|  |  |
| --- | --- |
| 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 | Interval |
| Hair Color | Nominal |
| Socioeconomic Status | Nominal |
| Fahrenheit Temperature | Ratio |
| Height | Interval |
| Type of living accommodation | Ordinal |
| Level of Agreement | Nominal |
| IQ(Intelligence Scale) | Interval |
| Sales Figures | Interval |
| Blood Group | Nominal |
| Time Of Day | Ordinal |
| Time on a Clock with Hands | Nominal |
| Number of Children | Nominal |
| 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?

Answer: The number of combinations which have two heads and one tail are:

{HHT, HTH, THH} which makes them 3 in number.

Therefore the Probability of getting two heads and one tails in the toss of three coins simultaneously is defined as:

P(Two heads and One Tail) = Number of desired outcomes

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 2and 3

Answer: a) Total number of outcomes = 36

Number of outcomes sum equal to 1=0

p (equal to 1) = 0/36

b) Total number of outcomes =36 Number of outcomes less than or equal to 4 = 6

p (less than or equal to 4) = 6/36 =1/6

c) Total number of outcomes =36 Number of outcomes sum is divisible by 2 and 3 =6

p (sum is divisible by 2 and 3)=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?

Answer: - There are 7 balls originally with 2 of them blue so the probability of the first ball not being blue is 5/7. This leaves 6 balls with 2 blue. The probability of the second ball not being blue assuming that the first wasn’t is 4/6. The probability that neither ball drawn was blue is (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

Answer: - Anticipated number of candies for a randomly selected child

= 1 \* 0.015 + 4\*0.20 + 3 \*0.65 + 5\*0.005 + 6 \*0.01 + 2 \* 0.12

= 0.015 + 0.8 + 1.95 + 0.025 + 0.06 + 0.24

= 3.090

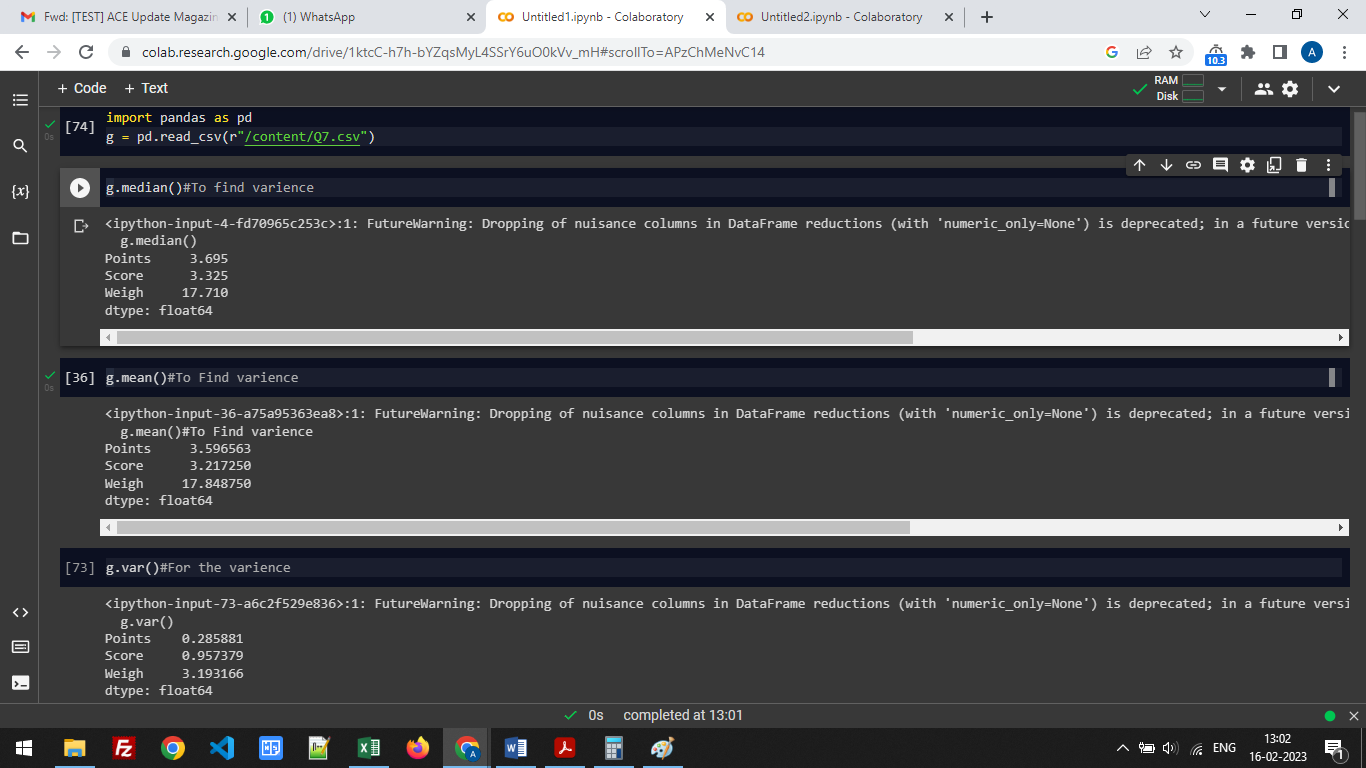
Q7) Calculate Mean, Median, Mode, Variance, Standard Deviation, Range & comment about the values / draw inferences, for the given dataset

