|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | Numerical - discrete |
| Results of rolling a dice | Categorical- nominal |
| Weight of a person | Numerical - continuous |
| Weight of Gold | Numerical - continuous |
| Distance between two places | Numerical - continuous |
| Length of a leaf | Numerical - continuous |
| Dog's weight | Numerical - continuous |
| Blue Color | Categorical- nominal |
| Number of kids | Numerical - discrete |
| Number of tickets in Indian railways | Numerical - discrete |
| Number of times married | Numerical - discrete |
| Gender (Male or Female) | Categorical- nominal |

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

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

= 3/8

Q4) Two Dice are rolled, find the probability that sum is

* Equal to 1 = “the sum of dices cannot be the 1 .beacuse on dice there is no 0 for that the dices have numbers 1,0”.
* Less than or equal to 4 = 1/6 [ (1,1), (1,2), (1,3), (2,1), (2,2), (3,1) ].
* Sum is divisible by 2 and 3 = 7/18 [(1,5), (2,4), (3,3), (4,2), (5,1), (2,6), (6,2), (3,5), (5,3), (4,4), (6,4), (4,6), (5,5), (6,6) ]

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?

= 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

= [ So, on average, a randomly selected child is expected to have about 3.09 candies. ]

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.

Out put =

count,32.0,32.0,32.0

**mean**,

points = 3.5965625,

score = 3.2172500000000004,

Weigh = 17.848750000000003

**std**,

points = 0.5346787360709715,

score = 0.9784574429896966,

weigh = 1.7869432360968431

**min**,

points = 2.76,

score = 1.513,

weigh = 14.5

**25%,**

Points = 3.08,

Score = 2.58125,

Weigh = 16.8925

**50%,**

Points = 3.6950000000000003,

Score = 3.325,

Weigh = 17.71

**75%,**

Points = 3.92,

Score = 3.61,

Weigh = 18.9

**max,**

points = 4.93,

score = 5.424,

weigh = 22.9

**Use Q7.csv file**

Q8) Calculate Expected Value for the problem below

* 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?

