Predicting Billboard Performance

Youssef Lazrak
Owen Liu
Ilias Miraoui
Sophie Pealat
Jacques-Olivier Weulassagou

Music in our Lives

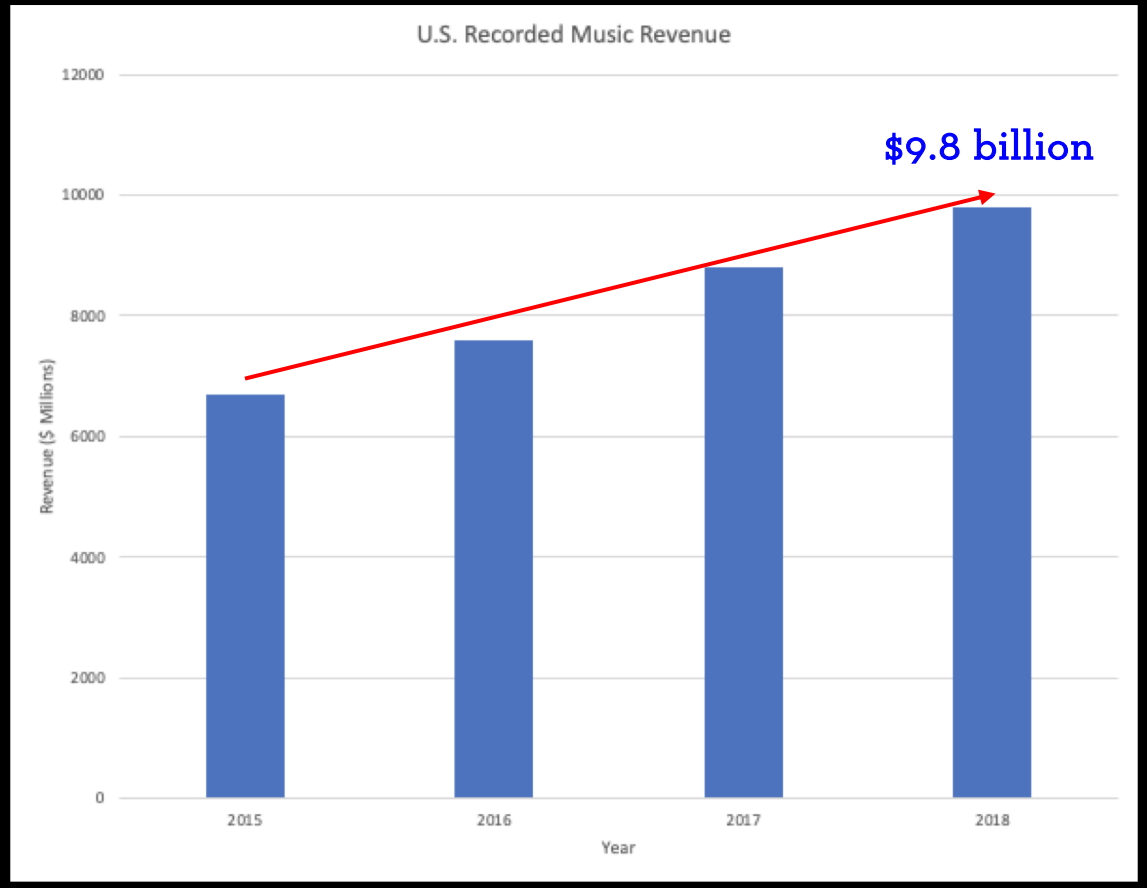
According to the IFPI 2019 Music Report:

- On average, people are spending around 18 hours a week listening to music
- 54% of respondents say they love or are fanatical about music

Potential Impacts

Predicting Music's popularity could help:

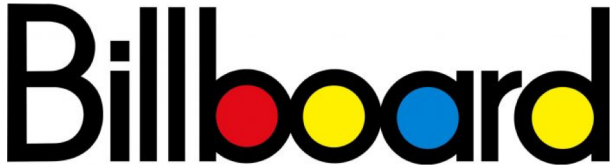
- Influence Music Creation
- Improve Record Companies' Portfolio Management
- Optimize Advertising Spend
- Influence Festival Organization



Sources of Data

- Billboard : Billboard charts from 1998 to 2019 (100 songs/chart)
- Spotify API : internal characteristics of songs
- RIAA website : song certifications history
- Google Trends : popularity of the artist

