**PRUEBA TÉCNICA Data Engineer**

# Python: modelado y ETL

El responsable de compras de una empresa nos pide un sistema de BI para la toma de decisiones dentro de su departamento, las cuales se dividen en 2 partes diferenciadas: la parte financiera y la parte de análisis de los proveedores.

En los que se refiere a la parte financiera, el responsable de compras quieres saber el importe total de las compras que realiza diariamente, siempre valoradas en euros, distribuidas por cada una de las secciones de la empresa. Es importante tener en cuenta que las compras se realizan a proveedores de distintos países, y que cada proveedor tiene asociada una moneda de referencia en la que se recogen sus facturas. Es por ello por lo que habrá que prestar atención al tipo de cambio, el cual fluctúa diariamente. La empresa dispone de un histórico de cambios diarios por divisa que le sirve para realizar las valoraciones correctamente. La fecha de referencia en este caso es la fecha de factura.

Sobre la segunda parte, la relativa a los proveedores, el responsable quiere poder obtener un ranking de proveedores según el importe total comprado en un periodo de tiempo determinado y según el número de referencias o productos distintos comprados. La ides es poder tener una medida de la importancia del proveedor: un proveedor al que se le compran muchos productos distintos, aunque no sean de un valor muy alto, es un proveedor importante.

Para medir la calidad del proveedor, en una primera fase, el responsable nos pide que midamos el tiempo de entrega o lead time real de las compras realizadas y que lo comparemos con el lead time teórico. Para calcular el lead time real disponemos en el sistema de la fecha de pedido y la fecha de recepción, de modo que se calcula por diferencia entre ambas fechas, expresada en días. Como lead time teórico se establecen 10 días para los proveedores españoles, 20 días para el resto de proveedores europeos (los que operan en EUR y no son España) y 45 para los no europeos (el resto).

Esta información se quiere consumir mediante una media anual (considerando como referencia la fecha de emisión del pedido), pues se entiende que en periodos más cortos puede haber fluctuaciones que distorsionen la medida. Además, es importante medir dicha calidad por producto-proveedor, ya que un proveedor puede ser bueno sirviendo un producto, pero no tanto sirviendo otro.

Según conversación con el equipo de IT de la compañía, toda la información se encuentra almacenada en un sistema de gestión de compras centralizado, del cual se extraen una serie de ficheros que constituirán nuestras fuentes de datos. Dichos ficheros son los siguientes (que se facilitan):

- Invoice_header: datos generales de las compras, con el proveedor al que se le hace la compra y la fecha de pedido, de recepción y de factura.
- Invoice_products: detalle de los productos pedidos en cada factura, su cantidad, su precio y la sección a la que van destinados.
- Products: maestro con todos los productos disponibles
- Suppliers: maestro con los datos de los proveedores
- Daily_currencies: histórico de tipos de cambio diario de cada una de las monedas.

**Se pide:**

- Diseñar un datamart/s dimensional/es apropiado/s para satisfacer el conjunto de requisitos del sistema. Para ello, será necesario tener en cuenta el proceso/s de negocio a modelar, la granularidad, las dimensiones, los hechos y las medidas implicadas y las relaciones. Justifica tus decisiones y si se considera necesario, representar dicho modelo mediante un diagrama.
- Implementar el diseño anterior mediante un script en python, usando la librería pandas para el manejo de los dataframes. Se debe partir del conjunto de ficheros facilitados junto con el enunciado, cargarlos en el script, realiza las tareas de trasformación que consideres necesarias (comprobación de tipos de datos, limpieza, columnas calculadas etc.) y finalmente escribir cada una de las tablas del datamart en su fichero correspondiente. Justifica y comenta adecuadamente las operaciones que realizas.

**Opcionalmente,** se desea incorporar al diseño y a la implementación la posibilidad de analizar la cifra de compras mensualizada en comparación con el presupuesto de compras de cada una de las secciones de la compañía. Dicho presupuesto se guarda en un fichero Excel externo, llamado purchase_budget.xls (que se facilita). Los análisis de la parte obligatoria no deben verse afectados por esta nueva funcionalidad.