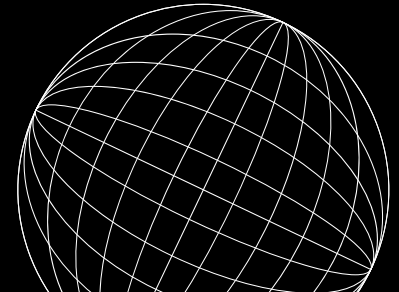2022

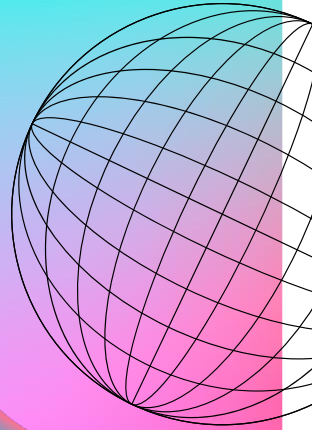# Predicting Song Popularity In Denver

By: Zain Elsell, Avery Fulton, Ayush Uniyal

Presentation Outline:

1   **Project Goals & Applications**

2   **Process & Methodology**

3   **Project Results**

4   **Conclusion & Future Considerations**

1.1 Project Goals

- Identify what quantitative values influence the popularity of a song specifically in Denver, Colorado.

- Build models that attempt to predict ranking of song given information on that particular tracks audio.

Real Applications

- Allow for artist's to break into untapped markets and increase viewership.
- Estimate general trends of music in a demographic
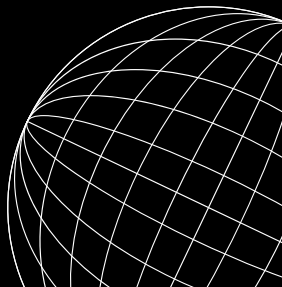- Could be used in conjunction with a song recommendation engine.

2

Process & Methods

2.1 Finding A Data Stream

Spotify City Charts Vs Charts Local Pulse

- Contains the most played songs in a region.
- Tends to showcase more the "average" music taste.

- Contains songs that are uniquely played within a region.
- Top songs that are uniquely popular in Denver, USA.

USA Top

Spotify Charts

USA

Weekly Top Songs USA
Your weekly update of the most played tracks right now.

Weekly ∨   **Week of Apr 21** ∨   **Select City** ∨

#		TRACK		PEAK	PREV	STREAK	STREAMS	
1	↑ 1	**As It Was** Harry Styles		1	2	3	15,105,490	More
2	↓ 1	**First Class** Jack Harlow		1	1	2	14,940,431	
3	NEW	**Don't Think Jesus** Morgan Wallen		3	—	1	6,189,833	
4	↓ 1	**Heat Waves** Glass Animals		1	3	72	6,081,673	

Denver Top Songs

Spotify Charts

USA

Denver
Your weekly update of the most popular tracks in Denver, USA.

Denver ∨   **Week of Apr 21** ∨

#		TRACK		PEAK	PREV	STREAK	
1	–	**First Class** Jack Harlow		1	1	2	
2	–	**As It Was** Harry Styles		1	2	3	
3	–	**Heat Waves** Glass Animals		1	3	27	More ∨
4	↑ 2	**STAY (with Justin Bieber)** The Kid LAROI, Justin Bieber		2	6	27	

Denver LP

Spotify Charts

USA

Local Pulse Denver
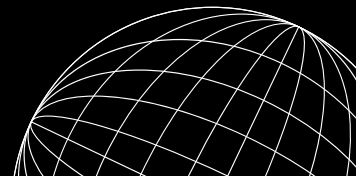Top songs that are uniquely popular in Denver, USA.

