

Analytics Jumpstart

Introduction



Justin Rothbart, Instructor/Data Scientist at Direct Auto Insurance

Selam Tekie, Instructor/Data Intelligence Analyst at Service Source

Mahesh Rao, Instructor at NSS

Mary van Valkenburg, Instructor/Program Manager at NSS



- **Your name**
- **The place you call home**
- **Something people are usually surprised to discover about you**



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!**



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**

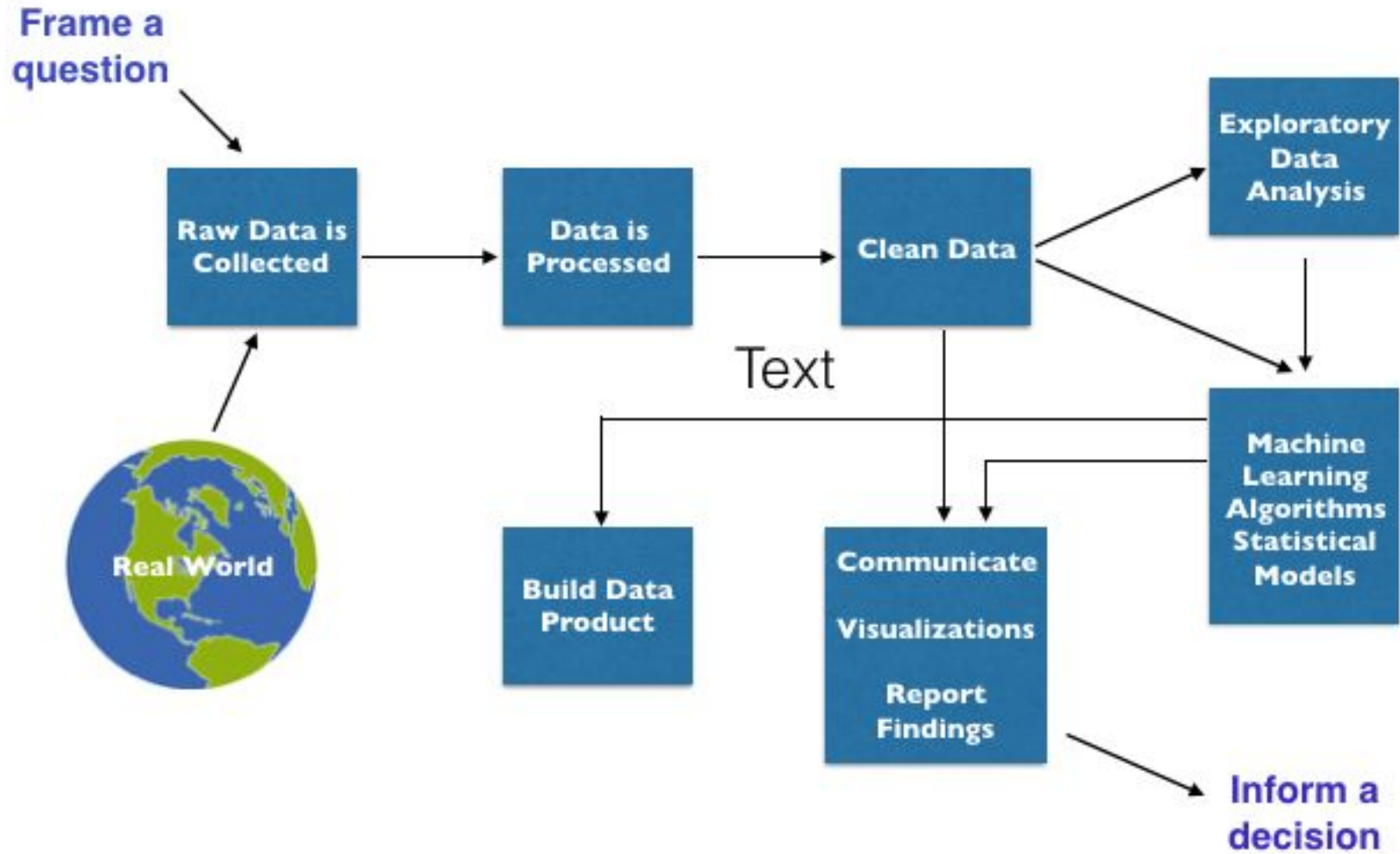


Goals for today

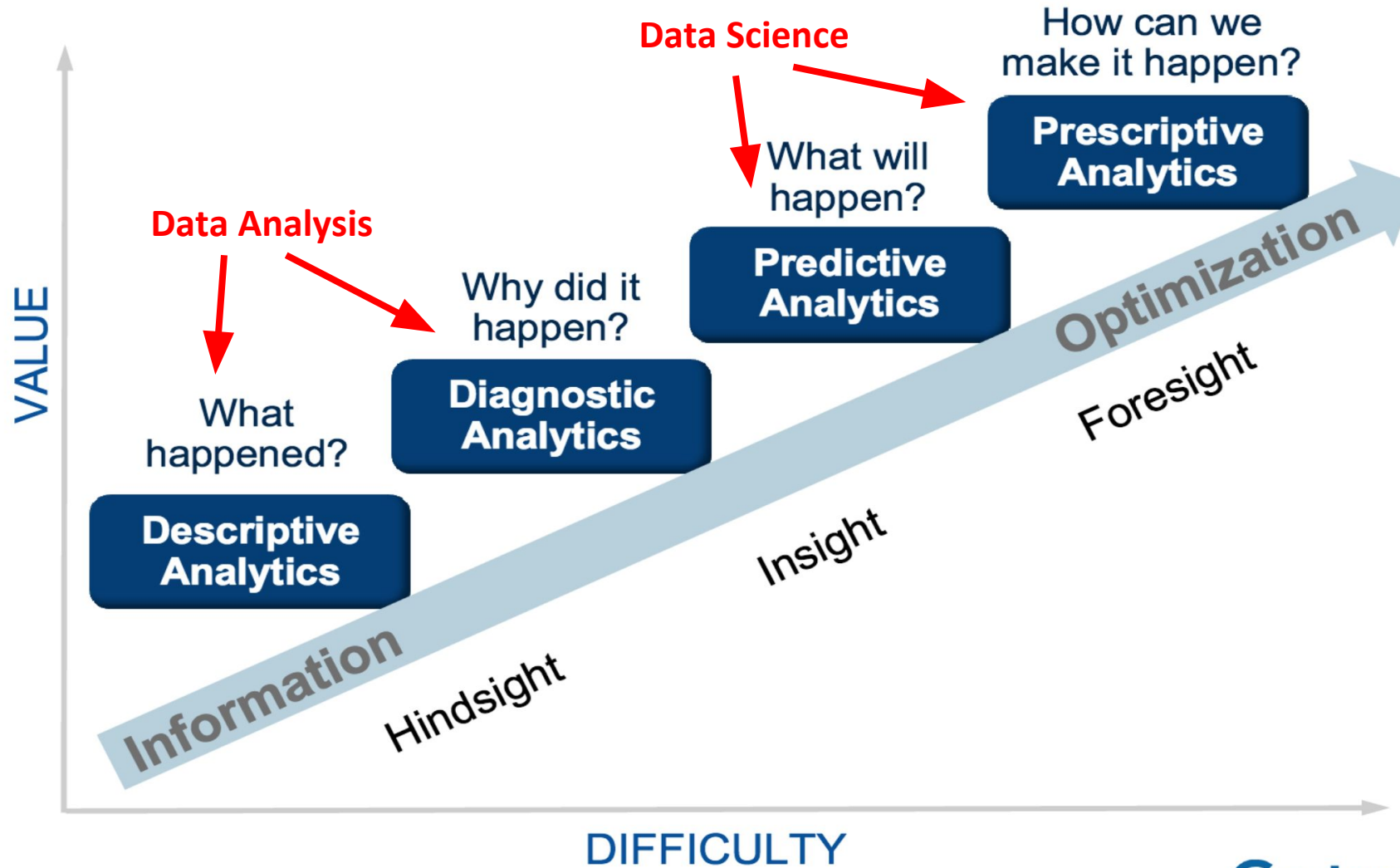
- Define the Data Science Process – *a basic mental model for analysis*
- Define *different types* of analytics and differentiate between Data Analytics and Data Science
- Jupyter Notebook walkthrough/orientation
- Analysis Guide walkthrough
- Get started on the project!



