

# Solución Challenge Estimación Precio / m<sup>2</sup>

DaCodes

Camilo Romero - 29 Ago 2023

# Contexto

## Estimación Precio / m<sup>2</sup>

Luego de una investigación previo al estudio de la data presentada para la estimación del precio por m<sup>2</sup>, se evidencia que depende mayoritariamente de:

- Ubicación / zona de la propiedad
- Amenities alrededor de la propiedad
- Ingreso estimado de personas por zona
- Características de la propiedad (número de pisos, baños, cuartos, año de construcción).

# Contexto

## Estimación Precio / m2

Se intentó solucionar de la siguiente forma:

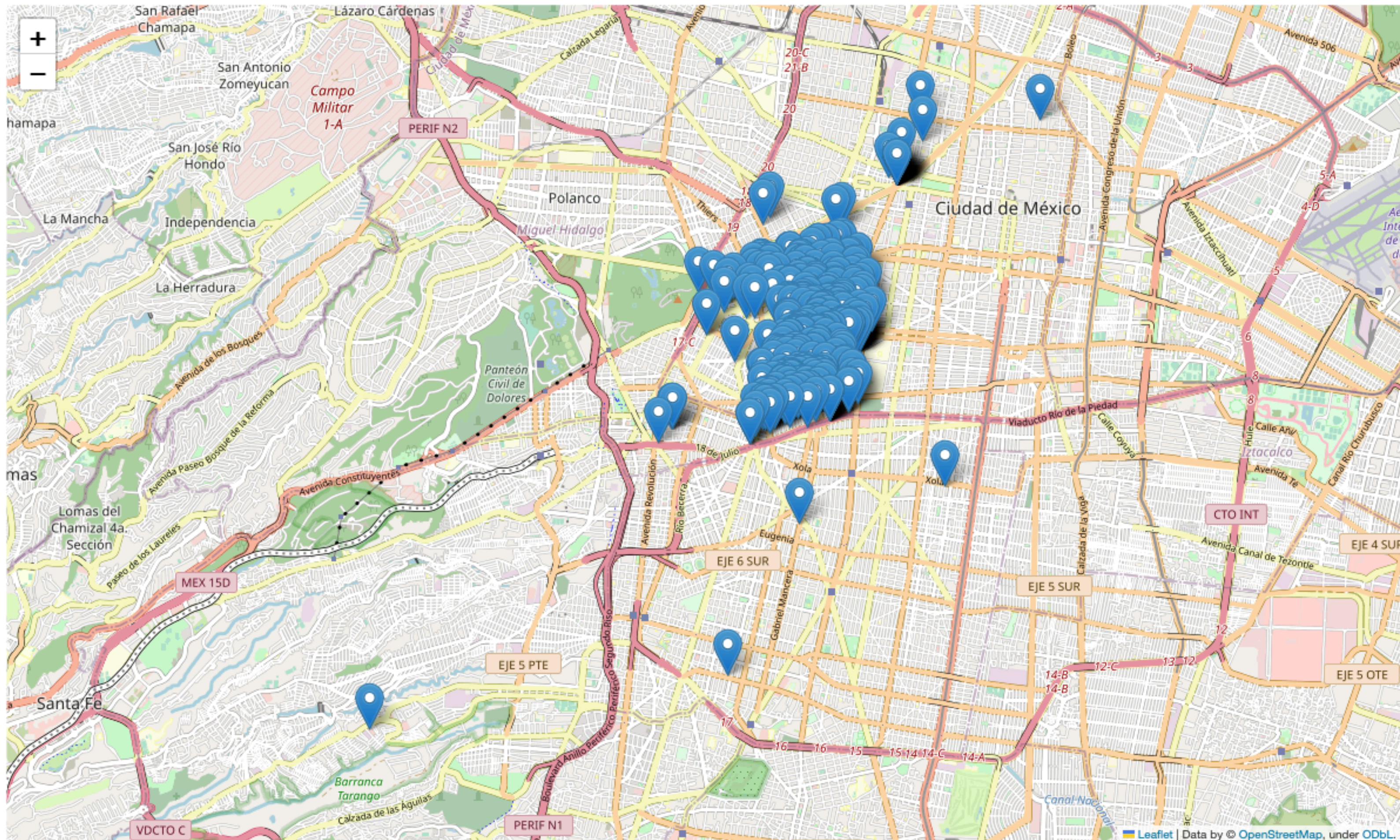
- **Ubicación / zona de la propiedad** (Uso OpenStreetMap para extraer el código postal, barrio, distrito)
- **Amenities alrededor de la propiedad** (Se intentó utilizar google maps API para construir un grafo con amenities en un rango de 1km. No se lo pudo realizar por el costo).
- **Ingreso estimado de personas por zona** (Se desconoce si existen datos abiertos en MX).
- **Características de la propiedad** (número de pisos, baños, cuartos, año de construcción) (Incluidas en el dataset).



