

# DATA MINING TUTORIAL

---

Introduction to Python Libraries

# Python

- In the last few years there is an increasing community that creates **Data Mining tools in Python**
  - Python is overwhelmingly used today for data science tasks
  - It is also heavily used in industry
  - We will use Python for this class.
- There are tons of resources online for Python.
- For an introduction you can also look at the slides of the [Introduction to Programming](#) course by prof. N. Mamoulis
- I assume you have installed Python to your laptop by now and you have a good knowledge of programming in python.

# Anaconda

- Installing libraries in Python can be complicated, so you should download the **Anaconda Scientific Python** distribution which will install most of the libraries that we will use.
  - Use Python 3.0
- Installing Anaconda installs a lot of libraries and also:
  - Anaconda Navigator
  - **Jupyter Notebook**: An interactive web-based interface for running python.
  - Anaconda Powershell: terminal for running commands

# Jupyter Notebook

- Installing Anaconda will also install Jupyter Notebook.
- If you wish to install it in a different way, together with the relevant libraries you are free to do so.
- We will use Notebook for our examples and it is **required** for the assignments.
  - In almost all assignments you are required to submit a Notebook.

File

All Apps Documents Email Web More

Feedback ...

Paste

Clipboard

2

3

4


5

6






7

Slide 5 of


**Best match**


 **Anaconda Navigator (Anaconda3)**  
App

**Apps**







-  **Anaconda Powershell Prompt (Anaconda3)** >
-  **Anaconda Prompt (Anaconda3)** >
-  **Jupyter Notebook (Anaconda3)** >
-  **Spyder (Anaconda3)** >
-  **Reset Spyder Settings (Anaconda3)** >

**Search the web**

-  **anac** - See web results >



**Anaconda Navigator (Anaconda3)**  
App

-  Open
-  Run as administrator
-  Open file location
-  Pin to Start
-  Pin to taskbar
-  Uninstall

anacnda Navigator (Anaconda3)

Taskbar icons: Windows, Search, Task View, Edge, File Explorer, Mail, Chrome, Word, Jupyter Notebook, Anaconda Navigator, PowerPoint

# The Anaconda Navigator


Anaconda Navigator


File Help


 ANACONDA NAVIGATOR

Sign in to Anaconda Cloud

 Home

 Environments

 Learning

 Community

Documentation

Developer Blog



Applications on

base (root)

Channels

Refresh



JupyterLab

1.0.2

An extensible environment for interactive and reproducible computing, based on the Jupyter Notebook and Architecture.

Launch



Jupyter Notebook

6.0.0

Web-based, interactive computing notebook environment. Edit and run human-readable docs while describing the data analysis.

Launch

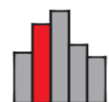


Spyder

3.3.6

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

Launch



Glueviz

0.15.2



Orange 3

3.23.1



RStudio

1.1.456

File

All Apps Documents Email Web More

Feedback ...

Paste

Clipboard

5

6

7


8

9


10


Slide 7 of 10


**Best match**

 **Anaconda Navigator (Anaconda3)** >  
App


**Apps**

 **Anaconda Prompt (Anaconda3)** >


 **Jupyter Notebook (Anaconda3)**

 **Anaconda Powershell Prompt (Anaconda3)** >







**Search the web**

 ana - See web results >

**Folders (1+)**



**Jupyter Notebook (Anaconda3)**  
App

-  Open
-  Run as administrator
-  Open file location
-  Pin to Start
-  Pin to taskbar
-  Uninstall

Windows taskbar: Jupyter Notebook (Anaconda3)

Taskbar icons: File Explorer, Edge, Mail, Chrome, Anaconda Navigator, Adobe Reader, Teams, PowerPoint

Files Running Clusters

Select items to perform actions on them.



