

Data Analytics



Dr. Umber Noreen

Outline

1

What is Power BI

2

Data Visualization

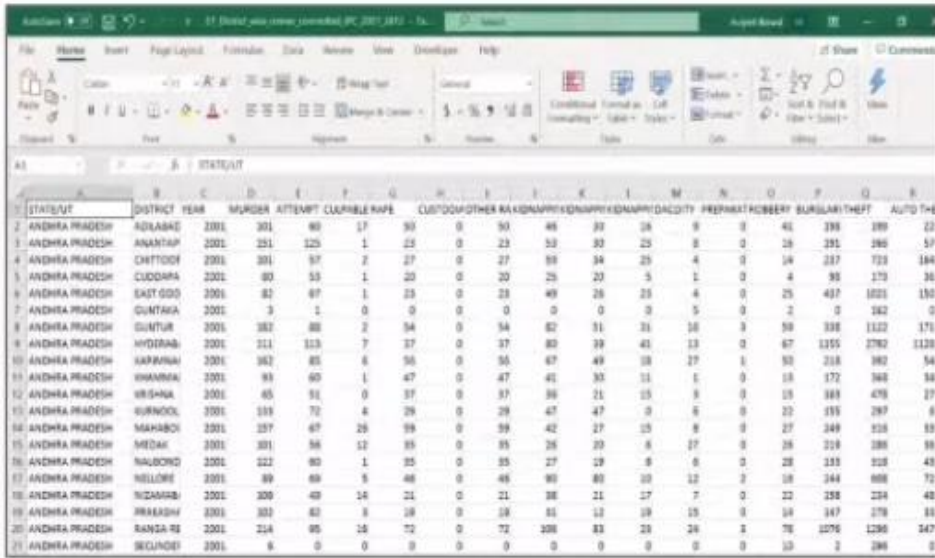
3

Python programming

4

Libraries

MS Power BI is the Business Intelligence tool to visualize your data and share insights across your organization



STATE/UT	DISTRICT	YEAR	MURDER	ATTEMPT	CLARIBLE	RAPE	CLUB/DOO/OTHER	RAKIDAN	PRITIDAN	PRIDAN	PRIDAN	DACCITY	PREPNAIT	HOBBERY	BURGLARI	THEFT	AUTO THE
ANDHRA PRADESH	AGILABAD	2001	301	60	17	93	0	93	46	37	28	9	0	41	399	189	22
ANDHRA PRADESH	ANANTAP	2001	251	125	1	25	0	23	52	30	25	8	0	26	391	260	57
ANDHRA PRADESH	CHITTOOR	2001	301	57	2	27	0	27	98	34	25	4	0	14	237	723	164
ANDHRA PRADESH	CUDDAPUR	2001	90	53	1	20	0	20	26	20	5	1	0	4	99	173	36
ANDHRA PRADESH	SATYADRA	2001	82	87	1	23	0	23	49	26	23	4	0	25	437	1021	150
ANDHRA PRADESH	GUNTUR	2001	3	1	0	0	0	0	0	0	0	0	0	2	0	162	0
ANDHRA PRADESH	GUNTUR	2001	382	88	2	94	0	94	82	51	31	18	3	58	338	1122	171
ANDHRA PRADESH	HYDRABAD	2001	211	113	7	27	0	27	83	39	41	13	0	67	1355	2782	1128
ANDHRA PRADESH	KAPURTHALA	2001	367	85	6	56	0	56	67	49	38	27	1	50	318	382	54
ANDHRA PRADESH	KHARSA	2001	93	60	1	47	0	47	41	30	11	1	0	13	172	360	54
ANDHRA PRADESH	KRISHNA	2001	65	51	0	37	0	37	36	21	15	5	0	13	383	476	27
ANDHRA PRADESH	KURNOOL	2001	133	72	4	26	0	26	47	47	0	6	0	22	355	287	6
ANDHRA PRADESH	MAHABUB	2001	137	67	29	59	0	59	42	27	15	8	0	27	349	518	83
ANDHRA PRADESH	MEDAK	2001	301	56	12	35	0	35	26	20	6	27	0	26	319	386	56
ANDHRA PRADESH	NALAND	2001	122	80	1	35	0	35	27	19	8	6	0	28	313	518	43
ANDHRA PRADESH	NELLORE	2001	89	89	5	46	0	46	90	80	10	12	2	18	344	608	72
ANDHRA PRADESH	NIZAMABAD	2001	209	49	14	21	0	21	38	21	17	7	0	22	258	234	48
ANDHRA PRADESH	RAJAHMUNDRAM	2001	302	62	8	18	0	18	51	12	19	15	8	14	347	278	83
ANDHRA PRADESH	RANGA REDDY	2001	214	95	18	72	0	72	308	83	28	26	1	76	1076	1296	147
ANDHRA PRADESH	SECUNDERABAD	2001	6	0	0	0	0	0	0	0	0	0	0	10	2	286	0

Data



Visualization

Power BI

Python Programming



Data Analytics



Python Programming



Why Python



- 1 Beginner friendly
- 2 Versatile and flexible
- 3 Most mature package libraries around
- 4 Career Opportunities
- 5 Startups and Corporates Python for both
- 6 Most popular in ML and Data Science

Python Programming



Why
Python

?

