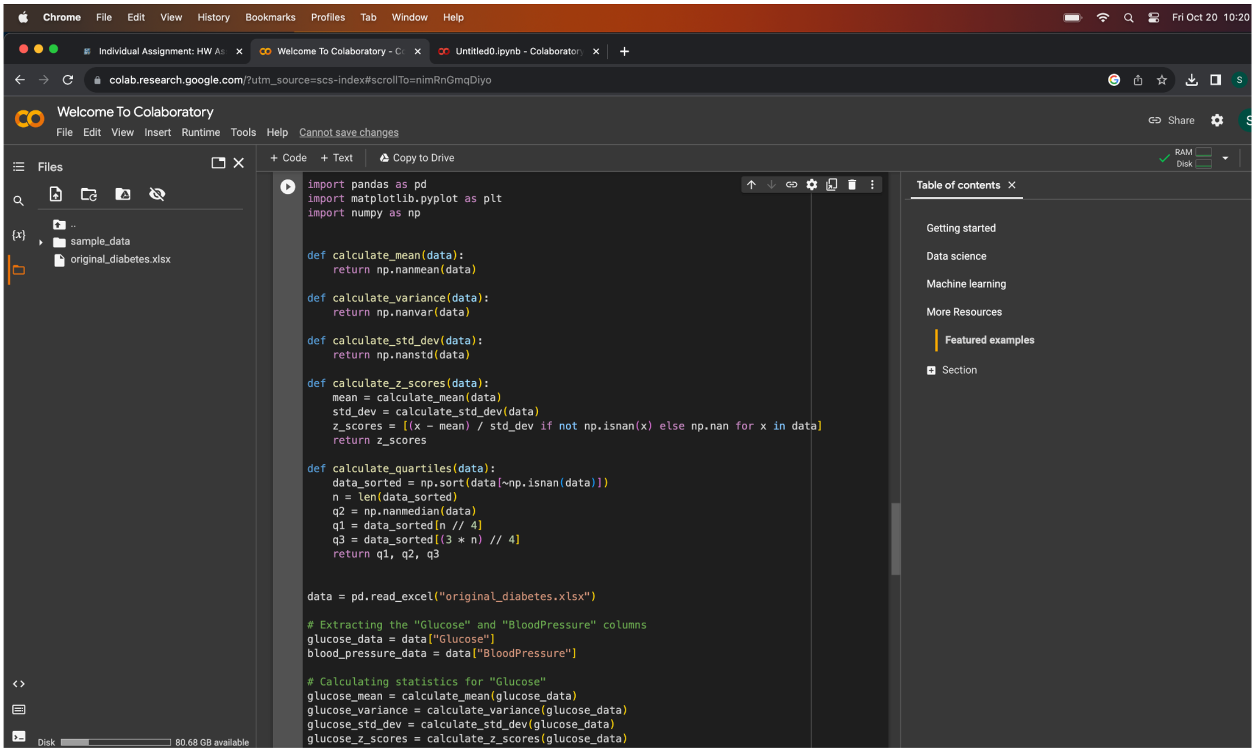
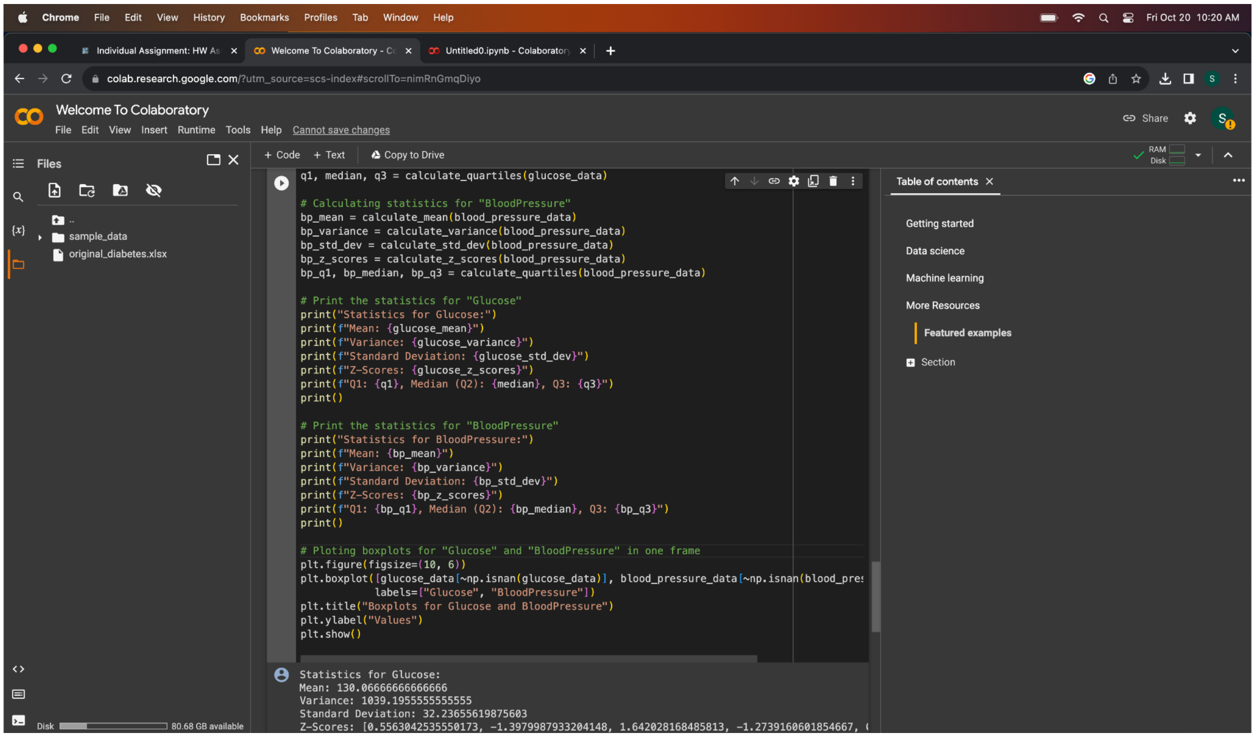
Question 1

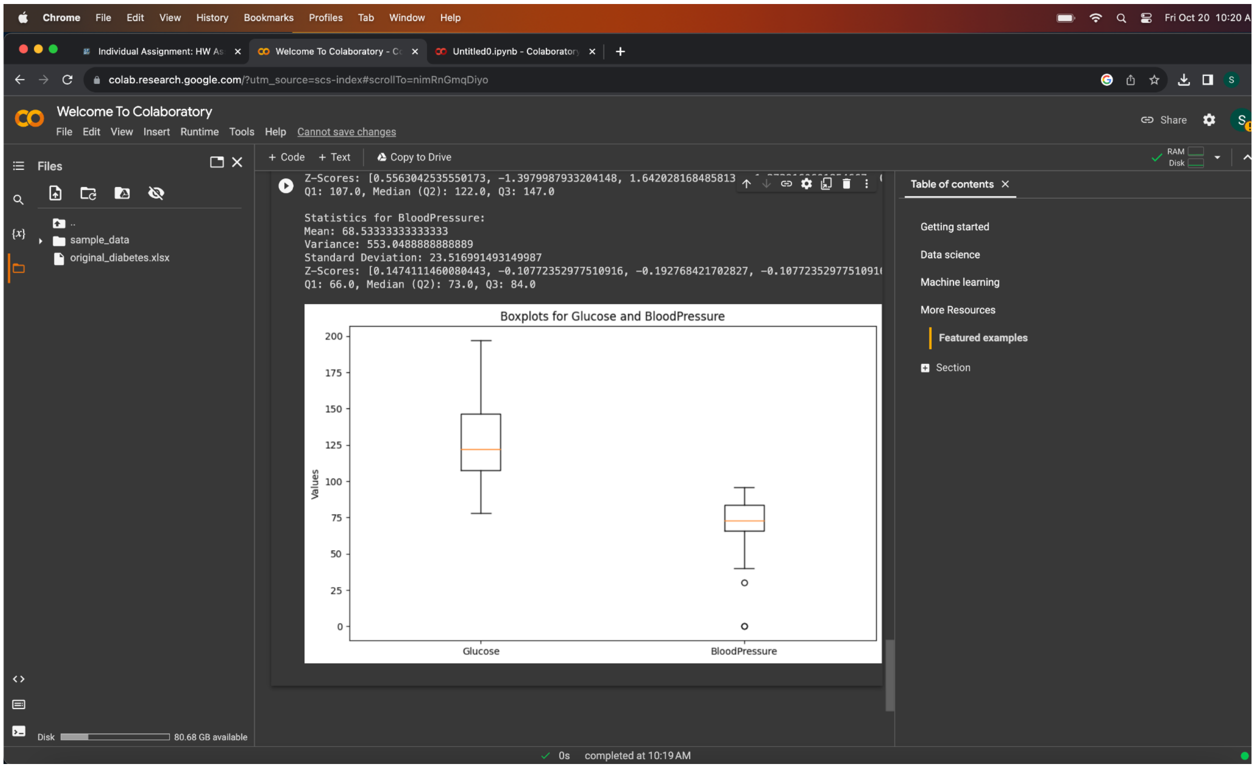

Chrome File Edit View History Bookmarks Profiles Tab Window Help
Fri Oct 20 10:20
Individual Assignment: HW As
Welcome To Colaboratory - Co
Untitled0.ipynb - Colaboratory
colab.research.google.com/?utm_source=scs-index#scrollTo=nimRnGmqDiyo
Welcome To Colaboratory
File Edit View Insert Runtime Tools Help Cannot save changes
Share
+ Code + Text Copy to Drive
RAM
Disk
Files
..
