Budget vs. Box Office: ROI in the Film Industry

Author: Chris Kucewicz

Introduction and Contents

1. Business Understanding

2. Data Understanding

3. Data Preparation

4. Data Analysis

5. Limitations

6. Recommendations

7. Next Steps

Business Understanding

Background

Company looking to produce movies in order to diversify their revenue streams

Goals

Provide three recommendations on which movie features are the most *cost efficient*

Success Criteria

Define '*cost efficient*' to be measured using return on investment (ROI):

$$ROI = \frac{\text{Net profit}}{\text{Cost of investment}}$$

Data Understanding

IMDB's film database (8 tables total):

movie_basics:

- **6 features** with **146,144 entries**
- Each entry represents a different movie

persons:

- **5 features** with **606,648 entries**.
- Each entry represents a person who has worked in/on a movie

principals:

- **6 features** and **1,028,186 entries**
- Each entry represents a movie with a corresponding individual who was part of the movie

Data Understanding

The Numbers' *budgets* dataset:

- 6 features and 5,782 entries
- Each entry represents a different movie

Data Preparation

- Cleaned four datasets: *budgets*, *movie_basics*, *persons*, *principals*

- Converted columns to correct data types with help of custom *get_info()* function

- Filtered out outliers and irrelevant columns, and addressed NaN values

Data Preparation

- Created three new dataframes (*top_people_budgets*, *top_roi_movie_basics*, *budgets_no_outliers*)
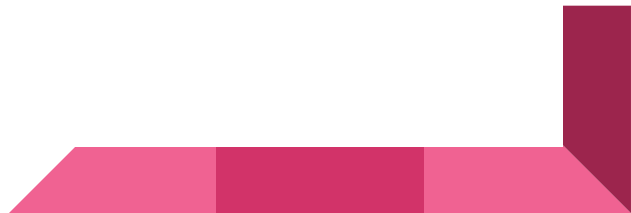
- Focused on movies with the top 25% ROIs

Data Analysis

- **37%** of all movies did *NOT* achieve **positive ROI**

- Typical movie:
    - **$16 million production budget**

    - **$26 million worldwide gross revenue**

    - **66% ROI**
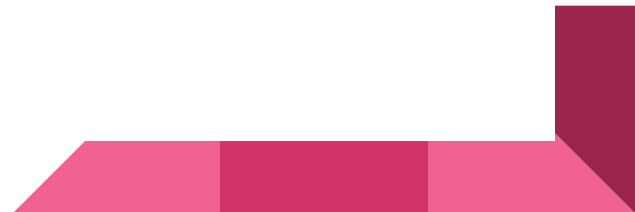
Data Analysis: Top 25% ROI Movies