MAJOR COMPANIES
THAT USE



python

Google NETFLIX facebook. Instagram

amazon Quora # slack intel NASA

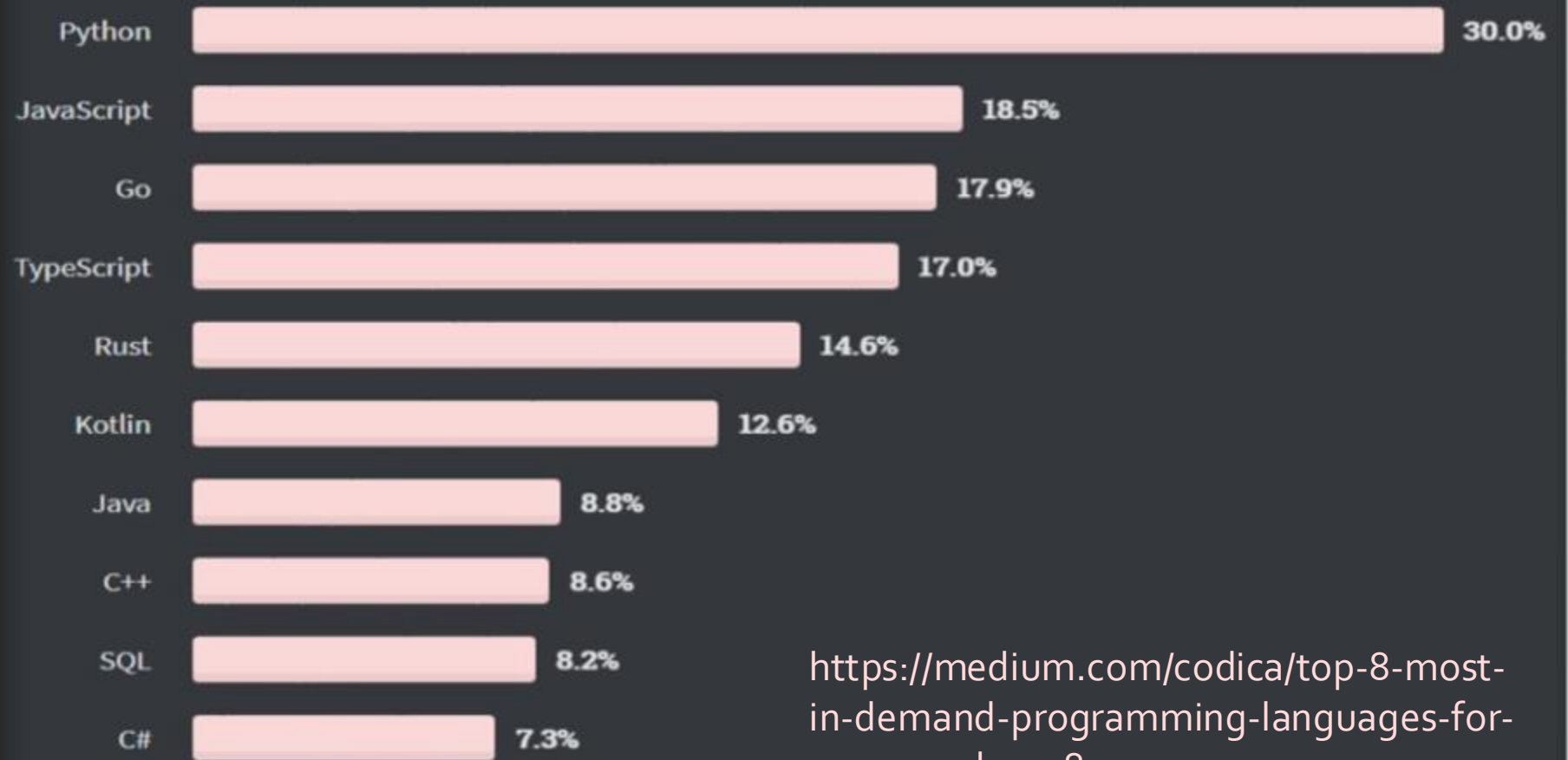
Dropbox ebay Spotify Capital One

Python Programming



Why
Python
?

The most wanted top programming languages



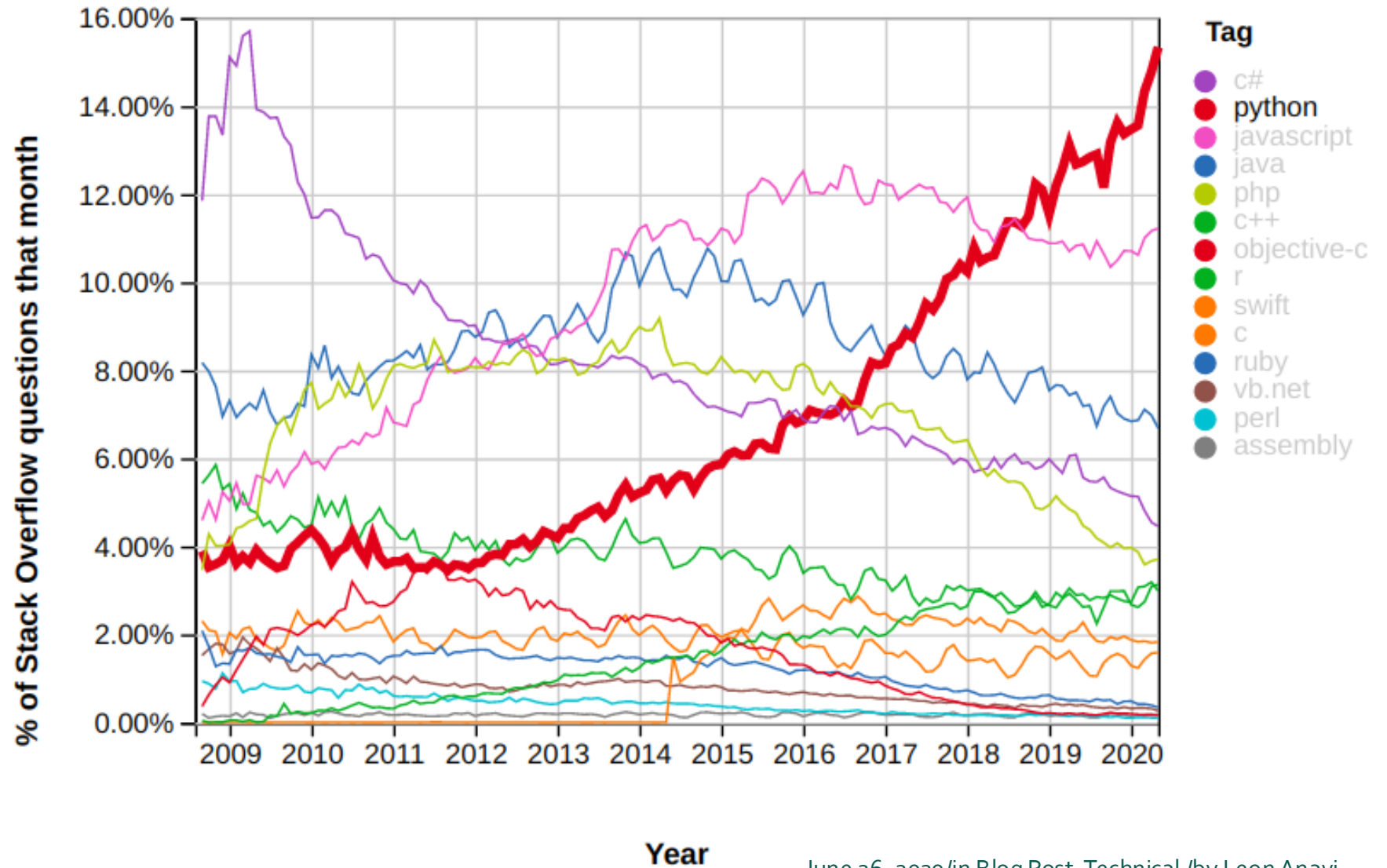
<https://medium.com/codica/top-8-most-in-demand-programming-languages-for-2021-50cd4c3a8c34>

