



University  
of Glasgow

# Python for Data Analysis

AccFin PhD workshop

1<sup>st</sup> Nov 2022

Jiawen Liang

[Jiawen.Liang@glasgow.ac.uk](mailto:Jiawen.Liang@glasgow.ac.uk)

**WORLD  
CHANGING  
GLASGOW**





# Your programming background?



# Python introduction

- Developed a large and active computing and data analysis community, e.g., Stack Overflow Python.
- It is one of the most popular languages for
  - data science
  - machine learning
  - software development
- There are useful packages:
  - NumPy, pandas, matplotlib, Scipy, scikit-learn, statsmodels



# Python introduction

- Anaconda Python
    - an open source package management system and environment management system
    - runs on Windows, macOS, Linux and z/OS.
    - A distribution of the Python and R programming languages applications
    - Simply the package management and deployment.
    - Helps you find and install packages.
- 

## Conda





# Python introduction

## Pycharm

- It is a dedicated Python Integrated Development Environment (IDE) providing a wide range of essential tools for Python developers, tightly integrated to create a convenient environment for productive Python, web, and data science development.
- PyCharm is available in three editions:
  - *Community* (free and open-sourced): for smart and intelligent Python development
  - *Professional* (paid, but free for students and teachers): for professional Python, web, and data science development
  - *Edu* (free and open-sourced): for learning Python and related technologies with integrated educational tools.

Python workshop – Linear\_regression.py

Current File ▾

Linear\_regression.py

Project

Python workshop [workshop]

- .ipynb\_checkpoints
- data
- Python for data analysis
  - 2-Mean-Variance-Optimis
  - Linear\_regression.py
- Python workshop\_solution
- S2\_Python\_Basics.ipynb
- S3\_Data\_Analysis.ipynb
- S4\_Statistical\_Inference.ipynb

External Libraries

Scratches and Consoles

```
1 import matplotlib.pyplot as plt
2
3 X = [[1],[2], [4], [5]]
4 Y = [2, 4, 6, 8]
5 plt.scatter(X, Y)
6 plt.show()
7
8 from sklearn.linear_model import LinearRegression
9 regr = LinearRegression()
10 regr.fit(X, Y)
11
12 y = regr.predict([[1.5]])
13 print(y)
14
15 plt.scatter(X, Y)
16 plt.plot(X, regr.predict(X)) #连线图
17 plt.show()
18
19 print('coefficient a:' + str(regr.coef_[0]))
20 print('intercept b:' + str(regr.intercept_))
```

Bookmarks

Structure

Version Control Run Python Packages TODO Python Console Problems Terminal Services

Looks like you're using code cells in the editor: Would you like to turn scientific mode on? // Use scientific mode Keep current layout // // Scientific mode provides... (26 minutes ago) 1:1 (1 char, 1 line break) LF UTF-8 4 spaces Python 3.9 (FinTech)



# Python introduction

## Spyder

- a complicated IDE like Pycharm.
- a free and open source scientific IDE for Python analysis, debugging, editing, and profiling.





# Python introduction

## Spyder vs. Pycharm

- Similarities:
  - advanced text editing software programs
  - built-in debugger
- Differences:
  - Spyder is a scientific IDE. It has a more extensive library to download when you install it with Anaconda.
  - PyCharm is more flexible. It has many plug-ins for Python programming, various supported languages, web development, and data science. It also provides a free education edition to learn Python.

File Edit View Insert Cell Kernel Help

ianitagomez/Local/Dev-Spyder/spyder/spyder/plugins/plots/plugin.py

Plots

```

1  # -*- coding: utf-8 -*-
2
3  # Copyright © Spyder Project Contributors
4  # Licensed under the terms of the MIT License
5  # (see spyder/_init__.py for details)
6
7  """
8      Plots Plugin.
9  """
10
11 # Third party imports
12 from qtpy.QtCore import Signal
13
14 # Local imports
15 from spyder.api.plugins import Plugins, SpyderDockablePlugin
16 from spyder.api.translations import get_translation
17 from spyder.plugins.plots.widgets.main_widget import PlotsWidget
18
19
20 # Localization
21 _ = get_translation('spyder')
22
23
24 class Plots(SpyderDockablePlugin):
25     """
26         Plots plugin.
27     """
28     NAME = 'plots'
29     REQUIRE_S = [Plugins.IPythonConsole]
30     TABIFY_S = [Plugins.VariableExplorer, Plugins.Help]
31     WIDGET_C = PlotsWidget
32     CONF_SECTION = NAME
33     CONF_FILE = False
34     DISABLE_ACTIONS_WHEN_HIDDEN = False
35
36     # --- SpyderDockablePlugin API
37     def get_name(self):
38         return _('Plots')
39
40     def get_description(self):
41         return _('Display, explore and save console generated plots.')
42
43     def get_icon(self):
44         return self.create_icon('hist')
45
46     def register(self):
47         # Plugins
48         ipyconsole = self.get_plugin(Plugins.IPythonConsole)
49
50         # Signals
51         ipyconsole.sig_shellwidget_changed.connect(self.set_shellwidget)
52         ipyconsole.sig_shellwidget_process_started.connect(
53             self.add_shellwidget)
54         ipyconsole.sig_shellwidget_process_finished.connect(
55             self.remove_shellwidget)
56

```

Variable Explorer

Name	Type	Size	Value
bool	bool	1	True
data	Array of str128	(3, 3)	ndarray object of numpy module
datetime_object	datetime	1	2021-04-14 17:35:14.687085
df	DataFrame	(2, 2)	Column names: Col1, Col2
filename	str	53	/Users/Documents/spyder/spyder/tests/test_dont_use.py
li	list	5	['abcd', 745, 2.23, 'efgh', 70.2]
myset	set	3	{'2', '1', '3'}
r	float	1	6.46567886443
t	tuple	5	('abcd', 745, 2.23, 'efgh', 70.2)
tinylist	list	2	[123, 'efgh']
x	float64	1	1.1235123099439

Help Variable Explorer Files Code Analysis

Plots

IPython console

History

conda: spyder.dev Python 3.8.5 LSP Python: ready master Line 10, Col 1 UTF-8 LF RW Mem 64%



# Python introduction

- Jupyter notebook
  - an interactive computational environment
  - Ipython kernel
  - A notebook user interface
  - Text editing: code and markdown.
  - Easy to start
  - Good for learning programming
- Jupyter is a good choice,
  - if you are a Python learner
  - if your project is not complex enough
  - if you want to visualize your codes



# Before installing Python

- Apply for a free educational license which is needed to activate Pycharm Professional
- <https://www.jetbrains.com/community/education/#students>

[Free License Programs](#)

[Academic Licensing](#) [Open Source](#) [User Groups](#) [Events Partnership](#) [Developer Recognition](#)

## Individual licenses for students and teachers

Get free access to all JetBrains IDEs for personal use at school or at home.

