

HAITIAN TUBES POPULARITY

Analysis and Prediction





BACKGROUND ...

**Music is
integral
part of
our life**

**Important
Part of
the
country
economy**

**Many
concerned
stakeholder**

Concerned Stakeholder



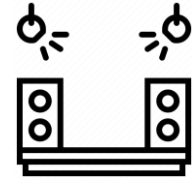
Venues Owner



Music
distributors



Artists



Producers

Hypothesis and desired Outcomes

Understand which feature of a song are important to predict the song popularity

Using song features, build a model which can predict if a given song will be above or below the mean of song popularity scores of the data



Data



Youtube Data Api
90 Channels
exported



Instagram
Api



Data processing

- Download the Api data in many json file and then convert them in dataframe stored in CSV file
- Extract all the statistics data regarding each song on youtube, using python, such as, Duration, publication age, like, dislikes, views and comments count.
- Download related main and featuring artist features from Instagram data API, such as, number of followers



Data exploration

- Traditional Data Cleansing, such as, remove NaNs, or song missing statistics features or song with a duration bigger than 5mn 30 or lesser than 3 mn
- Change duration column format from string to seconds
- Separate name from title, and artist from artist features.



Data processing

- After data cleansing, left with
Rows and 8 important features for
the case study
- Evaluate features relevance, using
correlation matrix



Data Exploration

Data Trends over the year.- fig1



Data Exploration

Numbers of song over the
years.- fig 2



Data Exploration

**Most popular song sorted by
contents_Views and year.- fig 3**



Data Exploration

Statistics



Data Exploration

Features importance.-

Using Random forest



Result

Model evaluation



Conclusion

Model evaluation