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

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

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

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

P(s)=8

P(x)=3/8

=0.375

=37.5%

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

🡪

The no of possible event occures

S={[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]}

N=36

I] probability Equal to 1

P(a)=0/36

P(a)=0

II]probability of Less than or equal to 4

P(b)=6/36

P(b)=1/6

P(b)=0.1667

III] Sum is divisible by 2 and 3

P(C)=6/36

P(c)=1/6

P(c)=0.1667

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?

🡪

Total number of balls =7 balls

N =(Event 2 balls are drawn randomly from bag)

N = 7! / 2! \* 5!

= (7\*6\*5\*4\*3\*2\*1) / (2\*1) \* (5\*4\*3\*2\*1)

N = (7\*6)/ (2\*1) = 21

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

N (Event (None of the balls drawn is blue) = 5! / 2! \* 3! = (5\*4) / (2\*1)

= 10

P (None of the balls drawn is blue) = N (Event (None of the balls drawn is blue) /

N (Event (2 balls are drawn randomly from

bag)

= 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

🡪

Expected no of candies=(1\*0.015)+(4\*0.20)+(3\*0.65)+(5\*0.005)+(6\*0.01)+(2\*0.120)

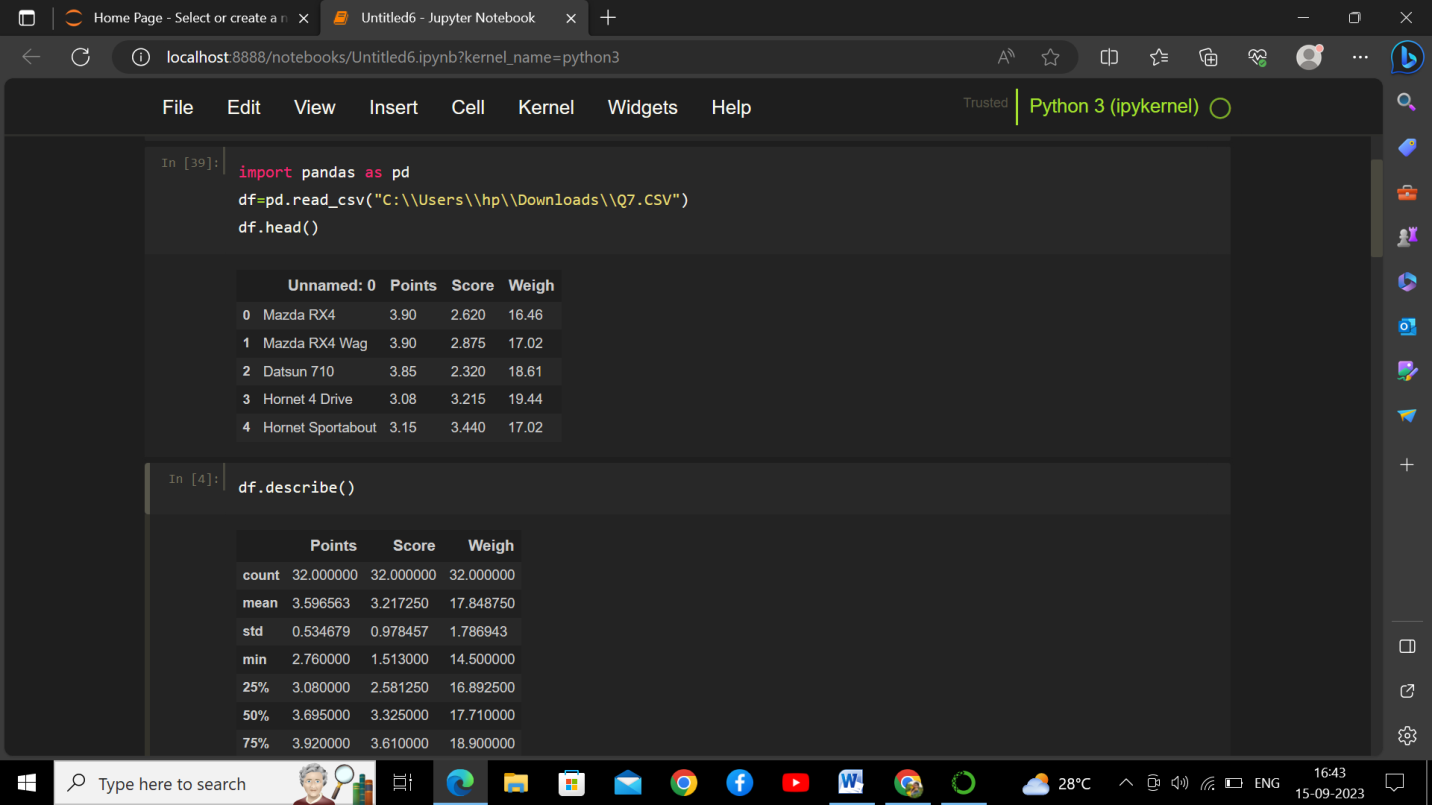
Expected no of candies=0.015+0.8+1.96+0.025+0.06+0.24

