|  |  |
| --- | --- |
| 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 |  |
| 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 outcome/Total no of outcome=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),(5,3),(5,4),(5,5),(5,6),(6,1),(6,2),(6,3),(6,4),(6,5),(6,6)}

N(s)=36

1. Equals to 1

n(E)= Total favourable outcomes=0

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

1. Less than or equals to 4

Possibility of getting number less than or equals to 4

A={(1,3),(2,2),(3,1)}

P(A)=n(A)/n(S)=3/36=1/12

P(C)= n(c)/n(s)=5/36

1. Sum is divided by 2 and 3

Total no of possible outcomes=36

Favourable outcomes= sum is divided by 2 and 3

Favourable outcomes={(1,5),(3,3),(4,2),(5,2),(6,6)=5

P(C)=n(c)/n(s)=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

Let S be a 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 two 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 no of candies for a randomly selected child:

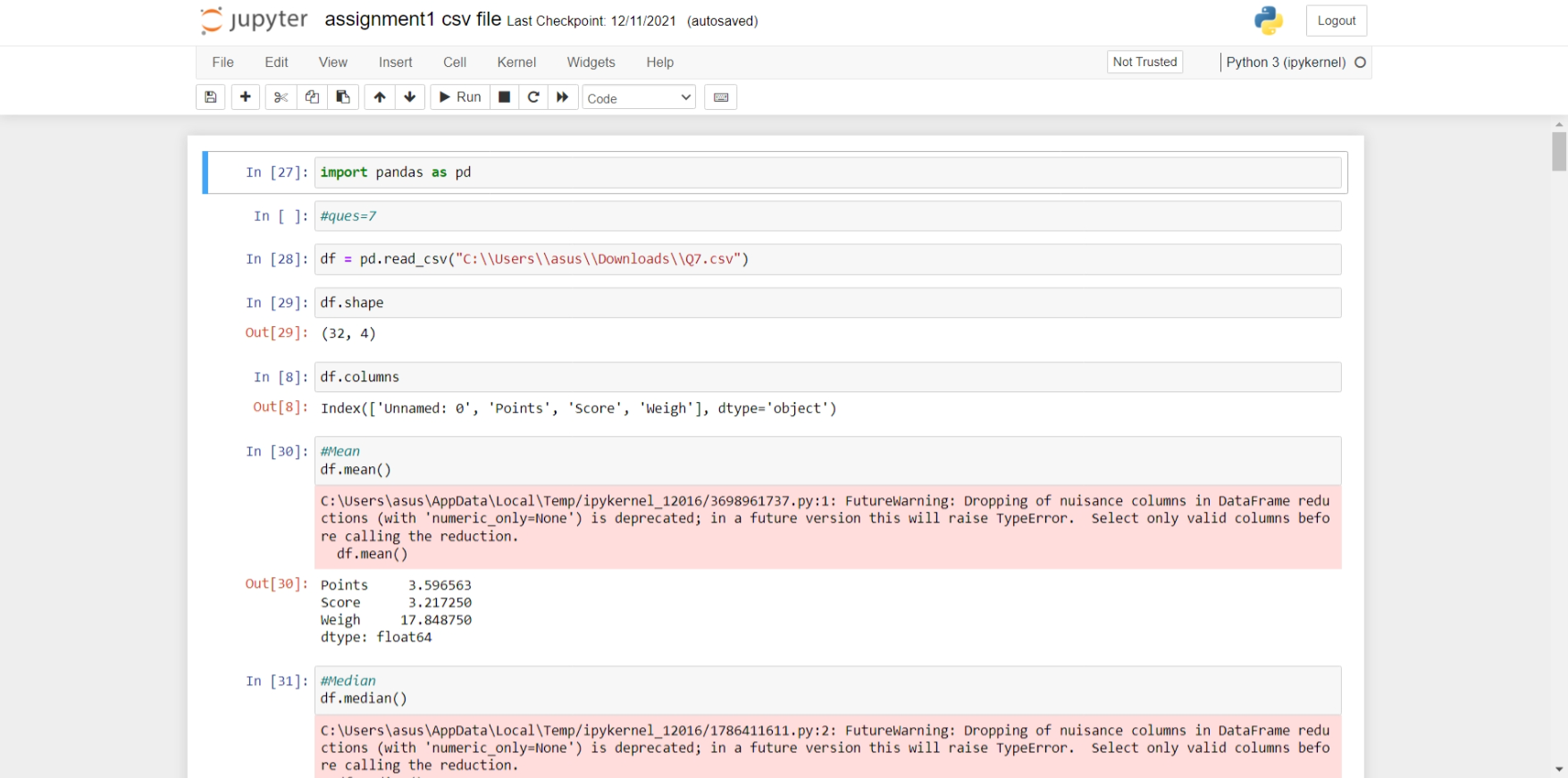
= 1\*0.015+4\*0.20+3\*0.65+5\*0.005+6\*0.01+2\*0.120 =3.09

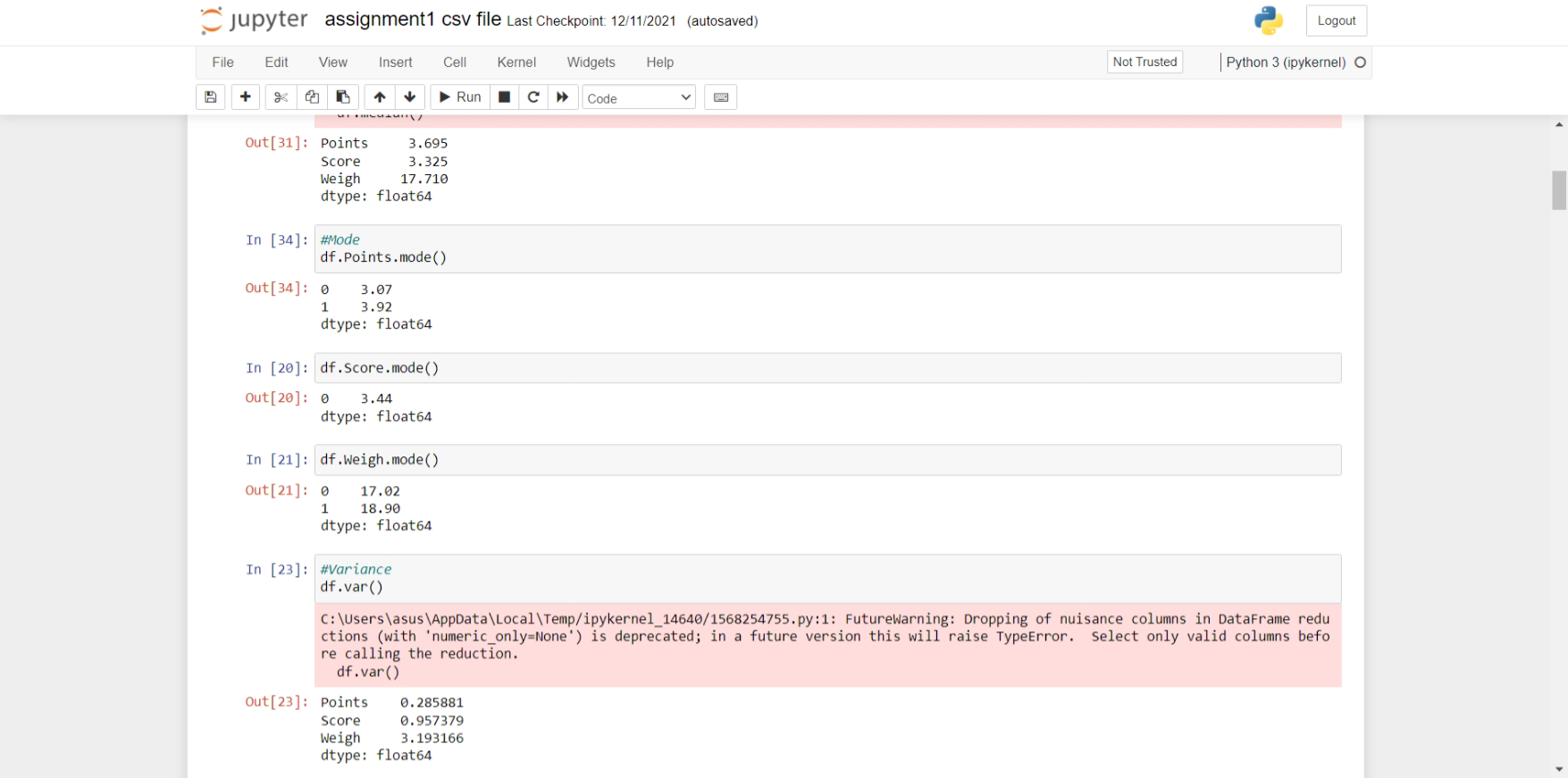
Q7) Calculate Mean, Median, Mode, Variance, Standard Deviation, Range & comment about the values / draw inferences, for the given dataset

