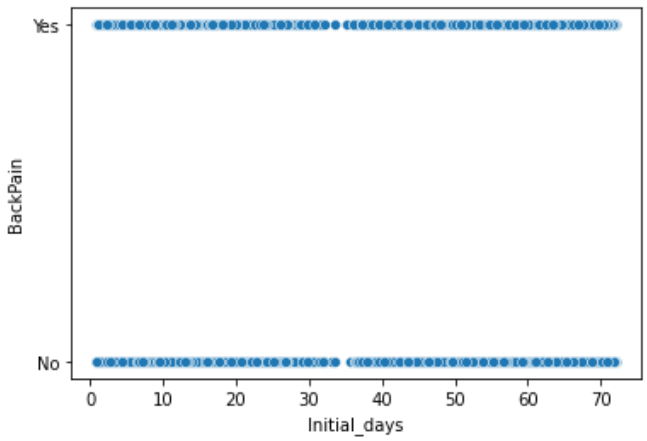

HighBlood
Yes
No
Initial_days

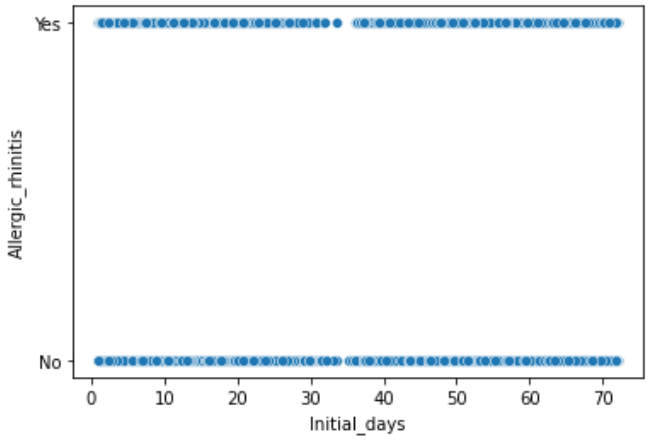

Stroke
No
Yes
Initial_days

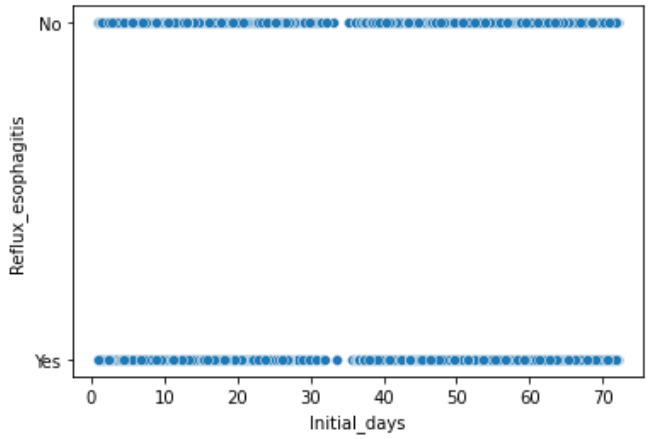

Yes
No
Arthritis
0
10
20
30
40
50
60
70
Initial_days

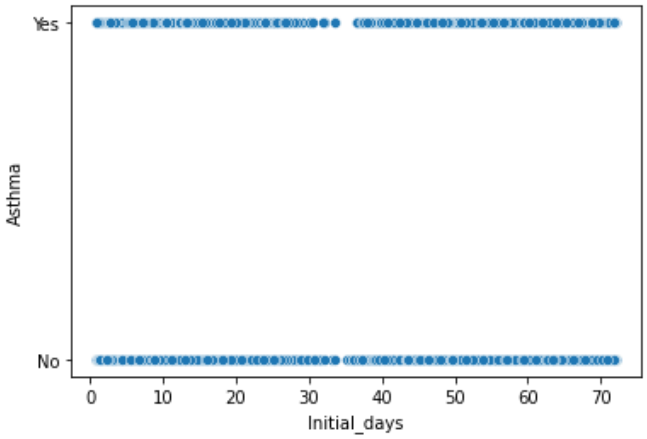

Yes
No
Diabetes
0
10
20
30
40
50
60
70
Initial_days

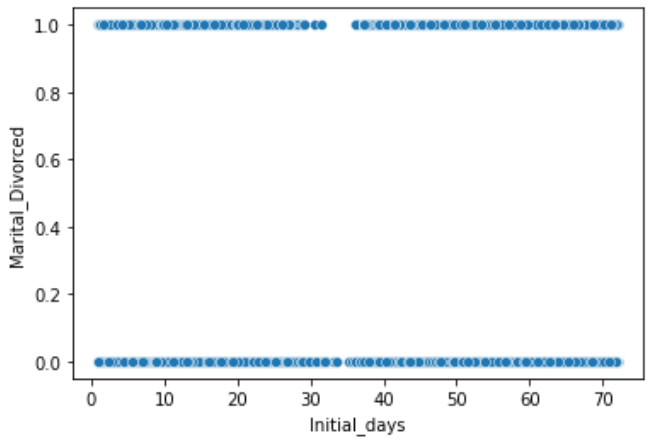

No
Yes
Hyperlipidemia
0
10
20
30
40
50
60
70
Initial_days

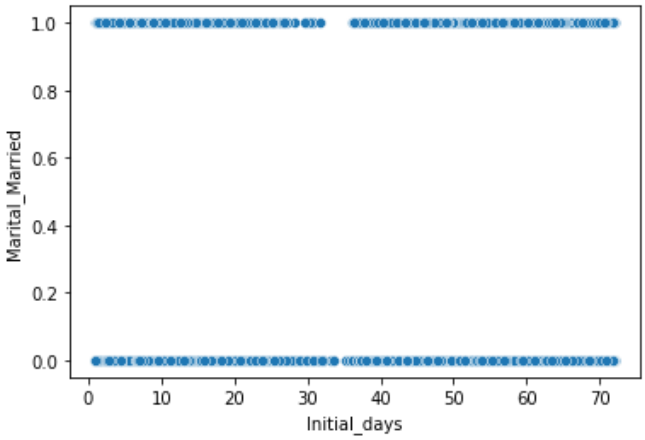

Yes
No
BackPain
0
10
20
30
40
50
60
70
Initial_days

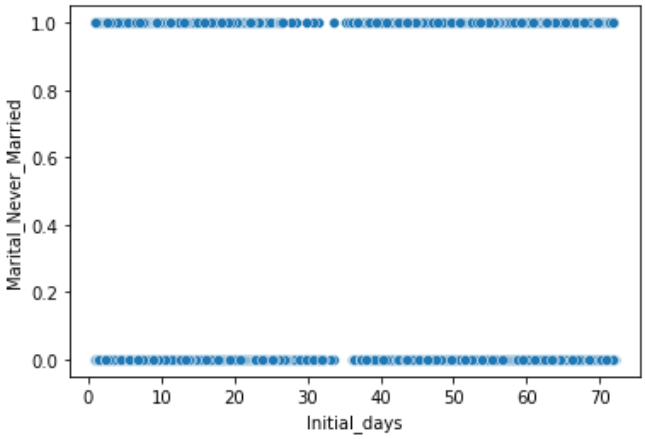

Yes
No
Allergic_rhinitis
0
10
20
30
40
50
60
70
Initial_days

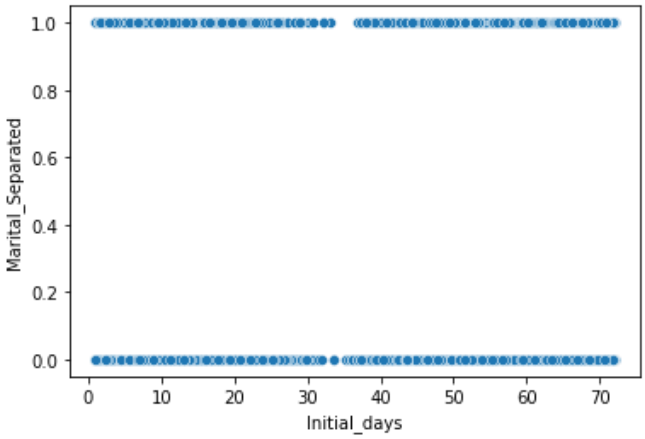

No
Yes
Reflux_esophagitis
0
10
20
30
40
50
60
70
Initial_days

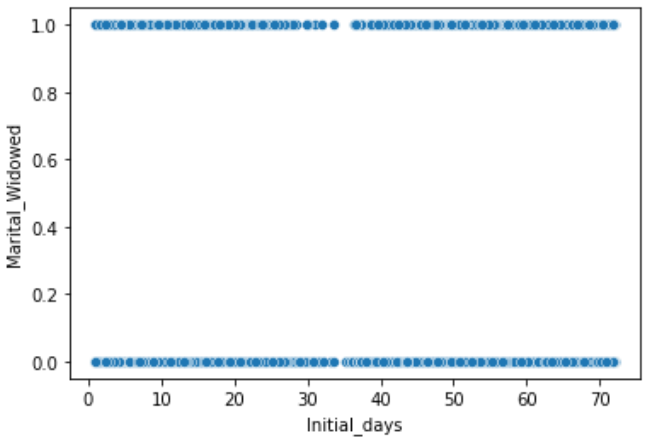

Yes
No
Asthma
0
10
20
30
40
50
60
70
Initial_days

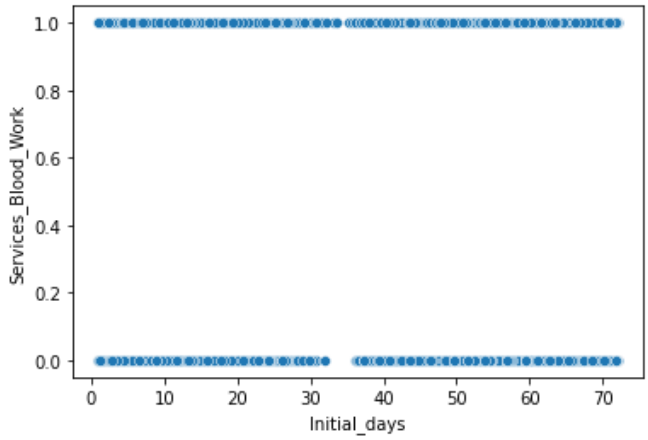

1.0
0.8
0.6
0.4
0.2
0.0
Marital_Divorced
0
10
20
30
40
50
60
70
Initial_days

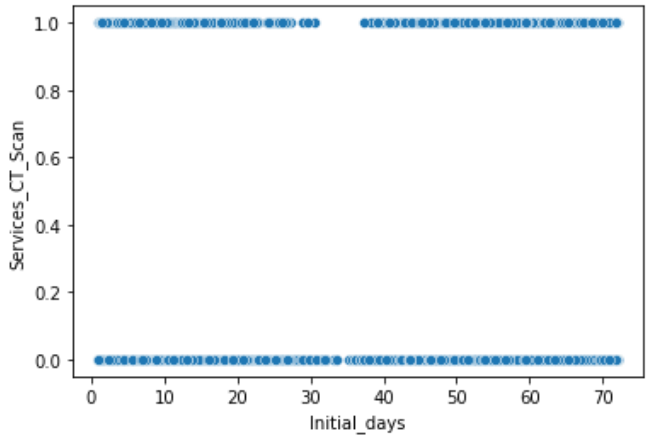

Marital_Married
Initial_days

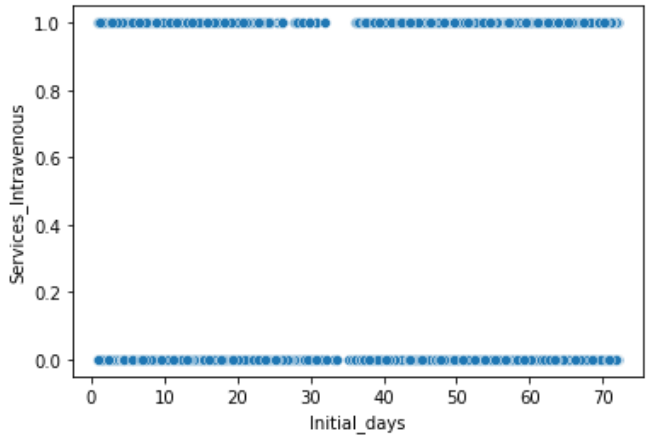

Marital_Never_Married
Initial_days

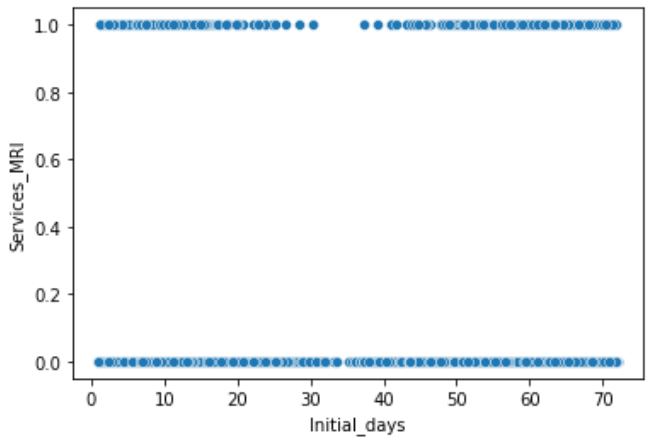

Marital_Separated
Initial_days

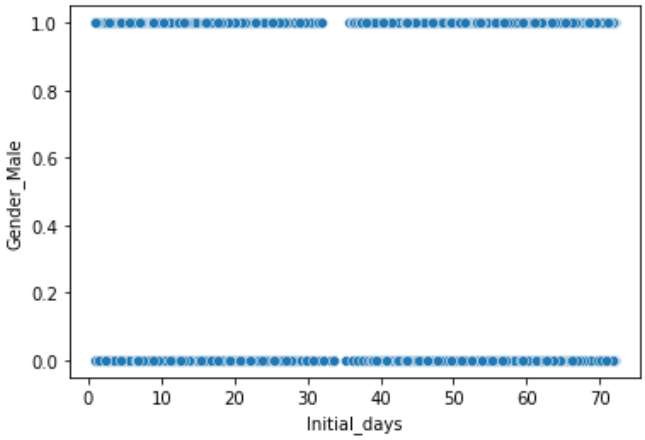

Marital_Widowed
