|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | Discrete |
| Results of rolling a dice | Discrete |
| Weight of a person | Continue |
| Weight of Gold | Continue |
| Distance between two places | Continue |
| Length of a leaf | Continue |
| Dog's weight | Continue |
| 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 | Ratio |
| Blood Group | Nominal |
| Time Of Day | Interval |
| Time on a Clock with Hands | Interval |
| Number of Children | Nominal |
| 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?

**Ans** – Three coins are tossed so the sample space is (TTT,TTH, THT, HTT, THH HTH, HHT, HHH)

n(s) = 8

No of times two head and one tails are occurs =3

P =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 – Two dice are rolled so sample space is (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

for

1. n(a) = 0 n(p)=0
2. n(b)= (1,3) (2,2) (3,1) = 3 n(p) 3/36 = 1/1
3. n(c)= (1,5) (2,4) (3,2) (3,3) (6,6) =5 n(p)=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 events = nCr =7C2=7!/2!\*5!=21

Interested events =5C2=5!/2!\*3!=10

Probability that none of the ball is blue=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

=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

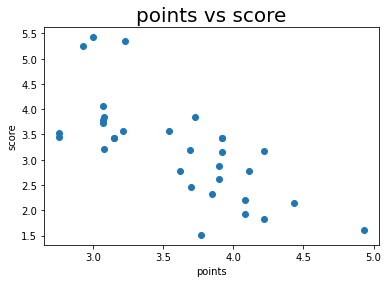
=**3.09**

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

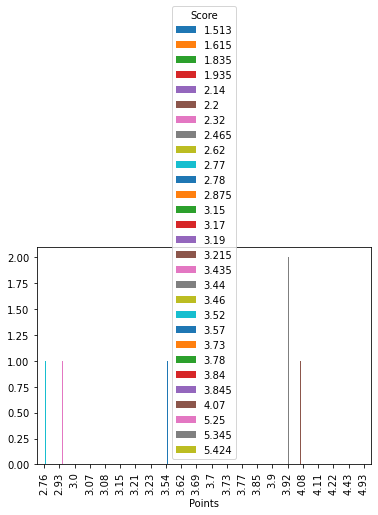
* For Points,Score,Weigh>

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

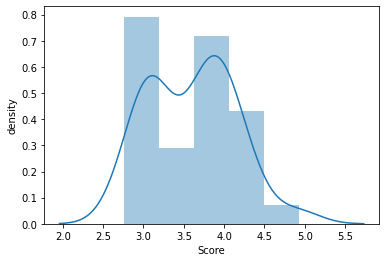
|  | **Points** | **Score** | **Weigh** |
| --- | --- | --- | --- |
| **count** | 32.000000 | 32.000000 | 32.000000 |
| **mean** | 3.596563 | 3.217250 | 17.848750 |
| **std** | 0.534679 | 0.978457 | 1.786943 |
| **min** | 2.760000 | 1.513000 | 14.500000 |
| **25%** | 3.080000 | 2.581250 | 16.892500 |
| **50%** | 3.695000 | 3.325000 | 17.710000 |
| **75%** | 3.920000 | 3.610000 | 18.900000 |
| **max** | 4.930000 | 5.424000 | 22.900000 |



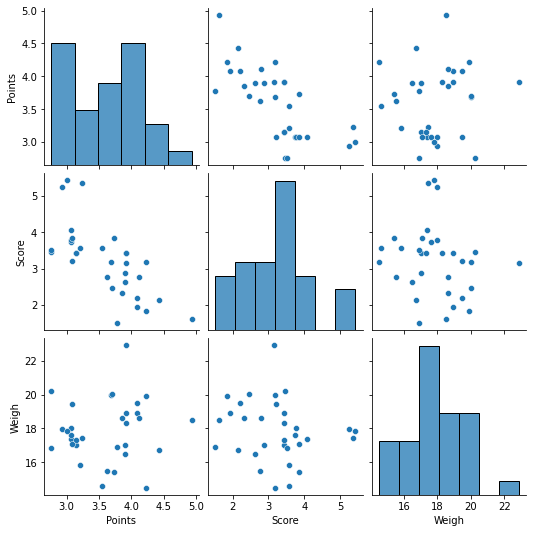
**Scatter Plot**



**Bar plot**

****

**Density Plot**



**Pairplot**

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 -** Expected value = Σ(probability\*value)= **ΣP(x)\*E(x)**