Upload New ↕ ↻

<input type="checkbox"/> 0	Name ↓	Last Modified	File size
<input type="checkbox"/> /			
<input type="checkbox"/> 3D Objects		21 days ago	
<input type="checkbox"/> Contacts		21 days ago	
<input type="checkbox"/> Documents		3 months ago	
<input type="checkbox"/> Downloads		4 hours ago	
<input type="checkbox"/> Dropbox (Personal)		8 days ago	
<input type="checkbox"/> Dropbox (Uol)		2 months ago	
<input type="checkbox"/> Favorites		10 days ago	
<input type="checkbox"/> Google Drive		3 days ago	
<input type="checkbox"/> Links		21 days ago	
<input type="checkbox"/> Music		21 days ago	
<input type="checkbox"/> OneDrive		2 days ago	
<input type="checkbox"/> Roaming		5 months ago	
<input type="checkbox"/> Saved Games		21 days ago	
<input type="checkbox"/> Searches		21 days ago	







← All Apps Documents Email Web More Feedback ...


**Best match**


-  **Anaconda Powershell Prompt (Anaconda3)**  
App
-  **Anaconda Navigator (Anaconda3)** >  
App

**Apps**

-  **Anaconda Prompt (Anaconda3)** >
-  **Jupyter Notebook (Anaconda3)** >
-  **Spyder (Anaconda3)** >
-  **Reset Spyder Settings (Anaconda3)** >





**Search the web**

-  **anac** - See web results >






### Anaconda Powershell Prompt (Anaconda3)

App

-  Open
-  Run as Administrator
-  Run ISE as Administrator
-  Windows PowerShell ISE

↓

0 \$

  Anaconda Powershell Prompt (Anaconda3) | 

Anaconda Powershell Prompt (Anaconda3)



```
(base) PS C:\Users\tsapa> █
```

# Installing Packages

- You can install packages from the Anaconda terminal using the command:
  - `conda install <name of package>`
- For example, [Seaborn](#) is a package for Statistical Data Visualization.
  - `conda install seaborn`
- [panda-datareader](#) is a package for loading online datasets.
  - `conda install pandas-datareader`

# Notebooks

- Jupyter Notebook offers an interactive web-based interface for running code.
- The Notebook runs inside a browser.
- It allows you to interact with the code, running different parts of the code
- The results also appear in the browser, so you can have together the code and the results
- You can also add text, commenting on the results.
- We will now see some details on how to create notebooks

# Changing the notebook default directory

- This used to be important before but now Jupyter Notebook takes you to the home directory
- From the Anaconda terminal type the command:  
➤ `jupyter notebook --generate-config`
- This will generate `.jupyter/jupyter_notebook_config.py` file under your home directory.
- Find, un-comment and modify the line  
`# c.NotebookApp.notebook_dir = ''` in the config file to point to the desired directory

Select items to perform actions on them.

Upload New ↕ ↻

<input type="checkbox"/>	0	▼	📁 /	Name	
<input type="checkbox"/>	📁			3D Objects	
<input type="checkbox"/>	📁			Contacts	
<input type="checkbox"/>	📁			Documents	
<input type="checkbox"/>	📁			Downloads	
<input type="checkbox"/>	📁			Dropbox (Personal)	10 months ago
<input type="checkbox"/>	📁			Dropbox (Uol)	2 months ago
<input type="checkbox"/>	📁			Favorites	8 months ago
<input type="checkbox"/>	📁			gensim-data	a year ago
<input type="checkbox"/>	📁			Google Drive	4 days ago
<input type="checkbox"/>	📁			IdeaProjects	18 days ago
<input type="checkbox"/>	📁			Links	8 months ago
<input type="checkbox"/>	📁			Music	8 months ago
<input type="checkbox"/>	📁			OneDrive	2 days ago
<input type="checkbox"/>	📁			OneDrive - ΠΑΝΕΠΙΣΤΗΜΙΟ ΙΩΑΝΝΙΝΩΝ	2 days ago
<input type="checkbox"/>	📁			Roaming	2 years ago

