

Capstone Project:

Best Bubble Tea Store Location in Toronto

Jerry(Ziyuan) Yan

1.0 Introduction & Business Problem:

Bubble tea has now become an increasingly popular drinks among young people. A friend of mine is looking to invest in a bubble tea store in a Toronto. They would like to find a great location for their bubble tea store. They come to me to seek help to pick a perfect place for their new store in the city of Toronto.

With that in mind, we have some constraints for the desired location. We want to find the desired place with two major conditions:

First, we would love to pick a location that has higher than average monthly average apartment rent since we believe higher income population tend to have higher chance of purchasing bubble tea.

Second, we would like our store to close to at least 5 Asian restaurants. This is because there is high correlation between Asian restaurant and milk tea store. This suggests majority of Asian restaurants' consumers tend to also like milk tea. Therefore, we want to place our store near Asian restaurants.

2.0 Data:

We would like to employ Toronto's neighborhood information (latitude and longitude) of City of Toronto are scrapped from Wikipedia page. (https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M)

Wikipedia page provides geo coordinate for each neighborhood in the city of Toronto.

Venues information are provided by Foursquare developer API. We will use venues information for each neighborhood to aggregate food/drink service provider.

Housing and rent data for each neighborhood are from 2018 October CMHC Rental Market Survey.

3.0 Methodology

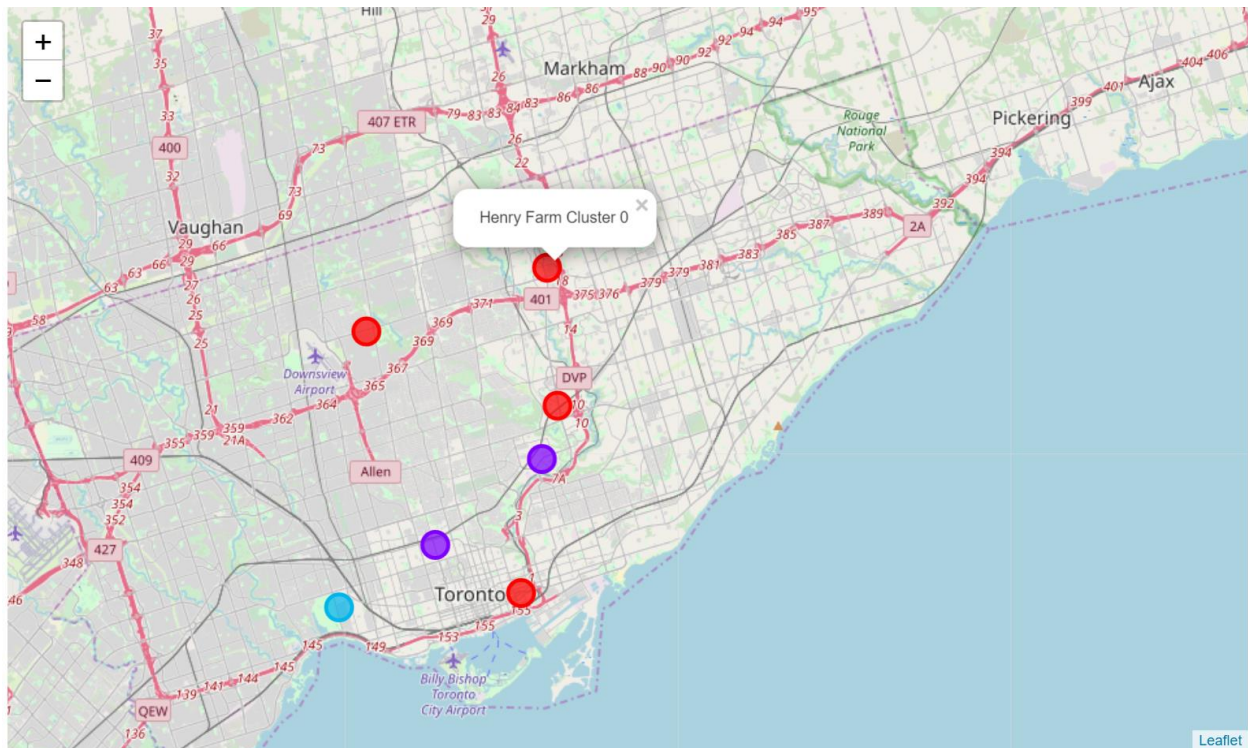
To find the desired location to meet the requirement, we want to first employ KMeans clustering to perform clustering for all neighborhood's in the city of Toronto. After clustering, we would like to identify the neighborhood centroid that meets requirement of higher than average housing cost.