# Ubicación de propiedades

## Distrito Federal

Numero de propiedades: 944

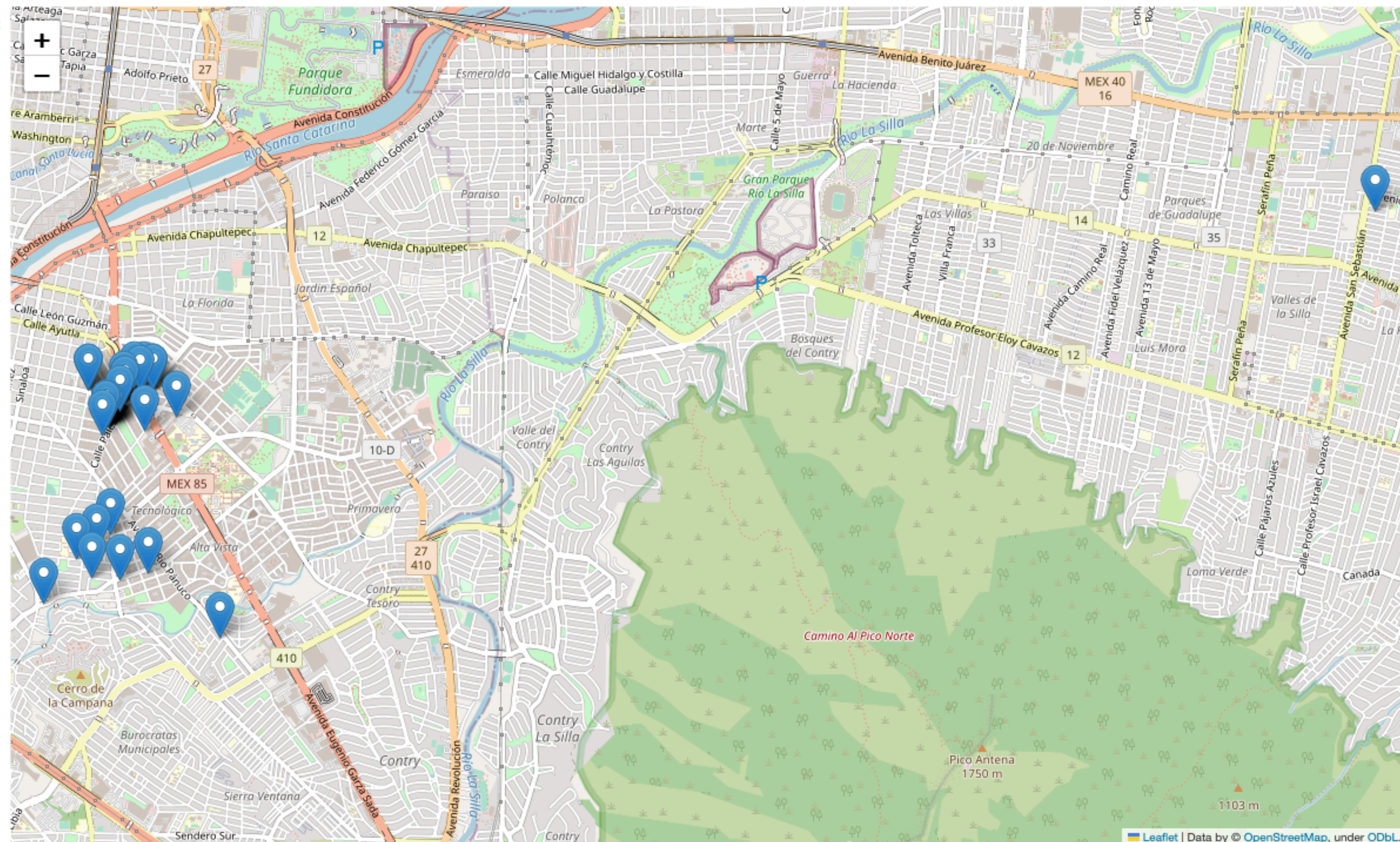




# Ubicación de propiedades

## Nuevo León

Numero de propiedades: 36

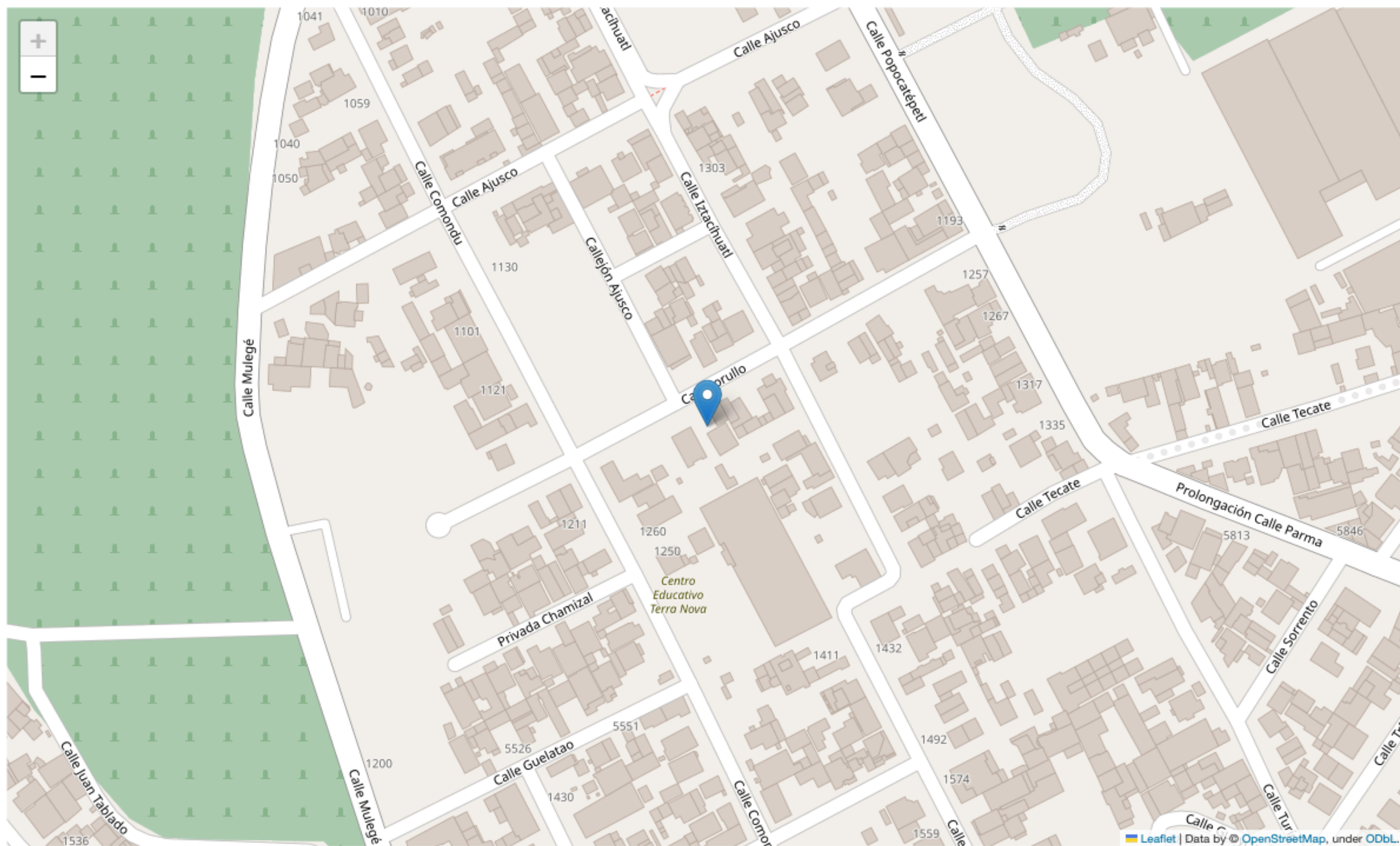




