|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | Discrete |
| Results of rolling a dice | Discrete |
| Weight of a person | Continues |
| Weight of Gold | Continues |
| Distance between two places | Continues |
| Length of a leaf | Continues |
| Dog's weight | Continues |
| 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 | Interval |
| Blood Group | Nominal |
| Time Of Day | Ratio |
| Time on a Clock with Hands | Ratio |
| Number of Children | Nominal |
| Religious Preference | Nominal |
| Barometer Pressure | Interval, Ordinal |
| SAT Scores | Interval |
| Years of Education | Interval |

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

**Ans**: {HHH, HHT, HTH, THH, TTH, THT, HTT, TTT} = 8

{HHT, HTH, THH} =3

=3/8

The Probability that two heads and one tail are obtained as 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:

{(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)}

Total probability is 36

1. Equal to 1 = 0/36
2. Less than or equal to 4= 6/36=1/6
3. Sum is divisible by 2 and 3 = 6/36 =1/6

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

=3.09

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

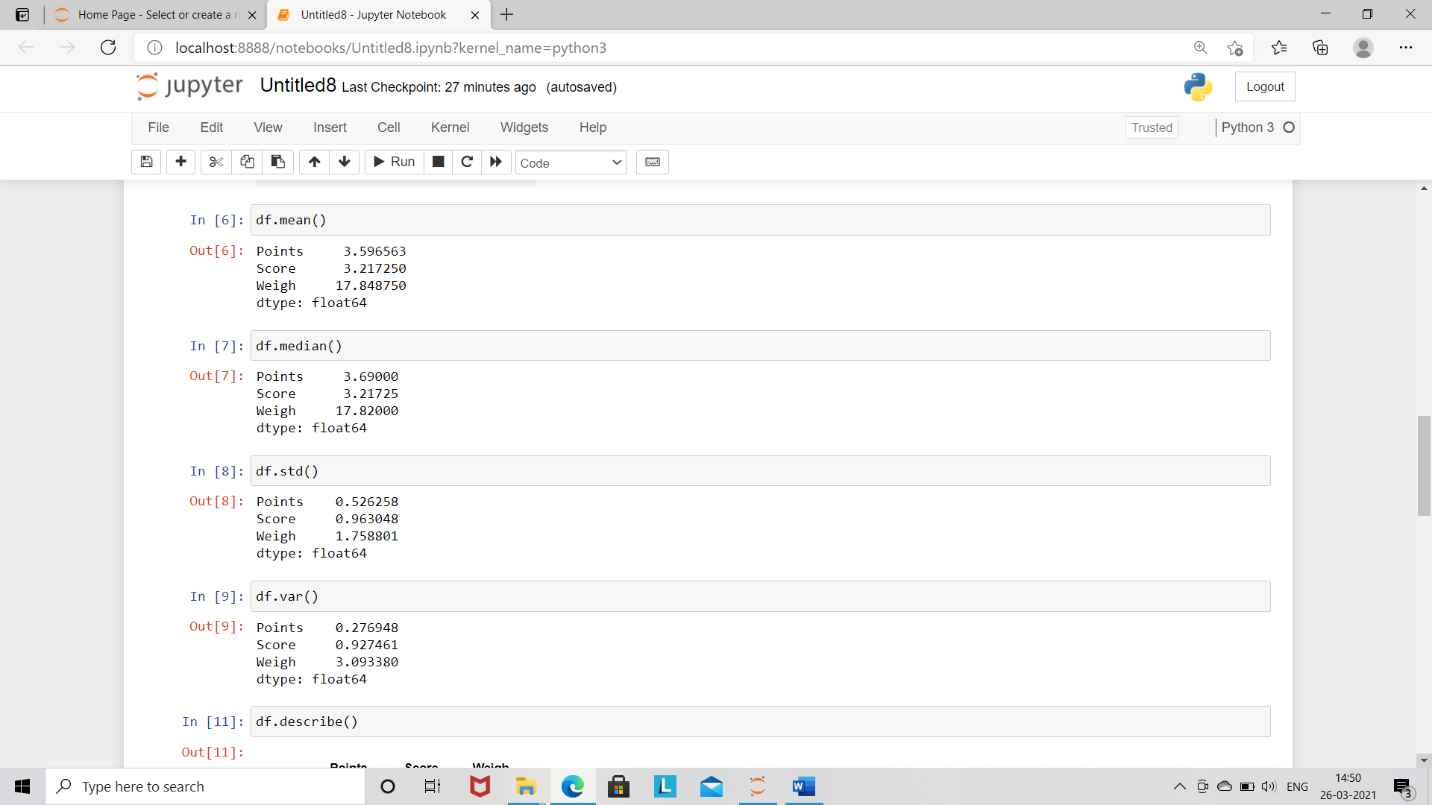
* For Points, Score, Weigh>

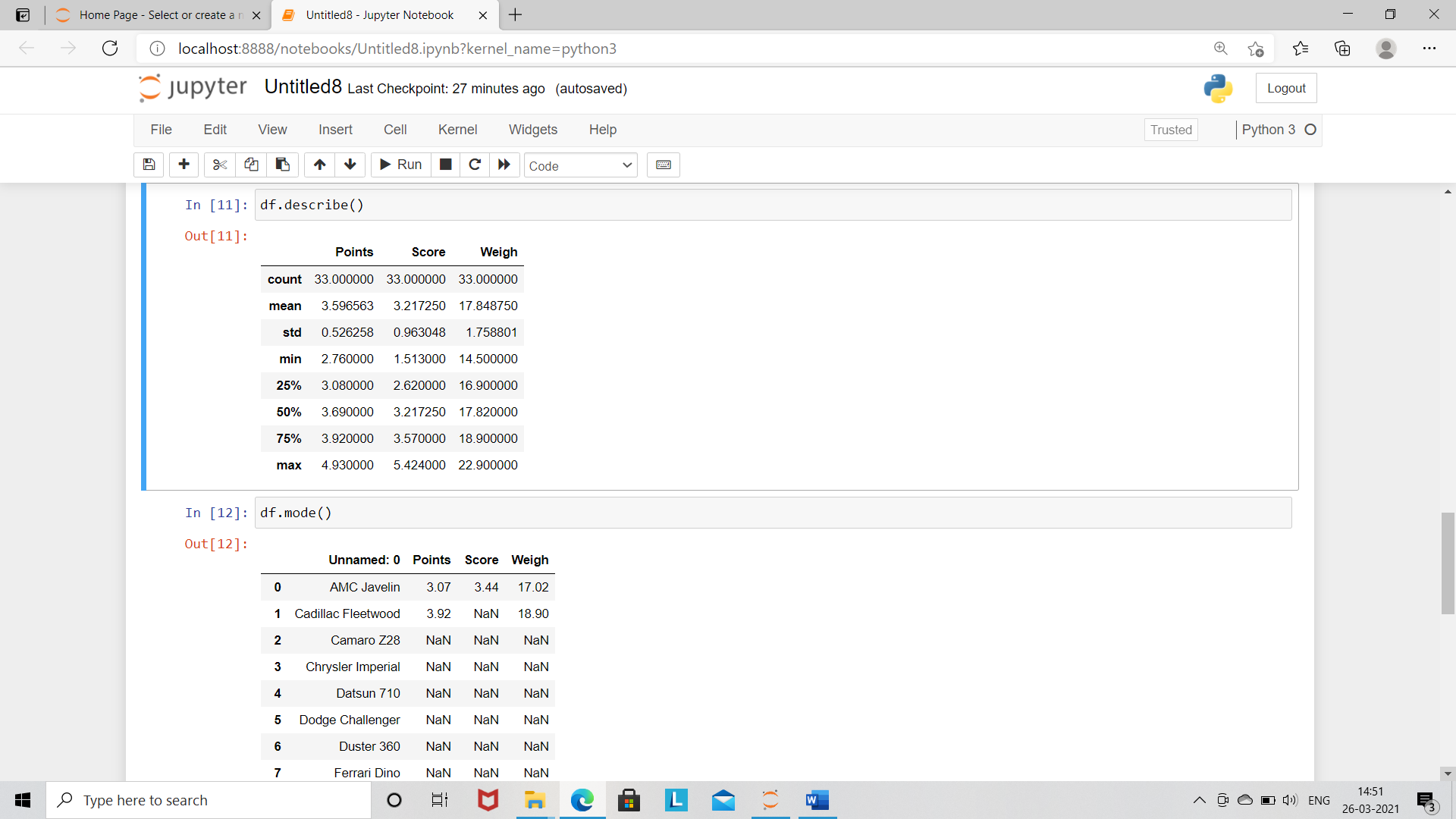
Find Mean, Median, Mode, Variance, Standard Deviation, and Range and also Comment about the values/ Draw some inferences.

