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

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

**Solution:-** No of outcomes = 8 (HHH, TTT, TTH, HHT, THT, THH, HTH, THT)

No of favorable outcomes =3(HHT, THH, HTH)

P = No of favorable outcomes

No of outcomes

= 3/8 = 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

**Solution:-** a)No of outcomes = 36

No of favorable outcomes = 0

P=0

b) No of outcomes = 36

No of favorable outcomes = 6

P= 6/36 = 16.6%

c) No of outcomes = 36

No of favorable outcomes = 8

P = 2/36 = 16.66%

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**:- No of outcomes = 21

No of favorable outcomes = 10 , P=10/21 = 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

**Solution**:- P = Favorable outcomes

Total outcomes

Favorable outcomes = P \* Total outcomes

A= 1\*0.015= 0.015

B= 4\*0.20 = 0.8

C= 3\*0.65 = 1.95

D= 5\*0.005= 0.025

E= 6\*0.01 = 0.06

F= 2\*0.120 = 0.24

Expected number of candies

for a randomly selected child =(0.015+0.8+1.95+0.025+0.06+0.24)

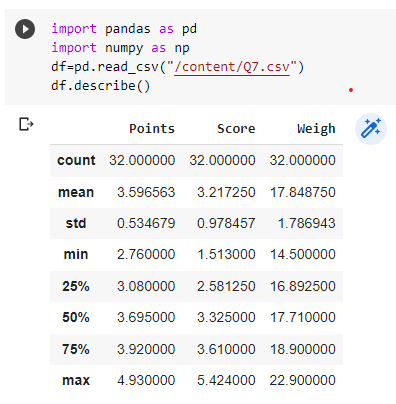
= 3.09

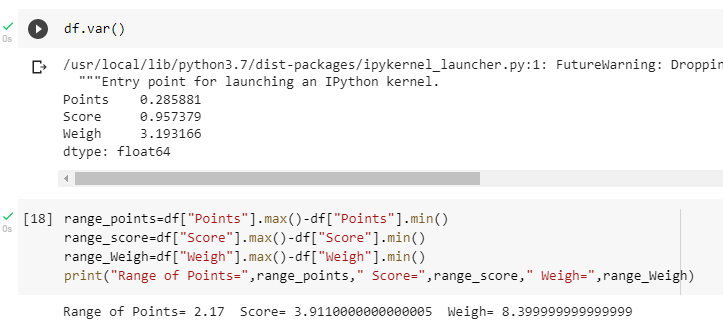
Q7) Calculate Mean, Median, Mode, Variance, Standard Deviation, Range & comment about the values / draw inferences, for the given dataset

* For Points,Score,Weigh>

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

**Use Q7.csv file**

****

****

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?

