## Build and run docker container running spark and jupyter

1. Modify the docker image to make sure when container is built and runs spark and jupyter when start.
    a. Add below lines to the end of given docker image:
    ```
    # Start Spark Master and Jupyter Notebook when the container runs
    CMD ["/bin/bash", "-c", "start-master.sh && start-worker.sh spark://localhost:7077 && jupyter notebook --ip=0.0.0.0 --allow-root --NotebookApp.token=""]
    ```

2. Change directory to where your docker image is located :
    ```
    cd /Users/main/Desktop/docker
    ```
3. Build a image in docker:
    a. Make sure using dot (.) this build docker container based on your docker image file
    ```
    docker build -t spark-python .
    ```
4. Create running container with exposing port:
    ```
    docker run -d --name spark-container -p 8888:8888 -p 4040:4040 -p 7077:7077 -v $(pwd)/workspace:/workspace spark-python
    ```

5. Now check if jupter notebook or spark are running:
    a. Jupyter: http://127.0.0.1:8888
    b. Create new python notebook and run:
    ```
    from pyspark.sql import SparkSession

    spark = SparkSession.builder.appName("test").getOrCreate()
    spark.range(10).show()
    ```

    c. Spark master: http://127.0.0.1:8080 (will not sure if job not running)
    d. Spark jobs: http://127.0.0.1:4040 (will not sure if job not running)
    e. Check spark if something wrong:
    ```
    docker exec -it spark-container jps
    ```
6. Attach VS code to container:
    a. Open VS Code.
    b. Go to Extensions (Cmd + Shift + X).
    c. Install the following:
        i. Docker (by Microsoft)
        ii. Remote - Containers (by Microsoft)
7. Open VS Code.
    a. Press Cmd + Shift + P (Command Palette) and search:
        i. Remote-Containers: Attach to Running Container
        ii. Select your container (spark-container).
    b. Once inside the container:
        i. Open the workspace/ directory inside the container:
        ii. Click File > Open Folder
        iii. Select /workspace
    c. Once inside workspace
        i. Create test.py file and add print('hello word'). Save
        ii. In vs code open terminal use: ctri + `
            1. Run: python3 test.py
8. Extra: change port to Jupyter: http://127.0.0.1:9999

    ```
    docker stop spark-container
    docker rm spark-container
    docker run -d --name spark-container -p 9999:8888 -p 4040:4040 -p 7077:7077 -v $(pwd)/workspace:/workspace spark-python
    ```