The Battle of neighborhoods

Capstone project

Background & Business Problem

People who move to Vancouver, BC ask themselves:

Where exactly should I be looking to buy if the distance between my future home and the following locations of interest is the most important factor for me?:

- nearest beach
- nearest school
- nearest coffee shop
- nearest restaurant
- nearest park

Data

The city of Vancouver publishes all its block numbers and corresponding location data on its Open Data Portal: link

We use the Foursquare API to identify venues of interest as well as their location data: link

Methodology

Gather data on blocks & venues of interest

1.1) Identify one central location point per neighborhood (centroid)

1.2) Request all "Coffee Shop", "Restaurant", "School", "Beach", "Park" around each centroid

1.3) Merge with results of other neighborhoods into one DataFrame and remove duplicates

Identify the closest venues of interest for each block

2.1) Develop function to calculate the distance of the closest venue[i] for each block

2.2) Calculate the distance of the closest venue for all venue types for each block

Cluster the blocks based on walkability using K-means

3.1) Cluster Vancouver's blocks using k-means.

3.1) Compare the means of the features of each cluster to understand the different characteristics of the areas within Vancouver as far as walking distances are concerned.

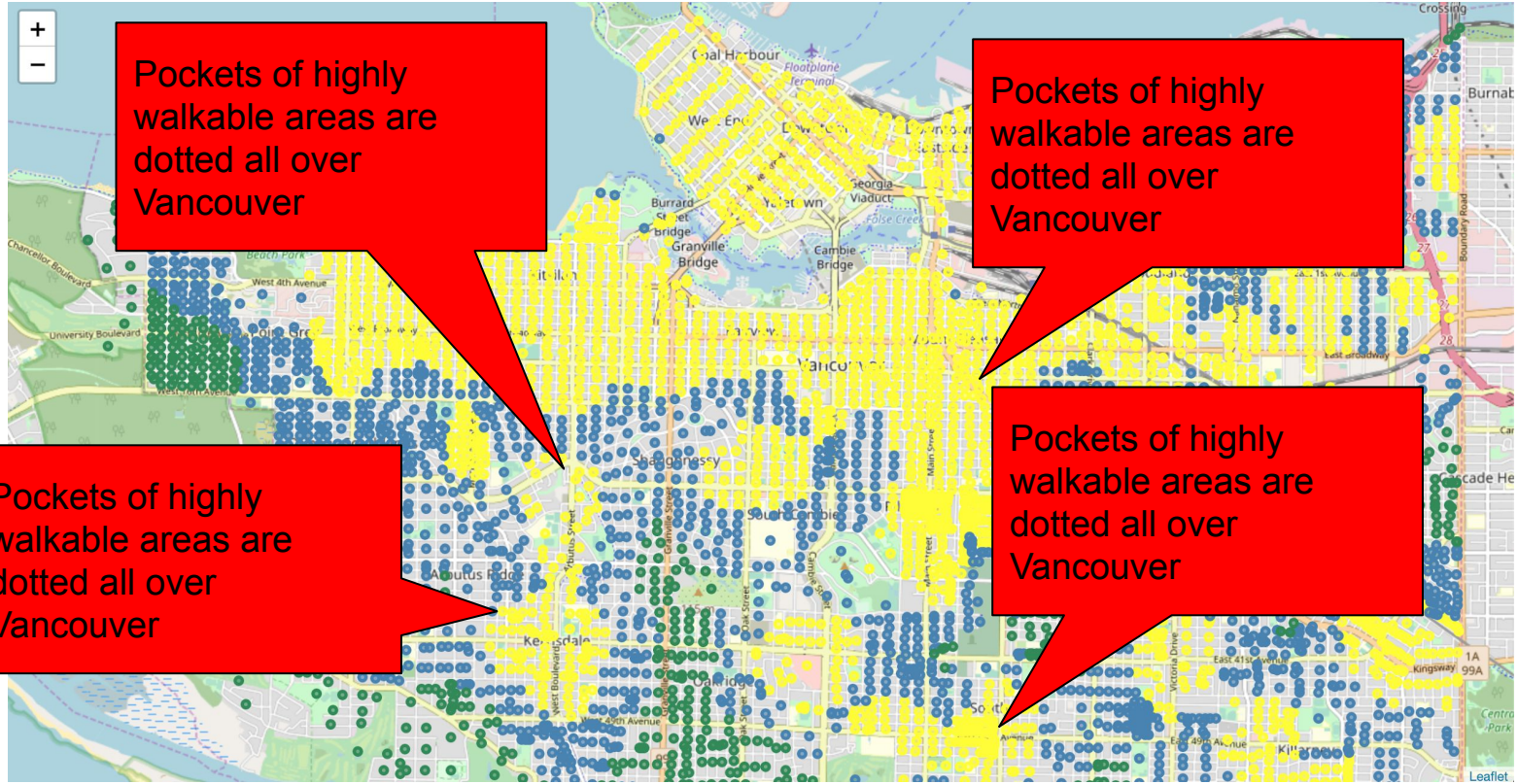
Analysis

cluster	Distance to nearest Beach	Distance to nearest Coffee Shop	Distance to nearest Restaurant	Distance to nearest School	Distance to nearest Park
0	1608.130602	296.579779	266.787993	367.357096	310.580154
1	1831.407743	725.483793	544.035172	441.267086	371.192606
2	1953.724817	1282.119105	689.000175	724.043411	497.749299

We can characterize the clusters as follows:

Cluster 0: Close to all venues of interest
Cluster 1: Medium distance to venues of interest
Cluster 2: Furthest away from venues of interest

Visualization & Discussion (1/2)



Visualization & Discussion (2/2)

Cluster 0
Cluster 1
Cluster 2

Future directions

This analysis is very limited:

1) Small number of features considered.
2) All features are weighted evenly which may or may not reflect your personal preferences.

Future improvements:

1) Addition of features including
    a) distance to grocery shops
    b) crime rates
    c) land value
    d) access to public transit
    e) average commute
2) Ability to assign personalized weights to features