

Data Science 101

Basic DataScience Project Using [Movielens Dataset](#)

About the Dataset:

- GroupLens Research has collected and made available rating data sets from the MovieLens web site (<http://movielens.org>). The data sets were collected over various periods of time, depending on the size of the set.
- Stable benchmark dataset.
- 20 million ratings and 465,000 tag applications applied to 27,000 movies by 138,000 users.
- Source: <https://grouplens.org/datasets/movielens/>

Motivation

I am a movie freak, yes, I am! While exploring the movielens dataset and ideas for data science mini-project I saw a question hundred times that is “Does number of Sci-Fi movie releases per year increased?” and I was like maybe..

But I wanted to dig deeper and actually find the answer of this question.

So I gained my motivation from exploring the web about the basic projects and finally decided to work on movielens dataset.

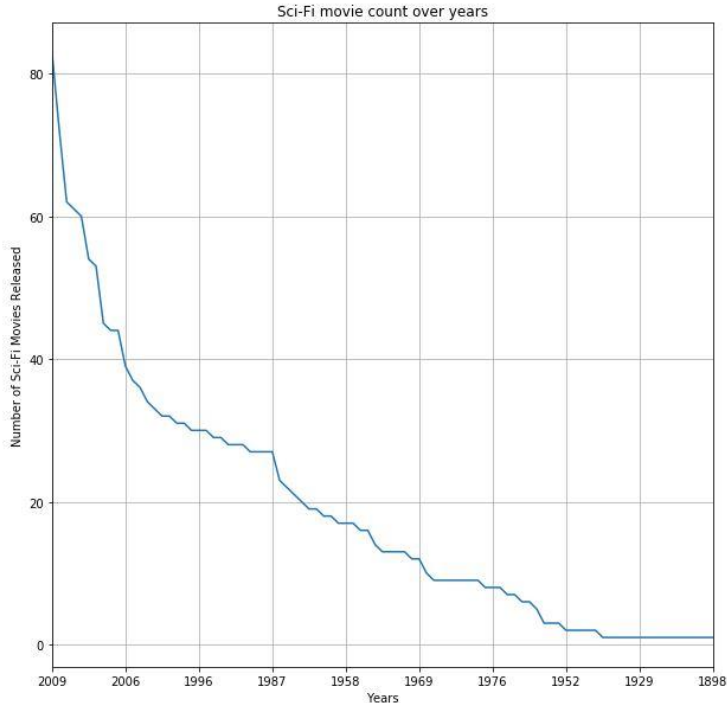
Research Question(s)

Which movie genres are becoming more popular and, are also rated higher over years?

- Does Sci-Fi movie releases are actually increasing over year?
- If the answer of previous question is yes, then movie genre is lagging behind in the race?
- Which genre combos are most popular and highly rated? (JUST FOR FUN)

To Generate Findings:

I plotted number o sci-fi movie released vs years.



AFTER THIS GRAPH

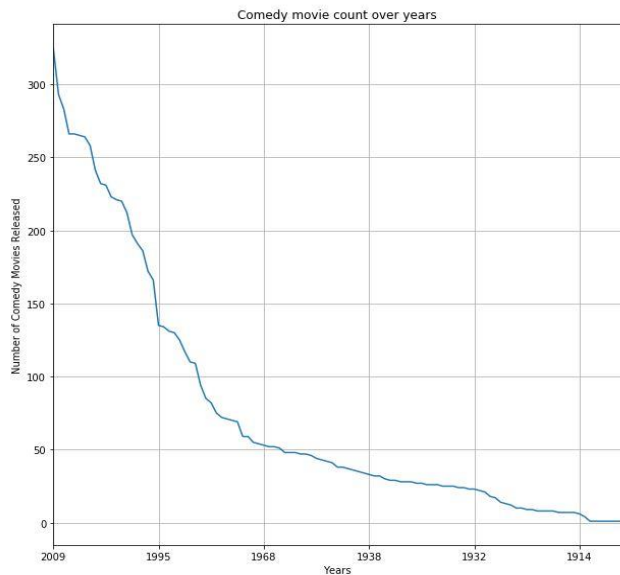
Does number of Sci-Fi movies releases increased over year?

YES!!

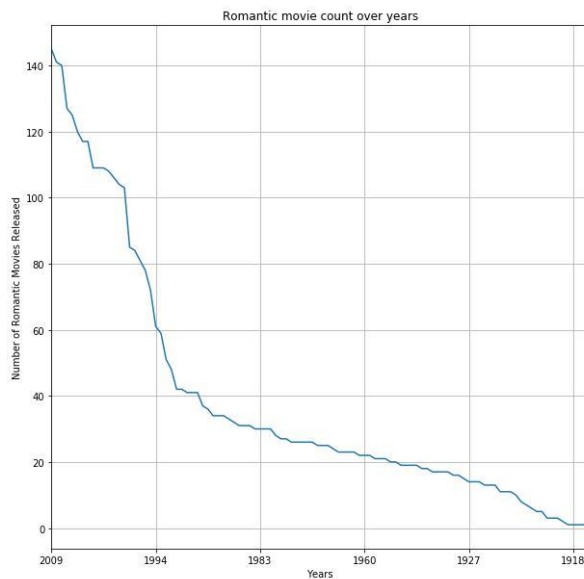
But Wait, just move to next slides first before concluding anything...

Just see these results also...

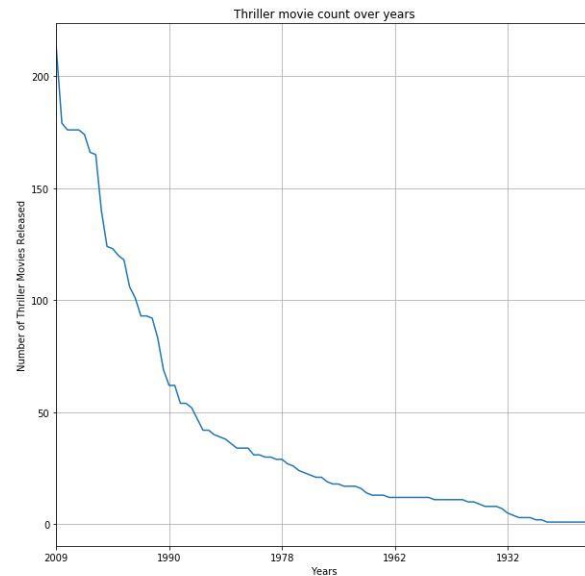
COMEDY MOVIES



ROMANTIC MOVIES

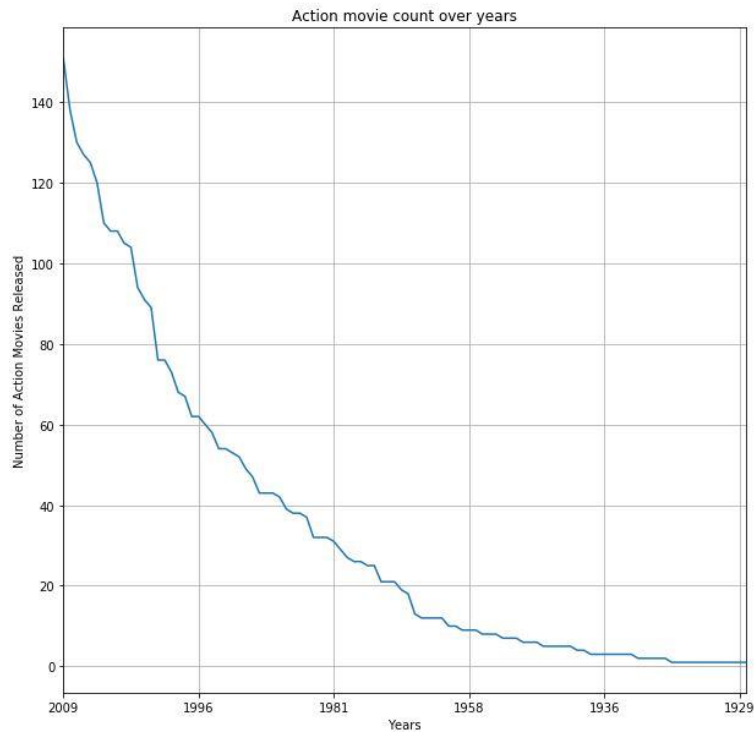


THRILLER MOVIES

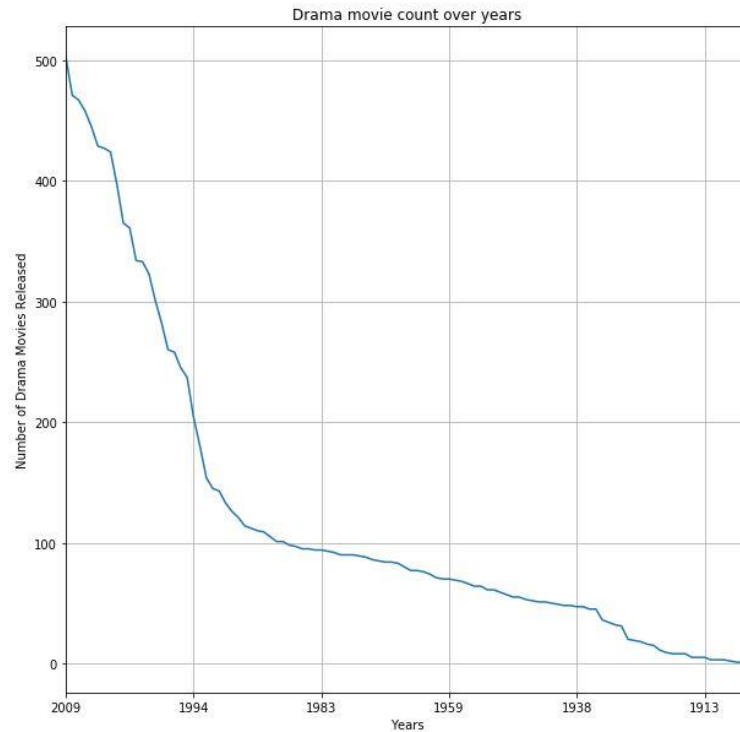


And these...

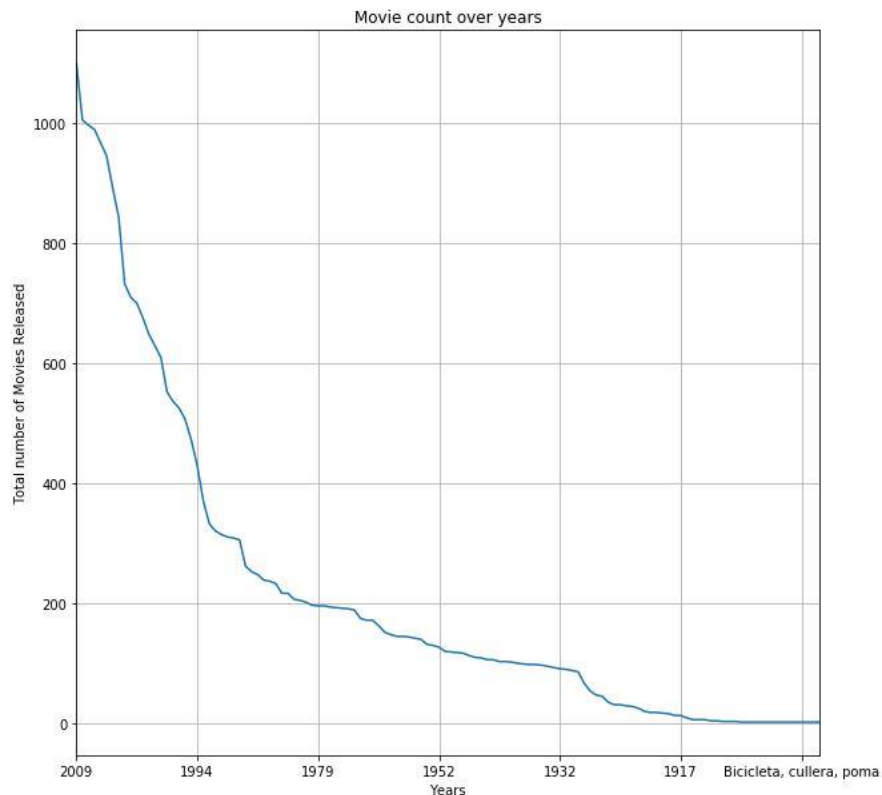
ACTION MOVIES



DRAMA MOVIES



Now see the overall movies released per year.



What does all these means?

When we compare the Sci-Fi graph with all other genres graph , **the shape is similar.**

And the trend in overall movie release graph , **again the shape is similar.**

So what is the answer?

I find that over years:

- Number of overall movie releases per year **INCREASED**.
- And YES, Sci-Fi movie release per year increased.
- But, so does every other genre, I mean **number of movie release of all genres has INCREASED**.
- And, ratio of number of movie released according to genres has not changed!!!

For all those who are here to find

Most popular movie genres combo(by number of movie release till date and then avg. rating) are:

1) *Comedy Drama*

2) *Drama Romance*

3) *Comedy Romance*

References

I only used [Movielens](#) and its [ml-20m.zip](#) dataset for my work.

Also some google searches.

[Click here](#) to download the my jupyter notebook and [click here](#) for the html version.

P.S. All the findings are based on movielens dataset assuming it has ratings for all the movies from the year 1917 and correct genres.

Thanks.