The Battle of the Neighborhood

Capstone Project

Content

- Introduction
- Data Selection
- Conclusion

Introduction

In this project, we are providing analysis for pizza place owners to find the proper location to open new pizza place in Chicago.

We will use machine learning powers to find a few most promising neighborhoods where there are not many pizza place yet.

Data Selection

1) Chicago neighborhood location

- We found a wikipage'https://en.wikipedia.org/wiki/Community_areas_in_Chicago' lists all the community areas in Chicao and its population of 2017. we will read the community table to a panda dataframe
- Geopy https://geopy.readthedocs.io/en/stable/ is a python library which enables geographic coordinates to be attained from addresses, cities, countries, and landmarks. This library will be utilized in order to coordinates of each of the Arlington neighborhoods.

2) Chicago community income per capita

- City of Chicago data portal provides a dataset for contains a selection of six socioeconomic indicators of public health significance and a “hardship index,” by Chicago community area, for the years 2008 – 2012. We will use the income per capita for each neighborhood for our analysis https://data.cityofchicago.org/Health-Human-Services/Census-Data-Selected-socioeconomic-indicators-in-C/kn9c-c2s2

3) Four Square Venue Data

- Four Square https://foursquare.com is a social networking service which provides an API with a feature that allows developers to find information about venues listed in their data that are nearby geographic coordinates. The API will be used to find venues that are nearby each neighborhood in Arlington in order provide data on the venue composition of them.

Conclusion

- As mentioned the purpose of this project was to explore Chicagao neighborhoods in order to find neighborhoods that could provide potentially successful pizza place locations. In order to determine possible solution neighborhoods, we use the K-means clustering algorithm to cluster the neighborhoods based on their basic demographics and types or categories of nearby venues.
- After the clustering, we can assume that if a single cluster has a much greater number of nearby pizza places in comparison to the other clusters, then the cluster has attributes that are favorable to pizza places.
- Comparing the feature of each cluster can give us also provide us with an understanding of what features a neighborhood needs in order to be potentially successful pizza place location.
- As a result of the clustering it can determined that there are three neighborhoods that provide a good solution West Garfield Park, Pullman,and West Englewood