Python Programming



Why
Python

?



Python Programming



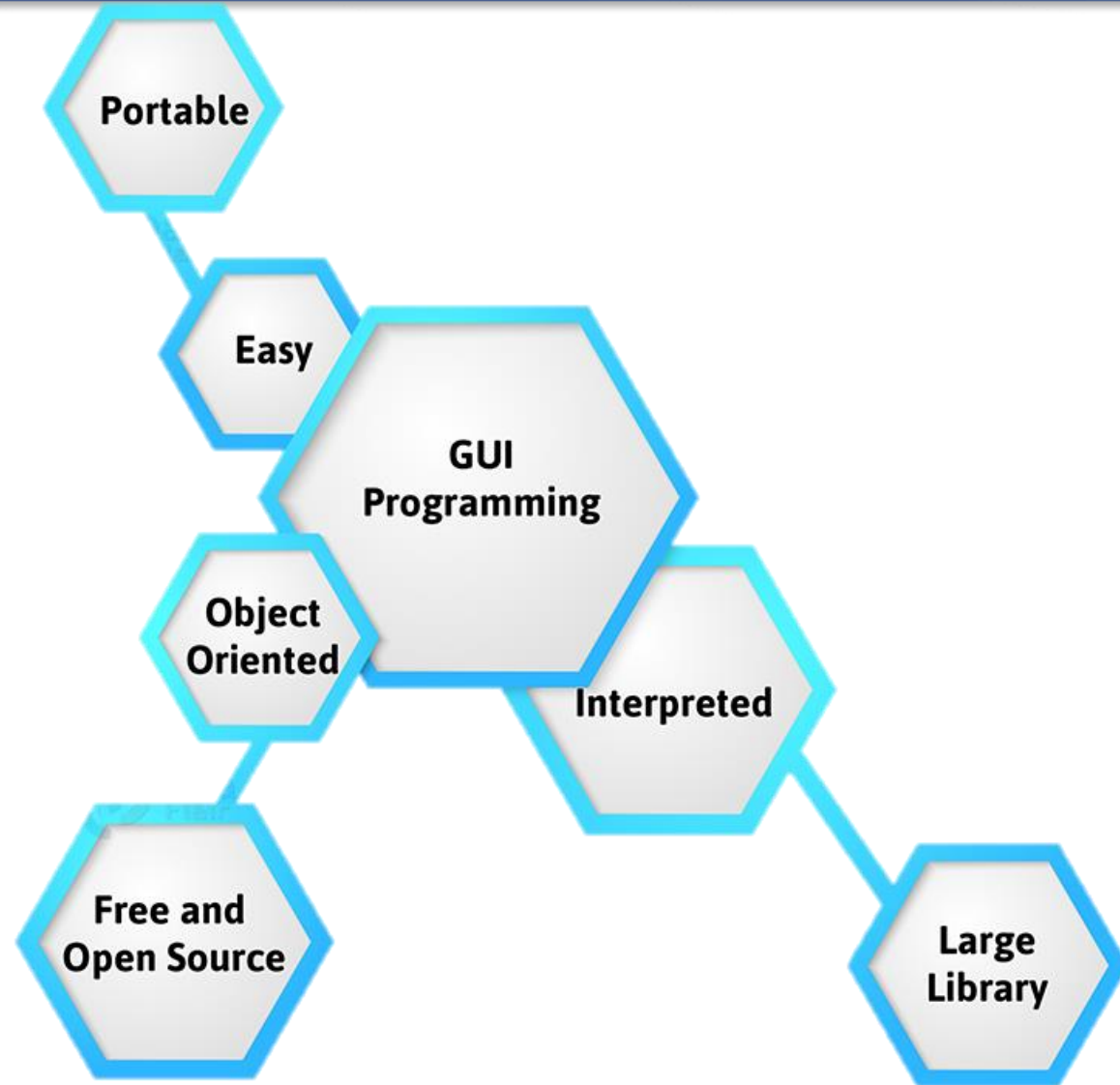
Python
IDE



Python Programming



Python Features



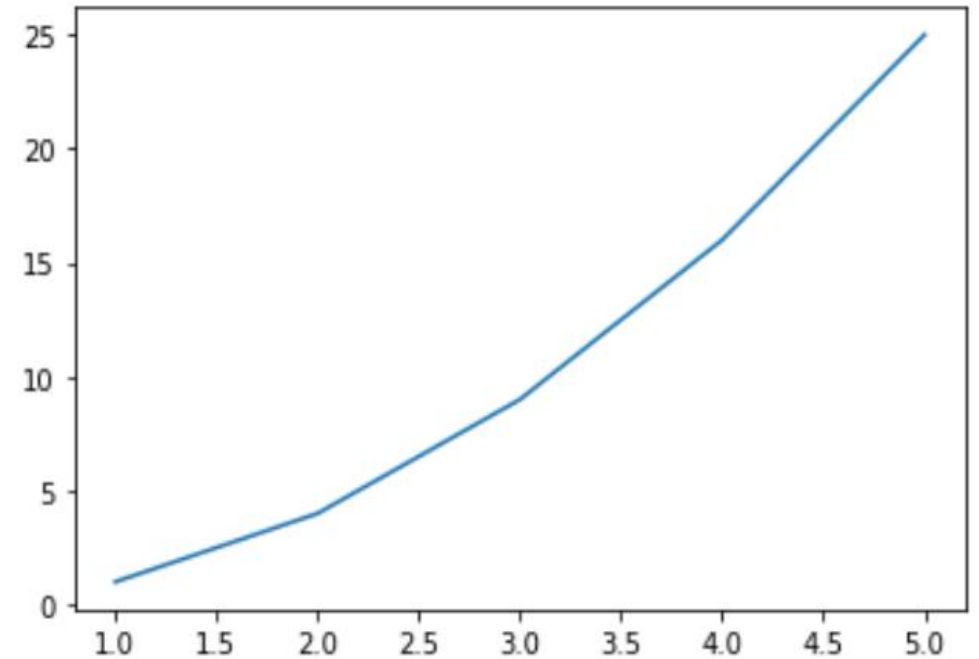
Matplotlib is a comprehensive library for creating static, animated, and interactive visualizations in Python

Library Matplotlib

```
import matplotlib.pyplot as plt

X=[1,2,3,4,5]
Y=[1,4,9,16,25]
