|  |  |
| --- | --- |
| 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) | 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 | 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 | Ratio |
| 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?

**Ans)** 3(HHT,HTH,THH)

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)** a)0

b)6{(1,1),(2,1),(3,1),(1,2),(2,2),(1,3)}

c)6{(1,5),(2,4),(4,2),(5,1),(3,3),(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?

**Ans)** Total(s)=7balls

N(s)=(7C2)=7\*6/2\*1=21

Total balls are drawn at random=N(A)=(5C2)=5\*4/2\*1=10

The probability that none of the balls drawn is blue=N(A)/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)** Expected number=

1\*0.015+4\*0.20+3\*0.65+5\*0.005+6\*0.01+2\*0.120=**3.09**

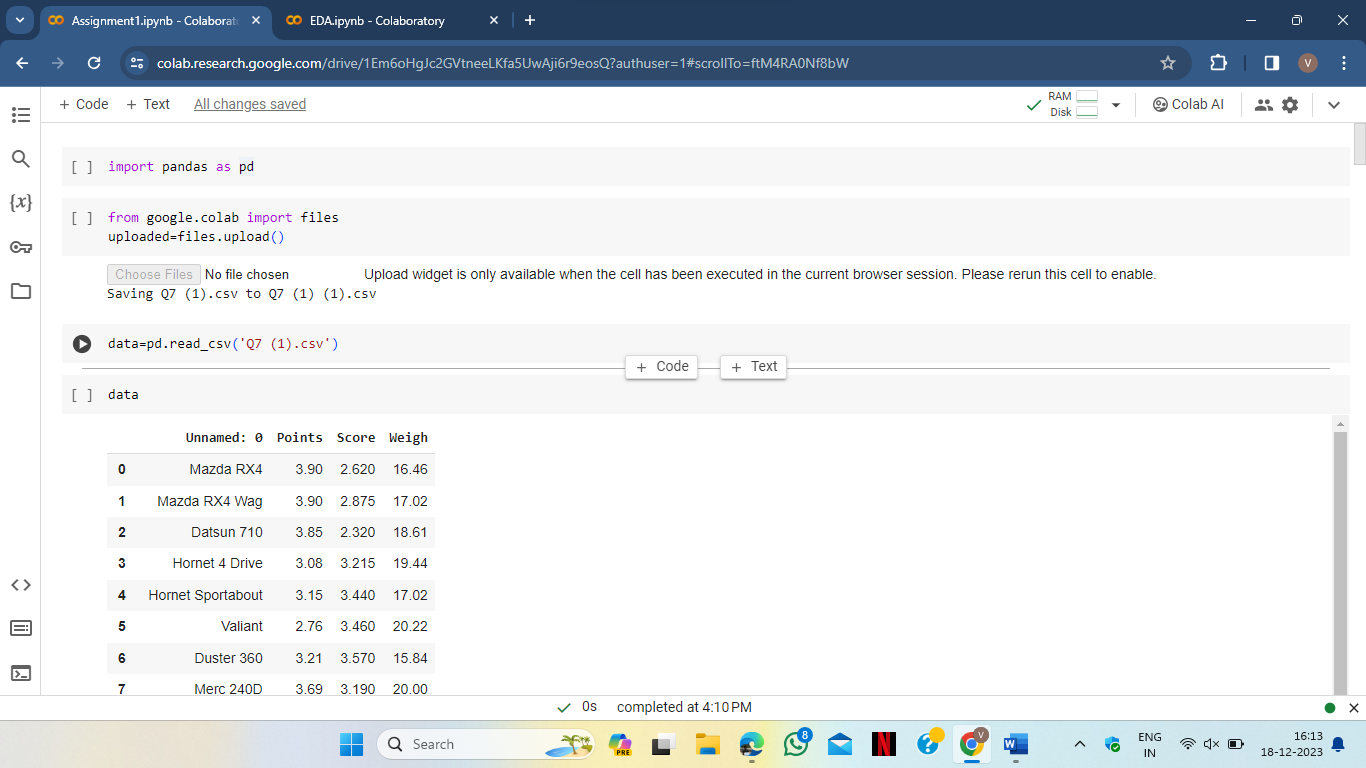
Q7) Calculate Mean, Median, Mode, Variance, Standard Deviation, Range & comment about the values / draw inferences, for the given dataset

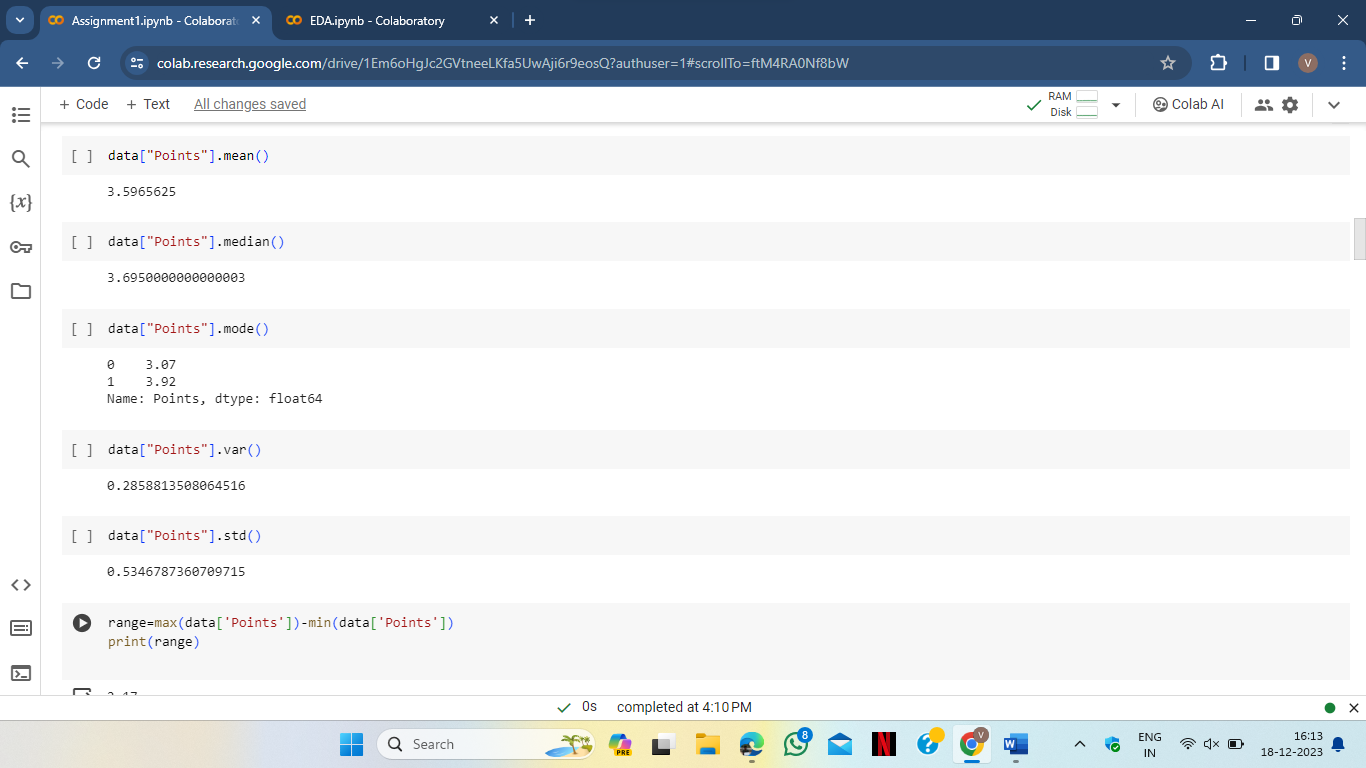
* For Points,Score,Weigh>

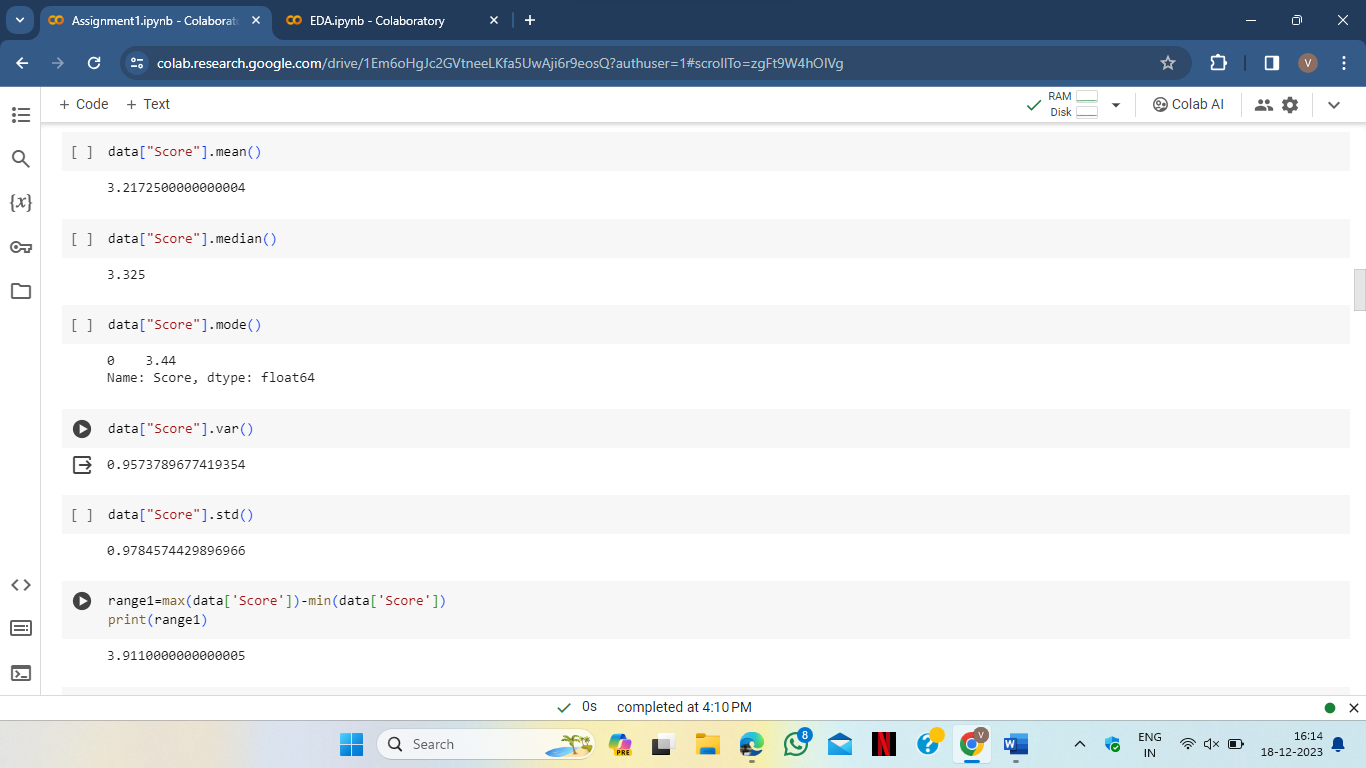
Find Mean, Median, Mode, Variance, Standard Deviation, and Range and also Comment about the values/ Draw some inferences.

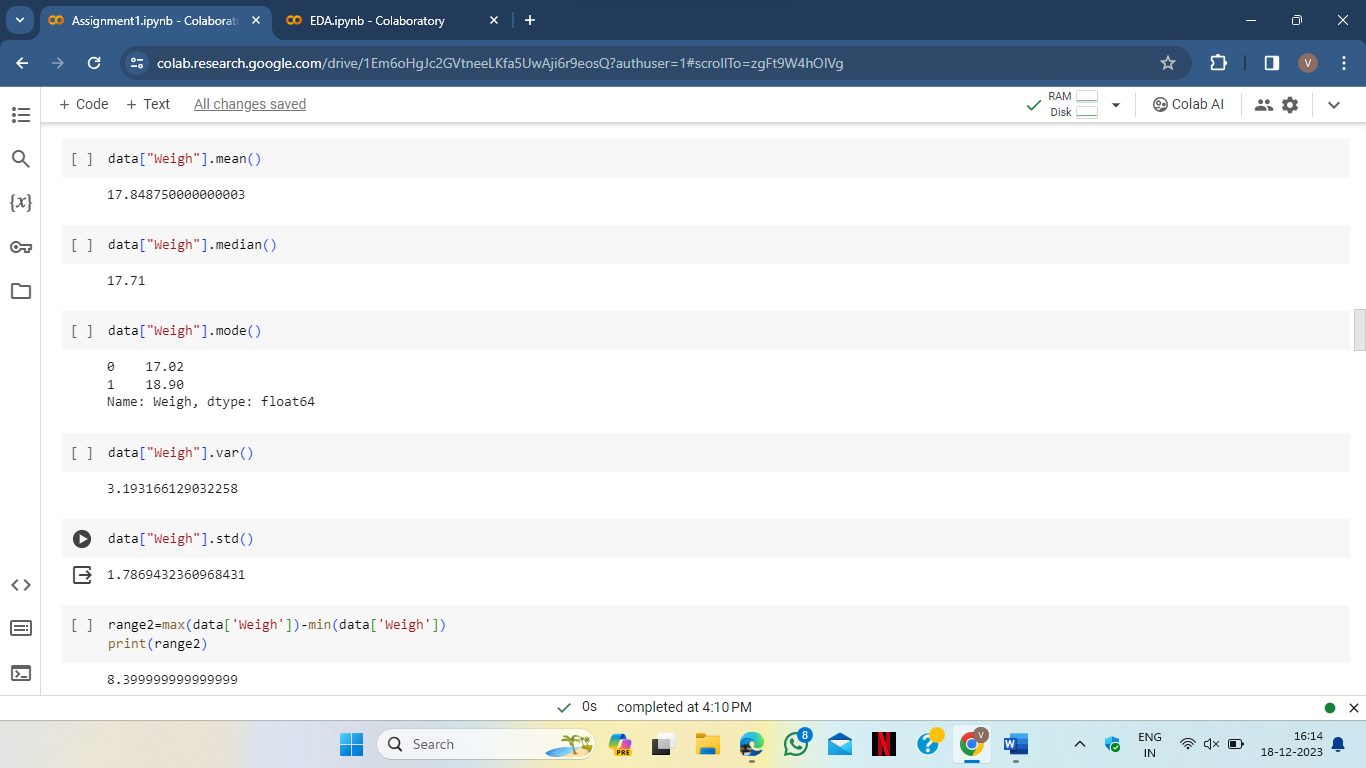
**Use Q7.csv file**

**Ans)**

****

****

****

****

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)** weight(x) = (108,110,123,134,135,145,167,187,199)

Total=9 patients

Total weight of the patients(x) is (108+110+123+134+135+145+167+187+199) = 1318

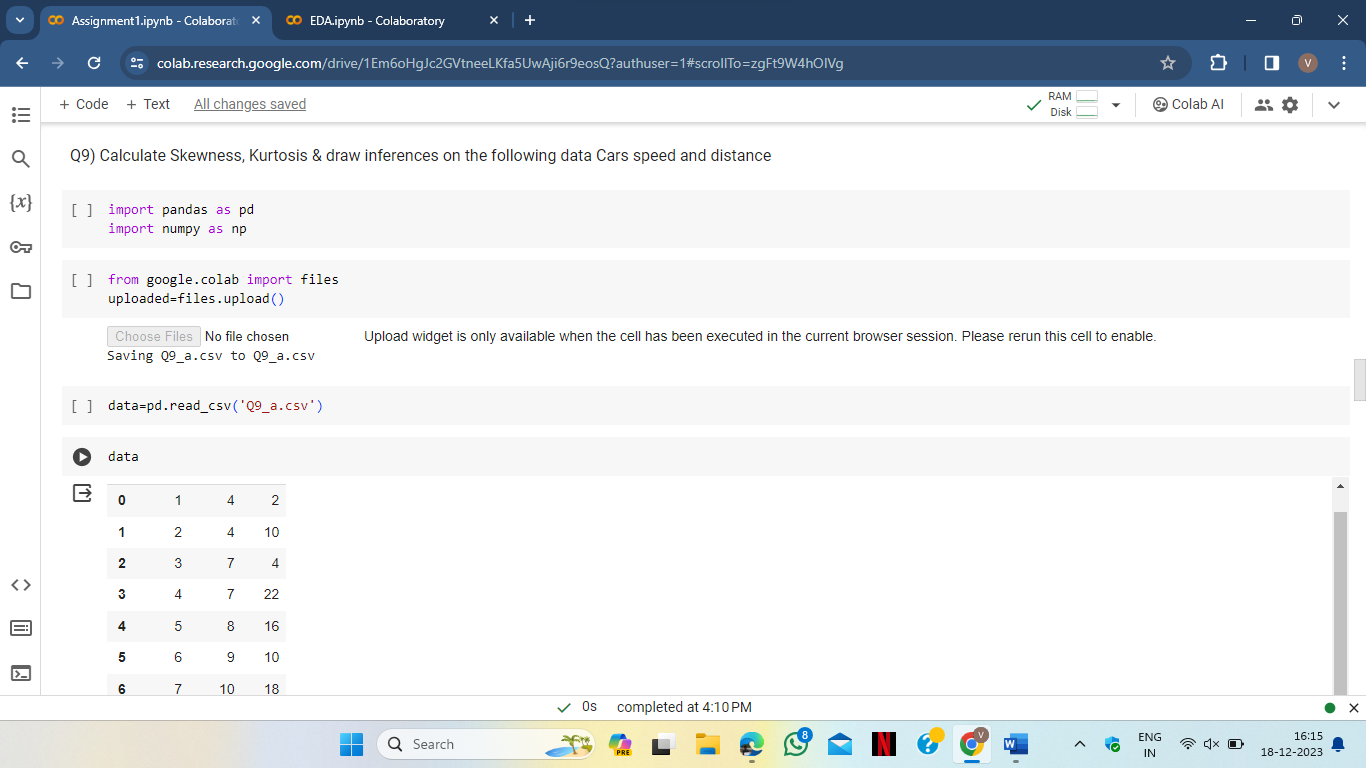
Expected Value of the weight is 1318/9 = 146.44

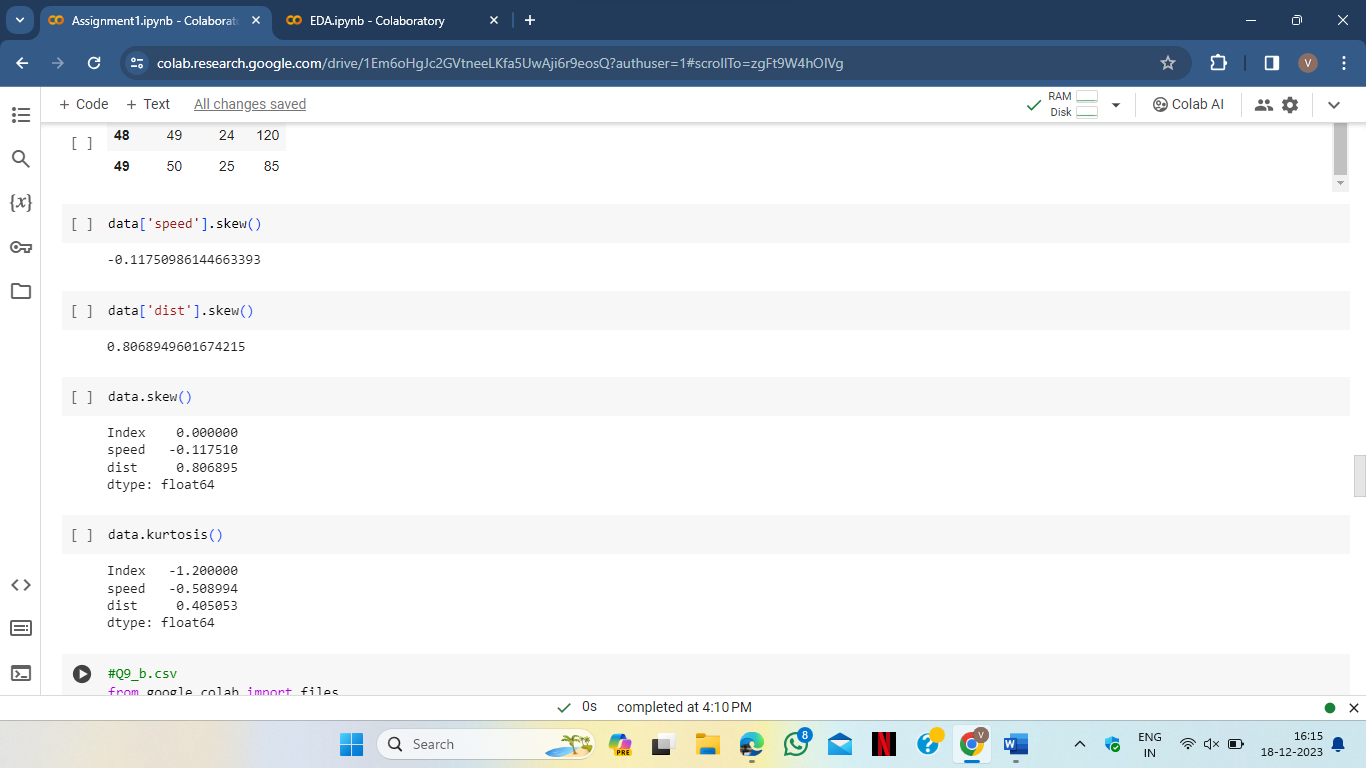
**Q9) Calculate Skewness, Kurtosis & draw inferences on the following data**

**Cars speed and distance**

**Use Q9\_a.csv**

**Ans)**

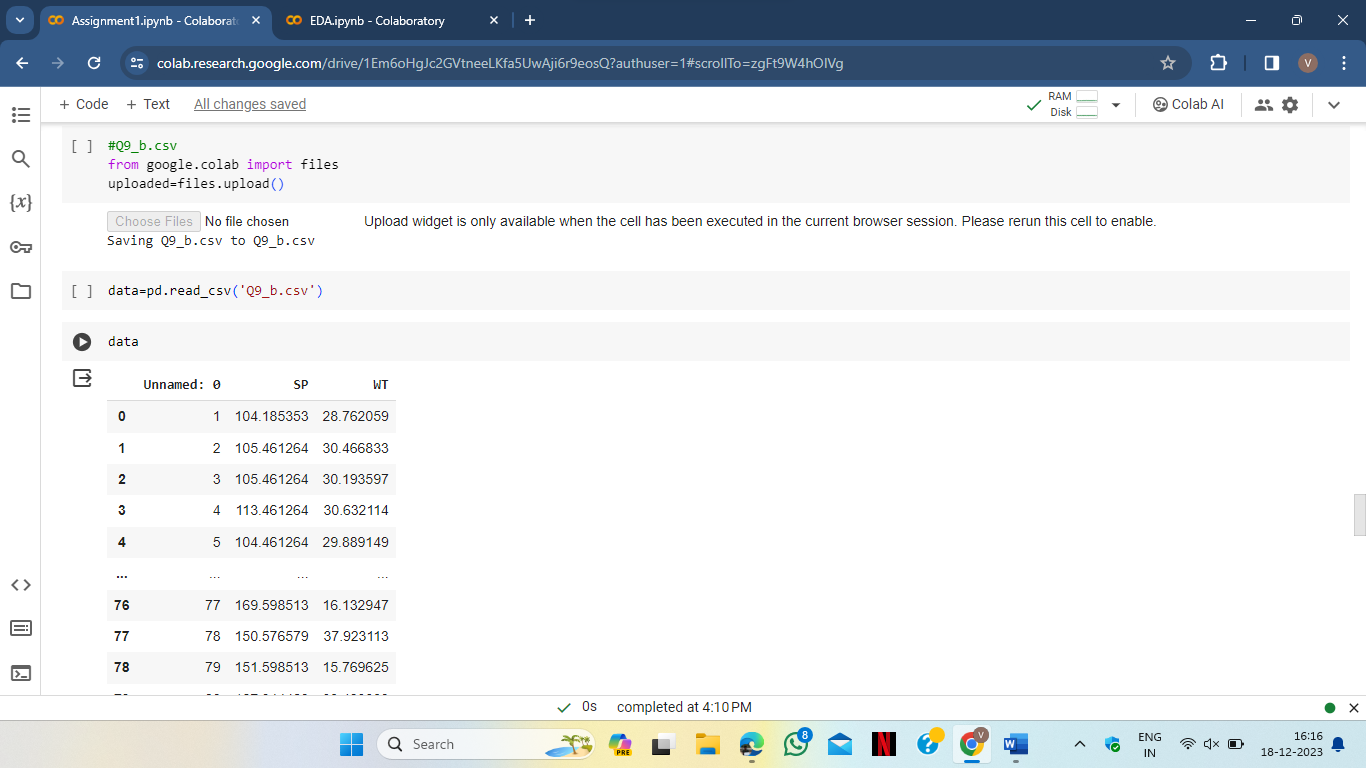
****

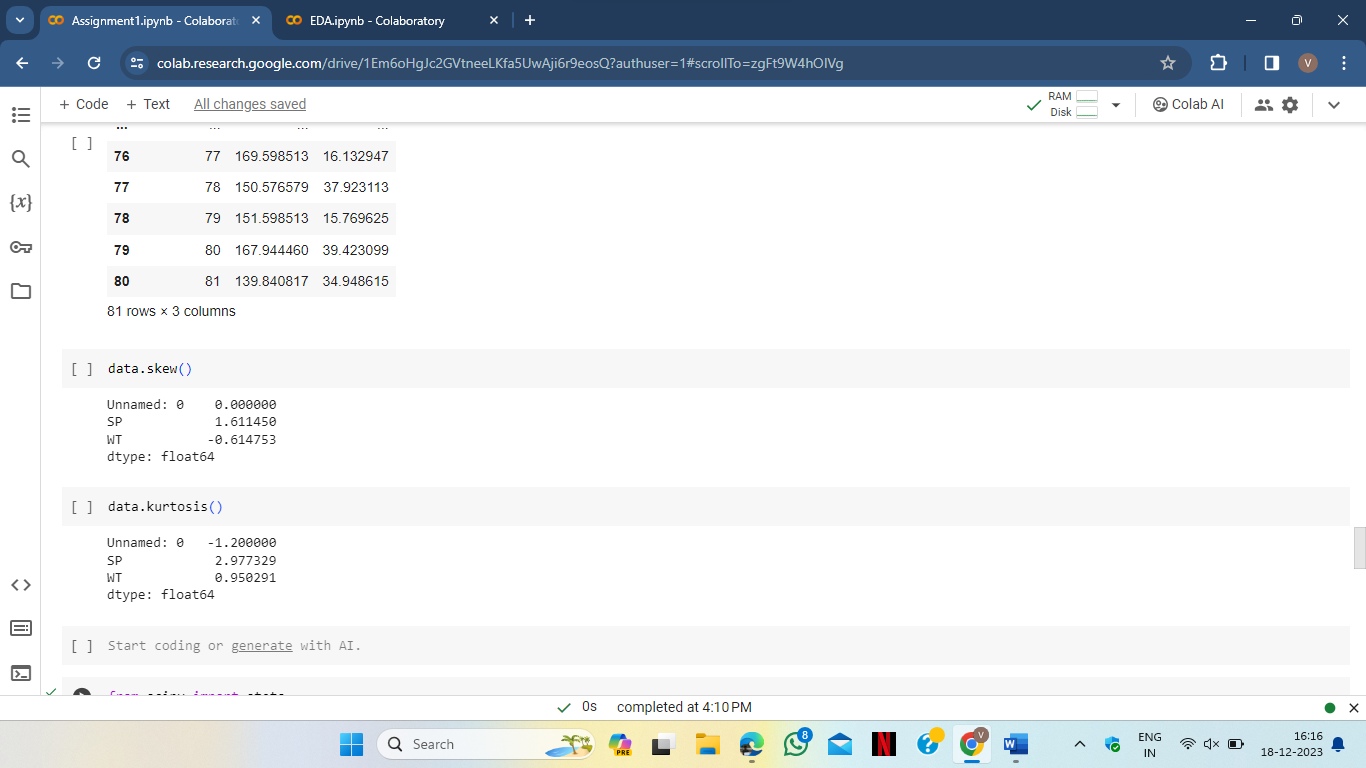
****

**SP and Weight(WT)**

**Use Q9\_b.csv**

**Ans)**

****

****

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



**Ans)** Here we see that the major chickweight fall in the category of 100 gram

(measures in x) as the Frequency which is 200.

The highest frequency lies between 50-100g.

The plot is Right skewed which show that there is less concentration of chick

Weight in the 300-400g.



**Ans)**

* Median is less than mean.
* Right skewed.
* We have outlier on the upperside of the boxplot.

**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?

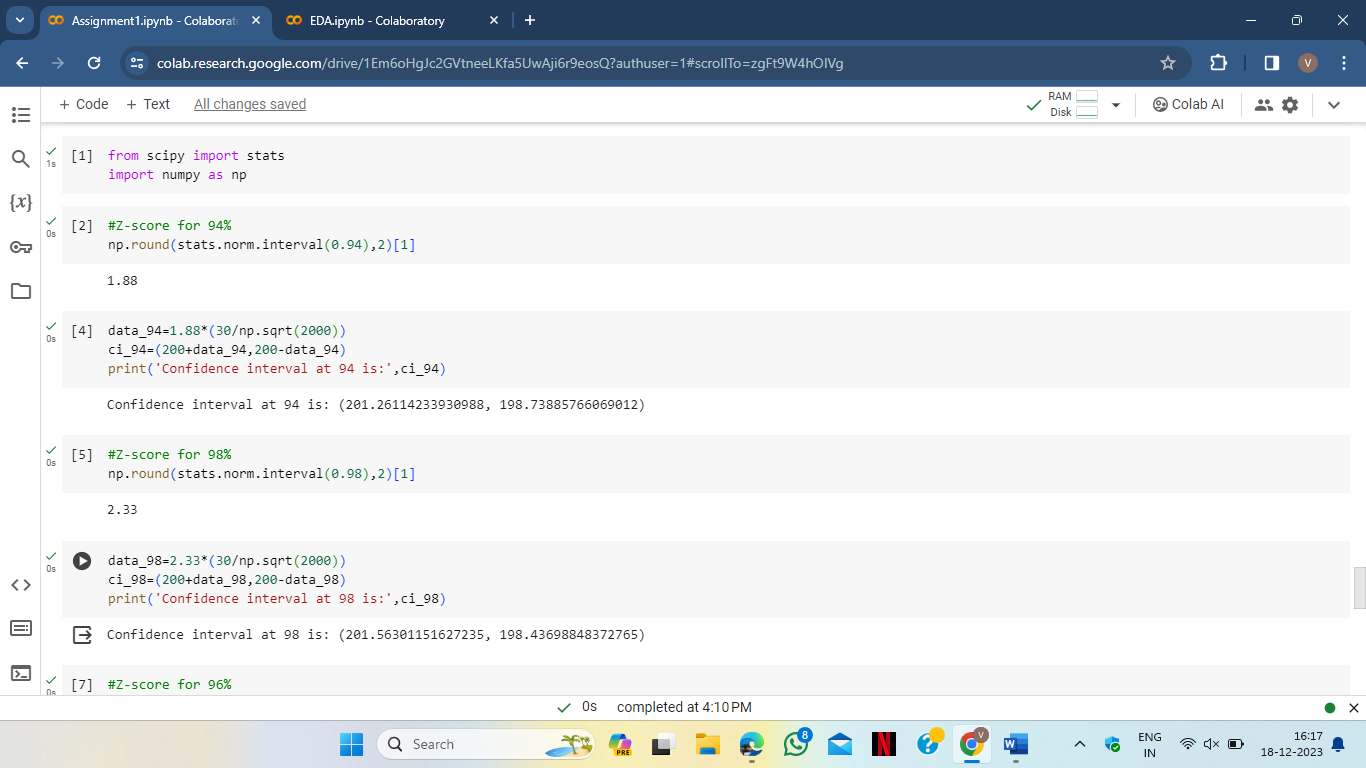
**Ans)**

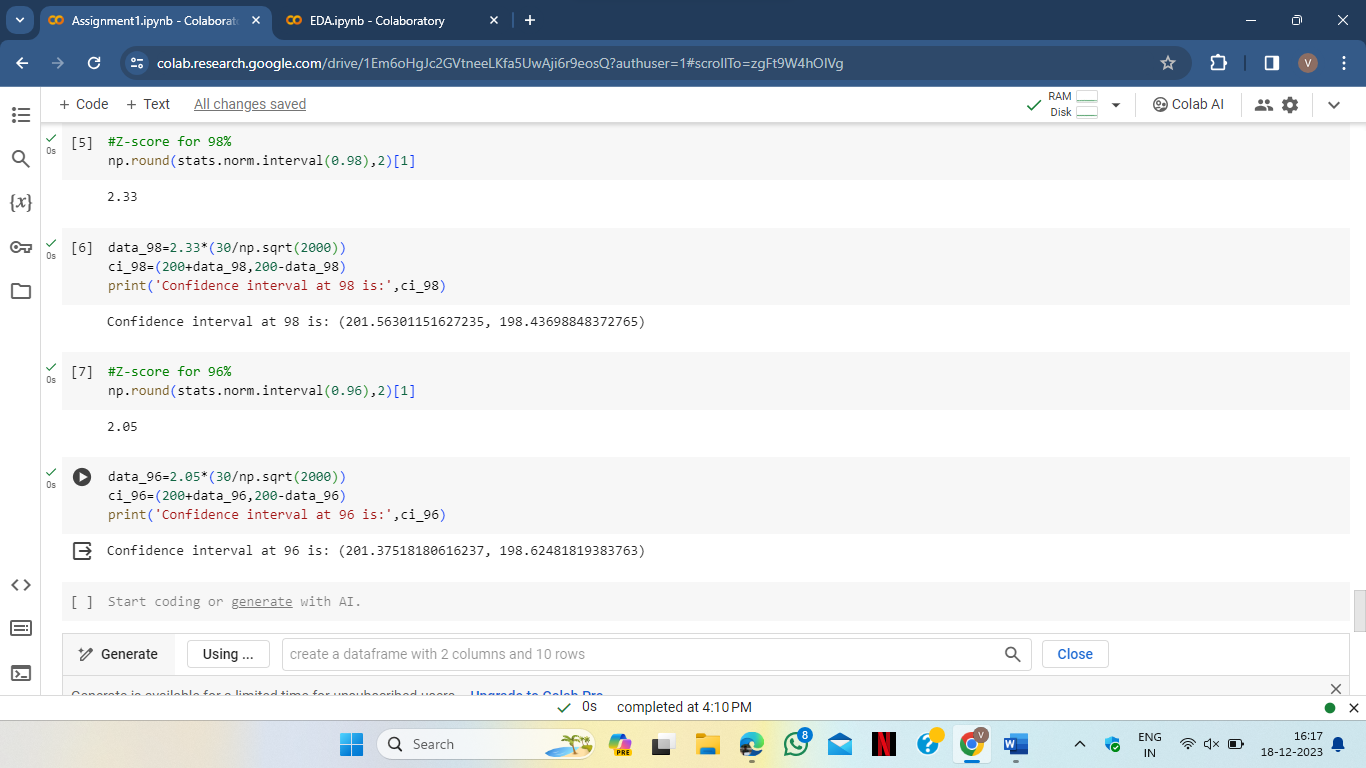
**n=2000**

**X=200**

**S=30**

**We use Z-score to find 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**

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

**Ans)**

* Mean=41
* Median=40.5
* Variance=25.53
* Std deviation=5.0526

1. What can we say about the student marks?

**Ans)** Mean > Median,this implies that the distribution slightly skewed

towards right.

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

**Ans)**

* The skewness will be symmetrical.
* Both sides of the plot must be equal.
* Normally distributed.

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

**Ans)** when mean > median

* The skewness will be positive(+ve) or right skewed.
* Most of the data will be lying on left side of the plot.

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

**Ans)** when mean > median

* The skewness will be Negative(-ve) or left skewed.
* Most of the data will be lying on right side of the plot.

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

**Ans)** A positive kurtosis value indicates that the distribution has heavier tail

then the normal distribution.

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

**Ans)** A Negative kurtosis value indicates that the distribution has lighter tail

then the normal distribution.

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



What can we say about the distribution of the data?

**Ans)** The data is not actually equally distributed on the line.

* The median value 15.3( Approx.x)
* 25% of data lies between 0-10.
* 50% of the data lies between 10-18.
* 25% of the data lies between 18-20.

What is nature of skewness of the data?

**Ans)** The data will be left skewed. Median is greater then the mean.

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

**Ans)** IQR means Inter Quartile Range

* Here Q1=10
* Q2=15.3
* Q3=18
* IQR=Q3-Q1=8.

**Q19)** Comment on the below Boxplot visualizations?



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

**Ans)**

* Here boxplot 2 is highly distributed and boxplot 1 is slightly distributed.
* Whiskers in this plane shows that boxplot 2 spread across the data and boxplot 1 spread 250-280 approx.
* The data spread equal on both of the boxplot.

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

Data \_set: Cars.csv

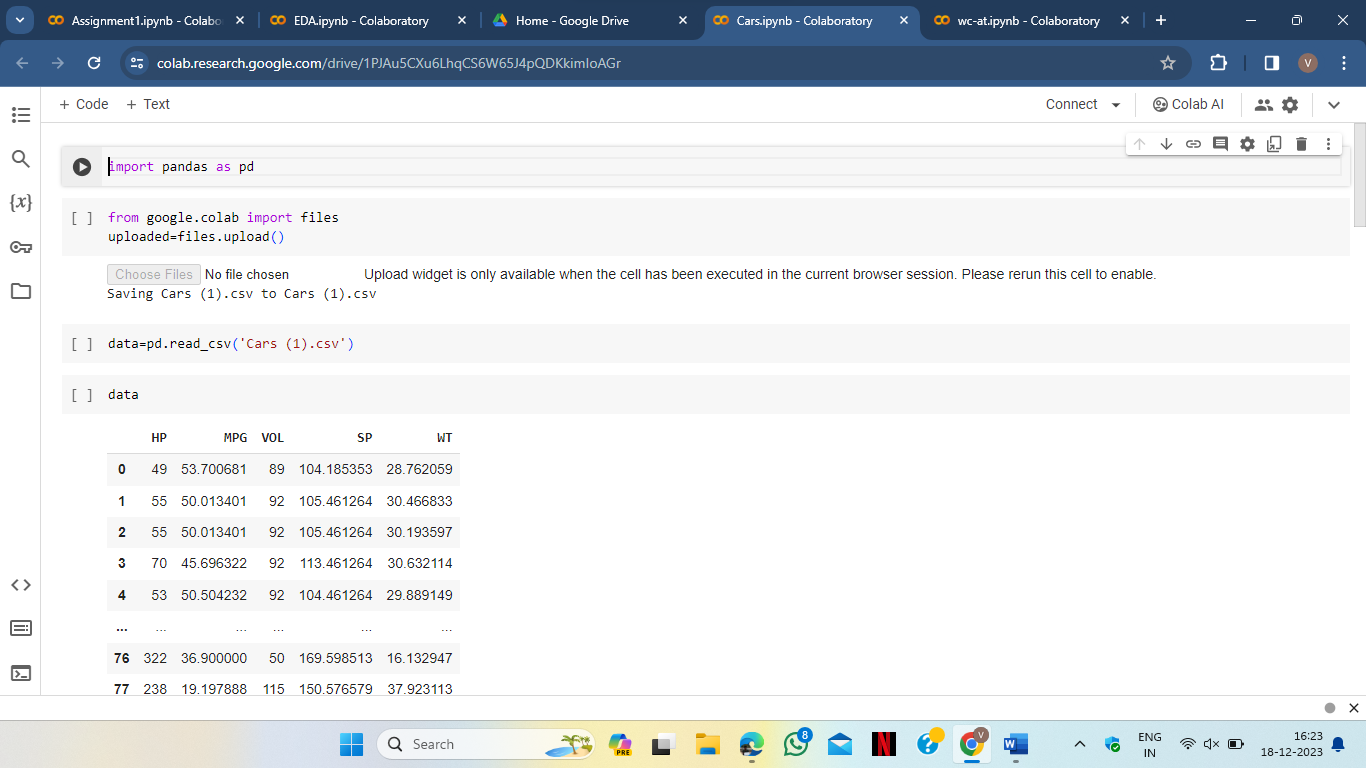
Calculate the probability of MPG of Cars for the below cases.

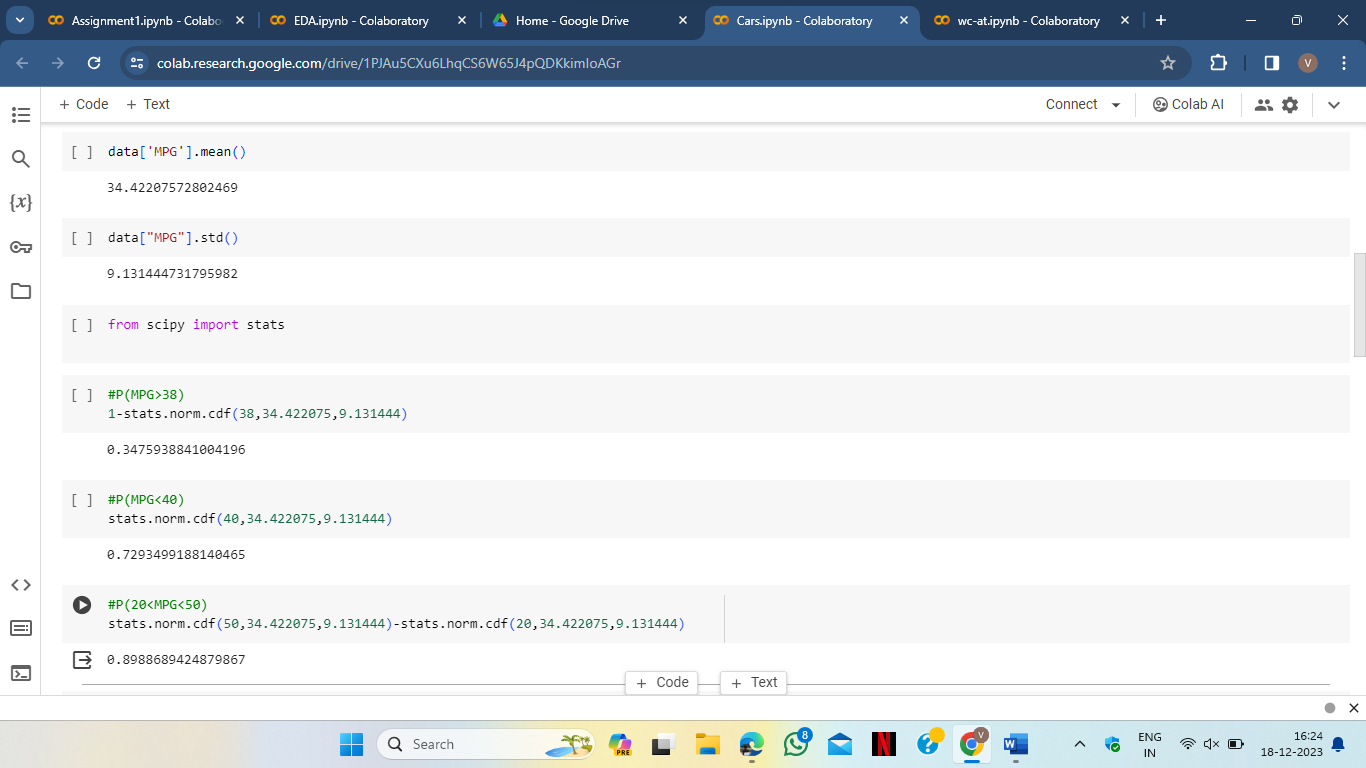
MPG <- Cars$MPG

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

c. P (20<MPG<50)

**Ans)**

****

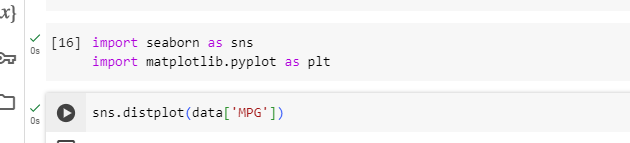
****

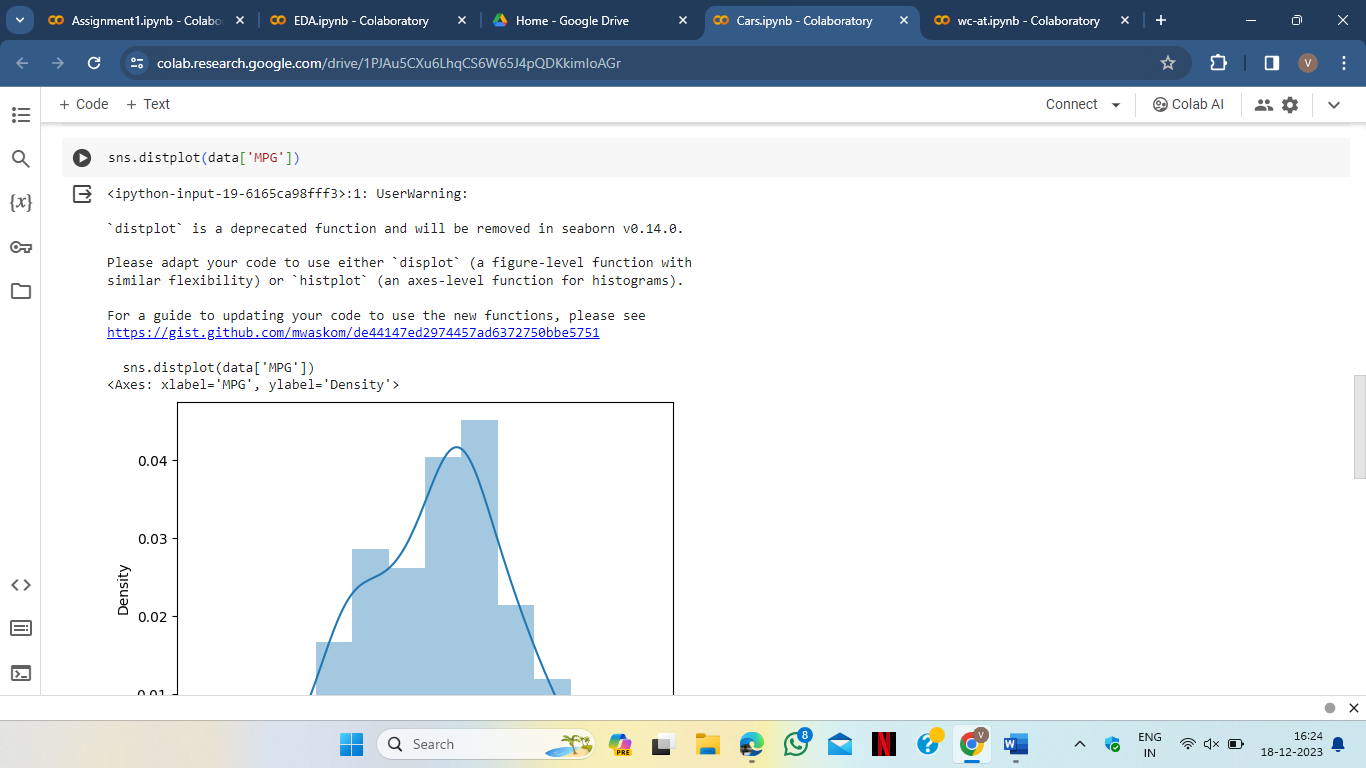
Q 21) Check whether the data follows normal distribution

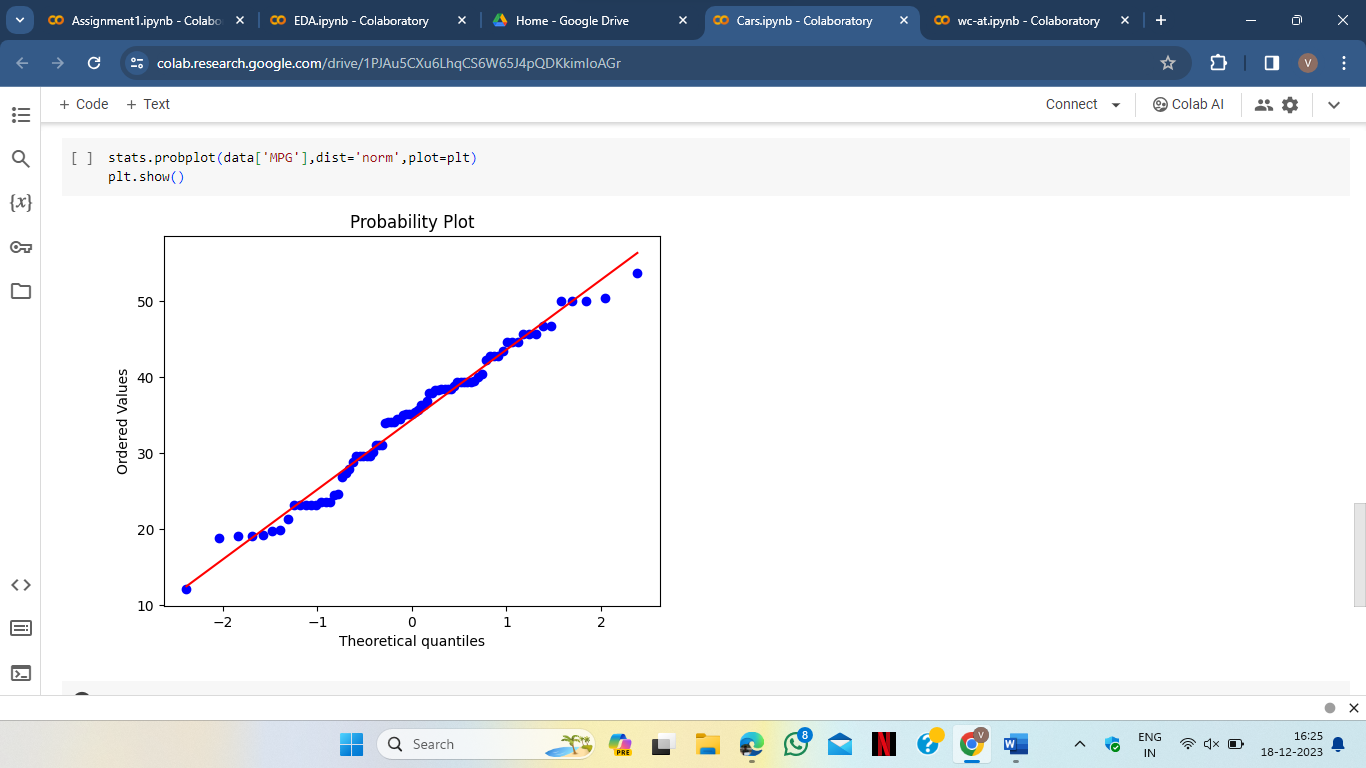
1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

**Ans)**

****

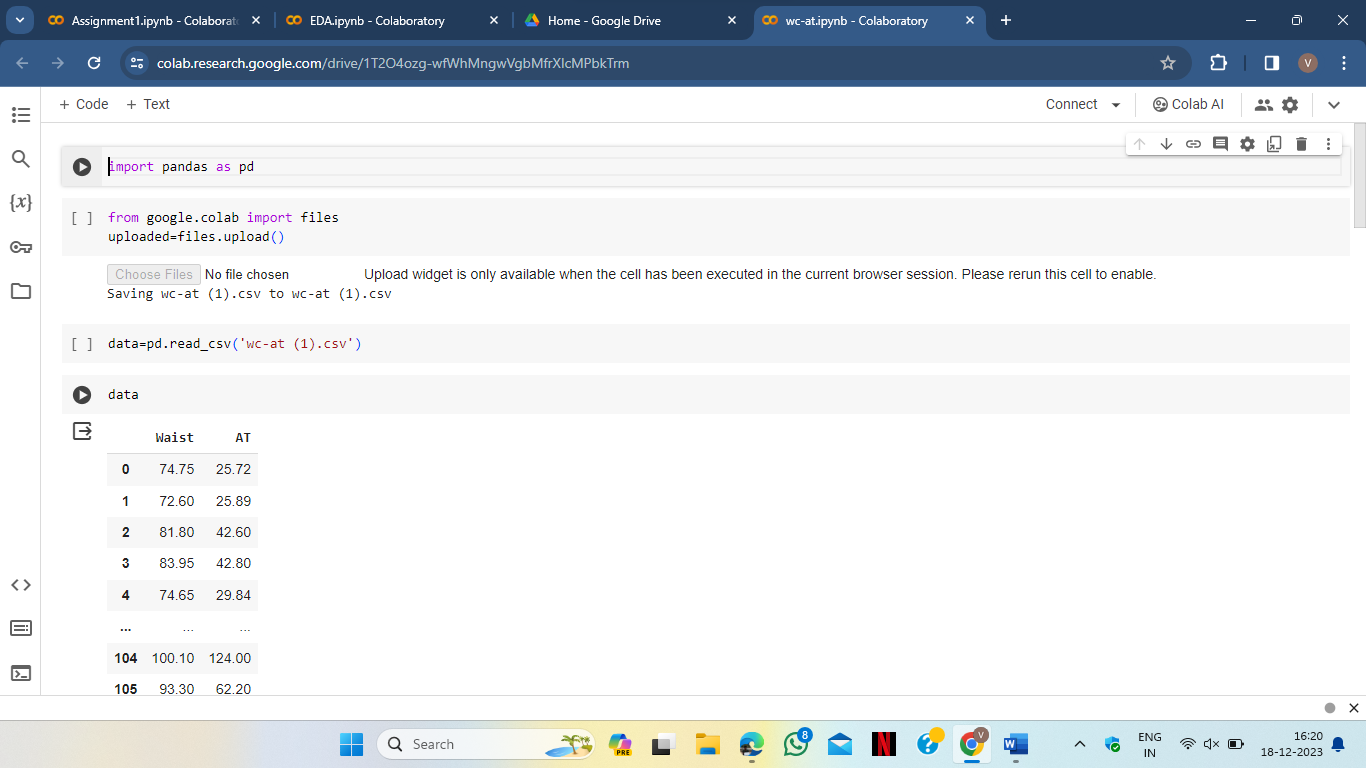


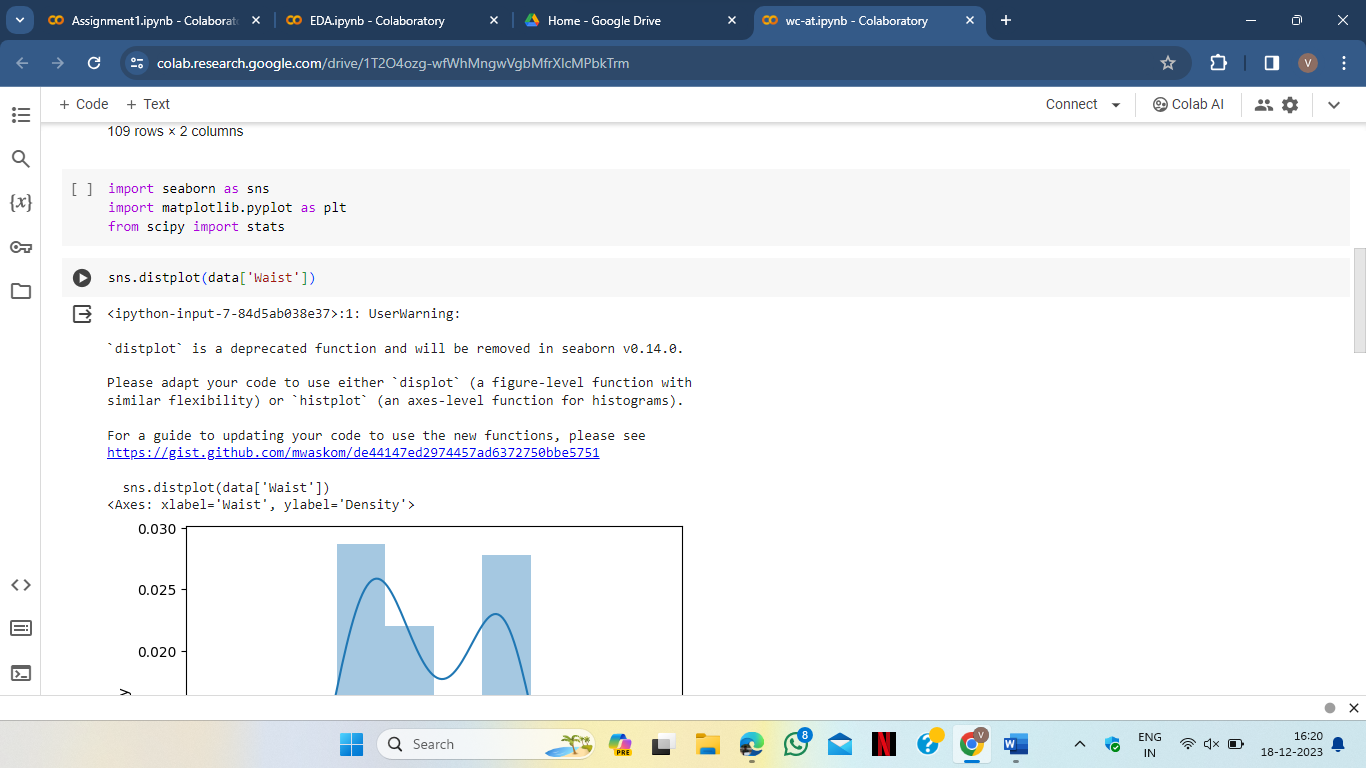


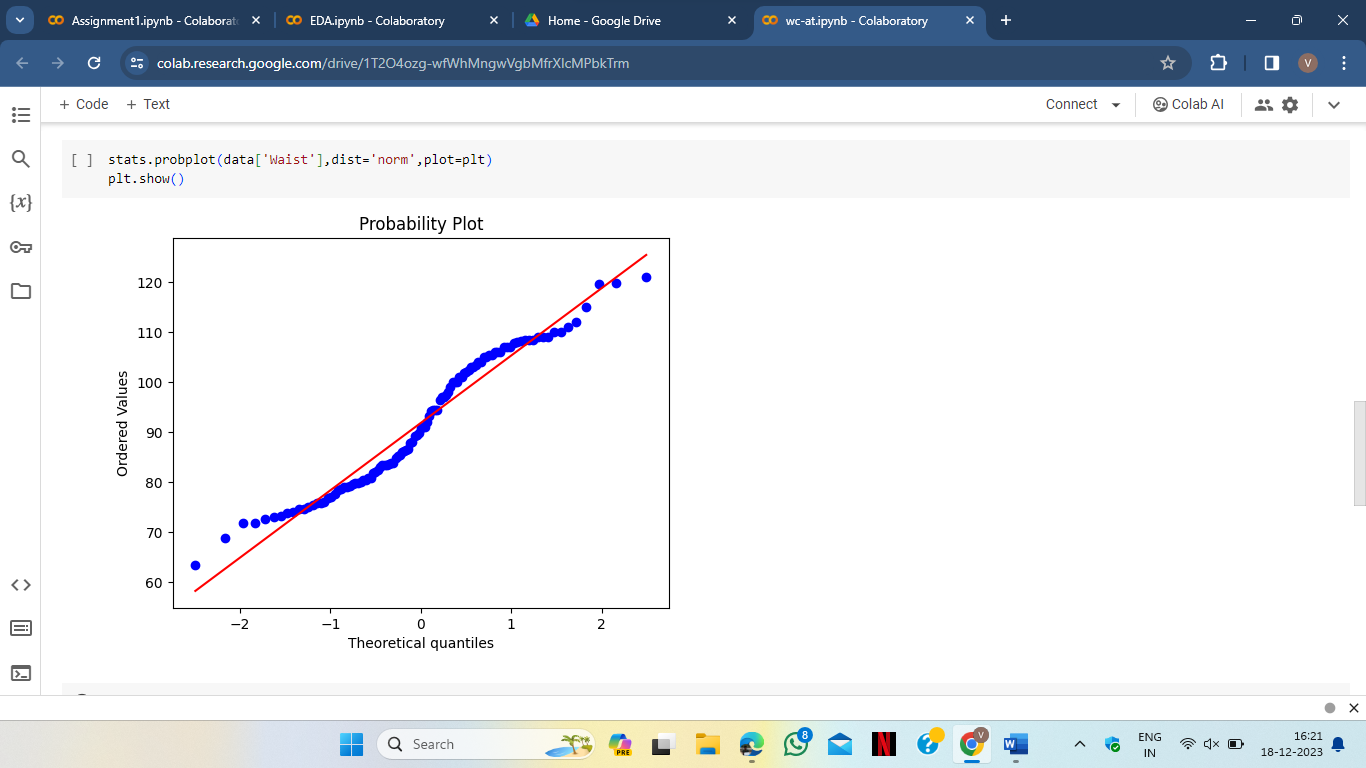
**b)**Check Whether the Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set follows Normal Distribution

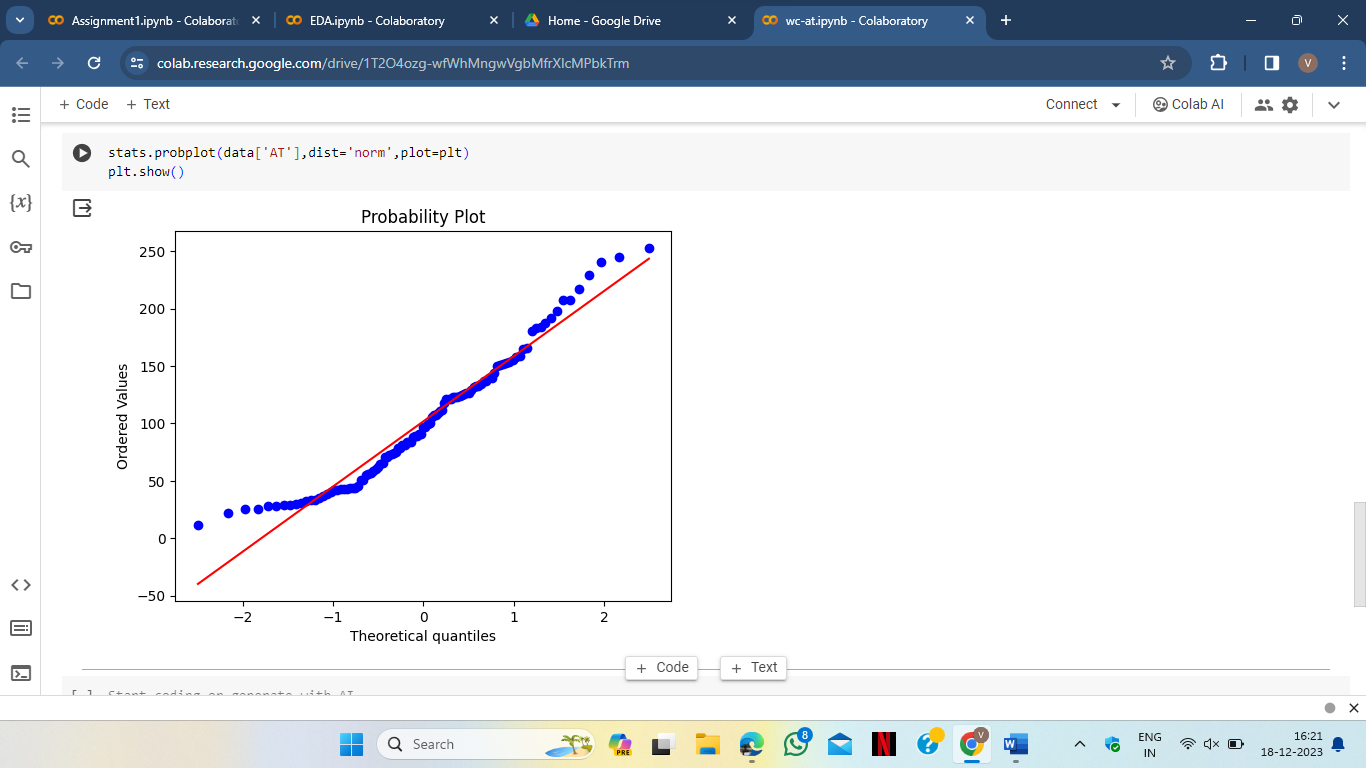
Dataset: wc-at.csv

**Ans)**









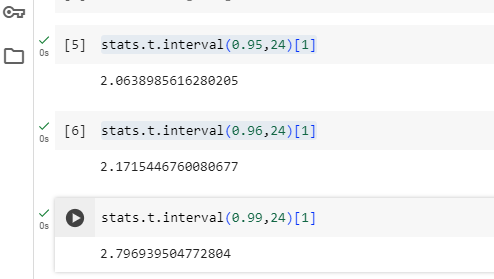
**Q 22)** Calculate the Z scores of 90% confidence interval,94% confidence interval, 60% confidence interval

**Ans)**

****

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

**Ans)**

****

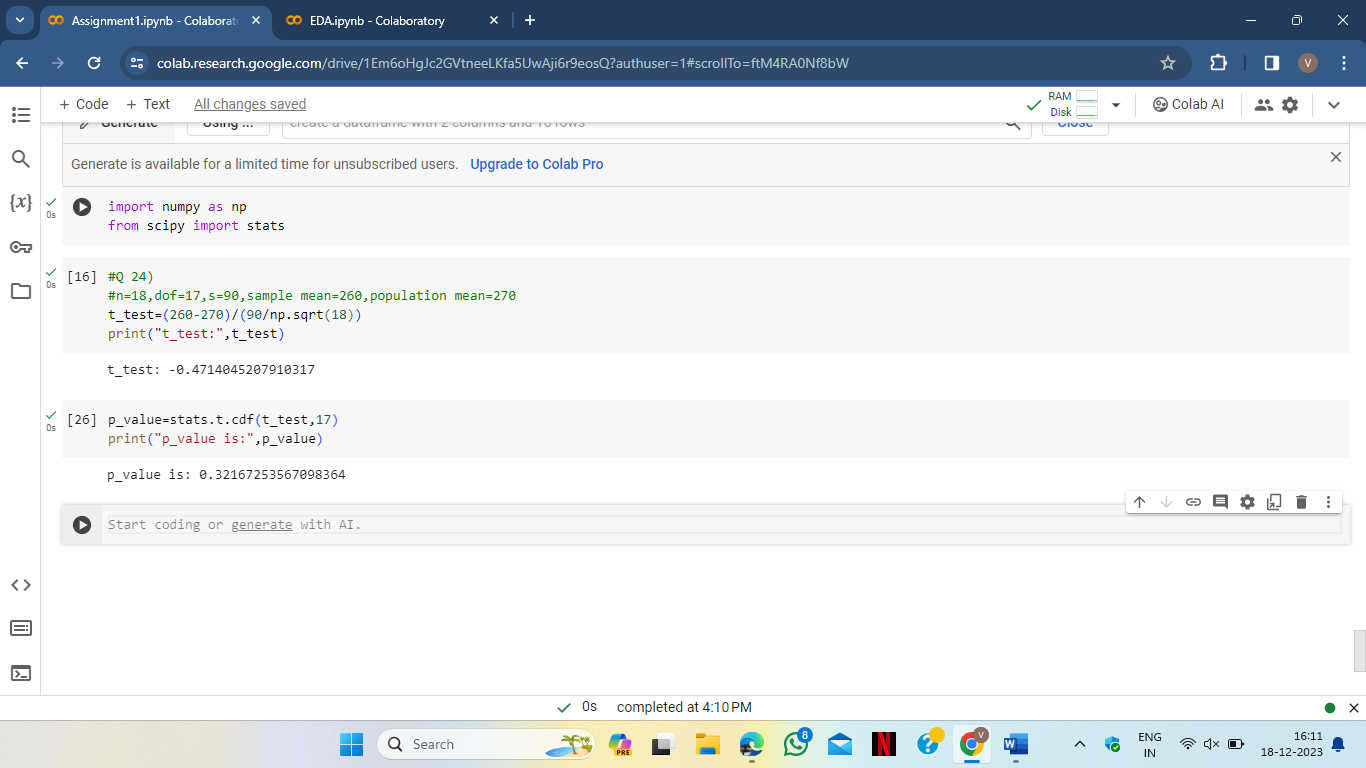
Q 24**)** A Government company claims that an average light bulb lasts 270 days. A researcher randomly selects 18 bulbs for testing. The sampled bulbs last an average of 260 days, with a standard deviation of 90 days. If the CEO's claim were true, what is the probability that 18 randomly selected bulbs would have an average life of no more than 260 days

Hint:

rcode 🡪 pt(tscore,df)

df 🡪 degrees of freedom

**Ans)**

****