

## PROYECTO PARA CLUSTERIZAR LOS BARRIOS DE TORONTO Y NUEVA YORK

### 1. Introducción.

Los métodos de clusterización o agrupamiento nos sirven de mucho ya que nos permiten agrupar ya sea observaciones o individuos en clases, de modo que los objetos que se encuentren en una misma clase van a tener características similares o siguen un mismo patrón a diferencia de otros objetos que pertenecen a otra clase.

Este proceso o algoritmo de clusterización nos puede ayudar para agrupar a nuestro cliente, productos entre otras cosas.

### 2. Público objetivo:

El presente trabajo busca ser una guía de ayuda para aquellas personas que estén interesadas en querer abrir una franquicia de una pizzería y busca ayudarlos a escoger el mejor lugar.

### 3. Problema

Se pretende identificar que vecindarios de Toronto son similares o comparten características con vecindarios de Nueva York para así tomar una decisión y ver en que vecindarios es más factible poner una franquicia de pizzerías.

### 4. Datos a usar

Los datos que vamos a usar son datos de vecindarios de NY y Toronto, estos datos están en un dataframe y este tiene las siguientes características

Cantidad de filas y columnas: 409 filas y 5 columnas.

Nombre de columnas:

**City:** Nombre de las ciudades (Nueva York y Toronto)

**Borough:** Nombre de los distritos

**Neighborhood:** Nombre de los vecindarios

**Latitude:** Latitud del vecindario

**Longitude:** Longitud del vecindario

```
City Borough Neighborhood Latitude Longitude
0 NY Bronx Wakefield 40.894705 -73.847201
1 NY Bronx Co-op City 40.874294 -73.829939
2 NY Bronx Eastchester 40.887556 -73.827806
3 NY Bronx Fieldston 40.895437 -73.905643
4 NY Bronx Riverdale 40.890834 -73.912585
(409, 5)
```

5. También haremos uso del API de Forsquare para obtener datos de los lugares más populares que contiene cada vecindario.

Para este caso vamos a llamar a los 100 mejores lugares ubicados en un radio de 500 metros de los vecindarios, al realizar esta operación nos resulta un Data Frame con las siguientes características:

- 12089 filas y 7 columnas
- **Neighborhood**: Nombre del vecindario
- **Neighborhood Latitude**: Latitud del vecindario
- **Neighborhood Longitude**: Longitud del vecindario
- **Venue**: Lugar popular en el vecindario
- **Venue Latitude**: Latitud del lugar
- **Venue Longitude**: longitud del lugar
- **Venue Category**: Categoría del lugar

(12089, 7)							
	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Wakefield	40.894705	-73.847201	Lollipops Gelato	40.894123	-73.845892	Dessert Shop
1	Wakefield	40.894705	-73.847201	Carvel Ice Cream	40.890487	-73.848568	Ice Cream Shop
2	Wakefield	40.894705	-73.847201	Walgreens	40.896528	-73.844700	Pharmacy
3	Wakefield	40.894705	-73.847201	Rite Aid	40.896649	-73.844846	Pharmacy
4	Wakefield	40.894705	-73.847201	Dunkin'	40.890459	-73.849089	Donut Shop

Hacemos un pequeño análisis de la columna Category

Al hacer un análisis vemos lo siguiente:

- Hay un total de **12089** filas
- **463** valores únicos de categorías
- Pizza Place es el valor que se mas repite con un total de **510**

Si hacemos un top 20 de los valores más comunes en las categorías nos da lo siguiente:

Venue Category	Cantidad
Pizza Place	510
Coffee Shop	486
Italian Restaurant	340
Deli / Bodega	275
Café	272
Bakery	261
Bar	253
Sandwich Place	236
Park	226
Chinese Restaurant	221
Grocery Store	212
Pharmacy	196
Mexican Restaurant	187
American Restaurant	185
Bank	169
Donut Shop	169
Ice Cream Shop	158
Restaurant	151
Gym	145
Sushi Restaurant	142

6. A continuación, validamos cuantos lugares nos devuelve por cada vecindario

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Agincourt	5	5	5	5	5	5
Alderwood, Long Branch	9	9	9	9	9	9
Allerton	26	26	26	26	26	26
Annadale	14	14	14	14	14	14
Arden Heights	4	4	4	4	4	4

7. A continuación, vamos a mostrar los 10 mejores lugares por cada vecindario para efectos prácticos solo mostraremos 5 filas.

