|  |  |
| --- | --- |
| 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?

Ans = 3/8

{HHT,HTH,THH}

1/8+1/8+1/8 = 3/8

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

Ans =

1. = There is not any outcome corresponding to 1. Probability is 0
2. = {(1,1),(1,2),(1,3),(2,1),(2,2),(3,1)}

6 outcomes

= 6/36

= 1/6

1. = {(1,5), (2,4), (3,3), (4,2), (5,1), (6,6)}

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

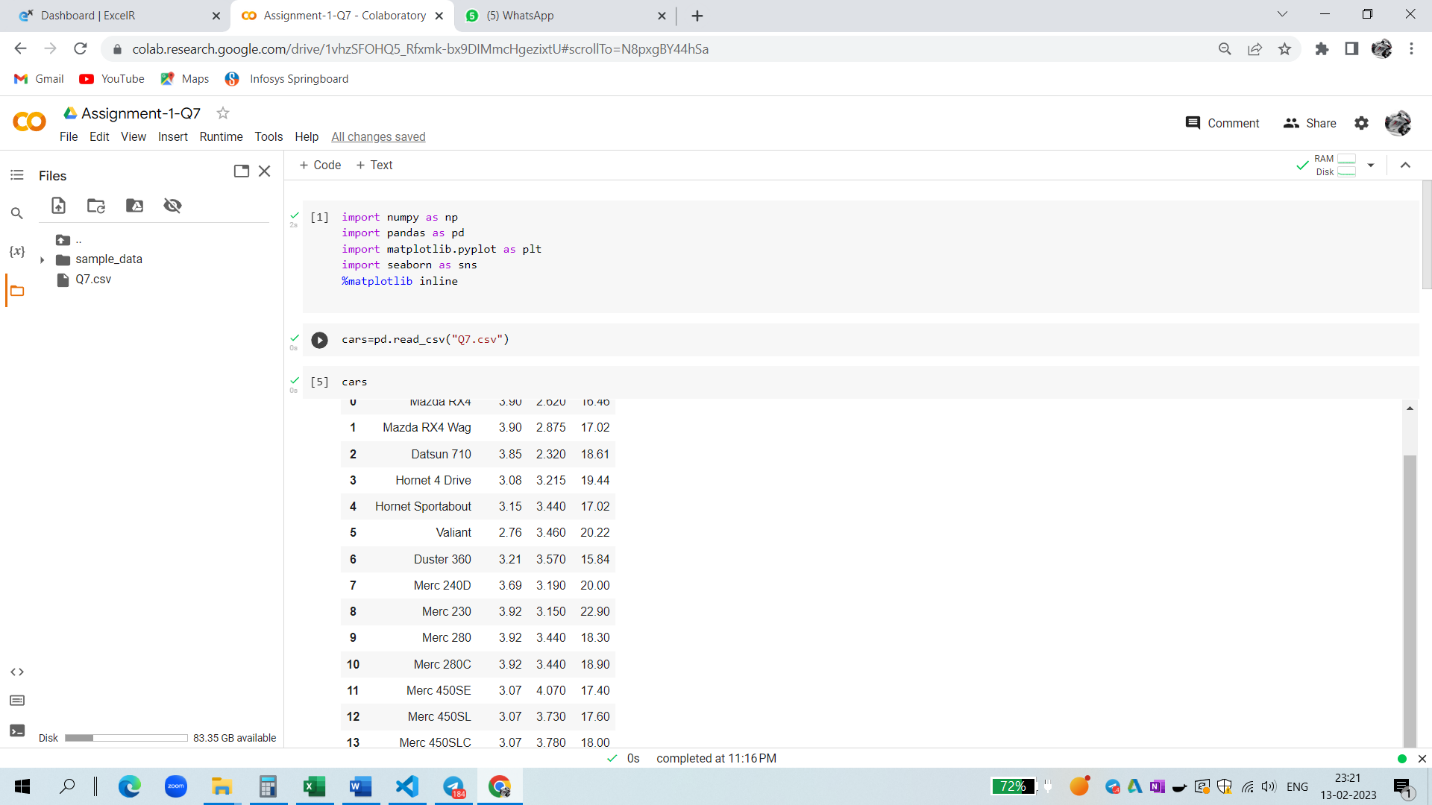
= 3.09

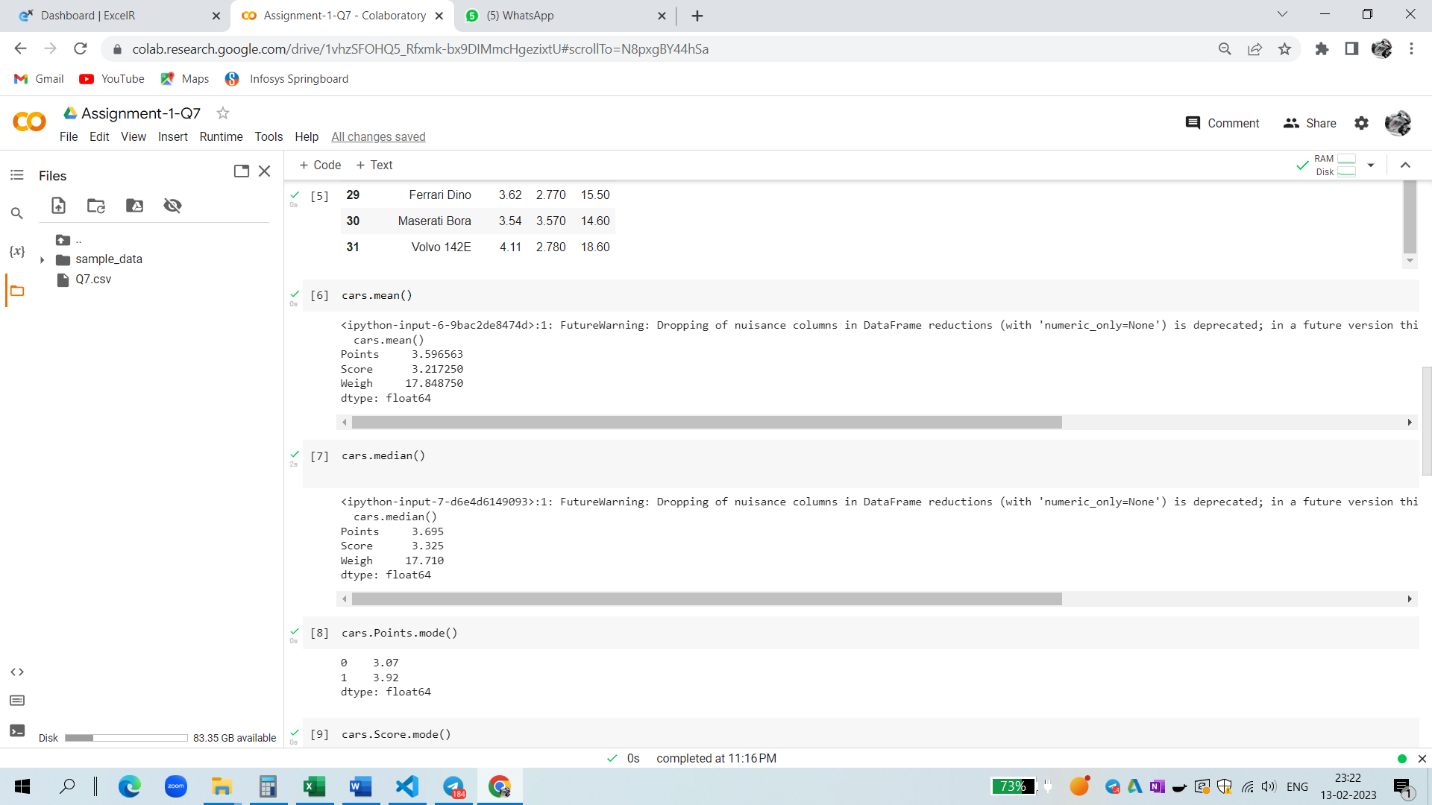
Q7) Calculate Mean, Median, Mode, Variance, Standard Deviation, Range & comment about the values / draw inferences, for the given dataset

