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

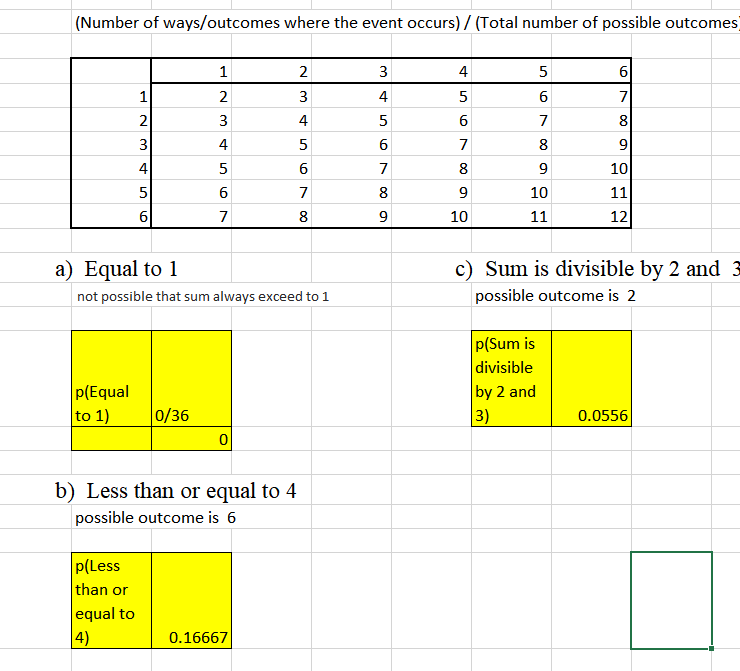
Q3) Three Coins are tossed, find the probability that two heads and one tail are obtained?

|  |
| --- |
|  |

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

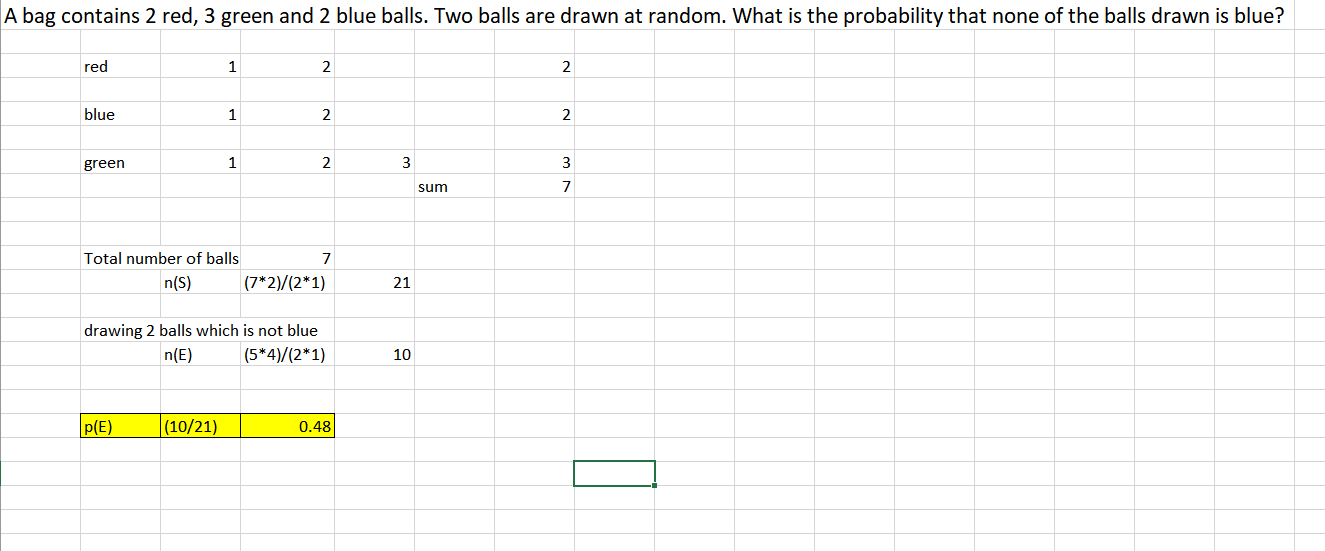
**Ans.**



**Plz refer Assignment calculation excel sheet**

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.**



**Plz refer Assignment calculation excel sheet**

Q6) Calculate the Expected number of candies for a randomly selected child

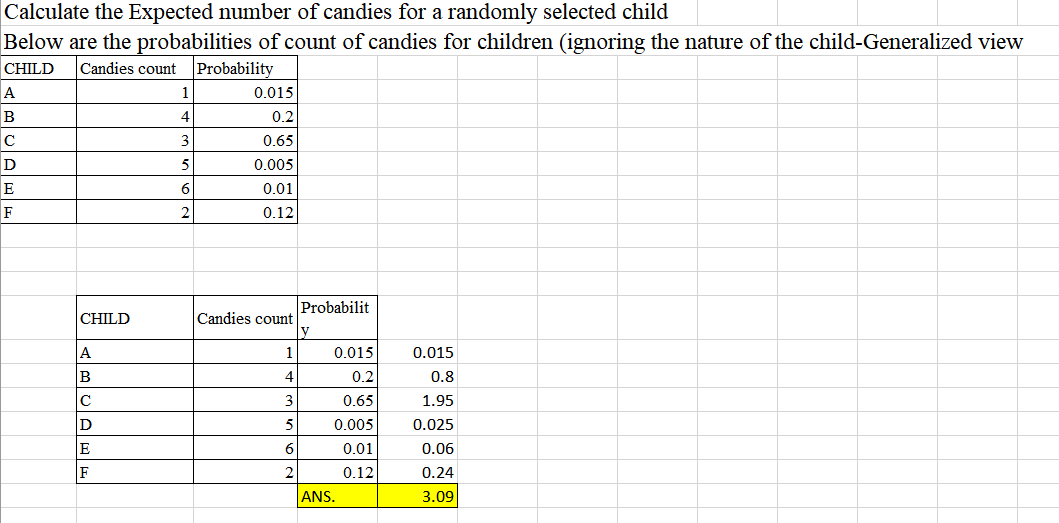
Below are the probabilities of count of candies for children (ignoring the nature of the child-Generalized view)

|  |  |  |
| --- | --- | --- |
| 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.**



**Plz refer Assignment calculation excel sheet**

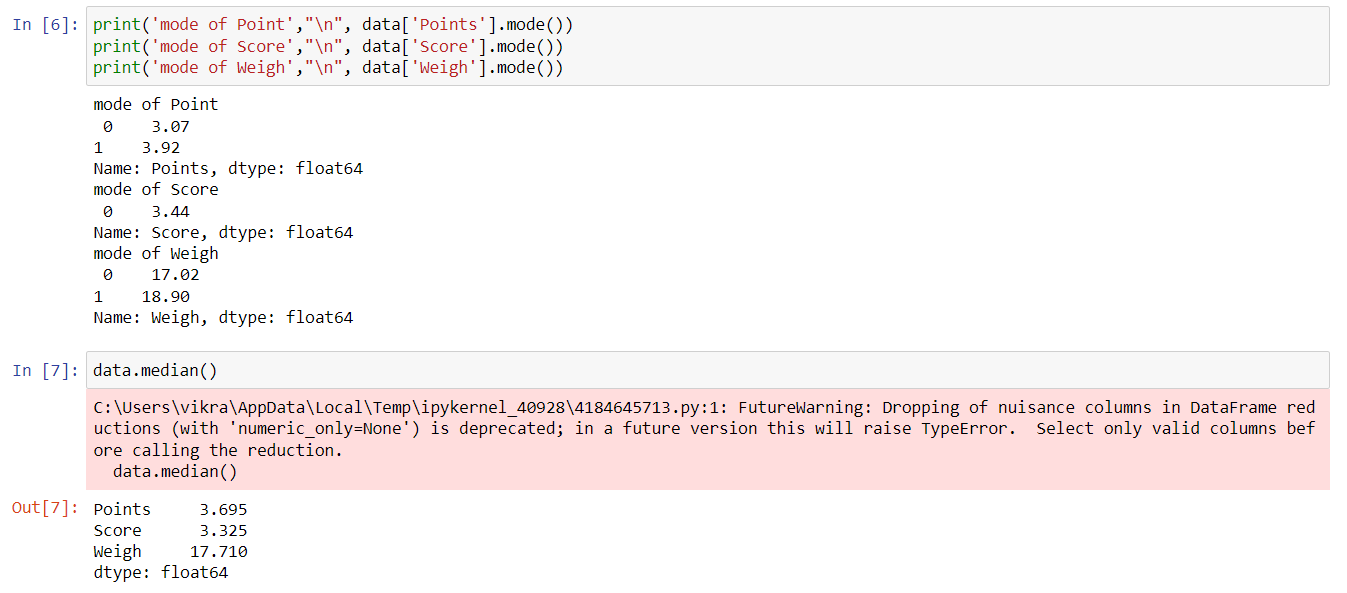
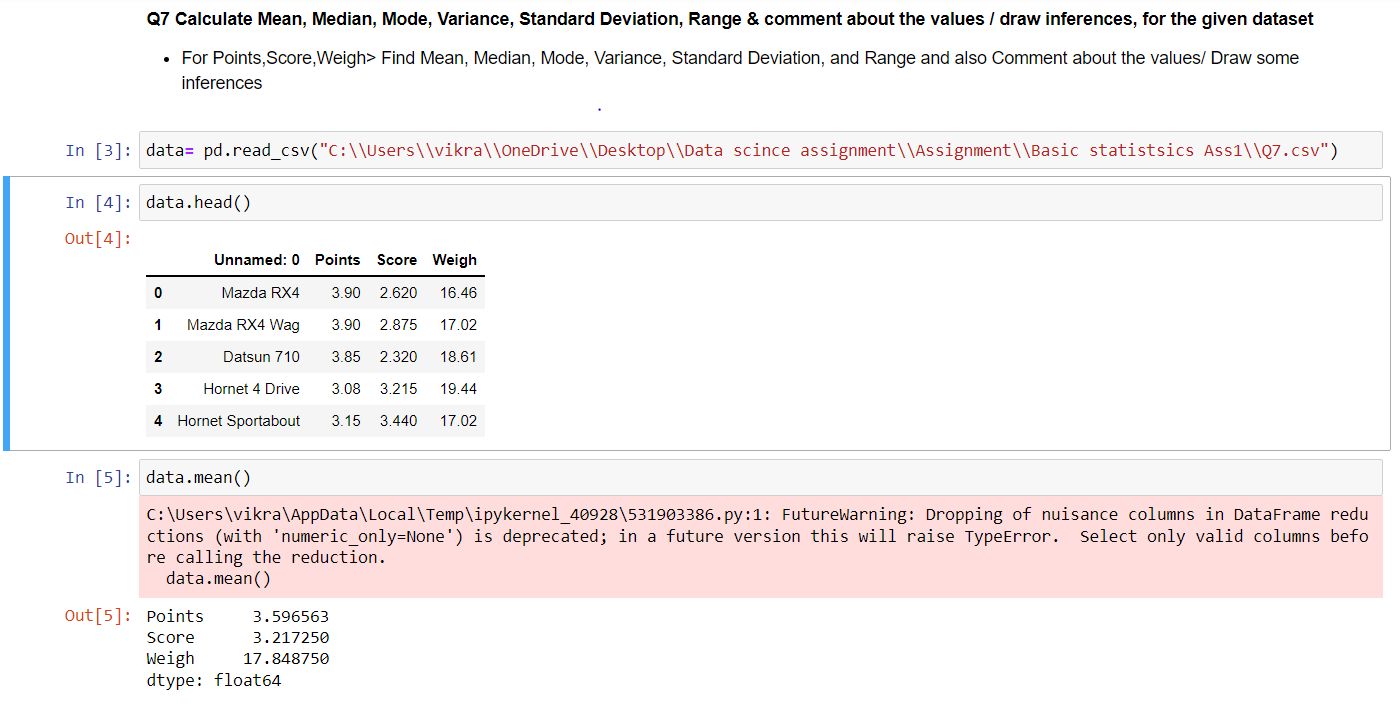
Q7) Calculate Mean, Median, Mode, Variance, Standard Deviation, Range & comment about the values / draw inferences, for the given dataset

