Create your work environment (python virtualenv, for instance).

Install and run cookiecutter for data-science:

https://github.com/drivendata/cookiecutter-data-science

With:

cookiecutter https://github.com/drivendata/cookiecutter-data-science

This is a template that creates some directories, skeleton files, etc. to bootstrap your data science project.

Once created, download raw original data:

- http://codeandbeer.org/virtual/BigData/Datasets/measures.tgz
- http://codeandbeer.org/virtual/BigData/Datasets/cryptocurrencypricehistory.tgz
- http://codeandbeer.org/virtual/BigData/Datasets/iris.data

Tasks

- Using:
  - jupyter:
    - Anaconda
    - Docker
    - Local install (virtualenvs)
    - http://beakerx.com/
  - openRefine
  - Python Scripts
- Read the data. Do it as 3 different projects (in subdirs).
- Clean and prepare the data:
  - temperatures: missing data, diferent formats, several sensors
  - cryptocoins: different coins
- Explore (simple)

Deliverable

- Git repository with the scripts, notebooks, etc. used (not the data)

Help

https://github.com/pandas-dev/pandas/blob/master/doc/cheatsheet/Pandas_Cheat_Sheet.pdf