| 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 | Ordinal |
| Level of Agreement | Ordinal |
| IQ(Intelligence Scale) | Ratio |
| Sales Figures | Interval |
| Blood Group | Nominal |
| Time Of Day | Ratio |
| Time on a Clock with Hands | Ratio |
| Number of Children | Ordinal |
| Religious Preference | Nominal |
| Barometer Pressure | Ratio |
| SAT Scores | Ratio |
| Years of Education | Interval |

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

Ans: The number of possible outcomes

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

The probability of getting Two heads and One tail is 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: Two dices were rolled, total possible cases =36

## a) Total Favourable cases (Having sum =1) = 0

## As minimum sum is 2 for outcome (1,1).

## Probability = 0/36

=0

b) Total Favourable cases = 6

Probabilty = 6/36

= 0.16667

c) Total Favourable cases = 5

Probabilty = 5/36

= 0.1388

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: Sample space = 7c2 = 21.

Let event E be none of the balls is blue = all balls are either red or green or both.

n(E) = 5c2

=10 p(E)

=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

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 \* 0.12

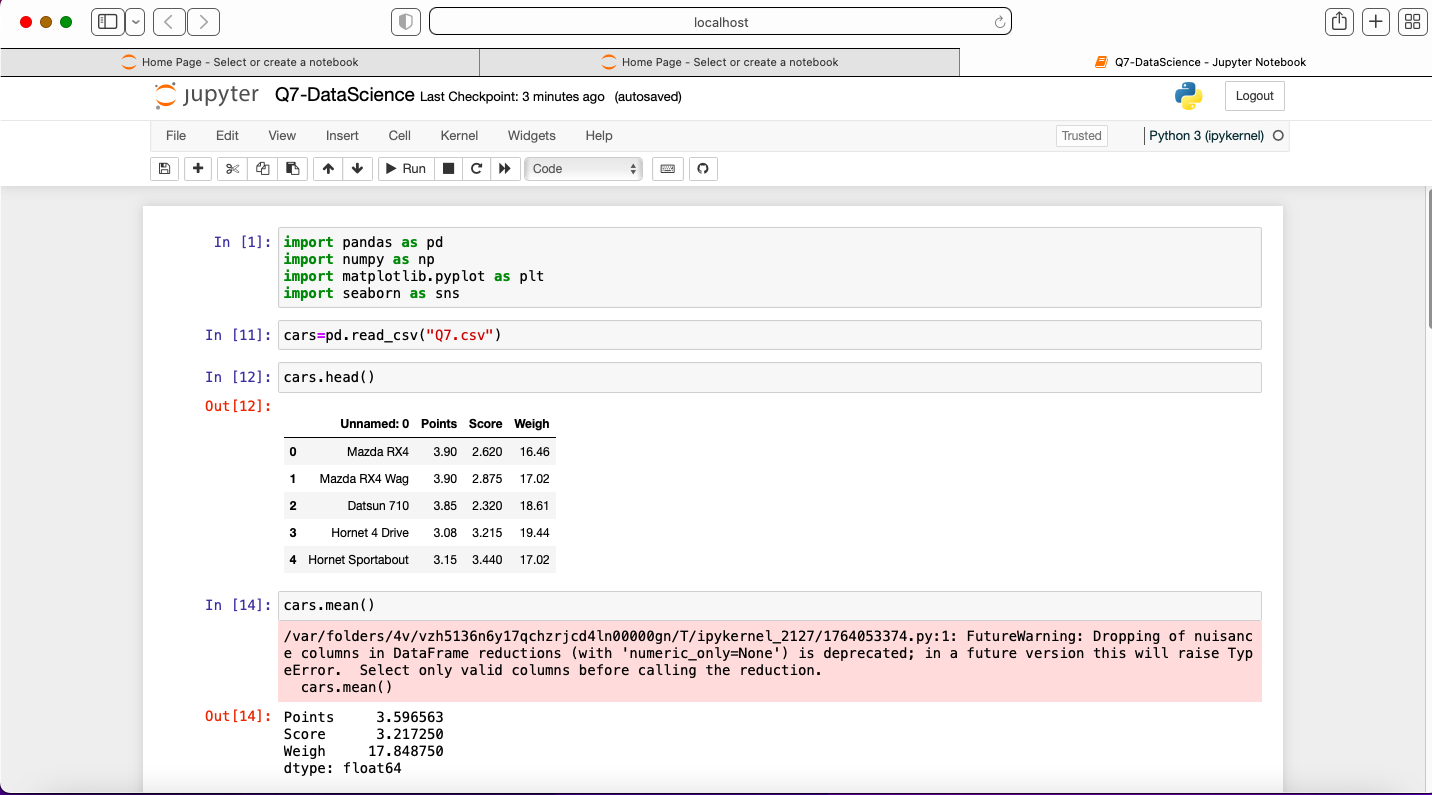
= 0.015 + 0.8 + 1.95 + 0.025 + 0.06 + 0.24

