BEST MOVIE ANALYSIS

Author:

Paul Gitonga Njoki

Client:

Microsoft

Project Overview

With all major corporations developing original visual studios. Microsoft wants to join in and has chosen to open a new movie studio, but they don't know anything about virtual video creation. Microsoft has tasked me with determining what steps they want to take in order to enter this field. I was given many data files to evaluate and make recommendations to the head of Microsoft's new movie studio based on my findings in order to succeed in the field of movie development

Business Problem

As a company, Microsoft wants to start creating original video content but lacks knowledge about movie creation to move forward with its plan.

Objectives

The Project has the following objectives:

a. Identifying which movie genres do well in the dataset and garner the most public attention.
b. Assessing the ideal moment to release a film.
c. Determining the most popular Studios.

Using multiple data frames obtained from the Box Office, it was possible to identify correlations and trends in the data, allowing for better business decisions.

Data mining will help identify movie trends based on many factors, develop smarter movie creation methods, and accurately predict movie performance.

CRISP DM

I will follow the CRISP-DM process for this task

The **CR**oss **I**ndustry **S**tandard **P**rocess for **D**ata **M**ining (*CRISP-DM*) is a process model that servesas the base for a data science process. It has six sequential phases:

1. Business understanding – What does the business need?
2. Data understanding – What data do we have / need? Is it clean?
3. Data preparation – How do we organize the data for modeling?
4. Modeling – What modeling techniques should we apply?
5. Evaluation – Which model best meets the business objectives?
6. Deployment – How do stakeholders access the results?

Data and Analysis Overview

I intend to do this analysis on the data sets containing vast movie genres. When we study the distinct data files, we can see that the data includes many different sorts of information about each movie, such as the release date, the Studio, average rating, rating, gross domestic and foreign, and many other details acquired from multiple movie websites.

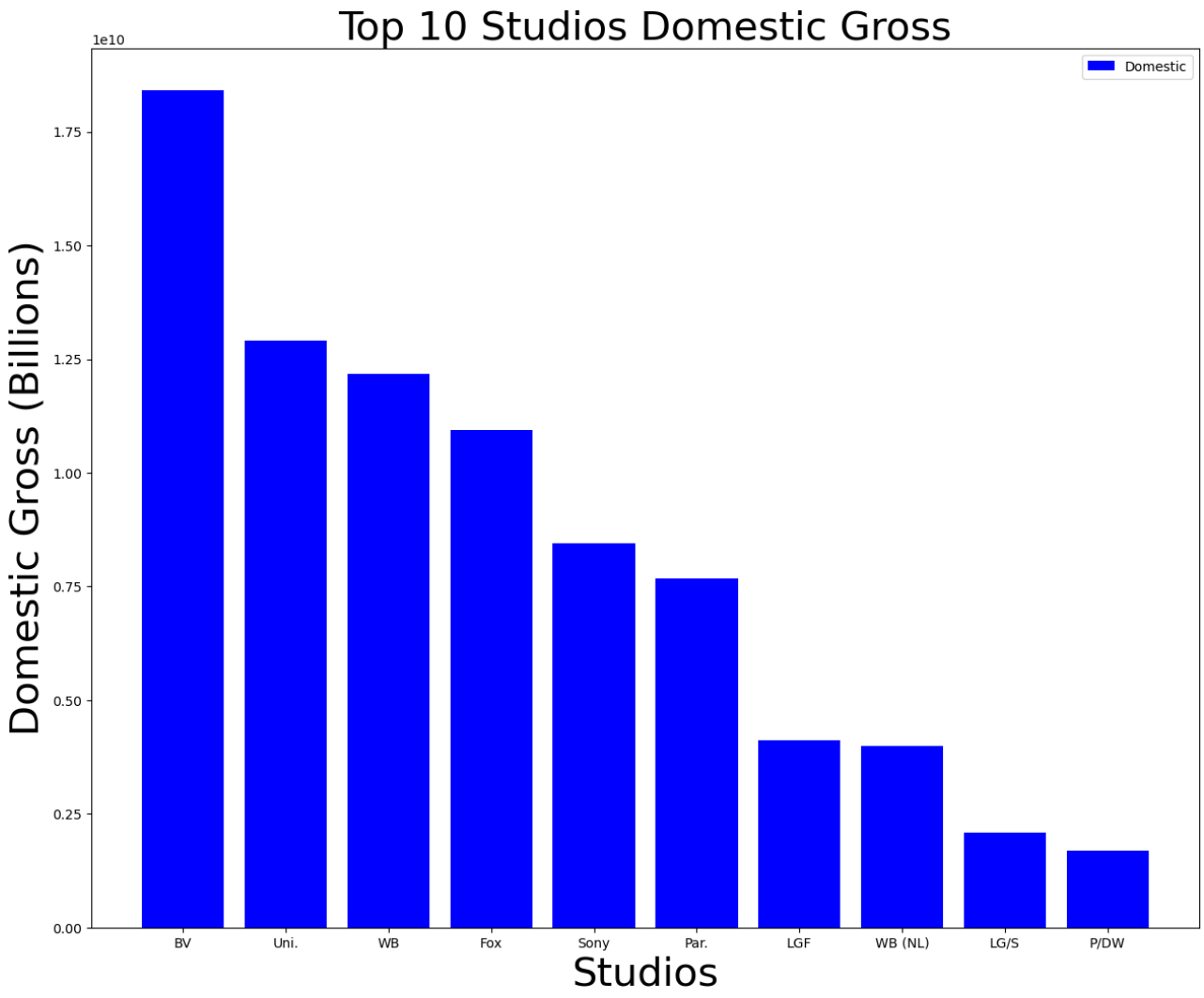
I utilized three different data sources for my analysis in order to have the most comprehensiveview of the current movie performance.

