**Question no = 1 Identify the data type of the following:**

|  |  |
| --- | --- |
| **Activity** | **Data Type** |
| **Number of beating from wife** | **Discrete** |
| **Result of rolling a dice** | **Discrete** |
| **Weight of a person** | **Continuous** |
| **Distance between two places** | **Continuous** |
| **Length of a leaf** | **Continuous** |
| **Dogs weight** | **Continuous** |
| **Blue Color** | **Discrete** |
| **Number of Kids** | **Discrete** |
| **Number of times married** | **Discrete** |
| **Gender (Male or Female)** | **Discrete** |
| **Number of tickets in Indian railways** | **Discrete** |
|  |  |

**Question no = 2 Identify the data type which were among the following:**

**Nominal , ordinal , interval , ratio**

|  |  |
| --- | --- |
| **Data** | **Data type** |
| **Gender** | **Nominal** |
| **High school class ranking** | **Ordinal** |
| **Celsius Temperature** | **Ratio** |
| **Weight** | **Ratio** |
| **Hair color** | **Nominal** |
| **Socio economic status** | **Ordinal** |
| **Fahrenheit Temperature** | **Ratio** |
| **Height** | **Ratio** |
| **Type of living accommodation** | **Ordinal** |
| **Level of agreement** | **Ordinal** |
| **IQ(Intelligence Scale)** | **Interval** |
| **Sales Figure** | **Ratio** |
| **Blood Group** | **Nominal** |
| **Time of Day** | **Ordinal** |
| **Time on a clock with Hands** | **Ratio** |
| **Number of children** | **Ratio** |
| **Religious Preference** | **Nominal** |
| **Barometer Pressure** | **Nominal** |
| **SAT Scores** | **Ordinal** |
| **Years of Education** | **Ratio** |

**Question = 3 Three coins are tossed , find the probability that two head and one tail are obtained.**

**Total possible event :8 , No of desired events :3**

**0 3/8 = 37.5% of probability**

**Question no = 4 Two dice are rolled , find the probability that sum is?**

1. **= probability of sum equals to 1 = 0**
2. **= probability of sum is less than or equals to 4 = 0.167**
3. **= 5/36 = 0.139**

**Question no = 5 A bag contain 2 red , 3 green and 2 blue , two balls are drawn at random , what is the probability that none of the balls drawn is blue.**

**Total number of balls = 7**

**2 balls are drawn randomly from the bag = 21**

**If none of them drawn 2 balls are blue = 7-2 = 5**

**N = none of ball drawn is blue = 10**

**Probability of drawn blue balls = 10/21**

**Answer = 0.476**

**Question no = 6 Calculated the expected number of candies for a randomly selected child?**

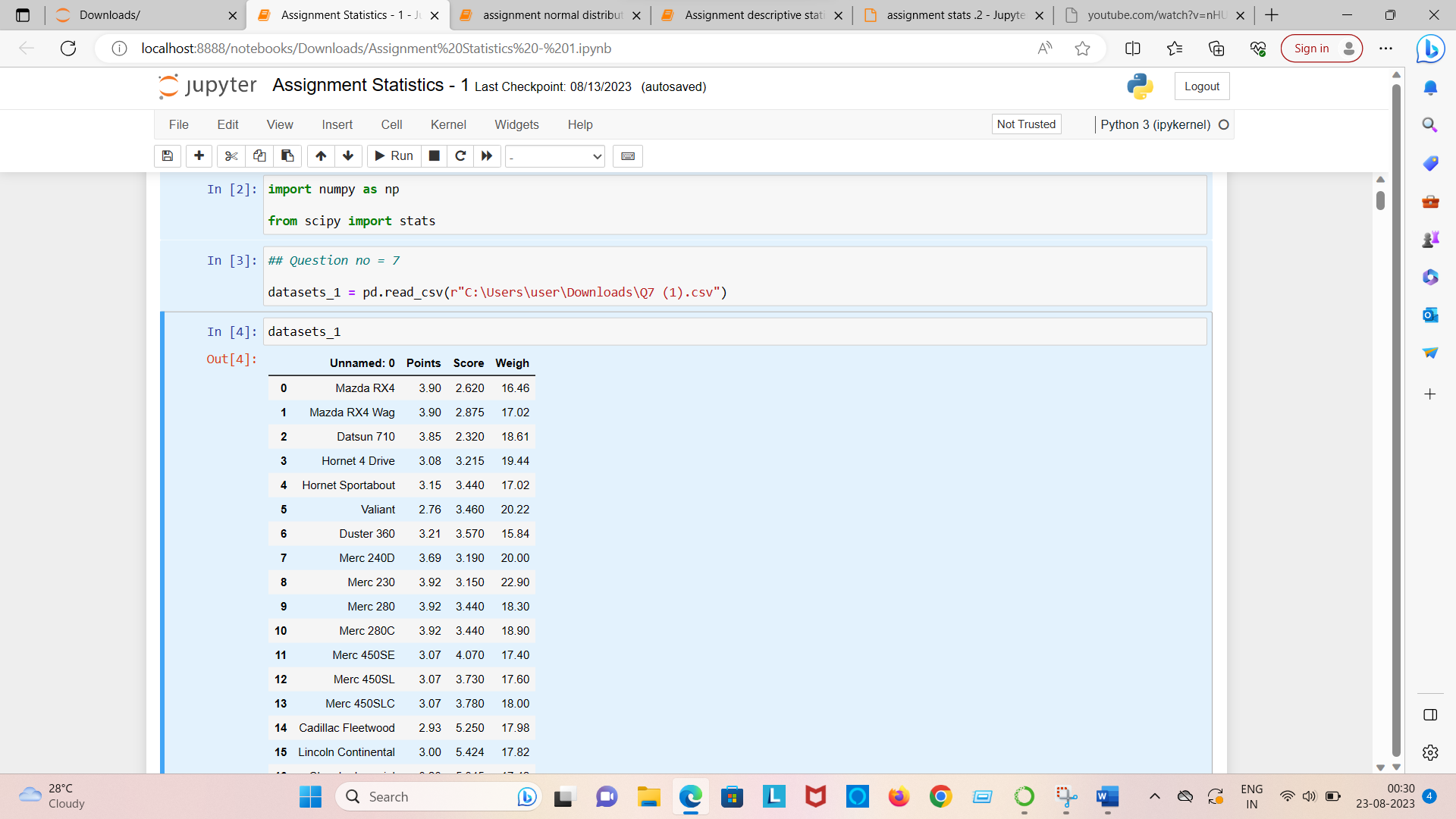
**Child A – probability of having 1 candy = 0.015**

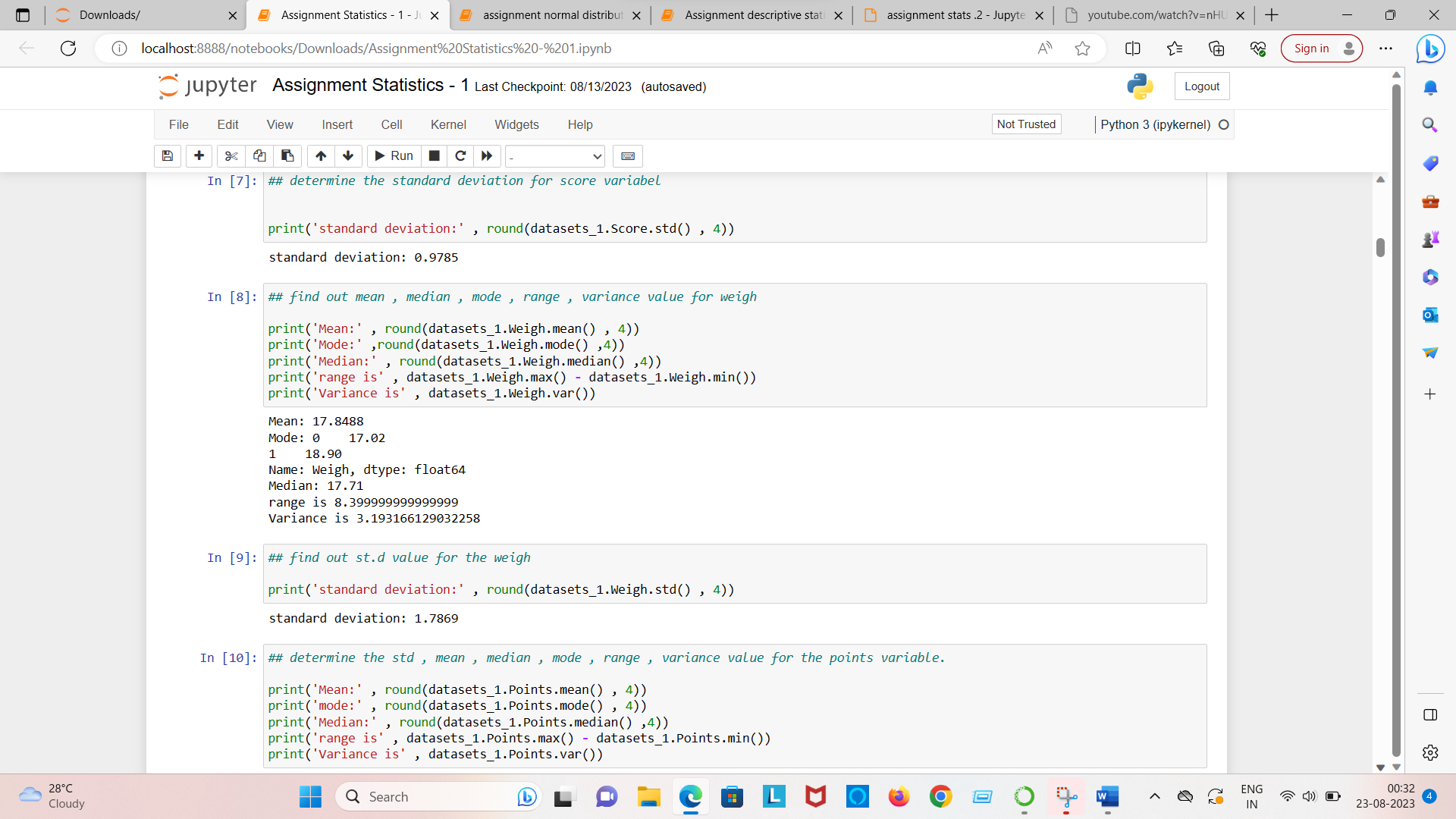
**Child B – probability of having 4 candies = 0.20**

**Probability = 3.09**

**Question no = 7 Calculate the mean , mode , median , variance , standard deviation , range & comment about the values /draw inferences for the given datasets.**

