|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | Discrete |
| Results of rolling a dice | Discrete |
| Weight of a person | Continuous |
| Weight of Gold | Continuous |
| Distance between two places | Continuous |
| Length of a leaf | Continuous |
| Dog's weight | Continuous |
| Blue Color | Discrete |
| Number of kids | Discrete |
| Number of tickets in Indian railways | Discrete |
| Number of times married | Discrete |
| Gender (Male or Female) | Discrete |

Q1) Identify the Data type for the Following:

Q2) Identify the Data types, which were among the following

Nominal, Ordinal, Interval, Ratio.

|  |  |
| --- | --- |
| Data | Data Type |
| Gender | Nominal |
| High School Class Ranking | Ordinal |
| Celsius Temperature | Interval |
| Weight | Ratio |
| Hair Color | Nominal |
| Socioeconomic Status | Ordinal |
| Fahrenheit Temperature | Interval |
| Height | Ratio |
| Type of living accommodation | Ordinal |
| Level of Agreement | Ordinal |
| IQ(Intelligence Scale) | Ratio |
| Sales Figures | Ratio |
| Blood Group | Ordinal |
| Time Of Day | Ratio |
| Time on a Clock with Hands | Interval |
| Number of Children | Nominal |
| Religious Preference | Nominal |
| Barometer Pressure | Interval |
| SAT Scores | Interval |
| Years of Education | Ordinal |

Q3) Three Coins are tossed, find the probability that two heads and one tail are obtained?

**Answer**:

P(3) = Total Outcomes = (HHT, HTH, THT, THH, HTT, TTH, HHH, TTT) = 8

P(2 H, 1T) = Possible outcomes / Total Outcomes = 3/8

Q4) Two Dice are rolled, find the probability that the sum is

1. Equal to 1
2. Less than or equal to 4
3. Sum is divisible by 2 and 3

**Answer**:

P(2) = 36

1. 0
2. Possible outcomes = (1,1), (1,2), (1,3), (2,1), (3,1),(2,2)

Possible outcomes/ Total Outcomes = 6/36 = 1/6

1. Divisible by 2 and 3 = 6 and 12

Possible outcomes = (1,5), (5,1), (2,4), (4,2), (3,3), (6,6)

6/36 = 1/6

Q5) A bag contains 2 red, 3 green and 2 blue balls. Two balls are drawn at random. What is the probability that none of the balls drawn is blue?

Answer:

2 r, 3 g, 2 b

Total balls = 7

Non blue balls = 5/7

Total Bag = 7C2 = 7!/2!\*5! = 21

Probability(non blue balls) = 5C2 = 5!/3!\*2! = 10

Probability of no blue balls being drawn = 10/21

Q6) Calculate the Expected number of candies for a randomly selected child

Below are the probabilities of count of candies for children (ignoring the nature of the child-Generalized view)

|  |  |  |
| --- | --- | --- |
| CHILD | Candies count | Probability |
| A | 1 | 0.015 |
| B | 4 | 0.20 |
| C | 3 | 0.65 |
| D | 5 | 0.005 |
| E | 6 | 0.01 |
| F | 2 | 0.120 |

Child A – probability of having 1 candy = 0.015.

Child B – probability of having 4 candies = 0.20

**Answer:**

0.015+0.8+1.95+0.025+0.06+0.24 = 3.09

Q7) Calculate Mean, Median, Mode, Variance, Standard Deviation, Range & comment about the values / draw inferences, for the given dataset

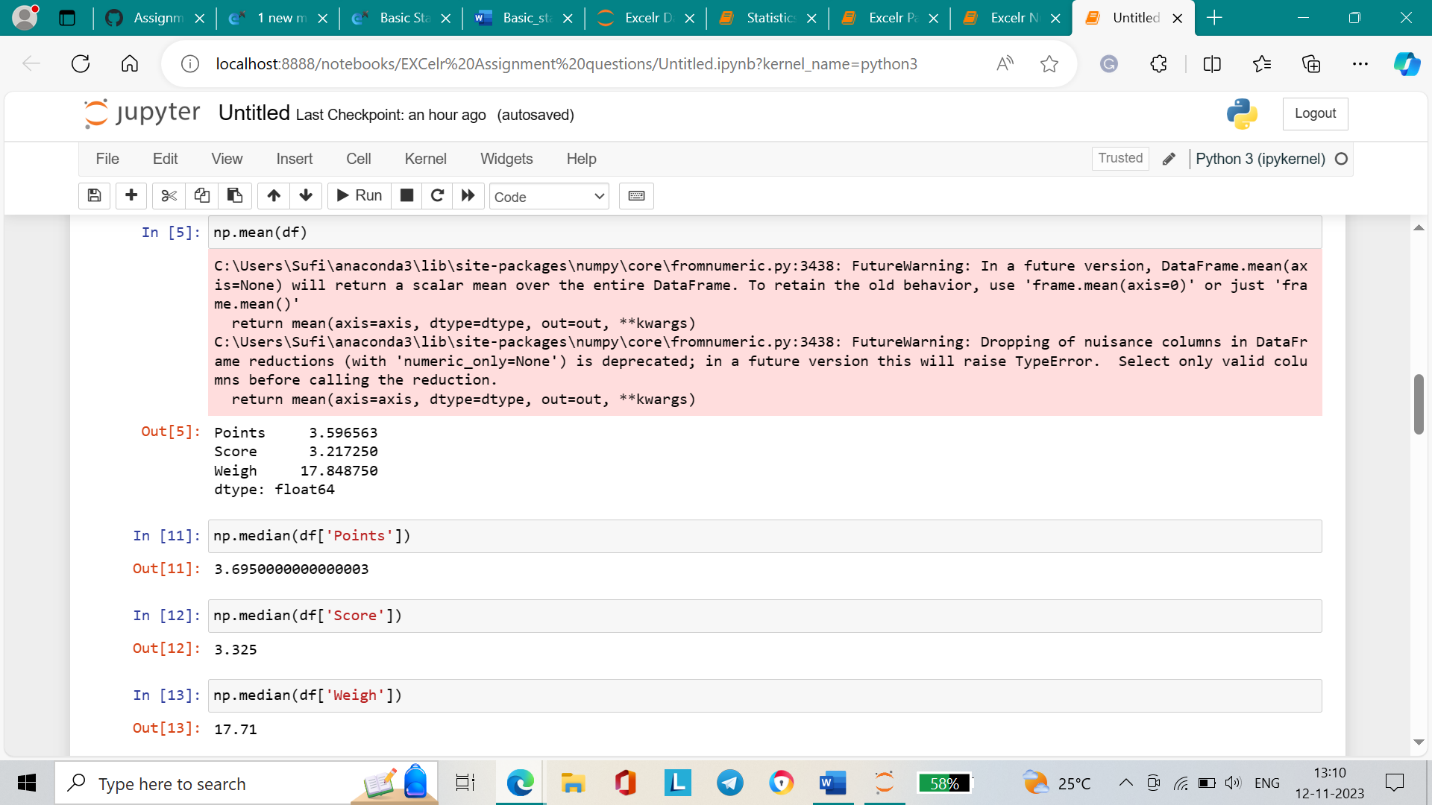
* For Points, Score, Weigh>

Find Mean, Median, Mode, Variance, Standard Deviation, and Range, Comment about the values/ Draw some inferences.