The next step is to use information provided by four squares to identify neighborhood that meet the requirement of Asian restaurant using percentile rank.

The highest-ranking neighborhood will become the target location.

4.0 Result & Discussion

The graph below shows the clustered neighborhood after filtering and selection. Three clusters are chosen to group the similar neighborhood together.



Here are some details on our cluster of interest, Cluster 0:

Exam Cluster

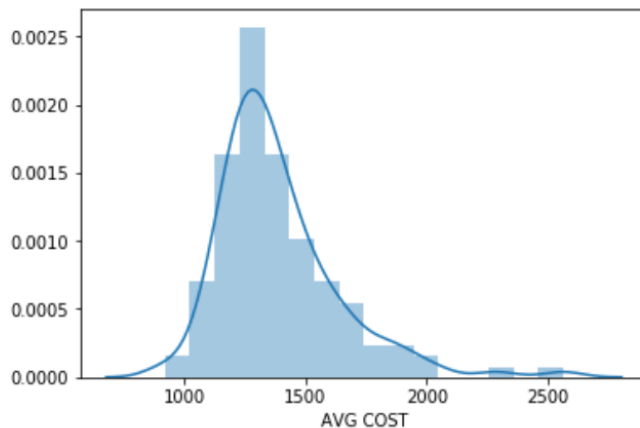
```
In [40]: tor_merged.loc[tor_merged['Cluster Labels'] == 0, tor_merged.columns[[2] + list(range(5, tor_merged.shape[1]))]].merge(df_rer
```

Out[40]:

	Neighborhood	Cluster Labels	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	Bachelor Be
0	Regent Park	0	Coffee Shop	Park	Bakery	Café	Pub	Theater	Mexican Restaurant	Breakfast Spot	Hotel	Farmers Market	975
1	Flemington Park	0	Gym	Asian Restaurant	Coffee Shop	Beer Store	Sporting Goods Shop	General Entertainment	Italian Restaurant	Japanese Restaurant	Fast Food Restaurant	Discount Store	**
2	Bathurst Manor	0	Coffee Shop	Shopping Mall	Supermarket	Deli / Bodega	Pet Store	Pharmacy	Pizza Place	Diner	Restaurant	Sandwich Place	**
3	Henry Farm	0	Clothing Store	Coffee Shop	Fast Food Restaurant	Food Court	Juice Bar	Bus Station	Asian Restaurant	Restaurant	Toy / Game Store	Jewelry Store	**

As shown in the table, cluster 0 consists of mainly restaurant related venues, which is a plus for our new bubble tea place.

One important factor that we need to consider is the average cost of living each neighborhood. We want to select a neighborhood that has higher than average living cost since we believe higher disposable income group are more likely to be our potential customers.



The above histogram shows the distribution of average living cost for the city of Toronto by different neighborhood. The mean cost of living is CAD 1,388.

Two choice in the cluster 0 meet the requirement:

Selection1 : Regent Park with average cost of living CAD 1585

Selection2 : Henry Farm with average cost of living CAD 1634

5.0 Conclusion

According to our analysis, we narrow down our potential bubble tea location to two neighborhoods in Toronto: Regent Park and Henry Farm. The two selections were chosen first by K-means to group similar venues that contains a higher than average Asian restaurant. After meet the first criterion, we examined average rent in the city of Toronto by neighborhood. We filtered out selections that have lower than average cost of living. This enable us to narrow down the two finalists.

Finally since Henry Farm has more Asian restaurant than Regent Park, I suggest the ideal bubble tea location to be Henry Farm.