

# TED Talks and #TED Tweets

IST 652  
Raya Young

# TED Talk database from kaggle.com

For this project, I used a dataset of TED Talks (TED and TEDx), updated as of May 2020. English table contained 4,005 rows of data.

These attributes include:

Passed into pandas dataframe for cleaning and exploration.

Talk ID	Title	Speaker 1	Occupations	About Speakers	Views	Date Recorded	Date Published	Comments	Duration	Topics	Related Talks	Description
0001	Averting the climate crisis	Al Gore	{0: ['climate advocate']}	{0: 'Nobel Laureate Al Gore focused the world'...	3523392	2006-02-25	2006-06-27	207020	00:16:17	['alternative energy', 'cars', 'climate change...]	{243: 'New thinking on the climate crisis', 54...	With the same humor and humanity he exuded in ...
0002	Simple designs to save a life	Amy Smith	{0: ['inventor', 'engineer']}	{0: 'Amy Smith designs cheap, practical fixes ...	1724438	2006-02-24	2006-08-15	100010	00:15:06	['MacArthur grant', 'alternative energy', 'des...]	{1561: 'Energy from floating algae pods', 1072...	Fumes from indoor cooking fires kill more than...
0003	How to rebuild a broken state	Ashraf Ghani	{0: ['president-elect of afghanistan']}	{0: 'Ashraf Ghani, Afghanistan's new president...	981920	2005-07-12	2006-10-18	7050	00:18:45	['business', 'corruption', 'culture', 'econom...]	{127: 'Want to help Africa? Do business here'...	Ashraf Ghani's passionate and powerful 10-minu...
0004	The real future of space exploration	Burt Rutan	{0: ['aircraft engineer']}	{0: 'In 2004, legendary spacecraft designer Bu...	2427994	2006-02-24	2006-10-25	109060	00:19:37	['NASA', 'aircraft', 'business', 'design', 'en...]	{141: 'Inside the world's deepest caves', 264...	In this passionate talk, legendary spacecraft ...
0005	Great cars are great art	Chris Bangle	{0: ['car designer']}	{0: 'Car design is a ubiquitous but often over...	978483	2002-02-02	2007-04-05	8010	00:20:04	['business', 'cars', 'design', 'industrial des...]	{4: 'The real future of space exploration', 26...	American designer Chris Bangle explains his ph...

# TED Talk database from kaggle.com

Speaker 1	Title
Alex Gendler	34
Iseult Gillespie	19
Emma Bryce	12
Daniel Finkel	11
Hans Rosling	10

With over 4,000 rows of data some speakers appeared several times.

While most talks are in the 10-20 minute range, further exploration showed that speakers with the most talks had a duration of under 10 minutes.

TED has a short animated series of lessons on their website.

Most other speakers appeared only once in the dataset.

# TED Talk database from kaggle.com

In contrast, the talk with the longest duration was over an hour, on the topic of climate change and environmental activism:

Talk ID	Title	Speaker 1	Occupations	About Speakers	Views	Date Recorded	Date Published	Comments	Duration	Topics	Related Talks	Description	Transcript
54715	How we can turn the tide on climate	Chris Anderson	{0: ['head of ted'], 1: ['climate advocate']}	{0: 'After a long career in journalism and pub...	1493370	2019-12-04	2019-12-12	4040	01:05:22	['climate change', 'environment', 'activism', ...	{32560: 'The disarming case to act right now o...	Witness the unveiling of Countdown, a major gl...	[Citizens of the world] [We face a global cris...

# TED Talk database from kaggle.com

Next, we explore the popularity and success of the talks themselves:

The average number of views is 2,148,006, and this talk by Sir Ken Robinson titled “Do Schools Kill Creativity” garnered 6,5051,954 as of May 2020

Talk ID	Title	Speaker 1	Occupations	About Speakers	Views	Date Recorded	Date Published	Comments	Duration	Topics	Related Talks	Description	Transcript
0066	Do schools kill creativity?	Sir Ken Robinson	{0: ['author', 'educator']}	{0: "Creativity expert Sir Ken Robinson challe...	65051954	2006-02-25	2006-06-27	40903010	00:19:24	['children', 'creativity', 'culture', 'dance', ...	{865: 'Bring on the learning revolution!', 173...	Sir Ken Robinson makes an entertaining and pro...	Good morning. How are you? (Audience) Good. It...

# TED Talk database from kaggle.com

Meanwhile, the average number of comments per talk (scraped from TED.com) was 239073.9, and the talk with the most comments was titled “Militant Atheism” by Richard Dawkins with 60,404,090 comments

	Title	Speaker 1	Occupations	About Speakers	Views	Date Recorded	Date Published	Comments	Duration	Topics	Related Talks	Description	Transcript
Talk ID													
0113	Militant atheism	Richard Dawkins	{0: ['evolutionary biologist']}	{0: 'Oxford professor Richard Dawkins has help...	5788514	2002-02- 02	2007-04- 16	60404090	00:29:10	['God', 'atheism', 'culture', 'religion', 'sci...	{86: 'Letting go of God', 94: "Let's teach rel...	Richard Dawkins urges all atheists to openly s...	That splendid music, the coming-in music, "The...

