1. nthcpimmQuitamos las columnas que no entendemos el significado o que tienen información redundante

No sirve para analizar, solo para consultar:

* ***Listing\_url***
* ***Host\_url***

Campos de texto, no predeterminados por airbnb:

* ***Summary***
* ***Space***
* ***Neighbourhood overview***
* ***Notes***
* ***Transit***
* ***Access***
* ***Interaction***
* ***House rules***

Variables redundantes:

* ***Scrape\_id:*** se puede obtener con *last\_scrapped*
* ***Calendar\_last\_scrapped:*** se puede obtener con *last\_scrapped*
* ***Jurisdiction\_name:*** se puede obtener con city
* ***Country:*** se puede obtener con city
* ***Country\_code:*** se puede obtener con city
* ***Street:*** se puede obtener con state o city
* ***Market:*** se puede obtener con state o city

Información no relevante o no fiable, por no saber cómo se obtiene

**URLs:**

* ***Thumbnail\_url***
* ***Picture\_url***
* ***Medium\_url***
* ***Xl\_picture\_url***
* ***Host\_thumbnail\_url***
* ***Host\_picture\_url***

**Desconocimiento del método de cálculo:**

* ***Is\_location\_exact***
* ***Maximum\_maximum\_nights***
* ***Maximum\_minimum\_nights***
* ***Minimum\_minimum\_nights***
* ***Minimum\_maximum\_nights***
* ***Is\_business\_travel\_ready***
* ***Host\_verifications***
* ***Host\_identity\_verified***

**Referencia al período del scrapping**

* ***Maximum\_nights:*** usamos la variable *maximum\_nights\_avg\_ntm* como referencia
* ***Minimum\_nights:*** usamos la variable *minimum\_nights\_avg\_ntm* como referencia
* ***Calendar\_updated:*** referencia a las variables anteriores
* ***Has\_availability:*** referencia al período de scrapping, información no relevante

**Información no relevante**

* ***Require\_guest\_profile\_picture***
* ***Require\_guest\_phone\_verification***
* ***Host\_has\_profile\_pic***

1. Quitamos por nulos

Experiences\_offered: 100% nulos

-------------

Hemos cambiado los formatos de las fechas de los siguientes campos: Last\_scraped, host\_since, first\_review, last\_review

Cambiamos t por True y f por False (pasar de ser object a Boolean)

* Host\_is\_superhost
* Requires\_license
* Instant\_bookable

Categorical : cancellation\_policy (menos estricto a mas estricto)

Hemos cambiado los valores de la columna **city** por “Madrid” para unificarlos. Hemos cogido solo los que tenían los siguientes valores: “Madrid”, “MADRID”, “Centro, “Madrid Center”, “Madrid, Comunidad de Madrid, ES”, “Madrid”, “ Madrid”

Quitamos la de …..neighbourhood\_group\_cleansed