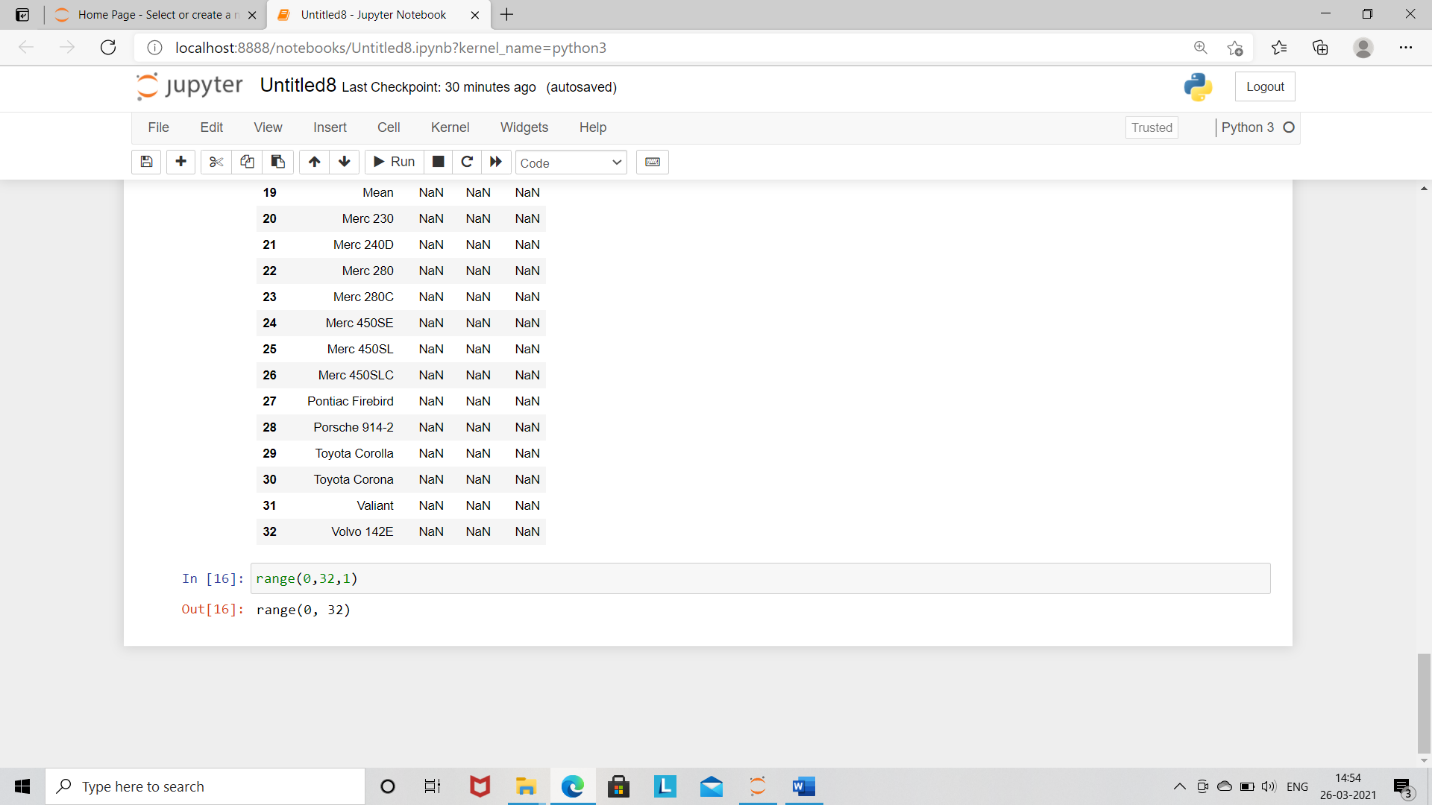
Use Q7.csv file

ANS:

|  |  |  |  |
| --- | --- | --- | --- |
|  | points | score | weigh |
| mean | 3.596563 | 3.217250 | 17.848750 |
| median | 3.69000 | 3.21725 | 17.82000 |
| Standard deviation | 0.526258 | 0.963048 | 1.758801 |
| variance | 0.276948 | 0.927461 | 3.093380 |
| mode | 3.92 | 3.44 | 17.02 |
| range | 2.17 | 3.911 | 8.4 |







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:**

Since we have nine total outcomes so probability of occurrence of each value selected randomly will be 1/9

=1/9 = 0.11

Therefore, the expected value of the weight of that patient will be

= sum (each occurrence – probability)

108\*0.11+110\*0.11+123\*0.11+134\*0.11+135\*0.11+145\*0.11+167\*0.11+187\*0.11+199\*0.11