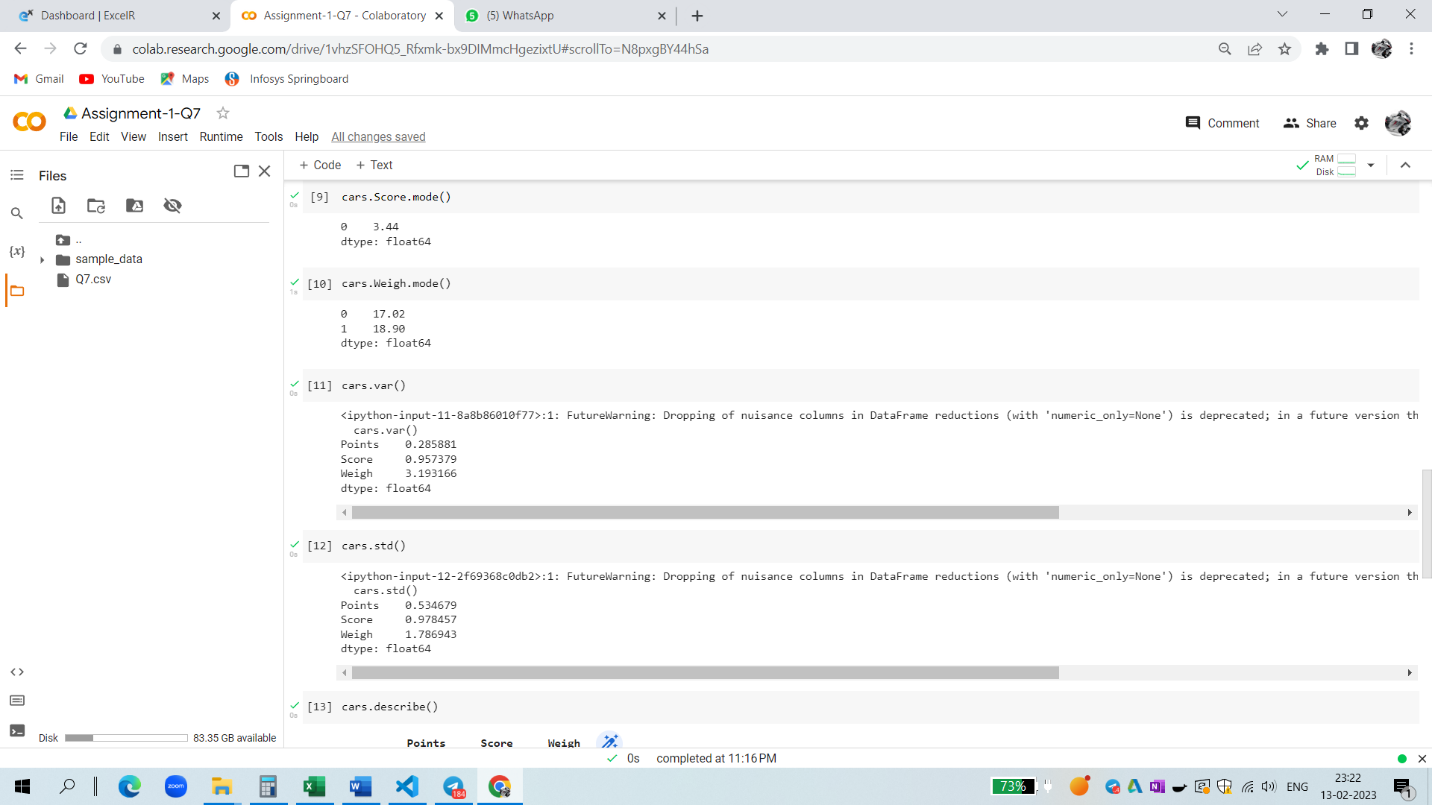
* For Points,Score,Weigh>

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?

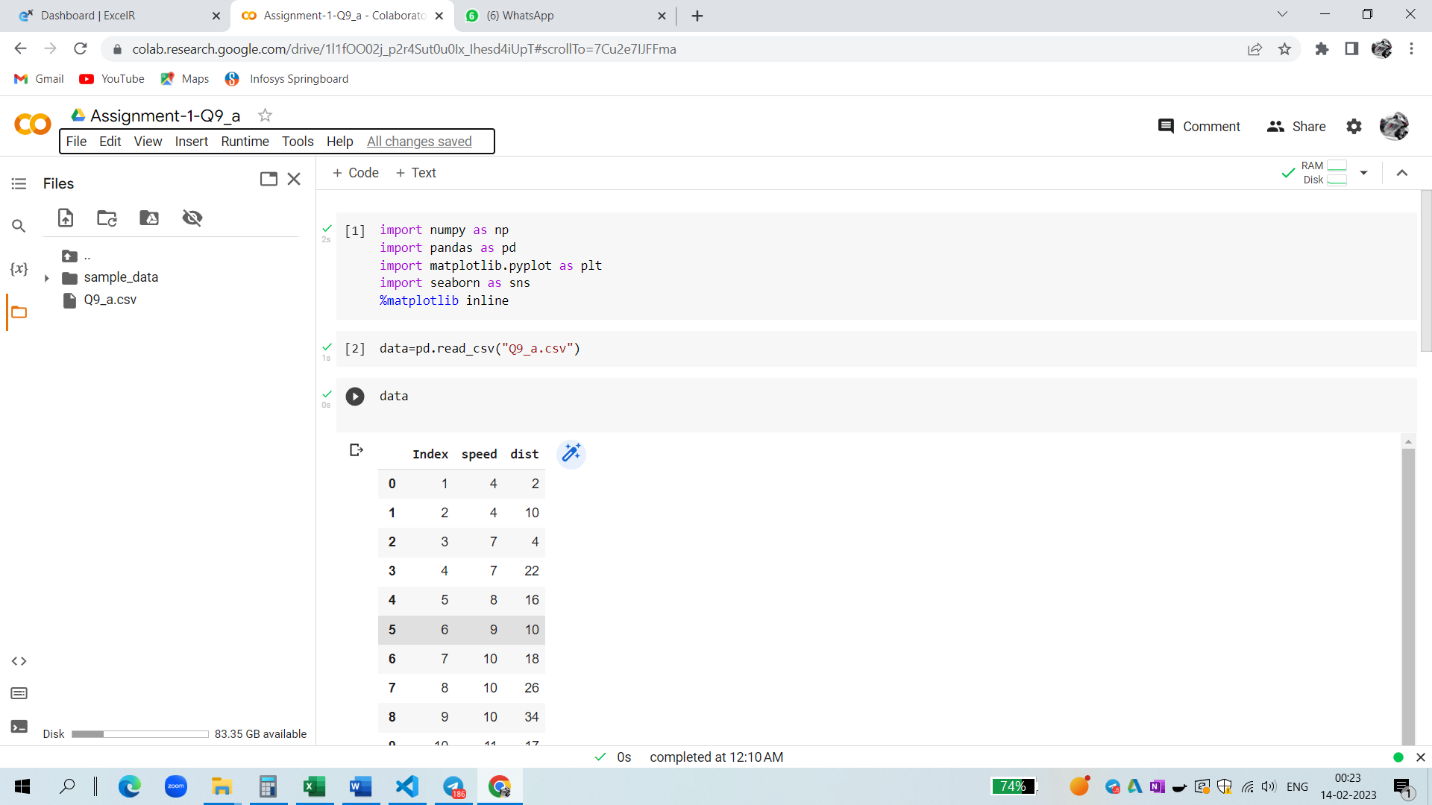
Expected value= (108+110+123+134+135+145+167+187+199)/9

