Data Section

Chicagoland primarily consists of Villages/Cities which are assigned to either one or multiple US Postal Codes. With essentially the entire Chicagoland are in scope for this analysis, the first step in the process would be to create a Pandas Data frame that contains all of the zip codes within this region with a description of the Village/City. This dataset would also include the latitude and longitude of each specific zip code. 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, with the following fields:

https://public.opendatasoft.com/explore/dataset/us-zip-code-latitude-and-longitude/table/

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

This data frame would then be matched up with the rich location data of the Foursquare API to explore the Villages within the Chicagoland area to find the clusters that have a high density of Italian and Mexican restaurants, schools and church organizations. The income demographic would also be taken into consideration to refine the suggested target market(s). A thorough analysis of each village will be done utilizing the K-means clustering algorithm 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.