IBM DATASCIENCE CAPTSTONE PROJECT

THE BATTLE OF NEIGHBORHOODS CHOOSING A LOCATION FOR OPENING A WAFFLE SHOP IN DOWNTOWN TORONTO

BY RAGHU GOPALKRISHNAN

BACKGROUND

- A SUCCESSFUL EUROPEAN WAFFLE MANUFACTURER WANTS TO EXPAND THEIR FOOTPRINT IN NORTH AMERICA
- NORTH AMERICA IS A POTENTIAL MARKET AND PRESENTS NUMEROUS OPPORTUNITIES TO EXPAND THE WAFFLE BUSINESS
- THEIR MARKET RESEARCH TEAM HAS IDENTIFIED THEY HAVE TAKEN A DECISION TO INVEST IN DOWNTOWN TORONTO.
- THEY HAVE ASKED OUR DATASCIENCE COMPANY TO FIND THE RIGHT NEIGHBOURHOOD TO START A FRANCHISE BASED ON RELEVANT DATA, EXPLORATION AND ANALYSIS.

PROBLEM STATEMENT

- SINCE TORONTO IS A LARGE AREA CHOOSING THE RIGHT NEIGHBOURHOOD TO OPEN THE WAFFLE SHOP IS CRUCIAL
- DUE TO HIGH REAL ESTATE PRICES IT IS VITAL THAT THE RIGHT LOCATION BE CHOSEN TO AVOID BUSINESS LOSSES.
- IF A RIGHT NEIGHBOURHOOD IT COULD BE DETRIMENTAL TO THEIR EXPANSION PLANS IN NORTH AMERICA.

DATA REQUIREMENTS

- LIST OF NEIGHBOURHOODS IN TORONTO ACCORDING TO BOROUGHS. THIS INFORMATION WILL BE SOURCED FROM WIKIPEDIA.
- GEOSPATIAL INFORMATION PER POSTAL CODE FROM A CSV FILE.
- LIST OF VENUES PER NEIGHBOURHOOD, THIS INFORMATION WILL BE OBTAINED FORM THE FOURSQUARE API.

AUDIENCE

- THE TARGET AUDIENCE FOR THIS PROJECT WOULD BE THE MANAGEMENT FOR THE WAFFLE COMPANY WHO ARE INTERESTED IN OPENING THEIR FRANCHISE IN NA
- THIS PAPER WILL ALSO INTEREST STUDENTS OF THE DATASCIENCE FIELD AS A REFERENCE

DATA DESCRIPTION

- LIST OF NEIGHBOURHOODS IN TORONTO ACCORDING TO BOROUGHS. THIS INFORMATION WILL BE SOURCED FROM WIKIPEDIA

Note: There are no rural FSAs in Toronto, hence no postal codes start with M0.

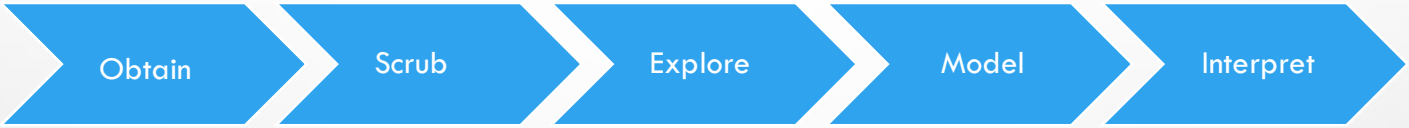
Postcode	Borough	Neighbourhood
M1A	Not assigned	Not assigned
M2A	Not assigned	Not assigned
M3A	North York	Parkwoods
M4A	North York	Victoria Village

- GEOSPATIAL INFORMATION PER POSTAL CODES FROM A CSV FILE.

1	Postal Coc	Latitude	Longitude
2	M1B	43.80669	-79.1944
3	M1C	43.78454	-79.1605

- LIST OF VENUES PER NEIGHBOURHOOD
  - THIS WILL BE OBTAINED USING THE FOURSQUARE API.

METHODOLOGY



Step	Task
Obtain	Read data from Wikipedia, convert to DF Obtain Geospatial Data
Scrub	Remove unwanted data, retain only needed columns. Data wrangling
Explore	Use Foursquare API to obtain venue details
Model	Use K-means clustering methodology
Interpret	Interpret clustering results, draw conclusions

SCOPE

- WE WILL LIMIT THE SCOPE OF THIS PROJECT TO THE 'DOWNTOWN TORONTO' AREA
- RENTAL PRICES, OTHER INFLUENCING FACTORS WHICH MIGHT BE DETRIMENTAL TO OPEN A SHOP IN AN AREA ARE NOT WITHIN THE SCOPE OF THIS PROJECT.
- CUSTOMER DEMAND, TASTE, PREFERENCES ETC ARE NOT WITHIN THE SCOPE OF THIS PROJECT.

