# DATA MINING TUTORIAL

Introduction to Python Libraries

# Python

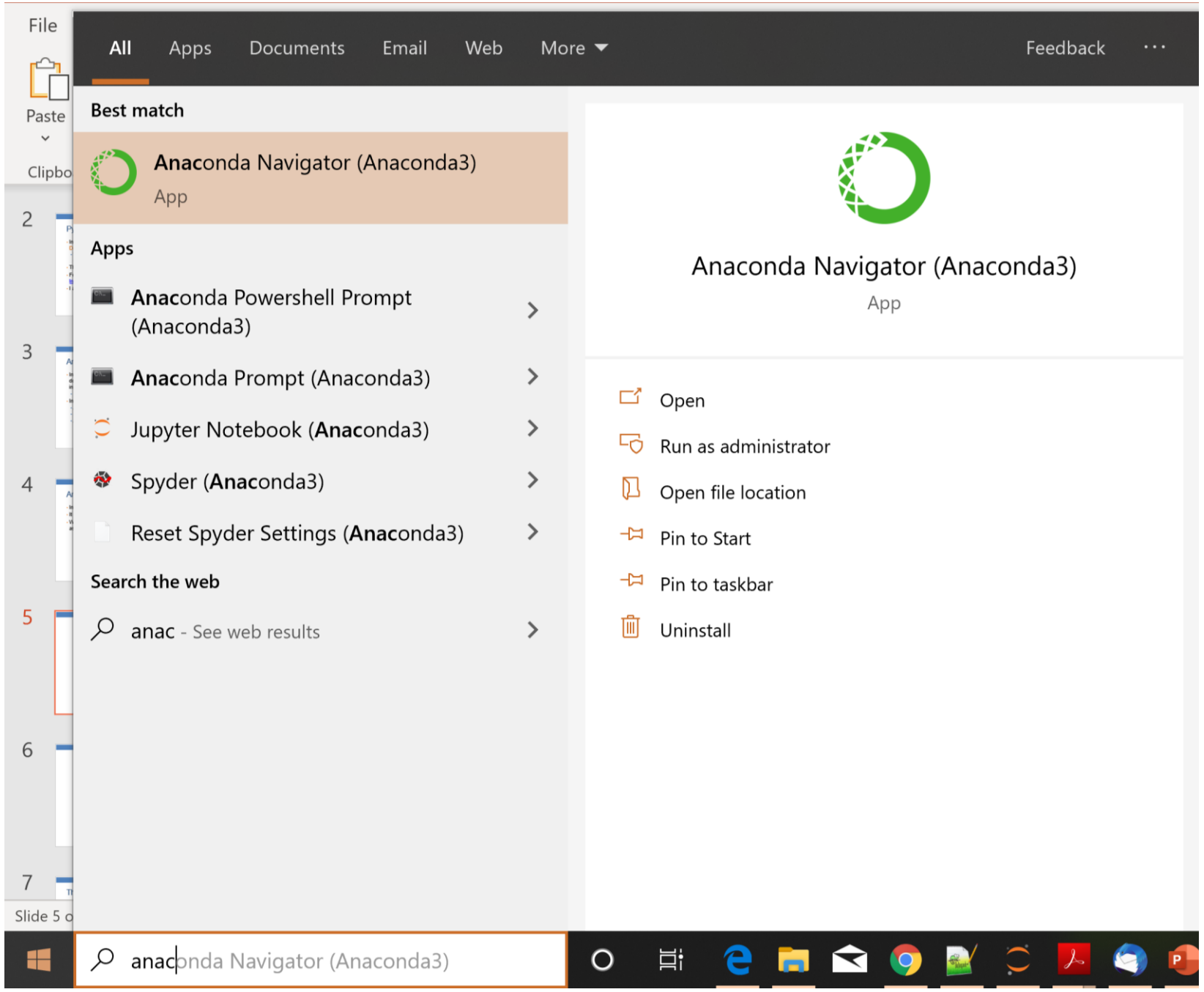
- In the last few years there is an increasing community that creates Data Mining tools in Python
  - Python is overwhelmingly used today for data science tasks
  - It is also heavily used in industry
  - We will use Python for this class.
- There are tons of resources online for Python.
- For an introduction 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

