

Sean Devine
February 2020

Capstone Project
Battle of the Neighborhoods

Report

1. Introduction

A key part of any business is expansion. In Los Angeles, California, stakeholders have decided to open a second branch of their company. Business at their current location is going well so they would like to open their second branch in a neighborhood that is very similar to their current one. To find a neighborhood that is similar to their current one, the Foursquare API will be utilized in combination with k-means clustering. From the similar neighborhoods, the business stakeholders would like to then select the one with the lowest amount of crime.

2. Data

In order to find the neighborhood best suited for the business, the foursquare API will be used to find the most common venues by neighborhood. This data will then be used to cluster the neighborhoods using k-means. On the cluster map, similar neighborhoods can be easily identified. Then, a second layer of the map will be created which identifies crime by neighborhood. This data is made available by the US government and spans from 2010 until 2020. With both layers of the map, the stakeholders can easily choose a neighborhood that is both similar to their current location and has a low crime rate.

In summary, the data needed is location and venue data from the foursquare API and crime data with coordinates from the US government.