**Location of John’s Ice Cream and Pizza Shop**

**2. Data**

In order to identify the towns that meet John’s criteria for his shop location (i.e., the three largest towns in McLean County, Illinois that have neither an ice cream shop, nor a pizza restaurant), two main types of data will be used: Foursquare data (for venues information) and data from Wikipedia (for town names, populations, and locations). They are explained in further detail as follows.

**2.1 Data required**

**List of towns in McLean County**

McLean County is located in central Illinois, about 2 hours south of Chicago. It is a fairly rural area with lots of farms. According to Wikipedia, McLean County is the largest county by land area in Illinois and has a population of 169,572. A list of its 19 incorporated communities is available here:

<https://illinois.hometownlocator.com/counties/cities,cfips,113,c,mclean.cfm>

**List of populations and locations (latitudes and longitudes) for each town in McLean County**

Unlike the vast amounts of data available for big cities like NYC or Toronto, I have not found a single file or webpage that provides all necessary population and location data for all towns. Therefore, this information may need to be separated retrieved from different Wikipedia webpages. For example, the populations and location data for the communities of Anchor and Arrowsmith can be found here:

<https://en.wikipedia.org/wiki/Anchor,_Illinois>

<https://en.wikipedia.org/wiki/Arrowsmith,_Illinois>

**Existing venues for each town in McLean County**

The Foursquare API (<https://developer.foursquare.com/>) provides location based data with information about venues, users, etc. We have been using Foursquare data in various assignments in this course. Through these earlier assignments, I already have the credentials to retrieve data from Foursquare.

CLIENT\_ID = 'OTVHEHACLVFKWMQAAEARJ2PGX1SZAWEI2Y1V5Y4JQPUEZZZ3'

CLIENT\_SECRET = 'QQIEDAOXB4VDE5DMGFQ0ZGEZ3PJUQ0S2UIEIZZGRNNNE3BUQ'

VERSION = '20180605'

**2.2 Data Sources, Data Processing, and Tools Used**

As outlined in 2.1, the sources for the needed data are as follows:

* List of towns in McLean County will be retrieved from HomeTownLocator.com
* List of populations and locations (latitudes and longitudes) will be retrieved from individual Wikipedia pages for each town
* Information about existing venues in McLean County will be retrieved using Foursquare API.

Data processing and analytical tools used will include:

* Pandas library will be imported to create and analyze data frames
* Json library will be imported to process data in JSON files
* Requests library will be imported to handle requests
* Geolocation data (latitudes and longitudes) will be found with algorithm coding and Nominatim.
* Folium mapping will be needed to generate maps of McLean County.