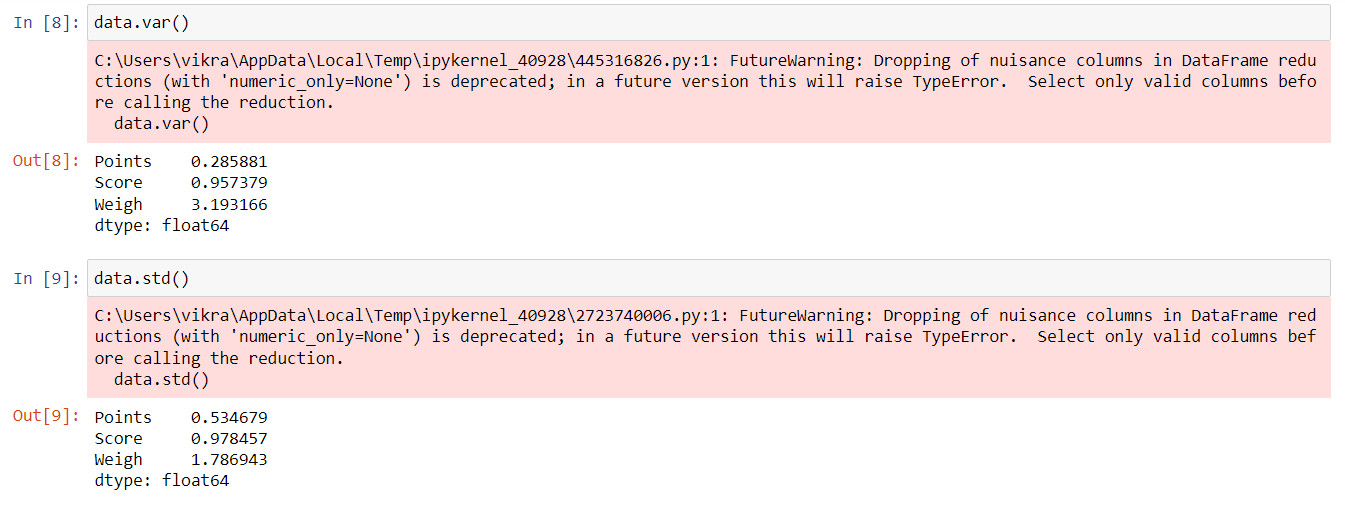
* For Points,Score,Weigh>

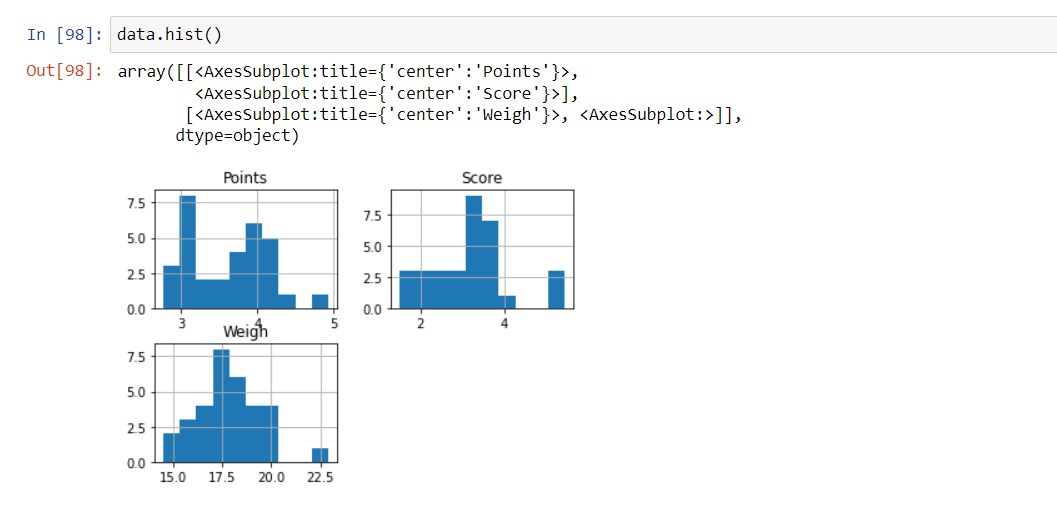
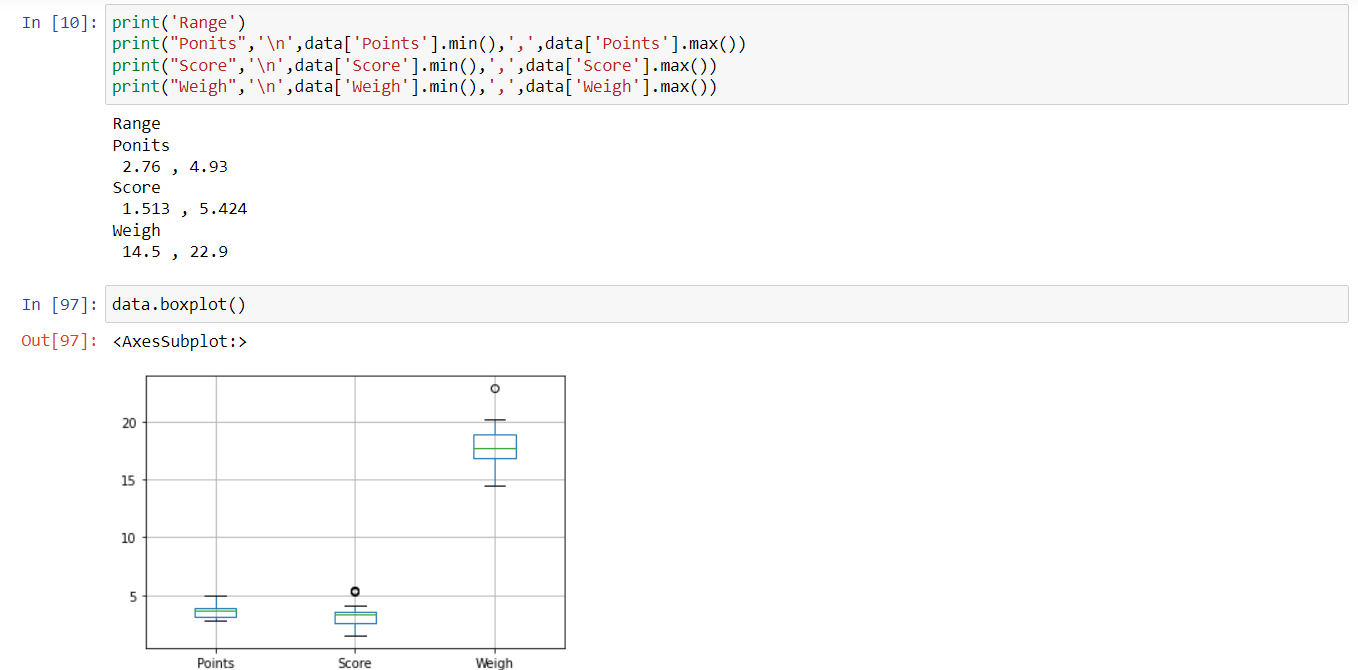
Find Mean, Median, Mode, Variance, Standard Deviation, and Range and also Comment about the values/ Draw some inferences.

