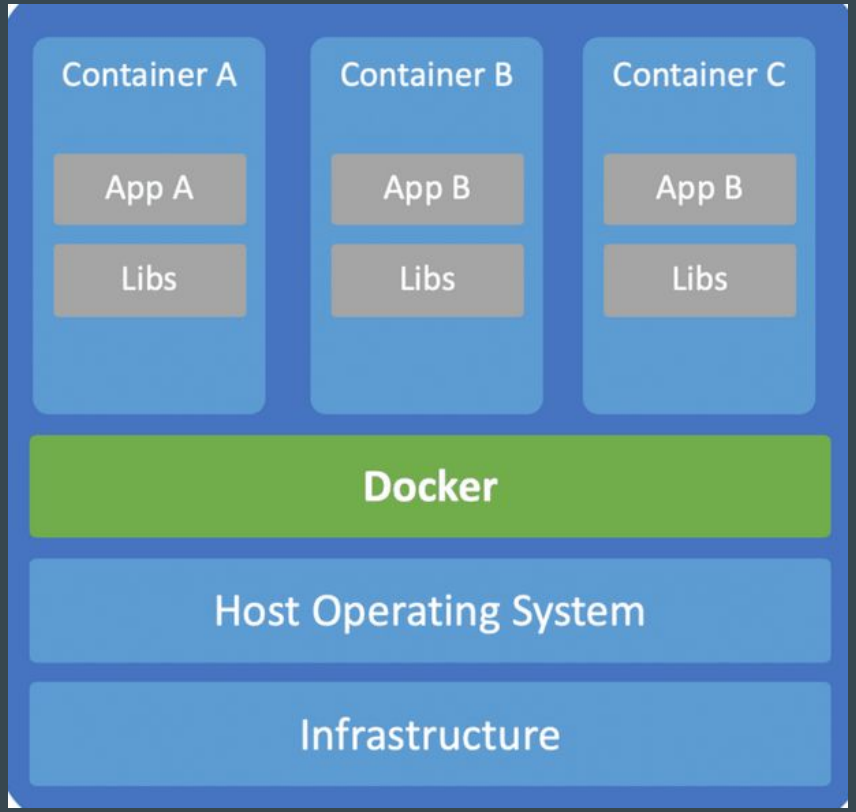
Контейнеризация рабочей среды для ML

Павел Вешкин

О чем поговорим

- Кратко о docker
- Зачем использовать docker? Какие задачи можно решить?
- Colab, Kaggle, . . .
- Jupyter lab
- Dev container в vscode

Кратко о docker



- Образ
- Dockerfile
- Контейнер
- Volumes
- Docker-compose.yml

Docker

Плюсы

- Низкий уровень потребления ресурсов.
- Скорость развертывания.
- Простое скрытие процессов.
- Простое масштабирование.
- Простой и удобный запуск.

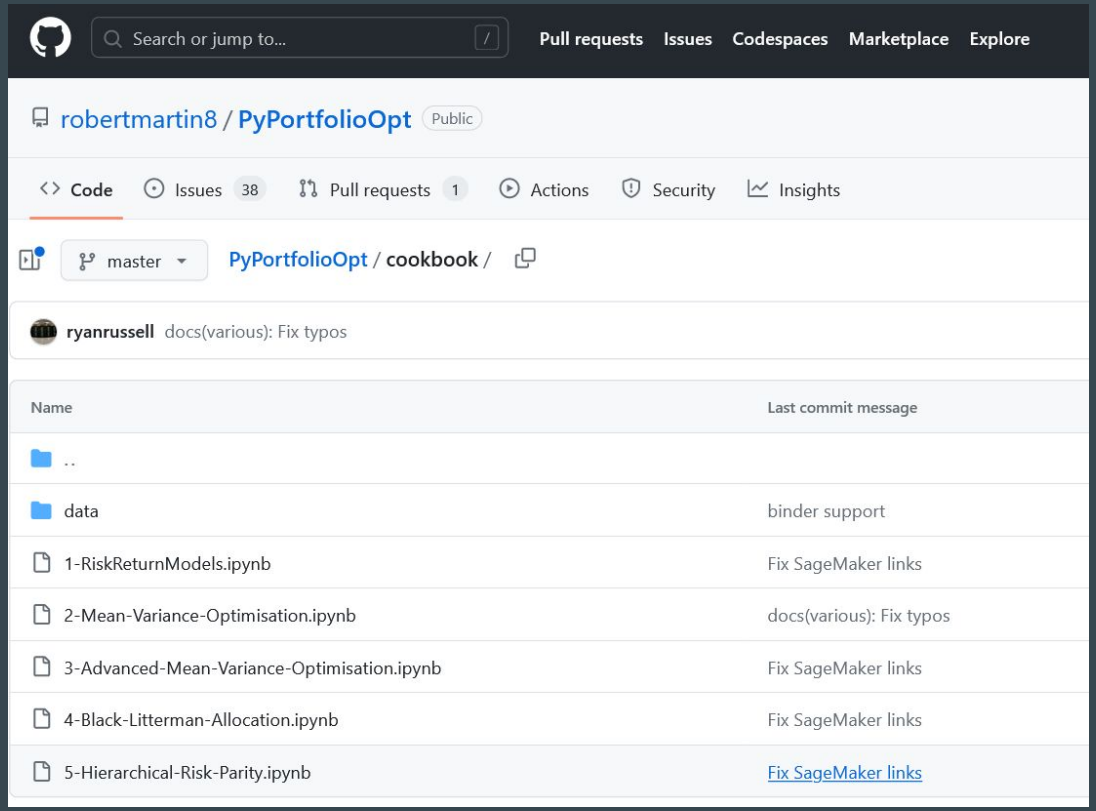
Минусы

- иногда требует много памяти
- исключительно linux внутри

Какие задачи можно решить?

- запустить проект из коробки не засоряя систему
- запускать ноутбук из преднастроенной среды если не хватает pip
- передать проект полностью без проблем с совместимостью
- перенести проект в облако

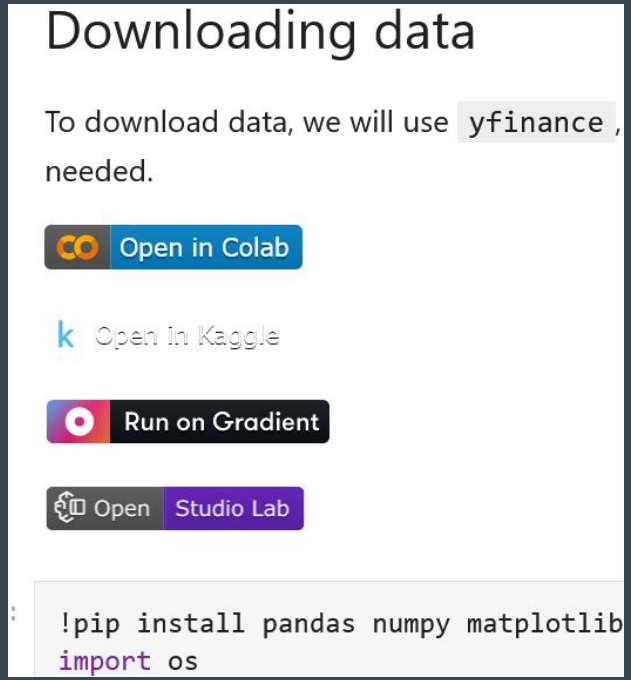
Github и ноутбуки



Colab, kaggle kernels

- Colab - это бесплатная интерактивная облачная среда для работы с кодом
- Kaggle kernels - облачная среда разработки по типу Jupyter Notebook

Смотрим ноутбуки с github



Если нет иконок

оригинальный адрес ноутбука:

https://github.com/robertmartin8/PyPortfolioOpt/blob/master/cookbook/2-Mean-Variance-Optimisation.ipynb

адрес для запуска в ноутбука в colab:

https://colab.research.google.com/github/robertmartin8/PyPortfolioOpt/blob/master/cookbook/2-Mean-Variance-Optimisation.ipynb

