Capstone Project Report

Sergii Baidachnyi

Problem

- Help employees to move from one region to another
  - We are moving a team from Toronto to New York right now
- Employees and their families have access to the same venues including shops, coffee shops, parks and so on.
- An employee can share existing district or desired one in own native city
- We need to provide more than one district to choose for moving

Available Data

- District details for New York (https://geo.nyu.edu/catalog/nyu_2451_34572)
- District details for Toronto (https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M)
- Access to Foursquare API to explore information about the most popular places in a particular area

How to make recommendation

- We are going to use K-Means clustering algorithm to split all districts in New York and Toronto alltogether in several clusters based on the most popular venues in Foursquare. Number of clusters cannot be recommended in advance and we might have potentially several cases:
  - some districts from Toronto are alone in a particular cluster: in this case we will need to decrease number of clusters to make recommendations exactly for these disctricts. All other districts from Toronto (that already have some areas from New York in their clusters) should be ignored because there recommendation is made already;
  - some districts from New York doesn't have any districts from Toronto in their clusters: these clusters might be ignored because we don't care about New York districts that are not similar to Toronto districts;
  - lots of districs from Toronto at the same clusters: in this case we need to increase number of clusters to make better splitting/recommendation for them. All districts from other clusters should be ignored because they already have recommendation;

Results

- As the result we have some clusters that have mix of districts from Toronto and New York;
- All employees that live in a district of Toronto (Westmount, for example) should have a list of districts in New York that we would like to recommend to live (Marble Hill or Chinatown in the case of Westmount)

```
In [75]: eto_merged.loc[eto_merged['Cluster Labels'] == 1, eto_merged.columns[[0] +[1]+ list(range(5, eto_merged.shape[1]))]]
```

		Markland Wood, Ol...					Center					
8	Etobicoke	Westmount	Pizza Place	Sandwich Place	Supermarket	Chinese Restaurant	Discount Store	Intersection	Middle Eastern Restaurant	Coffee Shop	Breakfast Spot	Ice Cream Shop
9	Etobicoke	Kingsview Village, Martin Grove Gardens, Richv...	Supermarket	Pizza Place	Supplement Shop	Mobile Phone Shop	Chinese Restaurant	Beer Store	Coffee Shop	Pharmacy	Sandwich Place	Bank
11	Etobicoke	Northwest	Coffee Shop	Lounge	Dog Run	Yoga Studio	Food & Drink Shop	Filipino Restaurant	Fish & Chips Shop	Fish Market	Flea Market	Flower Shop
12	Manhattan	Marble Hill	Park	Pizza Place	Spanish Restaurant	Supermarket	Mexican Restaurant	Donut Shop	Café	Sandwich Place	Athletics & Sports	Bar
13	Manhattan	Chinatown	Chinese Restaurant	Cocktail Bar	Ice Cream Shop	Café	Wine Bar	Sandwich Place	American Restaurant	Shoe Store	Optical Shop	Thai Restaurant
14	Manhattan	Washington Heights	Pizza Place	Latin American Restaurant	Café	Bakery	Grocery Store	Mexican Restaurant	Deli / Bodega	Tapas Restaurant	Bar	Park

Future Improvements

- Add additional details to existing data:
  - Time to commute
  - Closest schools
  - Car availability
  - And so on
- Additional details will help us to make better clustering

Summary

- Thanks to K-Means clustering we could build a table of all available districts (in Toronto and New York) and assign a cluster number to each of them
- Once any HR member wants to make a recommendation for relocation, it will be possible to find a cluster number for the current district of an employee and filter all data in New York by the same cluster number to make the recommendation