

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

Clipboard

2

3

4

5

6

7

Slide 5 of 7

AllAppsDocumentsEmailWebMore

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





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



Notebook

 6.0.0

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

Launch

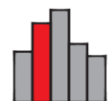


Spyder

3.3.6

Scientific PYTHON Development Environment. 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

Paste

Clipboard

5

6

7

8

9

10

Slide 7 of 10

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 c X

localhost:8889/tree

QuitLogout

jupyter

FilesRunningClusters

Select items to perform actions on them.

UploadNew↺

0

▼

📁 /

Name ▼

Last Modified

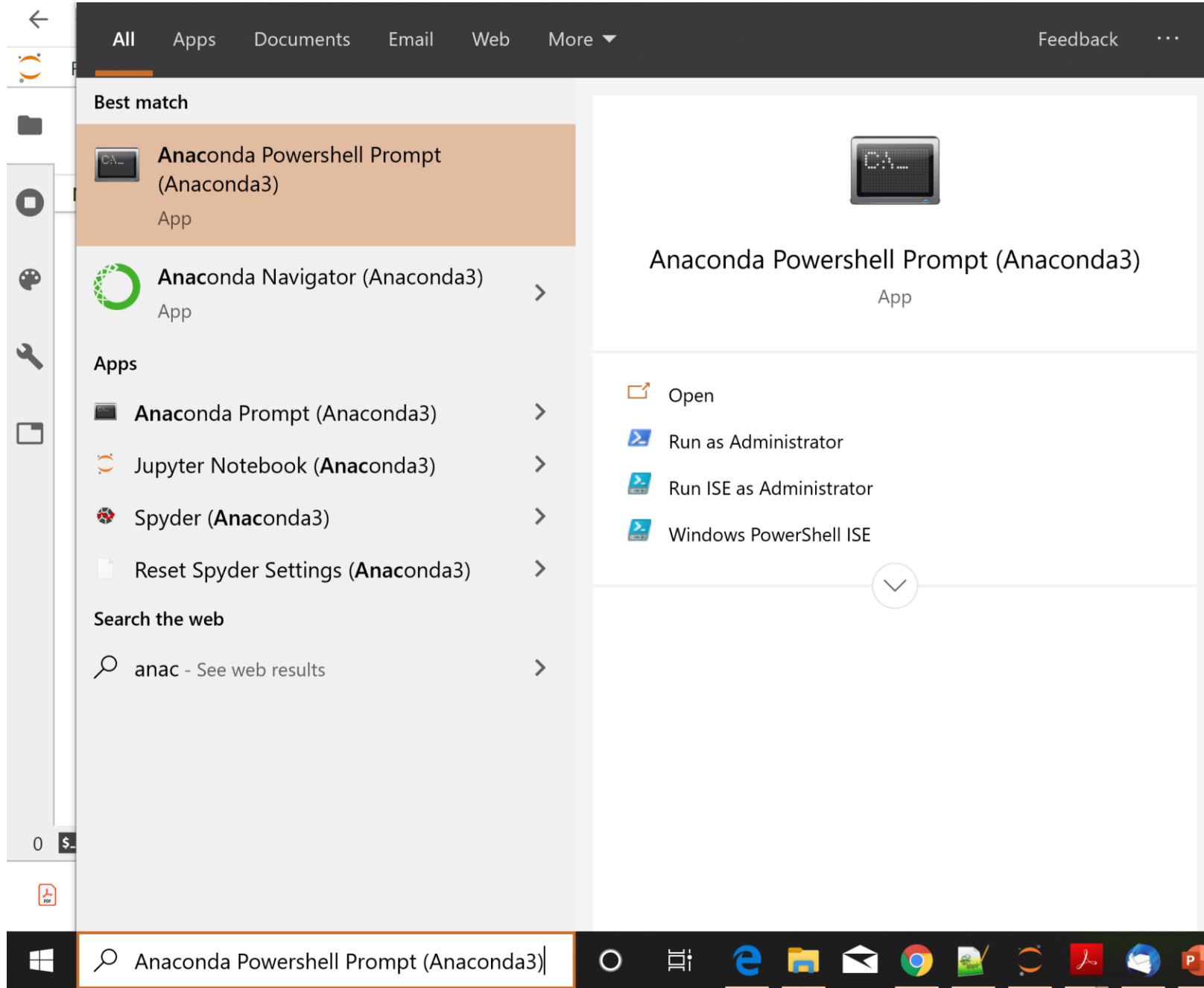
File size

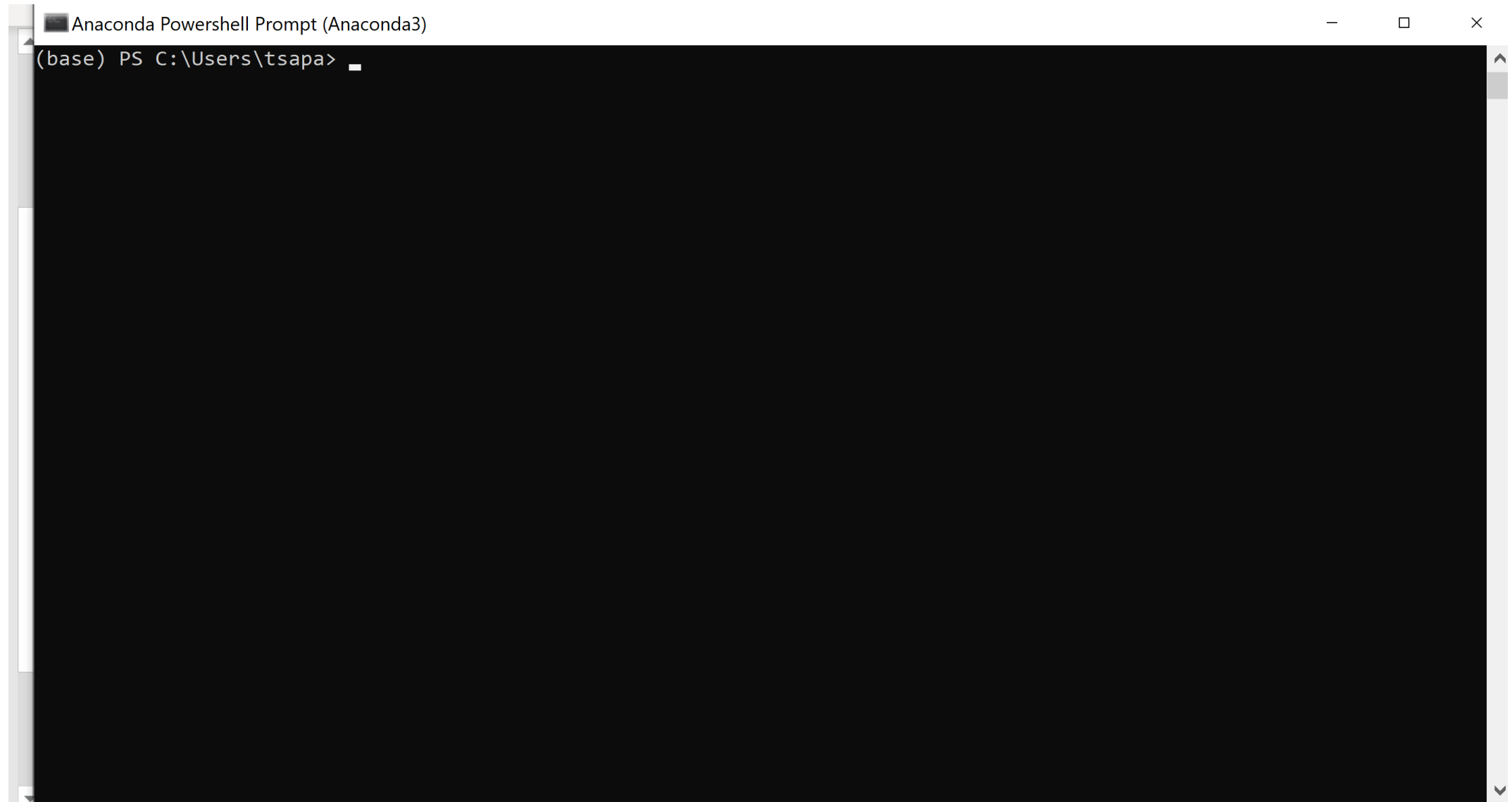
☐	📁 3D Objects	21 days ago
☐	📁 Contacts	21 days ago
☐	📁 Documents	3 months ago
☐	📁 Downloads	4 hours ago
☐	📁 Dropbox (Personal)	8 days ago
☐	📁 Dropbox (Uol)	2 months ago
☐	📁 Favorites	10 days ago
☐	📁 Google Drive	3 days ago
☐	📁 Links	21 days ago
☐	📁 Music	21 days ago
☐	📁 OneDrive	2 days ago
☐	📁 Roaming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

Type here to search

🔍📁📧🌐📄🔄📄📧📄

11:09 PM10/31/2019





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



Quit

Logout

Files

Running

Clusters

Nbextensions

Select items to perform actions on them.

☐	0		/		
☐			3D Objects		
☐			Contacts		
☐			Documents		
☐			Downloads		
☐			Dropbox (Personal)		10 months ago
☐			Dropbox (Uol)		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

Upload

New

Notebook:

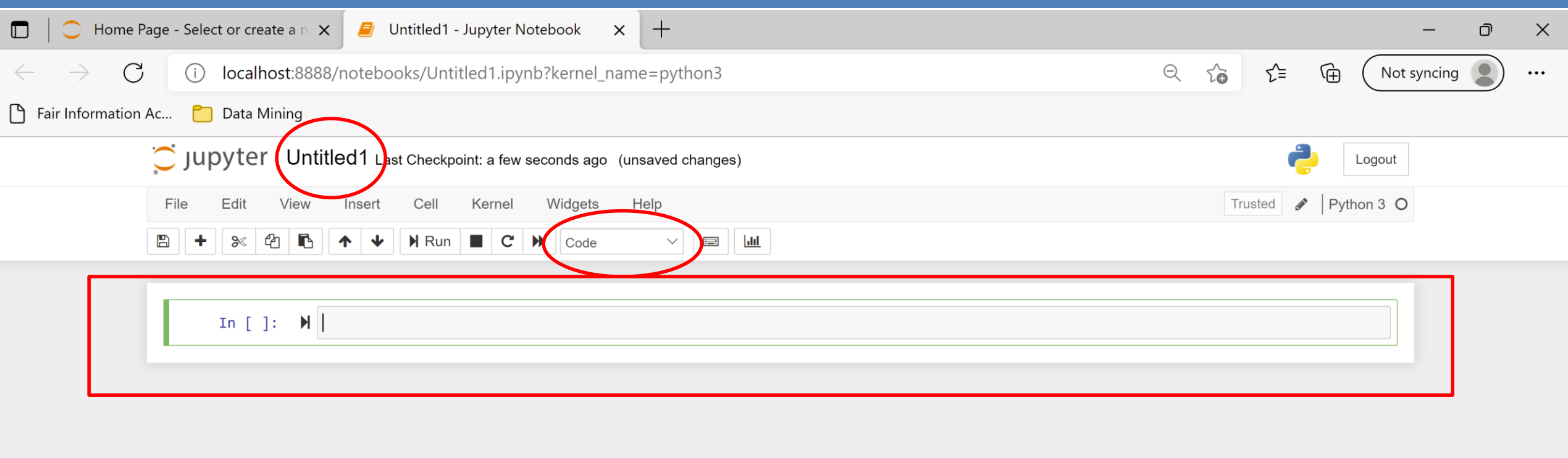
Python 3

Other:

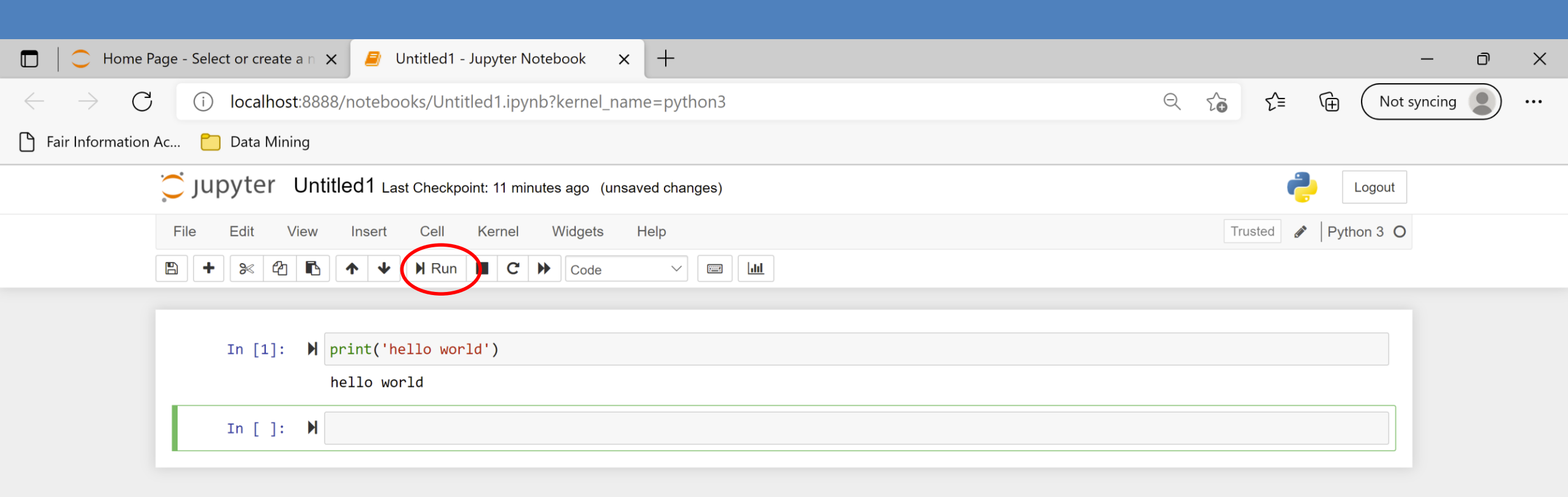
Text File

Folder

Terminal



- 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

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

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 []:

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: 33 minutes ago (autosaved)

Logout

File Edit View Insert Cell Kernel Widgets Help

Trusted Python 3

New Notebook
Open...

Make a Copy...
Save as...
Rename...
Save and Checkpoint

Revert to Checkpoint

Print Preview

Download as

Trusted Notebook

Close and Halt

Run

Code

```
print('hello world')
```

hello world

Running the code

You can run the code using the Run button or with Ctrl+Enter

ate that now we have both the code and the output in the notebook

AsciiDoc (.asciidoc)
HTML (.html)
~~LaTeX (.tex)~~
Markdown (.md)
Notebook (.ipynb)
PDF via LaTeX (.pdf)
reST (.rst)
Python (.py)
Reveal.js slides (.slides.html)

- 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 screenshot shows a Jupyter Notebook interface with three code cells. The first cell (In [1]:) contains the code `x = 2`. The second cell (In [3]:) contains the code `y = 2*x+z` and `print(y)`, with the output `10` displayed below it. The third cell (In [2]:) contains the code `x = 4` and `z = 2`. A red oval highlights the input prompts `In [1]:`, `In [3]:`, and `In [2]:` to show the execution order. The notebook title is `Untitled1` and the kernel is `python3`.

- 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

