**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: [xx]

# 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:

## Origen de los datos:

## Proceso de limpieza de datos:

Combina dos data frames utilizando la función bind\_rows() del paquete dplyr. Los data frames combinados son train y test.

Convierte la combinación de los data frames a un objeto sf utilizando la función st\_as\_sf() del paquete sf. El objeto sf representa datos espaciales, por lo que se espera que tenga coordenadas geográficas. En este caso, la función st\_as\_sf() espera que las coordenadas geográficas se encuentren en las columnas lon y lat, por lo que estas columnas se pasan como argumentos a la función coords=c(“lon”,“lat”). Además, se especifica que el sistema de referencia de coordenadas (CRS) es WGS84, el cual se indica con el código crs=4326.

creamos un mapa interactivo de chapinero que muestra la ubicación de las casas. 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).

El siguiente código de R utiliza la función “gCentroid()” del paquete “rgeos” para calcular los centroides de los objetos espaciales contenidos en la columna “geometry” de un objeto espacial llamado “parques\_chapinero”.

Para ello, primero se convierte la columna “geometry” del objeto espacial “parques\_chapinero” a un objeto “Spatial” utilizando la función “as()”. Luego, se pasa este objeto “Spatial” como argumento a la función “gCentroid()”, junto con la opción “byid = T”, que indica que se debe calcular el centroide para cada objeto individual en lugar de uno para todos los objetos.

El resultado de este código es un nuevo objeto espacial llamado “centroides\_chapinero”, que contiene los centroides de cada objeto espacial en la columna “geometry” del objeto espacial original “parques\_chapinero”. Este nuevo objeto puede ser utilizado para realizar operaciones espaciales adicionales, como la visualización o el análisis de la distribución espacial de los centroides.

El siguiente código de R utiliza el paquete “sf” para realizar operaciones espaciales con objetos espaciales.

En la primera línea, se utiliza la función “st\_as\_sf()” para convertir el objeto espacial “centroides\_chapinero” a un objeto espacial de la clase “sf”. Esto se hace para poder utilizar las funciones de la librería “sf” en los siguientes pasos del código.

En la segunda línea, se utiliza la función “st\_distance()” para calcular la distancia entre dos conjuntos de objetos espaciales: el objeto espacial “df\_chapinero” (que se asume que contiene los parques de Chapinero en formato “sf”) y el objeto espacial “centroides\_chapinero\_sf” (que contiene los centroides de los parques de Chapinero en formato “sf”).

El resultado de esta línea de código es una matriz de distancia que contiene las distancias euclidianas entre cada punto en “df\_chapinero” y cada centroide en “centroides\_chapinero\_sf”. Esta matriz de distancia puede ser utilizada para realizar análisis de agrupamiento espacial o para visualizar la distribución de los parques de Chapinero en relación a sus centroides.

La nueva columna contendrá el valor mínimo de cada fila de una matriz llamada “dist\_bar”, que se asume que tiene el mismo número de filas que el dataframe “df\_chapinero”.

La función “apply()” se utiliza para aplicar una función a cada fila de la matriz “dist\_bar”. En este caso, la función que se está aplicando es “min()”, que devuelve el valor mínimo de un vector. El argumento “1” en la función “apply()” indica que se debe aplicar la función a cada fila de la matriz.

Por lo tanto, el resultado de este código es que se añade una nueva columna a “df\_chapinero” que contiene el valor mínimo de cada fila de la matriz “dist\_bar”. Esto puede ser útil para realizar análisis o visualizaciones que involucren la distancia mínima entre cada parque en “df\_chapinero” y una serie de barras en “dist\_bar”.

***Chapinero - distancia al parque de la 93***

## 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 l 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)