Probability of selecting each patient =1/9

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

=145.33

Expected value of the weight of that patient is **145.33**

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

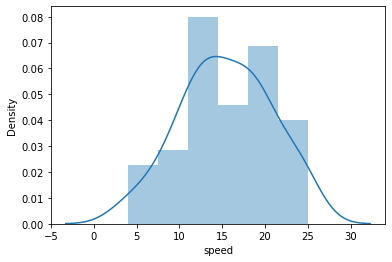
**Cars speed and distance**

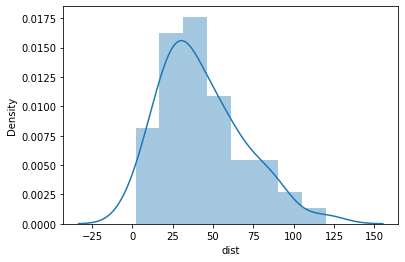
**Ans-Skewness** speed -0.117510

Distance 0.806895

**Kurtosis** speed -0.508994

Distance 0.405053



****

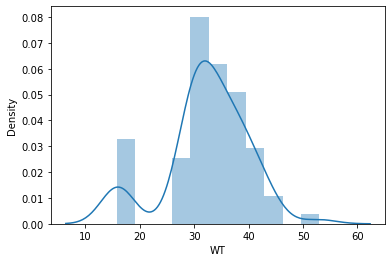
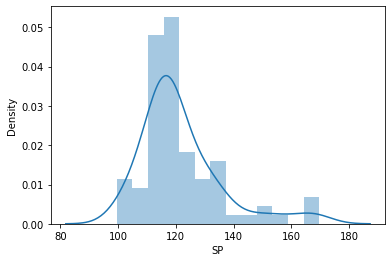
**SP and Weight(WT)**

**Ans- Skewness** SP 1.611450

WT -0.614753

**Kurtosis** SP 2.977329

WT 0.950291



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



Ans = The histogram shows the data is Right side skewed or positively skewed



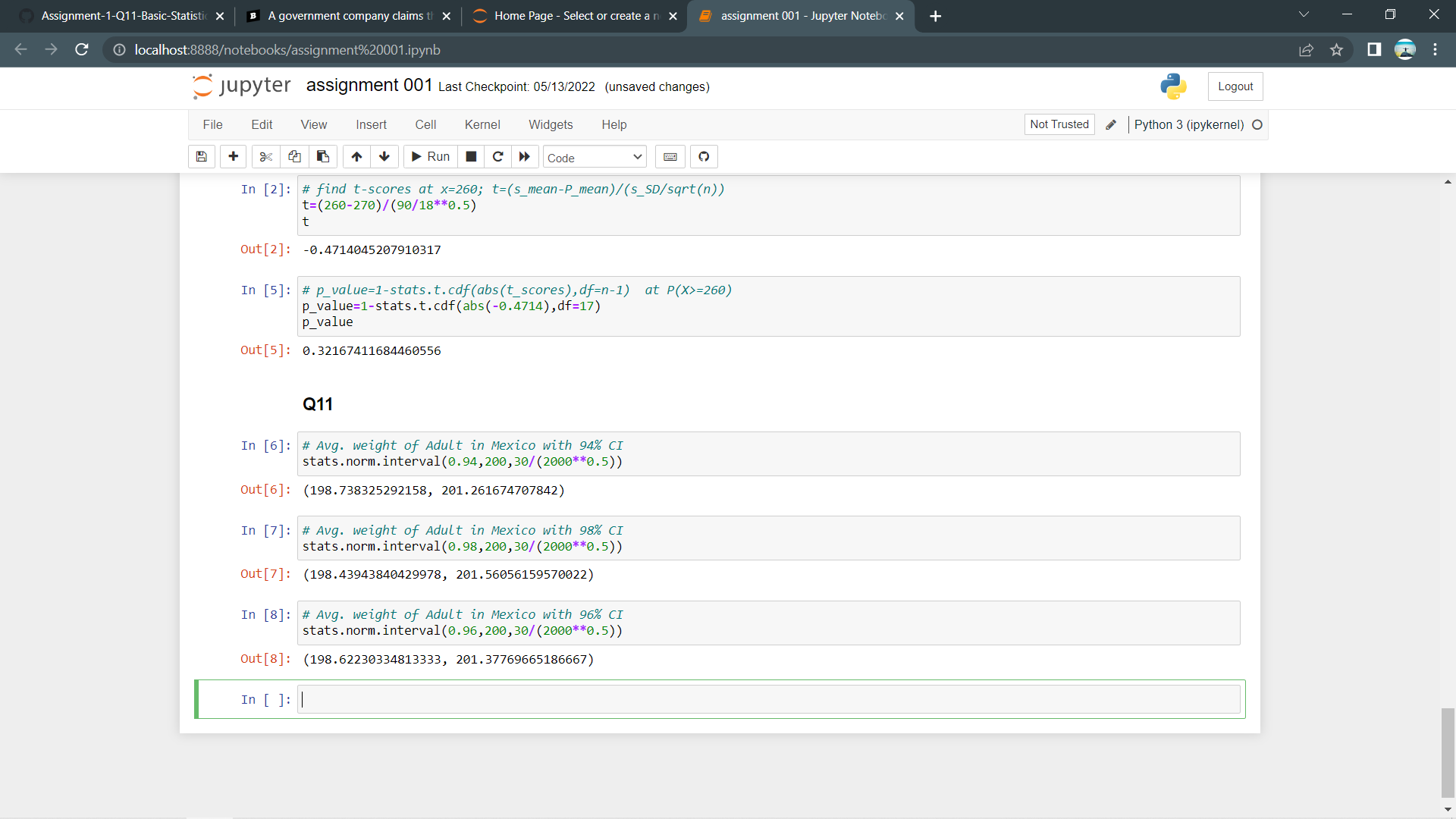
Ans = The interface for this box plot is positively skewed

**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** – Given = sample mean x̅ = 200

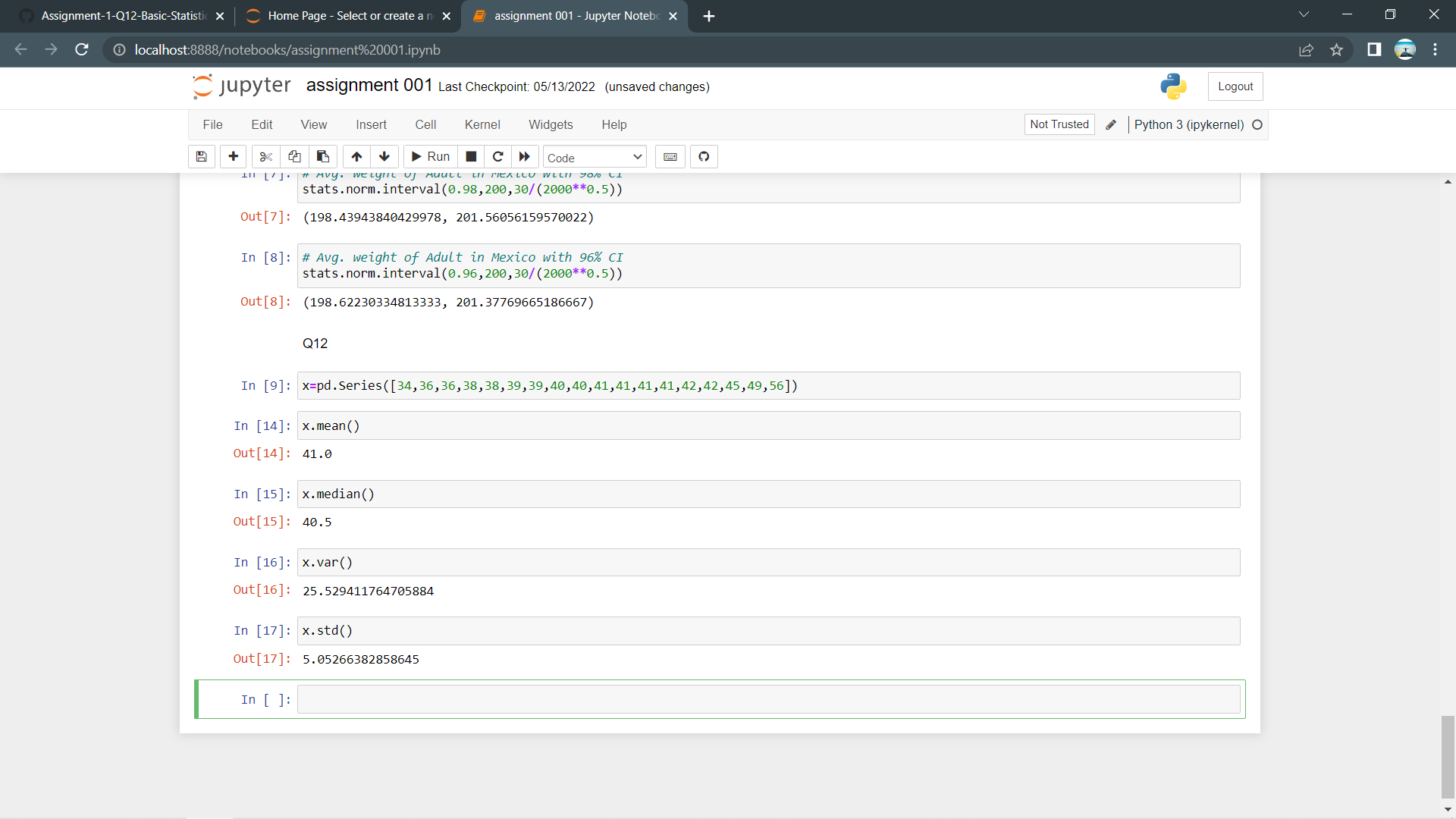
Sample standard deviation S=30

Sample size n= 2000

We have sample standard deviation so. t-distribution is used

**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 = The scores are in uniformly distribution data in ascending order

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**.

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

Ans - Negative value of kurtosis indicate that **a distribution is flat and has thin tails**.

