CAPSTONE PROJECT - BATTLE OF NEIGHBOURHOODS

SIMKA JANEVSKA

MAY,2021

INTRODUCTION

THE PURPOSE OF THIS PROJECT IS TO HELP PEOPLE IN EXPLORING BETTER FACILITIES AROUND THEIR NEIGHBORHOOD

THE AIM OF THIS PROJECT IS TO DEMONSTRATE HOW FOURSQUARE DATA CAN BE USED TO EXPLORE AND COMPARE NEIGHBOURHOODS OR CITIES OF CHOICE AND WHICH PROBLEMS CAN IT SOLVE.

PROBLEM DESCRIPTION

FOR THIS PROJECT, I CHOSE A HYPOTHETICAL BUSINESS PROBLEM. IMAGINE THAT I HAVE A FRIEND, WHO GOT JOB OFFER IN NORTH YORK, TORONTO, SO HE AND HIS FAMILY ARE MOVING TO NORTH YORK. IT IS ONLY NATURAL, THAT HE WANTS TO MOVE TO A NEIGHBOURHOOD, WHICH HAS EASY ACCESS TO SCHOOL, CAFE, SUPER MARKET, MEDICAL SHOPS, GROCERY SHOPS, MALL, THEATRE, HOSPITAL, ETC.. HE ASKED ME TO ASSIST HIM IN CREATING A DETAILED ANALYSIS OF THE AREAS IN NORTH YORK, AND CHOOSE WHICH IS BEST FOR HIM.

DATA DESCRIPTION

- DATA LINK: HTTPS://EN.WIKIPEDIA.ORG/WIKI/LIST_OF_POSTAL_CODES_OF_CANADA:_M
- FOURSQUARE API

METHODOLOGY

ALL THE ANALYSIS WAS DONE USING PYTHON-PROGRAMMING LANGUAGE WITH THE FOLLOWING PYTHON'S LIBRARIES:

- PANDAS: FOR CREATING AND MANIPULATING DATAFRAMES.
- FOLIUM: PYTHON VISUALIZATION LIBRARY WOULD BE USED TO VISUALIZE THE NEIGHBORHOODS CLUSTER DISTRIBUTION OF USING INTERACTIVE LEAFLET MAP.
- SCIKIT LEARN: FOR IMPORTING K-MEANS CLUSTERING.
- JSON: LIBRARY TO HANDLE JSON FILES.
- XML: TO SEPARATE DATA FROM PRESENTATION AND XML STORES DATA IN PLAIN TEXT FORMAT.
- GEOCODER: TO RETRIEVE LOCATION DATA.
- BEAUTIFUL SOUP AND REQUESTS: TO SCRAP AND LIBRARY TO HANDLE HTTP REQUESTS.
- MATPLOTLIB: PYTHON PLOTTING MODULE.

RESULTS

FOLIUM LIBRARY WAS USED TO VISUALIZE MAP OF NORTH YORK AND NEIGHBOURHOODS

DATAFRAME THAT DISPLAY THE TOP 10 VENUES FOR EACH NEIGHBORHOOD:

	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	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Bathurst Manor, Wilson Heights, Downsview North	Coffee Shop	Bank	Gas Station	Restaurant	Fried Chicken Joint	Diner	Deli / Bodega	Ice Cream Shop	Mobile Phone Shop	Bridal Shop
1	Bayview Village	Japanese Restaurant	Chinese Restaurant	Café	Bank	Discount Store	Construction & Landscaping	Convenience Store	Cosmetics Shop	Deli / Bodega	Department Store
2	Bedford Park, Lawrence Manor East	Coffee Shop	Italian Restaurant	Restaurant	Sandwich Place	Butcher	Grocery Store	Juice Bar	Comfort Food Restaurant	Café	Fast Food Restaurant
3	Don Mills North	Caribbean Restaurant	Café	Gym	Japanese Restaurant	Women's Store	Diner	Comfort Food Restaurant	Construction & Landscaping	Convenience Store	Cosmetics Shop
4	Don Mills South	Coffee Shop	Clothing Store	Restaurant	Gym	Italian Restaurant	Discount Store	Bike Shop	Beer Store	Sandwich Place	Grocery Store

- I'VE DECIDED TO USE K-MEANS CLUSTERING ALGORITHM TO GROUP NEIGHBOURHOODS INTO 5 DISTINCT GROUPS, BASED ON TOP 10 VENUES IN EACH NEIGHBOURHOOD.
- MOST OF THE NEIGHBORHOODS FALL INTO CLUSTER 2 WHICH HAS MOST VARIETY OF VENUES LIKE:COFFEE SHOP, CLOTHING STORE, DIFFEREN KINDS OF RESTAURANTS, PARK ETC. CLUSTER 1 HAS MOST COMMON VENUE FABRIC SHOP, CLUSTER 3 HAS ONE MOST COMMON VENUE AS FOOD TRUCK, AND CLUSTER 4 AND FIVE HAS PARK AS MOST COMMON VENUE. NOW WE CAN CONCLUDE THAT BEST PLACE TO LIVE IN NORTH YORK WILL BE CLUSTER 2.

CONCLUSION

FOURSQUARE DATA AND K-MEANS ALGORITHM, ALONGSIDE WITH FOLIUM MAPPING TOOLS CAN BE POWERFUL TOOLS HELPING PEOPLE LIVING IN LARGE CITIES TO MAKE EDUCATED DECISIONS. THIS APPROACH CAN BE REPLICATED FOR VARIOUS OTHER TASKS, E.G. CHOOSING A RESTAURANT OR A STORE LOCATION.