 gthein98 / ai-capstone

Public

☆ 0 stars

🍴 1 fork

☆ Star

▼

👁 Unwatch ▼

<> Code

🔍 Issues

🔗 Pull requests

🎬 Actions

📁 Projects


📖 Wiki

🛡 Security

📈 Insights

🔗 master ▼

...




Your master branch isn't protected

Protect this branch from force pushing, deletion, or require status checks before merging.

Protect this branch

×

 racday Adding PDFs ...

1 minute ago 🕒 77

View code

☰ README.md

✎

Semester 1 Capstone Project

Project Overview

In this project we performed exploratory data analysis and used statistical methods to provide business insights to Computing Visions for recommendations on movie performance based off available data. We provided three different recommendations, including the movie production budget, movie genre, and movie release month.

Business Understanding

Stakeholder - Computing Vision

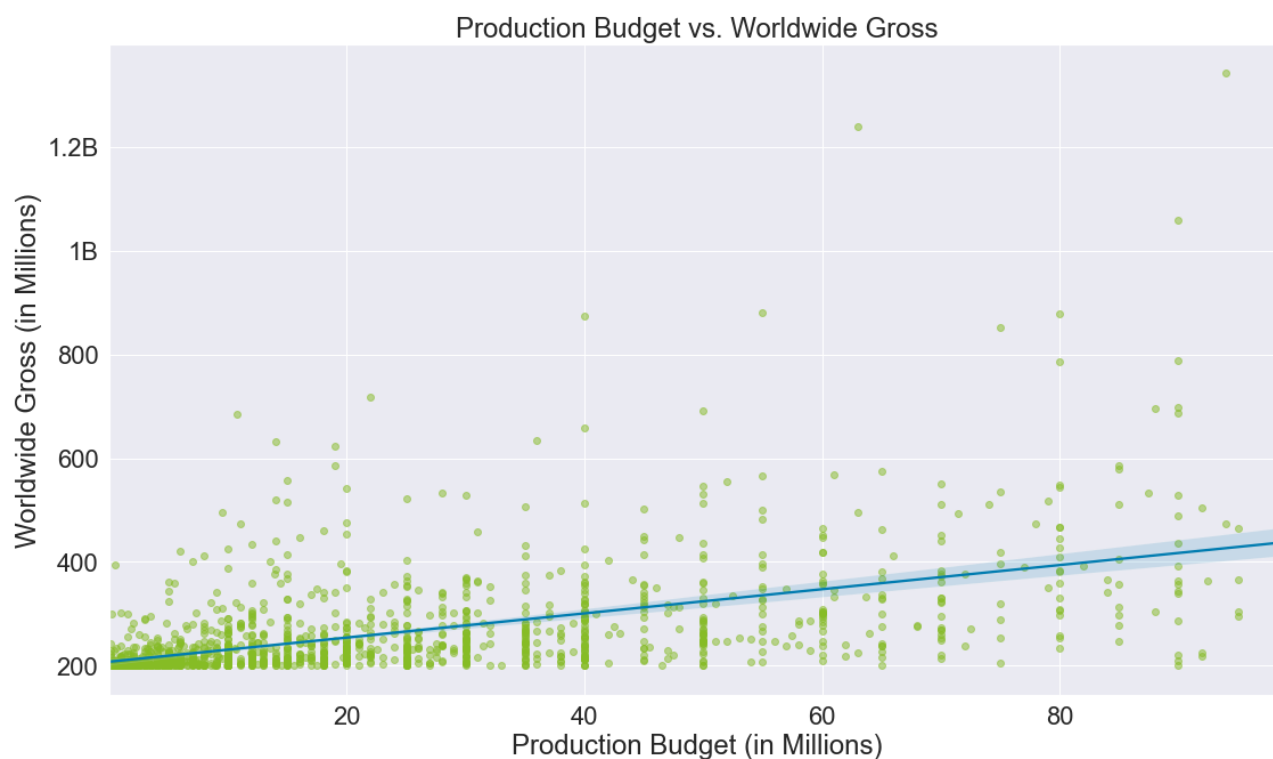
Business Question - What movie attributes will help create a successful movie?

