

## Reporte de Evaluación

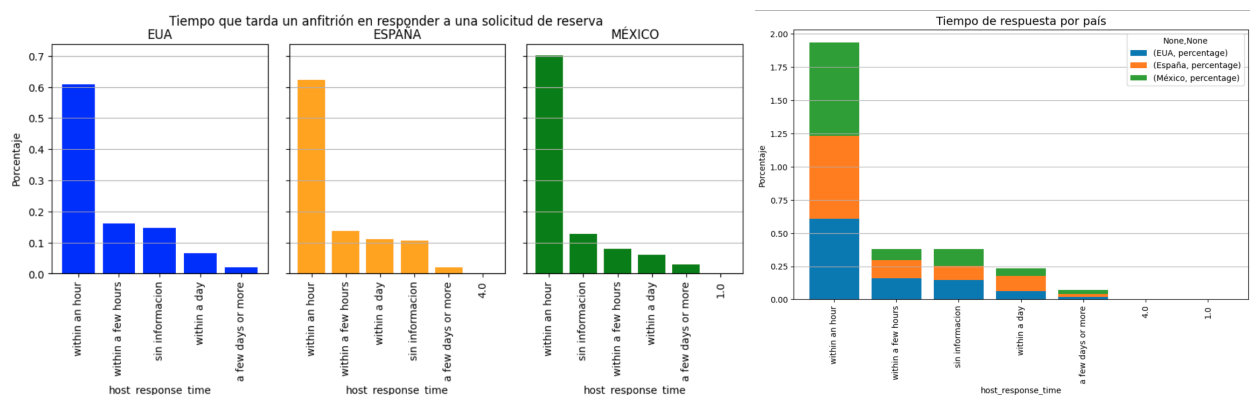
En esta evaluación se trabajó con 3 archivos (Girona\_España, DF\_Mexico y California\_EUA), en la primera etapa a cada archivo se le aplicaron métodos para reemplazar valores nulos ya fuera estableciendo un string en específico como “sin información”, con la media o mediana de la columna debido a la naturaleza de sus datos (cuantitativos) y por los métodos de backward fill y forward fill. La técnica de “forward fill” implica reemplazar los valores faltantes en una serie de datos con el valor de la observación más reciente anterior a los datos faltantes. La técnica de “backward fill” implica reemplazar los valores faltantes en una serie de datos con el valor de la observación más reciente posterior a los datos faltantes.

Una vez limpiados los archivos y que quedarán sin datos nulos, decidí aplicar el método de Rango Intercuartílico para eliminar Outliers de cada dataframe, decidí usar este método ya que la distribución de los datos no necesariamente es normal y puede contener valores atípicos que afecten a la media y desviación estándar. El método del IQR es más robusto y se basa en la mediana y el rango intercuartílico para identificar los valores atípicos.

Una vez eliminados los Outliers por medio del método de Rango Intercuartílico este da como resultado un archivo limpio de cada dataframe, así que podemos pasar a la siguiente etapa de extracción de datos la cuál consistió en aplicar diversos filtros a cada uno de los archivos limpios, como por ejemplo: que se mostrarán únicamente aquellos registros en los que la tasa de host\_acceptance\_rate fuera mayor al 50%, o sólo los registros que no hayan verificado identidad, etc.

Por último para la etapa 3 de Extracción de Características y Visualización de Datos escogí 5 variables para desarrollar un análisis comparativo respecto a las 3 ciudades elegidas para el servicio Airbnb de los cuáles los resultados fueron los siguientes:

### ● VARIABLE 1 : host\_response\_time



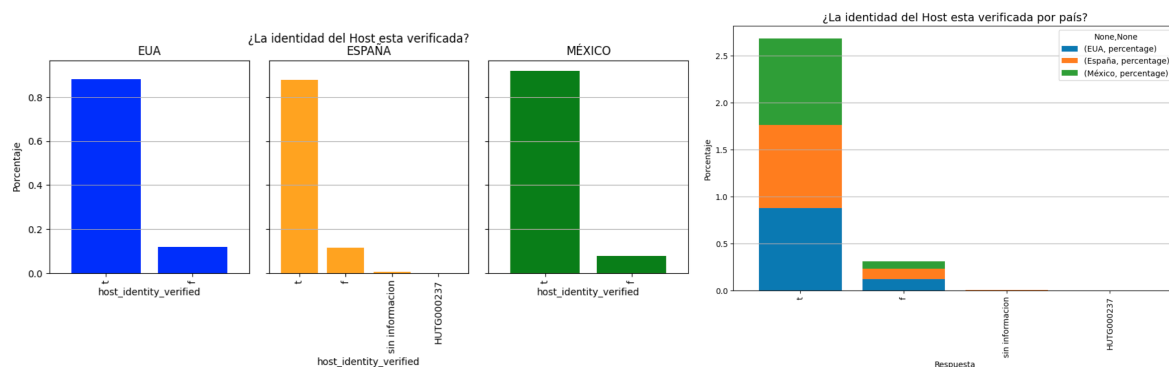
Las gráficas muestran la distribución de los tiempos de respuesta de los anfitriones en tres países: Estados Unidos (EUA), España (ESP) y México (MEX). Se puede observar que en los tres países, la mayoría de los anfitriones responden dentro de una hora o

unas pocas horas, siendo MEX el país con la tasa más alta de respuestas dentro de una hora (70.16%).