Expected value will be =143.88

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

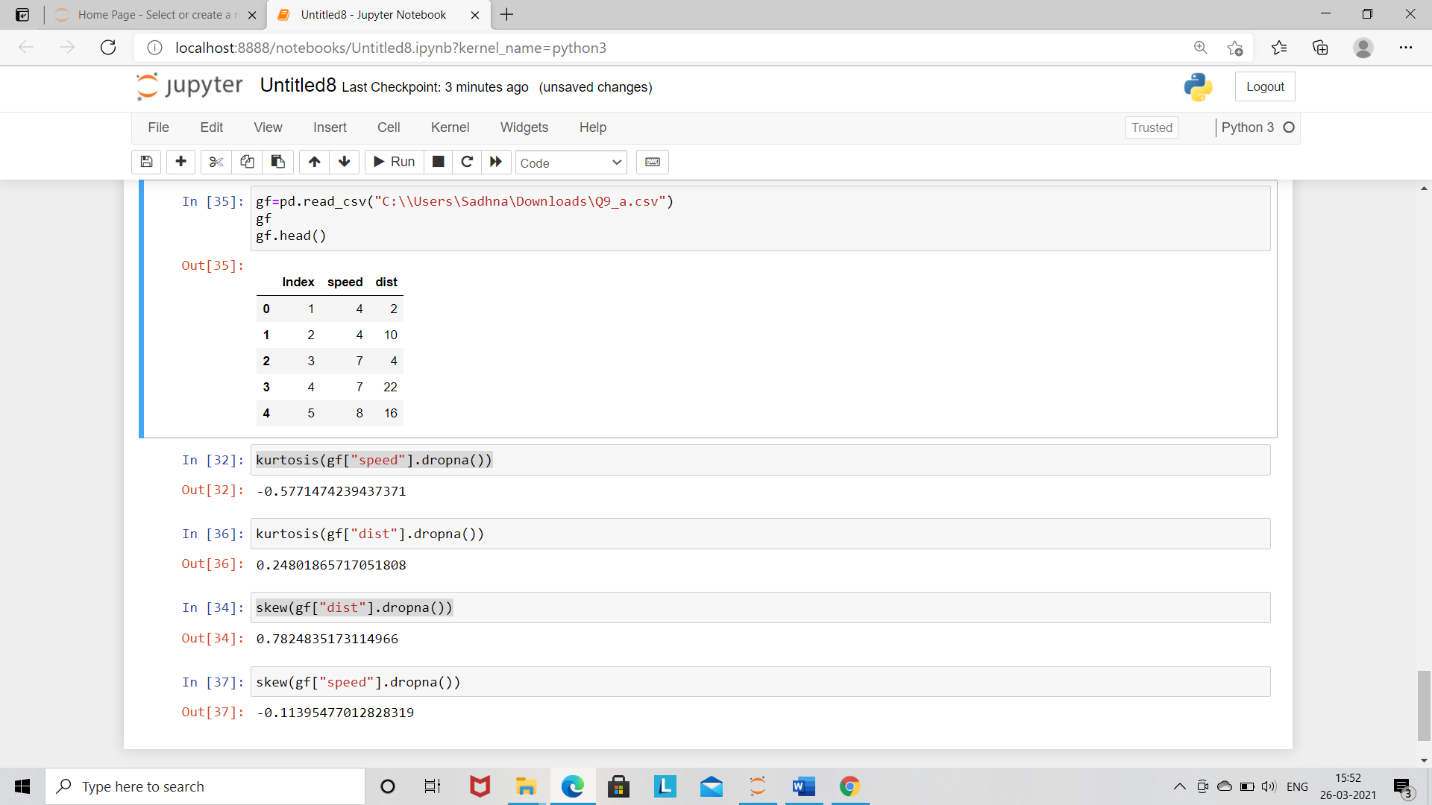
**Cars speed and distance**

**Use Q9\_a.csv**

**ANS:**

Skewness: dist.=0.78248351, speed=-0.11395477

Kurtosis: dist.=0.2480186, speed=-05771474



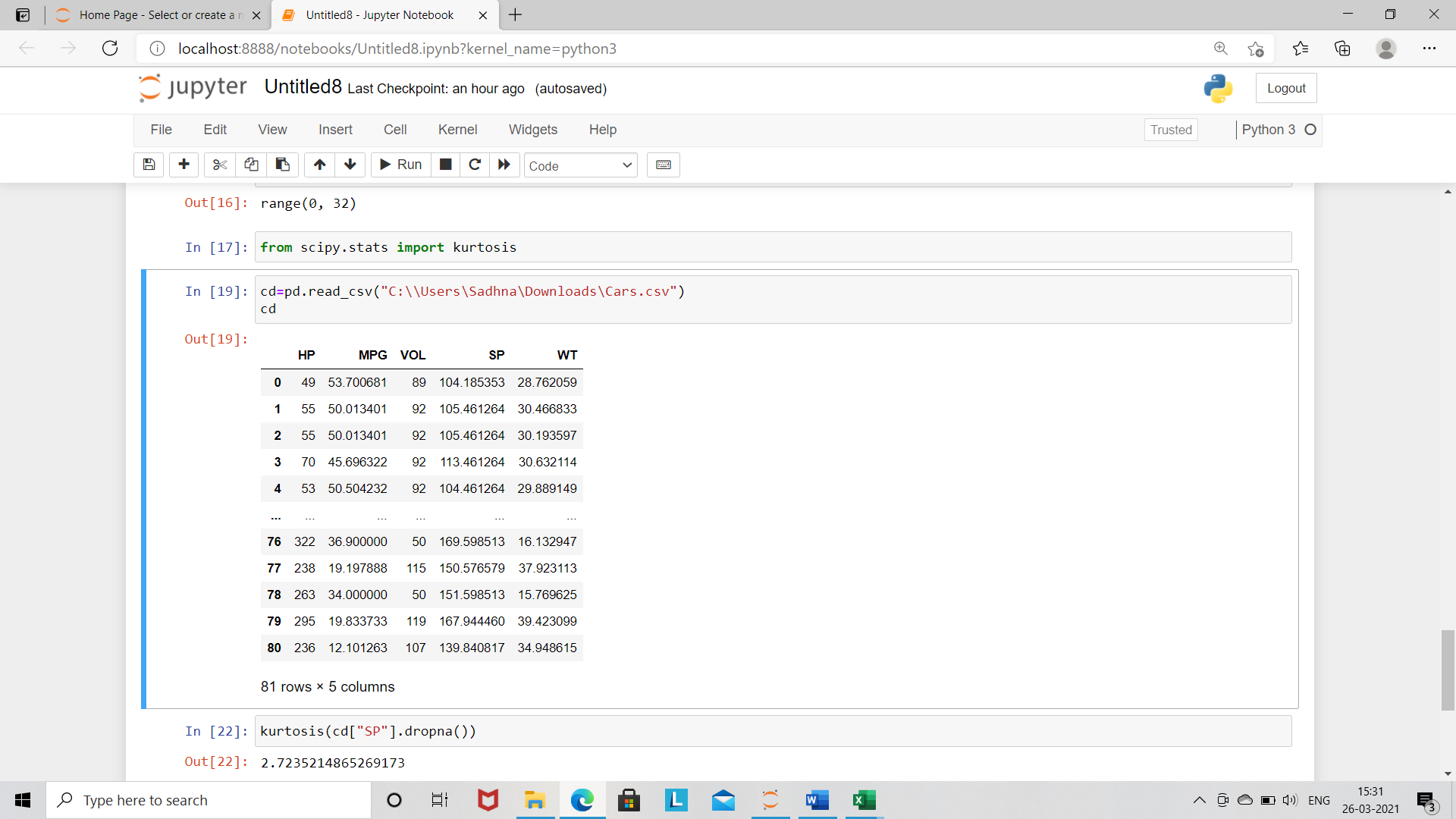
**SP and Weight (WT)**

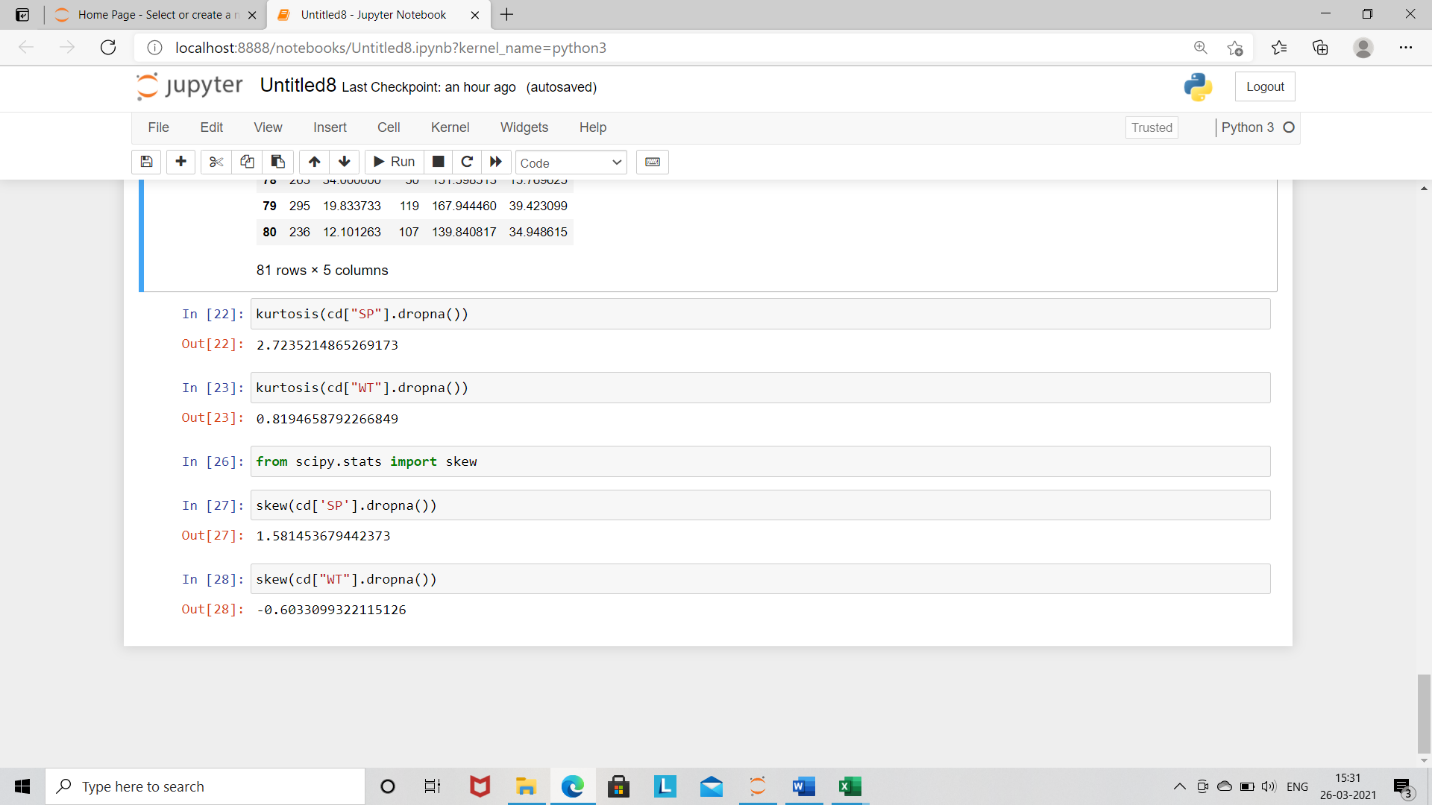
**Use Q9\_b.csv**

**Ans:**

Skewness: for SP = 1.58145367, WT= -0.6033099

Kurtosis: for SP = 2.723521, WT = 0.81946587





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



ANS: The above graph is the histogram graph. The graph plot between the frequency vs chick

Weight.

Data start from zero, its increases with respect to frequency then after 100 it falls down

slightly to 400 with respect to frequency. So, the graph called as “**Positive Skewness**".



Ans: The above graph is boxplot graph. There is a decrease value in the negative side of the graph and increase value in the positive side of the graph. So, we observed that outliers present at the positive side of the graph.

**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,

n=2000

sample mean=200

sample std=30

find=confidence interval of 94%,98%,96%

here we see sample is greater than 30 therefore we use central limit

theorem

std error= std/sqrt(n)= 30/sqrt (2000)