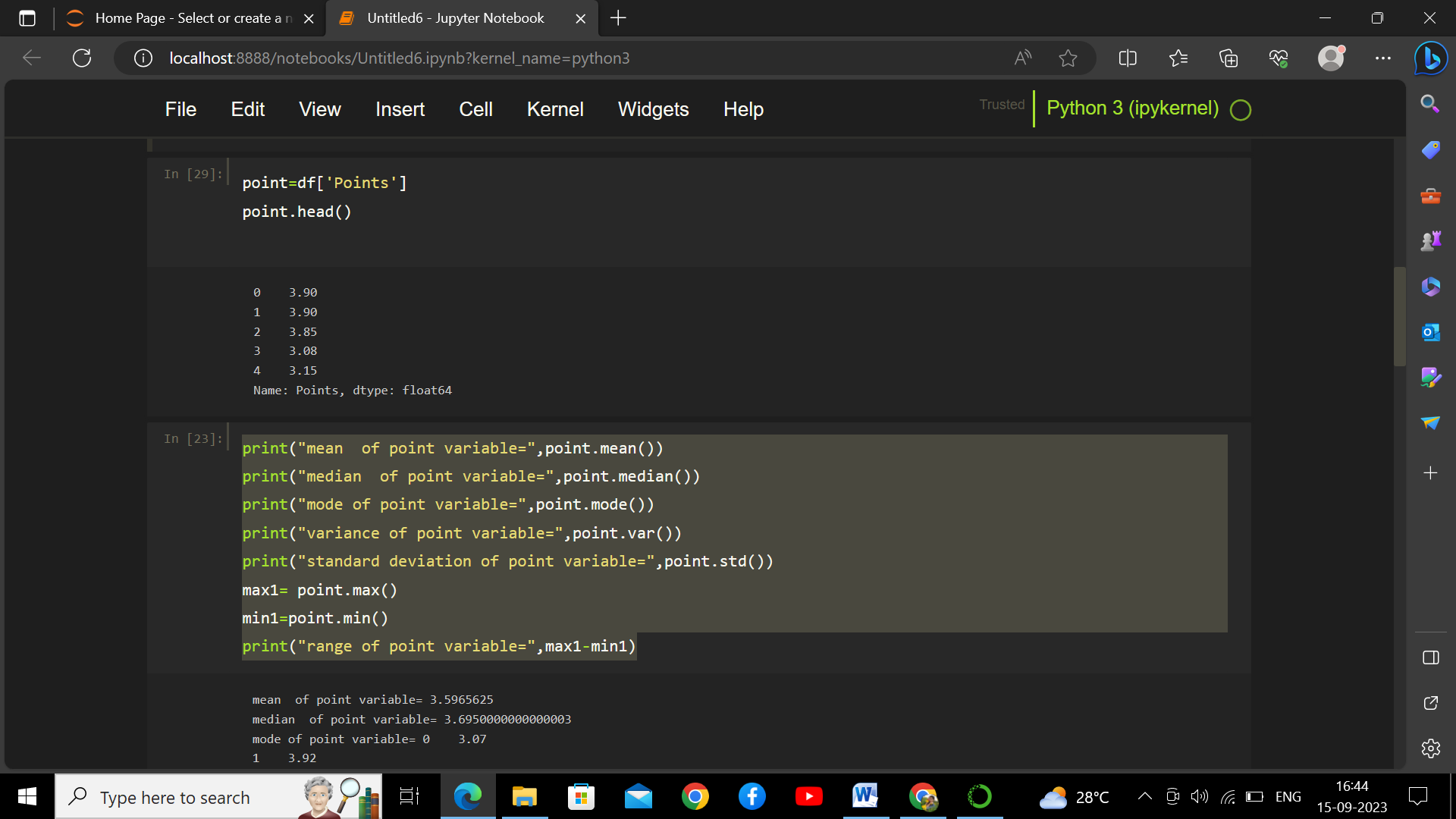
=3.10

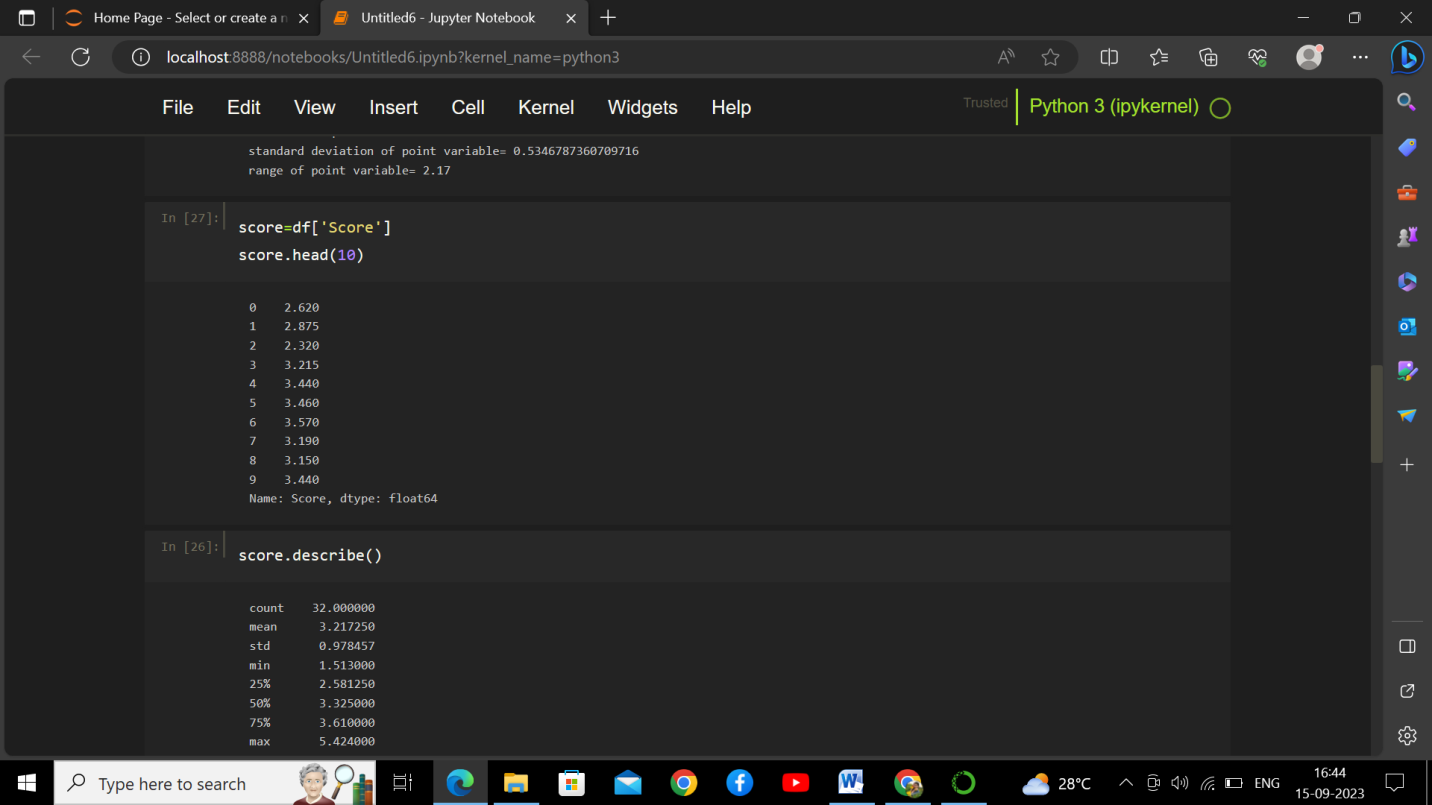
Q7) Calculate Mean, Median, Mode, Variance, Standard Deviation, Range & comment about the values / draw inferences, for the given dataset

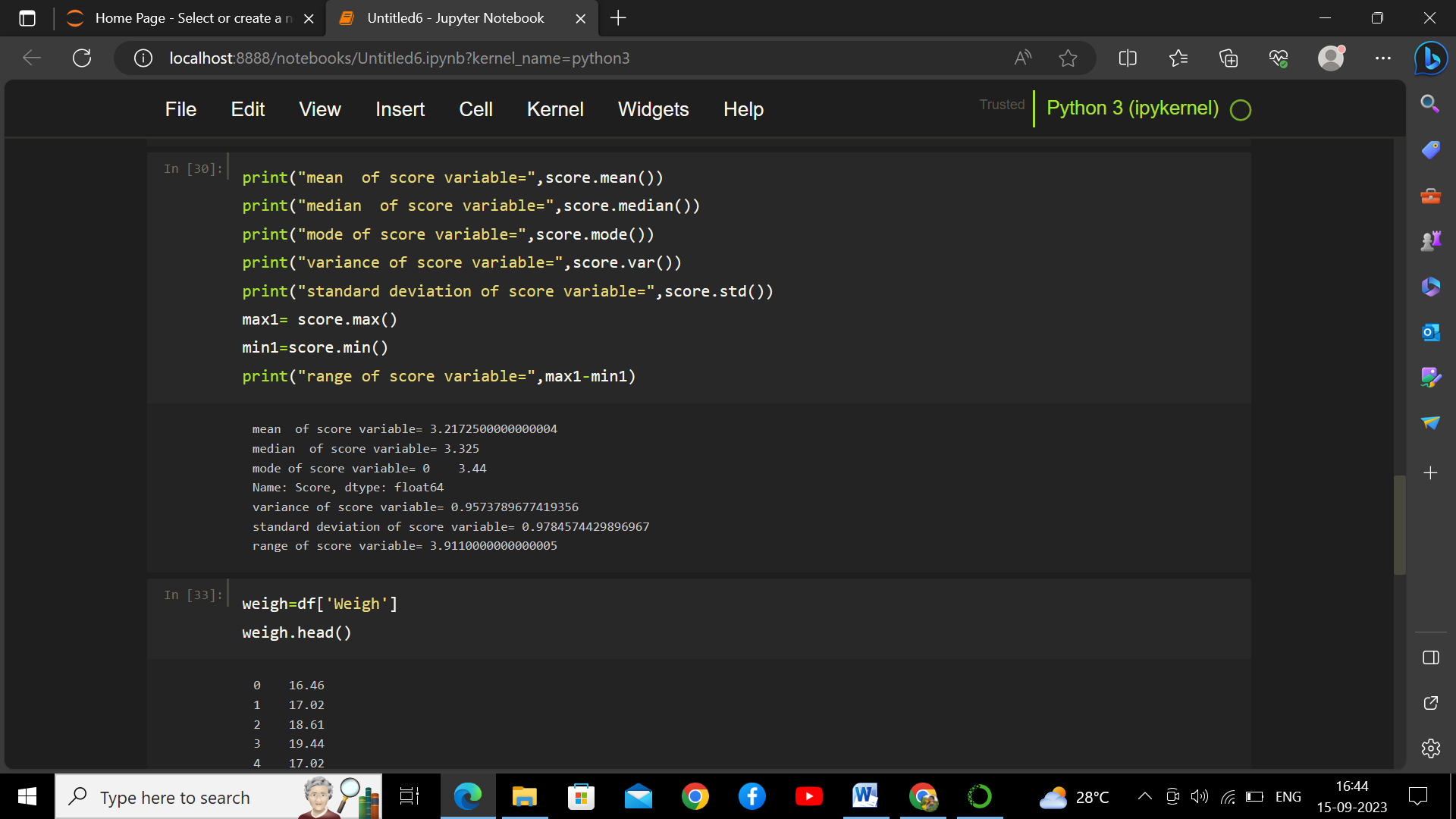
* For Points,Score,Weigh>

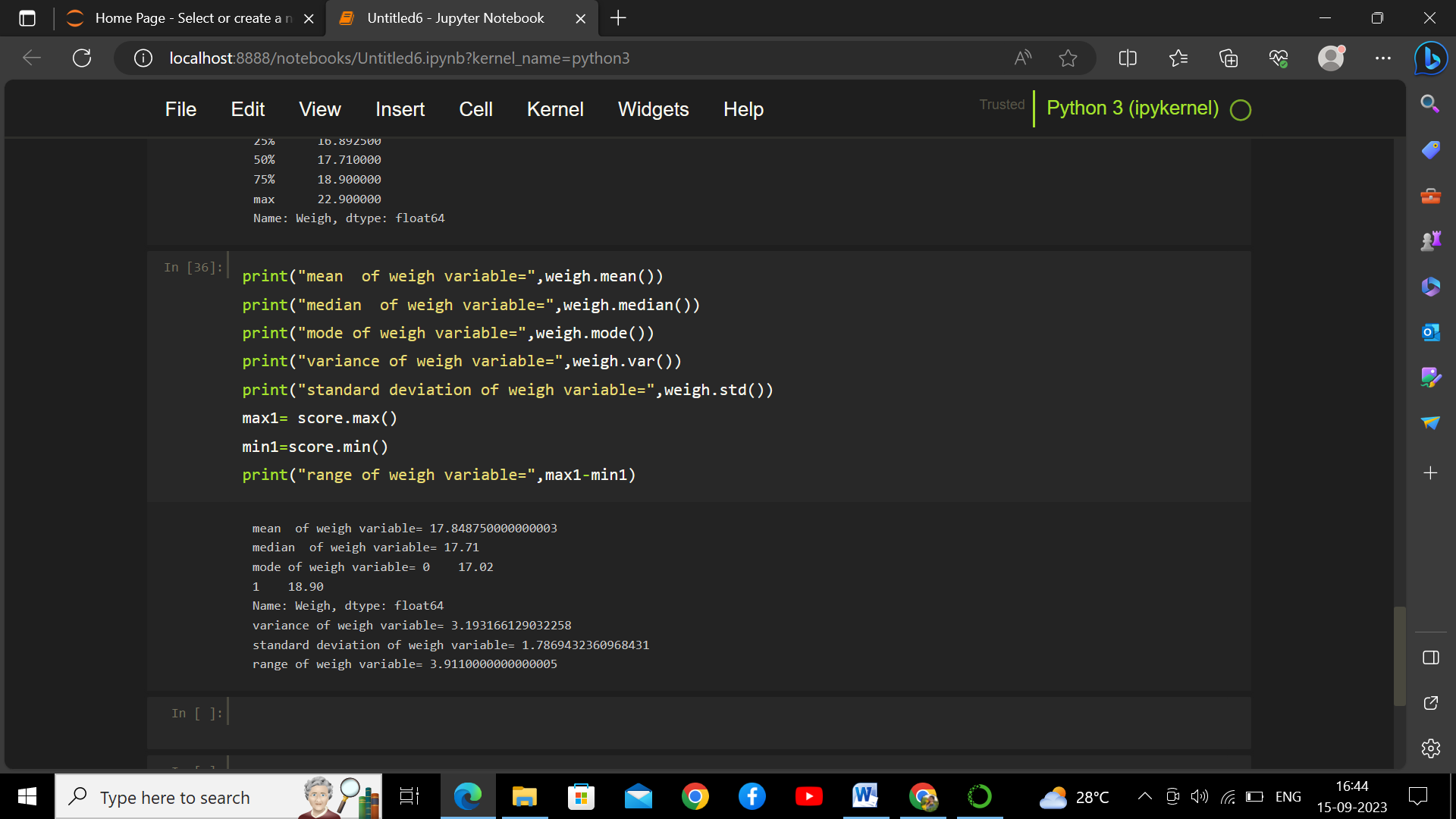
Find Mean, Median, Mode, Variance, Standard Deviation, and Range and also Comment about the values/ Draw some inferences.

🡪****

****

****

****

****

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?

