

# What predicts the popularity of TED Talks?

This dataset was developed using web scraping techniques, which extract data from websites. Web scraping is largely an automated solution; it is also an area of research that is rapidly growing. Data from web scraping is typically analyzed using text processing and artificial intelligence tools.

The data are from TED, a nonpartisan and nonprofit organization. TED spreads ideas, primarily via short talks that can be accessed on the internet. As noted on its website, TED was initiated in 1984 as a conference where technology, entertainment, and design ideas were shared. At present, TED Talks cover topics ranging from science to business to global issues. More information about TED can be found at the following website: <https://www.ted.com/>. Learning about the organization and its talks may be useful to develop your data analytic strategy.

This case study is currently a data competition on Kaggle (<https://www.kaggle.com/>). You may wish to check out what others have done with these data, although the analyses to date have been primarily descriptive in nature.

Your analysis in this case study will focus on the use of inferential techniques to analyze the data. As well, you should consider innovative approaches to measure popularity of the talks, beyond the conventional measure of the number of views of a talk.

## **Research Question**

The questions to consider when analyzing these data is: What characteristics of TED Talks predict their popularity?

## **Description of the Dataset:**

This dataset contains information about audio-video recordings of TED Talks uploaded to the official TED website. The data cover the period from 2006 to September 21st, 2017.

**Number of Records:** 2550

**Number of Columns:** 17

Column Name	Description
Comments	The number of first level comments made on the talk
Description	A description of what the talk is about
Duration	The duration of the talk in seconds
Event	The TED event where the talk took place
Film_date	The Unix timestamp of the filming
Languages	The number of languages in which the talk is available
Main_speaker	The first named speaker of the talk
Name	The official name of the TED Talk. Includes both the title and the speaker
Num_speaker	The number of speakers in the talk
Published_date	The Unix timestamp for the publication of the talk on TED.com
Ratings	A string dictionary of the ratings given to the talk (e.g., inspiring, fascinating, jaw dropping, etc.) and their frequency
Related_talks	A list of dictionaries of recommended talks to watch next
Speaker_occupation	The occupation of the main speaker
Tags	The themes associated with the talk
Title	The title of the talk
Url	The URL of the talk
Views	The number of views on the talk.