Principles of Urban Informatics
Assignment 10

Posted on: 11/17/2014
Due Date: 11/24/2014

Introduction

The purpose of this assignment is to get you familiarized with *pandas*. It is mostly about reading and learning pandas through examples and lessons. It is a preparation for a future assignment, when you will use *pandas* in a different domain.

Steps to complete this assignment include having *pandas* installed in your dev environment, following online tutorials that explain fundamental concepts of the library, and solving programming lessons to exercise the concepts. *Tasks* are mandatory, but no deliverable (code/report) is expected, while *Problems* expect code deliverables. *Problem 2* is optional (extra points).

Task 1 - Installing

We suggest installing *pandas* using *pip*, but if you prefer other method, there are further instructions at: `http://pandas.pydata.org/pandas-docs/stable/install.html`

Task 2 - Learning fundamentals

Follow tutorials at:

- Loading data: `http://goo.gl/L7CCCD`
- Data summary: `http://goo.gl/l4XlJQ`
- Querying/selection: `http://goo.gl/0gkiMg`

Problem 1 - Complaints over time

In lesson *Querying/selection* (`http://goo.gl/0gkiMg`) you learned how to plot a bar chart with the number of complaints per borough. Now, use the *creation date* column to plot a timeseries with the number of complaints over time, for each borough.

You can download the 311 complaints from `https://nycopendata.socrata.com/Social-Services/311-Service-Requests-from-2010-to-Present/erm2-nwe9`. Use data for complaints created between 06/01/2014 and 08/31/2014.

Your program should run as *python problem1.py [csv file]*, where *[csv file]* should be replaced by the .csv file downloaded from the NYC Open Data. The expected output is a plot with 5 timeseries, one for each NYC borough. Add labels and colors to identify the curve for each borough.

[Extra points] Problem 2 - Baby names in 1880

Complete the lesson at: `http://goo.gl/fqhFUm`. You can notice that the final plot shows the names ordered by lexicographical order. Your must modify the code to show the names ordered by number of births.

Your program should run as *python problem2.py*.

How to submit your assignment?

Your assignment should be submitted using the NYU Classes system. You should submit all the requested files in each problem in a zip file named NetID_assignment_10.zip, where you should change *NetID* by your NYU Net Id.

Grading

Your programs should be executable, i.e., you must not send only the missing code, you must include the complete code for each lesson. It is part of the assignment to understand (and research if necessary) the concepts asked.

Try to test your code before submitting: your script solves the problem when you achieve the desired results in each lesson.

References

- *pandas reference*: `http://pandas.pydata.org`