

INFORME PROYECTO FINAL BOOTCAMP MUJERES EN TECH

Big Data 2022-2023



KEEPCODING
Tech School

Grupo Data Witches:

Carmen Rey González

María Berraquero García

Laura Daniela Morales Gutiérrez

María Victoria Ramello

Introducción

En este proyecto nos hemos centrado en la información recogida acerca de los alojamientos disponibles en la Comunidad de Madrid desde la plataforma de Airbnb.

Esta es una compañía internacional que permite a los usuarios ofertar alojamientos para que otros hagan uso de ellos a cambio de cierto precio. Dado su constante crecimiento económico y el aumento de su popularidad, resulta de gran interés estudiar qué factores influyen en la estimación del precio de los alojamientos.

Objetivo

Inicialmente partimos de la incógnita principal de orientar nuestro proyecto de cara al anfitrión o cliente de Airbnb. Valorando las posibilidades que nos ofrecía cada una, nos decantamos por la primera opción, ya que consideramos más interesante explorar los datos desde dicha perspectiva.

El primer objetivo que propusimos fue pronosticar la rentabilidad de un alojamiento en función a sus características a través de un modelo predictivo.

Sin embargo, pronto observamos que esto no iba a resultar fácil con los datos de los que disponíamos en el Data Set.

Por lo tanto, escogimos un objetivo menos ambicioso:

“Determinar los parámetros más relevantes para predecir el precio del alquiler de un alojamiento de la Comunidad de Madrid registrado en la plataforma de Airbnb y establecer un modelo capaz de pronosticar dicho precio en función a esos factores.”

Selección de las métricas

El primer filtrado se llevó seleccionando aquellas columnas que por su título intuíamos que nos resultarían útiles a lo largo del proyecto, también seleccionamos las filas correspondientes a la Comunidad de Madrid.

Aunque la correlación de algunas de estas variables con el precio, que es la métrica principal para nuestro objetivo, ha resultado ser moderada o buena; realmente, en la mayoría de los casos no se ha encontrado una correlación fuerte.

Se ha explorado la tasa de ocupación calculada, 'Occupancy', para conocer cuánto tiempo pasa un espacio rentado en relación con cuánto tiempo pasa disponible. Sin embargo, no resultaba fiable del todo. Se realiza una interpretación subjetiva y extrapolación de ciertos datos. Además, no encontramos relevancia suficiente para su uso en el modelo de predicción.

También analizamos la columna 'Amenities' donde cada alojamiento lista los servicios que se ofrecen. Se pretendía conocer si existía algún servicio en específico que influyera en el precio o si el número total de servicios ofrecidos tenía algún impacto en este. Con esto, se pudo observar que existía cierto impacto en la media del precio de aquellos alojamientos con servicios para cuidados de niños.

Sin embargo, hay que tener en cuenta que esto puede deberse a propiedades con precios muy extremos que casualmente ofrezcan estos servicios y que pueden existir propiedades que no hayan descrito todas las funciones que ofrecen. Podría resultar de interés hacer un análisis estadístico más profundo en el futuro.

En el resto de las columnas analizadas vimos un comportamiento similar al esperado. Aunque también se detectaron outliers para precios más altos que se alejan de la mediana, especialmente en las columnas de 'Property Type' y 'Room Type'.

Visualización de las métricas

Hemos utilizado Tableau para crear los Dashboards del precio y de los barrios por disponibilidad de habitaciones.

En ellas se pueden visualizar las variables más relevantes: Precio por código postal, por barrio, tipos de habitaciones y un evolutivo del precio a través de los años.

Se ha añadido un mapa con el promedio de precio por código postal y la información del precio y del barrio y un filtro en barrio por habitaciones disponibles que nos muestra la disponibilidad mediante un gráfico de barras.

Además, se agregó un filtro general por tipo de propiedad que nos filtra todos los gráficos para poder ir comparando.

Por último, se realizó un gráfico de tendencia donde podemos ver la evolución de los precios desde el año de comienzo de Airbnb

Conclusiones y reflexión

Con el desarrollo de este proyecto hemos tenido la oportunidad de poner en práctica lo estudiado a lo largo de cada uno de los bloques del Bootcamp.

Además, el procesamiento de los datos desde su extracción hasta la obtención del modelo predictivo nos ha permitido conocer las tareas a las que nos tendremos que enfrentar en nuestro futuro laboral de una forma más realista.

Por otra parte, podemos finalizar afirmando que la variable del precio depende de muchos factores y de formas diversas y complejas. Sería necesario llevar a cabo un análisis estadístico más profundo de los datos outliers y del tipo de correlación que existe entre los diferentes elementos. Con una investigación más detallada y, una vez que ya conocemos mejor los datos con lo que trabajamos, tal vez podríamos lograr crear un modelo predictivo de mayor calidad.

Como propuesta de mejora y lección aprendida, creemos que la principal el haber comenzado analizando los datos del Data Set.

A lo largo del proyecto, hemos observado ciertas dificultades al ver que, cierta información en la que habíamos invertido tiempo y esfuerzo, no tenía el efecto esperado. Este error surge la idea preconcebida y subjetiva de como creíamos que se determinaba el precio de los arrendamientos, sin fundamento en los datos.

Debido a esto y a una falta de organización del tiempo más efectiva, ha habido ciertos puntos en el desarrollo donde creemos que se han perdido oportunidades para acercarnos a resultados más acertados que los obtenidos.

Sin embargo, el proyecto realizado tiene posibilidad de perfeccionamiento con más tiempo de desarrollo si continuamos indagando en la observación estadística y análisis.