**TED TALKS ANALYSIS**

**Problem Statement:**  
The main objective of the project is to explore the TED Talks data and generate interesting insights from it. This comprises mainly understanding the trend of popularity of ted talks over the years in terms of views, comments and ratings, also to add with exploring its possible drivers like occupation of the speaker, duration of the ted talk, number of speakers etc.  
  
**Implementation:**  
The data has been obtained from Kaggle which covers all audio-video recordings of TED Talks uploaded to the official TED.com website until September 21st, 2017. After cleaning the data and bringing it to an easily usable format using text mining techniques it will be analyzed graphically to generate inferences.

**How does it help the consumer?**  
The analysis will be useful to the consumer in understanding where TED Talk is heading over the years with respect to their popularity and identify its drivers. It will ultimately help them design the best ted talks and avoid the mistakes of the worst ones.

**Packages used:**

1. **jsonlite**: Used to read in data that exist as JSON
2. **stringr**: Used for character manipulation and cleaning the data
3. **tidytext**: Used to tidy text
4. **tidyverse**: Used for data manipulation and visualizing data better using ggplot
5. **wordcloud**: creating a word cloud
6. **DT**: To display the data on HTML in a scrollable format
7. **Lubridate**: To work with dates easily
8. **reshape2** : To reshape the data to a matrix to be used in wordcloud comparison
9. **plotly**: To plot interactive box plots

**After performing various tasks like Data Cleaning, Data Exploratory Analysis on the data, we have a n ample amount of conclusions to make on :**

* *Number of talks have been increasing every year with average number of talks of 212.5.*
* *In 2012, there was a peak in the number of talks with more than 300 talks that year which is the highest number of talks till date*
* *We can say on an average we can expect around 210 or more talks every year*
* *The highest average number of views was ~ 4.1 million observed in 2006 which had the lowest number of talks.*
* *This was kind of expected as it was the year when TED Talks started offering free viewing online.*
* *Although 2012 had the highest number of talks the views were below average.*
* *As expected the 2006 which had the maximum average views also observed the maximum average number of comments with 363 comments per talk followed by 2013 with 289 comments per talk. Post 2013 we see a steady decrease in the number of comments with 2016 having only 81 comments per talk which can be a possible reason for concern. Let’s look at ratings to confirm if less comments also come with lesser ratings.*
* *The average number of ratings also show its highest value in the 2006 but, 2013 which had the second highest views and comments did not receive many ratings. On the other hand, 2010 showed a peak in the number of ratings.*
* *From the graph above we observe that the percentage of negative, positive and neutral ratings have stayed more or less the same over the years. The year 2009 is the only year which have slightly higher ratings which can be further deep dived to find the reason behind it.*
* *The good news is that the percentage of negative ratings have been reducing since 2010 with the least in 2017.*
* *With an average of 1.8 million views, TED Talks are clearly very popular.*

**Summary:**

* The main purpose of this analysis was to help the readers to see the popularity trend TED talks over time and what makes the TED talks best or worst.

**Factors affecting popularity of TED talks:**

* *Month of publishing does not affect the popularity of the talk.*
* *Best talks fall both in single and co-speaker talks however, majority of them being single speaker talks.*
* *The views of the TED Talks increase with increase in the number of languages the talk is available in.*
* *TED talks that have words like - issues, technology, politics are less viewed whereas TED talks that have words like psychology, brain, work, culture, humor ,love and happiness are most viewed.*
* *Duration of the talks is not related to the number of views however, the least viewed category contain the longest talks.*

**Insights from the analysis:**

* *Over the years TED talks have grown with respect to all the 4 parameters - number of talks they conduct, number of views, number of comments and the % of positive rating*
* *2006 being the first year when TED talks was launched for free viewing online it was the year of highest average views inspite of the least number of talks*
* *Another successful year for them was 2013 when the average number of views and comments peaked*

Thus, now we know the ingredients to create the BEST TED talk.

**5.4 Limitations of the analysis or future scope:**

Another very important factor that affect the popularity of TED talks is the content of the talk or transcripts. This, can be further added into the analysis. As a further scope a regression model or a neural network can be built to predict the popularity of a talk based on its characteristics.