Introduction/Business Problem: *Discussion of the business problem and who would be interested in this project.*

The client currently has a new business idea to open an innovative bakery format in the greater Chicagoland area. The bakery concept is to be named the Scoli Dough Co… utilizing an 'all things flour' approach that would specialize in a highly focussed selection of quality Mexican style flour tortillas. Italian and American breads, pizza doughs as well as traditional donuts,. The primary focus is to become the value go-to, servicing business customers including; restaurants and fast-casual eateries. The client is also possibly looking to devote a small section of the facility to a small menu of items for dine-in and take-out, at extremely competitive prices.

The client already has a good understanding of the greater Chicagoland region and a short list of a few areas that they would be interested in placing their bakery concept. These locations fall within five Chicagoland counties (excluding all zip codes within the City of Chicago.) They are ultimately looking for a fresh set of unbiased and objective recommendations of the best possible Villages to house their business that align with a target market of a high density of Mexican, Italian, and American Restaurants - that are also currently not within close proximity to a high number of existing bakeries.

Data Section: *Description of the data that will be used to solve the problem and the source of the data.*

Chicagoland consists of many Counties that house numerous Villages/Cities. These Village/Cities are assigned to either one or multiple US Postal Codes. The first step of data collection was to locate the site that housed all of the Counties that exist within Illinois including the classification for the Village name and Zip Code. The exploration is limited to the following counties; McHenry, Lake, Kane, DuPage, and Cook (excluding the zip codes that are assigned to the City of Chicago.) This data was located on the

internet and then extracted to a .csv file.

Site for the County Data:

https://www.zipcodestogo.com/Illinois/

- Zip
- Village/City
- County
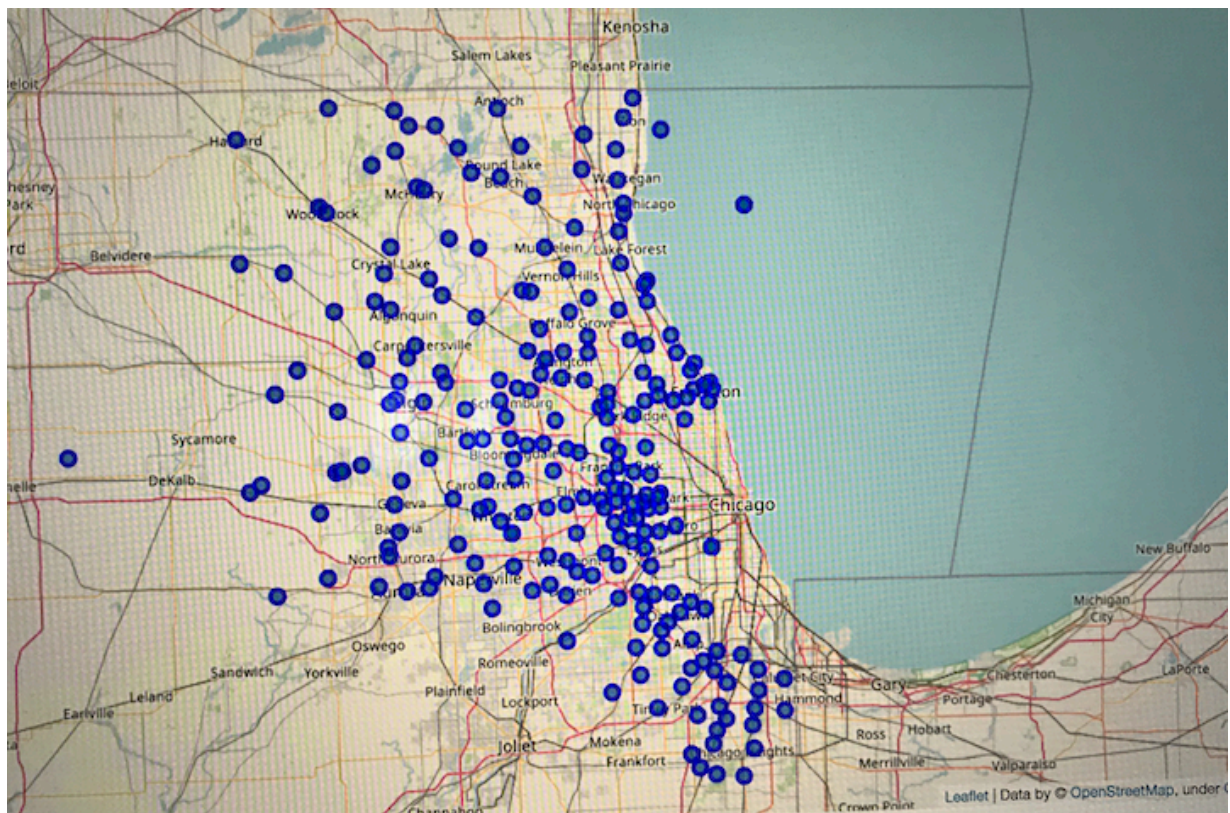
The second step in the process is to locate an additional dataset that would include the latitude and longitude of each specific zip code within the United States. The latitude and longitude given for the Postal Code is represented as the geographic centroid of the ZIP code, where the location given can generally be expected to lie somewhere within the ZIP code's "boundaries". The data for this portion of the project's requirement is available at the following URL's, with the following fields (and was also extracted as

