**Hallazgos de la Data Sucia o Problemáticas Detectadas**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Instacart\_Orders:**

1. **Valores Nulos (NaN, NULL):**
   * No se encontraron valores nulos en las columnas order\_id, user\_id, order\_number, order\_dow, order\_hour\_of\_day ni days\_since\_prior\_order. Esto es una buena noticia, ya que no es necesario realizar ninguna imputación ni eliminación de valores nulos.
2. **Registros Duplicados:**
   * Se encontraron **478,952 registros duplicados** en la tabla, lo que podría generar redundancia e inconsistencias en el análisis.
   * **Acción recomendada:** Eliminar los duplicados para garantizar que cada orden sea única, basándose en la columna order\_id.
3. **Valores Atípicos (Outliers):**
   * **En order\_number:** Se encontraron **60,362 registros** que presentan valores atípicos, lo que puede indicar un comportamiento anómalo en las órdenes.
     + Ejemplos de valores atípicos: 56, 63, 86, 78, etc.
   * **En order\_hour\_of\_day:** Se encontraron **6,360 registros** con valores atípicos.
     + Es posible que algunos valores de horas no sean realistas o estén mal formateados (por ejemplo, horas fuera del rango [0-23]).
   * **Acción recomendada:** Evaluar estos valores atípicos, dependiendo de su contexto y el análisis que se quiere realizar, se puede:
     + Eliminar los registros con valores atípicos si no son representativos.
     + Imputar valores razonables basados en la media o mediana de la columna.
4. **Inconsistencias en Tipos de Datos:**
   * Todos los tipos de datos parecen ser correctos y consistentes:
     + order\_id, user\_id, order\_number, order\_dow, order\_hour\_of\_day: int64.
     + days\_since\_prior\_order: float64.
   * **Acción recomendada:** No es necesario realizar ninguna conversión en este caso.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Order\_Products:**

1. **Valores Nulos (NaN, NULL):**
   * No se encontraron valores nulos en las columnas order\_id, product\_id, add\_to\_cart\_order ni reordered. Esto es positivo ya que no hay necesidad de imputación ni eliminación de valores nulos.
2. **Registros Duplicados:**
   * Se encontraron **449,158 registros duplicados** en la tabla. Esto puede deberse a que un pedido (order\_id) incluye varios productos (product\_id), pero algunos registros pueden repetirse innecesariamente.
   * **Acción recomendada:** Eliminar los registros duplicados basados en la combinación de las columnas order\_id y product\_id.
3. **Valores Atípicos (Outliers):**
   * **En add\_to\_cart\_order:** Se encontraron **59,362 registros** con valores atípicos.
     + Ejemplos de valores atípicos: 17.0, 35.0, 28.0, 16.0, etc. Estos valores representan el orden de los productos en el carrito, y pueden ser válidos, pero es importante evaluar si algunos son inesperados o erróneos.
   * **En order\_id, product\_id y reordered:** No se encontraron valores atípicos en estas columnas, lo que es positivo.
4. **Inconsistencias en Tipos de Datos:**
   * Los tipos de datos parecen correctos:
     + order\_id, product\_id: int64.
     + add\_to\_cart\_order: float64 (probablemente se debería tratar como un int64 ya que representa el orden de los productos).
     + reordered: int64.

**Products:**

1. **Valores Nulos (NaN, NULL):**
   * No se encontraron valores nulos en las columnas product\_id, product\_name, aisle\_id ni department\_id. Esto es positivo, ya que no es necesario realizar ninguna imputación ni eliminación de valores nulos.
2. **Registros Duplicados:**
   * Se encontraron **47,357 registros duplicados** en la tabla. Estos registros duplicados pueden ser el resultado de que ciertos productos se repitan innecesariamente en la base de datos.
   * **Acción recomendada:** Eliminar los registros duplicados basados en la columna product\_id para asegurar que cada producto esté representado una sola vez.
3. **Valores Atípicos (Outliers):**
   * **En product\_id, aisle\_id y department\_id:** No se encontraron valores atípicos en estas columnas, lo cual es positivo ya que las IDs no deben tener valores inesperados ni erróneos.
4. **Inconsistencias en Tipos de Datos:**
   * Los tipos de datos parecen correctos:
     + product\_id: int64.
     + product\_name: object (cadena de texto).
     + aisle\_id: int64.
     + department\_id: int64.