**Finding a Suitable Location to Open a Pastry Shop in Toronto**

Reginald P. Chua

January 25, 2021

1. **Introduction**

Ontario has the strongest economy in Canada, contributing the highest GDP per capita in the country. Toronto, its capital, has one of the top food scenes in Ontario. As aspiring pastry chefs based in Toronto, we want to open our first Pastry shop here. Which neighborhoods in Toronto can we consider?

Our Pastry shop concept focuses on products that go well with either coffee or tea. While there will be a small area reserved for diners, we are considering take-aways to drive most purchases. The product selection, for now, will focus on pastry products only. As we are entering this industry and market, we want to take the opportunity to position ourselves in neighborhoods with coffee shops and tea houses that serve products which complement our pastries so that we can more easily tap our market. This can also set the foundation to expand our reach later.

1. **Data Overview and Limitations**

The analysis uses data gathered from and limited to the following sources:

* **Wikipedia**
  + - An existing Wikipedia page contains information on postal codes, boroughs, and neighborhoods in Toronto. We will use web scraping to utilize the data.
    - <https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M>
* **Geospatial Data**
* A csv file on the web contains the coordinates of each postal code in Toronto, as provided by the IBM Data Science Capstone project.
* <https://cocl.us/Geospatial_data>
* **Geopy library**
  + Neighborhood coordinates are supplied through this library.
* **FourSquare**
  + This API grants access to data about establishments within a neighborhood. We utilize the data on the categories of establishments in each neighborhood, focusing on competitor and complementary industries including the following establishments: Bakery, Café, Coffee Shop, Donut Shop, Pastry Shop, Tea Room

1. **Target Audience**

The information here may be of interest to the following groups of people, while being mindful of the context and data sources stated above:

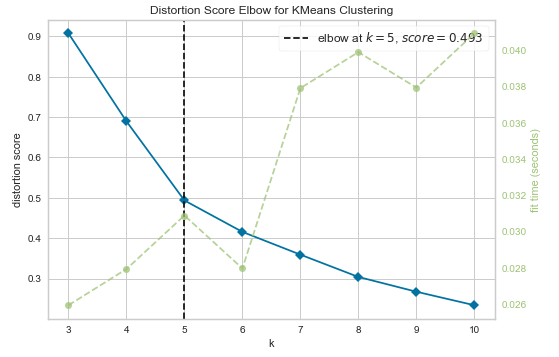
1. Entrepreneurs in search of a brief analysis involving the pastry shop industry in Toronto.
2. Data Science students and learners
3. Junior Data Analysts or Junior Data Science Professionals
4. **Methodology and Procedure**

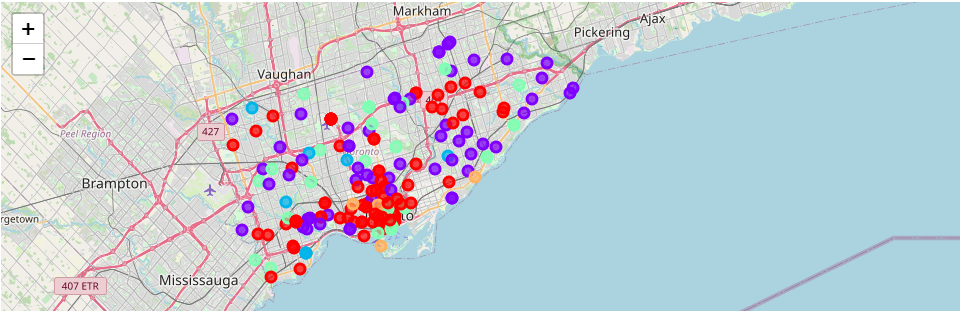
With the data described above, this study uses K-Means Clustering to group similar neighborhoods into clusters to help identify potential locations to open our first pastry shop.

We carry out the following procedures:

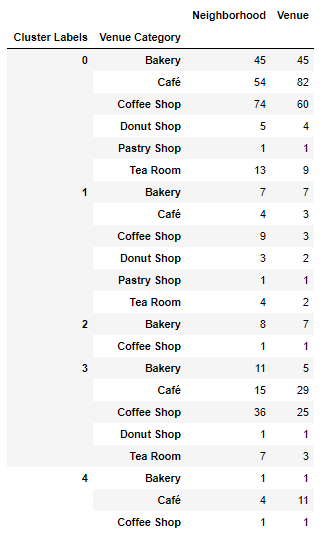
1. Data acquisition and staging
   * Scrape data from Wikipedia. This contains data based on postal codes within Toronto.
   * Obtain longitudes and latitudes for each postal code using geospatial data.
   * Refine the data from postal code level down to neighborhood level.
   * Obtain the coordinates per neighborhood.
   * Blend FourSquare data into each neighborhood to have a look into the types of establishments within each neighborhood.
2. Data transformation
   * Perform one-hot encoding in preparation for clustering.
   * Calculate for the means for each establishment type in each neighborhood.
3. Cluster and examine
   * Use the K Elbow Visualizer to identify the best k to use for clustering.
   * Perform K-Means Clustering.
   * Study each cluster with the aid of visualization and by examining the data points.
4. **Data Exploration and Analysis**

With the aid of the K Elbow Visualizer, we choose k = 5 and execute K-Means clustering then use Folium to visualize the results geographically.