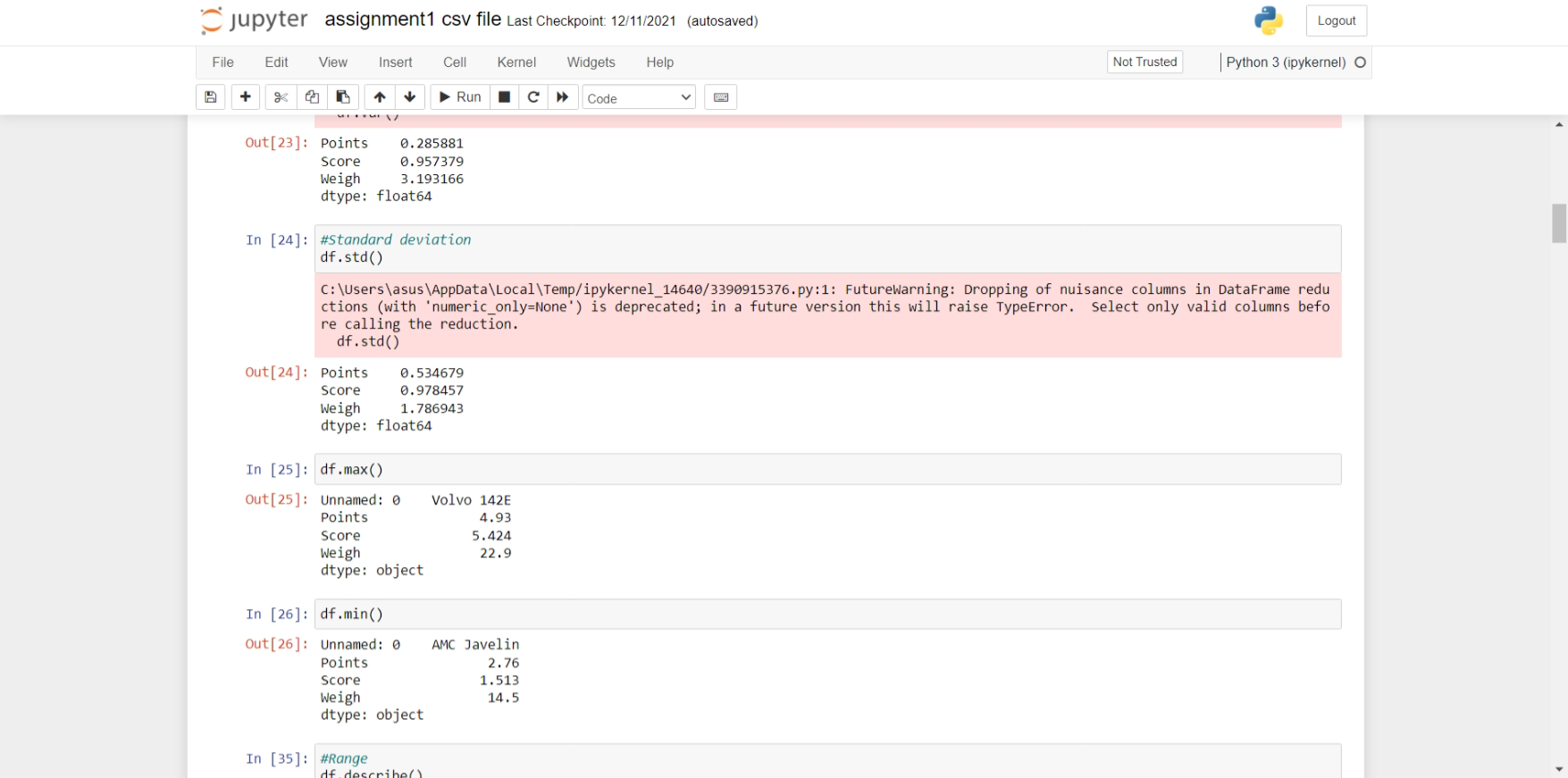
* For Points,Score,Weigh>

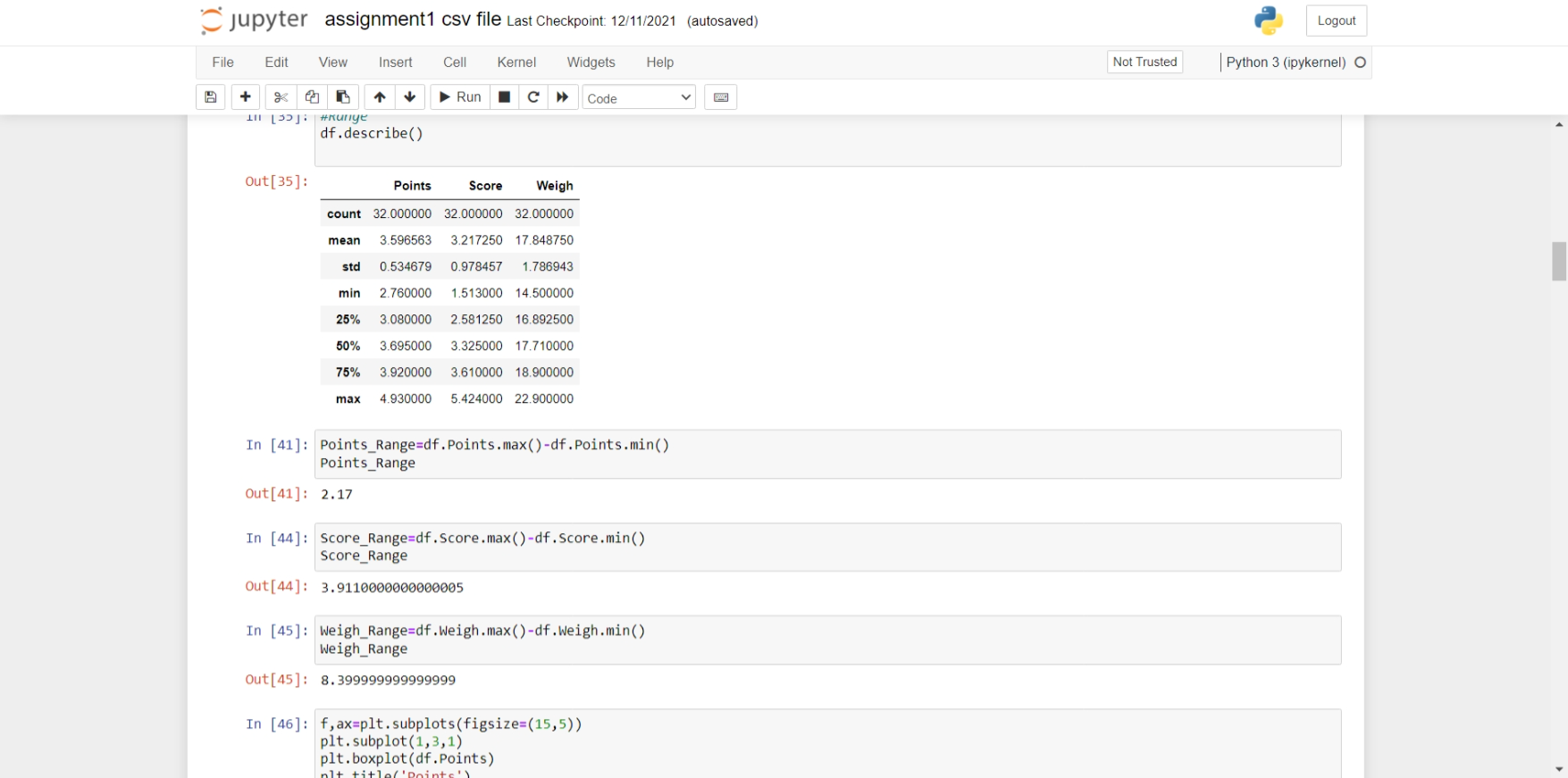
Find Mean, Median, Mode, Variance, Standard Deviation, and Range and also Comment about the values/ Draw some inferences.

