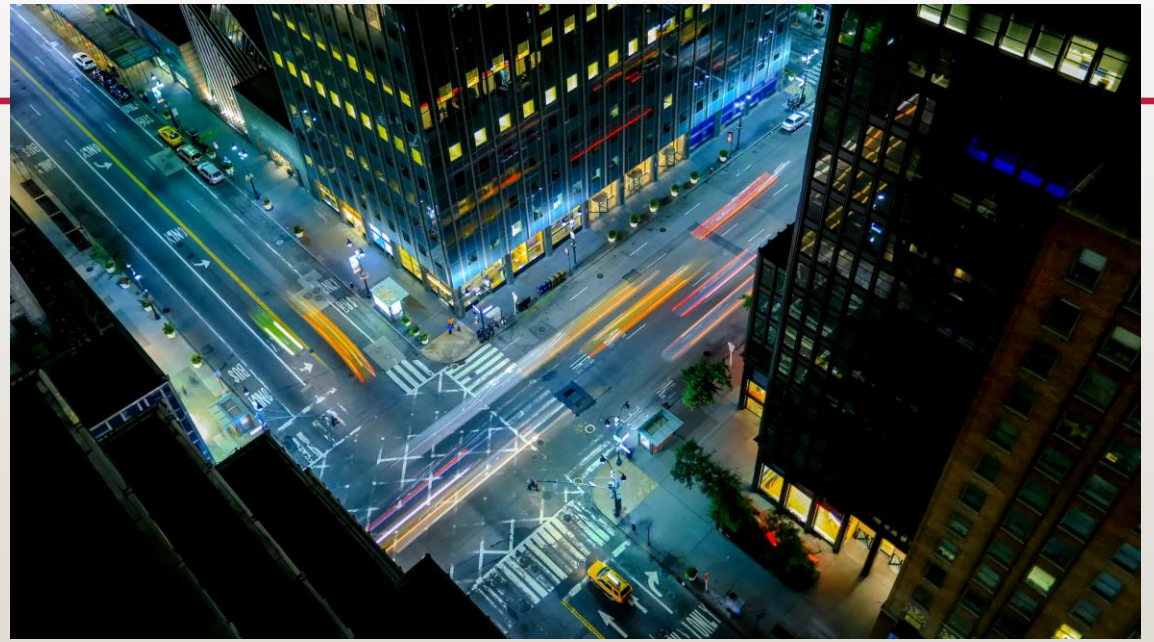
Different Places -Different Customers- Toronto

Data based classification of Toronto locations



BACKGROUND

- Traditional market research for shortlisting preferred location for shops/commercial spaces are
  - Time consuming
  - Keeps changing based on end use of shops
  - Static – can not be used on any other places
- Using data science – creating a dynamic and more flexible model to predict preferred location for shops/commercial spaces

PROBLEM

- Marketing research firm wants to highlight key business activities and preferred locations of shops based on customer preference to its clients
- Short listed Main Toronto - Needs further drill down for preferred location

SOLUTION LOGIC

- Typically , an area where most of shopping / commercial outlets are located , suggests that it has more commercial activities in and around that area.
- Further, based on simliar grouping, we can identify customer specific locations i.e what type of customer visit that place and/or what activity they perform.
- **Approach** :
- Demarcate Main Toronto area on the basis of pin codes and then using Foursquare API n identify which area has maximum stores/venues located within 500 meters of radius.
- Further, with using clustering algorithm, we can divide each location based on type of venues /shops present , indicating what type of customers /activity is preferred in each locations
-

DATA SOURCE

- Wikipedia - https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
- http://cocl.us/Geospatial_data : For latitude and longitude
- Foursquare APIs

DATA WRANGLING

- We used different methods to format data in our desired format
- From Wikipedia

	PostalCode	Borough	Neighborhood	Latitude	Longitude
0	M5A	Downtown Toronto	Regent Park, Harbourfront	43.654260	-79.360636
1	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government	43.662301	-79.389494
2	M5B	Downtown Toronto	Garden District, Ryerson	43.657162	-79.378937
3	M5C	Downtown Toronto	St. James Town	43.651494	-79.375418
4	M4E	East Toronto	The Beaches	43.676357	-79.293031

DATA WRANGLING

- Final data From foursquare

	name	categories	lat	lng
0	Downtown Toronto	Neighborhood	43.653232	-79.385296
1	Nathan Phillips Square	Plaza	43.652270	-79.383516
2	UNIQLO ユニクロ	Clothing Store	43.655910	-79.380641
3	Elgin And Winter Garden Theatres	Theater	43.653394	-79.378507
4	Richmond Station	American Restaurant	43.651569	-79.379266

METHODS

- We used Folium for plotting map , pandas for data frame statistics and sklearn for kmeans machine learning algorithm.
- Final map of locations

