Project 2 - Proposal

1. **Team Members:**
   a) Atharva Patil (B01014288)
   b) Shirin Syeda (B01022280)
   c) Shreyas Kadam (B01035884)

2. **Which NoSQL database do you want to use?**
   For this project, we will be using MongoDB as the NoSQL database.

3. **Which public dataset will you use? Provide the link. Does it have non tabular as well as tabular data?**
   For this project, we will be using the Airline Reviews dataset from Kaggle website.
   The URL for this dataset is:-
   https://www.kaggle.com/datasets/juhibhojani/airline-reviews
   This dataset contains a csv file containing data for airline reviews. The CSV file has 20 attributes and over 5000 samples.

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**
   1. Food & Beverages Feedback Analysis: **(newly added)**
      - Assess the impact of food and beverages' quality—as expressed in reviews—on passengers' perceptions of value for money.
      - This involves sentiment analysis of the reviews for mentions of food and beverages and correlating these sentiments with the structured rating for value for money
      - Here, we will be using structured data - Food & Beverages, Value for money and Unstructured data - Review

   2. Analysis of Sentiment Versus Value for Money:
      - Investigate the relationship between sentiment expressed in review and the 'valueForMoneyRating'.
      - Determine if higher satisfaction correlates with perceptions of value for money, or if expectations vary significantly across different price points.

   3. Sentiment by Route Popularity:
      - Categorize routes by popularity based on the number of reviews.
      - Analyze sentiments to determine if more popular routes have better or worse reviews, possibly indicating operational focus areas.

   4. Impact of Flight Delays on Sentiment:
      - Use text analysis on review to identify mentions of delays.
      - Compare sentiment of reviews mentioning delays versus those that do not to quantify the impact of punctuality on customer satisfaction.

Additional Analysis Task (incase the first one is rejected):

1. Relationship between Comfort of Seat and its Recommendation Chances:
   - Perform sentiment analysis on the review texts focusing on mentions of seat comfort and analyze its correlation with the likelihood of recommending the airline.
   - The task will help analyze if seat comfort significantly influences passenger's willingness to recommend the airline.
   - Here, we will be using Structured data — Seat Comfort, Recommended and Unstructured data — Review