Course: BAX-431 Data Visualization

Professor: Mak Ahmad

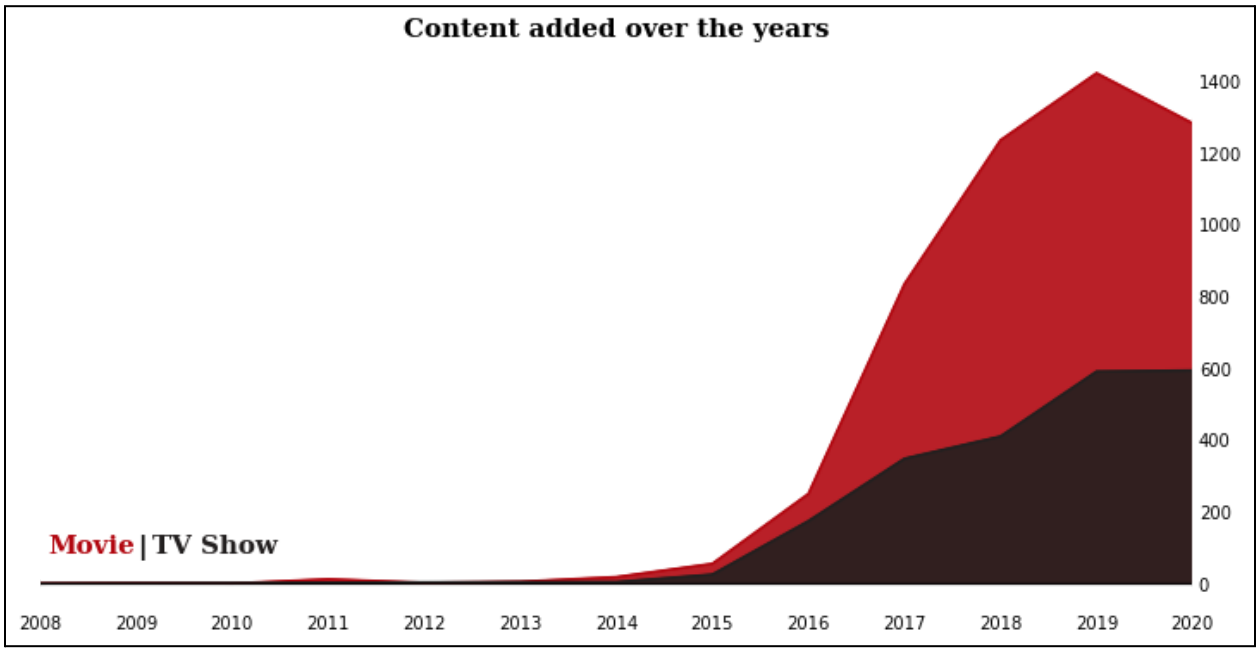
Assignment: Exploratory Visualization by Team 6

Introduction

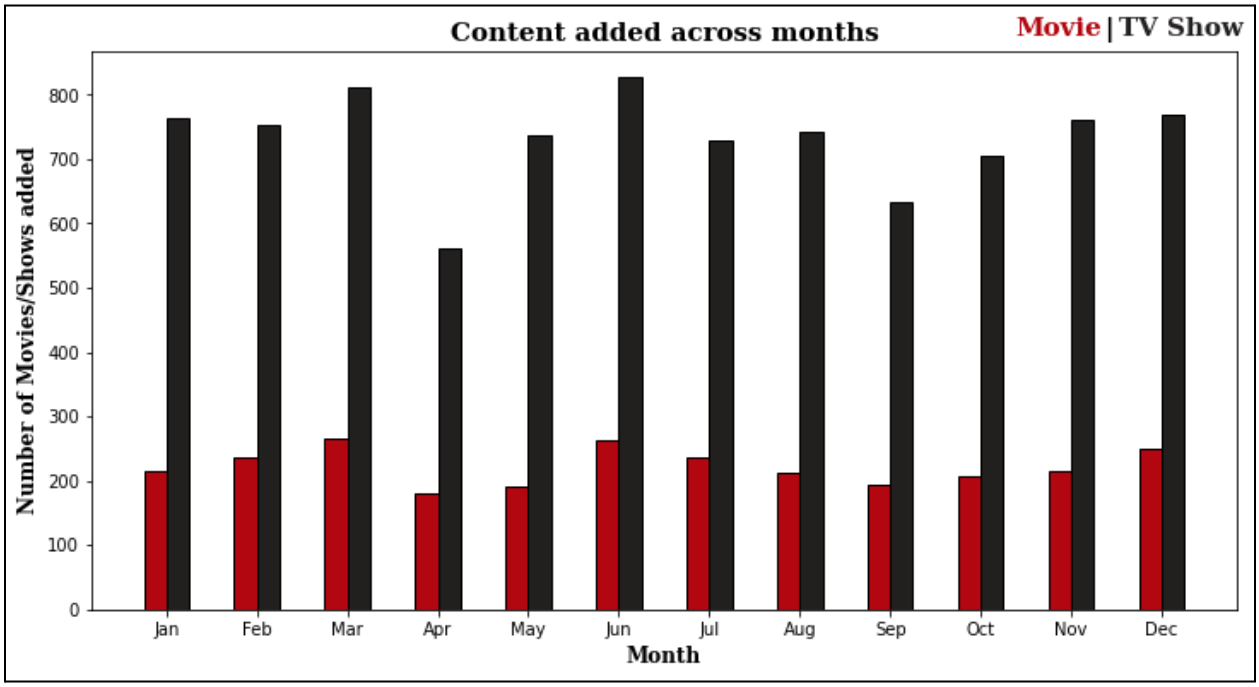
Through this exploratory exercise, we want to build visualizations that showcase how content (Movies/TV shows) hosted on Netflix has progressed over the years. This will tell our users the popularity of content based on their genres and geography. These insights can help content producers understand their target audience and preferences in what they consume. Since Netflix is not just a streaming service anymore but they have also been producing original content in recent years, our conclusions can thus prove to be instrumental in driving content that works well with the audience. This can also help Netflix fine-tune its recommendations based on trends in content consumption in different consumer segments.