🡪value = Sum (X \* Probability of X)

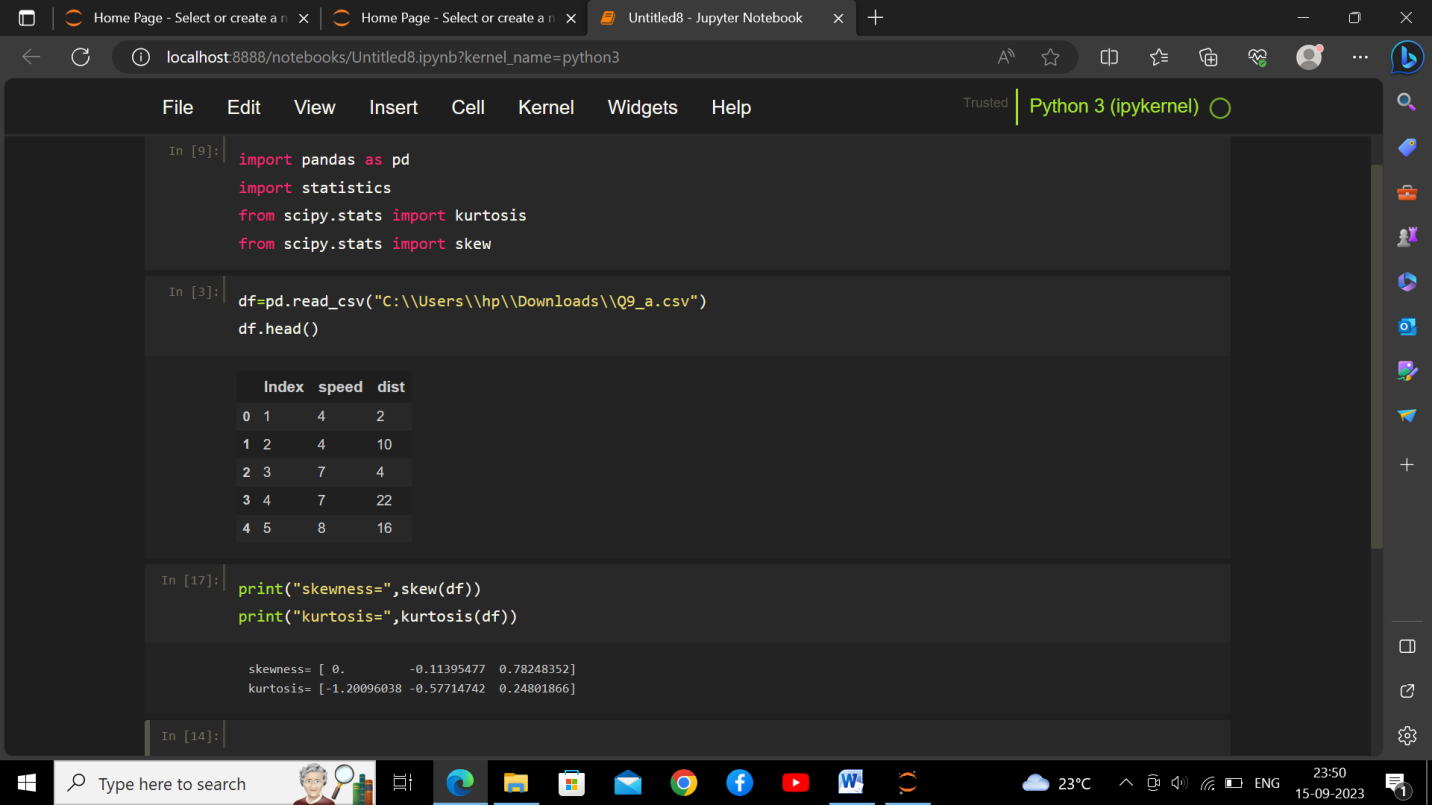
= (1/9)(108)+ (1/9)(110)+ (1/9)(123)+ (1/9)(134)+ (1/9)(145)+ (1/9)(167)+ (1/9)(187)+ (1/9)(199)

The expected weight of patient=145.33

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

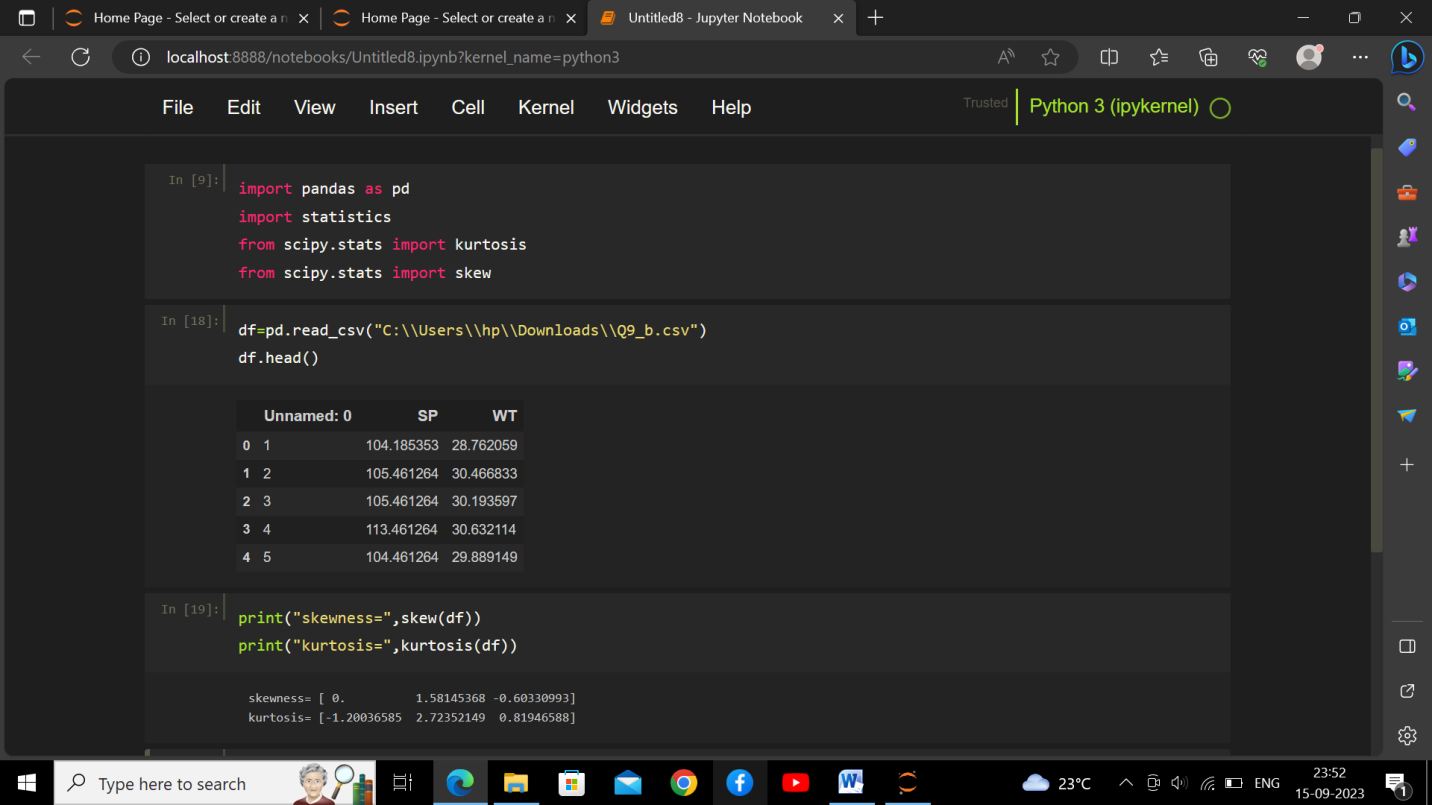
**Cars speed and distance**

**Use Q9\_a.csv**

****

**SP and Weight(WT)**

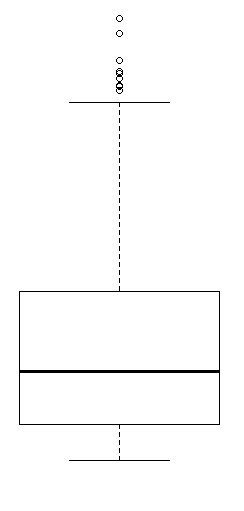
**Use Q9\_b.csv**

****

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



🡪 The histograms peak has right skew and tail is on right. Mean > Median. We have outliers on the higher side.



🡪 The boxplot has outliers on the maximum side.

**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?

🡪conf\_94 =stats.t.interval(alpha = 0.94, df=1999, loc=200, scale=30/np.sqrt(2000))

print(np.round(conf\_94,0))

print(conf\_94)

For 94% confidence interval Range is [ 198.73 – 201.26]

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

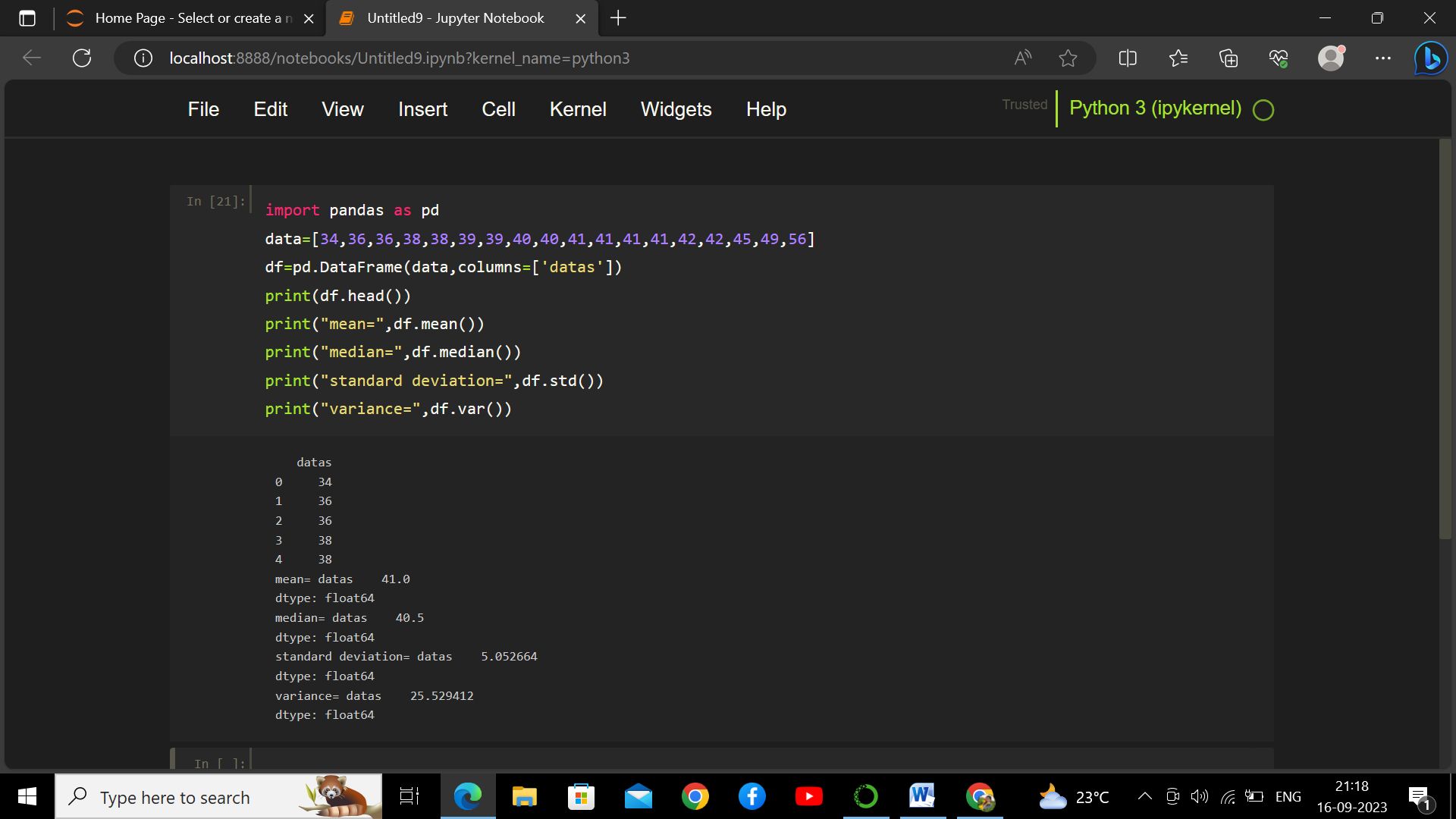
For 96% confidence interval range is [198.62 – 201.37]

