

Music Prediction

Sara Yun

Business Question

Which components affect larger audience

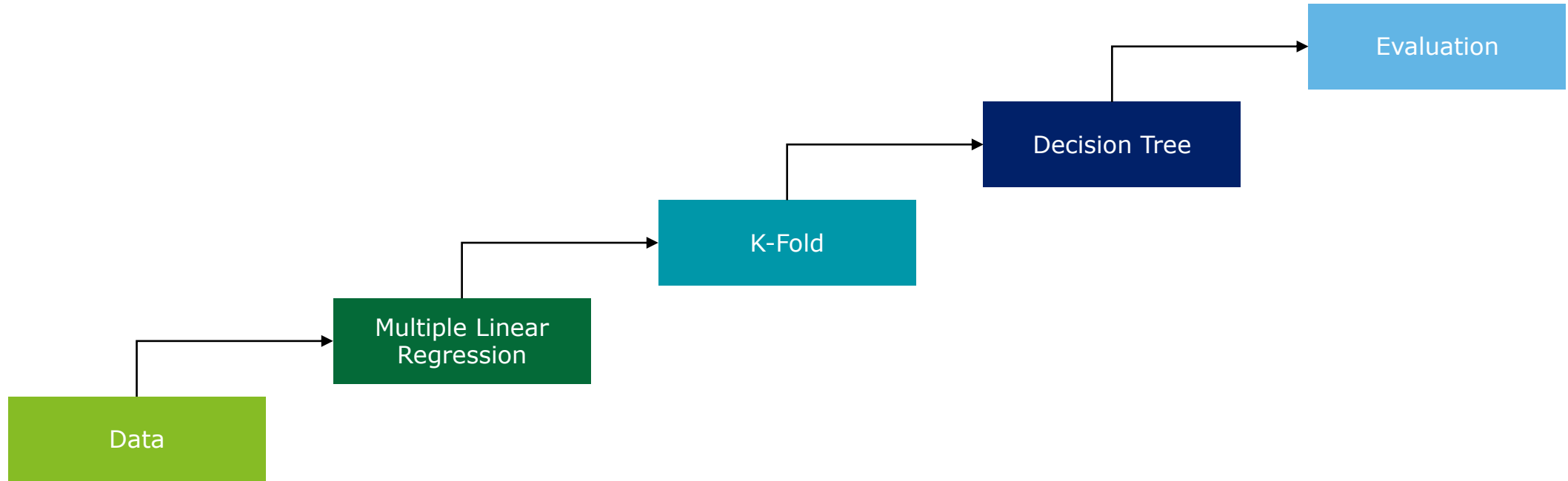
Definition of Hit:

- Billboard top 100

Factors of Music

- Considering factor are danceability, energy, key, loudness, liveness, tempo, duration time, sections, target, genre, acousticness, etc..

Overview



Data



01

Billboard top 100

Basic music list that considered as a hit

The original data were from 1990s to 2019. To increase the reliability of data, used data from 2000 to 2019

Because the data was too big, drop unnecessary information such as uri, writing credits, and lyrics.

<https://www.kaggle.com/datasets/daniel2255/data-on-songs-from-billboard-19992019>

02

Sportify Hit predictor Dataset

Data with extra features and components of the music

The dataset is from 1960 to 2019. To increase the reliability and combine with first data, only use data from 2000 to 2019

Drop unnecessary information and duplicate columns

<https://www.kaggle.com/datasets/theoverman/the-spotify-hit-predictor-dataset?select=dataset-of-00s.csv>

03

Combine two data

Due to the duplicates name of music and the name of artist, combine two data using not only artist name but also the music name

During modelling the data, used weeks on chart as target value

Drop unnecessary duplicate information for data size efficiency

Multiple linear regression

Target =
danceability

2000 to 2009
R square: 0.968

features	coef
valence	0.3375
time signature	0.1213
target	0.1139
instrumentalness	0.0747

2010 to 2019
R square: 0.964

features	coef
valence	0.2106
time signature	0.1583
target	0.0985
instrumentalness	0.0211

