

A close-up photograph of a white ceramic cup filled with a latte, featuring a heart-shaped latte art design on the surface. The cup sits on a matching white saucer with a silver spoon resting on it. In the background, a golden-brown croissant is visible on a white plate, and a glass of orange juice is partially seen. The scene is set on a table with a brown and white checkered pattern.

Café en Toronto City

Batalla de Vecindarios
Arianne Encinar Manzano

Mayo de 2025

¿Quieres abrir una cafetería en la ciudad de Toronto?



Identificar las mejores zonas de la ciudad hará que la inversión sea más rentable y duradera.



La elección del lugar marcará el punto de partida del negocio y permitirá proyectar mejor a futuro.



Estudiar la densidad de población, la competencia de la zona y los lugares de interés próximos que permitan analizar la posible clientela ayudará a escoger los lugares óptimos.

Fuentes de datos

Para el análisis se han utilizado 4 fuentes de datos diferentes:



Wikipedia – Listado de los FSAs específicos de la Ciudad de Toronto.



Mendeley Data -
Coordenadas de los diferentes FSAs de Toronto (Necesaria para poder extraer las coordenadas base de los FSAs para hacer búsquedas con Foursquare más precisas).



Datos censales de Toronto – Densidad de población por FSA.



Datos de Foursquare – API de datos para la extracción de competencia y lugares de interés.



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.

Para el análisis se han empleado 6 fases:

- 1. Preparación de los datos**
- 2. Análisis Exploratorio**
- 3. Normalización y puntuación compuesta**
- 4. Agrupamiento con aprendizaje automático**
- 5. Visualización geográfica**
- 6. Discusión de los resultados – reflejado en el apartado de resultados**

1. Preparación de datos

Se extraen los valores de los FSAs pertenecientes a la Ciudad de Toronto y se cargan los datos de geolocalización del CSV descargado de la fuente de datos. Con la lectura de ambos y su unión se tiene listo el listado de FSAs con unas coordenadas. Aprovechando la lista, la recorremos y llamamos recurrentemente a la API de Foursquare con el propósito de recoger información de las cafeterías cercanas a las coordenadas y los lugares de interés.

Con los nuevos listados , uno para las cafeterías y el otro para los lugares de interés, agruparemos los resultados por FSA aprovechando los datos devueltos por el API y almacenaremos el conteo, y lo juntaremos con la información de los FSAs del principio. De este modo, se obtiene un listado de los FSAs de la ciudad de Toronto junto con el recuento de cafeterías y el conteo de los lugares de interés.

Por último, se cargan los datos censales de Toronto y se juntan con el listado actual para obtener la densidad de población por FSA. En aquellos casos en los que no se tienen registros de los lugares de interés o cafeterías porque la API no haya encontrado ninguno, se rellenarán con 0.

2. Análisis Exploratorio

Se llevó a cabo un análisis exploratorio para comprender la distribución y comportamiento de las variables clave del estudio. Se analizaron histogramas de densidad poblacional y distribuciones de cafeterías por FSA, así como la relación entre población, competencia (número de cafeterías) y atractivo (número de lugares de interés). Este análisis permitió identificar outliers, zonas con escasa densidad o sin presencia de cafeterías, y estableció las bases para un análisis combinado posterior.



3. Normalización y puntuación compuesta

Dado que las variables clave presentaban diferentes escalas, se aplicó una normalización Min-Max para estandarizarlas en un rango común [0,1]. Adicionalmente, se invirtió la escala del número de cafeterías para reflejar que una menor competencia incrementa la idoneidad de la ubicación (campo competencia_invertida).

La fórmula compuesta empleada para calcular la puntuación final fue:

$$\text{score} = 0.4 \times \text{densidad_normalizada} + 0.4 \times \text{competencia_invertida} + 0.2 \times \text{interés_normalizado}$$

Esta métrica permitió generar un ranking objetivo de los FSAs más atractivos para ubicar una nueva cafetería.



