Project Report

**Identifying Locations for John’s Ice Cream and Pizza Shop**

**1. Introduction**

**1.1 Background**

A friend of mine, John, is a successful business manager in New York City. He is very passionate about ice cream, and since his childhood, he has dreamed of one day owning and operating an ice cream shop. As John’s 45th birthday approaches, he increasingly finds his current job unsatisfying, and has been thinking more and more about following his passion, moving closer to his family in Illinois, and having a slower pace of life. Having worked full time for 20 years, John has accumulated a good amount of savings that would provide the initial investment to get his ice cream shop started.

As the demand for ice cream is seasonal, John needs to offer more than just ice cream so that he can attract customers during colder months and have stable cash flow. Fortunately, John worked in a pizza restaurant during his high school years and is quite familiar with the pizza business. So, he has decided that in addition to ice cream, his shop will also make and sell pizza.

**1.2 Business Problem**

As John’s business plan gradually takes shape, he needs to decide where to open his ice cream and pizza shop. John has no doubt in his mind that the shop will be in McLean County, Illinois, where he grew up and many of his relatives still live. But he wonders which town in the county would be most ideal for his business.

After hearing that I am taking a data science course online and know how to use Foursquare data to explore venues and locations, **John has asked that I suggest two or three towns in McLean County for his future business that meet the following criteria**:

1. The town must be located in McLean County, Illinois
2. The town must not already have an ice cream shop.
3. The town must not already have a pizza restaurant.
4. The town has a substantial population to sustain the business.

**1.3 Interested Audience**

The interested audience is my friend **John**, who needs suggestions of three possible towns to locate his future 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.

**3. Methodology**

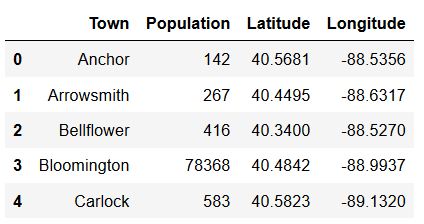
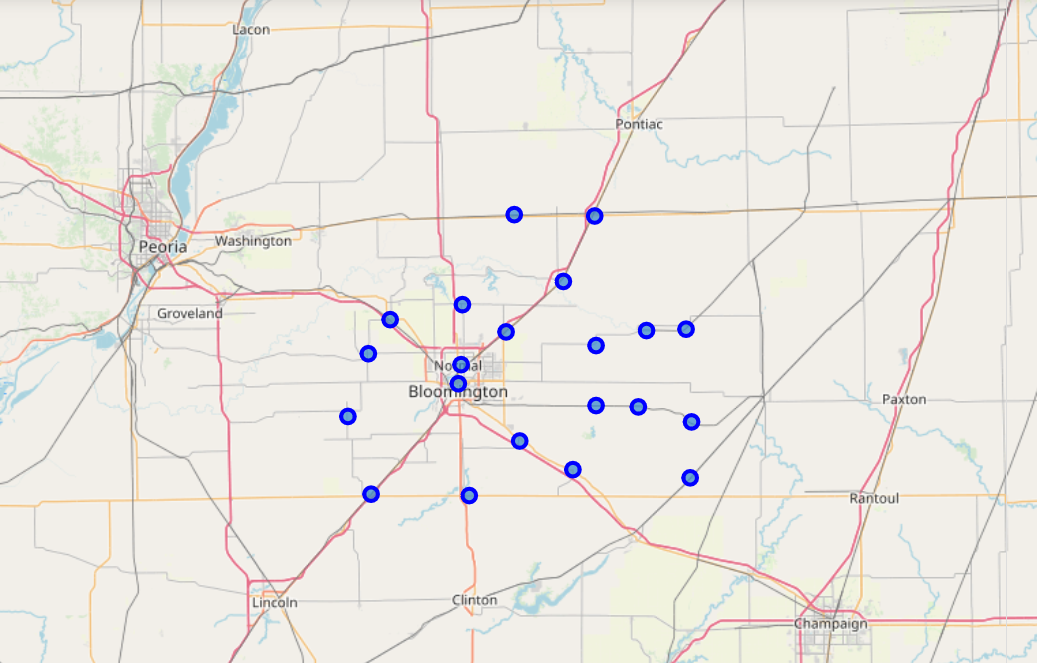
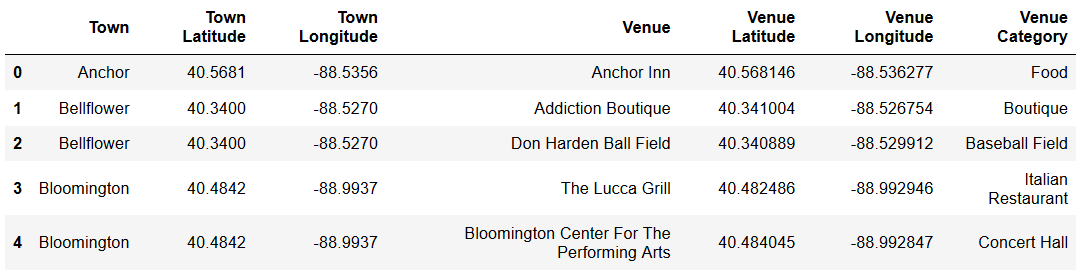
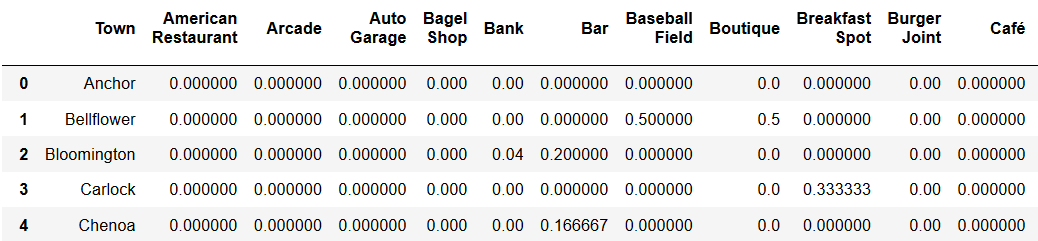
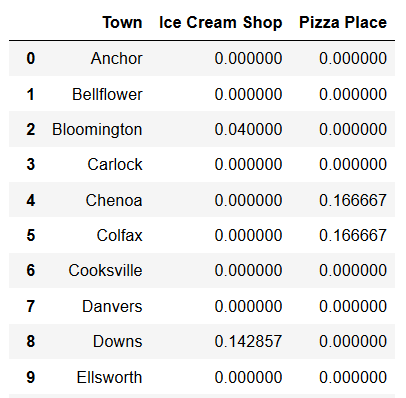
**3.1 Overview of methodology**

To identify the three largest towns in McLean County that currently have neither an ice cream shop nor a pizza restaurant, we will first construct a data frame that contains a list of the towns in the county along with their respective populations and location data (latitudes and longitudes).

We then use Foursquare API to get venues data to calculate the total numbers of these two categories of venues for each town. Based on these sums, we can then produce a list of towns that have neither type of restaurants. Finally, we can sort these towns based on their populations and identify the 3 towns with the largest populations.

**3.2 Detailed steps taken**

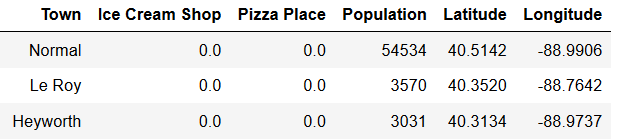
To describe this analytic process in detail, the following steps are taken:

1. Gather information from the Wikipedia pages for each town in McLean County (including name, population, latitude, and longitude) and compiled it in a CSV file
2. Read the CSV file into a pandas data frame  
     
   
3. Create a folium map for McLean county with all its towns superimposed on it  
     
   
4. Construct the GET request and send it to Foursquare to get JSON results for each town. (A limit of 100 venues will be set, and the radius will be 500 meters.)
5. Clean the JSON results and structure it into a data frame
6. Display all venues in all towns in McLean County  
     
   
7. Use the groupby method to calculate the total numbers of venues in each category for all towns  
     
   
8. Create a new data frame that only includes three columns: town names, total number of ice cream shops, and total number of pizza places  
     
   
9. Remove towns that have positive totals of ice cream shops and pizza restaurants
10. Sort the remaining towns by population and display the 3 towns with the largest populations

The above steps would yield the three largest towns in McLean County without an ice cream shop or a pizza restaurant. The results are discussed in the next section.

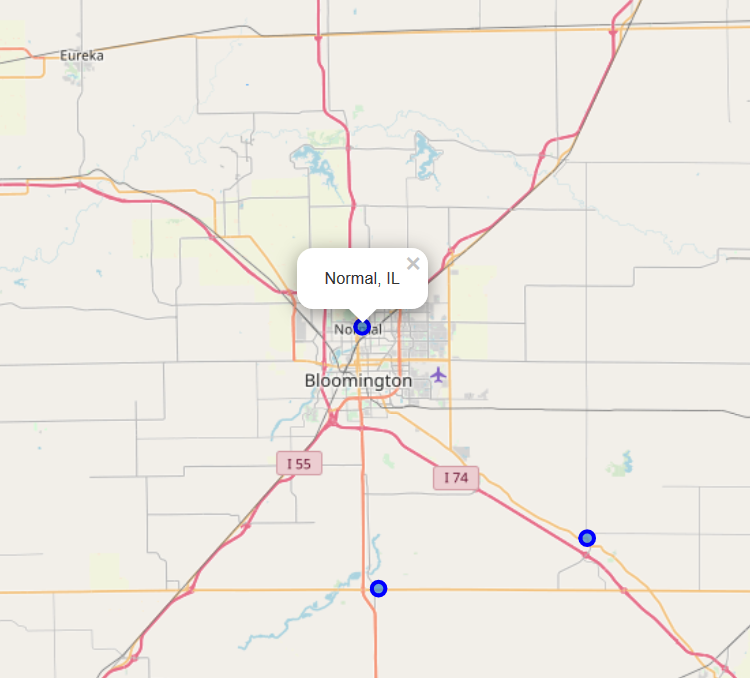
1. **Results**

After following the steps described above, the analysis led to three towns with the largest populations that currently do not have an ice cream shop nor a pizza restaurant. They are Normal, Le Roy, and Heyworth.



While both Le Roy and Heyworth are small towns with similar populations (both between 3,000 and 4,000), Normal is a much larger community with over 54,000 residents.

The following folium map displays the locations of these three towns.



1. **Discussions**

John is preparing to start a new ice cream and pizza shop in McLean County, Illinois, but needs recommendation regarding exactly where in the county to open this business. He has indicated that an ideal community for his shop will need to meet the following criteria:

1. The town must be located in McLean County, Illinois
2. The town must not already have an ice cream shop.
3. The town must not already have a pizza restaurant.
4. The town has a substantial population to sustain the business.

Based on these criteria, we gathered and analyzed data about each town (name, population, location) in McLean County as well as venues information from Foursquare API, we were able to identify the three towns with the largest populations but without either an ice cream shop or a pizza shop. They are Normal, Le Roy, and Heyworth.

Though all three communities meet John’s criteria, Normal is likely the most ideal choice given its much larger population (over 54,000), which is 15 times of Le Roy’s and 18 times of Heyworth’s.

**6. Conclusions and Recommendation**

Three conclusions will be presented to John:

1. The three towns that meet his selection criteria are Normal, Le Roy, and Heyworth.
2. Among them, Normal is likely the most ideal location for the shop given its larger population. Thus, **Normal will be recommended as the top choice for his shop**.
3. Le Roy will be recommended as the second choice, and Heyworth the third.