

AMAZON PRIME DATASET

1. ABOUT THE DATASET

- This tabular dataset consists of listings of all the movies and tv shows available on Prime, along with details such as - cast, directors, ratings, release year, duration, description etc.
- Totally it has 8807 rows and 12 columns.
- On observing the number of missing entries in each column of data frame, we can find that there are many null values present in the dataset. Such as director, cast, country date added, release year rating, duration.
- To make it suitable for EDA, Data cleaning needs to be done.
- **Data cleaning:**
 1. Drop the "show_id" column
 2. Check for duplicate records of movies/TV shows and delete duplicates
 3. Replace NaN values in "cast" and "director" columns with "Unavailable"
 4. Replace NaN values in "rating" with mode of the column.
- Once we make sure that our data is ready, move on to EDA.

2. EDA:

- On analyzing the dataset, below is the detailed part of the EDA

Questions:

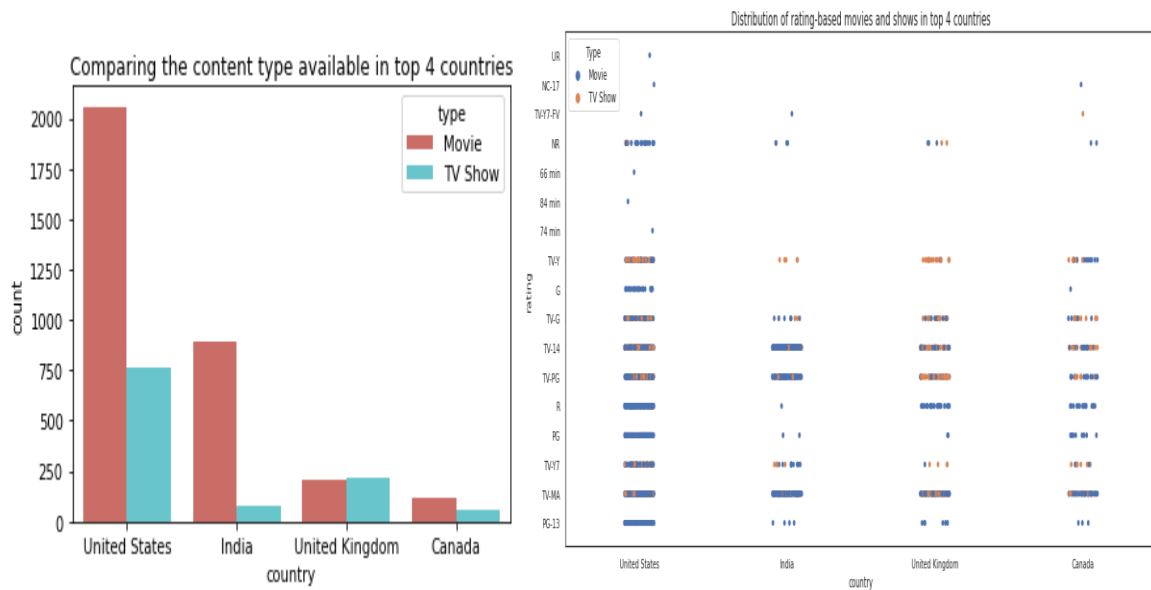
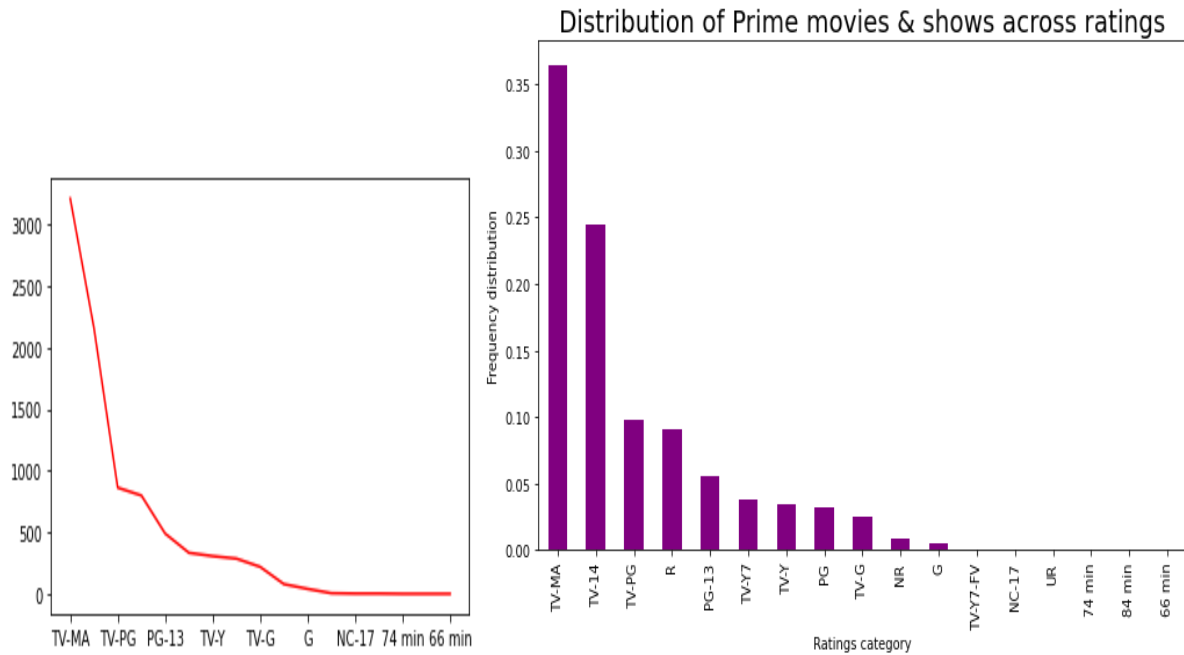
1. What does each category in ratings define?
2. How many movies and TV shows does Amazon Prime have based on each category of ratings?
3. Which are the highest watched rating categories in TV shows and movies?
4. What is the most watched category in each country?
5. Which is more popular on Prime? Movies or TV shows?
6. What was the focus of Prime on type of content in recent years?

INSIGHTS AND INFERENCE:

1. **Rating:** PG-13: Some material may be inappropriate for children under 13. Parents are urged to be cautious. Some material may be inappropriate for pre-teenagers. PG: Some material may not be suitable for children, may contain some material parents might not like for their young children

Here, I have been listed only 2, but there are about 17 unique ratings.

2. On finding the most popular category, TV-MA stands first.
3. Movies and shows of rating TV-MA are more popular in Prime, followed by TV-14 and others.
4. Though there are many countries, I have selected 4 countries for the rest of the analysis.
5. I observe that in India, Amazon Prime has considerably more content in Movies as compared to TV Shows. Next in line is the United States but here the number of TV Shows produced is the highest out of the top four countries.
6. Below are some visualizations for better understanding.



Final Insights:

- From the above analysis I can conclude that in Amazon Prime the content focus has been on Movies for up to 81% and the remaining 19% 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.