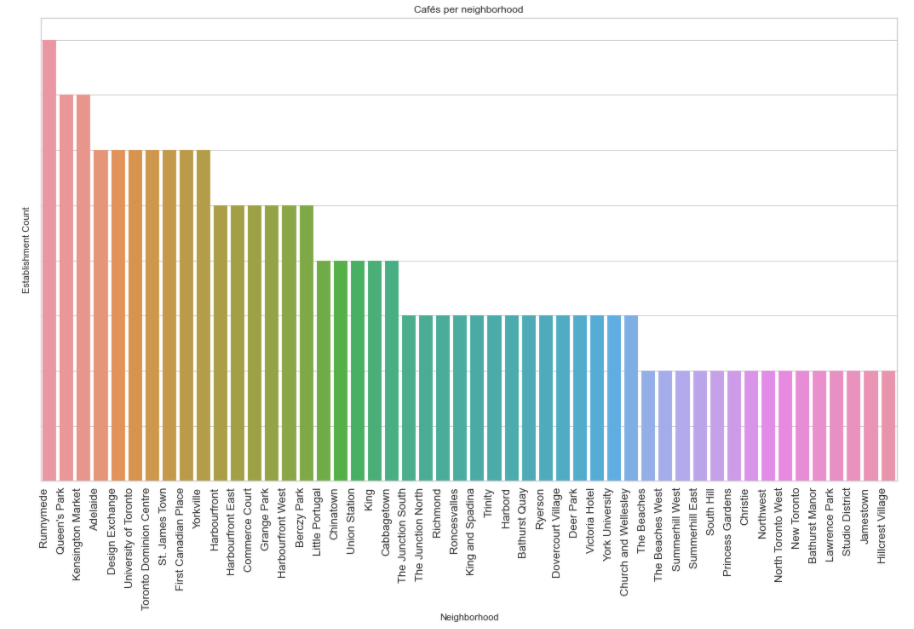
Inspecting the clusters (0-4), as we are looking to enter areas packed with Coffee Shops, Cafés, or Tea Rooms, as they complement our products, and with minimal competition, Cluster 3 is of interest.

