|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | Discrete |
| Results of rolling a dice | Discrete |
| Weight of a person | Continoues |
| Weight of Gold | Continuous |
| Distance between two places | Continuous |
| Length of a leaf | Continuous |
| Dog's weight | Continuous |
| Blue Color | Dicrete |
| 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 | 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 | ratio |
| Religious Preference | nominal |
| Barometer Pressure | Interval |
| SAT Scores | Interval |
| Years of Education | ratio |

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

P(2 H,1T)= {HHT,HTH,THH}= 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
4. 0
5. (1,2),

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?

7c2= (7\*6)/(2\*1)=21

5c2=(5\*4)/(2\*1)=10

10/21

P(€)=N(€)/n(s)

P(5/7, 4/6)=20/42=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.

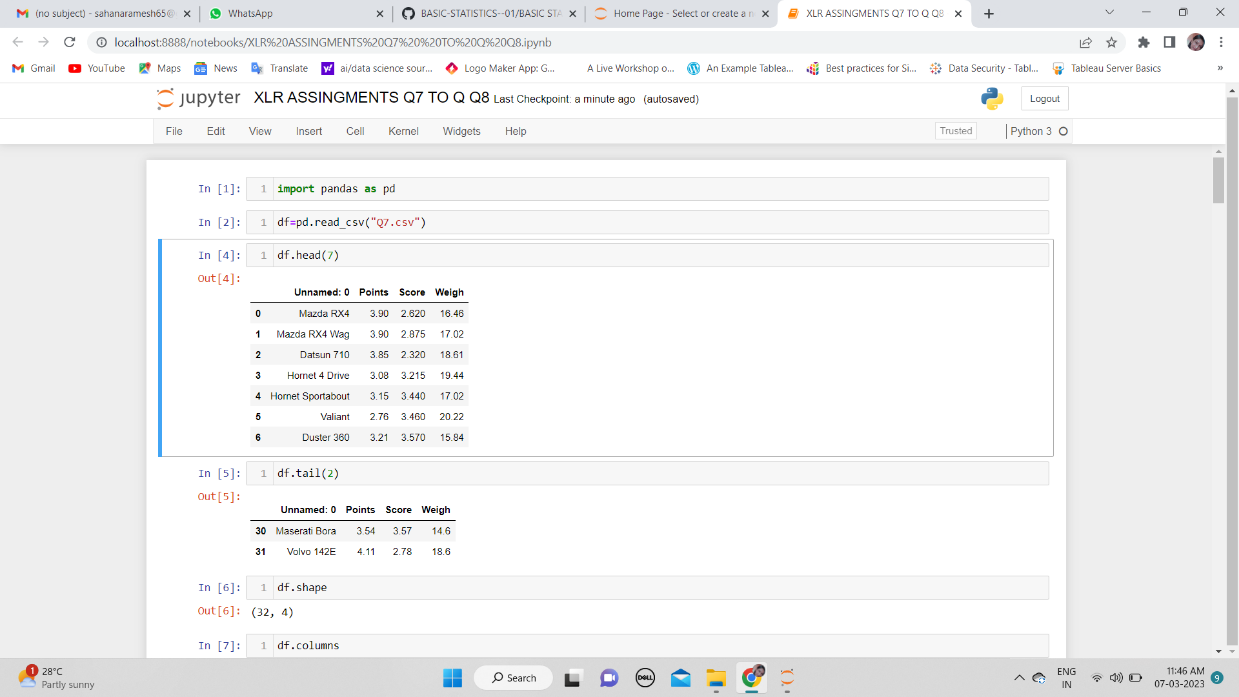
Child B – probability of having 4 candies = 0.20

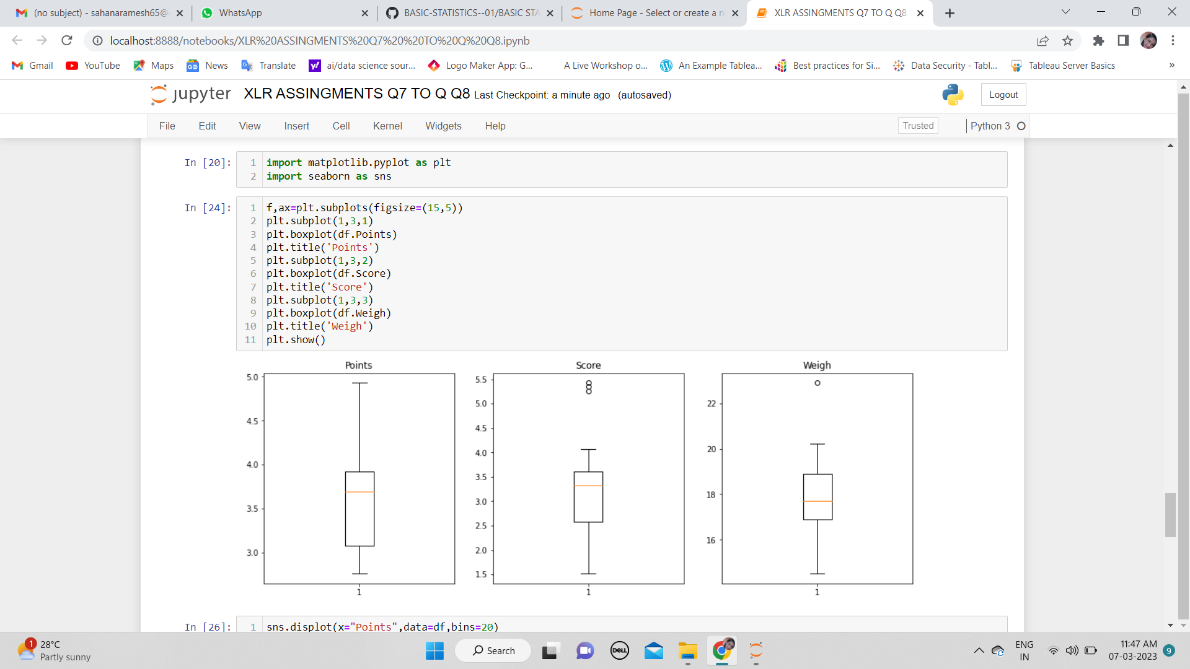
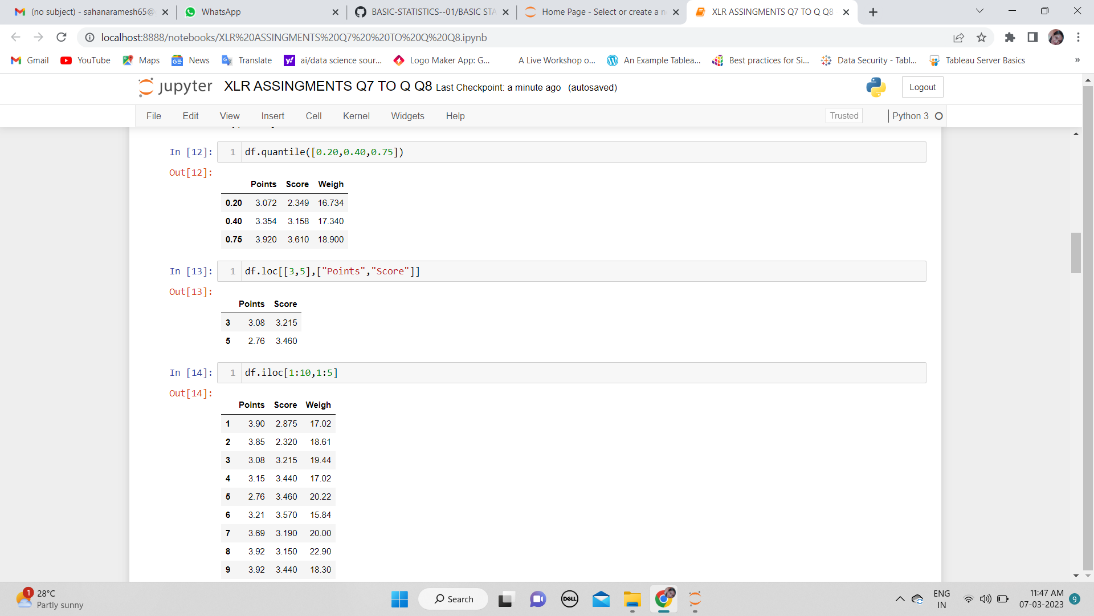
P(no. of cadies for randomly selected child)= 1\*0.015+4\*0.20+3\*0.65+5\*0.05+6\*0.01+2\*0.120= 3.09

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 **

****

ANS: As we notice in Points dataset:

1)Data **is** concentrated around Median**.** There **are** no outliers**. So,** the distribution **is** Right skewed**.**

In Score dataset, Data **is** concentrated around Median. There are 3 Outliers**. So the,** The distribution **is** Left skewed**.**

In Weigh datasets: data **is** concentrated around Median, There **is** 1 Outlier**.** The distribution **is** Left skewed**.**

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?

np. array= [108, 110, 123, 134, 135, 145, 167, 187, 199]

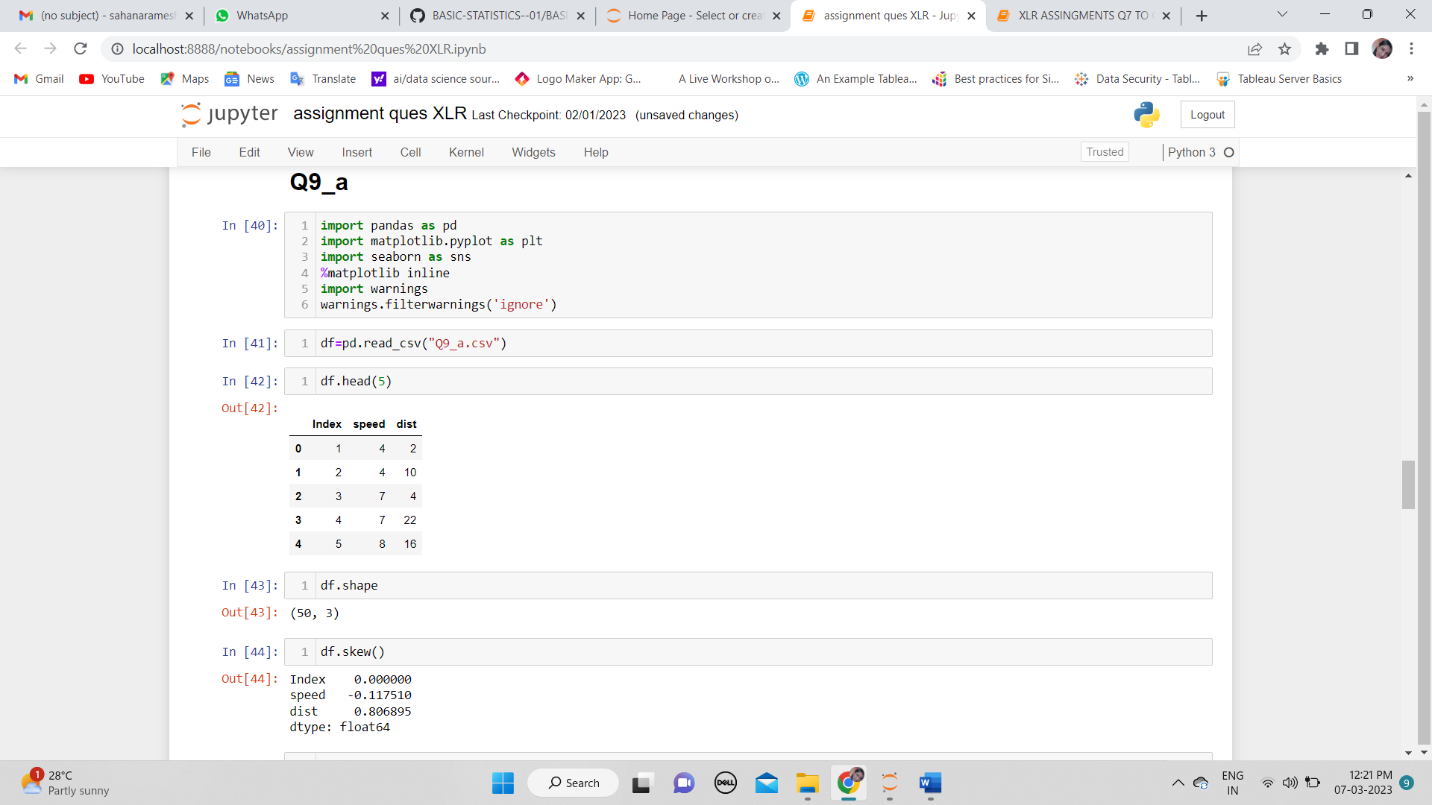
np. mean () = 145.33

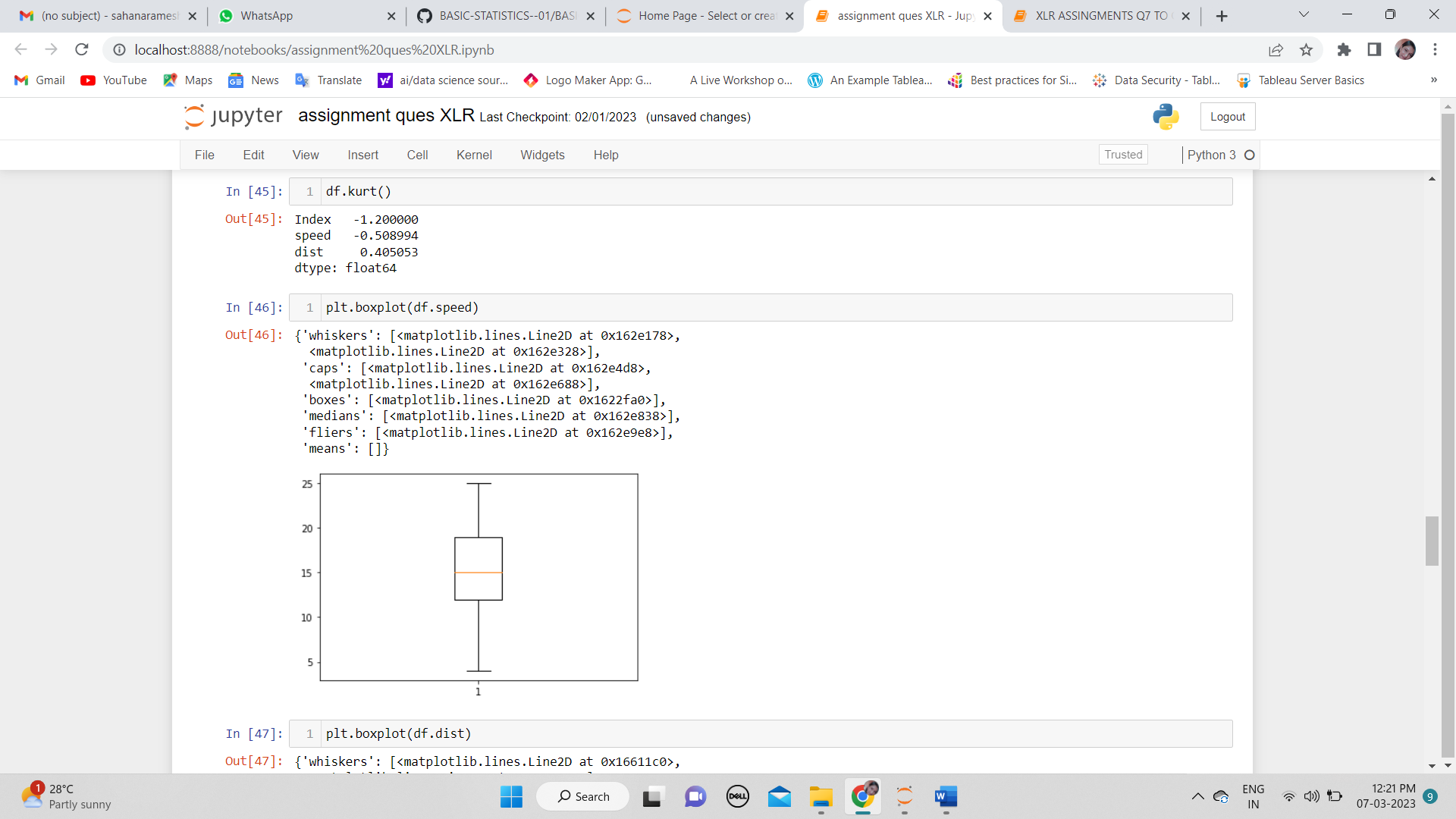
**Q9) Calculate Skewness, Kurtosis & draw inferences on the following data**

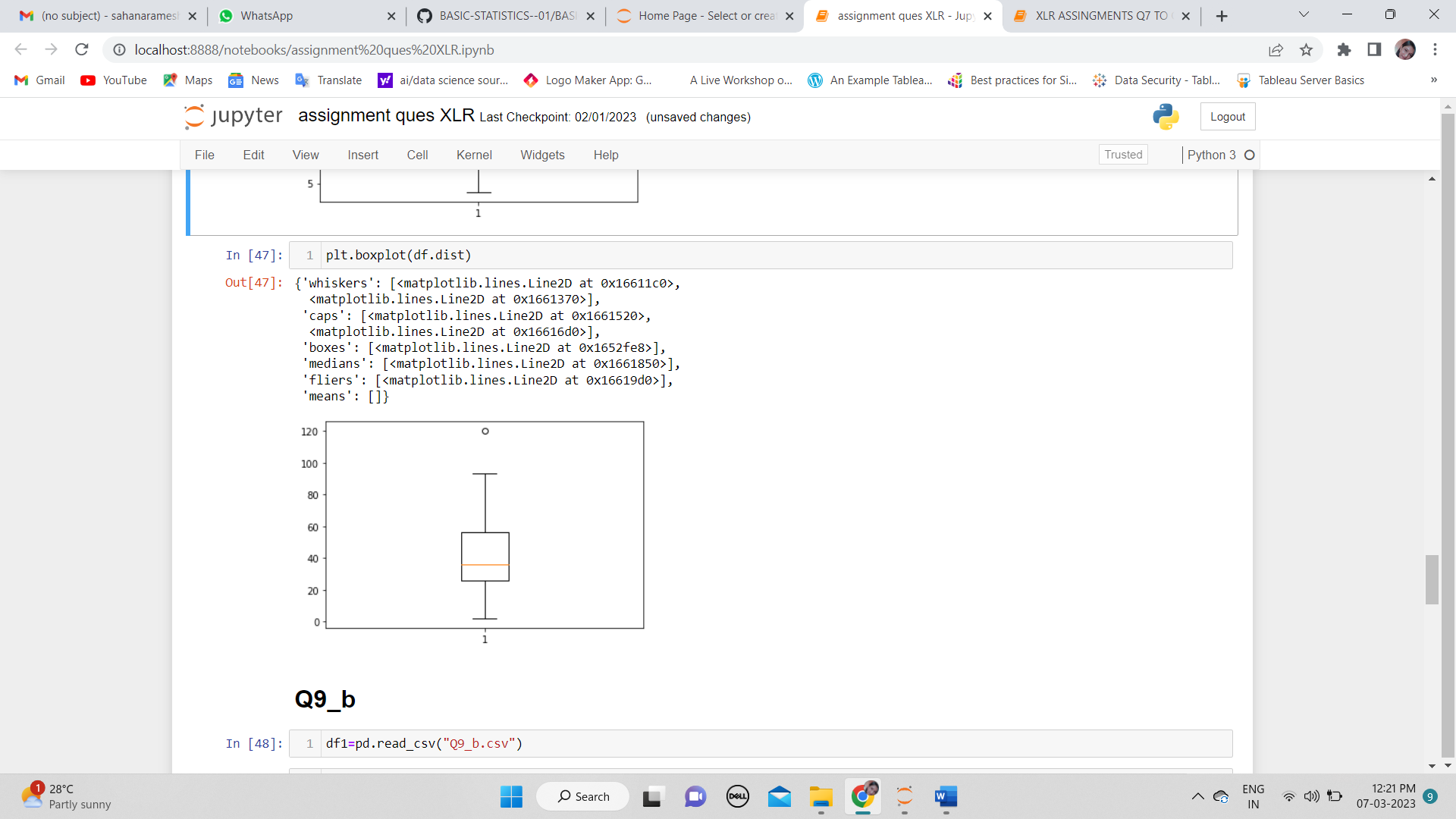
**Cars speed and distance**

**Use Q9\_a.csv**

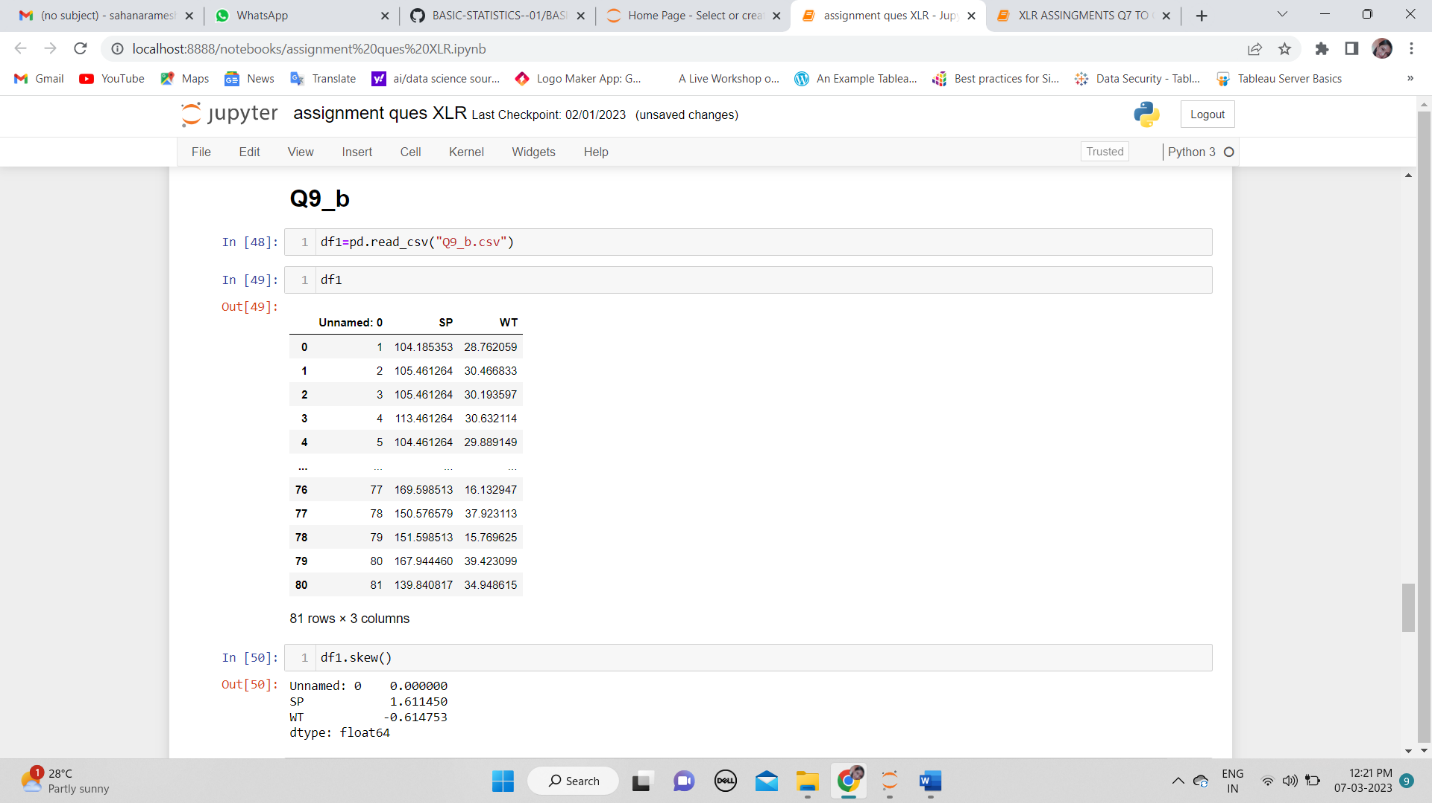
**SP and Weight(WT)**

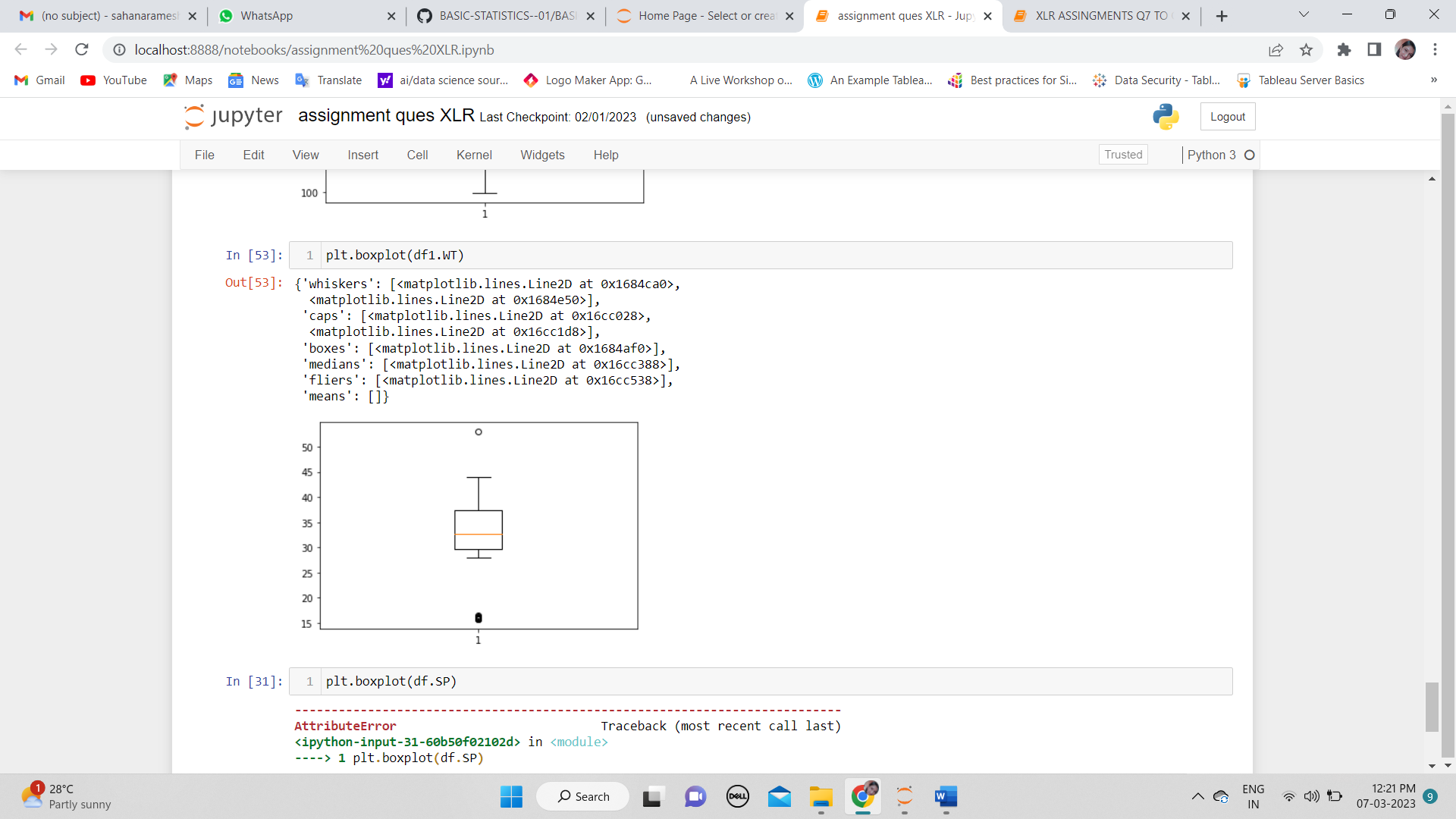
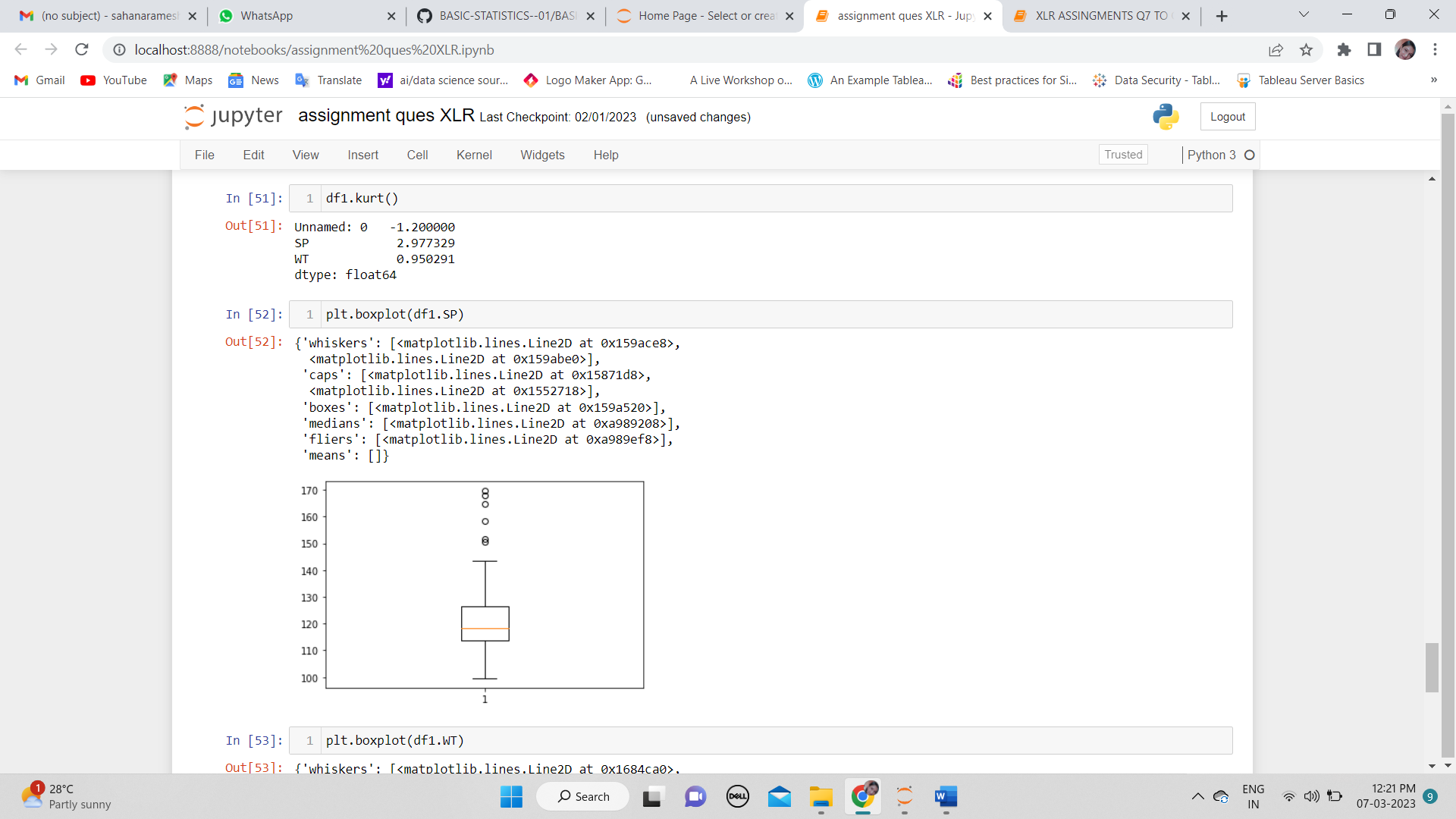
**Use Q9\_b.csv**

****

****

ANS: In speed and distance we can observe that Speed has negative skewness (negative value) that is left skewed. Distance has positive skewness (positive value) that is right skewed

****

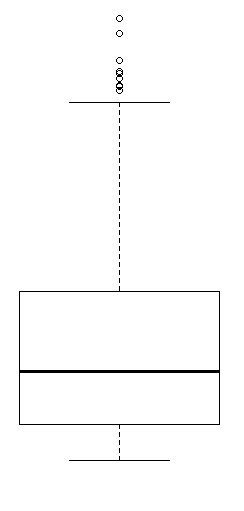
****

ANS: In SP and WT we notice that SP has more outliers, the values are positive so it is right skewed and positive skewness. For WT there is only one outlier, the value is negative so it lies in left skewed and it is negative skewness

