**Project ETL**

**Additional Notes:**

**Yelp API – data issue**

We ran into an issue with getting data at the end.

First, the Yelp API returned 50 restaurants at a time, and we decided to execute additional loops to get the rest of the data. We used the API offset parameter. But as it turns out, the returned data was a repeat of the initial data.

The returned data from the API was over 2000 restaurants, but when we dropped the duplicates, we only had the original 50.

We should have thought about an alternative way of getting the data using different latitudes and longitudes.

Perhaps one way was reading a csv file with city, state, latitude and longitude. If we had cities from 10 states, then we would end up with approximately 5000 restaurants. The code would have a for loop for reading a file and doing an API call for each coordinate.

That way we would be processing more data and stored it on MongoDB.

**Google API – data issue**

We also ran into an issue with getting data from Google.

The Google API only returned 60 restaurants at a time doing a search with coordinates. We used the same coordinates for Yelp and Google, thinking that most of the restaurants would match, but it turned out that very few were the same.

It turns out that if we had done an API call of each restaurant individually, we would end up with more reviews.