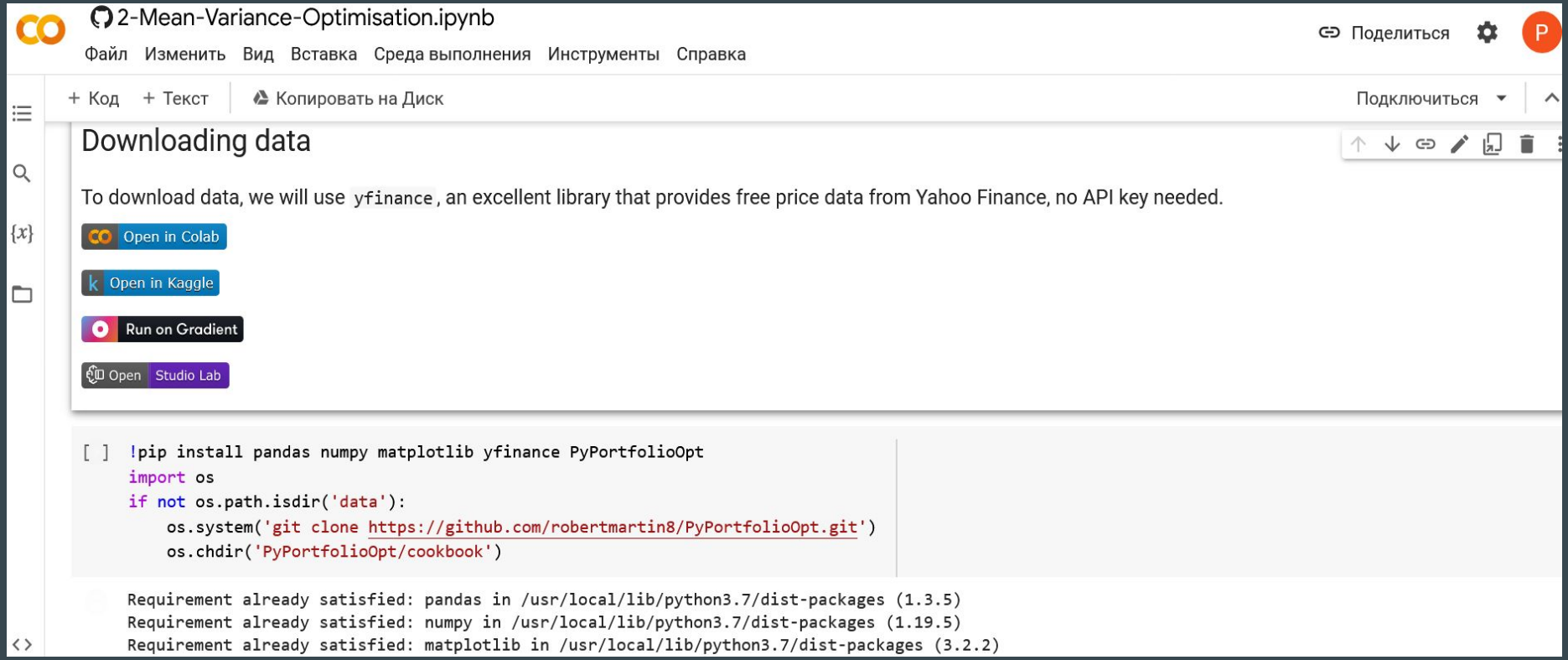
Действие меняем в адресе:

https://github.com/

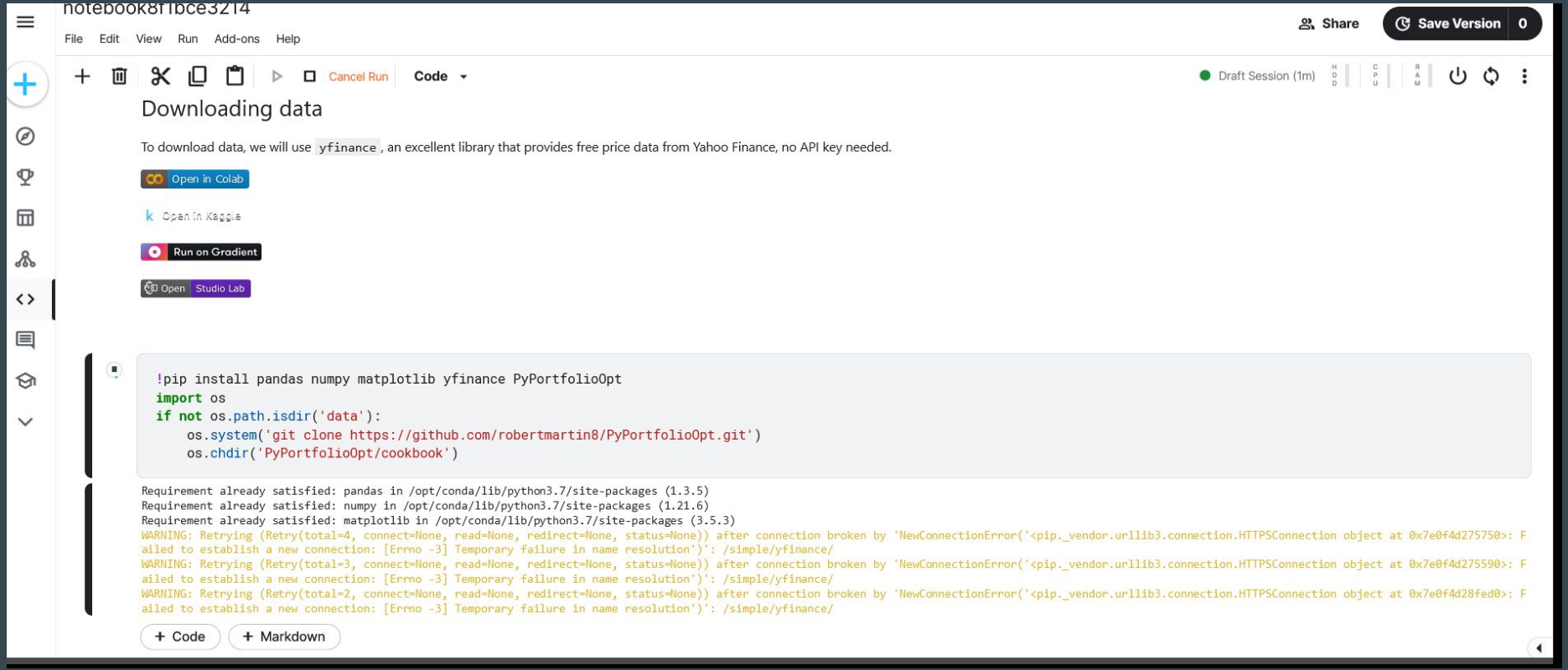
на

https://colab.research.google.com/github/

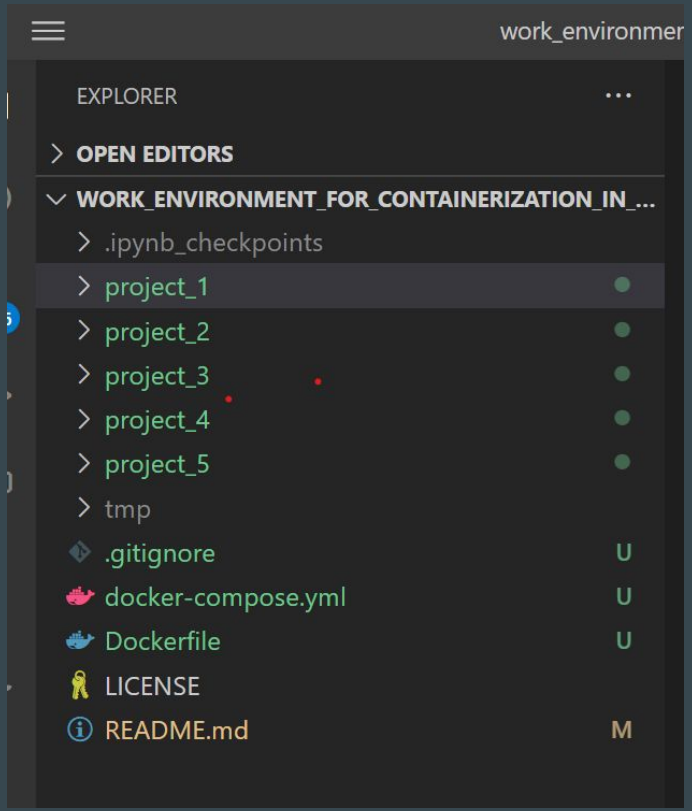
ноутбук в colab



ноутбук в Kaggle kernels



Пример


work_environmer
EXPLORER
OPEN EDITORS
WORK_ENVIRONMENT_FOR_CONTAINERIZATION_IN_...
.ipynb_checkpoints
project_1
project_2
project_3
project_4
project_5
tmp
.gitignore
U
docker-compose.yml
U
Dockerfile
U
LICENSE
README.md
M

Jupyter lab

- jupyter образы
  - Где взять и что выбрать
  - настройка Dockerfile
    - пользователь
    - выбор образа
    - общая папка
    -

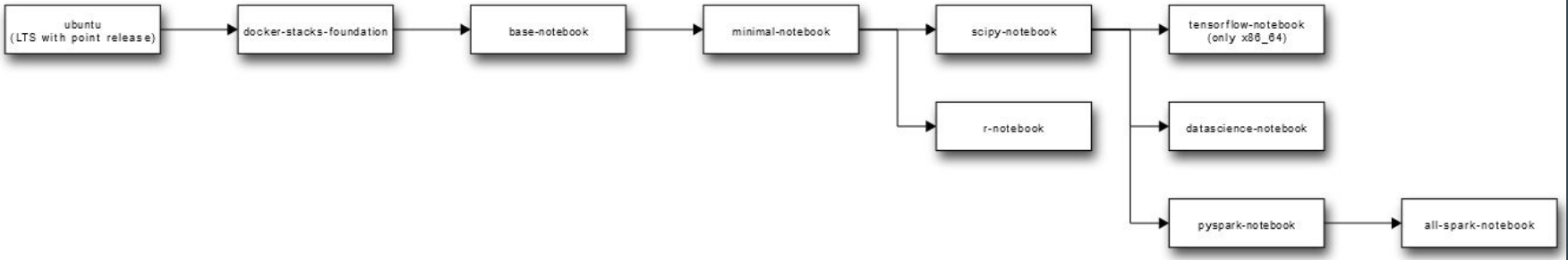
- общий docker compose для работы с исследовательскими ноутбуками

- локальный docker compose для проекта

jupyter образы Jupyter Docker Stacks

Image Relationships

The following diagram depicts the build dependency tree of the core images. (i.e., the `FROM` statements in their Dockerfiles). Any given image inherits the complete content of all ancestor images pointing to it.



