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

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

Ans - {HHH, HHT, HTH, THH, HTT, TTH, THT, TTT}

[P (HHT, THH, HTH)]

P (Two heads and One Tail) = Number of Desired Outcomes

(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

Ans - {(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)} = 36 = n(S)

1. Equal to 1 = 0
2. P(Less than or equal to 4) = P{(1,1) (1,2) (1,3) (2,1) (2,2) (3,1)} = 6 = n(A)

P(A) = n(A)/n(S) = 6/36 = 1/6

1. Sum is divisible by 2 and 3

(1,5) (3,3) (4,2) (5,1) (6,6)

Total no possible outcomes = 36

No of favorable outcomes = 5

P = favorable outcomes / total no possible outcomes = 5/36

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 no of balls (2+3+2) = 7

S be the Sample Space

n(S) = 2 balls are drawn at random out of 7

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

E = event of drawing 2 balls from which none of the balls drawn is blue,

n(E) = No of ways of ball drawing from (2+3) , none is blue

**nCr**= 5C2 = (5 \* 4\* 3) / (1\* 2\* 3) = 10

p(E) = n(E) / n(S) = 10 /2

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

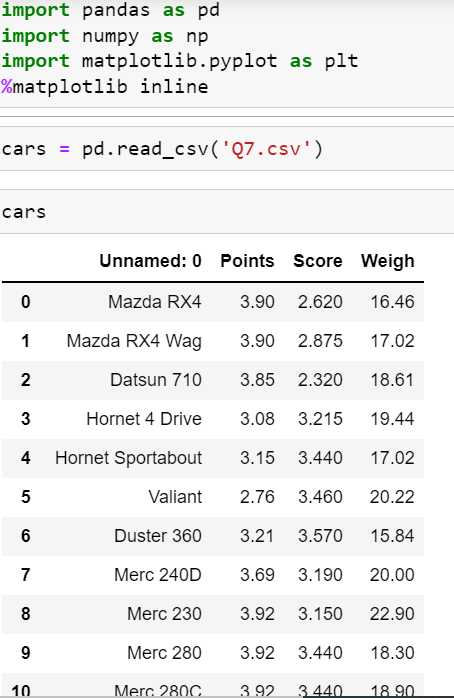
Expected number of candies for a randomly selected child = 3.09

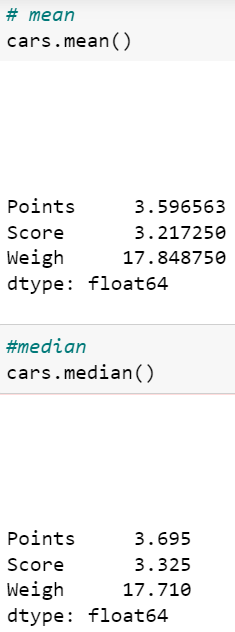
Q7) Calculate Mean, Median, Mode, Variance, Standard Deviation, Range & comment about the values / draw inferences, for the given dataset

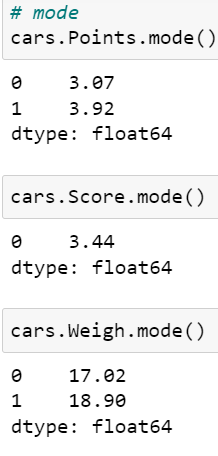
* For Points, Score, Weigh>

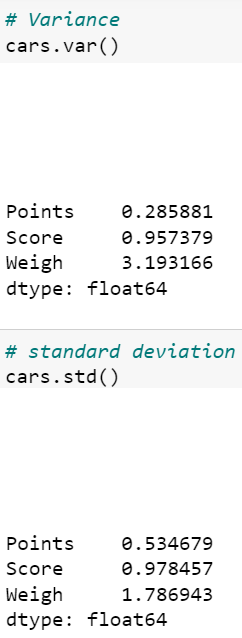
Find Mean, Median, Mode, Variance, Standard Deviation, and Range and also Comment about the values/ Draw some inferences.

