EXPERIMENT 3

NAME	Shreya Shetty
UID	2019140059
CLASS	TE IT
BATCH	A
SUBJECT	Big Data Analytics Lab

AIM: Explore and present interactive data insights from real world dataset (Dashboards) using Tableau
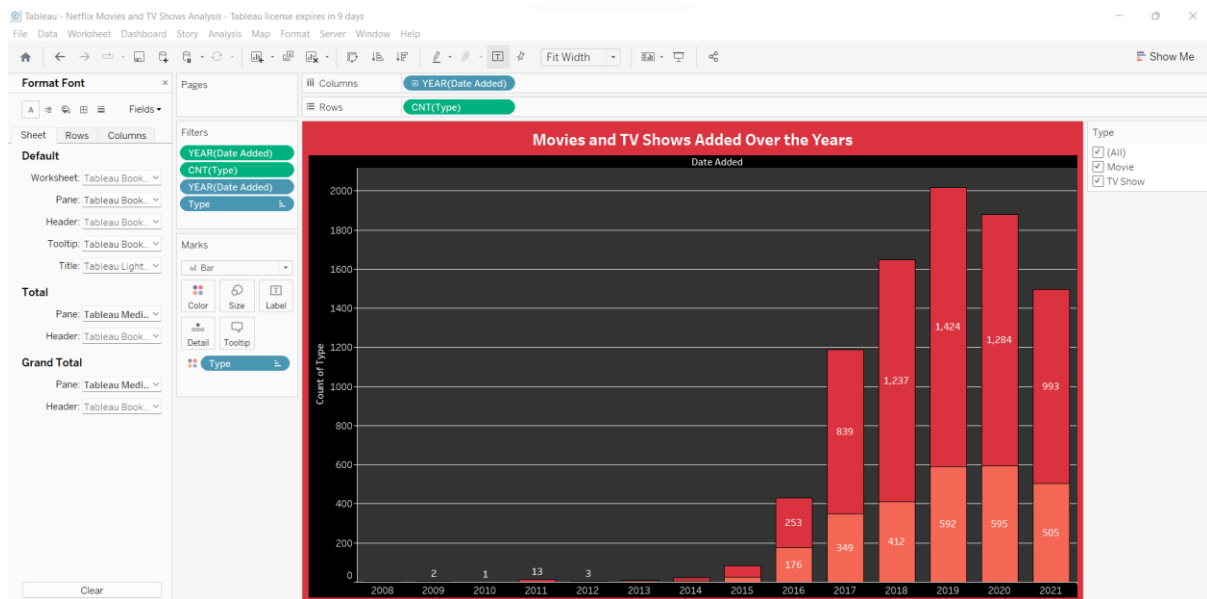
DATASET: Netflix Movies and TV Shows

KAGGLE LINK: https://www.kaggle.com/shivamb/netflix-shows/

There are 8807 unique rows in the dataset chosen. Following are the 12 attributes in this dataset-

| Sr. No.	Attribute Name	Attribute Description	Datatype
1	Show Id	Unique ID for every Movie / Tv Show	String
2	Type	Identifier - A Movie or TV Show	String
3	Title	Title of the Movie / Tv Show	String
4	Director	Director of the Movie	String
5	Cast	Actors involved in the movie / show	String
6	Country	Country where the movie / show was produced	String
7	Date Added	Date it was added on Netflix	Date
8	Release Year	Actual Release year of the move / show	Integer
9	Rating	TV Rating of the movie / show	String
10	Duration	Total Duration - in minutes or number of seasons	String
11	Listed In	Genre of Movie / TV Show	String
12	Description	The summary description	String

Importing the Netflix Movies and TV Show Dataset Downloaded from Kaggle:

PLOT 1: Movies and TV Show Count on World Map

In the above plot, a map is plotted where total count of distinct number of Titles of both TV Shows and Movies in each country is shown. Here, Count of Titles is marked with respect to country attribute. It can be observed that The United States, India, and the United Kingdom have a higher percentage of content. This shows that the platform has a good presence in these countries.

## PLOT 2: Movies and TV Shows added over the years

In the above chart, a bar graph is plotted between count of Type attribute (Movie & TV Show) and Date Added. It helps to visualize the number of TV Shows and Movies added to Netflix over the years. Maximum TV Shows and Movies were added in the year 2019. Also, there is an increase in the amount of content added on Netflix from 2016 onwards. This is because of the launch of the Netflix app in 2016 and expansion in Asia.

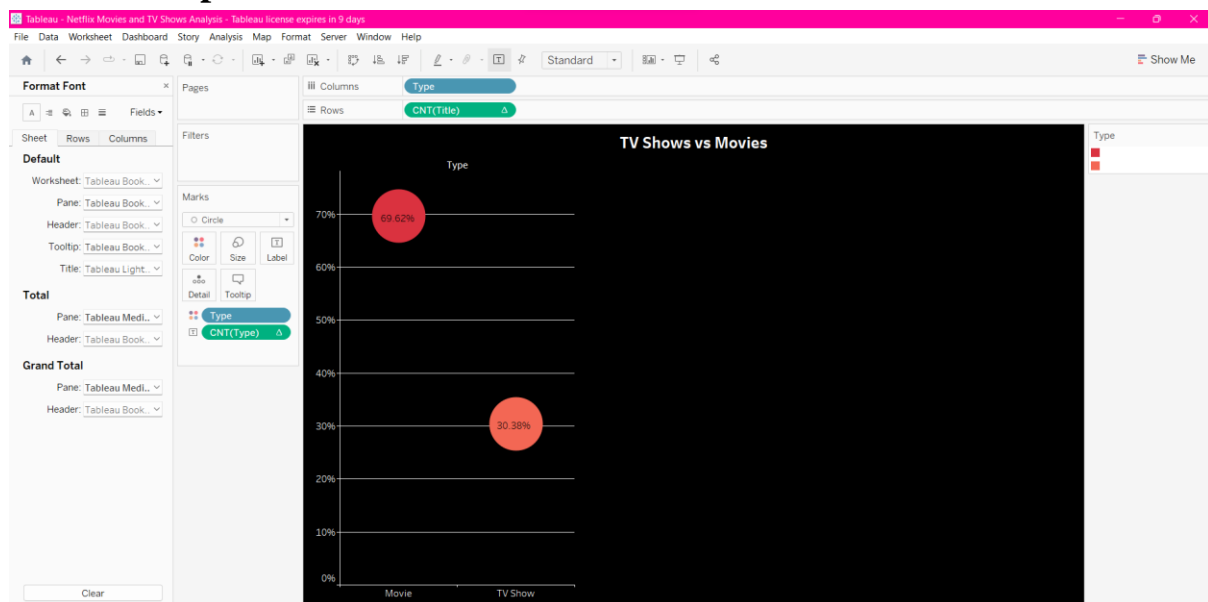## PLOT 3: Top 5 directors (By number of Titles)
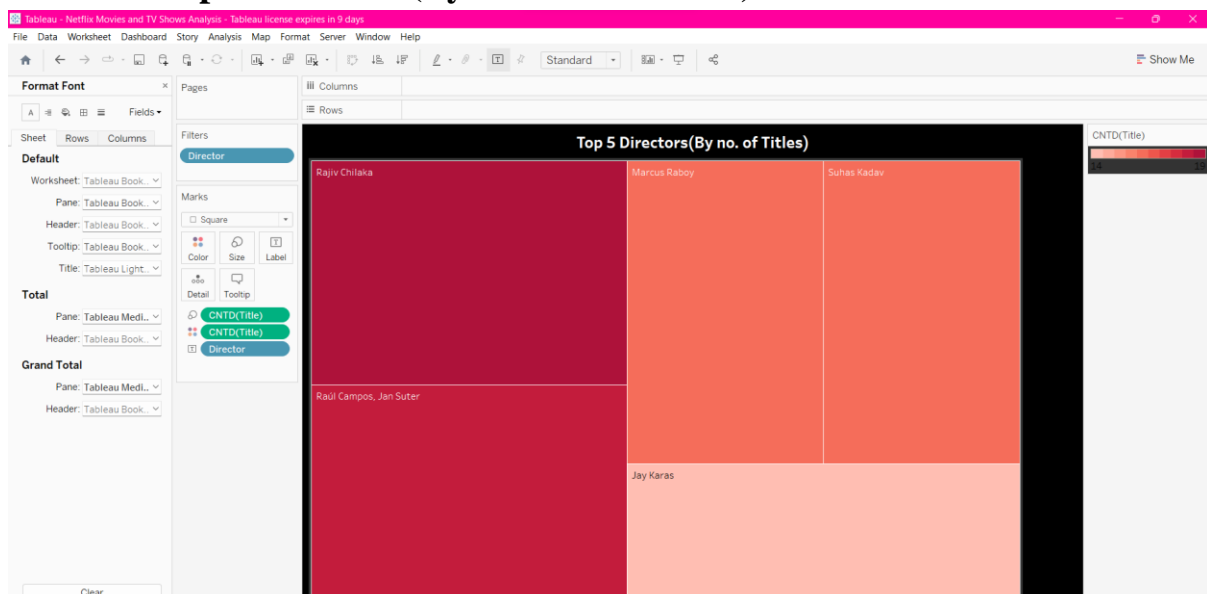
The above chart was plotted between the directors and number of titles. It helps us to identify the director having maximum titles. Here, the director with maximum Titles is Rajiv Chilaka.
