PLOTTING WITH PANDAS:

    SYNTAX : DataFrame.plot(kind = ' ')

* Pandas has predefined function called as
Plot.

The kinds of plot produced are:

    ↳ line : Line plot

    ↳ Bar : vertical bar

    ↳ barh : Horizantal bar

    ↳ hist : Histogram

    ↳ box : Boxplot

    ↳ kde/pdf : Kernel density estimation plot

    ↳ density : same as kde

    ↳ area : area plot

    ↳ pie : Piechart

    ↳ scatter : scatter plot

    ↳ hexbin : Hexbin plot

→ hist or kde - can produce only 1 column or 1 variable plot

→ scatter or hexbin - can produce 2 columns or 2 variables plot

→ box : Depending upon the situation, both 1 and 2 varibles or columns (numeric a characters) can be plotted.

SYNTAX: df.slcm.plot (kind = 'hist') or

df ['slcm'].plot (kind = 'hist')

similarly with the rest kind of plots.

VISUALIZATION IN MATPLOTLIB :

The matplotlib is a python 2D plotting library for data visualization and the creation of interactive graphics/plot. A plot is a graphical representation of data which shows the relationship between two variables or the distribution of data.

↳ subplots are used to display multiple plots in the same window.
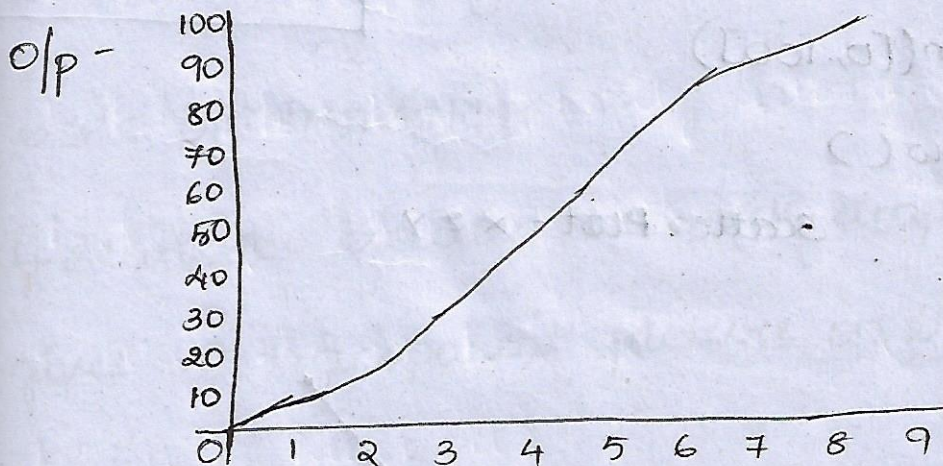
BASIC PLOTTING WITH MATPLOTLIB:

· import numpy as np

import pandas as pd

import matplotlib.pyplot as plt.

I/p - x = np.arange(10)

Y = x ** 2

plt.plot(x, y)

plt.show()

O/p -
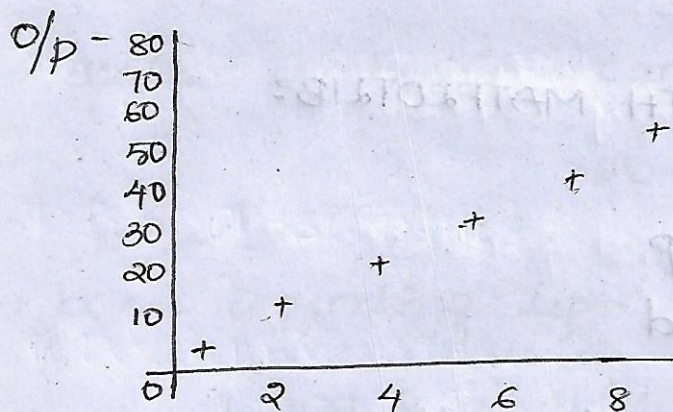
I/P – plt.plot(x, y, 'b+')
plt.show()

O/P –

LABELING:

I/P – plt.plot(x, y, 'bo', x, y, 'k-')

    plt.xlabel('x')
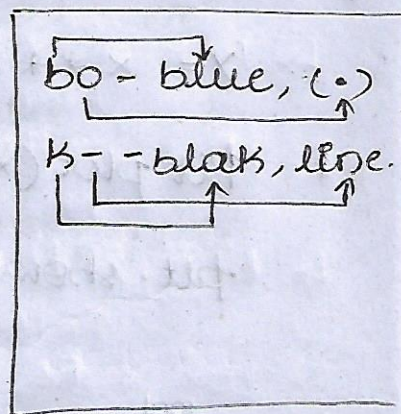
    plt.ylabel('y')

    plt.title('scatterplot – x z y')
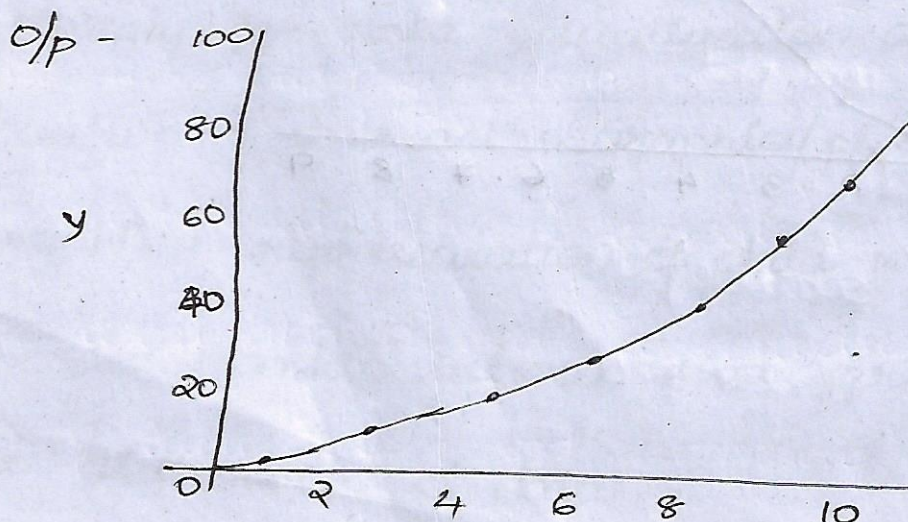
    plt.xlim([0, 10])

    plt.ylim([0, 100])

    plt.show()

bo – blue, (.)
k- – blak, line.

scatter Plot – x z y

O/P –

NOTE :

In the above output, if line is removed it is called as scatterplot.

STATISTICS: It is a form of mathematical analysis that uses quantified models, representations and synopses for a given set of experimental data or real-life studies.

↳ It studies methodologies, review, analyze and draw conclusions from data.

↳ These are used to make better-informed business decisions.

USES OF STATISTICS:

↳ statistics executes the work simply and gives a transparent picture on the work we do regularly.

↳ The statistical methods to help us to examine different areas such as medicine, business, economies, social science and others.

↳ statistics equips us with different kinds of organised data with the help of graphs, tables, diagrams and charts.

TYPES OF STATISTICS:

1. Descriptive statistics

2. Inferential statistics. — ESTIMATION (OR) GUESSING

3 Probability. — CAN'T PREDICT FOR CERTAINITY

1. DESCRIPTIVE STATISTICS:

In this type of statistics, the data is summarised through the given observations. the summarisation is one form a sample of population using parameters such as the mean or standard deviation.

Descriptive statistics is a way to organise, represent and describe a collection of data using table, graphs and summary measures.

→ Descriptive statistics are used to synopsize data from a sample exercising the mean or standard deviation.

↳ In simple, descriptive statistics is defined as some knowledge of my data to learn Properties and parameters.

The Descriptive statistics is further classified into some categories.

MEAN : It is also called as central tendency or measure of central tendency.

→ mean is the mathematical average of a group of two or more numerals.

PROPERTIES :

↳ Univariate Analysis ⟷ Measure of central Tendency

↳

MEAN   MEDIAN

MEDIAN : Measure of central tendency.

→ Sort all the values either in ascending or descending order.

→ Pick the middle value :

ODD : $\left(\dfrac{n+1}{2}\right)^{th}$ term

EVEN : Average of middle most values.

$$\dfrac{\left(\dfrac{n}{2}\right)^{th} + \left[\left(\dfrac{n}{2}\right)+1\right]^{th}}{2}$$

$$\boxed{\text{MEAN} : \mu = \sum_{i=1}^{n} \dfrac{x}{n}}$$

I/P – 10, 20, 30, 40, 50. Calculate the mean, median.

O/P –   ~~Middle~~

$$= \dfrac{\text{Average of given numbers}}{n}$$

$$= \dfrac{10 + 20 + 30 + 40 + 50}{5}$$

$\mu = 30$.

MEDIAN : 10, 20, 30, 40, 50.

ODD - middle value = 30.

I/p - 10, 20, 30, 40, 50, 60.

EVEN - middle most values average

$$= \frac{30+40}{2}$$

$$= 35$$

NOTE :

→ Mean is done faster than the median

→ If outliers are there, then median is prefered since its easy and best.