



DATA SCIENCE PROFESSIONAL CERTIFICATION



K-means Clustering for Desirable Living Locale

January 20, 2019

Wade Hooper

K-means Clustering for Desirable Living Locale

Revision Version	Date	Comments
<i>BOTN-REP-WH-001</i>	01/20/2019	Added Introduction and Data sections

Table of Contents

1.0	Introduction/Business Problem	1
2.0	Data.....	1

K-means Clustering for Desirable Living Locale

1.0 Introduction/Business Problem

Moving for many people can be a complicated and stressful ordeal. In a large city such as Ottawa, there may be different requirements when moving to a new neighbourhood; whether these are comforts similar to home (foods, shops, etc.) or safety (such as a similar lack of high crime levels in the surrounding area).

This paper will focus on providing an analysis of neighbourhood safety as compared to available services and stores. Simply put, if a person had certain preferences to shops or restaurants in a new neighbourhood, would that same neighbourhood be one where they might feel safe to live, commute, and work?

2.0 Data

To visualize the distribution of nearby restaurants, stores, or shops, the Foursquare API will be used to retrieve location data for each venue from nearby neighbourhoods ([shown here](#)). Crime incidents will be visualized via data provided by the Ottawa Police Service as compiled by Carleton University ([shown here](#)); this dataset lists incident occurrences for specific crimes (assault, theft, etc.) for the 2016 year within each neighbourhood ward (or borough).

The Foursquare data will be used to cluster neighbourhoods across Ottawa via a k-means algorithm. As the clusters would highlight trends, doing so would gain insights into popular venues of each neighbourhood; such as if neighbourhoods in Ottawa's downtown core have a large variation in available restaurant cuisine (e.g. Asian, French, etc.), or if the neighbourhoods further from the city center have nearby fitness facilities. The resulting clusters will be displayed on an overhead map of the Ottawa region.

After the city has been clustered with available venues, the crime data will be incorporated and displayed on the overhead map. Depending on the data spread (e.g. if neighbourhoods have non-null values), incidents may also be clustered via a k-means algorithm to show a relationship between crime level and a desired neighbourhood; an example being a high number of thefts for neighbourhoods bordering canals.