|  |  |
| --- | --- |
| 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 | Nominal |
| Level of Agreement | Ordinal |
| IQ(Intelligence Scale) | Ratio |
| Sales Figures | Ratio |
| Blood Group | Nominal |
| Time Of Day | Interval |
| Time on a Clock with Hands | Interval |
| Number of Children | Nominal |
| Religious Preference | Ordinal |
| Barometer Pressure | Interval |
| SAT Scores | Ordinal |
| Years of Education | Interval |

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

{HHH, HHT, HTH, THH, TTT, TTH, THT, HTT}

Total Number of Outcome=8

Probability of two heads and one tail is = {HHT, THH, HTH}

P= number of favorable outcomes/total num of outcomes

P=3/8=0.375

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

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

Total outcomes =6 \*6 =36

Possibilities = (1,1), (1,2), (1,3), (1,4), (1,5), (1,6)

(2,1), (2,2), (2,3), (2,4), (2,5), (2,6)

(3,1), (3,2), (3,3), (3,4), (3,5), (3,6)

(4,1), (4,2), (4,3), (4,4), (4,5), (4,6)

(5,1), (5,2), (5,3), (5,4), (5,5), (5,6)

(6,1), (6,2), (6,3), (6,4), (6,5), (6,6)

1. No of favorable outcomes=0

Probability of getting sum is equal to 1 r=0

1. No of favorable outcomes=6 (1,1),(1,2), (1,3), (2,1), (2,2), (3,1)

=6/36 = 1/6

1. Num of favorable outcomes=6

(1,5), (2,4), (2,5), (2,6), (3,2), (3,6)

Probability of getting sum is divisible by 2 and 3 is

=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?

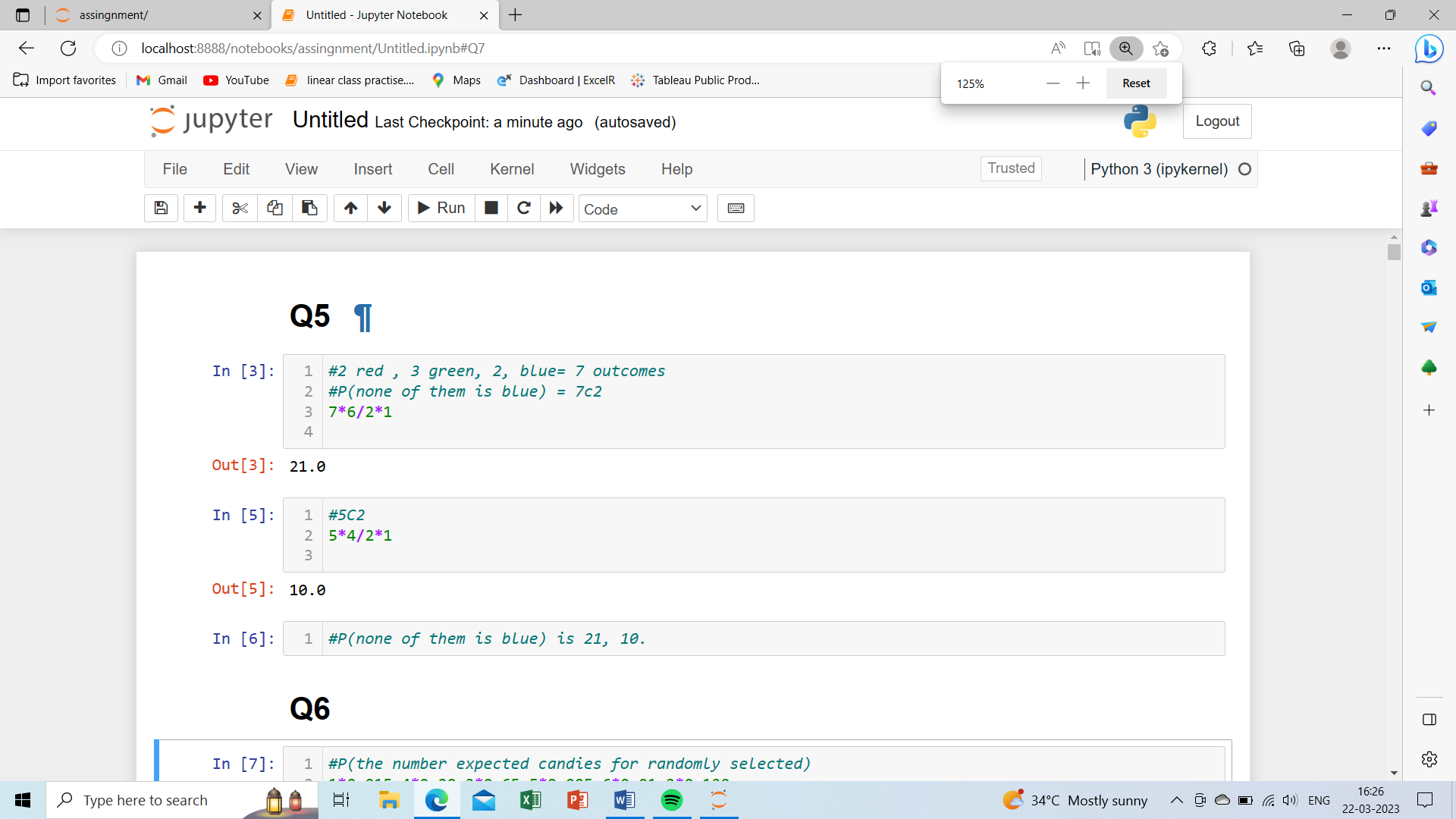
2 red , 3 green, 2, blue= 7 outcomes

P(none of them is blue) = 7c2

7\*6!/2\*1 = 42/2 = 21

5c2 = 5\*4/2\*1 = 20/2 = 10

P(none of them is blue) is 21, 10.



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)

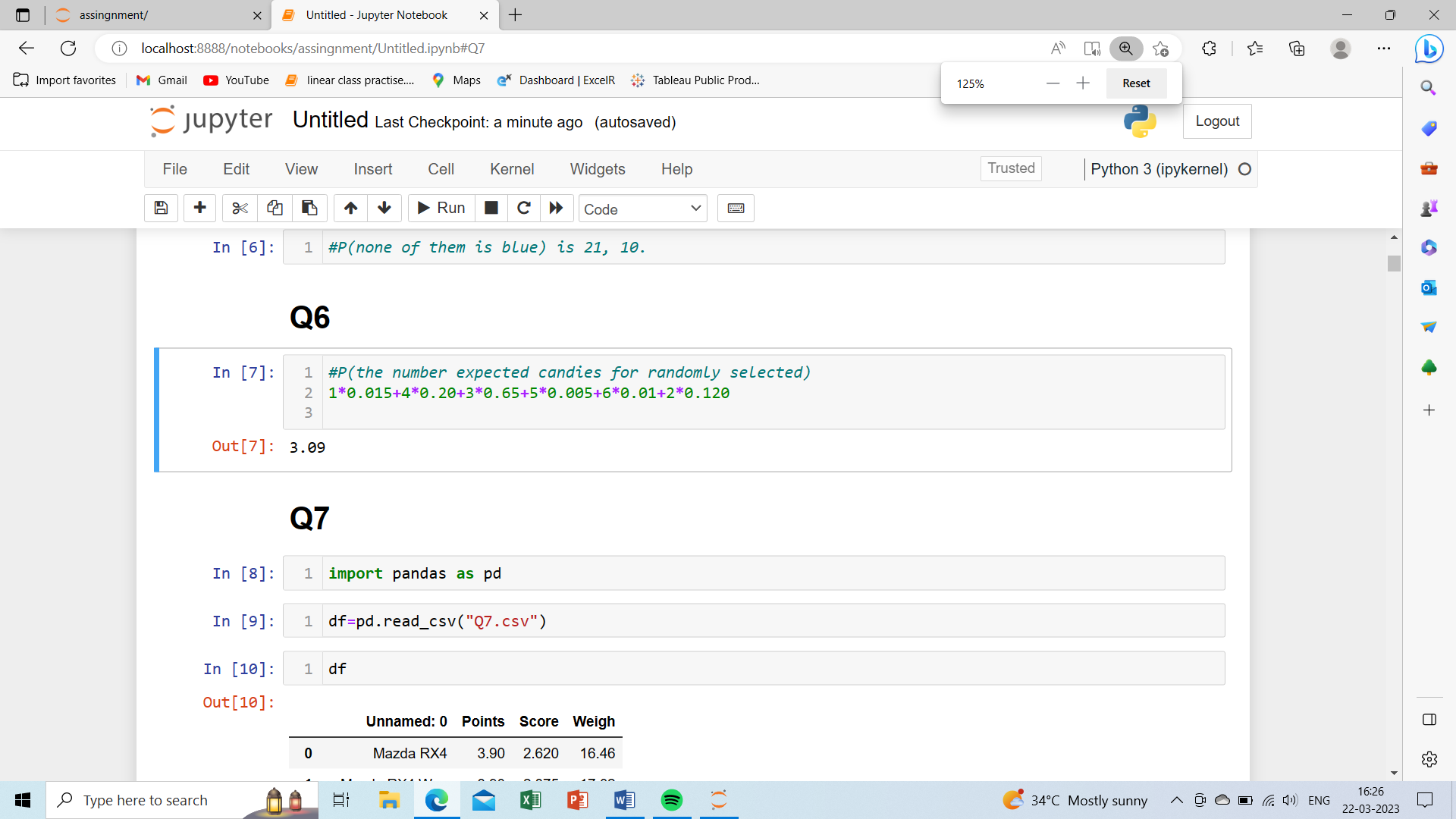
|  |  |  |
| --- | --- | --- |
| CHILD | Candies count | Probability |
| A | 1 | 0.015 |
| B | 4 | 0.20 |
| C | 3 | 0.65 |
| D | 5 | 0.005 |
| E | 6 | 0.01 |
| F | 2 | 0.120 |

Child A – probability of having 1 candy = 0.015.

