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

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

Sol:- {HHH , THH , TTH , TTT , HTT , HHT , THT , HTH } these are the number of possible outcomes.

So out of these outcomes we have the probability of getting two heads and one tail is 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:- total possible outcomes = 62 =36

1. Favorable outcomes(sum equal to 1) = 0

Required probability =0/36 =0

1. Favorable outcomes(less than or sum equal to 4) = 6

Required probability =6/36 =1/6

c)(sum is divisible by 2 and 3) /(two dice rolled) =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?

Ans:-total number of balls = 7 balls

Let S be the sample space

Then , n(S) = number of ways of drawing 2 balls out of 7 = 7c2

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

=21

Let E =event of drawing 2 balls ,none of which is blue

n(E)=number of ways of drawing 2 balls out of (5) balls =5c2

=(5\*4)/(2\*1)

=10

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

Ans :- 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\*0.120

=3.076

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**

**Ans :- Points:-**

**Mean =**3.596

**Median=**3.695

**Mode=**3.07

**Variance=**0.285

**Standard deviation=**0.534

**Range=**2.17

**Score:-**

**Mean = 3.21**

**Median=**3.32

**Mode=3.44**

**Variance=**0.95

**Standard deviation=**0.97

**Range=3.91**

**Weigh:-**

**Mean =**17.84

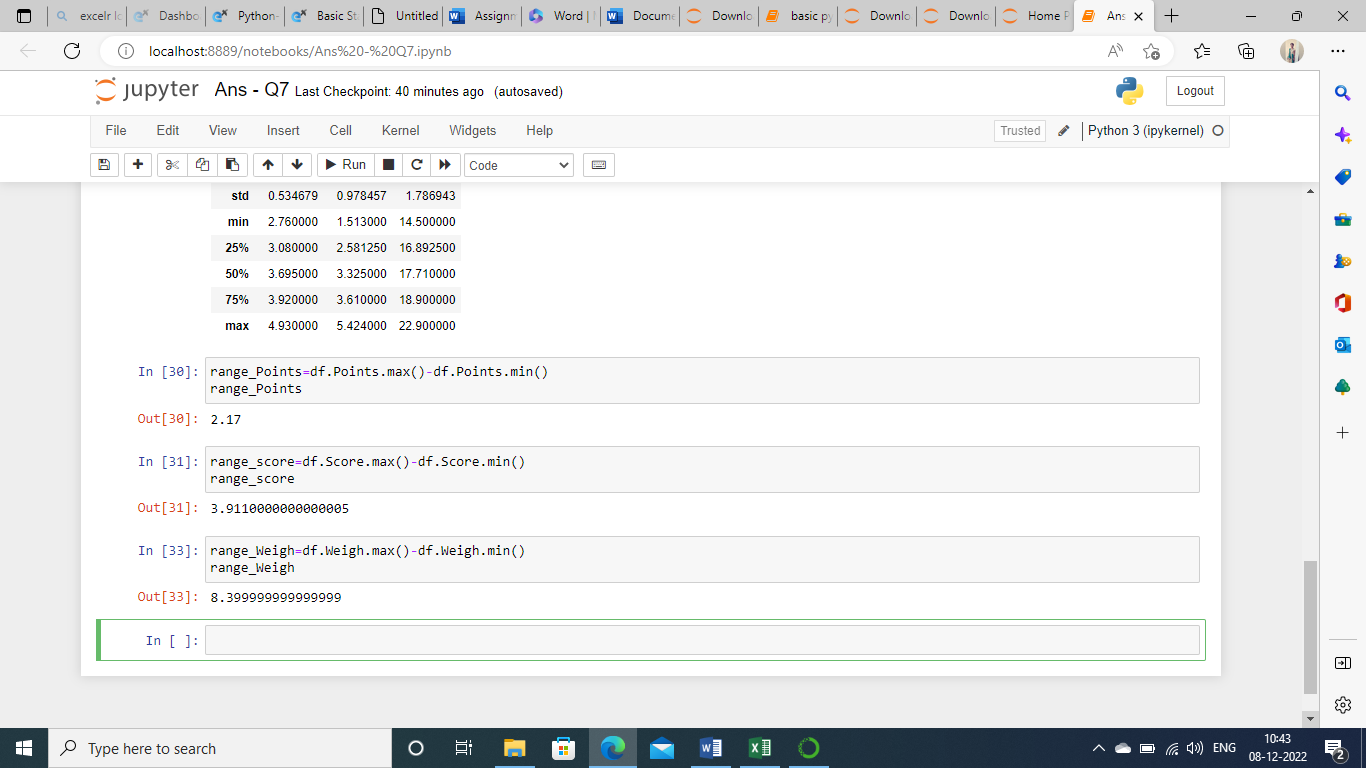
**Median=**17.71

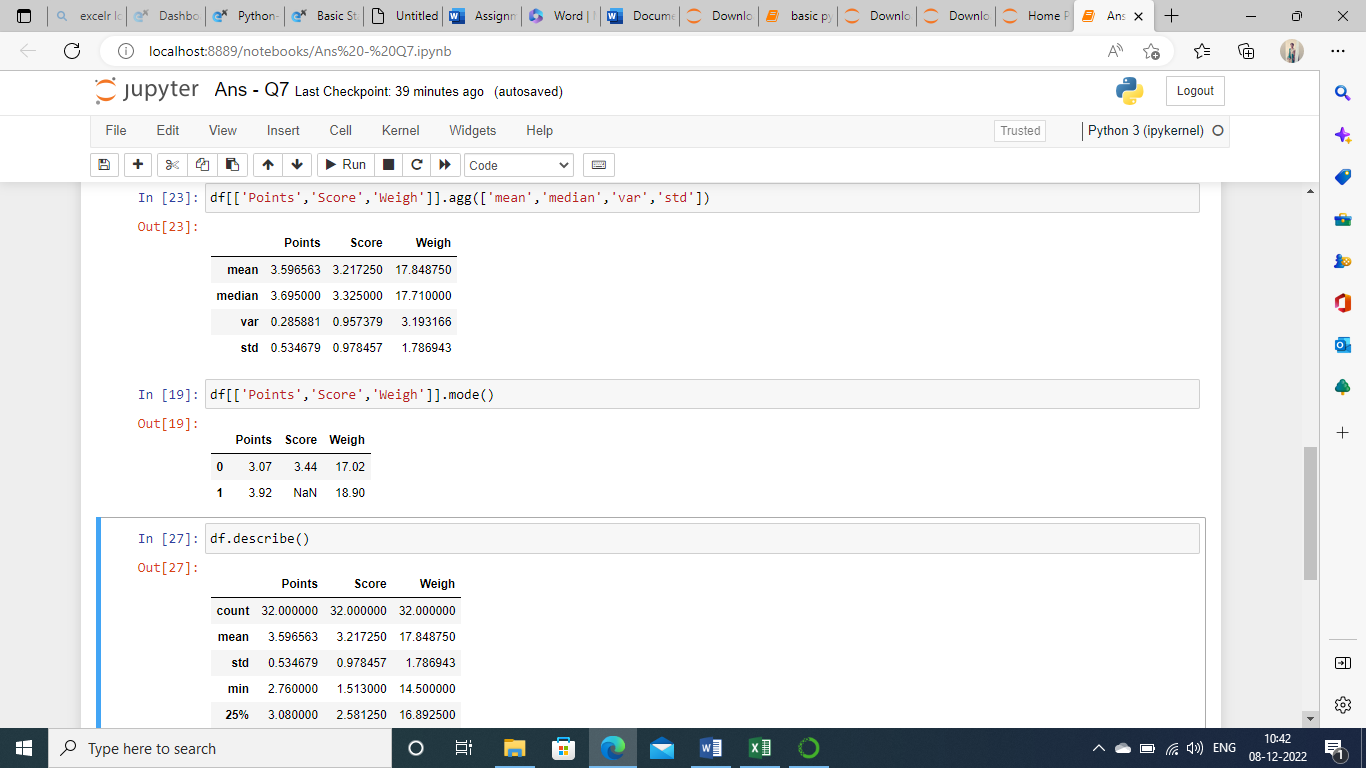
**Mode=**17.02

**Variance=**3.19

**Standard deviation=**1.78

**Range =8.399**





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?