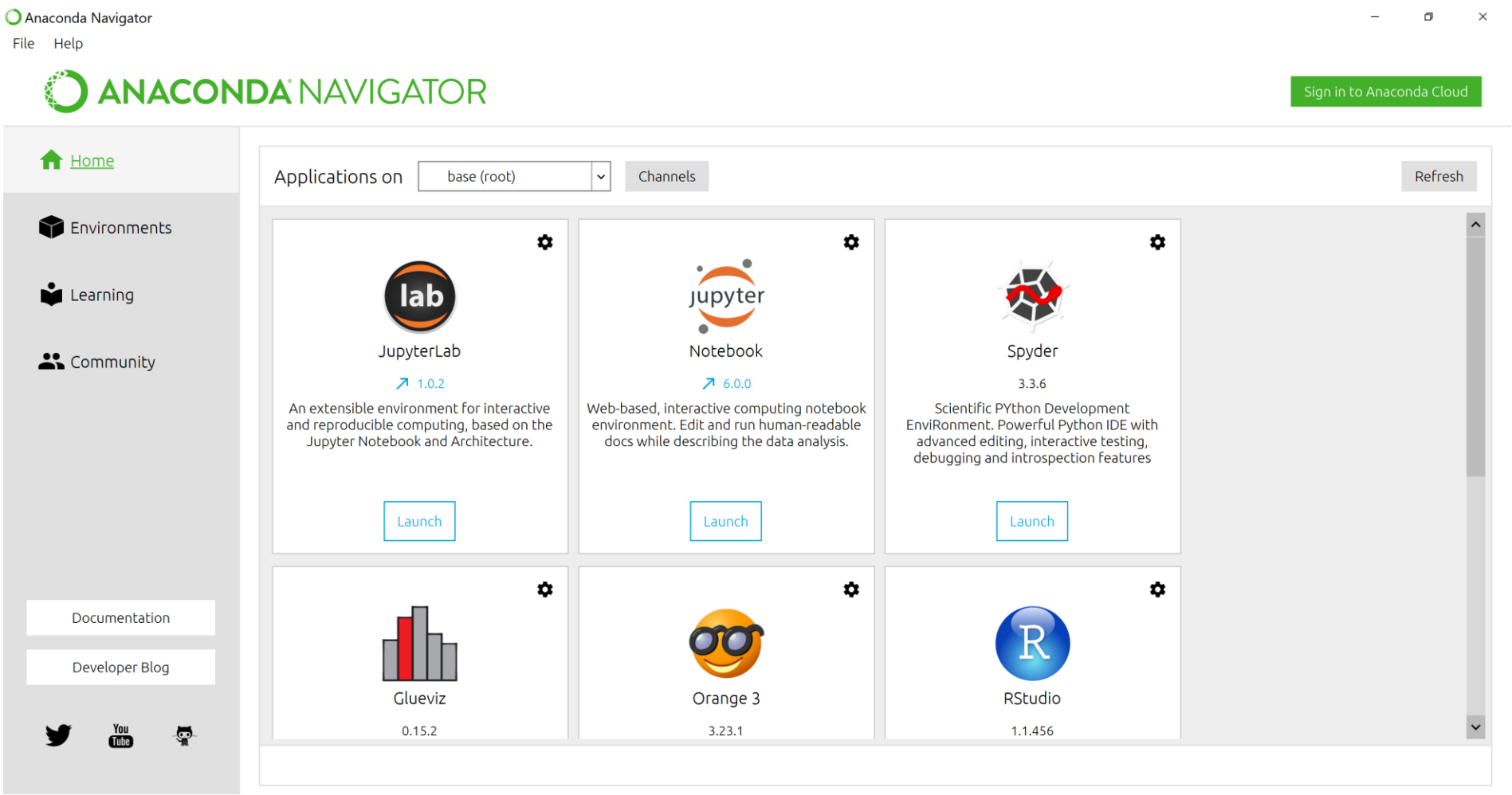
- Installing libraries in Python can be complicated, so you should download the Anaconda Scientific Python distribution which will install most of the libraries that we will use.
  - Use Python 3.0
- Installing Anaconda installs a lot of libraries and also:
  - Anaconda Navigator
  - Jupyter Notebook: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

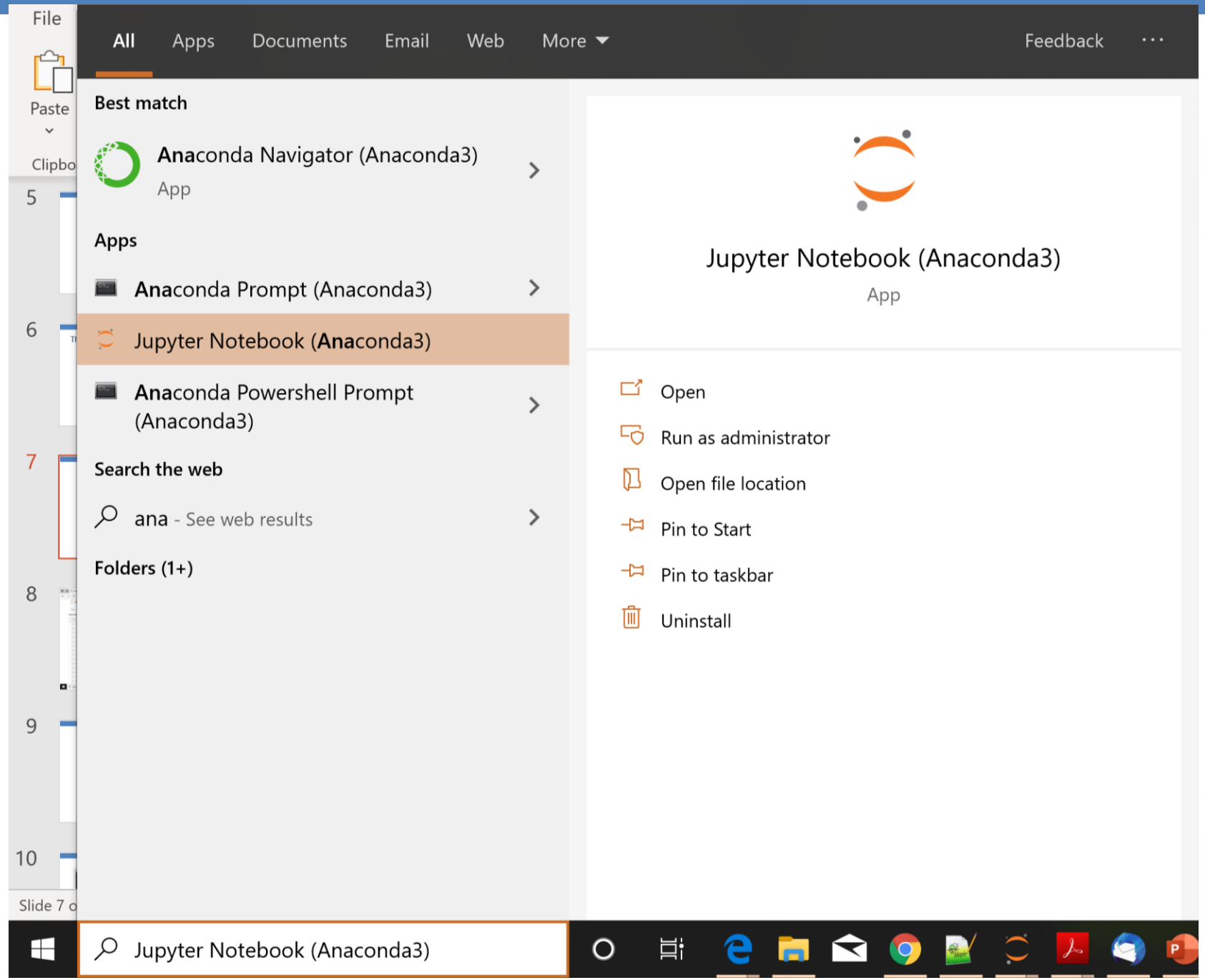
# Jupyter Notebook

- 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
Paste
Clipbo
2
3
4
5
6
7
Slide 5 o
All
Apps
Documents
Email
Web
More
Feedback
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
anaconda Navigator (Anaconda3)

# The Anaconda Navigator




File
Paste
Clipbo
5
6
7
8
9
10
Slide 7 o
All
Apps
Documents
Email
Web
More
Feedback
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
Jupyter Notebook (Anaconda3)

Home Page - Select or ( 

localhost:8889/tree

jupyter

Quit Logout

Files Running Clusters

Select items to perform actions on them.

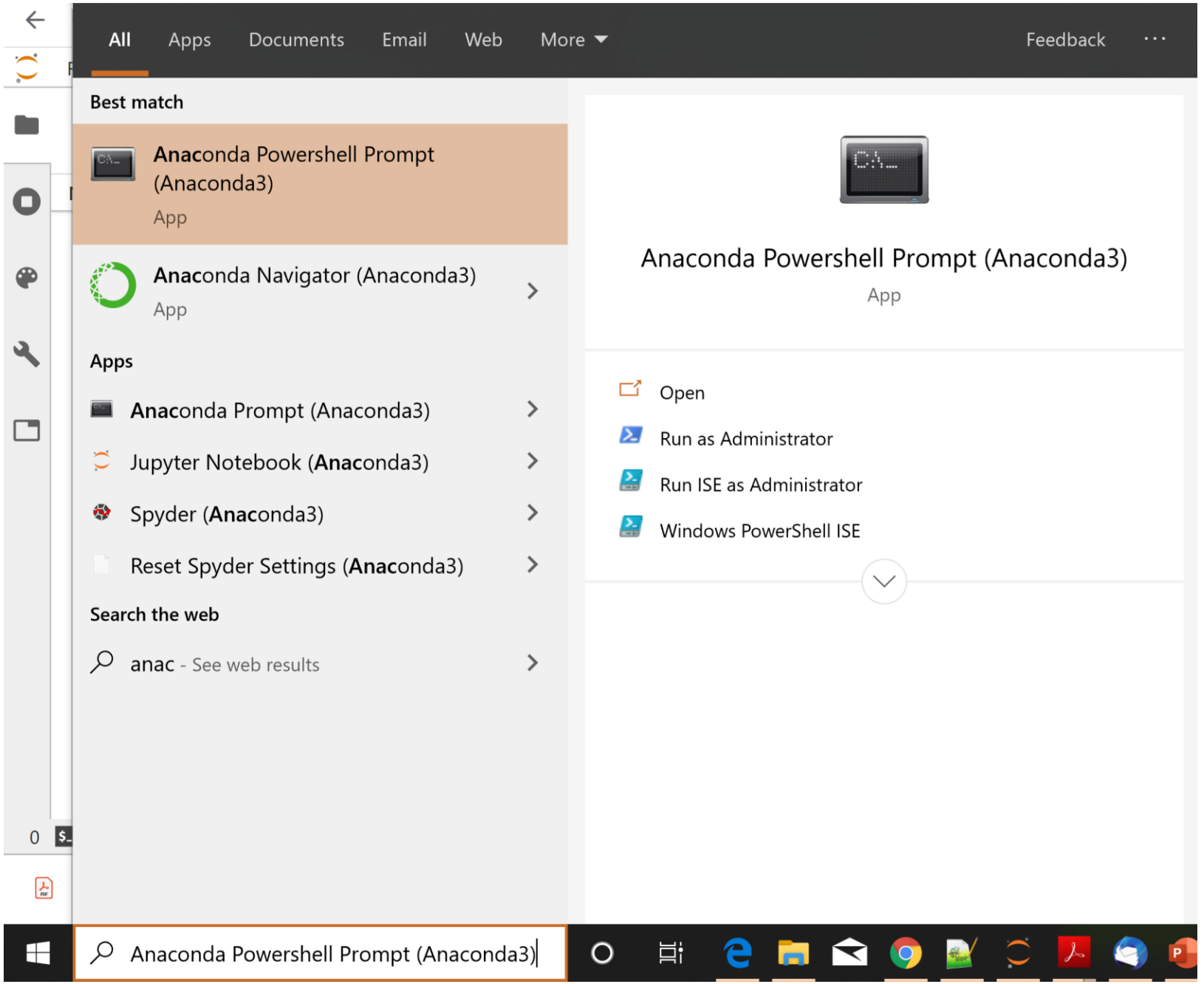
Upload New

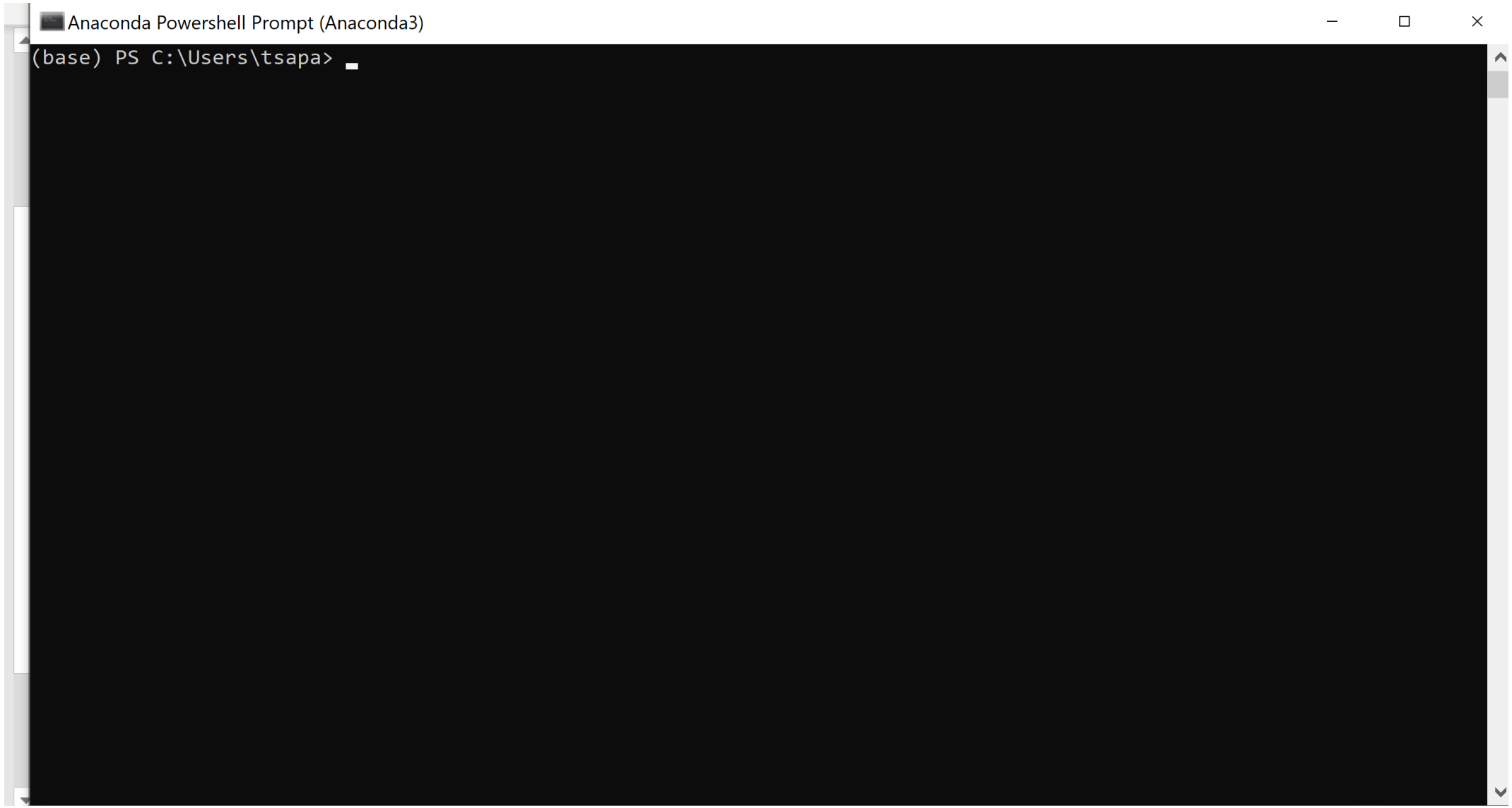
0 / Name Last Modified File size

| | | |
|---|---|---|
| 3D Objects | | 21 days ago |
| Contacts | | 21 days ago |
| Documents | | 3 months ago |
| Downloads | | 4 hours ago |
| Dropbox (Personal) | | 8 days ago |
| Dropbox (UoI) | | 2 months ago |
| Favorites | | 10 days ago |
| Google Drive | | 3 days ago |
| Links | | 21 days ago |
| Music | | 21 days ago |
| OneDrive | | 2 days ago |
| Roaming | | 5 months ago |
| Saved Games | | 21 days ago |
| Searches | | 21 days ago |

Type here to search

ENG 11:09 PM 10/31/2019


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
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

