**Taller Grupal No. 3**

Problem Set No. 3 – Making Money with ML?

# Introducción:

Por medio de este documento, se presentan los resultados al estudio y desarrollo de un modelo predictivo tendiente a la definición del precio de inmuebles en la localidad de Chapinero, Bogotá D.C., y su aplicación en la estrategia de compra de dichos bienes por el menor valor posible. Todo lo anterior, en el marco del Problem Set No. 3 del curso Big Data de la Facultad de Economía de la Universidad de Los Andes.

## El mercado inmobiliario: .

El mercado inmobiliario es uno de los mercados fundamentales para el desarrollo económico de una nación, toda vez que permite la generación de espacios para vivienda, comercio e industria a través de actividades de construcción que generan impacto en la producción, la prestación de servicios y el uso de instrumentos financieros, entre otros aspectos[[1]](#footnote-2). Es así, que el mercado inmobiliario tiene una relación importante con los consumidores y las firmas de inversión, ya que estos agentes acuden a los inmuebles, con recursos de ahorro o por medio del acceso a crédito, como una forma de inversión, bien sea para uso directo o a manera de renta[[2]](#footnote-3). En el contexto colombiano, podemos encontrar las siguientes cifras:

* Según el centro de estudios económicos del Banco BBVA, el sector de la construcción y venta de inmuebles representa el 13,5% del PIB nacional, y el 7,3% del empleo en el país[[3]](#footnote-4).
* En materia de vivienda para 2021, entre el 50% y 70% de la compra de este tipo de bienes, tuvo una vocación de inversión para renta, y no para uso propio. En materia de inmuebles comerciales y de industria, se encontró una disminución en su comercialización entre 2020 y 2021. Para 2022, inicia un crecimiento de las transacciones de venta y arriendo[[4]](#footnote-5).
* Según la Asobancaria, para 2022, se realizaron 128.200 operaciones de financiación para la compra de inmuebles, cuyo valor fue aproximadamente de 15.8 billones de pesos. Igualmente, aumentó el uso de instrumentos como el leasing habitacional y el crédito hipotecario[[5]](#footnote-6).

Todo esto, genera una necesidad importante: ¿Cómo definir el precio real de un inmueble? Al respecto, se debe considerar, no solo el avaluó comercial del bien, generalmente provisto por las autoridades territoriales, sino también las utilidades o servicios a los que puede acceder el poseedor del inmueble dentro de una zona determinada. Es así como, por ejemplo, el valor que le puede dar una pareja con hijos a un inmueble que se ubica en una zona con oferta de colegios es distinta a la de unas personas sin hijos respecto al mismo bien. Se resalta dentro de las investigaciones realizadas, la del autor Sherwin Rosen frente al uso de precios hedónicos en mercados implícitos y productos diferenciados[[6]](#footnote-7).

En este marco, es que un modelo predictivo que considere estas variables es importante para conocer las dinámicas del mercado y que consumidores y firmas conozcan de la mejor manera posible el verdadero precio del bien a ser adquirido o arrendado.

## Resumen de los resultados:

## Estructura del documento:

Para llegar a lo anterior, y considerando lo expuesto, el documento se dividirá en los siguientes acápites, a saber:

# Enlace a repositorio de GitHub:

El enlace de Github donde podrá encontrarse el repositorio con las respuestas del taller es el siguiente:

|  |
| --- |
| **Enlace al repositorio en GitHub** |
| <https://github.com/Carlosvergara1995/Problem_Set_3_Making_Money_with_ML.git> |

# Descripción de los datos:

A continuación, se presenta información sobre los datos utilizados dentro del ejercicio, así como las acciones tomadas para su limpieza y procesamiento.

## Origen de los datos:

Los datos utilizados para efectos de este análisis fueron sustraídos de la página web propertati.com.co, con información para 2021[[7]](#footnote-8). En este bloque de datos, se encuentran variables relacionados con la fecha de publicación del inmueble para venta y arriendo, su superficie, precio, el título, la descripción del inmueble y otros datos suministrados por el usuario. La información entregada consistió en dos (2) bases de datos: Una base de testeo (test) y una base de entrenamiento (training).

La información, a su vez fue complementada con datos del aplicativo OpenStreetMap para la ciudad de Bogotá y las Unidades de Planeación Zonales y localidad de Chapinero[[8]](#footnote-9).

## Proceso de limpieza y preparación de datos:

El proceso de limpieza de datos se dio con fundamento en las instrucciones del Taller, pero también buscando complementar la información que se encontraba dentro de las bases de datos provenientes del portal Propertati. Así, y dada el alto número de *missing values* se tomaron los datos, tanto de la descripción como del título de los inmuebles para completar los datos faltantes.

Adicionalmente, y considerando que la base de datos contaba con información de carácter geográfico, particularmente la latitud y longitud, se tomaron dichas bases, para convertirlas en data frames a través de funciones para la obtención de datos espaciales, convirtiendo las variables de latitud y longitud en una variable de coordenadas (coords=c(“lon”,”lat”) bajo el sistema de referencia WGS84. Con ello, se pudo delimitar el espacio de la variable de entrenamiento a la localidad de Chapinero, Bogotá D.C. con la correspondiente ubicación de los inmuebles.

