

1-Page Proposal for the NoSQL Project

1. Team members in alphabetical order of the last name:
 - Patil Shubham (B00957264)
 - Shetty Disha (B00965377)
2. What problem you want to work on?
 - We are going to explore the **Stack Overflow Developer Survey 2022** which was conducted in May-2022 with over 70,000 responses from over 180 countries all around the world, to understand and get some insight of all the software developers experience from learning to code to their favorite technologies and the workplace experience of professional developers.
3. Which NoSQL database you want to use?
 - We will be using the **MongoDB** NoSQL database.
4. Which public dataset will you use? Provide the URL of the dataset. Does the dataset have at least 20 attributes and 10,000 samples?
 - We will be using the **Stack Overflow Developer Survey 2022** dataset, which is made yearly available by the Stack Overflow and Stack Exchange, Inc company on their official website.

Dataset URL: - <https://insights.stackoverflow.com/survey>

The dataset has almost 85 attributes and more than 70,000 responses or samples.
5. Describe the N+1 non-trivial data analysis tasks you plan to design and implement. Do they extract new and valuable information from the raw data in the dataset? Why are the tasks non-trivial?
 - The non-trivial data analysis tasks that we plan to design and implement on **Stack Overflow Developer Survey 2022** dataset are: -
 - To analyze what is the impact of company size (i.e., MNC's or Startups or medium scaled IT business) on the job satisfaction of developers/employees?
 - To analyze whether is there a correlation between the age of developers and the programming languages they are proficient in or interested in learning?
 - To analyze Are the methods of learning software development different across various countries?
 - The data analysis tasks described above will involve extraction of new and valuable information from the raw dataset by defining analytical coding for retrieval which will involve consideration of multiple attributes of the dataset and their responses. This analysis will also help us to identify the employees' satisfaction with working in various sized companies, their interest in technology with increasing age and technological trends across multiple countries.
 - Implementing the tasks are non-trivial because the data analysis tasks involve the complex questions which cannot be answered through simple statistical analysis meaning analyzing each question involve consideration of multiple attributes to reach to the conclusion with careful planning and execution hence making the tasks challenging.

6. Why is it doable in this semester? Describe your plan, strength, and familiarity with the language/tool of your choice.
- For this project we will be using the MongoDB database application and Python as a coding language which we have studied in our undergrad & have knowledge. Along with it, we have professional experience working on MongoDB as a database technology and developing applications using python language.
 - Considering these factors, the project seems to be doable this semester because of the familiarity in using technologies and applications.