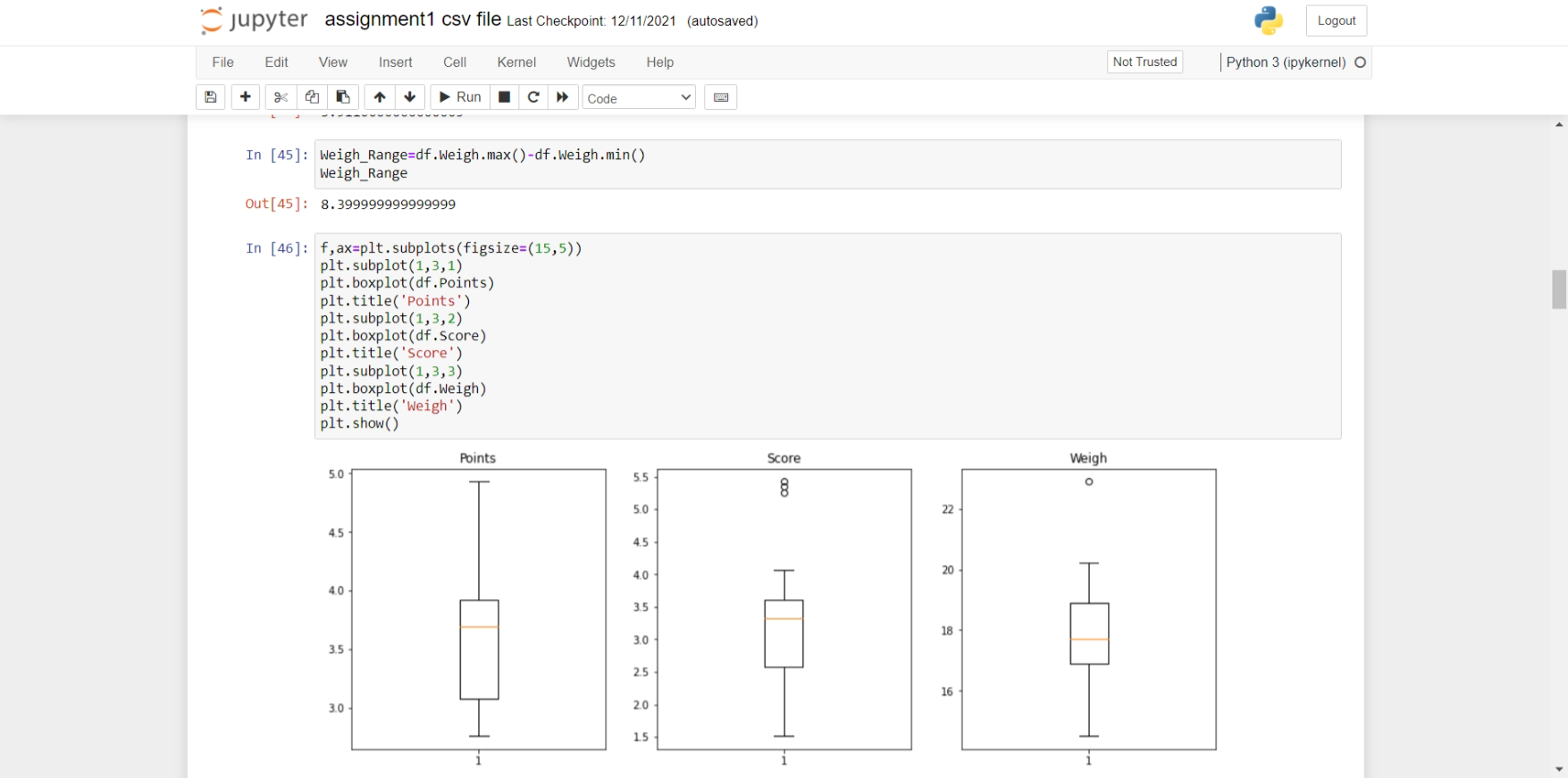
**Use Q7.csv file**

**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= Total no of observation = 9

Sum of total observation = 108+110+123+134+135+145+167+187+199