sample_data
original_diabetes.xlsx
Disk 80.68 GB available
Table of contents
Getting started
Data science
Machine learning
More Resources
Featured examples
Section

```
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np


def calculate_mean(data):
    return np.nanmean(data)

def calculate_variance(data):
    return np.nanvar(data)

def calculate_std_dev(data):
    return np.nanstd(data)

def calculate_z_scores(data):
    mean = calculate_mean(data)
    std_dev = calculate_std_dev(data)
    z_scores = [(x - mean) / std_dev if not np.isnan(x) else np.nan for x in data]
    return z_scores

def calculate_quartiles(data):
    data_sorted = np.sort(data[~np.isnan(data)])
    n = len(data_sorted)
    q2 = np.nanmedian(data)
    q1 = data_sorted[n // 4]
    q3 = data_sorted[(3 * n) // 4]
    return q1, q2, q3


data = pd.read_excel("original_diabetes.xlsx")

# Extracting the "Glucose" and "BloodPressure" columns
glucose_data = data["Glucose"]
blood_pressure_data = data["BloodPressure"]

# Calculating statistics for "Glucose"
glucose_mean = calculate_mean(glucose_data)
glucose_variance = calculate_variance(glucose_data)
glucose_std_dev = calculate_std_dev(glucose_data)
glucose_z_scores = calculate_z_scores(glucose_data)
```


Chrome File Edit View History Bookmarks Profiles Tab Window Help
Fri Oct 20 10:20 AM
Individual Assignment: HW As
Welcome To Colaboratory - Co
Untitled0.ipynb - Colaboratory
colab.research.google.com/?utm_source=scs-index#scrollTo=nimRnGmqDiyo
Welcome To Colaboratory
File Edit View Insert Runtime Tools Help Cannot save changes
Share
+ Code + Text Copy to Drive
RAM
Disk
Files
..