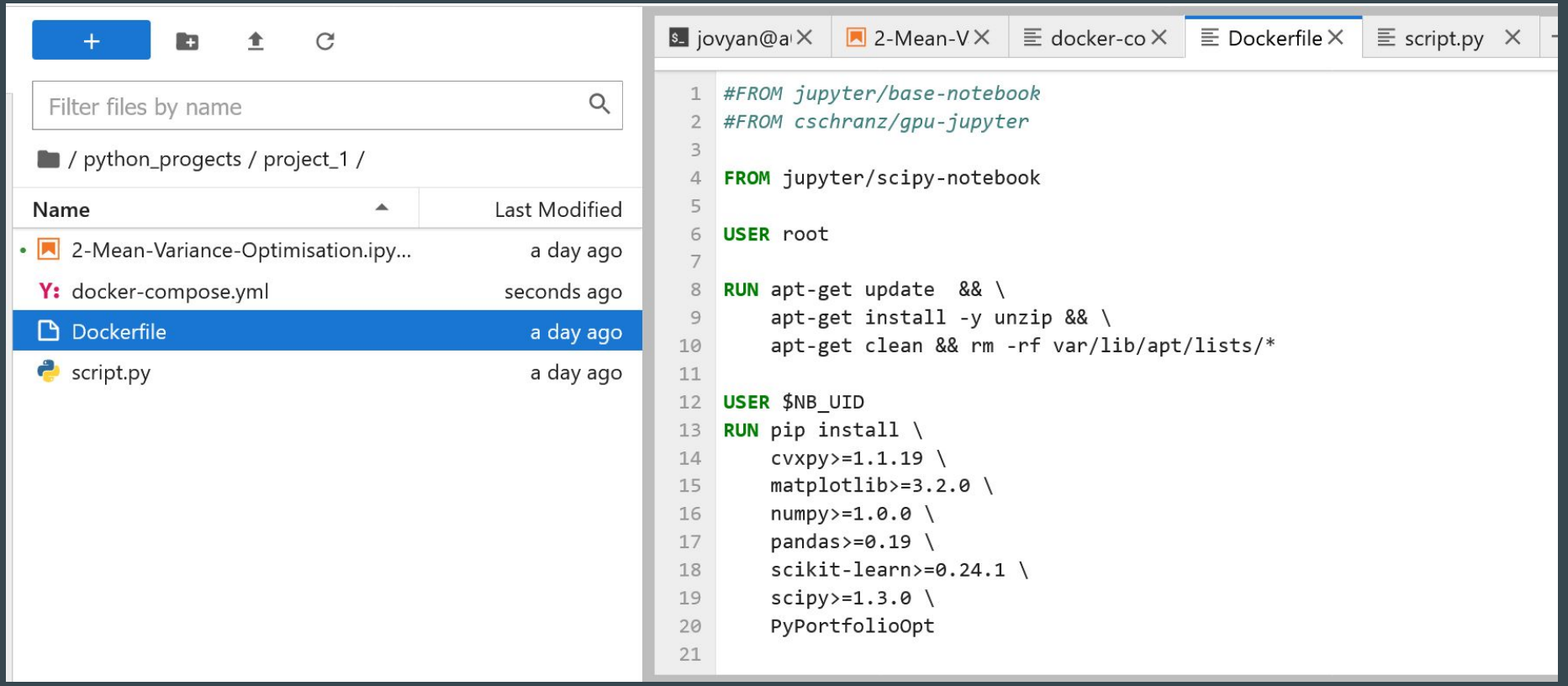
Настройка Dockerfile для всех проектов

```
4  FROM jupyter/scipy-notebook
5
6  USER root
7
8  RUN apt-get update  && \
9      apt-get install -y unzip && \
10     apt-get clean && rm -rf var/lib/apt/lists/*
11
12 USER $NB_UID
13 RUN pip install \
14     cvxpy>=1.1.19 \
15     matplotlib>=3.2.0 \
16     numpy>=1.0.0 \
17     pandas>=0.19 \
18     scikit-learn>=0.24.1 \
19     scipy>=1.3.0 \
20     PyPortfolioOpt
```

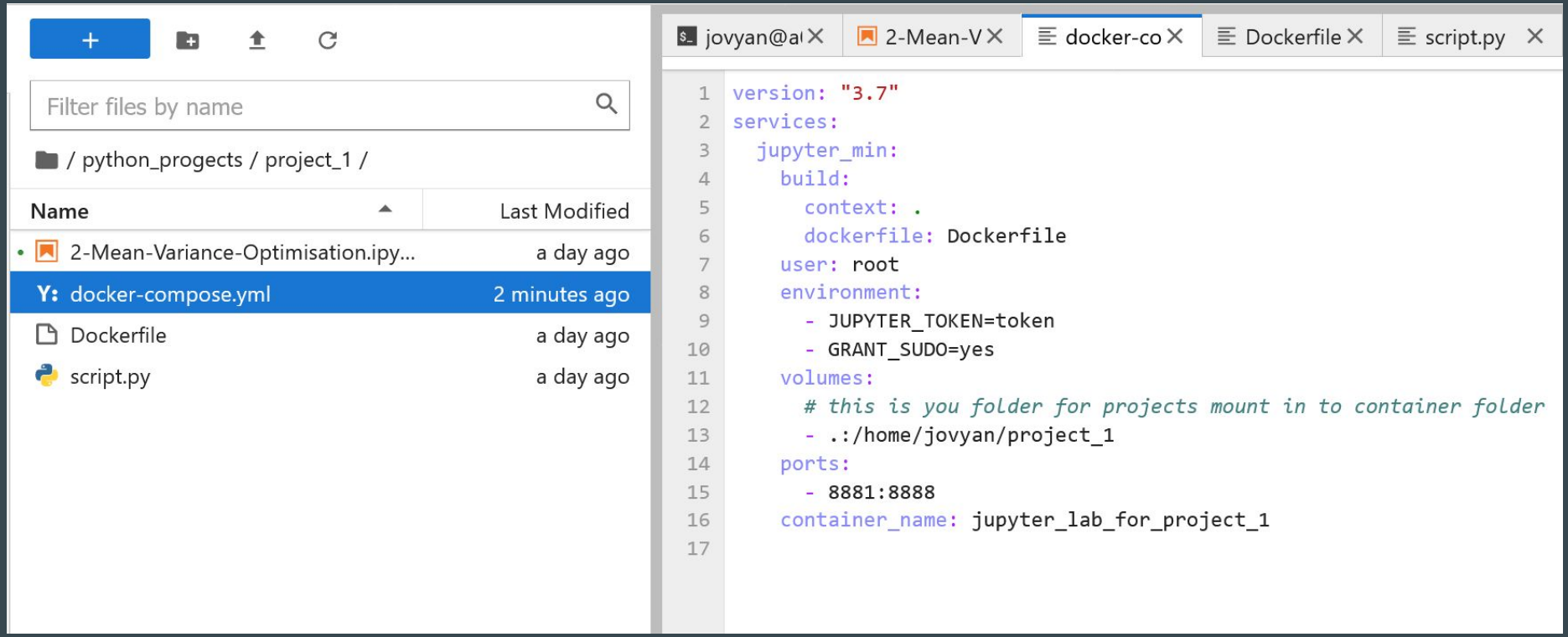
Настройка docker-compose.yml для всех проектов

```yaml
version: "3.7"
services:
  jupyter_lab_all:
    build:
      context: .
      dockerfile: Dockerfile
    user: root
    environment:
      - JUPYTER_TOKEN=token
      - GRANT_SUDO=yes
    volumes:
      # this is you folder for projects mount into container folder
      - .:/home/jovyan/python_progects
    ports:
      - 8880:8888
    container_name: jupyter_lab_for_all_projects
```