Ans:- 108+110+123+134+135+145+167+187+199=1308

1308/9=145.33

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

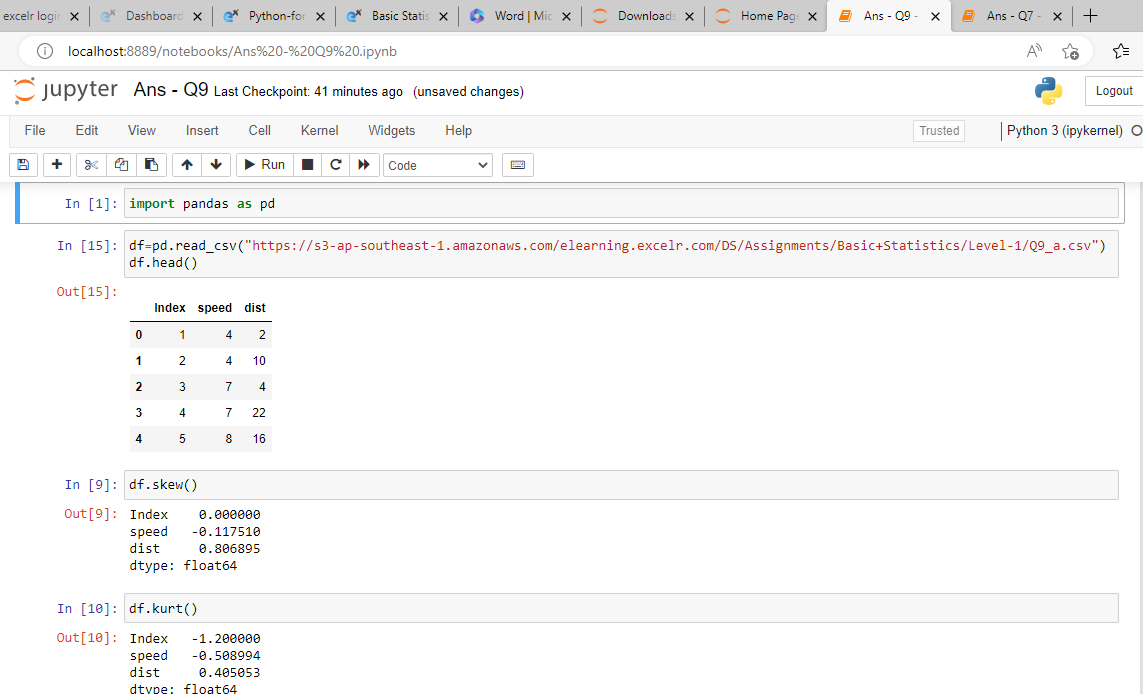
**Cars speed and distance**

**Use Q9\_a.csv**

**SP and Weight(WT)**

**Use Q9\_b.csv**

**Sol :-Q9\_a.csv**

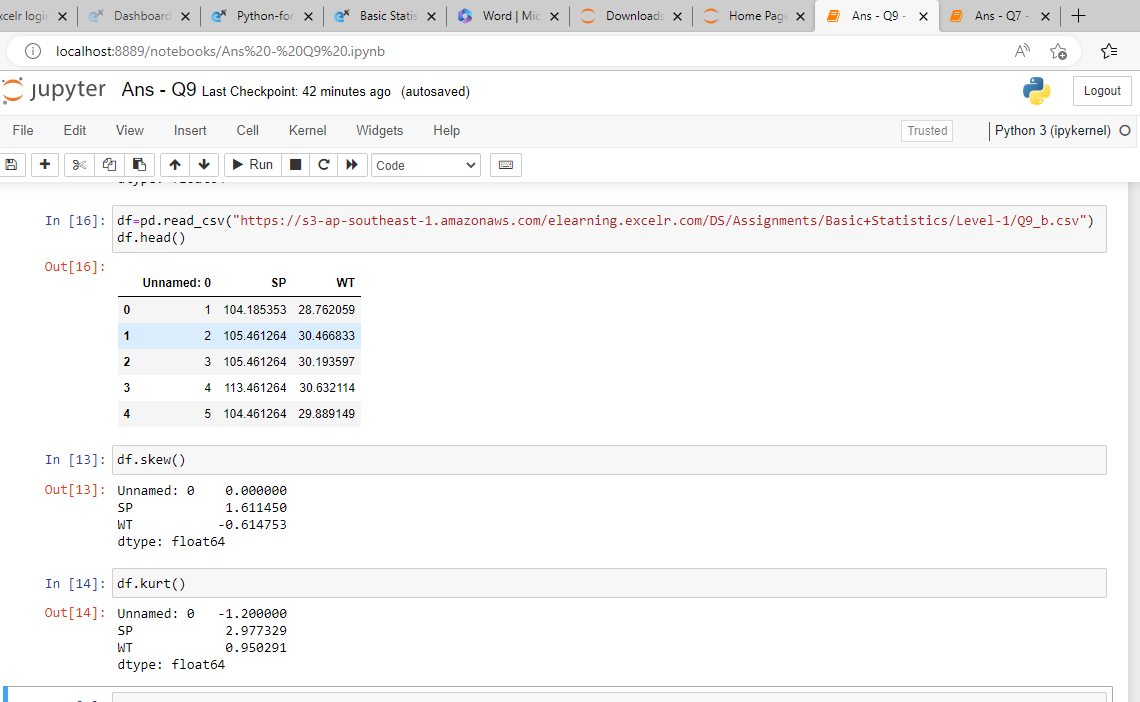
****

Skewness for speed = -0.117510 , skewness value is negative so it is left skewed . since magnitude is slightly greater than 0 it is slightly left skewed and for distance = 0.806895 , right skewed (positive) slight magnitude to right.

Kurtosis for speed = -0.508

Kurtosis for distance = 0.405

Q9\_b.csv :-



Skewness for SP = 1.611 kurtosis for SP = 2.977

Skewness for WT = -0.611 kurtosis for WT = 0.950

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



Sol:- The most of the data points are concentrated in the range 50 – 100 with frequency 200.

And least range of weight is 400 somewere around 0-10 .

So the expected value the above distribution is 75.

Skewness – we can notice a long tail towards right so it is heavily right skewed .



Sol:- Medican is less than mean right skewed and we have outlier on the upper side of the box plot and there is less data points between Q1 and 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?

Sol :- x+/-(z1-alpa\*(sigma/sqrt(n)))

Degree of freedom =2000-1=1999

Confidence interval =94%

(1-sigma/2)=1-0.03=0.97

confidence interval for 94%=1.882

confidence interval for 98%=2.33

confidence interval for 96%=2.05

**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 :-

1) Mean=41

Median=40

Variance =24.111

Standard deviation=4.910

2)we don’t have any outliers and the data is slightly skewed towards right because mean is greater than median.

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

Sol:- No skewness is present we have a perfect symmetrical distribution.

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

Sol :- skewness and tail is towards right.

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

Sol :- skewness and tail is towards left.

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

Sol:-positive kurtosis means the curve is more peaked and it is leptokurtic.

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

Sol:-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?

Sol:-the above boxplot is not normally distributed the median is towards the higher value.

What is nature of skewness of the data?

Sol:- the data is a skewed towards left .the whisker range of minimum value is greater than maximum.

What will be the IQR of the data (approximately)?   
The Interquartile Range = Q3 upper quartile –Q1 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.

Sol:- first there are no outliers for both. Second both the box plot shares the same median that is approximately in a range between 275 to 250 and they are normally distributed with zero to no skewness neither at the minimum or maximum whisker range .

