|  |  |
| --- | --- |
| 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 | Categorical |
| Number of kids | Discrete |
| Number of tickets in Indian railways | Discrete |
| Number of times married | Discrete |
| Gender (Male or Female) | 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 | Ordinal |
| Fahrenheit Temperature | Interval |
| Height | Ratio |
| Type of living accommodation | Nominal |
| Level of Agreement | Ordinal |
| IQ(Intelligence Scale) | Interval |
| Sales Figures | Ratio |
| Blood Group | Nominal |
| Time Of Day | Ordinal |
| Time on a Clock with Hands | Interval |
| Number of Children | Nominal |
| Religious Preference | Nominal |
| Barometer Pressure | Ratio |
| SAT Scores | Interval |
| Years of Education | Ratio |

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

Sol: The total possible outcomes are:

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

Therefore, Number of favorable outcomes =3/P(two head and one tail)

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

Sol:

1. The Total number of outcomes = 36

Number of outcomes sum equal to ‘1’= 0

p(equal to 1) = 0/36

(Not possible , Minimum sum is 2)

1. Total number of outcomes =36

Number of outcomes less than or equal to ‘4’ = 6

p(less than or equal to 4) = 6/36 =1/6

1. Total number of outcomes =36

Number of outcomes sum is divisible by 2 and 3 =6

p(sum is divisible by 2 and 3)=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?

Sol:

Total Number of balls = 7

Number of ways of drawing out 2 balls out of 7 is n(S)= 7C2 = (7\*6)/2 = 21

Number of ways of drawing 2 balls out of 5 n(S) =5C2 = (5\*4)/2 = 10

Therefore P(E) = 𝑛(𝐸)/𝑛(𝑆) = 10/21 = 0.476 = 47.6%

**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

**Sol:-**

Expected value = Probability\*candies

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

= 3.09

The Expected number of candies for a randomly selected child is 3.09

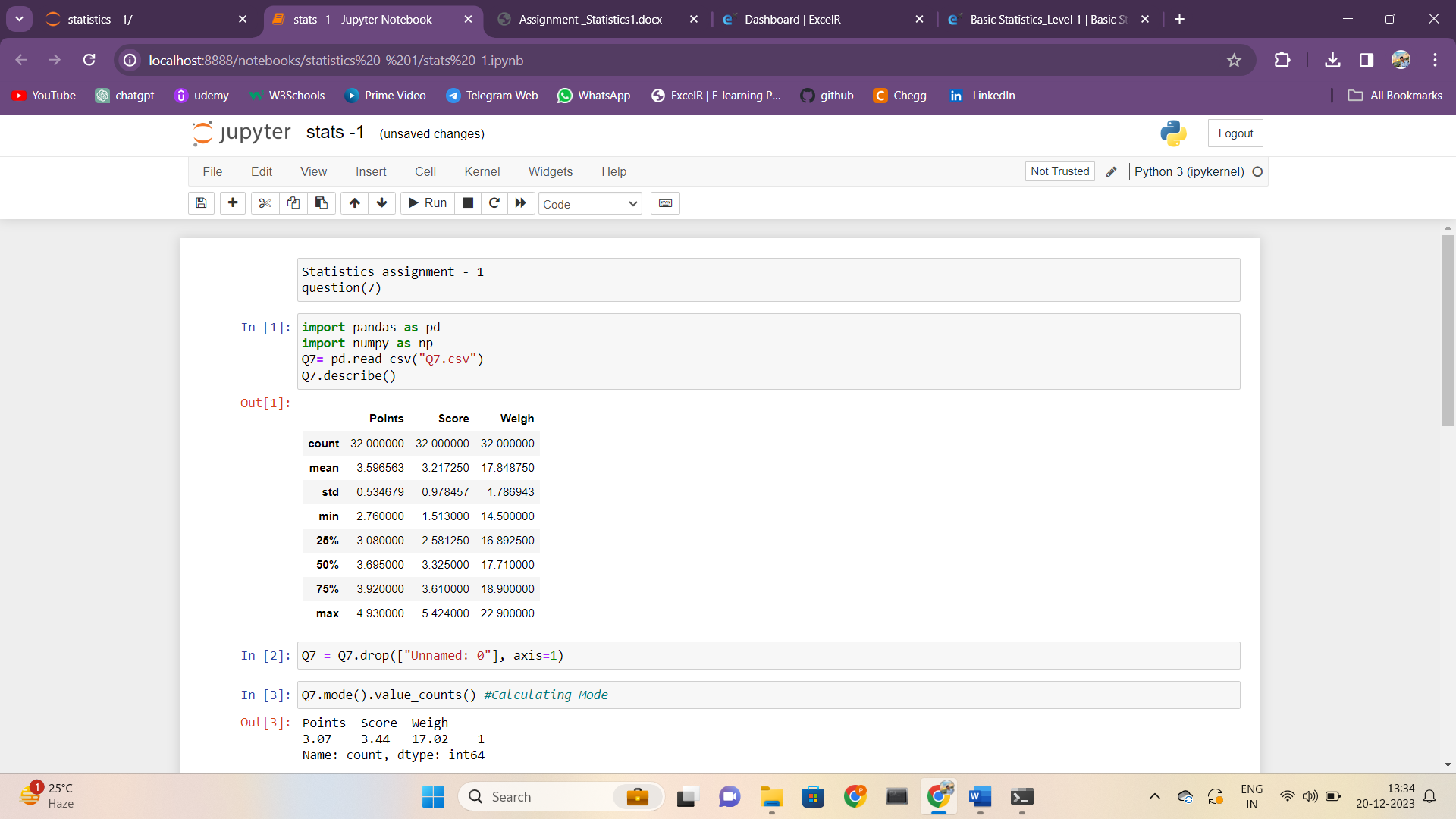
Q7) Calculate Mean, Median, Mode, Variance, Standard Deviation, Range & comment about the values / draw inferences, for the given dataset.

