DATA MINING TUTORIAL

Introduction to Python Libraries

Python

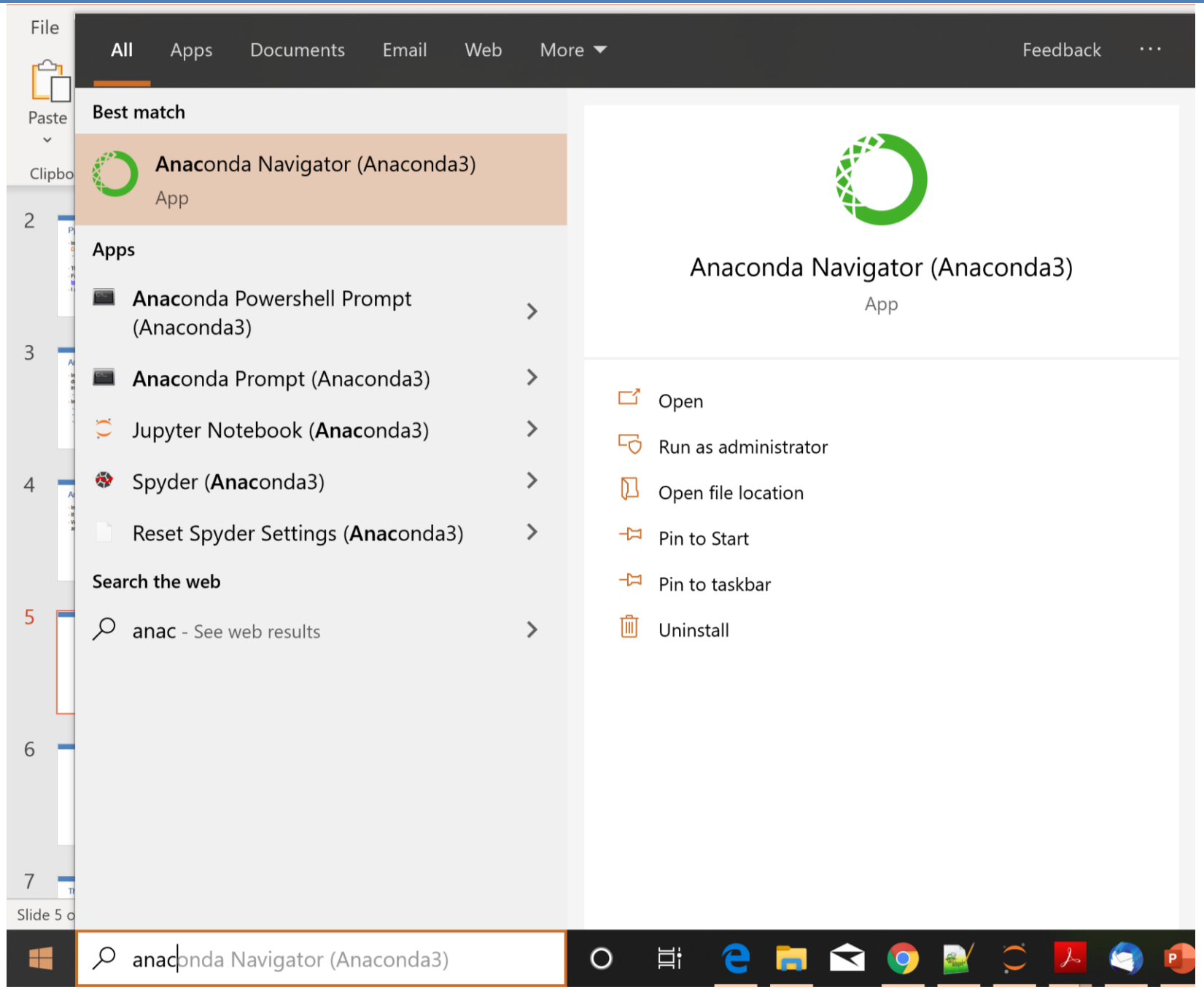
- In the last few years there is an increasing community that creates Data Mining tools in Python
  - There are also tools in other languages but we will use Python whenever we can for a common point of reference.
- There are tons of resources online for Python.
- For an introduction you can also look at the slides of the Introduction to Programming course by prof. N. Mamoulis
- I assume you have installed Python to your laptop by now.

