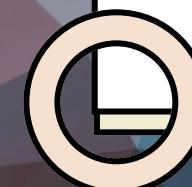




FINDING LOCATIONS FOR A COLOMBIAN RESTAURAN T IN TORONTO

BY JULIAN GOMEZ





Background

- “Diversity is our strength” is the official motto of the city of Toronto
 - The city stands apart from multicultural cities like New York, Los Angeles, or Singapore with approximately a 47% proportion of foreign-born residents.
 - The city also has the largest Latin American community in Canada, where a lot come from Colombia.
 - People from Colombia move to Canada to pursue a better quality of life, including job opportunities, education, and starting businesses, including opening a restaurant.



Problem and Interest

- With 140 neighborhoods to choose from within the city of Toronto, there is data to help narrow down to five neighborhoods to open a Colombian restaurant.
 - We use graphics and statistical analysis on city demographics, neighborhood profiles, and venue data to select the best locations for the restaurant.
- Any immigrants from Colombia looking to share the culture with everyone living and visiting Toronto would be interested in looking for a restaurant location before choosing a property and securing funding.
 - Any existing Colombian restaurant owners considering relocating for any reason may also be interested in neighborhood analysis to improve their brand.





Data Acquisition and Cleaning

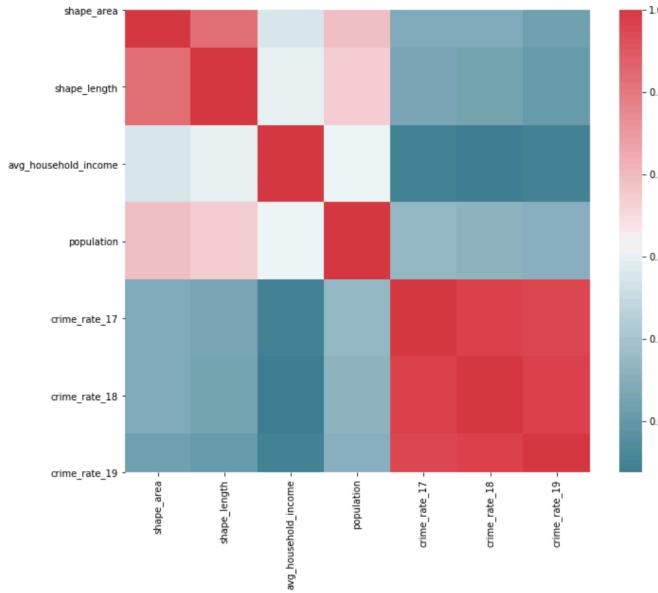
- The project analysis consists in gathering four datasets involving all Toronto neighborhoods and combine them into one.
 - Toronto neighborhood profiles - neighborhood data from the City of Toronto's Open Data Portal
 - 16 variables, 140 observations (neighborhoods)
 - Dropped eight columns - geometry, parent area ID, object ID, etc.
 - Variables - area code, latitude, longitude, area (sq. m), shape length
 - Average Household Income Data - 2016 Toronto Neighborhood Census
 - 1 variable, 140 observations
 - Filtered the excel file to Income topic
 - Joined with the neighborhood profile data
 - Toronto Crime Dataset - Toronto Police Service Portal
 - 61 variables, 140 observations
 - Variables include neighborhood, population, number of incidents for theft, breaking and entering, homicide, assault, ad crime rate change since 2014.
 - Calculated annual crime rates by dividing the total number of crime incidents and neighborhood population and multiplying by 1000 since the neighborhood populations are in thousands.
 - Kept neighborhood name, population, neighborhood ID, and calculated annual crime rates since 2016.
 - Joined to Toronto neighborhood profiles
 - Toronto venue data - Foursquare Places API
 - Sourced API for places in the city of Toronto
 - 7 variables, 5684 observations
 - Variables include neighborhood name, neighborhood coordinates, venue name, venue coordinates, and venue category
 - One duplicate deleted
 - Includes 351 different venue categories





