|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | Discreate |
| Results of rolling a dice | Discreate |
| 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) |  |

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

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

Ans - Three coin are tossed, the no. of possible combination = 2^3 = 8

The combination are HHH, THH, HTH, HHT, TTH, THT, HTT, TTT.

Ask – find the prob. That two head and one tail are obtained?

No. of combination that have two heads and one tail= 3(HHT, HTH, TTH)

The probability (P) of two head and one tail when three coin are tossed

P(two head and one tail) = no. of possible combination

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

Ans – two dice are rolled,

n(S) = { (1,1),(1,2),(1,3),(1,4),(1,5),(1,6),

(2,1),(2,2),(2,3),(2,4),(2,5),(2,6),

(3,1),(3,2),(3,3),(3,4),(3,5),(3,6),

(4,1),(4,2),(4,3),(4,4),(4,5),(4,6),

(5,1),(5,2),(5,3),(5,4),(5,5),(5,6),

(6,1),(6,2),(6,3),(6,4),(6,5),(6,6) }

n(S) = 36.

Conditions - 1) Sum is equal to 1 - n(A) = ‘0’ zero or null

2) Sum is less than or equal to 4 – n(B) = 6

n(B)/n(S) = 6/36 = 1/6 = 0.1666 = 16.66%

3) sum is divisible by 2 and 3 – n(C) = 6

n(C) / n(S) = 6/36 = 1/6 = 0.1666 = 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?

Ans- total number of balls= 2+3+2 =7

S be the sample space.

n(S) = no. of ways of drawing 2 balls out of 7

= 7!/2! = (7\*6)/(2\*1) = 21

E = drawing 2 ball, none of which is blue

n(E) = no. of ways of drawing 2 balls out of 2+3 balls

=5!/2! = (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 - the probabilities of count of candies for children (ignoring the nature of the child)

Child A – probability of having 1 candy = 0.015.

Child B – probability of having 4 candies = 0.20

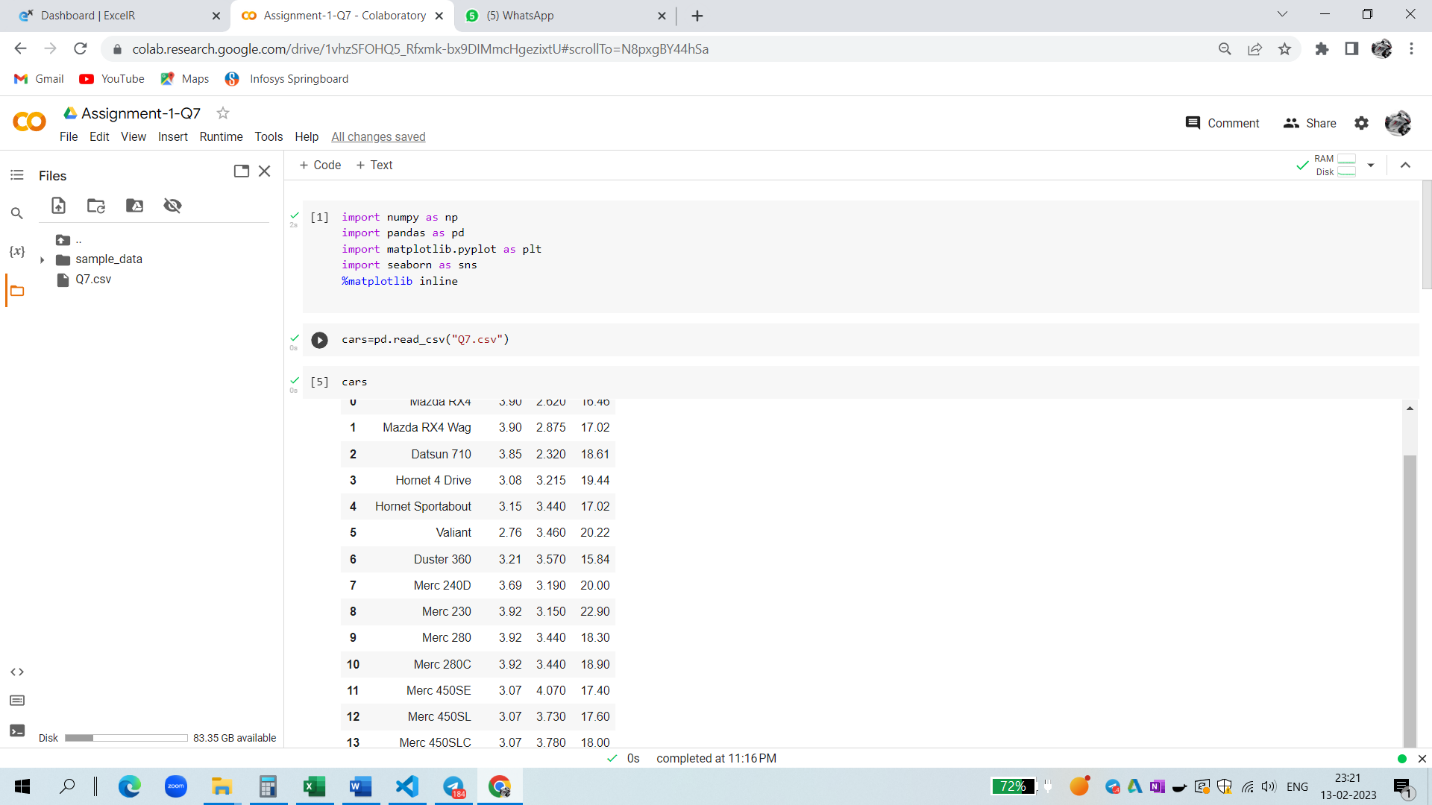
i.e (0.015\*1)+(0.20\*4)+(0.65\*3)+(0.005\*5)+(0.01\*6)+(0.120\*2)

