Activity	Data Type
Number of beatings from Wife	Ordinal
Results of rolling a dice	Discrete data
Weight of a person	Continuous 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	Nominal
Number of kids	Discrete data
Number of tickets in Indian railways	Discrete data
Number of times married	Discrete data
Gender (Male or Female)	Nominal data

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	Ordinal
Level of Agreement	Ordinal
IQ(Intelligence Scale)	Ratio
Sales Figures	Ratio
Blood Group	Nominal
Time Of Day	Ordinal
Time on a Clock with Hands	Interval
Number of Children	Nominal
Religious Preference	Nominal
Barometer Pressure	Interval
SAT Scores	Interval
Years of Education	Ordinal

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

Ans: P (Two heads and one tail)

= N (Event (Two heads and one tail)) / N (Event (Three coins tossed))

= 3/8

= 0.375

= 37.5%

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

a) Equal to 1

b) Less than or equal to 4

c) Sum is divisible by 2 and 3

a) Equal to 1

Ans: Number of possible outcomes for the above event is N (Event (Two dice rolled))

= 6^2

= 36

P (sum is Equal to 1) = '0'

b) Less than or equal to 4

Ans: P (Sum is less than or equal to 4)

= N (Event (Sum is less than or equal to 4)) / N (Event (Two dice rolled))

= 6 / 36

= 1/6

= 0.166

= 16.66%

c) Sum is divisible by 2 and 3

Ans: P (Sum is divisible by 2 and 3)

= N (Event (Sum is divisible by 2 and 3)) / N(Event (Two dice rolled))

= 6 / 36

= 1/6

= 0.16

= 16.66%

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 balls =7

balls N (Event (2 balls are drawn randomly from bag) = 7! / 2! * 5!

= (7654321) / (21) * (5432*1)

:- N (Event (2 balls are drawn randomly from bag)

= (76)/ (21) = 21

:- If none of them drawn 2 balls are blue

= 7 – 2 = 5

:- N (Event (None of the balls drawn is blue)

= 5! / 2! * 3!

= (54) / (21)

= 10

:- P (None of the balls drawn is blue)

= N (Event (None of the balls drawn is blue) / N (Event (2 balls are drawn randomly from bag)

= 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:-

CHILD	Candies count	Probability	Expected values
A	1	0.015	0.015
B	4	0.20	0.8
C	3	0.65	1.95
D	5	0.005	0.025
E	6	0.01	0.06
F	2	0.120	0.24

Expected number of candies for a randomly selected child

= 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

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.

Ans:- Mean for Points = 3.59, Score = 3.21 and Weigh = 17.84

Median for Points = 3.69, Score = 3.32 and Weigh = 17.71

Mode for Points = 3.07, Score = 3.44 and Weigh = 17.02

Variance for Points = 0.28, Score = 0.95, Weigh = 3.19

Standard Deviation for Points = 0.53, Score = 0.97, Weigh = 1.78

Range [Min-Max] for Points [3.59 – 4.93], Score [3.21 – 5.42] and Weigh [17.84 – 22.9]

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?

Ans:-

= (1/9)(108) + (1/9)(110) + (1/9)(123) + (1/9)(134) + (1/9)(145) + (1/9)(167) + (1/9)(187) + (1/9)(199)

= 145.33

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

Cars speed and distance

Use Q9_a.csv

Ans :

```
     from scipy.stats import skew
     from scipy.stats import kurtosis
     import pandas as pd
     import numpy as np
 Q_9 = pd.read_csv("/content/sample_data/Q9_a.csv")
       print(skew(Q_9 ,axis = 0, bias=True))
       print(kurtosis(Q_9 ,axis = 0, bias=True))
```

For Cars Speed Skewness value = **-0.11395477** and Kurtosis value = **0.57714742**

Skewness value = **0.78248352** and Kurtosis value = **0.24801866** for Cars Distance

SP and Weight(WT)

Use Q9_b.csv

Ans:

```
      from scipy.stats import kurtosis
      import pandas as pd
      import numpy as np

Q9_B = pd.read_csv("/content/sample_data/Q9_b (1).csv")
      print(skew(Q9_B ,axis=0, bias=True))
      print(kurtosis(Q9_B ,axis=0, bias=True))
```

