

IBM

Applied Data Science Capstone

Project Week 5

Introduction

In this project, I'll create a hypothetical scenario for a concept that my friend is looking to open a Chinese Restaurant in Toronto. As a data scientist, he wanted me to recommend him where to open it. I plan to determine the best place to open a Chinese restaurant by identifying where there aren't many Chinese restaurants and many Asians.

Business Problem

The objective is find the most suitable location to open Chinese restaurant in Toronto, Canada. By using data science methods and tools, this project aims to determine that if my friend wants to open a Chinese restaurant in Toronto, which is the best location to do that?

Target audience

My friend who wants to find the location to open a Chinese restaurant.

Data

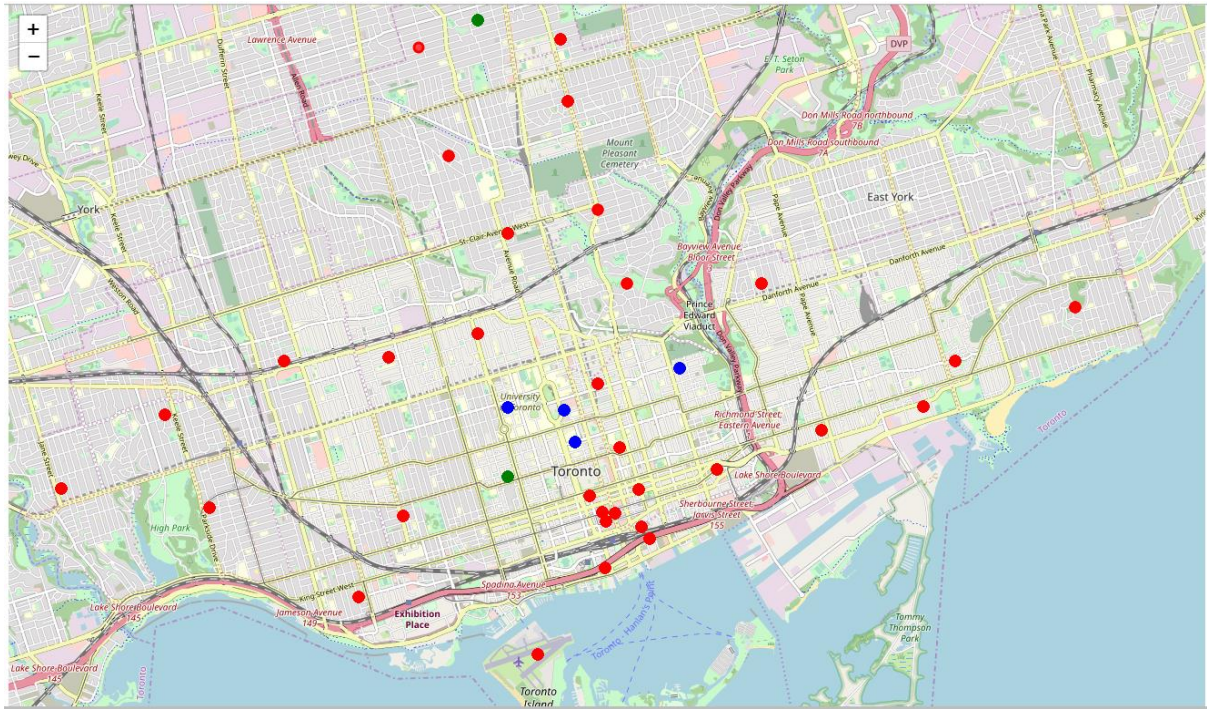
We will need below data:

- List of neighborhoods in Toronto, Canada.
- Latitude and Longitude of these neighborhoods.
- Venue data related to Chinese restaurant.

Methodology

- Get list of neighborhoods in Toronto. Use data source from wiki and process to get data by BeautifulSoup
- Use panda dataframe to merge list of neighborhoods data with coordinate data.
- Use foursquare to grouping data with venues
- Cluster data and I'll be already to recommend the location to open Chinese restaurant

Result



This picture is the result after I use k-means to show up how many Chinese restaurants in each areas order by categories cluster:

Cluster 0 (green): Neighborhoods with less Chinese restaurant.

Cluster 1 (blue): Neighborhoods with no Chinese restaurant.

Cluster 2 (red): Neighborhoods with more Chinese restaurant.

After what is seen on the simulator map, I recommend open Chinese restaurant near Toronto Islands, Union Station or Garden District. It might be a good location as there're not a lot of Chinese restaurant in these areas.