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

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

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

Solution: Total number of possible outcomes is 23 = 8

No. of Favorable outcomes = 3

P(Two head & one tail) = 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

Solution:

Total number of possible outcomes is 62=36

1. Equal to 1:

Favorable outcome (sum equal to 1) = 0

Required probability= 0/36 = **0**

1. Less than or equal to 4

Favorable outcome (Less than or equal to 4) = 6

Required probability = 6/36 = **1/6= 0.166**

1. Sum is divisible by 2 and 3:

Favorable outcome (Sum is divisible by 2 and 3) = 5

Required probability= **5/36**

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?

**Solution:**

Total number of balls = (2+3+2) = 7

Sample(S)= Number of ways of drawing 2 balls out of 7 = 7C2 = 21

Event(E) = Event of 2 balls, none of which is blue = 5C2 = 10

So, P(E)= n(E)/n(S) =**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

**Solution:**

Expected number of candies for a randomly selected child

= 1 \* 0.015 + 4\*0.20 + 3 \*0.65 + 5\*0.005 + 6 \*0.01 + 2 \* 0.12

= 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 and Comment about the values/ Draw some inferences.

**Use Q7.csv file**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Points** | **Score** | **Weigh** |
| **Mean** | 3.558 | 3.070 | 17.763 |
| **Median** | 3.695 | 3.325 | 17.71 |
| **Mode** | 3.92 | 3.44 | 17.02 |
| **Variance** | 0.285 | 0.957 | 3.193 |
| **Standard Deviation** | 0.526 | 0.963 | 1.759 |
| **Range** | 2.17 | 3.911 | 8.4 |

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?

**Solution:**

Expected 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)

**= 145.33**

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

**Cars speed and distance**

**Use Q9\_a.csv**

|  |  |  |
| --- | --- | --- |
|  | **Speed** | **Distance** |
| **skewness** | -0.117510 | 0.806895 |
| **kurtosis** | -0.508994 | 0.405053 |

Skewness: From the above table we can say that ‘dist’ is positively skewed whereas ‘speed’ is negatively skewed.

Kurtosis: From the above table we can say, kurtosis of ‘dist’ has Heavier tail than ‘speed’. Whereas, the tail of ‘speed’ is light.

**SP and Weight (WT)**

**Use Q9\_b.csv**

|  |  |  |
| --- | --- | --- |
|  | SP | Weight |
| **skewness** | 1.61 | -0.61 |
| **kurtosis** | 2.98 | 0.95 |

Skewness: From the above table we can say that, ’SP’ is positively skewed whereas WT is negatively skewed

Kurtosis: From the above table we can say, Kurtosis of ‘SP’ has heavier tail than ‘WT’ which has comparatively lighter tail

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



**Ans:** The histograms peak has right skew (positively skewed) and tail is on right. Mean > Median. We have outliers on the higher side.



**Ans:** There are around 8 outliers in the above box plot.

The distribution is not equal (Symmetric) and the interface for this box plot is positively skewed.

**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?

Ans:

Confidence interval: X (Z1-α.σ/sqrt(n))

