**MICROSOFT RESEARCH ABOUT THE FILM INDUSTRY**

**Introduction**

In today’s world data science has emerged to be a critical driver and background of innovation, efficiency and strong informed decision making. It cuts across all range of industries and has become the backbone of most thriving companies in the corporate world. The huge growth in volumes of data and its complexity within an organization has made data science an undisputed tool for collecting, analyzing and extracting valuable insights and unlocking infinite covered patterns within the data

Opening up a new film studio is quite an exciting venture and at the same time tiresome. Tasked to investigate the market for your new film industry, there is a lot of research to do about the industry. The film industry is very broad and has many players in it taking part. Based on my research, key things to take note of in the industry are;

* Revenue streams- this can include revenue generated from ticket sales in cinemas, which is split between studios and movie theatres
* Value chain – what is comprised here is production of films, distribution and marketing. The last two items can affect the rate on investment depending on how you do it.
* Market trends- film industry is diverse and wide, sticking to one category can throw you out of the business. There are many categories of films out and each has its own uniqueness.
* Key players in the industry- there are giant studios in this line of business and have dominated it over the years and give no signs of stepping out soon.
* Marketing and promotions – effective marketing practices and use of well known platforms are critical for a film success
* Challenges- film industry has a few challenges faced by each and everyone in the industry. Such challenges include competition from film giants, high cost of producing some movies, piracy by third parties

Understanding the above industry dynamics and business aspects is vital for our venture into the industry. Having a good foundation when investing on such a competitive environment is very important.

**Data understanding**

The primary objective of this analysis is to unravel the relationship and underlying trends and patterns within the industry’s data landscape. We seek to know and understand key factors that influence other studios performance in the market, the impact of different genres on films success and the changing behavior of audience over time. These insights offer the potential to optimize decision making processes, from marketing strategies to budget allocation

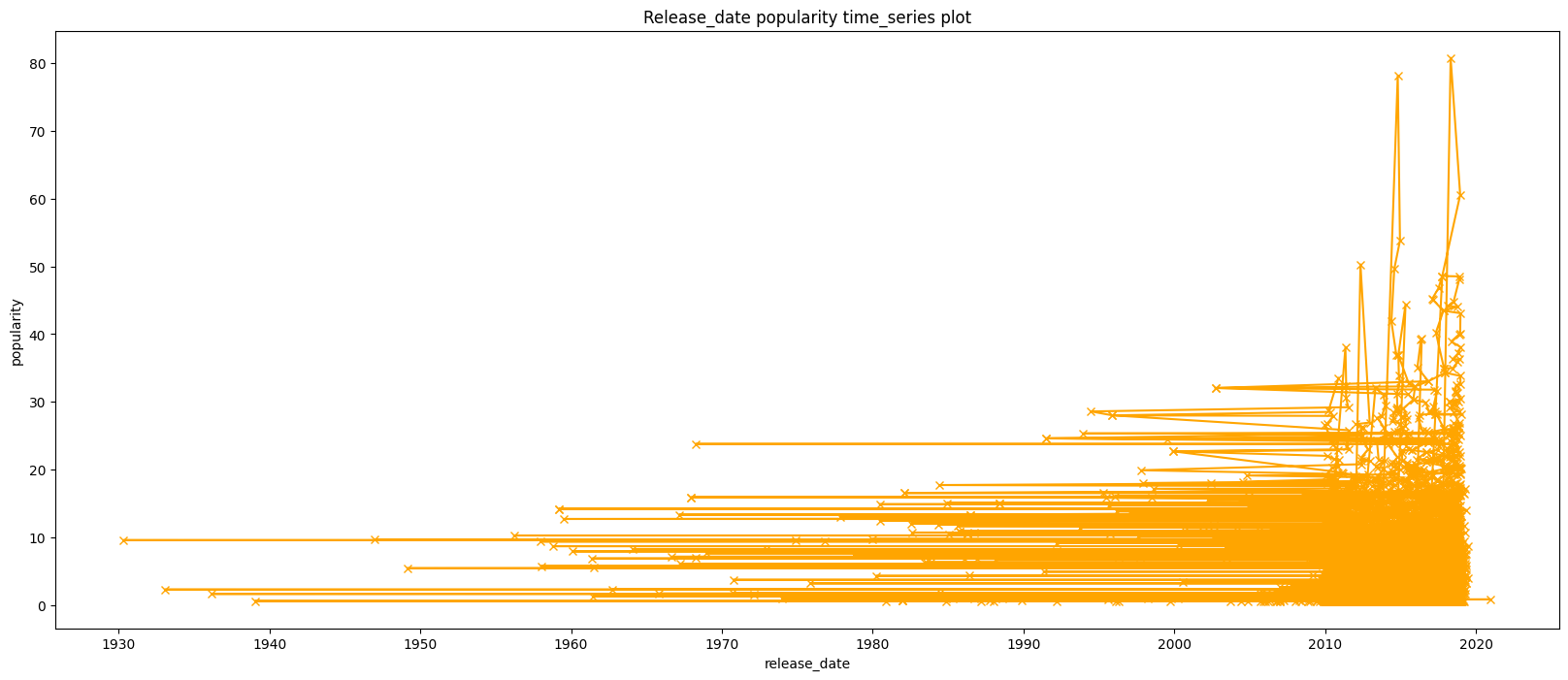
**Data sources and methods**

To make this analysis a breakthrough I have curated comprehensive data that covers a wide range of movies released over the past several years. This datasets compasses of genre, viewer rating, release dates, runtime and budget among other factors. I have employed summary statistics to provide overview over key metrics like genre prevalence. I have also used scatterplots to explore potential relation between variables offering insights which might drive success. I have also used histograms to visualize distribution of some attributes

**Analysis**

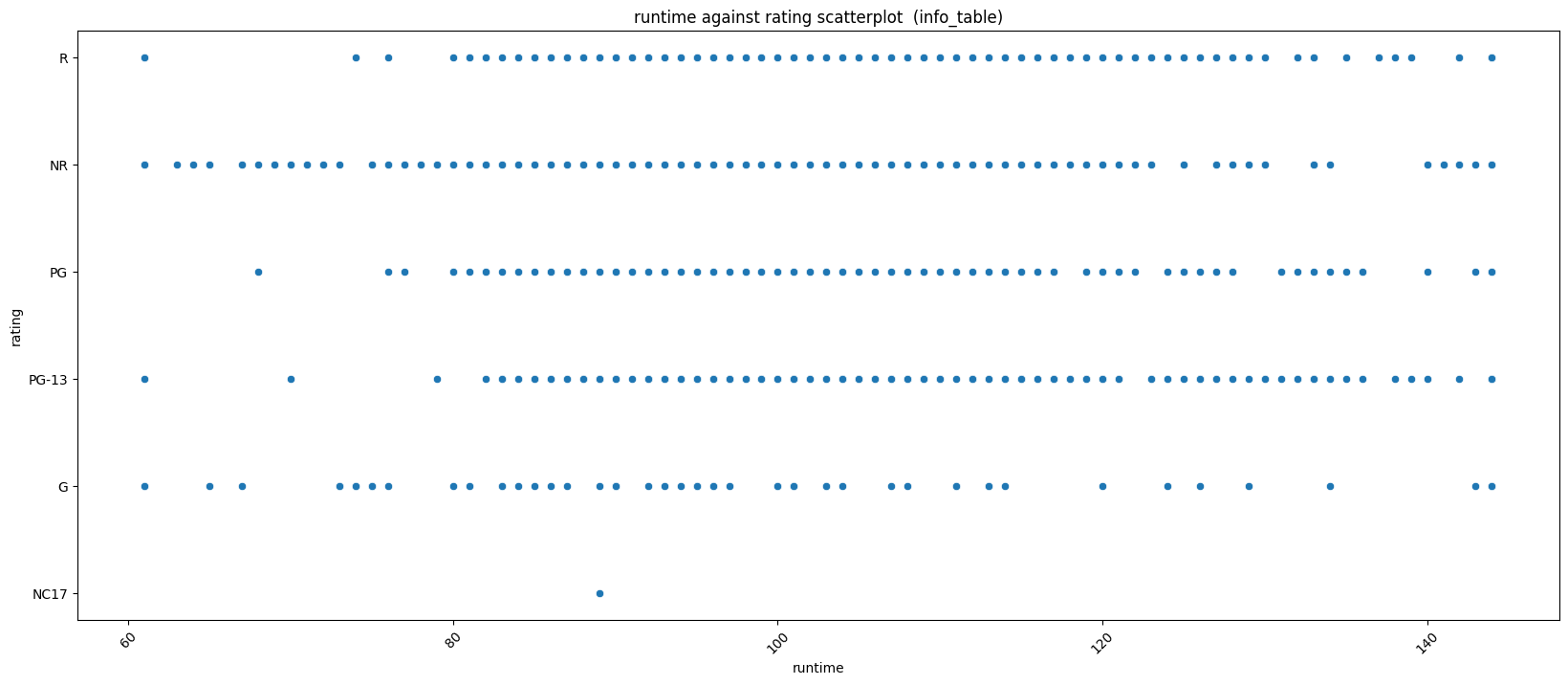
1. Popularity and release date popularity

We can see that popularity of movies increase as time increases. We can see from 1980 going backwards popularity of movies is low, but from 2000 upto the recent years popularity has risen. We can give credit to technology of today’s social media advertisements. Investing on ways of making our movies become popular in the market will increase our return on investment.



1. Genre relationship with viewer rating.

Knowing where most films fall in terms of viewer rating (General, Parental guidance, Rated etc.) can help us know which type of viewer ratings most audience love to watch. This can help us know which types of movies to produce and which viewer rating to attach that movie. From the scatter we can see that;

* We have five viewer ratings; No children under 17(NC17), General, Parental guidance for children under 13 years (PG13), parental guidance (PG), Not rated (NR) and Rated(R)
* Movies with viewer rating of NC17 has one genre, G has few genres, PG13 has little genres almost equal to G but PG, NR and R have the most genres represented there.
* From this data we can depict which genres are more popular with audience and studios in the industry

1. Runtime relationship runtime

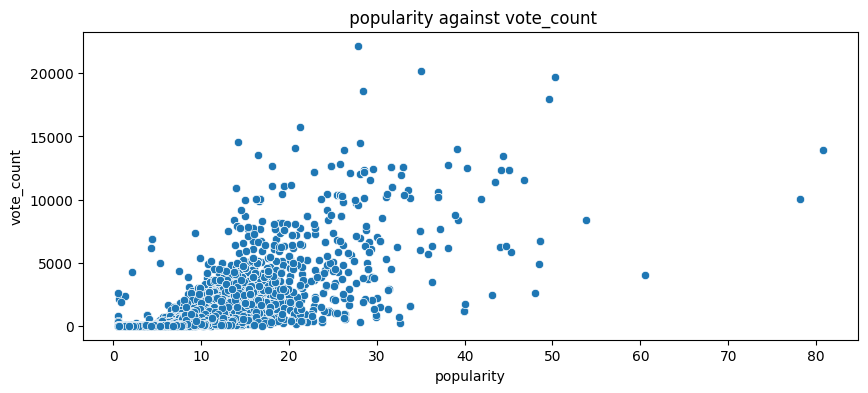
Studying relationship between runtime and viewer rating will help us know the amount of runtime movies to allocate to which viewer rated movies.

* From the analysis we can see that movies with NC 17 has a runtime of 90 min, G is kind of scattered but we can see it averages 75 to 100 min, the PG13 and PG movies are almost similar and have a runtime starting from 80 min to 140 min.
* NR has runtime starting from 60 to 140 taking the most runtime range from the different viewer ratings.
* Not R is similar to PG13 and PG as it has a runtime starting from 80 min to 140 min.
* From the above data we can know which rating to associate with what runtime.

1. Vote count relationship with popularity

From vote and popularity we can see that movies with vote counts of less that 5000 are the most popular and have a popularity ranging from 0 to 30.

From this insights we could further focus our search on this movies to know which genres fall under this category. This can help us have a stepping stone which movies to start producing. Looking at the scatter plot we can see a negative correlation relationship between the two.

While vote count increases the popularity decreases,

**Conclusion and recommendation**

Having seen the relationship between some key factors, we have been able to gain some insight from the data. Even though we have not conducted analyses on key factors like budget and return on investment, we have gotten a glimpse of what to start with when entering the market.

Here are few recommendations I would like to make based on and outside my research;

* I would recommend that you invest on sales and marketing strategies including targeted advertisements to build anticipation for upcoming releases.
* Based on the research I have done, I have seen a potential on the market if you invest your money on the movies with viewer ratings of PG, NR and R as the y have most category of genres meaning that it Is more popular in the market.
* Pick genre that has low budget and mostly runtime of one and half hours.

**Challenges**

* I had less time to conduct my research and would ask for more time in the future
* Reading data from huge data sets which requires a lot of time

**Next steps**

1. Explore and investigate Microsoft’s products and their market share.
2. Have better visualizations that offer more insight
3. Explore other records from other databases
4. Advice more on return on investment.