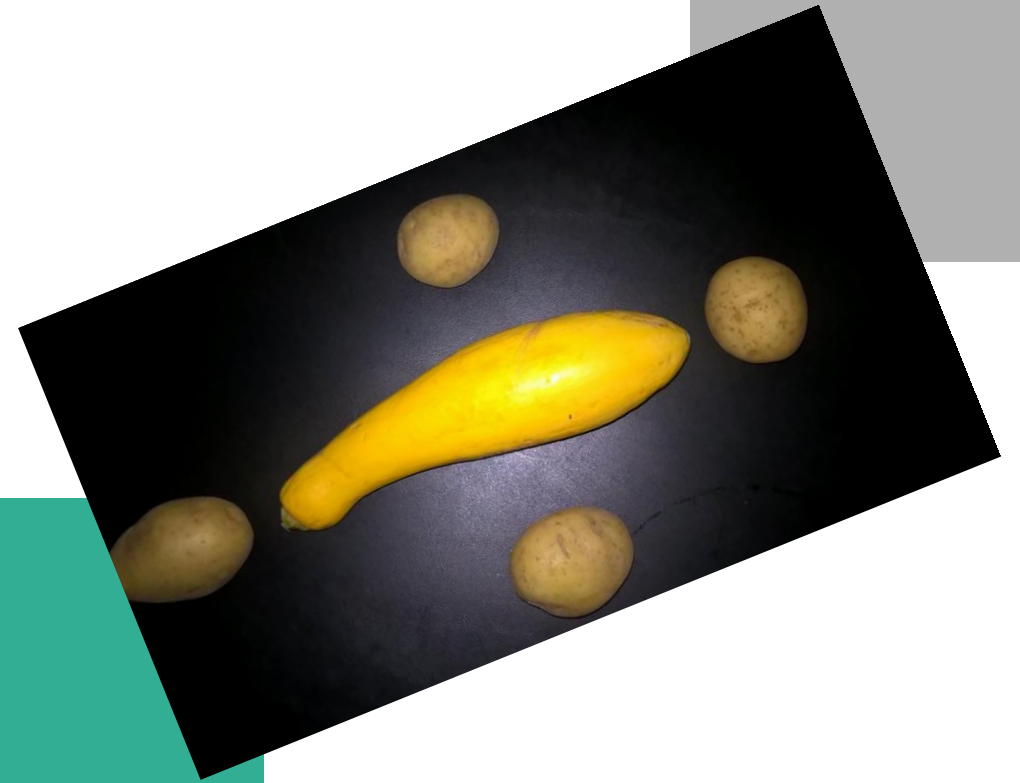
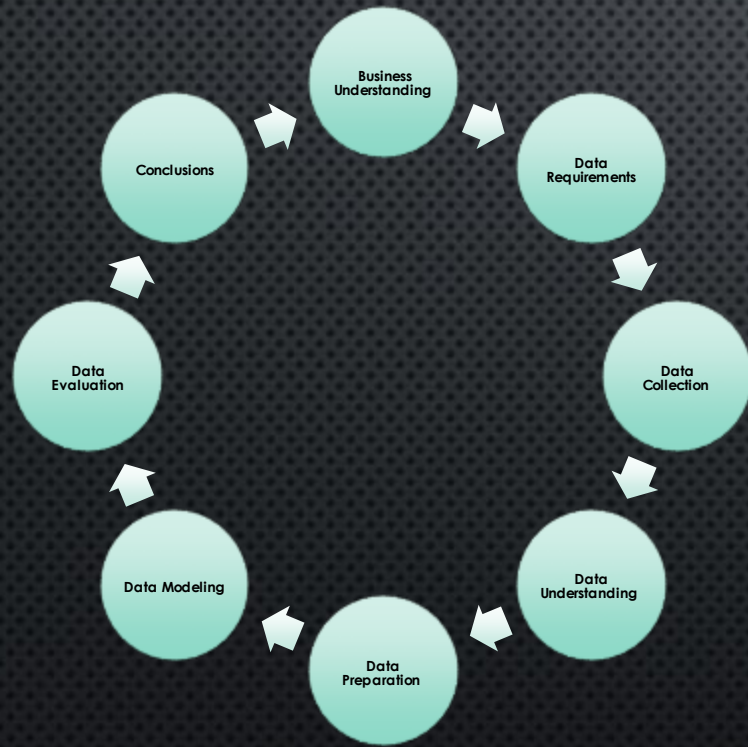