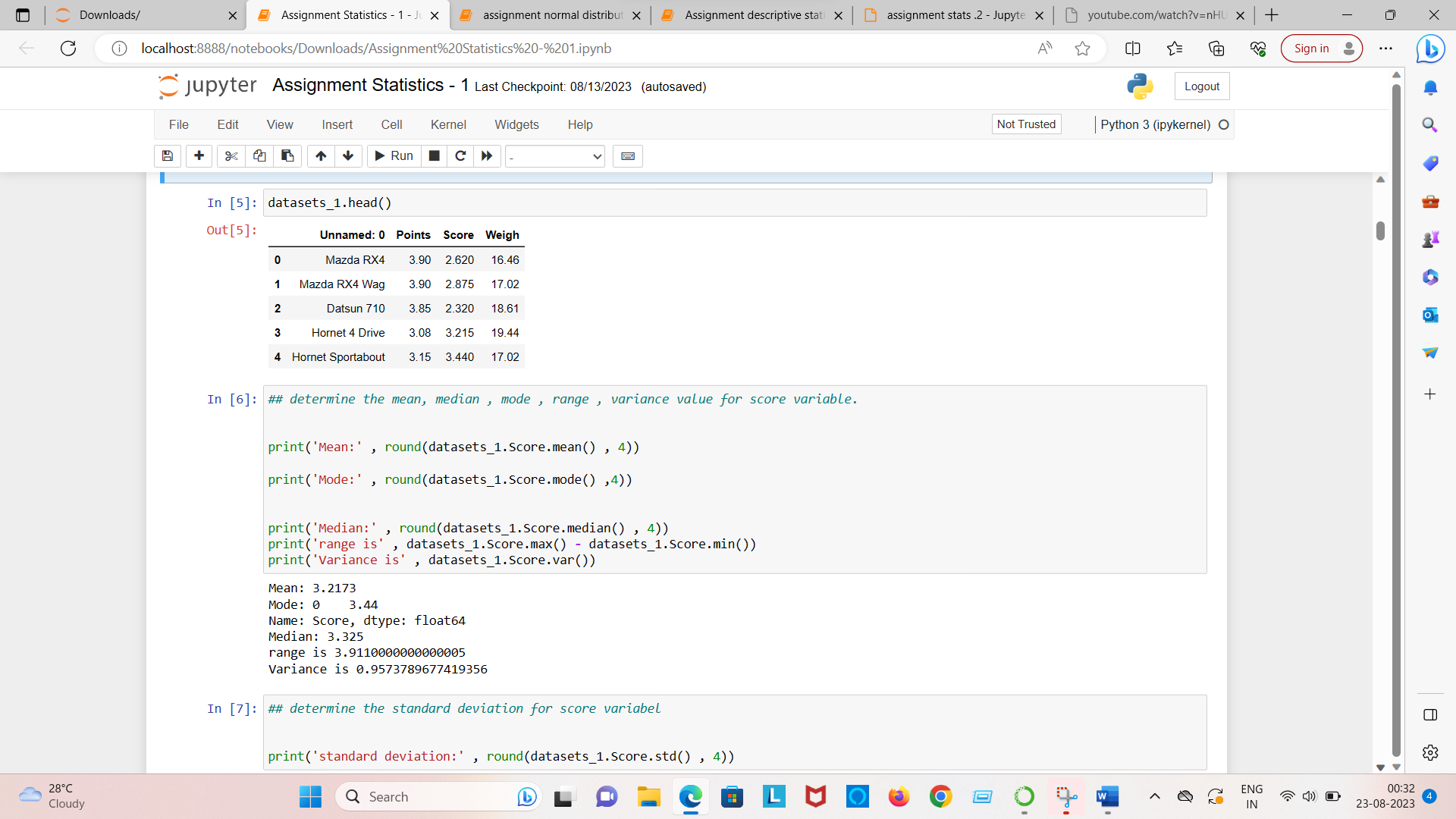
**For the point , score , weigh>**

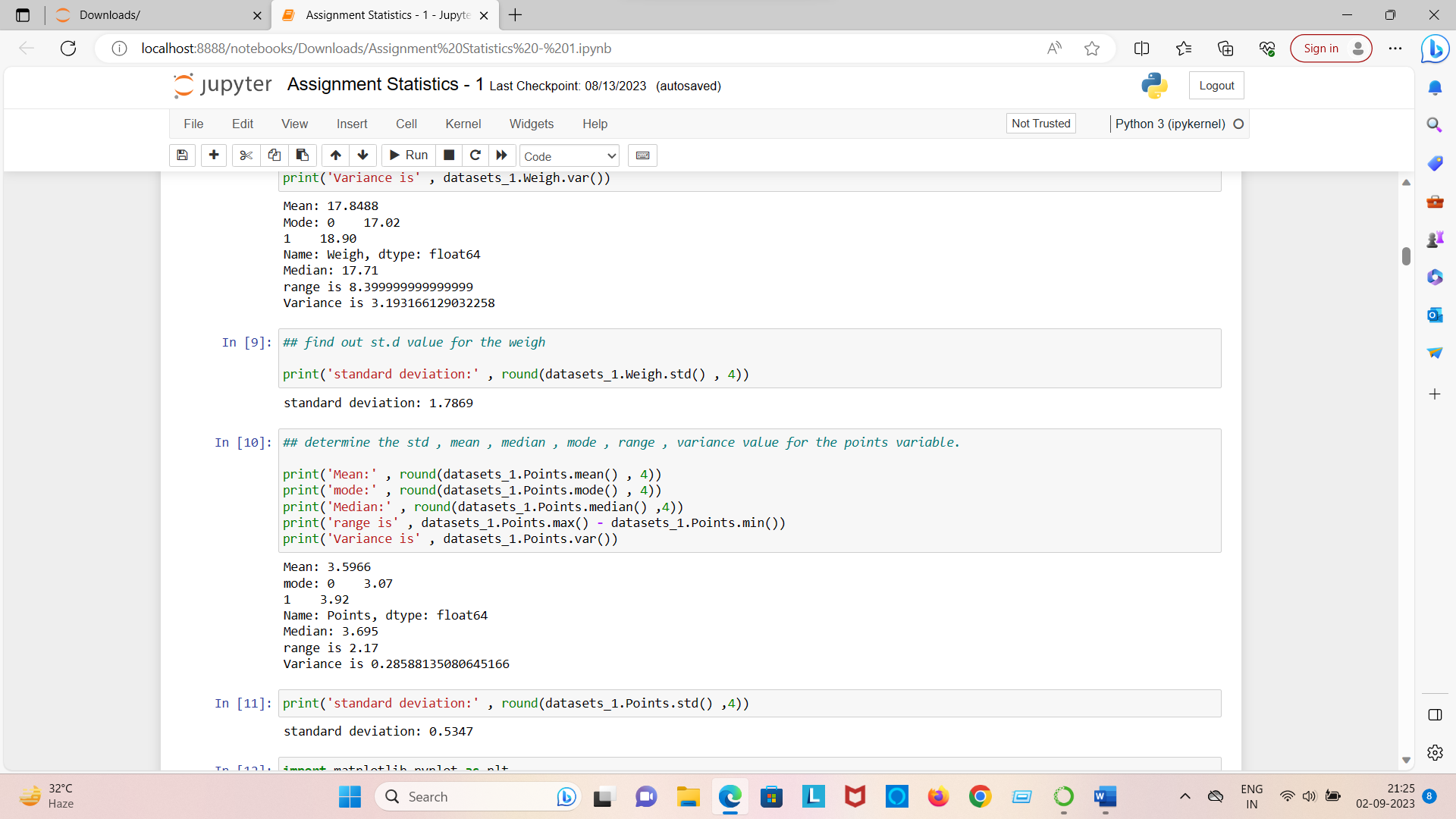
****

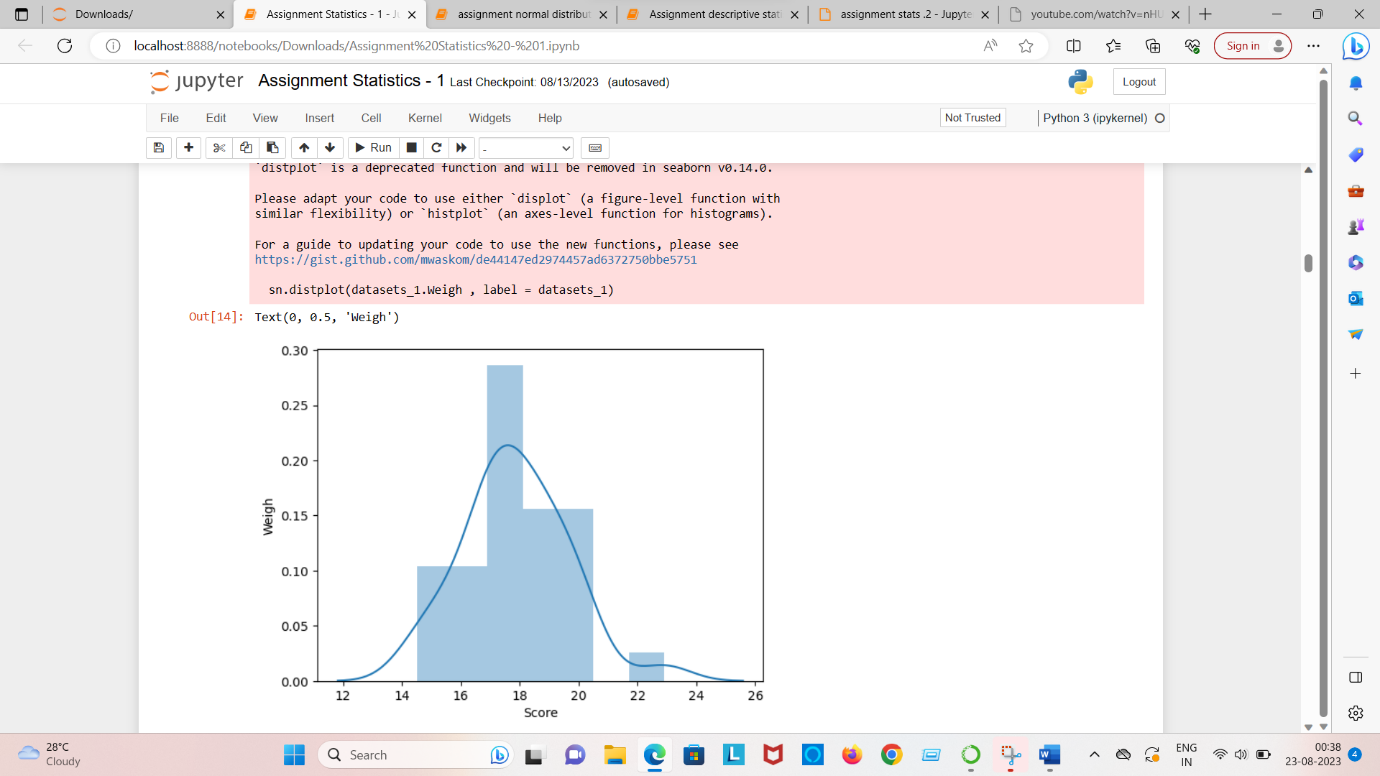
****

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Mean | Median | Mode | Variance | Std.Dev | Range |
| Points | 3.59 | 3.70 | 3.92 | 0.29 | 0.53 | 2.17 |
| Score | 3.22 | 3.33 | 3.44 | 0.96 | 0.98 | 3.91 |
| Weigh | 17.85 | 17.71 | 17.02 | 3.19 | 1.79 | 8.40 |

**No case has variable Mean = Median = Mode**

****





**Question = 8 Calculated expected the value for the problem below**

1. **The weight (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**

Expected value = sum(x\*probability of X)

