**Compras.csv**

dtypes: float64(1), int64(7), object(1)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Nombre | Tipo de dato | Formato | Descripción | Ejemplo |
| IdCompra | Int64 | NNNNN | Id de la compra | 1 |
| Fecha | Object | M/DD/YYYY | Fecha en que se realizó la compra | 1/30/2015 |
| Fecha\_año | Int64 | YYYY | Año en que se realizó la compra | 2015 |
| Fecha\_Mes | Int64 | M | Mes en que se realizó la compra | 1 |
| Fecha\_periodo | Int64 | YYYYMM | Año y mes en que se realizó la compra | 201501 |
| IdProducto | Int64 | NNNN | Id del producto que se compro | 42978 |
| Cantidad | Int64 | D | Cantidad de productos que se compraron | 6 |
| Precio | float | NN,NN | Precio unitario del producto que se compro | 81.24 |
| IdProveedor | Int64 | NN | Id del proveedor al que se le hizo la compra | 10 |

**¿Hay datos faltantes?**:Cantidad de registros 11539

* IdCompra 11539
* Fecha 11539
* Fecha\_Año 11539
* Fecha\_Mes 11539
* Fecha\_Periodo 11539
* IdProducto 11539
* Cantidad 11539
* Precio 11172
* IdProveedor 11539

Se puede visualizar una falta de datos en la columna “Precio”

**Gasto.csv**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Nombre | Tipo de dato | Formato | Descripción | Ejemplo |
| IdGasto | Int | NNNN | Id del gasto que se hizo | 8636 |
| IdSucursal | Int | NN | Id de la sucursal que realizo el gasto | 24 |
| IdTipoDeGasto | Int | N | Id del tipo de gasto que se hizo | 4 |
| Fecha | Object | YYYY-MM-DD | Fecha de cuando se hizo el gasto | 2020-12-01 |
| Monto | Float | NNN.NN | Monto de dinero que se gasto | 1156.31 |

dtypes: float64(1), int64(3), object(1)

**¿Hay datos faltantes?**: Cantidad de registros 8640

* IdGasto 8640
* IdSucursal 8640
* IdTipoGasto 8640
* Fecha 8640
* Monto 8640

No se encontraron datos faltantes en ninguna columna

**Localidades.csv**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Nombre | Tipo de dato | Formato | Descripción | Ejemplo |
| Categoría | Object | ------ ----- | Tipo de localidad | Localidad simple |
| Centroide\_lat | Float | -NN.NNNN | Coordenadas latitud de la localidad | -35.033073 |
| Centroie\_lon | Float | -NN.NNNN | Coordenadas longitud de la localidad | -60.280620 |
| Departamento\_id | Float | DDDD.O | Id del departamento donde se encuentra la localidad | 6021.0 |
| Departamento\_nombre | Object | ----------- | Nombre del departamento donde se encuentra la localidad | Alberti |
| Fuente | Object | NNNNN | Fuente de donde se extrajeron los datos | INDEC |
| Id | Int | NNNNNNNNN | Id de la localidad | 6021020000 |
| Localidad\_censal\_id | int | NNNNNNNN | Id de la localidad según el censo | 6021020 |
| Localidad\_censal\_nombre | Object | ----------------- | Nombre de la localidad | Mechita |
| Municipio\_id | Float | NNNNNN | Id del municipio | 60021.0 |
| Municipio\_nombre | Object | ------------------ | Nombre del municipio | Alberti |
| Nombre | Object | ------------------ | Nombre de la localidad | VILLA GRISOLIA |
| Provincia\_id | Int | N | Id de la provincia donde se encuentra la localidad | 6 |
| Provincia\_nombre | Object | --------------- | Nombre de la provincia donde se encuentra la localidad | Buenos Aires |

dtypes: float64(4), int64(3), object(7)

**¿Hay datos faltantes?**: Cantidad de registros 4142

* Categoría 4142
* Centroide\_lat 4142
* Centroide\_lon 4142
* Departamento\_id 4141
* Departamento\_nombre 4141
* Fuente 4142
* Id 4142
* Localidad\_censal\_id 4142
* Localidad\_censal\_nombre 4142
* Municipio\_id 3599
* Municipio\_nombre 3648
* Nombre 4142
* Provincia\_id 4142
* Provincia\_nombre 4142

Se visualizan datos faltantes en las columnas “Municipio\_nombre”, “municipio\_id”, “departamento\_id” y “departamento\_nombre”

Las columnas tienen nombres muy largos y se podría dividir en varias tablas para una mejor distribución y control de los datos

**Proveedores.csv**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Nombre | Tipo de dato | Formato | Descripción | Ejemplo |
| IDProveedor | Int | NN | Id del proveedor del negocio | 5 |
| Nombre | Object | ------------ | Nombre del proveedor | Bell S.A |
| Address | Object | ------------ | Dirección de donde se encuentra el proveedor | MORENO 331 |
| City | Object | ------------- | Nombre de la ciudad o barrio donde se encuentra | AVELLANEDA |
| State | Object | -------------- | Nombre de la provincia en donde se localiza el proveedor | BUENOS AIRES |
| Country | Object | ----------------- | País donde está el proveedor | ARGENTINA |
| Departamen | Object | ---------------- | Departamento donde se encuentra el proveedor | LANUS |

**¿Hay Datos faltantes?**: Cantidad de registros 14

* IDProveedor 14
* Nombre 12
* Address 14
* City 14
* State 14
* Country 14
* Departamen 14

Se encontraron datos faltantes en la columna “Nombre”

Las columnas tienen nombres en inglés, habría que cambiarlo para que se puedan complementar con las demás tablas. La mayoría de los registros están en mayúsculas

