BATTLE OF NEIGHBORHOODS

Applied Data Science Capstone Project

Olga Tolmacheva

2019

PROBLEM STATEMENT

What is the objective of this project?

- Apply Data Science techniques to select a neighborhood in the city of Mississauga, well suited for a new Pizza Restaurant.

Who might benefit from this project?

- Entrepreneurs interested in launching new Pizza business;
- Established Pizza businesses looking to expand in Mississauga.

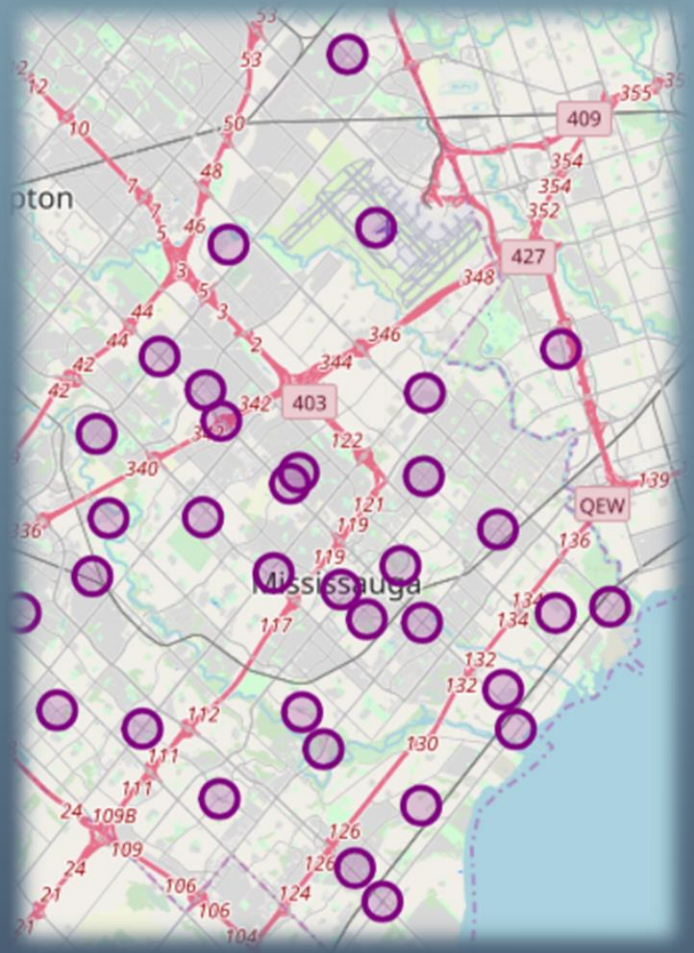
DATA

- **List of Mississauga neighborhoods**
- **Geographical coordinates of the neighborhoods**
- **Foursquare location data**
- **Population Age by Postal Code**

METHODOLOGY

- **Data Capture**
- **Data Analysis**
- **Data Segmentation**
- **Additional Analysis**