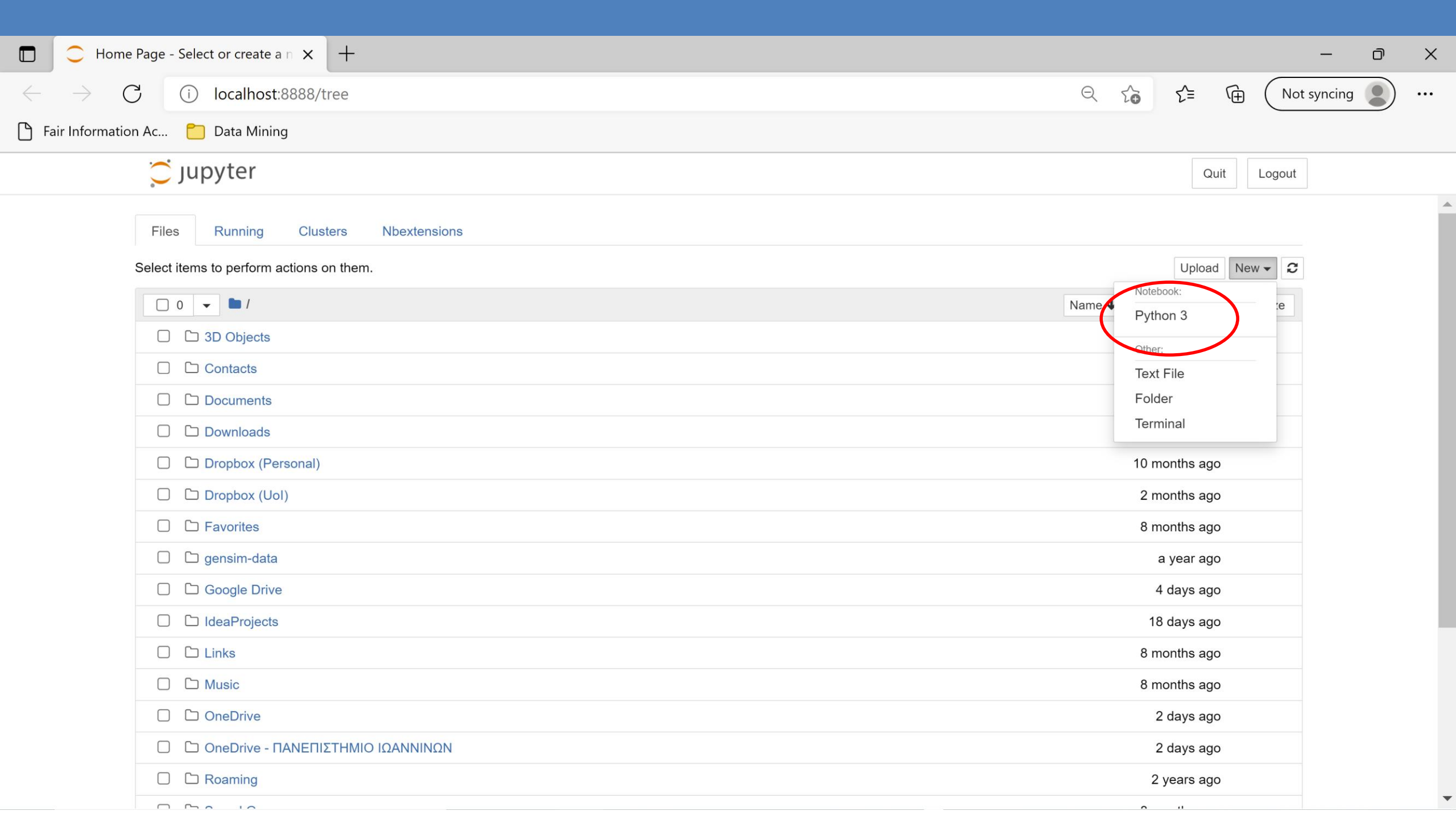
- This used to be important before but now Jupyter Notebook takes you to the home directory

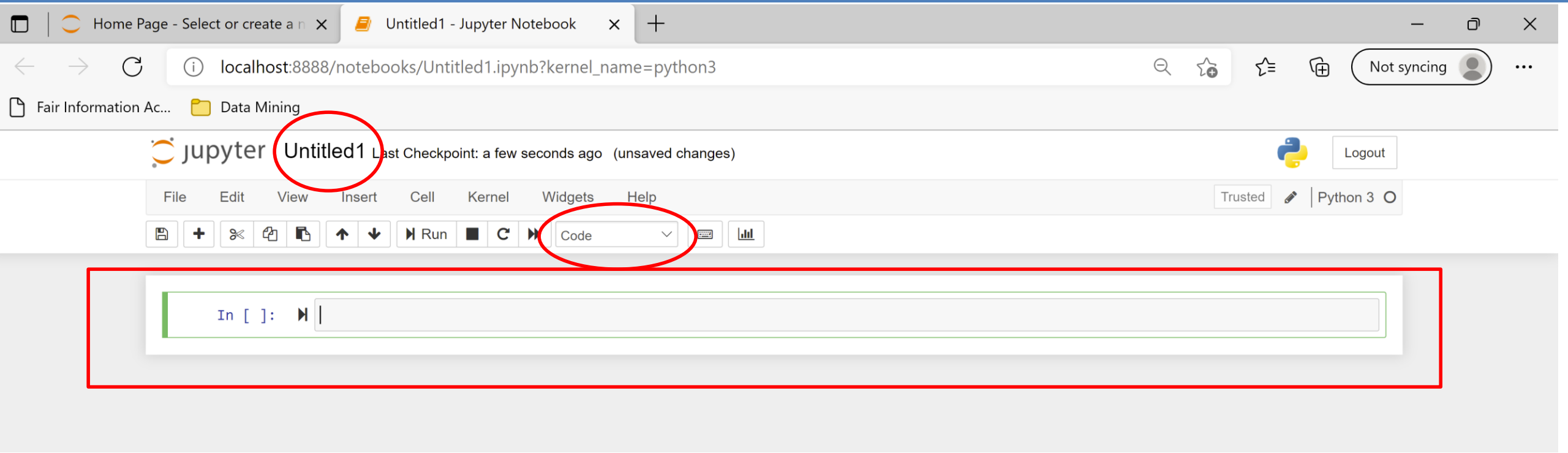
- From the Anaconda terminal type the command:
  - `jupyter notebook --generate-config`

- This will generate `.jupiter/jupyter_notebook_config.py` file under your home directory.

