



Tecnológico de Monterrey

Materia: Analítica de Datos y Herramientas de
Inteligencia Artificial II

Valores Nulos

Grupo: 101

Estudiantes:

Abner Palma García - A01735312

Edwin Nepomuceno Rivera - A01174706

Luis Alberto Mirón Toruño - A01735489

Paula Gabriela Armenta Nazario - A01735618

Fecha: 21 de abril del 2023

Datos de facturación

Antes de comenzar con el procedimiento de sustitución de valores nulos, se procedió a hacer un análisis de cuantos datos de este tipo existen en el archivo. En análisis se encontró con que existían 10588 valores nulos, los cuales se repartían en las siguientes columnas:

CVE_CLPV: 1

CVE_VEND: 48

FECHA_ENT: 2

FECHA_CANCELA: 10537

Teniendo esta información procedemos a la sustitución de valores nulos

En el caso de la columna CVE_CLPV, decidimos utilizar el método de sustitución de valores nulos, en este caso con mediana, ya que al ser datos que no tienen una gran diferencia entre sí, además de que solo es un dato nulo el que existe dentro de dicha columna, por lo que en un análisis de otro tipo no se ve afectado.

Continuando con el proceso en la columna CVE_VEND, se decidió utilizar de nueva cuenta el método de sustitución de valores nulos, en este caso con mediana, ya que, al momento de sustituir los valores nulos con este método, no altera de manera negativa los datos para un posterior análisis.

Así mismos para la columna FECHA_ENT, se utilizó el método de sustituir valores nulos por un string en concreto, el dato que se utilizó en cuestión fue "2024-07-31 00:00:00", para así nosotros saber que la fecha de esas facturas es porque no tenían información.

En la última columna en donde existían más datos nulos, es la de FECHA_CANCELA, en esta columna se decidió utilizar el método de sustituir valores nulos por un valor número en concreto, el cual fue el de "0000" ya que esas columnas que tienen valores nulos es porque su factura no fue cancelada sino, emitida, es por ello que optamos por ponerle un "0000".

Detalles de precios

En este archivo también se procedió a hacer un análisis de este para poder de esta manera encontrar los valores nulos que existen dentro de este mismo, siendo el resultado el hallazgo de tan solo dos valores nulos, encontrados en la columna "NOMBRE".

Para poder hacer la sustitución se decidió implementar el método de sustituir valores nulos por un string en este caso al ser el nombre de los vendedores y no tener mas datos o como consultar los nombres, se ha decidido sustituir el valor nulo, por "DESCONOCIDO", ya que esto nos ayudará a identificar de forma más óptima este dato.

Gastos y costos

Para este caso el documento viene dividido en diferentes hojas, que representan los años, así que para poder realizar el análisis correspondiente se dividió de la misma manera.

Para ello primero seleccionamos las columnas de interés las cuales fueron, FECHA, FOLIO, PROVEEDOR, DESCRIPCIÓN, TIPO GASTO, IMPORTE, IVA y TOTAL MX. Posteriormente se identificaron los datos nulos del dataframe los cuales se establecieron de la siguiente manera, en las siguientes columnas:

2020

FOLIO: 189

IMPORTE: 34

IVA: 268

Para la columna de folio se utilizó la de poder sustituir el valor, por un string en concreto en este caso se decidió utilizar el formato "XXXX" para poder sustituirlo y por consiguiente poder identificarlo en un análisis posterior.

Para la columna IMPORTE e IVA de importe se procedió a sustituir los valores nulos que se encontraron por el valor "0" esto debido a que se desconoce el por qué no se tiene el registro de esta información.

2021, 2022

Para esta hoja de cálculo se volvió a realizar los mismos análisis de valores nulos y estos se encontraron en las siguientes columnas:

2021-FOLIO: 147

2022-FOLIO:102

Al igual que en la hoja anterior se utilizó el método de sustituir los valores nulos encontrados por un string en concreto en este caso también se utilizó el método "XXXX" para sustituirlo y ser capaces de identificarlos más adelante.

2023

Para este año los valores nulos se encontraron en las siguientes columnas:

FOLIO:13

TIPO GASTO: 8

Para la columna FOLIO se utilizó el método de sustituir los valores nulos encontrados por un string en concreto en este caso también se utilizó el método "XXXX" para sustituirlo y ser capaces de identificarlos más adelante. Y de igual manera para la columna TIPO GASTO se utilizó el mismo método y en este caso se decidió utilizar el string "Desconocido"