

Data Science Dashboard

Tanya Garg

102016109

3CS12

Dataset: Netflix

Link: <https://www.kaggle.com/datasets/shivamb/netflix-shows>

Queries in R:

```
ccdf<-read.csv('dashboard/netflix_cast_and_crew.csv')
```

```
dordf<-read.csv('dashboard/netflix_date_of_release.csv')
```

```
desc1df<-read.csv('dashboard/netflix_description1.csv')
```

```
desc2df<-read.csv('dashboard/netflix_description2.csv')
```

```
tcdf<-read.csv('dashboard/netflix_type_countries.csv')
```

#row-wise combination

```
descdf<-rbind(desc1df,desc2df)
```

#left outer join

```
dfcctcdf<-merge(x=ccdf,y=tcdf,all.x=TRUE)
```

#inner join

```
nfwddf<-merge(dfcctcdf,dordf)
```

```
nfdf<-merge(nfwddf,descdf)
```

```
write.csv(nfdf,"Netflix_analytics_dataset.csv", row.names=FALSE)
```

```
data<-read.csv("Netflix_analytics_dataset.csv")
```

```
library(xlsx)
```

```
library(openxlsx)
```

```
library(dplyr)
```

```
library(tidyr)
```

```
library(modeest)
```

```
colSums(is.na(data))
```

```
df<-na.omit(data)
```

```
df1<- df %>%
```

```
  separate(date_added,c("Month"))
```

```
df1<-df1[!is.na(df1$Month),]
```

```
df1$Month[df1$Month==""]<-mfv(df1$Month)
```

```
df1$Month<-factor(df1$Month,levels = month.name)
```

```
df2<-separate_rows(df1,country,show_id , convert = TRUE, sep = ' , ')
```

```
country_count<-sort(table(df2$country),decreasing=TRUE)[1:10]
```

```
country_count<-data.frame(country_count)
```

```
write.xlsx(df,"netflix/df.xlsx")
```

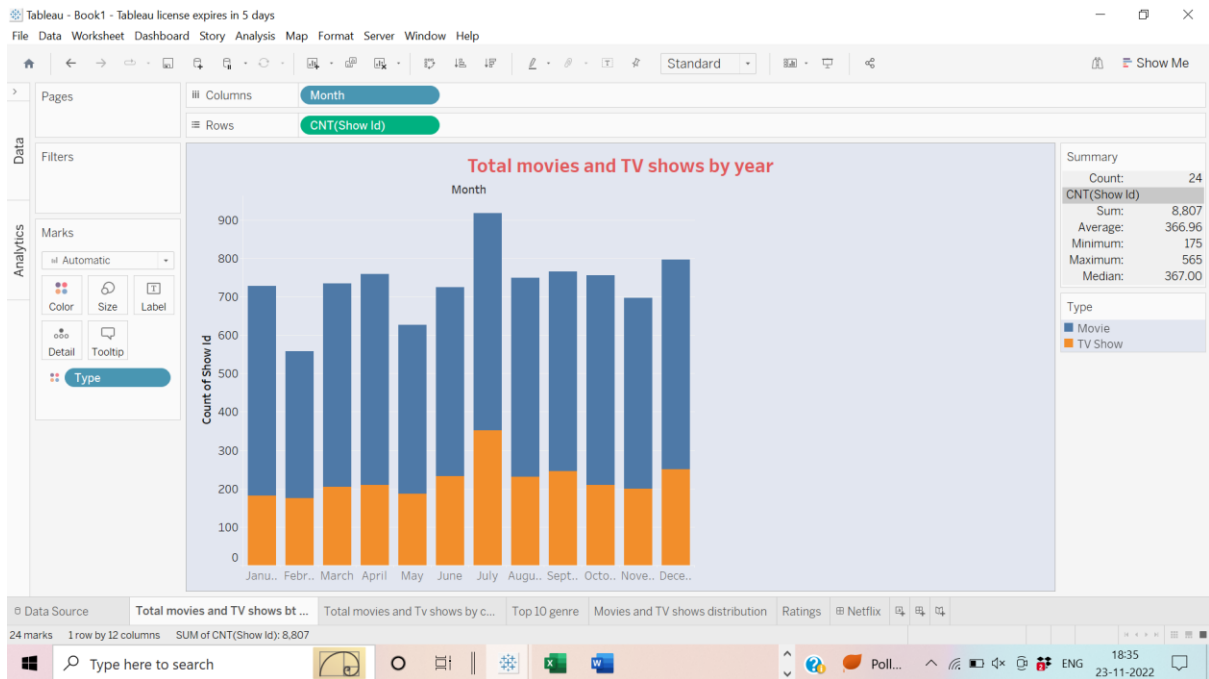
```
write.xlsx(df1,"netflix/df1.xlsx")
```

```
write.xlsx(df2,"netflix/df2.xlsx")
```

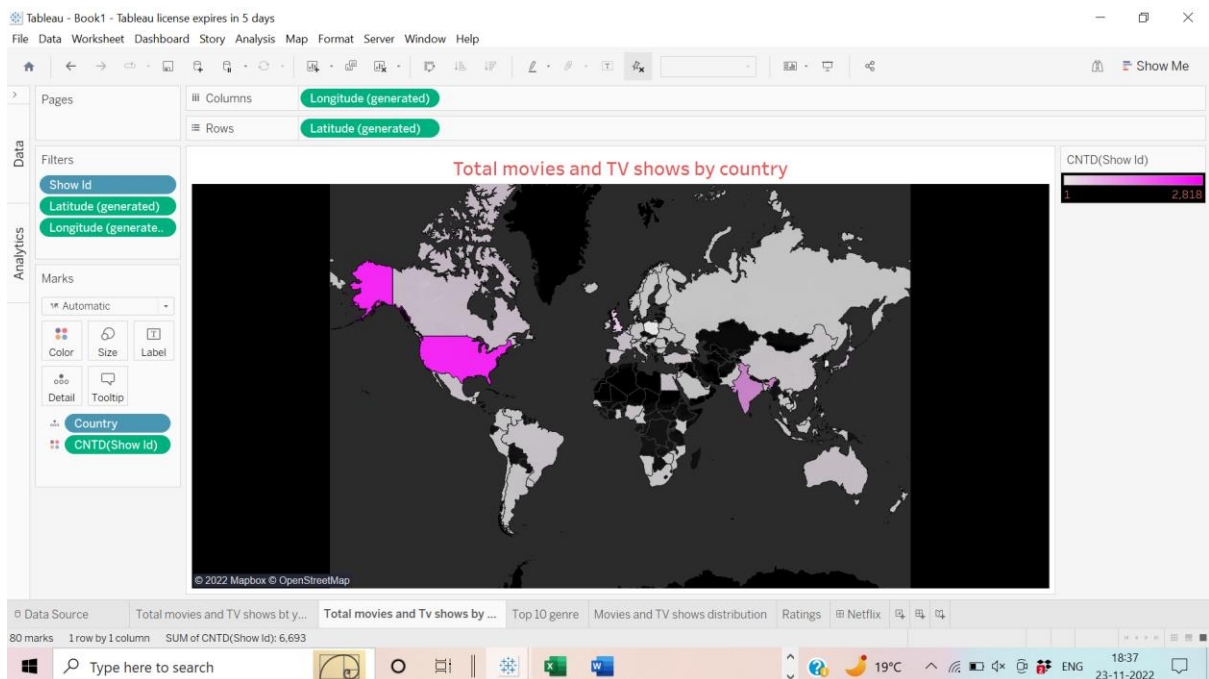
```
write.xlsx(country_count,"netflix/country_count.xlsx")
```

Sheets in Dashboard:

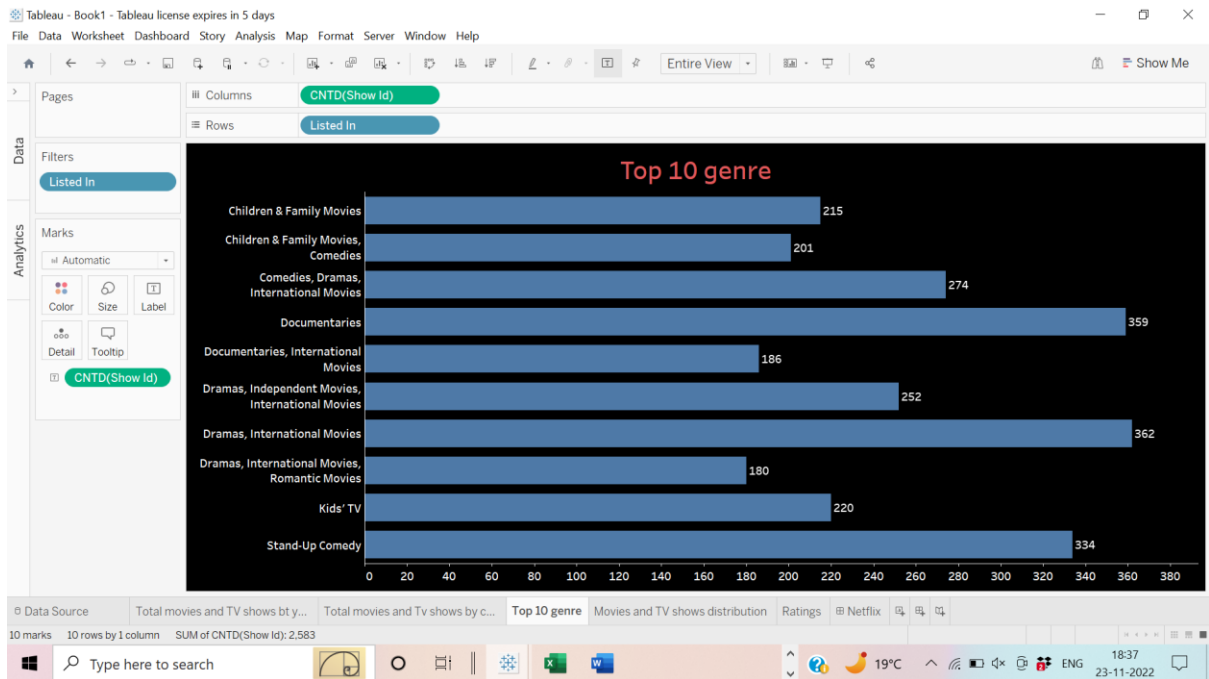
Total Movies and TV shows by year:-



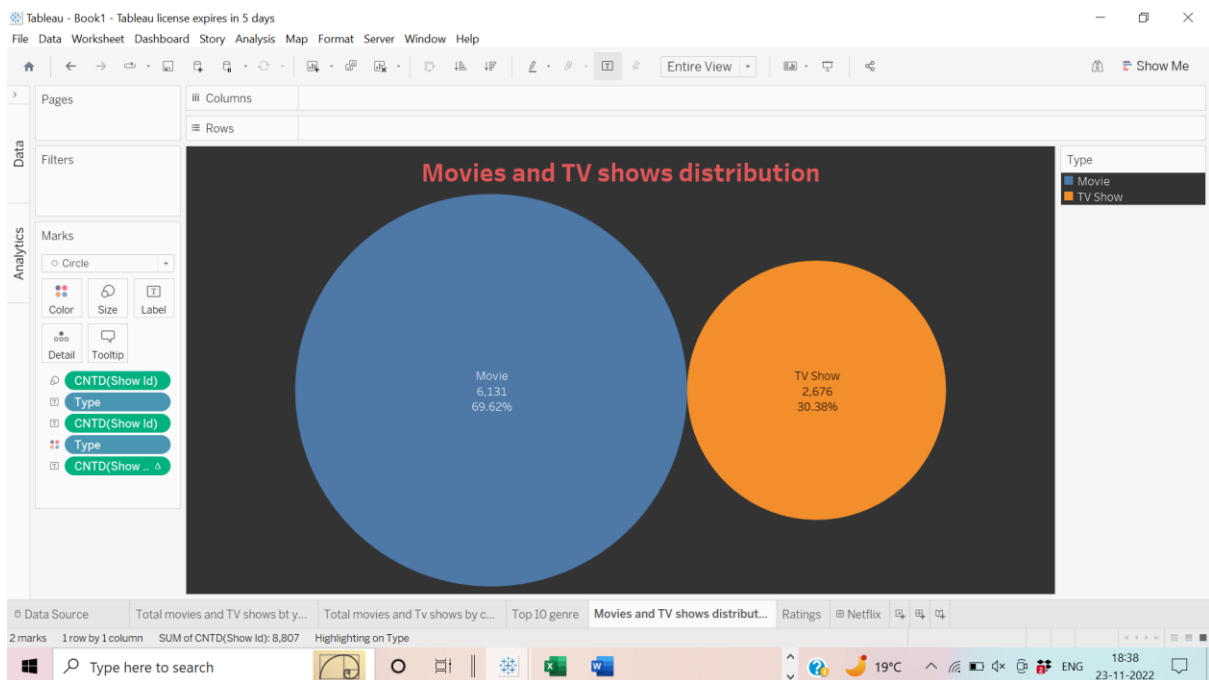
Total Movies and TV shows by country:-



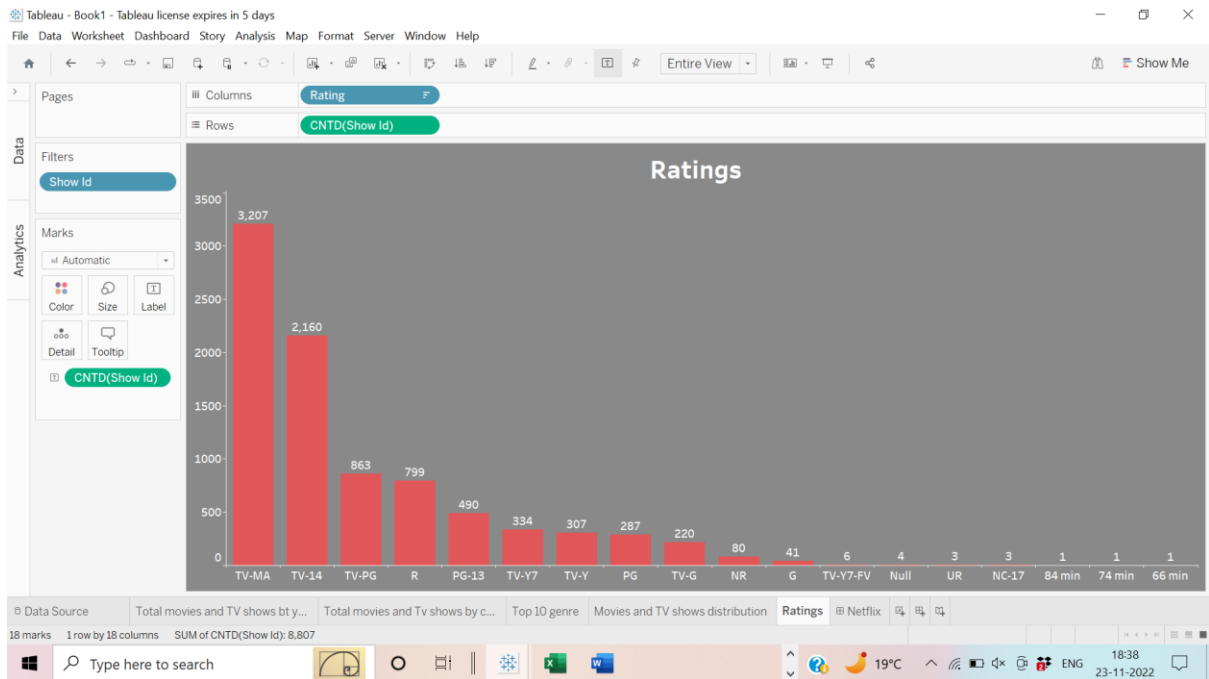
Top 10 genre:-



Movies and TV shows distribution:-



Ratings:-



Dashboard:-