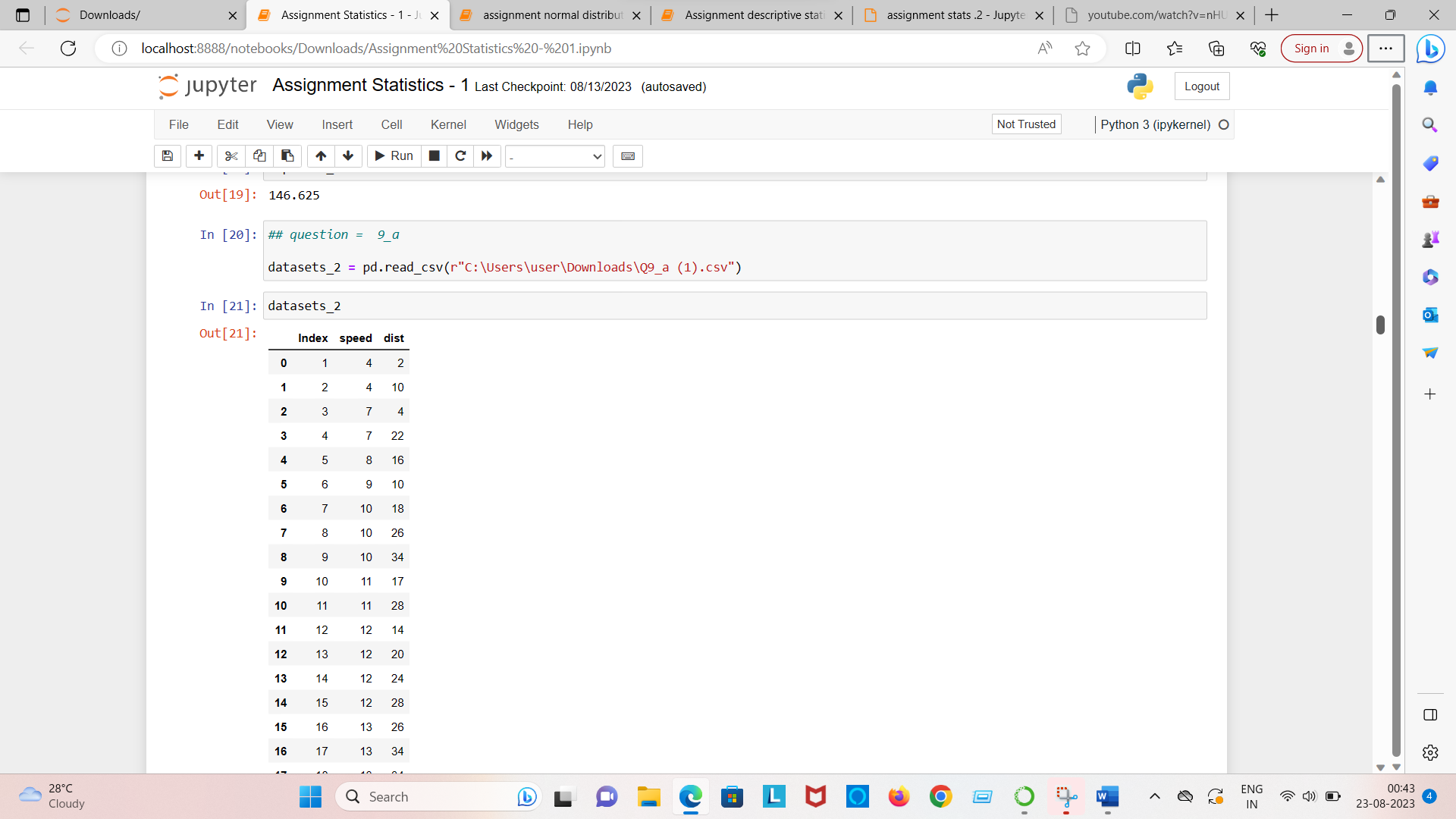
145.33

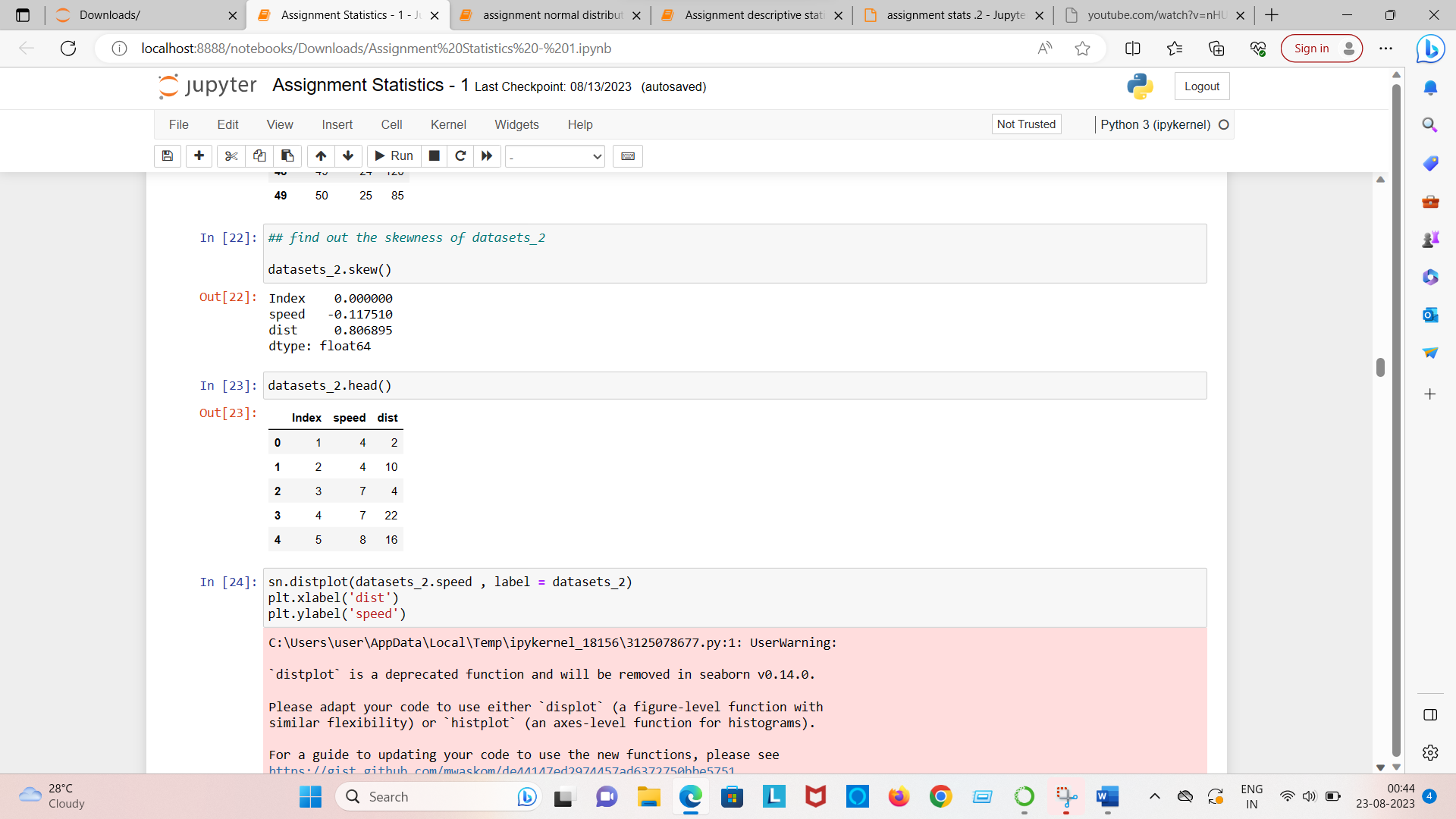
**Question no = 9 Calculate skewness , kurtosis & draw inference on the following data**

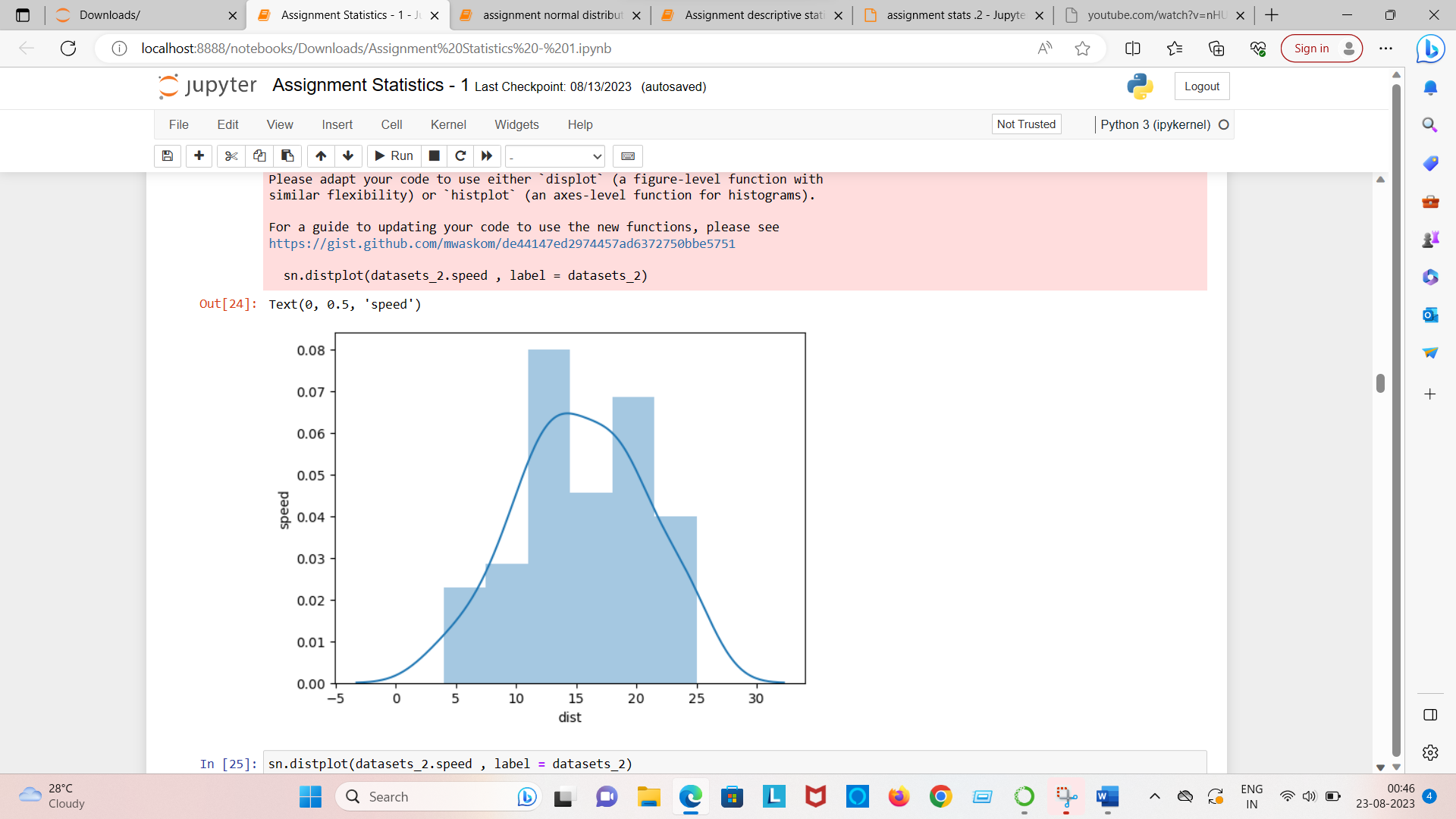
|  |  |  |
| --- | --- | --- |
|  | Skewness | Kurtosis |
| Speed | -0.11 | 2.42 |
| Dist | 0.76 | 3.25 |

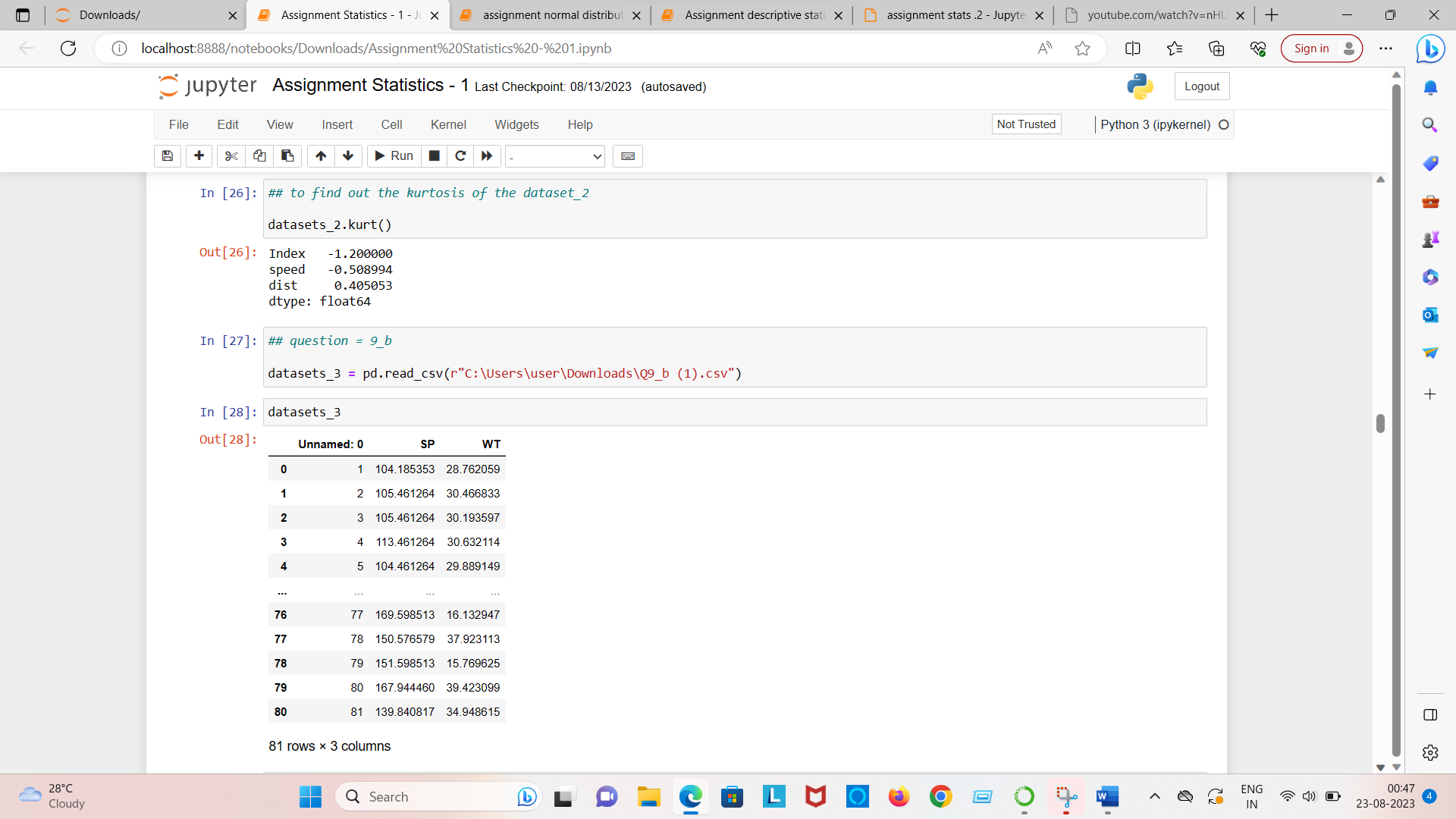
Dist is positively skewed where as speed is negatively skewed

1. Thus dist has distribution of data concentrated on the left whereas speed has distribution on the right .As seen in the graph
2. Both dist and speed has positive kurtosis







