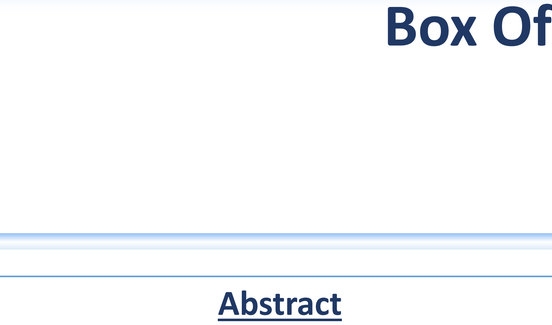
Box Office Breakdown: Analyzing the Factors that Drive Movie Revenue

Authors: Zane Brown, Dr. Robert Kelley

Bellarmine University Data Science Program

Abstract

This data science project is focused on analyzing movie data that includes a popularity score and vote average score provided by The Movie Database, as well as budget, revenue, runtime and genre, to try to gain insight into what drives a movies success. A predictive regression model will also be created to try to accurately predict movie revenue. The tools used to accomplish this include Python, SQLite3, and Power BI. Further analysis will be conducted to determine if there are any trend that hold true across a larger sample size and to identify other factors that contribute to a movie's success.



Objectives

- Explore how the revenue generated by movies in different genres differ from each other. For example, a certain genre may generate more revenue than another during specific parts of the year.
- Create a model that can accurately predict what the revenue of a movie will be based on a variety of different independent variables.
- Identify independent variables that could be significant in determining what a movies revenue could be.

Materials and Methods

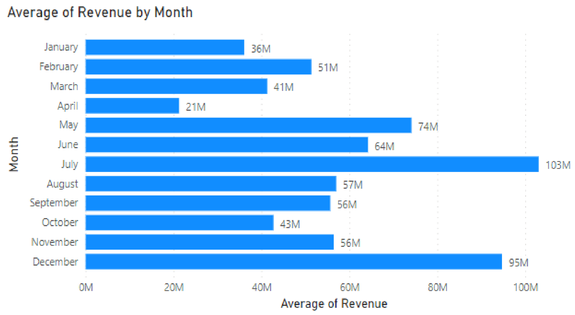
The Movie Database (TMDb): This is an API that I used to get data on popular/trending movies, as well as other movies provided from the IMDb file.

IMDb: Provided a large text file with over 1 million movie ids that allowed me to get more data from TMDb API. I was planning on using this more but only a paid version of the API was available.

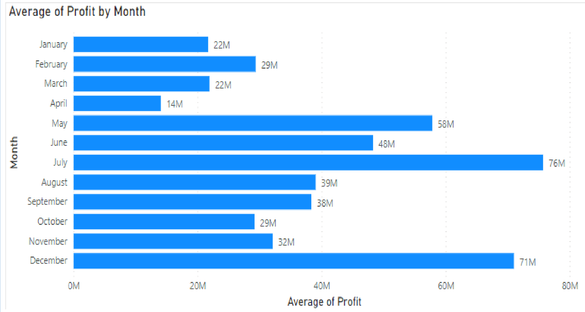
Python: Used pandas library to extract various data from TMDb API and arrange them into a single data frame. Additionally, used to clean up the data and get rid of movies with null values. As well as establish a connection to allow data to be pushed to a data store, SQLite3.

Power BI: Data visualization tool used to explore trends in the data that have to do with the revenue generated by the movie. This includes but not limited to average revenue by month, revenue by genre, and average

Results



July has the highest average revenue with an average of 103 million.



July has the highest average profit as well with an average of 76 million.

Conclusions

References

- The Movie Database (TMDb) : https://developers.themoviedb.org/3/getting-started/introduction
- IMDb: https://www.imdb.com/interfaces/
- Python: https://www.anaconda.com/
- SQLite3: https://www3.sqlite.org/index.html
- Power BI: https://powerbi.microsoft.com/en-us/

Contact

Zane J. Brown

Cell: (502)-410-9955

Email: zjbrown152@gmail.com

www.PosterPresentations.com