ECE 143 : Project Proposal for Group 9

Exploratory Analysis to Find Trends in Movie Data

Eeshaanee Arya, Feifan Xu, Kai Wang, Sina Malekian

Github Repository: https://github.com/Feifan1/Project-for-ECE-143

1. PROBLEM STATEMENT

In this project, we plan to investigate a movies dataset to find trends in the data. The analysis can help a film production company like Warner Bros., Fox Star Studios, etc. to decide which movie titles can mean profits if funded for production, how much to bid on produced movies, when to release, how much to invest in marketing and PR, etc. This information will not only help the studio before a title is released, but also helpful after the release. An example for this could be - the analysis can help with insights on how much a studio should negotiate with on-demand streaming companies for a second window streaming rights, etc. That said, it is not a prediction problem - it involves studying the analysis and understanding the conclusions.

2. DATASET

The dataset has been acquired from IMDB API : http://www.omdbapi.com/, and it contains around 30 attributes for around 40,000 movie titles. Some of these attributes include:
Year of Release, Runtime Duration, Genre, Language, Country of Production, Awards, IMDB Rating, Budget, Gross income and so forth.

3. TIMELINE

Our project is divided into mainly three phases:

1) Planning and Data pre-processing -
This phase is the most crucial to our project as it involves identification of the attributes that will be compared in order to derive conclusions. We want to validate obvious trends, but at the same time, we want to uncover some interesting and unexpected finds. One example for this could be - how does the revenue change
This phase will also involve validating the dataset and pre-processing it to extract important information which will be of interest.

2) Programming for data analysis and data visualisation -
This phase has two main components: Analysis and Visualization. The former would comprise of programming for data analysis of the pre-processed data. The latter one will involve selection and implementation of visualization techniques to represent the trends, for eg. bar graphs, pie charts, trend charts, boxplots,etc. This phase will also include trials and testing with the data to find interesting trends.

3) Conclusion and Reporting -

This phase will comprise of concluding the analysis, and reporting the findings.

4. SCHEDULE

Timeline	Task	Contributor
Week 5	i) Planning ii) Data pre-processing	Eeshaanee, Sina
Week 6-8	i) Programming for data analysis ii) Data visualisation	Eeshaanee, Sina Kai, Feifan
Week 9-10	Conclusion, Reporting and Presentation	Eeshaanee, Sina, Kai, Feifan