# Ubicación de propiedades

## Baja California

Numero de propiedades: 1



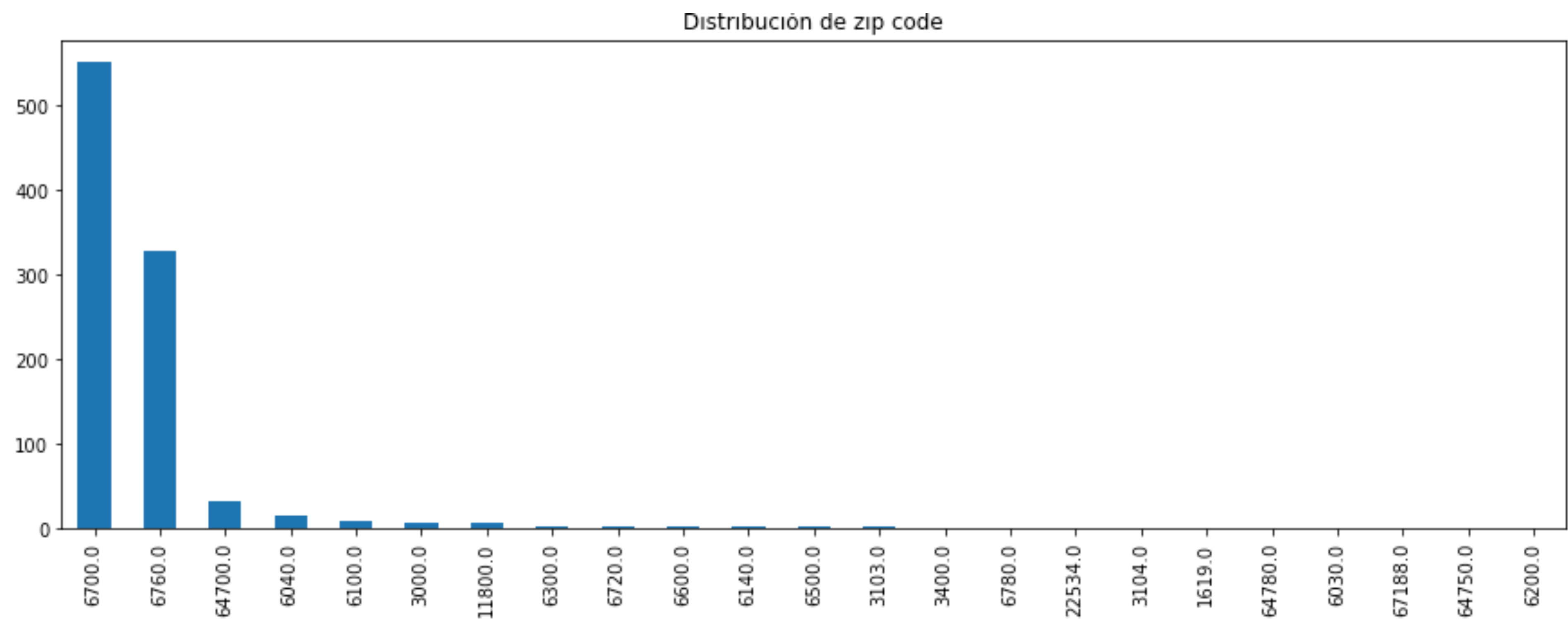
# Ubicación de propiedades

Existen ciertas novedades con las propiedades provistas:

- Desbalanceo entre las propiedades y su ubicación.
- Las propiedades de CDMX tienen precios promedios similares de acuerdo a fuentes externas.
- Baja California solo tiene 1 registro y será complicado por algún algoritmo que aprenda de esta región.

