**Definición y validación de los datos**

En el presente documento trabajaremos con datos de Airbnb sacados del portal OpenDataSoft, que a su vez lo ha sacado del portal Inside Airbnb. Inside Airbnb es una iniciativa que se dedica a proporcionar información y datos relevantes sobre el impacto de Airbnb en las comunidades residenciales. En este caso la zona de estudio se trata de la capital de España, Madrid.

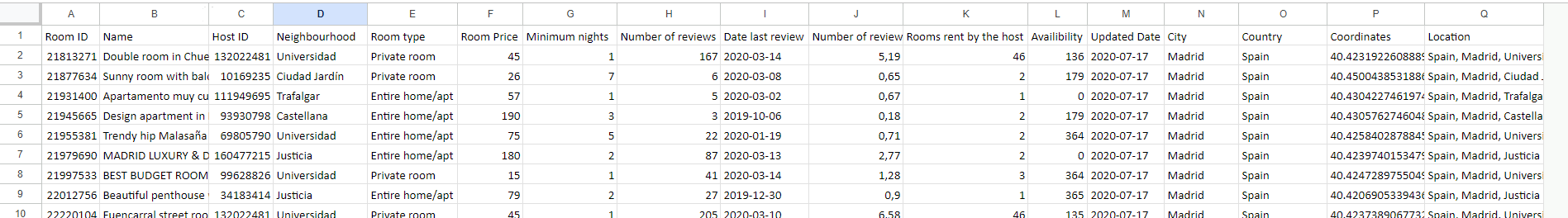
En el Dataset de entrada tenemos las siguientes 16 columnas con un total de 21.278 registros, el cual tienen la siguiente tipología de datos:

| **Room ID** | **Name** | **Host ID** | **Neighbourhood** | **Room type** | **Room Price** | **Minimum nights** |
| --- | --- | --- | --- | --- | --- | --- |
| *Double* | *Character* | *Double* | *Character* | *Character* | *Double* | *Double* |

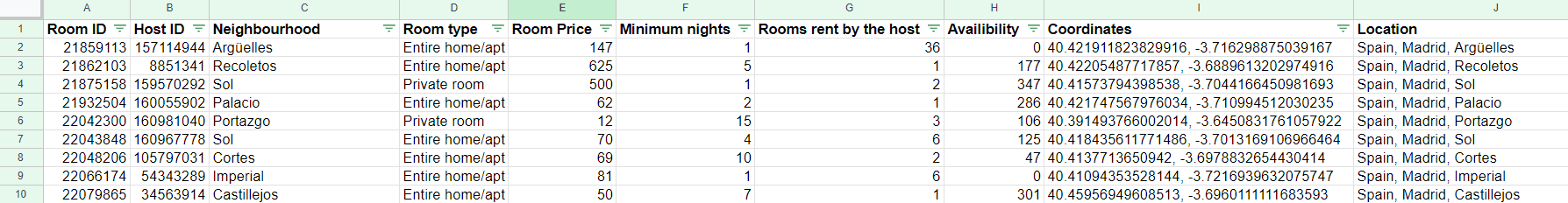
| **Number of reviews** | **Date last review** | **Number of reviews per month** | **Rooms rent by the host** |
| --- | --- | --- | --- |
| *Double* | *Double* | *Double* | *Double* |

| **Availibility** | **Updated Date** | **City** | **Country** | **Coordinates** | **Location** |
| --- | --- | --- | --- | --- | --- |
| *Character* | *Character* | *Character* | *Character* | *Character* | *Character* |

A continuación se enseña una pequeña muestra de los datos del fichero:



Seguidamente hemos procedido a hacer una primera selección de los datos que hemos querido utilizar, siendo estos los siguientes (10 columnas con 21.256 registros):



| **Room ID** | **Host ID** | **Neighbourhood** | **Room type** | **Room Price** | **Minimum nights** | **Rooms rent by the host** | **Availibility** | **Coordinates** | **Location** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *Double* | *Double* | *Character* | *Character* | *Double* | *Double* | *Double* | *Character* | *Character* | *Character* |