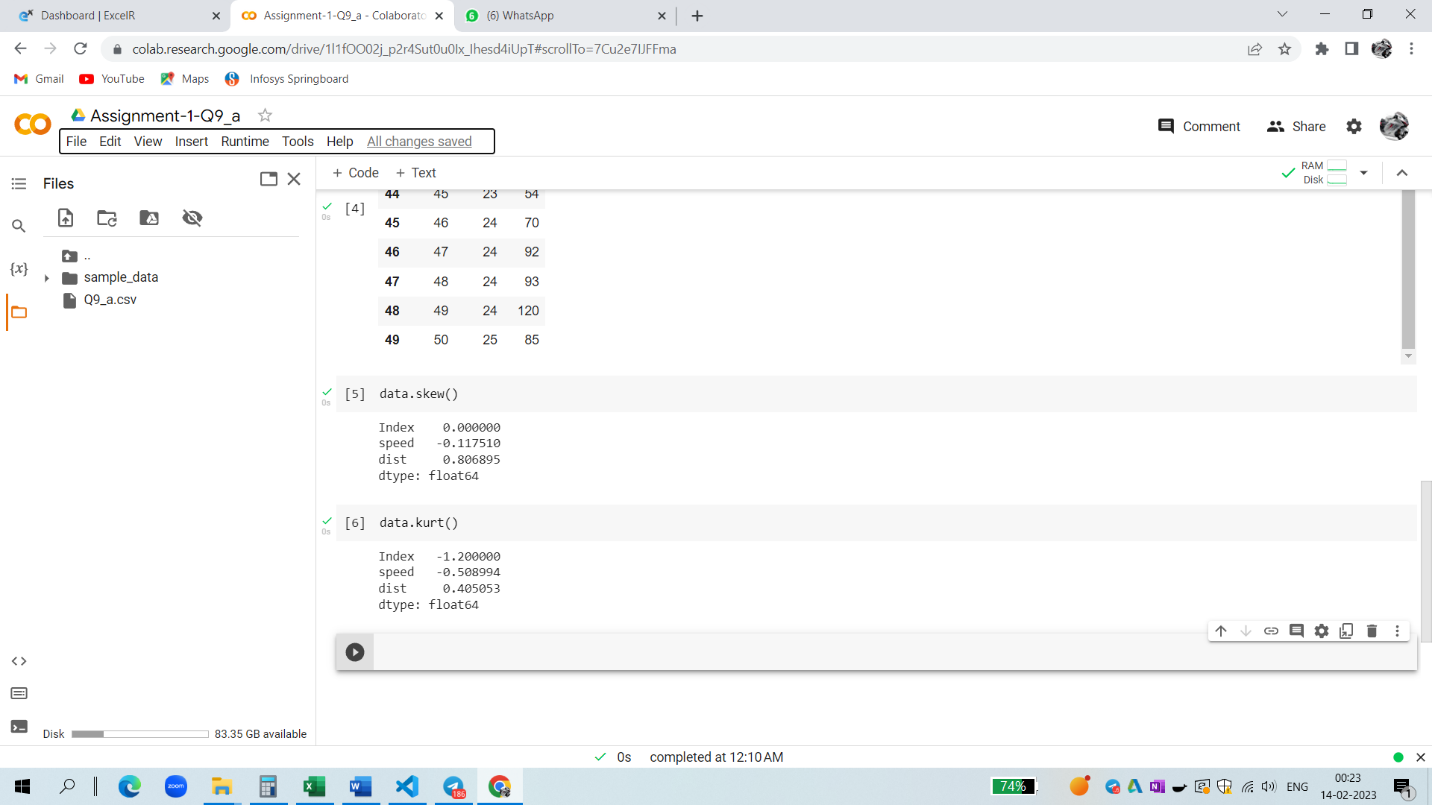
= 145.333

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

**Cars speed and distance**

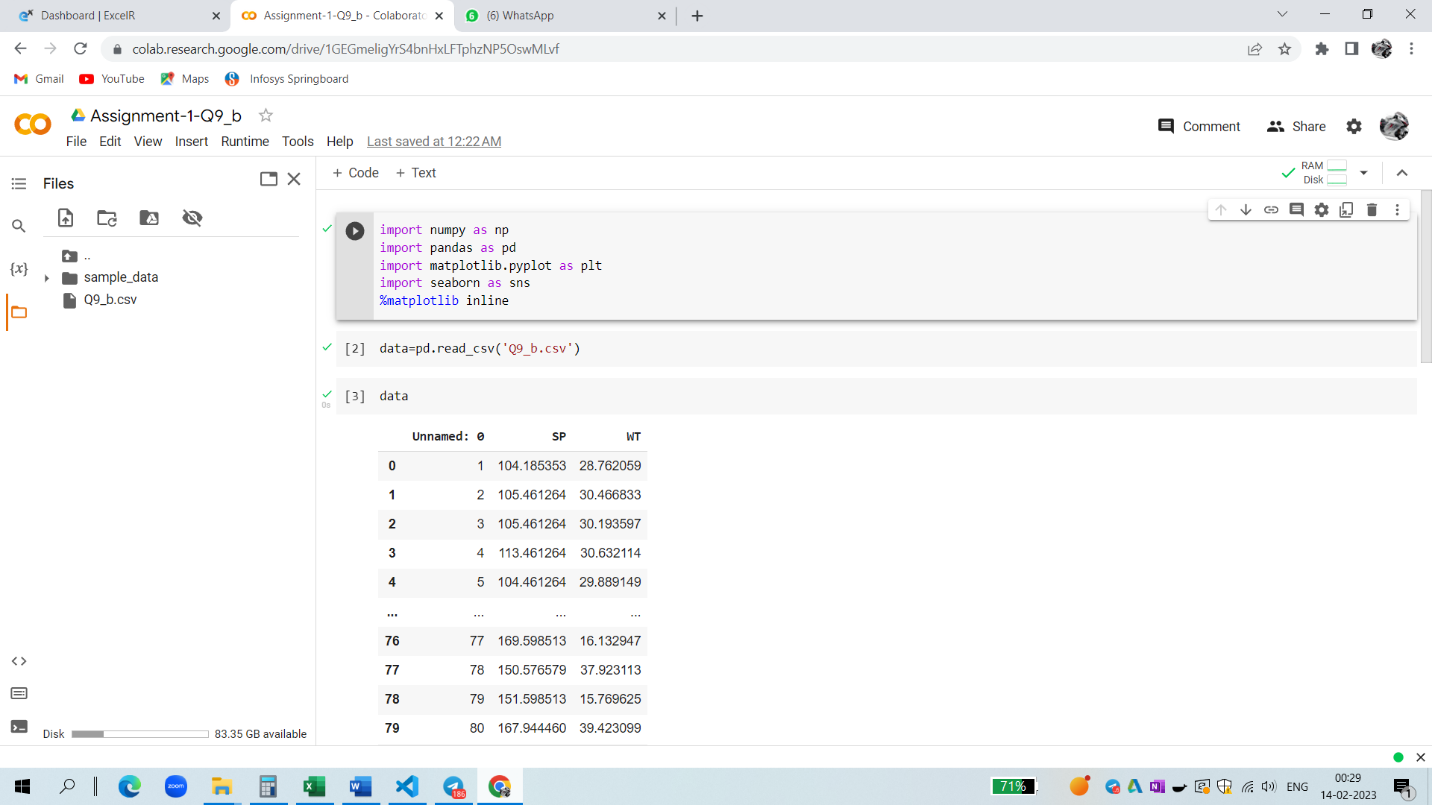
**Use Q9\_a.csv**

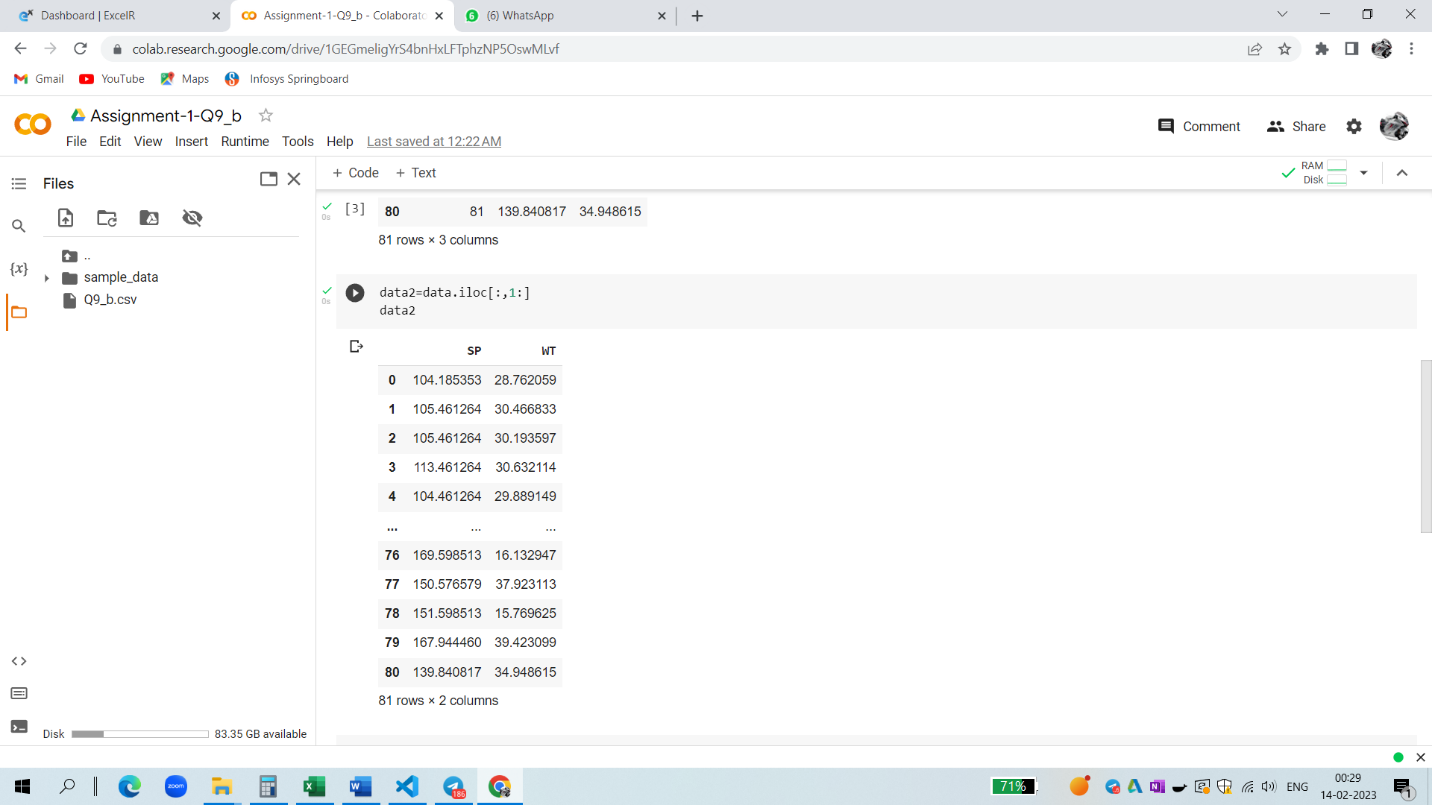
****

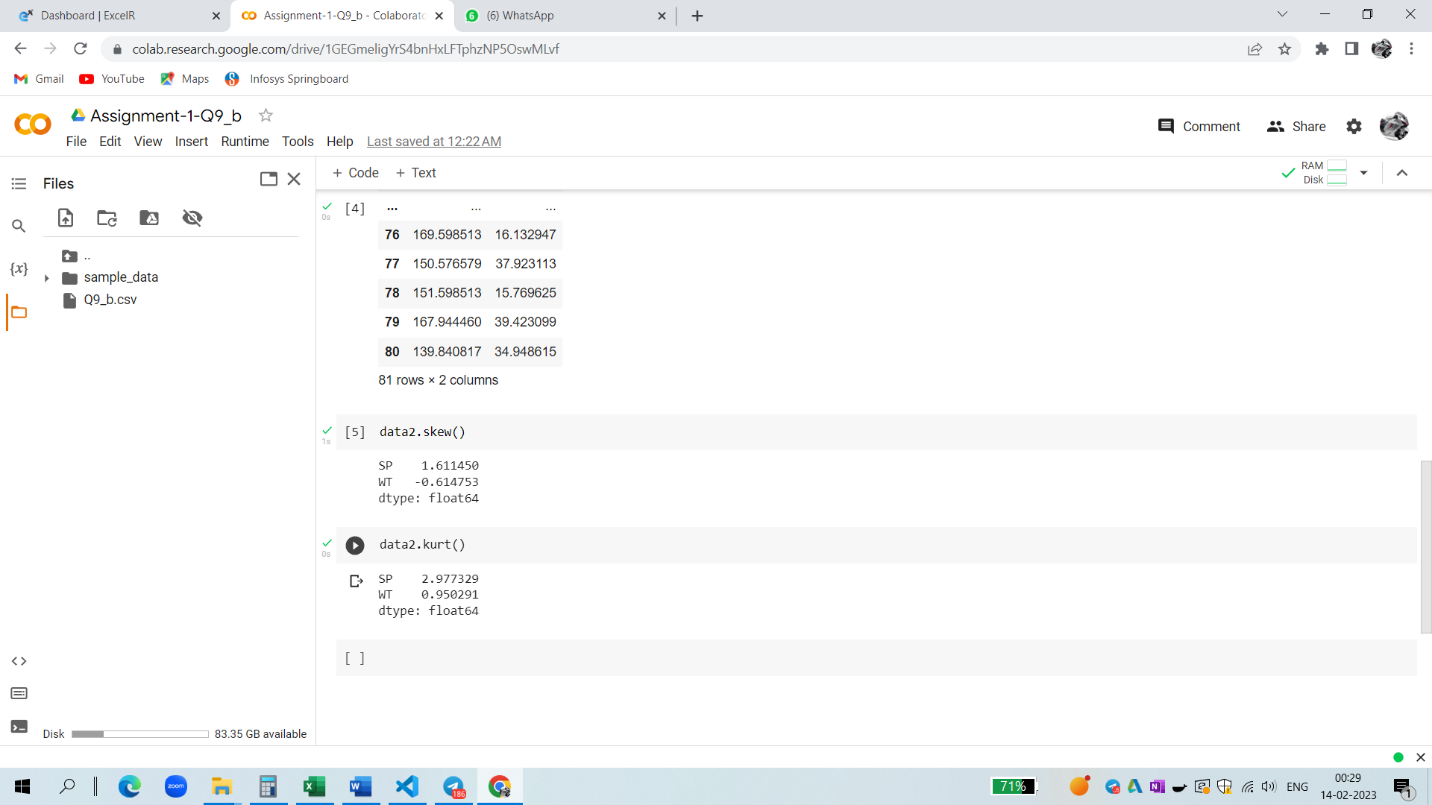
****

**SP and Weight(WT)**

**Use Q9\_b.csv**

****

****

****

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



Histogram = The most of the data points are concerated 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



Box plot = Median is less than mean right skewed and we have outlier on the upper side of 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?

Ans:

X+/-(Z1-α.σ/sqrt(n)

Degrees of freedom= 2000-1= 1999

Confidence interval= 94%

(1-σ/2)= 1-0.03) =0.97

for confidence interval for 94% is 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.

Ans=

Mean= 41

Median = 40.5

Variance = 24.111

Standard deviation = 4.910

1. What can we say about the student marks?

Ans=

We can say that the student marks have a median of 40 and a variance of 24.11

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

Ans= Symmetric

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

Ans= Right Skewness

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

Ans= Left Skewness

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

Ans= The data is normally distributed and kurtosis value is 0

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

Ans= The distribution of the data has lighter tails and a flatter-peaks than the normal distribution.

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)?

Ans – 1) It is not normally distributed the median is towards the higher value

2) the data is a skewness towards left

3)Inter quantile range = UQR (upper quantile range)-LQR(lower quantile range)

Q19) Comment on the below Boxplot visualizations?



