Aryan Mohan

500092142

Batch- 2

Machine Learning Lab

Experiment- 3

Objective: Preparing synthetic data set to plot scatter plot using M.S.Excel and Python

Syntax:

import pandas as pd

import matplotlib.pyplot as plt

Experiment 3: Prepairing synthetic dataset to plot a graph to find out the pattern/relationship with scatterplot in ms excel/python. Study fatures of scatter plot.

uidsalcar_df=pd.read_csv('C:/Users/aryan/OneDrive/Documents/SalaryCarPrice.csv')

print(uidsalcar_df)

Output:

	User ID	Annual_Salary	Car_Price
0	1	120000	205000
1	2	135000	354000
2	3	145000	205000
3	4	155000	354000
4	5	240000	399000
5	6	460000	354000
6	7	520000	399000
7	8	560000	397000
8	9	620000	354000
9	10	700000	599000
10	11	760000	599000
11	12	840000	402000
12	13	890000	470000
13	14	950000	649000
14	15	1050000	599000
15	16	1120000	425000
16	17	1250000	963000
17	18	1390000	1064000
18	19	1450000	999000
19	20	1520000	769000

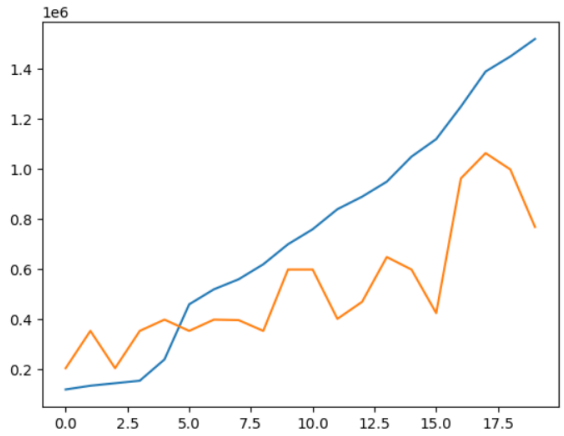
Syntax:

plt.plot(uidsalcar_df)

plt.show

Output:

```
<function matplotlib.pyplot.show(close=None, block=None)>
```



Syntax:

columns=["Annual_Salary", "Car_Price"]

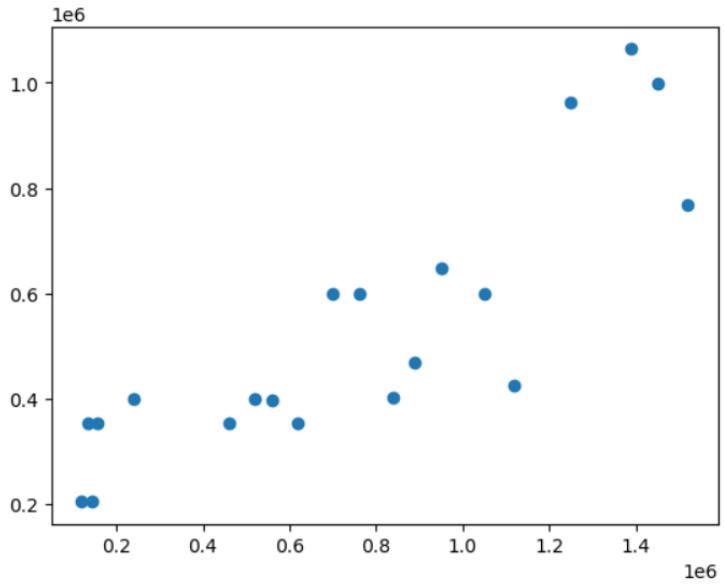
uidsalcar_df=pd.read_csv('C:/Users/aryan/OneDrive/Documents/SalaryCarPrice.csv', usecols=columns)

plt.scatter(uidsalcar_df.Annual_Salary, uidsalcar_df.Car_Price)

plt.show

Output:

```
<function matplotlib.pyplot.show(close=None, block=None)>
```



Features of scatter plot:

A scatter plot is a graphical representation of two variables that uses dots to represent individual data points. The position of each dot on the x- and y-axes represents values for an individual data point. The following are some of the features of a scatter plot:

- X- and Y-Axes: The x-axis and y-axis represent the two variables being compared. The x-axis is typically used to represent the independent variable and the y-axis is used to represent the dependent variable.
- Data Points: Each dot on the scatter plot represents a single data point and its position on the x- and y-axes represents the values of the two variables for that data point.
- Trends and Patterns: Scatter plots can be used to identify trends and patterns in the data. For example, a positive correlation between two variables will result in a upward slope in the scatter plot, while a negative correlation will result in a downward slope.
- Outliers: Scatter plots can also be used to identify outliers, which are data points that are significantly different from the rest of the data.
- Density: The density of data points on a scatter plot can provide additional information about the relationship between the two variables. For example, a high density of data points in a certain area of the plot may indicate a strong relationship between the two variables in that region.
- Customization: Scatter plots can be customized with different colors, shapes, and sizes to highlight specific data points or groups of data points.