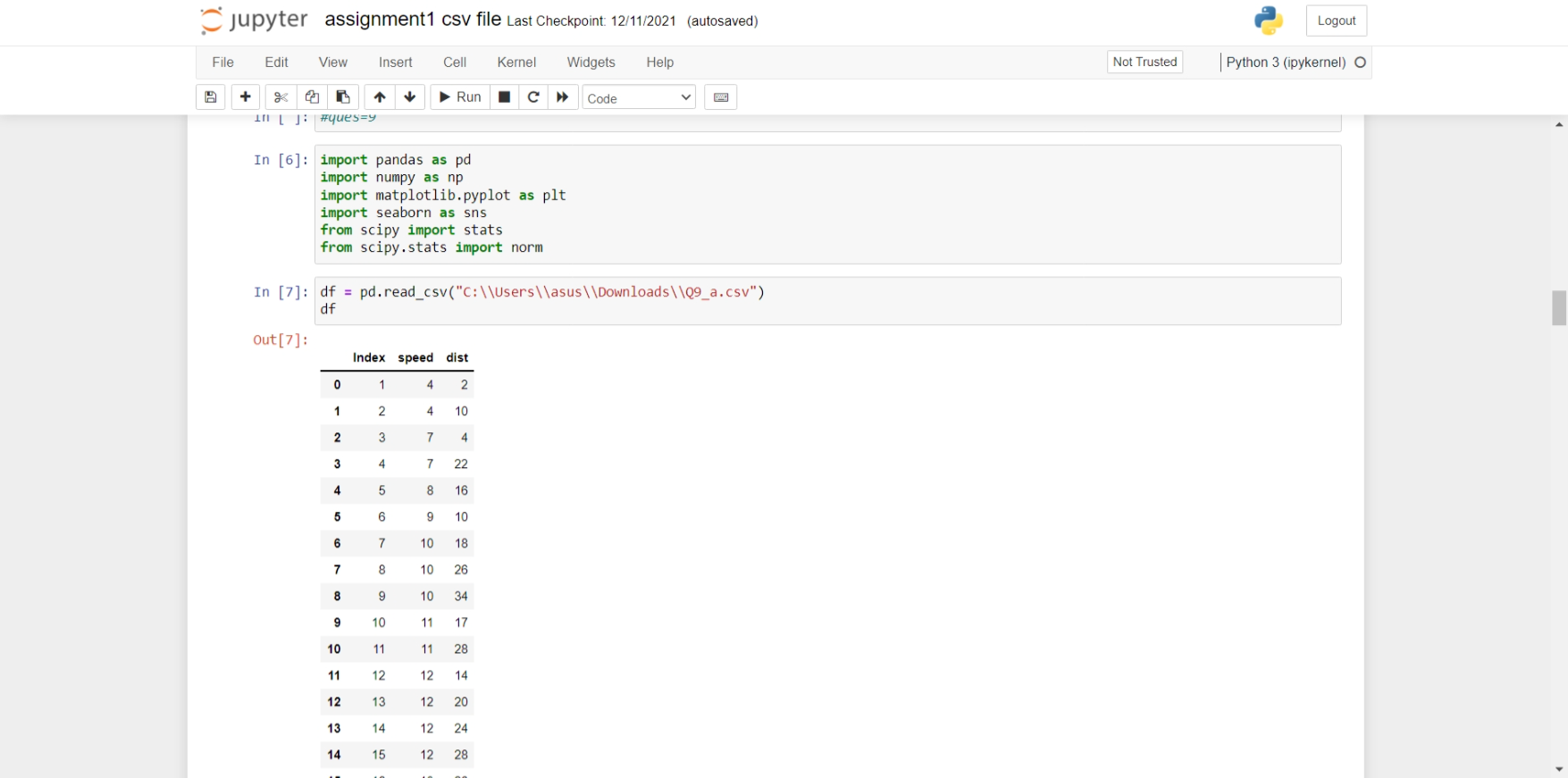
=1308

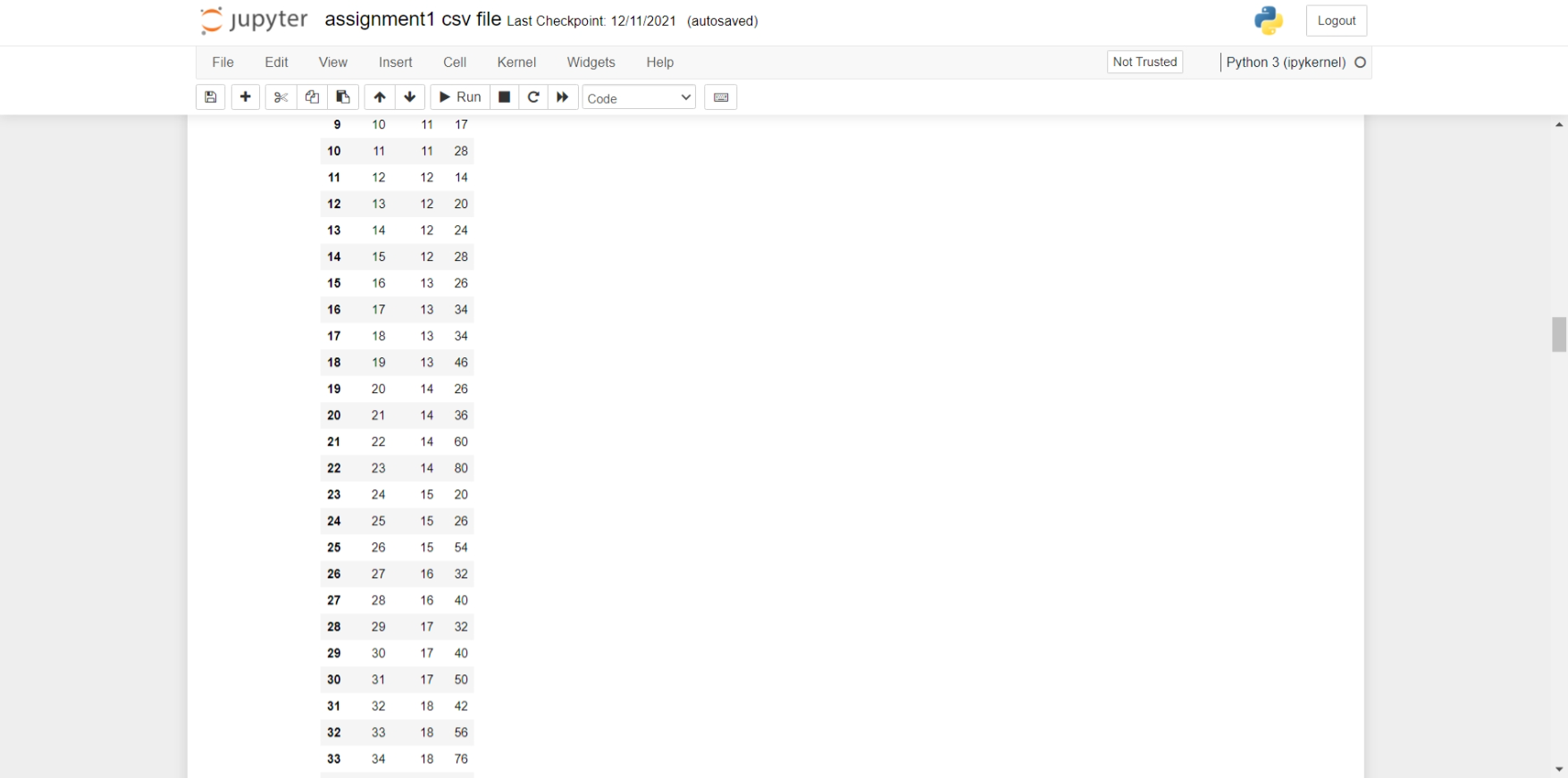
Expected value = 1308/9 = 145.333

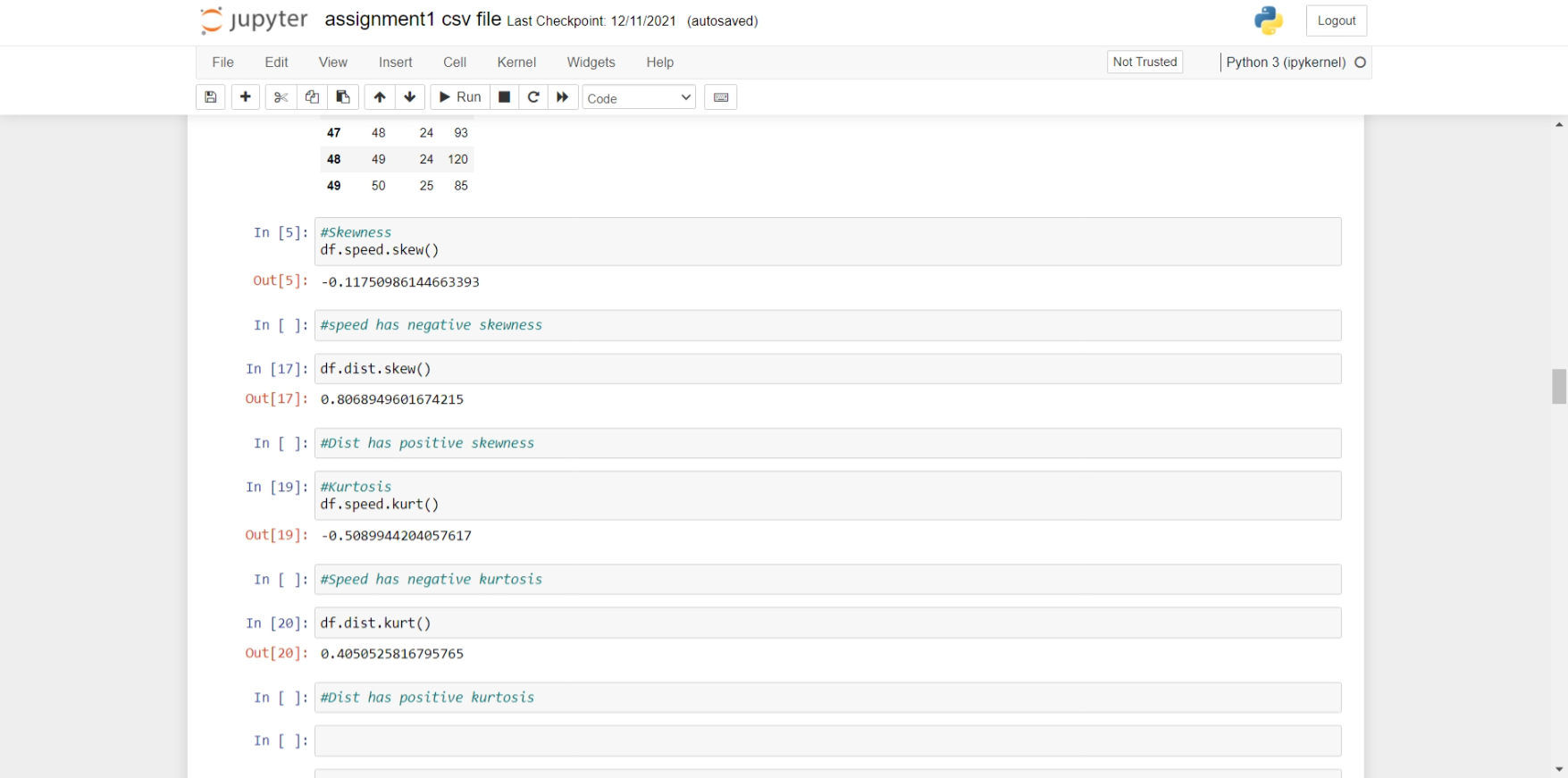
**Q9) Calculate Skewness, Kurtosis & draw inferences on the following data**