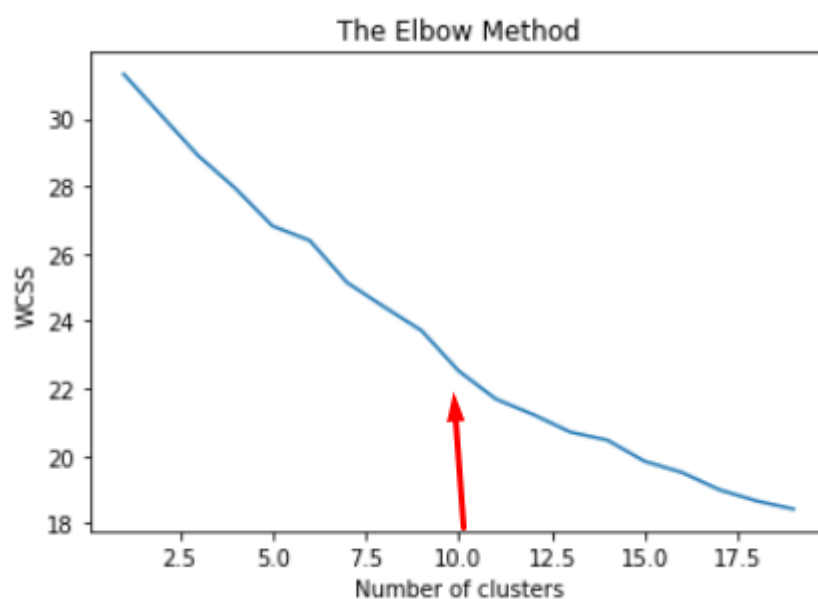
	Neighborhood	1st mas comun	2nd mas comun	3rd mas comun	4th mas comun	5th mas comun	6th mas comun	7th mas comun	8th mas comun	9th mas comun	10th mas comun
0	Agincourt	Clothing Store	Latin American Restaurant	Lounge	Breakfast Spot	College Auditorium	Club House	Cocktail Bar	Coffee Shop	College Academic Building	College Arts Building
1	Alderwood, Long Branch	Dance Studio	Gym	Athletics & Sports	Coffee Shop	Nail Salon	Club House	Cocktail Bar	College Academic Building	College Arts Building	College Auditorium
2	Allerton	Deli / Bodega	Chinese Restaurant	Discount Store	Intersection	Grocery Store	Gas Station	Breakfast Spot	Fast Food Restaurant	Bakery	Fried Chicken Joint
3	Annadale	Dance Studio	American Restaurant	Deli / Bodega	Food	Diner	Cosmetics Shop	Clothing Store	Club House	Cocktail Bar	Coffee Shop
4	Arden Heights	Bus Stop	Coffee Shop	Comfort Food Restaurant	Clothing Store	Club House	Cocktail Bar	College Academic Building	College Arts Building	College Auditorium	College Basketball Court

