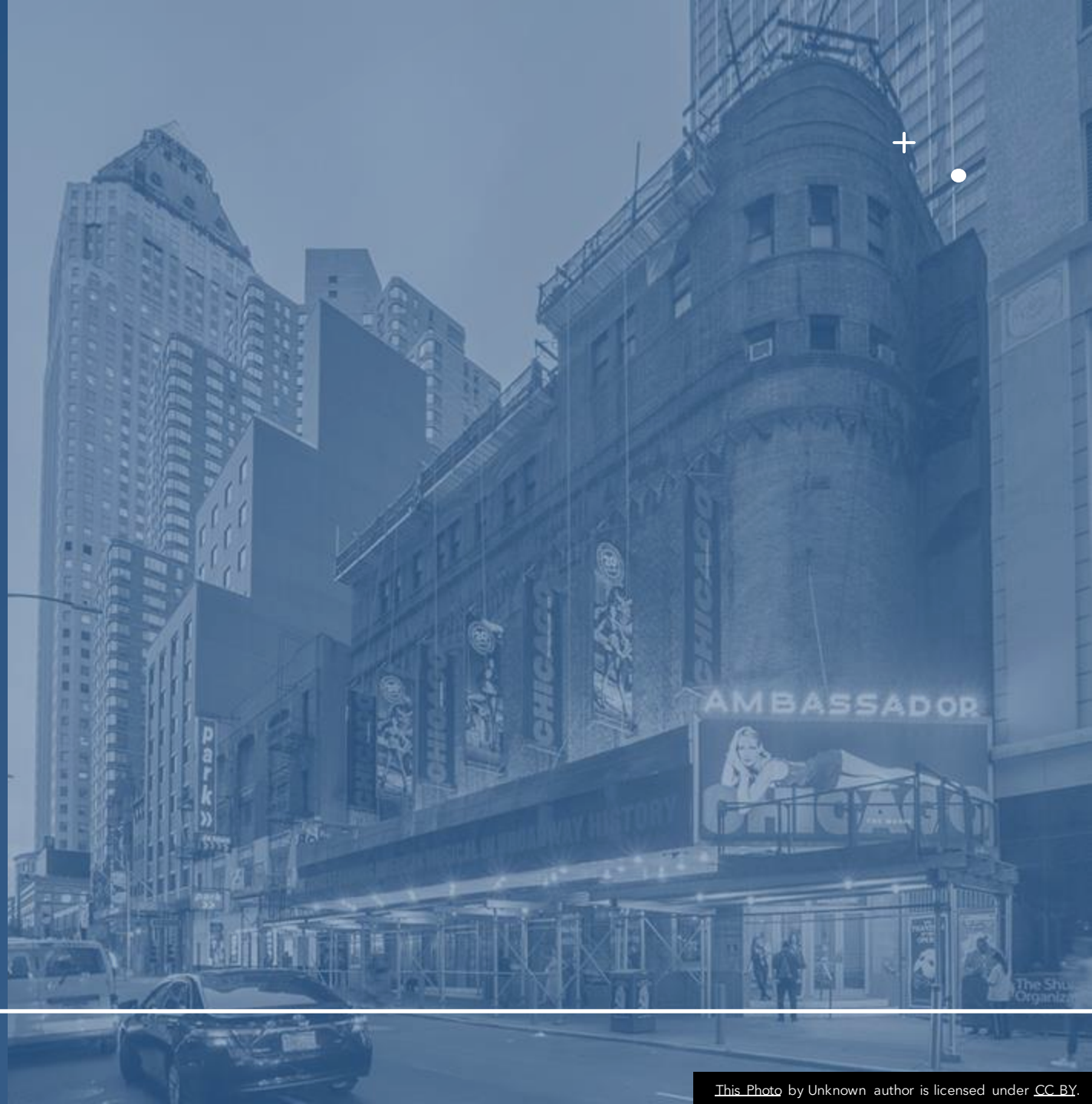


MOVIE EXPLORATORY DATA ANALYSIS

Pauline Ndung'u



BACKGROUND

- The film industry is an ever evolving industry, entering the industry is challenging without a roadmap to follow on. To new studios coming up, they need to know the history in the film-making business, which in a data analyst's world is information. This is helpful to understand the audience, the financial market and the seasons within which they are most likely to succeed in.