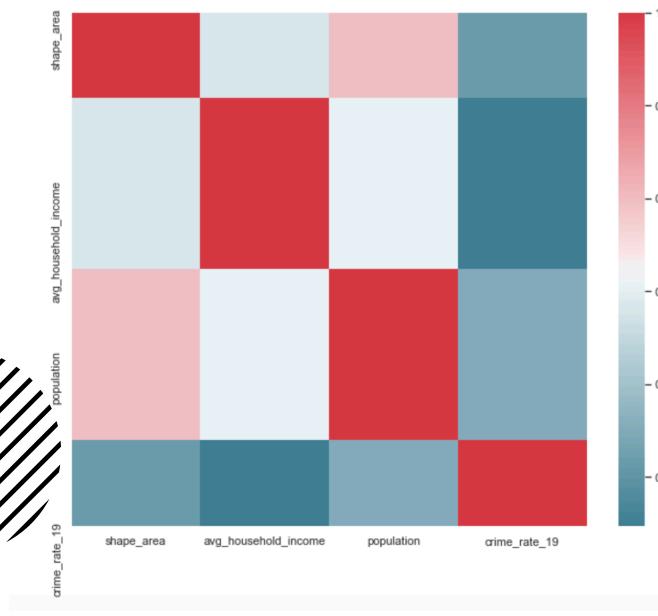
PART 3: EXPLORATORY DATA ANALYSIS

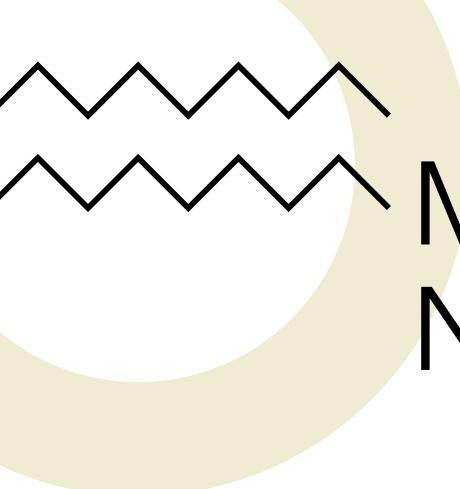




Correlations

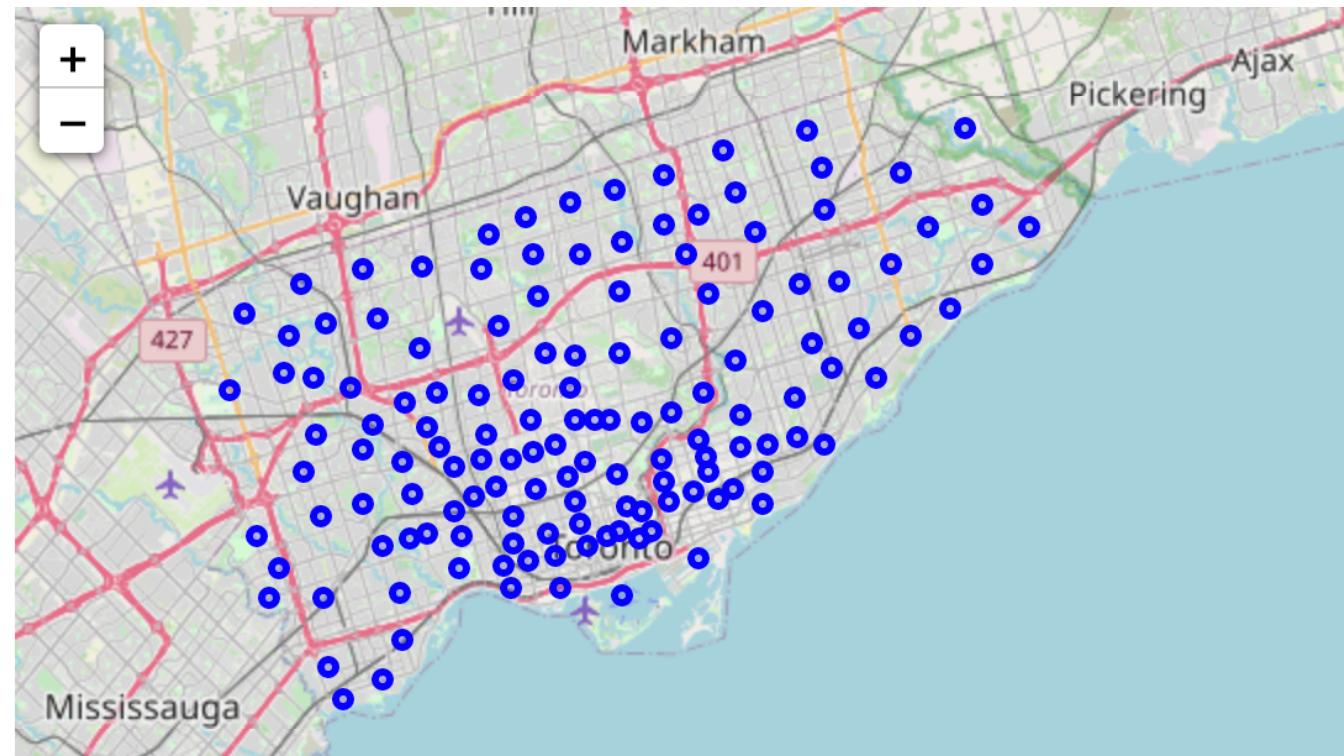
- The correlation plots from the combined Toronto dataset.
- Four out of the seven variables in the upper plot have high correlations.
- Used variance inflation factor (VIF) to find variables that yield a value over 10.
 - These variables cause multicollinearity and need to drop them from the dataset
- Dropped crime rates from 2017 to 2018 and shape length, resulting in the lower plot.
- Although shape area and population have a correlation of 0.6, the VIF for either of these variables is less than 10.





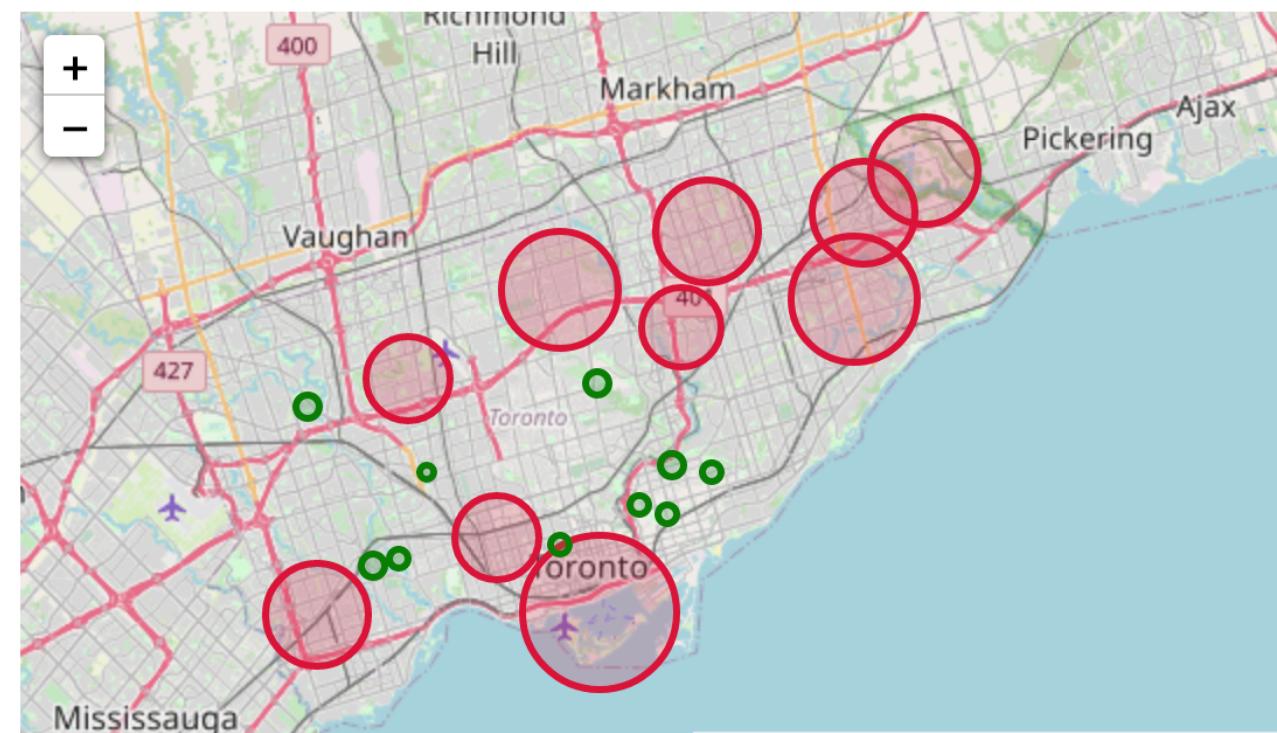
Mapping Neighborhoods

- Plot Neighborhoods with the Folium package
- All neighborhoods in the Toronto Area



