

3/3/2025

## LAB 1.

\* Write a python program to

- Method 1 Initializing values directly into DataFrame.

```
data = { "USN" : [1, 2, 3], "Name" : ["A", "B", "C"],  
        "Marks" : [100, 100, 100] }
```

- Method 2

Importing datasets from sklearn.datasets

```
from sklearn.datasets import load_diabetes  
diabetes = load_diabetes()
```

```
df = pd.DataFrame(diabetes.data, columns=  
                  diabetes.feature_names
```

```
df.head()
```

- Method 3

Importing datasets from a specific .csv file

```
import pandas as pd
```

```
df = pd.DataFrame("sample_sales_data.csv")
```

```
df
```

- Method 4

Downloading datasets from existing dataset repositories like Kaggle, UCI, Mendely.

```
import pandas as pd
```

```
df = pd.read_csv("content/Dataset_of_Diabetes.csv")  
df.head()
```



output:

	ID	No_pation	Gender	Age	Urea
			M	21	4.7
0	502	1001		25	4.6
1	735	2001	M	42	2.1
2	420	3001	F	49	4.9
3	101	4001	F		

## \* Stock Market Analysis

```
1. import yfinance as yf
import pandas as pd
import matplotlib.pyplot as plt
```

```
tickers = ["HDFCBANK.NS", "ICICIBANK.NS",
            "KOTAKBANK.NS"]
```

```
2) data = yf.download(tickers, start="2024-01-01",
                       end="2024-12-30",
                       group_by="ticker")
```

```
data.head()
```

```
data.shape
```

```
3) Plot the closing price and daily
    returns for all 3 Banks
```

```
h = data["HDFCBANK.NS"]
```

```
i = data["ICICIBANK.NS"]
```

```
k = data["KOTAKBANK.NS"]
```



```
plt.figure(figsize=(12,6))
plt.subplot(2,1,1)
h['close'].plot(title="HDFC")
plt.subplot(2,1,2)
h['Daily Return'].plot(title="HDFC", color =
'orange')

plt.tight_layout()
plt.show()
```

```
i['close'].plot(title="ICICI")
plt.subplot(2,1,2)
i['Daily Return'].plot(title="ICICI", color='red')
```

```
k['close'].plot(title="KOTAK")
plt.subplot(2,1,2)
k['Daily Return'].plot(title="KOTAK", color =
"blue")
```



## LAB 1

i) Loading Dataset.

```
df = pd.DataFrame("load.csv")
```

ii) display columns  
df.columns

iii) Statistical information of all numerical  
df.describe()

iv) To display the count of unique labels for Ocean Proximity  
df['ocean-proximity'].value\_counts()

v) To display which attributes in a dataset have missing values count greater than zero  
missing\_values = df.isnull().sum()  
mv = missing\_values[missing\_values > 0]

1) which columns in the dataset had missing values? How did you handle them?

ocean-proximity had null values replaced with fill technique  
df['ocean-proximity'].fill()

2) which categorical columns did you identify in the dataset? How did you encode them?



df = pd.read\_csv("content/Mobiles.csv",  
encoding="latin-1")  
df.head()

3) Difference Between Min-Max scaling and standardization? what would you use one over the other?

- ans -
- Min max scaling scales to a fixed range, usually  $[0, 1]$
  - Standardization centers data around 0 with a SD of 1
  - Min-max used when data has fixed range
  - Standardization when data has outliers.

Gur 03.03.2021