**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.

🡪



1. What can we say about the student marks?

🡪

* The student data is slightly skewed towards the right side because the Mean is slightly greater than median.
* The data does not contain outlier.

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

🡪If mean and median are equal then no skewed or zero skewed.

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

🡪If mean is greater than median then the data is right skewed.

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

🡪 If mean is less than median then the data is left skewed.

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

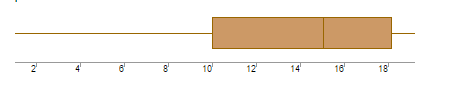
🡪positive kurtosis indicates the distribution is picked and having thick tails.

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

🡪 negative kurtosis indicates the distribution is flatter or less than normal and having thin tails.

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

1]What can we say about the distribution of the data?



🡪The above box plot is not normally distributed, the median is towards the higher point.

2]What is nature of skewness of the data?

🡪The data is skewed towards left side.

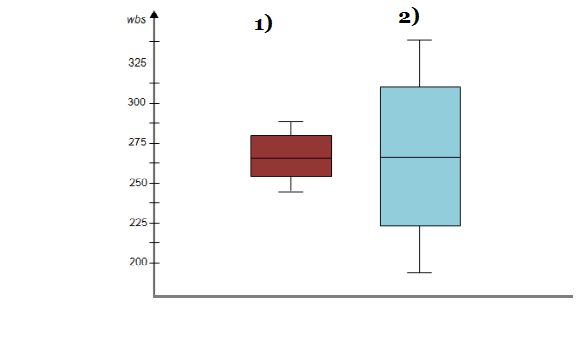
3]What will be the IQR of the data (approximately)?

🡪Inter quartile range(Q3) =upper limit(Q1) - lower limit(Q2)

=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.

🡪There are no outlier present in both boxplot1 and boxplot2.

Both boxplot share the same median which lies between 275 to 250.

They are normally distributed with zero to no skewness.

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

Data \_set: Cars.csv

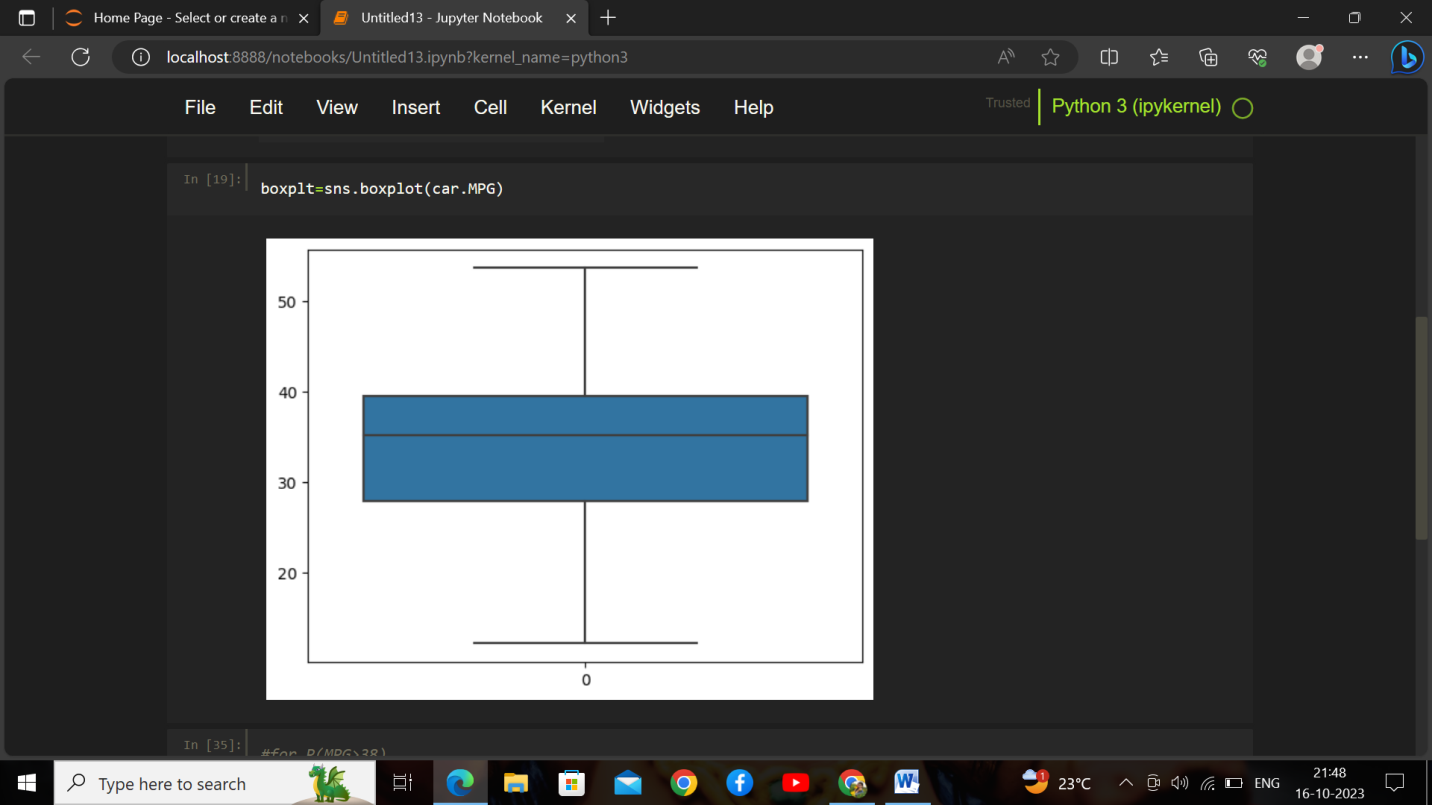
Calculate the probability of MPG of Cars for the below cases.

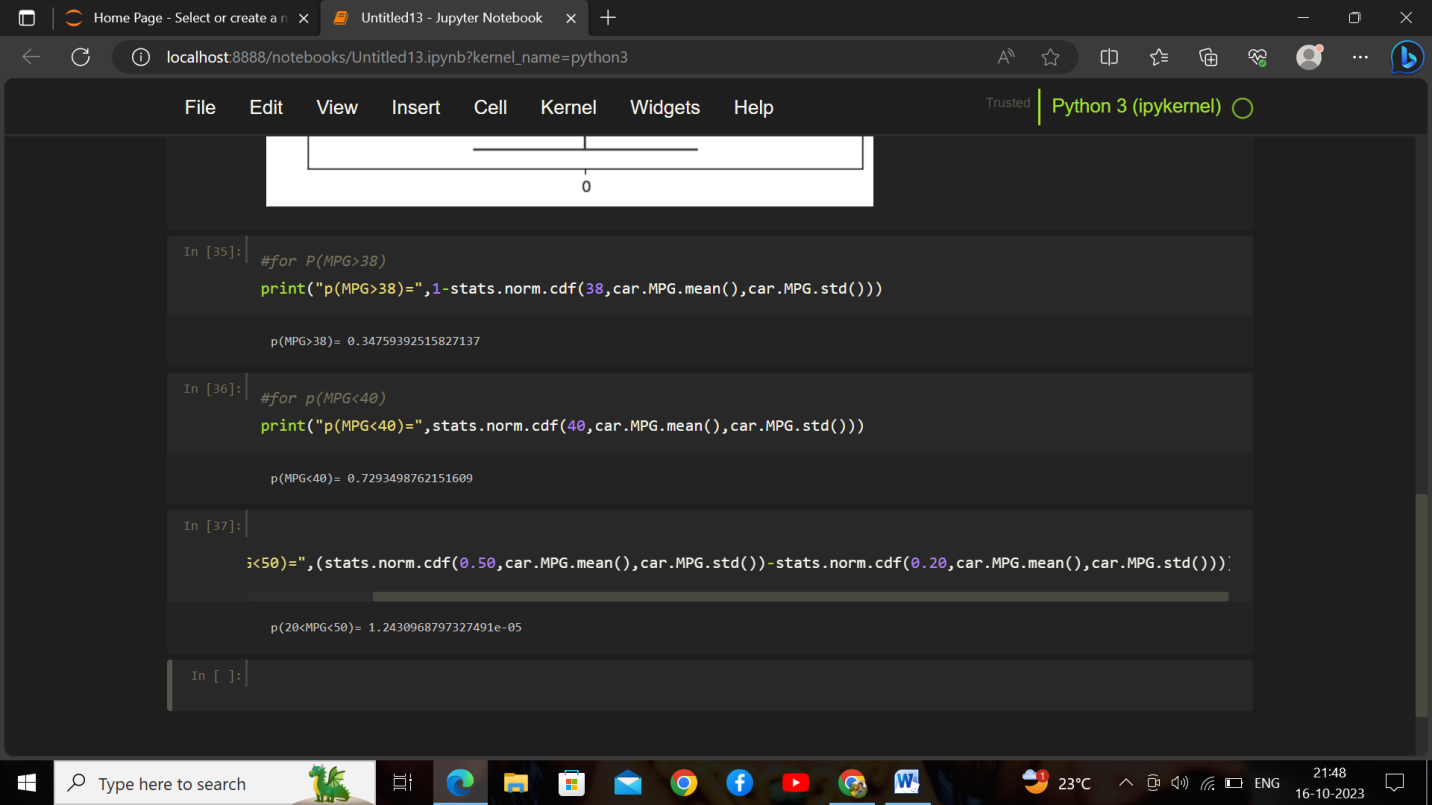
MPG <- Cars$MPG

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

c. P (20<MPG<50)



