# **Predicting Billboard Hits Using Spotify Data**

Dataset used: songs\_complete\_data.csv

Pre-processing Data: checking for null values and replacing, deleting the column (or) attribute which has more null values (Lyrics), Removing the duplicate rows.

Exploratory Data Analysis: Plotting the graphs for features vs time and feature vs feature. Used line plots for feature vs time, used scatterplots and heatmap for correlation of features in feature vs feature.

After dropping certain features which did not contribute much to the Target attribute (Top100), the features remained are:

* Danceability
* Liveness
* Instrumentalness
* Speechiness
* Acousticness
* Loudness
* Valence
* Tempo
* Energy
* Genre

After Pre-processing the final dataset is converted to csv file (final\_dataset.csv) and will be using this for modelling.

In the modelling, first the dataset will be divided into training and testing parts, then using the machine learning algorithm model gets build and gets trained with training dataset.

Finally with the help of testing he model gets evaluated and the accuracy of model can be calculated.