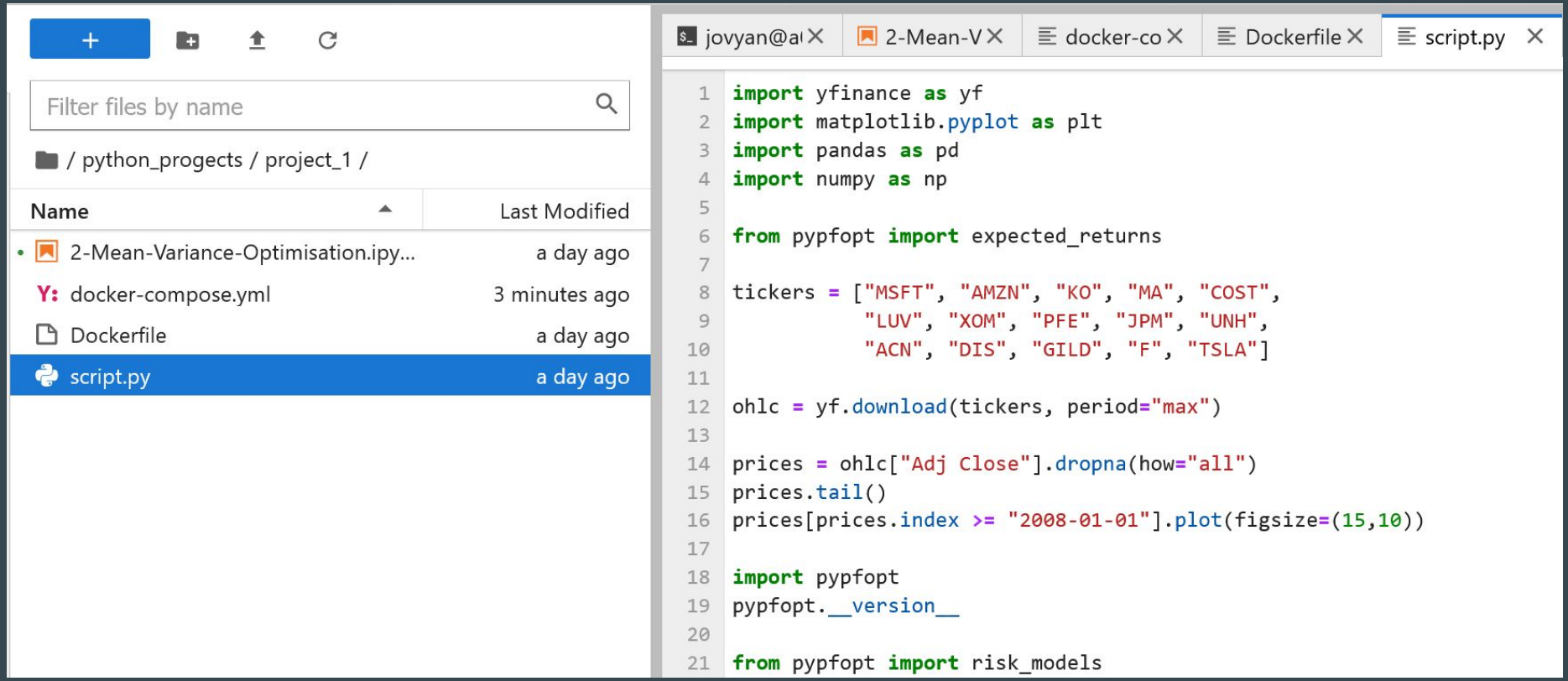
Настройка Dockerfile для конкретного проекта



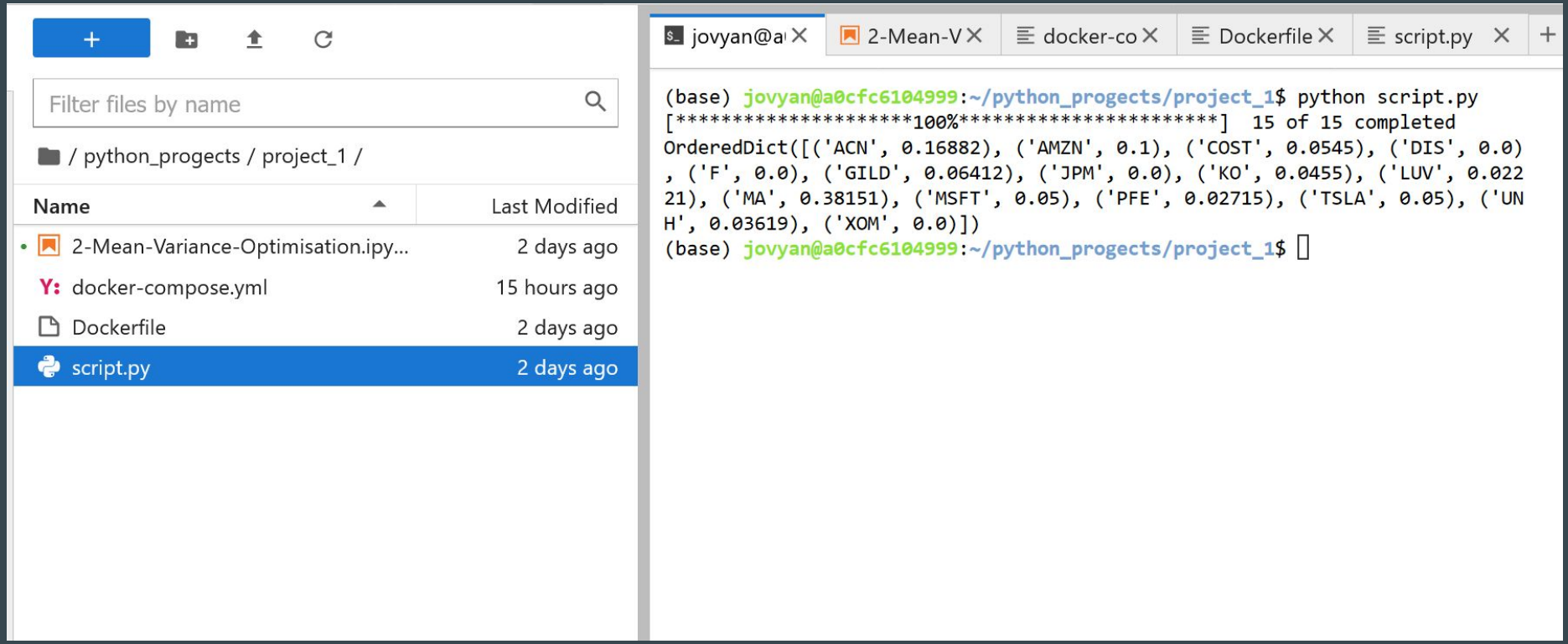
Настройка docker-compose.yml для конкретного проекта



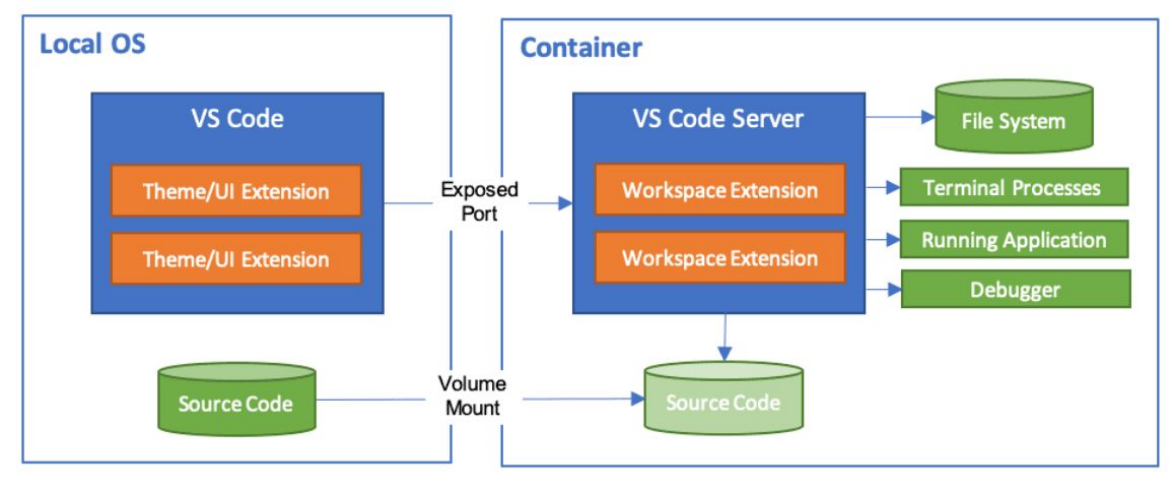
script.py для дальнейшего использования как сервиса