Child B – probability of having 4 candies = 0.20

P(the number expected candies for randomly selected)

1\*0.015+4\*0.20+3\*0.65+5\*0.005+6\*0.01+2\*0.120 = 3.09



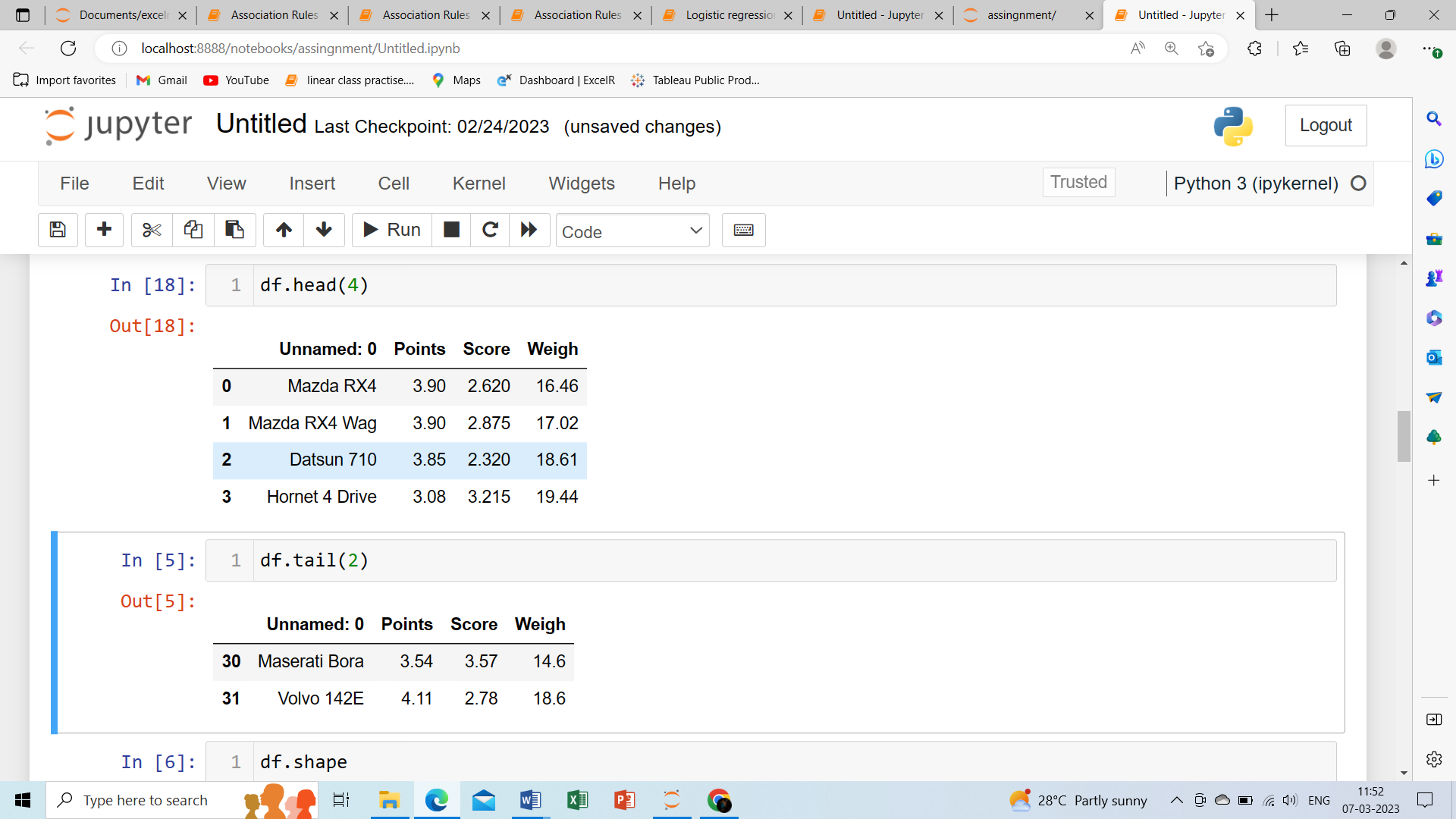
Q7) Calculate Mean, Median, Mode, Variance, Standard Deviation, Range & comment about the values / draw inferences, for the given dataset

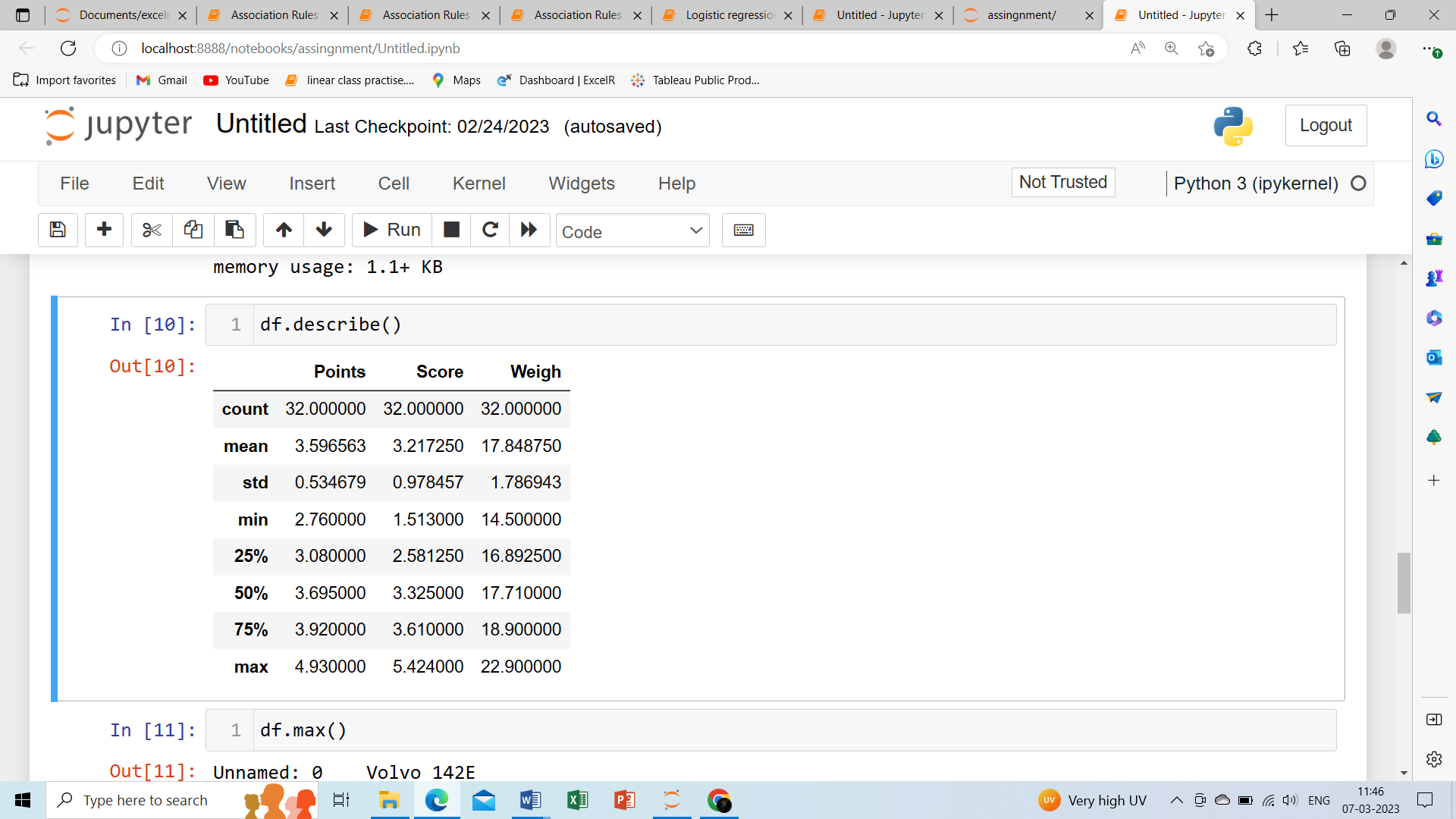
* For Points, Score, Weigh>

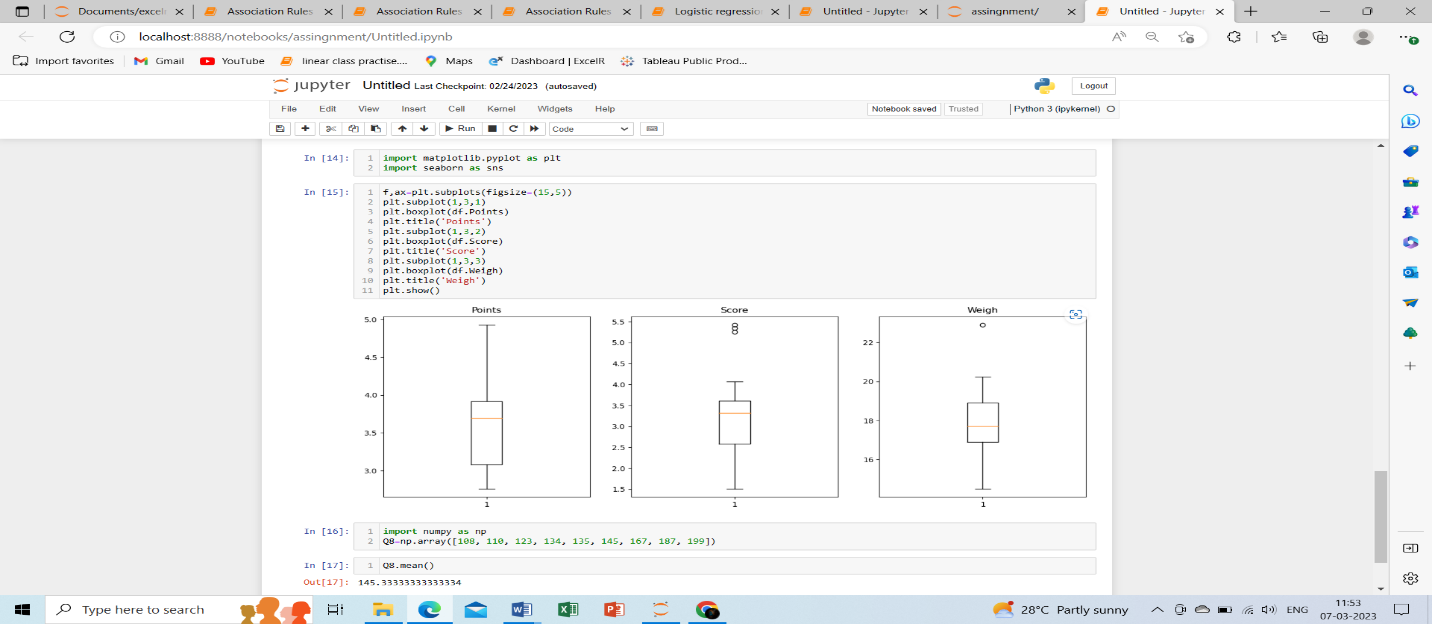
Find Mean, Median, Mode, Variance, Standard Deviation, and Range and also Comment about the values/ Draw some inferences.

