Coursera
Capstone Project

Improving in-patient satisfaction through personalized hospital selection

Capstone project submission

June 2021

Pre-empting in-patient satisfaction through personalized hospital selection

- Some people love being in a hospital, some hate it, others are indifferent. However, what is common to all is that their in-patient experience can be improved with the welcome escape to their favorite retreat, be-it a cafe, theatre or certain type of cuisine.
- Given the choice of which hospital to be admitted to, this will likely bear some weight on the patient's selection of hospital:
  - For medical professionals that are making these referrals, getting the best outcome for their patient includes making them feel comfortable in what will be a unfamiliar place.
  - For healthcare referral apps and also local search-and-discovery apps there is an unmet demand for pre-emptive hospital selection based on personalised nearby venues of interest.
- Pre-empting the choice of hospital can be achieved by knowing the patient's preference for their top 5 nearby venues and matching that to the analyses presented here.

Data

To conduct the analysis we use the Foursquare API to locate hospitals in Toronto, ON:

We use the search function to find all hospitals in Toronto and then use the getNearbyVenues function to get the most common venue categories within walking distance (500m) to each hospital.

The results are retuned as a JSON file which is filtered for necessary columns such as venue ID, name and location data. The JSON is then converted into a pandas dataframe for further cleansing of data including dropping of NaNs.

The hospital data has the following format:

index	id	name	lat	lng	address	postalCode
3	5de14daedb954c00086c07b2	Sunnybrook Hospital	43.721838	-79.375978	2075 Bayview Ave.	M4N 3M5

The venues data is extracted using the getNearbyVenues function, which is then grouped by hospital and venue count. Unique categories are extracted for analysis.

Methodology

This section describes how we explore venue data near hospitals in the Toronto area and group hospital with similar features using k-means clustering. K-means is vastly used for clustering in many data science applications, especially useful if you need to quickly discover insights from unlabeled data. In this notebook, you will learn how to use k-Means for hospital segmentation.

1. We use the search function to find all hospitals in Toronto
2. Then use the getNearbyVenues function to get the most common venue categories near to each hospital
3. We use this feature to group the hospitals into clusters and use the k-means clustering algorithm to complete this task.
4. The Folium library is used to visualize the hospitals in Toronto and their emerging clusters.
5. Finally, we analyze the cluster to find the dominant features characterizing it

The results are presented in the next section.

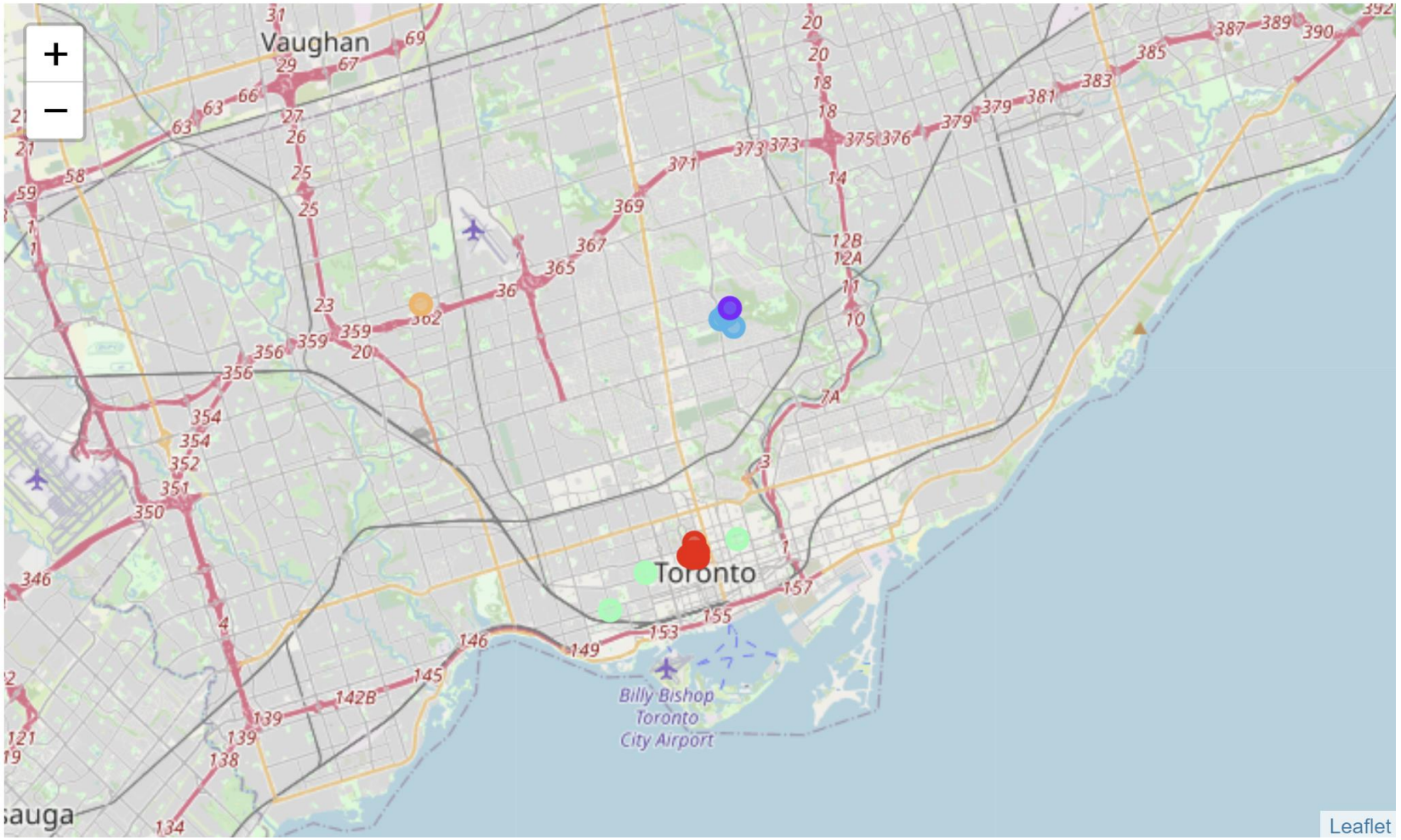
Analysis

- We encode venue categorical features as a one-hot numeric array and group by hospital to determine the frequency of occurrence of each venue category per hospital
- We select the top 10 venues per hospital, as shown below, and generate k-means clusters (shown in the following results section)

Hospital	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
CAMH	Bar	Men's Store	Art Gallery	Vietnamese Restaurant	Vegetarian / Vegan Restaurant	Theater	Seafood Restaurant	Café	Restaurant	Furniture / Home Store

Results: K-means

- Five k-means clusters are generated together with longitude and latitude location for each hospital



Results: Cluster 1

The discriminating venue categories that distinguish cluster 1 are fast food, deli and coffee shops

Hospital	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
Veterans Centre: Sunnybrook	Food Court	Deli / Bodega	Coffee Shop	Thai Restaurant	Café	Yoga Studio	Escape Room	Electronics Store	Donut Shop	Doner Restaurant

Results: Cluster 2

The discriminating venue categories that distinguish cluster 2 are bars, Italian restaurants, coffee shops and art venues

Hospital	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
CAMH	Bar	Men's Store	Art Gallery	Vietnamese Restaurant	Vegetarian / Vegan Restaurant	Theater	Seafood Restaurant	Café	Restaurant	Furniture / Home Store
Dundas Euclid Animal Hospital	Bar	Italian Restaurant	Cocktail Bar	Restaurant	Pizza Place	Sandwich Place	Art Gallery	Taco Place	Hobby Shop	New American Restaurant
Sherbourne Health Centre	Coffee Shop	Restaurant	Pub	Diner	Grocery Store	Gastropub	Japanese Restaurant	Italian Restaurant	Hotel	Breakfast Spot
Toronto Western Hospital	Bar	Café	Vegetarian / Vegan Restaurant	Taco Place	Mexican Restaurant	Burger Joint	Comfort Food Restaurant	Ramen Restaurant	Japanese Restaurant	Poutine Place

Results: Cluster 3

The discriminating venue categories that distinguish cluster 3 are delicatessen, coffee and thai food

Hospital	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
Sunnybrook Hospital	Deli / Bodega	Coffee Shop	Thai Restaurant	Food Court	Yoga Studio	Dog Run	Escape Room	Electronics Store	Donut Shop	Doner Restaurant

Results: Cluster 4

The discriminating venue categories that distinguish cluster 4 are parks, grocery stores and Vietnamese restaurants and pharmacy

Hospital	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
Humber River Hospital	Park	Grocery Store	Vietnamese Restaurant	Pharmacy	Coffee Shop	Distribution Center	Electronics Store	Donut Shop	Doner Restaurant	Dog Run

Results: Cluster 5

The discriminating venue categories that distinguish cluster 5 are coffee shops and Asian food

Hospital	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
Toronto General Hospital	Coffee Shop	Sushi Restaurant	Chinese Restaurant	Spa	Bubble Tea Shop	Gym / Fitness Center	Ice Cream Shop	Italian Restaurant	Japanese Restaurant	Café
The Hospital for Sick Children (SickKids)	Coffee Shop	Sandwich Place	Italian Restaurant	Café	Bubble Tea Shop	Burger Joint	Donut Shop	Portuguese Restaurant	Poke Place	Pizza Place
R. Fraser Elliot Building - Toronto General Ho...	Coffee Shop	Sandwich Place	Burger Joint	Café	Italian Restaurant	Bubble Tea Shop	Yoga Studio	Middle Eastern Restaurant	Poke Place	Pizza Place
Women's College Hospital	Coffee Shop	Sushi Restaurant	Café	Sandwich Place	Bookstore	Park	Beer Bar	Distribution Center	Smoothie Shop	Pizza Place

Discussion

The results from the data analysis show that there is explanatory information within foursquare to help an app developer to match in-patient user preferences to the clustered groups and prioritize the choice of hospital.

The application of k-means clustering was straightforward and showed that a good amount of explanatory data can be collected using this method from the foursquare dataset. Further analysis can be done to optimize the number of clusters used in the analysis. There are possible improvements to the foursquare dataset that can be made, such as a additional query items to specify the type of hospital or medical centers. Some venue categories, such as coffee shops, are generic and can be removed before performing the clustering in-order to focus on discriminating venue categories. Additionally, foursquare user tips can be incorporated to find anecdotal data for hospital experience and nearby venues.