# Client Project Retrospective

Our primary objective for this project is to build a web-scraping program that can download separate stories from the health care review website (https://www.careopinion.org.au) and save various relevant sections of each page into separate files. As mentioned before, Care Opinion Australia is an independent site where individuals can share their experiences regarding the Australian health-care system. By using web-scraping our client will be able to study how Australians feel and think about their health-care experiences and perform further analysis on this data.

Currently, with the help of the beautifulsoup package from the python library, our team has already demonstrated to the client the functionality of our web-scraping program to save different relevant sections of each page layout; these sections include: story id and story, username, title, about (location), good tag, improved tag, feel tag, etc. Subsequently, as an extra feature, we have created a database to store the downloaded information in a manner that it can be easily accessible using a unique ID. Our database has been designed so that it can be filtered using specific tags that have been scraped or keywords. The database we have created also incorporates an easy to use graphical user interface that connects back with the database using the tkinter python package.

With regular organised meetings with our client and providing continuous status updates of our progress throughout the project, we have maintained a positive working relationship with the client. Our client has conveyed his contentment regarding the overall progress of the project as it stands, specifically being pleased by the numerous files of information created via the web-scraping program.

Through the course of this project, we were also made aware of a duplication error concerning the story id, response id and update id; for instance, each story on the website may contain responses from the health care providers or updates from the original poster, because the response and updates are under the story, they will subsequently have the same story id. Thereby, our web-scraping program will find and scrape the same page multiple times separating them into multiple files. Our group has resolved this issue by implementing a filtering system in our program to avoid the duplication of these stories. Our client has responded positively to these changes.

As mentioned earlier our group has also created a graphical user interface as another extra add-on feature for the client, this will allow for newly published stories to be scraped and imported into the existing database via the input of a url and allow the client to interact with the database easily.