plt.plot(X,Y)

plt.show()
```



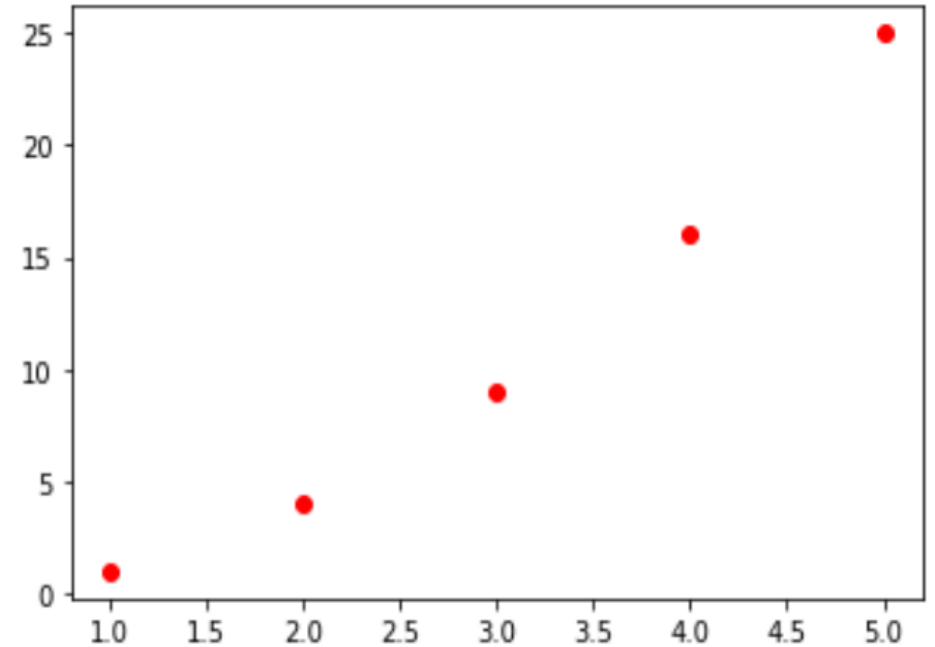
Line Graph

Library Matplotlib

```
import matplotlib.pyplot as plt

X=[1,2,3,4,5]
Y=[1,4,9,16,25]
plt.plot(X,Y, 'ro')

plt.show()
```



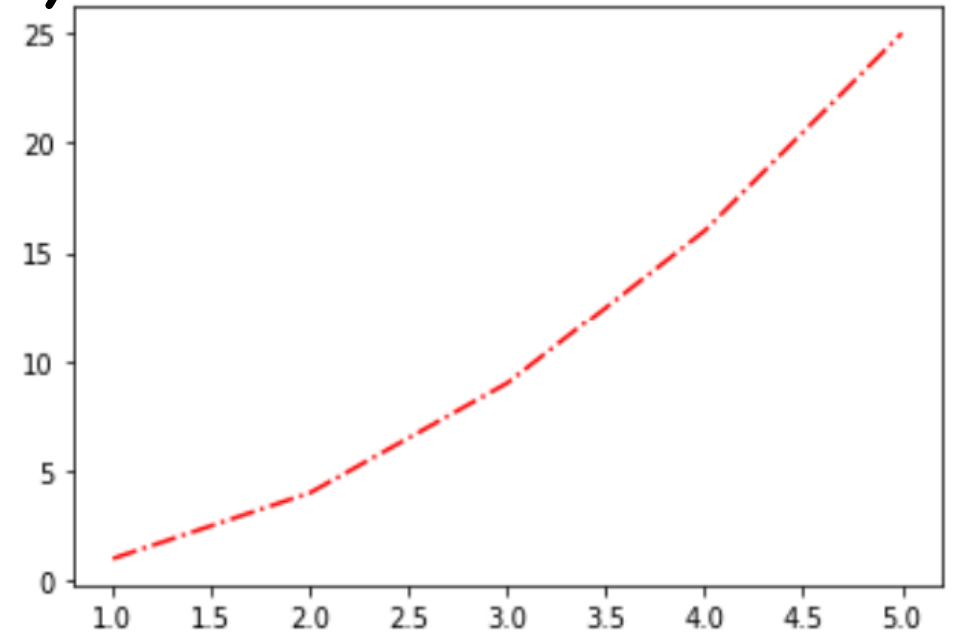
Line Graph

Library Matplotlib

```
import matplotlib.pyplot as plt

X=[1,2,3,4,5]
Y=[1,4,9,16,25]
plt.plot(X,Y, 'r-.')

plt.show()
```