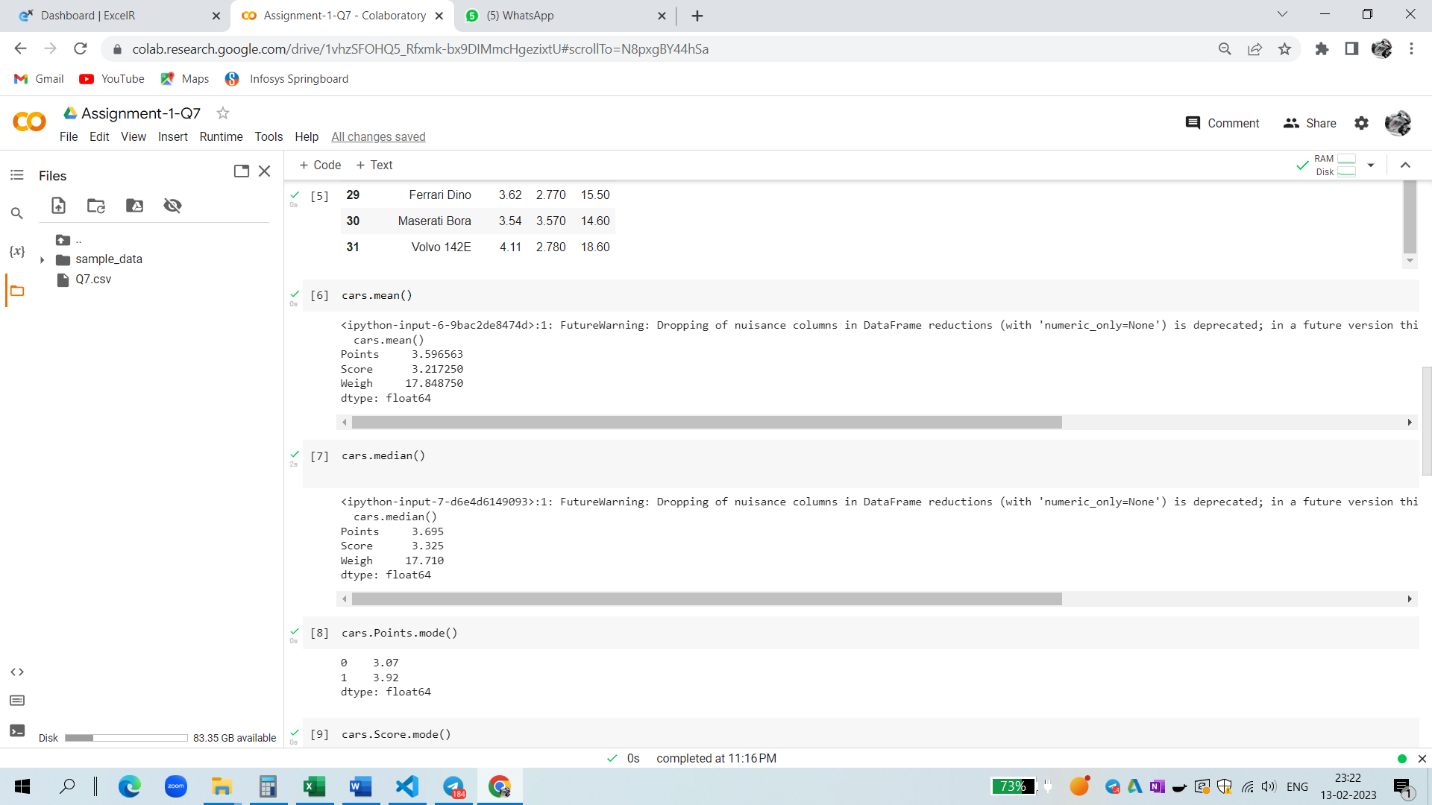
= 3.09

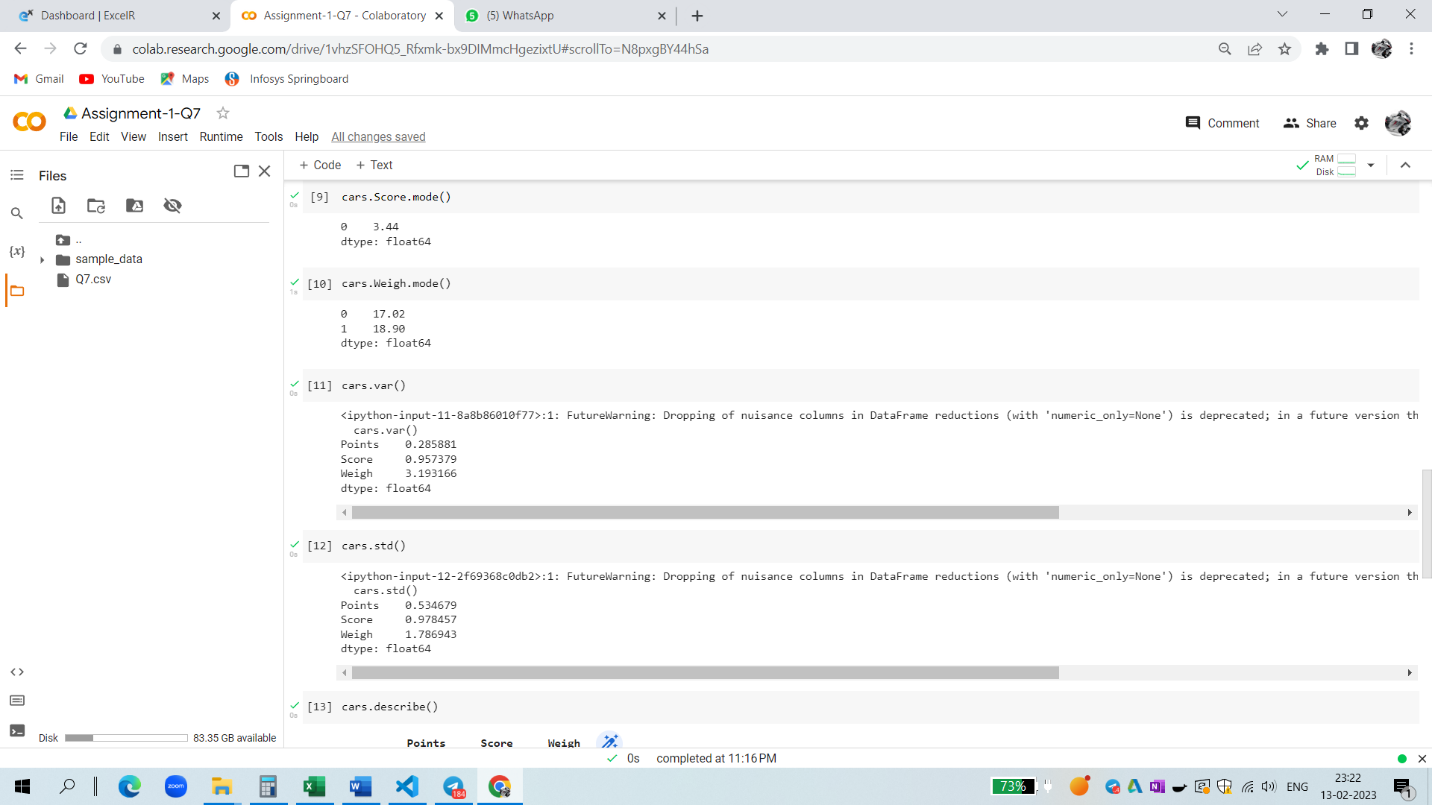
Q7) Calculate Mean, Median, Mode, Variance, Standard Deviation, Range & comment about the values / draw inferences, for the given dataset

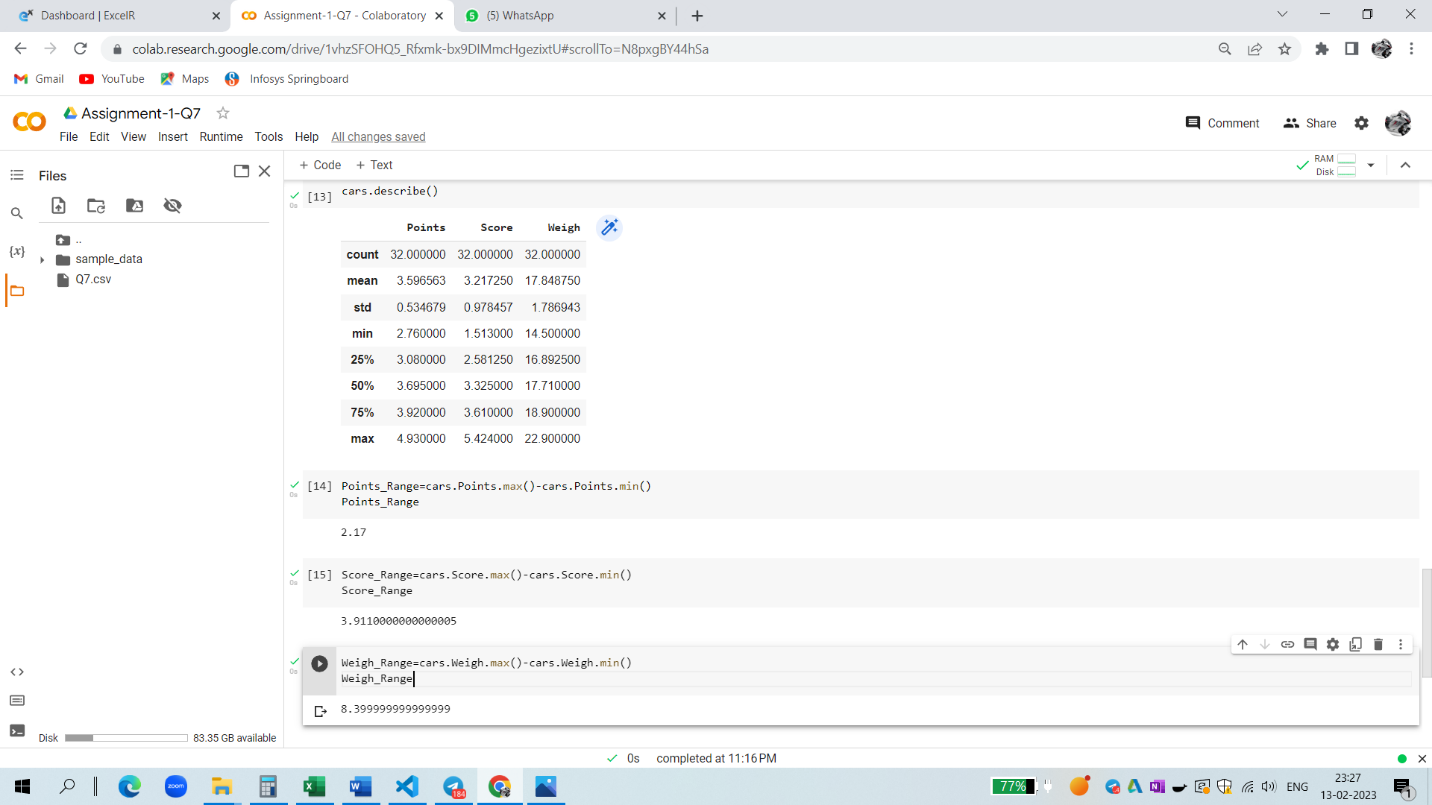
* For Points,Score,Weigh>

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

**Use Q7.csv file ANSWER – (GOOGLE COLLAB)**







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 - Expected Value  =  ∑ ( probability  \* Value )

 ∑ P(x).E(x)

there are 9 patients

Probability of selecting each patient = 1/9

Ex  108, 110, 123, 134, 135, 145, 167, 187, 199

P(x)  1/9  1/9   1/9  1/9   1/9   1/9   1/9   1/9  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)

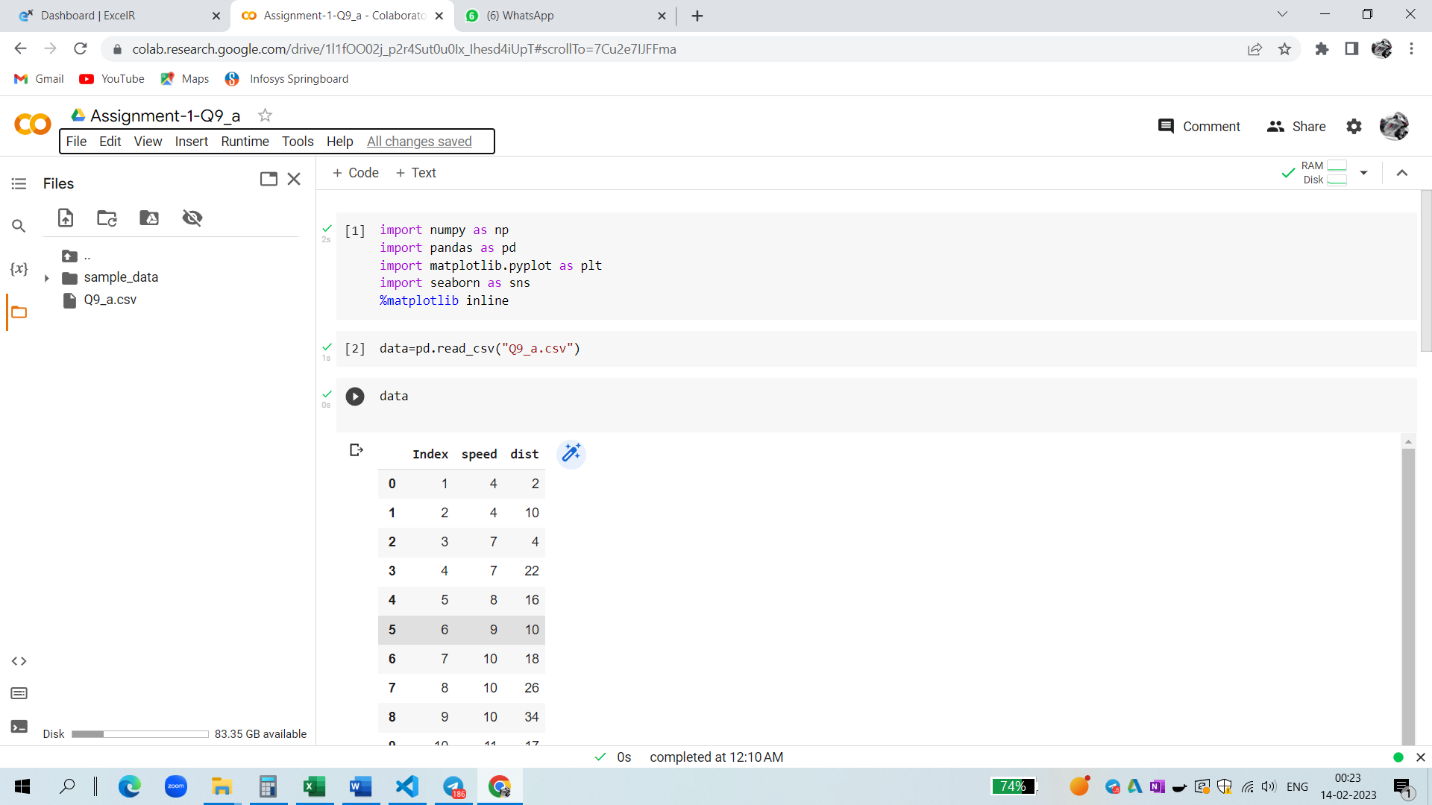
= (1/9)\* (  1308)