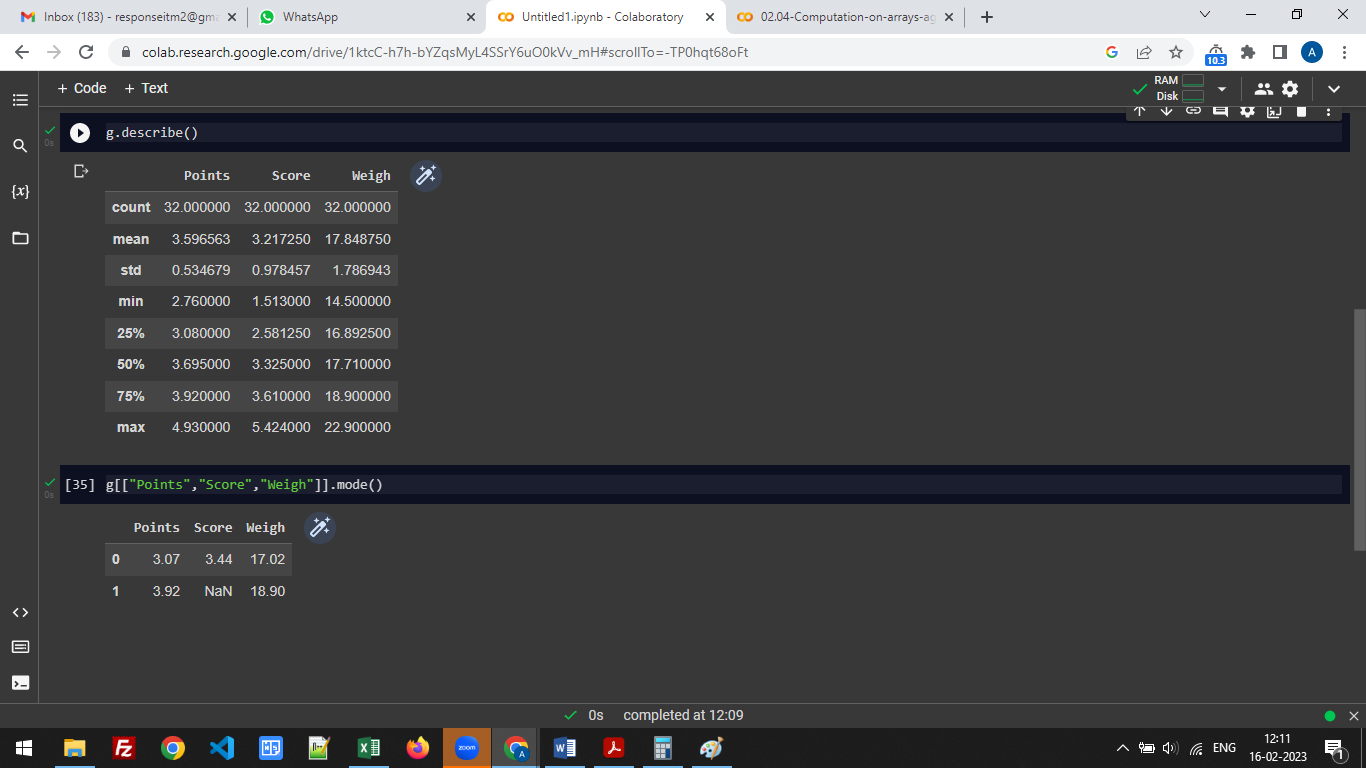
* For Points,Score,Weigh>

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

**Use Q7.csv file**

**Answer:**

****

****

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?

Answer: - Total number of patients = 9

So, the probability of selecting each patient is = 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)

