

main

Go to file

Add file

Code

About



tjkyner post-review edits ...

1 minute ago 31

data	added final merged df export	3 days ago
images	post-review edits	1 minute ago
notebooks	post-review edits	1 minute ago
submissions	post-review edits	1 minute ago
.gitignore	added .gitignore file	5 days ago
LICENSE	Initial commit	7 days ago
README.md	post-review edits	1 minute ago

README.md



Microsoft Studios - Movies Analysis

Student name: T.J. Kyner
Student pace: Full time
Instructor: Abhineet Kulkarni
Cohort: 040521

The Business Problem

The goal of this project is to conduct an exploratory data analysis in response to the following business problem:

Microsoft sees all the big companies creating original video content and they want to get in on the fun. They have decided to create a new movie studio, but they don't know anything about creating movies. You are charged with exploring what types of films are currently doing the best at the box office. You must then translate those findings into actionable insights that the head of Microsoft's new movie studio can use to help decide what type of films to create.

Repository Structure

```
├── data/
│   ├── cleaned/    # Cleaned csv files from `preprocessing.ipynb`
│   └── ...          # The raw csv files are housed directly under the `data` directory
├── images/         # Contains exported images of plots and illustrations used in the presentation
├── notebooks/      # Contains the `preprocessing.ipynb` and `analysis.ipynb` notebooks
├── submissions/    # Contains files used for the project submissions (presentation.pdf, etc.)
├── .gitignore
├── LICENSE
└── README.md
```

Data Preprocessing

Cursory data exploration and manipulation occurs in `preprocessing.ipynb`. The information included in each of the available data files is reviewed before settling on several for further exploration:

- `imdb.title.basics.csv`
- `imdb.title.ratings.csv`

This repository contains my phase 1 project for Flatiron School's data science program.

Readme

GPL-3.0 License

Releases

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

Packages

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

Languages

Jupyter Notebook 100.0%

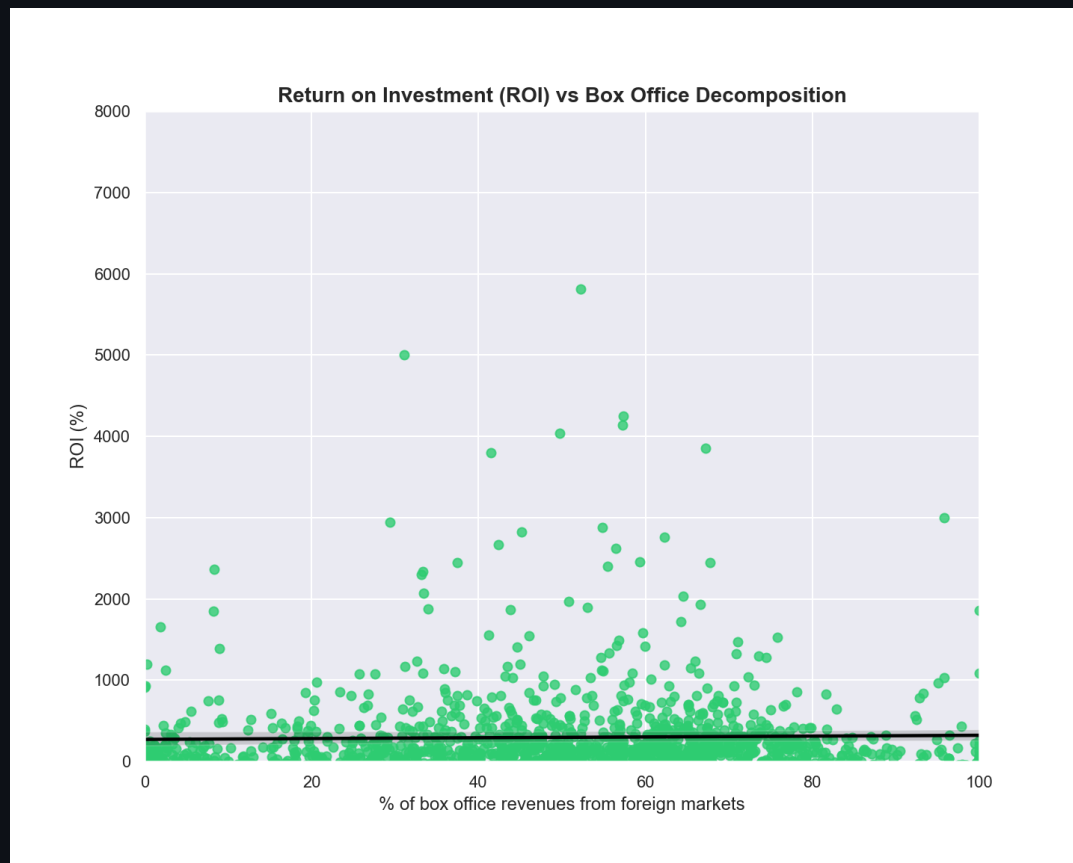
- `rt.movie_info.tsv`
- `tn.movie_budgets.csv`