### Who can get free individual licenses for education

Students and faculty from accredited educational institutions (high schools, colleges, and universities) are welcome to apply.

Students need to be enrolled in an accredited educational program that takes one or more years of full-time study to complete.

Not sure about the license terms? [Check out the FAQ](#) or read the full terms [here](#).

[Apply now](#)



# Getting started with Python

- If you don't want to download Conda and Pycharm, or you find it difficult to install Apps, please use Google Colaboratory Python Notebooks
- <https://colab.research.google.com/notebooks/intro.ipynb>

The screenshot shows the Google Colaboratory interface. On the left, there's a sidebar with icons for files, notebooks, and other tools. The main area displays the text "Welcome to Colab!" followed by a brief introduction. At the top, there's a navigation bar with "File" selected, showing options like "New notebook", "Open notebook", "Upload notebook", "Save a copy in Drive", "Save a copy as a GitHub Gist", "Save a copy in GitHub", "Save", "Download", and "Print". There are also "Code" and "Text" buttons, and a "Copy to Drive" button.



University  
of Glasgow

# Download and Installation (MacOS)

Download and

Install Anacoda Distribution for MacOS, Windows or Linus

<https://www.anaconda.com/products/distribution>



Products

Pricing

Solutions

Resources

Partners

Blog

Company

Contact Sales

Individual Edition is now

## ANACONDA DISTRIBUTION

The world's most popular open-source Python distribution platform

### Anaconda Distribution

Download



For MacOS

Python 3.9 • 64-Bit Graphical Installer • 688 MB

Get Additional Installers



anaconda - Google Search

https://www.google.co.uk/search?q=anaconda&ei=wK5aY-iLOdSUgQat5LGQBA&ved=0ahUKEwjo652o54D7AhVUsAKHS1yDEIQ4dUDCA8&uact=5&oq=anaconda&gs\_lcp=Cgdnd3Mtd2l6EAMyCggAELEDEIMBEEMy8AgAEEMyB... A

Not syncing

Google anaconda

All Images Videos News Shopping More Tools

About 102,000,000 results (0.57 seconds)

<https://www.anaconda.com>

**Anaconda | The World's Most Popular Data Science Platform**

Anaconda is the birthplace of Python data science. We are a movement of data scientists, data-driven enterprises, and open source communities.

**Installation**  
On Windows, macOS, and Linux, it is best to install Anaconda for ...

**Anaconda Distribution**  
Conda is an open-source package and environment management ...

**Installing on Windows**  
Note. Using Anaconda in a commercial setting? You may ...

**Open Source**  
An open-source Python natural language toolkit for symbolic ...

[More results from anaconda.com »](#)

**People also ask :**

Is anaconda a real snake?  
What is the biggest anaconda ever?  
Is anaconda a Python?  
Is anaconda a poisonous snake?

**Anaconda**  ANACONDA

Computer program

Anaconda is a distribution of the Python and R programming languages for scientific computing, that aims to simplify package management and deployment. The distribution includes data-science packages suitable for Windows, Linux, and macOS.

[Wikipedia](#)

**Developer:** Continuum analytics

**License:** Freemium (Miniconda and the Individual Edition are free software, but the other editions are software as a service)

**Operating system:** Windows, macOS, Linux

**Stable release:** 2022.10 / 18 October 2022; 8 days ago

**Initial release:** 0.8.0/17 July 2012; 10 years ago

**Programming language:** Python

Install

# Start coding immediately

Introducing Anaconda Notebooks + Learning

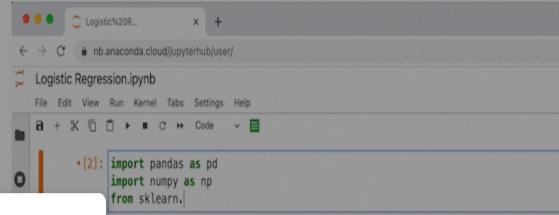
Master the foundations of data science and spin anytime, anywhere! Expert-led tutorials. Cloud-h the cloud today.

Get Started

Do you want to allow downloads on  
“www.anaconda.com”?

You can change which websites can download files in  
Websites Preferences.

Cancel    Allow



The screenshot shows a Jupyter Notebook interface with the title 'Logistic Regression.ipynb'. The code cell at the top contains the following Python imports:

```
import pandas as pd
import numpy as np
from sklearn.
```



## Welcome to the Anaconda3 Installer

**● Introduction**

- Read Me
- Licence
- Destination Select
- Installation Type
- Installation
- DataSpell
- Summary

You will be guided through the steps necessary to install this software.

[Go Back](#)[Continue](#)



● Introduction

● Read Me

● Lic

● De

● Ins

● Ins

● Da

● Su

## Software Licence Agreement

=====  
End User License Agreement - Anaconda Distribution  
=====

**To continue installing the software you must agree to the terms of the software licence agreement.**

Click Agree to continue or click Disagree to cancel the installation and quit the Installer.

Read Licence

Disagree

Agree



as Anaconda Individual Edition),

\* Modify and create derivative works of sample source code delivered in Anaconda Distribution from Anaconda's repository, and;

\* Redistribute code files in source (if provided to you by Anaconda as source) and binary forms, with or without modification subject to the requirements set forth below, and;

Print...

Save...

Go Back

Continue



## Install Anaconda3



### Select a Destination

- Introduction
- Read Me
- Licence
- Destination Select**
- Installation Type
- Installation
- DataSpell
- Summary



How do you want to install this software?