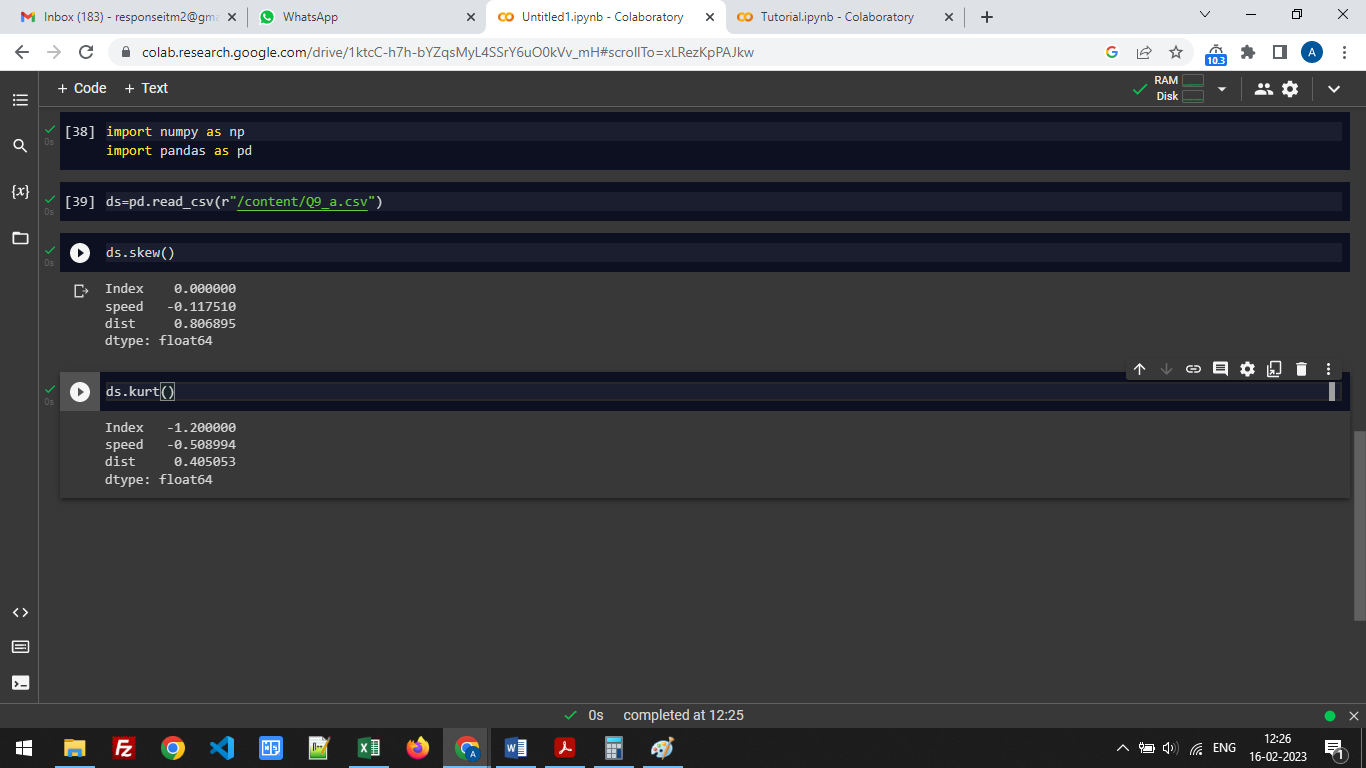
= 145.33

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

**Cars speed and distance**

**Use Q9\_a.csv**

**Answer:-**

****

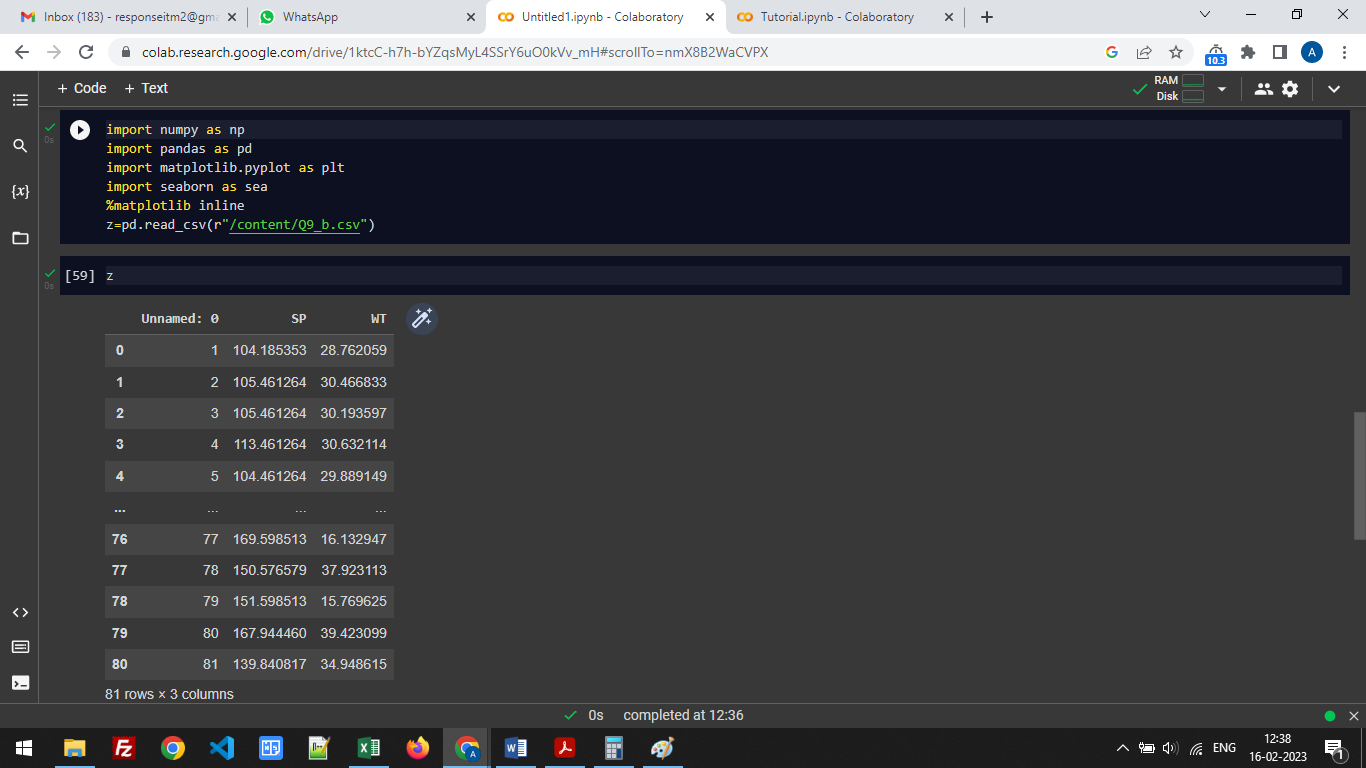
Skewness: speed -0.117510 dist 0.806895

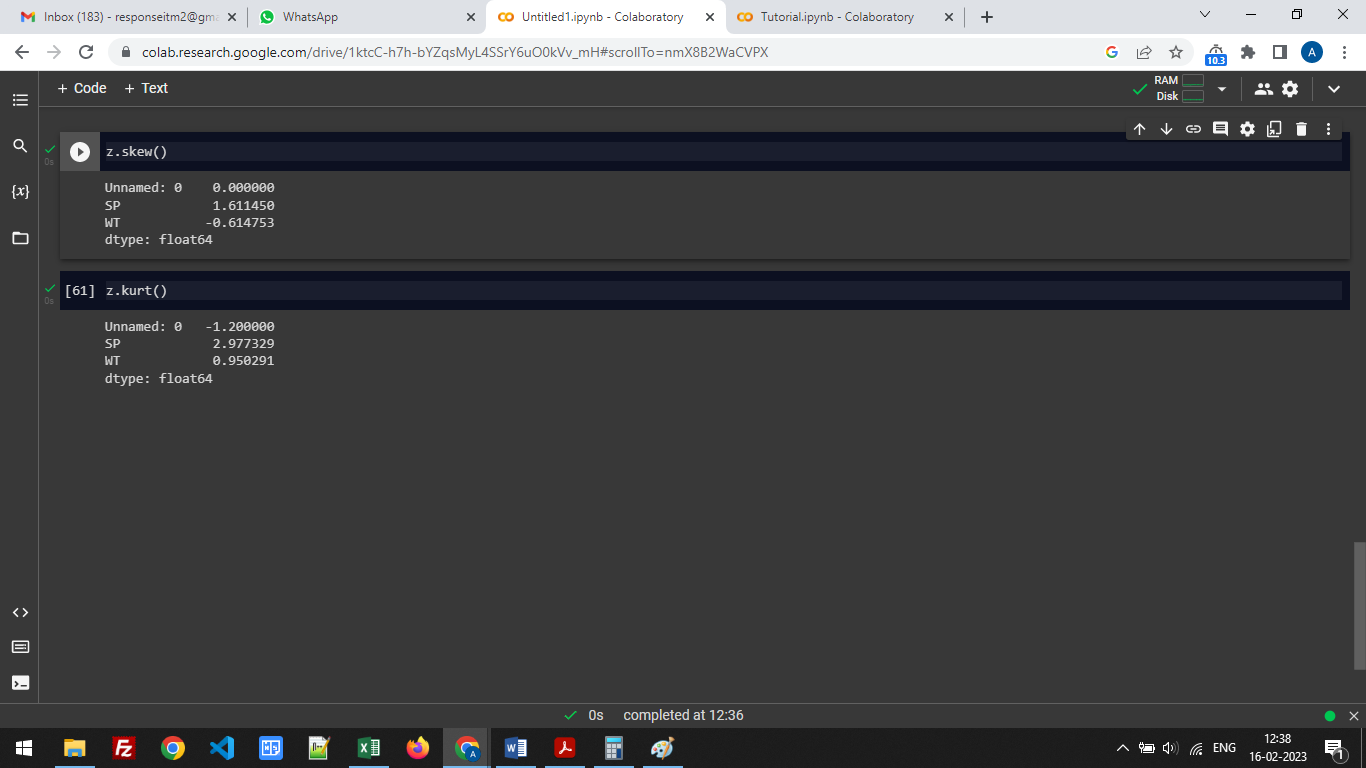
Kurtosis: speed -0.508994 dist 0.405053

**SP and Weight(WT)**

**Use Q9\_b.csv**

**Answer:-**

****

****

Skewness: SP - 1.611450 WT - -0.614753

Kurtosis: SP - 2.977329 WT - 0.950291

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



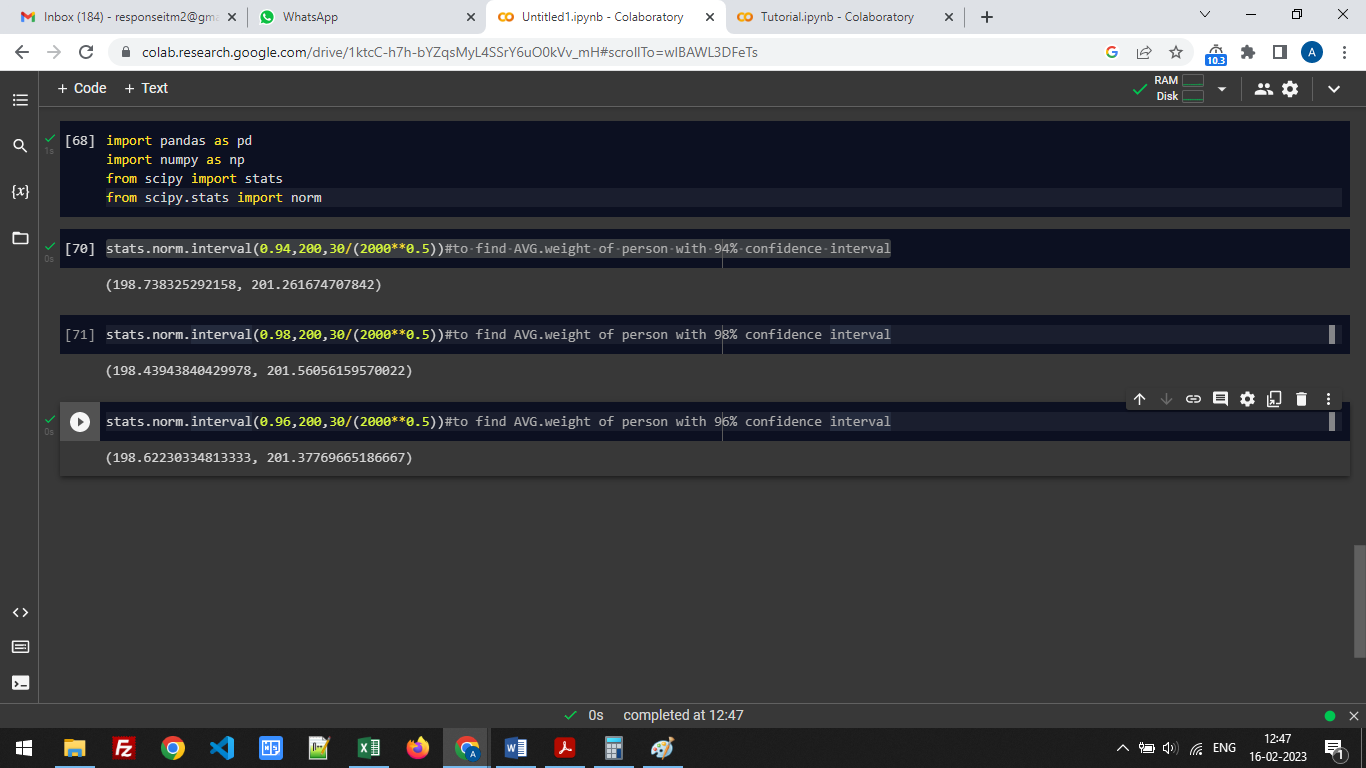