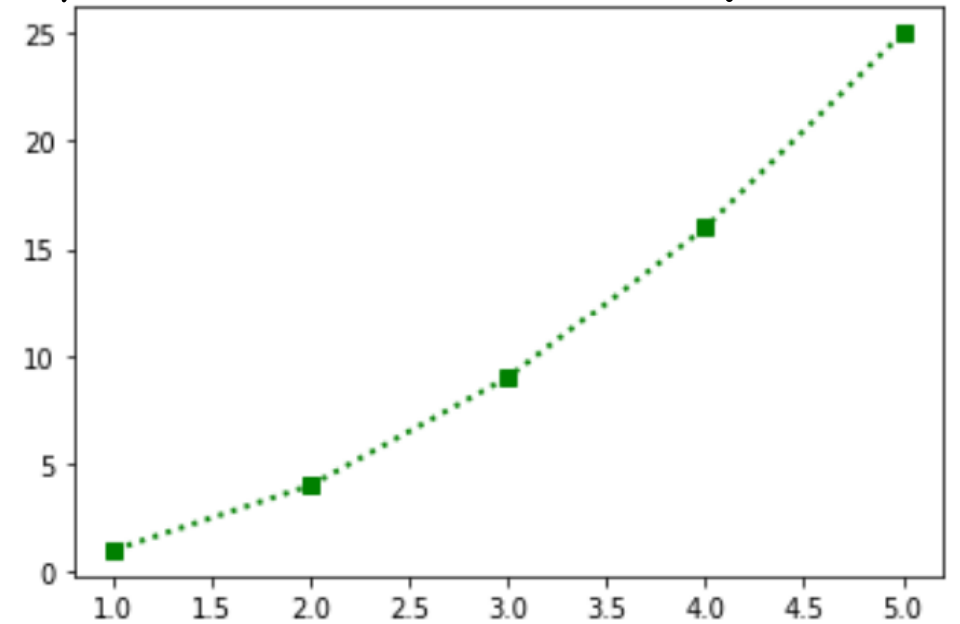
Line Graph

Library Matplotlib

```
import matplotlib.pyplot as plt

X=[1,2,3,4,5]
Y=[1,4,9,16,25]
plt.plot(X,Y, 'r-.', linewidth=2.0)

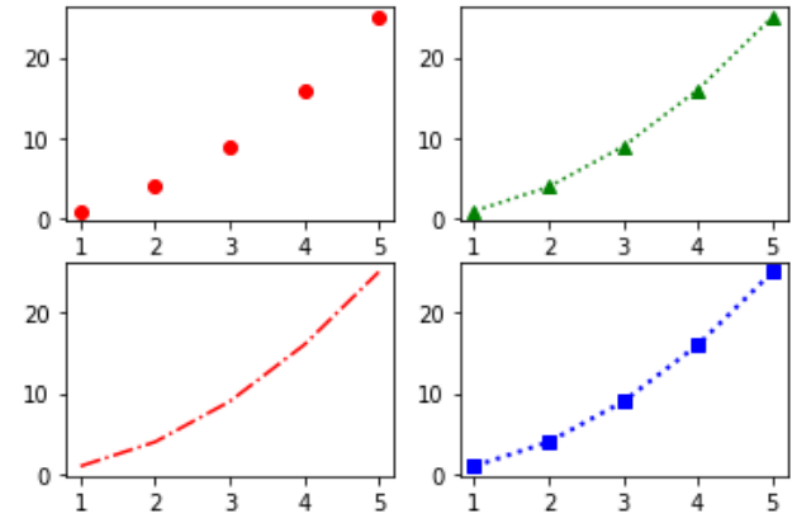
plt.show()
```



Line Graph

Library Matplotlib

```
import matplotlib.pyplot as plt
X=[1,2,3,4,5]
Y=[1,4,9,16,25]
plt.subplot(221)
plt.plot(X,Y,'ro')
plt.subplot(222)
plt.plot(X,Y,'g:^')
plt.subplot(223)
plt.plot(X,Y,'r-.')
plt.subplot(224)
plt.plot(X,Y,'b:s',linewidth=2.0)
plt.show()
```

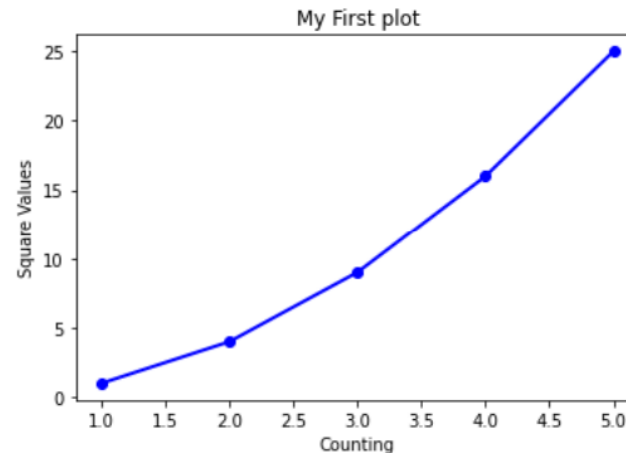


Line Graph

Library Matplotlib

```
import matplotlib.pyplot as plt

X=[1,2,3,4,5]
Y=[1,4,9,16,25]
plt.title("My First plot")
plt.xlabel("Counting")
plt.ylabel("Square Values")
plt.plot(X,Y, 'r-', linewidth=2.0)
plt.show()
```



Line Graph

Library Matplotlib

```
import matplotlib.pyplot as plt

X=[1,2,3,4,5]
Y=[1,4,9,16,25]
plt.title("My First plot", fontsize=20)
plt.xlabel("Counting", color="r")
plt.ylabel("Square Values", color="g")
plt.plot(X,Y,'b-o',linewidth=2.0)

plt.show()
```



Line Graph

Library Matplotlib

```
import matplotlib.pyplot as plt

X=[1,2,3,4,5]
Y=[1,4,9,16,25]
plt.title("My First plot", fontsize=20)
plt.xlabel("Counting", color="r")
plt.ylabel("Square Values", color="g")
plt.plot(X,Y,'b-o',linewidth=2.0)
plt.legend(['First series'])
plt.show()
```