DATA COLLECTION AND PROCESSING

- STEP-1WE USED THE PYTHON WIKIPEDIA API TO SCRAPE THE TABULAR DATA FOR THE LIST OF NEIGHBOURHOODS IN TORONTO. WE CREATED A DATAFRAME AS FOLLOWS

	Postalcode	Borough	Neighborhood
1	M1A	Not assign	Not assigned
2	M2A	Not assign	Not assigned
3	M3A	North Yorl	Parkwoods
4	M4A	North Yorl	Victoria Village
5	M5A	Downtowı	Harbourfront
6	M5A	Downtowı	Regent Park

- STEP-2 HERE WE DID SOME DATA CLEANING AND DATA WRANGLING TO REMOVE UNWANTED DATA LIKE ROWS WITH BOROUGHS AS ‘NOT ASSIGNED’ AND SOME DATA WRANGLING

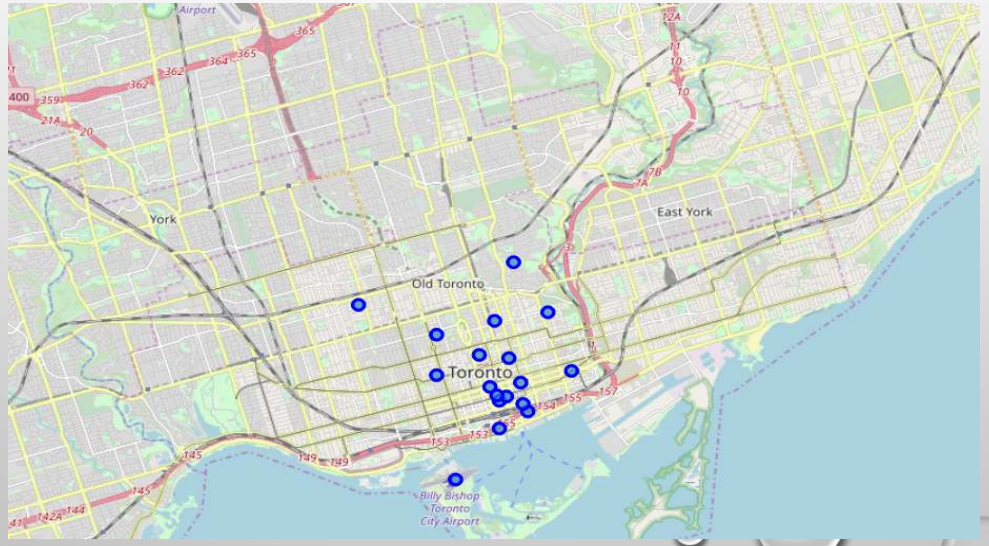
DATA COLLECTION AND PROCESSING CONTD..

- STEP-3WE DID SOME DATA AGGREGATION TO GROUP ON THE POSTAL CODE
- STEP-4 WE USED THE GEOSPATIAL FILE TO GET THE LONGITUDE AND LATITUDE FOR THE POSTAL CODES BY MERGING TWO DATAFRAMES ON THE POSTAL CODE COLUMN

	Postalcode	Borough	Neighborhood
1	M1B	Scarborou	Rouge , Malvern
2	M1C	Scarborou	Highland Creek , Rouge Hill , Port Union
3	M1E	Scarborou	Guildwood , Morningside , West Hill
4	M1G	Scarborou	Woburn
5	M1H	Scarborou	Cedarbrae
6	M1J	Scarborou	Scarborough Village
7	M1K	Scarborou	East Birchmount Park , Ionview , Kennedy Park

DATA COLLECTION AND PROCESSING CONTD..

- STEP-5 WE CREATED A MAP OF DOWNTOWN TORONTO AND SUPERIMPOSED THE NEIGHRBOURHOODS USING FOLIUM



DATA COLLECTION AND PROCESSING CONTD..

- STEP-7 WE USE THE FOURSQUARE API TO EXPLORE VENUES CLOSE TO OUR NEIGHBOURHOODS

	Neighborhood	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	
	Adelaide , King , Richmond	Steakhouse	Café	Pizza Place	Hotel	Asian Restaurant	Seafood Restaurant	Noodle House	Op
	Berczy Park	Café	Cocktail Bar	Beer Bar	Farmers Market	Seafood Restaurant	Park	Italian Restaurant	Ch
	CN Tower , Bathurst Quay , Island airport , Ha...	Airport Service	Airport Lounge	Airport Terminal	Airport	Airport Food Court	Airport Gate	Boat or Ferry	
	Cabbagetown , St. James Town	Bakery	Restaurant	Italian Restaurant	Café	Coffee Shop	Market	Pub	
	Central Bay Street	Coffee Shop	Italian	Bubble Tea	Spa	Seafood	Bar	Sandwich	

DATA COLLECTION AND PROCESSING CONTD..