a .csv file):

- Zip
- Village/City
- State
- Latitude
- Longitude
- Timezone
- Daylight savings time flag
- Geopoint

With these two datasets now available, a Pandas Data frame was created by merging these two sources of data on the zip code field and creating a master Pandas DataFrame that contains all of the zip codes within this region with a description of the Village/City along with the classification of the County. This DataFrame of the Villages/ Zip Codes and their Latitude and Longitudes was then matched up with the rich location data of the Foursquare API to enable the exploration of the Villages within all seven Counties that are in focus within the Chicagoland area.

Methodology: *The main component of the report where you discuss and describe any exploratory data analysis that you did, any inferential statistical testing that you performed, if any, and what machine learnings were used and why.*

The primary purpose of this analysis is the find the Villages within the Chicagoland are

that have a high density of the following Venue Categories, due to the fact that this would align with the target market of the clients products:

FourSquare API Venue Categories of Opportunity, ranked by most desirable:
1. Mexican Restaurant. Taco Place, Burrito Place
2. Italian Restaurant, Pizza Place
3. Sandwich Place
4. American Restaurant, Steakhouse, BBQ Joint, Burger Joint, Hot Dog Joint

Ultimately taking into consideration the Venue Categories that are classified as a "Bakery" to determine if there are Villages that coincide with the target market that aren't being serviced by the competition.

A thorough analysis of each village was also done utilizing the K-means clustering algorithm (utilizing five clusters) as well as the Folium library for visualization to determine the emerging clusters and ultimately provide worthwhile suggestions for the new business concept to potentially locate.

Out[57]:

	Neighborhood	Timezone	Daylight savings time flag	geopoint	County	...	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Alsip	-6	1	41.675861,-87.73189	Cook	...	Grocery Store	Park	Pub	Donut Shop	Coffee Shop
9	Blue Island	-6	1	41.656592,-87.68154	Cook	...	Pizza Place	American Restaurant	Italian Restaurant	Supermarket	Breakfast Spot
10	Bridgeview	-6	1	41.742432,-87.80678	Cook	...	Mexican Restaurant	Ice Cream Shop	Sandwich Place	Breakfast Spot	Convenience Store
14	Calumet City	-6	1	41.614188,-87.54638	Cook	...	Grocery Store	American Restaurant	Pizza Place	Seafood Restaurant	Restaurant
19	Country Club Hills	-6	1	41.561134,-87.72398	Cook	...	BBQ Joint	Mexican Restaurant	Breakfast Spot	Pizza Place	Donut Shop
24	Dolton	-6	1	41.626839,-87.59865	Cook	...	Food	Pharmacy	Discount Store	Donut Shop	Caribbean Restaurant
41	Glenwood	-6	1	41.544584,-87.61289	Cook	...	Ice Cream Shop	Pharmacy	American Restaurant	Discount Store	Fried Chicken Joint

In [53]: # Cluster 2

```
chicagoland_merged.loc[chicagoland_merged['Cluster Labels'] == 1, chicagoland_merged.columns[[1] + list(range(5, chicag
```

Results:

199 Unique Village/Cities exist within the five counties of the greater Chicagoland area (excluding the City of Chicago.). This is represented by 273 Unique Zip Code/Village combinations.