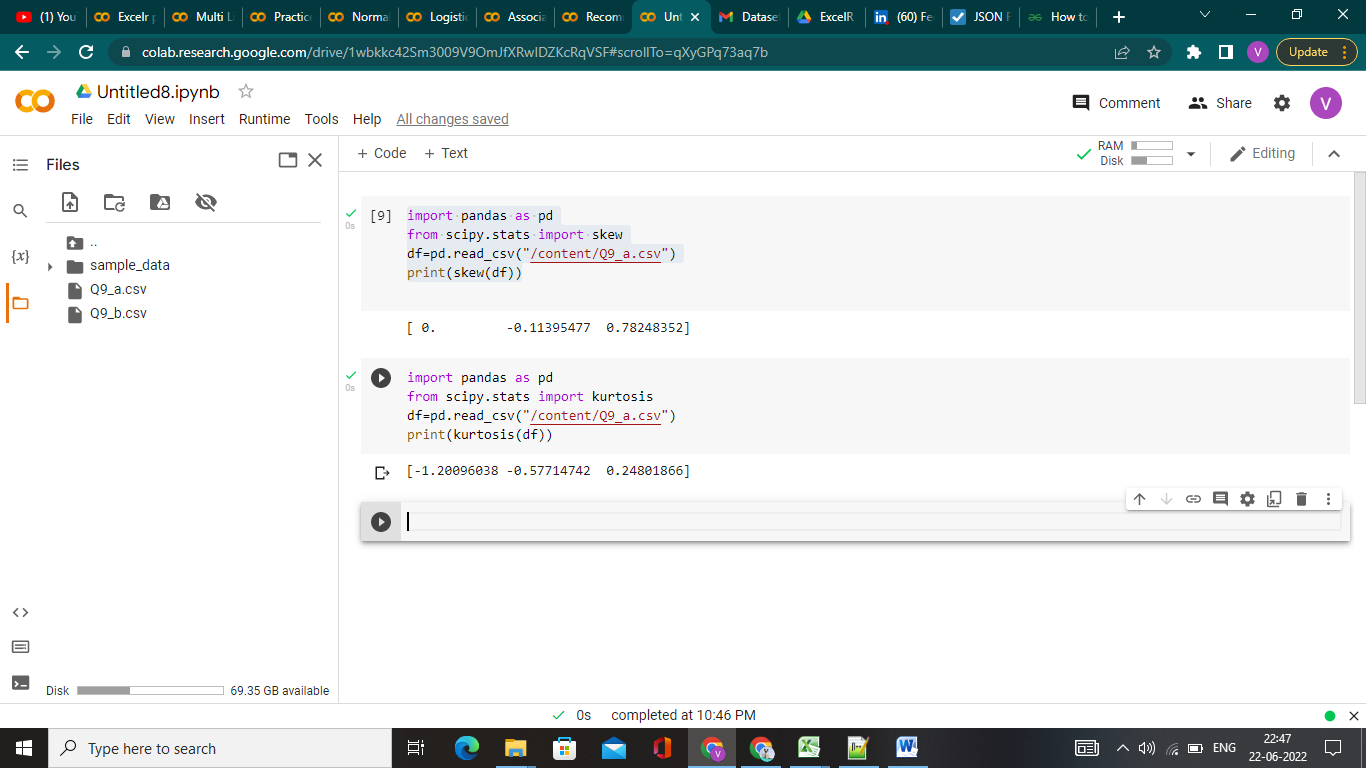
**Solution:** 108+110+123+134+135+145+167+187+199 = 145.33

9

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

**Cars speed and distance**

**Use Q9\_a.csv**

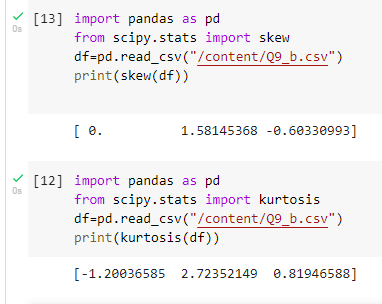
****

**Skewness:** Skewness of speed is nearly zero that means distribution of speed is symmetrical but skewness of distance is positively skewed hence it is asymmetrical.

**Kurtosis:** Speed is negative kurtosis and distance is positive kurtosis. Speed has more spread data and heavy tails and distance is less spread.

**SP and Weight(WT)**

**Use Q9\_b.csv**

****

**Skewness:**

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



Histogram:-

1. Data is positively skewed.
2. Most of the data is between 50 to 150.

Boxplot:-

1. The data is right skewed.
2. There are outlier.

**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?

**Solution:**  Sample mean=200 , Standard deviation=30, Sample=2000

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| CI 94% = | |  |  | | --- | --- | | 200 ± 1.8808× | 30 | |  | | √2000 | |
| = | 200 ± 1.262 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| CI 98%= | |  |  | | --- | --- | | 200 ± 2.3263× | 30 | |  | | √2000 | |
| = | 200 ± 1.561 |

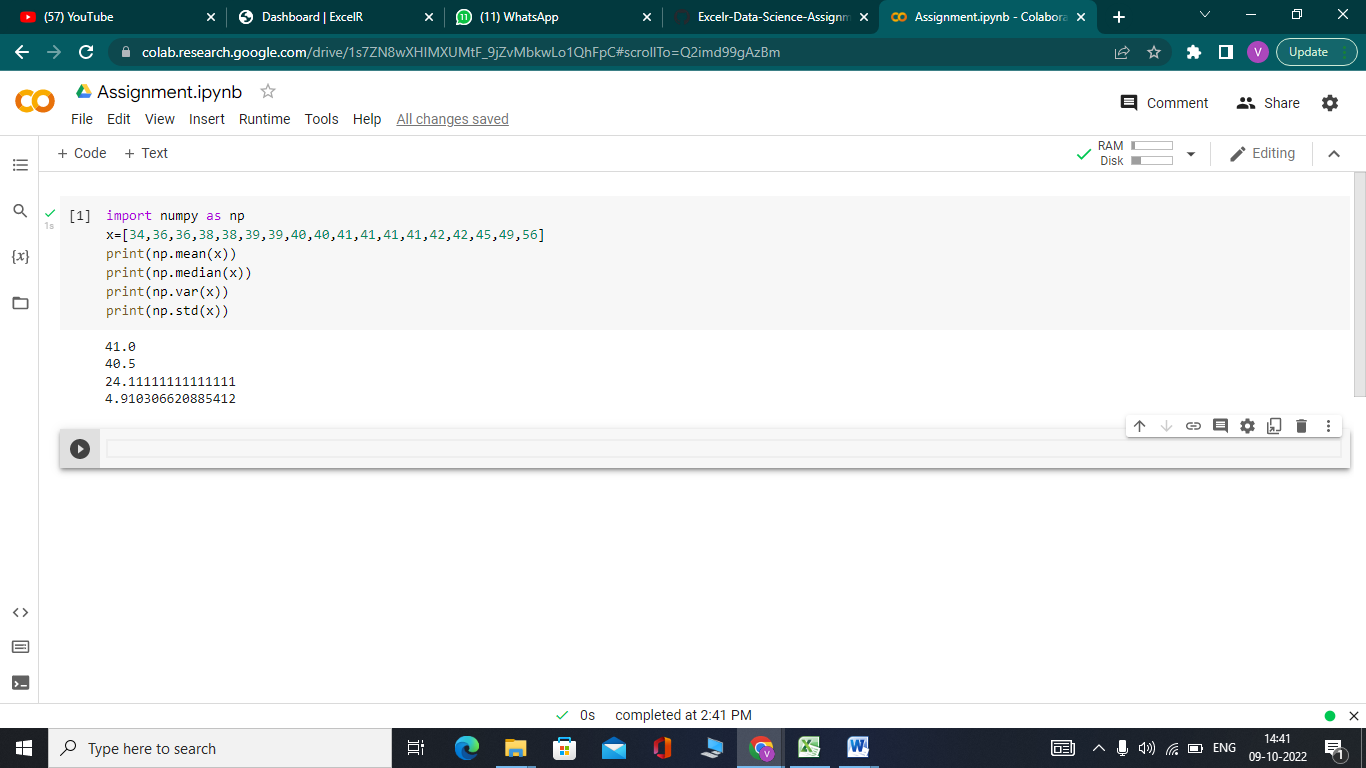
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| CI 96%== | |  |  | | --- | --- | | 200 ± 2.0537× | 30 | |  | | √2000 | |
| = | 200 ± 1.378 |

**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?

1)Solution:-



2)Answer:- There are equal number of students who are equal or above average and below average.

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

Answer:- When mean and median of given data are equal then skewness will be zero.

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

Answer:-When mean is greater than median then the nature of skewness will be positive.

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

Answer:-When median is greater than mean then the nature of skewness will be negative.

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

Answer:- Positive kurtosis indicate that a distribution is peaked and has thick tails.

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

Answer:- Negative kurtosis indicate that a distribution is flat and has thin tails.

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



What can we say about the distribution of the data?

What is nature of skewness of the data?

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

Answer 1:- Data is irregularly distributed hence it is not normal distribution.

Answer 2:-Nature of skewness is negative.

Answer 3:-IQR= 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.

Answer:- Boxplot 2 has more samples than boxplot 1 but in both samples data is centered at same level at somewhere between 275-250.

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.

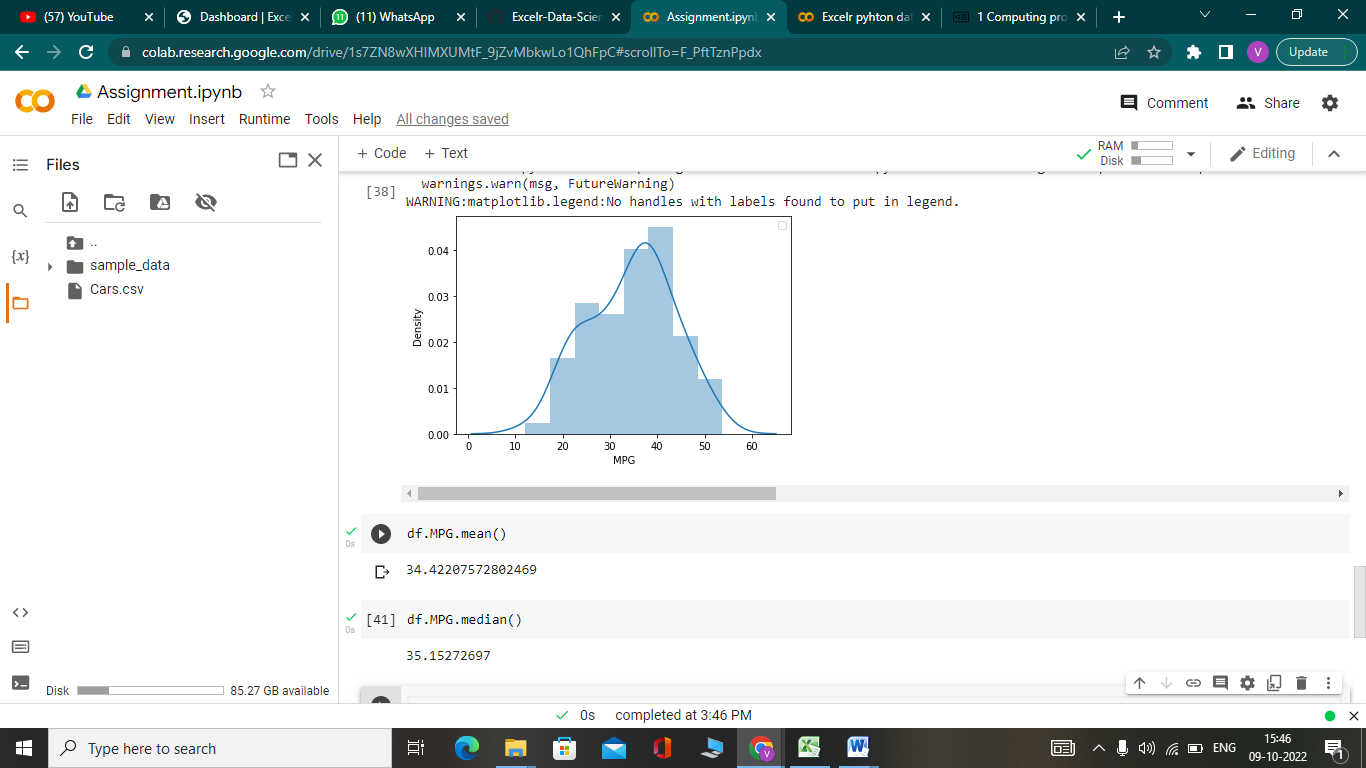
MPG <- Cars$MPG

* 1. P(MPG>38)
  2. P(MPG<40)
  3. P (20<MPG<50)

Answer:= 

Q 21) Check whether the data follows normal distribution

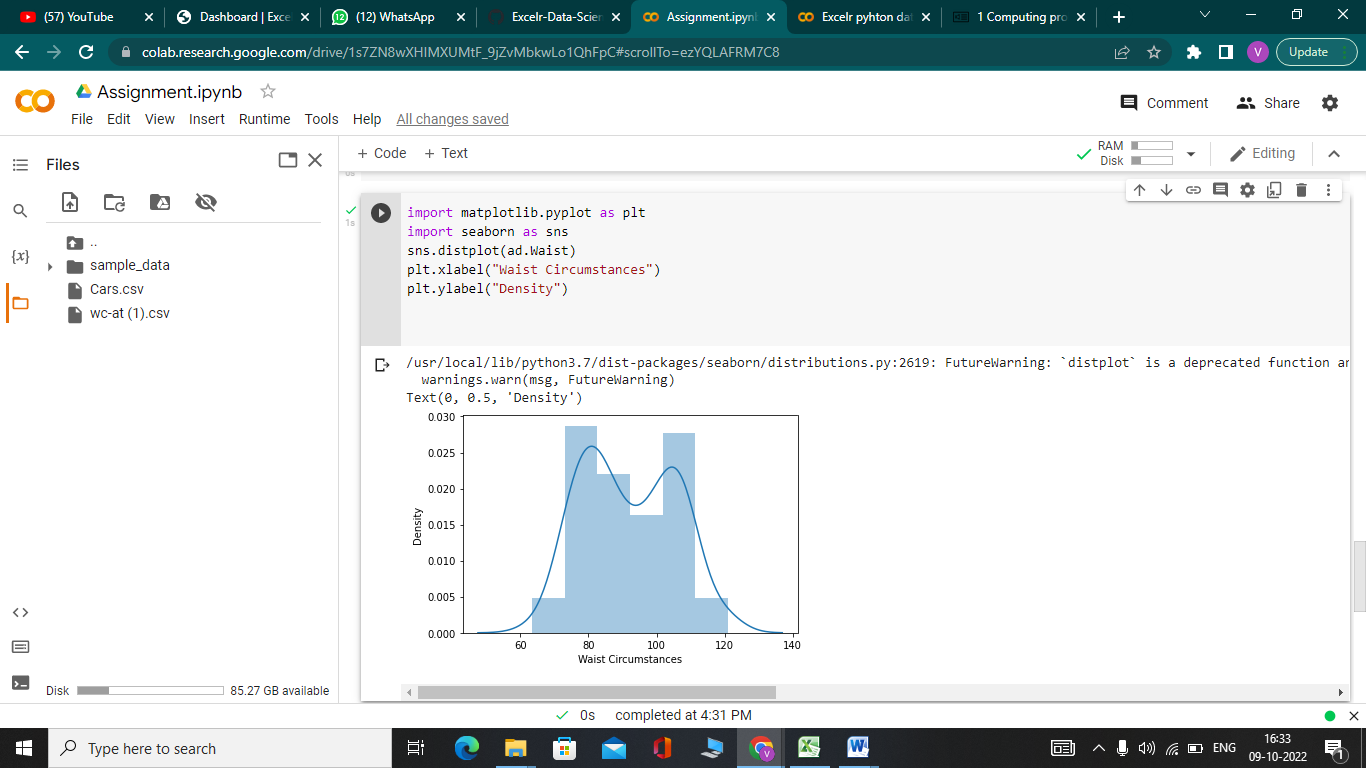
1. Check whether the MPG of Cars follows Normal Distribution

