IBM CAPSTONE PROJECT

WHERE SHOULD I LIVE?

TABLE OF CONTENTS

1) Introduction / Business Problem
2) Data
3) Methodology
4) Results
5) Discussion
6) Conclusion

INTRODUCTION / BUSINESS PROBLEM

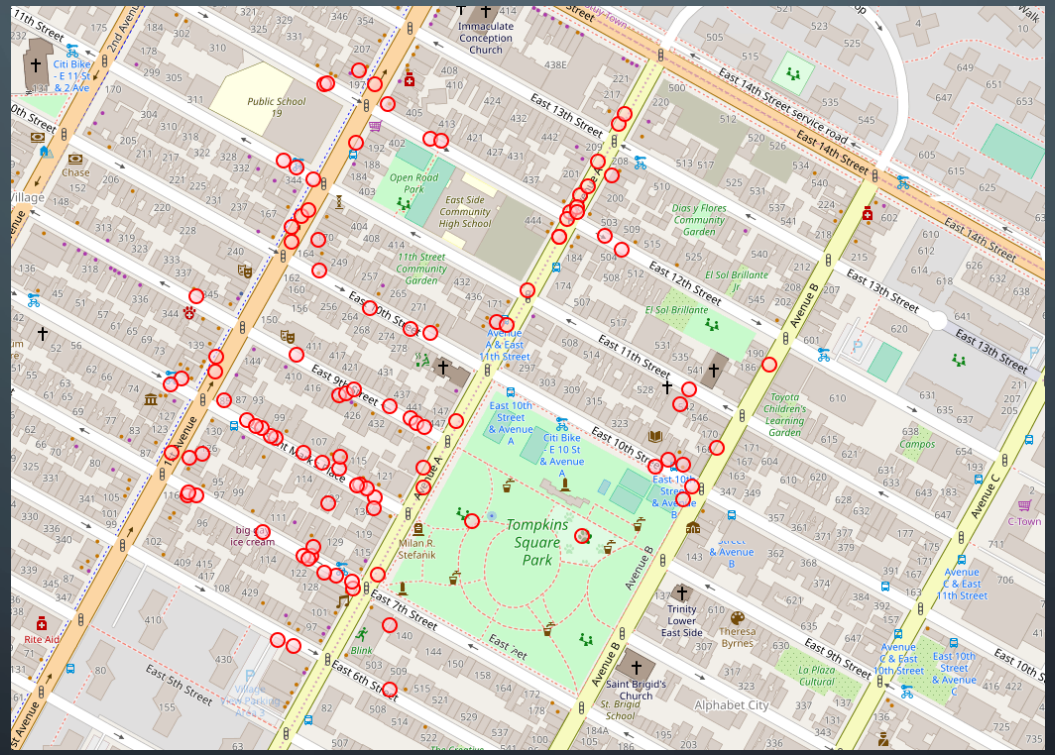
- Finding a place to live in a new city may be time consuming and expensive.
- Business Problem: How do I travel to a new city and find a neighborhood in which to live that offers similar comforts and amenities to the one I left?
- Target Audience:
  - People moving to another city and with a desire to maintain similar living conditions.
  - Real estate agents/relocation consultants who desire to offer a customized service.

DATA

- New York City Data
  - NYU Spatial Data Repository
  - Provides Borough, Neighborhood, Latitude, and Longitude data
- City of Toronto Data
  - “List of Postal Codes of Canada: M” Wikipedia page
  - Instructor furnished Geospatial_Coordinates.csv file
  - Provides Borough, Neighborhood, Latitude, and Longitude data
- FourSquare Data
  - Connect Using API to Gather Venue Information

