**Data Cleanup & Exploration**

* Describe the exploration and cleanup process
* Discuss insights you had while exploring the data that you didn't anticipate
* Discuss any problems that arose after exploring the data, and how you resolved them
* Present and discuss interesting figures developed during exploration, ideally with the help of Jupyter Notebook

Population: CENSUS csv

* Our population data were retrieved from the 2019 census dataset csv.
* Rural City are random cities with a population of less than 50,000.
* While, Urban City are random cities with a population more than 50,000.
* **Any city names(“NAME”) with a state name(“STNAME) result (i.e. Alabama) from both city results, city name was dropped from the column with a loc function.**
* **Only incorporated locations with population over 1,000 people.**
* **We eliminated all summed city/town data.**
* **Removed counties.**
* 100 Rural and Urban City samples were then drawn from these dataframes.

Restaurants: Yelp API data

* Our results for vegan restaurants were drawn from the Yelp API businesses search.
* Dictionaries collected from the API were: cities searched, city, state, latitude, longitude, restaurant name, category (i.e. pizza, vegan, seafood, donuts), and transactions (i.e. pick-up or delivery).

**Problem #1**

* API response: first Yelp API pull, our results were limited to 20 – 40 results

Resolution: API response: Parameters for the request were edited.

‘location’ : f “{city}, {state}” -- ADDED AS A STRING

‘radius’ : 8047 -- INCREASED FROM 4000 (is this correct?)

‘limit’ : 50 -- ADDED LIMIT PARAMETER AS SUGGESTED FROM API DOCUMENTATION

‘offset’ : offset[i] – ADDED PARAMETER AND SET ARANGE FUNCTION

**Problem #2**

* Limited to 5000 API calls per day

Resolution: Divided pull requests between available API keys, and saved copies of csv files for accessibility and further data retrieval and clean-up.

**Problem #3**

* Not all sub-categories were pulled for businesses

Resolution: Created a separate sub-category loop to collect all values and extracted results into different category columns.

**Problem #4**

* Duplicated results for our final urban and clean population dataframes.

Resolution: Census csv was further cleaned to eliminate duplicated population results from city results. Selective manipulation of pertinent population data was also applied.

DATA INSIGHTS:

* Population data from census.gov provided us a reliable source of current total population in U.S. cities.
* Summary distribution of vegan restaurants across the U.S. is significant in coastal states; however, still at a low percentage in comparison to other restaurant options.
* Vegan restaurants are only in existence at 0.49% within our gathered average number of urban restaurants of 558. Per capita urban vegan restaurants are at .000035.
* While, vegan restaurants are at 0.34% within the average number of rural restaurants of 155. Per capita rural vegan restaurants are at .000242.
* For our urban city summary, there is a weak correlation between population and vegan options.
* For our rural city summary, there is no correlation between population and vegan options.

INTERESTING FIGURES:

* Shown in vegan restaurant heatmap distribution
* Top 5 urban vegan cities
* Top 5 rural vegan cities