Capstone - Project 2
Project Showcase
Mark Hassan

Flatiron Data Science

December 2025

Project Overview

Business Problem

- Customer: FutureProduct Advisors (FPA), a consumer products consultancy
- Requirement: FPA clients want music that appeals to a broad audience but is inexpensive to license
- Impact: FPA consultants spend 4+ hours per week researching new music

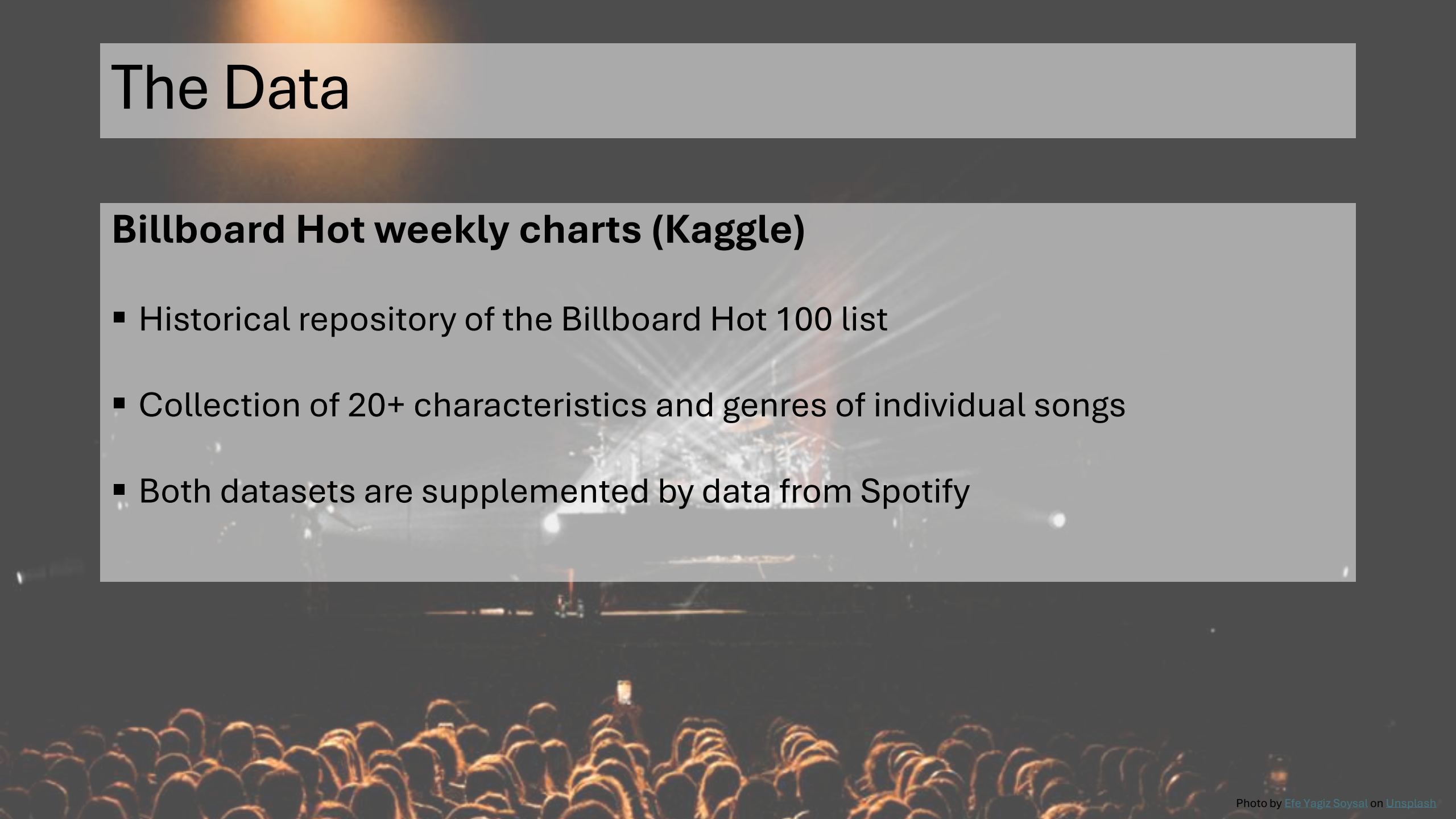
Goals

- Build data tools that can evaluate the Billboard Hot 100 list and predict:
  - The highest position each song on the list will achieve
  - The largest number of places each song will climb in one week

