**GitHub Link**.

<https://github.com/HasiniRangana/DrivenData-PumpItUpChallenge.git>

**Data Preprocessing Techniques.**

Under the data preprocessing done these things.

1. Data cleaning.

Missing values were handled under this category. First identify the missing values correctly by using various techniques. And then handle missing values correctly. Missing values replacing with mean value is used to numerical data columns.

1. Data transformation.

Normalization and categorical encoding done under the data transformation. Min max normalization used to normalize the dataset. Not only that label encoding and one hot encoding used as categorical encoding techniques.

1. Data reduction.

Under the data reduction reducing the number of features. There were columns with same data and removed those data columns.

**Feature Engineering Techniques.**

Under the feature engineering done these things.

1. Feature selection.

* Select most relevant features and discard the rest.
* By using this limited the number of features and prevent overfitting.

1. Feature extraction.

* Build new columns from raw data by reformatting, combining and transforming primary features into new ones.
* In here there were some spelling mistakes in data and I corrected those errors and combining that data types.

Under the feature engineering done creation of new features by using mathematical transformation. And also done target encoding.

