



Capstone Project - The Battle of Neighborhoods

Danilo Evangelista (https://github.com/danilodn/)

1. Notes

Capstone project report for the "Applied Data Science Capstone" course from the "IBM Data Science Professional Certificate" on Coursera (https://www.coursera.org/professional-certificates/ibm-data-science).

2. Introduction/Business Problem

Stakeholders want to open a new coffee shop branch either on Vancouver or Toronto and need the subsidiary information on the city/competitors to decide about the expanding strategy.

Their intention is to open one Café shop right away and elaborate a long-term strategy to allocate several stores over the next years in these two cities.

The plan should consider the competitors' location and neighborhood characteristics.

3. Dataset

The first set of data used was the list of boroughs and neighborhoods of Toronto and Vancouver, extracted from Wikipedia (https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M and https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_V), as seen in the following images:



Toronto's list of Postal Cods, Boroughs and Neighborhoods







Vancouver's list of Postal Cods, Boroughs and Neighborhoods

That way, we could get 29 different postal code neighborhoods from Vancouver and 102 from Toronto.

We also retrieved the list of venues for both cities from Foursquare, each limited to 100 occurrences from 1000 meters of each neighborhood centroid, which resulted in 1772 different category venues for Vancouver and 4951 for Toronto.