The Data Science Process



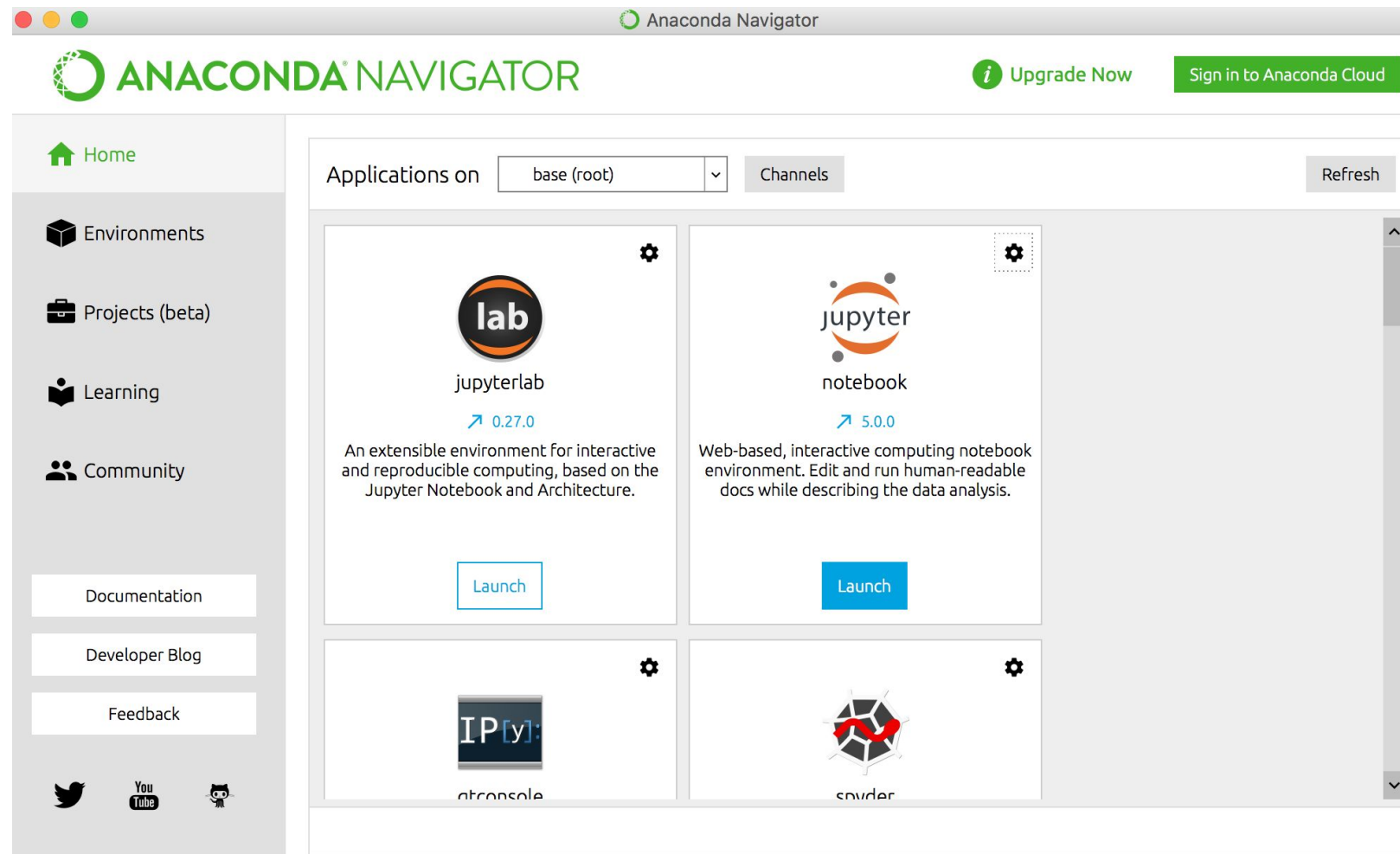
Gartner Analytic Ascendancy Model



