**What Makes A TED Talk Popular?**

1. **Introduction**

TED is a platform for sharing ideas, and TED Talks are influential videos from expert speakers on education, business, science, tech, creativity and so on. [1] In recent few years, TED Talks appeal to global citizens due to various merits, such as they are short but focused, shot like movies, highlight diverse speakers, discuss specific inspiring and informative topics, go beyond classrooms and offices. [2] And all these advantages are exactly the motivation that make our group willing to explore what else makes a TED talk popular with our knowledge.

1. **Objectives and Expected Contributions**

The main objective of our work is to investigate what make a TED Talk popular within the following aspects:

* Is there any relation between speech duration and speech popularity? How does the duration affect the popularity?
* Which aspects are people interested in? Which tags for a TED Talk can make it have more views? Does number of tags for a TED talks affect the views?
* Which speakers are popular?
* Are there any specific words can make a topic gain more click? How about the sentiment for these words?
* How does the speech description affect view?

We expect our work can make following contributions:

* This work can provide insights about what makes a TED Talk popular. Through sentiment analysis, we could provide tips for speakers about how to create the topics and descriptions that can inspire people’s interest most.
* We can give advice about which aspects’ speeches should be made more.

1. **Methodology**
   1. **Data Collection**

Asfalkfj

* 1. **Data Preprocessing**

The scraped durations for each video were in seconds, which was not easy to read while easy for data analysis. So firstly, we transformed the number for the durations to a more straightforward way, for example from 1262 seconds to 0:21:02 (0 hours 21 minutes 02 seconds)and call them as lengths.

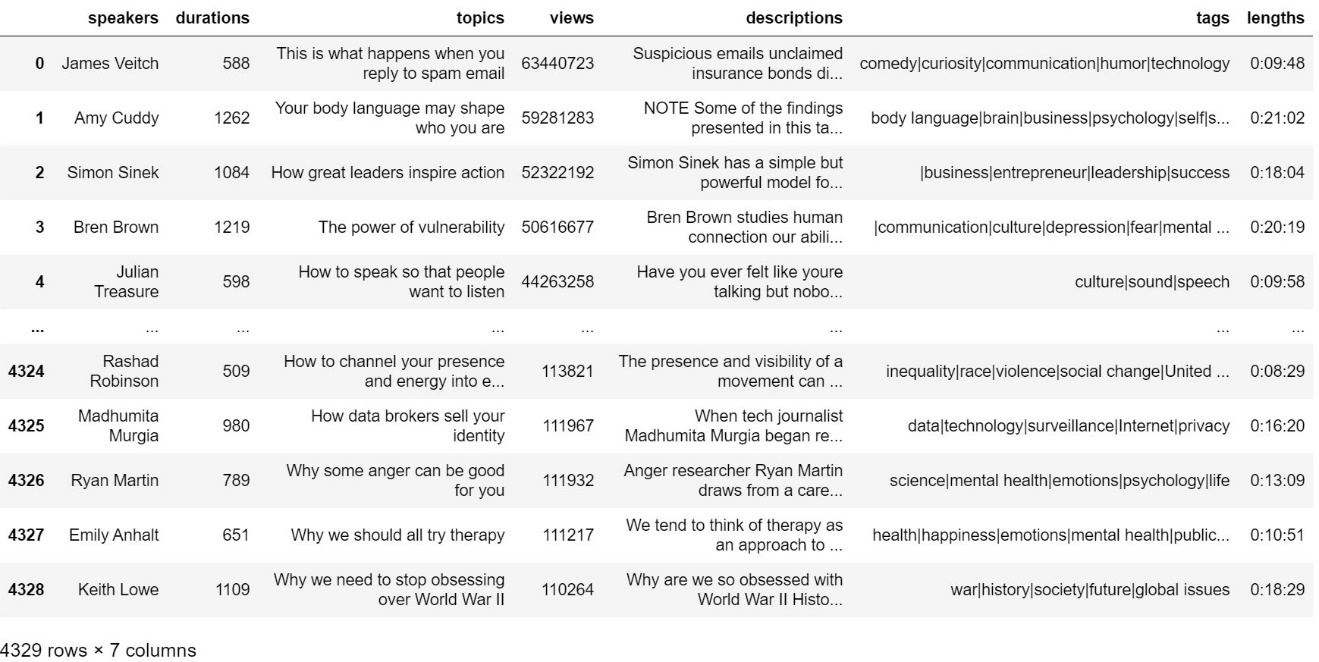
Secondly, we found there are some meaningless symbols in the text of speaker column, descriptions column and topics column, such as ‘?€?’, we removed these meaningless symbol for future better interpretation and text mining.