CAPSTONE INTRODUCTION TO DATA SCIENCE

GROCERY SHOPPING IN FOR RAW VEGETABLES EMBATTLED NEIGHBORHOOD



BUSINESS UNDERSTANDING

- INTRODUCTION
- OVGUIDE
 - A SPECIALIZATION ARM OF ACME INC RENTERS AND REAL ESTATE MANAGEMENT COMPANY.
 - WANTS TO OFFER A HOLISTIC PICTURE FOR ITS RENTER'S GUIDE.
 - THE PROPERTIES INCLUDED NEIGHBORHOODS IN TORONTO, ONTARIO AND NEW YORK, NEW YORK.
 - WANTS TO PROMOTE ITS PROPERTIES AS PEDESTRIAN FRIENDLY FOR LOCAL SHOPPING .
 - WANTS TO FOCUS ON EVERYDAY LIVING.
 - THE QUESTION FOR THE STUDY WAS WHETHER THE PUBLICLY LISTED DATA SUPPORTED THE TARGETED THE PERCEPTION.
 - STUDY DECIDED TO FOCUS ON GROCERY SHOPPING FOR RAW VEGETABLES AS THE MOST BASIC REQUIREMENT.

DATA REQUIREMENTS

- NEIGHBORHOOD DATA
- BOROUGH AND NEIGHBORHOOD DATA FOR TORONTO, ONTARIO IN CANADA AS WELL AS LATITUDE AND LONGITUDE FOR EACH OF THESE NEIGHBORHOODS.
- BUSINESS ADDRESS WITH OFFERINGS NAMELY GROCERS OFFERING RAW VEGETABLES IN THE NEIGHBORHOOD ASSOCIATED WITH THE POSTAL OR ZIP CODES AS WELL AS LATITUDE AND LONGITUDE OF SUCH ADDRESS.
- ROUTING ENGINE DATA FOR FOOT TRAFFIC TO THE GROCERS FROM POSTAL/ZIP CODE LATITUDE AND LONGITUDE LISTINGS AND FOR AVERAGE GAIT AND WEIGHT AND FITNESS (5 KM/H).
- THE DATA WILL LIST TRAVEL TIMES AND THE DISTANCE COVERED.
- DATA SOURCES
 - WIKIPEDIA AND CITY OF NEW YORK WEBSITE FOR COUNTY AND POSTAL OR ZIP CODES.
 - FOURSQUARE FOR BUSINESS LISTINGS IN THESE NEIGHBORHOODS.
 - GEONAMES.ORG, GRAPHHOPPER AND OPENSTREETMAP FOR GEOGRAPHICAL INFORMATION SYSTEMS (GIS) DATA.

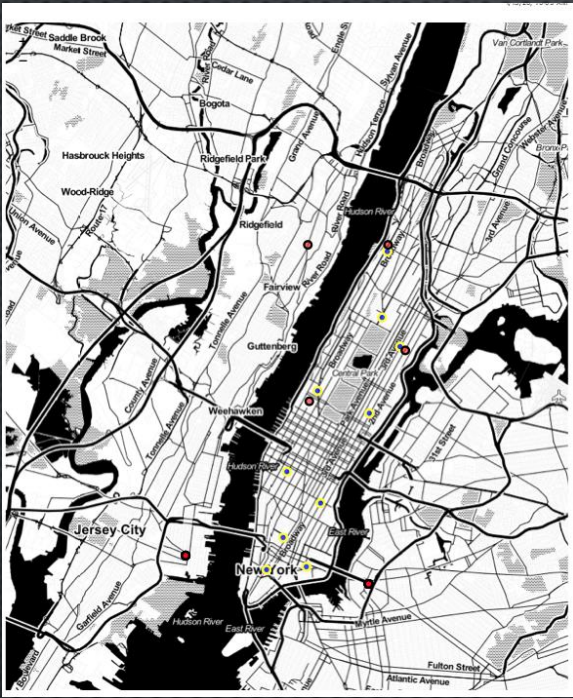
DATA COLLECTION

New York Neighborhood Data

	Borough	Neighborhood	ZIP Codes
0	Bronx	Central Bronx	10453, 10457, 10460
1		Bronx Park and Fordham	10458, 10467, 10468
2		High Bridge and Morrisania	10451, 10452, 10456
3		Hunts Point and Mott Haven	10454, 10455, 10459, 10474
4		Kingsbridge and Riverdale	10463, 10471
5		Northeast Bronx	10466, 10469, 10470, 10475
6	Brooklyn	Southeast Bronx	10461, 10462, 10464, 10465, 10472, 10473
7		Central Brooklyn	11212, 11213, 11216, 11233, 11238
8		Southwest Brooklyn	11209, 11214, 11228
9		Borough Park	11204, 11218, 11219, 11230
10		Canarsie and Flatlands	11234, 11236, 11239
11		Southern Brooklyn	11223, 11224, 11229, 11235
12		Northwest Brooklyn	11201, 11205, 11215, 11217, 11231
13		Flatbush	11203, 11210, 11225, 11226
14		East New York and New Lots	11207, 11208
15		Greenpoint	11211, 11222

