Applied DS Capstone project week5

Neighborhood search for business

Introduction

Business Problem

For this assignment I have chosen to solve the following business problem. Identify best neighborhood for establishing a day care center for children in Toronto city, Ontario, Canada to help small business owners to make a informed decision based on data analysis and insights derived from statistical analysis.

Background of the Problem

- There are various considerations to be taking into establishing a new business and particularly for a daycare center for children. Some of the features, I would like to use to determine the best neighborhood for a daycare center in Toronto are population, crime rate and existing businesses in neighborhoods to determine their feasibility for establishing a new business.

- Apart from this in a real-world scenario, there are various considerations to be taken before finalizing a location, like state laws, zoning laws, parking, and many other stringent safety requirements like, enough space for parking, safe play area, rest area for staff etc.

- For the scope of this assignment though, I will be exploring only potential neighborhoods to establish the day care, but will not go in depth about location etc, as that is beyond scope of this exercise.

Data

1. Scrape Postal code data from following Wikipedia page

2. Extract data from the csv file that has the geographical coordinates of each postal code and is located at http://cocl.us/Geospatial_data.

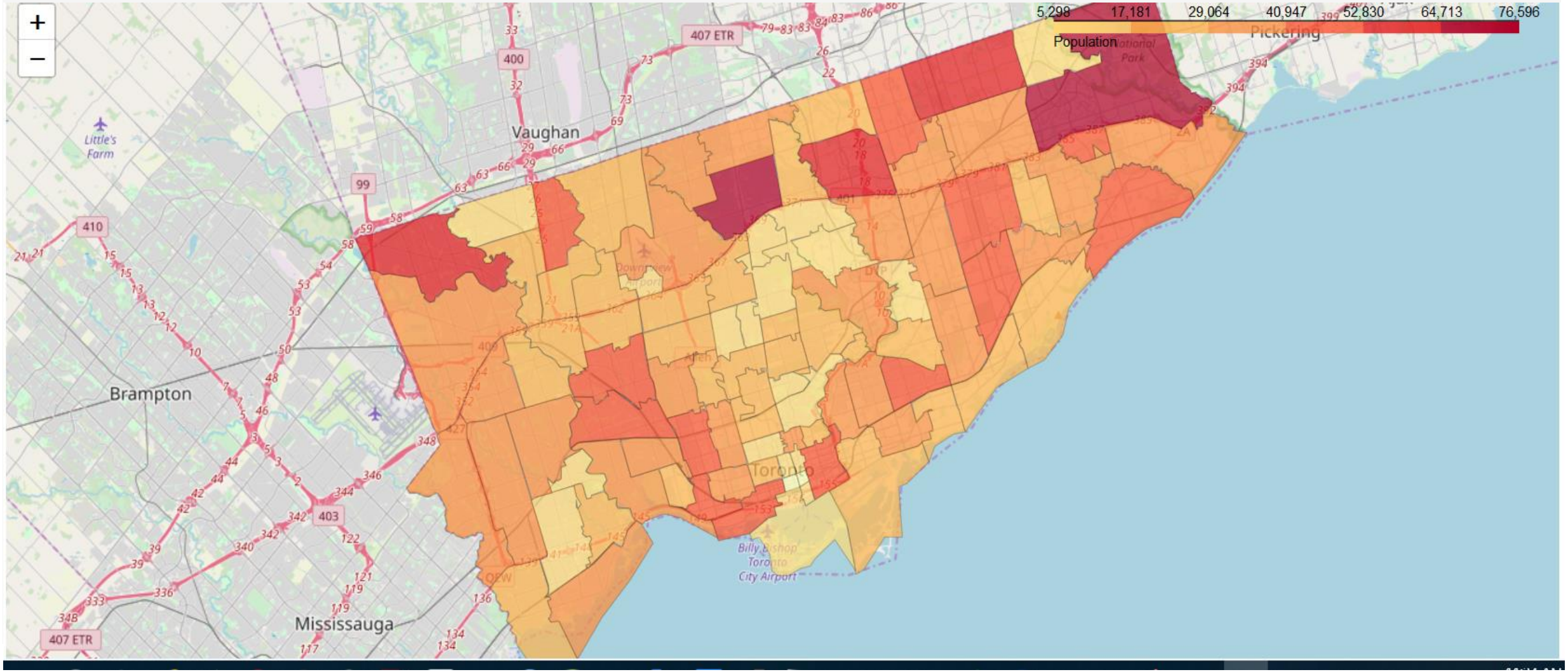
3. Using Foursquare API and Toronto latitude and longitude information from above

4. Get population data for each neighborhood in Toronto, Canada at Canada Statistics

5. Get Major Crime Indicator data for each neighborhood in Toronto, Canada from Toronto police open data

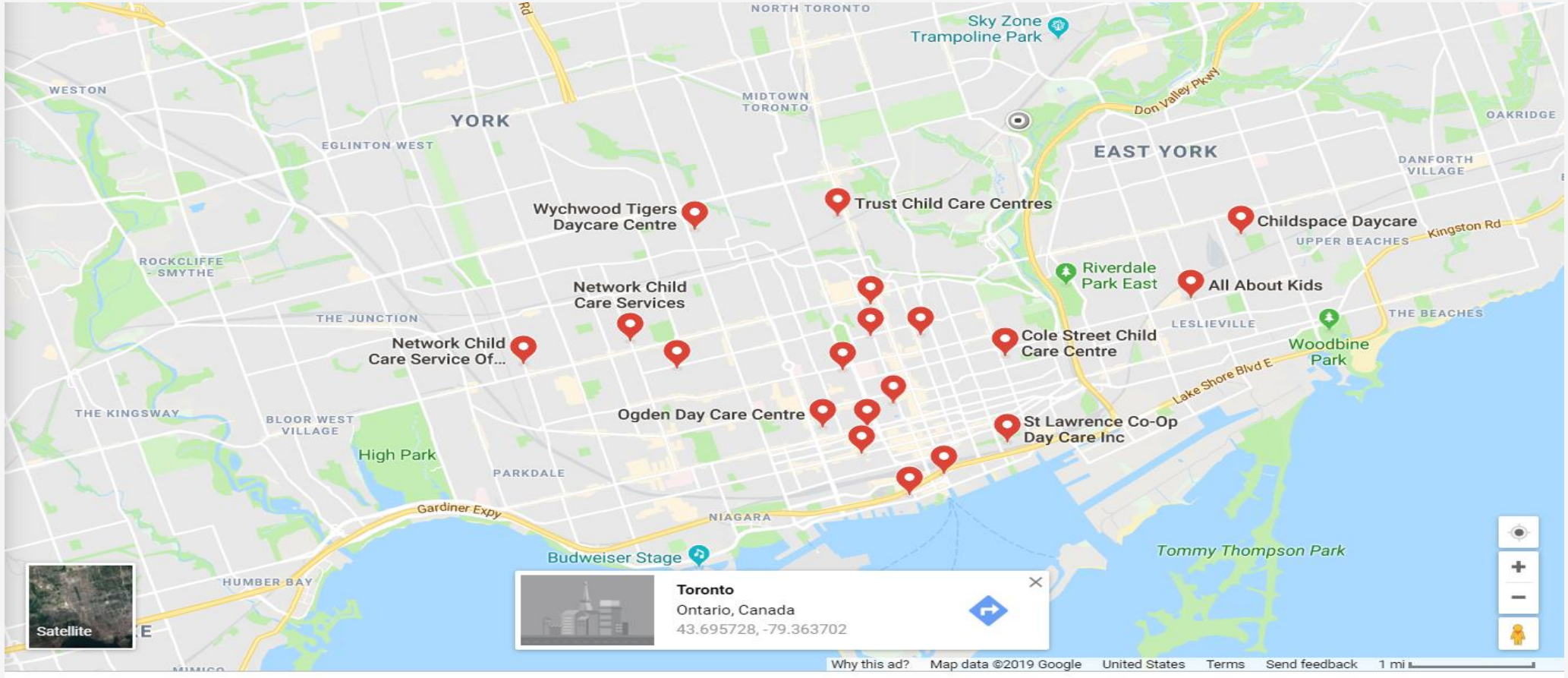
Data Analysis

Choropleth map of population distribution in Toronto neighborhoods



Results

Existing Child day care locations based on the data from Google Maps



Results

Identified potential neighborhoods for child day care business per ML and my analysis

Conclusion

- In a real-world scenario, first finding the neighborhoods would help short list potential location options, but after this exercise I felt, it would be interesting to find out the best location for a specific business than finding best neighborhood.
- With just the population and existing venue data, we were able to reduce the potential neighborhoods from 103 to 28 in the end.
- Adding couple of more data sets like number of schools in a neighborhood, office locations and transit options etc. would have enabled us to get more precise conclusions.