

```

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

import seaborn as sns

df = pd.read_csv("diabetes_012_health_indicators_BRFSS2015.csv")

print("Missing Values Before Handling:") print(df.isnull().sum())

numerical_cols = df.select_dtypes(include=np.number).columns

def detect_outliers_iqr(data):    Q1 =
data.quantile(0.25)
    Q3 = data.quantile(0.75)    IQR = Q3 -
Q1
    lower_bound = Q1 - 1.5 * IQR    upper_bound = Q3 +
1.5 * IQR
    outliers = data[(data < lower_bound) | (data > upper_bound)]    return outliers

for col in numerical_cols:
    outliers = detect_outliers_iqr(df[col])    if not
outliers.empty:
        median = df[col].median()    df.loc[outliers.index, col] = median
print(f"Outliers handled in column: {col}")

plt.figure(figsize=(8, 6))
sns.countplot(x='Diabetes_012', data=df) plt.title('Distribution of Diabetes (0,
1, 2)') plt.show()

plt.figure(figsize=(15, 12))
sns.heatmap(df.corr(), annot=True, cmap='coolwarm', fmt=".2f") plt.title('Correlation
Heatmap') plt.show()

plt.figure(figsize=(8, 6))
sns.boxplot(x='BMI', data=df)
plt.title('Boxplot of BMI') plt.show()

plt.figure(figsize=(8, 6))
sns.histplot(df['Age'], bins=20, kde=True)
plt.title('Distribution of Age') plt.show()

plt.figure(figsize=(8, 6))
sns.scatterplot(x='BMI', y='PhysHlth', data=df) plt.title('BMI vs.
Physical Health') plt.show()

```

```
df.to_csv("diabetes_012_health_indicators_BRFSS2015_modified.csv", index=False)
```

```

1 import pandas as pd
2 import numpy as np
3 import matplotlib.pyplot as plt
4 import seaborn as sns
5
6
7 df = pd.read_csv("diabetes_012_health_indicators_BRFSS2015.csv")
8
9
10 print("Missing Values Before Handling:")
11 print(df.isnull().sum())
12
13
14 numerical_cols = df.select_dtypes(include=np.number).columns
15
16 def detect_outliers_iqr(data):
17     Q1 = data.quantile(0.25)
18     Q3 = data.quantile(0.75)
19     IQR = Q3 - Q1
20     lower_bound = Q1 - 1.5 * IQR
21     upper_bound = Q3 + 1.5 * IQR
22     outliers = data[(data < lower_bound) | (data > upper_bound)]
23     return outliers
24
25 for col in numerical_cols:
26     outliers = detect_outliers_iqr(df[col])
27     if not outliers.empty:
28         median = df[col].median()
29         df.loc[outliers.index, col] = median
30         print(f"Outliers handled in column: {col}")
31
32
33 plt.figure(figsize=(8, 6))
34 sns.countplot(x='Diabetes_012', data=df)
35 plt.title('Distribution of Diabetes (0, 1, 2)')
36 plt.show()
37
38
39 plt.figure(figsize=(15, 12))
40 sns.heatmap(df.corr(), annot=True, cmap='coolwarm', fmt=".2f")
41 plt.title('Correlation Heatmap')
42
43 plt.figure(figsize=(8, 6))
44 sns.histplot(df['Age'], bins=20, kde=True)
45 plt.title('Distribution of Age')
46 plt.show()
47
48
49 plt.figure(figsize=(8, 6))
50 sns.scatterplot(x='BMI', y='PhysHlth', data=df)
51 plt.title('BMI vs. Physical Health')
52 plt.show()
53
54 df.to_csv("diabetes_012_health_indicators_BRFSS2015_modified.csv", index=False)

```

OUT PUT

Missing Values Before Handling:

Diabetes_012 0

HighBP 0

HighChol 0

CholCheck 0

BMI 0

Smoker 0

Stroke 0

HeartDiseaseorAttack 0

PhysActivity 0

Fruits 0

Veggies 0

HvyAlcoholConsump 0

AnyHealthcare 0

NoDocbcCost 0

GenHlth 0

MentHlth 0

PhysHlth 0

DiffWalk 0

Sex 0

Age 0

Education 0

Income 0

dtype: int64

Outliers handled in column: Diabetes_012

Outliers handled in column: CholCheck

Outliers handled in column: BMI

Outliers handled in column: Stroke

Outliers handled in column:

HeartDiseaseorAttack

Outliers handled in column: PhysActivity

Outliers handled in column: Veggies

Outliers handled in column:

HvyAlcoholConsump

Outliers handled in column:

AnyHealthcare

Outliers handled in column:

NoDocbcCost

Outliers handled in column: GenHlth

Outliers handled in column: MentHlth

Outliers handled in column: PhysHlth

Outliers handled in column: DiffWalk

Console 1/A X Console 2/A X

Missing Values Before Handling:

Diabetes_012	0
HighBP	0
HighChol	0
CholCheck	0
BMI	0
Smoker	0
Stroke	0
HeartDiseaseorAttack	0
PhysActivity	0
Fruits	0
Veggies	0
HvyAlcoholConsump	0
AnyHealthcare	0
NoDocbcCost	0
GenHlth	0
MentHlth	0
PhysHlth	0
DiffWalk	0
Sex	0
Age	0
Education	0
Income	0

dtype: int64

Outliers handled in column: Diabetes_012

Outliers handled in column: CholCheck

Outliers handled in column: BMI

Outliers handled in column: Stroke

Outliers handled in column: HeartDiseaseorAttack

Outliers handled in column: PhysActivity

Outliers handled in column: Veggies

Outliers handled in column: HvyAlcoholConsump

Outliers handled in column: AnyHealthcare

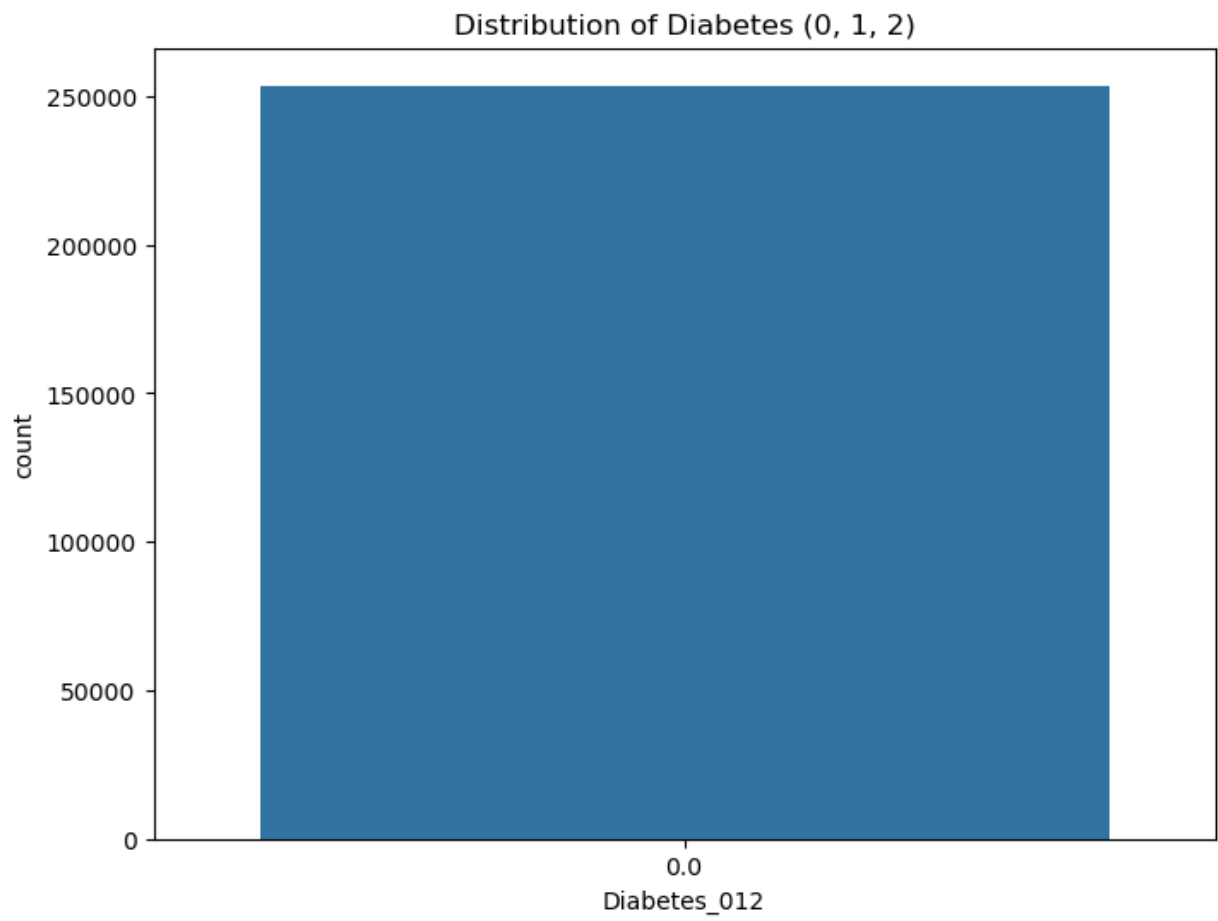
Outliers handled in column: NoDocbcCost

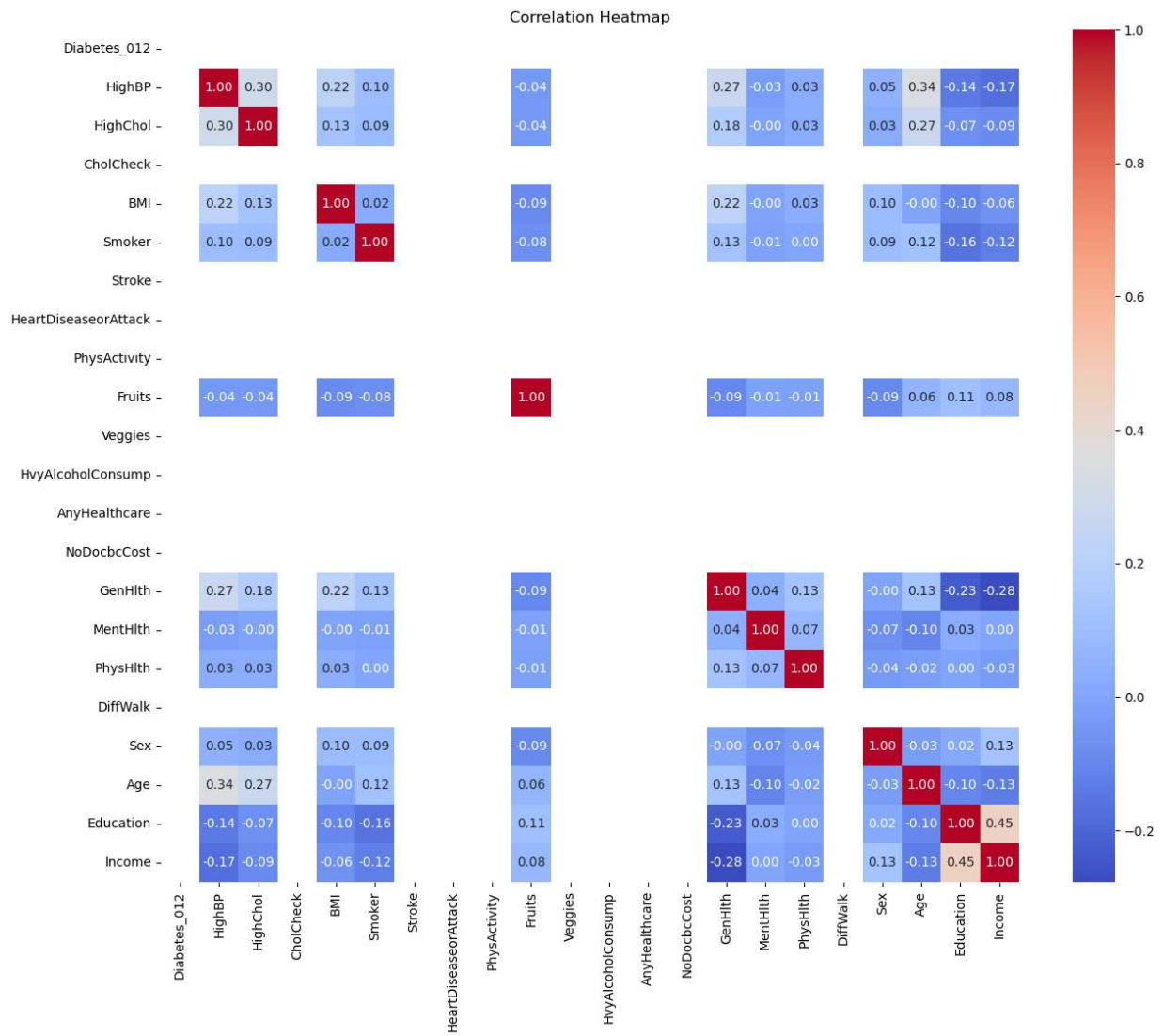
Outliers handled in column: GenHlth