**Cars speed and distance**

**Use Q9\_a.csv**

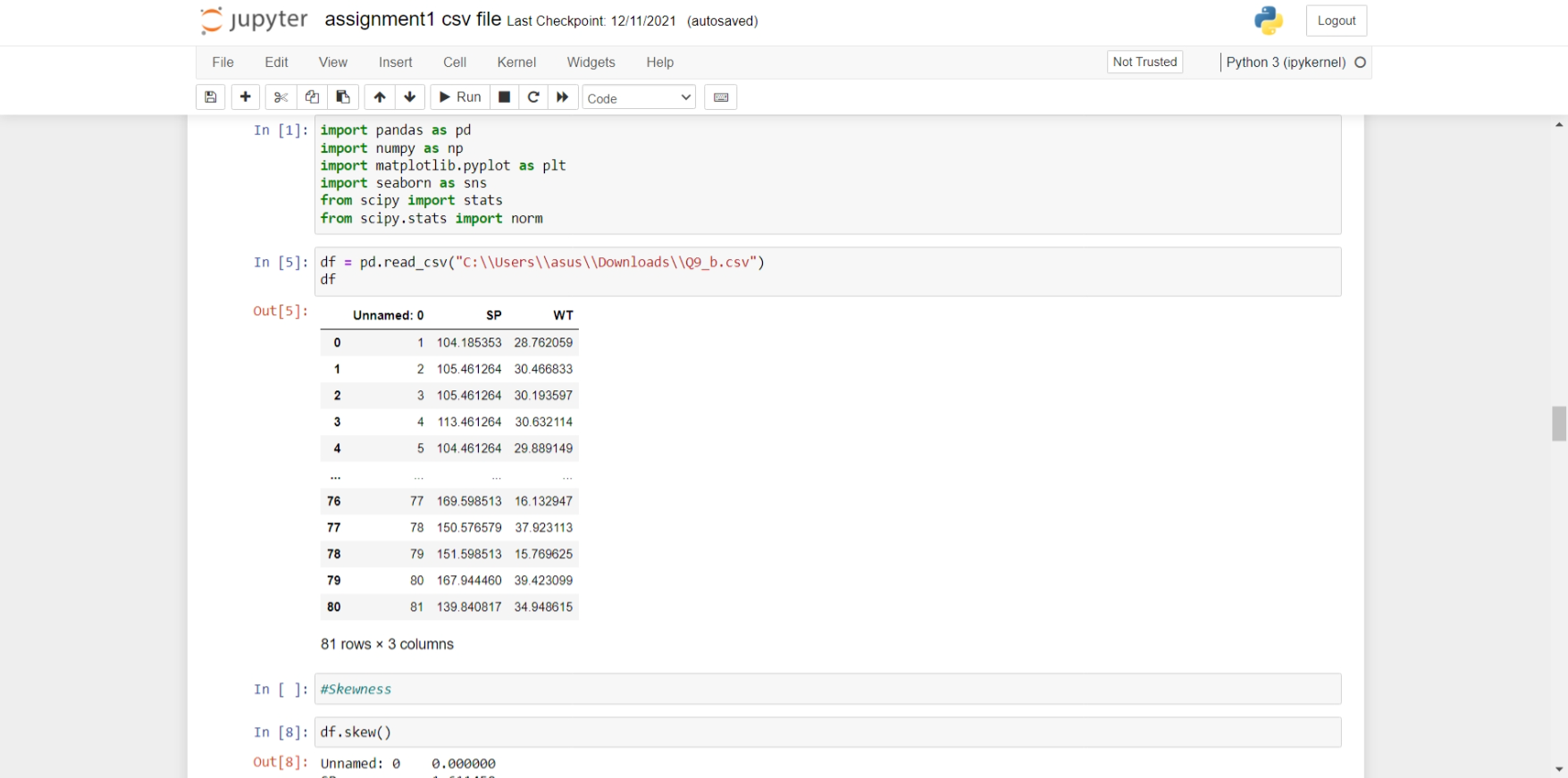
**Ans-**

****

****

**SP and Weight(WT)**

**Use Q9\_b.csv**

**Ans-**

****

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



Ans= The distribution is positively skewed.Mean>Median



Ans= The above distribution shows that the distribution has many 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?

Ans=(a) = 94% confidence interval

X-bar=200

Std=30

N=2000

Interval estimate = X-bar±z\*std/sqrt(n)

2000±1.88\*30/sqrt(2000)

=(198.74/201.26)

(b)= 98% confidence interval

X-bar=200

Std=30

n=2000

Interval estimate=X-bar±z\*std/sqrt(n)

2000±2.33\*30/sqrt(2000)

=(198.44/201.56)

(c)=96% confidence interval

X-bar=2000

Std=30

n=2000

Interval estimate=x-bar±z\*std/sqrt(n)

2000±2.05\*30/sqrt(2000)

=(198.6/201.38)

**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

Toward 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

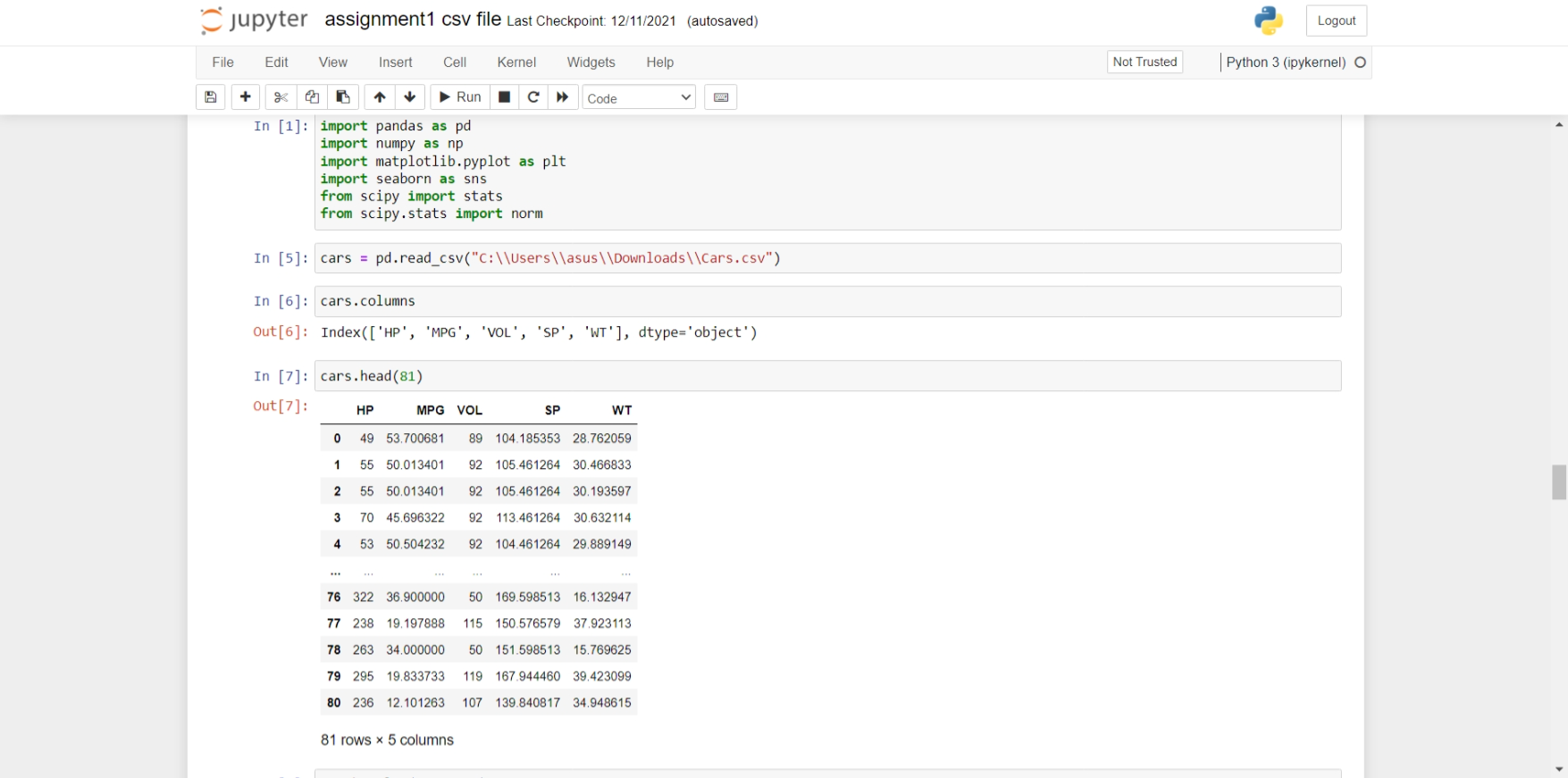
Data \_set: Cars.csv

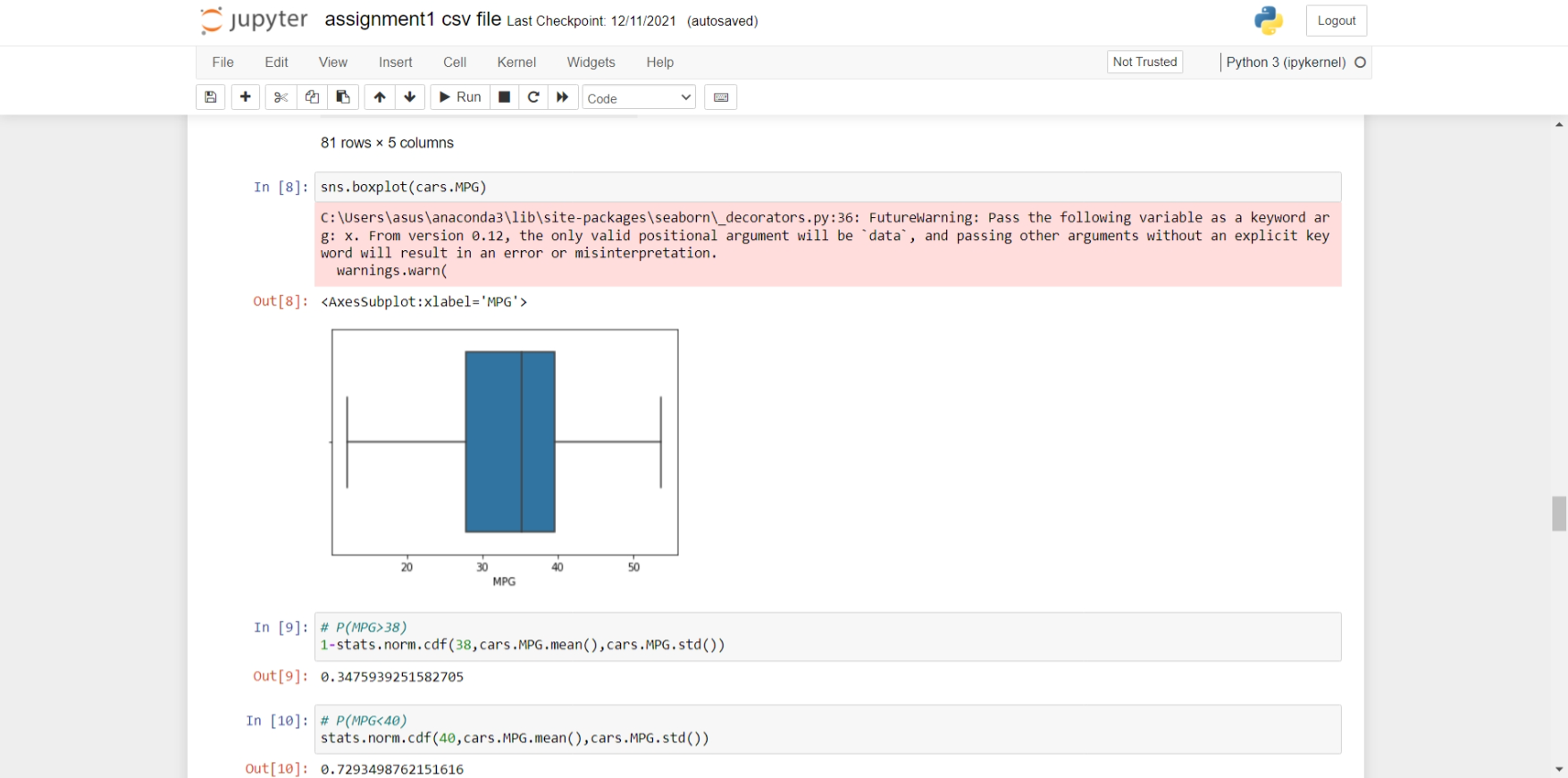
Calculate the probability of MPG of Cars for the below cases.

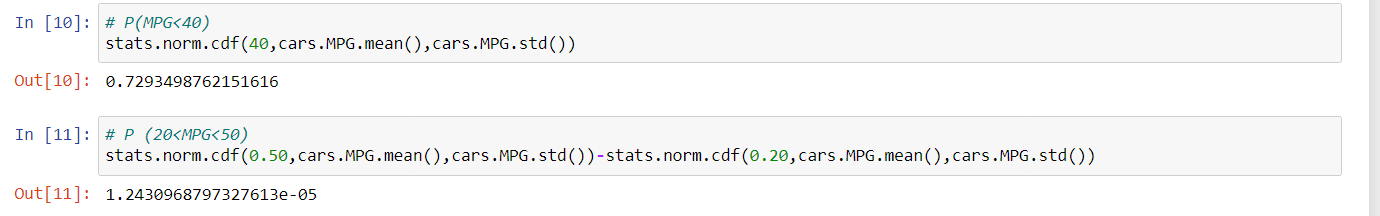
MPG <- Cars$MPG

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

c. P (20<MPG<50)

