

Materia: Ciencia de Datos Aplicada
Actividad: Taller 1
Estudiante: Alvaro Andres Castiblanco Lopez
Código: 202413971

Actividades:

1. Selección de data set de trabajo.

El set de datos utilizados corresponde a la ciudad de Berlín Alemania. Se utilizaron dos de los repositorios listings.csv.gz y neighbourhoods.geojson. los demás datasets usados en el análisis son generados en el proceso de manipulación de datos.

2. Entendimiento de los datos:

Realizando la lectura de los datos se evidencian datos relevantes para el análisis y la generación de recomendaciones que se esperan obtener.

Se proceden a generar un nuevo dataset con los siguientes datos: id, name, description, host_id, latitude, longitude, property_type, room_type, accomodates, price, availability_365, review_scores_location, review_scores_value.

De estos datos se tomaran para el análisis los siguientes teniendo en cuenta la relevancia de los mismos para el análisis

- Valores de arriendo (valor del arriendo por noche).
- Ocupación (días vacantes de las propiedades en un periodo de 365 días).
- Número de personas que se pueden hospedar.
- Valoración de la Ubicación
- Valoración de lo Propiedad

3. Estrategia de análisis:

Se tomará el dataset y se realizara una depuración de la data excluyendo todos los datos que sean hoteles o habitaciones de un hotel debido a que estos datos no serán relevantes para asesoría que se desea realizar.

Posterior a la depuración de la información se tomará como primer criterio el valor de arriendo por noche ordenando la información por los lugares con mayor valor

Un segundo criterio de evaluación relevante es el tiempo de ocupación que se tiene sobre el predio, este dato es fundamental para que el cliente pueda ponderar un valor aproximado de recaudo por año.

El criterio de validación es el numero de personas para la cual esta diseñado el inmueble y la repetición del mismo en la toma de arriendo.

Los dos últimos valores nos darán una comprensión de la calificación de los sectores como de las propiedades con el fin de ubicar propiedades con características y ubicaciones óptimas para un retorno de inversión optimo.

4. Desarrollo de estrategia.

Posterior a la depuración de los datos se seleccionarán los primeros 100 datos de las propiedades con mayor valor de arrendamiento se realizará una grafica para validar la posición de cada una en la ciudad de Berlín.

Teniendo en cuenta las 100 propiedades con mayor valor de arrendamiento se validara cual es la ocupación en un periodo de un año de las mismas, esto con el fin de tener en cuenta la ocupación y no se adquiera una propiedad que sea alquilada o sea alquilada muy poco, teniendo en cuenta esta premisa se realiza una extracción de los datos tomando el top 30 de las propiedades con mayor ocupación en un periodo de un año. Obteniendo el resultado que para estos 30 registros el valor de arrendamiento oscila de \$300 a \$350, pese a que el valor mas distante tiene una ocupación cerca a 233 días, con el fin de realizar una comparación más precisa se toma le dataset original depurado y se realizara una organización por ocupación teniendo como resultado que existen propiedades que tiene una ocupación todo el año. Realizando una comparación con las propiedades con un valor de arrendamiento más elevado podemos obtener conclusiones frente a una taza de retorno.

Tomamos el dataset generado por valor más alto y realizamos el análisis de la capacidad de hospedaje frente al volumen de arrendamiento.

Al ultimo dataset generado se evalúan la calificación del sector y de la propiedad. De estos datos se obtiene valores nulos, por lo cual no todos pueden ser comparados

5. Generación de resultados.

<https://github.com/acastiblancol/Entregas4101>

Presentación Análisis de resultados “Berlin”