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

Ans : 1)there is no outliers

2)both box plots median is approximately in range between 250 to 275

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

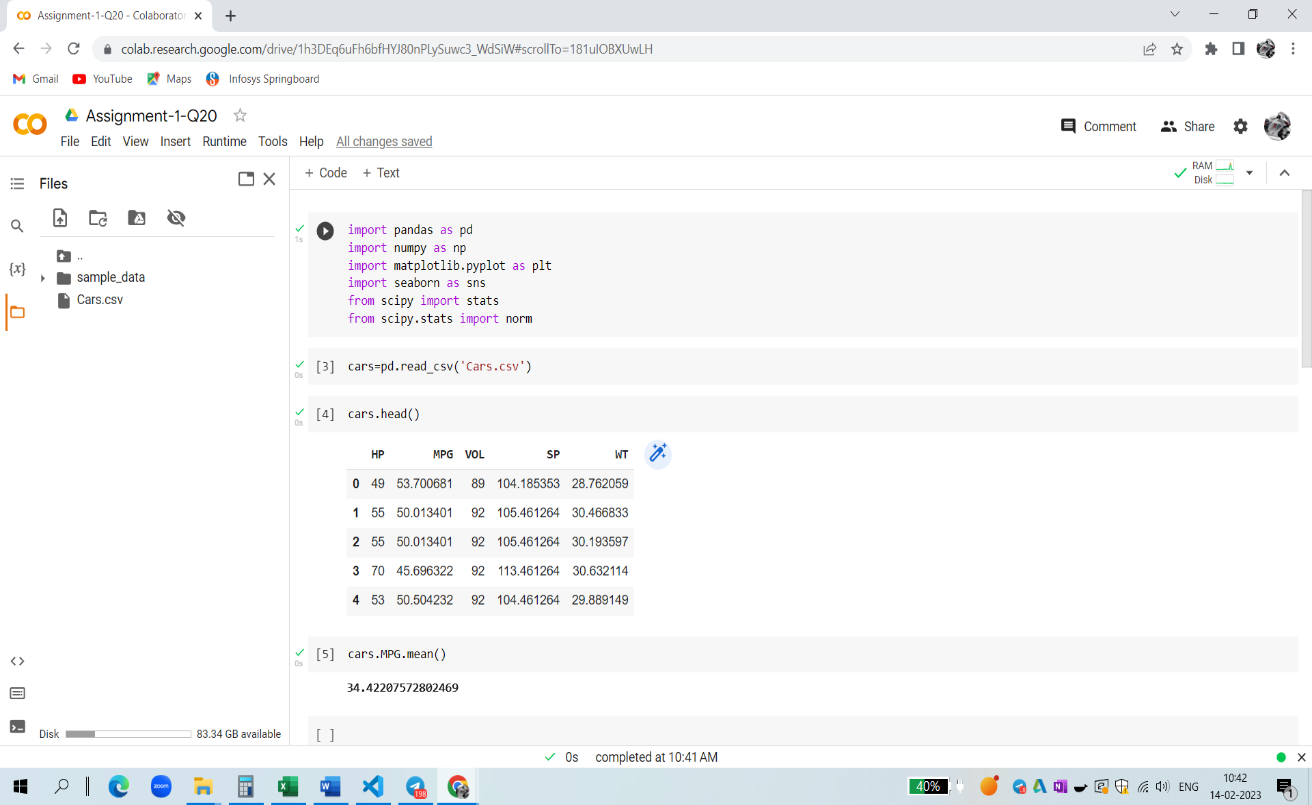
Data \_set: Cars.csv

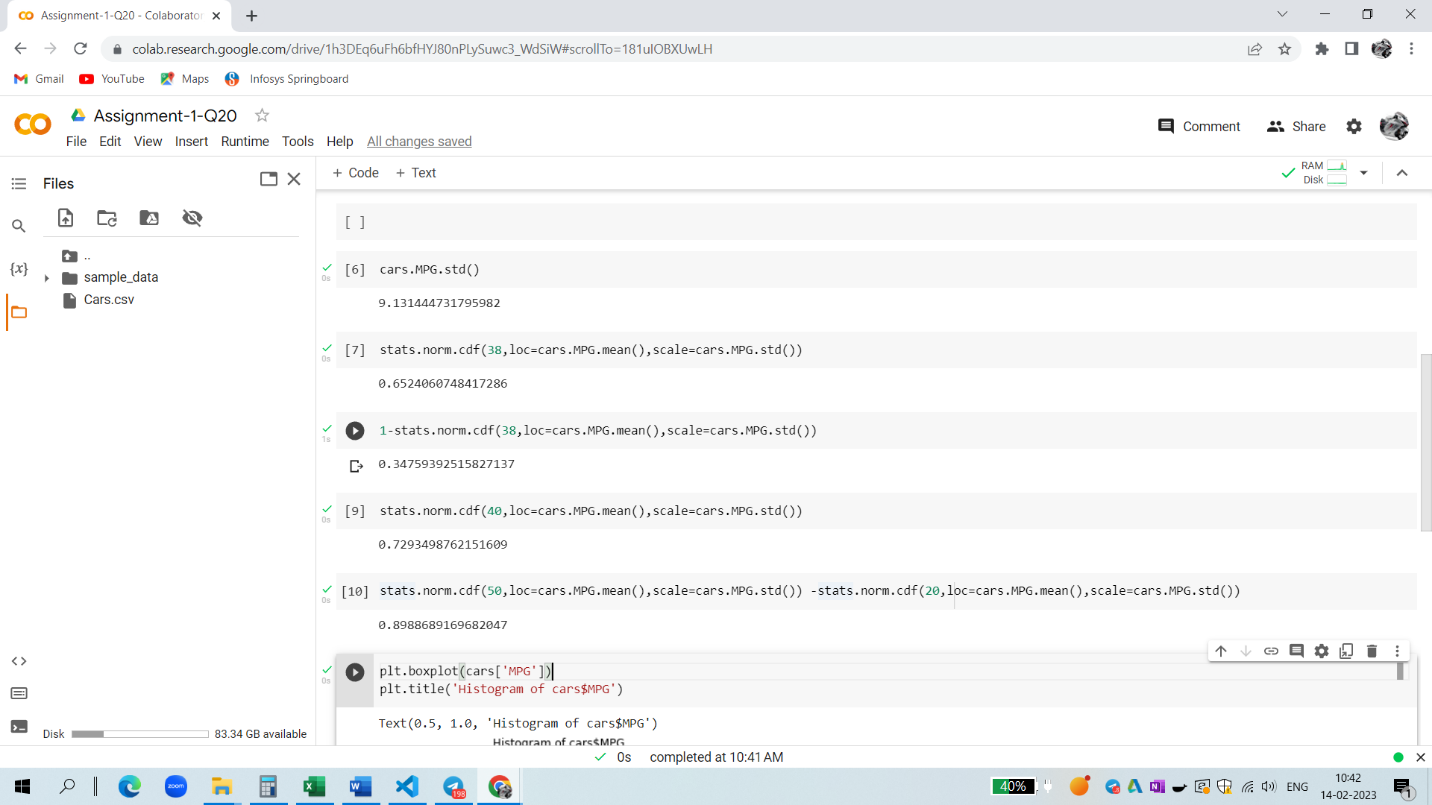
Calculate the probability of MPG of Cars for the below cases.

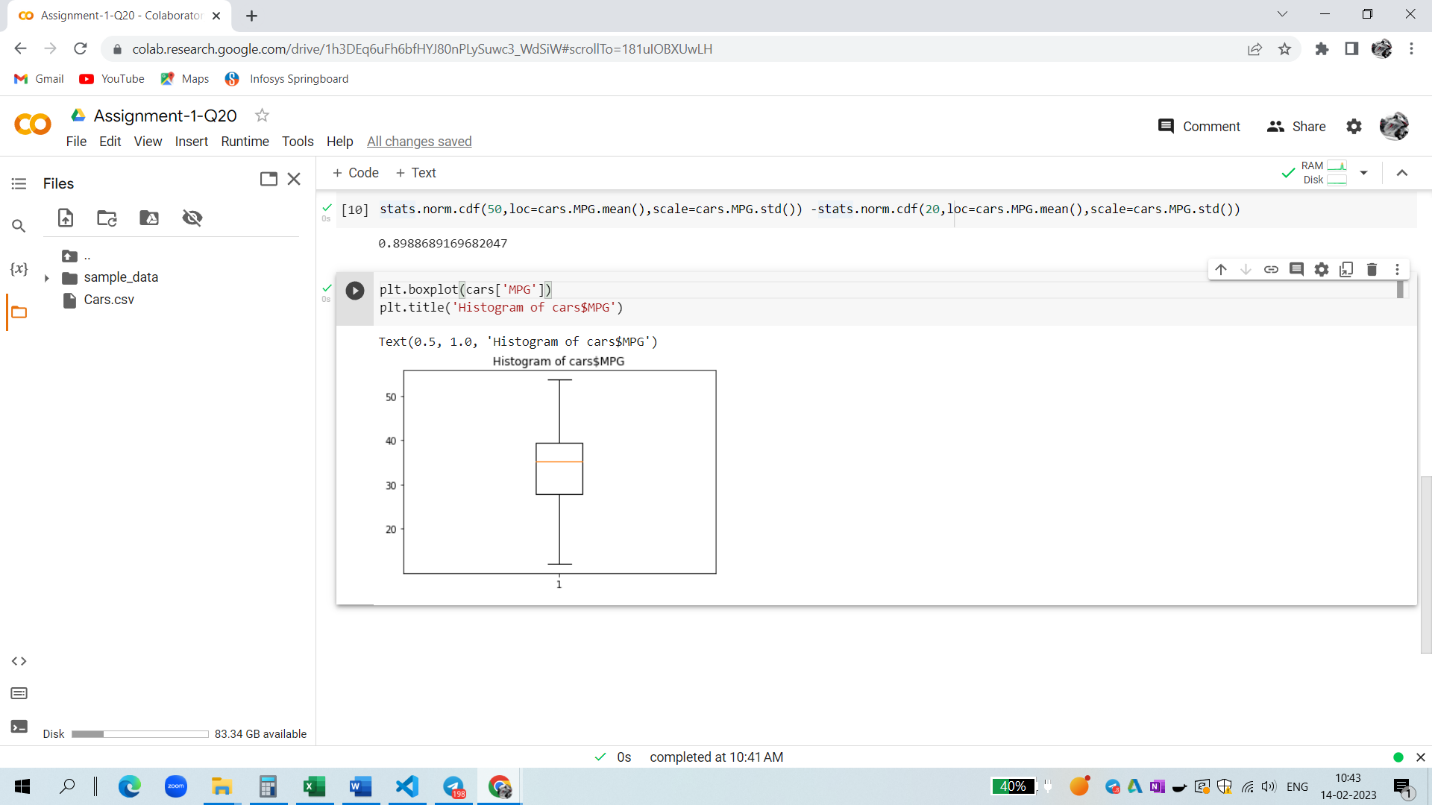
MPG <- Cars$MPG

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

c. P (20<MPG<50)