* 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

**Sol:-**

****

**A screenshot of a computer

Description automatically generated**

**Inferences :**

From the given data we can draw the following inferences

• Points and Score has Median greater than Mean. Therefore its Right

Skewed Distribution

• Weigh has Mean greater than Median. Therefore, it is Left Skewed

distribution

• We can also observe that Points and Weigh has more than one mode.

Therefore, they are Multimodal whereas Score has only one mode

• When boxplot is plotted for the following data set, we can clearly see there

are outliers in the given dataset.

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?

Sol:-

Expected Value = Σ (Probability \* Value) = ΣP(X).E(X)

From given Info There are 9 patients

Probability of selecting each patient = 1/9.

Expected Value = (1/9) (108) + (1/9) (110) + (1/9) (123) + (1/9) (134) + (1/9)

(135) + (1/9) (145) + (1/9) (167) + (1/9) (187) + (1/9) (199)

= (1/9) (108 + 110 + 123 + 134 + 135 + 145 + 167 + 187 + 199)

= (1/9) (1308)

= 145.33

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

**Cars speed and distance**

**Use Q9\_a.csv**

**-->** Skewness Of the data Kurtosis of the data

speed -0.117510 speed -0.50899

dist 0.806895 dist 0.405043

**SP and Weight(WT)**

**Use Q9\_b.csv**

--> Skewness of the data Kurtosis of the data

SP 1.611450 SP 2.977329

WT -0.614753 WT 0.950291

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



Solution:

Inferences based on Histogram are as follows:

• From the given histogram the skewness is positive and right tailed.

• And it is also evident that the Median is greater than Mean of the data.

Inferences based on Boxplot are as follows:

• Boxplots are mainly used to detect outliers.

• From the given boxplot we can see there are more outliers in the given data.

• The range of whisker is very wide in the upper quartile region. We should treat these outliers to get normal distribution.

