

La batalla de los Vecindarios:

Mi cafetería en Toronto



Arianne Encinar Manzano

Proyecto final – módulo Capstone

1. Introducción

En este documento se van a plantear diferentes aspectos a tener en cuenta a la hora de escoger el sitio ideal para una cafetería en la ciudad de Toronto. Se analizarán los diferentes vecindarios de Toronto en busca de información como número de cafeterías en la zona, densidad de población, ... que sirvan como apoyo para tomar estas decisiones.

El objetivo del proyecto es identificar aquellos vecindarios donde abrir una cafetería podría tener mayor potencial de éxito, lo que implica buscar áreas con alta afluencia de personas, pero con pocas cafeterías existentes. Un análisis de este tipo es valioso para empresas y emprendedores permitiéndoles minimizar la competencia y maximizar la visibilidad y clientela potencial.

2. Ingesta de Datos

Para llevar a cabo este análisis se utilizan cuatro fuentes de datos principales:

1. API de Foursquare (versión 3)

Se emplea la API de Foursquare para obtener información geolocalizada de lugares en la ciudad de Toronto. Esta API permite consultar lugares por categorías como cafeterías, oficinas, tiendas, parques u otros puntos de interés. En particular, se recopila:

- Ubicación de cafeterías: permite identificar cuántas cafeterías hay por zona (FSA), lo cual es fundamental para evaluar la competencia.
- Lugares de interés: como oficinas, zonas comerciales o atracciones turísticas, ya que incrementan la afluencia potencial de clientes.

Para cada lugar se recopilan datos como la dirección, nombre, categoría y coordenadas geográficas.

2. Datos censales de Toronto

Se utilizan datos abiertos del Gobierno de Toronto sobre perfiles de vecindario, específicamente los que incluyen la densidad de población por código postal (FSA – Forward Sortation Area).

Esta información permite evaluar la cantidad de personas que viven en cada zona, una variable clave para estimar la demanda potencial.

3. Coordenadas geográficas por FSA

Con el fin de representar los datos espacialmente, se incorpora un conjunto de datos externo con las coordenadas (latitud y longitud) promedio asociadas a cada FSA.

Este dataset fue obtenido de Mendeley Data:

[Canadian Forward Sortation Areas with Coordinates](#)

Su uso permite mapear y visualizar cada zona en un mapa interactivo, así como realizar un análisis espacial más preciso.

4. Información de los FSAs pertenecientes a la Ciudad de Toronto.

Con el fin de extraer la información relativa a los FSAs de la Ciudad de Toronto se utiliza el recopilado en la página web de la [Wikipedia](#), que facilita múltiple información de entre la que se extrae el listado de FSAs objetivo.

Todos los datos obtenidos se integran en un único conjunto de datos estructurado, uniendo la información demográfica, geográfica y comercial, utilizando el código postal (FSA) como clave principal de integración.

3. Descripción del problema y análisis

Abrir una cafetería puede parecer una idea sencilla, pero elegir dónde abrirla marca la diferencia entre el éxito y el fracaso. La ciudad de Toronto, con su diversidad cultural, gran población urbana y alto flujo de visitantes, es un lugar con mucho potencial... pero también con mucha competencia.

En este contexto, el presente análisis busca identificar las zonas más prometedoras de Toronto para establecer una nueva cafetería, teniendo en cuenta tres factores clave:

- La densidad de población (es decir, cuántas personas viven en la zona)
- El número de cafeterías existentes (para estimar la competencia)
- La presencia de otros lugares de interés, como oficinas, tiendas o parques (que atraen flujo de personas)

La idea es sencilla: si una zona tiene mucha gente, pocos competidores y alta actividad, entonces es una buena candidata para abrir una nueva cafetería.

Este tipo de análisis puede ser muy útil para emprendedores, pequeñas empresas o cadenas que buscan expandirse de manera inteligente, basando sus decisiones en datos reales y no solo en intuición.

Además, al usar datos abiertos y herramientas de análisis geoespacial, se ofrece una metodología que podría replicarse en otras ciudades o negocios similares. Eso sí, claro está, son las modificaciones pertinentes y la inclusión de fuentes de información adaptadas a las nuevas áreas geográficas.

4. Análisis del potencial de ubicación: competencia, atractivo turístico y densidad de población

Para este análisis se han considerado tres variables clave: la densidad de cafeterías por zona, la cantidad de lugares de interés y la densidad de población en los distintos códigos postales (FSA) de la ciudad de Toronto.

En primer lugar, la densidad de cafeterías permite evaluar el nivel de competencia en cada área. Zonas con una alta concentración de cafeterías pueden presentar mayores desafíos para un nuevo negocio debido a la competencia establecida.

En segundo lugar, se ha considerado el número de lugares de interés en cada zona. Áreas con mayor atractivo —como oficinas, parques, tiendas y otros puntos de reunión— suelen registrar un flujo mayor de personas, lo que incrementa la clientela potencial.

Por último, se ha tenido en cuenta la densidad de población por FSA. Una mayor cantidad de residentes en una zona sugiere un mayor número de personas que transitan o viven en el área, lo que aumenta las posibilidades de captar clientes habituales.

Combinando estas tres variables, se ha buscado identificar zonas con baja competencia, alto atractivo y elevada densidad de población, que puedan representar ubicaciones óptimas para la apertura de una nueva cafetería.

5. Metodología

El enfoque metodológico seguido en este análisis combina técnicas de limpieza y transformación de datos, análisis exploratorio, normalización de variables clave, y algoritmos de aprendizaje automático no supervisado (clustering) para identificar zonas óptimas para la apertura de una nueva cafetería en la ciudad de Toronto.

Los pasos que se han seguido en este análisis son los siguientes:

Paso 1: Preparación de los datos

La fase inicial consistió en recopilar y cruzar distintas fuentes de información:

- Datos de densidad de población por FSA (Forward Sortation Area) en Toronto.
- Localizaciones geográficas de cafeterías y otros puntos de interés extraídos desde la API de Foursquare.
- Coordenadas reales y nombres de vecindarios asociados a cada FSA mediante un dataset adicional publicado en Mendeley Data.

Para poder combinar estas fuentes, se realizó un proceso de limpieza y estandarización, incluyendo:

- Extracción del FSA desde direcciones postales.
- Normalización de nombres de columnas y formatos.
- Agregación del número de cafeterías y puntos de interés por FSA.
- Imputación de valores nulos en los vecindarios a partir de otras columnas.

Paso 2: Análisis exploratorio

Se realizó un análisis exploratorio para entender la distribución de las principales variables:

- Histograma de densidad de población.
- Distribución de cafeterías por FSA.

- Relación entre densidad de población, competencia (cafeterías) y atractivo (lugares de interés).

Este paso permitió detectar outliers, zonas con muy baja densidad o sin cafeterías, y preparó el terreno para una evaluación combinada.

Paso 3: Normalización y puntuación compuesta

Dado que las variables principales tienen diferentes escalas se procedió a una **normalización min-max** (para llevarlas al mismo rango [0, 1]).

Se invirtió la escala del número de cafeterías, para que valores bajos (menos competencia) aporten positivamente a la puntuación final (este campo se ha denominado *competencia_invertida*). La fórmula compuesta usada fue:

$$score = 0,4 * densidad_{normalizada} + 0,4 * competencia_{invertida} + 0,2 * interés_{normalizado}$$

Esta puntuación se utilizó para generar un ranking de los FSAs más atractivos.

Paso 4: Agrupamiento con aprendizaje automático

Para complementar el análisis, se utilizó el algoritmo de **K-Means Clustering** con el fin de segmentar los FSAs en grupos homogéneos según sus características combinadas:

- Densidad de población
- Número de cafeterías
- Número de lugares de interés

Se evaluó visualmente la separación de clusters mediante gráficos 2D y se utilizó k=5 como valor base para facilitar la interpretación (zonas favorables, neutras y poco recomendables).

