



 tjkyner / flatiron-ds-phase-1-project

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


 GPL-3.0 License

 0 stars  0 forks

 Star

 Unwatch ▼

 Code

 Issues

 Pull requests

 Actions

 Projects

 Wiki

 Security

 main ▼

...



tjkyner restructured to submissions folder ...

2 minutes ago

 29

[View code](#)

 README.md



# Microsoft Studios - Movies Analysis

Student name: T.J. Kyner

Student pace: Full time

Instructor: Abhineet Kulkarni

Cohort: 040521

## The Task

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:

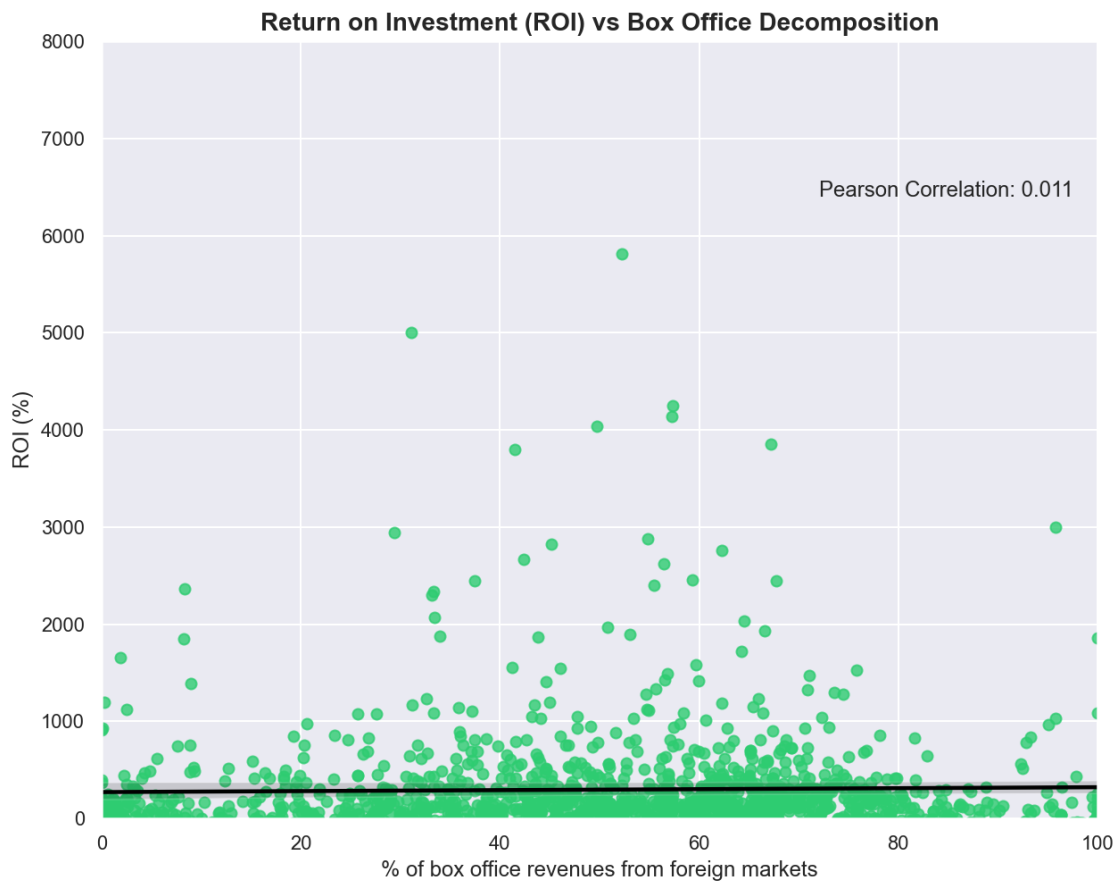
- `imdb.title.basics.csv`
- `imdb.title.ratings.csv`
- `rt.movie_info.tsv`
- `tn.movie_budgets.csv`

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.