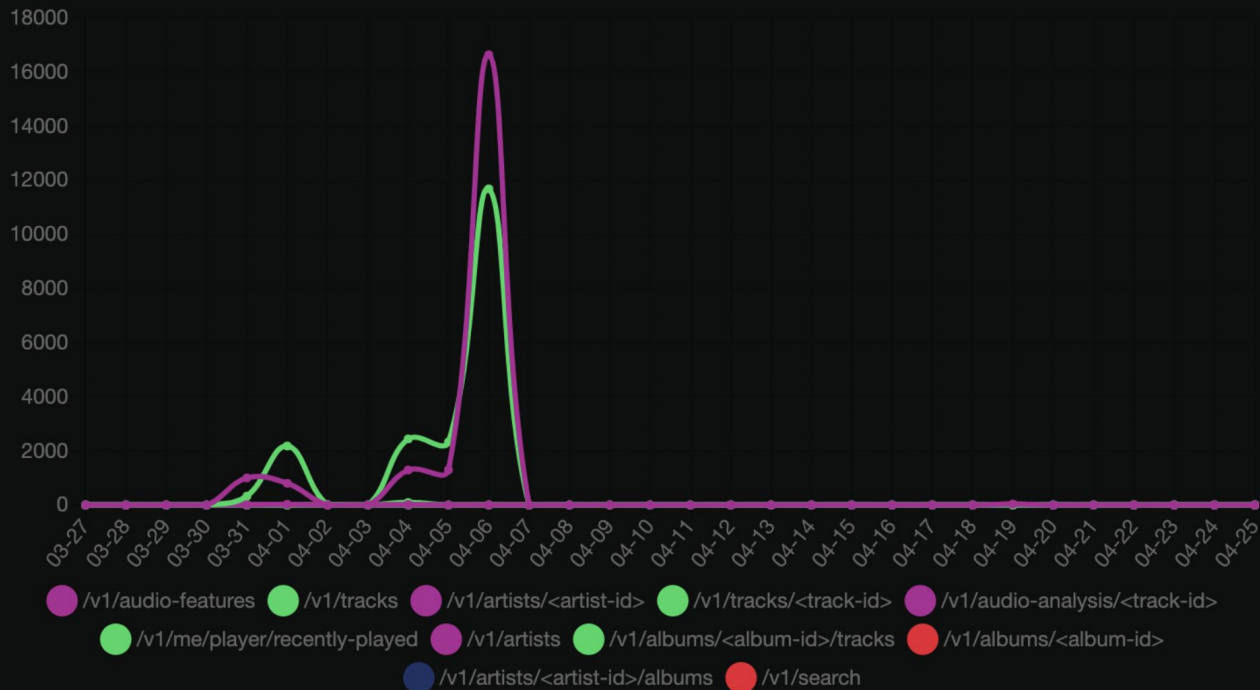**Denver** ∨   **Week of Apr 21** ∨

#		TRACK		PEAK	PREV	STREAK	
1	NEW	**No Patience** Shamir Jamal		1	—	1	More ∨
2	NEW	**Over the Edge** Tucker Nichol		2	—	1	
3	NEW	**The Devil's Opp** Young Roddy		3	—	1	
4	NEW	**Take A Trip** Liquid Stranger, LSDREAM		4	—	1	

Spotify API Requests Over Time

Number of Requests/Endpoint

24

Weeks of data collection

783

Unique Songs

2400

Data Points

2.2 Creating Clean Data Frame

Set start and end date for weeks we want to scrape.

⟹

Extract HTML table class

⟹

Extract song UUID from HTML code

⬇

Generate a data frame with cleaned data.

⟸

Clean data for input into machine learning models.

⟸

Take UUIDS and send to Spotify API
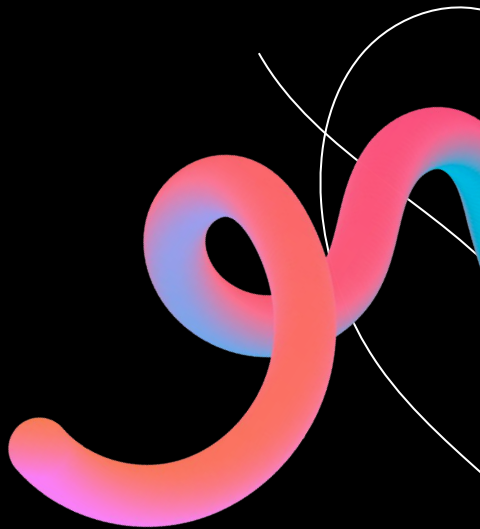
What are we even looking for?

