**Data Visualization**

On analysis of the YouTube dataset generated through YouTube API, we identified different types of relationships between the predictor variables such as views, genre, subscriber and average revenue per YouTube channel. These relationships are represented using various types of visualization plots for better understanding and analysis by the brand advertisement teams to identify the right type of channels for better market reach. The plots also help us understand the different type of genres that makes these YouTube channels popular.

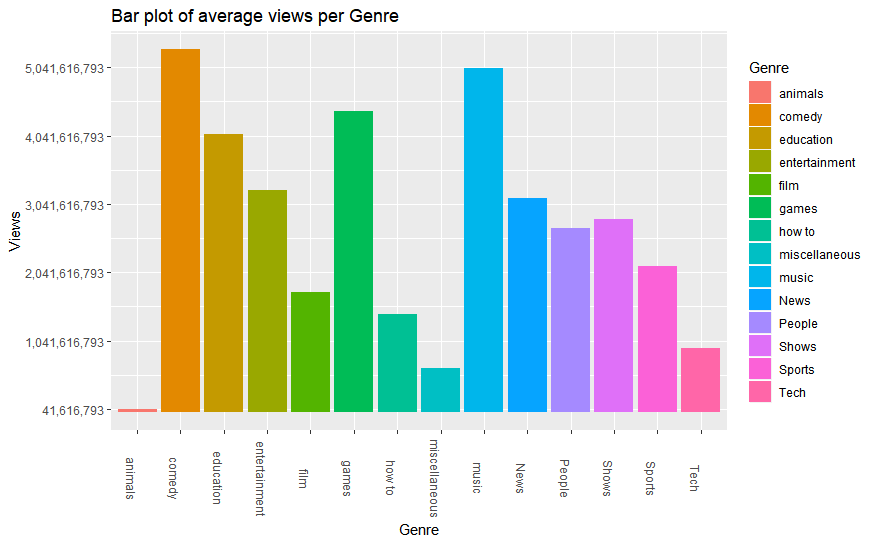


Fig 1. Bar plot of average views per genre.

Figure above represents a bar plot of the average number of views per YouTube Genre. These genres represent the different type of video content available in the YouTube community. The bar plot above helps us identify the top YouTube Genres that have generated significant amount of views and have made these genres popular. We see that genres Comedy, Music and Games are the top three genres that have managed to attract maximum viewers where Comedy has generated approximately 5.3 billion views, Music has generated 5.04 billion views and Games genre has generated approximately 4.3 billion view. These types of genres have generated billions of views over the years and are the current trending Genres in the YouTube community. YouTube channels featuring video content in these genres are likely to gain a greater number of views and attract more users and possible subscribers than other YouTube channels featuring other genre content. The bar plot also shows us that animals genre content have generated the least number of views. This representation will help the brand advertisement teams in analyzing the different type of genres that they can consider for featuring their products for more advertisement and possible increase in sales value.

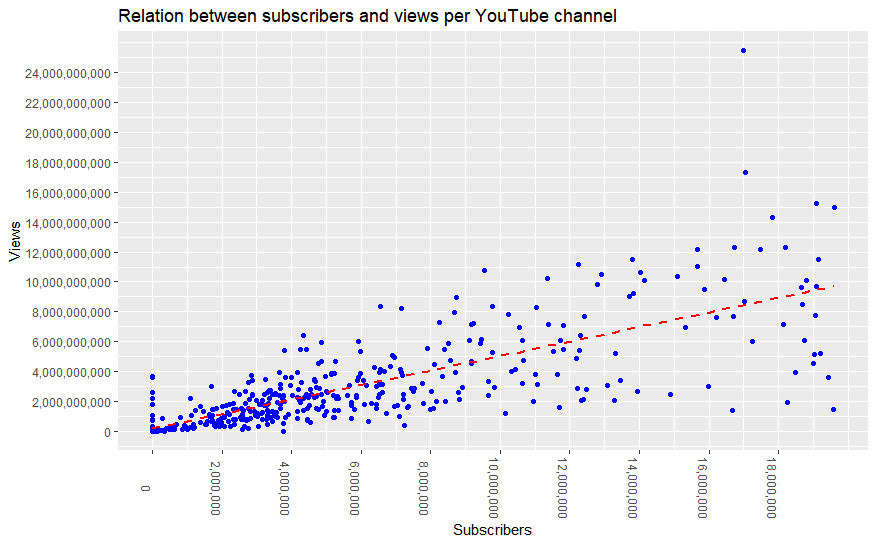


Fig 2. Relationship between subscriber and views per YouTube channel

Figure above represents the correlation between the number of subscribers and the number of views per YouTube channel. We see that majority of the YouTube channels have less than 5 million subscribers and the number of views generated by these channels range between 0 to 4 million. The scatterplot with the best fit line above helps us identify the relation between subscriber and views to identify if these factors actually have an impact on YouTube channel's popularity. It also helps us analyze the strength of the subscriber variable as a predictor towards the outcome variable as number of views.

Figure above shows us that the number of subscribers and the number of views have a positive correlation between them. This indicates that with an increment in the number of subscribers, there would be a strong positive increment in the number of views too. Therefore, the variable "number of subscribers" can be considered as a strong predictor of the YouTube channel's "number of views". This indicates that with an increment in the number of subscribers, there would be a strong positive increment in the number of views too. There are also certain outlier data such as the channels with 17 million subscribers and with more than 24 billion views. These are YouTube channels with possible viral content that has generated a large number of views. The above representation will help the marketeers to identify if subscribers and views can be considered as effective measures to decide on their content partners for brand advertisement.

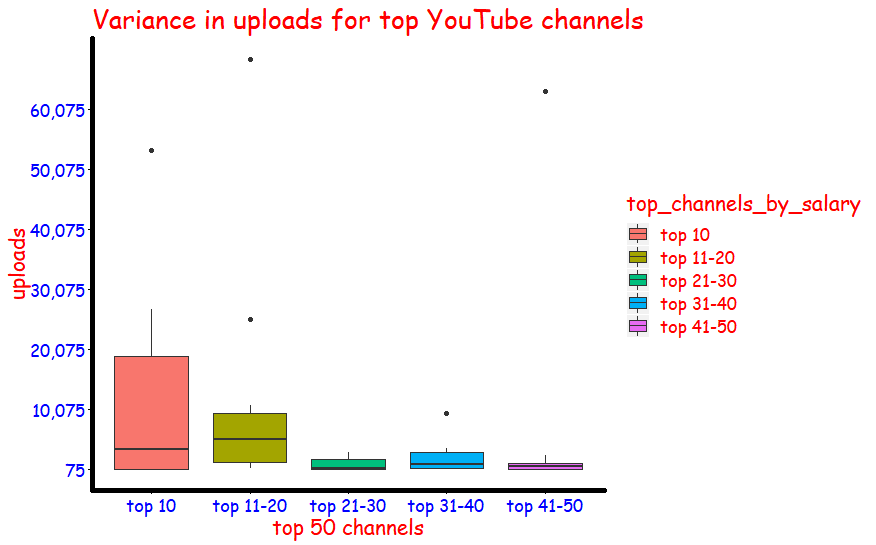


Fig 3. Variance in uploads for top YouTube channels

Figure above represents a Box Plot of the top YouTube channels categorized by the average revenue generated by these YouTube channels. The above boxplot visualization helps us identify the number of uploads and their variability by the top 50 YouTube channels. As seen from the boxplot above, we see that the top 10 channels have the maximum variability in terms of videos uploaded. The range of values covered above the median of the top 10 channels are on an average greater than the number of uploads by the rest of the top channels. As observed, the upper quartile range for categories top 11-20, top 21-30, top 31-40 and top41-50 are higher than the lower quartile range of the top 10 channels. However, the revenue generated by these channels is still less than the channels covered in the top 10 bin. This might be because the top 10 channels may have brand collaborations and businesses linked to their channels which help them in revenue generation however we might not be able to analyze those factors on the basis of the current dataset available.

This box plot can help the brand advertisement teams to evaluate the top 50 channels and their number of uploads and identify the channels with the required number of uploads for maximum outreach.