



Investing in the Toronto Hotel Industry

Applied Data Science Capstone

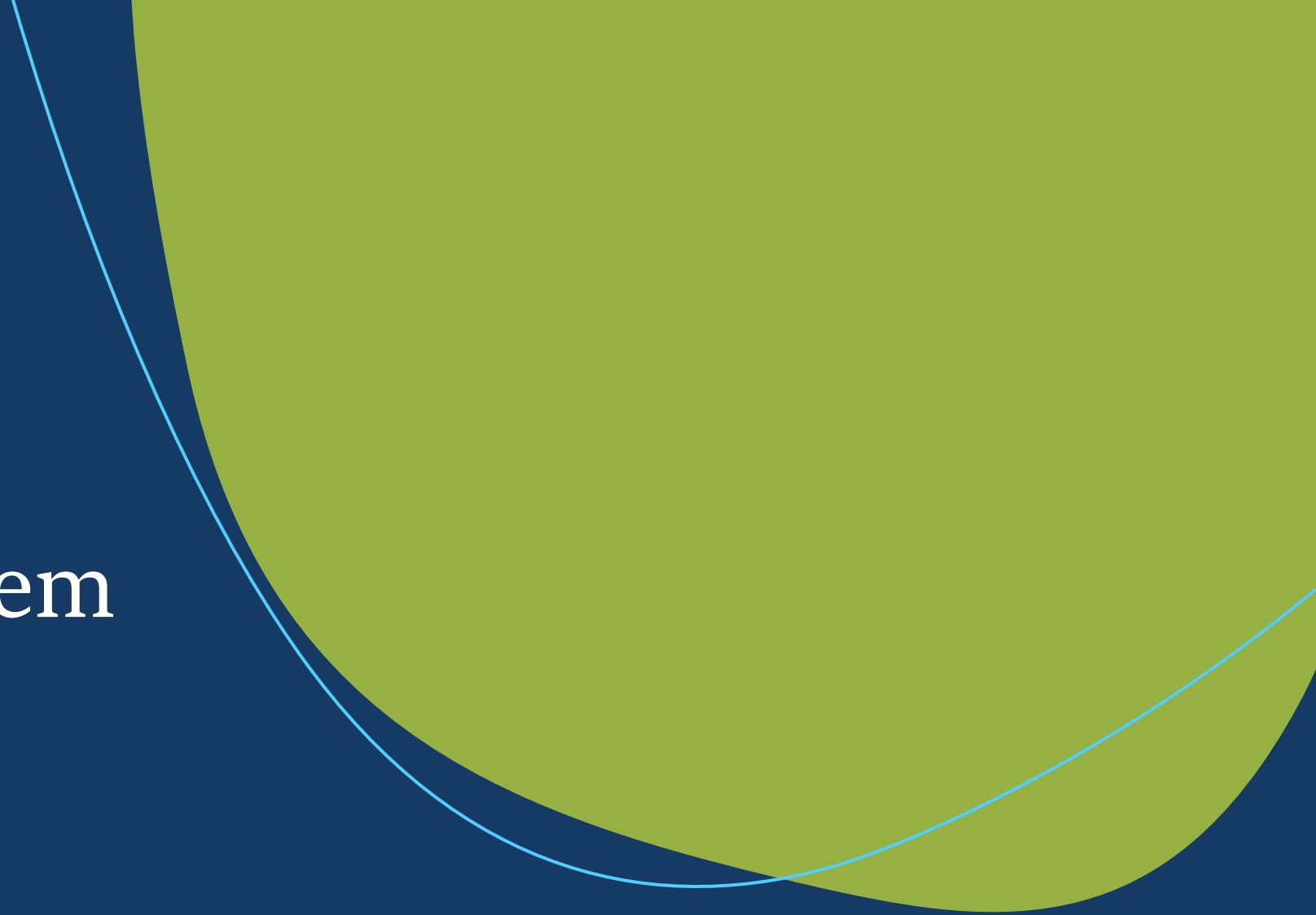


Summary

The goal of this project was to examine potential regions for investment in a hotel in Toronto, taking into consideration the number of hotels and the diversity of top venues in the neighborhood .