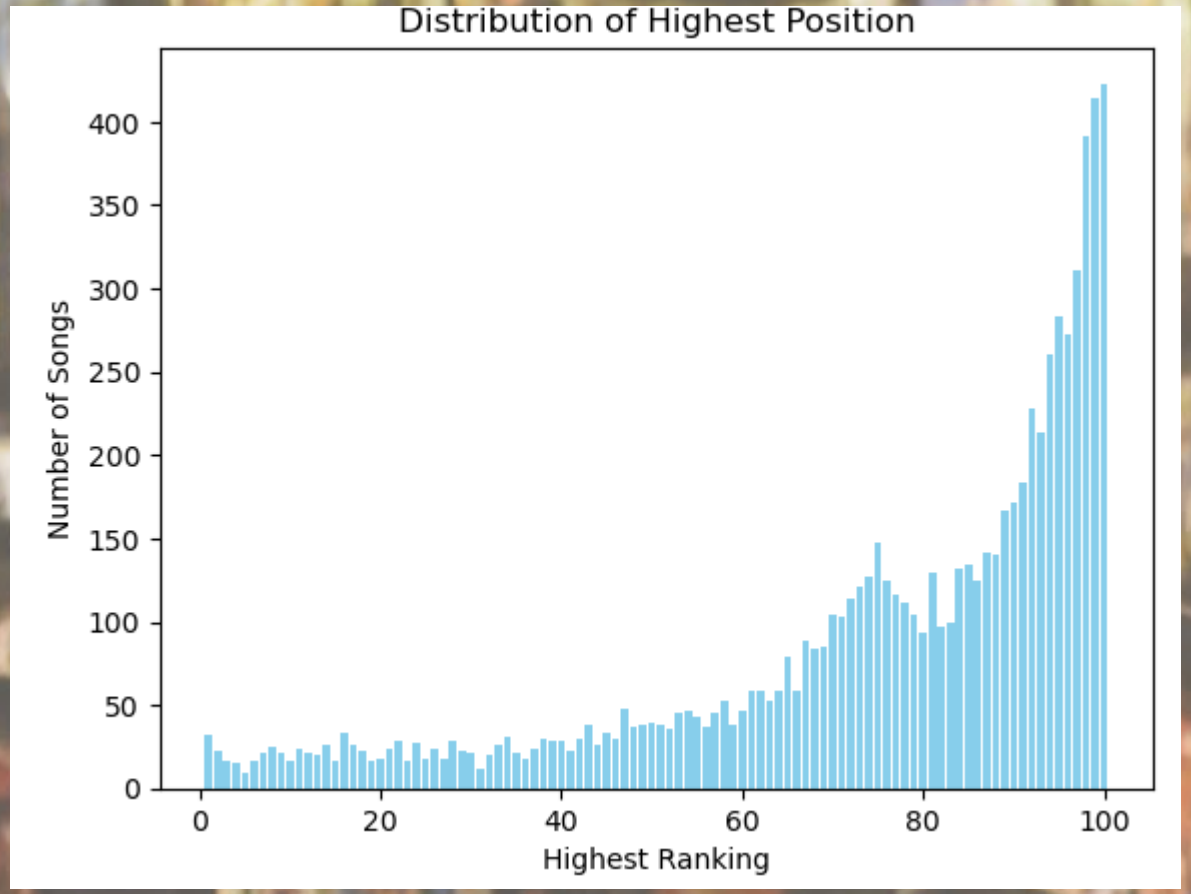
The Data

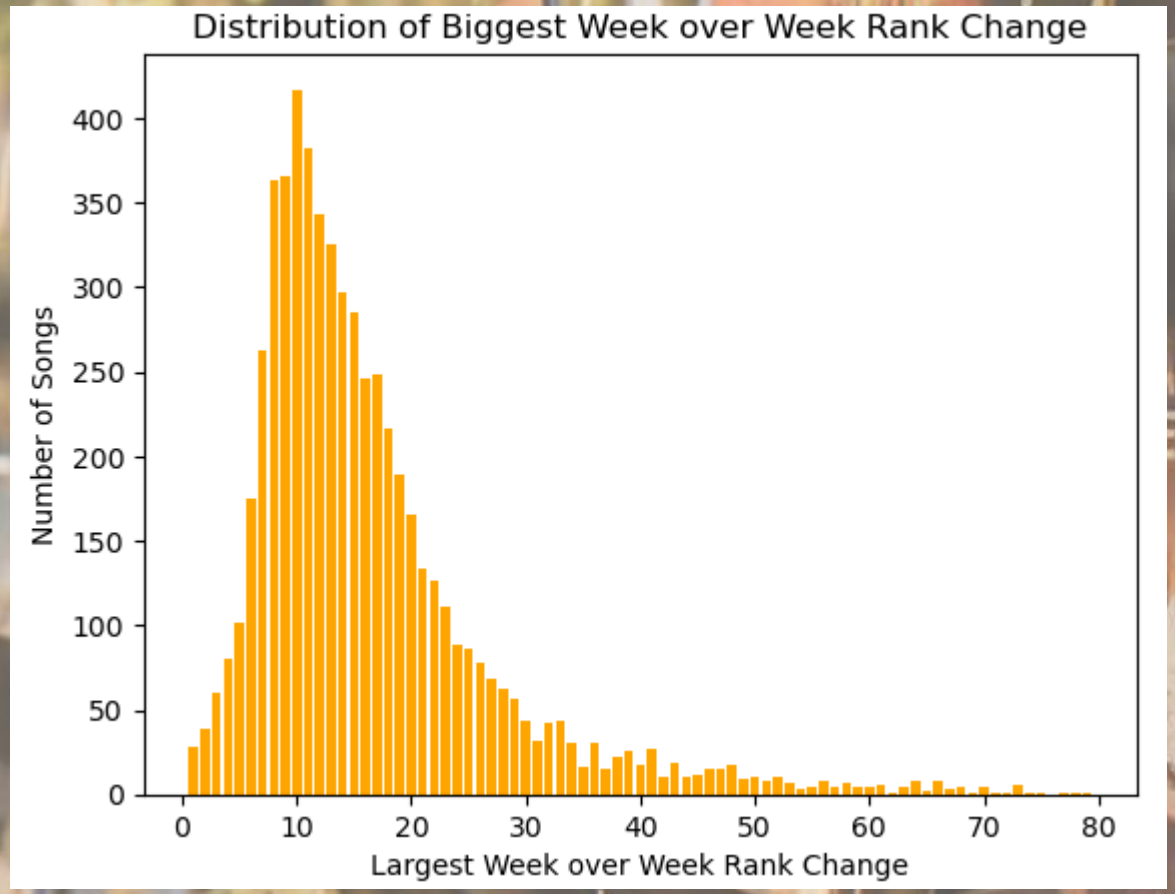
Billboard Hot weekly charts (Kaggle)

- Historical repository of the Billboard Hot 100 list
- Collection of 20+ characteristics and genres of individual songs
- Both datasets are supplemented by data from Spotify



The Data (continued)





Modeling Approach

Methodologies to Explore

- XGBoost: decision tree-based model
- k-Nearest Neighbors (k-NN): classification based on “neighbor” categories
- Deep Learning: artificial neural network that learns patterns in data

Data Features

- Including different combinations of data for each model
- Example: modeling with and without genre information

Modeling Approach (continued)

Evaluation Criteria

- Ability to learn from historical data (“fit quality”)
- Accuracy when making predictions on “new” data
- Complexity and resources needed to execute