Initial_days

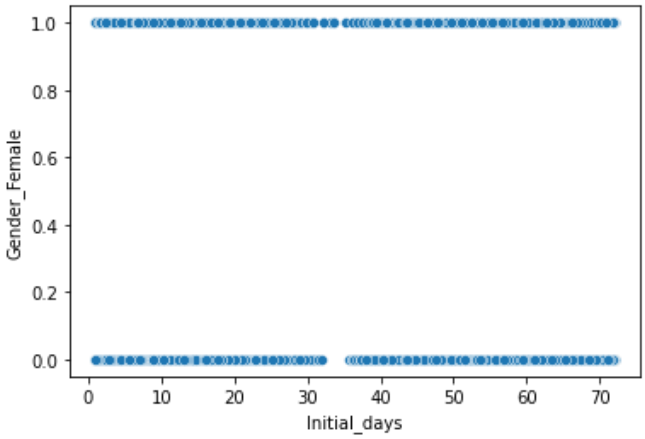

Services_Blood_Work
Initial_days

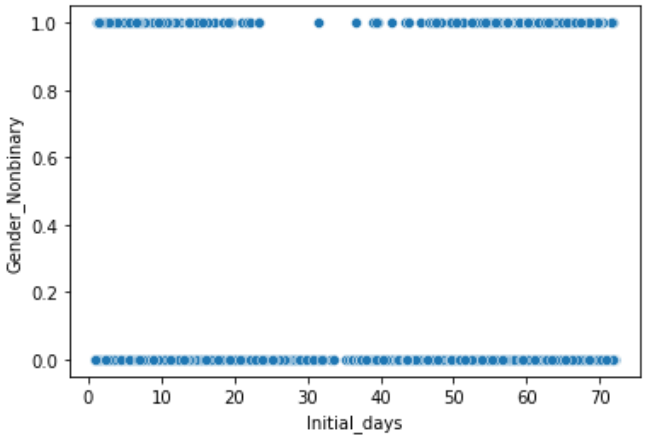

Services_CT_Scan
Initial_days

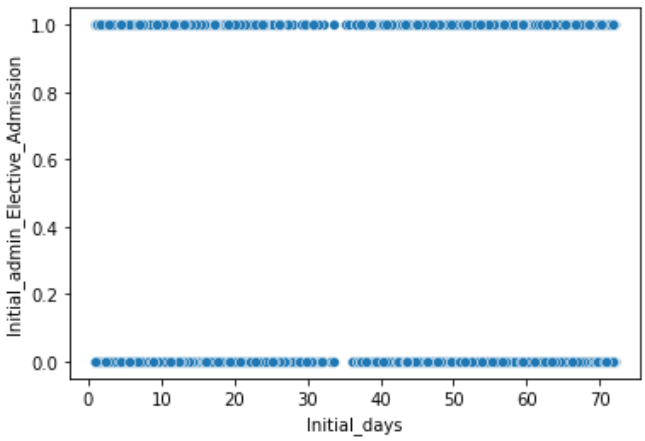

Services_Intravenous
Initial_days

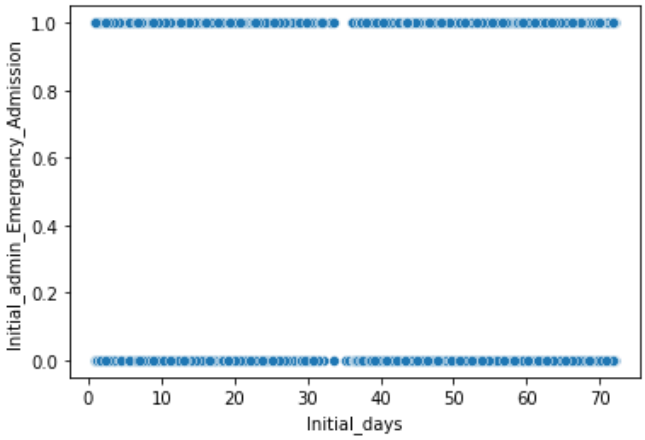

Services_MRI
Initial_days

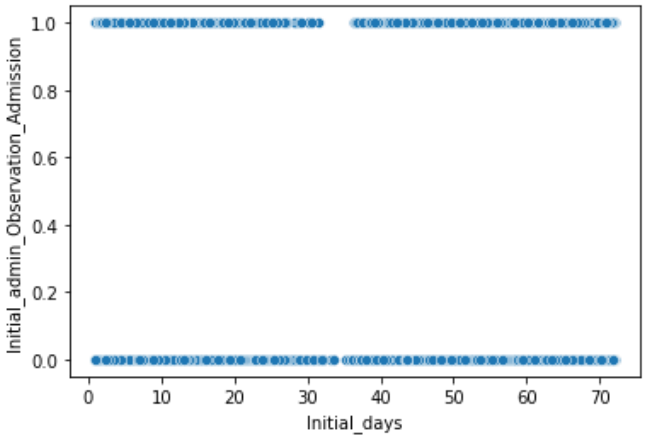

Gender_Male
Initial_days

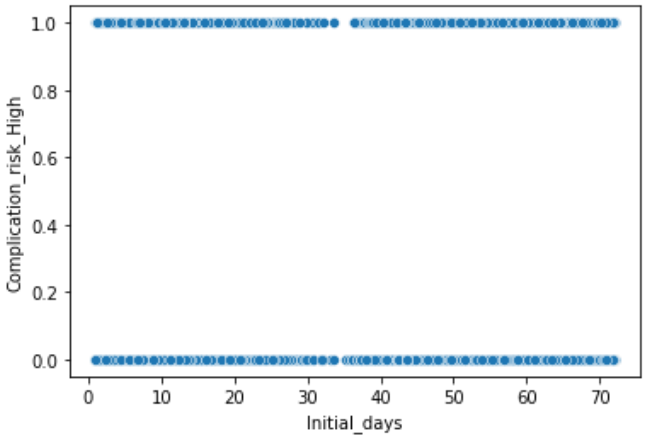

