Along with this document, you will receive a dataset with 4 tables:

* course
* course\_attendance
* enrollment
* schedule

Your task:

1. Using PySpark, extract the dataset into a dataframe, clean, and transform the dataset.
2. Using PySpark SQL, analyze the dataset and provide a minimum of 3 graphs.
3. Load your analysis result into a file (csv, parquet, json).
4. Please give documentation (txt file) about your script, how to run, and your dataset analysis.
5. **Bonus Point** if you manage to dockerize your project (make your script run with docker), and implement a Spark master and worker node.

What to submit:

* Your python file (.py) that contains your answer.
* A folder that contains the exported graph and the data.
* A txt file that contains the documentation of your answer, and your analysis explanation.
* **Optional:** a docker file, a Makefile, and/or docker-compose file

Gabungkan seluruh file dalam bentuk zip/rar