

IBM APPLIED DATA SCIENCE CAPSTONE

WEEK 4

INTRODUCTION

This is my capstone project for IBM DataScience Professional Certificate. In the last months, people everywhere is using more than usual their PCs, notebooks, tablets, etc, and it means that hardware or software problems has increased a lot.

I've created an scenario for people who has skill of IT and wants to open a business in Toronto-Canada or for somebody who wants to open a different bussiness.

I've analized that there are not enough Computer Repair Centers in Toronto area. Therefore, it could be a great opportunity for a entrepreneur who is based in Canada.

BUSINESS PROBLEM

The main objective of this capstone project is to find the most suitable location for the entrepreneur to open a Computer Repair Center in Toronto, Canada. I've used data science methods and tools along with machine learning algorithms such as clustering.

This project aims to provide a solution to answer the business question: In Toronto, if an entrepreneur wants to open an Computer Repair Center, where should they consider opening it?

DATA

I've used the following data:

- List of neighborhoods in Toronto, Canada
- Latitude and Longitude of these neighborhoods
- Venue data related to each borough along with the top 6 most common venues. This will help us find neighborhoods that are more suitable to open an Repair Computer Center.

EXTRACTING THE DATA

- The scrapping of Toronto neighborhoods via Wikipedia.
- Getting Latitude and Longitude data of these neighborhoods via Geocoder package.
- Using Foursquare API to get venue data related to these neighborhoods.

METHODOLOGY

- First, I got the list of neighborhoods in Toronto, Canada. This is possible by extracting the list of neighborhoods from Wikipedia:

“https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M”

- Second, I've made an analysis with the top 6 most common venues and the top 10 venues for each borough in Toronto - Canada. I've identified that there are a few business dedicated to repair computers near to top 10 venues.

- Third and final step, I will focus on most promising areas and within those create clusters of locations where an entrepreneur can open a business related to computer repairs. I've presented a map of all such locations but also create clusters (using k-means clustering) of those locations to identify general zones / neighborhoods / addresses which should be a starting point for final 'street level' exploration and search for optimal venue location by stakeholders.