PROJECT BASED LEARNING-I

GROUP-4

PROJECT-8

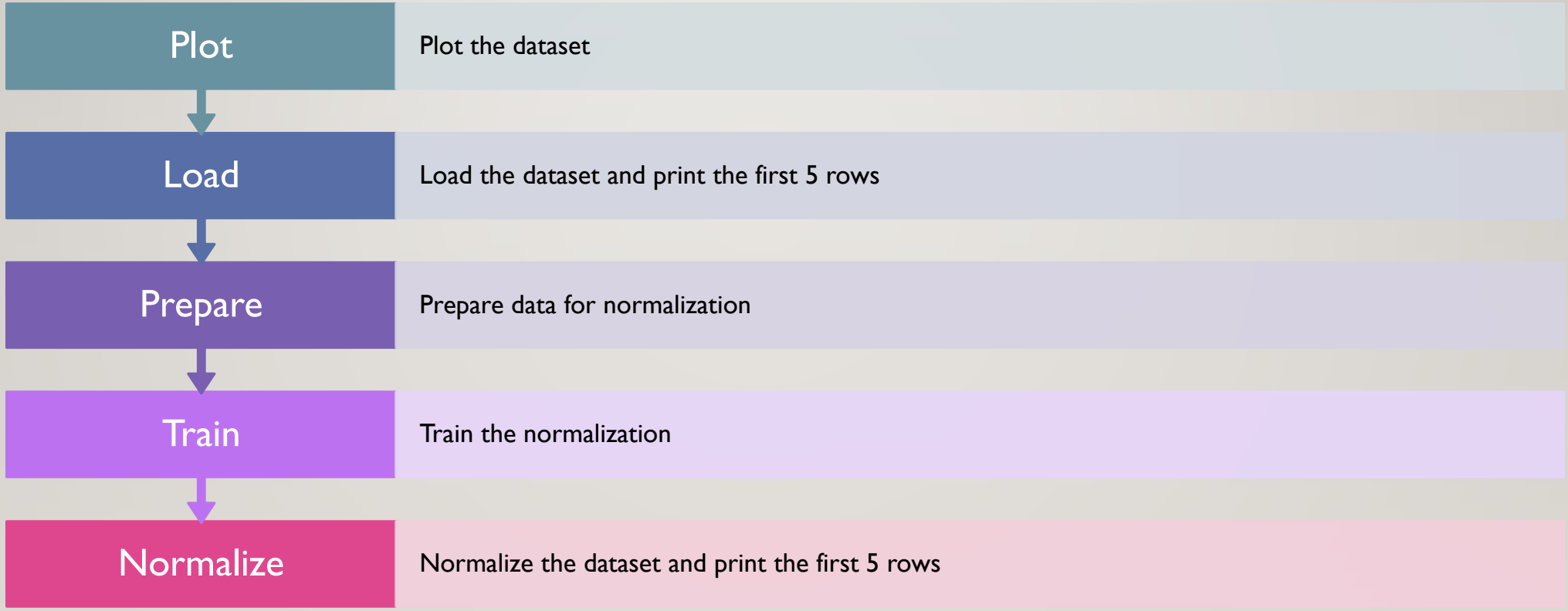
NORMALIZATION:

- The data normalization is **a basic element of data mining(data preprocessing).**
- It means transforming the data, namely converting the source data in to another format that allows processing data effectively.
- The main purpose of data normalization is to minimize or even exclude duplicated data.
- Normalization techniques are linear scaling, clipping log scaling, z-score and even more

REQUIRED PACKAGES

- NUMPY
- PANDAS
- SKLEARN
- MIN-MAX SCALER

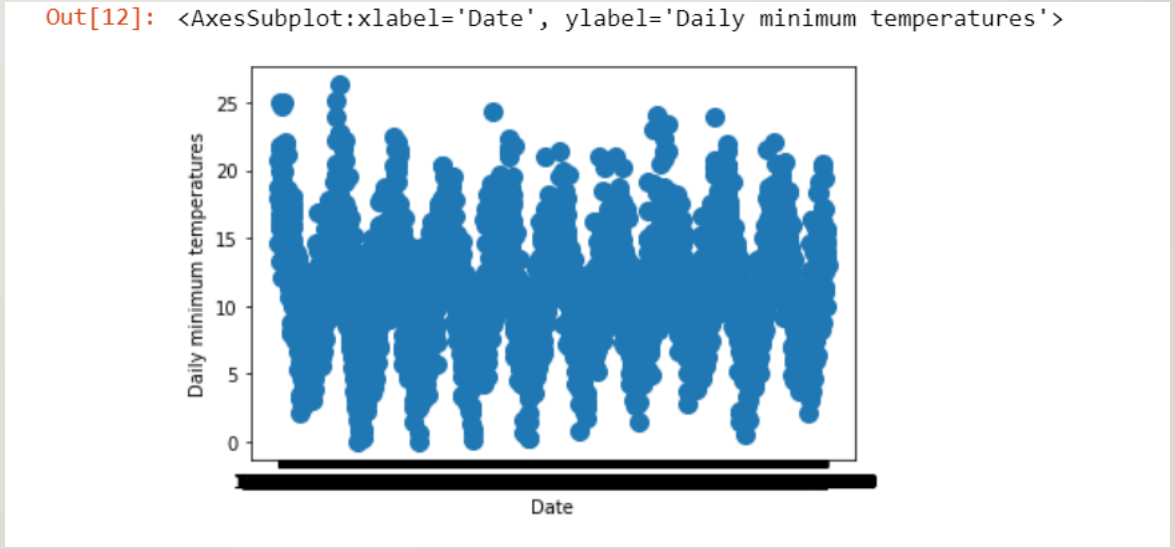
STEPS NEED TO BE TAKEN:



PLOTTING THE DATASET:

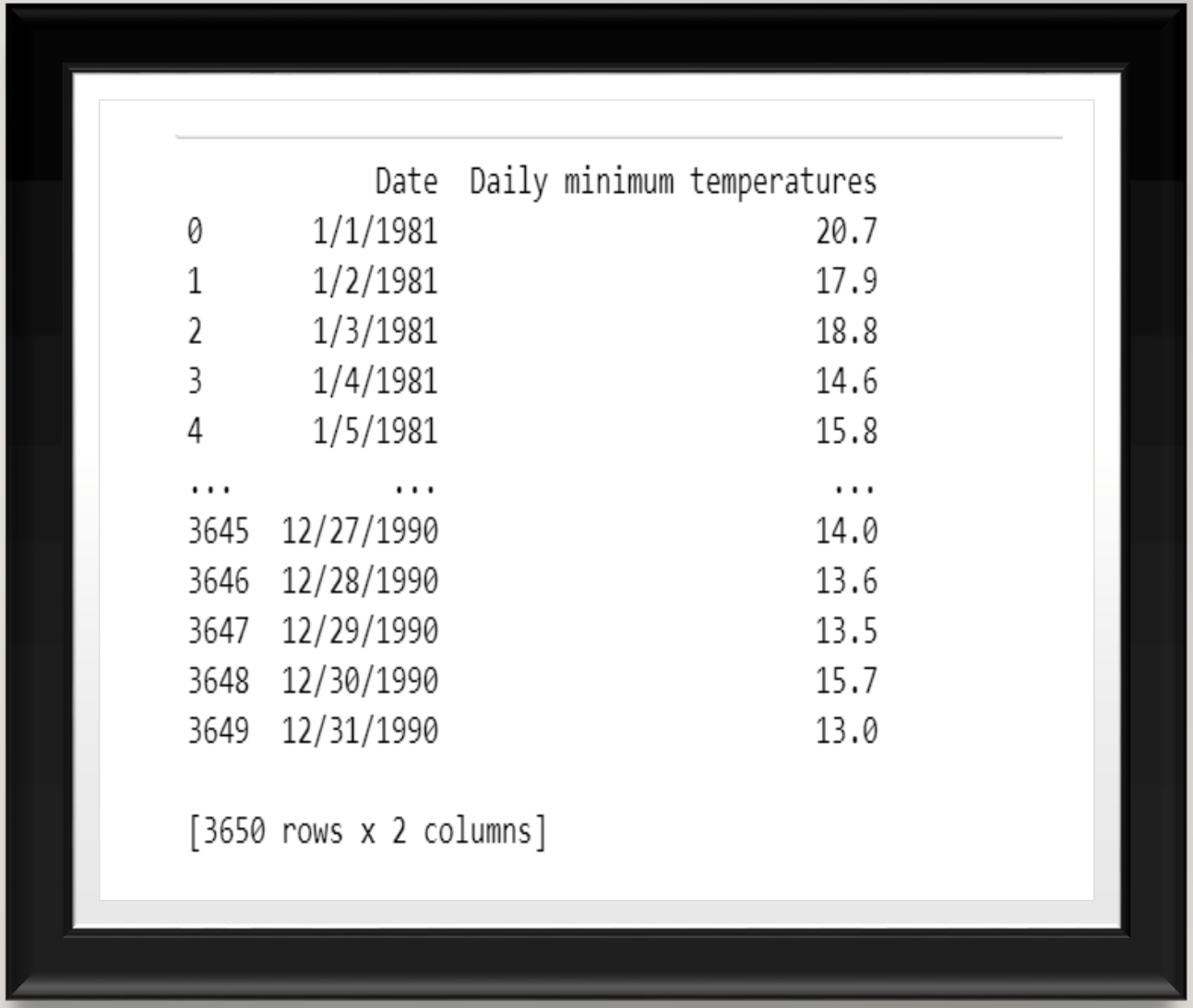
```
data.plot.scatter(x='Date',y='Daily minimum temperatures',s=100)
```

```
Out[12]: <AxesSubplot:xlabel='Date', ylabel='Daily minimum temperatures'>
```



LOADING THE DATASET:

- import numpy as np
- import pandas as pd
- data=pd.read_csv("daily-minimum-temperatures.csv")
- print(data)



	Date	Daily minimum temperatures
0	1/1/1981	20.7
1	1/2/1981	17.9
2	1/3/1981	18.8
3	1/4/1981	14.6
4	1/5/1981	15.8
...	...	...
3645	12/27/1990	14.0
3646	12/28/1990	13.6
3647	12/29/1990	13.5
3648	12/30/1990	15.7
3649	12/31/1990	13.0

[3650 rows x 2 columns]

PREPARE DATA FOR NORMALIZATION:

- from sklearn import preprocessing
- values= preprocessing.normalize(df)
- print(values)

TRAINING THE NORMALIZATION:

- series =pd.read_csv("daily-minimum-temperatures.csv", header=0, index_col=0)
- values = series.values
- values = values.reshape((len(values), 1))

```
scaler = MinMaxScaler(feature_range=(0, 1))
scaler = scaler.fit(values)
print('Min: %f, Max: %f' % (scaler.data_min_,
scaler.data_max_))
```

NORMALIZE AND PRINT FIRST 5 ROWS:

```
normalized = scaler.transform(values)
df1=pd.DataFrame(normalized)
df1.head()
```

Out[15]:

	0
0	0.787072
1	0.680608
2	0.714829
3	0.555133
4	0.600760

INFERENCE AND ANALYSIS:

- In this project we applied normalization techniques on the dataset.
- Some normalization techniques are zscore, min-max scaling, decimal scaling, etc…
- Here, in this project I have applied min-max scaling which results in scaled data values.


Thank
You!