

INFORME FINAL

Capstone

2023

Proyecto de final de carrera - La batalla de los vecindarios: encontrar un lugar mejor lugar en Scarborough, Toronto

1. Introducción:

El propósito de este proyecto es ayudar a las personas a explorar mejores instalaciones en su vecindario. Ayudará a las personas a tomar decisiones inteligentes y eficientes al seleccionar un vecindario excelente entre varios vecindarios en Scarborough, Toronto.

Muchas personas están migrando a varios estados de Canadá y necesitan mucha investigación para encontrar buenos precios de vivienda y escuelas con buena reputación para sus hijos. Este proyecto es para esas personas que buscan vecindarios mejores. Con el fin de acceder con facilidad a Cafeterías, Escuelas, Supermercados, farmacias, tiendas de abarrotes, centros comerciales, teatros, hospitales, personas con intereses similares, etc.

Este proyecto tiene como objetivo crear un análisis de características para las personas que migran a Scarborough para buscar el mejor vecindario como análisis comparativo entre vecindarios. Las características incluyen el precio medio de la vivienda y mejores escuelas de acuerdo a las calificaciones, tasas de crimen de esa área específica, conectividad vial, condiciones climáticas, buena gestión para emergencias, recursos de agua tanto frescos como residuos y excrementos transportados en alcantarillados y servicios recreativos.

Esto ayudará a las personas a conocer el área y el vecindario antes de mudarse a una nueva ciudad, estado, país o lugar por su trabajo o para comenzar una nueva vida.

2. Sección de datos

Enlace de datos: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

Utilizaremos el conjunto de datos de Scarborough que recolectamos de Wikipedia en la semana 3. El conjunto de datos consta de latitud y longitud, códigos postales.

Datos de la API de Foursquare:

Necesitaremos datos sobre diferentes lugares en diferentes vecindarios de ese distrito específico. Para obtener esa información, usaremos la información de ubicación de "Foursquare". Foursquare es un proveedor de datos de ubicación con información sobre

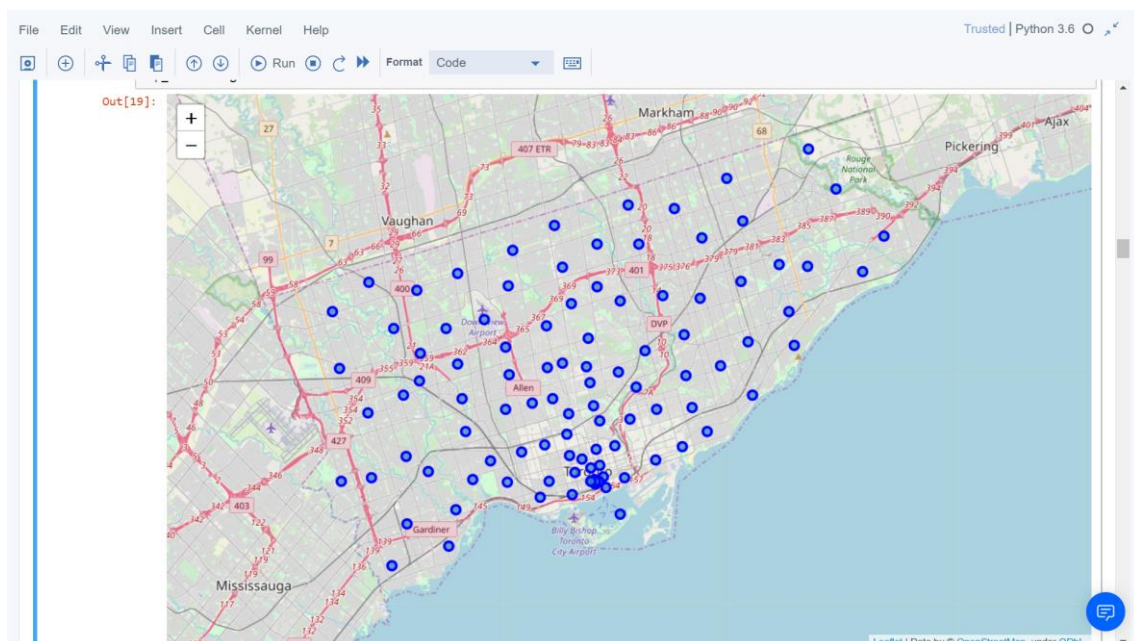
todo tipo de lugares y eventos dentro de un área de interés. Esa información incluye nombres de lugares, ubicaciones, menús e incluso fotos.

, la plataforma de ubicación de Foursquare se utilizará como la única fuente de datos ya que toda la información requerida se puede obtener a través de la API.

Después de encontrar la lista de vecindarios, nos conectamos a la API de Foursquare para recolectar información sobre los lugares dentro de cada vecindario. Para cada vecindario, hemos elegido un radio de 100 metros.

Los datos recolectados de Foursquare contenían información de lugares dentro de una distancia especificada de la longitud y latitud de los códigos postales. La información obtenida por lugar es la siguiente:

1. Neighborhood
2. Neighborhood Latitude
3. Neighborhood Longitude
4. Venue
5. Name of the venue e.g. the name of a store or restaurant
6. Venue Latitude
7. Venue Longitude
8. Venue Category



3. Sección de metodología

Enfoque de agrupamiento:

Para comparar las similitudes de dos ciudades, decidimos explorar vecindarios, segmentarlos y agruparlos en clusters para encontrar vecindarios similares en una gran ciudad como Nueva York y Toronto. Para poder hacer eso, necesitamos agrupar datos que es una forma de aprendizaje automático no supervisado: algoritmo de agrupamiento k-means.

