In [1]:
```
import pandas as pd

import numpy as np

import missingno as msno

import matplotlib.pyplot as plt

from pandas import DataFrame

import scipy.stats as stats

import seaborn

from sklearn.decomposition import PCA

#Loading the CSV of the default dataset
df = pd.read_csv(r'C:\Users\mmorg\WGU\D206\D206 Assessment Data\medical_raw_data.csv')

#Object or string is qualitative, int64 or float64 is quantitative
df.info()

#Detect duplicate rows
print(df.duplicated())

#Detect null values
print(df.isnull().sum())

#Missing data matrix showing where null values exist
msno.matrix(df, labels=True)
plt.title('Missing Data Matrix')
plt.show()

#Detection of outliers for quantitative values
boxplot=seaborn.boxplot(x='Children',data=df)
plt.show()
boxplot=seaborn.boxplot(x='Age',data=df)
plt.show()
boxplot=seaborn.boxplot(x='Income',data=df)
plt.show()
boxplot=seaborn.boxplot(x='VitD_levels',data=df)
plt.show()
boxplot=seaborn.boxplot(x='Doc_visits',data=df)
plt.show()
boxplot=seaborn.boxplot(x='Full_meals_eaten',data=df)
plt.show()
boxplot=seaborn.boxplot(x='VitD_supp',data=df)
plt.show()
boxplot=seaborn.boxplot(x='Overweight',data=df)
plt.show()
boxplot=seaborn.boxplot(x='Anxiety',data=df)
plt.show()
boxplot=seaborn.boxplot(x='Initial_days',data=df)
plt.show()
boxplot=seaborn.boxplot(x='TotalCharge',data=df)
plt.show()
boxplot=seaborn.boxplot(x='Additional_charges',data=df)
plt.show()
boxplot=seaborn.boxplot(x='Item1',data=df)
plt.show()
boxplot=seaborn.boxplot(x='Item2',data=df)
plt.show()
boxplot=seaborn.boxplot(x='Item3',data=df)
plt.show()
boxplot=seaborn.boxplot(x='Item4',data=df)
plt.show()
boxplot=seaborn.boxplot(x='Item5',data=df)
```

```
plt.show()
boxplot=seaborn.boxplot(x='Item6',data=df)
plt.show()
boxplot=seaborn.boxplot(x='Item7',data=df)
plt.show()
boxplot=seaborn.boxplot(x='Item8',data=df)
plt.show()

#Visualizing distribution shapes
print('Children Original: ')
plt.hist(df['Children'])
plt.show()
print('Age Original: ')
plt.hist(df['Age'])
plt.show()
print('Income Original: ')
plt.hist(df['Income'])
plt.show()
print('Overweight Original: ')
plt.hist(df['Overweight'])
plt.show()
print('Anxiety Original: ')
plt.hist(df['Anxiety'])
plt.show()
print('Initial_days Original: ')
plt.hist(df['Initial_days'])
plt.show()

#Fixing missing data based on distribution shapes
df['Children'].fillna(method='ffill', inplace=True)
df['Age'].fillna(method='ffill', inplace=True)
df['Income'].fillna(df['Income'].median(), inplace=True)
df['Soft_drink'].fillna("No", inplace=True)
df['Overweight'].fillna(0, inplace=True)
df['Anxiety'].fillna(0, inplace=True)
df['Initial_days'].fillna(method='ffill', inplace=True)

#Checking to make sure all missing data has been replaced with selected values
print(df.isnull().sum())

#Visualizing distribution shapes after changes
print('Children Modified: ')
plt.hist(df['Children'])
plt.show()
print('Age Modified: ')
plt.hist(df['Age'])
plt.show()
print('Income Modified: ')
plt.hist(df['Income'])
plt.show()
print('Overweight Modified: ')
plt.hist(df['Overweight'])
plt.show()
print('Anxiety Modified: ')
plt.hist(df['Anxiety'])
plt.show()
print('Initial_days Modified: ')
plt.hist(df['Initial_days'])
plt.show()

#Checking Z-scores because why not
##df['ChildrenZScore']=stats.zscore(df['Children'])
##df[['Children','ChildrenZScore']].head
##plt.hist(df['ChildrenZScore'])
##plt.show()
##
##df['AgeZScore']=stats.zscore(df['Age'])
```

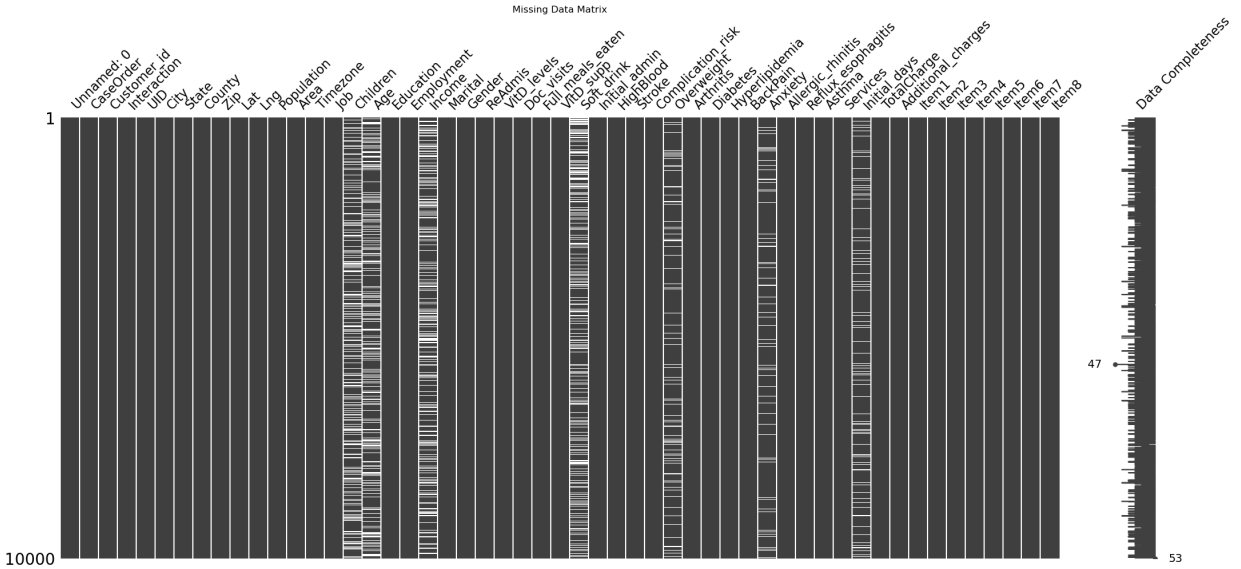
```
##df[['Age','AgeZScore']].head
##plt.hist(df['AgeZScore'])
##plt.show()

##df['IncomeZScore']=stats.zscore(df['Income'])
##df[['Income','IncomeZScore']].head
##plt.hist(df['IncomeZScore'])
##plt.show()

##df['IncomeZScore']=stats.zscore(df['Income'])
##df[['Income','IncomeZScore']].head
##plt.hist(df['IncomeZScore'])
##plt.show()

##df['VitD_levelsZScore']=stats.zscore(df['VitD_levels'])
##df[['VitD_levels','VitD_levelsZScore']].head
##plt.hist(df['VitD_levelsZScore'])
##plt.show()

##df['Doc_visitsZScore']=stats.zscore(df['Doc_visits'])
##df[['Doc_visits','Doc_visitsZScore']].head
##plt.hist(df['Doc_visitsZScore'])
##plt.show()

#Check for unique values to re-express categorical variables
print('Education: ', df.Education.unique(), "\n")
print('Marital: ', df.Marital.unique())
print('Gender: ', df.Gender.unique())
print('ReAdmis: ', df.ReAdmis.unique())
print('Soft_drink: ', df.Soft_drink.unique())
print('Initial_admin: ', df.Initial_admin.unique())
print('HighBlood: ', df.HighBlood.unique())
print('Stroke: ', df.Stroke.unique())
print('Complication_risk: ', df.Complication_risk.unique())
print('Arthritis: ', df.Arthritis.unique())
print('Diabetes: ', df.Diabetes.unique())
print('Hyperlipidemia: ', df.Hyperlipidemia.unique())
print('BackPain: ', df.BackPain.unique())
print('Allergic_rhinitis: ', df.Allergic_rhinitis.unique())
print('Reflux_esophagitis: ', df.Reflux_esophagitis.unique())
print('Ashtma: ', df.Asthma.unique())
print('Services: ', df.Services.unique())

#Turn categorical values into quantitative data
df['Education_numeric'] = df['Education']
dict_edu = {"Education_numeric": {"No Schooling Completed": 0,"Nursery School to 8th Grade": 1,'
df.replace(dict_edu, inplace=True)

df['Marital_numeric'] = df['Marital']
dict_marital = {"Marital_numeric": {"Never Married": 0,"Separated": 1,"Widowed": 2,"Divorced": 3
df.replace(dict_marital, inplace=True)

df['Gender_numeric'] = df['Gender']
dict_gender = {"Gender_numeric": {"Prefer not to answer": 0,"Male": 1,"Female": 2}}
df.replace(dict_gender, inplace=True)

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['Initial_admin_numeric'] = df['Initial_admin']
dict_Initial_admin = {"Initial_admin_numeric": {"Emergency Admission": 0, "Elective Admission":
df.replace(dict_Initial_admin, inplace=True)
```

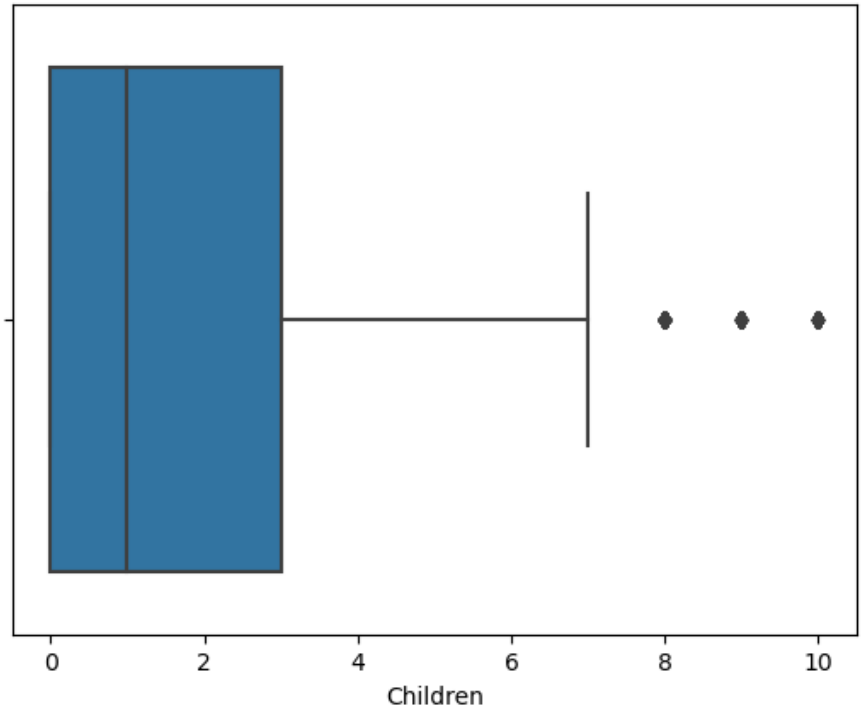
```
df['HighBlood_numeric'] = df['HighBlood']
dict_HighBlood = {"HighBlood_numeric": {"No": 0, "Yes": 1}}
df.replace(dict_HighBlood, inplace=True)

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

df['Complication_risk_numeric'] = df['Complication_risk']
dict_complication = {"Complication_risk_numeric": {"Low": 0, "Medium": 1, "High": 2}}
df.replace(dict_complication, 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['Services_numeric'] = df['Services']
dict_services = {"Services_numeric": {"Blood Work": 0, "Intravenous": 1, "CT Scan": 2, "MRI": 3}
df.replace(dict_services, inplace=True)

#Check for new columns
df.info()

#Checking data in new columns to make sure replacements worked
print('Education_numeric: ', df.Education_numeric.unique())
print('Marital_numeric: ', df.Marital_numeric.unique())
print('Gender_numeric: ', df.Gender_numeric.unique())
print('ReAdmis_numeric: ', df.ReAdmis_numeric.unique())
print('Soft_drink_numeric: ', df.Soft_drink_numeric.unique())
print('Initial_admin_numeric: ', df.Initial_admin_numeric.unique())
print('HighBlood_numeric: ', df.HighBlood_numeric.unique())
print('Stroke_numeric: ', df.Stroke_numeric.unique())
print('Complication_risk_numeric: ', df.Complication_risk_numeric.unique())
print('Arthritis_numeric: ', df.Arthritis_numeric.unique())
print('Diabetes_numeric: ', df.Diabetes_numeric.unique())
print('Hyperlipidemia_numeric: ', df.Hyperlipidemia_numeric.unique())
print('BackPain_numeric: ', df.BackPain_numeric.unique())
print('Allergic_rhinitis_numeric: ', df.Allergic_rhinitis_numeric.unique())
print('Reflux_esophagitis_numeric: ', df.Reflux_esophagitis_numeric.unique())
print('Ashtma_numeric: ', df.Asthma_numeric.unique())
print('Services_numeric: ', df.Services_numeric.unique(), '\n')
```