Paso 5: Visualización geográfica

Por último, se incorporaron mapas interactivos (mediante Plotly y Folium) para representar de manera visual los resultados sobre el plano real de Toronto:

- Puntuación total por FSA.
- Agrupamiento por cluster.
- Diagramas de dispersión y comparativas multivariadas.

Estas herramientas permiten navegar y profundizar en el análisis a nivel visual, facilitando la toma de decisiones por parte de un posible inversor o emprendedor.

6. Discusión de los resultados

Tras el análisis combinado de las variables clave (densidad de población, número de cafeterías y número de lugares de interés por zona (FSA)) se han obtenido indicadores relevantes que permiten identificar las áreas con mayor potencial para ubicar una nueva cafetería en Toronto.

El análisis exploratorio reveló que algunos FSAs presentan una alta densidad de población, pero también una gran cantidad de cafeterías, lo que eleva la competencia. Por ejemplo, las zonas con códigos postales M5A y M9C tienen una densidad de población elevada, de 48978 y 38752 habitantes respectivamente, pero también un número mayor de cafeterías en la zona (5 cafeterías cada una), por lo que el nivel de competencia sería mayor.

Otros FSAs, en cambio, combinan una densidad razonable de población con una baja presencia de cafeterías, lo que los convierte en opciones potenciales al ofrecer mercado disponible sin saturación. Por ejemplo, M2J y M9V tienen una densidad de población elevada, 61761 y 53878 habitantes respectivamente, pero no hay ninguna cafetería cercana a esas zonas.

La presencia de lugares de interés, como oficinas, parques y tiendas, varía notablemente por zona y tiende a concentrarse en áreas céntricas, aunque algunos FSAs periféricos también muestran buen equilibrio entre densidad y atractivo.

Por otra parte, se construyó una puntuación para cada FSA normalizando las tres variables y sumándolas, de forma que:

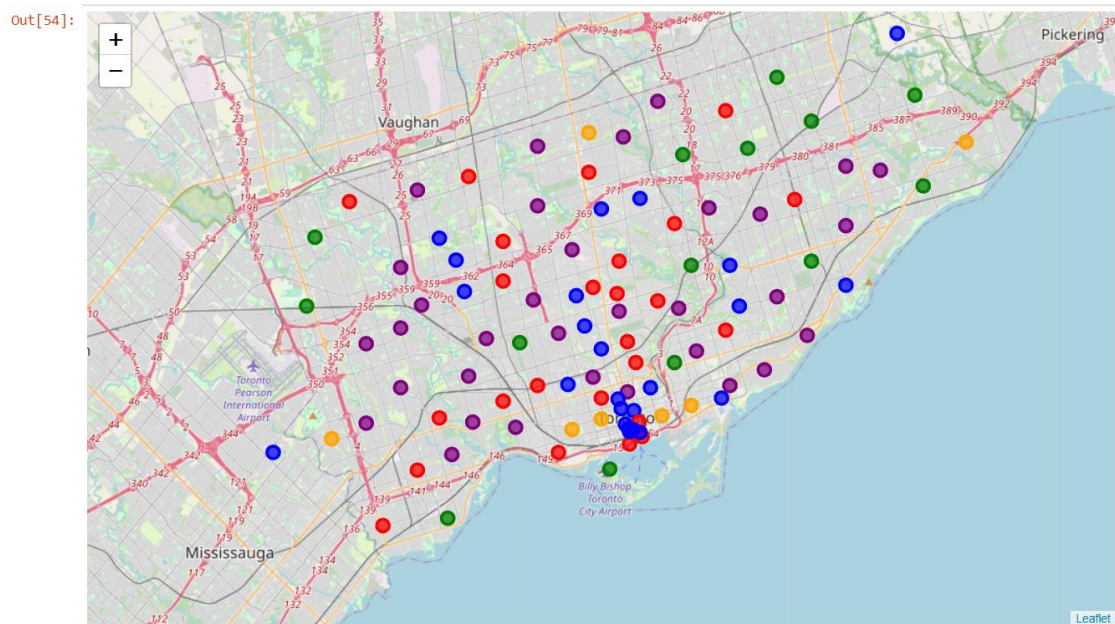
- Menos cafeterías → Más puntuación
- Más población y más lugares de interés → Más puntuación