Gender_Female
Initial_days

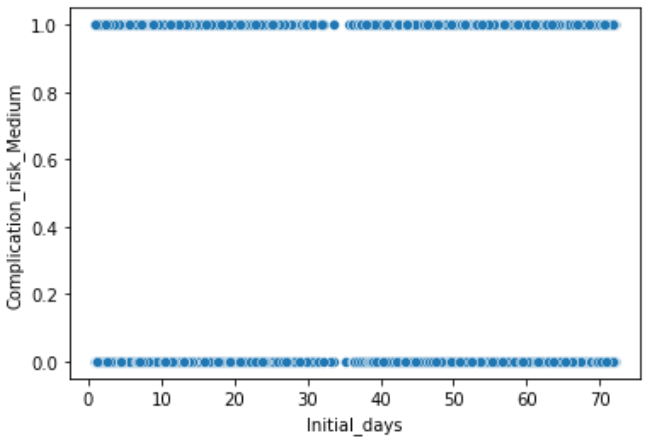

Gender_Nonbinary
Initial_days

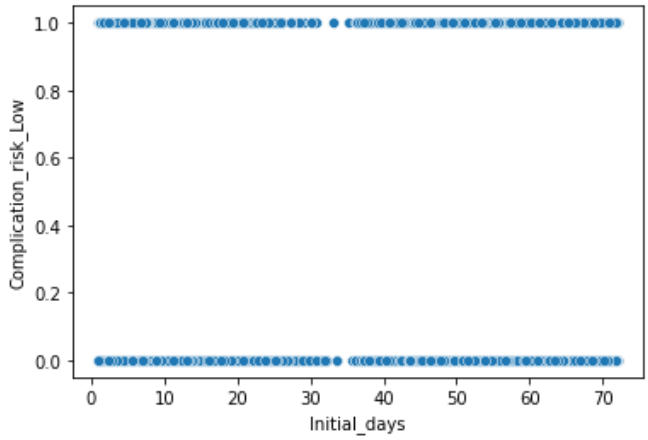

Initial_admin_Elective_Admission
Initial_days


Initial_admin_Emergency_Admission
Initial_days
0.0
0.2
0.4
0.6
0.8
1.0
0
10
20
30
40
50
60
70


Initial_admin_Observation_Admission
Initial_days
0.0
0.2
0.4
0.6
0.8
1.0
0
10
20
30
40
50
60
70


Complication_risk_High
Initial_days
0.0
0.2
0.4
0.6
0.8
1.0
0
10
20
30
40
50
60
70


Complication_risk_Medium
Initial_days
0.0
0.2
0.4
0.6
0.8
1.0
0
10
20
30
40
50
60
70



In [9]:
```
df.to_csv(r'C:\Users\mmorg\Desktop\D208 Assessment Files\Cleaned208data.csv')
```

In [10]:
```
mdl_initial_vs_variables = ols("Initial_days ~ vitD_supp + Children + Income + Full_meals_eaten + Additi

mdl_initial_vs_variables.summary()
```

Out[10]:

OLS Regression Results

Dep. Variable:	Initial_days	**R-squared:**	1.000
Model:	OLS	**Adj. R-squared:**	1.000
Method:	Least Squares	**F-statistic:**	9.258e+05
Date:	Thu, 24 Nov 2022	**Prob (F-statistic):**	0.00
Time:	21:10:32	**Log-Likelihood:**	-7060.1
No. Observations:	10000	**AIC:**	1.418e+04
Df Residuals:	9968	**BIC:**	1.441e+04
Df Model:	31		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	-10.8883	0.023	-471.390	0.000	-10.934	-10.843
vitD_supp	0.0001	0.008	0.019	0.985	-0.015	0.015
Children	-0.0016	0.002	-0.710	0.478	-0.006	0.003
Income	-9.605e-09	1.72e-07	-0.056	0.956	-3.48e-07	3.28e-07
Full_meals_eaten	-0.0042	0.005	-0.861	0.389	-0.014	0.005
Additional_charges	-2.634e-06	3.03e-06	-0.869	0.385	-8.57e-06	3.31e-06
TotalCharge	0.0122	2.28e-06	5345.969	0.000	0.012	0.012
VitD_levels	-0.0017	0.002	-0.702	0.482	-0.006	0.003
Age	0.0005	0.001	0.662	0.508	-0.001	0.002
Doc_visits	0.0009	0.005	0.191	0.848	-0.008	0.010
HighBlood_numeric	-1.3551	0.028	-48.387	0.000	-1.410	-1.300
Stroke_numeric	0.0181	0.012	1.463	0.144	-0.006	0.042
Arthritis_numeric	-0.8892	0.010	-86.563	0.000	-0.909	-0.869
Diabetes_numeric	-0.9146	0.011	-82.879	0.000	-0.936	-0.893
Hyperlipidemia_numeric	-1.1332	0.010	-108.907	0.000	-1.154	-1.113
BackPain_numeric	-1.0474	0.010	-104.629	0.000	-1.067	-1.028
Allergic_rhinitis_numeric	-0.7427	0.010	-73.806	0.000	-0.762	-0.723
Reflux_esophagitis_numeric	-0.7201	0.010	-72.092	0.000	-0.740	-0.701
Asthma_numeric	-0.0128	0.011	-1.179	0.238	-0.034	0.008
Marital_Divorced	-2.1805	0.011	-199.011	0.000	-2.202	-2.159
Marital_Married	-2.1820	0.011	-200.749	0.000	-2.203	-2.161
Marital_Never_Married	-2.1642	0.011	-199.295	0.000	-2.186	-2.143
Marital_Separated	-2.1874	0.011	-201.002	0.000	-2.209	-2.166
Marital_Widowed	-2.1742	0.011	-201.888	0.000	-2.195	-2.153
Services_Blood_Work	-2.7271	0.010	-264.686	0.000	-2.747	-2.707
Services_CT_Scan	-2.7189	0.014	-195.632	0.000	-2.746	-2.692
Services_Intravenous	-2.7327	0.011	-246.551	0.000	-2.754	-2.711
Services_MRI	-2.7095	0.021	-130.861	0.000	-2.750	-2.669
Gender_Female	-3.6287	0.013	-271.876	0.000	-3.655	-3.603
Gender_Male	-3.6210	0.013	-272.503	0.000	-3.647	-3.595

Gender_Nonbinary	-3.6386	0.025	-143.538	0.000	-3.688	-3.589
Initial_admin_Elective_Admission	-1.5435	0.011	-141.986	0.000	-1.565	-1.522
Initial_admin_Emergency_Admission	-7.8003	0.010	-769.676	0.000	-7.820	-7.780
Initial_admin_Observation_Admission	-1.5445	0.011	-139.863	0.000	-1.566	-1.523
Complication_risk_High	-6.9899	0.011	-663.987	0.000	-7.011	-6.969
Complication_risk_Low	-1.9465	0.011	-172.527	0.000	-1.969	-1.924
Complication_risk_Medium	-1.9519	0.010	-194.182	0.000	-1.972	-1.932

Omnibus:	90391.781	**Durbin-Watson:**	2.015
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	1792.347
Skew:	-0.761	**Prob(JB):**	0.00
Kurtosis:	1.592	**Cond. No.**	2.91e+17

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The smallest eigenvalue is 3.06e-22. This might indicate that there are
strong multicollinearity problems or that the design matrix is singular.

In [1]:
```
# 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)
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
Input In [1], in <cell line: 4>()
      1 # Checking for the VIF values of the variables.
      2 from statsmodels.stats.outliers_influence import variance_inflation_factor
----> 4 X = df[['Initial_days', 'vitD_supp', 'Children', 'Income', 'Full_meals_eaten', 'Additional_charg
es', 'TotalCharge', 'VitD_levels', 'Age', 'Doc_visits', 'HighBlood_numeric', 'Stroke_numeric', 'Arthriti
s_numeric', 'Diabetes_numeric', 'Hyperlipidemia_numeric', 'BackPain_numeric', 'Allergic_rhinitis_numeri
c', 'Reflux_esophagitis_numeric', 'Asthma_numeric', 'Marital_Married', 'Marital_Never_Married', 'Marital
_Separated', 'Marital_Widowed', 'Services_Blood_Work', 'Services_CT_Scan', 'Services_Intravenous', 'Serv
ices_MRI', 'Gender_Male', 'Gender_Nonbinary', 'Initial_admin_Elective_Admission', 'Initial_admin_Emergen
cy_Admission', 'Initial_admin_Observation_Admission', 'Complication_risk_High', 'Complication_risk_Low',
'Complication_risk_Medium']]
      7 # VIF dataframe
      8 vif_data = pd.DataFrame()

NameError: name 'df' is not defined
```