**Use Q7.csv file**

Solution:







Ans:

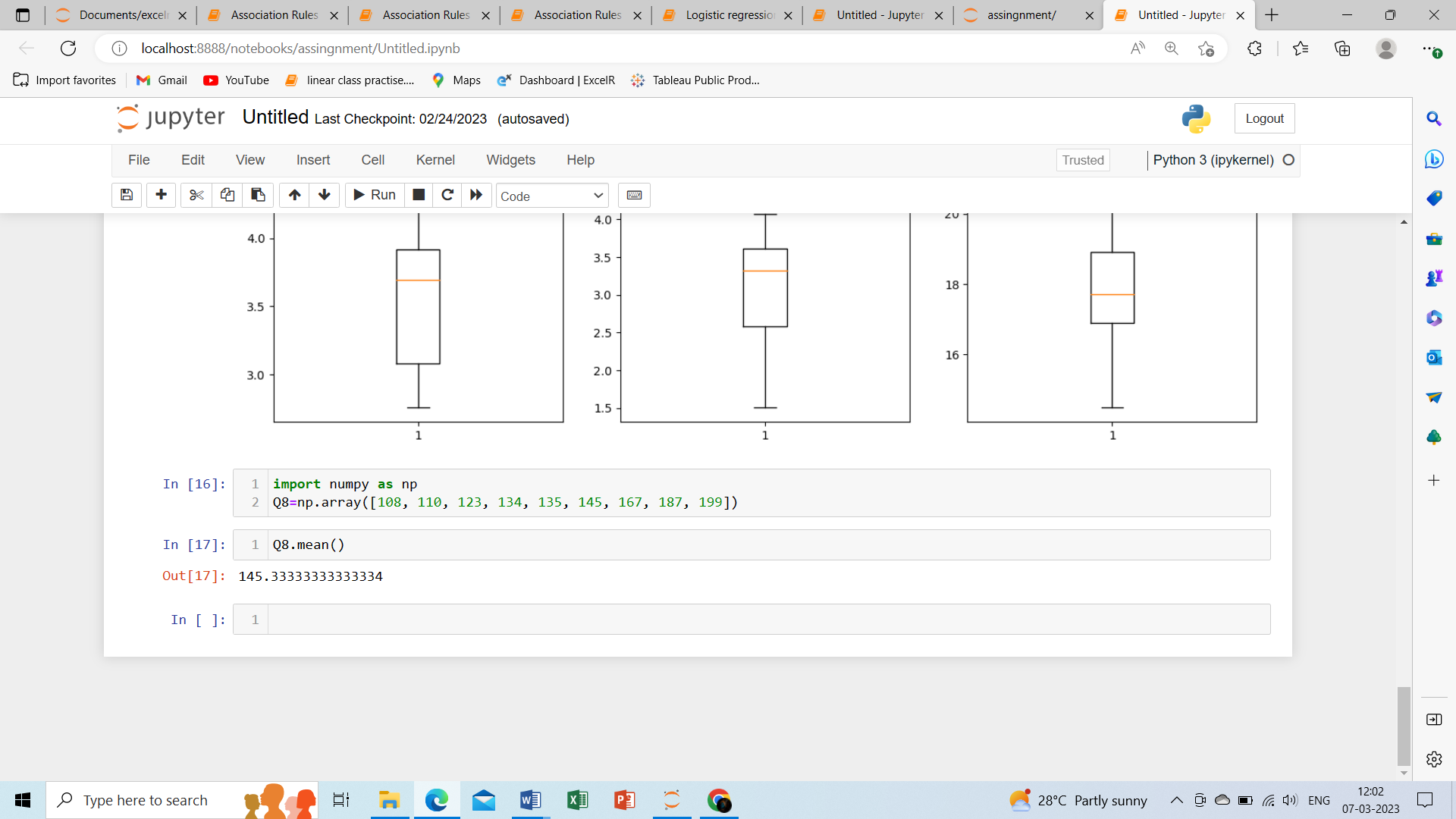
1. In case of points and scores, the median>mean, so the distribution is right skewed.
2. In case of weight the mean > median, so the distribution is left skewed.

Q8) Calculate Expected Value for the problem below

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

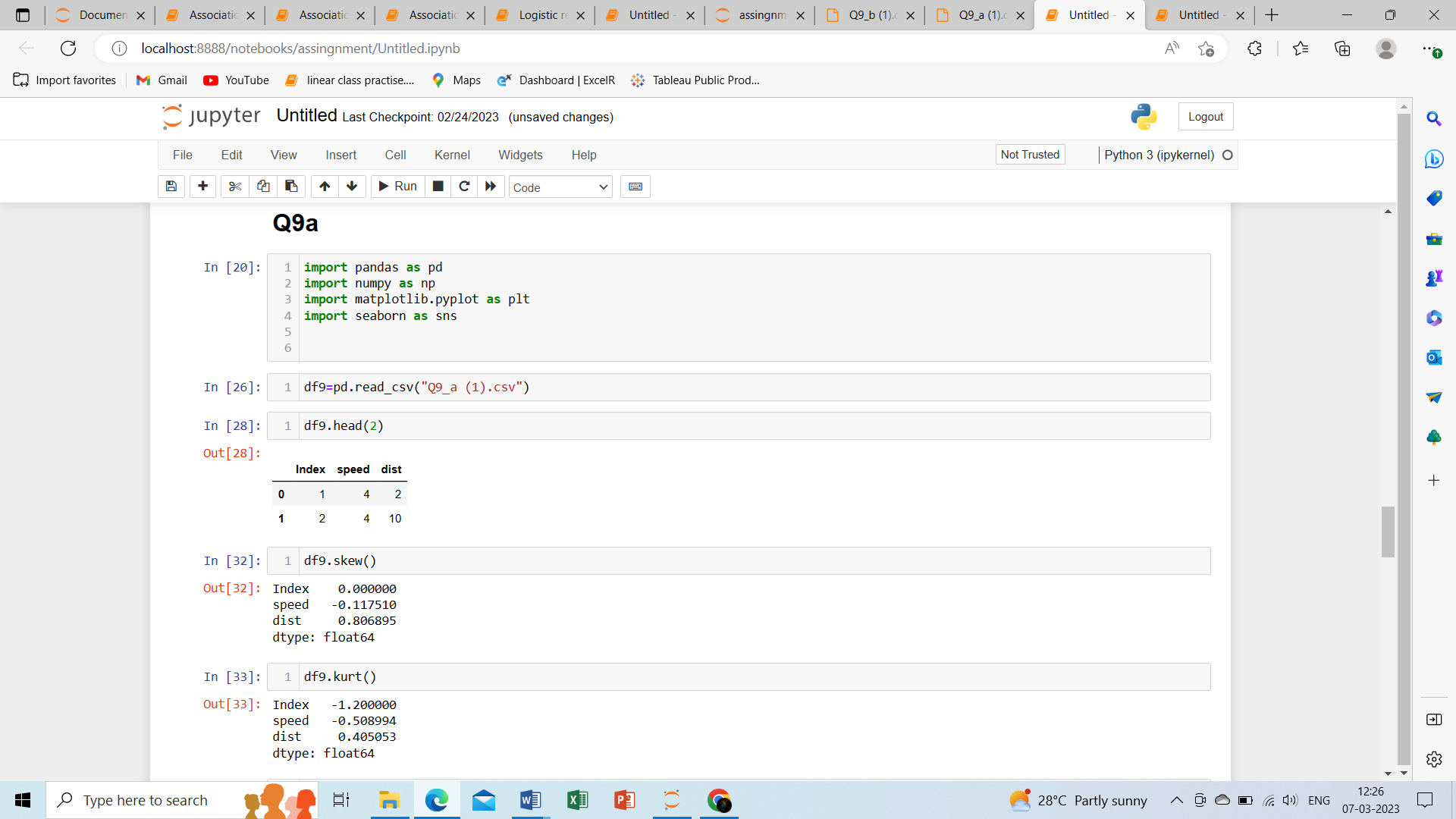
108, 110, 123, 134, 135, 145, 167, 187, 199

