

Please list out changes in the directions of your project if the final project is different from your original proposal (based on your stage 1 proposal submission).

We keep our overall goal of developing a sleeping monitor platform in our development but make some small modifications to details that we haven't considered in the proposal. For example, we decided to use a machine learning method to decide what sleeping disability the patient has, but in practice, we found it hard to achieve after combining two datasets and adjusting the length of two datasets to be the same with auto-generated data. So we modified it to use ChatGPT to find out whether the data of patients shows he/she has sleeping disabilities.

Discuss what you think your application achieved or failed to achieve regarding its usefulness.

We achieve the goal of generating a fully functional website with FrontEnd, API, SQL, and machine learning to collect user data and give diagnoses and suggestions. We finish the function of using the Machine Learning method to group variables, so users only need to fill in one of the different types of data in the same group. We also achieve the goal of recording the willingness of users to provide their contact information to help future users with similar problems. But we give up the function of using Machine Learning to find out the sleeping disorder of users because simple ML models are not accurate enough. Also, to achieve certain accuracy, the efficiency of the model can be slow, which will influence the user experience. So we give up simple models and use ChatGPT to generate the disorder type. Another one we give up is the visualization of user data. When running the ML model used to group variables, we found that the correlation between each variable was not strong enough to make the data visualization meaningful.

Discuss if you changed the schema or source of the data for your application.

Our group decided to keep our original sources of data and we have manipulated the grouping of these data to form data tables suitable for our application. We have modified the schema of our dataset to complement our application. Our changes include switching weak entities to strong entities to reduce the reliance on the 'User' table, which is the only strong entity in our previous design. This ensures easy data management and manipulation for our application. Another change we made was changing the data type of our data for the improvement of our application's performance.

Discuss what you change to your ER diagram and/or your table implementations. What are some differences between the original design and the final design? Why? What do you think is a more suitable design?

Some of the ER diagram and table implementation was changed while we implemented our application and database. Specifically, we have updated our relational schema between our tables

such that the tables other than the 'User' table were changed to being strong entities with UserId being their primary key. Initially, when we designed our ER diagram, all tables were assigned to be a weak entity other than the 'User' table, however, when considering the usefulness and effectiveness of our application structure, we decided to switch those tables to being strong entities. We have also modified some of our datatypes for the efficiency of memory utilization and dynamic data management. So many datatypes were changed to VARCHAR, STRING, and DECIMAL. Our group believes that the current design is a more suitable design because it closely aligns with the design of our application and serves its purpose.

Discuss what functionalities you added or removed. Why?

We add the function of using ChatGPT to decide the type of mental illness of users. And correspondingly, we remove the idea of using ml models to do so. We also remove the visualization of user data, since the correlation between data is not obvious and the graph may not be meaningful. We also add the register, login, and corresponding detecting of the format of username/whether it already exists to make the website more formal.

Explain how you think your advanced database programs complement your application.

Our advanced database programs include transactions, stored procedures, constraints, and triggers. We create transactions whenever the user submits their inputted health and body data. The transaction yields two results, one is the data was committed successfully, otherwise if an error occurs, the operation is aborted. Our database responds by either storing the data submitted or it leaves the database in its original state. The stored procedure is used to return relevant average data based on information from people of similar age to the user. This was used as some hints to assist users when they input their data. This allows us to create personalized custom experiences for our users who are using our application. We also implemented a trigger that deletes other tables before deleting the User table using the UserId. With this functionality, we can ensure that all associated data of the user is deleted ensuring user data security and consistency. The constraint is implemented to ensure the primary keys and foreign keys criteria are satisfied ensuring the user's unique identification. We believe that these advanced database programs complement our application because they improve our application's performance and allow customized and personalized user experience.

Each team member should describe one technical challenge that the team encountered. This should be sufficiently detailed such that another future team could use this as helpful advice if they were to start a similar project or where to maintain your project.

Hezi Jiang: The challenge I encountered was when working on the machine learning part of grouping related variables. Since our dataset has both numerical and categorical variables, it is

hard to find an algorithm to get comparable results for the relationship between numerical-categorical variables, numerical-numerical variables, and categorical-categorical variables. And there is no common standard for the correlation result to decide whether the two variables are related. So the standard I use may not be that accurate.

Selena Wang: We encountered a challenge during our project when the datatype we had planned to use didn't match the datatype of the data in the database. Because of this, we couldn't do the comparison and analysis we'd originally planned for the report. To fix it, we adjusted the datatype to match what the database used and tried more analysis methods.

Xiaoyu Li: Another challenge we faced was working with a cloud-based database. Initially, setting up the connection and ensuring seamless communication with the cloud database was difficult. We encountered issues like latency and unexpected configuration errors because of the lack of related knowledge. Then we tried to learn it by using online sources and finally solved it.

Feihao Fang: One of the challenges we faced was incorporating ChatGPT into our project. Initially, we had planned to include some machine learning to analyze the results, but we quickly realized that machine learning was much more difficult than expected. So, we decided to change our original idea to use ChatGPT instead to perform the comparison and analysis. This shift allowed us to overcome the difficulties and still achieve meaningful results in a more efficient way.

Are there other things that changed comparing the final application with the original proposal?

Initially, we had planned to include some machine learning methods to analyze the results, but we quickly realized that machine learning was much more difficult than expected. So, we decided to change our original idea to use ChatGPT instead to perform the comparison and analysis. This shift allowed us to overcome the difficulties and still achieve meaningful results in a more efficient way.

Describe future work that you think, other than the interface, that the application can improve on.

In the future, the application can improve on several aspects related to database management and data analysis. One is expanding the database to include a more comprehensive dataset. By integrating data from a larger and more diverse population, the application can provide more accurate and personalized health reports. Another improvement could focus on optimizing data analysis algorithms. Currently, the application processes inputs based on a limited set of parameters, but incorporating advanced analytical models, such as predictive analytics, could enhance the accuracy of results provided in the reports.

Describe the final division of labor and how well you managed teamwork.

Hezi Jiang mainly works on UI interaction and helps with the building of SQL. She also worked on finding, constructing, and combining the two datasets. *Selena Wang* built the database on GCP based on the schema designed previously with a few reasonable changes applied to adjust to our implementation and also helped with constructing the SQL script. She also worked on manipulating and processing the data. *Feihao Fang* worked on the backend implementation of the program including data and integrated our database to handle user data and requests. He also thoroughly tested the program's functionality. *Xiaoyu Li* contributed to the front-end design and implementation, ensuring a seamless and user-friendly interface by integrating the backend API with intuitive UI components.

All of our group members equally contributed to the implementation of advanced database program features and the final report.