Data Engineering and Advanced Analytics Challenge

Evaluation criteria:

Your response to the challenge will be evaluated according to the following criteria:

- Python programming skills
- SQL knowledge
- Data handling methods

Challenge:

- You are kicking off a project with one of France's top retailers.
- You have already found out during the kick-off discussion that they have little-to-no data hygiene.
- You have therefore asked them to provide you with an initial data sample while their teams are working on the full data extract.
- They sent you their transactional data of the past 3 months split into 3 csv files, as well as the product information for this time range.
- In addition, the client provided you with a data model shown in the appendix section below.

Tasks:

a) Consolidate and load the data into a database

b) Write & run SQL queries in the database to retrieve the following information from the dataset:
   - The total margin (*Marge_Nette_Magasin*) generated by signature products (*SIGNATURE PRODUCT?*) over the last 2 months
   - Revenue (*CA_Net_TTC*) split per day
   - Top 10 products (*Modele_Couleur_Ref*) in terms of units sold (*Quantite_Vendue*)
   - Number of transactions (*Numero_Transaction*) per store (*Point_de_Vente*)

Stack:

- Python
- SQL
- SQLite database

Implementation Notes:

- All data handling code should be written in Python
- All data should be loaded into an SQL database (via SQL statements). You can use SQLite or a containerized SQL server as your database target.
- All database queries should be in native SQL and not via pandas or any other python libraries.