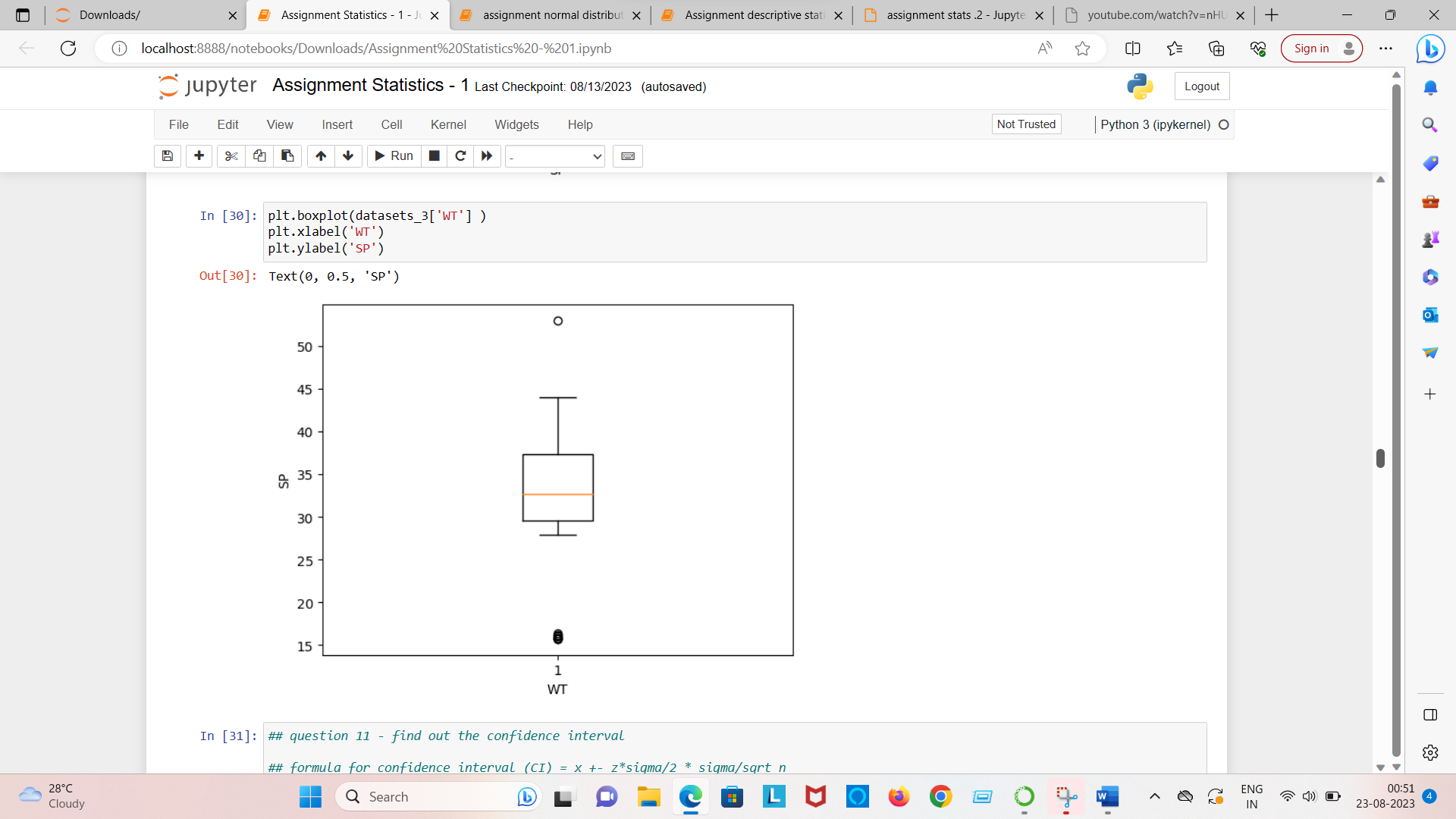


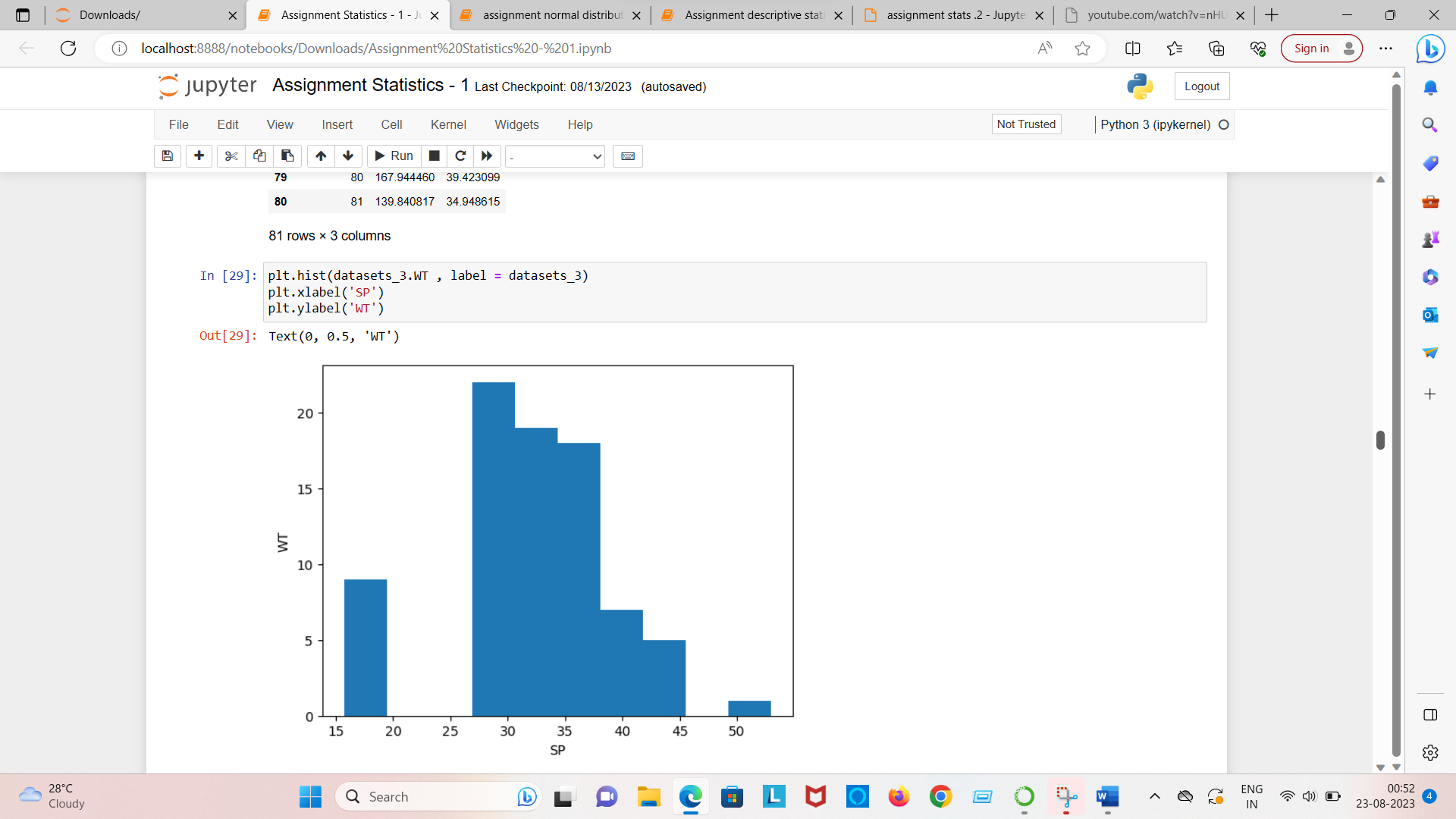
**Question no = 10 Draw the inferences about the following boxplot & histogram.**

Ans = Majority of the chicks has weight in range 50-150 , followed by 100-150 and 150-200

b.) the data is positively skewed

c.) Categories can be define (Under eight >50;Avg weight 51-150;Over weight <150)

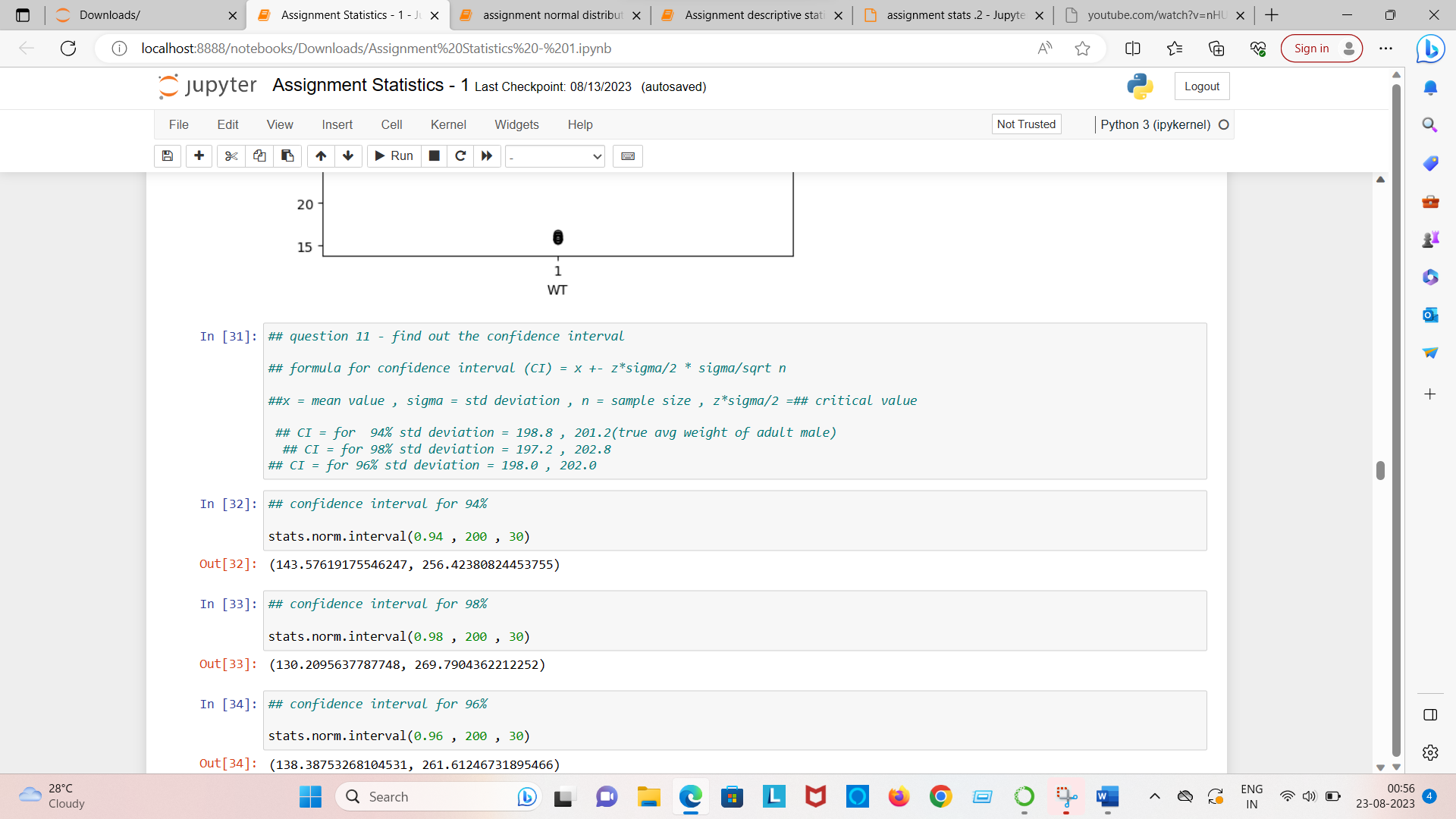




Answer no = 10 – the histograms peak has right skew and tail is on right . mean is greater than the median . outlier present on the higher side.

