

Analytics Jumpstart

Introduction



Mike Schriefer, Instructor/ Data Engineer at Stratasan

Dalila Paquiot, Data Scientist at XSOLIS

Julia Romano, Integration Engineer at PlayMaker Health

Mary van Valkenburg, Data Science + Analytics Program Manager



- **Your name**
- **The place you call home**
- **Something new in the last (pandemic) year**



Classroom *guidelines*

- Ask lots of questions
- Help each other; learn from each other
- Get comfortable with discomfort. Making mistakes, figuring them out, and then correcting them is part of the learning process
- After working through the assignment, form your own ideas and do your own exploration beyond what has been suggested



Class format

1. **Concepts/Code Lecture**
2. **Coding practice**
3. **Interactive with instruction team and other students!**

The weekly targets are for help in pacing lectures. Feel free to work ahead, but be aware you may need to use Google or the pandas API!!



Goals for the class

- **Get hands-on experience of what it might be like to work as a data analyst or data scientist**
- **Get an idea of whether or not this might be a good fit for a career**
- **Learn some tools to help you on personal analysis projects**
- **Make discoveries and have fun**

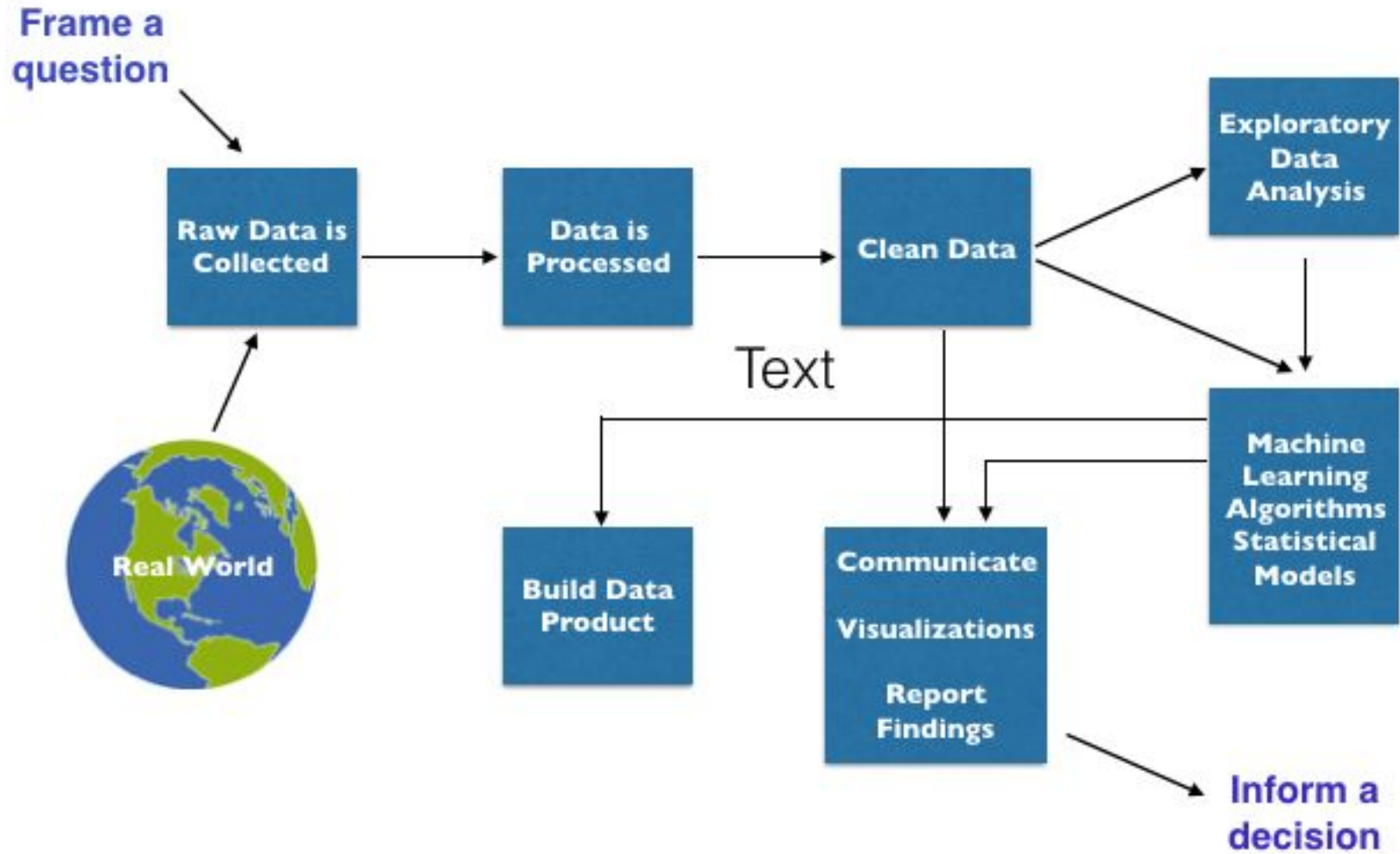


Goals for today

- Get an idea of what to expect
- Define the Data Science Process – *a basic mental model for analysis*
- Folder Setup
- Jupyter Notebook walkthrough/orientation
- Analysis Guide walkthrough