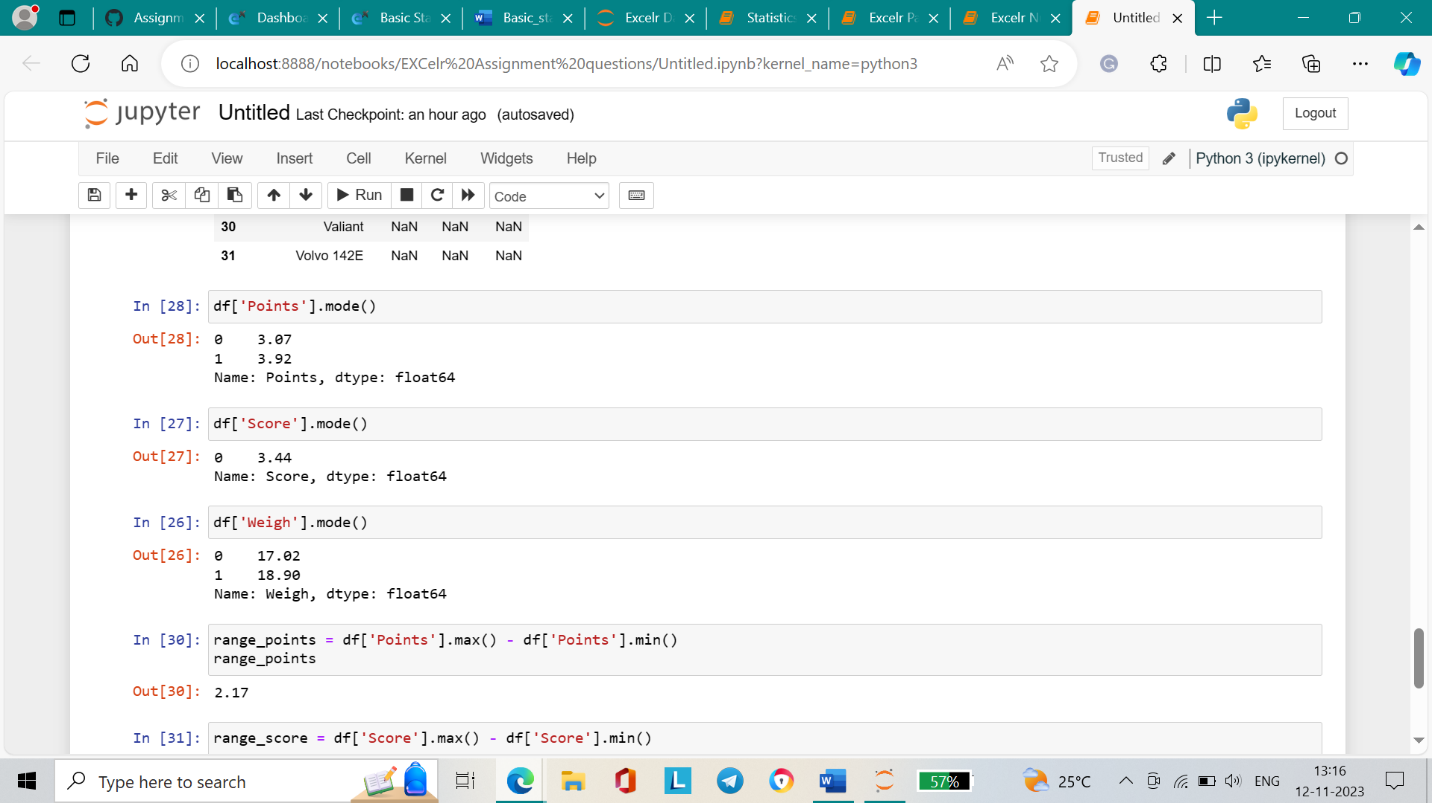
**Use Q7.csv file**

**Answer:**

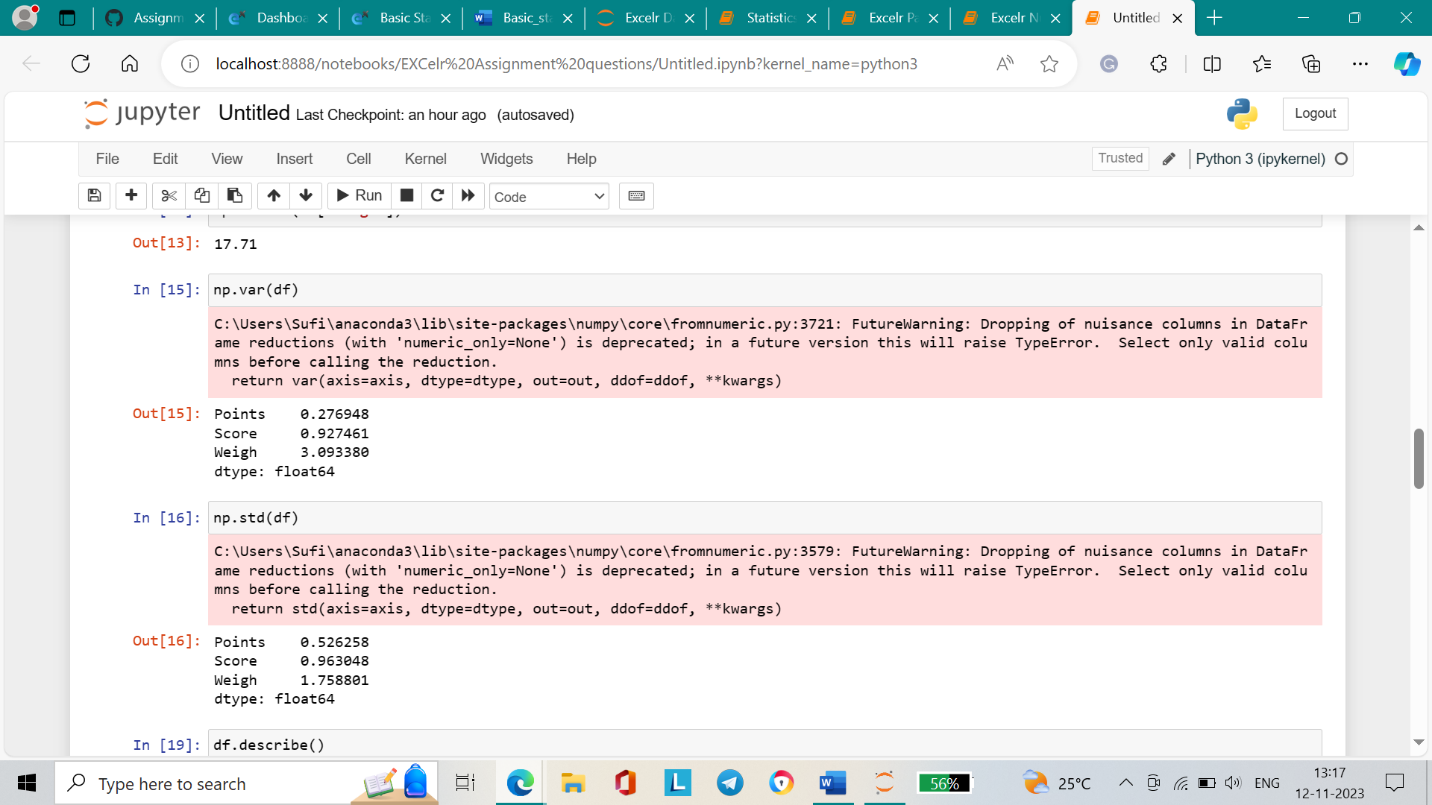
**Mean and Median:**



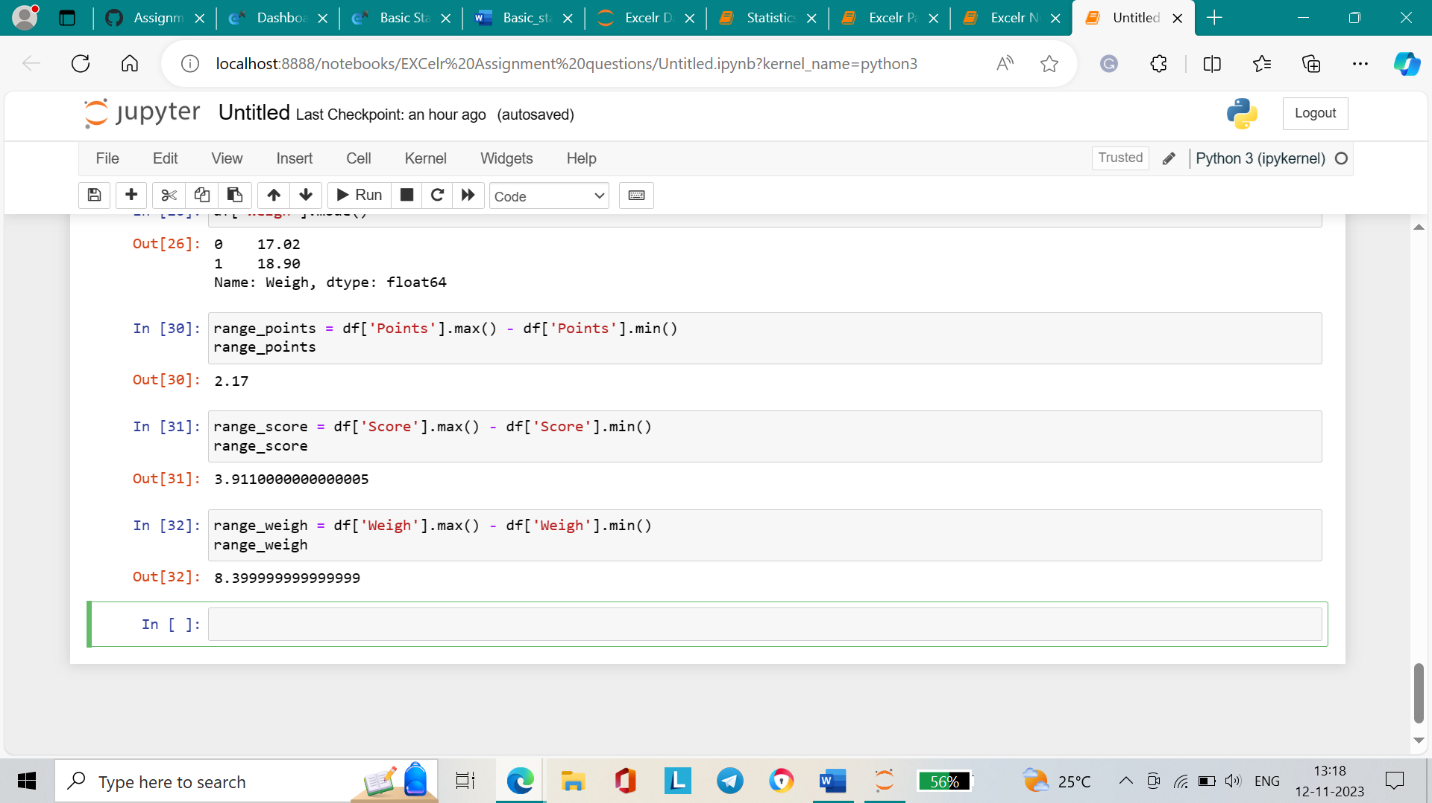
**Mode:**

****

**Variance and Standard Deviation:**

****

**Range:**

****

**Inference:**

The standard deviation and variance are highest for ‘weigh’, and this column also has the highest range value which means it has the most varying values.

Q8) Calculate the Expected Value for the problem below

1. The weights (X) of patients at a clinic (in pounds), are

108, 110, 123, 134, 135, 145, 167, 187, 199

Assume one of the patients is chosen at random. What is the Expected Value of the Weight of that patient?

**Answer:**

**Expected value = Sum of ( n \* probability)**

**Expected value =** (1/9)(108)+ (1/9)(110)+ (1/9)(123)+ (1/9)(134)+ (1/9)(145)+ (1/9)(167)+ (1/9)(187)+ (1/9)(199) = 145.33