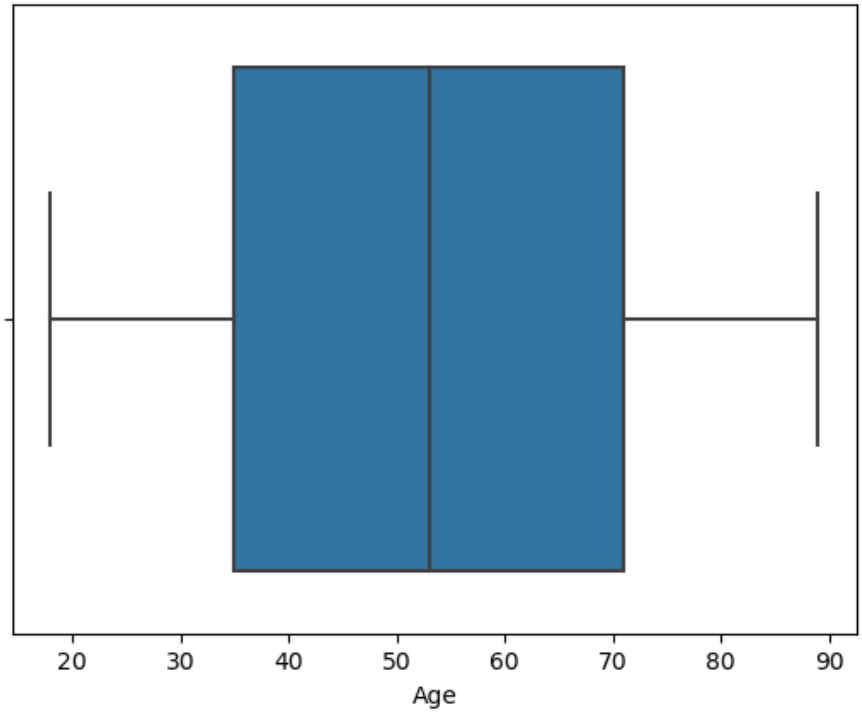
```
##Running PCA
test_pca = X_stand_df
test_pca_normalized=(test_pca-test_pca.mean())/test_pca.std()
pca = PCA(n_components=test_pca.shape[1])
pca.fit(test_pca_normalized)
print(pca)

test_pca2 = pd.DataFrame(pca.transform(test_pca_normalized),columns=['PC1','PC2','PC3', 'PC4',
                                                                     'PC10', 'PC11', 'PC12'])


loadings = pd.DataFrame(pca.components_.T,
columns = ['PC1','PC2','PC3', 'PC4', 'PC5', 'PC6', 'PC7', 'PC8', 'PC9','PC10', 'PC11', 'PC12',
index=test_pca_normalized.columns)
loadings
print(loadings)

cov_matrix = np.dot(test_pca_normalized.T, test_pca_normalized) / test_pca.shape[0]
eigenvalues = [np.dot(eigenvector.T, np.dot(cov_matrix, eigenvector)) for eigenvector in pca.cor

plt.plot(eigenvalues)
plt.xlabel('number of components')
plt.ylabel('eigenvalues')
plt.axhline(y=1, color="red")
plt.title('Scree Plot')
plt.show()

#Extracting cleaned data
df.to_csv(r'C:\Users\mmorg\Desktop\Cleaned206data.csv')
```

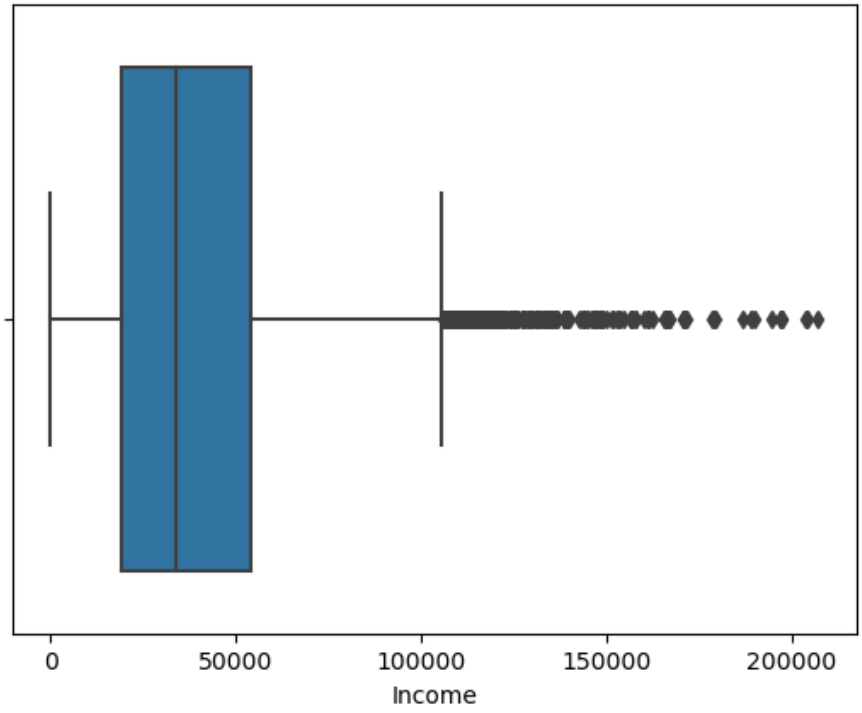
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10000 entries, 0 to 9999
Data columns (total 53 columns):
 #   Column              Non-Null Count  Dtype  
---  ------              --------------  -----  
 0   Unnamed: 0          10000 non-null  int64  
 1   CaseOrder           10000 non-null  int64  
 2   Customer_id         10000 non-null  object 
 3   Interaction         10000 non-null  object 
 4   UID                 10000 non-null  object 
 5   City                10000 non-null  object 
 6   State               10000 non-null  object 
 7   County              10000 non-null  object 
 8   Zip                 10000 non-null  int64  
 9   Lat                 10000 non-null  float64
 10  Lng                 10000 non-null  float64
 11  Population          10000 non-null  int64  
 12  Area                10000 non-null  object 
 13  Timezone            10000 non-null  object 
 14  Job                 10000 non-null  object 
 15  Children            7412 non-null   float64
 16  Age                 7586 non-null   float64
 17  Education           10000 non-null  object 
 18  Employment          10000 non-null  object 
 19  Income              7536 non-null   float64
 20  Marital             10000 non-null  object 
 21  Gender              10000 non-null  object 
 22  ReAdmis             10000 non-null  object 
 23  VitD_levels         10000 non-null  float64
 24  Doc_visits          10000 non-null  int64  
 25  Full_meals_eaten    10000 non-null  int64  
 26  VitD_supp           10000 non-null  int64  
 27  Soft_drink          7533 non-null   object 
 28  Initial_admin       10000 non-null  object 
 29  HighBlood           10000 non-null  object 
 30  Stroke              10000 non-null  object 
 31  Complication_risk   10000 non-null  object 
 32  Overweight          9018 non-null   float64
 33  Arthritis           10000 non-null  object 
 34  Diabetes            10000 non-null  object 
 35  Hyperlipidemia      10000 non-null  object 
 36  BackPain            10000 non-null  object 
 37  Anxiety             9016 non-null   float64
 38  Allergic_rhinitis   10000 non-null  object 
 39  Reflux_esophagitis  10000 non-null  object 
