import pandas as pd

import numpy as np

data = pd.read\_csv("heart.csv")

data.head()

check the number of null values

data.isnull().sum()

data.loc[1,"Age"] = np.nan

data.loc[4,"ID"] = np.nan

data.loc[5,"Age"] = np.nan

data.loc[6,"Age"] = np.nan

data.loc[5,"Gender"] = np.nan

adding an outlier

data.loc[7,"Age"] = 800

# M1 : IQR

Q1 = data["Age"].quantile(0.25)

Q3 = data["Age"].quantile(0.75)

IQR = Q3 -Q1

lower\_bound = Q1 - 1.5 \* IQR

upper\_bound = Q3 + 1.5 \* IQR

outliers = data[(data["Age"] < lower\_bound) | (data["Age"] > upper\_bound)]

print("Detected Outliers:", outliers)

# M2 : Z Score

as the 7th is the outlier print the 7th row and as the value is 50 i.e. greater that +-3 it is the outlier

z\_score =( data["Age"] - data["Age"].mean()) / data["Age"].std()

print(z\_score.head(8))

# M3 : Binning

## width (based on range of each bin)

age\_min = min(data["Age"])

age\_max = max(data["Age"])

age\_width = age\_max - age\_min

print("min age" , age\_min ,"max age" , age\_max , "width" , age\_width)

bins = 3

bin\_width = (age\_width/bins)

print("bin width is " , bin\_width)

bin\_label = ("l1" , "l2" , "l3")

equal\_bin\_width = pd.cut(data["Age"] , bins , labels = bin\_label)

print(equal\_bin\_width.value\_counts())

## Frequency (based on the no.of each bin element)

frequency = 3

equal\_bin\_frequency = pd.qcut(data["Age"] , frequency , duplicates = "drop")

print(equal\_bin\_frequency.value\_counts())

# M4 : visualization

import seaborn as sns

sns.boxplot(data["Age"])

sns.histplot(data["Age"])

sns.scatterplot(x=data["ID"], y= data["Age"],alpha=0.5, s=10, color='blue'

adding duplicate

data.loc[len(data)] = data.loc[0]

data

checking the duplicate row

print("Duplicate Rows:", data.duplicated())

droping the duplicate row

data = data.drop\_duplicates()

print("Duplicate Rows:", data.duplicated())