Coursera Capstone Project Presentation

BY PIRTHAM SANGIONE

Table of Contents

- Objective
- Methodology
- Results
- Benefits
- Limitations
- Q/A

Objective

▶ As a data scientist with 'Entrepreneurs R US', we were put in charge of creating an interactive tool to help entrepreneurs in the Greater Toronto region.

▶ This online service will provide the entrepreneurs with all the necessary information that will help them in picking the location for their new business ventures.

▶ The user will be able to research their area of interest based on their desire to support a specific ethnic group/community in Toronto.
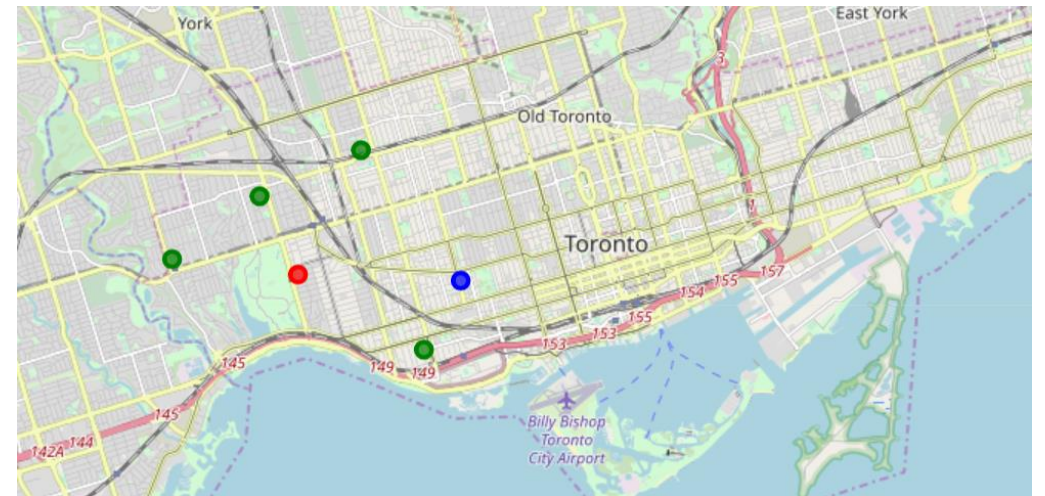
Methodology

▶ We followed a precise plan of attack shown below

Scripting Language	Python 3.6
IDE	https://labs.cognitiveclass.ai/tools/jupyterlab/
Input	1. Wikipedia page containing a table that lists all the postal codes starting with the letter "M". 2. Webpage on Statistics Canada website containing census information from the year 2016. Data was downloaded as a CSV file named "population.CSV". 3. Location service that leveraged location data for venues in Toronto.
Output	1. Pandas dataframe with relevant data 2. Clustered venue locations 3. Clusters visualized on a map of Toronto

Results

	PostalCode	Borough	Neighborhood	Population, 2016	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
76	M6H	West Toronto	Dovercourt Village, Dufferin	44950.0	43.669005	-79.442259	1	Discount Store	Bakery	Pharmacy
77	M6J	West Toronto	Little Portugal, Trinity	32684.0	43.647927	-79.419750	2	Bar	Asian Restaurant	Coffee Shop
78	M6K	West Toronto	Brockton, Exhibition Place, Parkdale Village	40957.0	43.636847	-79.428191	1	Coffee Shop	Café	Breakfast Spot
82	M6P	West Toronto	High Park, The Junction South	40035.0	43.661608	-79.464763	1	Mexican Restaurant	Café	Gastropub
83	M6R	West Toronto	Parkdale, Roncesvalles	19857.0	43.648960	-79.456325	0	Gift Shop	Breakfast Spot	Coffee Shop
84	M6S	West Toronto	Runnymede, Swansea	34299.0	43.651571	-79.484450	1	Café	Coffee Shop	Italian Restaurant

- On the right, we have the outputted table of clustered venues in the area of interest (West Toronto)

- We then visualized these neighborhoods on a map of Toronto with a different color representing a different cluster

Benefits

▶ The interactive tool provides flexibility to the user, allowing them to choose from a variety of Toronto boroughs for detailed analysis

▶ Efficient processing and analysis of information using advanced machine learning methods (K- Means clustering)

▶ Ability to access commercial location data via Foursquare API. Free to sign up!

Limitations

▶ Foursquare API location data has a limit on number of "calls" made for certain pieces of information (ratings, reviews, etc)

▶ Limited to the Greater Toronto region, with postal codes ending in the letter "M"

▶ Additional datasets need to be accessed and wrangled in order to output a more precise dataframe of information. (Ex: Toronto real estate prices)

Q/A

- Does anyone have any questions???

Thank you!

- On behalf of everyone here at Entrepeneurs R Us, we thank you for allowing us to be a part of your business idea. We hope to see you again soon!