- STEP-8 WE USE THE K-MEANS ALGORITHM TO CLUSTER THE NEIGHBOURHOODS USLING CLUSTER=5

15	M5W	Downtown Toronto	Stn A PO Boxes 25 The Esplanade	-79.374846	2.38	0	Café	Restaurant	Cocktail Bar	Farmers Market	Seafood Restaurant	Beer Bar	Pub	Chee Sh

Cluster-2

```
3]: Downtown_Toronto_merged.loc[Downtown_Toronto_merged['Cluster Labels'] == 1, Downtown_Toronto_merged.columns[[0] + [1] + [2] + lis
```

3]:

	Postalcode	Borough	Neighborhood	Longitude	Rental Index	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 Mo Comm Ven
14			CN Tower												

ANALYSIS OF DATA

- AFTER THE LAST STEP OF CLUSTERING USING K-MEANS ALGORITHM WE BREAKDOWN THE NEIGHBOURHOODS INTO CLUSTERS OF SIMILARITY TO HELP US WITH OUR ANALYSIS.
- IN THE ANALYSIS STEP WE ANALYSE THE NEIGHBOURHOODS ON VARIOUS PARAMETERS OF RENTAL INDEX, VENUES CLOSEBY ETC TO HELP US MAKE A DETERMINATION ON WHICH NEIGHBOURHOOD IS SUITABLE TO OPEN OUR WAFFLE BUSINESS.

DISCUSSION OF RESULTS

- THE FINDINGS FROM OUR ANALYSIS ARE
- CLUSTER 1 AND 4 HAS A HIGH DENSITY OF CAFES, RESTAURANTS, COFFEE SHOPS
- CLUSTER 2,3,5 HAS ONLY ONE NEIGHBORHOOD
- IN CLUSTER 1 CHURCH AND WELLESLEY, RYERSON , GARDEN DISTRICT SEEM SUITABLE VENUES DUE TO THEIR LOW RENTAL INDEX
- IN CLUSTER 5 CABBAGETOWN AND HARBOUR FRONT SEEM SUITABLE VENUES. HARBOUR FRONT IN PARTICULAR SEEMS INTERESTING SINCE THERE ARE BREAKFAST SPOTS AND BAKERY SHOPS AND WAFFLE COULD BE A POTENTIAL ITEM WHICH WILL BE CONSUMED.
- THE OTHER NEIGHBOURHOODS HAVE A RELATIVELY HIGHER RENTAL INDEX AND COULD BE A RISKY PROPOSITION IF THE BUSINESS DOESN'T TAKE OFF WELL

CONCLUSION

- THE PURPOSE OF OUR PROJECT WAS TO IDENTIFY SUITABLE NEIGHBORHOOD'S IN THE DOWNTOWN TORONTO AREA WHICH ARE SUITABLE TO OPEN A WAFFLE FRANCHISE SO THAT THE CLIENT MANAGEMENT OF THE WAFFLE COMPANY CAN NARROW DOWN TO SUITABLE AREAS TO START THEIR FRANCHISE BUSINESS.
- USING THE FOURSQUARE API WE WERE ABLE TO IDENTIFY VENUES CLOSER TO THE NEIGHBORHOOD'S AND THEN ALONGWITH THE K-MEANS CLUSTERING ALGORITHM WE WERE ABLE TO GROUP THE DATA ON FEATURE SIMILARITY IN ORDER TO FIND SUITABLE LOCATIONS.
- THE FINAL DECISION TO SELECT A NEIGHBOURHOOD WILL BE DONE THE EUROPEAN WAFFLE MANUFACTURER BASED ON COMMERCIAL/LEGAL PARAMETERS, MARKET RESEARCH, CUSTOMER TASTES ETC.

REFERENCES

- HTTPS://EN.WIKIPEDIA.ORG/WIKI/LIST_OF_POSTAL_CODES_OF_CANADA:_M
- HTTPS://FOURSQUARE.COM/
- HTTPS://EN.WIKIPEDIA.ORG/WIKI/K-MEANS_CLUSTERING
- HTTPS://SCIKIT-LEARN.ORG/STABLE/MODULES/GENERATED/SKLEARN.CLUSTER.KMEANS.HTML