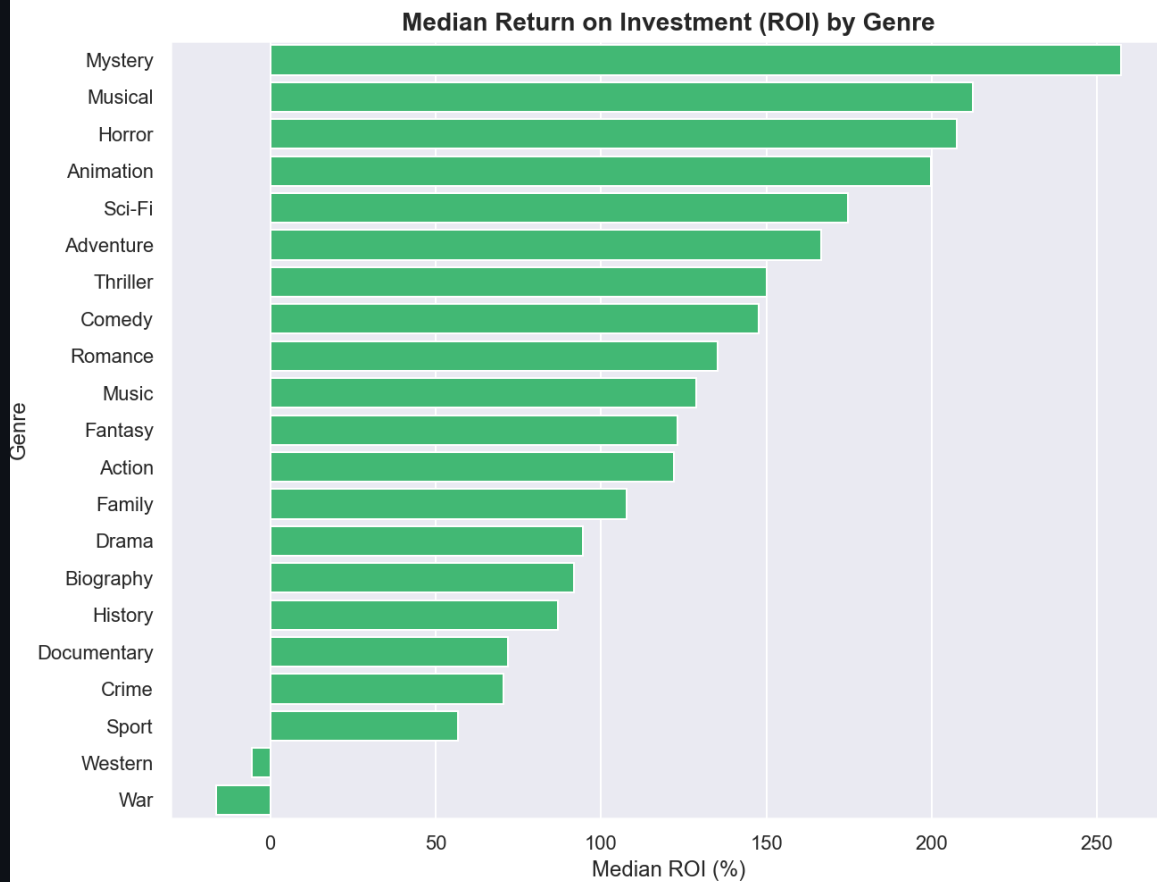
## Analysis

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.

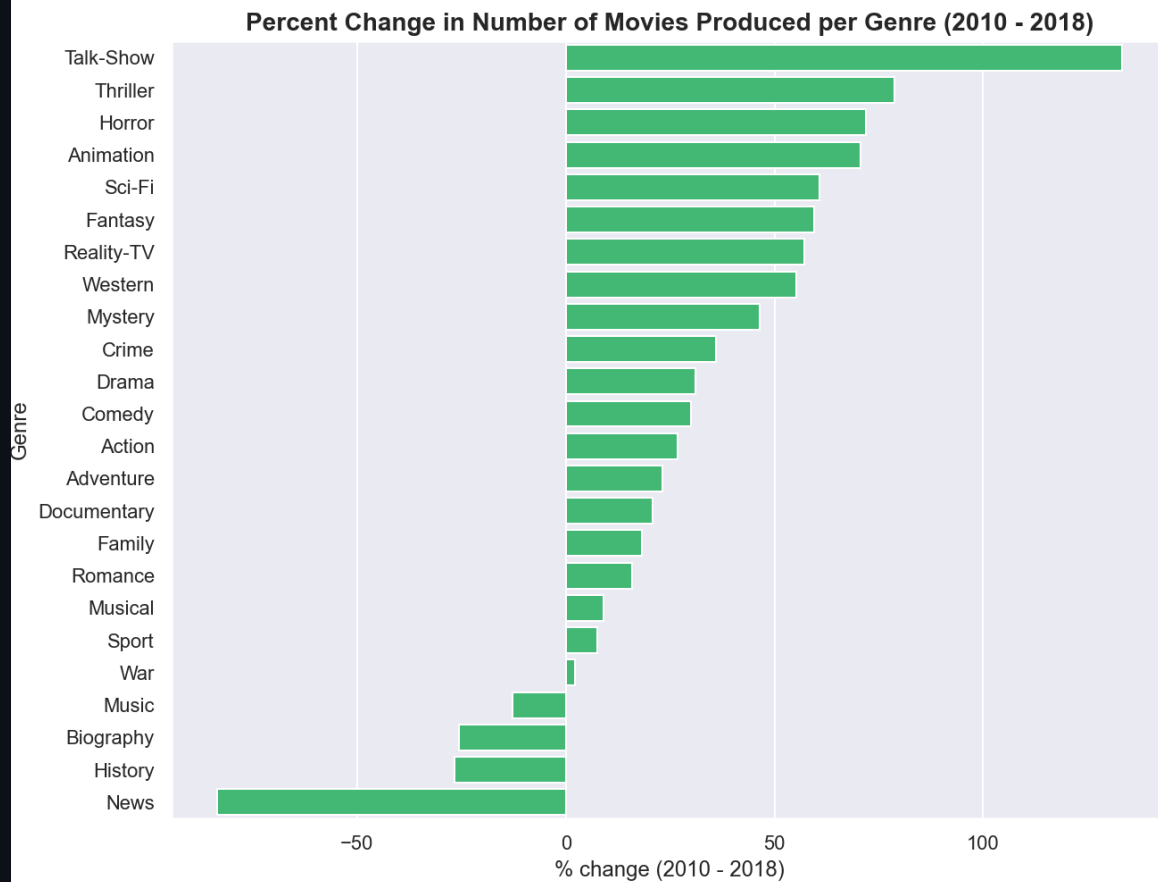
**Question 1: How important are foreign box office results to the overall return on investment?**



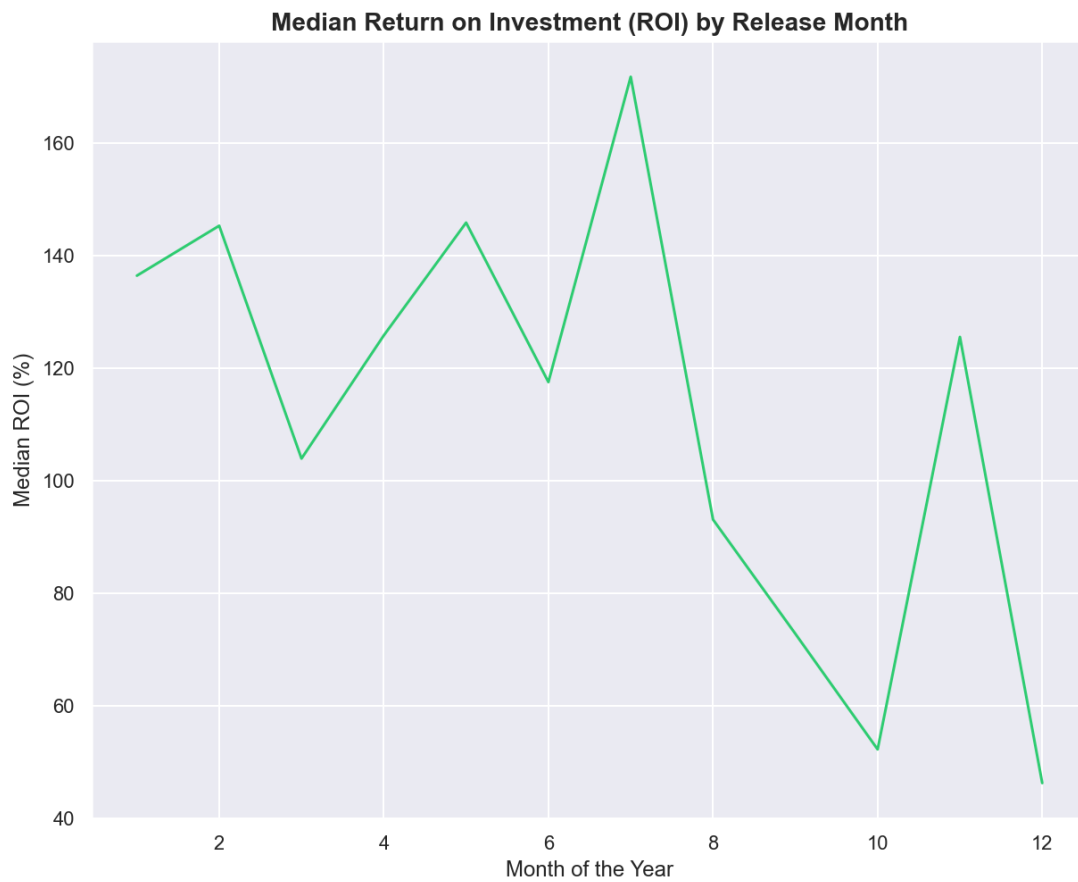
**Question 2: Which genres have the highest median return on investment?**



