PHASE ONE PROJECT

README.md

Project Overview

For this project, you will use exploratory data analysis to generate insights for a business stakeholder.

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.

Introduction

The project was assigned as the final part of the Module 1 in Flatiron School's Data Science Program. The brief is detailed below:

"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 the problem is they don't know anything about creating movies. They have hired you to help them better understand the movie industry. Your team is charged with doing data analysis and creating a presentation that explores what type of films are currently doing the best at the box office. You must then translate those findings into actionable insights that the CEO can use when deciding what type of films they should be creating."

The project centres around the hypothetical scenario in which Microsoft is entering the movie production business. As a Data Scientist I have been tasked with analysing public movie databases (Rotten Tomatoes, IMDB, TMDB, and Box Office Mojo) and determining what characteristics of films make them do"best" at box office. It was my

decision to use top studio and gross domestic earning as the predominant indicators of success. I explored several characteristic's impact to bom.movie_gross and studio.

Python and Data Science

To do Data Science with Python, I used Python with the following software libraries:

Numpy: it contains many functions for numerical computation (vectors, matrices, polynomials, etc.).

Matplotlib: it allows to trace and visualize data in the form of graphs.

Pandas: it allows to manipulate and analyze data (dataframes).

Seaborn: it complements Matplotlib by providing attractive statistical graphics.

The first task I , a Data Scientist took was to prepare the data, this step may took a long time understand each of them and clean a bit. Then, the dataset was ready, I explored the data and analyze it. Once this step was done, I modelled the data, adapted and validate it.

Preparing the Data

The Datasets

The datasets included a wide range of related data as below in .csv files:

Box Office Mojo: bom.movie_gross.csv.gz

IMDB: imdb.title.crew.csv.gz, imdb.title.akas.csv.gz, imdb.title.ratings.csv.gz, imdb.name.basics.csv.gz, imdb.title.basics.csv.gz, imdb.title.principals.csv.gz
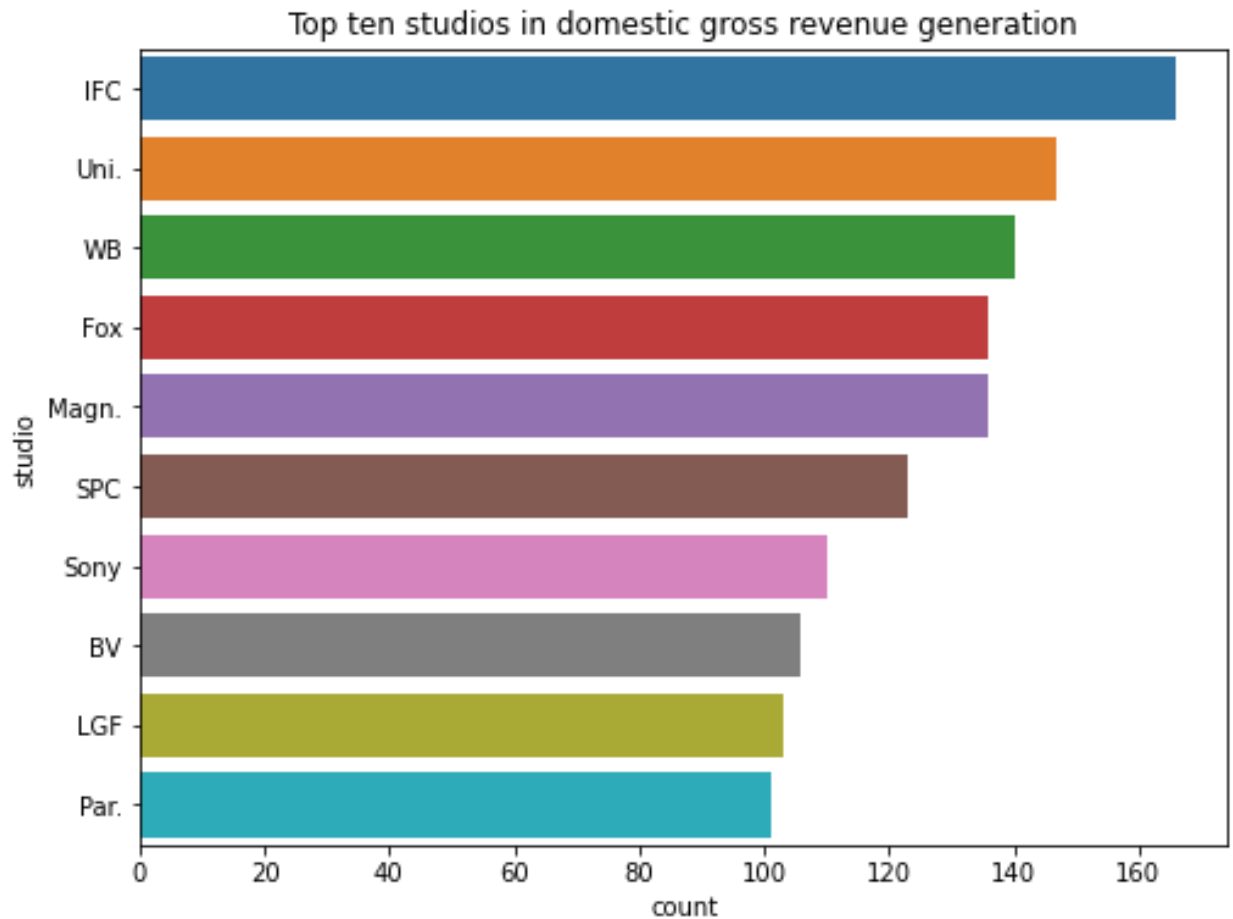
TMDB: tmdb.movies.csv.gz

