



Music Hits Predictions

Mehika Patel



Outline

- Context & Background
- Process
- Model Evaluation Metrics
- Final Visualizations
- Future Work

Context

- 100 songs selected on various criteria by billboard as the “top 100” songs of the week each week.
- What is the peak a song can reach on the chart once it has entered?



Why do we care?

- Music industry's standard record chart in the U.S.
- Rankings based on sales (physical, digital, radio play, & streams)
- Investors & Top Record Labels want to invest in songs that will be the best
- Musical competition among top artists



Data

- 2761 Data Entries
- Target Variable: Peak Position on Top 100
- Features:
 - Number of weeks on charts, artist followers on spotify, artist popularity, song danceability, energy, key, loudness, mode, speechiness, acousticness, instrumentalness, liveness, valence, tempo, length of song, time signature

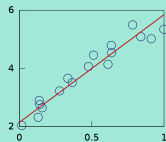


Process Outline

1. Data Gathering



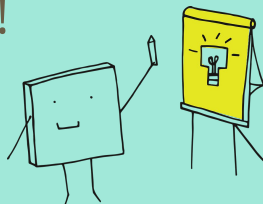
2. Model Building



3. Compare model evaluation metrics with CV



4. Visualize and finalize model!

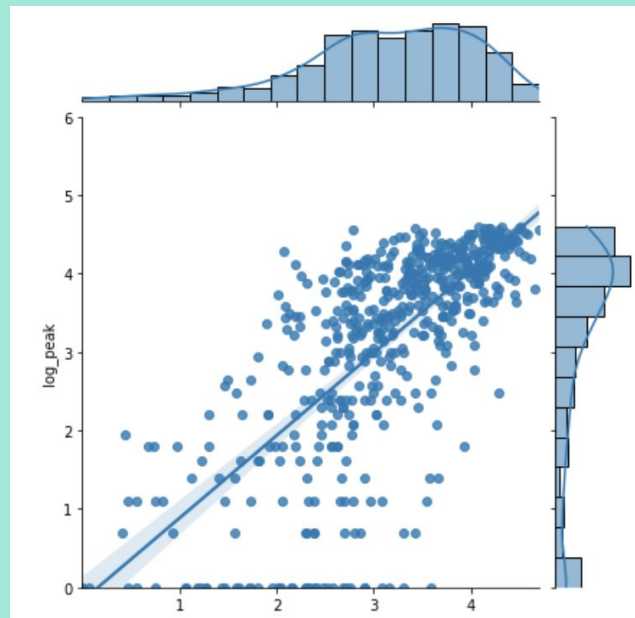


Final Model Evaluation Metrics

- Mean absolute error= .68
- SSE= 482.96
- RMSE= .93
- R2 Value from final test = .50

- Intercept: 4.222894599606599
- Danceability:-9.18010339e-01
- Popularity:2.66326807e-02
- Followers:-4.04140264e-08
- Num_wks:-1.06595401e-01
- Mode:1.71709497e-01
- Popularity Squared:-2.09622088e-04
- Followers Squared:3.23195872e-16
- Num_Wks Squared: 7.87829545e-04

*Unstandardized coefficients



Future Predictions

- Assume the following song will be on the charts one more week. Will they reach a new peak?:
- Film Out -BTS (81)-2
 - Peak 28 in week 2
- Arcade -Duncan Laurence (100)-2
 - Peak 71 in week 2
- 4 Da Gang- 42 Dugg (67)-2
 - Peak 44 in week 2

Future Work

- Different features that are more easily accessible
- Different target variable might be more useful--like play time ratio on spotify etc.
- Look at producer/record label
- Explore more interaction terms