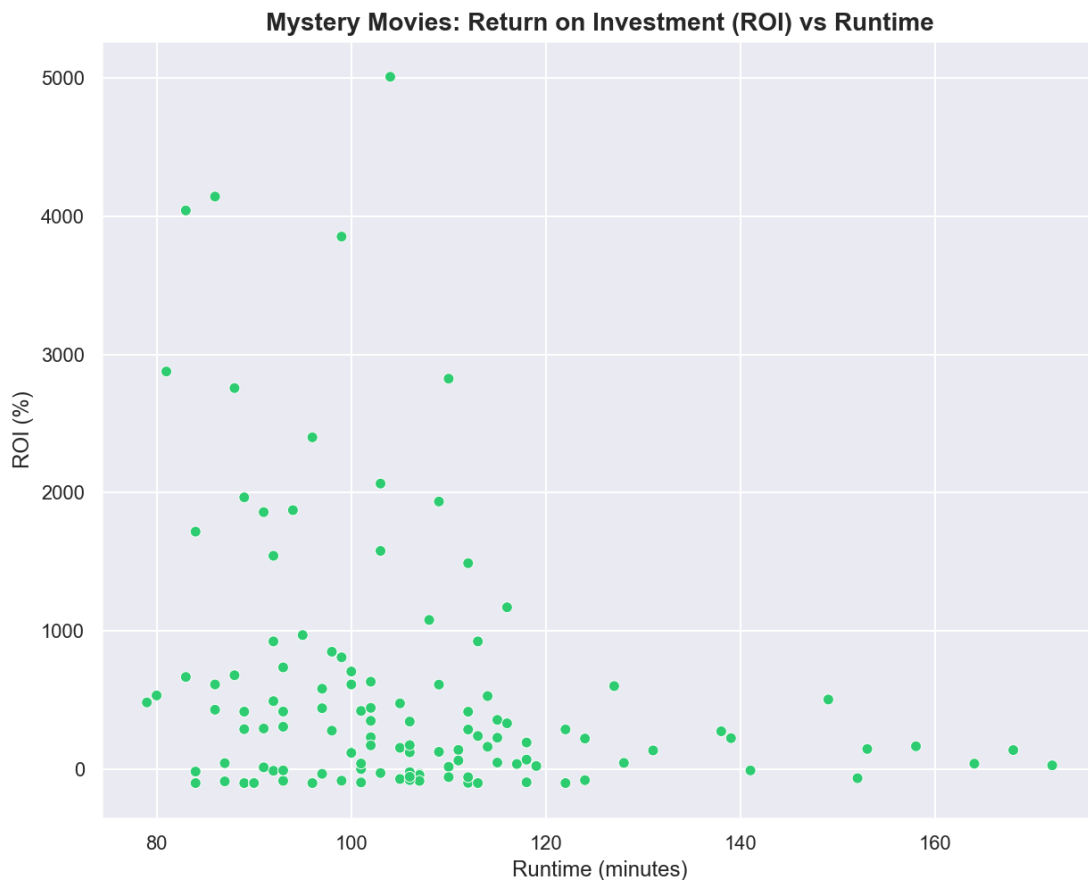
**Question 3: Are there any noticeable trends in the number of movies produced in a particular genre?**



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?**



## Recommendations

Based on the data explored thus far, Microsoft Studios should produce a movie in the **Mystery** genre with a runtime **under 2 hours** and an anticipated release date in **May or July**.

## 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?

No releases published

[Create a new release](#)

---

## Packages

No packages published

[Publish your first package](#)

---

## Languages



● Jupyter Notebook 100.0%