**Question no = 11**

**Suppose we want to estimate the average weight of an adult male in Mexico we draw a random sample of 2000 men from a population of 3000000 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 pound . calculate 94% , 98% , 96% confidence interval ?**

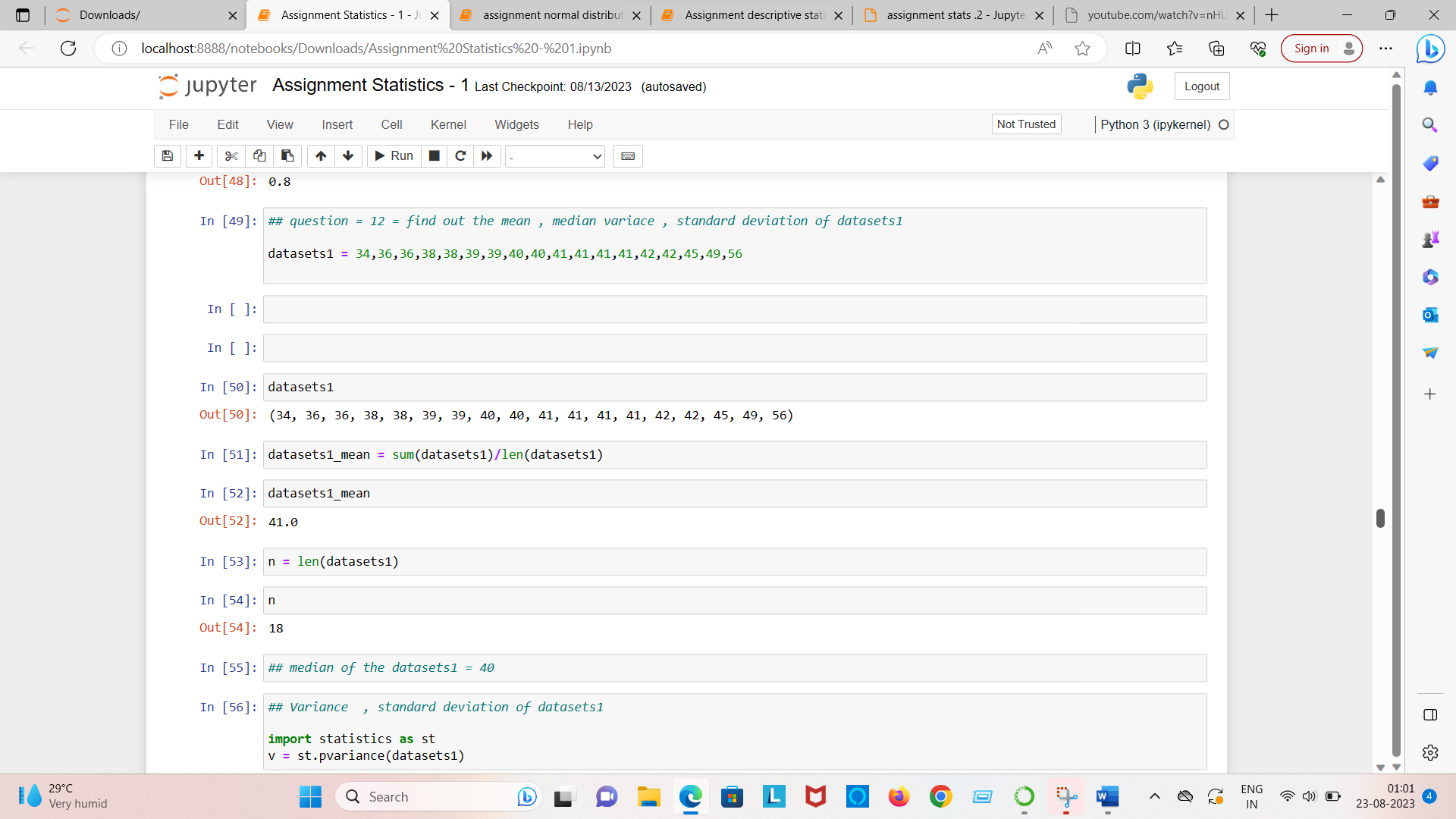
****

|  |  |  |  |
| --- | --- | --- | --- |
|  | 94% | 98% | 96% |
| Upper | 201.04 | 201.38 | 201.7 |
| Lower | 198.96 | 198.62 | 198.83 |

**Question no = 12 Below are the scores obtained by a student in tests?**

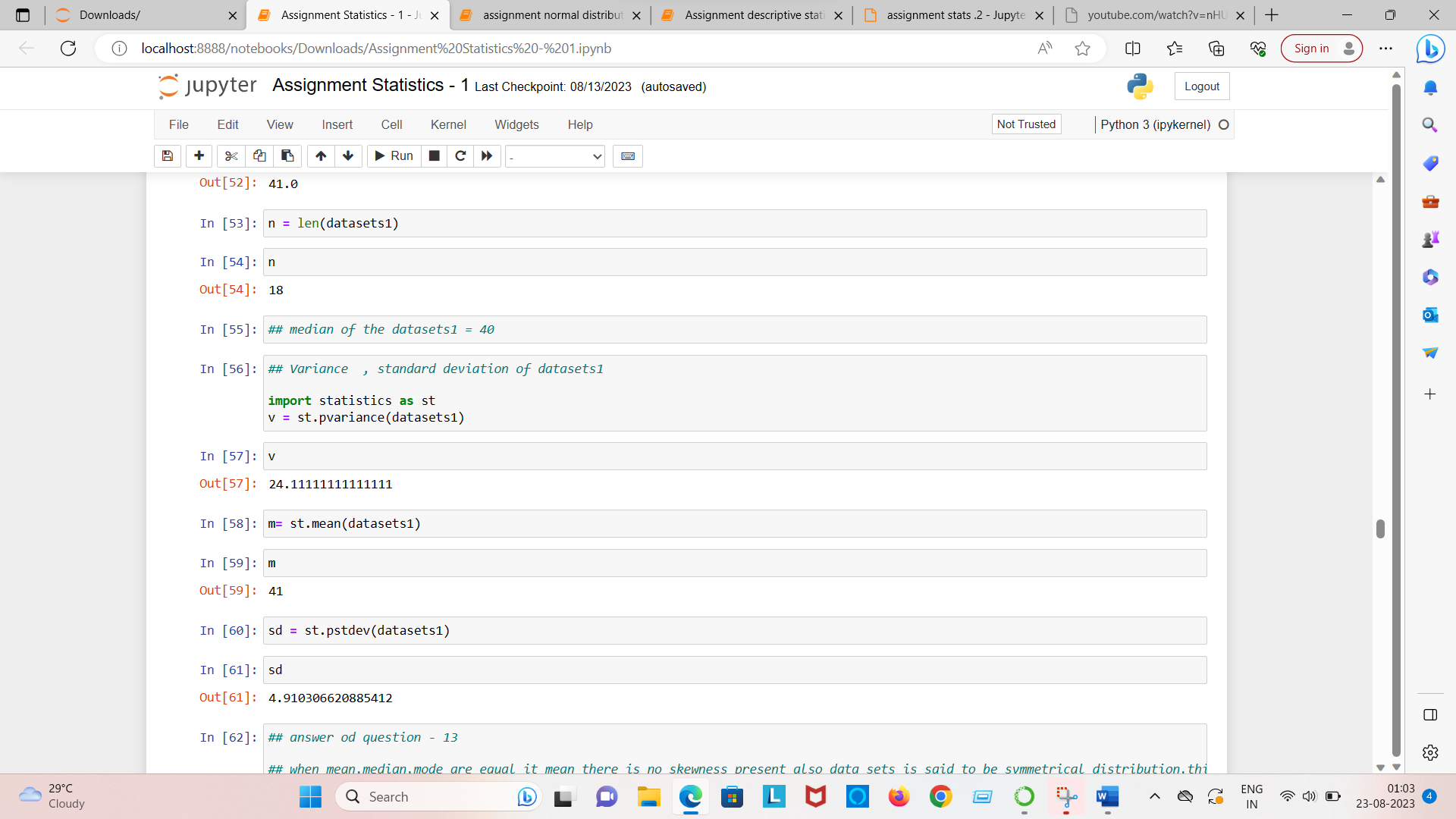
**Answer =**

|  |  |
| --- | --- |
| **Mean** | **41** |
| **Median** | **40.50** |
| **Variance** | **25.53** |
| **Std Deviation** | **5.05** |

****

What can we say about the student marks

1. Not normally distributed
2. Data has outlier
3. Majority of the data students scored between 35-45 Marks



Question no = 13 what is the nature of skewness , when mean , median of data are equal?

Answer no = skewness = 0 , No skewness present , it is perfect bell shaped curve

Question no = 14 what is the nature of the skewness , when mean>median?

Answer = positive skewed, data is distributed more on left

Question no = 15 what is the nature of skewness when median>mean

Answer = Negative skewed , most of data is distributed right side

Question no = 16 what does the positive kurtosis value indicates for a data?

Answer = High and narrow peak on central part of data

Question no = 17 what does negative kurtosis value indicates for a data?

Answer = wider peak on the central part of the data .

Question = 18 Answer the below question using the below boxplot visualization.

What can we say about the distribution of the data?

What is nature of skewness of the data?

What will be the IQR of the data (approximately)?

**Question no = 18 what can we say about the distributed of the data**

**Ans = The data is not symmetric,Data is more concentrated towards right side.**

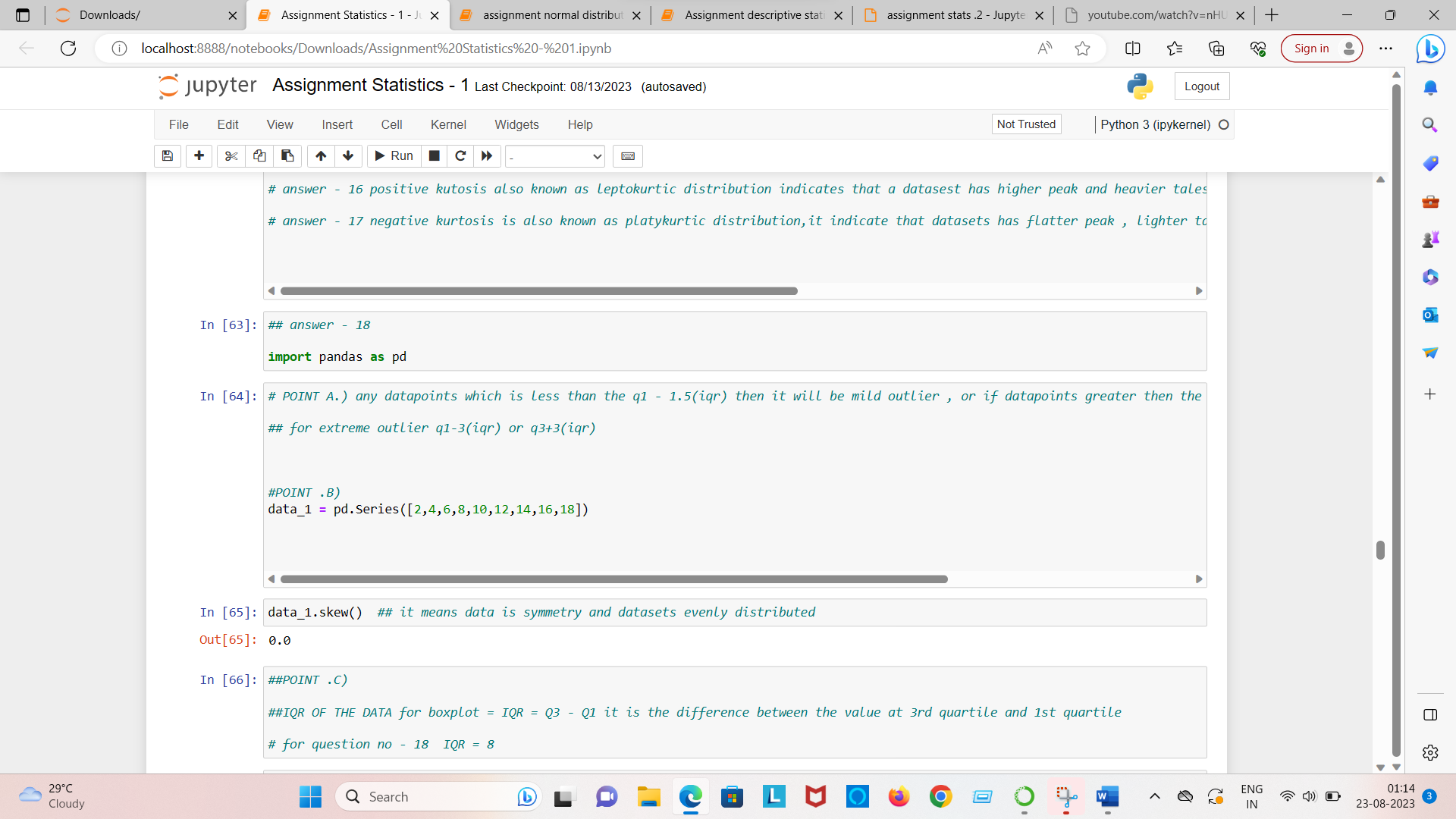
What is nature of skewness of the data

Ans = skewness = Negative

What will be the IQR of the data (approximately)