Answer: From the above mention Boxplot & Histogram we conclude that:-

* The Chick weight data is right skewed.
* More than 50% of Chick Weight is between 50 to 150.
* Most of the chick weight is between 50 to 100.
* Outlier is at upper side so there are outliers

**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?

Answer:-



94% of confidence interval

(198.738325292158, 201.261674707842)

98% of confidence interval

(198.43943840429978, 201.56056159570022)

96% of confidence interval

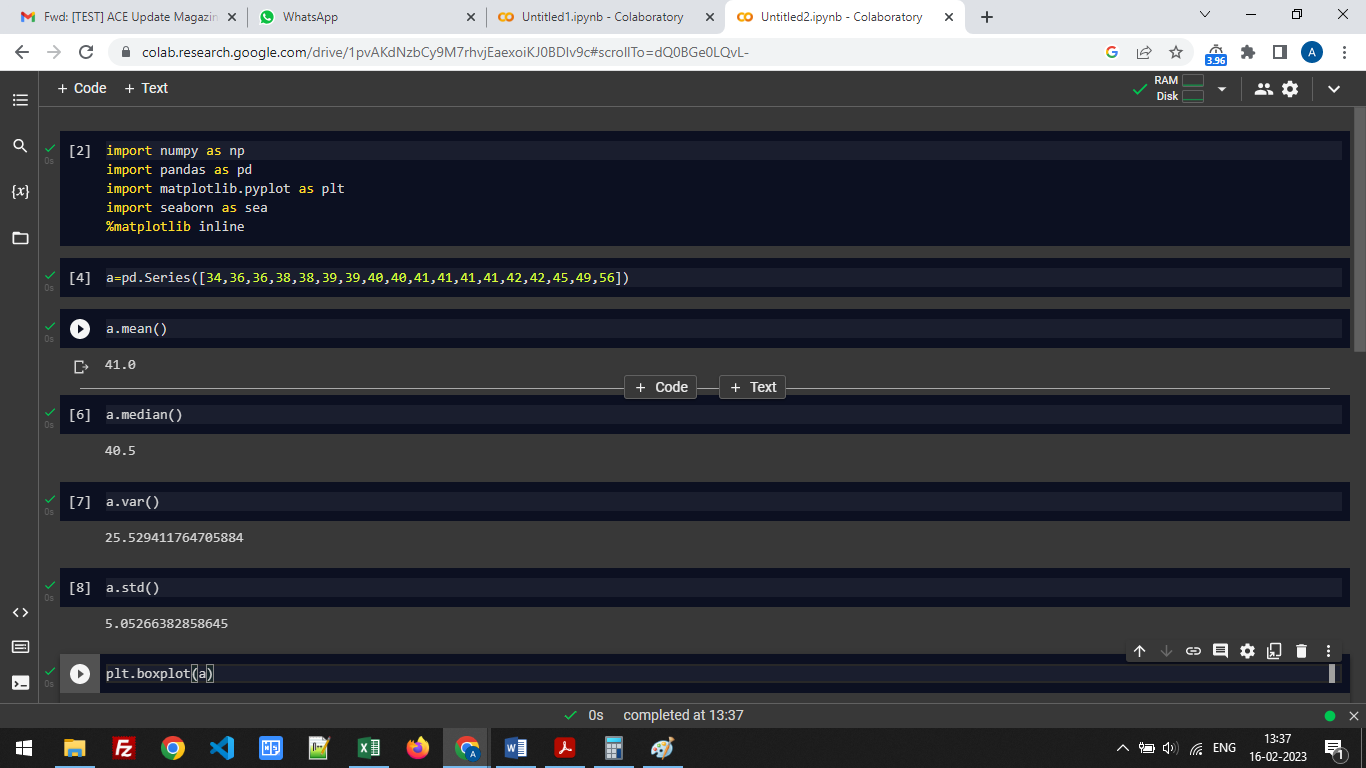
(198.62230334813333, 201.37769665186667)

