**Q1) Identify the Data type for the Following:**

|  |  |
| --- | --- |
| **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** | *Categorical* |
| **Number of kids** | *Discrete* |
| **Number of tickets in Indian railways** | *Discrete* |
| **Number of times married** | *Discrete* |
| **Gender (Male or Female)** | *Categorical* |

**Q2) Identify the Data types, which were among the following**

**Nominal, Ordinal, Interval, Ratio.**

|  |  |
| --- | --- |
| **Data** | *Data Type* |
| **Gender** | *Nominal* |
| **High School Class Ranking** | *Nominal* |
| **Celsius Temperature** | *Interval* |
| **Weight** | *Interval* |
| **Hair Color** | *Nominal* |
| **Socioeconomic Status** | *Nominal* |
| **Fahrenheit Temperature** | *Interval* |
| **Height** | *Interval* |
| **Type of living accommodation** | *Nominal* |
| **Level of Agreement** | *Ratio* |
| **IQ(Intelligence Scale)** | *Ordinal* |
| **Sales Figures** | *Ratio* |
| **Blood Group** | *Nominal* |
| **Time Of Day** | *Interval* |
| **Time on a Clock with Hands** | *Ratio* |
| **Number of Children** | *Ordinal* |
| **Religious Preference** | *Ordinal* |
| **Barometer Pressure** | *Interval* |
| **SAT Scores** | *Ordinal* |
| **Years of Education** | *Ratio* |

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

**Ans :**

*The probability of getting two heads and one tail on tossing coins at once is equal to 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 :**

*Sum = (1).*

*The minimum possible sum is (1, 1) = (2).*

*P(sum = 1) = (0)/(36) = 0*

*Sum <= (4)*

*combinations : (1,1), (1,2), (1,3), ( 2, 1), (2, 2), (3, 1)*

*P(sum <= 4) = (6/36) = 6*

*Sum divisible by (2 & 3)*

*combinations : (1,5), (3,3), (4,2), (5,1), (6,6)*

*Number of outcomes : 5*

*probability = 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 :**

*Sample Space = 7c2 = 21*

*Let Event A be non of the ball is blue = all balls are red or green or both*

*n(A) = 5c2 = 10*

*p(A) = 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 :**

*By solution we get,*

*=  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*

*=       3.090*

*=  3.09*

*So, Expected number of candies for a randomly selected child is 3.09*

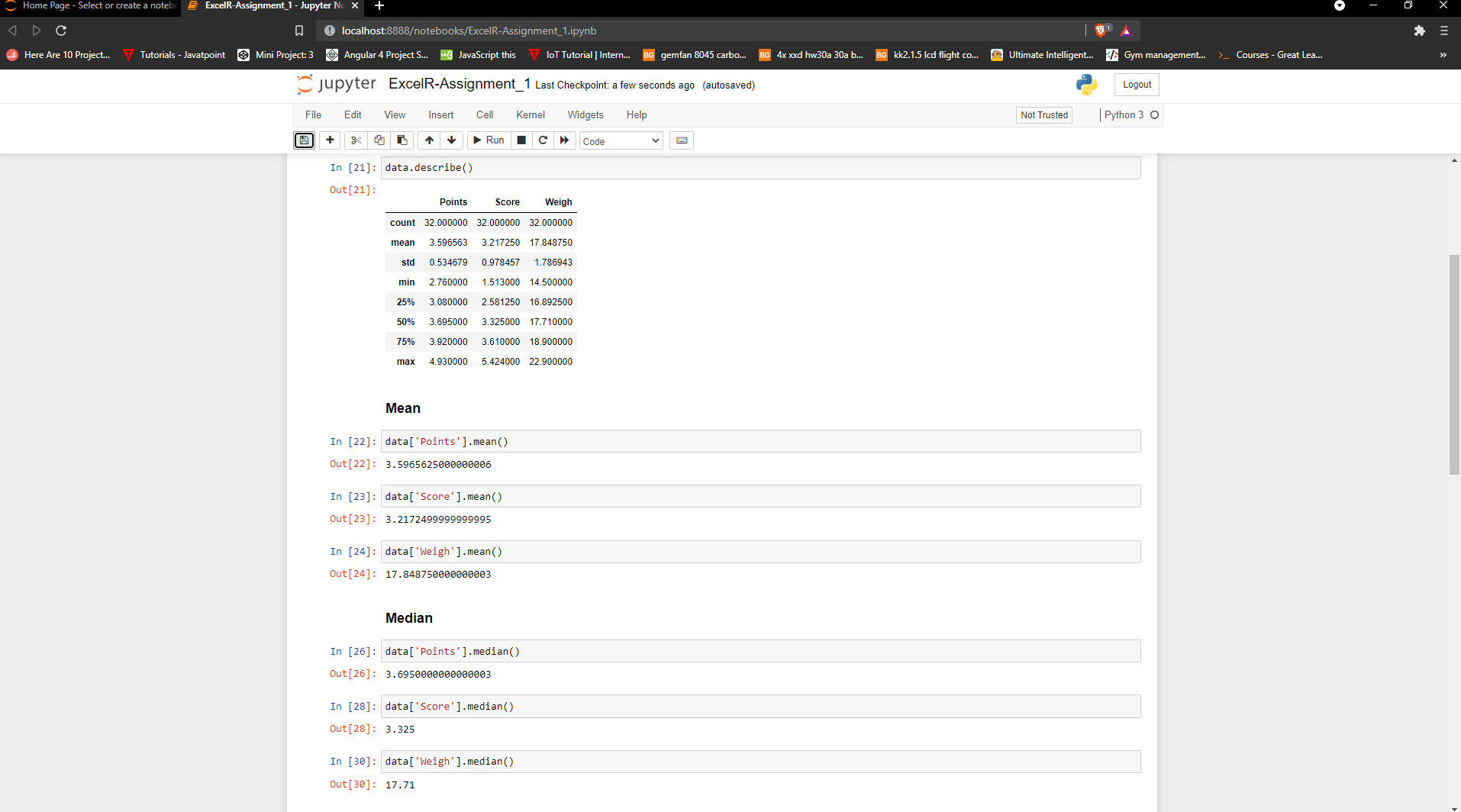
**Q7) Calculate Mean, Median, Mode, Variance, Standard Deviation, Range & comment about the values / draw inferences, for the given dataset**

