

# Proyecto Final - Análisis de barrios en Scarborough, Toronto

Allan Gajardo

# Introducción

## ○ Presentación del problema

Scarborough es un antiguo distrito que se encuentra ubicado en Toronto, Canada. Y resulta ser un destino popular para inmigrantes. De hecho es una de las áreas mas multiculturales consideradas en Canada.

## • Problema a Resolver

Este proyecto tiene como fin ser una herramienta de información para quienes busquen información relevante sobre la distribución de buenos índices sobre los barrios de Scarborough. En concreto, se generará un análisis comparativo de diferentes barrios

# Datos

- Se utilizarán datos provenientes de Foursquare, utilizando su API para extraer la información necesaria para realizar este proyecto.
- Debido a las limitaciones del nivel usuario de las credenciales de dicha API, se fijarán parámetros

# Metodología

- Se utilizará el algoritmo de K-means que cumple con formar grupos de datos que comparten características similares

IBM Cloud Pak for Data

Mis proyectos / Capstone Project / Proyecto Final - Análisis de barri...

File Edit View Insert Cell Kernel Help

Not Trusted | Python 3.6

### Usando cluster K-means

```
In [77]: # Usando K-means para agrupar Los barrios en 3 clusters
Scarborough_grouped_clustering = Scarborough_grouped.drop('Neighborhood', 1)
kmeans = KMeans(n_clusters=3, random_state=0).fit(Scarborough_grouped_clustering)
kmeans.labels_

Out[77]: array([2, 0, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 0,
                2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2,
                2, 2, 2, 1, 2, 2, 2, 2, 2, 1, 2, 1, 2, 2, 2, 2, 2, 2, 1, 2, 2, 2,
                1, 2, 2, 2, 2, 0, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 0,
                2, 2, 2, 1, 2, 2, 2, 1, 1])
```

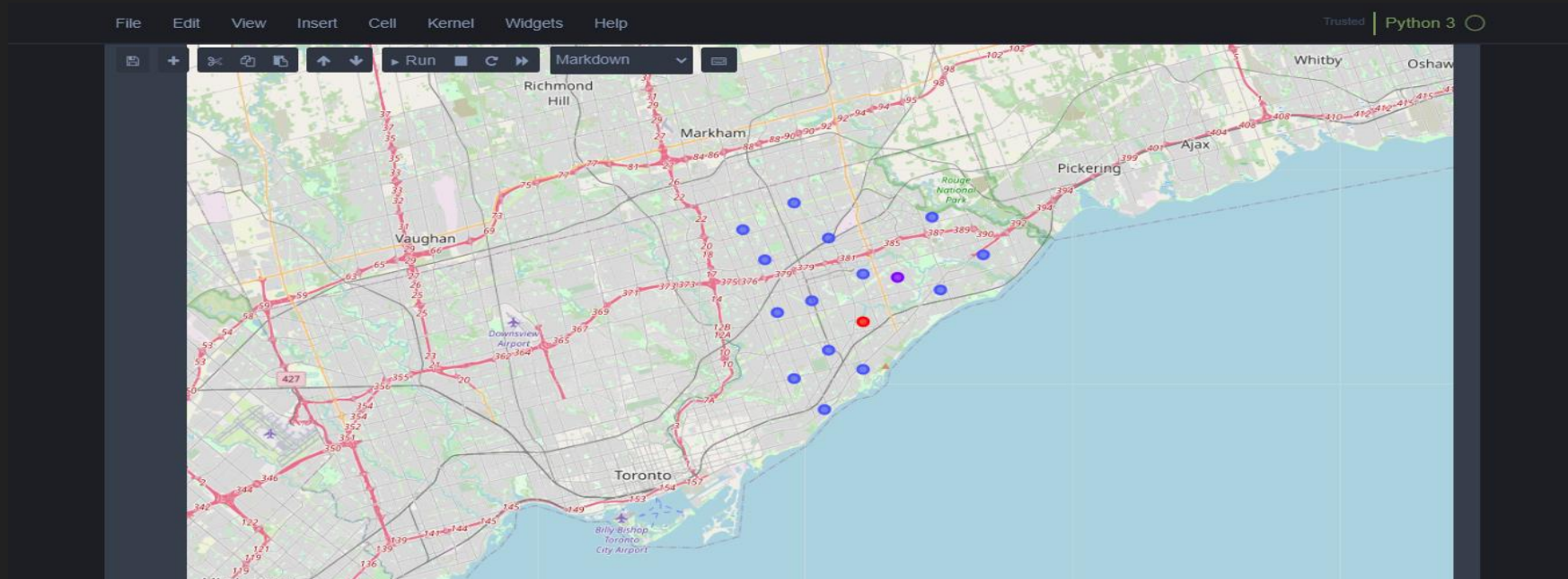
```
In [78]: neighborhoods_venues_sorted.insert(0, 'Cluster Labels', kmeans.labels_)
Scarborough_merged = df_2.iloc[:,16:]
# uniendo toronto_grouped con toronto_data para agregar Latitude/Longitude a cada barrio
Scarborough_merged = Scarborough_merged.join(neighborhoods_venues_sorted.set_index('Neighborhood'), on='Neighborhood')
Scarborough_merged.head()
```

```
Out[78]:
```

