

COMP41680 Assignment 1: Data Collection & Preparation

Deadline: 5pm, Friday 23rd March 2018

Overview:

The objective of this assignment is to collect a dataset from one or more open web APIs of your choice, and use Python to pre-process and analyse the collected data.

The assignment should be implemented as a single IPython Notebook (not a script). Your code should be clearly documented, using comments and Markdown cells to explain your code and results.

Tasks:

In this assignment you should complete all of the following tasks:

1. Choose one or more open web APIs as your source of data. If you decide to use more than one API, the APIs should be related in some way.
2. Collect data from your chosen API(s) using Python. Your dataset should contain at least 100 records/items in total. Depending on the API(s), you may need to repeat the collection process multiple times to download sufficient data.
3. Parse the collected data, and store it in an appropriate file format for subsequent analysis (e.g. JSON, CSV, XML).
4. Load and represent your data as a Pandas DataFrame. Apply any pre-processing and quality checking steps that may be required to clean and filter the data before analysis.
5. Analyse and summarise the cleaned dataset, using tables and plots where appropriate. Based on your results, what interpretations or insights can be made about the dataset? What further analysis might be done on the data? Detail this using Markdown cells in your notebook.

Guidelines:

- Submit your assignment via the COMP41680 Moodle page. Your submission should clearly state your full name and student ID number. Include your full dataset (or a sample of it) with your submission. Your submission should be in the form of a single ZIP file containing the IPython notebook and the data.
- This assignment should be completed individually. Any evidence of plagiarism will result in a 0 grade.
- Hard deadline: Submit before 5pm on Friday 23rd March 2018
 - 1-5 days late: 10% deduction from overall mark
 - 6-10 days late: 20% deduction from overall mark
 - Assignments will not be accepted after 10 days without an extenuating circumstances form and/or a medical certificate.