

# ВЕБ-СКРАПИНГ (СБОР ДАННЫХ)

## Лабораторная работа

### Основная цель

В разработанной вами игре (угадай слово) очень скучный список предметов (блюд). Ваша цель — получить информацию из другого источника и добавить её на свой сайт. На этот раз мы берём информацию со страницы местных рецептов с большим количеством разнообразных блюд. Ручной сбор данных может занять месяцы, поэтому мы воспользуемся методом, называемым веб-скрапингом. В этой части мы подготовим все необходимые элементы перед получением информации с других сайтов (это мы сделаем в следующей лабораторной работе).

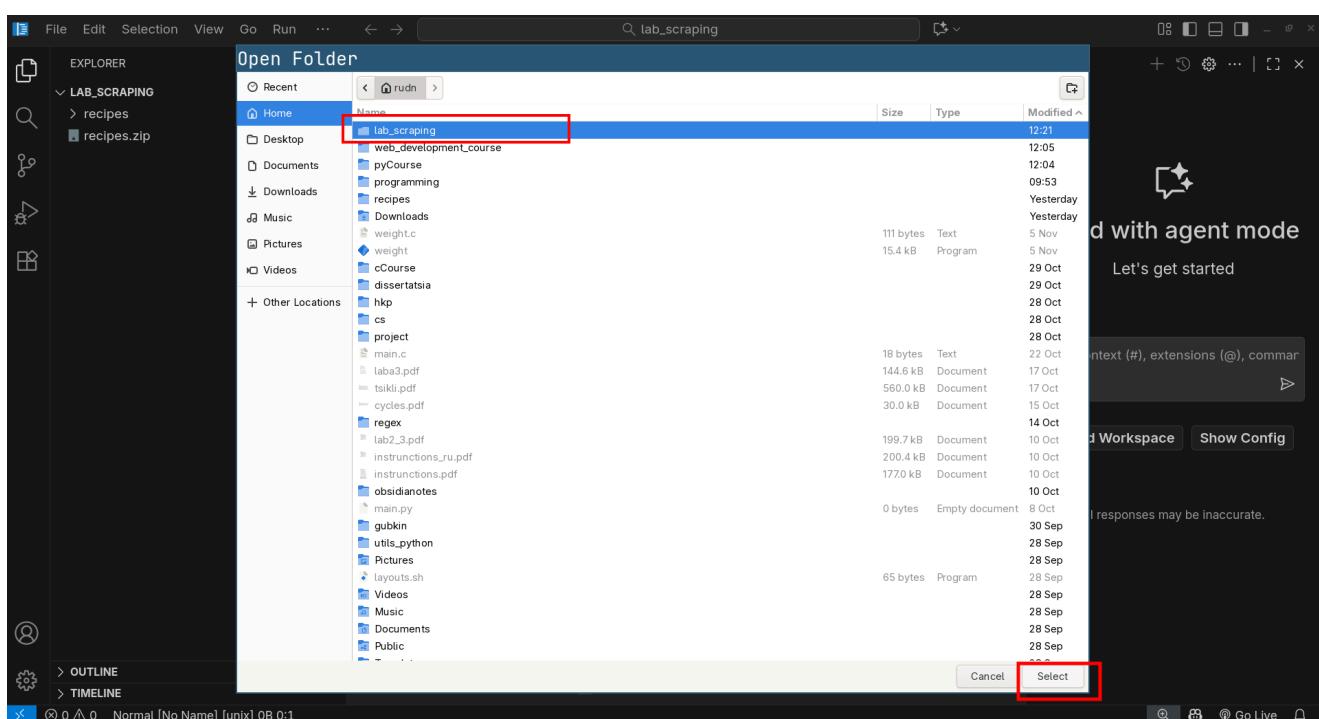
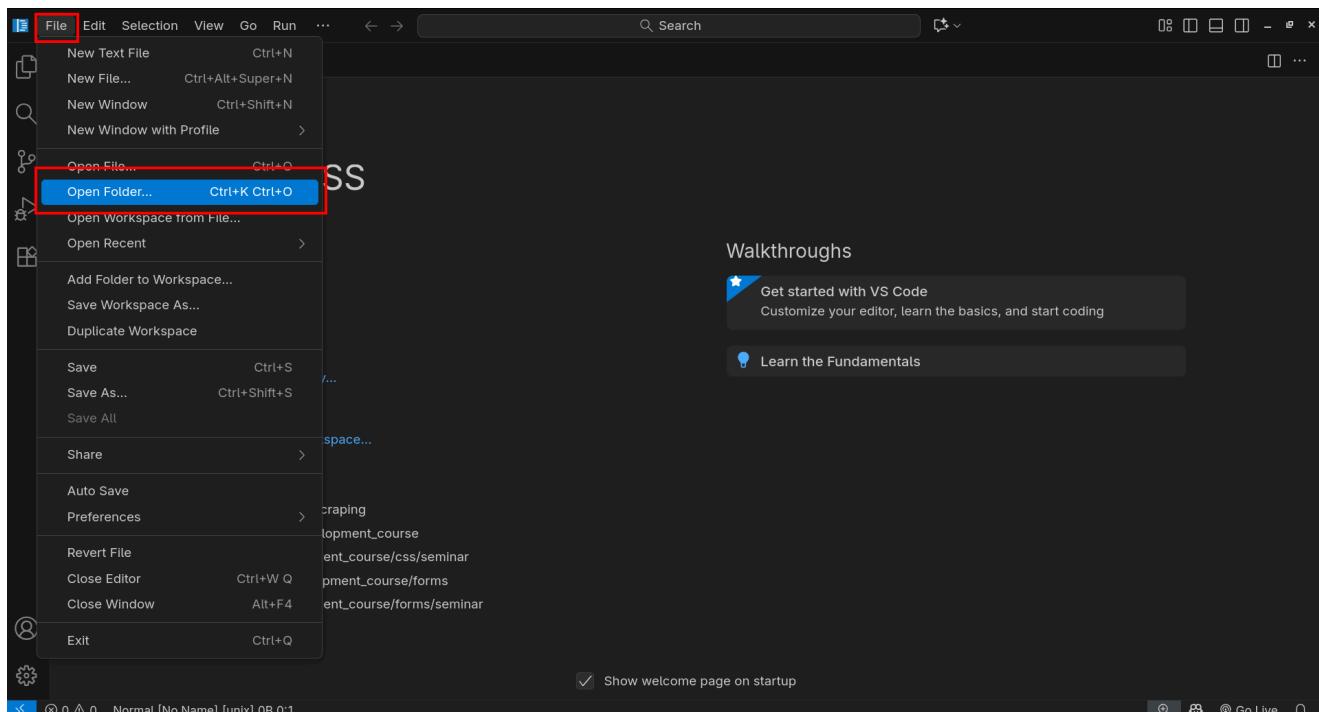
### Важная информация перед началом работы

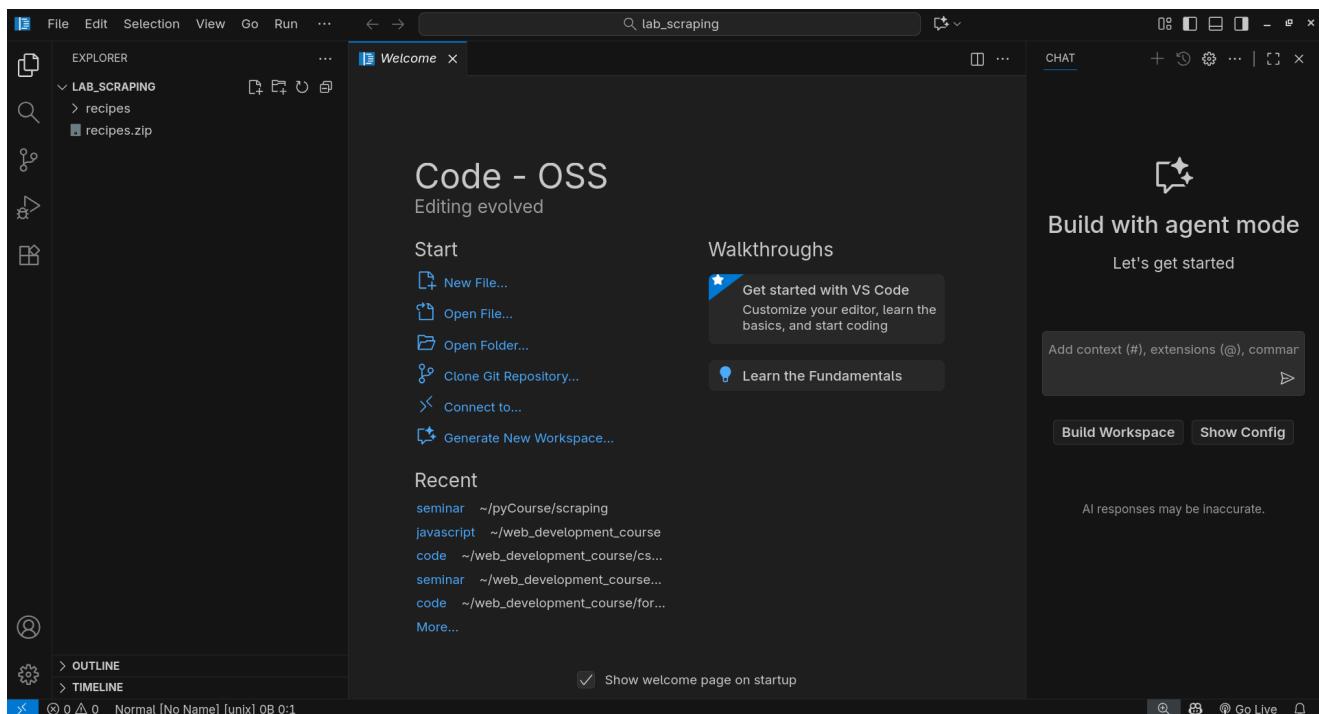
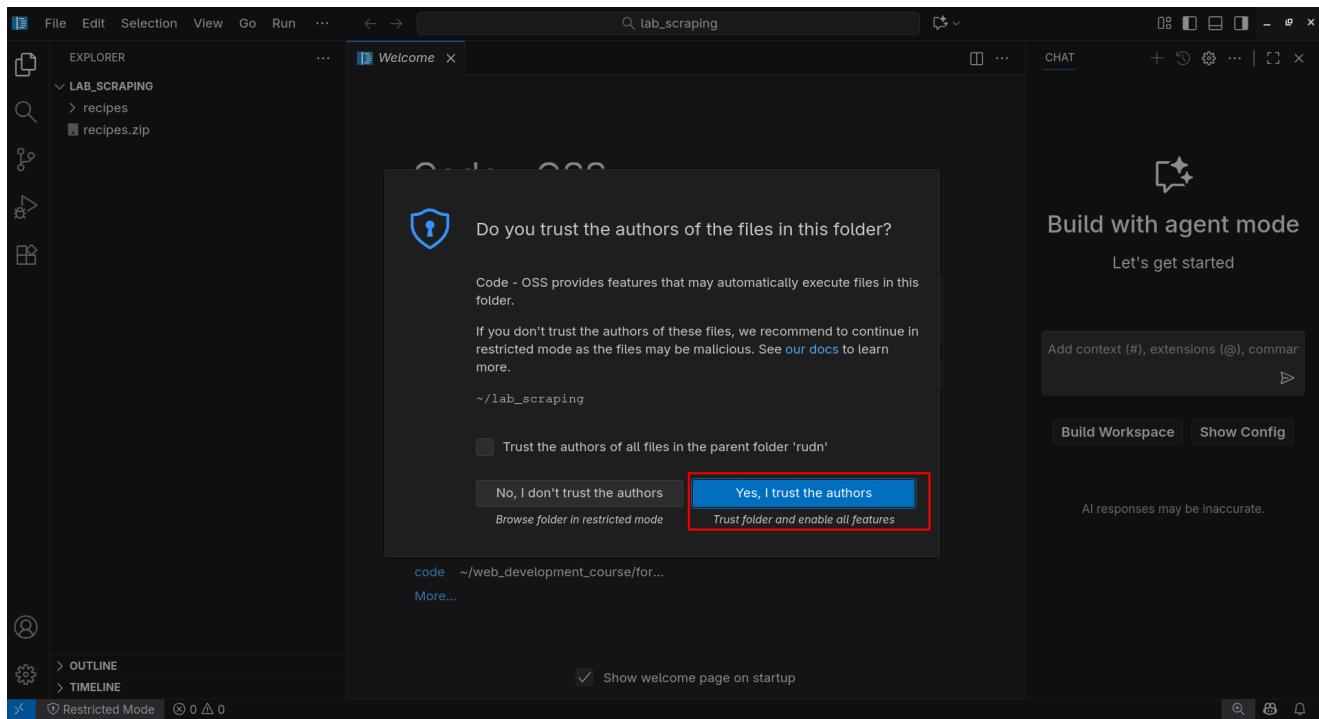
- Веб-скрапинг имеет некоторые этические аспекты. Публичное использование информации с других сайтов без разрешения может привести к неудобствам или юридическим проблемам. Поэтому я рекомендую использовать общедоступную информацию, предназначенную для свободного использования.
- На сайтах установлено специальное программное обеспечение, блокирующее веб-скрапинг. Поэтому перед началом работы убедитесь, что сайт, который вы собираетесь скrapить, разрешает такую практику.
- Практически все сайты имеют защиту от DDoS-атак, что означает ограничение на количество запросов, которые вы можете сделать за определённый промежуток времени (до блокировки вашего IP-адреса). Поэтому использование библиотек, таких как `request`, следует обдумывать и ограничивать несколькими попытками (не используйте их циклами). Также рекомендуется по возможности добавлять паузы между запросами.
- Используйте веб-скрапинг в основном для автоматизации скучных задач или при ограниченном объёме информации. Не создавайте точные копии сайта и не используйте всю доступную информацию.
- Эта лаборатория использует общедоступную и ограниченную информацию в образовательных целях и Локальная среда, то есть мы отправляем запросы не к реальным веб-сайтам в интернете, а к локальному веб-сайту, работающему на вашем компьютере.

После прочтения всей этой информации мы можем начать

### НАСТРОЙКА РАБОЧЕЙ ПАПКИ

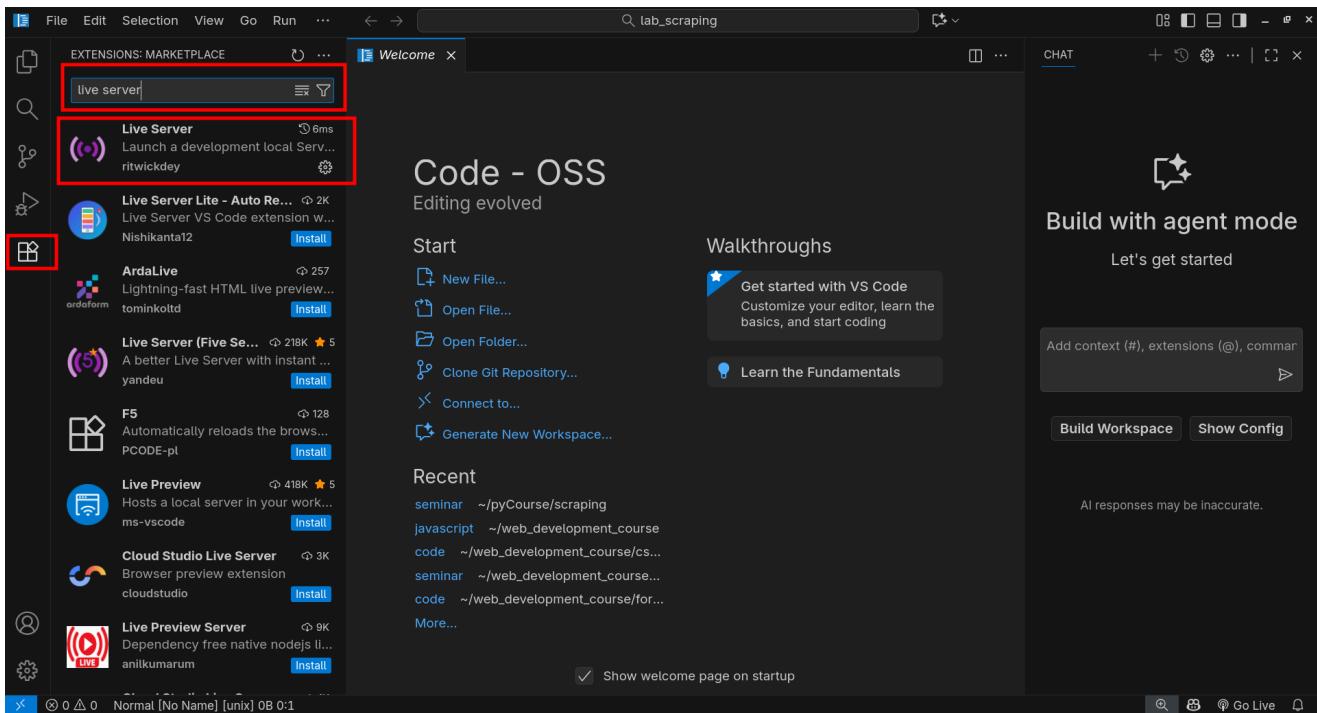
1. Создайте новую папку с именем `lab_scraping`.
2. Скопируйте файл `recipes.zip` в папку `lab_scraping`.
3. Распакуйте файл `recipes.zip` в папку `lab_scraping`. Это создаст новую папку с именем `recipes`.
4. Откройте `vscode`, затем нажмите `File` → `open folder` и выберите папку `lab_scraping`. Эта папка откроется в `vscode`. Если `vscode` предложит вам доверять авторам, нажмите `yes`.



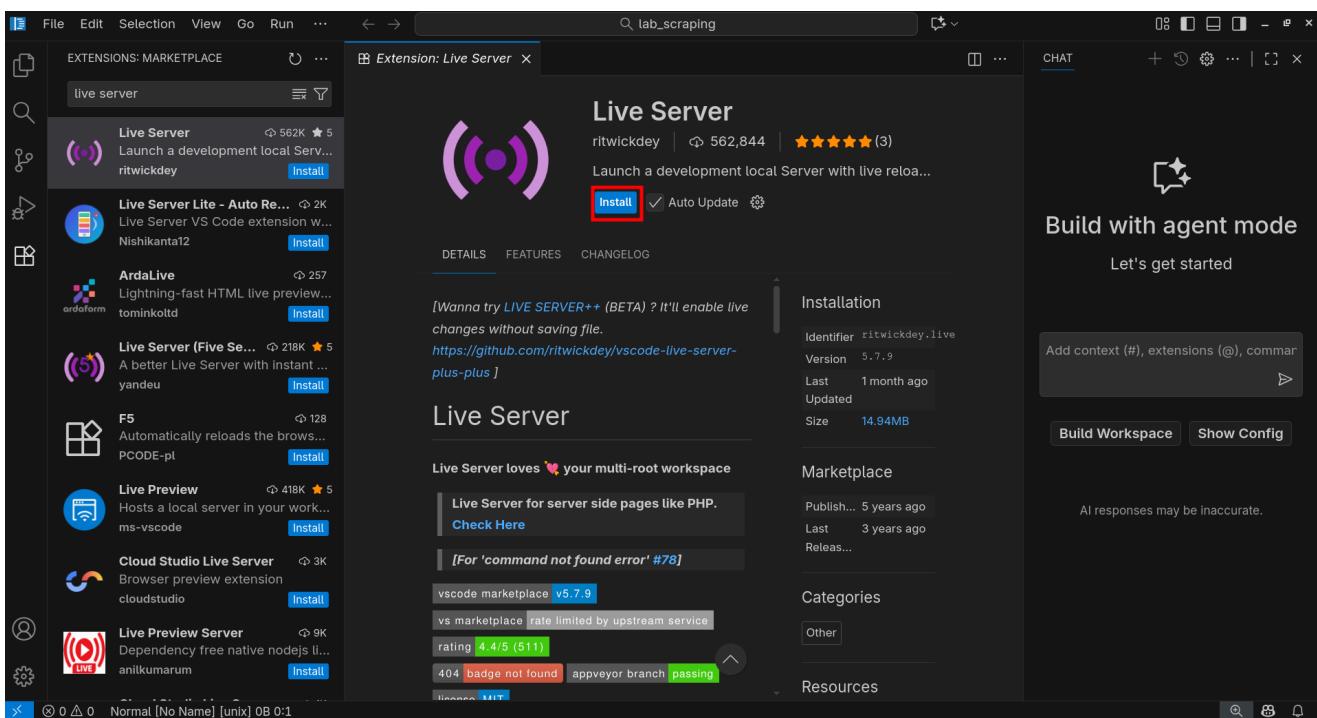


5. Нажмите кнопку «Расширения», затем в строке поиска напишите «live server».

Затем выберите расширение live server (как показано на рисунке).

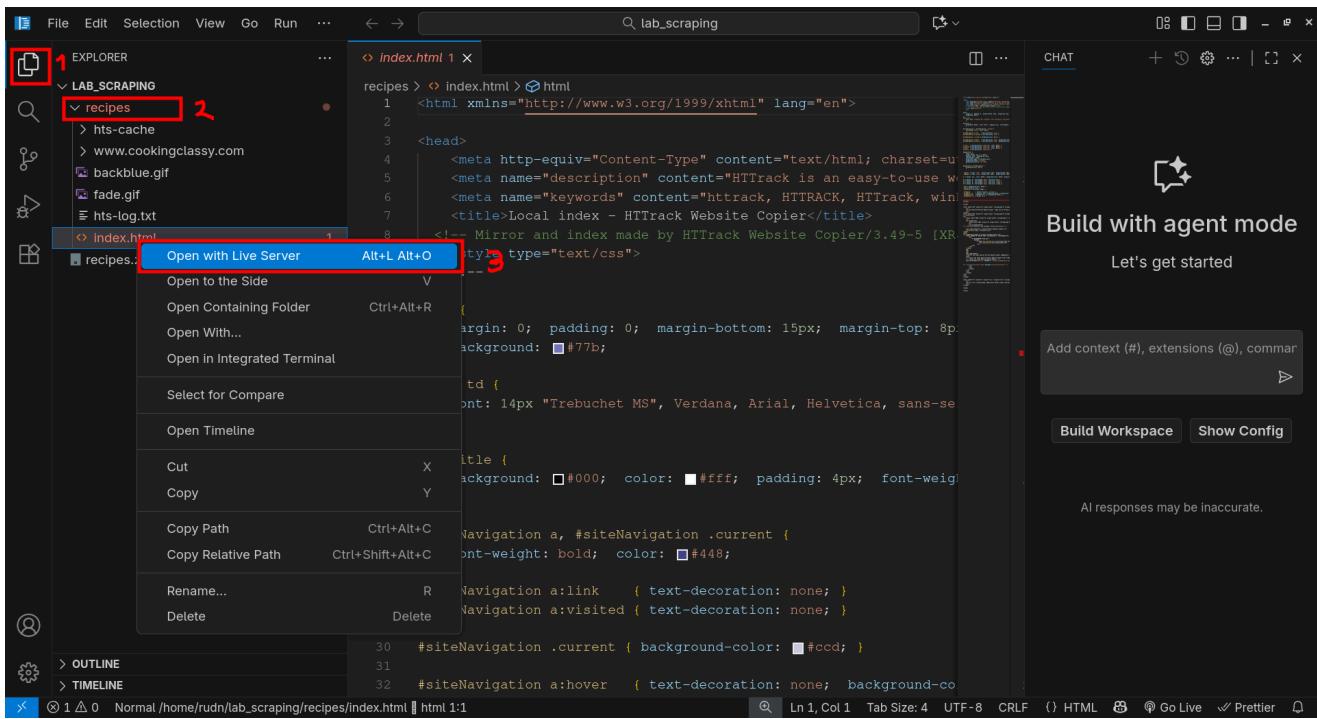


## 6. Нажмите «Установить» и дождитесь установки расширения.



## 5. В «vscode» перейдите в папку «recipes», чтобы отобразить её содержимое.

Затем найдите в папке «recipes» файл «index.html», щёлкните по нему правой кнопкой мыши и выберите «Открыть с помощью Live Server». Всё это нужно сделать после установки расширения «live server», иначе оно не заработает.

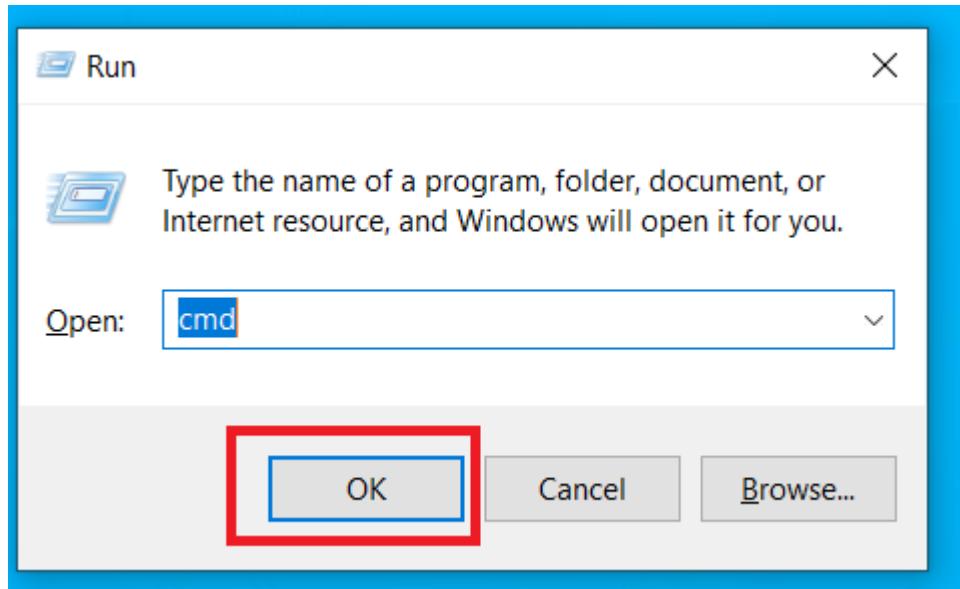


После этого откроется браузер со страницей, с которой мы будем скачивать рецепты. Эта веб-страница развернута в автономном режиме, то есть вся информация берётся из папки `recipes`, которая находится на вашем локальном компьютере.

## УСТАНОВКА PYTHON

6. Убедитесь, что в вашей системе установлен Python:

- нажмите клавишу Windows, а затем клавишу `r` (`win + r`)
- откройте терминал, введя `cmd`, а затем нажмите `Enter`



- введите в терминале `python --version`

```
C:\Windows\system32\cmd.exe
Microsoft Windows [Version 10.0.18363.592]
(c) 2019 Microsoft Corporation. All rights reserved.

C:\Users\RUDN>python --version
Python 3.13.9
номер версии
C:\Users\RUDN>
```

Если терминал выдаёт версию, значит, Python установлен. Если нет, перейдите к шагу 7

7. Используйте установщик, предоставленный TUIS. Если он не работает, скачайте его с сайта <https://www.python.org/downloads/>, выбрав свою операционную систему.

The screenshot shows the Python.org Downloads page. At the top, there's a yellow button labeled "Download Python 3.14.0". Below it, a message says "Looking for Python with a different OS? Python for [Windows](#) [Linux](#) / [Unix](#), [macOS](#), [Android](#), [other](#)". To the right, there's a cartoon illustration of two packages hanging from parachutes. The main content area is titled "Active Python Releases" and contains a table of Python versions with their maintenance status, first release date, end of support date, and release schedule. At the bottom, there's a link "Looking for a specific release?".

Python version	Maintenance status	First released	End of support	Release schedule
3.15	pre-release	2026-10-07 (planned)	2031-10	PEP 790
3.14	bugfix	2025-10-07	2030-10	PEP 745
3.13	bugfix	2024-10-07	2029-10	PEP 719
3.12	security	2023-10-02	2028-10	PEP 693
3.11	security	2022-10-24	2027-10	PEP 664
3.10	security	2021-10-04	2026-10	PEP 619
3.9	end of life, last release was 3.9.25	2020-10-05	2025-10-31	PEP 596

www.python.org/downloads/windows/

Python >> Downloads >> Windows

## Python Releases for Windows

• Latest Python install manager - Python install manager 25.0  
• Latest Python 3 Release - Python 3.14.0

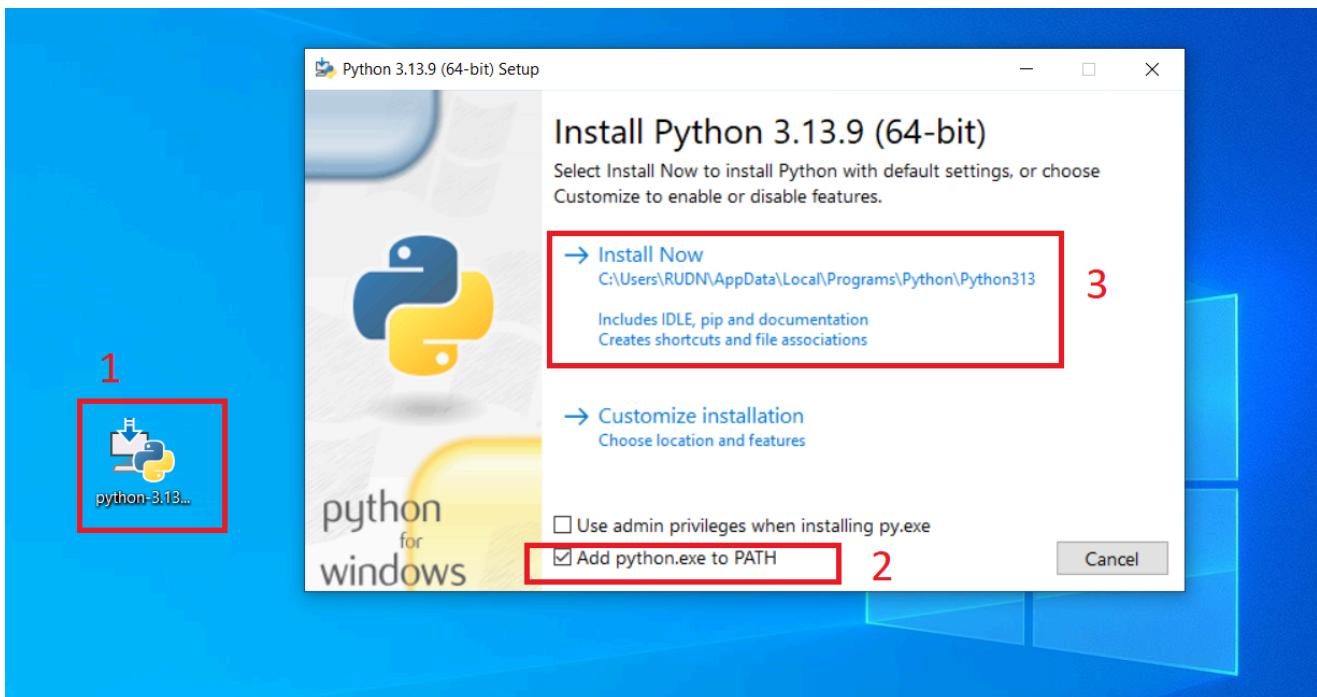
**Stable Releases**

- Python 3.13.9 - Oct. 14, 2025
  - Note that Python 3.13.9 cannot be used on Windows 7 or earlier.
  - Download Windows installer (64-bit)
  - Download Windows installer (32-bit)
  - Download Windows installer (ARM64)
  - Download Windows embeddable package (64-bit)
  - Download Windows embeddable package (32-bit)
  - Download Windows embeddable package (ARM64)
  - Download Windows release manifest
- Python install manager 25.0 - Oct. 8, 2025
  - Download Installer (MSIX)
  - Download MSI package
- Python 3.14.0 - Oct. 7, 2025
  - Download using the Python install manager.
  - Download Windows installer (64-bit)
  - Download Windows installer (32-bit)
  - Download Windows installer (ARM64)
  - Download Windows embeddable package (64-bit)
  - Download Windows embeddable package (32-bit)

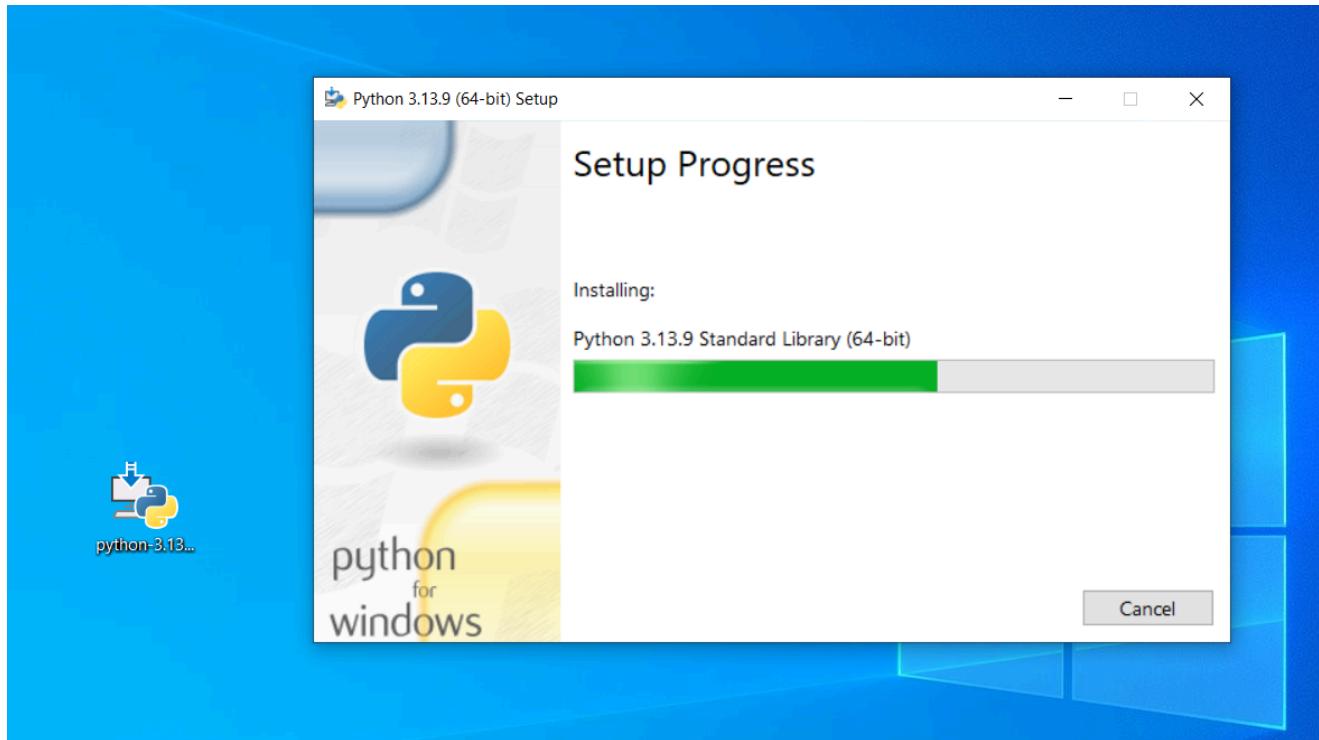
**Pre-releases**

- Python install manager 25.1 beta 1 - Oct. 27, 2025
  - Download Installer (MSIX)
  - Download MSI package
- Python 3.15.0a1 - Oct. 14, 2025
  - Download Windows installer (64-bit)
  - Download Windows installer (32-bit)
  - Download Windows installer (ARM64)
  - Download Windows embeddable package (64-bit)
  - Download Windows embeddable package (32-bit)
  - Download Windows embeddable package (ARM64)
  - Download Windows release manifest
- Python install manager 25.0 beta 15 - Sept. 18, 2025
  - Download Installer (MSIX)
  - Download MSI package
- Python 3.14.0rc3 - Sept. 18, 2025
  - Download Windows installer (64-bit)
  - Download Windows installer (32-bit)
  - Download Windows installer (ARM64)
  - Download Windows embeddable package (64-bit)
  - Download Windows embeddable package (32-bit)

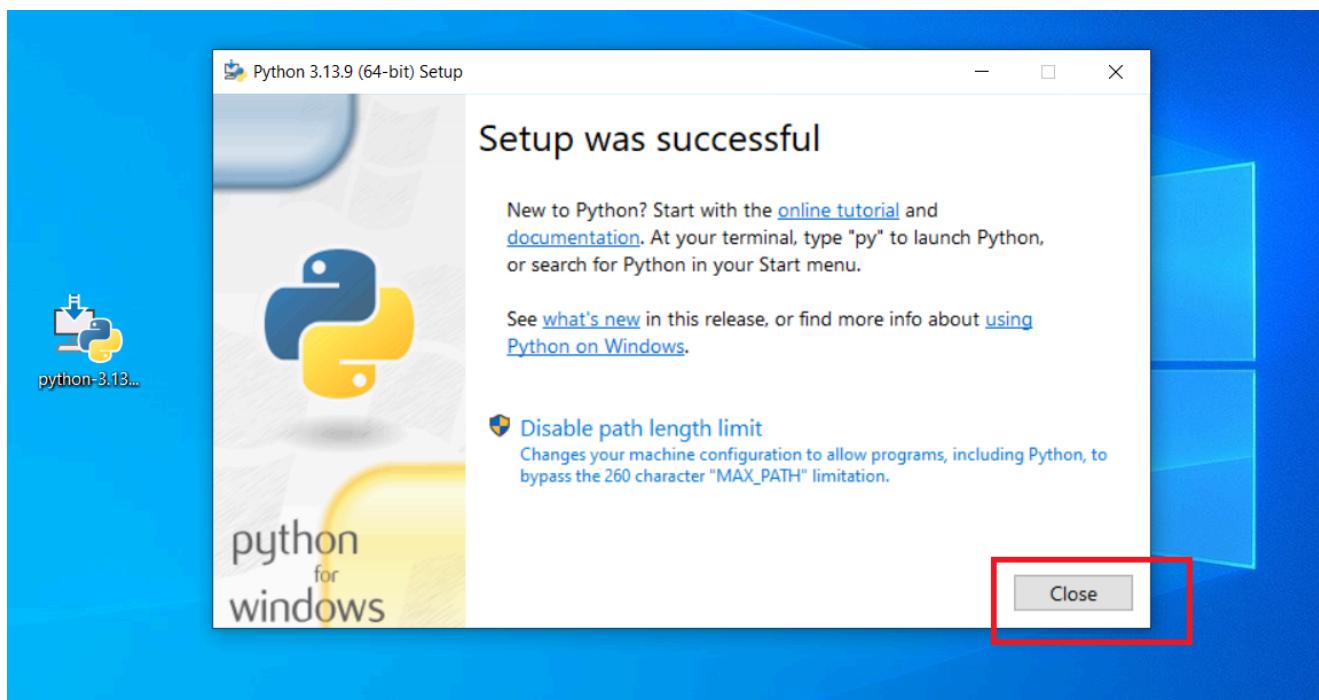
8. Дважды щёлкните по установщику Python, затем установите флажок «Добавить Python в переменную окружения PATH» и нажмите «Установить сейчас».



Python должен начать установку в вашей операционной системе.

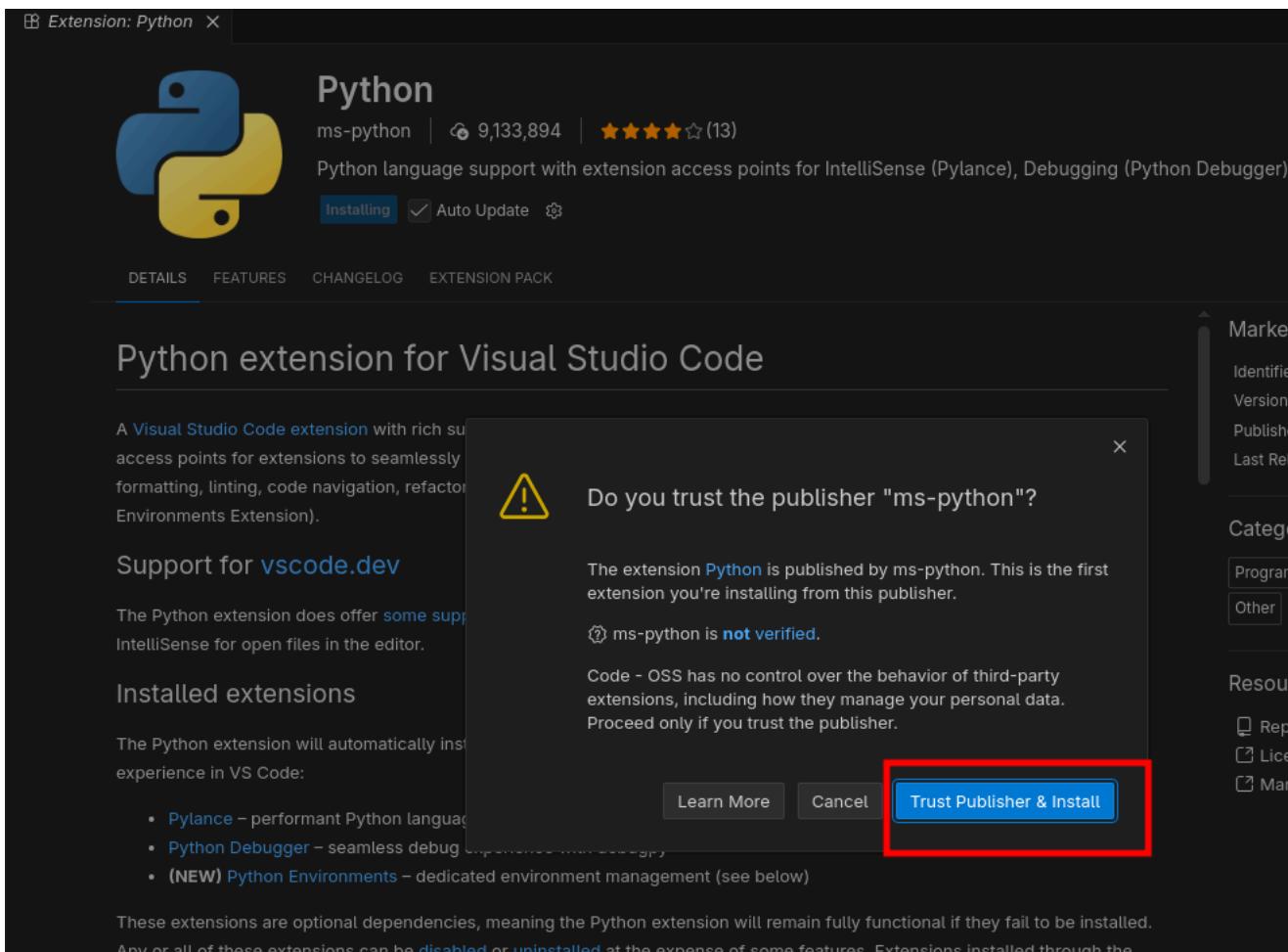
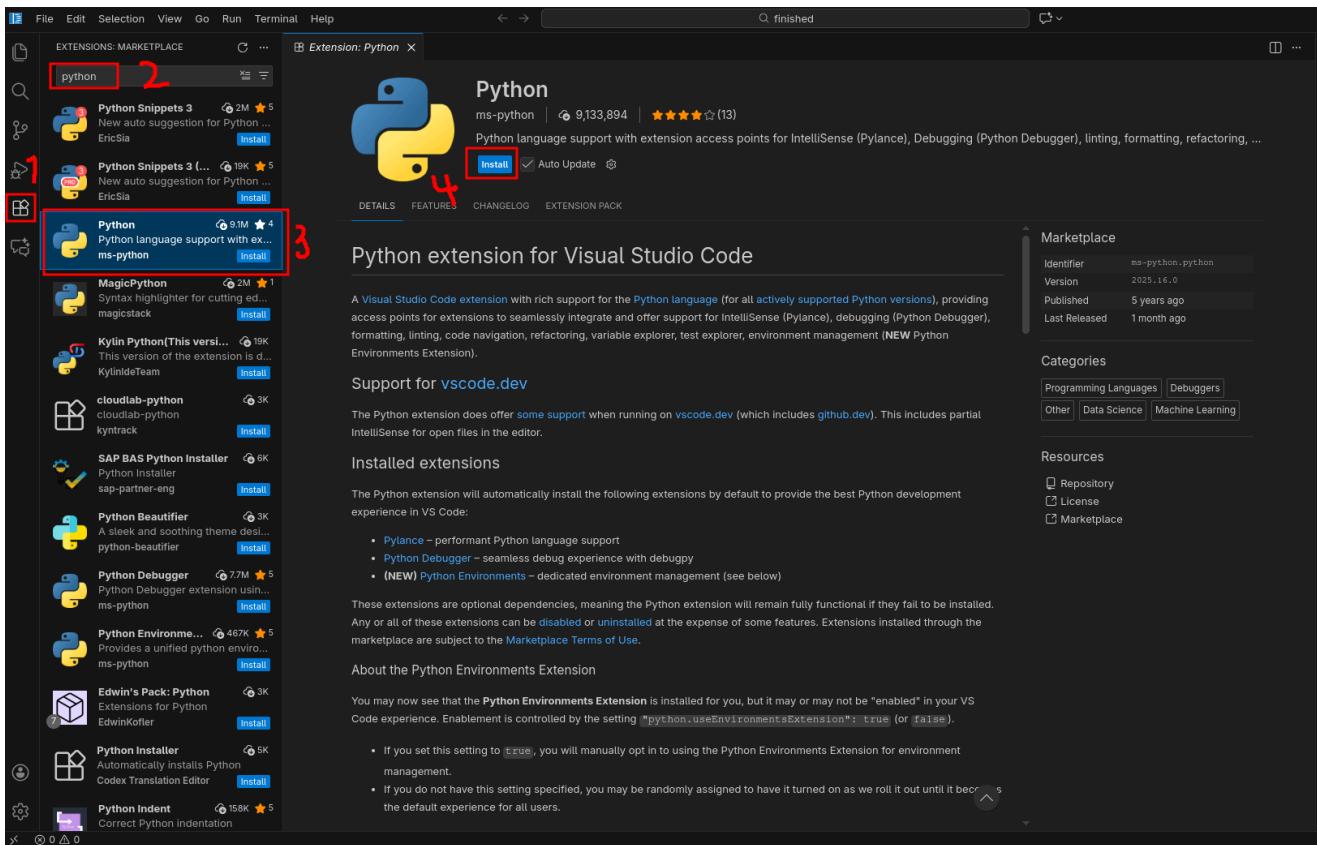


9. После завершения установки нажмите кнопку «Закрыть» и повторите шаг 6 (убедитесь, что Python установлен в вашей системе).

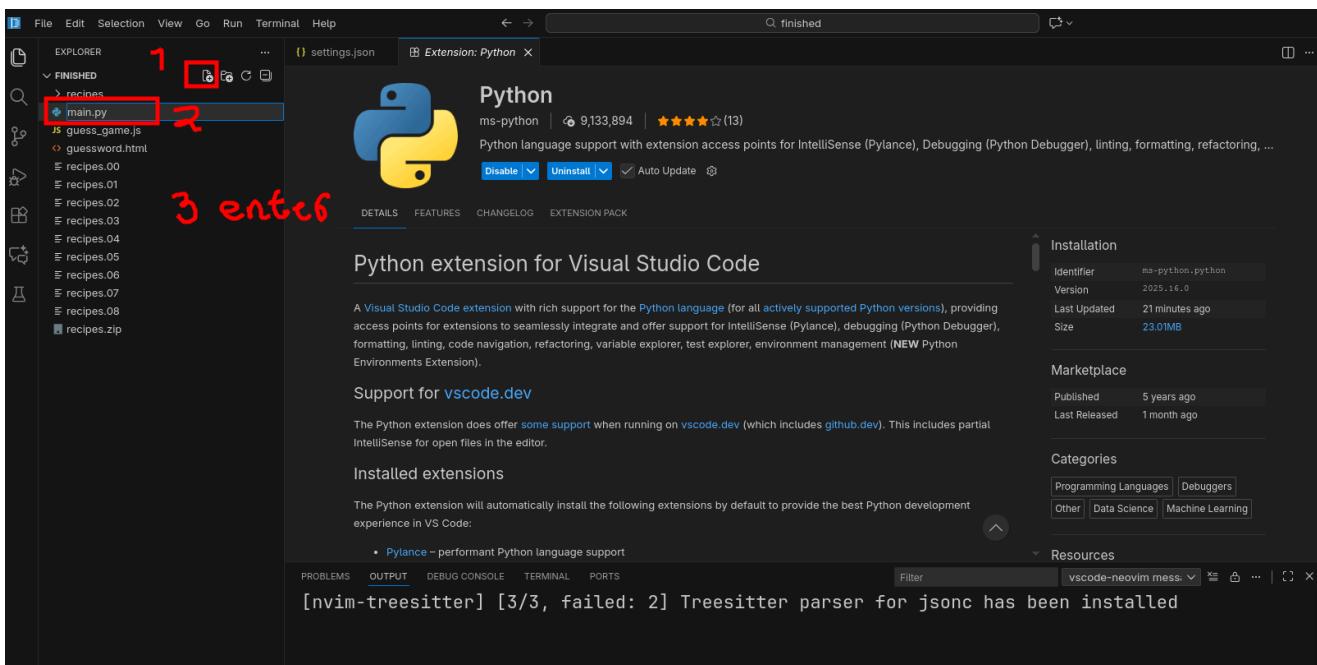


## ПРОВЕРКА РАБОТЫ PYTHON В VS CODE

10. В VS CODE перейдите в раздел «Расширения», затем введите в строке поиска `python`, затем выберите расширение «python» (показано на рисунке) без каких-либо дополнительных настроек и установите его.



11. Вернитесь в проводник и создайте новый файл Python с именем `main.py`, нажав кнопку «Создать файл», затем указав имя файла и его расширение (не забудьте расширение «ру») и нажав Enter.



12. Выберите файл `main.py`, щёлкнув по нему, и введите в него следующую строку кода:

```
1 print("hello from python")
```

13. Сохраните файл, нажав `CTRL+S`, затем нажмите кнопку «Run python file» и убедитесь, что в консоли выведено сообщение «hello from python».

The screenshot shows the Visual Studio Code interface. The Explorer sidebar on the left lists files and folders, including a 'FINISHED' folder containing 'recipes' (recipes.00 to recipes.08) and 'guess\_word.html'. The main workspace shows a 'main.py' file with the code: 

```
1 print("hello from python")
```

. The terminal at the bottom displays the output of running the script: 

```
/bin/python /home/rudn/web_development_course/scraping/seminar/finished/main.py
Welcome to fish, the friendly interactive shell
Type help for instructions on how to use fish
•rudn@archlinux ~ ~/s/s/finished (main)> /bin/python /home/rudn/web_development_course/
scraping/seminar/finished/main.py
hello from python
○ rudn@archlinux ~ ~/s/s/finished (main)>
```

. A red box highlights the word 'hello' in the terminal output, and a red number '2' is placed next to it.

## ЗАГРУЗИТЕ СВОИ РЕЗУЛЬТАТЫ В TUIS

14. Прикрепите к TUIS следующие скриншоты:

- Работающий сайт с рецептами (шаг 5)
- Установка версии Python (шаг 6)
- Работающий Python в VScode (шаг 13) python working on vscode (step 13)