METHODS

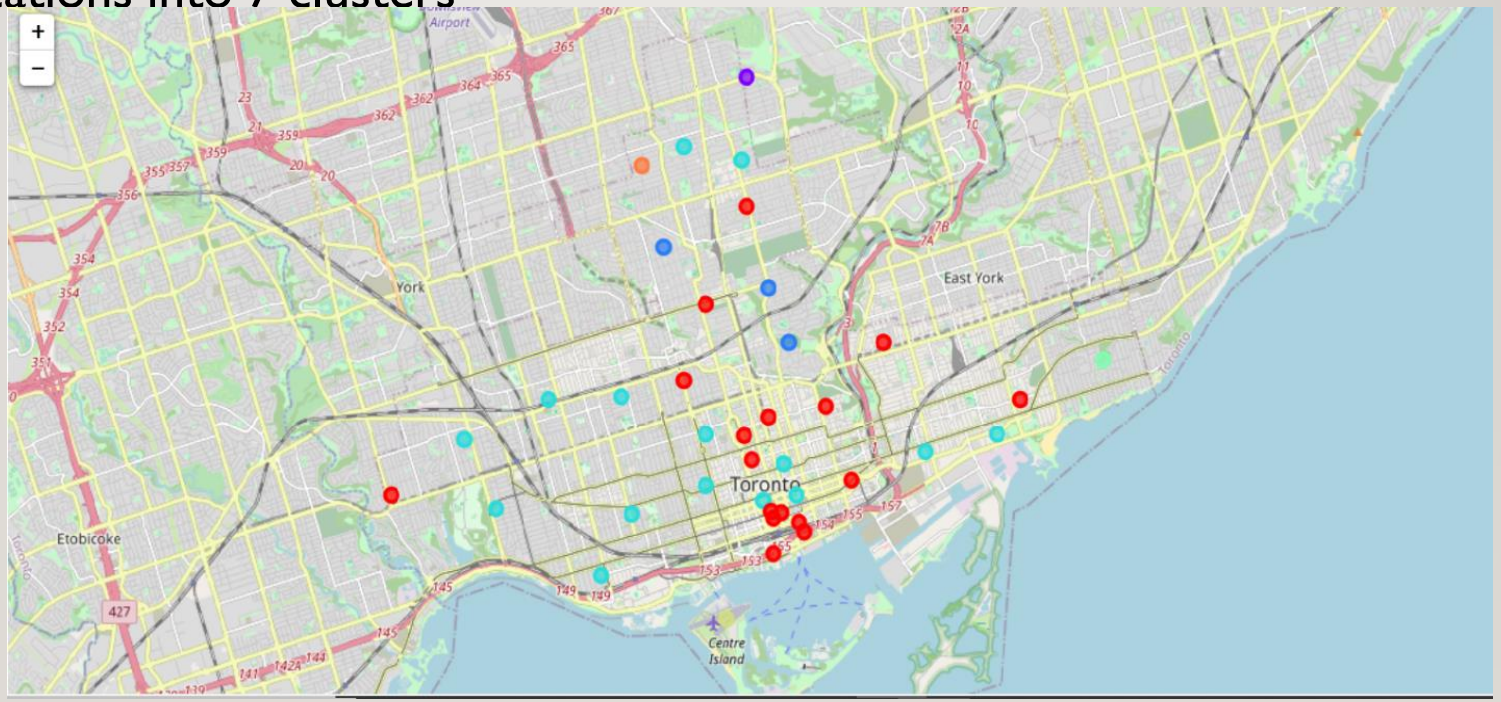
- Final clustering

```
# set number of clusters
kclusters = 7

main_toronto_grouped_clustering = main_toronto_grouped.drop('Neighborhood', 1)

# run k-means clustering
kmeans = KMeans(n_clusters=kclusters, random_state=0).fit(main_toronto_grouped_clustering)

# check cluster labels generated for each row in the dataframe
kmeans.labels_[0:38]
```

```
[175]: array([0, 3, 3, 5, 0, 3, 0, 0, 0, 3, 3, 0, 2, 3, 0, 3, 0, 3, 1, 3, 2, 3,
               3, 0, 0, 3, 2, 6, 0, 3, 0, 0, 3, 0, 0, 4, 0, 0], dtype=int32)
```

FINDINGS

- Dividing all locations into 7 clusters



FINDINGS

- Office Space : first cluster has lots of meeting joints like coffee shops , cafe , restaurant and hotel. this indicate that this cluster has lots of commercial activity in a day and has lots of office. If some one wants to target office goers can open store in this cluster.
- Open Space : This cluster clearly suggests that it has lots of open spaces and is ideal for adventure sports shops complementing trail and park already present in each location .
- Mixed Spaces : This above cluster has cafe , hotel , bar, discount store , clothing store etc . indicating it has mixed customers are with part residential and part commerical space presence and hence we name it Mixed use

FINDINGS

- Airport : This cluster clearly indicates the presence of Airport and allied services and anybody interested in airport related business activity must prefer this
- Residential space : This cluster has home service , garden and yoga studio, clearly indicating that its a residential area and accordingly somebody should plan a shop/outlet in this area
- Note : Cluster 2 and cluster 5 did not yielded in any meaningful insight and hence were excluded from the presentation.

SOLUTION/CONCLUSION

Sr. No	Name of clusters	Total number of locations	Suggestions
1	Office spaces	17	ideal for joints like coffee shops , restaurant ete
2	Open Spaces	3	Ideal for adventure sports shops
3	Mixed spaces	15	has mix of residential and commercial, good for shopping malls with joints
4	Air port	1	good for airport related joints/ services- duty free shops
5	Residential	1	good for home service, home decor etc

FUTURE

- We can further evolve this project so that it can be used for multiple cities and/or further drilling of shortlisted locations.
- Also we can use additional data like per capita income for each location to make further in depth finding of customer choices and locations .

THANK YOU