

The slide features two abstract green geometric shapes. On the left is a solid green triangle pointing downwards. On the right is a complex, multi-layered green shape composed of several overlapping triangles and polygons in various shades of green, creating a dynamic, layered effect.

Customized recommendation engine for identification of residential communities in Toronto

Problem Statement

- Which neighborhood in Toronto should be picked for one of my clients for purchasing a new house or condo?



Introduction

- Choosing a location when looking for a home is very important! After all, you can always update or fix your house, but you can't easily change its location and the vibe of the community!
- In this project, I would like to create a hypothetical business scenario with an aim of finding a compatible location for one of my clients.
- According to National Association of realtors, there is a significant home search process nearly among all generations of home buyers, ranging from Millennials/Gen Yers, Gen Xers, Younger Baby Boomers and Older Baby Boomers, except for the Silent Generation. In fact, buyers typically searched for 10 weeks and looked for at a median of 10 homes. This project aims to provide customized recommendation engines for home buyers during their home search process.

What challenges does this project aim to tackle and who are the customers?

- What if the customer doesn't have time to go through a thorough evaluation of the local communities?
- What if the customer lives in a different location, and they have no clue about the local neighborhood?
- Or perhaps, although they do have a general idea of the local neighborhood, they cannot make up their mind about the multiple choices.
- It is also possible that in the middle of the process, the customer thought of new important metrics that they or their significant others care about, how to make a decision effectively and efficiently in such a scenario?
- In this project, I aim to create a solution to address these issues.

Data

In this project, I will use four sets of raw data.

- **Wikipedia**
- **Geographical coordinates**
- **Data retrieved from FourSquare**
- **A dataset provided by the customer as his own expectation/rating.**

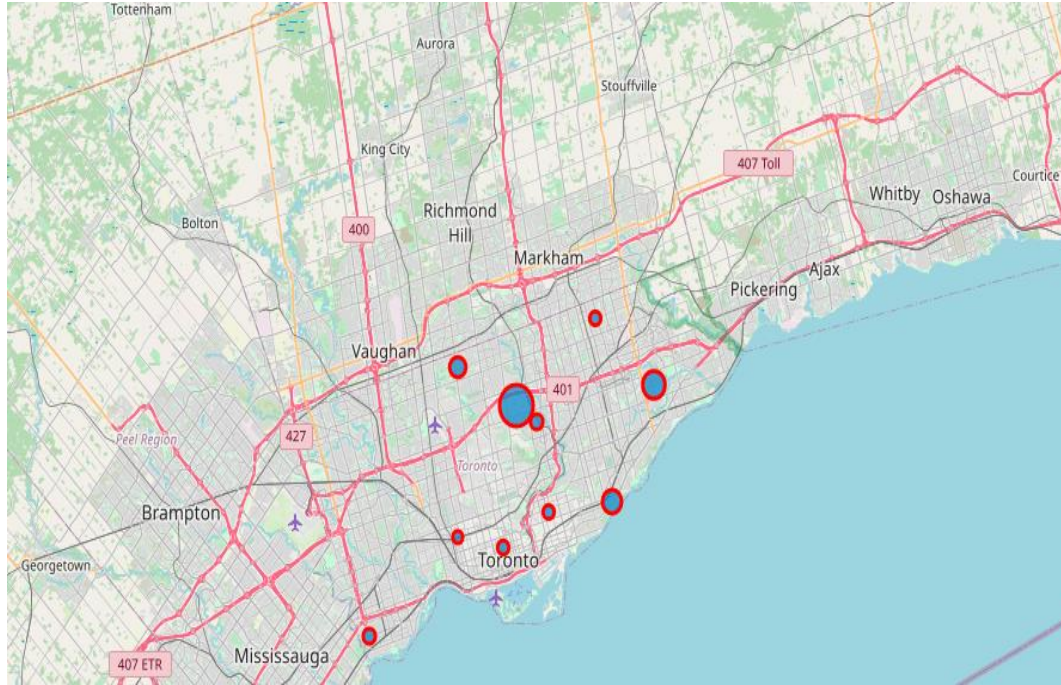


Methods

- Retrieve data and perform data format cleaning and preprocessing.
- Interpret customers' expectation when doing texts analysis. The alternative key words should be identified and considered.
- Filter the dataset based on customers' expectation to reduce computational complexity.
- Weighting the venues based on the actual frequency of occurrence of each venue as well as the customers' personal preference (the rating dataset). This will be performed by doing linear algebra operations of the venue matrix.
- Generate recommended candidate venues for the customer.



Results and conclusion



In this figure, the size of the circle represents the significance/rank of the area.

The area with the highest score has the biggest size to guide the eyes.

The top 10 recommended areas for the customer have the postal code M2L, M1G, M1N, M3B, M8W, M7A, M2R, M1V, M4J and M6H.

Discussion and Future Direction

- When using texts as the information to build recommendation, it is important to clarify ambiguity. For example, in this project, Café was further expanded to Coffee Shop, Gaming Café and Cafeteria; health care was expanded to Pharmacy and Hospital; gym was expanded to Climbing Gym, College Gym, Fitness Center. If these alternative words were not considered, we could potentially underestimate the significance of some neighborhoods.
- Moreover, we need to adjust the weight of each venue based on the preference of the customer. This process helps create a more customized recommendation.
- Future direction: communicate with customer and obtain more details of ratings of other important metrics such as price or house type. Also, if the customer could be more specific about the expectation of each metric, we could make the recommendation engine more precise.