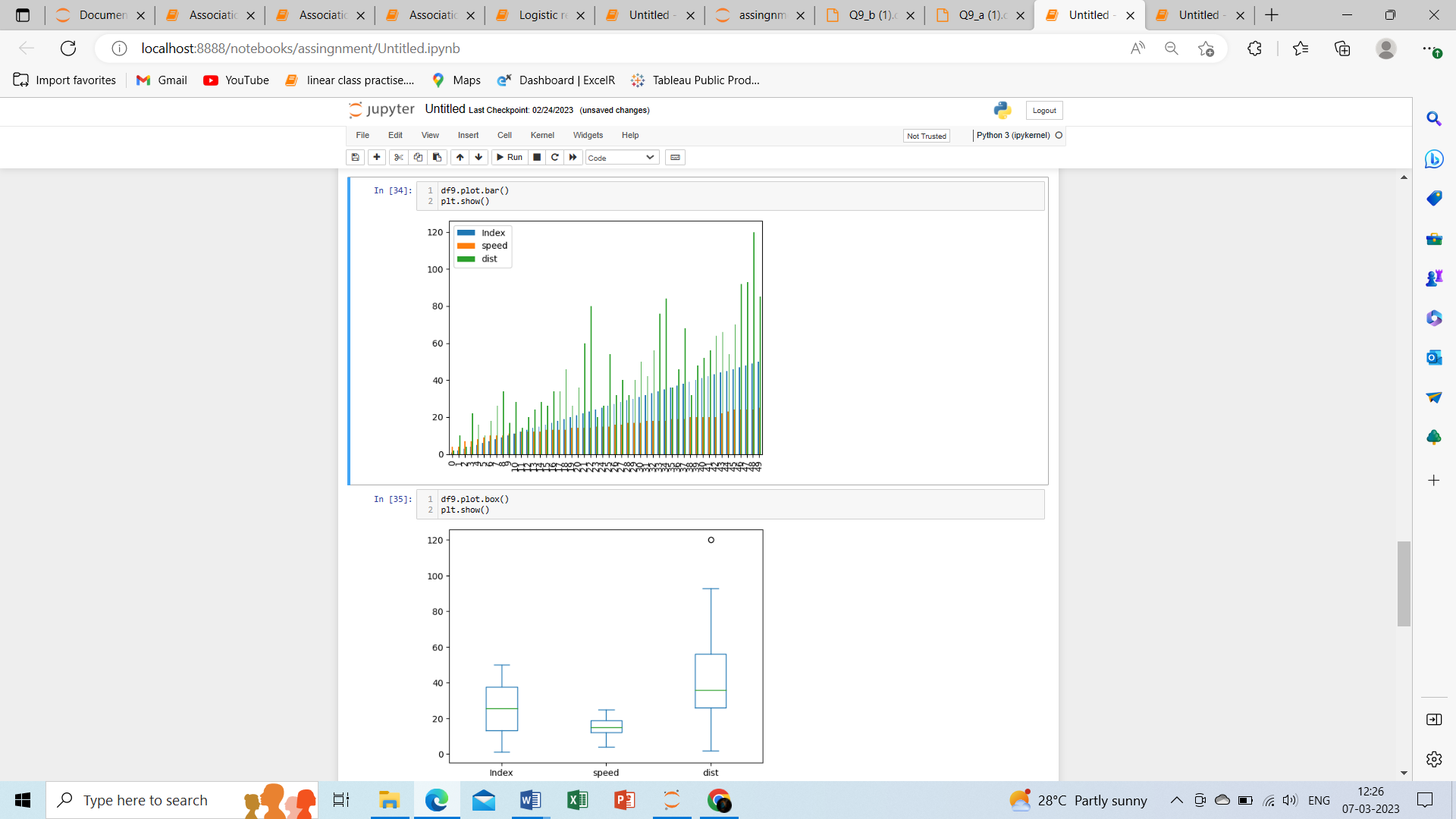
Assume one of the patients is chosen at random. What is the Expected Value of the Weight of that patient?



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

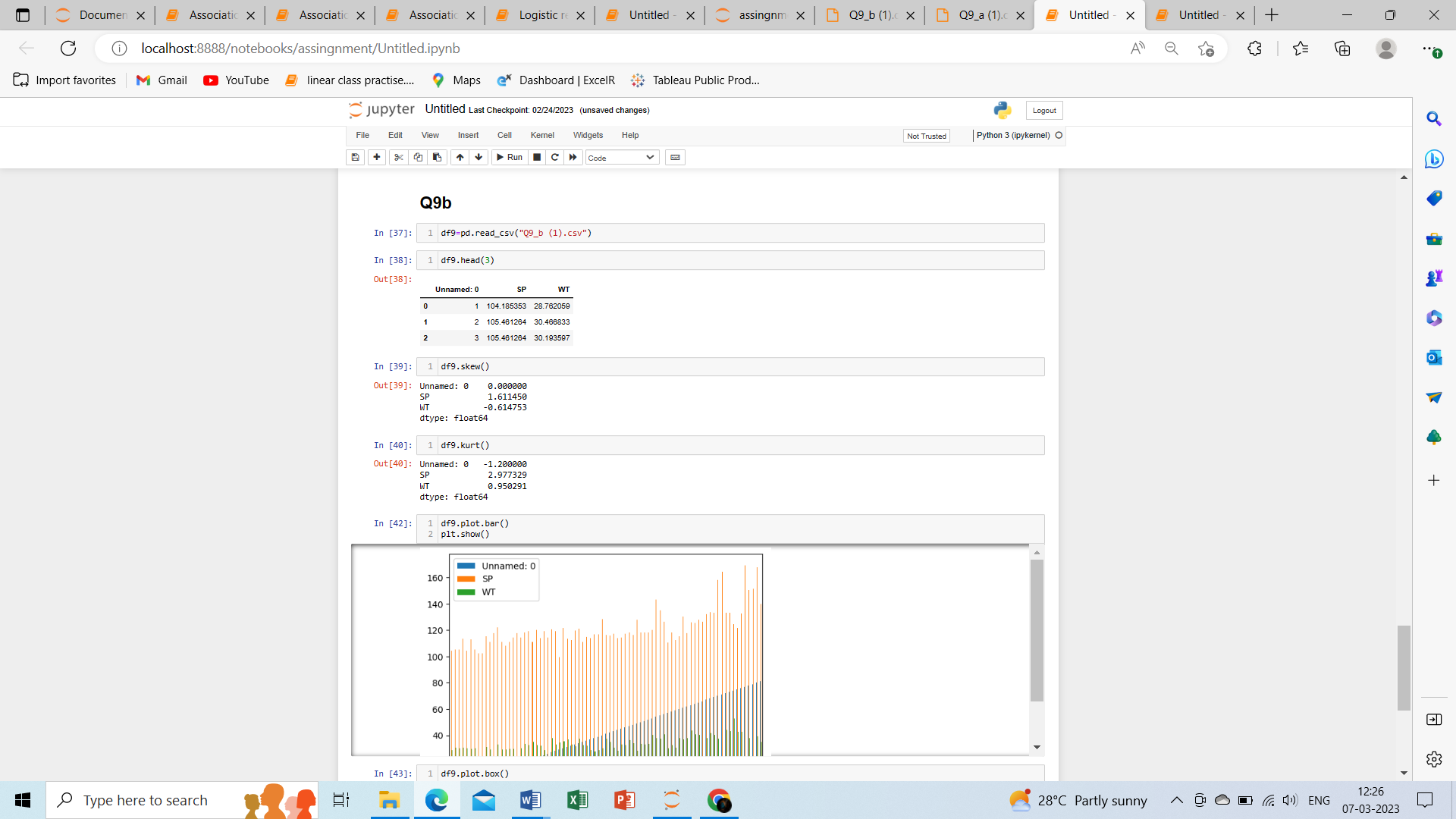
**Cars speed and distance Use Q9\_a.csv SP and Weight(WT)**

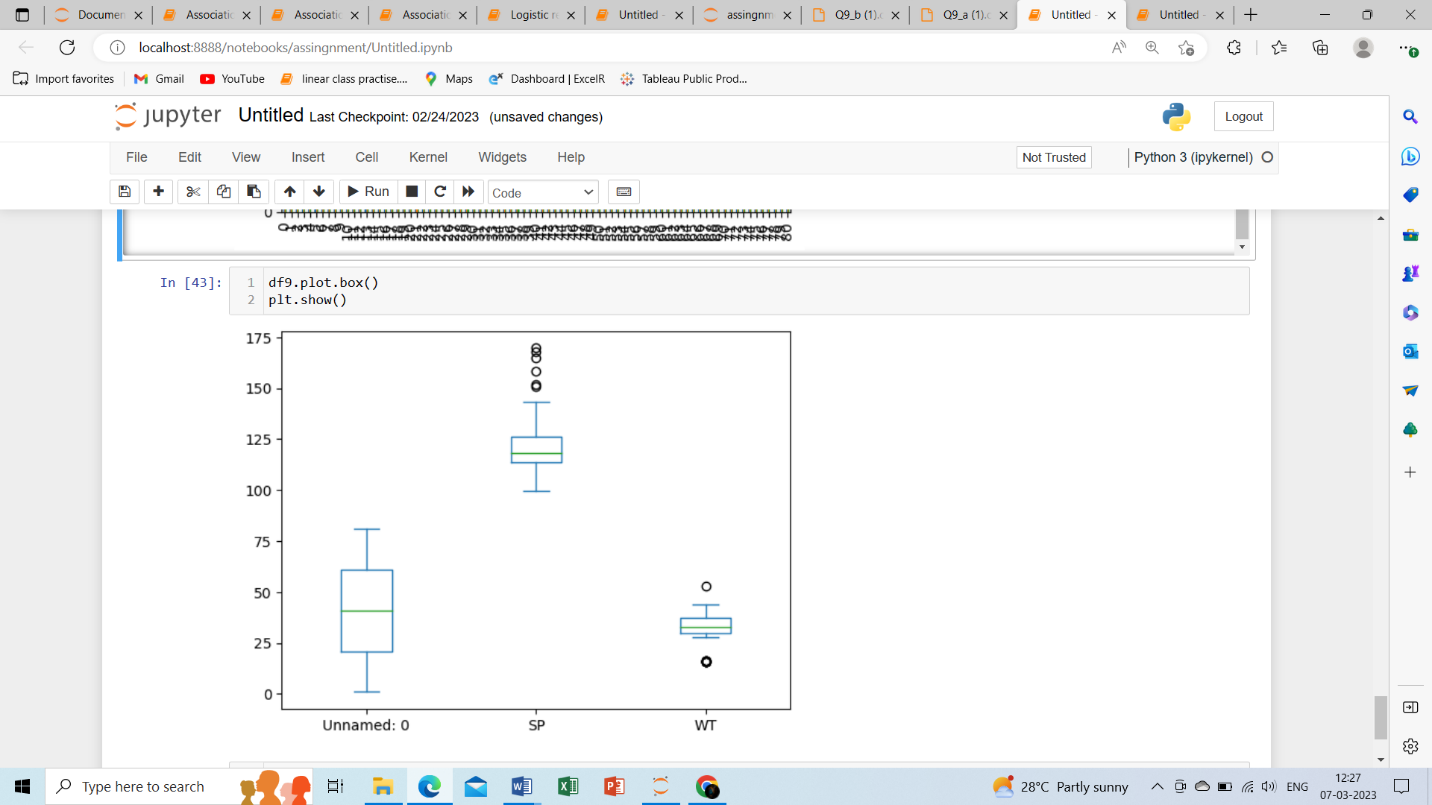
****

****

**Ans :** For speed and distance we notice that the speed has negative skewness

(-value) so it is right skewed, and the distance has positive skewness(+value) So it is left skewed.

. ****

****

**Ans:** For Wt and Spwe notice that Wt has negative skewness(-value) so left skewed, Sp has positive skewness(+value) so right skewed.