- Rotten Tomatoes Data: The dataset was provided in CSV format, having 1560 rows and 12 columns. According to the data, Drama is the most produced genre by value counts, followed by comedy.
- The Box Office Mojo Data: This was provided as zipped data in CSV format, with 5 columns and a collection of 3387 movies. The data set was taken from the Box Office website and spanned from 2010-2018. According to the Mojo data, most films were shot at the IFC studio.

I will start my analysis with a descriptive analysis of each data set. This allows me to identify trends in data relevant to what has to be known for a film to be successful. This analysis will be conducted mostly through the review of graphs featuring particular attributes.

Results

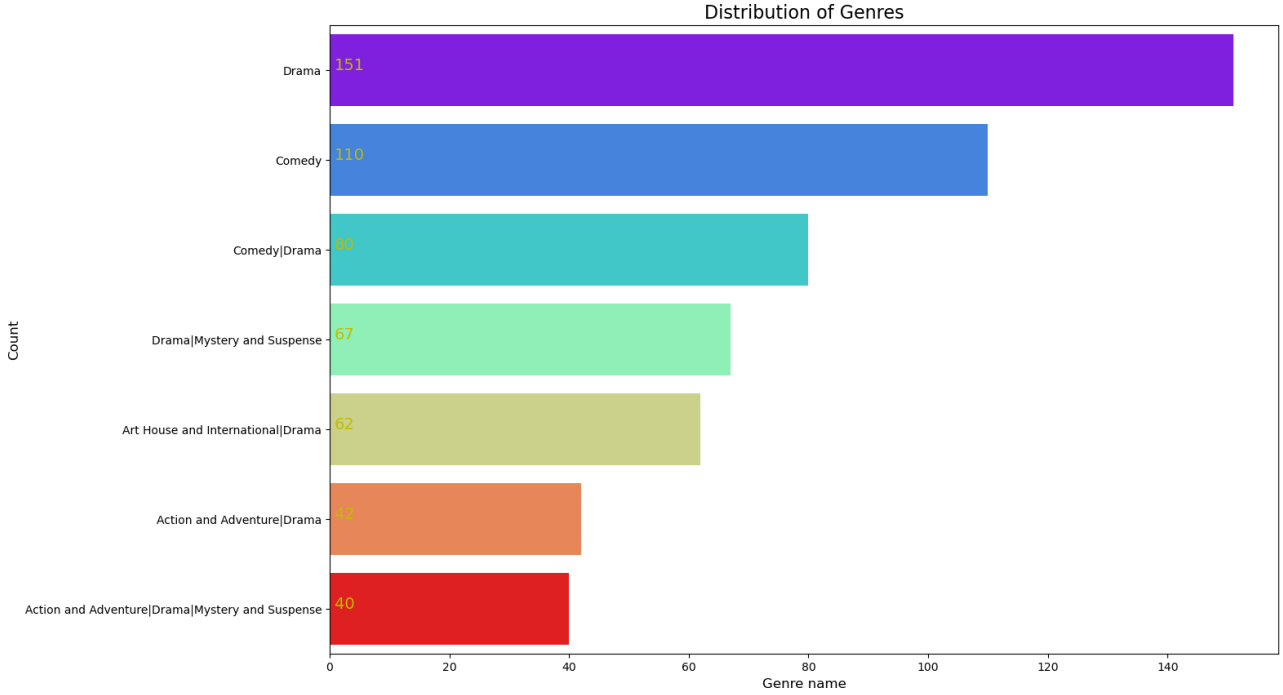
According to BOM data, on determining top 10 studios with their respective domestic gross:



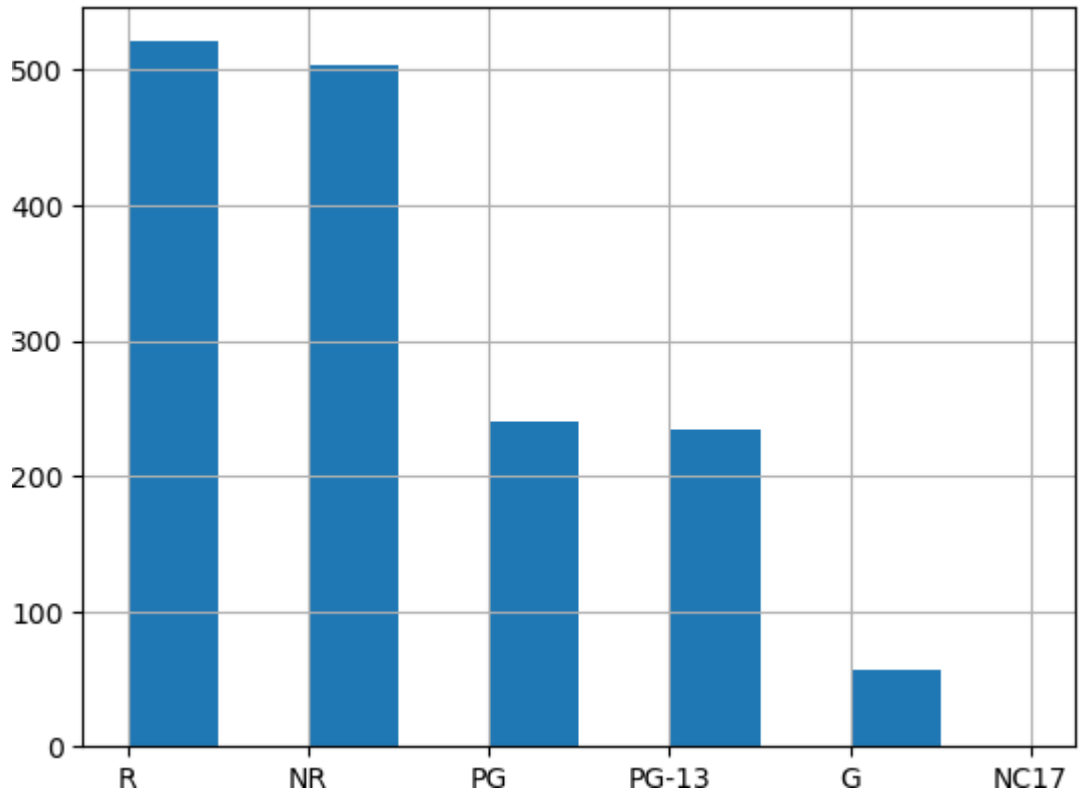
The domestic gross depends on the choice of Studio.

Using the Rotten Tomatoes Data, I was able to obtain the count of each genre. Drama has the most counts followed by comedy and the least was Comedy| Horror| Mystery and Suspense.