= 145.33

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

**Cars speed and distance**

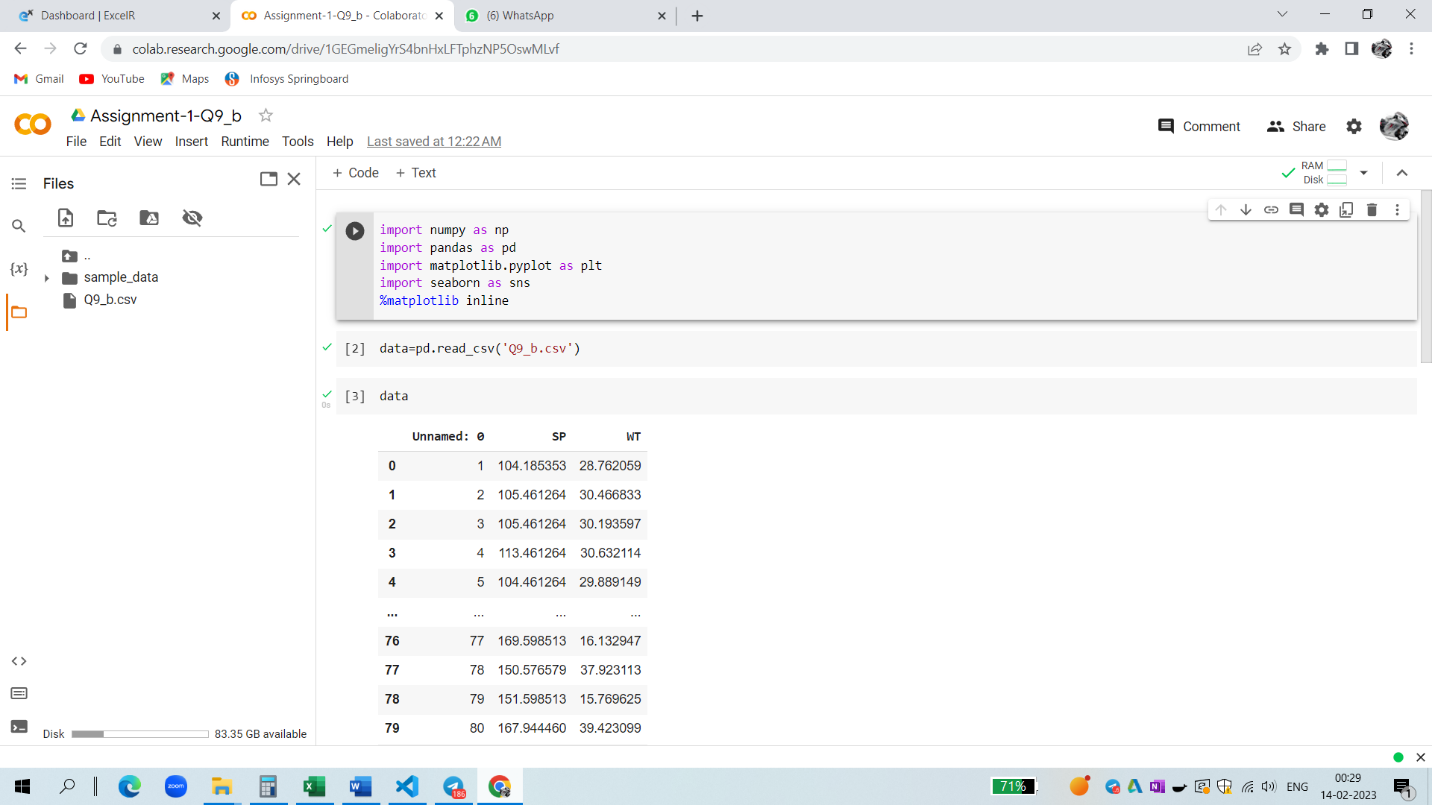
**Use Q9\_a.csv**

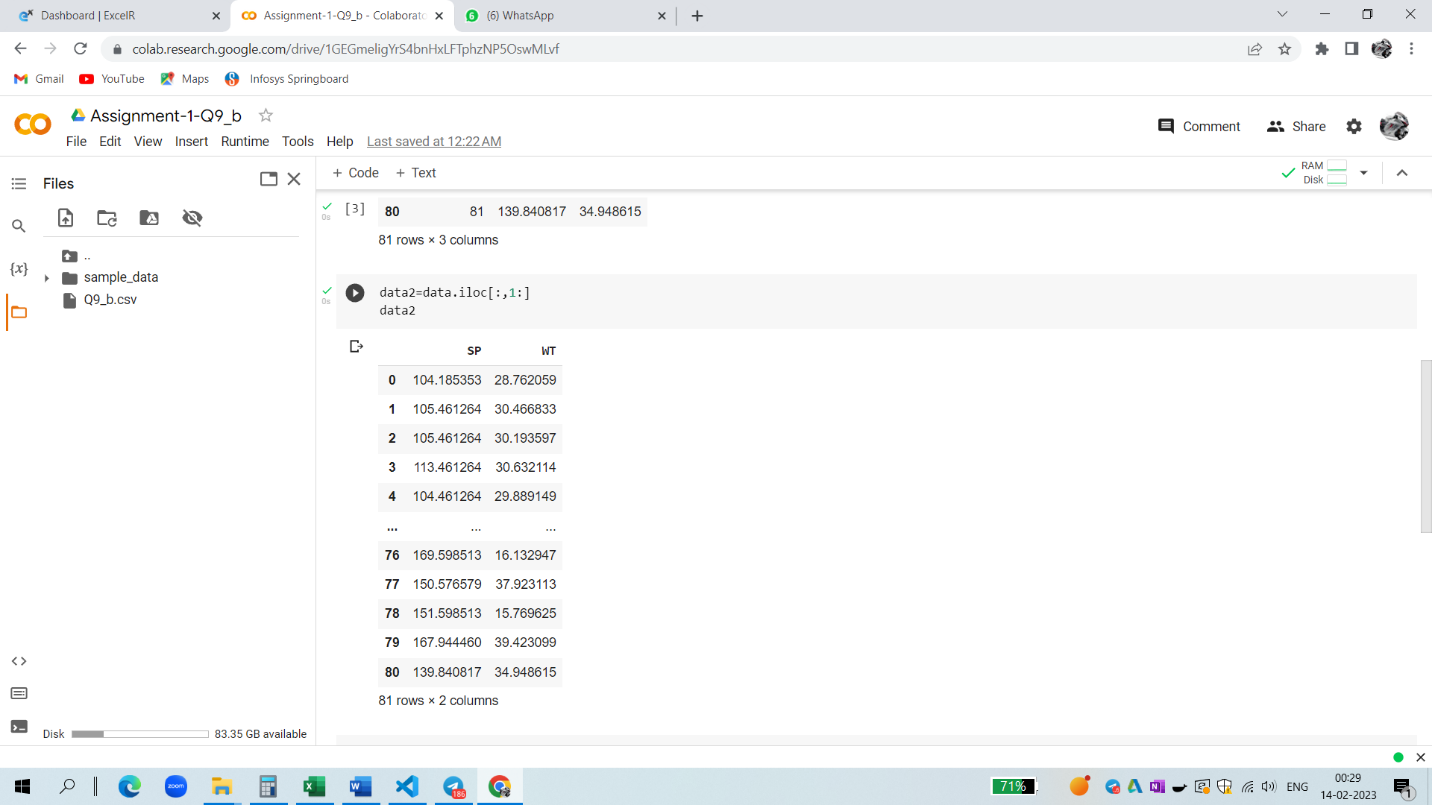
**Ans - **

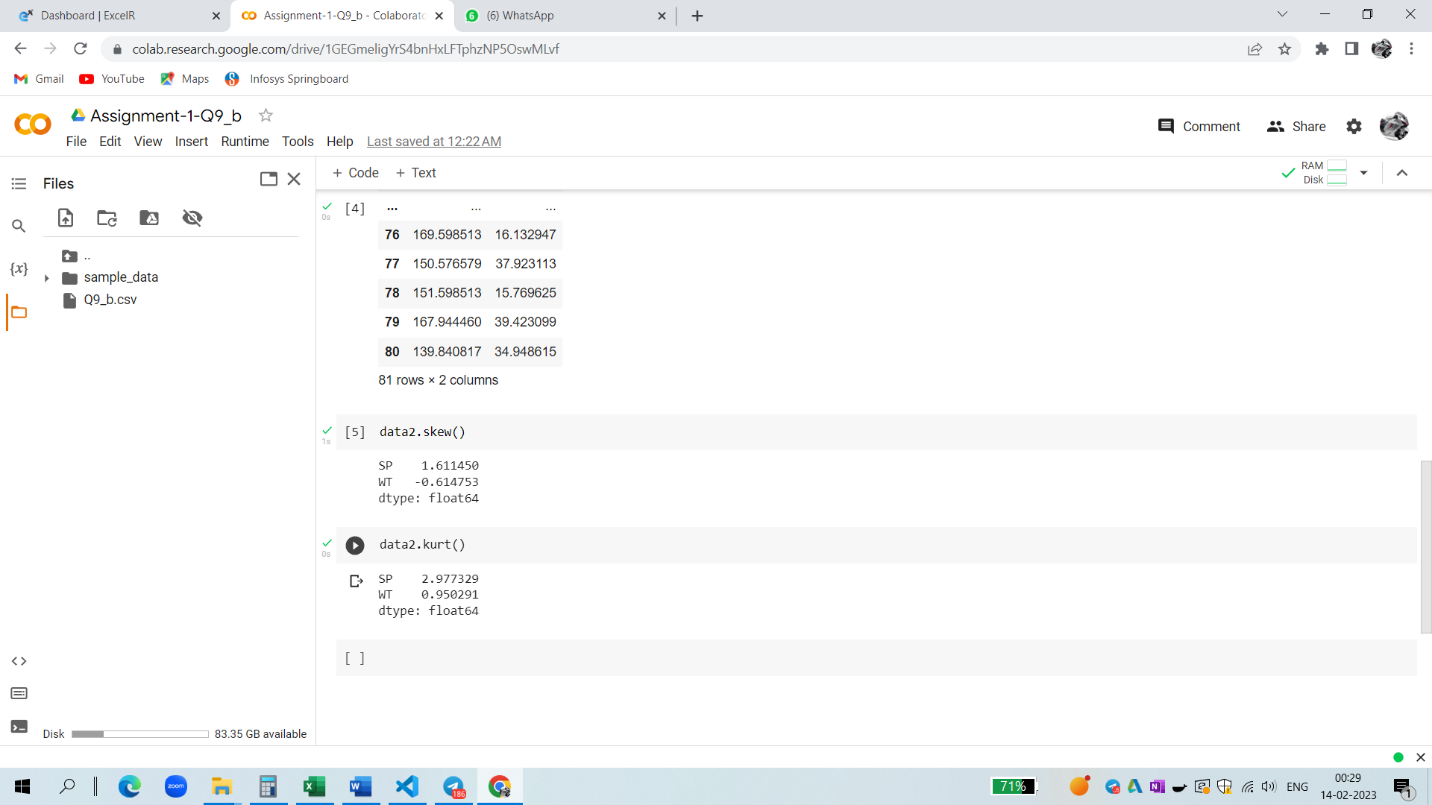
****

**SP and Weight(WT)**

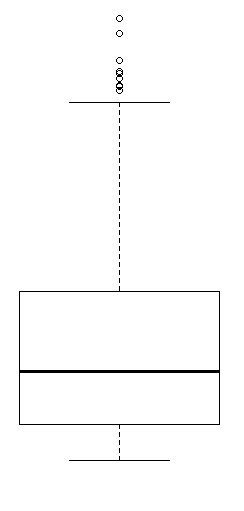
**Use Q9\_b.csv**

**Ans - **

****

****

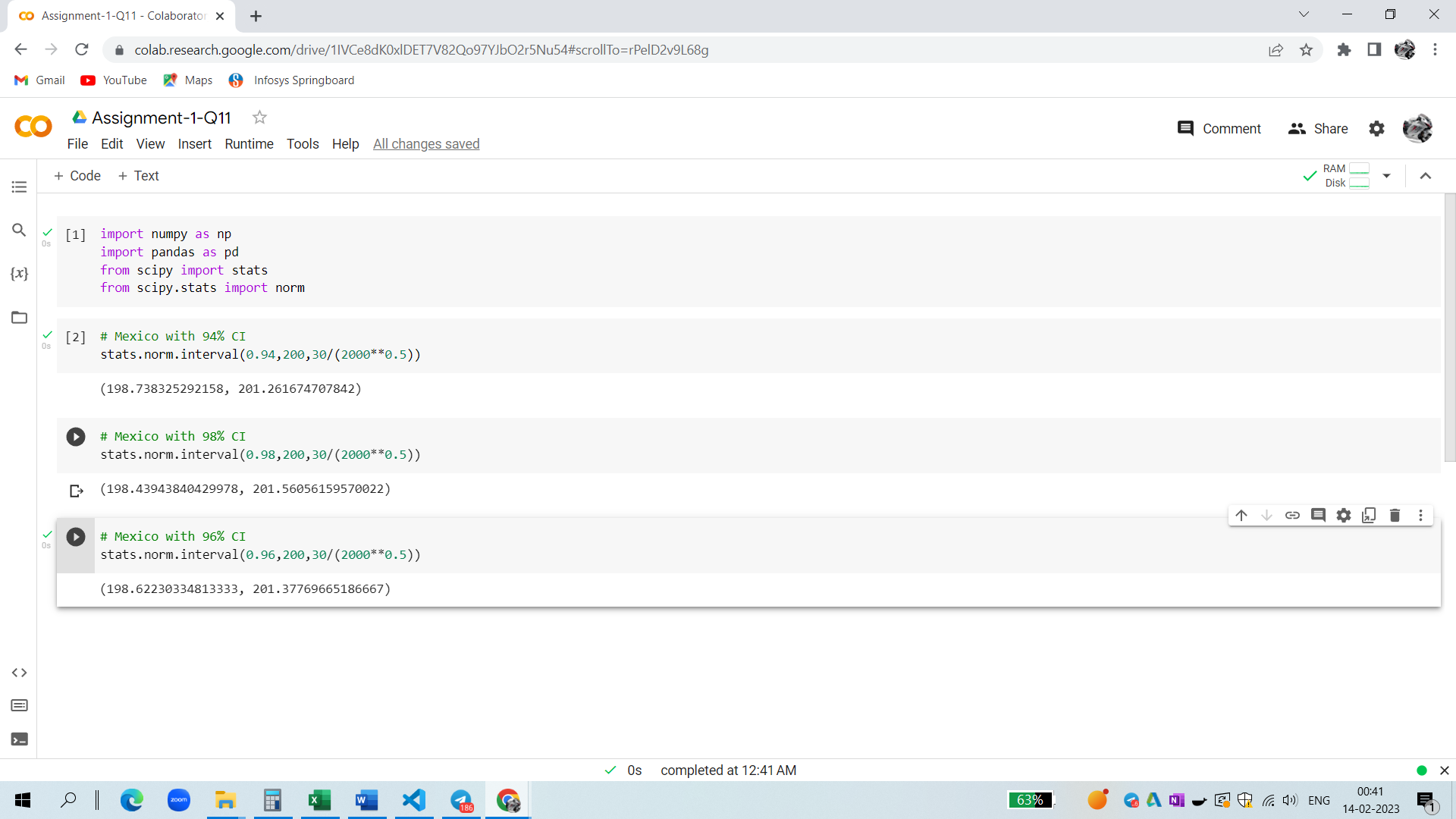
