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	Continous
Weight of Gold	Continous
Distance between two places	Continous
Length of a leaf	Continous
Dog's weight	Continous
Blue Color	Discrete
Number of kids	Discrete
Number of tickets in Indian railways	Discrete
Number of times married	Discrete
Gender (Male or Female)	Discrete

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	nominal
Fahrenheit Temperature	interval
Height	ratio
Type of living accommodation	nominal
Level of Agreement	Interval
IQ(Intelligence Scale)	interval
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	Ratio
SAT Scores	Interval
Years of Education	nominal

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

P = {(HHH), (HHT), (HTT), (TTH), (THT), (HTT), (TTT), (HTH)}

NO OF HEADS=2 & TAILS =1 HAS PROBABILITY 3/8

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

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

Ans

X={(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),(5,4),(5,6)
(5,1),(5,2),(5,3),(5,4),(5,5),(5,6)
(6,1),(6,2),(6,3),(6,4),(6,5),(6,6)

And N(X)=36

- A=1{}=0
- B={(1,1),(1,2),(1,3),(2,1),(2,2),(3,1)}=6
  Probability=B/N(X)
  =6/36=1/6
- C={(1,1),(1,2),(1,3),(1,5),(2,1),(2,2)
  (2,4),(2,6),(3,1),(3,3),(3,8),(3,6)
  (4,2),(4,4),(4,5),(4,6),(5,1),(5,3),
  (5,4),(5,5),(6,2),(6,3),(6,4),(6,6)}=24
  N(C)=24

P(C) =N(C)/N(X)

=24/36=2/3

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 number of events= ===21

Interested events===10

Probability that none of the balls 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 = E(x)

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

	points	score	weight
Mean	3.536563	3.21725	17.84875
Median	3.695	3.325	17.71
Mode	3.92	3.44	18.9
Variance	0.285881	0.957379	3.193166
standard deviation	0.534679	0.978457	1.786943
Range	2.17	3.911	8.4

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?

Expected Value = P*Value

Here Probability = 1/9 (there are total 9 values)

= 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 = 145.333

Points Score Weigh Mean- 3.596563 3.21725 17.84875

Median- 3.695 3.325 17.71 Mode- 3.92 3.44 17.02

Variance- 0.285881 0.957379 3.193166

std.Dev.- 0.534679 0.978457 1.786943 Range- 2.17 3.911 8.4

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

Cars speed and distance

Use Q9_a.csv

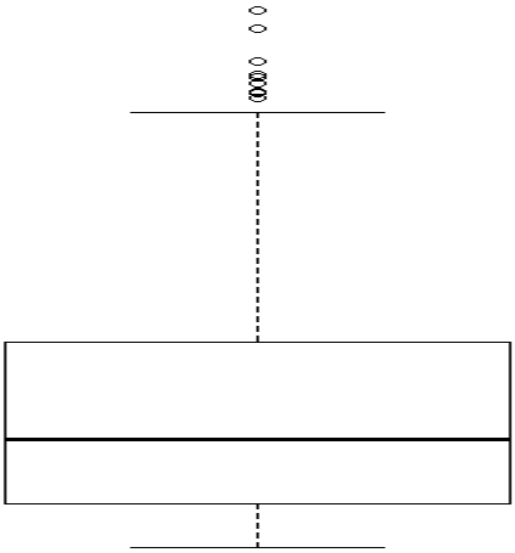
	speed	distance
skewness	-0.11055	0.782484
kurtosis	2.422853	3.248019

SP and Weight(WT)

Use Q9_b.csv

	SP	WT
Skewness	1.581455	-0.60331
kurtosis	5.723521	3.819766

Q10) Draw inferences about the following boxplot & histogram



Ans :Histogram indicates positive skewness and has some flatness representing Platykurtic. Boxplot whiskers are not even as well as its symmetry is not even. It has outliers on the upper extreme side.