INTRODUCTION

The goal of my project is to analyze what type of movies are doing well at the box office in an attempt to be of assistance to a new entrant in the film-making world as well as existing film-production companies.

Some of the metrics I will be analyzing to achieve my goal are:

- Popularity
- Profitability
- Rating
- Runtime
- Release Month





TOOLS

In order to analyze data better and manipulate it, I used the following libraries:

- Pandas
- Matplotlib
- Seaborn

For extracting data from the IMDB database, I used the sqlite3 package



DATA MINING

The data sources used for this project was pulled from the following sources:

- Box Office Mojo
- IMDb
- Rotten Tomatoes
- The Movie Database
- The Numbers.

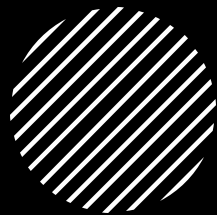
The dataset is relevant in finding out what types of films to create, as it provides a wealth of information on past movies, including box office earnings, audience and critic ratings, and genre classifications.

By analyzing the data, filmmakers can identify trends in the types of movies that have been successful in the past, and use this information to inform decisions on future movie productions.





DATA CLEANING



With the help of pandas, a Python Data Analysis Library, these are the steps I followed when cleaning the data:

- Identify any issues in the data, such as missing values, duplicates, inconsistent formatting, incorrect data types, or other anomalies.
- Remove or impute missing values
- Remove duplicates
- Correct inconsistent data formatting
- Check for outliers



Pitch deck title

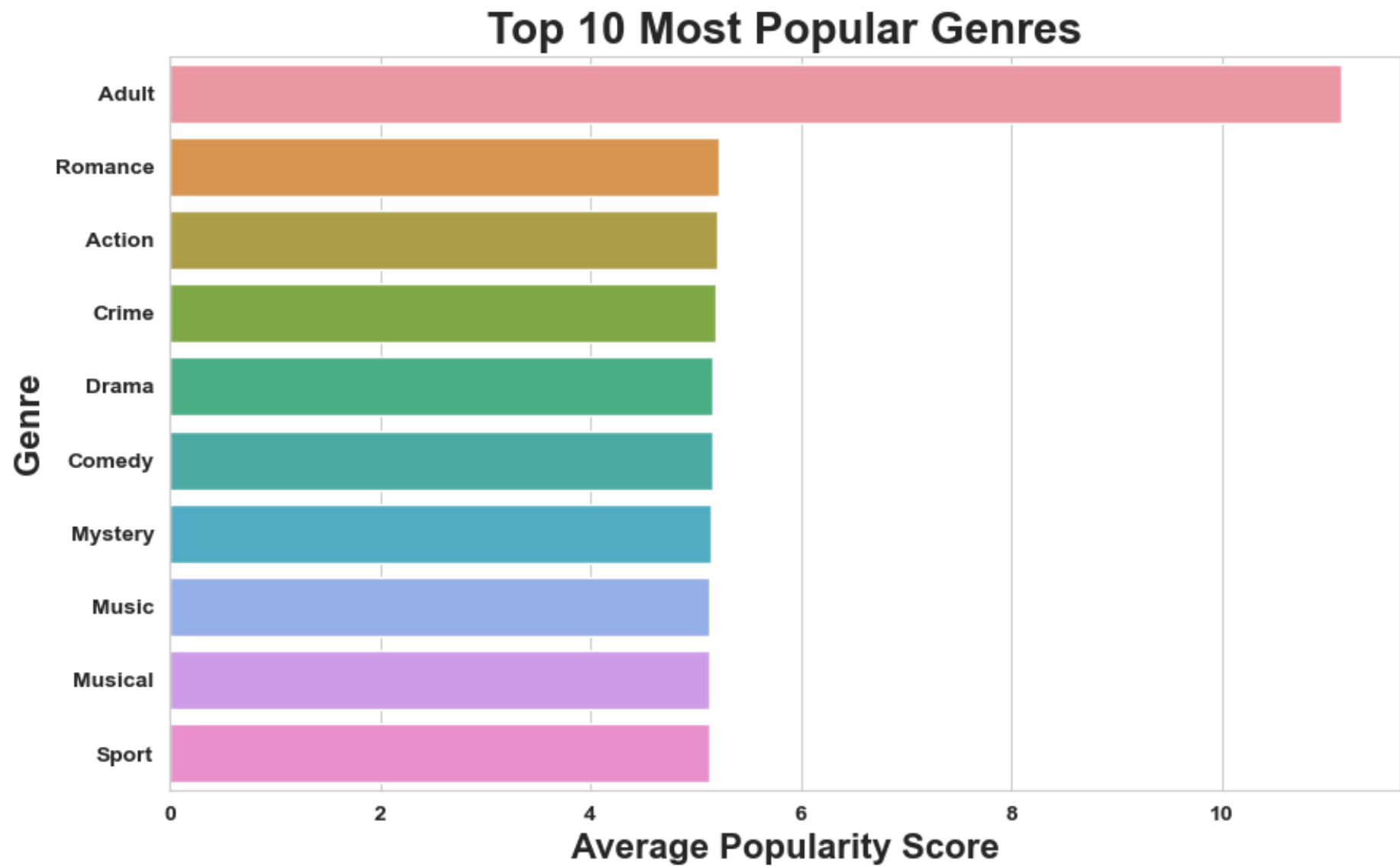
DATA EXPLORATION

With the help of Matplotlib and seaborn, I created some visualizations in order to understand the data better.

