Suitable Toronto Neighborhood(s)
for
Ethnic Cuisine Restaurant Chain

Author: Robert Ferdinand
Email: UOFLOUISIANA@GMAIL.COM

Problem

➢ US-based, IPBS Foods, Inc. (specializing in south Asian cuisine) plans to open new restaurant in Toronto, Canada.

➢IPBS seeking Toronto neighborhood(s) that have substantial populations leaning towards ethnic cuisine, in particular south Asian, so that business is:

- Profitable.

- Attractive to potential employees.

- Outreaching to local-area community.

- Helpful to Toronto economy leading to bolstering own standing in the United States.

Approach

➤ IPBS recruits US-based data science company DataTechWiz, Inc.

➤DataTechWiz will help IPBS decide on most appropriate Toronto neighborhood(s) to open their new restaurant in.

➤DataTechWiz and IPBS meet to decide on needs of IPBS as well as compensatory expectations of DataTechWiz with 10% bonus to DataTechWiz if business profits in first year of operation.

➤Mutual Commitment on continuous communication during project.

➤DataTech Wiz plans to leverage FOURSQUARE API location data via pandas package in PYTHON programming language in IBM's Watson Studio platform.

DataTechWiz's Strategy

➢ Obtain following available Toronto area neighborhoods data:

   (a) Data Set 1: Postal Codes for each Toronto borough and neighborhood (CSV file format).

   (b) Data Set 2: Geographical coordinates (latitudes and longitudes) for each Toronto Postal Code (CSV file format).

➢Read Data Sets 1 and 2 as pandas Data Frames (say TORONTO_DF1 and TORONTO_DF2, respectively).

➢Clean TORONTO_DF1 and TORONTO_DF2 as follows:

   (a) Remove all extra white spaces.

   (b) Drop missing (NaN) data columns.

DataTechWiz's Strategy (Continued)

➢ Merge TORONTO_DF1 and TORONTO_DF2 on common "Postal Code" key (just like joining tables in MySQL) to give Data Frame TORONTO_DF3.

➢ TORONTO_DF3 maps Toronto Neighborhoods to respective latitudes and longitudes.

Using FOURSQUARE API Location Data

➢Enter FOURSQUARE API credentials in PYTHON.

➢Apply FOURSQUARE API location data to TORONTO_DF3 to obtain Data Frame of Toronto neighborhoods' restaurant venue categories. Call this Data Frame TORONTO_DF4.

➢Proceed to using unsupervised, unlabeled, K-Means Clustering Machine Learing Algorithm on TORONTO_DF4 for the clustering or segmentation of Neighborhoods.

K-Means Clustering Machine Learning Algorithm

➤ Perform "One-Hot Encoding" on TORONTO_DF4. Converts categorical to numeric data to give Data Frame TORONTO_DF5.

➤ Group TORONTO_DF5 on "Neighborhood" key on mean of numerical data obtained from the "One-Hot Encoding". Resulting Data Frame TORONTO__DF6

➤ Apply K-Means Clustering Algorithm on TORONTO_DF6 with k = 5 clusters. Produces 5 clusters of Neighborhoods.

➤ Cluster descriptions in IPYTHON notebook on GitHub.

➤ Cluster Map Provided in next slide.

Toronto Neighborhoods Cluster Map
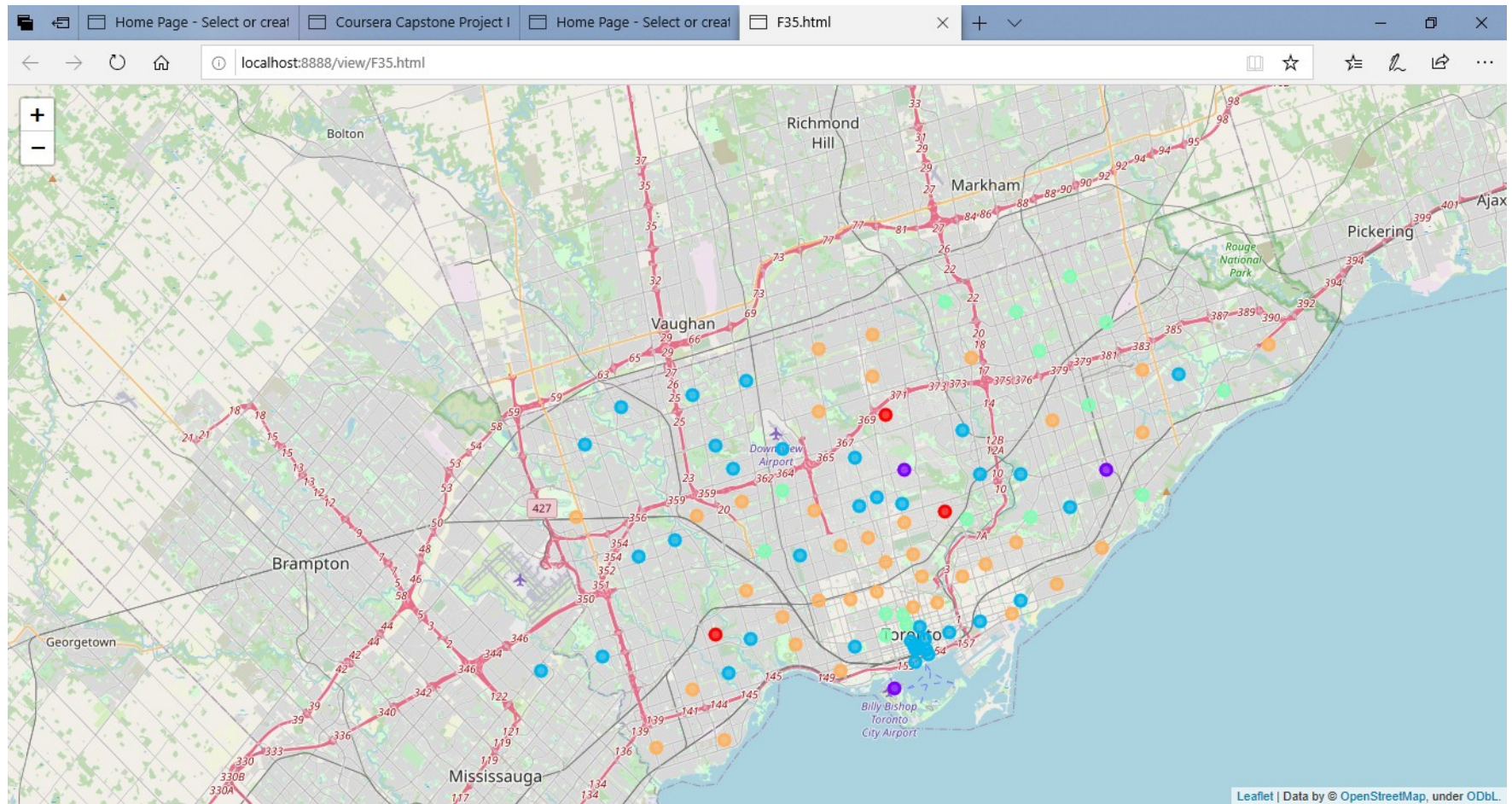
Neighborhoods Recommended by DataTechWiz Inc.
for
New IPBS Restaurant

➤Top Neighborhoods Recommended:

- Leaside
- Victoria Village
- India Bazaar
- The Beaches West
- Canada Gateway Processing Center (Mississauga area).

Acknowledgments

➢Coursera Inc.

➢Course Instructor Dr. Aklson

➢PYTHON programming language

➢FOURSQUARE API

➢PANDAS

➢IBM Watson Studio Platform