import numpy as np

import pandas as pd

import statsmodels.api as sm

import matplotlib.pyplot as plt

# 1) Download the dataset and confirm theres no missing data

mydf = pd.read\_csv('Environs.csv')

print("1) Download the dataset and confirm theres no missing data \*\*\*\*\*\*\*\*\*\*")

print(mydf.info())

# 2) Find the best predictor for Design

print("\n\n2) Find the best predictor for Design \*\*\*\*\*\*\*\*\*\*")

y = mydf.Design

##X = mydf.Pressure

##X = mydf.Time

##X = mydf['Heat Factor']

##X = mydf[ ['Time', 'Pressure'] ]

##X = mydf[ ['Time', 'Heat Factor'] ]

##X = mydf[ ['Pressure', 'Heat Factor'] ]

X = mydf[ ['Pressure', 'Heat Factor', 'Time'] ]

X = sm.add\_constant(X)

LRModel\_Radio = sm.OLS(y,X).fit()

print(LRModel\_Radio.summary())

# 3) Finding the equation

y\_int = 389.1659

pressure\_slope = 5.3185

heat\_slope = -3.0165

time\_slope = 2.1247

print("\n\n3) Find the equation to calculate the design value \*\*\*\*\*\*\*\*\*\*")

print('y =', pressure\_slope, '\* pressure +', time\_slope, '\* time +', heat\_slope, '\* heat factor +', y\_int)

# 4) Predicting the Design value for the three features

a\_values = [34.85, 31.84, 38.0]

b\_values = [34.6, 35.56, 33.99]

c\_values = [128.48, 132.08, 133.0]

y\_predicts = []

def predict(a,b,c):

y = pressure\_slope \* a + heat\_slope \* b + time\_slope \* c + y\_int

return y

for i in range(len(a\_values)):

a = a\_values[i]

b = b\_values[i]

c = c\_values[i]

y\_predicts.append(predict(a,b,c))

print("\n\n4) Use the equation to predict the design value \*\*\*\*\*\*\*\*\*\*")

print('Predicted values:', y\_predicts)