Los FSAs con mayor puntuación según esta métrica son:

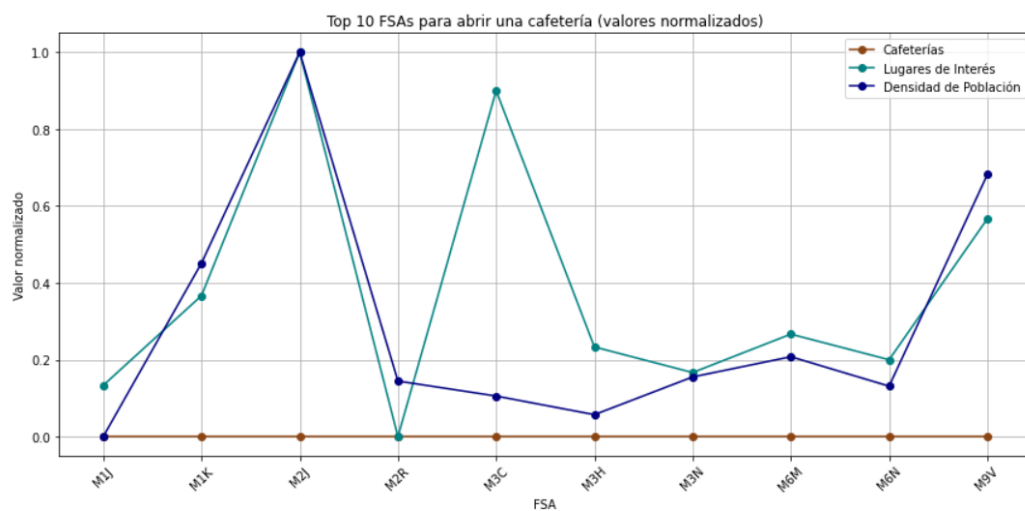
Código Postal	Cafeterías	Lugares de interés	Población	Puntuación
M2J	0	46	61761	0.968
M9V	0	33	53878	0.866
M1K	0	27	48175	0.806
M6M	0	24	42148	0.757
M3N	0	21	40846	0.737
M2R	0	16	40581	0.714
M6N	0	22	40252	0.737
M3C	0	43	39616	0.821
M3H	0	23	38416	0.730
M1J	0	20	37002	0.709

Adicionalmente, a través de un algoritmo de clustering (K-Means con k=5) se agruparon los FSAs en 5 categorías, permitiendo refinar aún más el análisis y verificar si los FSAs con mejor puntuación coincidían con pertenecer al mejor cluster. Mediante este análisis se comprueba que M2J pertenece al cluster 2, el mejor de los cluster en relación al número de habitantes, las cafeterías que serían competencia y el número de lugares de interés.

Visualización de los FSA con un color que representa el cluster al que pertenecen:

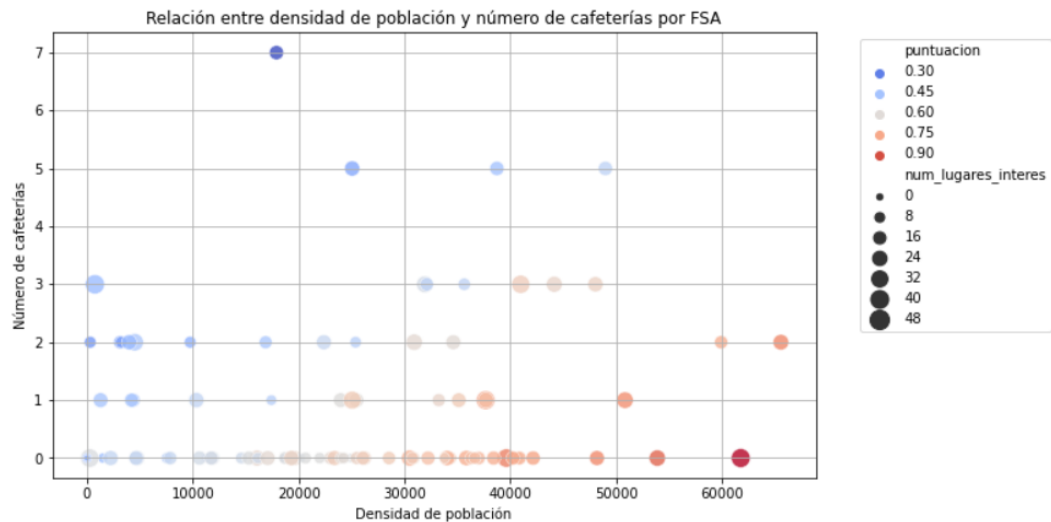


Por otro lado, se representaron los 10 mejores resultados en base a su scoring para comparar su potencial gráficamente:

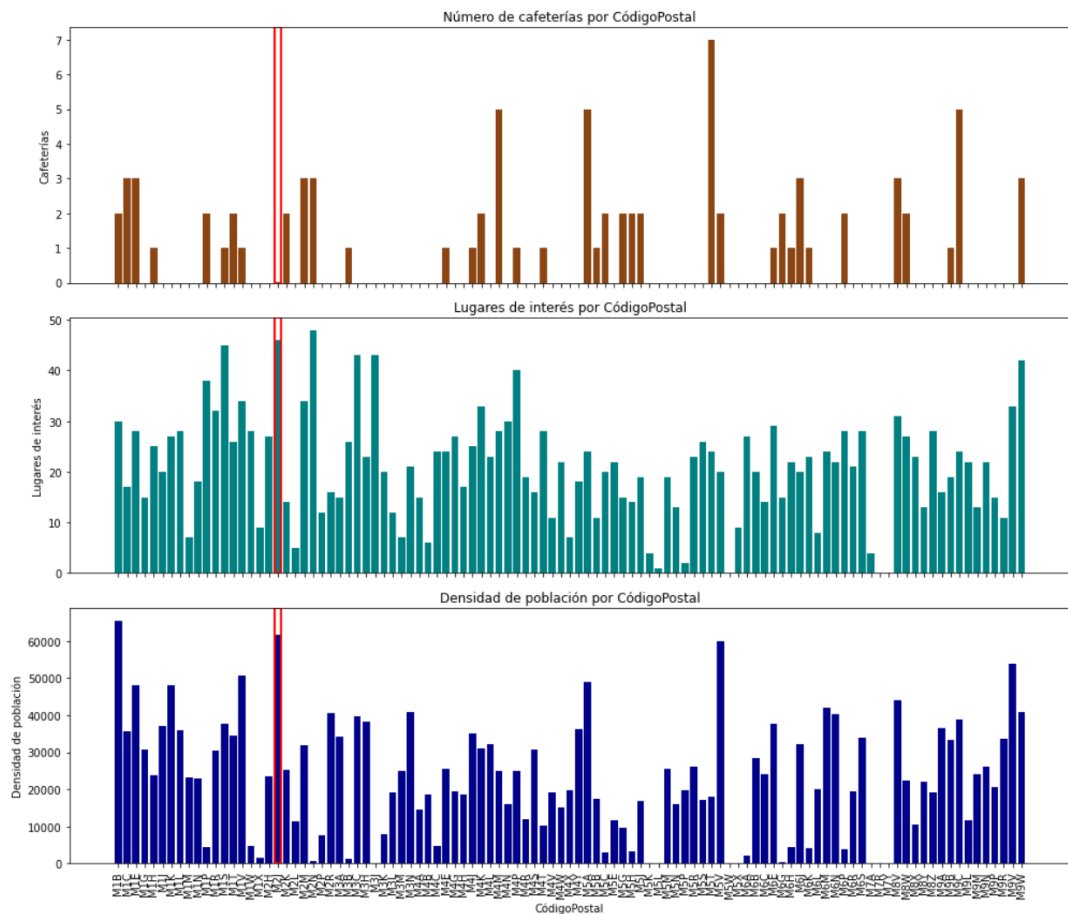


En la gráfica destaca sobre los demás el FSA M2J, coincidiendo con el análisis previo.