Notebook:  
**Python 3**  
Other:  
Text File  
Folder  
Terminal

Home Page - Select or create a n x Untitled1 - Jupyter Notebook x +

localhost:8888/notebooks/Untitled1.ipynb?kernel\_name=python3

Fair Information Ac... Data Mining

jupyter Untitled1 Last Checkpoint: a few seconds ago (unsaved changes)

Logout

File Edit View Insert Cell Kernel Widgets Help

Trusted Python 3

Code

In [ ]: |

- The notebook is organized in cells
- In each cell you can write either code or text
- The default behavior is code

The image shows a Jupyter Notebook interface in a browser. The browser address bar displays the URL `localhost:8888/notebooks/Untitled1.ipynb?kernel_name=python3`. The Jupyter interface includes a menu bar with options: File, Edit, View, Insert, Cell, Kernel, Widgets, Help. A toolbar below the menu bar contains various icons, with the 'Run' button (a play icon) circled in red. The notebook content shows a code cell with the input `In [1]: print('hello world')` and the output `hello world`. Below the code cell is an empty input field for the next cell, labeled `In [ ]:`. The interface also shows a 'Logout' button and a 'Python 3' kernel indicator.

- You can run the code using the Run button or with Ctrl+Enter
- Note that now we have both the code and the output in the notebook



Home Page - Select or create a n... x Untitled1 - Jupyter Notebook x +

localhost:8888/notebooks/Untitled1.ipynb?kernel\_name=python3

Fair Information Ac... Data Mining

jupyter Untitled1 Last Checkpoint: 27 minutes ago (autosaved) Logout

File Edit View Insert Cell Kernel Widgets Help Notebook saved Trusted Python 3

In [1]: `print('hello world')`  
hello world

**## Running the code**

- You can run the code using the Run button or with Ctrl+Enter
- Note that now we have both the code and the output in the notebook

- You can also write text in Markdown language
  - You can combine HTML, and Latex, and there are some other commands
- You can learn more about Markdown by searching online, e.g.:
  - [Learn How to Write Markdown & LaTeX in The Jupyter Notebook | by Khelifi Ahmed Aziz | Towards Data Science](#)
- You need to Run the Markdown cell as well

