**Google Drive Link:** https://drive.google.com/file/d/1L5IaQbMbS5j2GUQ72zcUw3A8UaooWudU/view?usp=sharing

1. Defining Problem Statement and Analysing basic metrics.

Analyze the data and generate insights and give recommendations that could help Netflix decide which type of shows/movies to produce based on the country, cast, directors, genre, duration, etc and how to grow the business.

1. Observations on the shape of data, data types of all the attributes, conversion of categorical attributes to 'category' (If required), missing value detection, statistical summary.

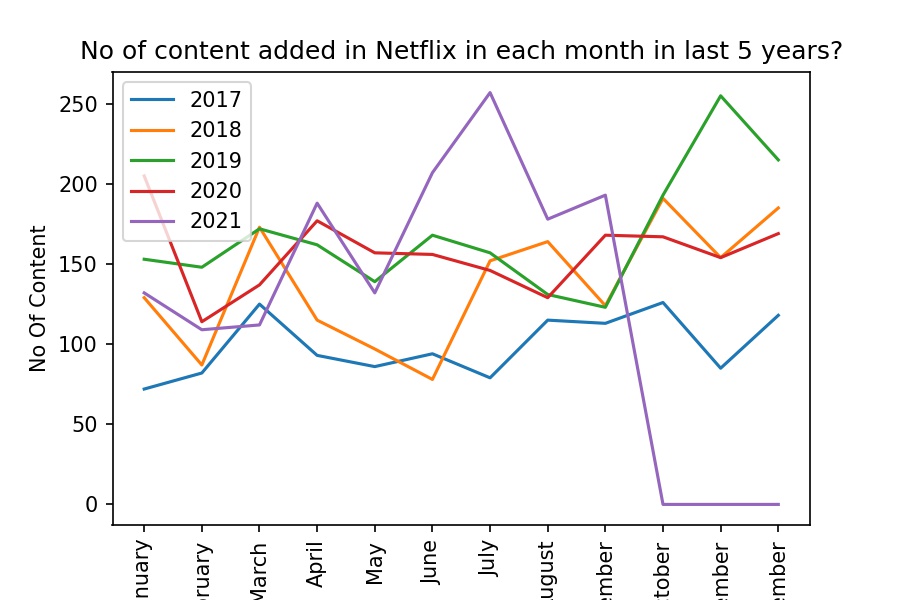
The given dataset has 8807 rows and 12 columns, out of which we can drop 2 columns – “description” and “title” which are nominal columns and would not provide much scope for deriving insights from the data, which would leave us with 10 columns. The grain of the data is the column “show\_id” which is a unique identifier for every row. All the columns are of object type except release year, which is of the type integer. From the column “date\_added”, we can make 2 new columns – “Year Added” of type integer and “Month Added” of type object and finally drop the original column “date\_added”. We would finally be left with **11 columns** in total.

Using the info method, we observed that the columns – “director”, “cast”, “country”, “date\_added”, “rating” and “duration” had **Nan** values

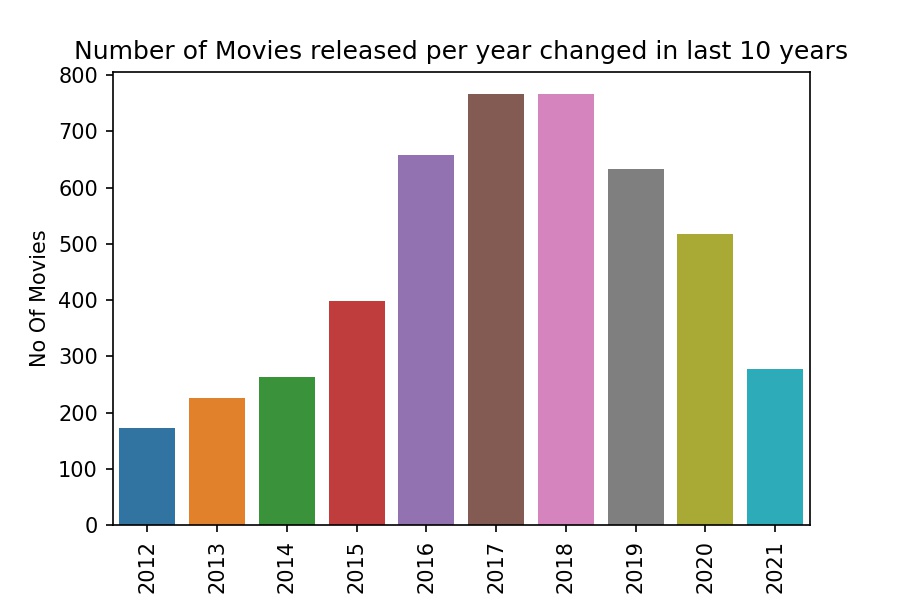
The columns – “director”, “cast”, “country”, “date\_added”, “rating” and “duration” had **Nan** values. All of the Nan values were filled with a certain value depending on the count of missing values and nature of the column, and none of the rows were dropped.