Orientation to Jupyter Notebook



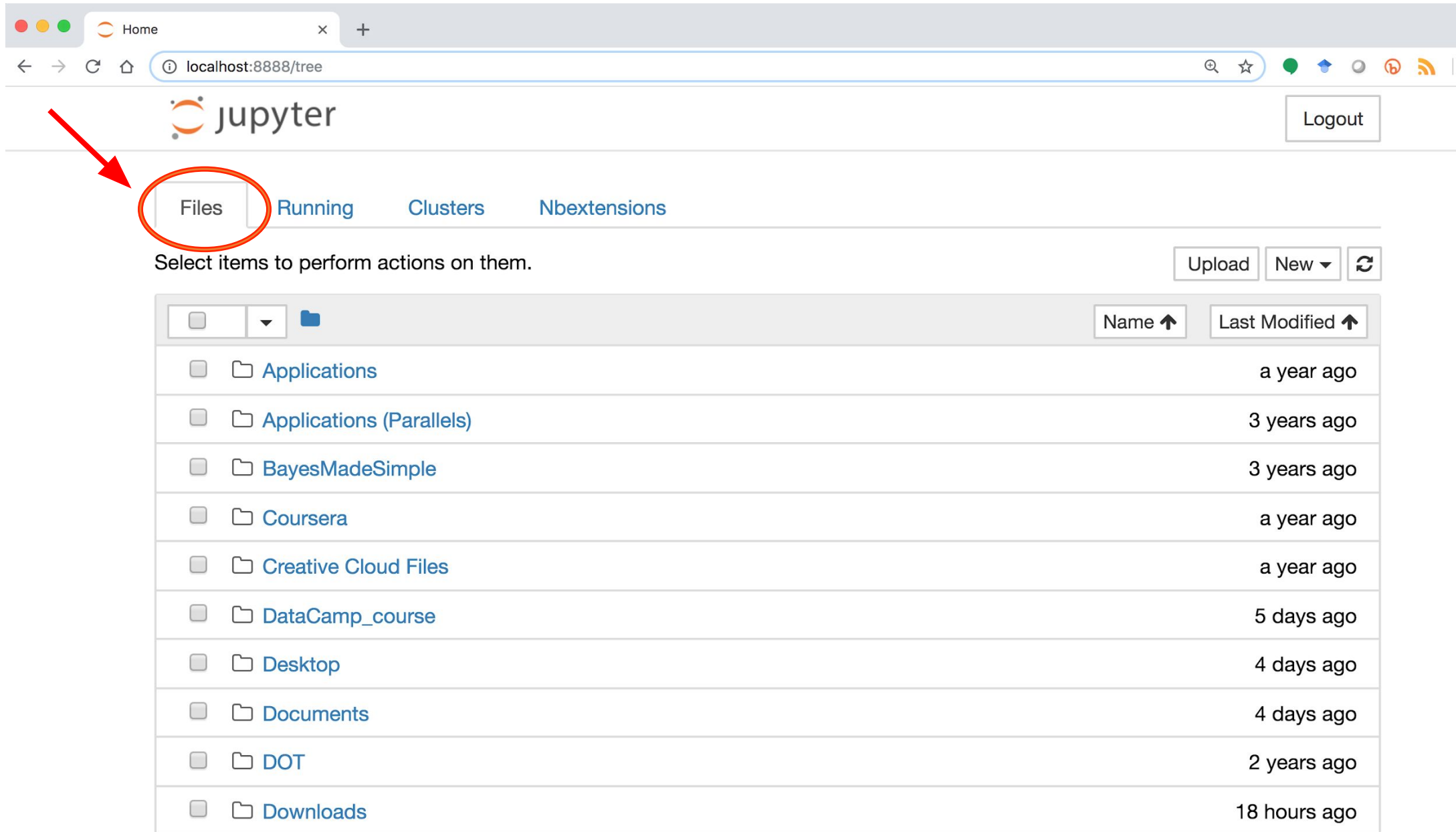
Open Anaconda Navigator, install, and launch Jupyter Notebook



