Organizing projects

Input , procedure, output

Python

Functions

Functions in classes: methods

First order classification: making blocks so data can be manipulated (then sorted, in the example)

Second order: bringing up more atributes

Code to store errors in logs

Data quality is about perspective and how clean it has to be dependes on project parameters.

Revise rules for .js data

.csv module in python

REGEX: regular expressions

Bandit: stats

Relational DB (usually normalized): faster and safer (for changes, ie)

Tldrlrgal.com

Ipython notebook: pip install “ipython[notebook]”

Os.path.join(‘zz’,’yy’,xx’) ----> so the code builds a path that is consequent with the operationd system (c:// for windows or users// for mac)

(in this example this will open or read or whatever: zz, yy, xx files)

.describe() ----> general stats

import matplotlib.pyplot as plt ----> import matplot from pandas

<https://dl.dropboxusercontent.com/u/36407457/Pivot%20Table%20DomConsAirfare.xlsx>

boto ----> importing on Amazon servers (with python)

loads ---> opossite of dump for .jason

floats are treated as aproximatiosn, so, when using numbers they must be converted to integrers (int) ------> java script things that translates to mongo and others. So numbers should be transformed into integrers before uploading to mongo.