**NoSQL Project 1-Page Proposal**

1. **Team members (in alphabetical order of the last name):**

Alison Lee, Ji Hoon Park

1. **What problem you want to work on?**

We want to analyze public sentiment on COVID-19 topics across time and different geographical locations.

1. **Which NoSQL database you want to use?**

MongoDB

1. **What data will you use (if you use a public data set)? If you want to generate your own data briefly explain why it is the best way. For example, you can say there is no existing public data set (make sure it’s true though).**

We’ll use the [COVID-19 Twitter chatter dataset for scientific use](https://github.com/thepanacealab/covid19_twitter).

1. **Why is it an important/interesting problem?**

COVID-19 has transformed from a public health issue to a social and political issue. Therefore, getting an idea of how the general public has been reacting to COVID-19 news and events may provide a useful perspective on how to combat misinformation and enact positive change.

1. **Why is it significant (non-trivial)?**

We want to perform sentiment analysis on the tweets and relate that data to geographical locations. We’d also like to see how tweets reflect current events by analyzing the most frequent keywords in tweets change over time.

1. **Why is it doable in this semester? (Describe your plan, strength, familiarity with the language/tool of your choice, etc.)**

We plan to first gain familiarity with the tools necessary for this project as well as looking into similar projects and any resources that may be useful. We then plan on using the background knowledge that we gathered to come up with an appropriate design as well as a design roadmap to track milestones and stretch goals. Finally, we plan on delegating the work and implementing our parts individually while keeping each other updated throughout the project.

The tasks we would like to focus on:

1. Sentiment over time, graphed on a timeline. We could try to correlate major news dates (for example, when the first vaccine was administered) to the sentiment on that date.
2. A mapping of sentiment on a USA or world map.
3. What are the most popular topics for each sentiment category? Could do a word cloud or a bar chart. For example, the most popular topics for “Negative” could be “allergic (reaction).” This could be used to analyze causes of concern, happiness, etc.
4. Comparing geographic sentiment data with number of COVID-19 cases, and vaccination rates for the geographical location.