- Correlations to Average Rank Over Weeks Trending
- Or Pure Song Ranking within a tracks "Audio Features"
  - What are the audio Features?
    - Energy
    - Loudness
    - "Instrumentalness"
    - "Acousticness"
    - "Danceability"
    - Liveness
    - Duration
    - Tempo
    - Valence
      - And Many More

2.3 Analyze Data Set & Build Models

How did we decide on model features and target?
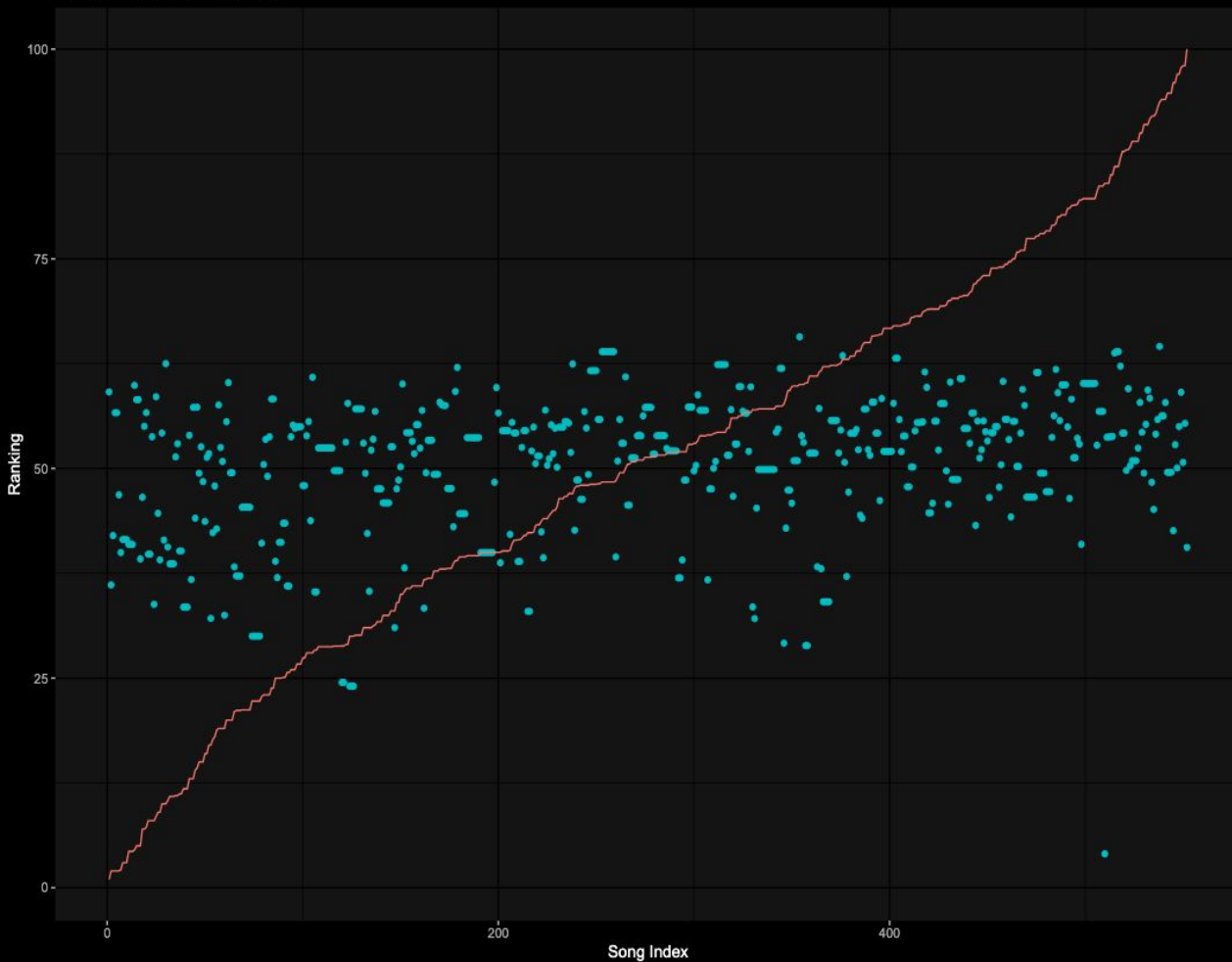
- Using Correlation Matrices
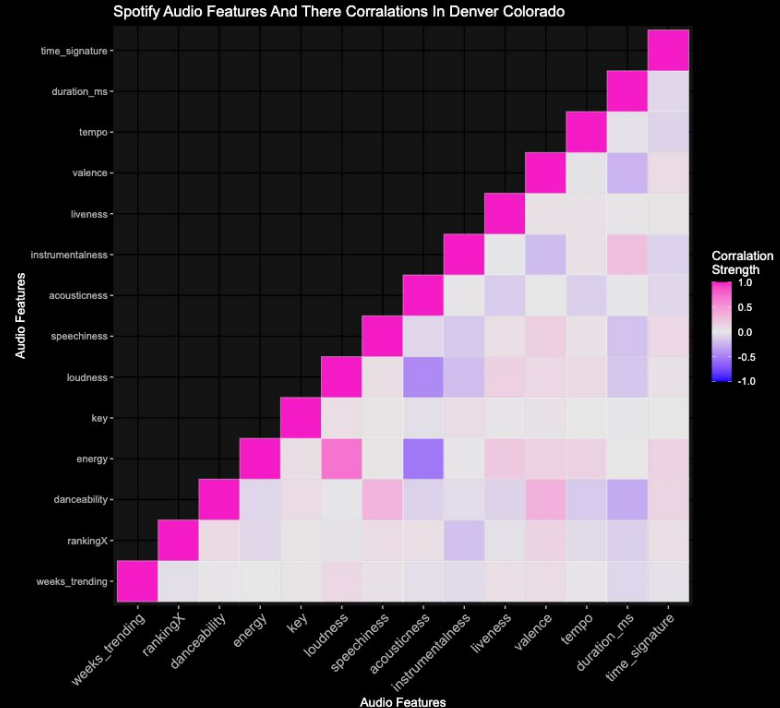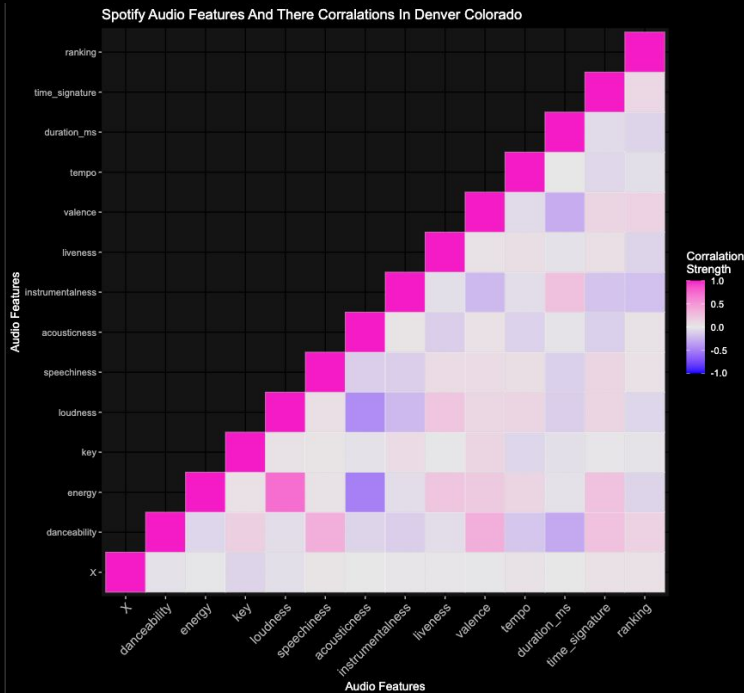
3

