**Documentación del proceso ETL pedidos globales**

# Descripción de archivos

## Dev.yaml / PRD.yaml

Este archivo que almacena las configuraciones por país, así como la ruta en donde se encuentra el archivo fuente global.

A continuación, se describen cada uno de las claves de este archivo:

|  |  |
| --- | --- |
| Clave | Descripción |
| globalPath | Identifica el path en donde se encuentra el archivo con la información global |
| [Nombre País] | Bajo el nombre de cada país se almacenan todas sus propias configuraciones. |
| processedPath | Almacena el path en el que se guardará la información procesada. |
| Abreviatura | Abreviatura de cada país |
| unidad\_reducida | Guarda cual es la unidad reducida que se maneja en cada país. (Esta configuración se podría hacer, también, a nivel de producto y o marca) |
| Factor\_conversion\_ST | Guarda el factor para convertir las unidades a las unidades reducidas utilizadas en cada país |
| moneda | Guarda el tipo de moneda a utilizar en cada país. |

## Main.ipynb

Archivo desarrollado en Python en donde se hace el llamado a las funciones para la carga, procesamiento y partición de la información.

En este archivo se ingresa el rango de fechas que se desea procesar, así como también se define el país y el ambiente (dev/prd)

## DataUtilities.py

Como una de las buenas prácticas en el desarrollo se crea esta librería con todas las funciones utilizadas para el proceso de ETL, con esto se busca ordenar y reutilizar el código.

A continuación, se describe cada una de las funciones:

|  |  |
| --- | --- |
| Función | Descripción |
| load\_config | Esta función recibe como parámentro el país y el ambiente en el que se desea trabajar, con base a estos parámetros lee el archivo de configuración e inicializa las variables globales con las configuraciones encontradas |
| date\_validate | Esta función solicita al usuario el ingreso del rango de fechas al mismo tiempo que valida que este rango esté correcto. De estar correcto el rango ingresado inicializa las variables globales de fecha inicio y fecha fin |
| data\_validation | Esta función revisa los registros de cada columna en busca de datos nulos o vacíos e informa al usuario de lo encontrado. Acá se puede agregar la opción de eliminar los datos nulos o vacíos encontrados |
| data\_process | Esta función carga la información del archivo global con base al ambiente (dev/prd), país y rango de fechas seleccionado. También excluye los registros con no coinciden con el tipo de entrega definido y realiza la conversión a unidades reducidas, agrega una columna por cada tipo de entrega valido encontrado en el set de datos y por último guarda un nuevo archivo csv por cada fecha proceso única encontrada. |