Degrees of freedom= 2000-1= 1999

Confidence interval= 94%

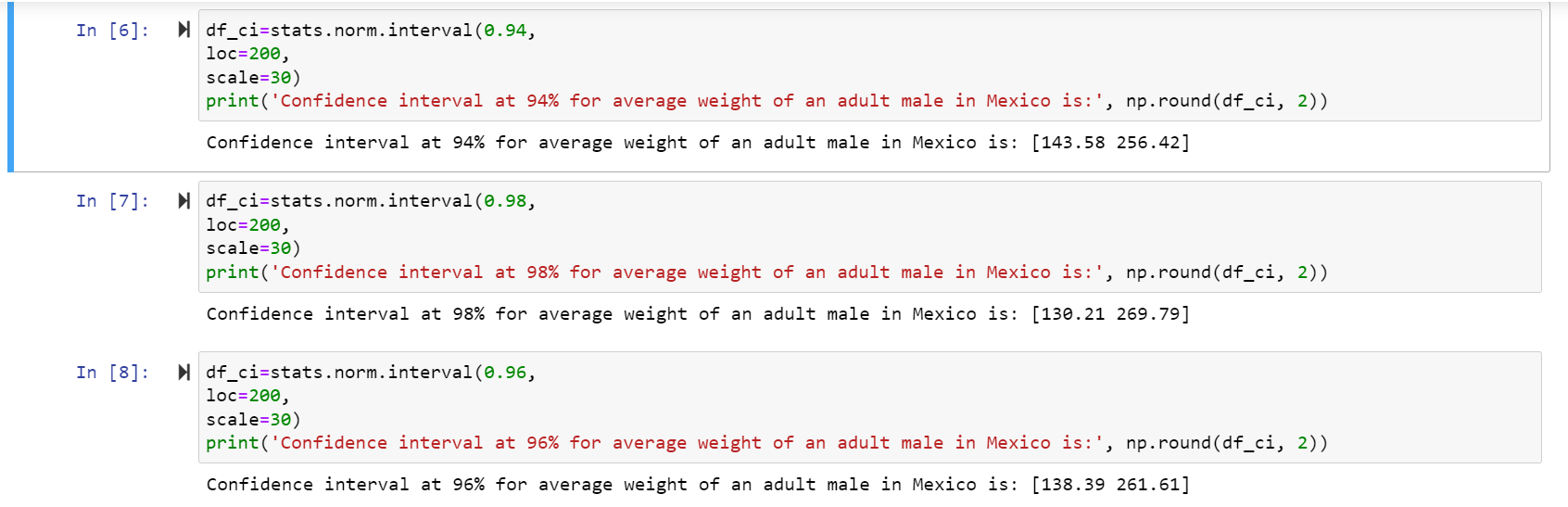
=(1-σ/2)= 1-0.03 =0.97

for confidene interval for 94% is 1.882

Confidence interval for 98%= 2.33

Confidence interval for 96% = 2.05

Using Coding:

****

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

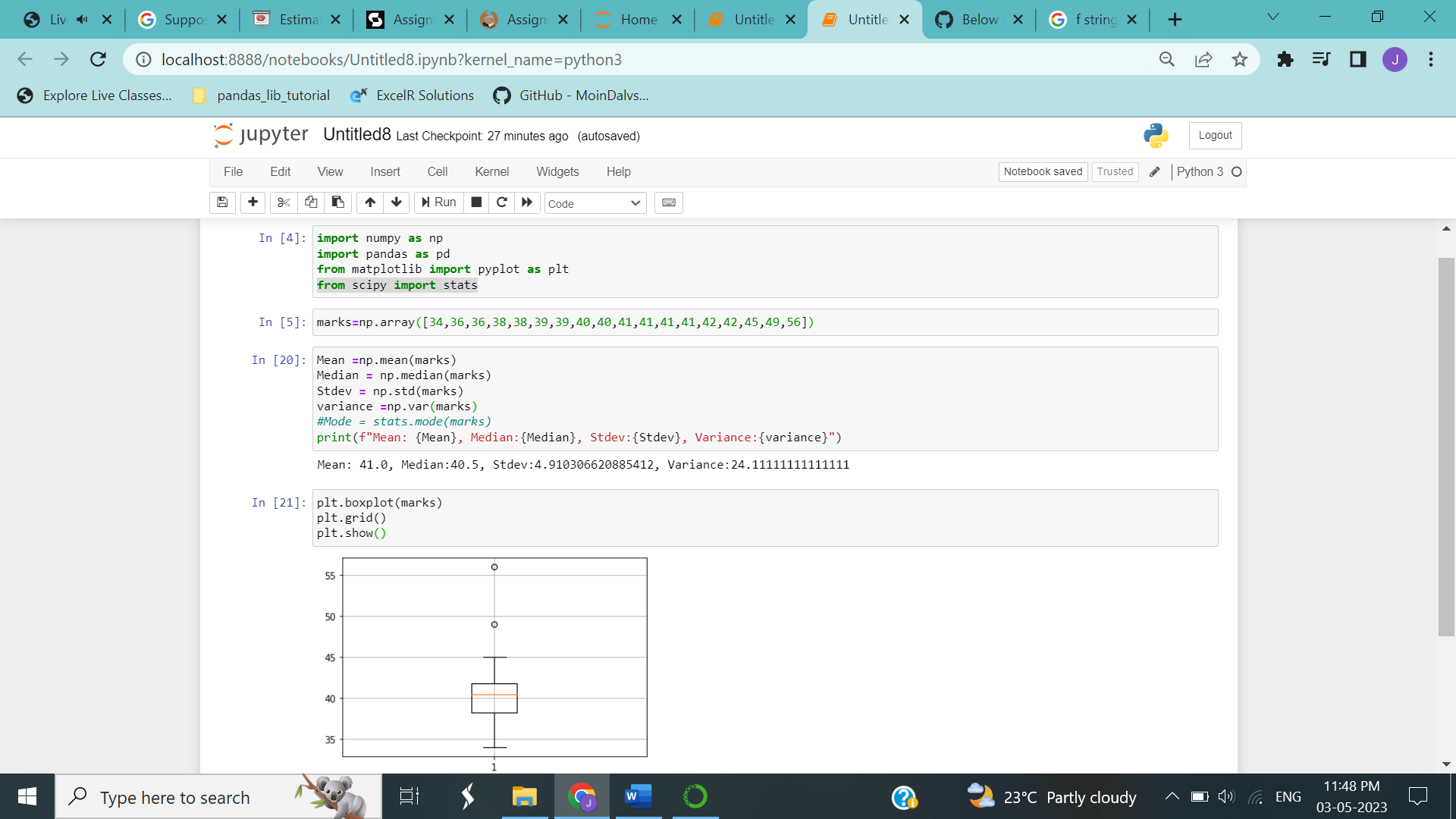
**34,3 6,36,38,38,39,39,40,40,41,41,41,41,42,42,45,49,56**