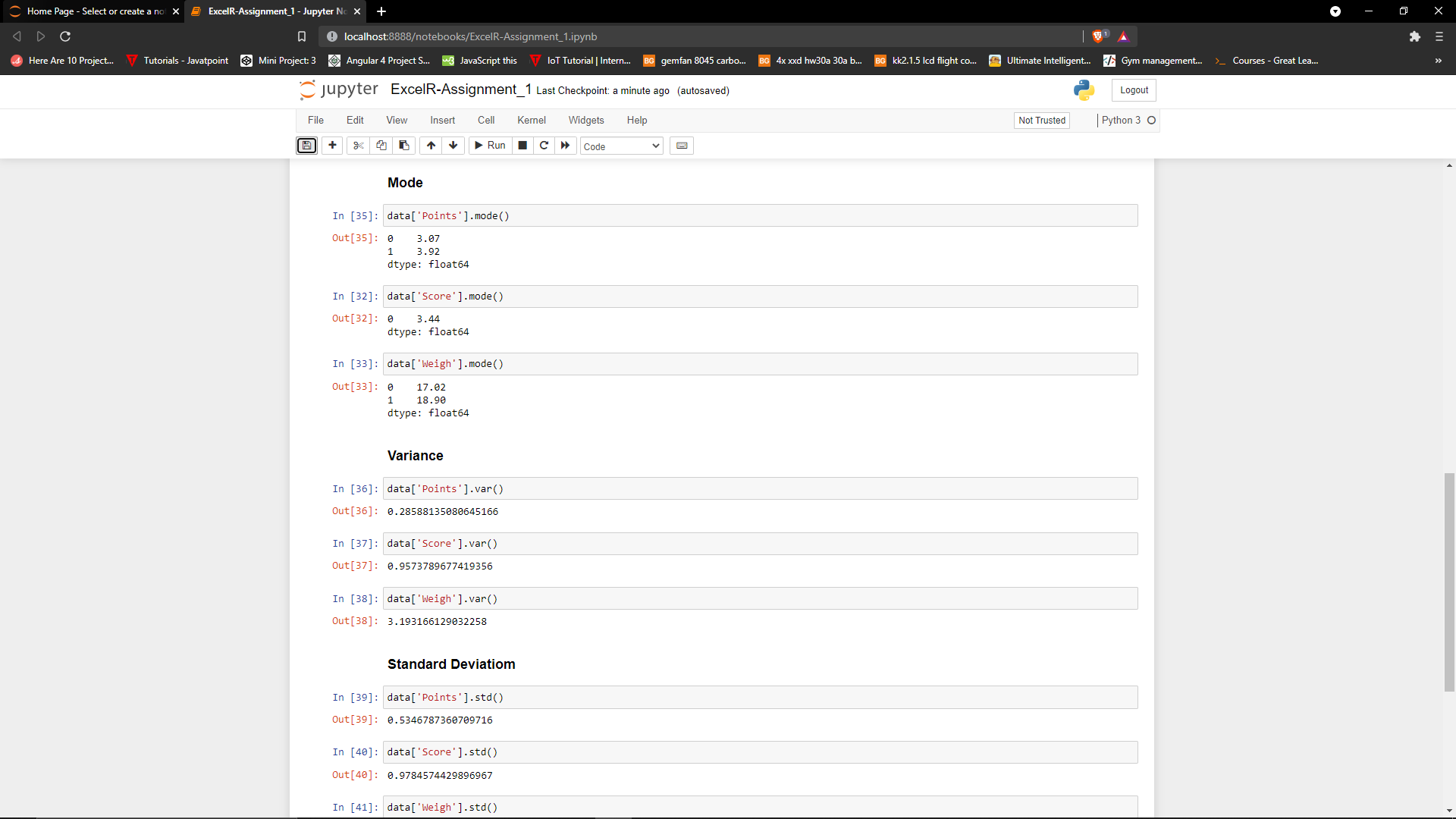
* **For Points,Score,Weigh>**

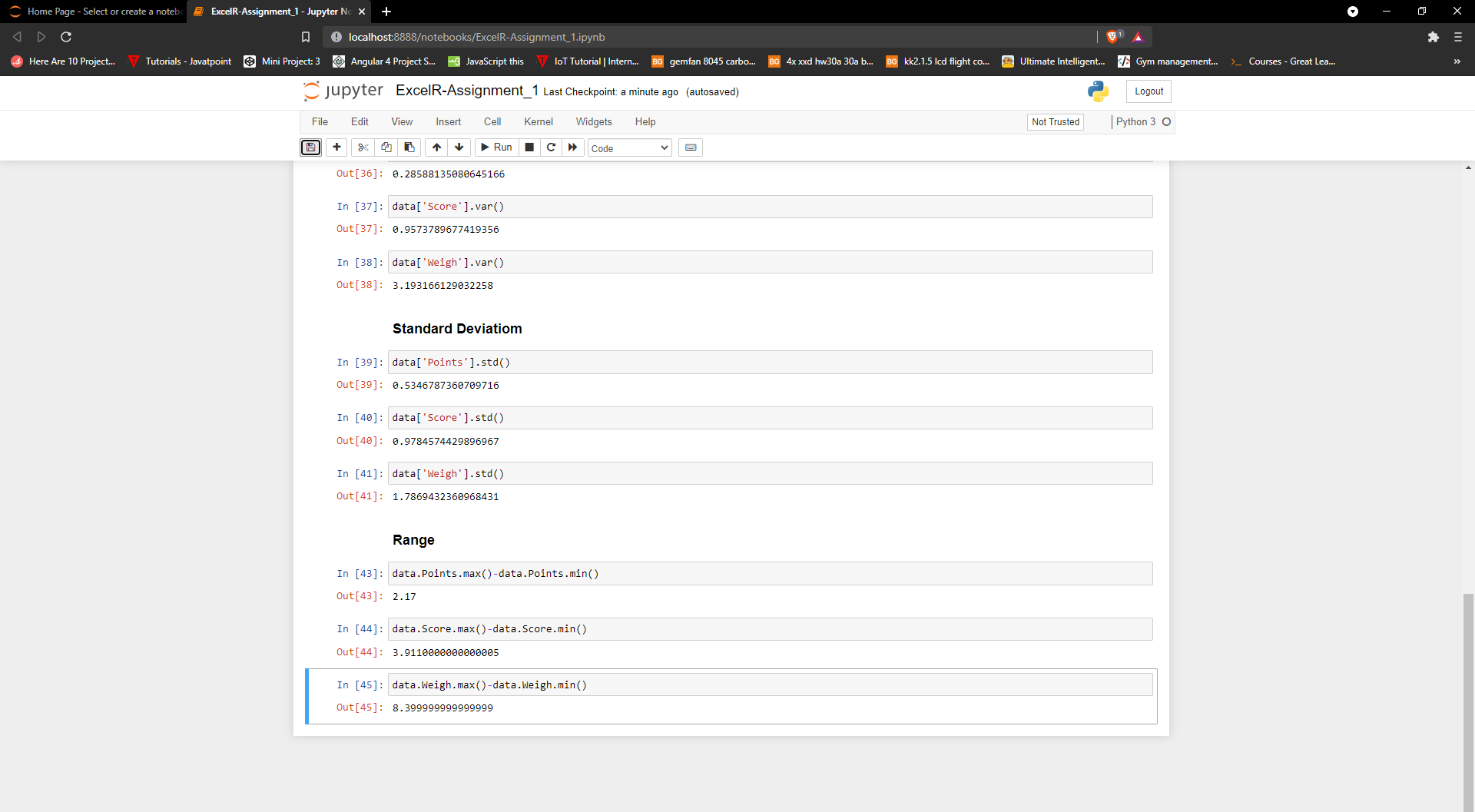
**Find Mean, Median, Mode, Variance, Standard Deviation, and Range and also Comment about the values/ Draw some inferences.**