Target =
energy

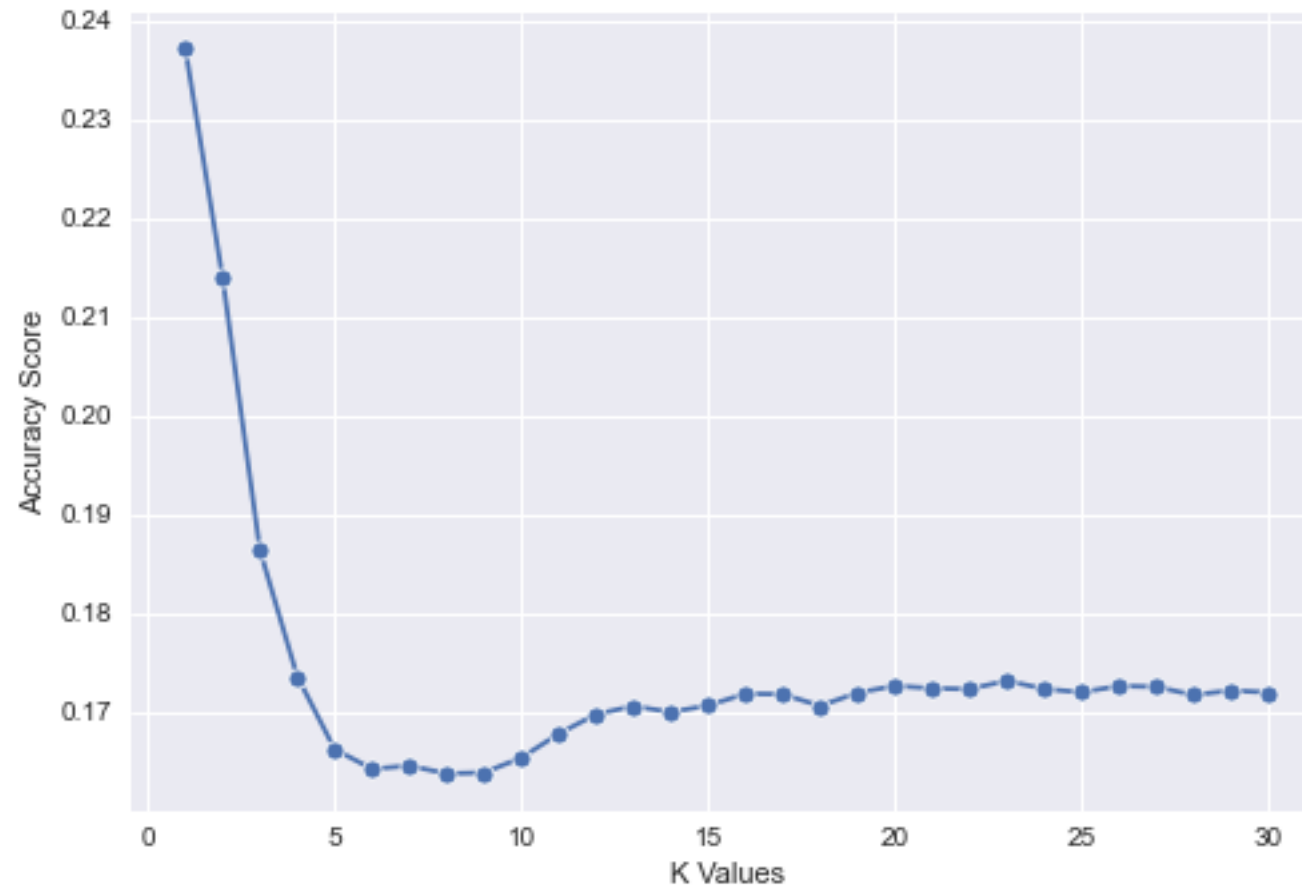
2000 to 2009
R square: 0.977

features	coef
valence	0.2221
time signature	0.1773
target	0.0373
instrumentalness	0.1511

