

```
In [ ]: import pandas as pd
import numpy as np
```

```
In [ ]: rainfall_df = pd.read_excel('./Data/RainfallCL3175771.xlsx', header=8)
```

```
In [ ]: rainfall_df.shape
```

```
Out[ ]: (1826, 9)
```

```
In [ ]: rainfall_df.head()
```

```
Out[ ]:
```

	SL	DISTRICT	UPAZILA	STATION ID	STATION NAME	DATE	RAIN FALL(mm)	LATITUDE	LONGITUDE
0	1	Bandarban	Lama	CL317	Lama	01-jan-2017	0.0	21.81	92.19
1	2	Bandarban	Lama	CL317	Lama	02-jan-2017	0.0	21.81	92.19
2	3	Bandarban	Lama	CL317	Lama	03-jan-2017	0.0	21.81	92.19
3	4	Bandarban	Lama	CL317	Lama	04-jan-2017	0.0	21.81	92.19
4	5	Bandarban	Lama	CL317	Lama	05-jan-2017	0.0	21.81	92.19

```
In [ ]: rainfall_df.columns
```

```
Out[ ]: Index(['SL', 'DISTRICT', 'UPAZILA', 'STATION ID', 'STATION NAME', 'DATE',  
            'RAIN FALL(mm)', 'LATITUDE', 'LONGITUDE'],  
            dtype='object')
```

```
In [ ]: rainfall_df.dtypes
```

```
Out[ ]: SL                int64  
DISTRICT                object  
UPAZILA                 object  
STATION ID              object  
STATION NAME            object  
DATE                   object  
RAIN FALL(mm)           float64  
LATITUDE                float64  
LONGITUDE               float64  
dtype: object
```

```
In [ ]: rainfall_df['RAIN FALL(mm)'].value_counts()
```

```
Out[ ]: 0.0      1159  
4.0        34  
6.0        32
```

```
2.0      25
3.0      23
```

```
...
16.5     1
172.0    1
93.0     1
160.0    1
7.6      1
```

Name: RAIN FALL(mm), Length: 180, dtype: int64

```
In [ ]: rainfall_df.tail()
```

```
Out[ ]:
```

	SL	DISTRICT	UPAZILA	STATION ID	STATION NAME	DATE	RAIN FALL(mm)	LATITUDE	LONGITUDE
1821	1822	Bandarban	Lama	CL317	Lama	27-dec-2021	0.0	21.81	92.19
1822	1823	Bandarban	Lama	CL317	Lama	28-dec-2021	0.0	21.81	92.19
1823	1824	Bandarban	Lama	CL317	Lama	29-dec-2021	0.0	21.81	92.19
1824	1825	Bandarban	Lama	CL317	Lama	30-dec-2021	0.0	21.81	92.19
1825	1826	Bandarban	Lama	CL317	Lama	31-dec-2021	0.0	21.81	92.19

```
In [ ]: waterLevel_df = pd.read_excel('./Data/Water_LevelSW2035770.xlsx', header=8)
```

```
In [ ]: waterLevel_df.head()
```

```
Out[ ]:
```

	SL	DISTRICT	UPAZILA	RIVER	STATION ID	STATION NAME	STATION TYPE	DATA TYPE	DATE/TIME	WATE LEVEL(m)
0	1	Bandarban	Lama	Matamuhuri	SW203	Lama	Non Tidal	Regular	1 January, 2017 06:00:am	6.2
1	2	Bandarban	Lama	Matamuhuri	SW203	Lama	Non Tidal	Regular	1 January, 2017 09:00:am	6.2
2	3	Bandarban	Lama	Matamuhuri	SW203	Lama	Non Tidal	Regular	1 January, 2017 12:00:pm	6.2
3	4	Bandarban	Lama	Matamuhuri	SW203	Lama	Non Tidal	Regular	1 January, 2017 03:00:pm	6.2
4	5	Bandarban	Lama	Matamuhuri	SW203	Lama	Non Tidal	Regular	1 January, 2017 06:00:pm	6.2

```
In [ ]: waterLevel_df.shape
```

Out[ ]: (9130, 12)

