Demographic Trends and Box Office Performance:

A Data-Driven Analysis of Four-Quadrant Films

Marl Jonson

Department of Computer Science, San Jose State University

CS 100W: Technical Writing Workshop

Dr. Gregory D. Tomlinson

May 12, 2025

TABLE OF CONTENTS

I.   Introduction and Background Research

In film marketing, a "four-quadrant film" is a film that appeals to viewers of both sexes aged over and under 25. These groups are the four major demographic groups of a film-going audience: men under 25, women under 25, men over 25, and women over 25. The goal of a four-quadrant film is to maintain universal appeal, doing so by combining genres and portraying narrative to which people of different genders, ages, and cultural backgrounds can relate.

According to [1], it is a rare occurrence for a film studio to produce a film focused on only one quadrant. Of the films that have grossed 1 billion dollars, all 50 fall under the classification of a four-quadrant film. The popular strategy of producing and marketing a film to target a diverse range of demographic groups has proven successful in maximizing global box office returns, as delineated by [2] and [3].

Data analysis can be used to represent demographic trends in the box office performance of films by interpreting and visualizing data of both film viewership and the information in the films themselves. For example, a simple proportional analysis can be applied to film data to interpret the increase in percentage of female leads in major films within the past 50 years. The results of this analysis may suggest that action films with female leads are more likely to generate more revenue than those with male leads.

To abide by copyright law, this project will use datasets from the Internet Movie Database website (developer.imdb.com) that are designated as freely usable. The goal of the project is to use Python and data visualization libraries to discover trends based on box office performances and viewer demographics that can guide large-scale film production and marketing decisions within the corporate film industry. Insights developed in this project have numerous real-world applications. Film studios like Universal, Sony, and Paramount can use this project's insights to

make more inclusive casting decisions in future film developments and productions. Advocacy groups can use this project's insights to strengthen their arguments for the representation of underrepresented communities in film roles.

II.   Required Personnel

The first team member needed is a project manager, who may also serve as a lead technical developer. This team member should have six months of practical experience working in a leadership-based technology role. Preferred qualifications include an Agile or Coursera project management certificate. The duties of the project manager are to plan the overall project by establishing a timeline of expected progress, tracking team members to ensure their completion of duties, documenting the work so that it can be reproduced, and tailoring the final proposal for presentation to a predetermined audience of committee members.

The second team member needed is a data analyst. This team member should have six months of practical experience working with Python programming, moderate familiarity with the Jupyter Notebook file format, and experience in computer programming or data analysis equivalent to at least one college-level course. The duties of the data analyst are to select relevant Python libraries for data manipulation, numerical operations, and data visualization. They are also responsible for applying programming methods to handle missing data values, statistical outliers, and data type conversions within the public data sets used for the project.

The third team member needed is a UI/UX and graphic designer. This team member should have two years of experience designing slideshows that are visually appealing and easy to understand. The designer should be very comfortable navigating digital presentation programs such as Microsoft PowerPoint (preferred), Google Slides, or Canva. The designer will maintain a

relationship with the data analyst; the data analyst's data visualizations will be used by the designer in the designer's slideshow.

The fourth team member needed is a data engineer. This team member should have six months of practical experience working with Microsoft Excel, Google Sheets, and CSV files. The data engineer does not need to have experience managing databases using SQL or NoSQL, as the project is simple enough to maintain the data files in their original format.

The fifth team member needed is a subject matter expert. This team member should be enrolled in (or graduated from) an undergraduate degree program specializing in marketing, film studies, theatre, or the arts. The subject matter expert should maintain a comprehensive knowledge of contemporary trends and viewer demographics in the Hollywood film industry. Additional knowledge of multi-cultural or foreign film industries is recommended.

