

INFORME

El objeto de la elaboración de este informe radica en la comprensión y el acercamiento a los set de datos que la empresa ha facilitado para su posterior tratamiento, elaborando un diccionario de datos.

En la base de datos de la empresa se encuentran siete datasets o conjuntos de datos. Ellos son:

- Venta
- Compra
- Gasto
- Clientes
- Localidades
- Proveedores
- Sucursales

A continuación, se procede a explicar brevemente cómo se encuentra conformado cada uno a partir de una descripción de la información que nos brinda cada columna.

VENTA

Esta tabla contiene 46180 instancias y está compuesta por 10 columnas. A saber:

- IdVenta: identificador utilizado para registrar cada operación
- Fecha: fecha en que se realizó la venta
- Fecha_entrega: fecha en que fue entregado el producto
- IdCanal: medio por el cual se realizó la venta
- IdCliente: identificador del cliente
- IdSucursal: identificador de la sucursal que hizo la venta
- IdEmpleado: identificador del empleado que realizó la venta
- IdProducto: identificador del producto
- Precio: precio del producto
- Cantidad: unidades vendidas

COMPRA

Este dataset consta de 11539 registros y 9 columnas. Ellas son:

- IdCompra: identificador de la operación de compra
- Fecha: fecha en que se hizo la compra
- Fecha_Año: año de la compra
- Fecha_Mes: mes de la compra

- Fecha_Periodo: conjunción de las dos columnas anteriores
- IdProducto: identificador del producto comprado
- Cantidad: unidades compradas de ese producto
- Precio: precio de compra del producto
- IdProveedor: identificador del proveedor del producto

GASTO

Para este dataset se cuenta con 8640 instancias con 5 columnas. Los atributos son:

- IdGasto: identificador del registro
- IdSucursal: identificador de la sucursal
- IdTipoGasto: identificador del tipo de erosión
- Fecha: fecha en que se realizó la operación de compra
- Monto: valor en unidad monetaria de la compra

CLIENTES

En este set de datos de 2672 registros y 10 columnas se registran los clientes de la empresa. Los features son:

- ID: identificador del cliente
- Provincia: provincia de residencia del cliente
- Nombre_y_Apellido: datos del cliente
- Domicilio: dirección
- Teléfono: teléfono del cliente
- Edad: edad del cliente
- Localidad: localidad de residencia
- X: longitud
- Y: latitud
- col10: columna sin información

LOCALIDADES

En este caso, se toma como dataset el provisto por el Estado argentino a través de la página:

https://datos.gob.ar/ar/dataset/jgm-servicio-normalizacion-datos-geograficos/archivo/jgm_8.14.

Con ello, se logra conectar con la API que permite realizar correctamente la normalización de todos los registros de las demás tablas. Vale aquí destacar que además la tabla original de localidades provista por la empresa contaba con muchas inconsistencias e incongruencias. Esencialmente en lo relativo a la localización de la Ciudad Autónoma de Buenos Aires. Esto se pudo resolver importando la tabla de localidades que facilita el portal de Datos Abiertos del Estado Nacional argentino.

Las columnas que se utilizaron son las siguientes:

- Centroide_lat: centro de latitud de la localidad
- Centroida_lon: centro de longitud de la localidad
- Departamento_nombre: nombre del departamento
- Municipio_nombre: nombre del municipio en que se encuentra la localidad
- Nombre: localidad
- Provincia: provincia a la que pertenece la localidad

PROVEEDORES

El dataset de proveedores incluye un total de 14 filas que se estructuran en 7 campos. Estos campos son:

- IDProveedor: identificador del proveedor
- Nombre: nombre del proveedor
- Address: su dirección
- City: ciudad en la que se encuentra
- State: provincia
- Country: país
- departamen: departamento

SUCURSALES

Por último, se encuentra un set de datos que representa a las sucursales que tiene la empresa. En total son 31 y tenemos 7 columnas que nos brindan más información de cada una de ellas:

- ID: identificador de la sucursal
- Sucursal: nombre
- Dirección: domicilio
- Localidad: localidad en la que se encuentra la sucursal
- Provincia: provincia
- Latitud: coordenada de latitud
- Longitud: coordenada de longitud