Q1) Identify the Data type for the Following:

Activity	Data Type
Number of beatings from Wife	Discrete data
Results of rolling a dice	Discrete data
Weight of a person	Discrete data
Weight of Gold	Continuous data
Distance between two places	Continuous data
Length of a leaf	Continuous data
Dog's weight	Continuous data
Blue Color	Discrete data
Number of kids	Discrete data
Number of tickets in Indian railways	Discrete data
Number of times married	Discrete data
Gender (Male or Female)	Discrete data

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

Nominal, Ordinal, Interval, Ratio.

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

Barometer Pressure	CONTINUOUS data------Ratio
SAT Scores	CONTINUOUS data-----Interval
Years of Education	CONTINUOUS data------Interval

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

sol: let s be the sample space then
s={H,T},{H,T},{T,H}--------->Event Eis{HHT,HTH,THH}

Let x be the random variable having the two heads and one tails p(x=2)
=P(HHT)+P(HTH)+P(THH)=3/8

Q4) Two Dice are rolled, find the probability that sum is

a) Equal to 1 sol:1=0
b) Less than or equal to 4 sol:N(E)=6/36=1/6
c) Sum is divisible by 2 and 3

solC:6/36=1/6=0.16

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?

sol: No of balls=(2+3+2)=7-------->n(s)=no of ways of drawing 2balls out of 7=7c2=21
Let E=event of drawn 2balls non of which is blue
n(E)=no of ways of drawing 2balls out of (2+3)balls 5c2=10
so probability that none of the ball drawn is blue n(E)/n(s)=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

sol: Child+(candies count(probability))
=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

=3.090
=3.09

the 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

Q8) Calculate Expected Value for the problem below

a) 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?
sol:Expected value =sum(x) probabilityof x=(1/9)(108)+(1/9)(110)+(1/9)(1/23)+(1/9)(1/134)+(1/9)(1/145)+(1/9)(1/67)+(1/9)(1/187)+(1/9)(1/199)=145.33

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

Cars speed and distance

Use Q9_a.csv car speed skweness value=-0.12 and kurtosis value=0.81
cars distance skweness value=0.81 and kurtosis value=0.41

SP and Weight(WT)

sp skweness=1.61 and kurtosis=0.95
Use Q9_b.csv WTskweness=1.61andkurtosis=0.95

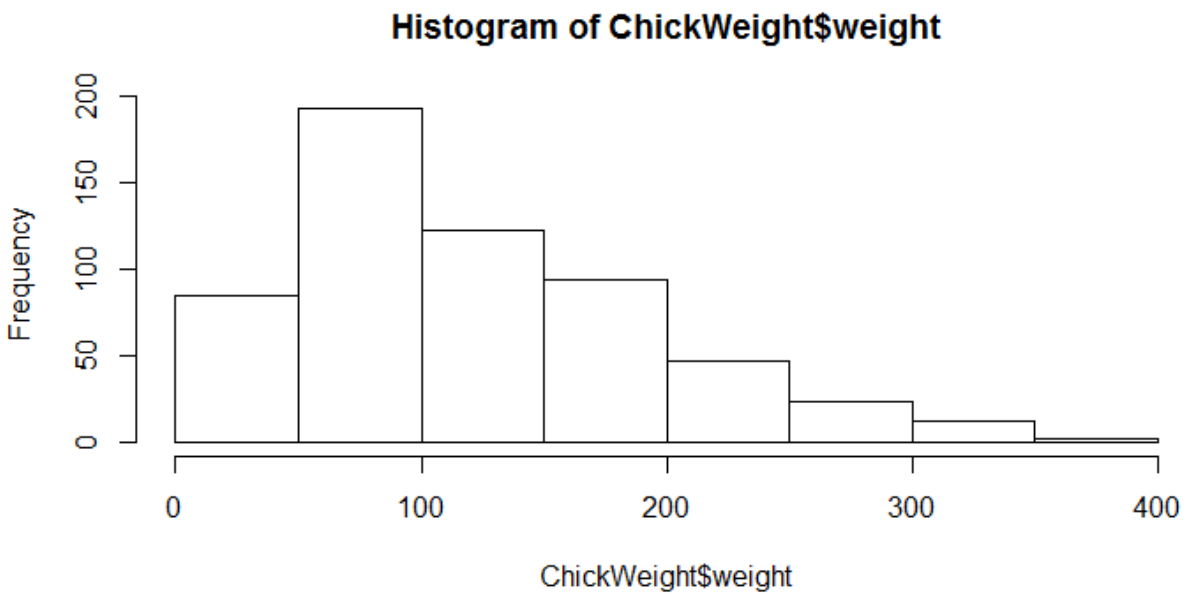
Q10) Draw inferences about the following boxplot & histogram

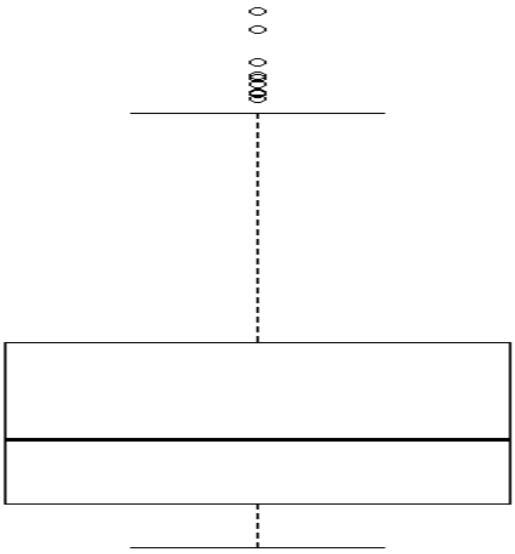
The histogram shows that :

