|  |  |
| --- | --- |
| 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 | Ordinal |
| Fahrenheit Temperature | Interval |
| 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 | Ratio |
| 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?

Answer:

Sample space(S)=(TTT),(HHH),(HTT),(THT),(TTH),(THH),(HTH),(HHT)

Total no.of outcomes=8

Favourable outcomes=3

Probability=3/8

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

Answer:

Sample Space(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)

Total no.of outcomes=36

1. Favorable outcomes=0

Probability=0/36=0

Sum less than or equal to 4=(1,1),(1,2),(1,3),(2,1),(2,2),(1,3)

b)Favorable outcomes=6

probability=6/36=1/6

c)sum divisible by 2 and 3 = (1,5),(2,4),(3,3),(4,2),(5,1),(6,6)

Favourable outcomes=6

Probability=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?

Answer:

Total balls 2R+3G+2B=7

Total no.of outcomes=7C2=7!/(7-2)!(2!)

=21

Faourable outcomes=5C2=5!/(5-2)!(2!)

=10

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

Answer:

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

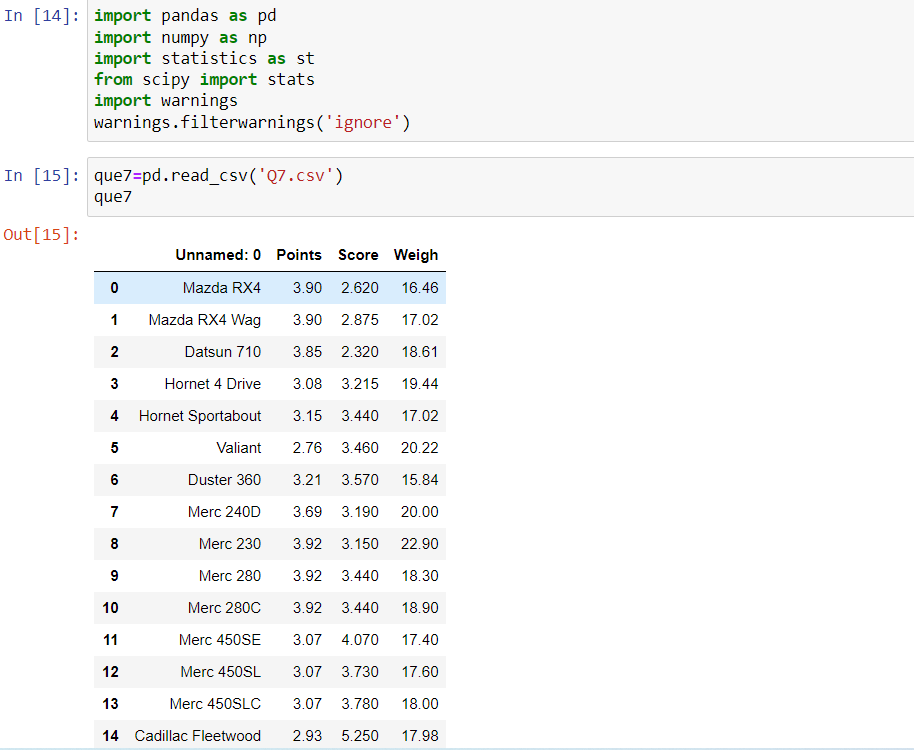
=0.015+0.8+1.95+0.025+0.06+0.24

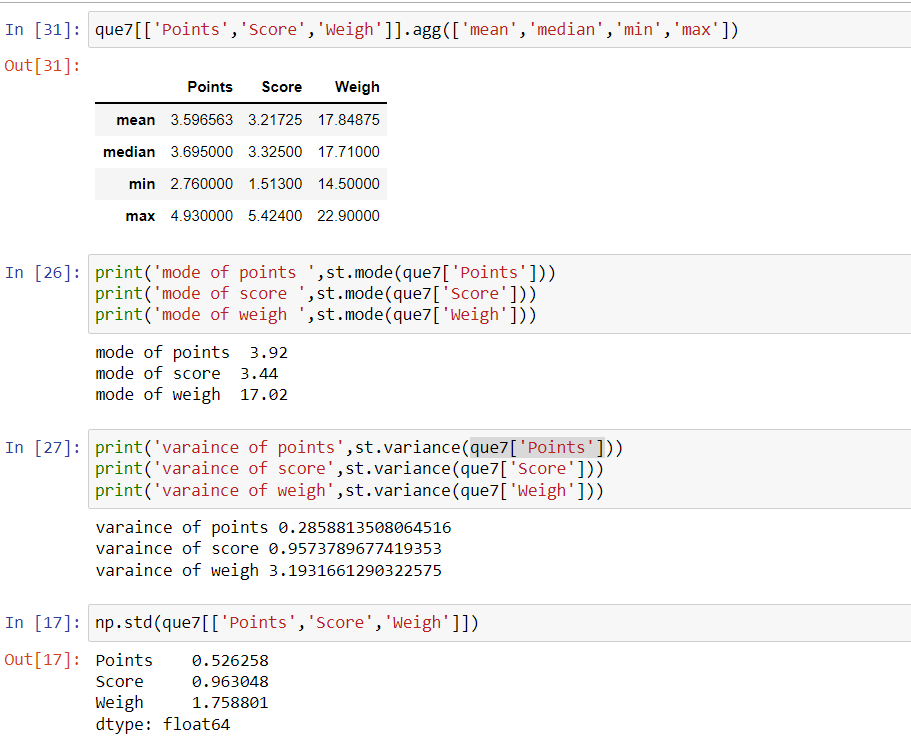
=3.090

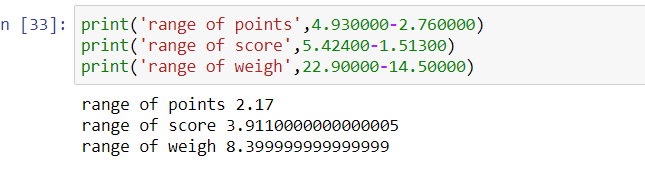
Q7) Calculate Mean, Median, Mode, Variance, Standard Deviation, Range & comment about the values / draw inferences, for the given dataset

* For Points,Score,Weight>

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







**We calculated that mean, median, mode,standard deviation ,variance and range of required columns we saw that variance for weigh is very high which is not good because if the variance is high there is more deviation from central value and even points and score have high variance**

**We can say that more frequently occurring values are 3.92,3.44.17.02 for points,score and weigh respectively**

**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?

Answer:

Expected Value  =  sum of ( probability  \* Value )

there are 9 patients

Probability of selecting each patient = 1/9

Ex  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) + (1/9)110  + (1/9)123 +

(1/9)134 + (1/9)135 + (1/9)145 + (1/9(167) + (1/9)187 +

(1/9)199

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

= (1/9)(  1308)

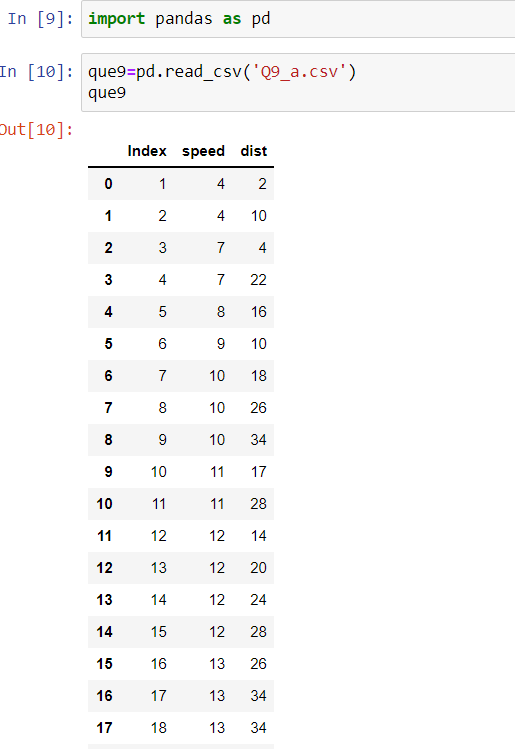
= 145.33

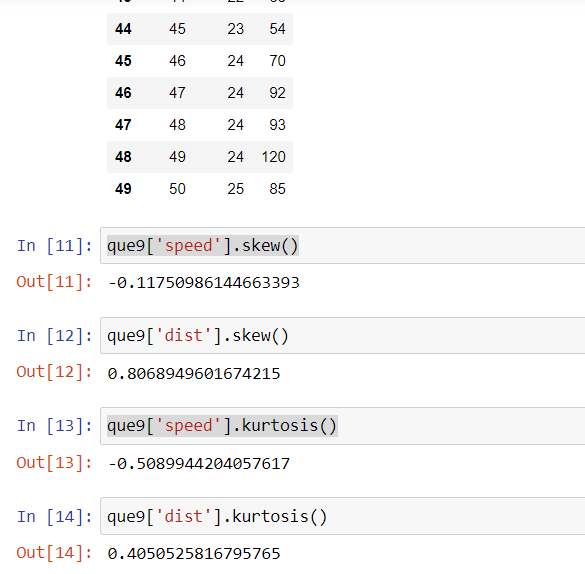
Expected Value of the Weight of that patient = 145.33

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

**Cars speed and distance**

**Use Q9\_a.csv**

****

****

**We calculated the skewness and kurtosis of given data ,we found that skewness of speed is -0.117 so it is negative skewness and the data is fairly symmetrical.The skewness of dist is 0.806 so it is positive skewness and the data is moderately skewed. We found that kurtosis of speed is -0.508 so it is platykurtic that it is lack of outliers.The kurtosis of dist is 0.405 so it is platykurtic it has profusion outliers.**

**Use Q9\_b.csv**

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

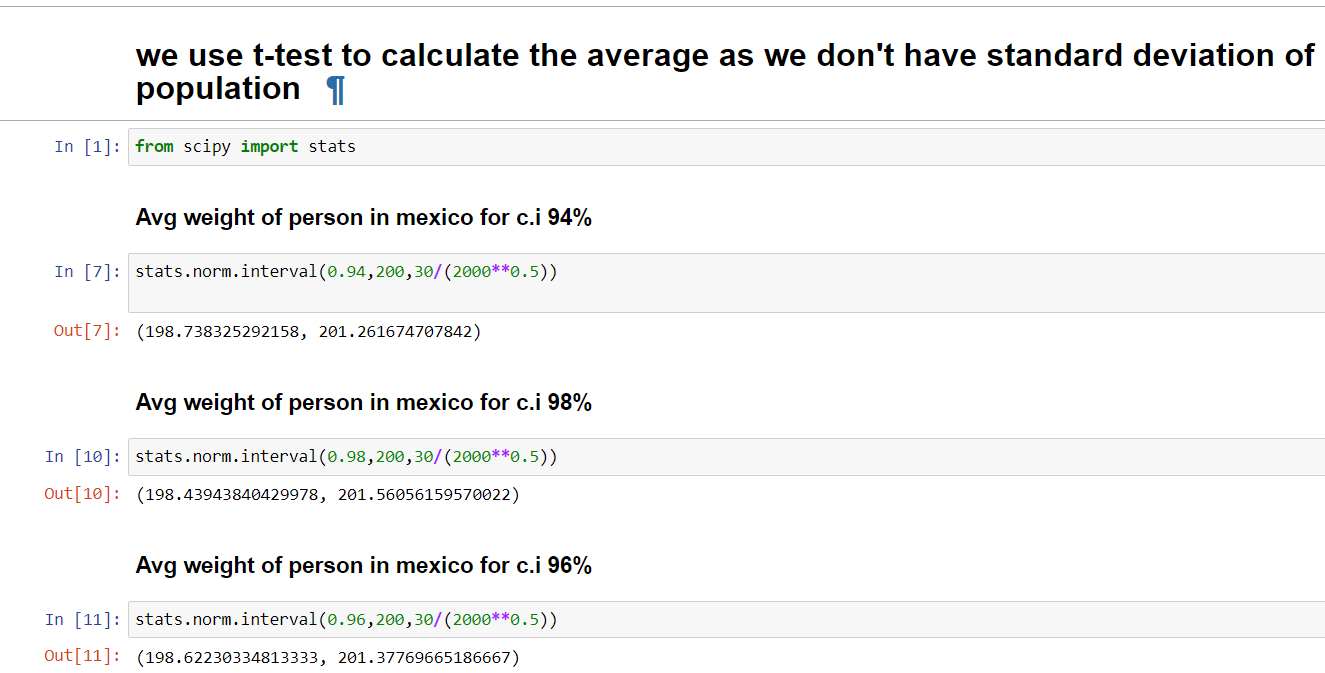




HISTOGRAM: we know that histogram is used to understand distribution of one discrete data. From above histogram we can say that it positive or right skewed because the tail is longer towards right so we can say that mean>median>mode we can also say that most of the ChickWeight$weight lies between 50-100

BOXPLOT:we know that boxplot is used for both univariate and bivariate analysis.From above boxplot we can say that it is positive (or) right skewed because we can clearly see that Q2-Q1<Q3-Q2 and even whiskers . So that mean>median>mode and we can see that there are nearly 8 outliers

**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?

It is normal distribution that is skewness is 0

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

It is positive (or) right skewed

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

It is negative (or) left skewed

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

A positive kurtosis indicates that a distribution is peaked and poses thick tails. It is also known as leptokurtic

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

A negative kurtosis indicates that the distribution has lighter tails than the normal distribution. It is also known as platykurtic

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



What can we say about the distribution of the data?

We can say that median is approximately 15.2, minimum value is less than 2 and maximum value is greater than 19.It is not symmentrical distribution

What is nature of skewness of the data?

It is negative (or) left skewed

What will be the IQR of the data (approximately)?   
 Q3-Q1=18.5-10=8.5

Q19) Comment on the below Boxplot visualizations?



We have two boxplots for Boxplot1 The minimum value is approximately 240,maximum value is 287.5 and median is 262.5 we can also see that it is normal distribution

For Boxplot2 The minimum value 187.5 ,maximum value is 325 and median is 262.5 we can see that is normal distribution.50% of data lies between 250 and 281

For Boxplot1 the minimum value is greater than Boxplot2 , maximum value is less than Boxplot1. Both Boxplots have same median and they are normal distributions.50% of data lies between 225 and 312.5 and Boxplot2 is more dispersed than Boxplot10

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

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

Data \_set: Cars.csv

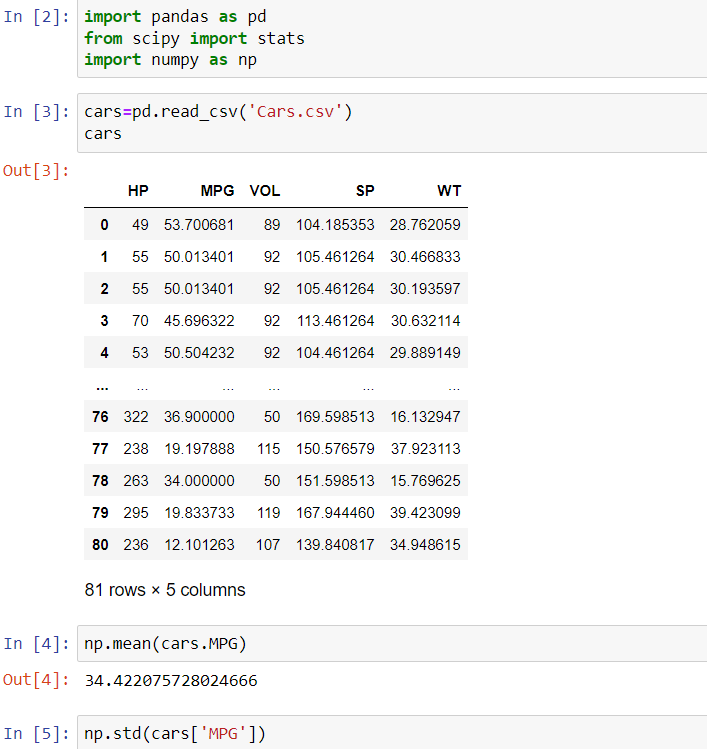
Calculate the probability of MPG of Cars for the below cases.

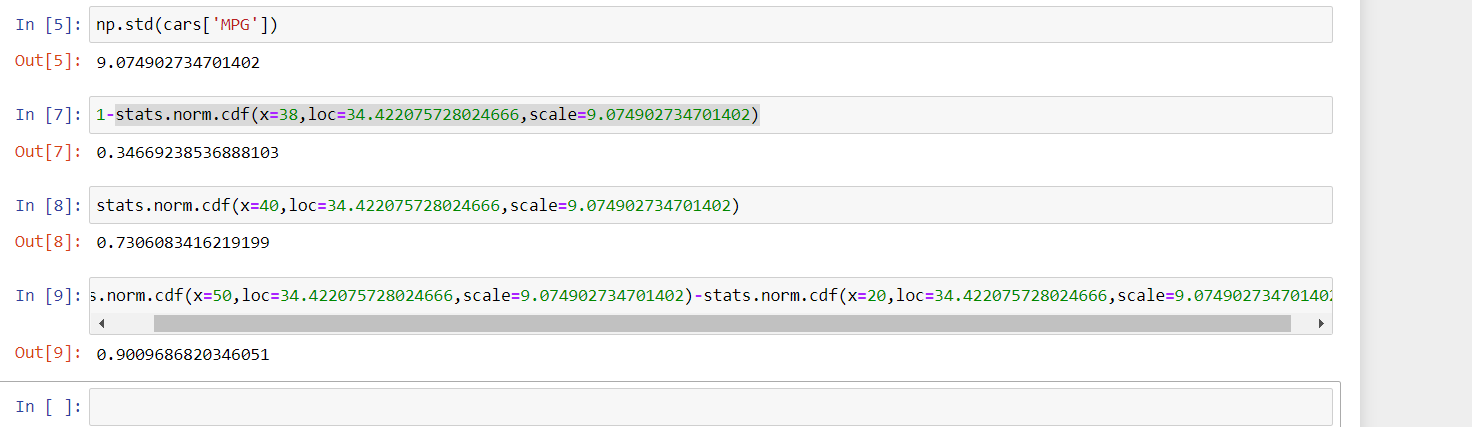
MPG <- Cars$MPG

a. P(MPG>38)

b. P(MPG<40)

c. P (20<MPG<50)





