

Prueba - Python para el análisis de datos

En esta prueba validaremos nuestros conocimientos sobre Python. Para lograrlo, necesitarás aplicar lo aprendido en las unidades anteriores.

Lee todo el documento antes de comenzar el desarrollo **individual**, para asegurarte de tener el máximo de puntaje y enfocar bien los esfuerzos.

Prerrequisitos

1. Para realizar lo solicitado debes crear la base de datos **classicmodels** en tu motor PostgreSQL. Para esto, abre una ventana de terminal y ejecuta la siguiente instrucción:

```
psql -h localhost -p 5432 -U postgres -c "CREATE DATABASE classicmodels;"
```

2. Una vez creada la base de datos, debes importar el archivo **classicmodels.sql** a esta base de datos.

```
psql -h localhost -p 5432 -U postgres -d classicmodels -f classicmodels.sql
```

Descripción

El área comercial de una empresa pide realizar un cierre de año de las ventas, tanto para revisar si las metas fueron cumplidas, como para poder planificar el siguiente año. Para ello, considerarán los datos del dataset **classicmodels.sql** para responder algunas preguntas, realizando las siguientes tareas.

1. Genera una función llamada `leer_tabla(tabla, engine)` y utilízala para leer tablas completas desde la base de datos en dataframes independientes. Utilizando esta función, importa las siguientes tablas:
 - order
 - orderdetails
 - customers
 - products
 - employees
2. Realiza el cruce entre los DataFrames, asegurándote de utilizar correctamente el parámetro **validate** para asegurar la integridad referencial.

3. Agrega las siguientes columnas, considerando su nombre y la fórmula asociada
 - **venta:** `quantityOrdered*priceEach`
 - **costo:** `quantityOrdered*buyPrice`
 - **ganancia:** considerando las columnas anteriores
4. ¿Cuál fue el total de ventas por línea de productos? Incluye una fila de totales.
5. ¿Cuántos clientes distintos hicieron compras?
6. ¿Existen clientes que aún no han hecho ninguna compra? ¿Cuántos son?
7. Se solicita la creación de dos reportes, que respondan las preguntas dadas
 - ¿Cuáles fueron los 10 clientes que reportan mayores ventas brutas en dinero durante el año 2005? Genera un DataFrame y guárdalo en una tabla de Postgre llamada `top_10_clientes_2005`, en la que se especifique el nombre del cliente y su correspondiente venta, costo y ganancia.
 - ¿Cuál fue el top 10 de artículos más vendidos durante el año 2005? Genera un DataFrame y guárdalo en una tabla de Postgre llamada `top_10_productos_2005`, en la que se especifique el nombre del producto y su correspondiente venta, costo y ganancia.

Para este punto debes aplicar el principio DRY, por lo que se deben utilizar funciones para realizar el filtrado por fechas, generar tablas pivote y escribir tabla en Postgre. Las funciones deben estar en un archivo separado llamado **funciones.py** y ser importadas al Jupyter Notebook. En este archivo se debe incluir:

- Una función que permita filtrar un DataFrame por fechas, indicando dataframe, columna para filtrar, fecha inicio y fecha fin. La función debe retornar un DataFrame.
- Una función que permita generar reportes dependiendo de parámetros de entrada como dataframe, filas, columnas, valores y medida (funcion_agrupadora). Utilizar `fill_value = 0`. Esta función debe retornar un DataFrame pivotado.
- Una función que permita escribir en la base de datos a través del guardado de un DataFrame dependiendo de parámetros de entrada como DataFrame, nombre de la tabla, engine y comportamiento en caso de que exista la tabla (`if_exists`).

Requerimientos

Dentro del archivo de Jupyter Notebook debes ir ejecutando las siguientes acciones y explicar lo que estás haciendo:

1. Genera y aplica funciones de Python para realizar consultas a bases de datos SQL. **(2 Punto)**
2. Manipular DataFrames para realizar cruces de tablas, generar columnas y realizar cálculos utilizando comandos de Python. **(3 Puntos)**
3. Genera reportes a partir de DataFrames, aplicando filtros y exportando tablas por medio de comandos de Python. **(3 Puntos)**
4. Crear y aplicar funciones en Python aplicando el principio DRY y buenas prácticas. **(2 Puntos)**



¡Mucho éxito!

Consideraciones y recomendaciones

- Debes entregar tu trabajo en un archivo de Jupyter Notebook, con todo el código y las explicaciones respectivas para desarrollar tu trabajo.