Hypothesis 1: More content is added during the Holiday period in the Americas and Europe

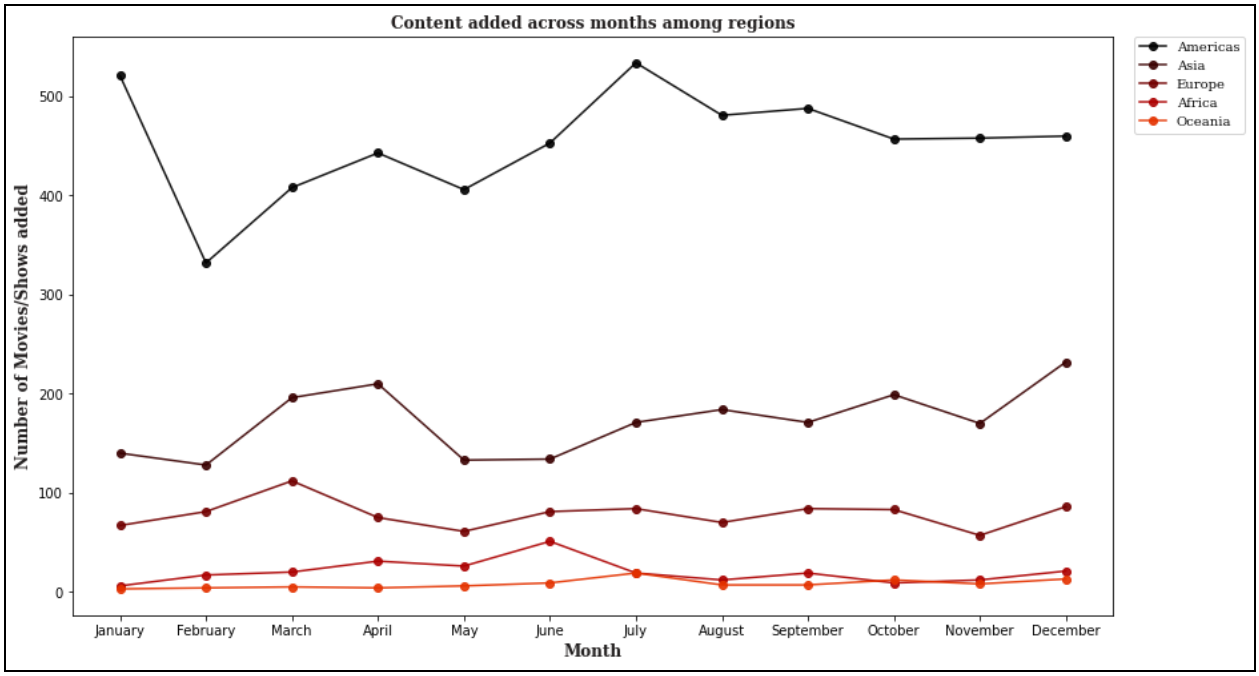
We see a slow start for Netflix over several years. Things began to pick up in 2015, with a rapid increase from 2016 for both movies and TV Shows. However, we see a much more rapid growth for movies than shows. Content additions slowed down in 2020, likely due to the COVID-19 pandemic.



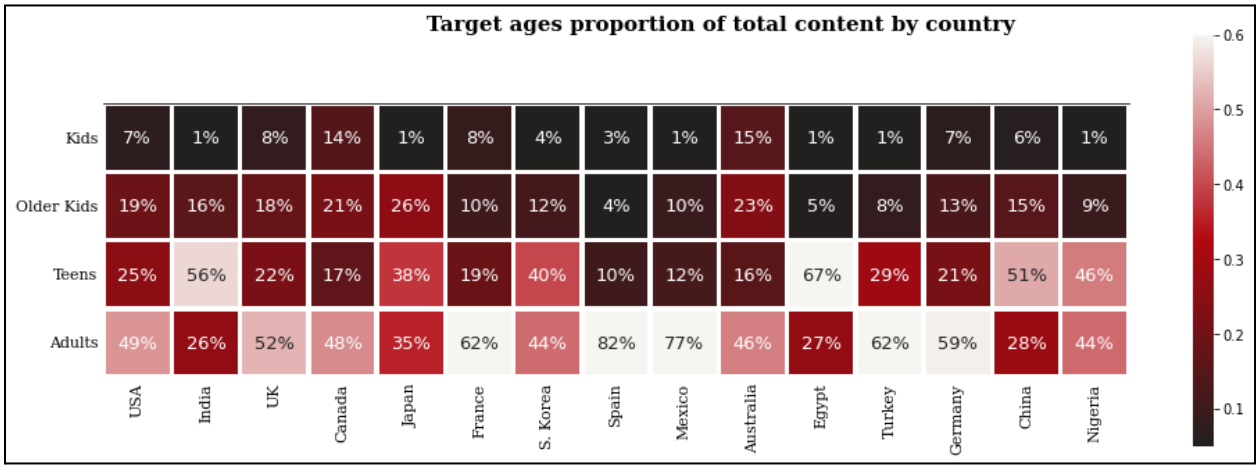
We see that December and July have the highest number of content added, which also coincides with the winter and summer vacation period, where people will be directed towards watching online content. This will be helpful for our users like Netflix and Content producers to know so they can add more relevant content based on their target audience during this period.



We observe that the Americas region has the highest number of content added over the months, with the highest during December, July, and September. The months of December and July are vacation periods; hence our consumers would watch more content online therefore, our users can develop more content that can be streamed during those months. We can see a content slump in February for the Americas. Similarly for the Asia region, we see an uptick in December, April and March. While for Europe, Oceania and Africa regions we do not necessarily see a pattern for specific months with an almost equal amount of content added over the months. Based on our hypothesis, we do see strong evidence for the Americas region but not for Europe.

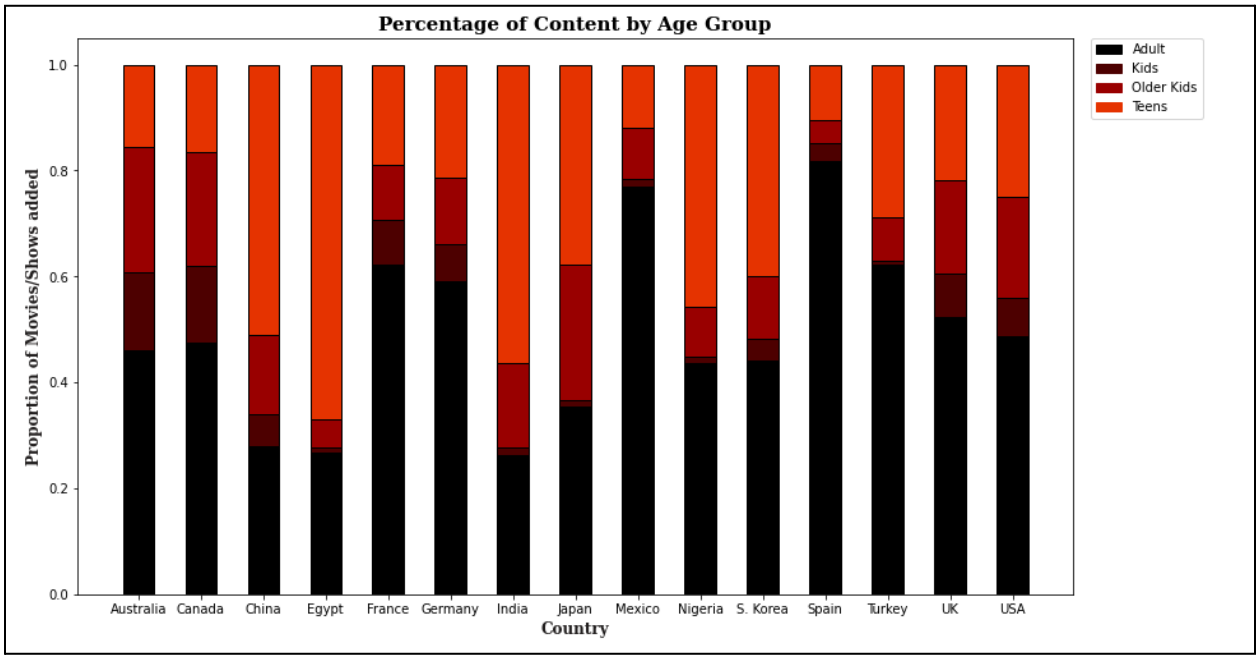


Hypothesis 2: Western countries produce more mature content in comparison with eastern counties that are more conservative

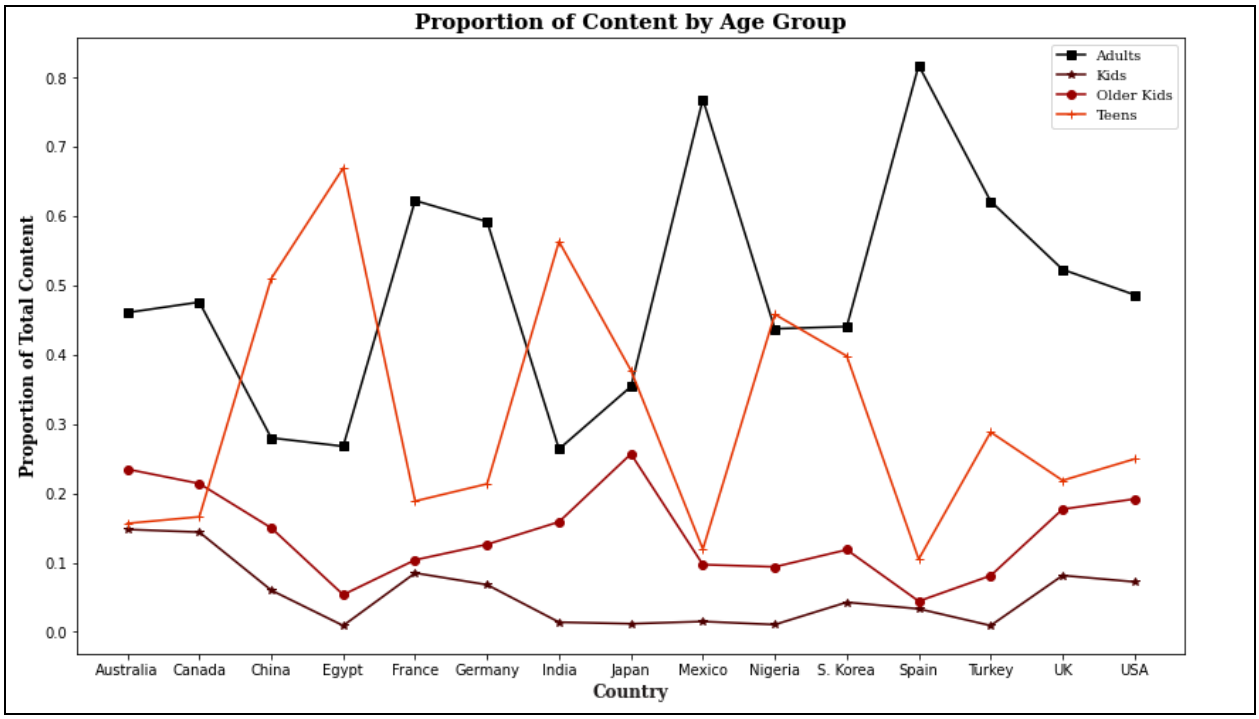


Now we can see that there are significant differences in the target ages between different countries. For India, Japan, South Korea, Egypt, and China, which are mostly in Asia, it is true that most Netflix contents are for teenagers. However, in some Western countries like the USA, UK, Canada, France, Spain

and Australia, a large proportion of the movies/TV shows are for adults. It shows that the target ages significantly differ between Western and Eastern cultures. People in the east tend to be more conservative about movie content, and Netflix should pay attention to this difference and produce or purchase movies/TV shows accordingly. One big problem with the plot is that the heatmap can clearly show the target age proportion. However, we still need to compare the numbers manually. To be more efficient, we first choose also to make a cumulative bar chart.

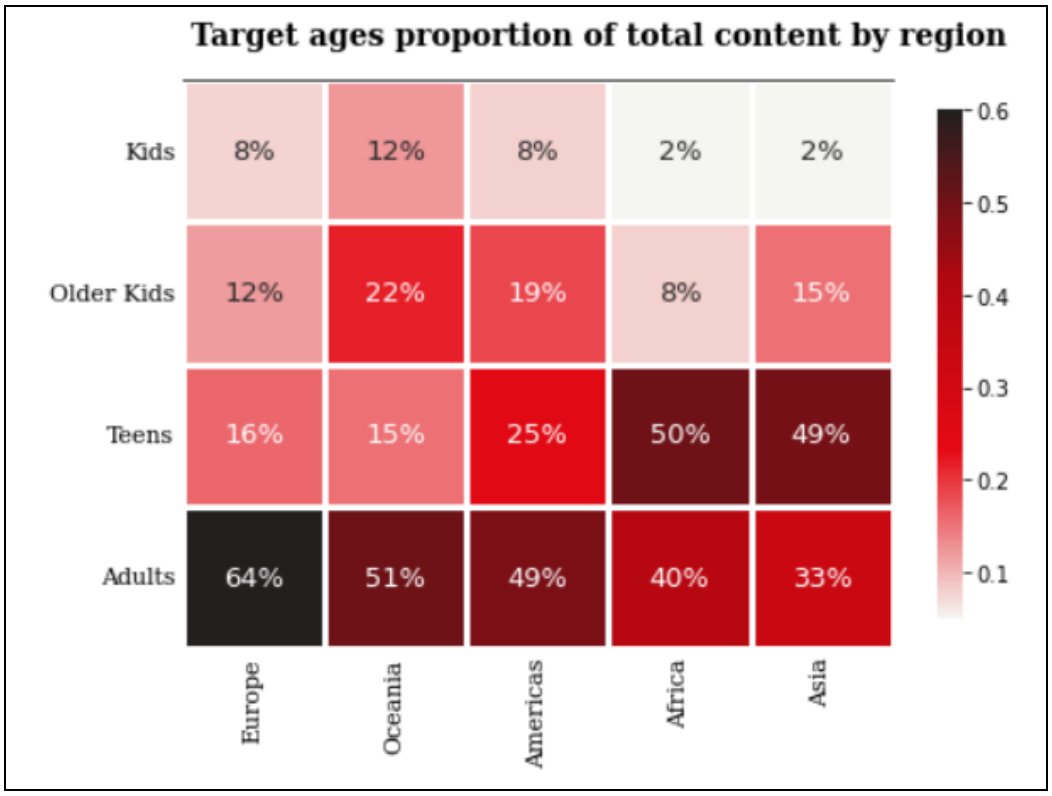


Here is the cumulative bar chart. The result support our hypothesis well, but it is still not straightforward. To improve upon this limitation, we chose to make a line chart to observe the difference between western countries and eastern countries.



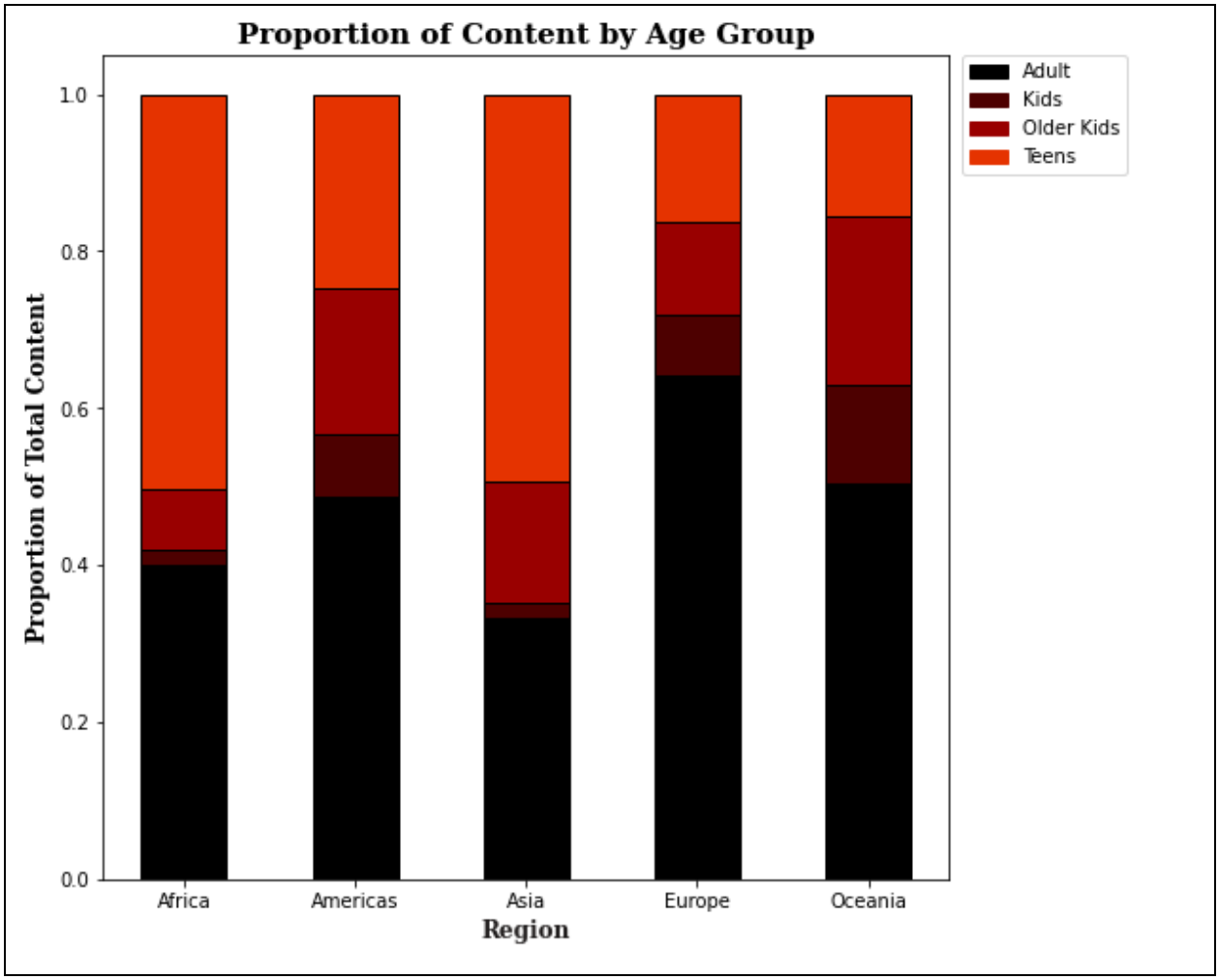
Based on the line chart, we can clearly confirm that Eastern countries may be more conservative about video content than western countries. Majority of content falls into the age category of teenagers (13-18) in eastern countries but not in western countries.

Further analyzing with respect to regions

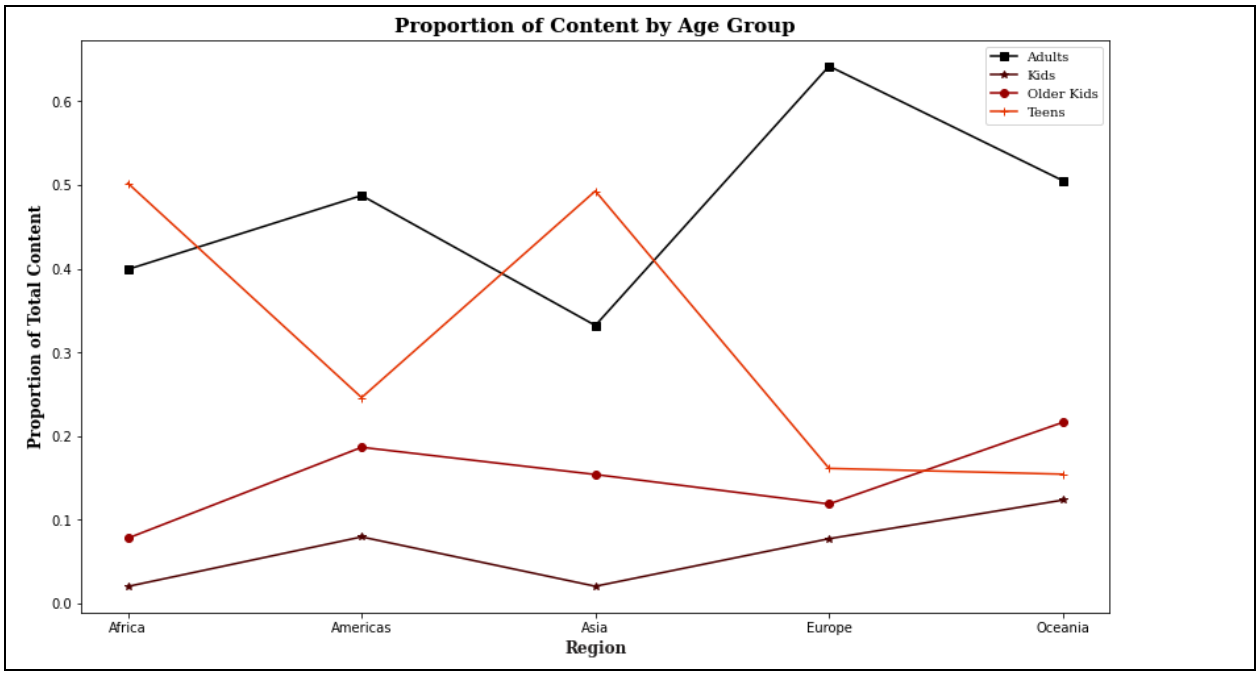


This visualization shows proportion of total content produced for each age category in each region.

