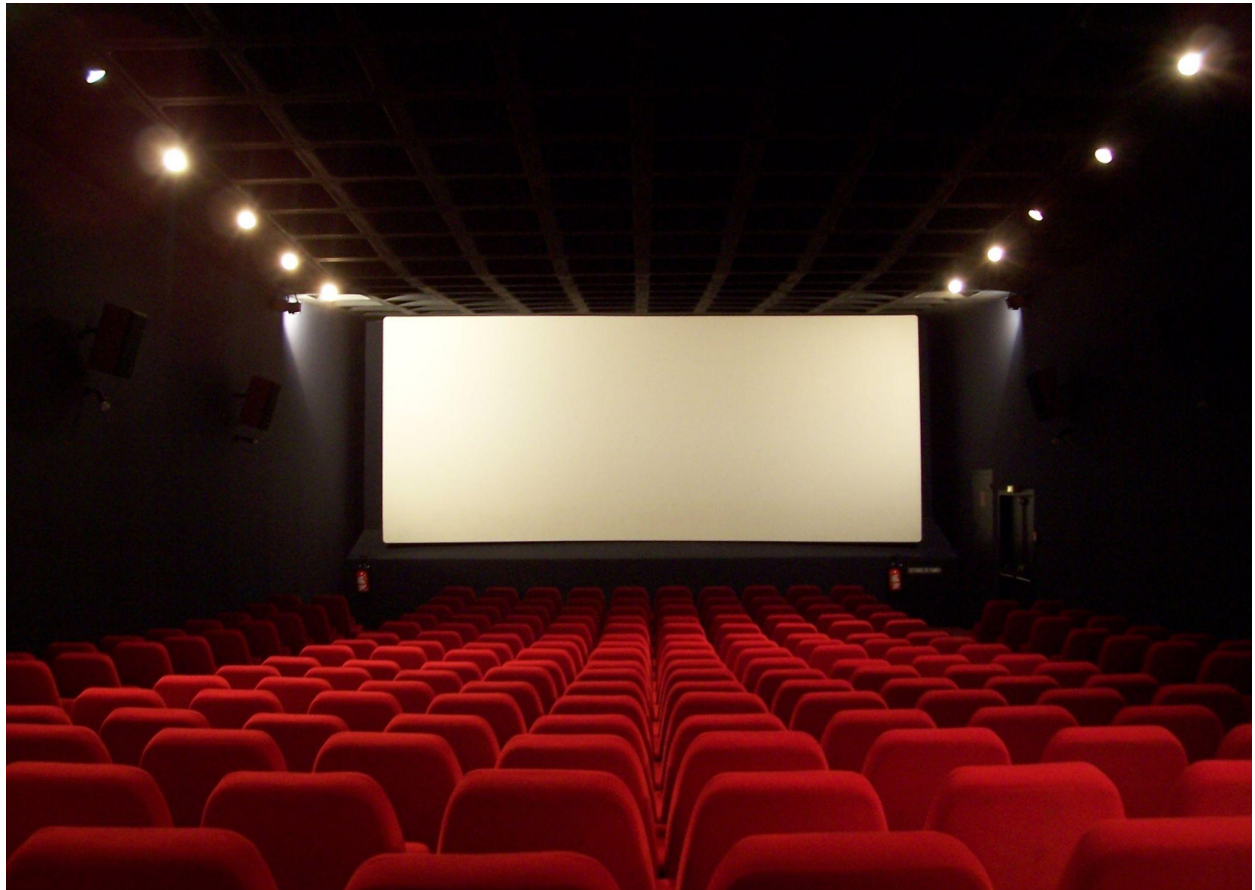# Investigating IMDb movies Dataset

Introduction

In our world, cinema and movies play an important role whether by getting us from a bad mood when we watch them or we work in the industry and it's our job that we love and makes us proud when it comes to light. So in this project I will walk you through some facts and interesting ideas about movies.

First : Cleaning

The dataset was easy to clean. It had a number of small issues. I used pandas libraries to erase or replace these issues to make the data frame easy to handle.

Issues:

- Replacing missing values of budget and revenue column with the mean of each column.

  Using boolean indexing method.

- Deleting some columns as they have a large number of missing values or it won't be needed like (homepage, id, ……).

  Using the 'drop' function to delete them.

- Replacing the character '|' with ', ' in some columns (cast,genres,…) to make it nicer and cleaner.

