1-Page Proposal for the NoSQL Project

1. List your team member(s):

- Tejas Dnyaneshwar Abhang
- Kaushal Vinay Nerkar
- Nikitha Reddy Birudala

2. Which NoSQL database do you want to use?

MongoDB

3. Which public dataset will you use? Provide the link. Does it have non-tabular as well as tabular data?

- **Dataset : TMDB 5000 Movie Dataset**
- https://www.kaggle.com/datasets/tmdb/tmdb-movie-metadata/data
- This dataset has 20 column and around 5000 rows
- Yes, This dataset has both tabular and 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.

1. **Impact of Director and Production Team on Revenue and Ratings**
   Find the influence of the director on movie revenue and user ratings. Focus on a specific director or set of directors to analyze the revenue distribution of their movies, and explore how user ratings vary based on their involvement.
   Attributes Used:
   a. revenue (Structured)
   b. crew (Unstructured)
   c. vote_average (Structured)
   d. vote_count (Structured)
   e. budget (Structured)

2. **Identifying trends in actor rating based on movie popularity, revenue, budget, user votes.**
   Generate actor ratings in relation to movie popularity, revenue, budget, and user votes, aiming to uncover trends that show the connection between actor performance and various aspects of a film's success
   Attributes Used:

a. budget (Structured)
b. vote_count (Structured)
c. vote_average (Structured)
d. revenue (Structured)
e. popularity (Structured)
f. cast (Unstructured)

3. **Explore how frequently certain cast and crew members collaborate across multiple movies.** Identify pairs of cast or crew members who have worked together most frequently, considering diverse production companies.
   Attributes Used:
   a. cast (Unstructured)
   b. crew (Unstructured)
   c. production_companies (Unstructured)

4. **Identify the most dominant genres in each decade based on the number of movies and analyze the evolution of genre popularity over time wrt production company**.
   Attributes Used:
   a. release_date (Structured)
   b. genres (Unstructured)
   c. popularity (Structured)

5. **Analyze the relationship between movie revenue, genres, and user ratings to identify trends in audience preferences.**
   Focus on a specific genre, explore the revenue distribution across movies within that genre, and correlate it with user ratings. Also, incorporate the movie overview to extract revenue and ratings related patterns.
   Attributes Used:
   a. revenue (Structured)
   b. genres (Unstructured)
   c. vote_average (Structured)
   d. overview (Unstructured)