At the end of the above processes, all the 11 columns had 8807 non-null values.

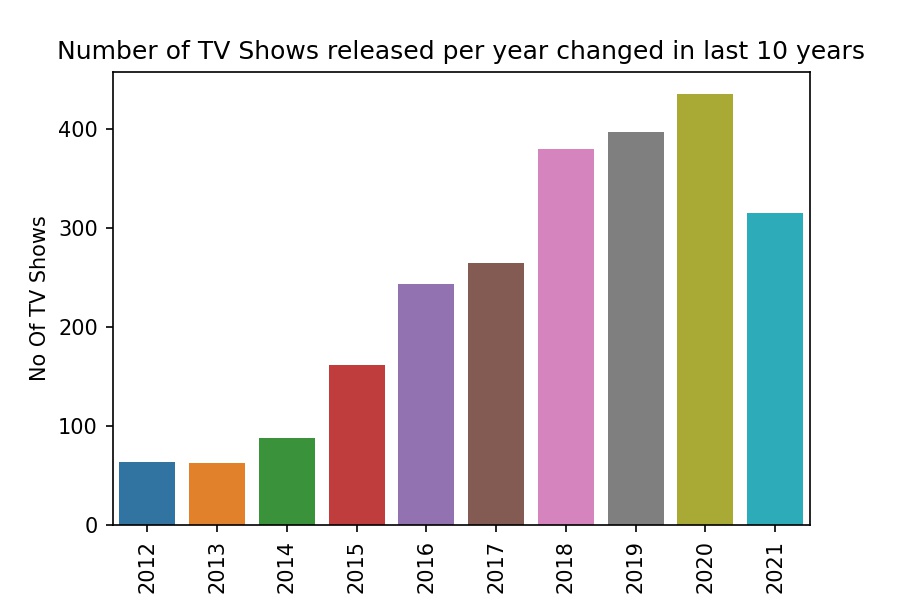
1. Business Insights (10 Points) - Should include patterns observed in the data along with what you can infer from it



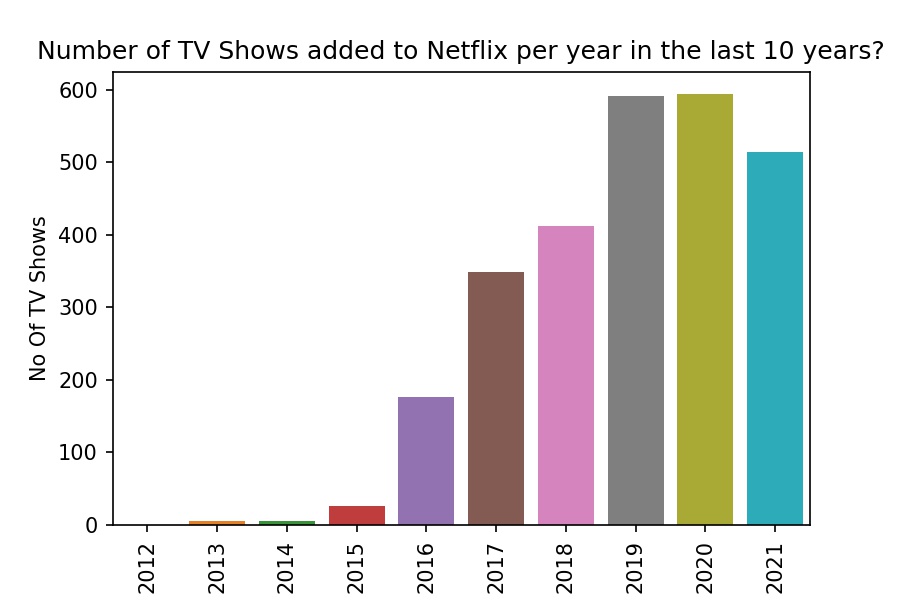
From the above graph, we can infer that there is no data available for the months – October, November, December for the year 2021(purple line). A lot of movies were released in July 2021 and November 2019.



