

## Evaluación: Reporte

### Etapas

En esta primera etapa, primeramente empecé cargando todos mis archivos de csv. Para posteriormente dividir mis datos en 2 dataframe, los numéricos y los de dato tipo string. Puesto que iba a reemplazar valores nulos dependiendo de la naturaleza de mis datos. Para los de tipo int y float, los reemplacé con un número 0, y para los de tipo string, los reemplacé con un before fill. Las que no se rellenaron, las vuelvo a rellenar con un doble guión para así tener 2 dataframes completamente limpios.

Posteriormente, utilicé solo mi dataframe de tipo numérico, para eliminar valores extremos con el método de rango intercuartil. Notamos una nueva distribución de los datos al graficarlos en un gráfico de tipo caja.

Finalmente, ya con mis 2 dataframes preprocesados y limpios, los concatenamos y volvimos a tener un solo dataframe por país, completamente limpio.

### Etapas

En esta etapa hicimos la extracción de diferentes datos. Las extracciones realizadas fueron las siguientes:

1. Obtener información de los únicos alojamientos en Airbnb que reciben una tasa de aceptación mayor al 50%, que para un servicio, una tasa de aceptación de solo la mitad, es bastante baja. Por lo que se realizó este filtro.
2. Obtener solo con alojamientos que su host tuvo un desempeño excepcional.
3. Obtener solo alojamientos con identidades no verificadas. Significa que estas respuestas son completamente anónimas, y de usuarios que no tienen identidad comprobada por airbnb.
4. Obtener solo alojamientos con característica de cuarto privado o cuarto en hotel.
5. Obtener alojamientos con más de un baño.
6. Obtener alojamientos mayor a 10,000 pesos y con característica de ser casa completa.
7. Los airbnb más limpios, con calificación arriba de 4.5.
8. Los mejores airbnb, con las calificaciones generales arriba de 4.9.
9. Los airbnb más solicitados, con menos de 100 días de disponibilidad al año.
10. Los alojamientos con hosts más atentos, que tienden a responder en un rango de 1 hr.

### Etapas

Finalmente, realizamos hallazgos de ciertas variables por medio de tablas de frecuencia.

- host\_response\_time: En esta variable se comparó como es el tiempo de respuesta de los host por país, siendo los hosts mexicanos los más atentos, ya que el 80% aproximado responde dentro de una hora, siendo los más desconsiderados los hosts españoles, ya que solo el 69% de los hosts responden en un rango de una hora.
- reviews\_scores\_location: En esta variable, encontramos la calificación dada al alojamiento dependiendo de su localización. Podemos ver, que en la ciudad de Girona, los alojamientos tienden a estar en zonas bastante buenas, siendo la ciudad con mayor frecuencia en las calificaciones más altas. También recordemos, que las

ciudades europeas, tienden a estar diseñadas para el peatón, mientras que las ciudades en América, son más extensas, por lo que es algo lógico que las localizaciones de Girona, tengan calificaciones más altas.

- `host_identity_verified`: En esta variable comparamos los hosts que tenían una identidad verificada ante airbnb. Como persona que renta un alojamiento, queremos un alojamiento donde nuestro anfitrión esté verificado por airbnb. La ciudad con los alojamientos más seguros, es la ciudad de México, con un valor de 92%.
- `property_type`: Aquí podemos comparar cuál es el tipo de alojamiento que más se repite. En California, es la ciudad que ofrece menos cantidad de alojamientos enteros, mientras que Girona, es la que ofrece más de este tipo. Además, la ciudad de México, es la que más tipo de alojamientos tiene en oferta.
- `reviews_score_value`: Esta variable es la calificación general otorgada a cada alojamiento. Girona, es la única ciudad entre las comparadas con alojamientos menores a 4. En cuanto a California y México, tienen la misma distribución en calificaciones.