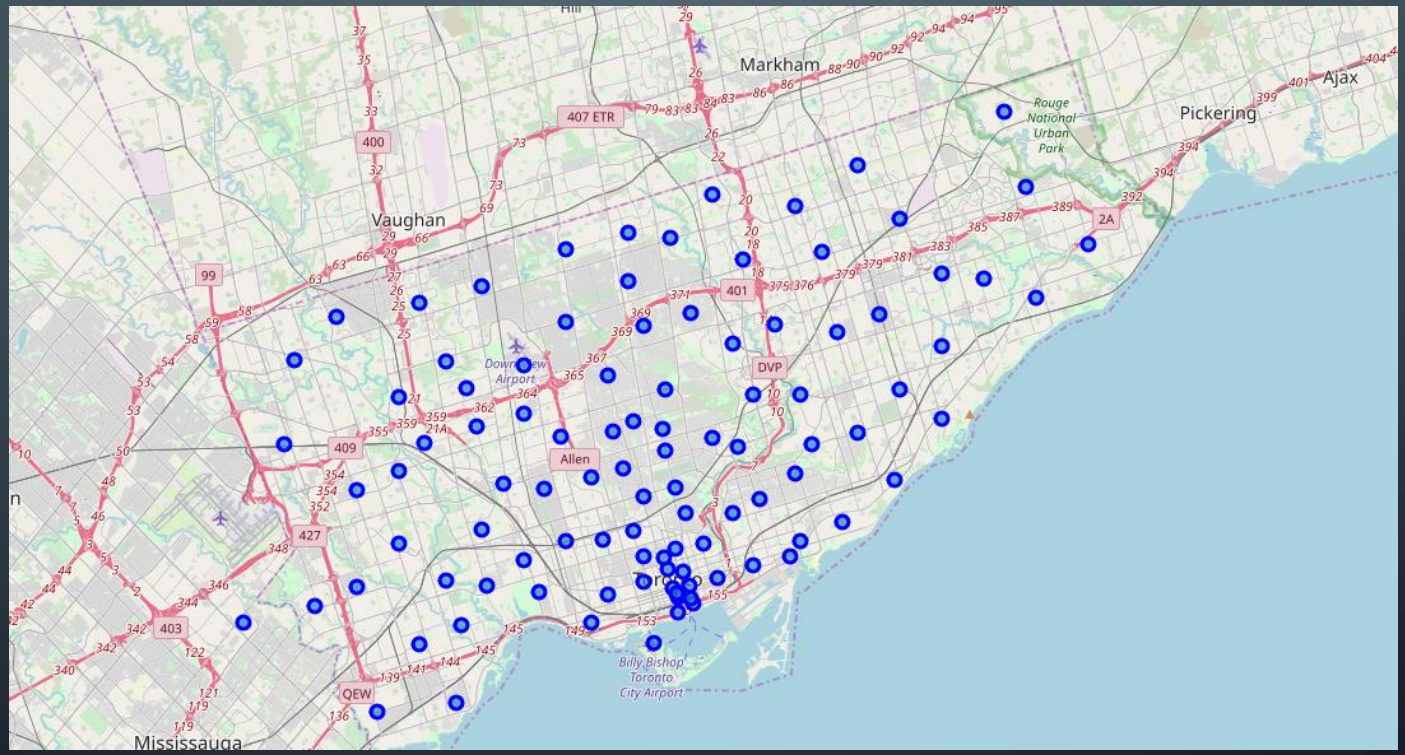
DATA (CONTINUED)

Map of Venues within the East Village of New York City



DATA (CONTINUED)