Mapping Neighborhoods: Population

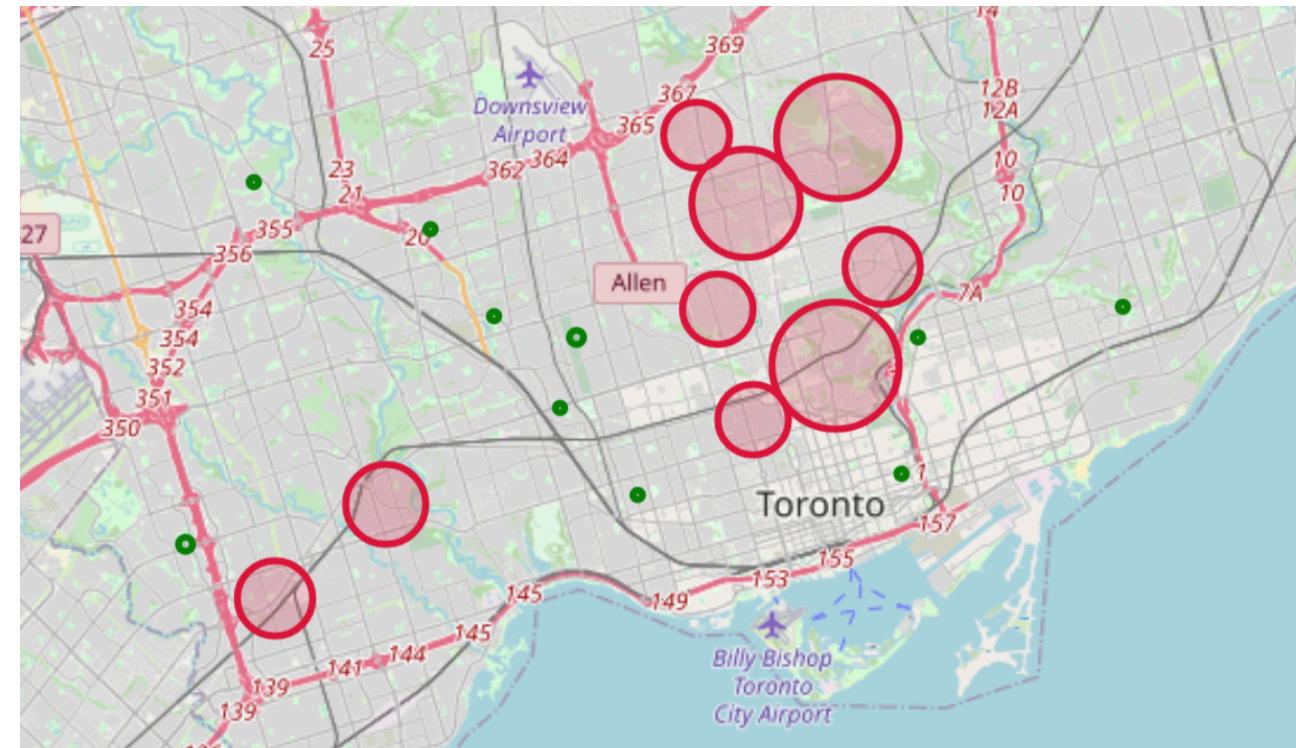
- Most Populated Neighborhoods (from greatest)
 - Waterfront Communities - the Island
 - Woburn
 - Willowdale East
 - Rouge
 - L'Amoreux
 - Islington-City Centre West
 - Malvern
 - Dovercourt-Wallace Emerson-Junction
 - Downsview-Roding-CBF
 - Parkwoods-Donalda
- Least Populated Neighborhoods (from smallest)
 - Beechborough-Greenbrook
 - University
 - Blake Jones
 - Playter Estates - Danforth
 - Woodbine-Lumsden
 - Lambton Baby Point
 - Old East York
 - Bridle Path - Sunnybrook - York Mills
 - Kingsway South
 - Elms-Old Rexdale





Mapping Neighborhoods: Household Income

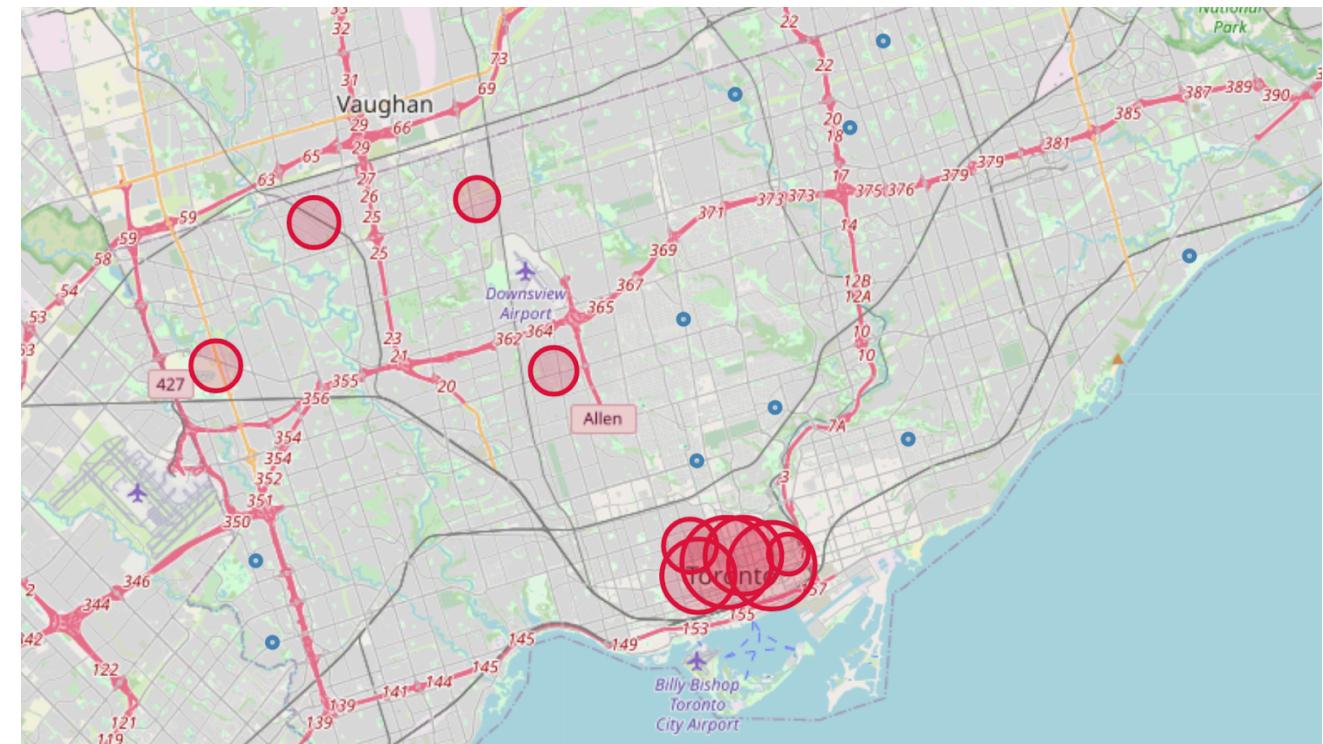
- Neighborhoods with the highest average household income
 - Rosedale-Moore Park
 - Bride Path-Sunnybrook-York Mills
 - Lawrence Park South
 - Kingsway South
 - Islington-City Centre West
 - Leaside-Bennington
 - Annex
 - Forest Hill South
 - Rouge
 - Bedford Park-Nortown
- Neighborhoods with the lowest average household income
 - Taylor-Massey
 - Regent Park
 - Beechborough-Greenbrook
 - Rustic
 - Elms-Old Rexdale
 - Broadview North
 - Weston-Pellam Park
 - Dufferin Grove
 - Etobicoke West Mall
 - Caledonia-Fairbank