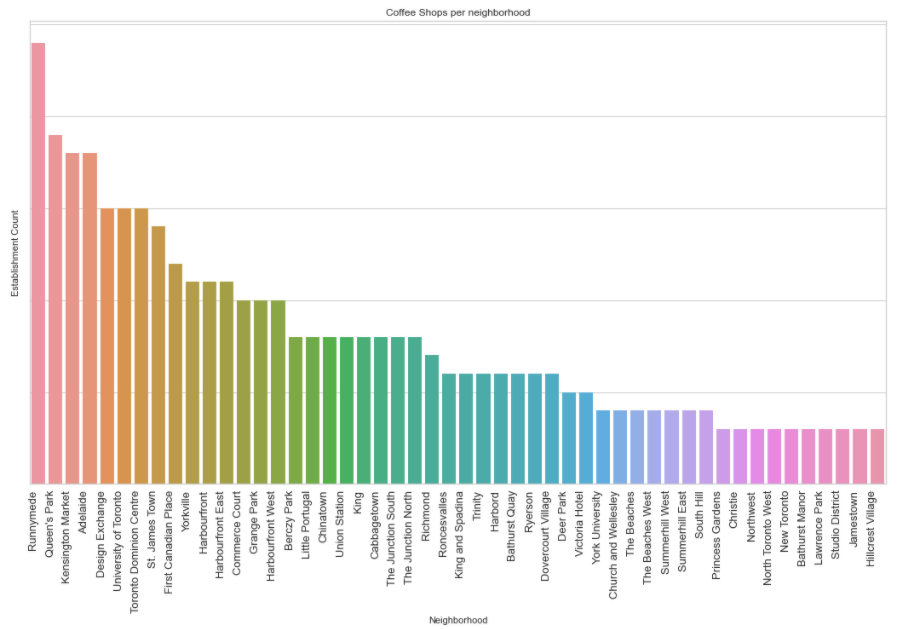


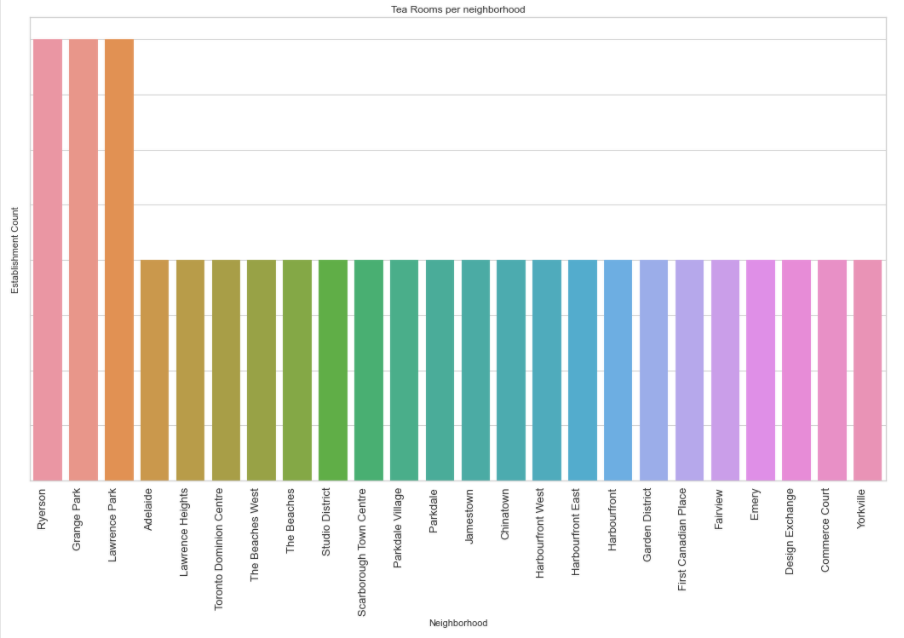
Within Cluster 3, we would like to prioritize neighborhoods that are densely populated with coffee shops, and with minimal competition: Oriole, Queen's Park, East Toronto, Mount Dennis, North Park, St Phillips, Martin Grove Gardens, Scarborough Village, Scarborough Village West, The Queensway East, Jamestown.