 40  Asthma              10000 non-null  object 
 41  Services            10000 non-null  object 
 42  Initial_days        8944 non-null   float64
 43  TotalCharge         10000 non-null  float64
 44  Additional_charges  10000 non-null  float64
 45  Item1               10000 non-null  int64  
 46  Item2               10000 non-null  int64  
 47  Item3               10000 non-null  int64  
 48  Item4               10000 non-null  int64  
 49  Item5               10000 non-null  int64  
 50  Item6               10000 non-null  int64  
 51  Item7               10000 non-null  int64  
 52  Item8               10000 non-null  int64  
dtypes: float64(11), int64(15), object(27)
memory usage: 4.0+ MB
0       False
1       False
2       False
3       False
4       False
        ...
9995    False
```

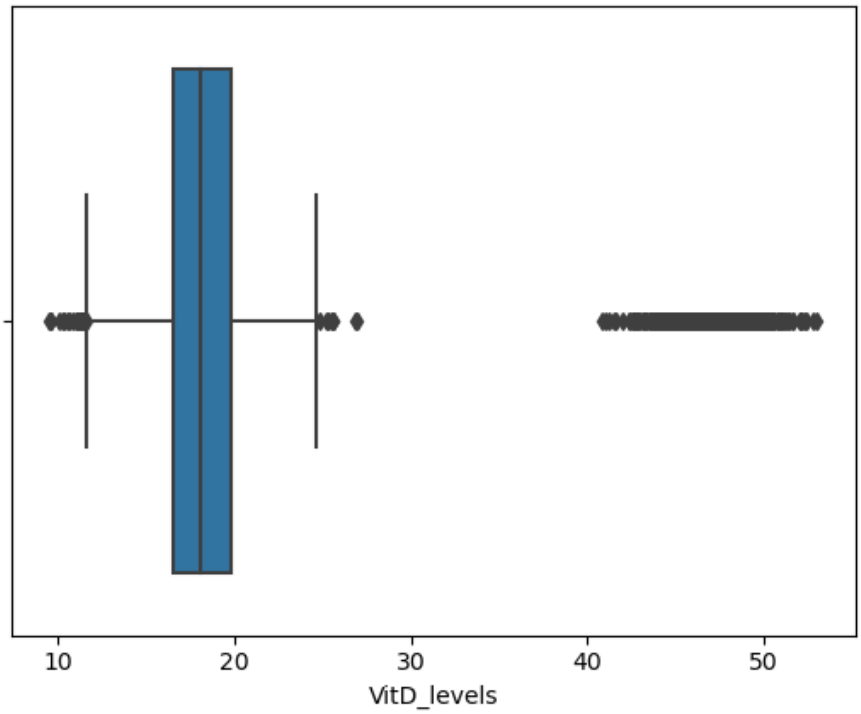
```
9996    False
9997    False
9998    False
9999    False
Length: 10000, dtype: bool
Unnamed: 0                0
CaseOrder                 0
Customer_id               0
Interaction               0
UID                       0
City                      0
State                     0
County                    0
Zip                       0
Lat                       0
Lng                       0
Population                0
Area                      0
Timezone                  0
Job                       0
Children               2588
Age                    2414
Education                 0
Employment                0
Income                 2464
Marital                   0
Gender                    0
ReAdmis                   0
VitD_levels               0
Doc_visits                0
Full_meals_eaten          0
VitD_supp                 0
Soft_drink             2467
Initial_admin             0
HighBlood                 0
Stroke                    0
Complication_risk         0
Overweight              982
Arthritis                 0
Diabetes                  0
Hyperlipidemia            0
BackPain                  0
Anxiety                 984
Allergic_rhinitis         0
Reflux_esophagitis        0
Asthma                    0
Services                  0
Initial_days           1056
TotalCharge               0