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



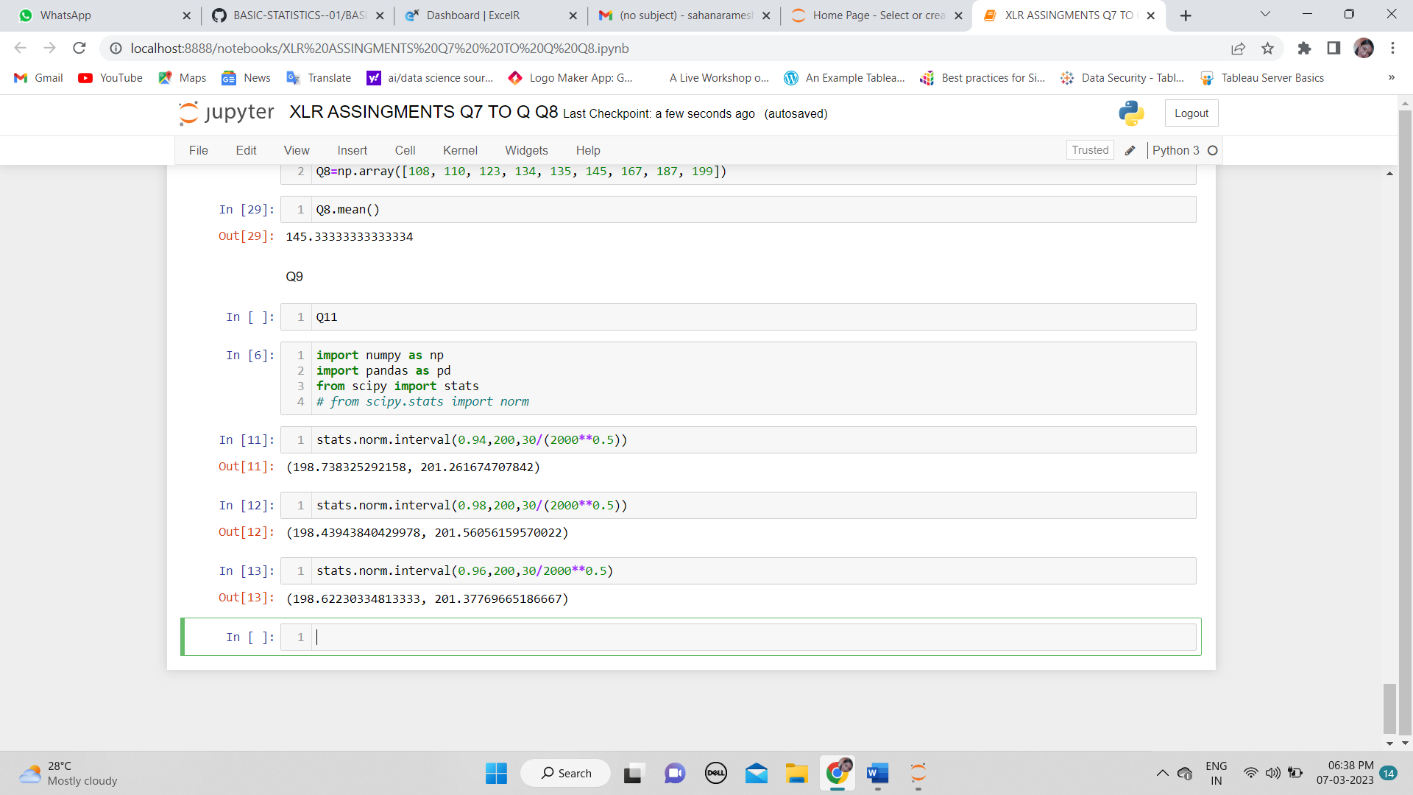
the data points lies in thee range of 50-100 with frequency 200. And least range of wieght is 400 somewhere around 0-10. So, the expected values will be 75. Skewnees- we can see the long tail towards the right so it is stronglyright skewed

median is lesser than means right skewed and we have outlier on the upper side of the box plot and there is less data points bw 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: n= 2000, avg mean is 200, S=30.

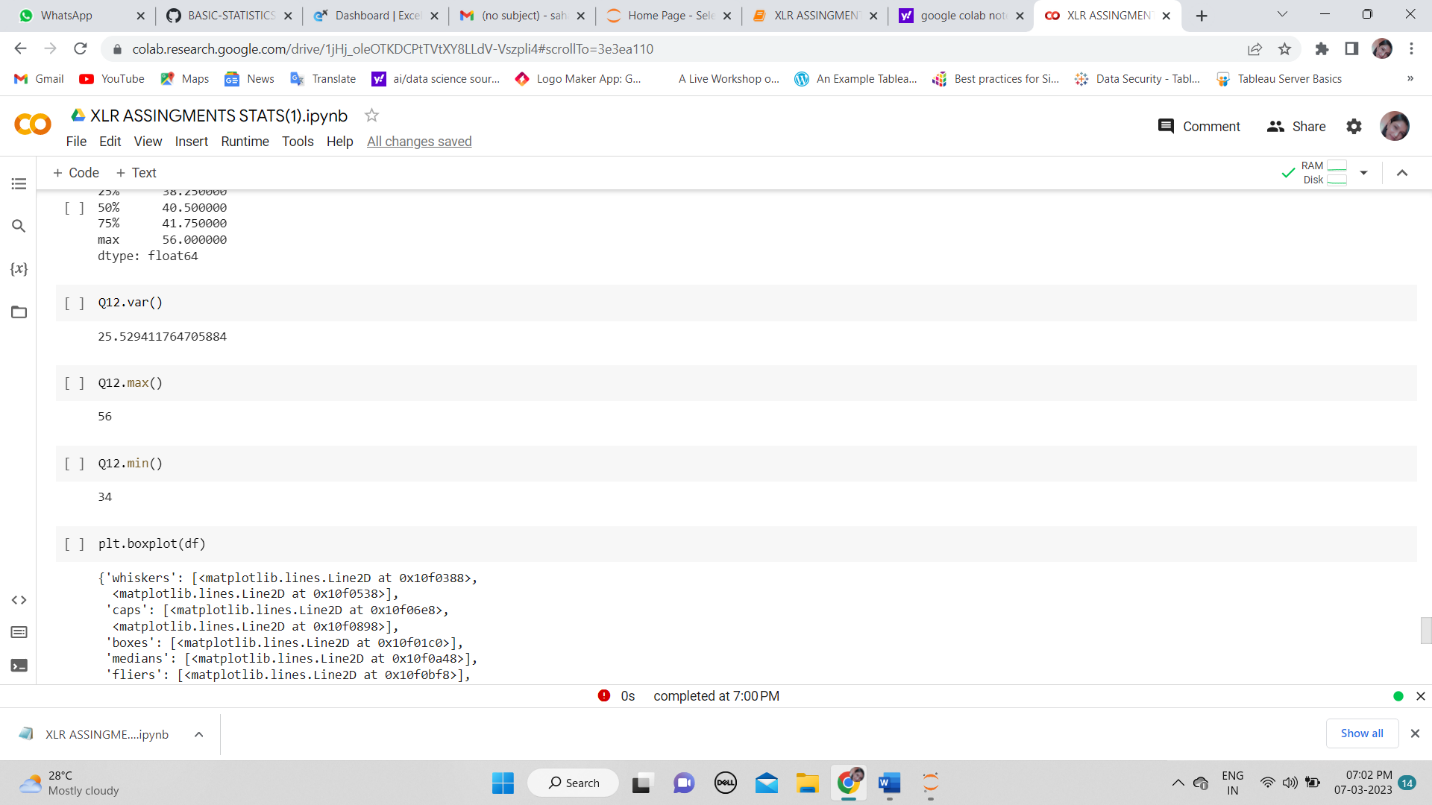
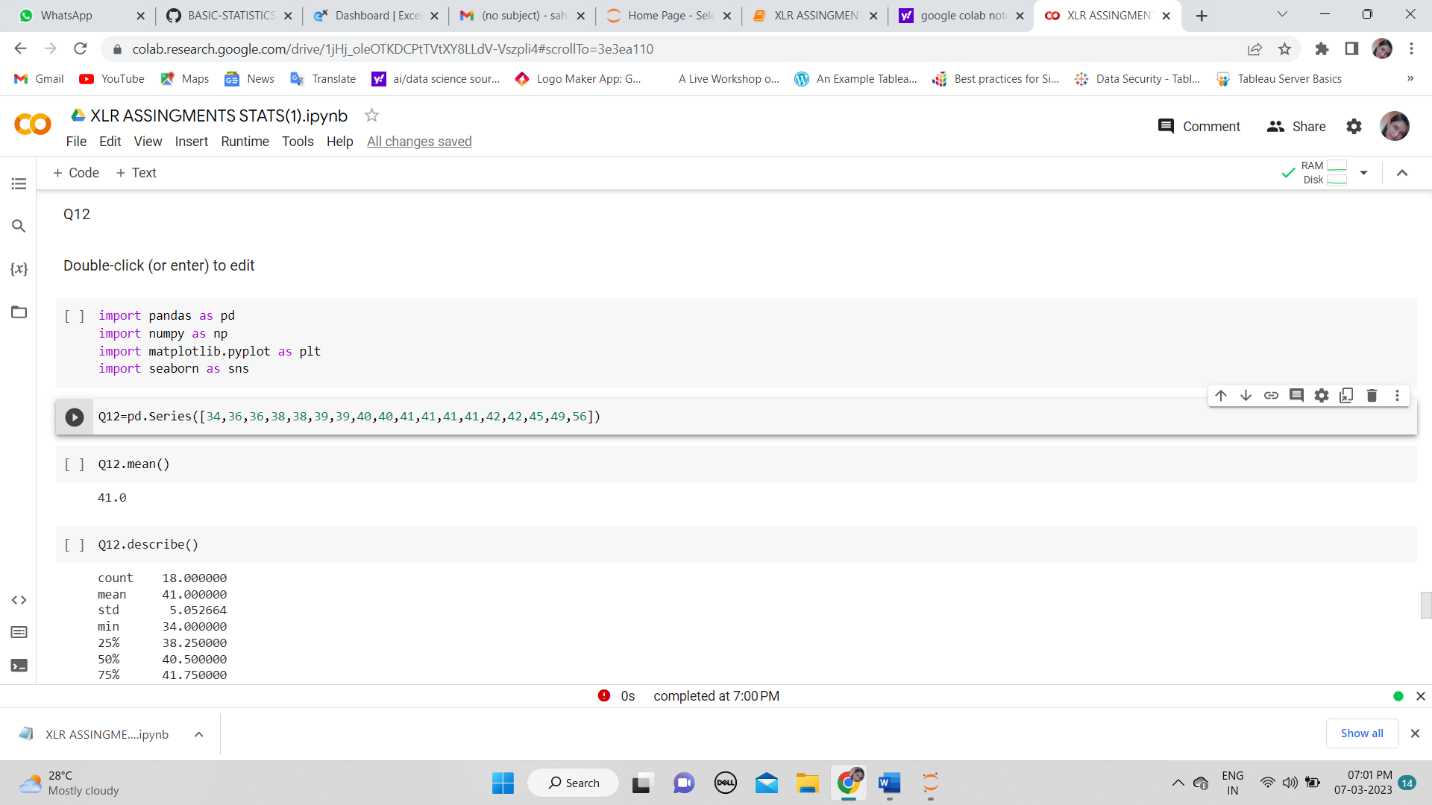
X-2s/n >=mean>=-x-2s/n