Anaconda

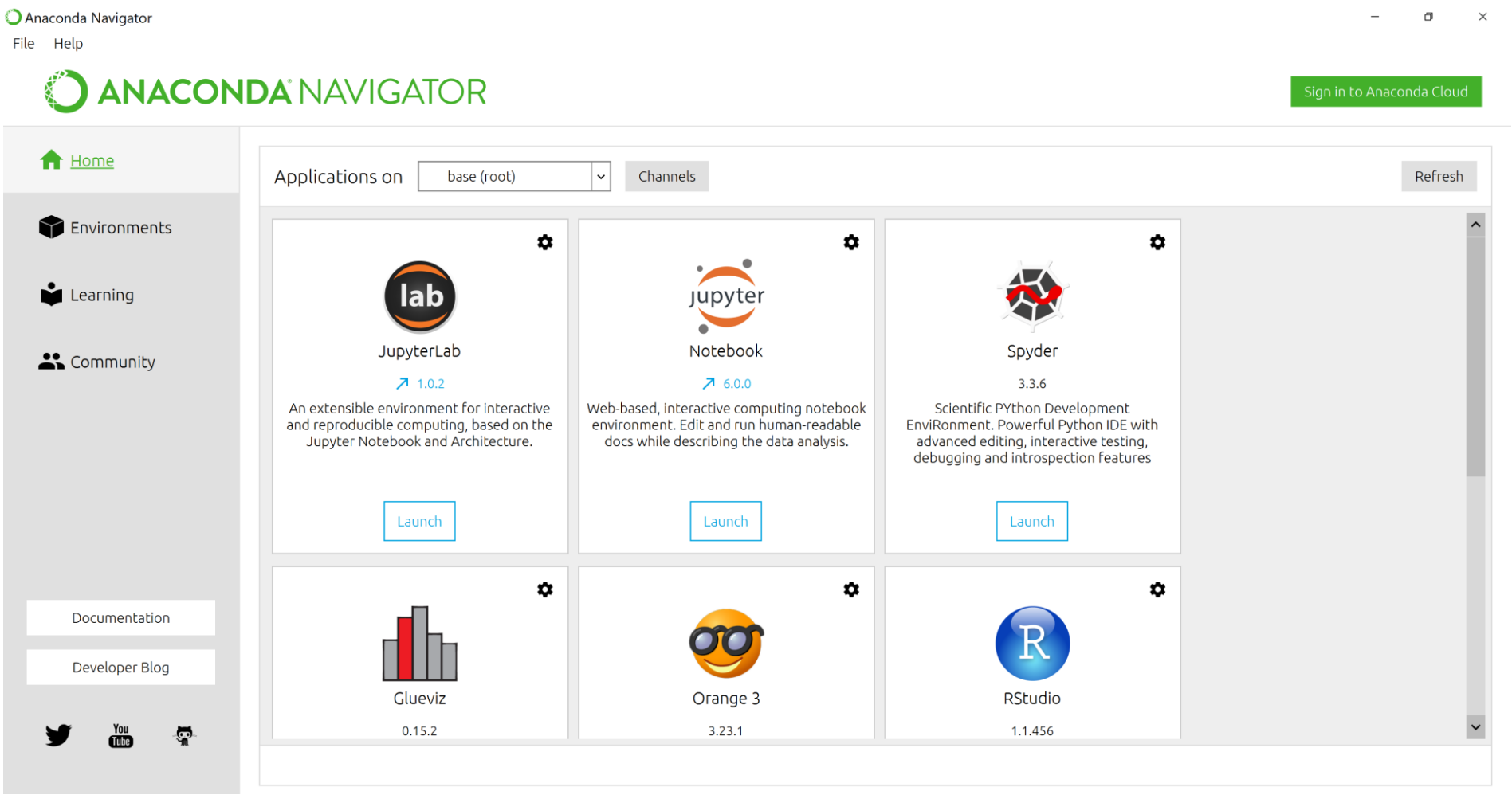
- Installing libraries in Python can be complicated, so you should download the Anaconda Scientific Python distribution which will install most of the libraries that we will use.
  - Use Python 3.0
- Installing Anaconda installs a lot of libraries and also:
  - Anaconda Navigator
  - Jupyter Notebook: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