The Numbers: tn.movie_budgets.csv.gz

After initial analysis and some data cleaning, the datasets which I used for our analysis was: bom.movie_gross.csv.gz for **financial data**
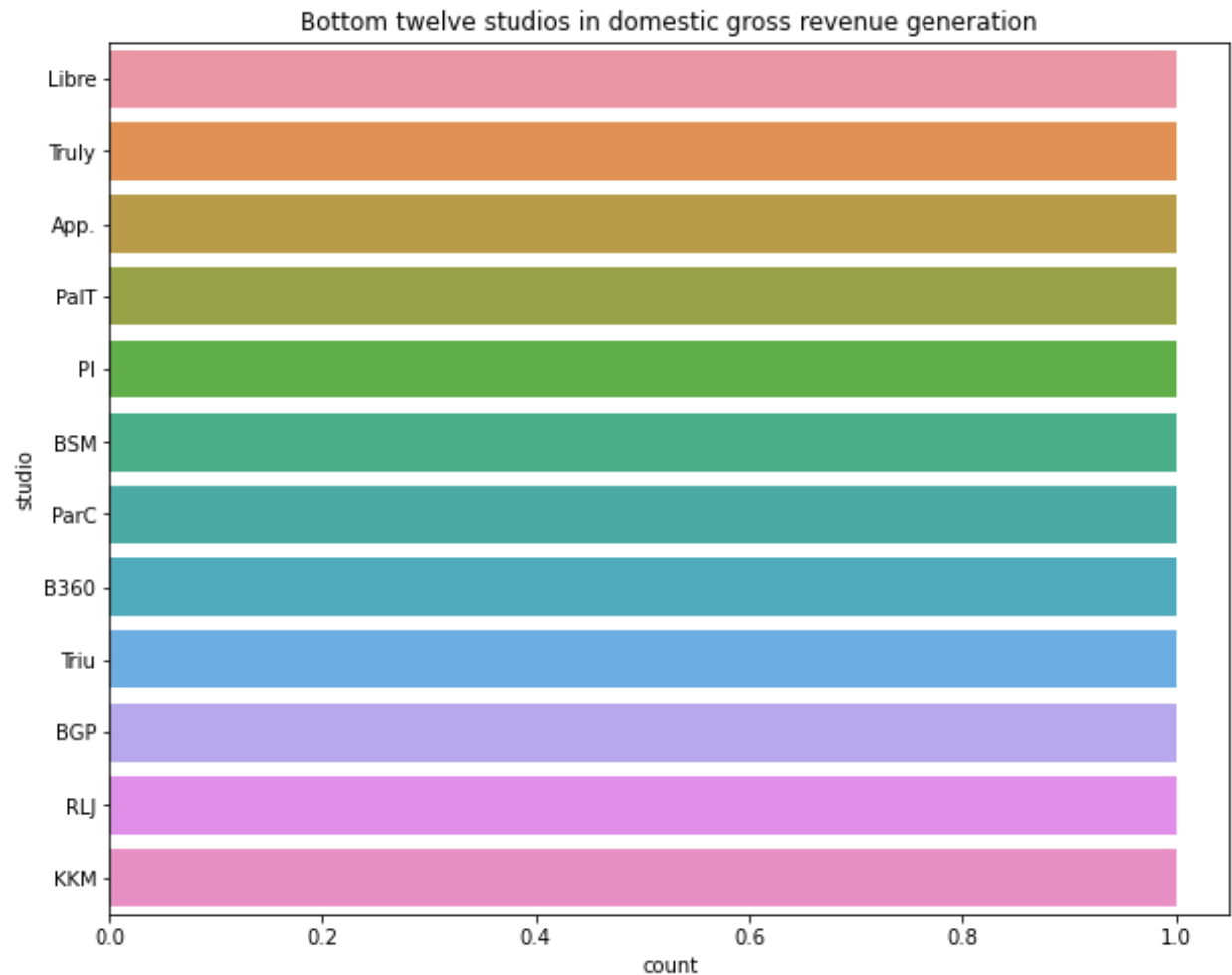
Which studios of movies have ROI?

