

# Week 9 Progress Report

## DATA 450 Capstone

Andrew Nemkov

Week ending: 3/30/23

### Time Log

Day	Time	# Hours	Task(s)	GH Commit(s)
Mon.				
Tue.				
Wed.				
Thu.				
Fri.	[12:00pm-4:00pm & 12:00am-1:00am]	[5]	[finished part 1 and 2 for question 2]	<a href="#">Link 1</a>
Sat.				
Sun.	[4:00pm-7:00pm]	[3]	[finished part 3 of question 2 and started question 3]	<a href="#">Link 2</a>

**Total # of hours worked this week: 8**

### What you accomplished (or attempted) this week:

This week I completed all my visualizations for question 2, which related with analyzing word frequency and speech patterns. I was able to complete my graphics on the top 50 most and least frequent nouns in the three Baki series, ordinary and normalized distributions of each series' part of speech, and a visualization of the TF-IDF scores for each arc of the manga. In addition I wrote down all the interesting points from my visualizations for question 1 and 2 to help me in drafting out my poster and seeing what information I should put on it. Overall the results and my thinking back of my analysis was really interesting, allowing me to learn

and note several things that I previously didn't know (or have enough proof of) for the Baki manga.

**Any setbacks/roadblocks you experienced:**

I did just recently experience one setback in my coding regarding my final question. I was thinking of grouping my word\_US english words in my dataset into separate themes, and then perform analysis on this grouping. This task seems a lot harder than what I originally imaged, as there are many things and issues to consider. At the moment I am seeing if using NLTK and the process of LDA would help me properly group my words.

**What you plan to work on next:**

For next week I plan to continue chipping away at my final question for my project and put most of my effort refining my poster and the information I present on it. I will have to most likely schedule myself specific time in the week to do these things, as it seems my weekly amount of work keeps constantly increasing as the semester progresses.

**Are you on track? If not, how will you get back on track?**

Considering all my issues I am in fact on track for completion of my project. This is because I have most of the confusing parts done, as I really only have a single issue that needs addressing.