Moreover, we also removed some needless suffixes, such as ‘(TED)’ in lecturers’ name, ‘TEDx’, ‘TED-Ed’, ‘TEDMD’ etc in tags to avoid meaningless high-frequency speakers’ name or tags.

In some cases, if there are multiple speakers attend one same talk, the raw data used ‘+’ to connect speakers’ names. We replaced ‘+’ with the word ‘and’ to remain the same meaning.

1. **Exploratory Data Analysis**

To start with, we generated a data overview to display popular topics and the relational information about the topics.



Next, we tokenized the tags column and got the count of each unique tag, then sorted with descending sequence. Fig 1 below shows the top 15 tags. The topics of science, culture, health are far much popular than other topics. The 8th issue tags are global issues for most cases.

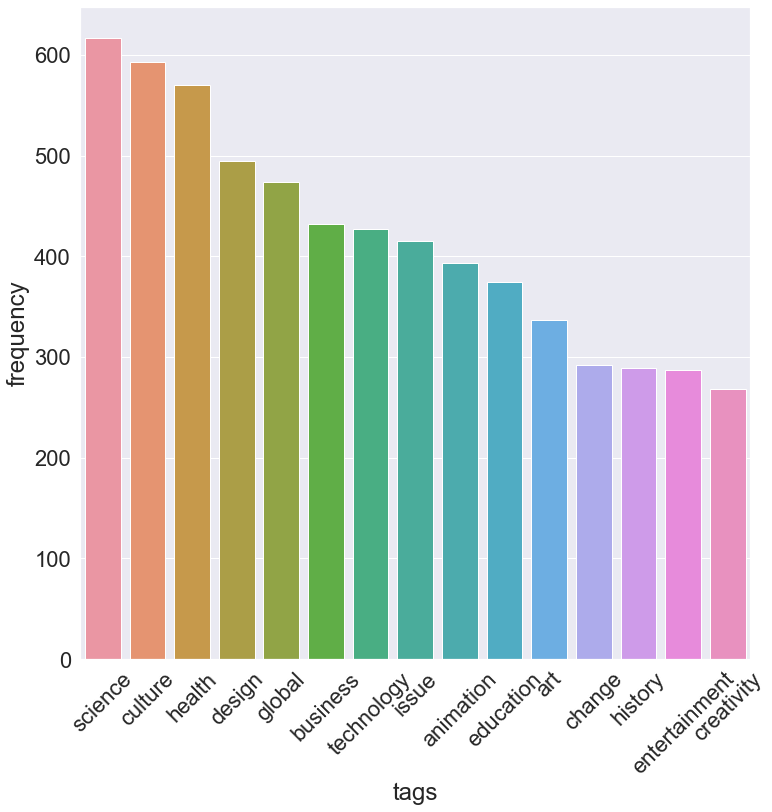
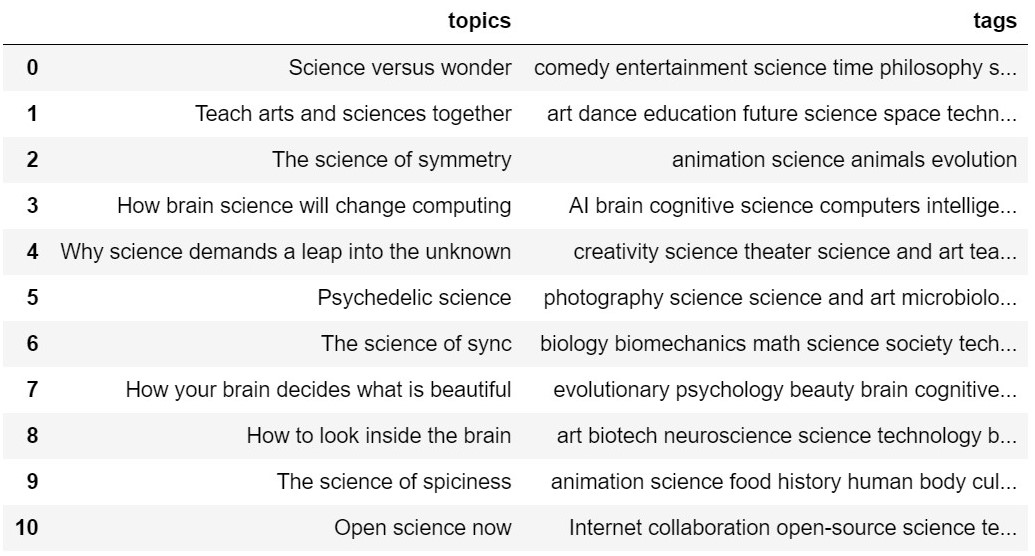
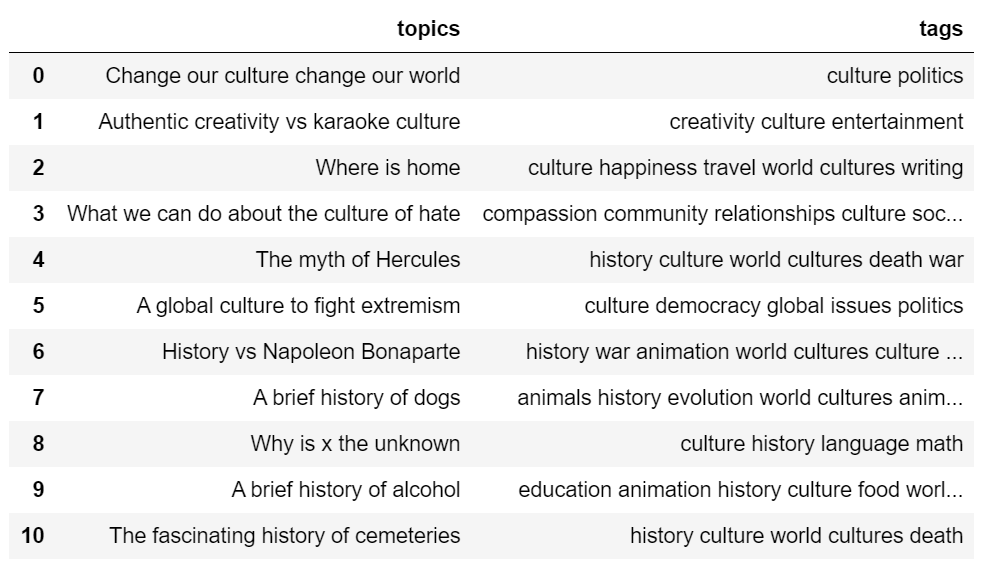


