

Locating Potential Venues for Homeless Shelters in the City of Toronto

by

Yael Valdez Navarro, MASc.

A capstone project
presented to Coursera
in the completion of
the
IBM Data Science Certificate

Introduction

Background

There has been a rising trend in Canada, a vulnerable group of people with diverse issues such as drug abuse, mental health and fleeing abusive situations face the problem of homelessness.

Multiple charity groups organize soup kitchens and shelters to alleviate the daily hazards faced by this community. The fact is most of these locations operate at almost 100% occupancy rate every day. To tackle this, a business problem is proposed.

Business Problem

By considering the challenges presented in the previous section, the problem statement is then, by analyzing multiple sources of data (i.e. socio-economic and geographic), find the best possible locations for establishing new homeless shelters on the city of Toronto.

There are multiple criteria that are given more importance to the establishment of the best location. These following factors can also be considered our assumptions for the problem definition:

- **Accessibility:** The ideal venue should be close to public transport, train stations or bus stops.
- **General income by neighborhood:** It has been shown that low income areas are hot-spots for homelessness, specially with increasingly prohibitive rent costs.
- **Closeness to other shelters:** To maximize area covered, the ideal venue should not be too close to other venues.

Data

The data compiled to solve the proposed problem is gathered from various sources. An outline is displayed according to the criteria chosen.

- Foursquare: Will help with the retrieval of public transportation stations.
- Stat-Can: Two main datasets are associated with this source, mainly
 - [Homeless shelter locations:](#) Provides the longitude and latitude locations of most homeless shelters in the city of Toronto.
 - [Socio-Economic Data:](#) Access to multiple socio-economic factors, mainly, family income according to the neighborhood.

Alongside the datasets, several Python libraries will be used to exploit their functionality and their integration with the project.

- Pandas: Used for store large table-like arrays of data
- Requests: Used to handle external queries and requests.
- Numpy: Used to handle vectors and matrices in a more intuitive manner.
- Matplotlib: Used to graphically display different data.
- Folium: Used to display maps and visualize location-based data.
- Scikit-learn: Used to perform clustering analysis, such as, k-means.

In the following week the problem will be developed in Python. This document will be updated with the final product accomplishments.