20XX

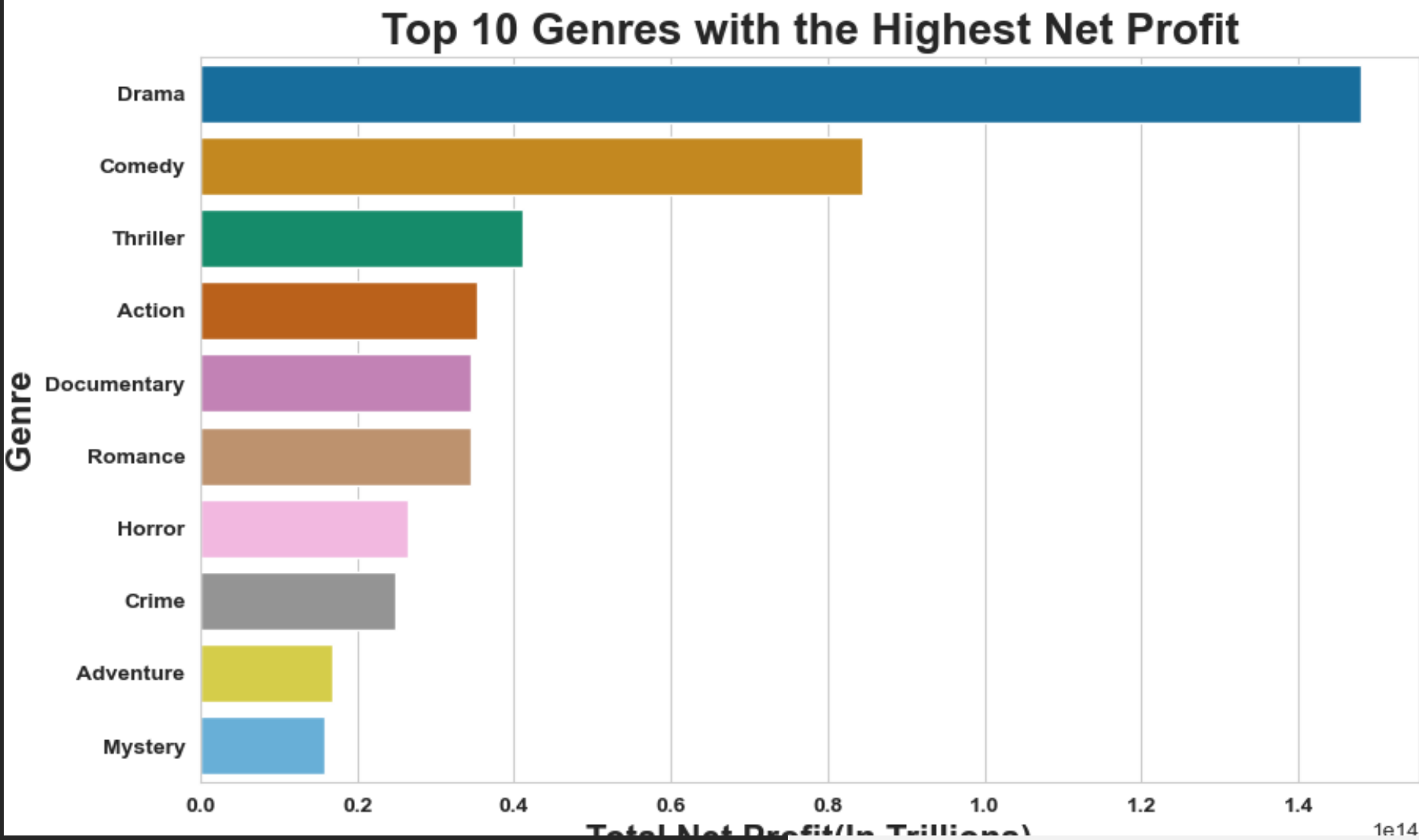


POPULARITY



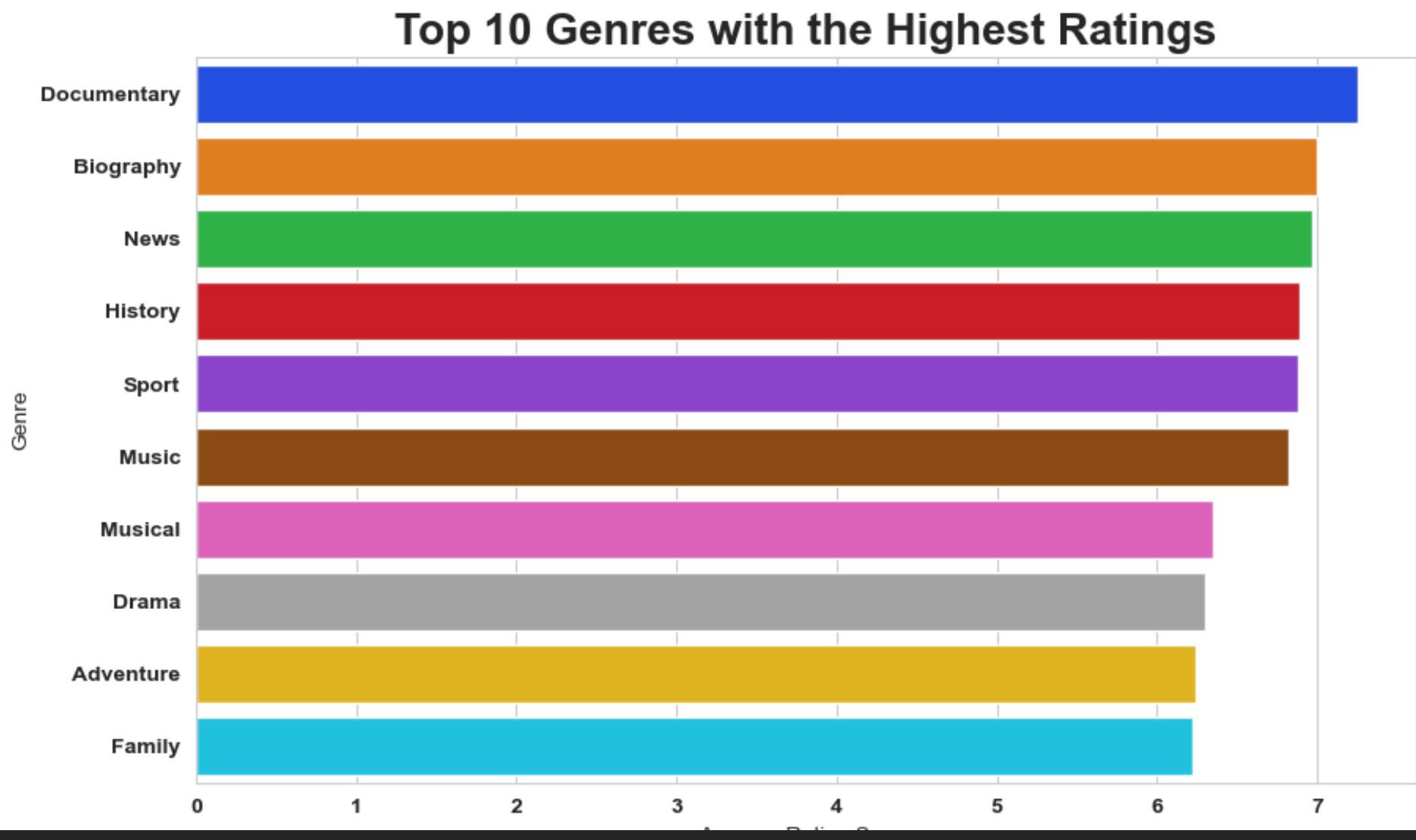


PROFITABILITY





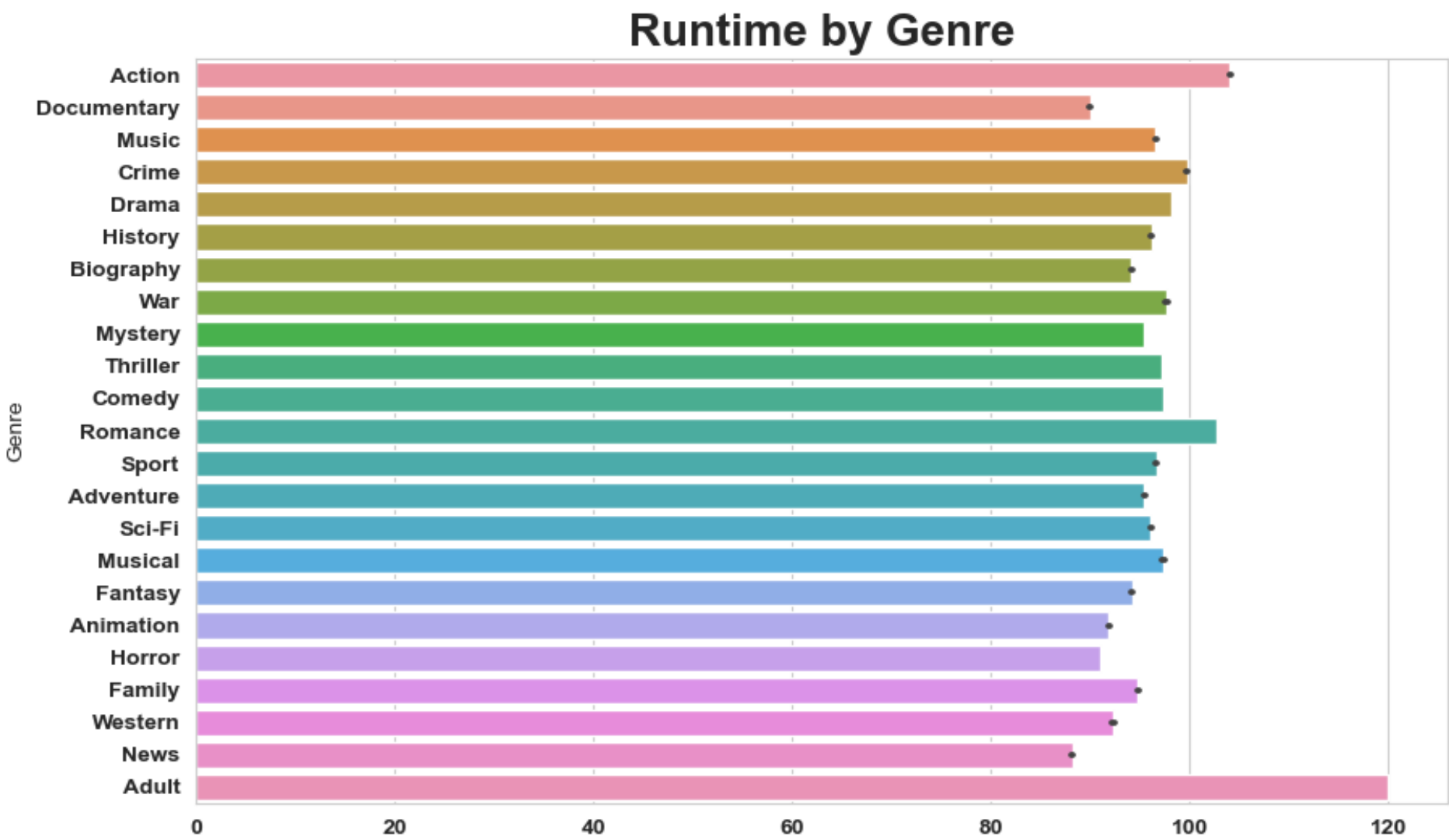
RATING





RUNTIME

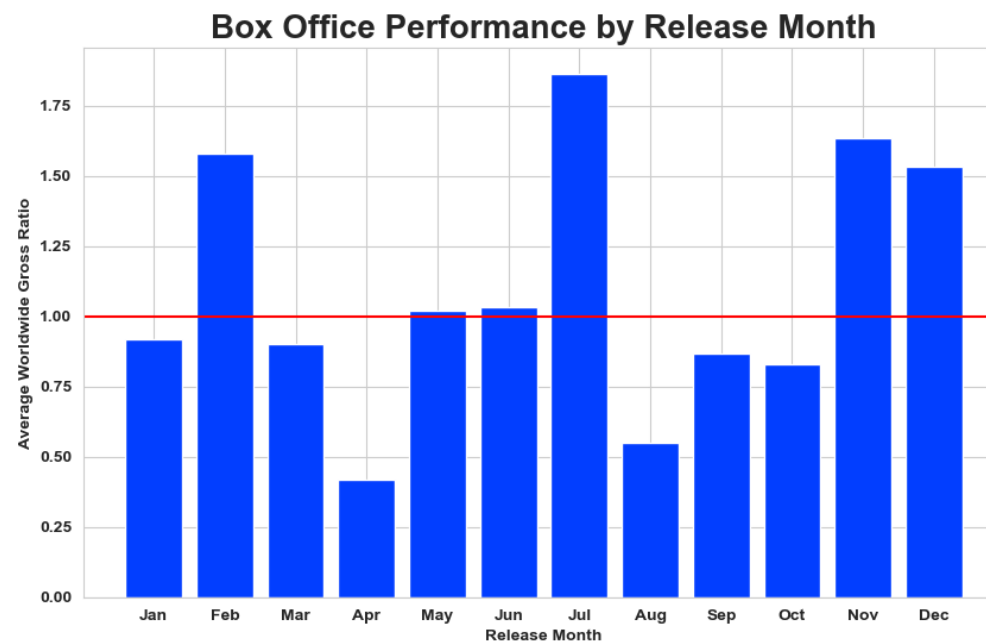






RELEASE MONTH

RELEASE MONTH

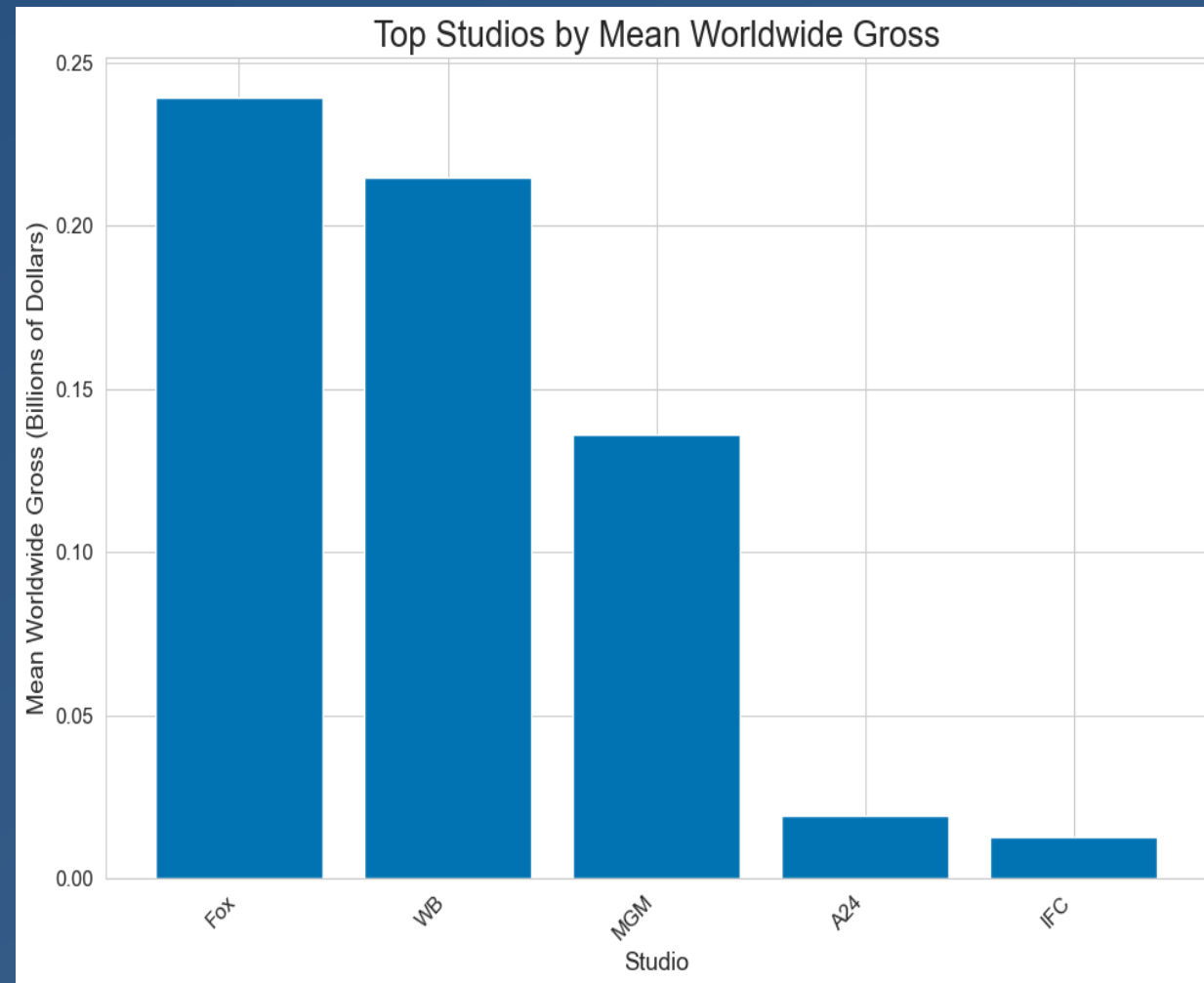




TOP STUDIOS

- Fox
- Warner Bros
- MGM

TOP STUDIOS





FINDINGS

- Drama, Comedy, and Documentary are the most popular genres, with the highest number of movies in each category.
- Action, Adventure, Crime, and Thriller are also popular genres, with a significant number of movies in each category.
- Adult, News, and Western are the least popular genres, with the lowest number of movies in each category.
- The mean rating for each genre is relatively consistent across all genres, ranging from 4.89 to 5.23.
