Data Science Project Structure

<project-name>

README.md

* + Welcome/overview – what’s this project? What is it for?
    - TODO – notes from the author, to the author/knowledgeable user
    - Gotchas – any nasty surprises someone should know about
  + Setup/Installation
  + Running

src (code)

data

scripts (load data …) ?

Python

Advantages: (i) Powerful built-in data types (e.g. list, dictionary); (ii) Expansive ecosystem of data science libraries

Text

Description automatically generated

CLI scripts

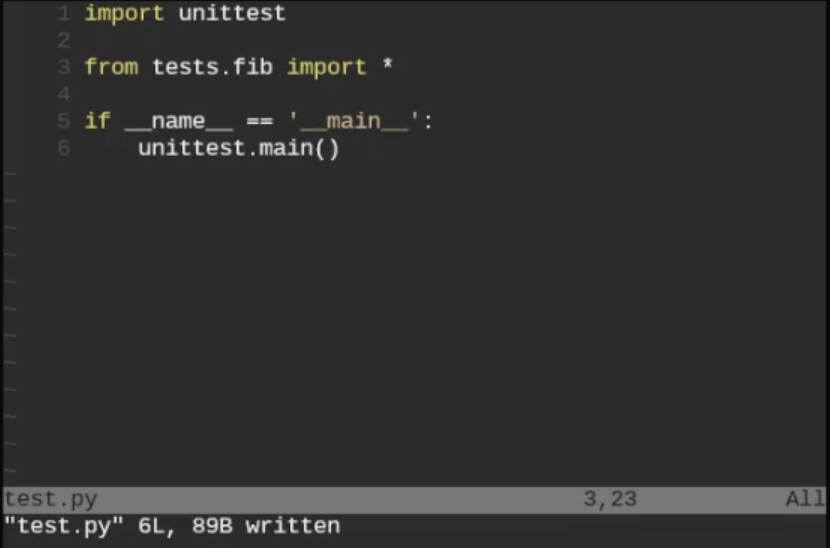
Text

Description automatically generated

Unit testing

Text

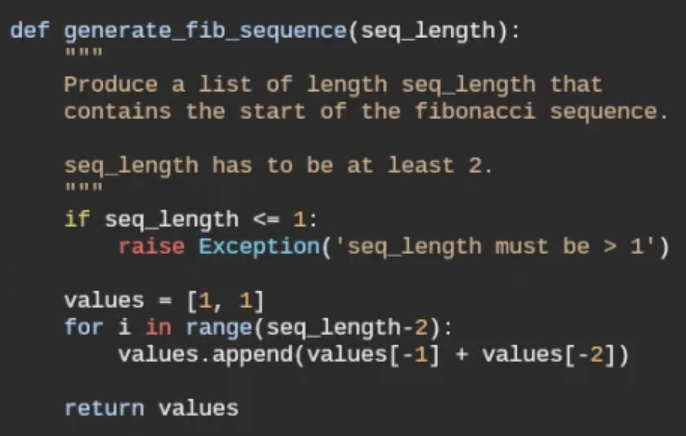
Description automatically generated

Text

Description automatically generated

Exception

Text

Description automatically generated

Resource Referencing

**NEVER use explicit paths (Linus: / ; Windows: \)**

Build paths from

* Input arguments
* Current script location = os.path.dirname(\_\_file\_\_)
* Path separators = os.path.sep
* Path combination functions: os.path.join(...)

Text

Description automatically generated

Version control

* Backup
* Multi-machine work
* Collaboration & Sharing
* Tracking progress

.gitignore (file name this file will be ignored by git add)

Super large files (like data) shouldn’t be checked into a repo

* Use README to explain how to get there, where to setup, etc…

Jupyter

1. Create an SLL certificate

openssl req -x509 -nodes -days 365 -newkey rsa:1024 -keyout mycert.pem -out mycert.pem

mkdir ~/.jupyter/^C

cp mycert.pem ~/.jupyter/

jupyter notebook --generate-config

c.NotebookApp.certfile = '/home/yingw/.jupyter/mycert.pem'

c.NotebookApp.ip = '\*'

c.NotebookApp.open\_browser = False

c.NotebookApp.password = <hashed pw – 981005>

2. Configure Jupyter to use the SSL certificate

3. Attach Jupyter to your external IP address

4. Set your password

5. Open the appropriate port (typically 8888) on your EC2

<https://stackoverflow.com/questions/34869580/how-to-upload-local-system-files-to-amazon-ec2-using-ssh>

Dashboard

bokeh serve --address='\*' --port=8080 --allow-websocket-origin=52.86.231.158:8080 --auth=auth.py circle

52.86.231.158:8080/circle?login=letmein

Text

Description automatically generatedText

Description automatically generated

bokeh serve --address='\*' --port=8080 --allow-websocket-origin=52.86.231.158:8080 --auth=nyc\_dash/auth.py nyc\_dash

52.86.231.158:8080/nyc\_dash?username=nyc&password=iheartnyc