**Q10) Draw inferences about the following boxplot & histogram**



Ans- This graph histogram showing that the distribution of the data is Asymmetric. This is right skewed and we also say that the this is positively skewed.

The boxplot showing positively skewness of the 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?

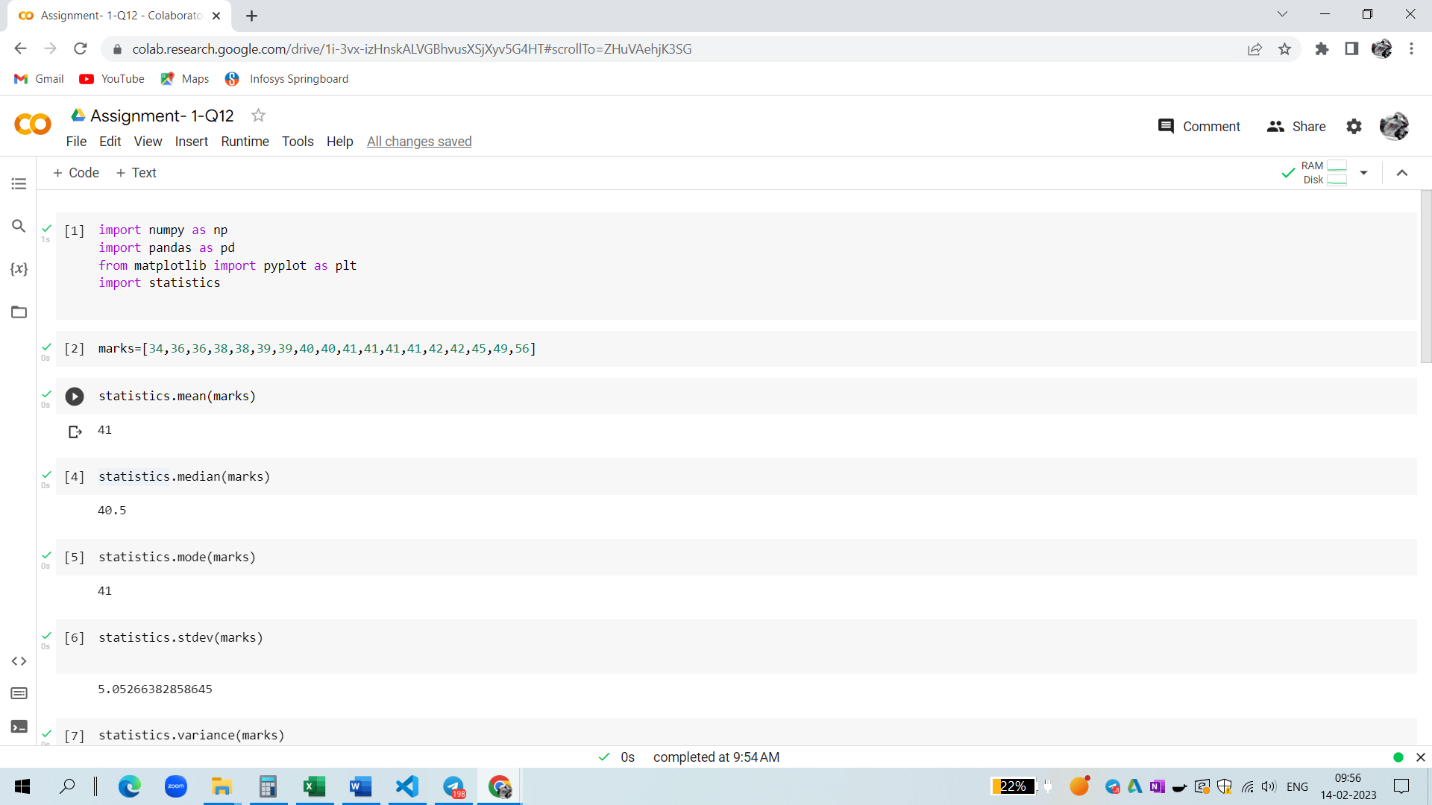
Ans - 

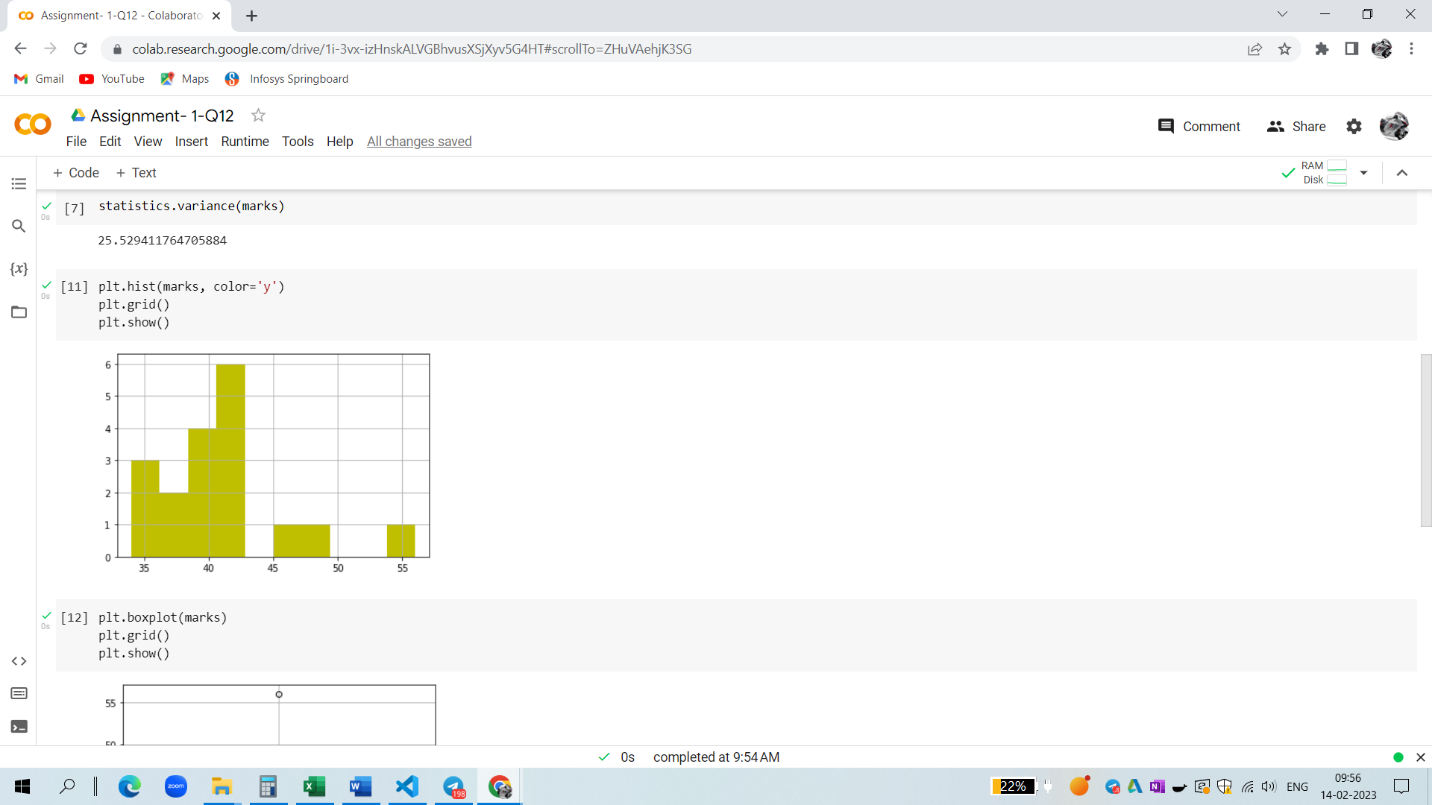
Q12) Below are the scores obtained by a student in tests

