**Project Report**

***Problem statement***

Is there opportunity to increase numbers of people who watch TED Talks online by 20 % through checking and calculating numbers of view and comments on talks which provoke the most online discussion by the next years?

**TED Conferences, LLC** (**Technology, Entertainment, Design**[[7]](https://en.wikipedia.org/wiki/TED_(conference)" \l "cite_note-about_TED-7)) is an American-Canadian non-profit[[7]](https://en.wikipedia.org/wiki/TED_(conference)#cite_note-about_TED-7) media organization that posts international talks online for free distribution under the slogan "ideas worth spreading".

People usually enjoy TED talks because of the engaging style of the speakers. Each speaker has a different style which, is interesting, but they accomplish the same goal: captivate the audience and make them to discuss different topics of the viewed talks.

Recently, the Executive team expressed a concern in a decline in numbers of people who discuss topics online.

The problem is very relevant for TEDx since it can potentially increase its popularity as it gives recommendations of interest to their clients.

The data is available on:

<https://www.kaggle.com/datasets/rounakbanik/ted-talks>

The above link has contents which include two CSV files, as below:

ted\_main.csv - Contains data on actual TED Talk metadata and TED Talk speakers.

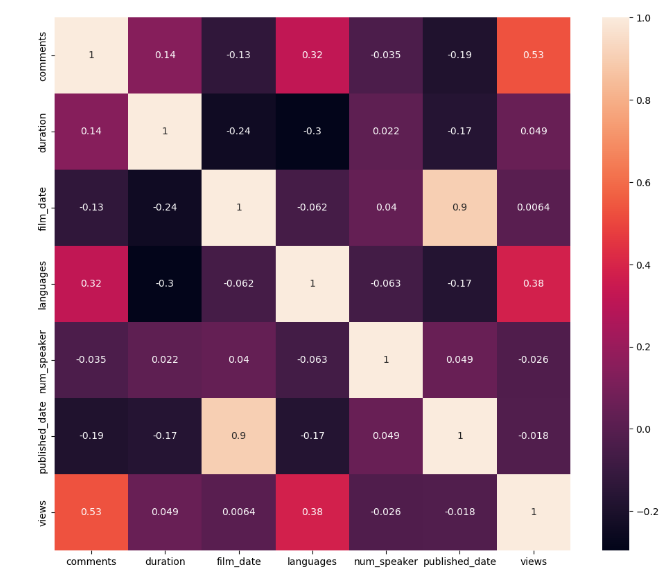
transcripts.csv - Contains transcript and URL information for TED Talks

[TED: Ideas Worth Spreading](https://www.ted.com/).

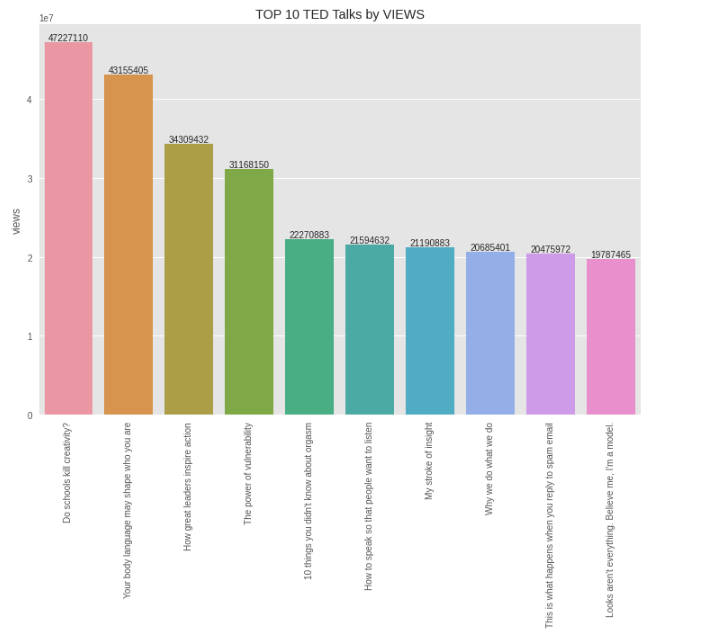
I used the following csv file to analyse our data: '/kaggle/input/ted-talks/ted \_main.csv'

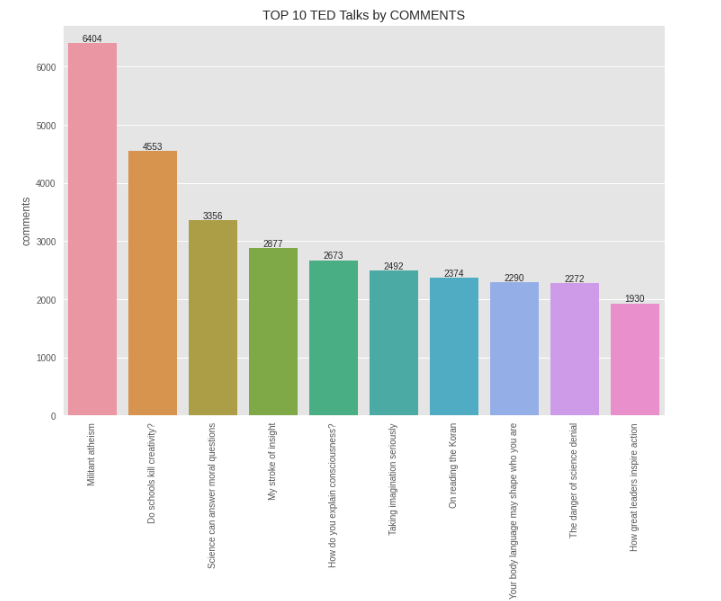
About the DataSet features:

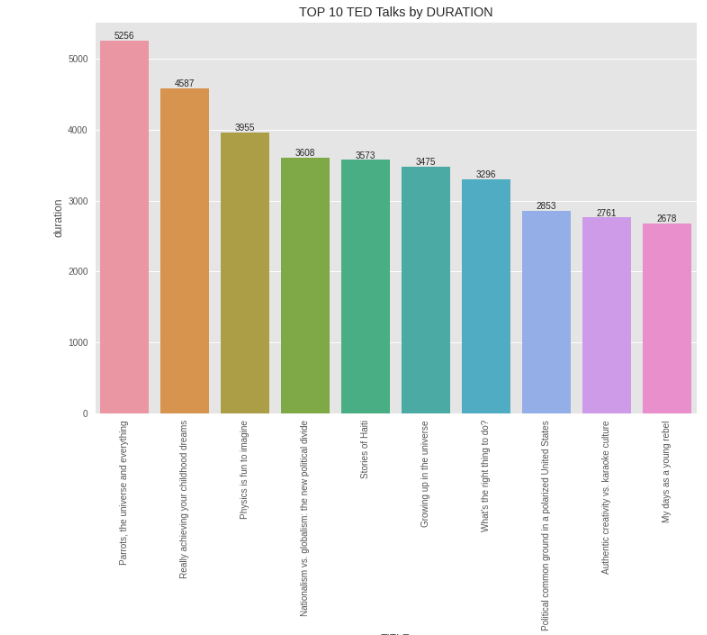
* comments: The Number of comments of the talk
* description: A summary of what the talk was about
* duration: The Duration of the talk in seconds
* event: The Event where the talk took place
* **film\_date**: The Date on which the talk was filmed/recorded.
* **languages**: The number of languages in which the TED talk is available.
* **main\_speaker**: The main speaker of the talk
* **name**: Includes Title and name of the speaker.
* **num\_speaker**: Number of speakers in the talk
* **published\_date**: Date when the TED Talk was published
* **ratings**: A Dictionary of various ratings given like Courageous, Inspiring, etc
* **related\_talks**: A list of talks recomended to watch next
* **speaker\_occupation**: Occupation of the speaker
* **tags**: Different themes related to talk
* **title**: Title of the TED Talk
* **url**: The URL of the TED Talk
* **views**:Number of views on the TED Talk



From the heatmap, the number of views correlates well with language and comments.









Observations:

* Do schools kill creativity? by Ken Robinson is the most viewed talk with 47.2 Million views and 4553 comments.
* Militant atheism by Richard Dawkins has the most comments i.e 6404.
* Parrots, the universe and everything by Douglas Adams has the maximum duration with 1.46hr

Based on above charts were able to see the main topics which took attention of viewers. These analysis could help us to prioritize talks and increase numbers of people who would discuss what they watched.

The above mentioned charts could help the Executive team tackle the concern about why they had a decline on online discussion and might help them to decide their future steps and plans in order to implement new ways to achieve their set goals.