4. Agrupamiento con aprendizaje automático

Para enriquecer el análisis, se aplicó el algoritmo de *K-means Clustering* con el objetivo de agrupar los FSAs en segmentos homogéneos según sus características clave:

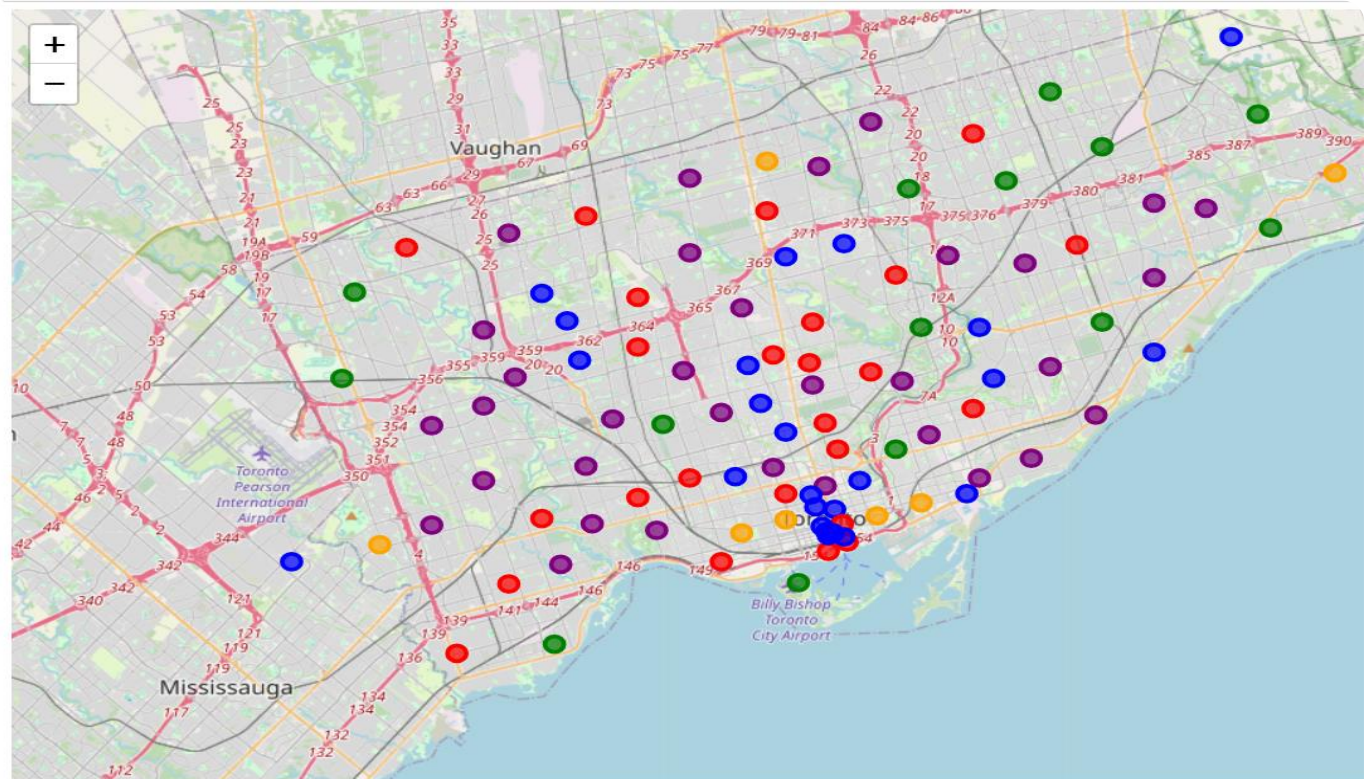
- Densidad de población
- Número de cafeterías
- Número de lugares de interés
- La segmentación permitió identificar patrones latentes en los datos y clasificar las zonas según su atractivo comercial. Se utilizó un valor de $k = 5$, seleccionado por su equilibrio entre simplicidad y diferenciación entre grupos. La validez de los clusters fue evaluada mediante representaciones gráficas bidimensionales, que mostraron una separación razonable entre segmentos.