Run button circled in red

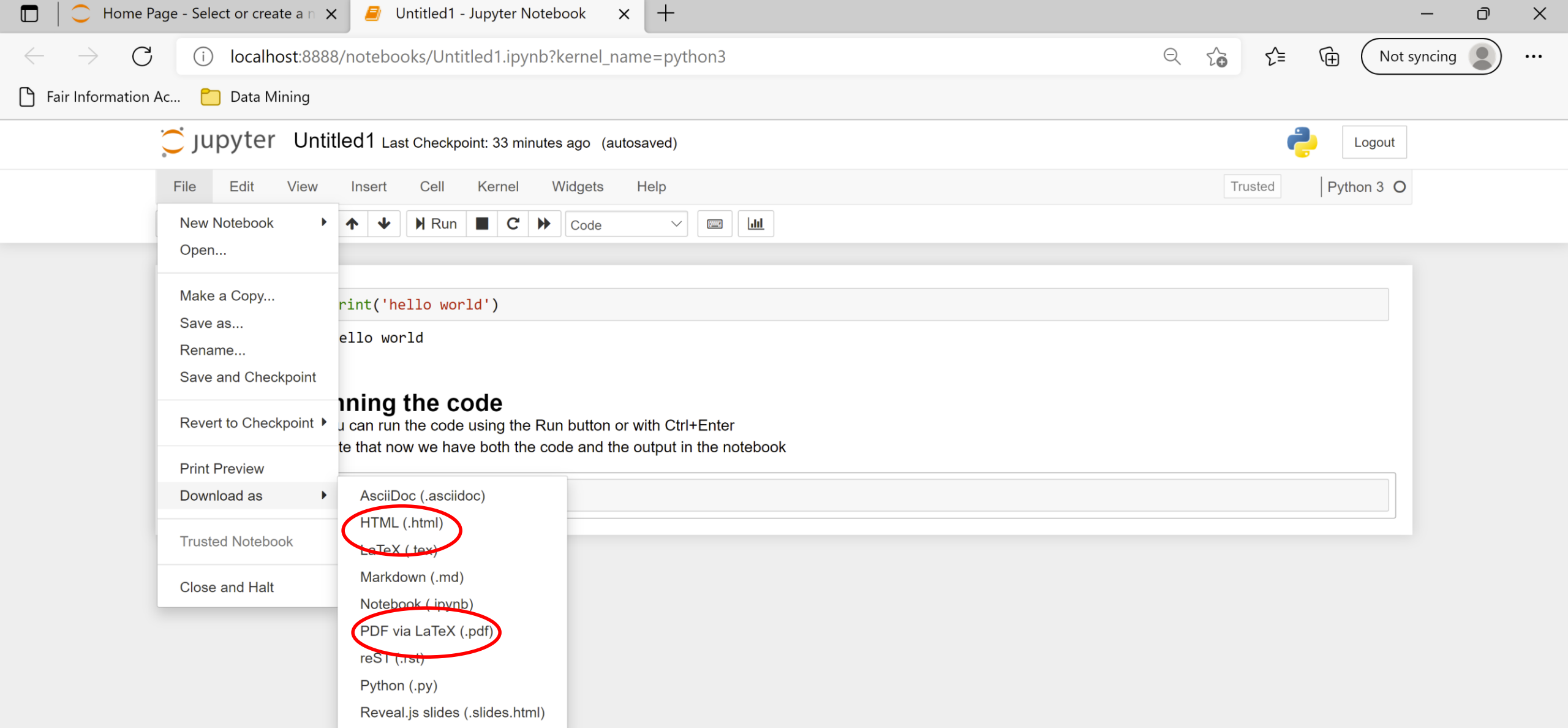
```
In [1]: print('hello world')
```

hello world

### Running the code

- You can run the code using the Run button or with Ctrl+Enter
- Note that now we have both the code and the output in the notebook

```
In [ ]:
```



- You can export the notebook into HTML or PDF

# Attention!

- A notebook is run interactively, each time running a specific cell
- The state of the program remains in memory while the notebook is running
- Each cell has access to the current state of the memory
- You can jump between cells in a non-linear way
- You should be aware of the state of the memory of the notebook when you run a specific cell.