- The areas with the fewest hotels are: Bathurst Manor, Wilson Heights, Downsview North; Caledonia-Fairbanks; Davisville; High Park, The Junction South; Wexford, Maryvale.
- Neighborhoods in cluster label 2 (solid blue) has most diversity in venues.
- Top venues in Cluster 2 neighborhoods include coffee shops, restaurants, and many other attractions. Cluster 2 appears to be convenient spots for tourists (and thus hotels).
- Thus, hotel investments would be neighborhoods with 1) few hotels, and 2) in Cluster 2.
- Final areas of potential hotel investment: *Bathurst Manor, Wilson Heights, Downsview North; Davisville; High Park, The Junction South; and Wexford and Maryvale*

Business Problem

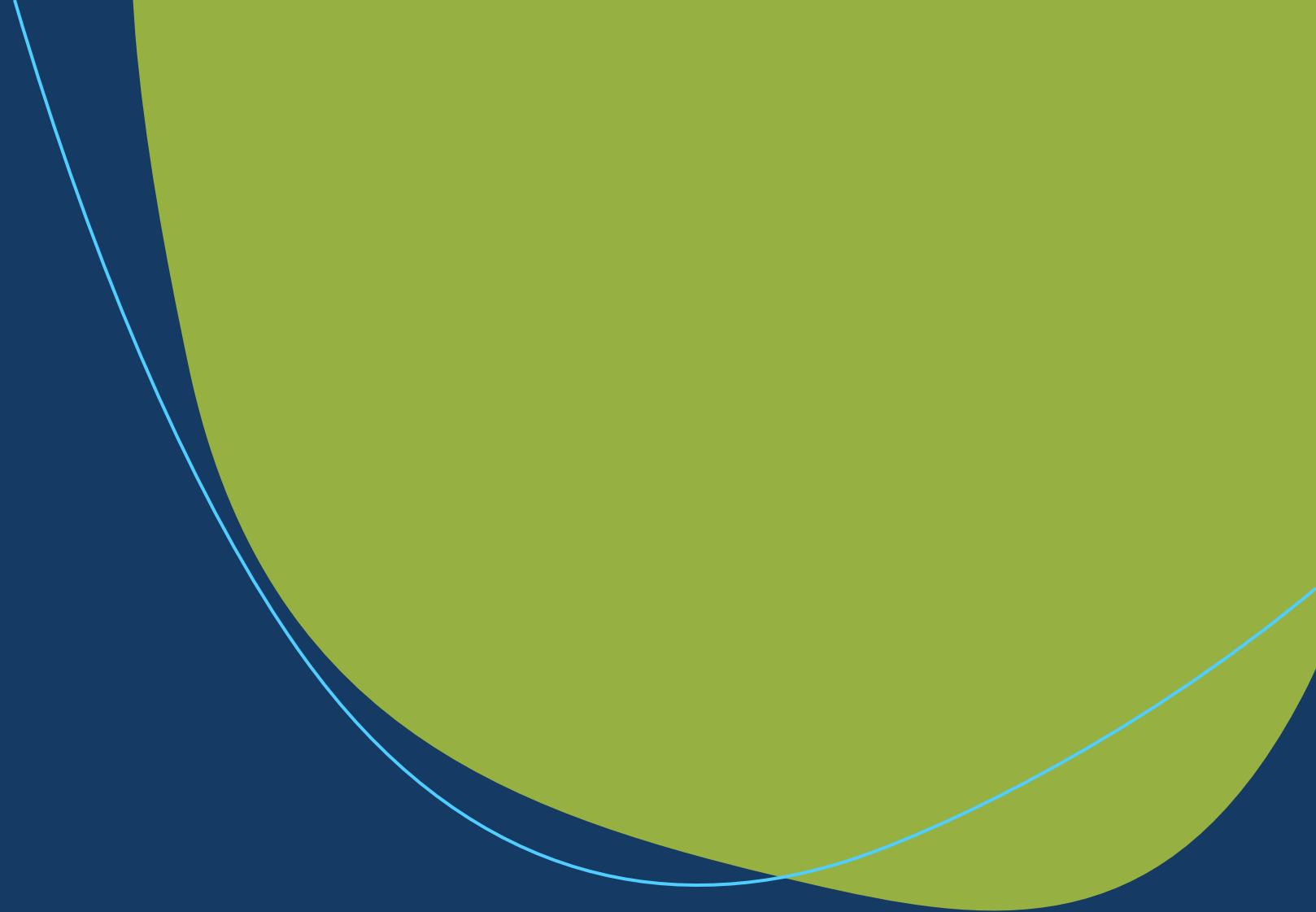




Business Problem

- Goal: identify where are the most optimal neighborhoods to buy or build a hotel
- Qualifications: a region with 1) low number of hotels AND 2) high diversity in top venues and attractions in the region.
- *Intended audience:* For investors and developers who are interested in gaining a deeper view of where optimal locations for might be starting up a hotel business, especially in the context of surrounding venues and attractions.