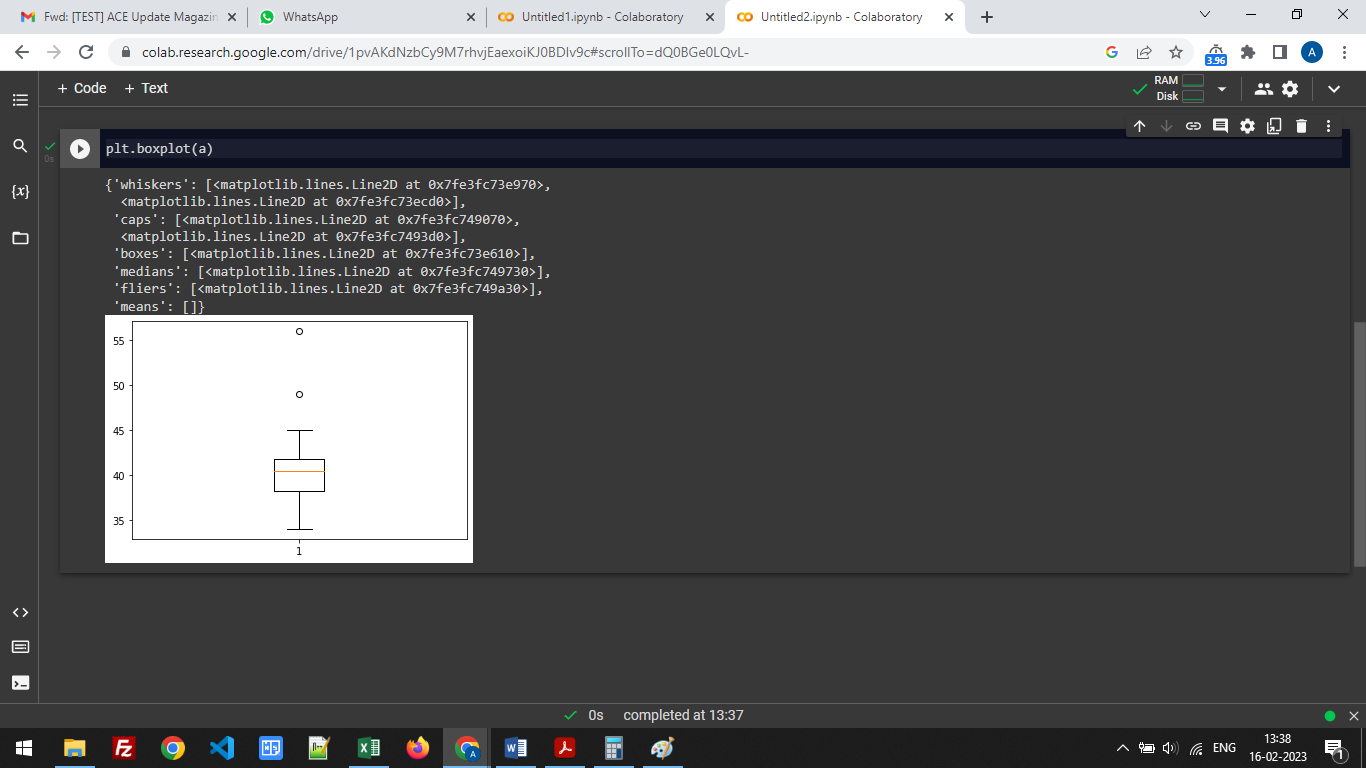
**Q12)**Below are the scores obtained by a student in tests

**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?

Answer:-





Mean = 41, Median = 40.5, variance= 25.529411764705884, Std. Deviation = 5.05266382858645

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

Answer:- The data is normalized as

If the mean is greater than the median, then data is positively skewed.

If the mean is less than the median, the data is negatively skewed.

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

Answer:- If the mean is greater than the median, then data is positively skewed.

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

Answer:- If the mean is less than the median, the data is negatively skewed.

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

Answer: - It means that the distribution is peaked and possesses thick tails.

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

Answer: - It means that the distribution has lighter tails than the normal distribution.

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



What can we say about the distribution of the data?

Answer:- The mentioned boxplot is not normally distributed as the median is to the higher value

What is nature of skewness of the data?

Answer: - The data is skewed to the left, that means the nature of skewness is negative.

What will be the IQR of the data (approximately)?   
Answer:- The IQR of the data is

18-10=8

Q19) Comment on the below Boxplot visualizations?



1. There are no outliers present.

2. 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

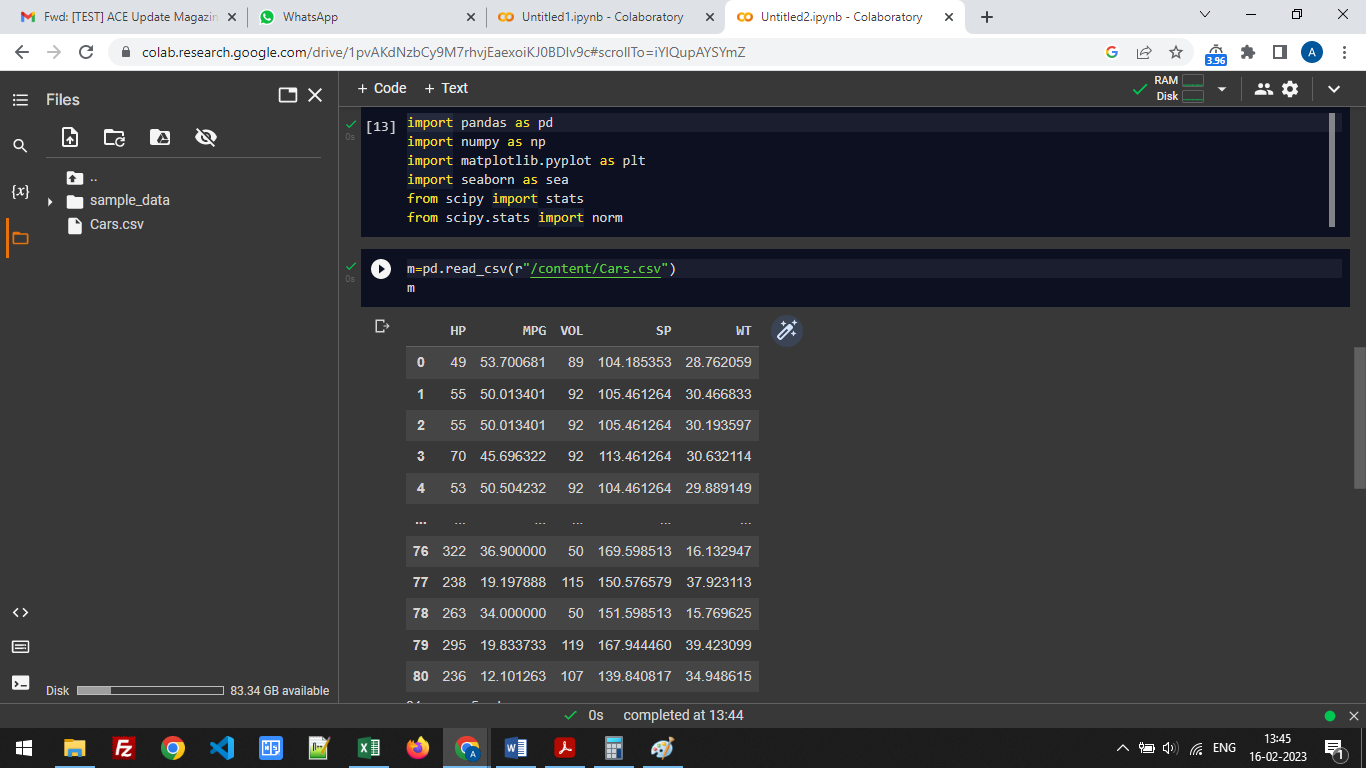
Calculate the probability of MPG ofCars for the below cases.