8. Para saber la calidad óptima de clúster hacemos uso del grafico del codo:

**Método del codo:** Este método utiliza los valores de la inercia obtenidos tras aplicar el K-means a diferente número de Clusters (desde 1 a N Clusters), siendo la inercia la suma de las distancias al cuadrado de cada objeto del Cluster a su centroide:

$$Inercia = \sum_{i=0}^N \|x_i - \mu\|^2$$

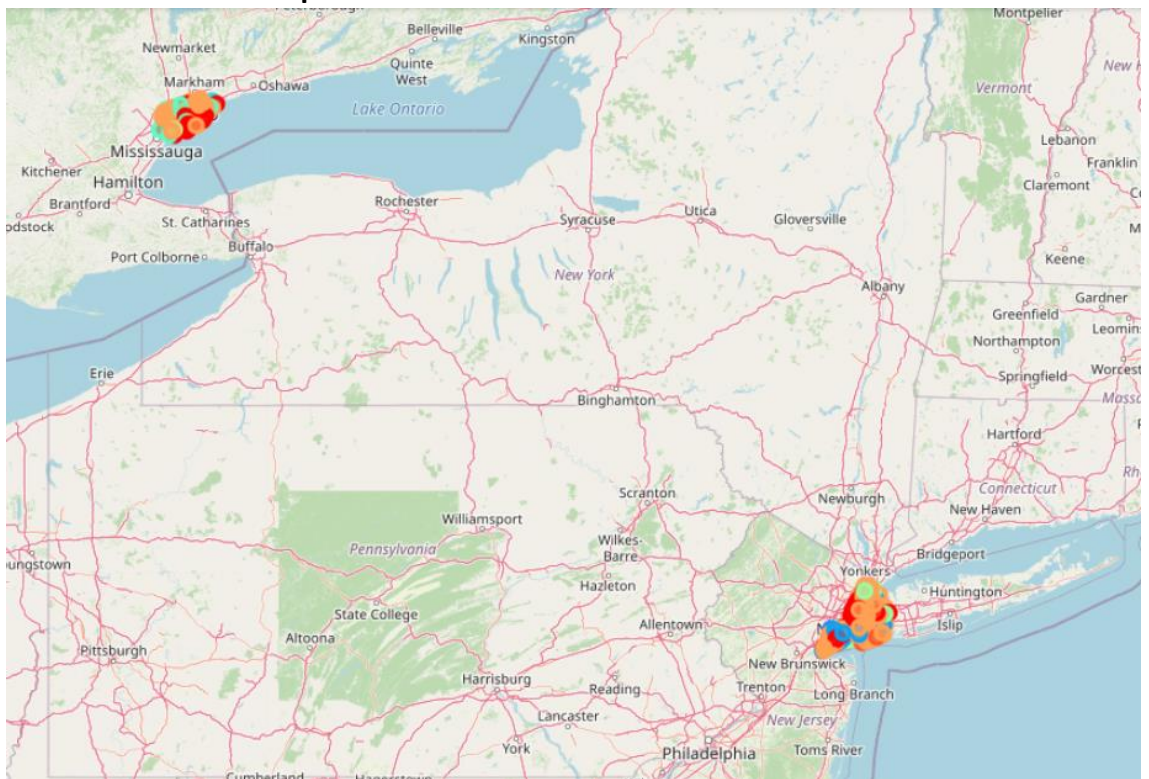
En este la gráfica no nos ayuda mucho, pero optaremos por elegir 1º Cluster que es donde se nota una leve curvatura



## 9. Observamos un resultado previo de lo que nos arroja al correr el cluster

	City	Borough	Neighborhood	Latitude	Longitude	Cluster_Labels	1st mas comun	2nd mas comun	3rd mas comun	4th mas comun	5th mas comun	6th mas comun	7th mas comun	8th mas comun
0	NY	Bronx	Wakefield	40.894705	-73.847201	8.0	Gas Station	Dessert Shop	Laundromat	Food	Deli / Bodega	Ice Cream Shop	Donut Shop	Cafe
1	NY	Bronx	Co-op City	40.874294	-73.829939	6.0	Bus Station	Fast Food Restaurant	Grocery Store	Basketball Court	Baseball Field	Fried Chicken Joint	Bagel Shop	Diner
2	NY	Bronx	Eastchester	40.887556	-73.827806	6.0	Bus Station	Caribbean Restaurant	Diner	Deli / Bodega	Bus Stop	Automotive Shop	Bowling Alley	Cosplay
3	NY	Bronx	Fieldston	40.895437	-73.905643	6.0	Business Service	Bus Station	College Basketball Court	Club House	Cocktail Bar	Coffee Shop	College Academic Building	Cafe
4	NY	Bronx	Riverdale	40.890834	-73.912585	6.0	Bus Station	Gym	Baseball Field	Bank	College Bookstore	Cocktail Bar	Coffee Shop	Cafe

