AirBnb Data enunciado práctica TADM 21\_22. MADM

Grupo y nombre de cada usuario

2021-2022

Table of Contents

# Taller evaluable en grupos datos AirBnb

Enlace a estos datos de [AirBnb](http://insideairbnb.com/get-the-data.html) Generad un proyecto nuevo. Bajad lo datos de AirBnb a un carpeta/directorio que se llame AirBnb y dentro de AirBnb crear una carpeta/directorio que se llame model\_AirBnb.

Podéis (tenéis) que utilizar las ayudas del taller de estos datos.

## Instrucciones

* Entregad en grupos de prácticas.
* Se puede hacer con R o python.
* Hay que entregar el Rmd/notebook junto con su salida en html/pdf
* Máxima longitud: 10 páginas en pdf.
* Hay que cuidar la presentación, ortografía y redacción.
* Fecha entrega dd de mm de yyyy.

## Contexto de los datos

La página [Inside .Airbnb](http://insideairbnb.com/) <http://insideairbnb.com/> contiene información sobre los datos de los apartementos o residencias vacacionales puesto en alquiler en diversas locacizaciones del mundo.

Los datos recogidos están repartidos por diversas regiones, provicias, departamento, condados… del mundo. Los datos son [Open Source](https://www.redhat.com/es/topics/open-source/what-is-open-source) y los podemos usar ver [About Inside Airbnb](/http://insideairbnb.com/about.html) /<http://insideairbnb.com/about.html>.

En resumen el acceso y los diccionarios de datos y otras utilidades son accesiblea desde la [página principal](http://insideairbnb.com/get-the-data.html) o en los siguientes enlaces:

**Data Resources**

* [Get](http://insideairbnb.com/get-the-data.html) the data
* View [Data Dictionary](https://docs.google.com/spreadsheets/d/1iWCNJcSutYqpULSQHlNyGInUvHg2BoUGoNRIGa6Szc4/edit#gid=982310896)
* Read [Data Policies](http://insideairbnb.com/data-policies.html) including aligning data availability to the mission, Community Guidelines and policies on Archived and New Data
* Make a [Data Request](http://insideairbnb.com/data-requests.html) for Archived Data or Data for a new region

¡¡Atención!! el último servicio es de pago para datos de más de un año de antiguedad.

### Acceso a los datos

En el enlace [Get the data](http://insideairbnb.com/get-the-data.html) podemos descargar para cada ciudad los ficheros:

|  |  |
| --- | --- |
| File Name | Description |
| listings.csv.gz | Detailed Listings data for Name City. |
| calendar.csv.gz | Detailed Calendar Data for listings in Name City. |
| reviews.csv.gz | Detailed Review Data for listings in Name City. |
| listings.csv | Summary information and metrics for listings Name City (good for visualisations). |
| reviews.csv | Summary Review data and Listing ID (to facilitate time based analytics and visualisations linked to a listing) N/A Name City. |
| neighbourhoods.csv | Neighbourhood list for geo filter. Sourced from city or open source GIS files N/A Name City. |
| neighbourhoods.geojson | GeoJSON file of neighbourhoods of the city. |

Definid una carpeta datay dentro una carpeta por zona Mallorca, ValenciaBarcelonetc. Bajad los datos podeía utilizar el programitadownload\_city\_inside\_airbnb.R` de la raíz del github de la práctica.

### Especificación de las tablas de datos

Para entender cada tabla de datos tenemos que acceder al [Diccionario de datos](https://docs.google.com/spreadsheets/d/1iWCNJcSutYqpULSQHlNyGInUvHg2BoUGoNRIGa6Szc4/edit#gid=982310896).

Tenemos que comprender qué variables vamos a cargar y el tipo de datos. Como hay datos de topdo tipo tenemos que ir con espceial atención

* A los id de enteros largos que se puedan confundir con variables numéricas: Hay que leerlos como cadenas de caracteres.
* A ls variables numéricas que puedan contenewr caractres especiales: símnolo de dolar, símbolo de euro, %, separadores de miles…
* Variables que sean listas; por ejmplos extras de la vivienda [wifi,TV, piscina,….].
* Otros tipos especiales de variables: latitud , longitud, texto etc.

Como el problema es de datos sin una estructura clara cada grupo tendrá que estudiar las zonas:

* Mallorca
* Valencia
* Barcelona
* Varias ciudad más hasta completar (junto con las tres anteriores) el número de miembros del grupo.

### Bibliografía y sofware adicional

* Gráficos dinámicos con plotly: <https://plotly.com/r/animations/>
* MAPAS de España: <https://www.cienciadedatos.net/documentos/58_mapas_con_r.html> fijos

## Cuestión 1: Contexto del problema y modelo de datos (25%)

1. Cargar fichero listing.csv, calendar y reviews de cada ciudad. Tenés que estudiarlas y decidid qué tipo de dato y qué varibles cargáis. Hay que el explicar las transformaciones que ralicéis para manipular los datos; por ejemplo 50$ lo tranformo a 50, “2020-01-30” lño leo en tipo date….
2. Definid un **modelo de datos** con todas las tablas. Por ejemplo unid todos los listings de vuestras ciudades en una sola tabla, añadienco una variable que especifique la zona: Mallorca, Valencia, Barcelona, CiudadX, CiudadY….
3. Guarada el modelo de datos en ficheros .csv o ,Robj para la segunda parte de la práctica.
4. Redacatar un informe exlicando los tres apartados anteriores.

## Cuestión 2: Análisis exploratorio (EDA). (50%)

En las siguientes preguntas aplica todo lo que hemos visto acerca de la documentación en el EDA: Título de gráficos, etiquetas de los ejes, coloreado con información, leyendas, tablas bien presentadas (knitr)…

1. Calcual la frecuencia del número de rewies por apartamento. Es decir cuántos partamento tinen 1 review, 2 reviews, 3 reviews y así sucesivamente. Sigue al frecuencia de reviews por apartamentos vacacionales y el rango de reviews una “power law” (relación potencial).
2. Analizar el número de reviews por zona, barrio día de las semana y por meses.
3. De cda ciudad selecciona los 5 zonas/barrios con más apartamentos vacacionales. De estas zonas y para cada ciudad compara los precios medios (de todo el periodo), el número de habitaciones y el número de ocupación.

## Cuestión 3: Presentación final. (25%)

Presentación final