Project Results

Elastic Model Performance

Ranking

Song Index

colour
Actual Rank
Predicted Rank

/////////////////////////////////
Mean squared error:
504.5795
Root mean squared
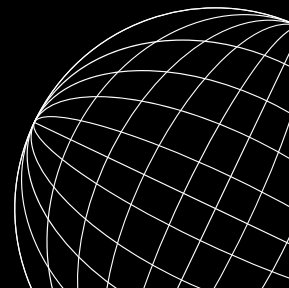error: **22.46285**
Normalized root mean
squared error:
0.2268975
R^2 value of model:
0.09332926
Adjusted R^2 value of
model: **0.07486004**
/////////////////////////////////

How well does **EN** predict a song in the top X tracks?
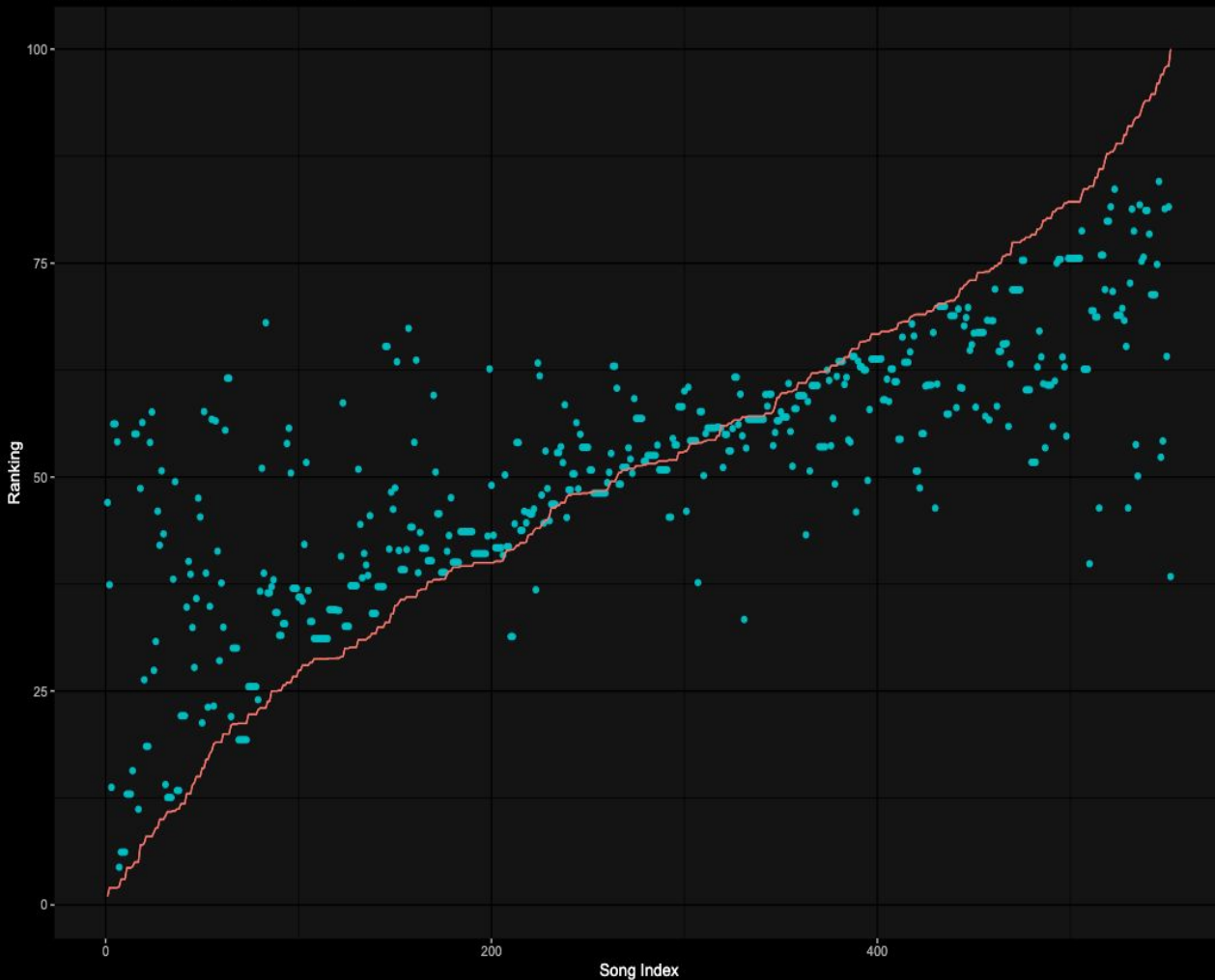
52%

