|  |  |
| --- | --- |
| 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-nominal |
| Religious Preference | Discrete-Ratio |
| Barometer Pressure | Discrete-Interval |
| SAT Scores | Discrete-interval |
| Years of Education | Discrete-ratio |

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

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

X (HHT) = 3

No. of outcomes (X) = 8

P(X=HHT) =X (HHT) / X

= 3/8

P(X=HHT) =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

Event occurred = { (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) (3,1) (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)

No. of outcomes (X) = 36

1. Equal to 1

X=36

Event of getting sum is equal to 1 X (E) = 0

P =X(E) / X

P=0 / 36

= 0

1. Less than or equal to 4

X = 36

Event occurred 6 times

X(E) = 6

P = X(E) / X

P =6/36

= 1/6

1. Sum is divisible by 2 & 3

X = 36

Event occurred 5 times

X(E) = 5

P=X(E)/X

=5/36

=0.138

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

X = way of drawing 2 balls out of 7 = 21

X(E)=Event of drawing 2 balls none of which is blue

2\*5 = 10

P = 10/21

= 0.467

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 no 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

=0.015+0.8+1.95+0.025+0.06+0.24

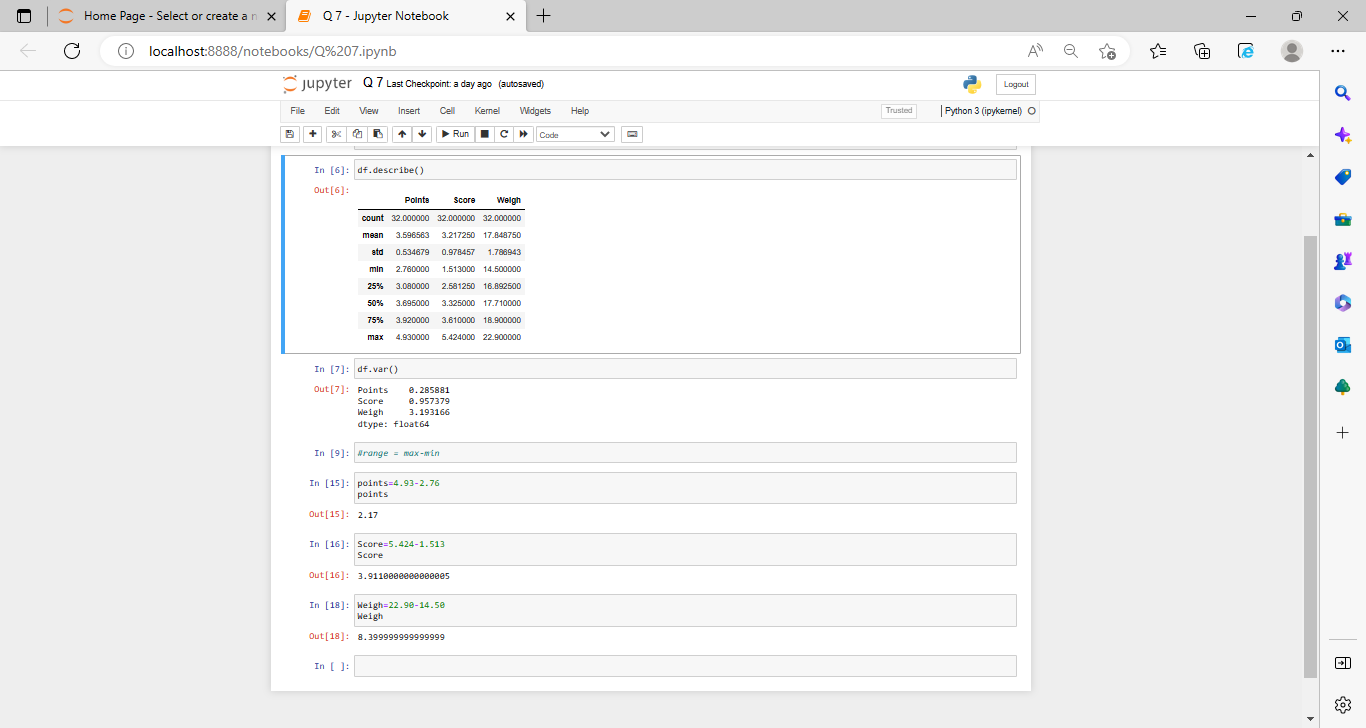
=3.080

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?

Expected value= ∑ ( probability  \* Value )

∑ P(x).E(x)

There are 9 patients.

Probability of selecting each patients=1/9

E(X) =108, 110, 123, 134, 135, 145, 167, 187, 199

P(X) =1/9,1/9,1/9,1/9,1/9,1/9,1/9,1/9,1/9

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

=1/9(1308)

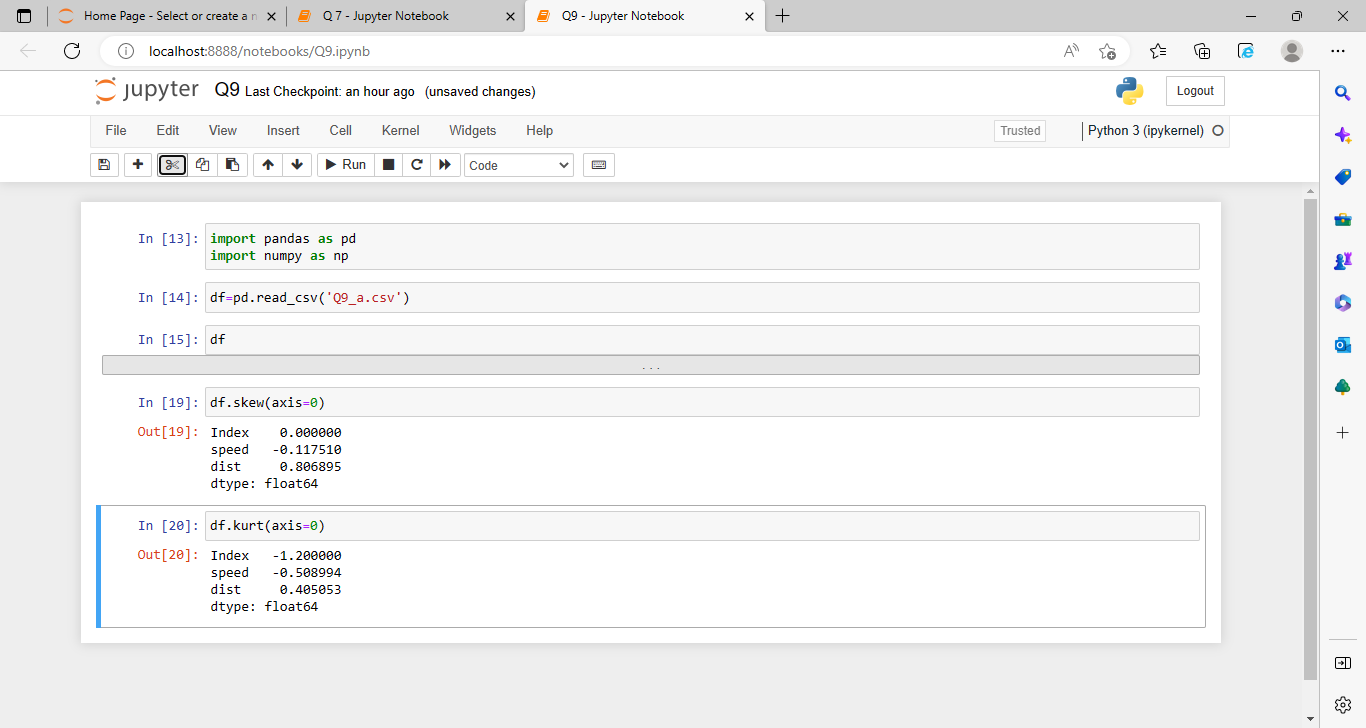
=145.33

Expected value of the weight of that person is 145.33.

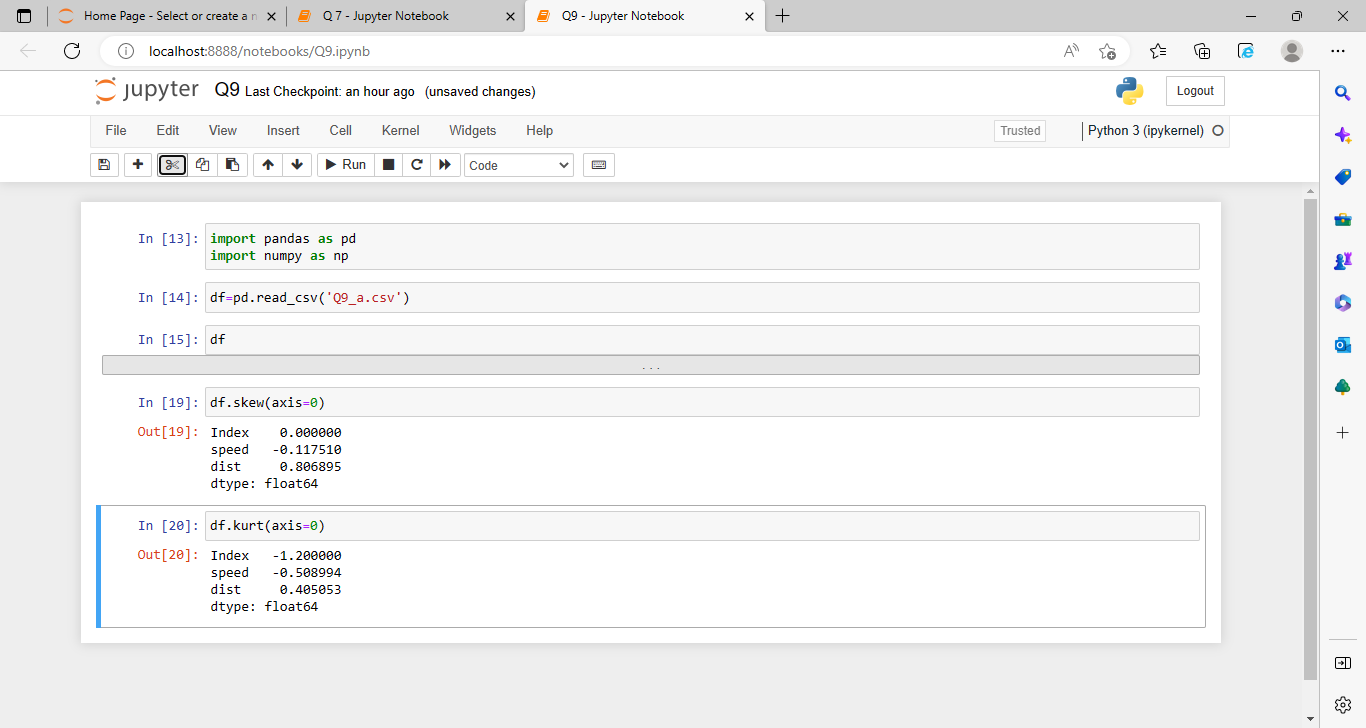
**Q9) Calculate Skewness, Kurtosis & draw inferences on the following data**

**Cars speed and distance**

**Use Q9\_a.csv**

****

**SP and Weight(WT)**

**Use Q9\_b.csv**

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

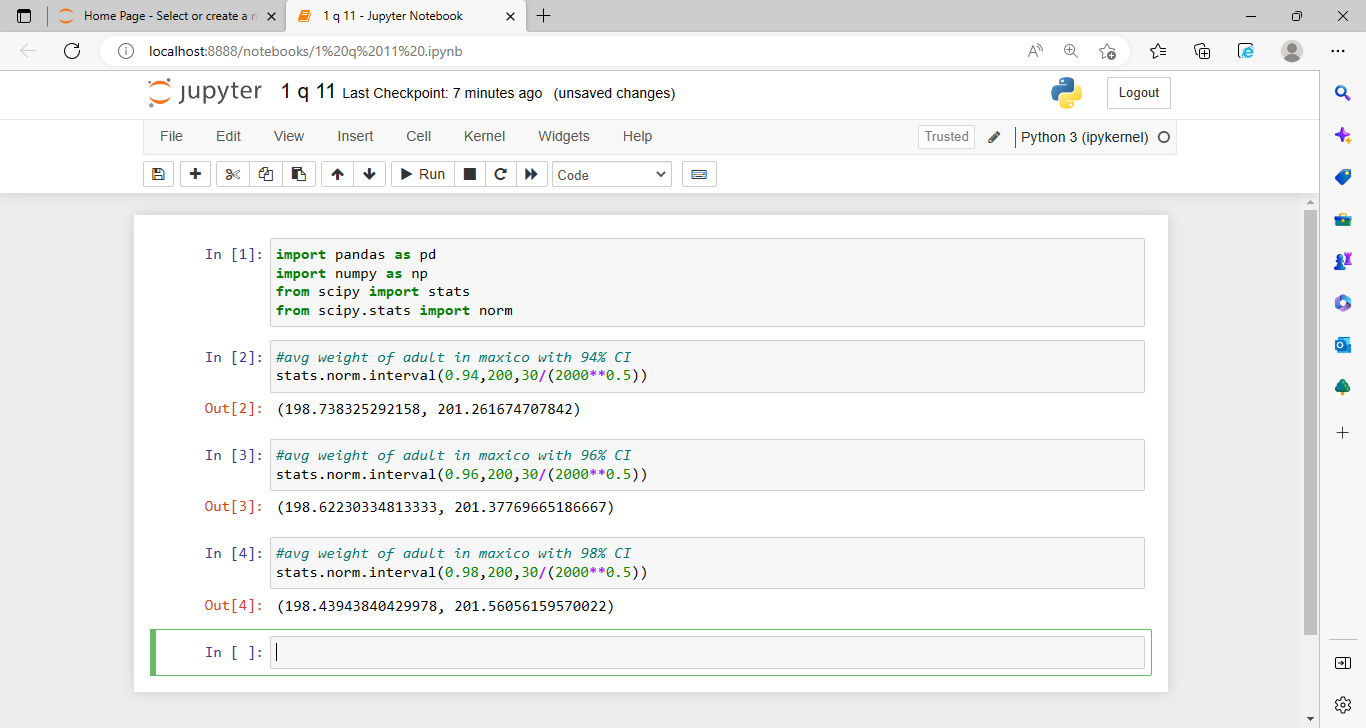


There is a right(+ve) skewness in histogram. That means there mean value >median value.



There is outlier at the upper extream

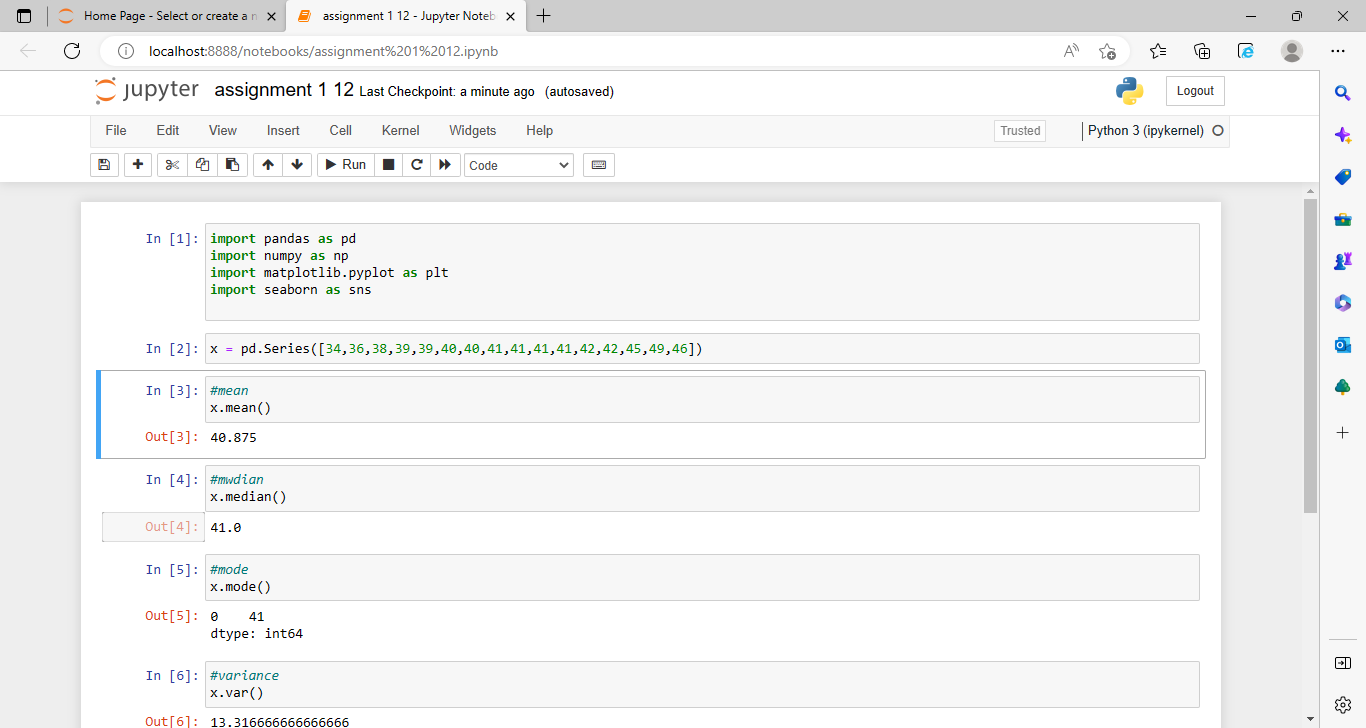
**Q11)** Suppose we want to estimate the average weight of an adult male in Mexico. We draw a random sample of 2,000 men from a population of 3,000,000 men and weigh them. We find that the average person in our sample weighs 200 pounds, and the standard deviation of the sample is 30 pounds. Calculate 94%,98%,96% confidence interval?

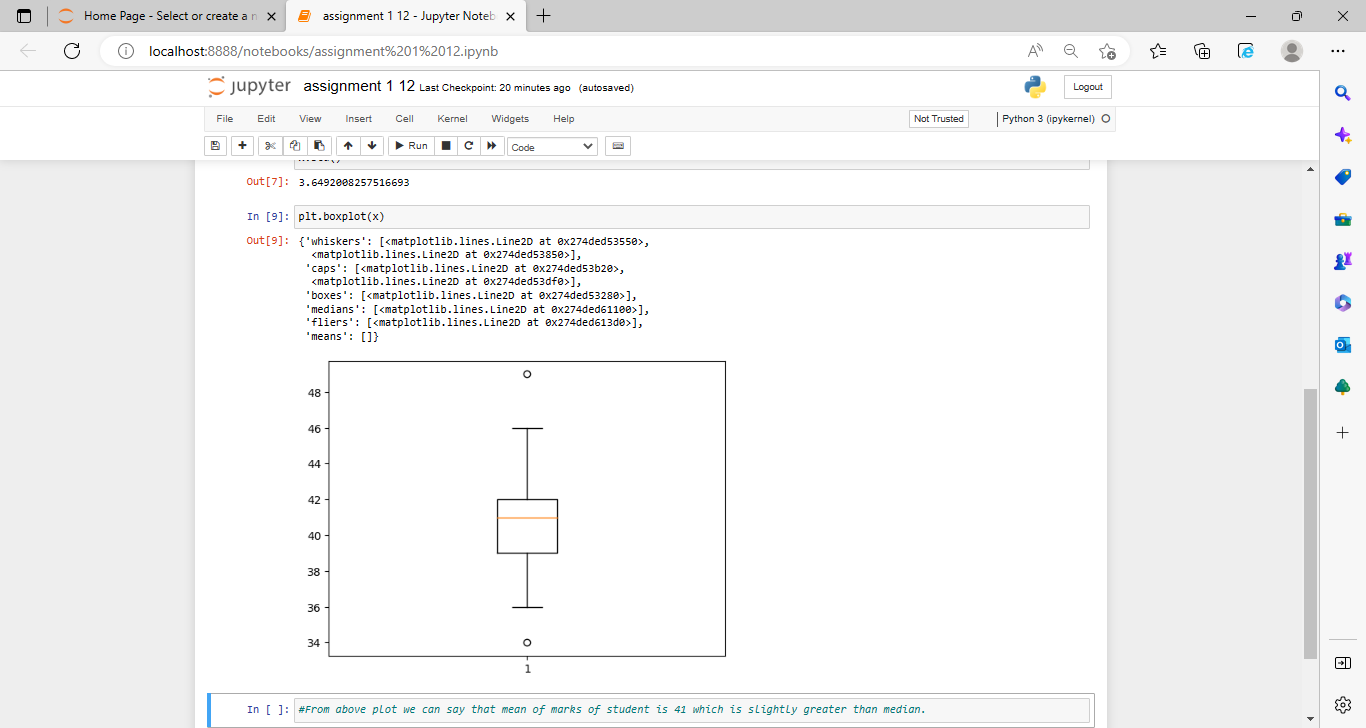


**Q12)** Below are the scores obtained by a student in tests

**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.

What can we say about the student mark?



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

Ans. Data are summetrically distributed with zero skewness that’s mean normal distribution.

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

Ans. The nature of skewness when mean >median is positive or right skewness means long tail at the right side of graph .

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

Ans. The nature of skewness when median > mean is left or negative skewness means long tail at the left side of graph.

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

Ans. Positive kurtosis value indicates a heavy tailed distribution of data.

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

Ans. Negative kurtosis value indicates a light tailed distribution of data.

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



What can we say about the distribution of the data?

Ans. The distribution of the data is negatively skewed (left skewness)

What is nature of skewness of the data?

Ans. The nature of skewness of the data is negative (left) skew.

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

Ans. The IQR of the data between 10-18 approximately.

Q19) Comment on the below Boxplot visualizations?



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

Ans. The distribution of the data for boxplot 1 and boxplot 2 symmetrical means normal distributed. And the median of both the boxplot is same but the ranges is different.

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

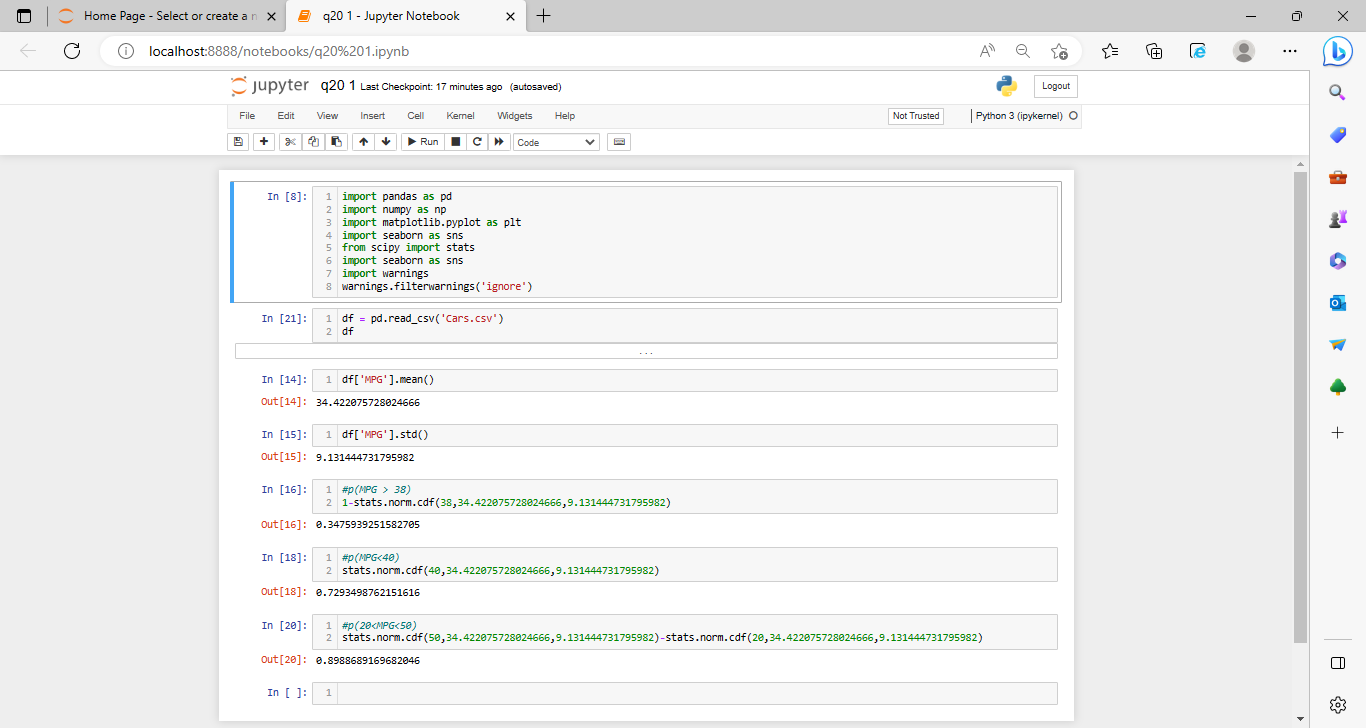
Data \_set: Cars.csv

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

MPG <- Cars$MPG

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

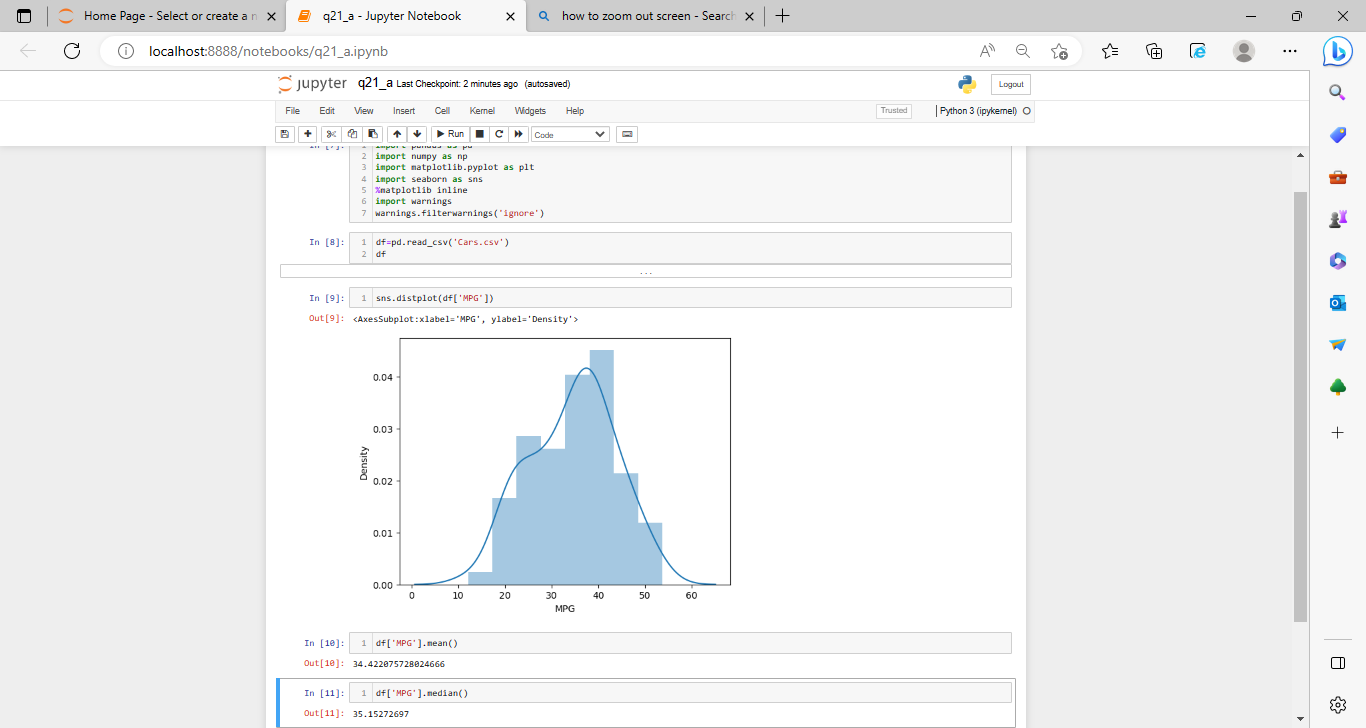
c. P (20<MPG<50)



Q 21) Check whether the data follows normal distribution

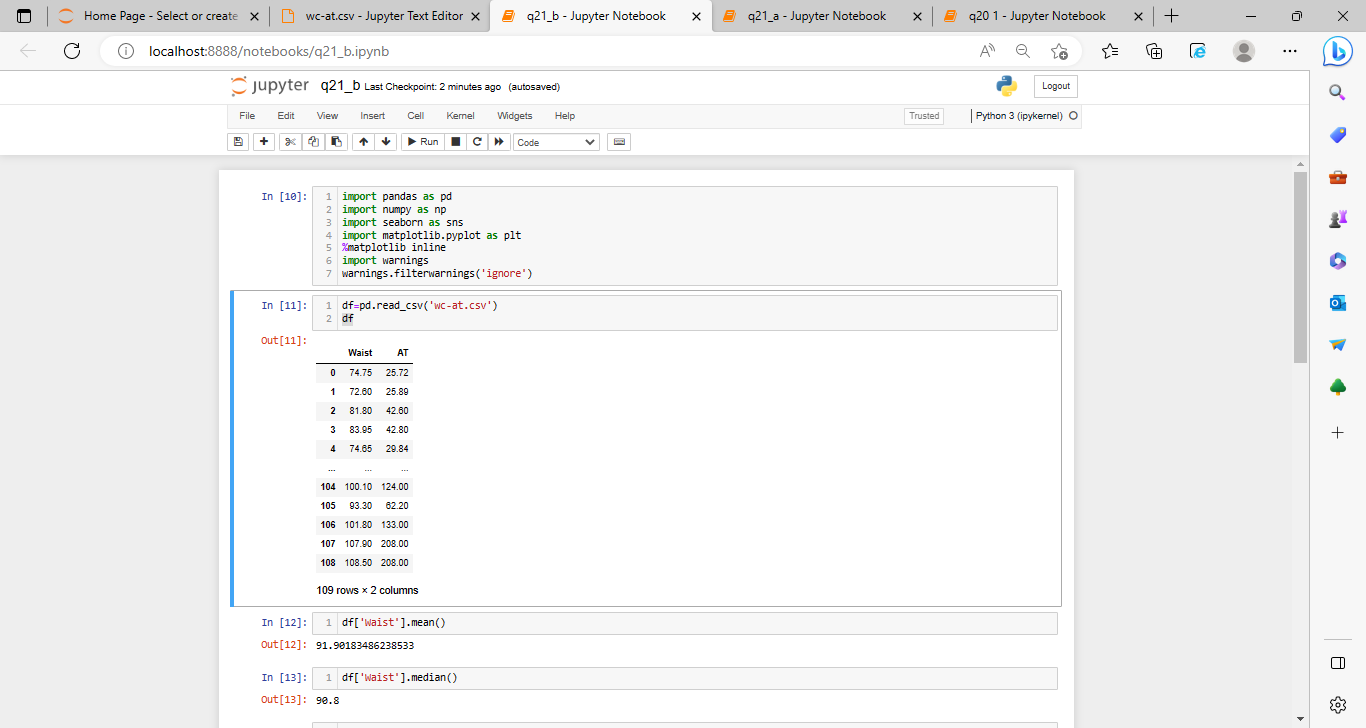
1. Check whether the MPG of Cars follows Normal Distribution

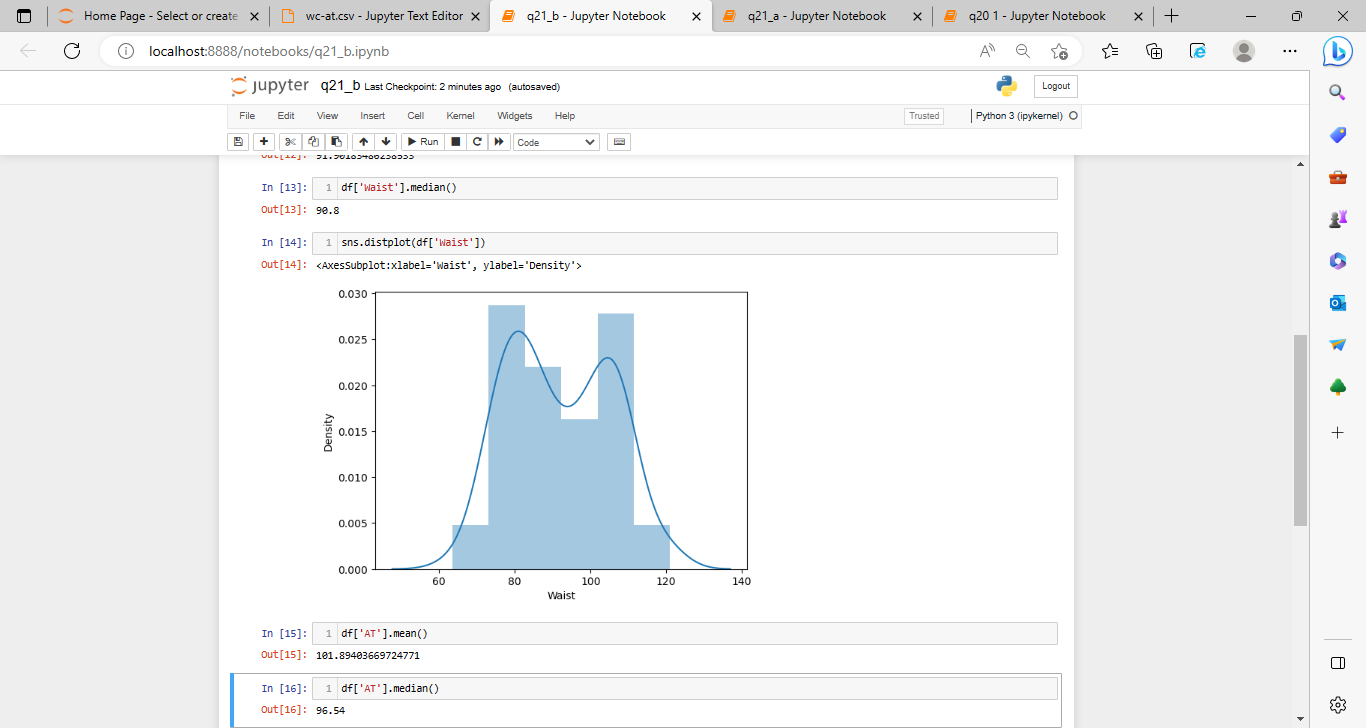
Dataset: Cars.csv

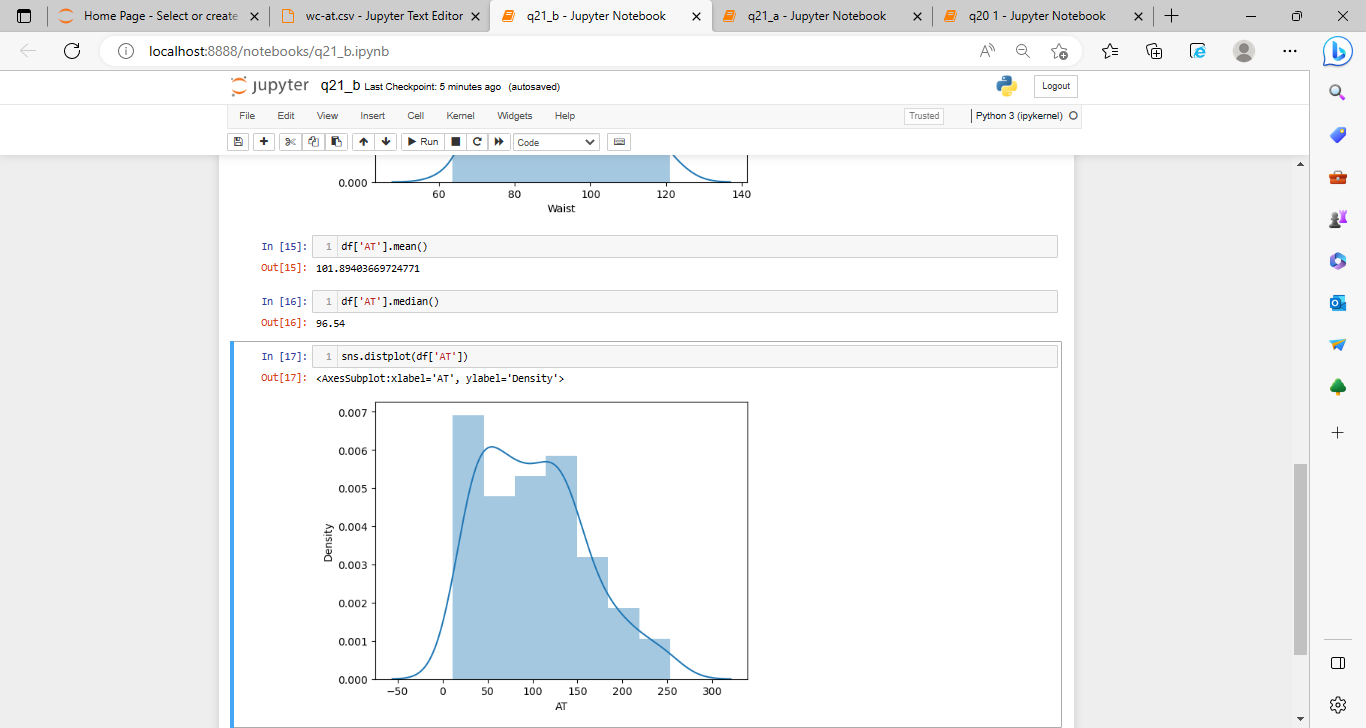


1. Check Whether the Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set follows Normal Distribution

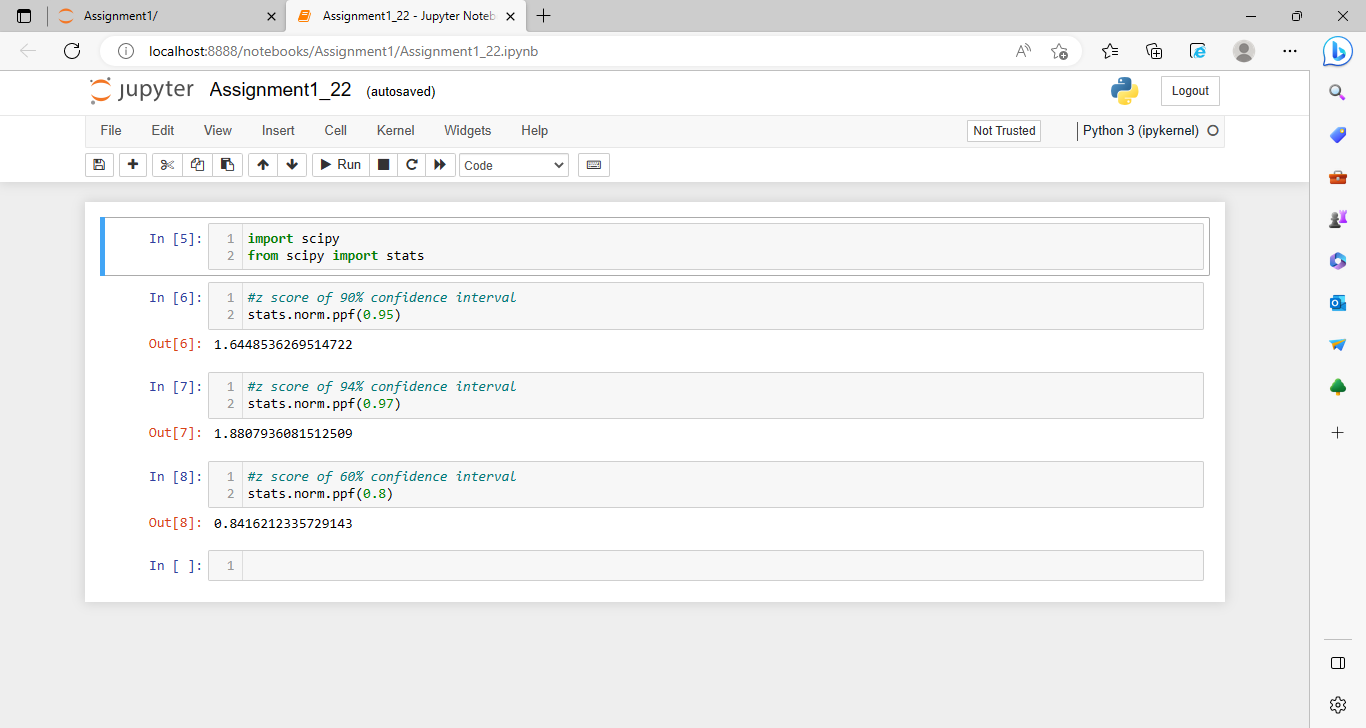
Dataset: wc-at.csv



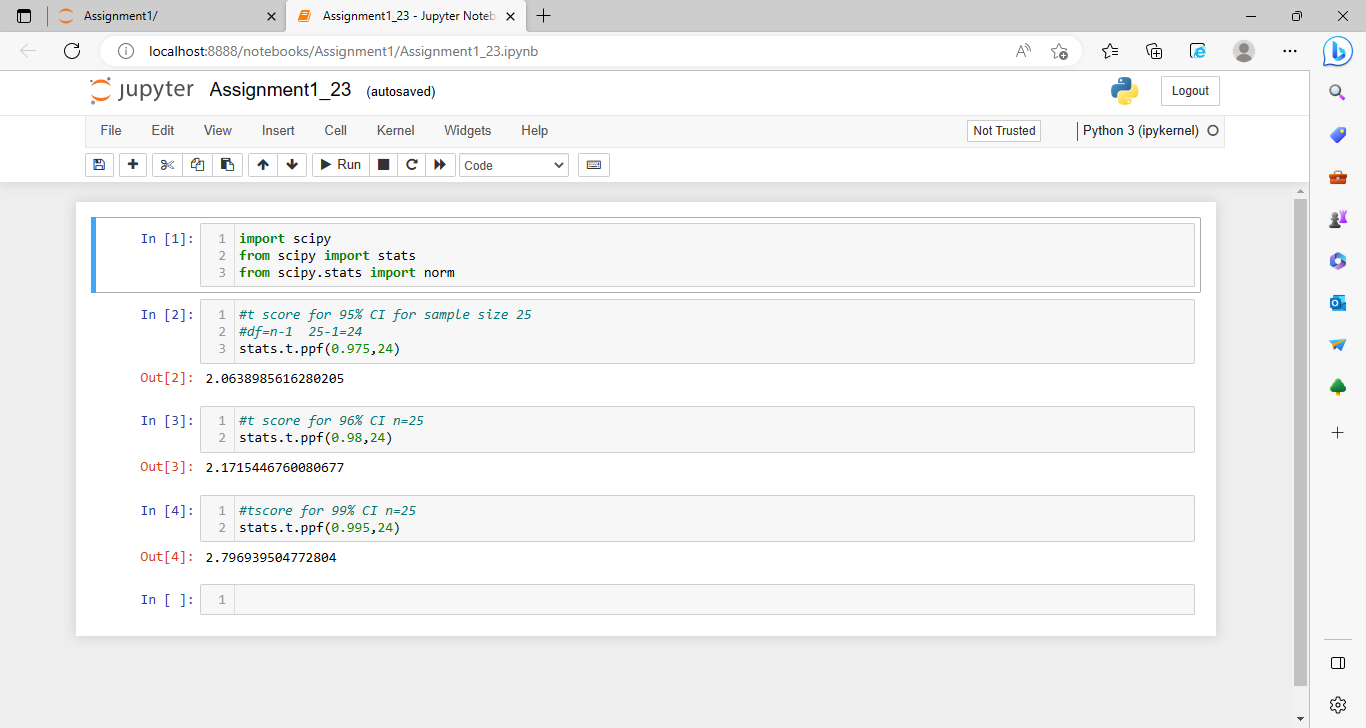




Q 22) Calculate the Z scores of 90% confidence interval,94% confidence interval, 60% confidence interval



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