- Get started on the project!






The Data Science Process



Organize your files in the structure shown

We will share a Zip file you can unzip and move to where you want (Documents, Desktop, etc)

Do not put your files on the cloud (One Drive, iCloud)

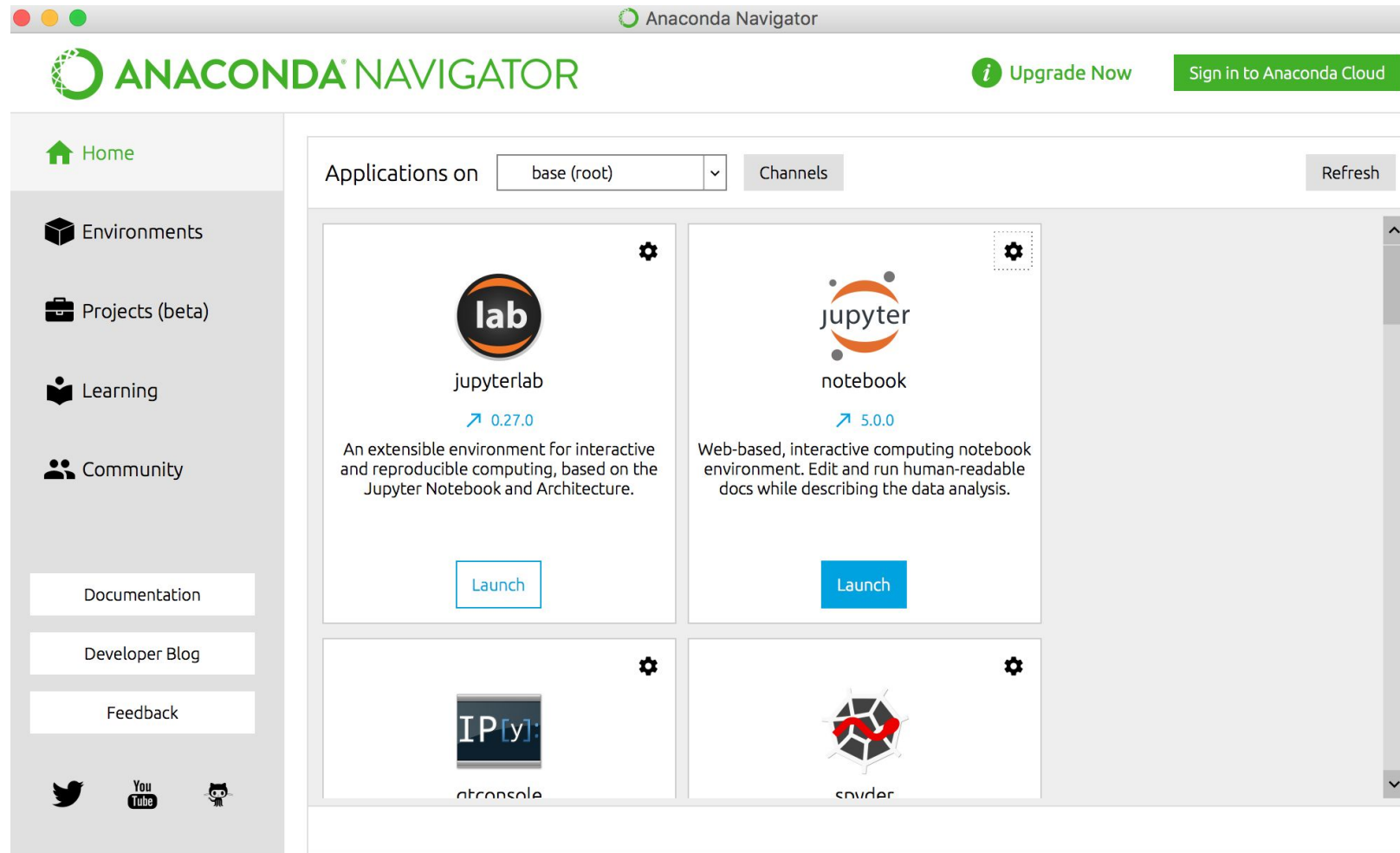
Name	Date Modified	Date Added	Size	Kind	^
▶ data	Today at 9:48 AM	Today, 9:48 AM	--	Folder	
▶ notebooks	Today at 7:48 AM	Today, 9:48 AM	--	Folder	
▶ slides	Today at 7:48 AM	Today, 9:48 AM	--	Folder	
 Analysis_Guide.docx	Jun 18, 2020 at 8:43 PM	Today, 9:48 AM	15 KB	Micros... (.docx)	
 Glossary.docx	Jun 5, 2020 at 7:43 AM	Today, 9:48 AM	10 KB	Micros... (.docx)	
 Syllabus.docx	Today at 7:45 AM	Today, 9:48 AM	7 KB	Micros... (.docx)	



