

# **IDENTIFYING UNIQUE LOCAL VENUES IN TORONTO**

Applied Data Science Capstone – Final Project

# CONTENTS

- Background and Introduction
- Business Problem and Value Delivered
- Data Acquisition and Cleaning
- Problem Solving Methodology
- Results and Observations
- Discussion
- Conclusion

# BACKGROUND AND INTRODUCTION

- Toronto is the largest city in Canada; it is a multicultural city with a diverse population and numerous local attractions
- In addition to millions of tourists and visitors from all over the world, several thousand people from different parts of Canada as well as the world move to Toronto every year





# BUSINESS PROBLEM AND VALUE DELIVERED

- While it is beneficial for Toronto's visitors to locate top food places, there is great value in identifying neighborhoods that offer some of the other popular yet unique venues such as pet stores, fitness centers, etc. that are an integral part of daily life
- As such information is less commonly available to newcomers than restaurant or historical landmark information, this project aims to identify top venues across different Toronto neighborhoods in four select categories:
  - Fitness
  - Pets
  - Outdoor Recreation
  - Transport
- Restaurants and other food places are not in scope for this analysis



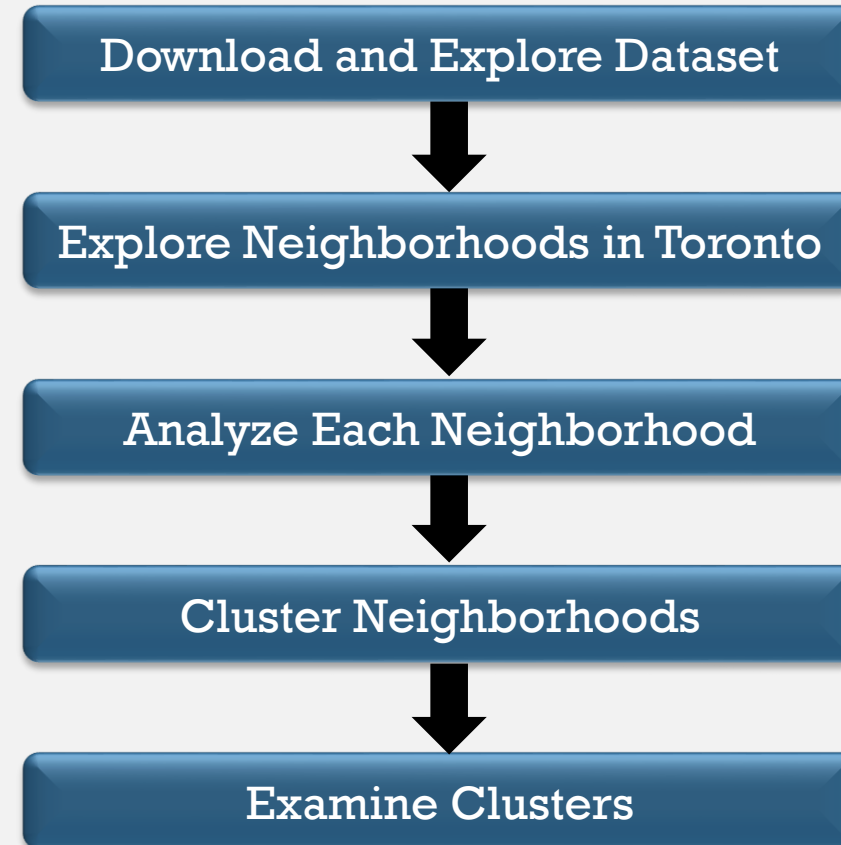
# DATA ACQUISITION AND CLEANING

- Neighborhoods in Toronto – **Wikipedia**
  - Cleaned the data and reduced it to boroughs that contain 'Toronto'
- Geographical coordinates (latitude, longitude) for Toronto – **Geospatial data (csv)**
  - Alternatively, **geopy** package can also be used
- Most common venues for given Boroughs of Toronto – **Foursquare API**
- Details of given venues of Toronto – **Foursquare API**



# PROBLEM SOLVING METHODOLOGY

- Download and Explore Dataset
- Explore Neighborhoods in Toronto
- Analyze Each Neighborhood
- Cluster Neighborhoods
- Examine Clusters



# METHODOLOGY –*Download and explore dataset*

- Read the data from the Wikipedia page and clean it
- Transform into a pandas dataframe
  - 3 essential features: Postal code, Borough, and Neighborhood
- Read geospatial data and merge with neighborhood dataframe
- Visualize Toronto map using *folium* library

	Postal code	Borough	Neighborhood
0	M1B	Scarborough	Malvern , Rouge
1	M1C	Scarborough	Rouge Hill , Port Union , Highland Creek
2	M1E	Scarborough	Guildwood , Morningside , West Hill
3	M1G	Scarborough	Woburn
4	M1H	Scarborough	Cedarbrae





# METHODOLOGY – *Explore Neighborhoods in Toronto*

- Define Foursquare credentials and version
- Get a neighborhood's name and geographical coordinates as a start
- Set a limit on the number of venues to return (100) and define a radius (500 meters)
- Clean and structure the data returned by the API (JSON file) and read into a dataframe
- Create a function to repeat the process for multiple Toronto neighborhoods

