[Task] Integrated Data Processing Solution

Project Structure

- **project_root** *(entry point folder)*
    - research_notebooks (*folder storing all research notebooks*)
        - `research_notebook_b.ipynb`
        - `research_notebook_a.ipynb`
    - **shared_functions** (*folder storing all user defined functions*)
        - `function_a.py`
        - `function_b.py`
    - **data_processing** *(folder which stores files related to data processing)*
        - docker-compose.yml (*describes below images, but not limited to them*)

          Kafka, Zookeeper, Processing Module
        - `processing_module`

          *Custom Module which is doing micro-batch processing and storing data back to the Kafka. Module is subscribed to Kafka topic and waiting to accumulate micro-batches of some window frames (for example 5 minutes). After window data is accumulated, it runs shared user defined function (**function_a.py** for example) against the accumulated data and send the result to the another kafka topic.*

Project consist of 3 main components, each of them has it's own folder in the project structure:

- Shared Processing Functions
- Data Processing Module

- Research Notebooks

Both **Research Notebooks** and **Data Processing Module** should rely on the shared processing functions as a source of truth for data processing.

Task

To implement solution to be able to process micro-batches in the processing module docker-compose environment with a help of user defined functions which are shared between research notebooks and data processing module.

- As a User, I should be able to run notebooks locally with imported functions shared functions
- Shared functions should be covered with unit tests and way of running tests documented in Readme
- Data Processing module can be implemented with any batch processing engine written in python, like **Faust** or **Nuclio** or any other.

The execution code for processing functions is up to you, you can take pieces of code from you past projects which were based on micro batching. You should be able to elaborate on your solution and describe algorithmic part of processing.