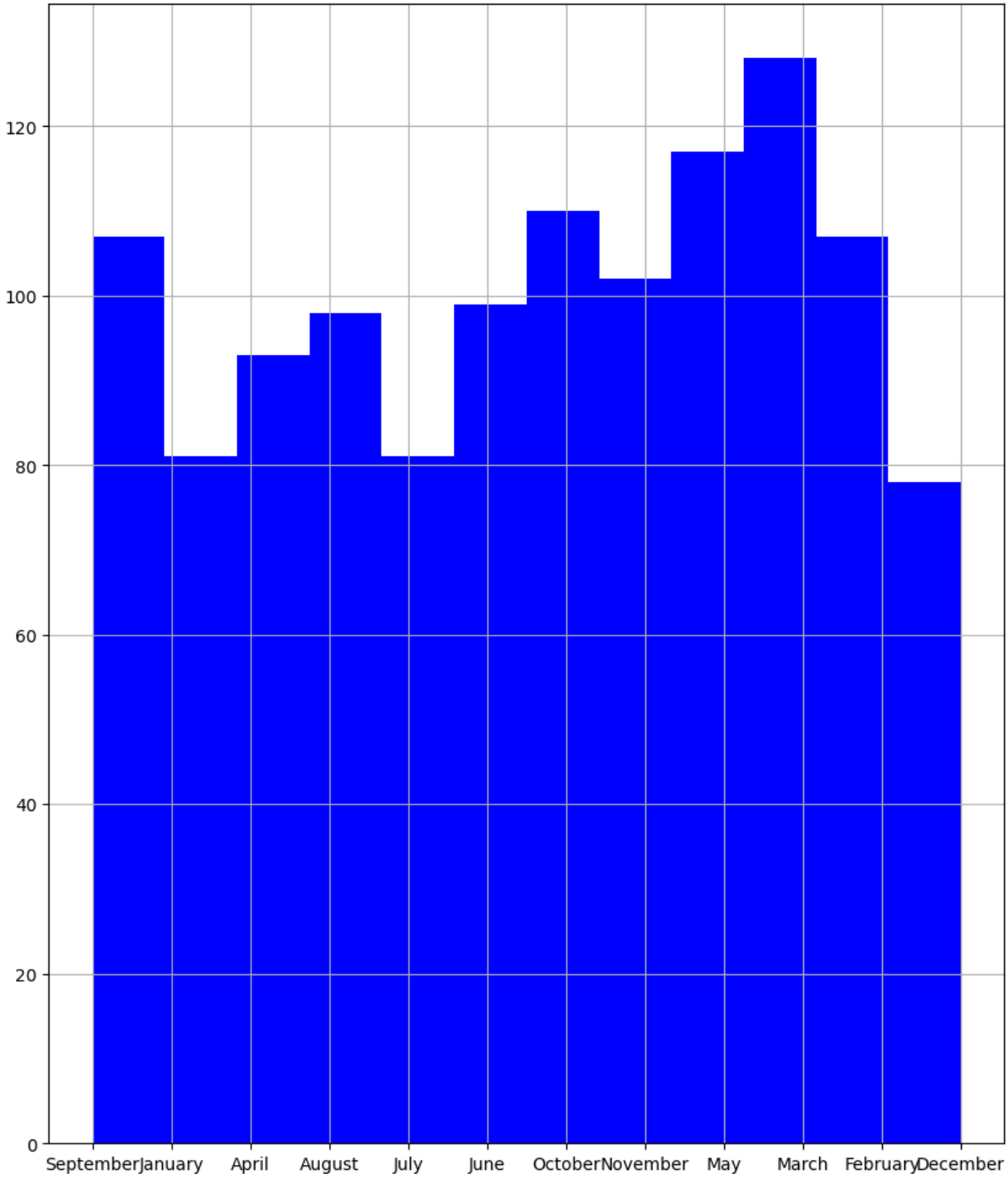
The diagram below illustrates the distribution of genres using the rotten tomatoes data:



Most movies were rated as R(Restricted)

