

RETO-TEST PYTHON FOR DATA: ENUNCIADOS

La empresa ficticia "Global Importaciones" ha decidido implementar un sistema de gestión de pedidos utilizando la base de datos Northwind como referencia. La empresa se dedica a la importación y distribución de productos a nivel internacional. Como analista de datos para la empresa internacional "Global Importaciones", te enfrentas al desafío de extraer información clave de la base de datos Northwind para optimizar la gestión de pedidos. La dirección ha solicitado insights específicos para tomar decisiones informadas.

Para poder resolver este ejercicio primero, es necesario que instales la BBDD de 'Northwind' en Postgres/DBeaver, a continuación deberás generar una conexión entre Python y Postgres/DBeaver para poder realizar los siguientes ejercicios.

Ejercicio 1. Familiarizarse con la Base de Datos:

Descripción del ejercicio:

El primer paso como analista de datos es conocer la estructura y la información básica de nuestra BBDD, Crea el esquema de northwind y observa con atención las tablas que tiene y cómo están relacionadas entre sí.

Ejercicio 2. Primeras consultas

Descripción del ejercicio:

Ahora vamos a sacar información básica de nuestra BBDD, para ello tienes que generar las Queries necesarias para responder a las siguientes preguntas:

- ¿Cuántos empleados tenemos contratados en 'Global Importaciones'? Indica su id, nombre, apellido, ciudad y país.
- ¿Qué productos tenemos? Indica el id del producto, id del proveedor, nombre del producto, precio por unidad, unidades en stock, unidades pedidas al

proveedor y productos descontinuados.

- ¿Tenemos productos descontinuados?. Indica el nombre del producto, y cantidad que nos queda en stock. (El atributo Discontinued es un booleano: si es igual a 1 el producto ha sido descontinuado.)
- ¿Qué proveedores tenemos? Indica el id de la compañía, nombre de la compañía, ciudad y país.
- ¿Qué pedidos hemos tenido? Indica el número de pedido, id del cliente, id del transportista, día del pedido, día requerido de llegada y día de llegada real.
- ¿Cuántos pedidos hemos tenido?
- ¿Cuántos clientes tenemos? Indica el id del cliente, nombre de la compañía, ciudad y país.
- ¿Con qué empresas de transporte trabajamos? Indica su id del transportista y el nombre de la compañía.
- ¿Cómo son las relaciones de reporte de resultados entre los empleados?

Ejercicio 3. Análisis de la empresa

Vamos a crear un par de DataFrames uno con la información de las tablas de pedidos y clientes y otro con la información de productos, proveedores y detalles de los pedidos para poder hacer un estudio de la evolución de nuestra empresa y qué cosas podemos mejorar de esta.

- Haz un estudio de la evolución de los pedidos realizados a lo largo del tiempo. Para ello primero realiza la query necesaria para obtener los meses, años y pedidos durante cada mes. A continuación crea una línea temporal para ver dicha evolución
- Investiga cuáles son los países donde tenemos más ventas (País origen de la compañía). No es necesario realizar una query para obtener el DataFrame. A raíz de estos datos genera una columna con el continente a partir del siguiente diccionario:

```
(continentes = {'Europe': ['Austria', 'Belgium', 'Denmark', 'Finland', 'France',  
'Germany', 'Ireland', 'Italy', 'Norway', 'Poland', 'Portugal', 'Spain', 'Sweden',  
'Switzerland', 'UK'],
```

```
'America': ['Argentina', 'Brazil', 'Canada', 'Mexico', 'USA', 'Venezuela'] })  
y estudia la distribución de pedidos por continente  
. Realiza la visualización que creas más conveniente .
```

- Sabemos que algunos pedidos han llegado con retraso, además hay pedidos que no ha sido registrada su llegada. Investiga si la compañía de transporte está relacionada con ello o no. Realiza un boxplot para ver la diferencia de rango intercuartílico.
- Hay bastante diferencia entre el precio pagado en cada pedido. Averigua la distribución media del precio del pedido por país de procedencia del cliente. Realiza la visualización que creas más conveniente para sacar conclusiones
- Investiga si existen clientes que no hayan pedido nunca. ¿Qué porcentaje de clientes no tienen pedidos registrados?
- Estudia los productos más demandados e investiga cuáles corre prisa hacer reestock (Los que quedan 20 o menos y no hay unidades pedidas). Realiza la visualización que creas más conveniente para sacar conclusiones.

Ejercicio 4. Queries Avanzadas

Nuestro jefe acaba de venir y nos ha hecho una serie de peticiones sobre la base de datos que tenemos que poder contestar.

- Quiere saber cuándo fue la última vez que se pidió un producto de cada categoría.
- Necesita saber si existe algún producto que nunca se haya vendido por su precio original.
- Quiere tener toda la información necesaria para poder identificar un tipo de producto. En concreto, tienen especial interés por los productos con categoría "Confections". Devuelve el ID del producto, el nombre del producto y su ID de categoría.
- Quiere saber si existe algún proveedor del que pueda prescindir ya que todos los productos que tiene se encuentran discontinuados.
- Extraer los clientes que compraron mas de 30 articulos "Chai" en un único pedido

- Indica los clientes cuya suma total de carga en los pedidos sea mayor de 1000
- Desde recursos humanos nos piden seleccionar los nombres de las ciudades con 5 o más empleadas de cara a estudiar la apertura de nuevas oficinas.