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



What can we say about the distribution of the data?

Ans = The data is distributed in De-assigned format

What is nature of skewness of the data?

Ans = Left side skewed

What will be the IQR of the data (approximately)?   
Ans = Q3-Q1

= 18-10

= 8 is IQR

Q19) Comment on the below Boxplot visualizations?



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

Ans = The box plot 1 designed with range = 3 , The second one range is = 1.5

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

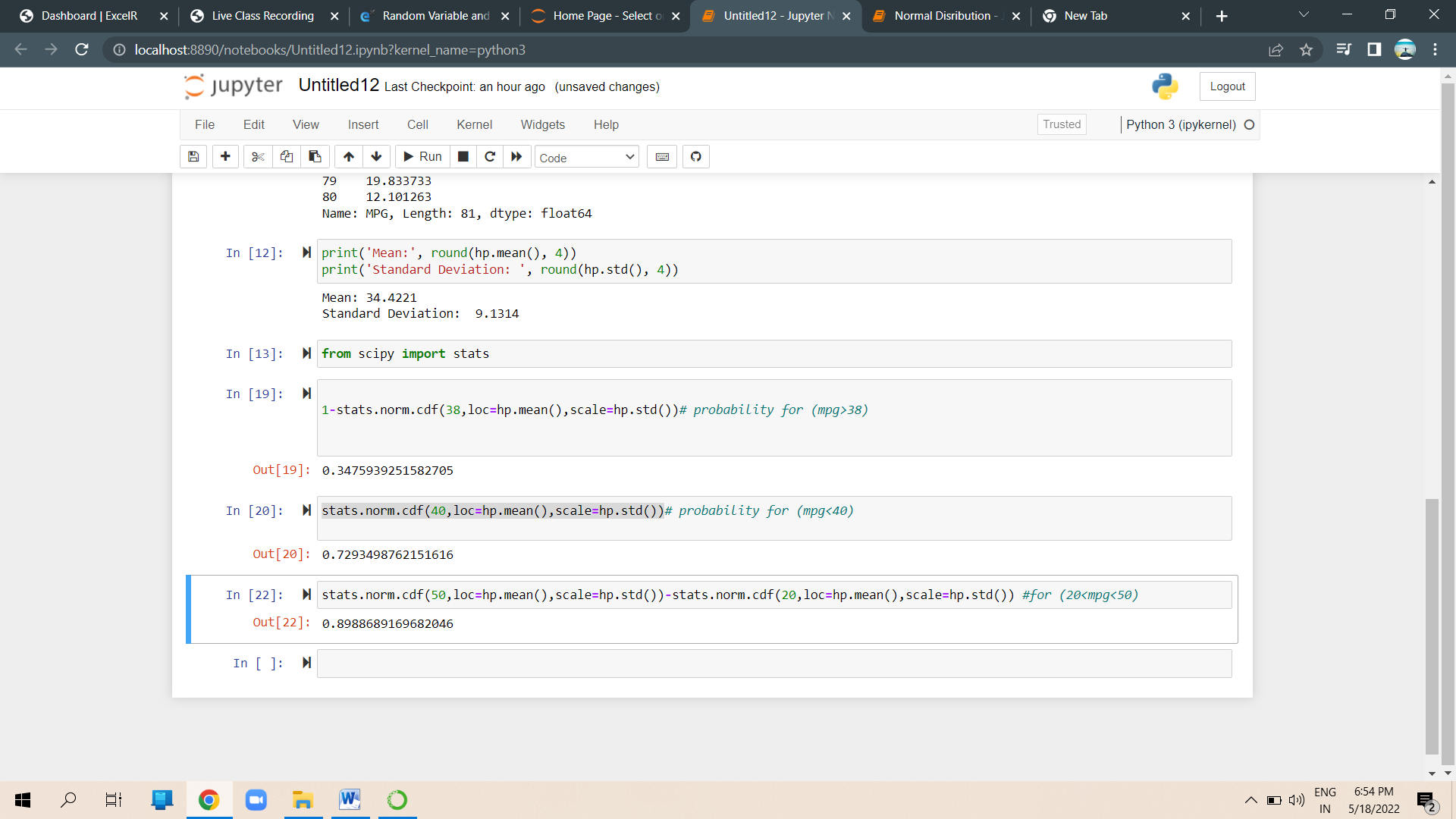
Data \_set: Cars.csv

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

MPG <- Cars$MPG

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

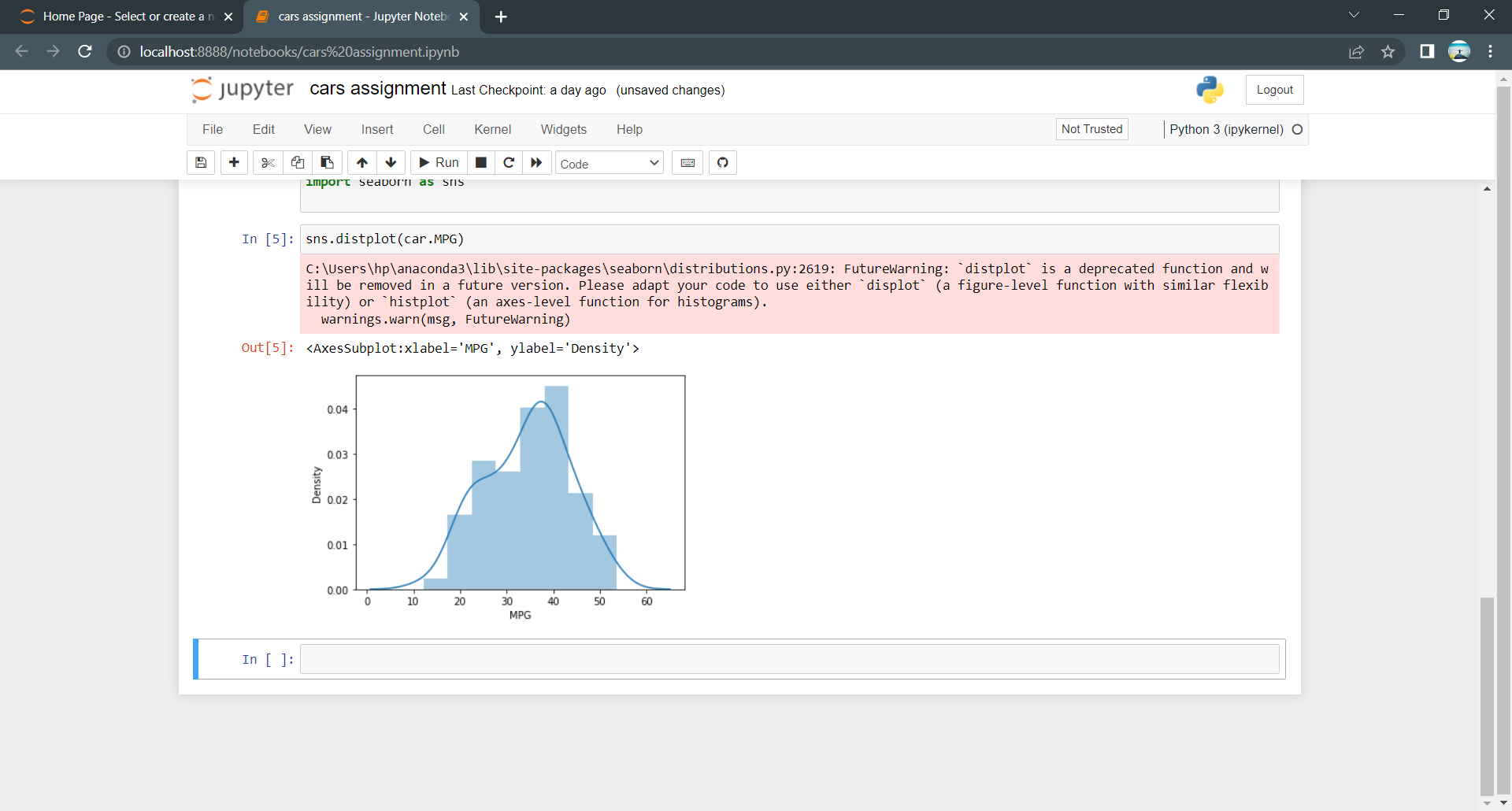
c. P (20<MPG<50)



Q 21) Check whether the data follows normal distribution

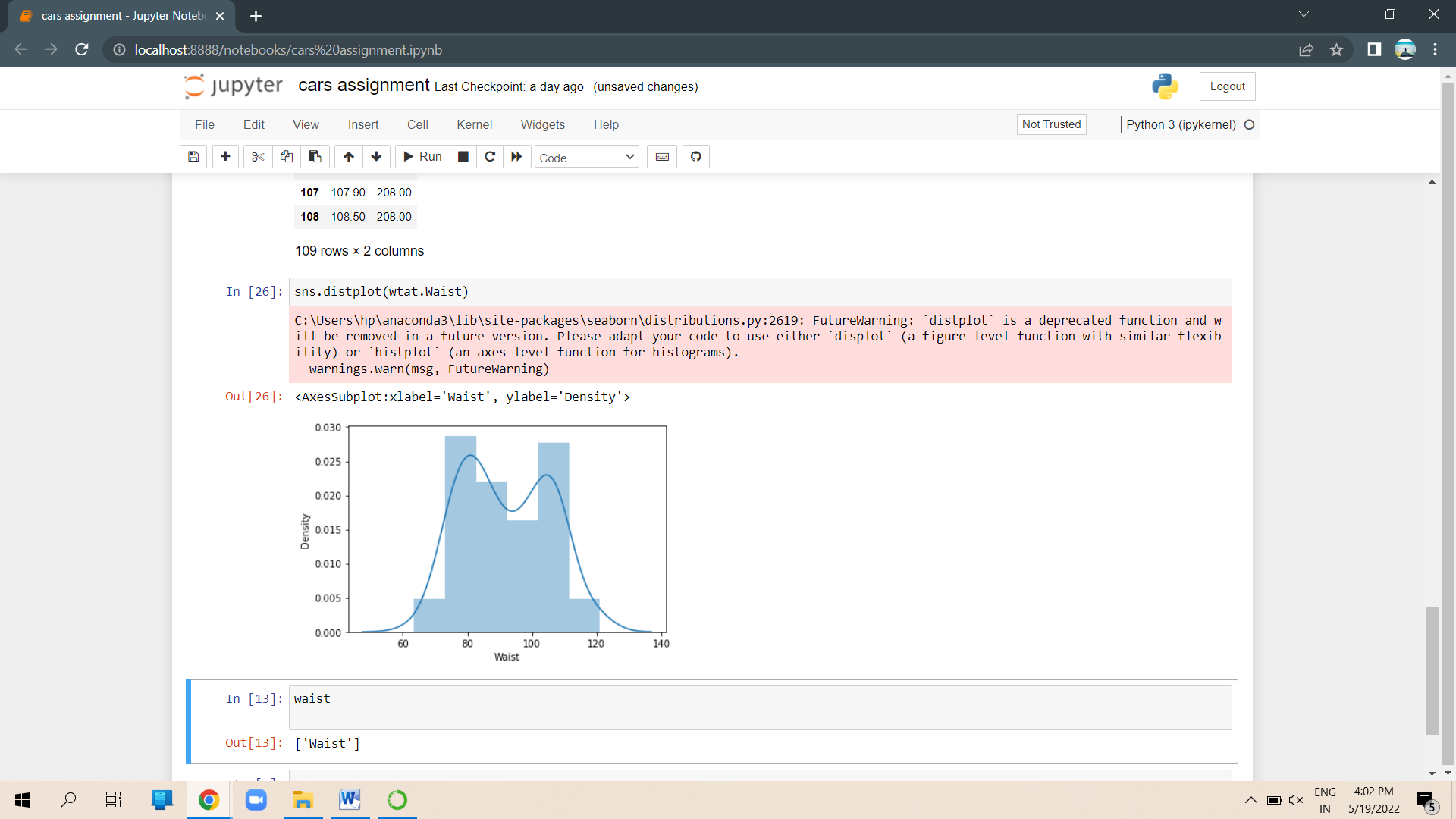
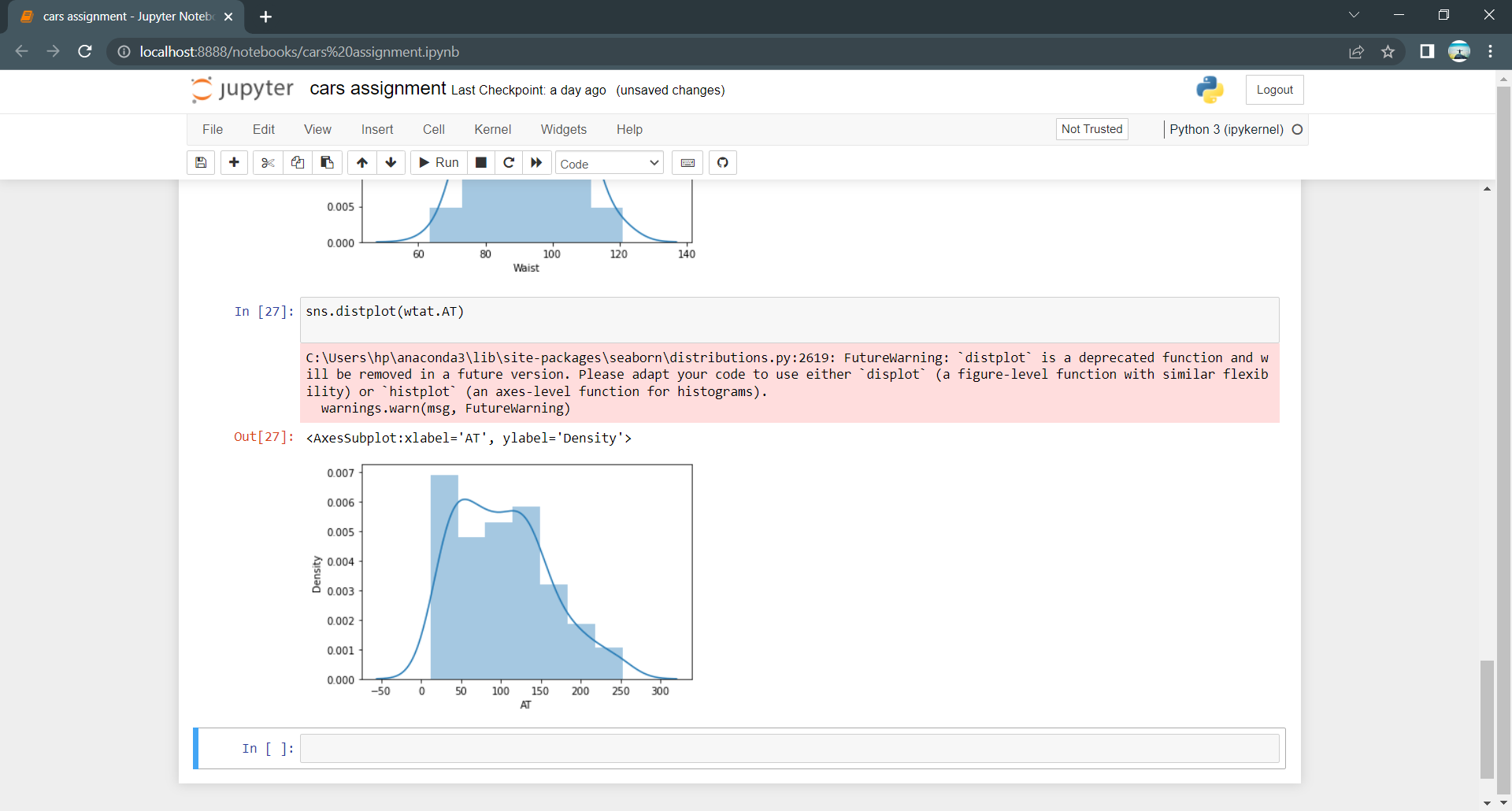
1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

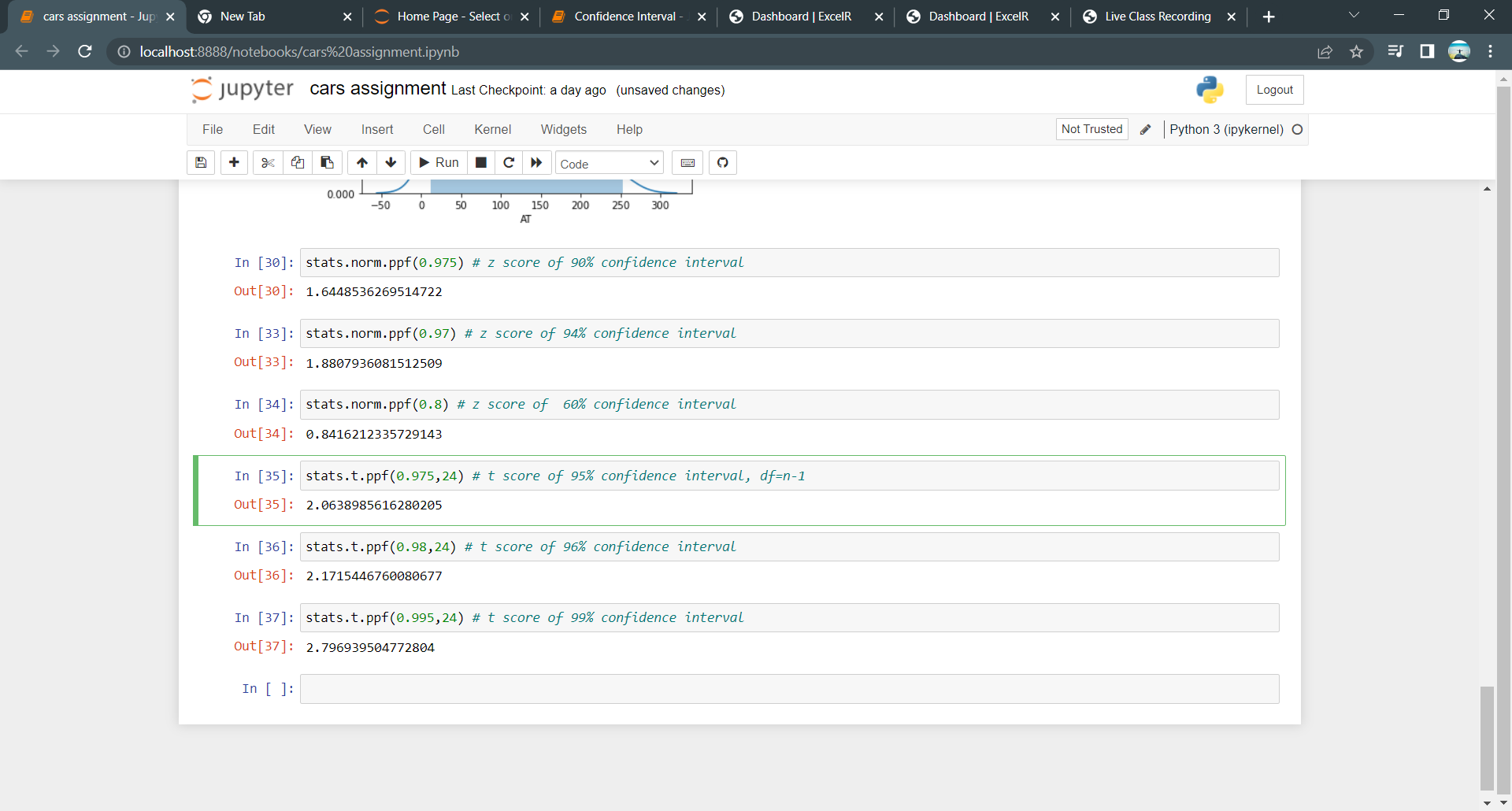


1. Check Whether the Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set follows Normal Distribution