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

Ans: 1) Mean= 41, Median= 40, variance= 24.111, Standard deviation= 4.910

2) From below plot we can say that mean of marks of student is 41 which is slightly greater than

median.Most of the students got marks in between 41-42, there are two outlier 49,56.



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

**Ans:** No skewness is present we have a perfect symmetrical distribution.

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

**Ans:** Skewness and tail is towards Right i.e., positively skewed.

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

**Ans:** Skewness and tail is towards left i.e., left skewed.

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

**Ans:** Positive kurtosis means the curve is more peaked and it is Leptokurtic.

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

**Ans:** Negative Kurtosis means the curve will be flatter and broader.

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



* What can we say about the distribution of the data?

**Ans:** The above Boxplot is not normally distributed the median is towards the higher value

* What is nature of skewness of the data?

**Ans:** The data is a skewed towards left i.e., left skewed. The whisker range of minimum value is greater than maximum.

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

**Ans:** The Inter Quantile Range = Q3 Upper quartile – Q1 Lower Quartile

= 18 – 10

**IQR = 8**

Q19) Comment on the below Boxplot visualizations?



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

Ans: a. The data of both the box plots is normal as it is equally distributed.

b. There are no presence of outliers in both the box plots.

c. The median of the both box plots are approximately equal 262.5.

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)

Ans:

1. P(MPG>38)

sd(MPG)=9.131445, mean(MPG)=34.42208

=1 - pnorm(38,mean(MPG),sd(MPG))

= 0.347

****

1. P(MPG,40)

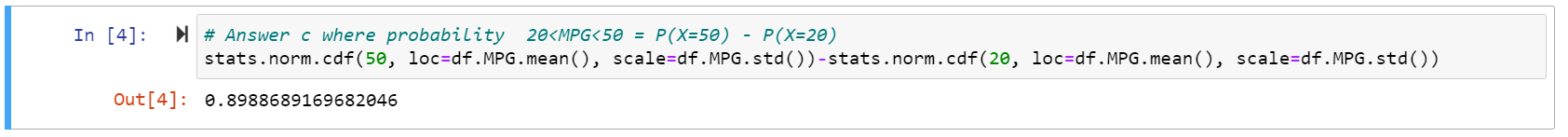
=pnorm(40,mean(MPG),sd(MPG))

= 0.7293

1. P(20<MPG<50)

=pnorm(50,mean(MPG),sd(MPG)) - pnorm(50,mean(MPG),sd(MPG))

