



Search or jump to...



[Pull requests](#) [Issues](#) [Marketplace](#) [Explore](#)



[AndreiMklv](#) / [Phase_I_Final_Project](#) Public

[Pin](#)

[Unwatch](#) **1**

[Fork](#) **0**

[Star](#) **0**

[Code](#) [Issues](#) [Pull requests](#) [Actions](#) [Projects](#) [Wiki](#) [Security](#) [Insights](#) [Settings](#)

[main](#) **1** branch **0** tags

[Go to file](#)

[Add file](#)

[Code](#)

About



No description, website, or topics provided.

- [Readme](#)
- [View license](#)
- 0** stars
- 1** watching
- 0** forks

Releases

No releases published
[Create a new release](#)

Packages

No packages published
[Publish your first package](#)

	AndreiMklv Final version	784d1f5 2 minutes ago	🕒 1 commit
📁	images	Final version	2 minutes ago
📄	20200727_CanadaFilmRebounds.jpg	Final version	2 minutes ago
📄	Action.png	Final version	2 minutes ago
📄	Animation.png	Final version	2 minutes ago
📄	CONTRIBUTING.md	Final version	2 minutes ago
📄	LICENSE.md	Final version	2 minutes ago
📄	README.md	Final version	2 minutes ago
📄	Sc-Fi.png	Final version	2 minutes ago
📄	action.xlsx	Final version	2 minutes ago
📄	animation.xlsx	Final version	2 minutes ago
📄	awesome.gif	Final version	2 minutes ago
📄	presentation.pdf	Final version	2 minutes ago
📄	project.ipynb	Final version	2 minutes ago
📄	project.pdf	Final version	2 minutes ago
📄	scifi.xlsx	Final version	2 minutes ago

[README.md](#)



Phase 1 Project Description

Overview

During this project, we will be investigating movie datasets using exploratory analysis for the customer, who is new to the movie industry. For the sake of research, we will use the following datasets (with formats):

- The Internet Movie DataBase (IDMB), SQL
- Rotten Tomatoes, TSV
- The Movie DataBase (TMDB), CSV
- The Numbers (TN), CSV
- Box Office Mojo (BOM), CSV We will employ Pandas, SQLITE3, MatPlotLib, and SeaBorn libraries.

Considering the customer strategy 'Comprehensive Products Revolutionising Services Across Globe' we choose worldwide box office as the primary metric for measuring films' success as it reflects global audience coverage most unambiguously in the given datasets.

We will research box offices by month of a film release, box office by genres in two different datasets, trending genres, ROI by genres, and top producers by box offices of films they took part in for the 2010-2018 years. We choose the latter assuming that the customer may want to consult or hire subject matter experts.

Business Problem

The customer wants to expand and create a movie studio, and the main challenge is that they don't have any expertise in the industry. Our primary goals in this project are generating insight and making further recommendations to the decision-makers. We assume that before investing in the movie studio, the customer may want to know which genres are trending globally and what are their cummlative box offices, or, in other words, what is the industry size by genres. We also think that the customer, who wants to produce movies by themself, may wish to hire subject matter experts (producers) to reduce potential risks accompanied by entering a highly-competitive industry.

Methods

For the analysis we used descriptive statistics. The following libraries also been used.

- Pandas as pd
- Sqlite3
- Matplotlib
- String
- Seaborn
- Numpy

Findings

Conclusions

After reviewing and analyzing the data, we can make the following suggestions.

Genres

The genres with the maximum median box offices and return of investments are:

- Action and Adventure;
- Sci-Fi;
- Animation.

The most trending genres are also Sci-Fi, Action, and Animation. The 'News' genre shows the maximum median ROI, but the box office is relatively low.

Release months

According to the maximum box offices by the month of a movie release, the top 3 months are:

- December;
- February;
- April.

Whereas the top worst are:

- August;
- September;
- January.

Repository Structure

├── README.md <- The top-level README for reviewers of this project ├── project.ipynb <- Narrative documentation of analysis in Jupyter notebook ├── presentation.pdf <- PDF version of project presentation ├── data <- Compressed datasets (must be unzipped before using) ├── images <- Both sourced externally and generated from code └── project.pdf <- PDF version of the project source code



© 2022 GitHub, Inc.

[Terms](#)

[Privacy](#)

[Security](#)

[Status](#)

[Docs](#)

[Contact GitHub](#)

[Pricing](#)

[API](#)

[Training](#)

[Blog](#)

[About](#)