

TED is devoted to spreading powerful ideas on just about any topic. These datasets contain over 4,000 TED talks including transcripts in many languages.

Founded in 1984 by Richard Salmen as a nonprofit organization that aimed at bringing experts from the fields of Technology, Entertainment, and Design together, TED Conferences have gone on to become the Mecca of ideas from virtually all walks of life. As of 2015, TED and its sister TEDx chapters have published more than 2000 talks for free consumption by the masses and its speaker list boasts of the likes of Al Gore, Jimmy Wales, Shahrukh Khan, and Bill Gates.

The main objective is to build a predictive model, which could help in predicting the views of the videos uploaded on the TEDx website.

Attribute	Description	Data Type
talk_id	Talk identification number provided by TED	int
title	Title of the talk	string
speaker_1	First speaker in TED's speaker list	string
speakers	Speakers in the talk	dictionary

occupations	*Occupations of the speakers	dictionary
about_speakers	*Blurb about each speaker	dictionary
views(Dependent Variable)	Count of views	int
recorded_date	Date the talk was recorded	string
published_date	Date the talk was published to TED.com	string
event	Event or medium in which the talk was given	string
native_lang	Language the talk was given in	string

available_lang	All available languages (lang_code) for a talk	list
comments	Count of comments	int
duration	Duration in seconds	int
topics	Related tags or topics for the talk	list
related_talks	Related talks (key='talk_id', value='title')	dictionary
url	URL of the talk	string
description	Description of the talk	string

transcript	Full transcript of the talk	string
------------	-----------------------------	--------

*The dictionary key maps to the speaker in 'speakers'.