Project 1

These datasets contain information about all audio-video recordings of TED Talks uploaded to the official TED.com website until September 21st, 2017. The TED main dataset contains information about all talks including number of views, number of comments, descriptions, speakers and titles. The TED transcripts dataset contains the transcripts for all talks available on TED.com.

The dataset has two csv files and the content for the csv files are listed below:

TED Main Dataset

- **name:** The official name of the TED Talk. Includes the title and the speaker.
- **title:** The title of the talk
- **description:** A blurb of what the talk is about.
- **main_speaker:** The first named speaker of the talk.
- **speaker_occupation:** The occupation of the main speaker.
- **num_speaker:** The number of speakers in the talk.
- **duration:** The duration of the talk in seconds.
- **event:** The TED/TEDx event where the talk took place.
- **film_date:** The Unix timestamp of the filming.
- **published_date:** The Unix timestamp for the publication of the talk on TED.com
- **comments:** The number of first level comments made on the talk.
- **tags:** The themes associated with the talk.
- **languages:** The number of languages in which the talk is available.
- **ratings:** A stringified dictionary of the various ratings given to the talk (inspiring, fascinating, jaw dropping, etc.)
- **related_talks:** A list of dictionaries of recommended talks to watch next.
- **url:** The URL of the talk.
- **views:** The number of views on the talk.

TED Transcripts Dataset

1. **url:** The URL of the talk
2. **transcript:** The official English transcript of the talk.

Acknowledgements

The data has been scraped from the official TED Website and is available under the Creative Commons License.

Project Ideas

Some of the questions that can be directly answered with this dataset:

1. How is each TED Talk related to every other TED Talk?
2. Which are the most viewed and most favorited Talks of all time? Are they mostly the same? What does this tell us?
3. What kind of topics attract the maximum discussion and debate (in the form of comments)?
4. Which months are most popular among TED and TEDx chapters?
5. Which themes are most popular amongst TEDsters?

You may also want to go a step ahead and predict if positive discussions from a TED talk have any gain on the author's scholarly reach (in terms of citation). However, to address this you may want to collect some data on your own that are not part of the existing dataset.