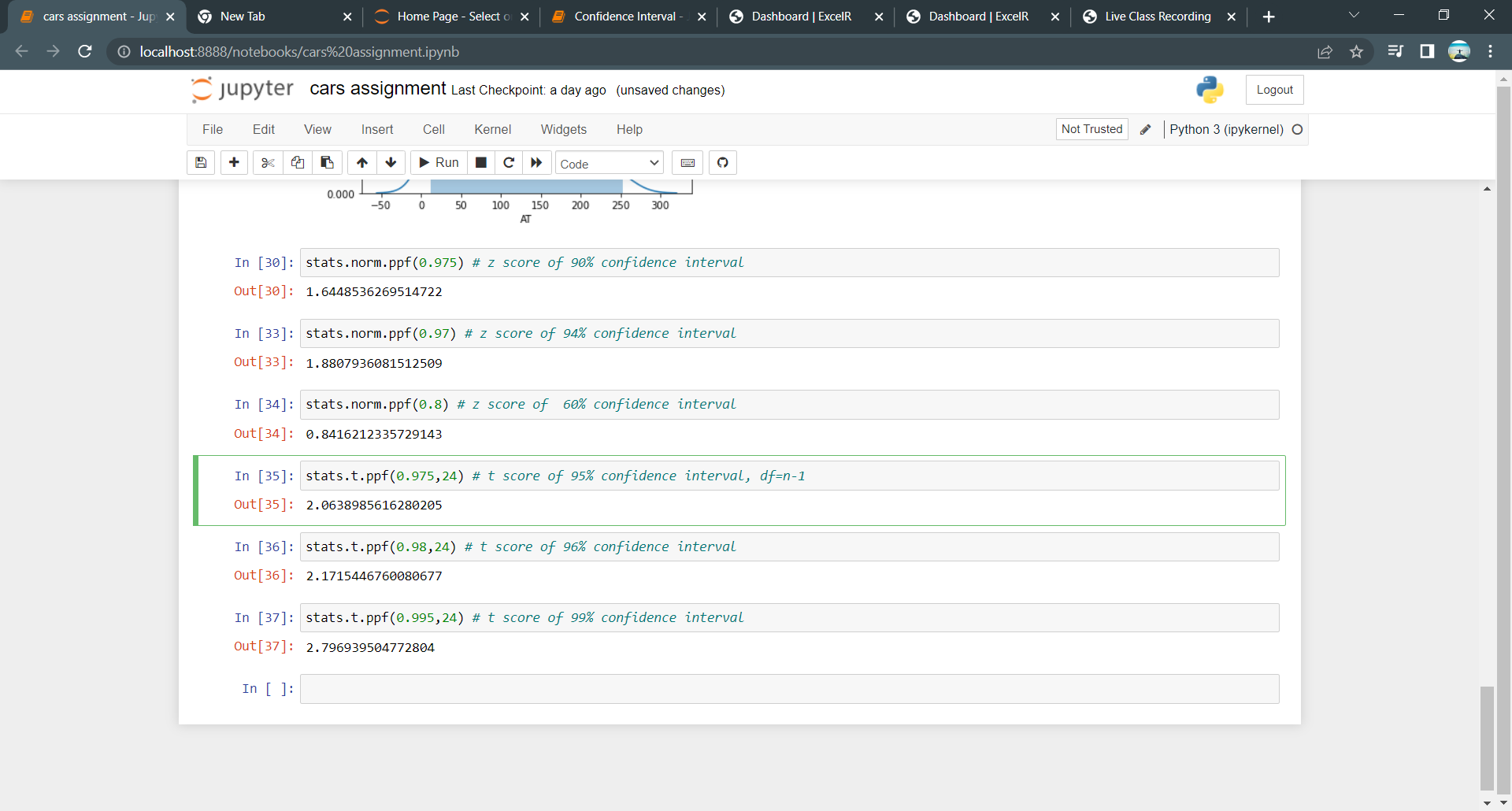
Dataset: wc-at.csv

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