**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 :**

*Probability of selecting each patient = 1/9*

*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)*

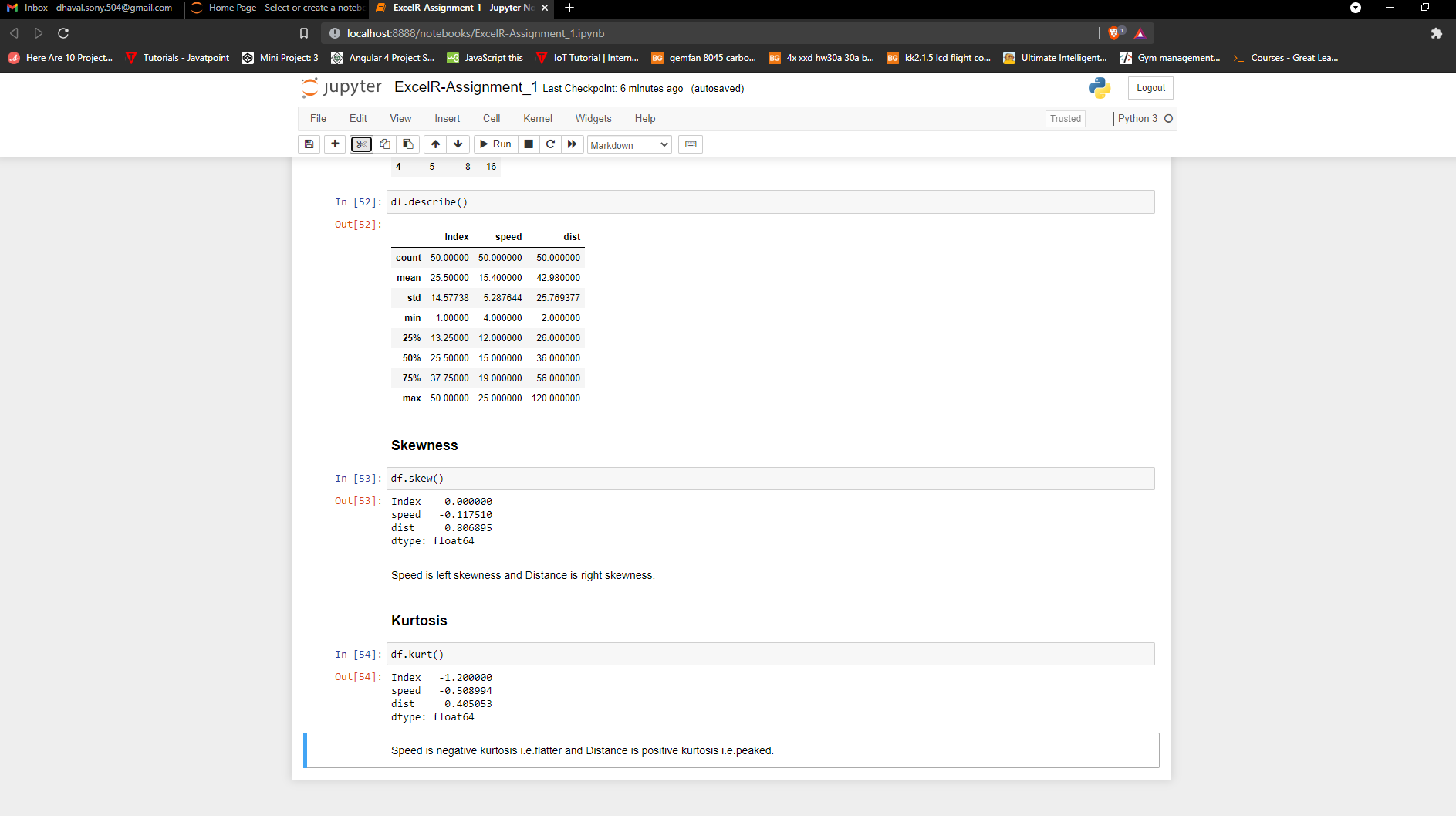
*= 145.33*

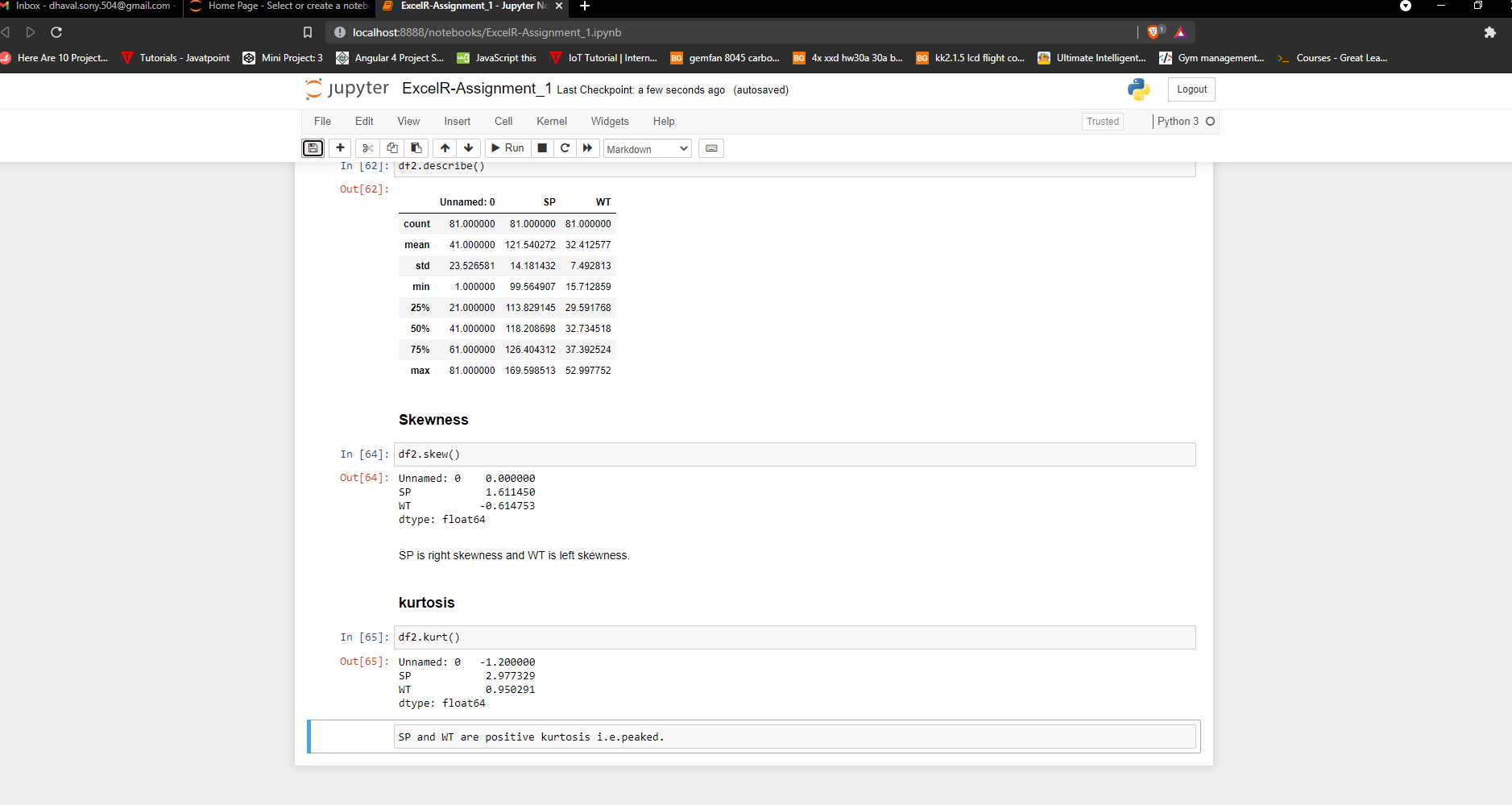
*Expected Value of the Weight of that patient = 145.33*

