# THE BATTLE OF NEIGHBORHOODS

## APPLIED DATA SCIENCE CAPSTONE PROJECT

PATRICIO LOZA

# INTRODUCTION

- The purpose of this Capstone Project is to help people in exploring better facilities around their neighborhood. It will help people making smart and efficient decision on selecting great neighborhood out of numbers of other neighborhoods in Scarborough, Toranto.

- Lots of people are migrating to various states of Canada and needed lots of research for good housing prices and reputated schools for their children. This project is for those people who are looking for better neighborhoods. For ease of accessing to Cafe, School, Super market, medical shops, grocery shops, mall, theatre, hospital, like minded people, etc.

- This Capstone Project aim to create an analysis of features for a people migrating to Scarborough to search a best neighborhood as a comparative analysis between neighborhoods. The features include median housing price and better school according to ratings, crime rates of that particular area, road connectivity, weather conditions, good management for emergency, water resources both freash and waste water and excrement conveyed in sewers and recreational facilities.

- It will help people to get awareness of the area and neighborhood before moving to a new city, state, country or place for their work or to start a new fresh life.
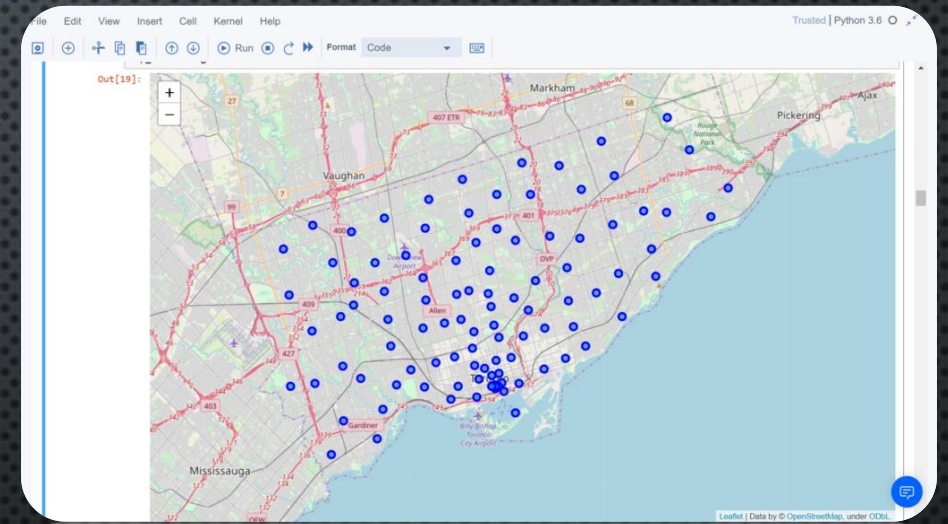
# DATA SECTION

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

  WILL USE SCARBOROUGH DATASET WHICH WE SCRAPPED FROM WIKIPEDIA ON WEEK 3. DATASET CONSISTING OF LATITUDE AND LONGITUDE, ZIP CODES.

- FOURSQUARE API DATA:

  WE WILL NEED DATA ABOUT DIFFERENT VENUES IN DIFFERENT NEIGHBORHOODS OF THAT SPECIFIC BOROUGH.

# 3. METHODOLOGY SECTION

- CLUSTERING APPROACH:

To compare the similarities of two cities, we decided to explore neighborhoods, segment them, and group them into clusters to find similar neighborhoods in a big city like New York and Toronto. To be able to do that, we need to cluster data which is a form of unsupervised machine learning: k-means clustering algorithm.

# 3. METHODOLOGY SECTION

- WORK FLOW:

USING CREDENTIALS OF FOURSQUARE API FEATURES OF NEAR-BY PLACES OF THE NEIGHBORHOODS WOULD BE MINED. DUE TO HTTP REQUEST LIMITATIONS THE NUMBER OF PLACES PER NEIGHBORHOOD PARAMETER WOULD REASONABLY BE SET TO 100 AND THE RADIUS PARAMETER WOULD BE SET TO 500.

# RESULTS SECTION

- THE LOCATION:

SCARBOROUGH IS A POPULAR DESTINATION FOR NEW IMMIGRANTS IN CANADA TO RESIDE. AS A RESULT, IT IS ONE OF THE MOST DIVERSE AND MULTICULTURAL AREAS IN THE GREATER TORONTO AREA, BEING HOME TO VARIOUS RELIGIOUS GROUPS AND PLACES OF WORSHIP. ALTHOUGH IMMIGRATION HAS BECOME A HOT TOPIC OVER THE PAST FEW YEARS WITH MORE GOVERNMENTS SEEKING MORE RESTRICTIONS ON IMMIGRANTS AND REFUGEES, THE GENERAL TREND OF IMMIGRATION INTO CANADA HAS BEEN ONE OF ON THE RISE.

