Proposal for the NoSQL Project

1. **Team member(s):**
   Dishant Bhatnagar, Sagar Naresh Sidhwa, Saloni Manoj Wandile

2. **Which NoSQL database do you want to use?**
   We would like to use MongoDB.

3. **Which public dataset will you use? Provide the link. Does it have non-tabular as well as tabular data?**
   Name of dataset-> Ok Cupid profiles dataset
   https://www.kaggle.com/datasets/yashsrivastava51213/okcupid-profiles-dataset
   Yes, this dataset contains tabular as well as non-tabular data.

4. **Concisely describe the N+1 (N = Number of students in a project group) nontrivial NoSQL queries you propose to implement using a bulleted list. (N =3)**

   a) Matching Preferences Exploration

   In the "Matching Preferences Exploration," we're diving into the OkCupid Matching preferences. Initially, we'll focus on structured attributes like age, height, and body type to identify patterns within different age groups and body types. Then, we'll turn our attention to unstructured attributes by analyzing the language in users' essays to understand how self-descriptions impact interactions. Essentially, we aim to uncover whether specific age groups, heights, or body types will attract distinct types of matches based on these attributes. This analysis provides valuable insights into the variation in dating preferences of OkCupid users, shedding light on how personal descriptions influence potential matches. For example, we'll explore whether users aged 19-35 with a specific body type compared to those aged 36-45 and how their language in essays and other features provides crucial insights into the dating profiles of OkCupid users.

   b) Profile Popularity Analysis

   In the "Profile Popularity Analysis," we delve into understanding the intricate factors that influence a user's popularity on the OkCupid dating platform. Our exploration encompasses both structured attributes, such as income, diet, drinks, and smoking habits, and the deeper insights provided by unstructured attributes, namely, users' essays. By examining structured attributes, we aim to discern whether variables like income level, dietary preferences, and smoking/drinking habits play a significant role in a user's attractiveness to potential matches. Additionally, we scrutinize unstructured attributes, particularly users' essays, to glean insights into their personalities, interests, and communication styles. Through this comprehensive analysis, we seek to uncover correlations and patterns that contribute to a user's popularity, shedding light on the nuanced dynamics at play in the online dating landscape. This understanding can inform OkCupid's recommendation algorithms and user experience enhancements, ultimately fostering more meaningful connections between users.

   c) Language and Communication Patterns

   In this task, we'll utilize data from the 'age', 'sex', 'speaks' and ‘essay’ columns to

determine which age groups and genders exhibit higher levels of multilingualism on our dating site platform. For instance, we may find that individuals between 25-34 years old and females tend to list more languages in their 'speaks' column. We'll employ aggregation frameworks or map reduce functions supported by MongoDB on the ‘essay0, essay7’ columns to uncover frequent words and common catchphrases within user profiles. As an example, we might discover that phrases like "love hiking" or "passionate about travel" are frequently mentioned in user descriptions, indicating common interests among members.

d) Analysis of Habits and Relationships

We will group users based on their relationship status, such as single, in a relationship, married, divorced, etc. Then, we will pair various lifestyle choices (e.g., diet, drinks, drugs, etc.) with each relationship status group. By analyzing the prevalence of these lifestyle choices within each group, we aim to understand how lifestyle choices impact relationship status. Additionally, we will assess the essay1 column to determine how users are describing their working habits and how it is affecting their relationship status.