**Steps for cleaning the Dataset**

1. Open the csv file of [Mouth of Placentia Bay Buoy](https://drive.google.com/file/d/1rinZ5XgK_f64NxtV-Vy0ZWCEp-oWajxJ/view?usp=sharing) and [Pilot Boarding Station / Red Island Shoal Buoy](https://drive.google.com/file/d/1WgaF7Q_jej-YlrAJEfuGnqhJ5l3xAXbJ/view?usp=sharing).
2. Check the data types.
3. Convert the timestamp into datetime format.
4. Match the columns of both the files and remove the extra columns.
5. Check for Null values.
6. If there are Null values, Remove those.
7. See the summary of the dataset.
8. Make sure that all the column are showing the same number of records.
9. Save the clean files.

**Steps for merging two files**

Merging will be done based on “Date & Time” parameter. In these cleaned files, the timestamp does not match, so grouping is required.

1. Group the dataset for 3hour frequency.

df.groupby (pd.Grouper(key='DATE & TIME', freq='3H')).mean()

## You can try 1H or 2H frequency and see the results where you get high accuracy

1. Check if there are Null values and if yes, drop those.
2. And merge both the files.

data = pd.concat([file1,file2], axis=1, join="inner")

1. See the co-relation between the variables.
   1. Create Heat map or co-relation matrix
2. Remove the non-co-related features.

**Steps for training the model**

1. In the given notebook, we are trying to predict the wave\_ht\_max of Red Shoal Buoy.
2. Since the target variable is continuous, so it is Regression Problem.
3. Split the dataset into training and testing.
4. We have applied support vector regressor and randon forest regressor. You can try other regressor model to see the performance.
5. Try some different approach to achieve good results.

**Note:** Above mentioned approach is our approach to clean and merge the files. You can try your own approach for cleaning and merging files to get better accuracy. We did it for two buoy files. you can add the data of land station also and see the results.

**Mouth of Placentia**

<https://dataplatform.cloud.ibm.com/analytics/notebooks/v2/1b0cfa7a-24a4-4836-a7e4-2517e2d637a0/view?access_token=ef5e2abde59e101aa8f86ee8b5939b69c0e7beb93b8c624e576036c69894b60b>

https://tinyurl.com/mouthofplacentia

**Red Shoal**

<https://dataplatform.cloud.ibm.com/analytics/notebooks/v2/02a713f6-aad4-4eff-bbd9-7e10b72b425d/view?access_token=46af5e05fd74c6e54d0e1abd8efacd3a62db52106500b8c93ac54e038162505e>

https://tinyurl.com/redshoal

**Merging and Prediction**

<https://dataplatform.cloud.ibm.com/analytics/notebooks/v2/0d2de5f9-ebde-4e67-adc0-569f9db8f3a2/view?access_token=845b2f4310ba0e5520fb8fbdc962ecf094f91295a976ce554b196320052017b5>

https://tinyurl.com/mergeandprediction