En general, ESP y EUA tienen una distribución similar de tiempos de respuesta, con alrededor del 60% de los anfitriones respondiendo dentro de una hora y un pequeño porcentaje que no proporciona información sobre sus tiempos de respuesta. Por otro lado, MEX tiene una mayor proporción de anfitriones que no proporcionan información sobre sus tiempos de respuesta (8.12%), pero también una mayor proporción de anfitriones que responden dentro de unas pocas horas (12.76%).

En términos de tiempos de respuesta más largos, EUA tiene la tasa más alta de anfitriones que tardan algunos días o más en responder, mientras que ESP y MEX tienen tasas similares de anfitriones que tardan más de un día en responder. En general, la mayoría de los anfitriones en los tres países responden rápidamente, pero todavía hay un pequeño porcentaje de anfitriones que tardan más tiempo en responder a las consultas de los huéspedes.

## ● VARIABLE 2 : host\_identity\_verified

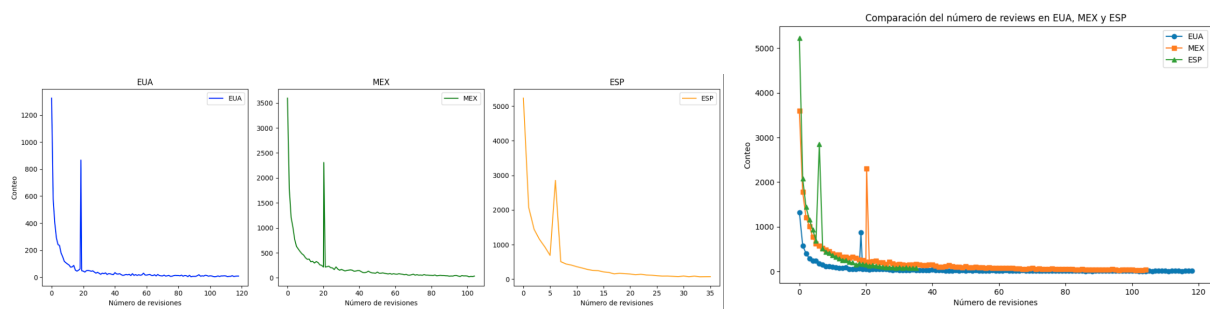


Las gráficas muestran la distribución de la verificación de la identidad de los anfitriones en tres países: Estados Unidos (EUA), España (ESP) y México (MEX). Se puede observar que en los tres países, la mayoría de los anfitriones han sido verificados en cuanto a su identidad, siendo MEX el país con la tasa más alta de anfitriones verificados (92.05%).

En EUA y ESP, alrededor del 88% de los anfitriones han sido verificados en cuanto a su identidad, mientras que en MEX, esta tasa es ligeramente más alta en comparación con los otros países. En cuanto a la verificación de la identidad falsa, EUA y ESP tienen tasas similares de anfitriones que no han sido verificados (11.97% y 11.50%, respectivamente), mientras que en MEX esta tasa es mucho más baja en comparación con los otros países (7.94%).

Es importante tener en cuenta que en ESP, hay un pequeño porcentaje de anfitriones que no proporcionan información sobre la verificación de su identidad y también hay un valor único (HUTG000237) que no es claro lo que representa. En general, parece que la mayoría de los anfitriones en estos tres países han sido verificados en cuanto a su identidad, lo que puede brindar más confianza a los huéspedes al hacer una reserva.

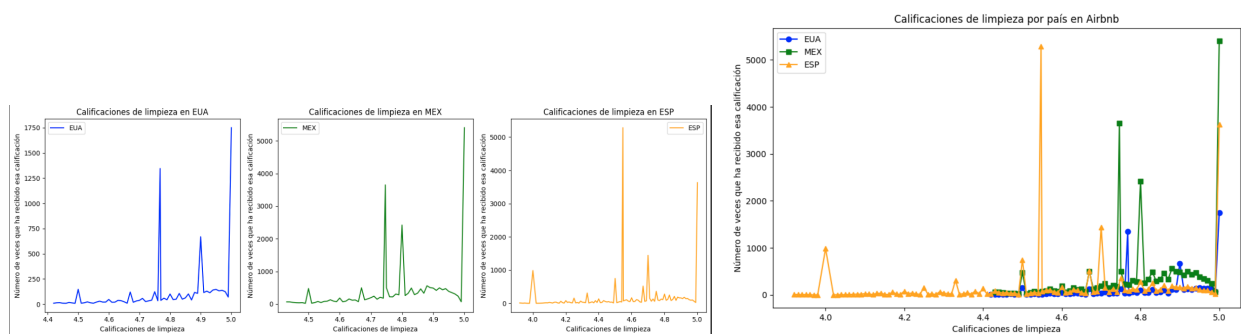
- **VARIABLE 3 : number\_of\_reviews**



Estos datos representan el número de revisiones (reviews) que ha tenido un listado en Airbnb en tres ciudades: España (ESP), Estados Unidos (EUA) y México (MEX). Para cada ciudad se ha registrado el conteo de listados que ha recibido cierto número de revisiones. Por ejemplo, en España hay 5229 listados que han recibido cero revisiones, 2074 listados que han recibido una revisión, 1443 que han recibido dos revisiones, y así sucesivamente.

En la tabla correspondiente a EUA, el conteo de listados que han recibido cierto número de revisiones va desde 0 hasta 118 revisiones. En la tabla de España, el conteo va desde 0 hasta 35 revisiones y por último en la tabla de México, el conteo va desde 0 hasta 104 revisiones.

- **VARIABLE 4 : review\_scores\_cleanliness**

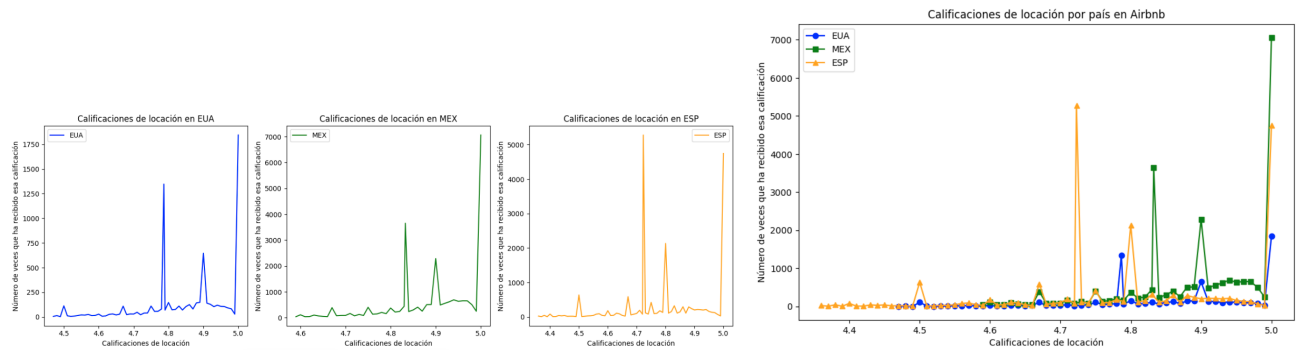


Esta es una variable numérica la cuál contiene información sobre las puntuaciones de limpieza de las listas de Airbnb en España, México y Estados Unidos.

Para comparar los tres países en términos de la puntuación de limpieza en las listas de Airbnb, podemos observar que en España, la mayoría de las listas tienen una puntuación de limpieza en el rango de 4,3 a 4,5, con la puntuación más común siendo 4. En los Estados Unidos, la mayoría de las listas tienen una puntuación de limpieza en el rango de 4,7 a 4,8, con la puntuación más común siendo 4.767581395. En México, la puntuación de limpieza más común también es 5, pero hay una amplia gama de puntuaciones diferentes entre 4.43 y 5.

Teniendo en cuenta estos datos, parece que las listas de Airbnb en los Estados Unidos son, en promedio, más limpias que las de España y México, aunque la variación en los datos de México hace que sea más difícil sacar conclusiones firmes. En general, es importante tener en cuenta que la limpieza de una lista de Airbnb depende del anfitrión y de cómo se mantiene la propiedad.

- VARIABLE 5 : review\_scores\_location



Al ser una variable numérica podemos comparar los tres países en términos de las calificaciones promedio de la ubicación de las propiedades. Podemos observar que en general, la mayoría de las evaluaciones de los tres países se concentran en el rango de 4.7 a 4.9, sin embargo, las propiedades en España tienen las calificaciones más altas de ubicación, con una calificación promedio, seguido de Estados Unidos, mientras que México tiene una calificación un poco más baja que los otros países.