**Use Q7.csv file**

**Ans**







**Plz refer the Data science assignment 1 python file.**

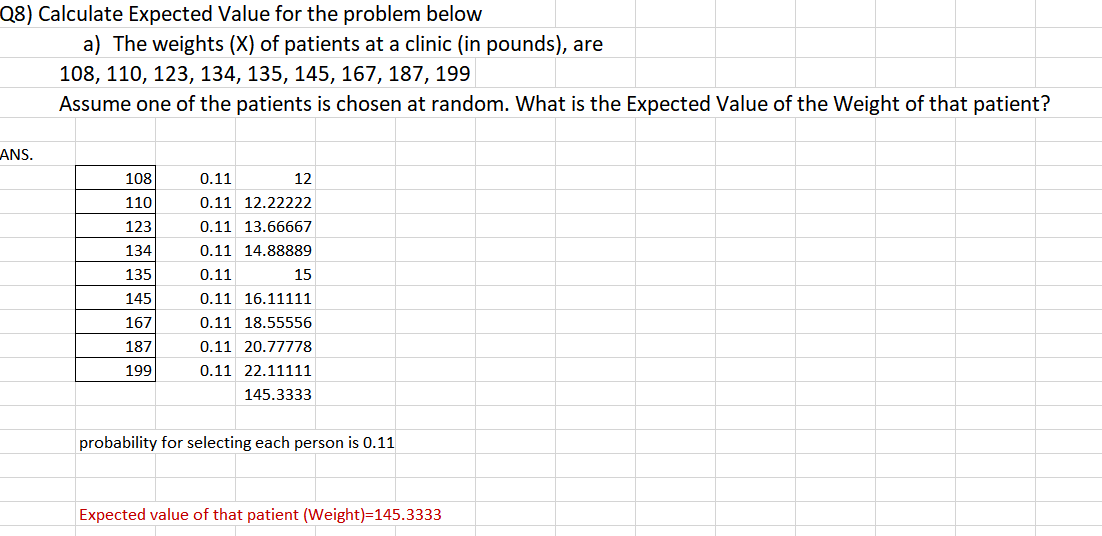
Q8) Calculate Expected Value for the problem below

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.**



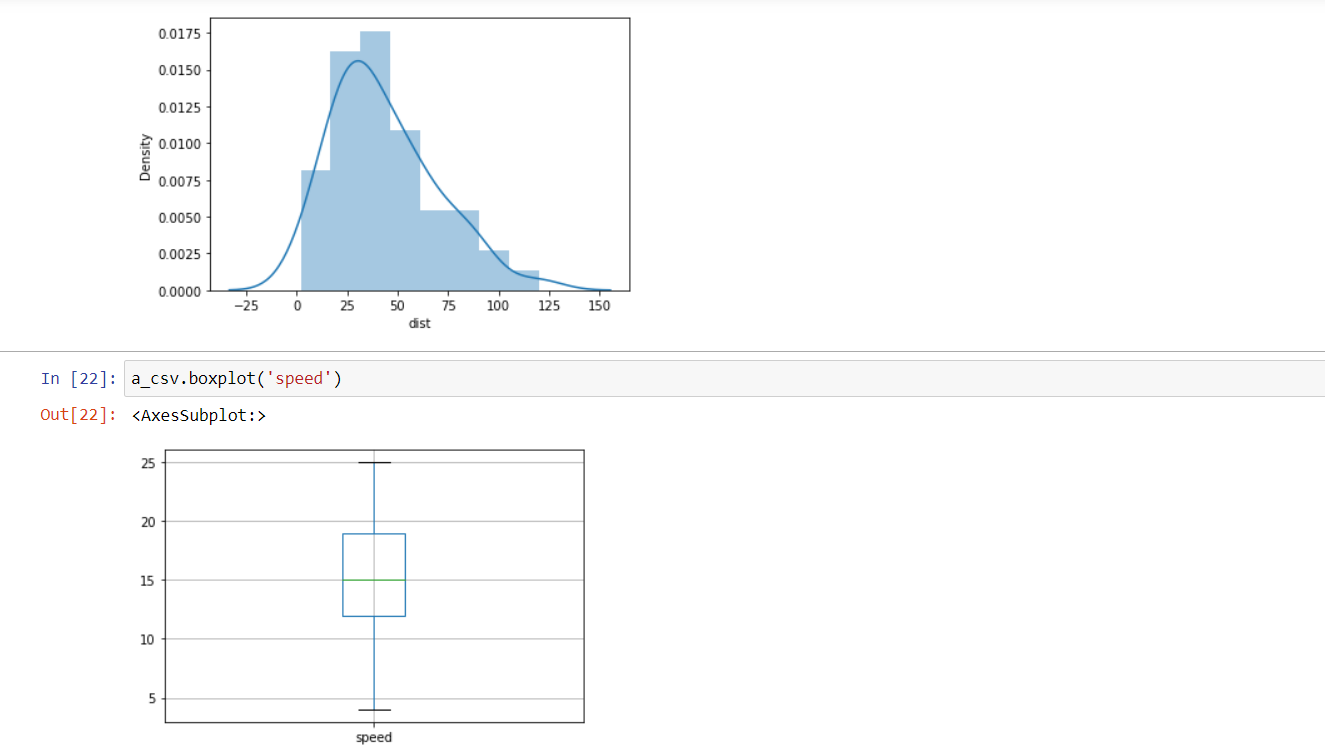
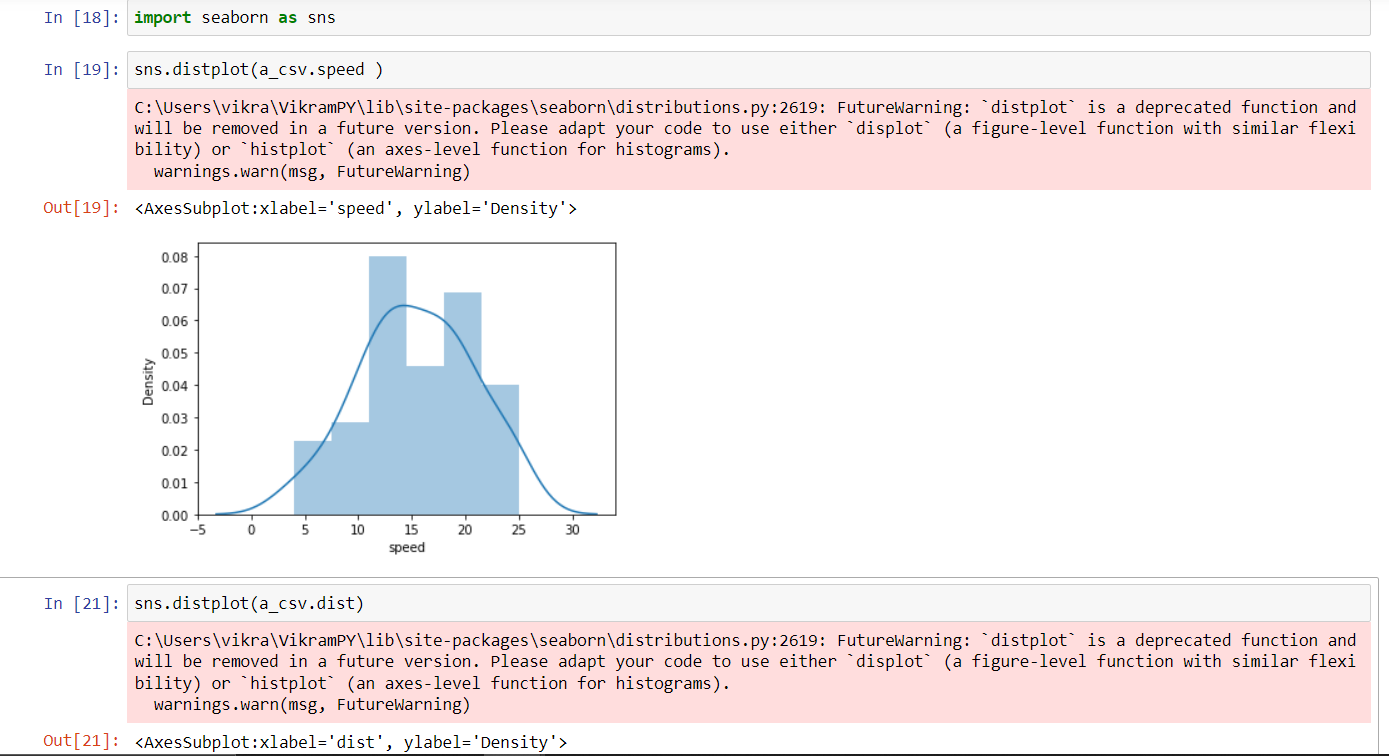
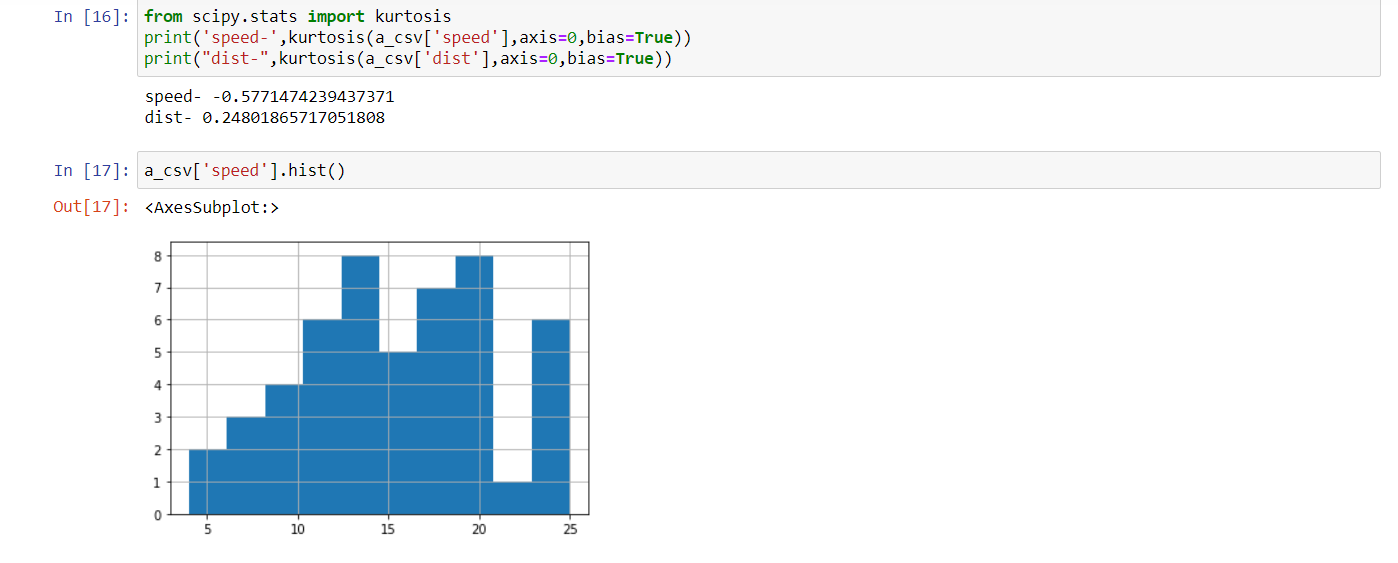
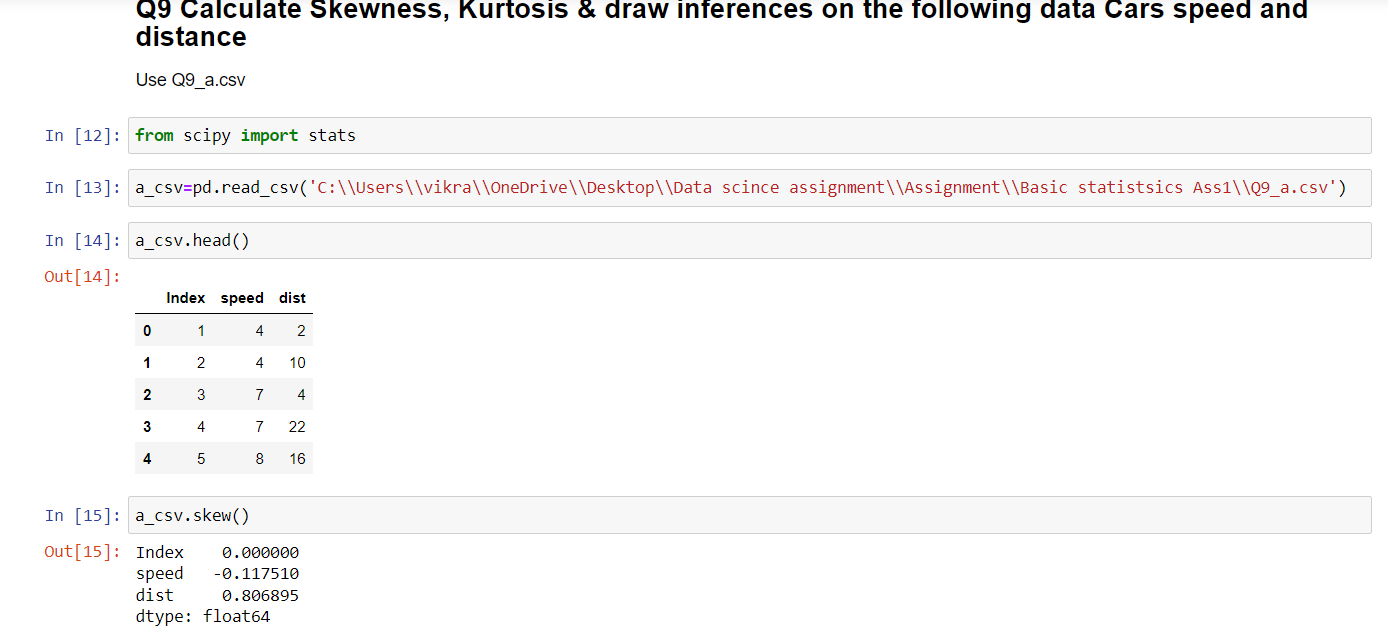
**Plz refer Assignment calculation excel sheet**

