The main preprocessing should be done to make the distribution of the dataset non-skewed!
Preprocessing steps:

- log-transform the target feature
- box cox transform the input feature

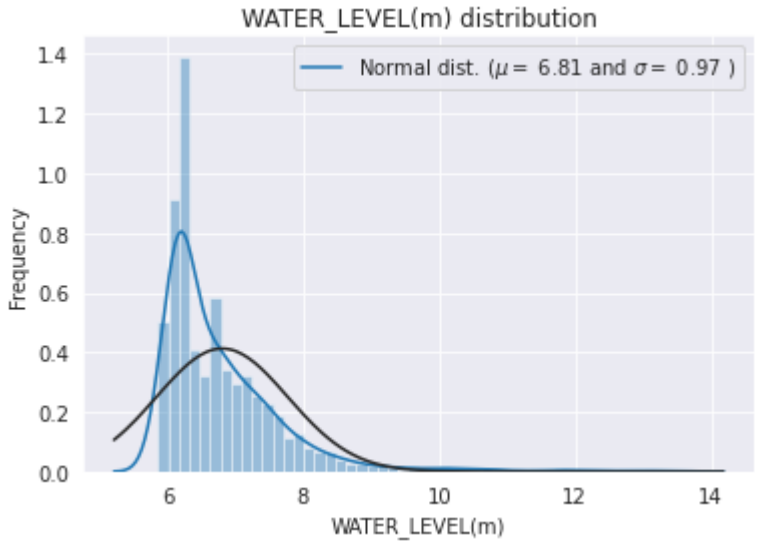
In [ ]:
```
# importing some dependencies
import numpy as np
import pandas as pd
%matplotlib inline
import matplotlib.pyplot as plt
import seaborn as sns
color = sns.color_palette()
sns.set_style('darkgrid')

import warnings
warnings.filterwarnings('ignore')


from scipy import stats
from scipy.stats import norm, skew
```

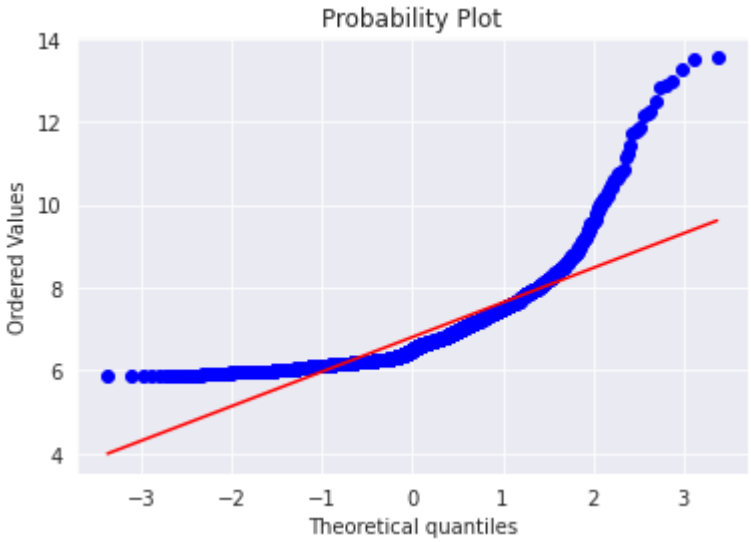
In [ ]:
```
# importing the dataset
df = pd.read_csv("./Data/data.csv",sep=",")
df.drop(['Unnamed: 0'], axis=1, inplace=True) # There were some formatting issue
                                              # writing the csv
```

Lets check the distribution of the target variable **WATER_LEVEL** more closely

In [ ]:
```
sns.distplot(df['WATER_LEVEL(m)'] , fit=norm);

# Get the fitted parameters used by the function
(mu, sigma) = norm.fit(df['WATER_LEVEL(m)'])
print( '\n mu = {:.2f} and sigma = {:.2f}\n'.format(mu, sigma))

#Now plot the distribution
plt.legend(['Normal dist. ($\mu=$ {:.2f} and $\sigma=$ {:.2f} )'.format(mu, sigm
            loc='best')
plt.ylabel('Frequency')
plt.title('WATER_LEVEL(m) distribution')

#Get also the QQ-plot
fig = plt.figure()
res = stats.probplot(df['WATER_LEVEL(m)'], plot=plt)
plt.show()
```

```
mu = 6.81 and sigma = 0.97
```