Each of these files are loaded into dataframes where summary statistics and information are then performed. Notes are taken on any potential outliers, data type inconsistencies, etc. Based on these observations and notes, the dataframes are cleaned accordingly. Finally, the dataframes are consolidated and exported for further use in the `analysis.ipynb` notebook.

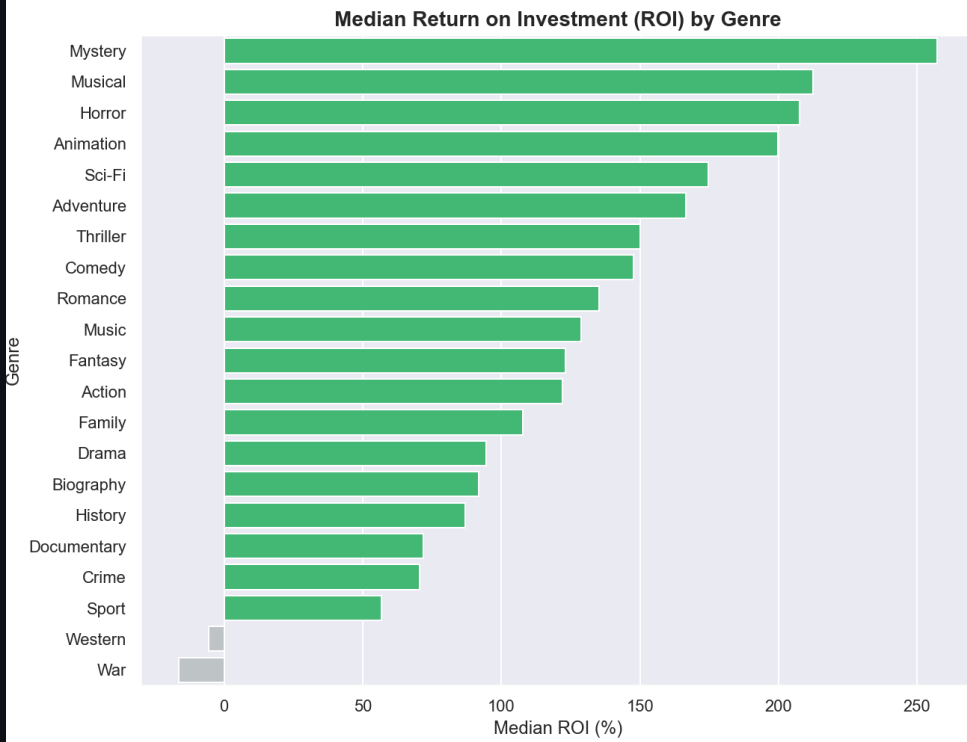
Analysis

This notebook revolves around the asking and answering of questions regarding the consolidated dataframe exported in `preprocessing.ipynb`. The answers to the following questions help form and provide evidence for my ultimate recommendation.

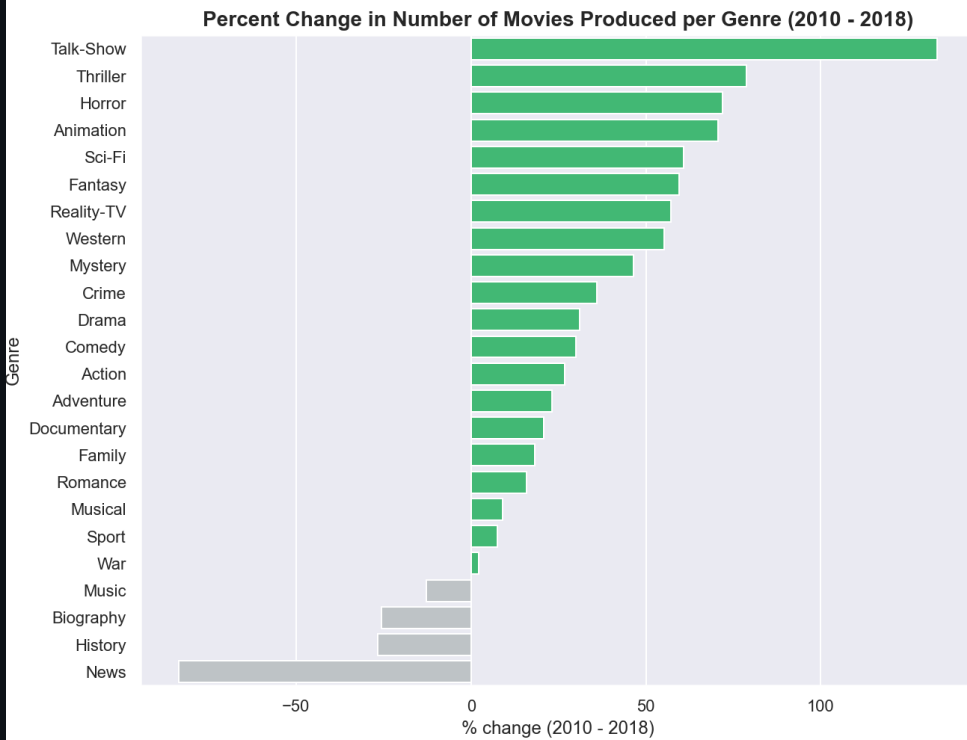
Question 1: How important are foreign box office results to the overall return on investment?



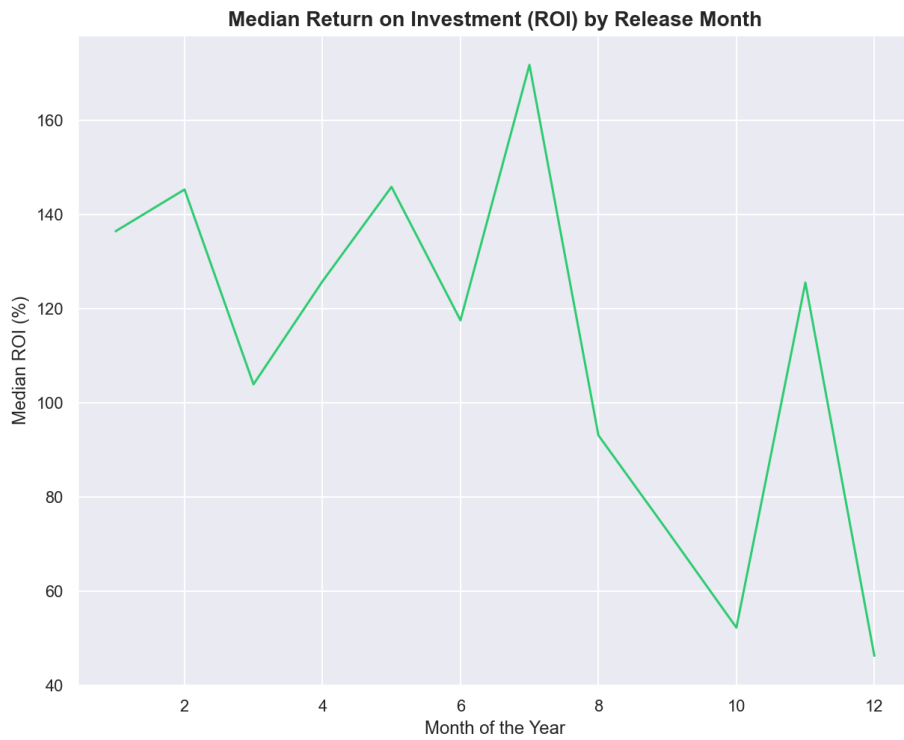
Question 2: Which genres have the highest median return on investment?