# TED Talk database from kaggle.com

```
from textblob import TextBlob  
  
text = str(views_tokens_list)  
  
blob = TextBlob(text)  
blob.tags  
  
blob.noun_phrases  
  
for sentence in blob.sentences:  
    print(sentence.sentiment.polarity)
```

```
0.7  
0.0  
0.7  
0.8  
0.2  
0.0  
0.4  
0.16666666666666666  
0.0  
0.0  
0.0  
0.0  
0.0  
0.65  
0.0  
0.0  
-0.05  
-0.05  
0.0  
...
```

```
text = str(comments_tokens_list)  
  
blob = TextBlob(text)  
blob.tags  
  
blob.noun_phrases  
  
for sentence in blob.sentences:  
    print(sentence.sentiment.polarity)
```

```
0.8333333333333334  
0.0  
0.0  
0.5416666666666667  
0.0  
0.0  
0.35  
0.2916666666666667  
-0.20625  
-0.02222222222222216  
0.05  
0.10714285714285714  
0.3  
0.0  
-0.3  
0.36767676767676766  
0.0  
0.13636363636363635  
0.0  
...
```

Though they are on unrelated subjects, the relative engagement and popularity of these talks is intriguing

The transcript of each talk was passed through the TextBlob package, and a sentiment polarity score was returned for each sentence.

Helpful next steps would be an overall score, as well as a frequency distribution of terms.

Later, a model could be trained to predict popularity / engagement based on certain attributes.

# #TED Tweets

Using a Twitter API, Tweets were scraped with the hashtags #TED, #TEDTalk, and #TEDx.

I wanted to know what was being said about the talks, which ones were popular, and if this popularity was driven by the news cycle.

40000 Tweets were pulled for each hashtag, with the expectation to match the number of rows in the English TED Talk .csv. However, only 4601 Tweets were available, with date stamps from 08-21 to 08-29



# #TED Tweets

	created_at
created_date	
2020-08-22	784
2020-08-26	604
2020-08-23	588
2020-08-25	507
2020-08-24	492
2020-08-28	442
2020-08-21	426
2020-08-27	425
2020-08-29	333

```
1 ,lang
2 en,3418
3 es,273
4 und,165
5 ko,125
6 ja,107
7 tr,104
8 fr,79
9 ar,76
10 it,40
```

Exploring the frequency of each attribute, and binning datetime by date, we note that the 8-22 was the day with the most Tweets at 784

At a glance, we see lots of languages represented: 25 total in the whole collection.

	created_date	text
1021	2020-08-22	RT @TEDxCDMX: Hoy más que nunca, recordemos es...
1022	2020-08-22	그럼 어째서 우리는 결코 사무엘 피어폰 랭리에 대해서 들어본 적이 없는가? \nTh...
1023	2020-08-22	RT @fahadalahmdi: هل تعرف ماهو الفرق بين السوا
1024	2020-08-22	@SirKenRobinson murió ayer 21 agosto 2020, gra...
1025	2020-08-22	RT @thewetbaguett3: Like for part 7! #ted #mov...
1026	2020-08-22	그들은 없었다. 우리가 성공의 레서피라 여기는 것들을 \nthey had none ...
1027	2020-08-22	Like for part 7! #ted #movie #moviescene #4u #...
1028	2020-08-22	ダニエル・H・コーエン: よい議論をするために\nhttps://t.co/llqGnURP...
1029	2020-08-22	RT @TEDxCDMX: Hoy más que nunca, recordemos es...
1030	2020-08-22	If you've never watched @SirKenRobinson's orig...

# #TED Tweets

Using the API, Tweets were scraped with the hashtags #TED, #tedtalk, and #TEDx.

The top 20 hashtags are displayed, still including all languages.

## Top 20 Frequency Hashtags

TedTalk 925

TED 704

Leadership 576

Management 382

TEDx 283

tedx 265

Tedx 210

ted 160

TEDTalk 107

教育 91

Ted 77

動画 74

tedxtcet 70

فهد\_الأحمدي 67

تيد 67

صباح\_الخير 67

talk 58

IFTHH 56

tedtalk 55

technology 52

# #TED Tweets

Out of the 4601 Tweets scraped, 3418 had the attribute lang = 'en', indicating English.

The average number of favorites in this dataset was 2, while the maximum was 818

	_id	created_at	text	entities	user	retweet_count	favorite_count	lang	created_date	created_hour
2920	5f4aa86f678bd9aa94f14169	2020-08-22 12:45:33	Good morning. If you haven't watched my #TedTa...	{ 'hashtags': [{'text': 'TedTalk', 'indices': [...	{ 'id': 17375057, 'id_str': '17375057', 'name':...	286	818	en	2020-08-22	12

# #TED Tweets

The average number of retweets in this collection was 28.8

while the maximum was 1235

	_id	created_at	text	entities	user	retweet_count	favorite_count	lang	created_date	created_hour
4600	5f4aa89bd6a56ce85825ae65	2020-08-21 05:44:10	RT @SandyAhlawat89: When civilian countrymen a...	{'hashtags': [], 'symbols': [], 'user_mentions...}	{'id': 2880746036, 'id_str': '2880746036', 'na...	1235	0	en	2020-08-21	5