Additional_charges        0
Item1                     0
Item2                     0
Item3                     0
Item4                     0
Item5                     0
Item6                     0
Item7                     0
Item8                     0
dtype: int64
```

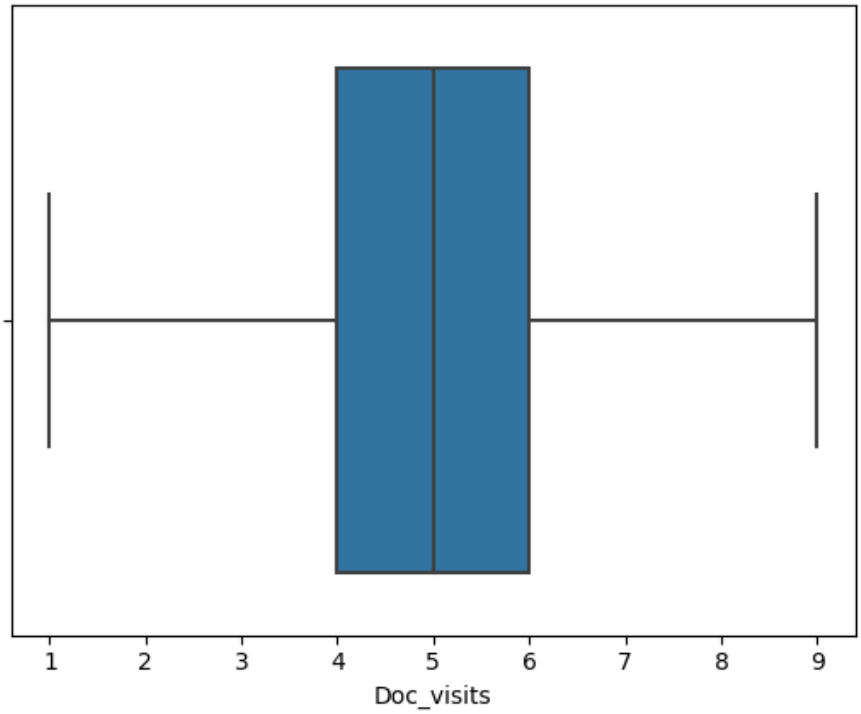

Missing Data Matrix
Unnamed: 0
CaseOrder
Customer_id
Interaction
UID
City
State
County
Zip
Lat
Lng
Population
Area
Timezone
Job
Children
Age
Education
Employment
Income
Marital
Gender
ReAdmis
VitD_levels
Doc_visits
Full_meals_eaten
VitD_supp
Soft_drink
Initial_admin
HighBlood
Stroke
Complication_risk
Overweight
Arthritis
Diabetes
Hyperlipidemia
BackPain
Anxiety
Allergic_rhinitis
Reflux_esophagitis
Asthma
Services
Initial_days
TotalCharge
Additional_charges
Item1
Item2
Item3
Item4
Item5
Item6
Item7
Item8
Data Completeness
1
10000
47
53

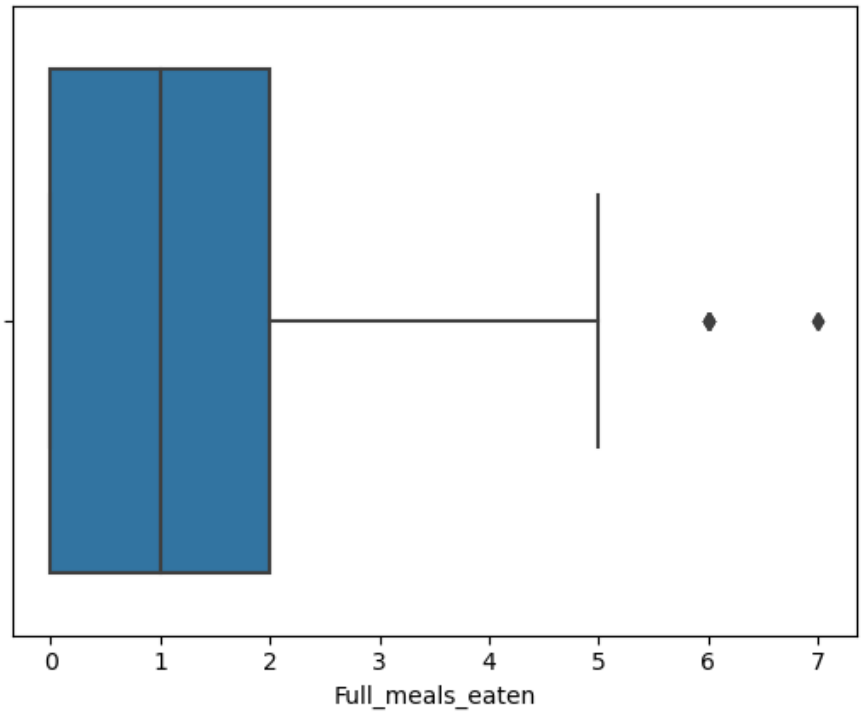

0
2
4
6
8
10
Children

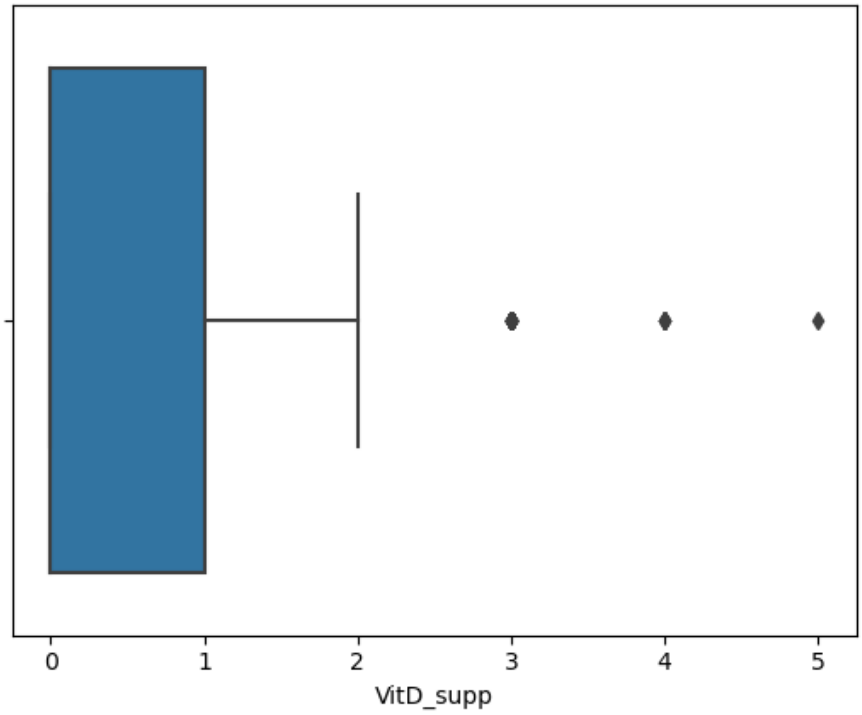

20
30
40
50
60
70
80
90
Age

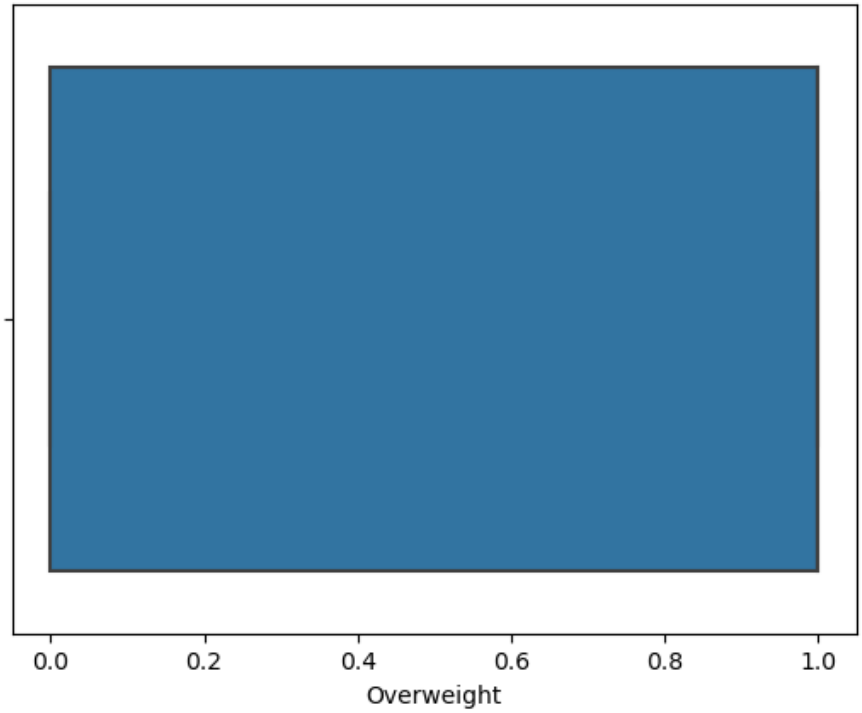

0
50000
100000
150000
200000
Income

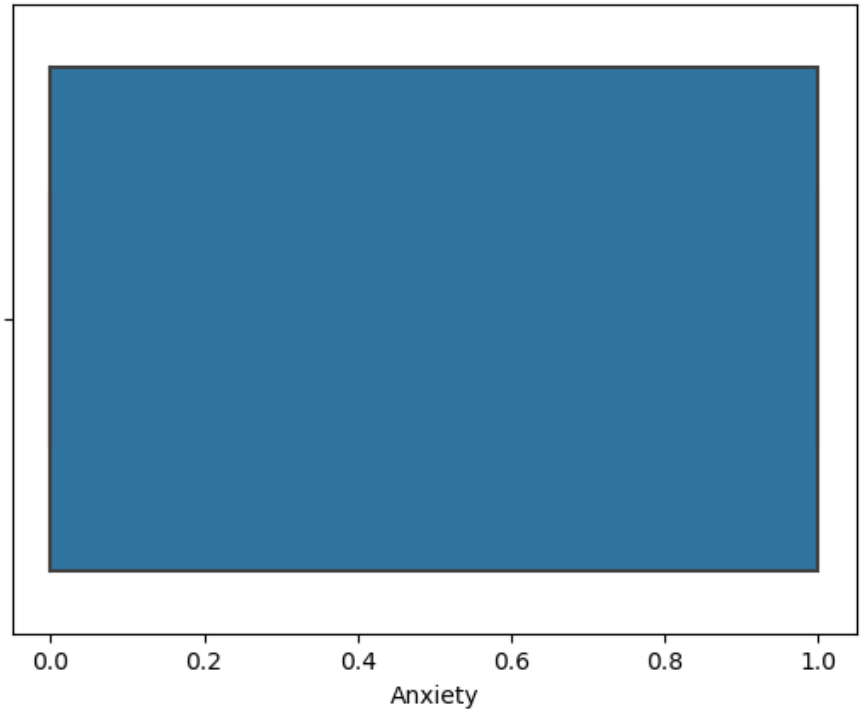

10
20
30
40
50
VitD_levels

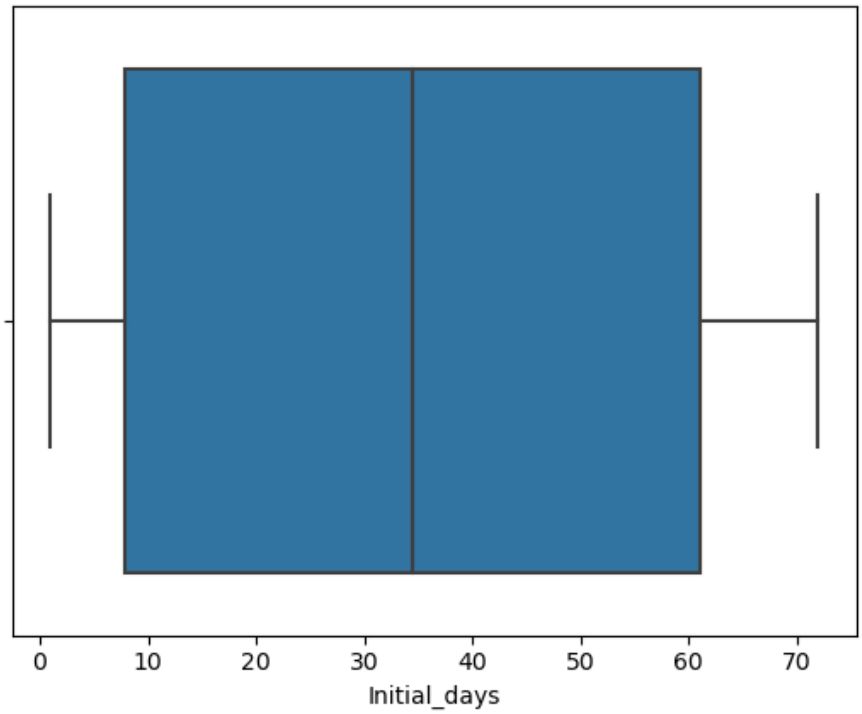

1
2
3
4
5
6
7
8
9
Doc_visits

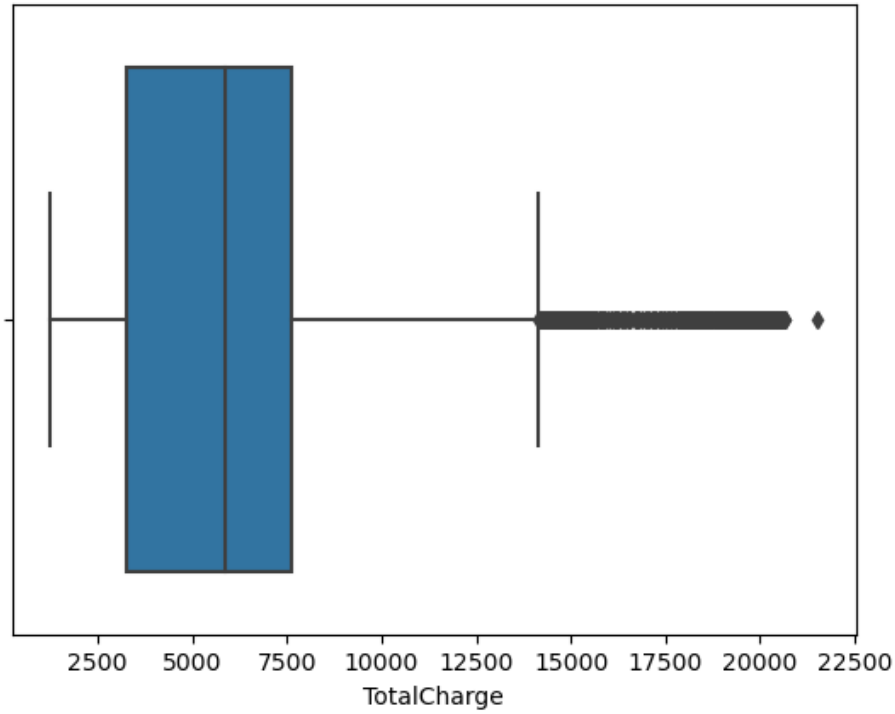

0
1
2
3
4
5
6
7
Full_meals_eaten

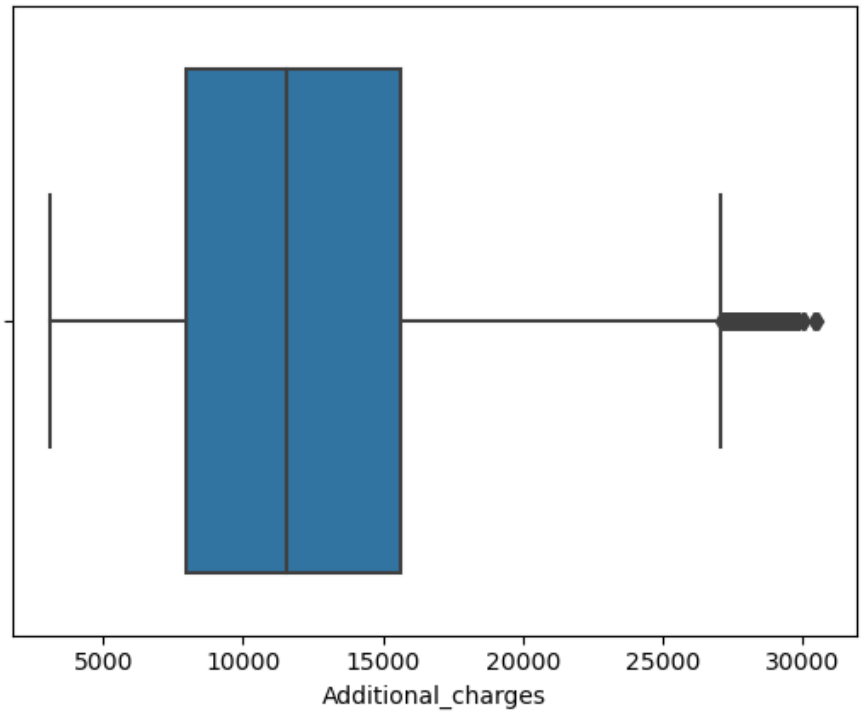

0
1
2
3
4
5
VitD_supp

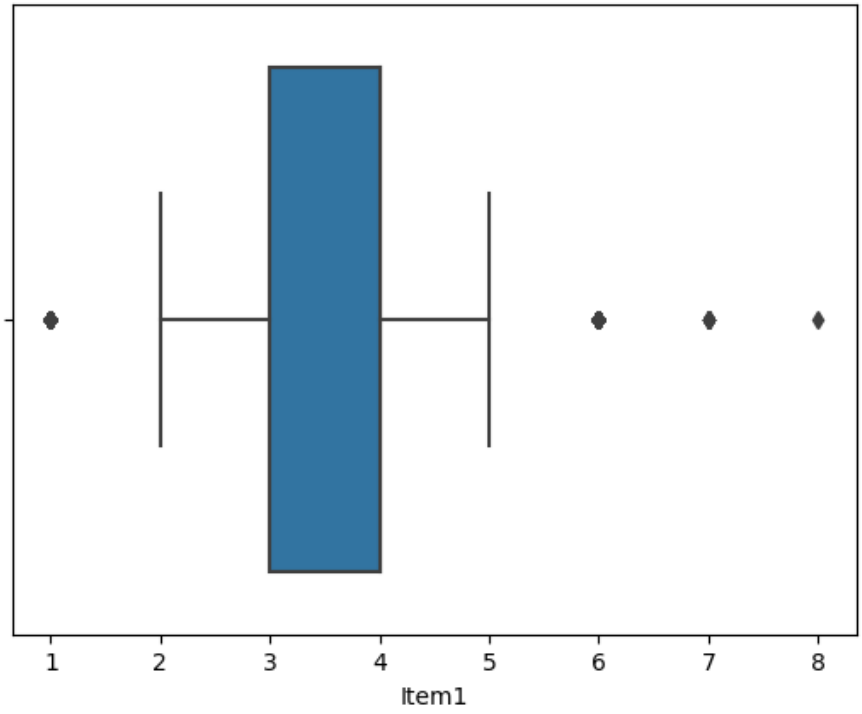

0.0
0.2
0.4
0.6
0.8
1.0
Overweight

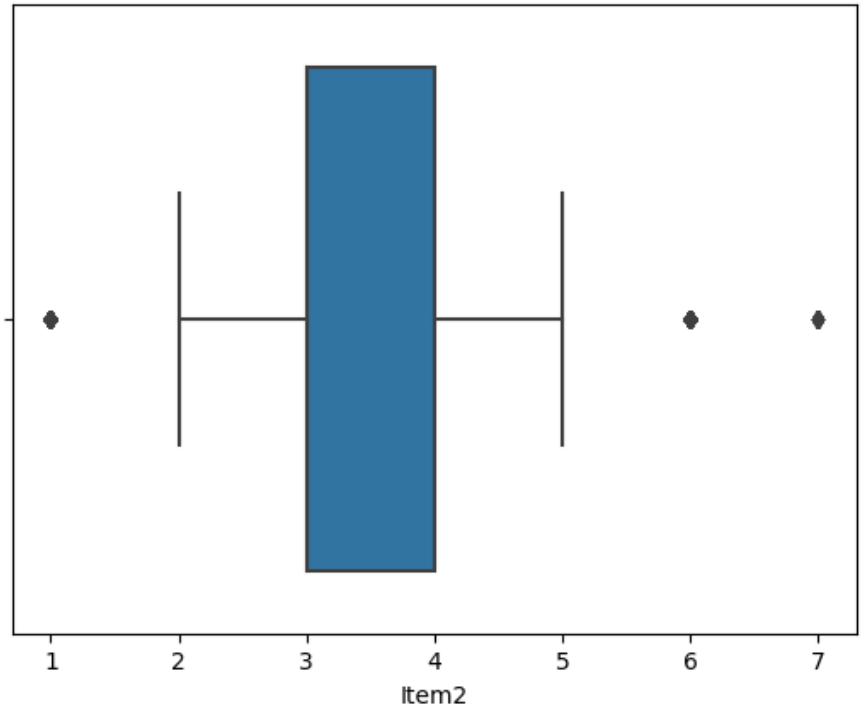

0.0
0.2
0.4
0.6
0.8
1.0
Anxiety

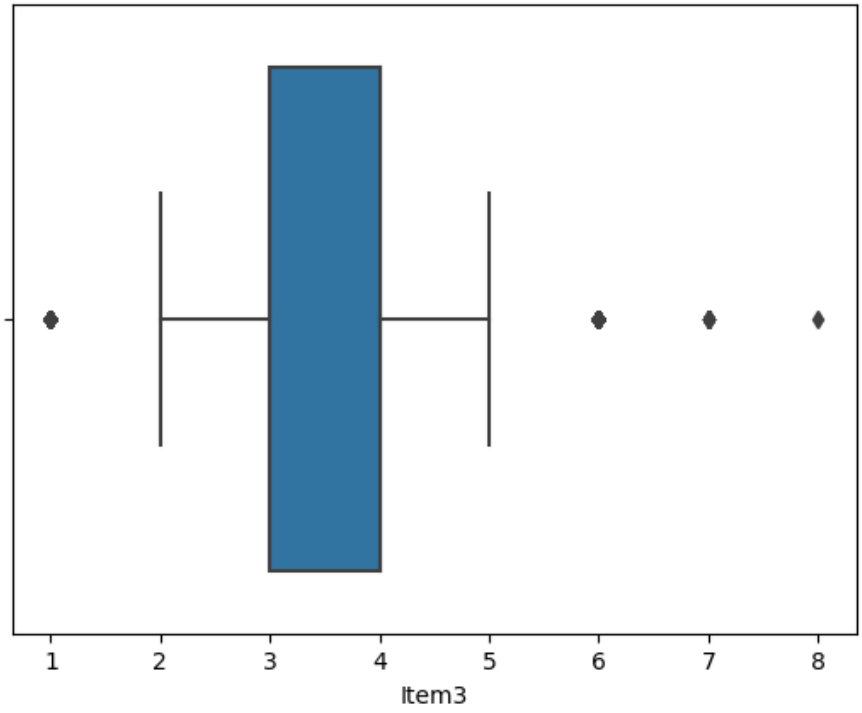

0
10
20
30
40
50
60
70
Initial_days

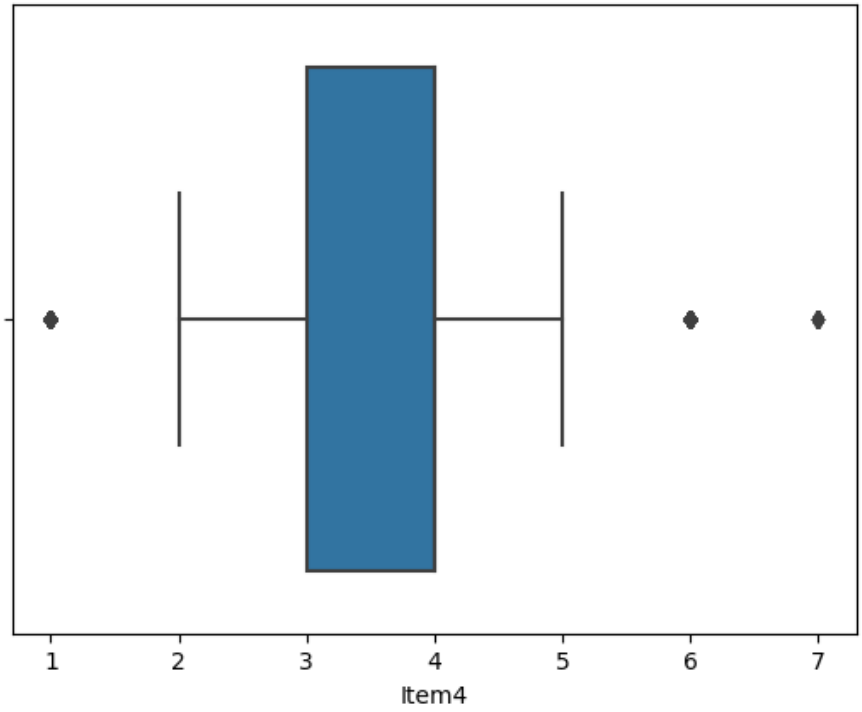

2500
5000
7500
10000
12500
15000
17500
20000
22500
TotalCharge

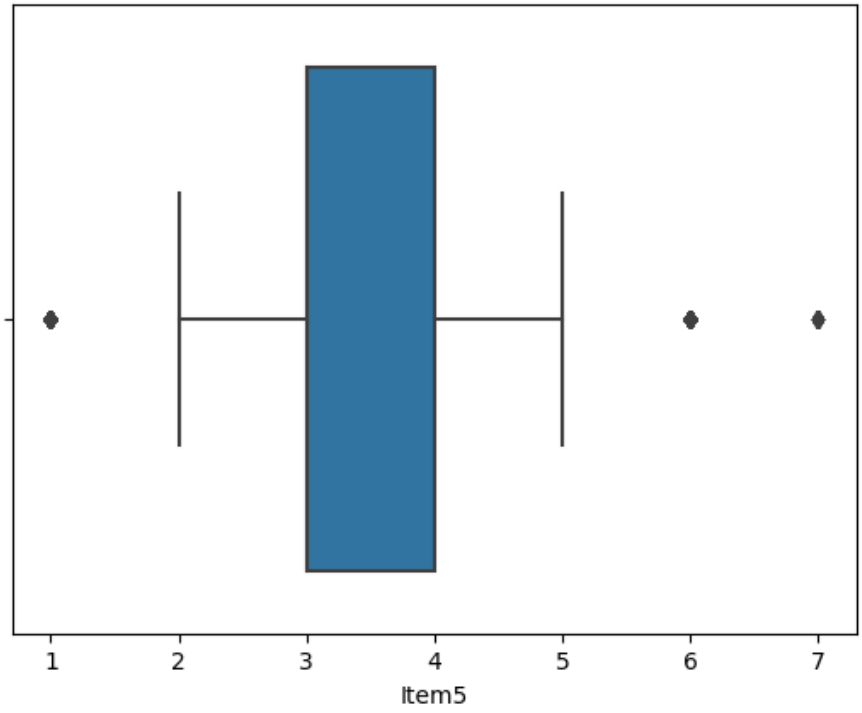

5000
10000
15000
20000
25000
30000
Additional_charges

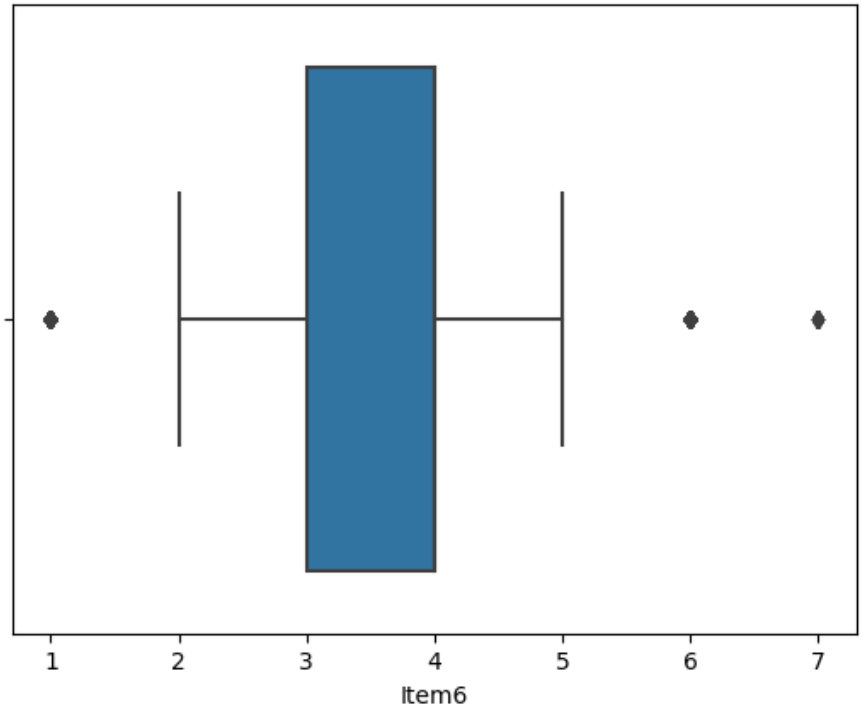

1
2
3
4
5
6
7
8
Item1

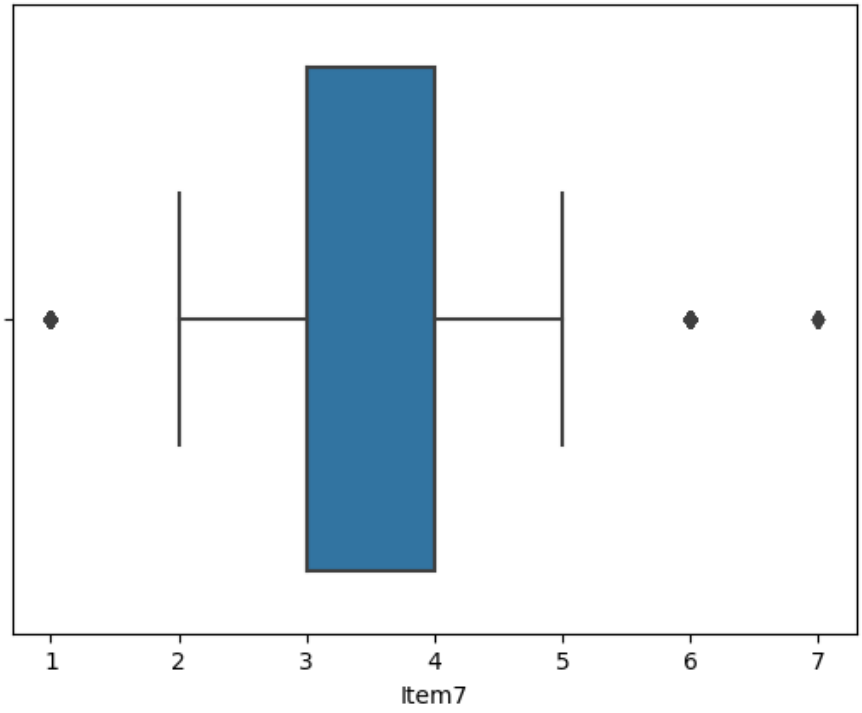

1
2
3
4
5
6
7
Item2

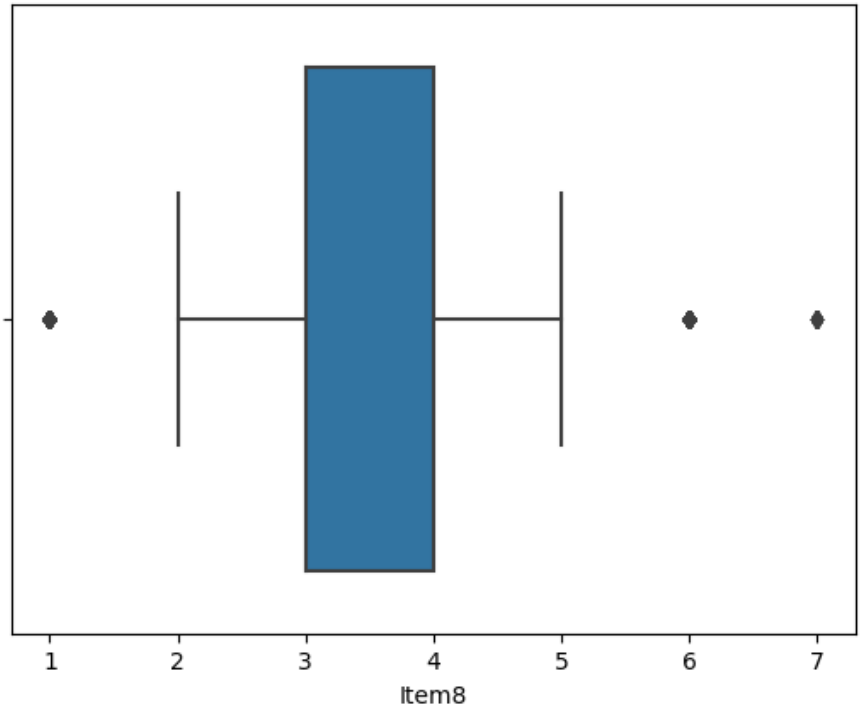

1
2
3
4
5
6
7
8
Item3


1
2
3
4
5
6
7
Item4

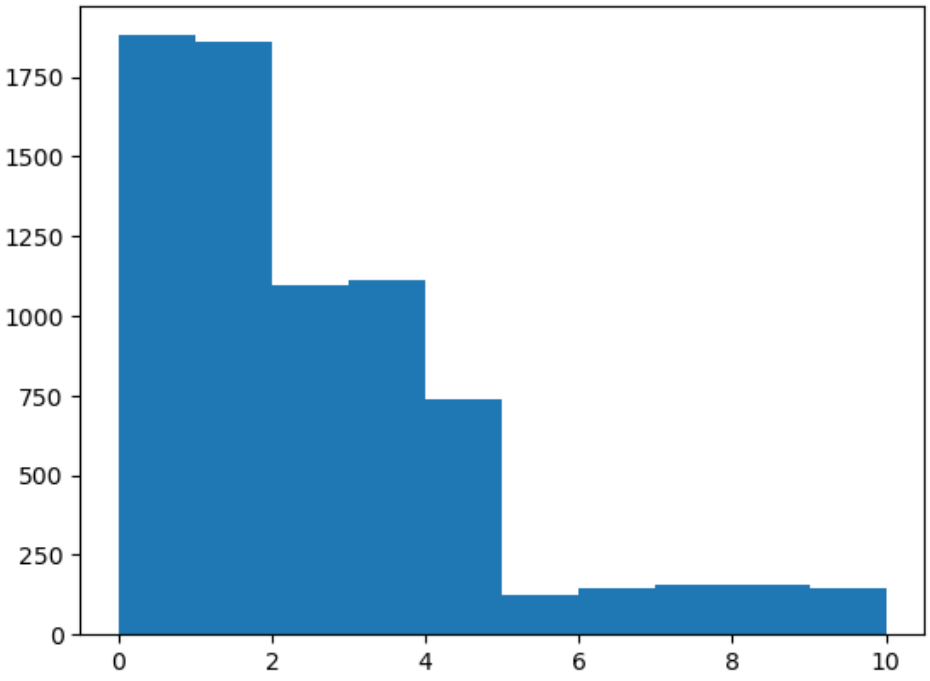

1
2
3
4
5
6
7
Item5


1
2
3
4
5
6
7
Item6

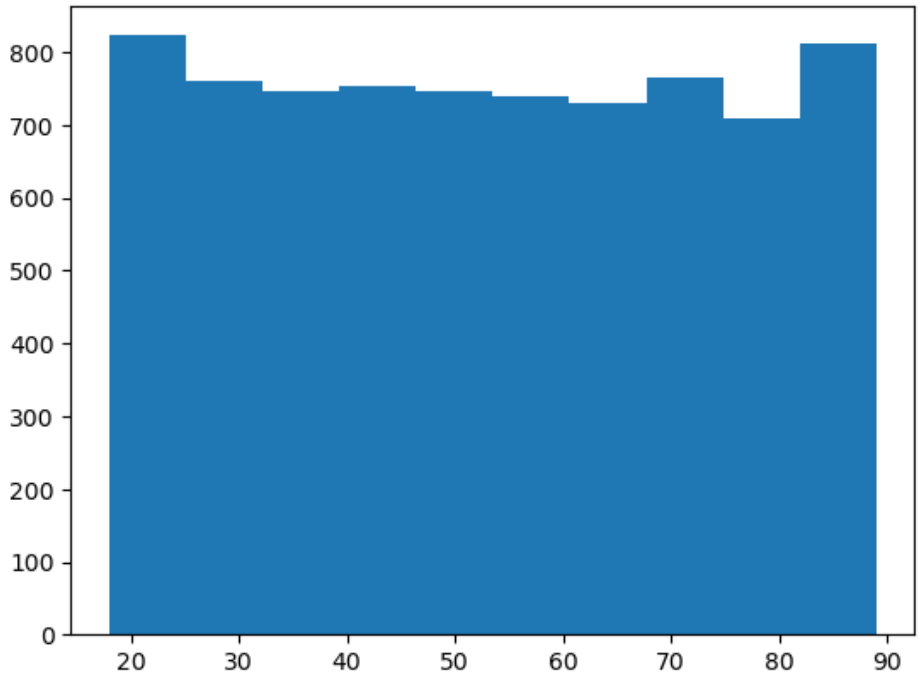

1
2
3
4
5
6
7
Item7



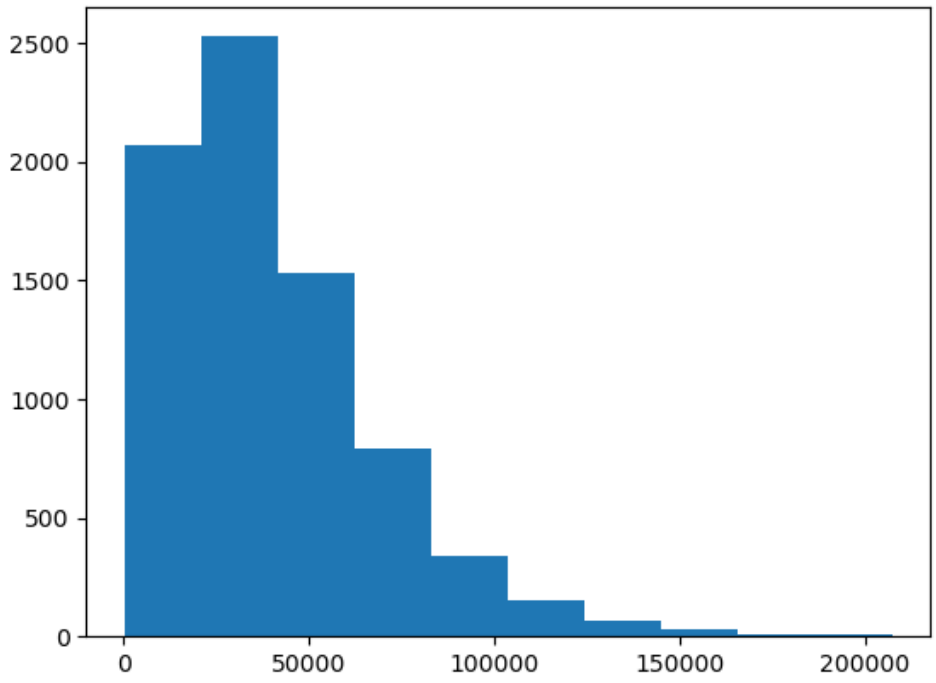

