CS 327E Final Presentation

By: Alexia Carmona and Amber Vasquez

Area of Interest and Questions We Set Out to Answer (Milestone 1)

- Interested in the movies topic
- IMDB primary dataset
- Bollywood secondary dataset
- How female leads or female directors' movies are received?
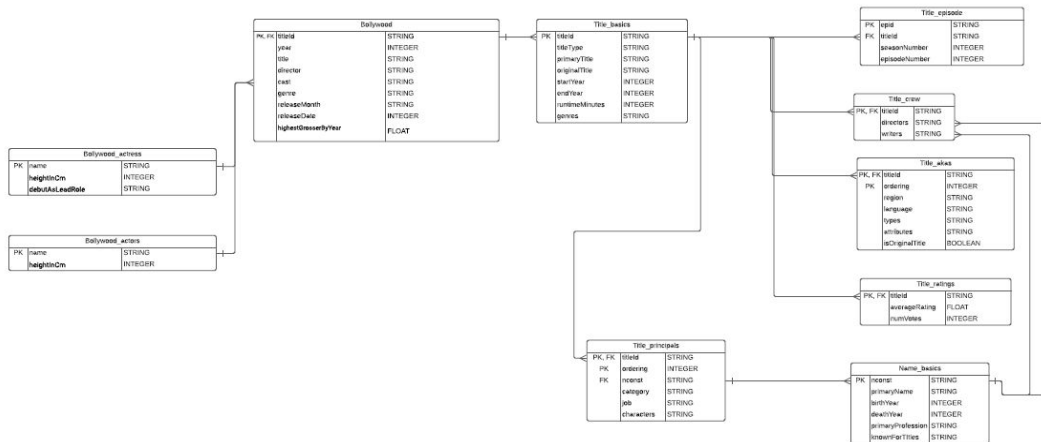- If Bollywood Industry surpasses Hollywood Industry?

Dataset Selections and Overview of the Raw Data (Milestone 1)

- Chose IMDB and Bollywood datasets
- Bollywood dataset - 3 tables
- IMDB - 7 tables

Staging and Modeled Tables (Milestone 1 and Milestone 2)

- Bollywood - 3 staging tables
- IMDB - 7 staging tables
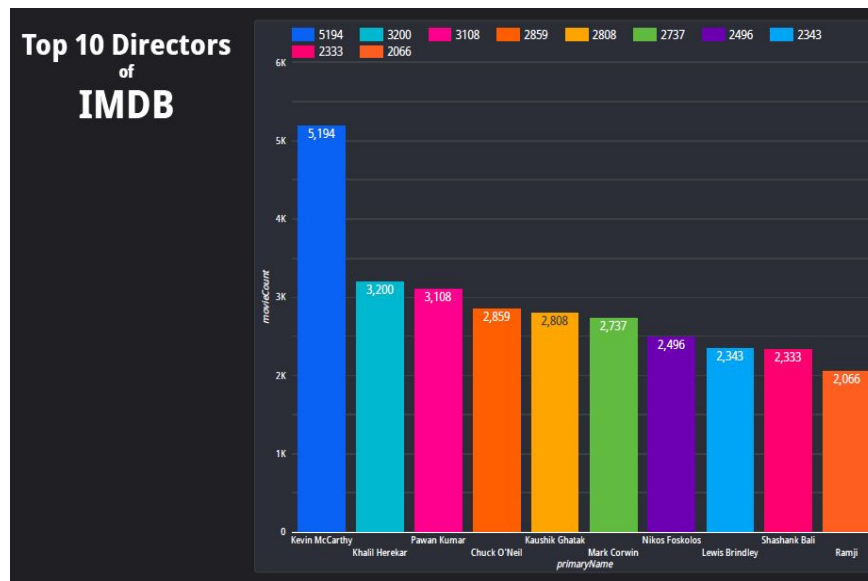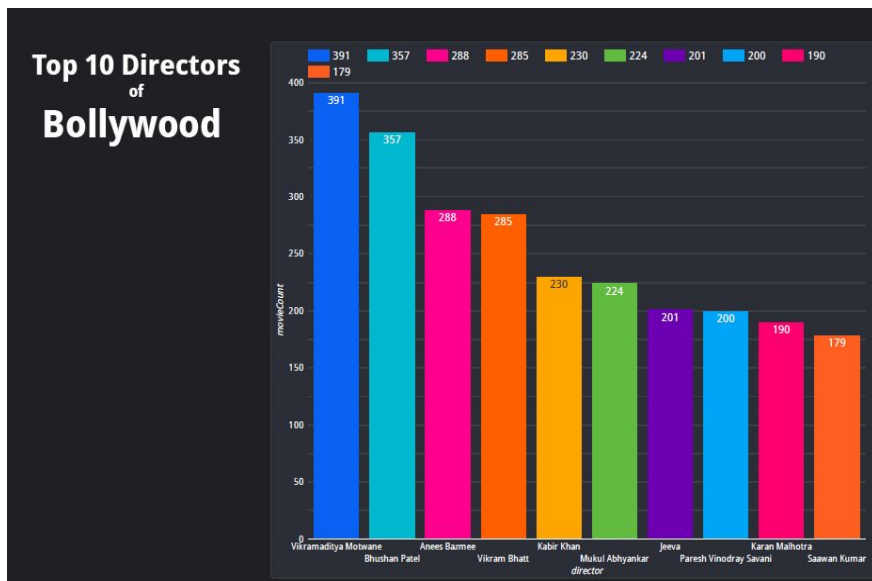- Parsing errors
- Modeled tables
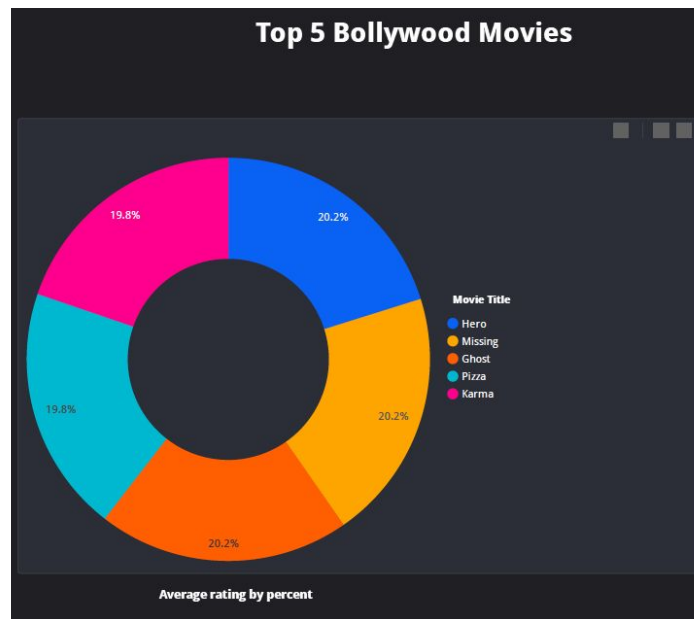
Beam Pipeline(s) (Milestone 3)

