Task_1:

Amazon prime is a popular streaming service that offers a vast catalog of movies, TV shows, and original contents. The data consist of contents added to Prime from 2008 to 2021. The oldest content is as old as 1925 and the newest as 2021 (see attachment prime.csv). The tasks that you need to perform are
• Undertake an exploratory data analysis and produce a 2 page report on your findings/insights
• Clean the data and produce good visualization for story telling
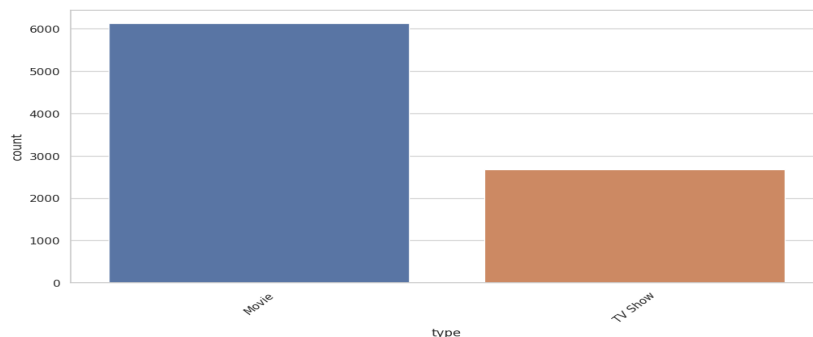• Submit your code and report via Github repo

Solution:
Basically it consists of 8807 rows,12 columns with data about the amazon prime services. As a task 1 listed, I performed an EDA on given data. Found some interesting points about the movies,released year,country and much more. I will show the task one by one.
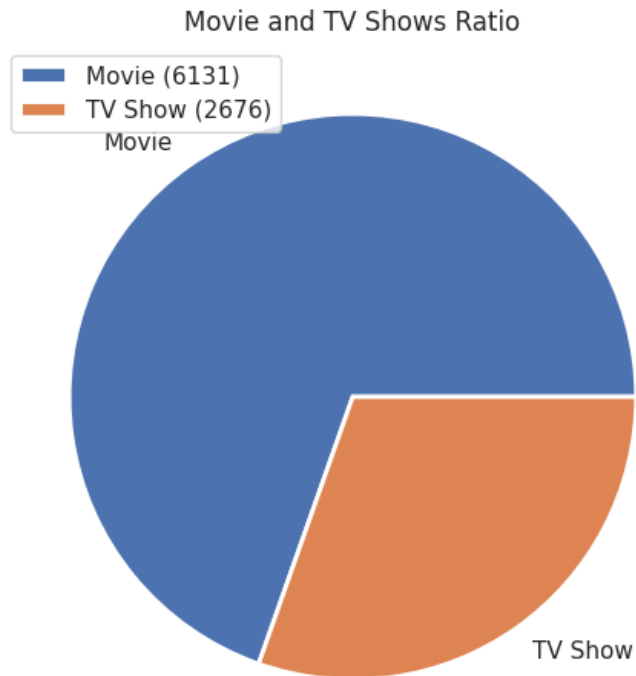Steps:
- The dataset comprises information on content ranging from 1925 to 2021. The initial analysis revealed that there were no missing values or duplicates in the dataset.
- Performed a data visualization in the dataset.Using matplotlib,seaborn,wordcloud(textual data).
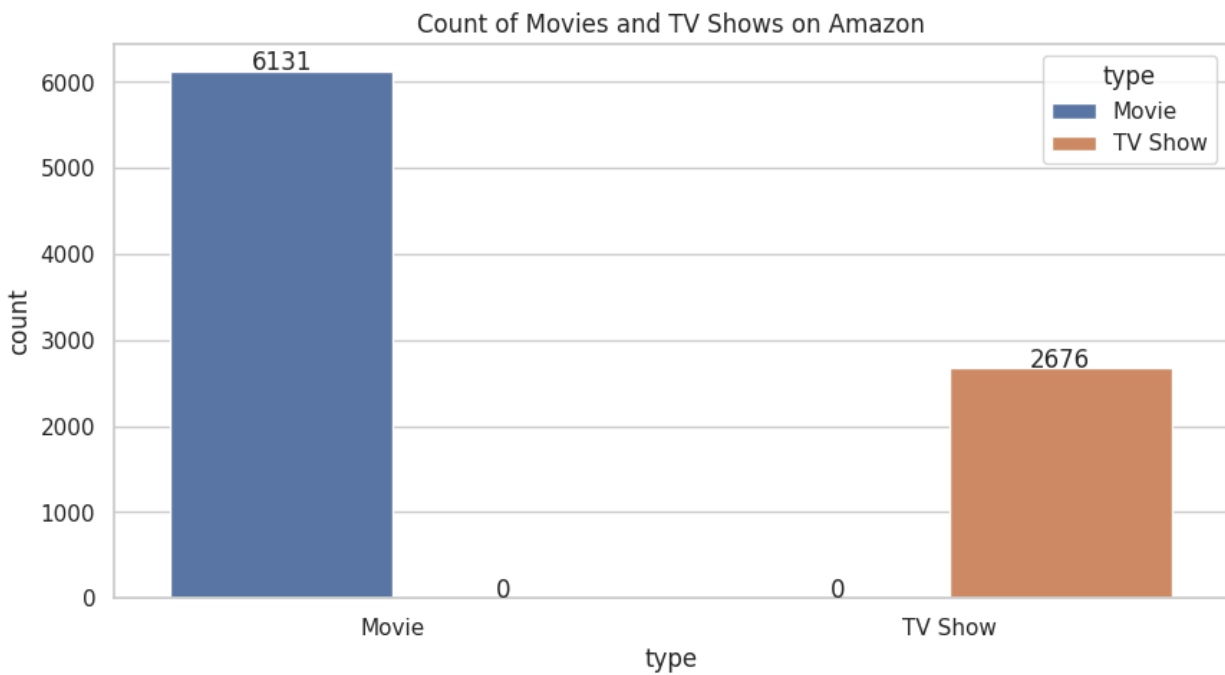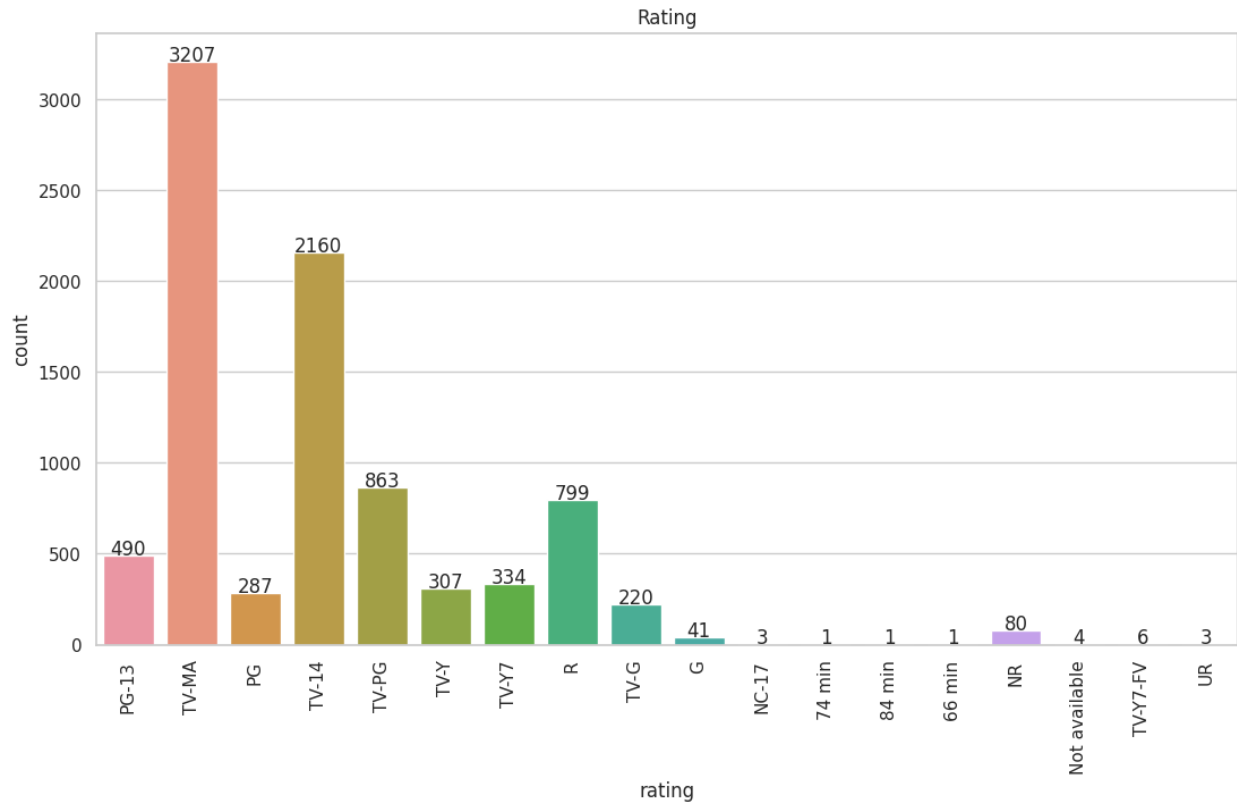
Key Findings:
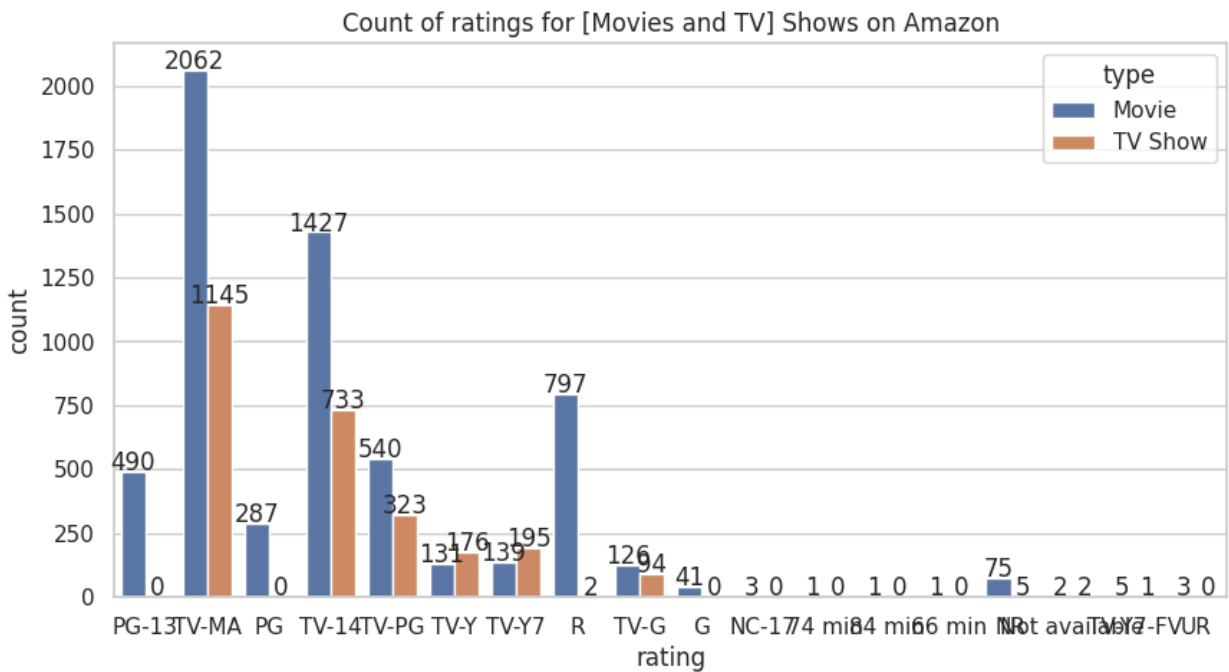