```
Children Original:
```



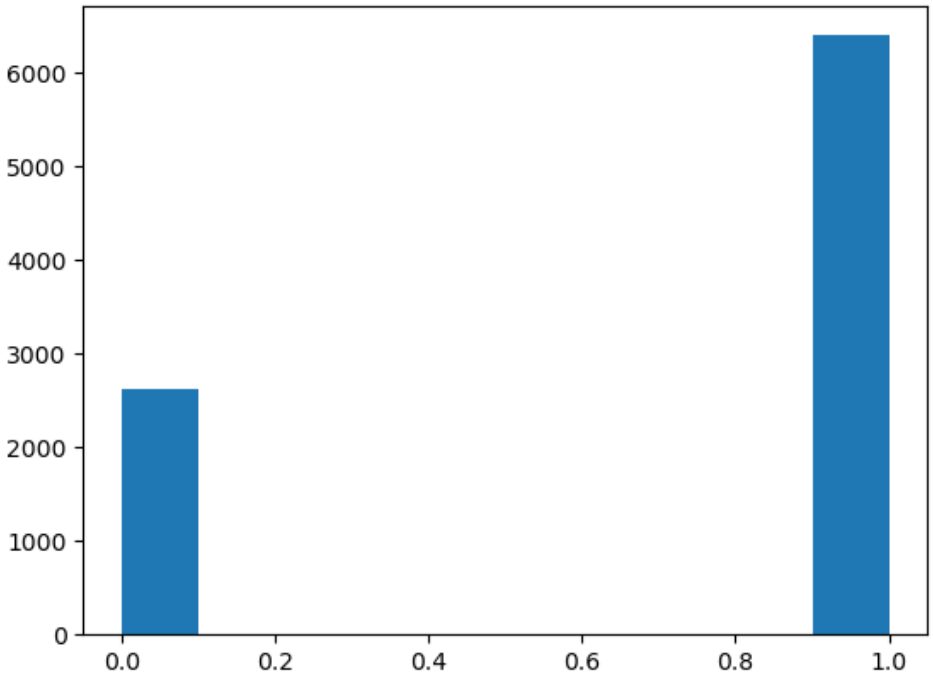

```
Age Original:
```



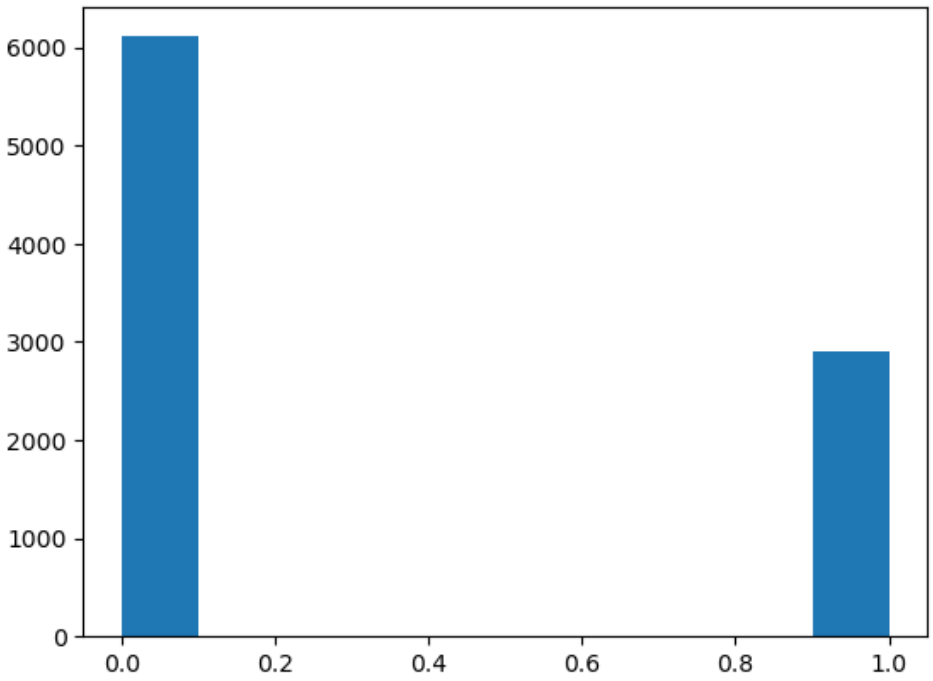
Income Original:



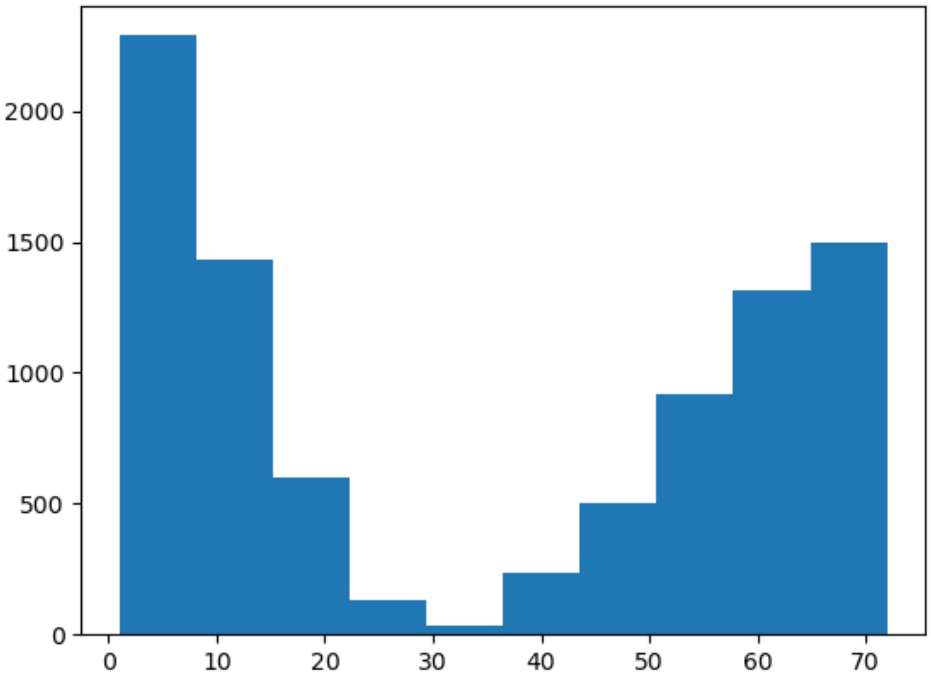
Overweight Original:



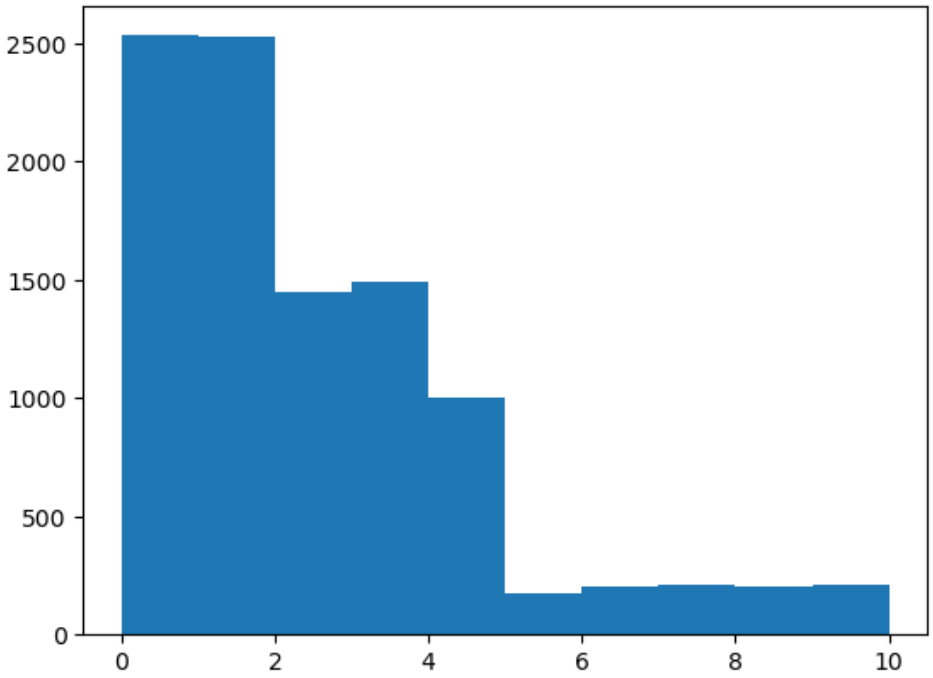