Outliers handled in column: MentHlth

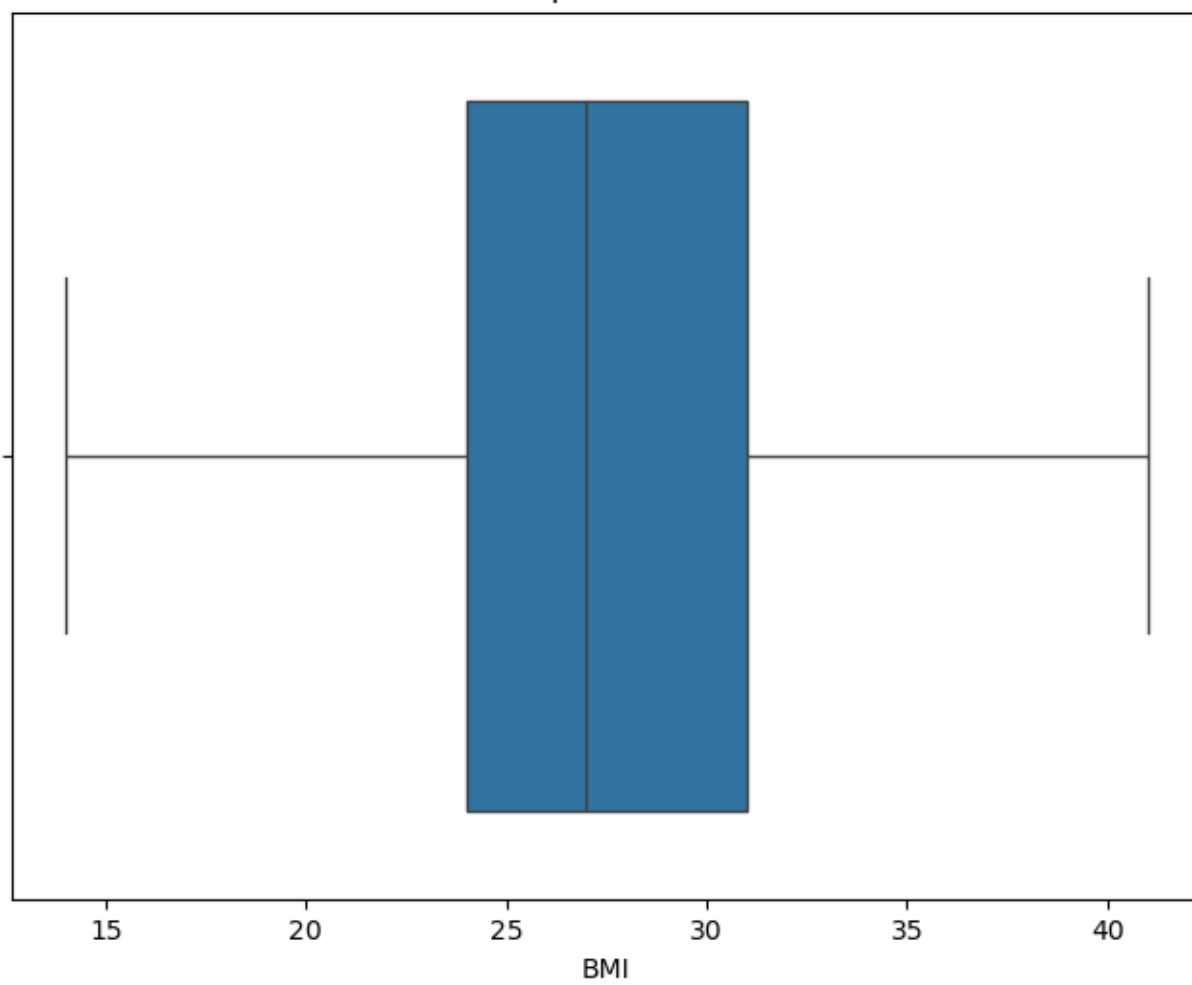
Outliers handled in column: PhysHlth

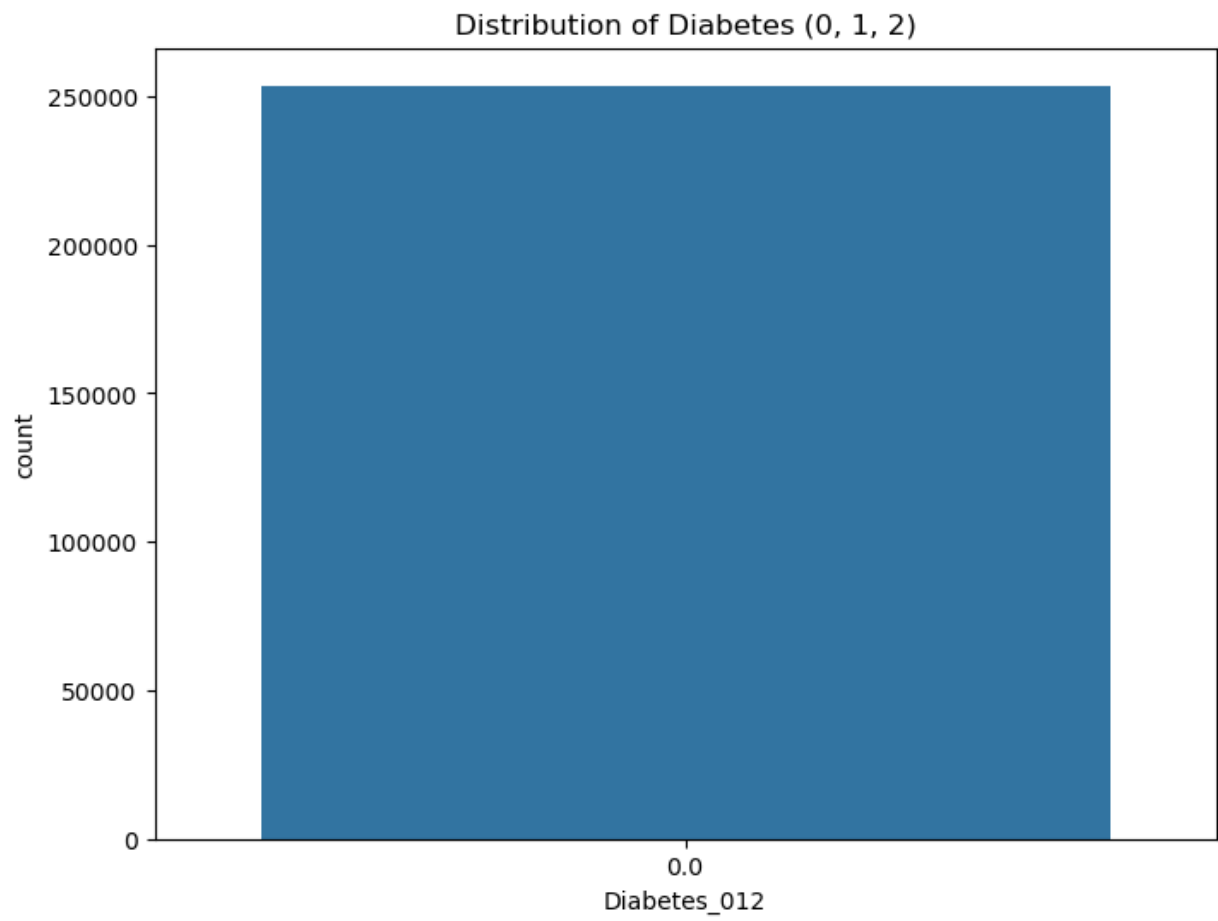
Outliers handled in column: DiffWalk