  Using for loop and "str.replace" function

- Turning the release date column to date time as it would be easier to work with.
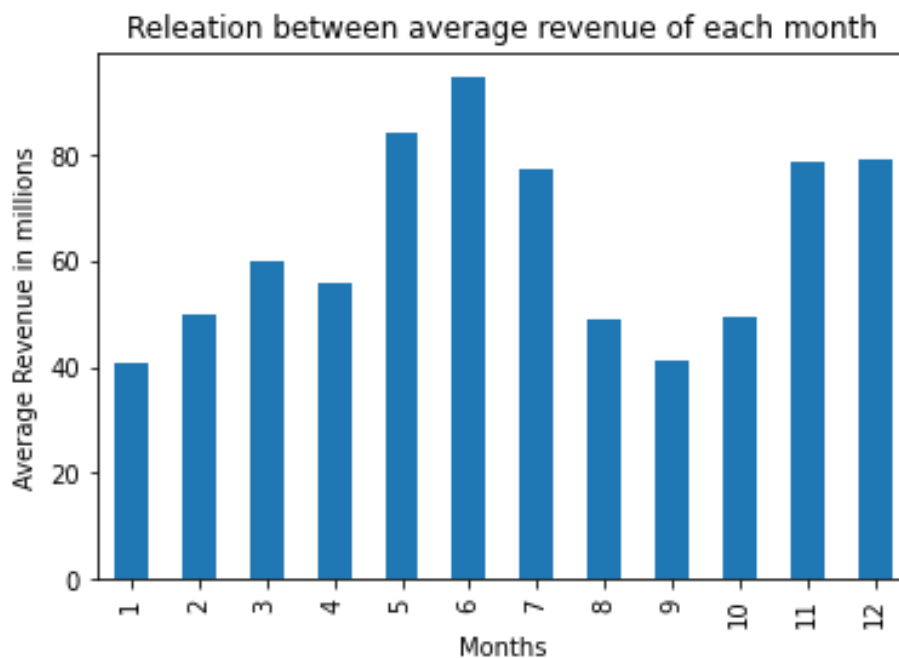
  Using "To_date" Function.

- Replacing missing values with the word "Unknown" in production company and director columns.

- swapping the popularity and title columns so when you open the data the name of the movie is the first thing you see.

After cleaning these Issues then I was ready to make my visualizations and explore the results I found from the graphs.

Visualizations:

A) Which month has the most average revenues?

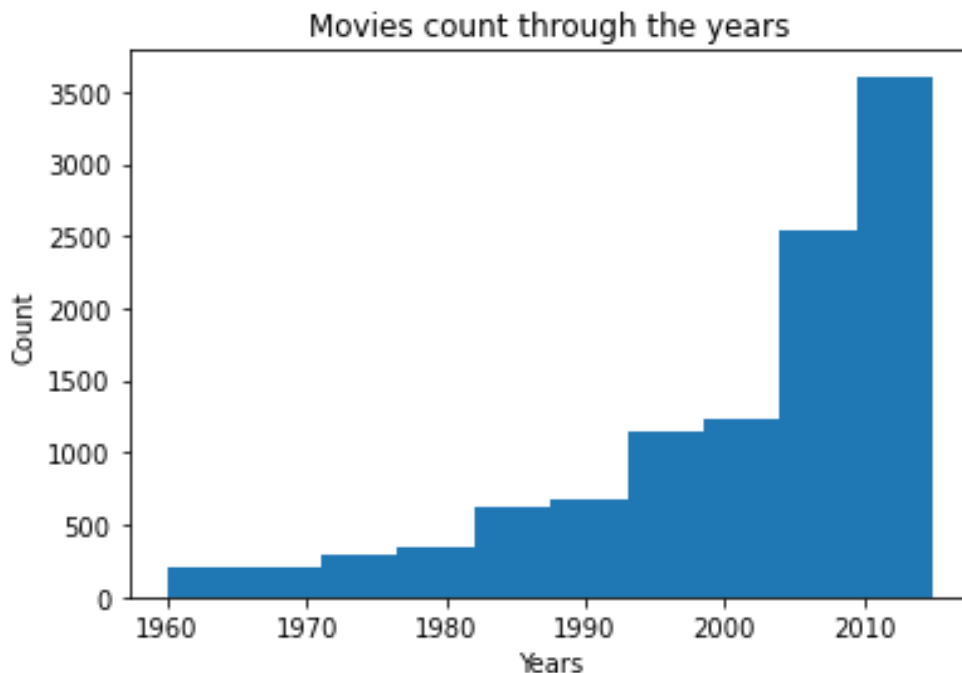Releation between average revenue of each month

This First Image shows the relation between the average revenues and the release month.

From this graph we can see that the most revenues are in june. But what does that mean?

June is the first month of summer and as a student myself when I finish my exams I want to go to the cinema and watch some movies with friends or family and that may be the reason behind this graph.

B) What happens with the number of movies through the years?

Movies count through the years

We can see from this graph that the number of movies increases with years and that means that the industry is growing and companies start to invest more in it.

And finally here are some interesting facts about movies:

- Most revenues : Avatar.

  It had the most revenues though it doesn't have the highest vote.

- Highest vote (9.2) : The Story of Film: An Odyssey.

  A documentary about movies. Seems funny right !

Conclusion:

These were my insights regarding the information I had

Taking into consideration that these results were based solely on the data at hand and had some limitations like there was a large number of zero values in the revenue and I took some approaches to solve these issues so someone may take different approaches and have different insights.

You can also find more information on IMDb([IMDb: Ratings, Reviews, and Where to Watch the Best Movies & TV Shows](#)) site or to try different sites like Rotten Tomatoes([Rotten Tomatoes: Movies | TV Shows | Movie Trailers | Reviews - Rotten Tomatoes](#)).

But in the end, The world of movies is massive and no one can predict what comes next and everyone may have their own opinions but sure we can all agree that we all enjoy it.