# Cardiovascular disease prediction - Code

# Import libraries

import numpy as np

import pandas as pd

import seaborn as sns

import matplotlib.pyplot as plt

from sklearn.model\_selection import train\_test\_split

from sklearn.preprocessing import StandardScaler, LabelEncoder, OneHotEncoder

from sklearn.ensemble import RandomForestClassifier

from sklearn.metrics import accuracy\_score, classification\_report, confusion\_matrix

from sklearn.pipeline import Pipeline

from sklearn.compose import ColumnTransformer

# Load the data

from google.colab import files

uploaded = files.upload()

# Read the data into a DataFrame

df = pd.read\_csv('CVD\_cleaned.csv')

# Display the first 10 rows of the DataFrame

print(df.head(10))

# Get the shape of the data

print(df.shape)

# Count the empty values in each column

print(df.isna().sum())

# View some basic statistics

print(df.describe())

# Get a count of the number of patients with Heart\_Disease

print(df['Heart\_Disease'].value\_counts())

# Visualize the count of patients with and without Heart Disease

plt.figure(figsize=(8, 6))

sns.countplot(x='Heart\_Disease', data=df)

plt.xlabel('Heart Disease')

plt.ylabel('Count')

plt.title('Count of Patients with and without Heart Disease')

plt.show()

# Create a bar chart of the number of patients in each age category

plt.figure(figsize=(10, 6))

df['Age\_Category'].value\_counts().plot(kind='bar')

plt.xlabel('Age Category')

plt.ylabel('Number of Patients')

plt.title('Age Category of Patients with Heart Disease')

plt.show()

# Create a pie chart of the percentage of patients in each age category

plt.figure(figsize=(8, 8))

df['Age\_Category'].value\_counts(normalize=True).plot(kind='pie', autopct='%1.1f%%')

plt.xlabel('Age Category')

plt.ylabel('Percentage of Patients')

plt.title('Age Category of Patients with Heart Disease')

plt.show()

# Get the correlation of the columns

print(df.corr())

# Visualize the correlation matrix

plt.figure(figsize=(20, 20))

sns.heatmap(df.corr(), annot=True, fmt='.2f', cmap='coolwarm')

plt.title('Correlation Matrix')

plt.show()

# Output

Saving CVD\_cleaned.csv to CVD\_cleaned.csv

Name General\_Health Checkup Exercise \

0 Isabelle Watson Poor Within the past 2 years No

1 Xander Delgado Very Good Within the past year No

2 Evelynn Sweeney Very Good Within the past year Yes

3 Amirah Green Poor Within the past year Yes

4 Aisha Valencia Good Within the past year No

5 Jagger Casey Good Within the past year No

6 Amara Huerta Fair Within the past year Yes

7 Khalid Dougherty Good Within the past year Yes

8 Kaya Horn Fair Within the past year No

9 Kylan Charles Fair Within the past year No

Heart\_Disease Skin\_Cancer Other\_Cancer Depression Diabetes Arthritis \

0 No No No No No Yes

1 Yes No No No Yes No

2 No No No No Yes No

3 Yes No No No Yes No

4 No No No No No No

5 No No No Yes No Yes

6 Yes No No No No Yes

7 No No No No No Yes

8 No No No Yes No No

9 No No No No Yes Yes

Sex Age\_Category Height\_(cm) Weight\_(kg) BMI Smoking\_History \

0 Female 70-74 150 32.66 14.54 Yes

1 Female 70-74 165 77.11 28.29 No

2 Female 60-64 163 88.45 33.47 No

3 Male 75-79 180 93.44 28.73 No

4 Male 80+ 191 88.45 24.37 Yes

5 Male 60-64 183 154.22 46.11 No

6 Male 60-64 175 69.85 22.74 Yes

7 Female 65-69 165 108.86 39.94 Yes

8 Female 65-69 163 72.57 27.46 Yes

9 Female 70-74 163 91.63 34.67 No

Alcohol\_Consumption Fruit\_Consumption Green\_Vegetables\_Consumption \

0 0 30 16

1 0 30 0

2 4 12 3

3 0 30 30

4 0 8 4

5 0 12 12

6 0 16 8

7 3 30 8

8 0 12 12

9 0 12 12

FriedPotato\_Consumption

0 12

1 4

2 16

3 8

4 0

5 12

6 0

7 8

8 4

9 1

(150, 20)