The Data



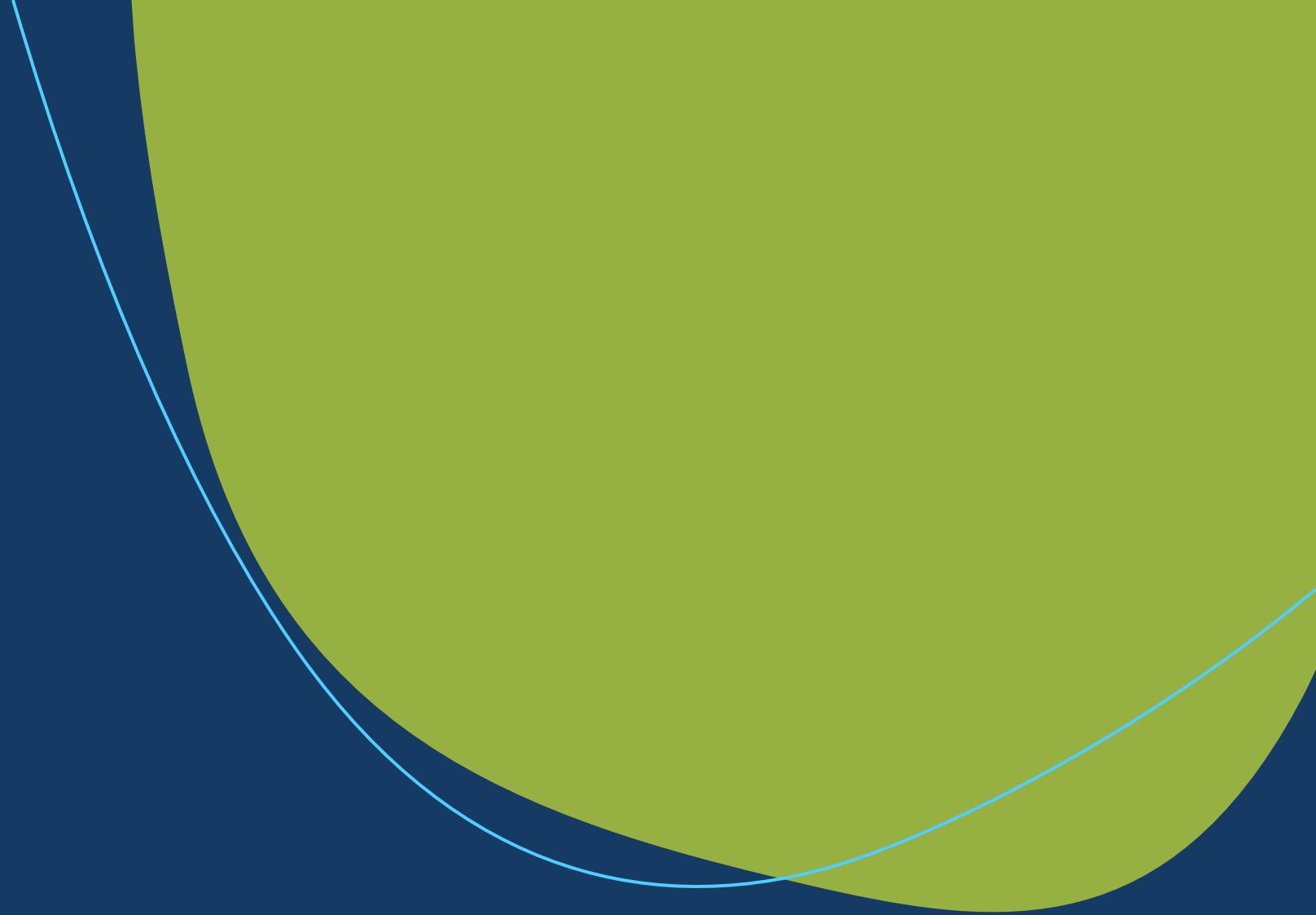
Data Sources

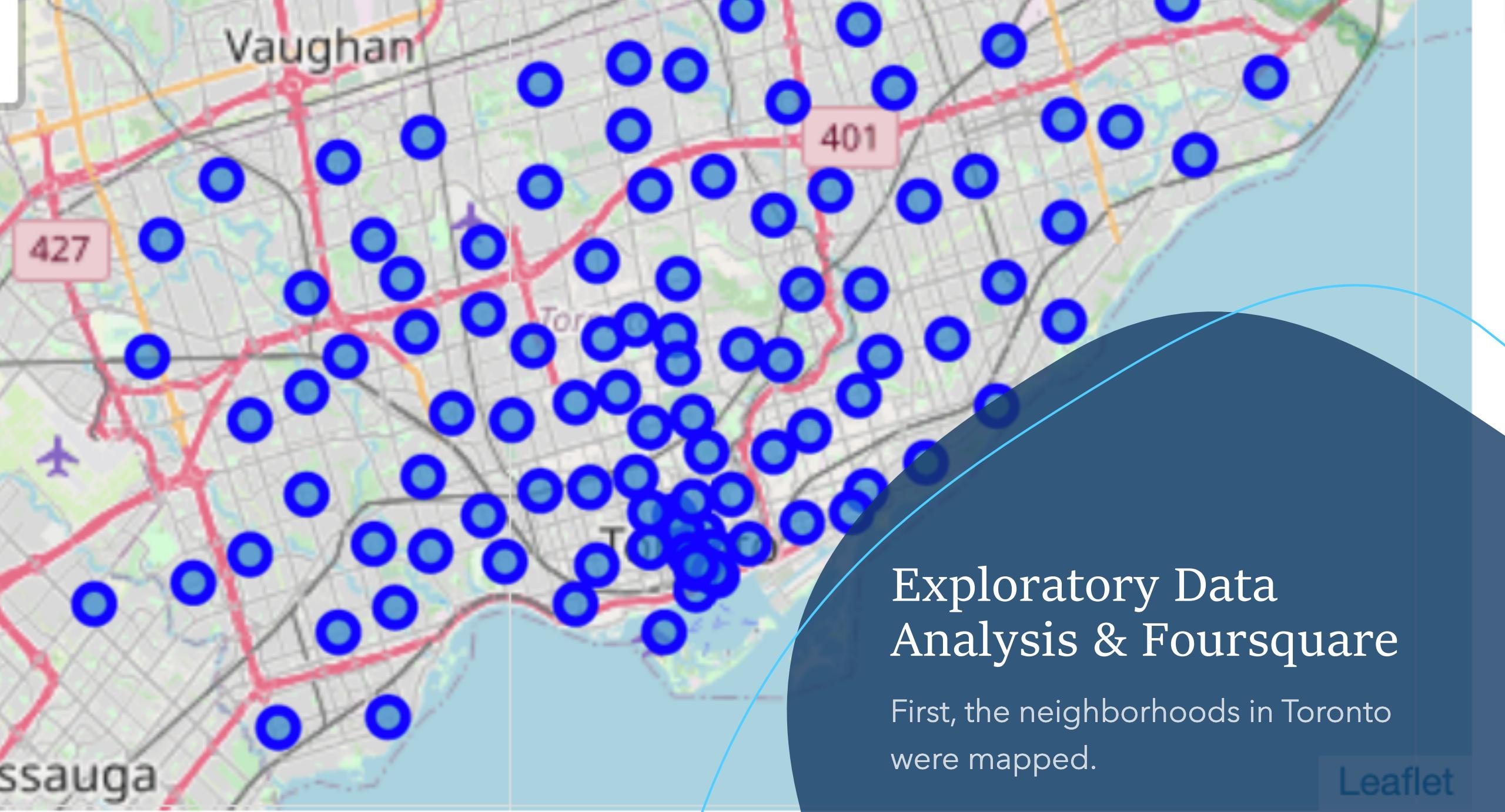
- Neighborhood and borough data: from [Wikipedia page](#)
- Data on common venues: from Foursquare API
- Longitude and latitude data: from Geocoder Python package & https://cocl.us/Geospatial_data

Data Cleaning

- After scraping the Wikipedia list for postal codes and neighborhoods in the Toronto area, the data was cleaned by dropping blank rows.
- The resulting data has 103 rows and 3 columns.
- Geospatial data (longitude, latitude) was pulled from GeoCoder and https://cocl.us/Geospatial_data and then joined to the previous dataframe
- This data is the basis of the neighborhood information for the subsequent analysis.

Methodology





Foursquare API

- Then, using Foursquare's API, and the list of geo data, a search for the top 100 venues in a radius of 600m was done for each neighborhood via Foursquare and the 'explore' function.
- This was repeated for all neighborhoods in Toronto, with the resulting dataframe having 2131 venues and 7 columns.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Parkwoods	43.753259	-79.329656	Brookbanks Park	43.751976	-79.332140	Park
1	Parkwoods	43.753259	-79.329656	Variety Store	43.751974	-79.333114	Food & Drink Shop
2	Parkwoods	43.753259	-79.329656	TTC stop - 44 Valley Woods	43.755402	-79.333741	Bus Stop
3	Parkwoods	43.753259	-79.329656	Bella Vita Catering & Private Chef Service	43.756651	-79.331524	BBQ Joint
4	Victoria Village	43.725882	-79.315572	Victoria Village Arena	43.723481	-79.315635	Hockey Arena

One hot encoding

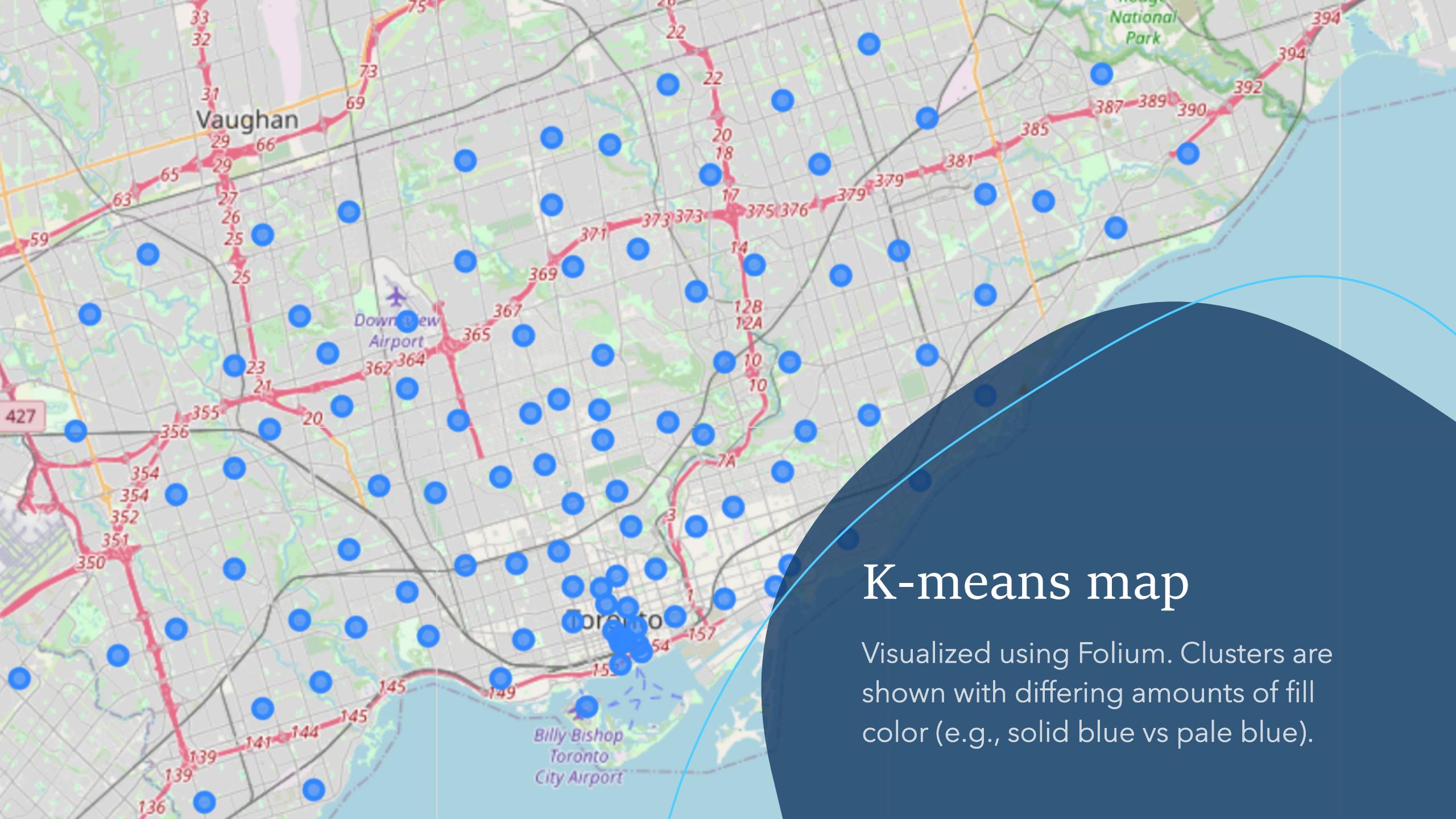
- Each neighborhood was then analyzed using one hot encoding to then determine the most common 10 venues in each area.
- Goal: identify areas with most diversity in type of venues, or which have venues that would be most of interest to tourists (e.g., restaurants, attractions, etc.).
- This would make them good opportunities for hotel investment.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Aigincourt	Lounge	Skating Rink	Breakfast Spot	Latin American Restaurant	Electronics Store	Eastern European Restaurant	Dumpling Restaurant	Drugstore	Ethiopian Restaurant	Donut Shop
1	Alderwood, Long Branch	Pizza Place	Gym	Coffee Shop	Skating Rink	Pharmacy	Pub	Sandwich Place	Dim Sum Restaurant	Deli / Bodega	Department Store
2	Bathurst Manor, Wilson Heights, Downsview North	Coffee Shop	Bank	Middle Eastern Restaurant	Pizza Place	Supermarket	Sushi Restaurant	Deli / Bodega	Shopping Mall	Restaurant	Mobile Phone Shop
3	Bayview Village	Japanese Restaurant	Café	Bank	Chinese Restaurant	Dessert Shop	Diner	Discount Store	Distribution Center	Dog Run	Women's Store
4	Bedford Park, Lawrence Manor East	Restaurant	Coffee Shop	Italian Restaurant	Sandwich Place	Juice Bar	Butcher	Sushi Restaurant	Pizza Place	Pharmacy	Indian Restaurant

