

## Initial Phase →

\* Initially, explored Kaggle to find dataset after gaining some insights about how I'd prepare such model and what would it contribute.

\* Picked up a dataset from Kaggle. Explored the number of rows and column in the csv file. I was looking for a column which would give me some info about whether skipped or not or such related information.

Got another dataset satisfying the demands.

## Data Exploration →

\* Now, we explore the dataset and check if the dataset is not biased.

\* Plotted data on a graph to get a better idea of the data.

## Defining Features & Target →

• Initialized two variables, one storing the target column, which is 'skip-30s' in this case, and the other storing all other columns which we refer to as "Feature".

## Standardization →

• As we have varied scale data, we try to scale down the data using `StandardScaler()` STL which standardizes using  $Z$ -Score

Normalisation.

Splitting Data →

We split data into two parts, one to train our model and the other part to test how good our model is doing.

I have split the data in a 80:20 ratio.

Last Steps →

- We train the model using Logistic Regression's principles.
- We predict an output from the model after training it and giving it testing sample features.
- We compare the predicted output by the actual output using various techniques.