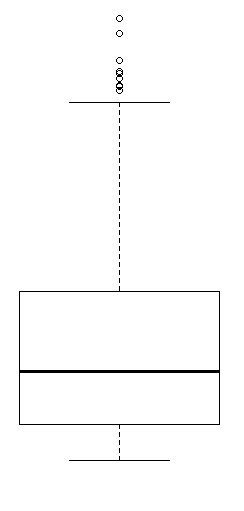
**Use Q9\_b.csv**

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



Ans: The most of data points are in b/w the range of frequency 50-100 with weight 200, And the least range of weight 400 lies b/w the 0-10, so the expected value of above the range is 75.

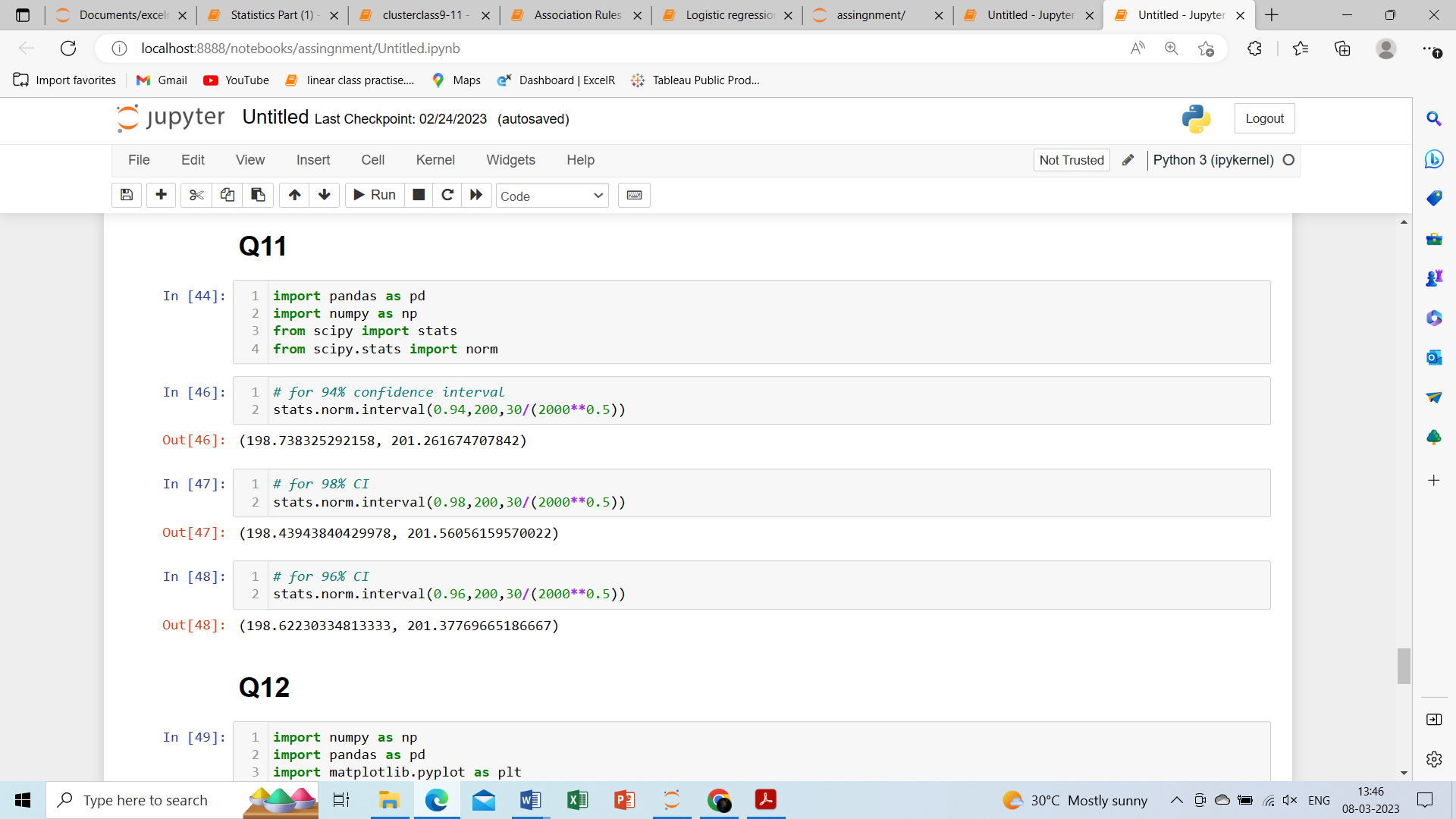
we can notice that long tail towards right so it is right skewed.



Ans: Here in this box plot we can infer number of outliers values are upper extrem ,

it shows right skewed, there is less data point in between the Q1 and the bottom point.

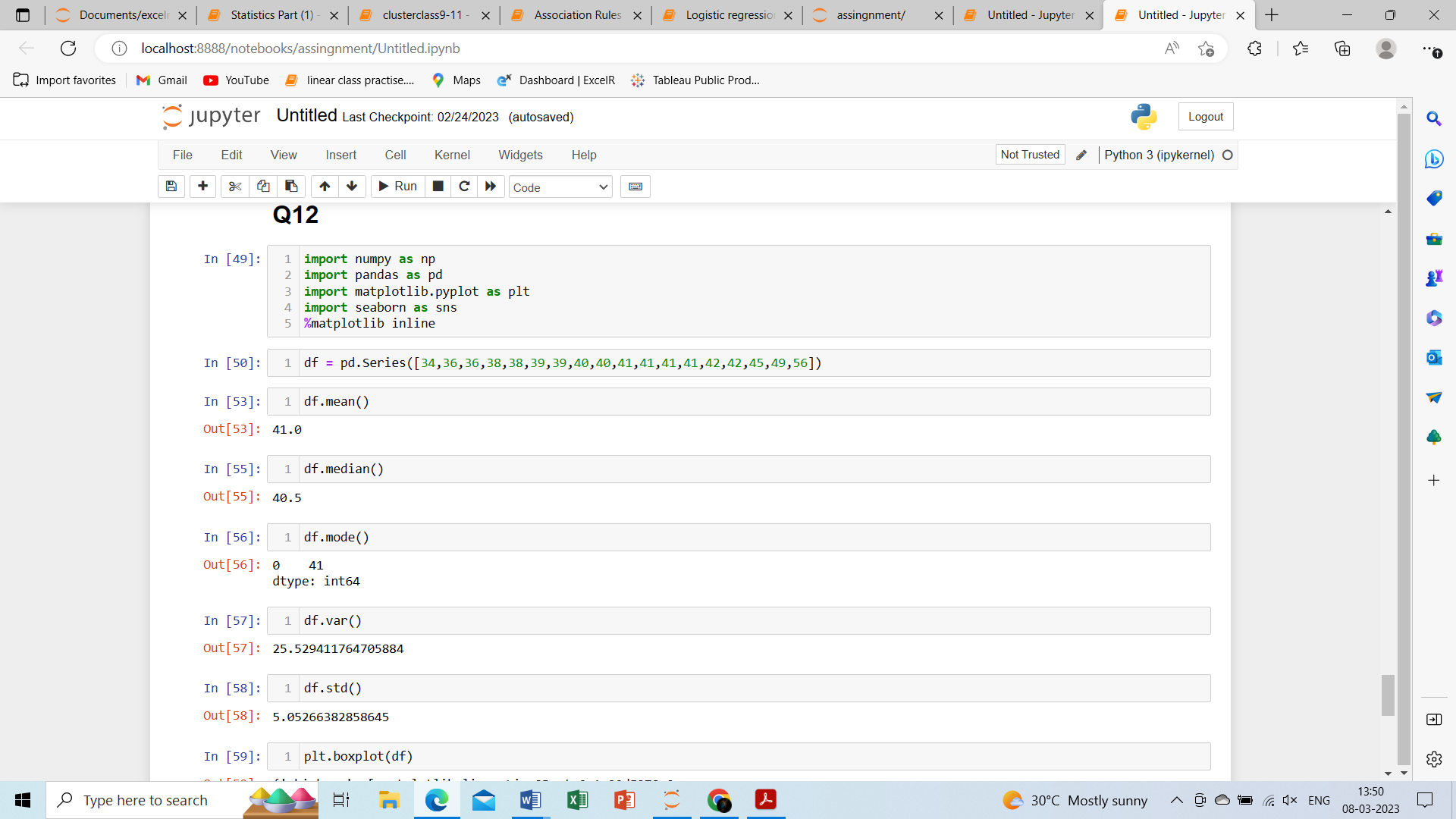
**Q11)** Suppose we want to estimate the average weight of an adult male in Mexico. We draw a random sample of 2,000 men from a population of 3,000,000 men and weigh them. We find that the average person in our sample weighs 200 pounds, and the standard deviation of the sample is 30 pounds. Calculate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