III.   Associated Technologies

The project will be implemented in the programming language Python. The project will use the virtual development environment Jupyter Notebook, which is an ideal choice because it represents all steps (lines of code) of the project in a sequential manner. The project will use the library Cinemagoer to import data from the Internet Movie Database (IMDb) and convert it to usable data. This data will be refined, filtered, and analyzed using the library Pandas. Then, the data will be visualized into user-friendly graphs and charts using the libraries Seaborn and Plotly.

For the presentation portion of the project, the graphic designer may use a slideshow design software of their choice to portray the project's narrative with the most impact. This may be a free software, like Microsoft PowerPoint, Google Slides, or Canva, or it may be a paid software, like Adobe Premiere Pro or Gamma.

IV.   Project Usefulness and Commitment to Ethical Outcomes

The researchers of this project are committed to maintaining transparency and high ethical standards in the analysis and presentation of the project and its data. The researchers understand that insights derived from this project may be interpreted in a discriminatory manner. The researchers will use any insights derived from this project in an ethically responsible manner that does not reinforce stereotypes of protected groups.

A possible ethical concern with the insights from this project is that it might produce unintended bias in the interpretation of the film data sourced from the IMDb database. Such bias might lead to misuse of the insights by a film production studio that may contradict the neutral ethical intentions of the team members. To mitigate this potential risk, this project will declare its technical and qualitative methodologies in the source code of the Jupyter Notebook. In addition, the slideshow and verbal presentation will explain the limitations and potential biases in the film data and resultant insights. One example of such bias is the data's emphasis toward a gender or racial preference that is reflective of the time period in which the data was collected.

V.   Timeline of Deliverables and Release Date

This project is designed to begin and end within the length of one academic semester, or 16 full weeks. The project manager may decide to accelerate or prolong the timeline of the project, according to the team's collective ability to work on the project.

The first deliverable is the Jupyter Notebook containing the data engineer's contributions. The Jupyter Notebook should contain the Cinemagoer package and IMDb data about the top 50 highest-grossing films. The film data should be formatted using the Python library Pandas. This deliverable should ideally be completed by week 4 of 18.

The second deliverable is the Jupyter Notebook containing the data analyst's contributions, which build upon the data engineer's progress. The Jupyter Notebook should contain visualizations of the imported IMDb data generated by the Python libraries Seaborn and Plotly. This deliverable should ideally be completed by week 8 of 18.

The third deliverable is the slideshow produced by the graphic designer. This slideshow should incorporate the visualizations produced in the Jupyter Notebook by the data analyst. The graphic designer should also develop a significant narrative that connects the insights in the slideshow to one another; they may consult the subject matter for assistance in developing this narrative. This deliverable should ideally be completed by week 12 of 18.

Between weeks 13 and 16, the project manager should organize a weekly team meeting at a designated time. This meeting can be held virtually, if necessary. The purpose of the team meetings is to review the team's progress toward the project's completion. It is also to rehearse and refine a short presentation pitch (5–10 minutes) to promote the project to a stakeholder audience. This rehearsal serves as an avenue for team members to practice both soft skills, like targeted communication and body language, and technical skills used in creating the project itself.

References

[1] Friend, T. (2009, January 11). *The Cobra*. The New Yorker. Retrieved April 14, 2025, from

https://www.newyorker.com/magazine/2009/01/19/the-cobra

[2] H. Wang and H. Zhang, "Movie genre preference prediction using machine learning for

customer-based information," *2018 IEEE 8th Annual Computing and Communication

Workshop and Conference* (CCWC), Las Vegas, NV, USA, 2018, pp. 110-116, doi:

10.1109/CCWC.2018.8301647.

[3] R. H. Amier and J. Setiawan, "Visualization and Prediction of Film Award Nominations by

Using of Visual Data Mining (VDM) and Exploratory Data Analysis (EDA) Method,"

2019 5th International Conference on New Media Studies (CONMEDIA), Bali,

Indonesia, 2019, pp. 84-88, doi: 10.1109/CONMEDIA46929.2019.8981822.