- A few array types
- Created 7 beam pipelines from Title Basics, Title Crew, Name Basics, Bollywood tables
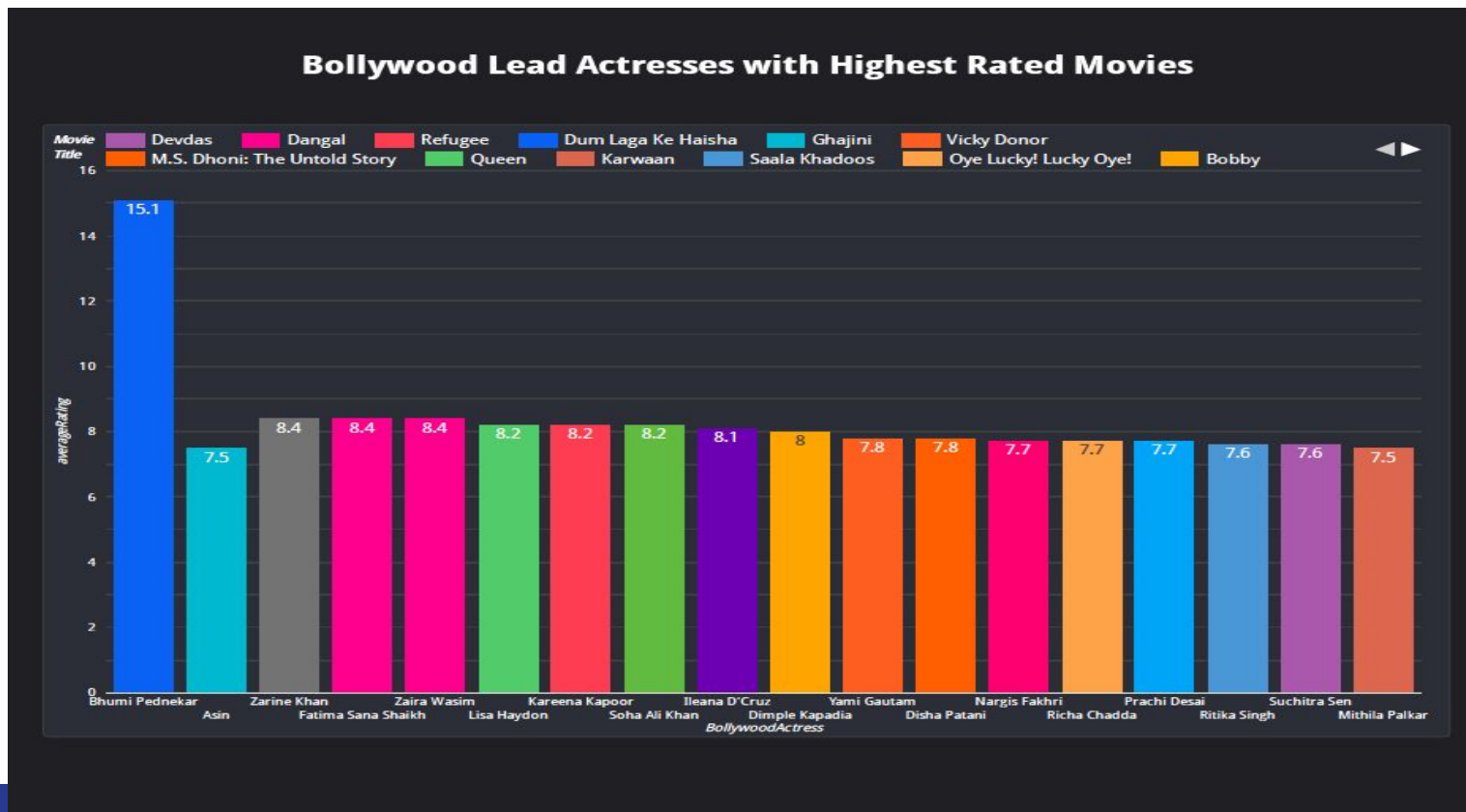
Cross-Dataset Queries Demo and Data Visualizations (Milestone 4)

Cross-Dataset Queries Demo and Data Visualizations (Milestone 4)

Cross-Dataset Queries Demo and Data Visualizations (Milestone 4)

Bollywood Lead Actresses with Highest Rated Movies

Future Improvements to Our Project

- Get more Bollywood data

- Explore Bollywood tv shows vs non Bollywood tv shows