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

**Cars speed and distance. Use Q9\_a.csv & SP and Weight(WT). Use Q9\_b.csv**

**Ans :**





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





**Ans :**

**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 :**

*S.E = 0.6708*

*Compute alpha = 0.05*

*critical probability = 0.975*

*degree of freedom = 1999*

*critical value (from t-chart) = 1.96*

*By compute margin of error(ME), we get values :*

*94% confident that the population mean falls within the interval 200 +- 1.84.*

*98% confident that the population mean falls within the interval 200 +- 1.92.*

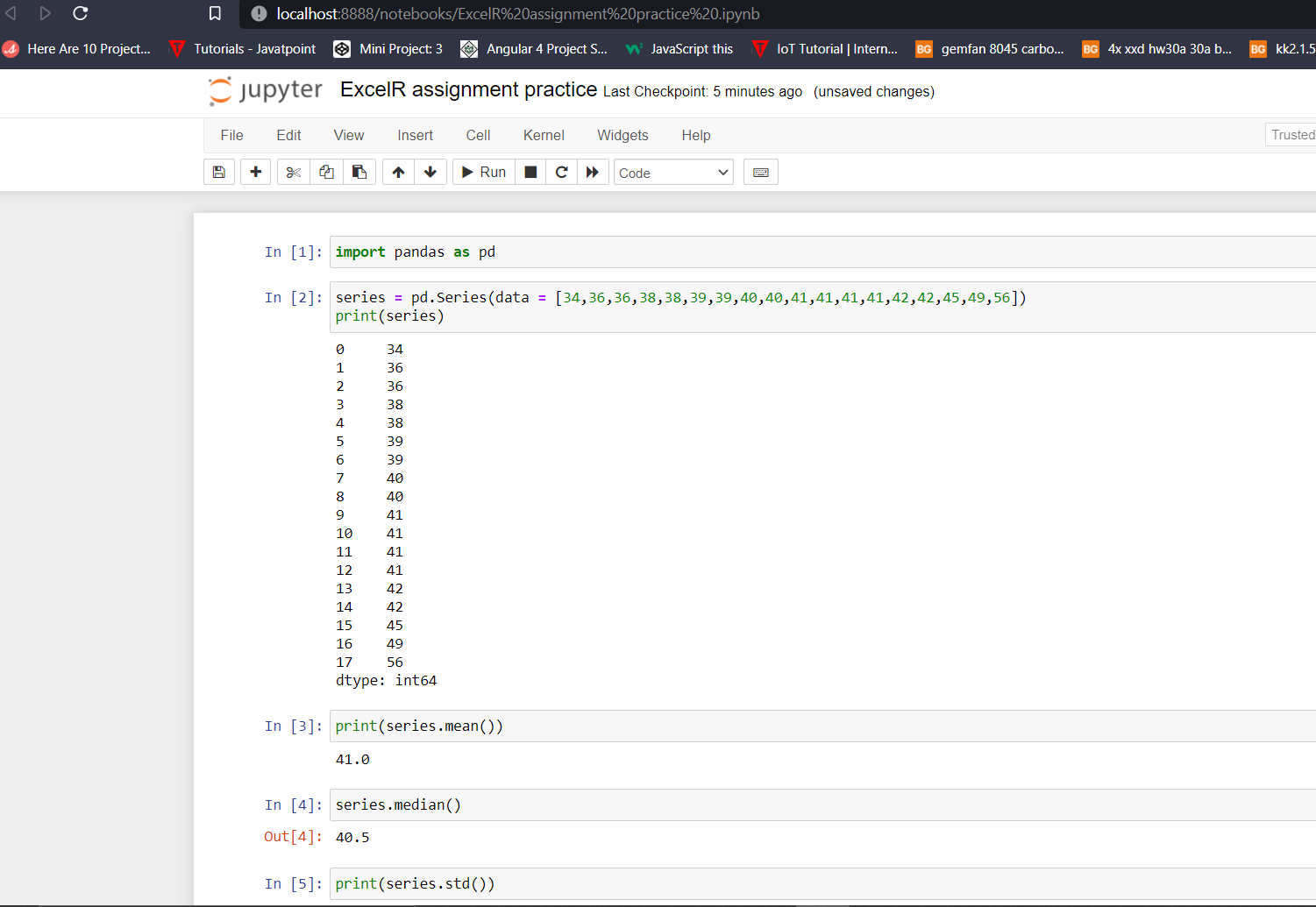
*96% confident that the population mean falls within the interval 200 +- 1.88.*

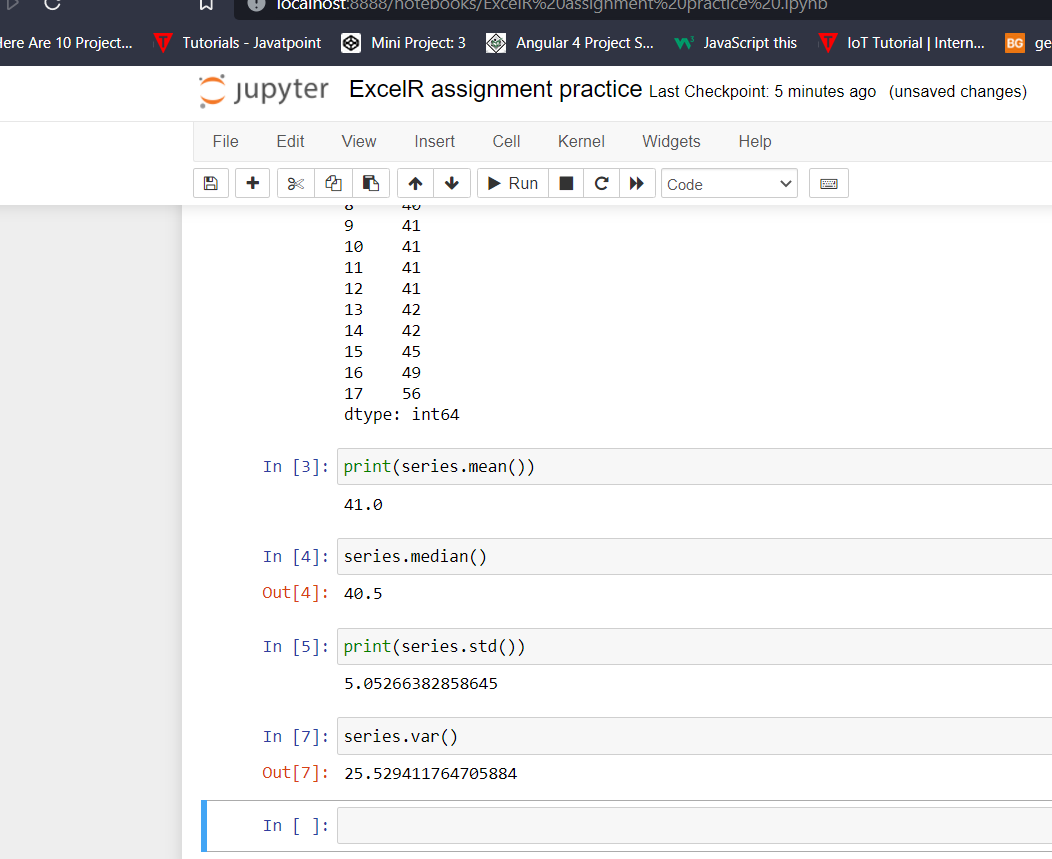
**Q12) Below are the scores obtained by a student in tests**

**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 :**





*According to student marks, mean marks scored is 41. Highest marks scored by a student is 56. Least marks scored by a student is 34.*

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

**Ans :**

*If the distribution is symmetric, then the mean is equal to the median, and the distribution has zero skewness. If the distribution is both symmetric and unimodal, then the mean = median.*

**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 a distribution is peaked and possess thick tails.*

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

**Ans :**

*Negative values 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?**

**What is nature of skewness of the data?**

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

**Ans :**

i) The mean is greater than median.

ii) Nature of skewness of the above data is positively skewed.

iii) The IQR of the data will be 15.

**Q19) Comment on the below Boxplot visualizations?**



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

**Ans :**

According to boxplot 2, data is more spread out. That means more values are away from mean. So, there will be a Bell-Shaped Graph.