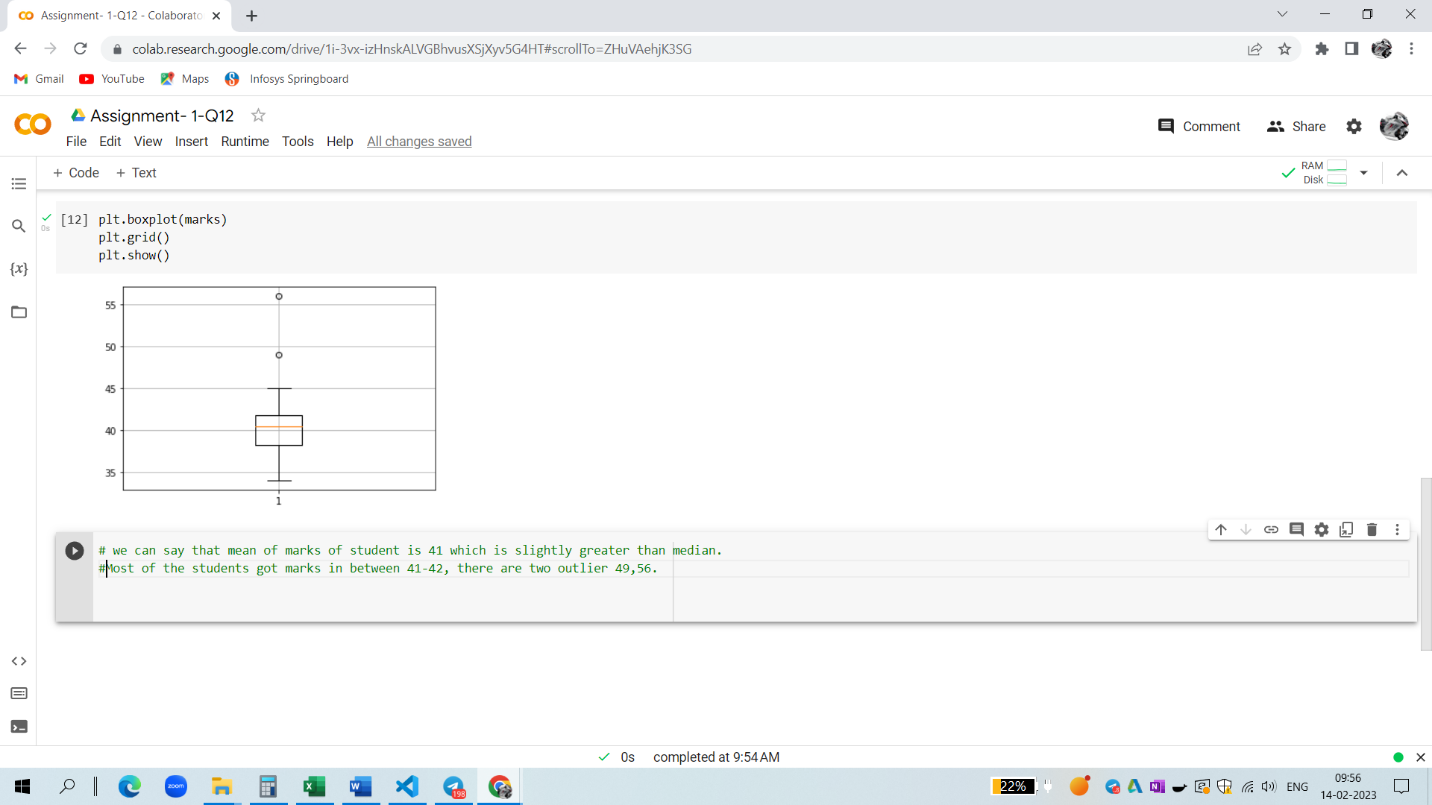
34,36,36,38,38,39,39,40,40,41,41,41,41,42,42,45,49,56

1) Find mean, median, variance, standard deviation.

2) What can we say about the student marks?

Ans - 



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

Ans – No skewness is present when we have a perfect symmetrical distribution.

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

Ans – skewness towards right .

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

Ans - skewness towards left .

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

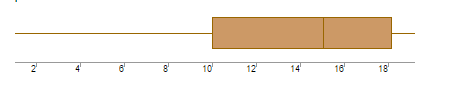
Ans – the curve it is leptokurtosis.

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

Ans – the curve is flatter

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

Ans -



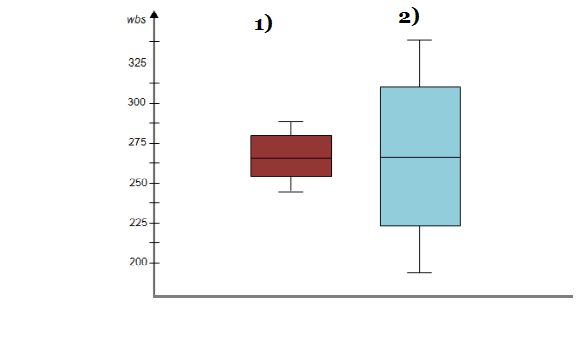
1. What can we say about the distribution of the data?
2. What is nature of skewness of the data?
3. What will be the IQR of the data (approximately)?

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

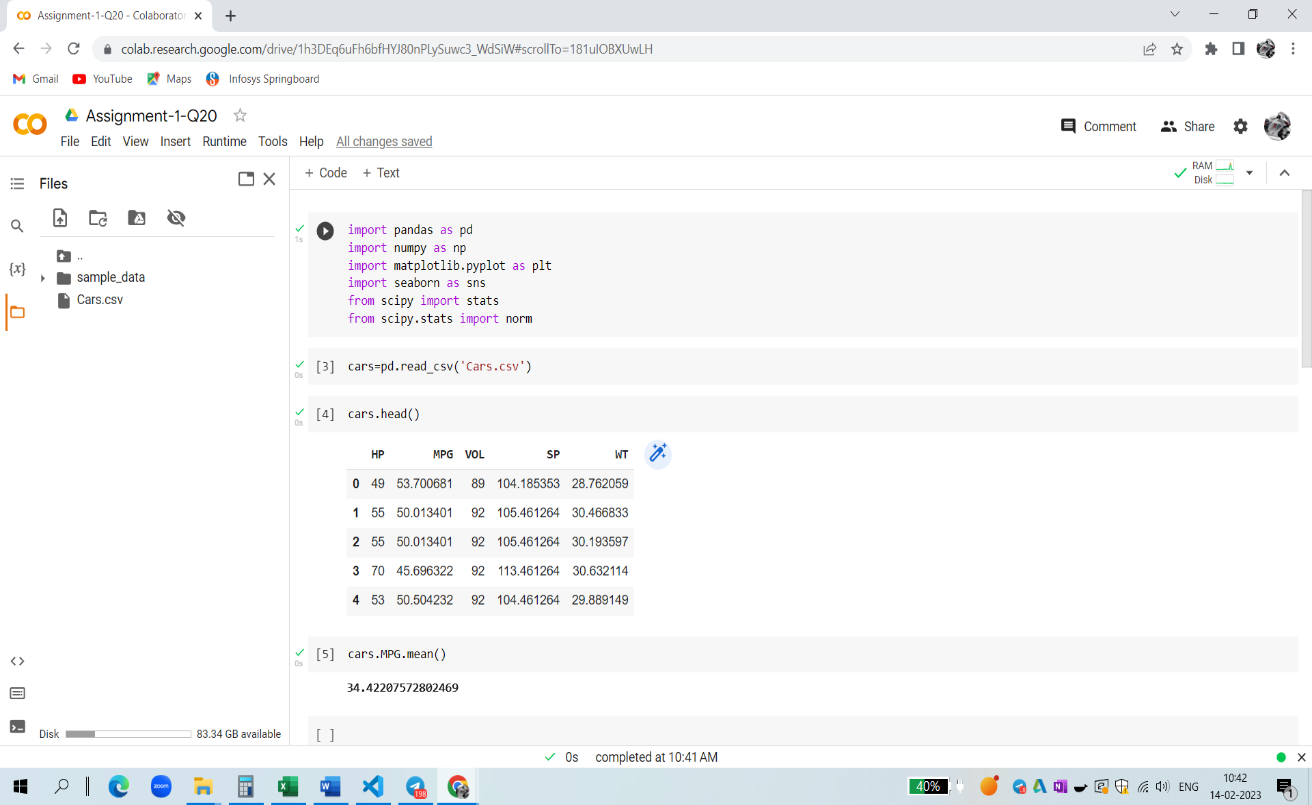
Q 20) Calculate probability from the given dataset for the below cases

