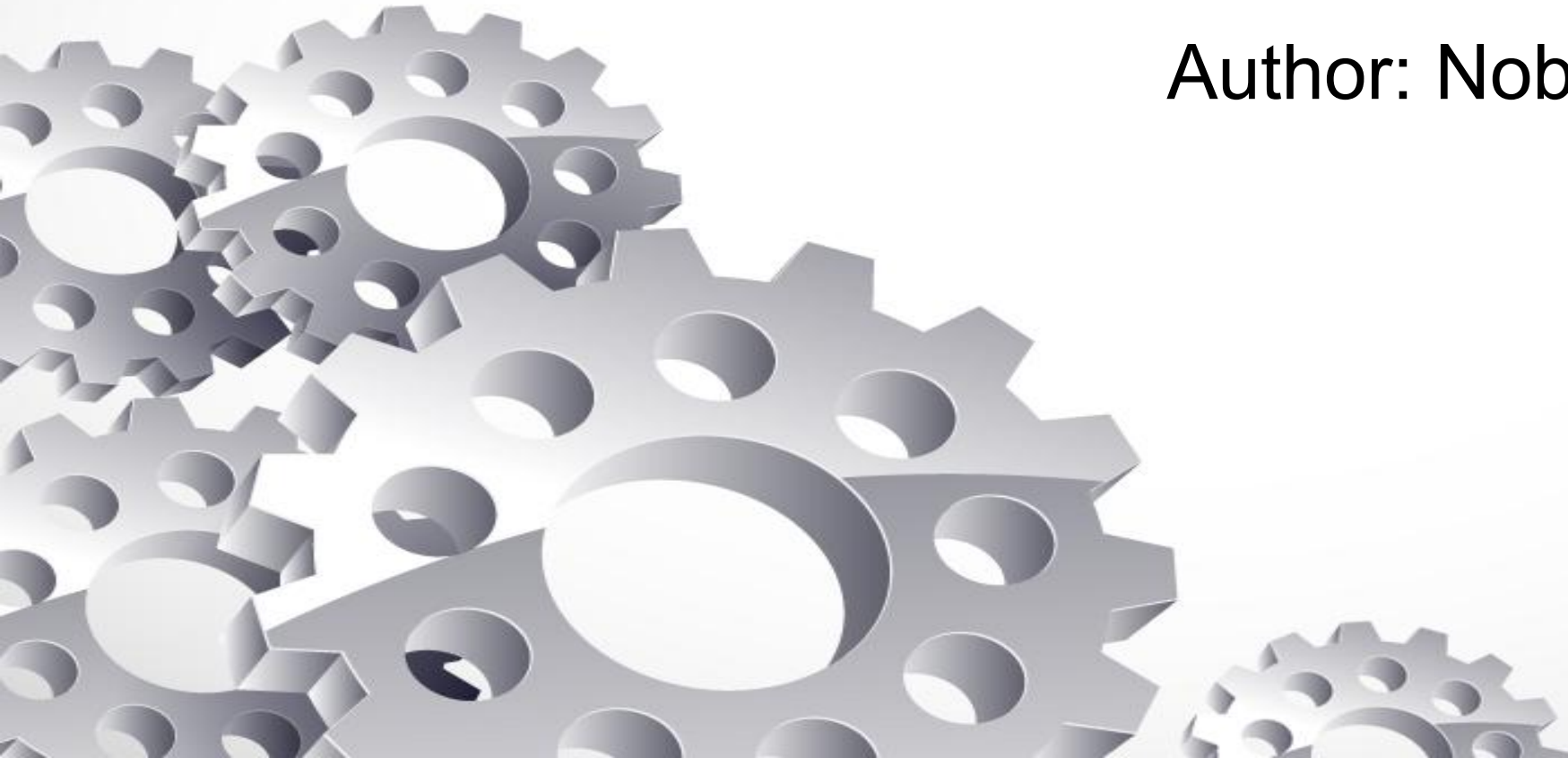


Data Science Phase 1 Project

Author: Nobert Akwir Ogwel



Business Background

- Microsoft is an American multinational computer technology corporation, started in 1975. Its main business was in developing, licensing, and supporting a range of software products and services catering to different requirements. Its big break through came when a partnership was formed with IBM, which resulted in microsoft providing a crucial operating system for IBM PCs. Since then, they have been providing operating systems for most of the PC's, at all consumer levels. They also entered the gaming and mobile phone market (i.e Xbox and windows mobile OS) and were succesful in capturing a large market share. They have also ventured into the videotelephony market with the aqcuisation of Skype and recently moved into cloud computing with Windows Azure Platform.