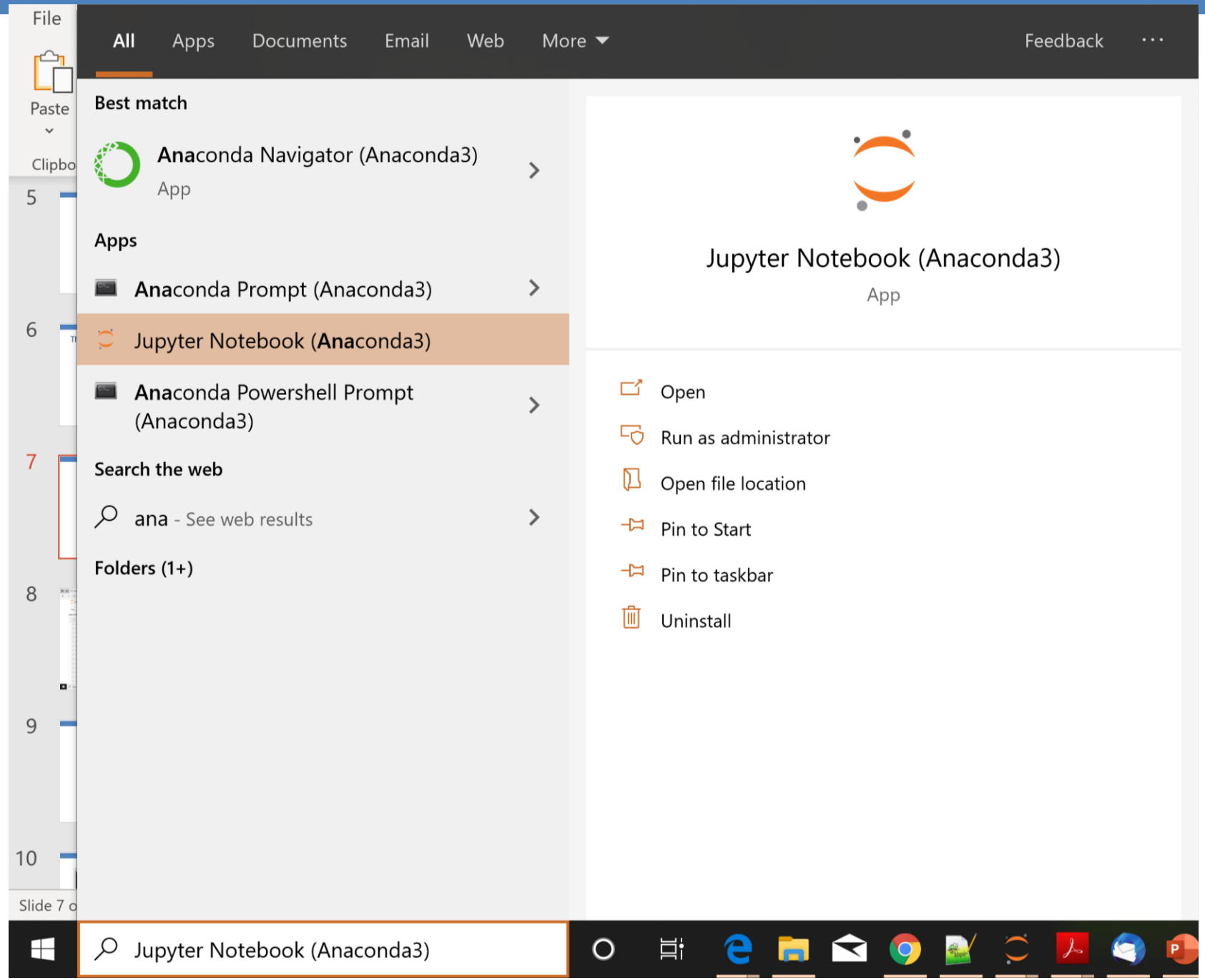
Anaconda

- Installing Anaconda will also install Jupyter Notebook,
- It is very convenient for loading and experimenting with data
- We will use it in our examples, and it is recommended for the assignments as well.