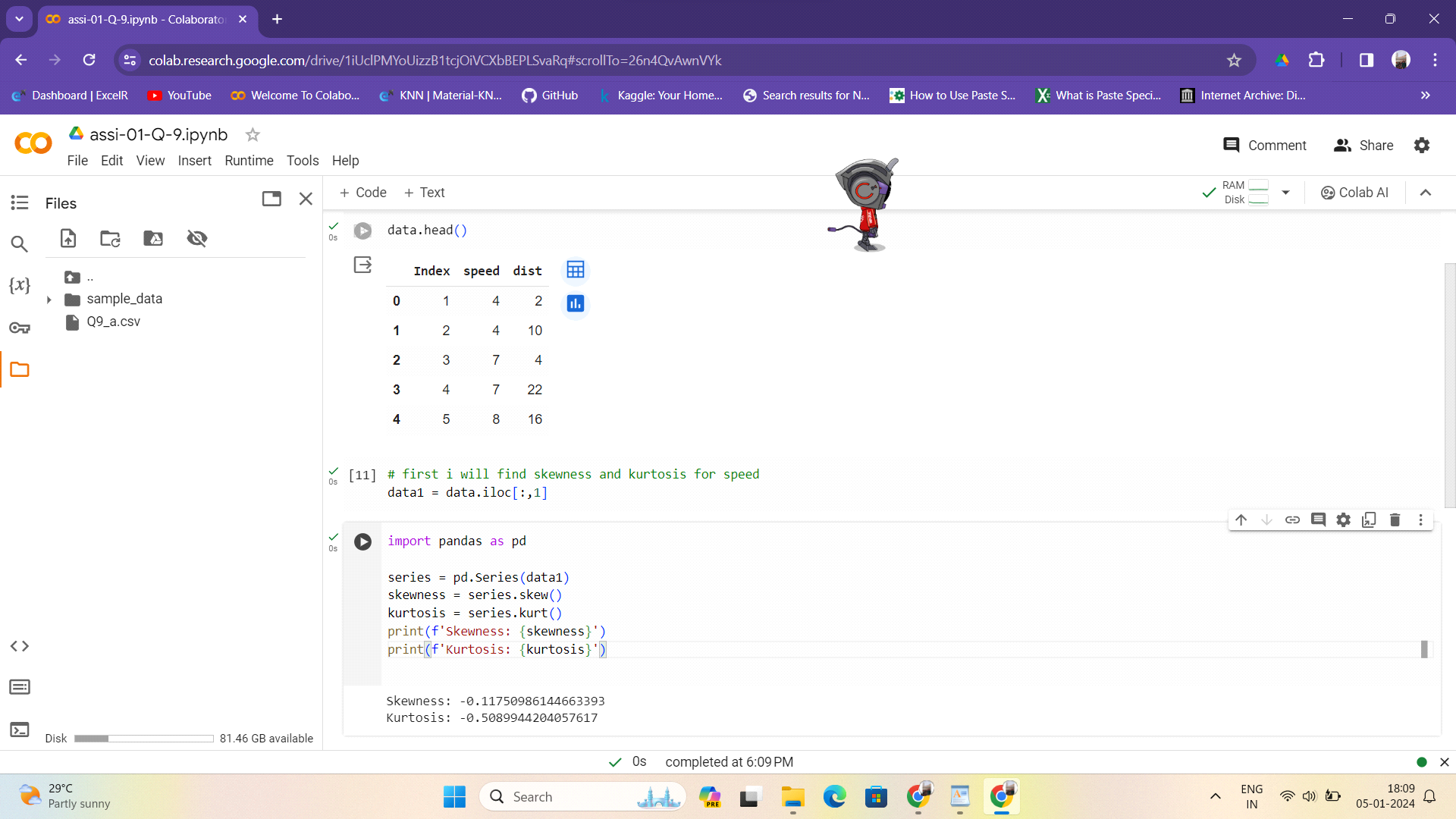
**[ Mean = 145.33 pounds ]**

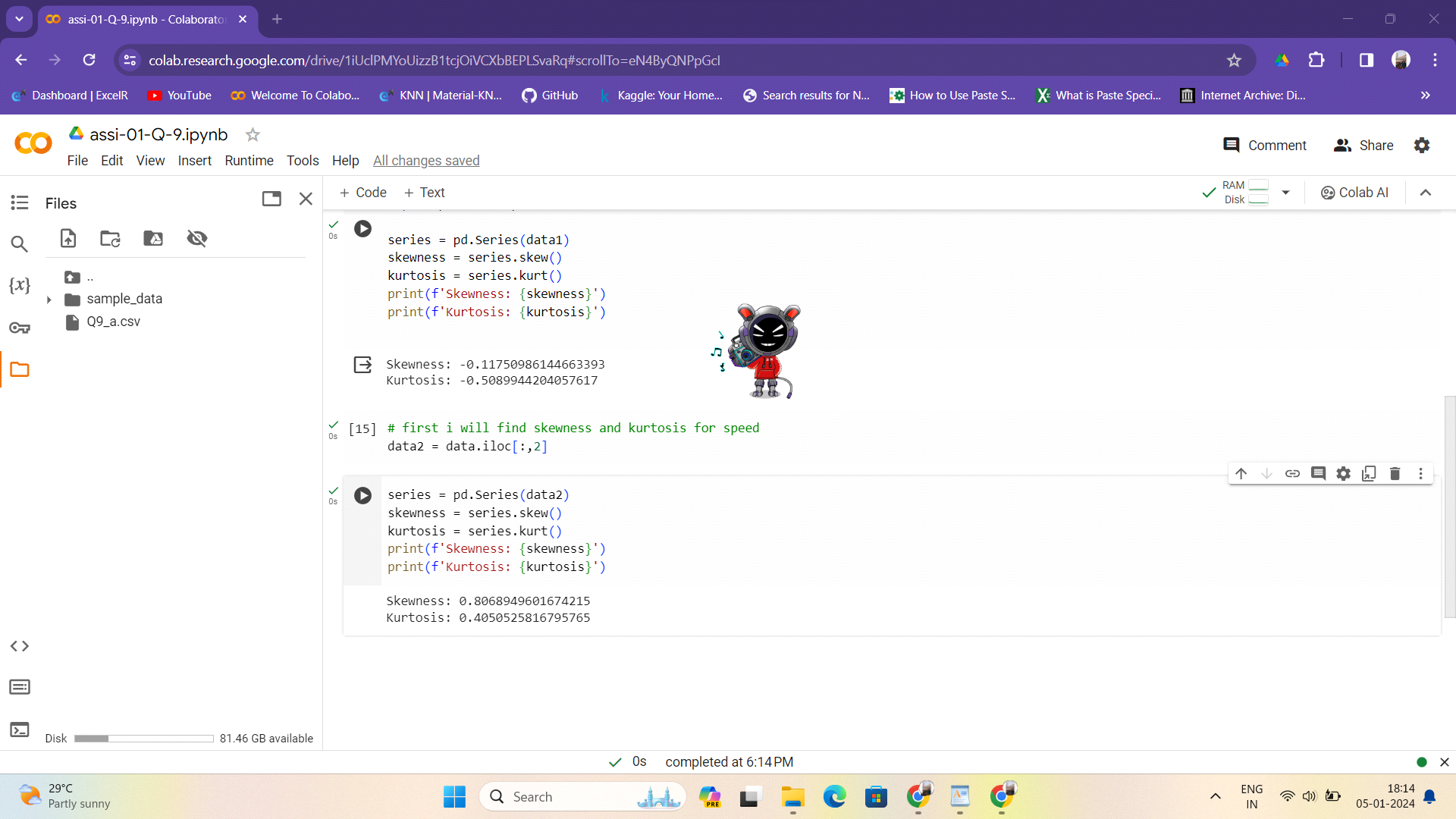
**“Like you guys have mentioned in this question that I have to take random value from this so I will go with the average value that value I find good if I have to take random value”**

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

**Cars speed and distance**

**Use Q9\_a.csv [**]

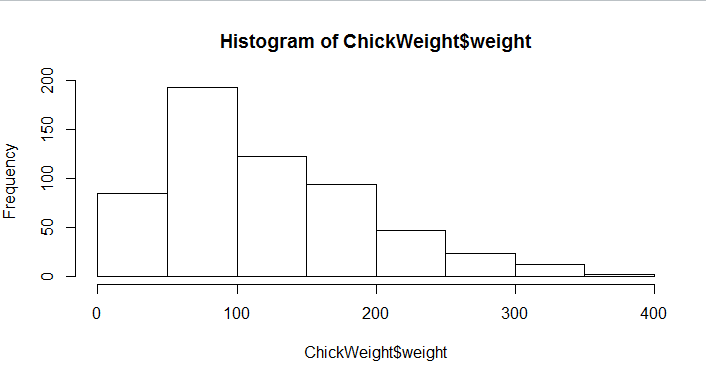


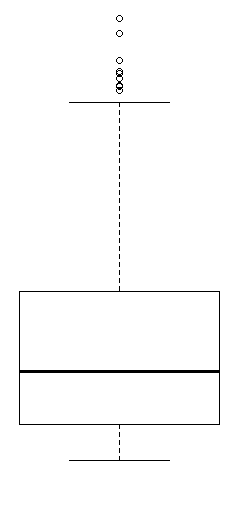


**SP and Weight(WT)**

**Use Q9\_b.csv**

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





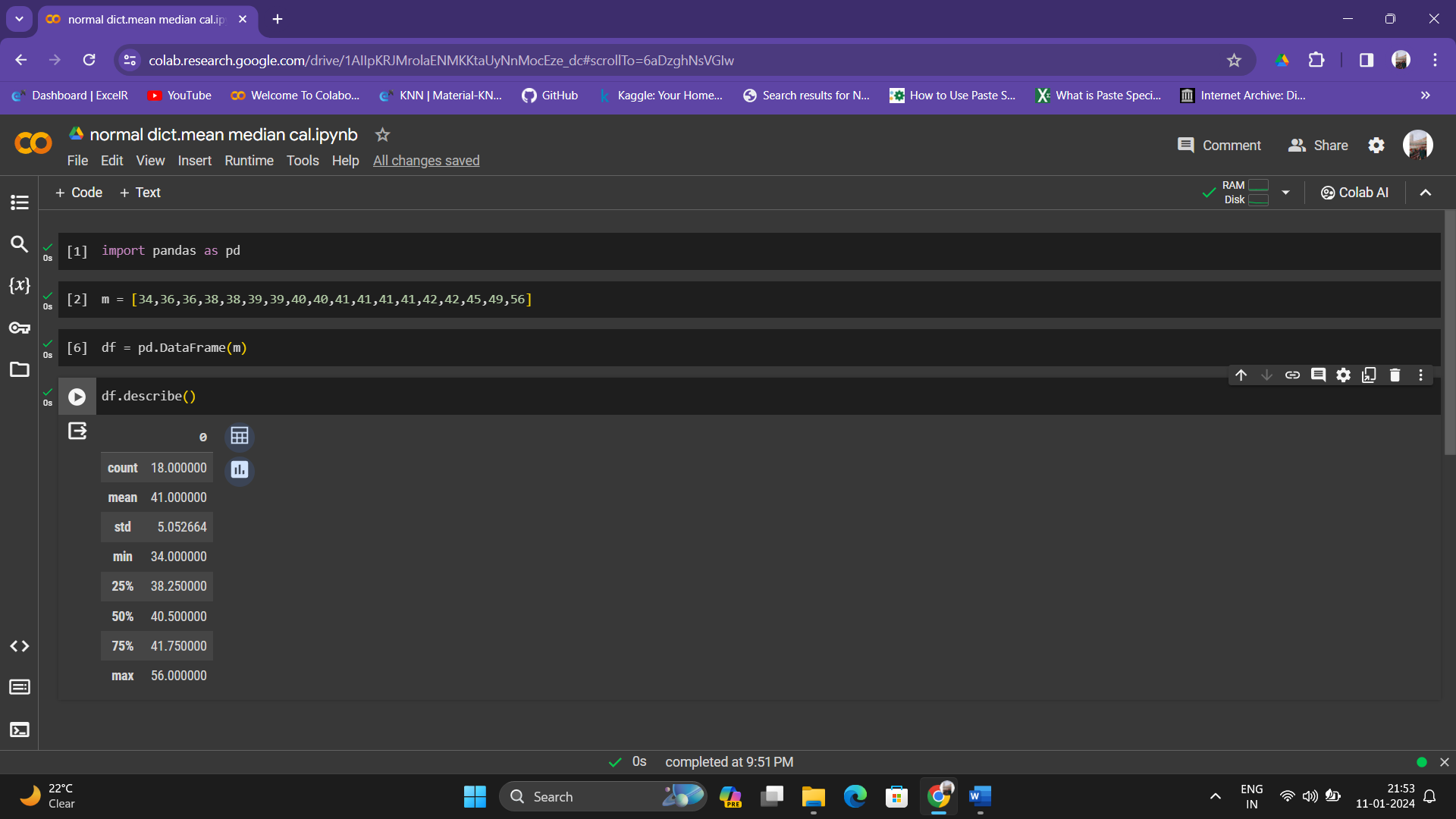
**Q11)** Suppose we want to estimate the average weight of an adult male in Mexico. We draw a random sample of 2,000 men from a population of 3,000,000 men and weigh them. We find that the average person in our sample weighs 200 pounds, and the standard deviation of the sample is 30 pounds. Calculate 94%,98%,96% confidence interval?



**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**

* Find mean, median, variance, standard deviation.



* What can we say about the student marks?
* Student need to focus on studies, there are some area that student to focus into several subjects.

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

* Its means the data is equally distributed.
* Normal distribution

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

* Its positively skewed data set.

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

* Yep its negatively skewed data

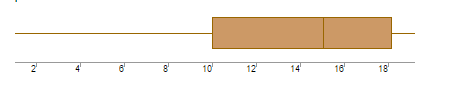
Q16) What does positive kurtosis value indicates for a data ?

* This means that the data has more frequent and extreme outliers than a normal distribution .

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

* This means that the data has thinner tails flatter shape than a normal distribution and having extreme outliers than a normal distribution.

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



What can we say about the distribution of the data?

* If the Median lies **closer to the Third Quartile** and if the **whisker at the** **upper end is shorter** then it has a **Negative Skew (Left Skew).**

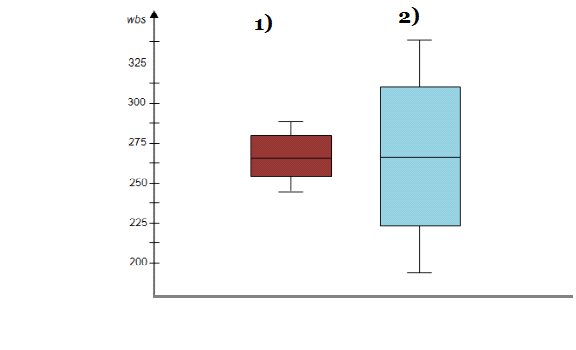
What is nature of skewness of the data?

* Negatively skewed data.

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

* 50% data will be IQR(inter quartile range).

Q19) Comment on the below Boxplot visualizations?



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

* In the first(1) boxplot I think so data is symmetrical or normal distributed data meadian .

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

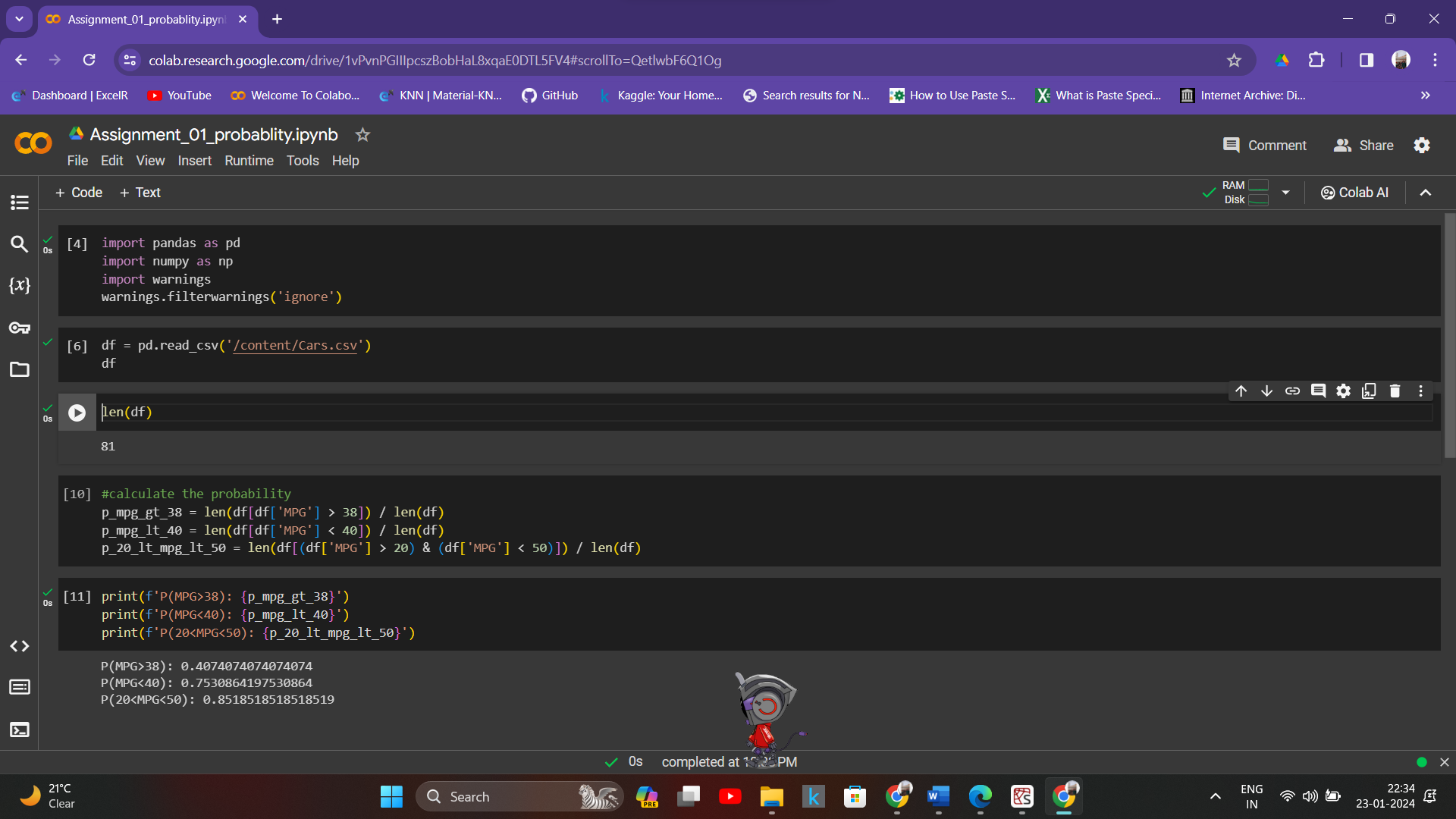
Data \_set: Cars.csv

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

MPG <- Cars$MPG

* P(MPG>38)
* P(MPG<40)

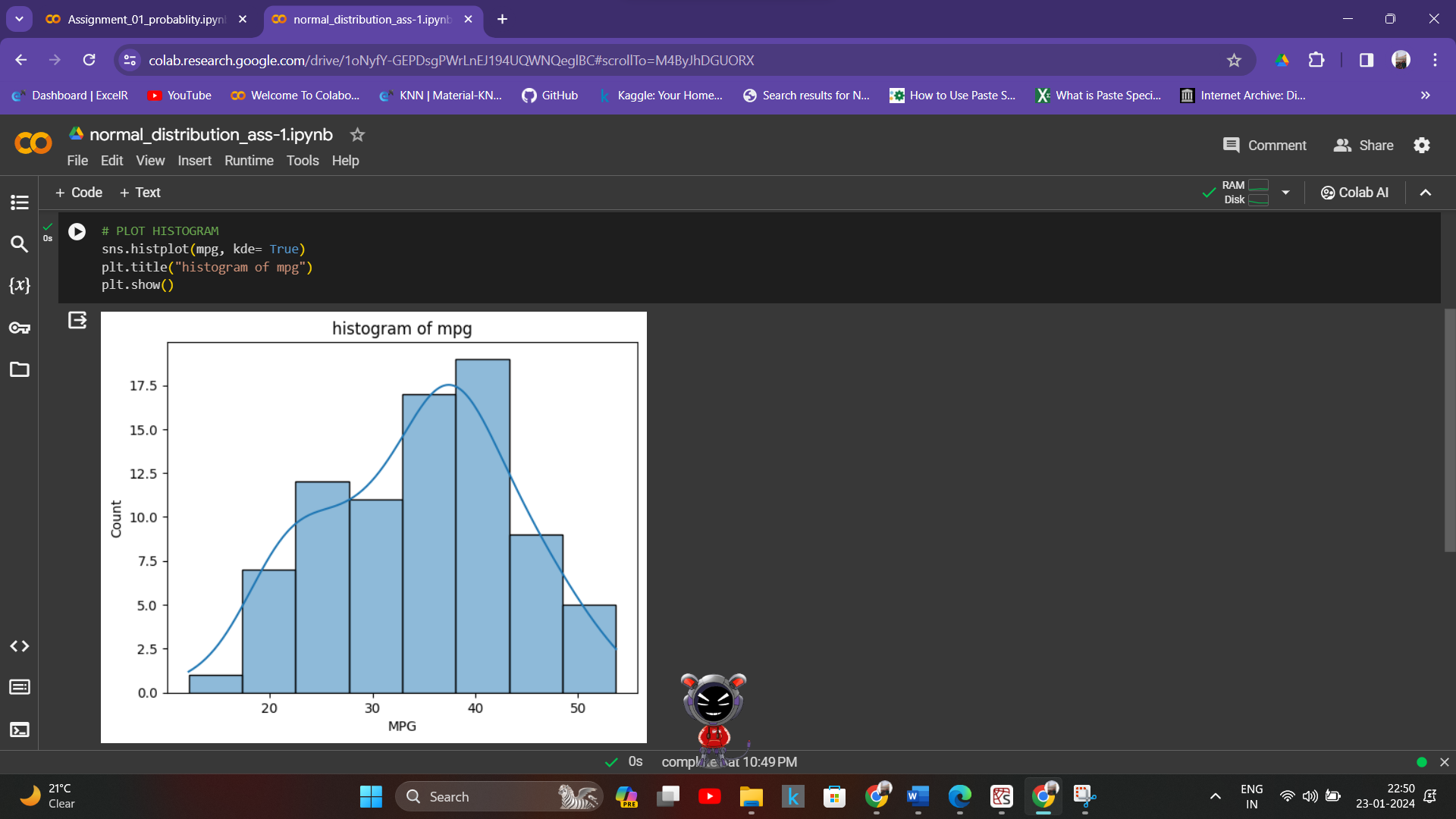
c. P (20<MPG<50)

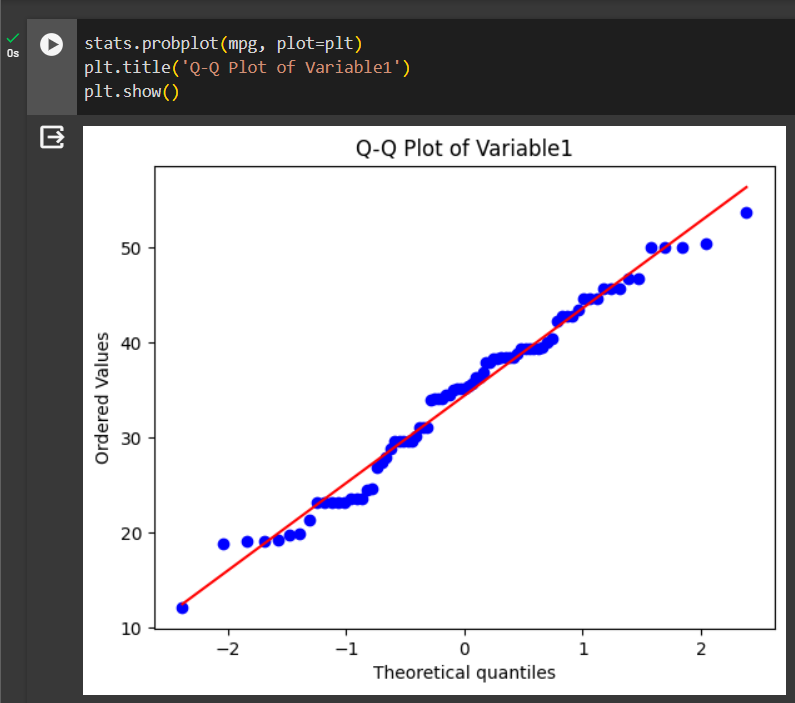


Q 21) Check whether the data follows normal distribution

* Check whether the MPG of Cars follows Normal Distribution

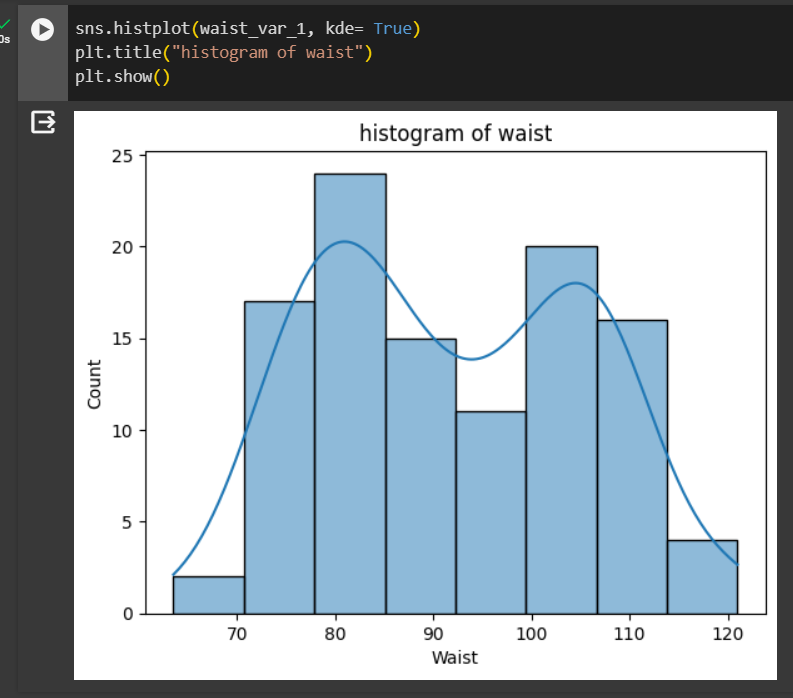
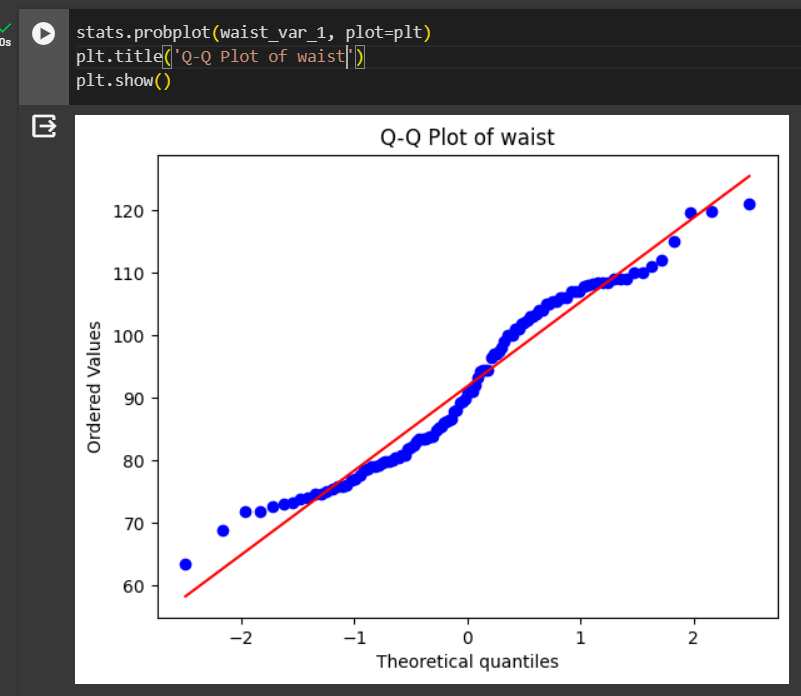
Dataset: Cars.csv

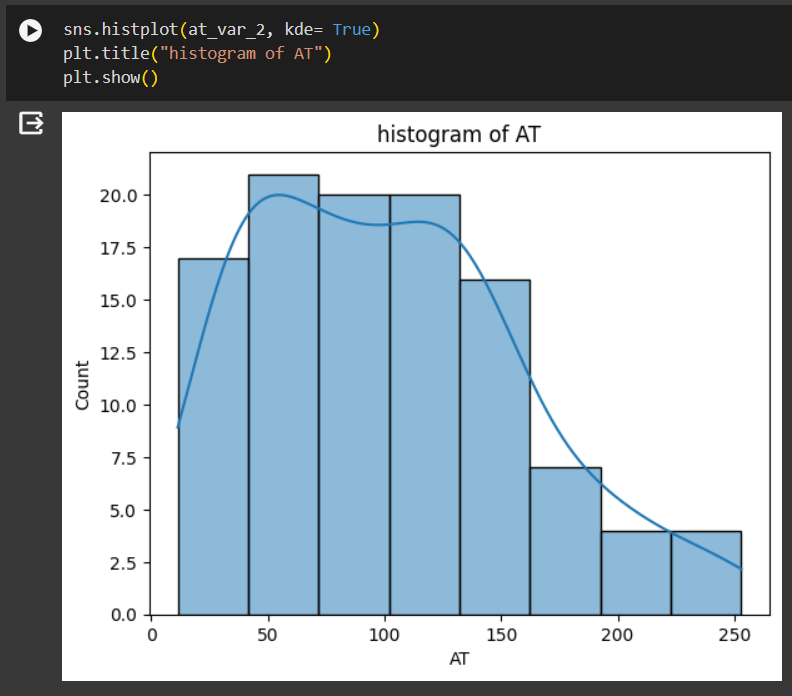
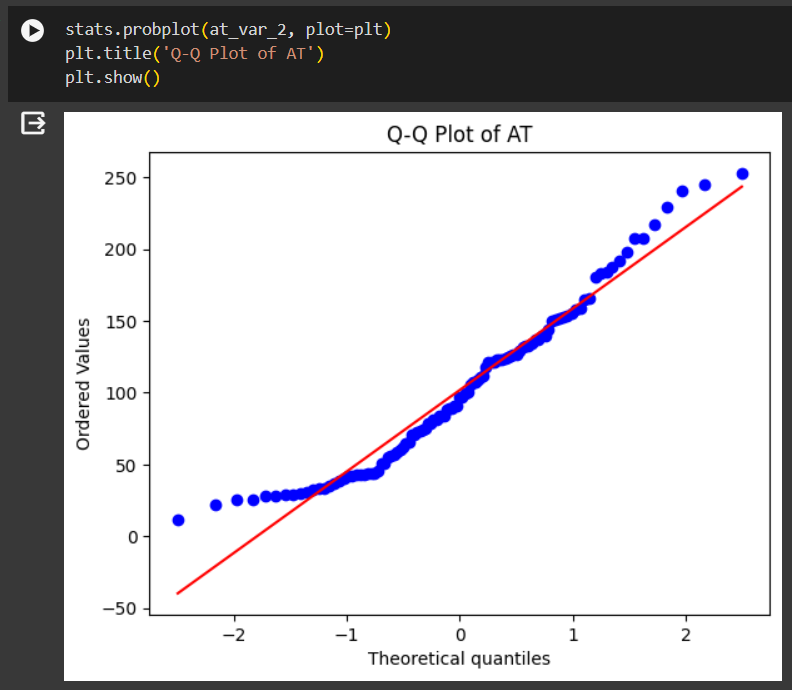




* Check Whether the Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set follows Normal Distribution

Dataset: wc-at.csv

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

* For = 90% = 1.654
* For = 94% = 1.55
* For = 60% = 0.814

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

* 95% = 2.064
* 96% =
* 99% = 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

* Probability that 18 randomly selected bulbs would have an average life of no more than 260 days is 0.31867594411696853