```
Anxiety Original:
```



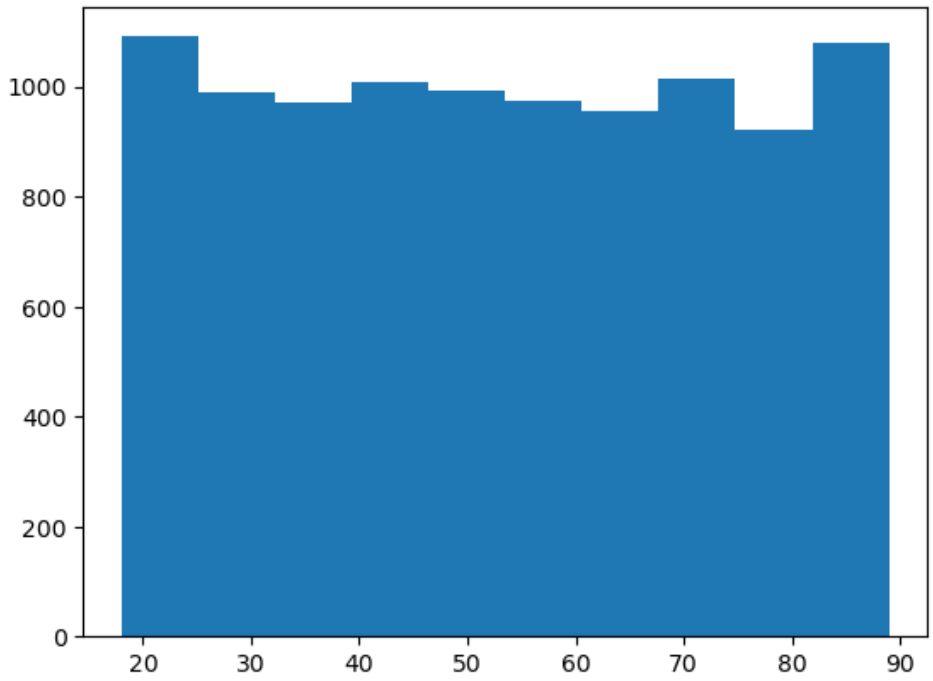

```
Initial_days Original:
```


2000
1500
1000
500
0
0
10
20
30
40
50
60
70

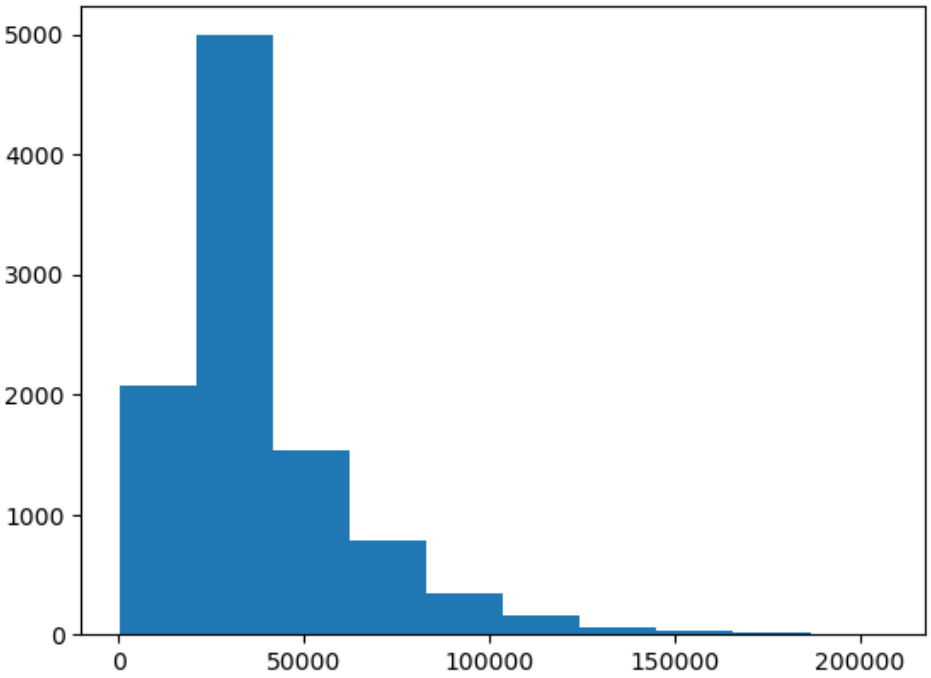
```
Unnamed: 0            0
CaseOrder             0
Customer_id           0
Interaction           0
UID                   0
City                  0
State                 0
County                0
Zip                   0
Lat                   0
Lng                   0
Population            0
Area                  0
Timezone              0
Job                   0
Children              0
Age                   0
Education             0
Employment            0
Income                0
Marital               0
Gender                0
ReAdmis               0
VitD_levels           0
Doc_visits            0
Full_meals_eaten      0
VitD_supp             0
Soft_drink            0
Initial_admin         0
HighBlood             0
Stroke                0
Complication_risk     0
Overweight            0
Arthritis             0
Diabetes              0
Hyperlipidemia        0
BackPain              0
Anxiety               0
Allergic_rhinitis     0
Reflux_esophagitis    0
Asthma                0
Services              0
Initial_days          0
TotalCharge           0
Additional_charges    0
Item1                 0
Item2                 0
Item3                 0
Item4                 0
Item5                 0
Item6                 0
Item7                 0
Item8                 0
dtype: int64
Children Modified:
```



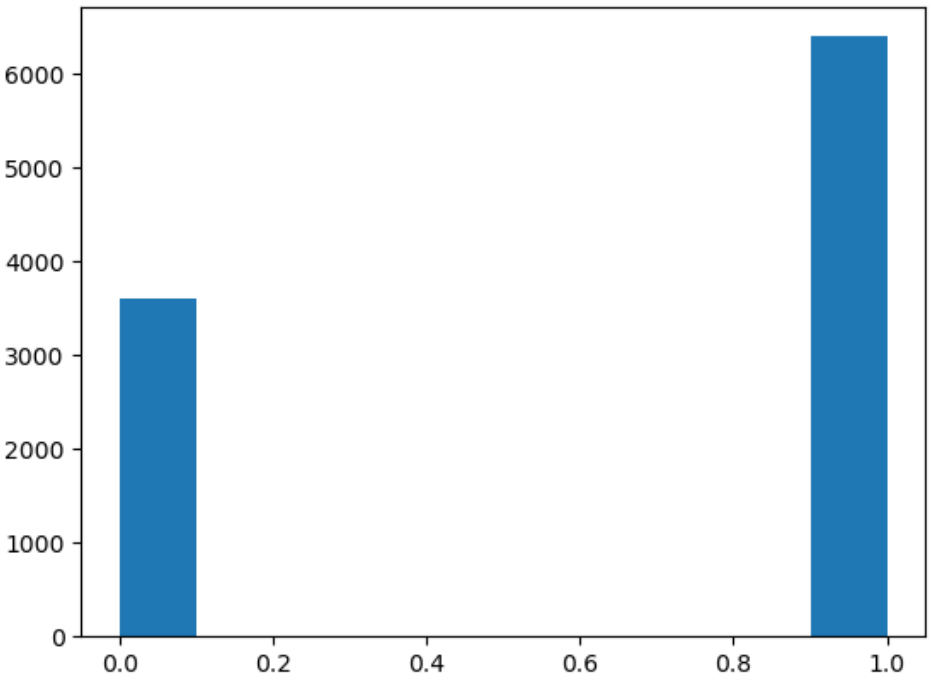
Age Modified:



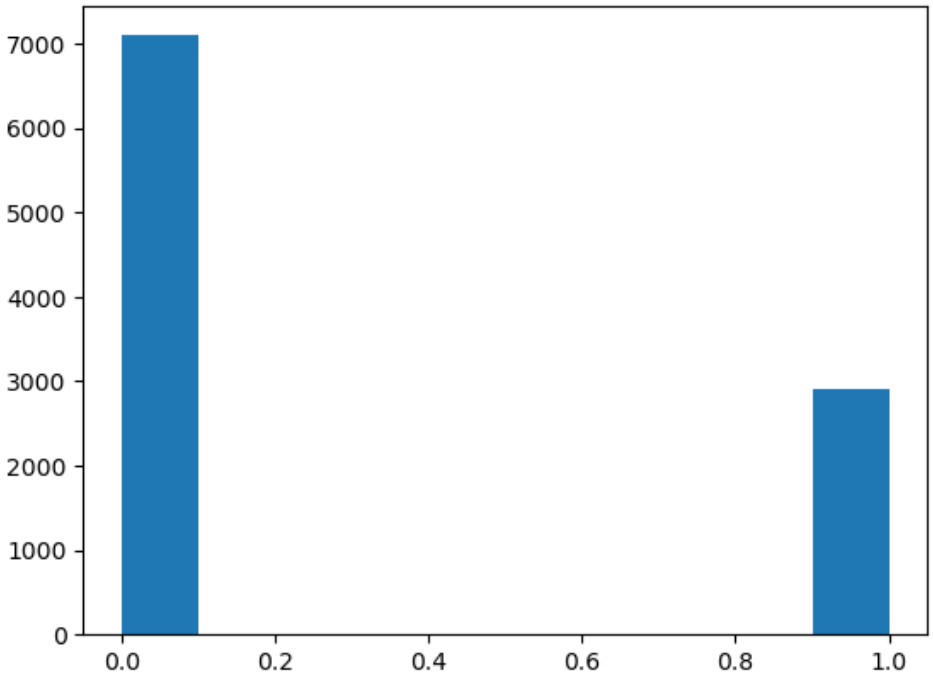
Income Modified:



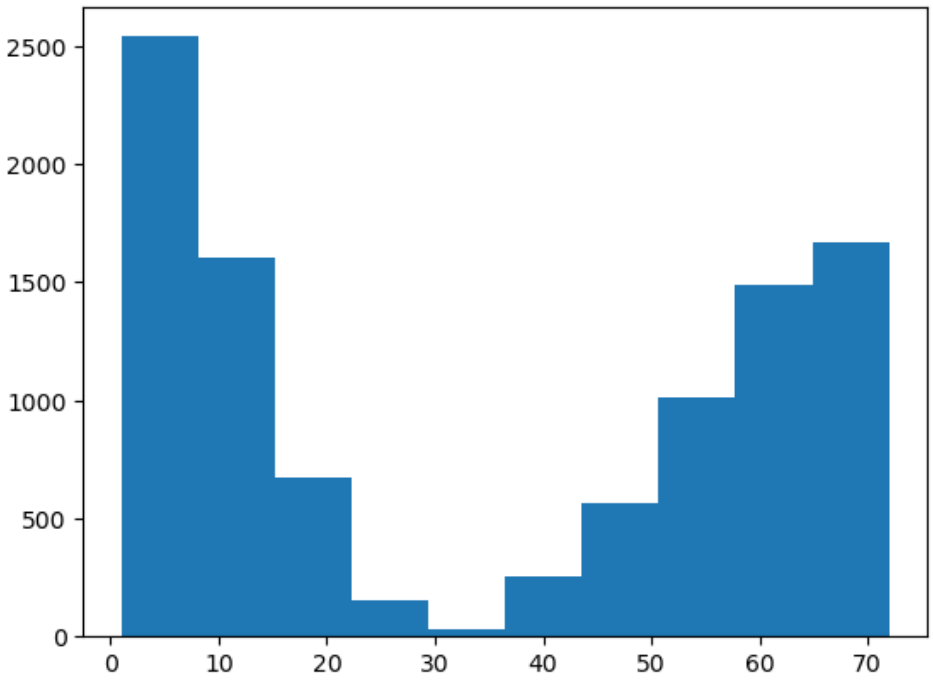
Overweight Modified:



Anxiety Modified:




```
Initial_days Modified:
```



```
Education:  ['Some College, Less than 1 Year'
 'Some College, 1 or More Years, No Degree'
 'GED or Alternative Credential' 'Regular High School Diploma'
 "Bachelor's Degree" "Master's Degree" 'Nursery School to 8th Grade'
 '9th Grade to 12th Grade, No Diploma' 'Doctorate Degree'
 "Associate's Degree" 'Professional School Degree'
 'No Schooling Completed']

Marital:  ['Divorced' 'Married' 'Widowed' 'Never Married' 'Separated']
Gender:  ['Male' 'Female' 'Prefer not to answer']
ReAdmis:  ['No' 'Yes']
Soft_drink:  ['No' 'Yes']
Initial_admin:  ['Emergency Admission' 'Elective Admission' 'Observation Admission']
HighBlood:  ['Yes' 'No']
Stroke:  ['No' 'Yes']
Complication_risk:  ['Medium' 'High' 'Low']
Arthritis:  ['Yes' 'No']
Diabetes:  ['Yes' 'No']
Hyperlipidemia:  ['No' 'Yes']
BackPain:  ['Yes' 'No']
Allergic_rhinitis:  ['Yes' 'No']
Reflux_esophagitis:  ['No' 'Yes']
Ashtma:  ['Yes' 'No']
Services:  ['Blood Work' 'Intravenous' 'CT Scan' 'MRI']
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10000 entries, 0 to 9999
Data columns (total 70 columns):
 #   Column                      Non-Null Count  Dtype
---  ------                      --------------  -----
 0   Unnamed: 0                  10000 non-null  int64
 1   CaseOrder                   10000 non-null  int64
 2   Customer_id                 10000 non-null  object
 3   Interaction                 10000 non-null  object
 4   UID                         10000 non-null  object
 5   City                        10000 non-null  object
 6   State                       10000 non-null  object
 7   County                      10000 non-null  object
 8   Zip                         10000 non-null  int64
 9   Lat                         10000 non-null  float64
 10  Lng                         10000 non-null  float64
 11  Population                  10000 non-null  int64
 12  Area                        10000 non-null  object
 13  Timezone                    10000 non-null  object
 14  Job                         10000 non-null  object
 15  Children                    10000 non-null  float64
 16  Age                         10000 non-null  float64
 17  Education                   10000 non-null  object
 18  Employment                  10000 non-null  object
 19  Income                      10000 non-null  float64
 20  Marital                     10000 non-null  object
 21  Gender                      10000 non-null  object
 22  ReAdmis                     10000 non-null  object
 23  VitD_levels                 10000 non-null  float64
 24  Doc_visits                  10000 non-null  int64
 25  Full_meals_eaten            10000 non-null  int64
 26  VitD_supp                   10000 non-null  int64
 27  Soft_drink                  10000 non-null  object
 28  Initial_admin               10000 non-null  object
 29  HighBlood                   10000 non-null  object
 30  Stroke                      10000 non-null  object
 31  Complication_risk           10000 non-null  object
 32  Overweight                  10000 non-null  float64
 33  Arthritis                   10000 non-null  object
 34  Diabetes                    10000 non-null  object
 35  Hyperlipidemia              10000 non-null  object
 36  BackPain                    10000 non-null  object
 37  Anxiety                     10000 non-null  float64
```