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.

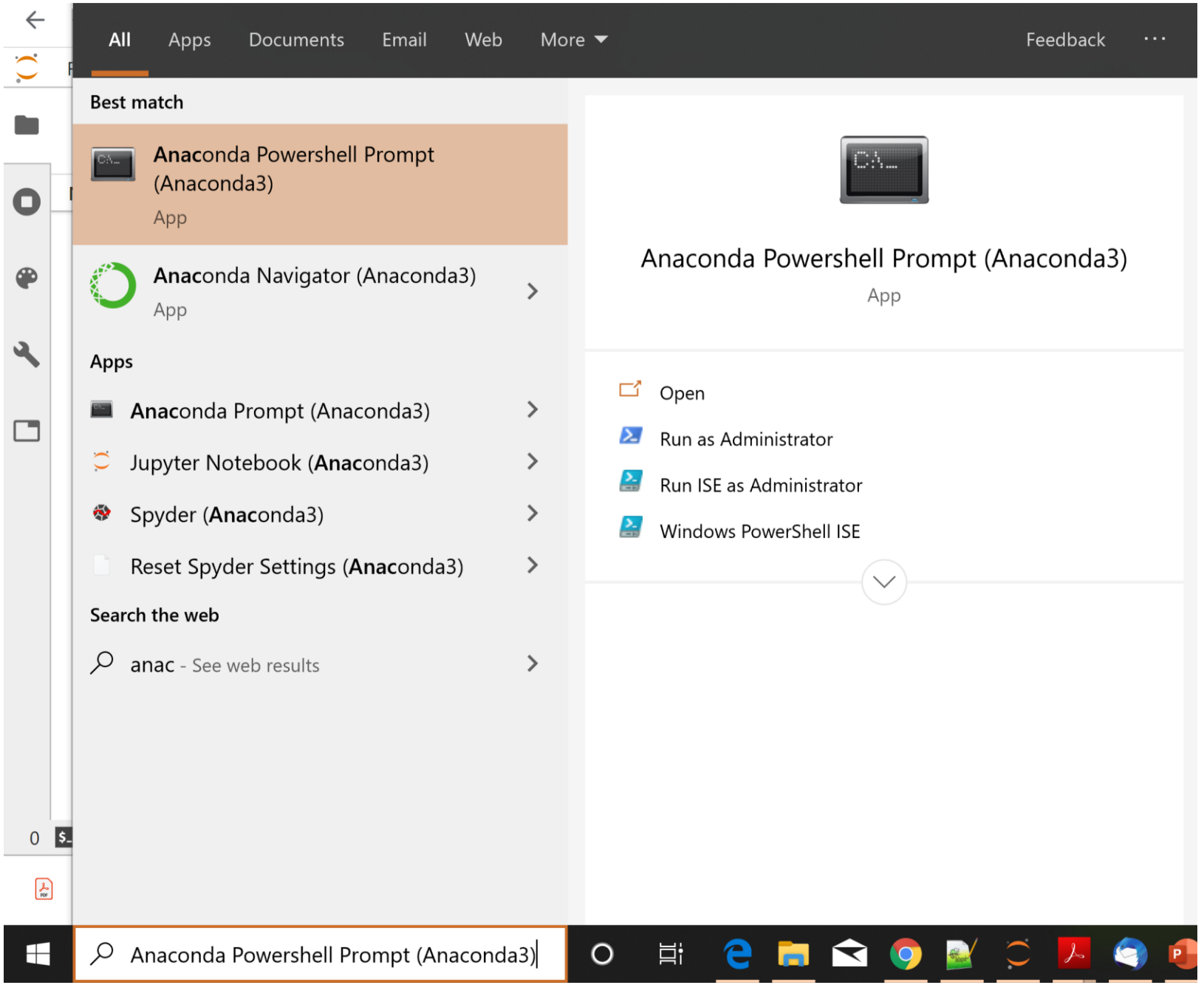
Upload New

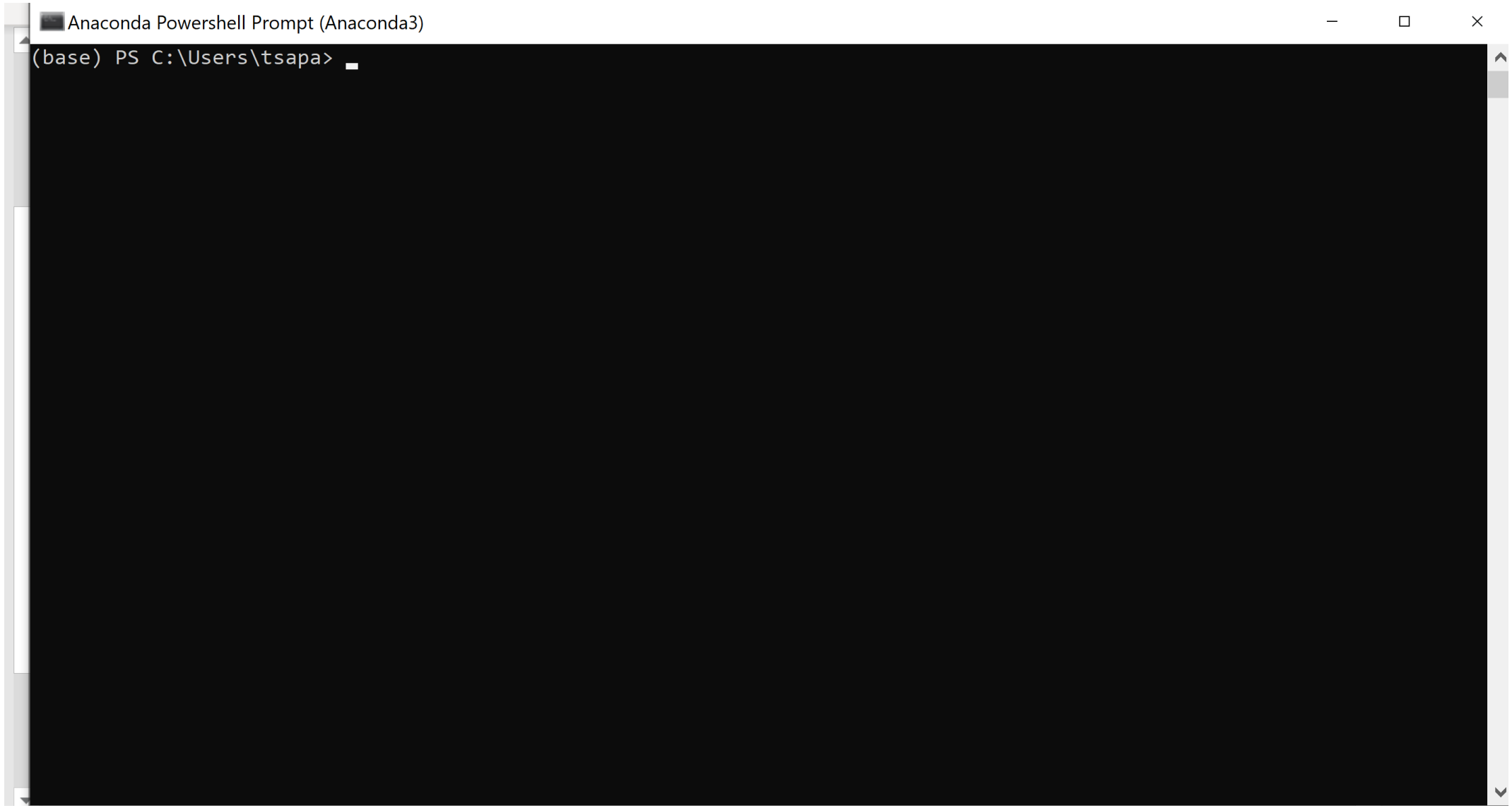
0 /

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`

Changing the notebook default 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

Pandas

- Python Data Analysis Library
  - A library for data analysis of (mostly) tabular data
  - Gives capabilities similar to Excel and SQL but also with some of the Matlab and R capabilities for data matrix manipulation.
- In this class we will cover:
  - Data structures
  - Basic operations
  - Plotting
- The full documentation here. The short version here.