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: AVG_WGT1 = stats.norm.interval(0.97, loc = 200, scale = 30)

print('Average weight of adult in Mexico at 94% confidence interval',

np.round(AVG_WGT1, 3))

Average weight of adult in Mexico at 94% confidence interval [134.897 265.103]

AVG_WGT2 = stats.norm.interval(0.99, loc = 200, scale = 30)

print('Average weight of adult in Mexico at 98% confidence interval',

np.round(AVG_WGT2, 3))

Average weight of adult in Mexico at 98% confidence interval [122.725 277.275]

AVG_WGT3 = stats.norm.interval(0.98, loc = 200, scale = 30)

print('Average weight of adult in Mexico at 96% confidence interval',

np.round(AVG_WGT3, 3))

Average weight of adult in Mexico at 96% confidence interval [130.21 269.79]

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?

Mean	41
Median	40.5
Variance	25.52941
Standard deviation	5.052664

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

Ans:When mean and median are equal that time it is symmetric about mean

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

Ans:Whenever mean is greater than mean it indicates positive skewness

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

Whenever mean is smaller than mean it indicates negative skewness

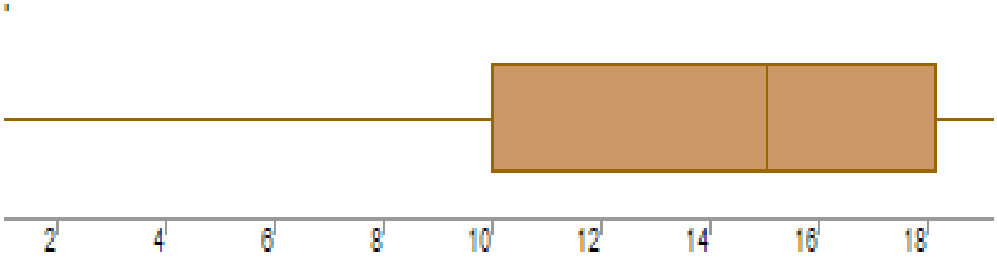
Q16) What does positive kurtosis value indicates for a data ?

Positive kurtosis refers to thin and high peak

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

Positive kurtosis refers to wide and short peak

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



What can we say about the distribution of the data?

Ans:Positive kurtosis refers to wide and short peak

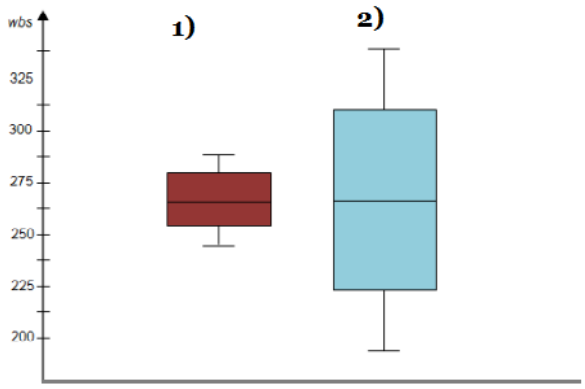
What is nature of skewness of the data?

Ans:Positive kurtosis refers to wide and short peak

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

Ans:Interquartile range of given data is 8

Q19) Comment on the below Boxplot visualizations?



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

1. The first boxplot is symmetric because Q3-Q2 = Q2-Q1 that is the whisker on both the sides of first boxplot have of same length.

2. In the second boxplot it is also a symmetric because Q3-Q2 = Q2-Q1 because the whisker on both the sides are equa.

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

a. P(MPG>38) ->34.75%
b. P(MPG<40)-> 72.93%
c. P (20<MPG<50) ->89.88%

Q 21) Check whether the data follows normal distribution

a) Check whether the MPG of Cars follows Normal Distribution
Dataset: Cars.csv

Ans:It Slightly follows normal distribution as mean and median shows little differenc

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

It doesnot follows normal distribution, when two curves are compared (AT&WC)

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

```
> qnorm(0.95)
[1] 1.644854
> qnorm(0.97)
[1] 1.880794 >
 qnorm(0.80)
 [1] 0.8416212
```

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

```
> qt(0.975, df = 24)
 [1] 2.063899
> qt(0.98, df = 24)
 [1] 2.171545
 > qt(0.995, df = 24)
 [1] 2.79694
```

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

μ = 270

n = 18

x = 260

S = 90

t formula = $x-\mu\ s\ \sqrt{n}$

= 260 − 270 90 √18

t value = -0.471404521

df = n-1 =

18-1

= 17

pt(-0.471404521,df=17)

[1] 0.3216725

From this we can conclude that outcome probability is 32.16%