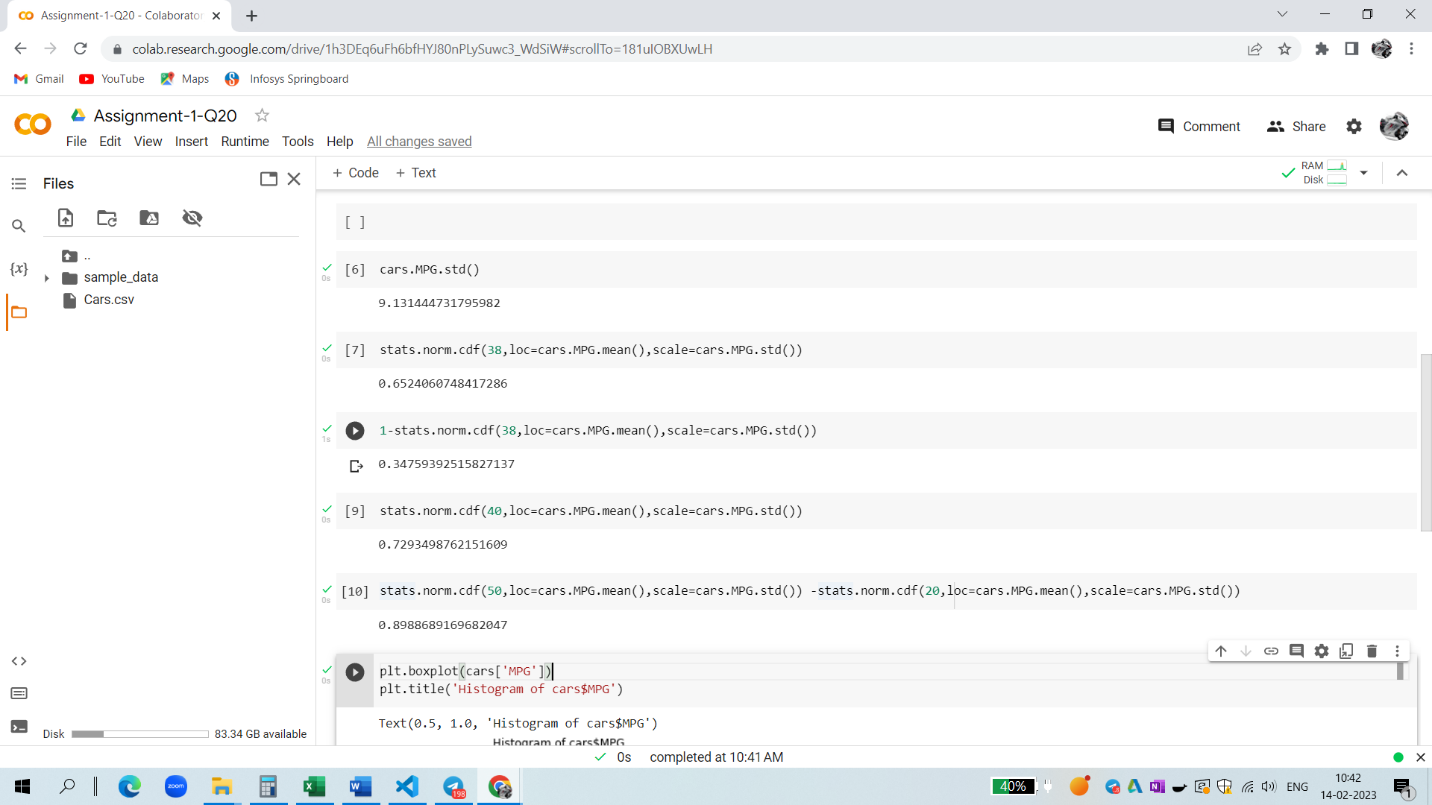
Data \_set: Cars.csv

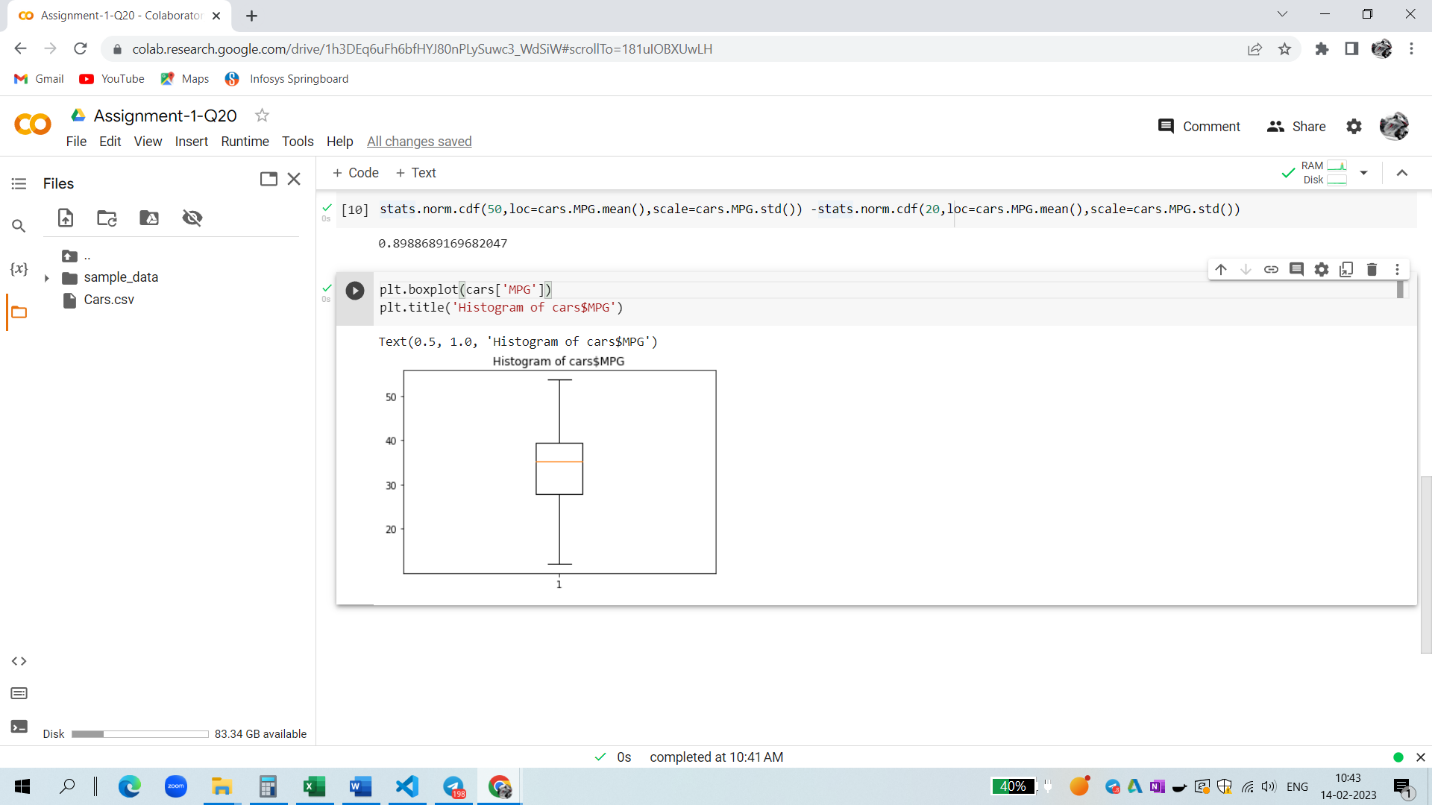
Calculate the probability of MPG of Cars for the below cases.

