**Actividad 5: Evaluación**

Miguel Saúl Fernández Avalos  
A01707491

**Etapas 1 y 2:**

En primer lugar, para manejar los datos se eliminaron columnas compuestas únicamente por valores NA ya que ninguna metodología de reemplazo iba a ser capaz de generar los datos, y con esto se eliminaron 3 columnas para California y Girona, y 4 para el DF.

Posteriormente, se realizaron las metodologías para eliminar valores nulos. Para eliminar datos cualitativos, se aplicaron los métodos de bfill y ffill, se aplicaron ambos, ya que en actividades anteriores se identificó que en caso de que dos instancias seguidas tengan NA, el aplicar sólo un método de éstos no permitía eliminar en su totalidad los NA, mientras que aplicando ambos métodos si se lograba.

Por su parte, para los datos cuantitativos se reemplazaron todos los NA por la mediana de su respectiva columna, se decidió la mediana y no la media porque a diferencia de ésta última, la mediana no se ve afectada por valores atípicos, por lo que es más representativa del centro de los datos.

Con esto se comprobó que ya no hubieran valores nulos, y se continuó con el procesamiento de valores atípicos.

Para le eliminación de valores nulos se utilizó la metodología de rango intercuartil, ya que éste método es más estricto para clasificar los datos como outliers por lo que aunque se pierda más información se tiene una distribución de datos más cercana a donde realmente se tiene información más representativa dentro del dataset, a comparación del método de desviación estándar, donde por la "holgura" que tiene para clasificar los datos como outliers la distribución puede continuar con afectaciones.

Para garantizar que la metodología de manejo de valores atípicos no generara ningún NA, se aplicaron los métodos de bfill y ffill de nueva cuenta, con lo que se obtuvo de nuevo un dataset sin NA.

De los filtros no hay mucho que hablar, se limpiaron los datos que no cumplieran con las condiciones solicitadas.

**Etapa 3**

De las columnas que se requería que se seleccionaran para analizar su tabla de frecuencias, se identificó que sólo 3 eran realmente categóricas para Python, ya que el resto las clasificaba como numéricas, esto complicó el análisis ya que la función sólo permitía usar variables no numéricas. Para solucionar esto, se creó una tabla más sencilla con el método value\_counts, que no afectó mucho ya que a final de cuentas se analizó la frecuencia y el resto de las columnas no eran tan relevantes.

Con esto, se decidió analizar las siguientes columnas, obteniendo los siguientes resultados:

* Host response time:
  + El DF tiene tanto el mayor porcentaje de arrendatarios con respuesta más rápida como de respuesta más lenta, con 80.09% y 3.49% respectivamente.
  + Girona es la ciudad con menor porcentaje de respuesta en el mismo día ya que poco más del 15% toma un día o más para responder.
  + En California se tiene el menor porcentaje de arrendatarios que responden en una hora menos del 70% una diferencia de más del 10% con el DF, y aún así tienen un nivel de respuesta similar durante el mismo día, es decir, el acumulado de respuestas en una hora con respuestas en varias horas.
* Host identity verified:
  + De nueva cuenta, el DF tuvo mejor desempeño en esta métrica, con más del 92% de los arrendatarios verificados, mientras que California tiene peor desempeño con el 87.7% de arrendatarios verificados.
* Property type:
  + En las 3 ciudades el tipo de propiedad más común es la renta de propiedad entera.
  + En DF y Girona se tienen 92 y 77 tipos de propiedades diferentes respectivamente.
* Review Score Cleanliness:
  + Los inquilinos son muy indulgentes, considerando que en las 3 ciudades las 5 estrellas son la calificación más común.
  + Los arrendatarios realizan tienen un buen sistema de limpieza, considerando que las calificaciones más bajas en las 3 ciudades rondan el 4.5 de calificación.
* Review scores value:
  + En DF los inquilinos son más indulgentes para evaluar a las propiedades, ya que, a diferencia de California y Girona, la calificación más común es 5 estrellas.
  + En Girona existen usuarios muy estrictos, ya que existen reviews de incluso menos de 4 estrellas, aunque son poco comunes.