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	The Beaches	43.676357	-79.293031	Glen Manor Ravine	43.676821	-79.293942	Trail
1	The Beaches	43.676357	-79.293031	The Big Carrot Natural Food Market	43.678879	-79.297734	Health Food Store
2	The Beaches	43.676357	-79.293031	Grover Pub and Grub	43.679181	-79.297215	Pub
3	The Beaches	43.676357	-79.293031	Glen Stewart Park	43.675278	-79.294647	Park
4	The Beaches	43.676357	-79.293031	Upper Beaches	43.680563	-79.292869	Neighborhood



## METHODOLOGY – *Analyze Neighborhoods*

- Use one-hot encoding technique
- Obtain mean of the frequency for each venue category and group by neighborhood
- Convert into a pandas dataframe for further analysis

[illegible]

# METHODOLOGY – *Cluster Neighborhoods*

- Apply the *k-means Clustering* algorithm, with  $k=5$
- Run algorithm to cluster Toronto neighborhoods into 5 clusters
- Create dataframe that includes these clusters as well as the top venues for each neighborhood
- Visualize neighborhood clusters using *folium* library



# METHODOLOGY – *Examine Clusters*

- Examine each cluster to determine the venue categories that distinguish them
- Examples of findings from cluster analysis
  - First cluster: restaurants and coffee shops are the top venue categories
  - Third cluster: airport and related locations are the top venues

	Borough	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 Most Common Venue	10th Most Common Venue
0	East Toronto	0	Park	Pub	Trail	Health Food Store	Yoga Studio	Diner	Deli / Bodega	Department Store	Dessert Shop	Distribution Center
1	East Toronto	0	Greek Restaurant	Coffee Shop	Italian Restaurant	Ice Cream Shop	Bookstore	Furniture / Home Store	Frozen Yogurt Shop	Cosmetics Shop	Brewery	Bubble Tea Shop
2	East Toronto	0	Fast Food Restaurant	Pizza Place	Sandwich Place	Italian Restaurant	Restaurant	Steakhouse	Ice Cream Shop	Board Shop	Brewery	Burrito Place
3	East Toronto	0	Café	Coffee Shop	Gastropub	Bakery	Brewery	American Restaurant	Convenience Store	Sandwich Place	Cheese Shop	Pet Store
4	Central Toronto	0	Park	Bus Line	Swim School	Deli / Bodega	Eastern European Restaurant	Dumpling Restaurant	Donut Shop	Doner Restaurant	Dog Run	Distribution Center
5	Central Toronto	0	Dance Studio	Hotel	Park	Breakfast Spot	Gym	Sandwich Place	Department Store	Food & Drink Shop	Donut Shop	Doner Restaurant
6	Central Toronto	0	Clothing Store	Coffee Shop	Yoga Studio	Bagel Shop	Mexican Restaurant	Rental Car Location	Spa	Restaurant	Chinese Restaurant	Toy / Game Store
7	Central Toronto	0	Sandwich Place	Dessert Shop	Pizza Place	Café	Gym	Coffee Shop	Italian Restaurant	Sushi Restaurant	Deli / Bodega	Bar

# RESULTS AND OBSERVATIONS

- Toronto neighborhood cluster analysis indicates that restaurants, coffee shops, bars, and other food places are top venue categories, as is common in large cities like Toronto that attract many domestic and international visitors
- This project focuses only on select venue types and 4 main boroughs (East Toronto, Central Toronto, Downtown Toronto, and East Toronto):

Type	Top Venue Categories (Foursquare API)	Frequency (excl. food places)
Fitness	Gym, Yoga Studio, Dance Studio, Swim School	17%
Pets	Pet Store, Dog Run	7%
Outdoor Recreation	Playground, Park, Trail	11%
Transport	Light Rail Station, Bus Line, Airport, Boat/Ferry	7%



# DISCUSSION

- Excluding restaurants and food places, the four selected venue types (fitness, pets, outdoor recreation, and transport) account for 43% of the top 10 venue categories (returned by Foursquare) in 39 Toronto neighborhoods across four boroughs (East Toronto, Central Toronto, Downtown Toronto, and East Toronto)
- Gyms and fitness centers (including yoga studios) emerged as the top venue type with even distribution across the four boroughs
- Transport facilities are concentrated primarily in Central Toronto and Downtown
- West Toronto ranks low with regards to the presence of outdoor recreation and transport facilities

# CONCLUSION

- Explored and analyzed different neighborhoods in the city of Toronto
- Utilized postal codes, boroughs, neighborhoods, and geographical coordinates data
- Segmented and clustered each of the Toronto neighborhoods with the help of k-means Clustering algorithm
- Studied the distribution of local establishments from less common yet top venue categories of four select venue types: *Fitness*, *Pets*, *Outdoor Recreation*, and *Transport*
- Uncovered insights on local spots that are not generally highlighted in other readily-available tourist materials like restaurants, for the benefit of new residents and visitors