

## **LAB-12**

- Imported the dataset that is in this folder
- Print Summary of the dataset
- Keep numerical and categorical variable names in two different list.
- Plot box plot for numerical columns
- Use Standard Deviation outlier detection method to replace the outliers for the column 'temp' with NULL. (consider a data point is outlier if it is two standard deviations away from the mean.)
- Count the number of outliers corresponding to the column 'temp'
- Replace those outliers with mean value of the corresponding attribute
- Identify the outliers of the column 'hum' using Z-Score (consider the threshold 3)
- Replace those outliers with mean of "hum"
- Using IQR of the column "windspeed" replace the data points that lie outside of the lower and the upper bound with a NULL value in the original dataframe
- Check for null values in the updated dataframe
- Drop the rows for which the windspeed value is missing
- Finally print the description and summary of the cleaned dataset that will be processed for datamining task.