Orientation to Jupyter Notebook



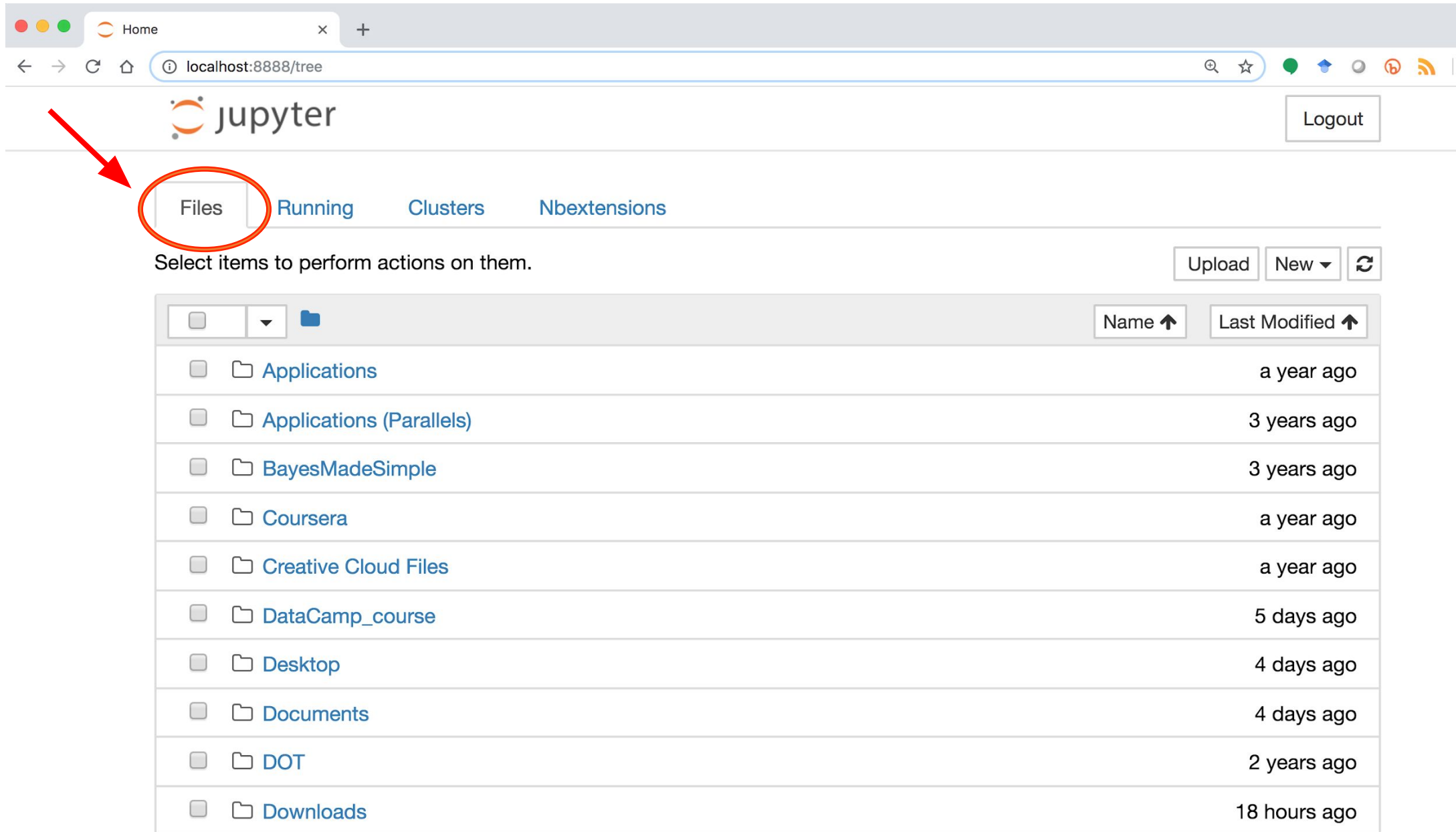
Open Anaconda Navigator, install, and launch Jupyter Notebook