## 10. Visualizamos en el mapa los cluster



11. Como podemos ver dibujado en el mapa hay barrios de Toronto y NY que pertenecen a un mismo cluster por lo que podemos decir que son similares.

## 12. Conclusiones.

Para nuestras conclusiones analizaremos unos de los cluster para ver las características que comparten los vecindarios

City	Borough	Cluster_Labels	1st mas comun	2nd mas comun	3rd mas comun	4th mas comun	5th mas comun	6th mas comun	7th mas comun	8th mas comun	9th mas comun	10th mas comun
NY	Bronx	0	Caribbean Restaurant	Bar	College Cafeteria	Cocktail Bar	Coffee Shop	College Academic Building	College Arts Building	College Auditorium	College Basketball Court	College Bookstore
NY	Bronx	0	Asian Restaurant	American Restaurant	Deli / Bodega	Bar	Italian Restaurant	Coffee Shop	Mobile Phone Shop	Club House	Cocktail Bar	College Academic Building
NY	Bronx	0	Italian Restaurant	Deli / Bodega	Bakery	Donut Shop	Bank	Dessert Shop	Liquor Store	Fish Market	Mexican Restaurant	Bar
NY	Brooklyn	0	Italian Restaurant	American Restaurant	Greek Restaurant	Bar	Chinese Restaurant	Bagel Shop	Ice Cream Shop	Hookah Bar	Grocery Store	Mediterranean Restaurant
NY	Brooklyn	0	Bar	Cocktail Bar	Coffee Shop	Deli / Bodega	French Restaurant	Café	Grocery Store	Bakery	Mexican Restaurant	Coworking Space
...	...	...	...	...	...	...	...	...	...	...	...	...
Toronto	Etobicoke	0	American Restaurant	Bakery	Coffee Shop	Mexican Restaurant	Fast Food Restaurant	Liquor Store	Gym	Fried Chicken Joint	Café	College Rec Center
Toronto	Downtown Toronto	0	Coffee Shop	Café	Italian Restaurant	Japanese Restaurant	Cocktail Bar	Beer Bar	Hotel	Creperie	Art Gallery	Breakfast Spot
Toronto	Downtown Toronto	0	Café	Coffee Shop	Chinese Restaurant	Bakery	Italian Restaurant	Butcher	General Entertainment	Gift Shop	Market	Convenience Store
Toronto	Downtown Toronto	0	Coffee Shop	Café	Japanese Restaurant	Hotel	Gym	Asian Restaurant	Deli / Bodega	American Restaurant	Gastropub	Concert Hall
Toronto	Downtown Toronto	0	Coffee Shop	Japanese Restaurant	Gay Bar	Café	Hotel	Bubble Tea Shop	Mediterranean Restaurant	Burger Joint	Men's Store	Hobby Shop

Este grupo se caracteriza por tener cerca a restaurantes especializados en comidas de otros países, así por ejemplo tenemos restaurantes de comida italiana, americana, asiática, mexicana y entre otros.

También se caracteriza por tener coffe shop

NOTA: Como podemos ver no hay pizzerías cerca por lo que sería una muy buena opción poner una franquicia de pizzas en los barrios mostrados en la imagen

**Bibliografía:**

- <https://jarroba.com/seleccion-del-numero-optimo-clusters/>
- [https://www.jacobsoft.com.mx/es\\_mx/k-means-clustering-con-python/](https://www.jacobsoft.com.mx/es_mx/k-means-clustering-con-python/)