Name 0

General\_Health 0

Checkup 0

Exercise 0

Heart\_Disease 0

Skin\_Cancer 0

Other\_Cancer 0

Depression 0

Diabetes 0

Arthritis 0

Sex 0

Age\_Category 0

Height\_(cm) 0

Weight\_(kg) 0

BMI 0

Smoking\_History 0

Alcohol\_Consumption 0

Fruit\_Consumption 0

Green\_Vegetables\_Consumption 0

FriedPotato\_Consumption 0

dtype: int64

Height\_(cm) Weight\_(kg) BMI Alcohol\_Consumption \

count 150.000000 150.000000 150.000000 150.000000

mean 168.020000 80.279467 28.317267 2.026667

std 9.886246 19.254784 5.721927 5.695809

min 150.000000 32.660000 14.060000 0.000000

25% 160.000000 68.040000 24.510000 0.000000

50% 165.000000 79.830000 27.910000 0.000000

75% 175.000000 89.582500 31.232500 0.000000

max 196.000000 154.220000 46.110000 30.000000

Fruit\_Consumption Green\_Vegetables\_Consumption \

count 150.000000 150.000000

mean 21.133333 10.920000

std 19.759516 10.381657

min 0.000000 0.000000

25% 8.000000 4.000000

50% 16.000000 8.000000

75% 30.000000 16.000000

max 90.000000 60.000000

FriedPotato\_Consumption

count 150.000000

mean 4.740000

std 5.815312

min 0.000000

25% 1.000000

50% 3.000000

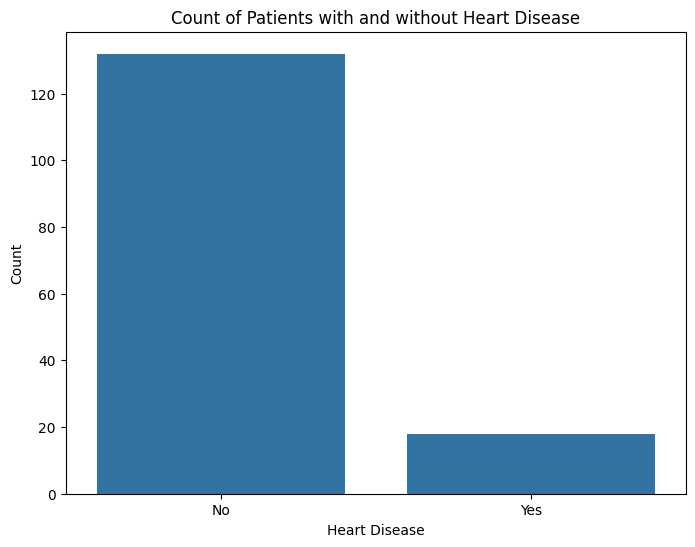
75% 8.000000

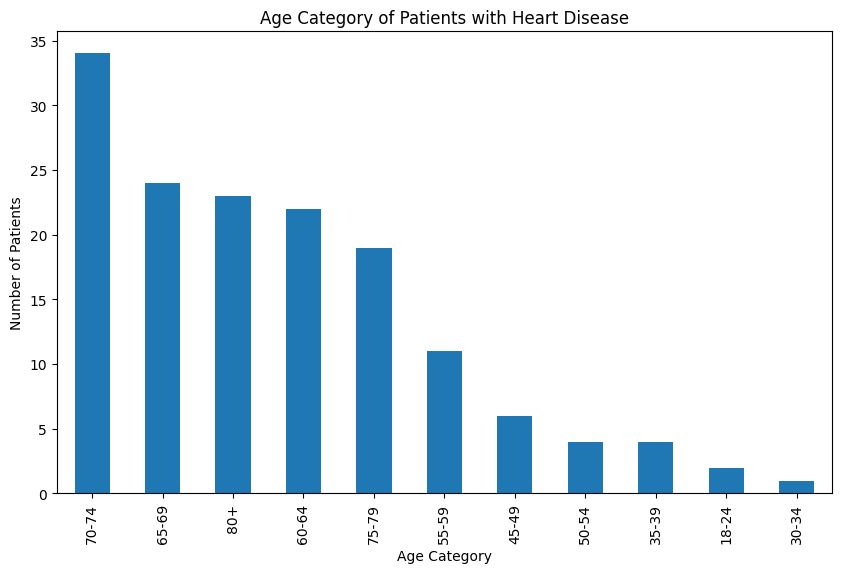
max 30.000000

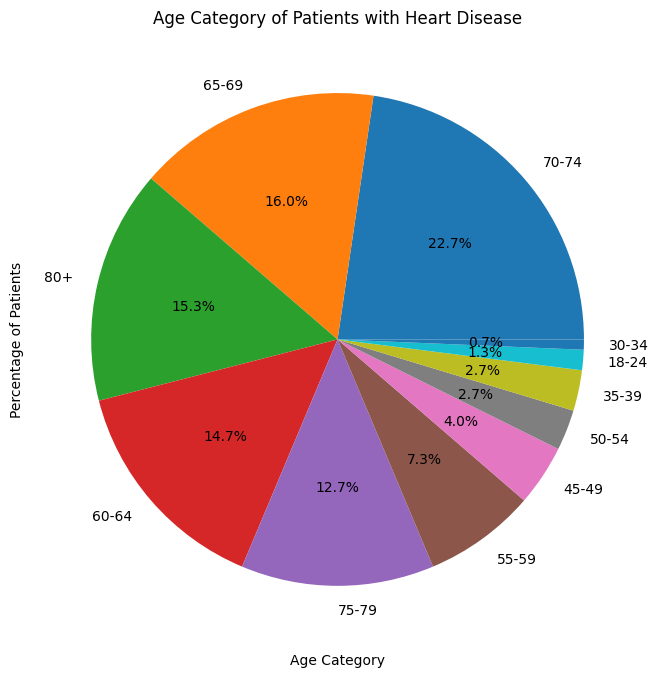
No 132

Yes 18

Name: Heart\_Disease, dtype: int64







<ipython-input-2-ea21368c8d27>:60: FutureWarning: The default value of numeric\_only in DataFrame.corr is deprecated. In a future version, it will default to False. Select only valid columns or specify the value of numeric\_only to silence this warning.

