



3032ICT / 7230ICT / 1117ICT

Big Data Analytics and Social Media

Assignment Specifications

Instructions

- **Structure:** This assignment is broken up into two milestones. You complete Milestone 1 first, then later Milestone 2, which includes a video presentation. The milestones are aligned to your progress through the contents of the course and the general data analytics cycle.
- **Due:** See course site on Learning@Griffith for the due dates of each milestone.
- **Late Submissions:** An assessment item submitted after the due time on the due date set by the Course Convenor, without an approved extension, will be penalised. Assessment items submitted after the due time on the due date will be penalized at a rate of 5 percent (%) for each calendar day the assessment item is late. Assessment items submitted more than seven calendar days after the due date will be awarded zero marks.
- **Extensions:** If for any valid reason (e.g., being sick) you need an extension, you must apply for an extension by the due date of the milestone through this online form: <https://www.griffith.edu.au/students/assessment-exams-grades/assessment-applications>

Overview

In this assignment, you are required to think about **a case study, in which you can apply social media analytics to gain insight about how a certain music artist or band can improve their popularity**. You will need to **describe the setting for your case study**, apply social media analytics using the tools introduced during the labs, and evaluate your findings and determine appropriate future actions.

You are required to use software for analysis and produce a written report. In the report, you need to explain your **analysis findings**. The report accounts for the majority of marks for each question. Simply pasting screenshots of your analysis outputs will not give you full marks. Accuracy and reproducibility of your code will be checked. You will also need to present your findings in a recorded video presentation and attend an interview to answer questions on your report/presentation.

- **Choose data sources and data** that are appropriate for your case study.
- Pay attention to **how much data you retrieve and how frequently you retrieve data**. If you try to get lots of data often, the APIs will impose a rate-limit on your account. However, you can still proceed after the rate-limit has ended.
- Use the **software** introduced in the labs (RStudio, Gephi, Tableau).
- Add **headings in your R scripts** so that we can easily find the code related to each question.
- **Export** all datasets as .RData files (so that we can re-run your code if needed).
- Add **screenshots of your results in the report**. Include in the screenshot the code you wrote to produce that result. (Usually this would be the last line of code.)
- Make **plots/visualisations** wherever possible, using R functions, Gephi, and Tableau. (Most results can be displayed as a plot/visualisation!)

Advice

1. Complete each part of this Assignment as you are going through the relevant learning module in the course. For example, Milestone 2 includes a part on “Network Analysis”. Work on that as you go through the “Network Analysis” module. Otherwise, you may not be able to catch up and finish by the due date. This is particularly important because you are also dependent on the API rate limits.
2. Make sure to read through all the information in detail & before you start (everything in this document, the marking rubric, and the submission instructions on the course site).
3. Make sure to address all parts of each Assignment question and use the marking rubric to guide you.
4. Start answering the Assignment questions by going back to the lab scripts and altering them to fit your case study based on the specifications. Then, use the rubric to improve your answer incrementally.
5. Once you are ready to submit a Milestone, make sure to follow the submission requirements, otherwise you will lose marks. The submission requirements are posted on the course site and further explained in this document and the marking rubric.

Milestone 1

- Choose a well-known artist or band. Assume you are the artist's/band's manager and want to help improve their popularity by using social media analytics.
- Your chosen artist/band should be well-known already so that there **exists enough social media data** that is somehow related to it. Otherwise, you may not be able to retrieve enough useful data for performing the analytical steps later.

Case Study Setting

- 1.1) Describe the artist/band you are managing. Make sure to reference your sources properly (don't plagiarise). Use APA referencing style.

For example:

- How many years have they been active?
- How many albums & songs have they published?

[1-2 paragraphs, 2 marks]

Data Selection & Exploration

- 1.2) Collect data about your artist/band from Twitter. Make sure to choose keywords for data retrieval that are most relevant to your artist/band. However, try not to be too narrow. As a rough guide, you should retrieve at least 1000 tweets. **List the keywords, explain** your search strategy, and **how much data you have collected**.

(=> see Lab 2.1 for help)

[1.5 marks]

- 1.3) List the **top 5 most influential users** for your artist/band. Find out **what other interests/characteristics they have besides** those related to your artist/band. **Do these 5 have something in common?** (=> Lab 2.1)

[2.5 marks]

- 1.4) List the top **10 most important terms** that appear together with your keyword(s) related to your artist/band. **Explain** the results. (=> Lab 2.1)

[1.5 marks]

- 1.5) Calculate how many unique user accounts there are in your dataset. Explain the code you have used for the calculation. What do the results tell you?

[2.5 marks]