CS 411 Final Project Report

1. **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).**
    a. Our overall direction for the final project stayed very similar to the initial direction of the project. Our initial goal was to create a more personalized way for students to rank universities and find the perfect university for themselves. We fulfilled this goal in our final implementation, and although we had numerous changes along the project, the overall direction and goal of the project remained the same.
2. **Discuss what you think your application achieved or failed to achieve regarding its usefulness.**
    a. Our application went past just fulfilling our goal. In order to make our project useful, we wanted to allow users to rank universities in a personalized manner. We initially implemented this by allowing users to weigh certain factors and create a linear sum of the best universities based on the weightings. To ensure our application goes beyond this, we created a machine learning function. This function uses cosine-similarity to find universities that have a similar vector to the preferences of the user. This allows the user to find universities that are not just rated high but instead actually find universities that have the same focuses as our users.
3. **Discuss if you changed the schema or source of the data for your application**
    a. We changed the schema of our application slightly. In order to keep track of the universities our users enjoy, we added a popularity column to universities. Additionally, we added triggers, stored procedures, and transactions that enhanced our application. Lastly, we also changed the source of our data. We generated synthetic data for our tables in order to meet the size requirements of the tables. We also found that the Kaggle dataset had many missing values, and was not sufficient to be used as a dataset.
4. **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?**
    a. As mentioned before, we added the popularity column to the Universities table in order to track what universities are the most popular amongst our users. Also, instead of tracking the rankings for users, we implemented real-time rankings based on personalization factors for users. This is more suitable as it still holds the functionality of providing personalized rankings, but it is more efficient as we have to store less data in our table.
5. **Discuss what functionalities you added or removed. Why?**

a. We decided to add functionalities to add, edit, and search for universities aside from just ranking universities based on alumni, research, popularity, and quality statistics in order to enhance the customization process for users. This allows our application to stay up to date, and allows the user to change the stats of universities if they wish. This allows users to create their personalized rankings and edit them according to specific preferences that only apply to one or two schools. By giving users more flexibility with these rankings, we allow them to have more freedom and more accurately represent their desires when searching for universities. To go even further, we added a machine learning algorithm that uses cosine similarity to find universities that align with the users preferences. This will help users find universities that are not the top universities but still match the user's preferences.

6. **Explain how you think your advanced database programs complement your application.**
    a. We have 2 stored procedures that we implemented. The first one is AddorUpdateFavorite. This procedure allows users to favorite a university, and then it updates the popularity score of that university. This implements the base structure of our application, allowing users to find universities they like and favorite them. Also, it allowed users to see what universities other users like. The other program is the GetTop5Universities procedure. This procedure allows users to set custom weightings for alumni, quality, and research, in order to find the best universities in a personalized way for users. These 2 programs help complete an application that allows users to find personalized universities that are good for them, and keep track of them.
7. **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.**
    a. Anika - A technical challenge that the team encountered was deciding what tables we needed in order to accurately and efficiently create rankings and favorites. The main aspect between these two features are the University table data and the Favorites table data, and we needed a way to connect the two and also save the rankings we would create based on the university data. We decided to create a separate table, "Choices", to save the rankings that occurred based on the weights of quality, research, and alumni fields.
    b. Noah - One technical challenge that the team encountered during the development of this project was allowing for indexing of advanced queries in Google Cloud Platform. We had not had prior experience with GCP, and therefore were unfamiliar with how to set up indexing and how to measure the performance changes that the indexing caused. One issue we faced was deciding which

columns to apply the indexing to. The instructions indicated to use indexing on variables involved with joins or other main clauses. While we were testing which variables had the most impact, we ran into trouble in removing indexing from variables that had indexing applied. This slowed down our development process, but we were able to get around this roadblock by looking at tutorials.

c. Pulkit - A technical challenge that the team encountered while developing this project was using React.js. Prior to this, I had experience with developing websites but had never specifically used React in development. Understanding the React development environment, including Node.js, npm, and package.json, was key in building my experience with the library. I also watch tutorials on the key aspects of React and how applications are typically built in the library. This allowed me to streamline the development process and build a reactive frontend to connect to the backend.
d. Mahathi - One technical challenge that the team encountered while developing this project was getting familiar with Node.js, specifically while implementing the server.js file. This is where we defined our core backend logic for the application and handled the requests that came in from our frontend. Setting up the file had multiple steps, like connecting to the database and handling different kinds of requests like GET, POST, etc. We overcame this challenge by carefully planning how our backend logic would work and interact with the frontend. We then watched tutorials in order to understand server.js files are typically set up, and how we could adapt those for our specific use case.

8. **Are there other things that changed comparing the final application with the original proposal?**
    a. Although our functionality is the same on our final application, we changed the logic slightly. For example, we offered different ways for users to get personalized rankings by allowing real time rankings based on weighted preferences. Furthermore, we added the machine learning program to take the personalization even further and create an application geared towards users.
9. **Describe future work that you think, other than the interface, that the application can improve on**
    a. The application has some areas of improvement that could be implemented with further development. First, we used synthetic data to populate our tables. By searching for datasets that meet the scope of our project, we could make our final project more applicable to the real world by providing the user with accurate data. We could also find data that has more features for universities to make our banking based on a larger variety of factors. Furthermore, we could create a login system o our application is more protected and users can only see their favorites on their portal.
10. **Describe the final division of labor and how well you managed teamwork.**

a. The final division of labor for our group was made up of group meetings at each project stage. We met virtually over Zoom or in-person at the CIF in order to plan, design, and work through the requirements of each stage. During the planning meetings, each team member was assigned a portion of the work to handle. We would work on them either at the CIF, or virtually in order to get help from each other. By planning and assigning tasks, we were able to ensure that each group member did their part in creating the final product. We also made sure to continually check in with each other to offer support to ensure a smooth and efficient workflow.