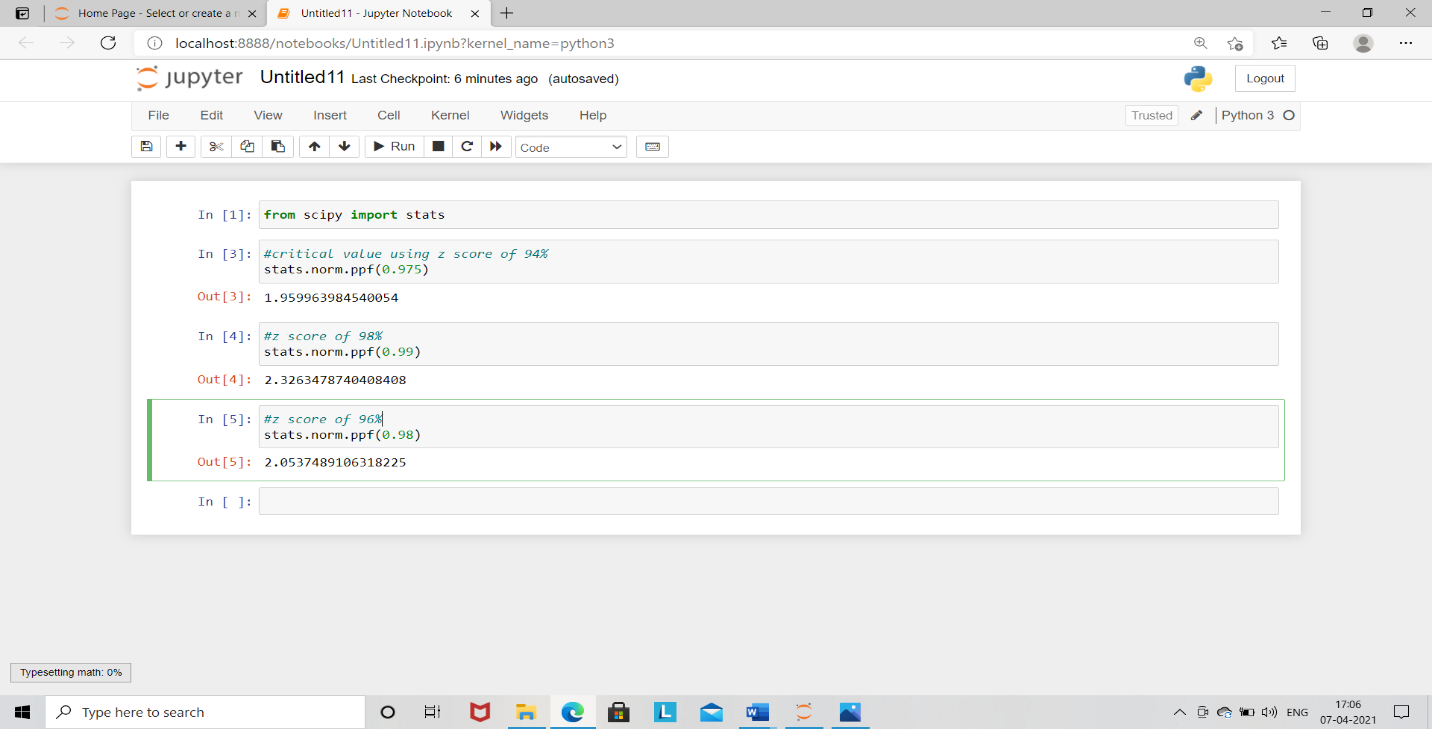
=0.670

significance= 1-alpha

1. 94%=1-94/100=0.06
2. 98%=1-98/100=0.02
3. 96%=1-96/100=0.04

Critical value

1. 94%=1.96
2. 98%=2.326
3. 96%=2.053



Margin error = critical value\*std error

1. 94%=1.96\*0.670=1.3132
2. 98%=2.326\*0.670=1.55842
3. 96%=2.053\*0.670=1.375

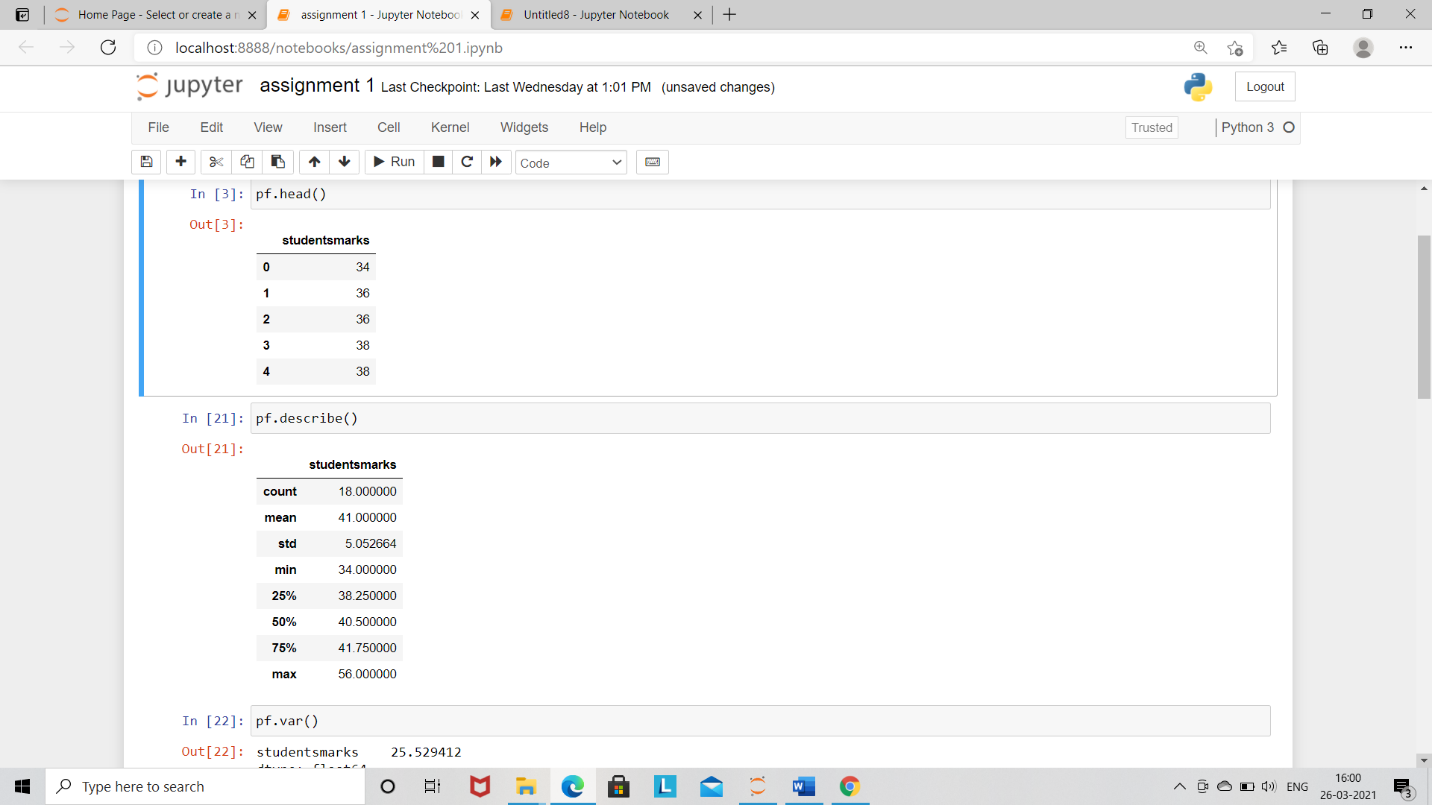
**Q12)** Below are the scores obtained by a student in tests

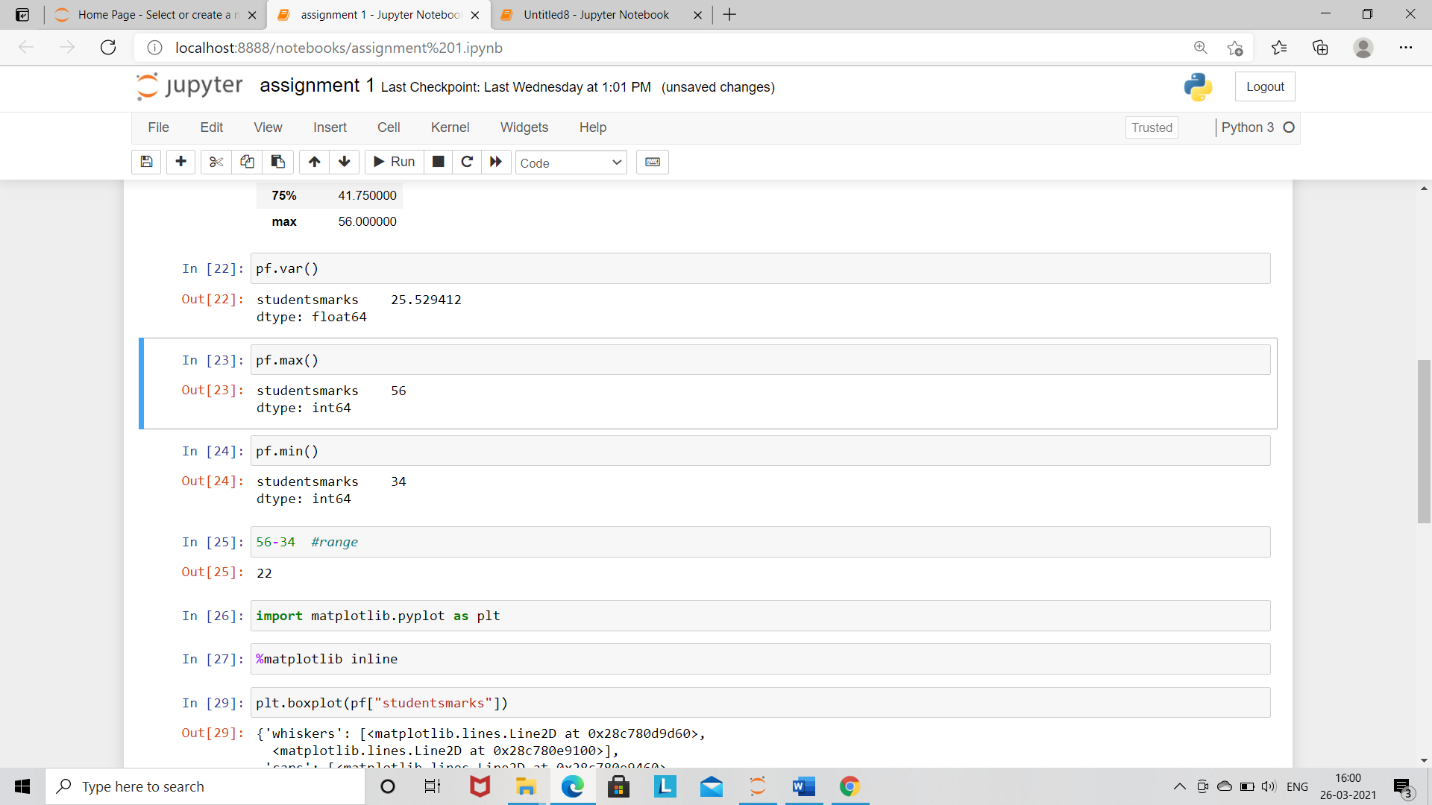
**34,36,36,38,38,39,39,40,40,41,41,41,41,42,42,45,49,56**

1. Find mean, median, variance, standard deviation.

ANS:

|  |  |
| --- | --- |
| Mean | 41 |
| Median | 40.5 |
| Variance | 25.529 |
| Standard deviation | 5.052 |





Q2) What can we say about the student marks?

Ans: There is very huge difference in last two scores compare to first 16

Scores. So, there is a possibility of outliers.

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

Ans: The nature of skewness is zero skewness

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

Ans: The nature of skewness is positive skewness

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

Ans: the nature of skewness is negative skewness

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

Ans: positive kurtosis value indicates the peakness of distribution

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

Ans: Negative kurtosis value indicates the flat distribution

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



1. What can we say about the distribution of the data?

Ans: Most of the data is toward left hand side from the median as compared

to right hand side.

1. What is nature of skewness of the data?

Ans: the nature of above boxplot is negative skewness

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

Ans: IQR= q3-q1

= 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:

1. The two boxes do not overlap with each other, say, box 1 is placed right side of box 2.
2. There are no outliers in box 1 and box 2.
3. These are median, the middle values of both boxes are same.
4. The whiskers show how big a range there is between those two extremes. Box2: larger ranges indicate wider distribution, that is, more scattered data.

Box1: shorter range indicate less scattered data.

1. From box1 the data is skewed left, box2 data is zero skewness.

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

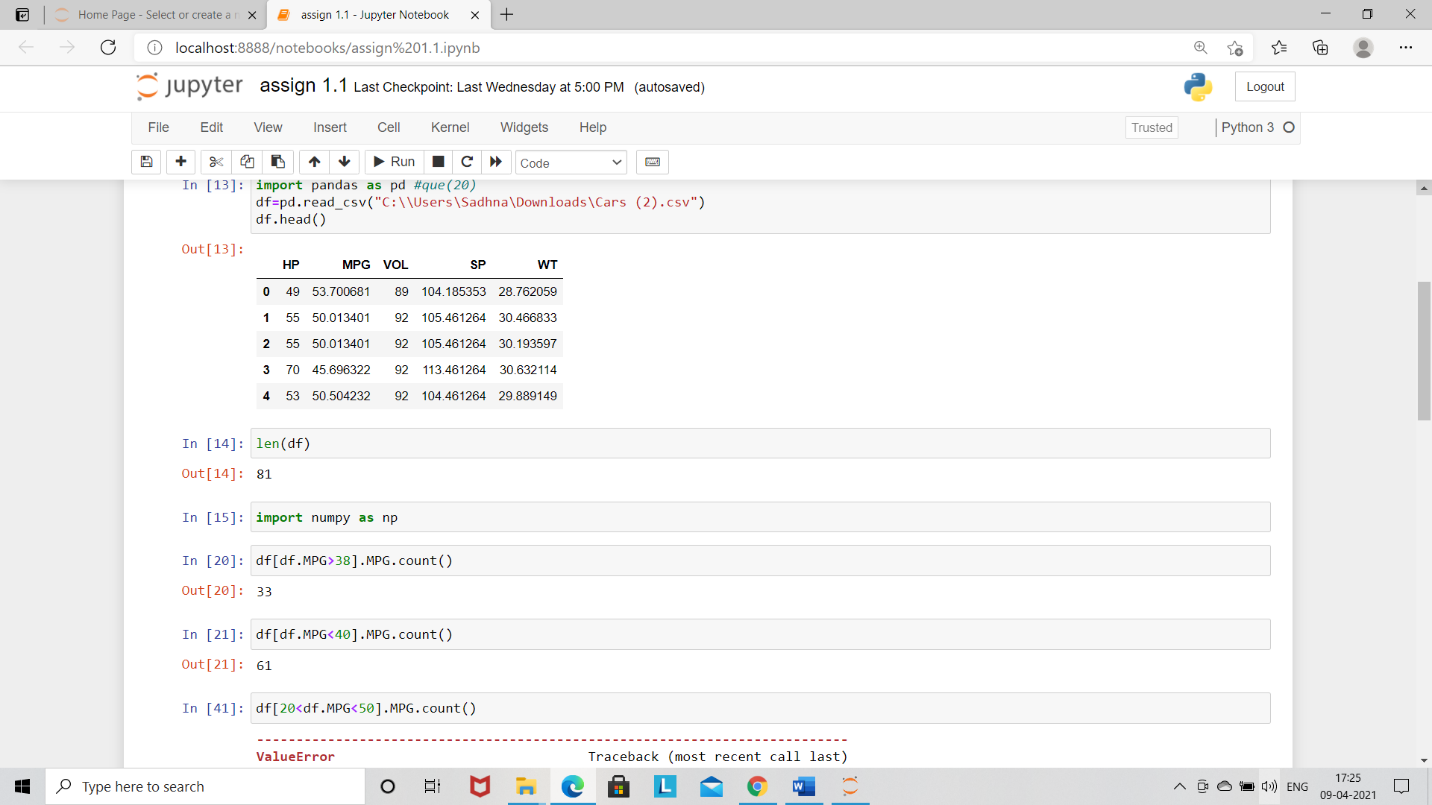
Data \_set: Cars.csv

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)



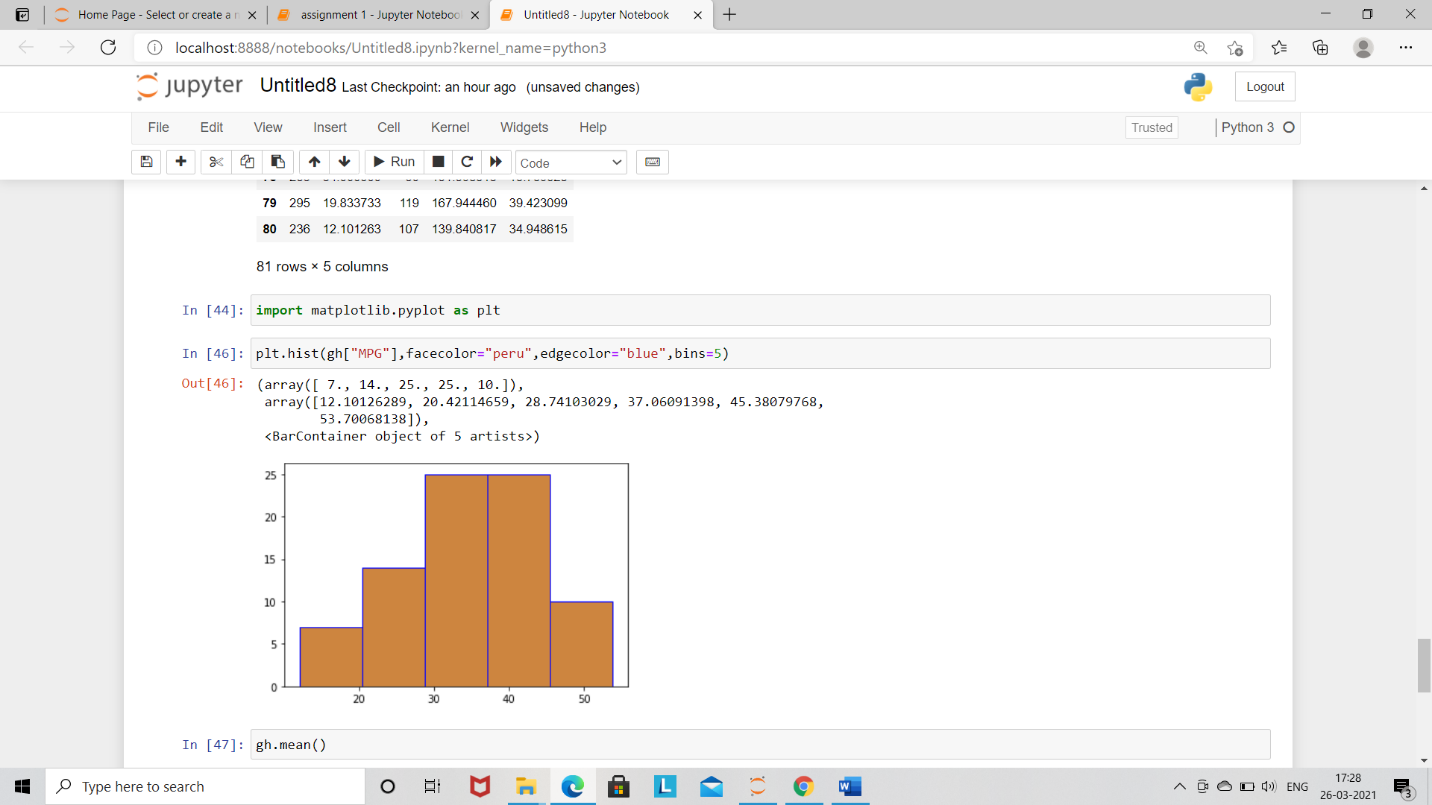
Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