Usando el enfoque de agrupamiento K-Means

```
In [36]: neighborhoods_venues_sorted.insert(0, 'Cluster Labels', kmeans.labels_)

Scarborough_merged = df_2.iloc[:,16,:]

# merge toronto_grouped with toronto_data to add Latitude/Longitude for each neighborhood
Scarborough_merged = Scarborough_merged.join(neighborhoods_venues_sorted.set_index('Neighborhood'), on='Neighborhood')

Scarborough_merged.head()# check the last columns!
```

Out[36]:

orough	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
orough	Rouge, Malvern	43.811525	-79.195517	0	Zoo Exhibit	Financial or Legal Service	Fast Food Restaurant	Construction & Landscaping	Fish & Chips Shop	Filipino Restaurant	Field	Fish Market	Farmers Market	Doner Restaurant
orough	Highland Creek, Rouge Hill, Port Union	43.785665	-79.158725	0	Bar	Falafel Restaurant	Donut Shop	Dumpling Restaurant	Eastern European Restaurant	Electronics Store	Elementary School	Ethiopian Restaurant	Event Space	Yoga Studio
orough	Guildwood, Morningside, West Hill	43.765815	-79.175193	2	Park	Gym / Fitness Center	Pool	Fried Chicken Joint	Indian Restaurant	Athletics & Sports	Ethiopian Restaurant	Donut Shop	Dumpling Restaurant	Eastern European Restaurant
orough	Woburn	43.768369	-79.217590	0	Coffee Shop	Fast Food Restaurant	Business Service	Park	Yoga Studio	Dumpling Restaurant	Eastern European Restaurant	Electronics Store	Elementary School	Ethiopian Restaurant
orough	Cedarbrae	43.769688	-79.239440	0	Flower Shop	Athletics & Sports	Thai Restaurant	Bank	Bakery	Caribbean Restaurant	Hakka Restaurant	Indian Restaurant	Eastern European Restaurant	Electronics Store

Los lugares más comunes cerca del vecindario

```
import numpy as np
num_top_venues = 10

indicators = ['st', 'nd', 'rd']

columns = ['Neighborhood']
for ind in np.arange(num_top_venues):
    try:
        columns.append('{} Most Common Venue'.format(ind+1, indicators[ind]))
    except:
        columns.append('{}th Most Common Venue'.format(ind+1))

neighborhoods_venues_sorted = pd.DataFrame(columns=columns)
neighborhoods_venues_sorted['Neighborhood'] = Scarborough_grouped['Neighborhood']

for ind in np.arange(Scarborough_grouped.shape[0]):
    neighborhoods_venues_sorted.iloc[ind, 1:] = return_most_common_venues(Scarborough_grouped.iloc[ind, :], num_top_venues)

neighborhoods_venues_sorted.head()
```

t[34]:

	Neighborhood	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	Adelaide, King, Richmond	Coffee Shop	Café	Hotel	Gastropub	Burger Joint	Asian Restaurant	Bar	Restaurant	American Restaurant	Steakhouse
1	Agincourt	Chinese Restaurant	Shopping Mall	Pizza Place	Supermarket	Sushi Restaurant	Breakfast Spot	Print Shop	Mediterranean Restaurant	Coffee Shop	Pool
2	Agincourt North, L'Amoreaux East, Milliken, ST...	Pharmacy	Sandwich Place	Sushi Restaurant	Doner Restaurant	Donut Shop	Dumpling Restaurant	Eastern European Restaurant	Electronics Store	Elementary School	Ethiopian Restaurant
3	Albion Gardens, Beaumont Heights, Humbergate, ...	Grocery Store	Park	Sandwich Place	Discount Store	Japanese Restaurant	Fried Chicken Joint	Beer Store	Hardware Store	Pizza Place	Fast Food Restaurant
4	Alderwood, Long Branch	Convenience Store	Pub	Sandwich Place	Coffee Shop	Gas Station	Dance Studio	Gym	Pharmacy	Pizza Place	Falafel Restaurant

Flujo de trabajo:

Usando las credenciales de la API de Foursquare, se extraerían las características de los lugares cercanos de los vecindarios. Debido a las limitaciones de las solicitudes http, el

número de lugares por vecindario se establecería razonablemente en 100 y el parámetro de radio se establecería en 500.

4. Sección de resultados

Mapa de Clusters en Scarborough



Precio medio de la vivienda por clusters en Scarborough

6. Sección de conclusiones

En conclusión, el proyecto de la batalla de los vecindarios buscó ayudar a las personas a explorar mejores instalaciones en su vecindario en Scarborough, Toronto. Se utilizó técnicas de web scraping y análisis de datos para recopilar y analizar información sobre precios de vivienda, calificaciones de escuelas, tasas de delitos y lugares comunes en diferentes vecindarios. Los resultados proporcionaron una visión general de las características de los vecindarios de Scarborough, pero se recomienda hacer más investigación y considerar otros factores importantes antes de tomar una decisión sobre dónde vivir.