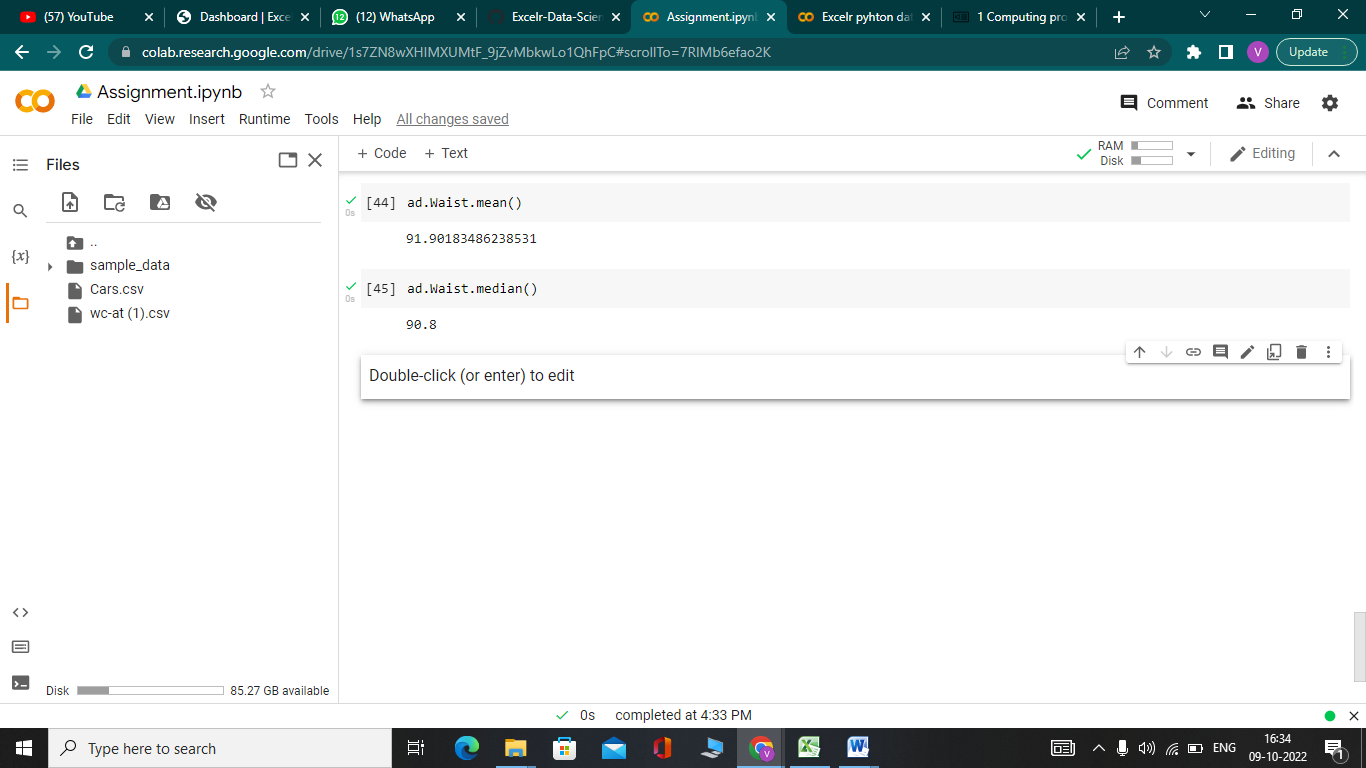


As Mean and median of the data is approximately similar then we can say it is a normal distribution.