**Sucursales.csv**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Nombre | Tipo de dato | Formato | Descripción | Ejemplo |
| ID | Int | NN | Id de la sucursal comercial | 4 |
| Sucursal | Object | ------------- | Nombre de la sucursal | Palermo 1 |
| Dirección | Object | -------------- | Dirección de la sucursal comercial | Av. Corrientes 2352 |
| Localidad | Object | ------------------ | Localidad donde se encuentra la sucursal | CABA |
| Provincia | Object | ---------------- | Provincia donde se encuentra la sucursal | Córdoba |
| Latitud | Object | ------------- | Coordenadas de latitud de donde se encuentra la sucursal | -34,5678060 |
| Longitud | Object | ----------- | Coordenadas de longitud donde se encuentra la sucursal | -64,1864580 |

**¿Hay datos faltantes?:** Cantidad de registros 31

* ID 31
* Sucursal 31
* Dirección 31
* Localidad 31
* Provincia 31
* Latitud 31
* Longitud 31

**Venta.csv**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Nombre | Tipo de dato | Formato | Descripción | Ejemplo |
| IdVenta | Int | NNNNN | Id de la venta que se realizo | 48241 |
| Fecha | Object | YYYY-MM-DD | Fecha en que se efectuó la venta | 2020-12-30 |
| Fecha\_entrega | Object | YYYY-MM-DD | Fecha en que se entregó el producto vendido | 2021-01-06 |
| IdCanal | Int | N | Id del canal por el cual se realizo la venta | 2 |
| IdCliente | Int | NNNN | Id del cliente que realizo la compra | 2389 |
| IdSucursal | Int | NN | Id de la sucursal donde se hizo la venta | 26 |
| IdEmpleado | Int | NNNN | Id del empleado que concreto la venta | 3836 |
| IdProducto | Int | NNNNN | Id del producto que se vendió | 42839 |
| Precio | Float | NNN.NN | Precio unitario del producto que se vendió | 52.0 |
| Cantidad | Float | NNN.NN | Cantidad de productos que llevo | 1.0 |

dtypes: float64(2), int64(6), object(2)

**¿Hay datos faltantes?:** Cantidad de registros 46180

* IdVenta 46180
* Fecha 46180
* Fecha\_Entrega 46180
* IdCanal 46180
* IdCliente 46180
* IdSucursal 46180
* IdEmpleado 46180
* IdProducto 46180
* Precio 45260
* Cantidad 45296

Se visualizan datos faltantes en las columnas “Precio” y “Cantidad”

A esta columna de ventas se le sumo el archivo “Ventas\_Dic2020.csv”, y completo un total de 46645 registros

**Clientes.csv**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Nombre | Tipo de dato | Formato | Descripción | Ejemplo |
| ID | Int | NN | Id del cliente | 5 |
| Provincia | Object | -------------- | Nombre de la provincia donde vive el cliente | Buenos Aires |
| Nombre\_y\_Apellido | Object | NN AA | Nombre y apellido del cliente | JAVIER ALONSO RIVEIRO |
| Domicilio | Object | -------------- | Domicilio donde reside el cliente | CALDERON DE LA BARCA 498 |
| Telefono | Objetc | NN-NNNN | Teléfono del cliente | 49-2883 |
| Edad | Float | NN.0 | Edad del cliente | 34.0 |
| Localidad | Object | ------------- | Ciudad donde vive el cliente | SANTA ROSA |
| X | Object | --------------- | Coordenadas de longitud donde vive el cliente | -59,12794068 |
| Y | Object | -------------- | Coordenadas de la latitud donde vive el cliente | -31,32123448 |
| Col10 | Float |  | La columna esta llena de np.Nan | Nan |

dtypes: float64(1), int64(2), object(7)

La columna “Col10” contiene todos valores Nan, recomiendo eliminarla

**¿Hay datos faltantes?:** Cantidad de registros 2384

* ID 2384
* Provincia 2359
* Nombre\_y\_Apellido 2365
* Domicilo 2341
* Telefono 2354
* Edad 2383
* Localidad 2359
* X 2384
* Y 2384
* Col10 0

Las únicas columnas que no presentan datos faltantes son “ID” , “X” e “Y”, las demás todas presentan falta de datos.

A esta tabla se le agregaron los registros del archivo “Clientes\_Dic2020.csv” sumando así un total de 2672 registros.

**TiposDeGasto.csv**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Nombre | Tipo de dato | Formato | Descripción | Ejemplo |
| idTipoDeGato | Int | N | Id del tipo de gasto que se tuvo q realizar | 4 |
| Descripción | Object | -------------- | Descripción del tipo de gasto que se realizo | Impuestos |
| Monto\_aproximado | Int | NNNNNN | Monto aproximado a pagar por el tipo de gasto | 300 |

**¿Hay datos faltantes?:** Cantidad de registros 4

* IdTipoGasto 4
* Descripcion 4
* Monto\_Aproximado 4

**ETL**

En la tabla de ventas se concatenaron los registros de diciembre 2020 para así tener una sola tabla con todos los registros.

En la tabla clientes se hizo lo mismo.

Se creo una tabla llamada productos con los datos de valor de cada producto y su respectivo ID.

De la tabla Localidad se extrajeron los datos de Provincias, Municipios y Departamentos para crear tablas de estas y así, dejar únicamente con los ID a la tabla localidad.

En la tabla de Proveedores se normalizaron los nombres de las columnas y todos los datos se pusieron en minúsculas.

Para finalizar el proceso todas las tablas se subieron a una base de datos en MySQLWorkbench.