WaeniKakenyi / **Movie-Genre-Recommendation**   Public

<> Code   ⊙ Issues   ⇄ Pull requests   ▷ Actions   ⊞ Projects   📖 Wiki   ⚠ Security   ⬚ Insights   ⚙ Settings

⅄ master ▾   **Movie-Genre-Recommendation** / README.md   Go to file   ···

WaeniKakenyi Updated repository structure   Latest commit eca3abd now   🕘 History

⚇ **1 contributor**

☰   122 lines (95 sloc) | 5.78 KB   <>   🗎   Raw   Blame   ✏ ▾   ⧉   🗑

Movie-Grenre-Recommendation

Overview

Microsoft intend to enter the movie studio scene and they have undertaken to research the industry to ensure that the content they create is marketable. They have charged the Data Scientist to give them insights on the market trend so that they can make a decision on what types of films to create. This will be achieved by analyzing what types of films are doing best at the box office and translating the findings into actionable insights that can then be rendered to the head of Microsoft's new movie studio.

Problem Statement

- The main goal of the project is to find the types of films doing best at the box office. This shall be achieved by analyzing the following:

- What genre would be recommended
- Who are the target audience
- Recommended directors for movie
- Movie length recommendation
- Who the competition is

Data

- The datasets in use were collected from the following sites:
 - Box Office Mojo
  - IMDB
  - The Numbers

Methods

- The datasets were imported and the dataset preparation was done.
  - Missing were handled by dropping affected columns or filling with the mean of the data depending on the scenario
  - Duplicated data were dropped and only the first instance of the duplicates was kept
  - The cleaned data was used to analyze and answer the problem statement
  - Data visualization methods were used to answer the problem statement with visuals
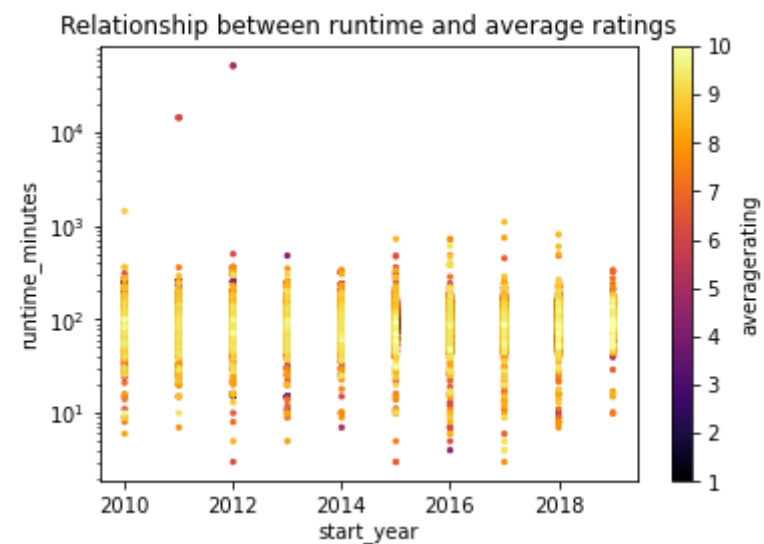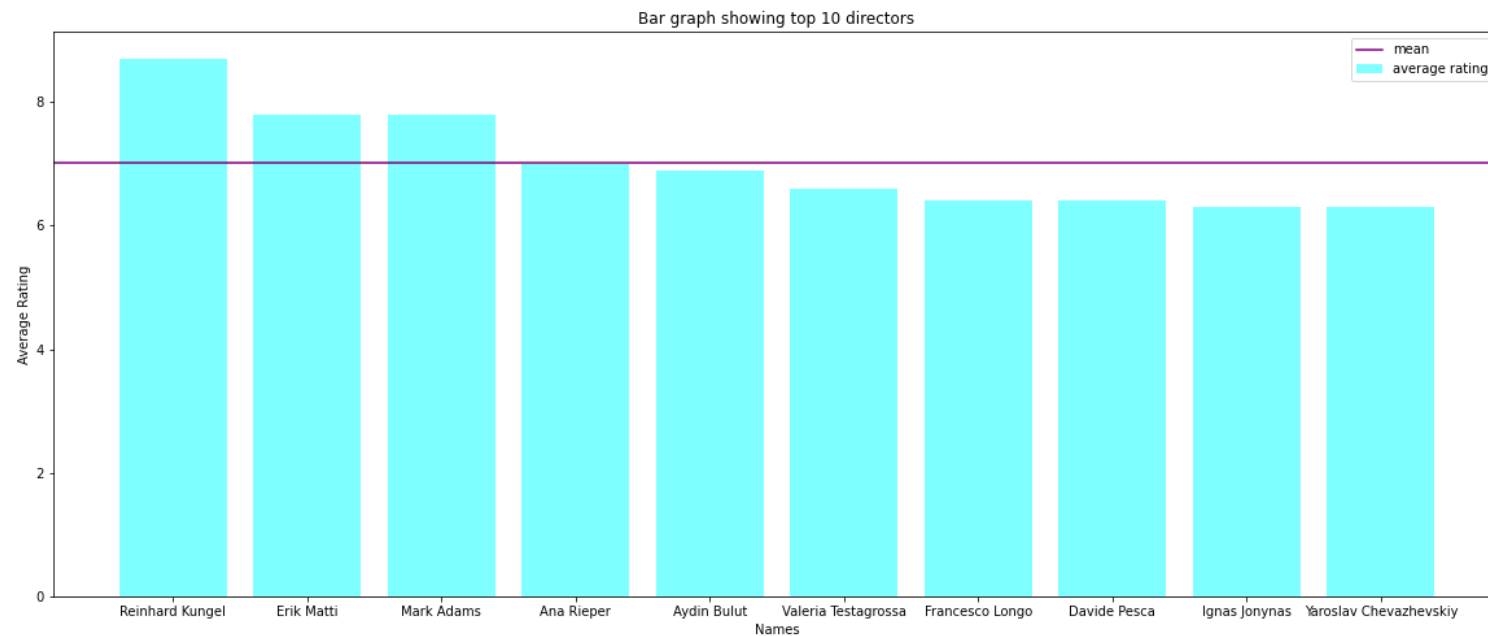 - A recommendation was made

