|  |  |
| --- | --- |
| 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 | Categorical |
| Number of kids | Discrete |
| Number of tickets in Indian railways | Discrete |
| Number of times married | Discrete |
| Gender (Male or Female) | Categorical |

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 | Ordinal |
| Religious Preference | Nominal |
| Barometer Pressure | Ratio |
| SAT Scores | Ratio |
| Years of Education | Interval |

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

Ans) We have different possibilities of trails

H H H

H H T

H T H

T H H

T T T

T H H

T H T

T T H

So, there are total 8 possibilities, in which we have 3 possibilities of two heads and one tail.

Probability = Total of of two heads and one tail.

Total number of possibilities

= 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 2 and 3

Possibilities

(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)

1. Equal to 1

Ans) 0/36= 0

1. Less than or equal to 4

Ans) possibilities (1,1), (1,2), (1,3), (2,1), (2,2), (3,1) = 6/36 = 1/6

1. Sum is divisible by 2 and 3

Ans) (1,5), (2,4), (3,3), (4,2),(5,1), (6,6) = 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?

Ans) Total balls = 7, Drawing 2 balls from 7 =

nCr = n!/r!(n-r)!.=7C2= 7\*6/2\*1 = 42/2 = 21

Total red and green balls are = 5

2 balls needs to drawn from 5 balls( Red and Green)/ = 5C2 = 5\*4/2\*1 = 20/2 = 10

Probability = 2 balls needs to drawn from 5 balls( Red and Green)/ 2 balls drawn from Total balls

Probability = 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 = E ( X ) = ∑ x P ( x ).

Probability of Expected number of candies for a randomly selected child

= each individual Candies count \*Probability of each individual =

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

= 3.09

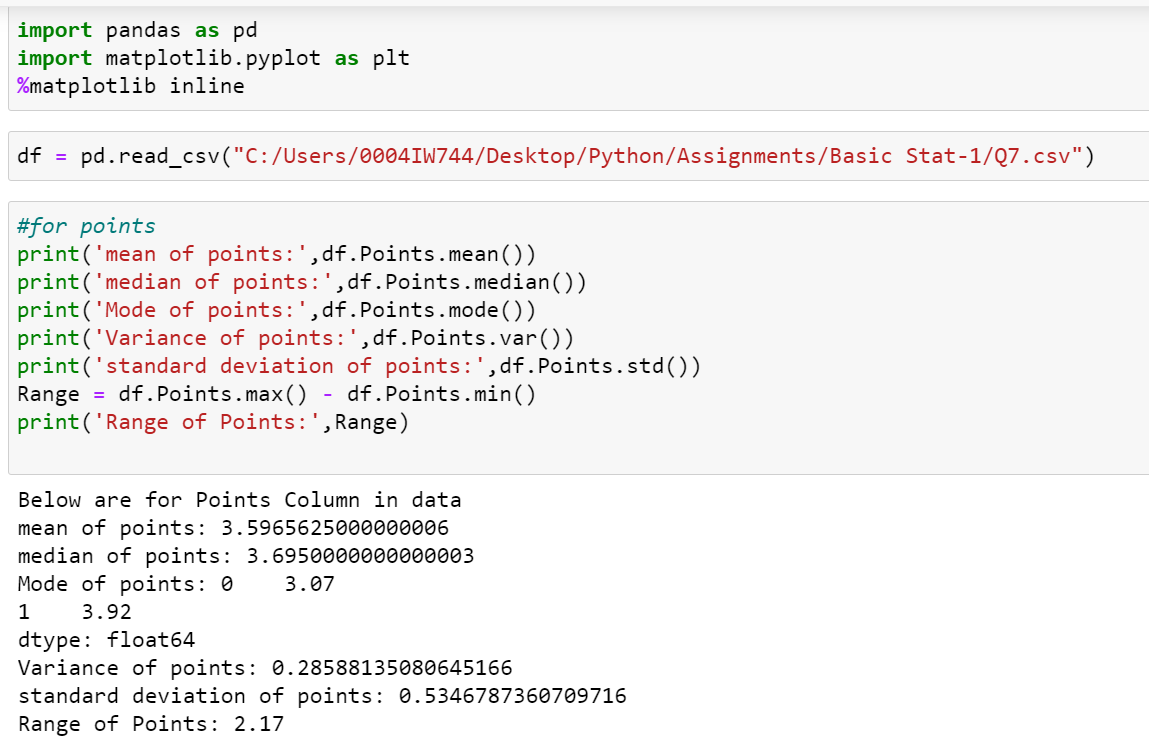
Q7) Calculate Mean, Median, Mode, Variance, Standard Deviation, Range & comment about the values / draw inferences, for the given dataset

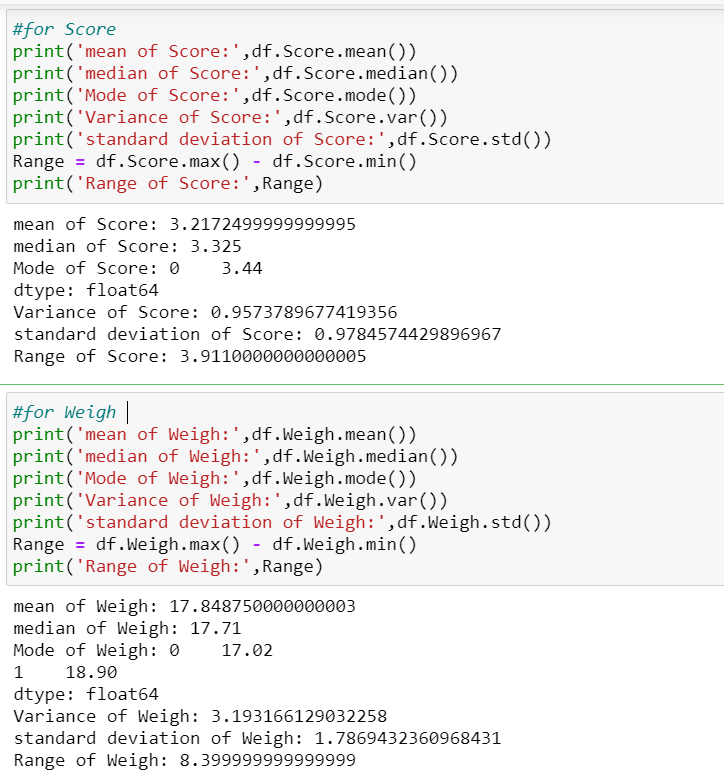
* For Points,Score,Weight

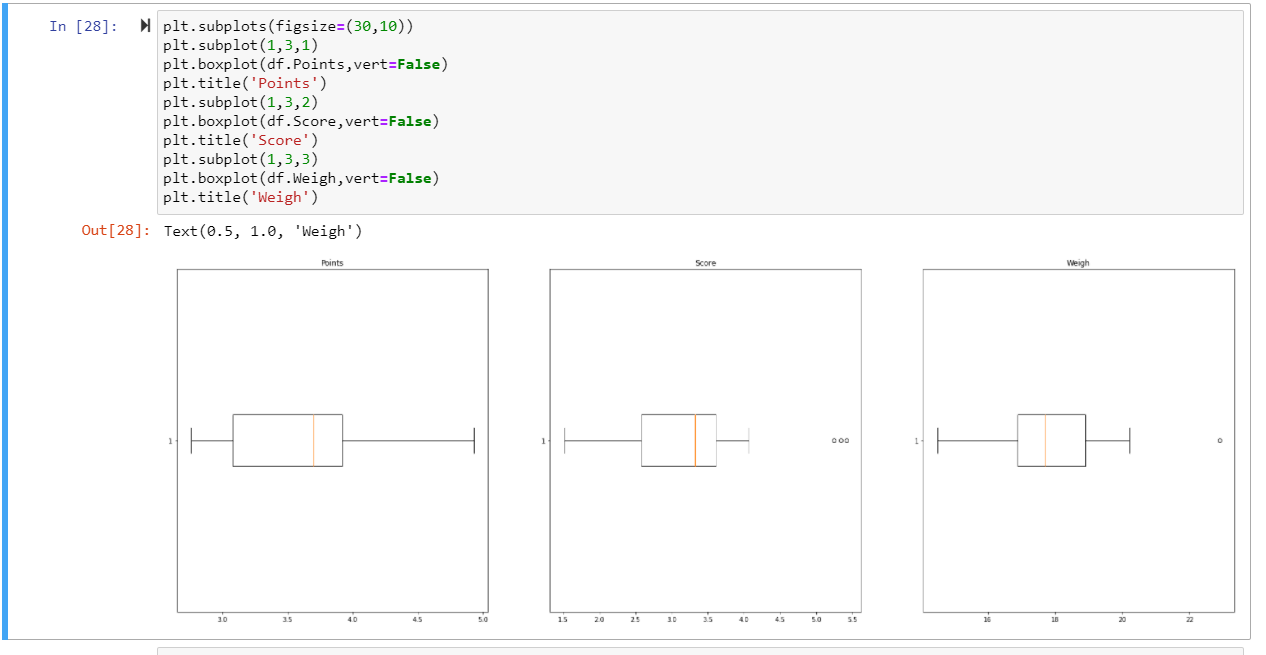
Find Mean, Median, Mode, Variance, Standard Deviation, and Range and also Comment about the values/ Draw some inferences.