print(df.corr())

<ipython-input-2-ea21368c8d27>:64: FutureWarning: The default value of numeric\_only in DataFrame.corr is deprecated. In a future version, it will default to False. Select only valid columns or specify the value of numeric\_only to silence this warning.

sns.heatmap(df.corr(), annot=True, fmt='.2f', cmap='coolwarm')

Height\_(cm) Weight\_(kg) BMI \

Height\_(cm) 1.000000 0.539424 0.069849

Weight\_(kg) 0.539424 1.000000 0.869939

BMI 0.069849 0.869939 1.000000

Alcohol\_Consumption 0.081156 0.044487 0.005126

Fruit\_Consumption -0.118955 -0.159535 -0.106356

Green\_Vegetables\_Consumption -0.142667 -0.106666 -0.039998

FriedPotato\_Consumption 0.136207 0.179911 0.125331

Alcohol\_Consumption Fruit\_Consumption \

Height\_(cm) 0.081156 -0.118955

Weight\_(kg) 0.044487 -0.159535

BMI 0.005126 -0.106356

Alcohol\_Consumption 1.000000 -0.089301

Fruit\_Consumption -0.089301 1.000000

Green\_Vegetables\_Consumption -0.010179 0.289792

FriedPotato\_Consumption 0.049448 0.006612

Green\_Vegetables\_Consumption \

Height\_(cm) -0.142667

Weight\_(kg) -0.106666

BMI -0.039998

Alcohol\_Consumption -0.010179

Fruit\_Consumption 0.289792

Green\_Vegetables\_Consumption 1.000000

FriedPotato\_Consumption 0.073134

FriedPotato\_Consumption

Height\_(cm) 0.136207

Weight\_(kg) 0.179911

BMI 0.125331

Alcohol\_Consumption 0.049448

Fruit\_Consumption 0.006612

Green\_Vegetables\_Consumption 0.073134

FriedPotato\_Consumption 1.000000