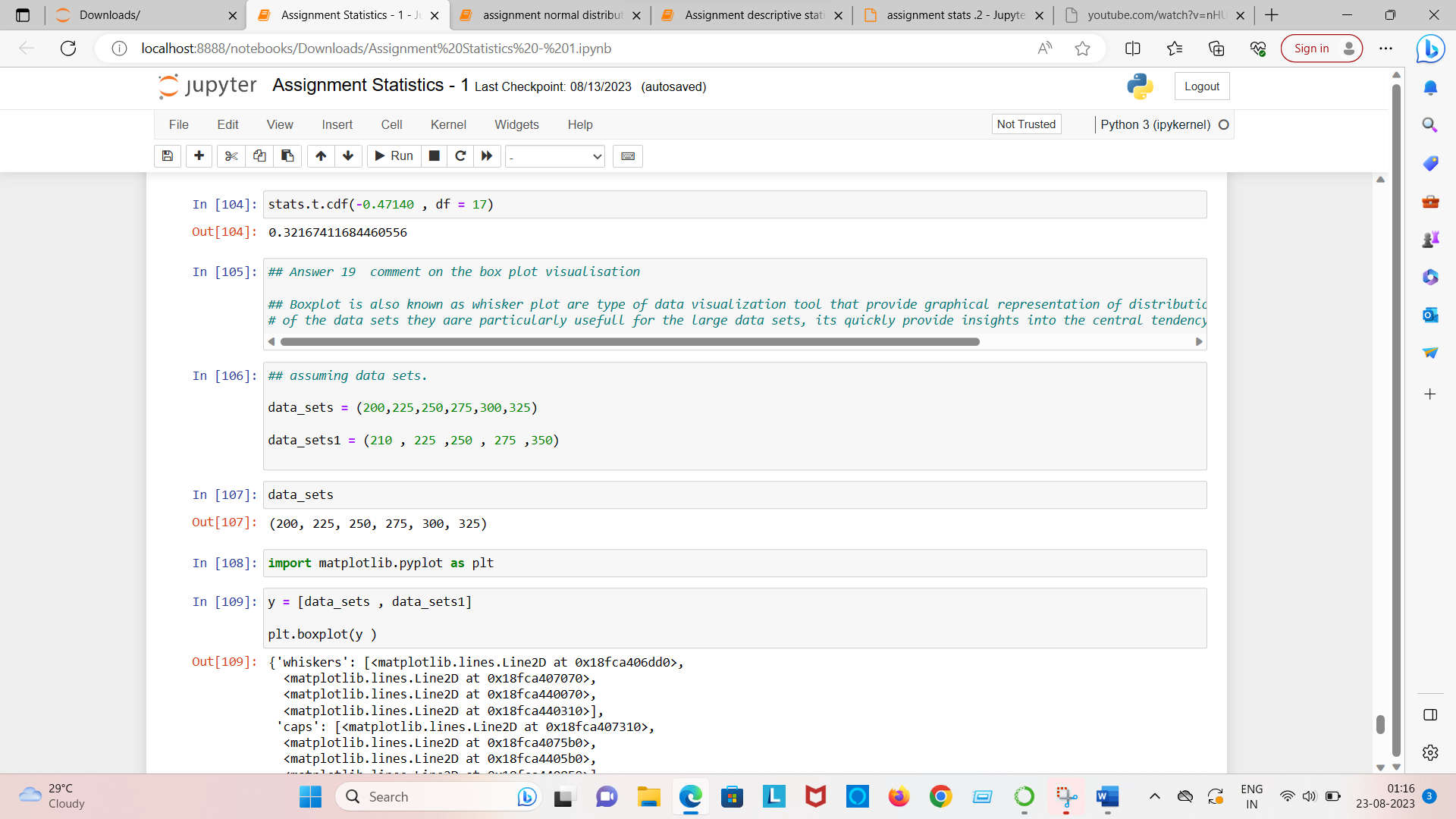
The IQR of this data is = Q3 – Q1

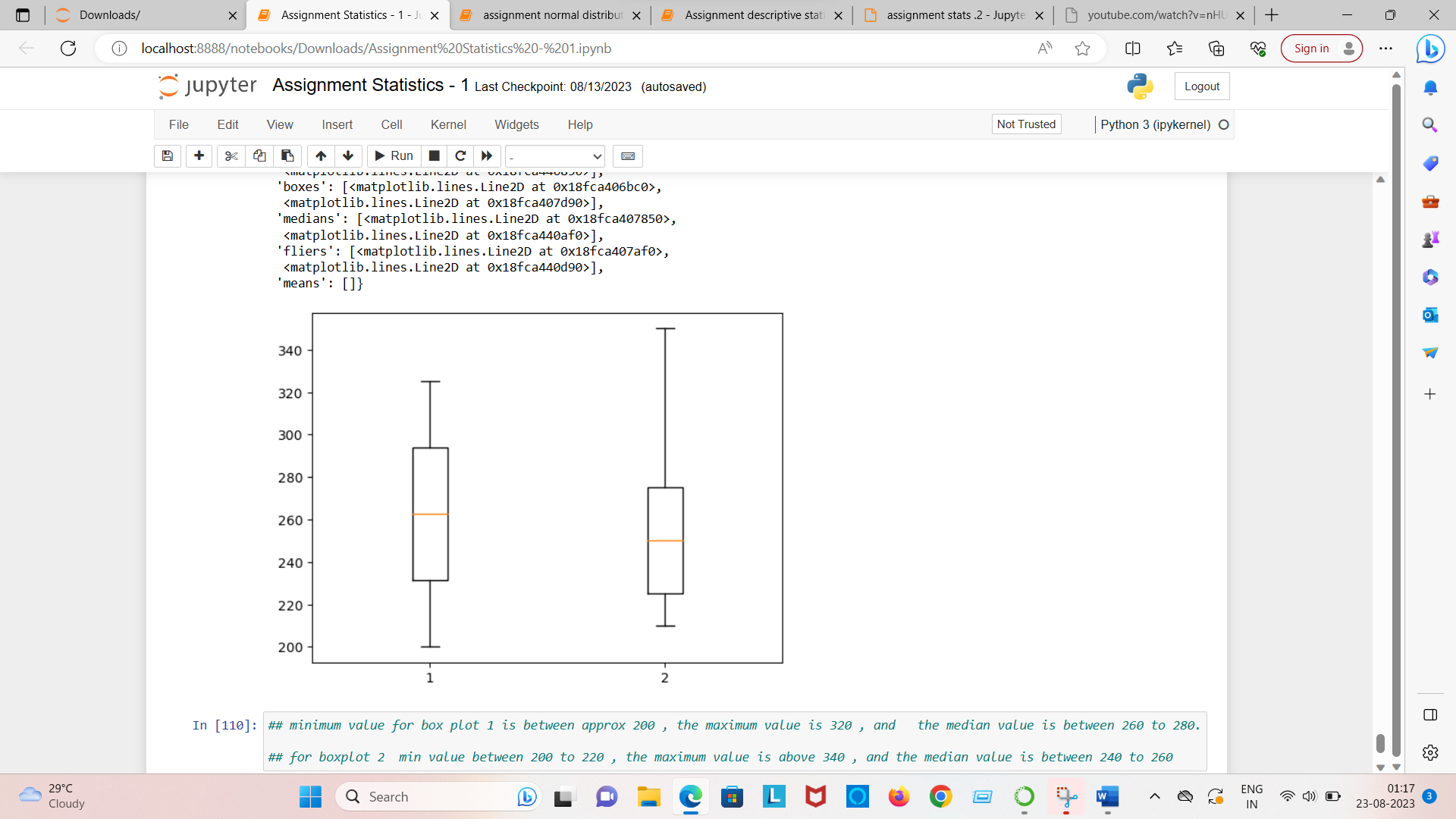
18 – 10 = 8



Question no = 19 Draw an inference from the distribution of data for boxplot 1 with respect boxplot 2

1. Data is normally distributed, No outlier, center around 262.5. comparatively, first graph has less range
2. Data is normally distributed , no outlier, center around 262.5 comparatively, second graph has more range





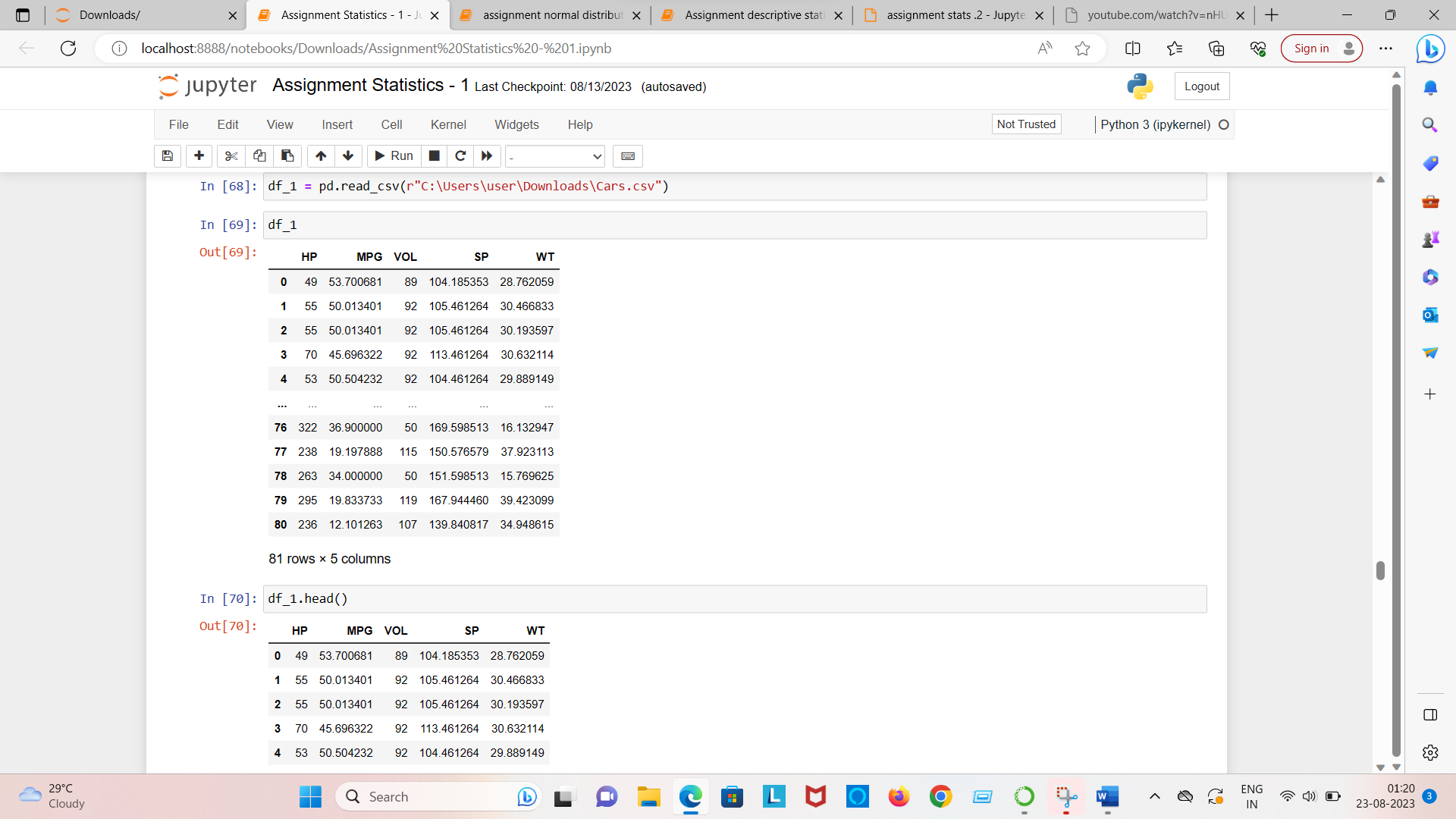
Question no = 20 Calculate the probability from the given data sets for the below cases.

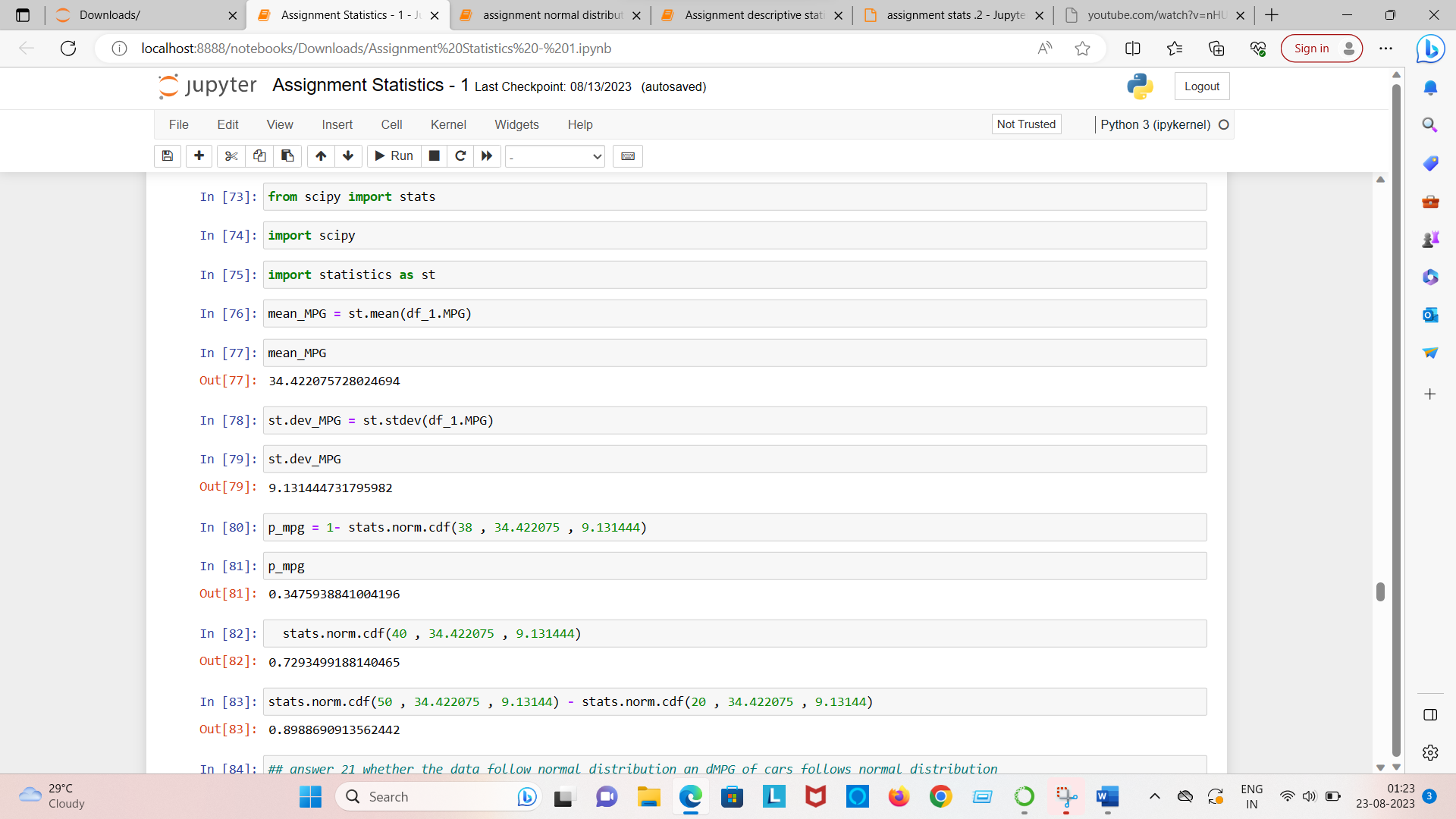
1. P(mpg>38)
2. P(mpg<40)
3. P(20<mpg<50)