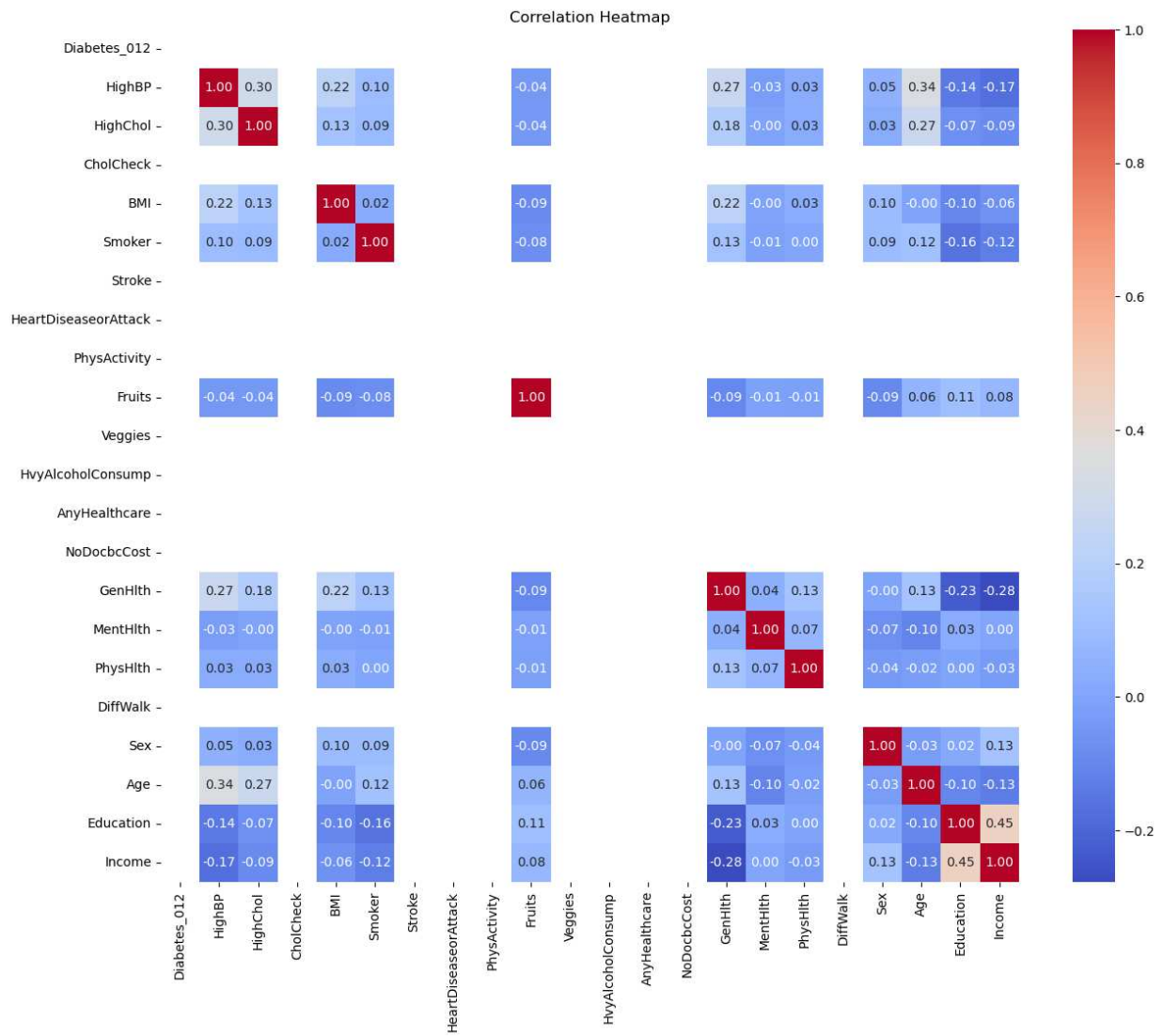




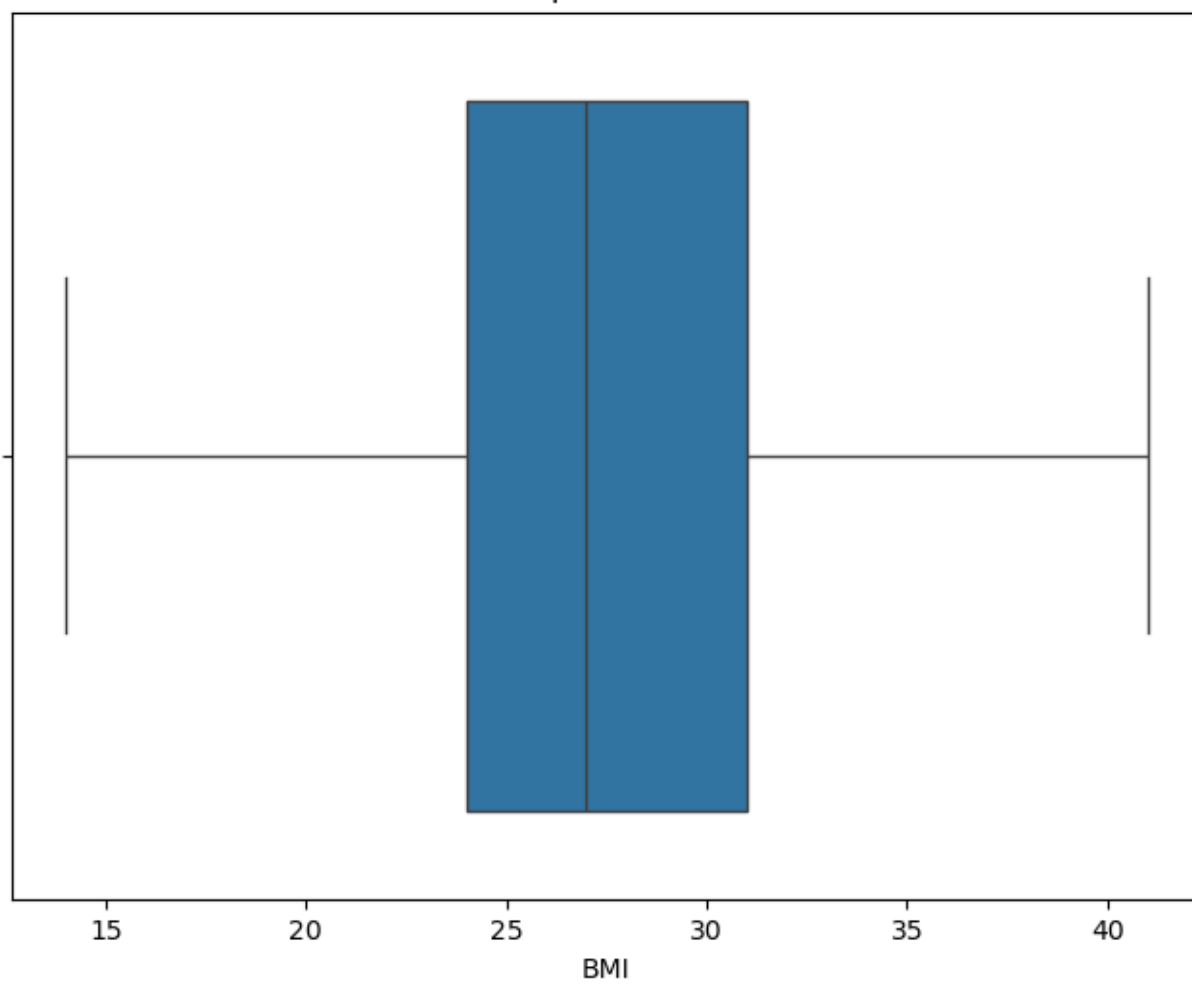
Boxplot of BMI

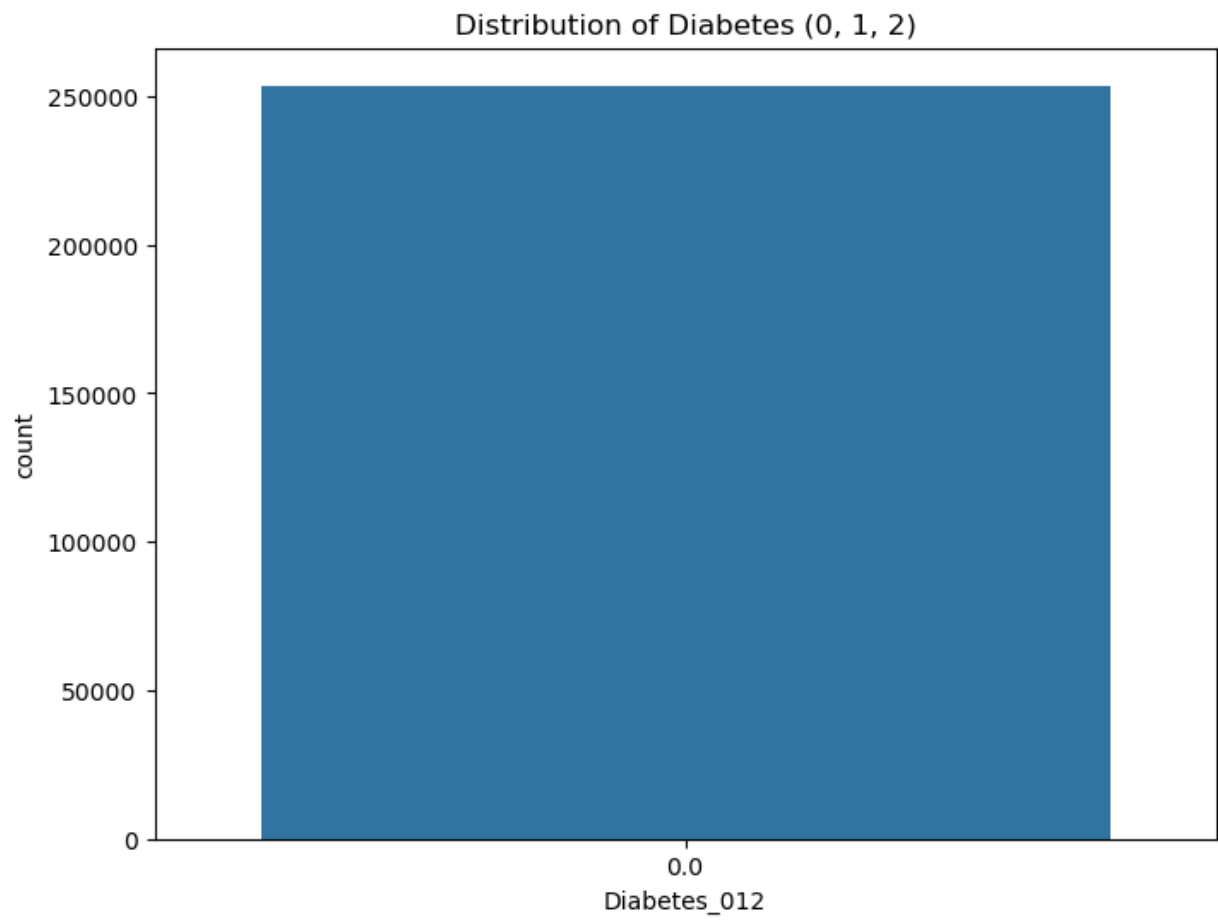


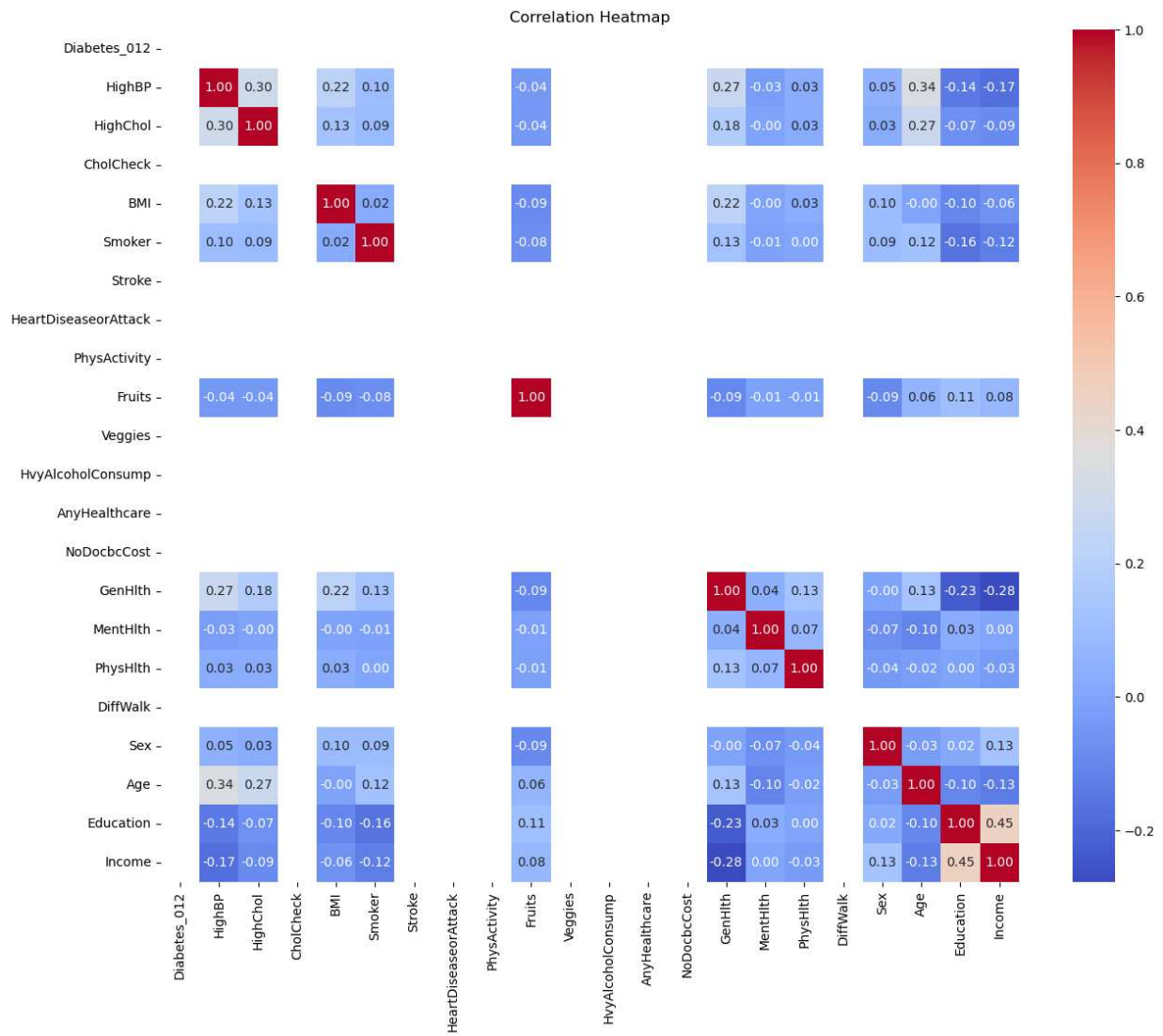




Boxplot of BMI







Boxplot of BMI