A new tab will open in your default browser. It's not actually connecting to the internet, just running on your machine



You will see the file structure on your computer and can navigate as normal

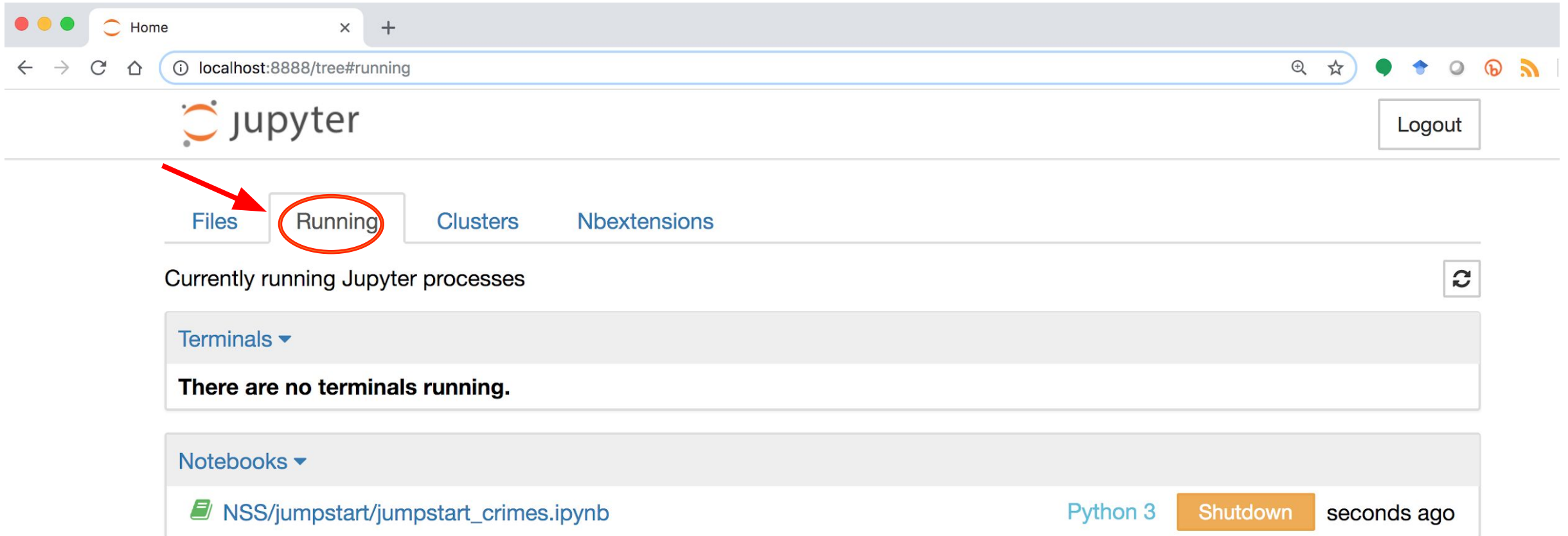


The screenshot shows the JupyterLab interface in a web browser. The address bar indicates the URL is `localhost:8888/tree`. The Jupyter logo is visible in the top left, and a 'Logout' button is in the top right. Below the logo, there are four tabs: 'Files', 'Running', 'Clusters', and 'Nbextensions'. The 'Files' tab is selected and circled in orange, with a red arrow pointing to it. Below the tabs, there is a text prompt 'Select items to perform actions on them.' and buttons for 'Upload', 'New', and a refresh icon. The main area displays a file browser view with a table of files and folders. The table has columns for 'Name' and 'Last Modified'. The files listed are:

Name	Last Modified
Applications	a year ago
Applications (Parallels)	3 years ago
BayesMadeSimple	3 years ago
Coursera	a year ago
Creative Cloud Files	a year ago
DataCamp_course	5 days ago
Desktop	4 days ago
Documents	4 days ago
DOT	2 years ago
Downloads	18 hours ago



See what notebooks are already running (should be empty if just opening Jupyter)



The screenshot shows a web browser window with the address bar at `localhost:8888/tree#running`. The Jupyter logo and a 'Logout' button are at the top right. Below the navigation bar, the 'Running' tab is selected and circled in red, with a red arrow pointing to it from the left. The 'Files' tab is also visible. The main content area is titled 'Currently running Jupyter processes' and contains two sections: 'Terminals' and 'Notebooks'. The 'Terminals' section shows 'There are no terminals running.' The 'Notebooks' section shows a single notebook entry: 'NSS/jumpstart/jumpstart_crimes.ipynb' with a 'Python 3' label, a 'Shutdown' button, and the text 'seconds ago'.

Home x +

localhost:8888/tree#running

jupyter Logout

Files **Running** Clusters Nbextensions

Currently running Jupyter processes

Terminals ▼

There are no terminals running.

Notebooks ▼

NSS/jumpstart/jumpstart_crimes.ipynb Python 3 Shutdown seconds ago



Navigate to analytics_jumpstart/notebooks and create a new Python 3 notebook

twareS x +

ree/Documents/NashvilleSoftwareSchool/analytics_jumpstart/notebooks

Home Trickster Euchre New folder Apple ESPN Google Maps YouTube remote

jupyter

Quit Logout

Files Running Clusters

Select items to perform actions on them.

0

/ Documents / NashvilleSoftwareSchool / analytics_jumpstart / notebooks

Name

..

The notebook list is empty.