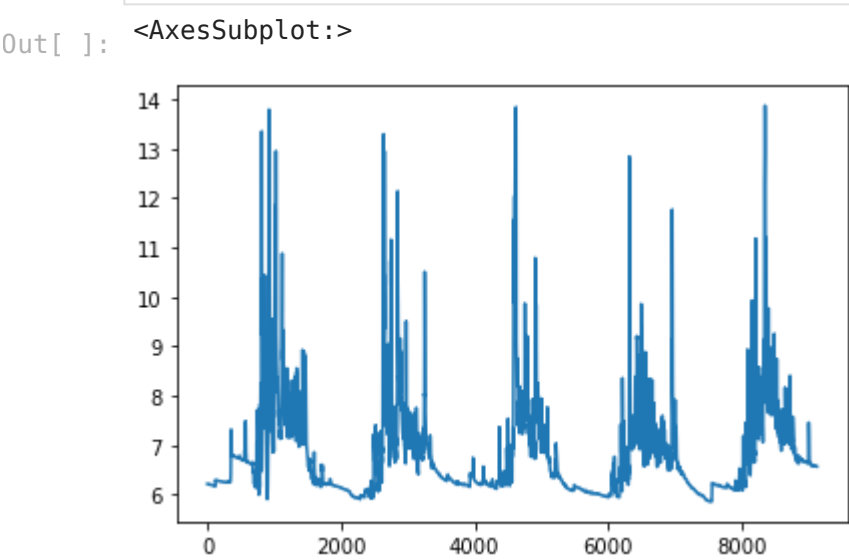
```
In [ ]: waterLevel_df.columns
```

Out[ ]: Index(['SL', 'DISTRICT', 'UPAZILA', 'RIVER', 'STATION ID', 'STATION NAME',  
 'STATION TYPE', 'DATA TYPE', 'DATE/TIME', 'WATER LEVEL(m)', 'LATITUDE',  
 'LONGITUDE'],  
 dtype='object')

```
In [ ]: waterLevel_df.dtypes
```

Out[ ]: SL int64  
DISTRICT object  
UPAZILA object  
RIVER object  
STATION ID object  
STATION NAME object  
STATION TYPE object  
DATA TYPE object  
DATE/TIME object  
WATER LEVEL(m) float64  
LATITUDE float64  
LONGITUDE float64  
dtype: object

```
In [ ]: waterLevel_df['WATER LEVEL(m)'].plot()
```



```
In [ ]: waterLevel_df.tail()
```

Out[ ]:

SL	DISTRICT	UPAZILA	RIVER	STATION ID	STATION NAME	STATION TYPE	DATA TYPE	DATE/TIME	WATER LEVEL(m)
----	----------	---------	-------	------------	--------------	--------------	-----------	-----------	----------------

	SL	DISTRICT	UPAZILA	RIVER	STATION ID	STATION NAME	STATION TYPE	DATA TYPE	DATE/TIME	WATER LEVEL
9125	9126	Bandarban	Lama	Matamuhuri	SW203	Lama	Non Tidal	Regular	31 December, 2021 06:00:am	
9126	9127	Bandarban	Lama	Matamuhuri	SW203	Lama	Non Tidal	Regular	31 December, 2021 09:00:am	
9127	9128	Bandarban	Lama	Matamuhuri	SW203	Lama	Non Tidal	Regular	31 December, 2021 12:00:pm	
9128	9129	Bandarban	Lama	Matamuhuri	SW203	Lama	Non Tidal	Regular	31 December, 2021 03:00:pm	
9129	9130	Bandarban	Lama	Matamuhuri	SW203	Lama	Non Tidal	Regular	31 December, 2021 06:00:pm	



```
In [ ]: (waterLevel_df['WATER LEVEL(m)'] > 12).value_counts()
```

```
Out[ ]: False    9071
        True     59
        Name: WATER LEVEL(m), dtype: int64
```

```
In [ ]: WATER_LAVEL = waterLevel_df.groupby(np.arange(len(waterLevel_df['WATER LEVEL(m)'],
['WATER LEVEL(m)']
```

```
In [ ]: WATER_LAVEL.shape
```

```
Out[ ]: (1826,)
```

```
In [ ]: WATER_LAVEL.head()
```

```
Out[ ]: 0    6.22
        1    6.22
        2    6.22
        3    6.21
        4    6.21
        Name: WATER LEVEL(m), dtype: float64
```

```
In [ ]: WATER_LAVEL.value_counts()
```

```
Out[ ]: 6.210    39
        6.200    35
        6.160    28
        6.190    28
        6.220    25
```

```

6.520      1
7.628      1
6.362      1
6.432      1
10.010     1
Name: WATER_LEVEL(m), Length: 788, dtype: int64

```

```
In [ ]: rainfall_df.head()
```

```
Out[ ]:
```

	SL	DISTRICT	UPAZILA	STATION ID	STATION NAME	DATE	RAIN FALL(mm)	LATITUDE	LONGITUDE
0	1	Bandarban	Lama	CL317	Lama	01-jan-2017	0.0	21.81	92.19
1	2	Bandarban	Lama	CL317	Lama	02-jan-2017	0.0	21.81	92.19
2	3	Bandarban	Lama	CL317	Lama	03-jan-2017	0.0	21.81	92.19
3	4	Bandarban	Lama	CL317	Lama	04-jan-2017	0.0	21.81	92.19
4	5	Bandarban	Lama	CL317	Lama	05-jan-2017	0.0	21.81	92.19