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

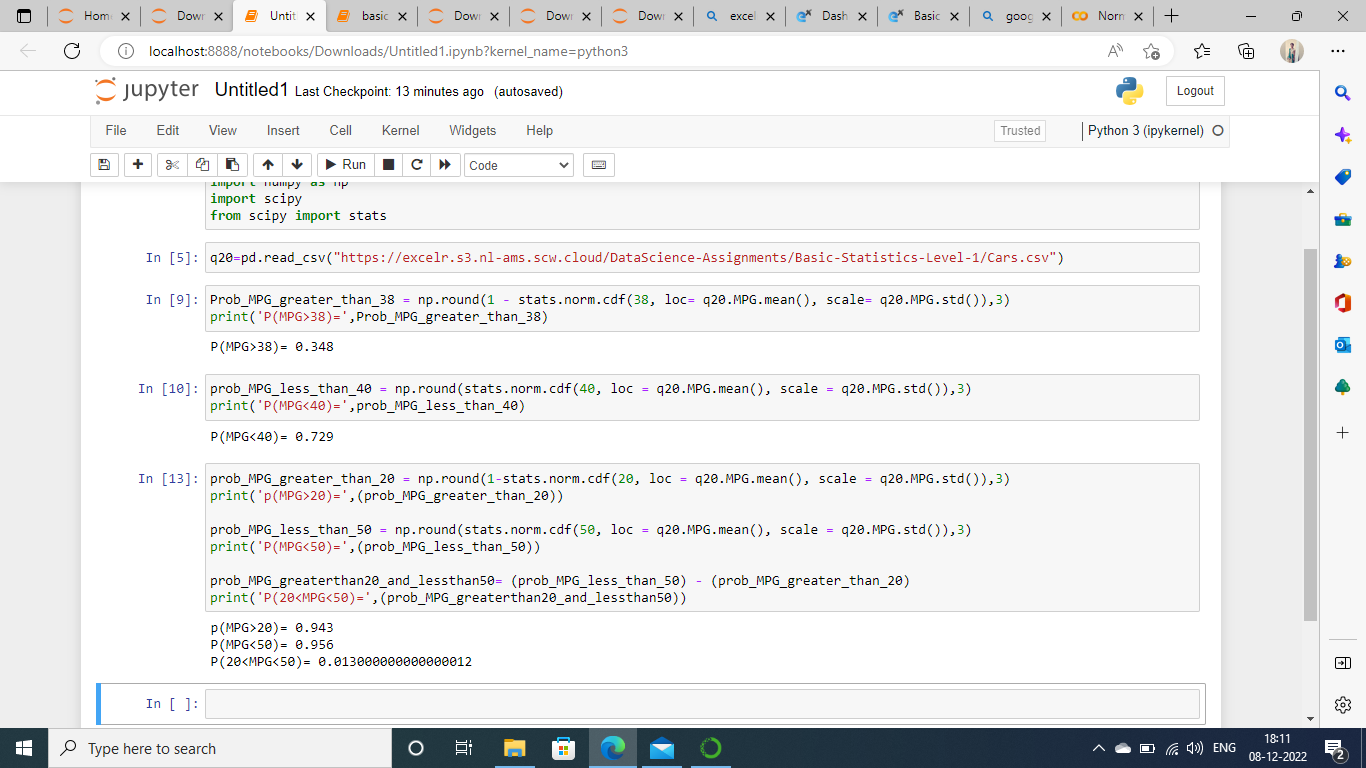
* 1. P(MPG>38)
  2. P(MPG<40)

c. P (20<MPG<50)