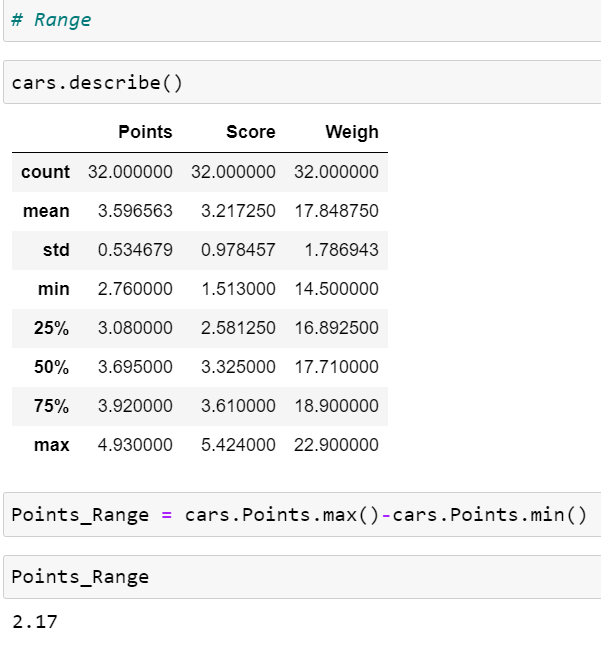
**Use Q7.csv file**

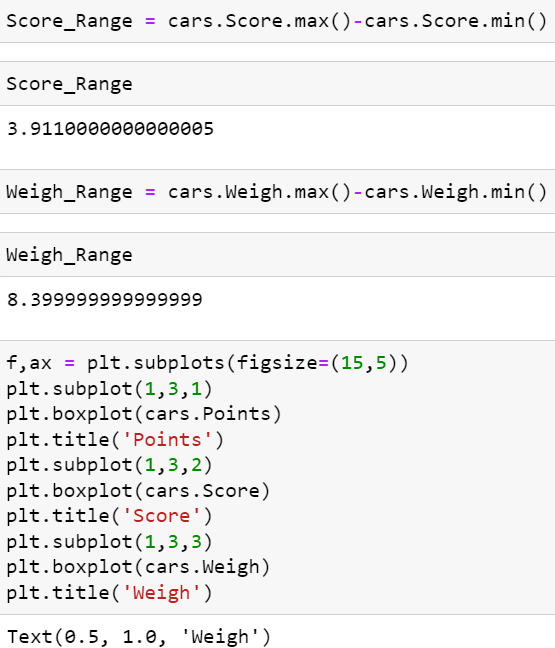
****

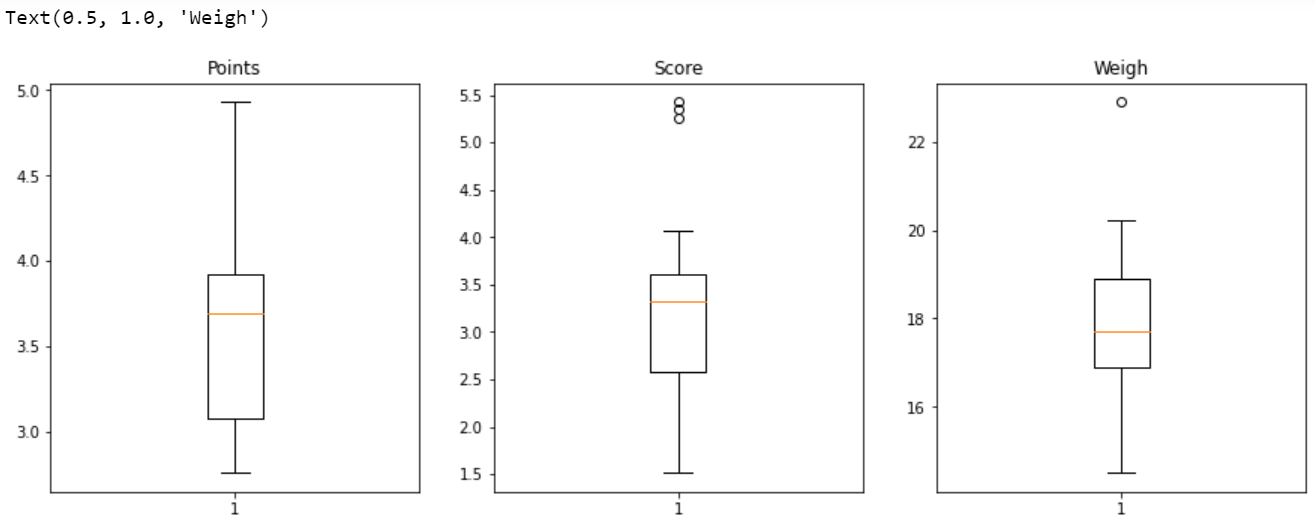
****

****

****

****

****

****

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 - The weights (X) of patients at a clinic (in pounds), are