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

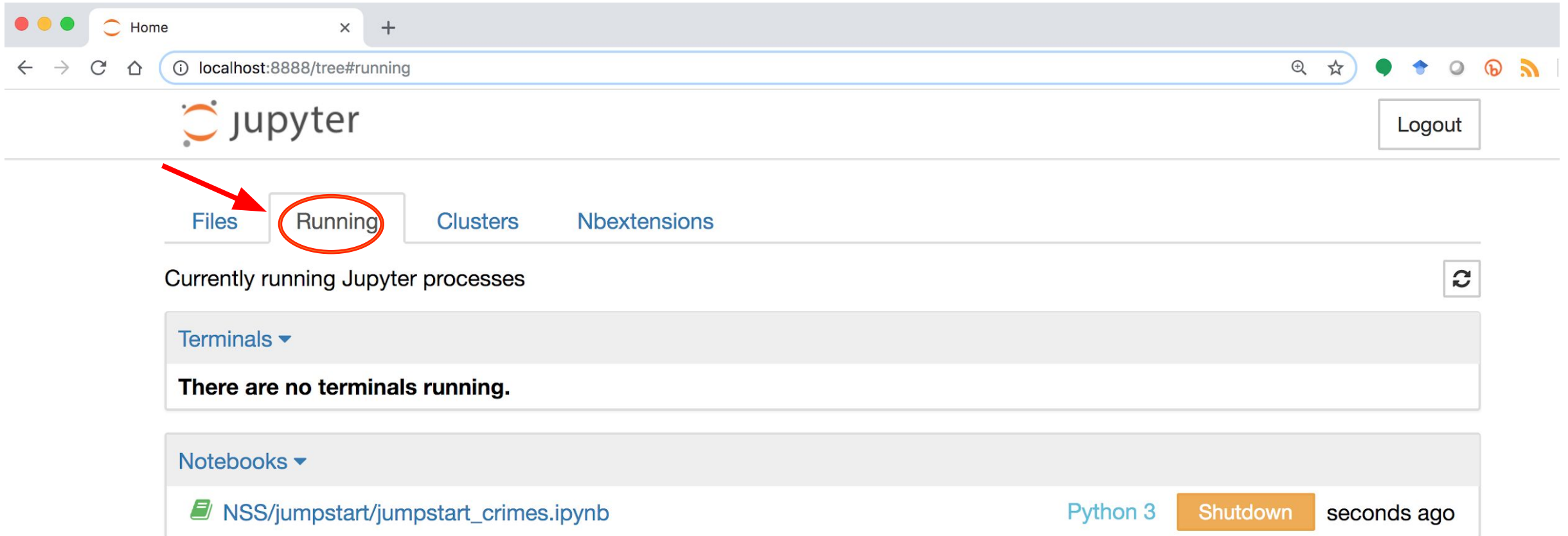


The screenshot shows the JupyterLab interface in a web browser. The address bar indicates the URL is `localhost:8888/tree`. The Jupyter logo is visible in the top left, and a "Logout" button is in the top right. Below the logo, there are four tabs: "Files", "Running", "Clusters", and "Nbextensions". The "Files" tab is selected and circled in orange, with a red arrow pointing to it. Below the tabs, there is a message: "Select items to perform actions on them." To the right of this message are buttons for "Upload", "New", and a refresh icon. The main area displays a list of files and folders. Each item has a checkbox, a folder icon, the name, and the last modified time. The files are sorted by name in ascending order.

	Name ↑	Last Modified ↑
<input type="checkbox"/>	Applications	a year ago
<input type="checkbox"/>	Applications (Parallels)	3 years ago
<input type="checkbox"/>	BayesMadeSimple	3 years ago
<input type="checkbox"/>	Coursera	a year ago
<input type="checkbox"/>	Creative Cloud Files	a year ago
<input type="checkbox"/>	DataCamp_course	5 days ago
<input type="checkbox"/>	Desktop	4 days ago
<input type="checkbox"/>	Documents	4 days ago
<input type="checkbox"/>	DOT	2 years ago
<input type="checkbox"/>	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 Jupyter interface. The address bar shows 'localhost:8888/tree#running'. The Jupyter logo is 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 'Running' tab is selected and circled in red, with a red arrow pointing to it. Below the tabs, the text 'Currently running Jupyter processes' is displayed. There are two expandable sections: 'Terminals' and 'Notebooks'. The 'Terminals' section is expanded and shows 'There are no terminals running.' The 'Notebooks' section is also expanded and shows a single notebook entry: 'NSS/jumpstart/jumpstart_crimes.ipynb'. To the right of the notebook name, it says 'Python 3' and 'Shutdown 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:

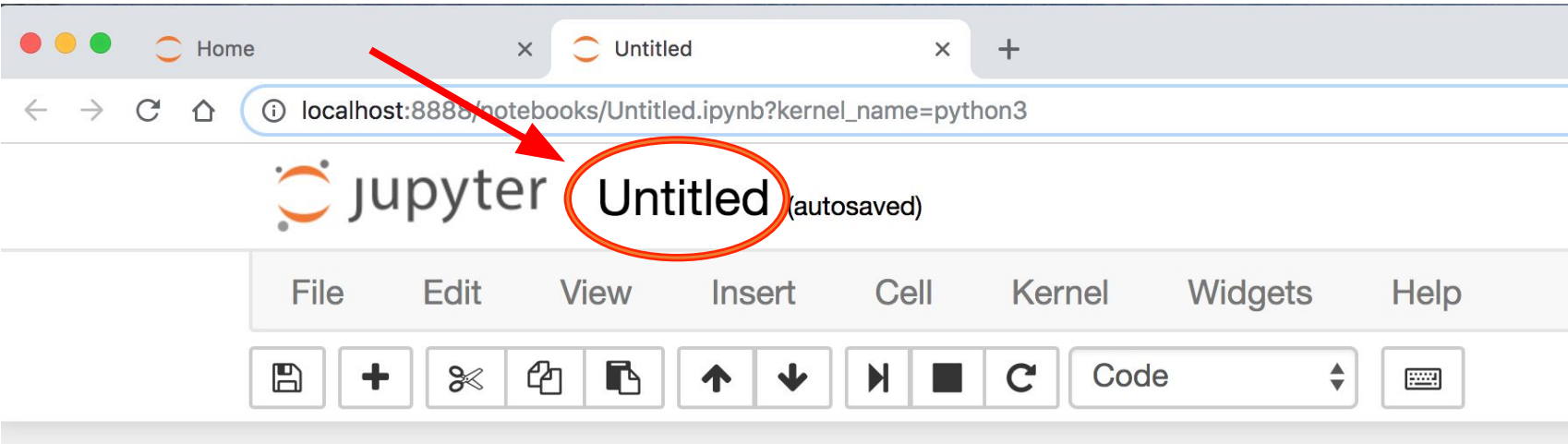
Text File

Folder

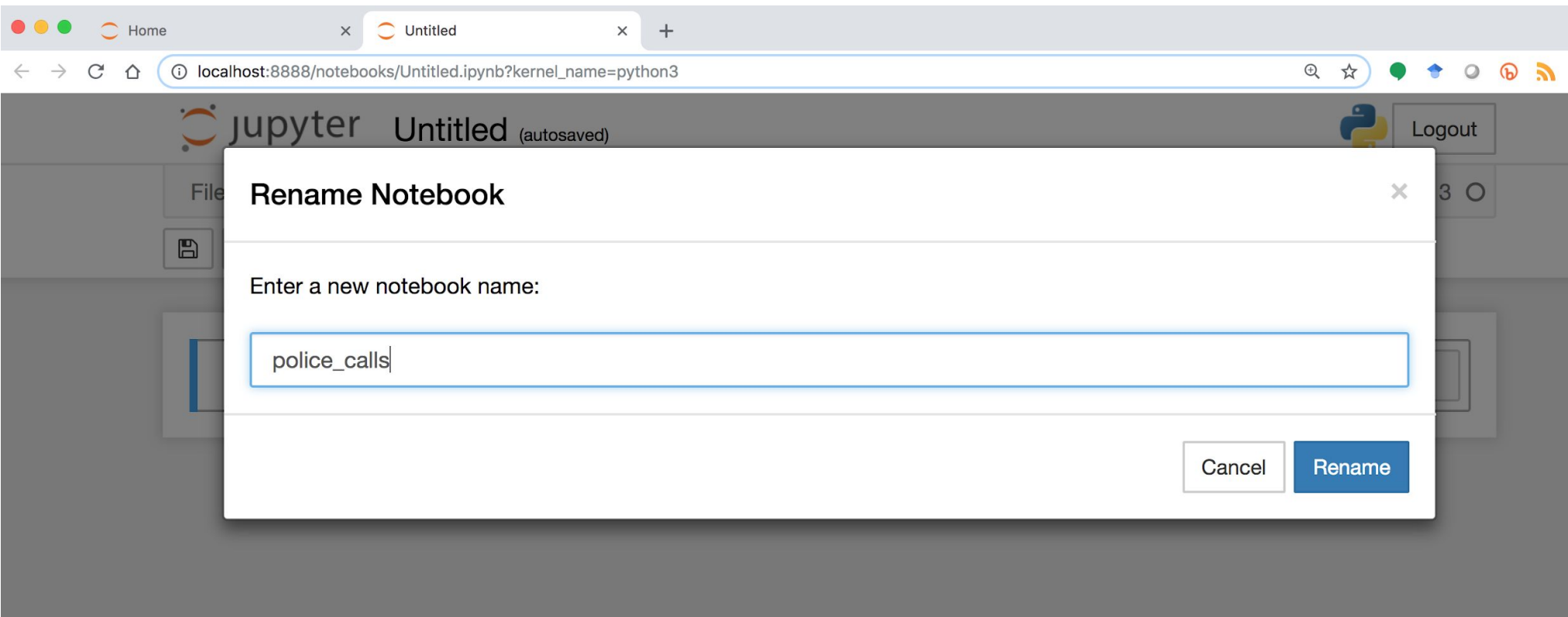
Terminal



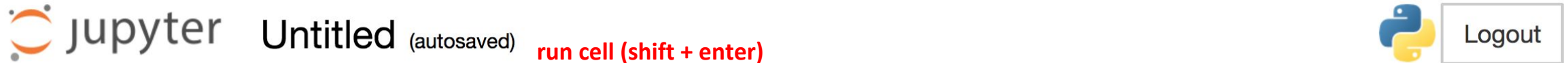
Title the notebook