DATA CAPTURE



MISSISSAUGA NEIGHBORHOODS

- List of Neighborhoods

  from Wikipedia page using *Beautiful Soup*

- Geo coordinates

  using *geopy Nominatim*

LOCATION DATA

- Foursquare API

  Using *Explore* function

DATA ANALYSIS

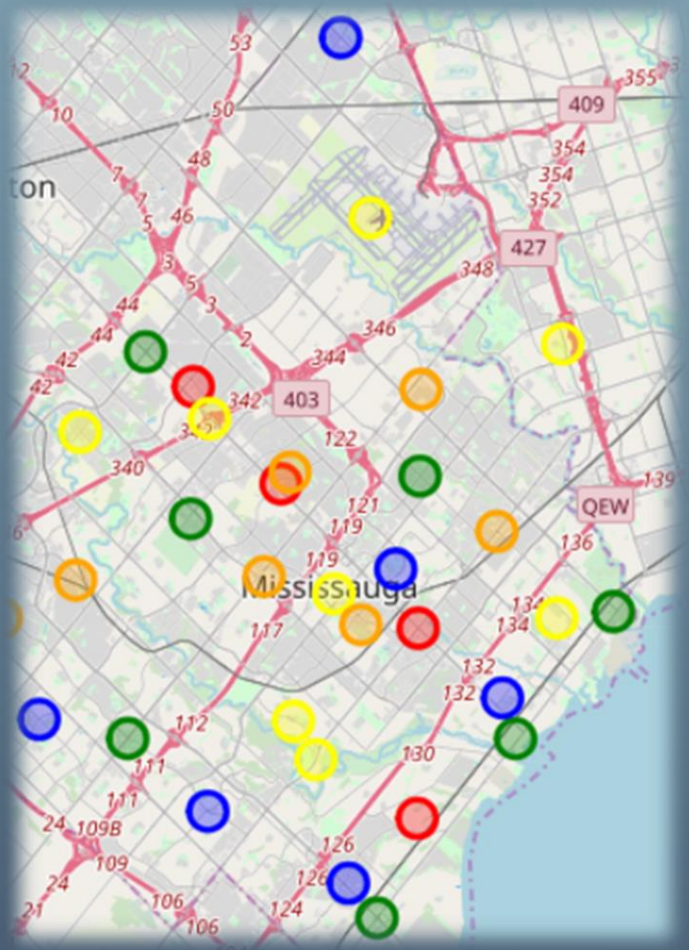
```
_venues['Venue Category'].unique()

y(['Park', 'Gym / Fitness Center', 'Pharmacy', 'Sandwich Place',
   'Bank', 'Tennis Court', 'Pizza Place', 'Skating Rink', 'Plaza',
   'Convenience Store', 'Trail', 'Burrito Place', 'Korean Restaurant
   'Indian Restaurant', 'Caribbean Restaurant',
   'Middle Eastern Restaurant', 'Café', 'Vietnamese Restaurant',
   'Portuguese Restaurant', 'Grocery Store', 'Fried Chicken Joint',
   'BBQ Joint', 'Mexican Restaurant', 'Fast Food Restaurant',
   'Chinese Restaurant', 'Mediterranean Restaurant', 'Supermarket',
   'Pakistani Restaurant', 'Paper / Office Supplies Store',
   'Coffee Shop', 'Bakery', 'Sushi Restaurant', 'Shopping Mall',
   'Bus Station', 'Train', 'Field', 'Performing Arts Venue',
   'Italian Restaurant', 'Burger Joint', 'Yoga Studio',
   'Electronics Store', 'College Gym', 'Sporting Goods Shop',
   'Cosmetics Shop', 'Department Store', 'Clothing Store',
   'Discount Store', 'Wings Joint', 'Deli / Bodega', 'Spa', 'Gym',
   'Event Service', 'Mountain', 'Construction & Landscaping', 'Farm
   'Hotel', 'Bar', 'Restaurant', 'Pool Hall', 'Irish Pub',
   'Ice Cream Shop', 'Tapas Restaurant', 'Poutine Place',
   'Harbor / Marina', 'Gastropub', 'Cocktail Bar',
   'Seafood Restaurant', 'Gas Station', 'Breakfast Spot',
   'Waterfront', 'Gift Shop', 'Motel', 'Health & Beauty Service',
   'Gun Range', 'Liquor Store', 'Japanese Restaurant',
   'Video Game Store', 'Salon / Barbershop', 'Platform',
   'Train Station', 'Falafel Restaurant', 'Bookstore',
   'Asian Restaurant', 'Tea Room', 'Beer Store', 'Optical Shop',
   'Pet Store', 'Record Shop', 'Shoe Store', "Women's Store",
   'Rental Car Location', 'Baseball Field', 'Airport',
   'Airport Terminal', 'Vegetarian / Vegan Restaurant', 'Diner',
   'Greek Restaurant', 'American Restaurant', 'Building',
   'Big Box Store', 'Transportation Service', 'Movie Theater',
   'New American Restaurant', 'Thai Restaurant', 'Dessert Shop',
   'Steakhouse', 'Smoothie Shop', 'Salad Place', 'Athletics & Sports
   'Afghan Restaurant', 'Shop & Service', 'Shopping Plaza',
   'Mobile Phone Shop', 'Pool', 'Donut Shop', 'Beer Bar',
   'Kids Store', 'Furniture / Home Store', 'Bubble Tea Shop',
   'Smoke Shop', 'Candy Store', 'Playground', 'Sports Club',
   'History Museum', 'Laser Tag', 'Thrift / Vintage Store',
   'Light Rail Station', 'Road'], dtype=object)
```

LOCATION DATA ANALYSIS

- **Fast Food**
  - Category contains ‘Pizza’, ‘Burger’, ‘Sandwich’
- **Restaurant**
  - Category contains ‘Restaurant’
- **Other Venue**
  - The rest

DATA SEGMENTATION



K-MEANS CLUSTERING

Label	Fast Food	Other Venue	Restaurant
4	0.188582	0.796266	0.015152
3	0.085826	0.591588	0.322586
1	0.075423	0.442082	0.482495
0	0.05894	0.739046	0.202014
2	0.004444	0.991111	0.004444

nue	Restaurant
5266	0.015152
1588	0.322586
2082	0.482495
9046	0.202014
1111	0.004444

	Neighborhood	Fast Food	Other Venue	Restaurant	Label	Posta
0	Mineola	0.250000	0.750000	0.000000	4	
1	Southdown	0.250000	0.750000	0.000000	4	
2	Churchill Meadows	0.181818	0.727273	0.090909	4	
3	Mississauga Valley	0.181818	0.818182	0.000000	4	
4	Malton	0.142857	0.857143	0.000000	4	

… still too broad. Requires …

ADDITIONAL ANALYSIS

ADDITIONAL ANALYSIS

	PostalCode	Age	Population
0	L4T	1	450
1	L4T	1	450
2	L4T	2	455
3	L4T	3	500
4	L4T	4	465

...

	PostalCode	Age	Population
1712	L5W	96	0
1713	L5W	97	5
1714	L5W	98	0
1715	L5W	99	0
1716	L5W	100	5

DEMOGRAPHIC DATA

- Age Demographic

  From Statistics Canada web site using *Beautiful Soup*

- Bin to Create Age Groups

  Babies and Toddlers: 0-4 years
  Children: 5-12 years
  Teens: 13-18 years
  Youth: 18-24 years
  Young Adults: 25-34 years
  Adults: 35-64 years
  Younger Seniors: 64-80 years
  Older Seniors: 80+ years

AgeGroup	AgeGroupTotal	AreaTotal	AgeGroupPct
Babies and Toddlers	2320	38445	0.060346
Children	4230	38445	0.110027
Teens	3010	38445	0.078294
Youth	3515	38445	0.091429
Young Adults	5605	38445	0.145793

PostalCode	TargetAgePopulation	T
L4T	12130	
L4W	4985	
L4X	5475	
L4Z	11960	
L5A	13310	

Define Target Age Group as Teens, Youth and Young Adults

Aggregate by Postal Code and determine share of population in Target Age Group

Normalize Target Age and Fast Food Preference characteristics and plot for comparison

```
df_top = df_resu
df_top.head(1)
```

	Neighborhood
1	Southdown

RESULTS

Based on high score in both the Target Age Share and the Fast Food Preference characteristics, **Southdown** is the recommended location for the new Pizza Place.

Thank You!