

Контейнеризация рабочей среды для ML

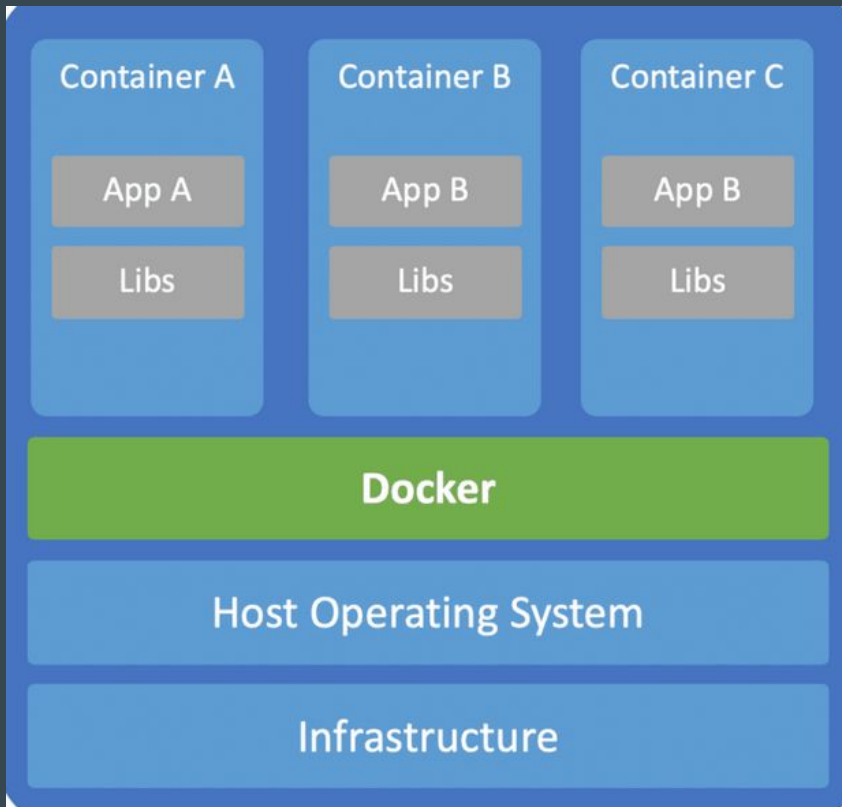
...

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 скрывание 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


 Search or jump to... 








[Pull requests](#) [Issues](#) [Codespaces](#) [Marketplace](#) [Explore](#)

 [robertmartin8](#) / [PyPortfolioOpt](#) Public

[Code](#) [Issues](#) 38 [Pull requests](#) 1 [Actions](#) [Security](#) [Insights](#)

  master [PyPortfolioOpt](#) / [cookbook](#) / 

 [ryanrussell](#) docs(various): Fix typos

Name	Last commit message
 ..	
 data	binder support
 1-RiskReturnModels.ipynb	Fix SageMaker links
 2-Mean-Variance-Optimisation.ipynb	docs(various): Fix typos
 3-Advanced-Mean-Variance-Optimisation.ipynb	Fix SageMaker links
 4-Black-Litterman-Allocation.ipynb	Fix SageMaker links
 5-Hierarchical-Risk-Parity.ipynb	Fix SageMaker links

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

Downloading data

To download data, we will use `yfinance`, needed.

 Open in Colab

 Open in Kaggle

 Run on Gradient

 Open Studio Lab

```
!pip install pandas numpy matplotlib  
import os
```

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

 2-Mean-Variance-Optimisation.ipynb

Файл Изменить Вид Вставка Среда выполнения Инструменты Справка


Поделиться ⚙️ Р


+ Код + Текст Копировать на Диск


Подключиться



Downloading data

To download data, we will use `yfinance`, an excellent library that provides free price data from Yahoo Finance, no API key needed.

 Open in Colab

 Open in Kaggle

 Run on Gradient

 Open  Studio Lab

```
[ ] !pip install pandas numpy matplotlib yfinance PyPortfolioOpt
import os
if not os.path.isdir('data'):
    os.system('git clone https://github.com/robertmartin8/PyPortfolioOpt.git')
    os.chdir('PyPortfolioOpt/cookbook')
```

Requirement already satisfied: pandas in /usr/local/lib/python3.7/dist-packages (1.3.5)
Requirement already satisfied: numpy in /usr/local/lib/python3.7/dist-packages (1.19.5)
Requirement already satisfied: matplotlib in /usr/local/lib/python3.7/dist-packages (3.2.2)

ноутбук в Kaggle kernels

notebook8f1bce3214

File Edit View Run Add-ons Help

Share Save Version 0

Download data

To download data, we will use `yfinance`, an excellent library that provides free price data from Yahoo Finance, no API key needed.

Open in Colab

Open in Kaggle

Run on Gradient

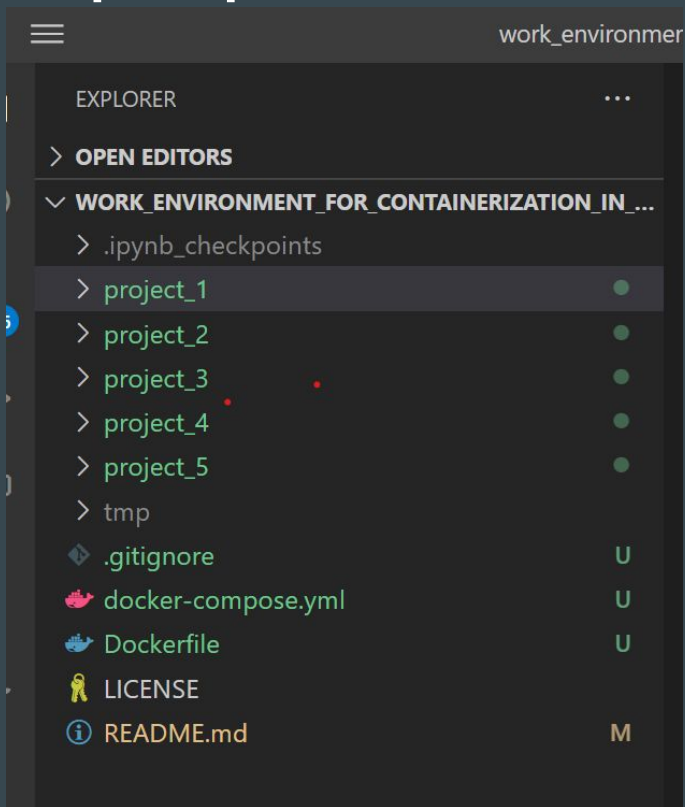
Open Studio Lab

```
!pip install pandas numpy matplotlib yfinance PyPortfolioOpt
import os
if not os.path.isdir('data'):
    os.system('git clone https://github.com/robertmartin8/PyPortfolioOpt.git')
    os.chdir('PyPortfolioOpt/cookbook')
```

Requirement already satisfied: pandas in /opt/conda/lib/python3.7/site-packages (1.3.5)
Requirement already satisfied: numpy in /opt/conda/lib/python3.7/site-packages (1.21.6)
Requirement already satisfied: matplotlib in /opt/conda/lib/python3.7/site-packages (3.5.3)
WARNING: Retrying (Retry(total=4, connect=None, read=None, redirect=None, status=None)) after connection broken by 'NewConnectionError(<pip._vendor.urllib3.connection.HTTPSConnection object at 0x7e0f4d275750>: Failed to establish a new connection: [Errno -3] Temporary failure in name resolution)': /simple/yfinance/
WARNING: Retrying (Retry(total=3, connect=None, read=None, redirect=None, status=None)) after connection broken by 'NewConnectionError(<pip._vendor.urllib3.connection.HTTPSConnection object at 0x7e0f4d275750>: Failed to establish a new connection: [Errno -3] Temporary failure in name resolution)': /simple/yfinance/
WARNING: Retrying (Retry(total=2, connect=None, read=None, redirect=None, status=None)) after connection broken by 'NewConnectionError(<pip._vendor.urllib3.connection.HTTPSConnection object at 0x7e0f4d28fed0>: Failed to establish a new connection: [Errno -3] Temporary failure in name resolution)': /simple/yfinance/

+ Code + Markdown

Пример



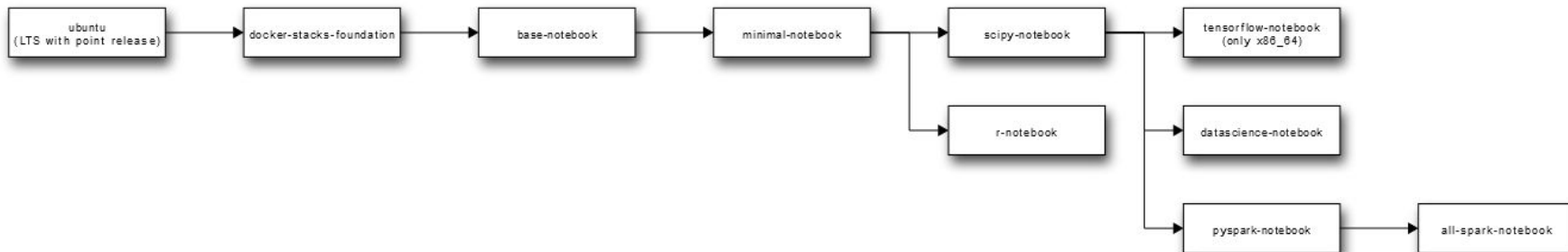
Jupyter lab

- 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.



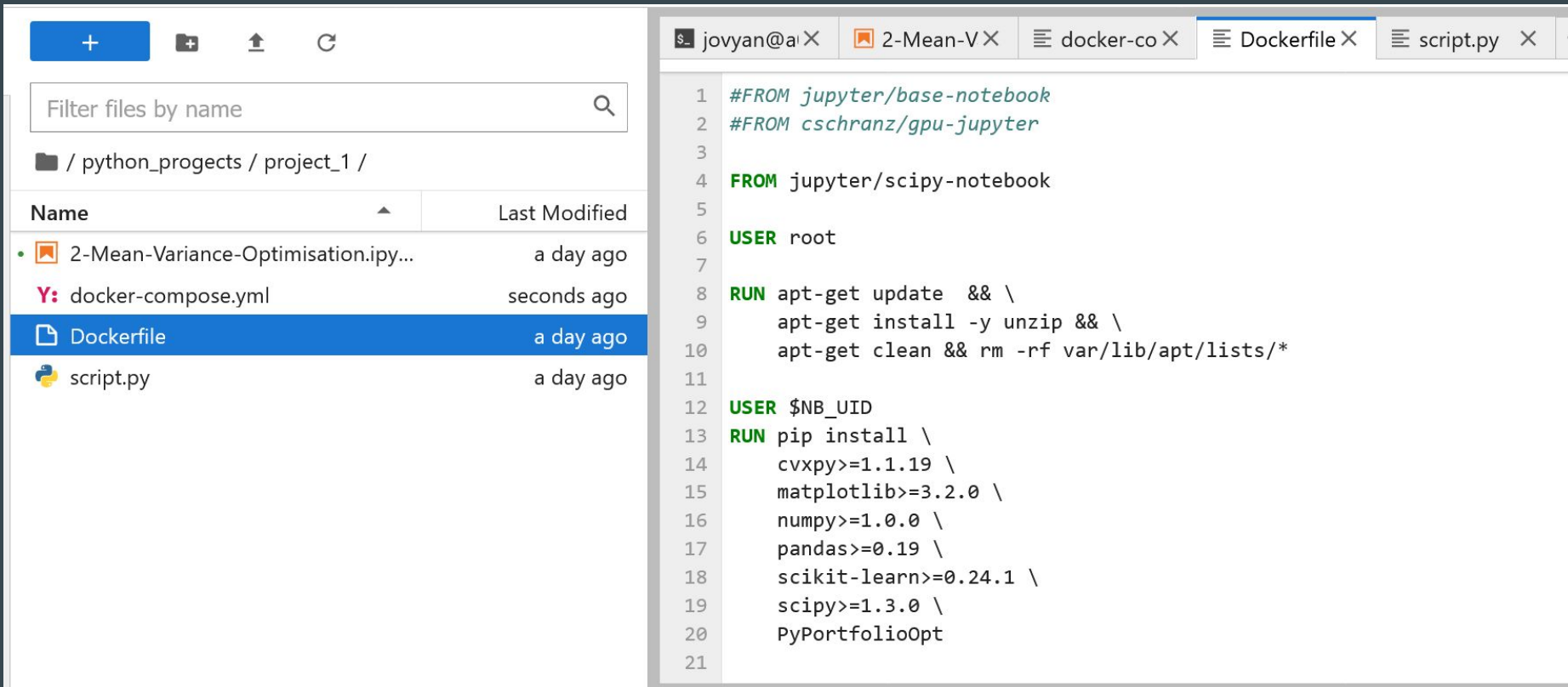
Настройка Dockerfile для всех проектов

```
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

```
1 version: "3.7"
2 services:
3   jupyter_lab_all:
4     build:
5       context: .
6       dockerfile: Dockerfile
7     user: root
8     environment:
9       - JUPYTER_TOKEN=token
10      - GRANT_SUDO=yes
11     volumes:
12       # this is you folder for projects mount into container folder
13       - ./home/jovyan/python_projects
14     ports:
15       - 8880:8888
16     container_name: jupyter_lab_for_all_projects
17
```

Настройка Dockerfile для конкретного проекта



The image shows a code editor interface with a file explorer on the left and a code editor on the right.

File Explorer (Left):

- Search bar: Filter files by name
- Path: / python_projects / project_1 /
- Table of files:

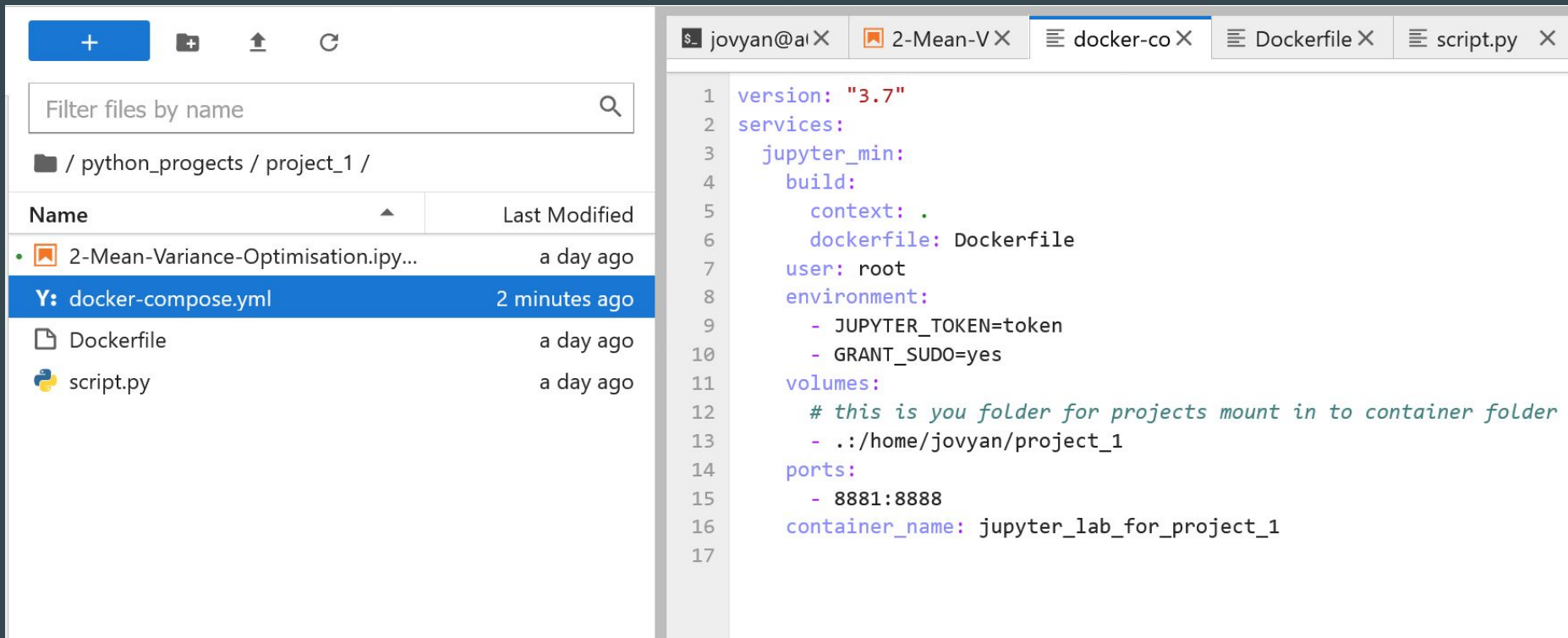
Name	Last Modified
2-Mean-Variance-Optimisation.ipynb	a day ago
Y: docker-compose.yml	seconds ago
Dockerfile	a day ago
script.py	a day ago

Code Editor (Right):

Tab: Dockerfile X

```
1 #FROM jupyter/base-notebook
2 #FROM cschranz/gpu-jupyter
3
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
21
```


Настройка docker-compose.yml для конкретного проекта



The image shows a code editor interface with a file explorer on the left and a code editor on the right.

File Explorer (Left):

- Search bar: Filter files by name
- Path: / python_projects / project_1 /
- Table of files:

Name	Last Modified
2-Mean-Variance-Optimisation.ipynb	a day ago
Y: docker-compose.yml	2 minutes ago
Dockerfile	a day ago
script.py	a day ago

Code Editor (Right):

Tab: docker-co X

```
1 version: "3.7"
2 services:
3   jupyter_min:
4     build:
5       context: .
6       dockerfile: Dockerfile
7     user: root
8     environment:
9       - JUPYTER_TOKEN=token
10      - GRANT_SUDO=yes
11     volumes:
12       # this is you folder for projects mount in to container folder
13       - ../home/jovyan/project_1
14     ports:
15       - 8881:8888
16     container_name: jupyter_lab_for_project_1
17
```

script.py для дальнейшего использования как сервиса

Filter files by name

/ python_projects / project_1 /

Name	Last Modified
2-Mean-Variance-Optimisation.ipynb	a day ago
Y: docker-compose.yml	3 minutes ago
Dockerfile	a day ago
script.py	a day ago

jovyan@al

2-Mean-V

docker-co

Dockerfile

script.py

```
1 import yfinance as yf
2 import matplotlib.pyplot as plt
3 import pandas as pd
4 import numpy as np
5
6 from pypfopt import expected_returns
7
8 tickers = ["MSFT", "AMZN", "KO", "MA", "COST",
9            "LUV", "XOM", "PFE", "JPM", "UNH",
10           "ACN", "DIS", "GILD", "F", "TSLA"]
11
12 ohlc = yf.download(tickers, period="max")
13
14 prices = ohlc["Adj Close"].dropna(how="all")
15 prices.tail()
16 prices[prices.index >= "2008-01-01"].plot(figsize=(15,10))
17
18 import pypfopt
19 pypfopt.__version__
20
21 from pypfopt import risk_models
```

Зануек script.py

Filter files by name

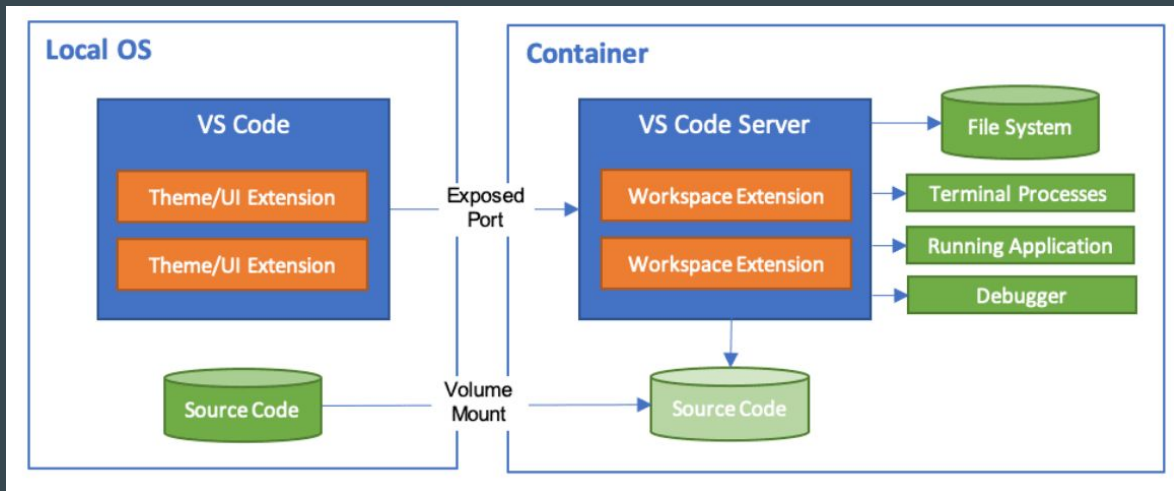
/ python_projects / project_1 /

Name	Last Modified
• 2-Mean-Variance-Optimisation.ipynb	2 days ago
Y: docker-compose.yml	15 hours ago
Dockerfile	2 days ago
script.py	2 days ago

jovyan@a0cfc6104999 2-Mean-V docker-co Dockerfile script.py

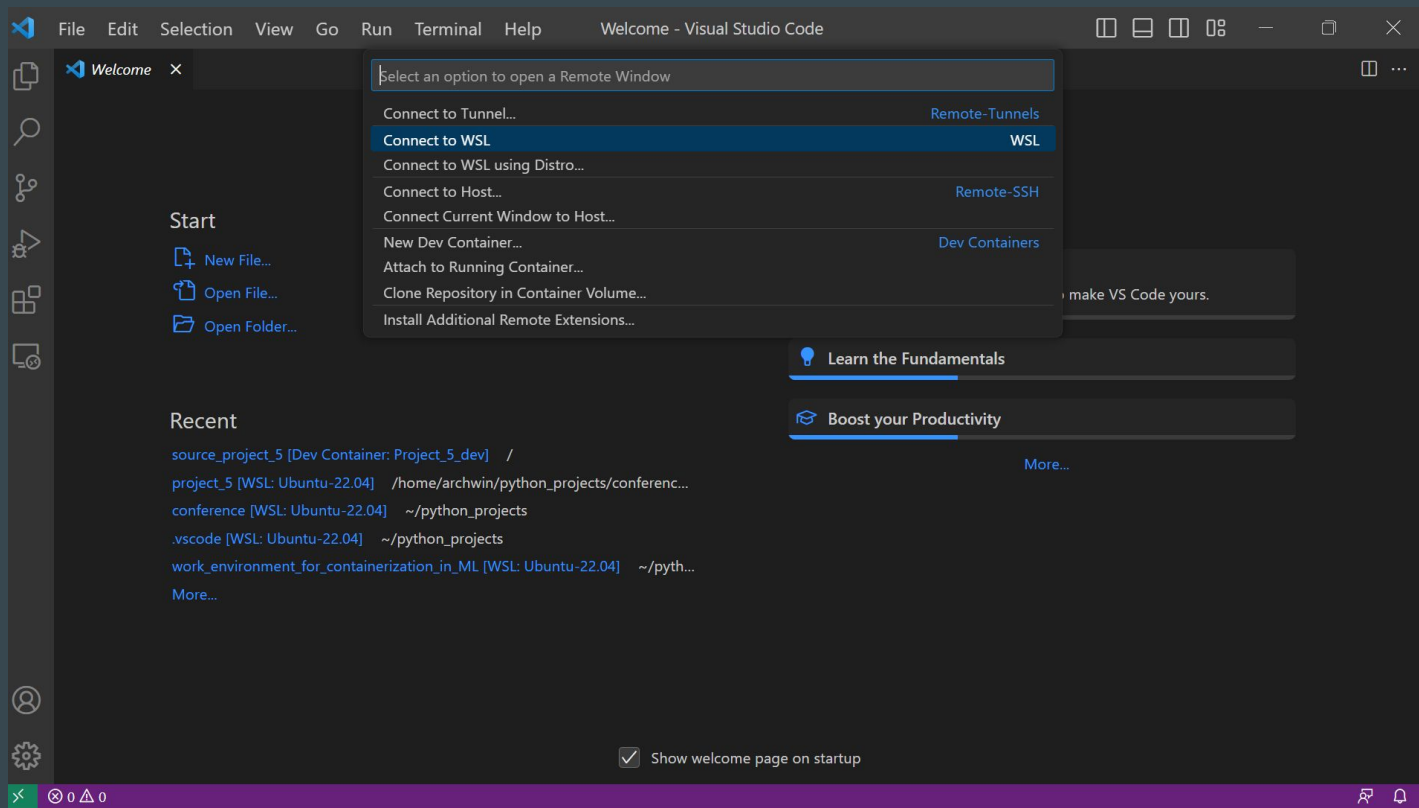
```
(base) jovyan@a0cfc6104999:~/python_projects/project_1$ python script.py
[*****100%*****] 15 of 15 completed
OrderedDict([('ACN', 0.16882), ('AMZN', 0.1), ('COST', 0.0545), ('DIS', 0.0),
('F', 0.0), ('GILD', 0.06412), ('JPM', 0.0), ('KO', 0.0455), ('LUV', 0.022
21), ('MA', 0.38151), ('MSFT', 0.05), ('PFE', 0.02715), ('TSLA', 0.05), ('UN
H', 0.03619), ('XOM', 0.0)])
(base) jovyan@a0cfc6104999:~/python_projects/project_1$
```

Dev_container в vscode

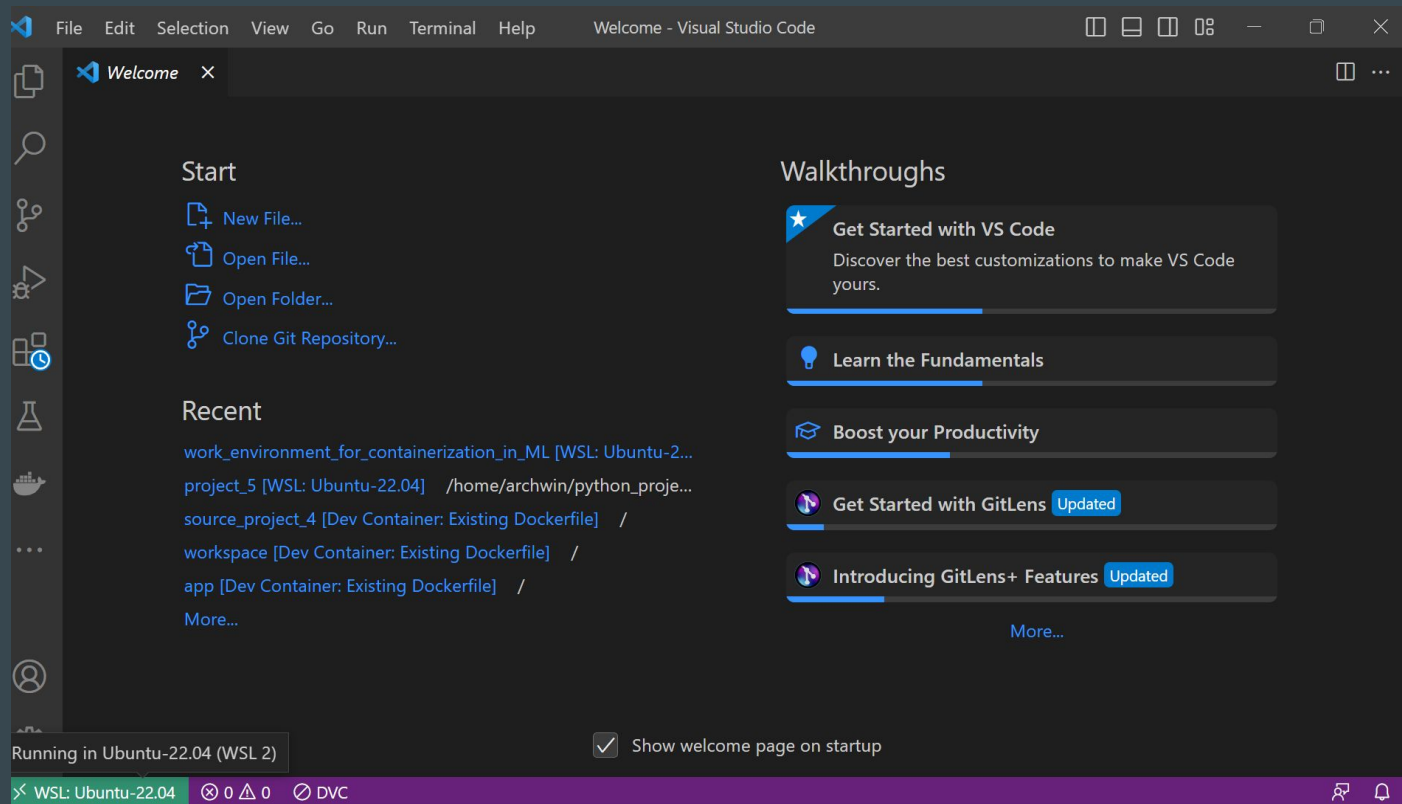


- создаем проект с нуля
- качаем проект с github
- присоединяемся к уже работающему docker контейнеру

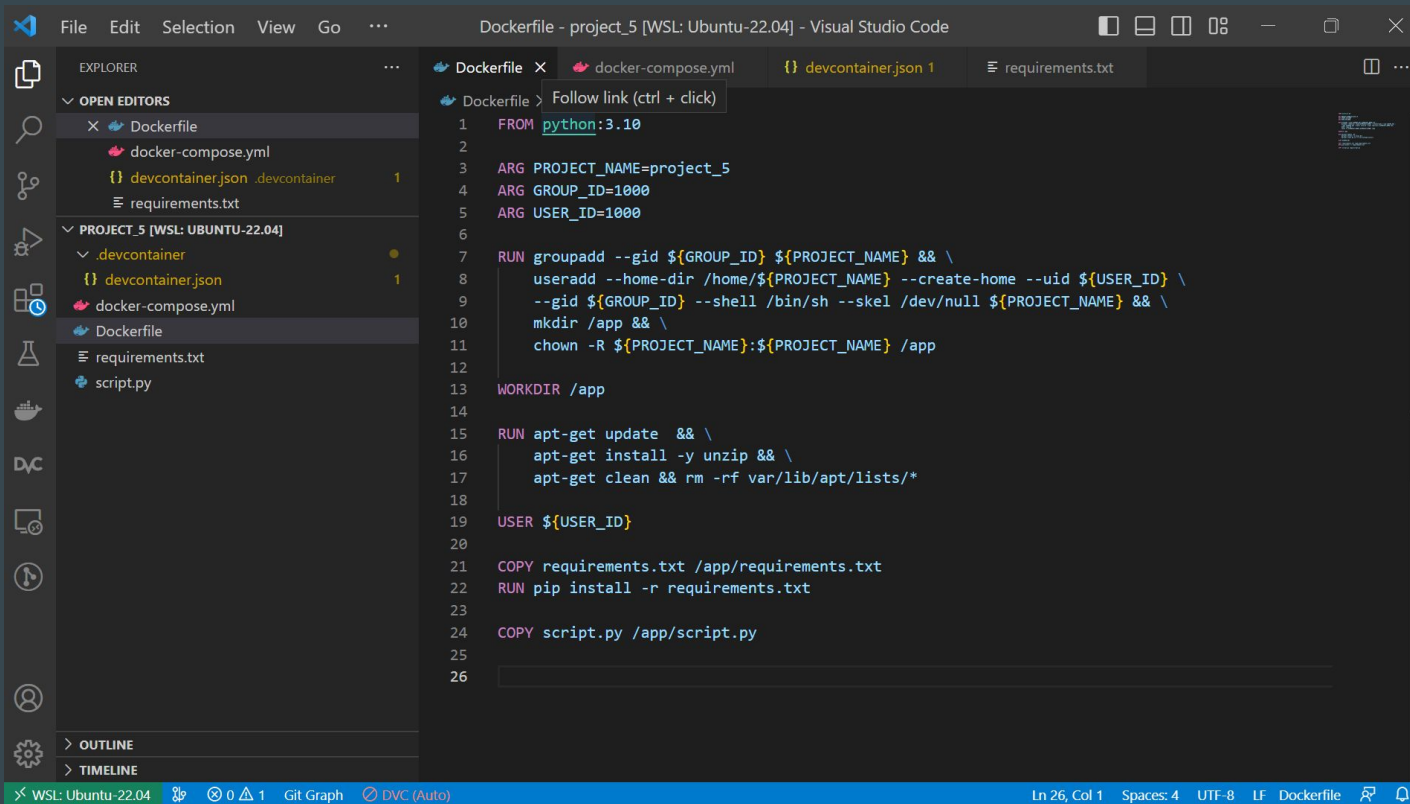
Vscode



WSL2 в vscode

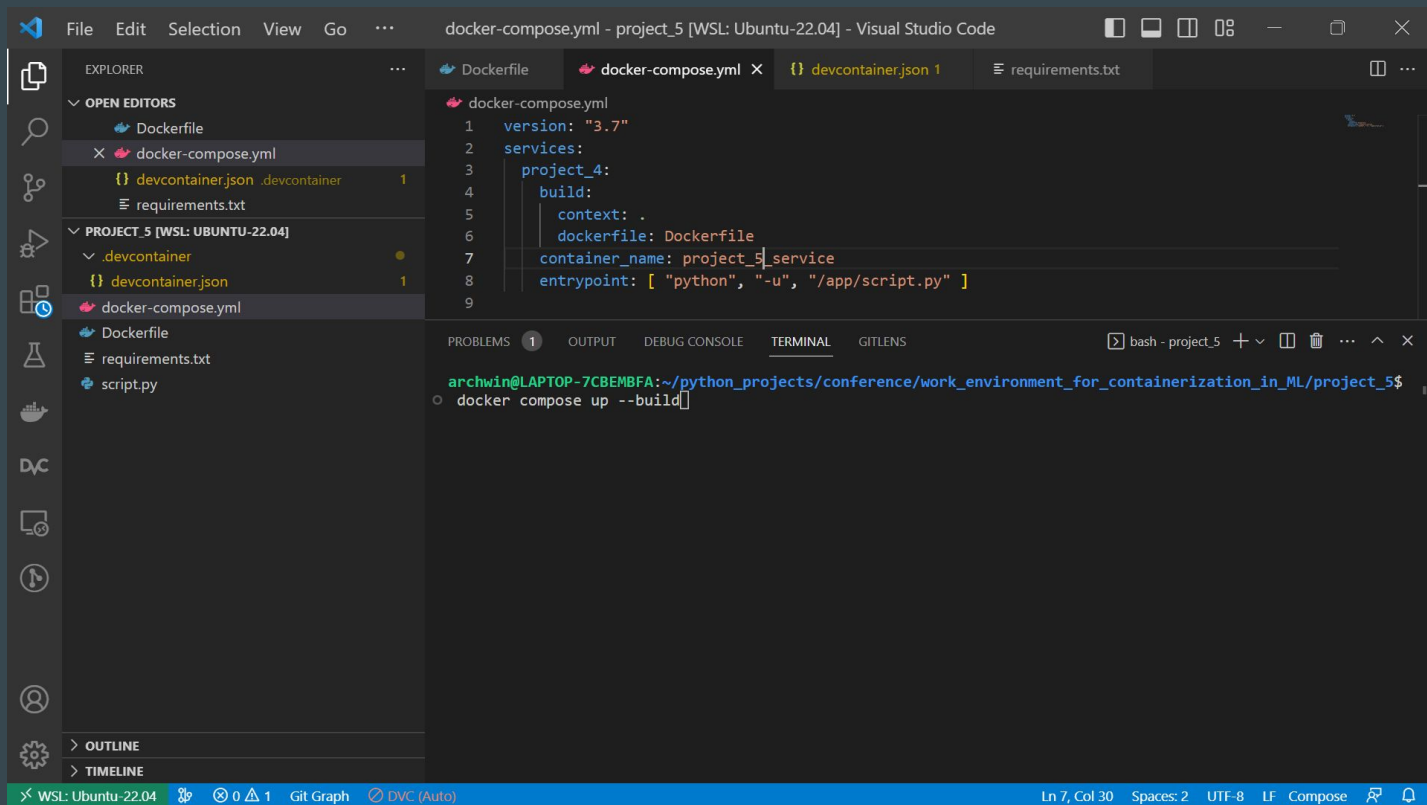


Обновленный Dockerfile

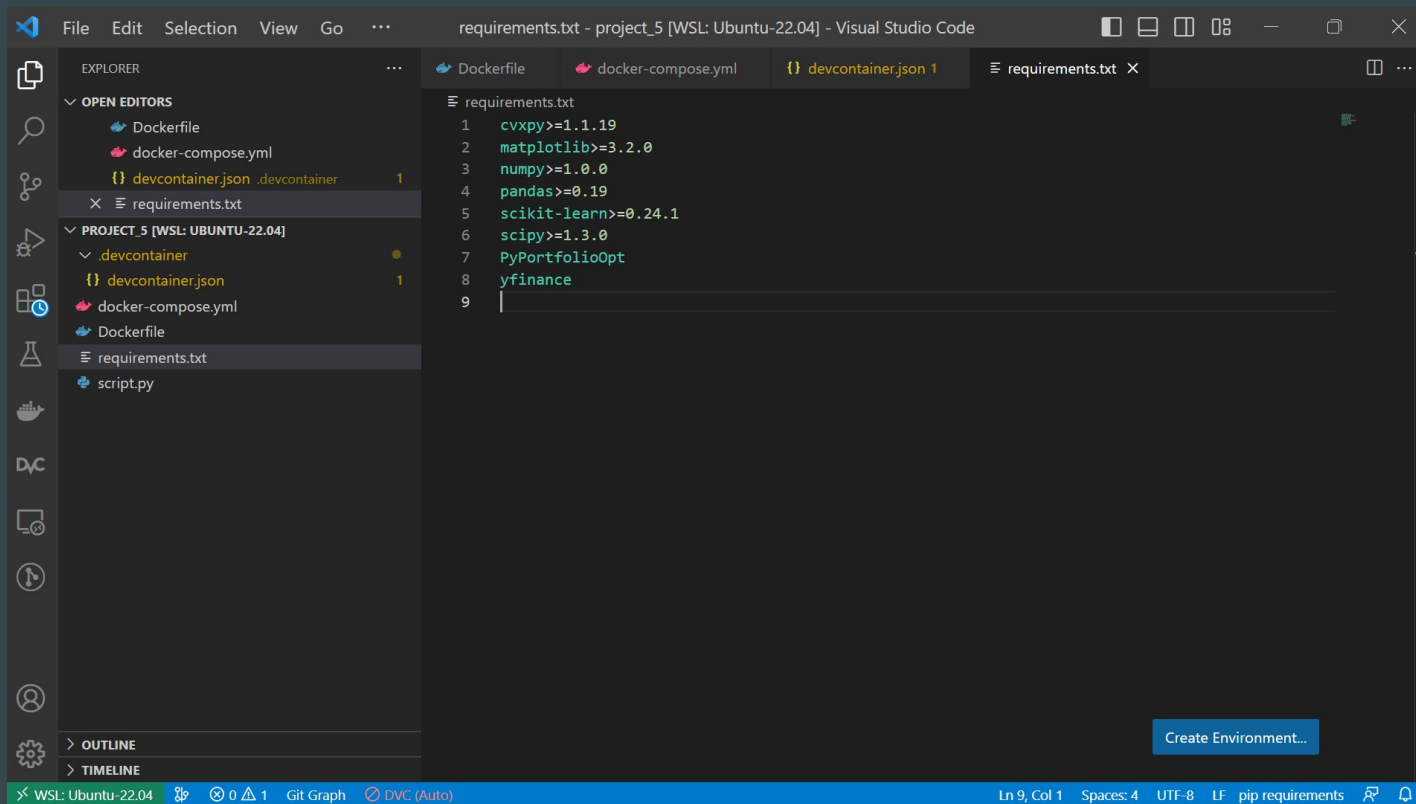


```
1 FROM python:3.10
2
3 ARG PROJECT_NAME=project_5
4 ARG GROUP_ID=1000
5 ARG USER_ID=1000
6
7 RUN groupadd --gid ${GROUP_ID} ${PROJECT_NAME} && \
8     useradd --home-dir /home/${PROJECT_NAME} --create-home --uid ${USER_ID} \
9         --gid ${GROUP_ID} --shell /bin/sh --skel /dev/null ${PROJECT_NAME} && \
10     mkdir /app && \
11     chown -R ${PROJECT_NAME}:${PROJECT_NAME} /app
12
13 WORKDIR /app
14
15 RUN apt-get update && \
16     apt-get install -y unzip && \
17     apt-get clean && rm -rf var/lib/apt/lists/*
18
19 USER ${USER_ID}
20
21 COPY requirements.txt /app/requirements.txt
22 RUN pip install -r requirements.txt
23
24 COPY script.py /app/script.py
25
26
```

Обновленный docker-compose.yml



requirements.txt

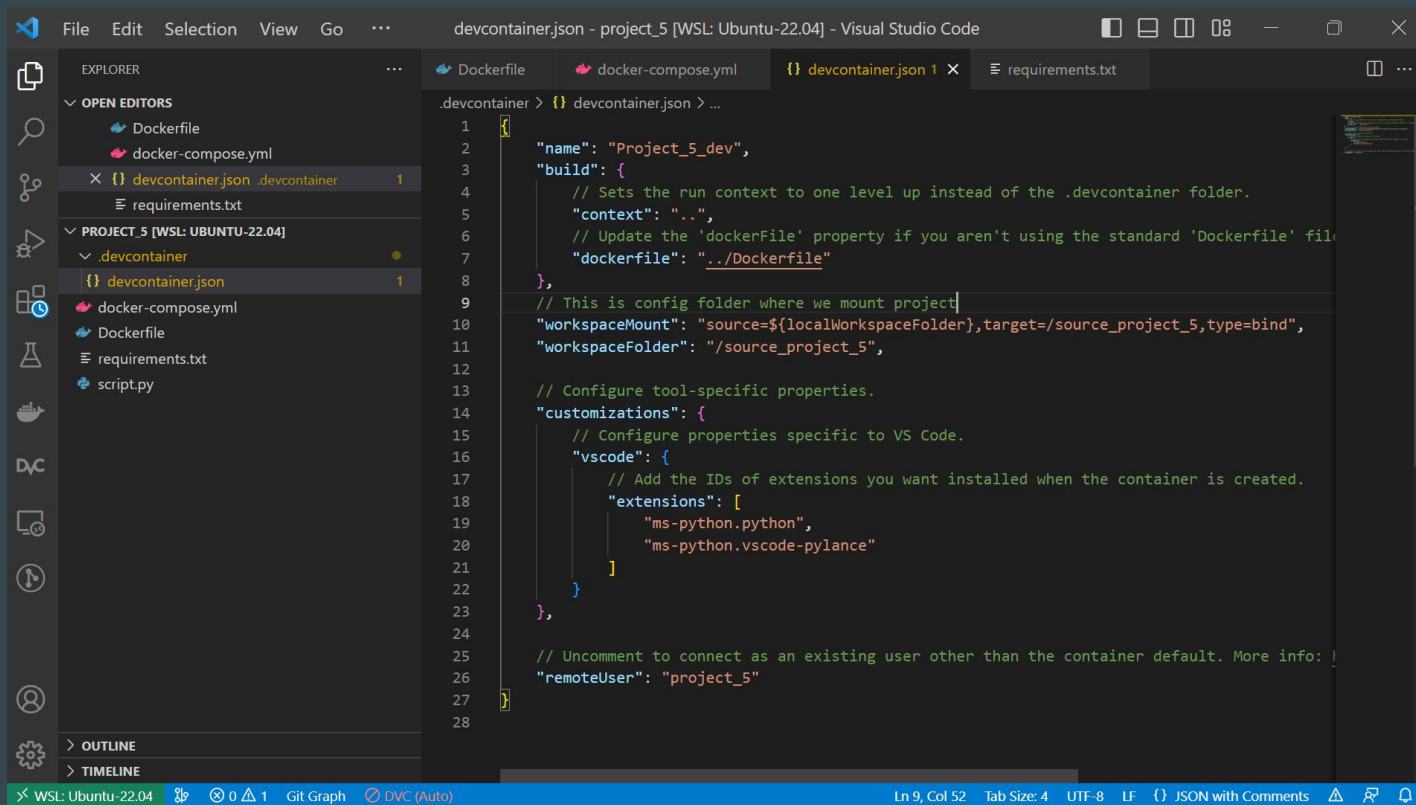


The screenshot shows the Visual Studio Code interface with a project named 'project_5' open. The Explorer sidebar on the left shows the file structure, including 'requirements.txt' under the 'PROJECT_5' folder. The main editor area displays the contents of 'requirements.txt'.

```
requirements.txt
1 cvxpy>=1.1.19
2 matplotlib>=3.2.0
3 numpy>=1.0.0
4 pandas>=0.19
5 scikit-learn>=0.24.1
6 scipy>=1.3.0
7 PyPortfolioOpt
8 yfinance
9
```

At the bottom right of the editor, there is a button labeled 'Create Environment...'. The status bar at the bottom indicates the current environment is 'WSL: Ubuntu-22.04' and the file encoding is 'UTF-8'.

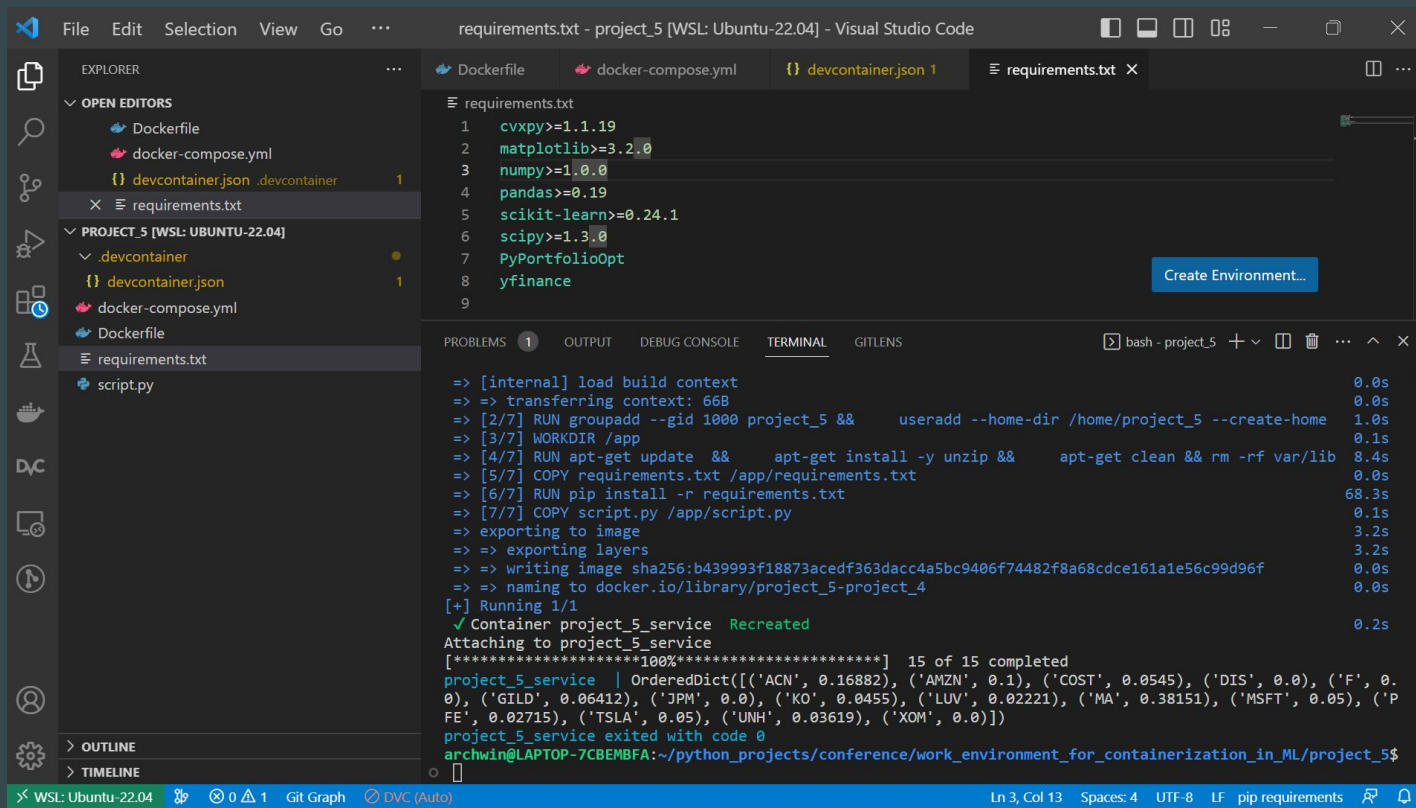
Настройка devcontainer.json



```
1 {
2   "name": "Project_5_dev",
3   "build": {
4     // Sets the run context to one level up instead of the .devcontainer folder.
5     "context": "..",
6     // Update the 'dockerFile' property if you aren't using the standard 'Dockerfile' file
7     "dockerfile": "../Dockerfile"
8   },
9   // This is config folder where we mount project
10  "workspaceMount": "source=${localWorkspaceFolder},target=/source_project_5,type=bind",
11  "workspaceFolder": "/source_project_5",
12
13  // Configure tool-specific properties.
14  "customizations": {
15    // Configure properties specific to VS Code.
16    "vscode": {
17      // Add the IDs of extensions you want installed when the container is created.
18      "extensions": [
19        "ms-python.python",
20        "ms-python.vscode-pylance"
21      ]
22    }
23  },
24
25  // Uncomment to connect as an existing user other than the container default. More info: !
26  "remoteUser": "project_5"
27 }
```

WSL: Ubuntu-22.04 | Git Graph | DVC (Auto) | Ln 9, Col 52 | Tab Size: 4 | UTF-8 | LF | JSON with Comments

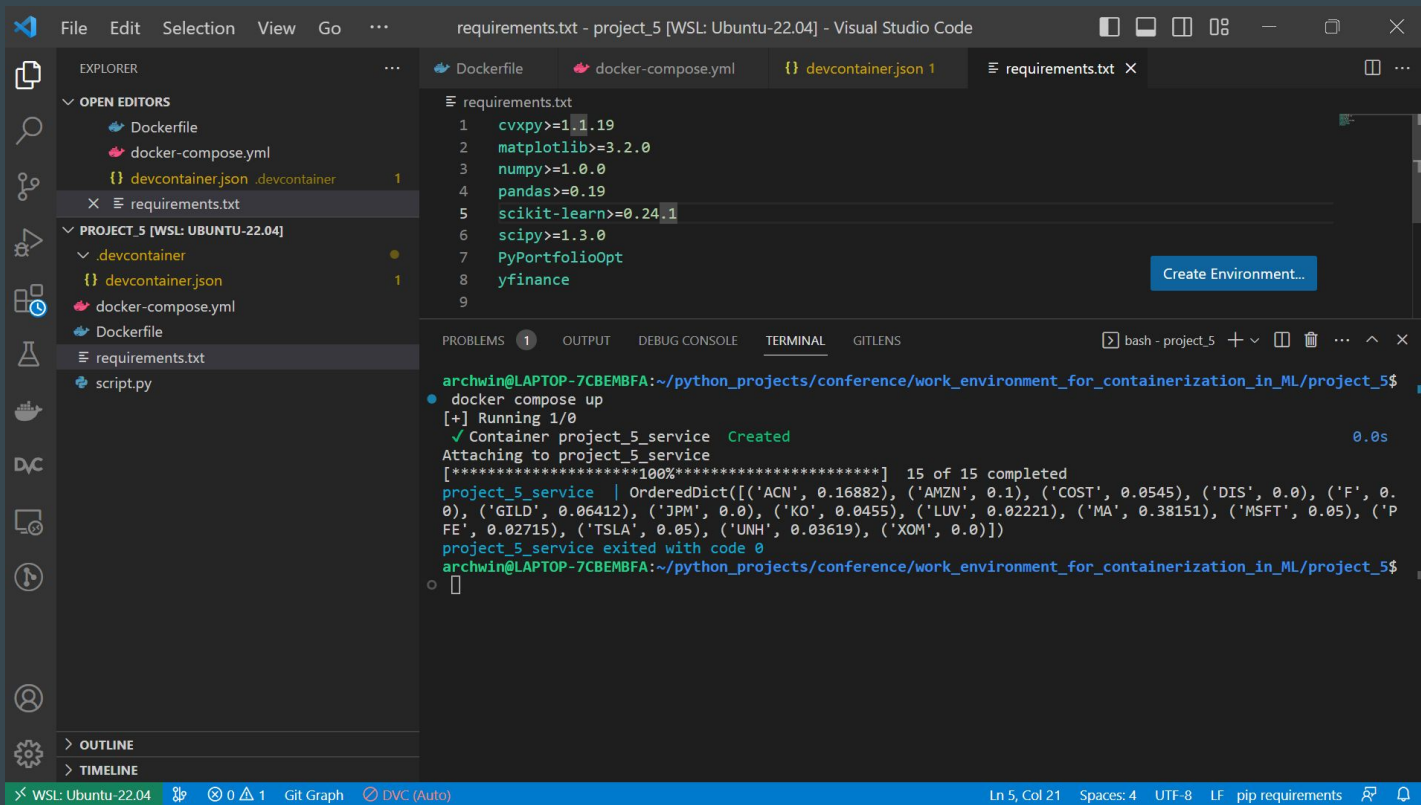
Создаем и запускаем сервис через docker-compose up --build



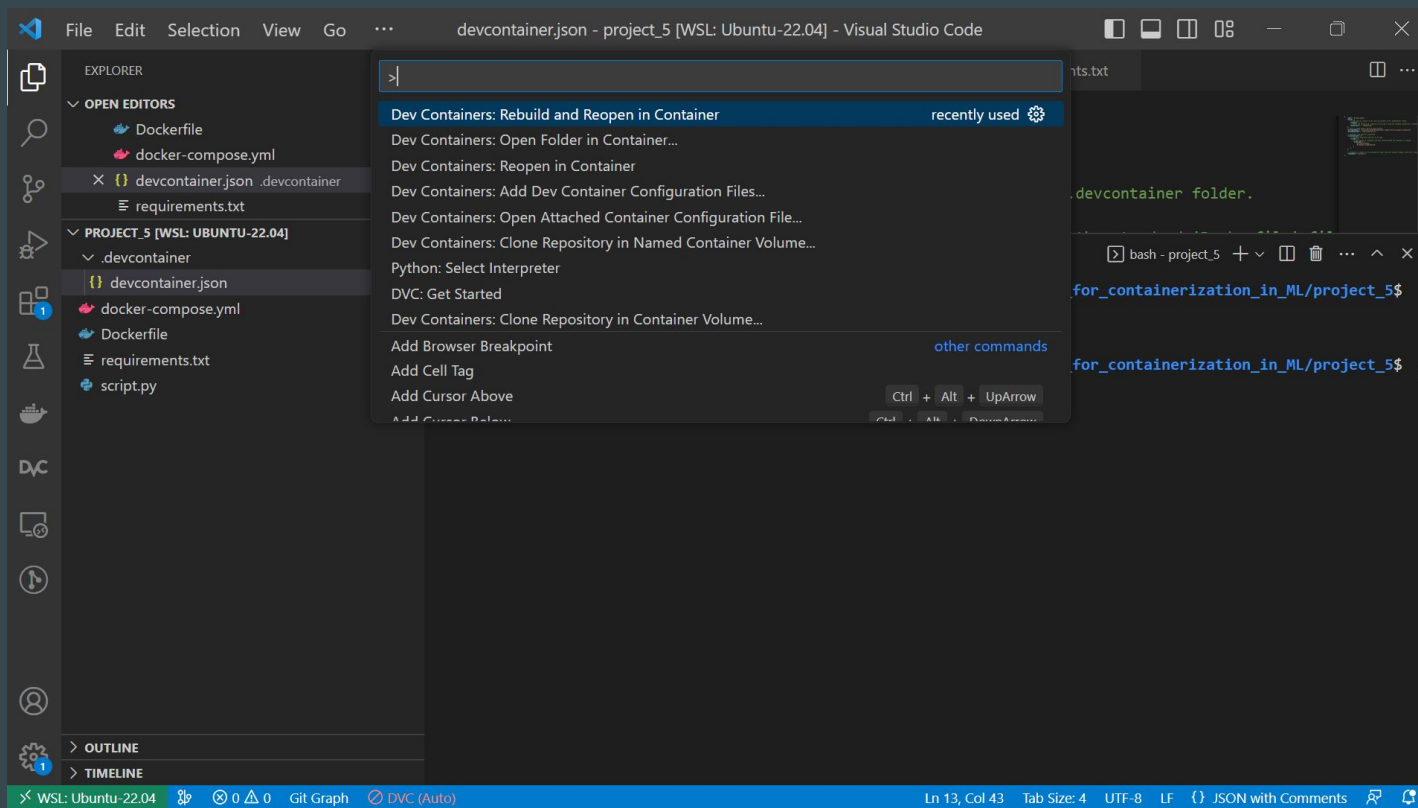
```
requirements.txt
1 cvxpy>=1.1.19
2 matplotlib>=3.2.0
3 numpy>=1.0.0
4 pandas>=0.19
5 scikit-learn>=0.24.1
6 scipy>=1.3.0
7 PyPortfolioOpt
8 yfinance
```

```
bash - project_5
=> [internal] load build context 0.0s
=> => transferring context: 66B 0.0s
=> [2/7] RUN groupadd --gid 1000 project_5 && useradd --home-dir /home/project_5 --create-home 1.0s
=> [3/7] WORKDIR /app 0.1s
=> [4/7] RUN apt-get update && apt-get install -y unzip && apt-get clean && rm -rf var/lib 8.4s
=> [5/7] COPY requirements.txt /app/requirements.txt 0.0s
=> [6/7] RUN pip install -r requirements.txt 68.3s
=> [7/7] COPY script.py /app/script.py 0.1s
=> exporting to image 3.2s
=> => exporting layers 3.2s
=> => writing image sha256:b439993f18873acedf363dacc4a5bc9406f74482f8a68cdce161a1e56c99d96f 0.0s
=> => naming to docker.io/library/project_5-project_4 0.0s
[+] Running 1/1
✔ Container project_5_service Recreated 0.2s
Attaching to project_5_service
[*****100%*****] 15 of 15 completed
project_5_service | OrderedDict([('ACN', 0.16882), ('AMZN', 0.1), ('COST', 0.0545), ('DIS', 0.0), ('F', 0.0), ('GILD', 0.06412), ('JPM', 0.0), ('KO', 0.0455), ('LUV', 0.02221), ('MA', 0.38151), ('MSFT', 0.05), ('PFE', 0.02715), ('TSLA', 0.05), ('UNH', 0.03619), ('XOM', 0.0)])
project_5_service exited with code 0
archwin@LAPTOP-7CBEMBFA:~/python_projects/conference/work_environment_for_containerization_in_ML/project_5$
```

Запускаем сервис без пересоздания



Запускаем и заходим в dev container



Внутри dev container

```
File Edit Selection View Go ... source_project_5 [Dev Container: Project_5_dev] - Visual Studio Code

EXPLORER
OPEN EDITORS
SOURCE_PROJECT_5 [DEV CONTAINER: PROJECT_5_DEV]
  .devcontainer
  docker-compose.yml
  Dockerfile
  requirements.txt
  script.py

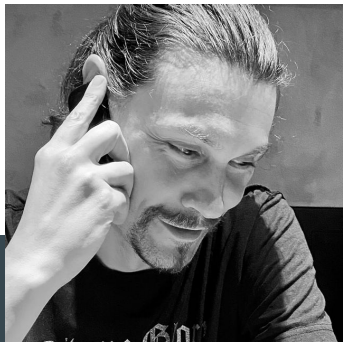
PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS 1
sh - source_project_5

$ pwd
/source_project_5
$
$ ls /
app boot etc lib media opt root sbin srv tmp var
bin dev home lib64 mnt proc run source_project_5 sys usr vscode
$
$ ls /app
requirements.txt script.py
$
$ ls /source_project_5/
Dockerfile docker-compose.yml requirements.txt script.py
$
$ python /app/script.py
[*****100%*****] 15 of 15 completed
OrderedDict([('ACN', 0.16882), ('AMZN', 0.1), ('COST', 0.0545), ('DIS', 0.0), ('F', 0.0), ('GILD', 0.06412), ('JPM', 0.0), ('KO', 0.0455), ('LUV', 0.02221), ('MA', 0.38151), ('MSFT', 0.05), ('PFE', 0.02715), ('TSLA', 0.05), ('UNH', 0.03619), ('XOM', 0.0)])
$
$ python /source_project_5/script.py
[*****100%*****] 15 of 15 completed
OrderedDict([('ACN', 0.16882), ('AMZN', 0.1), ('COST', 0.0545), ('DIS', 0.0), ('F', 0.0), ('GILD', 0.06412), ('JPM', 0.0), ('KO', 0.0455), ('LUV', 0.02221), ('MA', 0.38151), ('MSFT', 0.05), ('PFE', 0.02715), ('TSLA', 0.05), ('UNH', 0.03619), ('XOM', 0.0)])
$
```

Материалы и ссылки:

- Репозиторий 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