Lets create a main dataframe now

```
In [ ]: df = rainfall_df.copy()
```

```
In [ ]: df.shape , WATER_LAVEL.shape
```

```
Out[ ]: ((1826, 9), (1826,))
```

Lets add our water level per day series to our newly created dataframe

```
In [ ]: df = pd.concat([df,WATER_LAVEL], axis=1)
```

```
In [ ]: df.shape
```

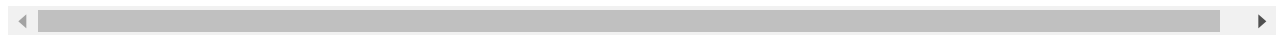
```
Out[ ]: (1826, 10)
```

```
In [ ]: df.head()
```

```
Out[ ]:
```

	SL	DISTRICT	UPAZILA	STATION ID	STATION NAME	DATE	RAIN FALL(mm)	LATITUDE	LONGITUDE	WATE LEVEL(m)
0	1	Bandarban	Lama	CL317	Lama	01-jan-2017	0.0	21.81	92.19	6.2

	SL	DISTRICT	UPAZILA	STATION ID	STATION NAME	DATE	RAIN FALL(mm)	LATITUDE	LONGITUDE	WATE LEVEL(m)
1	2	Bandarban	Lama	CL317	Lama	02-jan-2017	0.0	21.81	92.19	6.2
2	3	Bandarban	Lama	CL317	Lama	03-jan-2017	0.0	21.81	92.19	6.2
3	4	Bandarban	Lama	CL317	Lama	04-jan-2017	0.0	21.81	92.19	6.2
4	5	Bandarban	Lama	CL317	Lama	05-jan-2017	0.0	21.81	92.19	6.2



Lets rename all the columns

In [ ]:

```
df.columns
```

Out[ ]:

```
Index(['SL', 'DISTRICT', 'UPAZILA', 'STATION ID', 'STATION NAME', 'DATE',  
      'RAIN FALL(mm)', 'LATITUDE', 'LONGITUDE', 'WATER LEVEL(m)'],  
      dtype='object')
```

In [ ]:

```
df.columns = df.columns.str.replace(' ', '_')
```

In [ ]:

```
df.columns
```

Out[ ]:

```
Index(['SL', 'DISTRICT', 'UPAZILA', 'STATION_ID', 'STATION_NAME', 'DATE',  
      'RAIN_FALL(mm)', 'LATITUDE', 'LONGITUDE', 'WATER_LEVEL(m)'],  
      dtype='object')
```

As we are using default range indexing we dont need SL columns

In [ ]:

```
df.drop(['SL'],axis=1, inplace=True)
```

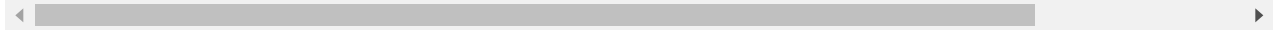
In [ ]:

```
df.head()
```

Out[ ]:

	DISTRICT	UPAZILA	STATION_ID	STATION_NAME	DATE	RAIN_FALL(mm)	LATITUDE	LONGITUDE
0	Bandarban	Lama	CL317	Lama	01-jan-2017	0.0	21.81	92.1
1	Bandarban	Lama	CL317	Lama	02-jan-2017	0.0	21.81	92.1
2	Bandarban	Lama	CL317	Lama	03-jan-2017	0.0	21.81	92.1
3	Bandarban	Lama	CL317	Lama	04-jan-2017	0.0	21.81	92.1

	DISTRICT	UPAZILA	STATION_ID	STATION_NAME	DATE	RAIN_FALL(mm)	LATITUDE	LONGITUDE
4	Bandarban	Lama	CL317	Lama	05-jan-2017	0.0	21.81	92.1



```
In [ ]: df.shape
```

```
Out[ ]: (1826, 9)
```

```
In [ ]: df.columns
```

```
Out[ ]: Index(['DISTRICT', 'UPAZILA', 'STATION_ID', 'STATION_NAME', 'DATE',
              'RAIN_FALL(mm)', 'LATITUDE', 'LONGITUDE', 'WATER_LEVEL(m)'],
              dtype='object')
```

```
In [ ]: df.index
```

```
Out[ ]: RangeIndex(start=0, stop=1826, step=1)
```

Now lets finally create a CSV

```
In [ ]: df.to_csv('./Data/data.csv')
```

```
In [ ]:
```