```
 38  Allergic_rhinitis           10000 non-null  object
 39  Reflux_esophagitis          10000 non-null  object
 40  Asthma                      10000 non-null  object
 41  Services                    10000 non-null  object
 42  Initial_days                10000 non-null  float64
 43  TotalCharge                 10000 non-null  float64
 44  Additional_charges          10000 non-null  float64
 45  Item1                       10000 non-null  int64
 46  Item2                       10000 non-null  int64
 47  Item3                       10000 non-null  int64
 48  Item4                       10000 non-null  int64
 49  Item5                       10000 non-null  int64
 50  Item6                       10000 non-null  int64
 51  Item7                       10000 non-null  int64
 52  Item8                       10000 non-null  int64
 53  Education_numeric           10000 non-null  int64
 54  Marital_numeric             10000 non-null  int64
 55  Gender_numeric              10000 non-null  int64
 56  ReAdmis_numeric             10000 non-null  int64
 57  Soft_drink_numeric          10000 non-null  int64
 58  Initial_admin_numeric       10000 non-null  int64
 59  HighBlood_numeric           10000 non-null  int64
 60  Stroke_numeric              10000 non-null  int64
 61  Complication_risk_numeric   10000 non-null  int64
 62  Arthritis_numeric           10000 non-null  int64
 63  Diabetes_numeric            10000 non-null  int64
 64  Hyperlipidemia_numeric      10000 non-null  int64
 65  BackPain_numeric            10000 non-null  int64
 66  Allergic_rhinitis_numeric   10000 non-null  int64
 67  Reflux_esophagitis_numeric  10000 non-null  int64
 68  Asthma_numeric              10000 non-null  int64
 69  Services_numeric            10000 non-null  int64
dtypes: float64(11), int64(32), object(27)
memory usage: 5.3+ MB
Education_numeric:  [ 5  6  3  4  9 10  1  2 11  7  8  0]
Marital_numeric:  [3 4 2 0 1]
Gender_numeric:  [1 2 0]
ReAdmis_numeric:  [0 1]
Soft_drink_numeric:  [0 1]
Initial_admin_numeric:  [0 1 2]
HighBlood_numeric:  [1 0]
Stroke_numeric:  [0 1]
Complication_risk_numeric:  [1 2 0]
Arthritis_numeric:  [1 0]
Diabetes_numeric:  [1 0]
Hyperlipidemia_numeric:  [0 1]
BackPain_numeric:  [1 0]
Allergic_rhinitis_numeric:  [1 0]
Reflux_esophagitis_numeric:  [0 1]
Ashtma_numeric:  [1 0]
Services_numeric:  [0 1 2 3]

PCA(n_components=20)
                                 PC1       PC2       PC3       PC4       PC5  \
ReAdmis_numeric             0.019851  0.127795  0.175456 -0.353865 -0.032540
Services_numeric            0.705512  0.009579 -0.013558 -0.004255 -0.010995
Age                         0.013203 -0.229001 -0.143450  0.074654  0.354792
Gender_numeric              0.036482 -0.036906  0.062988  0.395736 -0.274561
Doc_visits                 -0.001959 -0.261453  0.133387  0.008292 -0.242440
Soft_drink_numeric          0.011052  0.114656  0.543563  0.084041  0.164635
Initial_admin_numeric      -0.021101 -0.112728 -0.255054  0.400224 -0.167758
HighBlood_numeric           0.007924 -0.394309 -0.055899 -0.298067 -0.065738
Stroke_numeric             -0.003465 -0.196086  0.016540  0.251773 -0.072744
Complication_risk_numeric   0.004593 -0.140210 -0.025062 -0.192025 -0.258244
Arthritis_numeric          -0.004227  0.047176 -0.436380 -0.255660  0.114494
Diabetes_numeric            0.034947  0.094009 -0.086680  0.241931  0.356749
Hyperlipidemia_numeric     -0.004628  0.427117  0.150956 -0.177230  0.298589
```

BackPain_numeric	0.008511	-0.236902	0.538287	0.062751	0.062242
Allergic_rhinitis_numeric	0.006058	-0.242265	-0.058299	-0.164063	0.247245
Reflux_esophagitis_numeric	0.004447	0.157243	0.136674	0.022118	-0.331555
Asthma_numeric	0.004309	-0.159516	0.043155	0.303615	0.411165
Services_numeric	0.705512	0.009579	-0.013558	-0.004255	-0.010995
Education_numeric	-0.019142	0.238284	0.026542	-0.072749	-0.164102
Marital_numeric	0.012822	-0.454559	0.163839	-0.268373	0.081588

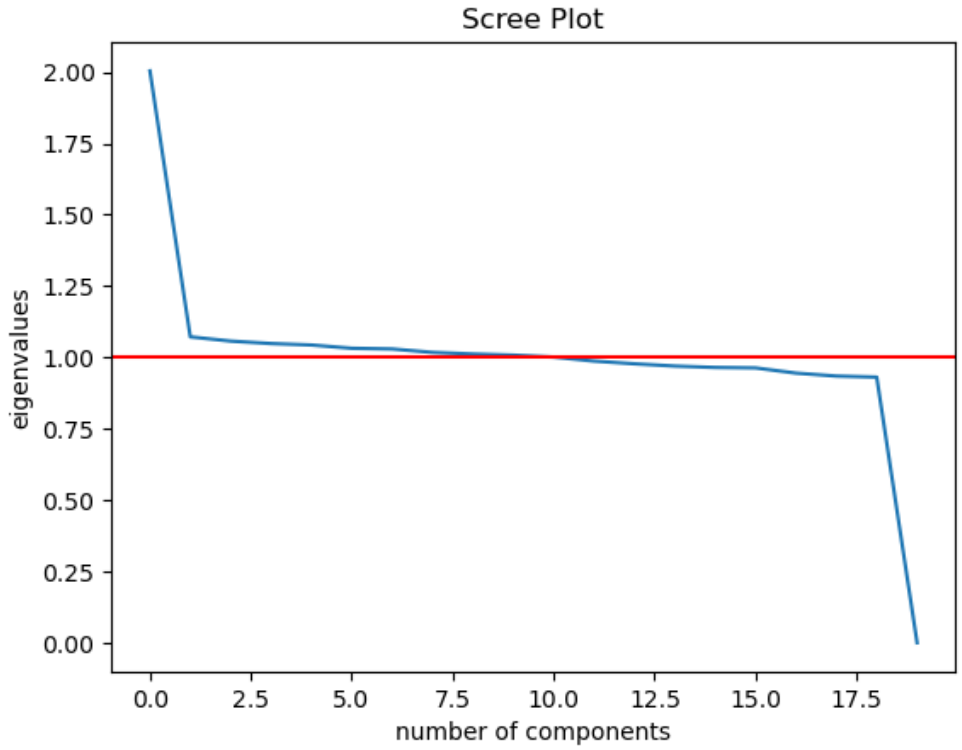
	PC6	PC7	PC8	PC9	PC10 \
ReAdmis_numeric	-0.143015	-0.119035	0.181851	0.522127	0.108240
Services_numeric	0.004155	0.017485	0.019507	-0.009729	-0.010693
Age	-0.062761	-0.149631	-0.112612	0.494622	0.017862
Gender_numeric	-0.219364	-0.189680	-0.461167	0.033749	-0.077140
Doc_visits	0.246096	-0.582819	0.107615	-0.035123	-0.031306
Soft_drink_numeric	0.196500	-0.273601	-0.004180	-0.054392	0.139432
Initial_admin_numeric	0.020542	0.095840	0.392278	0.006889	0.073086
HighBlood_numeric	0.095530	0.093574	0.124019	-0.150690	0.437013
Stroke_numeric	0.110094	0.149226	0.225268	0.426483	0.119244
Complication_risk_numeric	0.494101	0.105350	-0.316720	0.033632	0.347314
Arthritis_numeric	-0.369964	-0.313207	0.087425	0.013103	0.228266
Diabetes_numeric	0.222501	-0.425421	0.146032	-0.072290	0.276184
Hyperlipidemia_numeric	0.159638	0.191373	-0.063646	0.029092	0.110516
BackPain_numeric	-0.338007	0.069437	0.161794	0.074169	0.033480
Allergic_rhinitis_numeric	-0.144564	-0.223779	-0.321020	-0.215393	-0.121809
Reflux_esophagitis_numeric	-0.441239	-0.036442	-0.120240	-0.084374	0.545925
Asthma_numeric	-0.102526	0.238795	0.013412	-0.289638	0.330642
Services_numeric	0.004155	0.017485	0.019507	-0.009729	-0.010693
Education_numeric	-0.035461	-0.156908	0.460780	-0.308875	-0.114648
Marital_numeric	-0.094869	0.086213	0.140313	-0.163975	-0.230519

	PC11	PC12	PC13	PC14	PC15 \
ReAdmis_numeric	0.014549	0.449140	0.051658	0.140656	-0.415099
Services_numeric	-0.007358	0.000096	0.003527	-0.021920	0.016814
Age	-0.024631	-0.458214	-0.116199	-0.376736	-0.229437
Gender_numeric	0.136788	-0.143189	0.042206	0.288660	-0.347084
Doc_visits	-0.230820	0.045052	-0.021763	0.047394	0.185108
Soft_drink_numeric	-0.071909	-0.002802	-0.102073	-0.294438	0.185735
Initial_admin_numeric	-0.368419	-0.069619	0.460711	-0.110073	-0.115627
HighBlood_numeric	0.086318	-0.038997	0.218765	0.120616	-0.131129
Stroke_numeric	0.598561	0.017625	0.010997	0.040444	0.435130
Complication_risk_numeric	-0.026769	-0.101371	-0.070682	-0.220261	-0.220000
Arthritis_numeric	-0.169559	-0.078617	-0.158997	-0.032542	0.260781
Diabetes_numeric	0.164254	-0.046727	0.067213	0.448329	-0.164821
Hyperlipidemia_numeric	-0.058372	-0.400797	0.374411	0.211075	0.080255
BackPain_numeric	-0.202252	-0.095895	0.236883	-0.071264	-0.040638
Allergic_rhinitis_numeric	0.331937	0.225528	0.566124	-0.252433	0.069993
Reflux_esophagitis_numeric	0.098464	-0.189384	-0.021739	-0.020853	0.170788
Asthma_numeric	-0.016352	0.322078	-0.316182	-0.086194	-0.191273
Services_numeric	-0.007358	0.000096	0.003527	-0.021920	0.016814
Education_numeric	0.444033	-0.238318	-0.060882	-0.354637	-0.377126
Marital_numeric	0.048836	-0.351229	-0.234861	0.373097	-0.028496

	PC16	PC17	PC18	PC19 \
ReAdmis_numeric	-0.050733	-0.058256	-0.108383	0.244485
Services_numeric	-0.002500	0.008656	-0.000445	0.003786
Age	0.062460	0.164435	-0.238058	-0.071583
Gender_numeric	0.275468	-0.028774	0.269179	0.252051
Doc_visits	-0.116821	0.541442	-0.035362	0.201534
Soft_drink_numeric	0.376105	-0.450375	-0.065626	0.189284
Initial_admin_numeric	-0.047727	-0.267675	-0.149689	0.299899
HighBlood_numeric	0.604621	0.125542	-0.007621	-0.176800
Stroke_numeric	-0.028099	0.008687	0.177146	0.171509
Complication_risk_numeric	-0.401106	-0.207756	0.276240	0.074581
Arthritis_numeric	0.045635	-0.178694	0.480117	0.200406
Diabetes_numeric	-0.238413	-0.191279	-0.025169	-0.327723
Hyperlipidemia_numeric	-0.008969	0.305429	0.105164	0.367868
BackPain_numeric	-0.225011	0.043870	0.458607	-0.344312

```
Allergic_rhinitis_numeric  -0.177501 -0.073691 -0.092135  0.137754
Reflux_esophagitis_numeric -0.224839  0.037990 -0.449507 -0.028934
Asthma_numeric             -0.085380  0.278721  0.054463  0.338494
Services_numeric           -0.002500  0.008656 -0.000445  0.003786
Education_numeric          -0.021570  0.108999  0.140777  0.073220
Marital_numeric            -0.183192 -0.285961 -0.193797  0.312618
```

```
                                PC20
ReAdmis_numeric            -3.569954e-17
Services_numeric            7.071068e-01
Age                         1.543904e-16
Gender_numeric              1.110223e-16
Doc_visits                 -2.081668e-17
Soft_drink_numeric          1.110223e-16
Initial_admin_numeric       8.326673e-17
HighBlood_numeric          -1.040834e-17
Stroke_numeric              5.724587e-17
Complication_risk_numeric  -9.714451e-17
Arthritis_numeric          -2.775558e-17
Diabetes_numeric           -3.469447e-17
Hyperlipidemia_numeric     -3.469447e-17
BackPain_numeric            2.775558e-17
Allergic_rhinitis_numeric   4.857226e-17
Reflux_esophagitis_numeric -8.326673e-17
Asthma_numeric              2.220446e-16
Services_numeric           -7.071068e-01
Education_numeric           4.510281e-17
Marital_numeric            -2.775558e-17
```



In [ ]: