# Project Proposal: Ted Talks

**Summary:**

This data set is about the Ted Talks. As TED is one of the best learning platforms from the experiences of the best people from different fields. So, we have collected this data from the Kaggle to drive data analysis on the Ted Talks. As we all know Ted Talks are organized by a specific author and this would be rated based on the number of views and likes for that program. We want to explore the data set and find the popular TED talks Speaker with more views and likes, Month & Year-wise analysis on the Ted talk frequency, Popular Ted Talk shows etc.

This data set consists of Show title, Author/ Speaker of the show, Date at which Show was conducted, Number of views, Number of likes, link to the Ted talk. This data set contains the data recorded from Feb 1984 to Feb 2022. This data set can be divided into two tables namely; the main table & the Author table. In the main table, the title of the show serves as a primary key while in the Author Table, Author\_id serves as a unique identifier. There is a possibility that one author could performer one or more shows. Therefore, we need to create a Relational Database table.

Steps that will be followed in this project:

1. Collect data and divided big data into two different tables (main table, Author table).
2. Loaded the data files into the RDS.
3. Building the relationship among the tables using Entity- Relationship Diagram.
4. SQL codes will be written to query the data to check the popular shows based on the views & likes, to find most popular Ted talk speaker, Month & Year-wise Ted shows frequency.
5. Using this cleaned data and Data Visualization tool, will perform the data analysis and display the important information using graphs.