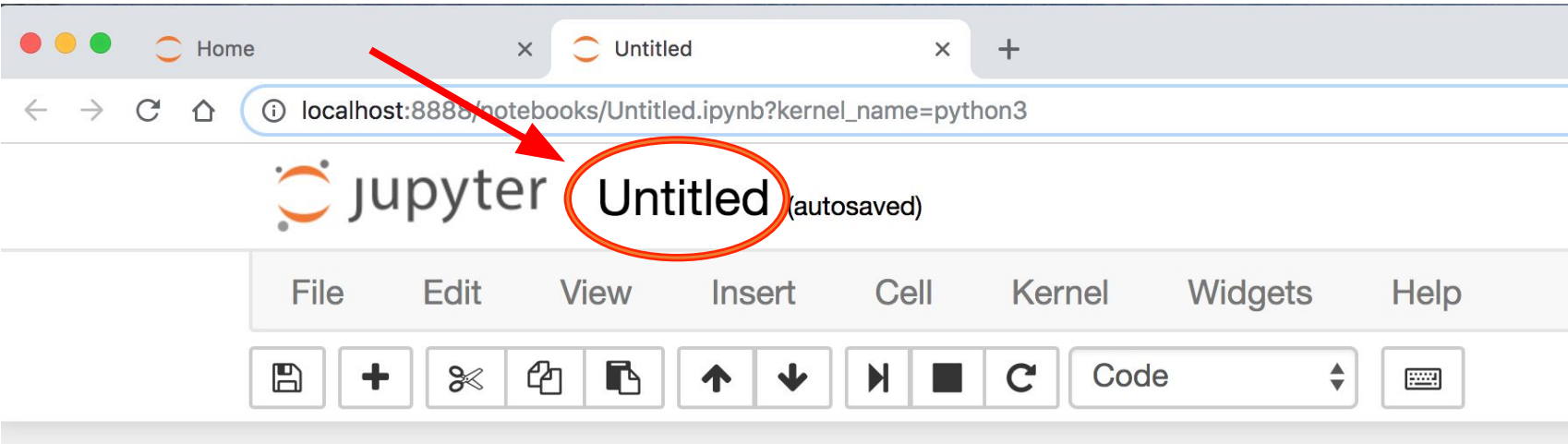
Upload New

Notebook:
Python 3

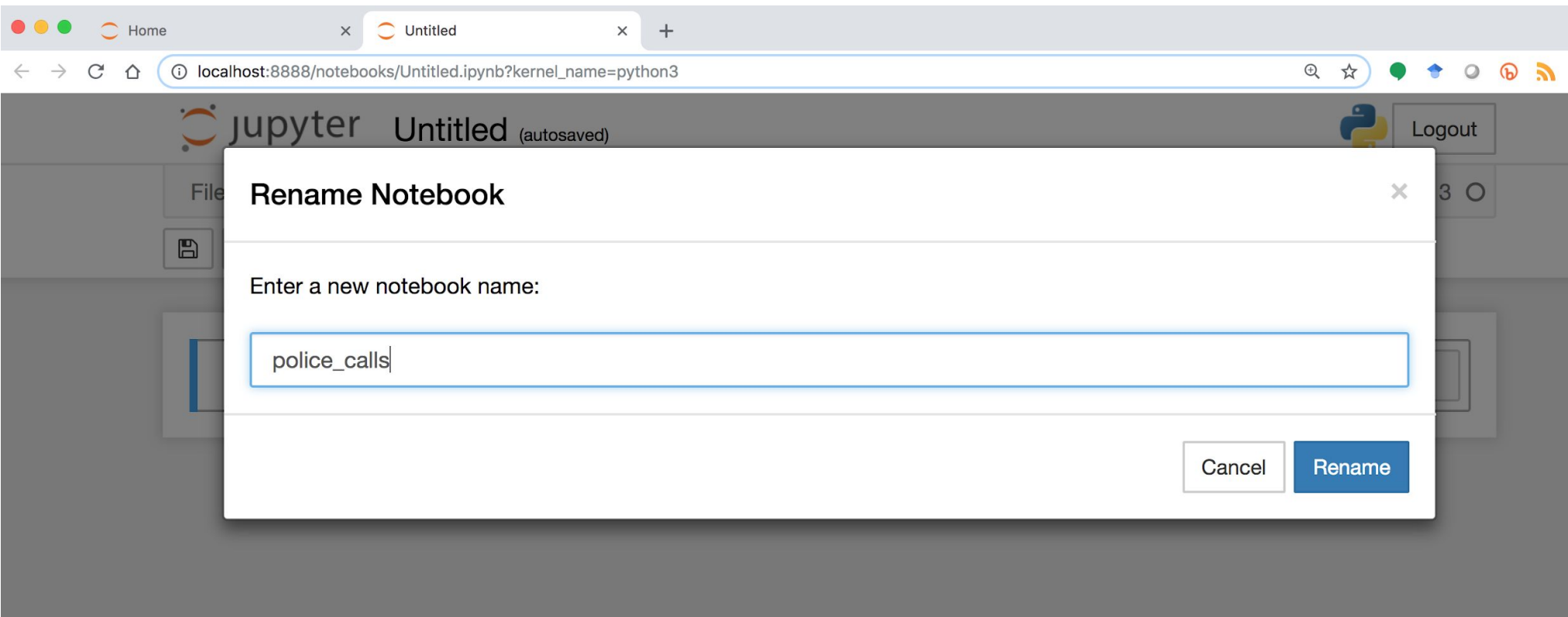
Other:
Text File
Folder
Terminal



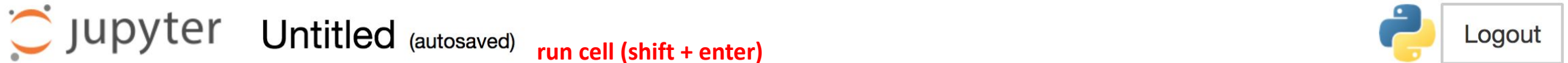
Title the notebook