K-means clustering

- Used to group neighborhoods by type of venues to facilitate the search for an area suitable for hotel investment.
- $k = 5$ was chosen as it was the “kink” identified the elbow method, making it the optimal k value.

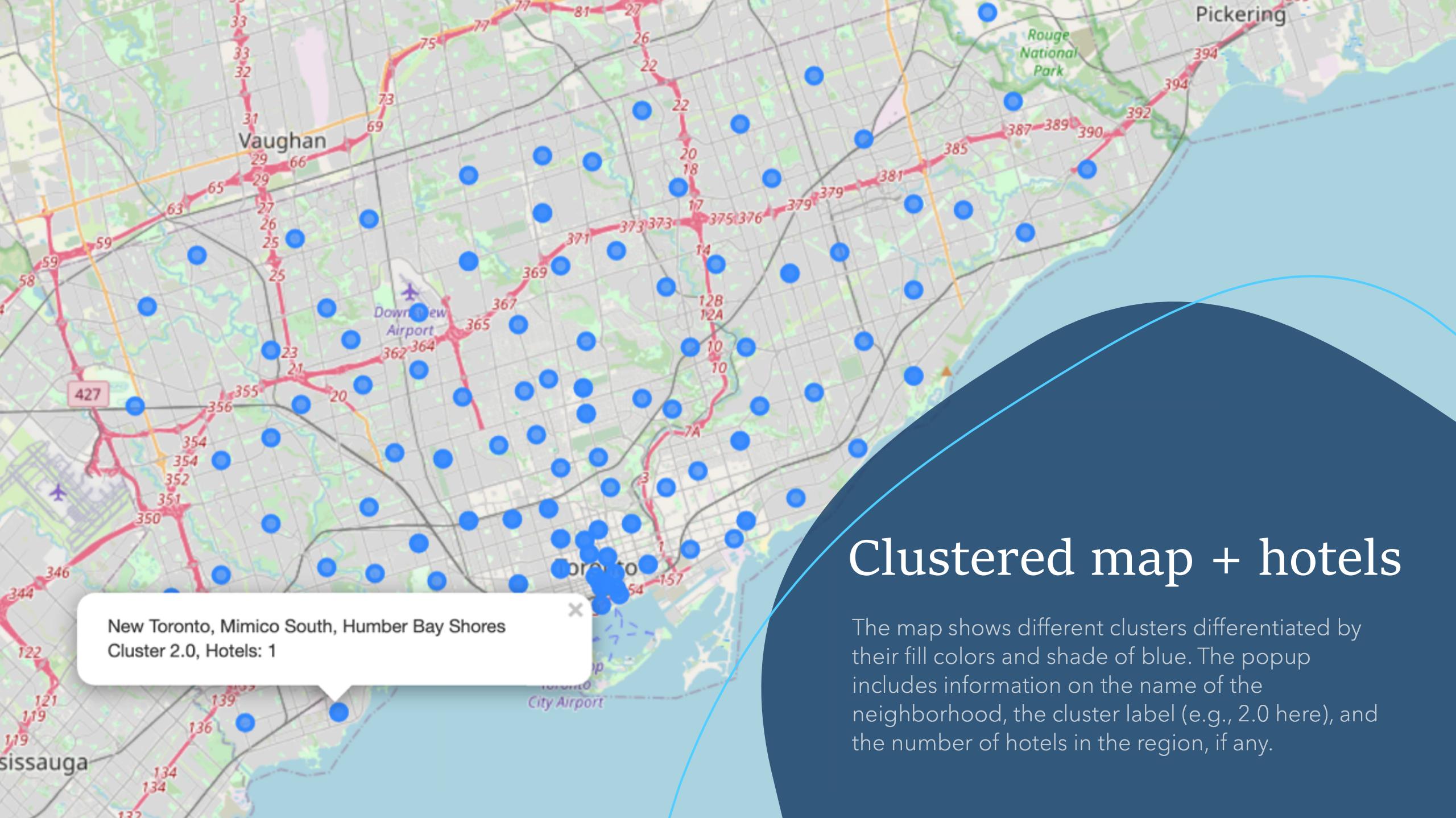
	Postal Code	Borough	Neighbourhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd M Com Ve
0	M3A	North York	Parkwoods	43.753259	-79.329656	2.0	Park	BBQ Joint	Food Drink S
1	M4A	North York	Victoria Village	43.725882	-79.315572	2.0	Portuguese Restaurant	Financial or Legal Service	Intersec
2	M5A	Downtown Toronto	Regent Park, Harbourfront	43.654260	-79.360636	2.0	Coffee Shop	Park	Bal
3	M6A	North York	Lawrence Manor, Lawrence Heights	43.718518	-79.464763	2.0	Clothing Store	Furniture / Home Store	Accesso S
4	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government	43.662301	-79.389494	2.0	Coffee Shop	Diner	Distribu Ce