Запуск script.py

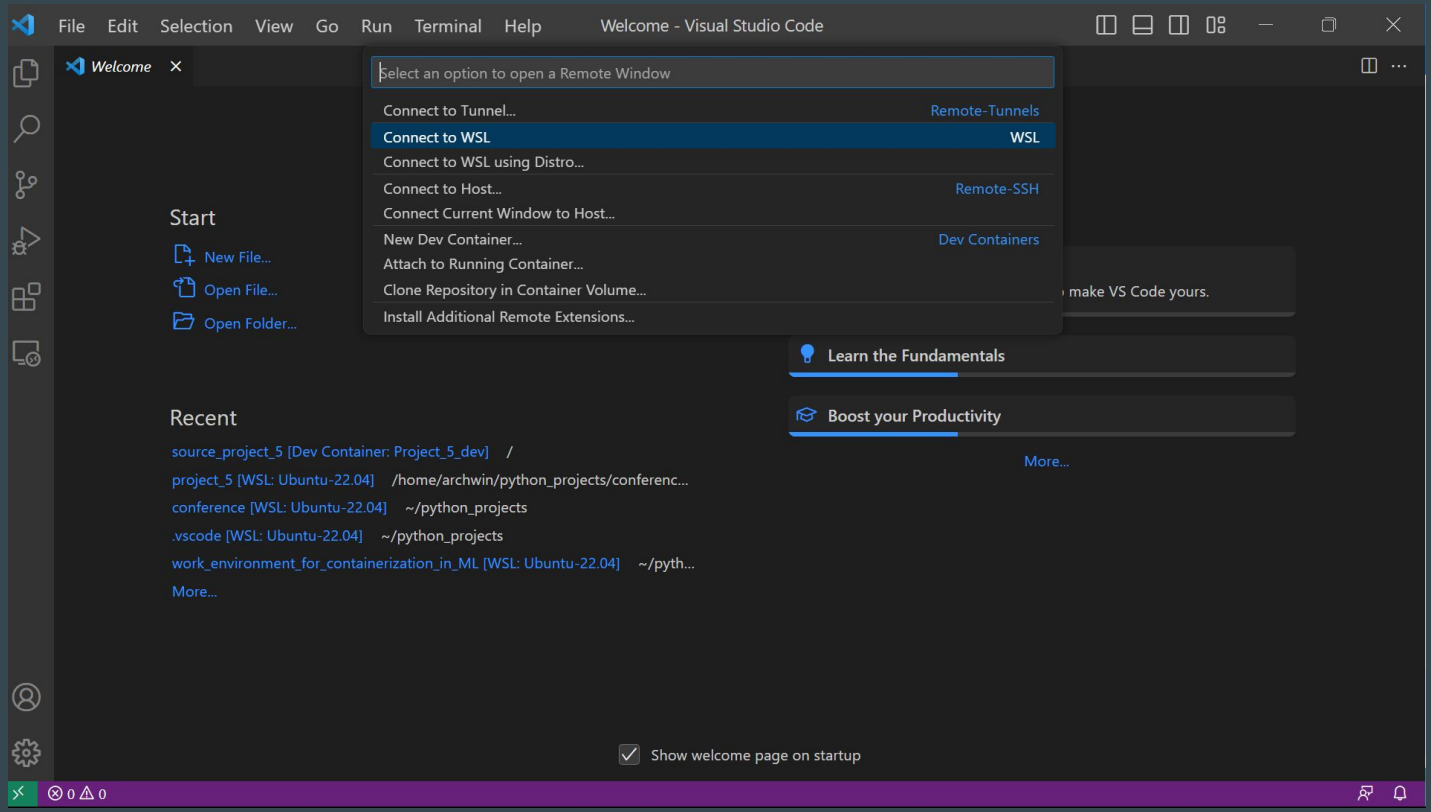


Dev_container в vscode

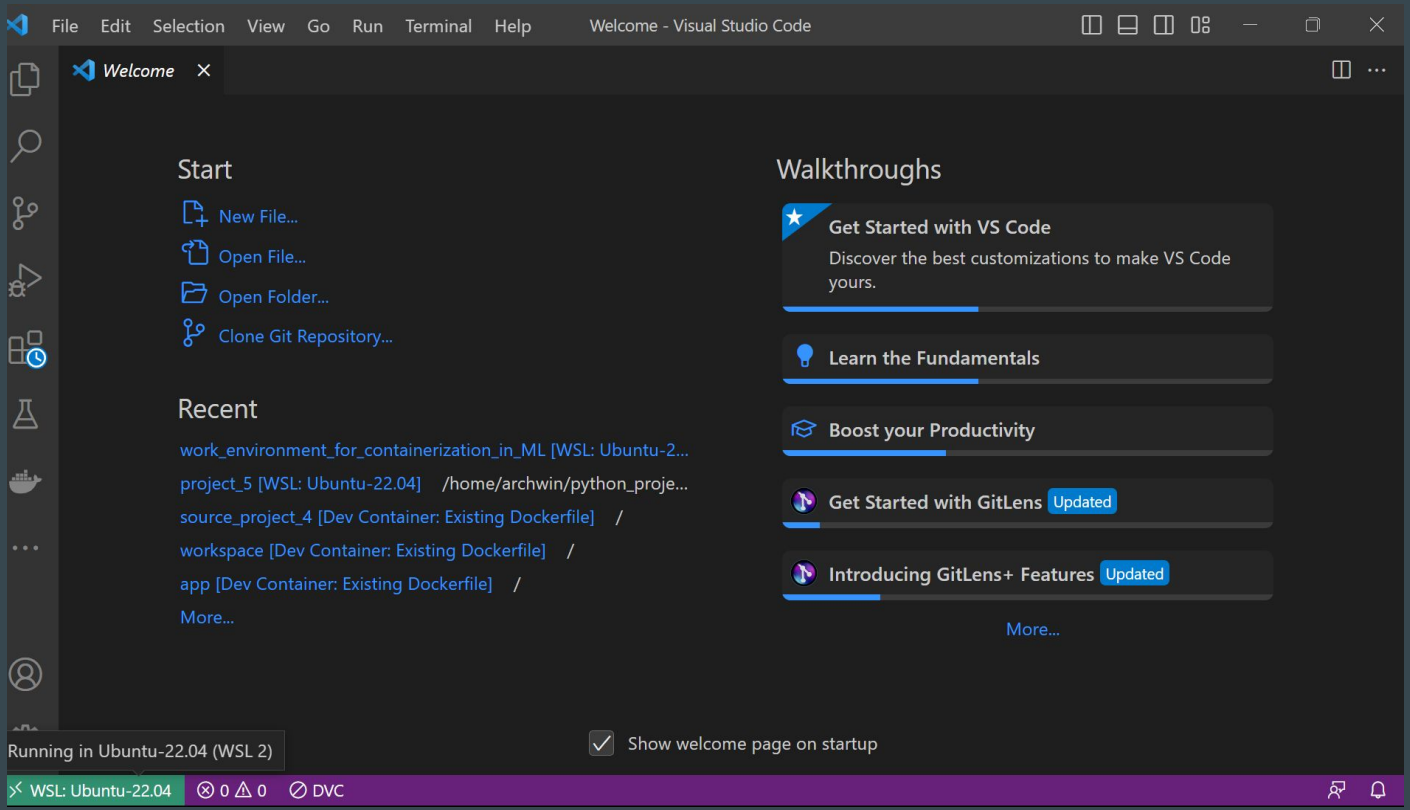


- создаем проект с нуля
- качаем проект с github
- присоединяемся к уже работающему docker контейнеру

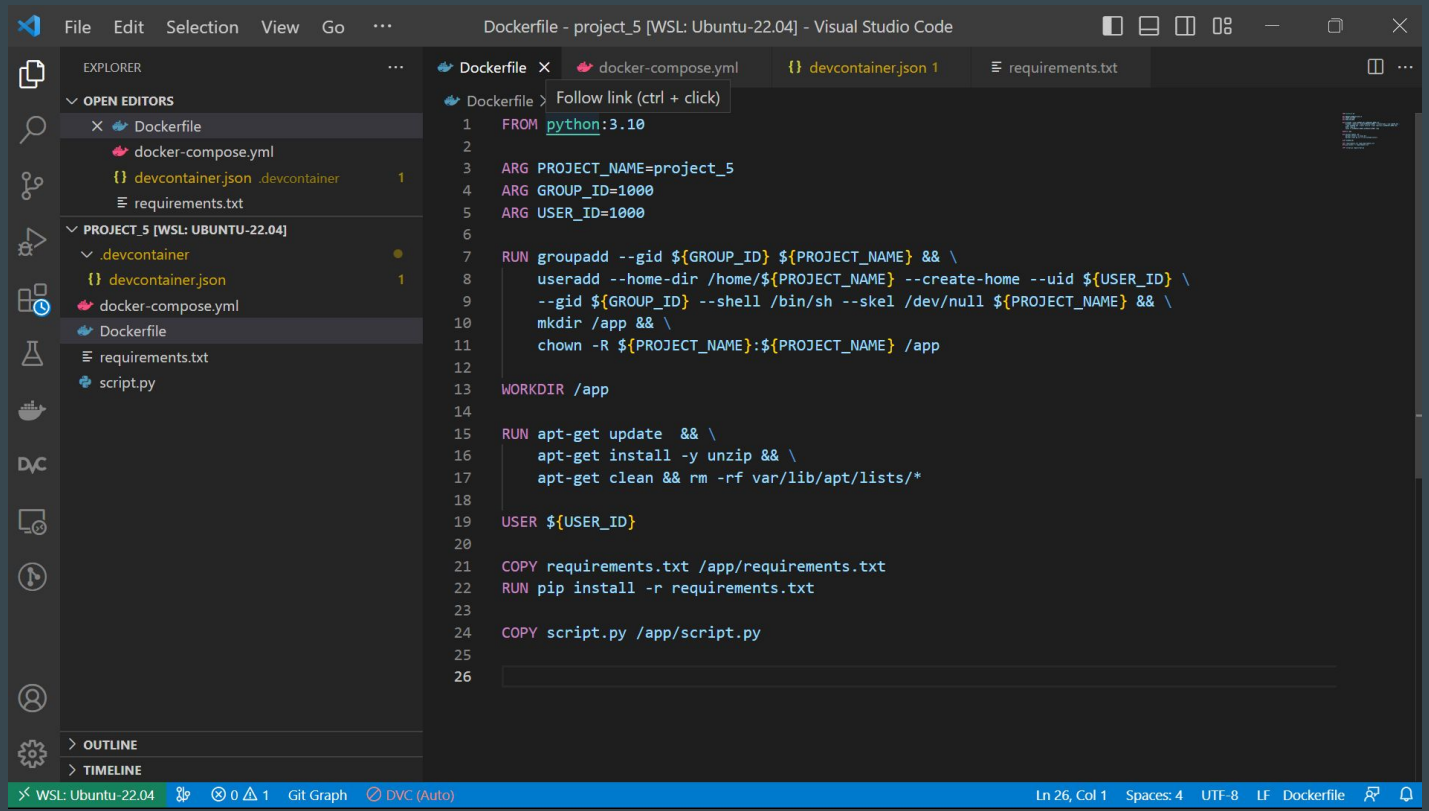
Vscode


File Edit Selection View Go Run Terminal Help
Welcome - Visual Studio Code
Welcome
Select an option to open a Remote Window
Connect to Tunnel... Remote-Tunnels
Connect to WSL WSL
Connect to WSL using Distro...
Connect to Host... Remote-SSH
Connect Current Window to Host...
New Dev Container... Dev Containers
Attach to Running Container...
Clone Repository in Container Volume...
Install Additional Remote Extensions...
Start
New File...
Open File...
Open Folder...
make VS Code yours.
Learn the Fundamentals
Boost your Productivity
More...
Recent
source_project_5 [Dev Container: Project_5_dev] /
project_5 [WSL: Ubuntu-22.04] /home/archwin/python_projects/conferenc...
conference [WSL: Ubuntu-22.04] ~/python_projects
.vscode [WSL: Ubuntu-22.04] ~/python_projects
work_environment_for_containerization_in_ML [WSL: Ubuntu-22.04] ~/pyth...
More...
Show welcome page on startup
0 0