28 variables : 8 categorical, 20 continuous

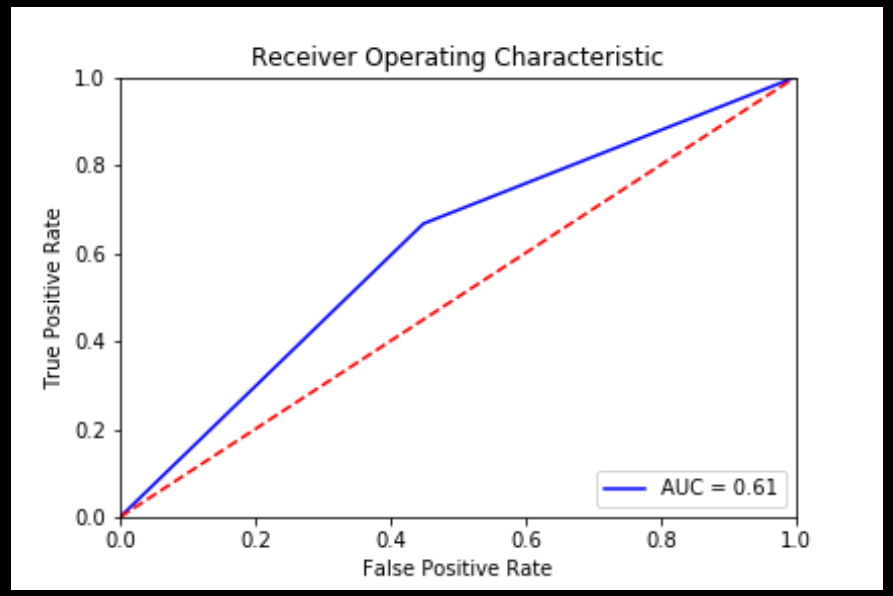


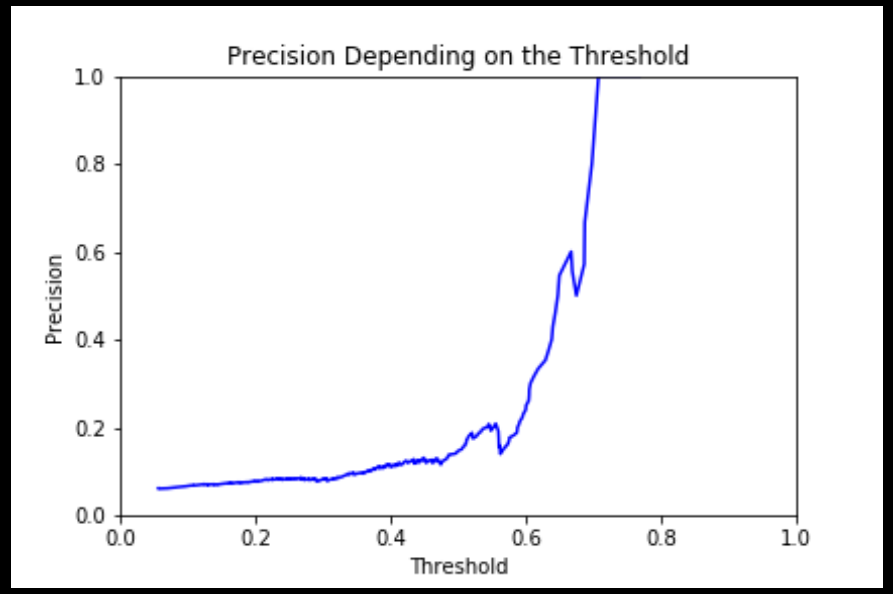
Logistic Regression

- Split the Data 80/20 (6287/1602)
- Predicting 1 if the song reaches the top 10, 0 otherwise
- Despite attempting undersampling and oversampling, results are still mitigated in our attempt to maximize precision:
  - With a threshold of p=0.7 for test set:

    [1506, 0]

    [92 , 4]
- Potential solutions:
  - Attempt more complex models
  - Attempt PCA to avoid overfitting on more complex methods
  - Try to Predict a larger "Top" (Top 25? Top 50?)





The Way Forward

- Include data from Google Trend : train model + predictions
- Additional Variable selection
- Principal Component Analysis for Complex Models
- Run additional models to improve performance (CART, RF, Boosting, Neural Network)
- Is inference possible?