58%

49%

X = 60

X = 50

X = 30

*percentages given as balanced accuracy

Random Forest Model Performance

Random Forest "R²" Visualized
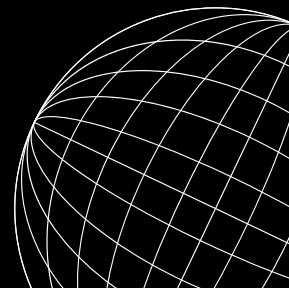
/////////////////////////////////
Mean squared error: 527.9183
Root mean squared error: 22.97647
Normalized root mean squared error: 0.2220856
Variance Explained: 0.8083897
/////////////////////////////////

How well does **RF** predict a song is with in the top X tracks?

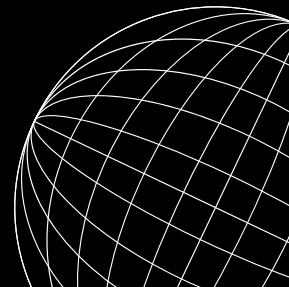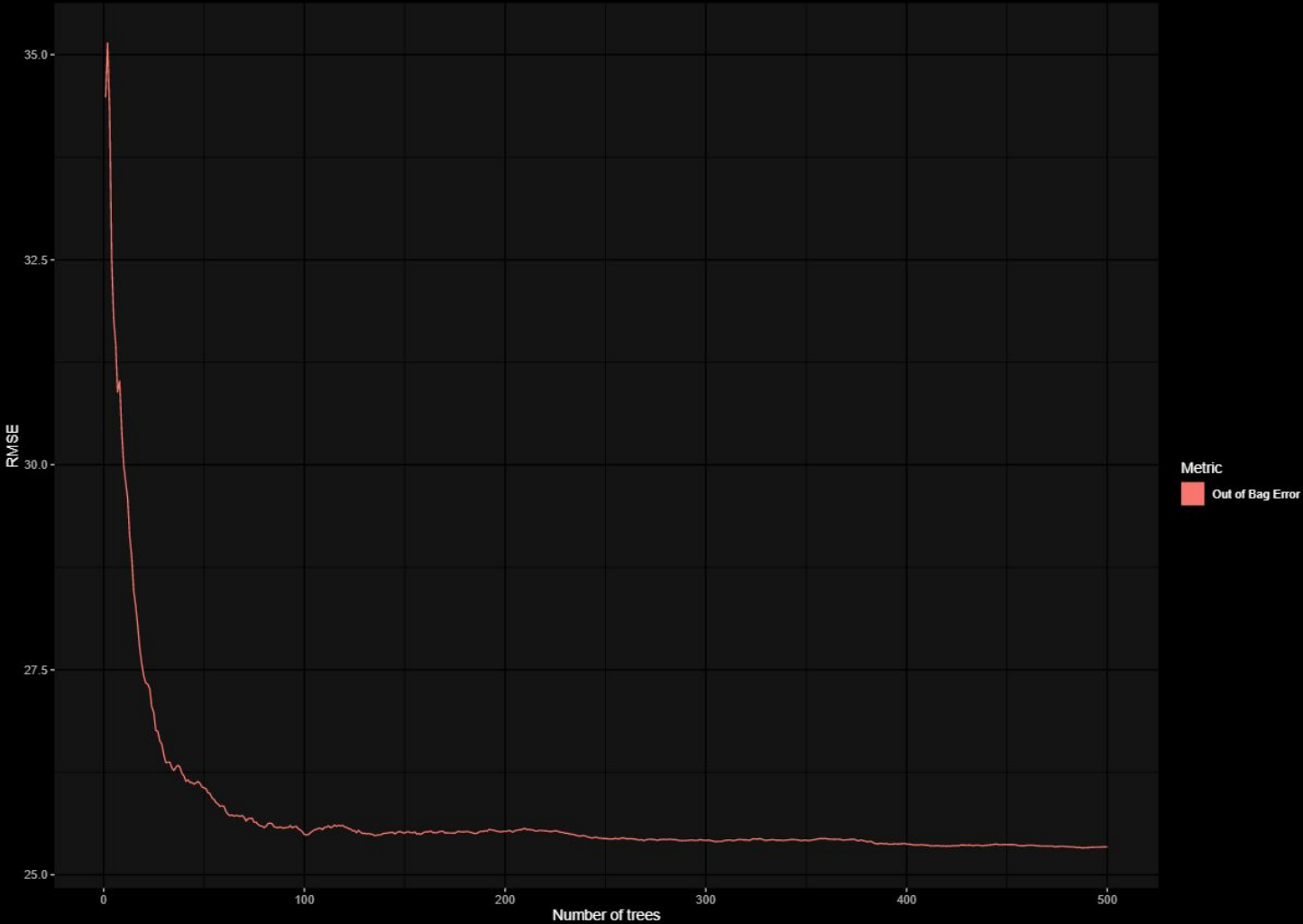| 67%	86%	76%
X = 60	X = 50	X = 30