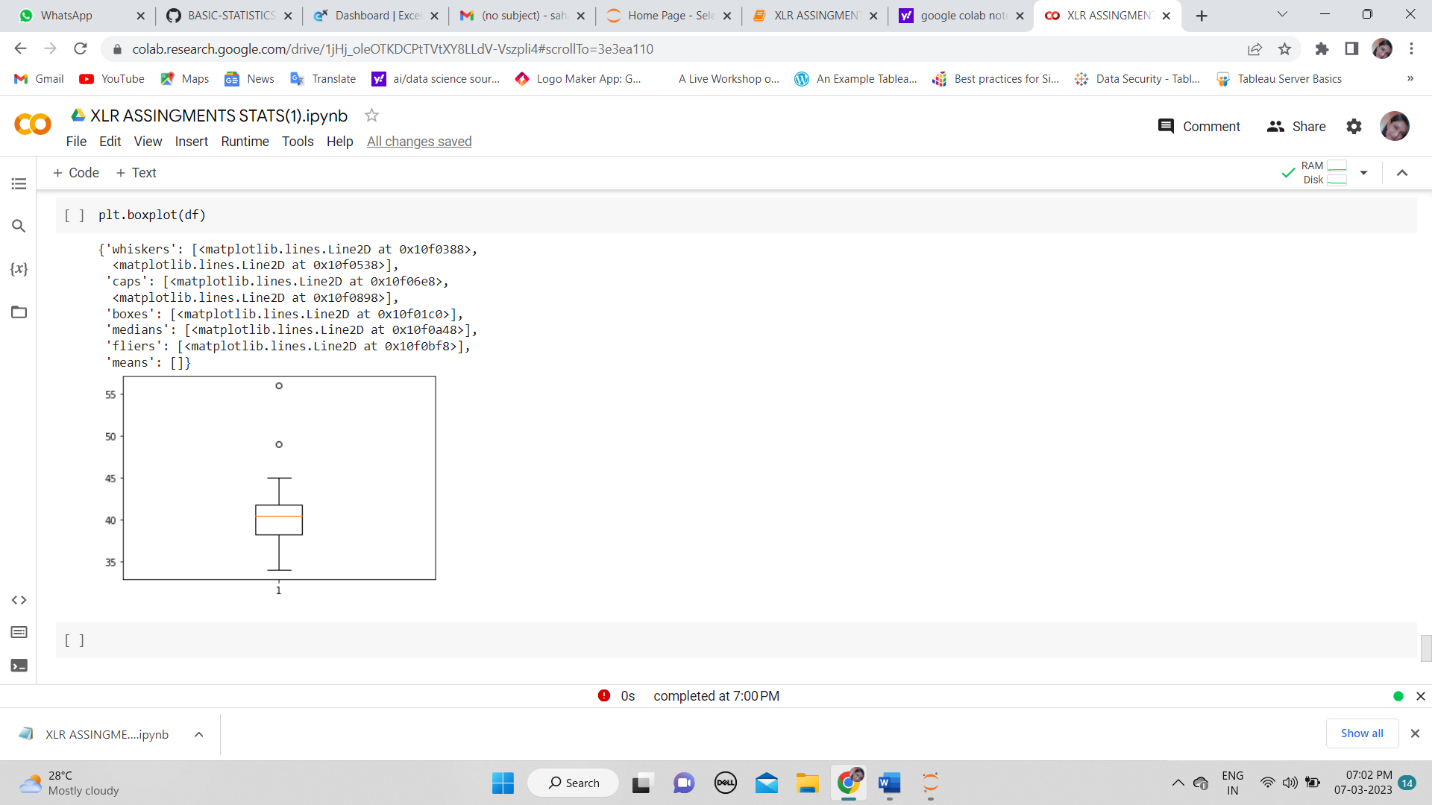
200-2(30)/2000= 200-1.34= 198.66 

**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: We can say that the maximum student marks is 56 and min is 34. The outliers is above 45 to 55. The majority student scores between 45 and 60

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

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

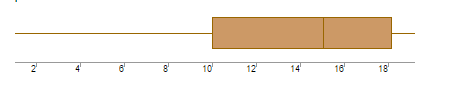
Q15) What is the nature of skewness when median > mean? negative

Q16) What does positive kurtosis value indicates for a data ? when the is normal distributed the value of positive kurtosis is 0

Q17) What does negative kurtosis value indicates for a data? It has thinner tails and flat distribution

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

What can we say about the distribution of the data?



ANS: The distribution is low to high

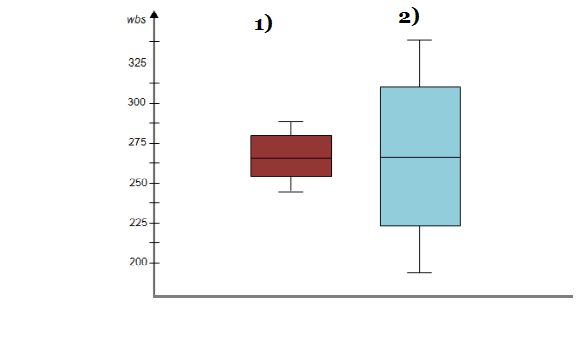
What is nature of skewness of the data?

ANS: The data points is lying between 10 to 18. So, the mean will be 15. It is negative skewness because median >mean

What will be the IQR of the data (approximately)?   
ANS: IQR= Q3-Q1, Q3=18 and Q1 =10, Q2=15.5(mean)

IQR=18-10=8 (approx.)

Q19) Comment on the below Boxplot visualizations?



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

let’s take an example of gender and and salary with respect to 1), 2). 1)is male and 2) is female. So, here we can see 2) female has more range is high (salary) compare to 1) male.

Mean and median are equal. Hence, distribution is symmetrical.

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.