Data Understanding and Analysis

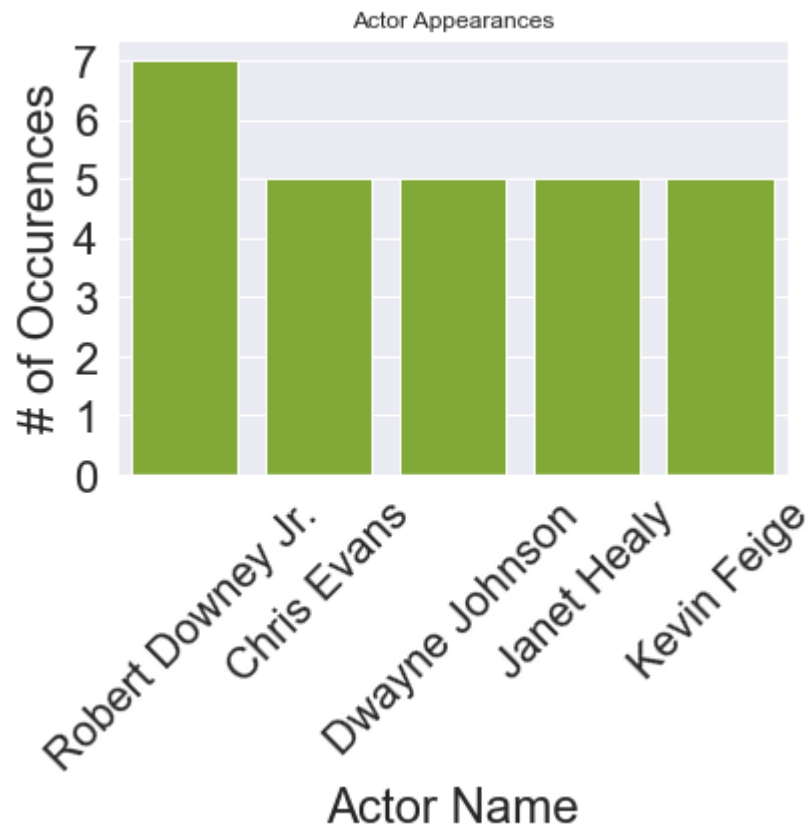
Sources of data

IMDB - Movie title, movie genre, movie actors
The Numbers - Movie title, movie release date, movie worldwide gross, movie production cost

Production Budget

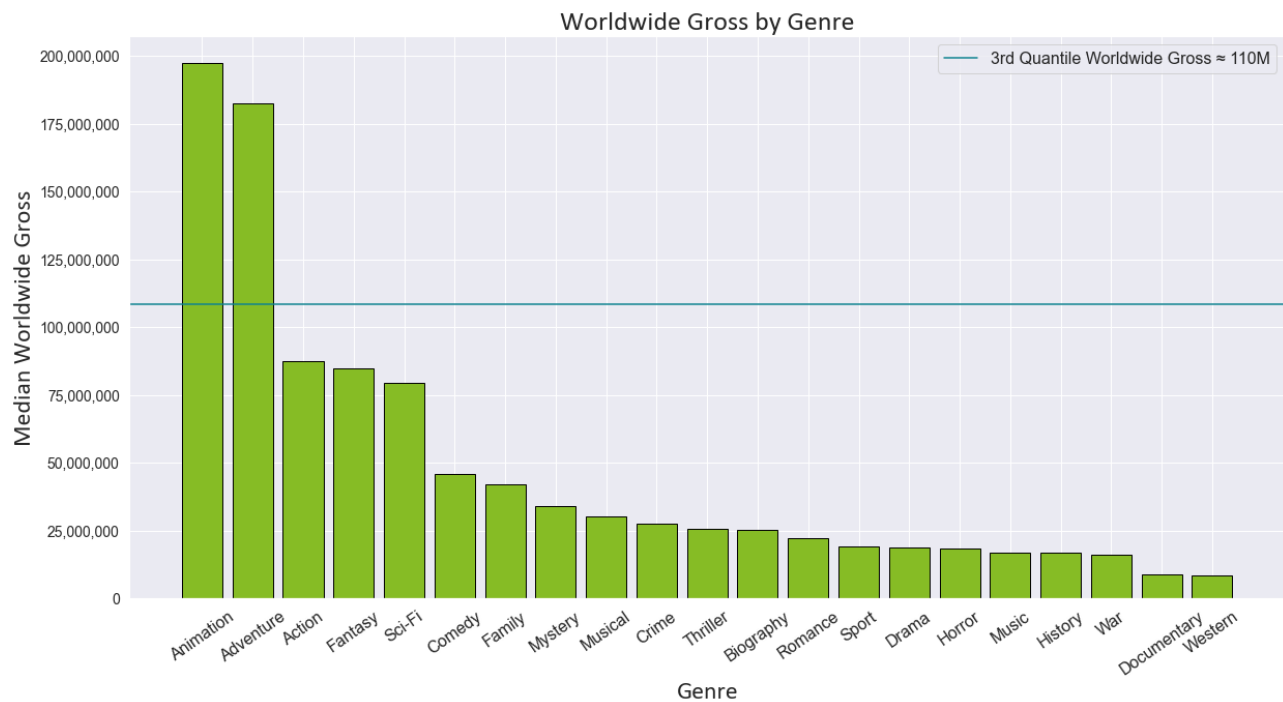


Movie production budget vs. Worldwide Gross of the movie. We can see the positive correlation between the two

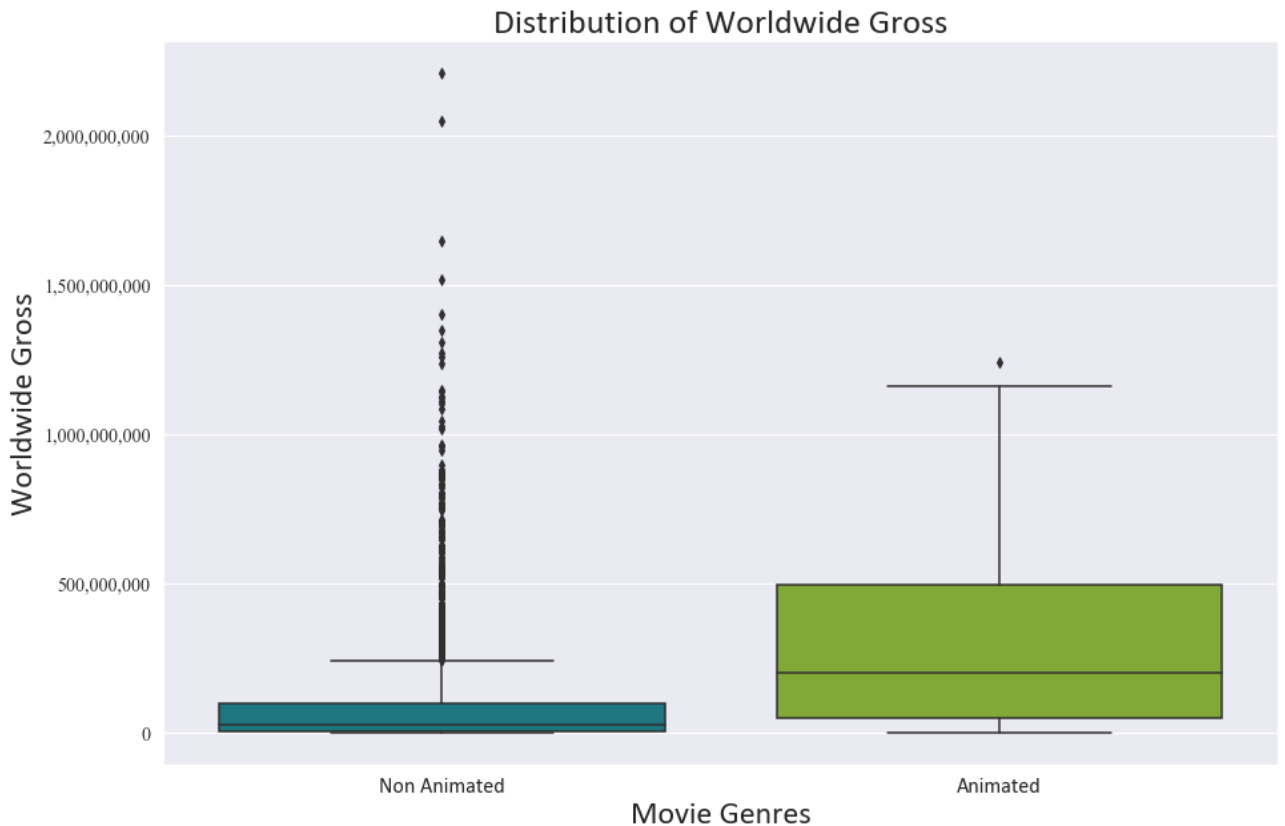


Most popular actors in most grossing movies. Popular actors grant visibility and therefore higher gross.

Genre

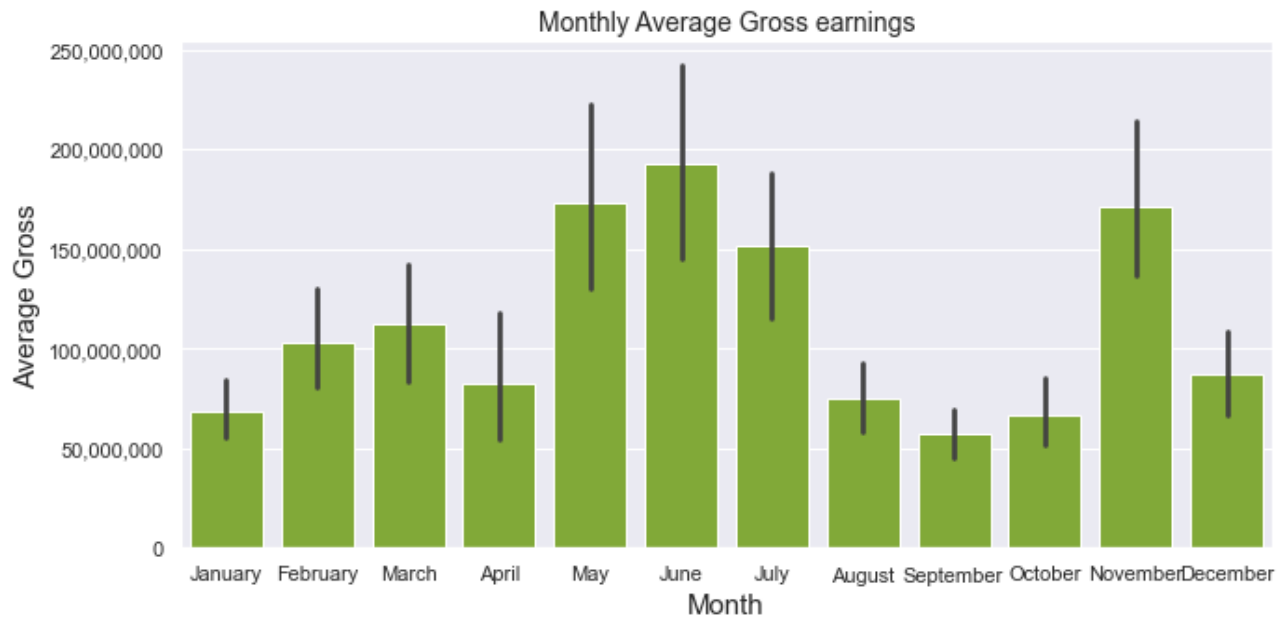


Median Worldwide Gross by Genre w/3rd quartile line. The three highest grossing genres were Animation, Adventure, and Action