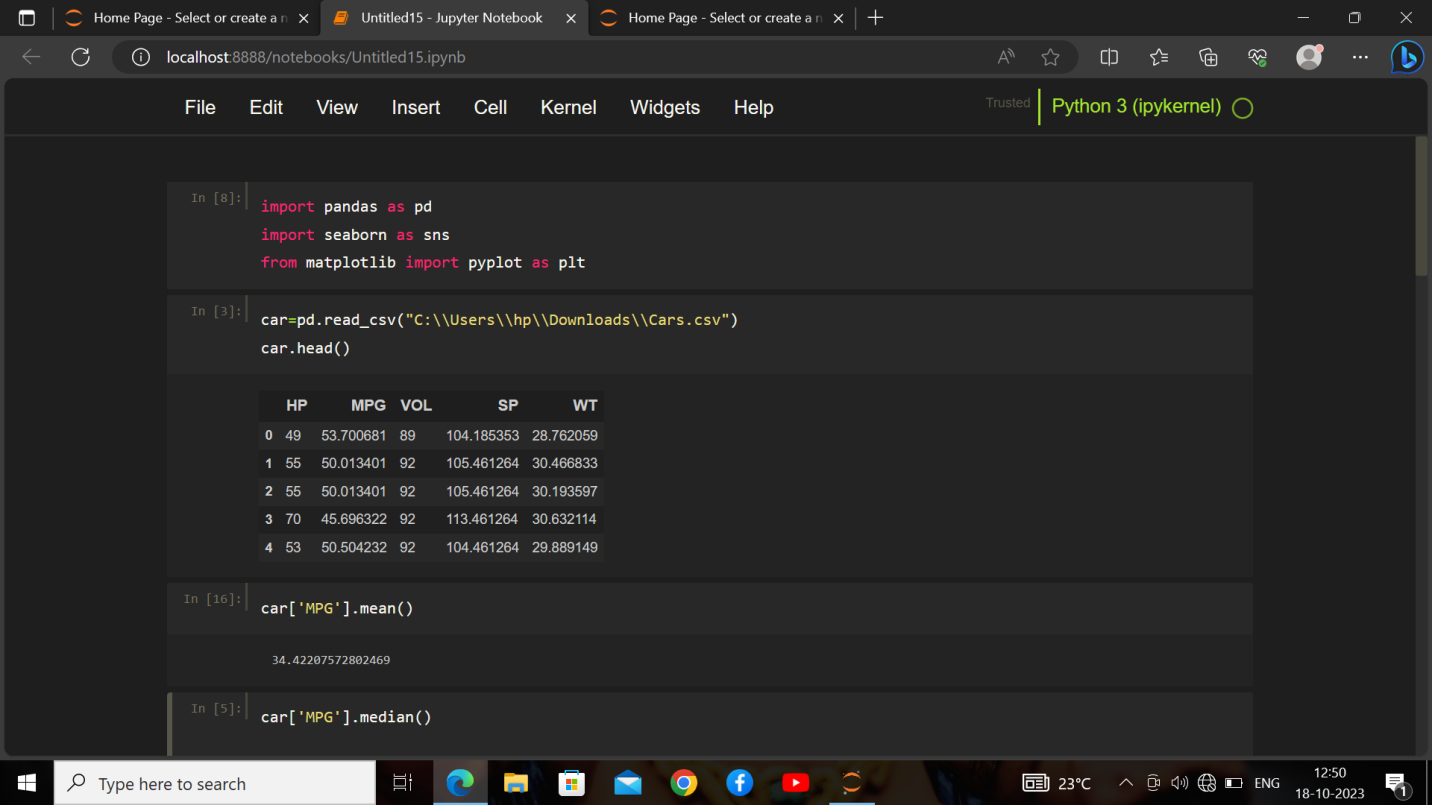


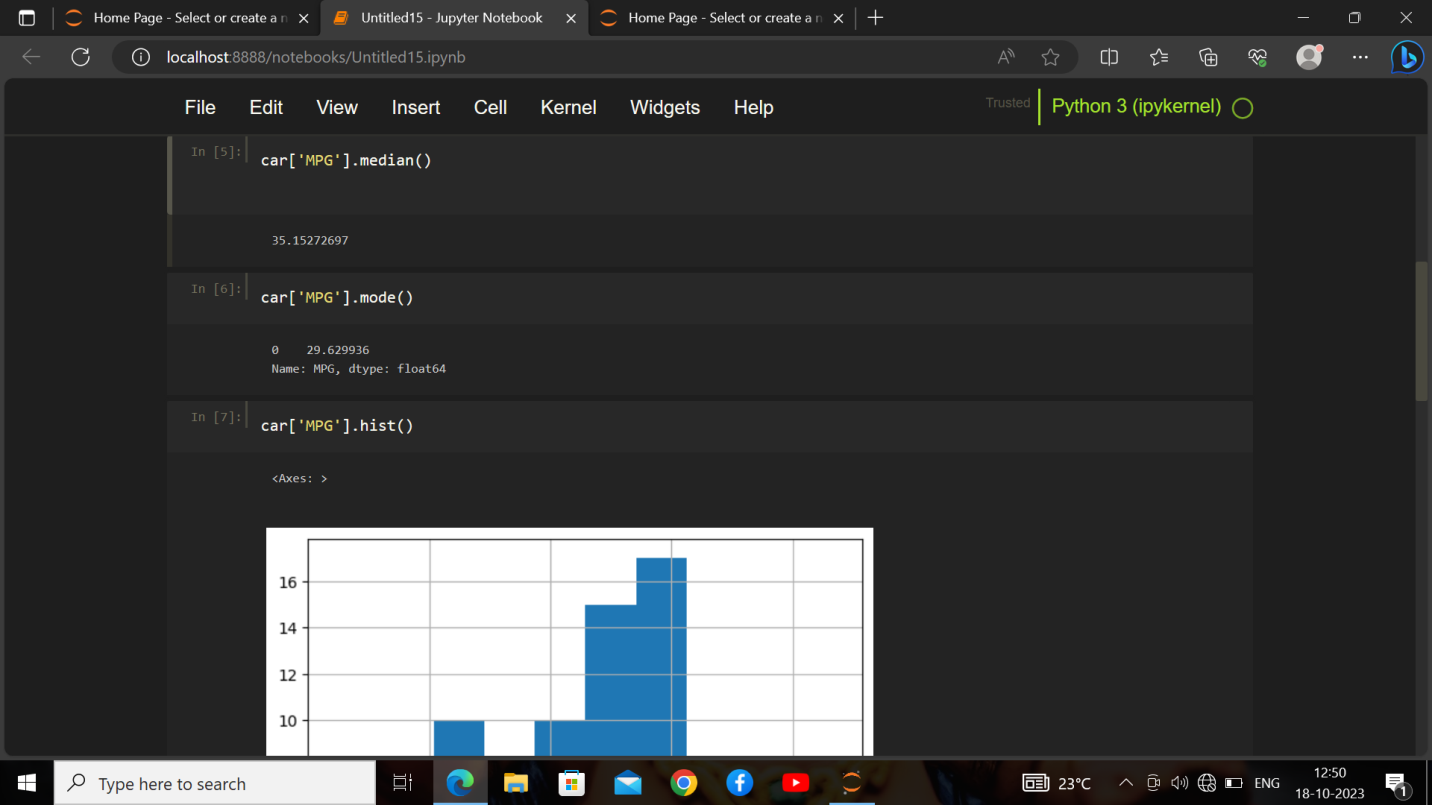


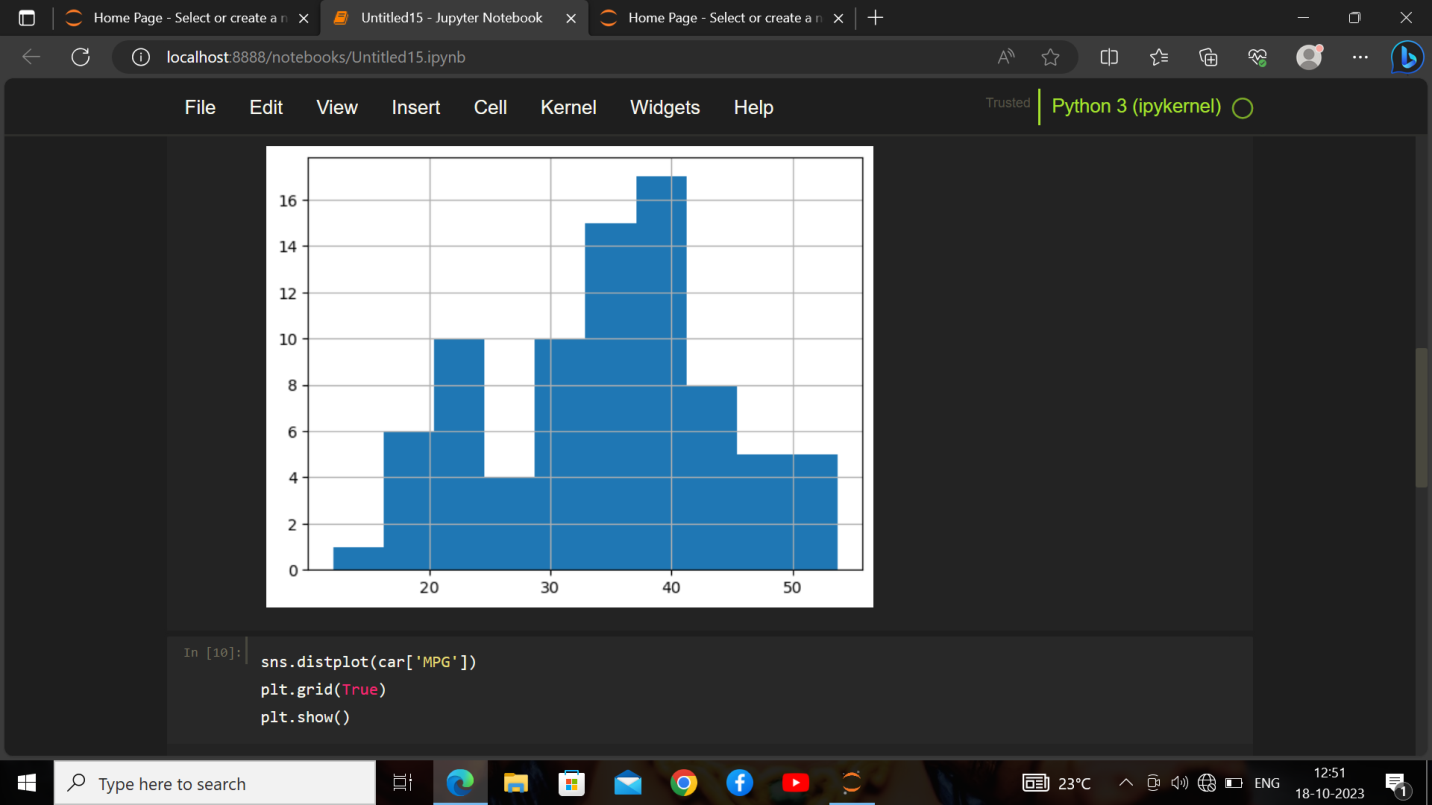
Q 21) Check whether the data follows normal distribution