# A simple example

jupyter Untitled1 Last Checkpoint: a few seconds ago (autosaved) Python 3 Logout

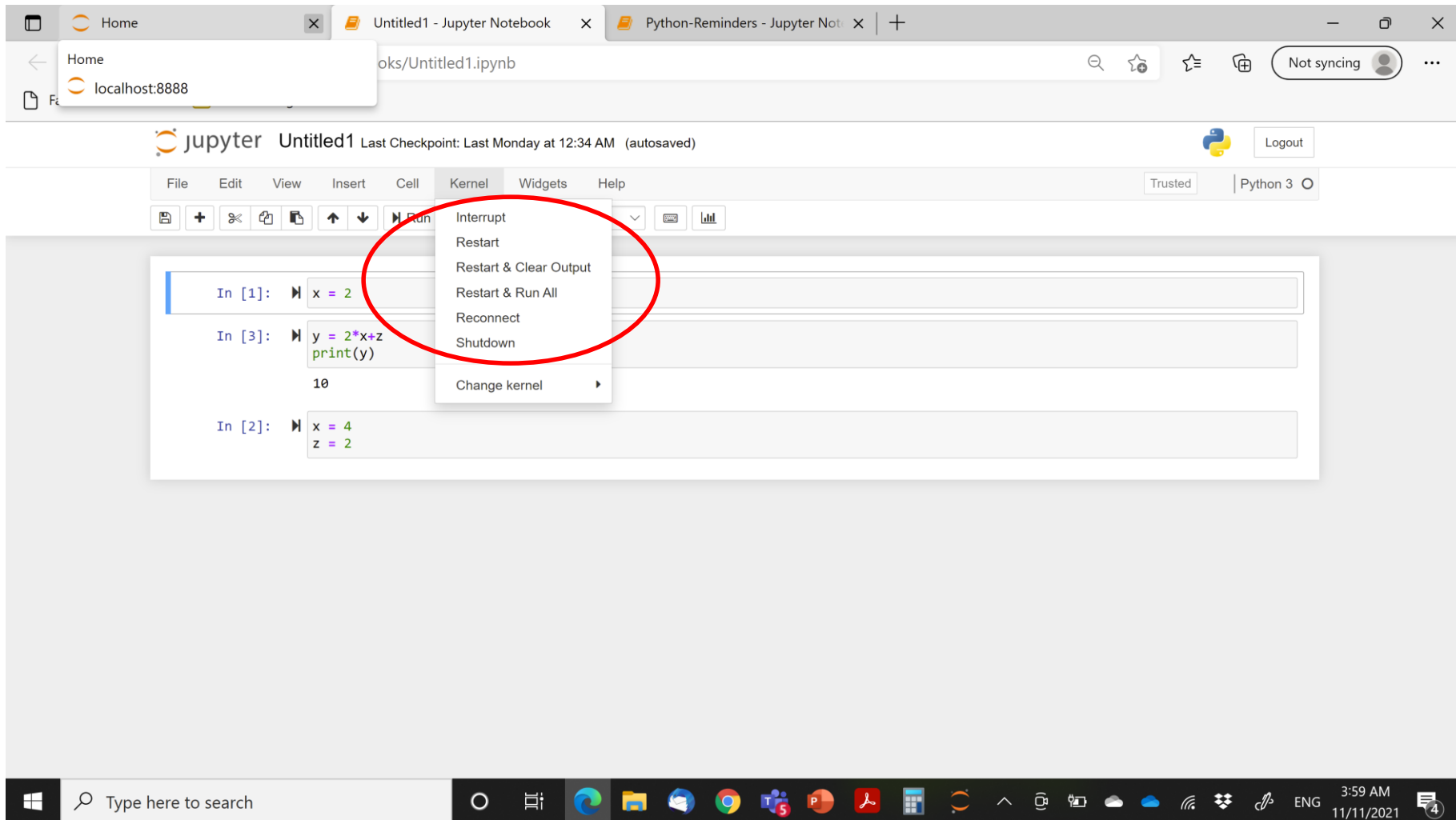
File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

Run Code

```
In [1]: x = 2
In [3]: y = 2*x+z
         print(y)
         10
In [2]: x = 4
         z = 2
```

- The order in which the cells are executed is shown in the increasing numbers (not always useful)
- The second in order cell is executed third
  - So it has access to z, and uses the value 4 for x

# Restarting



The screenshot displays a Jupyter Notebook interface in a web browser. The browser tabs include 'Home', 'Untitled1 - Jupyter Notebook', and 'Python-Reminders - Jupyter Notebook'. The address bar shows 'localhost:8888'. The Jupyter interface includes a menu bar with 'File', 'Edit', 'View', 'Insert', 'Cell', 'Kernel', 'Widgets', and 'Help'. The 'Kernel' menu is open, showing options: 'Interrupt', 'Restart', 'Restart & Clear Output', 'Restart & Run All', 'Reconnect', 'Shutdown', and 'Change kernel'. A red circle highlights the 'Restart' and 'Restart & Clear Output' options. The notebook content shows three code cells:

```
In [1]: x = 2
```

```
In [3]: y = 2*x+z
print(y)
10
```

```
In [2]: x = 4
z = 2
```

The Windows taskbar at the bottom shows the search bar, system tray, and various application icons. The system clock indicates 3:59 AM on 11/11/2021.