

IBM Applied Data Science Capstone Project

Business Problem

This project aimed to find the best neighborhood for opening a new mall in Denver, Colorado by comparing the Toronto. Here are some questions that will be answered to decide the best neighborhood for the new mall.

1. Should the new mall be close to restaurants or more residential areas?
2. Should the new mall be in more multicultural neighborhoods?
3. Should the new mall be accessible with variety of public transportations?

Data Collection and Data Sources

For Toronto, previous data set will be used. For Denver, I will scrap the data from https://en.wikipedia.org/wiki/List_of_neighborhoods_in_Denver then I will get the geographic coordinates with geocoder.

Methodology

These libraries below is required to complete this project:

Pandas: For creating and manipulating data frames.

Folium: Python visualization to visualize the neighborhoods cluster distribution of using interactive leaflet map.

Scikit Learn: For K-Mean clustering.

JSON: For importing JSON data.

XML: For scraping data from the web site.

Geocoder: To retrieve Location Data.

Beautiful Soup and Requests: To scrap and library to handle http requests.

Matplotlib: Python Plotting Module.