**Problem Statement**

1. The attached ‘**case study.csv’** data is wind turbine data with below columns.

|  |  |
| --- | --- |
| **Columns** | **Description** |
| WTGID | Wind turbine Machine ID |
| Loc | Location of the Wind turbine Machine |
| MonthStartDate | month aggregate start date date |
| f1 | feature 1 |
| f2 | feature 2 |
| f3 | feature 3 |
| f4 | feature 4 |
| f5 | feature 5 |
| f6 | feature 6 |
| target | Target to predict based on independent variables |
| dataset | Data for Training(model) and Validation(val) |

1. Perform exploratory data analysis.
2. Predict ‘target’ variable using independent variables only using ‘model data (from dataset column). You can choose any time series model or classification model to predict target variable.
3. Explain why you choose the model over other models with details and also explain which features are important in classifying the ‘target’.
4. Apply the model on ‘val’ data in the dataset and estimate how accurately model is able to classify the target classes on validation dataset.
5. Suggestion to improve the accuracy of the model on validation dataset.
6. Explain all the above details in nice power point presentation and also share code file(.ipynb or .py file)

Note: All the details need to be explained in detail in power point presentation and also expectation is to communicate results effectively during F2F discussion.