Docker для рабочей среды DS

•••

Павел Вешкин

О чем поговорим

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

Установка Docker

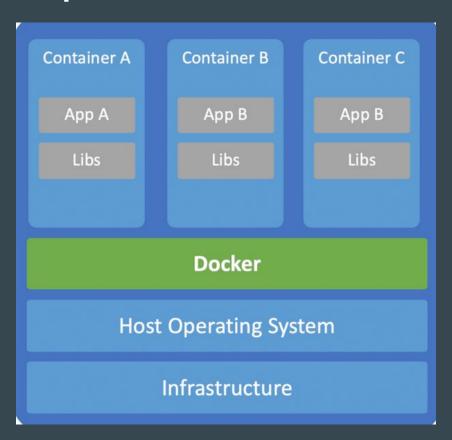
Windows

- Проверяем установлена ли WSL
- заходим на сайт Docker качаем установочную программу и запускаем
- https://docs.docker.com/desktop/install/windows-install/

Linux

- Добавляем репозиторий
- Запускаем команду установки
- https://docs.docker.com/engine/install/ubuntu/

Кратко o 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.

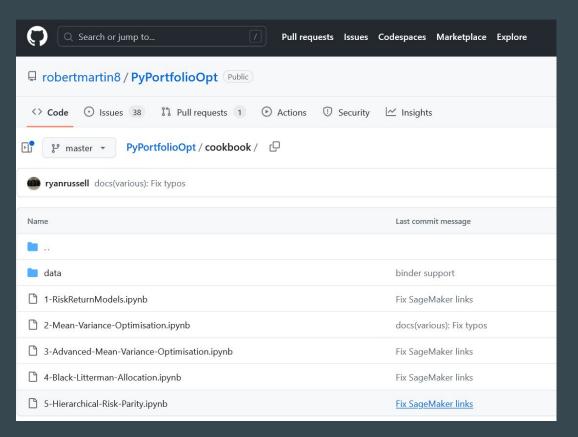
Минусы

- 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 рір
- передать проект полностью без проблем с совместимостью
- перенести проект в облако

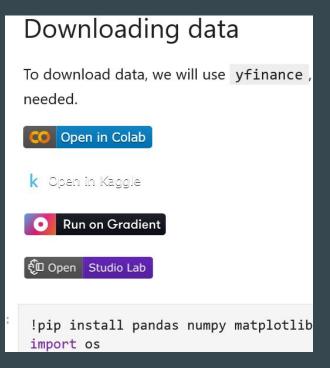
Github и ноутбуки



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/PvPortfolioOpt/blob/master/cookbook/2-Mean-Variance-Optimisation.ipvnb

адрес для запуска в ноутбука в colab:

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

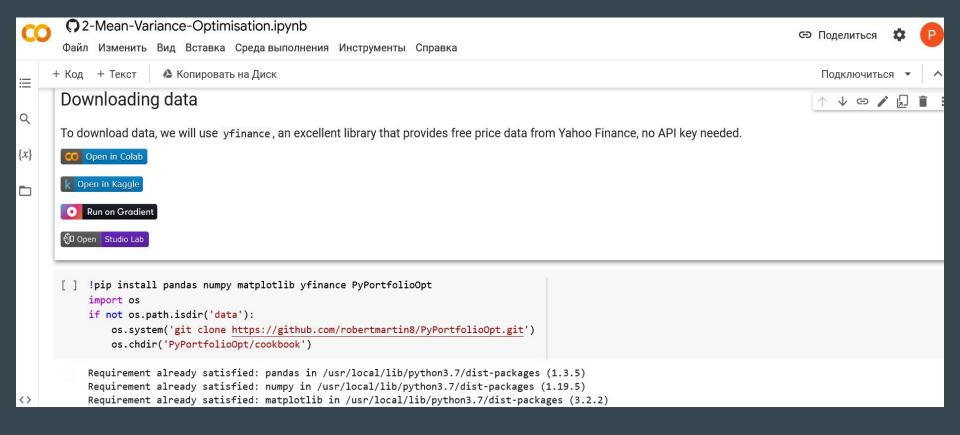
Действие меняем в адресе:

https://github.com/

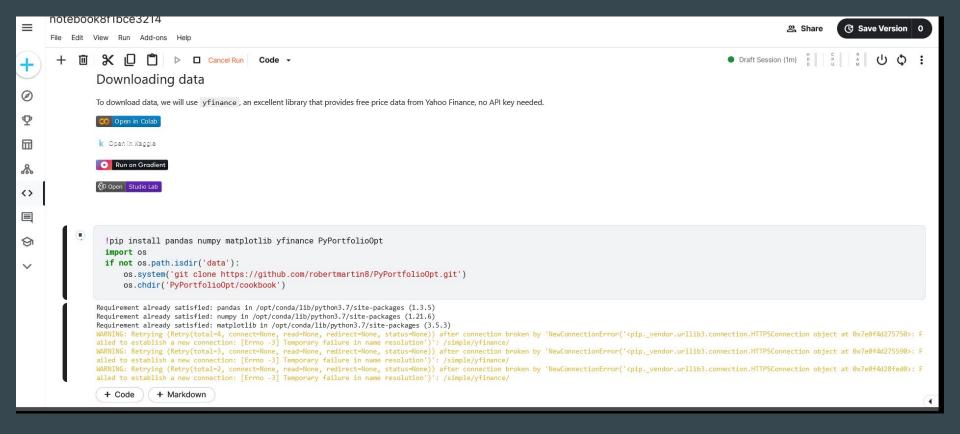
на

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

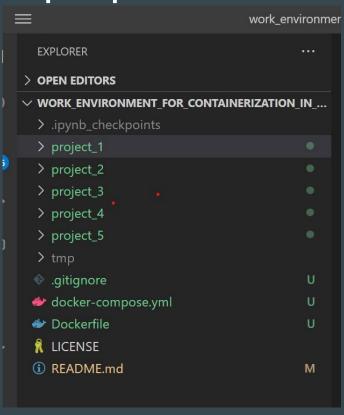
ноутбук в colab



ноутбук в Kaggle kernels



Пример



Jupyter lab

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

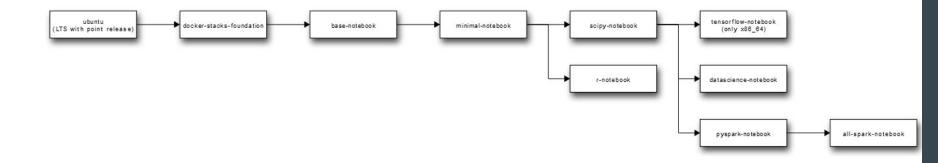
_

- 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

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

Hactpoйка docker-compose.yml для всех проектов

```
version: "3.7"
   services:
     jupyter_lab_all:
        build:
       context: .
          dockerfile: Dockerfile
      user: root
        environment:

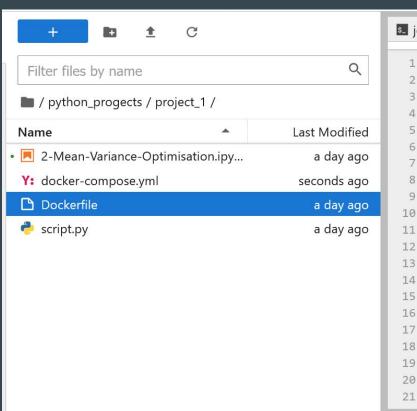
    JUPYTER TOKEN=token

 9

    GRANT SUDO=yes

10
11
       volumes:
          # this is you folder for projects mount into container folder
          - .:/home/jovyan/python progects
13
14
        ports:
15
          - 8880:8888
        container name: jupyter lab for all projects
16
```