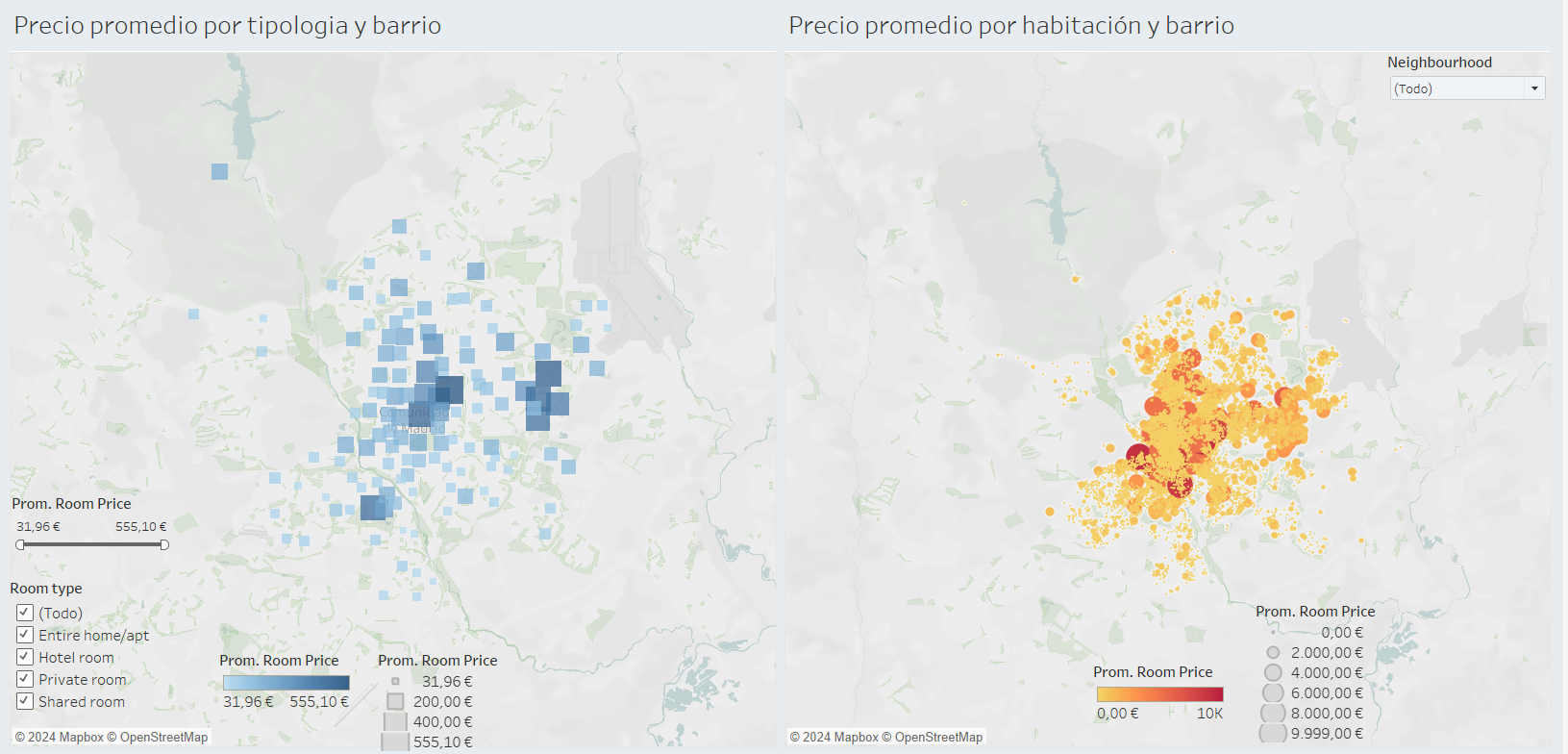
Los demás los hemos eliminado debido a que los consideramos irelevantes para nuestro análisis.

Después de este primer contacto con el Dataset, hemos procedido a ver que los datos que tenemos son los necesarios y a proceder a su normalización y revisión de su calidad, y que por lo tanto no contengan errores, tales como que sean los precisos de la zona de estudio, como que no tengan duplicidades, que estén en el formato correcto que no tenga valores nulos, entre otros.

**Visualización de datos en Tableau**

Para la visualización de datos, se ha planteado la pregunta siguiente:

* En qué zonas de Madrid y qué tipologías són las más caras para alquilar un apartamento dependiendo de su tipología.



Primeramente se han recalculado los campos de Latitud y Longitud creados anteriormente en el Dataset, seguidamente se ha procedido a hacer un dashboard con dos visualizaciones interactivas.

En el primer mapa se puede interactuar con el precio promedio que un usuario esté buscando según la tipología de habitación que necesite. Cómo más grande sea un cuadrado más caro será el precio, así cómo su tamaño

En el segundo caso se puede interactuar por el barrio y se muestra el promedio del precio de las habitaciones, siendo las zonas más caras las que tengan los círculos más grandes y más oscuros.