For SP Skewness = **1.58145** kurtosis = **2.7235**

For WT Skewness = **-0.6033** Kurtosis = **0.819465**

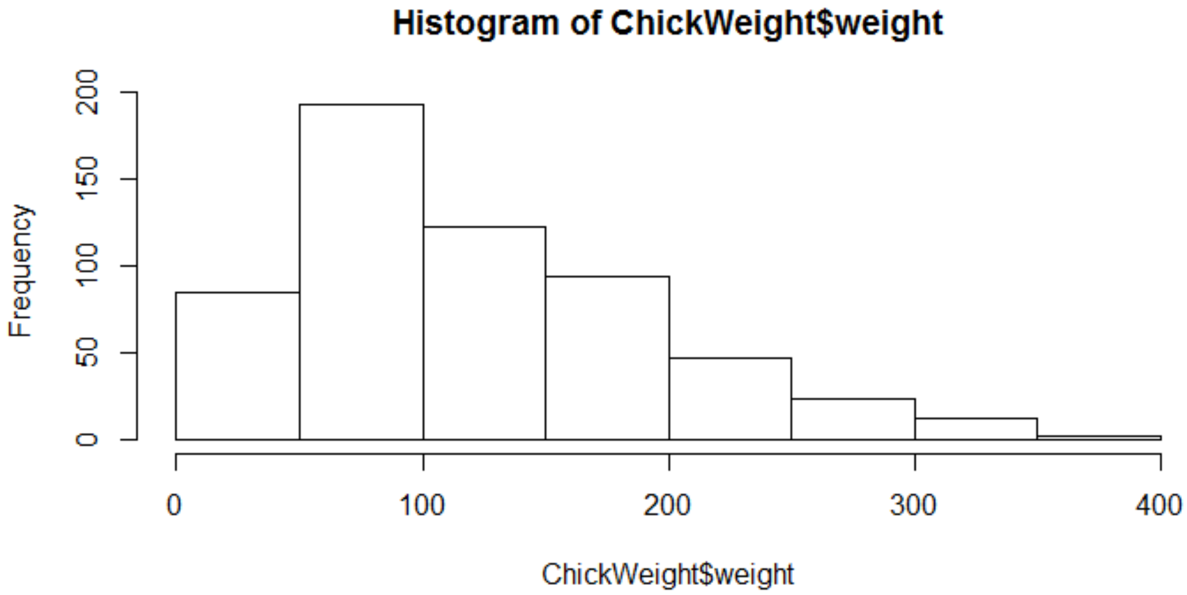
Q10) Draw inferences about the following boxplot & histogram

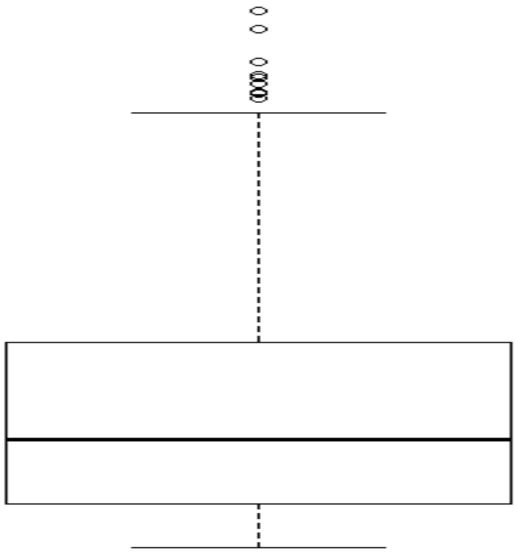
Ans: The histograms peak has right skew and tail is on right.

Mean > Median. We have outliers on the higher side.

Ans:- On the basis of boxplot and histogram we concluded that Its right skewed data set so that most number of data points on lower side and median will be on the left of the mean of the data set and the histogram has log tail so that there some of the data point in higher side has outliers and this outliers we can see in the box plot as well.

The boxplot has outliers on the maximum side.


Histogram of ChickWeight$weight
Frequency
0
50
100
150
200
0
100
200
300
400
ChickWeight$weight



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:
```
from scipy import stats
import numpy as np
sample_mean = 200
sample_std = 30
sample_size = 2000
df = sample_size - 1

t_94 = stats.t.ppf (0.97, df)
t_98 = stats.t.ppf (0.99, df)
t_96 = stats.t.ppf (0.99, df)

ci_94 = (sample_mean - t_94 * sample_std / np.sqrt(sample_size),
ci_98 = (sample_mean - t_98 * sample_std / np.sqrt(sample_size),
ci_96 = (sample_mean - t_96 * sample_std / np.sqrt(sample_size),

print("94% Confidence Interval:", ci_94)
print("98% Confidence Interval:", ci_98)
print("96% Confidence Interval:", ci_96)
```

For 94% confidence interval Range is **[198.73 – 201.26]**

For 98% confidence interval range is **[198.43 – 201.56]**

For 96% confidence interval range is **[198.62 – 201.37]**

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 = 25.52 and Standard Deviation = 5.05

2) What can we say about the student marks?

Ans:-

1] there is some deviation in the marks so that sum of the students scores high marks and some of the student score low marks in respective of the other student.

2] mean and median approx. similar so that the data is equally distributed both side of the mean points so that scores of the students in the exam equally distributed almost % student got low marks and 50% students got high marks.