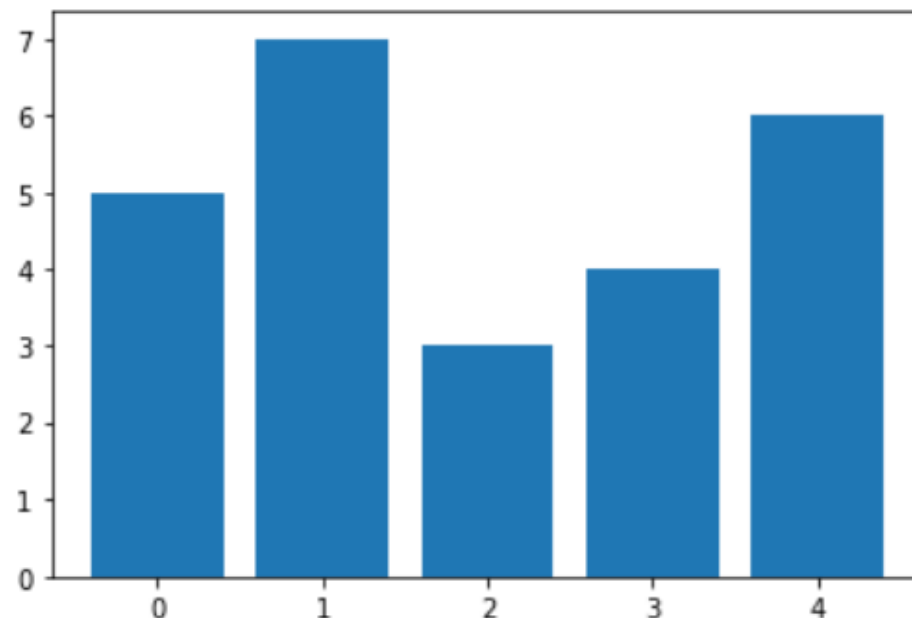


Bar Charts

Library
Matplotlib

```
import matplotlib.pyplot as plt

index = [0,1,2,3,4]
values = [5,7,3,4,6]
plt.bar(index,values)
plt.show()
```

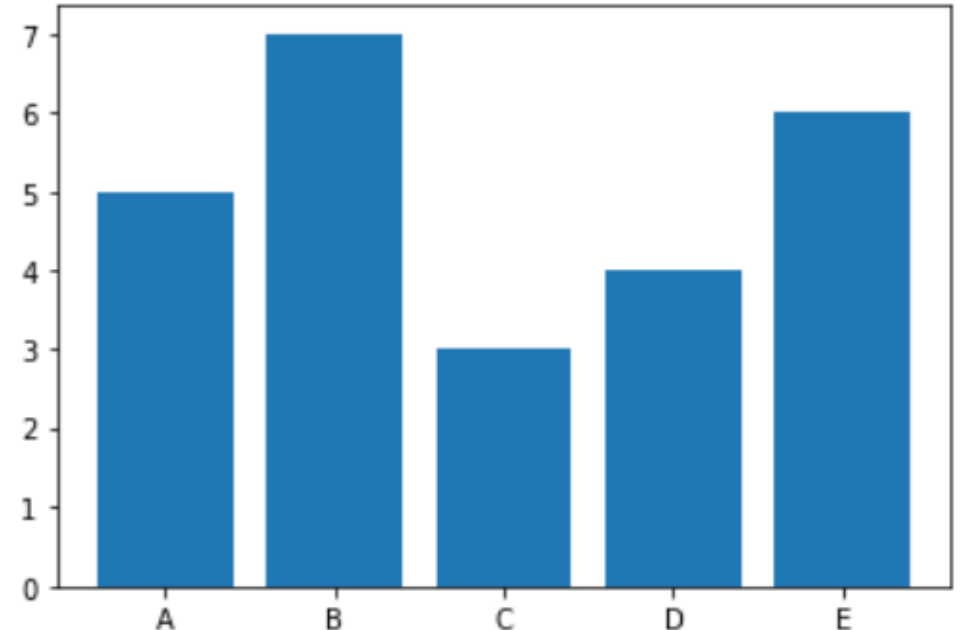


Bar Charts

Library Matplotlib

```
import matplotlib.pyplot as plt

index = ['A', 'B', 'C', 'D', 'E']
values = [5, 7, 3, 4, 6]
plt.bar(index, values)
plt.show()
```

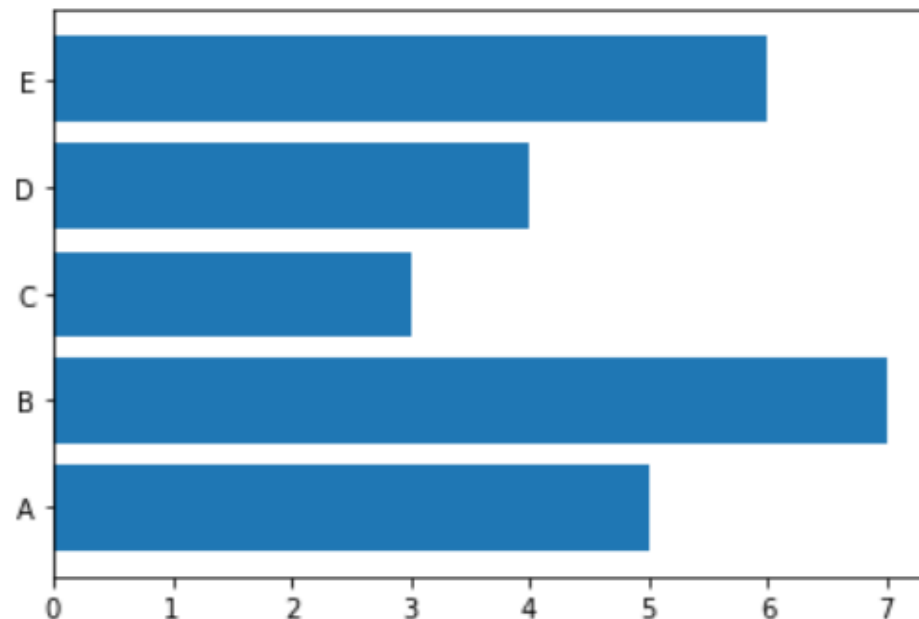


Bar Charts

Library Matplotlib

```
import matplotlib.pyplot as plt

index = ['A', 'B', 'C', 'D', 'E']
values = [5, 7, 3, 4, 6]
plt.barh(index, values)
plt.show()
```

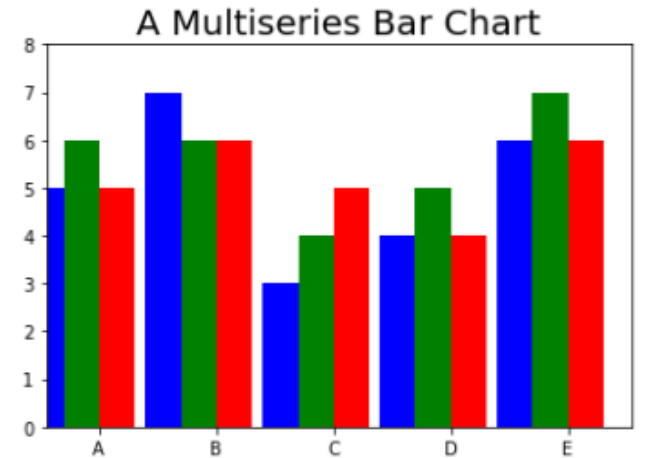


Multiple Bar Charts

Library Matplotlib

```
import matplotlib.pyplot as plt
import numpy as np

index = np.arange(5)
values1 = [5, 7, 3, 4, 6]
values2 = [6, 6, 4, 5, 7]
values3 = [5, 6, 5, 4, 6]
plt.axis([0, 5, 0, 8])
plt.title('A Multiseries Bar Chart', fontsize=20)
bw=0.3
plt.bar(index, values1, bw, color='b')
plt.bar(index+bw, values2, bw, color='g')
plt.bar(index+2*bw, values3, bw, color='r')
plt.xticks(index+1.5*bw, ['A', 'B', 'C', 'D', 'E'])
plt.show()
```



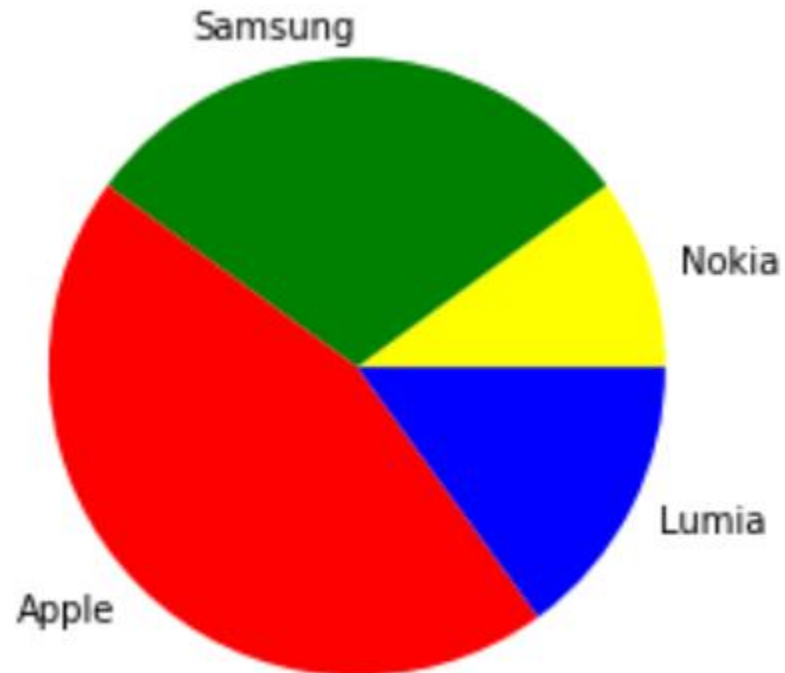
Pie Charts

Library Matplotlib

```
import matplotlib.pyplot as plt

labels = ['Nokia', 'Samsung', 'Apple', 'Lumia']
values = [10, 30, 45, 15]
colors = ['yellow', 'green', 'red', 'blue']
plt.pie(values, labels=labels, colors=colors)

plt.show()
```



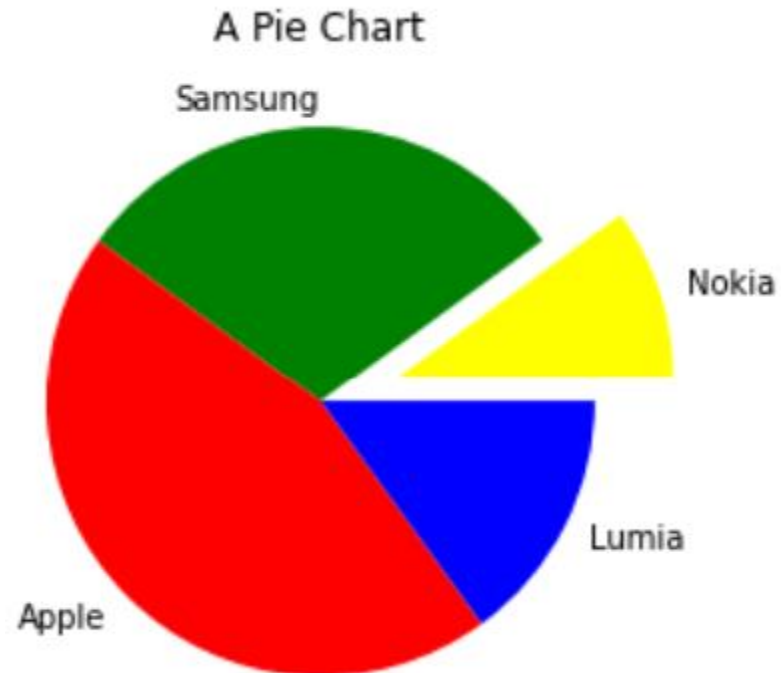
Pie Charts

Library Matplotlib

```
import matplotlib.pyplot as plt

labels = ['Nokia', 'Samsung', 'Apple', 'Lumia']
values = [10, 30, 45, 15]
colors = ['yellow', 'green', 'red', 'blue']

explode = [0.3, 0, 0, 0]
plt.title('A Pie Chart')
plt.pie(values, labels=labels, colors=colors, explode=explode)
plt.show()
```



Pie Charts

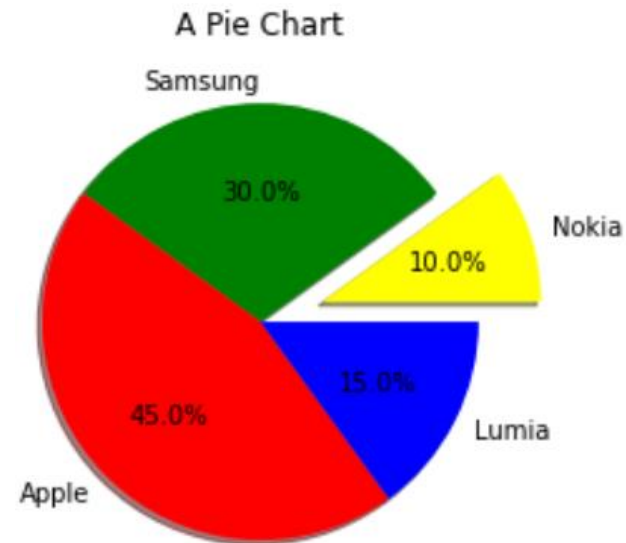
Library Matplotlib

```
import matplotlib.pyplot as plt

labels = ['Nokia', 'Samsung', 'Apple', 'Lumia']
values = [10, 30, 45, 15]
colors = ['yellow', 'green', 'red', 'blue']

explode = [0.3, 0, 0, 0]
plt.title('A Pie Chart')

plt.pie(values, labels=labels, colors=colors, explode=explode, shadow=True, autopct='%1.1f%%')
plt.show()
```



Data Frame: A data frame is a multi-dimensional array where the rows and the columns can be labeled.

```
1 import pandas as pd
2 import numpy as np
3
4 myarray = np.array([[1, 2, 3], [4, 5, 6]])
5 rownames = ['a', 'b']
6 colnames = ['one', 'two', 'three']
7 mydataframe = pd.DataFrame(myarray, index=rownames, columns=colnames)
8
9 print(mydataframe)
```

	one	two	three
a	1	2	3
b	4	5	6

Load CSV File with Pandas: You can load your CSV data using Pandas and the `pandas.read_csv()` function.

How To Load Machine Learning Data

- `import pandas as pd`
- `# import pandas library to read dataset`
- `df = pd.read_csv('salary_data.csv')`
- `# if program file and Dataset both are in Same folder`
- `df = pd.read_csv('D:\\03 LECTURES\\000 WCC\\o FALL 2021\\Machine Learning\\Week-02-D-21-10-2021\\salary_data.csv')`
- `# Windows format for file path`
- `df = pd.read_csv('D:/03 LECTURES/000 WCC/o FALL 2021/Machine Learning/Week-02-D-21-10-2021/salary_data.csv')`
- `# Linux/MacOS format for file path`

Read Dataset and display it

- `import pandas as pd`
- `df = pd.read_csv('salary_data.csv')`
- `print(df)`

View the First and last 6 rows of your data using the `head()` and `tail()` function on the Pandas Data Frame

Understanding Your Data With Descriptive Statistics

- `import pandas as pd`
- `df = pd.read_csv('salary_data.csv')`
- `#print(df)`
- `print(df.head(6))`
- `print(df.tail(6))`
- `print(df.sample(6))`

Dimensions of Your Dataset: View the shape and size of your dataset by printing the shape property on the Pandas DataFrame

Understanding Your Data With Descriptive Statistics

- `import pandas as pd`
- `df = pd.read_csv('salary_data.csv')`
- `print(df.shape)`

Data Type For Each Attribute: list the data types used by the DataFrame to characterize each attribute using the `dtypes` property

Understanding Your Data With Descriptive Statistics

- `import pandas as pd`
- `df = pd.read_csv('salary_data.csv')`
- `print(df.dtypes)`

Descriptive Statistics

The `describe()` function on the Pandas DataFrame lists 8 statistical properties of each attribute. They are:

- Count
- Mean
- Standard Deviation
- Minimum Value
- 25th Percentile
- 50th Percentile (Median)
- 75th Percentile
- Maximum Value

Understanding Your Data With Descriptive Statistics

Understanding Your Data With Descriptive Statistics

- `import pandas as pd`
- `df = pd.read_csv('salary_data.csv')`
- `print(df. describe())`

Understanding Your Data With Descriptive Statistics

- `import pandas as pd`
- `df = pd.read_csv('salary_data.csv')`
- `print(df. describe())`

Understanding Your Data With Descriptive Statistics

- `import pandas as pd`
- `df = pd.read_csv('salary_data.csv')`
- `print(df.info())`

Understanding Your Data With Descriptive Statistics

- `import pandas as pd`
- `df = pd.read_csv('salary_data.csv')`
- `print(df.columns)`

Understanding Your Data With Descriptive Statistics

- `import pandas as pd`
- `df = pd.read_csv('salary_data.csv')`
- `print(df. size)`
- `print(df. values)`

Visualizing the dataset: The library `matplotlib.pyplot` is used for Visualizing the Dataset

```
import pandas as pd

df = pd.read_csv('salary_data.csv')

import matplotlib.pyplot as plt

plt.xlabel('Years Experience', fontsize=16, color='g')
plt.ylabel('Salary', fontsize=16, color='b')
plt.title("Salary of Employees \n with Experience", fontsize=20, color='r')
plt.scatter(df.YearsExperience,df.Salary,color='b',marker='s')
plt.axis([0,14, 20000, 140000])
plt.grid(True)

plt.show()
```