sol:-a) P(MPG>38)=0.348

b)P(MPG<40)=0.729

c)P(20<MPG<50)=0.0130000000

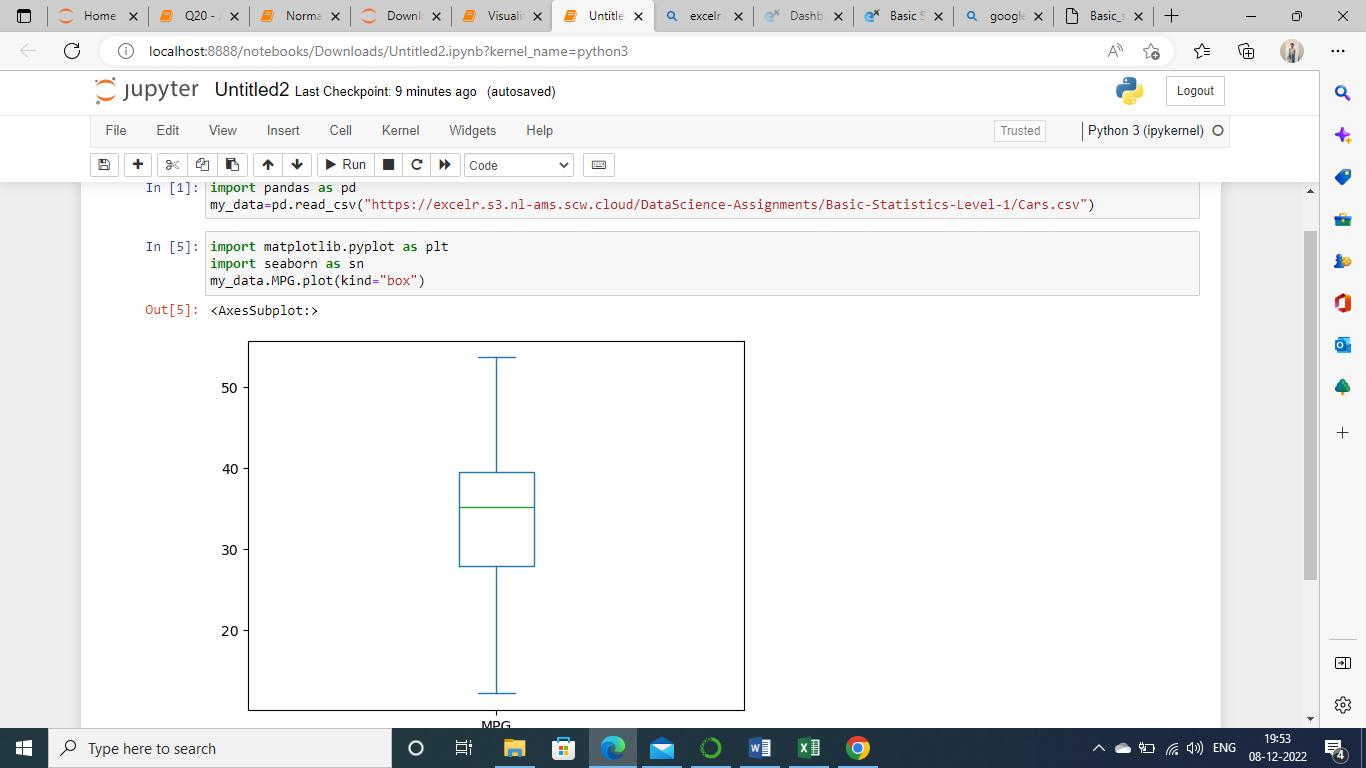


Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Sol:-MPG of cars follows normal distribution



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

Dataset: wc-at.csv

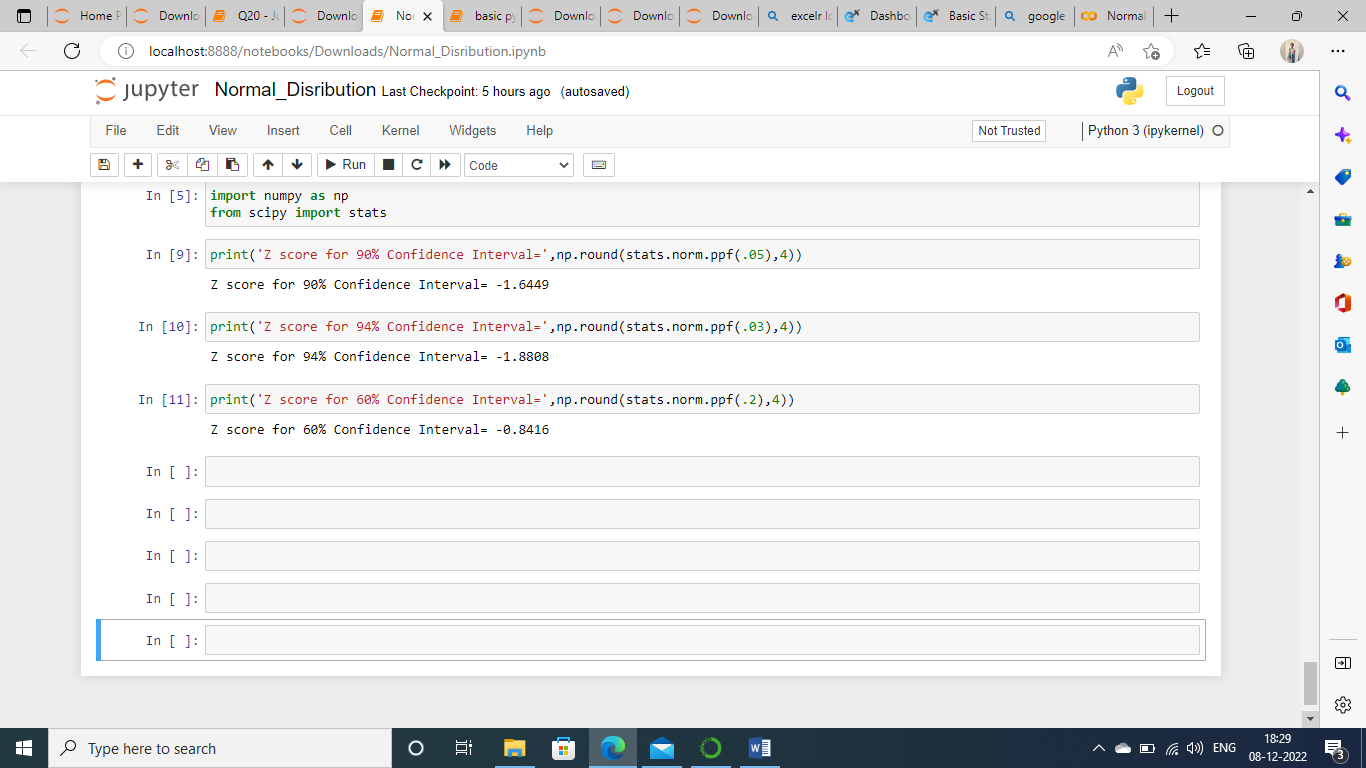
Sol:-Adipose tissue (AT) and Waist does not follow Normal distribution.