Ans-

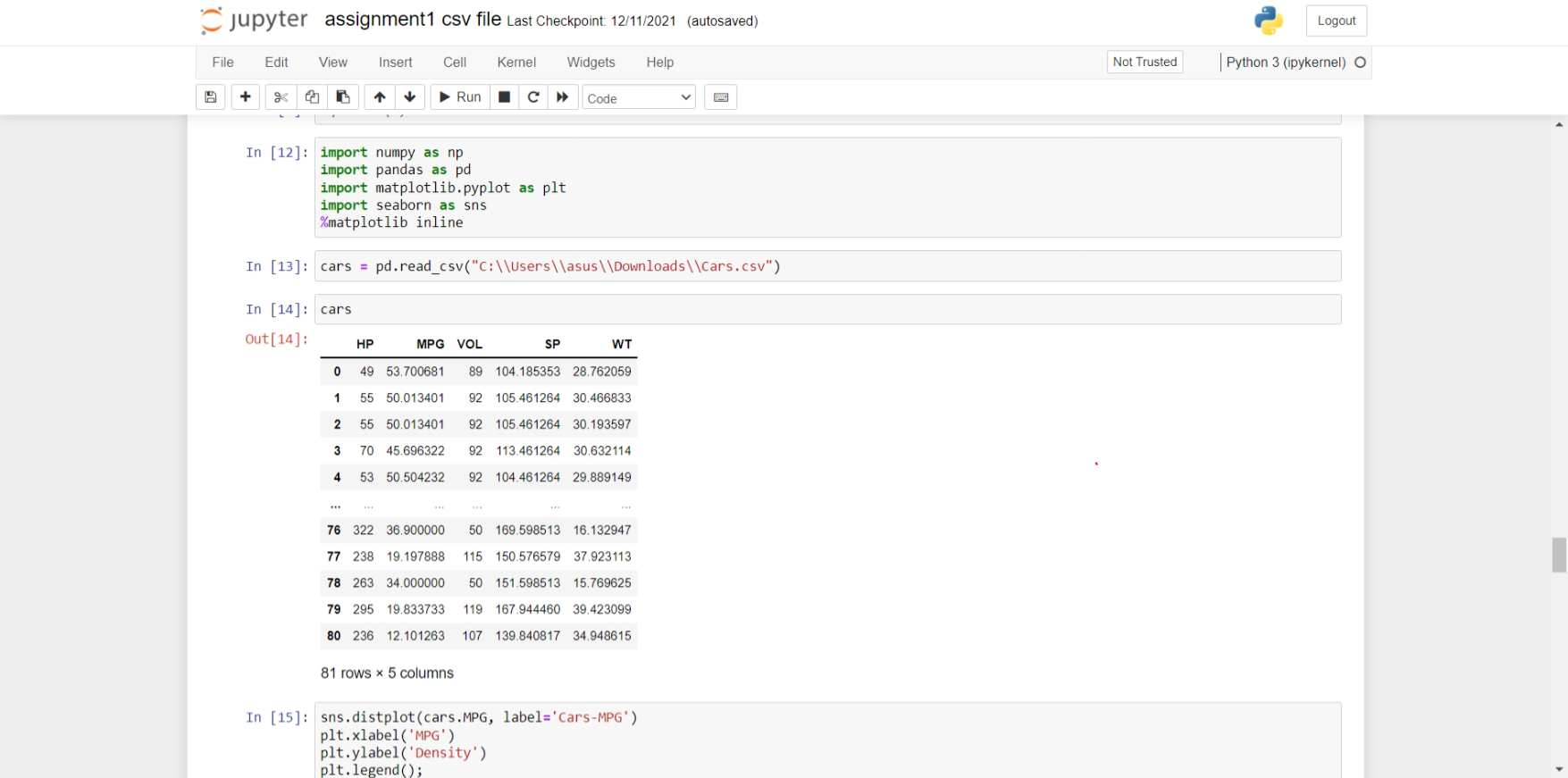


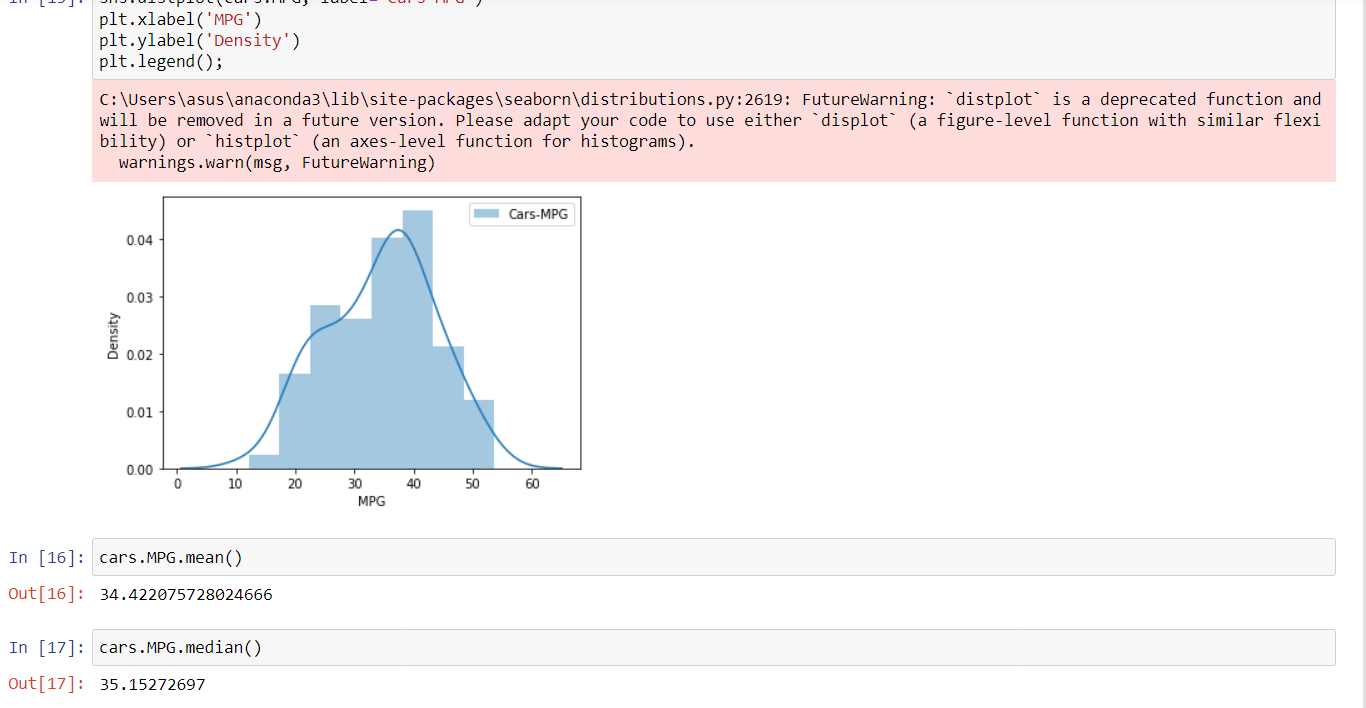


Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

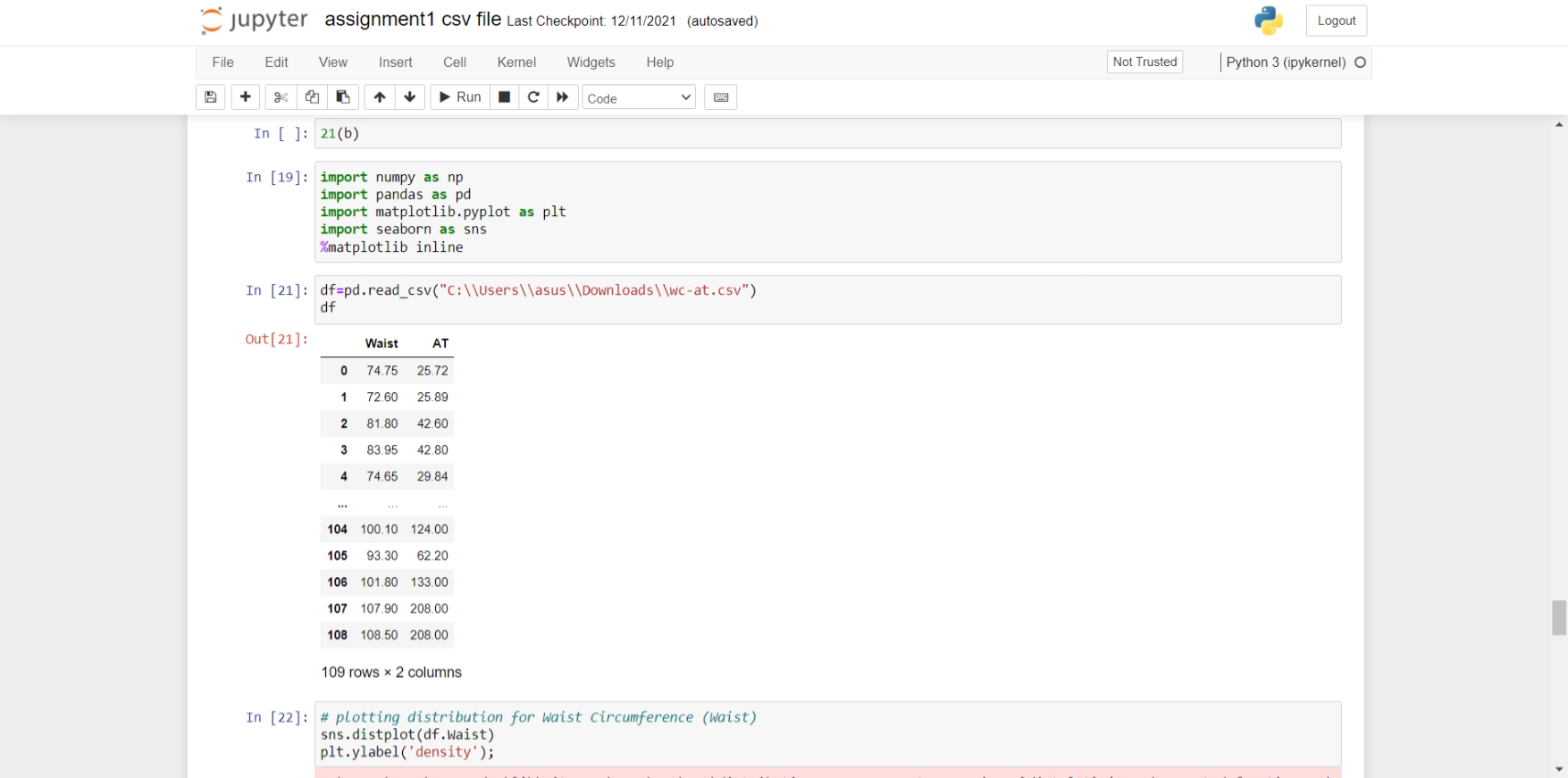
Dataset: Cars.csv

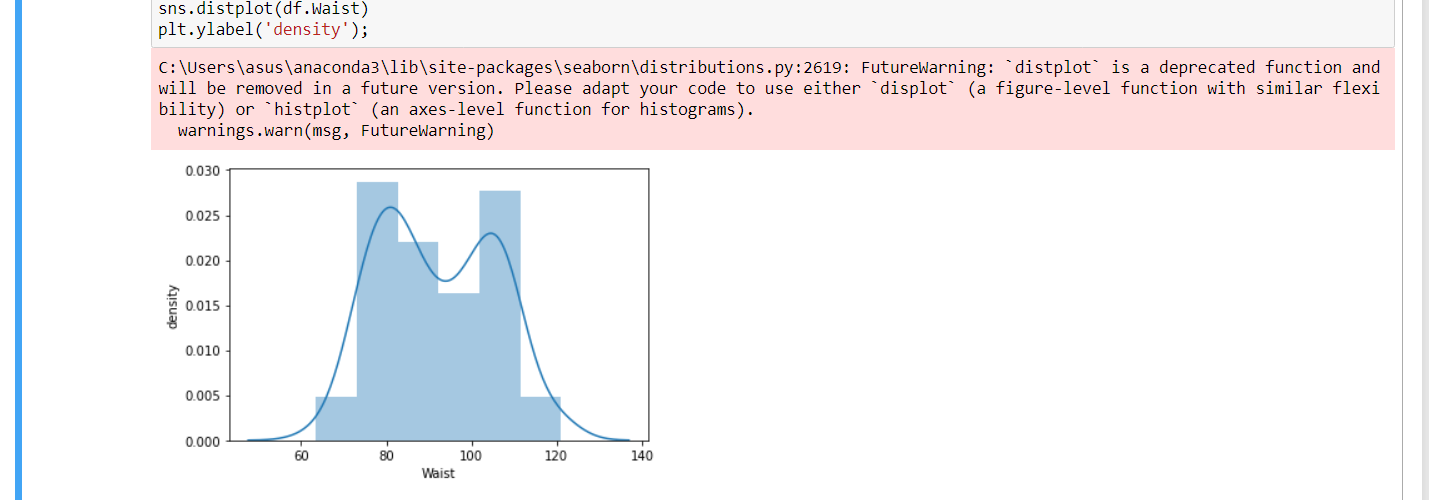
Ans-

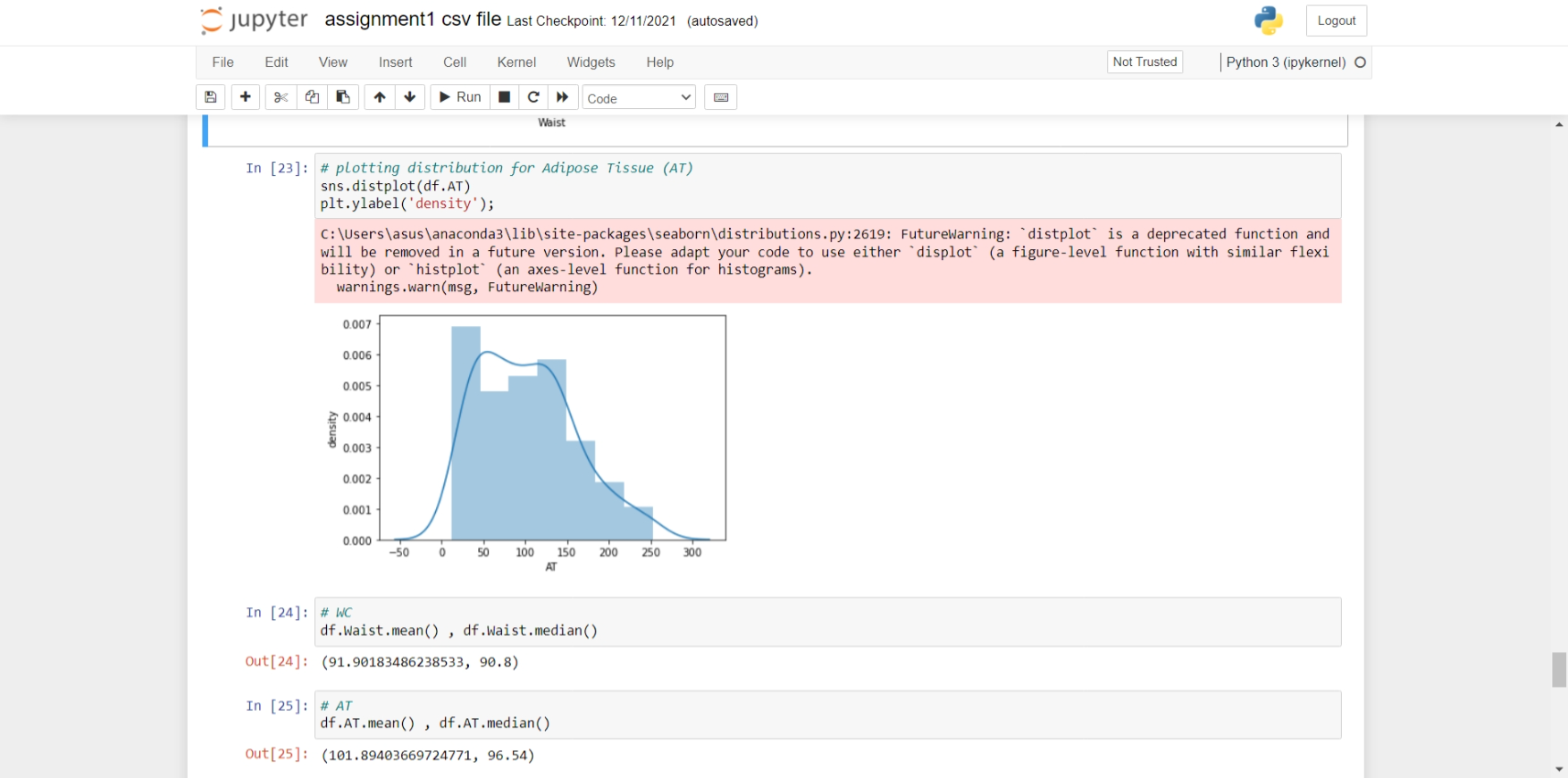


1. Check Whether the Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set follows Normal Distribution

Dataset: wc-at.csv

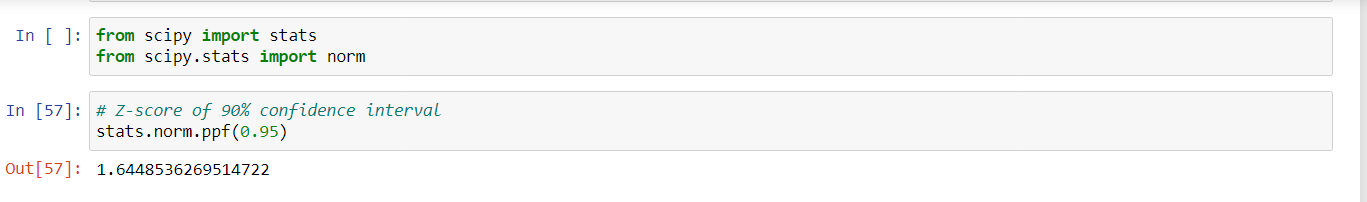
Ans-





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

Ans = (a) Z scores of 90% confidence interval=0.95



