

Capstone Project - The Battle of Neighborhoods (Week 1)

To explore and analyze
The venues of neighborhoods of New York City and City of Toronto
Using Four Square Venue data



1.Introduction

To explore and analyze, the venues of the neighborhoods of New York City and City of Toronto using Four Square Venue data to ensure the ease of decision making.

1.1 Background

At times when we are visiting or moving to a new city, it becomes very difficult to discover or to choose restaurants, stores and other local business venues at the neighborhoods from user perspective and to start up a venue from a business point of view.

Decision making will be a huge task with lot of criteria in mind to select the venue based on distance, price tier, ratings and sometimes we often stuck at certain stages in deciding. In which the user contributions will also been an important element such as user likes of a venue, upload of venue photos by the user and user tips of a venue, which will be counted to the comparisons and the decision that we finally make out.

About the two cities

New York City is made up of five major areas or “boroughs” sitting where the Hudson River meets the Atlantic Ocean. some separated by rivers and connected via ferry or bridge. The five boroughs of **New York are Manhattan, Brooklyn, Queens, Staten Island and the Bronx.**

Toronto is a city of neighborhoods. each with its own style, vibe and scene. We might find ourself in a shopping mecca in the morning, a historic market around lunchtime, and surrounded by popular bars at night. One thing Toronto doesn't have a shortage of is shopping, whether it be outlet shopping, thrift or on trend pieces, there's a neighborhood for it. The neighborhoods are **East York, Etobicoke, North York, Old City of Toronto, Scarborough and the York**

1.2 Business Problem

To explore four-square venue data, to get new insights to recommend venue owners to start a new venue. To compare and conclude the better city on venues. To cluster & segment the venues to enhance user experience on selecting the venues basis the below analysis.

- To compare and distinguish venue similarities of the two cities using frequency distribution all columns with percentile and by various methods, parameters.
- To analyze if there is any correlation exist between the user likes, rating, price tier, tips, photos and distance.
- To perform clustering and segmentation for the venues based on four square venue details.

1.3 Target Audience

Business personnel who wants to startup a new venue, this analysis will be a providing a detailed insight for them on the opportunity available to launch a venue on certain venue categories to target the users in fulfilling business requirements.

Stakeholders to add new venues to the site in the respective city or neighborhoods having less than two or NIL venues in the categorical segment.

The analysis will give an overview to the existing venue owners to have an insight of the user behavior, patterns and trends, which will be useful to enhance their business or to revisit the existing business model.

The analysis will also be useful for the venue owners to view as how the users likes, tips and photos of the venue will bring effectiveness on the Price Tier and Ratings of the venue.

The stakeholders both internal and external will be benefited on the venue segmentation to select their venues in time and to classify their venues better.

2. Data Description

2.1 Cities to be Analyzed

- New York City
- City of Toronto

2.2 Data Source

2.2.1 Wikipedia: To download cities neighborhood list. Links given below,

Link 1: New York City

https://en.wikipedia.org/wiki/Neighborhoods_in_New_York_City

Link 2: City of Toronto

https://en.wikipedia.org/wiki/Demographics_of_Toronto_neighbourhoods

2.2.2 Geopy Library: To download location coordinates, latitude & longitude.

```
from geopy.geocoders import Nominatim
geolocator = Nominatim()
geolocator = Nominatim(user_agent="LN-Capstone-test")
adrs='New York'
location = geolocator.geocode(adrs)
tlatitude = location.latitude
tlongitude = location.longitude
```

2.2.3 FourSquare: To download places data & Augment basic venues details

a: Link 3 using Lat & Long: <https://api.foursquare.com/v2/venues/explore?>

b: Response details: Venue ID, name, category, id, address, distance, lat and lng.

c: Limit: 100

a: Link 4 using Venue ID: <https://api.foursquare.com/v2/venues/{}?>

b: Response details: likes.count, photos.count, rating, ratingSignals, reasons.count, tips.count, verified and price.tier

c: Limit: 50

2.3 Wikipedia Source Data Frame for two cities

2.3.1 City of Toronto

```
dftn.head()
```

	Neighborhood	Borough	Population	Average Income	lanper	lannam
0	Agincourt	Scarborough	44,577	25,750	19.3	Cantonese
1	Alderwood	Etobicoke	11,656	35,239	6.2	Polish
2	Alexandra Park	Old City of Toronto	4,355	19,687	17.9	Cantonese
3	Allenby	Old City of Toronto	2,513	245,592	1.4	Russian
4	Amesbury	North York	17,318	27,546	6.1	Spanish

2.3.2 New York City

```
dfny.head()
```

	Borough	code	Neighborhood
0	Bronx	CB 1	Melrose
1	Bronx	CB 1	Mott Haven
2	Bronx	CB 1	Port Morris
3	Bronx	CB 2	Hunts Point
4	Bronx	CB 2	Longwood

2.4 GEOSPY to fetch location Coordinates – Latitude and Longitude

2.4.1 City of Toronto – Coordinates data frame

```
dftnbrgl1.head(10)
```

	Borough	Latitude	Longitude
0	Scarborough,Toronto	43.773077	-79.257774
1	Etobicoke,Toronto	43.671459	-79.552492
2	Old City of Toronto,Toronto	43.653963	-79.387207
3	North York,Toronto	43.770817	-79.413300
4	East York,Toronto	43.691339	-79.327821
5	York,Toronto	43.689619	-79.479188

2.4.2 New York City – Coordinates data frame

```
dfnybcsv1.head(10)
```

	Borough	Latitude	Longitude
1	Bronx ,New York	40.850485	-73.840404
2	Brooklyn ,New York	40.650104	-73.949582
3	Manhattan ,New York	40.789624	-73.959894
4	Queens ,New York	40.652493	-73.791421
5	Staten Island ,New York	40.583456	-74.149605

2.5 FourSquare (Part 1): To download venue id based on latitude and longitude

2.5.1 City of Toronto data frame of venue id, name, category & distance

	name	categories	distance	lat	Ing	id	Borough
0	Knuckle Sandwich	Sandwich Place	545	43.696194	-79.328749	56df62a4498e96e8608b5e94	East York
1	East York Farmers' Market	Farmers Market	110	43.690482	-79.328509	4e3816b18877541e90eba62f	East York
2	Mon K Patisserie	Pastry Shop	636	43.696922	-79.329520	51b0bcb6498e4f0309a58a65	East York
3	The Wren	American Restaurant	987	43.682467	-79.328079	5155ca12e4b0e05526806bdf	East York
4	Rendez-Vous Restaurant Bar & Cafe	Ethiopian Restaurant	976	43.682570	-79.327544	4ad9221cf964a520671821e3	East York

2.5.2 New York City data frame of venue id, name, category & distance

	name	categories	distance	lat	Ing	id	Borough
0	LA Fitness	Gym / Fitness Center	154	40.849739	-73.841949	53d1b12d498ea039475dec73	Bronx
1	Residence Inn by Marriott New York The Bronx a...	Hotel	190	40.850020	-73.842579	54932887498ee0902b1ed511	Bronx
2	Starbucks	Coffee Shop	325	40.851371	-73.844087	57b25375498e76f51c083656	Bronx
3	Zeppieri & Sons Italian Bakery	Bakery	796	40.847119	-73.832057	4c9205941adc370460a134d1	Bronx
4	Empire Bagels	Bagel Shop	840	40.849392	-73.830527	4c249bffa852c9285f52e36c	Bronx

2.6 FourSquare (Part 2): To download user interaction details based on venue id

2.6.1 City of Toronto data frame of user likes, rating, price tier, photo count & tips

```
dftrvopt.head()
```

Unnamed: 0	id	likes.count	name	photos.count	price.tier	rating	ratingSignals	reasons.count	tips.count	verified
0	0.0 56df62a4498e96e8608b5e94	12	Knuckle Sandwich	14	1.0	8.4	20.0	0	7	False
1	0.0 4e3816b18877541e90eba62f	8	East York Farmers' Market	37	NaN	7.8	12.0	0	5	False
2	0.0 51b0bcb6498e4f0309a58a65	19	Mon K Patisserie	38	NaN	8.1	29.0	0	14	False
3	0.0 5155ca12e4b0e05526806bdf	132	The Wren	175	2.0	8.8	183.0	1	48	False
4	0.0 4ad9221cf964a520671821e3	35	Rendez-Vous Restaurant Bar & Cafe	24	2.0	9.0	53.0	1	23	False

2.6.2 New York City data frame of user likes, rating, price tier, photo count & tips

```
dfnyvopt.head()
```

Unnamed: 0	id	likes.count	name	photos.count	price.tier	rating	ratingSignals	reasons.count	tips.count	verified
0	0.0 53d1b12d498ea039475dec73	43	LA Fitness	1	NaN	8.6	52	1	4	True
1	0.0 54932887498ee0902b1ed511	25	Residence Inn by Marriott New York The Bronx a...	43	NaN	8.7	27	0	0	True
2	0.0 57b25375498e76f51c083656	26	Starbucks	11	1.0	8.5	31	0	3	True
3	0.0 4c9205941adc370460a134d1	31	Zeppieri & Sons Italian Bakery	27	1.0	9.4	41	1	10	True
4	0.0 4c249bffa852c9285f52e36c	31	Empire Bagels	19	1.0	8.6	39	1	8	False