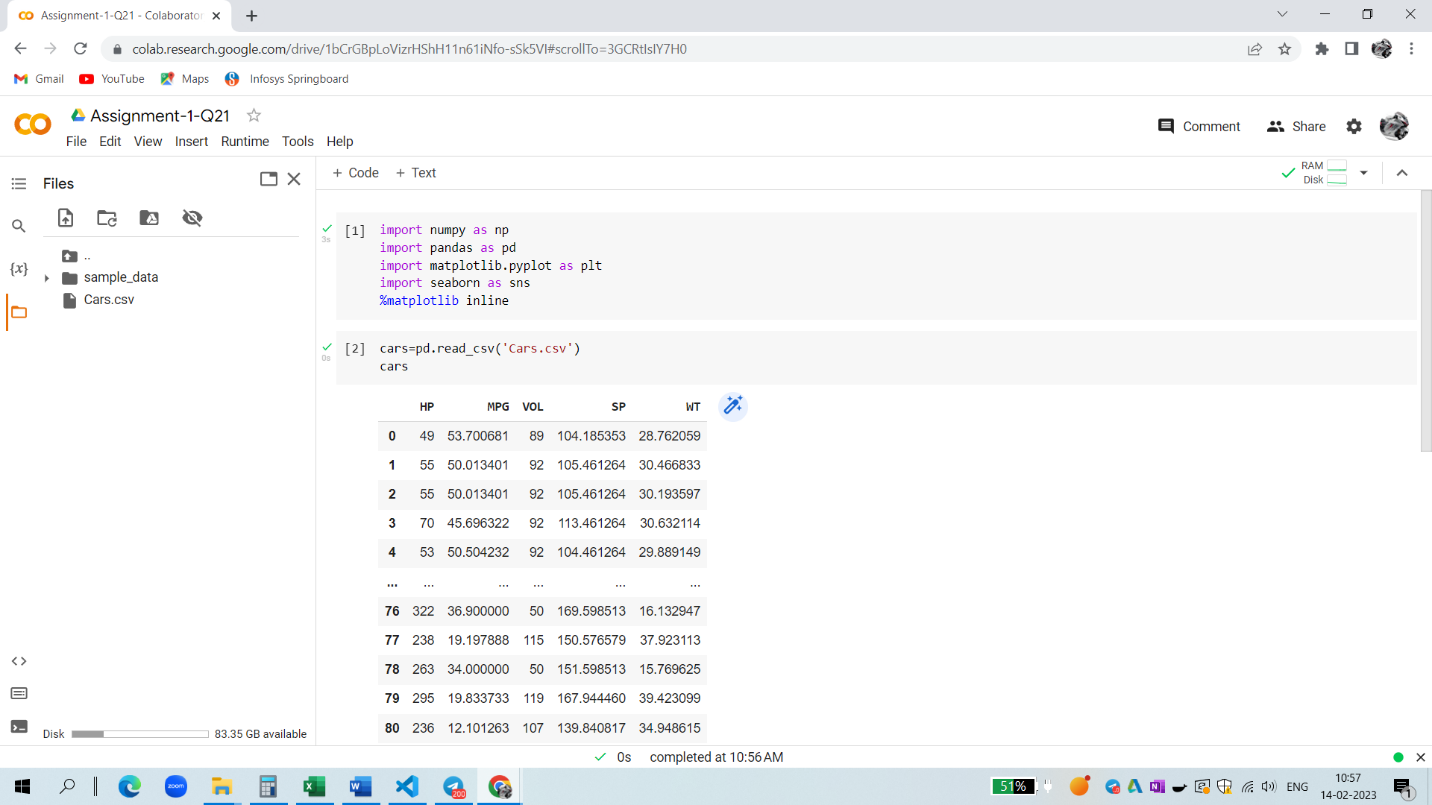


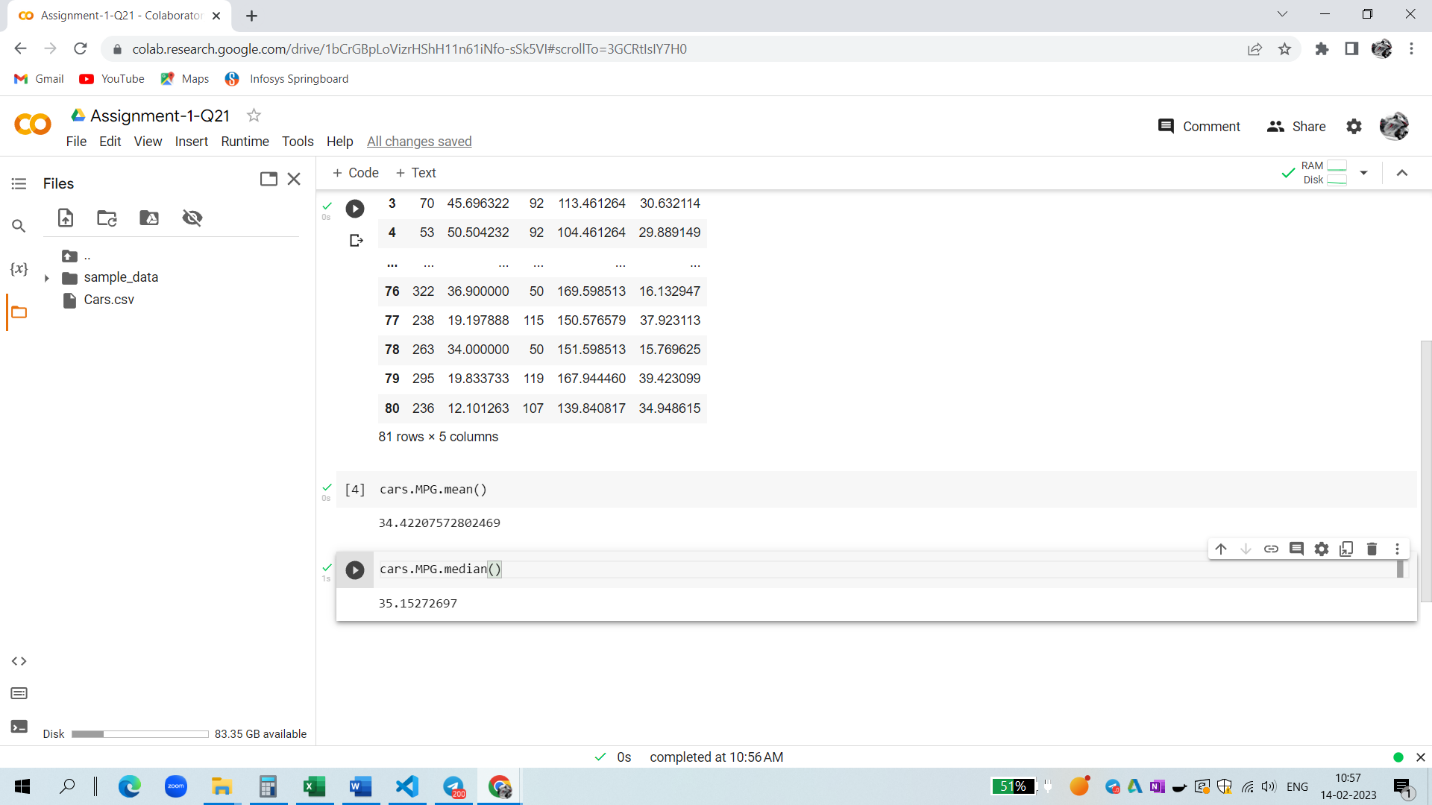


Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

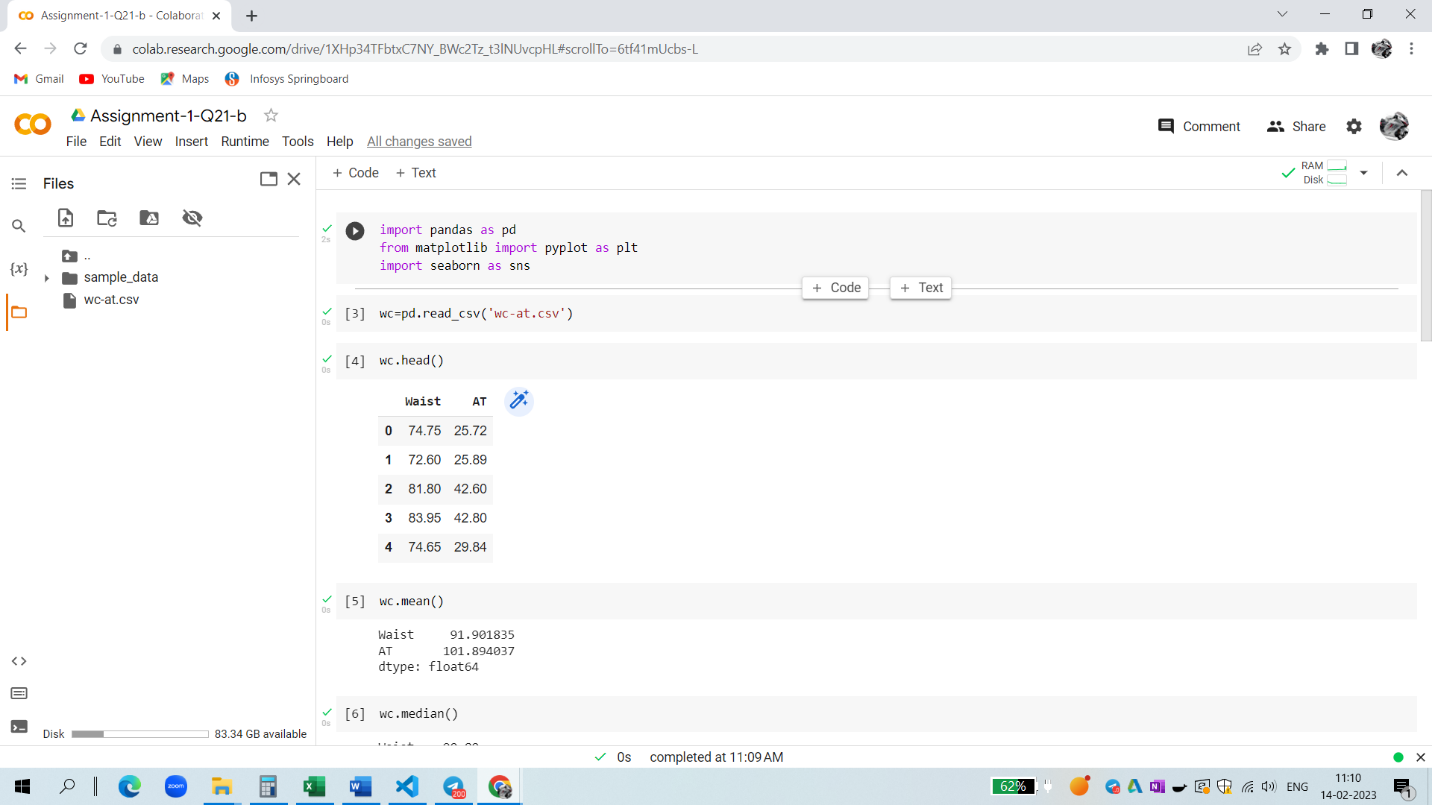
Dataset: Cars.csv