Hactpoйкa Dockerfile для конкретного проекта



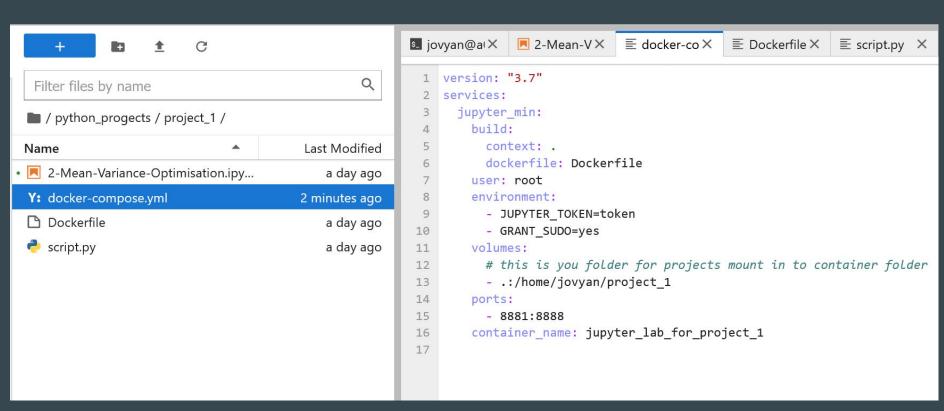
```
■ Dockerfile X
s_ jovyan@a ×
               ■ 2-Mean-V X

    docker-co ×

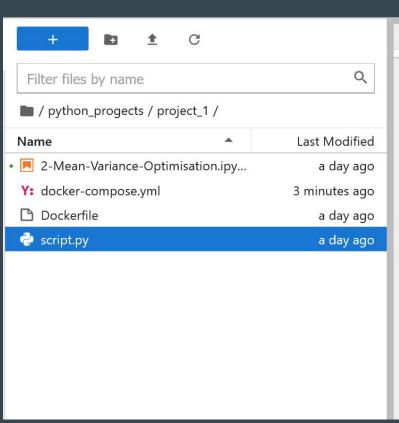
    ≡ script.py ×

    #FROM jupyter/base-notebook
    #FROM cschranz/gpu-jupyter
    FROM jupyter/scipy-notebook
    USER root
    RUN apt-get update && \
         apt-get install -y unzip && \
        apt-get clean && rm -rf var/lib/apt/lists/*
10
12
    USER $NB UID
    RUN pip install \
14
         cvxpy>=1.1.19 \
15
        matplotlib>=3.2.0 \
        numpy>=1.0.0 \
16
        pandas>=0.19 \
17
18
        scikit-learn>=0.24.1 \
         scipy>=1.3.0 \
19
        PyPortfolioOpt
20
21
```

Hactpoйкa docker-compose.yml для конкретного проекта



script.py для дальнейшего использования как сервиса



```
jovyan@a(X
               2-Mean-V X

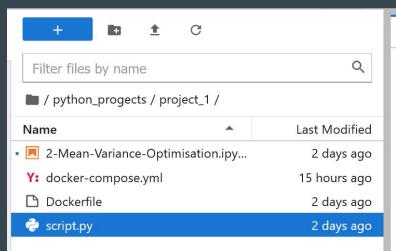
    docker-co X

    ■ Dockerfile X

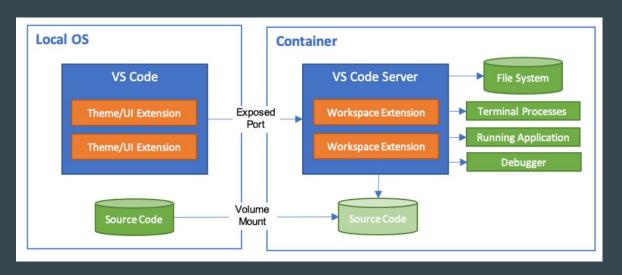
    ≡ script.py ×

    import yfinance as yf
    import matplotlib.pyplot as plt
    import pandas as pd
    import numpy as np
    from pypfopt import expected returns
    tickers = ["MSFT", "AMZN", "KO", "MA", "COST",
                "LUV", "XOM", "PFE", "JPM", "UNH",
                "ACN", "DIS", "GILD", "F", "TSLA"]
 10
11
    ohlc = yf.download(tickers, period="max")
13
    prices = ohlc["Adj Close"].dropna(how="all")
    prices.tail()
    prices[prices.index >= "2008-01-01"].plot(figsize=(15,10))
17
    import pypfopt
    pypfopt. version
 20
    from pypfopt import risk models
```