In [13]:
```
#Reduced model, dropping variables with infinite VIF
mdl_initial_vs_variables = ols("Initial_days ~ vitD_supp + Children + Income + Full_meals_eaten + Additi

mdl_initial_vs_variables.summary()
```

Out[13]:

OLS Regression Results

Dep. Variable:	Initial_days	**R-squared:**	0.003
Model:	OLS	**Adj. R-squared:**	0.001
Method:	Least Squares	**F-statistic:**	1.609
Date:	Mon, 21 Nov 2022	**Prob (F-statistic):**	0.0533
Time:	12:09:14	**Log-Likelihood:**	-46874.
No. Observations:	10000	**AIC:**	9.378e+04
Df Residuals:	9982	**BIC:**	9.391e+04
Df Model:	17		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	34.0439	2.885	11.800	0.000	28.389	39.699
vitD_supp	0.6559	0.419	1.566	0.117	-0.165	1.477
Children	0.2781	0.122	2.286	0.022	0.040	0.517
Income	-1.179e-05	9.23e-06	-1.278	0.201	-2.99e-05	6.3e-06
Full_meals_eaten	-0.4394	0.261	-1.683	0.093	-0.951	0.073
Additional_charges	-0.0002	0.000	-1.159	0.246	-0.000	0.000
VitD_levels	-0.0441	0.131	-0.338	0.735	-0.300	0.212
Age	0.0618	0.038	1.624	0.104	-0.013	0.136
Doc_visits	-0.1814	0.252	-0.720	0.471	-0.675	0.312
HighBlood_numeric	1.2442	1.472	0.845	0.398	-1.642	4.130
Stroke_numeric	-0.0671	0.661	-0.102	0.919	-1.363	1.229
Arthritis_numeric	1.0291	0.549	1.874	0.061	-0.048	2.106
Diabetes_numeric	-0.1275	0.591	-0.216	0.829	-1.285	1.030
Hyperlipidemia_numeric	-0.2355	0.557	-0.423	0.672	-1.327	0.856
BackPain_numeric	0.9249	0.535	1.729	0.084	-0.124	1.974
Allergic_rhinitis_numeric	0.2122	0.538	0.394	0.694	-0.843	1.268
Reflux_esophagitis_numeric	0.6491	0.534	1.215	0.225	-0.398	1.697
Asthma_numeric	-0.7605	0.580	-1.310	0.190	-1.898	0.377

Omnibus:	41465.433	**Durbin-Watson:**	0.164
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	1277.292
Skew:	0.070	**Prob(JB):**	4.36e-278
Kurtosis:	1.255	**Cond. No.**	5.60e+05

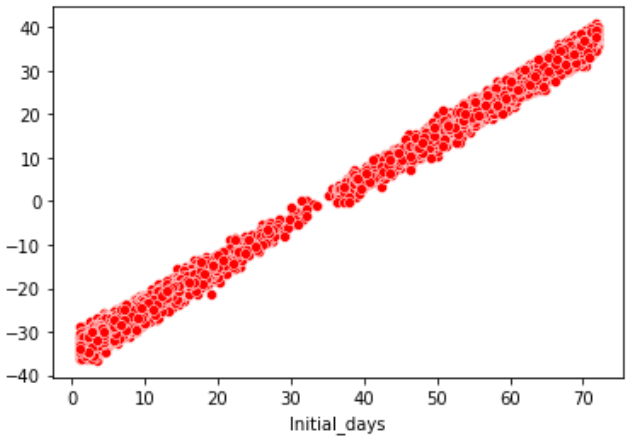
Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 5.6e+05. This might indicate that there are
strong multicollinearity or other numerical problems.

In [14]:
```
#Residual Standard Error
np.sqrt(mdl_initial_vs_variables.scale)
```

Out[14]:
```
26.295731213305718
```

In [15]:
```
#Residual plot
df['intercept'] = 1
residuals = df['Initial_days'] - mdl_initial_vs_variables.predict(df[['vitD_supp', 'Children', 'Income',
sns.scatterplot(x=df['Initial_days'], y=residuals, color='red')
```

Out[15]:
```
<AxesSubplot:xlabel='Initial_days'>
```



In [ ]: