|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | Discrete Data |
| Results of rolling a dice | Discrete Data |
| Weight of a person | Continuous Data |
| Weight of Gold | Continuous Data |
| Distance between two places | Continuous Data |
| Length of a leaf | Continuous Data |
| Dog's weight | Continuous Data |
| Blue Color | Discrete Data |
| Number of kids | Discrete Data |
| Number of tickets in Indian railways | Discrete Data |
| Number of times married | Discrete Data |
| Gender (Male or Female) | Discrete Data |

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

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

Ans- Total no of outcomes when three coins are tossed=8

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

Probability that two heads and one tail:

{HHT,HTH,THH}

Let E= two heads and one tail

P(E) = No of favourable outcomes/Total no of outcomes

=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

Ans- When two dice are rolled 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)}

N(s)=36

1. Equal to 1

n(E)=Total favourable outcomes=0

P(E)=n(E)/n(S)=0/36=0

1. Interested events ={(1,1)(1,2),(1,3),(2,1),(2,2),(3,1)}

Therefore Reqd.probability=6/36

1. Interested events ={(1,5)(2,4),(3,3),(4,2),(5,1),(6,6)}

Therefore Reqd.probability=6/36

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

Let S be the sample space

n(S)=No of ways of drawing 2 balls out of 7=7c2=(7\*6)/(2\*1)=21

Let E=Event of drawing 2 balls,none of which is blue

n(E)= No of ways of drawing 2 balls out of 5 balls=5c2=(5\*4)/(2\*1)=10

p(E)=n(E)/n(S)=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:

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

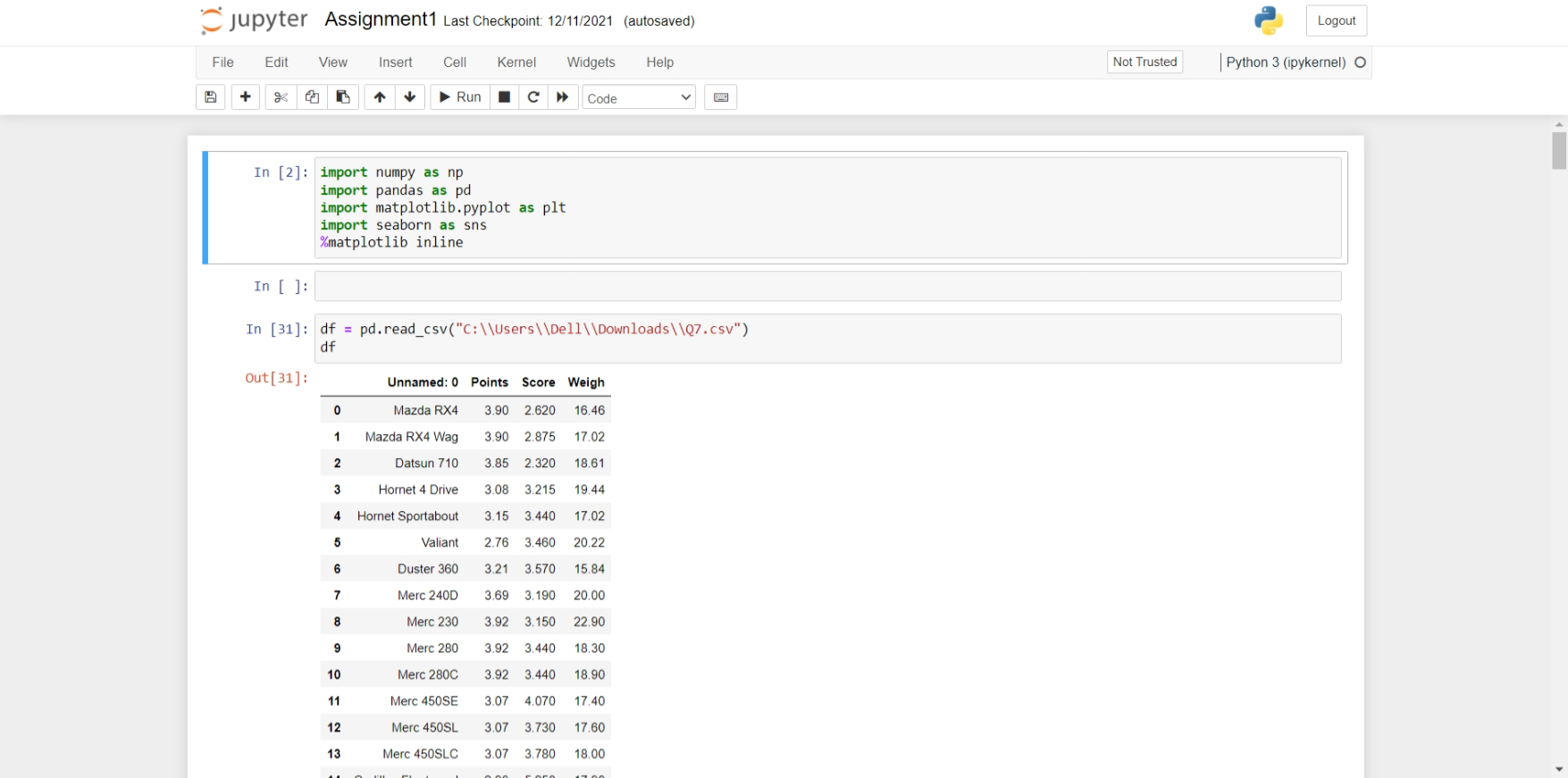
=3.09

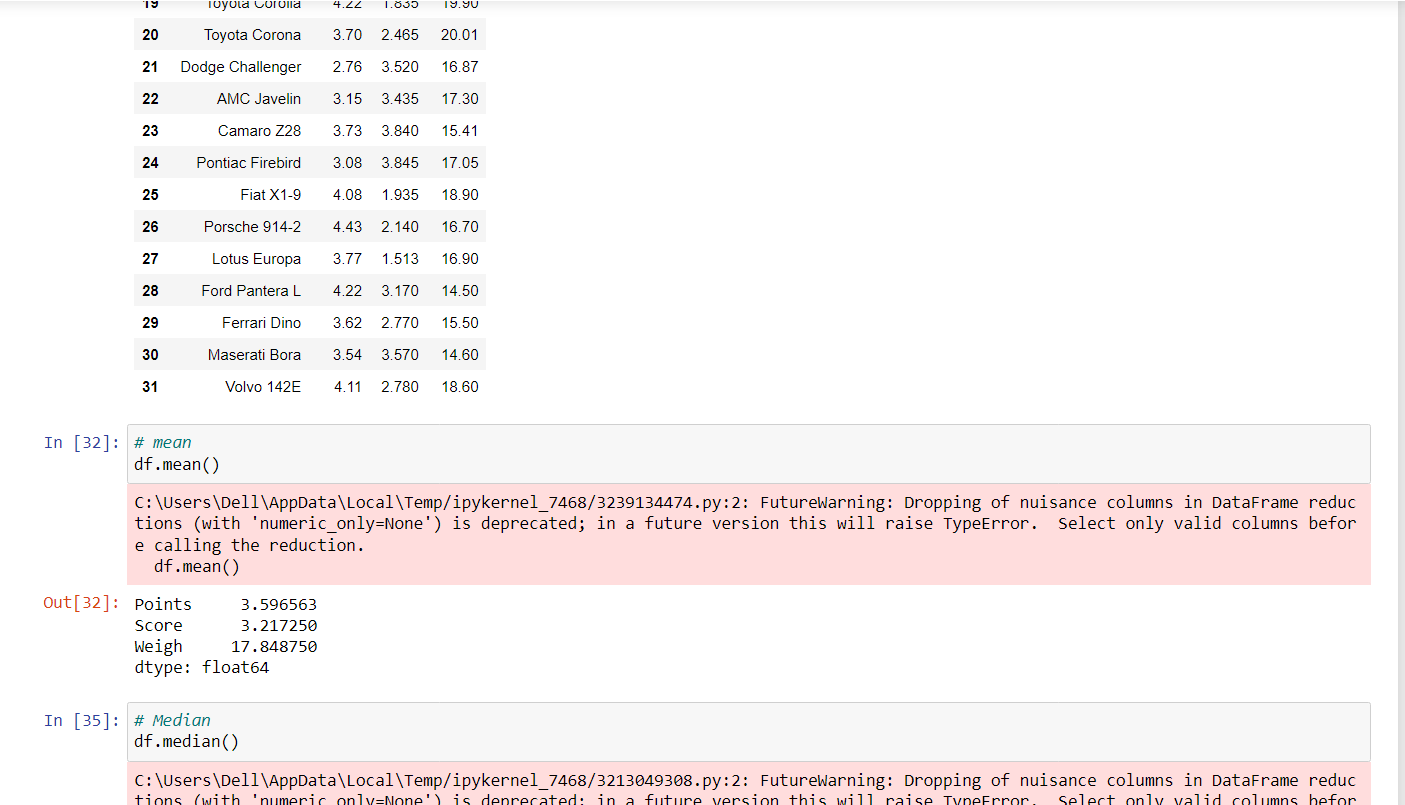
Q7) Calculate Mean, Median, Mode, Variance, Standard Deviation, Range & comment about the values / draw inferences, for the given dataset

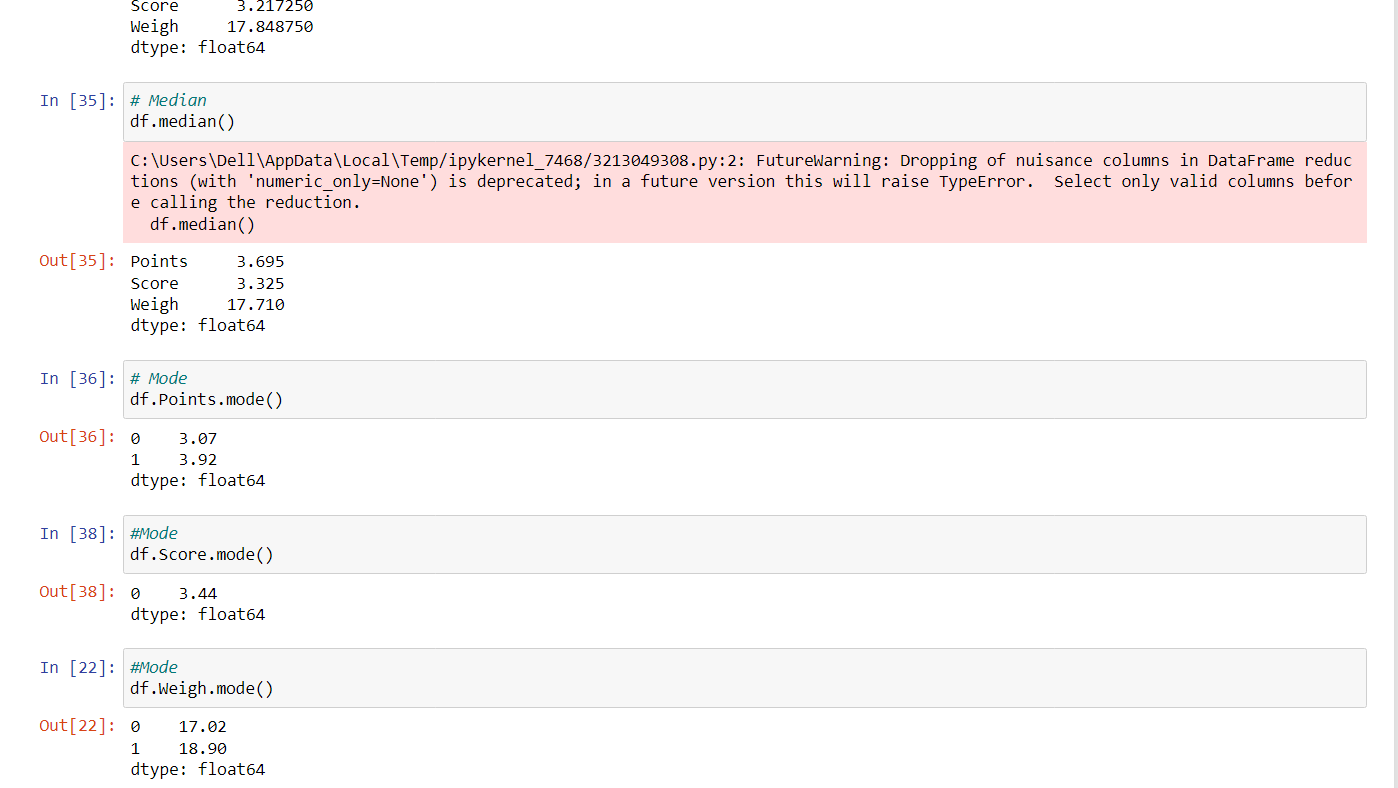
* For Points,Score,Weigh>

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

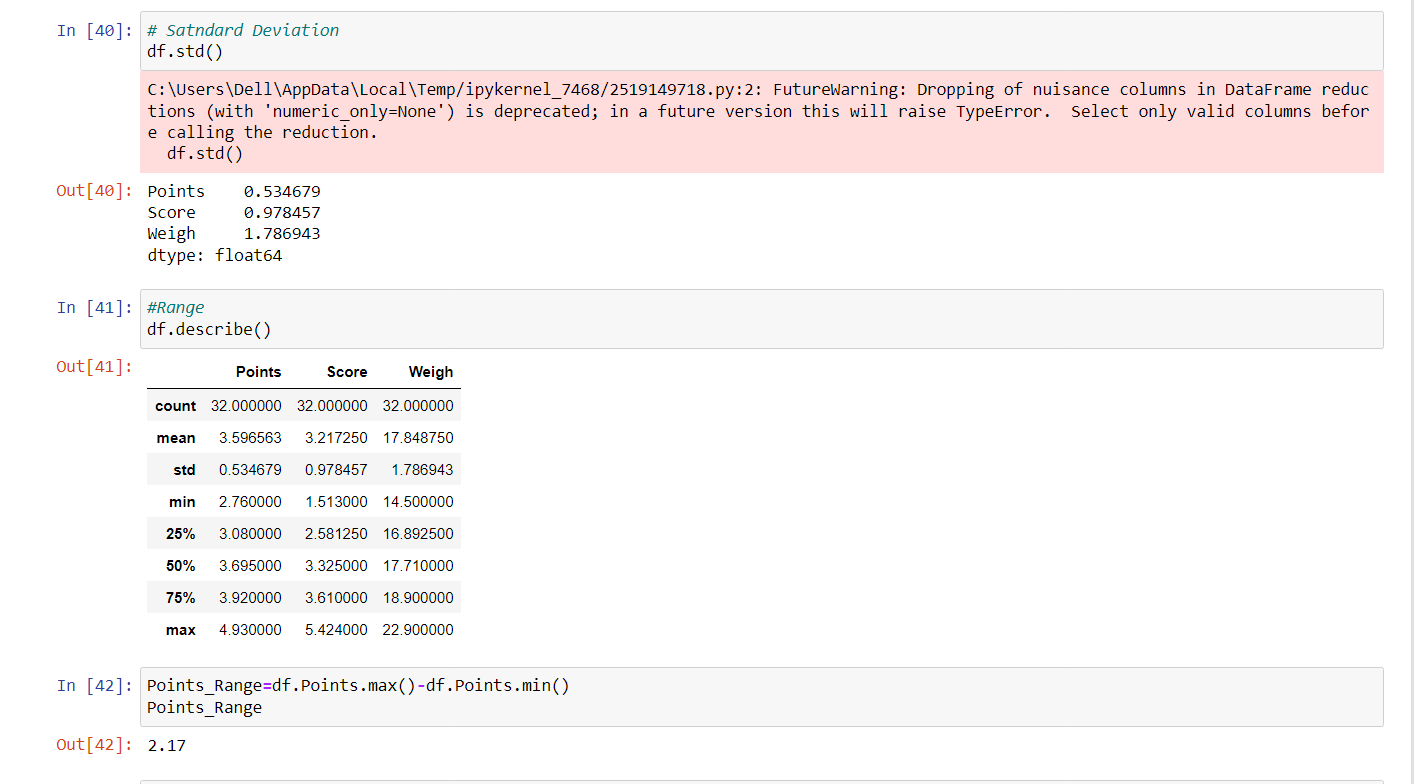
**Use Q7.csv file**

**ANS-**

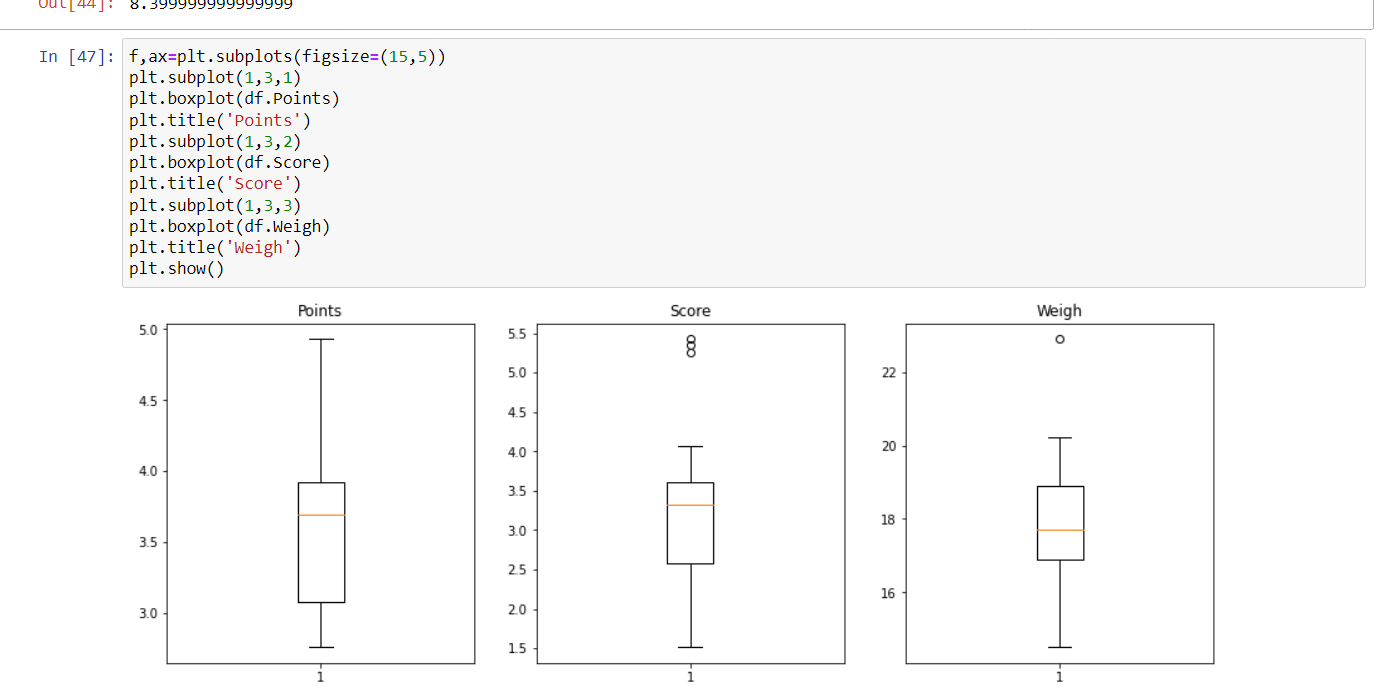
****

****

****

****

****

****

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- 108+110+123+134+135+145+167+187+199

= 1308

= 1308/9

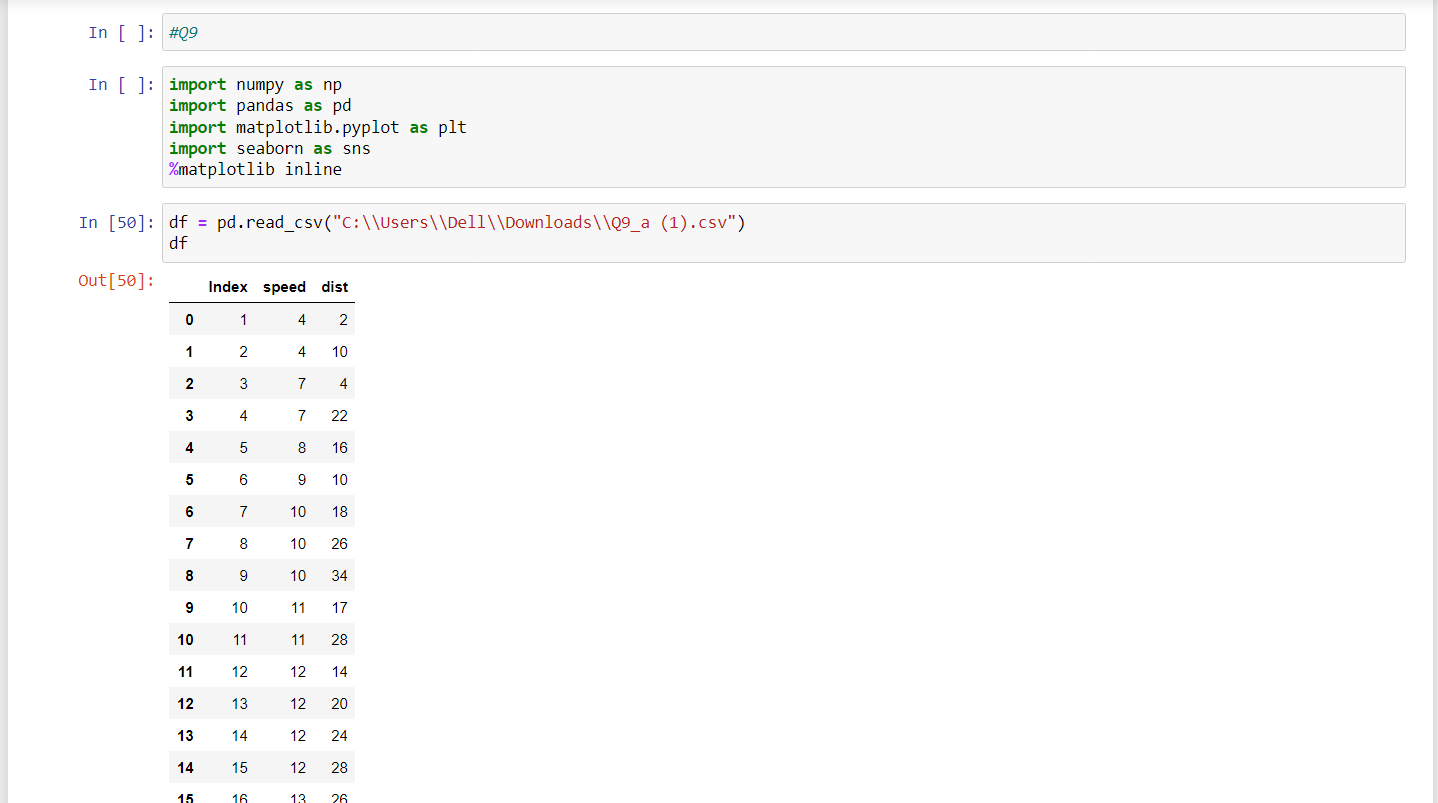
=145.333

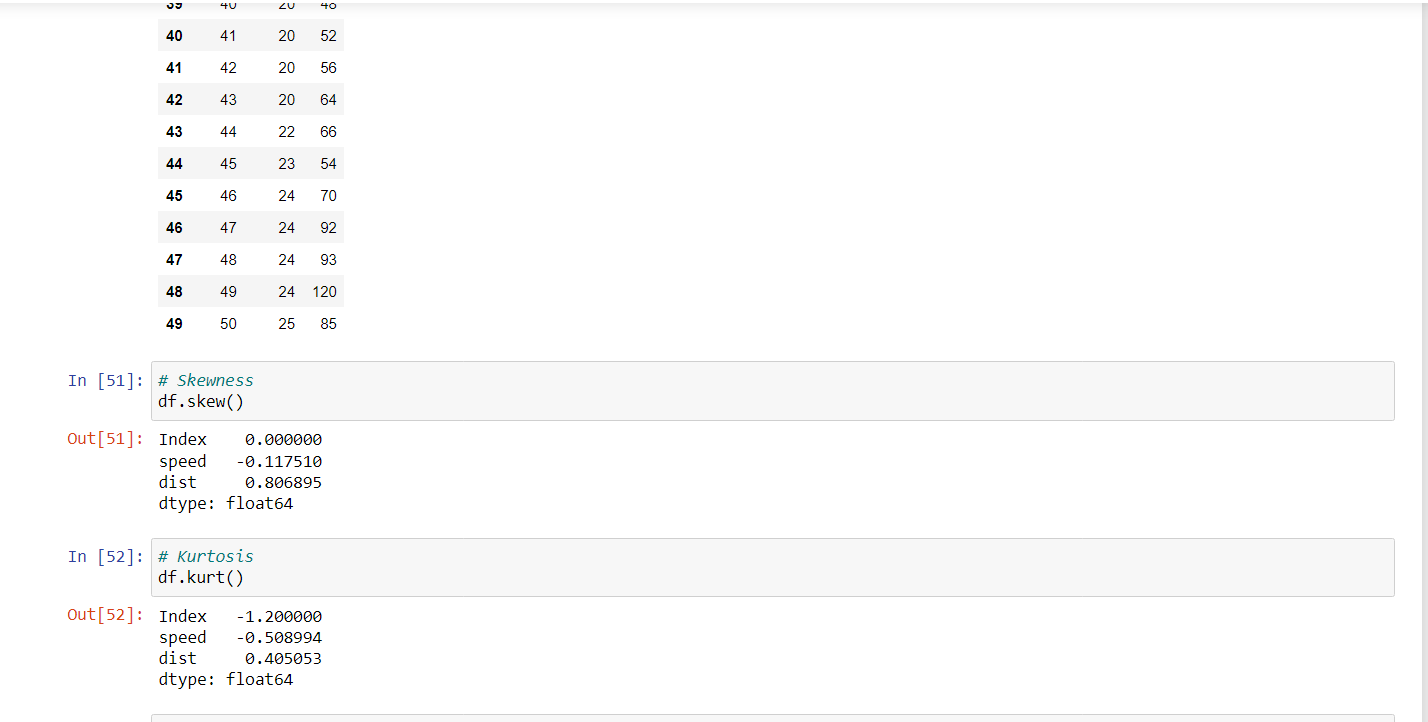
**Q9) Calculate Skewness, Kurtosis & draw inferences on the following data**

**Cars speed and distance**

**Use Q9\_a.csv**

**ANS-**

****

****

df.speed.skew() => -0.11750986144663393

If skewness value is negative (-) then the data is stated as negatively skewed or left skewed. Since our data is negatively skewed here, it means that it has a lower number of data points having low values.

And if value is lied between -0.5 to 0.5 then it is saying as fairly symmetrical

So, speed in data set Q9\_a.csv is fairly symmetrical.

df.dist.skew()=> 0.8068949601674215

Since our data is positively skewed here, it means that it has a higher number of data points having low values

If skewness value is positive (+) then the data is stated as positively skewed or right skewed

So, distance in data set Q9\_a.csv is right skewed or positive skewed.

df.speed.kurt() => -0.5089944204057617

Negative values of kurtosis indicate that a distribution is flat and has thin tails. Platykurtic distributions have negative kurtosis values.

A platykurtic distribution is flatter (less peaked) when compared with the normal distribution, with fewer values in its shorter (i.e. lighter and thinner) tails.

df.dist.kurt()=>0.4050525816795765

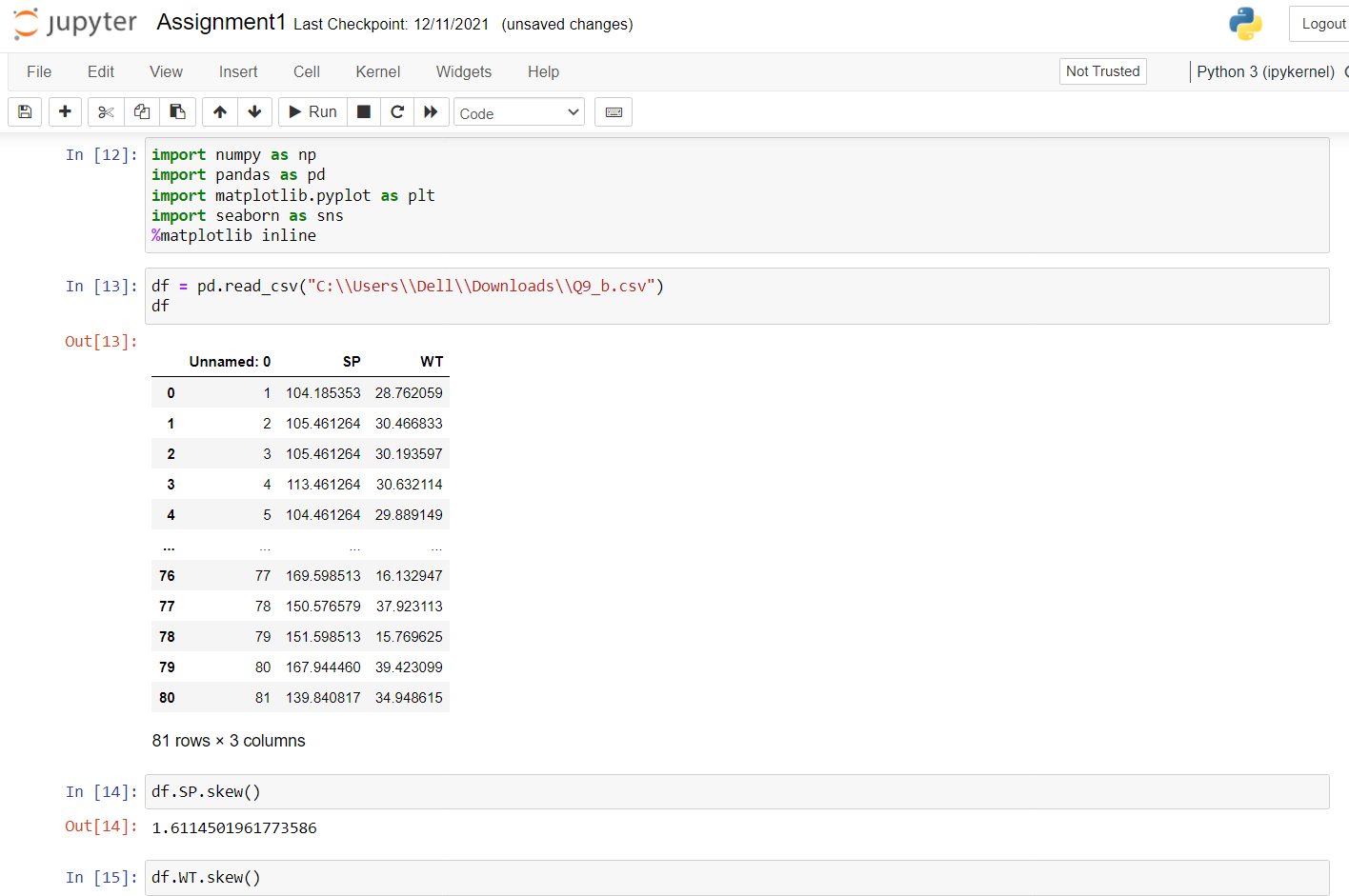
Positive values of kurtosis indicate that a distribution is peaked and possess thick tails. Leptokurtic distributions have positive kurtosis values.

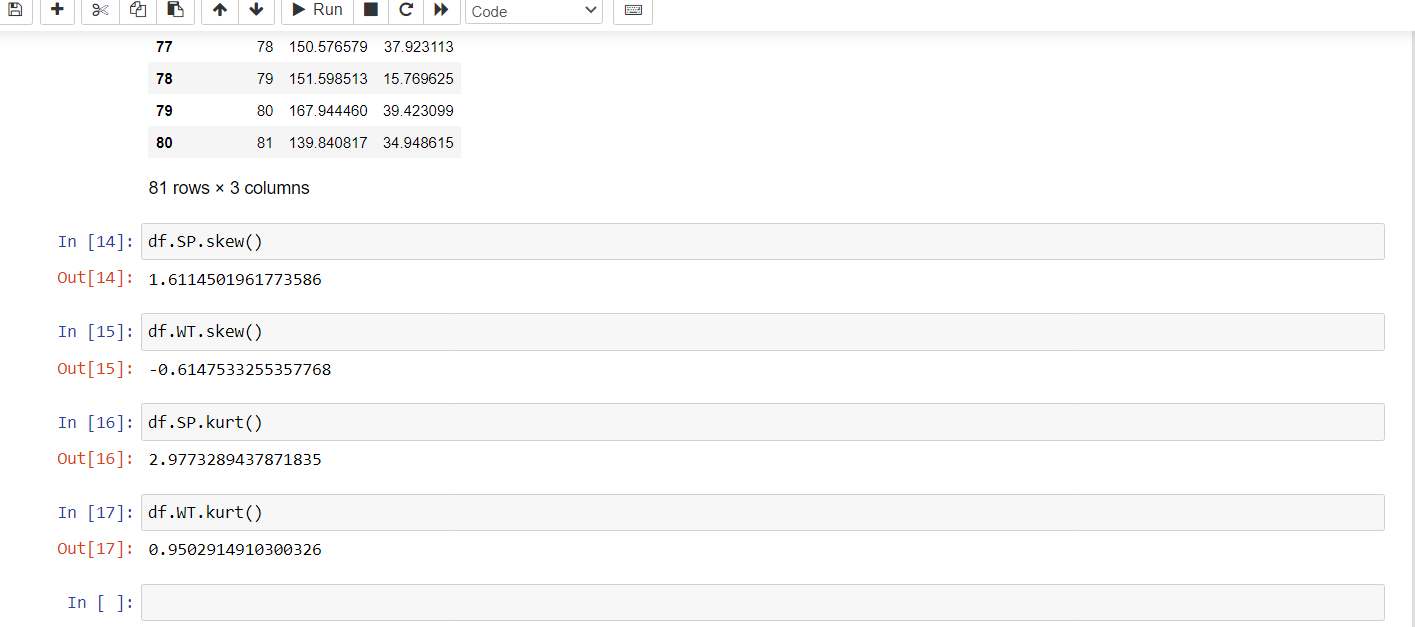
A leptokurtic distribution has a higher peak and taller (i.e. fatter and heavy) tails than a normal distribution.

If kurtosis = 0 then it is equal to normal distribution.

SP and Weight(WT)

Use Q9\_b.csv

ANS- 



df.SP.skew() => 1.6114501961773555

If skewness value is positive (+) then the data is stated as positively skewed or right skewed. Since our data is positively skewed here, it means that it has a higher number of data points having low values

So, distance in data set Q9\_a.csv is right skewed or positive skewed.

df.WT.skew() => -0.6147533255357768

A negatively skewed distribution is the distribution with the tail on its left side. The value of skewness for a negatively skewed distribution is less than zero.

Since our data is negatively skewed here, it means that it has a lower number of data points having low values.

df.SP.kurt() => 2.9773289437871764

Positive values of kurtosis indicate that a distribution is peaked and possess thick tails. Leptokurtic distributions have positive kurtosis values. In a data the obtained kurtosis is positive then middle values in that data frame are so high compared to the left and right side values.

df.WT.kurt() => 0.9502914910300326

Positive values of kurtosis indicate that a distribution is peaked and possess thick tails. Leptokurtic distributions have positive kurtosis values. In a data the obtained kurtosis is positive then middle values in that data frame are so high compared to the left and right side values.

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



The distribution is right Skewed. Mean>Median



The above boxplot suggests thst the distribution has lots of outliers upper extreme.

**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- At 94% confidence interval is (143.57619175546247, 256.42380824453755)

At 96% confidence interval is (138.38753268104531, 261.61246731895466)

At 98% confidence interval is (130.2095637787748, 269.7904362212252)

**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.

Ans-Mean=41

Median=40.5

Variance=25.53

Standard deviation=5.05

1. What can we say about the student marks?

Ans-Mean>median. This implies that the distribution is slightly skewed towards right. No outliers are present.

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

Ans-Skewness=0, Symmetric

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

Ans-Right Skewed

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

Ans-Left Skewed

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

Ans- Sharp peak, Thick tails

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

Ans- Broad peak, Thin tails.