- meaningful
- no spaces



Useful buttons (and shortcuts) for running code and moving cells around



run cell (shift + enter)



save add cell cut cell

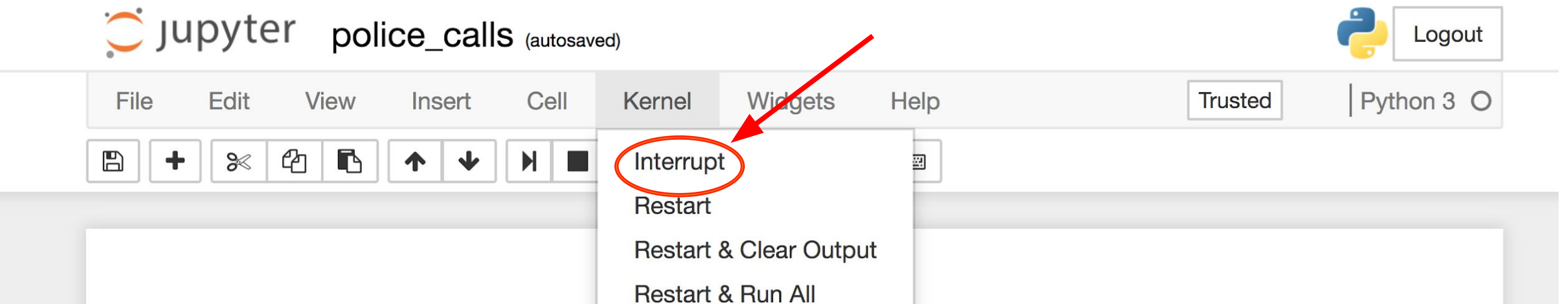
move cell (up/down)