Fig 1 Top 15 tags

In addition to find top 15 popular tags for TED talks, we also implement tfidf algorithm to find out top 10 topics that highly related to tags. We can use them as auto recommender to recommend related topics to tags or use them for deep text mining later.



science tag related topics



culture tag related topics



Health tag related topics

Furthermore, we use tokenized topics column to plot word cloud to find most frequent words in topics. These words are make, new, life, world, future, good, help, solve, brain, learn, power, art, human, and so on.

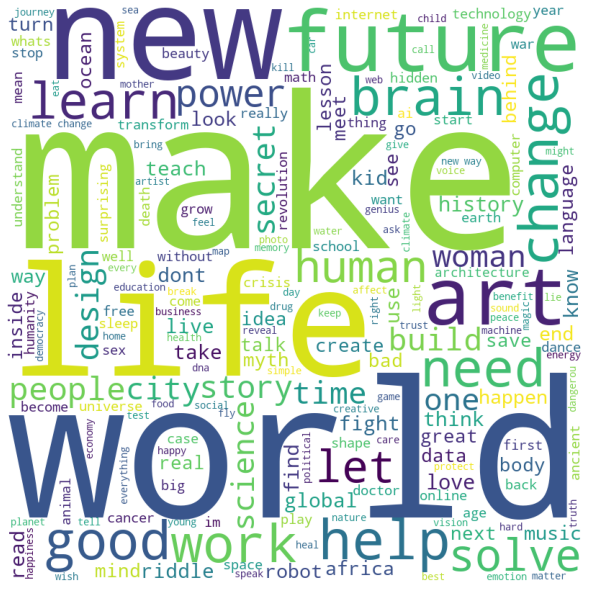


Fig 2 Topic word cloud

We sort our data by descending views and extracted the speakers who made the top 5 viewed TED Talks. Fig 2 shows the top 5 speakers. For the top 5 speakers, four of them are English- speakers and only one Bren Brown is French-speaker. English here is the most popular or most frequent language used in Ted Talks.

