

Dataset: Habitaciones de hoteles

Truman Tapia Mora / Julio Zamora Guerrero

Descripción del dataset

El conjunto de datos generado en esta actividad reúne los principales atributos de habitaciones de hoteles de distintas ciudades del mundo.

Imagen identificativa



Ilustración 1. GH Hotel en Tenerife

Contexto

El objetivo de la práctica es la creación de un dataset a partir de los datos contenidos en un sitio web. Para llevar esto a cabo hemos elegido Hoteles.com (<https://es.hoteles.com/>).

Hemos elegido Hoteles.com ya que es uno de los principales proveedores mundiales de alojamiento hotelero y alquiler vacacional que ofrece servicios de reserva a través de su propia red de sitios web localizados, permitiendo acceder de manera centralizada a información sobre alojamientos, precios, servicios y disponibilidad.

Hotels.com L.P. es una de las compañías de viajes que conforman Expedia Group.

Contenido

Cada fila del dataset corresponde a una habitación de hotel. Estos son sus atributos:

- Search City → Ciudad del hotel
- Search Timestamp → Marca de tiempo de la búsqueda
- Check-in → Fecha de entrada al hotel
- Accommodation Name → Nombre del hotel
- Accommodation Direction → Dirección del hotel
- Accommodation Services → Servicios disponibles del hotel

- Accommodation Image URL → Foto portada de hoteles.com del hotel
- Reservation Link → Link a hoteles.com de la habitación
- Room Price → Precio del hotel en USD

Agradecimientos

Los datos han sido recolectados desde Hoteles.com haciendo uso del lenguaje de programación Python y de técnicas de Web Scraping para extraer la información alojada en los HTML de los enlaces de búsqueda por localización.

Inspiración

El presente conjunto de datos tiene la finalidad principal de ser objeto de estudio para estudiantes, ya que puede ser referencia para distintos proyectos en el ámbito de la ciencia de datos, como por ejemplo:

- Análisis exploratorio de los precios por localización y fecha de check in.
- Aplicación de técnicas de aprendizaje no supervisado (clustering de precios).
- Aplicación de técnicas de aprendizaje supervisado, para predicción de precio de los hoteles en distintas fechas.

Por supuesto también tiene un uso para crear una herramienta de monitorización de hoteles frecuentes para una organización, como por ejemplo puede ser el caso de una aerolínea que frecuenta los mismos hoteles en sus principales destinos.

Licencia

La licencia escogida para la publicación de este conjunto de datos ha sido CC BY-SA 4.0 License, cuyas cláusulas principales son:

- Atribución. Se debe dar crédito al nombre del creador del conjunto de datos generado, indicando los cambios que se han realizado. De este modo se reconoce el trabajo ajeno.
- Share Alike. Las contribuciones realizadas a posteriori sobre el trabajo publicado bajo esta licencia deberán distribuirse bajo la misma.

Código

<https://github.com/truman-t/Scrape-hoteles.com>

Dataset

<https://zenodo.org/record/5649244>

Tabla de contribuciones

Contribuciones	Firma
Investigación previa	TTM; JZG
Redacción de las respuestas	TTM; JZG
Desarrollo del código	TTM; JZG