sample_data
original_diabetes.xlsx
Disk 80.68 GB available
Table of contents
Getting started
Data science
Machine learning
More Resources
Featured examples
Section

```
q1, median, q3 = calculate_quartiles(glucose_data)

# Calculating statistics for "BloodPressure"
bp_mean = calculate_mean(blood_pressure_data)
bp_variance = calculate_variance(blood_pressure_data)
bp_std_dev = calculate_std_dev(blood_pressure_data)
bp_z_scores = calculate_z_scores(blood_pressure_data)
bp_q1, bp_median, bp_q3 = calculate_quartiles(blood_pressure_data)

# Print the statistics for "Glucose"
print("Statistics for Glucose:")
print(f"Mean: {glucose_mean}")
print(f"Variance: {glucose_variance}")
print(f"Standard Deviation: {glucose_std_dev}")
print(f"Z-Scores: {glucose_z_scores}")
print(f"Q1: {q1}, Median (Q2): {median}, Q3: {q3}")
print()

# Print the statistics for "BloodPressure"
print("Statistics for BloodPressure:")
print(f"Mean: {bp_mean}")
print(f"Variance: {bp_variance}")
print(f"Standard Deviation: {bp_std_dev}")
print(f"Z-Scores: {bp_z_scores}")
print(f"Q1: {bp_q1}, Median (Q2): {bp_median}, Q3: {bp_q3}")
print()

# Ploting boxplots for "Glucose" and "BloodPressure" in one frame
plt.figure(figsize=(10, 6))
plt.boxplot([glucose_data[~np.isnan(glucose_data)], blood_pressure_data[~np.isnan(blood_pres
            labels=["Glucose", "BloodPressure"])
plt.title("Boxplots for Glucose and BloodPressure")
plt.ylabel("Values")
plt.show()
```

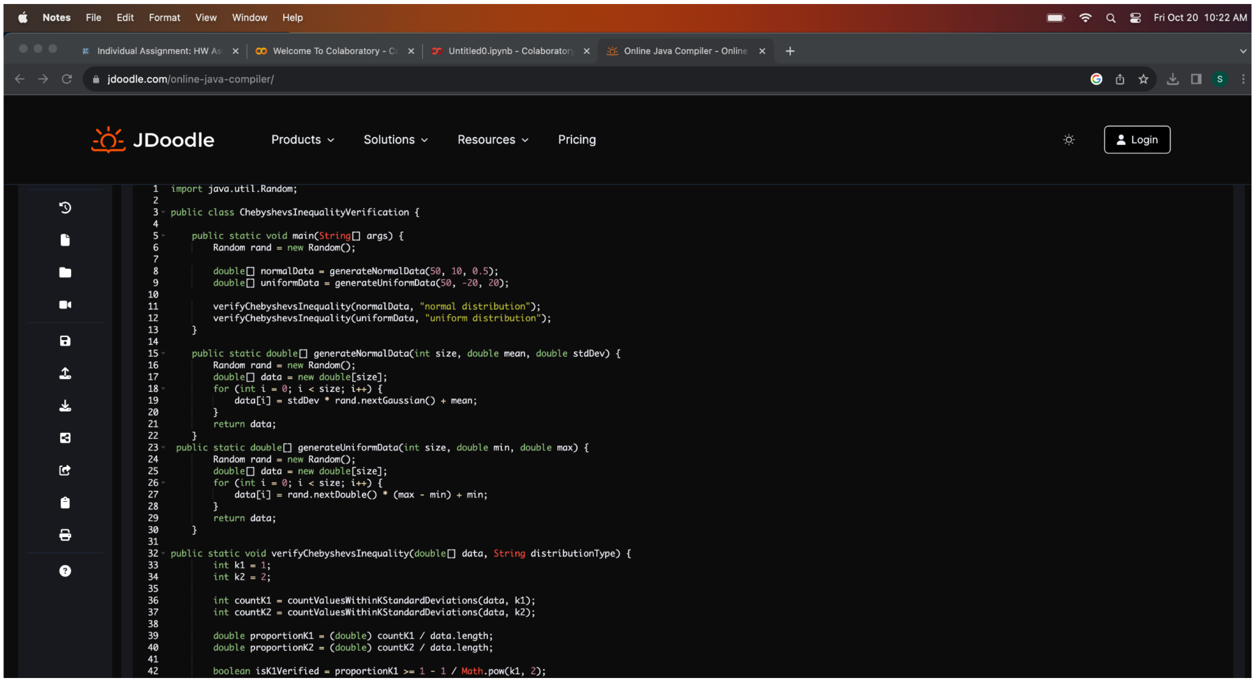
```
Statistics for Glucose:
Mean: 130.06666666666666
Variance: 1039.1955555555555
Standard Deviation: 32.23655619875603
Z-Scores: [0.5563042535550173, -1.3979987933204148, 1.642028168485813, -1.2739160601854667, (
```