The histogram of the chickweight is positively skewed

And the more number of the chickweigt is lies from 50to 150

And the more number chick weight is between 50 to 100





* In these boxplot the outliers falls at upperside

*The data of boxplot is rightly or 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?

sol: confidence interval=94=stats.interval alpha=0.94

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
2) What can we say about the student marks? Ans:there is no outliers and the data is some what rightly skewed because mean is greater than median

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

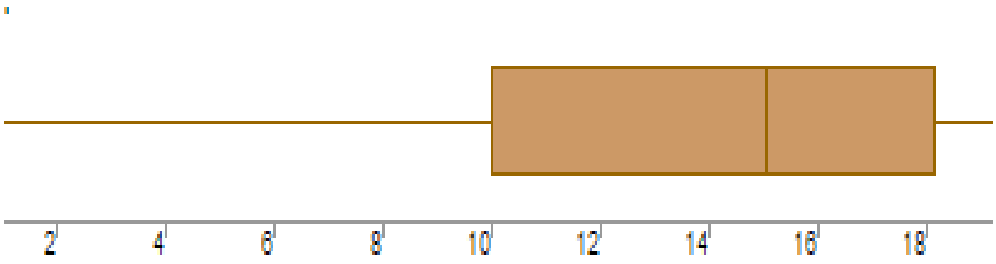
Q14) What is the nature of skewness when mean > median ? Ans:Rightly skewed

Q15) What is the nature of skewness when median > mean? Ans: left skewed

Q16) What does positive kurtosis value indicates for a data ? Ans:Peakedness of data

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

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?

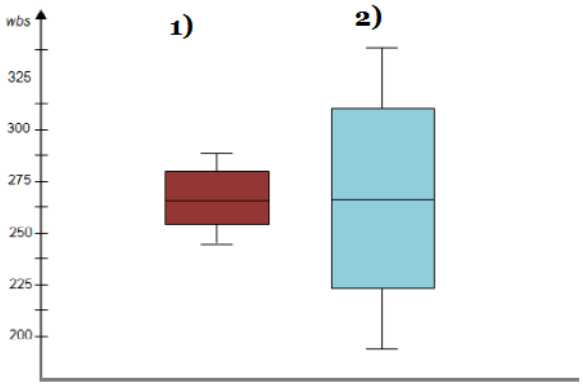
What will be the IQR of the data (approximately)?

1: The boxplot is not equally distrubuted and the median is towards right side

2:The data is skewed towards left and the whisker range of minimum value is greater than minimum

3:The iqr(INTER QUARTILE RANGE)=Q3
upper quartile is Q1 Lower quartile=18-10=8

Q19) Comment on the below Boxplot visualizations?



Ans:There is no outliers and the two boxplot has the same median that is in the range 275to250 and there are normally distributed with the zero to no skweness nor at the minimum or maximum whisker range

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

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

Data _set: Cars.csv

a)Ans: prob_MPG_greater_than_38=np.round (1-stats.norm.cdf(38,loc=q20.MPG.mean (),scale=q20.MPG.std(),3) print:(p(MPG>38)=,prob_MPG_greater_than_38) p(MPG>38)=0.348

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

MPG <- Cars$MPG

a. P(MPG>38)
b. P(MPG<40)
c. P (20<MPG<50)

b)Ans: p(MPG<40) prob_MPG_less_40=np.round (stats.norm.cdf(40,loc=q20.MPG.mean (),scale=q20.MPGstd()),3) print:(p(MPG<40)=prob_MPG_less_than_40)

c)Ans:prob_MPG_greater_than_20=np.round (1-stats.norm.cdf(20,loc=q20.MPG.mean (),scale=q20.MPG.std(),3) print:('p(MPG>20)0.943

prob_MPG_less_than_50=np.round(stats.norm.cdf(50,loc=q20.MPG.mean (),scale=q20MPG.std(),3) print:('p(MPG<50)=(prob_MPG_less_than_50))p (MPG<50)=0.956

pro_MPG_greater_than20_and_less_than50=(prob_MPG_less_than._50)-(prob_MPG_greater_than_20) print(P (MPG<20<50)=,(prob_MPG_greaterthan_20and_lessthan50))p(20<MPG<50)=0.0130000012

Q 21) Check whether the data follows normal distribution

a) Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

a)ANS: MPG of cars follows normal distribution

b)ans:adipose tissue(AT) and waist does not follows normal distribution

b) 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

22ans:
```
Z value for 90% confidence interval
print(zscore for60%confidence interval=np.rounds(stats.norm.ppf(0.5)
4))z score for 60% confidence intervel=1.6449

*Zvalue for94%confidence interval
print(z score for 60% confidence interval=np.round(stats.norm.ppf
(0.3),4))z score for 60% confidence interval=108808
*z value for 60%confidence interval=np.round(stats.norm.ppf(2),4))z
score for 60% confidence interval=0.8416
```

23Ans:
```
t score for 95% confidence interval
print(tscore for 95%confidence interval=np.round(stats.t.ppf
(0.025,df=24),4))t score for 95% confidence interval=-2.0639
t value for 94% confidence interval
print(tscore for 94% confidence interavl=np.rounds(stats.t.ppf(003,df=24),4))
t score for94%confidence interval=1.974
t value for 99% confidence interval
print(tscore for95%confidence interval=np.rounds(stats.t.ppf
(0.005,df=24),4)) t score for 95% confidence interval=-2.7969
```

24Ans:
```
import numpy as np import scipy as stats t_score=(x-popmean)/
(sample satndard deviation/square root of sample size)(260-270)/90/np.sqrt
(18))
t_score=0.471 stats.t.cdf(t_score,df=17)0.32=32%
```