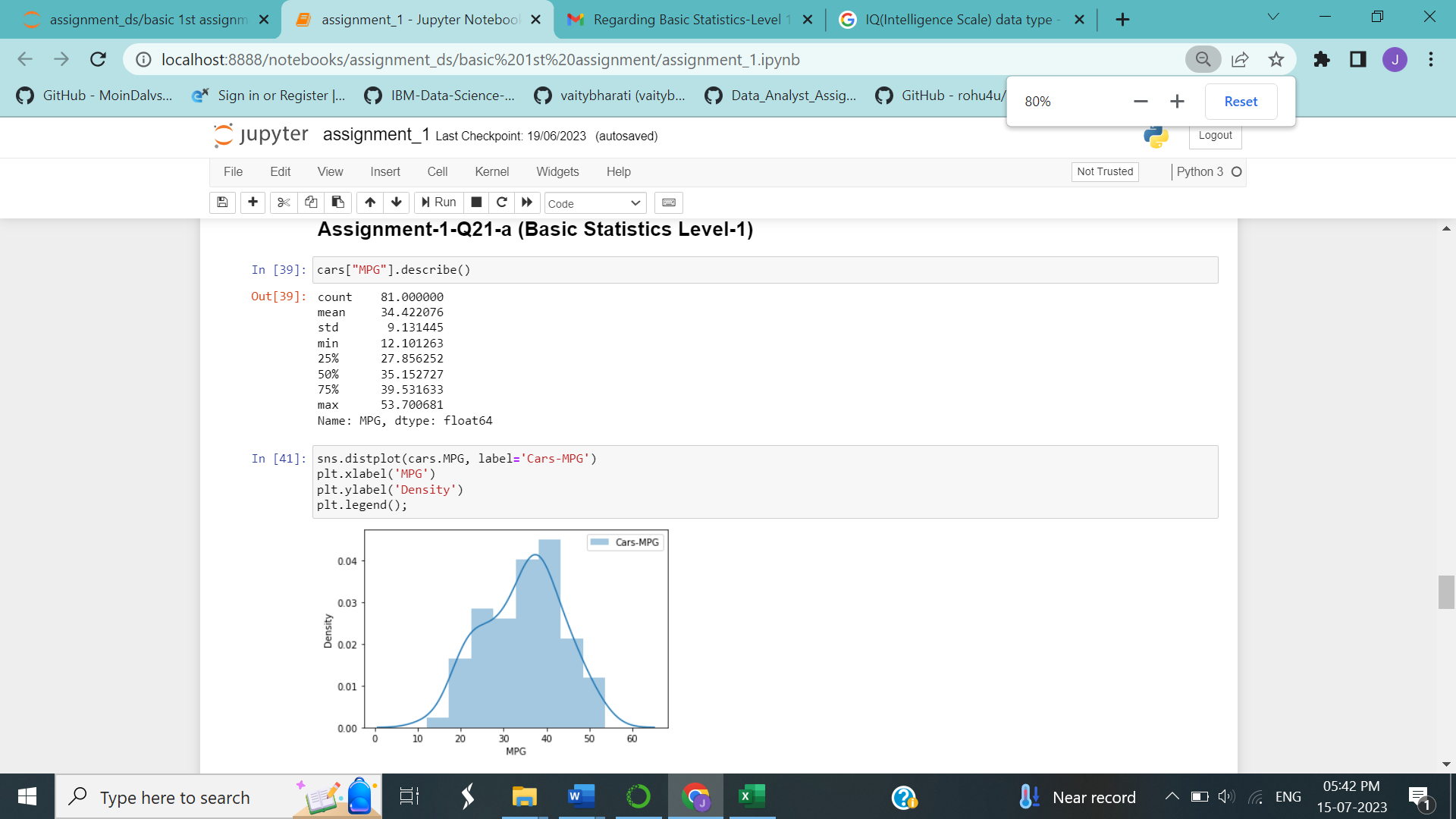
= 0.8988



Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