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

**1.Cars speed and distance**

**Use Q9\_a.csv**

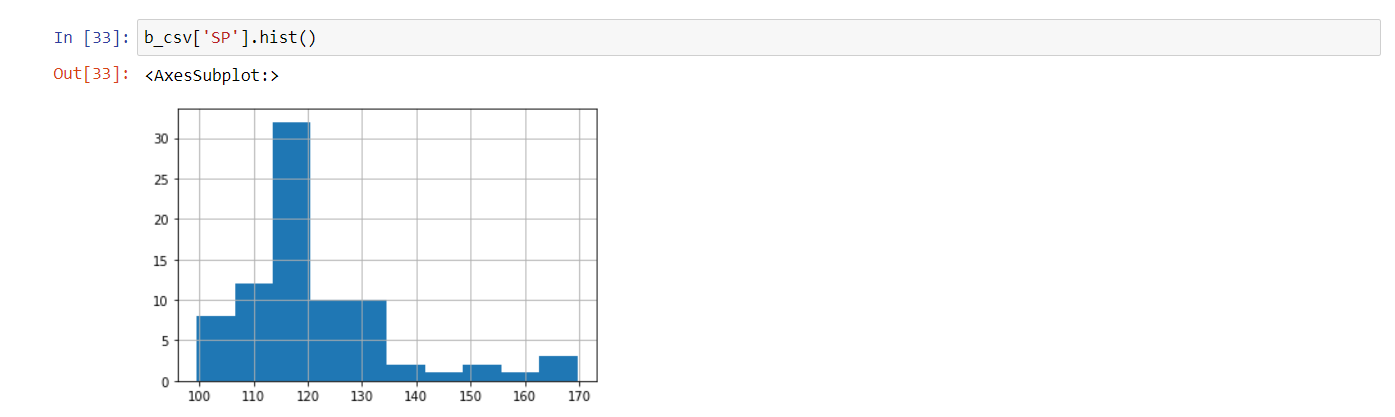
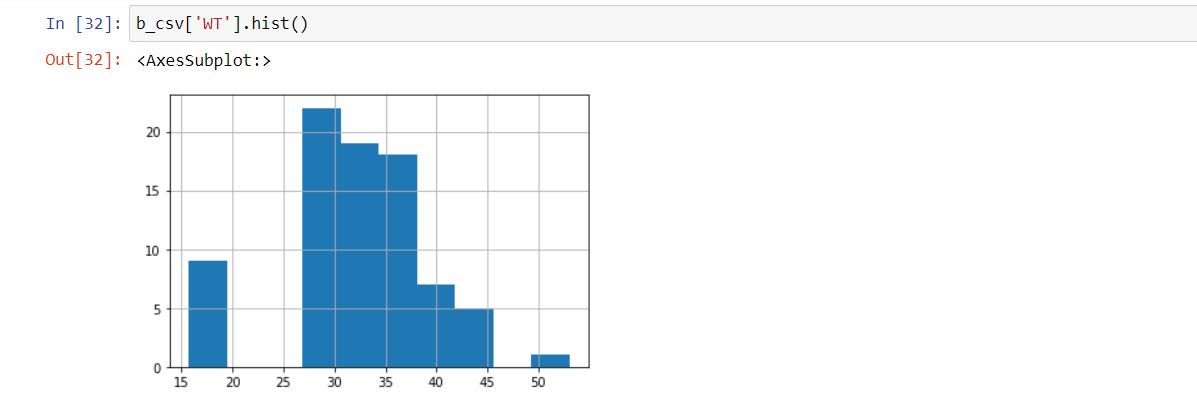
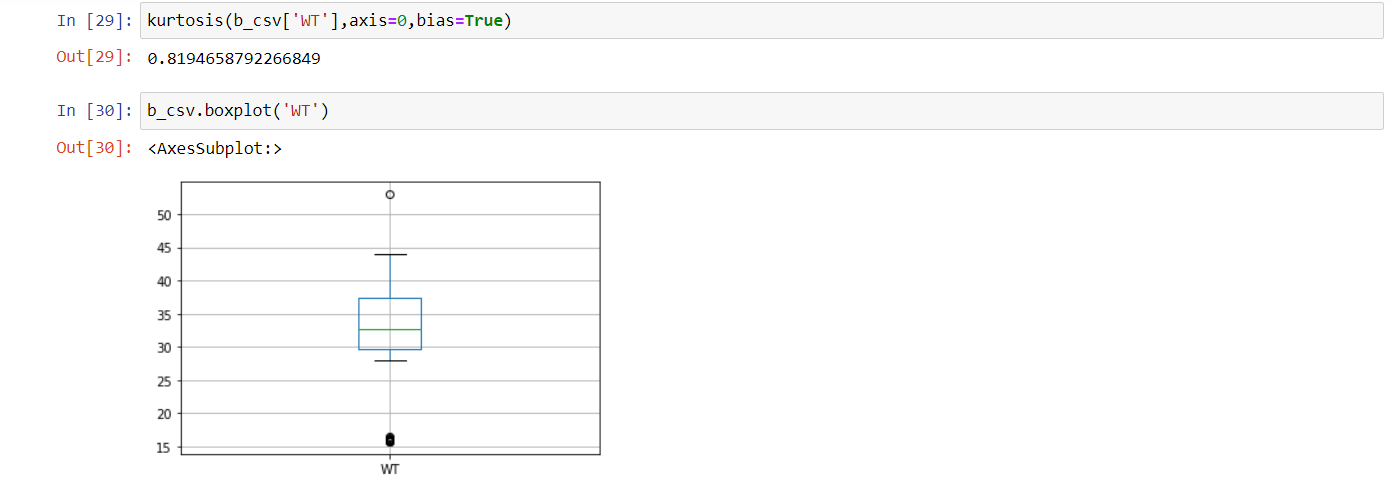
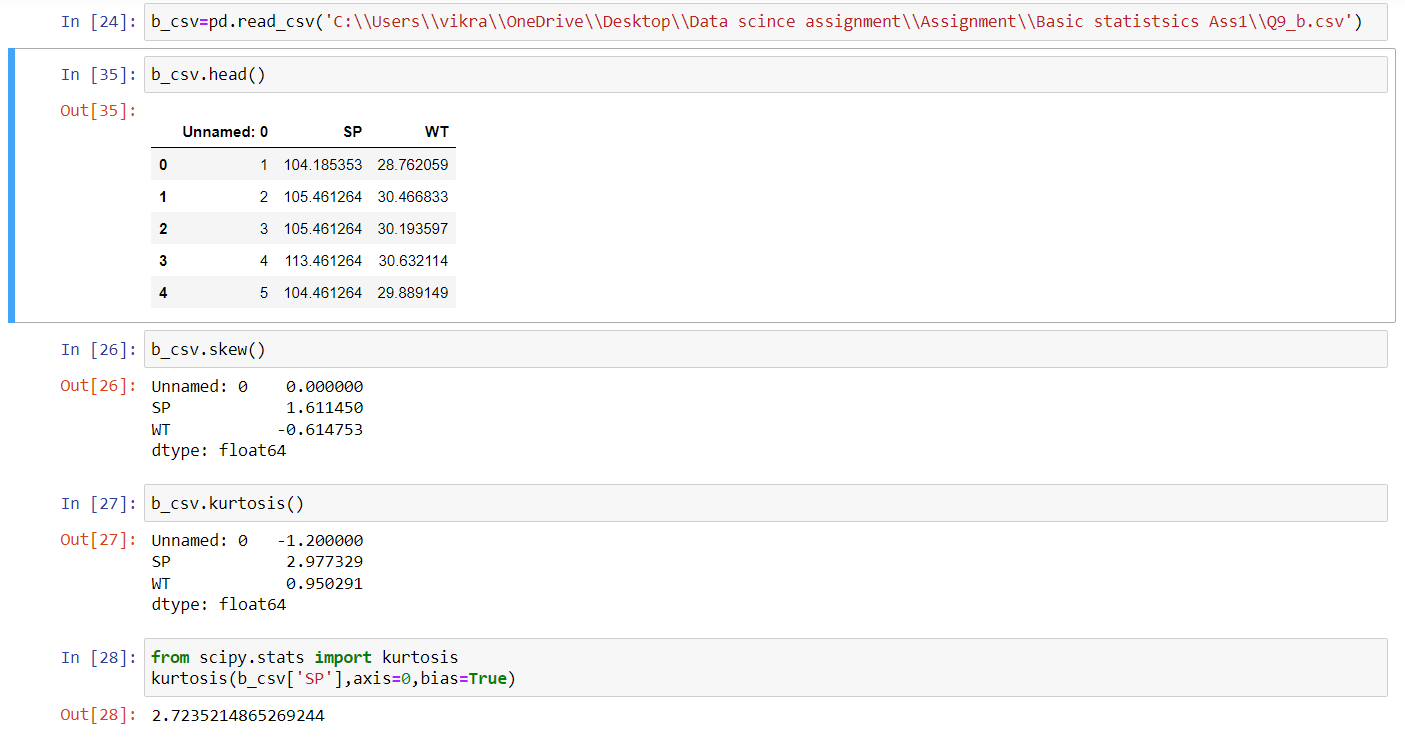
**Ans**