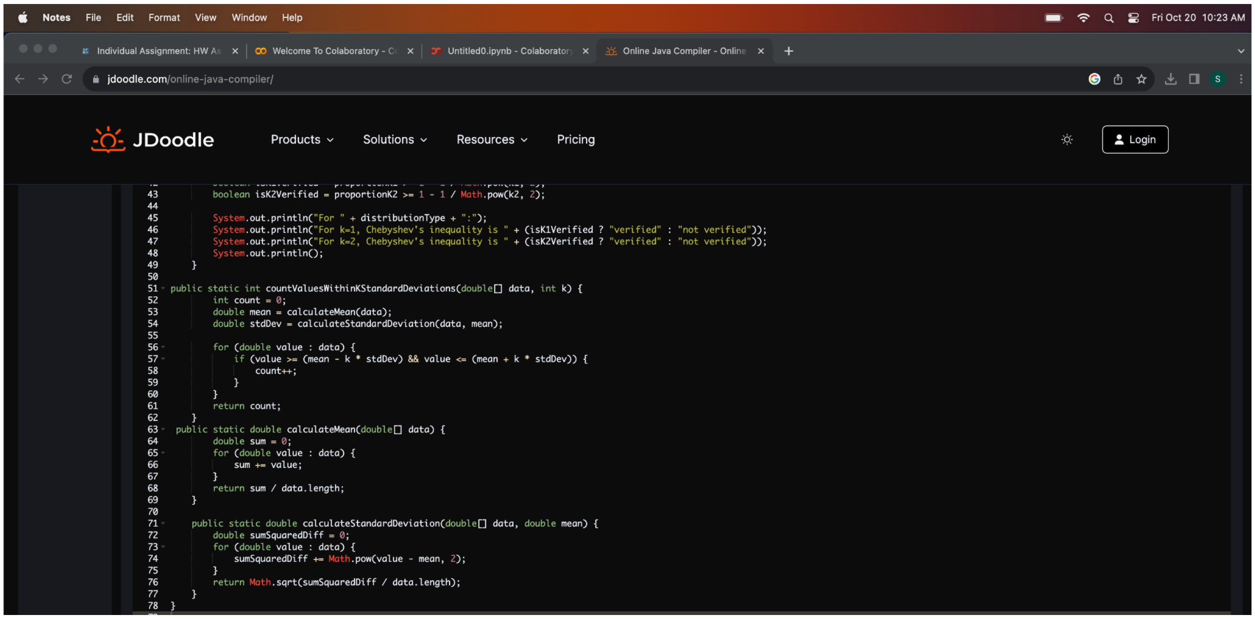
Chrome File Edit View History Bookmarks Profiles Tab Window Help
Fri Oct 20 10:20 A
Individual Assignment: HW As
Welcome To Colaboratory - C
Untitled0.ipynb - Colaborator
colab.research.google.com/?utm_source=scs-index#scrollTo=nimRnGmqDiyo
Welcome To Colaboratory
File Edit View Insert Runtime Tools Help Cannot save changes
Share
+ Code + Text
Copy to Drive
RAM
Disk
Files
..
sample_data
original_diabetes.xlsx
Disk 80.68 GB available
Table of contents
Getting started
Data science
Machine learning
More Resources
Featured examples
Section
Boxplots for Glucose and BloodPressure
Values
200
175
150
125
100
75
50
25
0
Glucose
BloodPressure
0s completed at 10:19 AM

```
Z-Scores: [0.5563042535550173, -1.3979987933204148, 1.642028168485813
Q1: 107.0, Median (Q2): 122.0, Q3: 147.0

Statistics for BloodPressure:
Mean: 68.53333333333333
Variance: 553.0488888888889
Standard Deviation: 23.516991493149987
Z-Scores: [0.1474111460080443, -0.10772352977510916, -0.192768421702827, -0.10772352977510916
Q1: 66.0, Median (Q2): 73.0, Q3: 84.0
```

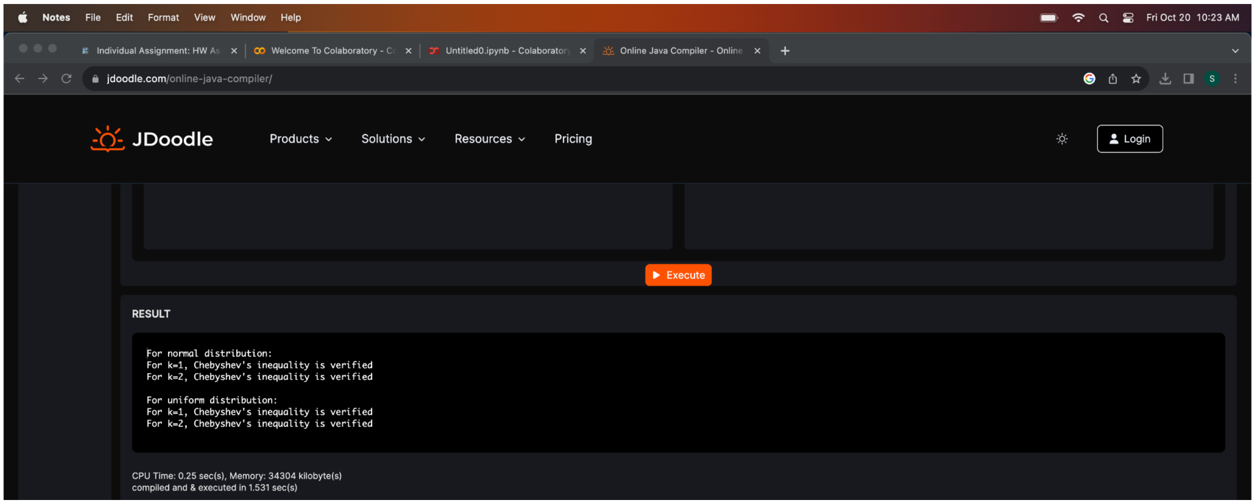
Question 2


Notes File Edit Format View Window Help
Fri Oct 20 10:22 AM
Individual Assignment: HW As
Welcome To Colaboratory - Co
Untitled0.ipynb - Colaboratory
Online Java Compiler - Online
jdoodle.com/online-java-compiler/
JDoodle
Products
Solutions
Resources
Pricing
Login

```java
import java.util.Random;

public class ChebyshevsInequalityVerification {

    public static void main(String[] args) {
        Random rand = new Random();

        double[] normalData = generateNormalData(50, 10, 0.5);
        double[] uniformData = generateUniformData(50, -20, 20);

        verifyChebyshevsInequality(normalData, "normal distribution");