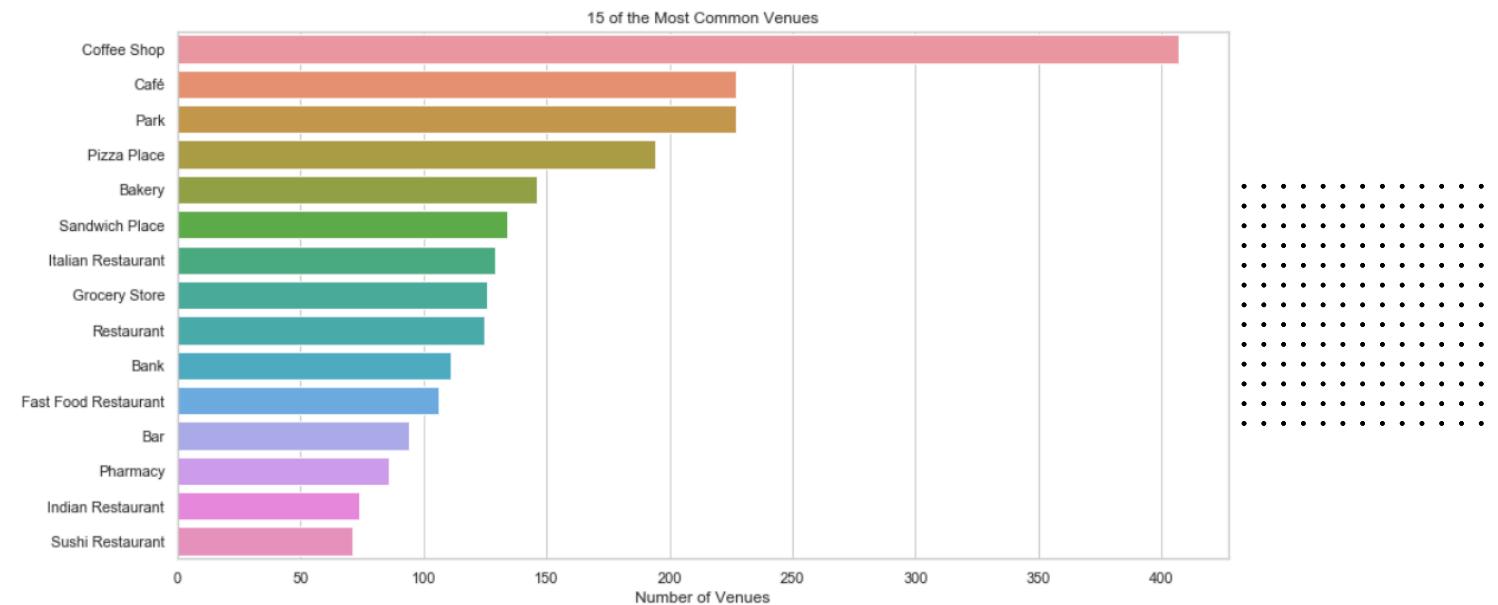
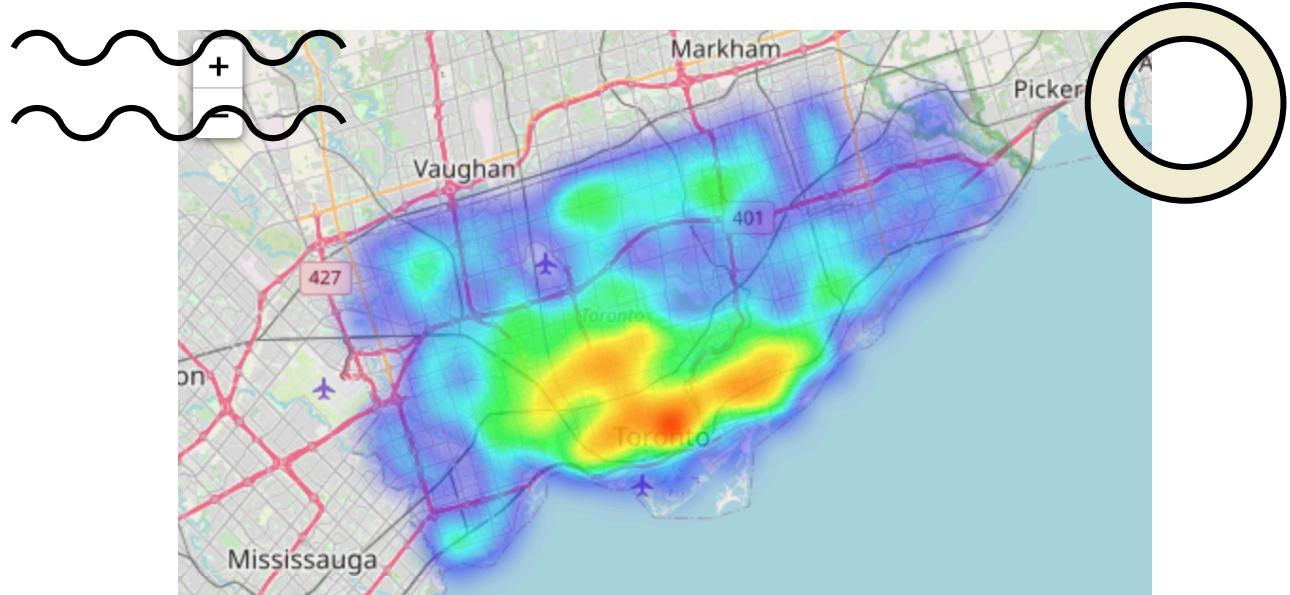
Mapping Neighborhoods: Neighborhoods Crime Rate in 2019

- The Neighborhoods with the highest crime rate
 - Bay Street Corridor
 - Moss Park
 - Church-Yonge Corridor
 - Keningston-Chinatown
 - University
 - West Humber-Clairville
 - Humber Summit
 - Yorkdale-Glen Park
 - York University Heights
 - Regent Park
- Neighborhoods with the lowest crime rate
 - Guildwood
 - Lawrence Park North
 - Pleasant View
 - Steeles
 - Markland Wood
 - Yonge St. Clair
 - Eringate-Centennial-West Deane
 - Leaside-Bennington
 - Bayview Woods-Steeles
 - Woodbine-Lumsden



Mapping Toronto Venues

- Toronto has the most venue density in the downtown area.
- Areas farther away from downtown have low venue density.
- The most common venues include coffee shops, cafes, parks, and pizza places.





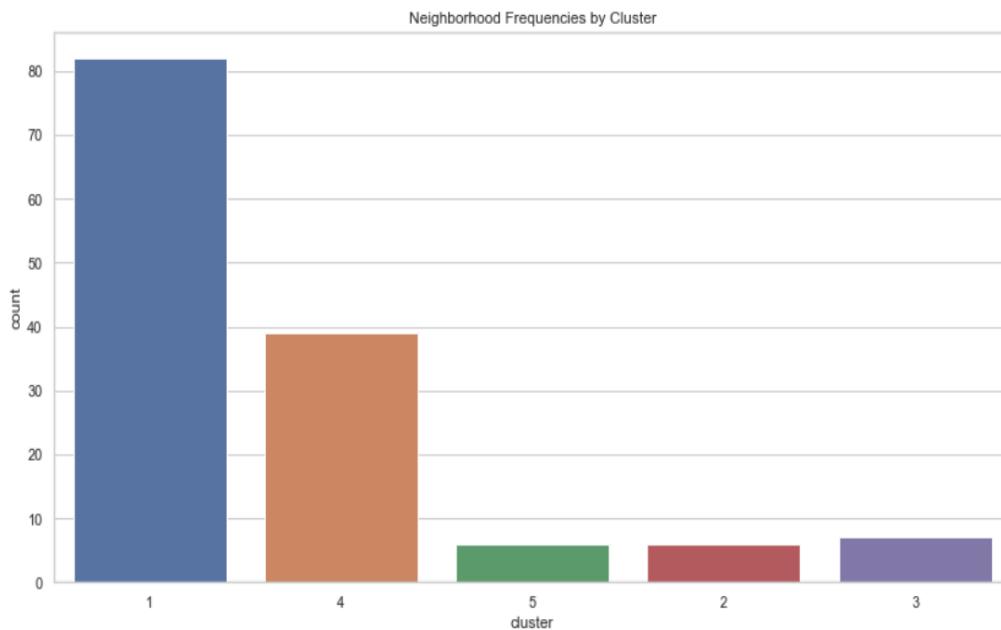
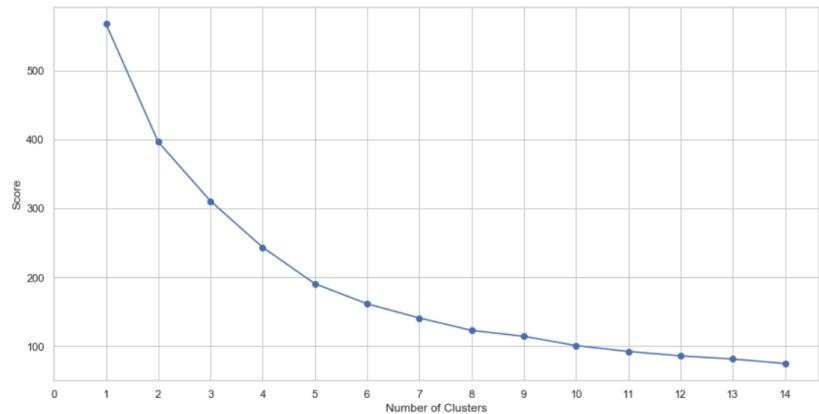
PART 4: MODELING





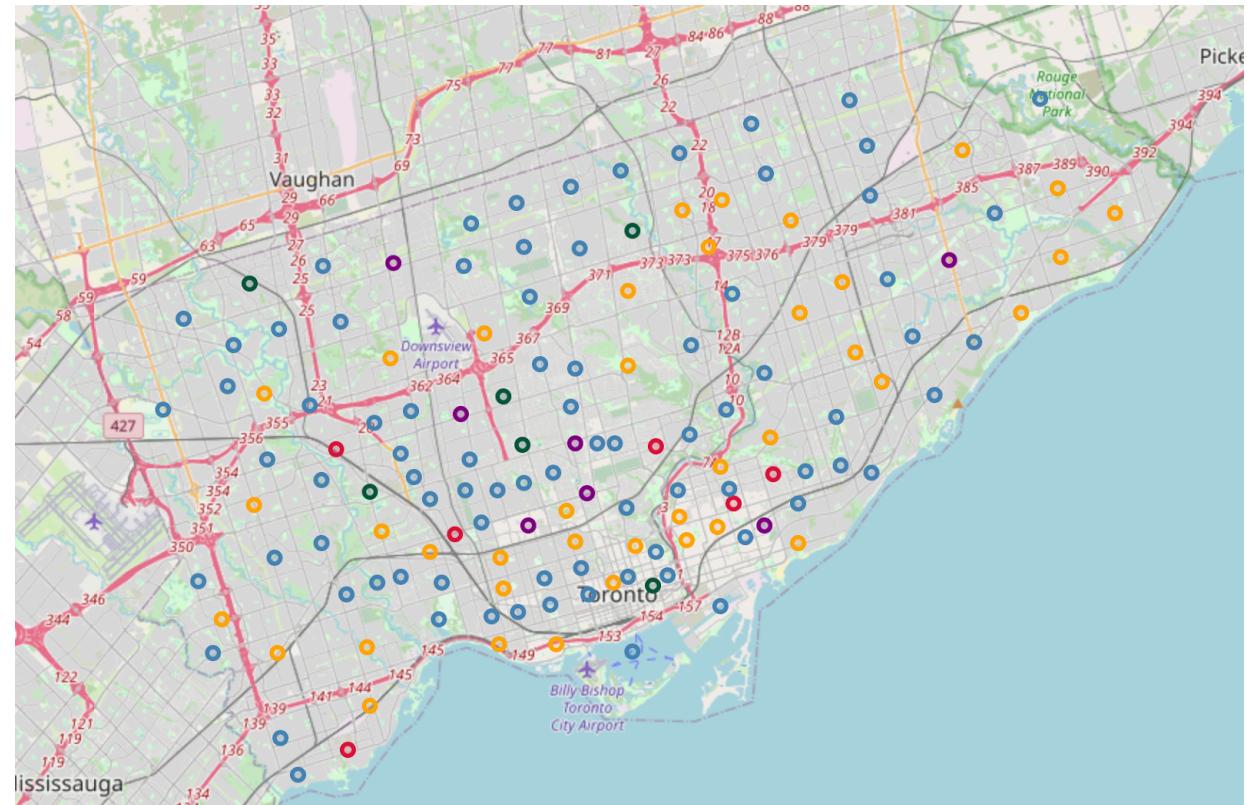
K-Means Clustering

- Partitions data into K distinct, non-overlapping segments.
 - Specify a desired number of clusters K to assign data to one of the K clusters.
 - Identifying the optimal value of K is a fundamental step.
 - Standardized demographics data and grouped neighborhoods by proportion of every venue category .
 - The scree plot shows an elbow over the optimal number of clusters
 - Result: 5 clusters
- Cluster indexed 1-5 in analysis instead of 0-4 in Python
- Cluster 1 has the most neighborhoods while 2 and 5 have the least

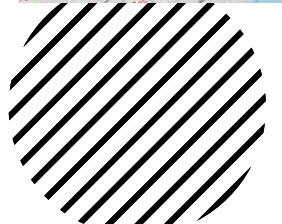
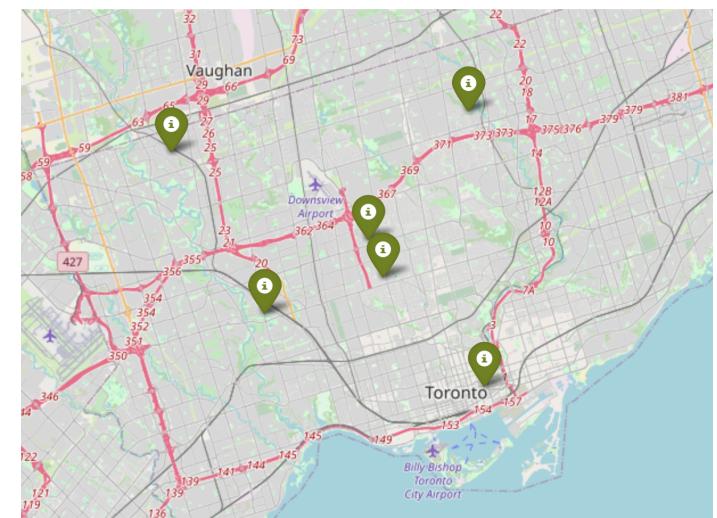
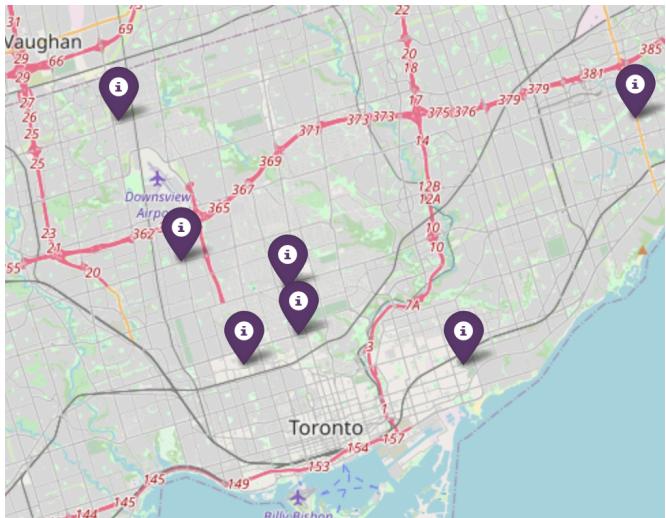
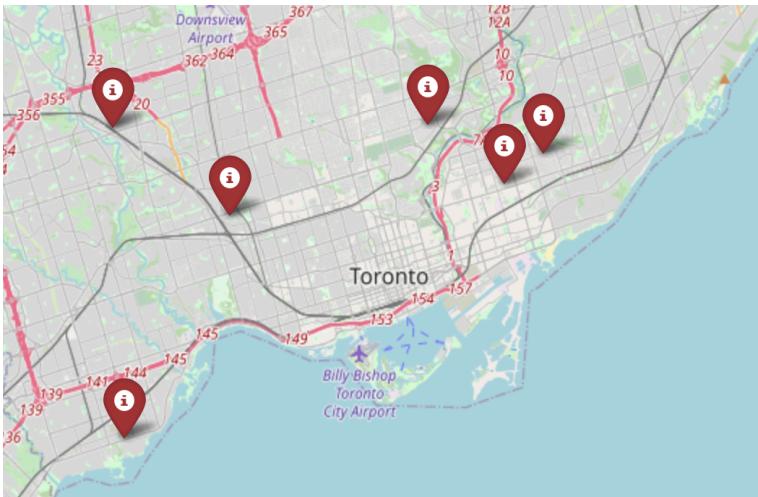
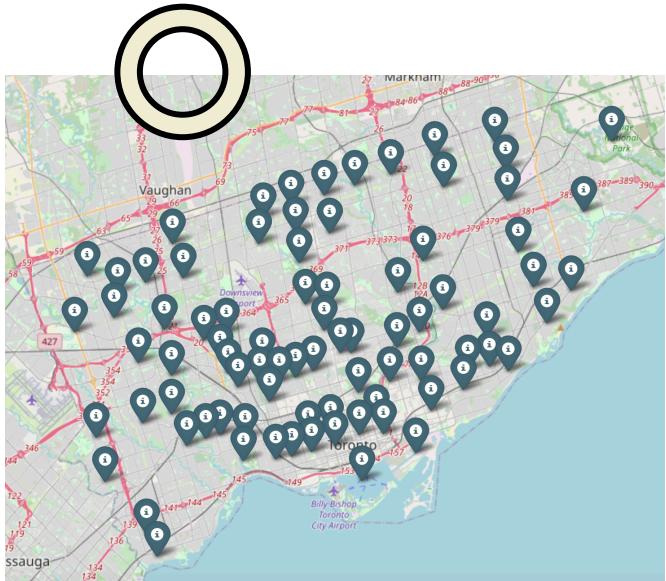


Cluster Maps

- Cluster 1 is in blue, cluster 2 in red, cluster 3 in purple, 4 in orange, and 5 in green.
- Assume that the best neighborhood(s) for the restaurant are in clusters 1 and 4

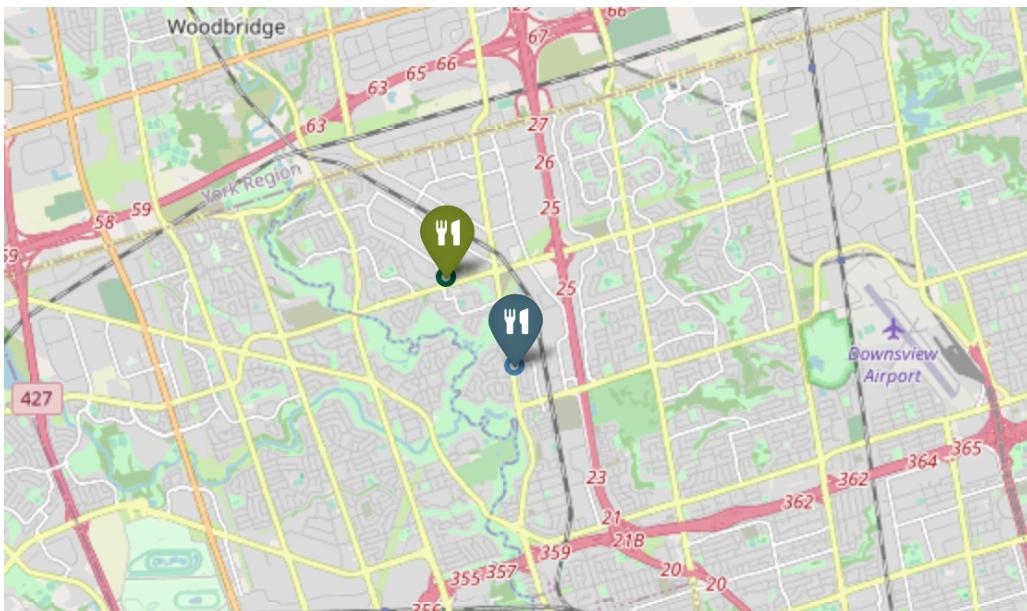
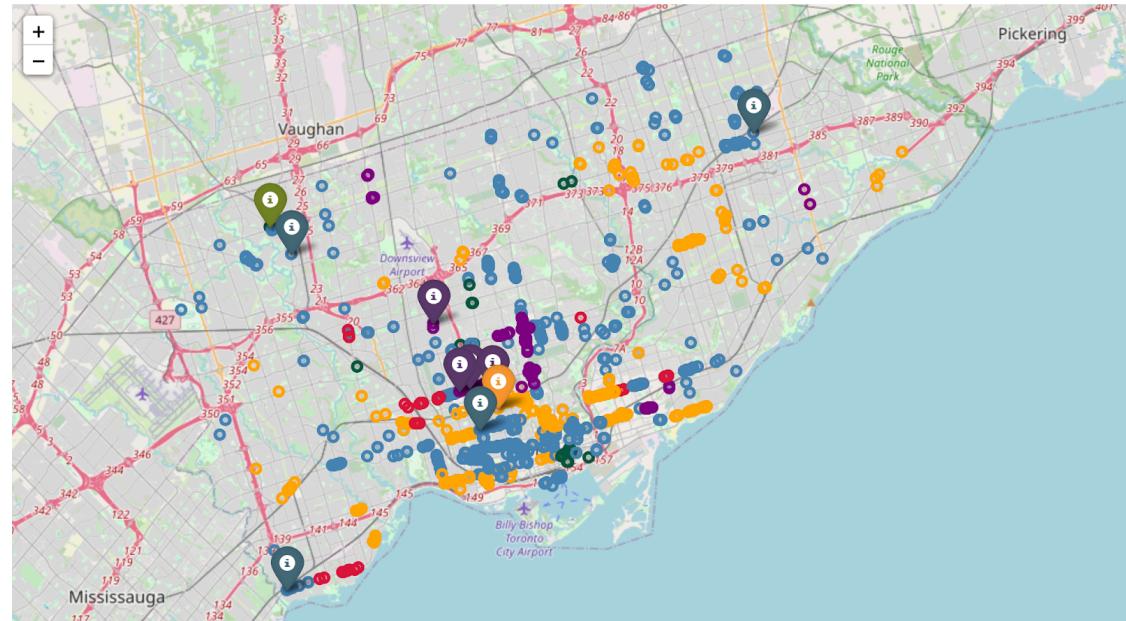