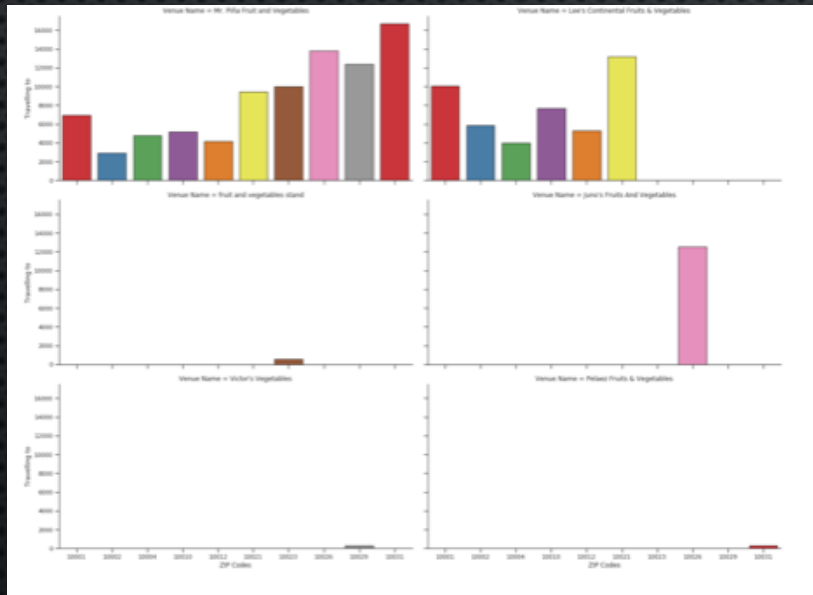
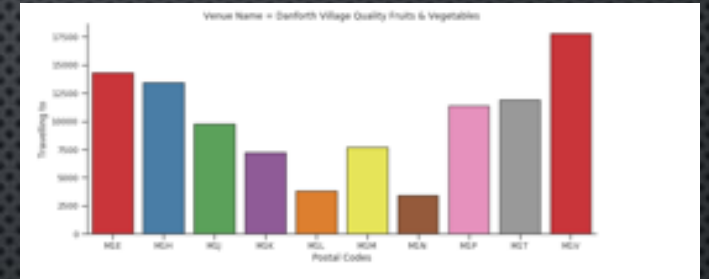
Toronto Ontario Neighborhood Data

	Postcode	Borough	Neighborhood	latitude	longitude
0	M1E	Guildwood,	Guildwood,	43.7678	-79.1866
1	M1H	Cedarbrae,	Cedarbrae,	43.7686	-79.2389
2	M1J	Scarborough Village,	Scarborough Village,	43.7464	-79.2323
3	M1K	East Birchmount Park, Ionview,	East Birchmount Park, Ionview,	43.7298	-79.2639
4	M1L	Clairlea,	Clairlea,	43.7122	-79.2843
5	M1M	Cliffcrest, Scarborough Village West,	Cliffcrest, Scarborough Village West,	43.7247	-79.2312
6	M1N	Birch Cliff, Cliffside West,	Birch Cliff, Cliffside West,	43.6952	-79.2646
7	M1P	Dorset Park, Scarborough Town Centre,	Dorset Park, Scarborough Town Centre,	43.7612	-79.2707
8	M1T	Clarks Corners, Sullivan,	Clarks Corners, Sullivan,	43.7812	-79.3036
9	M1V	Steeles East,	Steeles East,	43.8177	-79.2819