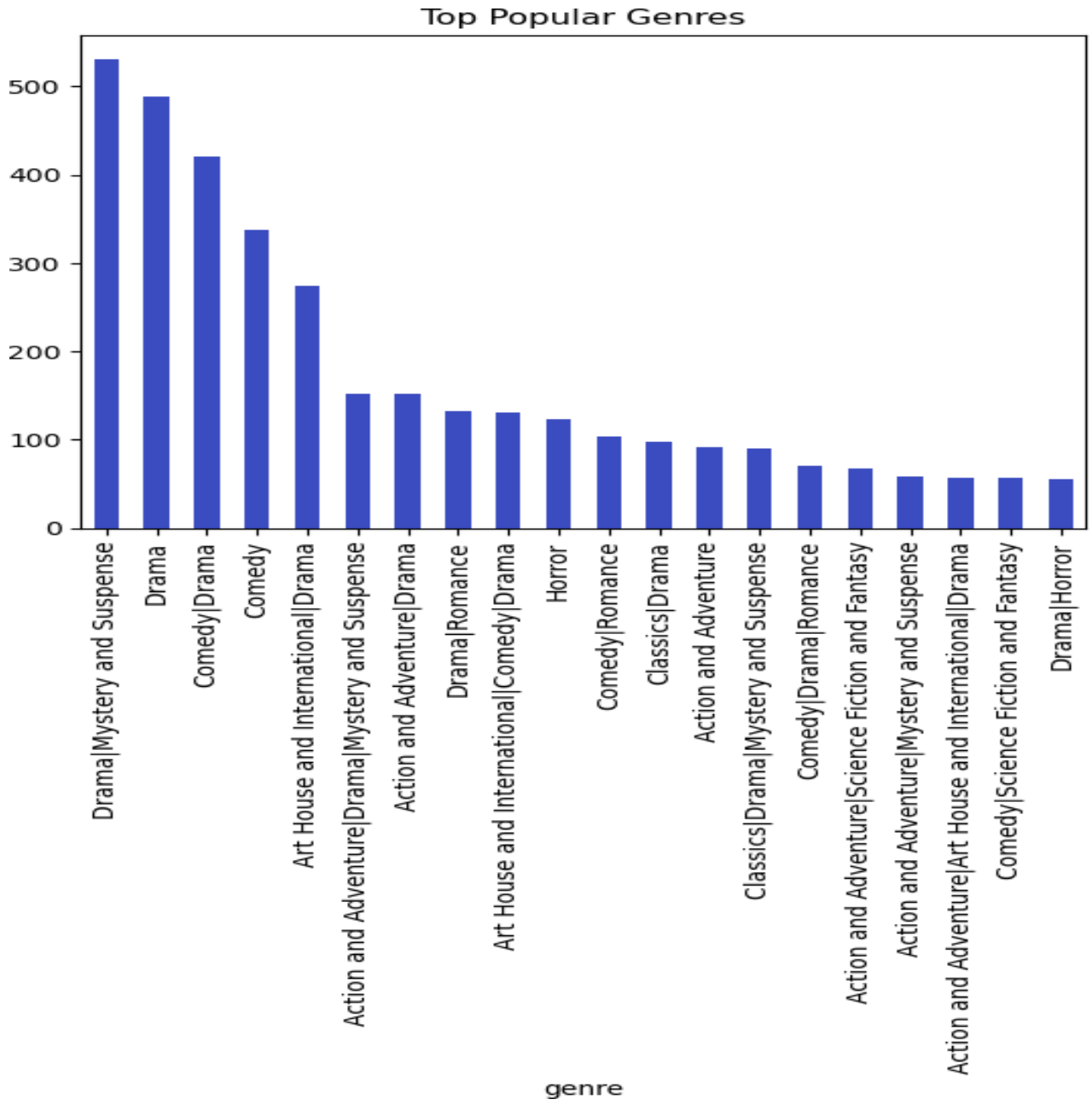


March, May, October, February, and September are the top 5 months in which most movies are released.As illustrated through the graph plotted below:

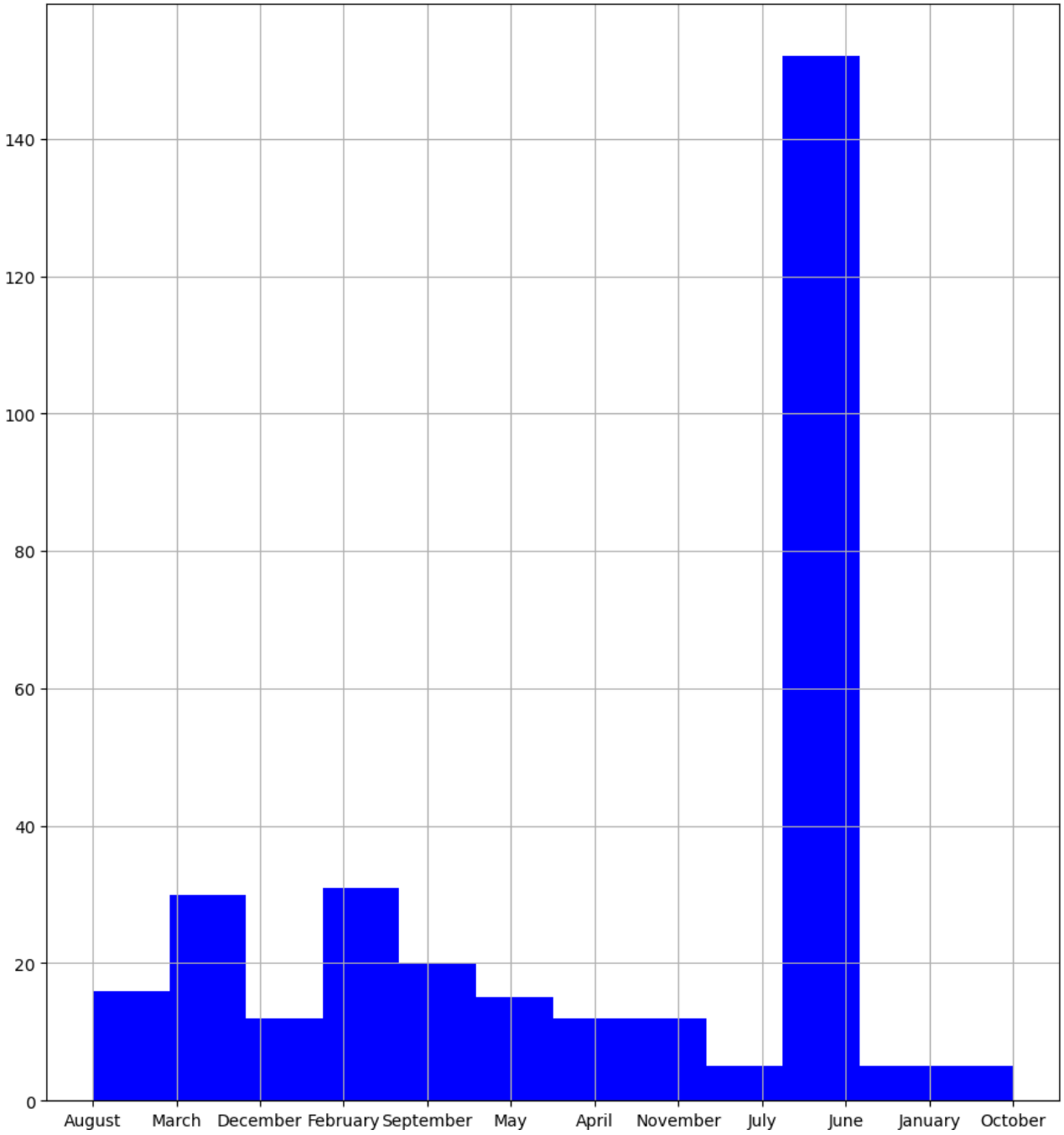


Upon merging two data sets, the BOM and The Rotten Tomatoes, the most popular genre is Drama| Mystery and Suspense, followed by Drama andComedy| Mystery and Suspense| Romance being the least popular.

Going by the illustration below:



I created a data frame for the Drama| Mystery and Suspense genres to analyse it further.



As per the graphical illustrations, we deduce that most Drama| Mystery and Suspense movies are released in June ascompared to any other time of the year.

Conclusion & Recommedations

1. Microsoft should produce a Drama, Comedy, Mystery, Fantasy and Suspense movie and release it in February, June or September.
2. From the Box Office Mojo, for Microsoft to have a good return on the domestic gross, they shouldgo for the top most-ranked studios.
3. In movie production, the Drama genre to be precise, for it to have a good average rating and receive the most public attention, Microsoft should consider having the best director in production.