2.7 Consolidation of final data frame from all 4 data sources

2.7.1 City of Toronto Final Data Frame

```
In [104]: FinVenueTor
```

	name	categories	distance	lat	lng	id	Borough	State	likes.count	photos.count	rating	ratingSignals	reasons.count
0	Knuckle Sandwich	Sandwich Place	545	43.696194	-79.326749	56df62a4498e96e8608b5e94	East York	Toronto	12	14	8.4	20.0	0
1	East York Farmers' Market	Farmers Market	110	43.690482	-79.328509	4e3816b18877541e90eba62f	East York	Toronto	8	37	7.8	12.0	0
2	Mon K Patisserie	Pastery Shop	636	43.696922	-79.329520	51b0bcb6498e4f0309a58a65	East York	Toronto	19	38	8.1	29.0	0
3	The Wren	American Restaurant	987	43.682467	-79.328078	5155ca12e4b0e05526806bdf	East York	Toronto	132	175	8.8	183.0	1
4	Rendez-Vous Restaurant Bar & Cafe	Ethiopian Restaurant	976	43.682578	-79.327544	4ad9221cf964a520671821e3	East York	Toronto	35	24	9.0	53.0	1
5	Red Rocket Coffee	Cafe	1003	43.682340	-79.328530	4f0f11fba4b035449917e927	East York	Toronto	64	119	8.7	92.0	1
6	Local 7304	Coffeehouse	8924	43.681407	-79.319947	240d9a8f0d0a708a2a8a4a8a8	East York	Toronto	33	77	8.9	45.0	1

2.7.2 New York City Final Data Frame

lat	lng	id	Borough	State	likes.count	photos.count	rating	ratingSignals	reasons.count	tips.count	verified	price.tier
3.896194	-79.328749	58df62a4498e98e8808b5e94	East York	Toronto	12	14	8.4	20.0	0	7	False	1.0
3.890482	-79.328509	4e3816b18877541e90eba82f	East York	Toronto	8	37	7.8	12.0	0	5	False	NaN
3.898922	-79.329520	51b0bcb8498e4f0309a58a65	East York	Toronto	19	38	8.1	29.0	0	14	False	NaN
3.882487	-79.328079	5155ca12e4b0e05526808bdf	East York	Toronto	132	175	8.8	183.0	1	48	False	2.0
3.882670	-79.327544	4ad9221cf984a520871821e3	East York	Toronto	35	24	9.0	53.0	1	23	False	2.0

2.7.3 Final Structure of the Consolidated Data Frame from 4 Data Source

City of Toronto data frame info

```
dfmsttn.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 568 entries, 0 to 567
Data columns (total 16 columns):
name                568 non-null object
categories          568 non-null object
distance            568 non-null int64
lat                 568 non-null float64
lng                 568 non-null float64
id                  568 non-null object
Borough             568 non-null object
State               568 non-null object
likes.count         568 non-null int64
photos.count        568 non-null int64
rating              553 non-null float64
ratingSignals        553 non-null float64
reasons.count       568 non-null int64
tips.count          568 non-null int64
verified            568 non-null bool
price.tier          355 non-null float64
dtypes: bool(1), float64(5), int64(5), object(5)
memory usage: 67.2+ KB
```

New York City data frame info

```
dfmstny.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 500 entries, 0 to 499
Data columns (total 16 columns):
name                500 non-null object
categories          500 non-null object
distance            500 non-null int64
lat                 500 non-null float64
lng                 500 non-null float64
id                  500 non-null object
Borough             500 non-null object
State               500 non-null object
likes.count         500 non-null int64
photos.count        500 non-null int64
rating              498 non-null float64
ratingSignals        498 non-null float64
reasons.count       500 non-null int64
tips.count          500 non-null int64
verified            500 non-null bool
price.tier          256 non-null float64
dtypes: bool(1), float64(5), int64(5), object(5)
memory usage: 59.2+ KB
```


2.8. How data will be used to solve the problem?

The above final data frame will be used to solve the problem statement

- Frequency distribution table with Venue category and Neighborhoods will be calculated. Top 15 venues will be filtered to identify opportunity of the missing slots or neighborhood having less than 2 venues will be recommended to business owner to start up new venues or opportunity to add more venue in the respective neighborhood to the site.
- Regression analysis will be performed to explore if there is any correlation exist between rating, price tier, user likes, tips, photo count and distance.
- Overall frequency tables will be calculated with percentile to distinguish between the two cities on all the columns and various visualization technique will be performed to observe the difference between two cities under various methods and parameters.
- K-Mean algorithm will be performed to cluster and segment the venues based on distance, price tier, user likes, photo count, tips and rating to enhance user experience in selecting the venues.