

## Daily Log

### Monday September 2

Holiday

### Tuesday September 3

Downloaded Two Corresponding CSV Files from Kaggle (rounakbanik/ted-data-analysis). Dataset contained around 2500 TED talks with transcripts, videos, and details for each. Both CSV files did not match so I had to delete extra entries in one of the files. Noticed that extra entries are ones that are in 0 languages and tried using Excel search function but the files still did not match. Had to undo this and started to manually delete.

### Thursday September 5

Finished up deleting extra entries from both files. Merging cells did not work so I had to delete a few data points (still have enough). Started looking into different libraries that convert audio to text. Many cost money so started to look at Github repositories to see if there were any good starting points. Found Potential Repository (<https://github.com/akras14/speech-to-text>)

## Timeline

Date	Goal	Met
Aug 22	N/A	N/A
Aug 29	Finalize Idea and Think of Good Starting Point for Project	Yes, Submitted Journal 0
Sep 5	Finish Formatting Sizeable Dataset	Yes, Dataset has 2461 Entries
Sep 12	Find Good Audio to Text API or Github Repository and Implement	
Sep 19	Finish Baseline Implementation of Audio to Text and Get Basic Results	

## Reflection

It was tedious to have to manually edit the CSV Files and look for discrepancies between the two. However, it began streamlined after I found the characteristic of the extra entries. The reason the Excel search function did not work was the shorter file also had a few formatting errors. I decided to delete those entries as merging cells was not keeping the whole transcript. I think finishing the dataset was a good starting point for the project and hope I can find a reliable audio to text API to build off of in the future. Below is a snap shot of the two CSV files side by side.

	A	B		A	B	C	D	E	F	G	H
1	transcript	url		1	comments	description	duration	event	film_date	languages	main_speaker name
2	Good mornin	https://ww		2	4553	Sir Ken Robin	1164	TED2006	1140825600	60	Ken Robinson Ken Robinson
3	Thank you so	https://ww		3	265	With the san	977	TED2006	1140825600	43	Al Gore Al Gore: Ave
4	(Music: "The	https://ww		4	124	New York Tir	1286	TED2006	1140739200	26	David Pogue David Pogue:
5	If you're here	https://ww		5	200	In an emotio	1116	TED2006	1140912000	35	Majora Carter Majora Carte
6	About 10 yea	https://ww		6	593	You've never	1190	TED2006	1140566400	48	Hans Rosling Hans Rosling
7	Thank you. I	https://ww		7	672	Tony Robbins	1305	TED2006	1138838400	36	Tony Robbins Tony Robbins
8	On Septemb	https://ww		8	919	When two yo	992	TED2006	1140739200	31	Julia Sweeney Julia Sweeney
9	I'm going to	https://ww		9	46	Architect Jos	1198	TED2006	1140652800	19	Joshua Prince Joshua Prince
10	It's wonderfu	https://ww		10	852	Philosopher I	1485	TED2006	1138838400	32	Dan Dennett Dan Dennett
11	I'm often ask	https://ww		11	900	Pastor Rick V	1262	TED2006	1140825600	31	Rick Warren Rick Warren:
12	I'm going to	https://ww		12	79	Accepting his	1414	TED2006	1140912000	27	Cameron Sinclair Cameron Sin
13	I can't help b	https://ww		13	55	Jehane Nouja	1538	TED2006	1140912000	20	Jehane Nouja Jehane Nouja
14	I'm the lucki	https://ww		14	71	Accepting th	1550	TED2006	1140652800	24	Larry Brilliant Larry Brilliant