

Problem Set 3: Making Money with ML?

Big Data y Machine Learning para Economía Aplicada

Gustavo Adolfo Castillo Álvarez (201812166),
Alexander Almeida Ramírez (202225165),
Jorge Luis Congacha Yunda (201920042) y
Jaime Orlando Buitrago González (200612390)

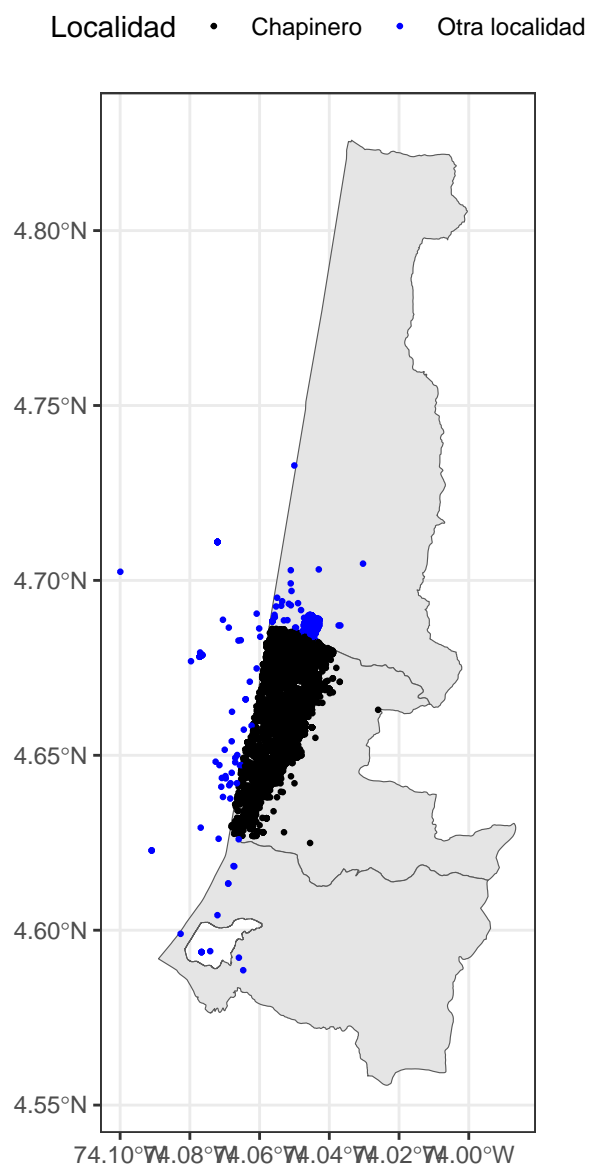
2024-05-13

1 Datos y Métodos

Los datos principales consistieron en los archivos descargados de la página de datos de la [competencia en Kaggle](#). competencia.

La información espacial de referencia se tomó del Mapa de Referencia para Bogotá D.C. del Ideca disponible [aquí](#). Al cruzar la capa de los polígonos las localidades con la tabla de *test* observamos que 593 propiedades se encuentran fuera de Chapinero (Figura 1), que es la localidad en la cuál se predecirán los precios de las viviendas.

Figura 1: Ubicación datos de validación



Características de la propiedad:

- Precio
- Fecha de publicación de clasificado
- Superficie total y superficie cubierta (80% y 76% valores faltantes)
- Número de habitaciones
- Número de baños
- Tipo de propiedad: apartamento o casa
- Ubicación georreferenciada

Spatial Ammenities: - parques - tiendas o centros comerciales -