toggle between markdown and code

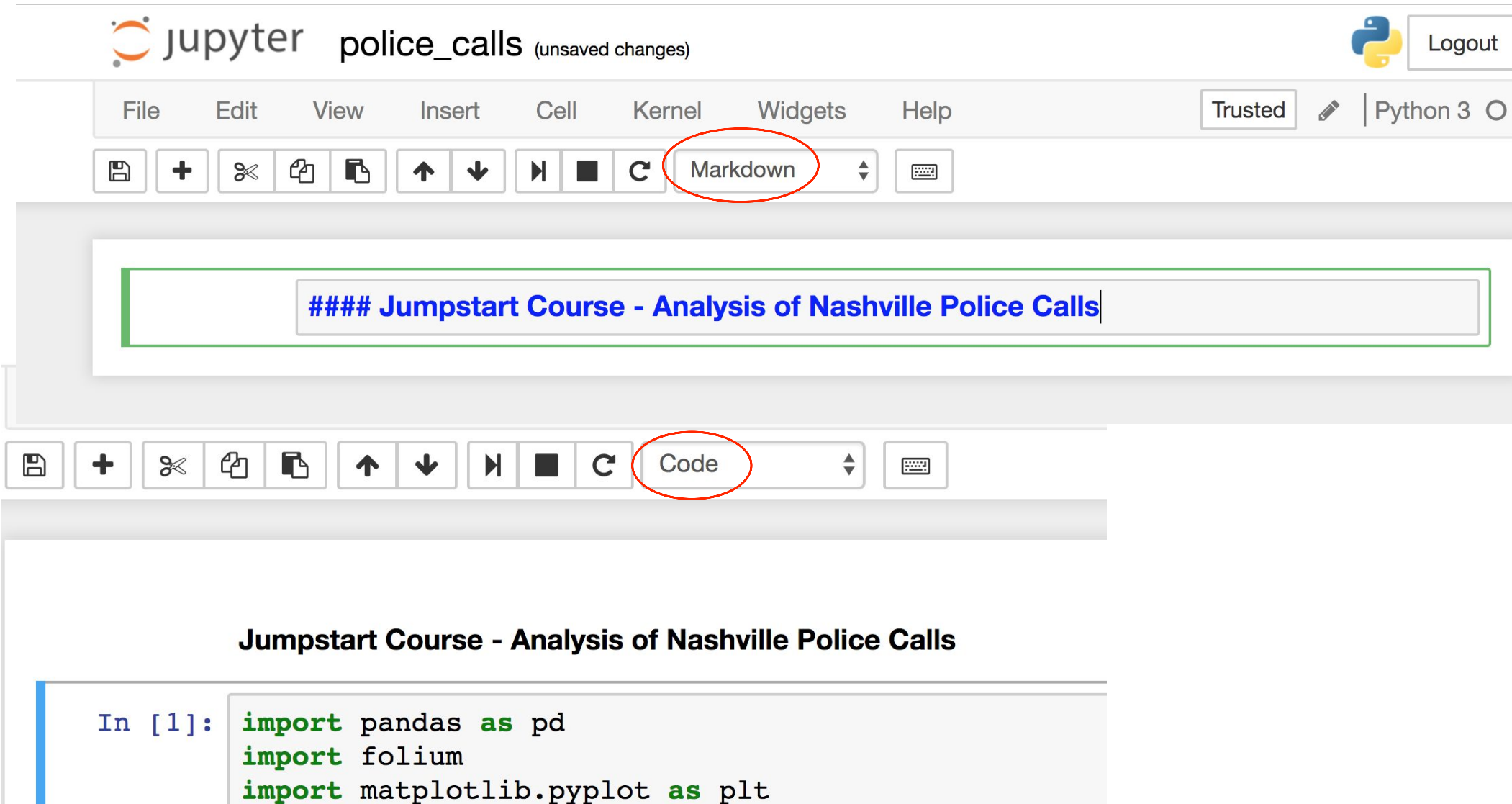
In []:

If your code is taking too long or is giving unexpected results, try restarting the kernel

Each time you run a piece of code in a Jupyter Notebook, that process is saved to a kernel. All the inputs, outputs, variables, etc. are saved. Even if you modify or delete a cell, the first time it was run was saved. This can sometimes lead to strange results. Restarting the kernel will clear out the memory so you can start fresh. Closing and opening the notebook will also do this. But remember to rerun all your code after you restart the kernel!



Change the format of the cell to add notes or run code



The screenshot displays the JupyterLab interface for a file named `police_calls` with unsaved changes. The top menu bar includes File, Edit, View, Insert, Cell, Kernel, Widgets, and Help. The 'Cell' menu is open, showing options like New, Copy, Paste, Cut, Undo, Redo, and a dropdown menu where 'Markdown' is currently selected and circled in red. Below the menu, a text input field contains the text `#### Jumpstart Course - Analysis of Nashville Police Calls`. Below this, another dropdown menu is shown, where 'Code' is selected and circled in red. The bottom section of the image shows the resulting code cell, titled 'Jumpstart Course - Analysis of Nashville Police Calls', containing the following Python code:

```
In [1]: import pandas as pd
import folium
import matplotlib.pyplot as plt
```

Next up:

- **launch Jupyter**
- **walk through analysis guide**

