TASK 1: Amazon prime

Dataset description:

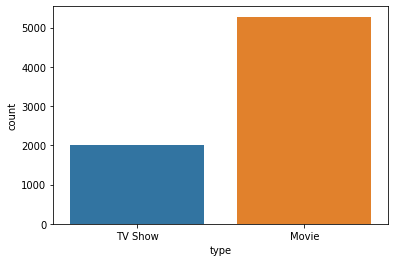
* The dataset is on a list of movies and tv shows in amazon prime. The dataset contains of 8807 records with 12 attributes.
* Out of the 12 attributes, there is 1 numerical attribute which is the release year and the rest are categorical attributes.

Dataset cleaning:

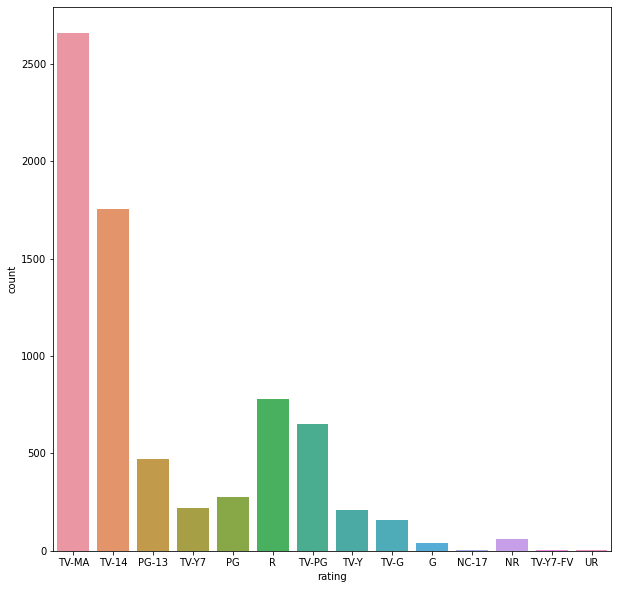
* There are few null values present in the dataset. The number of null values in those columns are; ‘director’: 2634, ‘cast’: 825, ‘country’: 831, ‘date\_added’: 10, ‘rating’: 4, ‘duration’: 3.
* Since director column has more than 2000 records with null values, it is better to impute those, and the other null values are removed.

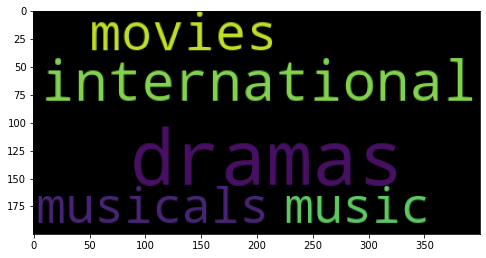
Exploratory data analysis:

There are more than 5000 movies in amazon and less half the number of tv shows which is around 2000.



The respective content control rating is also provided as, from which we see that movies and shows with ‘TV-MA’ rating has the highest count, followed by ‘TV-14’.





From category, we see more of drama category followed by international.

TASK 2: video games sales

DATASET DESCRIPTION:

The dataset contains sales data for video games from all around the world which to help the game developers in understanding the preferred game and the current trend.

The dataset has 1907 records with 13 attributes. Out of the 13 attributes, there are 4 categorical variables and the rest 9 being numerical variables, the rank being the most important variable.

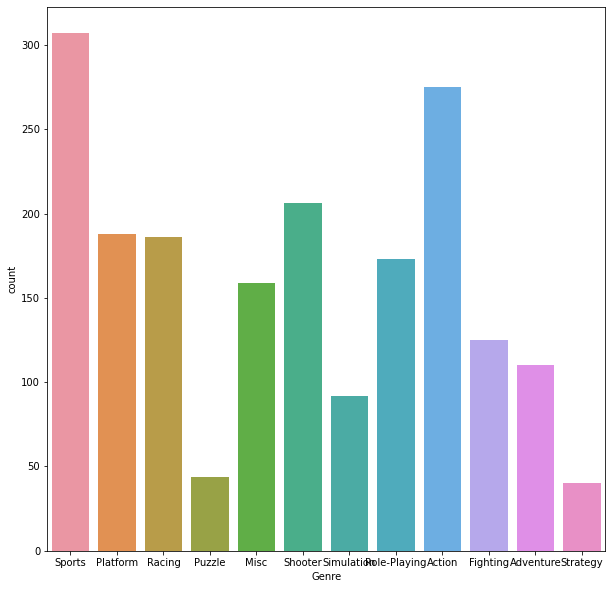
DATASET CLEANING:

There are 29 records having null values under the ‘year’ column and 2 records under ‘publisher’ attribute.

Null values under ‘Year’ are imputed with ‘unknown’ and the rest 2 are dropped.

EXPLORATORY DATA ANALYSIS:

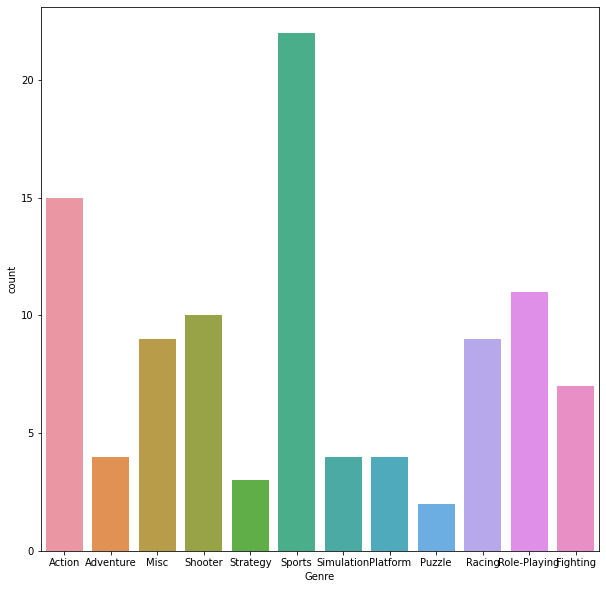
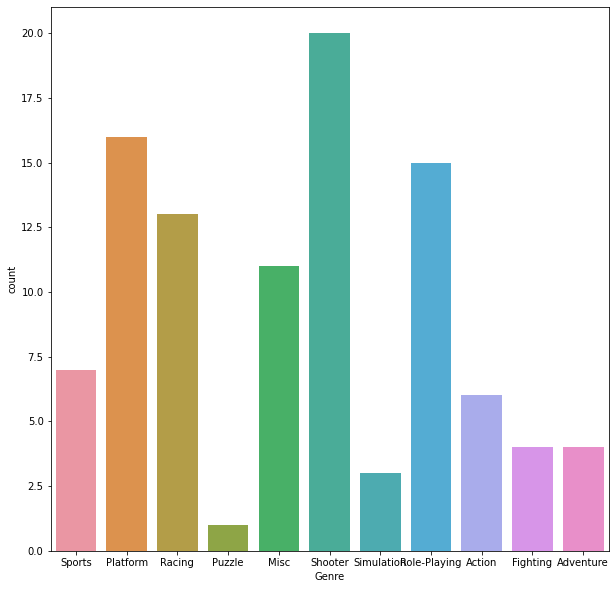
The number of games under each genre can be seen from the countplot. Sports being the highest, followed by action, followed by shooter.



To understand the most liked genre, we can get the first 100 games based on rank. And similarly for the least ranked, last 100 are taken.

For the first 100 ranked games, shooter>platform>role playing is the order of genre preferred and least would be puzzle.

For the last 100 ranked games sports> actoion>role playing will the order of preferrence and lowest again for puzzle.



The top 100 games were mostly published by Nintendo.

The sales average of top 100 games was calculated to be

global: 13.2036

north america: 6.6298

europe: 3.6818999999999993

Japan: 1.8686