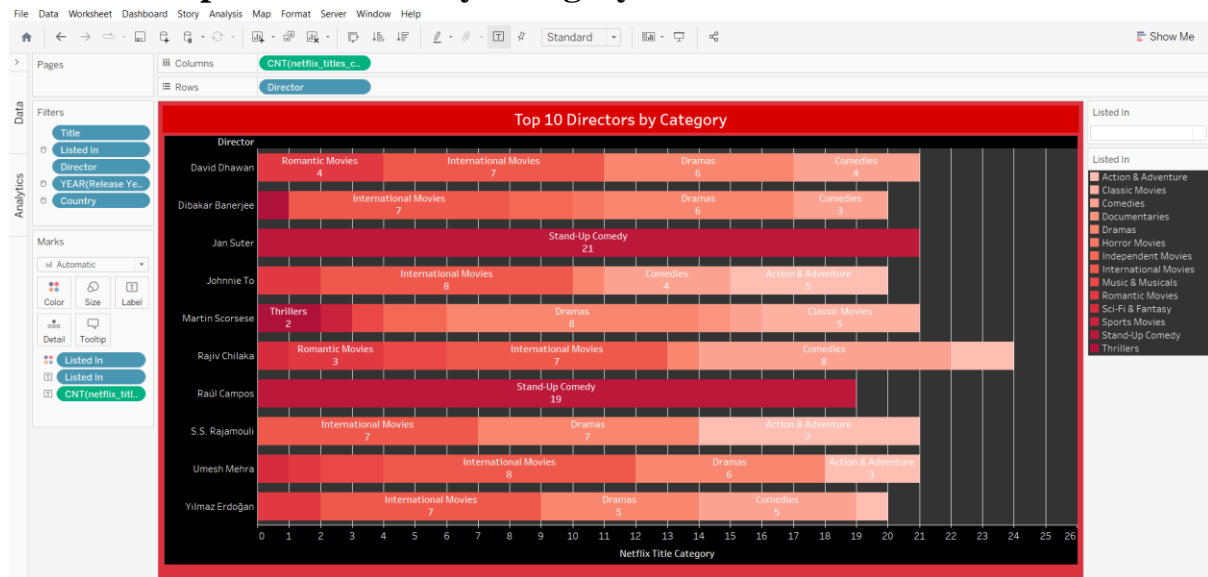
PLOT 4: Top Genres (By number of Titles)
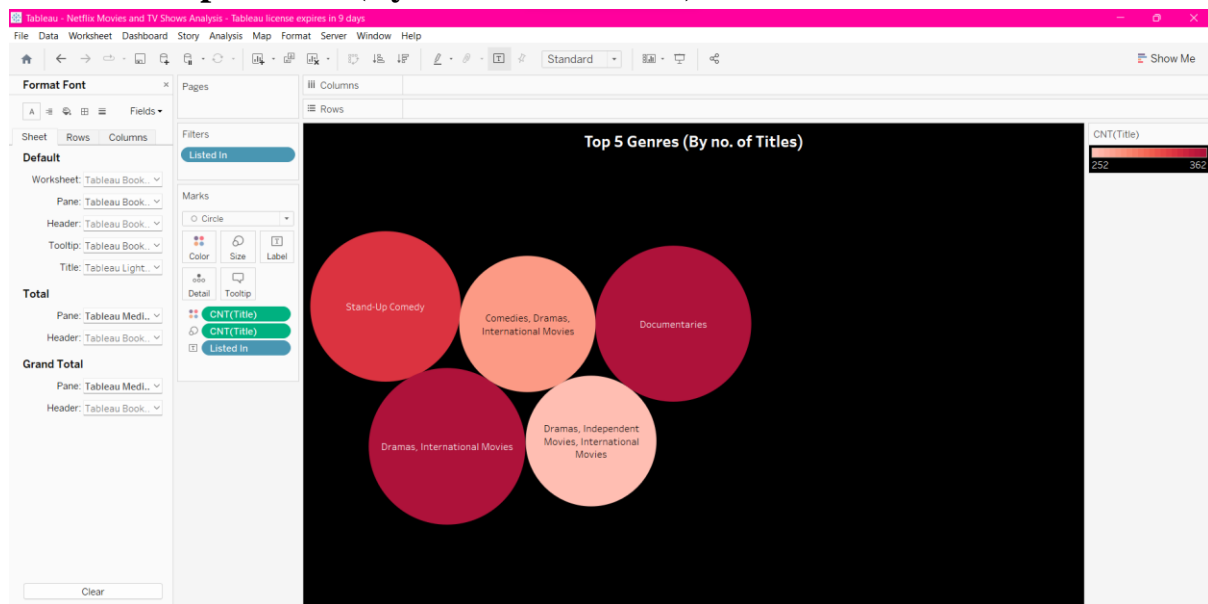
The above chart was plotted between the Genres i.e. Listed In Attribute and number of titles. It helps us to identify the genre having maximum titles. In this dataset, the Genre with a maximum of 334 Titles is Stand-Up Comedy.

PLOT 5: Top TV Shows V/S Movies

The above chart was plotted between the Types Attribute and percentage of types. It helps us to identify which type i.e. TV Shows or Movies has maximum titles. In this dataset, the percentage of Movies (69.62%) is way more than the percentage of TV Shows (30.38%).

## PLOT 6: Top 10 Directors By Category

The above chart was plotted between sub-categories of Genres and Directors. It helps us to identify the top 10 directors along with the number of titles in different genres (Listed In Attribute) with respect to each of the top directors.

PLOT 7: Top TV Show Categories
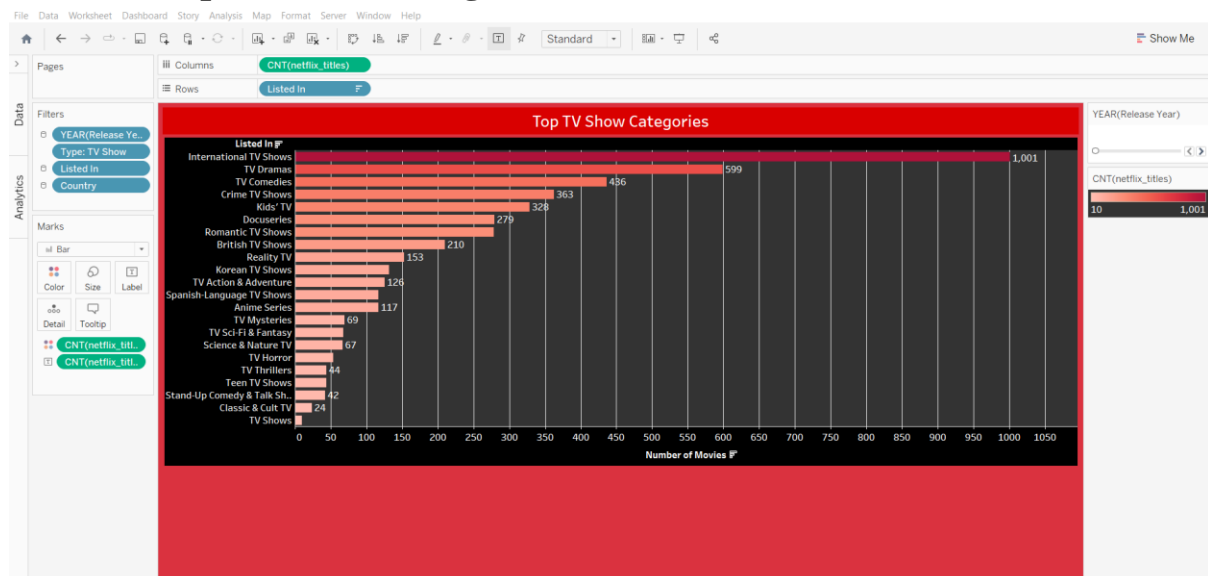
The above bar graph was plotted between Genres (Listed In) and Count of TV Shows (From Type: TV Shows). It helps us to identify the genres with highest number of TV Shows.

PLOT 8: Top Movie Categories
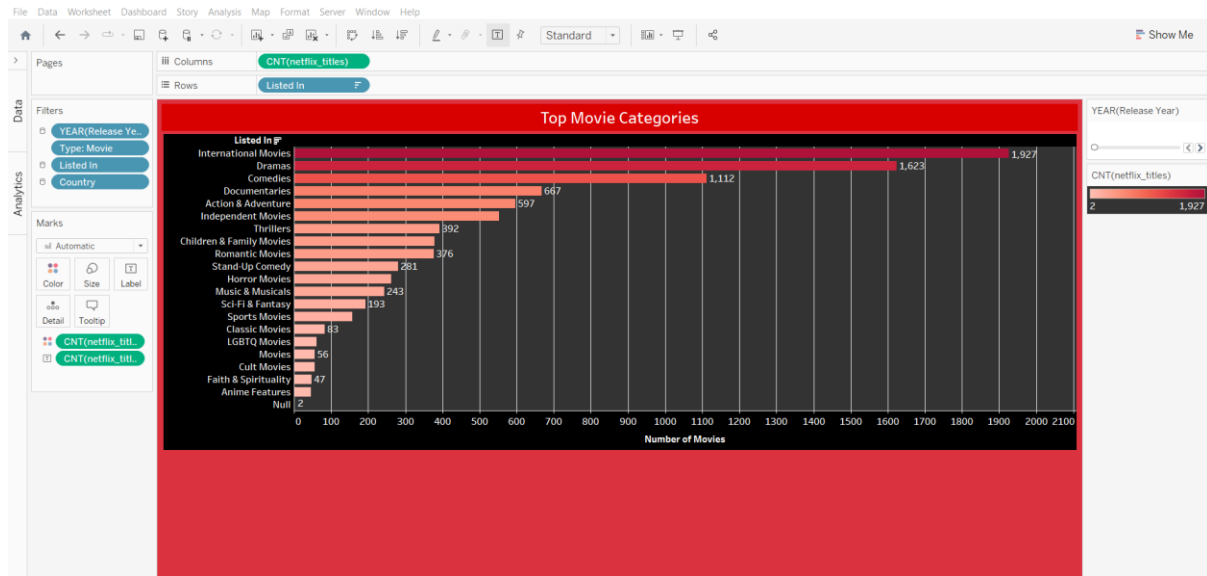
The above bar graph was plotted between Genres (Listed In) and Count of Movies (From Type: Movies). It helps us to identify the genres with highest number of Movies.

CONCLUSION:

In this experiment I learnt to use Tableau for Exploratory Data Analysis and plotted 8 different plots to draw insights from the Netflix Movies and TV Shows Dataset.