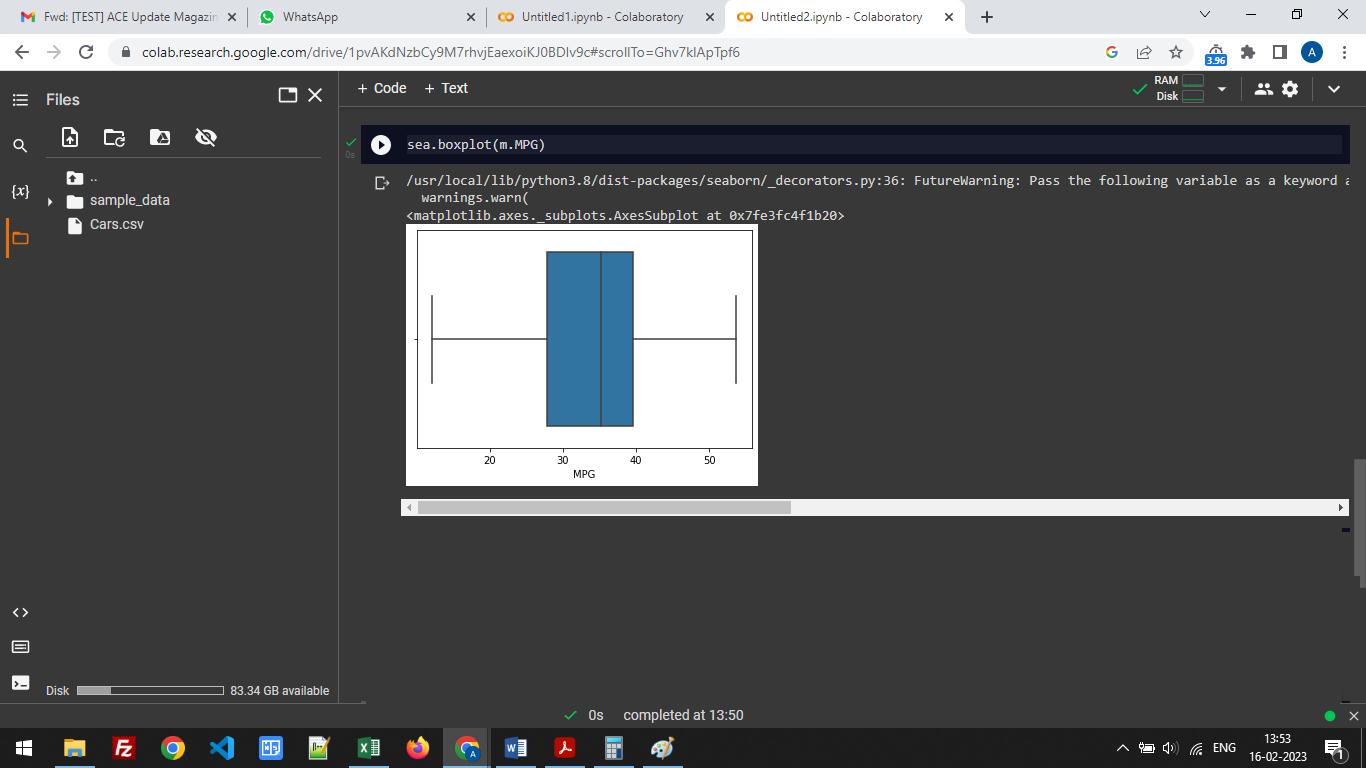
MPG<- Cars$MPG

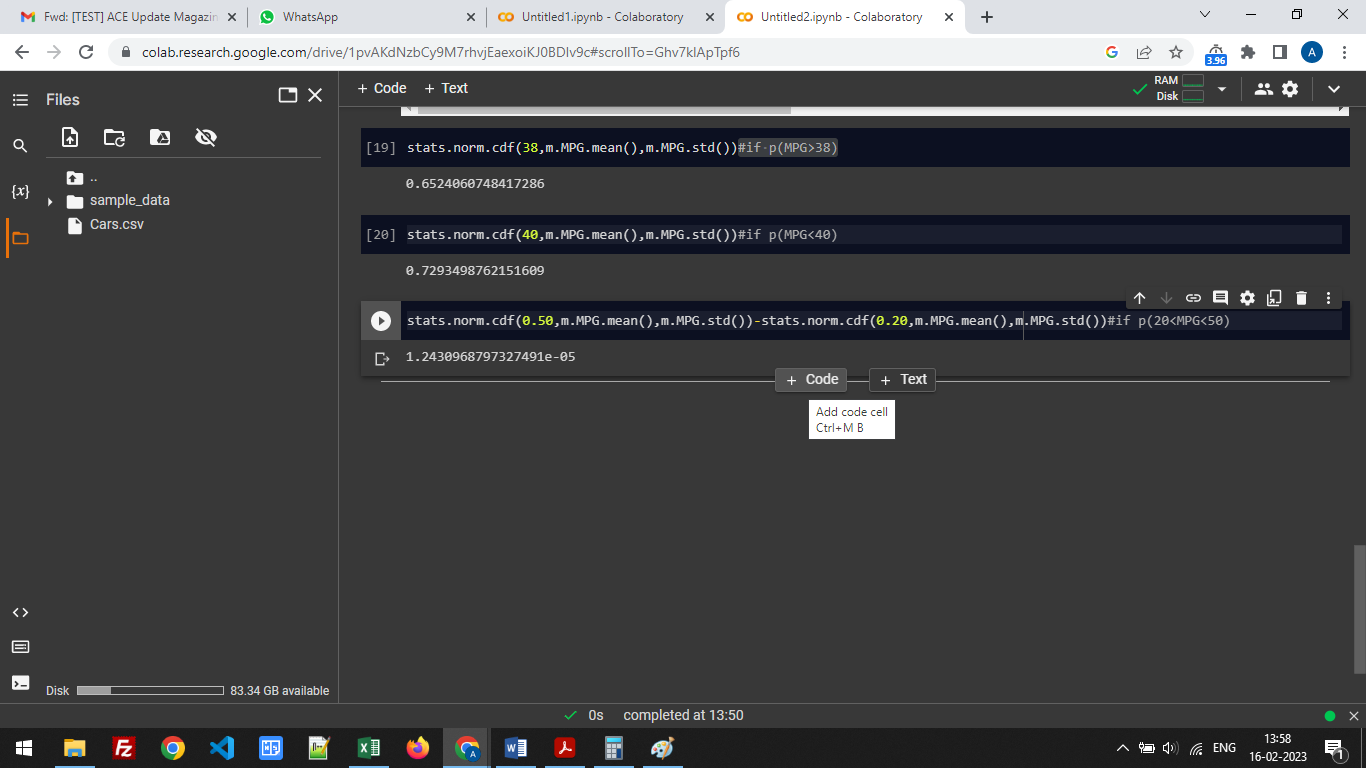
* 1. P(MPG>38)
  2. P(MPG<40)

c. P (20<MPG<50)