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

Ans:- No skewness is present we have a perfect symmetrical distribution

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

Ans:- Skewness and tail is towards Right

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

Ans:- Skewness and tail is towards left

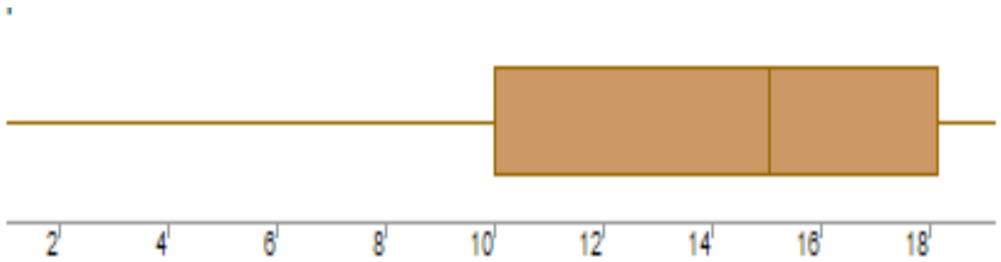
Q16) What does positive kurtosis value indicates for a data ?

Ans:- Positive kurtosis means the curve is more peaked and it is Leptokurtic

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

Ans:- Negative Kurtosis means the curve will be flatter and broader

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



What can we say about the distribution of the data?

Ans:- What can we say about the distribution of the data? Ans: The above Boxplot is not normally distributed the median is towards the higher value.

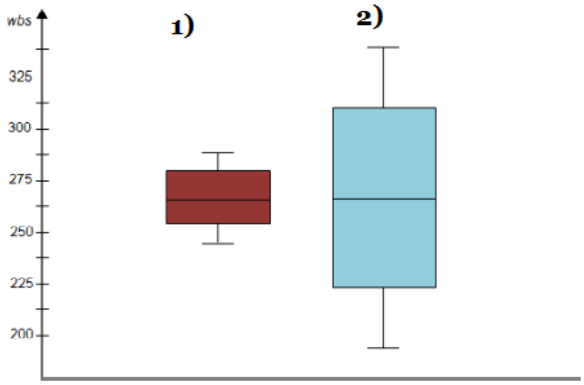
What is nature of skewness of the data?

Ans:- What is nature of skewness of the data? Ans: The data is a skewed towards left. The whisker range of minimum value is greater than maximum.

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

Ans:- What will be the IQR of the data (approximately)? Ans: The Inter Quantile Range = Q3 Upper quartile – Q1 Lower Quartile = 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:- First there are no outliers. Second both the box plot shares the same median that is approximately in a range between 275 to 250 and they are normally distributed with zero to no skewness neither at the minimum or maximum whisker range.

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)
b. P(MPG<40)
c. P (20<MPG<50)

A.P(MPG>38)

Ans:
```
( 30-mpg.mean())/mpg.std()
    probality = 1-stats.norm.cdf(-0.48426901305407655,mpg.mean(),mpg.std())
    probality*100
```

a. P(MPG>38) = **99.99%(0.99)**

B.P(MPG<40)

Ans:
```
(40-mpg.mean())/mpg.std()
    probality = stats.norm.cdf(0.6108479474833596,mpg.mean(),mpg.std())
    probality*100
```

b. P(MPG<40) = **0.010%(0.0010)**

C.P (20<MPG<50)

Ans:
```
print((20-mpg.mean())/mpg.std())
    (50-mpg.mean())/mpg.std()
    probality = stats.norm.cdf(1.7059649080207957,mpg.mean(),mpg.std())-
    stats.norm.cdf(-1.5793859735915128,mpg.mean(),mpg.std())
    probality*100
```

c. P (20<MPG<50) = **0.0129(0.00012)**

Q 21) Check whether the data follows normal distribution

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

Ans:- MPG of cars 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

Ans:- Adipose Tissue (AT) and Waist does not follow Normal Distribution

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

Ans:- **z value for 90% confidence interval**

```
print('Z score for 60% Conifidence Intervla
=',np.round(stats.norm.ppf(.05),4))
```

```
Z score for 60% Conifidence Intervla = 1.6449
```

z value for 94% confidence interval

```
print('Z score for 60% Conifidence Intervla
=',np.round(stats.norm.ppf(.03),4))
```

```
Z score for 60% Conifidence Intervla = 1.8808
```

z value for 60% confidence interval

```
print('Z score for 60% Conifidence Intervla
=',np.round(stats.norm.ppf(.2),4))
```

```
Z score for 60% Conifidence Intervla = 0.253
```

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% confidence interval

```
print('T score 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 96% confidence interval

```
print('T score for 94% Confidence Inteval =',np.round(stats.t.ppf(0.03,df=24),4))
T score for 94% Confidence Inteval = 2.171
```

t value for 99% Confidence Interval

```
print('T score for 95% Confidence Interval =',np.round(stats.t.ppf(0.005,df=24),4))
T score for 95% Confidence Interval = 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: rcode ◊ pt(tscore,df)
df ◊ degrees of freedom

Ans: import numpy as np

import scipy as stats

t_score = (x - pop mean) / (sample standard daviation / 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%**