# #TED Tweets

Placing the top Retweets into their own table, we can see an emerging pattern:

	_id	created_at	text	entities	user	retweet_count	favorite_count	lang	created_date	create
4600	5f4aa89bd6a56ce85825ae65	2020-08-21 05:44:10	RT @SandyAhlawat89: When civilian countrymen a...	{ 'hashtags': [], 'symbols': [], 'user_mentions...'	{ 'id': 2880746036, 'id_str': '2880746036', 'na...	1235	0	en	2020-08-21	
3298	5f4aa89bd6a56ce85825a94f	2020-08-28 21:57:27	RT @AttorneyGriggs: "From Civil Rights to Soci...	{ 'hashtags': [{'text': 'justicefighter', 'indi...	{ 'id': 32931825, 'id_str': '32931825', 'name':....	492	0	en	2020-08-28	
2821	5f4aa86f678bd9aa94f14106	2020-08-22 14:25:13	RT @hmcghee: Good morning. If you haven't watc...	{ 'hashtags': [{'text': 'TedTalk', 'indices': [...	{ 'id': 16586846, 'id_str': '16586846', 'name':....	286	0	en	2020-08-22	
2804	5f4aa86f678bd9aa94f140f5	2020-08-22 15:14:06	RT @hmcghee: Good morning. If you haven't watc...	{ 'hashtags': [{'text': 'TedTalk', 'indices': [...	{ 'id': 1476103248, 'id_str': '1476103248', 'na...	286	0	en	2020-08-22	
2805	5f4aa86f678bd9aa94f140f6	2020-08-22 15:12:45	RT @hmcghee: Good morning. If you haven't watc...	{ 'hashtags': [{'text': 'TedTalk', 'indices': [...	{ 'id': 1454712913, 'id_str': '1454712913', 'na...	286	0	en	2020-08-22	

Some RTs have slight variations in the text, but are essentially the same [2821, 2804, 2805]

# #TED Tweets

Consolidated further, we can see the most popular RT is in reference to speaker Heather C. McGee, with a talk titled “Racism Has a Cost for Everything.”

text	retweet_count
RT @hmcghee: Good morning. If you haven't watched my #TedTalk, please do — and let me know what you think.	50050
RT @hmcghee: Good morning. If you haven't watched my #TedTalk, please do — and let me know what you think. <a href="https://t.co/1sOAaRf0Ng">https://t.co/1sOAaRf0Ng</a>	15730
RT @maysoonzayid: TALK TO THE #DISABLED PERSON NOT THE NON DISABLED PARENTS. Thanks for coming to my #TedTalk	6810
RT @DuncanJWardle: Which is your favorite place to #ThinkDifferent? Mine is at my @TEDx talk on the Theory of #Creativity at @Tedx_AUK: <a href="http://t.co/1sOAaRf0Ng">http://t.co/1sOAaRf0Ng</a>	2288
RT @EmergMedDr: If you get bit by an insect , it stings and hurts but you don't need to seek help straight away in an A&E. Buy some antihi...	1634
...	...
RT @LollyDaskal: A caring attitude is one of the great leadership qualities of a great leader.~@LollyDaskal <a href="https://t.co/ppfSHzIgez">https://t.co/ppfSHzIgez</a> #Leade...	8
RT @LollyDaskal: You don't INSPIRE OTHERS by speaking about how amazing you are.~@LollyDaskal <a href="https://t.co/ppfSHzIgez">https://t.co/ppfSHzIgez</a> <a href="https://t.co/JFj35CfF76">https://t.co/JFj35CfF76</a> ARE....	8
RT @TEDxTCET: Even in these difficult times, a sister's love always shines. Here's to a quarantine Raksha Bandhan! #tedxtcet #tedxtcet2020...	8
A caring attitude is one of the great leadership qualities of a great leader.~@LollyDaskal <a href="https://t.co/ppfSHzIgez">https://t.co/ppfSHzIgez</a> <a href="https://t.co/JFj35CfF76">https://t.co/JFj35CfF76</a>	8
Get inspired for some family time! 🍷👉 What's your favourite #TEDtalk?~@LollyDaskal <a href="https://t.co/ObshPEhraS">https://t.co/ObshPEhraS</a>	8

# Conclusion and Next Steps

Tweets contents did appear to show a response to the news cycle, although this collection is incredibly small in scope.

Lots of opportunities for sentiment analysis, from Tweets, speech transcripts, and comments. What is being said about the talks? How many memes are present?

A prediction model could be trained on the attributes of the most popular tweets to predict future views and comments / audience engagement. TED could even promote past talks based on Google search trends and topics / keywords in their database of talks.

Can we anticipate the ebbs and flows of a certain talk (either new or aged), that will rise in popularity due to relevance, and predict the amount of time it will take to be discussed on Twitter? Will publish date and time affect the potential popularity of a talk?

May also be interesting to collect descriptive statistics on the speakers themselves: most common occupation, key terms in the About section, etc. What makes a speaker? Are there differences among TED and TEDx?