Milestone project 1:

Microsoft movies analysis

Author: Ngoc Ho

Summary

- This project analyses the movies using data from IBDM database, Box office Mojo and The Numbers. Results from this data analysis can guide Microsoft's decision on genre of movies, production budget and release timeline.
- Descriptive data analysis of title rating and domestic/worldwide gross showed the most popular and highest gross movie genre are Action, Adventure and Sci-Fi with median runtime between 80-100 minutes.
- Profit varies seasonally and the most profitable season is Summer suggesting this is a favourable season for a movie release.
- The budget required to make a top rating movie is around 30 millions.

Outline

- Business Problem
- Data
- Methods
- Results
- Conclusions

Business Problem

Microsoft is interested in building a movie studio.

To successfully set up a new movie studio, we have to consider:

- Properties of a popular and profitable movie such as their genre, runtime
- Business aspects such as budget, profit margin
- Strategy aspects such as best time to release, marketing strategies, etc,.

Data analytic questions that might help exploring the these points:

- Which movie genres has highest rating/most popular/ highest gross
- What is a typical runtime of top rating movies?
- Which season best to release a movie?
- What is the average budget for a movies? Does higher budget correlates to higher profit?

Data

In this project we use data from:

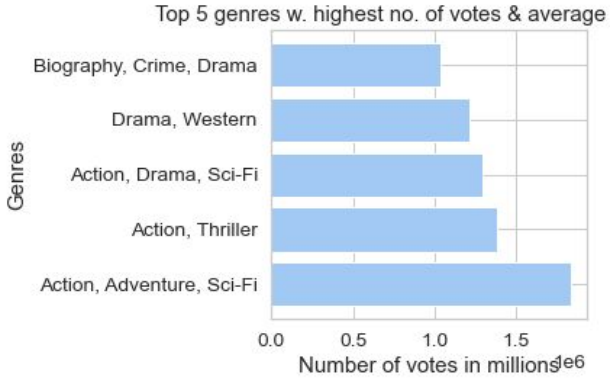
Database	Data info	Dimension	Target variables
IMDb	Title.basic : movie titles, release year, and genres Title.ratings: average movie rating and number of votes	146144 titles and 6 columns 73856 rows x 3 columns	genres, runtime minutes, average rating and number of votes.
Box Office Mojo	movie's production budget, domestic gross, worldwide gross and release date from 1915 to 2020.	3387 movies entries and 5 columns	domestic and foreign gross and studio
The Numbers	movie's production budget, domestic gross, worldwide gross and release date from 1915 to 2020.	5782 movies entries with 6 columns	production budget, domestic, worldwide gross and release date

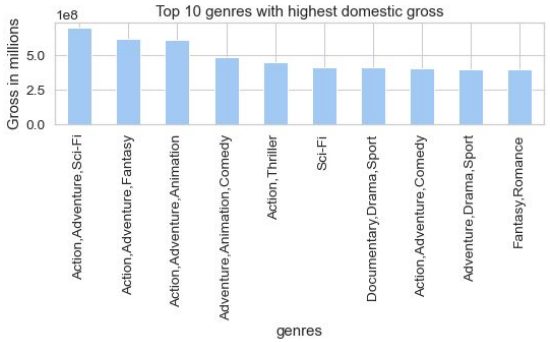
Methods

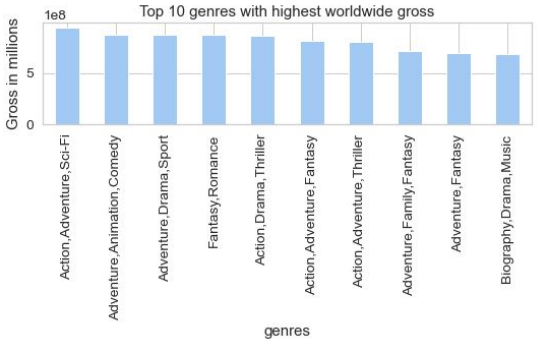
1. **Highest average rating and highest votes:** data was filtered to only movies with average rating >=8 then the top 5 genres with highest number of votes were identified
2. **Typical runtime:** find median runtime for movies with average rating >=8
3. **Most profitable release season:** ROI was calculated for all titles then classified into seasons according to release date
4. **Typical production budget:** median budget of all movies and top 20% most profitable were calculated
5. **Correlation between budget vs profit:** regression analysis of both variables

Results

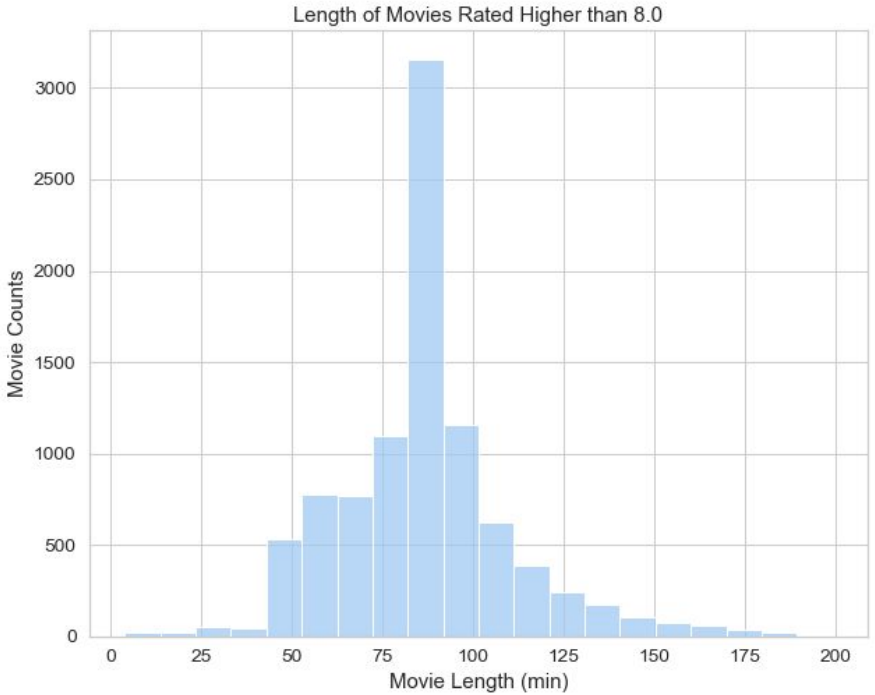
1. **Highest average rating with highest votes and gross:** Action/Adventure/Sci-fi