108, 110, 123, 134, 135, 145, 167, 187, 199

one of the patients is chosen at random

**To Find –** Expected Value

**Solution -** Expected Value = (probability \* value)

There are 9 patients,

Probability of selecting each patient = 1/ 9

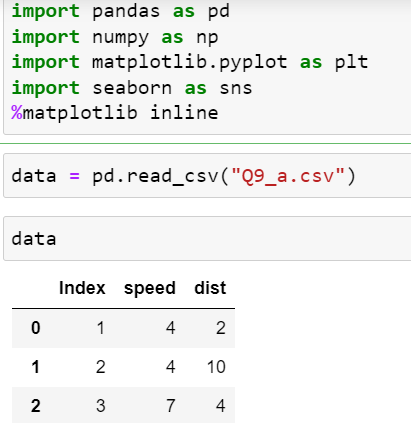
**Expected Value** = (1/9) \*(108) + (1/9) \*(110) + (1/9) \*(123) + (1/9) \*(134) + (1/9) \*(135) + (1/9) \*(145)  + (1/9) \*(167) + (1/9) \*(187) + (1/9) \*(199) = 145.33

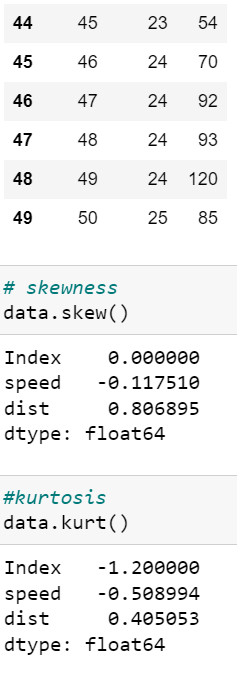
Expected Value of the Weight of that patient = 145.33

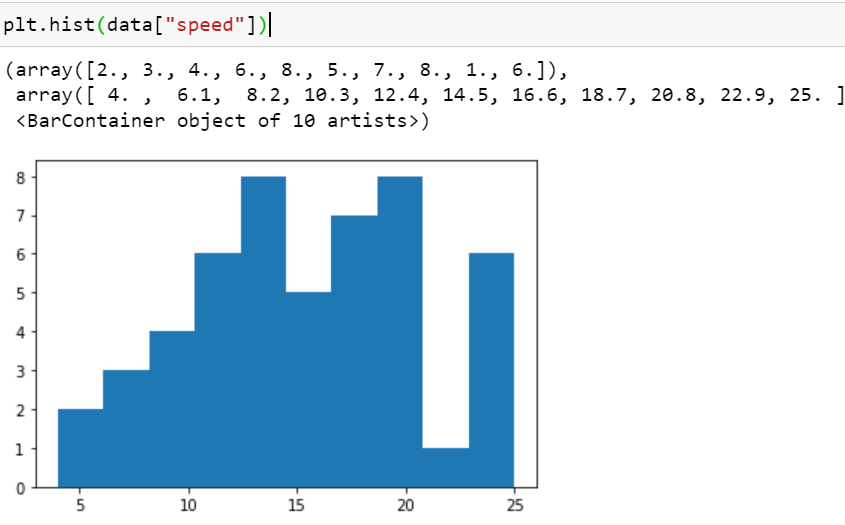
**Q9)Calculate Skewness, Kurtosis & draw inferences on the following data**

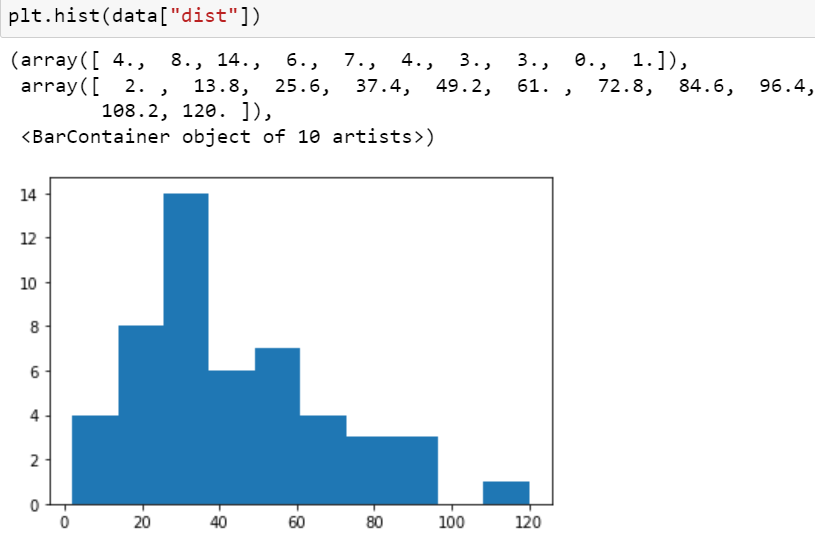
**Cars speed and distance**

**Use Q9\_a.csv**

****

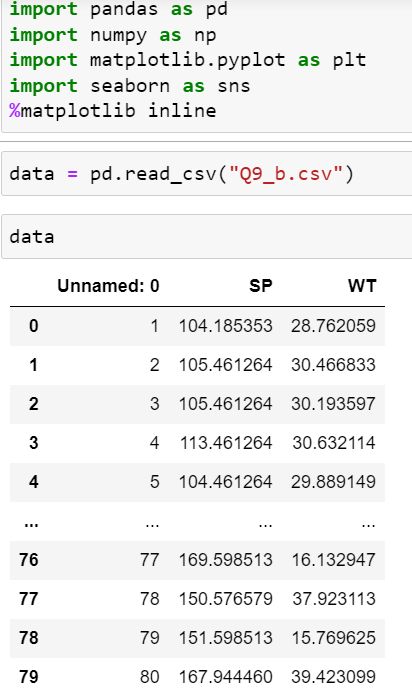
****

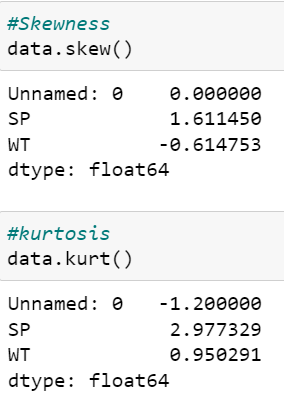
****

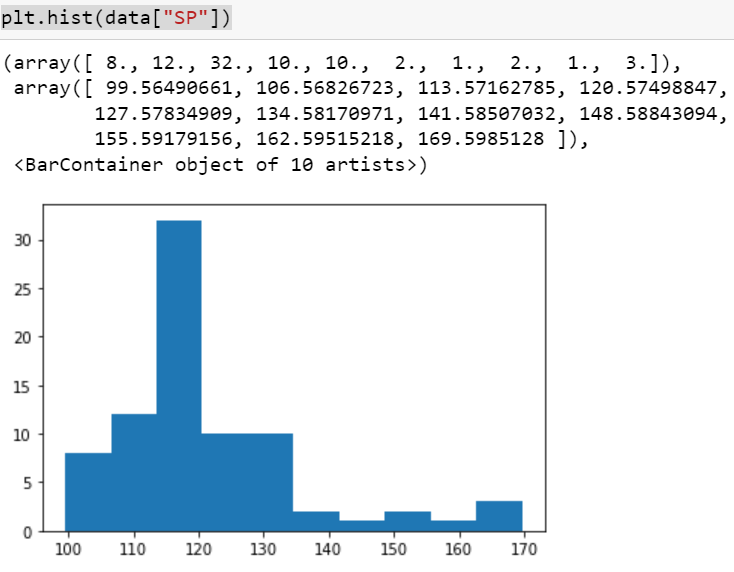
****

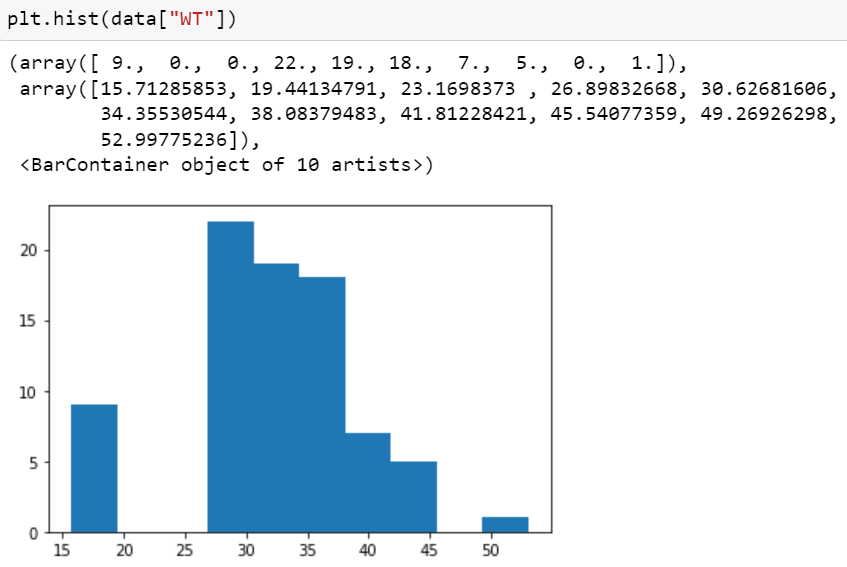
**SP and Weight(WT)**

**Use Q9\_b.csv**

****

****

****

****

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



**Histogram –**

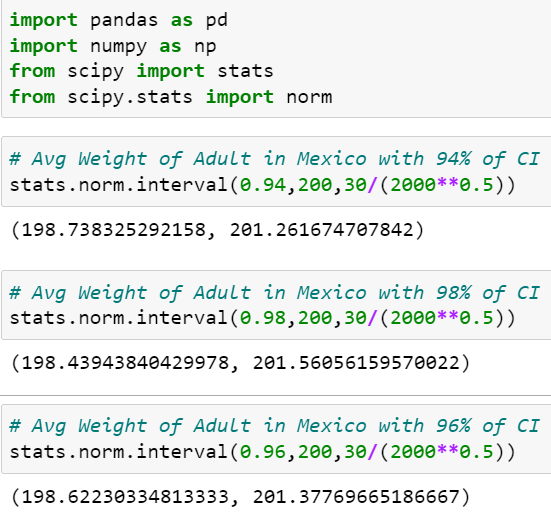
**Ans –** Right side skewed or Positively skewed



**Boxplot –**

1. The data is right skewed
2. There are outliers at the upper 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?

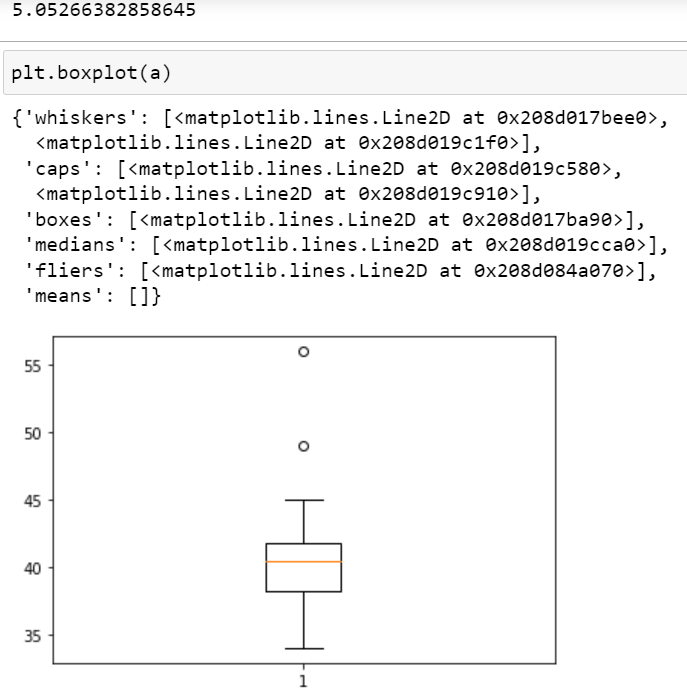


**Q12)** Below are the scores obtained by a student in tests

**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 – Normalized Skewnness

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

Ans – positively skewed.

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

Ans - negatively skewed.

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

Ans – peaked and possesses thick tails.

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

Ans - flat and has thin tails.

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



What can we say about the distribution of the data?

Ans - The distribution of data in De-assigned format.

What is nature of skewness of the data?

Ans - Left Skewed or Negatively Skewed

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

Ans - Q3 –Q1 = 18-10 = 8 is IQR of the data

Q19) Comment on the below Boxplot visualizations?



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

Ans -

a) The Boxplot designed with range = 3

b) The Second Boxplot with range = 1.5

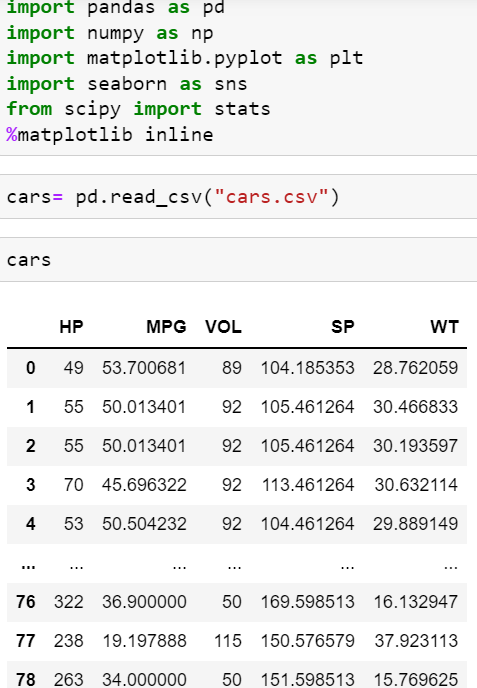
Q 20) Calculate probability from the given dataset for the below cases

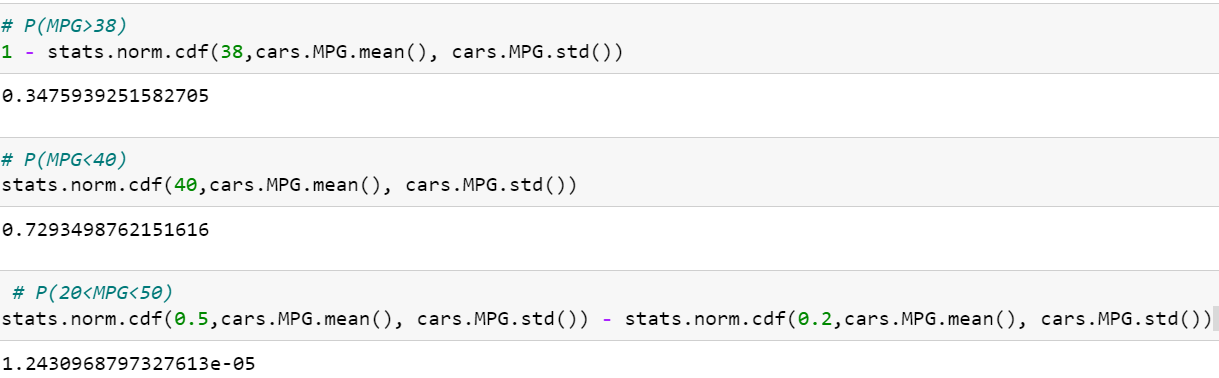
Data \_set: Cars.csv

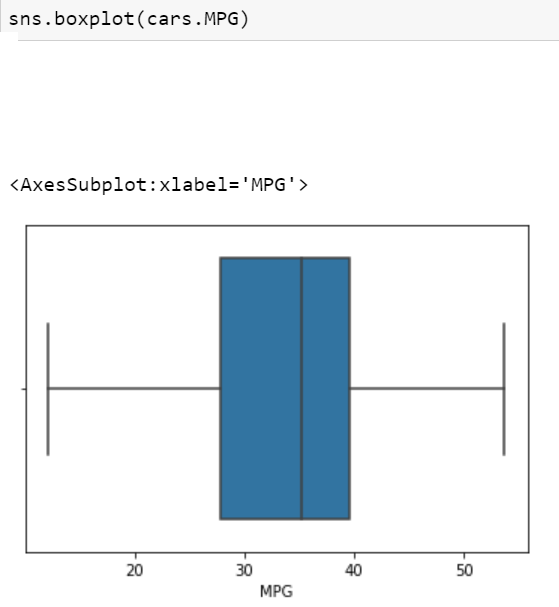
Calculate the probability of MPG of Cars for the below cases.

MPG <- Cars$MPG

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



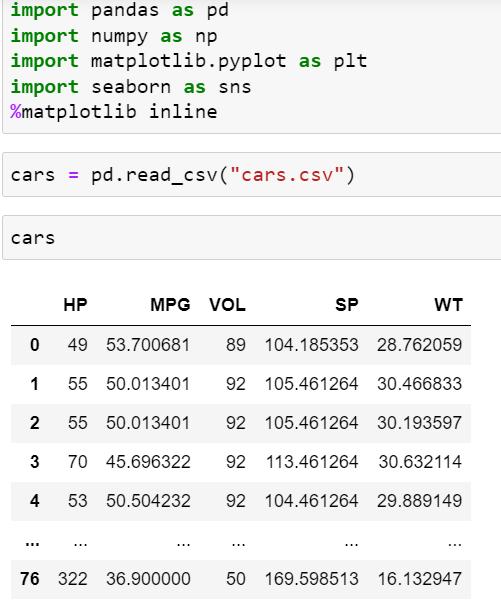


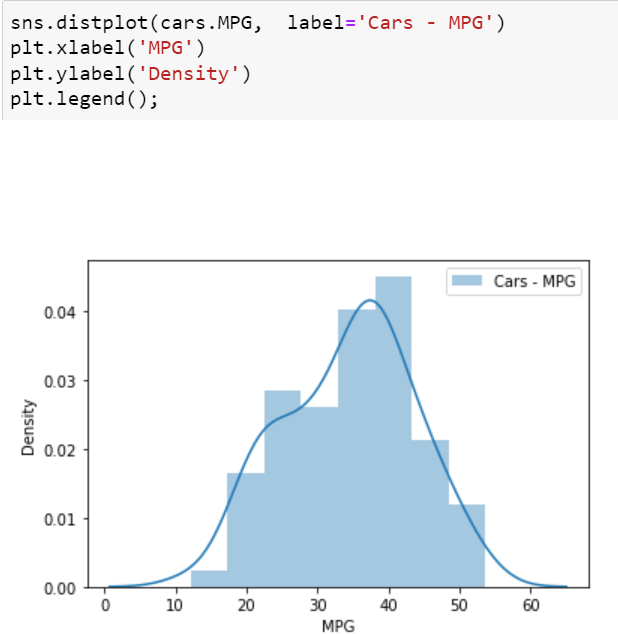


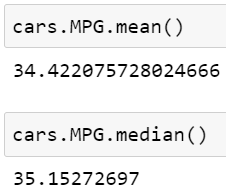
Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

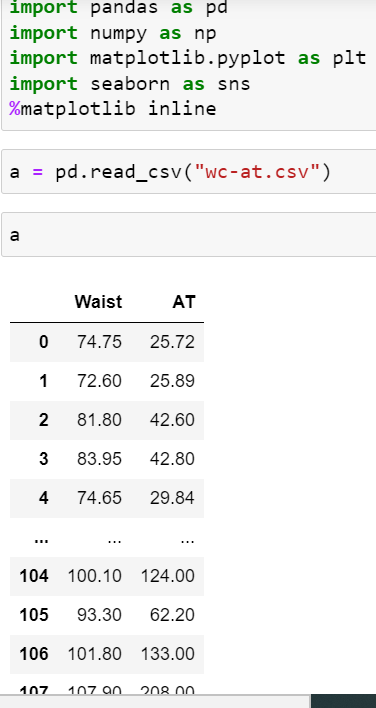


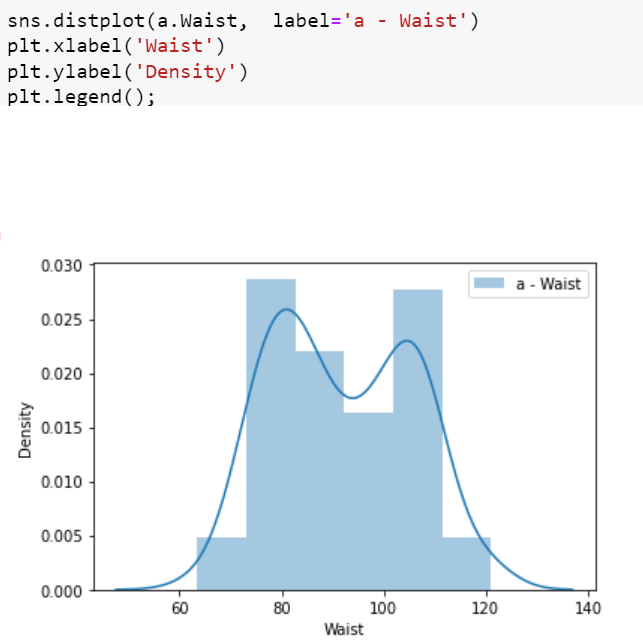


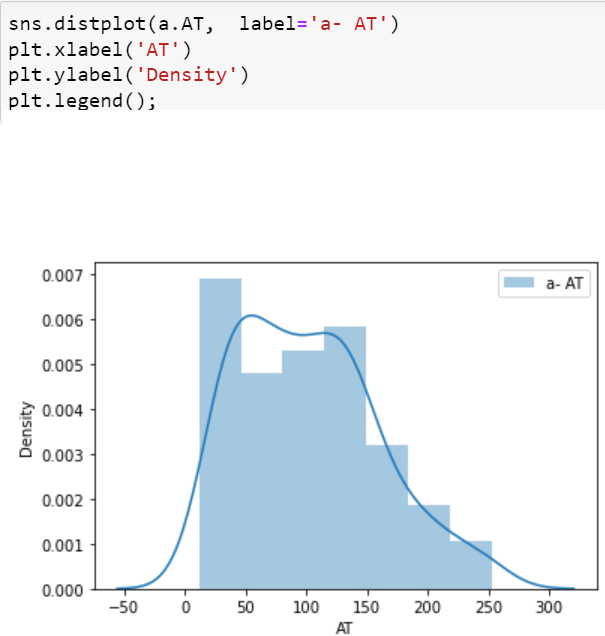


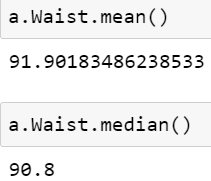
1. Check Whether the Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set follows Normal Distribution

Dataset: wc-at.csv

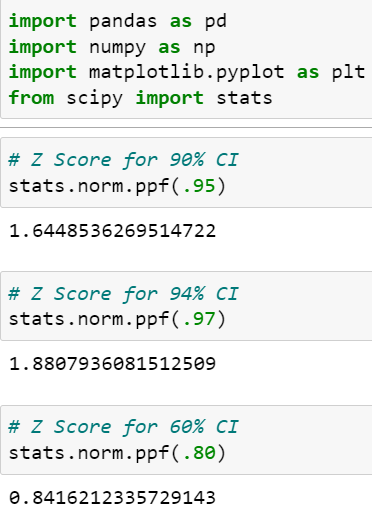




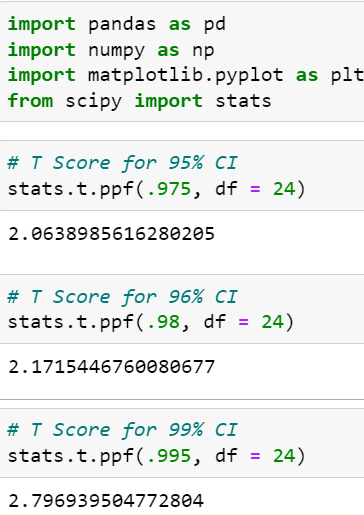




Q 22) Calculate the Z scores of 90% confidence interval,94% confidence interval, 60% confidence interval



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