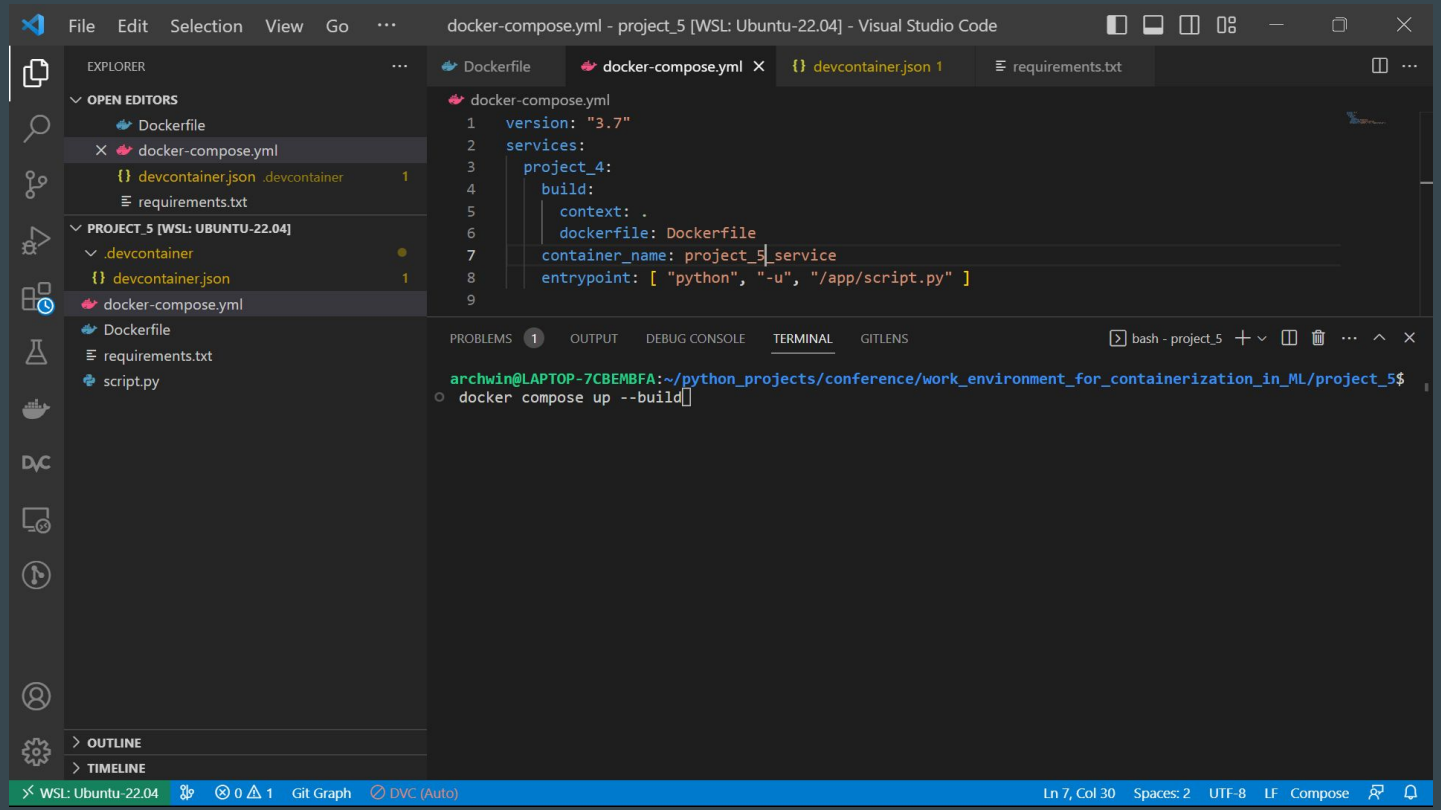
WSL2 в vscode



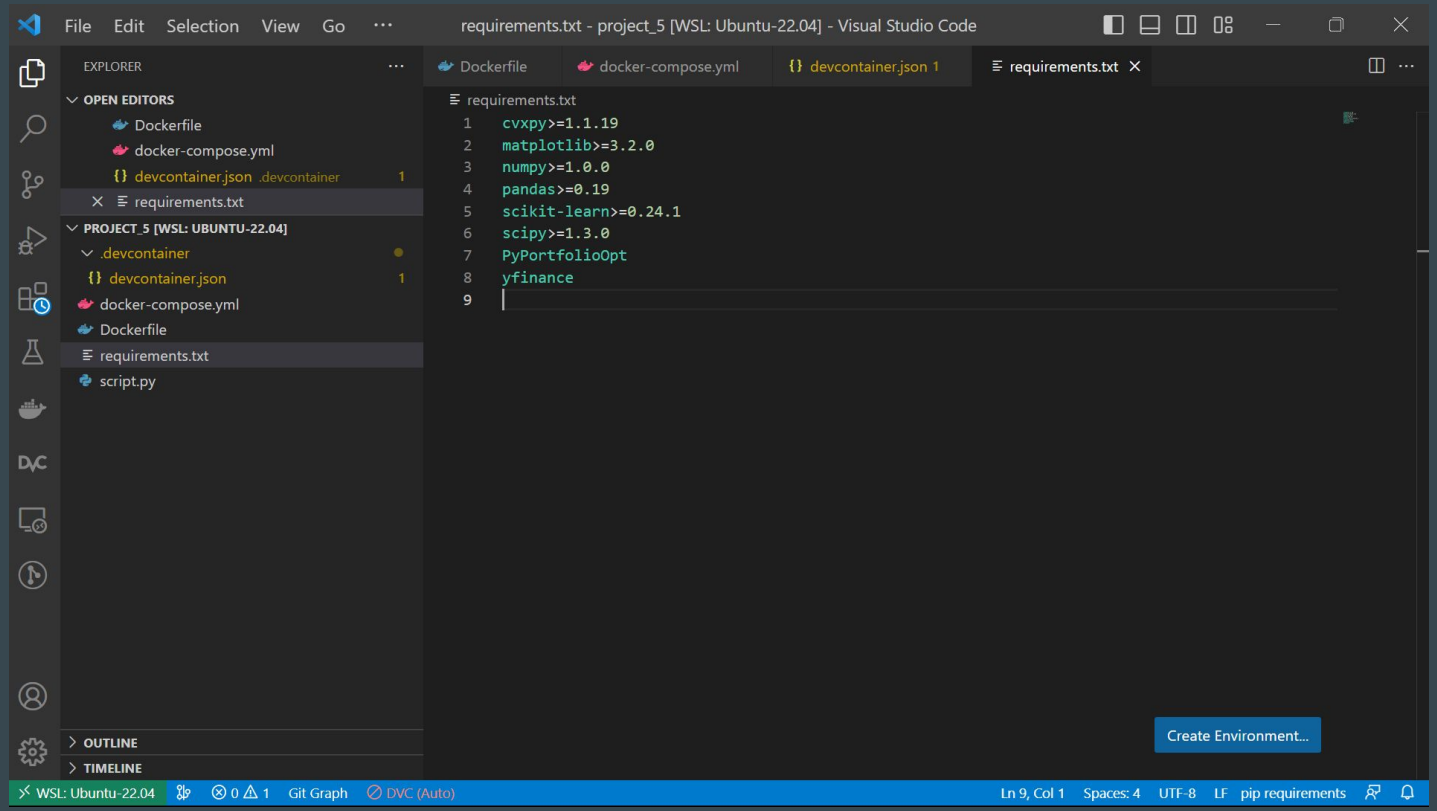
Обновленный Dockerfile



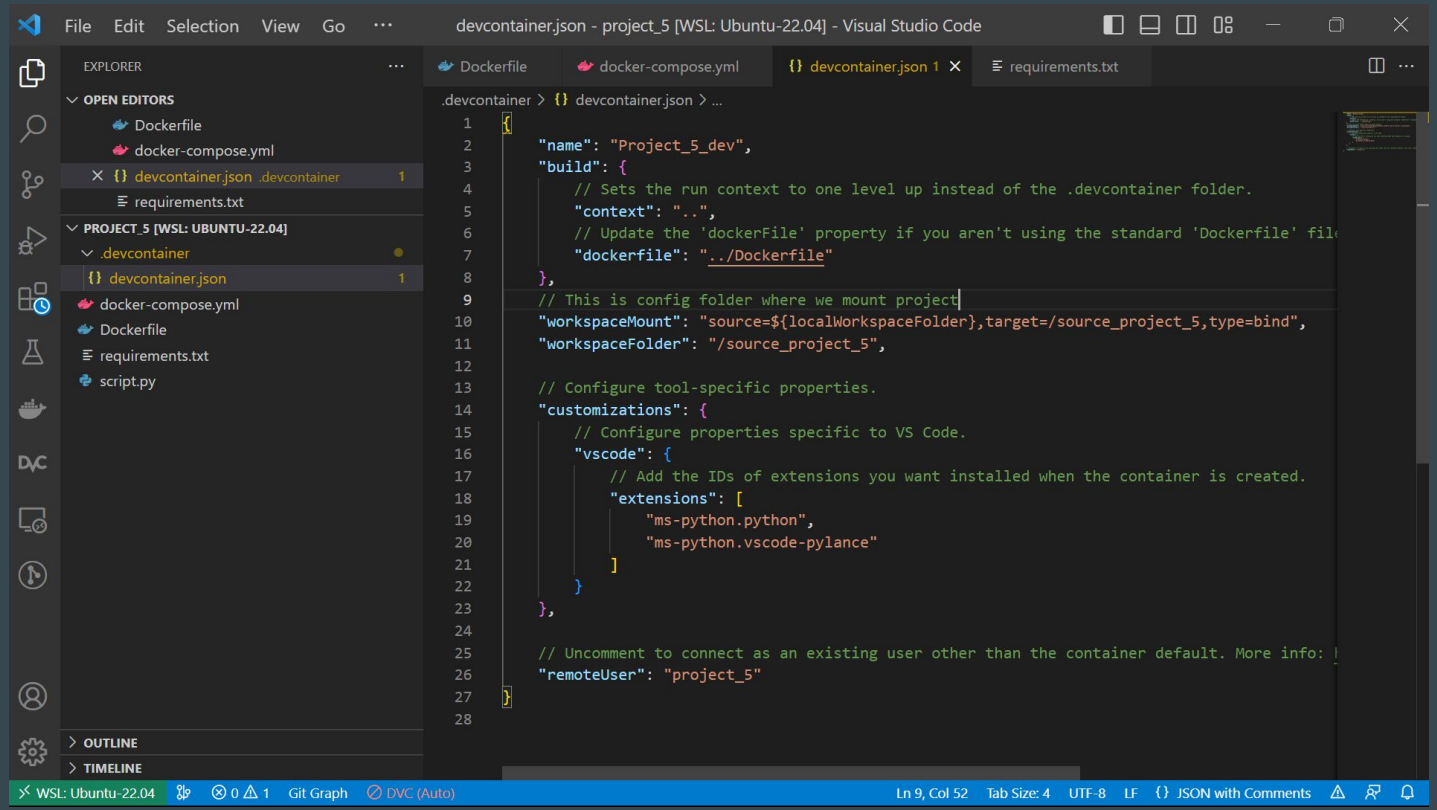
Обновленный docker-compose.yml



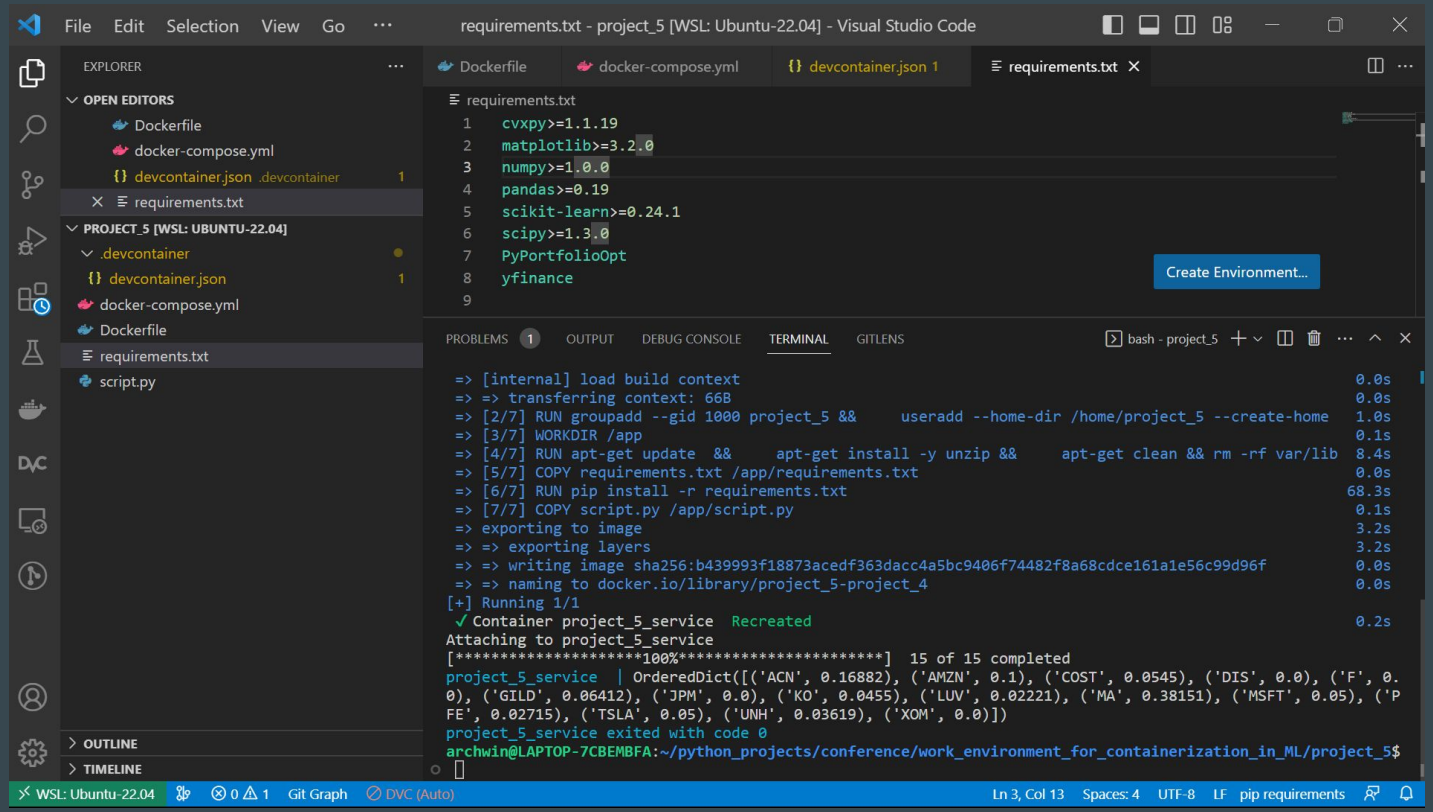
requirements.txt



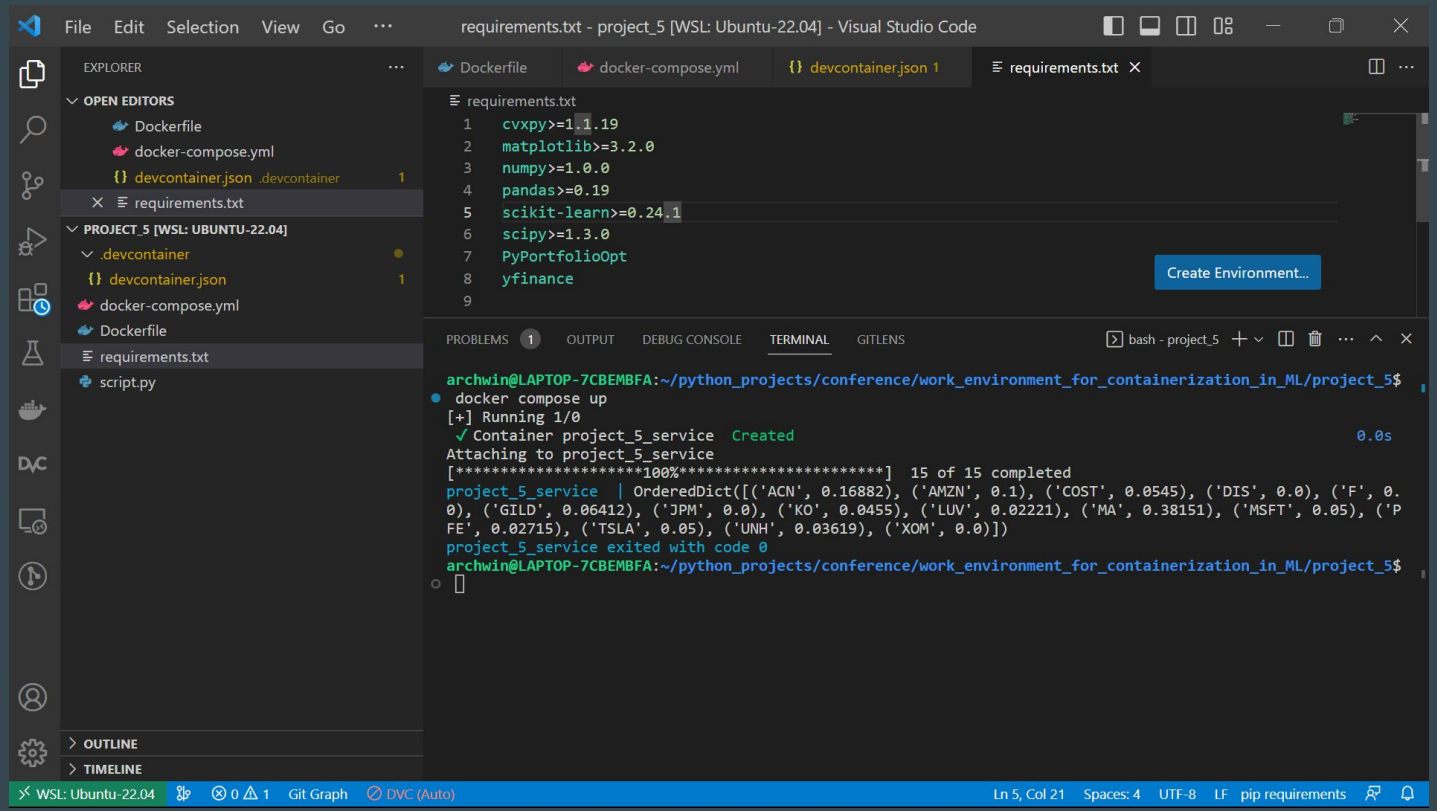
Настройка devcontainer.json



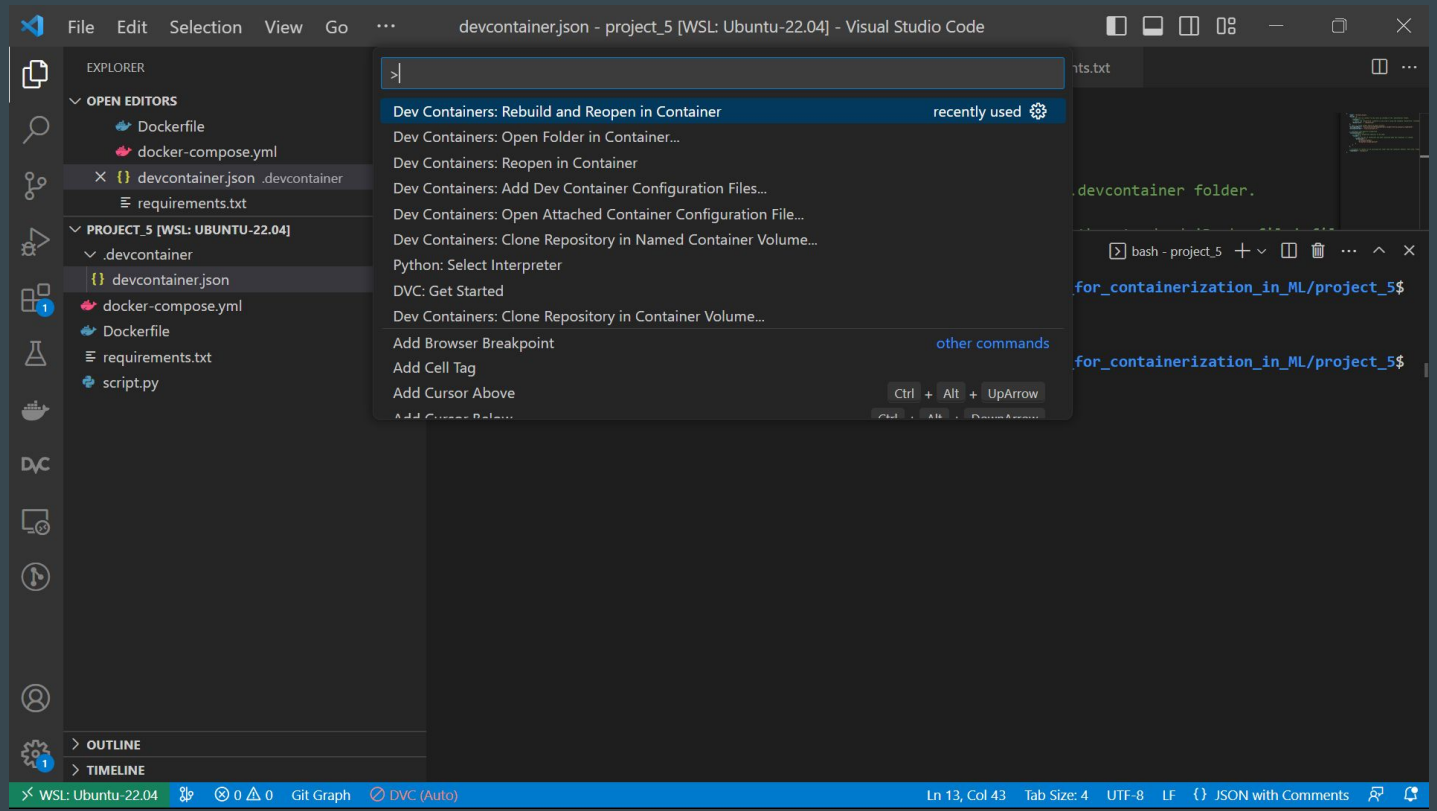
Создаем и запускаем сервис через docker-compose up –build