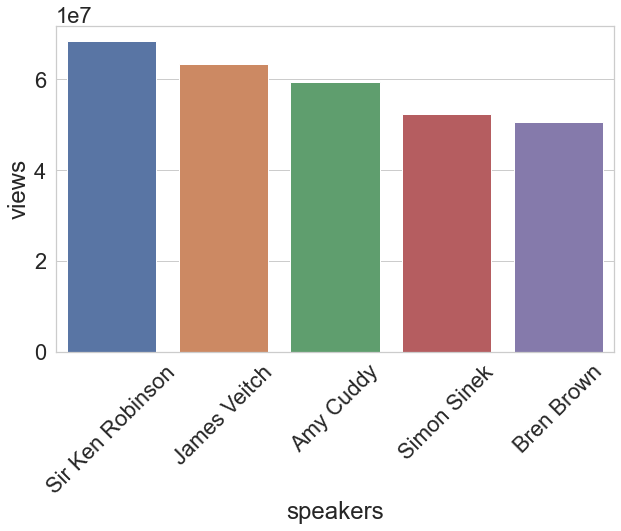


Fig 3 Top 5 speakers

We try to explore the data by implementing some naive sentiment analysis in both topics and description. We found that:

* count of positive words in topics are 1225
* count of negative words in topics are 1206
* count of positive words in descriptions are 1112
* count of negative words in descriptions are 927

It seems that positive and negative topics are evenly distributed. Then, we want to explore that whether audience preferences positive or negative topics by filtering data into top 500 popular topics. Then, we found that:

* count of positive words in top 500 topics are 195
* count of negative words in top 500 topics are 155

It seems that audience still evenly distributed but slightly trend to preference positive topics.

Moreover, we generated a figure to show the correlation between number of views and the number of tags in fig 3. The most viewed ted talks generally have 3 to 9 tags.

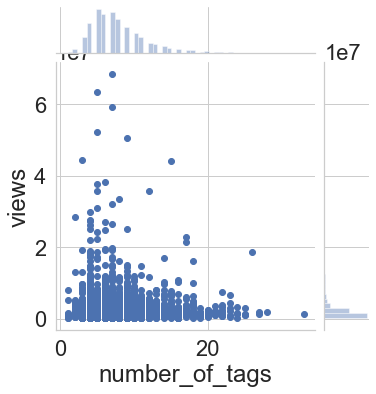


Fig 4 correlation between number of views and the number of tags

We generated a figure to show the correlation between number of views and the speech duration(unit: seconds) in fig 4. Most viewed talks have 60 to 2000 seconds of duration; That is, 1 minutes to around 33 minutes are the favorable length of talks.

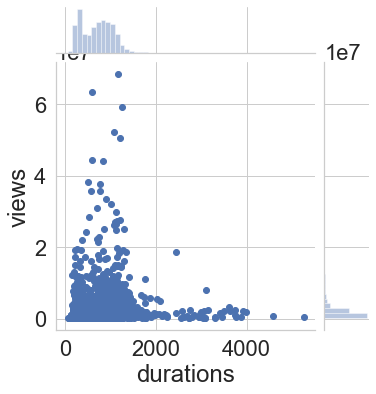


Fig 5 correlation between number of views and the speech durations

1. **Project Plan**

|  |  |  |
| --- | --- | --- |
| **Task** | **Assignee** | **Status** |
| data collection | Zhizheng Li, Yuhui Ren | complete |
| data preprocessing | Xin Zhao, Ruixi Wang | complete |
| exploratory data analysis | Ruixi Wang, Xin Zhao | complete |
| mid-term report | Yuhui Ren, Zhizheng Li | complete |
| feature extraction | Yuhui Ren | in progress |
| model selection/comparison i | Xin Zhao | in progress |
| model selection/comparison ii | Ruixi Wang | in progress |
| result analysis | Zhizheng li | in progress |
| poster creation | Yuhui Ren | in progress |
| research report writing | All | in progress |

**Reference**

[1] <https://www.ted.com/about/our-organization>

[2] https://thebiz.bentley.edu/9-reasons-ted-talks-are-so-popular/