READ_ME

Adriana Pitea

2024-12-19

Jupyter Notebook Docker Container

This repository contains a Docker image for running a Jupyter Notebook with Python and to read in the ***taxonomic_data.csv*** and perform the following tasks:

1. Calculate total species count for each taxonomic phylum and the average species count per species within each phylum. Write the results to *csv* files.
2. Generate a bar chart showing the total species count for each phylum.

How to Build and Run

1. Create a folder and store the Dockerfile from the git repository in it.

```
mkdir jupyter-docker
cd jupyter-docker
```

2. The Dockerfile uses a pre-built image for common data science libraries:

```
# Use the official Jupyter notebook image
FROM jupyter/scipy-notebook:latest

# Set the working directory in the container
WORKDIR /Users/ads13/Documents/Prepaire_test

# Expose port 8888 to access Jupyter from the browser
EXPOSE 8888

# Start Jupyter Notebook when the container launches
CMD ["start-notebook.sh", "--NotebookApp.token=''", "--NotebookApp.password=''"]
```

3. **Build the Docker Image**:

```
docker build -t jupyter-notebook-image .
```

4. **Run the Docker Image**:

```
docker run -p 8888:8888 -v ~/Prepaire_test:/home/jovyan/work jupyter/scipy-notebook:latest
```

To mount your local directory (for example, the directory with the Jupyter notebook) to the container so that any files you have in the folder will appear in the Jupyter environment, we use the -v flag to bind mount a directory. I also had the csv input file in this directory.

5. **Access the Jupyter Notebook**: Open a browser and go to http://127.0.0.1:8888 to access the Jupyter Notebook, open the *ipynb* and run the code.