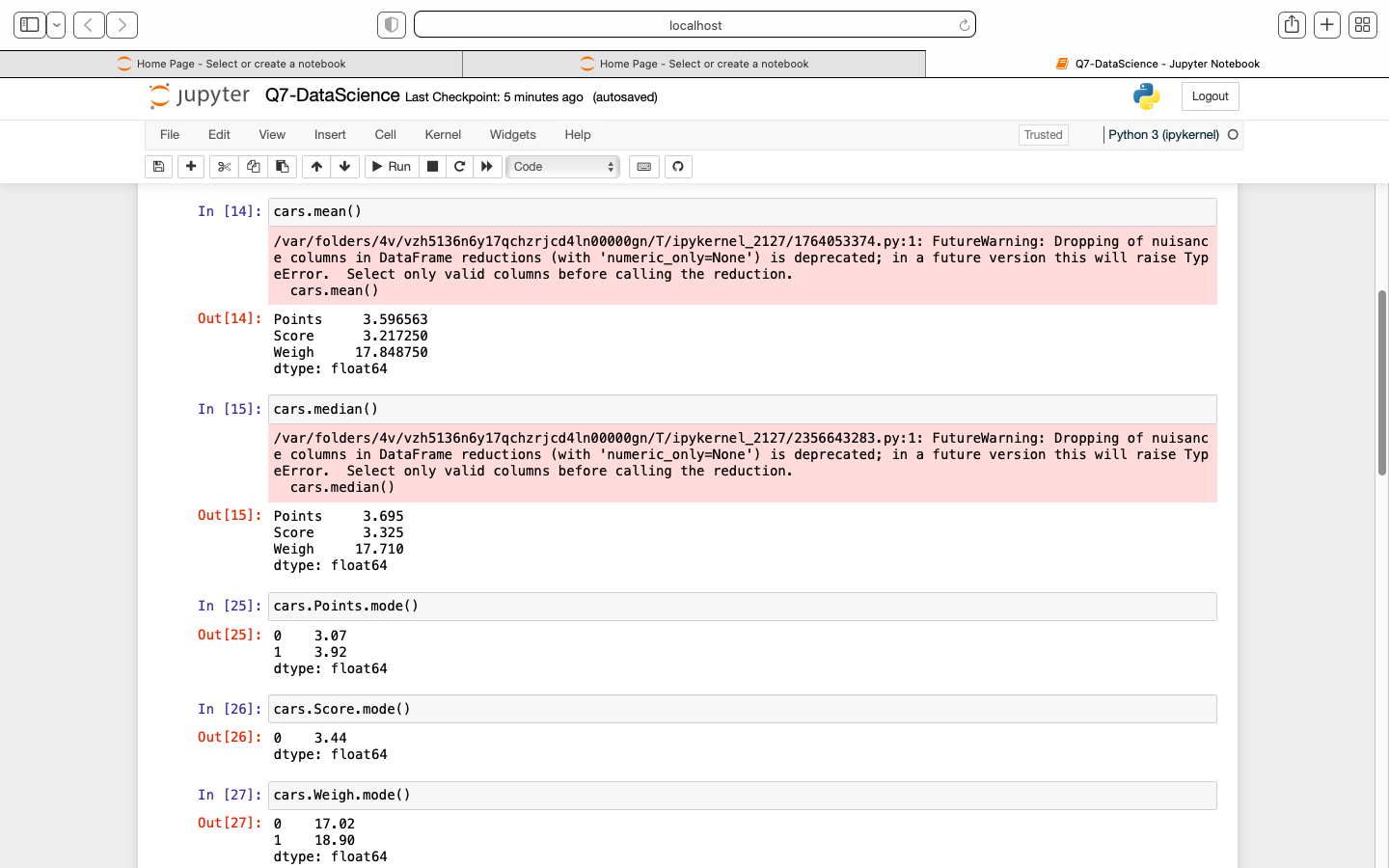
= 3.09

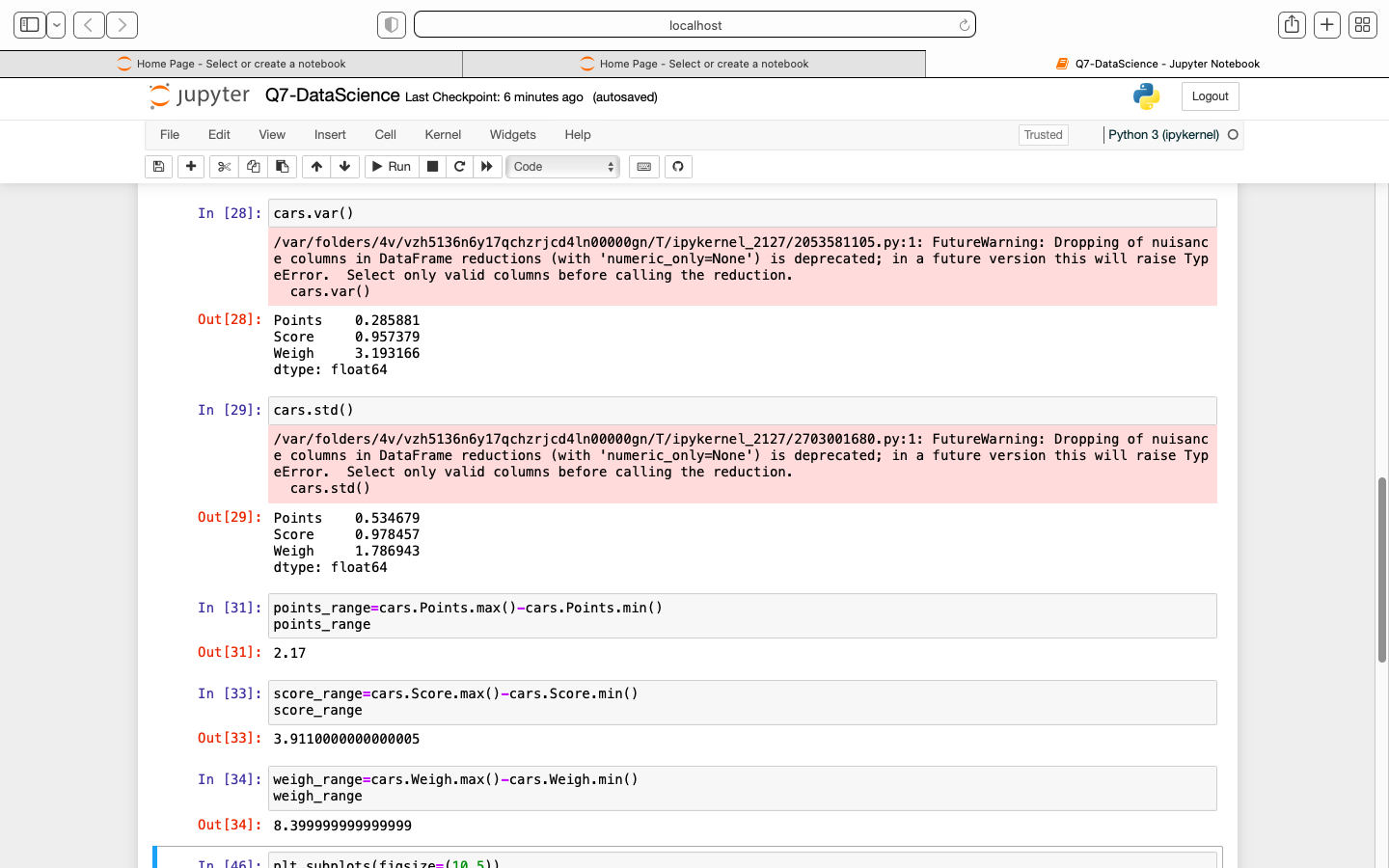
Q7) Calculate Mean, Median, Mode, Variance, Standard Deviation, Range & comment about the values / draw inferences, for the given dataset