	Postal Code	Borough	Neighborhood	Latitude	Longitude	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	M1B	Scarborough	Malvern, Rouge	43.808686	-79.194353	2	Fast Food Restaurant	Coffee Shop	Hobby Shop	Spa	Dumpling Restaurant	Discount Store	Distribution Center	Dive Bar	Dog Run	Doner Restaurant
1	M1C	Scarborough	Rouge Hill, Port Union, Highland Creek	43.784535	-79.180497	2	Breakfast Spot	Bar	Burger Joint	Women's Store	Dive Bar	Dog Run	Doner Restaurant	Donut Shop	Drugstore	Dumpling Restaurant
2	M1E	Scarborough	Guildwood, Morningside, West Hill	43.783573	-79.188711	2	Restaurant	Fast Food Restaurant	Park	Intersection	Medical Center	Sports Bar	Electronics Store	Beer Store	Moving Target	Bank
3	M1G	Scarborough	Woburn	43.770992	-79.216917	1	Coffee Shop	Park	Business Service	Women's Store	Drugstore	Discount Store	Distribution Center	Dive Bar	Dog Run	Doner Restaurant
4	M1H	Scarborough	Cedarbrae	43.773136	-79.239478	2	Bakery	Indian Restaurant	Coffee Shop	Bank	Gas Station	Caribbean Restaurant	Athletics & Sports	Fried Chicken Joint	Chinese Restaurant	Thai Restaurant

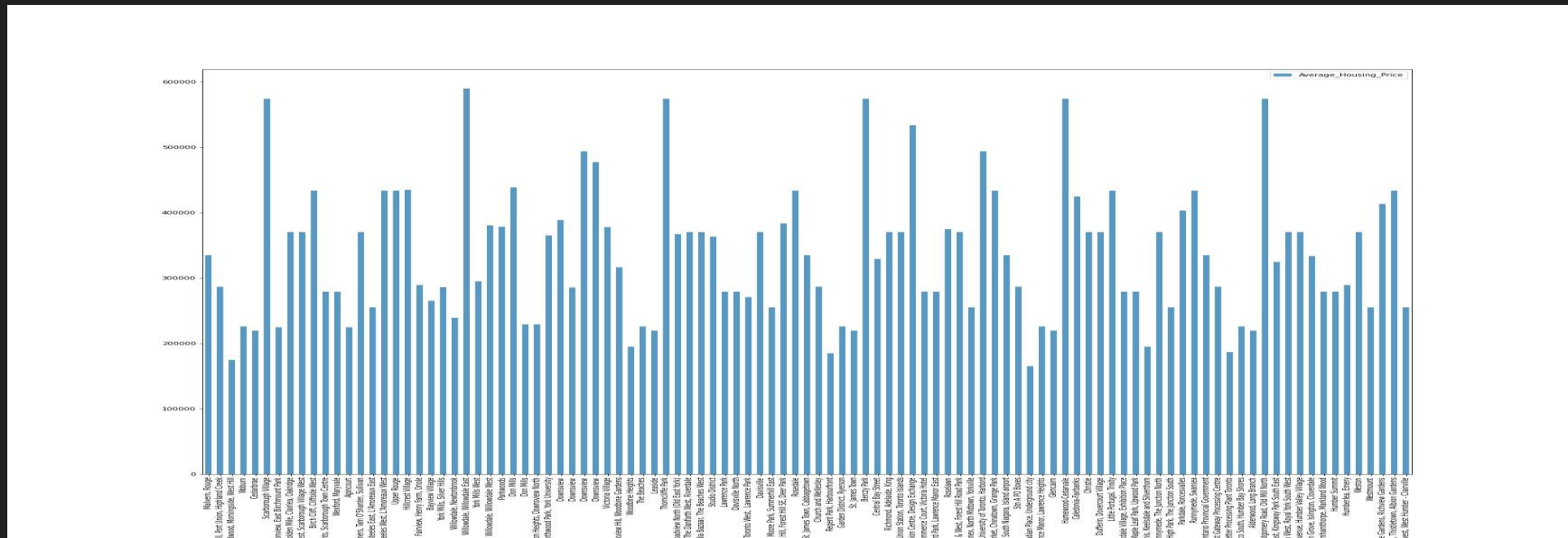
# Resultados

- Obtenemos los grupos distribuidos según características que comparten en el siguiente mapa generado con códigos de machine learning