Modeling Outcomes

XGBoost

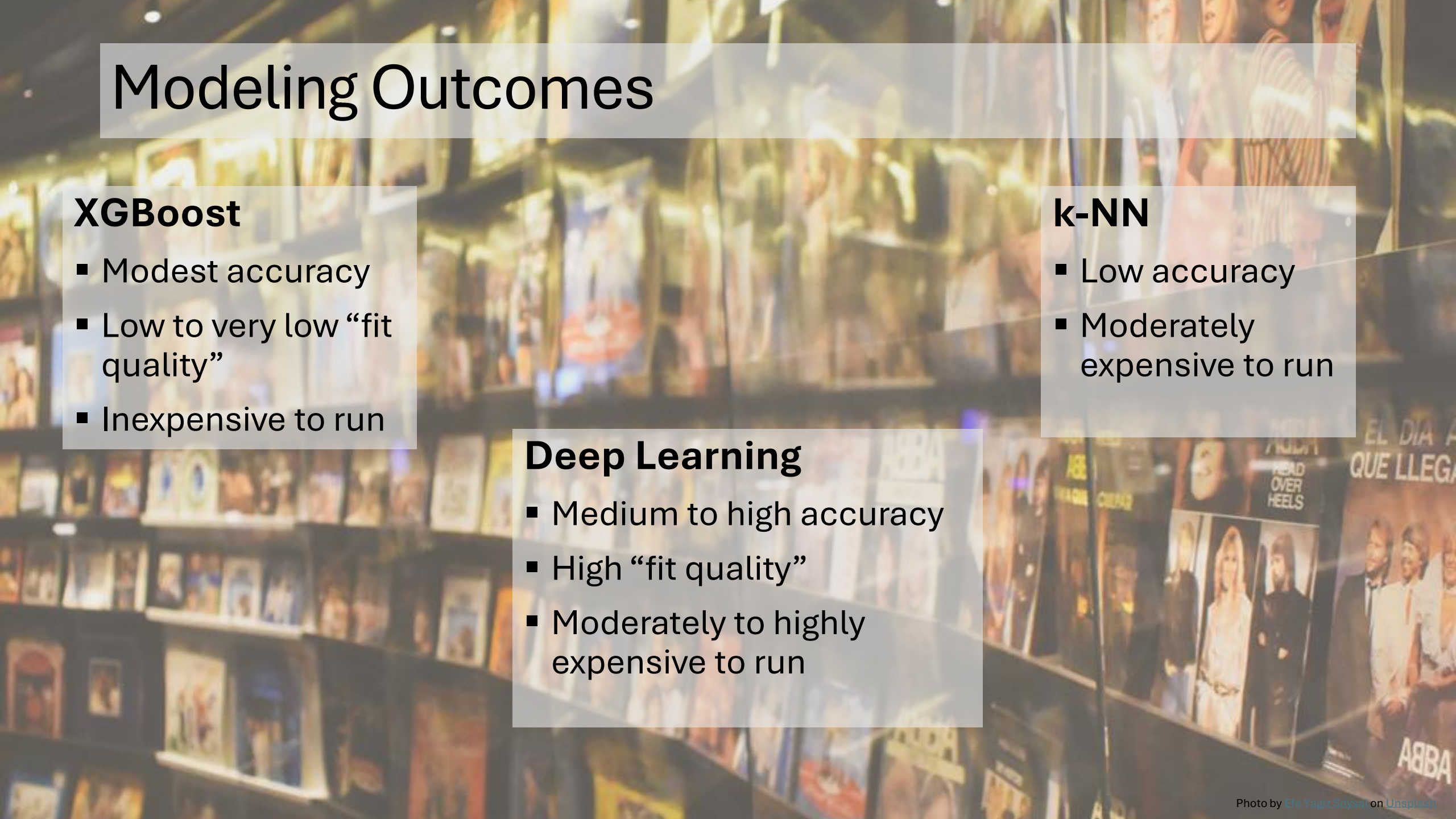
- Modest accuracy
- Low to very low “fit quality”
- Inexpensive to run

Deep Learning

- Medium to high accuracy
- High “fit quality”
- Moderately to highly expensive to run

k-NN

- Low accuracy
- Moderately expensive to run



Modeling Outcomes – Best Models

Predicting Highest Song Position

- "3-layer, regularized model" trained on data with genre *included*
- Predictions have an average error of 17 positions against an average Highest Song Position of 75

Predicting Biggest Position Change

- "3-layer, regularized model" trained on data with genre *excluded*
- Predictions have an average error of 9 positions against an average Biggest Position Change of 13

Conclusions

Highest Song Position

- The Highest Song Position model will deliver significant value to FPA consultants and their customers

- At minimum, this model allows FPA consultants to confidently predict whether a song will end up in the top, middle, or bottom of the Hot 100 list

Biggest Position Change

- The Biggest Position Change model delivers moderate value to FPA consultants and their customers

- This model gives directional guidance on the rate at which a song will climb the Hot 100 list

Next Steps

Moving to Production

- Both prediction models have been saved and end to end data processing is ready
- Once the external data source is configured, FPA will receive a weekly update of predictions on the Hot 100 list

Future Opportunities

- There is a wide array of additional data to explore (e.g. sentiment analysis of written reviews)
- Additional modeling approaches may also give new insights