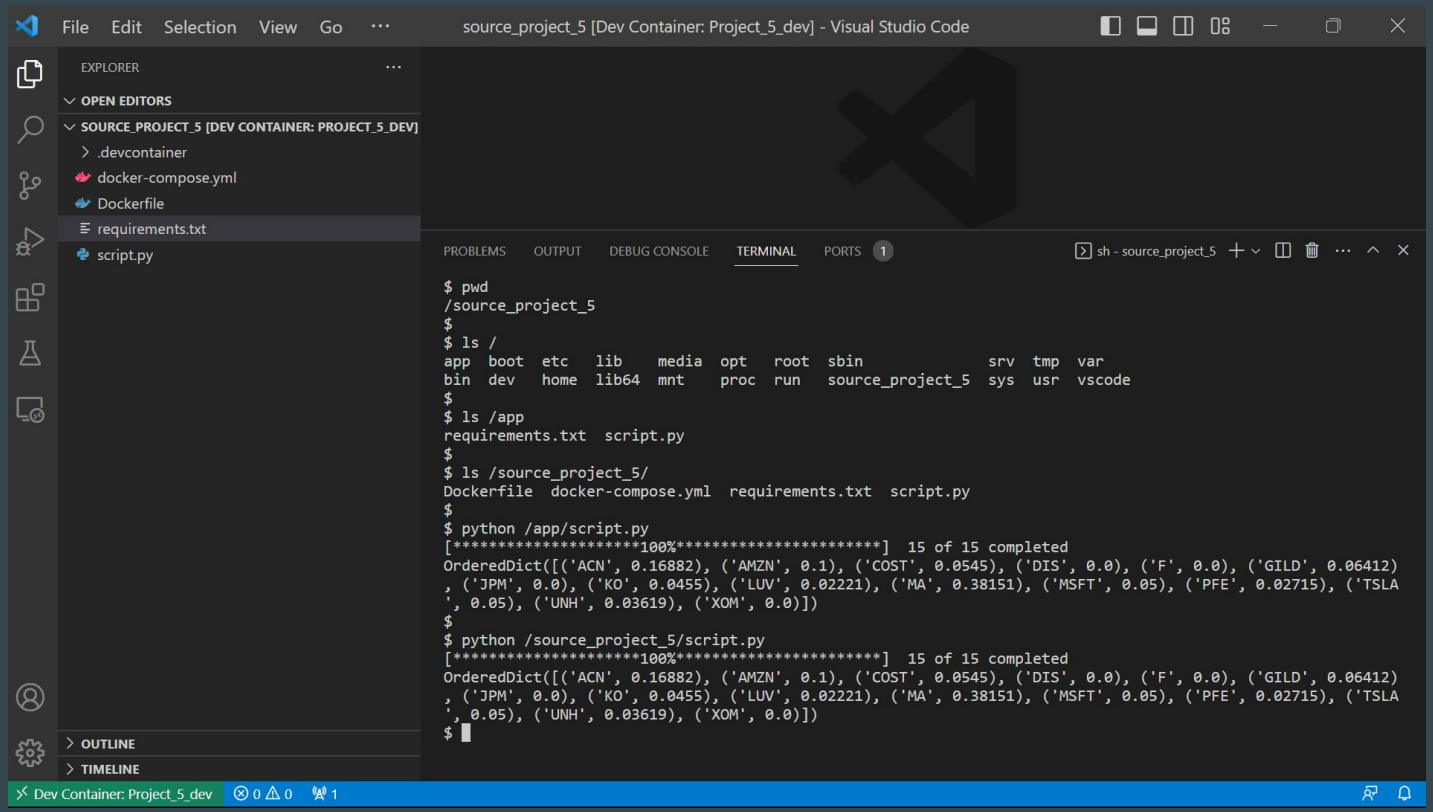
Запускаем сервис без пересоздания



Запускаем и заходим в dev container



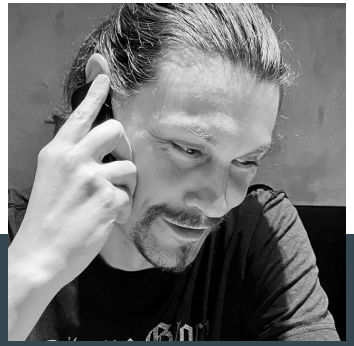
Внутри dev container



Материалы и ссылки:

- Репозитарий с примерами и презентацией
  - https://github.com/Archkras/work_environment_for_containerization_in_ML
- Как использовать jupyter lab в докере
  - https://youtu.be/UXxUcZDSNwA
- Как настроить dev container в vscode
  - https://youtu.be/UP-2vEd3NGw
- Как запустить ноутбук из github в colab
  - https://stackoverflow.com/questions/62596466/how-can-i-run-notebooks-of-a-github-project-in-google-colab
- Какой образ jupyter взять за основу
  - https://jupyter-docker-stacks.readthedocs.io/en/latest/using/selecting.html
- vscode
  - https://jupyter-docker-stacks.readthedocs.io/en/latest/using/selecting.html

Спасибо за ваше время и внимание.



Павел Вешкин

Контакты:

https://t.me/arch_kras

vpa_kras@rambler.ru