import pandas as pd

import numpy as np

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

data.head()

# Normalize caffine intake column

## min-max

min\_value = data["Caffeine Intake (mg/day)"].min()

max\_value = data["Caffeine Intake (mg/day)"].max()

for i in data["Caffeine Intake (mg/day)"]:

normalized = (i - min\_value) / (max\_value - min\_value)

print(normalized)

## z-score

sd = data["Caffeine Intake (mg/day)"].std()

mean = data["Caffeine Intake (mg/day)"].mean()

for i in data["Caffeine Intake (mg/day)"]:

normalized = (i - mean )/ sd

print(normalized)

## Decimal Scaling

import math

max\_abs\_value = data["Caffeine Intake (mg/day)"].abs().max()

j = math.ceil(math.log10(max\_abs\_value))

for i in data["Caffeine Intake (mg/day)"]:

normalized = i / 10\*\*j

print(normalized)

## Robust Scaling

introduce outlier

data.loc[3,"Caffeine Intake (mg/day)"] = 1000000

robust scaling

median = data["Caffeine Intake (mg/day)"].median()

Q1 = data["Caffeine Intake (mg/day)"].quantile(0.25)

Q3 = data["Caffeine Intake (mg/day)"].quantile(0.75)

IQR = Q3 - Q1

for i in data["Caffeine Intake (mg/day)"]:

normalized = (i - median) / IQR

print(normalized)