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

Sol:-Z score for 90% confidence interval = -1.6449

Z score for 94% confidence interval = -1.8808

Z score for 60% confidence interval = -0.8416

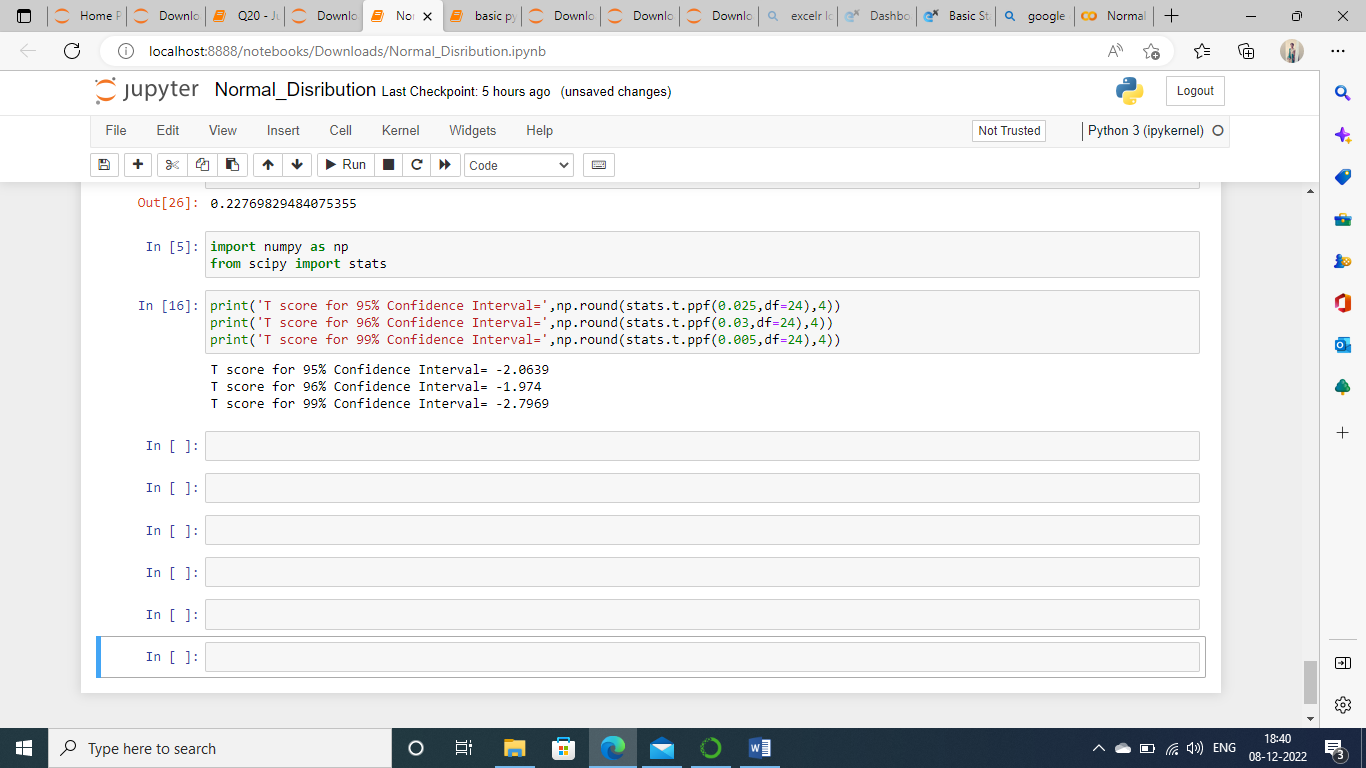


Q 23) Calculate the t scores of 95% confidence interval, 96% confidence interval, 99% confidence interval for sample size of 25

Sol:- t score for 95% confidenc interval = -2.0639

t score for 96% confidence interval = -1.974

t score for 99% confidence interval = -2.7969



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

sol:-

