Microsoft Movie Studio Analysis

Introduction

Microsoft has decided to venture into the movie industry by creating a new movie studio. To ensure the success of this new venture, it's crucial to understand the types of films that are currently performing well at the box office. This project aims to analyze various movie datasets to identify trends and provide actionable insights for Microsoft's new movie studio.

Data Overview

For this project, we used datasets from the following sources:

1. **Box Office Mojo**: Provides movie gross revenue data.
2. **IMDB**: Provides information on movie titles and ratings.
3. **Rotten Tomatoes**: Provides movie reviews and ratings.
4. **TheMovieDB**: Provides detailed information about movies.
5. **The Numbers**: Provides box office and movie financial data.

We primarily focused on the following files:

- `imdb.title.basics`
- `imdb.title.ratings`
- `bom.movie_gross`

Data Cleaning

The datasets were cleaned and preprocessed to ensure accurate analysis. The steps included:

1. **Loading the data**: Imported the datasets into the Jupyter Notebook.
2. **Inspecting the data**: Checked for missing values and data types.
3. **Handling missing values**: Filled or removed missing values as appropriate.
4. **Merging datasets**: Combined the datasets based on common columns to create a comprehensive dataset for analysis.

```
# Importing necessary libraries
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# Loading the datasets
title_basics = pd.read_csv('zippedData/imdb.title.basics.csv.gz')
title_ratings = pd.read_csv('zippedData/imdb.title.ratings.csv.gz')
movie_gross = pd.read_csv('zippedData/bom.movie_gross.csv.gz')

# Inspecting the datasets
print(title_basics.head())
print(title_ratings.head())
print(movie_gross.head())

# Merging datasets
movies = title_basics.merge(title_ratings, on='tconst').merge(movie_gross, left_on='primaryTitle', right_on='title')
```

```
<style type="text/css">@media print {
    *, :after, :before {background: 0 0 !important;color: #000 !important;box-shadow: none !important;text-shadow: none !im
    a, a:visited {text-decoration: underline}
    a[href]:after {content: " (" attr(href) ")"}
    abbr[title]:after {content: " (" attr(title) ")"}
    a[href^="#"]:after, a[href^="javascript:"]:after {content: ""}
    blockquote, pre {border: 1px solid #999;page-break-inside: avoid}
    thead {display: table-header-group}
    img, tr {page-break-inside: avoid}
    img {max-width: 100% !important}
    h2, h3, p {orphans: 3;widows: 3}
    h2, h3 {page-break-after: avoid}
}
html {font-size: 12px}
@media screen and (min-width: 32rem) and (max-width: 48rem) {
    html {font-size: 15px}
}
@media screen and (min-width: 48rem) {
    html {font-size: 16px}
}
body {line-height: 1.85}
.air-p, p {font-size: 1rem;margin-bottom: 1.3rem}
.air-h1, .air-h2, .air-h3, .air-h4, h1, h2, h3, h4 {margin: 1.414rem 0 .5rem;font-weight: inherit;line-height: 1.42}
.air-h1, h1 {margin-top: 0;font-size: 3.998rem}
.air-h2, h2 {font-size: 2.827rem}
.air-h3, h3 {font-size: 1.999rem}
.air-h4, h4 {font-size: 1.414rem}
.air-h5, h5 {font-size: 1.121rem}
.air-h6, h6 {font-size: .88rem}
.air-small, small {font-size: .707em}
canvas, iframe, img, select, svg, textarea, video {max-width: 100%}
body {color: #444;font-family: 'Open Sans', Helvetica, sans-serif;font-weight: 300;margin: 0;text-align: center}
img {border-radius: 50%;height: 200px;margin: 0 auto;width: 200px}
a, a:visited {color: #3498db}
```

```
a:active, a:focus, a:hover {color: #2980b9}
pre {background-color: #fafafa;padding: 1rem;text-align: left}
blockquote {margin: 0;border-left: 5px solid #7a7a7a;font-style: italic;padding: 1.33em;text-align: left}
li, ol, ul {text-align: left}
p {color: #777}</style>
```