**Data Analyst Job Search**

**Collaborators**

* Megan Moroney
* Nakouma Sako
* Wen Shon
* Matt Moormeier

**Summary**

Our ETL project is geared toward our fellow students who will be seeking employement as data analyst upon completion of our Data Visualization Bootcamp. We located a csv file on Kaggle of data analyst jobs on Glassdoor.com. We determined the top 50 cites with number of data analyst job postings to limit our job search. Then we web scrapped Zillow.com to calculate average apartment pricing for each of our 50 cities. The last dataset we brought in was current weather data for each city. Our goal is to provide quick information about the city that the most data analyst jobs are being offered.

**System Requirements**

* Chrome Web Browser
* Chromedriver
* Python environment running Python 3.7 with the following installations:
  + beautifulsoup4
  + numpy
  + pandas
  + pymongo
  + requests
  + splinter

**Data Sources**

We gathered data analyst job listing from Kaggle.com [HERE](https://www.kaggle.com/andrewmvd/data-analyst-jobs) . To determine information about the cost of living we scraped [Zillow.com](https://www.zillow.com/) for apartments in each of our cities and calculated summary information. Finally we utilized the [OpenWeatherMap.org](https://openweathermap.org/) api to get a snapshot of weather in each city.

**Transformation Steps**

We created three separate jupyter notebooks to develop our python code; GlassDoor, City, Zillow and Weather.

In the the GlassDoor notebook we:

* Import the DataAnalyst.csv file and drop unneccsary columns.
* Sort the dataframe by cities with the most jobs available and trim it to 50.
* Filter our analyst job dataframe and export to csv
* Create city dataframe and csv to export

In the Zillow notebook we:

In the Weather notebook we:

In Postgres we: