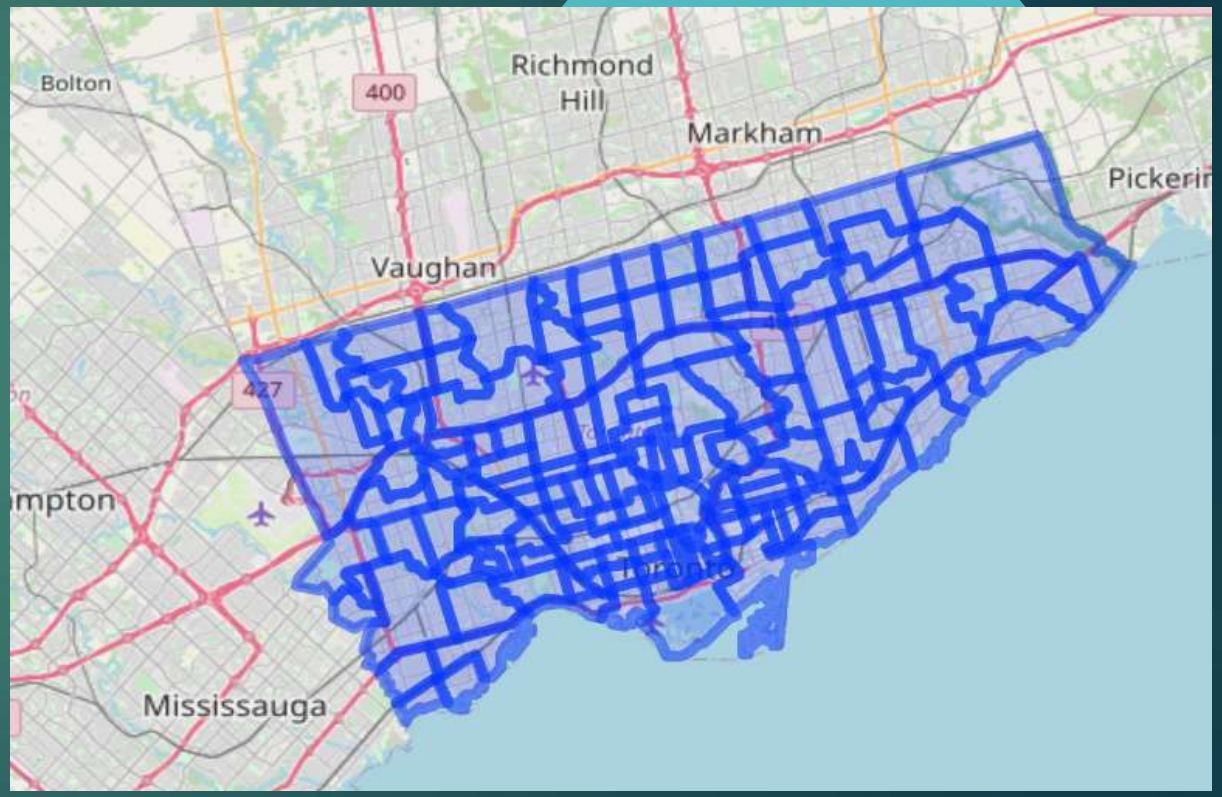
Understanding Demographics, Crime and Venues for Neighborhood Segmentation: A k-means approach

BATTLE OF NEIGHBORHOODS – PRESENTED BY MAIGHA

Introduction

- Neighborhoods in Toronto explored for livability based on features
- Total 140 neighborhoods in the area
- Key question:
  - ***Can we determine the attractiveness of a neighborhoods (top 10) based on the venues, lower crime rates, number of rented dwellings, average rent, etc?***



Data Sources and Preparation

- Multiple data sources were used.
  - Geospatial coordinated of neighborhoods
  - Foursquare for venues data
  - Crime data for that past 5 years
  - Wellbeing data

- Key features

Venues	Total population	Healthy food index
Early development instrument	Recent immigrants	Average family income
Tenant average rent	Rented Dwellings	Owned Dwellings
Assault	Auto Theft	Breaking and Entering
Homicide	Robbery	

Venues Analysis

- Top 10 venues for each neighborhood were determined

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
Agincourt North	Pizza Place	Discount Store	Fast Food Restaurant	Fried Chicken Joint	Frozen Yogurt Shop	Liquor Store	Sandwich Place	Beer Store	Chinese Restaurant	Bakery
Agincourt South-Malvern West	Chinese Restaurant	Mediterranean Restaurant	Bank	Pool Hall	Restaurant	Noodle House	Cantonese Restaurant	Seafood Restaurant	Shopping Mall	Motorcycle Shop
Alderwood	Pizza Place	Convenience Store	Pharmacy	Coffee Shop	Fast Food Restaurant	Electronics Store	Ethiopian Restaurant	Event Space	Falafel Restaurant	Farm
Annex	Sandwich Place	Café	Pub	Pharmacy	BBQ Joint	Social Club	Burger Joint	Pet Store	French Restaurant	Liquor Store
Banbury-Don Mills	Shoe Store	Pizza Place	Gourmet Shop	Coffee Shop	Movie Theater	Furniture / Home Store	Liquor Store	Sandwich Place	Cantonese Restaurant	Cosmetics Shop

Neighborhood Crime Rates

- The crime data included multiple fields like assault, theft, etc.
- The averages for each field over the last 5 years were used

LOWEST CRIME RATES

Lambton Baby Point
Woodbine-Lumsden
Yonge-St.Clair
Maple Leaf
Markland Wood
Guildwood
Casa Loma
Forest Hill South
Old East York
Kingsway South

HIGHEST CRIME RATES

Waterfront Communities-The Island
Bay Street Corridor
Church-Yonge Corridor
West Humber-Clairville
Moss Park
Downsview-Roding-CFB
York University Heights
Woburn
Kensington-Chinatown
West Hill

Most Populated Areas and Neighborhoods with Highest Immigration

TOP 10 POPULATION DENSE AREAS

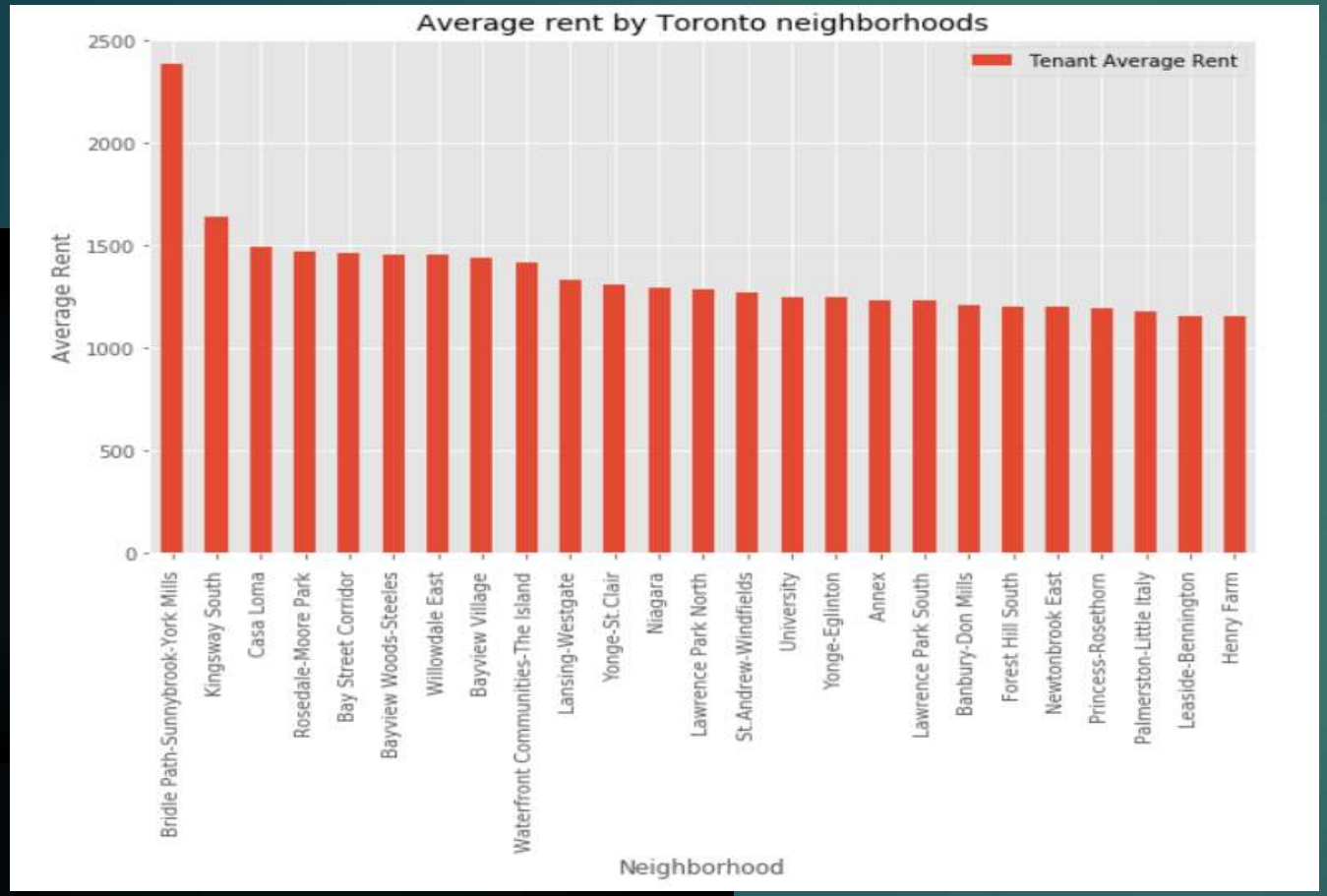
Waterfront Communities-The Island
Woburn
Willowdale East
Rouge
L'Amoreaux
Islington-City Centre West
Malvern
Dovercourt-Wallace Emerson-Junction
Downsview-Roding-CFB
Parkwoods-Donalda

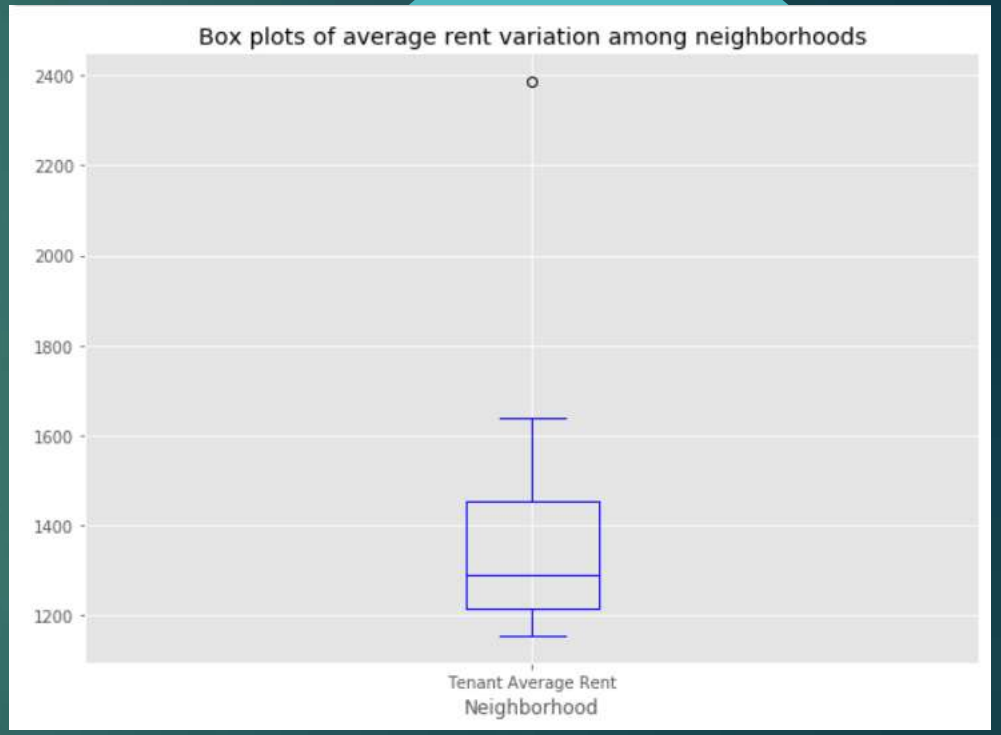
TOP 10 IMMIGRATIONS DENSE AREAS

Willowdale East
Woburn
Waterfront Communities-The Island
Mount Olive-Silverstone-Jamestown
Westminster-Branson
L'Amoreaux
Thorncliffe Park
Don Valley Village
Newtonbrook West
Downsview-Roding-CFB

- An interesting observation is that five of the most populated areas were also the ones with large immigrant population

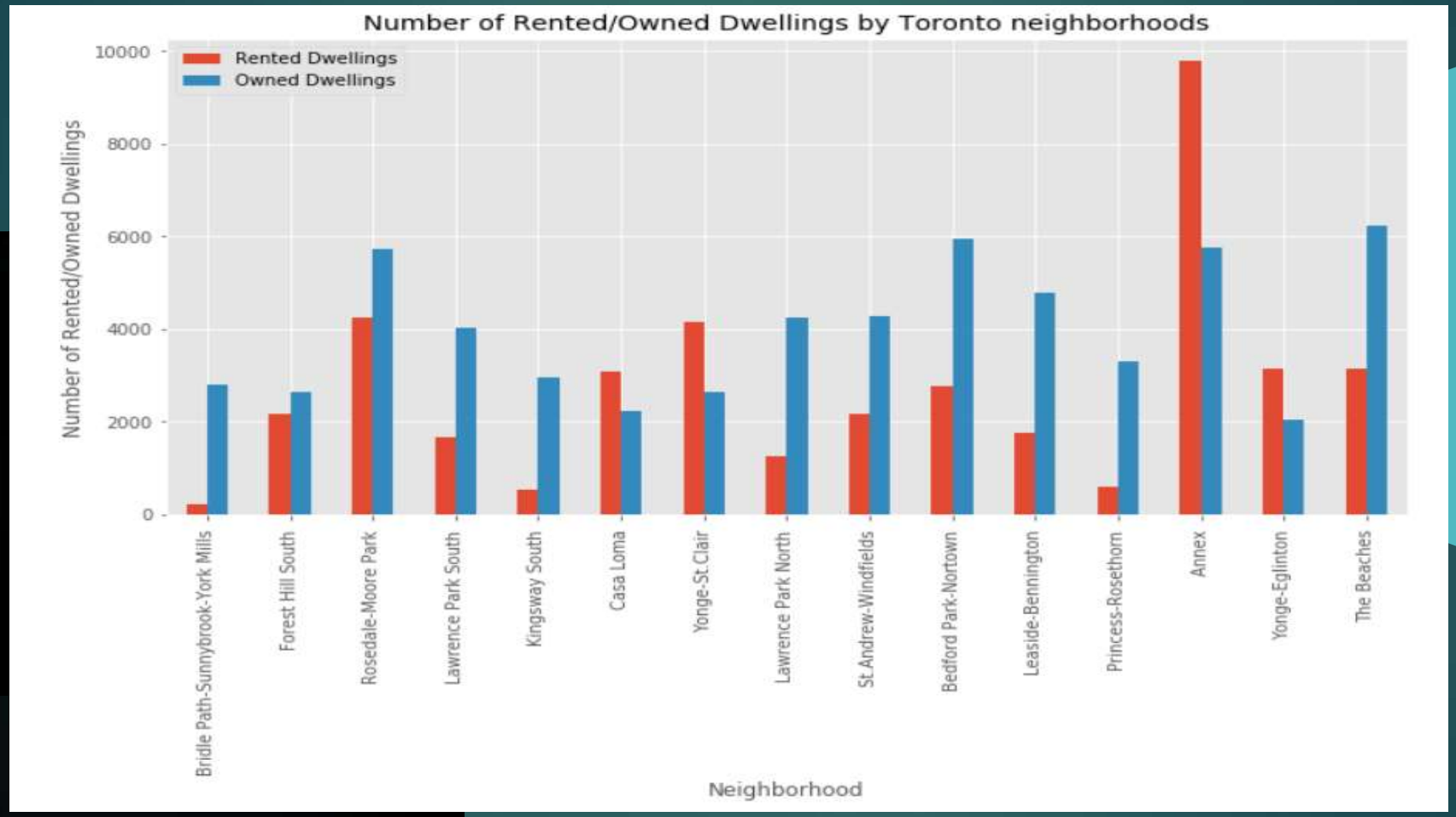
Average Tenant Rents and Distribution





- *Except for the first neighborhood in the chart, the rents of the other neighborhoods are distributed in a narrow range*

Rented vs Owned Dwellings

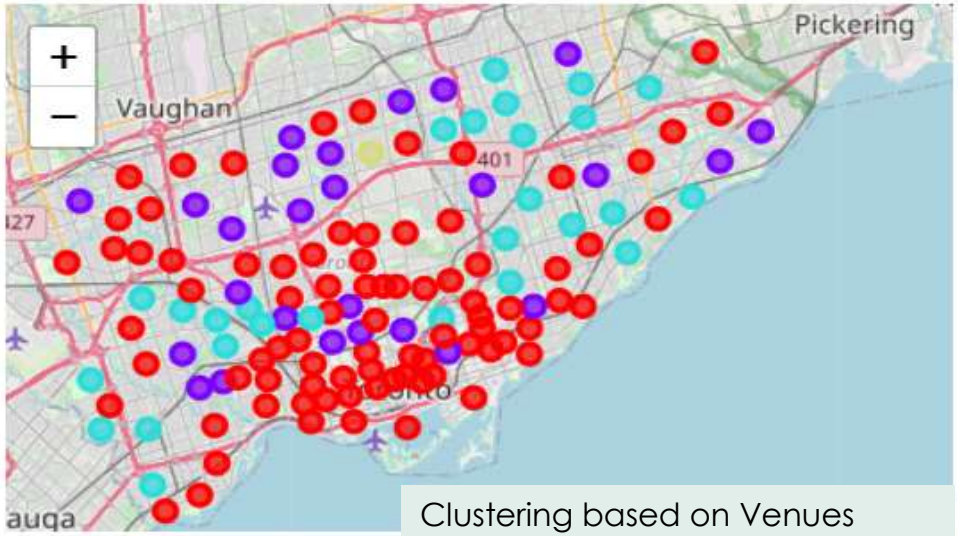


- Neighborhood with the highest rental average has the fewest rented units. It seems that mostly owned dwellings exist in the area

Neighborhood Clustering

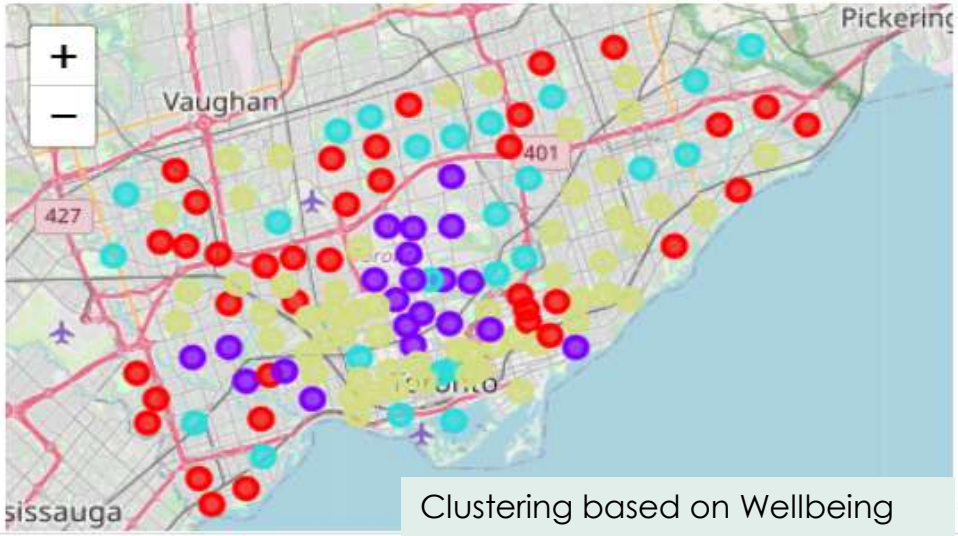
- Analysis was performed in 4 steps
  - The neighborhoods were clustered based on venues only
  - Clustering based on crime data – each feature taken individually
  - Clustering based on the features extracted from wellbeing data
  - Clustering based on combining the above features
- Data normalization was performed carefully

Results

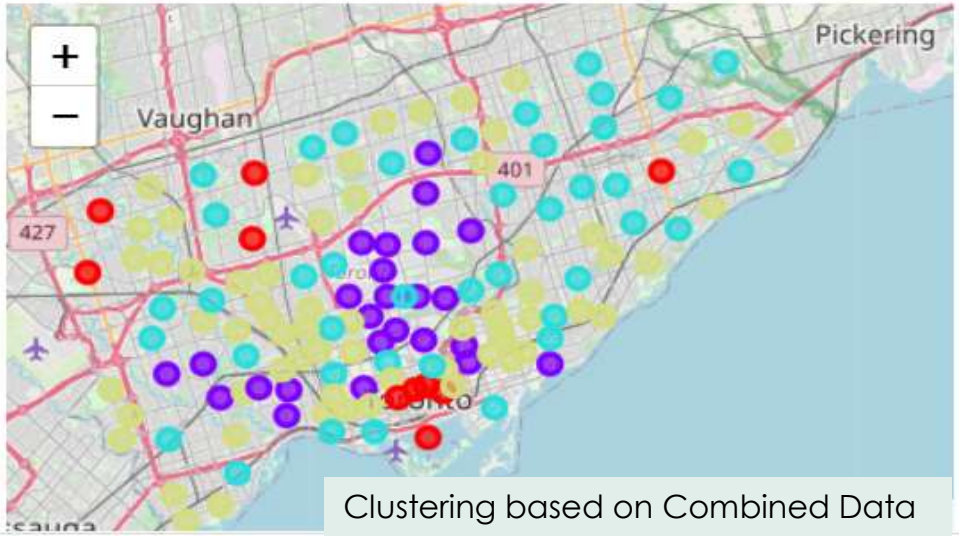


Clustering based on Venues

Clustering based on Crime



Clustering based on Wellbeing



Clustering based on Combined Data

Conclusion and Future Directions

- Using the clustering and initial data exploration, it may be possible to categorize and shortlist neighborhoods based on personalized criteria.
- The objective of this analysis was to show the possibility and some initial visualization.
- It is evident that a lot needs to be explored in depth for driving decisions based on this.
- Based on the above analysis, fewer features will be chosen for in-depth analysis to determine the underlying similarities between neighborhoods, their key features and characteristics.
- This study has probably just scratched the surface within a tiny scope.