**CAPSTONE PROJECT. IBM DATA SCIENCES PROFESSIONAL CERTIFICATE.**

THE BATTLE OF NEIGHBORHOODS: FEASIBILITY STUDY TO OPEN NEW RESTAURANTS IN CHICAGO.

bY: Romulo guedez

mAY 2020.

1. introduction

Chicago is the third largest city in the United States of America, with a population of nearly three million people, and is home to seventy-seven (77) community areas and one hundred (100) neighborhoods (<https://www.chicago.gov/city/en/about/facts.html>).

One of its most important facts is that the city is seeing as a foodie destination, with seven AAA Diamond-rated restaurants, 26 Michelin-starred restaurants, host of the annual James Beard awards and it even has 144 dog-friendly restaurants. In addition, Chicago is a very charmed city on the southwestern shore of Lake Michigan, the 5th largest body of fresh water in the world. Chicago´s extensive parklands, including 3,000 hectares of city parks attract an estimated 86 million visitors annually.

The facts above-mentioned call for the attention of entrepreneurs who wants to invest in a place with big opportunities, mainly in the sectors of tourism and services, with special attention in the restaurant’s business.

The application of data sciences and machine learning methodologies can be very powerful tools that allow us to analyze the available information regarding to the restaurant sector in the city of Chicago. In the particular case of analyze the feasibility whether open or not new restaurants, these tools help us to solve several doubts such as, for example, hidden patterns of kinds of restaurants distribution thorough the city, segmentation of the city depending on the most popular restaurant categories, etc.

1. business problem

An important entrepreneur based on the east coast, which its main specialization is the branch of restaurants, is interested to expand their business to other regions of the country.

Chicago, Illinois, has been considered as a foodie destination in the United States. This big American city is the primary target in the expansion policies currently carried out by an tourism sector entrepreneur, whose current business are mainly based on the east coast.

The main aim of this project is provide some recommendations to the client, in order to help them to resolve the problem of opening new restaurants in the city of Chicago.

The recommendations will be base in this feasibility study results, which consist to analyze the current distribution of restaurants categories along the city. This could led us to visualize some hidden patterns, such as a possible segmentation of the restaurant categories distribution through the city. This will be a very helpful tool in order to understand the current market and help to make decisions about, for instance, propose preliminary locations for a determined kind of restaurant in a specific place.

1. DATA SECTION

* Data 1: Neighborhoods Information.

Neighborhoods information is required to learn about the city´s urban organization and segment its populate areas. This defines the scope of this project, addressed to the city of Chicago, which is home of 77 community areas and more than 100 Neighborhoods.

The source of Chicago neighborhoods data is the follow internet address:

(<https://en.wikipedia.org/wiki/List_of_neighborhoods_in_Chicago>).

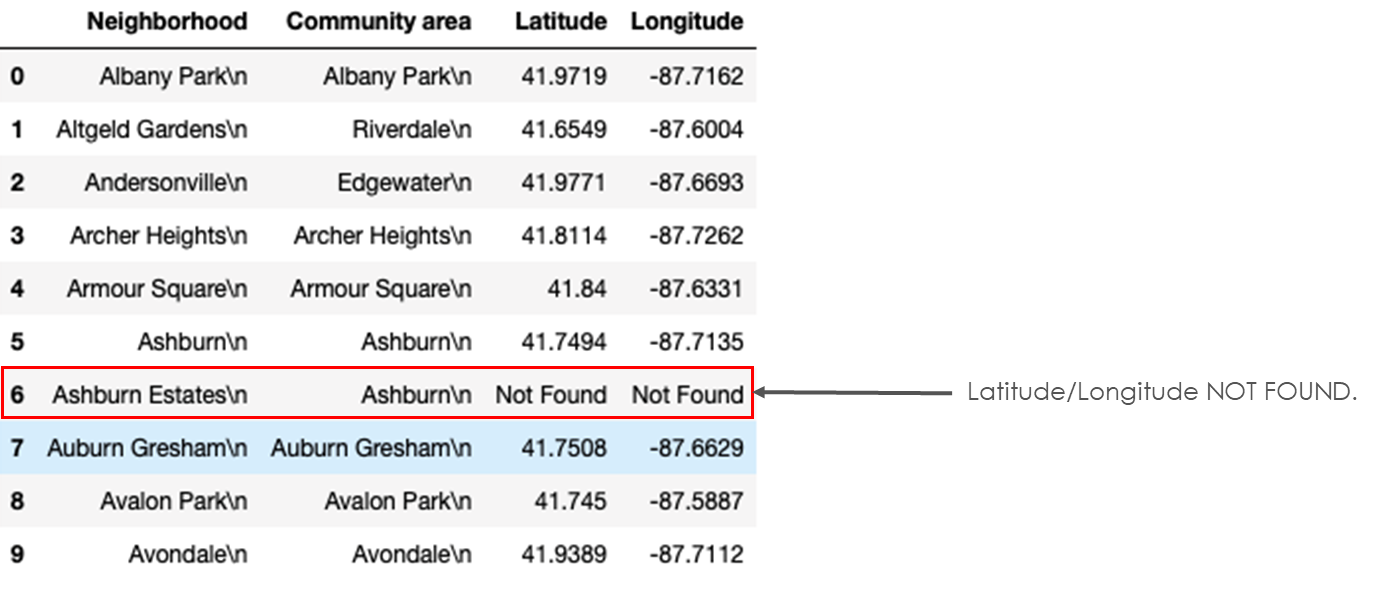
We used web-scraping techniques to extract the data from the above-mentioned Wikipedia page through the Python requests and BeatifulSoup packages. Then, we generated the dataframe with the Chicago neighborhood information.



List (Dataframe) of Chicago Neighborhoods.

* Data 2: Neighborhoods Coordinated (Longitude / Latitude).

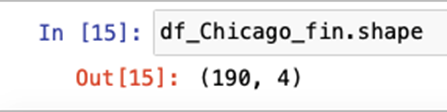
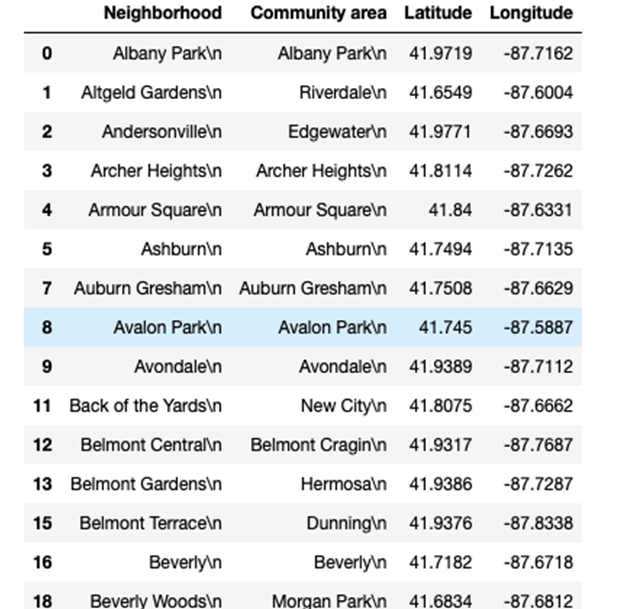
We could get the geographical coordinates of neighborhoods using Python Geocoder package (Nominatin), which allow us to know the latitude / longitude information.



Dataframe List integrating Neighborhoods names and Latitude/Longitude information.

.

There were some rows with neighborhood information with Latitude/Longitude Not Found. We performed data cleaning techniques to eliminate those rows without geographical coordinate information. After apply data cleaning, we have a dataframe with 190 rows or neighborhoods.

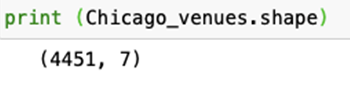


Neighborhoods Name and Latitude/Longitude List dataframe cleaned. Also, is show the dataframe dimensions.

* Data 3: Venues Information.

We used the integrate dataframe of Chicago neighborhoods & geographical coordinates as input for the Foursquare API ([www.foursquare.com](http://www.foursquare.com/)). Through this API we were able to find the basic venues information for each neighborhood. The basic venues information from Foursquare consist of Venue Name, Latitude, Longitude and Venue Category.





Dataframe listing Names, Latitude/Longitude of Neighborhoods and Venues Information in Chicago. We can also see the dimensions of dataframe, which have 4451 rows.