**Plz refer the Data science assignment 1 python file.**

**2. SP and Weight(WT)**

**Use Q9\_b.csv**



**Ans**

**Plz refer the Data science assignment 1 python file.**

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



**Ans**

**The chickWeight data is positively skewed data (Right skewed data). 50% of chickWeight is between 50 to 100**

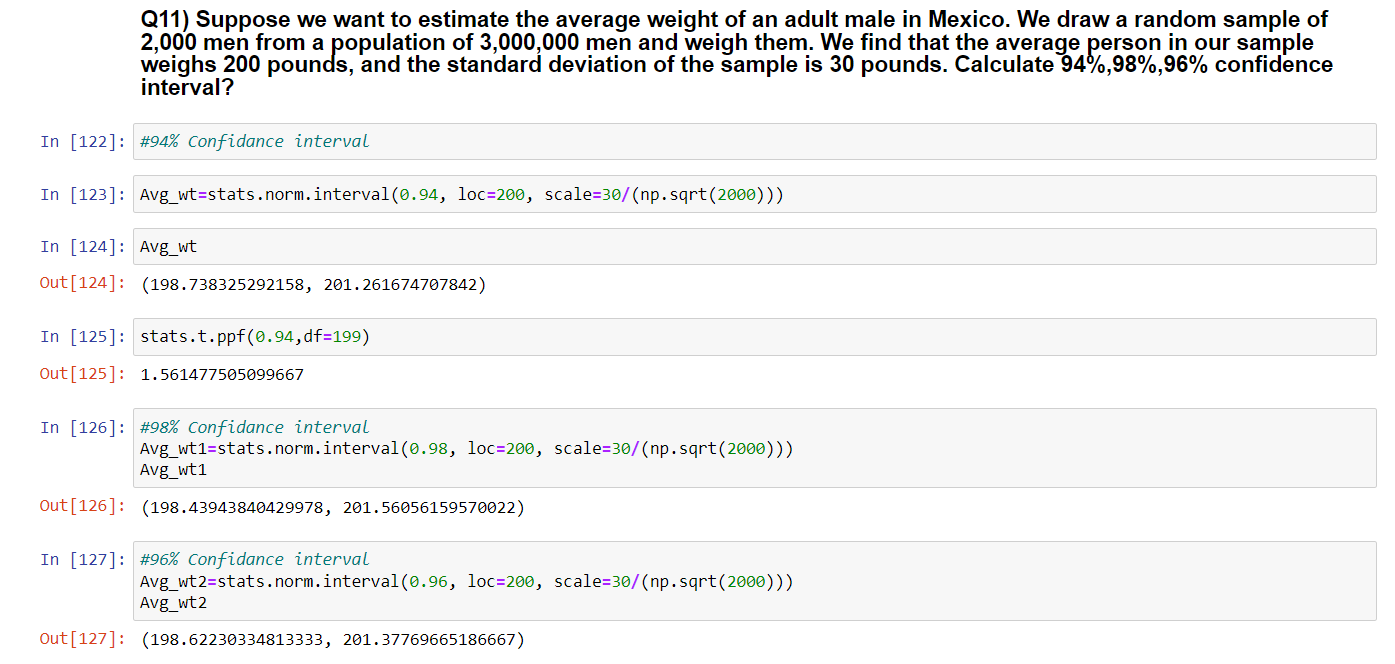


**Ans.**

**Data is positively skewed data (Right skewed data) and outliers at upper side of boxplot.**

**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**

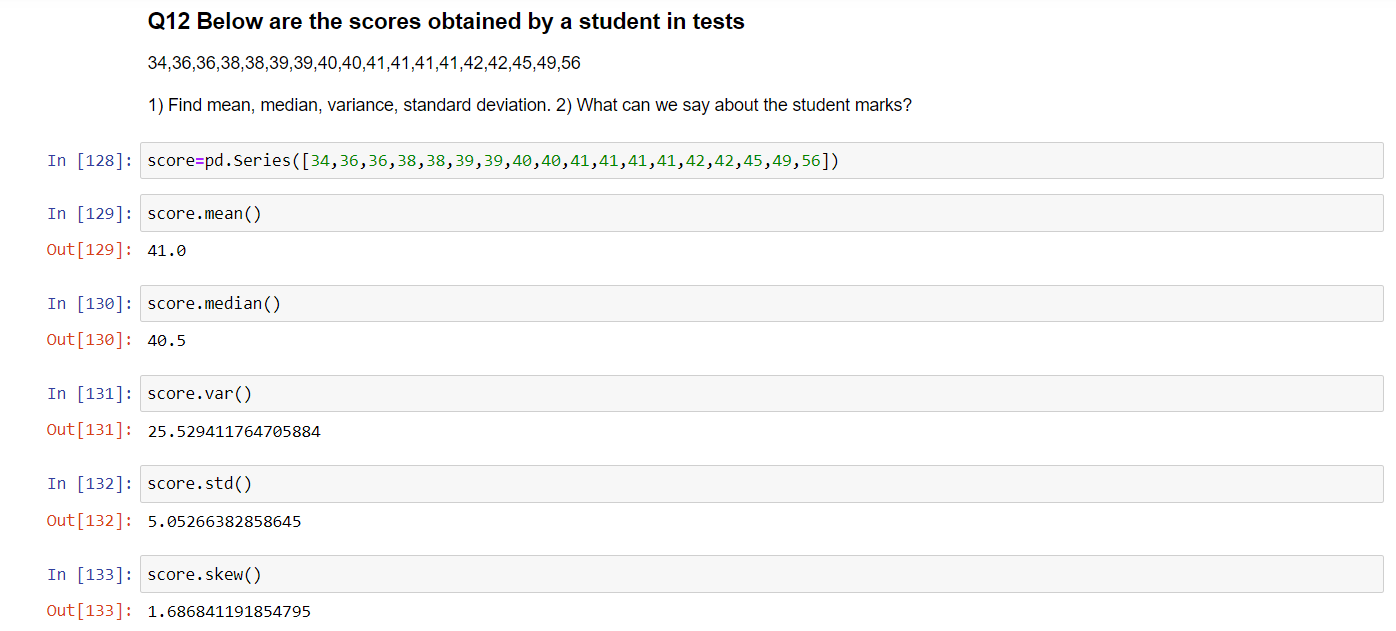


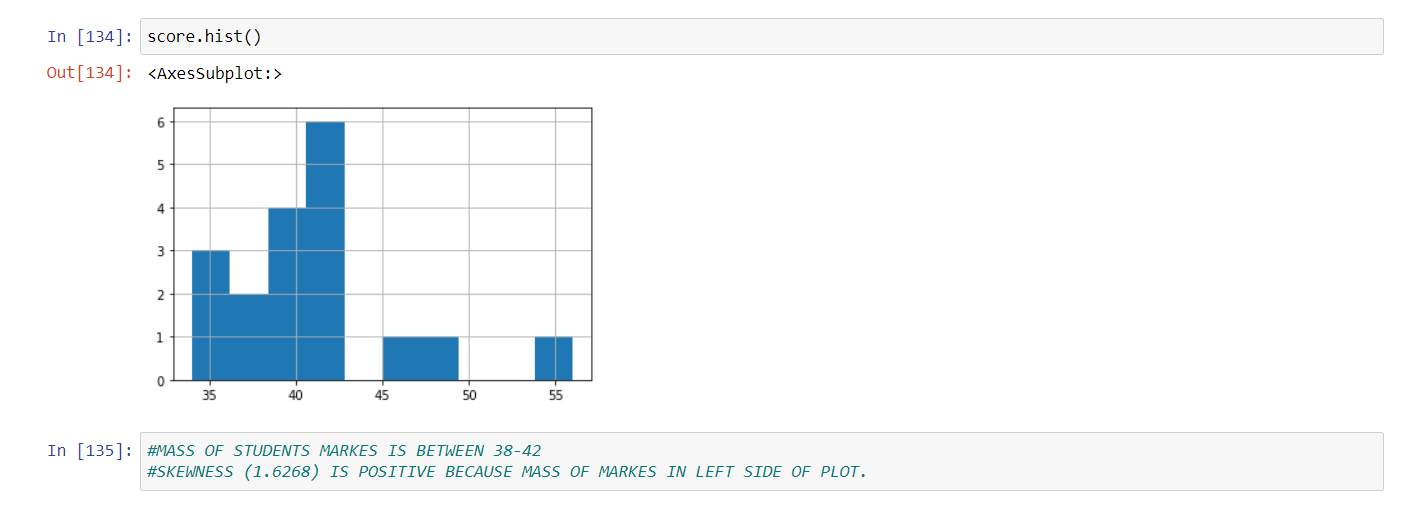
**Plz refer the Data science assignment 1 python file.**

**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?





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

**Ans.**

**There is no skewness and the data is normalized.**

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

**Ans.**

**Negative Skewness implies mass of the Distribution concentrated on right side**

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

**Ans.**

**Positive Skewness implies mass of the Distribution concentrated on left side.**

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