Problem Statement

- Microsoft sees all the big companies creating original video content and they want to get in on the fun. They have decided to create a new movie studio, but have no clue on how to create movies that'll do good in the boxoffice. We therefore are tasked with exploring what types of films are currently doing the best at the box office then translate our findings into actionable insights that will help the head of the new movie studio, decide what films to create.



Data



We used data from movie ratings sites, box office mojo and IMDB. The datasets span from the year 2010 to 2018.

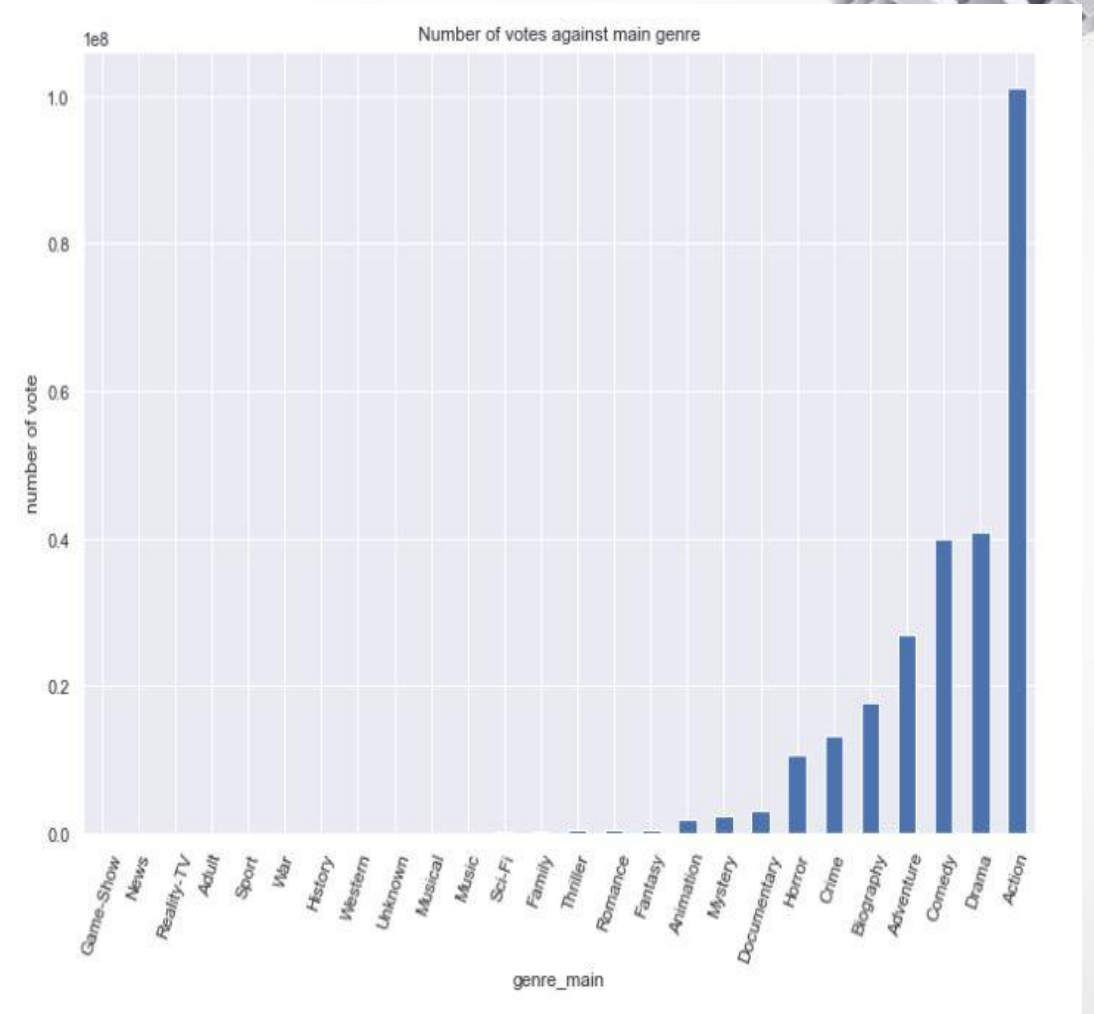
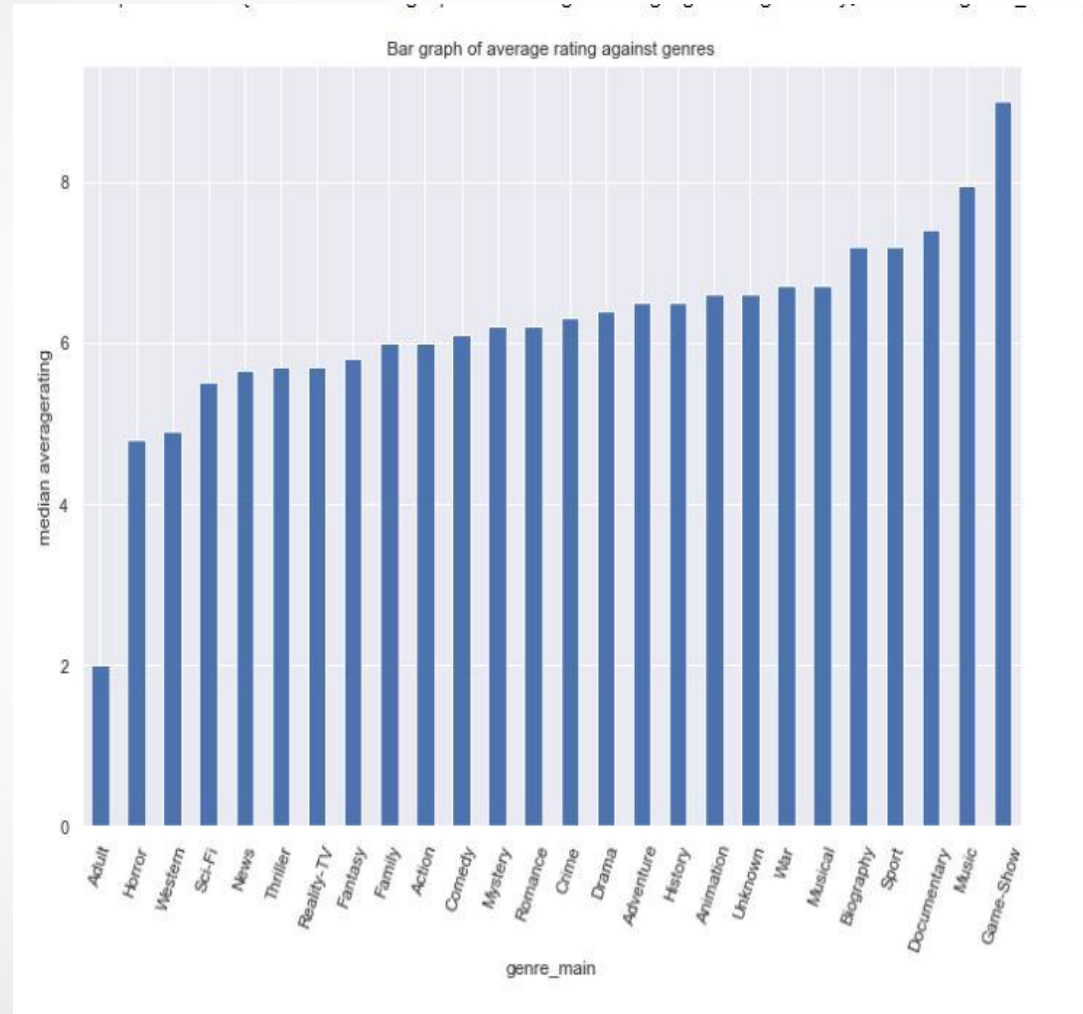
- * [Box Office Mojo](<https://www.boxofficemojo.com/>)
- * [IMDB](<https://www.imdb.com/>)
- * imdb.title.basics(containing the films information)
- * imdb.title.ratings(containing the films ratings)
- * bom.movie_gross(containing the films gross revenue)

Methods

We used both descriptive and exploratory analysis to gain useful insights from our datasets in order to make important recommendations to the Microsoft team on which films they should create.