**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?

**Sol:-**

A screenshot of a computer

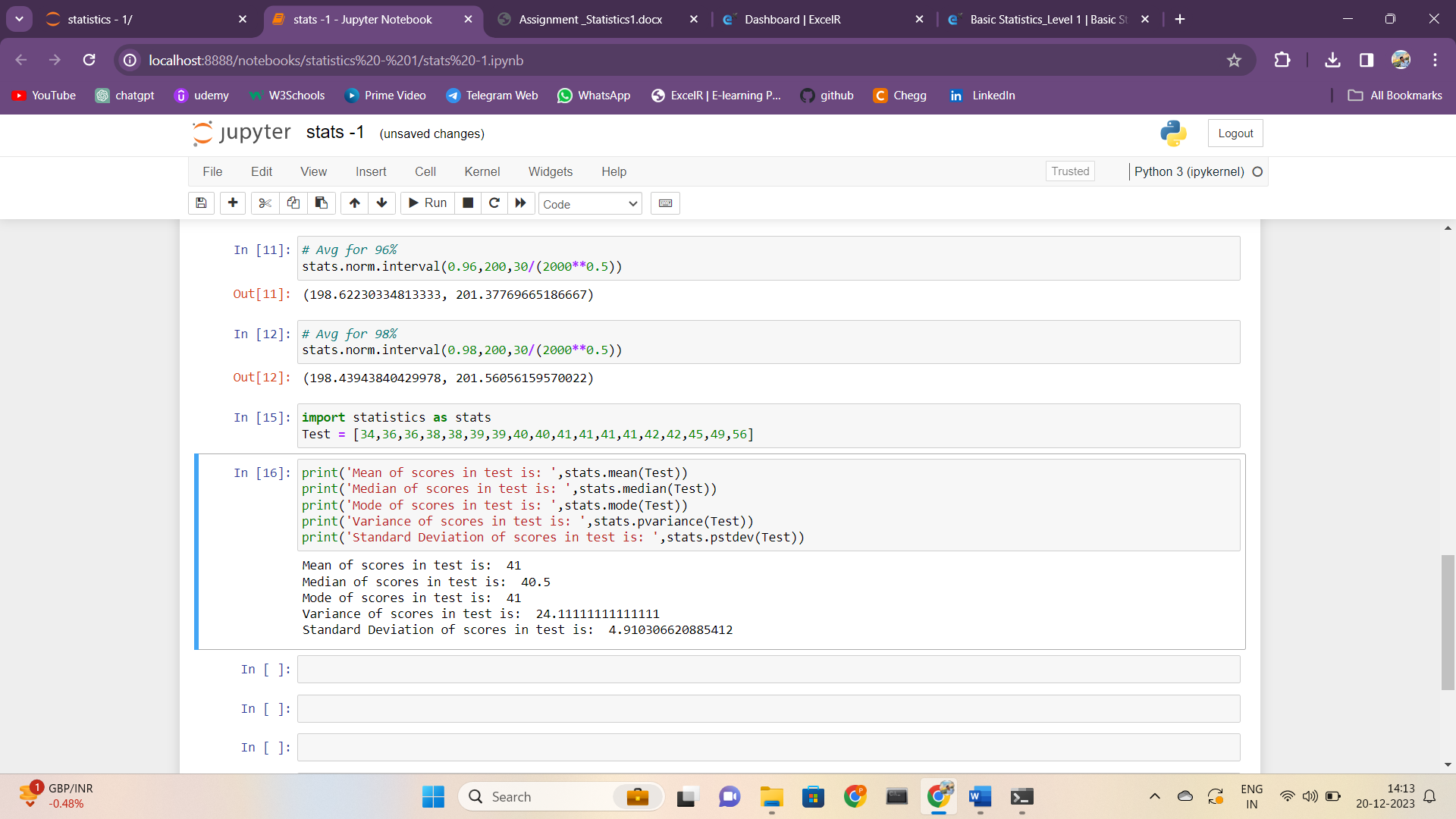
Description automatically generated

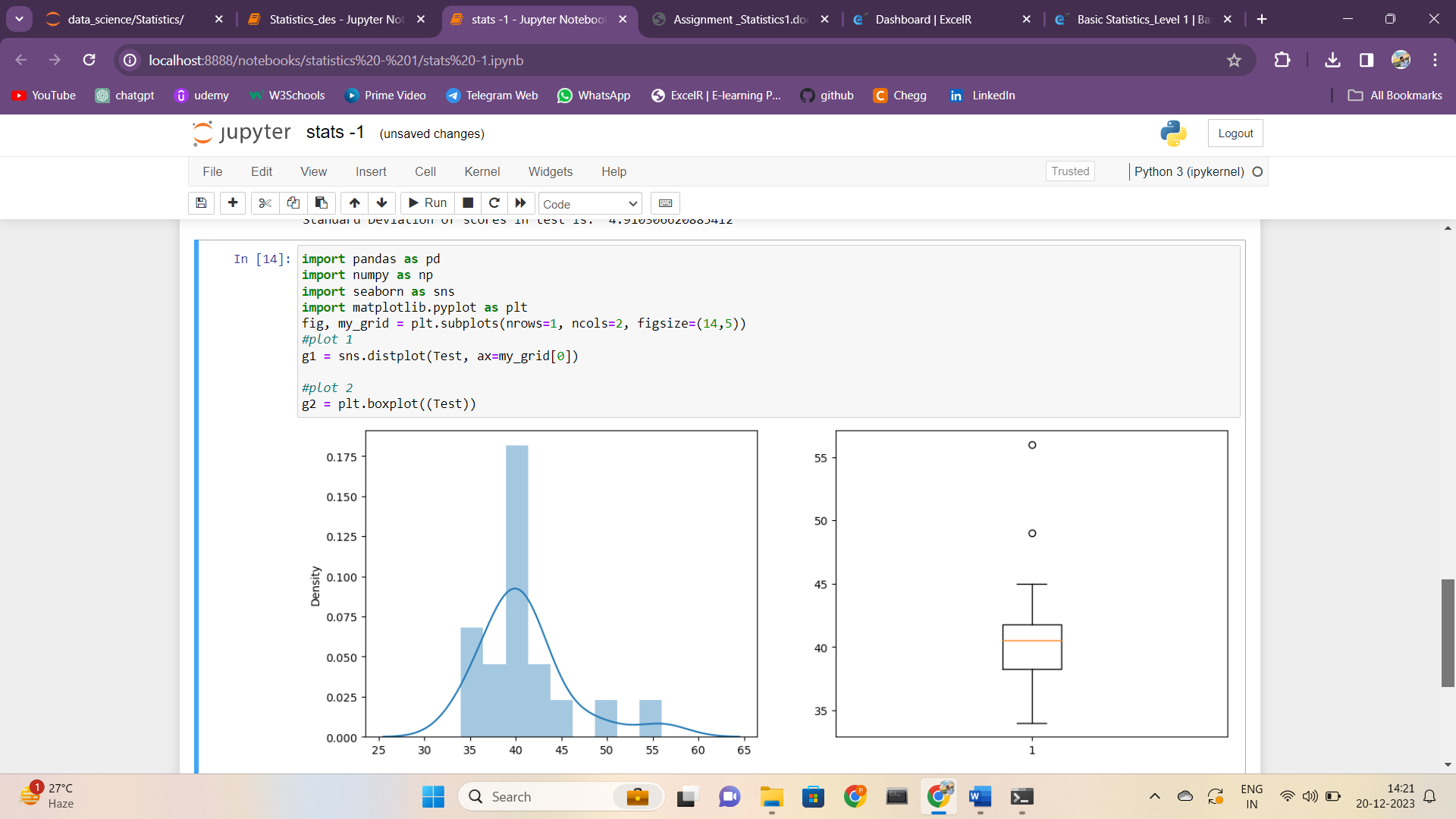
**Q12)** Below are the scores obtained by a student in tests.

**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?

Sol:-





**Inference:**

• From the graph we can see that the marks of students is symmetrically distributed as the mean and median are almost equal.

• From the boxplot we have two outlier values which are beyond the upper Quartile region.

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

Sol: - If the mean and median of data are equal this implies the skewness is zero. Therefore, the nature of skewness is Symmetrically Distributed.

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

Sol: - If the mean is greater than the median, the distribution is positively skewed

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

Sol: - If the mean is less than the median, the distribution is negatively skewed.

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

Sol: - Positive values of kurtosis indicate that distribution is peaked and possesses thick tails.

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

Sol: - A platykurtic distribution or negative kurtosis is flatter (less peaked) when

compared with the normal distribution.

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



What can we say about the distribution of data?

**Sol**:- From the boxplot it is clear that the data is not symmetrically distributed and most of the values lie below the lower quartile region.

What is nature of skewness of the data?

**Sol**:- The Distribution of the data is left skewed.

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

**Sol:-** Inter-Quartile Range = Upper Quartile – Lower Quartile

=18 – 10

=8

Therefore IQR = 8.

Q19) Comment on the below Boxplot visualizations?



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

Sol:- Following are the inferences from the given boxplot :

• Comparing both the boxplot 1 and 2, They are symmetrically distributed.