Answer a.) = 0.407

Answer b.) = 0.75

Answer c.) = 0.85





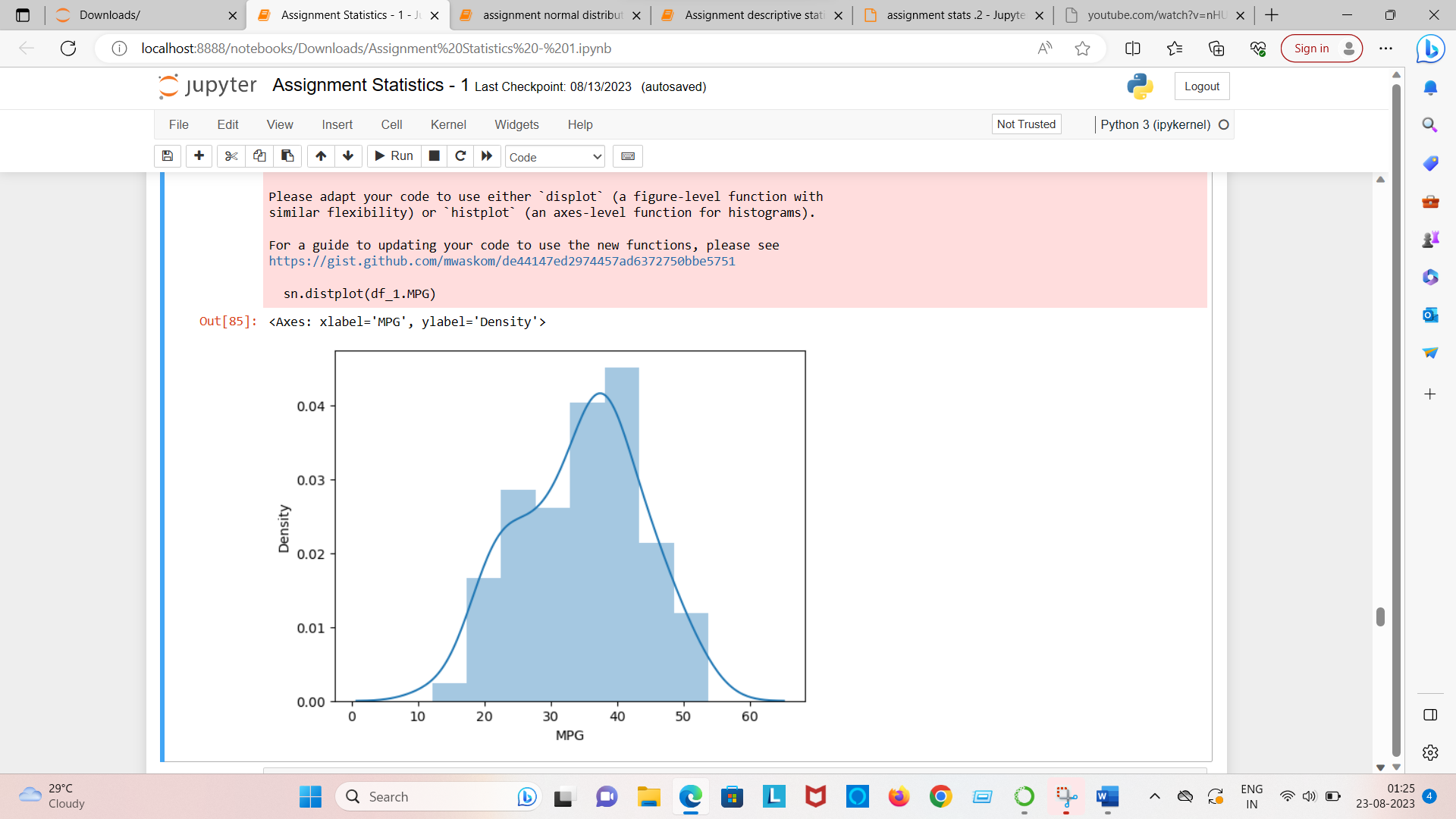
Question no = 21 Check whether data follows normal distribution

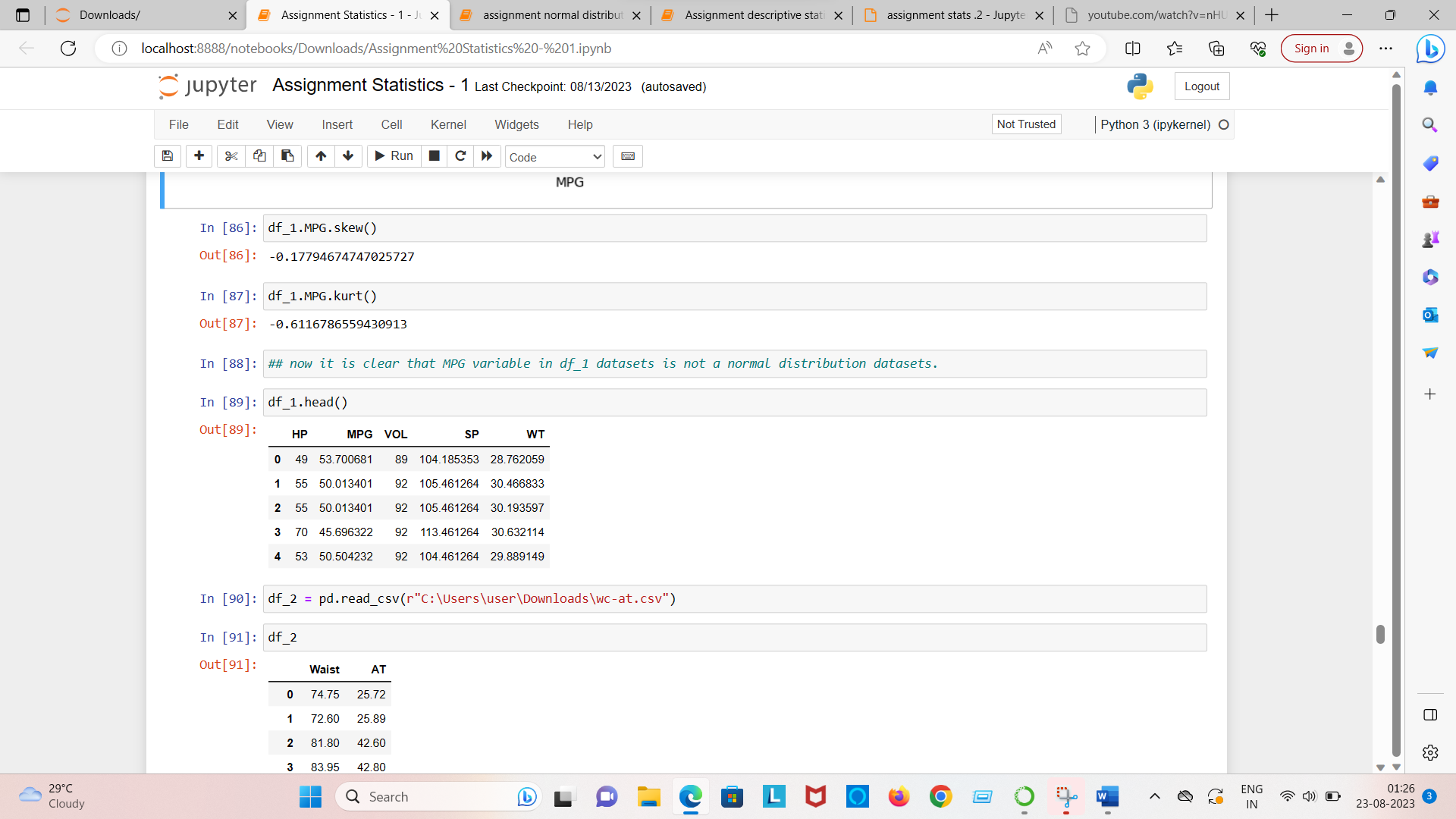
1. Check whether the MPG of cars follows normal distribution dataset : cars.csv
2. Check whether the adipose tissue (AT) and waist circumference (waist) from wc-at data set follows normal distribution .