        verifyChebyshevsInequality(uniformData, "uniform distribution");
    }

    public static double[] generateNormalData(int size, double mean, double stdDev) {
        Random rand = new Random();
        double[] data = new double[size];
        for (int i = 0; i < size; i++) {
            data[i] = stdDev * rand.nextGaussian() + mean;
        }
        return data;
    }
 public static double[] generateUniformData(int size, double min, double max) {
        Random rand = new Random();
        double[] data = new double[size];
        for (int i = 0; i < size; i++) {
            data[i] = rand.nextDouble() * (max - min) + min;
        }
        return data;
    }

public static void verifyChebyshevsInequality(double[] data, String distributionType) {
        int k1 = 1;
        int k2 = 2;

        int countK1 = countValuesWithinKStandardDeviations(data, k1);
        int countK2 = countValuesWithinKStandardDeviations(data, k2);

        double proportionK1 = (double) countK1 / data.length;
        double proportionK2 = (double) countK2 / data.length;

        boolean isK1Verified = proportionK1 >= 1 - 1 / Math.pow(k1, 2);
```


Notes File Edit Format View Window Help
Fri Oct 20 10:23 AM
Individual Assignment: HW As
Welcome To Colaboratory - Co
Untitled0.ipynb - Colaboratory
Online Java Compiler - Online
jdoodle.com/online-java-compiler/
JDoodle
Products
Solutions
Resources
Pricing
Login

```java
        boolean isK2Verified = proportionK2 >= 1 - 1 / Math.pow(k2, 2);

        System.out.println("For " + distributionType + ":");
        System.out.println("For k=1, Chebyshev's inequality is " + (isK1Verified ? "verified" : "not verified"));