Boxplot 2 has more wider number of data than boxplot 1. Variation is more

in boxplot 2 compared to boxplot 1.

• The medians are all at the same level. However, the box plots of 1 and 2 show very different

distributions of views.

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

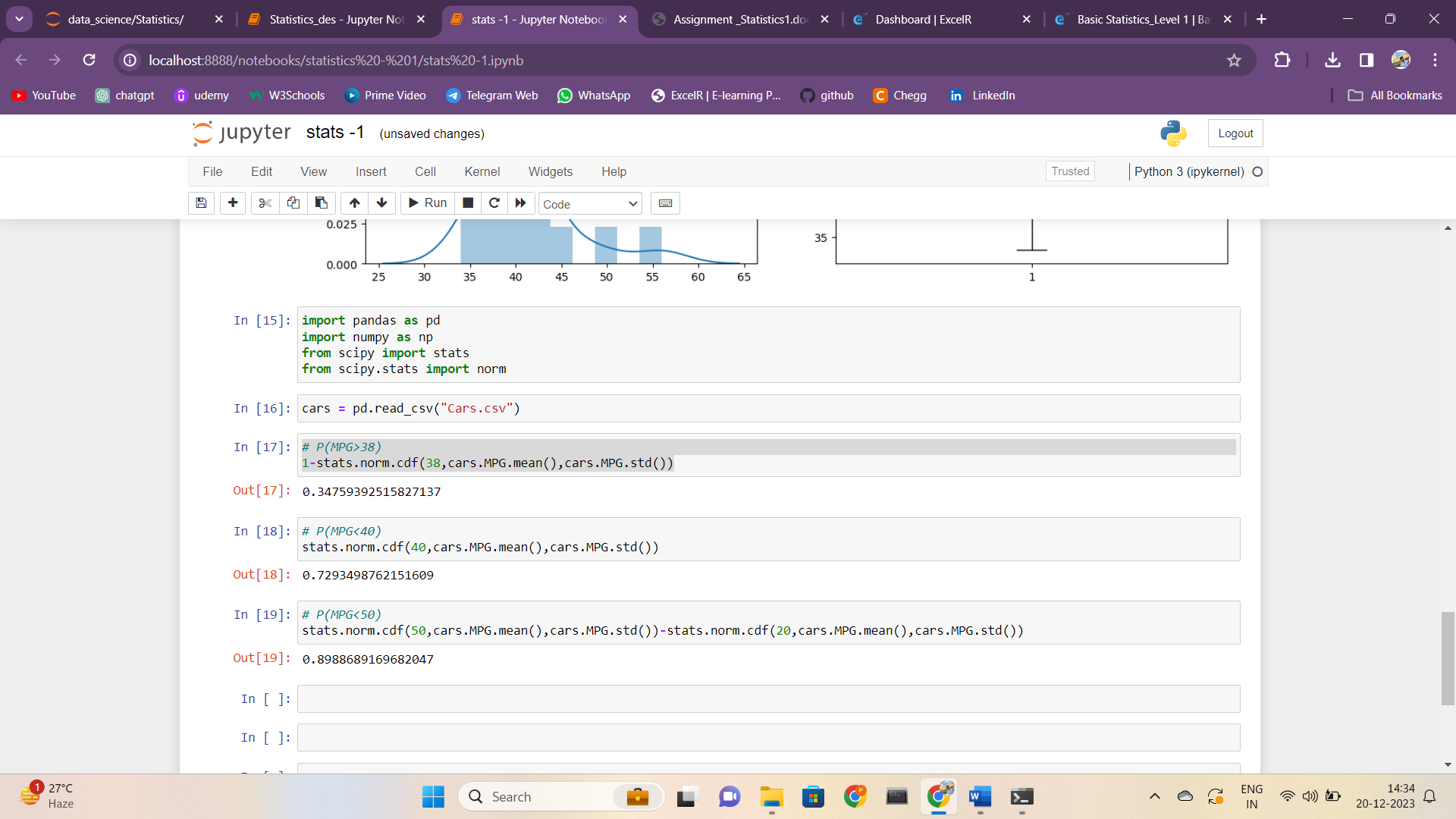
Data \_set: Cars.csv

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

MPG <- Cars$MPG

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

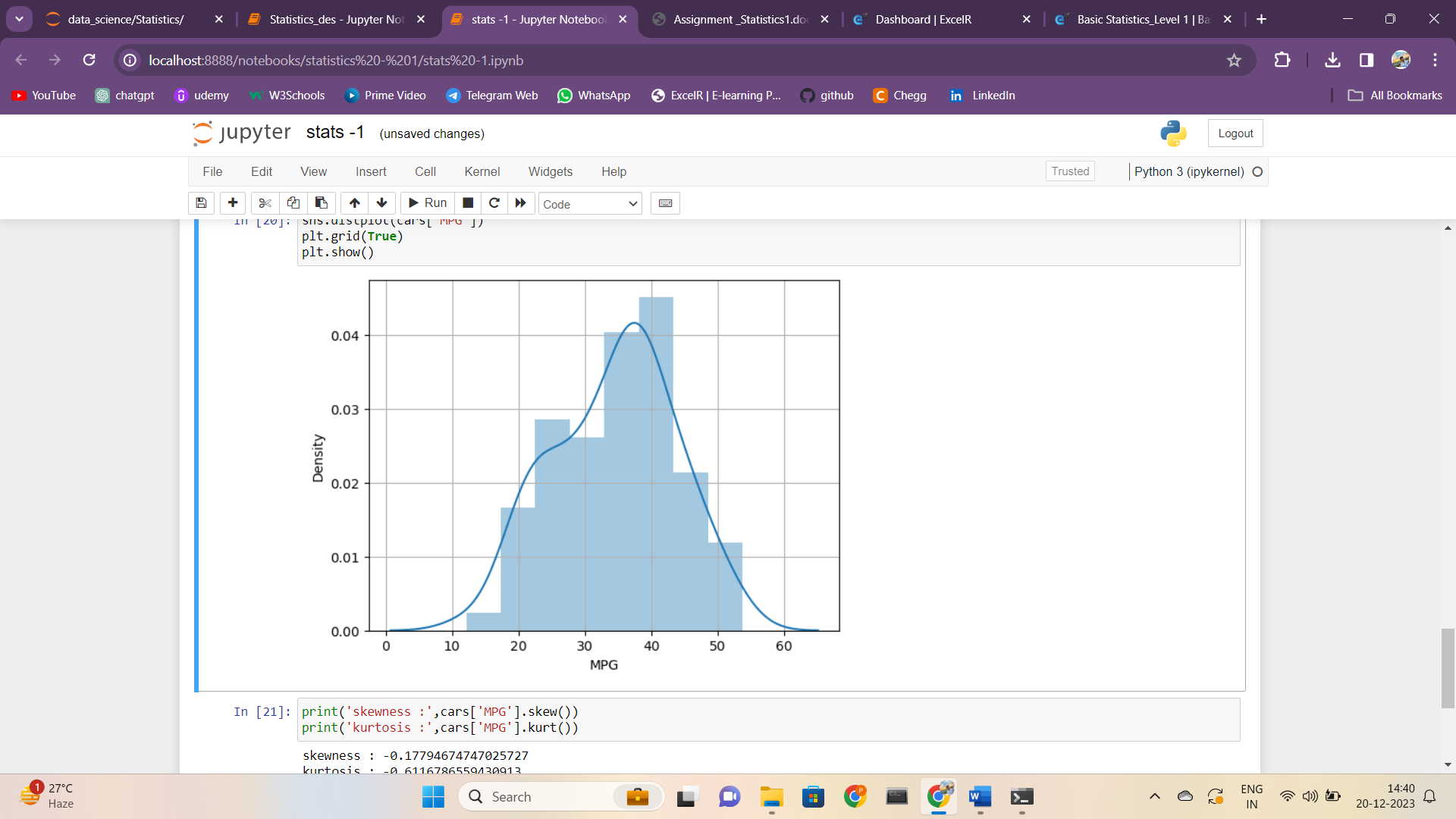
c. P (20<MPG<50)



Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

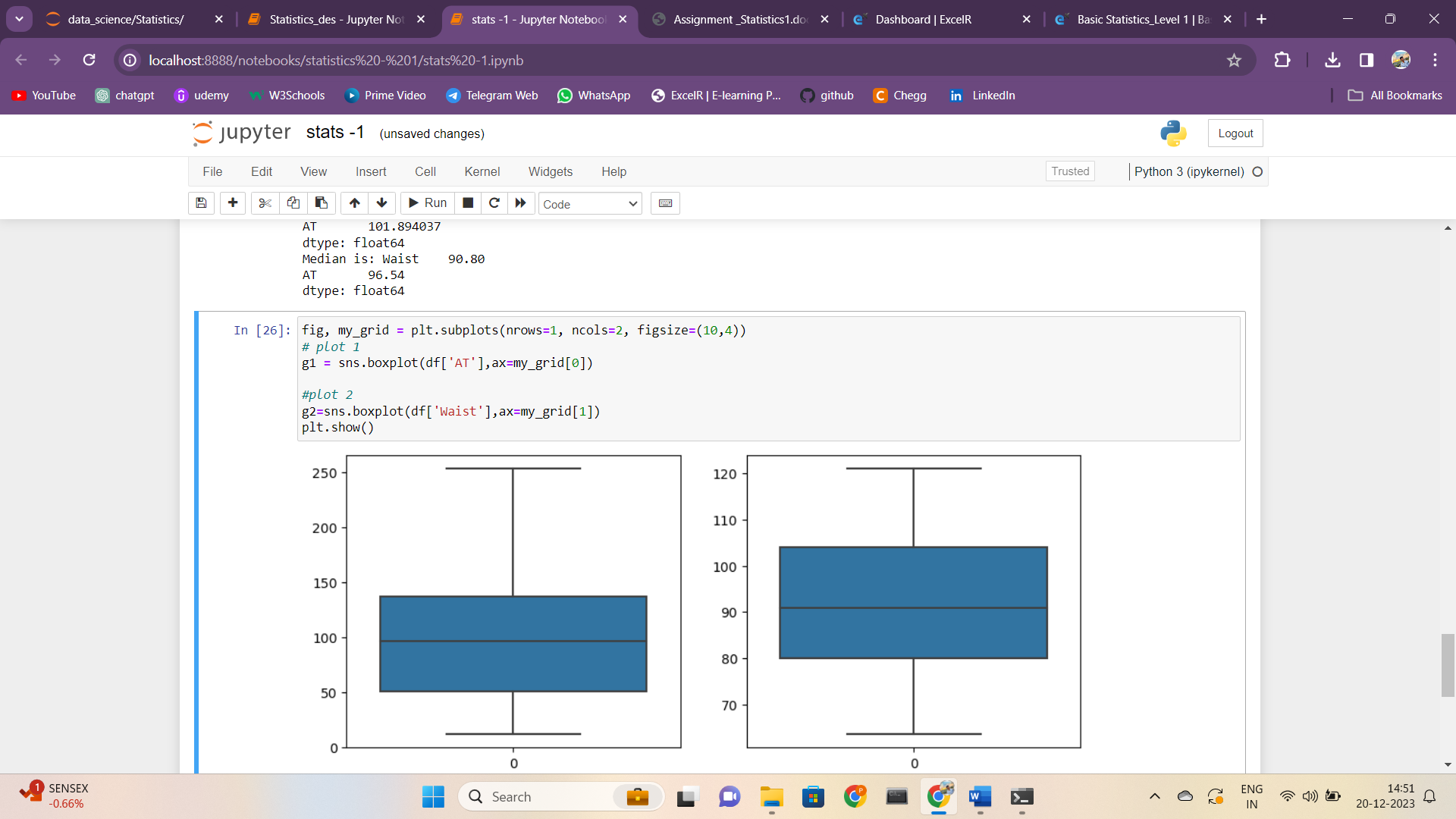


1. Check Whether the Adipose Tissue (AT) and Waist Circumference (Waist) from wc-at data set follows Normal Distribution

Dataset: wc-at.csv

A screenshot of a computer

Description automatically generated



From the boxplot we can inference about the distribution of data.

• In boxplot of AT, mean > median, right whisker is larger than left whisker, data is positively skewed.

• In boxplot of Waist mean > median, both the whisker are of same length, median is slightly shifted towards left. Data is symmetrically distributed.