Number of hotels

- Using Foursquare API again with the “categoryId” term, only the hotels within radius = 500m of the set neighborhood long/lat points were returned; limit = 100
- This was repeated for all neighborhoods in Toronto and the dataframe then cleaned.

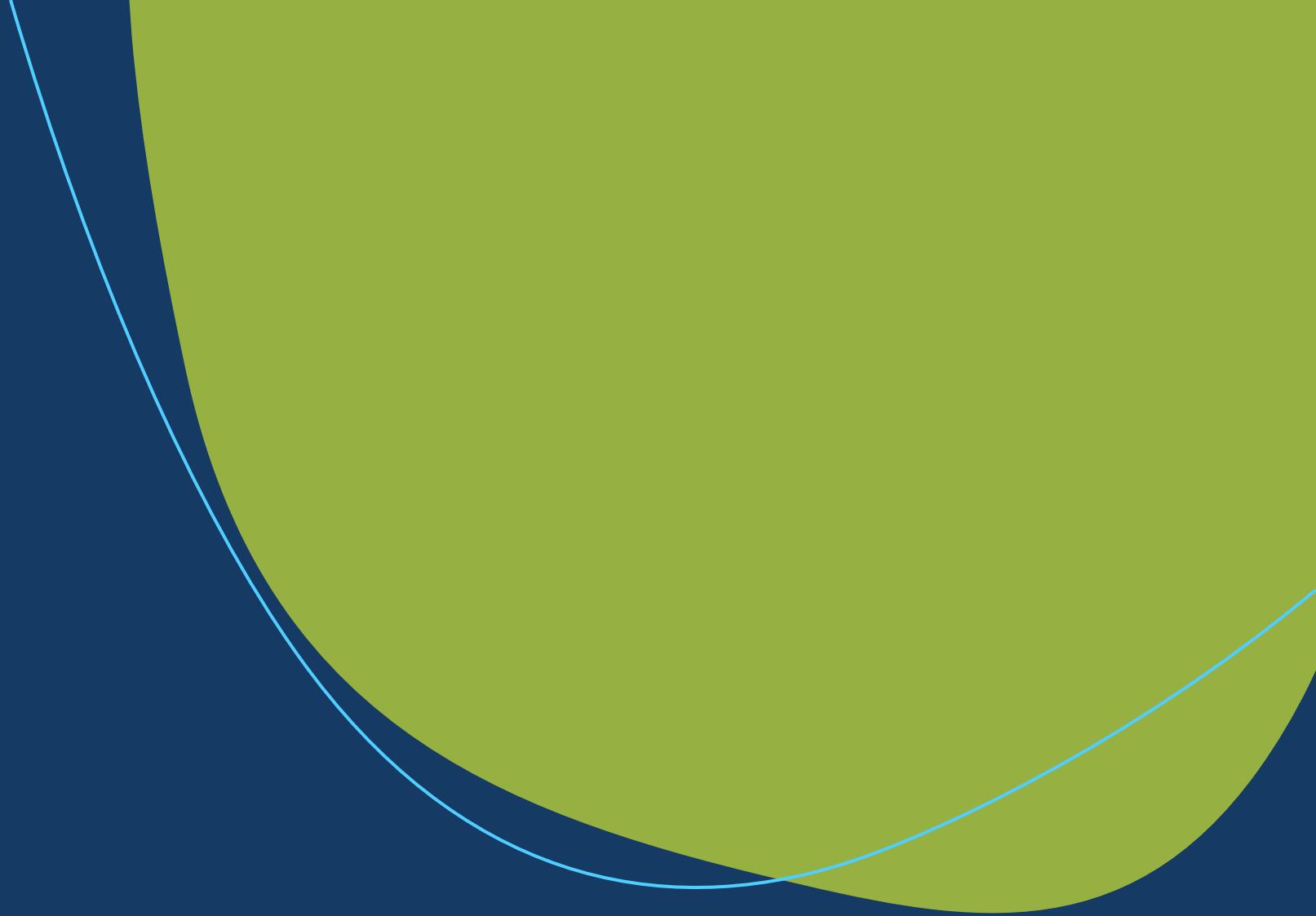
Neighborhood	Number of Category: Hotels
Bathurst Manor, Wilson Heights, Downsview North	1
Berczy Park	8
Brockton, Parkdale Village, Exhibition Place	2
Caledonia-Fairbanks	1
Canada Post Gateway Processing Centre	4
Central Bay Street	21
Christie	2
Church and Wellesley	21
Cliffside, Cliffcrest, Scarborough Village West	2
Commerce Court, Victoria Hotel	37
Davisville	1
Davisville North	2
Dufferin, Dovercourt Village	3
First Canadian Place, Underground city	41
Garden District, Ryerson	41



Clustered map + hotels

The map shows different clusters differentiated by their fill colors and shade of blue. The popup includes information on the name of the neighborhood, the cluster label (e.g., 2.0 here), and the number of hotels in the region, if any.

Results





Results: Number of Hotels

Areas with the least number of hotels:

- Bathurst Manor, Wilson Heights, Downsview North, 1 hotel
- Caledonia-Fairbanks, 1 hotel
- Davisville, 1 hotel
- High Park, The Junction South, 1 hotel
- Wexford, Maryvale, 1 hotel

Results: Diversity of venues

