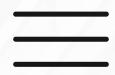


EXPLORATORY DATA ANALYSIS FOR MICROSOFT'S MOVIE STUDIO

BY : MOSCHINI ONYANGO

PROJECT OVERVIEW



Welcome to the **Exploratory Data Analysis (EDA)** presentation for Microsoft's new movies studio.

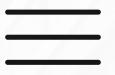
The main objectives for this analysis was to identify the types of films that are currently performing well in terms of their ratings, reviews and total annual revenue (both domestic and foreign) in the movies industry, and also to leverage the information from the insights of the analysis, into making a more informed decision, about Microsoft's movies production strategies.

The primary audience for this analysis is the "head of Microsoft's new movie studio", who may not have no extensive knowledge of the data science of the movies industry.

GOAL

The main goal of this project is to analyze movies datasets, from various sources and provide, actionable insights to help guide Microsoft's new studios in creating successful films. the main focus is on understanding the current trends in the movies industry and translating findings into recommendations for the type of films Microsoft's studios should consider producing in their movie production facility.

DATA SOURCES



The movie datasets used for this analysis were sourced from various platforms, including:

- (i) Box Office Mojo**
- (ii) IMDB**
- (v) The Numbers**
- (iii) Rotten Tomatoes**
- (iv) TheMovieDB**

The primary data for this analysis, was stored in the "im.db.zip" file, which is a zipped SQLite database containing relevant tables such as "movie_basics" and "movie_ratings." Additionally, i utilized the "bom.movie_gross.csv.gz" file, a compressed CSV file containing the movies office gross information.

METHODOLOGY

The analysis involved exploring data from different sources, including: **TSV files, SQLite databases, CSV files** and visualizing data from key areas of interest to the head of Microsoft's new movie studio, such as the type of movies, movie ratings, reviews, annual movie income.

the steps involved in the exploratory data analysis include:

step 1. importing libraries

using jupyter notebook, a python data analysis tool, data was imported through:

- import pandas as pd , to be able to read and manipulate the data.
- import numpy as np
- import sqlite3 to read the sql database data
- import matplotlib.pyplot as plt , to plot both the horizontal bar plot and line graph
- %matplotlib inline
- import seaborn as sns to increase the aesthetics of the plots