a. P(MPG>38)=34.66%

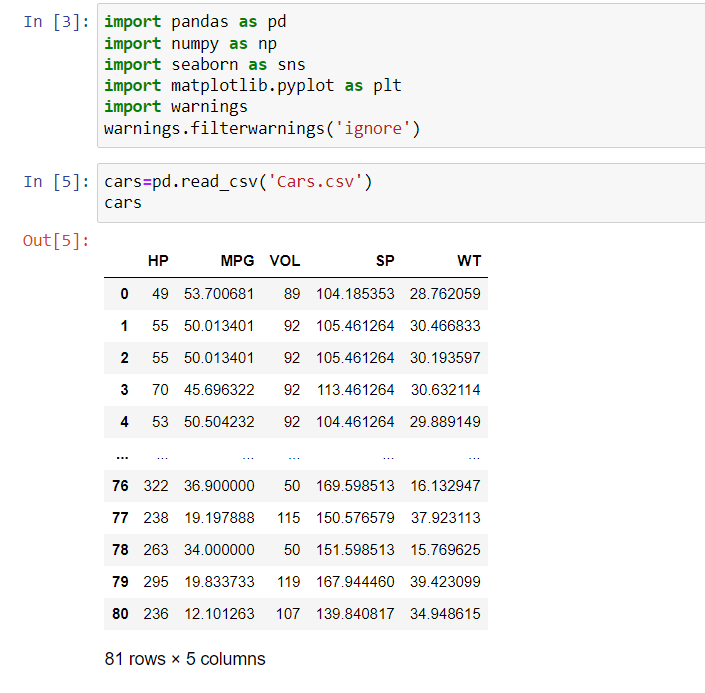
b. P(MPG<40)=73.06%

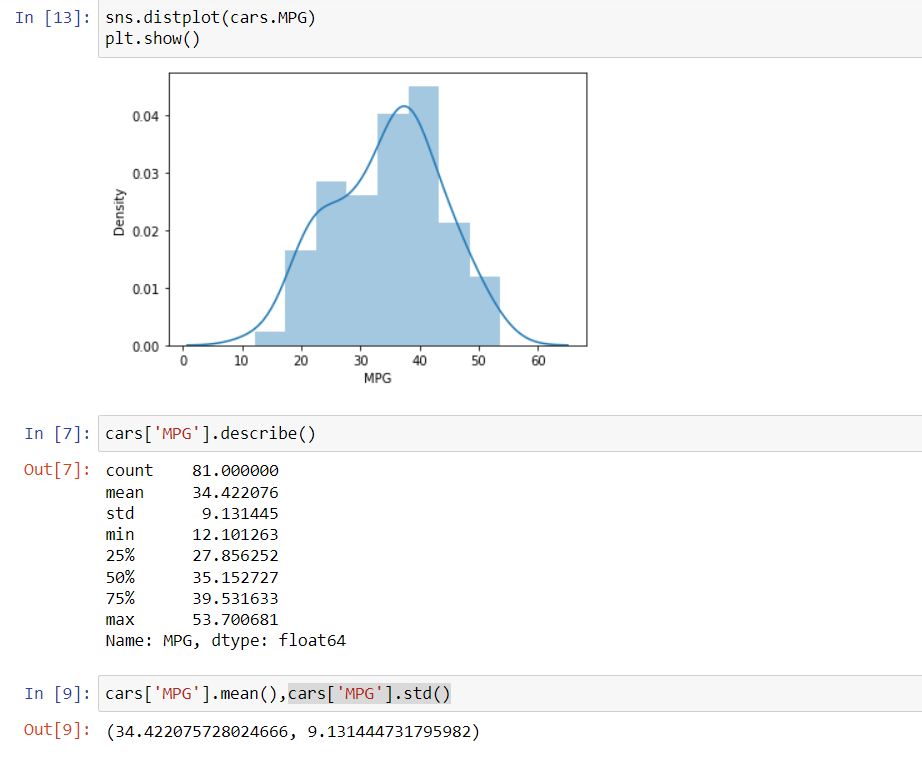
c. P (20<MPG<50)=90.09%

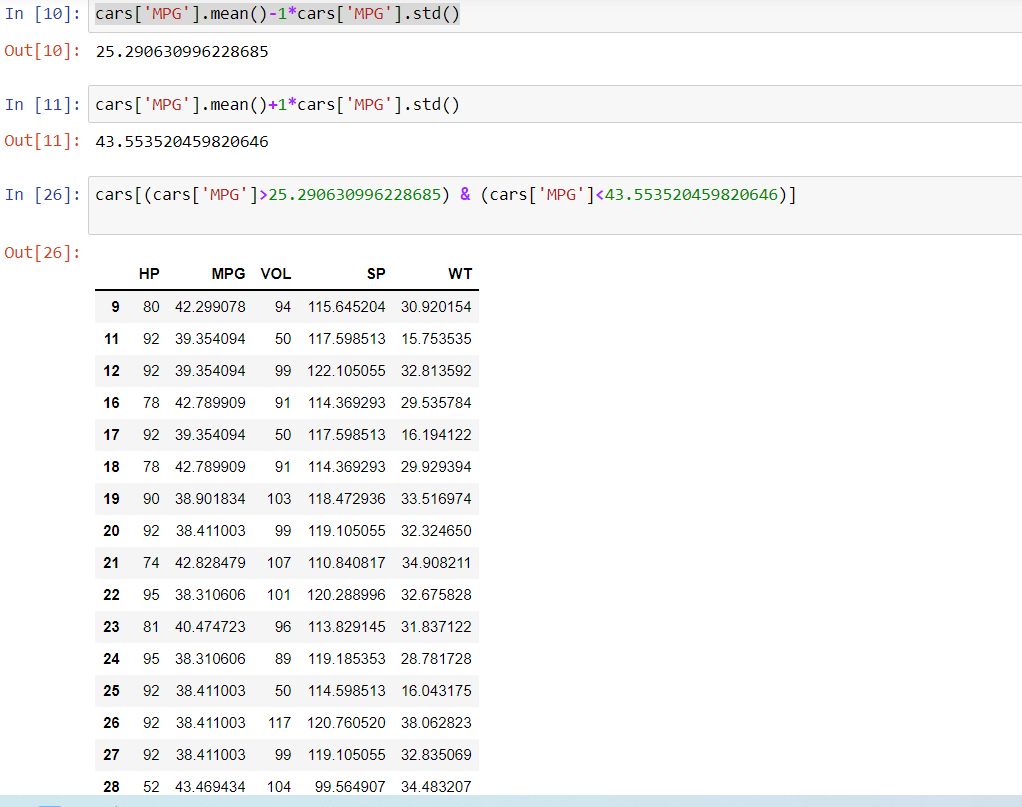
Q 21) Check whether the data follows normal distribution

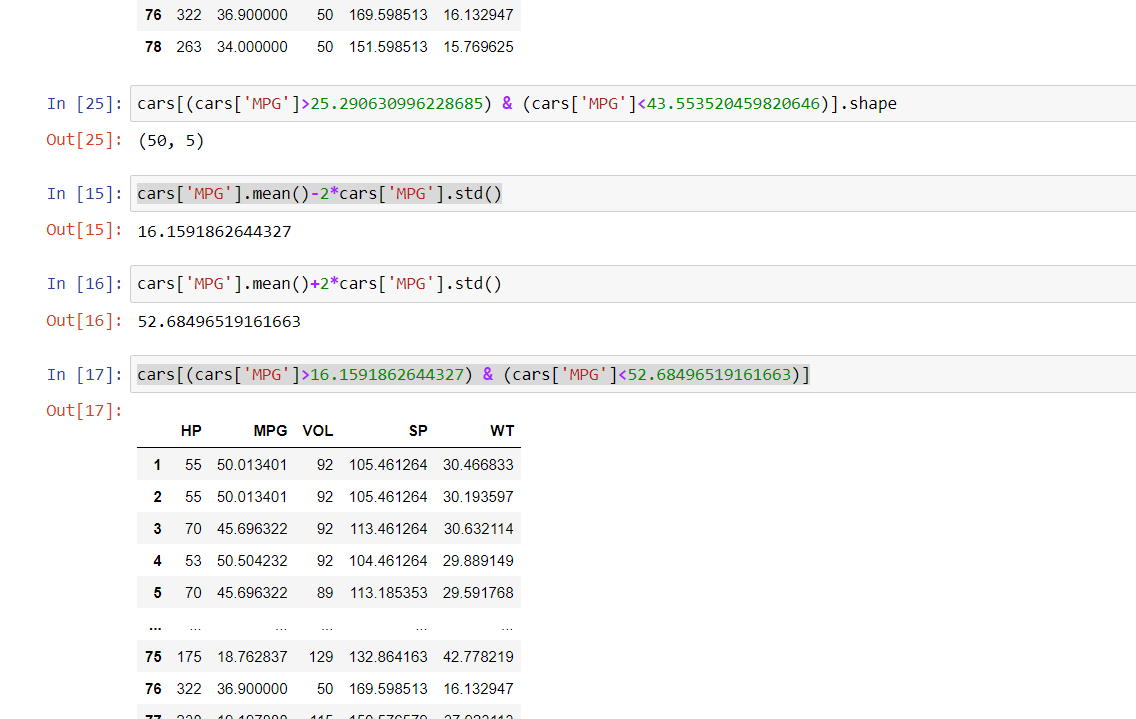
1. Check whether the MPG of Cars follows Normal Distribution

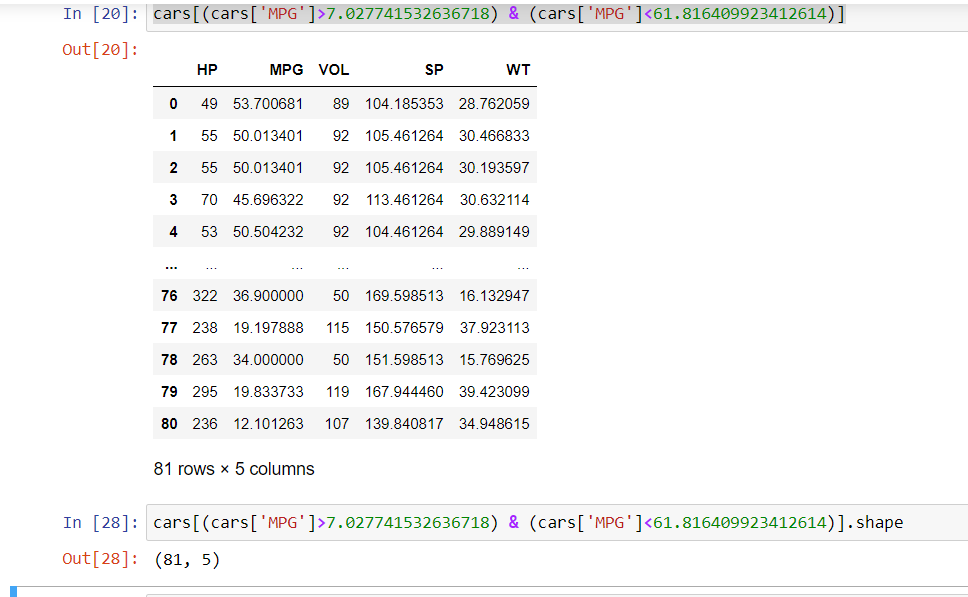
Dataset: Cars.csv







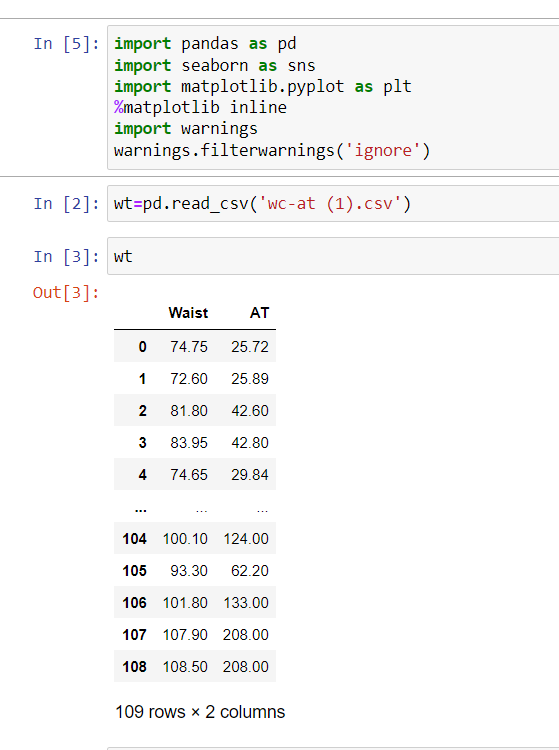


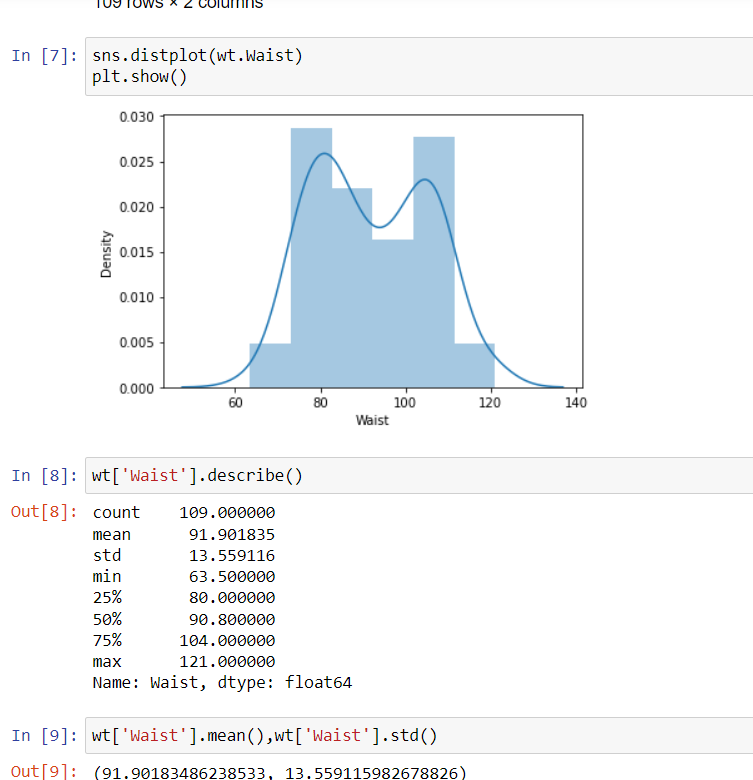


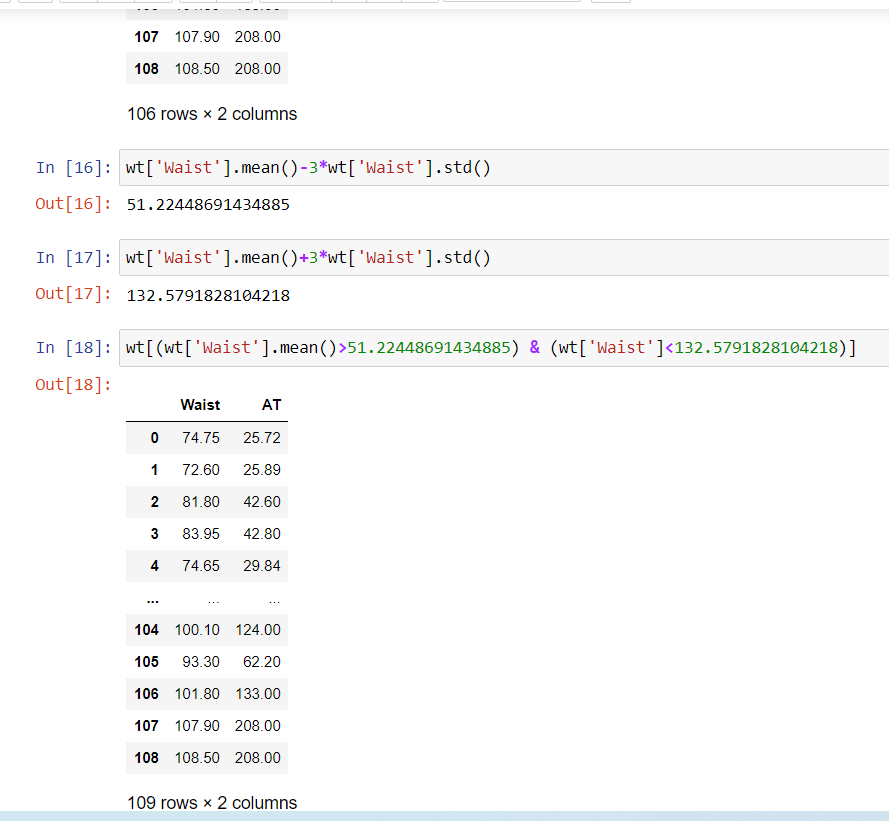
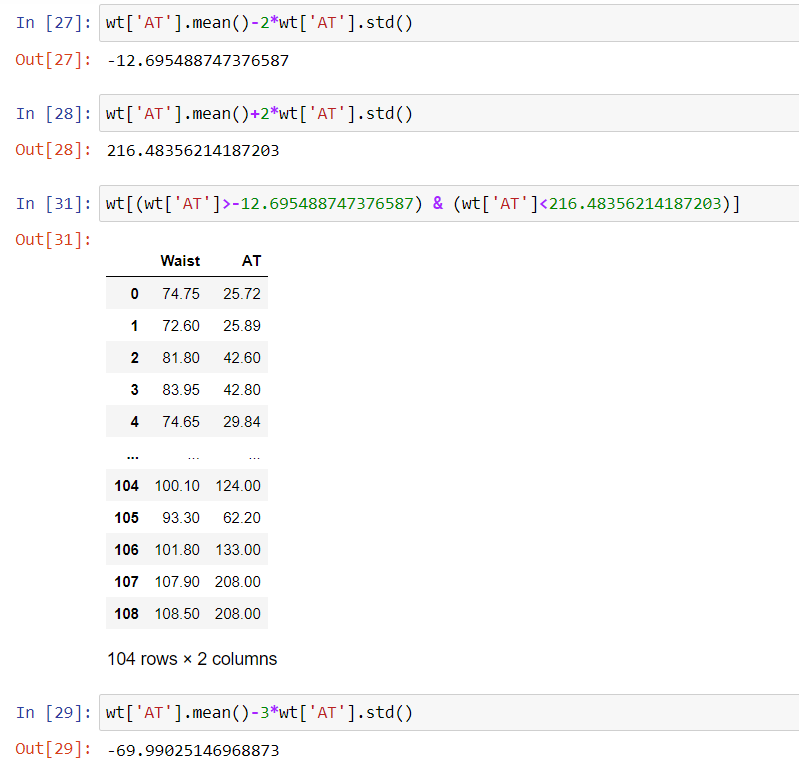
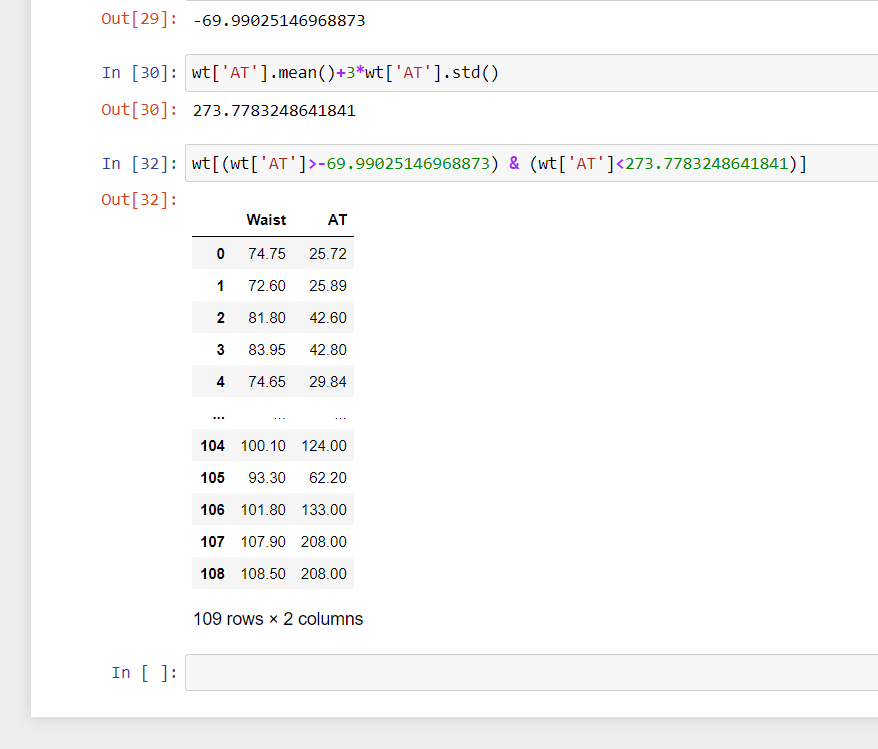
No, it doesn’t follow normal distribution

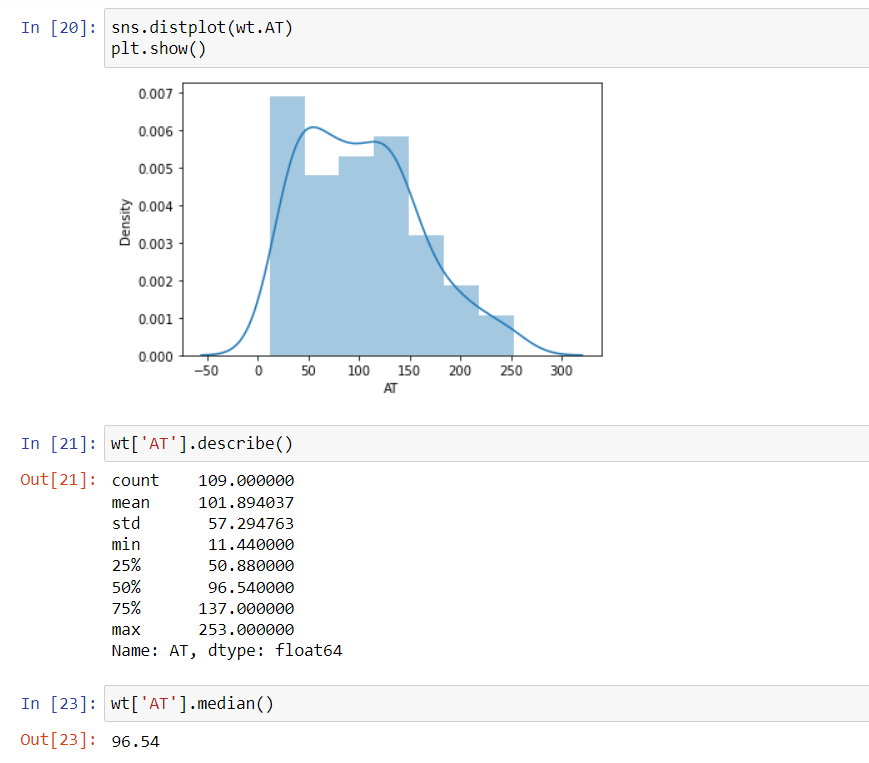
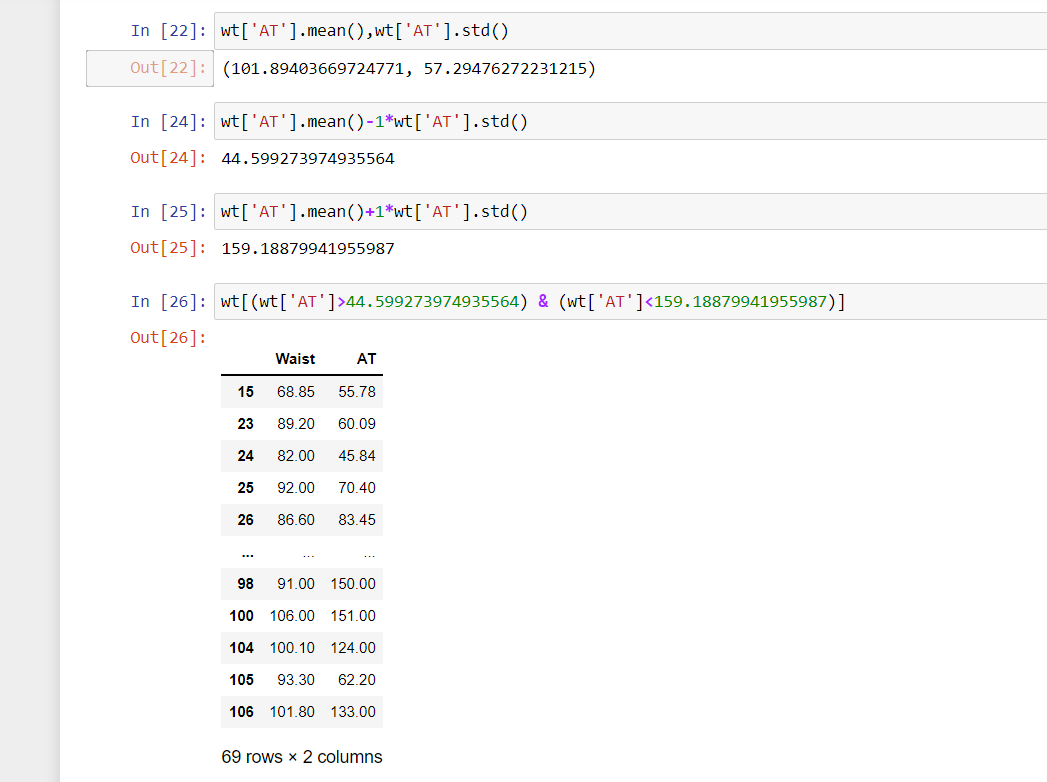
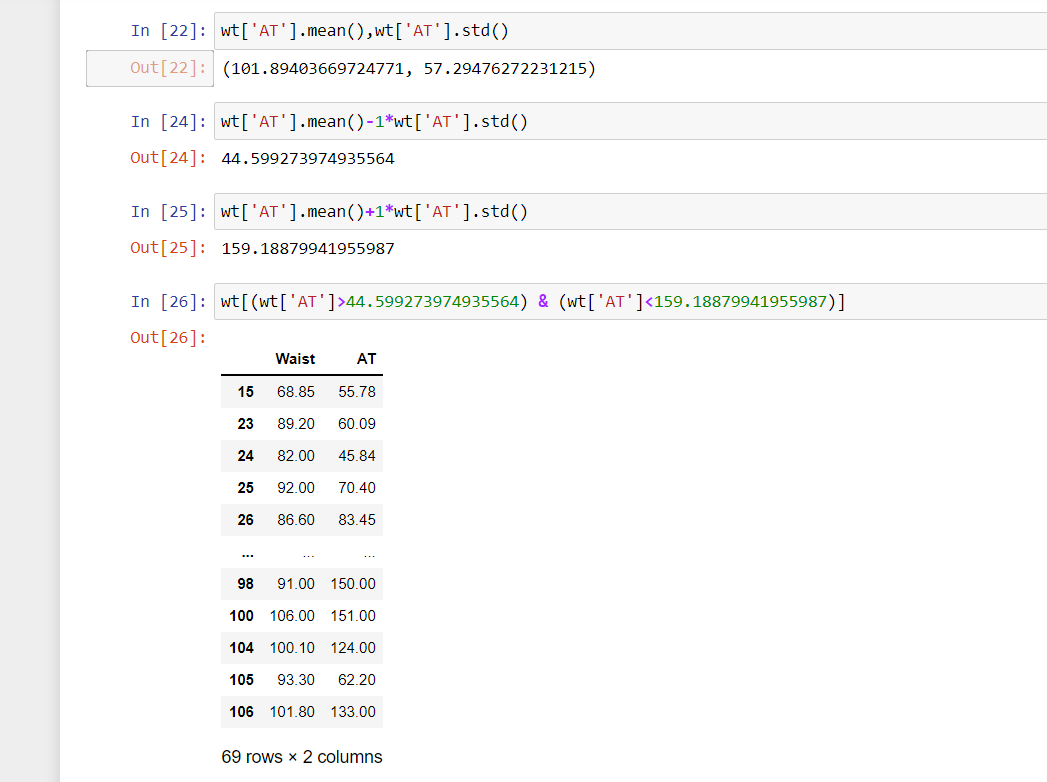
B) Check Whether the Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set follows Normal Distribution

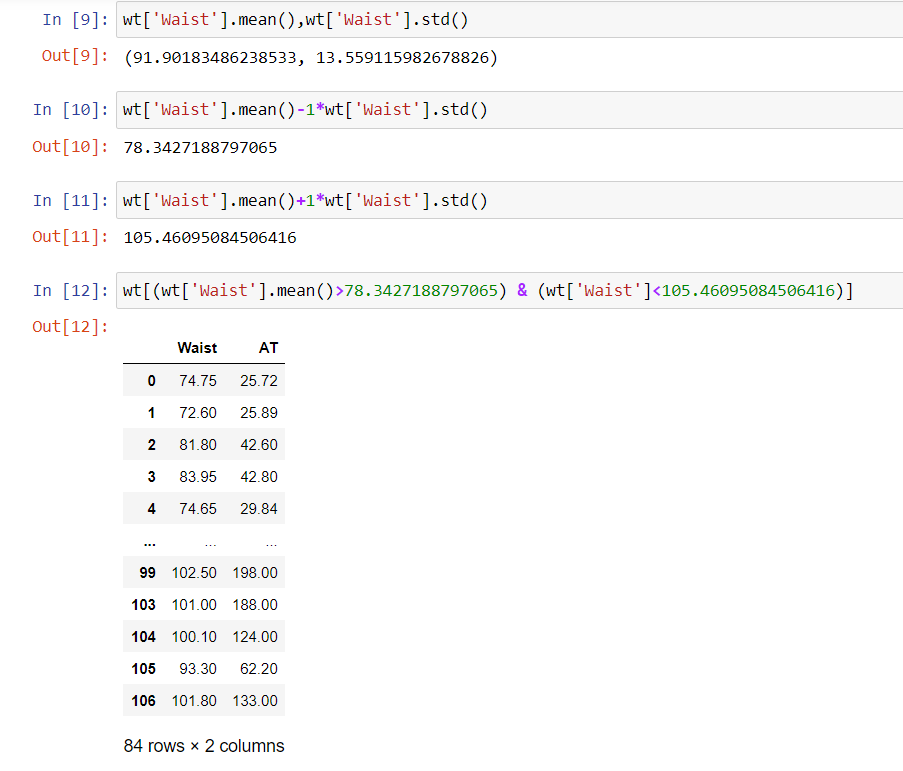
C) Dataset: wc-at.csv











No, both features doesn’t follow normal distribution

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