**Ans.**

**Positive Kurtosis indicates that thinner peak and wider tails.**

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

**Ans.**

**Negative kurtosis indicates that wider peak and thinner tails.**

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



What can we say about the distribution of the data?

**Ans.**

**Data is not normally distributed it will shows left skewed data.**

What is nature of skewness of the data?

**Ans.**

**Negative skewness (left skewed)**

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

**Ans.**

**IQR 10-18**

Q19) Comment on the below Boxplot visualizations?



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

**Ans.**

**The data of Boxplot 1 is 50% of the boxplot 1. Both the boxplot does not have any outliers. IQR of boxplot 1 is 250-280(Approximately) and IQR of boxplot 2 is 225-315 (Approximately). Min of boxplot 1 is 235 and Max is 290(approx.). Min of boxplot 2 is 180 and Max is 350(approx.)**

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)
  3. P (20<MPG<50)

**Ans**



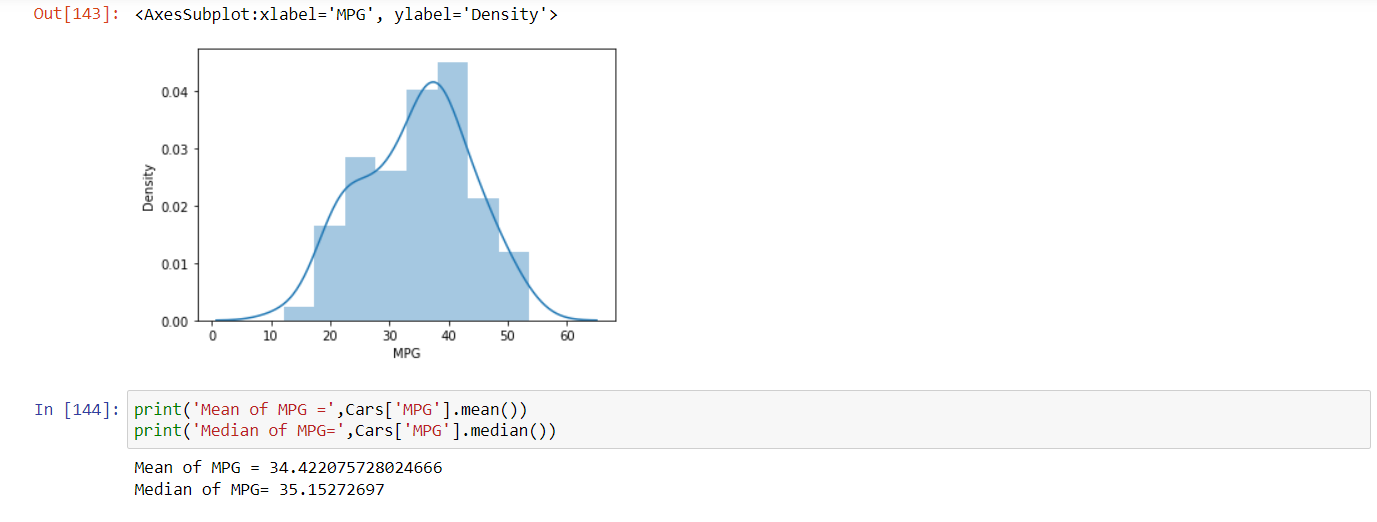
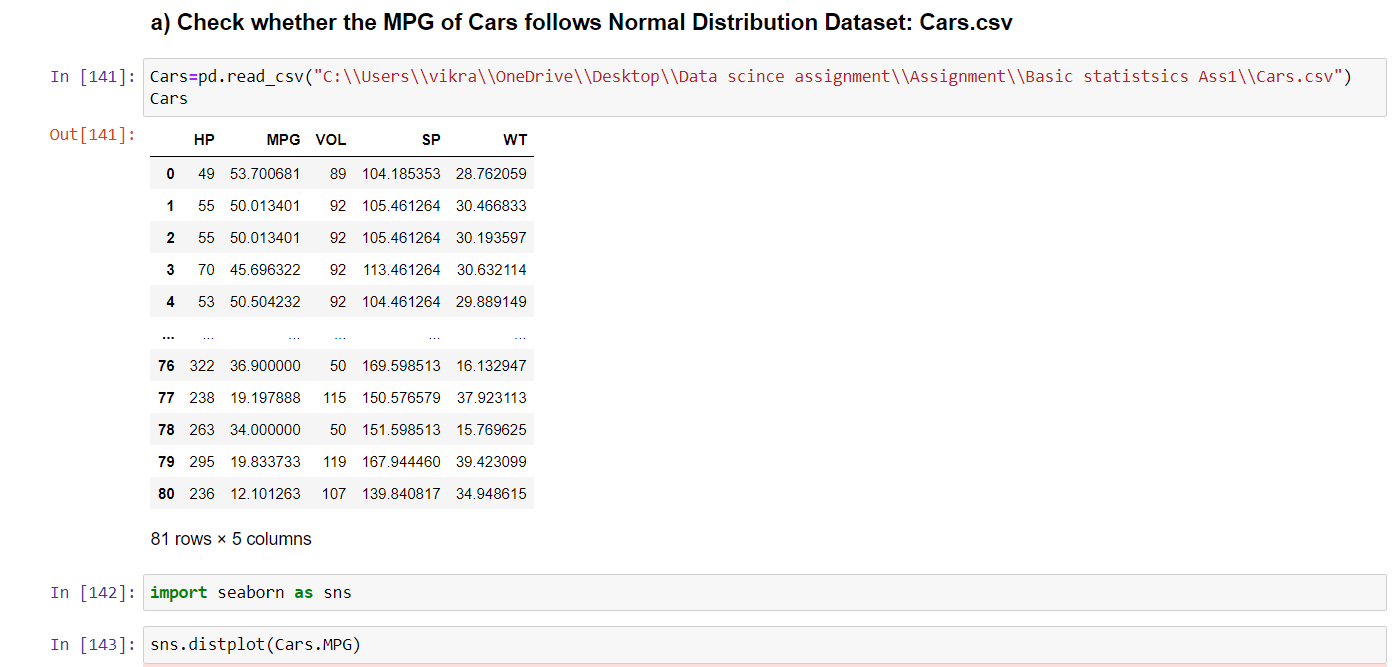
**Plz refer the Data science assignment 1 python file.**