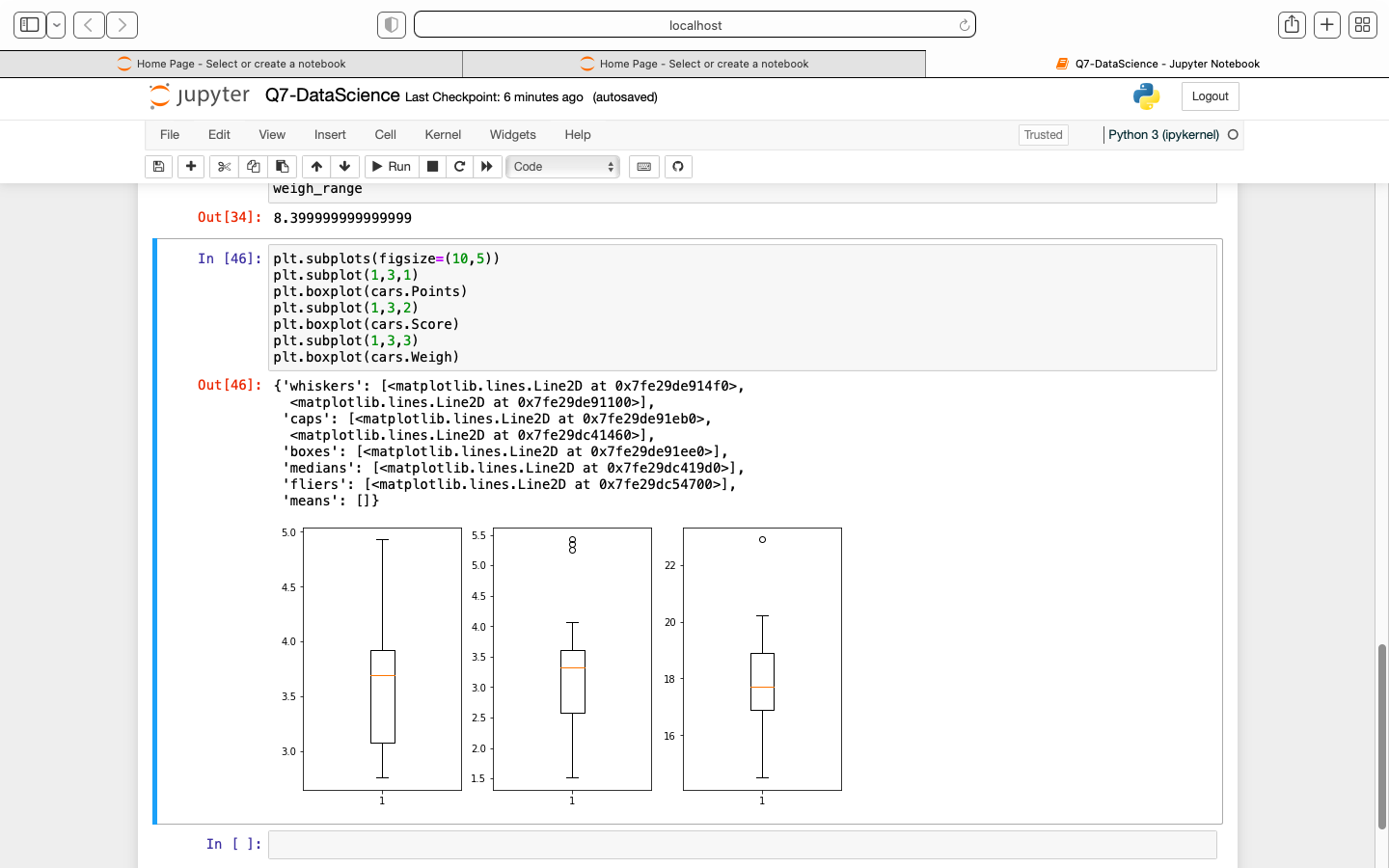
* For Points,Score,Weigh>

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

Ans- 







**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- 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

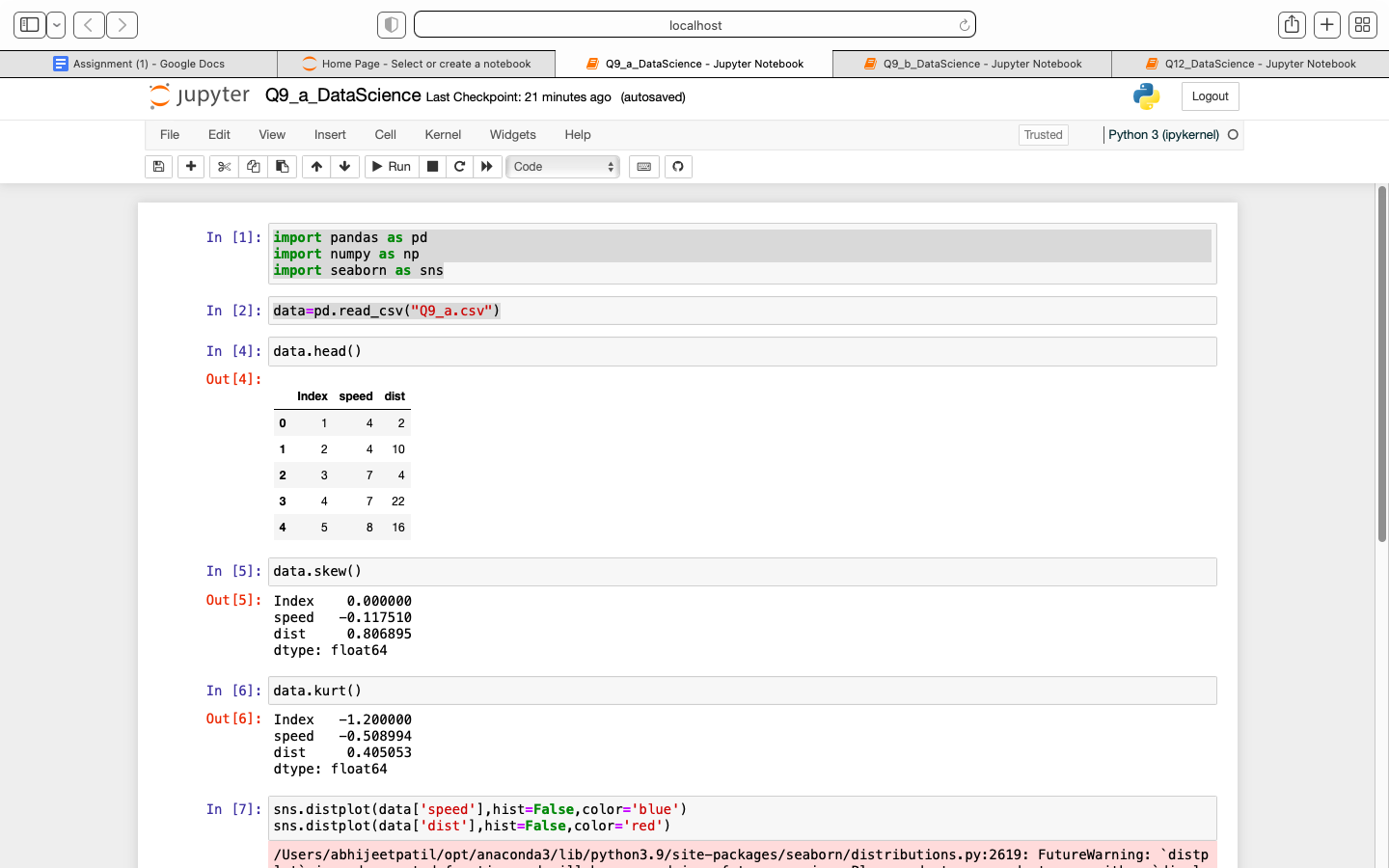
= (1/9) ( 108 + 110 + 123 + 134 + 135 + 145 + 167 + 187 + 199)

