**Tarea: Entendimiento de los datos**

**Introducción**

**Objetivo**

Realizar el proceso de entendimiento de datos de WWImporters a través de perfilamiento general, análisis de reglas de negocio, estadísticos y análisis de calidad de datos.

**¿Para qué?**

Practicar lo aprendido en el tutorial de entendimiento de los datos.

**¿Qué necesita?**

1. Notebook para trabajar: puede usar la seccion 5 “Tarea” al final del notebook del tutorial para realizar esta actividad
2. Acceso a la base de datos WWImportersTransactional

**Caso WWImporters**

Wide World Importers[^1] es un importador y distribuidor de productos de primera categoría, que funciona desde el área de la bahía de San Francisco. Como mayorista, los clientes de WWI son principalmente empresas que revenden a individuos. WWI vende a los clientes de venta al por menor en el Estados Unidos como tiendas especializadas, supermercados, almacenes informáticos, tiendas de atracción turística y otras personas. WWI también vende a otros mayoristas a través de una red de agentes que promocionan los productos en el nombre de WWI. A pesar de que todos los clientes de WWI se basan actualmente en el Estados Unidos, la empresa está intentando impulsar la expansión en otros países.

WWI compra mercancías de proveedores, incluidos los fabricantes de la novedad y el juguete, y otros mayoristas. Almacenan las mercancías en el almacén de WWI y reordenan a los proveedores según sea necesario para satisfacer los pedidos de los clientes. También compran grandes volúmenes de materiales de embalaje y los venden en cantidades más pequeñas, por comodidad para los clientes.

Los objetivos de Wide World Importers son:

1. Ofrecer a los clientes productos que deseen cuándo y dónde lo deseen
2. Dar un servicio de calidad que maximice las ganancias por ventas
3. Optimizar el manejo de inventario de los productos

Wide World Importers nos contrata como consultores para hacer un proyecto de analítica 1.0, en el cual quiere entender el comportamiento de los clientes o vendedores a partir de órdenes de venta con el fin de monitorear el logro de los objetivos estratégicos anteriores. Como resultados de la consultoría se tiene la documentación de los requerimientos analíticos en el siguiente formato:

| **Tema analítico** | **Análisis requeridos o inferidos** | **Categoría del análisis (\*)** | **Procesos de negocio** | **Fuentes de datos** |
| --- | --- | --- | --- | --- |
| **Comportamiento de clientes a partir de órdenes** **de compra** | Visualizar la cantidad de productos vendidos y el valor total recaudado en un **rango de fechas** por cliente, vendedor, ciudad, producto | Tablero de control | Ventas | Ordenes, detalles de órdenes, clientes, productos, ciudades, vendedores |

Adicionalmente, WWImporters desea realizar análisis sobre los datos de movimientos de inventarios y esa será la temática principal para el desarrollo de las tareas del curso.

**Enunciado**

Ahora que ya aprendió a usar PySpark para exploración y calidad de datos, lo invitamos a que realice el proceso de **entendimiento de datos** de la tabla movimientosCopia que se encuentra en la base de datos WWImportersTransactional, que registra el movimiento en inventario de un producto. La empresa no compartió información de las tablas de clientes ni de proveedores para esta actividad.

En esta actividad le recomendamos revisar el diccionario de datos suministrado y en particular la información relacionada con la tabla **movimientosCopia**. Si tienen alguna duda relacionada con esta tabla deben incluirla en las preguntas o aclaraciones que le entreguen al negocio. En relación a los datos Wide World Importers menciona:

* La cantidad máxima de productos movidos es 50 millones por transacción
* Existen 236.668 movimientos de producto realizados desde 2013
* Los movimientos están relacionados con 228.265 clientes
* El formato de fechas manejado es YYYY-MM-DD HH:MM:SS
* Actualmente tenemos 13 proveedores

Recuerde:

* Revisar la infografía de entendimiento de datos y de calidad de datos.
* Ser ordenado en la entrega
* Evitar tener imágenes o tablas sin análisis
* Incluir el significado de una fila (no es copia del diccionario de datos)
* Terminar la actividad con una conclusión orientada a las personas de negocio
* Incluir la lista de preguntas o aclaraciones para la organización, problemas de calidad y alternativas para la limpieza.
* La entrega corresponde a la URL de tu Wiki, asociada al notebook donde incluyes únicamente los elementos solicitados en esta tarea.

En número máximo de unidades en una transacción es y el minimo es