

SIT 112 | Data Science Concepts

Lecturer: Dr Truyen Tran | truyen.tran@deakin.edu.au

Assistant: Trang Pham | phtra@deakin.edu.au

ASSIGNMENT TWO Due: Friday, 5pm 15th September, 2017

This assignment contributes 30% to your final SIT112 mark. The assignment must be completed and submitted to *CloudDeakin* by **5pm, 15th September 2017**.

The theme for this assignment, and the subsequent data science project, is to continue to explore data problems related to Australia. In this assignment, you are provided with sets of keywords. Each set contains three keywords. You will need to select any set of three keywords and perform two tasks

- Crawling data from Twitter within Australia based your selected keywords
- Performing analytics tasks as instructed in the IPython notebook supplied.

The sets of keywords are as follows,

- computer, animal, movie
- game, bike, fiction
- food, holiday, song
- restaurant, singer, fun
- gift, science, actor
- study, actor, fashion

1. Assignment Resources

In the Assignment 2 folder, you will find the following files:

Filename	Description
Assignment2_notebook.ipynb	This is the IPython notebook which has been prepared and partially completed for this assignment. You will need this notebook to complete your assigned tasks.
Assignment2_instructions.pdf	This is the assignment instruction file which contains information for submission instructions.

2. Task Description

The description of assigned tasks has been given in details in the assignment notebook. **You are required to follow the instructions in the notebook to complete your tasks.**

3. Submission Instruction

You are required to submit **five** files:

- Your solution notebook: **[yourID]_assignment2_solution.ipynb**
- An exported version of your output: **[yourID]_assignment2_output.html**
- Three json data files storing tweets collected for three keywords, named **[yourID]_w1.json**, **[yourID]_w2.json**, **[yourID]_w3.json** where **w1**, **w2**, **w3** are three keywords you selected.

END OF ASSIGNMENT DESCRIPTION