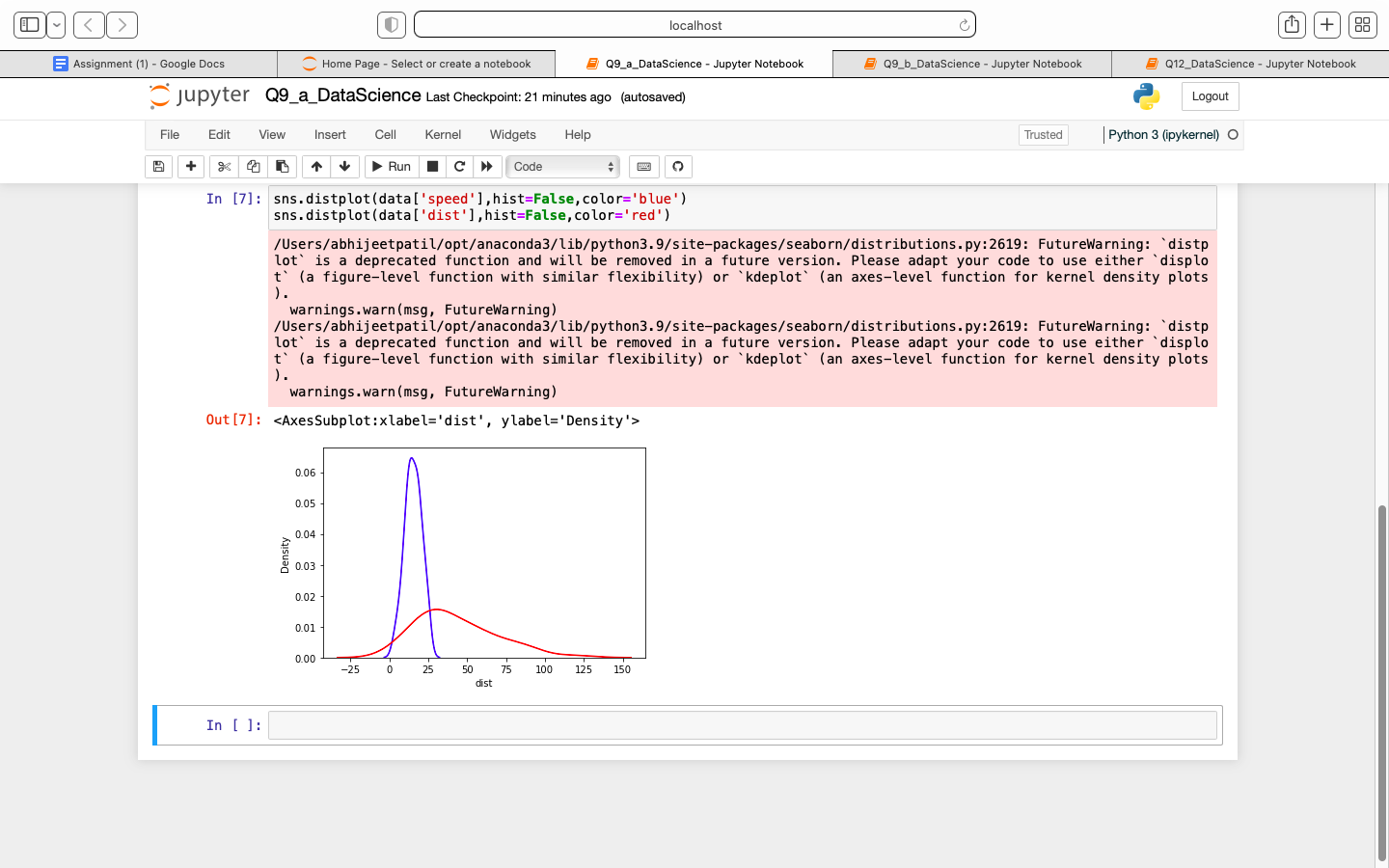
= 145.33

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

**Cars speed and distance**

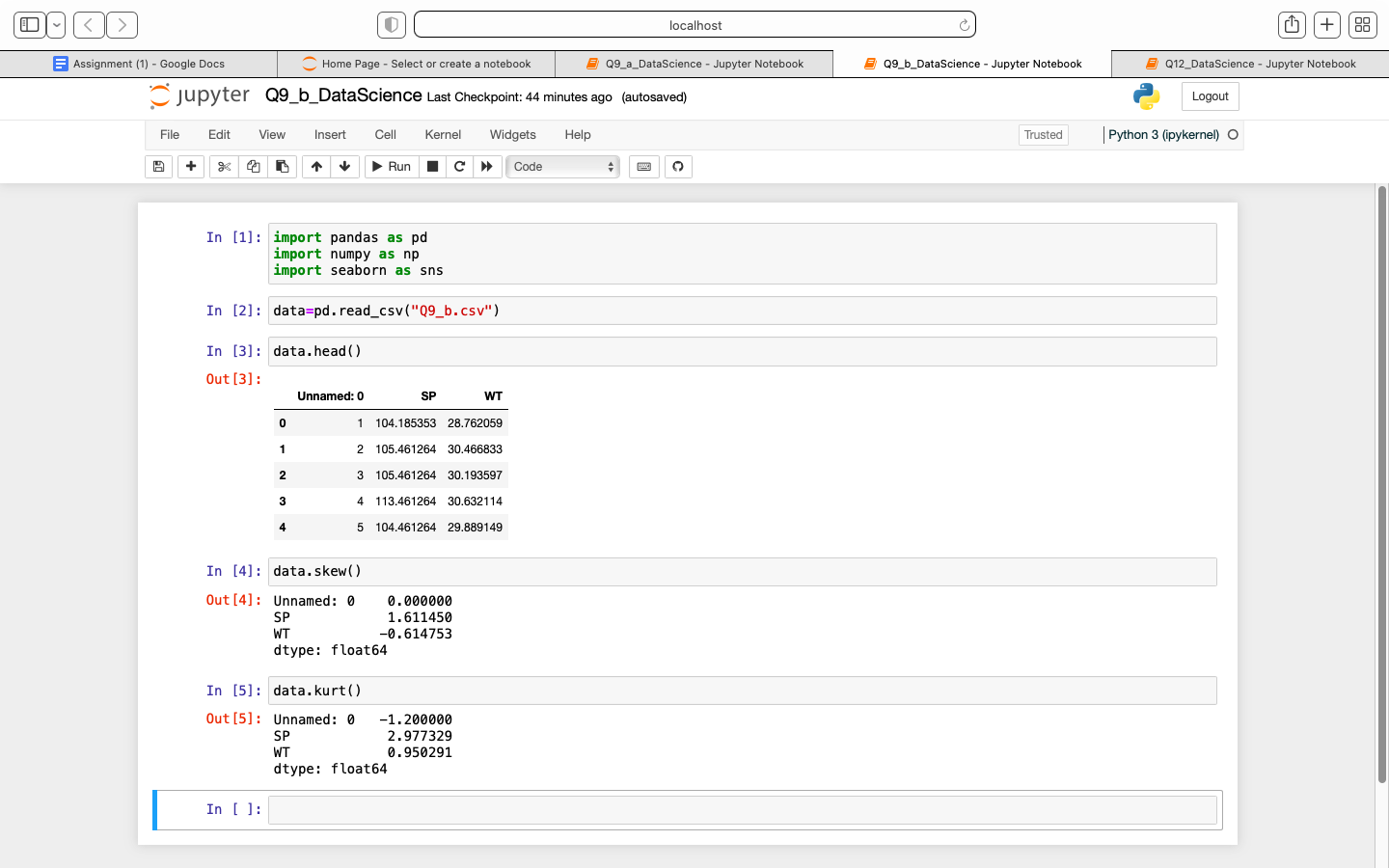
**Use Q9\_a.csv**

****

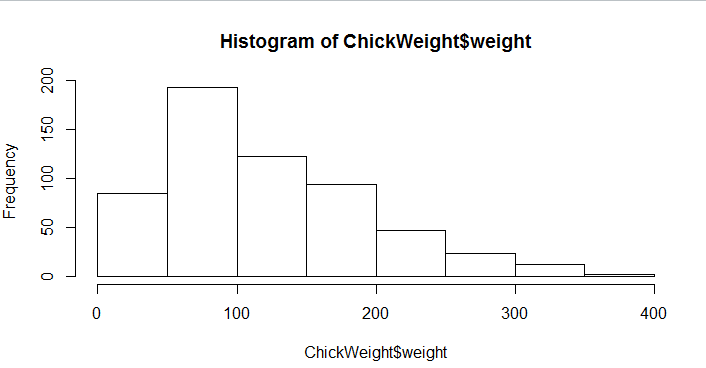
****

**SP and Weight(WT)**

**Use Q9\_b.csv**

****

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



Ans- Histogram:

- Chick weight data is right skewed or positively skewed.

- More than 50% Chick Weight is between 50 to 150.

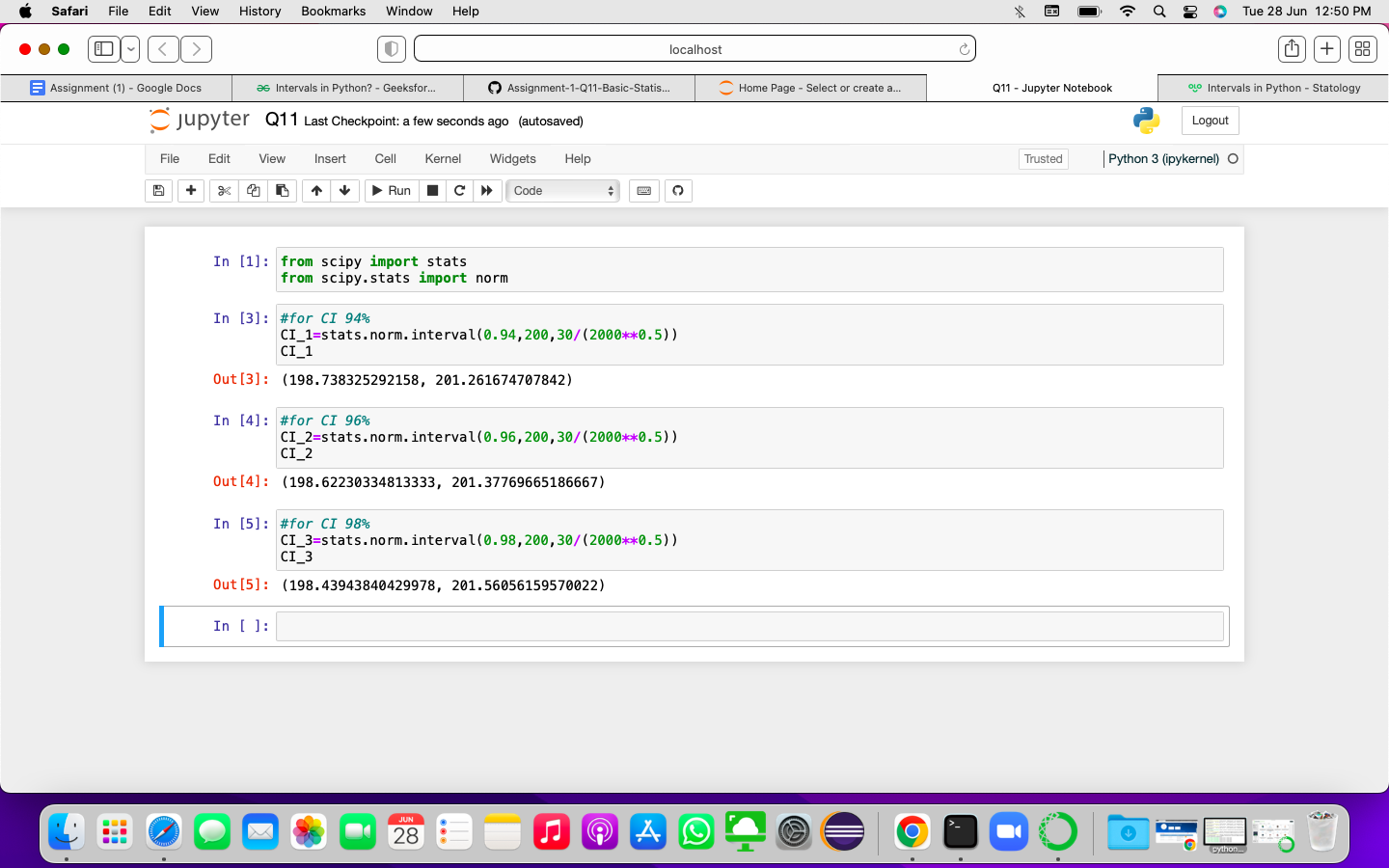
- Most of the chick weight is between 50 to 100.



Boxplot - The data is right skewed.

- There are outliers at upper side.

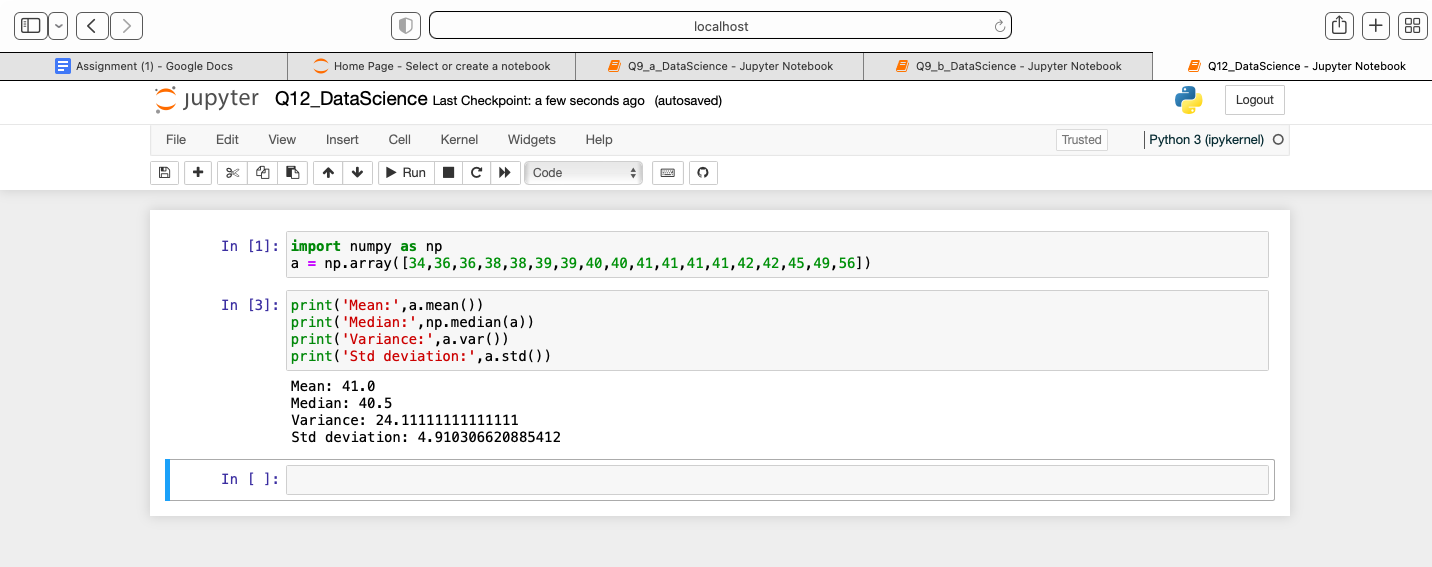
**Q11)** Suppose we want to estimate the average weight of an adult male in Mexico. We draw a random sample of 2,000 men from a population of 3,000,000 men and weigh them. We find that the average person in our sample weighs 200 pounds, and the standard deviation of the sample is 30 pounds. Calculate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

Ans- 

**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?

Ans- 

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

Ans- If the mean is equal to the median as well as the mode, hence the skewness is zero.

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

Ans- If the mean is greater than the median, the distribution is positively skewed.

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

Ans- If the mean is less than the median, the distribution is negatively skewed.

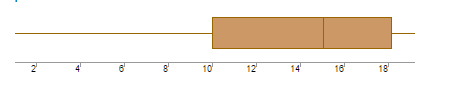
Q16) What does positive kurtosis value indicates for a data ?

Ans- Positive values of kurtosis indicate that distribution is peaked and possesses thick tails. An extreme positive kurtosis indicates a distribution where more of the numbers are located in the tails of the distribution instead of around the mean.

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

Ans- A negative kurtosis means that your distribution is flatter than a normal curve with the same mean and standard deviation.

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



What can we say about the distribution of the data?

Ans- The median is closer to the top of the box, and if the whisker is shorter on the upper end of the box, then the distribution is negatively skewed

What is nature of skewness of the data?

Ans- When the median is closer to the top of the box, and if the whisker is shorter on the upper end of the box, then the distribution is negatively skewed.

Hence above boxplot is negatively skewed.

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

Ans- The IQR of the data is approximately

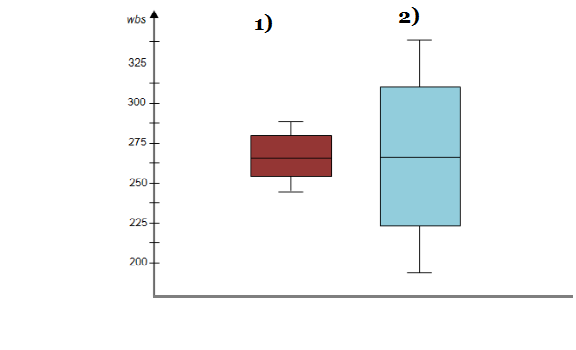
Q1= 18

Q3= 10

IQR= 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- By Observing Box plots, we can say that the mean of both plots are same.

Quartile range for box plot 1 is less as compared to box plot 2.

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

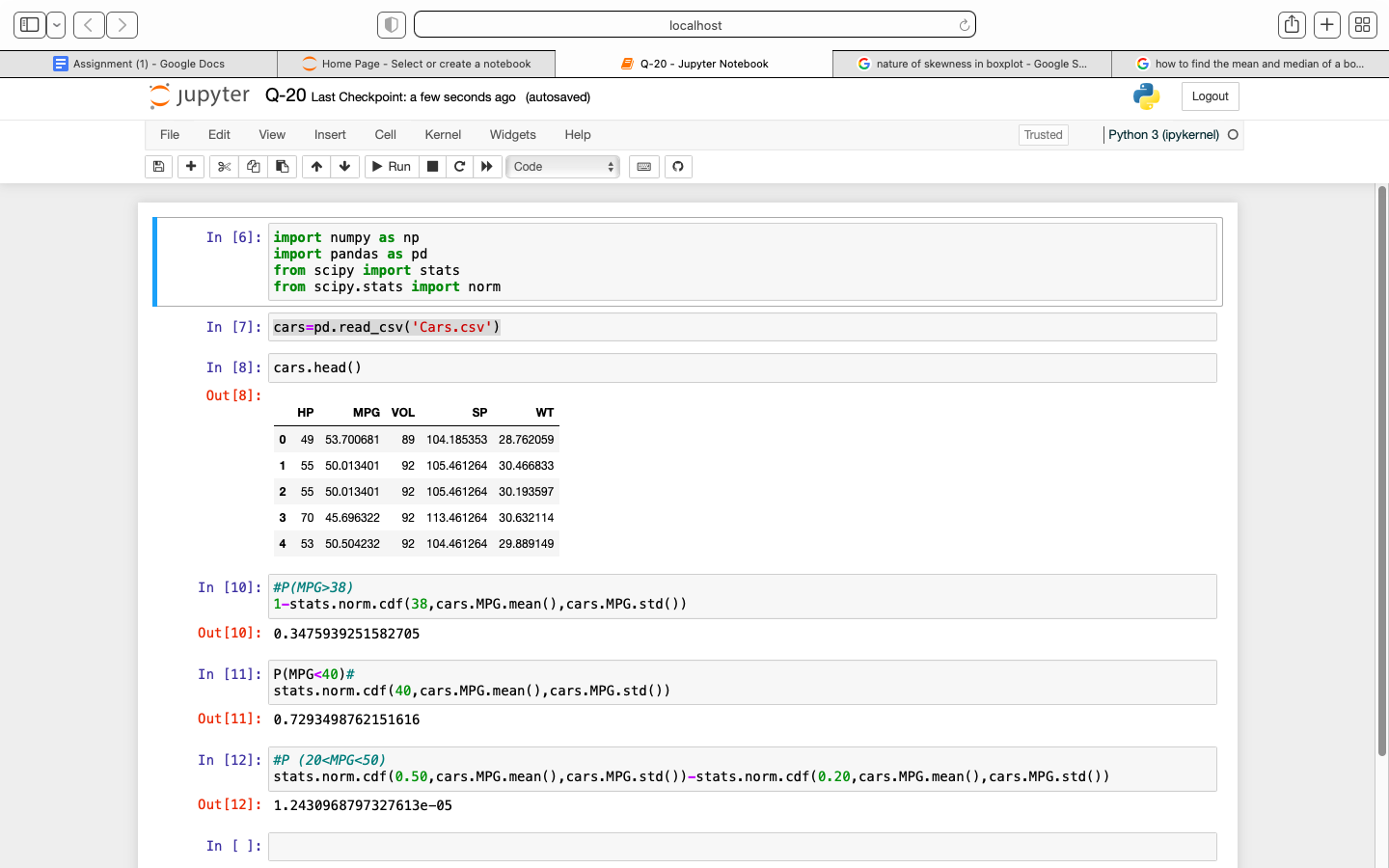
Data \_set: Cars.csv

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

MPG <- Cars$MPG

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

c. P (20<MPG<50)