- meaningful
- no spaces



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



save add cell cut cell

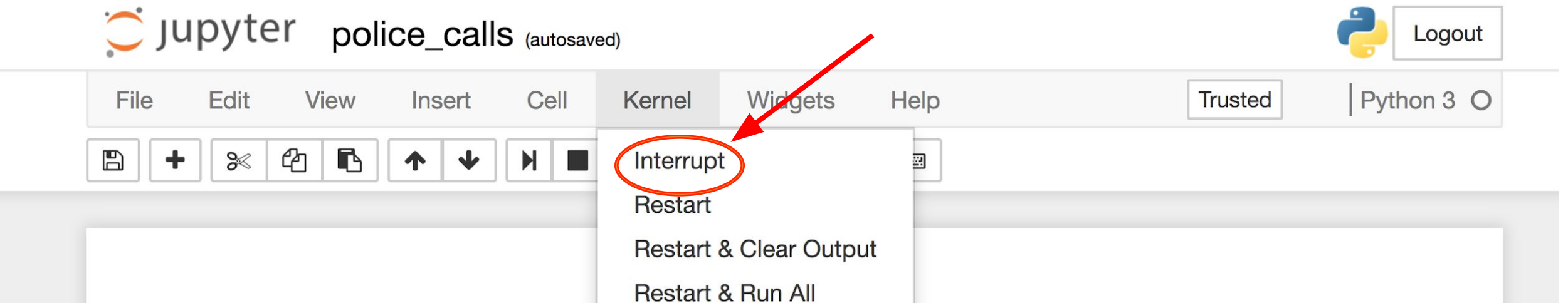
move cell (up/down)