- **More than 33%** of movies classified as: **Dramas** & **Comedies**

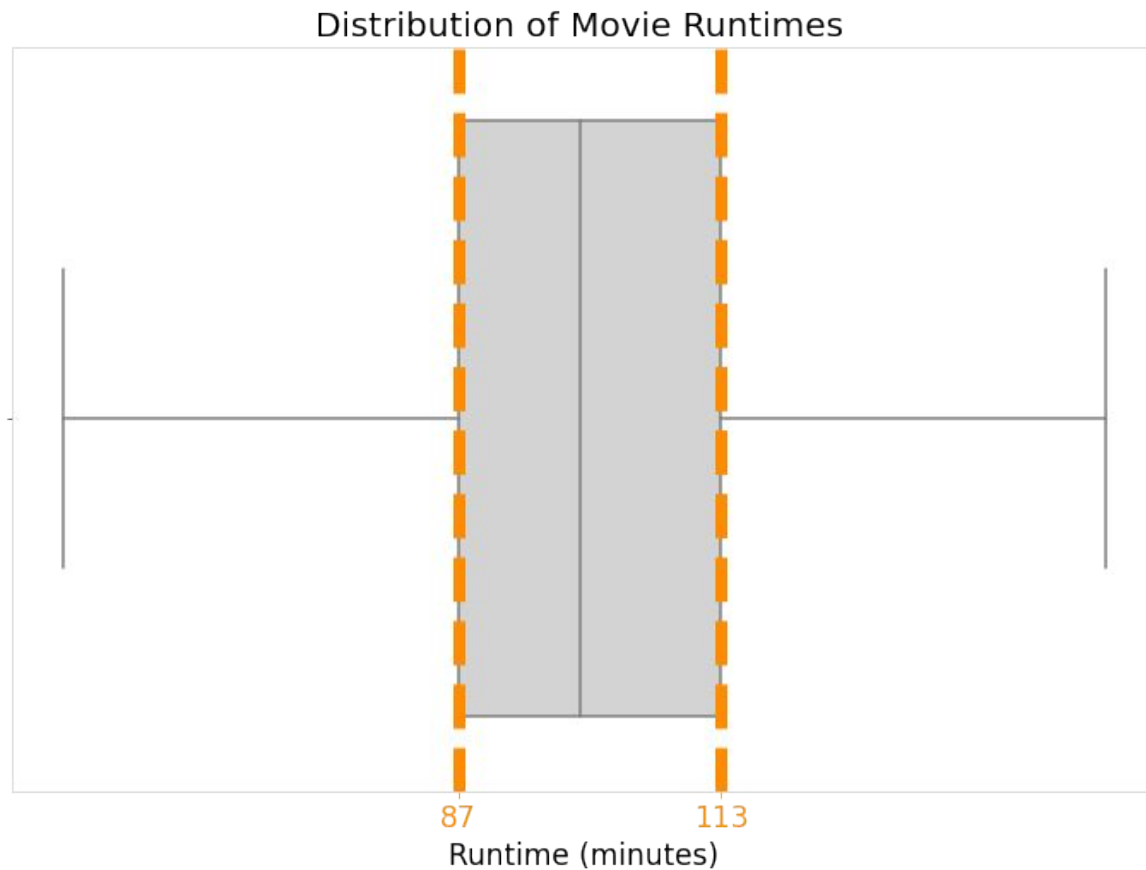- Middle 50% of movies had **runtimes** ranging **87** to **113 minutes**

Data Analysis: Top 25% ROI Movies

- Drama & comedy **professions** generating **highest ROI:**

  - **Composers**
    - Danny Elfman, Alexandre Desplat, Marco Beltrami, Thomas Newman, Theodore Shapiro