Cluster Maps



Restaurant Search: Clusters

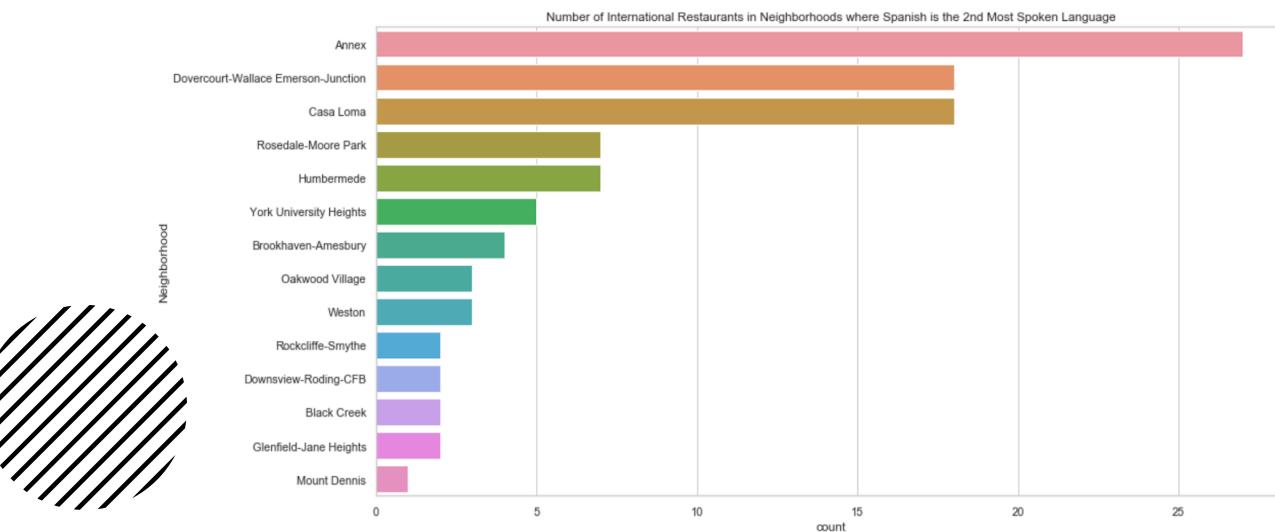
- Map shows all ethnic restaurants
- Pin markers represent Latin American/South American restaurants
 - Similarity to Colombian restaurants.
- Look at the top 10 venue categories in the five clusters of the previous section.
 - Clusters 1 and 5 have Latin American Restaurant as 7th and 3rd respectively
 - Humbermede (blue) and Humber Summit (green)
 - Humbermede's restaurant is named Mi Pueblo, serves some Colombian food in addition to dishes from other countries
 - Humber Summit has a mini-mall named Plaza Latina, which serves different dishes from all over Latin America

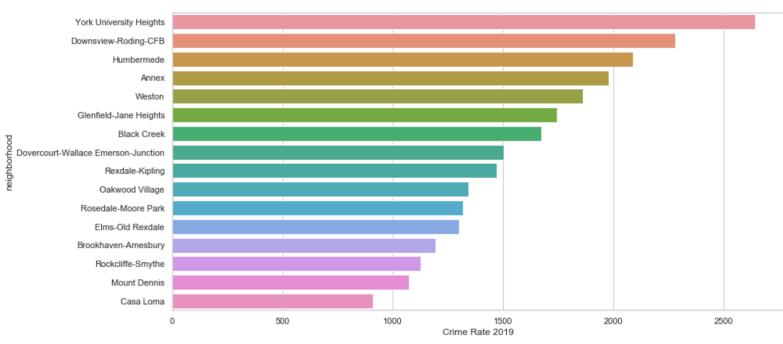
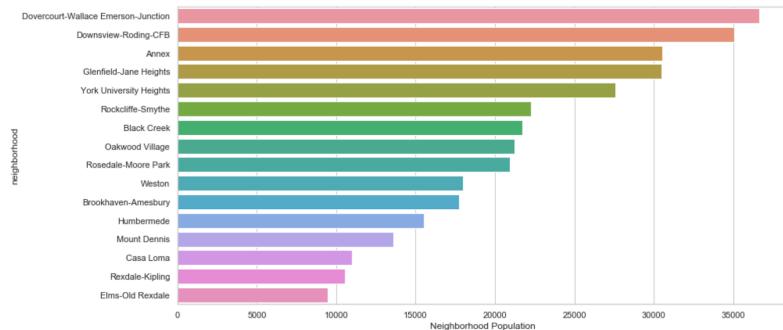
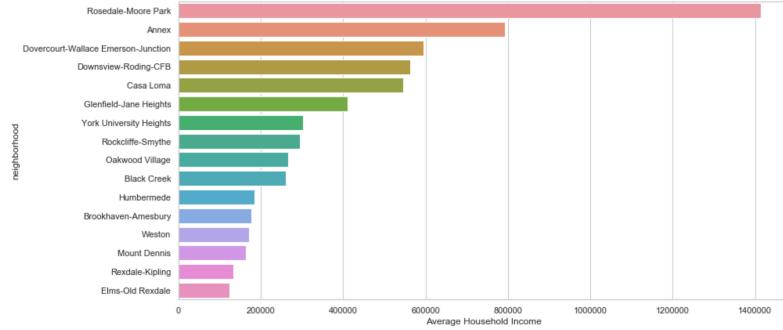




Restaurant Search: Spanish-Speaking Neighborhoods

- Looked in neighborhoods where Spanish is the most spoken language after English using the following components.
 - Other international restaurants operating
 - Location of other Latin American restaurants, if any,
 - Average household income and crime rate statistics.
- 14 out of 20 neighborhoods have at least one ethnic restaurant.
- Circle markers in the map represent all ethnic restaurants in the 14 neighborhoods
- Pin markers represent neighborhoods
- The four neighborhoods with the most ethnic restaurants are located close to downtown.





Neighborhood Shortlist

- Based on the three bar plots on the right and the clusters, the shortlist includes five neighborhoods for the restaurant.
 - Annex - most international cuisine venues
 - Dovercourt-Wallace Emerson-Junction is the most populated neighborhood
 - Humbermede is the Spanish-speaking neighborhood from the cluster results
 - Casa Loma has the lowest crime rate
 - Glenfield-Jane Heights has the largest population of Spanish Speakers



○ Final Results

- In order to choose one neighborhood out of the five, it would be Casa Loma
 - Although Annex has 27 venues contrary to 18 in Casa Loma, it has a higher crime rate, especially in breaking and entering as well as assault.
 - It is important to consider safety issues and crime statistics when choosing a location.
 - Even so, Annex ranks second no only for number of ethic restaurants but also because it has at least a Latin American restaurant compared to Casa Loma
 - Although Latin American restaurants are not among its 10 most common venues, Casa Loma has one named Gourmet Gringos.
 - Moreover, this neighborhood's venues include Mexican, Indian, French, Japanese, and Middle Eastern restaurants.



Discussion and Conclusion

- Gathered four datasets to compare and narrow down neighborhoods for a Colombian restaurant.
- In addition to neighborhood venue data, I used:
 - Neighborhood population
 - 2019 crime rate by 1000 people
 - Average household income
- Sourced through Spanish-speaking neighborhoods with demographic variables and number of ethnic restaurants.
- K-means clustering helped narrow down at least one neighborhood for the top 5 locations.
 - Proved to be quite useful in narrowing restaurant locations with data.