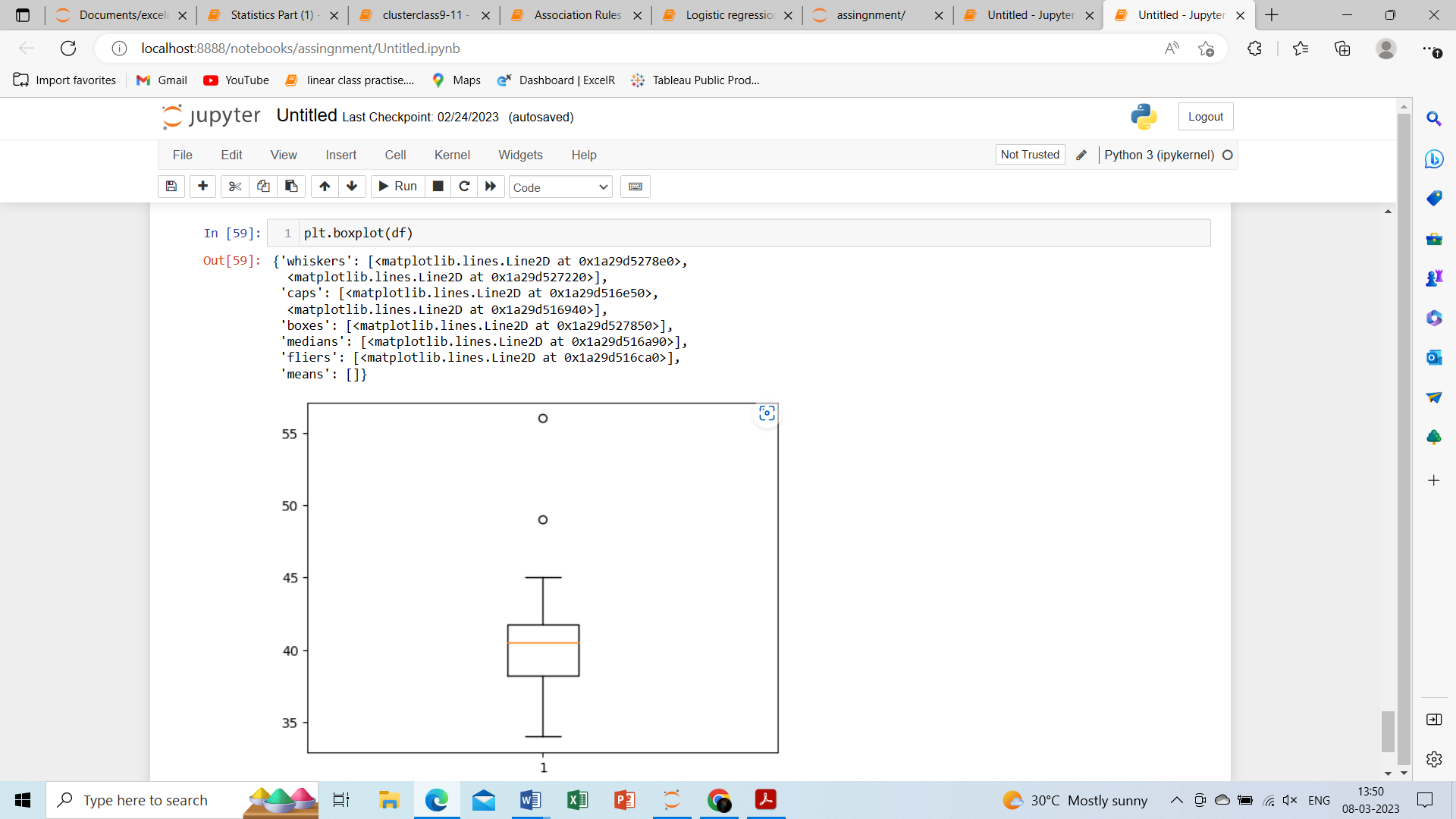
**Q12)** Below are the scores obtained by a student in tests

**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.



1. What can we say about the student marks?



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

Ans: When mean= median, then is no skewness its symmetric.

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

Ans: positive skewness (right skew distribution).

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

Ans: negative skewness (left skew distribution)

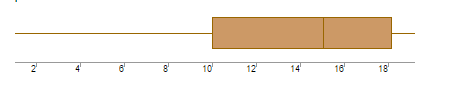
Q16) What does positive kurtosis value indicates for a data?

Ans: most of the number located in tail of the distribution.

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

Ans: there are less number located in tails of the distribution.

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



What can we say about the distribution of the data?

Ans: the boxplot is normally distributed, the median is towards the higher value.

What is nature of skewness of the data?

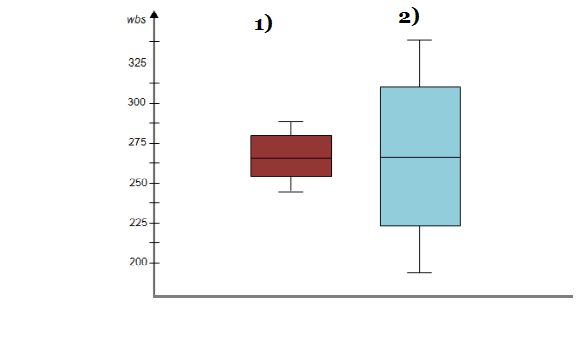
Ans: its left skewed

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

Ans: the IQR of data= upper quartile-lower quartile

=18-10

=8  
  
Q19) Comment on the below Boxplot visualizations?



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

Ans: By observing both the plots Whiskers level is high in boxplot2.

Mean and median are equal hence distribution is symmetric.

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

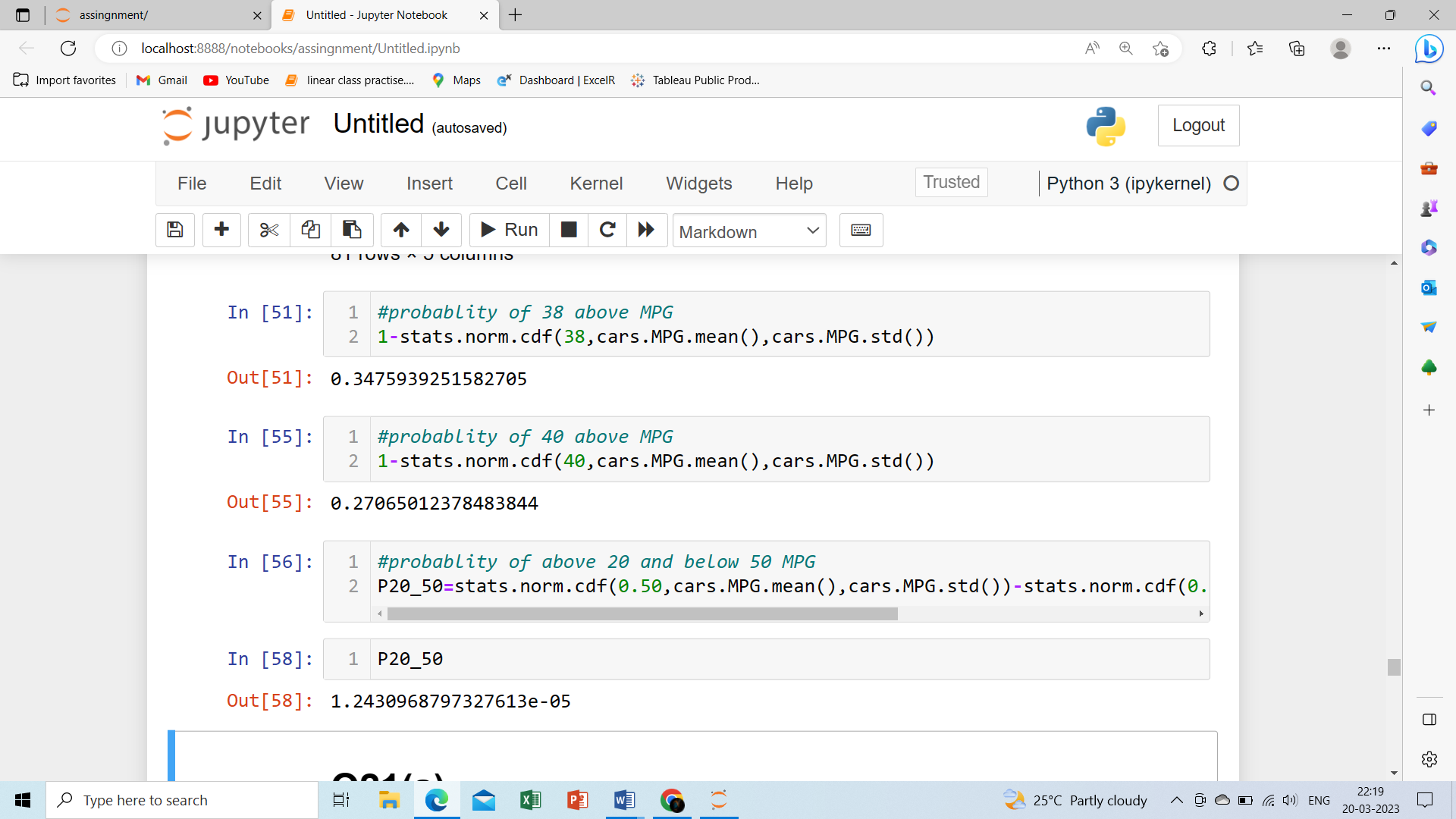
Data \_set: Cars.csv

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)

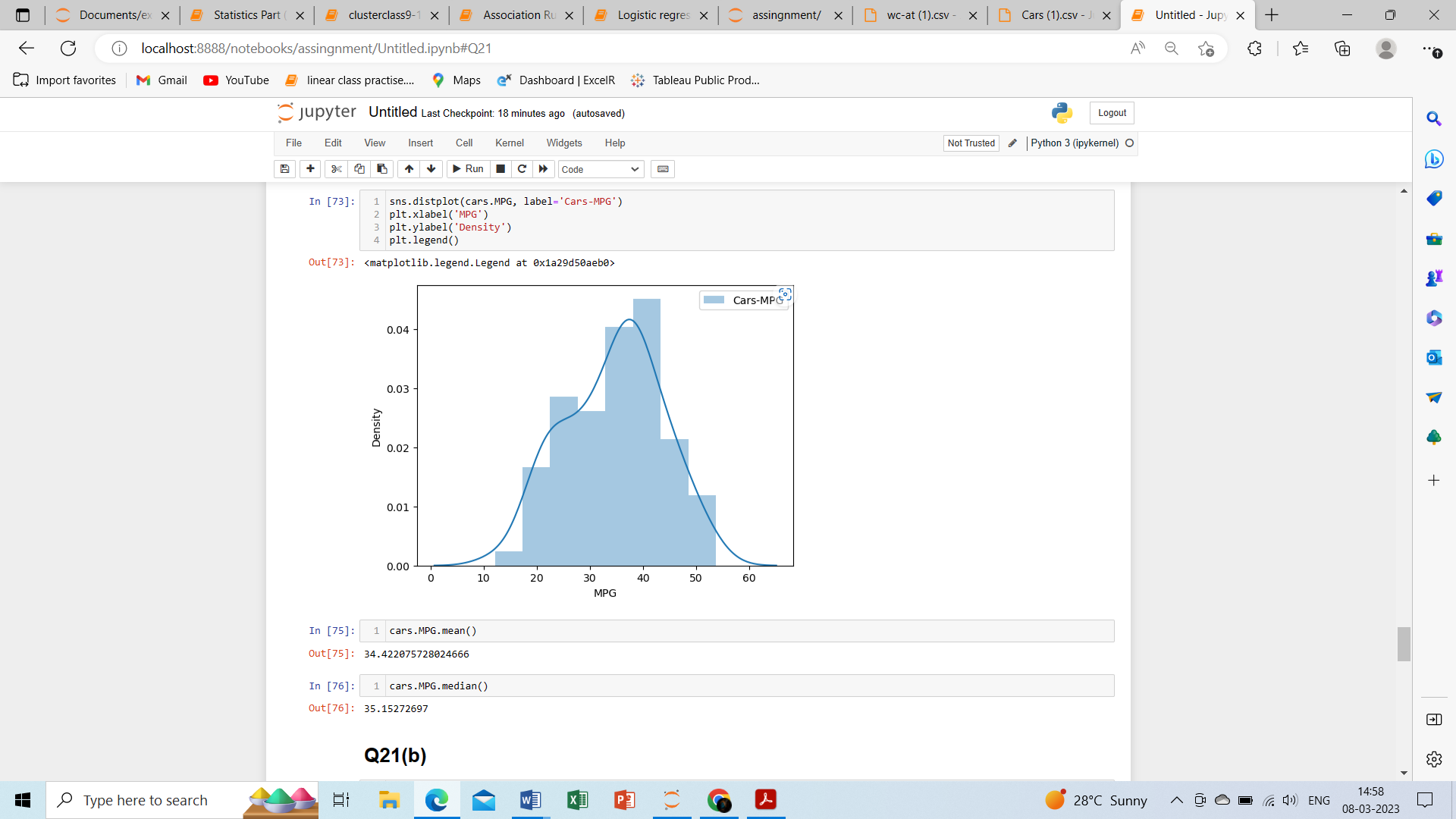
ANS:



Q 21) Check whether the data follows normal distribution

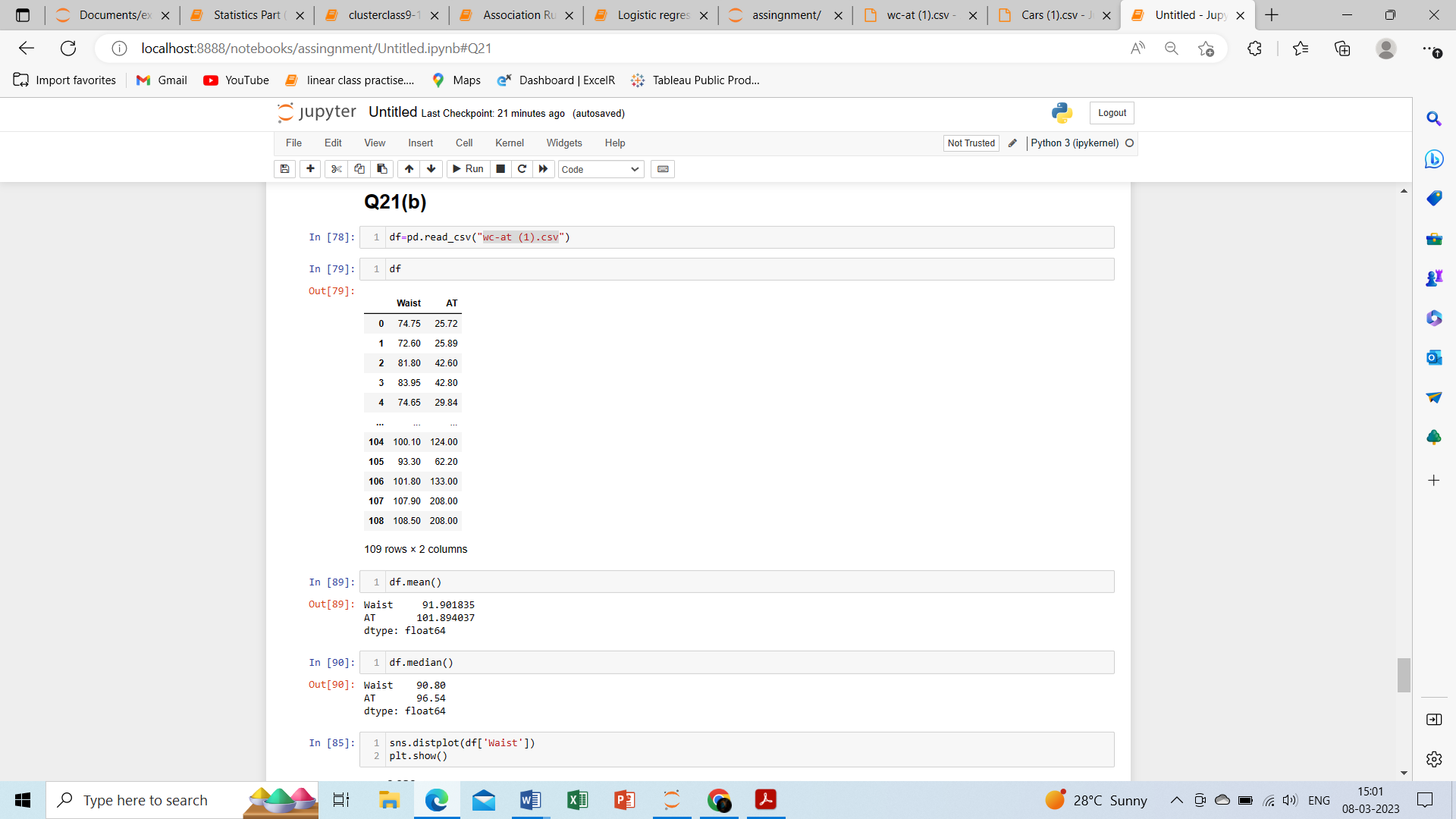
1. Check whether the MPG of Cars follows Normal Distribution