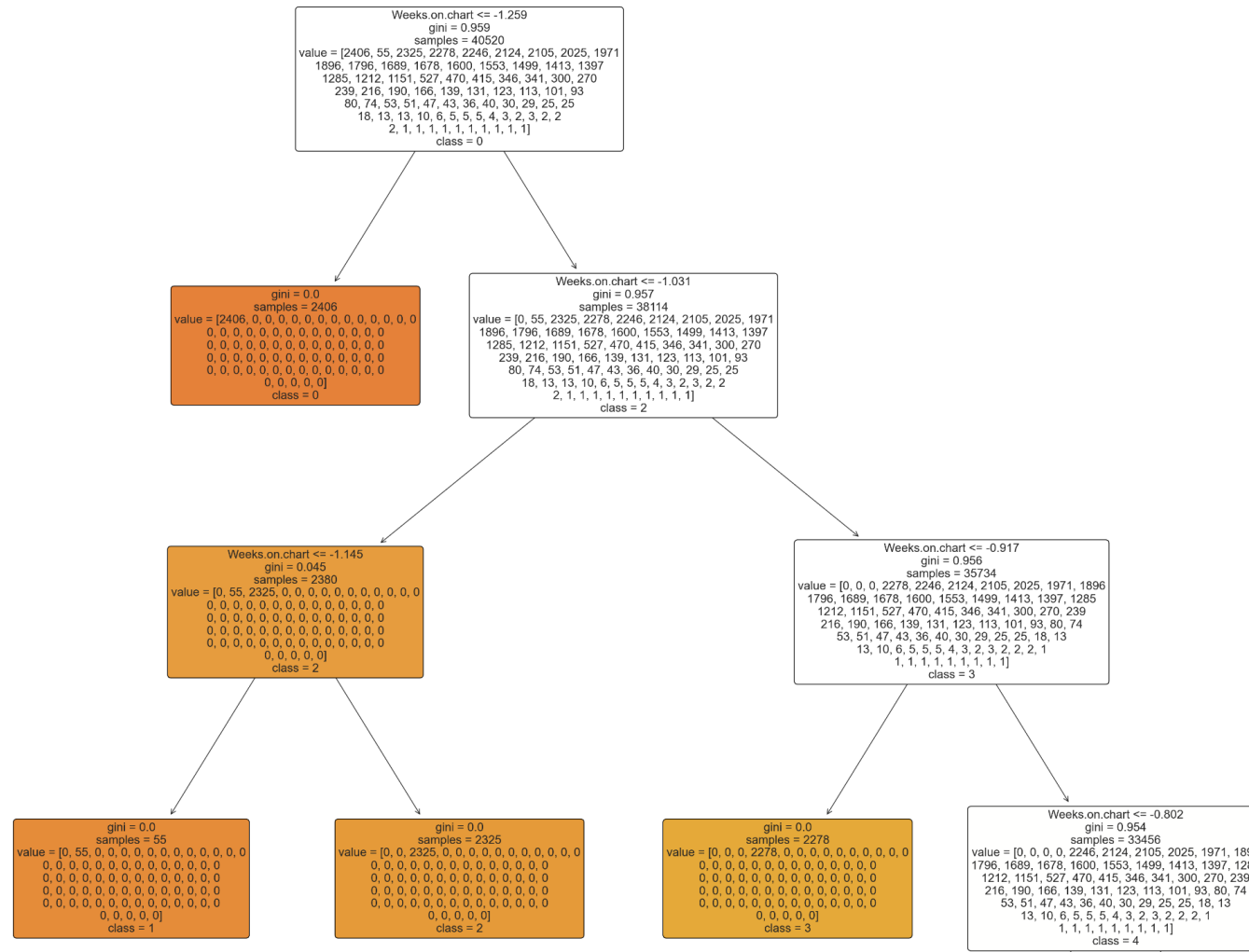
2010 to 2019
R square: 0.979

features	coef
valence	0.2107
time signature	0.1814
target	-0.0051
instrumentalness	0.1507

