index

July 24, 2024

1 Movie Market Analysis

1.1 Define Objectives

1.1.1 Project Overview

The primary objective of this project is to explore and analyze trends in the movie industry to provide actionable insights for a new movie studio. This analysis aims to inform the studio's decision-making process regarding the types of films to produce, based on current market trends and box office performance.

1.1.2 Key Business Questions

To guide the analysis, we will focus on the following key business questions:

1. **What types of films are currently performing well at the box office?**
   - Analyze the box office performance of different genres to identify which types of films generate the highest revenue.
2. **Which factors are associated with higher box office revenue?**
   - Investigate various factors such as budget, release date, and runtime to determine their impact on box office success.
3. **What actionable strategies can be derived from the analysis to guide the studio's film production decisions?**
   - Provide concrete recommendations based on the data analysis to help the studio decide on the types of films to produce.

1.1.3 Data Sources

To answer these questions, we will utilize datasets from the following sources: - **Box Office Mojo**: Data on box office gross. - **IMDB**: Movie metadata, including genres, ratings, and more (stored in a SQLite database). - **Rotten Tomatoes**: Movie reviews and ratings. - **TheMovieDB**: Additional movie details and metadata. - **The Numbers**: Box office revenue data.

1.1.4 Expected Deliverables

The expected deliverables for this project are: 1. **Non-Technical Presentation**: A slide deck presenting the analysis and findings to business stakeholders. 2. **Jupyter Notebook**: A well-documented notebook that presents the analysis in a clear and organized manner, suitable for a data science audience. 3. **GitHub Repository**: A repository containing all project files, including the notebook, presentation, and a detailed README file.

By addressing these objectives and utilizing the provided datasets, we aim to deliver insights that will guide the new movie studio in making informed production decisions.

1.2 Data Acquisition

1.2.1 Overview

In this section, we will acquire and load the datasets needed for our analysis. The data comes from multiple sources, including Box Office Mojo, IMDB, Rotten Tomatoes, TheMovieDB, and The Numbers. We will first download and unzip the datasets, then load them into our Jupyter Notebook for further analysis.

1.2.2 Datasets

The primary datasets we will use are:

1. **Box Office Mojo**: Contains data on box office gross.
   - File: `data/bom.movie_gross.csv`
2. **IMDB**: Contains movie metadata including genres, ratings, and more.
   - File: `data/im.db`
   - Relevant tables: `movie_basics`, `movie_ratings`

1.2.3 Data Loading

We will use Python libraries such as `pandas` and `sqlite3` to load the datasets into dataframes for analysis.

```python
[5]: import pandas as pd
     import matplotlib.pyplot as plt
     import sqlite3
     import os
     %matplotlib inline
```

```python
[45]: bom_df = pd.read_csv('data/bom.movie_gross.csv')
      rtmi_df = pd.read_csv('data/rt.movie_info.tsv', delimiter='\t', index_col=0)
      rtr_df = pd.read_csv('data/rt.reviews.tsv', delimiter='\t', index_col=0,
        encoding='ISO-8859-1')
      tmdb_df = pd.read_csv('data/tmdb.movies.csv', index_col=0)
      tnmb_df = pd.read_csv('data/tn.movie_budgets.csv')
```

```python
[49]: conn = sqlite3.connect('data/im.db')
      pd.read_sql(
      """
      SELECT name
      FROM sqlite_master
      WHERE type='table';
      """, conn
      )
```

```
[49]:            name
      0    movie_basics
      1       directors
      2        known_for
      3      movie_akas
      4   movie_ratings
      5         persons
      6      principals
      7         writers
```

1.2.4 Data Overview

Box Office Mojo Data

```
[50]: bom_df.head()
```

```
[50]:                                               title studio  domestic_gross  \
      0                                       Toy Story 3     BV     415000000.0
      1                         Alice in Wonderland (2010)     BV     334200000.0
      2   Harry Potter and the Deathly Hallows Part 1     WB     296000000.0
      3                                         Inception     WB     292600000.0
      4                                 Shrek Forever After    P/DW    238700000.0

         foreign_gross  year
      0     652000000  2010
      1     691300000  2010
      2     664300000  2010
      3     535700000  2010
      4     513900000  2010
```

IMDB Data

IMDB Data: Movie Basics

```
[56]: pd.read_sql(
      """
      SELECT *
      FROM movie_basics
      """, conn
      ).head()
```

```
[56]:      movie_id                      primary_title                original_title  \
      0  tt0063540                          Sunghursh                     Sunghursh
      1  tt0066787   One Day Before the Rainy Season          Ashad Ka Ek Din
      2  tt0069049       The Other Side of the Wind  The Other Side of the Wind
      3  tt0069204                    Sabse Bada Sukh             Sabse Bada Sukh
      4  tt0100275         The Wandering Soap Opera       La Telenovela Errante
```

```
   start_year  runtime_minutes                 genres
0        2013            175.0      Action,Crime,Drama
1        2019            114.0         Biography,Drama
2        2018            122.0                   Drama
3        2018              NaN            Comedy,Drama
4        2017             80.0  Comedy,Drama,Fantasy
```

