**Group 1 ETL Project Report Q&A**

* What data sources we chose, and why:

For the ETL project our group decided to combine average auto insurance with weather day. The first step was to collect average auto insurance data. Six tables, which included average auto insurance rates, were scraped from CarInsurance.com and AutoInsurance.org. After importing the tables as dataframes they were combined into a master table. In order to be utilized, the dataframe needed to be cleaned of duplicate and missing data. For example, some of the rows did not include ‘state’ data. To fill in the missing data, two additional sources were used:

* + - Zip-Codes API (<https://api.zip-codes.com>)
    - State Abbreviations Code Table (<http://app02.clerk.org/menu/ccis/Help/CCIS%20Codes/state_codes.html>).

The Zip-Codes API was used to find the state abbreviation for the zip codes that were missing a state. The State Abbreviations Code Table was used to change the state abbreviations to the full state name. The next step was to extract data for each zip code from Open Weather Map (<https://openweathermap.org/>).

* Details of the extraction, transformation, and loading steps:
* Explanation of the types of transformations performed, and why:
* Why you chose the type of final database

As a group we felt much more comfortable with what has been learned to date, the types of exercises, and language syntax regarding PostgreSQL over Mongo. It also makes sense in light of our data being quite structured and numerical. If the data was more unstructured and fractured, a non-SQL database would’ve made more sense.

* Schema of the tables/collections in the final database
* Flask API:
* Hypothetical use cases for our database:

This data could be used by insurance companies and/or customers to determine if weather affects insurance rates and evaluate their business strategy. For example, an insurance company may use the average auto insurance rate to determine if they are charging too much or not enough in areas with a lot of rain. If their rates are too high, this may explain why their business is low in these areas.