



- 1- creating new folder and moving into mkdir/cd
- 2- initialize git : git init
- 3- code crawler.py/dockerfile/requirements.txt (I copied my files from first folder as I had cache issues to rebuild and rerun docker which were reading an older version of my crawler.py)
- 4- git add . to add all folder to the tracking
- 5 - git commit -m 'first commit' to save changes to master brnch
- 6 - building image from scratch again with correct files, using --no-cache option to prevent issues met previously

```
(base) fitec@fitec-HP-ProBook-450-G5:~/Documents/eval_webscrapping_docker$ cd ..
(base) fitec@fitec-HP-ProBook-450-G5:~/Documents$ mkdir france_temperatures
(base) fitec@fitec-HP-ProBook-450-G5:~/Documents$ cd france_temperatures/
(base) fitec@fitec-HP-ProBook-450-G5:~/Documents/france_temperatures$ git init
Dépôt Git vide initialisé dans /home/fitec/Documents/france_temperatures/.git/
(base) fitec@fitec-HP-ProBook-450-G5:~/Documents/france_temperatures$ git code crawler.py
git : 'code' n'est pas une commande git. Voir 'git --help'.
```

La commande la plus ressemblante est
clone

```
(base) fitec@fitec-HP-ProBook-450-G5:~/Documents/france_temperatures$ git status
Sur la branche master
```

Aucun commit

Fichiers non suivis:

(utilisez "git add <fichier>..." pour inclure dans ce qui sera validé)

```
city.list.json
crawler.py
dockerfile
requirements.txt
```

aucune modification ajoutée à la validation mais des fichiers non suivis sont présents (utilisez "git add" pour les suivre)

```
(base) fitec@fitec-HP-ProBook-450-G5:~/Documents/france_temperatures$ git add .
(base) fitec@fitec-HP-ProBook-450-G5:~/Documents/france_temperatures$ git commit -m 'first commit with replicated file from first attempt'
```

[master (commit racine) 867ae0f] first commit with replicated file from first attempt

4 files changed, 1886411 insertions(+)

```
create mode 100644 city.list.json
create mode 100644 crawler.py
create mode 100644 dockerfile
create mode 100644 requirements.txt
```

```
(base) fitec@fitec-HP-ProBook-450-G5:~/Documents/france_temperatures$ sudo docker build --no-cache --tag mael/eval_mokhtar .
```

Sending build context to Docker daemon 34.83MB

Step 1/7 : FROM python:3.6-slim

--> 1c43175e9000

Step 2/7 : LABEL maintainer = "mael.falvet@gmail.com"

--> Running in 8e702525c2cd

Removing intermediate container 8e702525c2cd

--> c1f73f6104b5

Step 3/7 : WORKDIR /app/

--> Running in a3177ec1eaf8

Removing intermediate container a3177ec1eaf8

--> 8d7a47a60ab7

Step 4/7 : COPY crawler.py requirements.txt city.list.json /app/

--> 8b48bdb5dffe

Step 5/7 : RUN pip install -r requirements.txt

--> Running in 269c062ad194

Collecting pandas

Downloading pandas-1.1.5-cp36-cp36m-manylinux1_x86_64.whl (9.5 MB)

Collecting termcolor

Downloading termcolor-1.1.0.tar.gz (3.9 kB)



- 7- creating docker volume to stock csv
- 8- running docker image tied to created volume
- 9 - running calls for 1000 french cities due to time constraints instead of 19k, should be proof of work

```
Successfully built rec9c93f0bbcc
Successfully tagged mael/eval_mokhtar:latest
(base) fitec@fitec-HP-ProBook-450-G5:~/Documents/france_temperatures$ sudo volume create api_weather_data
sudo: volume : commande introuvable
(base) fitec@fitec-HP-ProBook-450-G5:~/Documents/france_temperatures$ sudo docker volume create api_weather_data
api_weather_data
(base) fitec@fitec-HP-ProBook-450-G5:~/Documents/france_temperatures$ sudo docker run -it -v api_weather_data:/app mael/eval_mokhtar
generation de l url
Invocation du API afin de recuperer les données
formatage des donnée
{'city': 'Forville', 'country': 'FR', 'temp': 0.96, 'temp_max': 3.43, 'temp_min': -7.41}
Affichage de données
-----
Current weather in: Forville, FR:
0.96 °C
Max: 3.43, Min: -7.41
-----
enregistrement des données à dans un fichier CSV
generation de l url
Invocation du API afin de recuperer les données
formatage des donnée
{'city': 'La Portanière', 'country': 'FR', 'temp': 6.09, 'temp_max': 8.79, 'temp_min': 3.38}
Affichage de données
-----
Current weather in: La Portanière, FR:
6.09 °C
Max: 8.79, Min: 3.38
-----
enregistrement des données à dans un fichier CSV
generation de l url
Invocation du API afin de recuperer les données
formatage des donnée
{'city': 'Wittenheim', 'country': 'FR', 'temp': -0.05, 'temp_max': 1.44, 'temp_min': -1.06}
Affichage de données
-----
Current weather in: Wittenheim, FR:
-0.05 °C
Max: 1.44, Min: -1.06
-----
enregistrement des données à dans un fichier CSV
generation de l url
Invocation du API afin de recuperer les données
formatage des donnée
{'city': 'Wattrelos', 'country': 'FR', 'temp': 6.39, 'temp_max': 7.1, 'temp_min': 5.5}
Affichage de données
-----
Current weather in: Wattrelos, FR:
6.39 °C
Max: 7.1, Min: 5.5
-----
enregistrement des données à dans un fichier CSV
generation de l url
Invocation du API afin de recuperer les données
```