  - **Directors**
    - David O. Russell, Steven Spielberg, Damien Chazelle

  - **Producers**
    - Simon Kinberg, Michael De Luca, Dana Brunetti
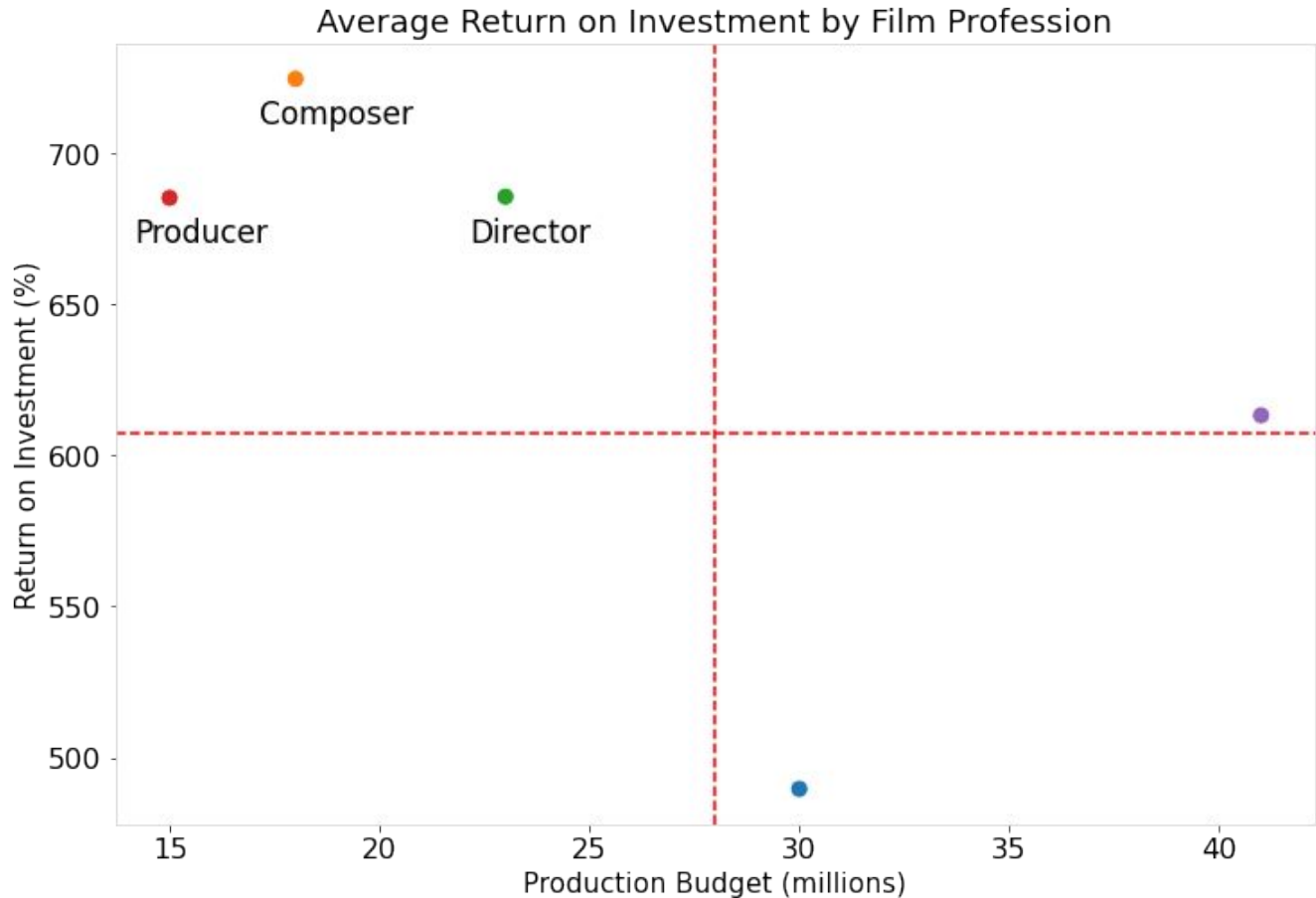
More than 33% of movies classified as **Dramas** & **comedies**

.