The specifics that were returned by the full Four Square API Call:

- A maximum of 100 locations per Village/Zip Code combination were specified.
- 5km radius for each Village/Zip Code combination.
- Based off of these two criteria, 23991 locations were returned by the FourSquare API (these location are not necessarily unique, as there could have been overlap within the concurrent radius circles.)
- 371 unique Venue Categories were returned by the FourSquare API.

All of the 199 unique villages were in scope within this analysis, and the following table represents the Villages that scored highest based off of the ranking system:

Neighborhood	Composite Score	Bakery Exist?	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
Elk Grove Village	92	NONE	Mexican Restaurant	Coffee Shop	Italian Restaurant
Melrose Park	89	NONE	Mexican Restaurant	Pizza Place	Italian Restaurant
Stone Park	86	NONE	Mexican Restaurant	Pizza Place	American Restaurant
Bedford Park	82	NONE	Mexican Restaurant	Taco Place	Coffee Shop
Glenview Nas	82	NONE	Mexican Restaurant	Taco Place	Coffee Shop
Oak Lawn	82	NONE	Mexican Restaurant	Pizza Place	Italian Restaurant
Palos Heights	79	NONE	Sandwich Place	Italian Restaurant	Mexican Restaurant
Palatine	78	NONE	Mexican Restaurant	Pizza Place	Italian Restaurant

Glendale Heights	76	NONE	Pizza Place	Fast Food Restaurant	Mexican Restaurant
Northbrook	75	NONE	Mexican Restaurant	Grocery Store	Coffee Shop
Saint Charles	75	NONE	Italian Restaurant	Mexican Restaurant	Pizza Place
Addison	73	9	Mexican Restaurant	Coffee Shop	Pizza Place
Lisle	73	NONE	Pizza Place	Mexican Restaurant	Italian Restaurant
Wheeling	72	10	Pizza Place	Italian Restaurant	Mexican Restaurant
Des Plaines	71	NONE	Mexican Restaurant	Breakfast Spot	Sandwich Place
Oak Forest	71	NONE	Sandwich Place	Pizza Place	Fast Food Restaurant
Itasca	70	NONE	Breakfast Spot	Mexican Restaurant	Italian Restaurant
Bellwood	69	NONE	Pizza Place	Mexican Restaurant	Hot Dog Joint
Midlothian	69	NONE	Sandwich Place	Mexican Restaurant	American Restaurant
Villa Park	69	NONE	Pizza Place	American Restaurant	Mexican Restaurant
Waukegan	68	NONE	Pizza Place	Mexican Restaurant	Sandwich Place
Wood Dale	68	NONE	Sandwich Place	Mexican Restaurant	Pizza Place
Bloomingdale	67	NONE	Italian Restaurant	Pizza Place	Mexican Restaurant
Alden	66	NONE	Mexican Restaurant	Fast Food Restaurant	Sandwich Place
Medinah	66	9	Italian Restaurant	Pizza Place	Mexican Restaurant

The ranking system took into account the 10 Most Common Venues per village and assigned a score based off of the Most Common Venue being classified as one of the desirable businesses. Where the most desirable type of business (number 1 from the

Methodology section above) was given a higher score per occurrence. The composite score then represents the total score of strength, that the Venue Categories that exist within a village align with the clients focussed business plan - and also wether a Venue Category of bakery was also ranked within the top 10.

Discussion: *Discussion of any observations you noted and any recommendations you can make based on the results.*

There are a significant amount of Villages within the greater Chicagoland area that have a high density of target Venue Categories including Mexican, Italian and American Restaurants. Currently, these Villages do not necessarily house many local bakeries to service their needs. Based off of these observations, there exists an opportunity for a viable business concepts to fill that need. I would suggest that the next step would be to utilize this focussed list of locations to pull in some additional demographics. This could include median income, price of real estate per sq. ft., local taxes, etc…

Conclusion:

The FourSquare API houses a very rich database of useful location data that is very accessible to call on for many data exploratory and analysis needs. Pairing this resource with the Anaconda Jupiter Notebook shell (via Python 3) there is a significant amount of data collection, modeling and visualization that can be achieved to gain a full understanding of a business problem. Utilizing the Greater Chicagoland Bakery Exploration as a sample approach, I am much better positioned to provide worthwhile suggestions for future Data Science projects.