Mean value = 108 + 110+ 123+134+135+145+167+187+199

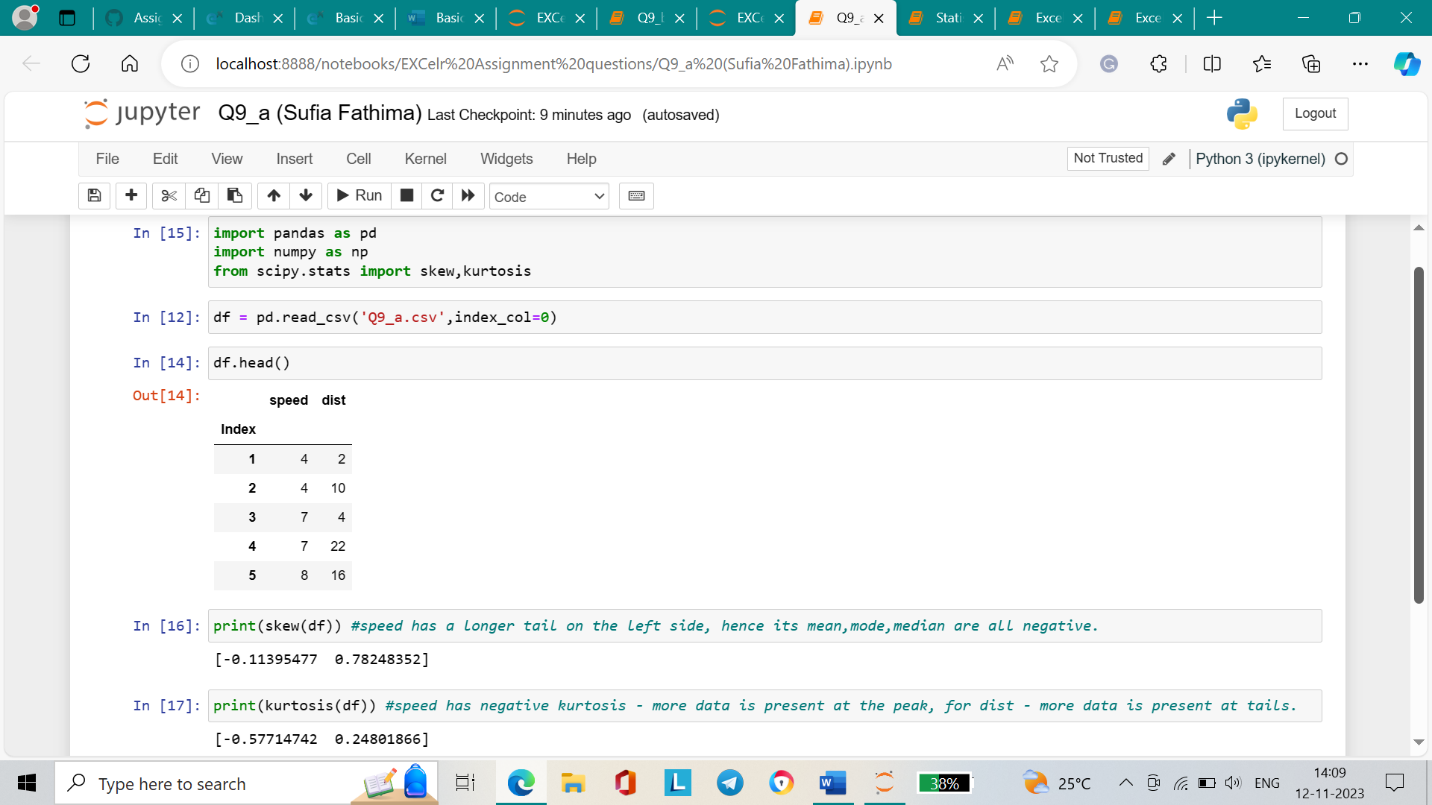
/ 9 = 145.33

**Q9) Calculate Skewness, Kurtosis & draw inferences on the following data**

**Cars speed and distance**

**Use Q9\_a.csv**

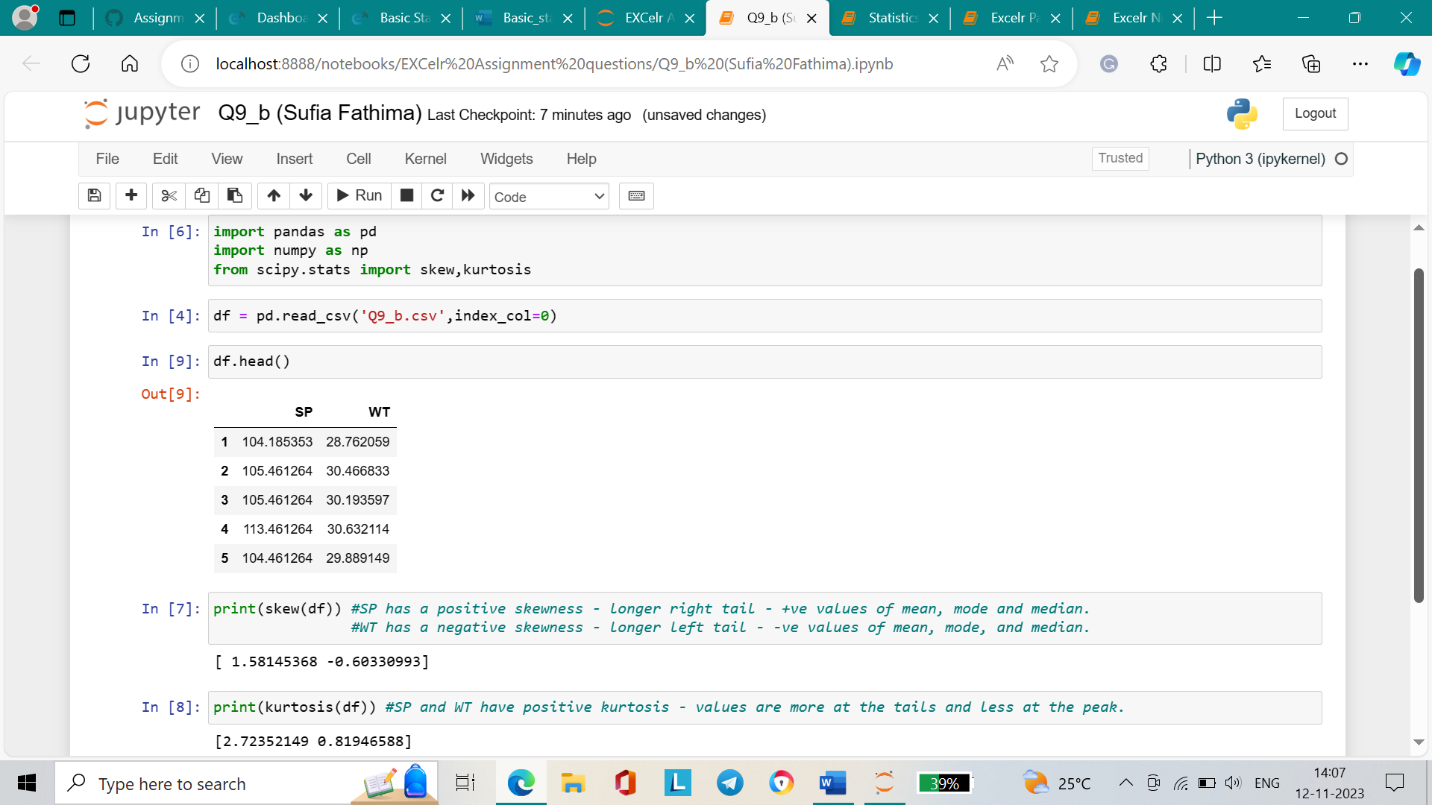
**Answer:**

****

**SP and Weight(WT)**

**Use Q9\_b.csv**

**Answer:**

****

**Q10) Draw inferences about the following boxplot & histogram**



**Inference:**

The histogram is right-skewed and has a tail on the right side, and the mean is greater than the median. It indicates that the data is not distributed normally. There are not many outliers.