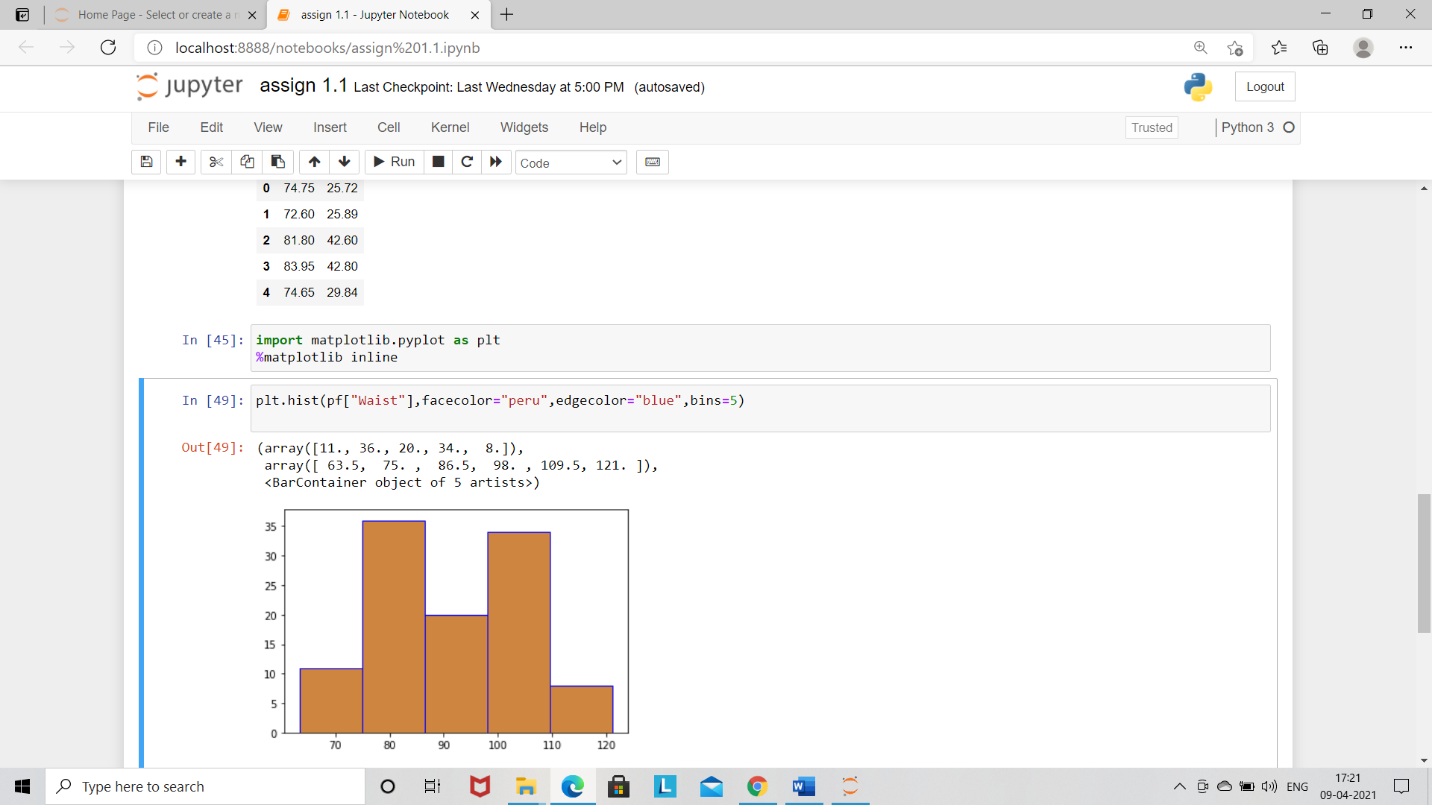
ANS:

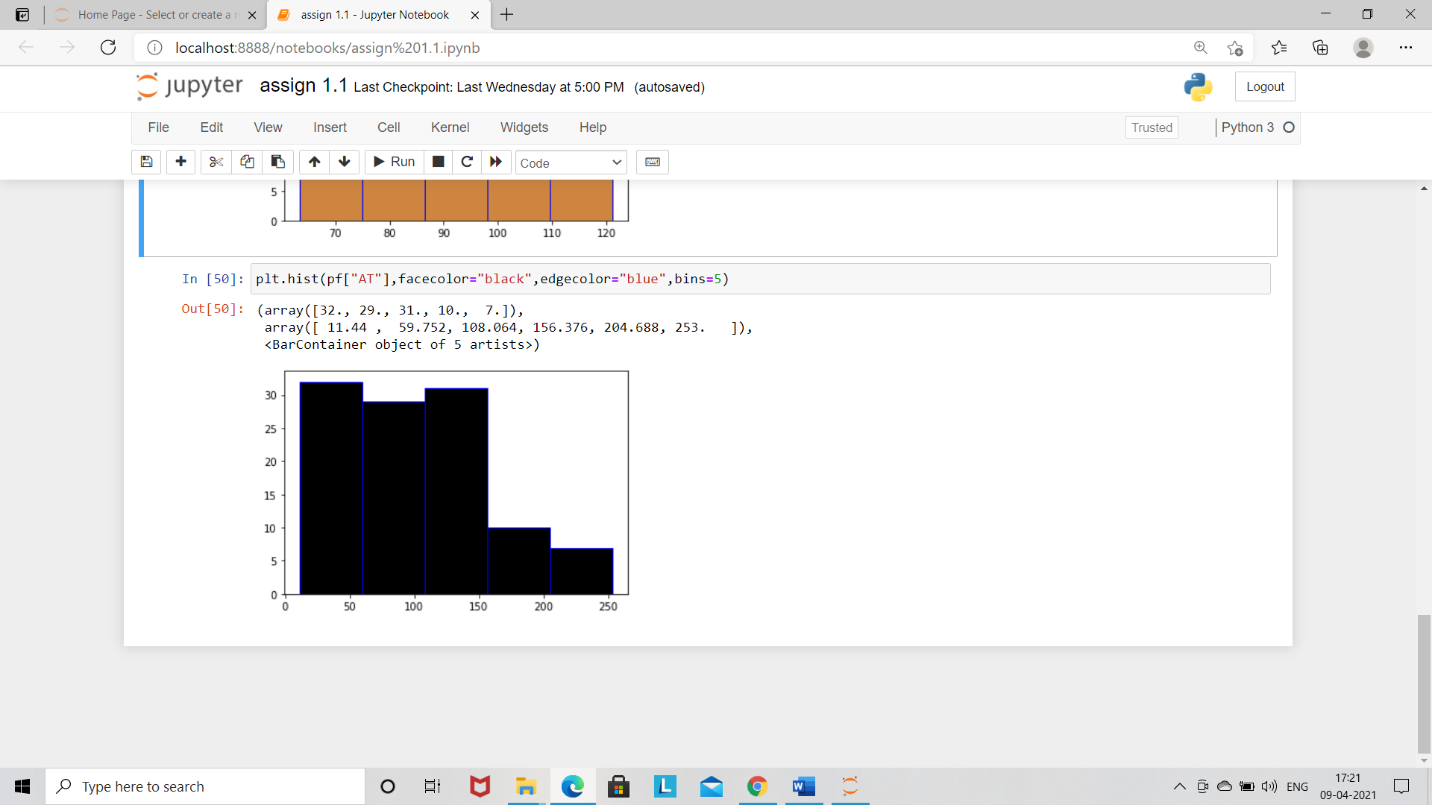
MPG of cars does not follow normal distribution because mean, median, mode is not equal to each other and skewness is not equal to zero.



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

Dataset: wc-at.csv





Both are not following a normal distribution.

