* REVIEWS\_PER\_MONTH : sacar los percentiles y clasificarlos de mejor a peor alojamientos.
* Una vez divididos en grupos:
  + Calcular las medias de reviews 🡪 útil para tableau
  + Calcular las medias de precios 🡪 útil para tableau
* NEIGHBOURHOOD\_CLEANSED: nos sirve para analizar los precios por barrio, como de masificado está etc
* PROPERTY\_TYPE: eliminar outliers : eliminar Villas (el precio se irá si calculamos por barrio)
* ROOM\_TYPE: importante para comparar
* ACCOMMODATES: sacar el precio por persona : PRICE\_CLEANSED / ACCOMODATE : PRICE\_CLEANSED\_PAX
* Ratio persona- cama - ACCOMODATED/BEDS
* PAX\_ HAB \_CLEANSED – ACCOMODATES/BEDROOMS
* BATHROOM, BEDROOM, BEDS : duda : como lo consideramos?
* BED\_TYPE: consideraremos camas reales.
  + Unificar tipos de cama en OTROS
  + REAL BED en BED

|  |  |
| --- | --- |
| 10.0 | 4 |
| AIRBED | 5 |
| COUCH | 15 |
| FUTON | 36 |
| PULL-OUT SOFA | 240 |
| REAL BED | 13026 |

* AMENITIES: identificar los que queremos y separarlos.
* REVIEWS SCORE: Mejor NAN – ver si luego lo puedo transformar otra vez en NaN
* SECURITY\_DEPOSIT: no usarlo para calcular el precio medio porque le das más peso a las casas que tienen fianza.
* CLEANING\_FEE: sumar al precio medio porque se paga sí o sí . NaN 🡪 0
* SECURITY\_DEPOSIT\_BOOL = Crear nuevo TRUE si hay fianza >0 y FALSE si no hay.
* MINIMUM\_NIGHTS : razonamiento : si hay un min de 10 noches igual no lo cojo
* CANCELLATION\_POLICY: transformarlo en columna con valor numérico: Para decir la media por barrio. i..e barrio Chamberí tiene una política de 3.5 por lo tanto es “estricto”
* REVIES SCORES : todas – con NaNs
* MONTHLY y WEEKLY: calcular los que tienen descuento y poner TRUE o FALSE
  + Importe del descuento también sacarlo.
  + Con el importe se podría sacar un % que sirva de recomendación de cuanto descuento deberían de hacer

Next steps:

* Cambios en Python hablados hoy
* Dividirnos por partes:
  + **SQL**: Lili
  + **Tableau**: Ana
  + **R :**