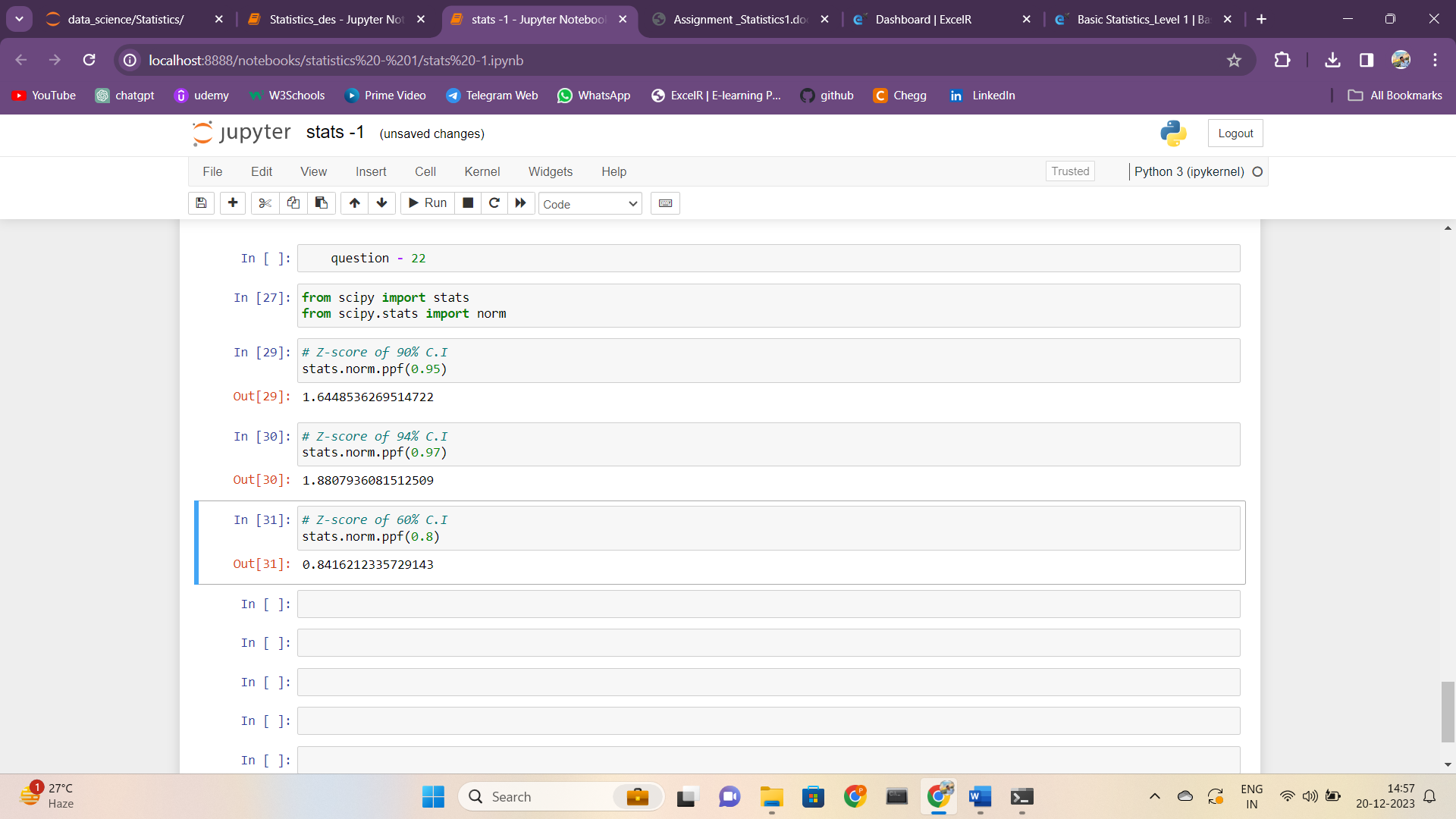
Q 22) Calculate the Z scores of 90% confidence interval,94% confidence interval, 60% confidence interval

Sol:-

For 90% confidence interval: A= (1+0.9)/2 = 0.95

For 94% confidence interval: A= (1+0.942)/2 = 0.97

For 60% confidence interval: A= (1+0.602)/2 = 0.80



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

Sol:-

For 95% confidence interval: A= (1+0.952)/2 =0.975

For 96% confidence interval: A= (1+0.962)/2 = 0.98

For 99% confidence interval: A= (1+0.992)/2 = 0.995

A screenshot of a computer

Description automatically generated

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

A screenshot of a computer

Description automatically generated