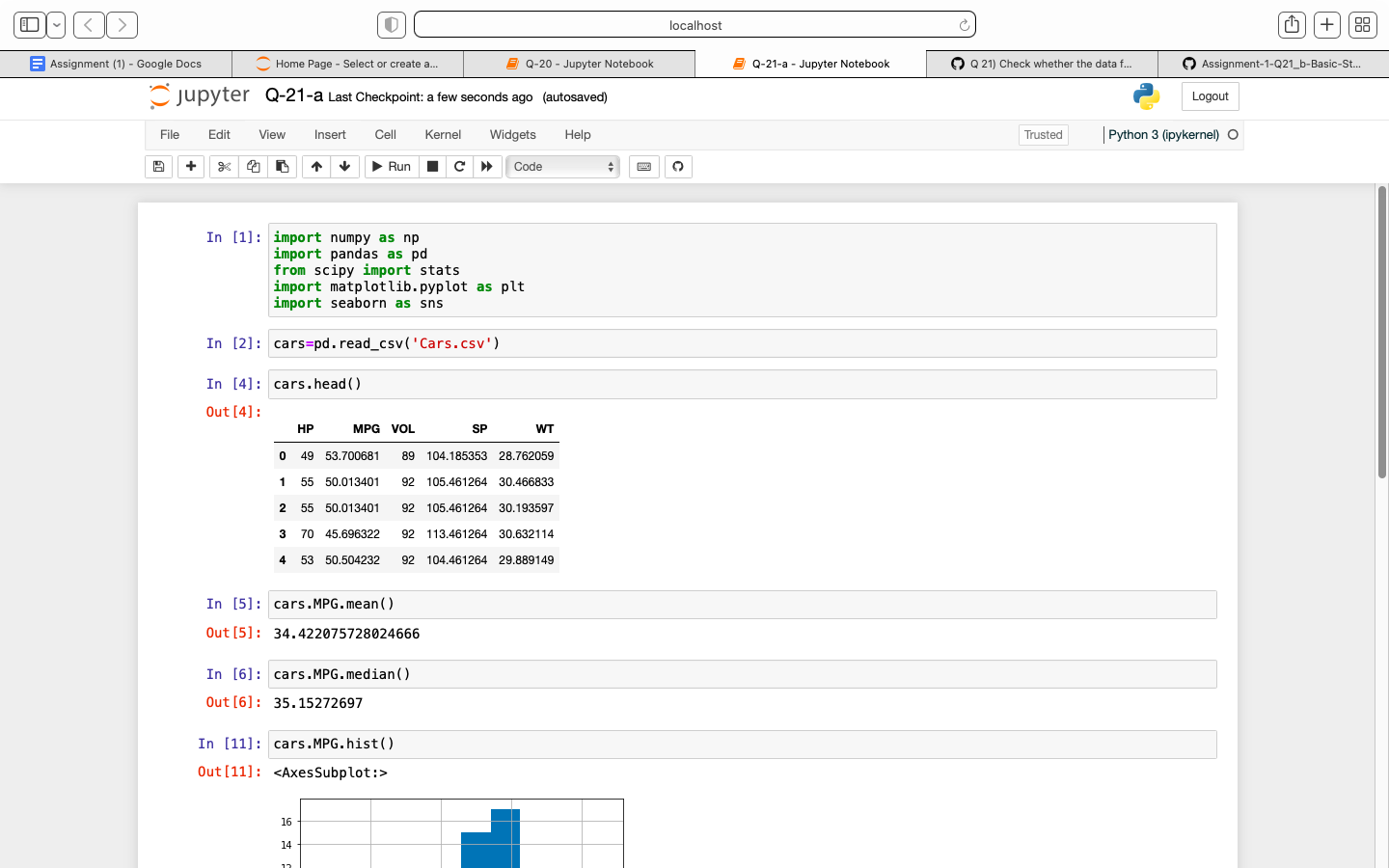


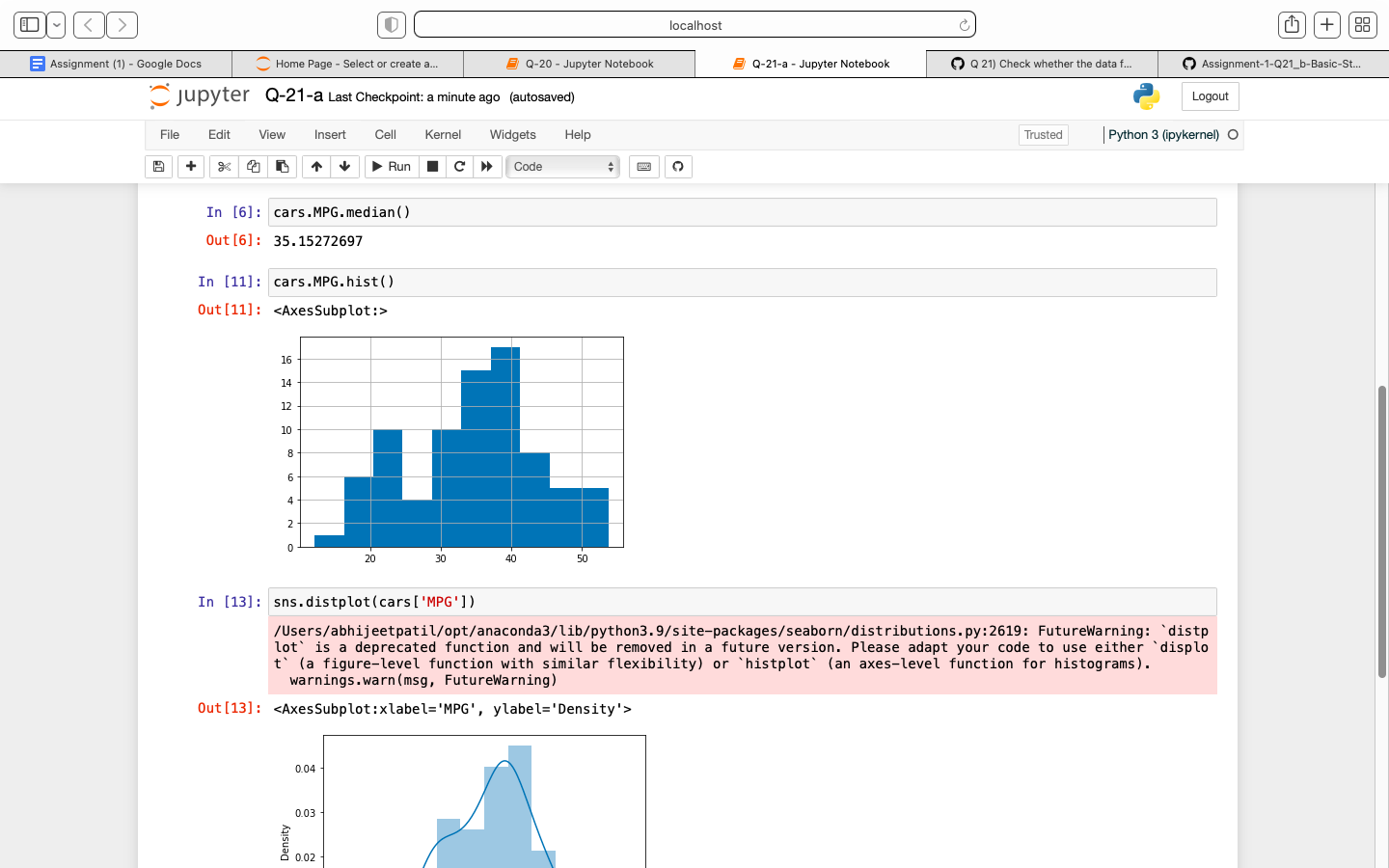
Q 21) Check whether the data follows normal distribution