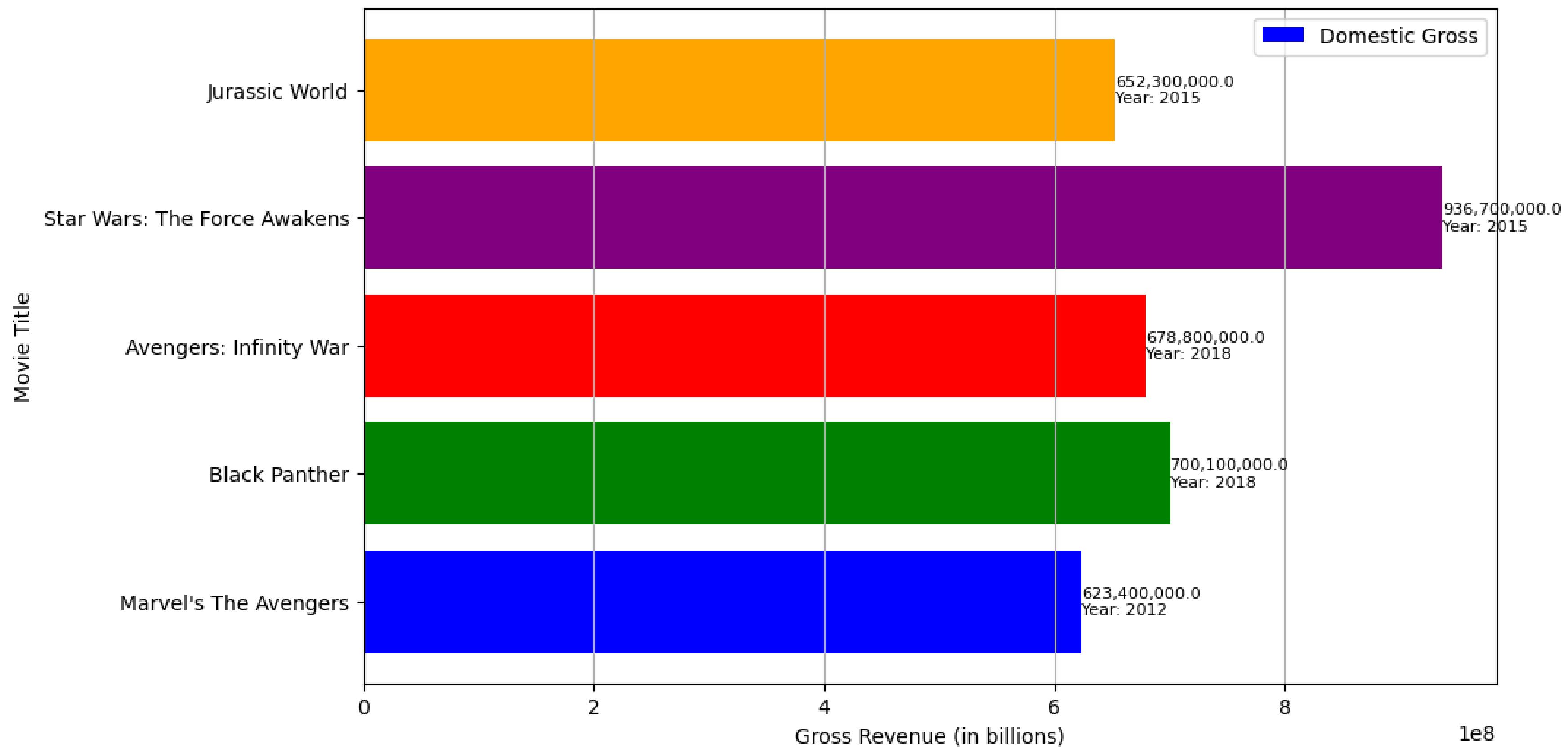
Step 2: Loading Data

- Data was loaded from various sources, including the movie budgets and movie gross datasets, and an initial examination of the data was done.