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

**Data \_set: Cars.csv**

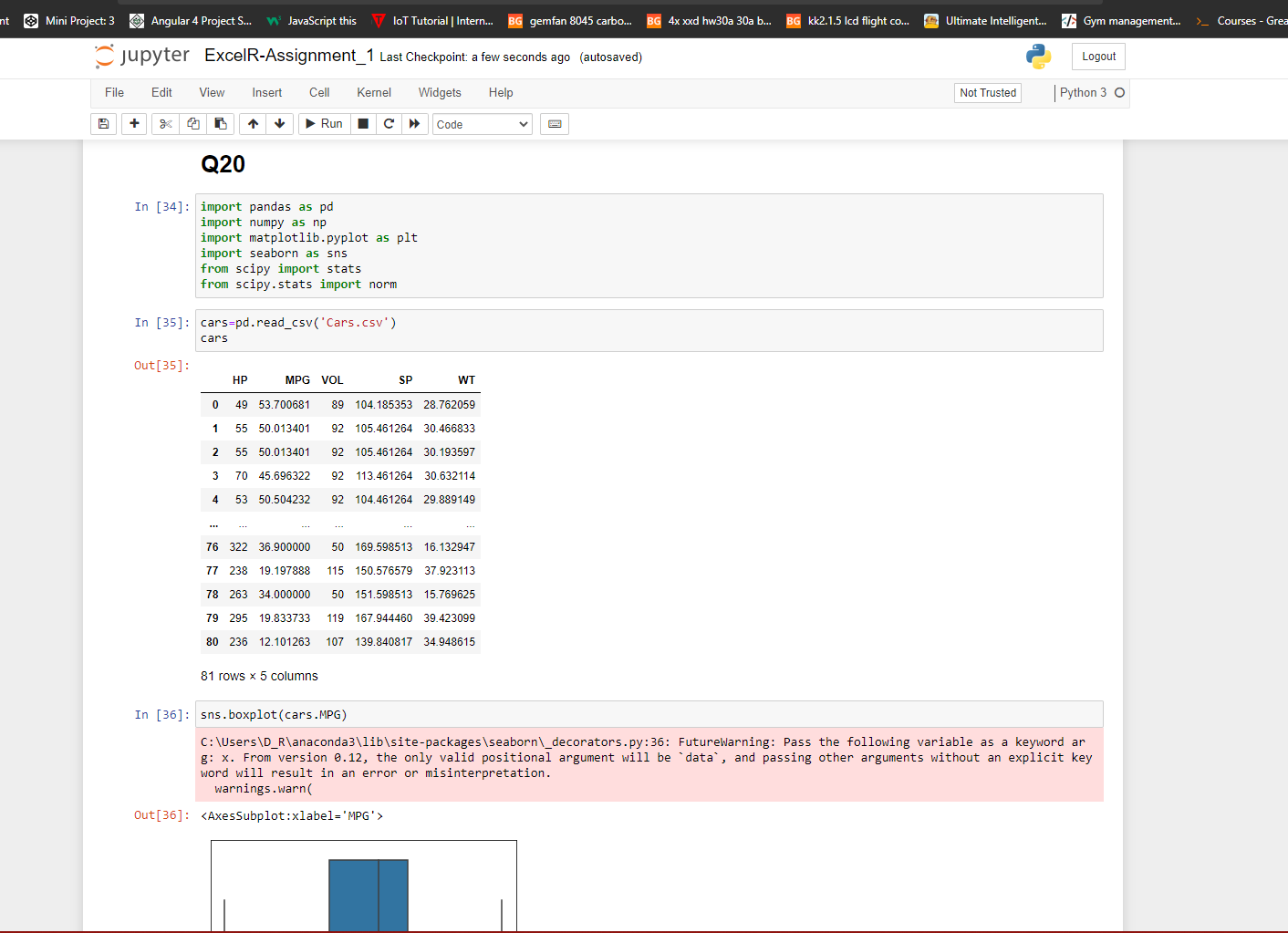
**Calculate the probability of MPG of Cars for the below cases.**