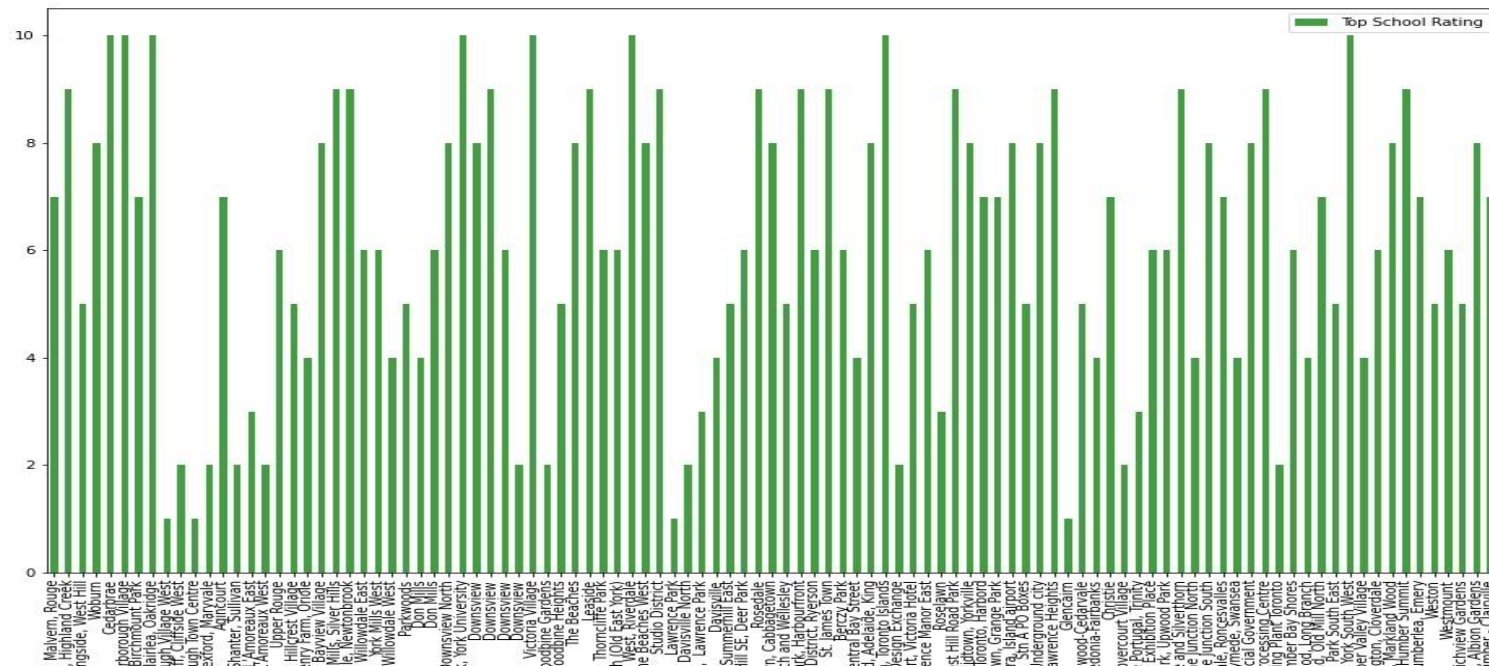
# Resultados

- Obtenemos además una gráfica que muestra el comportamiento de los precios de las casas en localidades de Scarborough



# Resultados

- Y finalmente se tiene una gráfica que muestra los puntajes de las escuelas, que podrían indicar un nivel superior de escolaridad por localidad



# Discusión

- Resulta necesario mencionar que dicho análisis realizado comprende una suposición de mejor lugares para vivir variables como lo socialmente más concurrido, y algunas otras estadísticas que ofrecen altos niveles y puntajes de lo que se considera bueno. Sin embargo, acotar que esta suposición puede variar para cada tipo de analista que tenga otras preferencias para tomar alguna decisión observando este trabajo.



# Conclusión

- La utilización del algoritmo K-means en este proyecto, y la utilización de gráficos, logra mostrar que efectivamente existen barrios muy similares aquellos que están cercanos entre sí. Y que efectivamente, se percibe una diferenciación entre barrios.
- Espero que este trabajo ayude a tomar decisiones con fundamento a quienes necesiten tomar en cuenta variables como las que aquí se tienen en consideración.