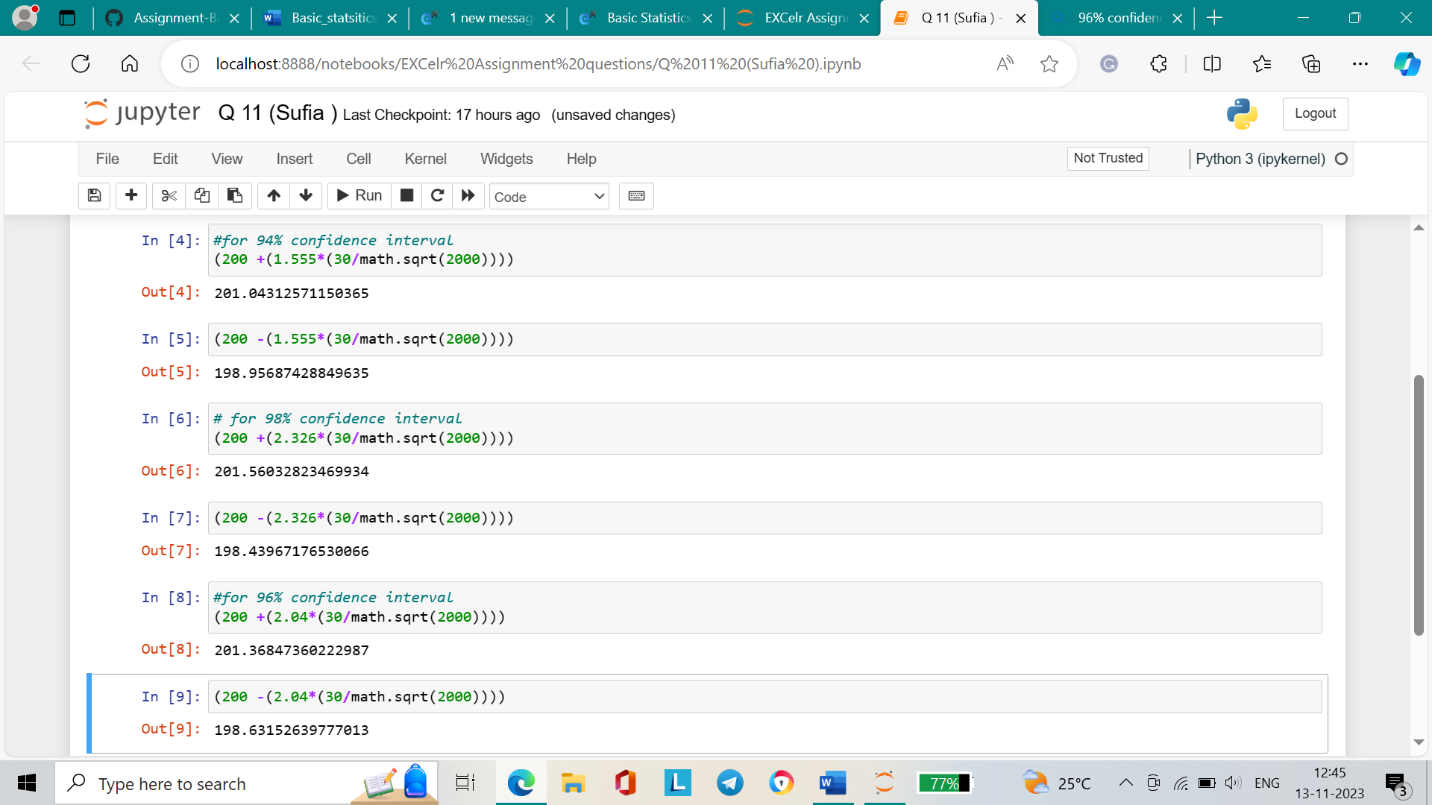


**Inference:**

The outliers exist at the very high end of the dataset**. T**he median lies towards the lower quartile hence the boxplot is skewed. The box plot is right-skewed (positively) as the upper whisker is longer.

**Q11)** Suppose we want to estimate the average weight of an adult male in Mexico. We draw a random sample of 2,000 men from a population of 3,000,000 men and weigh them. We find that the average person in our sample weighs 200 pounds, and the standard deviation of the sample is 30 pounds. Calculate 94%,98%, and 96% confidence interval?

**Answer**:



For 94% confidence interval Range is [ 198.95 – 201.04]

For 98% confidence interval range is [198.43 – 201.56]

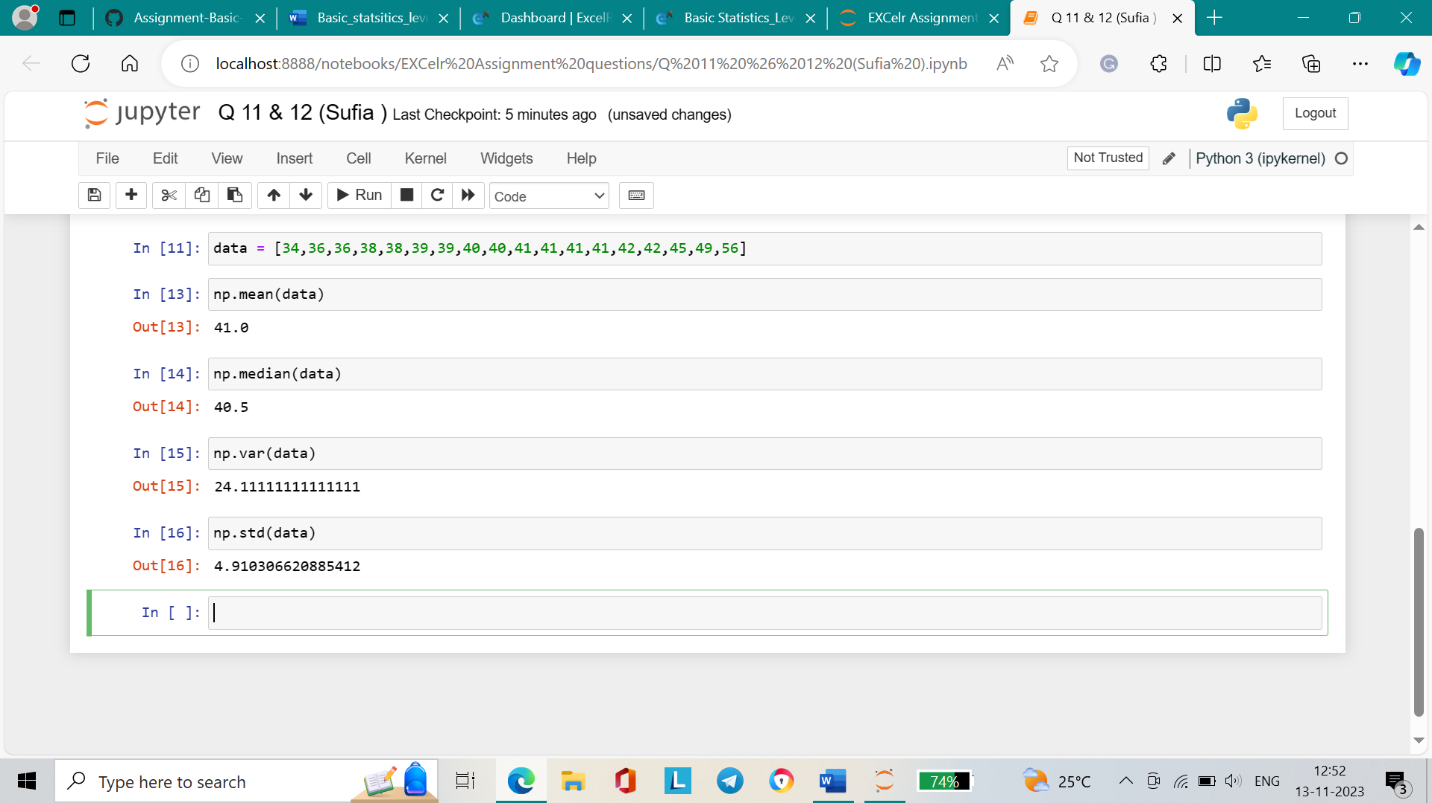
For 96% confidence interval range is [198.63 – 201.37]

**Q12)** Below are the scores obtained by a student in tests

**34,36,36,38,38,39,39,40,40,41,41,41,41,42,42,45,49,56**

1. Find mean, median, variance, standard deviation.
2. What can we say about the student marks?

Answer:



1. Since the mean is greater than the median, the data is right-skewed. Also, there are no outliers present in the data.

Q13) What is the nature of skewness when the mean, and median of data are equal?

**Answer:** When the mean = median, there is no skewness present. This means that the data is distributed in a symmetrical manner.

Q14) What is the nature of skewness when mean > median ?

**Answer:** When mean > median, then the nature of skewness is positive, i.e. the distribution is right-skewed.

Q15) What is the nature of skewness when median > mean?

**Answer:** When median > mean, the nature of skewness is negative, i.e. the distribution is left-skewed.

Q16) What does positive kurtosis value indicates for a data ?

