UCSD Wellness Study

Problem:

Using data to uncover trends in UCSD student's collective thoughts.

Dataset:

The data will come from UCSD Confessions, a very popular Facebook group amongst students. Students use a form to anonymously submit any text "confession". The admins of the group release a public Google Spreadsheet with content, timestamps, and some other metadata of all confessions. The module 'gspread' to pull this data from the cloud in Python. For efficiency, it could be first saved to a local database on the local machine and analyzed from there.

Proposed Solution and Real world Application :

Our proposed solution involves using tools such as simple word count, sentiment analysis, and semantic modelling etc. on anonymous text submissions to uncover the collective mental state of the UCSD student body. The solution can also be used to extract and visualise trends in student mental, emotional, and academic well-being throughout the school year.

The content that students submit to UCSD Confessions varies widely: from angry rants, to legitimate complaints about the university, to very explicit and intimate personal experiences, to restaurant reviews. By analyzing the trends in this data, with respect to time and cross checking with UCSD academic calendar, we can uncover increases in certain topics or emotions at various points throughout the quarter.

Analysis of this seemingly useless data could potentially have huge real world applications. It could be harnessed by UCSD wellness departments and student organizations alike to provide more relevant, timely services to the student body. For example, we could identify the weeks throughout the quarter system where students are the most stressed. This would allow more mental health resources and events to be allocated to these times, to more precisely respond to demand.

Ultimately, this system would help universities implement data-driven wellness programs, and create a positive feedback loop that will incentivize deeper participation in a collaborative message board like UCSD confessions.

Project steps

Step	Estimated completion time	Person(s) in charge (among the group of 3)
1. Extracting and cleaning up data	One week	Curtis Lee
2. Semantic Modelling	Two weeks	Tyler Farnan
3. Data visualization (to obtain data statistics and user behavior)	One week	Meihan Zheng