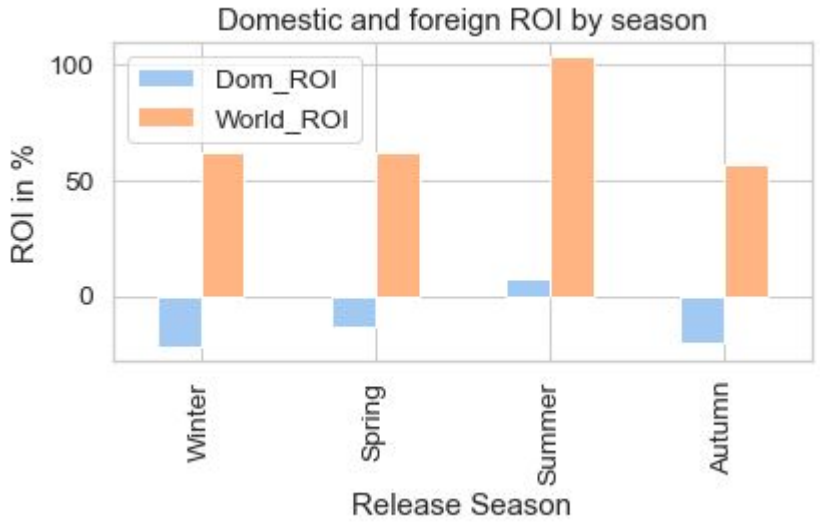




2. Typical runtime: 80-100 minutes



3. Most profitable release season: Summer with ROI 103% globally and is the only season with positive ROI value domestically.

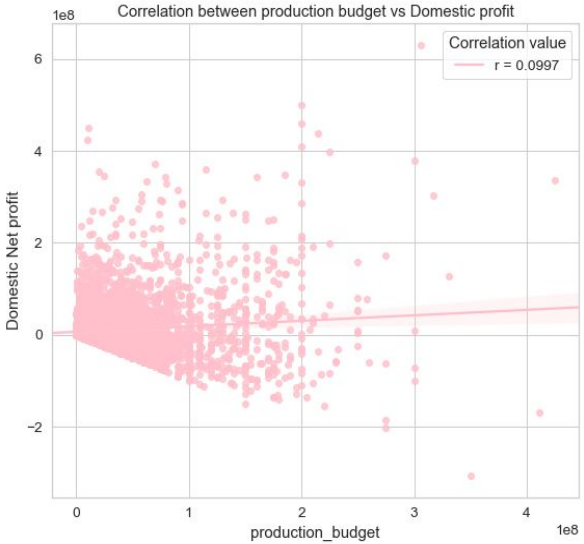


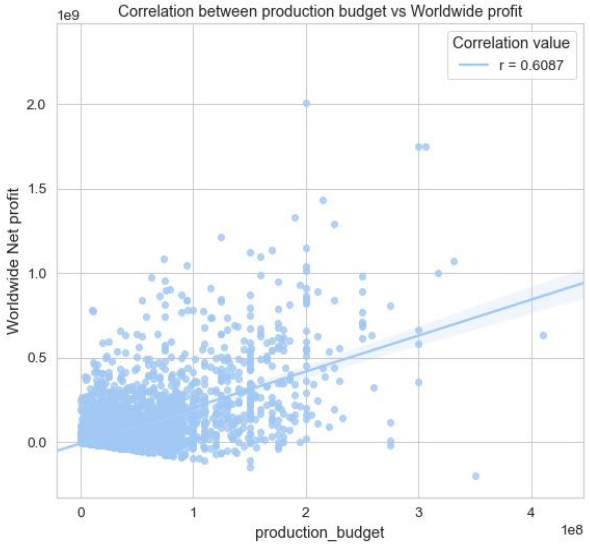
4. Typical production budget: All movies : 15-20 millions, Top 20% most profit: 30 millions

5. Correlation between budget vs profit:

Budget vs Domestic profit = non existent (r-value = 0.01)

Budget vs Global profit = positively mild - moderate (r-value = 0.61)





Conclusions

Recommendations:

- Genre: Action, Adventure, Sci-fi
- Runtime: 80-100 minutes
- Season to release: Summer (June - August)
- Budget: 30 millions USD (Higher budget does not correlate to higher profit)

Limitations:

- Runtime: did not account for movies ratings < 8
- Currency of data: no data after 2020 was included
- Foreign gross: missing data

Future considerations:

- Movies critics, reviews analysis
- Competitors
- Film crews analysis and recruitment

Thank You!

Email: yen.ho993@gmail.com
GitHub: @NBYHO
LinkedIn: www.linkedin.com/in/yen-ho993