(b) Z scores of 94% confidence interval=0.97



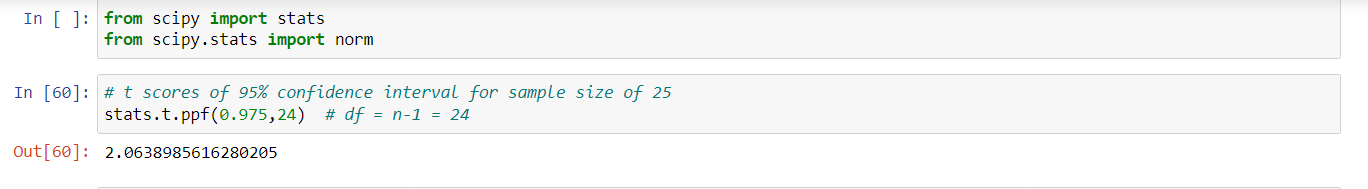
(c) Z scores of 60% confidence interval=0.8



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

Ans=Degree of freedom= n-1=24

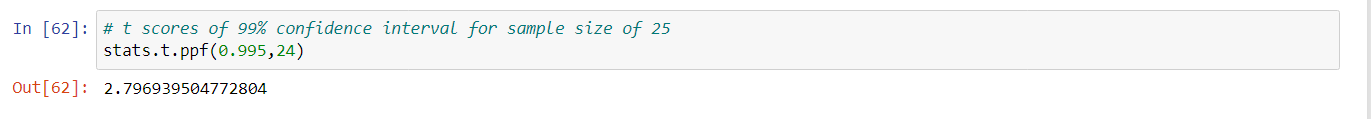
t scores of 95% confidence interval=(0.975,24)



t scores of 96% confidence interval=(0.98,24)



t scores of 99% confidence interval=(0.995,24)



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

Ans= µ=270,n=18, x-bar=260,s=90

tscore=(x-bar-µ)/(s/sqrt(n))

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

=-10/21.23

=-0.4714052