Right, I couldn't start without having a look at what studios make the most money. Based on the data available, the top 10 averaging movie studios IFC, Uni, WB, Fox, Magn, SPC, Sony, BV, LGF and Par with IFC leading the studios. The average of profit was taken to account for the gross of studios produced. I assumed that the counts would be the most reliable with the data we had, and another assumption was that studios equal chance of producing/circulating movie.

Top ten studios in domestic gross revenue generation

Taking a look at the top 10 studios against gross domestic counts, we are able to paint a picture of the total amount of profit produced by each studio was immensely great. IFC has produced the most amount profit. This graph could be considered a reflection of whether studios as productive at per time of this analysis.

The bottom twelve studios aginst their respective gross domestic earning was dismally low. The below figure gave insights the productivity or otherwise of studios as well.

Bottom twelve studios in domestic gross revenue generation

My expert recommendation would be, MS is looking for a gamble then the top 10 averaging movie studios IFC, Uni, WB, Fox, Magn, SPC, Sony, BV, LGF and Par could continue keeping them in the movies business with immensurable comfort. The leading profit studios IFC, Uni, WB show a more convincing story, and IFC appears to be extremely profitable.

What are the leading movie studios for business?

The movie production industry is a competitive arena, and to be the best you have to beat the best. I've taken a deeper dive into which studios will be our most successful

competitors. Based on the data available, IFC, followed by Universal Studios and Warner Brothers. Here I assumed that the Total Gross produced by a Studio would be an indicator of whether they are a market leader.

Assuarity

I've only considered two metrics for the 'best' types of movies, which we decided would be studio and grossing. There are multiple variables that can be explored. To mention a few: ROI and High gross assumes high Ratings increasing the likelihood of more views on streaming platforms.

Conclusion

Preparation, modelling, visualization and interpretation of the data, allowed me to conduct analysis of the movie studio rankings on the gross domestic earnings.

Note Studio influences domestic gross profit and industry leaders preferred IRF branding

Recommendations

Go for the top studios.

Be cautious when dealing with bottom studios.

The top 10 studios should continue with their respective strategies put in place to maintain their top gross domestic product earningsand even improve further.

The bottom studios should put in place interventions to ensure they improve their gross domestic earnings.

The bottom 12 studios shoulde be merged to reduce their operational costs and maximise on their gross domestic revenue generation.

The bottom 12 studios should be merged with the top 10 gross domestic revenue generating studios that they may well versed in dealings geared towards reaping the optimum gross domestic revenue from the film industry.

What a movie exploration you have made known to yourself.