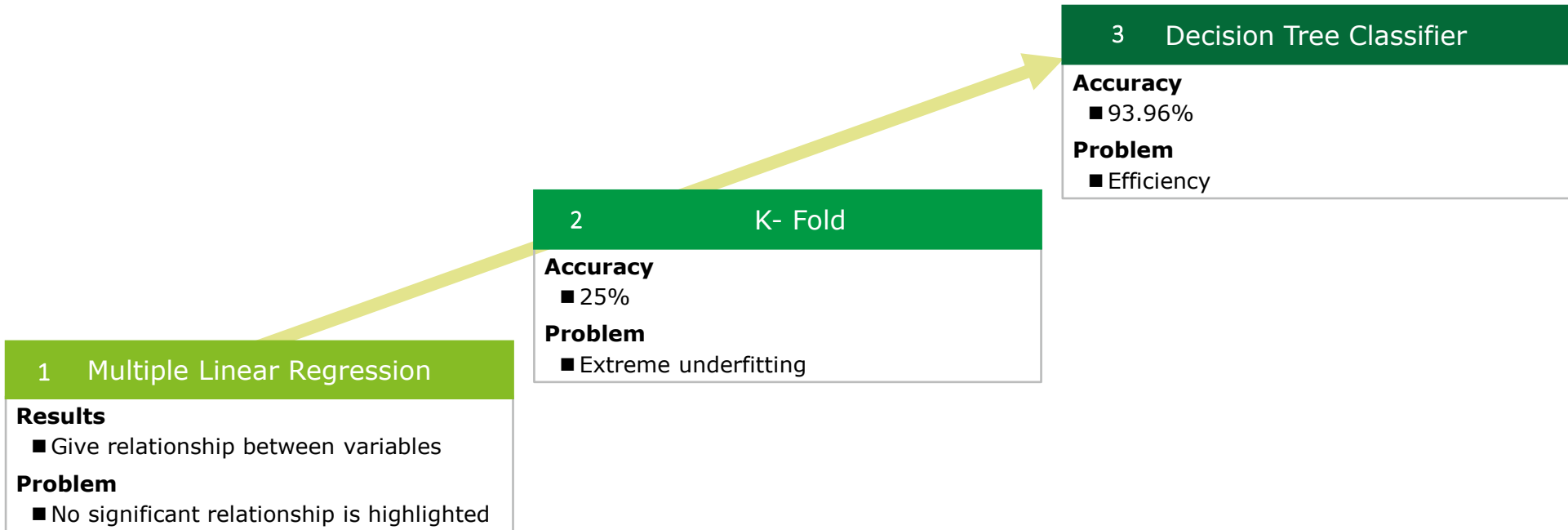
K - Fold



Decision Tree Classifier



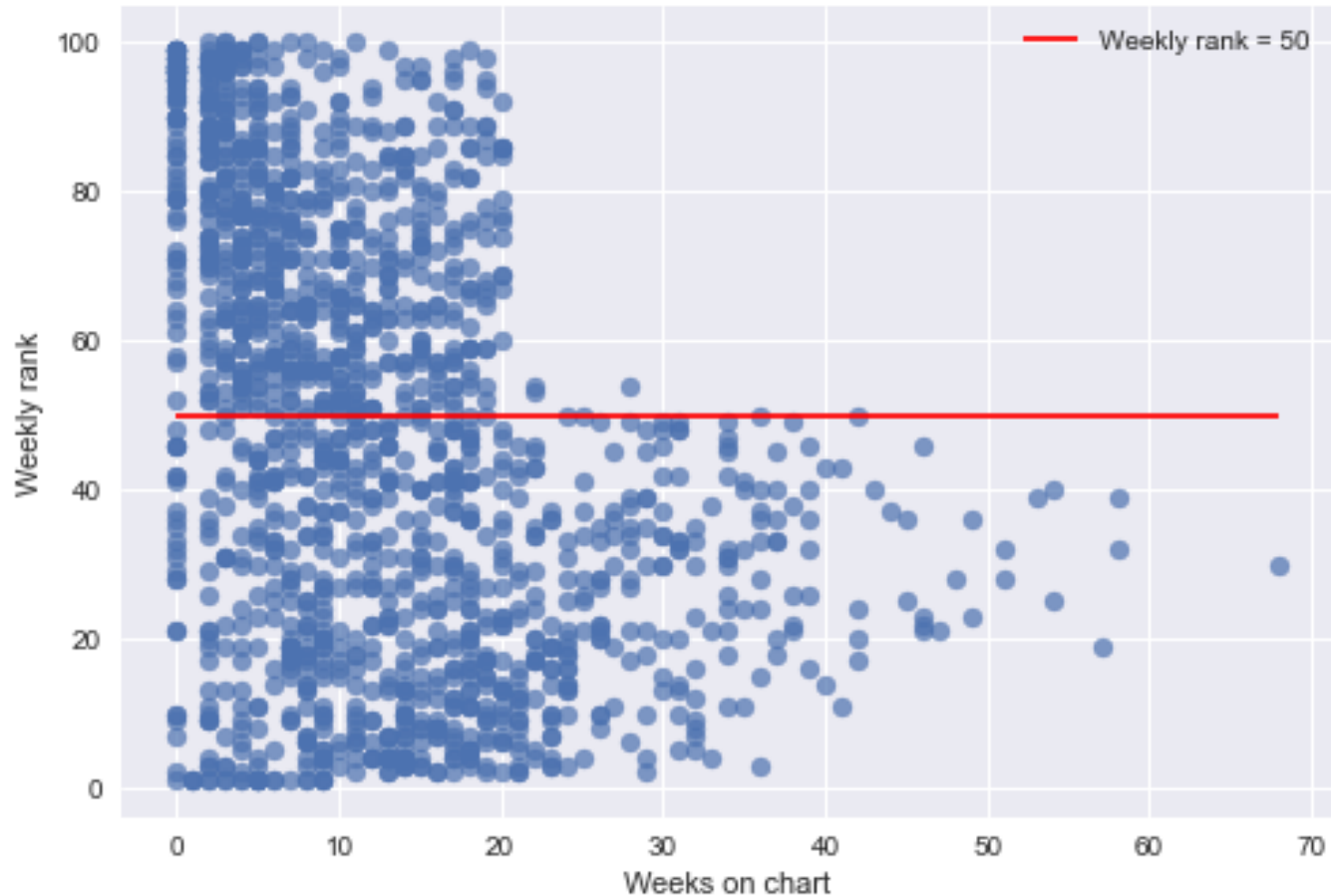