IMDB Data: Movie Ratings

```
[57]: pd.read_sql(
      """
      SELECT *
      FROM movie_ratings
      """, conn
      ).head()
```

```
[57]:     movie_id  averagerating  numvotes
      0  tt10356526            8.3        31
      1  tt10384606            8.9       559
      2   tt1042974            6.4        20
      3   tt1043726            4.2     50352
      4   tt1060240            6.5        21
```

Rotten Tomatoes

Rotten Tomatoes: Movie_Info

```
[58]: rtmi_df.head()
```

```
[58]:                                          synopsis rating  \
      id
      1   This gritty, fast-paced, and innovative police…      R
      3   New York City, not-too-distant-future: Eric Pa…      R
      5   Illeana Douglas delivers a superb performance …      R
      6   Michael Douglas runs afoul of a treacherous su…      R
      7                                               NaN     NR

                                    genre          director  \
      id
      1  Action and Adventure|Classics|Drama  William Friedkin
      3     Drama|Science Fiction and Fantasy  David Cronenberg
      5     Drama|Musical and Performing Arts    Allison Anders
      6             Drama|Mystery and Suspense     Barry Levinson
      7                        Drama|Romance    Rodney Bennett

                             writer  theater_date       dvd_date currency  \
      id
      1              Ernest Tidyman   Oct 9, 1971  Sep 25, 2001      NaN
```

```
3   David Cronenberg|Don DeLillo   Aug 17, 2012    Jan 1, 2013          $
5                  Allison Anders   Sep 13, 1996   Apr 18, 2000        NaN
6   Paul Attanasio|Michael Crichton   Dec 9, 1994   Aug 27, 1997        NaN
7                    Giles Cooper           NaN            NaN        NaN


    box_office       runtime              studio
id
1        NaN  104 minutes                 NaN
3    600,000  108 minutes   Entertainment One
5        NaN  116 minutes                 NaN
6        NaN  128 minutes                 NaN
7        NaN  200 minutes                 NaN
```

Rotten Tomatoes: Reviews

```
[59]:  rtr_df.head()
```

```
[59]:                                        review rating   fresh  \
    id
    3   A distinctly gallows take on contemporary fina…    3/5   fresh
    3   It's an allegory in search of a meaning that n…    NaN  rotten
    3   … life lived in a bubble in financial dealin…     NaN   fresh
    3   Continuing along a line introduced in last yea…   NaN   fresh
    3             … a perverse twist on neorealism…       NaN   fresh

              critic  top_critic          publisher               date
    id
    3       PJ Nabarro           0   Patrick Nabarro   November 10, 2018
    3   Annalee Newitz           0           io9.com        May 23, 2018
    3    Sean Axmaker           0  Stream on Demand    January 4, 2018
    3   Daniel Kasman           0              MUBI   November 16, 2017
    3             NaN           0      Cinema Scope    October 12, 2017
```

TheMovieDB

```
[60]:  tmdb_df.head()
```

```
[60]:            genre_ids     id original_language  \
    0      [12, 14, 10751]  12444                en
    1  [14, 12, 16, 10751]  10191                en
    2         [12, 28, 878]  10138                en
    3         [16, 35, 10751]    862                en
    4         [28, 878, 12]  27205                en


                               original_title  popularity release_date  \
    0  Harry Potter and the Deathly Hallows: Part 1     33.533   2010-11-19
    1                       How to Train Your Dragon     28.734   2010-03-26
    2                                     Iron Man 2     28.515   2010-05-07
```

```
3                                             Toy Story    28.005  1995-11-22
4                                             Inception    27.920  2010-07-16


                                            title  vote_average  vote_count
0  Harry Potter and the Deathly Hallows: Part 1           7.7       10788
1                     How to Train Your Dragon           7.7        7610
2                                  Iron Man 2           6.8       12368
3                                   Toy Story           7.9       10174
4                                   Inception           8.3       22186
```

The Numbers

```
[61]: tnmb_df.head()
```

```
[61]:    id   release_date                                  movie  \
      0   1   Dec 18, 2009                                 Avatar
      1   2   May 20, 2011  Pirates of the Caribbean: On Stranger Tides
      2   3    Jun 7, 2019                            Dark Phoenix
      3   4    May 1, 2015                 Avengers: Age of Ultron
      4   5   Dec 15, 2017       Star Wars Ep. VIII: The Last Jedi


         production_budget  domestic_gross  worldwide_gross
      0       $425,000,000    $760,507,625   $2,776,345,279
      1       $410,600,000    $241,063,875   $1,045,663,875
      2       $350,000,000     $42,762,350     $149,762,350
      3       $330,600,000    $459,005,868   $1,403,013,963
      4       $317,000,000    $620,181,382   $1,316,721,747
```

```
[ ]:
```