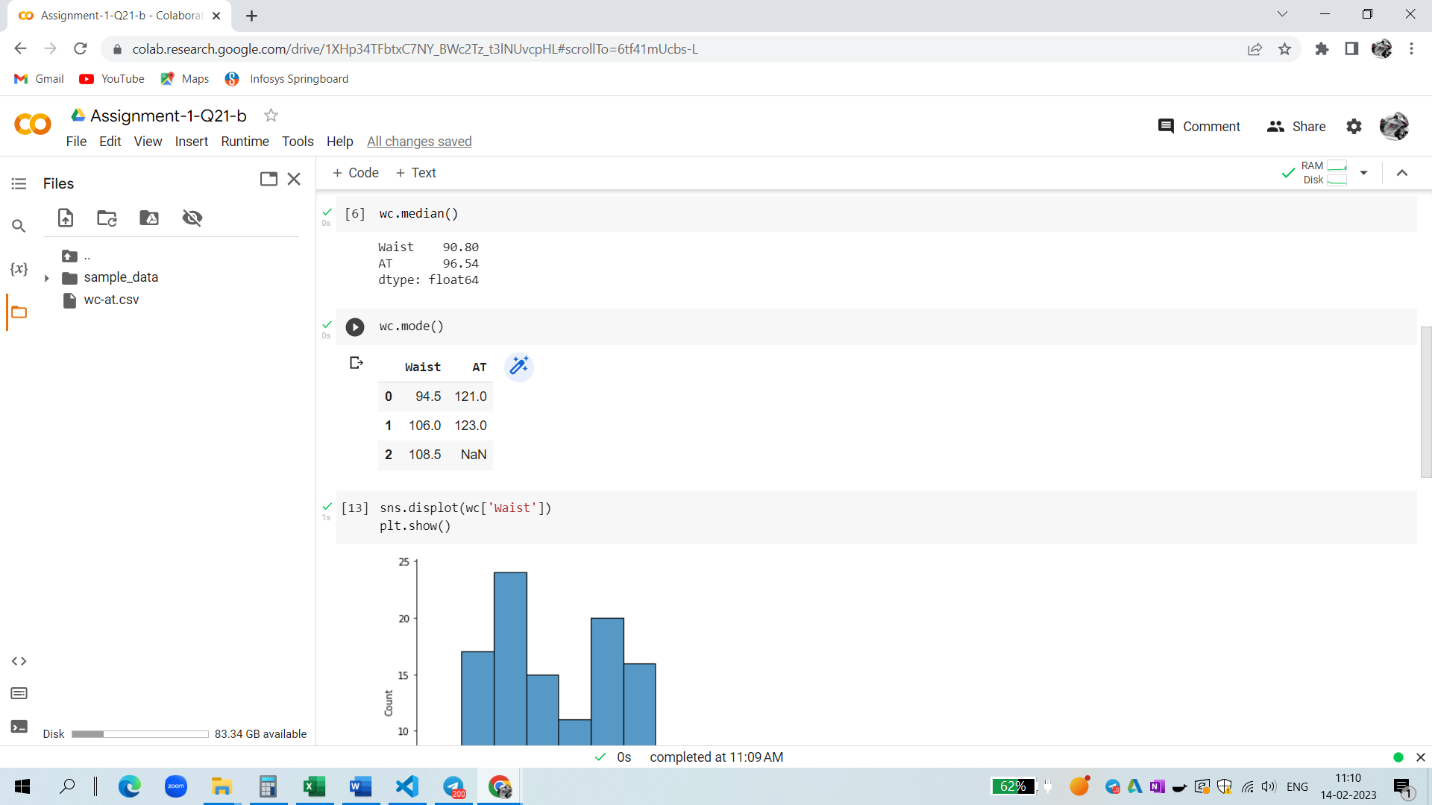


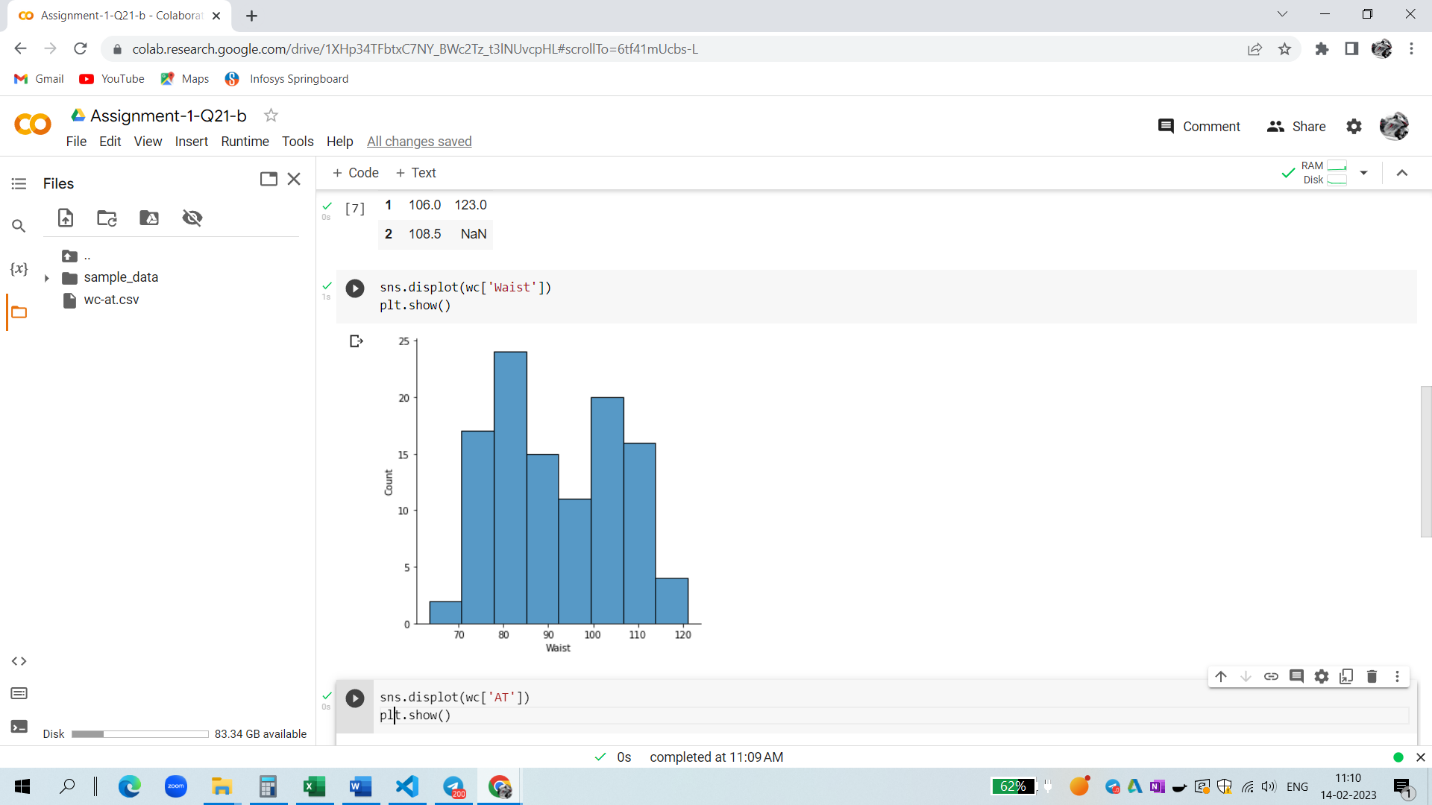


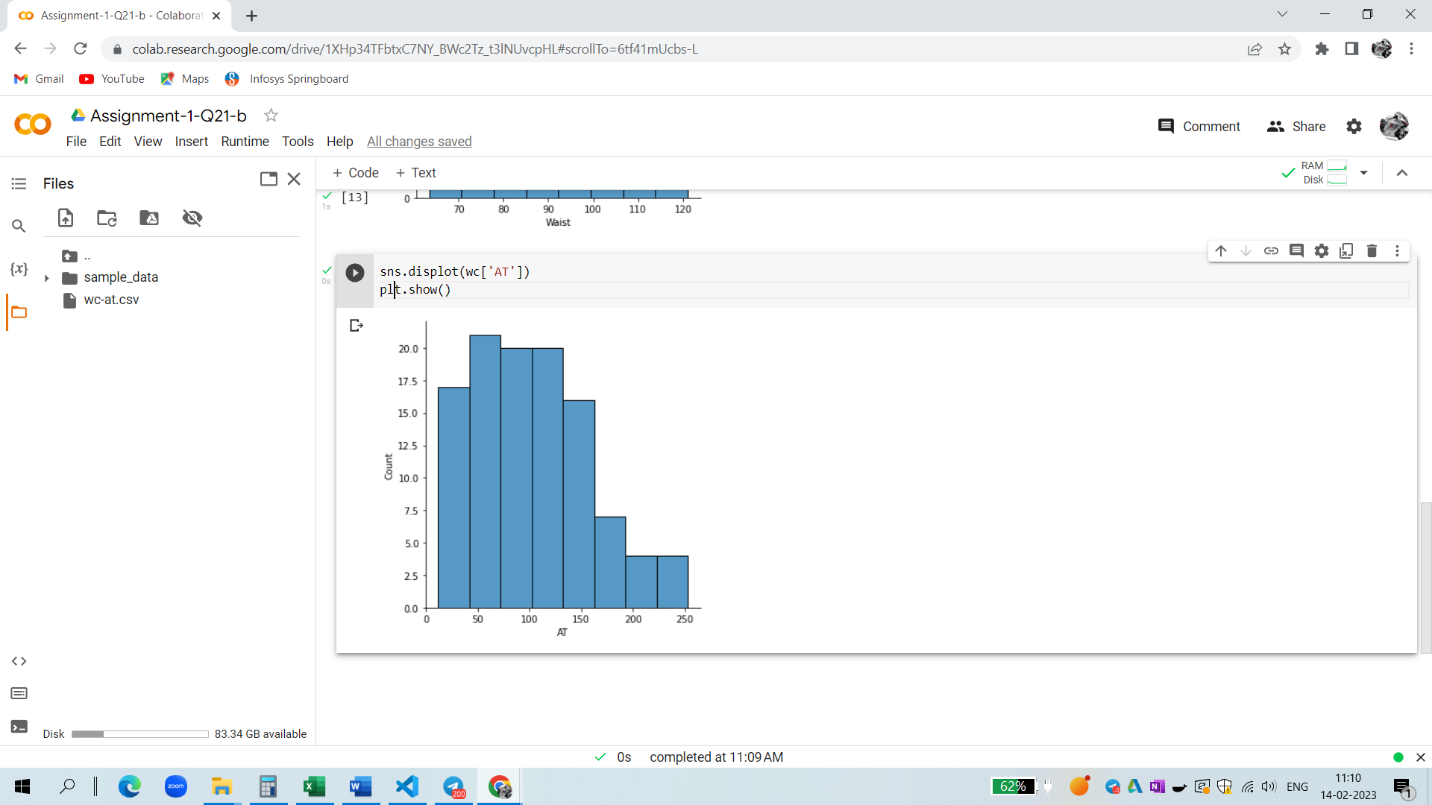
1. Check Whether the Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set follows Normal Distribution