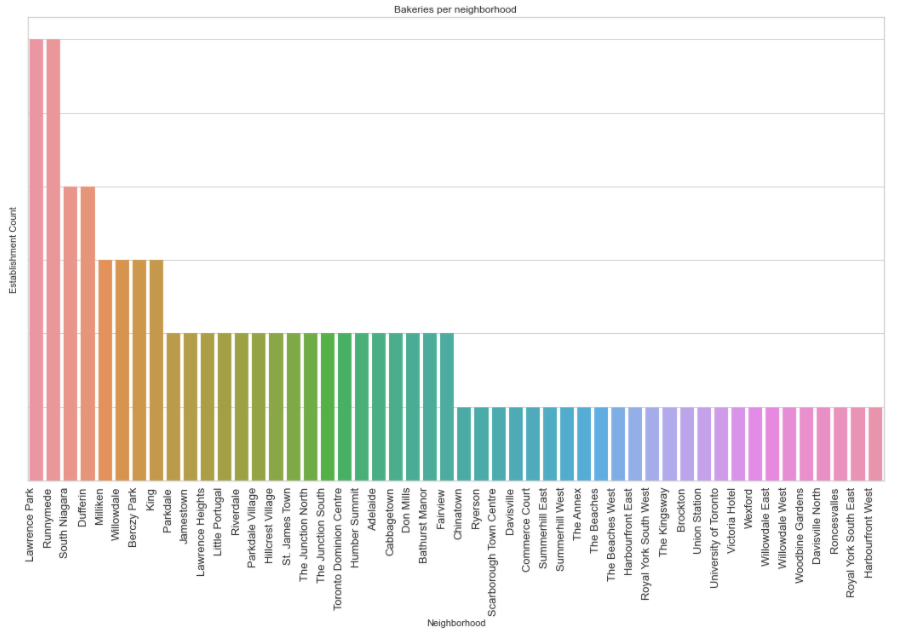


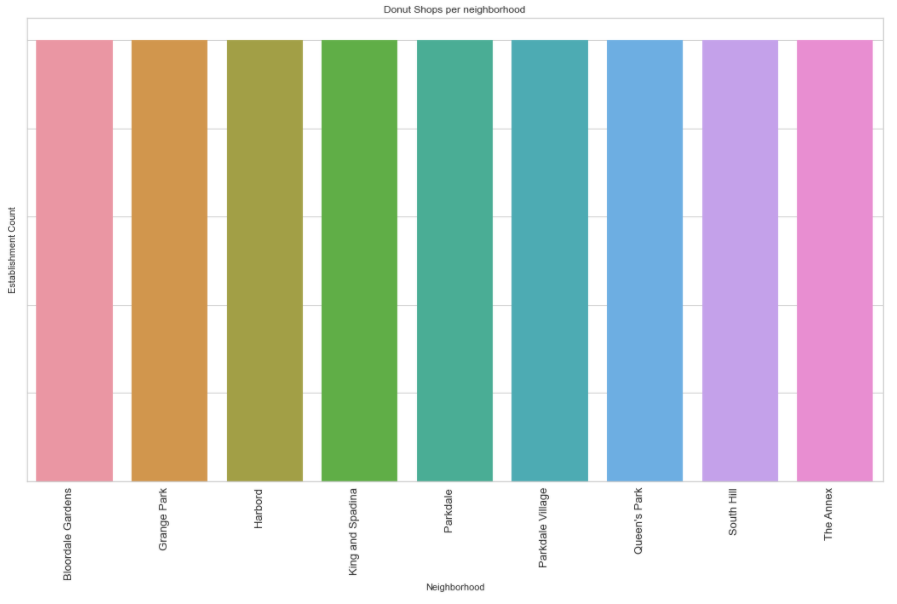
In the following graphs, we note that Runnymede (Cluster 0) has a high presence of Cafés, Coffee Shops, but also Bakeries.

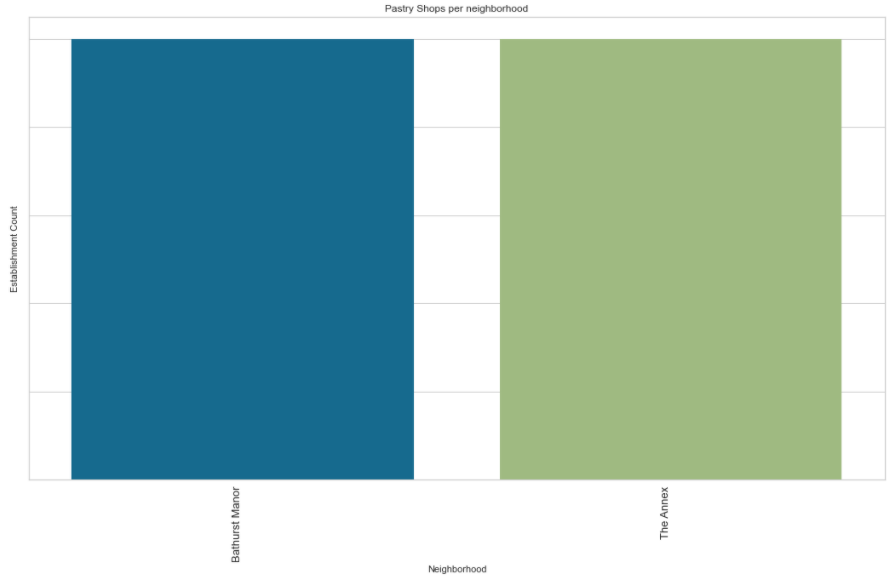












Lastly, we note that the volume of donut shops and pastry shops is much lower than the other types of establishments in this study.

1. **Recommendation**

We are looking to enter areas packed with Coffee Shops, Cafés, or Tea Rooms, as they complement our products, where competition is minimal, as our brand is a new entrant in the market. Given this backdrop, Cluster 3 is of interest.

Within Cluster 3, we would like to prioritize neighborhoods that are densely populated with coffee shops and pose little competition. The neighborhoods are: Oriole, Queen's Park, East Toronto, Mount Dennis, North Park, St Phillips, Martin Grove Gardens, Scarborough Village, Scarborough Village West, The Queensway East, Jamestown.

We note that Runnymede (Cluster 0) has a high presence of Cafés, Coffee Shops, but also Bakeries. While this neighborhood is of a lesser priority as of this analysis, one would be prudent to familiarize one’s self with the scene at this neighborhood for two reasons.

First, competition may come from establishments here seeking to expand elsewhere. As such, we should take steps to prevent or challenge such moves, should they choose to expand to our locations of interest.

Second, it would be prudent to earmark this location for further expansion later when the brand is more established and able to operate in a more competitive environment.

Moreover, as there is a notably low volume of donut shops and pastry shops, additional confirmation is required, for example via a visual inspection of the recommended neighborhoods, to validate the situation as the data suggests.

For further analysis, as the accuracy of the models has room for improvement, consider augmenting the study with more data from other sources, such as store ratings, a richer list of establishments and locations, foot traffic or population density.