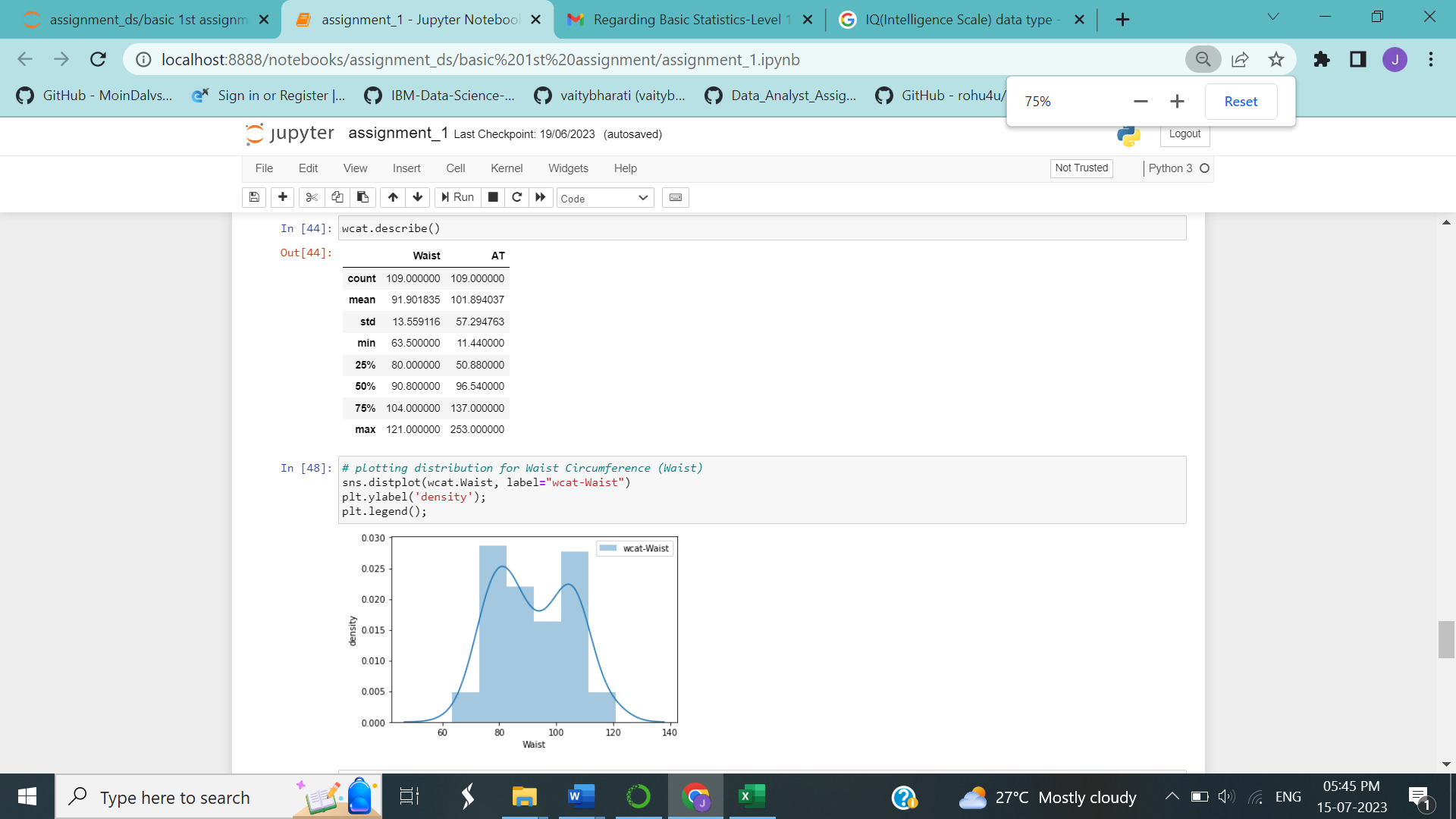
Dataset: Cars.csv



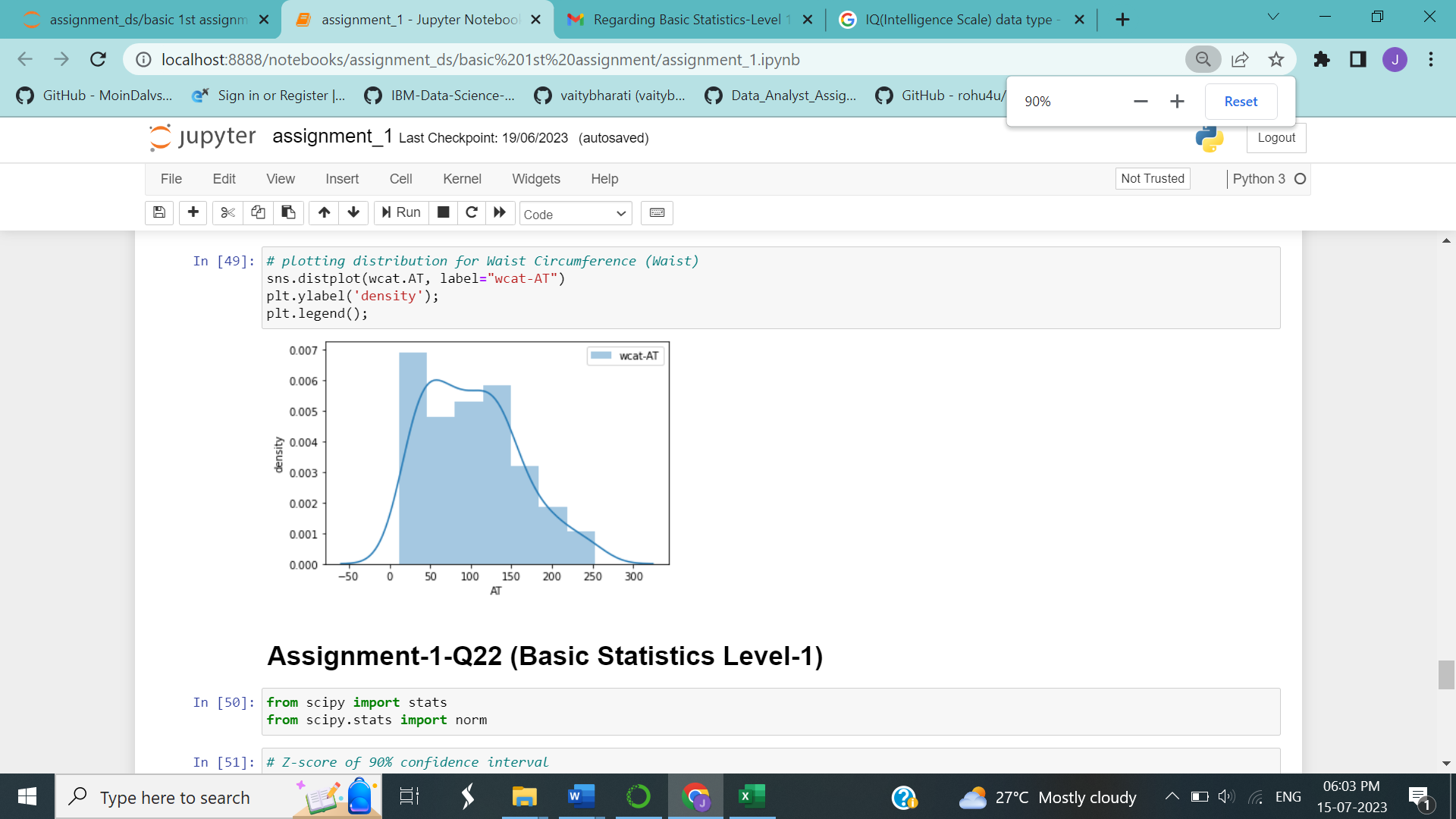
1. Check Whether the Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set follows Normal Distribution