En particular, utiliza la biblioteca “leaflet” para crear el mapa y “addTiles” para agregar un fondo de mapa. Luego, utiliza la función “addCircles” para agregar círculos a cada ubicación de la casa en el mapa, con el tamaño del círculo determinado por la variable especificada en el argumento “radius” y el color determinado por la variable especificada en el argumento “color”. En este caso, se utiliza la función “addCircles” en combinación con un objeto de datos llamado “house”, que se supone que es un objeto de datos espaciales que contiene la ubicación de las casas en las variables “lat” (latitud) y “lon” (longitud).

| **Extracto de los mapas interactivos de la Localidad de Chapinero, Bogotá D.C.** | |
| --- | --- |
| **Definición de la localidad** | **Ubicación de los inmuebles** |
|  |  |
| **Fuente**: Elaboración propia | |

Posteriormente, y para complementar la información de las bases de datos, se buscaron características relevantes (“amenities”), tanto para Bogotá D.C., como con Chapinero a través del código R de gCentroid para calcular los crentroides espaciales contenidos en la columna de coordenadas. Para efectos del ejercicio, esto se utilizó para calcular las distancias de cada inmueble con bares, escuelas, parques, restaurante y bancos.

## Estadísticas descriptivas:

# Modelo de clasificación:

# Modelo de regresiones:

# Modelo final:

# Conclusiones:

1. Al respecto, el sub-director del FMI para 2014 indicó, particularmente frente al mercado de inmuebles para vivienda: “*First, housing is an essential sector of the economy but also one that has been the source of vulnerabilities and crises. Hence, while the recent recovery in global housing markets is a welcome development, we need to guard against another unsustainable boom. Second, detecting over-valuation in housing markets is still more of an art than a science. Broad measures, such as house price to rent ratios, provide a first pass. But detailed analysis and judgment are needed to make a call about overvaluation. Third, the policy toolkit to manage housing booms is still under construction. A variety of tools have been used and the evidence suggests some short-run success. But more analysis and sharing of experience are needed on what works and what doesn’t. Conferences of this kind are useful in adding to our stock of knowledge*.” En este orden de ideas, la importancia del mercado inmobiliario, no solo tiene efectos financieros, sino también sociales y de bienestar, dadas las barreras de acceso, y especialmente la vivienda como parte de una garantía fundamental en el bienestar de los agentes. FMI. *Housing Markets, Financial Stability and the Economy.* (11 de junio de 2014). Disponible en: <https://www.imf.org/en/News/Articles/2015/09/28/04/53/sp060514>. [↑](#footnote-ref-2)
2. Frente a la relación del mercado de inmuebles y el gasto de los consumidores ver: Alan Carruth y Andrew Henley. *The Housing Market and Consumers Spending.* Fiscal Studies, vol. 11, No. 3. (1990). [↑](#footnote-ref-3)
3. BBVA Research. *Colombia. Situación Inmobiliaria.* (enero de 2023).Disponible en: <https://www.bbvaresearch.com/publicaciones/colombia-situacion-inmobiliaria-2022/> [↑](#footnote-ref-4)
4. Fondo Inmobiliario Colombia. *Panorama Inmobiliario 2022.* (marzo de 2022). Disponible en: <https://fondoinmobiliariocolombia.com/panorama-inmobiliario-2022/1794/> [↑](#footnote-ref-5)
5. InHaus Colombia. *El sector inmobiliario en cifras: del 2021 al 2022.* (noviembre de 2022). Disponible en: <https://www.lahaus.com/blog/tendencias-del-mercado/sector-inmobiliario-cifras-2021-2022>. [↑](#footnote-ref-6)
6. Sherwin Rosen. *Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition.* Journal of Political Economy, vol. 82, No. 1, (1974). [↑](#footnote-ref-7)
7. Properati es un portal inmobiliario para la presentación y búsqueda de anuncios clasificados para la venta y alquiler de bienes inmuebles, el cual tiene presencia sur-américa. El portal ofrece datos abiertos por medio de la licencia Creative Commons 3.0. Ver: <https://cloud.google.com/customers/properati?hl=es-419>; Storybench. *How an Argentine real estate platform is building visualizations for Latin American newsrooms.* (5 de julio de 2017), disponible en: <https://www.storybench.org/argentine-real-estate-platform-building-visualizations-latin-american-newsrooms/>. [↑](#footnote-ref-8)
8. OpenStreenMpas es una base de con datos geográficos de carácter colaborativo, la cual cuenta con licencias de uso abierto. El aplicativo fue creado en 2004 y se fundamenta en la metodología WSG 84 para la ubicación de latitudes y longitudes. Disponible en: <https://www.openstreetmap.org/about>. [↑](#footnote-ref-9)