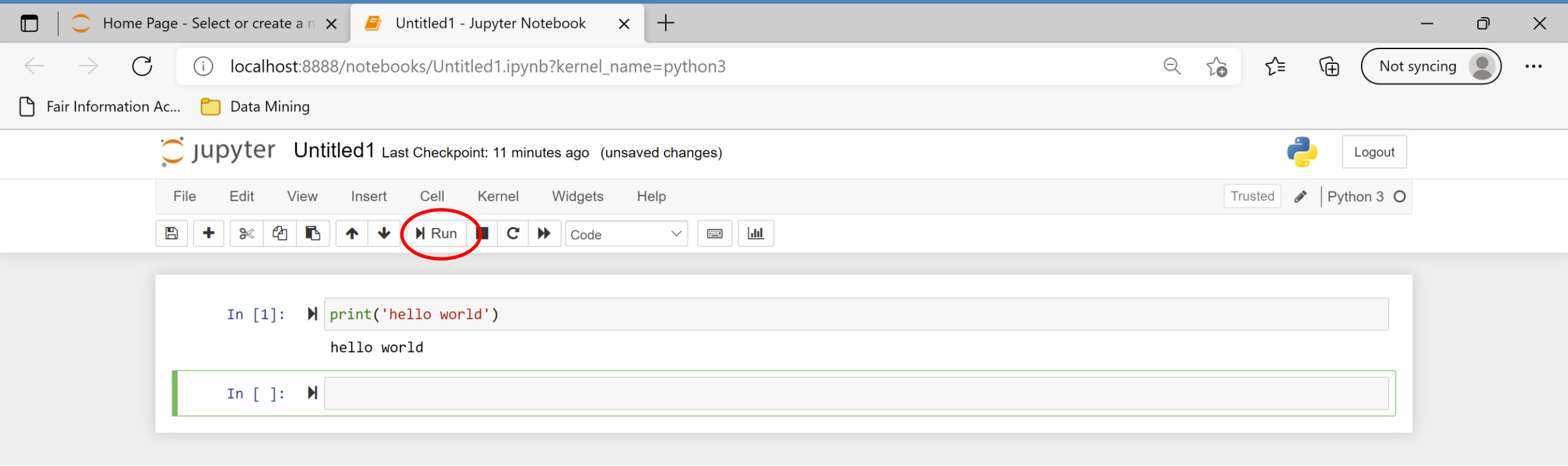
- Find, un-comment and modify the line

`# c.NotebookApp.notebook_dir = ''` in the config file to point to the desired directory

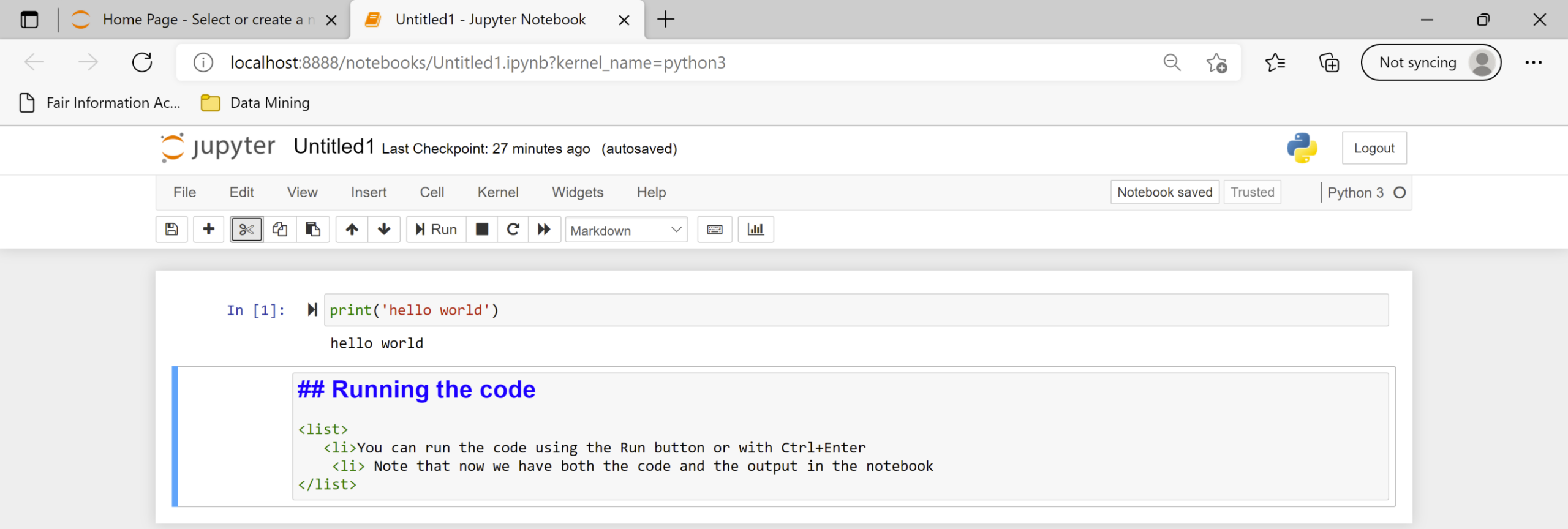

Home Page - Select or create a n
localhost:8888/tree
Not syncing
Fair Information Ac...
Data Mining
jupyter
Quit
Logout
Files
Running
Clusters
Nbextensions
Select items to perform actions on them.
Upload
New
Notebook:
Python 3
Other:
Text File
Folder
Terminal
0
/
Name
3D Objects
Contacts
Documents
Downloads
Dropbox (Personal)
10 months ago
Dropbox (UoI)
2 months ago
Favorites
8 months ago
gensim-data
a year ago
Google Drive
4 days ago
IdeaProjects
18 days ago
Links
8 months ago
Music
8 months ago
OneDrive
2 days ago
OneDrive - ΠΑΝΕΠΙΣΤΗΜΙΟ ΙΩΑΝΝΙΝΩΝ
2 days ago
Roaming
2 years ago




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

Home Page - Select or create a n | Untitled1 - Jupyter Notebook

localhost:8888/notebooks/Untitled1.ipynb?kernel_name=python3

Not syncing

Fair Information Ac... | Data Mining

jupyter Untitled1 Last Checkpoint: 31 minutes ago (unsaved changes)