**Use Q7.csv file**

**Solution:-**

****

****

****

Q8) Calculate Expected Value for the problem below

1. The weights (X) of patients at a 0063linic (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?

Solution:-

Expected value = Summation of the product of probability choosing each person individually and their weights

Probability of each patients choosing individually is = 1/9

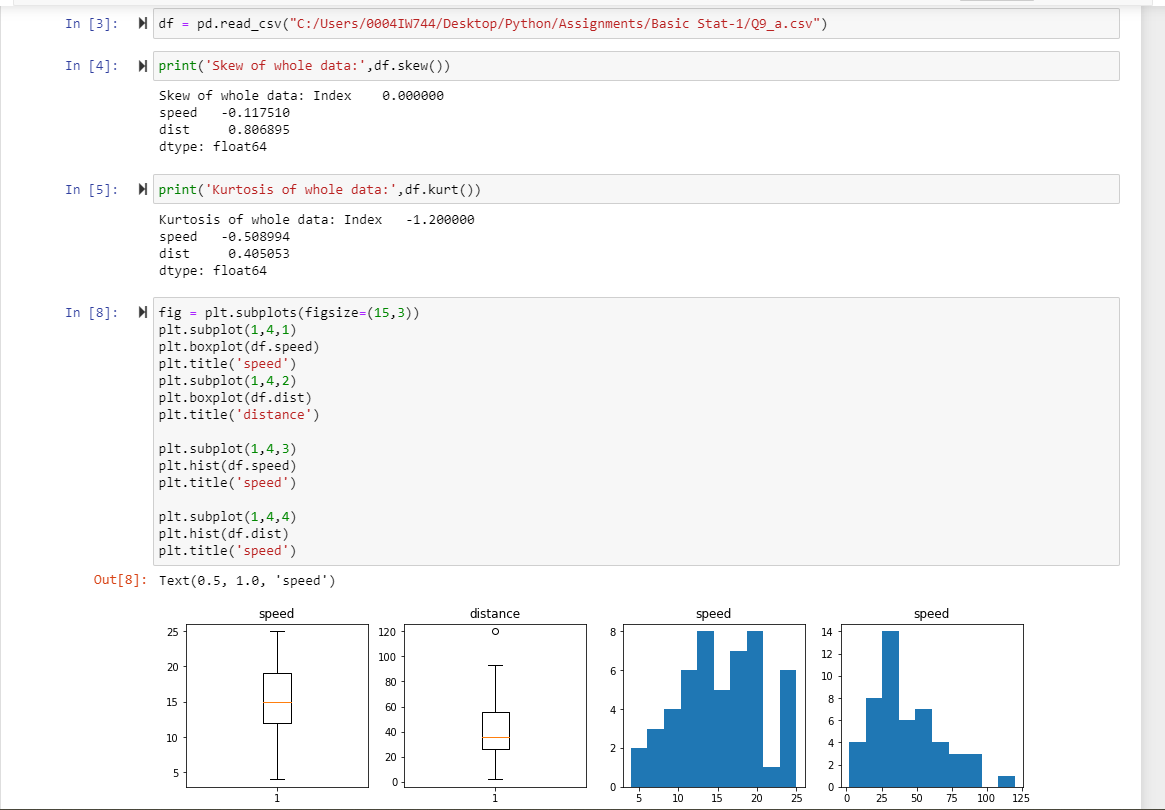
Expected Value = (108+110+ 123+134+ 135+ 145+ 167+ 187+ 199)\*1/9

=145.33

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

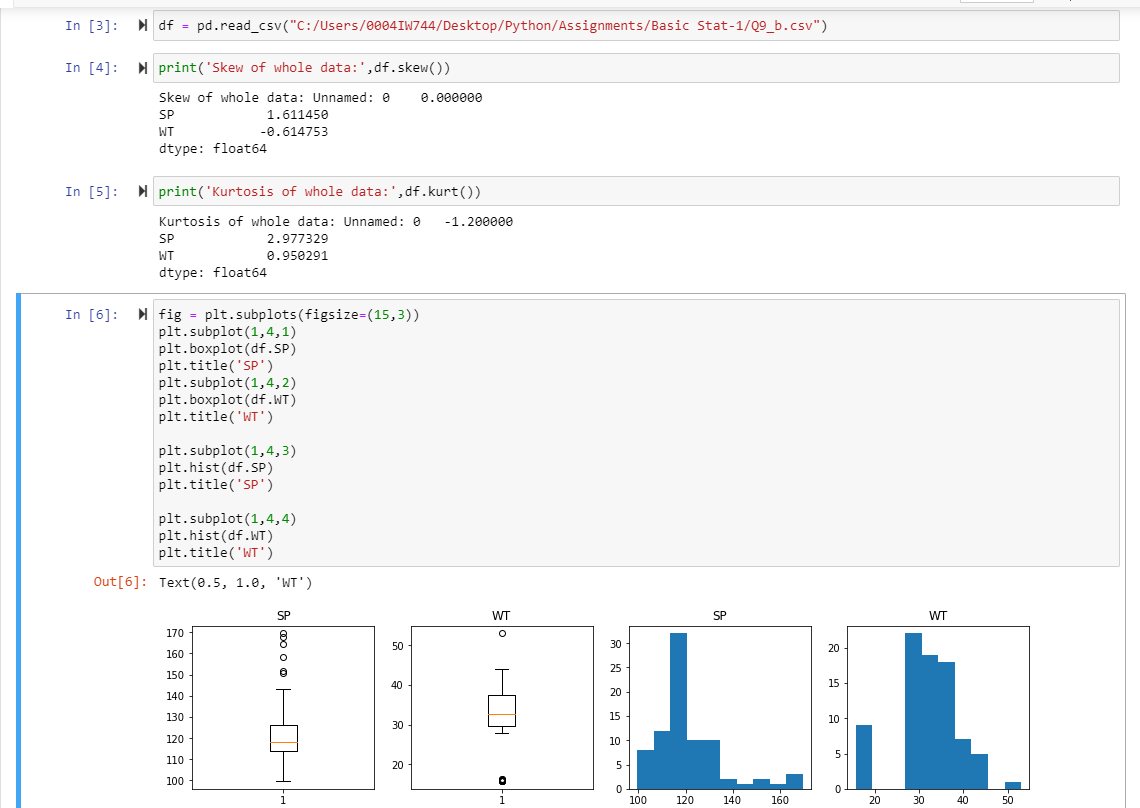
**Cars speed and distance**

**Use Q9\_a.csv**

****

**SP and Weight(WT)**

**Use Q9\_b.csv**

****

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



1) The data is skewed on the right side. So, data is positively skewed.

2) There are no outliers for the given data



1. Outliers exists
2. Data is distributed on the right
3. Positive skew

**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?

Sol:- 

**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**

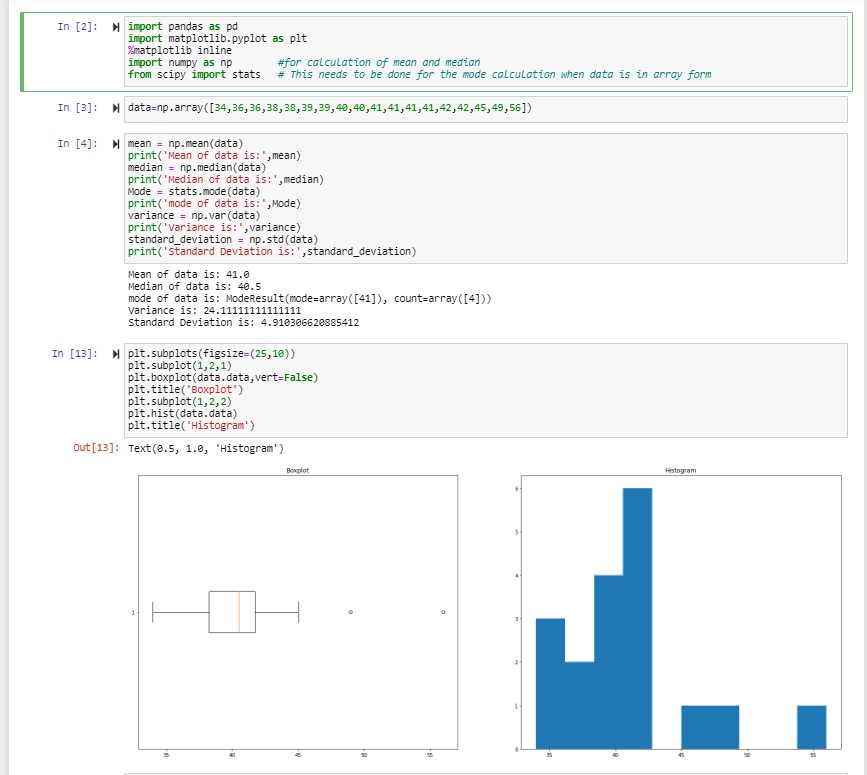
1. What can we say about the student marks?

Sol:- --

* Two outliers in the Student's marks: 49 and 56
* It is not following the Normal Distribution

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

* Mean of data is: 41.0
* Median of data is: 40.5
* mode of data is: ModeResult(mode=array([41]), count=array([4]))
* Variance is: 24.11111111111111
* Standard Deviation is: 4.910306620885412

Ans) 

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

Sol) **Symmetrical distribution**

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

Sol)  Right-skewed

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

Sol) Left-skewed

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

Sol) **Distribution is peaked and possess thick tails for given data**

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

Sol) **Distribution is not peaked and don’t have thick tails for given data**

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



What can we say about the distribution of the data?

Sol) No Outliers

q1= 18

Median= 15.2

Q3 = 10

What is nature of skewness of the data?

Sol)Negative Skewness

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

Sol) 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.

For the above data, we can say that

* Skewness is ‘0’
* Normal Distribution Exists for both
* There are no Outliers for both plots

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

Data \_set: Cars.csv

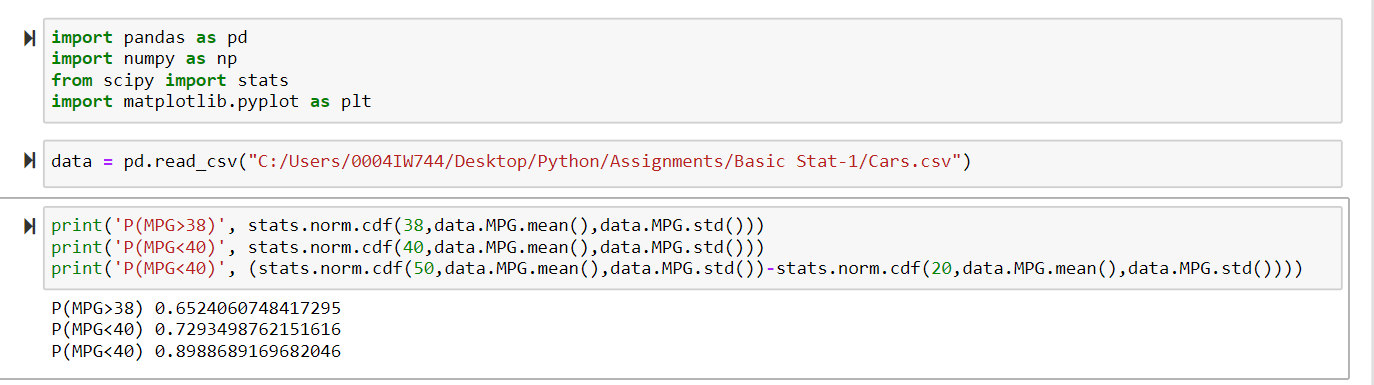
Calculate the probability of MPG of Cars for the below cases.

MPG <- Cars$MPG

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

c. P (20<MPG<50)

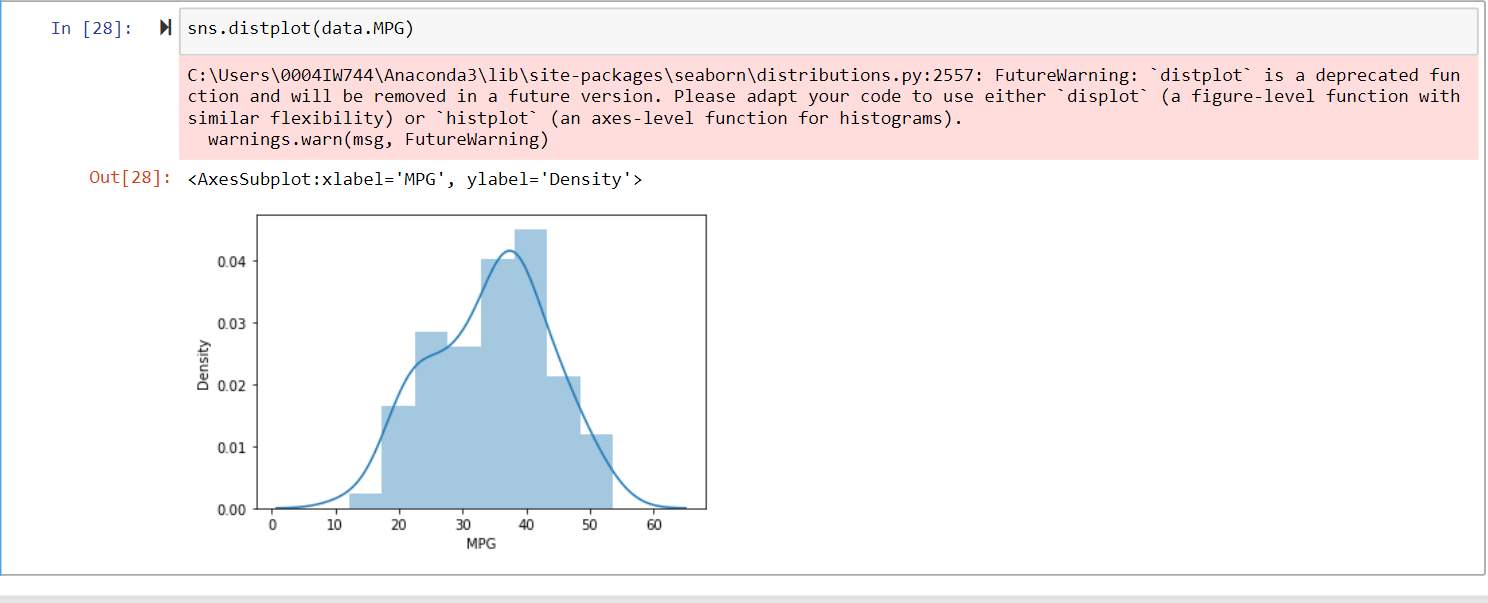
Sol:-



Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Sol:- BellShape – Yes(Approximately), Skewness – Negative

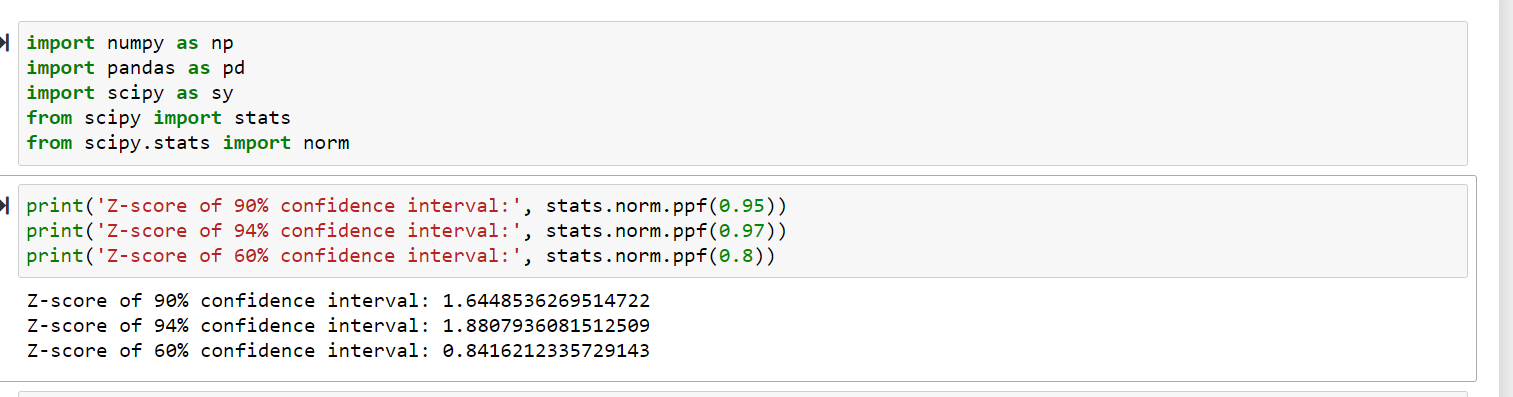
1. Check Whether the Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set follows Normal Distribution

Dataset: wc-at.csv

Sol:- 

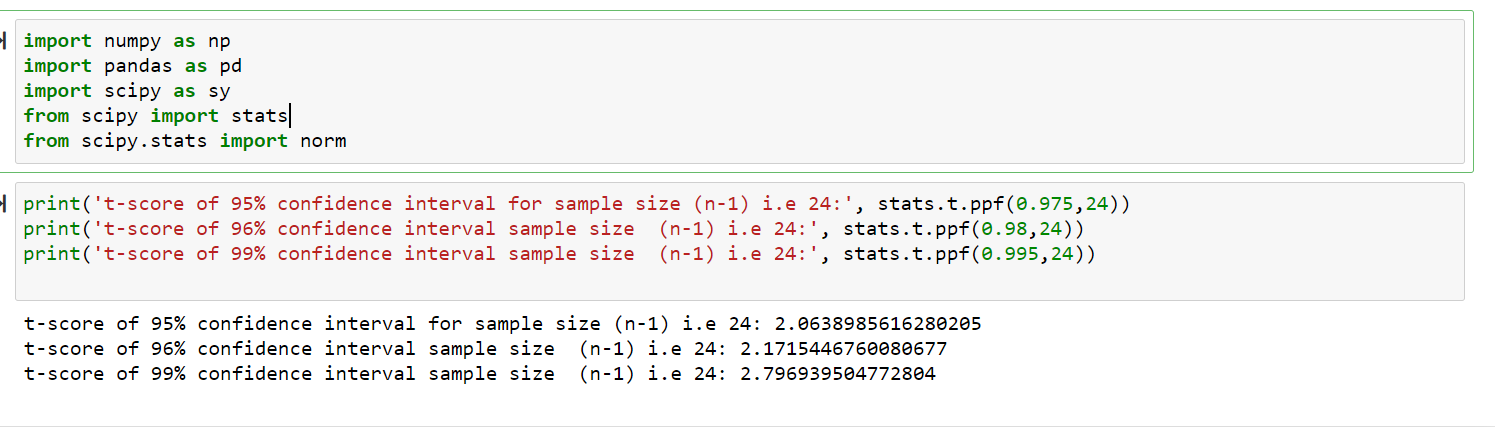
In Above:- Waist doesn’t following normal distribution and AT is following approximately

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

Sol:- 

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

Sol:- (df)= sample size – 1=24



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