ss

Q 21) Check whether the data follows normal distribution

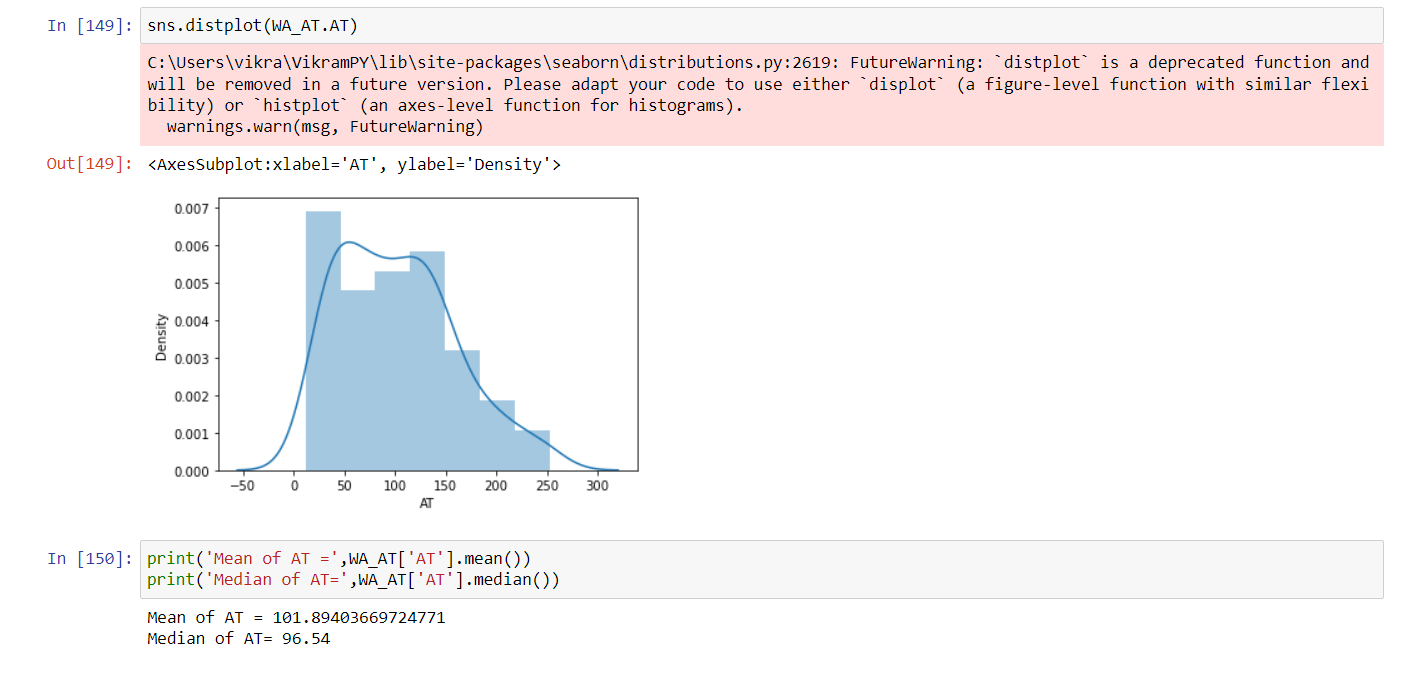
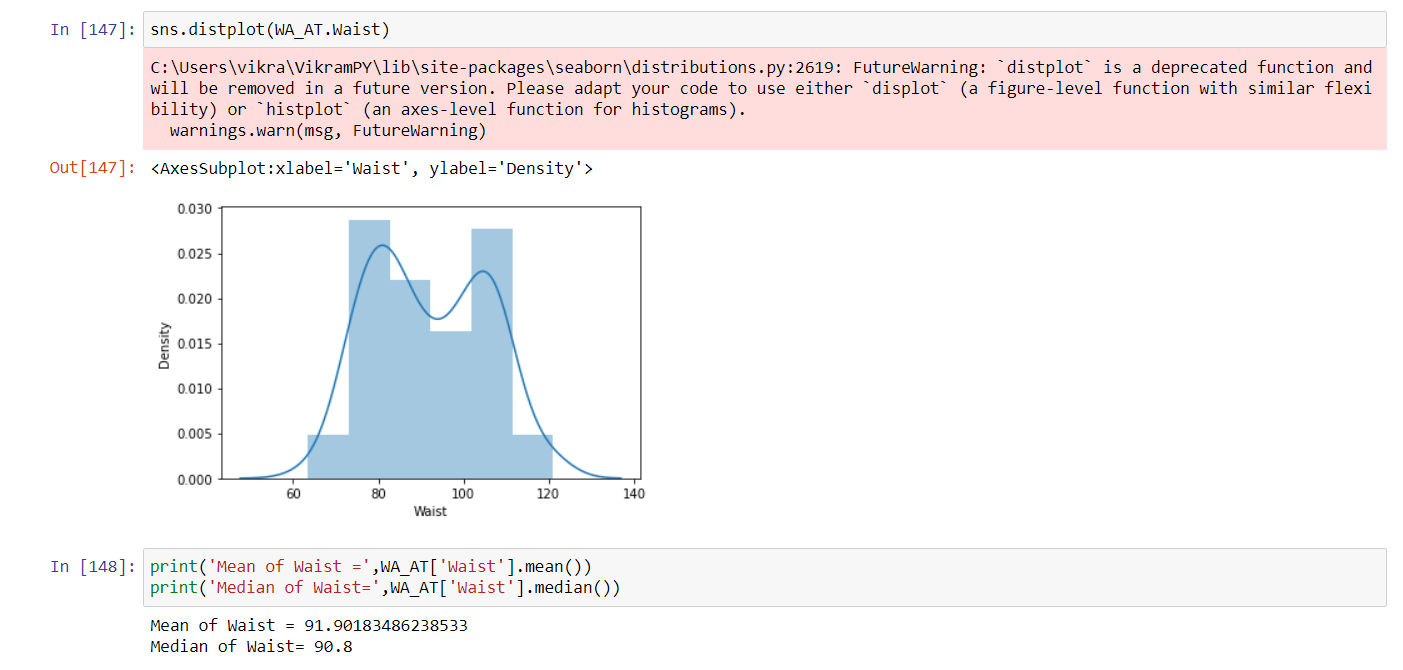
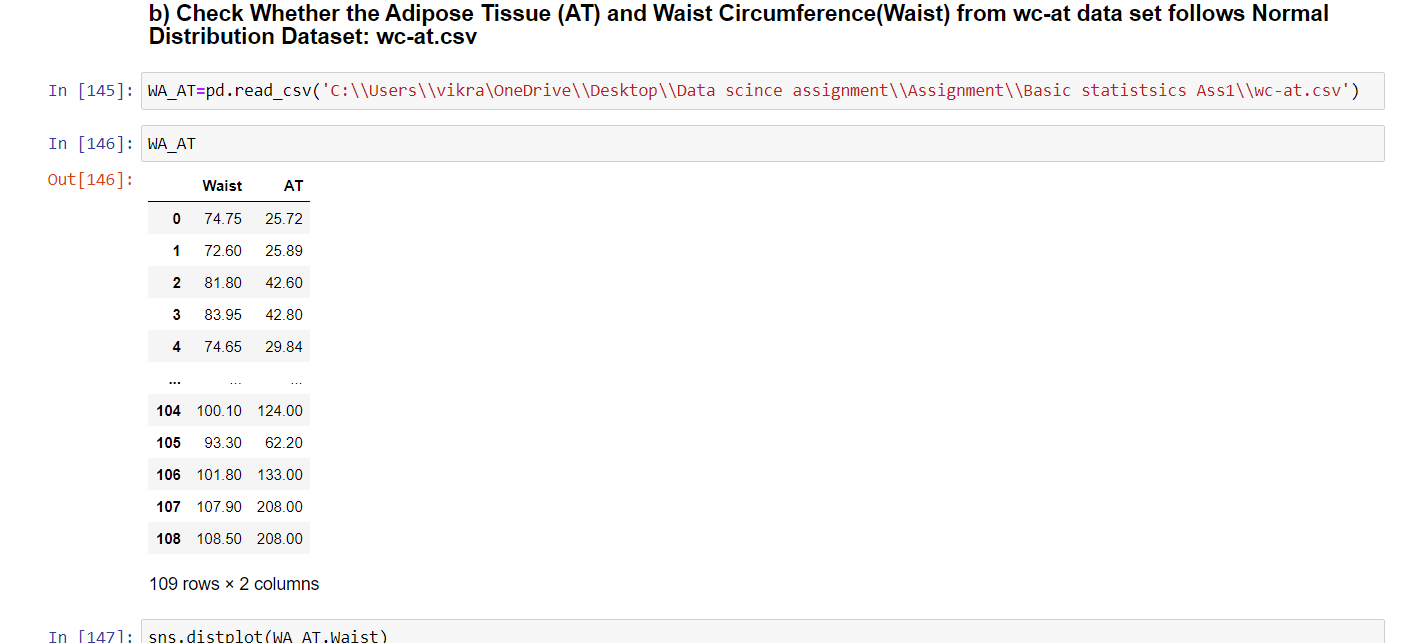
1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv



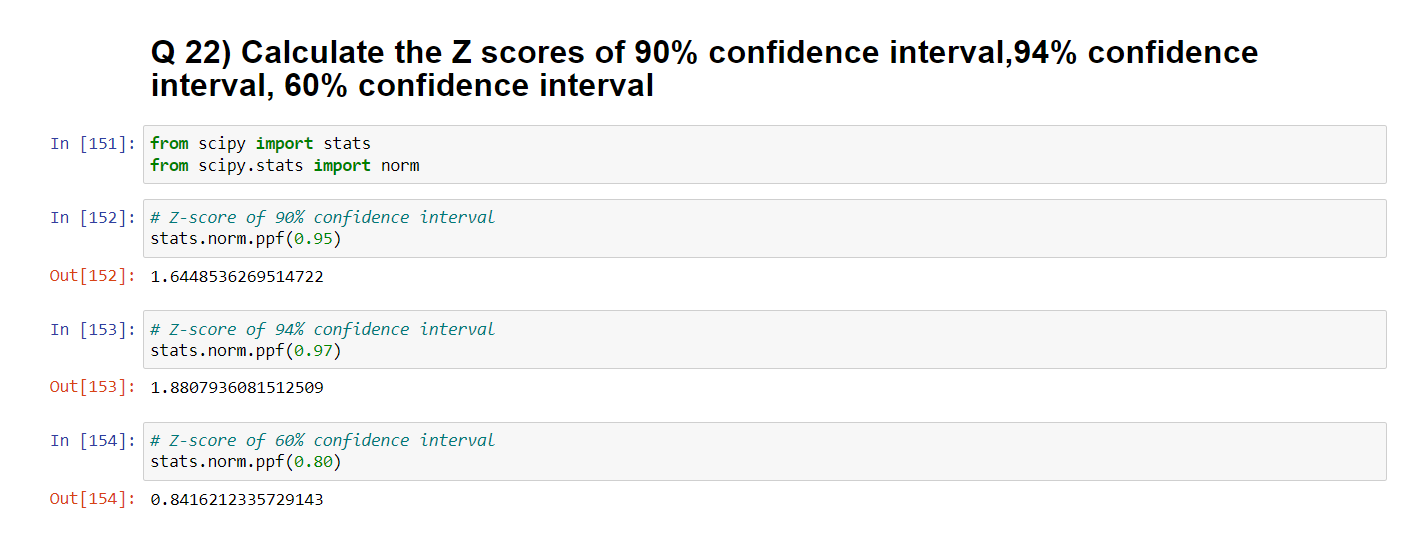
1. Check Whether the Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set follows Normal Distribution

Dataset: wc-at.csv



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

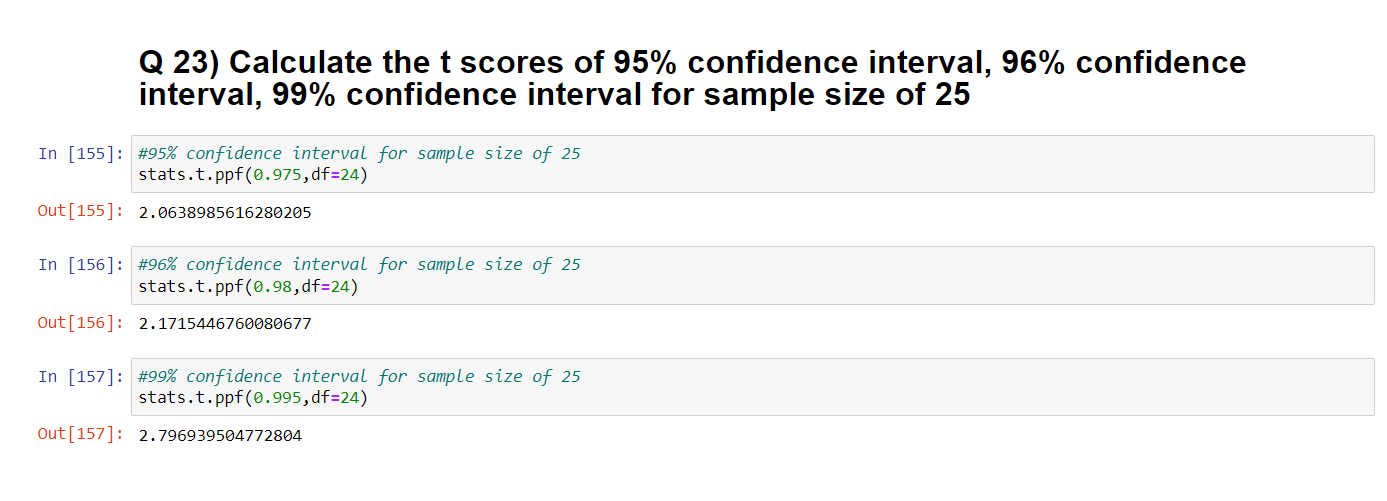
**Ans**



**Plz refer the Data science assignment 1 python file.**

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

**Ans**



**Plz refer the Data science assignment 1 python file.**

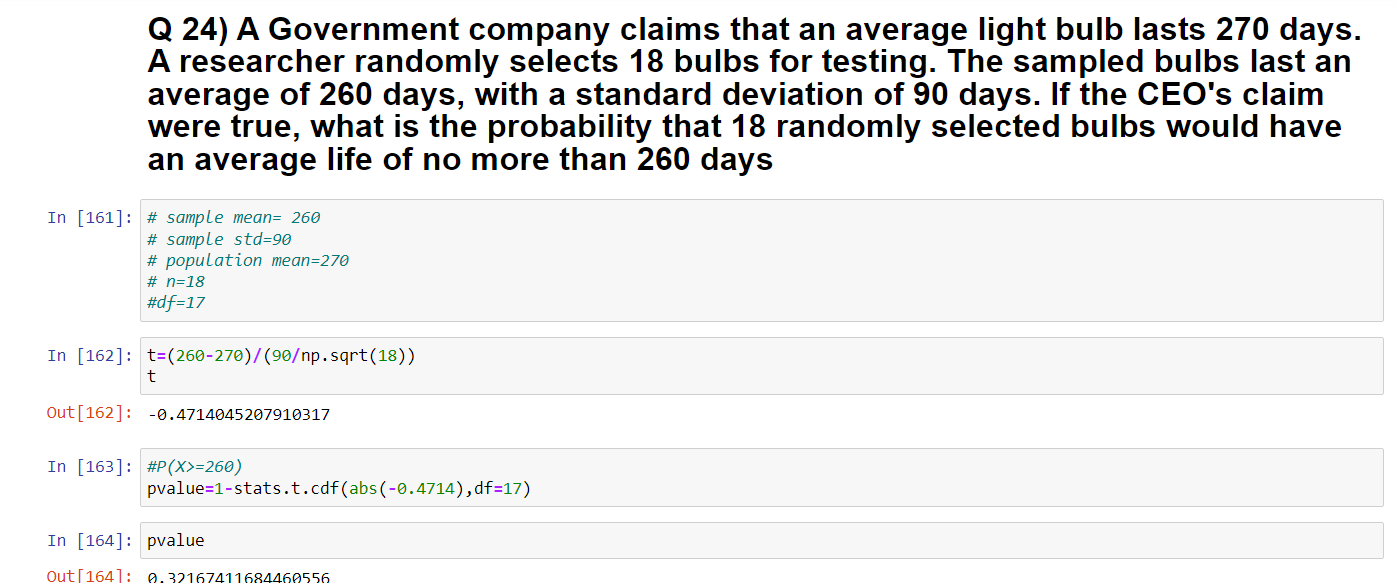
Q 24**)** A Government company claims that an average light bulb lasts 270 days. A researcher randomly selects 18 bulbs for testing. The sampled bulbs last an average of 260 days, with a standard deviation of 90 days. If the CEO's claim were true, what is the probability that 18 randomly selected bulbs would have an average life of no more than 260 days

Hint:

rcode 🡪 pt(tscore,df)

df 🡪 degrees of freedom

**Ans**



**Plz refer the Data science assignment 1 python file.**