Project 2 Group 2

Matt, Meagan,, Wenbo, Nakouma

**Database Purpose**

Geared toward students completing this Data Analytics course and looking for a job. Starting with a csv file from Kaggle on Data Analytics jobs as our anchor. Then adding data focused on the city that each job is located in to aid in deciding what cities to rule out when applying for a job. Scraping data from Zillow regarding home prices (avg, high, low), weather data (temp, precipitation) and population. Store this data in a MongoDB. See sample spreadsheet of database – last columns.

**Data Sources**

1. **Data Analysis Jobs.csv**

<https://www.kaggle.com/andrewmvd/data-analyst-jobs>

Possible Webscrape of Indeed

<https://medium.com/@msalmon00/web-scraping-job-postings-from-indeed-96bd588dcb4b>

<https://medium.com/@msalmon00/web-scraping-job-postings-from-indeed-96bd588dcb4b>

1. **Zillow.com** – Webscrape

<https://www.scrapehero.com/how-to-scrape-real-estate-listings-on-zillow-com-using-python-and-lxml/>

<https://stackoverflow.com/questions/46623658/whats-the-best-way-to-scrape-data-from-zillow>

1. **OpenWeather API**

<https://openweathermap.org/api>

1. **Census API**

<https://www.census.gov/data/developers.html>

**Database Structure**

Data\_Analyst\_Job Table

Housing Table

Weather Table

Population Table

All connected by City\_id

**Final Database**

Mongo Database

**Plan of Action**

DataAnalyst.csv (2253 records)

Import data into Dataframe

Clean data – remove null

Split Location cell City,State to separate cells – City-level is limit

Drop columns that are not needed (Easy Apply)

Census API

Bring in population data at city and state level

Create Summary table at City Level

Add to “DataAnalyst.csv” dataframe

Scrape Data from Zillow

<https://www.scrapehero.com/how-to-scrape-real-estate-listings-on-zillow-com-using-python-and-lxml/>

<https://stackoverflow.com/questions/46623658/whats-the-best-way-to-scrape-data-from-zillow>

<https://www.zillow.com/howto/api/APIOverview.htm>

Create summary tables at the City Level

Number of listings, Average Price, ect…

Add to “DataAnalyst.csv” dataframe

Maybe calculations like Salary to Avg. Home Price?

OpenWeather API

Get High, Low, AVG Temperatures, Preceipitation, Cloudiness, Wind by City

Indeed.com – Webscrape?

Assemble Database and export to MongoDB