**Coursera Capstone**

IBM Applied Data Science Capstone

Opening a new Russian restaurant in Chicago

, IL



**Intro**

Chicago is a quite diverse and multicultural city and as a result it offers a wide variety of different cuisines from different countries for its guests and residents. Therefore, opening a restaurant could be a very challenging task. Potential new owner must consider a lot of things. What kind of restaurant it will be, what cuisine to serve, how will be his target customers and e.c.t. One of the crucial elements of opening restaurant is a location. Picking a right location could be a very hard task and one of the most important decisions that will determine whether the restaurant will be successful or failure in a future.

**Business Problem**

The objective of this capstone is to analyze and select the best locations in city of Chicago to open new Russian restaurant. Using data science methods and ML techniques this projects goal is to provide a recommendation to the one of the business questions like where potentially new restaurant owner of Russian cuisine should open his business in Chicago.

**Data**

* List of suburbs in Chicago city.
* Venue data (Restaurants with specific cuisine)
* Coordinates

**Sources of data**

The neighborhood data will be taken form Wiki page (<https://en.wikipedia.org/wiki/List_of_neighborhoods_in_Chicago>). It contains a list of all neighborhoods and community areas. This data will be extracted by scraping techniques, we will use **beautifulsoup** package for this. Coordinates will be provided by **Geocoder** package. Foursquare will provide the restaurant data.

**Methodology**

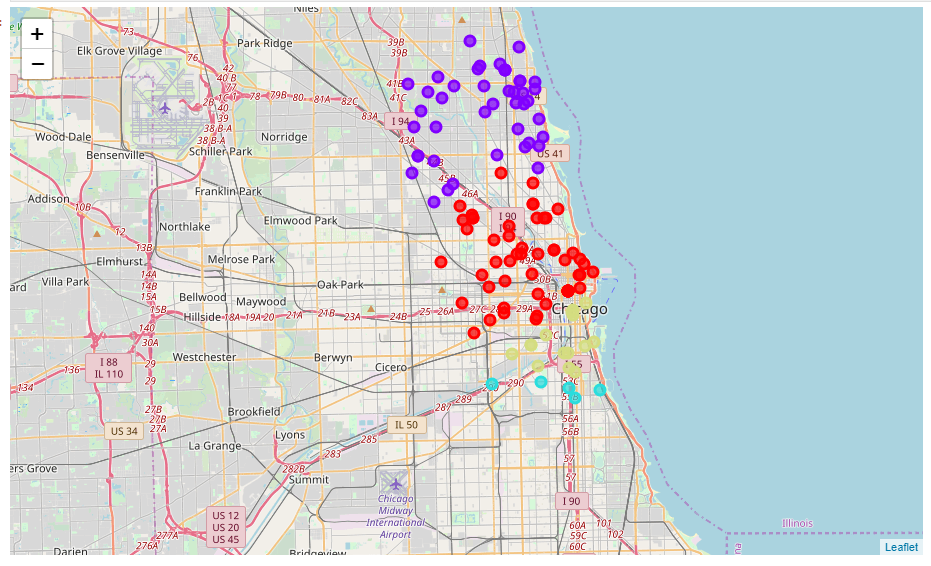
Initially, we need to get the list of Chicago area neighborhoods. Data could be attained from wiki page (link is provided above in Data section) by web scraping. Python request and beautifulsoup package will help to extract all data we need. This list only will have Chicago area neighborhood names and community area names. Next, we need to get coordinates (latitude and longitude) per row in a list of neighborhoods. All this data then exported to DataFrame and visualized in a map using Folium package.

Next, we will use FourSquare API to get Russian restaurants that are within 500 meters in each neighborhood.

**Results**

K-means clustering categorized Chicago neighborhood into 4 groups based on the density of Russian restaurant.

The results of the clustering are visualized on the map.



**Discussion**

Observing results and clusters we can notice that Russian restaurant are mostly located and distributed on the north side of the Chicago. Blue cluster group is the biggest one, it could be explained by the fact that one of the biggest Russian/ Ukrainian/ Polish communities’ that live on the north side neighborhoods of the Chicago. Second biggest group is red cluster in the center of the town. Third and fourth clusters are located way smaller and located on the south side of the town. Its also where China town located, and it explains why we don’t see a lot of Russian restaurants there.

We could conclude based on this information that its desirable to open new Russian restaurant on the north side of the town where potential customers live or in the downtown where we have a steady stream of tourists.

**Conclusion**

In this project we identified and stated a business case and requirements for data and what data is required. We extracted all necessary data and prepared for further analysis. We used machine learning clustering technique to group the data into 4 groups based on their similarities. Based on the result of that we propose options where new Russian restaurant could be potentially opened.