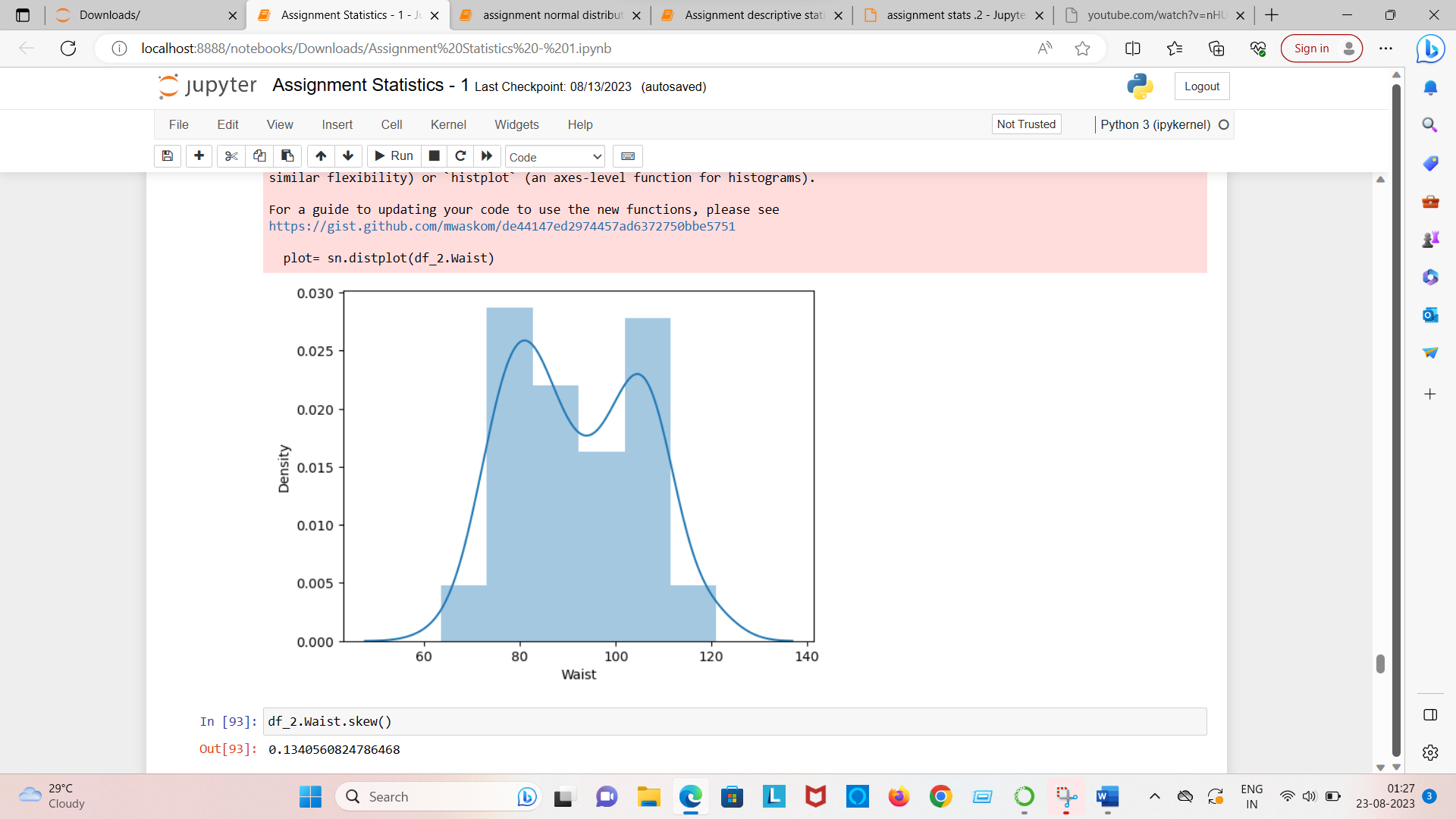
Data:wc-at.csv

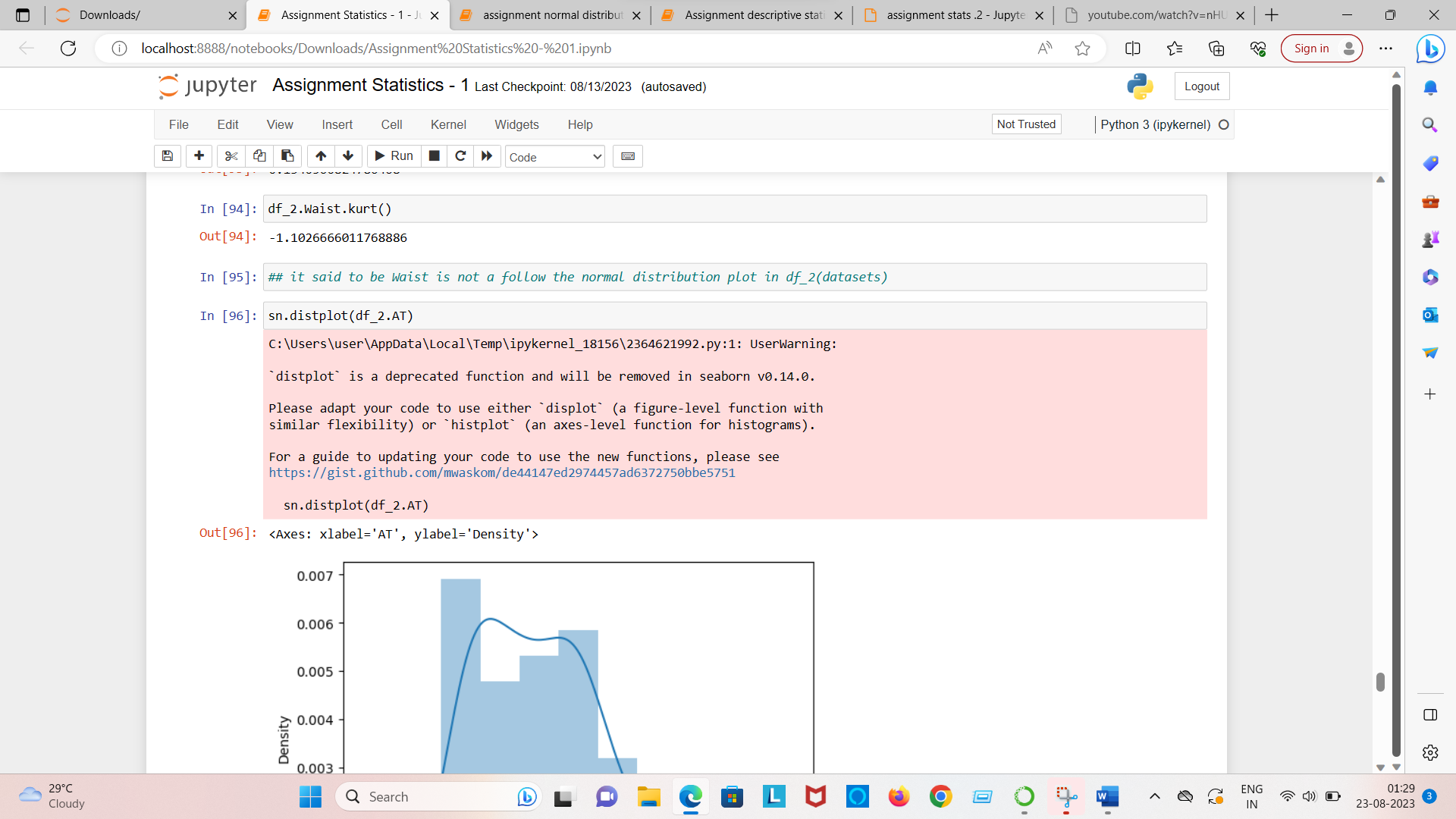
Answer a.) = MPG is normally distributed

Answer b.) = Both AT and Waist does not follow Normal Distribution



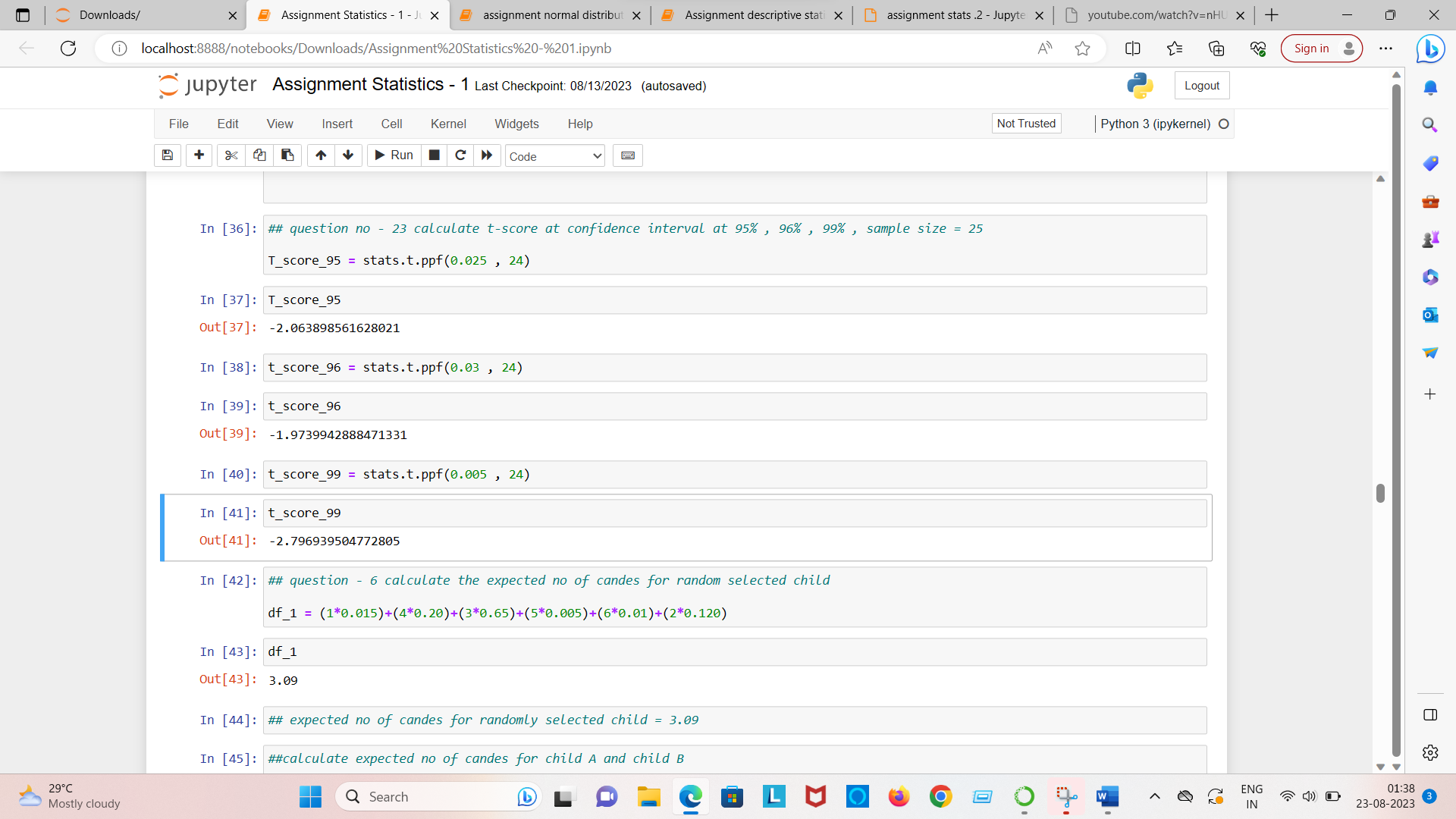






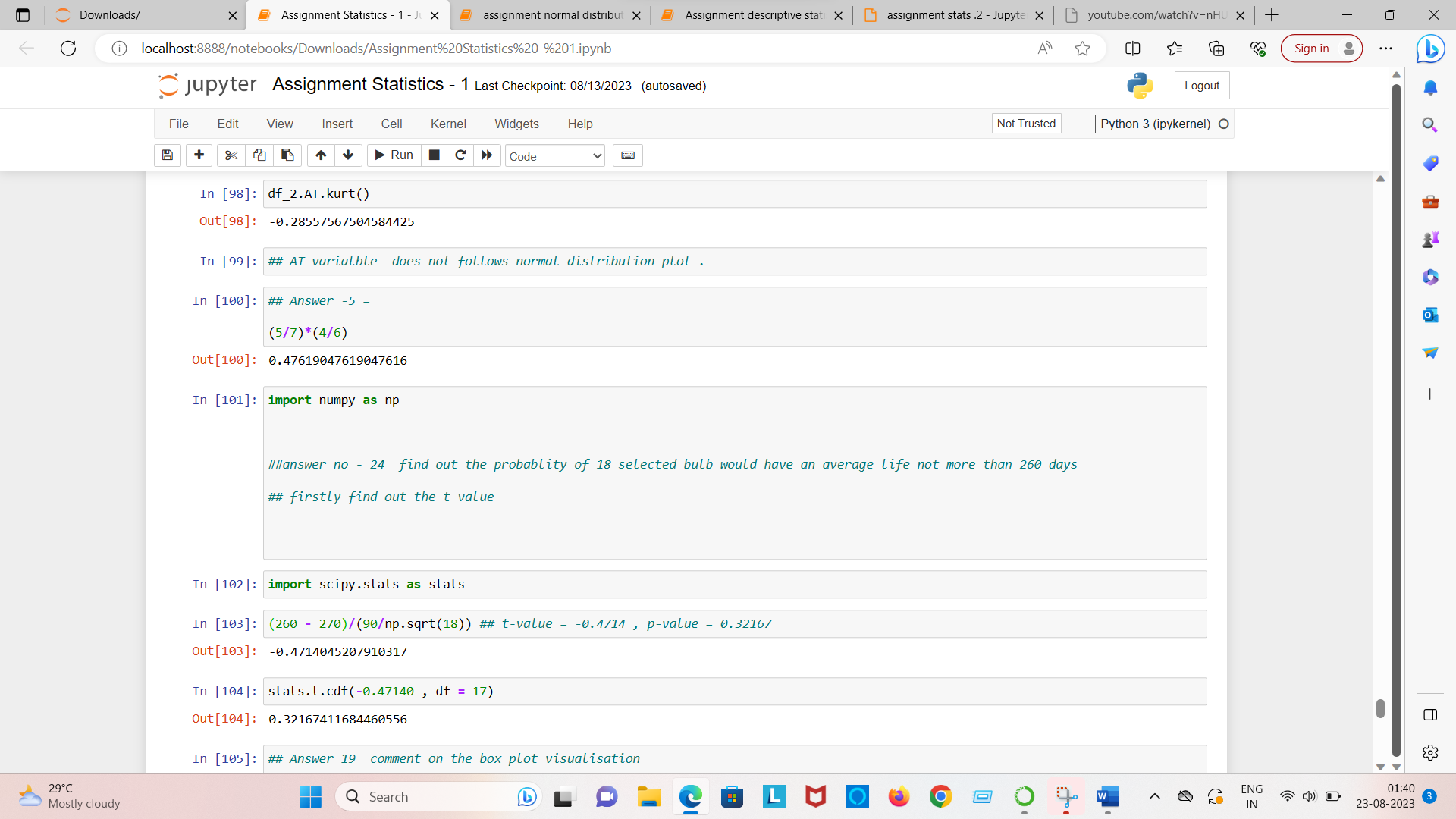
Question no = 23 Calculate the t-scores of 95% confidence interval , 96% confidence interval , 60% confidence interval for the sample size of 25?

|  |  |
| --- | --- |
| 95% | +-2.060 |
| 96% | +-2.16 |
| 99% | +-2.78 |



Question no = 24 A Government company claims that an average light bulb last 270 days , A researcher randomly select 18 bulbs for testing the sampled bulbs last average of 250 days with a standard deviation od 90 days if the CEO claims is true, what is the probability

that 18 randomly selected bulb would have an average life of no more than 260 days ?



t-score = -0.4714

degree of freedom = 17

p(t) = 0.3216

Question no = 22 = Calculate the z-score of of 90% confidence interval , 94% confidence interval , 60% confidence interval.

ANSWER

|  |  |
| --- | --- |
| 90% | +-1.711 |
| 94% | +-1.828 |
| 60% | +-2.492 |