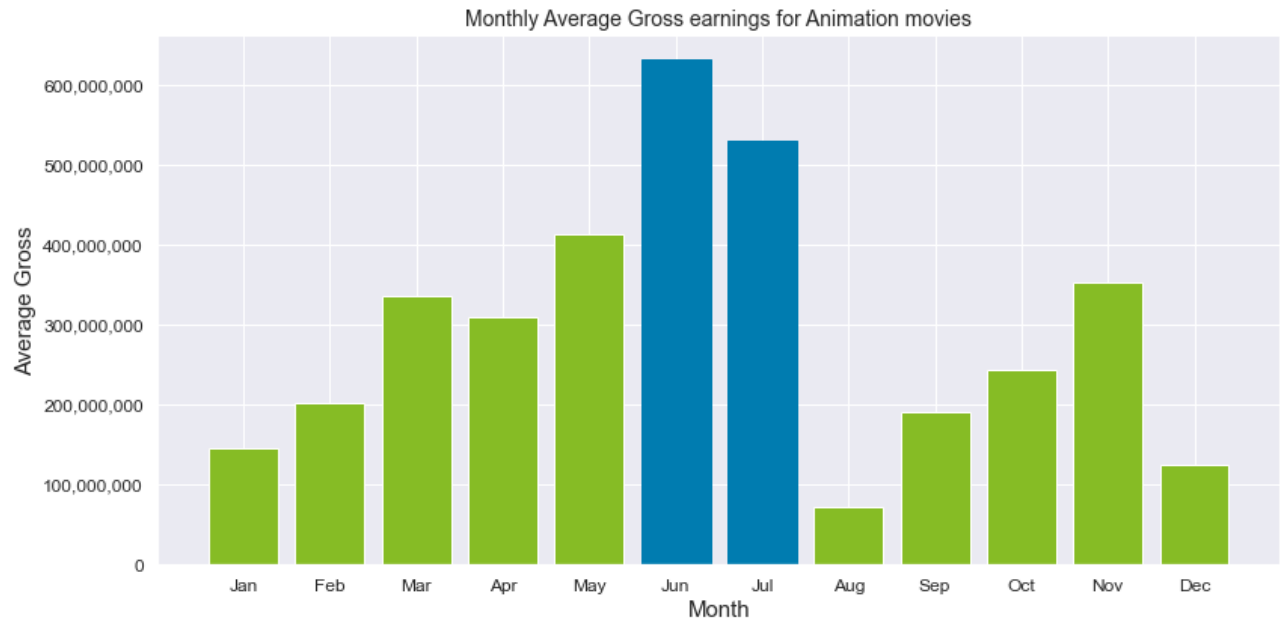


Distribution of Worldwide Gross for Animated vs. Nonanimated movies. Animated movies significantly perform better than other genres.

