Spotify Final Project

Yan He, Peter Valentine, Amany Yaakoub, Leona Yang

Agenda

- Selected Topic and Reason
- Description of Source Data
- Questions to be Answered
- Data Exploration Phase
- Analysis phase
- Dashboard Plan

Selected Topic and Reason

- Dataset of Spotify tracks Predict the popularity of a track
- High interest in music and to better understand how a track becomes a hit

Description of Source Data

Popularity	Duration	Key	Mode	Time Signature
Acoustic	Energy	Liveness	Speech	Valence
Danceability	Instrumental	Loudness	Tempo	

Questions to be Answered

- What features are the most predictive in determining track popularity?
- What are the optimal combination of features for popular tracks?
- Are there any differences across genres?

Data Exploration Phase

- Used Python (Pandas) to examine records, for example:
  - How many tracks were in each type of genre?
  - Which ones were underrepresented?
  - Which features likely had no bearing on whether it was a hit (such as key or tempo)?
  - How many tracks could be removed from various genres that were not likely to be musical “hits” such as comedy?

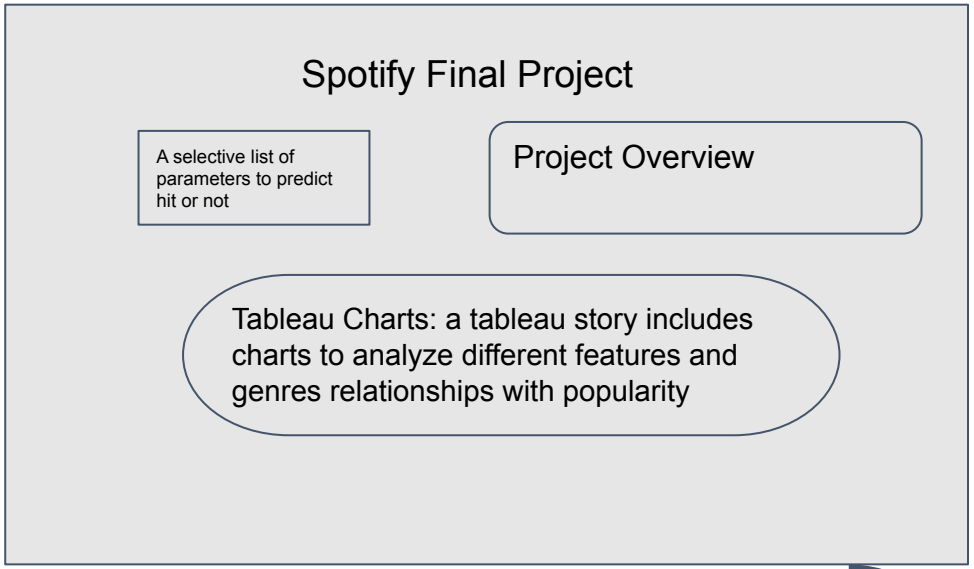
Analysis Phase

- Started with ETL
- Random Forest Machine Learning Model (used to help identify predictive features)
- Data will be analyzed/visualized using Tableau
- Website to be developed with dashboard and a prediction tool to predict tracks to be a hit or not?

Dashboard Plan

- Several dashboards to be done in Tableau
  - Artists dashboard bar graph
  - Features vs popularity scatter graphs and line chart
  - Average popularity vs genre bar graph
  - Top 100 artists (using sum)

Sample Dashboard



Thank you