        System.out.println("For k=2, Chebyshev's inequality is " + (isK2Verified ? "verified" : "not verified"));
        System.out.println();
    }

public static int countValuesWithinKStandardDeviations(double[] data, int k) {
        int count = 0;
        double mean = calculateMean(data);
        double stdDev = calculateStandardDeviation(data, mean);

        for (double value : data) {
            if (value >= (mean - k * stdDev) && value <= (mean + k * stdDev)) {
                count++;
            }
        }
        return count;
    }
 public static double calculateMean(double[] data) {
        double sum = 0;
        for (double value : data) {
            sum += value;
        }
        return sum / data.length;
    }

    public static double calculateStandardDeviation(double[] data, double mean) {
        double sumSquaredDiff = 0;
        for (double value : data) {
            sumSquaredDiff += Math.pow(value - mean, 2);
        }
        return Math.sqrt(sumSquaredDiff / data.length);
    }
}
```

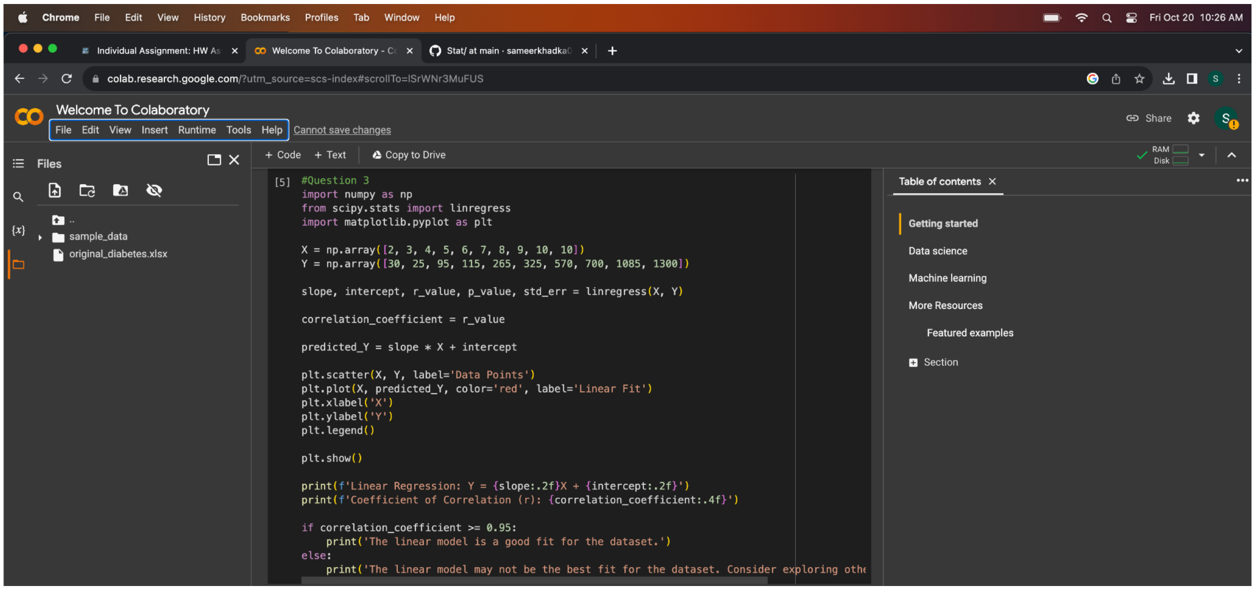

Notes File Edit Format View Window Help
Fri Oct 20 10:23 AM
Individual Assignment: HW As
Welcome To Colaboratory - Co
Untitled0.ipynb - Colaboratory
Online Java Compiler - Online
jdoodle.com/online-java-compiler/
JDoodle
Products
Solutions
Resources
Pricing
Login
Execute

RESULT

```
For normal distribution:
For k=1, Chebyshev's inequality is verified
For k=2, Chebyshev's inequality is verified

For uniform distribution:
For k=1, Chebyshev's inequality is verified
For k=2, Chebyshev's inequality is verified
```

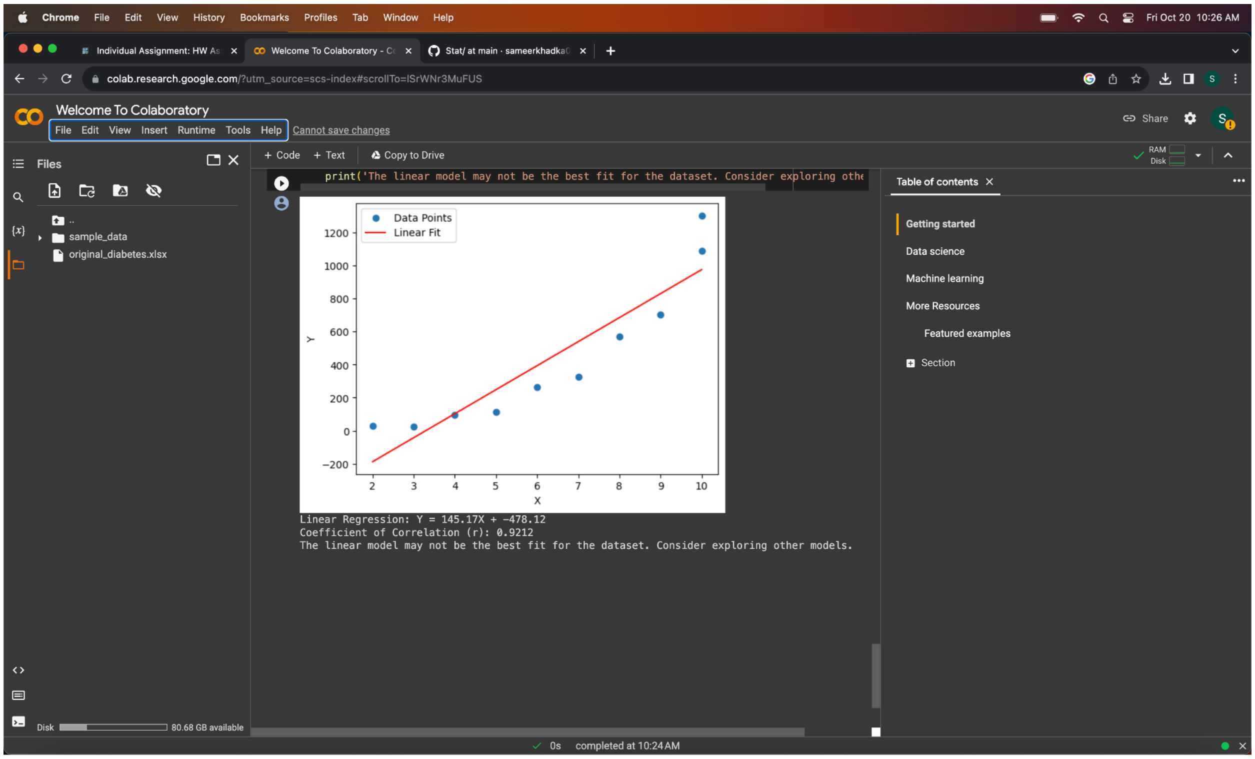
CPU Time: 0.25 sec(s), Memory: 34304 kilobyte(s)
compiled and & executed in 1.531 sec(s)

Question 3