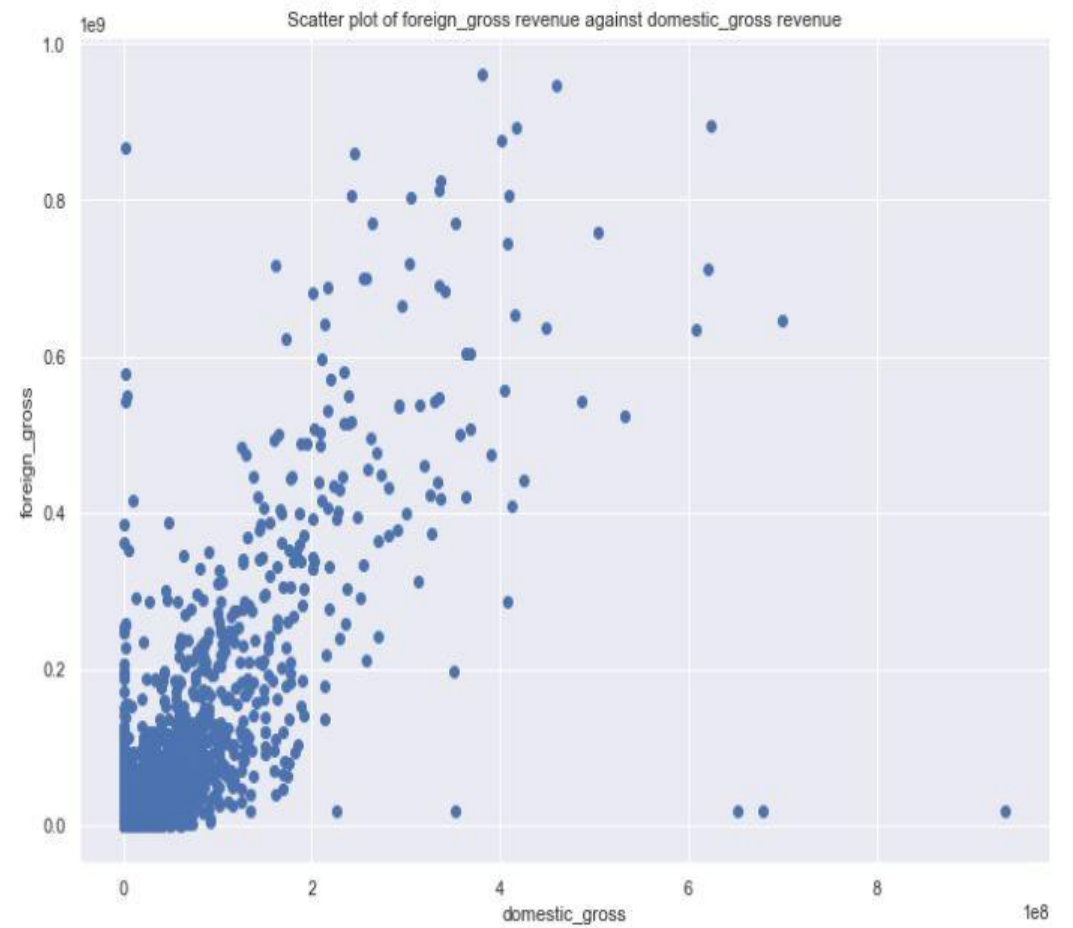
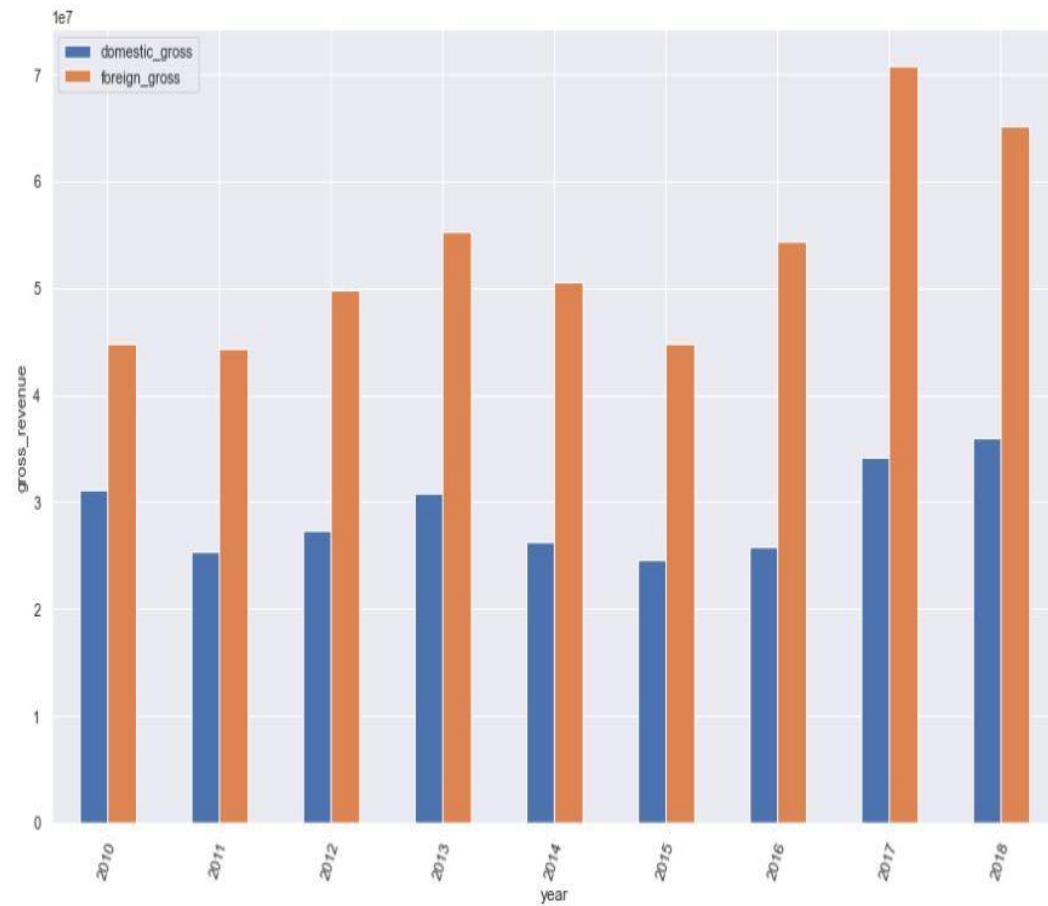


Results





From above bar graphs, we get to observe that the long films have the highest average ratings. But this is affected by the number of votes since very few people watch them. Therefore the second graph shows that the feature films have a lot of votes therefore their average ratings might be really a true picture of the whole population.





From above visualisations, we observe that the foreign market gross revenue is higher than that of the domestic market. But we have to put into consideration that the foreign market is quite large compared to the domestic market hence the disparity. The scatter plot put this into perspective since it shows that the relationship between these two markets is strongly positively correlated.

Conclusion and Recommendation



This analysis leads to three recommendations that will help the head of the new Microsoft movie studio decide what films to create.

- They should focus on creating long films. These are films with run times greater than 210 minutes. From our analysis, this type of film had the highest average ratings therefore meaning that the consumers of the subject films really enjoyed them.**
- As much as the foreign market brings in more gross revenue from the films, they should try focus their resources in both markets. This is because of the strong positive correlation between the gross revenues of the two markets. Also, they should create films that really capture audiences from both markets since the gross revenues in the two markets might have huge differences but we have to put into consideration that the foreign market is very large compared to the domestic one, hence the disparity.**
- As much as we observe that from average ratings, the longer films are highly rated, from the number of votes, we see that the feature films (60mins < runtime < 210mins) have the highest number of votes. This translates to more people watching the subject films. Therefore they should also focus on this feature films since they have a large audience hence more people watching.**

Future Plans



In the future, we should work on updated data so as to provide updated insights to the organisation.