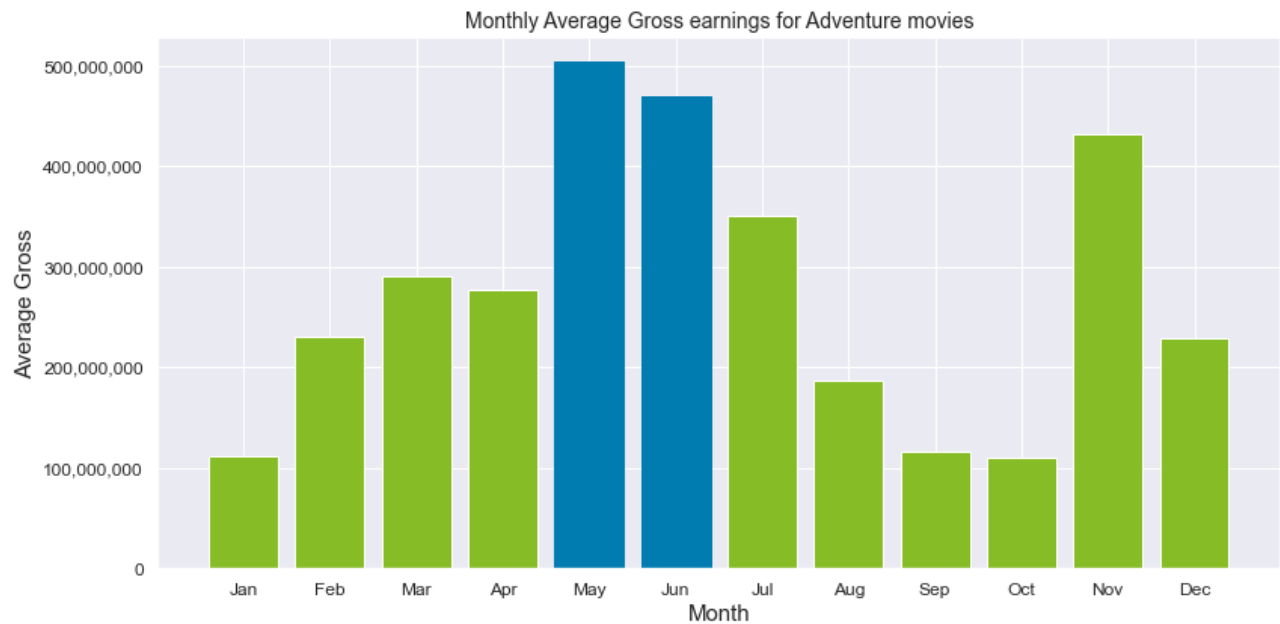
Release Month



From the graph above, we can see that the average gross earnings is significantly higher during the holiday seasons.



Animated movies released in the month of June yielded highest average gross earnings.



We can see that May has returned highest average gross earnings for Adventure movies, but having a very similar pattern to the Animated movies release time.

Statistical Communication

We performed a 1 sample t t-test.

Alpha = .05

H0: Sci-Fi Movies produce the same World Wide Gross as other genres of movies

HA: Sci-Fi Movies produce a higher World Wide Gross than other genres of movies

Rational

We are using a hypothesis test to determine if there was a significant difference between the Worldwide gross of Sci-Fi movies compared to the rest of the sample of movies. We used a 1 sample t-test instead of a graph so that we could see if the difference was statistically significant. With a p-value that is approaching 0, we can reject the null hypothesis

Limitations

The largest issue that comes here is due to the skewness of the data. This is likely due to the data collection hear, as the database with the production value and worldwide gross only has around 4k movies, compared to the 150k+ movies on the IMDB database. These movies also appear to be larger movies that have a much higher production value.

Recommendations

We would recommend that the company creates a sci-fi movie, as it has the highest return, and is statistically significant compared to the others.

Conclusion

After performing our analysis we were able to come to 3 conclusions for movie attributes that Computing Vision should consider when producing their movie.

Production Budget

After comparing production budget and worldwide gross, it was found that having a higher production budget can yield higher total gross.

Movie Genre

From the analysis, the top profitable genres are Animation, and Adventure. Computing Vision should create a movie that incorporates these genres.

Release Month

Performing analysis showed us that Computing Vision should release Animated movies in the months of June or July and an Adventure type movie in May or June .

Repository

index.ipynb - main notebook with our statistical analysis.

zippedData - Folder that contains the data sources.

sandbox - Contains files for early EDA.

gitignore - Blocks local files from being added to the repository.

Releases

No releases published

[Create a new release](#)

Packages

No packages published

[Publish your first package](#)

Contributors 5



Jniegocki



gthein98



hmoreddy



hamorris

Languages

● Jupyter Notebook 100.0%