|  |  |
| --- | --- |
| 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 | Interval |
| Weight | Ratio |
| Hair Color | Nominal |
| Socioeconomic Status | Nominal |
| Fahrenheit Temperature | Ratio |
| 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 | Nominal |
| 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?

ANS: Combinations | HHH,HHT,HTH,HTT,THH,THT,TTH,TTT |

No of combinations with 2 head and a tail : 3

Probability : 3/8 = 0.37

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:

1. 0
2. 1/6 = 0.16
3. 1/6 = 0.16

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: Probability : 10/21 = 0.47

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 of candies for a randomly selected child :

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

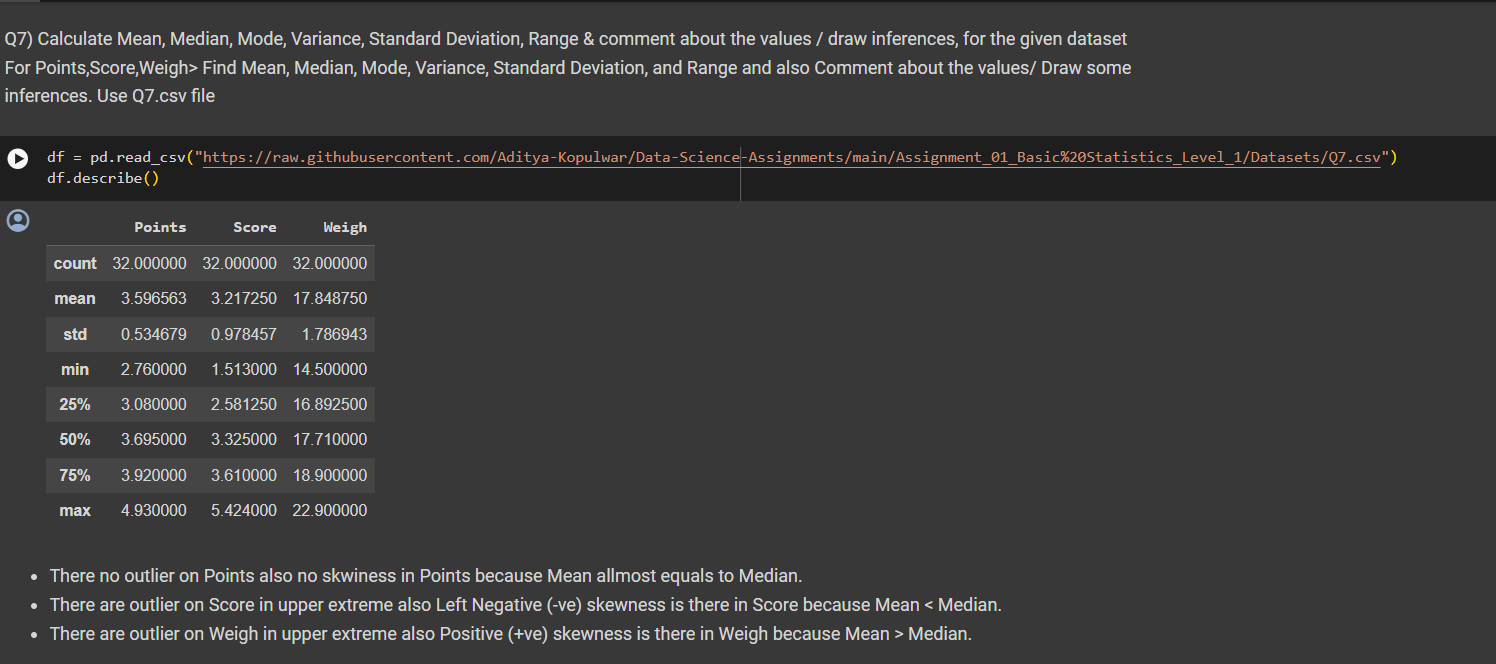
= 3.09

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**



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: Weights | 108, 110, 123, 134, 135, 145, 167, 187, 199 |

Probability : 1/9 = 0.11

Expected Value of the Weight of that patient :

0.11\*108 + 0.11\*110 + 0.11\*123 + 0.11\*134 + 0.11\*135 + 0.11\*145 + 0.11\*167 + 0.11\*187 + 0.11\*199

= 145.33

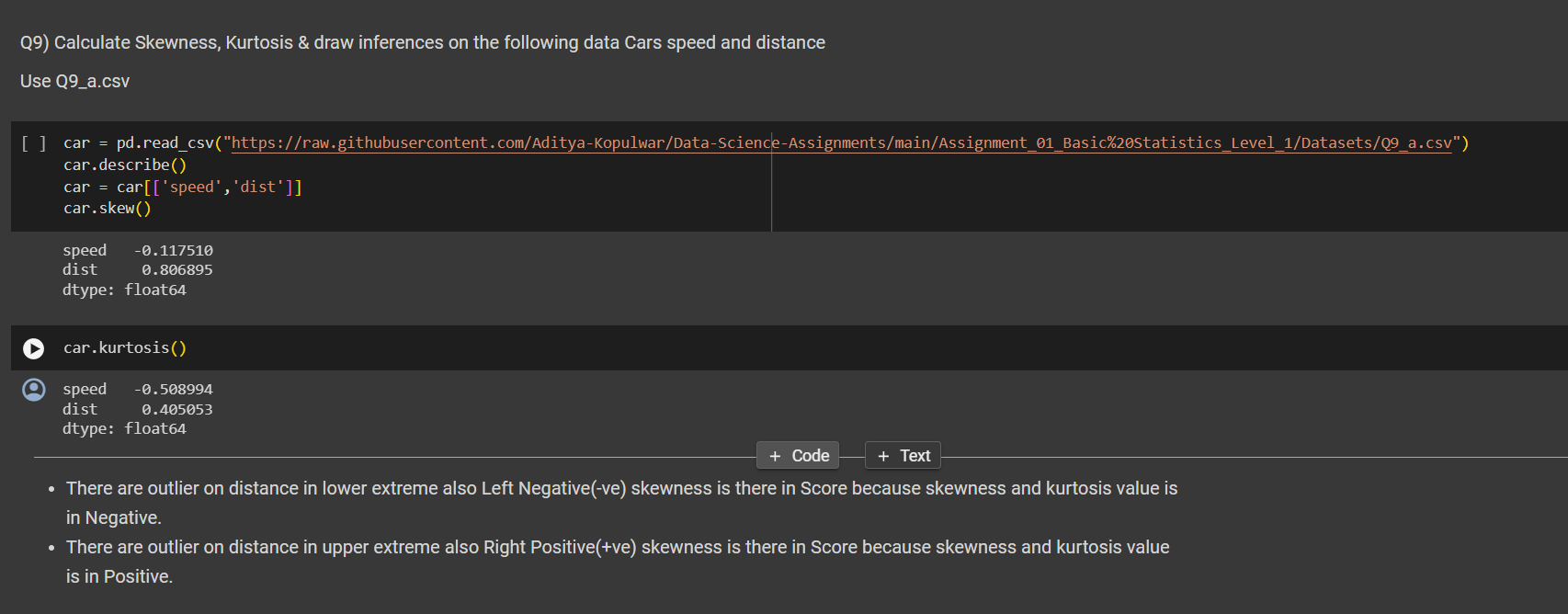
**Q9) Calculate Skewness, Kurtosis & draw inferences on the following data**

**Cars speed and distance**

**Use Q9\_a.csv**

**SP and Weight(WT)**

**Use Q9\_b.csv**

****

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



ANS: There are 8 outlier on weight in upper extreme also a Right(+ve) skewness in it.

**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: sample = 2,000, population = 3,000,000,

sample mean = 200, sample sd = 30

confidence interval 94% and alpha = 0.06 :

lower extreme : 200-st.t.ppf(0.97,1999)\*30/math.sqrt(2000)

=198.7376

upper extreme : 200+st.t.ppf(0.97,1999)\*30/math.sqrt(2000)

= 201.26

confidence interval 96% and alpha = 0.04 :

lower extreme : 200-st.t.ppf(0.98,1999)\*30/math.sqrt(2000)

=198.62

upper extreme : 200+st.t.ppf(0.98,1999)\*30/math.sqrt(2000)

= 201.378

confidence interval 98% and alpha = 0.02 :

lower extreme : 200-st.t.ppf(0.99,1999)\*30/math.sqrt(2000)

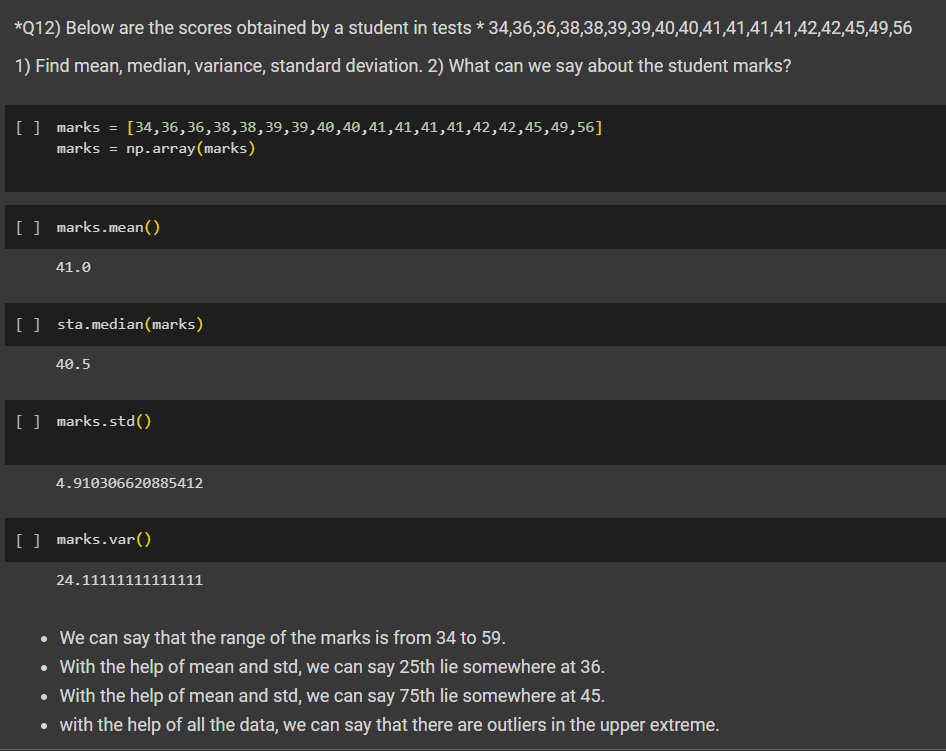
=198.438

upper extreme : 200+st.t.ppf(0.99,1999)\*30/math.sqrt(2000) = 201.561

**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?



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

ANS: As mean = median there will be no skewness and it will follow normal distribution

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

ANS: If mean > median then Right(Positive) skewness.

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

ANS: If mean < median then Left(Negative) skewness.

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

ANS: If there is +ve kurtosis then there is outlier in upper extreme.

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

ANS: If there is -ve kurtosis then there is outlier in lower extreme.

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



What can we say about the distribution of the data?

What is nature of skewness of the data?

What will be the IQR of the data ()?   
  
ANS:

* 1. It is not normal distributed
  2. There is Left (Negative) skewness in the plot
  3. The IQR (approximately) is from 10 to 18

Q19) Comment on the below Boxplot visualizations?



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

ANS: Boxplot 1(approximately):-

1. Median : 262.5
2. Min : 240
3. Q1 : 260
4. Q3 : 280
5. Max : 290

Boxplot 2(approximately):-

1. Median : 262.5
2. Min : below 200
3. Q1 : 220
4. Q3 : 310
5. Max : above 325

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

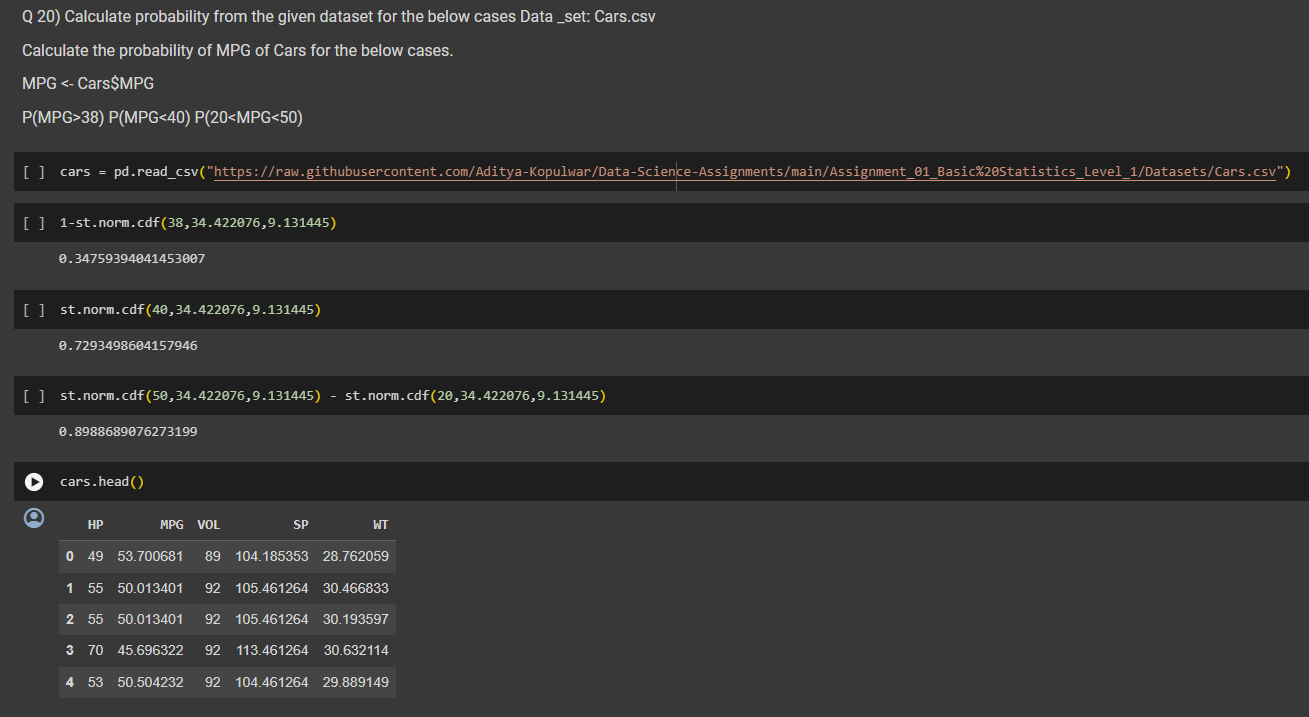
Data \_set: Cars.csv

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

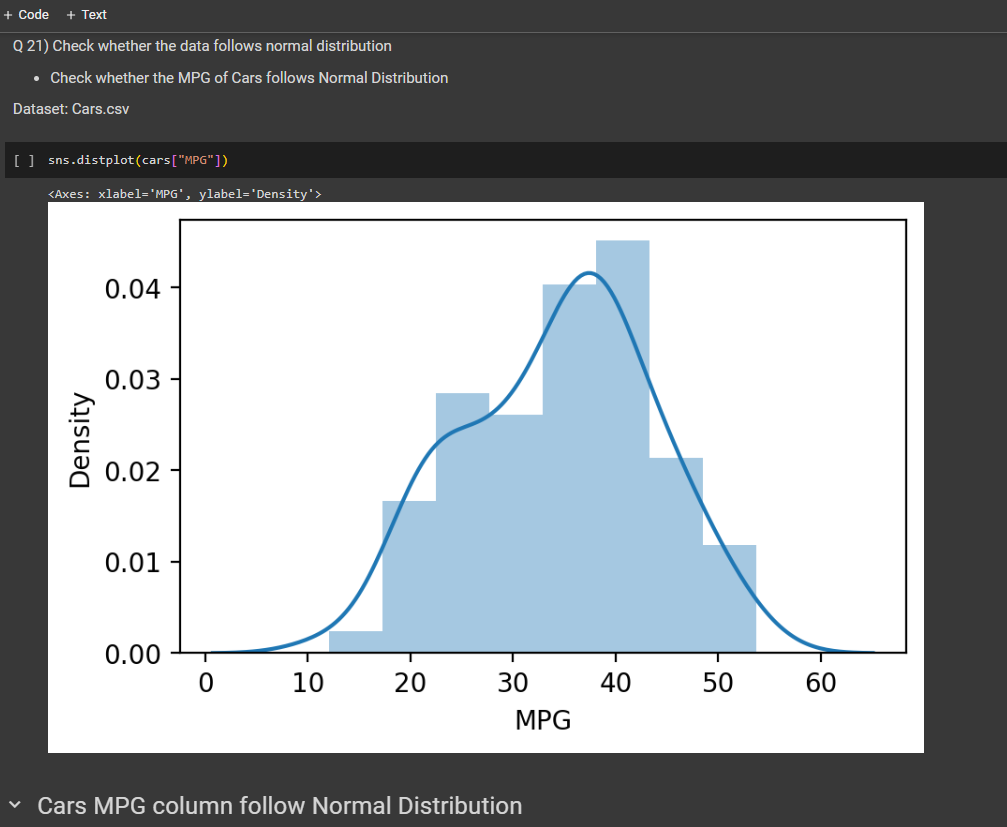
MPG <- Cars$MPG

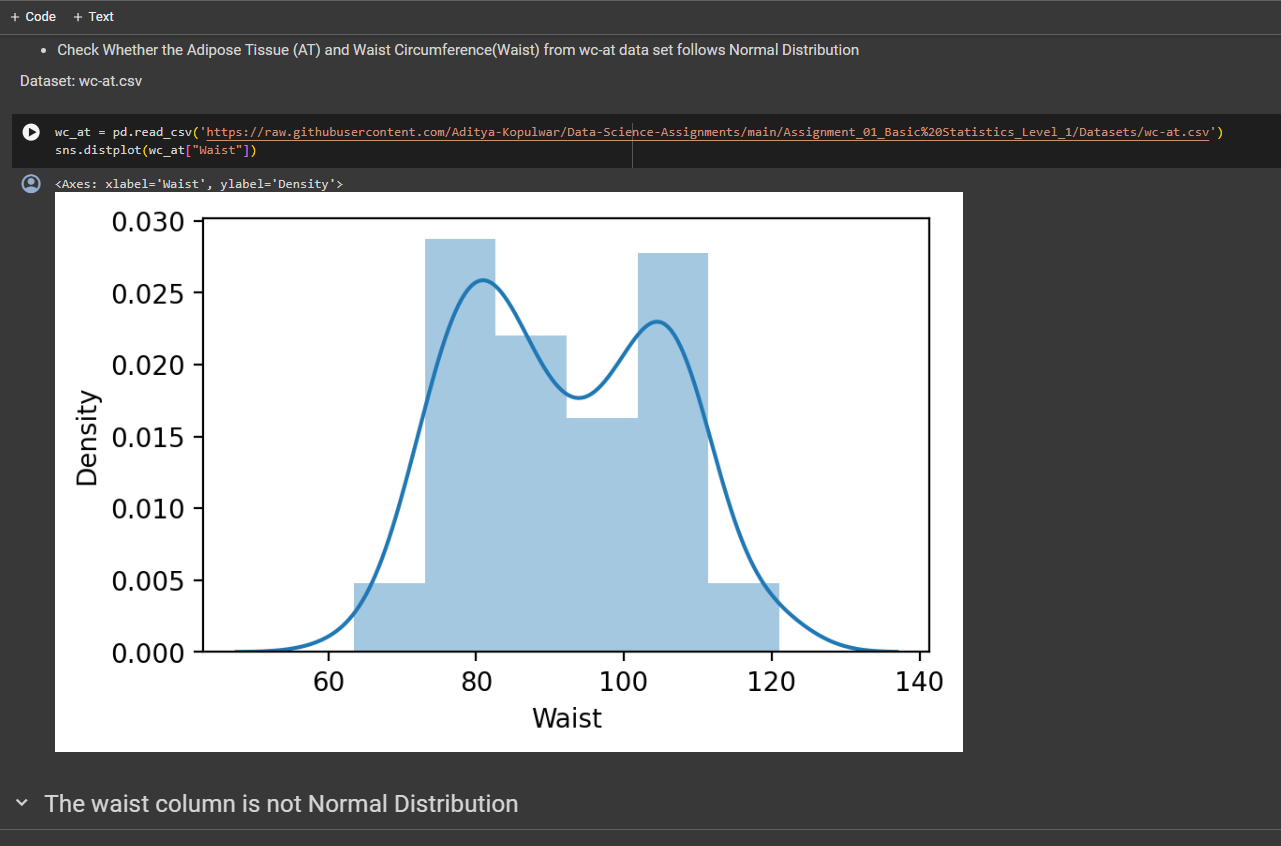
* 1. P(MPG>38)
  2. P(MPG<40)

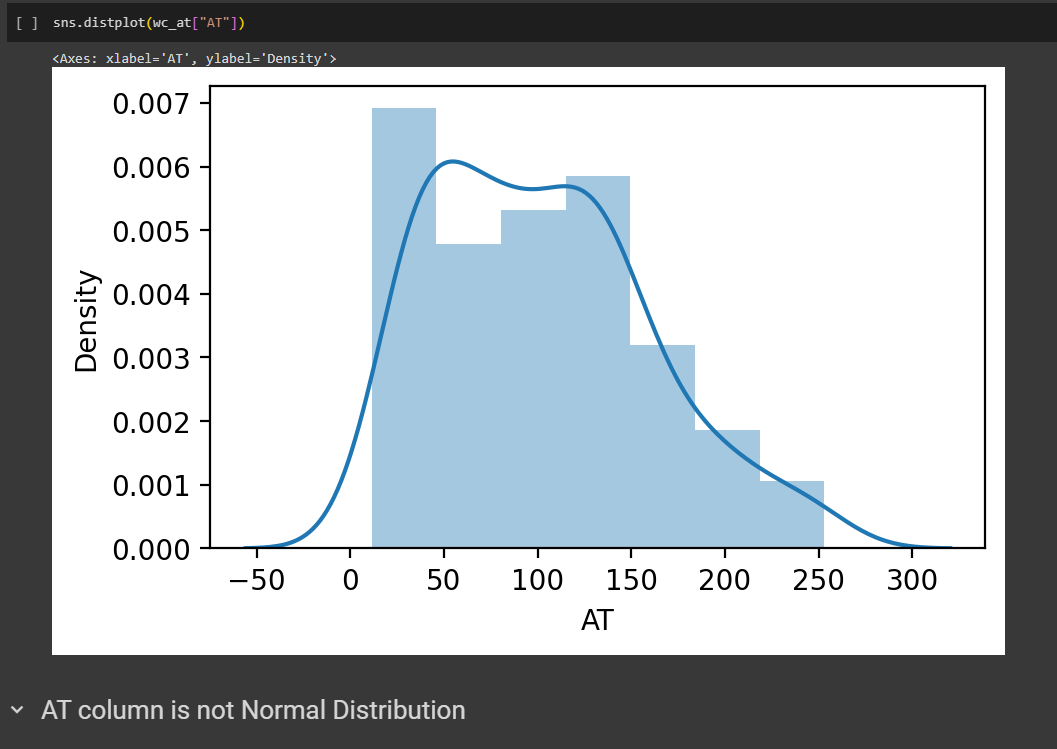
c. P (20<MPG<50)



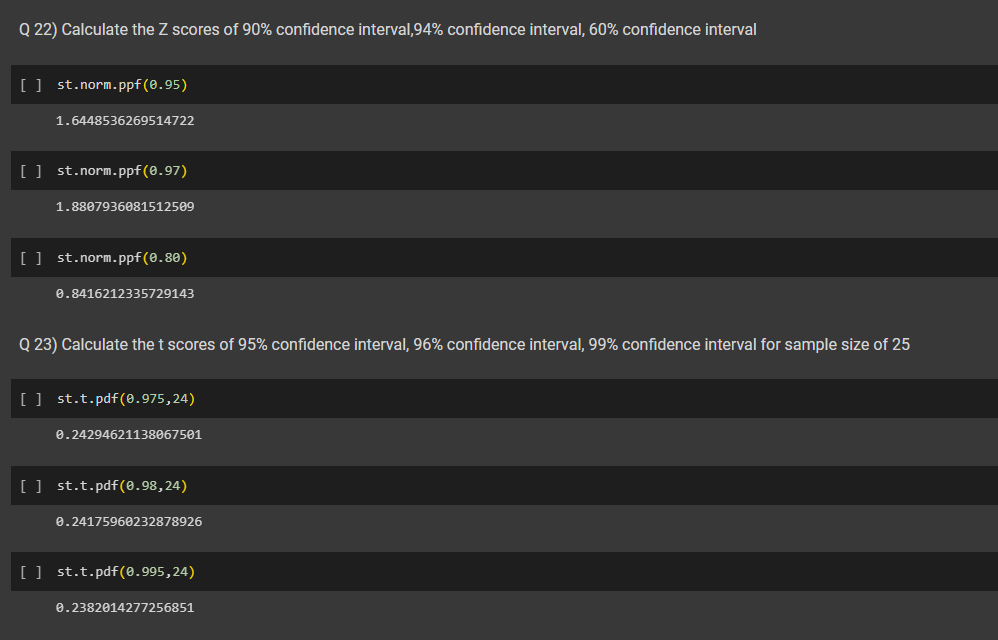
Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution Dataset: Cars.csv
2. Check Whether the Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set follows Normal Distribution

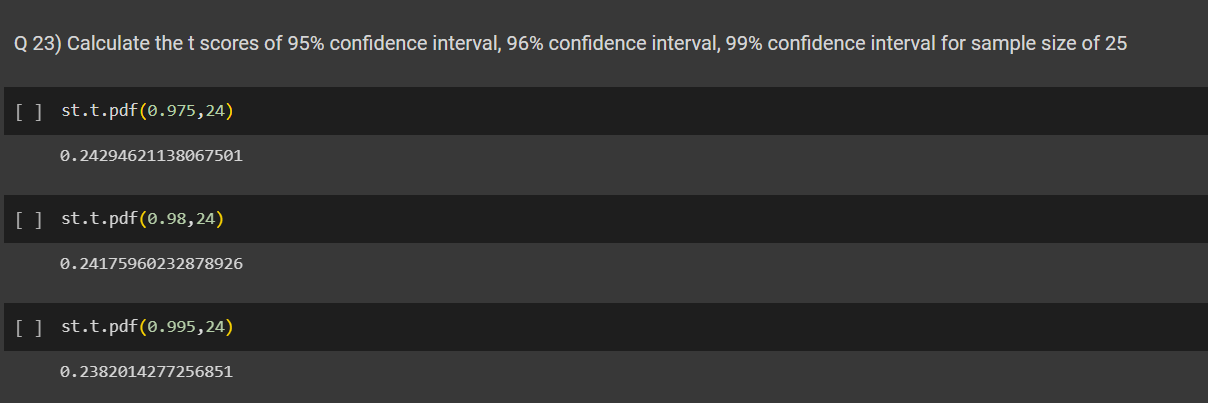
Dataset: wc-at.csv



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