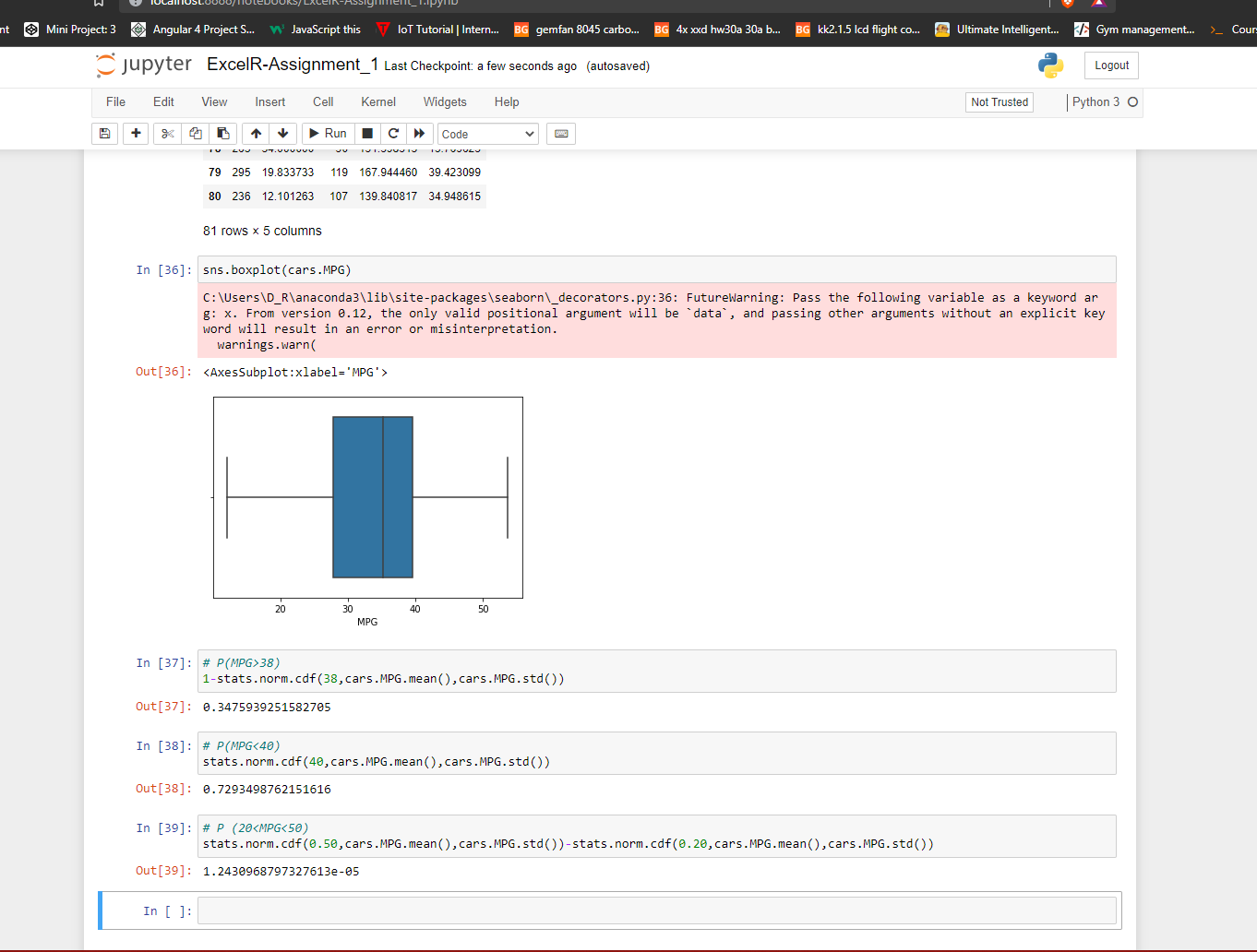
**MPG <- Cars$MPG**

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

**c. P (20<MPG<50)**

**Ans :**





**Q 21) Check whether the data follows normal distribution**

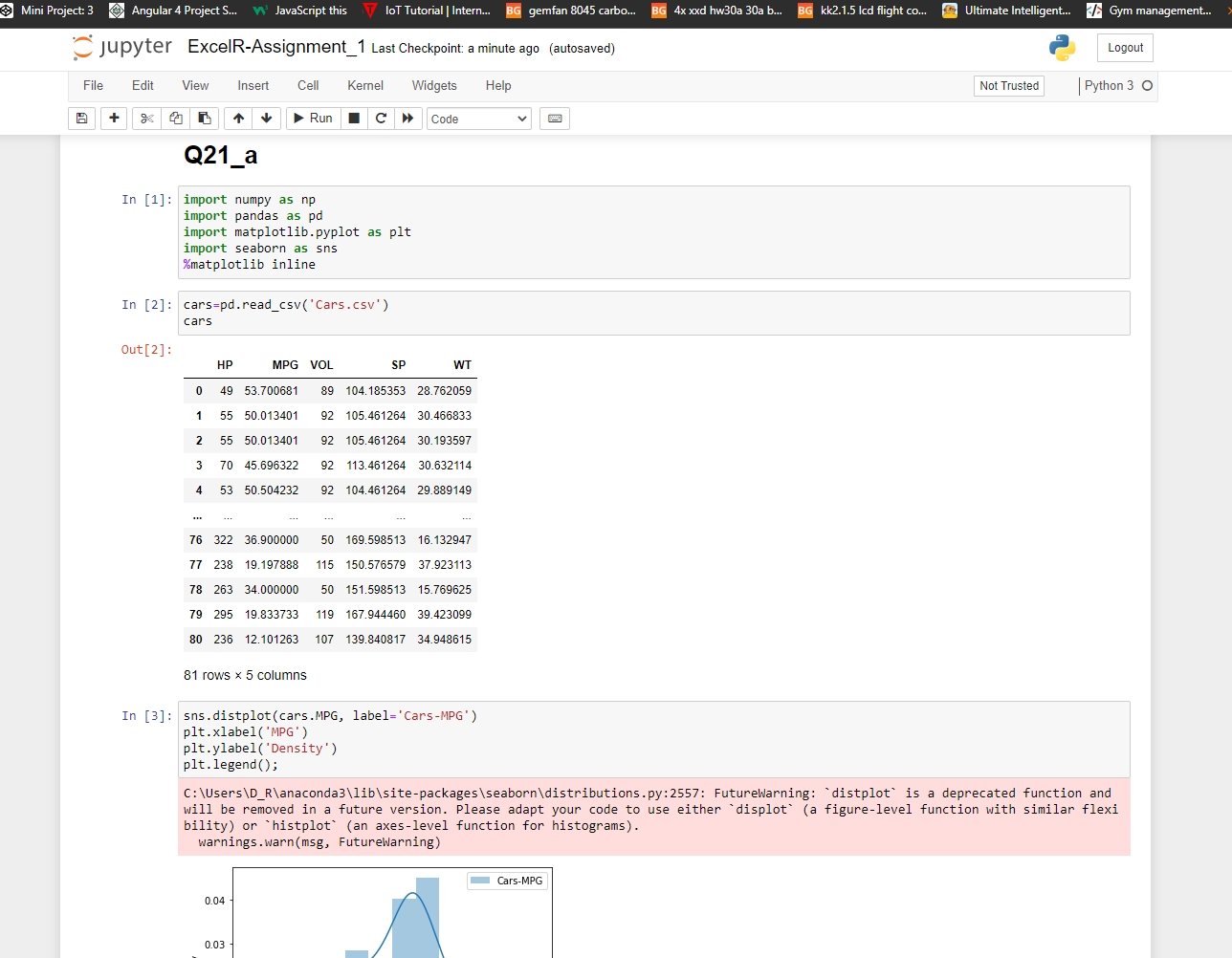
1. **Check whether the MPG of Cars follows Normal Distribution**

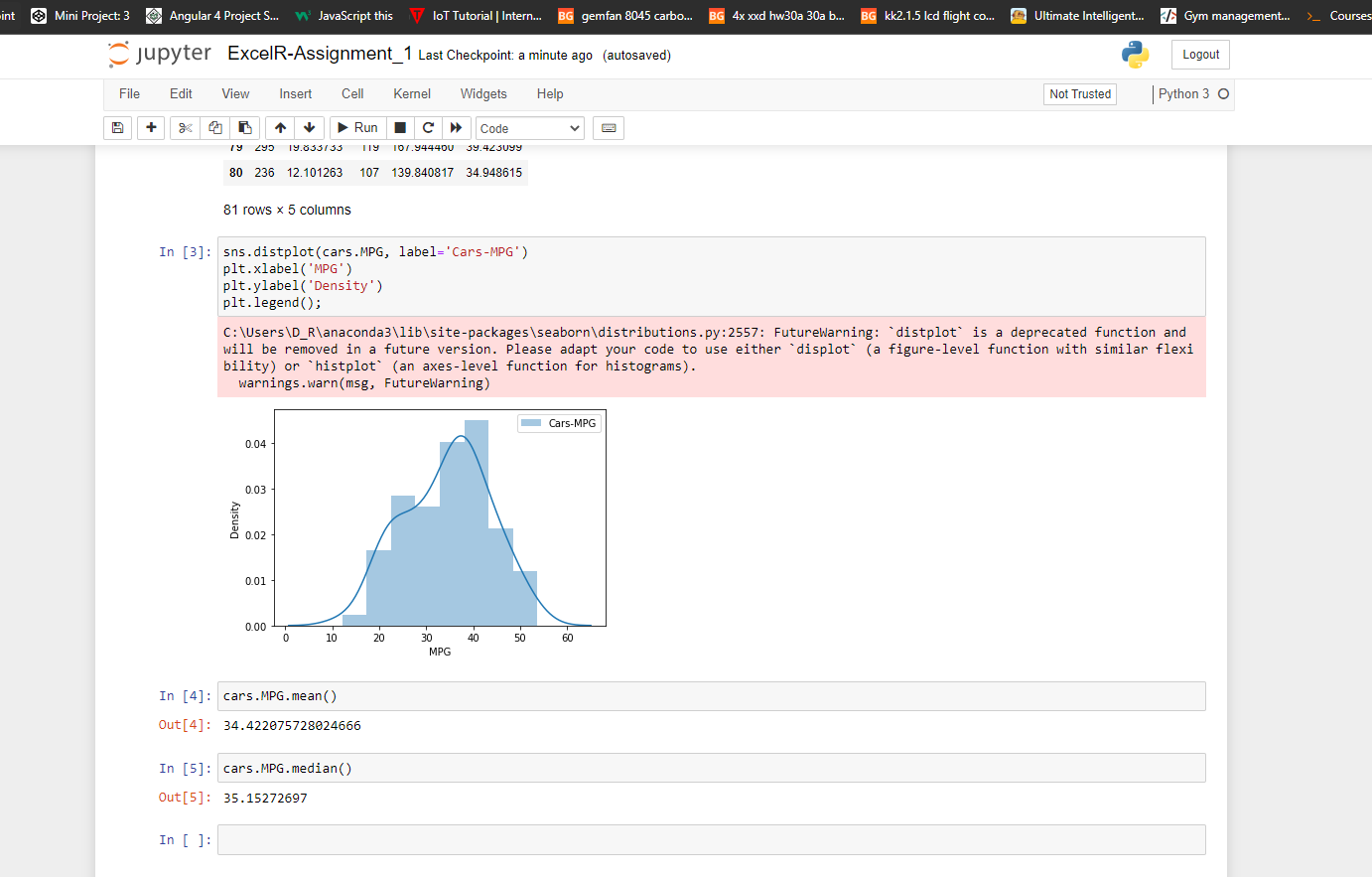
**Dataset: Cars.csv**

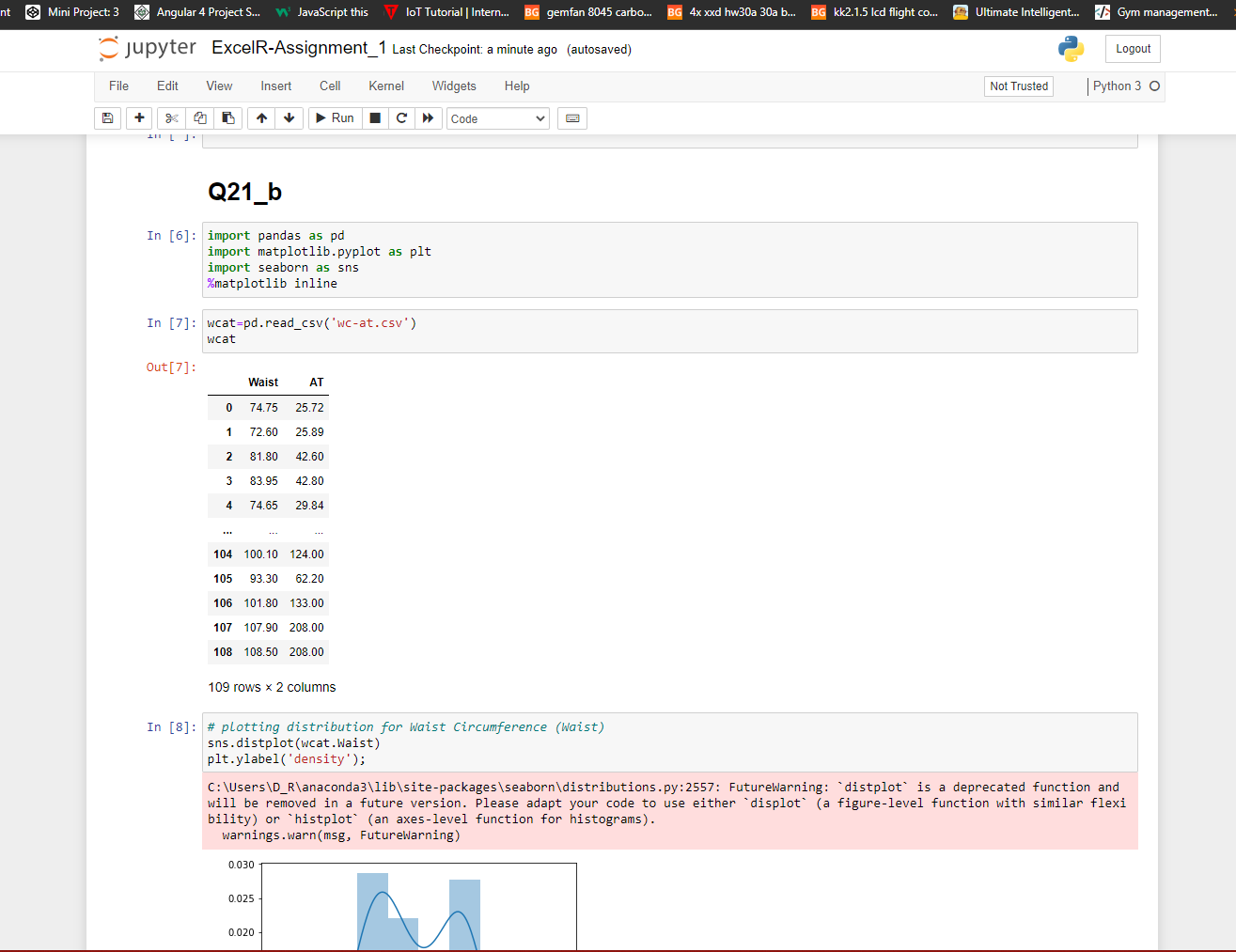
1. **Check Whether the Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set follows Normal Distribution**