Movie Genres in the Top Quartile of ROI
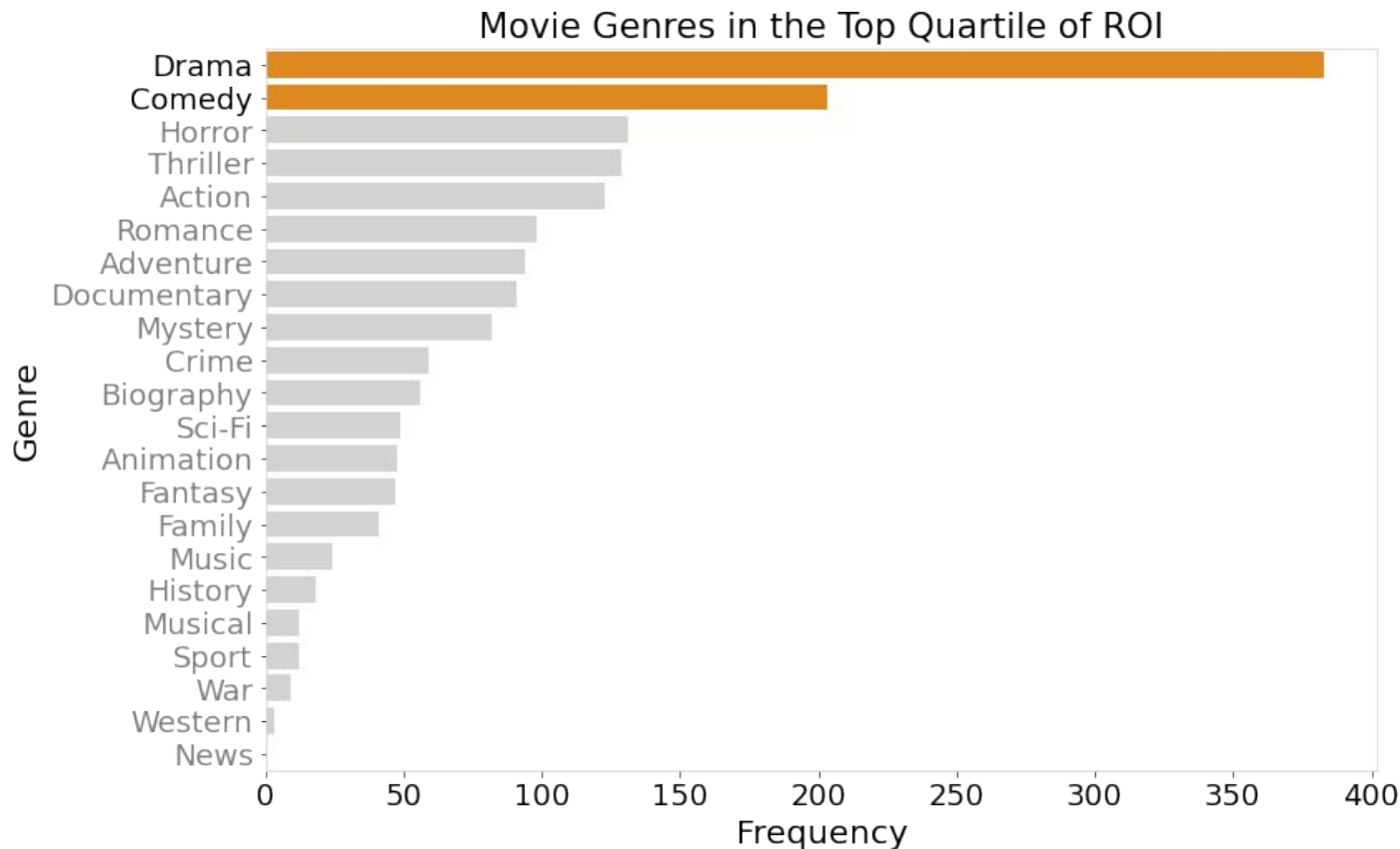
Runtimes between 87-113 minutes for the middle 50% of movies

Distribution of Movie Runtimes

87   113

Runtime (minutes)

Producers, writers, & directors generate the **highest ROI**

Average Return on Investment by Film Profession

Drama & comedy composers: Danny Elfman, Alexandre Desplat, Marco Beltrami, Thomas Newman, Theodore Shapiro