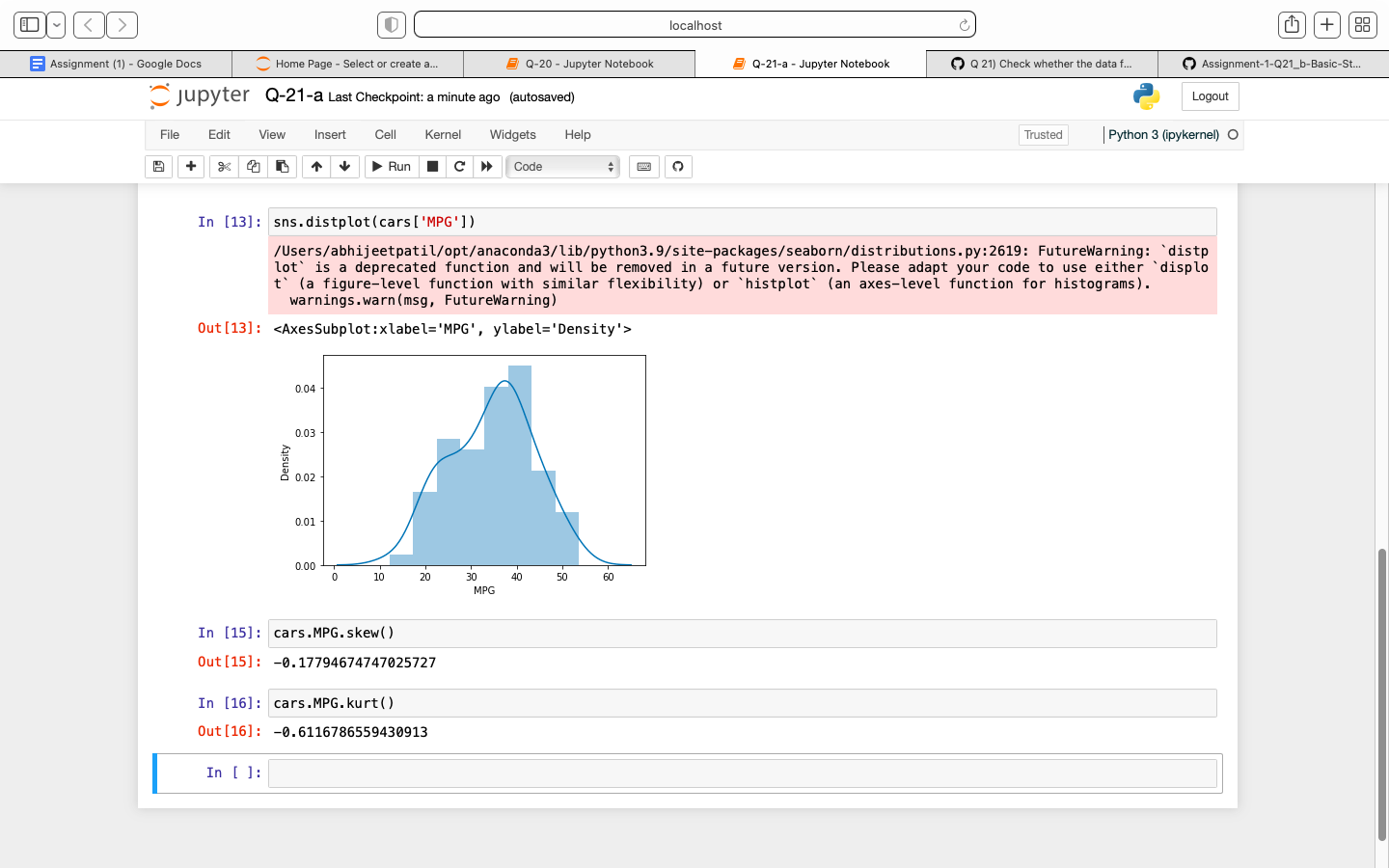
1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Ans- From below plot and values we can say that data is fairly symmetrical, i.e fairly normally distributed.







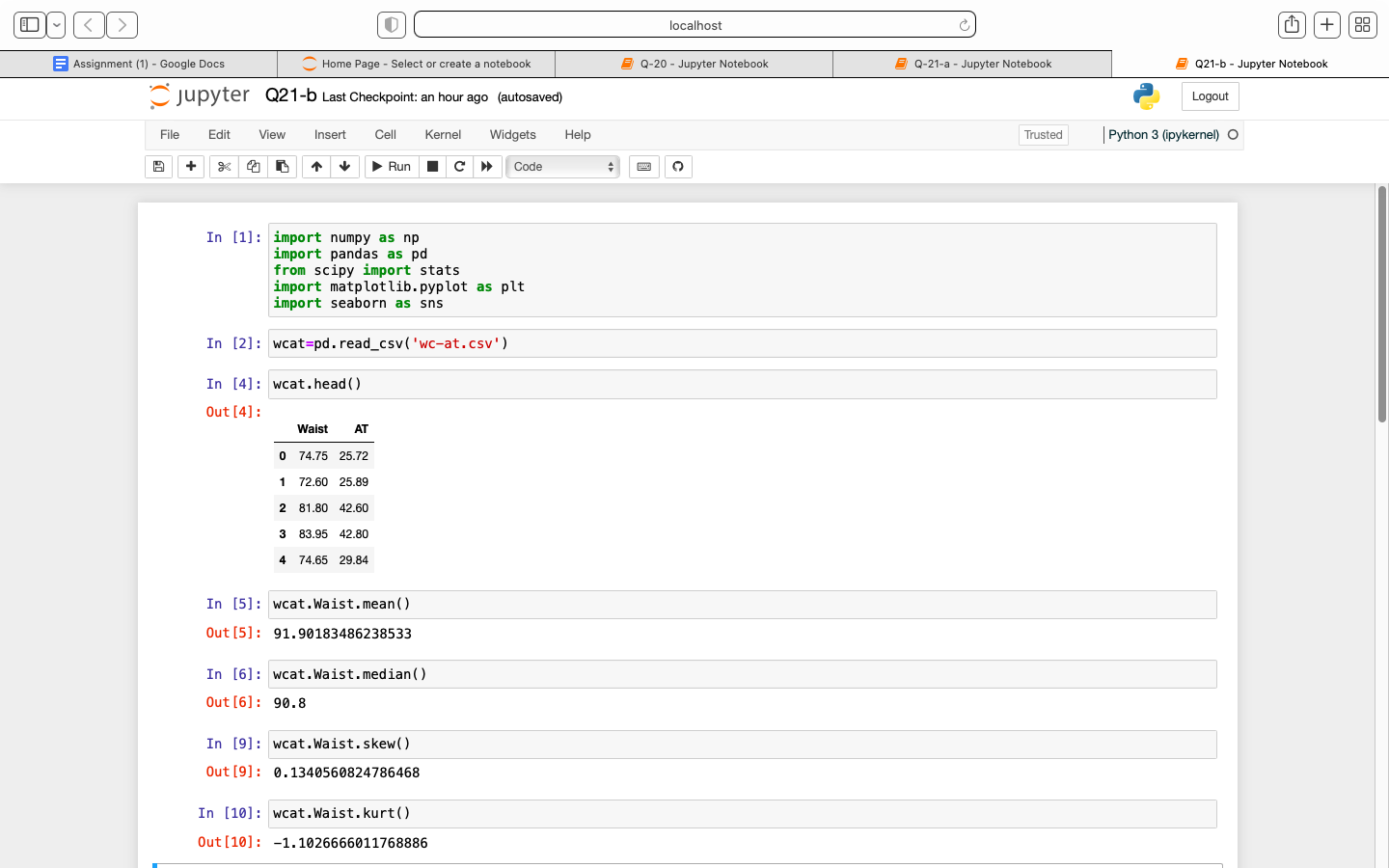
1. Check Whether the Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set follows Normal Distribution