- No of counts of movie,TV shows are shown in the figure

Movie and TV Shows Ratio

Movie (6131)
TV Show (2676)

Movie

TV Show

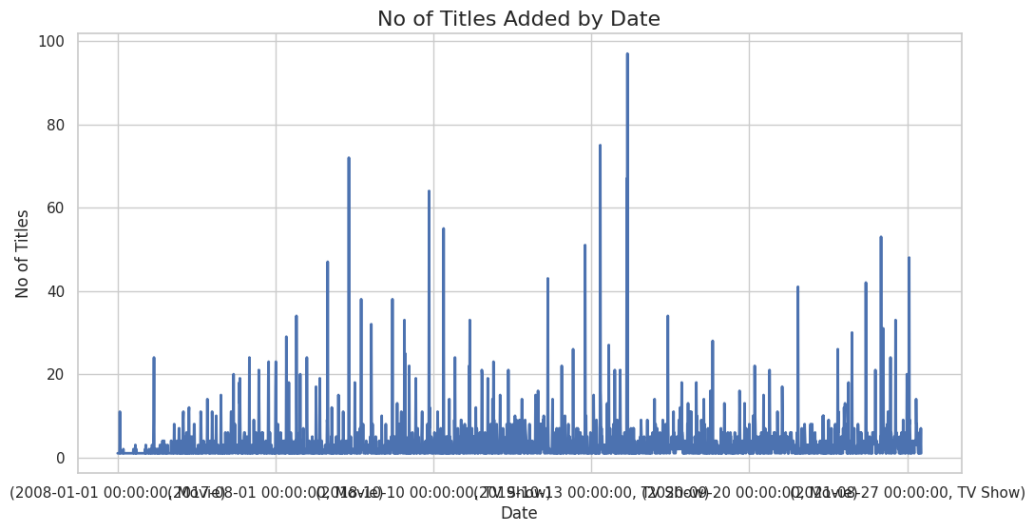- Ratio of both movie,TvShow. It tells you about the subscribers who are interested in watching movie.

Count of Movies and TV Shows on Amazon

Rating

- Distribution of count of the ratings

Count of ratings for [Movies and TV] Shows on Amazon
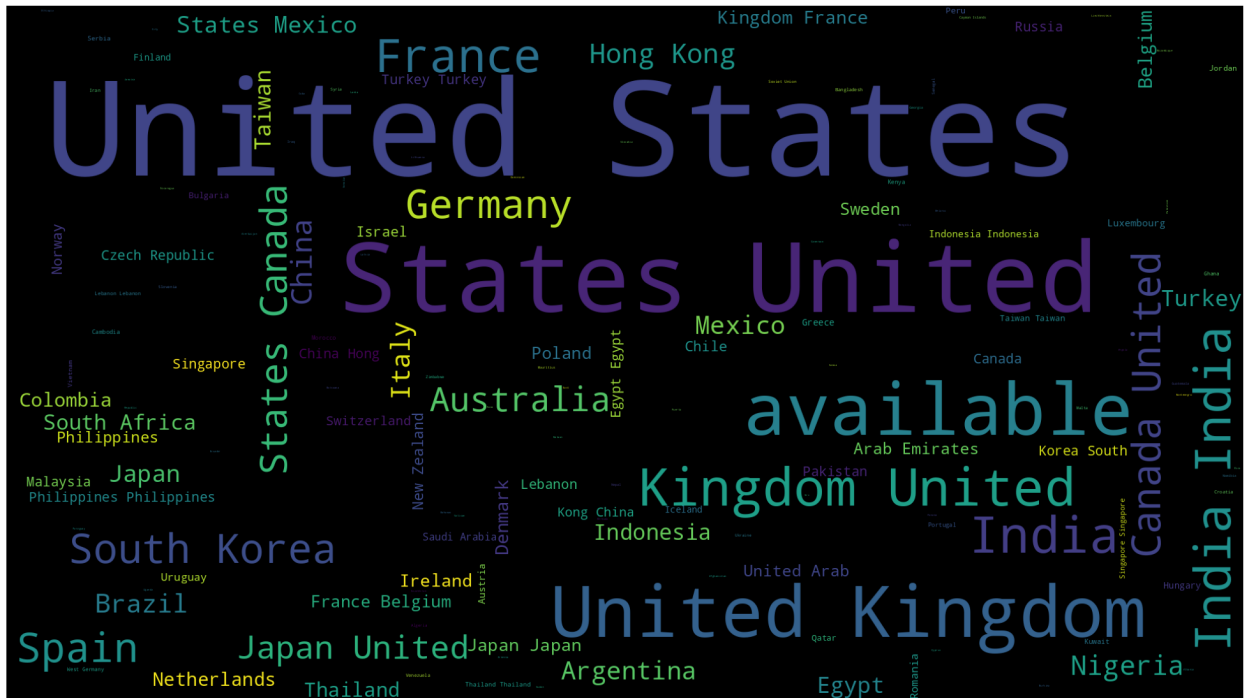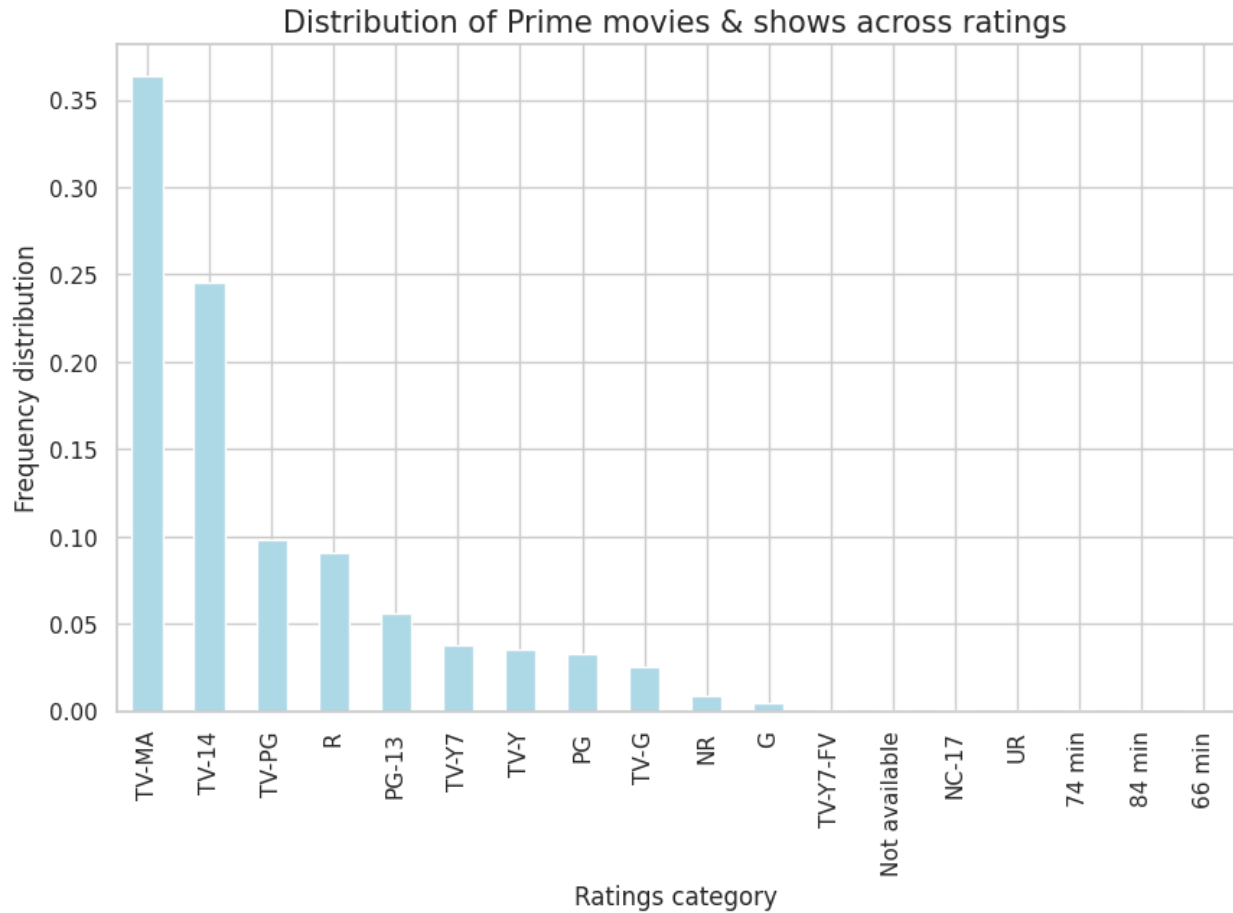
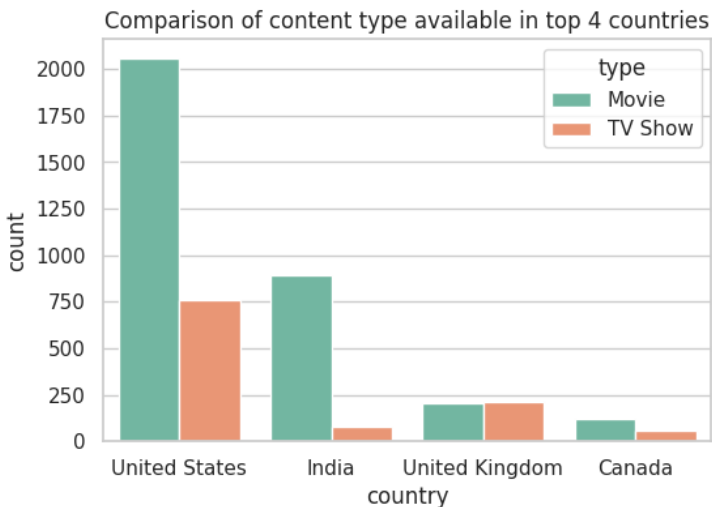- We can see the distribution of count of the rating over the movie and Tv shows
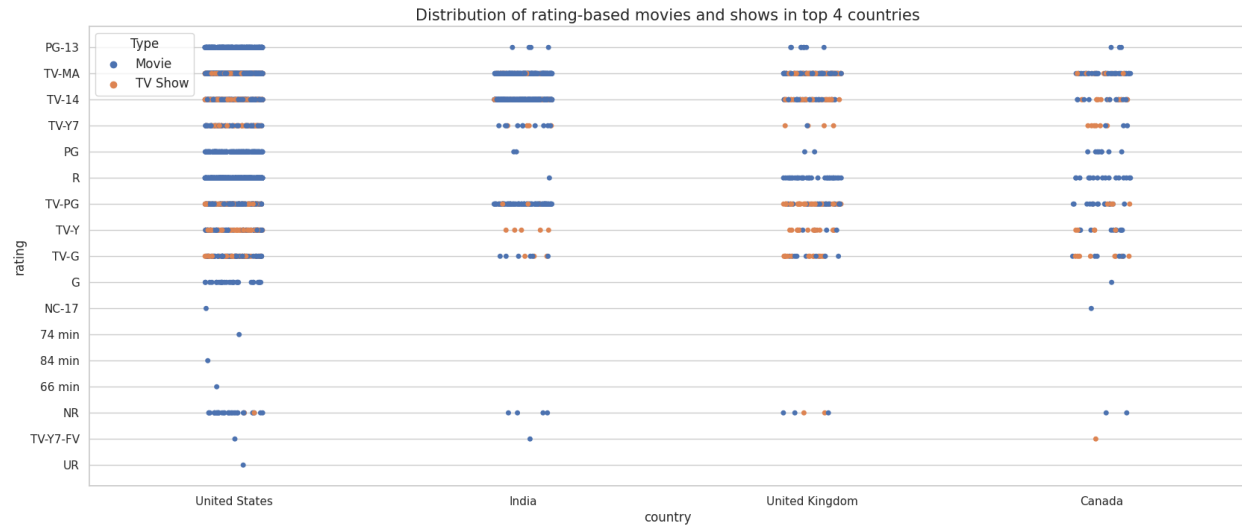
- Perform a word cloud visualization over country

- No of titles (movies, tvshows) added over a period of time

Distribution of Prime movies & shows across ratings

- How well the distribution of prime movies and shows across ratings with frequency distribution.

Comparison of content type available in top 4 countries

- After a comparison, I got the top 4 countries of the count of the view in prime movies and tv shows.

Distribution of rating-based movies and shows in top 4 countries

- From above analysis we can come to a conclusion that in Amazon Prime the content focus has been on Movies for upto 69% and the remaining 33% on TV Shows.

- Out of all the content available on Prime, Movies and shows of rating 13+ are more popular, followed by 16+ and others. To be more specific, we have found that 13+ rated movies are highly produced as compared to the rated TV shows.

- Amazon Prime has the highest content availability in United States, next in line is India, United Kingdom and Canada.