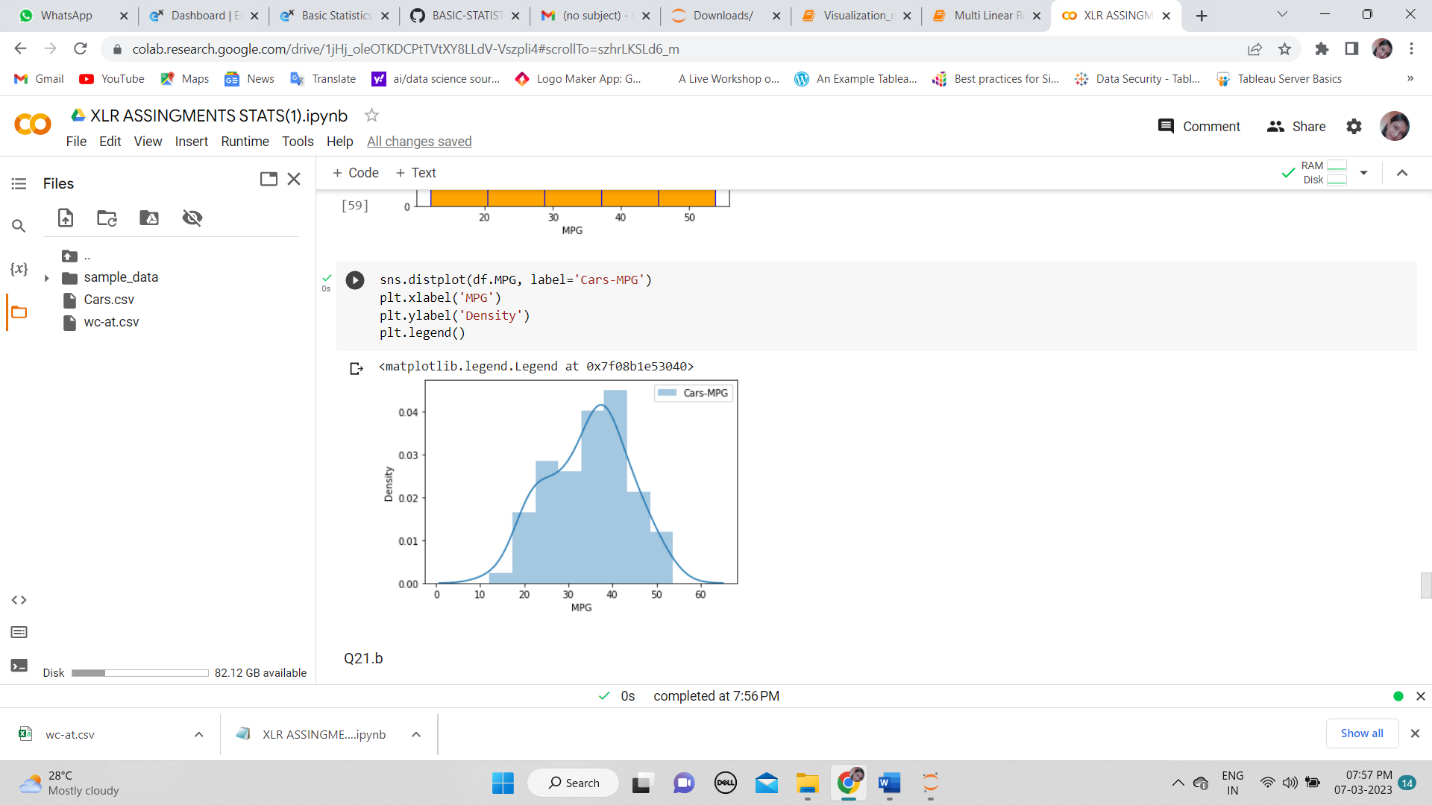
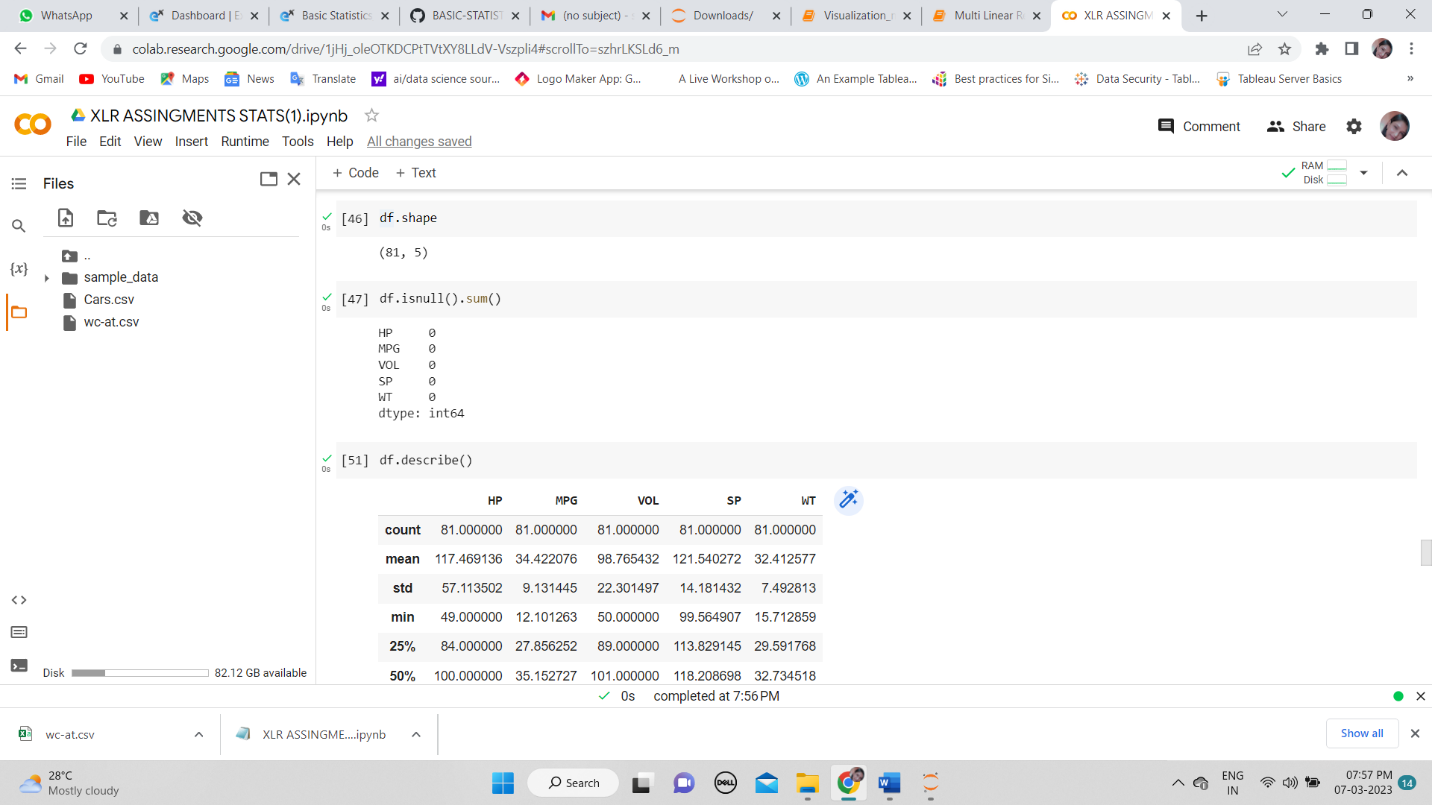
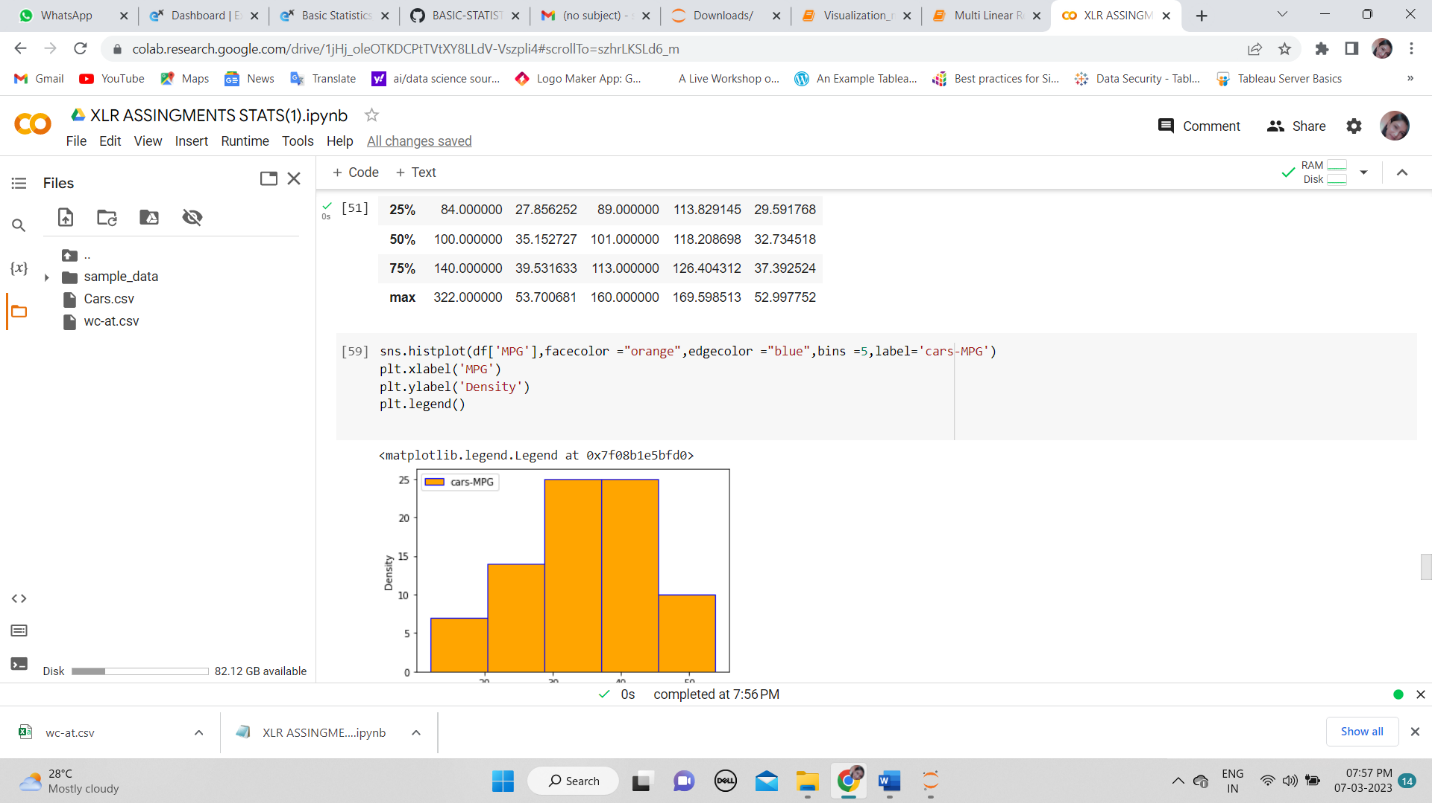
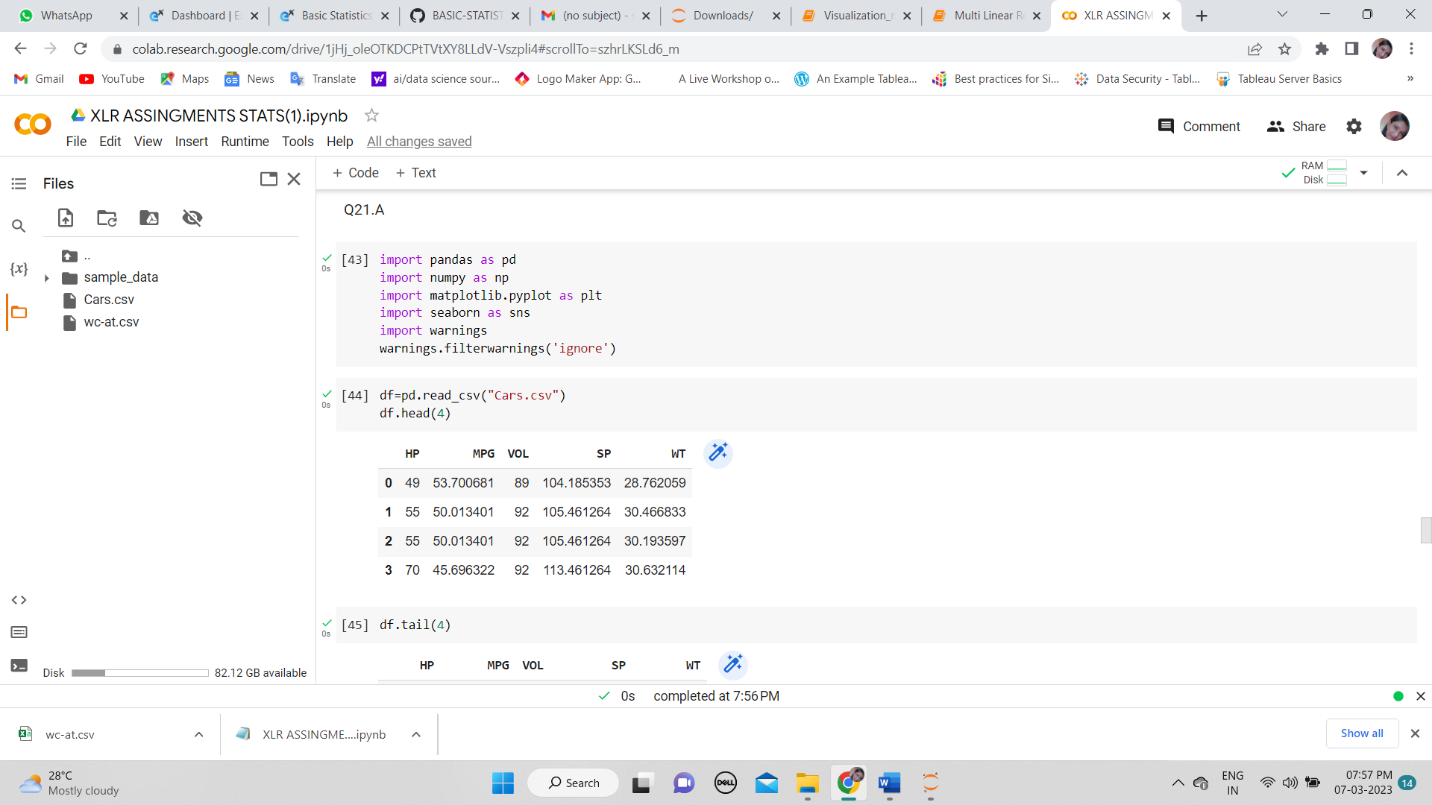
MPG <- Cars$MPG

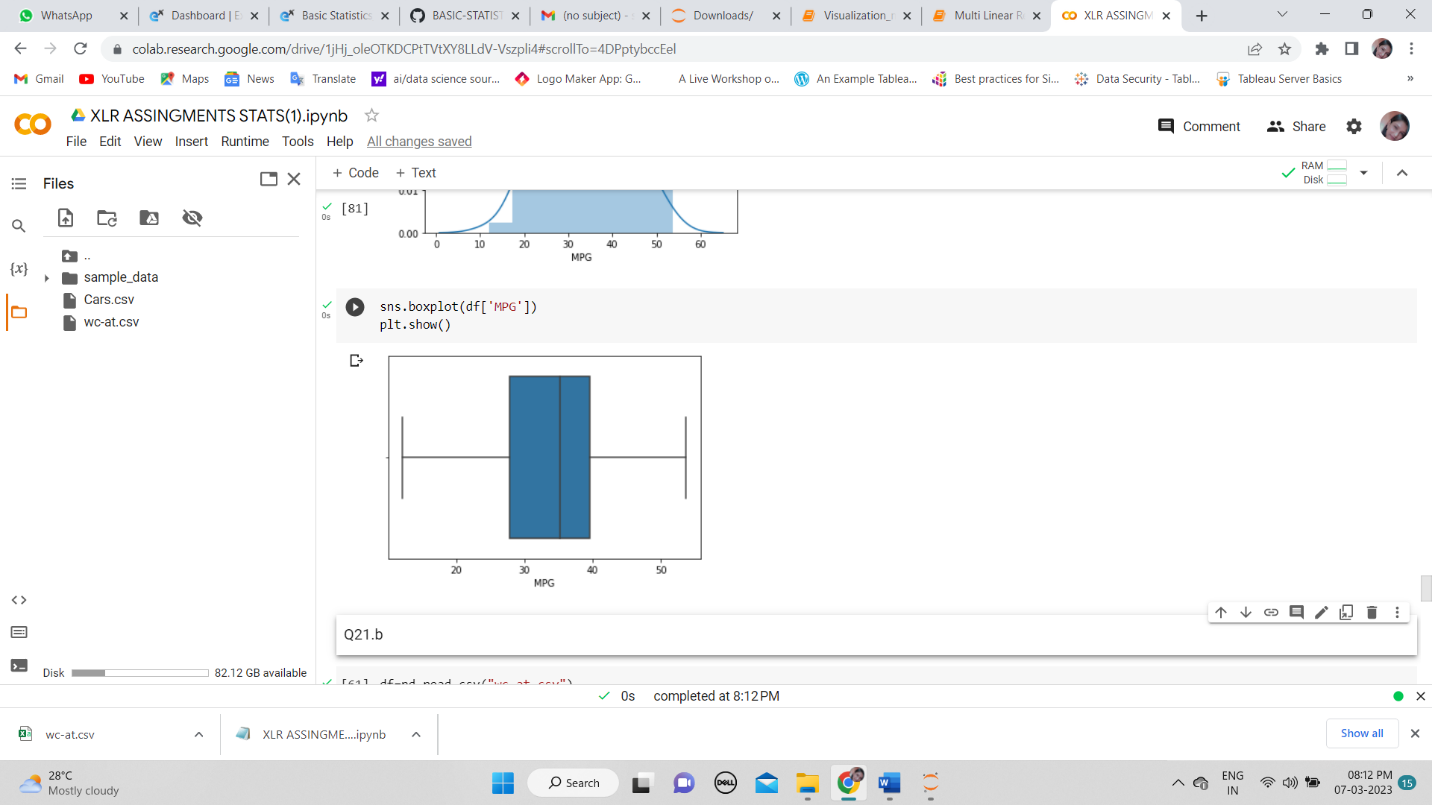
* 1. P(MPG>38) There are 33 observations in MPG >38
  2. P(MPG<40)61 observations<40
  3. P (20<MPG<50) 69 observations>20 and <50