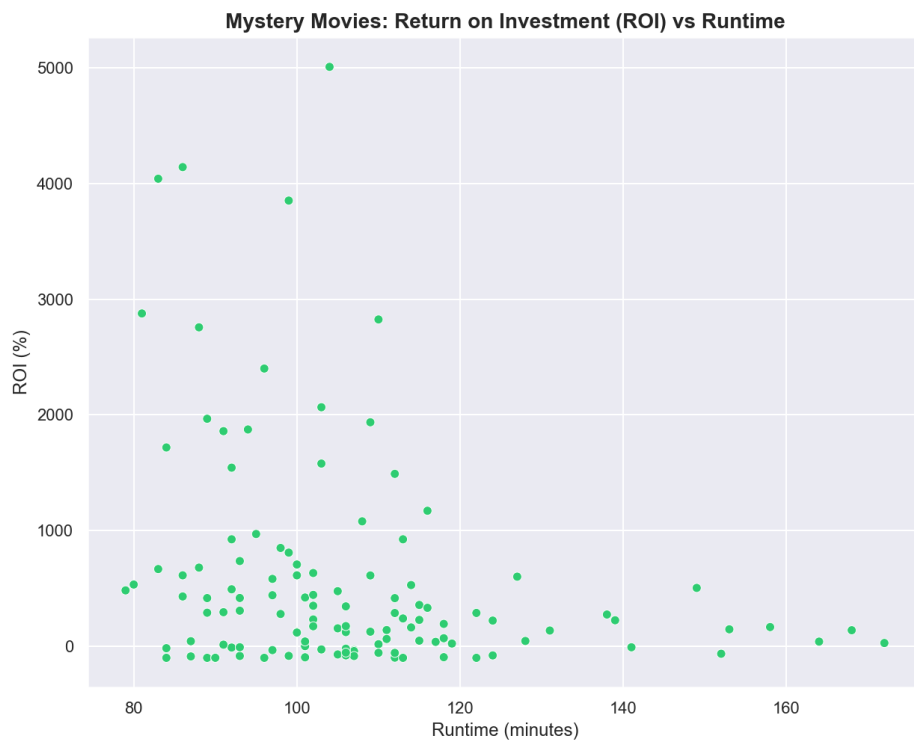
Question 3: Are there any noticeable trends in the number of movies produced in a particular genre?



Question 4: What months have the highest median return on investment?



Question 5: Does the runtime of a movie in the Mystery genre affect its ROI?



Conclusion & Recommendation

To maximize the potential return on investment, Microsoft Studios should produce a movie with the following characteristics:

- Mystery genre
- Runtime under 2 hours

- Released in May or July

The data supports this recommendation since the Mystery genre has the highest median ROI of any genre, is increasing in popularity, and is most profitable for runtimes under two hours. Furthermore, targeting a release date in May or July increases the probability of success given the higher median ROIs for movies released in those two months.

Next Steps

Next steps in the analysis include exploring what makes a good Mystery movie and answering the following questions:

- Are there any specific directors or writers that excel in the Mystery genre?
- What should the production budget be?
- Can box office data be found for movies older than 2010?