Answer:-







1. P(MPG>38) ---- 0.6524060748417286
2. P(MPG<40) ---- 0.7293498762151609

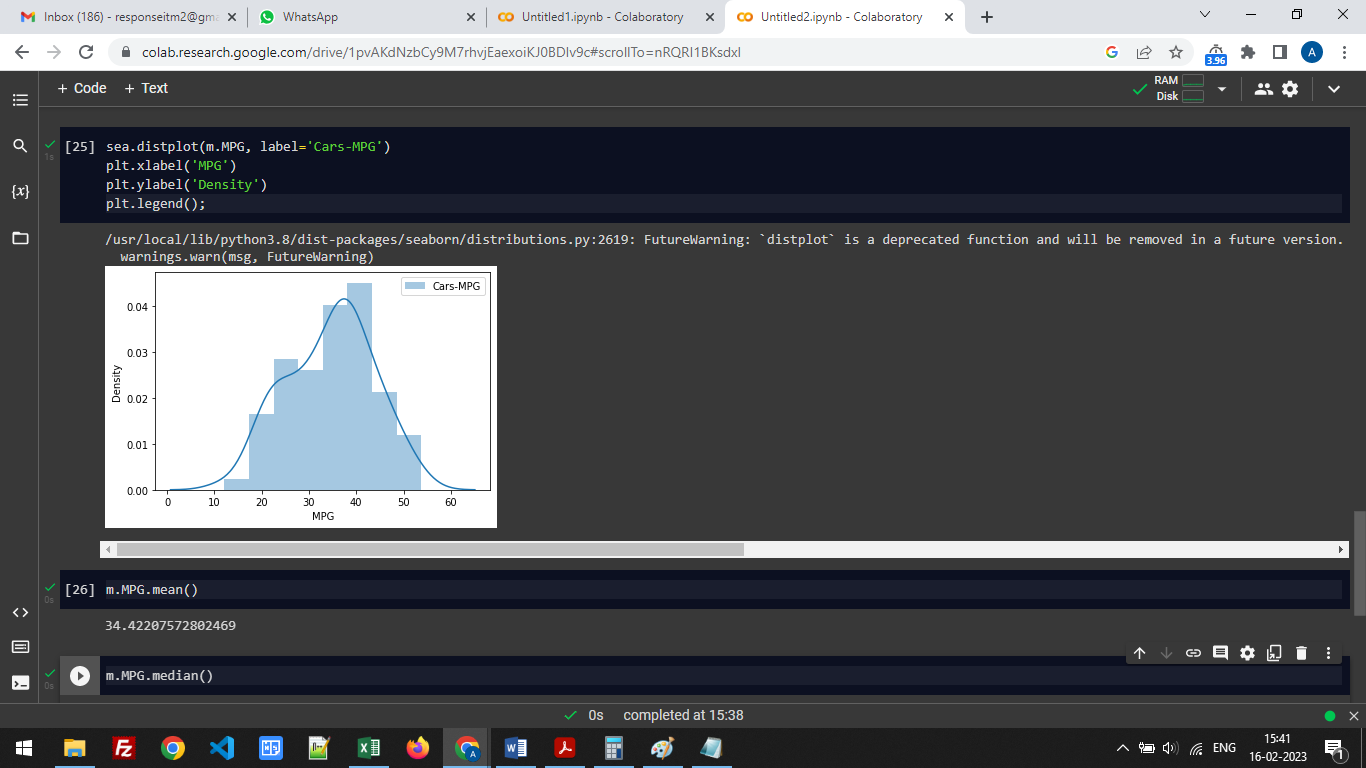
c. P (20<MPG<50) ---- 1.2430968797327491e-05

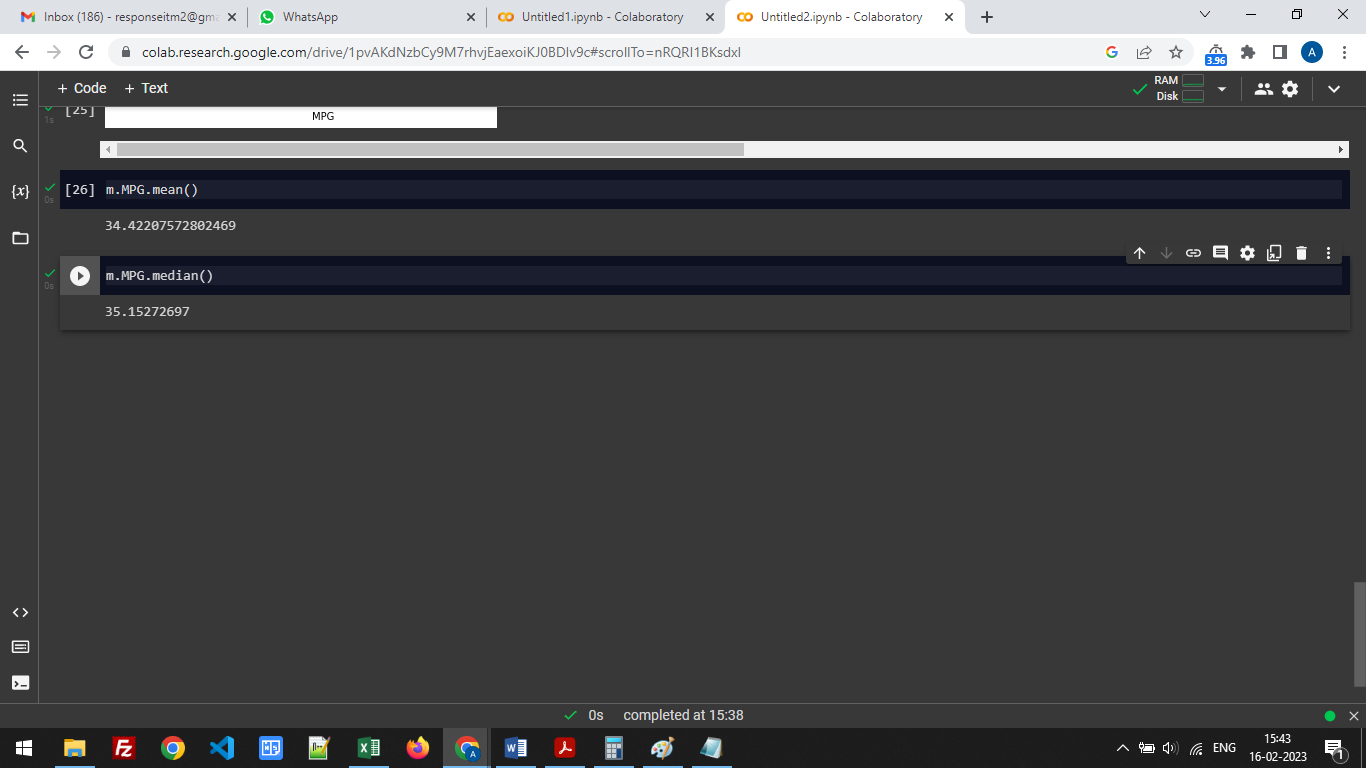
Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Answer:-