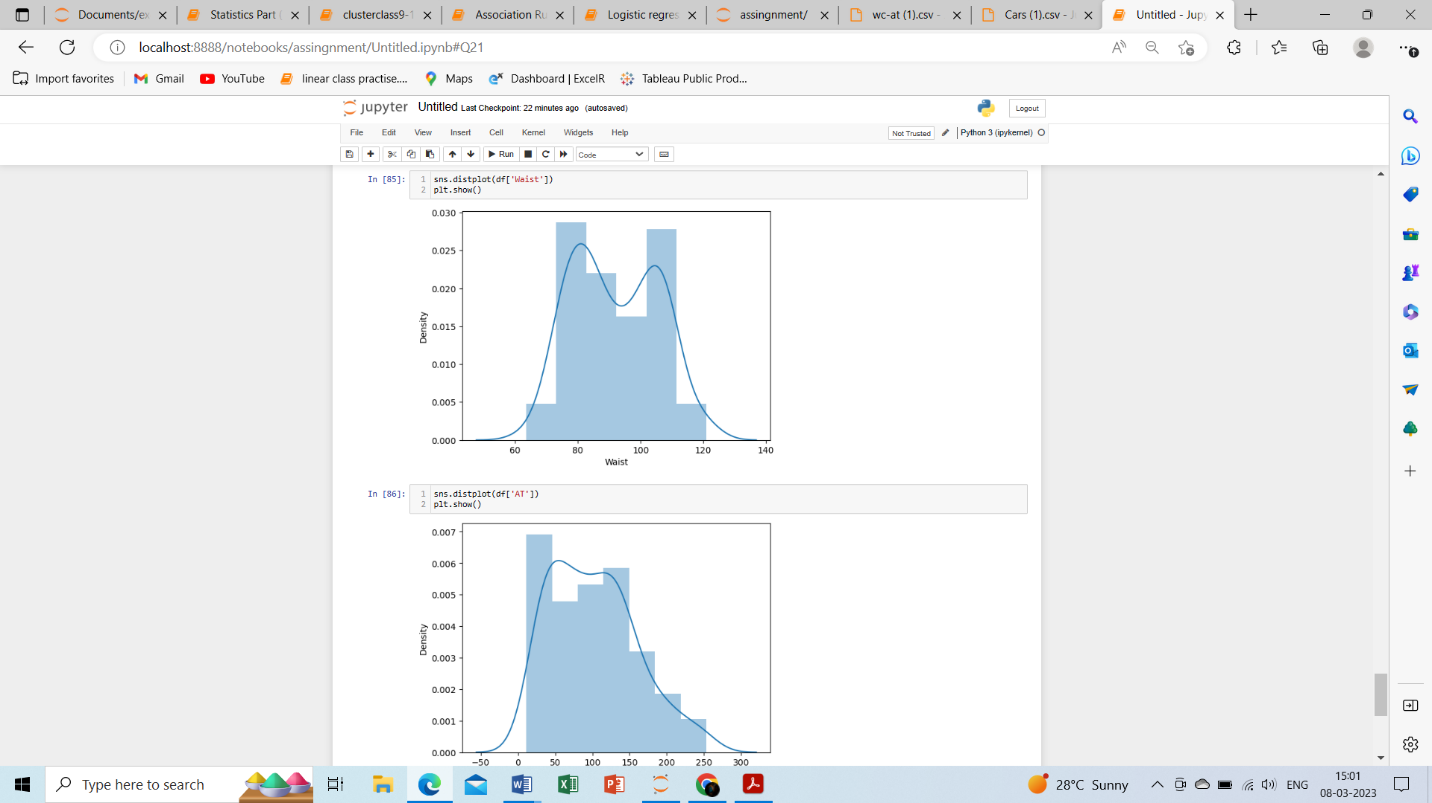
Dataset: Cars.csv

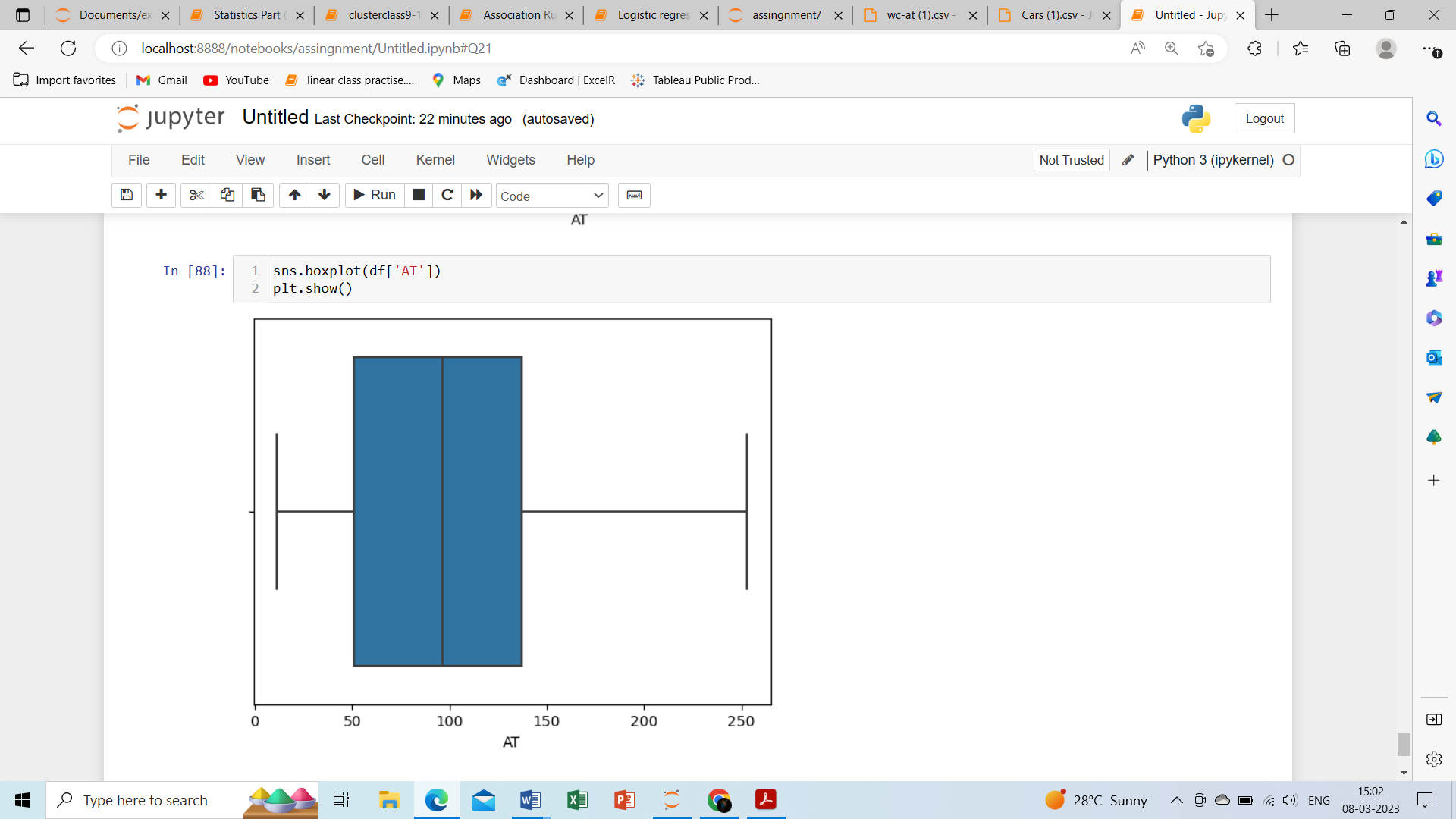


1. Check Whether the Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set follows Normal Distribution

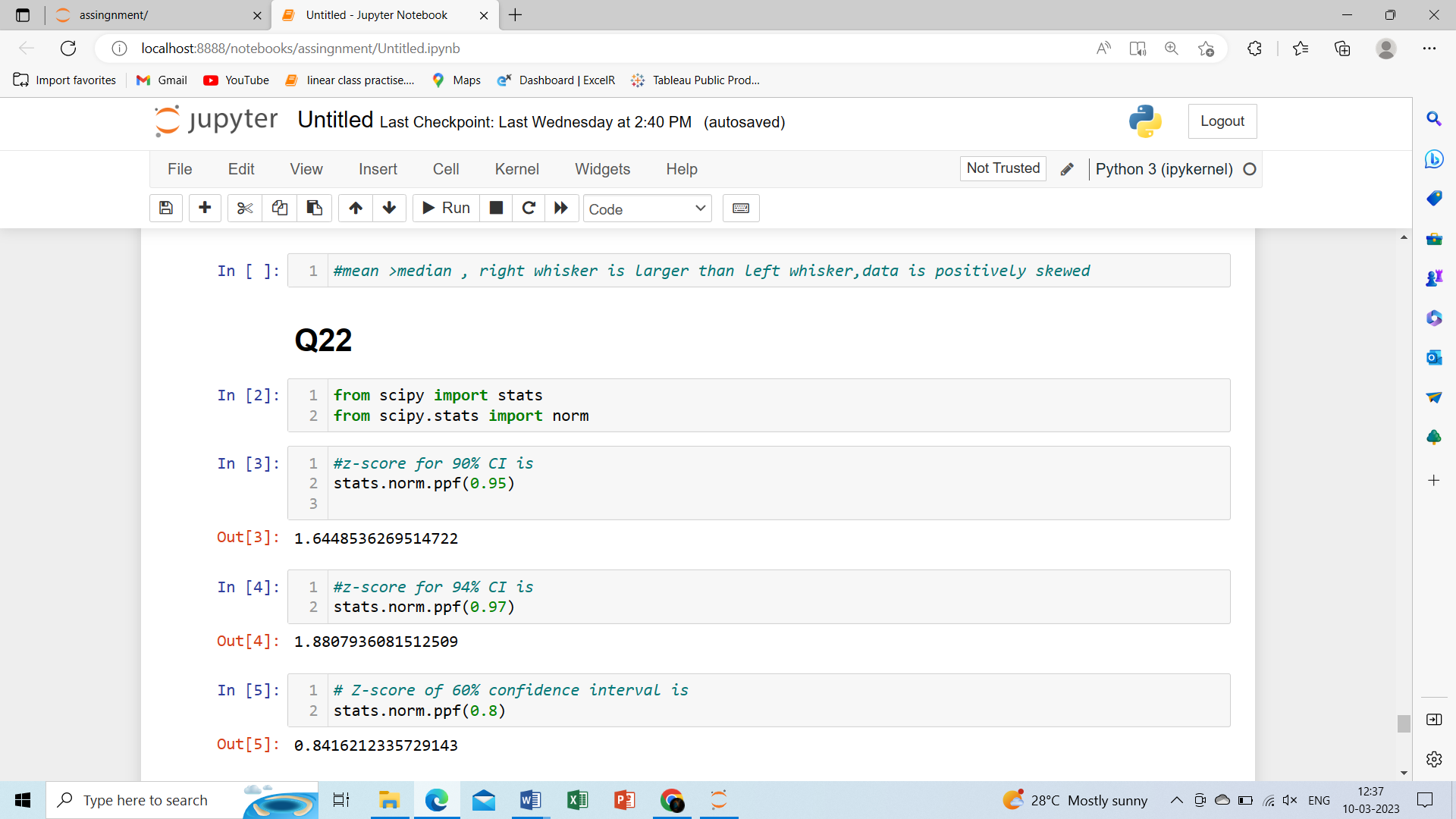
Dataset: wc-at.csv



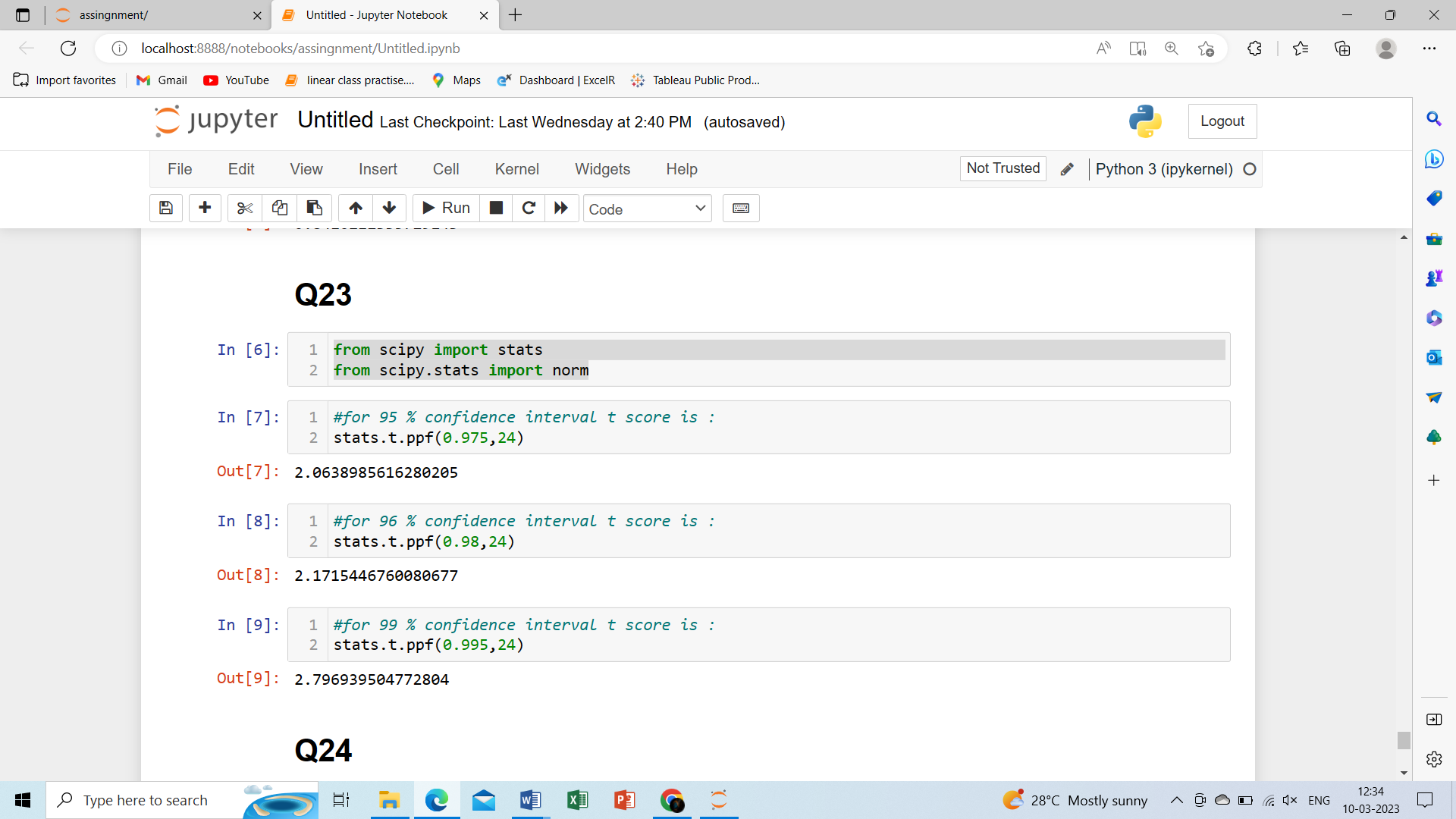




Q 22) Calculate the Z scores of 90% confidence interval,94% confidence interval, 60% confidence interval



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



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