Results



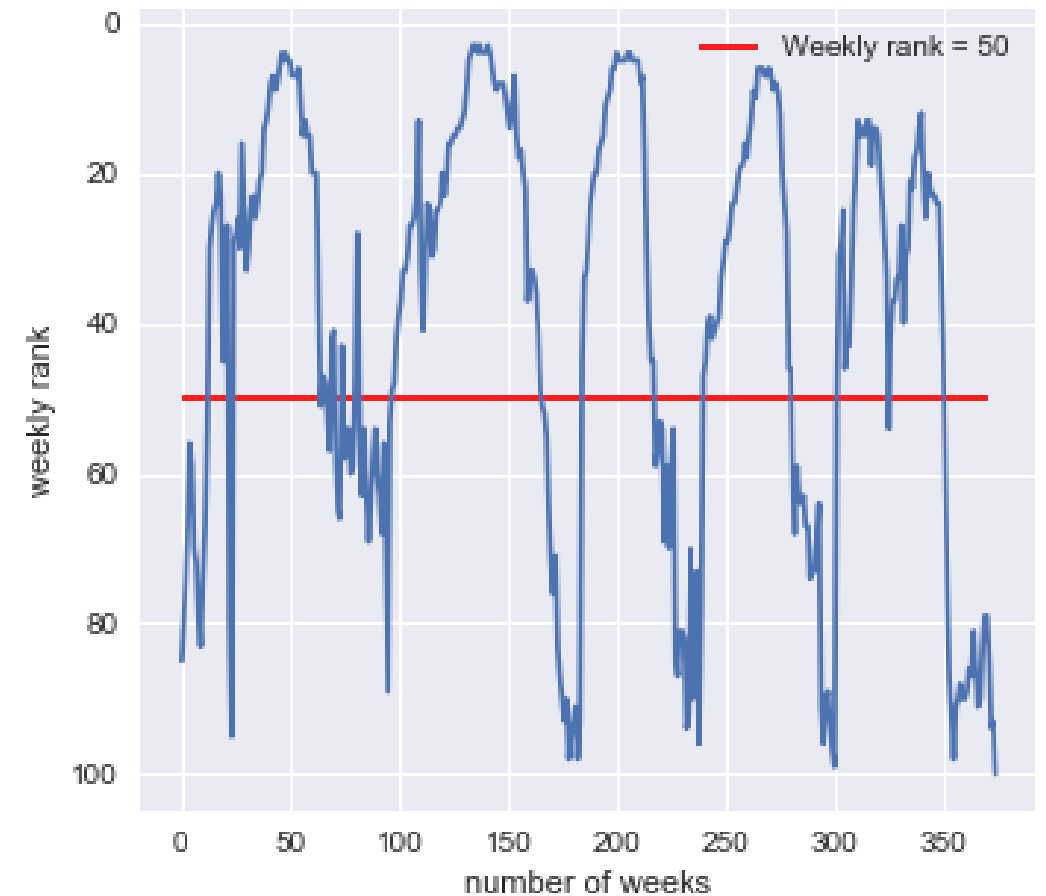
Evaluation

What can make model work better?