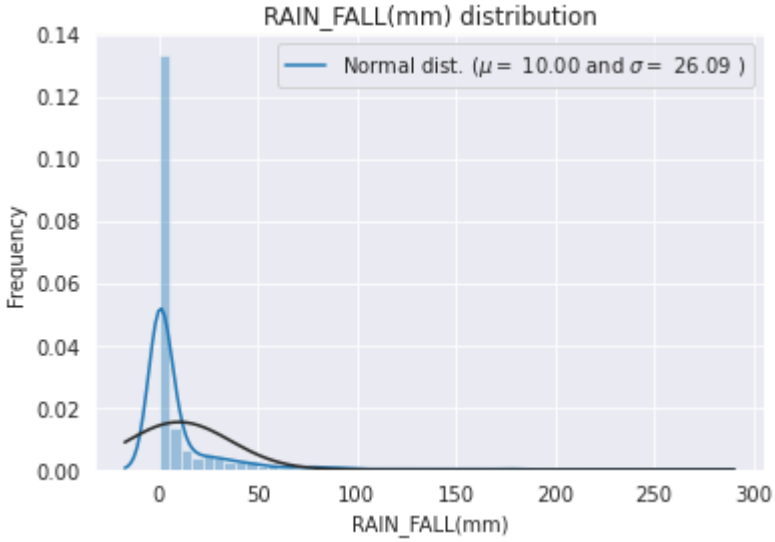


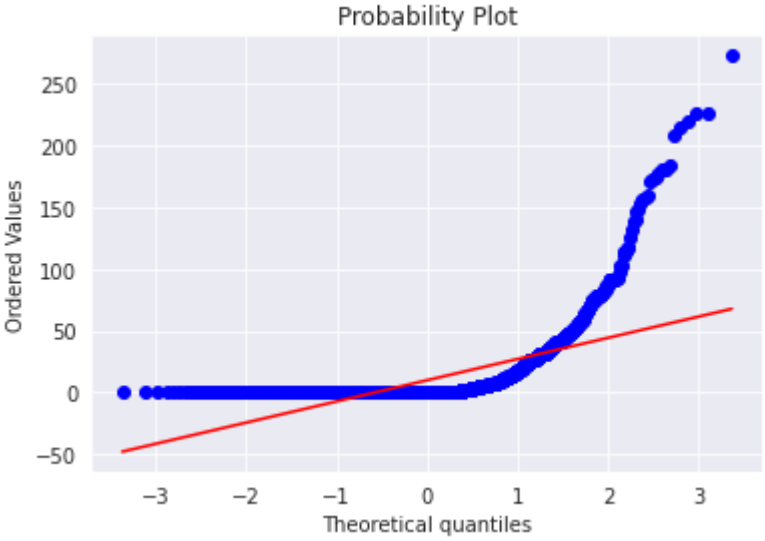
Lets check the distribution of the target variable **RAINFALL_LEVEL** more closely

In [ ]:

```
sns.distplot(df['RAIN_FALL(mm)'] , fit=norm);

# Get the fitted parameters used by the function
(mu, sigma) = norm.fit(df['RAIN_FALL(mm)'])
print( '\n mu = {:.2f} and sigma = {:.2f}\n'.format(mu, sigma))

#Now plot the distribution
plt.legend(['Normal dist. ($\mu=$ {:.2f} and $\sigma=$ {:.2f} )'.format(mu, sigm
            loc='best')
plt.ylabel('Frequency')
plt.title('RAIN_FALL(mm) distribution')

#Get also the QQ-plot
fig = plt.figure()
res = stats.probplot(df['RAIN_FALL(mm)'], plot=plt)
plt.show()
```

```
mu = 10.00 and sigma = 26.09
```


RAIN_FALL(mm) distribution
Normal dist. (μ = 10.00 and σ = 26.09)
Frequency
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0.00
0
50
100
150
200
250
300
RAIN_FALL(mm)


Probability Plot
Ordered Values
250
200
150
100
50
0
−50
−3
−2
−1
0
1
2
3
Theoretical quantiles