**Answer:** When the data has a positive kurtosis value, this means that there is lesser data at the peak and more data near the tails. This shows a sharper peak and heavier tales.

Q17) What does negative kurtosis value indicates for a data?

**Answer:** When the data has a negative kurtosis value, it means that more data is present at the peak making it look more flat and large. Lesser data is present near the tails.

Q18) Answer the questions below using the below boxplot visualization.



What can we say about the distribution of the data?

**Answer**: The data is not distributed normally as the median lies towards the higher value and the left whisker is comparatively longer.

What is the nature of the skewness of the data?

**Answer**: The data is left-skewed. The whisker range of minimum values is longer than the maximum value.

What will be the IQR of the data (approximately)?   
**Answer**: IQR = Q3 – Q1 = 18 – 10 = 8

Q19) Comment on the below Boxplot visualizations?



**Answer**:

* There are no outliers in both the boxplots.
* They share almost the same median and are approximately symmetrical with no skewness.

Draw an Inference from the distribution of data for Boxplot 1 with respect Boxplot 2.

**Answer**: The first box plot has less variability in data due to the size of the box and the short length of the whiskers. Whereas, the second box plot is more variable in data due to the longer box and whisker lengths.

Q 20) Calculate probability from the given dataset for the below cases

Data \_set: Cars.csv

Calculate the probability of MPG of Cars for the below cases.

MPG <- Cars$MPG

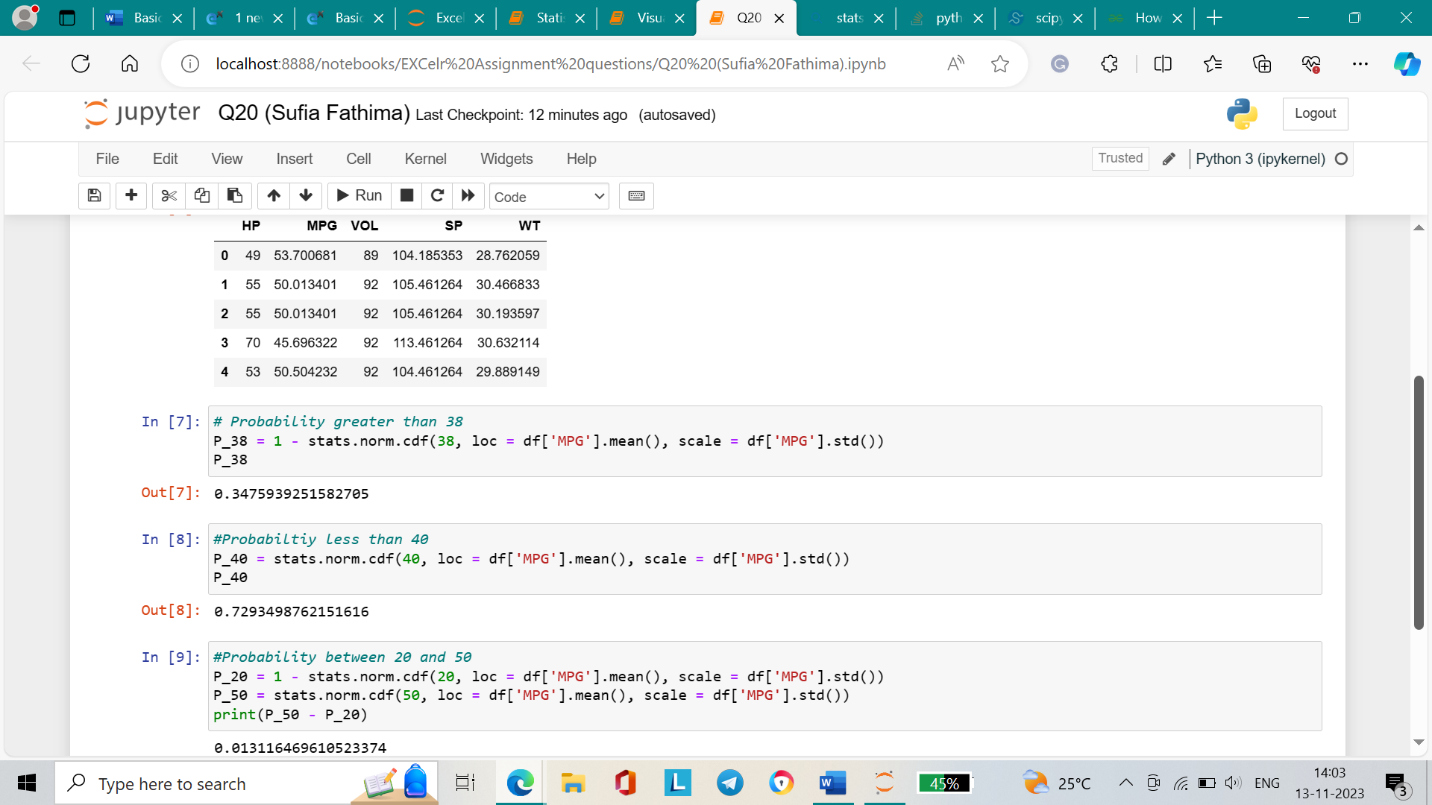
* 1. P(MPG>38)
  2. P(MPG<40)
  3. P (20<MPG<50)