MPG <- Cars$MPG

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

Ans - 

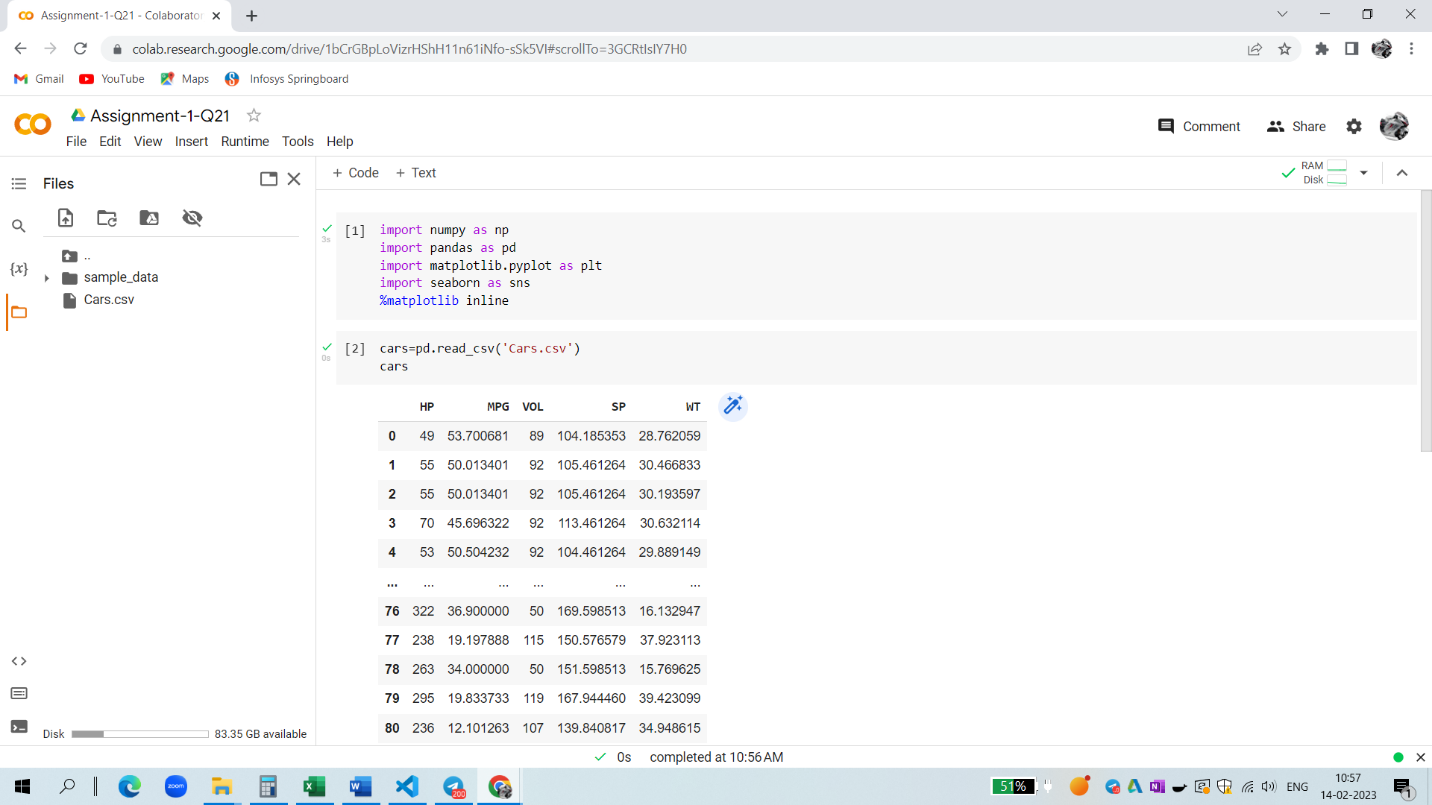


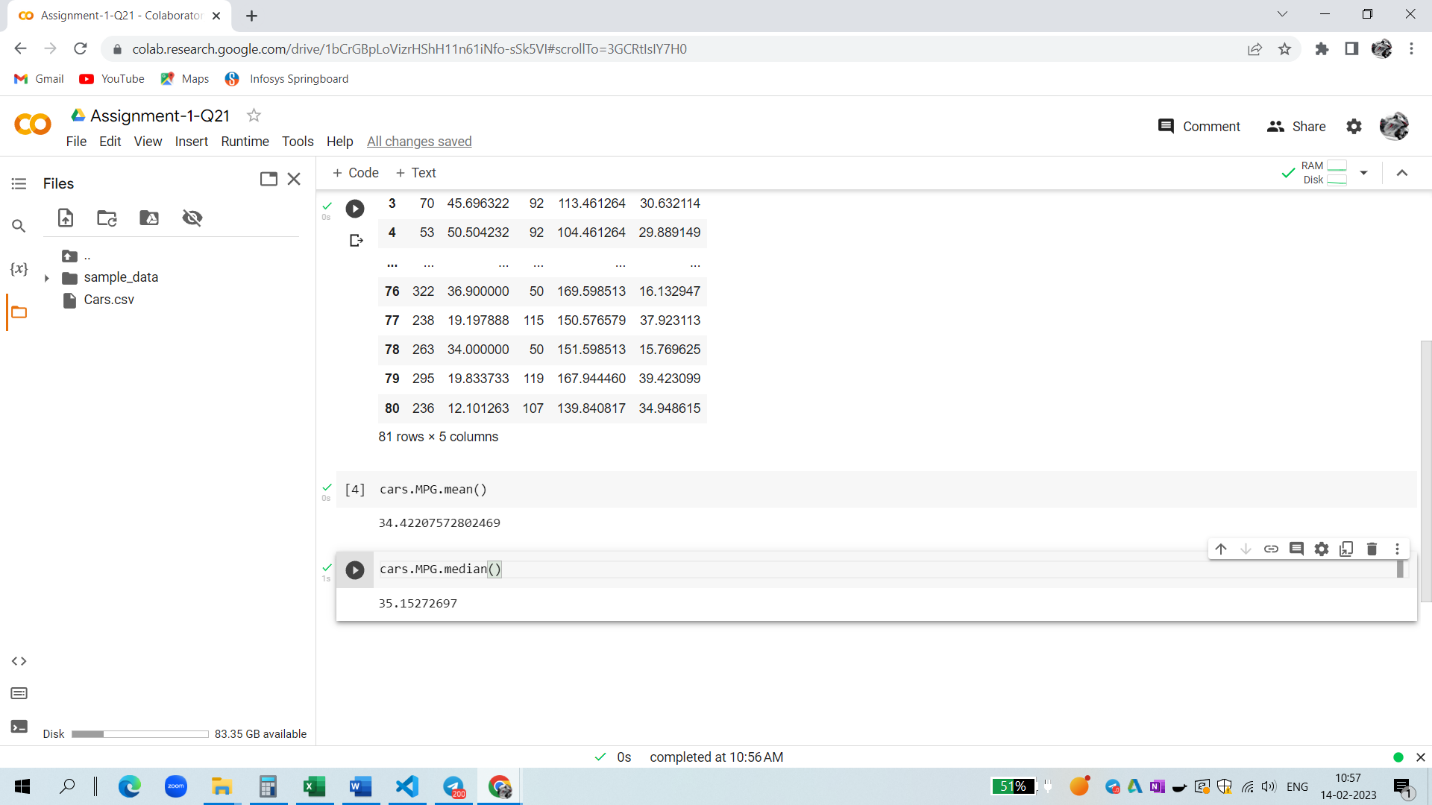


Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

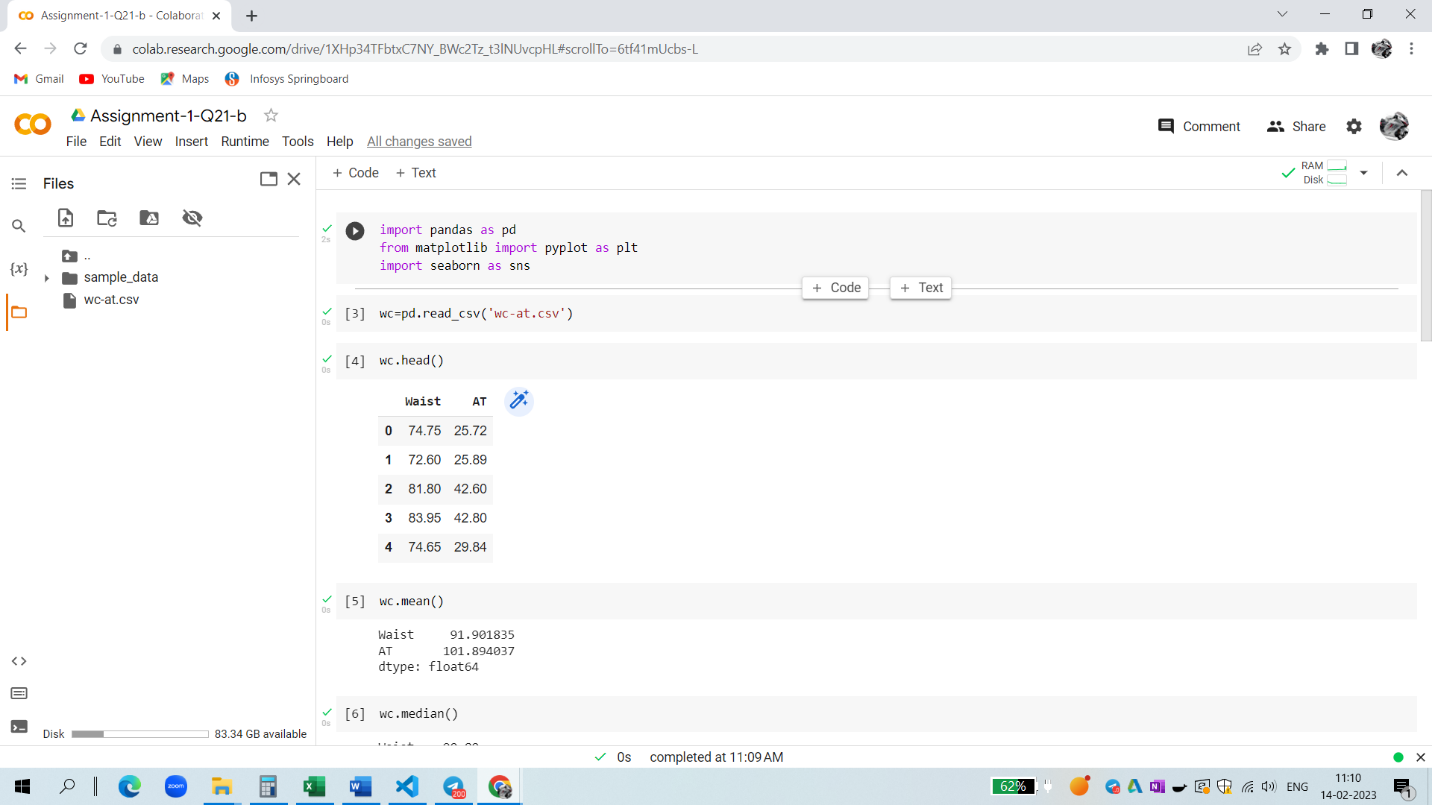
Dataset: Cars.csv

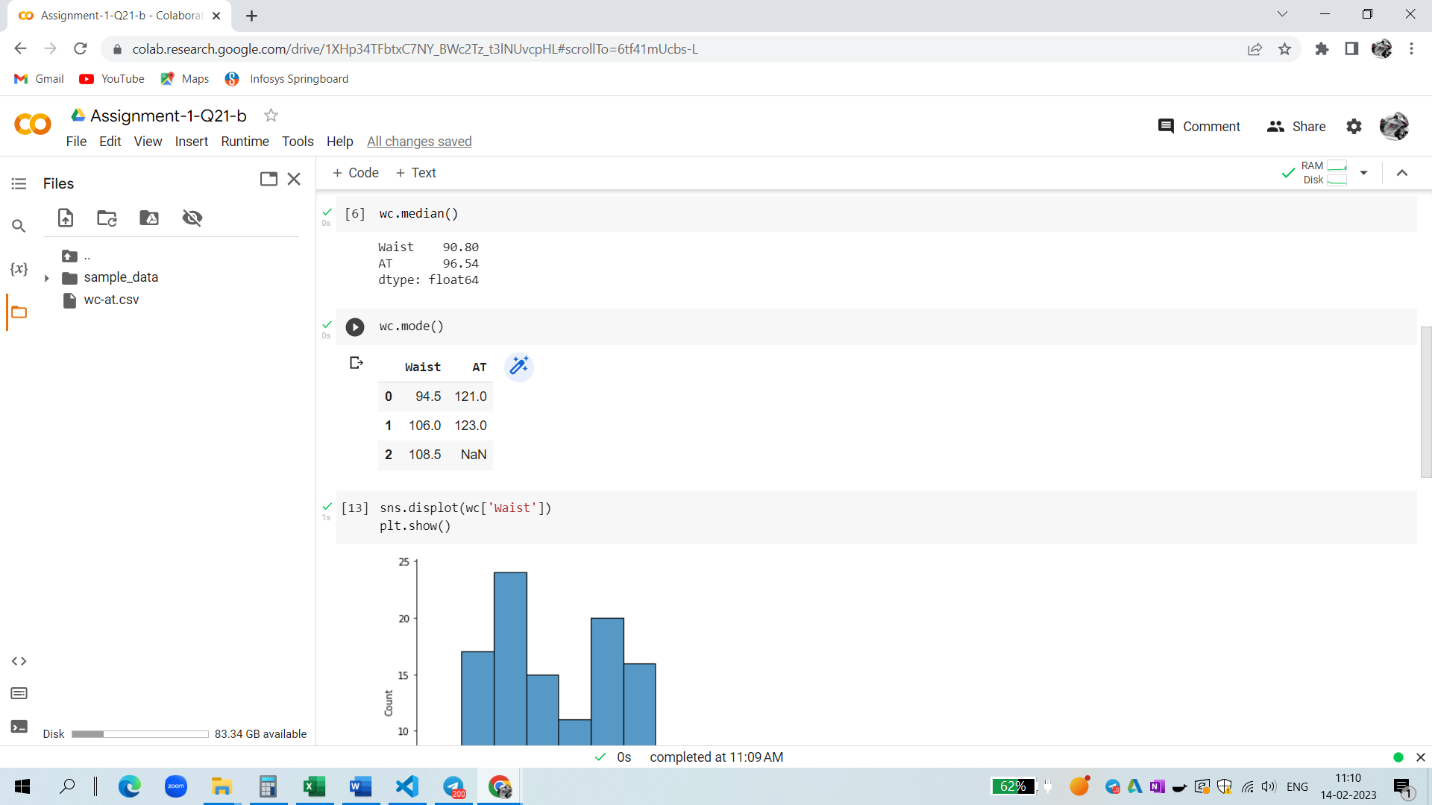


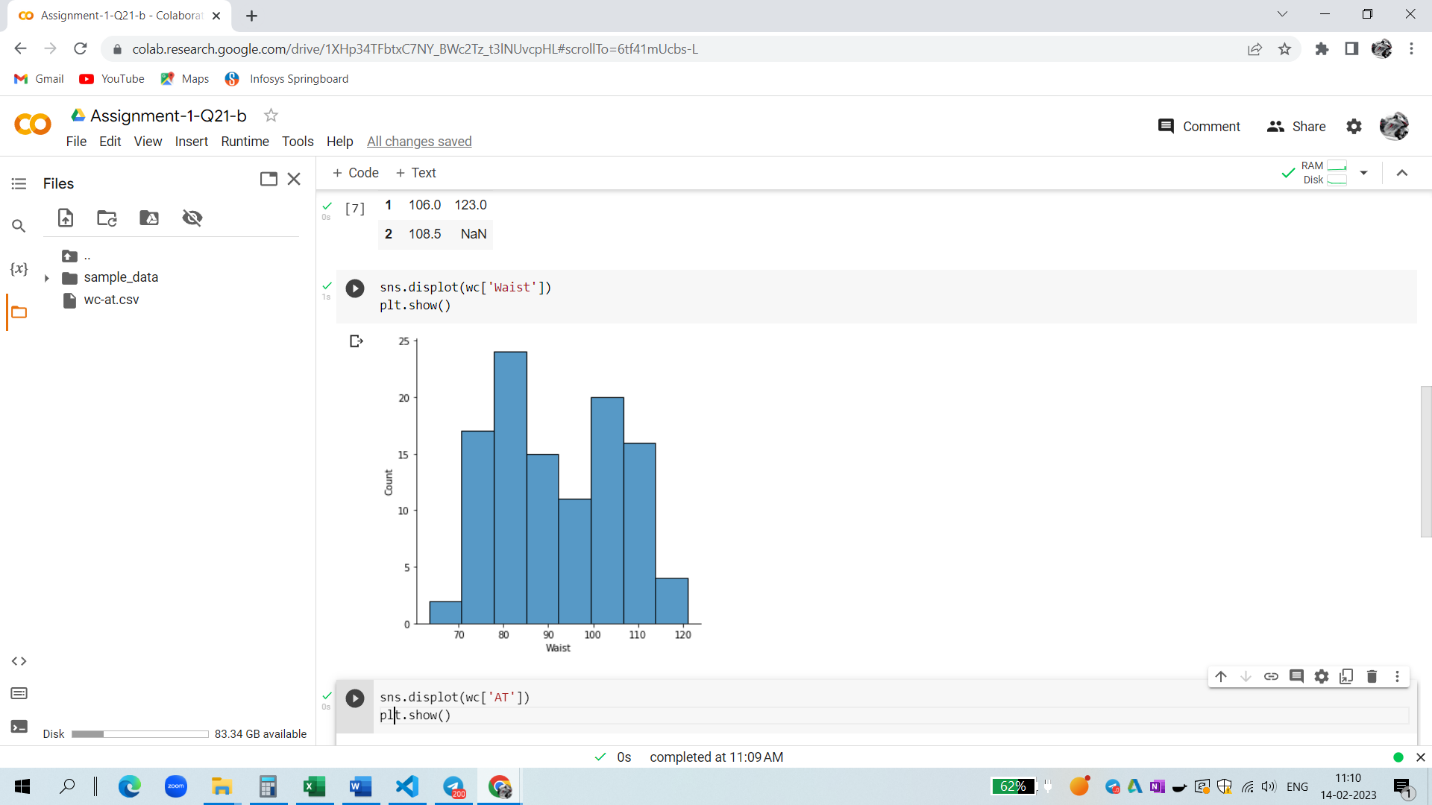