5. visualización geográfica

En último lugar, se incorporaron mapas interactivos para representar de manera gráfica los diferentes clusters generados por el algoritmo. 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.

Out[54]:

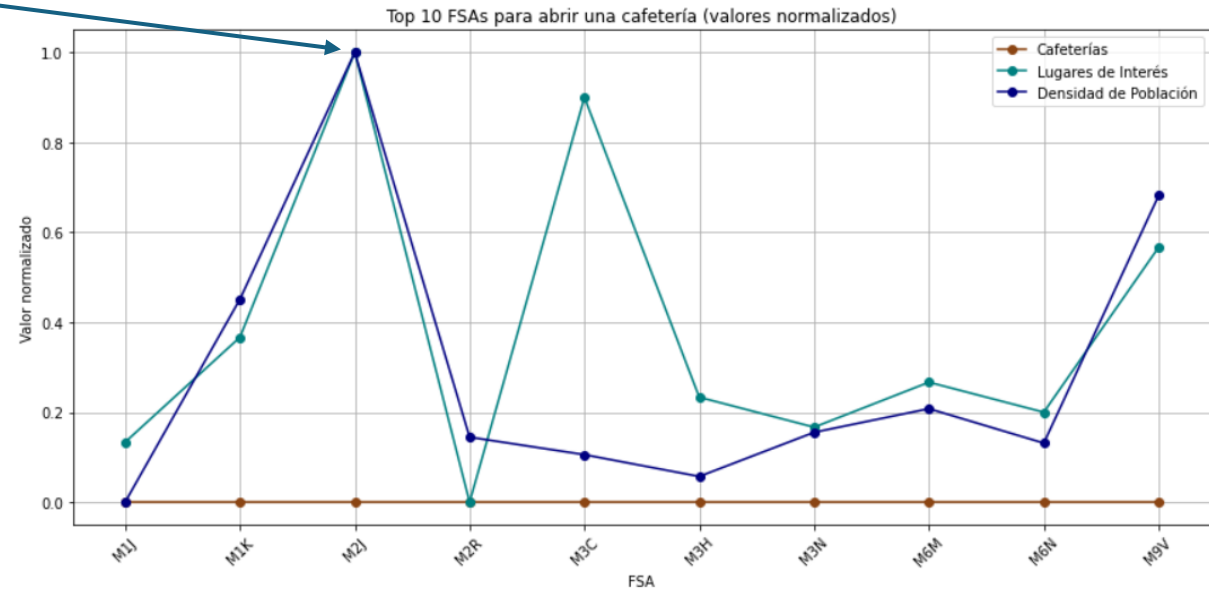


Resultados

El análisis exploratorio reveló la presencia de FSA altamente poblados, pero con un elevado número de cafeterías, lo que eleva la competencia. Otros FSAs, en cambio presentan una densidad razonable de población, pero una baja presencia de cafeterías, lo que los convierte en opciones potenciales.

Partiendo de estos datos y de la combinación con el scoring realizado, se destacan 10 posibles localizaciones muchas de ellas pertenecientes a los clusters 2 y 3:

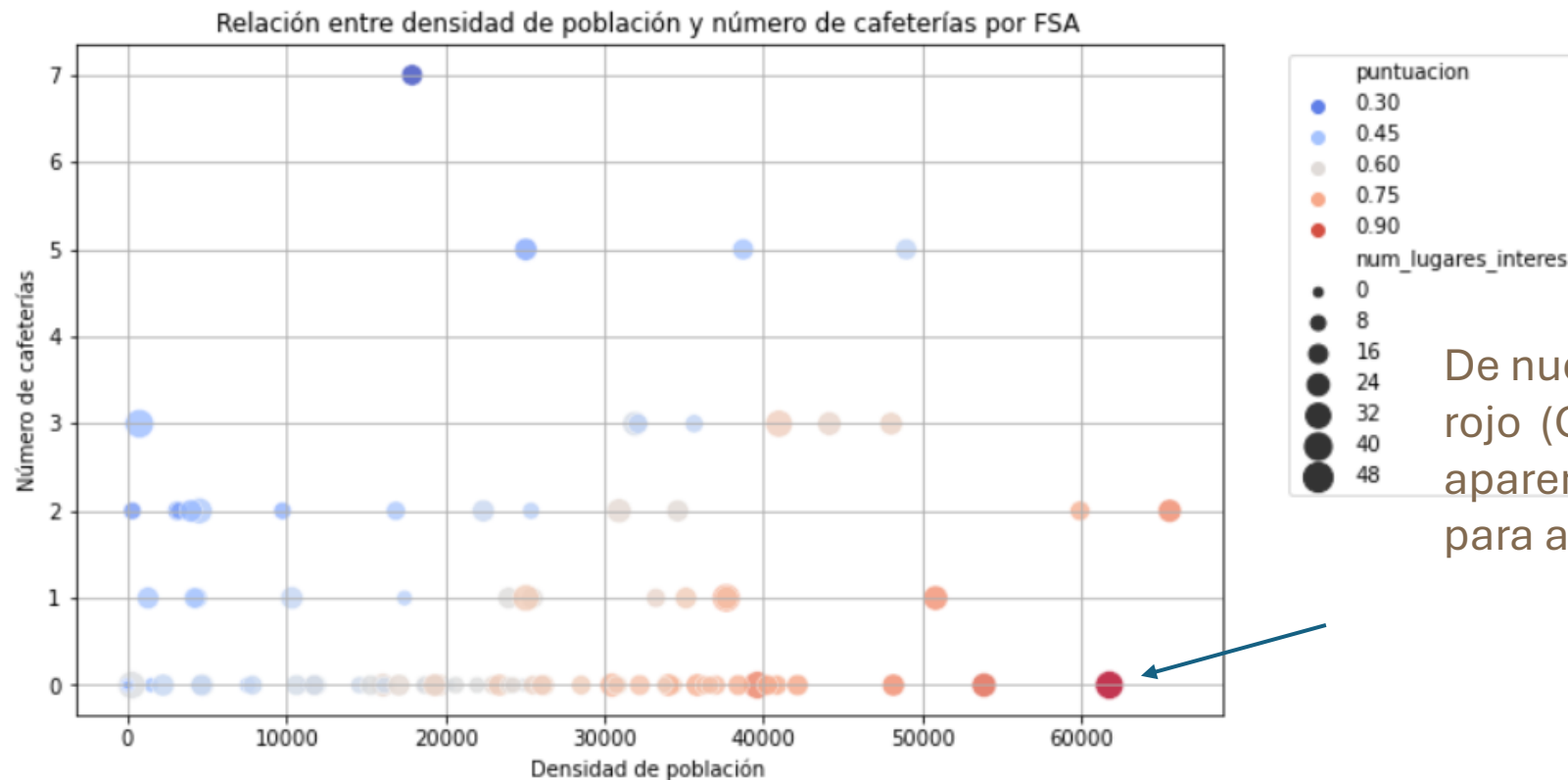
Punto máximo



| 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 |

Resultados

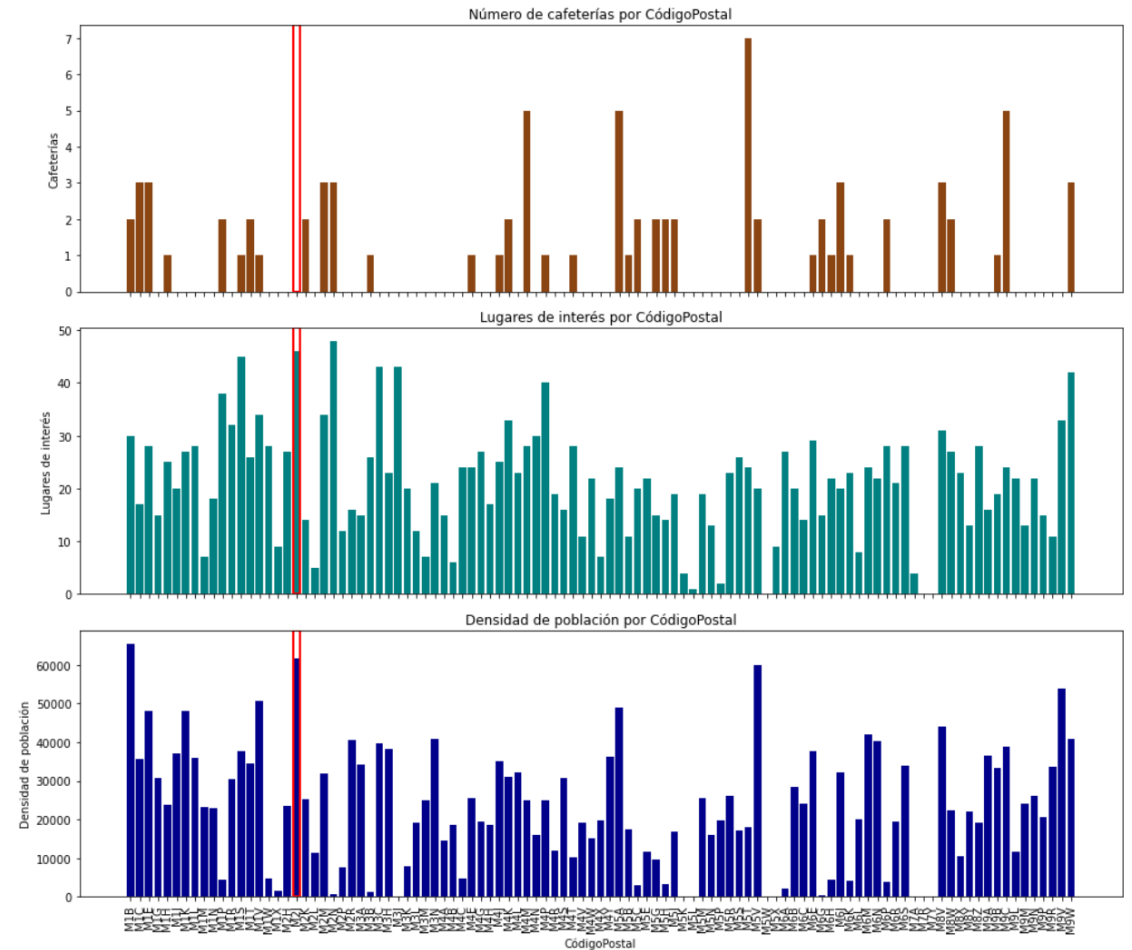
Las tres variables estudiadas se pueden visualizar conjuntamente en el siguiente gráfico de dispersión en el que el eje x muestra la densidad de población, el eje y el número de cafeterías existentes, el tamaño del nodo el número de lugares de interés y el color del nodo muestra la puntuación (cuanto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



De nuevo, se observa que el punto rojo (Correspondiente a M2J) aparenta ser la mejor ubicación para abrir una cafetería.

Resultados

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



Conclusiones y trabajo futuro

- El análisis exploratorio destaca en todas las gráficas que las propiedades del FSA M2J son las mejores para localizar una nueva cafetería, se trata de una zona con un gran numero de habitantes, sin presencia de cafeterías en los alrededores y presenta números lugares de interés que pueden atraer clientela potencial.
- Si se quisiese llevar a cabo el proyecto y profundizar más, sería interesante estudiar el nivel adquisitivo de la zona para definir el estilo de cafetería y poder personalizar el perfil del posible cliente.