Install for all users of this computer



Install for me only



Install on a specific disk...

Installing this software requires 721.9 MB of space.

You have chosen to install this software in your home folder.  
Only the current user will be able to use this software.

Go Back

Continue



## Install Anaconda3



### Standard Install on "Macintosh HD"

- Introduction
- Read Me
- Licence
- Destination Select
- Installation Type**
- Installation
- DataSpell
- Summary



This will take 721.9 MB of space on your computer.

Click Install to perform a standard installation of this software in your home folder. Only the current user of this computer will be able to use this software.

[Change Install Location...](#)

[Customise](#)

[Go Back](#)

[Install](#)



## Install Anaconda3



### Installing Anaconda3

- Introduction
- Read Me
- Licence
- Destination Select
- Installation Type
- Installation**
- DataSpell
- Summary



**Running package scripts...**

Install time remaining: Less than a minute

Go Back

Continue



## Install Anaconda3



### Anaconda + JetBrains

- Introduction
- Read Me
- Licence
- Destination Select
- Installation Type
- Installation
- DataSpell**
- Summary

Working with Python and Jupyter is a breeze in DataSpell. It is an IDE designed for exploratory data analysis and ML. Get better data insights with DataSpell.

DataSpell for Anaconda is available at:

<https://www.anaconda.com/dataspell>



ANACONDA<sup>®</sup>



Go Back

Continue



The installation was completed successfully.

- Introduction
- Read Me
- Licence
- Destination Select
- Installation Type
- Installation
- DataSpell
- Summary**

Thank you for installing **Anaconda Distribution**.

Here are some helpful tips and resources to get you started. We recommend you bookmark these links so you can refer back to them later.

[Anaconda Distribution Tutorial](#)

[Getting Started with Anaconda](#)

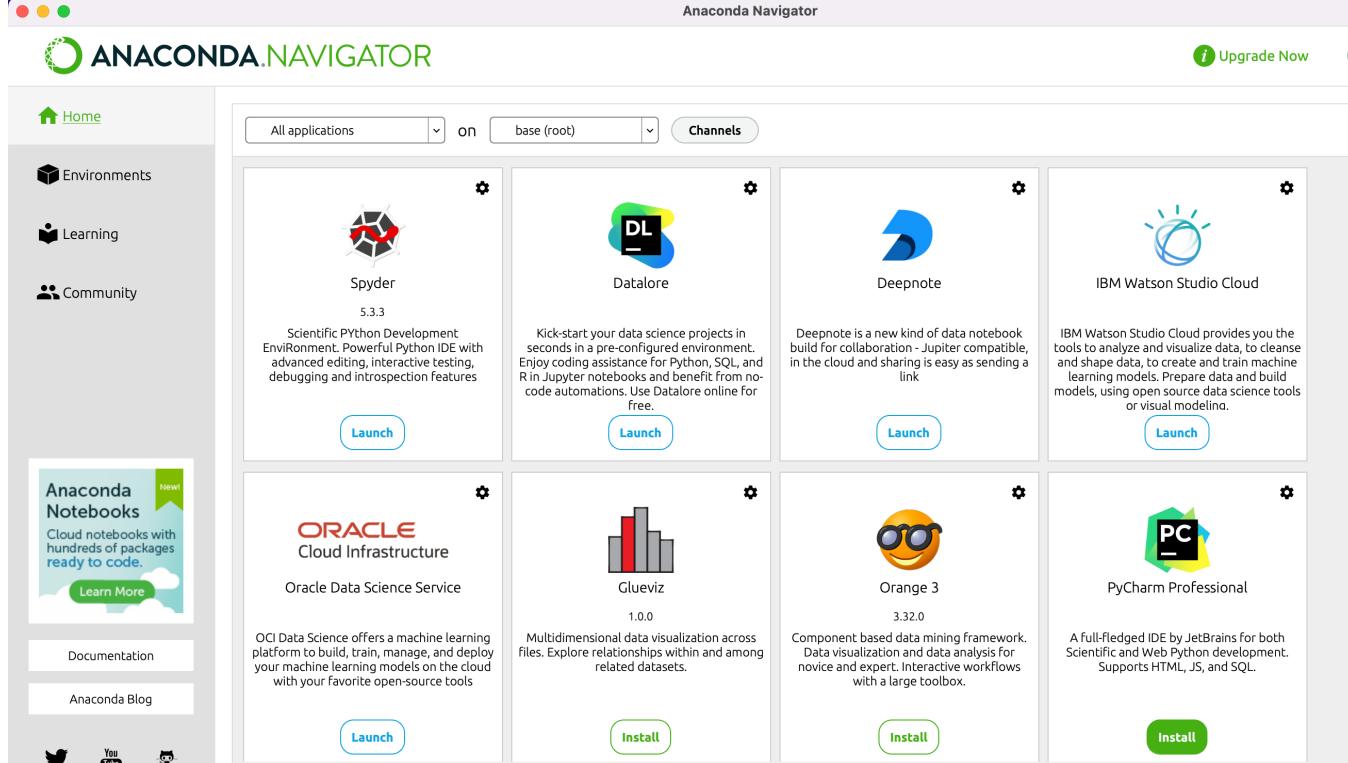


Go Back

Close

# Download and Installation

After installing Conda (Anaconda), open Conda > Home > Install and Lauch Pycharm Professional



# Ready to go Pro?

Working with Python and Jupyter notebooks is a breeze with PyCharm Professional.

Download now and have the best tools for data scientists and data engineers at your fingertips

 **PyCharm** **ANACONDA**[Download now](#)[.dmg ▾](#)

Free 30-day trial



PyCharm

.drag\_to(  )

Applications



# Download and Installation (Windows)

Next, I will show how to download and install Conda, and Pycharm on Windows.