Logout

File Edit View Insert Cell Kernel Widgets Help

Trusted | Python 3

Run | Code

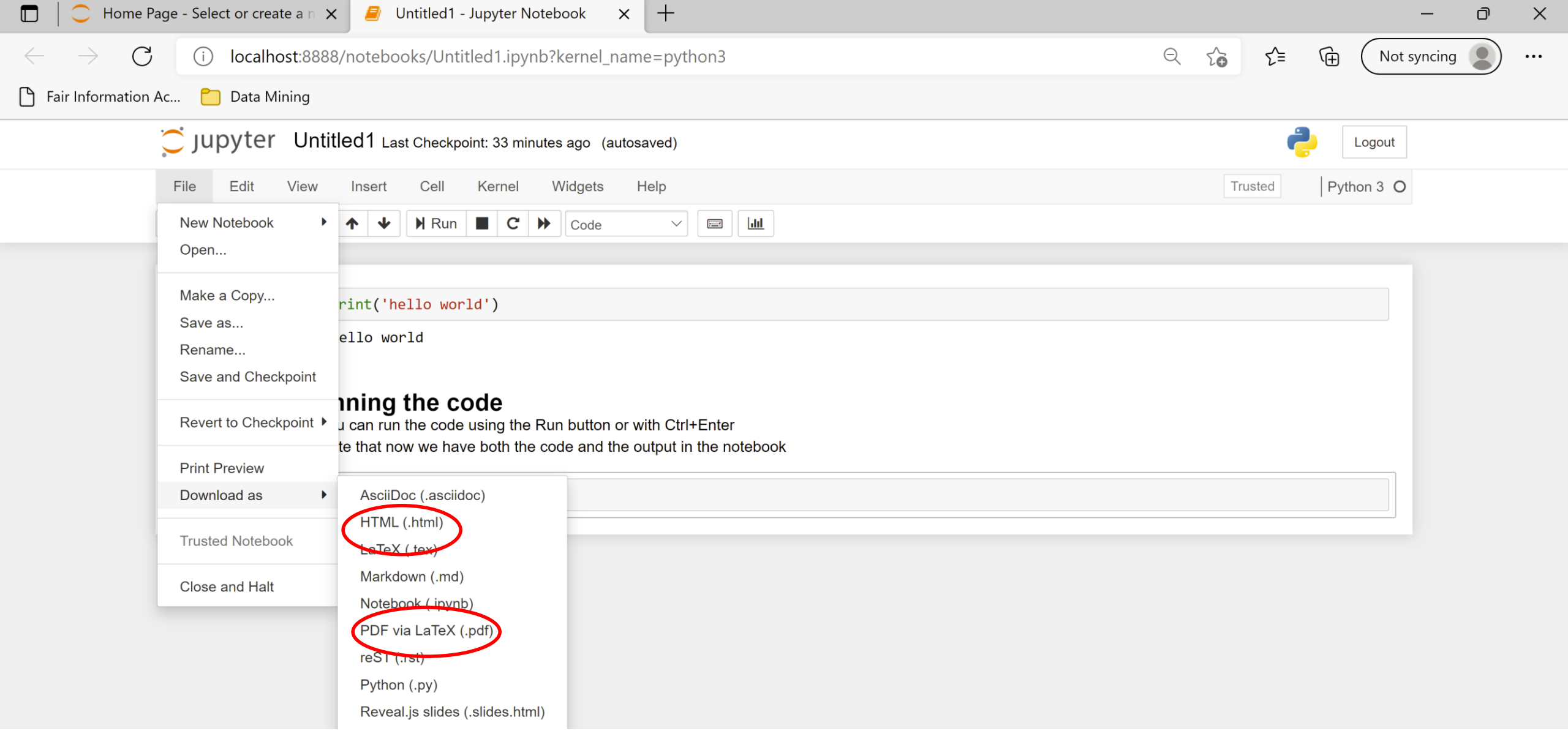
In [1]:
```
print('hello world')
```

```
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 [ ]:

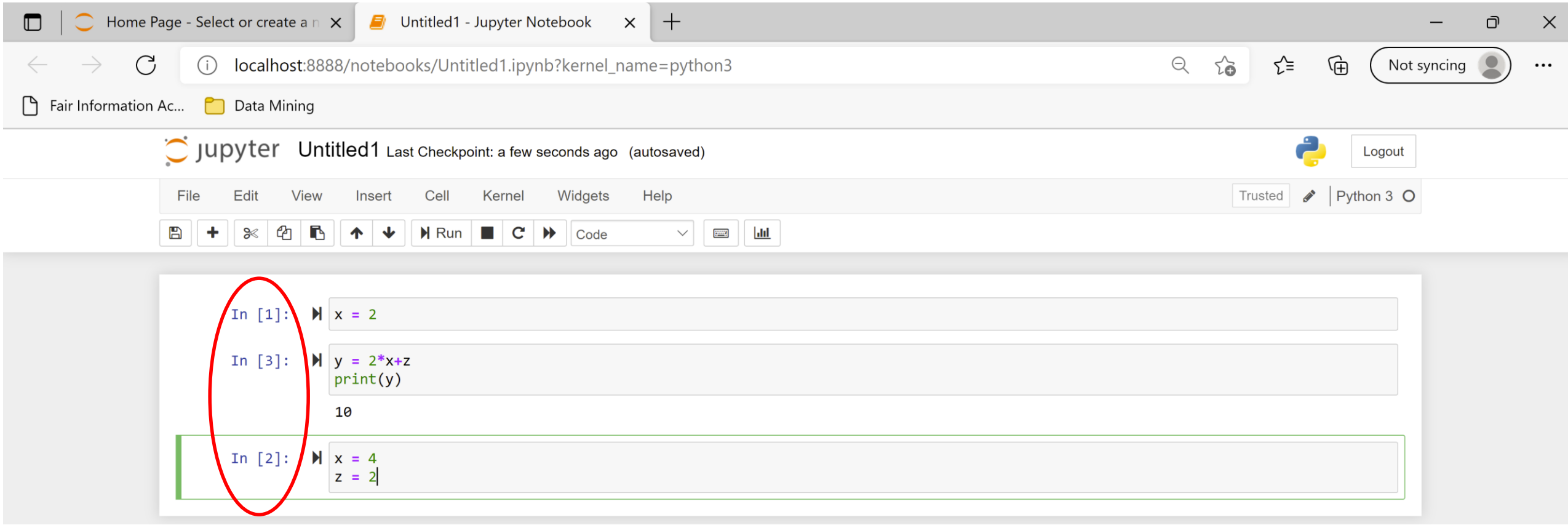



- You can export the notebook into HTML or PDF

# Attention!

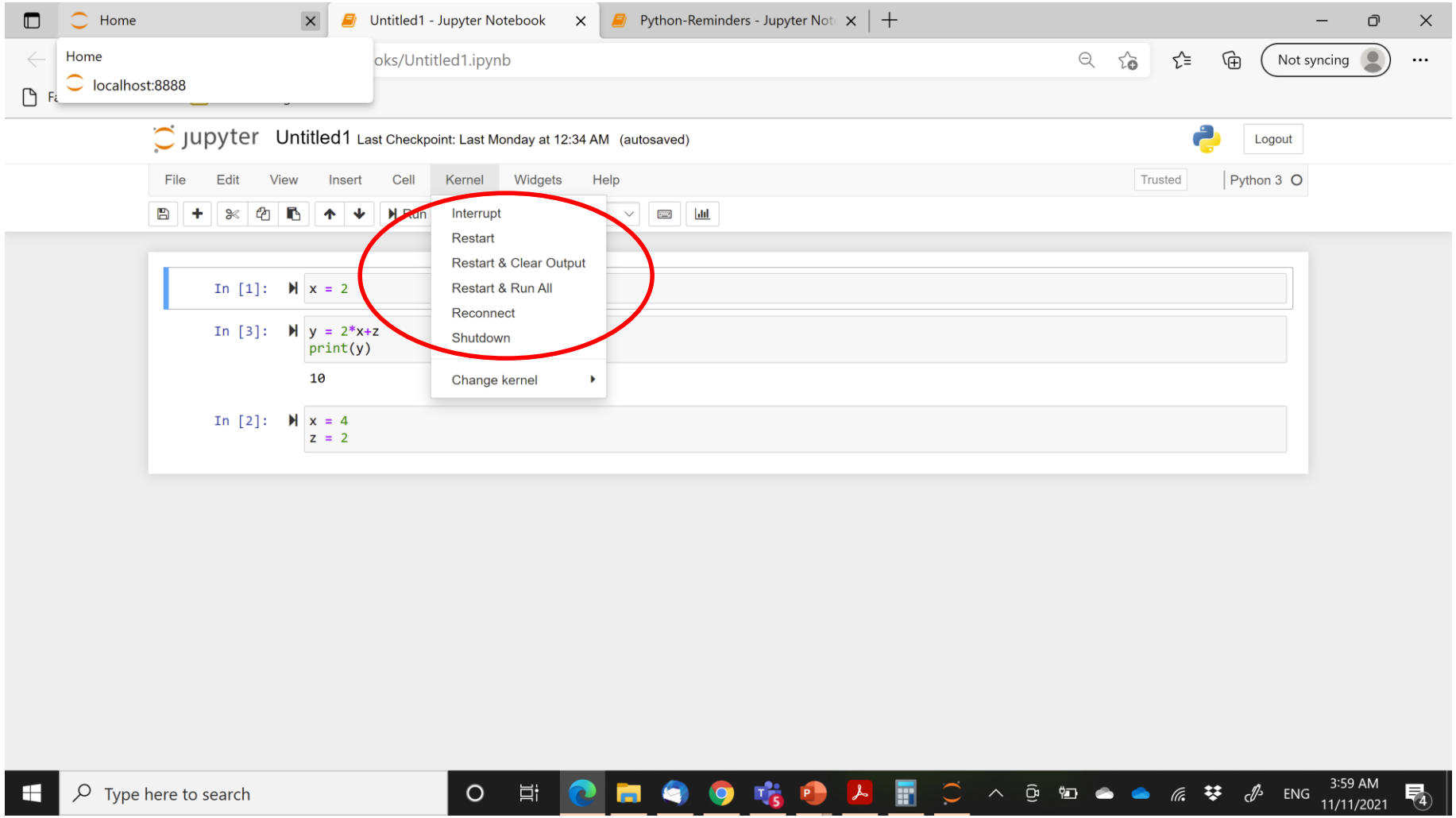
- A notebook is run interactively, each time running a specific cell
- The state of the program remains in memory while the notebook is running
- Each cell has access to the current state of the memory
- You can jump between cells in a non-linear way
- You should be aware of the state of the memory of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

# A simple example




- 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 as access to z, and uses the value 4 for x

# Restarting


Home
Untitled1 - Jupyter Notebook
Python-Reminders - Jupyter Not
Home
localhost:8888
oks/Untitled1.ipynb
Not syncing
jupyter Untitled1 Last Checkpoint: Last Monday at 12:34 AM (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3
Run
Interrupt
Restart
Restart & Clear Output
Restart & Run All
Reconnect
Shutdown
Change kernel
In [1]: x = 2
In [3]: y = 2*x+z
print(y)
10
In [2]: x = 4
z = 2
Type here to search
ENG
3:59 AM
11/11/2021