Map of Neighborhoods within the City of Toronto

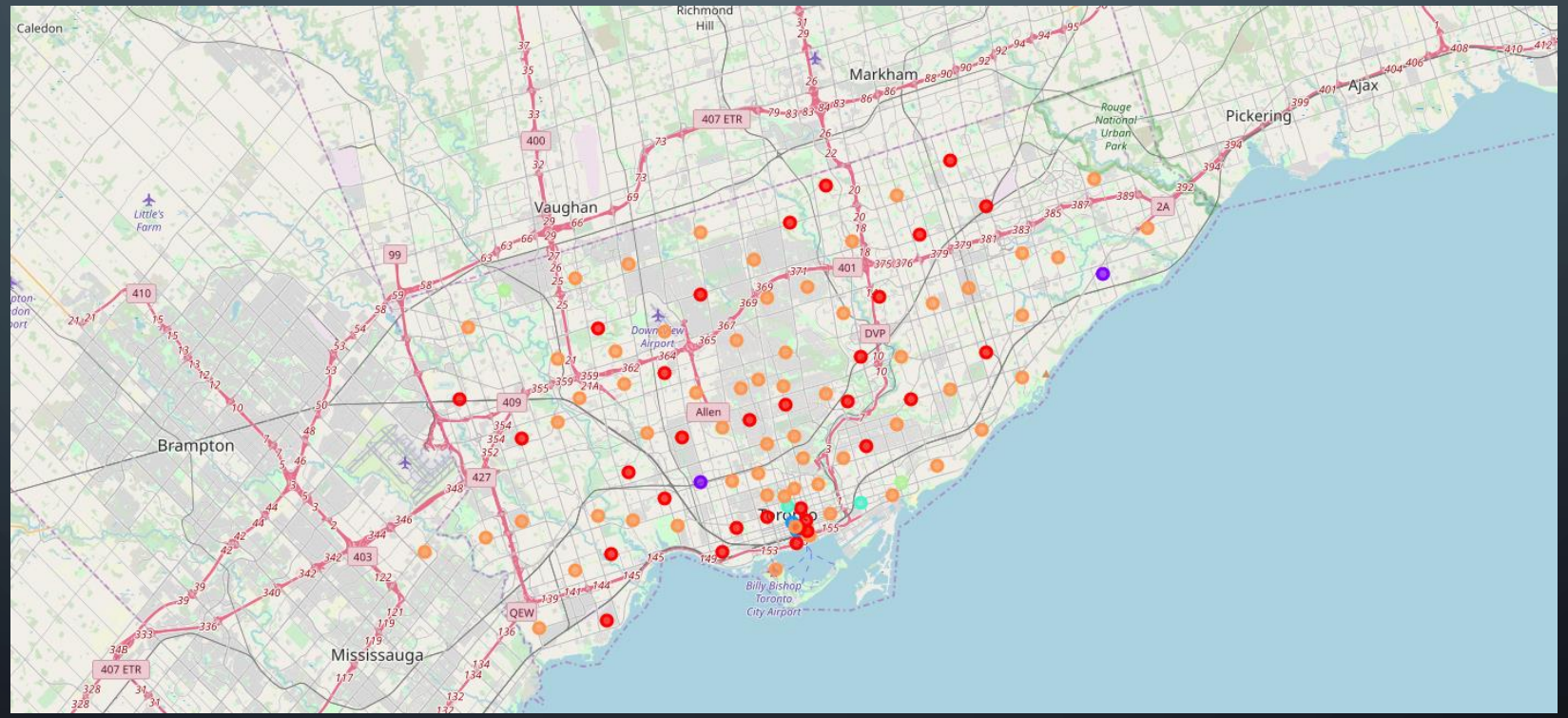


METHODOLOGY

1. Find top 10 venue categories in person's current neighborhood.
2. Find top 10 venue categories in neighborhoods within the target city.
3. Combine current neighborhood with those in the target city for clustering.
4. Use k-means machine learning algorithm to cluster neighborhoods.
5. Examine resulting clusters to find which one contains current neighborhood. This cluster presents those neighborhoods the individual should investigate for potential relocation.

RESULTS

The current neighborhood (East Village) is matched with 62 neighborhoods in Toronto. This cluster is identified by the color orange. These are the potential relocation areas.



DISCUSSION

- More neighborhoods (62) were returned than had been expected.
  - Tried different setting for “k.” None appeared to make a significant difference.
  - Including a procedure to find the optimum number for “k” may improve results.
- Alternative approach may provide improved results, such as:
  - Mean-Shift Clustering
  - Density-Based Spatial Clustering of Applications
  - Expectation–Maximization (EM) Clustering
  - Agglomerative Hierarchical Clustering.
- Future enhancement could include the introduction of a proxy for venue quality.

CONCLUSION

- Offers a systematic and reproducible approach utilizing a quantitative based methodology to compare the neighborhood of one city to those of another.
- Provides a beginning step for those looking to relocate and defines a list of possible relocation sites before one leaves home, potentially saving them both time and money.

QUESTIONS?