The screenshot shows a web browser window displaying the Anaconda website at <https://www.anaconda.com>. The page features a dark header with the Anaconda logo and navigation links for Products, Pricing, Solutions, Resources, Partners, Blog, and Company. A 'Contact Sales' button is also visible. Below the header, a large banner promotes the 'STATE OF DATA SCIENCE 2022' report, with a green 'Access the Report' button. The main content area contains the text: 'Data science technology for a better world.' followed by a description of Anaconda's offerings. At the bottom, there is a 'Download' button for Windows, a note about the Python 3.9 installer size (621 MB), and links for 'For Windows', 'Get Additional Installers', and 'Get Support'. A live chat bubble in the bottom right corner asks if the user has registered for an upcoming webinar. The browser's taskbar at the bottom shows various pinned icons and the system clock indicating 17:16 on 27/10/2022.

Anaconda | The World's Most Popular Data Science Platform

Products Pricing Solutions Resources Partners Blog Company Contact Sales

STATE OF DATA SCIENCE 2022

Access the Report

Data science technology for a better world.

Anaconda offers the easiest way to perform Python/R data science and machine learning on a single machine. Start working with thousands of open-source packages and libraries today.

Download Windows

For Windows

Python 3.9 • 64-Bit Graphical Installer • 621 MB

Get Additional Installers

Have you registered for our upcoming webinar?

Sunset 17:16 27/10/2022



Recycle Bin



Microsoft  
Edge



Anaconda3-2  
022.10-Wind  
ows-x86\_64



Type here to search



Temps drop



17:17  
27/10/2022



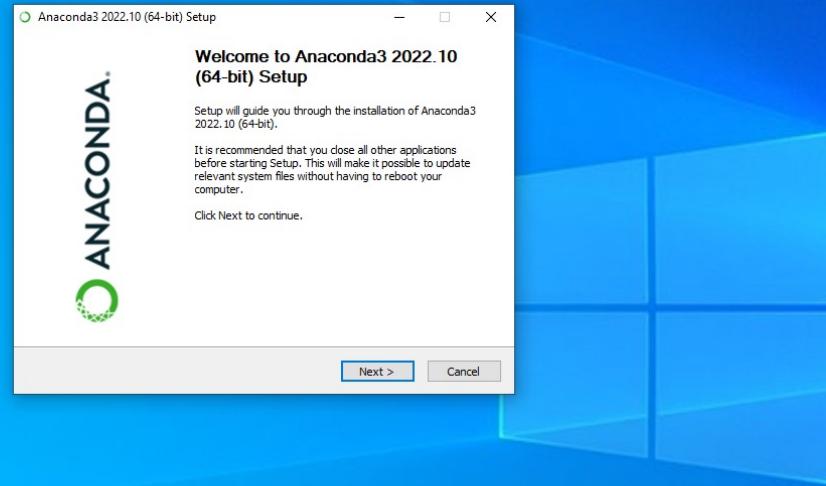


Recycle Bin



Microsoft Edge

Anaconda3...



Type here to search



Temp drop 17:18  
27/10/2022



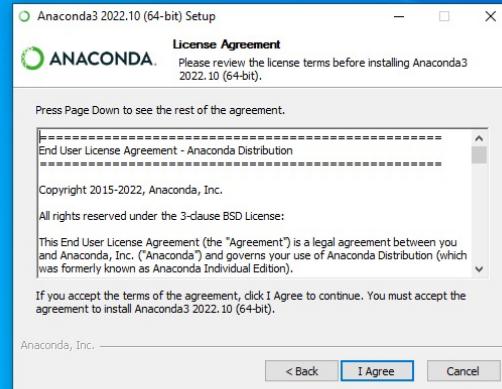
Recycle Bin



Microsoft Edge



Anaconda3...



Type here to search

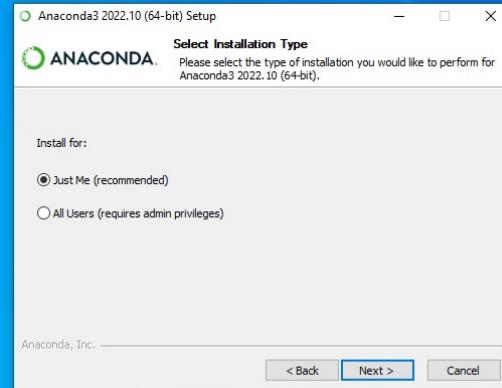
17:18  
Temps drop 27/10/2022 4



Recycle Bin



Microsoft Edge



Type here to search



16°C Cloudy 18:46  
27/10/2022 4

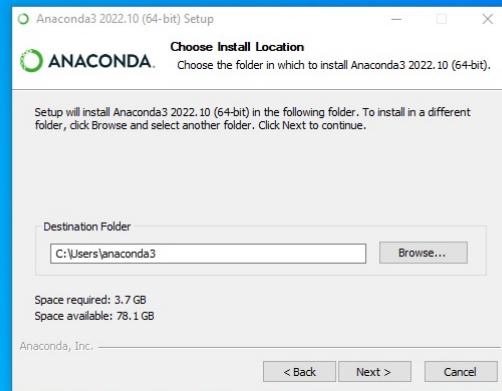


Recycle Bin



Microsoft Edge

Anaconda3-2  
022.10-Wind  
ows-x86\_64



Type here to search



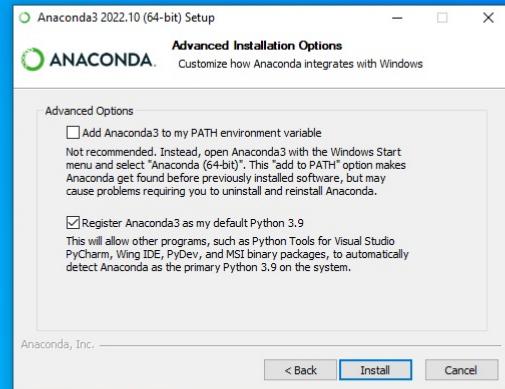
16°C Cloudy 27/10/2022 19:05



Recycle Bin



Microsoft Edge



Type here to search



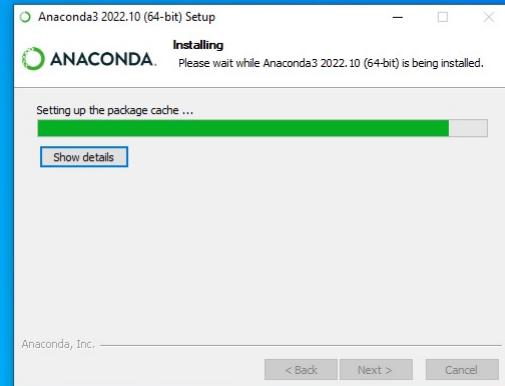
16°C Cloudy 19:05  
27/10/2022



Recycle Bin



Microsoft Edge



Type here to search



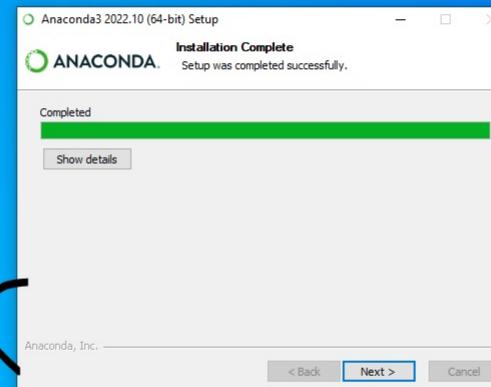
16°C Cloudy 19:05  
27/10/2022



Recycle Bin



Microsoft Edge



# ANACONDA.NAVIGATOR

Connect ▾

[Home](#)[Environments](#)[Learning](#)[Community](#)

Building Data Science Solutions with Anaconda  
Available on Amazon

[Documentation](#)[Anaconda Blog](#)

All applications on base (root) Channels

In one user-friendly environment

[Install](#)[Launch](#)[Launch](#)[Launch](#)[Launch](#)[Launch](#)

5.2.2

Scientific Python Development Environment. Powerful Python IDE with advanced editing, interactive testing, debugging and introspection features

[Launch](#)

Kick-start your data science projects in seconds in a pre-configured environment. Enjoy coding assistance for Python, SQL, and R in Jupyter notebooks and benefit from no-code automations. Use Datalore online for free.

[Launch](#)

Deepnote is a notebook built for collaboration. Create notebooks in your browser, spin up your conda environment in seconds and share with a link.

[Launch](#)

IBM Watson Studio Cloud provides you the tools to analyze and visualize data, to cleanse and shape data, to create and train machine learning models. Prepare data and build models, using open source data science tools or visual modeling.

[Launch](#)

Oracle Data Science Service

OCI Data Science offers a machine learning platform to build, train, manage, and deploy your machine learning models on the cloud with your favorite open-source tools

[Launch](#)

1.0.0

Multidimensional data visualization across files. Explore relationships within and among related datasets.

[Install](#)

3.32.0

Component based data mining framework. Data visualization and data analysis for novice and expert. Interactive workflows with a large toolbox.

[Install](#)

PyCharm Professional

A full-fledged IDE by JetBrains for both Scientific and Web Python development. Supports HTML, JS, and SQL.

[Install](#)

1.1.456

A set of integrated tools designed to help you be more productive with R. Includes R essentials and notebooks.

[Install](#)

Type here to search



16°C Cloudy



19:39

27/10/2022

JB PyCharm: the Python IDE for Pro x +

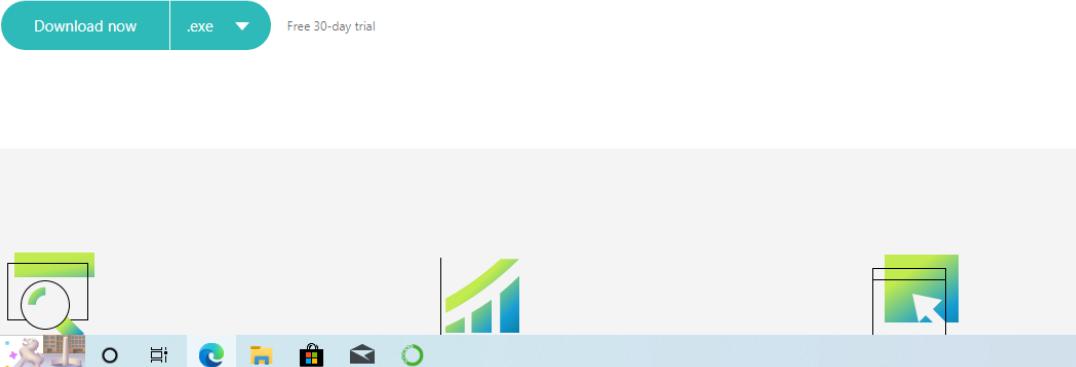
https://www.jetbrains.com/pycharm-anaconda/?=

JET BRAINS — English Not syncing

# Ready to go Pro?

Working with Python and Jupyter notebooks is a breeze with PyCharm Professional. Download now and have the best tools for data scientists and data engineers at your fingertips

Download now .exe ▾ Free 30-day trial



ANACONDA

[X] Cookies and IP addresses allow us to deliver and improve our web content and to provide you with a personalized experience. Our website uses cookies and collects your IP address for these purposes. [Learn more](#)

=-----|  
| JetBrains may use cookies and my IP address to collect individual statistics and to provide me with personalized offers and ads subject to the [Privacy Policy](#) and the [Terms of Use](#). JetBrains may use [third-party services](#) for this purpose. I can revoke my consent at any time by visiting the [Opt-Out page](#).  
[Y]es, I agree [N]o, thanks
~ root#

16°C Cloudy 27/10/2022



Recycle Bin



Microsoft  
Edge



Anaconda3...



pycharm-pro  
fessional 202  
2.2.3



Type here to search



16°C Cloudy



19:41  
27/10/2022



Recycle Bin



Microsoft Edge

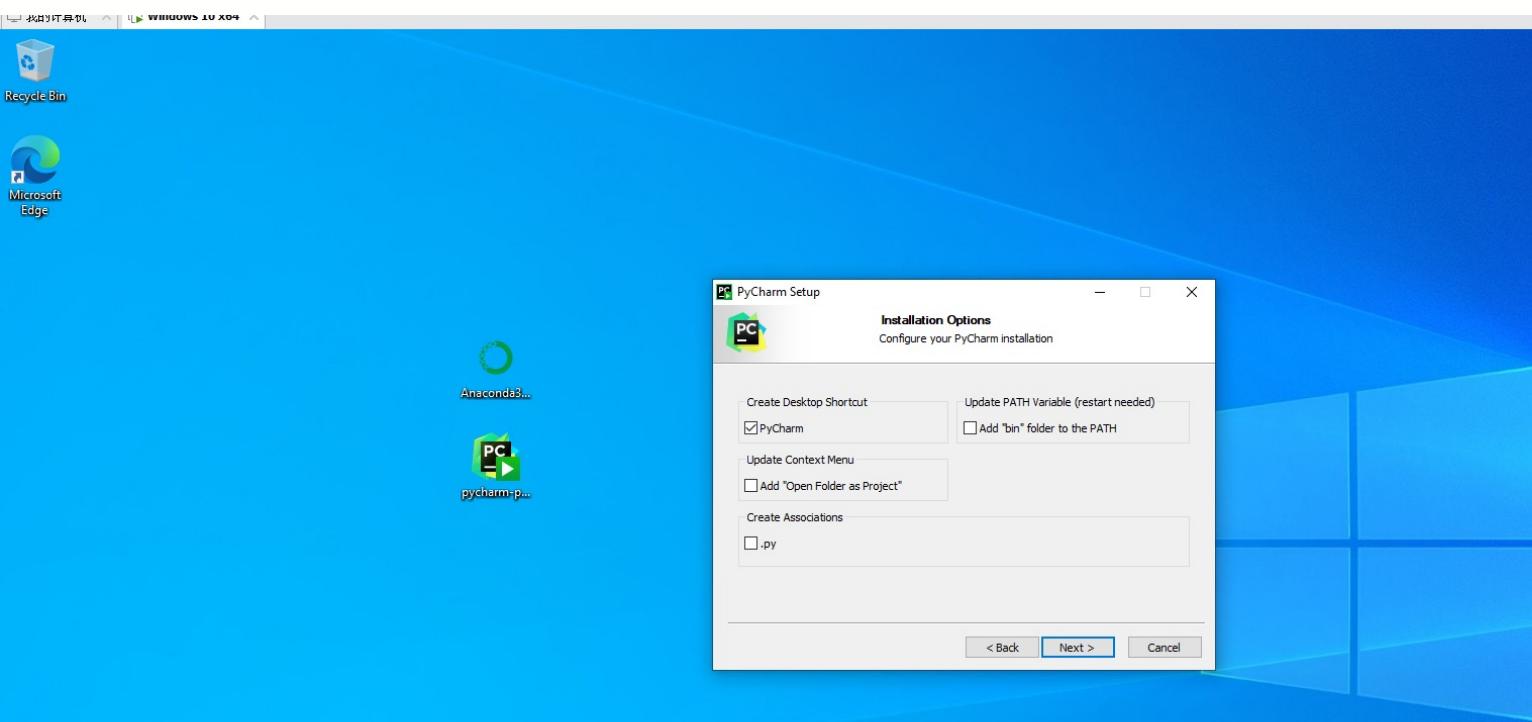


Anaconda3...



pycharm-p...







Recycle Bin



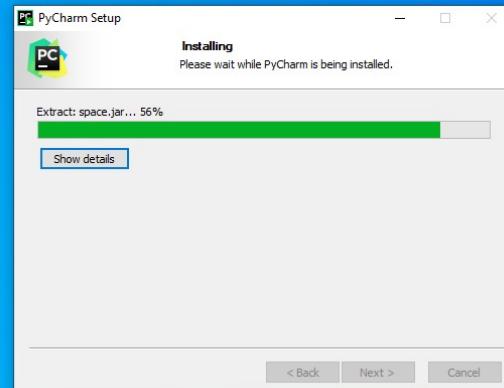
Microsoft Edge

PyCharm  
2022.2.3

Anaconda3...



pycharm-p...



Type here to search



16°C Cloudy 19:43  
27/10/2022



Recycle Bin



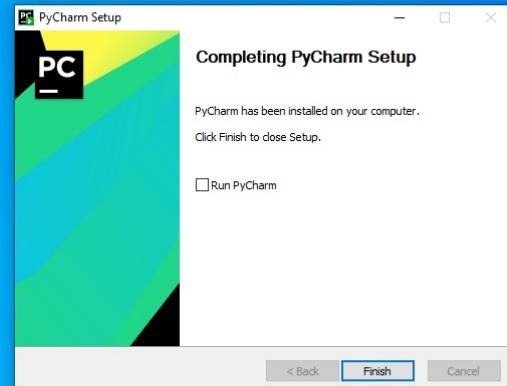
Microsoft Edge

PyCharm  
2022.2.3

Anaconda3...



pycharm-p...



Type here to search



16°C Cloudy 19:43  
27/10/2022

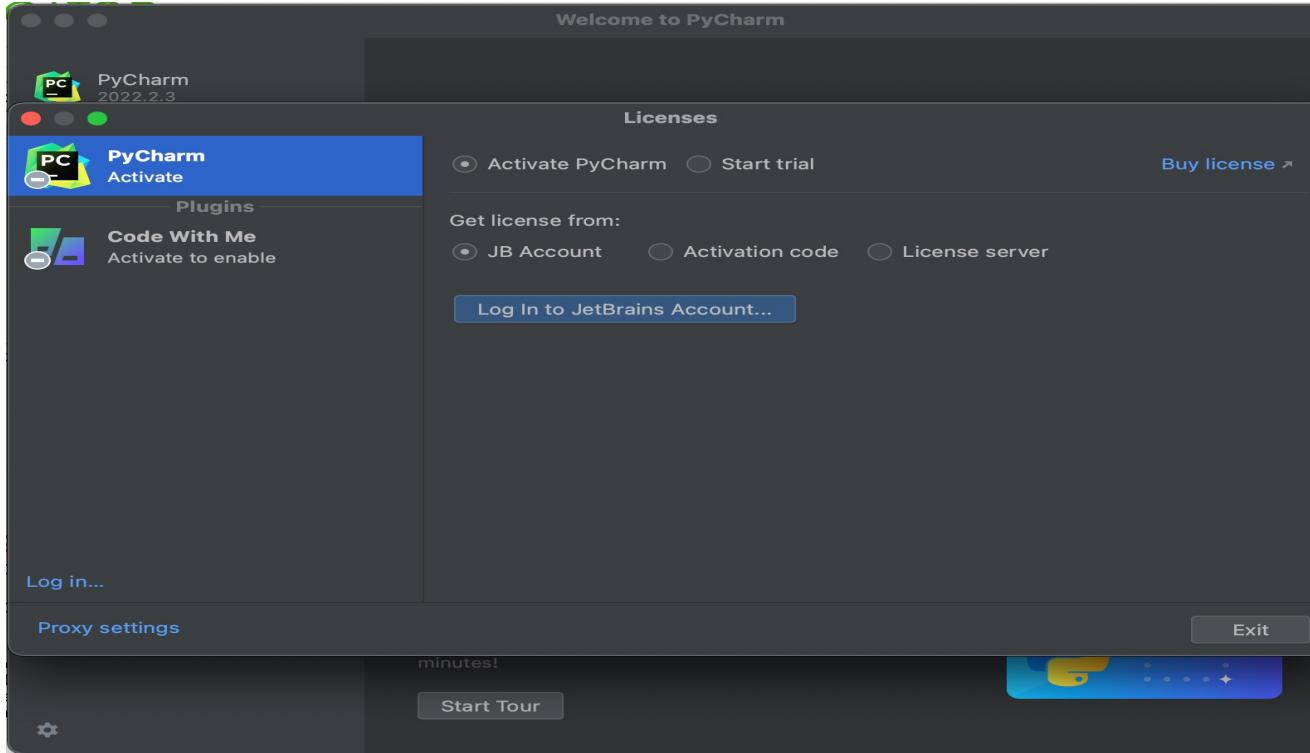


- Installation successful.
- Let's get started with Pycharm!



# Activate Pycharm

- Active your account with your username and password.
- You can also select free trial if you don't want to activate Pycharm.



[Sign In](#) [Create Account](#)

Username



Password

[Sign In](#)

[Forgot password?](#)

---

Or sign in with:

 Google



## **Authorization Successful**

You can close this page and return to the IDE



University  
of Glasgow

# Create a new project

Welcome to PyCharm

Welcome to PyCharm

Create a new project to start from scratch.  
Open existing project from disk or version control.

New Project    Open    Get from VCS

Take a quick onboarding tour

Get familiar with the PyCharm user interface and learn how to code in Python with smart assistance in just 7 minutes!

Start Tour

Projects

Remote Development (Beta)

- SSH
- JetBrains Space

Customize

Plugins

Learn PyCharm



# Setup an interpreter

New Project

Location: /Users/python/PycharmProjects/PythonWorkshop

▼ Python Interpreter: New Conda environment

New environment using Conda

Location: /Users/python/.conda/envs/PythonWorkshop

Python version: 3.10

Conda executable: /Users/python/opt/anaconda3/bin/conda

Make available to all projects

Previously configured interpreter

Interpreter: Python 3.10 (PythonWorkshop) ~/conda/envs/PythonWorkshop/bin/python

Create a main.py welcome script

Create a Python script that provides an entry point to coding in PyCharm.

⚡ Environment location directory is not empty

Create



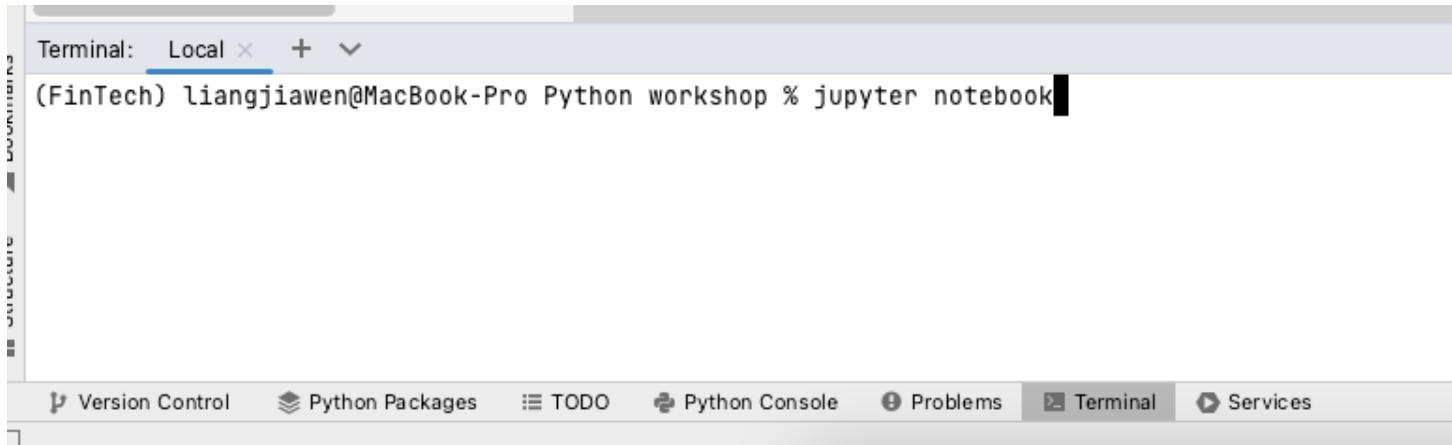
# Jupyter notebook

1. Terminal > type “pip install jupyter”>Click Enter

A screenshot of a terminal window within a code editor interface. The terminal tab is selected, showing the text "(FinTech) liangjiawen@MacBook-Pro Python workshop % pip install jupyter". The code editor has a sidebar with "Bookmarks" and "Structure" options, and a bottom navigation bar with icons for Version Control, Python Packages, TODO, Python Console, Problems, Terminal, and Services.

# Jupyter notebook

2. You can choose to open notebook on web by typing “jupyter notebook” in your terminal



The image shows a screenshot of a terminal window. The title bar says "Terminal: Local". The command "(FinTech) liangjiawen@MacBook-Pro Python workshop % jupyter notebook" is typed into the terminal. Below the terminal window is a navigation bar with icons for Version Control, Python Packages, TODO, Python Console, Problems, Terminal, and Services.