Results

The genre with the highest rating was Family
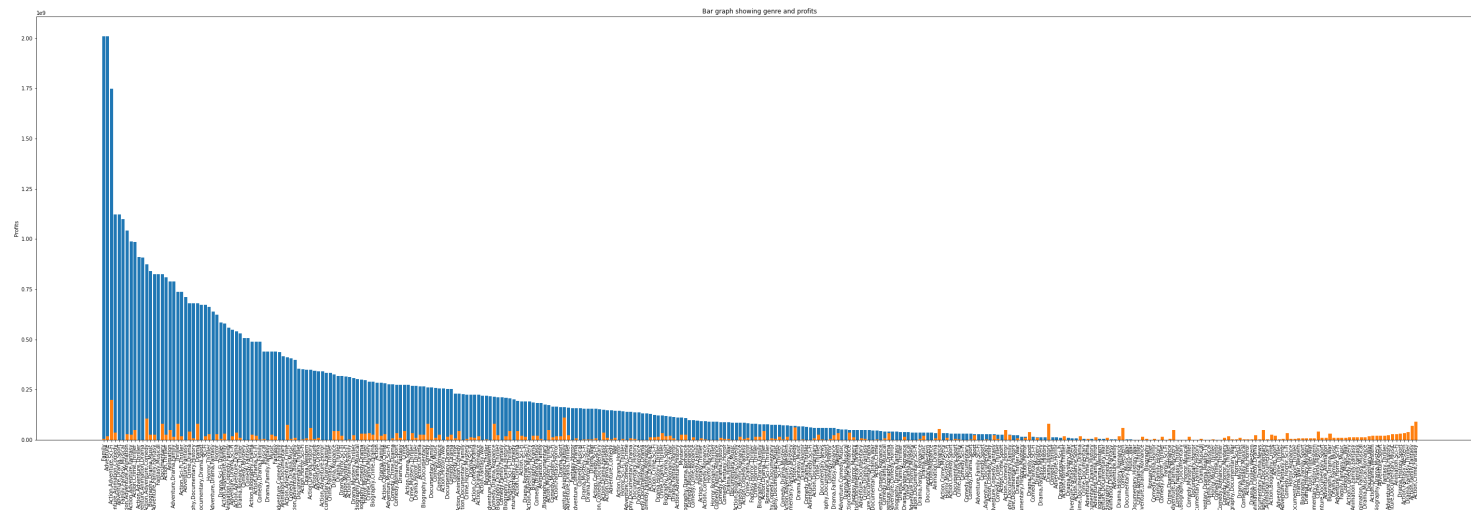
The movies released for sale to the worldwide audience did better than those released domestically
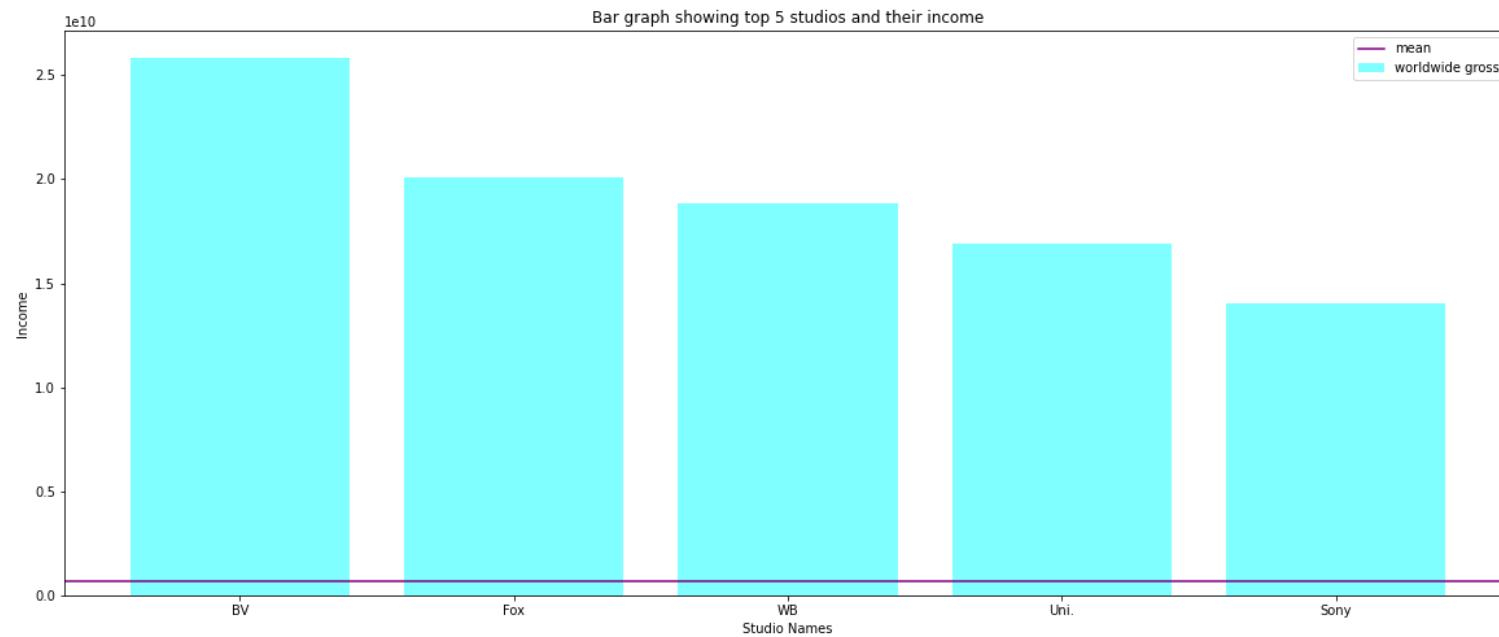
The top 10 directors were found to be

The recommended movie rating was between 40 to 120 minutes

The competitors as the top 5 studios

- The recommendation above would propel Microsoft to enter the movie production scene with confidence as well as the tools necessary to ensure they compete with the top 5 studios.

Conclusion

Microsoft recommendations for their new movie studio would be:

1. Genre Recommendations

These recommendations were based on the profits and losses made on each genre

- For the single genres, the following would be best:

- Family
- Adventure
- Fantasy musical

- For multicategory genres, the following would be best:
  - Action, Adventure, Scifi
  - Adventure, Animation, Comedy
  - Adventure, Drama, Sport
- Categories that brought losses:
  - Action, Adventure, Scifi
  - Adventure, Animation, Family
  - Action, Adventure, Comedy

2. Target Audience

The recommended audience would be worldwide as they brought in significantly more income to the company.

It is seen that the difference between foreign and worldwide sales was 54953833.13705276.

This means releasing the movies to the domestic audience would incur in losses when compared to releasing to the world audience.

3. Directors to be considered for the movie productions

The director recommendation is based on the average ratings they have

- The following directors would be best for consultation or hiring when working on the different films:
  - Reinhard Kungel
  - Erik Matti
  - Mark Adams
  - Ana Reiper
  - Aydin Bulut

- Valeria Testagrossa
- Fancesco Longo
- Davide Pesca
- Ignas Joynas

4. Movie Length

The movie length should range from 40 to 120 minutes. This is because movies that lie in this range had better ratings through the years compared to other times

5. Competition

- The top studios that they would be in competition with are:
  - Studio BV
  - Fox
  - Warner Bros (WB)
  - Universal Pictures (Uni)
  - Sony

Next Steps

- Further analyses could yield additional insights to further improve operations at Microsoft:
  - Web scraping to obtain current data from the movie sites. This would help to do analysis based on current events
  - Narrowing down to the specific genres needed. This would help rank specific movies better even those with a combination of genres.
  - Making a machine learning predictive model that would analyze the current movies and offer a suggestion on the route to take. This would automate the process and the genre recommendations would be done faster.

For more information

See the full analysis in the Jupyter Notebook or review this presentation.

For additional info, contact Medrine Waeni at medrinewaeni@gmail.com

Repository structure

```
├── data                                <- Both sourced externally
├── images                              <- Both sourced externally and generated from code
├── .gitattributes                      <- Contains Git LFS of im.db as it was too big to push
├── README.md                           <- The top-level README for reviewers of this project
├── Writeup.pdf                         <- PDF version of the write up
├── microsoft_movie_analysis.ipynb      <- Narrative documentation of analysis in Jupyter notebook
├── notebook.pdf                        <- PDF version of Jupyter notebook
├── presentation.pdf                    <- PDF version of project presentation
```