**Allica Bank - Data Engineer Home Task**

Task:

ABC Musical Instruments LTD is a UK musical instrument distributor. They sell and deliver orders to clients all over UK. Up until recently, their order processes have been managed on Excel. However, the management team want to move to a more robust system which would allow them to generate sale reports, manage their customers and their inventory better.

Your job is to build data model and engineer data pipelines for their new analytical data warehouse.

Including in this home task is a CSV file with sample data from some of the existing orders. Your task is to model a few tables to store this data and write pipelines to ingest data into these tables. The data might need to be transformed before inserted into the database.

Requirements:

* The data models and pipelines need to be written exclusively in Python. Some SQL codes are allowed but they need to be executed via Python.
* You can use existing libraries like sqlalchemy, pandas, etc. If you use existing libraries, please make sure to include a list of libraries used in requirements.txt.
* The data warehouse will be stored on SQLite database. The SQLite database file can be part of your submission. However, you need to make sure your codes can recreate this database using the data from the CSV file.
* Your codes must be clean, well-structured and easy to use. Please also include a Readme file in the submission with instructions on how to run the codes.

Assessment criteria:

* Data modelling:
  + The split between facts and dimension are done properly.
  + Correct data types, primary keys and foreign keys.
* Coding:
  + Your codes will be run against the sample file to test if the database is created and data is correctly ingested. Your codes will then be run against a few new sample files to see how it handles new data. One of the new files will contain errors to test how your codes handle errors.
  + We might delete the database file and rerun your codes to test resiliency of your pipelines. A good pipeline should be able to recreate the database from raw files.
  + Your coding styles and project structure will also be assessed. We want to see consistent coding standards, reusable classes and functions, codes written with usability in mind, simple and easy to follow instructions on how to execute pipelines.

Important notes:

* All data files in this exercise will be CSV to keep things simple.
* Don’t over complicate the task! You will only need between 3 to 5 tables.

Bonuses:

* Slowly Changing Dimension (SCD) implementation
* Unit tests