

ANÁLISIS DE LOS VALORES NULOS Actividad 5.1, 5.2 y 5.3

5.1

En la Actividad 5.1 experimenté con diferentes métodos de tratamiento de valores nulos para comprender cómo funcionan los filtros. Primero importé las librerías necesarias y cargué el archivo Ventas Totales, luego identifiqué los valores nulos por columna y los sumé a nivel de DataFrame, obteniendo un total de 42 valores faltantes. Probé el primer método de sustitución reemplazando los nulos por el promedio, aplicándolo en columnas como Salón John Bajo Ventas y tarjetas_debito, y funcionó correctamente. Después utilicé el sexto método, aplicando backward fill para completar registros con el valor siguiente disponible, y también dio resultado. Posteriormente probé el método de forward fill, sustituyendo nulos en columnas como subtotal_Ventas_Alimento_Bebidas, confirmando que los valores se completaron. Finalmente verifiqué con `.isnull().sum()` que ya no había nulos en esas columnas. Aunque al inicio no tenía todo tan ordenado ni bien identificado, fue un buen ejercicio que permitió aplicar diferentes técnicas y comprobar que todos los filtros funcionaron de forma satisfactoria.

5.2

En la Actividad 5.2 comenzamos importando las librerías Pandas, NumPy y Matplotlib, después cargamos el archivo Gastos y Costos 2023 en formato Excel aplicando `skiprows=5` para omitir las primeras filas, ya que el archivo no se cargaba correctamente. A continuación identificamos los valores nulos con `isnull()` y al sumar observamos un total de 6706 valores faltantes. Para tratarlos aplicamos primero el método de backward fill (`bfill`), verificamos que sí se reflejaran los cambios y posteriormente utilizamos también forward fill (`ffill`) para completar los registros; en este caso no fue posible aplicar el método de sustitución por promedio, por lo que los métodos de relleno hacia adelante y hacia atrás fueron los que permitieron dejar el dataset limpio y listo para continuar con el análisis.

5.3

En la Actividad 5.3 importamos la librería pandas y cargamos el archivo Tokyo.csv, el cual era muy pesado para visualizar en GitHub; posteriormente identificamos los valores nulos con `data.isnull().sum()` y copiamos la información en un nuevo DataFrame llamado `data2` para trabajar con mayor seguridad; al revisar los resultados observamos que la única variable con valores faltantes fue `reviews_per_month`, por lo que aplicamos el primer método de sustitución de valores nulos, reemplazando los registros vacíos con el promedio de la columna mediante `.fillna`; finalmente corroboramos de nuevo los valores nulos y comprobamos que el conteo fue cero, dejando el dataset limpio y listo para análisis posteriores

