DME solutions homework: lesson 1

Boris Wolvers

February 22, 2023

Setup

Set up a project directory structure

The first question is to create a new folder outside of the DME22 folder called **cleanup**. The second question is to initialize a new poetry environment. But lets combine these two questions. So use your terminal and go to your **code** folder, from there we create a new **cleanup** folder whilst at the same time creating a new poetry environment. Commands used:

```
cd ~/code
poetry new cleanup
```

If everything goes right, you should have a new **cleanup** folder outside of your DME22 folder.

Now go into your new created **cleanup** folder and add the libraries you need, e.g.:

```
cd cleanup
poetry add pandas
poetry add numpy
poetry add pydantic
poetry add loguru
poetry add logger
poetry add pyarrow
poetry add jupyter
poetry add click
poetry add black
```

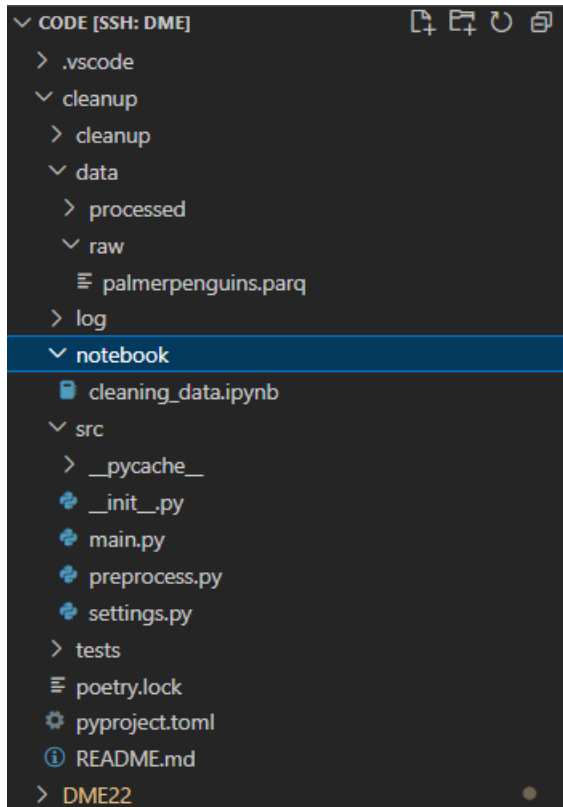Next we create a **data/raw** folder and copy **palmerpenguins.parq** into it:

```
mkdir data
cd data
mkdir raw
cp /home/admindme/code/DME22/data/raw/palmerpenguins.parq
/home/admindme/code/cleanup/data/raw
```

Finally we are going to create the final folders.

```
cd ~/code/cleanup/data
mkdir processed
cd ..
mkdir src
```

```
mkdir  notebook
```

To create the **main.py** file, we can go to the explorer, search for the src folder, right mouse click, create new file and save it as main.py. The final result of the folder structure should look something like this:

Create a prototype notebook

Go through the notebook file in the notebook folder.

Implement

The code can be seen and tested through the terminal, e.g.:

```
cd ~/code/cleanup
poetry shell
python src/main.py --file=palmerpenguins.parq
```

You can also format all of your python code inside the src folder by using black, e.g.:

```
cd ~/code/cleanup
black src
```