Total Gross Revenue Across Dramas & Comedies
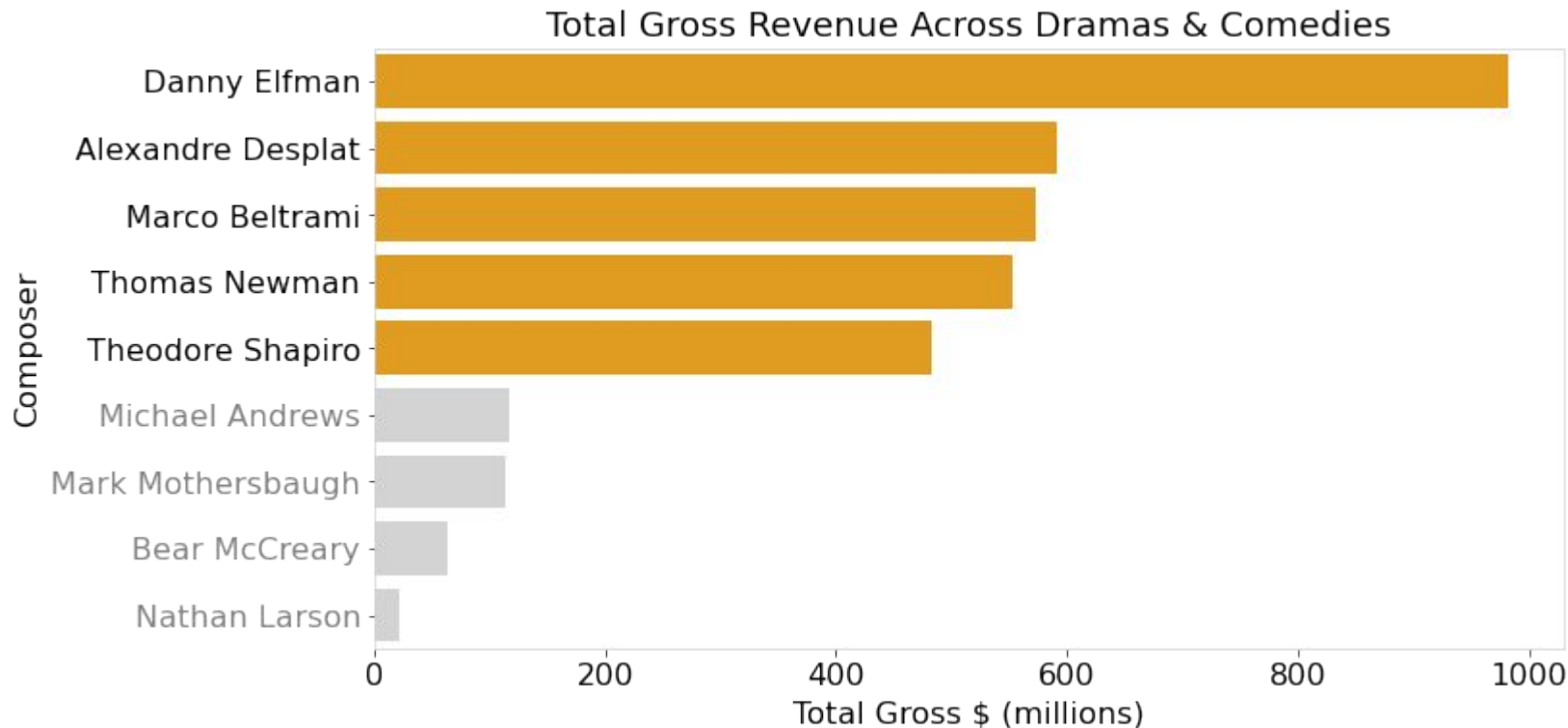
Drama & comedy **directors:** David O. Russell, Steven Spielberg, Damien Chazelle

Cumulative Gross Revenue Across Dramas & Comedies

Horizontal bar chart. Y-axis labeled "Director", X-axis labeled "Total Gross $ (millions)".