7- erreur qui gâche tout sur un bout de code rajouté en fin de script pour nettoyer le csv qui laisse apparaître le header entre chaque requete :

```
france_df = pd.read_csv('weatherOpenMap.csv')
france_df =
france_df[france_df['temp max'] != 'temp max']
france_df.to_csv('clean weather France.csv')
```

8- localiser le csv et copier via la commande

```
sudo cp /var/lib/docker/volumes/api_weather_data/_data/mael_weatherOpenmap.csv
/home/fitec/Documents/france_temperatures/mael_weatherOpenmap2.csv
```

```
-----
Current weather in: Sarreinsming, FR:
3.32 °C
Max: 5.53, Min: 1.36
-----
Enregistrement des données à dans un fichier CSV
Traceback (most recent call last):
  File "./crawler.py", line 171, in <module>
    france_df = pd.read_csv('weatherOpenMap.csv')
  File "/usr/local/lib/python3.6/site-packages/pandas/io/parsers.py", line 688, in read_csv
    return _read(filepath_or_buffer, kwds)
  File "/usr/local/lib/python3.6/site-packages/pandas/io/parsers.py", line 454, in _read
    parser = TextFileReader(fp_or_buf, **kwds)
  File "/usr/local/lib/python3.6/site-packages/pandas/io/parsers.py", line 948, in __init__
    self._make_engine(self.engine)
  File "/usr/local/lib/python3.6/site-packages/pandas/io/parsers.py", line 1180, in _make_engine
    self._engine = CParserWrapper(self.f, **self.options)
  File "/usr/local/lib/python3.6/site-packages/pandas/io/parsers.py", line 2010, in __init__
    self._reader = parsers.TextReader(src, **kwds)
  File "pandas/_libs/parsers.pyx", line 382, in pandas._libs.parsers.TextReader._cinit__
  File "pandas/_libs/parsers.pyx", line 674, in pandas._libs.parsers.TextReader._setup_parser_source
FileNotFoundError: [Errno 2] No such file or directory: 'weatherOpenMap.csv'
(base) fitec@fitec-HP-ProBook-450-G5:~/Documents/france_temperatures$ sudo docker volume ls
[sudo] Mot de passe de fitec :
DRIVER    VOLUME NAME
local     37f0d6918a28d50933a2971cce152af5d4f535c56a06564cd4d2fa864f422467
local     20196c19c507d469811839acdcf434f20a6c2bd2bf31a8e5f0fa02f3e13c0a0
local     1979240622d7c9986313f32ca5ea15cf9a247f61485c3fd7b00be7c1649c4e99
local     api_weather_data
local     c1b4766deb33ae232275736fc1cf148f209157ac225056750f775210179ba79d
local     france_temperature
local     jobs-volume
local     portainer_data
local     temperatures_france
(base) fitec@fitec-HP-ProBook-450-G5:~/Documents/france_temperatures$ sudo ls /var/lib/docker/volumes
/api_weather_data
(base) fitec@fitec-HP-ProBook-450-G5:~/Documents/france_temperatures$ sudo ls /var/lib/docker/volumes
/api_weather_data/_data
city.list.json  crawler.py  mael_weatherOpenMap.csv  requirements.txt
```



9- push du repo vers github en attendant de réparer l'erreur de récupération du csv

10- Recréer tout l'image/volume docker et relancer les appels d'api pour reconstruire le csv a rajouter plus tard au github

```
(base) fitec@fitec-HP-ProBook-450-G5:~/Documents/france_temperatures$ git remote add origin https://github.com/Spurda/france_temperatures.git
fatal: la distante origin existe déjà.
(base) fitec@fitec-HP-ProBook-450-G5:~/Documents/france_temperatures$ git remote remove origin
(base) fitec@fitec-HP-ProBook-450-G5:~/Documents/france_temperatures$ git remote add origin https://github.com/Spurda/france_temperatures.git
(base) fitec@fitec-HP-ProBook-450-G5:~/Documents/france_temperatures$ git branch -M main
(base) fitec@fitec-HP-ProBook-450-G5:~/Documents/france_temperatures$ git push -u origin main
Username for 'https://github.com': Spurda
Password for 'https://Spurda@github.com':
Énumération des objets: 9, fait.
Décompte des objets: 100% (9/9), fait.
Compression par delta en utilisant jusqu'à 8 fils d'exécution
Compression des objets: 100% (8/8), fait.
Écriture des objets: 100% (9/9), 4.10 Mio | 2.58 Mio/s, fait.
Total 9 (delta 2), réutilisés 0 (delta 0)
remote: Resolving deltas: 100% (2/2), done.
To https://github.com/Spurda/france_temperatures.git
 * [new branch]      main -> main
La branche 'main' est paramétrée pour suivre la branche distante 'main' depuis 'origin'.
(base) fitec@fitec-HP-ProBook-450-G5:~/Documents/france_temperatures$
```