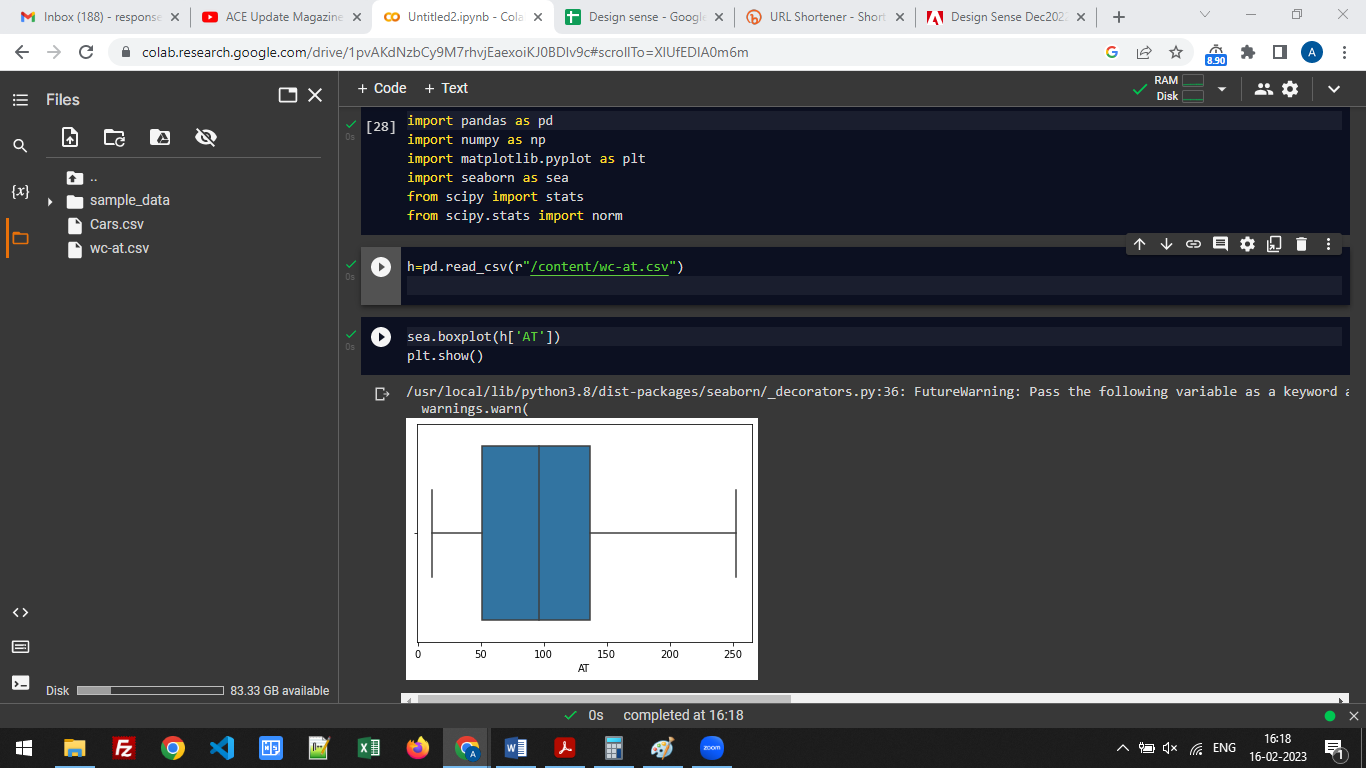


The distribution is normal.

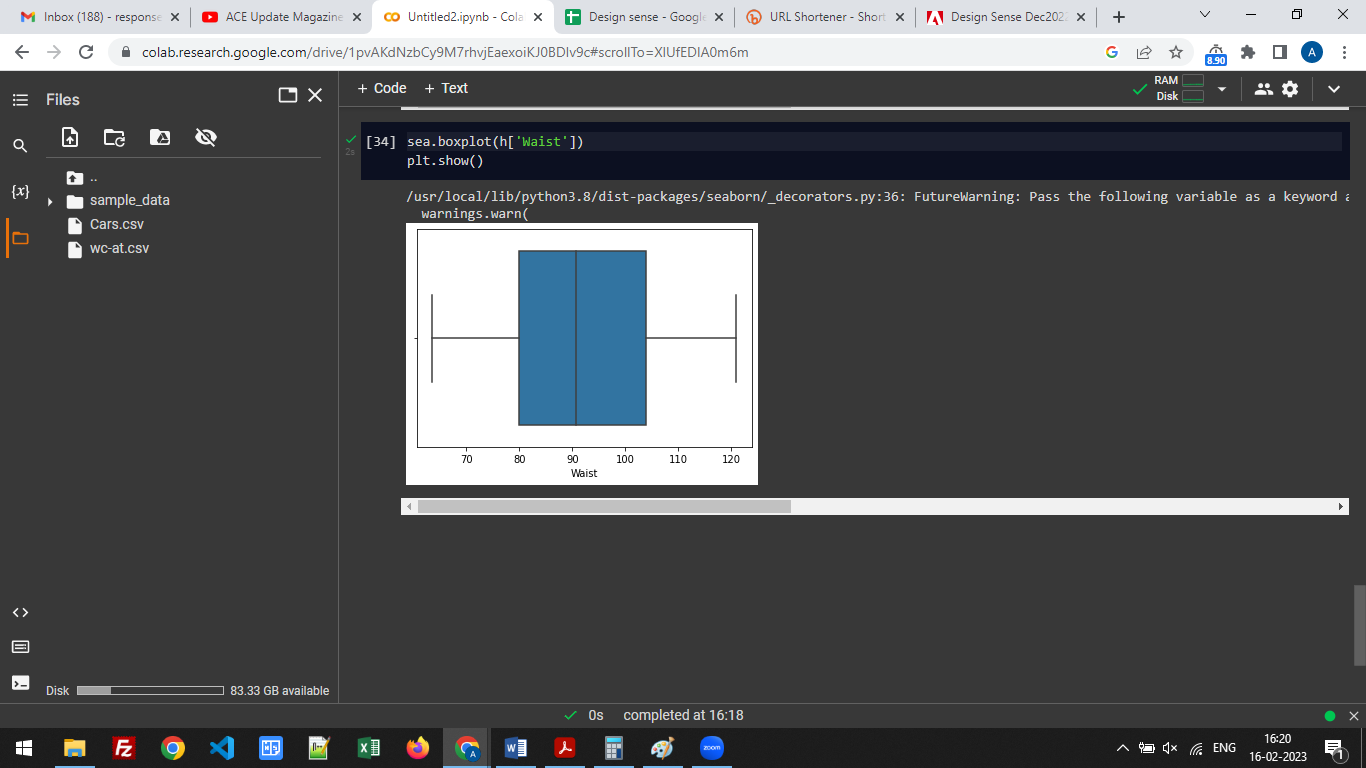
1. Check Whether the Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set follows Normal Distribution

Dataset: wc-at.csv

Answer:-



Adipose Tissue (AT) – Normal Distribution



Waist Circumference(Waist) - Normal Distribution

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



The Zscore of 90% confidence interval – 1.6448536269514722

The Zscore of 94% confidence interval - 1.8807936081512509

The Zscore of 60% confidence interval - 0.8416212335729143

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

Answer: - t scores of 95% confidence interval = 2.0638985616280205

t scores of 96% confidence interval = 2.1715446760080677

t scores of 99% confidence interval = 2.796939504772804

Q 24**)**A Government companyclaims 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

Answer: - t = - 0.471

The probability of the bulbs lasting less than 260 days on average = 0.32167411684460556