

PREDICT  
DANCEABILITY  
BASED ON LYRICS

# SONG ANALYTICS

## WHY THIS PROJECT



MUSIC LOVER



WANT TO UNDERSTAND THE  
MUSIC INDUSTRY BETTER

# DATA

- Billboard top 100 from 1960 to 2016
  - Songs information such as length, rank and danceability
  - Lyrics
  - Over 5700 observations

# MODEL

- CV of Random Forest

R square:  
0.5209

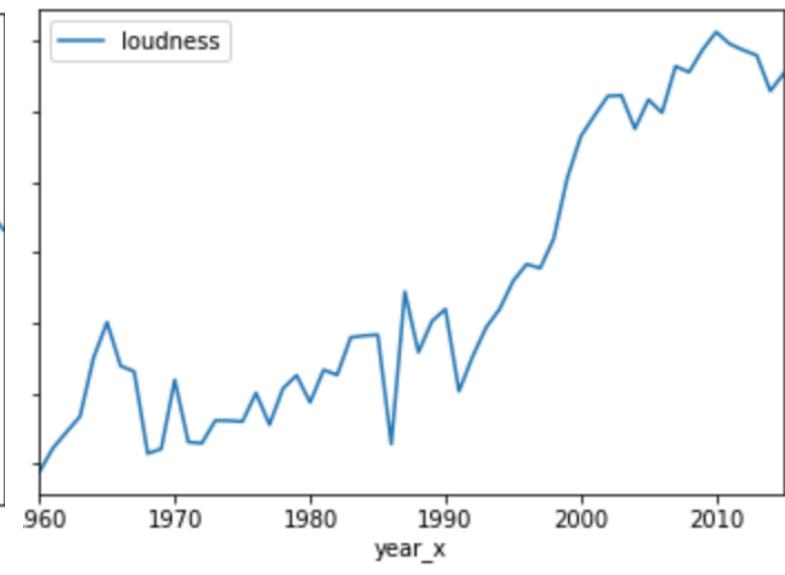
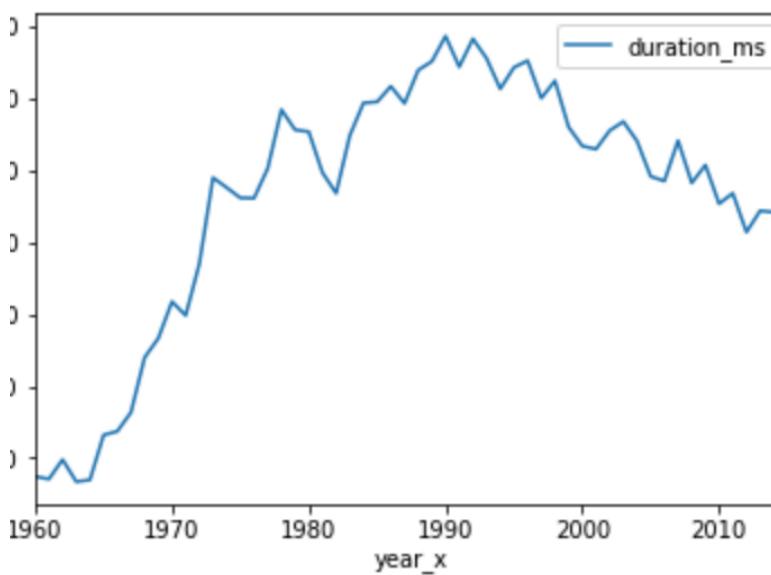
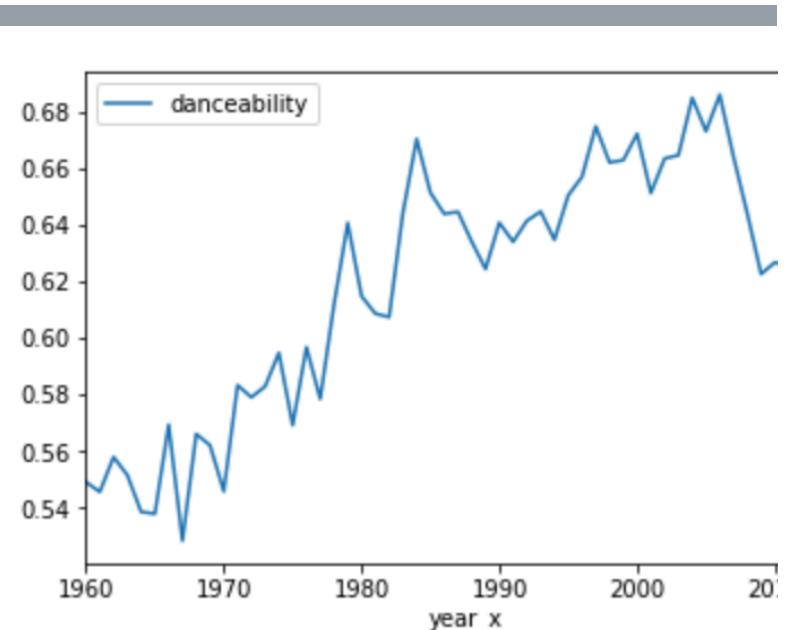
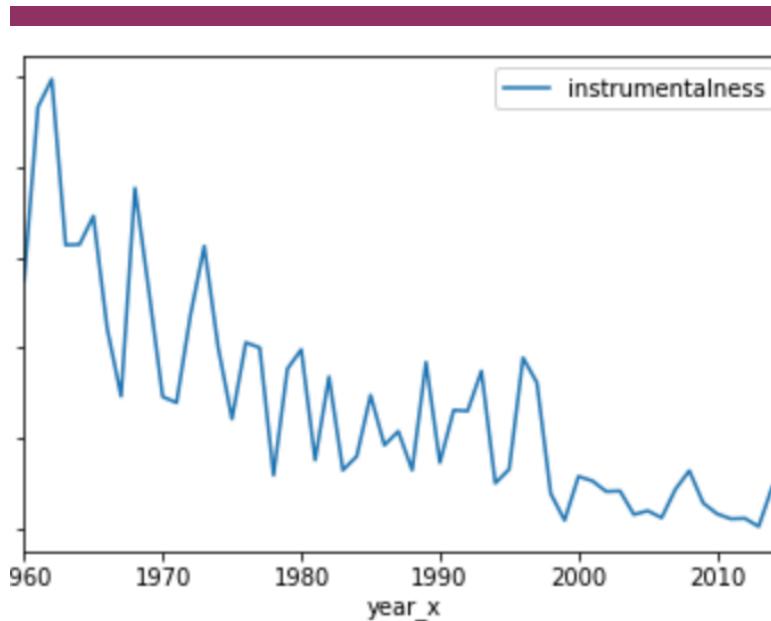
MSE: 0.0173

Test R  
square:  
0.2096

- Potential overfitting because of the tf-idf transformation of lyrics

# INSIGHT FROM DATA

- Danceable is the trend
- Duration of the songs peak during late 90s
- Songs are getting louder
- Song writers are getting rid of instruments





**THANKS &  
QUESTIONS**