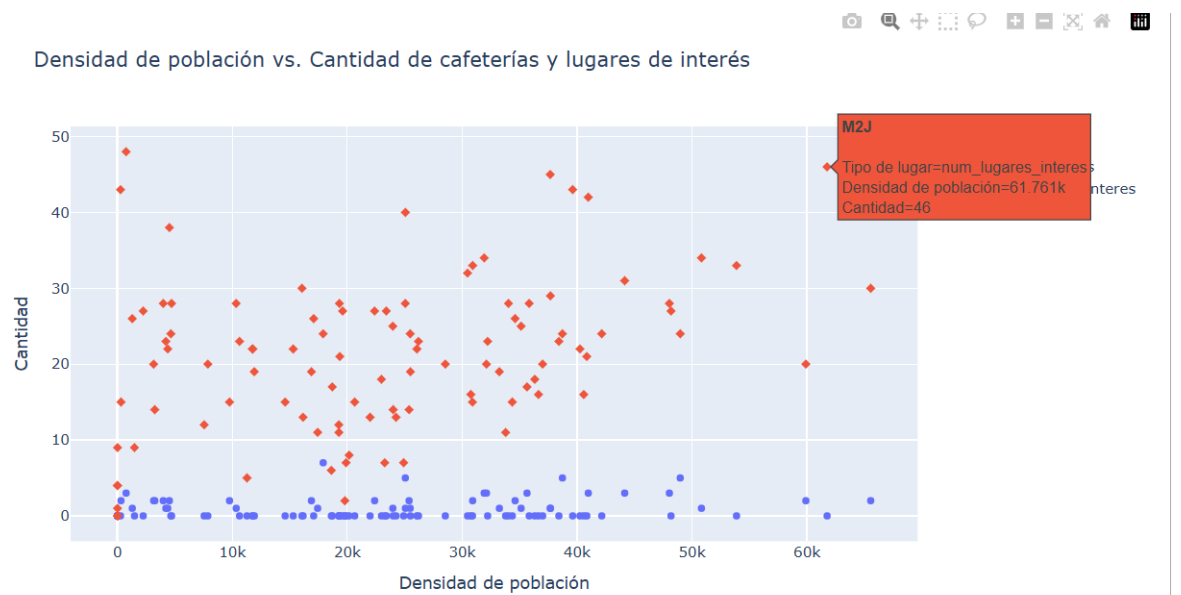
Las tres variables estudiadas se pueden visualizar conjuntamente en el siguiente gráfico de dispersión en el que el eje x muestra la densidad de población, el eje y el número de cafeterías existentes, el tamaño del nodo el número de lugares de interés y el color del nodo muestra la puntuación (cuanto más alta, mejor ubicación para abrir una cafetería nueva):



De nuevo, se observa que el punto rojo (correspondiente a M2J) es la mejor ubicación para abrir una cafetería. La información de cada variable de la opción preferida (M2J, señalada con un rectángulo rojo) se compara con la de las otras localizaciones a partir de los siguientes diagramas de barras:



Por último, se ha incluido un gráfico interactivo que representa la cantidad de cafeterías (en azul) y la cantidad de lugares de interés (en rojo) en el eje y frente a la densidad de población en el eje x. Clicando sobre cada punto se muestra su información:



En base a los análisis descritos y las gráficas explicadas, se concluye que M2J es la mejor ubicación potencial para abrir una nueva cafetería.