toggle between markdown and code

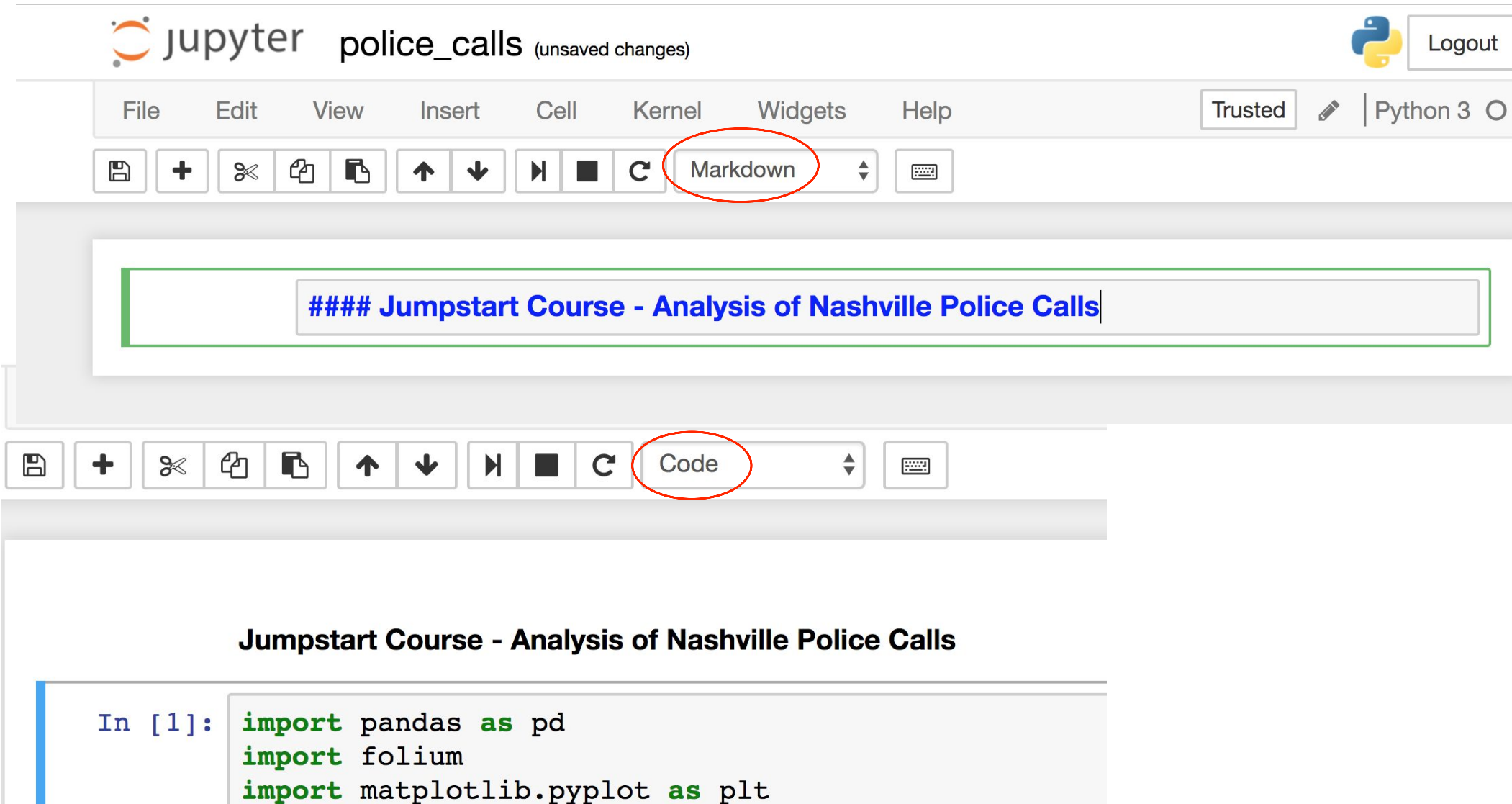
In []:

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 toolbar contains icons for saving, adding, deleting, and other cell actions. The main editor area shows a cell with the text `#### Jumpstart Course - Analysis of Nashville Police Calls`. Below this, another toolbar shows the 'Code' option circled in red. The bottom section of the image shows the rendered output of the code cell, which is the title **Jumpstart Course - Analysis of Nashville Police Calls**. Below the title, the code cell is shown with the following Python code:

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

Next up:

- **walk through analysis guide**

