Best Genre In Netflix

Daniel L Robelto B

2023-02-16

```
library(tidyverse)
best_movies_netflix <- read_csv("best_movies_netflix.csv")
view(best_movies_netflix)
```

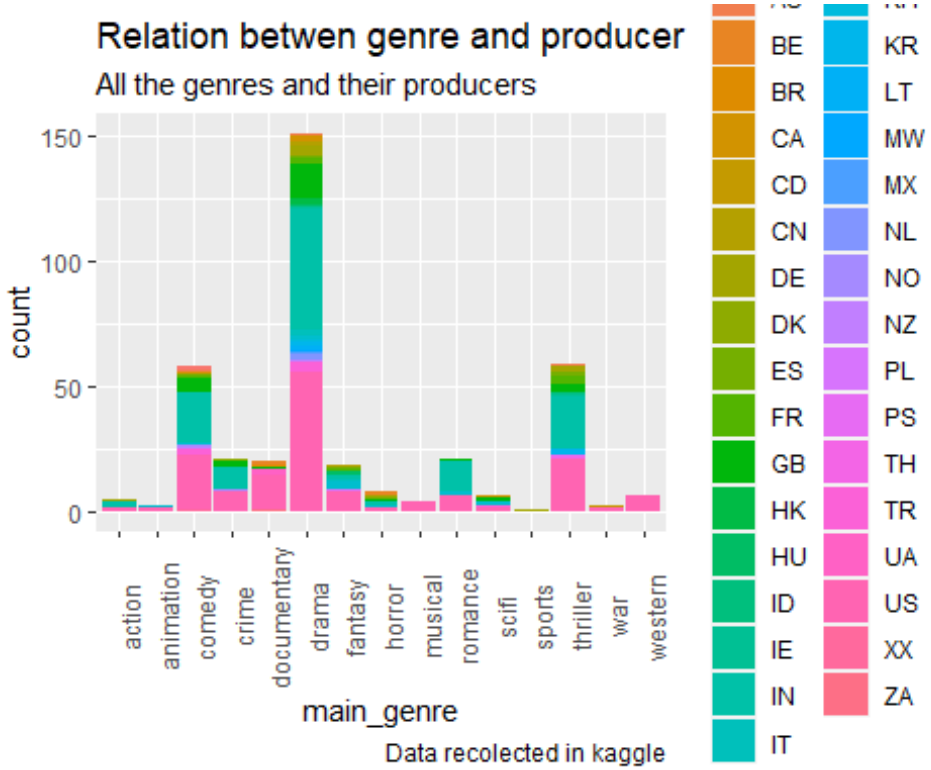
What is the top genre to use if you wanna make a movie?

Based on Netflix kaggle dataset there are in total 15 kind of genres in the top of all the platform.

```
best_movies_netflix %>%
  group_by(main_genre) %>%
  count(main_genre)%>%
  arrange(-n)
```

```
## # A tibble: 15 × 2
## # Groups:   main_genre [15]
##    main_genre      n
##    <chr>       <int>
##  1 drama         151
##  2 thriller       59
##  3 comedy         58
##  4 crime          21
##  5 romance        21
##  6 documentary    20
##  7 fantasy        19
##  8 horror          8
##  9 scifi           7
## 10 western         7
## 11 action          5
## 12 musical         4
## 13 animation       3
## 14 war             3
## 15 sports          1
```

```
ggplot(data = best_movies_netflix)+geom_bar(mapping = aes(x=main_genre, fill =
main_production))+labs(title = "Relation betwen genre and producer", subtitle = "All the
genres and their producers", caption = "Data recolected in kaggle")+theme(axis.text.x =
element_text(angle = 90))
```



Drama genre is the top one for lot, but this genre would be a good option to use if you are researching about the best genre to use if you are making a movie or serie.

One of the best options to be sure that one serie or movie was watched is the amount of votes that had where we can see the next:

```
best_movies_netflix %>%
  group_by(main_genre) %>%
  summarise(total_votes = sum(number_of_votes), mean_score = mean(score)) %>%
  arrange(-total_votes)
```

```
## # A tibble: 15 × 3
##    main_genre  total_votes mean_score
##    <chr>             <dbl>      <dbl>
##  1 drama          21804564       7.48
##  2 thriller        8398468       7.47
##  3 comedy          4703792       7.61
##  4 crime           3857710       7.47
##  5 scifi           3832671       7.77
##  6 western         2770109       7.56
##  7 fantasy         2639958       7.48
##  8 horror          1759154       7.35
##  9 romance         1428590       7.4
## 10 action           535189       7.7
## 11 documentary      504297       7.76
## 12 animation        341513       7.2
## 13 war              141682       7.8
## 14 musical           94236       7.15
## 15 sports            21558       7.4
```

with a big difference drama genre is the preferred genre the audience watch it's the 41% from all the votes with a mean score of 7.47 where the global mean is 7.51 from this data set.

based on the last, is time to select who would be the best producer now that we select drama genre like our focus.

```
best_movies_netflix %>%
  distinct(main_production, .keep_all = TRUE) %>%
  count(main_production) %>%
  summarise( total_productions = sum(n))
```

```
## # A tibble: 1 × 1
##   total_productions
##               <int>
## 1                35
```

There are 35 productions where the top of then are:

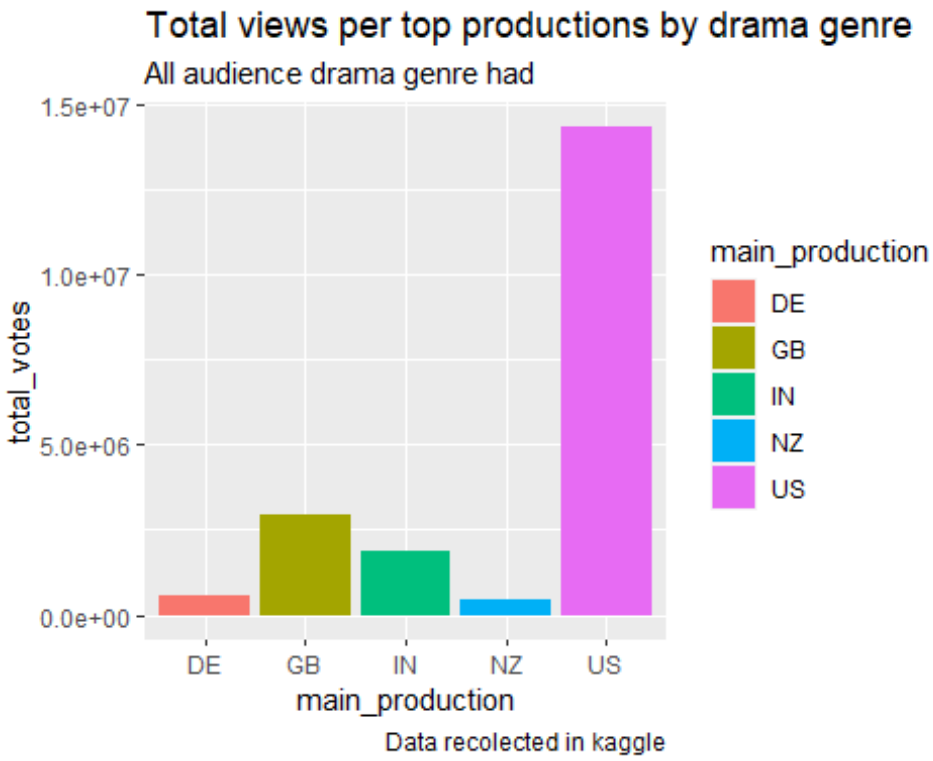
```
best_movies_netflix %>%
  filter(main_genre == "drama") %>%
  count(main_production) %>%
  arrange(-n) %>%
  head(5)
```

```
## # A tibble: 5 × 2
##   main_production     n
##   <chr>           <int>
## 1 US                 56
## 2 IN                 48
## 3 GB                 14
## 4 DE                  4
## 5 TR                  4
```

The are two big possibilities between US and IN, but more participation mean have more views?, that's why is necessary verify if our two possibilities have enough votes because sometimes have a big participation doesn't mean big audience.

```
drama_votes_df <- best_movies_netflix %>%
  filter(main_genre == "drama") %>%
  group_by(main_production) %>%
  summarise(mean_score = mean(score), total_votes = sum(number_of_votes)) %>%
  arrange(-total_votes) %>%
  head(5)

ggplot(data = drama_votes_df)+geom_col(mapping = aes(x=main_production, y=total_votes,
fill=main_production))+labs(title = "Total views per top productions by drama genre",
subtitle = "All audience drama genre had", caption = "Data recolected in kaggle")
```



As we can see, US producers had more participation and more total votes that's mean audience that watch series or movies which are produced by US are the preference for people that use Netflix. But if you are looking for ideas to make you movie here there is a brief summary about years with a mean_score more than 7.5, as you already know 7.5 is the total average from this data set.

```
best_movies_netflix %>%
  filter(main_genre=="drama") %>%
  group_by(release_year) %>%
  summarise(mean_score = mean(score)) %>%
  arrange(-mean_score) %>%
  head(8)
```

```
## # A tibble: 8 × 2
##   release_year mean_score
##          <dbl>      <dbl>
## 1         1984       8.3
## 2         1987       8.3
## 3         1994       8.05
## 4         1998       7.9
## 5         2003       7.82
## 6         1964       7.8
## 7         1990       7.8
## 8         2012       7.77
```

Now you know what would be the years you should use if you wanna extract ideas for your series or movie. Remember that all this analysis was made only based on Netflix and not global. And don't forget that you have more possibilities with US producer if you want a big audience