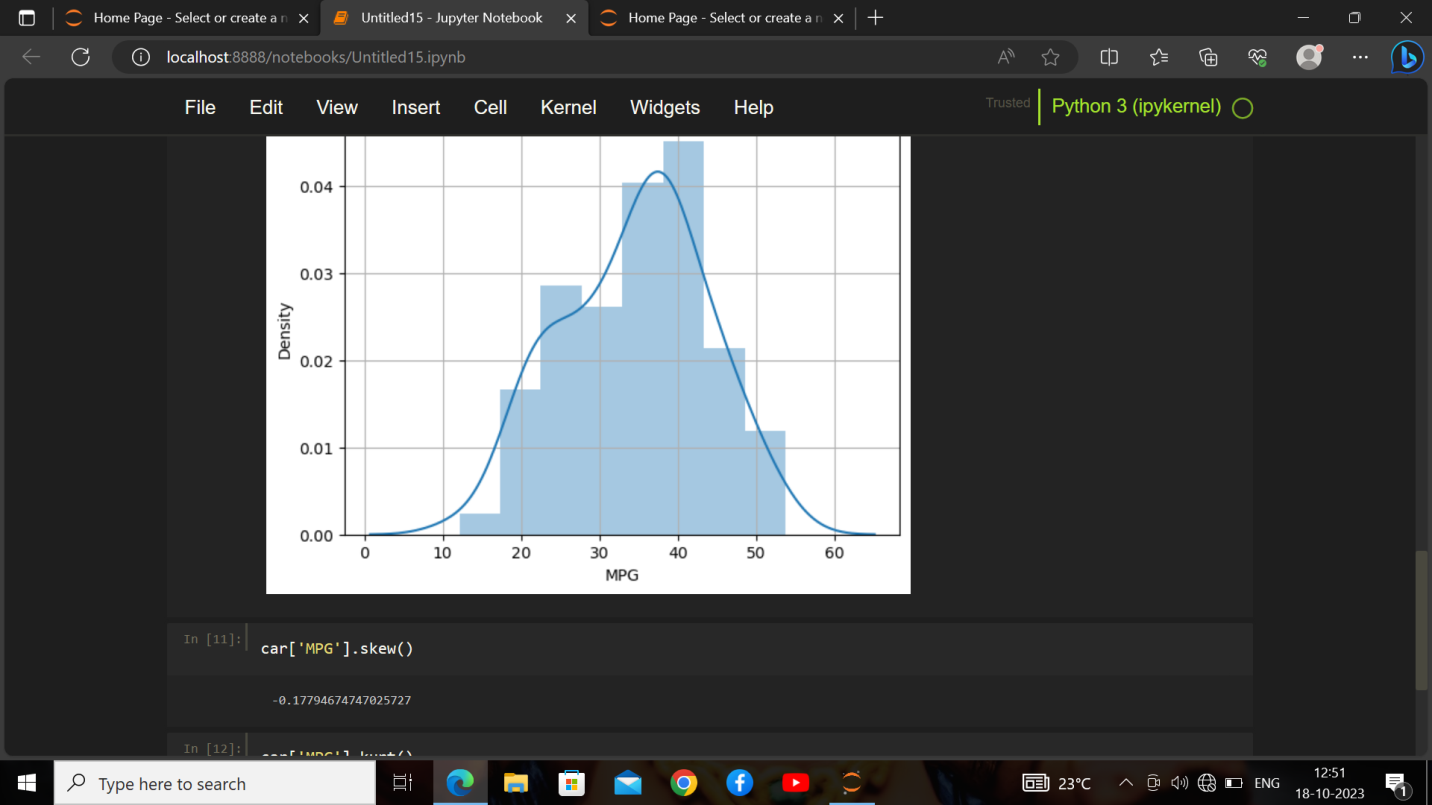
1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv



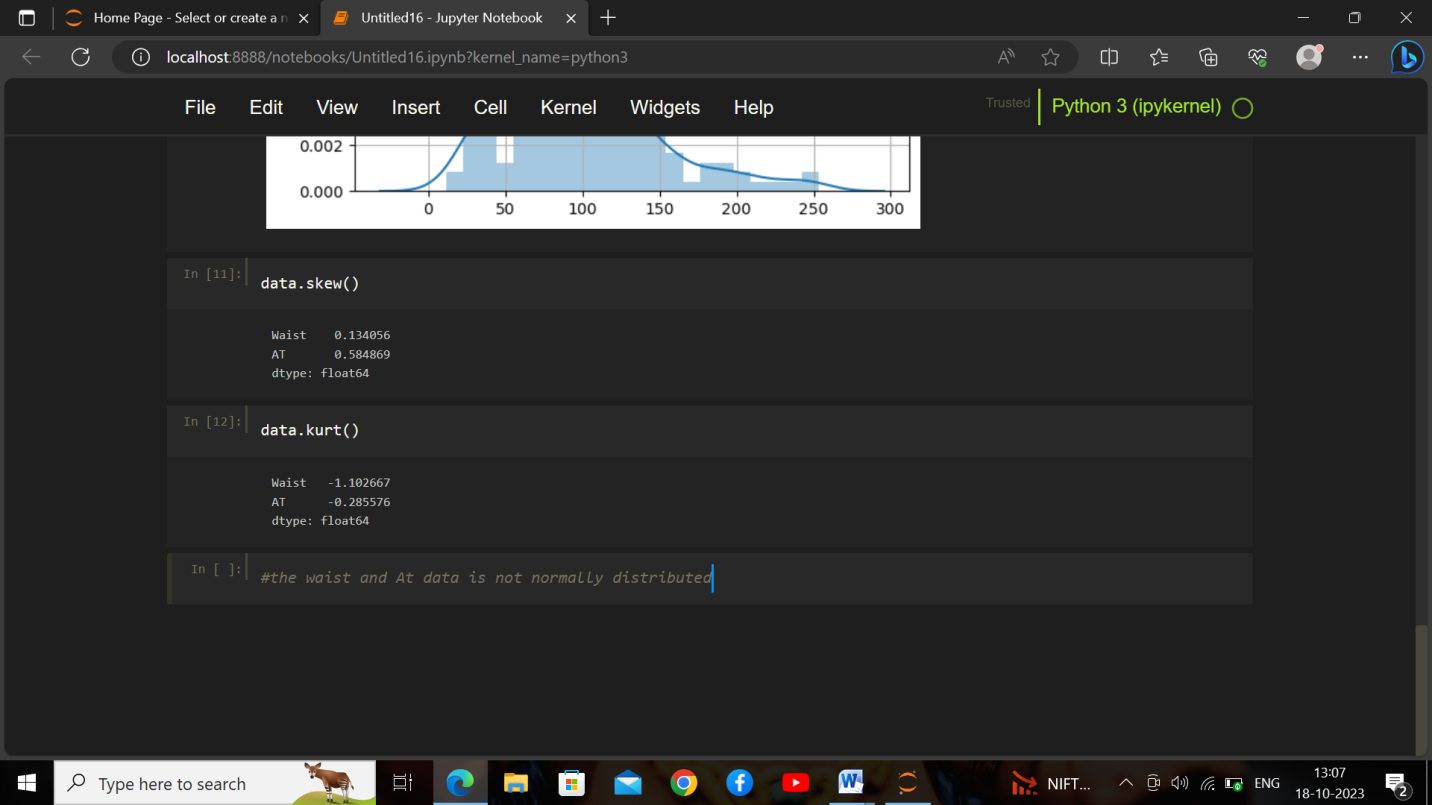
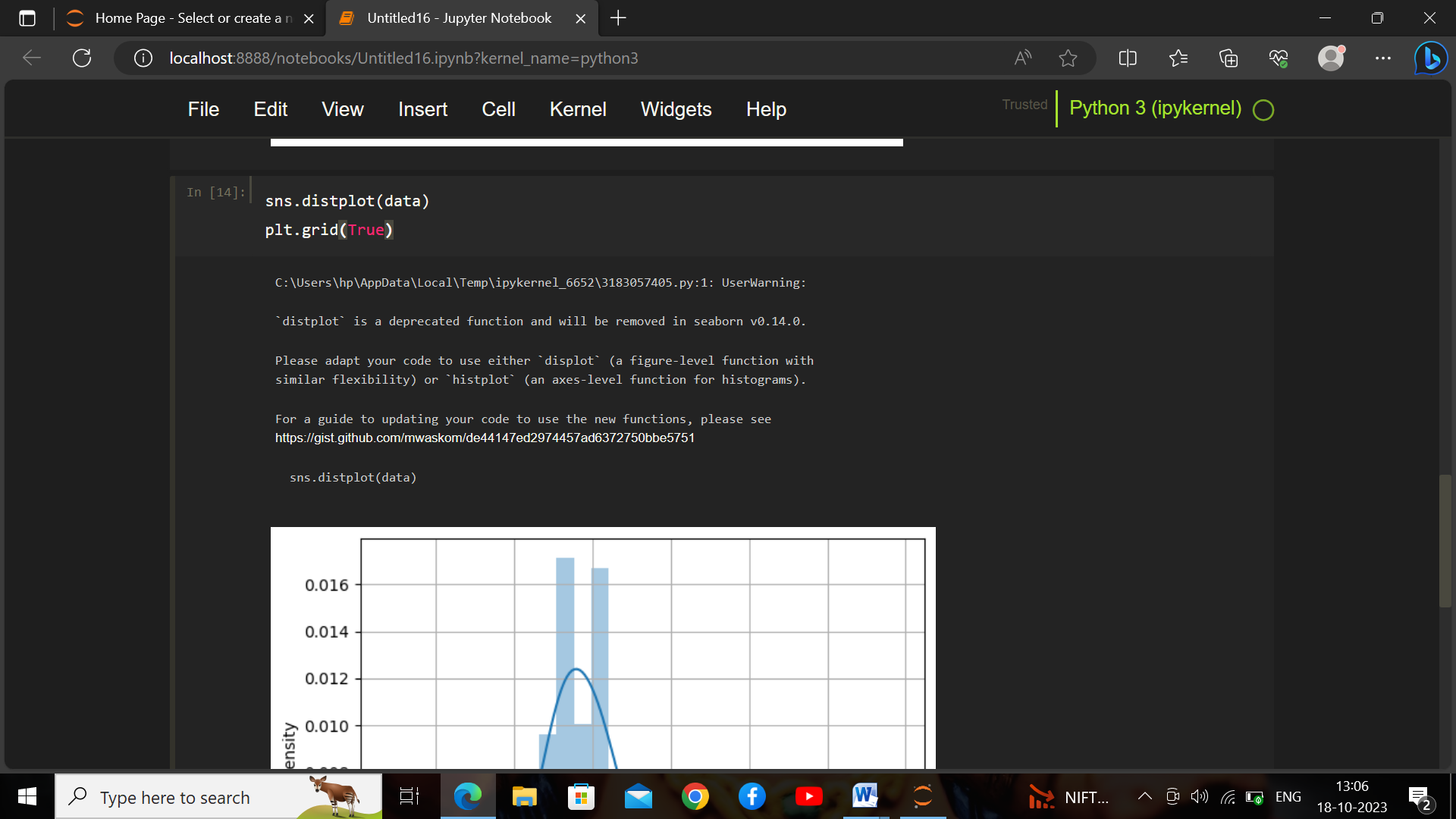
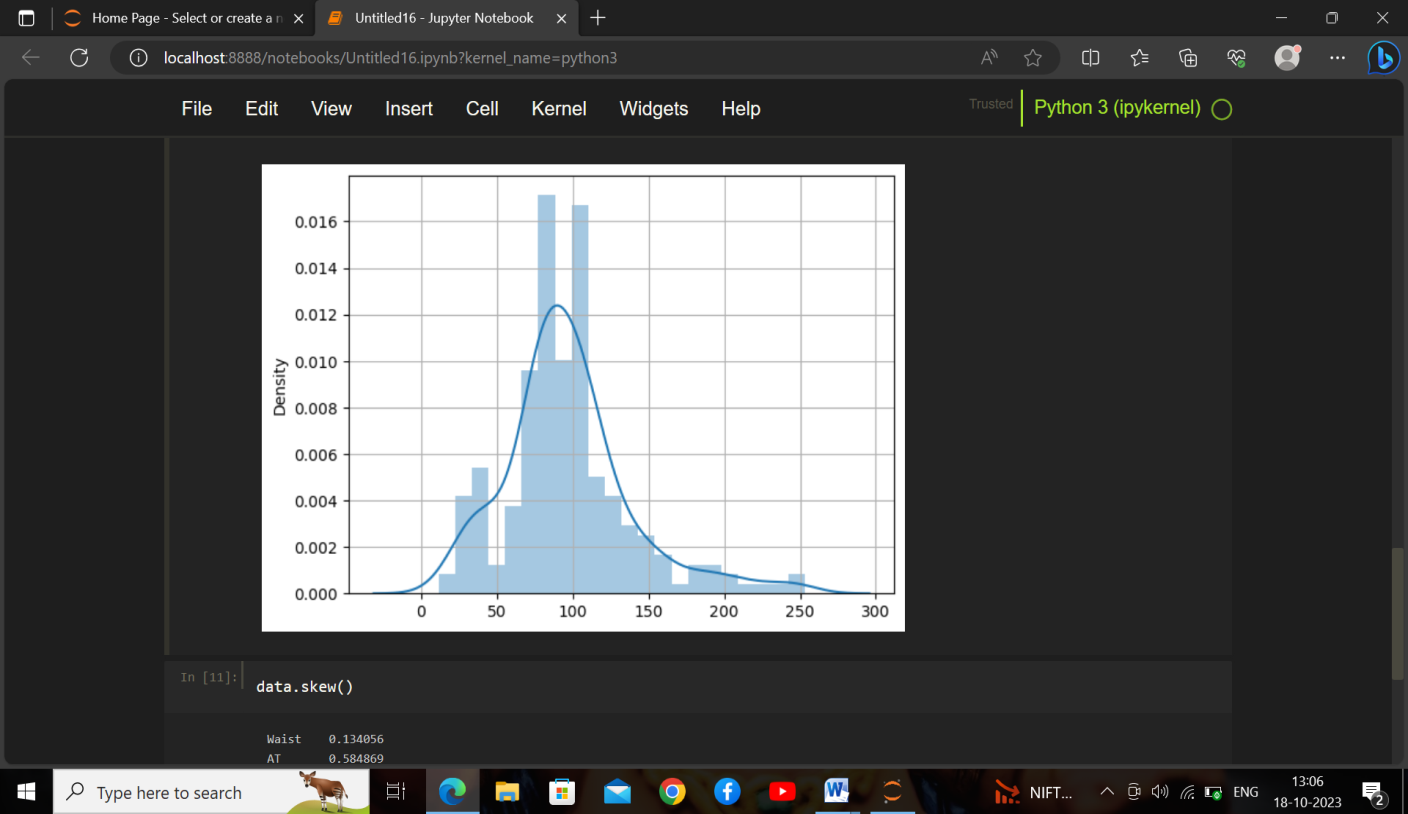
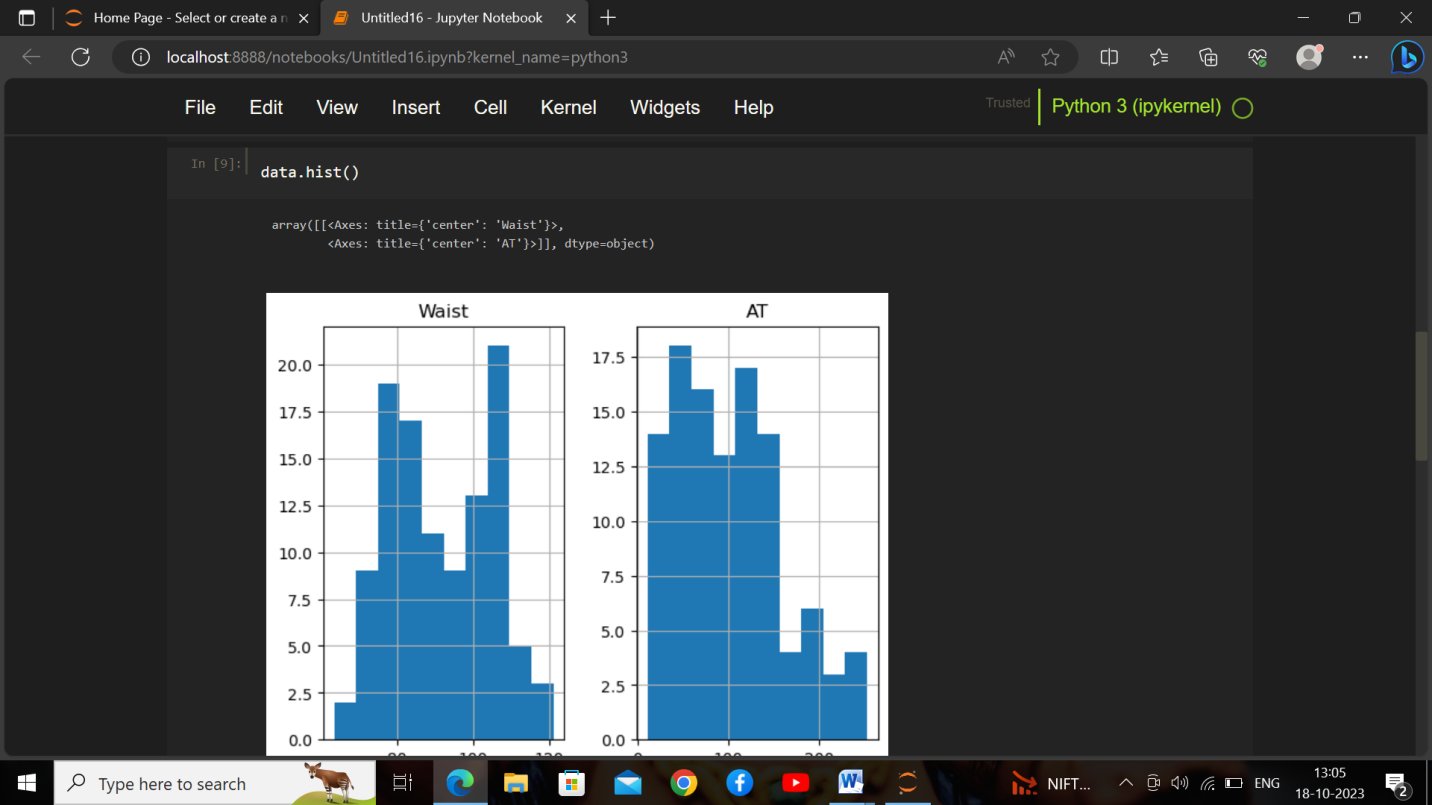
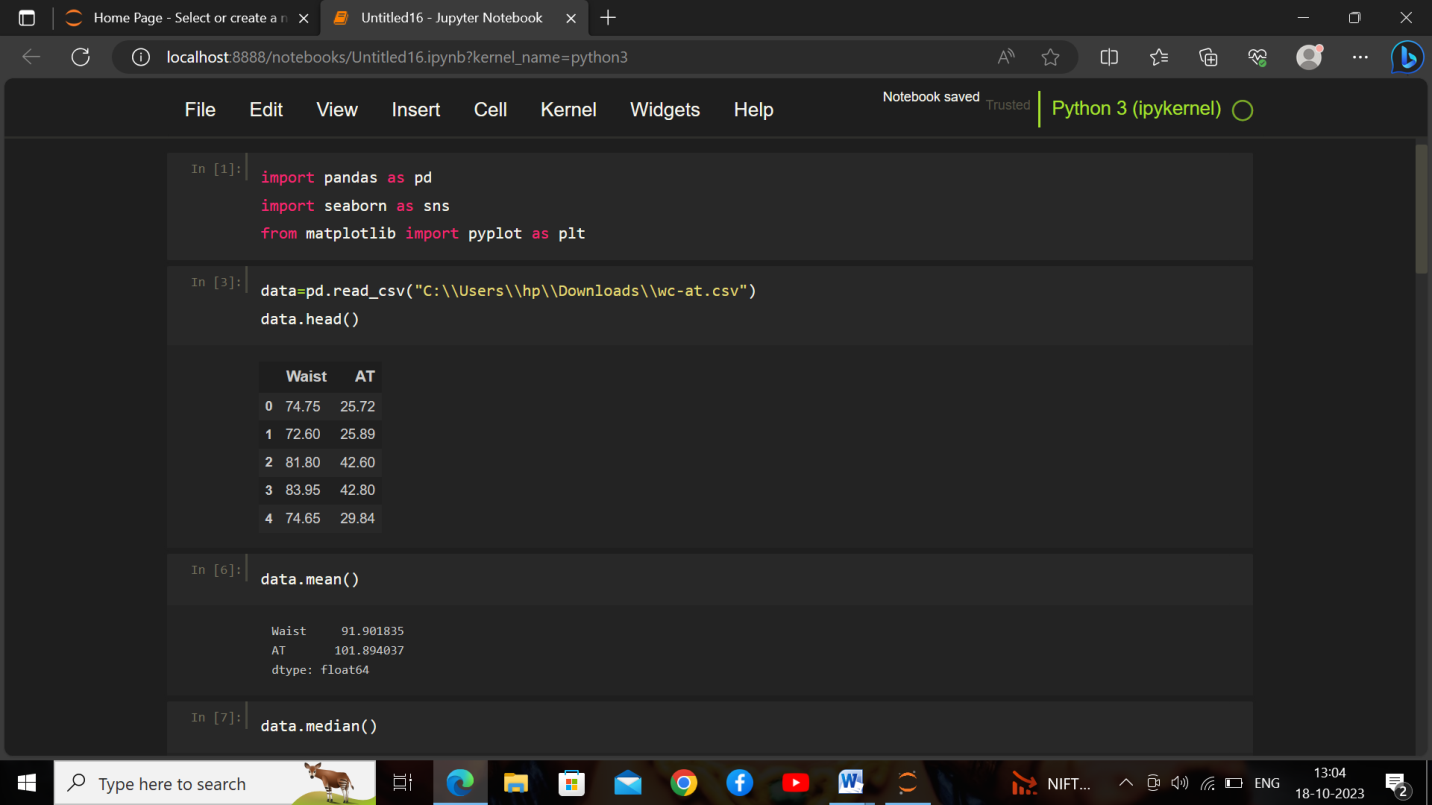




