Part-A

1) List the visual aids for EDA:-

Scatter plot
lineplot
Histogram
pie chart
Bar chart

These are various form of visual aids for EDA.

2) Benefits of Data Transformation:-

The Data Transformation is used to transfer any form of data's one format to another format.

It can also transfer the structures of an data into another structure

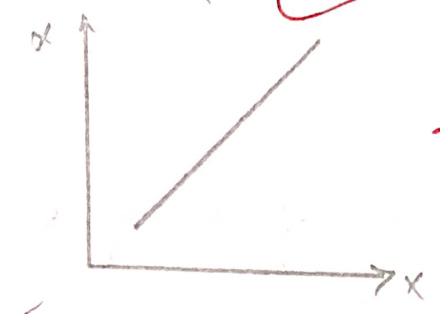
the Data Transformation is more useful to transfer the large storages of the big data.

3)

PLOT	Scatter
The plot is used to plot the lines in the graph.	the scatter used to plot the different form of scatter using markers in the graph

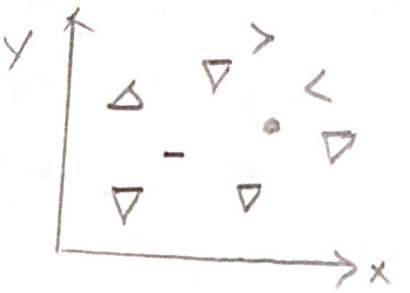
The plt.plot() is more essential for the large datas in the data graph to be plot.

Ex:-



The scatter() does not has much essential like plot. It will map the data as individuals

Ex:-

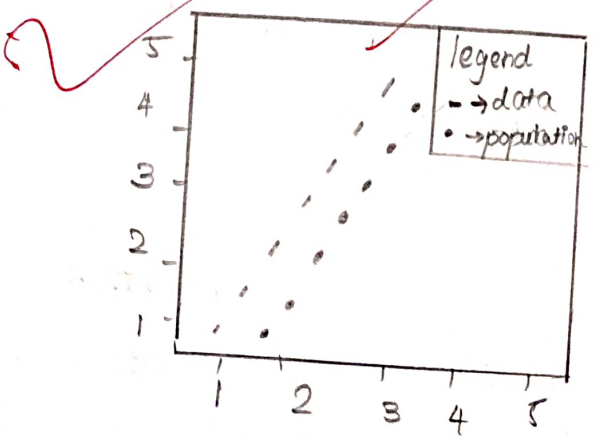


5)

4) purpose of legends:-

The legends are used to describe the data in the graph.

It can be used in the corner of graph which represents the data with the symbol.



6)

5) Standard Deviation:-

The standard deviation is defined as the statistical analysis which is a level of measures, measures the how far the data point from mean and desperation.

This measures are called as the standard deviation.

The standard deviation can be calculated by a formula,

$$s = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x}_i)^2}{n-1}}$$

Here, $s \rightarrow$ standard deviation

$\bar{x} \rightarrow$ Mean value

$n \rightarrow$ ~~Total value~~ No. of items in the data.

$x_i \rightarrow$ Given values in the data set

6) Evaluate the measure of central tendency:-

The various form of measure of central tendency are

Mean
Median
Mode
percentile
Quartile

propositions

Mean → $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{N}$

Median → Median = $\frac{x(n/2) + x(n/2+1)}{2}$

proposition → $\hat{p} = \frac{x}{n}$

7) Types of probability in contigency Table:

There are three types of probability in the contigency Table.

They are,

Joint probabity →
Marginal probability
combination probabity

Joint probability:-

$P(i\ row, j\ column) = \frac{\text{count in cell } (i,j)}{\text{Grand total}}$

Marginal:-

$P(i\ row) = \frac{\text{total count in row } (i)}{\text{Grand total}}$

$P(j\ column) = \frac{\text{total count in column}}{\text{Grand total,}}$

8) Outliers:-

Outliers are the data points that can diverge from the other observations.

The main task in the EDA to remove and determine the outliers.

Because, the outliers can cause more effects into the other data points of the data sets.

9) steps in Data cleaning:-

step-1 : Remove Duplicates and intermediate datas.

step-2 : Fix structural errors.

step-3 : Filter the unwanted outliers from the Data.

step-4 : Complete Deduplication process

step-5 : Estimate QA and Data value

10) Examples of Time series Analysis:-

Heart beat per minute
Monthly subscribers
stock price
Annual stock prices
Electrical source in brain
Rainwater measurements.

These are the few example for time series.