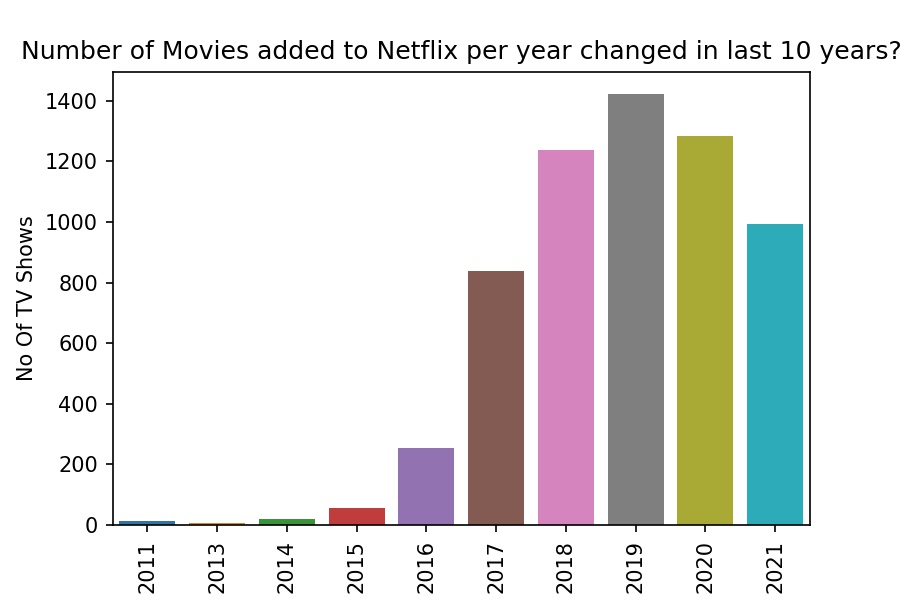
From the above graph, we can infer that there has been a reduction in the number of movies released year on year after 2018. We cannot infer for year 2021, as we do not have data for the last 3 months for 2021.

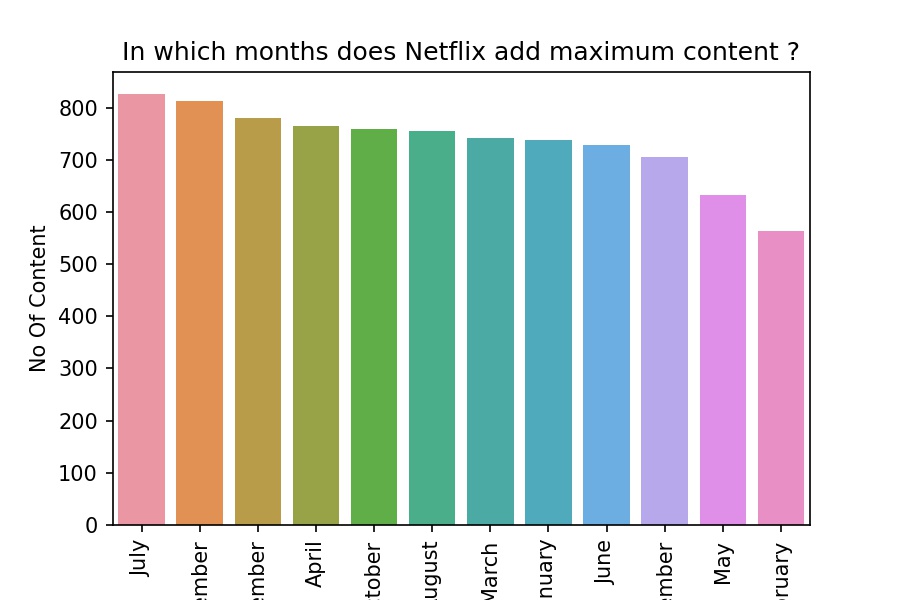


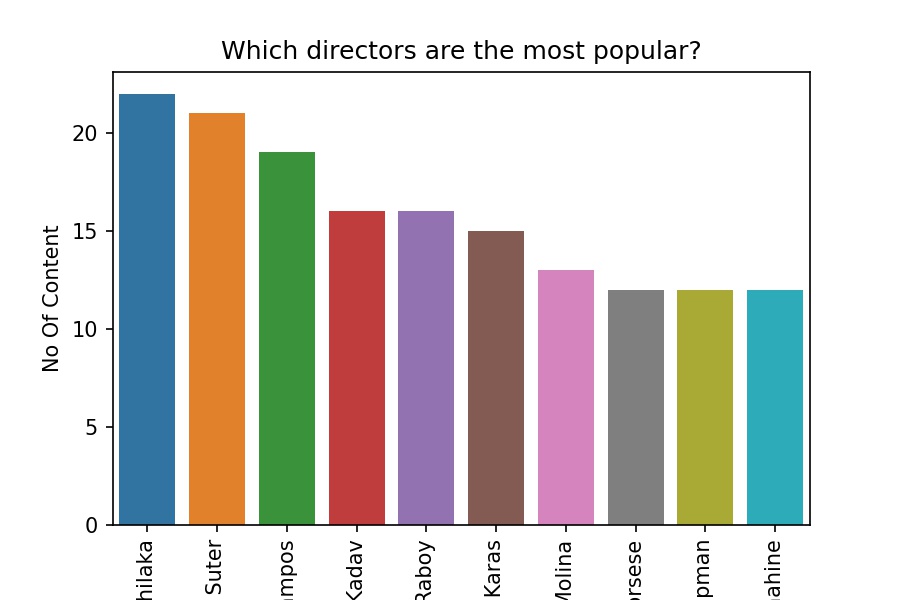
From the above graph, we can infer that there has been a steady increase in the number of Tv shows released year on year. We cannot infer for year 2021, as we do not have data for the last 3 months for 2021.



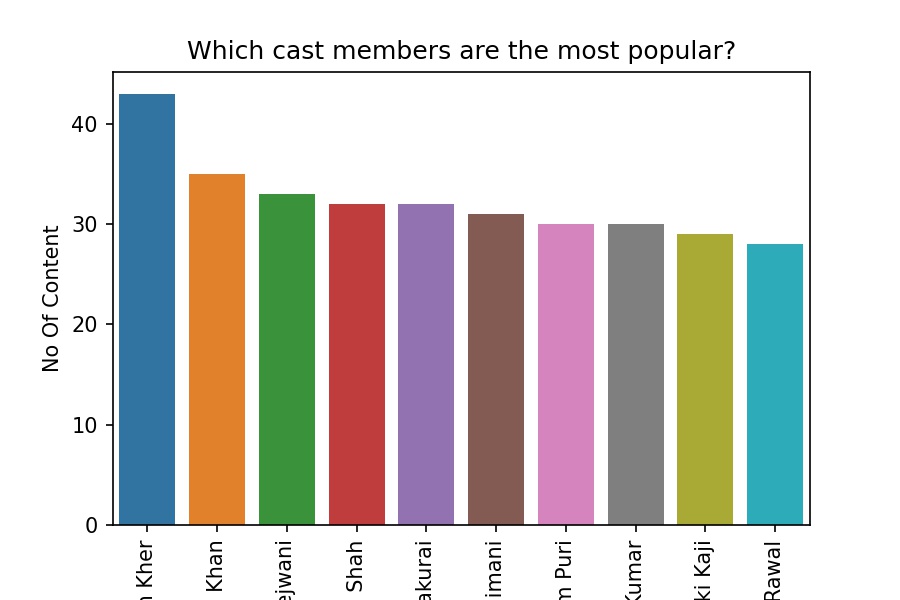
From the above graph, we can infer that there has been a steady increase in the number of TV shows added on Netflix year on year. One reason could be that there has been an increase in the no of TV show releases year on year. We cannot infer for year 2021, as we do not have data for the last 3 months for 2021.



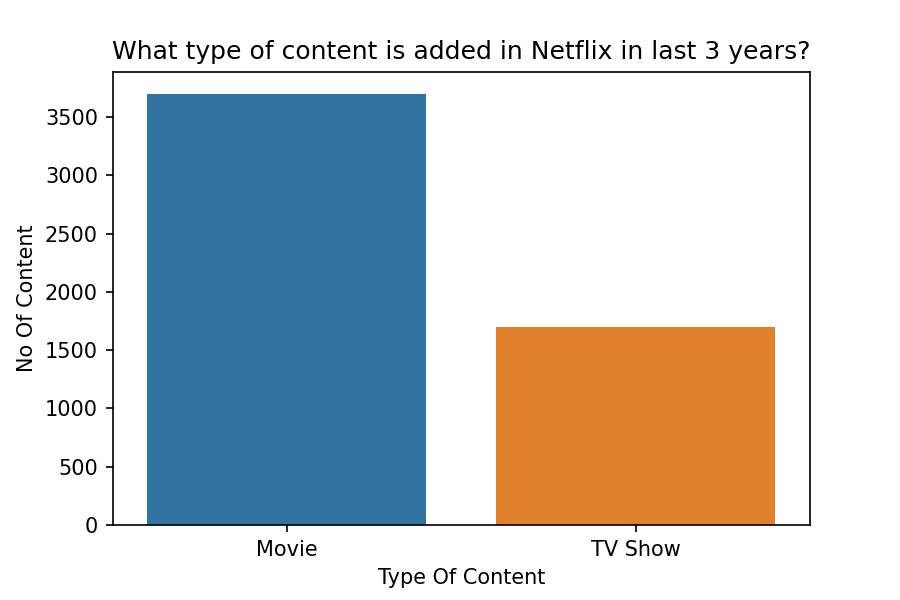
From the above graph, we can infer that there has been a decrease in the number of movies added on Netflix after 2019. Few reasons could be that there has been an increase in the no of TV show releases and TV show additions on Netflix year on year and also decrease in the no of movies released in last few years. We cannot infer for year 2021, as we do not have data for the last 3 months for 2021.

From the above graph, we can infer that most content(movie and TV show) are added on Netflix in the months July and December, and least number of content are added in the months May and February.

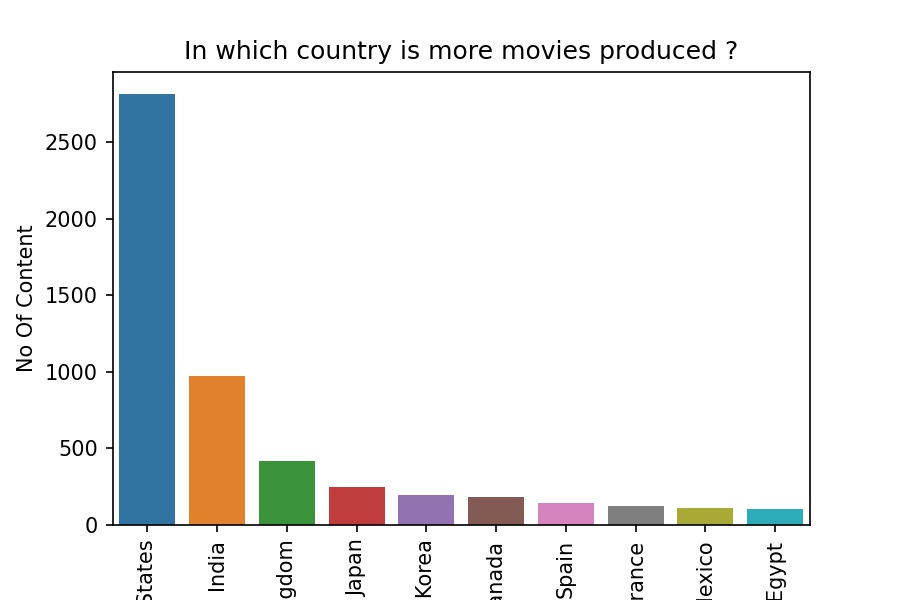
From the above graph, we can infer that the most popular directors are – Rajiv Chilaka, Jan Suter and Raul Campos.



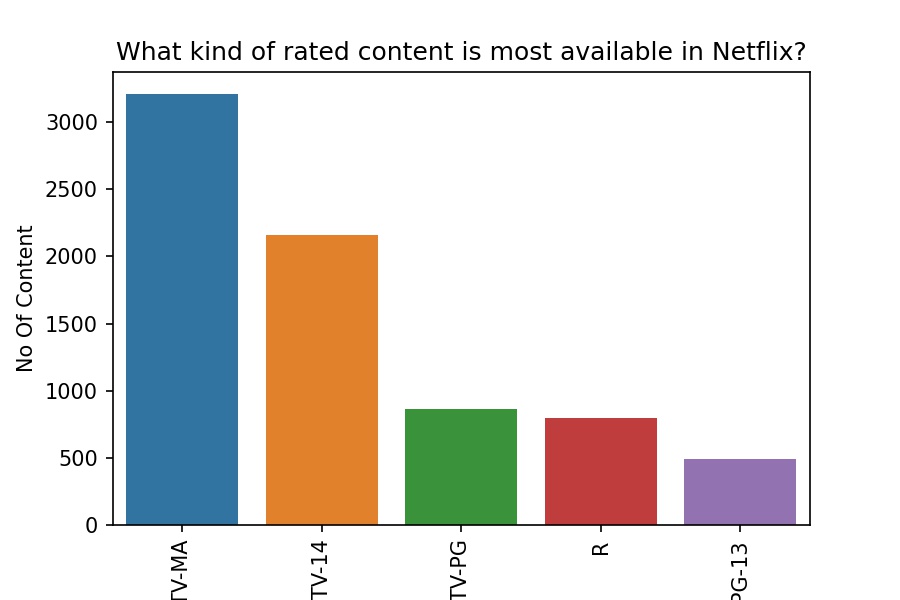
From the above graph, we can infer that the most popular casts are – Anupam Kher and Shahrukh Khan. Also, we get to infer that casts such as Julie Tejwani, Rupa Bhimani and Takahiro Sakurai are the most popular voice over artists for animated movies.



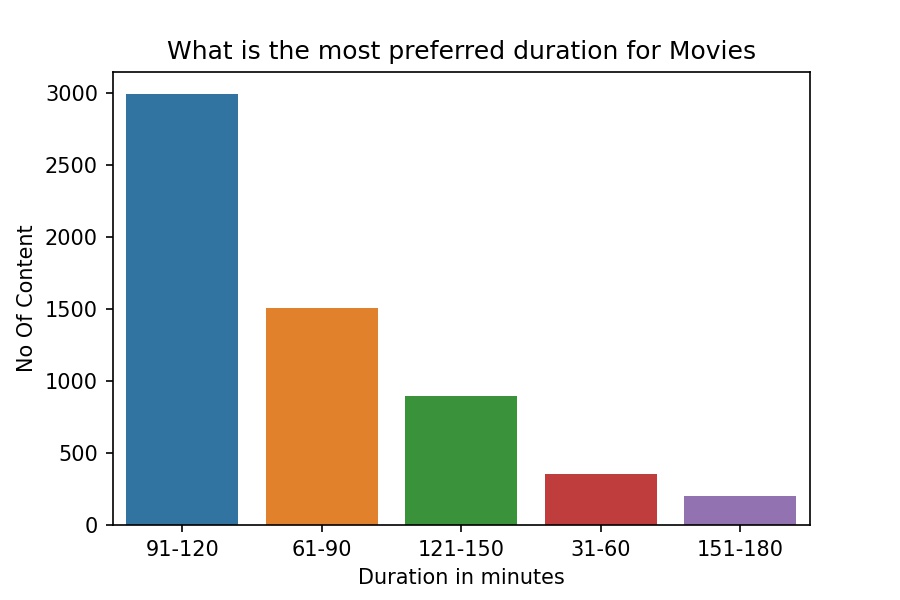
From the above graph, we can infer that in the last 3 year, the number of movies added in comparison to TV shows is 2 times.



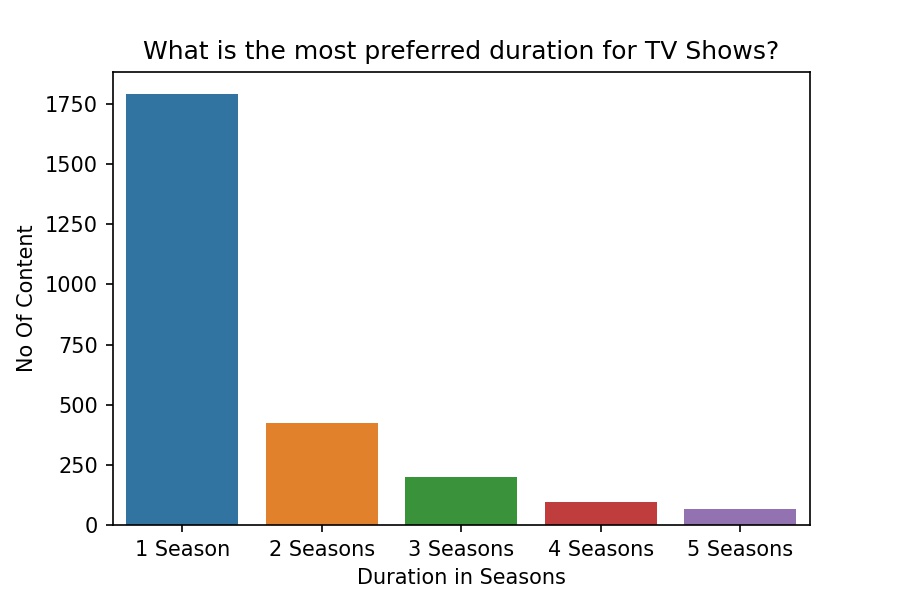
From the above graph, we can clearly infer that maximum number of movies are produced in the United States. India is ranked 2nd in the list.