Version Control    Python Packages    TODO    Python Console    Problems    Terminal    Services



University  
of Glasgow

# Jupyter notebook

## 3. Web surface

The screenshot shows the Jupyter Notebook web interface. At the top, there is a header with the Jupyter logo, the word "jupyter", and two buttons: "Quit" and "Logout". Below the header, there is a navigation bar with three tabs: "Files" (which is selected), "Running", and "Clusters". A message "Select items to perform actions on them." is displayed above the file list. The file list itself has a header row with columns for selection, name, last modified, and file size. There are two items listed:

	Name	Last Modified	File size
<input type="checkbox"/>	Python for data analysis	5 hours ago	
<input type="checkbox"/>	Python Basics.ipynb	2 minutes ago	581 B



University  
of Glasgow

# Jupyter notebook

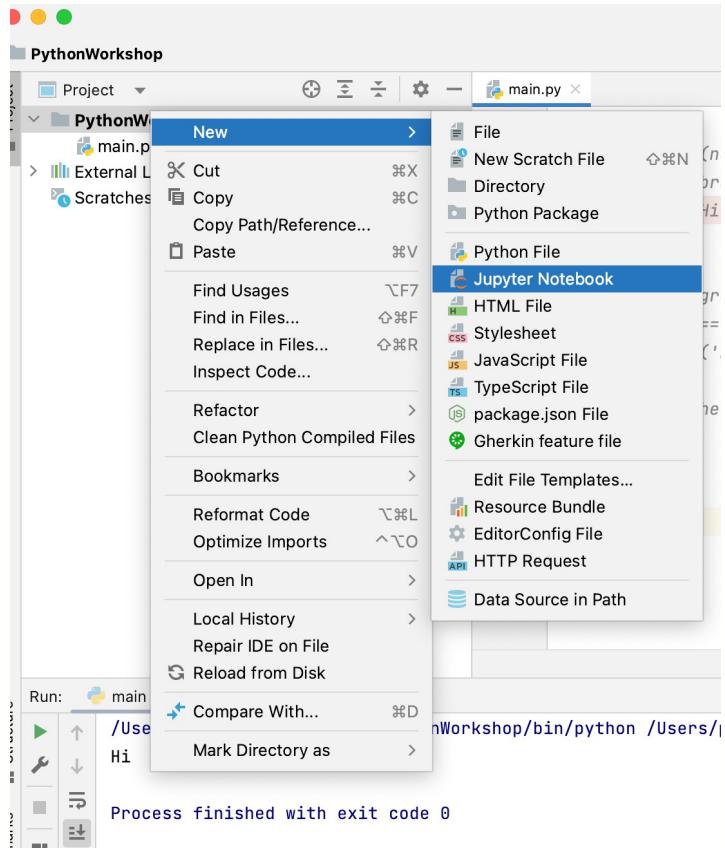
## 4. Create file on web

The screenshot shows the Jupyter Notebook web interface. At the top, there is a header with the Jupyter logo, the word "jupyter", and buttons for "Quit" and "Logout". Below the header, there is a navigation bar with tabs for "Files", "Running", and "Clusters". The "Files" tab is selected. A message "Select items to perform actions on them." is displayed above a file list. The file list shows a directory structure with 0 files and 1 folder named "/". Inside the "/" folder, there are two items: "Python for data analysis" and "Python Basics.ipynb". On the right side of the interface, there is a sidebar with a "New" button. A dropdown menu is open from the "New" button, showing options for creating a new notebook or file. The "Notebook" option is selected, showing a sub-menu with "Python 3 (ipykernel)" highlighted. Other options in the "Notebook" submenu include "Text File" (3 B) and "Folder". Below the "Notebook" submenu, there are options for "Terminal" and "Cluster".

# Jupyter notebook

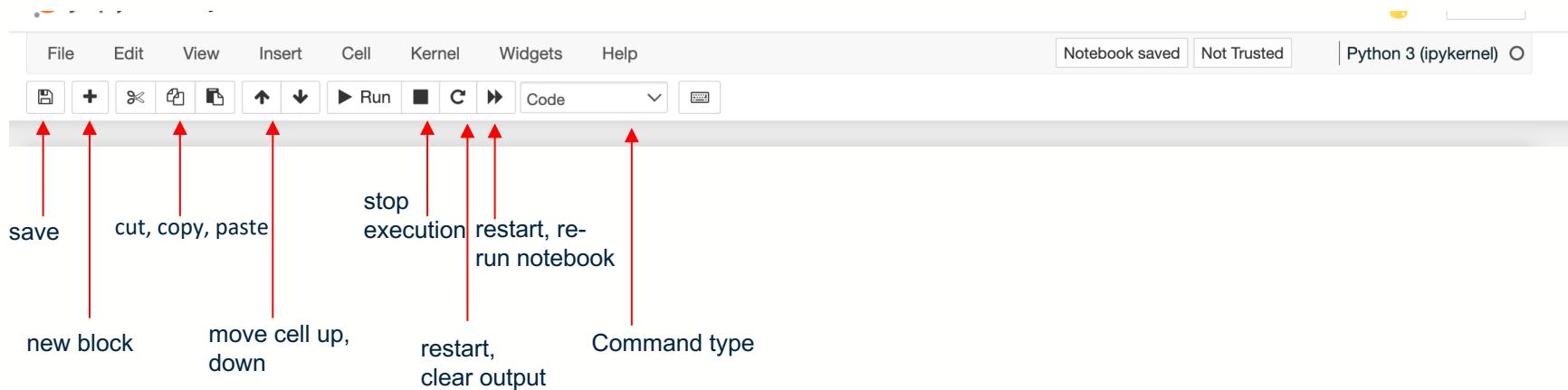
Another option to open jupyter notebook:

1. create a new jupyter notebook in Pycharm.
2. Select and double click the ipynb that you want to open in Pycharm.





# Jupyter notebook





## Running the cells

- By clicking the Run button
- Shift + Enter



University  
of Glasgow

# Thanks for attending this session!