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

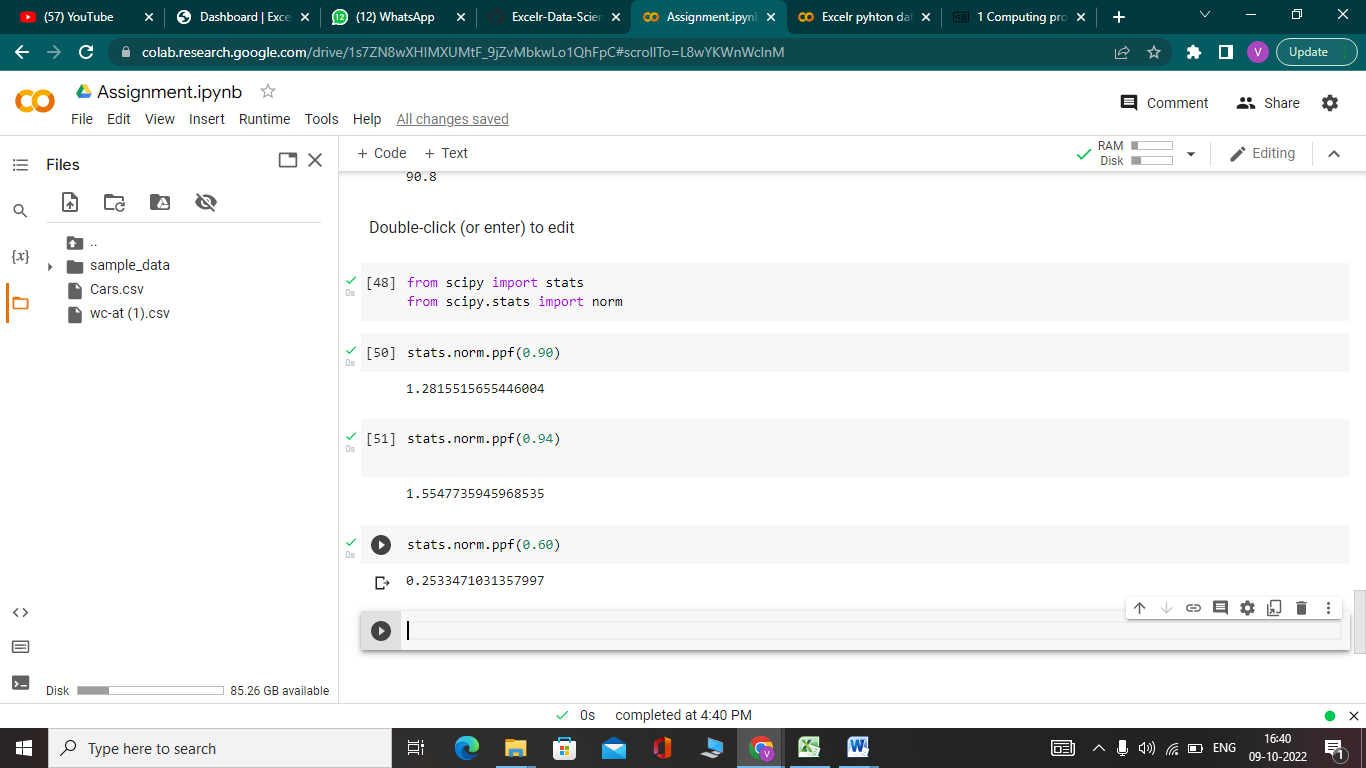
Dataset: wc-at.csv





As Mean and median of the data is approximately similar then we can say it is a normal distribution

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

Answer:- Degree of freedom is N-1. Therefore, the degree of freedom here will be 25-1=24.

Now, by looking at the t-table we can find the t-score:  
1. 95% of 24 = 2.064  
2. 96% of 24 = 2.172  
3. 99% of 24 = 2.797

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