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

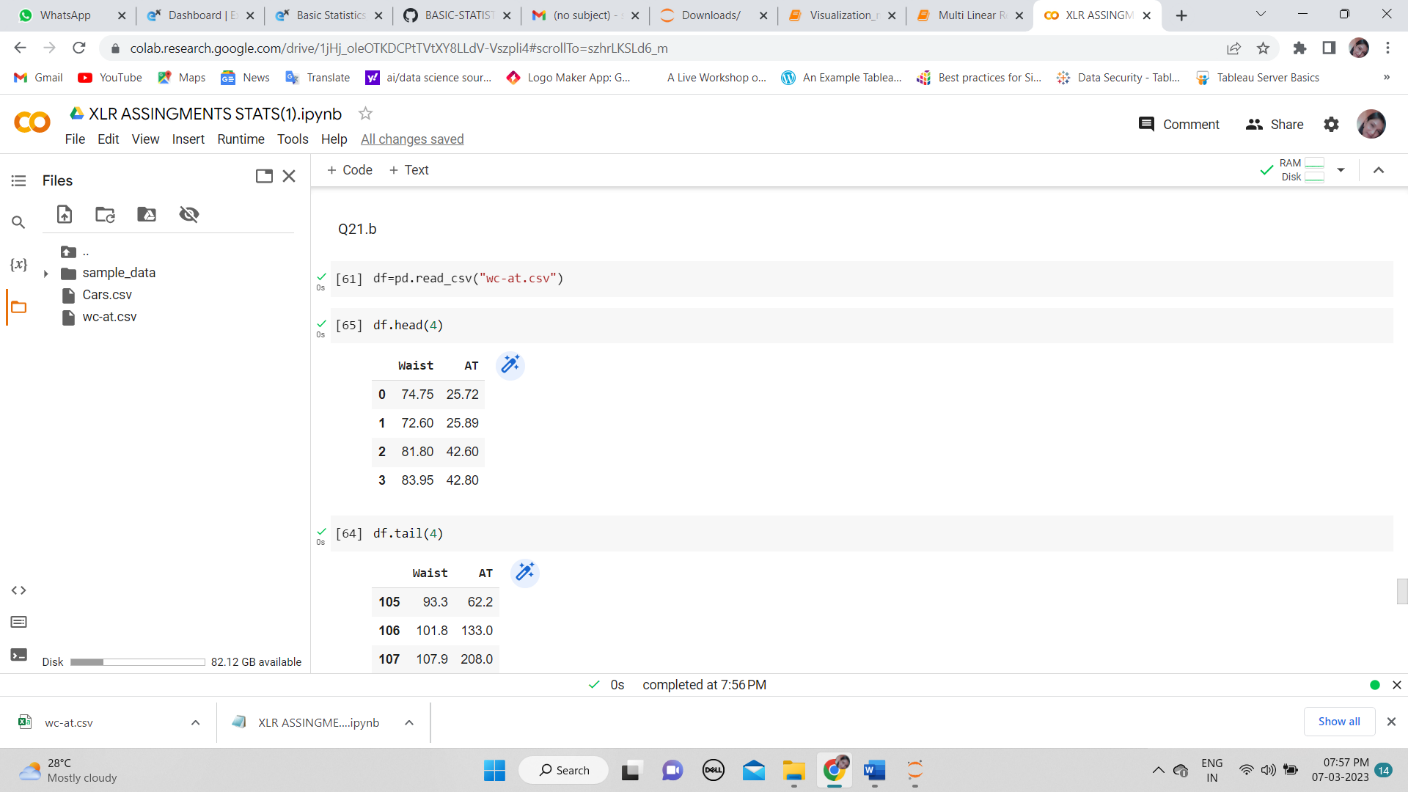
Dataset: Cars.csv

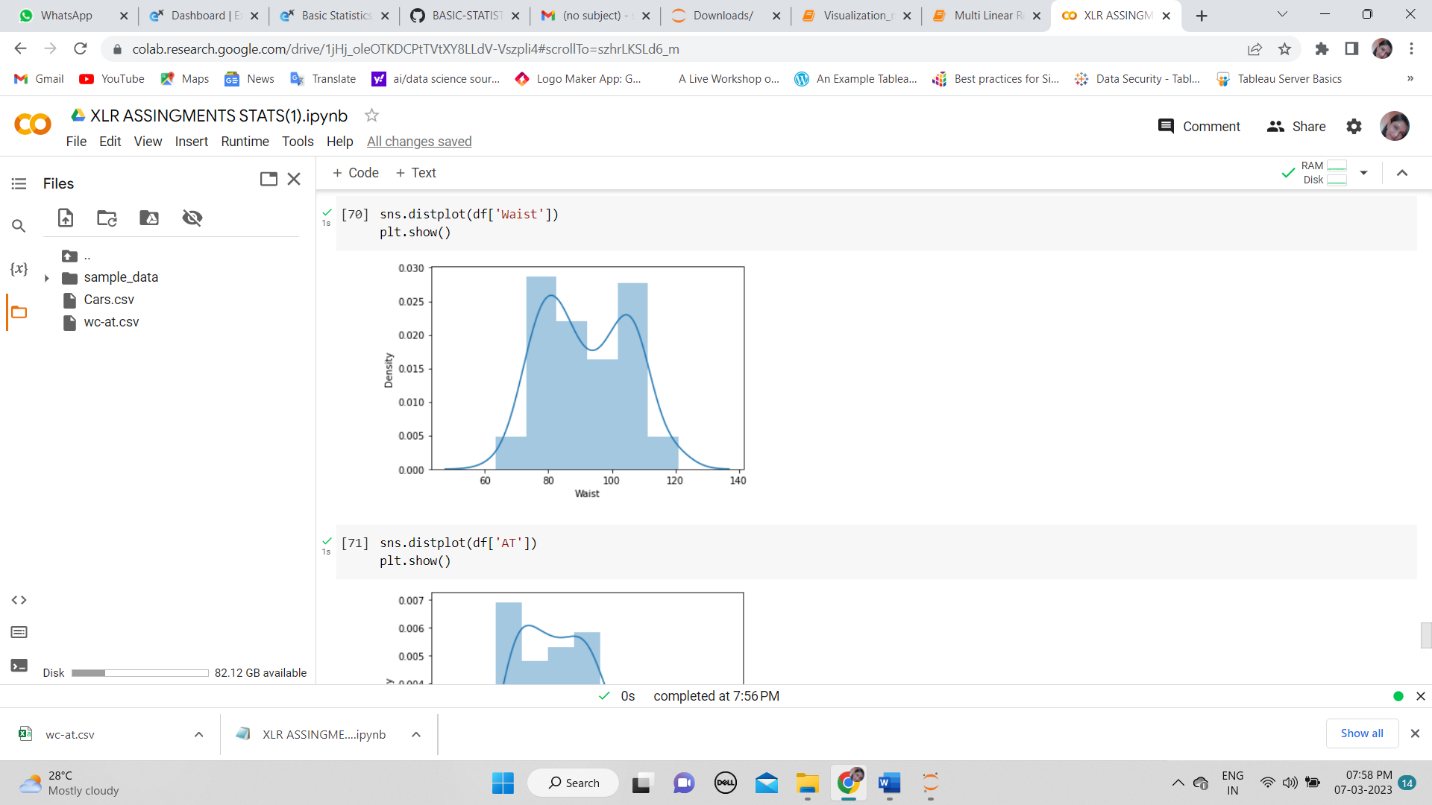
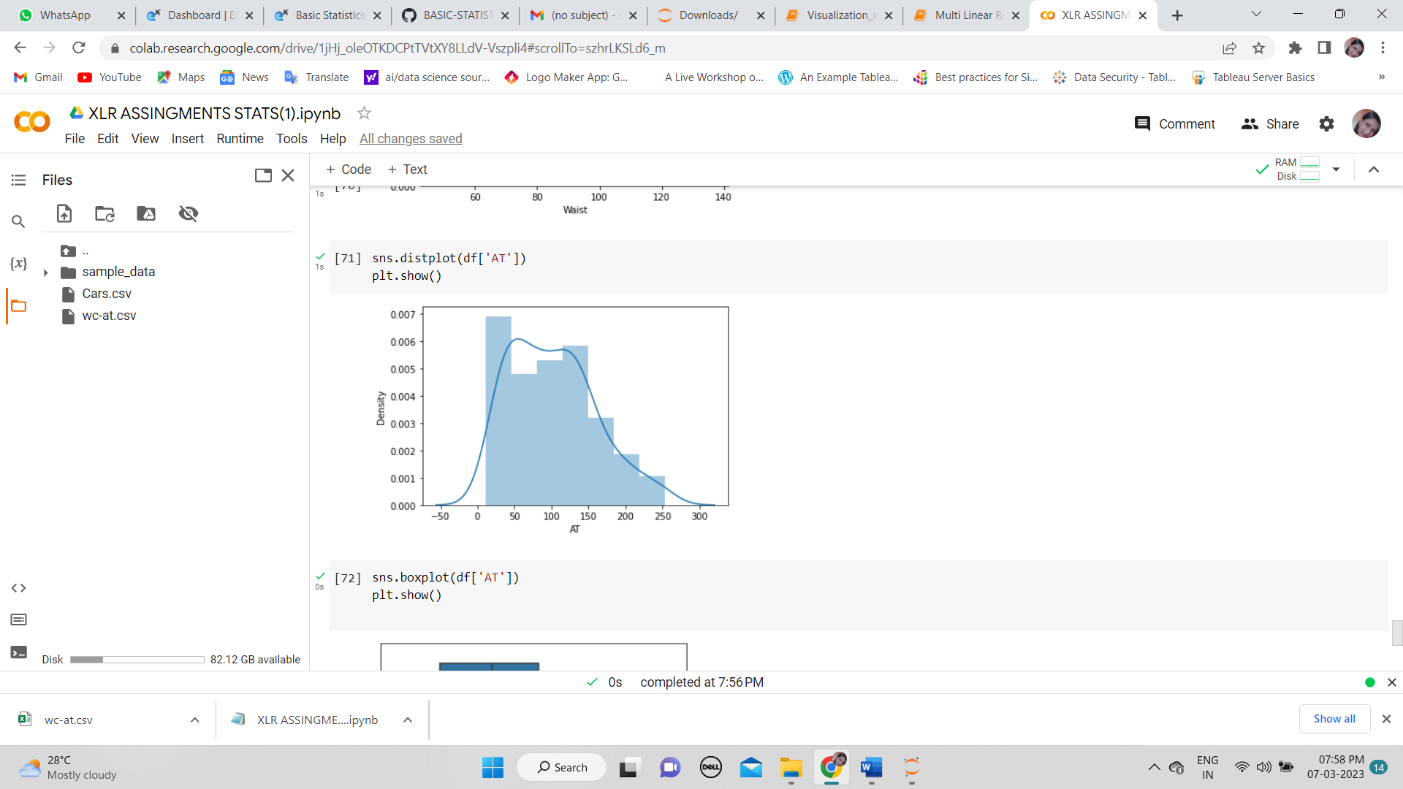
1. 