Dataset: wc-at.csv

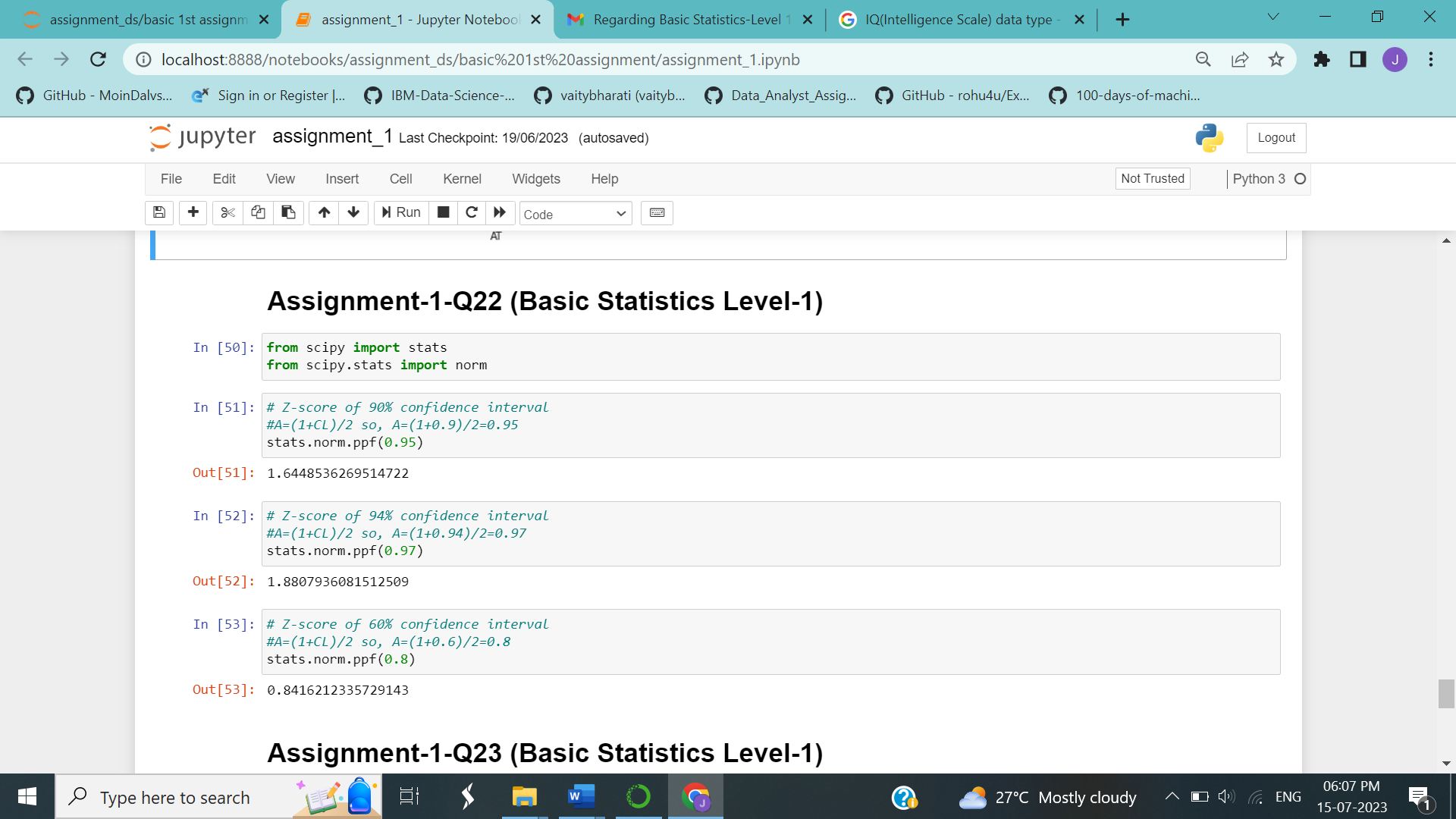
1)From the values and graph below, we can say that Waist of wc-at is not normally distributed as mean, median and mode are not exactly equal (Partially equal)



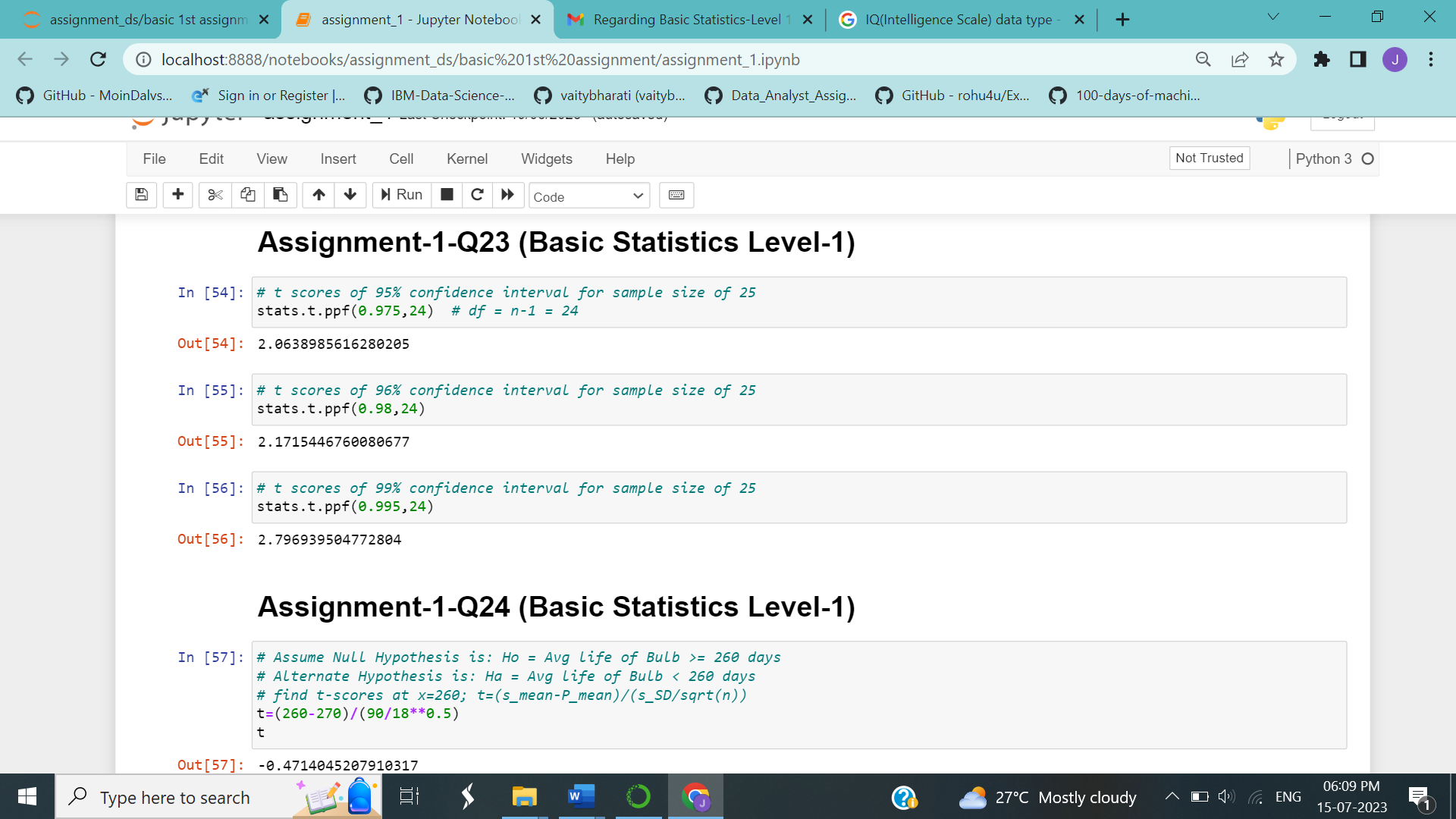
2) From the values and graph below, we can say that AT of wc-at is not normally distributed as mean, median and mode are not exactly equal.



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

Solution:

t – statistics : t= (x- μ)/(s/

x = mean of the sample of bulbs = 260

μ = population mean = 270

s = standard deviation of the sample = 9

n = number of items in the sample = 18

Find t-scores

t = (x̄- μ)/(S/n^0.5)

= (260- 270)/(9/

t = - 0.471

Now, we find P value: -

P value = stats.t.cdf (-0.4714045, df=17)

= 0.32167

Hence, Probability that 18 randomly selected bulbs would have an average life of no more than 260 days is **32.17%**

**Thus, as p-value > α, we accept Ho i.e. The CEO claims are false and the avg. life of bulb > =260 days**

