**Problem Statement:-**

A shoe size is an alphanumerical indication of the fitting size of a shoe for a person. Often it just consists of a number indicating the length because many shoemakers only provide a standard width for economic reasons. Several different shoe-size systems are used worldwide. These systems define in what they measure, what unit of measurement they use, and where the size 0 (or 1) is positioned. Only a few systems also take the width of the feet into account. Some regions use different shoe-size systems for different types of shoes (e.g., men’s, women’s, children’s, sport, or safety shoes). Perform a survey to collect data regarding the Height of a person and the Shoe size (at least 100 persons), anonymously. Perform linear regression over the data collected using Least Square Estimation.

**Calculation:-**

The dataset and the Table Are written in Excel Sheet Named **‘Dataset.xlsx** Sheet Name Is **‘Table’.**

**For US Size Prediction:-**

**No.of Observations** (n) = 109

**∑xi(**Sum of Height) **=** 18901

**∑yi(**Sum of Shoe Sizes) **=** 866.5

**=** 173.4036697

**=** 7.949541284

**∑xi2 =** 3297423

**∑yi2 =** 7126.75

**∑xi \* yi =** 151112.5

**∑ei2 =** 201.498

**Sxy**= ΣXiYi- [(ΣXi)(ΣYi) / n ]

=151112.5 – [(173.4036697)(7.949541284) / 109 ]

= 858.2201835

**Sxx** =ΣXi 2 - (ΣXi) 2 / n

**=** 3297423 – [30068.83267 / 109]

**=** 19920.23853

**B1 = Sxy /Sxx**

**=** 858.2201835 / 19920.23853

**=** 0.043082827

**B0 =** 1 / n ( ∑ yi – B1∑xi)

= [1 / 109 ] (866.5 – (0.043082827) (866.5)

= 0.478821041

**Regression Line :-**  = B0 + B1X

**=** 0.478821041 + 0.043082827(X)

**Graph:-**

**Code :-**

import matplotlib.pyplot as plt

import numpy as np

import pandas as pd

from sklearn import linear\_model

Dataset = pd.read\_excel("./Dataset.xlsx", "Python\_Data")

regression = linear\_model.LinearRegression()

regression.fit(Dataset[["Hight"]], Dataset["Shoe Size(US)"])

print(f"B0 = {regression.intercept\_} \nB1 = {regression.coef\_}")

plt.figure(figsize=(18, 9))

plt.scatter(x=Dataset["Hight"], y=Dataset["Shoe Size(US)"], color="red", marker="\*")

plt.plot(Dataset["Hight"], regression.predict(Dataset[["Hight"]]), color="blue")

plt.xlabel("Hight")

plt.ylabel("Shoe Size")

plt.title("Shoe Size Prediction")

plt.show()

**Output :-**