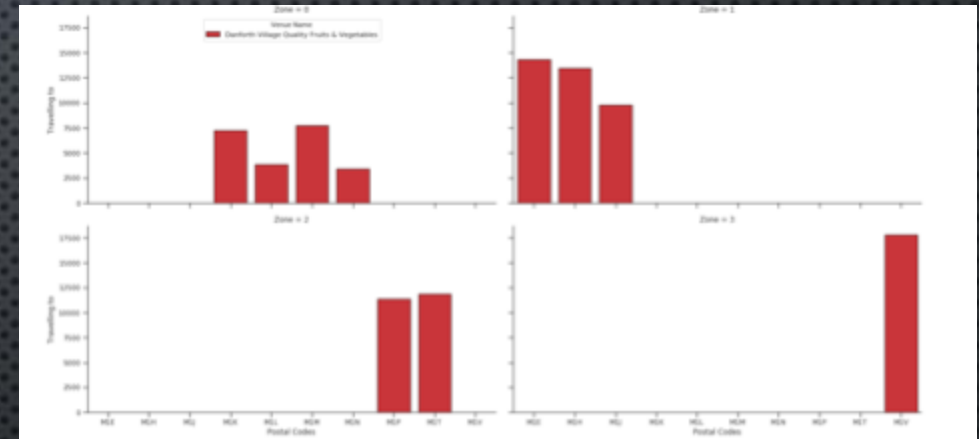
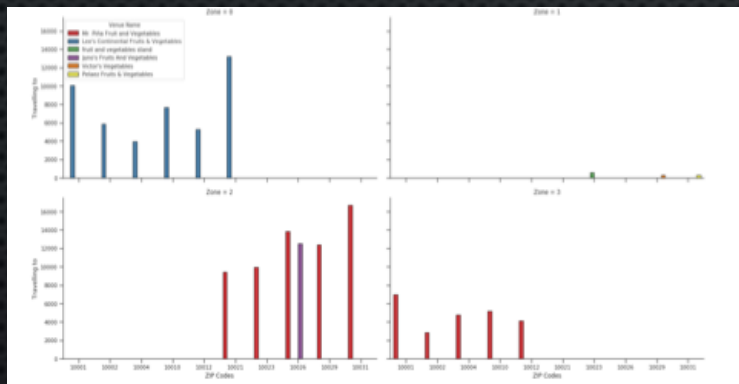
DATA UNDERSTANDING

Colorful Data and Map Plots of New York, NY and Toronto, ON Neighborhood and Grocers with listed physical address.



DATA MODELING & EVALUATION

Data plots of New York, NY and Toronto, ON neighborhood and Grocers as labeled/Zoned clusters using k-means clustering



CONCLUSIONS

- **CONCLUSIONS FROM THE DATA ANALYSIS OF TORONTO, ONTARIO.**

THE FOURSQUARE DATA FOR TORONTO, ONTARIO RETURNED ONLY ONE BUSINESS LISTED AS GROCER FOR VEGETABLES WITH AN ACTUAL PHYSICAL ADDRESS AND POSTAL CODE LISTED IN THE FOURSQUARE DATABASE.

TO INCREASE THE RELIABILITY THE FOURSQUARE DATABASE WAS LIMITED TO RETURN 3 BUSINESS LISTING PER QUERY OR CUSTOMER ZIP CODES IN THIS CASE.

THE DATA PLOTS SHOWS THAT THE TRAVELING DISTANCES TO THE SINGLE GROCER RANGE FROM 3.3 KM TO 17.5 KM FOR THE VARIOUS POSTAL CODES.

THE K-MEANS CLUSTERING IS PRIMARILY INFLUENCED BY THE WALKING DISTANCES TO THESE LOCATIONS. THE ZONE LABELED AS 0 TURNS OUT TO BE THE MOST HABITABLE ZONE AS FAR AS PEDESTRIAN WALKING TO THESE GROCERS ARE CONCERNED. THIS IS MOST PROMINENT IN THE BAR PLOT OF VARIOUS POSTAL CODES AND THE CORRESPONDING LABELS USING K-MEANS CLUSTERING.

- **CONCLUSIONS FROM THE DATA ANALYSIS OF NEW YORK, NEW YORK**

THE FOURSQUARE DATA FOR NEW YORK, NEW YORK RETURNED THREE BUSINESS LISTED AS GROCER FOR VEGETABLES WITH AN ACTUAL PHYSICAL ADDRESS AND ZIP CODE LISTED IN THE FOURSQUARE DATABASE.

TO INCREASE THE RELIABILITY THE FOURSQUARE DATABASE WAS LIMITED TO RETURN 2 BUSINESS LISTING PER QUERY OR CUSTOMER ZIP CODES IN THIS CASE.

THE DATA SHOW THAT THE WAKING DISTANCES TO THE GROCERS RANGE FROM 290 M TO 16.7 KM FOR VARIOUS POSTAL CODES.

THE K-MEANS CLUSTERING IS PRIMARILY INFLUENCED BY THE WALKING DISTANCES TO THESE LOCATIONS. THE ZONE LABELED AS 1 TURNS OUT TO BE LEAST HABITABLE AS FAR AS PEDESTRIAN WALKING TO THESE GROCERS ARE CONCERNED. ZONE 0 AND ZONE 3 OFFER ONLY ONE GROCER AS A CHOICE.

THANK YOU

NARSI SARATHY

NARSI16-EDU@YAHOO.COM

COURSERA IBM COGNITIVE CLASS IBM DATA SCIENCE PROFESSIONAL
CERTIFICATE