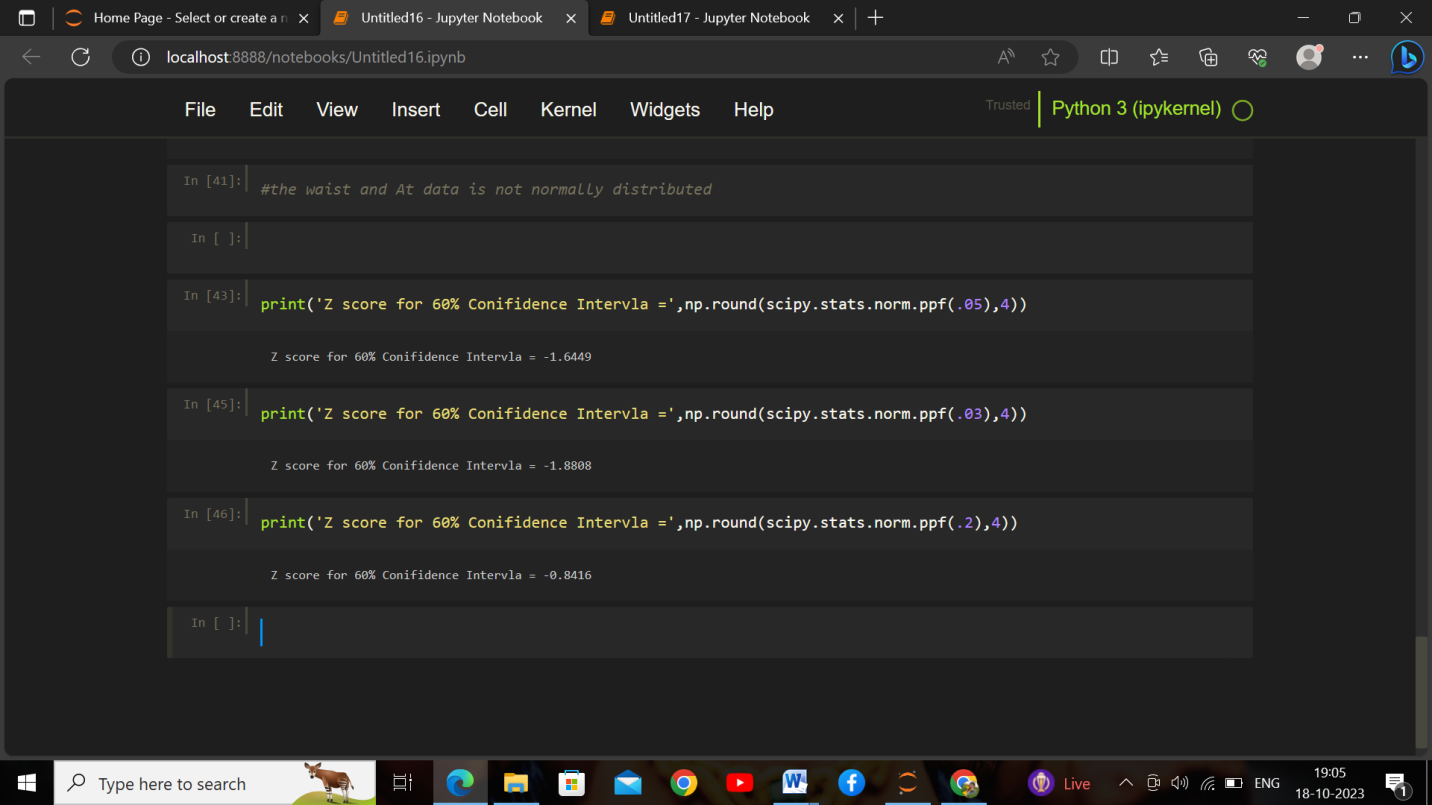


1. Check Whether the Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set follows Normal Distribution

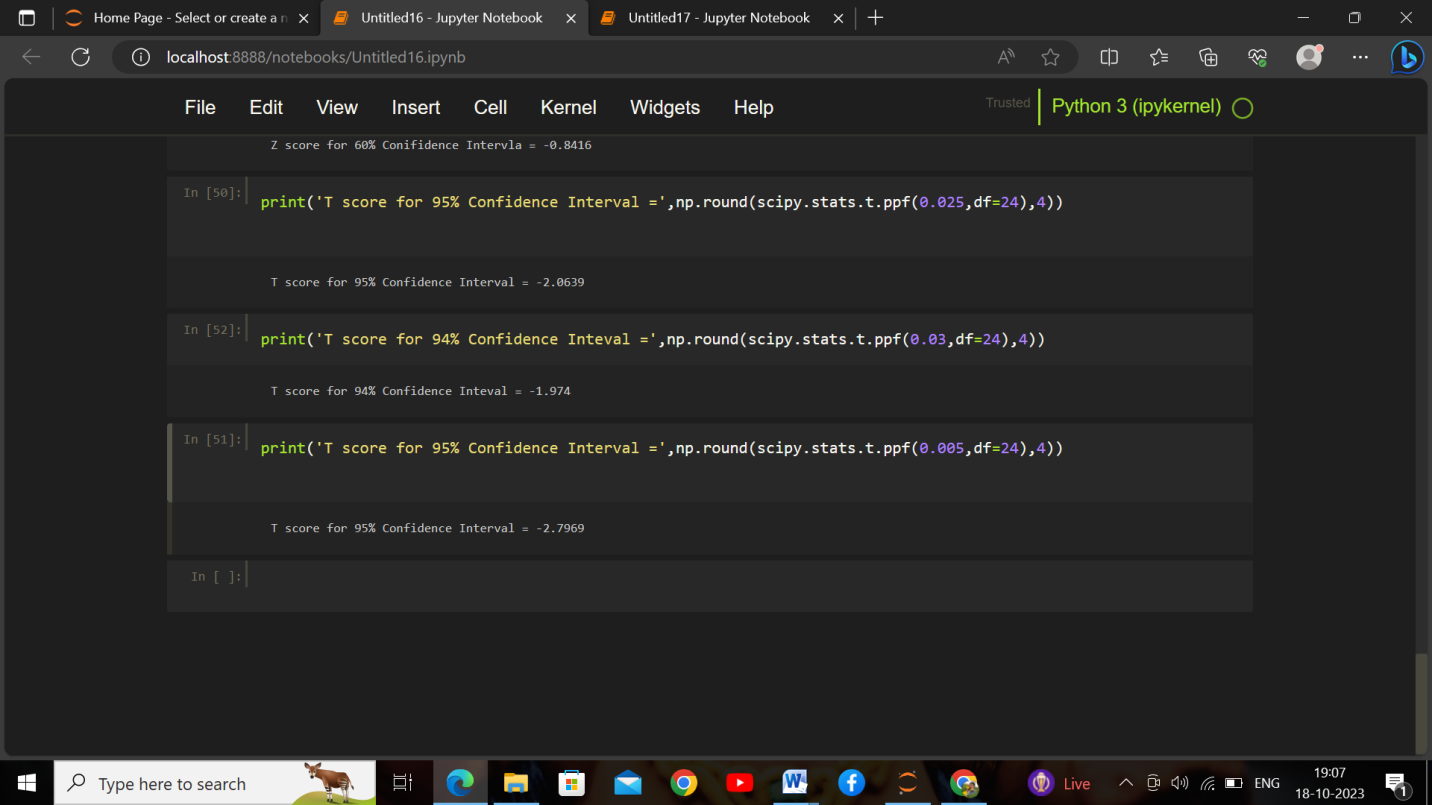
Dataset: wc-at.csv



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

