

COMP30760 Assignment 1

Deadline: Wednesday 11th November 2020

Overview:

The objective of this assignment is to collect a dataset from one or more open web APIs, and use Python to prepare, analyse, and derive insights from the collected data.

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

Tasks:

In this assignment you should complete the following tasks:

1. Data Identification and Collection:

- a) Choose one or more APIs from the list of public web APIs provided at the link below. If you decide to use more than one API, these APIs should be related in some way.

<http://mlg.ucd.ie/modules/COMP30760/assignment1.html>

- b) Collect data from your API(s) using Python. Depending on the choice of API, you might need to repeat the collection process multiple times to download sufficient data for analysis.
- c) Save the collected dataset in JSON format for subsequent analysis.

2. Data Preparation and Analysis:

- a) Load the stored JSON dataset, and represent it using an appropriate structure (i.e. records/items as rows, described by features as columns).
- b) Apply any preprocessing steps that might be required to clean, filter or engineer the dataset before analysis. Where more than one API is used, apply suitable data integration methods.
- c) Analyse, characterise, and summarise the cleaned dataset, using tables and visualisations where appropriate. Clearly explain each step of this process, and interpret the results which are produced. Markdown cells should be used for the explanations and interpretations.
- d) At the end of your notebook, summarise any insights which you gained from your analysis of the dataset. Suggest ideas for further analysis which could be performed on the data in future.

Guidelines:

- The assignment should be completed individually. Any evidence of plagiarism will result in a 0 grade.
- Submit your assignment via the COMP30760 Brightspace page. Your submission should be in the form of a single ZIP file which contains:
 1. Your Jupyter notebook. This should be a IPYNB file, not a HTML file.
 2. Your JSON dataset. If your dataset is too large to upload to Brightspace, please include a smaller sample of the data in the ZIP file.
- In your notebook please clearly state your full name, student number, and the API(s) which you have chosen for your assignment.
- Hard deadline: Submit by end of 11th November 2020
 - 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 extenuating circumstances.