

Equipo: Ricardo

Ricardo Herrera Guerrero A01422698

Camila Murillo A00828860

Sara Ibarra A01734979

Hannia Martínez A01552310

21 de Abril, 2023

Reporte Valores Nulos

Al momento de trabajar con bases de datos de grandes empresas, es importante conocer y tener un buen manejo de los mismos justamente para un control transparente tanto con los clientes y hasta los propios trabajadores.

En esta actividad, se recopilaron los datos de los archivos “Datos de Facturación”, “Detalles precios y productos fabricados 2022” y “Gastos 20-23”. Como primera instancia, en los 3 archivos se procedió a convertirlos en dataframes para una mejor manipulación, luego se realizó un conteo por columna para identificar si había datos nulos, y de ser así, cuántos y qué columna.

Con el primer archivo “Datos de Facturación”, se encontraron datos nulos en las columnas de “clave vendedor”, “fecha de entrada” y “fecha de cancelación”. Para la primera columna se reemplazó por la clave “SC” que significa “Sin Cliente” con la función “fillna()”. Luego, tanto para las fechas de entrada y cancelación, se procedió a reemplazar con la función previamente dicha por el valor “00:00:00” para que se entienda que no hay datos de la fecha en la que se emitieron.

Con el segundo archivo, “Detalle precios y productos fabricados 2022” se observa la información del cliente y del vendedor al igual que sus datos de factura y lo que esto implica. Al momento de leer los valores nulos del archivo se puede observar que únicamente se tienen dos datos nulos en la parte de “NOMBRE”, por lo tanto no hay una manera en la cual podamos saber cuál fue el nombre del vendedor, así que optamos que en los datos nulos apareciera “Sin registro”.

Con el tercer y último archivo, “Gastos 20-23” se cuenta con 4 hojas, las cuales se depuraron cada una de manera individual.

-2020

En el 2020 se tuvo que eliminar las primeras 5 filas ya que solo entorpecen el control y manejo de la información, además se consideró que no contenían

información valiosa para nuestro proyecto. Después detectó que existían datos nulos en la columna de Folios , y al no existir información para poder completar las celdas con datos nulos se tomó la decisión de llenar dichas celdas con “Sin registro”.De igual manera se detectaron datos nulos en la columna de gastos, al no existir gastos y con el fin de no tener datos nulos se optó por llenar las celdas con el valor de 0.La columna de TC (tipo de cambio) de igual manera contaba con valores nulos por lo que optó por llenarlos con un 0. En la columna de importe también se detectaron celdas con valores nulos y se decidió llenarlas con el valor 0 ya que esto no tendrá ninguna influencia en la tabla, sólo el evitar valores nulos en el data frame.En la columna de IVA se detectó 256 valores nulos y al hacer un análisis de las demás celdas dentro de la columna se detectó que los datos nulos tendrían que ser el 16% de los valores de las celdas en la columna de importe , por lo que se decidió llenar las celdas con datos nulos multiplicando 0.16 por los valores de las celdas de la columna de importe y por último en la columna de póliza se decidió por llenar las celdas con datos nulos con las palabras de “sin registro”.

-2021

En el 2021 se tuvo que eliminar las primeras 3 filas ya que solo entorpecen el control y manejo de la información, además se consideró que no contenían información valiosa para nuestro proyecto. Las siguientes columnas fueron llenadas con la frase “Sin registro” ya que no existe ningún registro sobre la información que pertenece a las columnas;

- Tipo
- Póliza
- MP
- Folio

Las siguientes columnas fueron llenadas con el valor 0 ya que esto no afectará la información del data frame, sólo evitará los valores nulos.

- Tipo de gasto
- TC(Tipo de cambio)
- Importe

Por último en la columna de IVA se detectó valores nulos y al hacer un análisis de las demás celdas dentro de la columna se detectó que los datos nulos tendrían que ser el 16% de los valores de las celdas en la columna de importe , por lo que se decidió llenar las celdas con datos nulos multiplicando 0.16 por los valores de las celdas de la columna de importe.

-2022

En el 2022 se tuvo que eliminar las primeras 3 filas ya que solo entorpecen el control y manejo de la información, además se consideró que no contenían

información valiosa para nuestro proyecto. Las siguientes columnas fueron llenadas con la frase "Sin registro" ya que no existe ningún registro sobre la información que pertenece a las columnas;

-Folio

-MP

-Póliza

Las siguientes columnas fueron llenadas con el valor 0 ya que esto no afectará la información del data frame, sólo evitará los valores nulos.

-TC

-Otros

-2023

En el 2023 se tuvo que eliminar las primeras 3 filas ya que solo entorpecen el control y manejo de la información, además se consideró que no contenían información valiosa para nuestro proyecto. Las siguientes columnas fueron llenadas con la frase "Sin registro" ya que no existe ningún registro sobre la información que pertenece a las columnas;

-Folio

-MP

-Póliza

Las siguientes columnas fueron llenadas con el valor 0 ya que esto no afectará la información del data frame, sólo evitará los valores nulos.

-Otros

-FP