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

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 | Ratio |
| Blood Group | nominal |
| Time Of Day | ordinal |
| Time on a Clock with Hands | interval |
| Number of Children | nominal |
| Religious Preference | nominal |
| Barometer Pressure | ratio |
| SAT Scores | nominal |
| Years of Education | nominal |

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

ANS:-

S={HHH,HHT,HTH,THH,TTT,TTH,THT,HTT}

n(S)=8

A=two heads and one tail

n(A)={HHT,THH,HTH}=3

p(A)=n(A)/n(S)= 3/8= 0.375=37.5%

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

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

1. Sum is Equal to 1

s(a)=0

P(a)=n(a)/n(s)=0/36=0

1. Less than or equal to 4

s (b)={(1,1),(1,2),(1,3),(2,1),(2,2),(3,1)}

n (b)=6

P(b)=n(b)/n (s)=6/36=1/6

1. Sum is divisible by 2 and 3

s (c) ={ (1,5), (2,4), (3,3), (4,2), (5,1), (6,6)}

n(c)=6

P(c)=6/36=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?

Total number of balls = (2 + 3 + 2) = 7

Let S be the sample space.

Then, n(S) = Number of ways of drawing 2 balls out of 7

=7C2

=(7×6)(2×1)

=21

Let E = Event of drawing 2 balls, none of which is blue.

∴n(E)= Number of ways of drawing 2 balls out of (2 + 3) balls.

=5C2

=(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:-

Expected Value =(1×0.015)+(4×0.20)+(3×0.65)+(5×0.005)+(6×0.01)+(2×0.120)

*E*(*X*)=(1×0.015)+(4×0.20)+(3×0.65)+(5×0.005)+(6×0.01)+(2×0.120)

=(0.015)+(0.80)+(1.95)+(0.025)+(0.06)+(0.24)

*E*(*X*)=(0.015)+(0.80)+(1.95)+(0.025)+(0.06)+(0.24)

=3.09

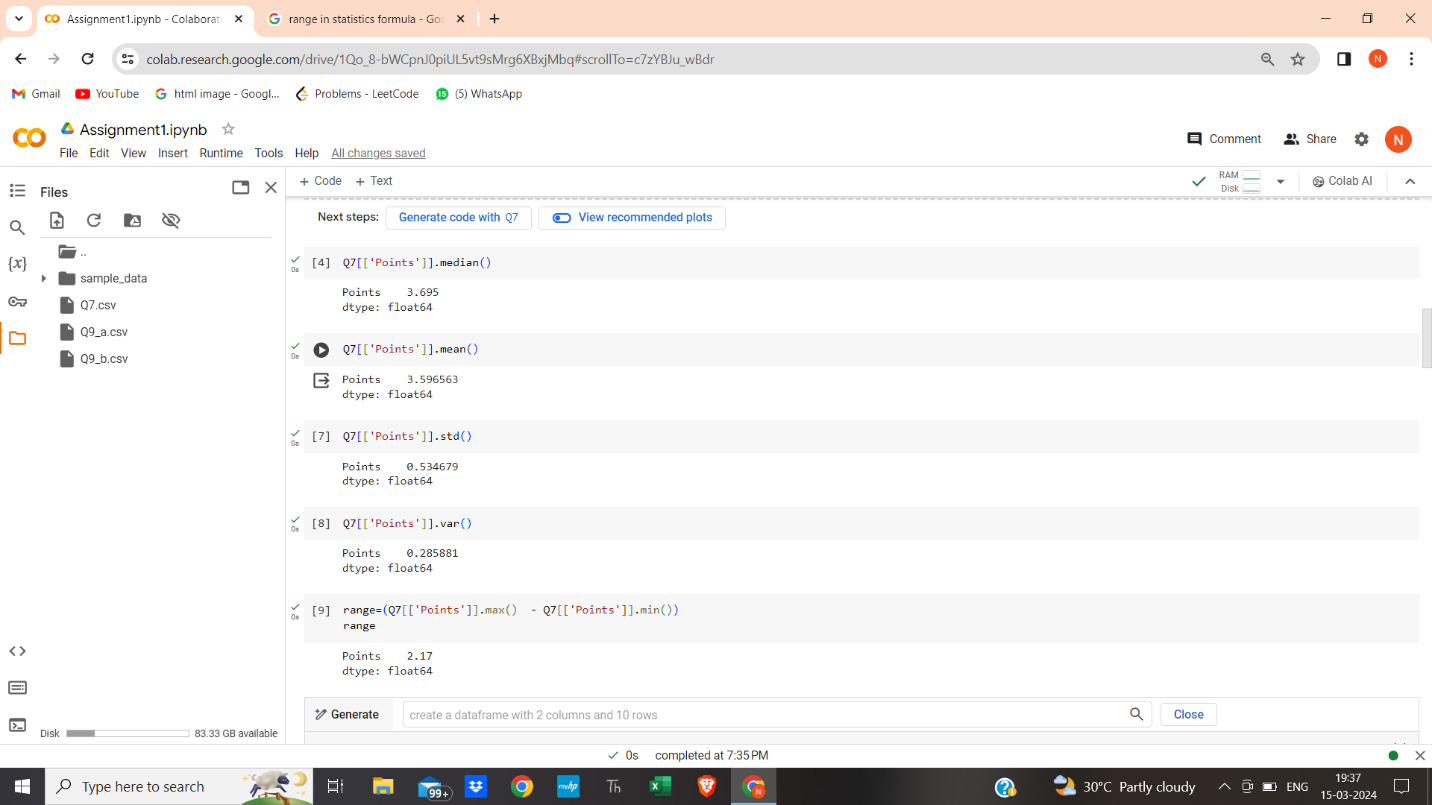
Therefore, the expected number of candies for a randomly selected child is 3.09

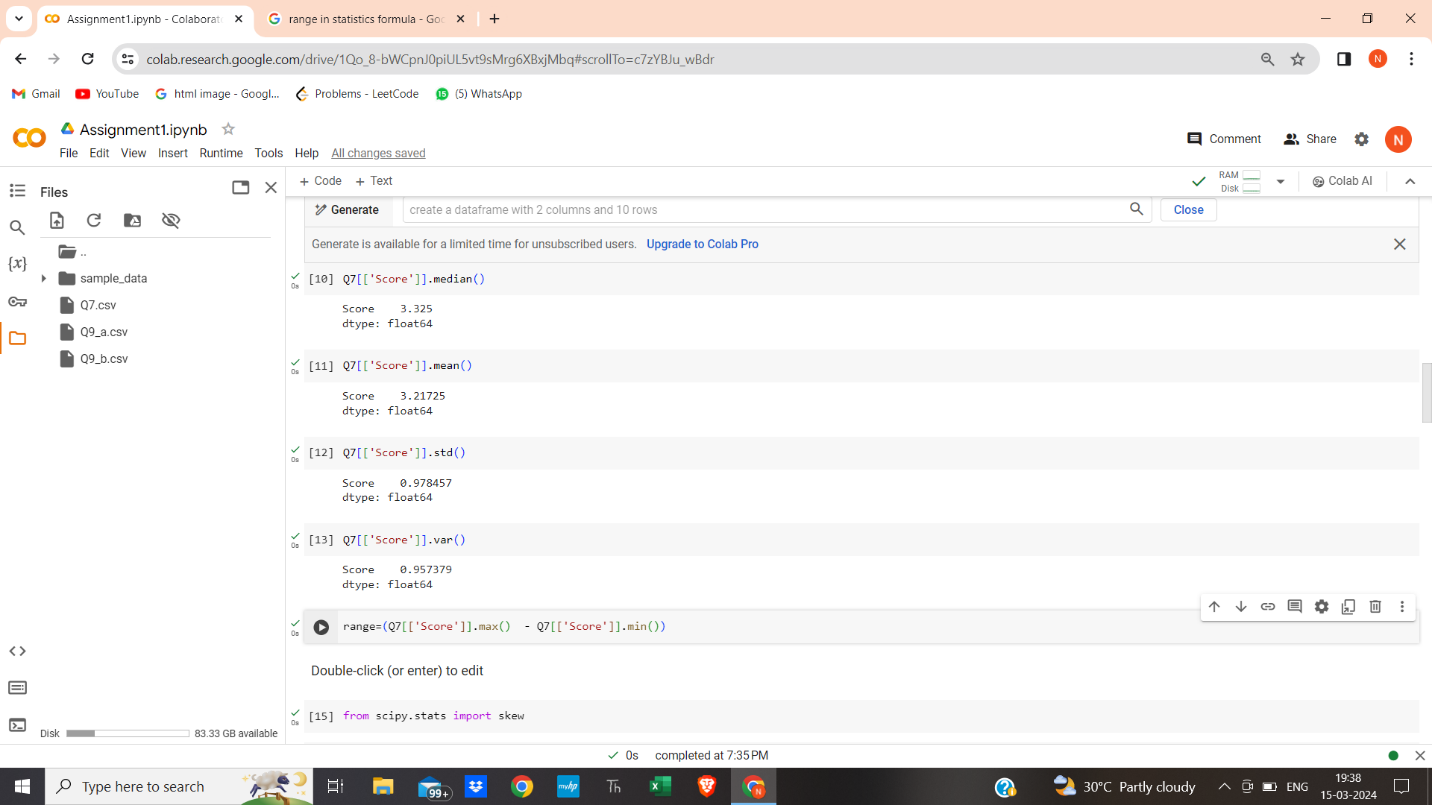
Q7) Calculate Mean, Median, Mode, Variance, Standard Deviation, Range & comment about the values / draw inferences, for the given dataset

* 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**

****

****

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=(108+110+123+134+135+145+167+187+199)/9

=145.333

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

**Cars speed and distance**

**Use Q9\_a.csv**

from scipy.stats import skew

skew(q9['dist'].dropna())

0.7824835173114966

skew(q9['speed'].dropna())

-0.11395477012828319

from scipy.stats import kurtosis

kurtosis(q9['dist'].dropna())

0.24801865717051808

kurtosis(q9['speed'].dropna())

-0.5771474239437371`

**SP and Weight(WT)**

**Use Q9\_b.csv**

skew(q9b['SP'].dropna())

1.5814536794423764

skew(q9b['WT'].dropna())

-0.6033099322115126

kurtosis(q9b['SP'].dropna())

2.7235214865269244

kurtosis(q9b['WT'].dropna())

0.8194658792266849

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



The histograms peak has right skew and tail is on right. Mean > Median. We have outliers on the higher side.



The boxplot has outliers on the maximum side.

**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=

conf\_94 = t.interval(0.94, 1999, 200, 30/np.sqrt(2000))

print(np.round(conf\_94,0))

print(conf\_94)

[199. 201.]

(198.7376089443071, 201.2623910556929)

2) conf\_98 = t.interval(0.98, 1999, 200, 30/np.sqrt(2000))

print(np.round(conf\_98,0))

print(conf\_98)

[198. 202.]

(198.4381860483216, 201.5618139516784)

3) conf\_96 = t.interval(0.96, 1999, 200, 30/np.sqrt(2000))

print(np.round(conf\_96,0))

print(conf\_96)

[199. 201.]

(198.6214037429732, 201.3785962570268)

**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:-

Mean=34+36+36+38+38+39+39+40+40+41+41+41+41+42+42+45+49+56/18

=​?? 38.0556

Median= 39+40/2=39.5​/ ?

Variance=(34−38.0556)2+(36−38.0556)2+(36−38.0556)2 +(38−38.0556)2+(39−38.0556)2 +(39−38.0556)2+(40−38.0556)2+(40−38.0556)2+(41−38.0556)2+(41−38.0556)2+(41−38.0556)2+(41−38.0556)2+(42−38.0556)2+(42−38.0556)2+(45−38.0556)2+(49−38.0556)2+(56−38.0556)2​

==381.278​//18=21.1828

Standard Deviation=4.599

+ +

+

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

Ans:- If mean and median of data are equal then graph is symmetric and graph appear same on both side.

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

If the mean is greater than the median, the distribution is positively skewed.

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

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

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

Positive kurtosis indicates heavier tails and a more peaked distribution

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

Negative kurtosis suggests lighter tails and a flatter distribution.

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



What can we say about the distribution of the data?

Ans:-Assymetric Distribution

What is nature of skewness of the data?

Ans:-Negative skewness

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

Ans:- IQR=18-10=8  
  
  
  
Q19) Comment on the below Boxplot visualizations?



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

Ans: First there are no outliers. Second 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.

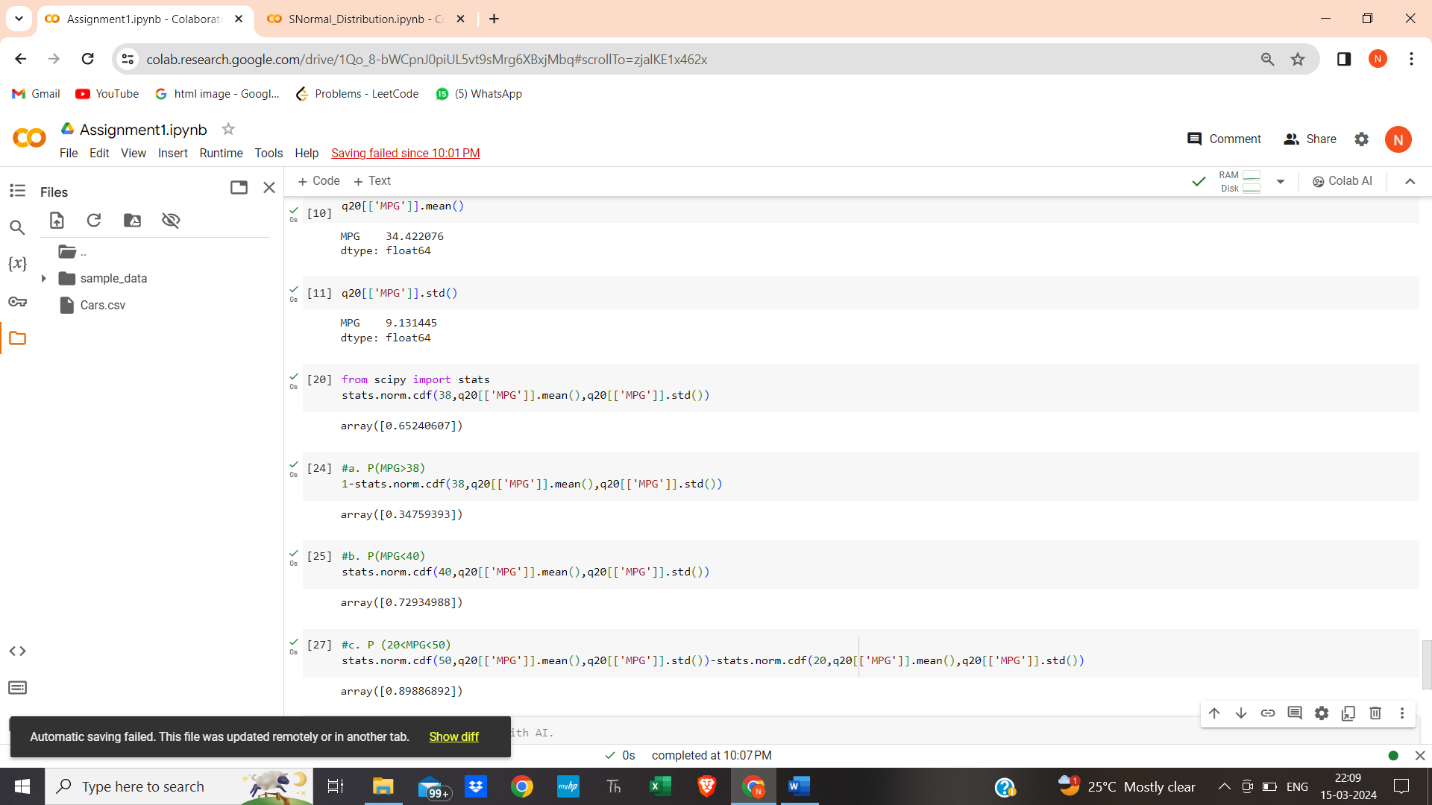
Q 20) Calculate probability from the given dataset for the below cases

Data \_set: Cars.csv

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

MPG <- Cars$MPG

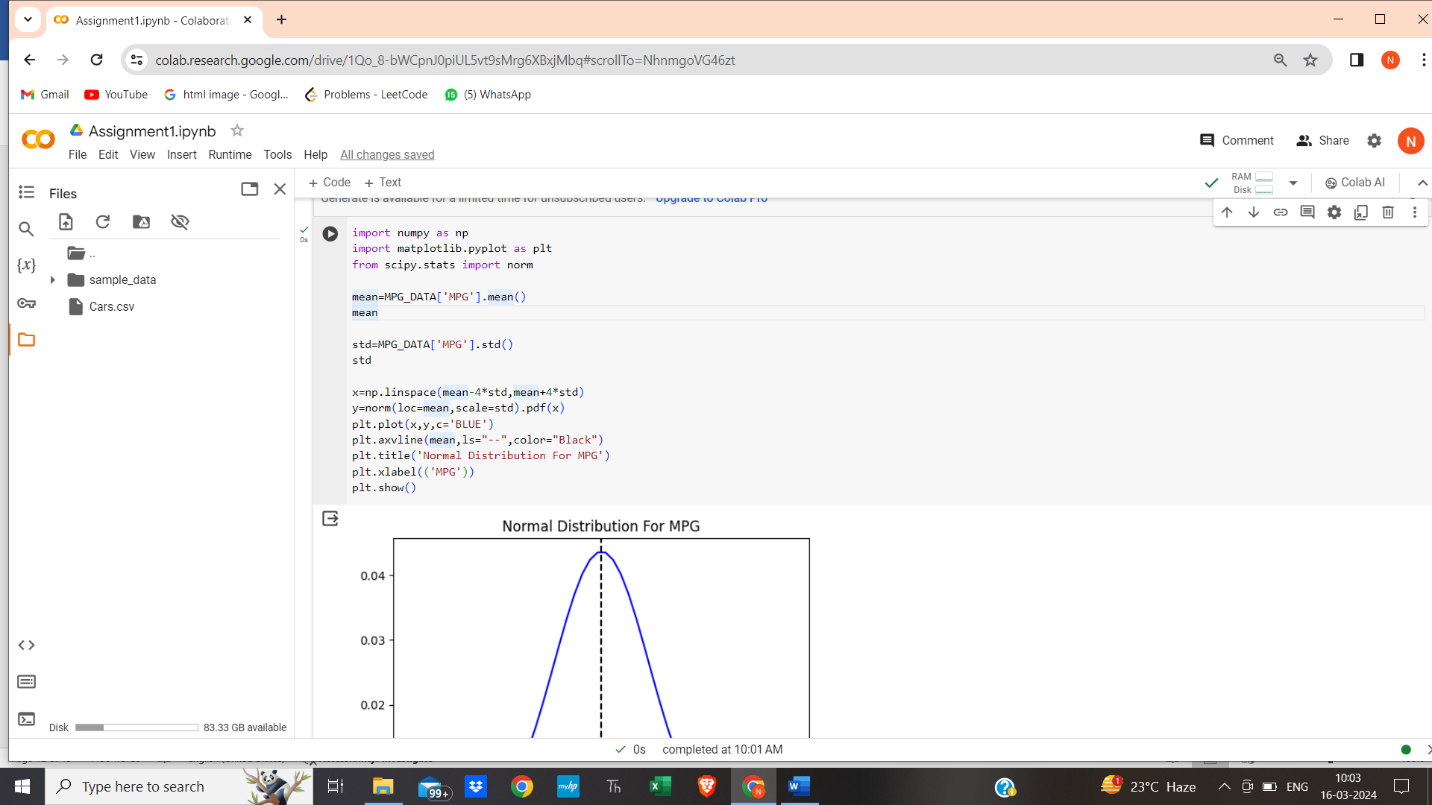
* 1. P(MPG>38)
  2. P(MPG<40)
  3. P (20<MPG<50)

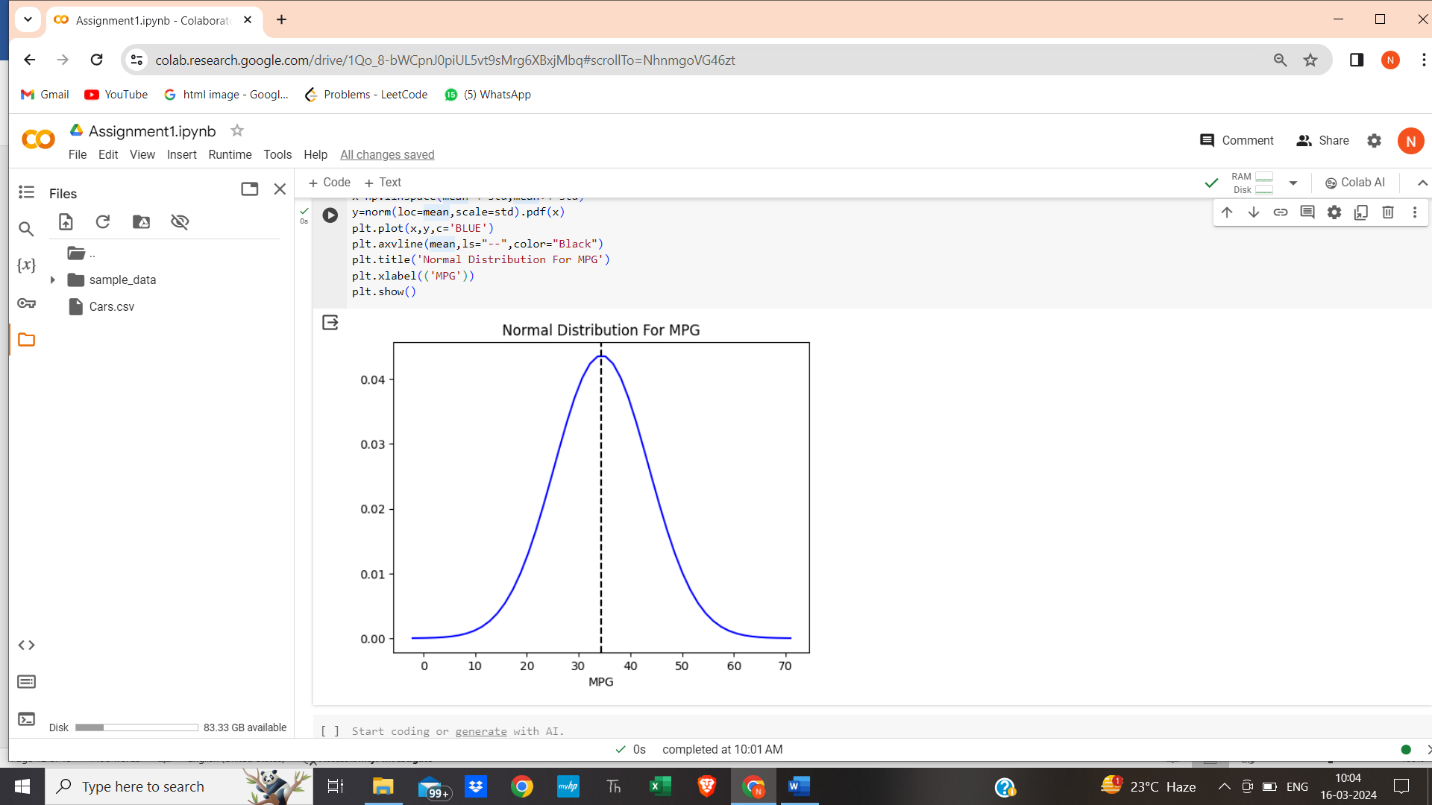


Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

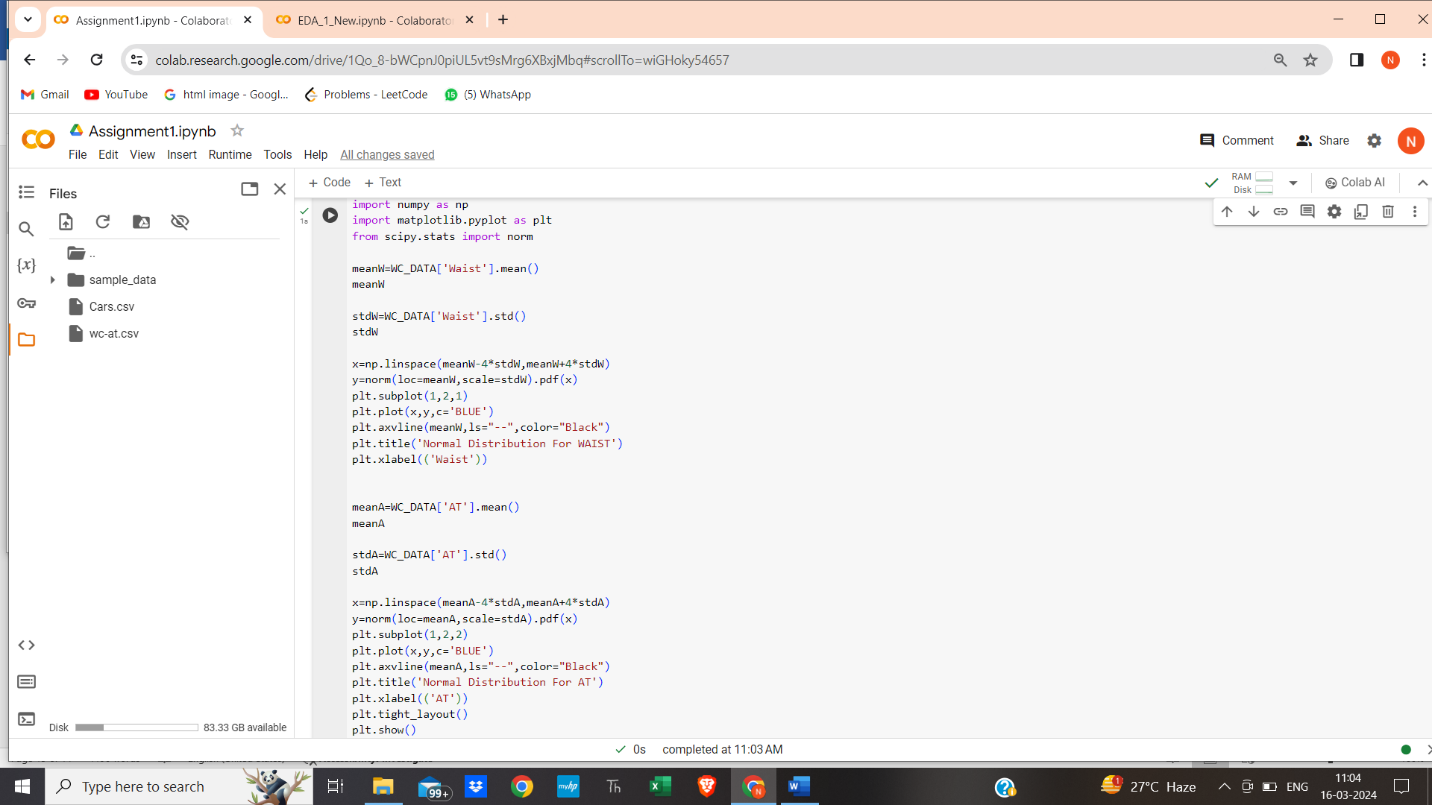
Dataset: Cars.csv

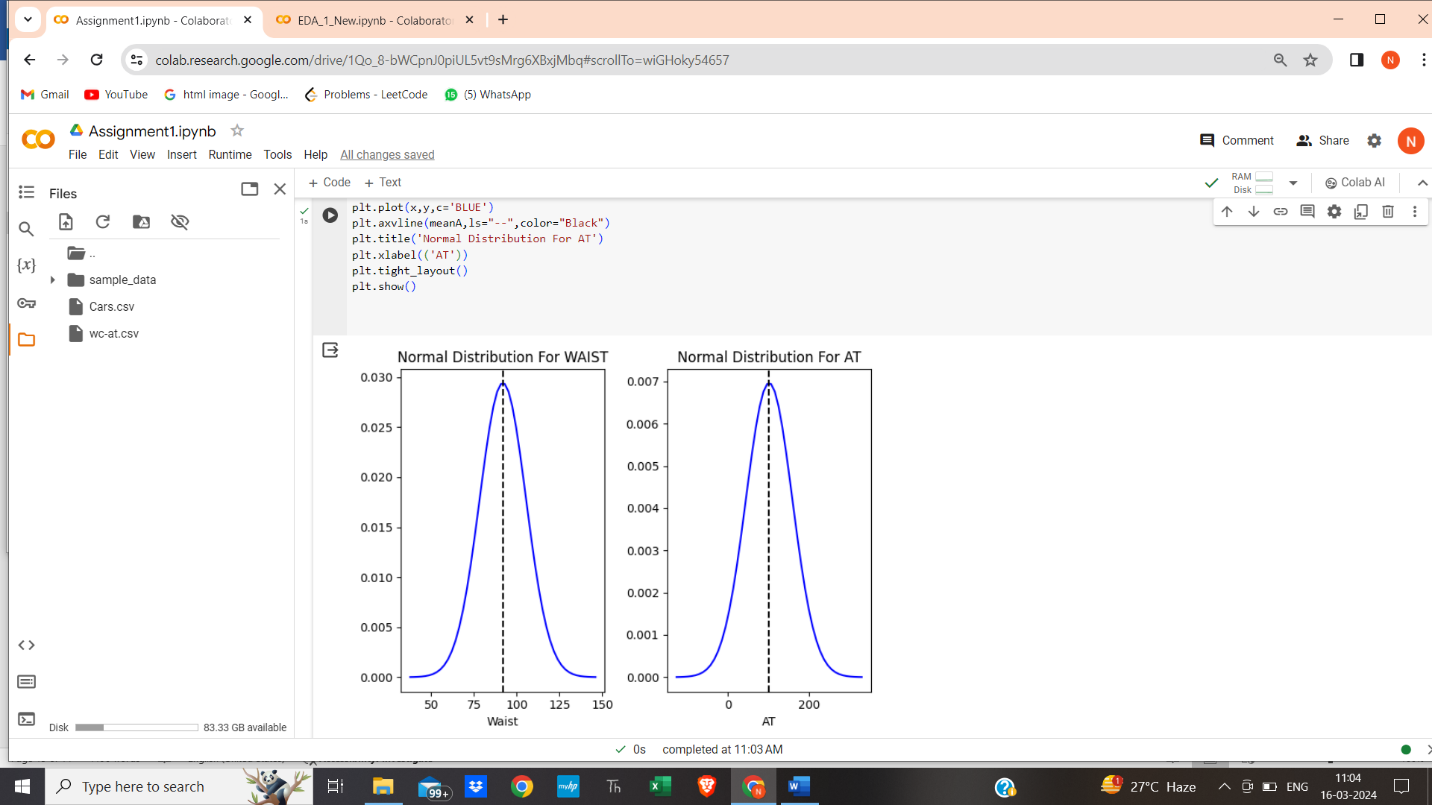




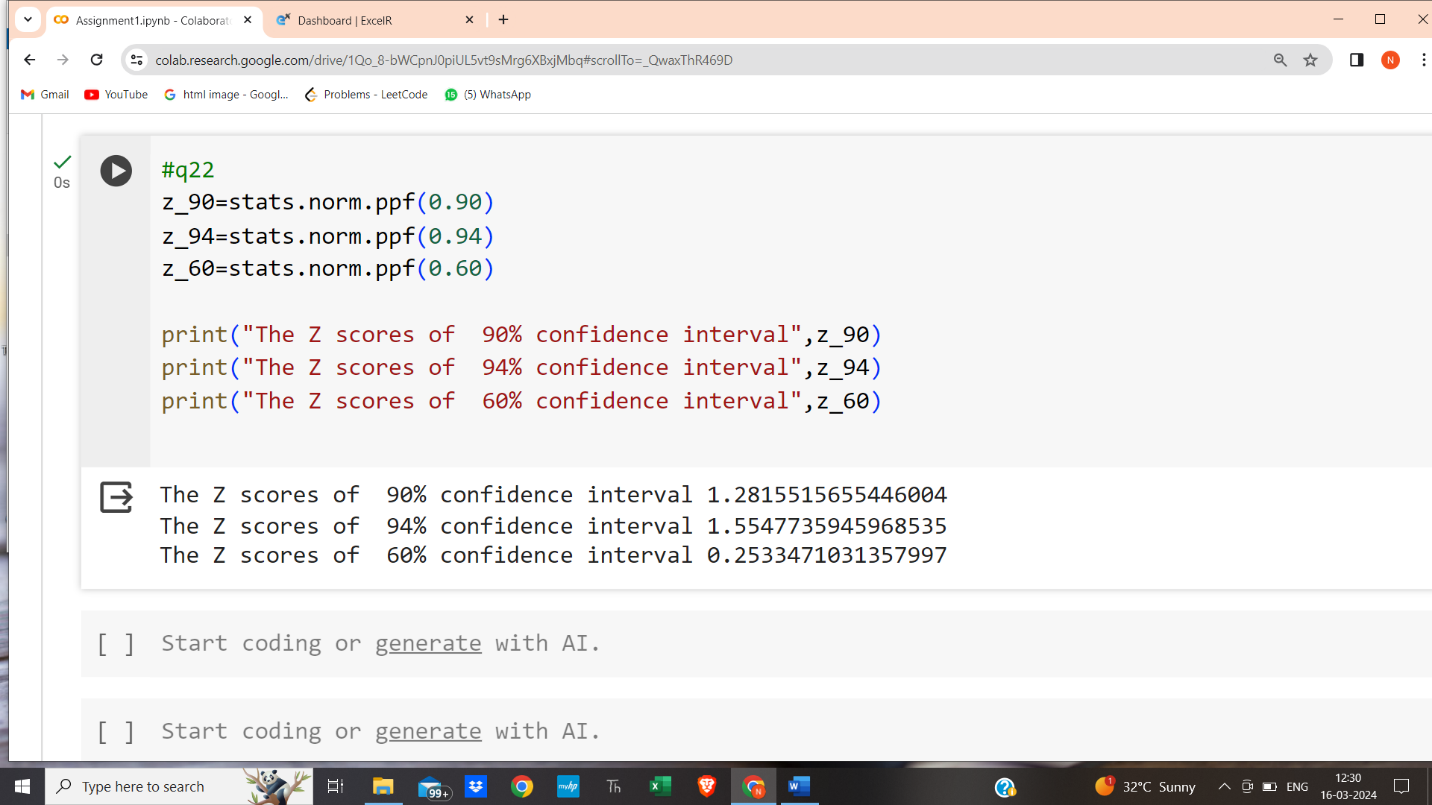
1. Check Whether the Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set follows Normal Distribution

Dataset: wc-at.csv

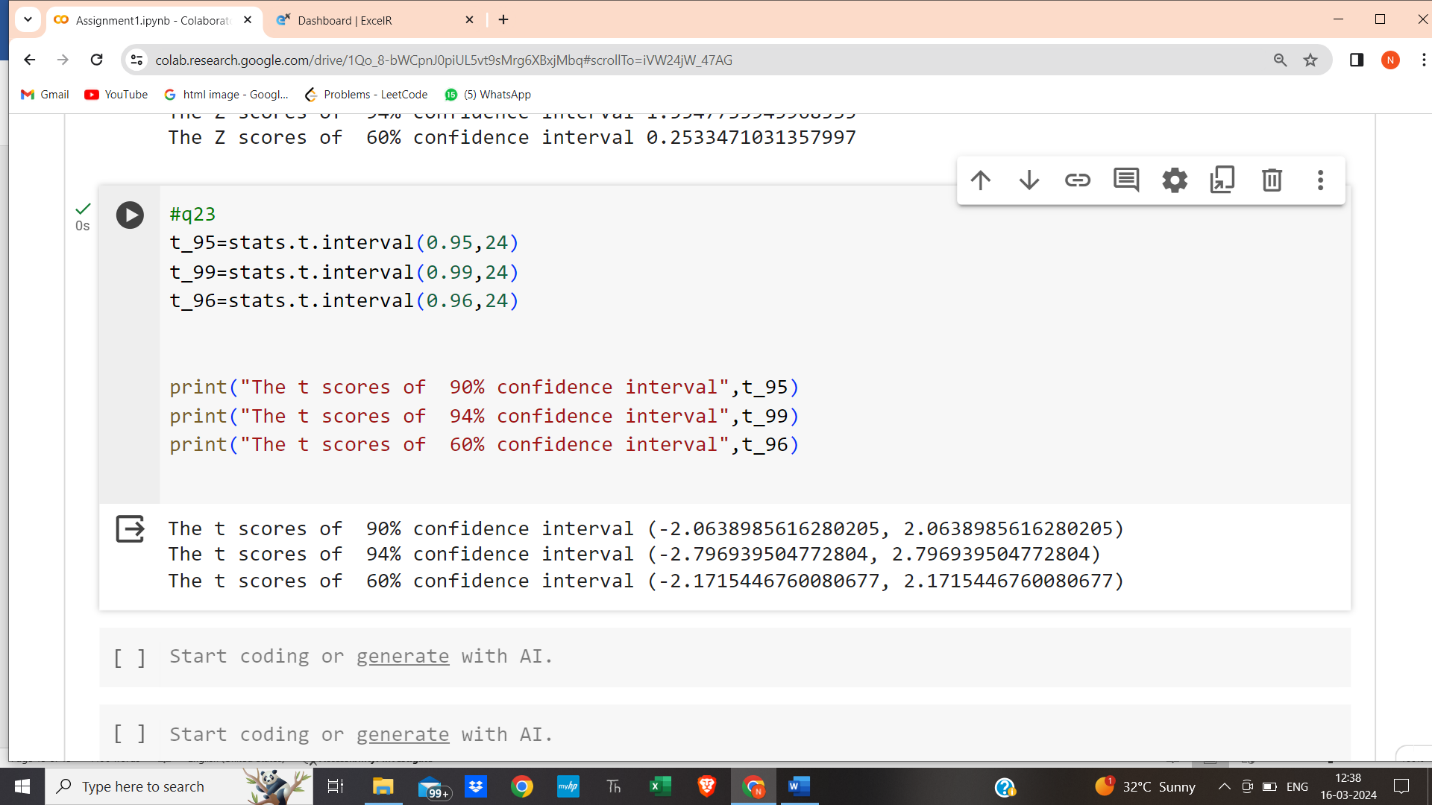




Q 22) Calculate the Z scores of 90% confidence interval,94% confidence interval, 60% confidence interval



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