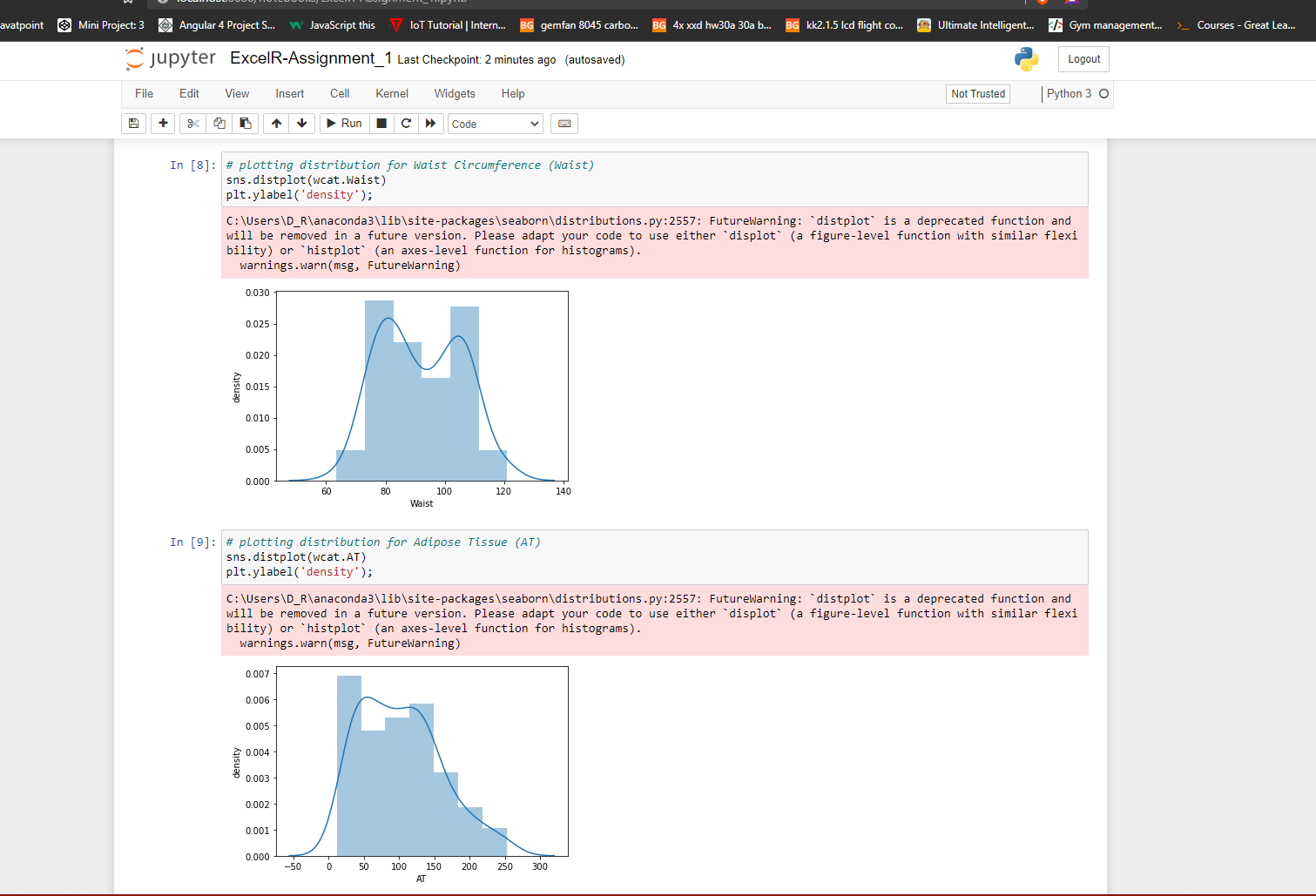
**Dataset: wc-at.csv**

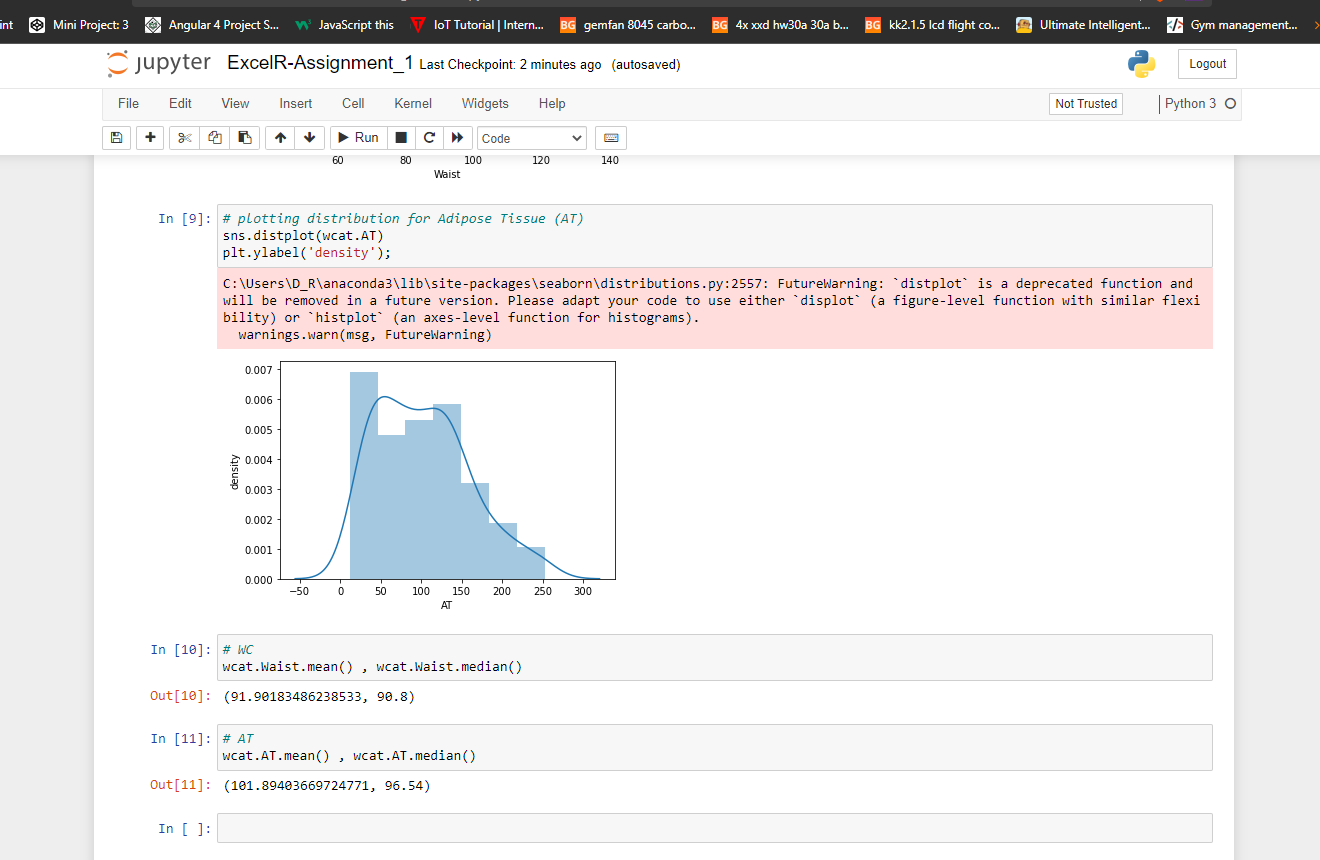
**Ans :**











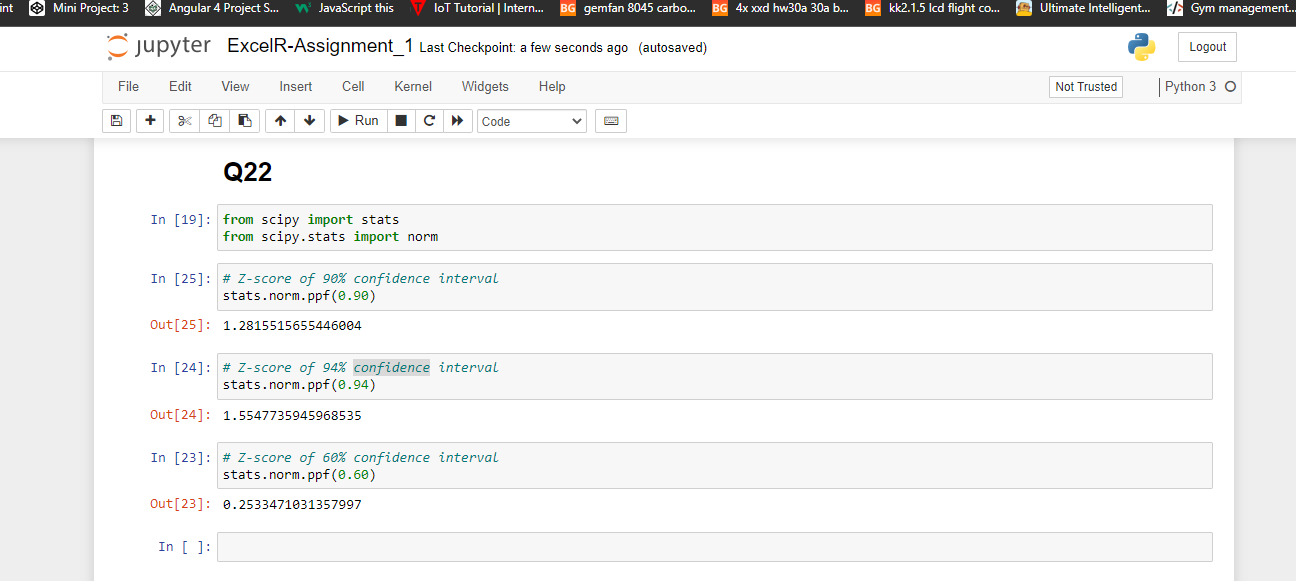
**Q 22) Calculate the Z scores of 90% confidence interval,94% confidence interval, 60% confidence interval**

**Ans :**

i) Z score of 90% confidence interval is 1.2

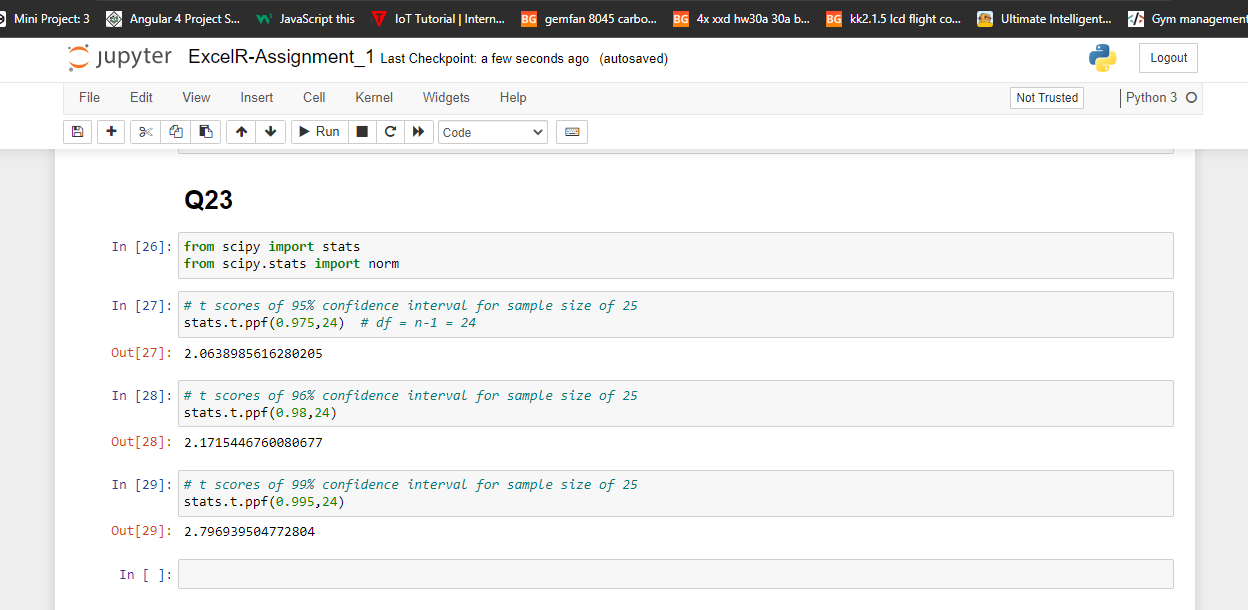
ii) Z score of 94% confidence interval is 1.5

iii) Z score of 60% confidence interval is 0.2



**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 :**