- The maximum rating for each genre varies, with Comedy, Romance, and Drama having the highest maximum ratings.
- Drama genre has the highest net profit among the listed genres.
- Comedy and Thriller genres follow Drama in terms of net profit.
- Adventure and Mystery genres have comparatively lower net profits.
- Other genres such as Action, Documentary, Romance, Horror, and Crime also have significant net profits.



FINDINGS

- The genre with the highest average rating is Documentary, with a mean rating of 7.16
- Followed closely by Sport with a mean rating of 7.65.
- The lowest-rated genres are Adventure and Family, with mean ratings of 5.84 and 5.34, respectively.
- The minimum runtime is 80 minutes
- The maximum runtime is 145 minutes
- The average (mean) runtime is approximately 96.35 minutes
- The median runtime is 101 minutes.
- The standard deviation of the runtime is approximately 15.4 minutes indicating that there is some variability in the runtime of the movies.



FINDINGS

- The minimum worldwide gross is \$62,375.0
- The maximum worldwide gross is \$1,146,894,640.0.
- The average worldwide gross for all movies in the dataset is \$130,176,284.8
- The months of November, December, and June had the highest gross
- November has the highest gross among all months.



FINDINGS

- Fox has earned the most with over \$300 trillion in worldwide gross earnings
- Followed by WB with over \$58 trillion.
- MGM, IFC, and A24 round out the top 5 with over \$14 trillion each in earnings.



RECOMMENDATIONS

- Consider popular genres like action, adventure, and animation, but ensure a unique and compelling storyline.
- For profitability, focus on Drama, Comedy, and Thriller genres and avoid Adventure and Mystery genres.
- Family films have a relatively high median and mean runtime and are a good choice for a family audience.
- For high-rated films, consider Documentary and Sports genres, but balance ratings and profitability.
- For longer films, choose Biography or Fantasy genres, and for shorter films, choose Documentary, Horror, or Western genres.
- Release films during November, December, or June for higher revenue potential.
- Aim to work with top-performing studios such as Fox, WB, and MGM, as they have the highest overall worldwide gross.
- Keep the production budget in check as it has a significant impact on the net profit of a movie.





THANK YOU

Pauline Ndung'u

paulinendungu.nj@gmail.com



This Photo by Unknown author is licensed under [CC BY](#).