# Ubicación de propiedades

Distribución de post code de las propiedades:





# Estimación de Precio / m2

No es aconsejable estimar el precio por m2 ya que al ser una medida de radio, la alteración del precio o m2 puede resultar en la misma medida. Por ejemplo:

$$\text{Propiedad}_1 = \text{precio} / \text{m2} = 1000 \text{ USD} / 2 \text{ m2} = 500 \text{ USD/m2}$$

$$\text{Propiedad}_2 = \text{precio} / \text{m2} = 500 \text{ USD} / 1 \text{ m2} = 500 \text{ USD/m2}$$

Por ende es mejor **estimar el precio final** y dividirlo para el metraje de la propiedad.



# Planteamiento del problema

- Se divide en bins la variable precio final para elaborar un problema de clasificación.
- Se divide en train y test para medir el performance del modelo.
- Se realiza un análisis de descriptivo corto por cada bin creado.
- Se utiliza feature importance por el método del árbol y sharp para medir la importancia de las variables.



# Modelamiento

- Se utilizó un modelo de Random Forest con ciertas columnas que no tenían valores missing y variabilidad del modelo.
- Como medida de evaluación del modelo, se escogió F1 score por el desbalance de las propiedad y variabilidad alta.
- Se escogió feature importance por el método del árbol y sharp para determinar que variables más impacta en el columna target.



# Conclusiones

- No existe, con las variables existentes y extraídas con OpenStreetMap, una forma de separar por grupos creados de propiedad, por lo que estimar su precio final tiene un margen de error considerable.
- Se recomienda contar con mayor número de registros para encontrar propiedades similares y así poder estimar su precio final.
- No es una buena idea estimar directamente el precio por m<sup>2</sup> ya que al ser una medida de radio, valores diferentes pueden resultar en el mismo valor final.



# Pregunta 3

DaCodes

Camilo Romero - 29 Ago 2023

# Contexto

## Learning to Run a Power Network (L2RPN)

Learning to Run a Power Network fue sin duda el mejor proyecto que he participado.

- Se basa en aplicar reinforcement learning (RL) para la solución de un “juego” idealizado para solventar obstáculos en una red (similar a una red de telecomunicaciones pero aplicado a la red eléctrica).
- Generalmente las redes eléctricos son muy complejas, con muchas decisiones por tomar y nadie entiende como funcionan animadas.



# Contexto

## Learning to Run a Power Network (L2RPN)

- El principal reto fue crear data sintética con propiedad a la data original (características estadísticas similares).
- Disponer de los mecanismo para ejecutar todo el conjunto de acciones para solventar el problema. Regularmente los problemas de RL tienen un campo de acciones pequeño (mover joystick up, down, left, right). Nuestro problema tenía un campo muy amplio lo que dificultaba el problema.
- Otro reto fue elaborar baselines del agentes de RL que se ajusten a la medida mínimo de score propuesto.

# Contexto

## Learning to Run a Power Network (L2RPN)

- Se trabajó con diversos investigadores del área (Google, Isabel Guyón) para organizar el challenge con un MVP como primera versión y lanzar la competencia.
- EL resultado fue satisfactorio. Diversos participantes desarrollaron metodologías para reducir el campo de acción del agente del RL.
- Otros participantes utilizaron analizar la data y ocupar únicamente ciertos tiempos para el entrenamiento del modelo.



# Contexto

## Learning to Run a Power Network (L2RPN)

- Se aprendió a modelar data sintética (vía útil cuando no se disponen de muchos datos).
- Se aprendió a modelar agentes de RL.
- Se realizó análisis profundo de la data disponible para complicar las reglas de juego o competencia.