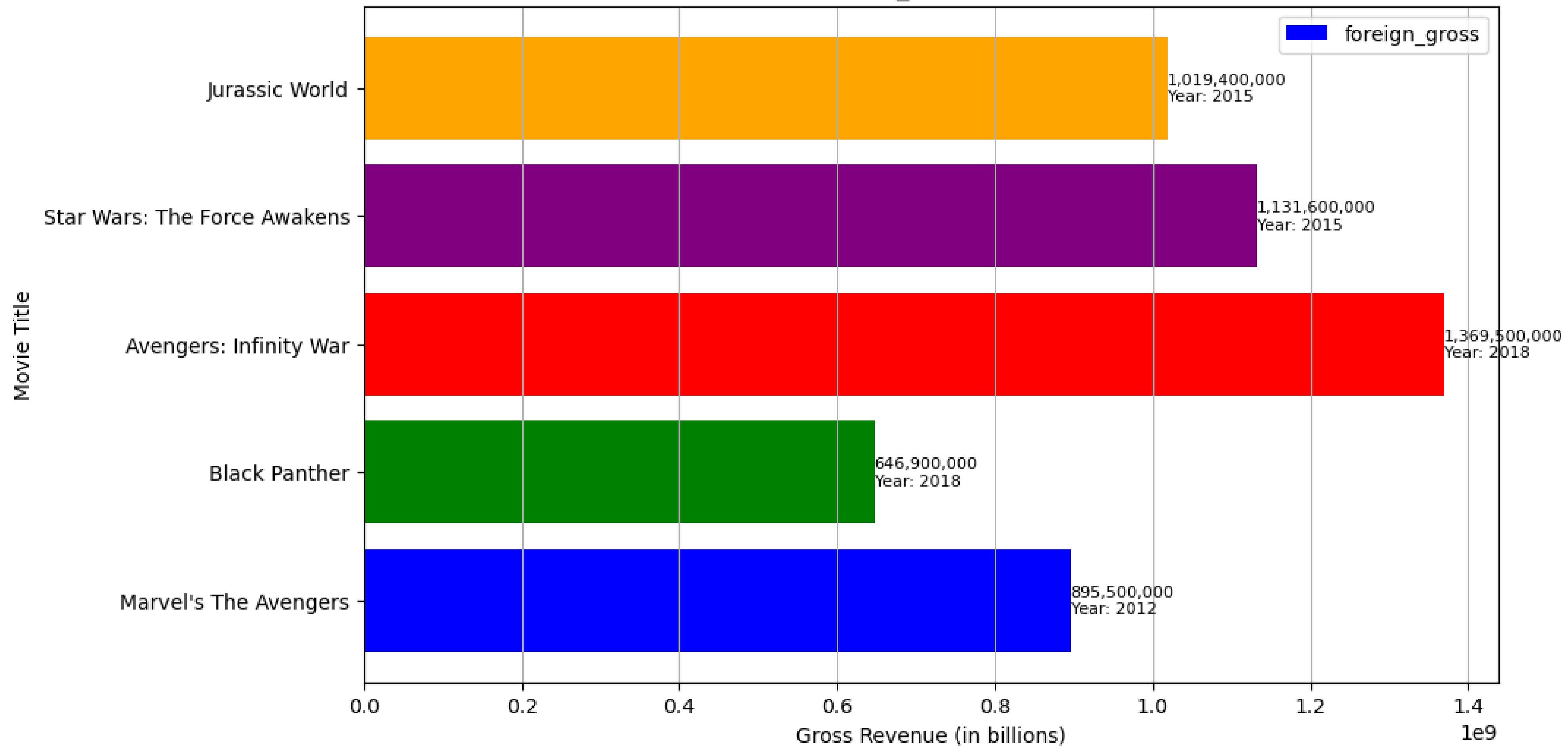
Step 3: Top Movies Analysis

- Data from the "bom.movie_gross" dataset was used to identify the top movies based on domestic and foreign gross. then the insights was used to visualize the information using horizontal bar plots.