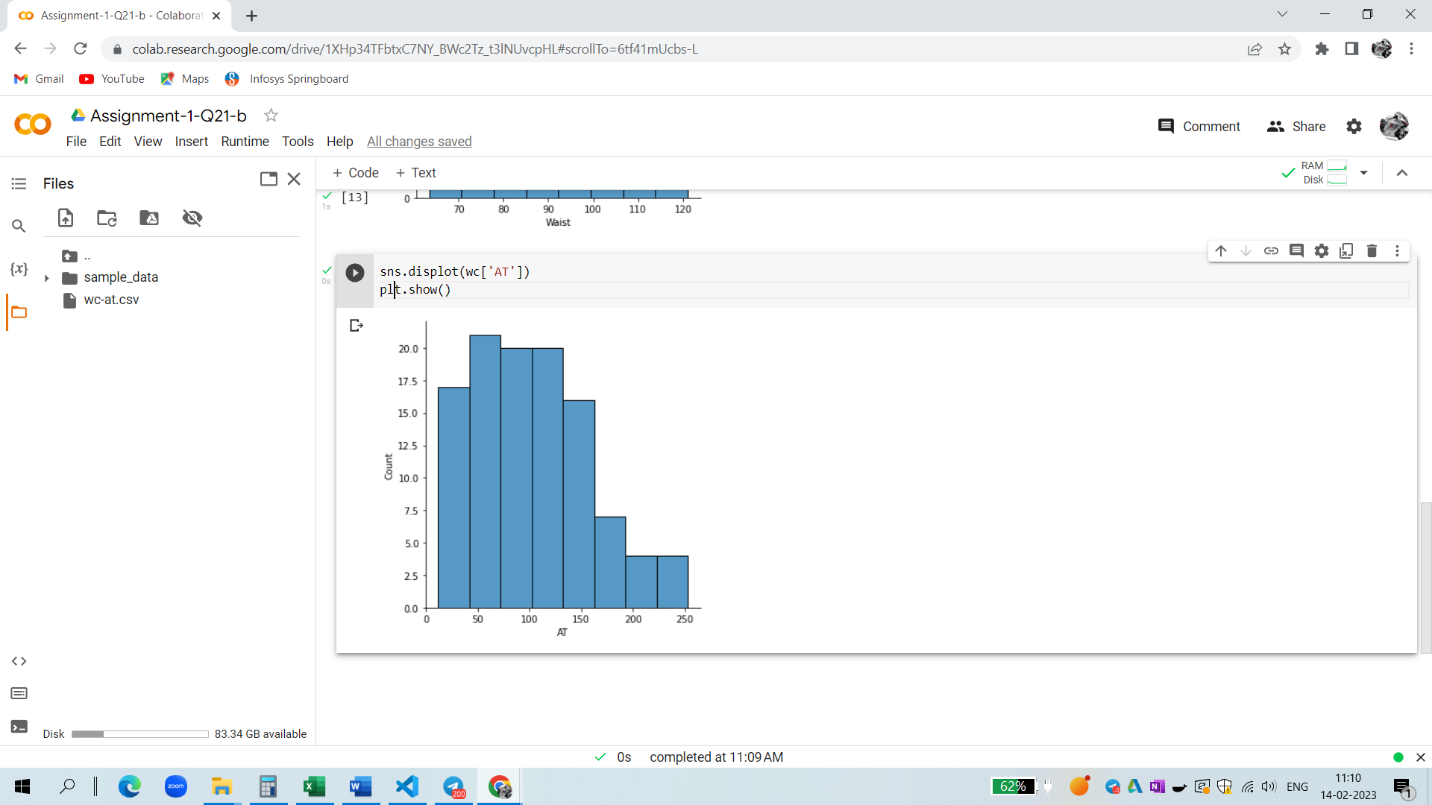
1. Check Whether the Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set follows Normal Distribution

Dataset: wc-at.csv

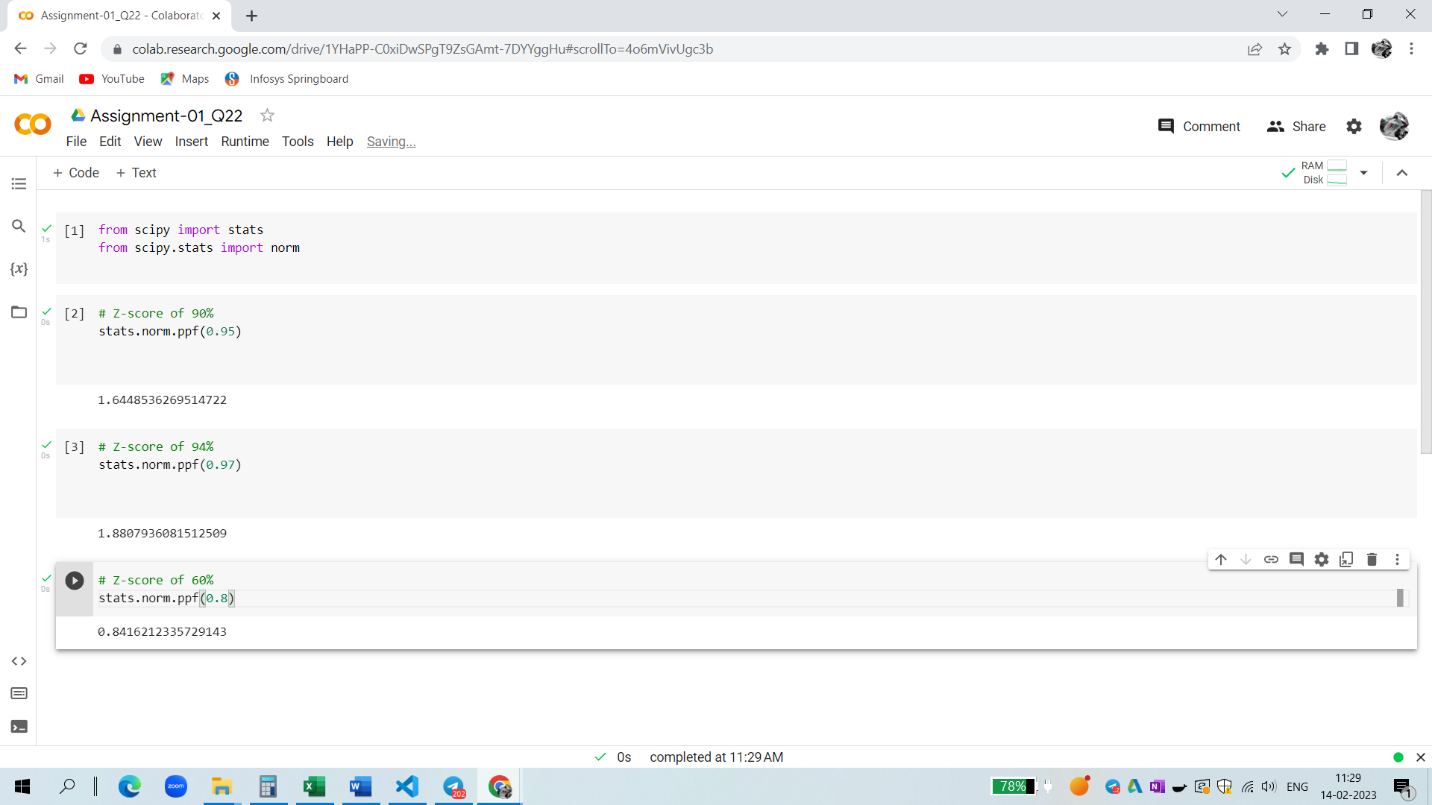








Q 22) Calculate the Z scores of 90% confidence interval,94% confidence interval, 60% confidence interval

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 - 