S30 ETL Assignment

Goal

The purpose of this assignment is to showcase your skills in analyzing business needs and in utilizing Python and SQL to provide solutions.

Assignment

Scenario:

Company XYZ held a promo sale for their signature items named: x,y,z. Sales are at an all-time high, but they want to create a marketing strategy to target age groups of people by looking at total quantities purchased.

They then created a database with these business rules:
- A sales receipt can have multiple items in an order.
- For every order, the clerk records all quantities for all items, including items not bought (which they denote with quantity=NULL).
- Each customer can do multiple sales transactions, and has his/her age stored in a database.

Refer to the image below for the table structures and relationships.
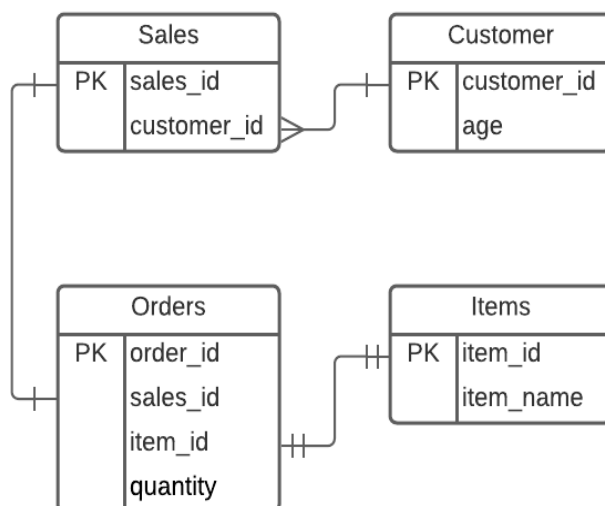
Objectives

Create a Python script that can:

1. connect to the SQLite3 database provided
2. extract the total quantities of each item bought per customer aged 18-35.
   - For each customer, get the sum of each item
   - Items with no purchase (total quantity=0) should be omitted from the final list
   - No decimal points allowed (The company doesn't sell half of an item ;))

   Challenge: Provide 2 solutions, one using purely SQL, the other using Pandas
3. store the query to a CSV file, delimiter should be the semicolon character (';')

Test case:
 Customer 1 bought Item X on multiple occasions, totaling 10 for Item X only
 Customer 2 bought one of each item only once, totaling 1 each Item
 Customer 3 bought Item Z on two occasions, totaling 2 for Item Z only
 Then the output file should look like the example below:

```
Customer;Age;Item;Quantity
1;21;x;10
2;23;x;1
2;23;y;1
2;23;z;1
3;35;z;2
```
(Note: Actual values will vary)

Delivery

Provide a link to a public repository where we can download the code containing all files.