Director	Total Gross $ (millions)
David O. Russell	~620
Steven Spielberg	~480
Damien Chazelle	~465
Jon M. Chu	~265
M. Night Shyamalan	~245
Ryan Coogler	~190
Steven Soderbergh	~185
Timur Bekmambetov	~165
David F. Sandberg	~150
Judd Apatow	~145
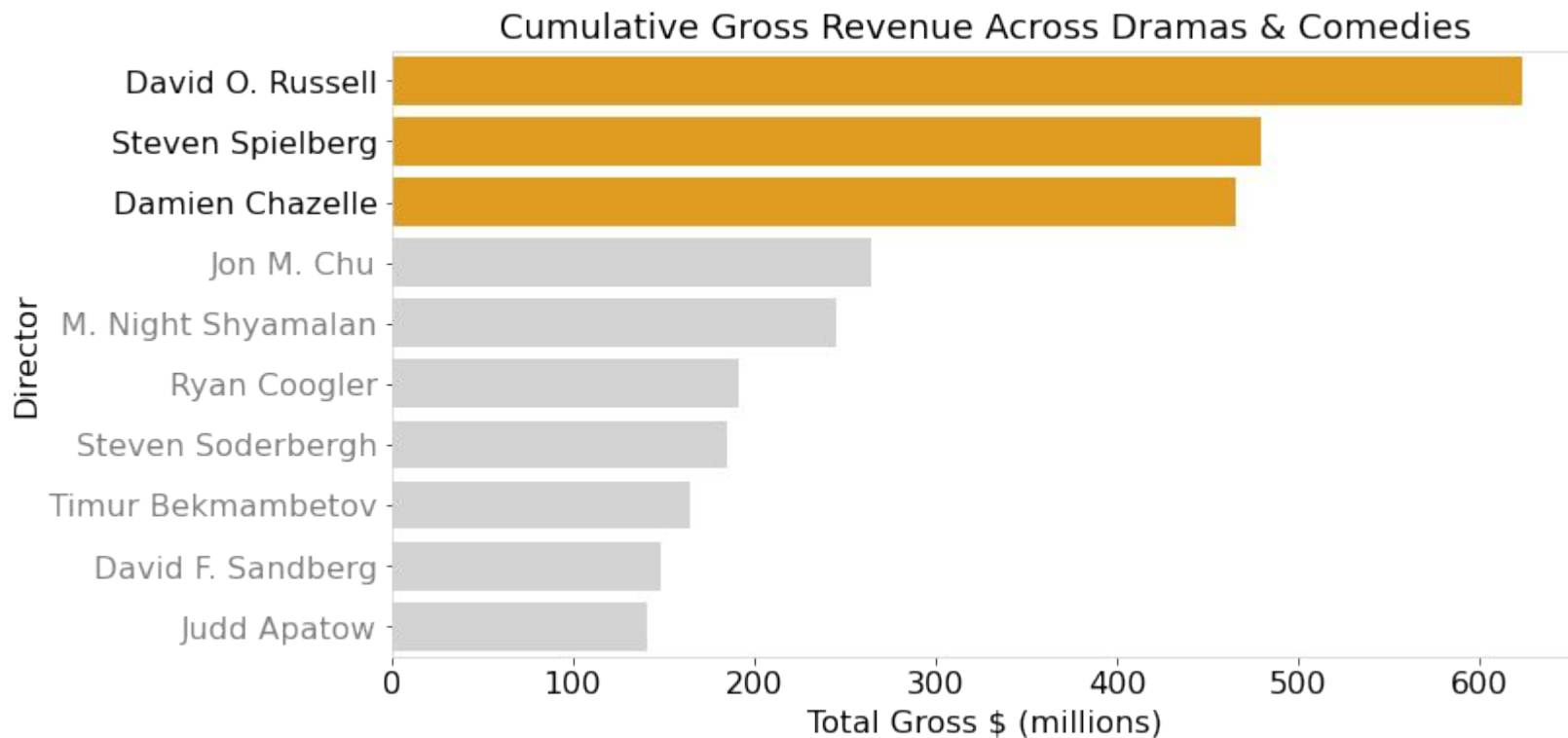
Drama & comedy **producers:** Simon Kinberg, Michael De Luca, Dana Brunetti

