



Columbia University

Spotify PAC

Presentation by **Ishu Jaswani**

Applied Analytics | 2022

Problems

Dwadling problem

At first, I didn't split the large data, the code was taking forever to run. I tried to also include performers as all factors in regression trees but it wasn't adding value, and taking a lot of time to run.



What i Did Right

Implementation

Data Wrangling

I one-hot encoded top 10 occurring genres to include in the model.

models

Ran and tuned all regression models taught in class. This helped me check how every model performs with the dataset.

What would I do differently

Wrangle the genre data to catch the essence of all predictors not only top 10.

Wrangle and clean the performance data in such a way that it has shorter bins.

Tune more hyper-parameters so that I can then find the most optimal model.



Models Performance

Page

4

