MOVIES AND SHOWS

SQUIRTLE

EREN YILMAZ 181180080 eren.yilmaz1203@gmail.com

MUHAMMED FARUK YILDIRIM 181180077 muhammedfaruk2000@hotmail.com

EGEMEN KARAAYTU 181180043 semih-egemen@hotmail.com

KARIM EMENOV 181180401 N0sferatj3w@protonmail.com

ABSTRACT

In this document, the work we have done on the 'Movies and Shows' project, which is the subject of the project we have chosen, will be mentioned sequentially. First, the information and data obtained as a result of extensive research-datasets will be focused on. Then, the working steps performed in accordance with the "data science roadmap" will be transferred on the selected data sets. After that, Exploratory Data Analysis and Visualizations, machine learning methods used and the results obtained will be given and the report will be completed. This report also summarizes what we have learned and have applied.

1.INTRODUCTION

A film or a show is an entertainment that involves interaction with a user amusement and Hollywood or other production companies.[1] Nowadays, most people have an intense interest in Netflix and platforms such as. According to the sources obtained after the scanning studies, movie genres, platforms, producers, etc. Ratings and gross quantities by genres, as well as data on viewing datas on the IMBD have been accessed. Various analysis, visualization and prediction studies have been carried out with these resources.

2.METHADOLOGY

Before we started working on the selected datasets, their contents were learned in detail. For each dataset, the data to be used was selected according to the information we want to obtain. For example, the information we wanted to obtain in our first data was mostly about the grosses, genre, watchtimes by platform and genre frequency by years.[2] In order for us to analyze this more clearly, the rows with the corresponding null value in the data field were deleted. Then, an encoding study was performed for columns such as, for example, genres, ratings which are categorical data that complicate our analysis. In other datasets, the steps were performed in a similar way and thus clearer results were obtained.

3. DATASETS

3.1 NETFLIX SHOWS

This tabular dataset consists of listings of all the movies and tv shows available on Netflix, along with details such as - cast, directors, ratings, release year, duration, etc.

3.2 IMBD DATASET OF TOP 1000 MOVIES AND TV SHOWS

We used this dataset to reach as much as information about so many movies. These information are such as gross, genres, runtimes.

3.3 DISNEY MOVIES AND TV SHOWS

We wanted to know if Disney+ has more focus on TV Shows than movies in recent years.[3] So we've used this dataset to answer.

3.4 MOVIE INDUSTRY

This dataset is the answer when we want to obtain if movie industry is dying.

3.5 Hollywood Theatrical Market Synopsis 1995 to 2021

This Dataset contains the data of market analysis built on The Numbers unique categorization system, which uses 6 different criteria to identify a movie.

4. EXPERIMENTS WITH THE RESULT

In line with the questions asked about movie industry and the results desired to be obtained, results were obtained with many tables and graphics in the analysis section.

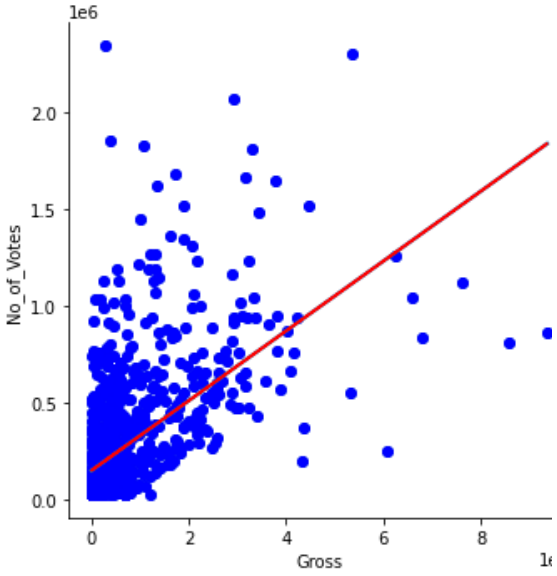

Fig 4.1

We've used regression by using dataset called 'IMDB TOP 1000 MOVIES AND TV SHOWS'.[4]