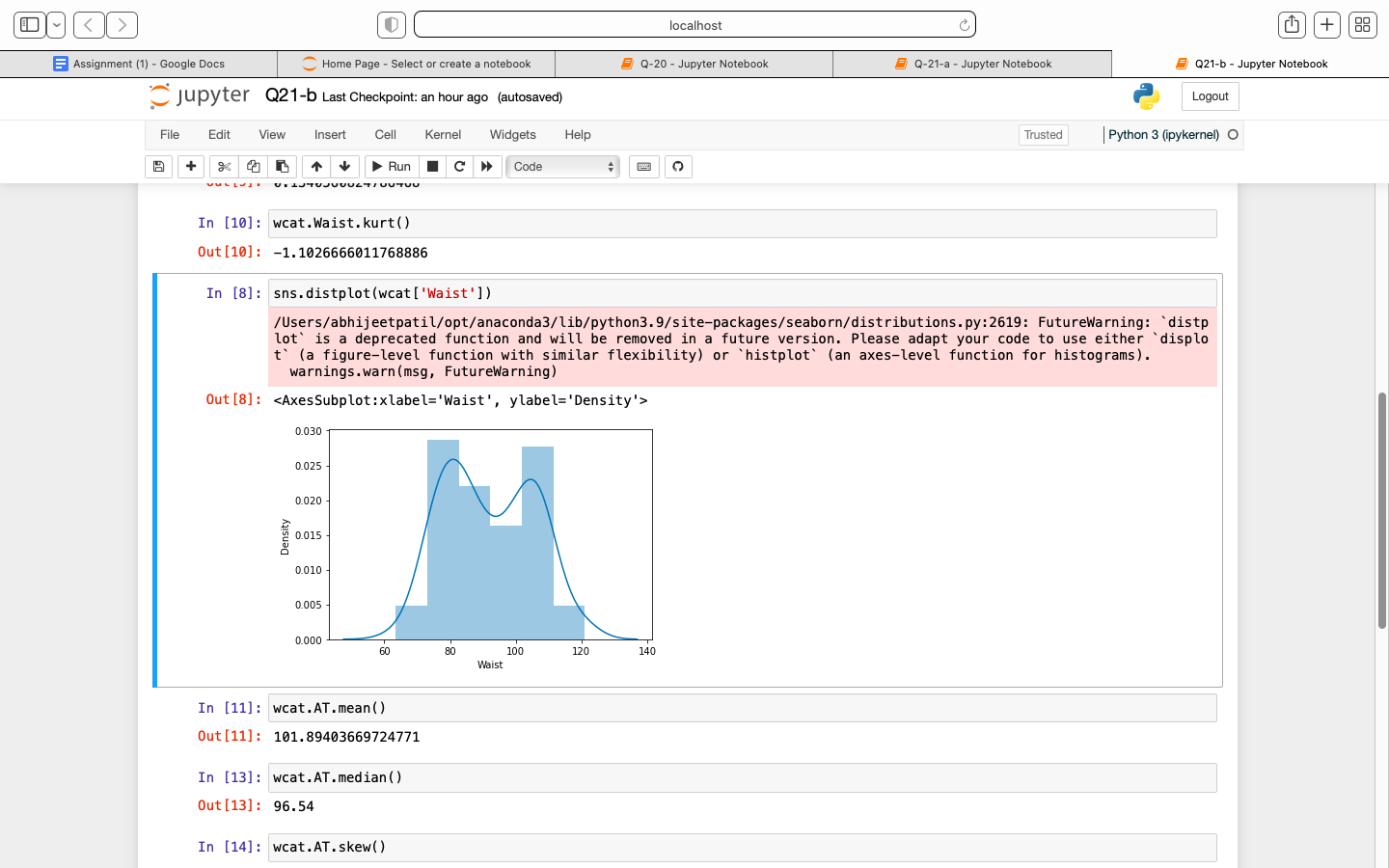
Dataset: wc-at.csv

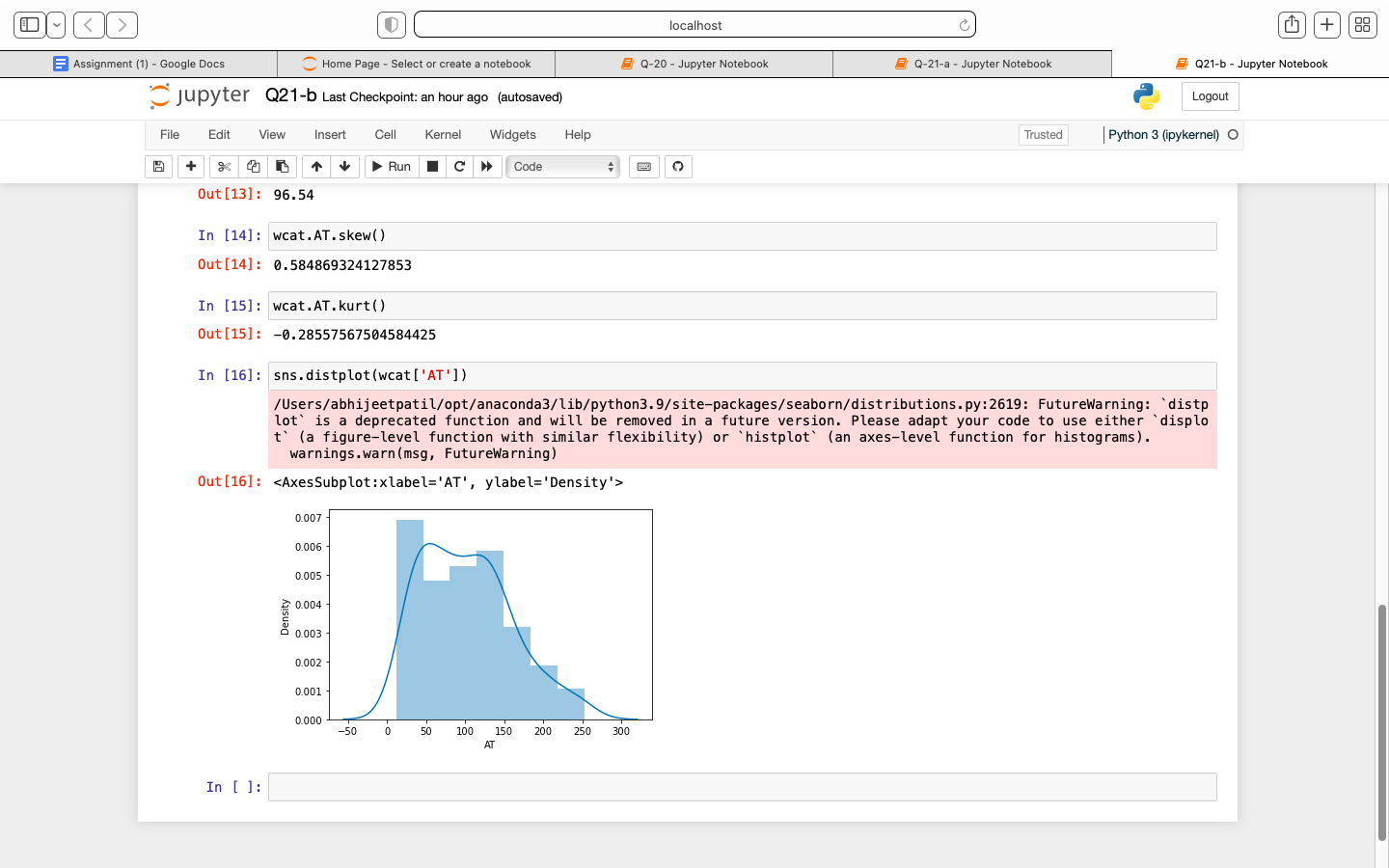
Ans- From below plots and values,

for Waist- data is fairly symmetrical

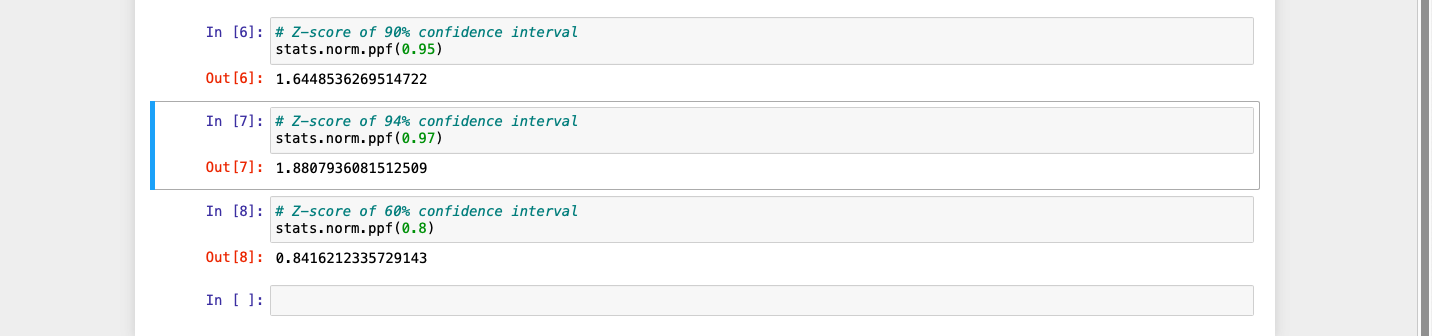
for AT- data is moderately skewed







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

Ans- 

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

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- 