Dataset: wc-at.csv



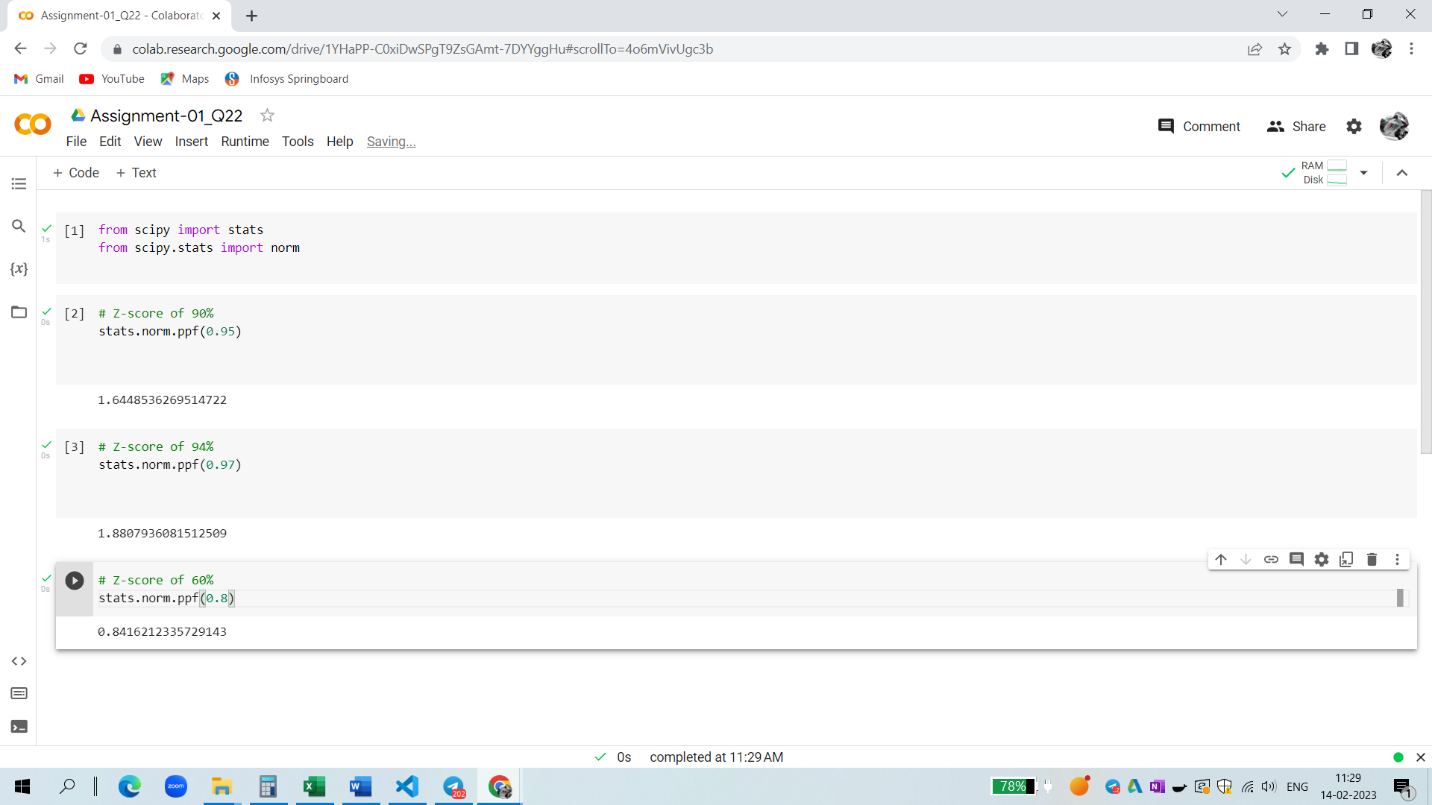






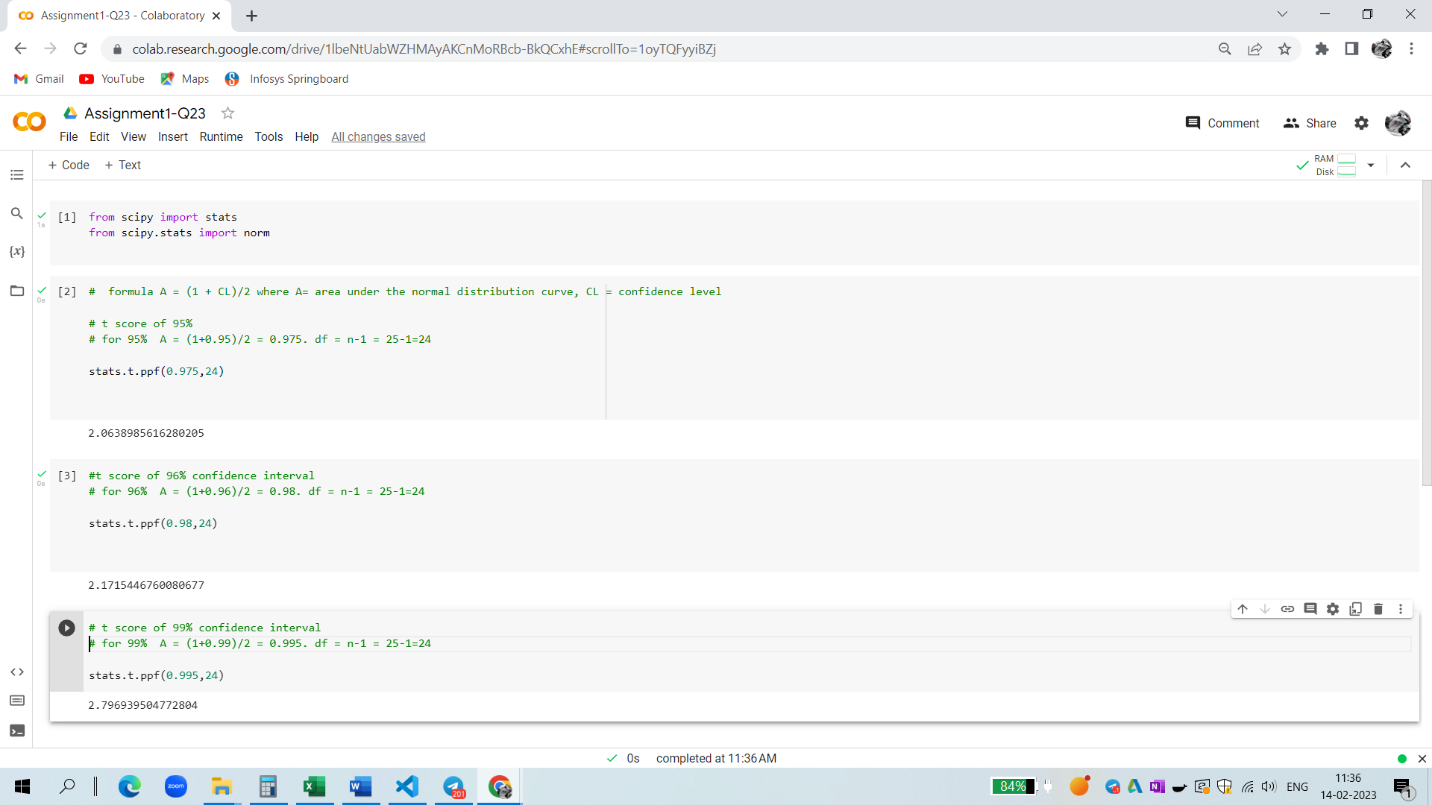
Q 22) Calculate the Z scores of 90% confidence interval,94% confidence interval, 60% confidence interval

Ans:



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

Ans:



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

Ans:-