- FOURSQUARE API:

THIS CAPSTONE PROJECT HAVE USED FOUR-SQUARE API AS ITS PRIME DATA GATHERING SOURCE AS IT HAS A DATABASE OF MILLIONS OF PLACES, ESPECIALLY THEIR PLACES API WHICH PROVIDES THE ABILITY TO PERFORM LOCATION SEARCH, LOCATION SHARING AND DETAILS ABOUT A BUSINESS.
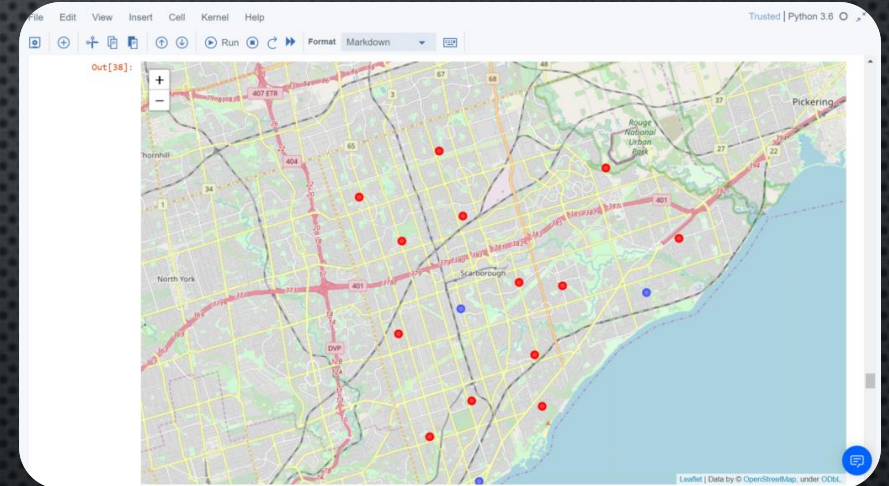
# DISCUSSION SECTION

- PROBLEM WHICH TRIED TO SOLVE:

- THE MAJOR PURPOSE OF THIS PROJECT, IS TO SUGGEST A BETTER NEIGHBORHOOD IN A NEW CITY FOR THE PERSON WHO ARE SHIFFTING THERE. SOCIAL PRESENCE IN SOCIETY IN TERMS OF LIKE MINDED PEOPLE. CONNECTIVITY TO THE AIRPORT, BUS STAND, CITY CENTER, MARKETS AND OTHER DAILY NEEDS THINGS NEARBY.

- SORTED LIST OF HOUSE IN TERMS OF HOUSING PRICES IN A ASCENDING OR DESCENDING ORDER

- SORTED LIST OF SCHOOLS IN TERMS OF LOCATION, FEES, RATING AND REVIEWS

# 6. CONCLUSION SECTION

- IN THIS CAPSTONE PROJECT, USING K-MEANS CLUSTER ALGORITHM I SEPARATED THE NEIGHBORHOOD INTO 10(TEN) DIFFERENT CLUSTERS AND FOR 103 DIFFERENT LATTITUDE AND LOGITUDE FROM DATASET, WHICH HAVE VERY-SIMILAR NEIGHBORHOODS AROUND THEM. USING THE CHARTS ABOVE RESULTS PRESENTED TO A PARTICULAR NEIGHBORHOOD BASED ON AVERAGE HOUSE PRICES AND SCHOOL RATING HAVE BEEN MADE.

- I FEEL REWARDED WITH THE EFFORTS AND BELIEVE THIS COURSE WITH ALL THE TOPICS COVERED IS WELL WORTHY OF APPRECIATION.
THIS PROJECT HAS SHOWN ME A PRACTICAL APPLICATION TO RESOLVE A REAL SITUATION THAT HAS IMPACTING PERSONAL AND FINANCIAL IMPACT USING DATA SCIENCE TOOLS.
THE MAPPING WITH FOLIUM IS A VERY POWERFUL TECHNIQUE TO CONSOLIDATE INFORMATION AND MAKE THE ANALYSIS AND DECISION BETTER WITH CONFIDENCE.