Cumulative Gross Revenue Across Dramas & Comedies
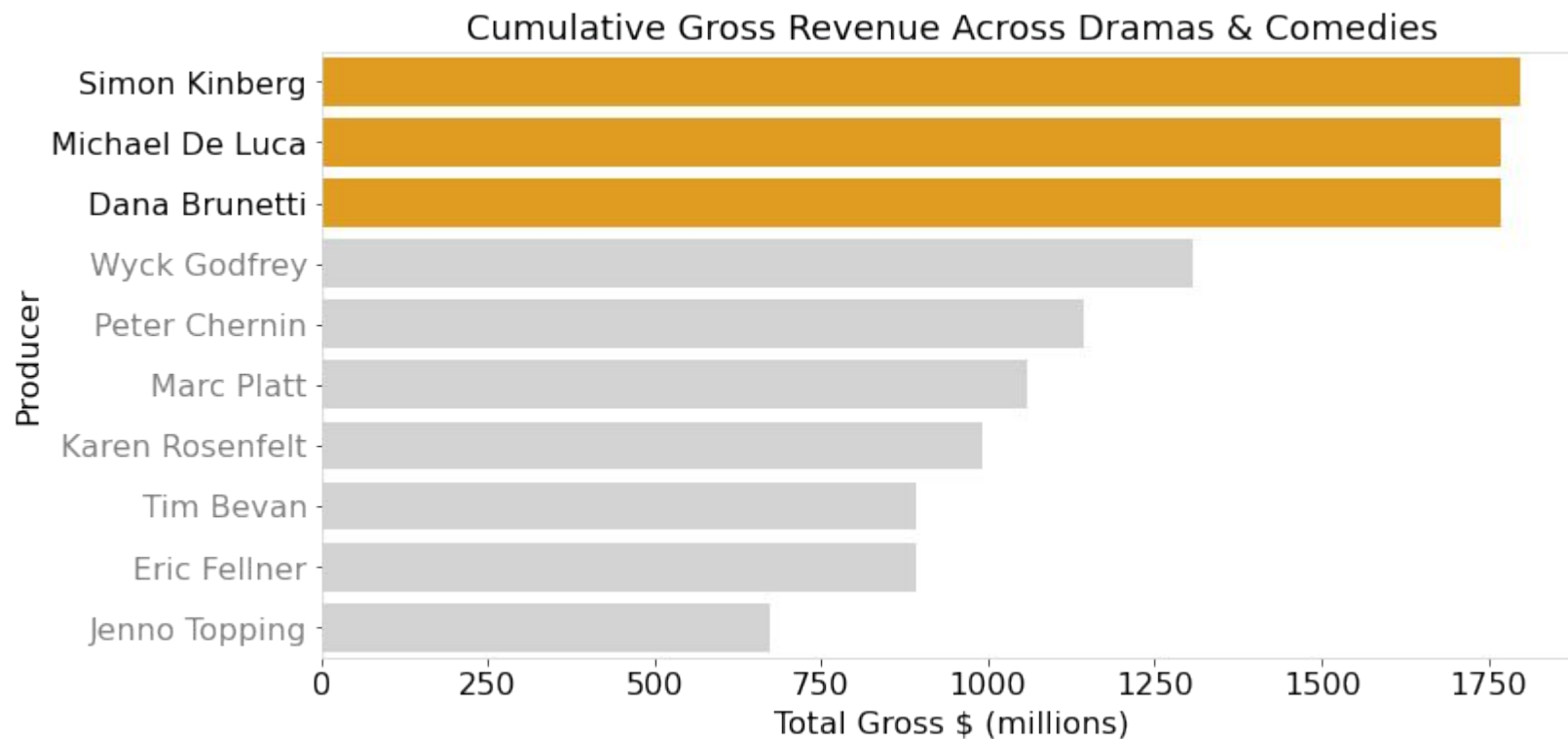
Limitations

- Limited amount of budget data = Less data for ROI analysis

Recommendations

| Recommendation 1: **Genre** | Recommendation 2: **Runtime** | Recommendation 3: **Hiring** |

Focus on creating:

- **Dramas**
- **Comedies**

Ensure movies are between **87** and **113** minutes long

Hire high-quality **composers, directors**, & **producers** specializing in **drama** & **comedy**

Next Steps

1. Moving forward I am interested in:

   a. Gathering budget data on a wider number of movies

   b. Performing regression analysis:
      i. Which factors most strongly correlate with movie ROI?

Thank you!

Github Repository:
https://github.com/ckucewicz/movie_analysis_project

Contact Chris Kucewicz at
cfkucewicz@gmail.com with additional questions