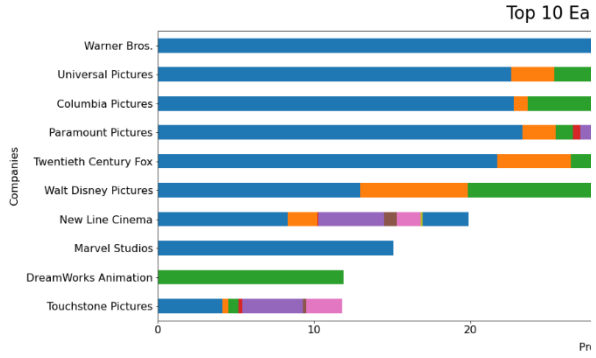

Fig 4.2

As we can see on fig 4.2 we may obtain which producer earns more and also we obviously can see effects of genres.

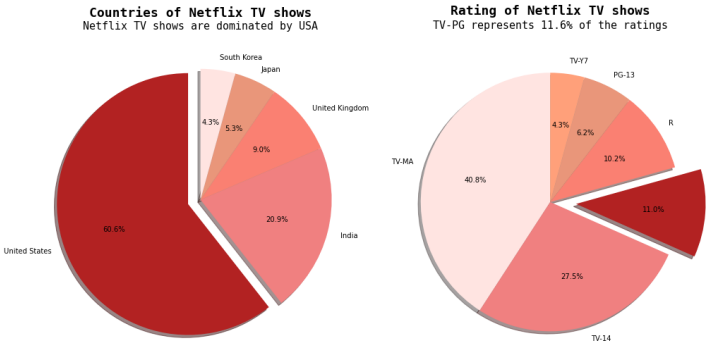

Fig 4.3

With these pie charts we can see regions of netflix shows and also we can observe ratings of these shows.[5]

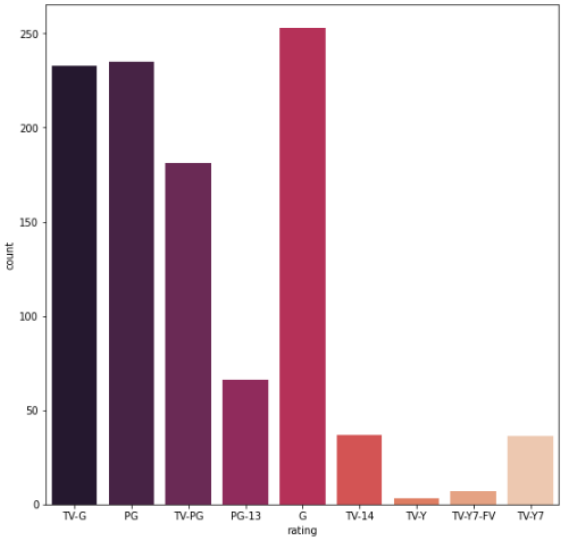

Fig 4.4

These chart shows us the Disney frequency of ratings such as TV-G, PG.

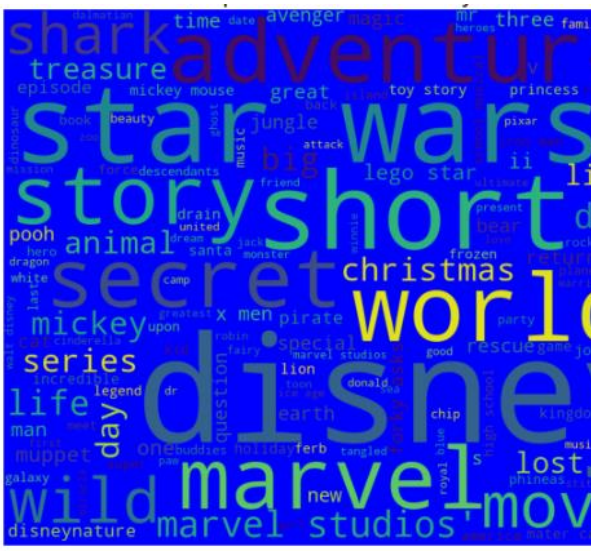

Fig 4.5

Most frequent words in Disney+ titles

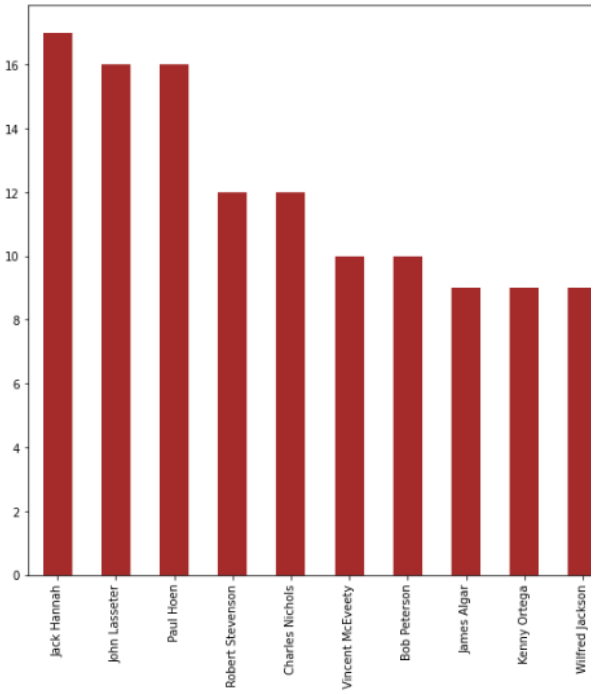

Fig 4.6

Top 10 directors with most disney Movies

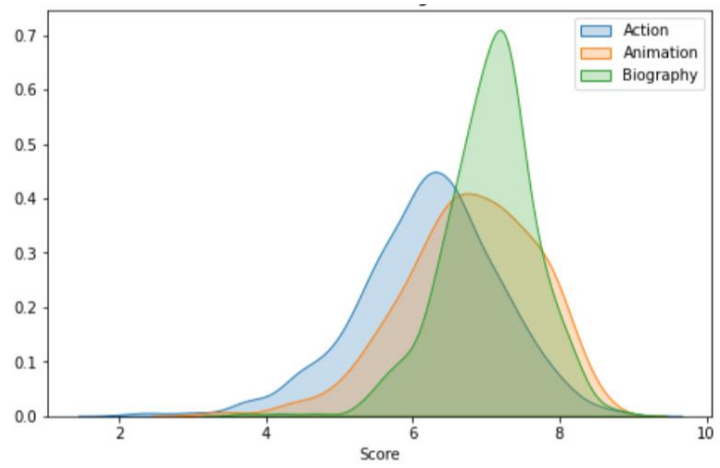

Fig 4.7

A quite beneficial visual which discloses the general audience preference on movie genres

5.CONCLUSIONS

With finishing this project we've pretty much understood how to get data-sets, manipulate and edit them and removed the empty datas.[6] Also before we go into the machine learning part, we got how we visualize our data by using like Python, pandas and seaborn tools. We determined the datasets which we used and rearranged it. We understood the methods we've used which are like regression, visualization. As we research on these technical terms, we had valuable information about Python terms and cinema industry. Then we got the answers to the questions we have prepared so far. Also, besides all of these coding stuff we were curious about lifetime grosses, so we had a research about it.[7] As a result, we can say that we are capable of both Python basics, jupyter etc. and cinema.

6.REFERENCES

[1] Cinema of the United States - Wikipedia

[2] How to Identify Movie Genres: Beginner's Guide to 13 Film Genres - 2023 - MasterClass

[3] Disney+ Movies and TV Shows | Kaggle

[4] IMDB Movies Dataset | Kaggle

[5] Netflix Movies and TV Shows | Kaggle

[6] Data Manipulation in Python using Pandas - GeeksforGeeks

[7] Top Lifetime Grosses - Box Office Mojo