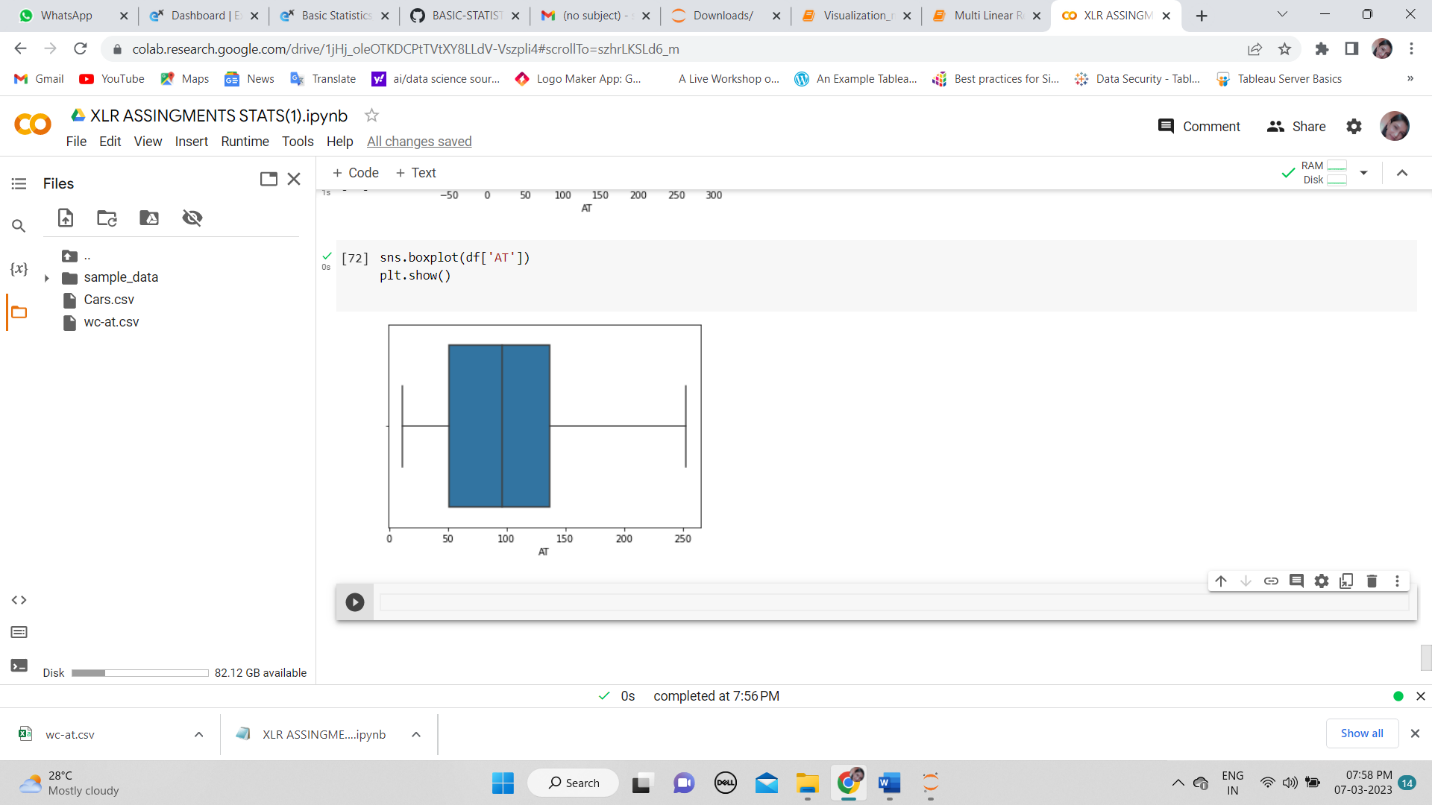
ANS: MPG follows normal distribution , we can observe from the boxplot that mean and median is going to be similar

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

Dataset: wc-at.csv

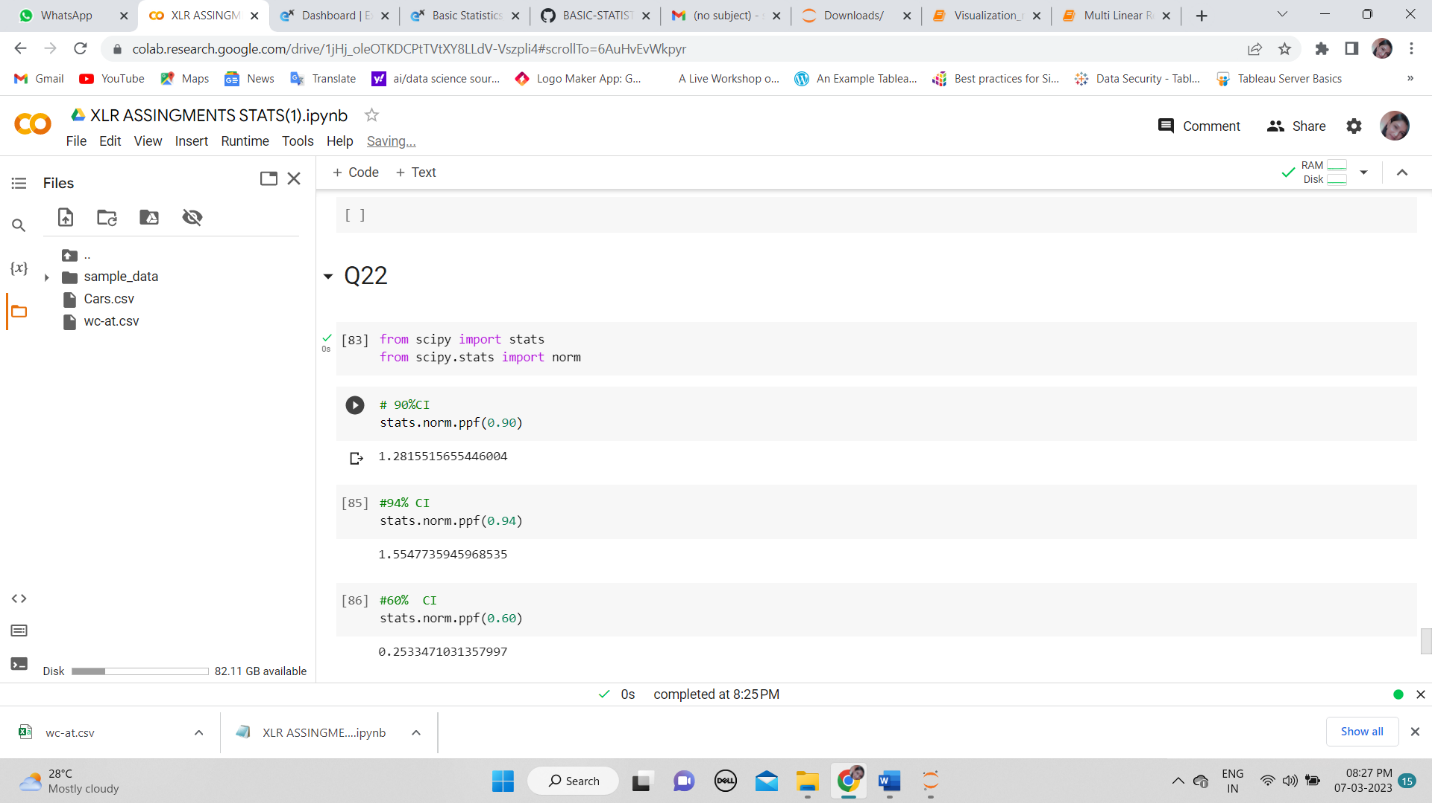




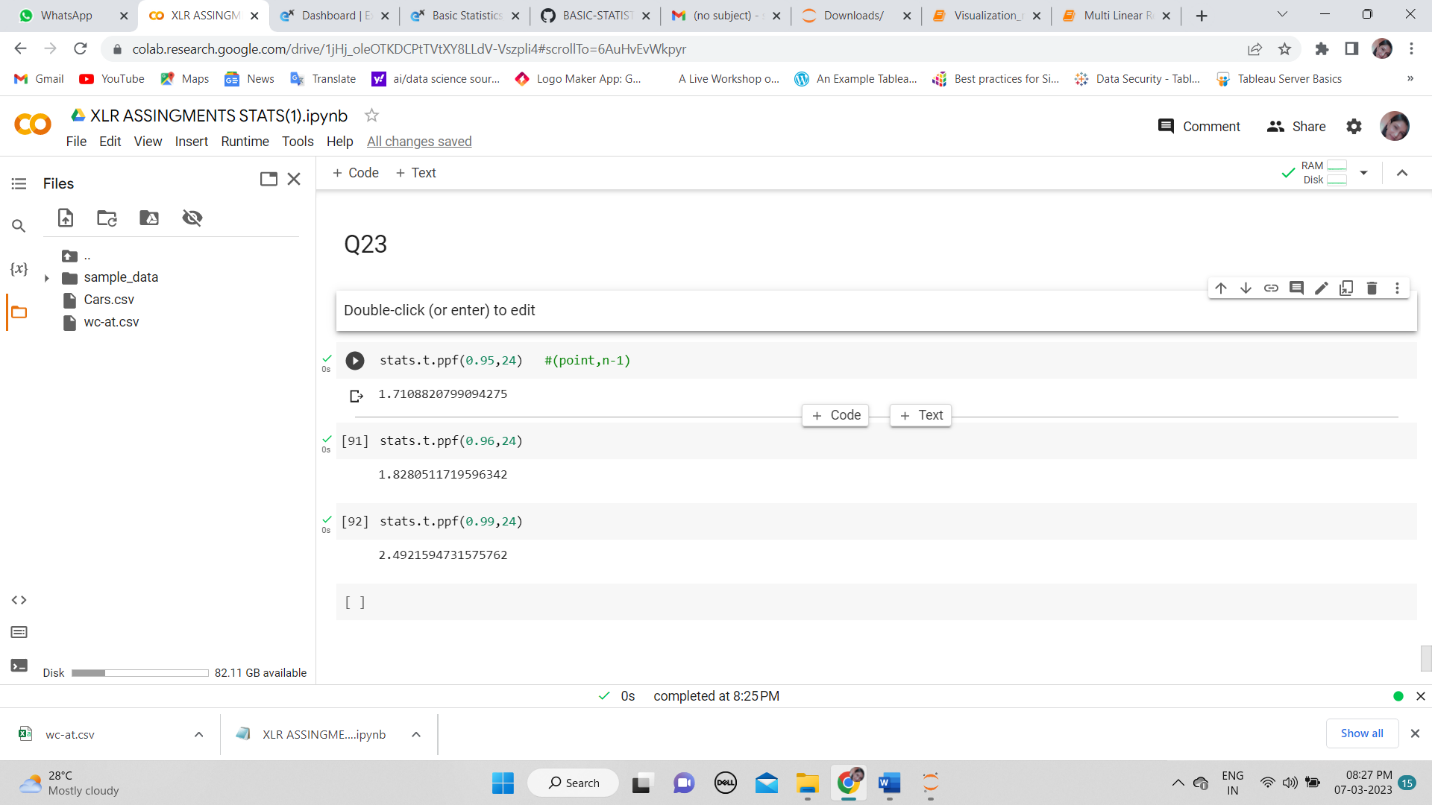


ANS: mean>median, the right side is more so data is positive skew.

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



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

