Analytics Engineer Technical Test (Take Home Challenge)

The goal of this technical test is to showcase an end to end ELT pipeline from a data source to data warehouse using SQL and DBT. The final data models should make it possible to answer the following questions:

- Top 10 stores per transacted amount
- Top 10 products sold
- Average transacted amount per store typology and country
- Percentage of transactions per device type
- Average time for a store to perform its 5 first transactions

Deliverables:

Please share the source code, data model design, SQLs to answer the above questions via public git repository including a README file explaining your assumptions, design and solution implementation details.

Usage of pandas libraries and Python notebooks are not permitted. For any data manipulations please use SQL.

Assumptions:

- The sample tables refer to a ficticious company that provides payment devices for retail stores.
- Each of the customers have one or multiple stores, which are physical locations where their business happens, those stores are classified by different typology.
- The company provides them with different types of devices (1 to 5), those devices are hardware needed to perform any kind of transactions, each of them are tied up to a specific store.
- A transaction is a payment made using the provided devices, currently the devices only handle payment made by card and in euros. Those transactions are made to pay for products sold inside the store, each product got a name and a SKU (stock keeping unit) which is unique.
- Everyday, our customer's customer walk into their favorite stores to buy products and pay for them using the devices that the company provides.

The company wants to target customers that will use their devices efficiently and benefit the most of their product. That's they need to answer the above questions to know which stores,

products and devices are the most efficient and also to know how long it takes for a store to adopt their devices.

In order to solve this problem, we have provided the following three datasets in csv format

- Stores
- Devices
- Transactions

The dataset supplied with this test contains only sample data. Your design and implementation should scale for larger volumes of data (millions to billions records).

Ideally, you should set up a fully functional dbt project that can connect to a cloud warehouse of your choice where the models can be run (Snowflake would be the preferred option – you can set up a trial account for this purpose).