Запуск script.py

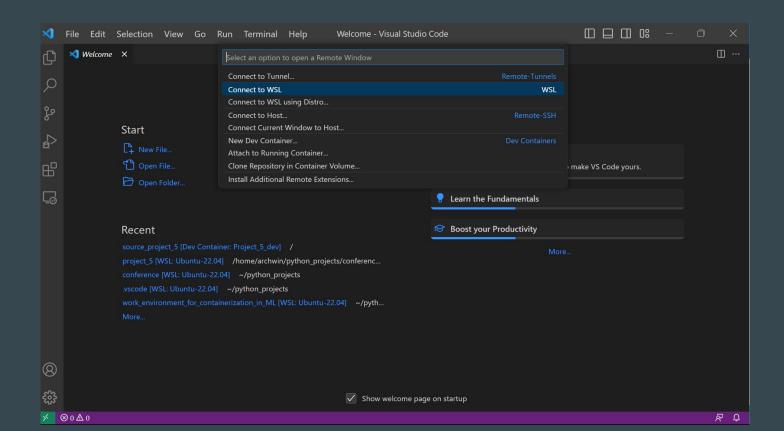


Dev_container B vscode

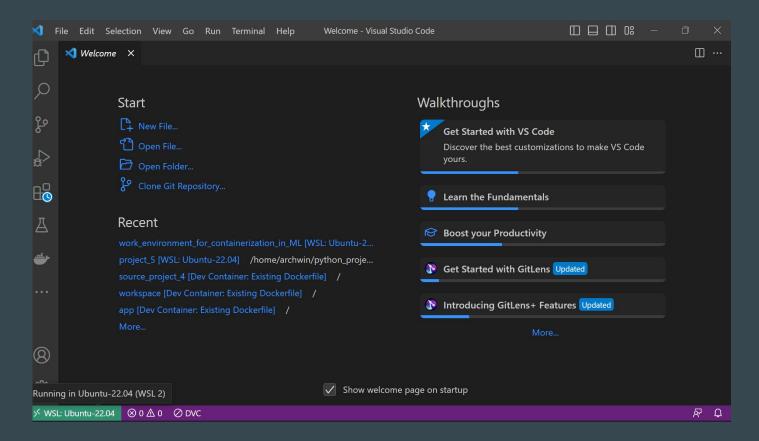


- создаем проект с нуля
- качаем проект с github
- присоединяемся к уже работающему docker контейнеру

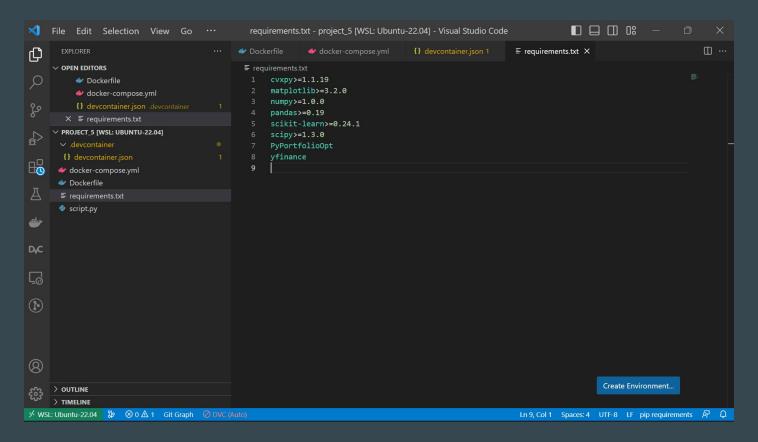
Vscode



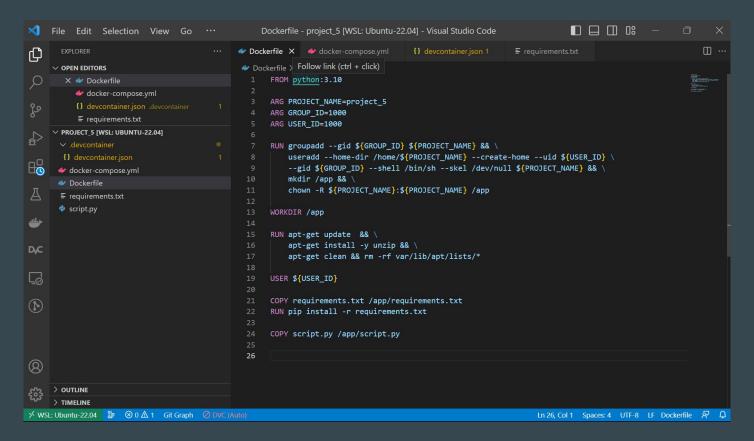
WSL2 B vscode



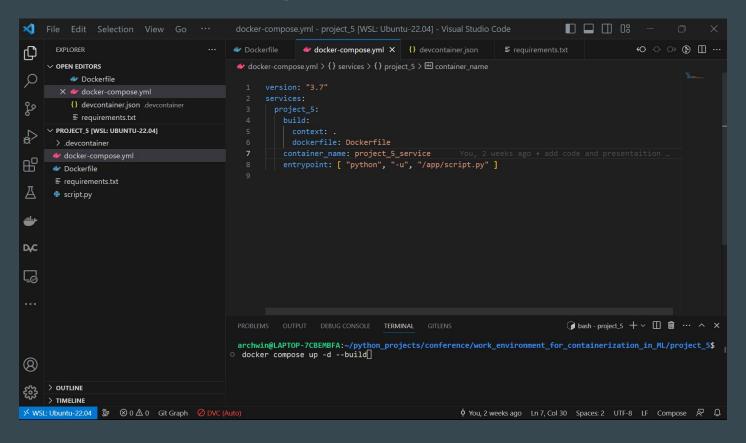
requirements.txt



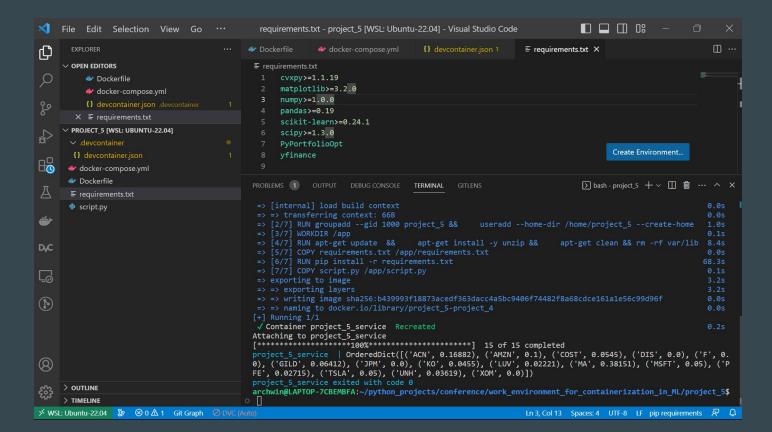
Обновленный Dockerfile



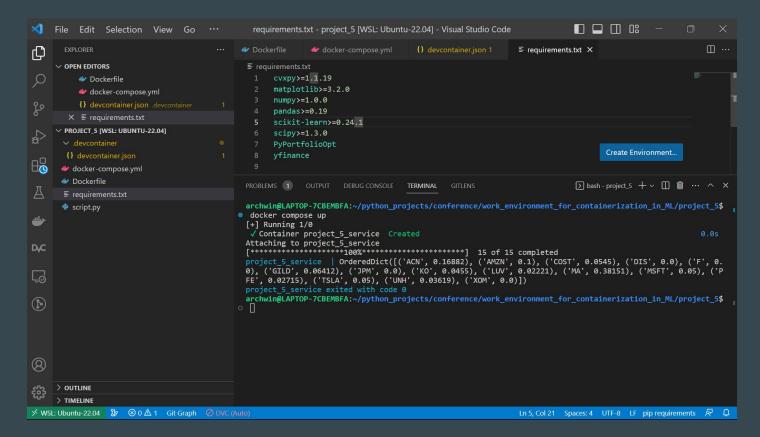
Обновленный docker-compose.yml



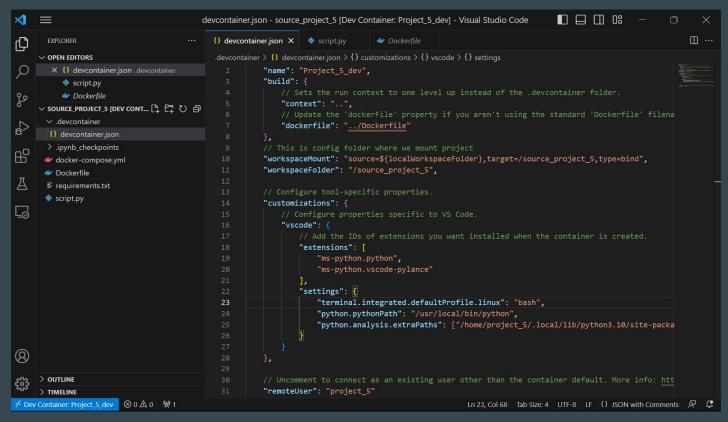
создаем и запускаем сервис через docker-compose up -build



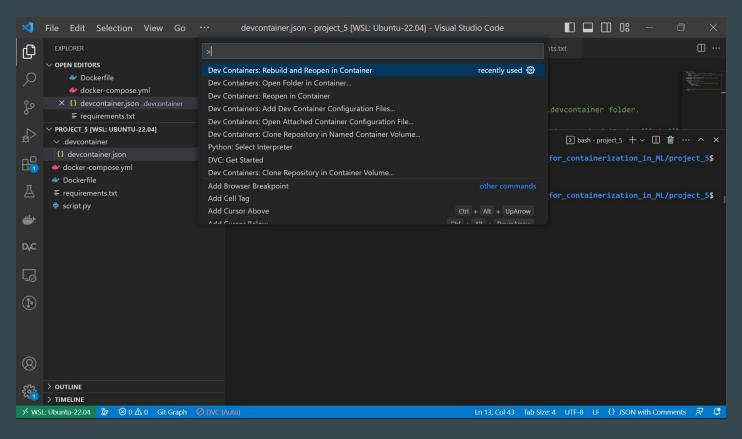
Запускаем сервис без пересоздания



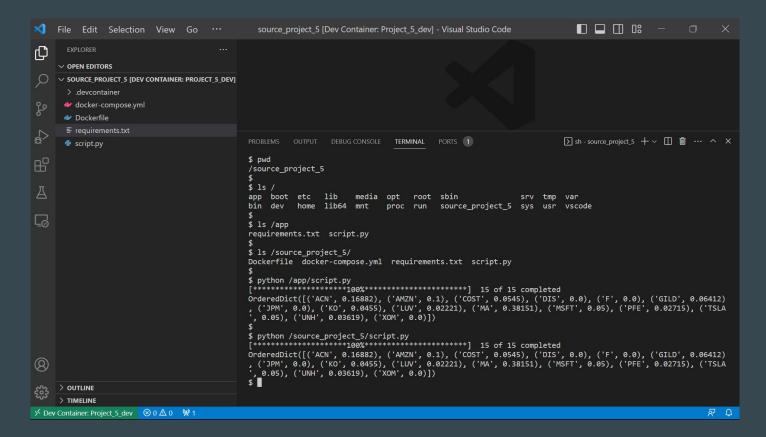
Hастройка devcontainer.json



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
 - <u>https://youtu.be/UXxUcZDSNwA</u>
- Как настроить dev container в vscode
 - https://youtu.be/UP-2vEd3NGw
- Как запустить ноутбук из github в colab
 - <u>https://stackoverflow.com/questions/62596466/how-can-i-run-notebooks-of-a-github-project-in-google-colab</u>
- Какой образ jupyter взять за основу
 - https://jupyter-docker-stacks.readthedocs.io/en/latest/using/selecting.html
- dev container vscode
 - <u>https://code.visualstudio.com/docs/devcontainers/containers</u>

Спасибо за ваше время и внимание.



Павел Вешкин

Контакты:

https://t.me/arch_kras

vpa_kras@rambler.ru