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



What can we say about the distribution of the data?

Ans-Not a normal distribution

What is nature of skewness of the data?

Ans-Left Skewed

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

Ans-18-10=8

Q19) Comment on the below Boxplot visualizations?



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

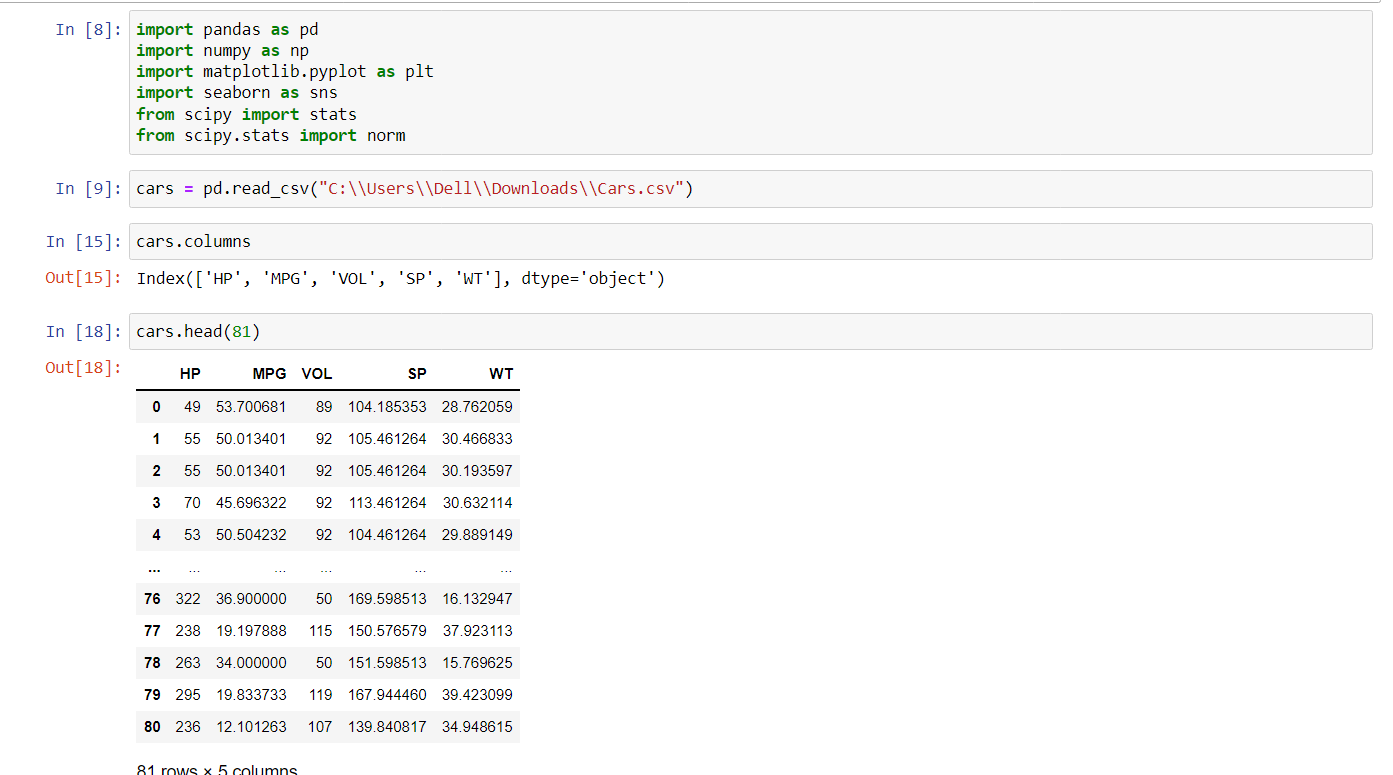
Ans- Both are normally distributed

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

Data \_set: Cars.csv

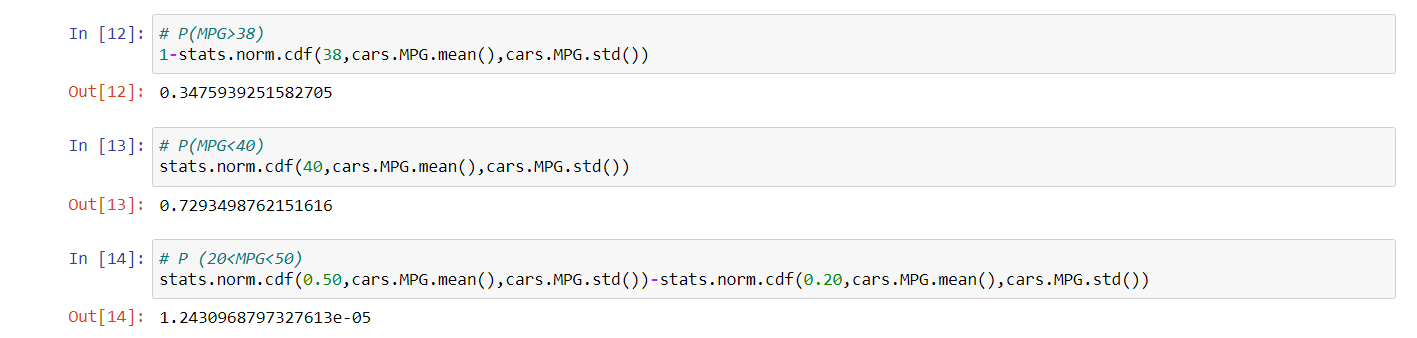
Calculate the probability of MPG of Cars for the below cases.

MPG <- Cars$MPG

ANS-



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

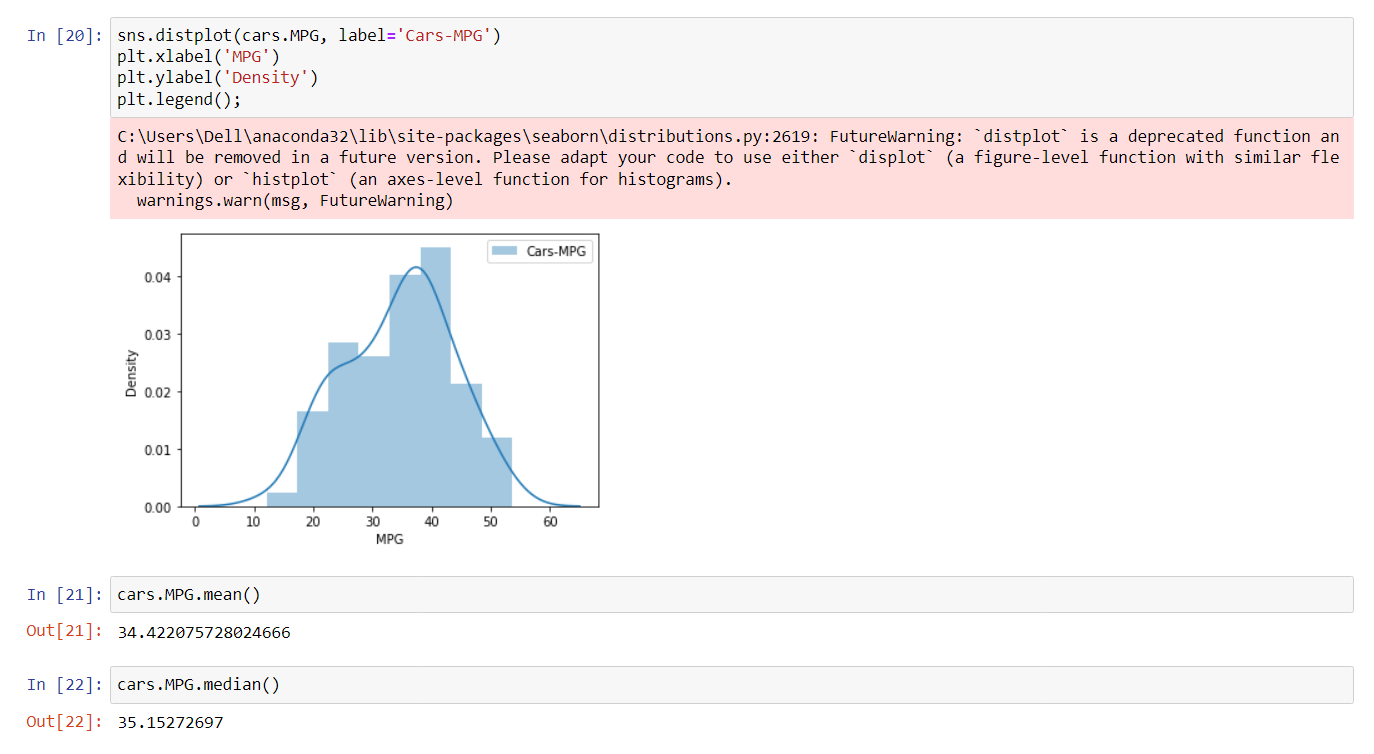
ANS-

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

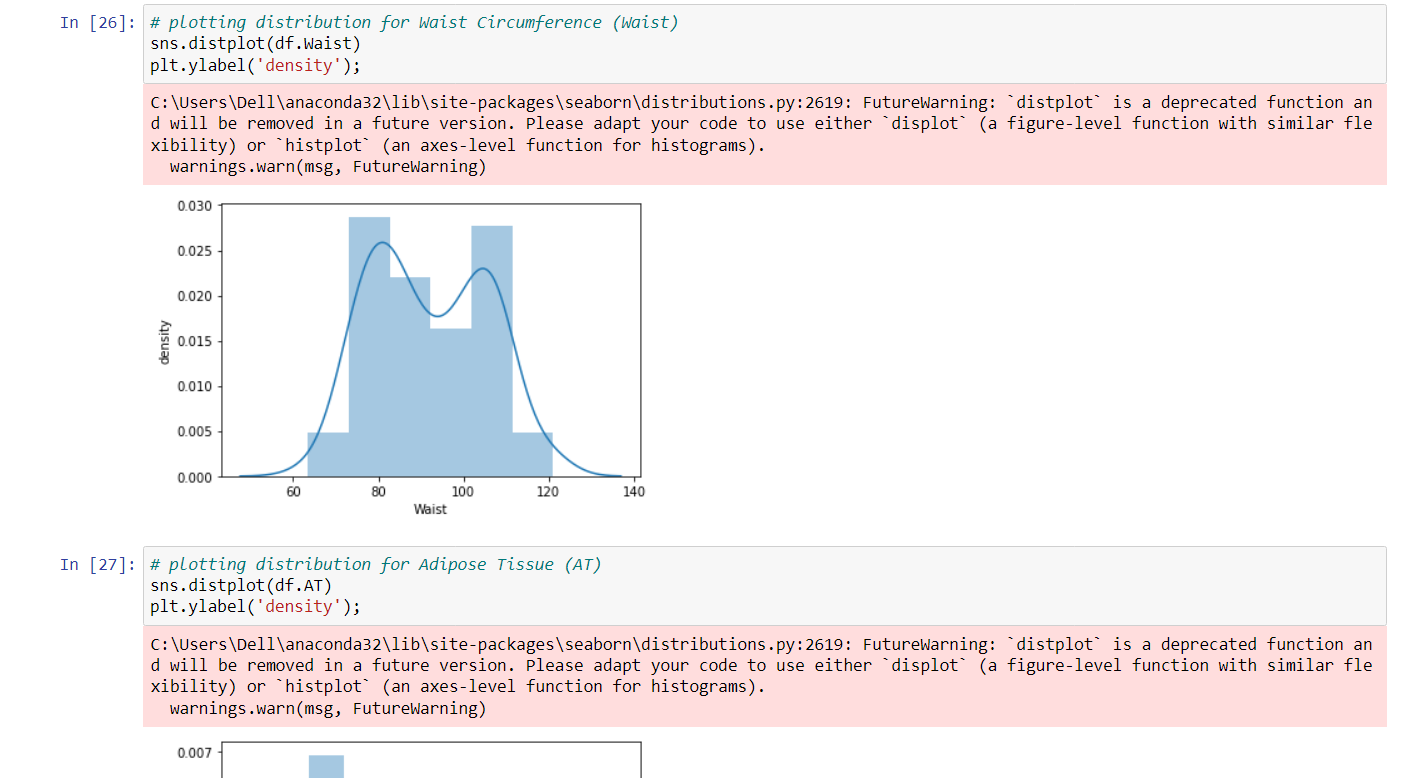
ANS-

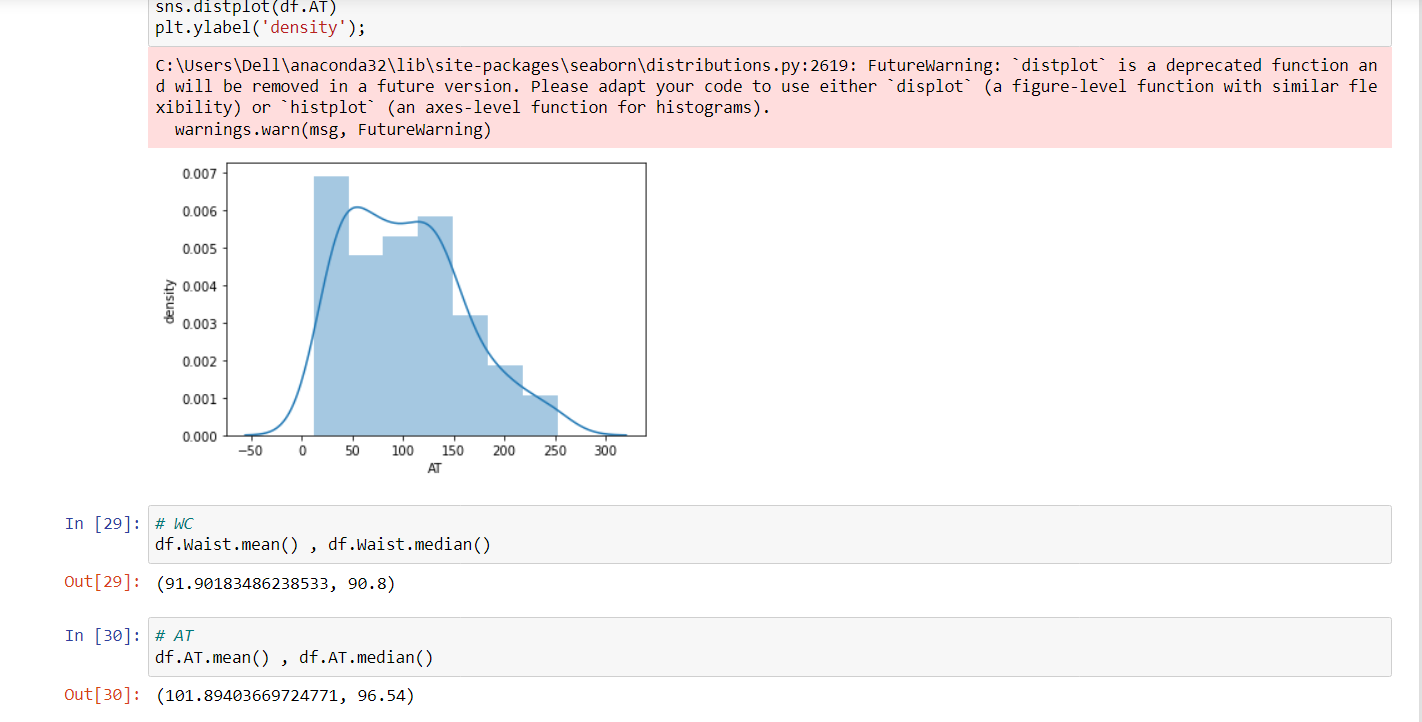


1. Check Whether the Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set follows Normal Distribution

Dataset: wc-at.csv

ANS-



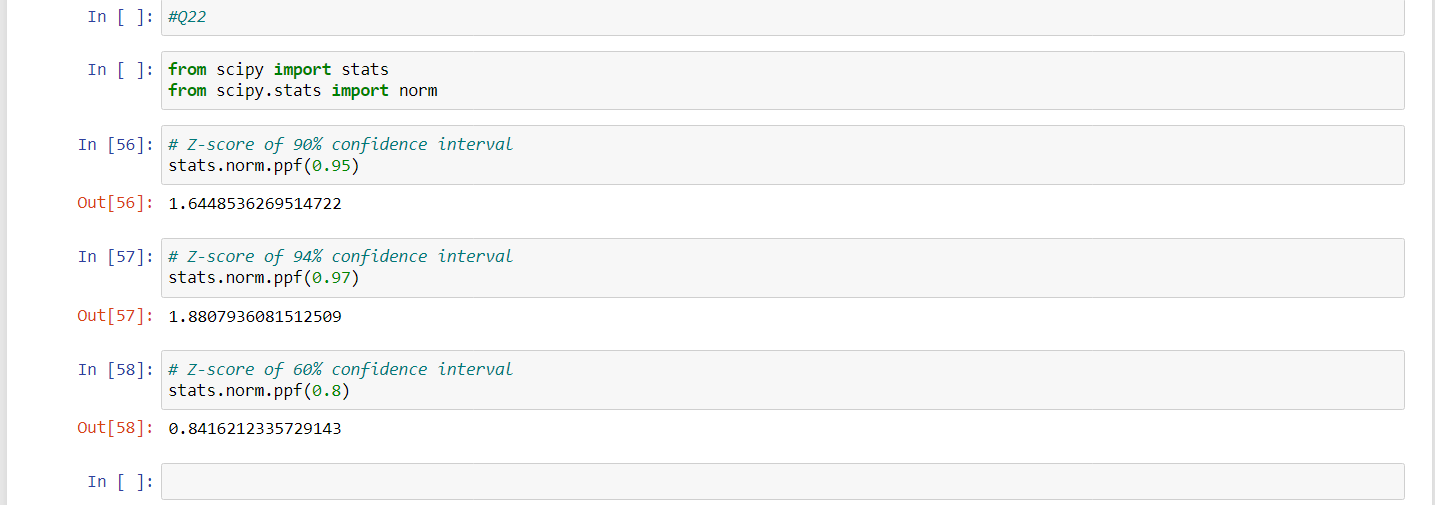


Q 22) Calculate the Z scores of 90% confidence interval,94% confidence interval, 60% confidence interval

ANS- Z scores of 90% confidence interval=0.95

Z scores of 94% confidence interval=0.97

Z scores of 60% confidence interval=0.8



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

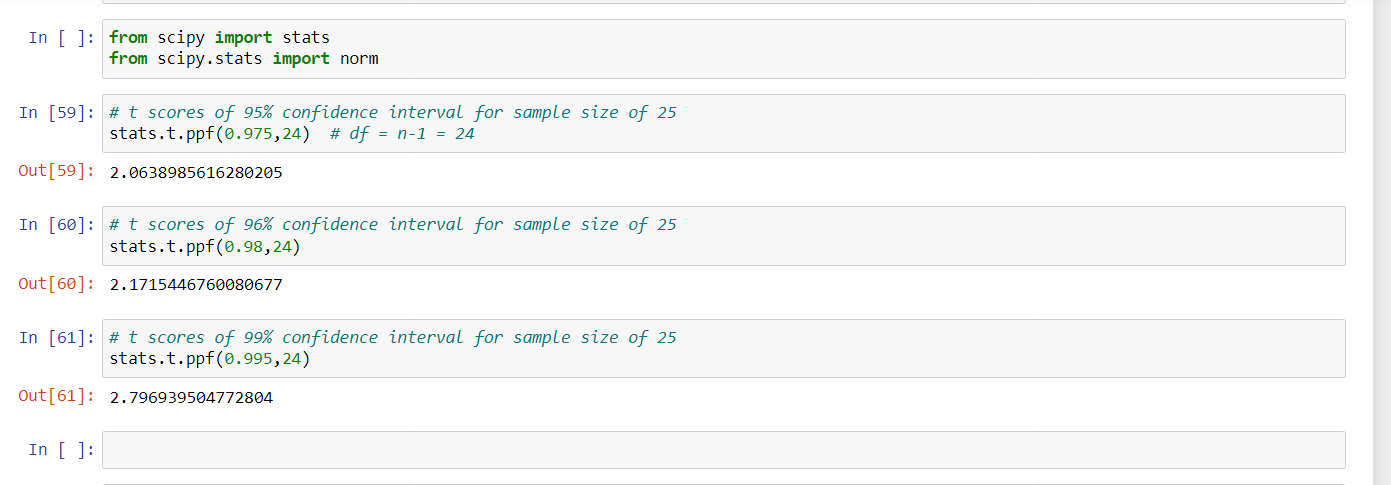
ANS- df - degrees of freedom

=n-1=24

t scores of 95% confidence interval= (0.975,24)

t scores of 95% confidence interval=(0.98,24)

t scores of 95% confidence interval=(0.995,24)

ANS-

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

ANS- µ=270,n=18,xbar=260,s=90

t score=(x bar-µ)/(s/sqrt(n))

=(260-270)/(90/sqrt(18))

=-10/21.23

=-0.4714052



df 🡪 degrees of freedom

t - statistics for the data is given as follows:



x = mean of the sample of bulbs =  260

μ = population mean = 270

s = standard deviation of the sample = 90

n = number of items in the sample = 18

t = (260-270)/(90/√18)

t = -10/(90/3√2)

t = -10/(30/√2)

t = (-1\*√2) / 3

t = - 0.471

For probability calculations, the number of degrees of freedom is n - 1, so here you need the t-distribution with 17 degrees of freedom.

The probability that t < - 0.471 with 17 degrees of freedom assuming the population mean is true, the t-value is less than the t-value obtained With 17 degrees of freedom and a t score of - 0.471, the probability of the bulbs lasting less than 260 days on average of 0.3218 assuming the mean life of the bulbs is 300 days.