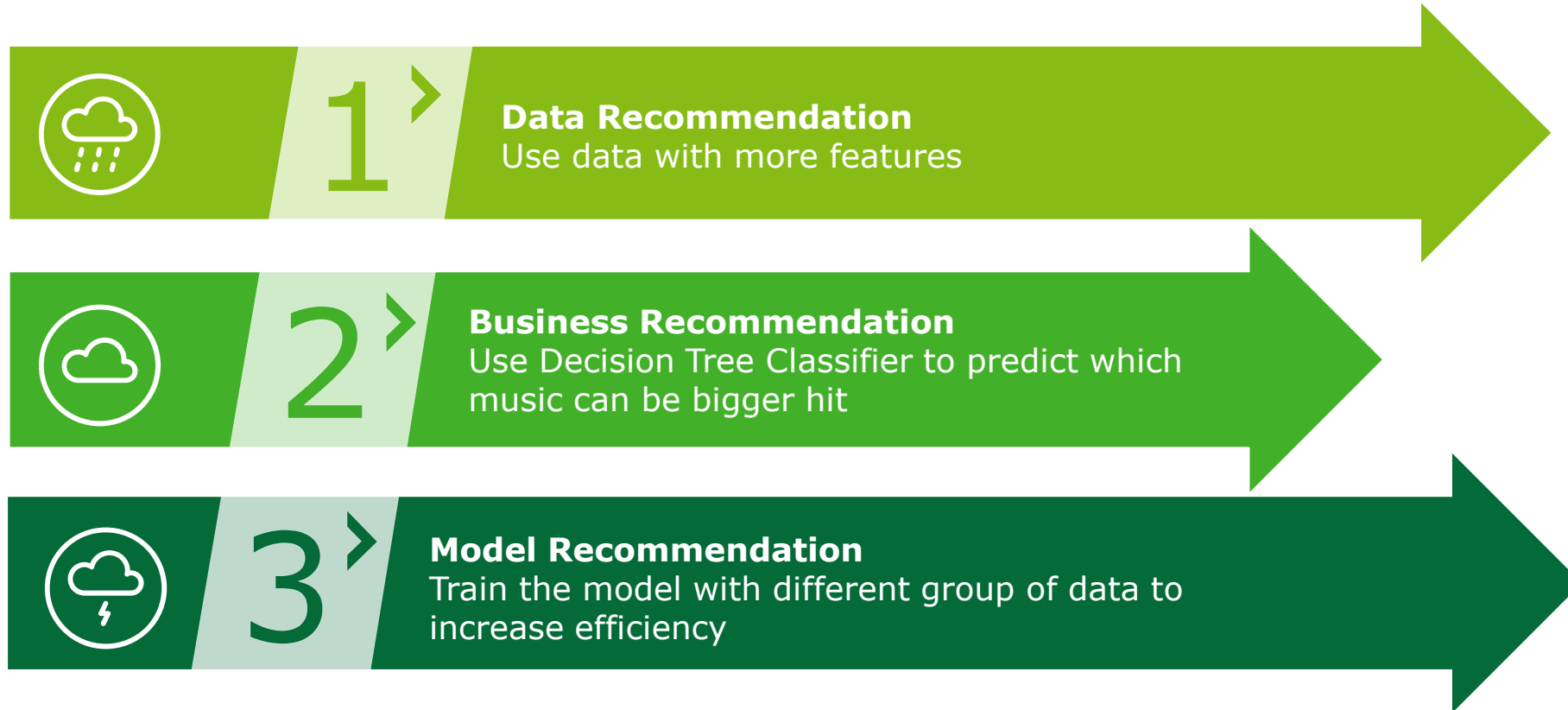
2010s weekly rank vs weeks on chart



Imagine Dragons' weekly rank



Recommendation





Thank you for your time
Questions?

Sara Yun

USDC Solution Analyst

saryun@deloitte.com

