Final Presentation
Box Office Prediction

DATA 515 Software Design For Data Science
Mar 14, 2024

Apratim Tripathi, Trisha Banerjee,
Rohit Chandiramani, Zongze Li

UNIVERSITY of WASHINGTON

Background

Our project introduces an innovative tool designed to aid the movie production industry. Aimed at producers, directors, casting directors, and industry insiders, this tool leverages predictive analytics to forecast the potential revenue of movies that are still in the conceptual phase.

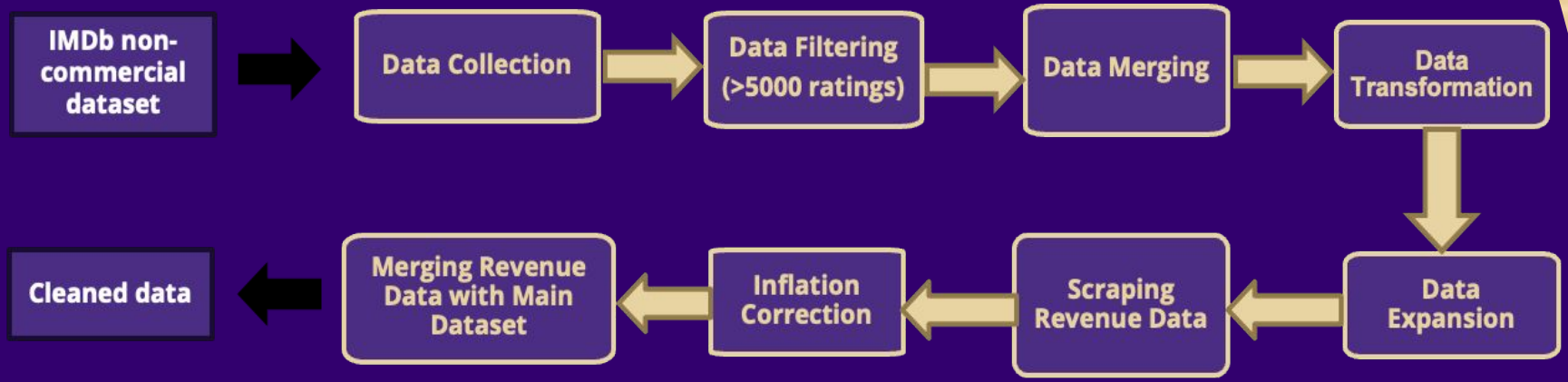
Key Features:

- **Revenue Prediction**: By inputting data such as actors, actresses, writers, directors, and genres, our tool provides an estimated revenue figure for hypothetical movies.
- **Analytics Dashboard:** Users can explore in-depth analytics showing the performance of actors and actresses across various genres over the years. This valuable insight assists in making informed decisions about casting and genre selection.
- **Strategic Decision Making**: The tool aids in strategizing the hiring process, focusing on maximizing profits by selecting the ideal combination of talent and genre based on historical data and predictive analysis.

Our Goal:

- To empower industry players with data-driven insights, enabling them to make strategic decisions that enhance profitability and success in the competitive landscape of movie production.

Data Used



- Unique combinations of actor-actress-director-producer-writer-genre
- Around 3,00,000 combinations
- Key fields: IMdb rating, Number of votes on IMdb, Year of release, Runtime, Infation-corrected Revenue

Use Cases - Scenarios

- Consultant analyzing performances
- Casting director predicting ratings
- Film producer predicting revenues

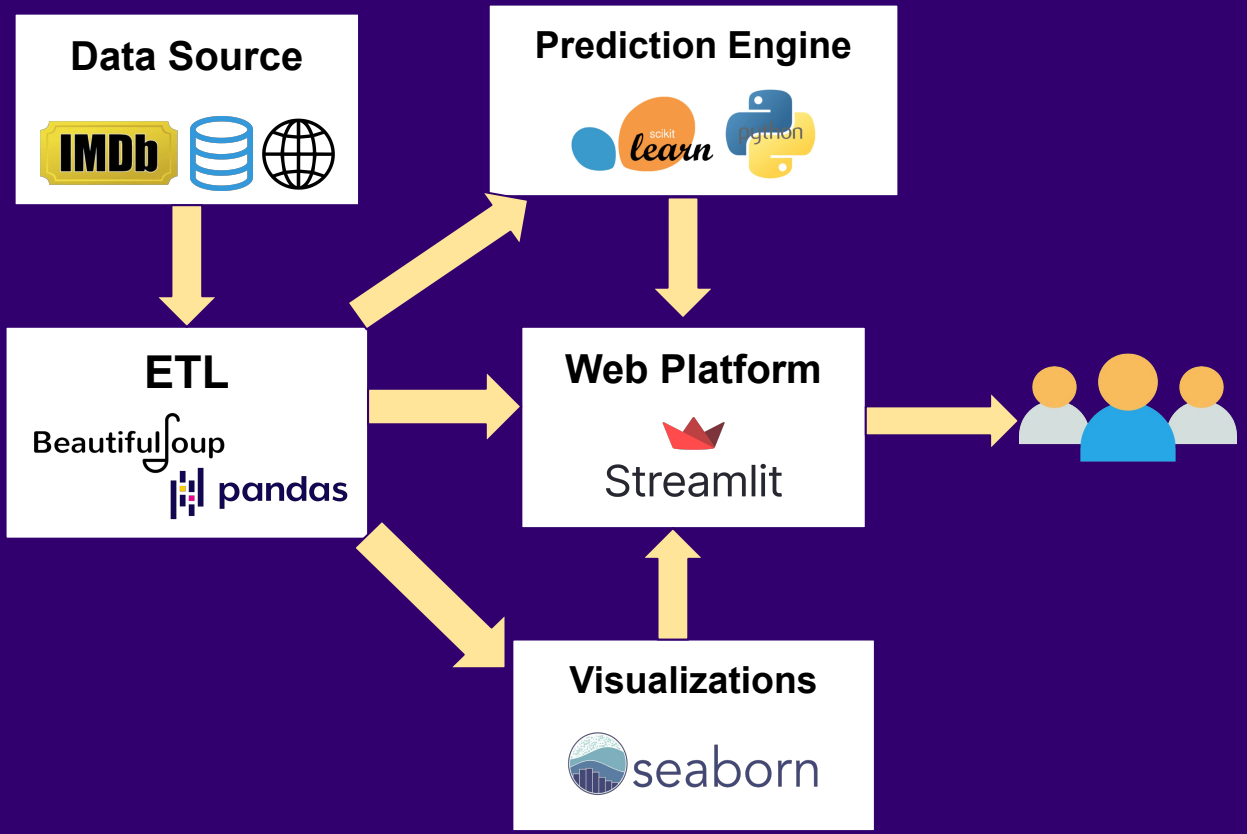
Use Case – Analyzing Performances

- User select Actors or Actresses
- User select a genre (e.x. Action, History, Romance, Sci-Fi)
- User select the top X actors/actresses to display, sorted by revenues
- Select an actor/actress to view his/her historical revenue trend (double clicks to highlight, single click to hide)

Use Case – Predicting Ratings & Revenues

- User select the Primary Genre
- User select the Actors, Actresses, Directors, Writers and Production House
- User enter Runtime in Minutes (default value is 150, with max being 300)
- User click predict button to see the predicted ratings / revenues

Design


Data Source
IMDb
Prediction Engine
scikit
learn
Python
ETL
Beautiful Soup
pandas
Web Platform
Streamlit
Visualizations
seaborn

What's next?

Lessons learnt

- Design Specification
  - Define use case
  - Design components and interaction
- Collaborating project with Github
  - Use branch
  - Commit, push, and pull regularly
- Streamlit limitations
  - Design and theming capabilities
  - Integrate external CSS files

Future works

- We would like to build an automated process where, it periodically pulls data from imdb datasets and add them to our model.
- We want to add a recommendation system which can recommend which cast and crew combination may work better together based on our selection.

Demo

Thank you!

Questions?