- Cluster label 2 (solid blue dot) appears to have areas with the most diversity in venues.
- Top venues in Cluster 2: coffee shops, restaurants, and many more.
- Thus areas in cluster label 2 would be good spots for tourists to visit (and thus a good spot to build a hotel).



Results: Overlapping neighborhoods

Areas where “least number of hotels” overlaps with “cluster label 2” (aka, high diversity of surrounding venues):

- Bathurst Manor, Wilson Heights, Downsview North, 1 hotel (cluster label 2)
- Davisville, 1 hotel (cluster label 2)
- High Park, The Junction South, 1 hotel (cluster label 2)
- Wexford, Maryvale, 1 hotel (cluster label 2)

Discussion & Future Direction



Discussion & Recommendations

- Recommended areas for hotel investment are: Bathurst Manor, Wilson Heights, Downsview North; Davisville; High Park, The Junction South; and Wexford and Maryvale .
- They are all in cluster label 2 and thus have the higher diversity of venues and attractions, and all have the least number of hotels (hotels = 1)
- Thus, there is opportunity for travellers to be attracted to the location, AND there is little competition for hotels.

Future Direction

- Examining financial returns on investment and cost of real estate by neighborhood
- Considering number of Airbnbs in neighborhood as competitor
- Traveller demand and knowledge of travel to specific neighborhoods