*percentages given as balanced accuracy

RMSE vs Number Of Trees Random Forest

4

Future Considerations & Conclusion

4 Future Considerations & Conclusions

- Audio feature analysis alone isn't very viable for rank.

- Reducing noise in a dataset is crucial.

- Further analysis could include researching additional model features.
  - Artist platform size
  - Cross referencing with other region specific charts
- Additionally, we didn't account for the variance of features depending on the song's genre.
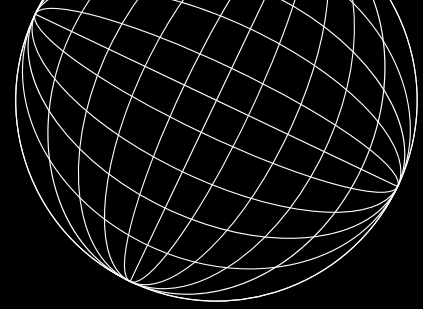
Thank You For Listening, Questions?

References

"Spotify Charts - Spotify Charts Are Made by Fans."
https://charts.spotify.com/charts/view/citypulsetrack-denver-weekly/latest.

"ModelMetrics - Evaluate models Fast".
 https://rdrr.io/cran/ModelMetrics/man/

"Spotify Platform API Documentation"
https://developer.spotify.com/documentation/web-api/

"SpotifyR- Wrapper Class For Spotify API"
https://github.com/charlie86/spotifyr

"Efficient Consulting - SlideGO"
https://slidesgo.com/theme/efficient-consulting-toolkit

"Visualize correlation matrix using correlogram"
http://www.sthda.com/english/wiki/visualize-correlation-matrix-using-correlogram

Random Forests and the Bias-Variance Tradeoff
https://towardsdatascience.com/random-forests-and-the-bias-variance-tradeoff-3b77fee339b4