# ETL Project

For the ETL project, I decided to use my own Google data in an attempt to demonstrate how much data Google has about my life and habits.

To obtain my Google data, I used my longest held and most active google account for my email address [Victoria.a.tonsall@gmail.com](mailto:Victoria.a.tonsall@gmail.com). On myaccount.google.com, I was able to request my personal data. Unsure of what to expect, I chose to receive my location history, search history, image search history and a variety of related data. Once I responded to the permissions request, I received an option to download the information or add them to Google Drive, Dropbox, etc. I opted to download the data and store it locally on my computer.

The collection houses data about my location and search history back to 2013. In an effort to narrow down a more specific range of information, I had hoped to use data from August 2015 through December 2016 – a time period representative of my pre-engagement, engagement, wedding planning and immediate post-nuptial life. Specifically, how my location and search activity changed during my engagement and wedding planning period, versus before and after.

My first challenge was determining how to extract the location history data and transform into useful information. The location data arrived to me in json format, while search history and image search history are in html. Using Jupyter Lab, I attempted to read in the data using the following code:

import pandas as pd

import numpy as np

Import json

json\_file = "Takeout/Location History/Location History.json"

location\_history\_df = pd.read\_json(json\_file, orient='index')

location\_history\_df.head()

This attempt returned one row of data with over two-million columns.

I made another attempt to use the “Copy” function in PostgreSQL, but the code proved to be too advanced for my skillset.

For my most recent attempt, I used the Python, in Jupyter Lab, to read in the file using “with open”. This was more successful than other attempts, but due to the size of the file, I received an error when requesting to print/display the data.

I do plan to use this data for my final project; thus, I plan to keep working to figure out how to parse it and write code to execute my desired goal.