Now, when we drill down to regions, we can see Europe produces the highest proportion of adult-rated content. Africa and Asia produce least proportion of Kids rated content and nearly half of the content is suitable for teens. At the same time, almost half of all the content produced in Americas and Oceania are adult-rated.

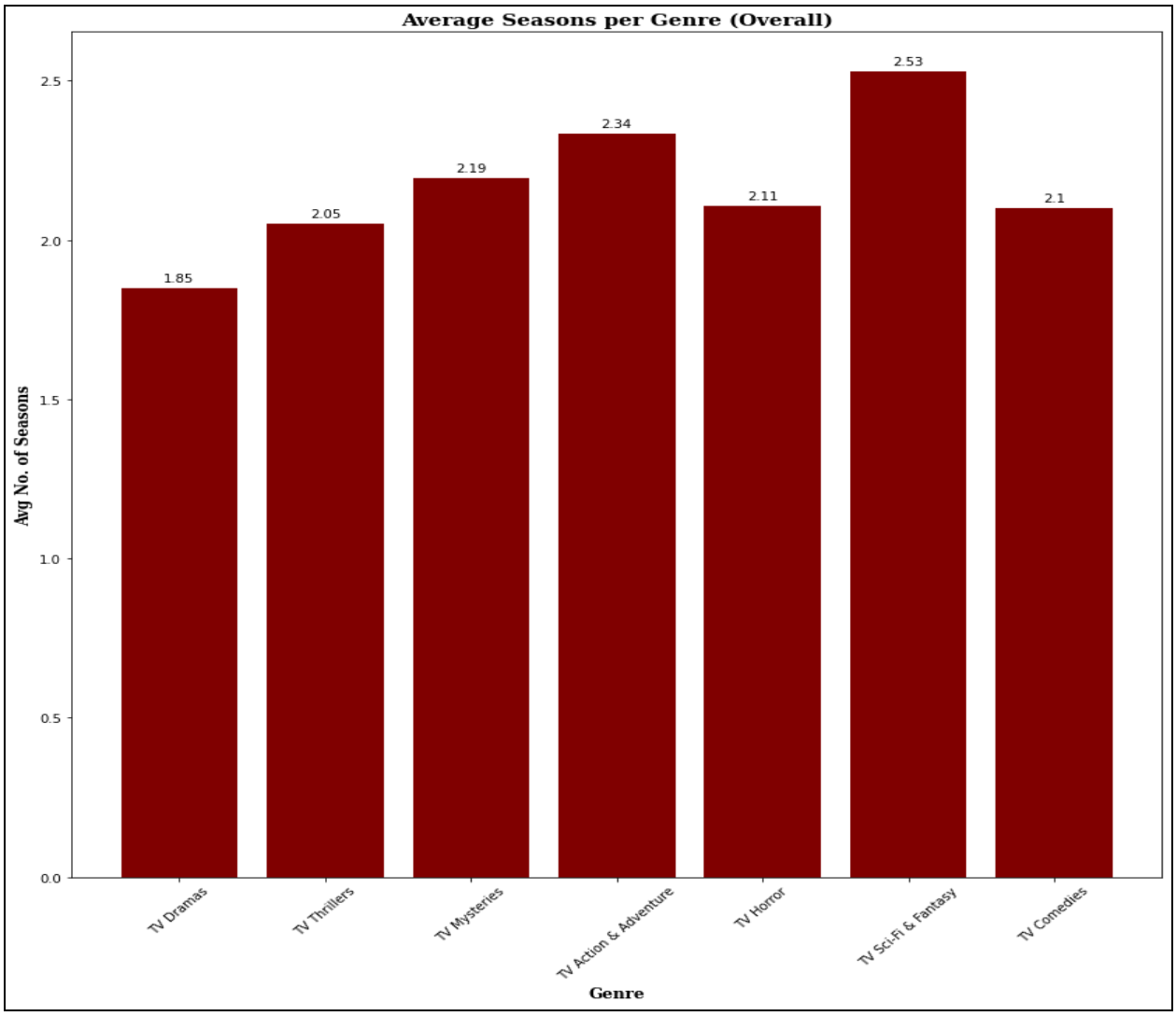


Based on the line chart and bar chart, we can more clearly see that Asia produces the least proportion of Adult rated content. Proportion of mature content produced is highest Western region (Europe, 64%) and lowest in Eastern region (Asia, 33%)

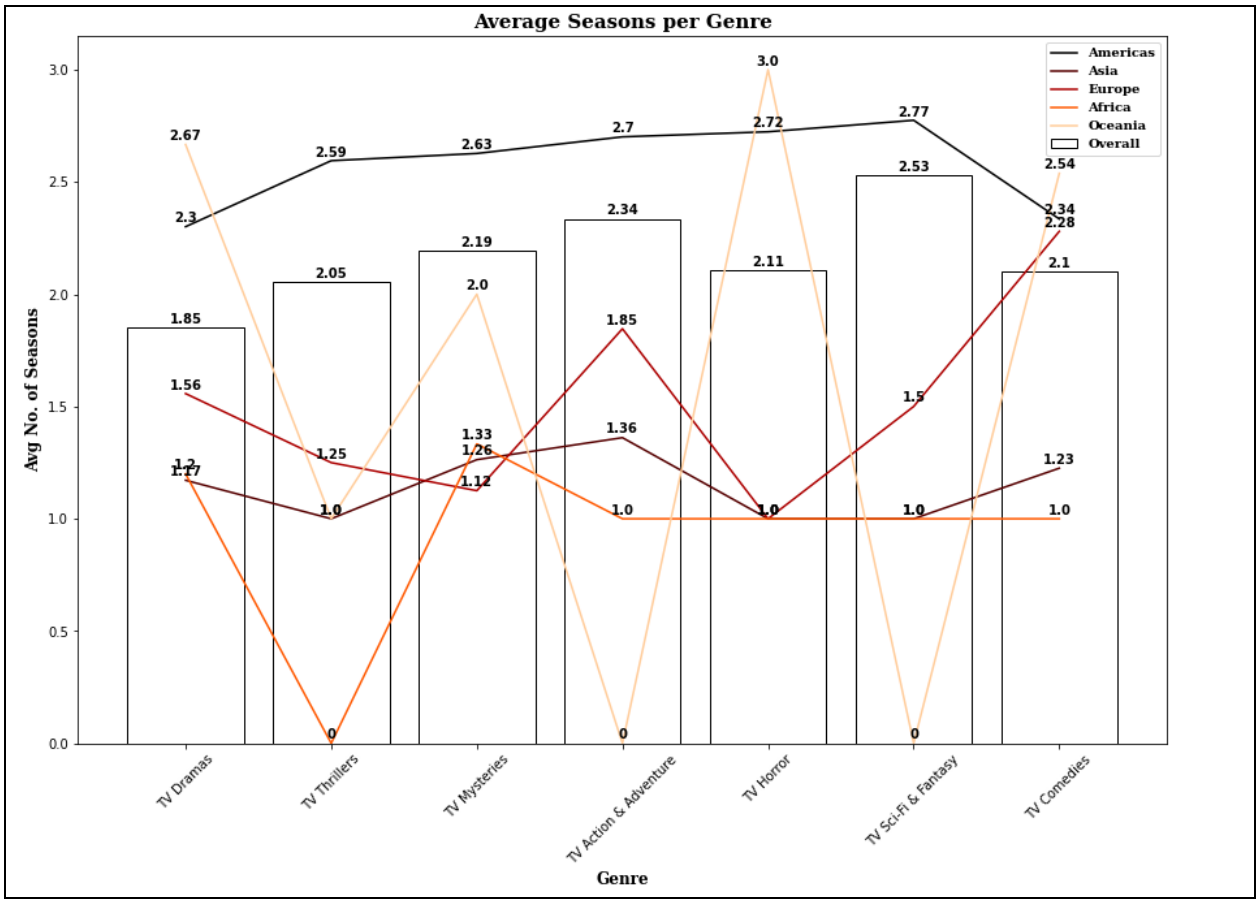


Hypothesis 3: Among all genres for TV Shows, comedies have the highest number of seasons

To assess the above hypothesis we first plot the average number of seasons for TV Shows by genre, looking at the plot we may be tempted to conclude that the hypothesis stated is false as clearly Sci-Fi and Fantasy has the highest average number of seasons at 2.65.



To explore further, this plot is extended to view the average number of seasons sub sectioned by region. Observing this plot we can see that our hypothesis indeed proves to be true for Europe at 2.28 seasons. Hence sub sectioning data by region gives deeper insights and helps make better region specific decisions for content production.



Conclusion

Through our exploratory visualization, we found the following insights:

1. There is a surge in content production during the summer and winter holidays for the Americas and Asian regions.
2. Western countries tend to have a higher proportion of mature content, while eastern countries have a higher proportion of teen-friendly content.
3. Amongst European countries, comedy TV shows are likely preferred as they run for more seasons