Horizontal Bar Plot of Domestic Gross with Movie Names and Year



Horizontal Bar Plot of foreign_gross with Movie Names and Year



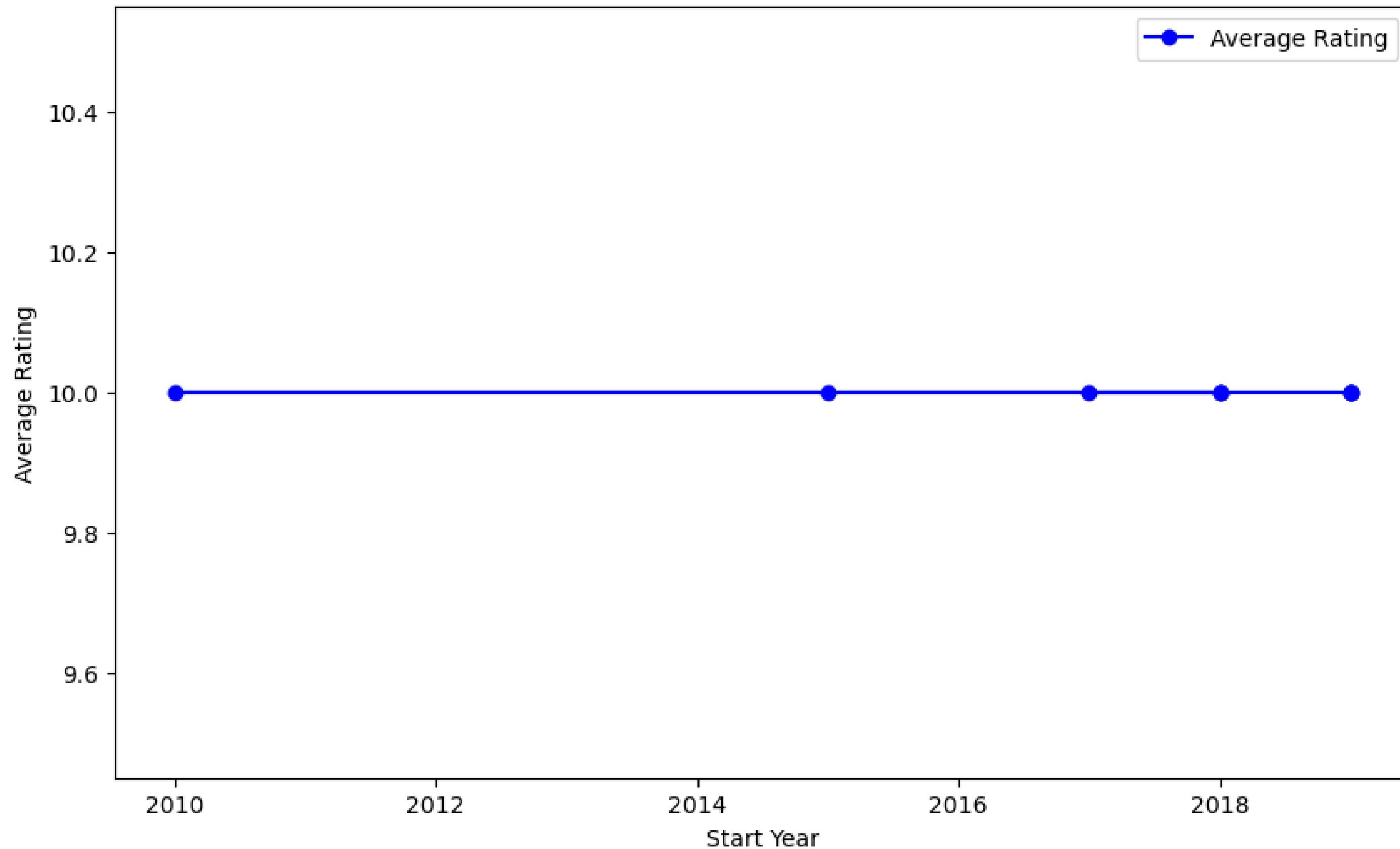
Step 4: Additional Data Sources

- Data explored data from Rotten Tomatoes, TMDB, and the "im.db" SQL database, specifically the "movie_basics" and "movie_ratings" tables.

Step 5: Movie Ratings Analysis

- Data collected, identified the movie with the highest rating, the top-rated movies were analyzed, and a line plot of average ratings for, over the years was created.

Line Plot of Average Ratings Over Years



RECOMMENDATION

Based on the analysis, it is recommended to Microsoft's movie studio, to consider focusing on the: **genres, release years, runtime and annual domestic and gross revenue**, that have historically resulted in successful films.

Q&A

1. Which specific tools or libraries were used in the analysis?

- tools like pandas for data manipulation, matplotlib for plotting the visualizations were utilized.

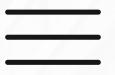
2. What data sources were used for the analysis?

- The movie datasets were sourced from Box Office Mojo, IMDB, Rotten Tomatoes, TheMovieDB, and The Numbers

3. Why is it important to consider both domestic and foreign gross revenue in the analysis?

- both domestic and foreign gross revenue provides a comprehensive view of a film's global success.

CONCLUSION



This exploratory data analysis provides valuable insights for Microsoft's new movie studio, guiding decision-making for the production of successful and well-received films. The detailed analysis of the movie trends, top-rated movies, and key features contributing to success will contribute to Microsoft's strategic planning in the competitive movie industry.

LEARN MORE

**Here is a GitHub link to the project
files: [click](#)**