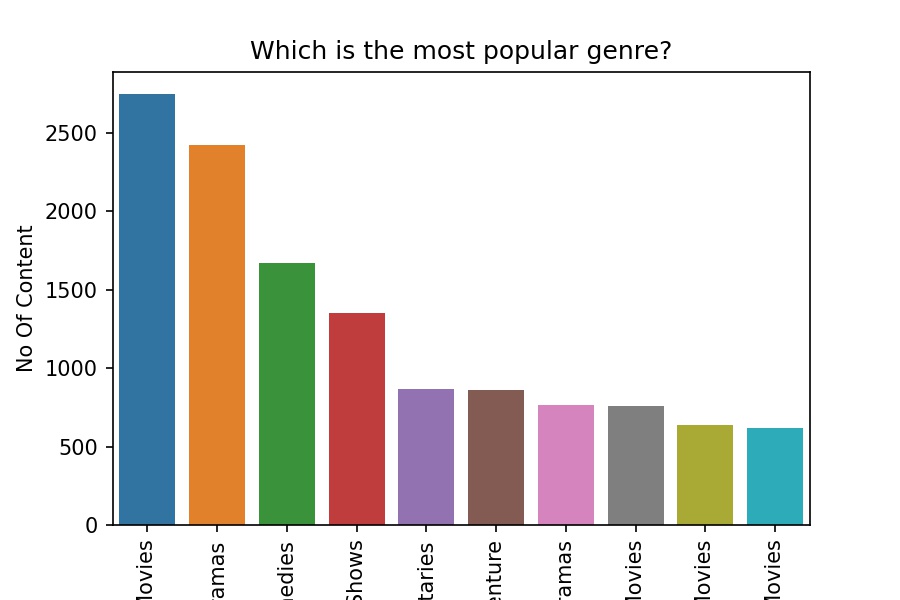
From the above graph, we can infer that most content produced are of the ratings – TV-MA(For Mature Audiences. May not be suitable for ages 17 and under) and TV-14(Parents strongly cautioned. May not be suitable for ages 14 and under).



From the above graph, we can infer that the most preferred duration of movies is in the range of 91 to 120 minutes.



From the above graph, we can infer that the most preferred duration of TV shows is 1 season.



From the above graph, we can infer that International Movies, Dramas and Comedies are the most popular genre.

1. Recommendations
2. Since from the last few years, there has been an increase in the no of TV shows being released and added on Netflix year on year, and decrease in the no of movies being released and added on Netflix year on year, it shows that there is an increasing demand for TV shows, and therefore Netflix should try to add more TV shows going forward.
3. Even though there has been an increasing trend for TV shows, but in the last 3 years, the no of movies added compared to TV shows is twice, therefore Netflix should try to add more TV shows on its platform.
4. The months May and February are the months in which content added on Netflix is the least. Therefore Netflix can try to add more content in these 2 months to ensure that people have enough content options during May and February.
5. Netflix should try to add more content for the most popular directors such as Rajiv Chilaka, Jan Suter and Raul Campos.
6. Netflix should try to add more content for the most popular casts such as Anupam Kher and Shahrukh Khan. Also, the presence of casts such as Julie Tejwani, Rupa Bhimani and Takahiro Sakurai in the top 10 tells us that animated movies are very popular and Netflix should try to add more animated movies as well.
7. Netflix should try to add more content having ratings – TV-MA and TV-14.
8. Netflix should try to add more content which are produced in the United States, as clearly it is the most popular country where content is produced.
9. Netflix should try to add more movies which have a duration of in the range of 91 to 120 minutes.
10. Netflix should try to add more TV shows which have a 1 season.
11. Netflix should try to add more content for the genres - International Movies, Dramas and Comedies.