**Answer:**

P(MPG>38) = 0.347

P(MPG<40) = 0.729

P (20<MPG<50) = 0.0131

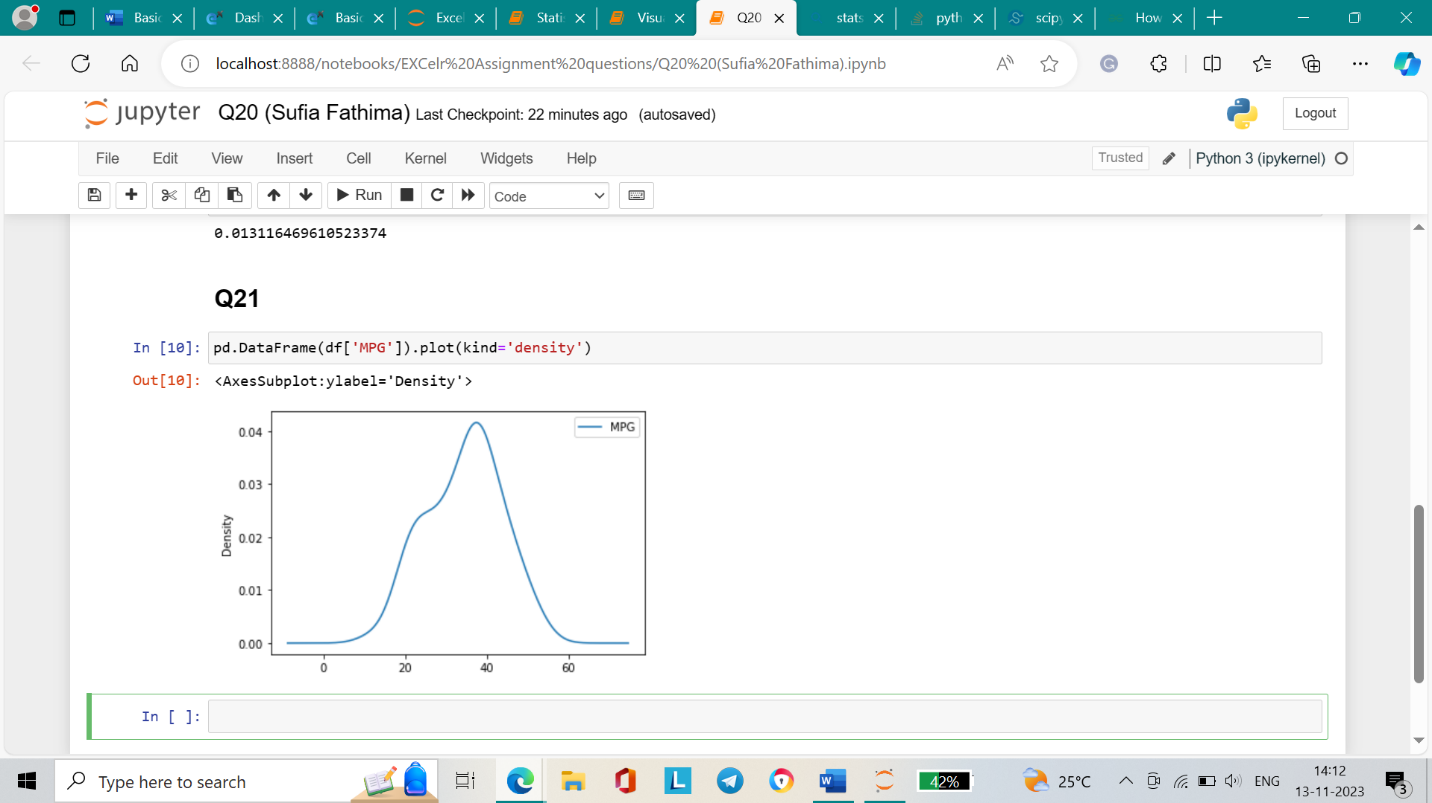


Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

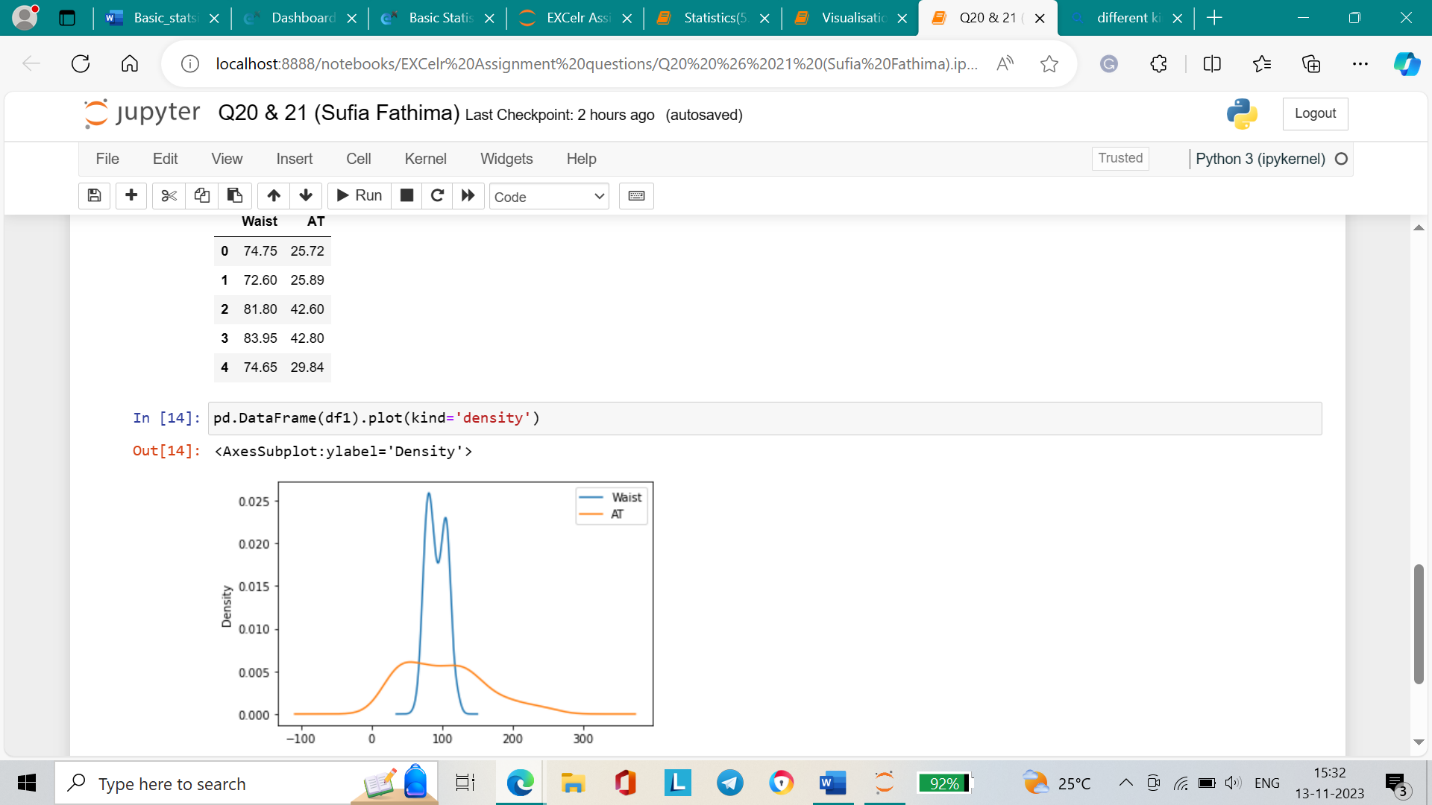
**Answer**:



1. Check Whether the Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set follows Normal Distribution

Dataset: wc-at.csv

**Answer**:



Q 22) Calculate the Z scores of 90% confidence interval,94% confidence interval, 60% confidence interval

**Answer**:

* **90%** CI = (1+0.90)/2 = 0.95

**Zscore table =** 1.65

* 94% CI = (1+0.94)/2 = 0.97

**Zscore table =** 1.91

* **60% CI =** (1+0.60)/2 = 0.8

**Zscore table** = 0.85

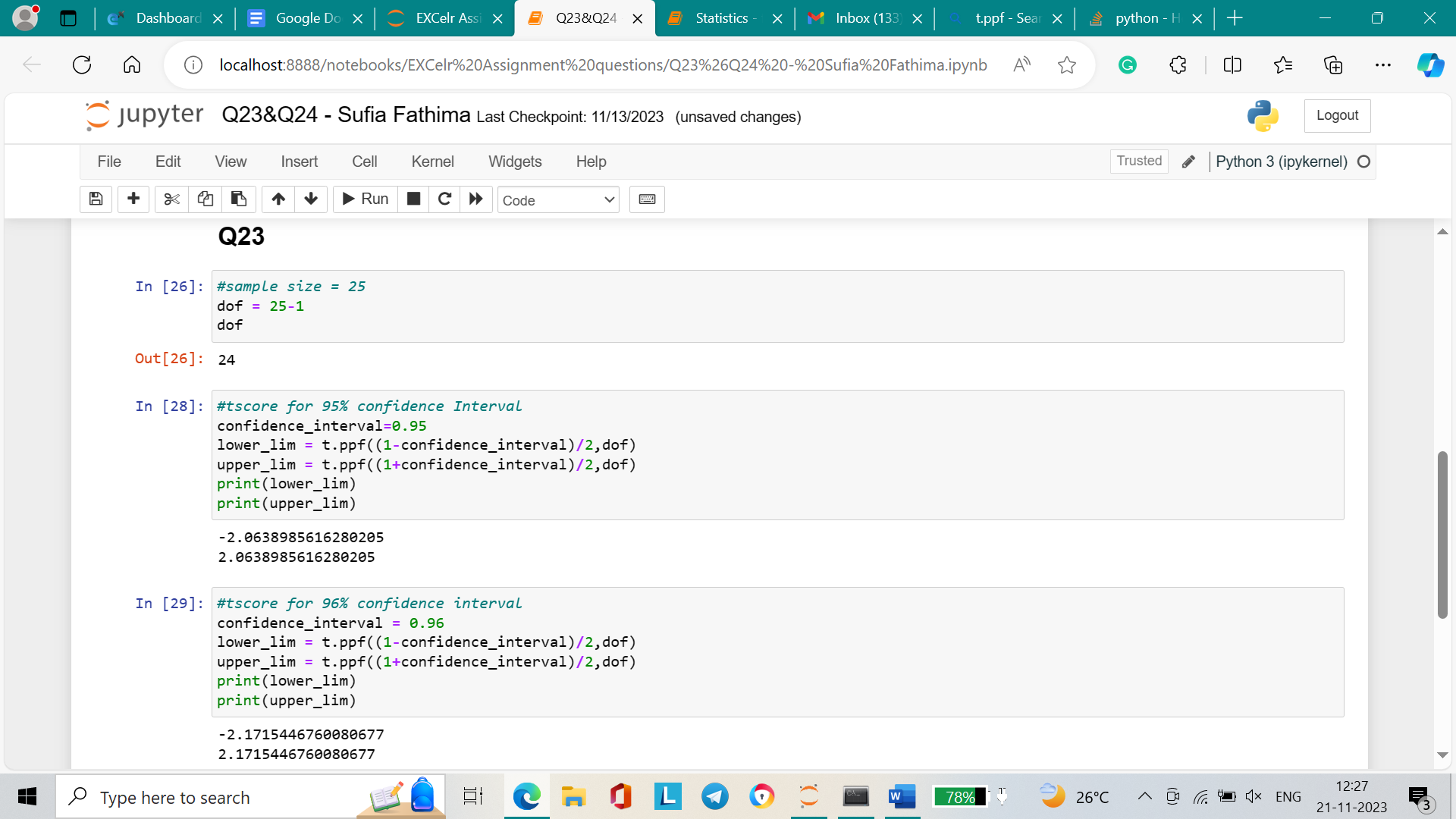
Q 23) Calculate the t scores of 95% confidence interval, 96% confidence interval, and 99% confidence interval for the sample size of 25

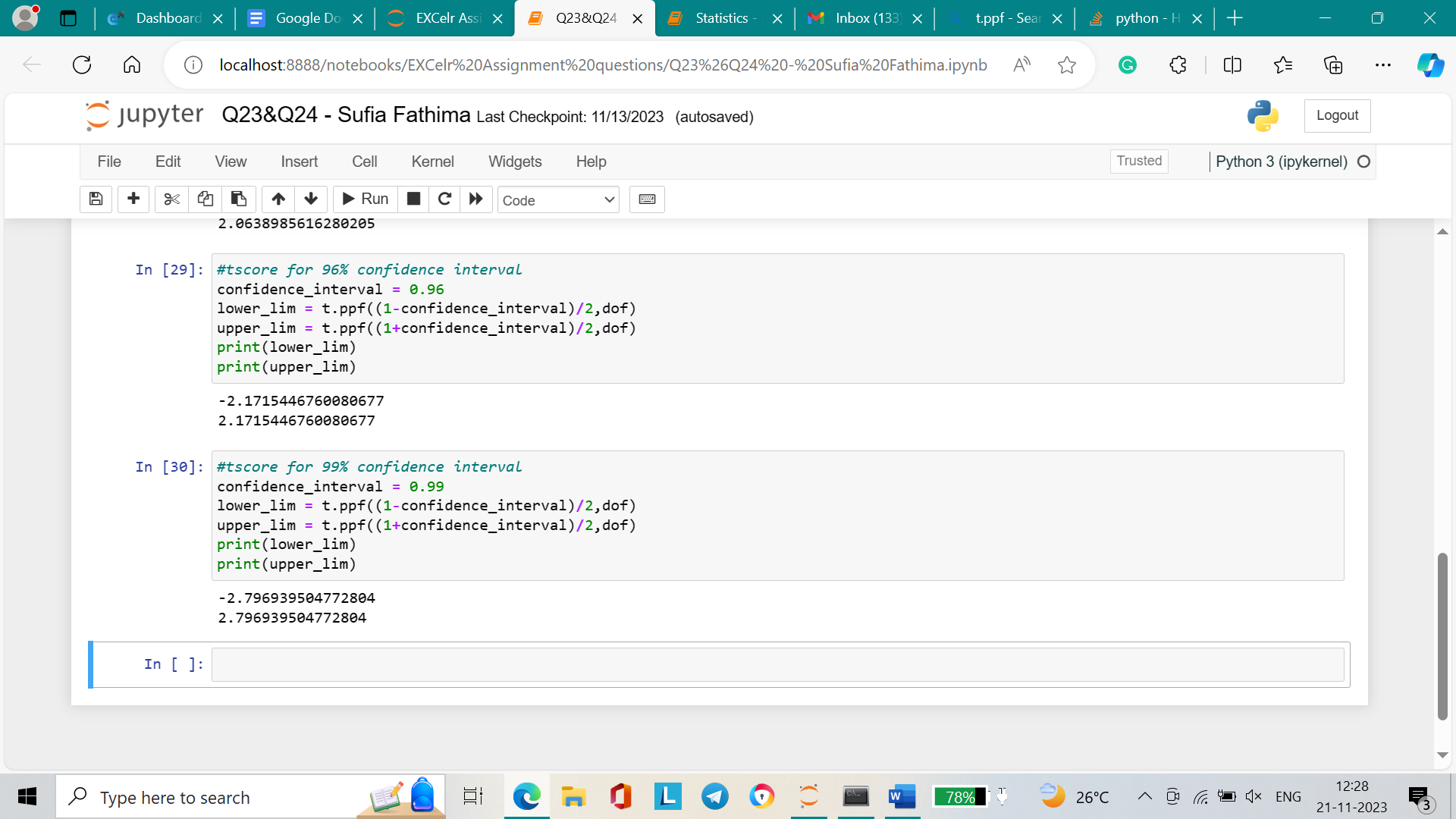
**Answer**:

* Tscore for 95% confidence interval:

-2.06 and 2.06

* Tscore for 96% confidence interval:
  + -2.17 and 2.17
* Tscore for 99% confidence interval:
  + -2.79 and 2.79





Q 24**)** A Government company claims that an average light bulb lasts 270 days. A researcher randomly selects 18 bulbs for testing. The sampled bulbs last an average of 260 days, with a standard deviation of 90 days. If the CEO's claim were true, what is the probability that 18 randomly selected bulbs would have an average life of no more than 260 days

Hint:

rcode 🡪 pt(tscore,df)

df 🡪 degrees of freedom

**Answer**:

Probability = 0.321