Chrome File Edit View History Bookmarks Profiles Tab Window Help
Fri Oct 20 10:26 AM
Individual Assignment: HW As
Welcome To Colaboratory - Co
Stat/ at main · sameerkhadka0
colab.research.google.com/?utm_source=scs-index#scrollTo=lSrWNr3MuFUS
Welcome To Colaboratory
File Edit View Insert Runtime Tools Help
Cannot save changes
Share
Files
+ Code + Text
Copy to Drive
RAM
Disk
..
sample_data
original_diabetes.xlsx
Table of contents
Getting started
Data science
Machine learning
More Resources
Featured examples
Section

```
[5] #Question 3
    import numpy as np
    from scipy.stats import linregress
    import matplotlib.pyplot as plt

    X = np.array([2, 3, 4, 5, 6, 7, 8, 9, 10, 10])
    Y = np.array([30, 25, 95, 115, 265, 325, 570, 700, 1085, 1300])

    slope, intercept, r_value, p_value, std_err = linregress(X, Y)

    correlation_coefficient = r_value

    predicted_Y = slope * X + intercept

    plt.scatter(X, Y, label='Data Points')
    plt.plot(X, predicted_Y, color='red', label='Linear Fit')
    plt.xlabel('X')
    plt.ylabel('Y')
    plt.legend()

    plt.show()

    print(f'Linear Regression: Y = {slope:.2f}X + {intercept:.2f}')
    print(f'Coefficient of Correlation (r): {correlation_coefficient:.4f}')

    if correlation_coefficient >= 0.95:
        print('The linear model is a good fit for the dataset.')
    else:
        print('The linear model may not be the best fit for the dataset. Consider exploring othe
```



```
        print('The linear model may not be the best fit for the dataset. Consider exploring othe
```

```
Linear Regression: Y = 145.17X + -478.12
Coefficient of Correlation (r): 0.9212
The linear model may not be the best fit for the dataset. Consider exploring other models.
```