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

Ans: Formula

1. Z score of 90% = 0.9+0.05

=0.95

By using z table= 0.8531

=1.95

1. Z score of 94% = 0.94+0.03

= 0.97

By using z table= 0.8340

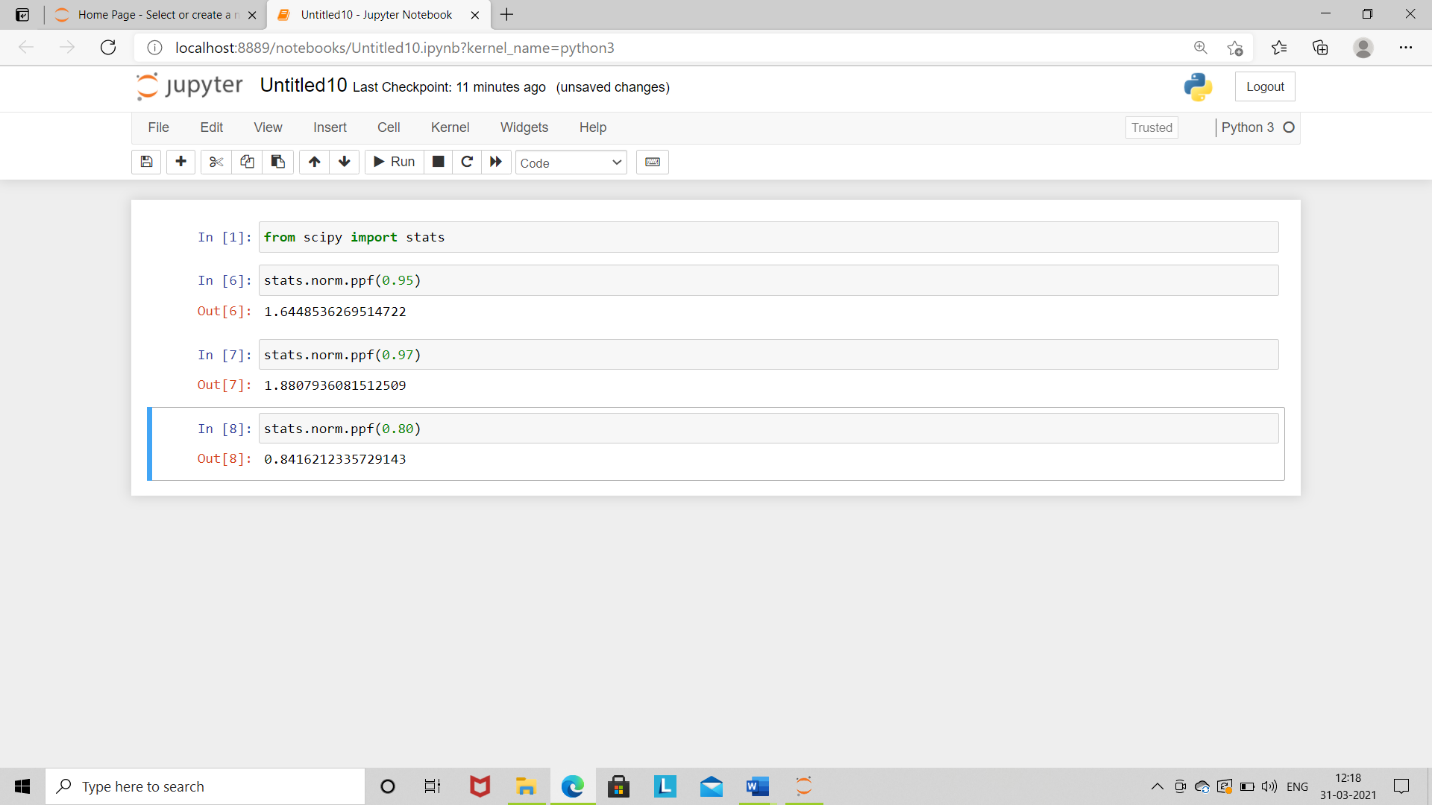
=1.834

1. Z score of 60% = 0.60+0.20

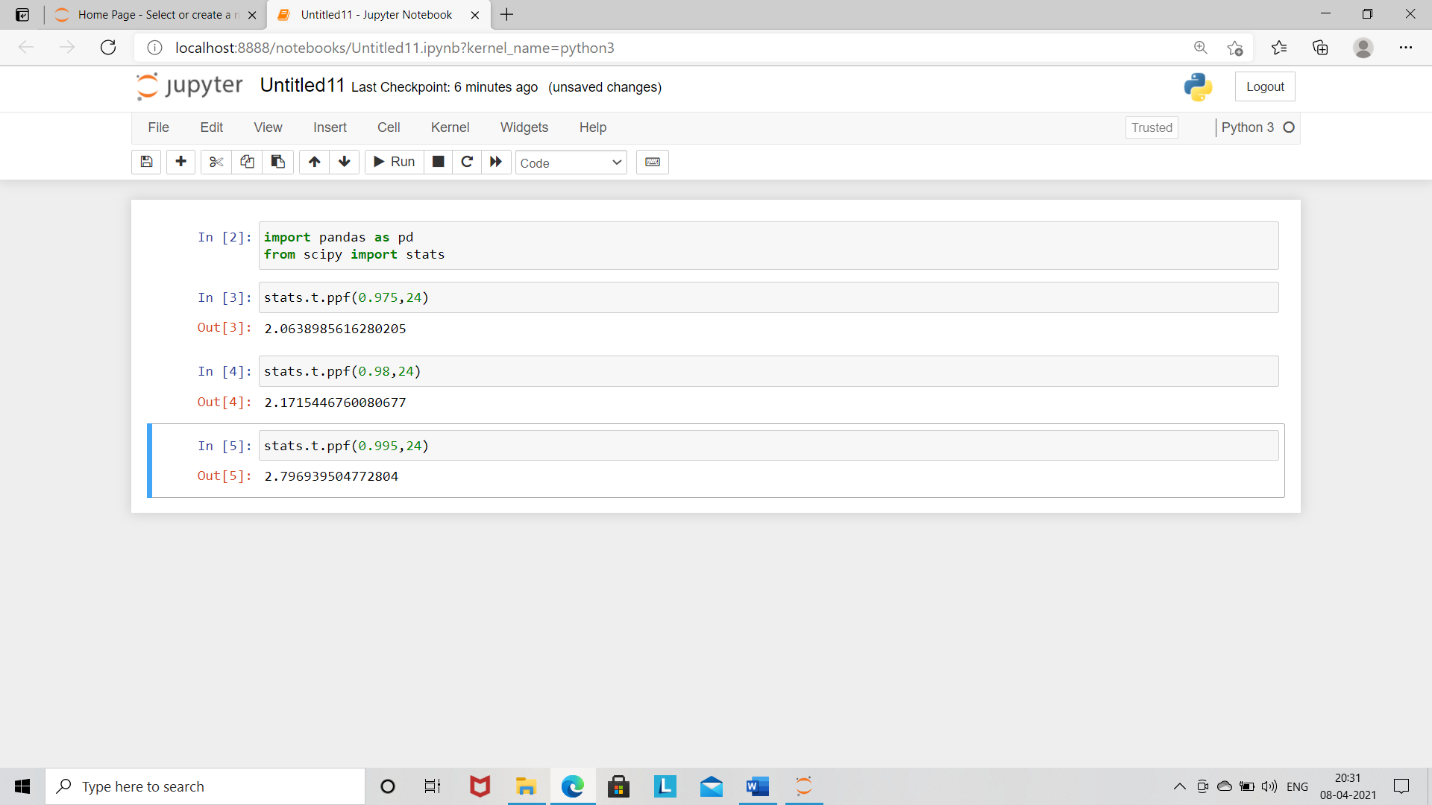
=0.80

By using z table= 0.7881

=1.898



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

Ans

T score for 95%= 2.063

96%=2.171

99%=2.7969

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: x = 260

Mu = 270

S = 90

 N = 18

t=(x-mu)/(s/sqrt(n))

t

[1]-0.4714045

Prob=pt(t,df=17)

Prob

[1]0.3216725

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