 Faculty of Engineering and Technology

# Coursework Title: AS1

# Data processing and visualisation Group task

**Module Name:** Computer Science Workshop

**Module Code:** 4216COMP

**Level:** 4

**Credit Rating:** 20

**Weighting:** 100% of module

**Maximum mark:** 100

**Lecturer:** Dr Pavel Semukhin

**Contact:** Email: p.semukhin@ljmu.ac.uk

Room: 607b

**Hand-out Date:** 18th Jan 2022 (week 17)

**Hand-in Date:** 5th Apr 2022 (week 28). Presentation in class on the 5th Apr 2022

**Hand-in Method:** Canvas Assignment Handler (presentation in person)

**Feedback Date:** Three weeks after submission, i.e., 26th Apr 2022

**Feedback Method:** feedback document, via email/canvas

## Introduction

This assignment is a group task, whereby you will establish a project team of *five or six students*. Your project team must *work independently from other teams.* Your task will be to identify a freely available dataset, then process and analyse the data using suitable Python packages and produce visual outputs e.g. plots, graphs, or maps.

**Learning Outcomes to be assessed**

1. Understand how to work as a team towards a shared goal.
2. Share their ideas on a problem in a structured way for the benefit of the student's technical team.
3. Analyse and contribute to the behaviour of a program written by someone else.
4. Utilize the facilities of modern IDEs and SCMs to collaborate with others and to streamline development in a group-working context.

**Detail of the task**

*Preamble:* This assessment requires you to work in a group. Your group membership will not be alterable during the semester (unless some special circumstances arise). To coordinate your efforts, every week your team will be expected to agree a plan for the next week and prepare a summary of the work that was done in the week before. This will be assessed by the lab instructors during the group sessions and form part of the final mark.

**Project workflow**

**Task 1) As a team:**

* Form a team of 5-6 students, choose your team’s name. The team leader role will rotate; each group member will take control for approximately two weeks.
* **Conduct regular weekly meetings and document the minutes. The minutes should have clearly identifiable actions that are assigned to individuals in the group. Each week these actions should be reviewed.**
* Setup a source code repository on GitHub and share it with the module leader (guest access is not sufficient). Your repository must be titled as follows: 2022-4216COMP-yourgroupname (where *yourgroupname* is replaced by your group’s name).
* Identify an open data set that you will be working on (e.g., on https://data.gov.uk/ or <https://github.com/awesomedata/awesome-public-datasets>). Agree your choice with the module leader or your lab instructor.
* Identify insights you would extract from the data and how you would enable exploration e.g. menu, keyword search (**requirements**). For example, if it is a population data, you may want to search and sort the data by towns, regions, years, etc. If it is a timetable data, you may want to search for an optimal route between two destinations. However, the requirements will depend on the data set you choose and will be different for each team.
* Identify algorithms that you will use to process and analyse the data. Design interfaces to interact with the application (**design**).
* Prepare a **specification document** that describes your insights into the chosen data set as well as your requirements and design.
* **Implement** the algorithms. At this stage the output of your algorithms should be in a text form, either as console output or saved in a text file. You don’t need to implement them from scratch. Instead, you can use existing Python packages (e.g., Pandas). You should use Git for source code management.
* Attempt performance optimisation of the implementation or implement more advanced algorithms (e.g., clustering).
* **Demonstration:** On the last week, each group will be required to give a 20 minute long demonstration of the accomplishments of the group. The demo must be done, as a group, and you should all contribute. During the demo, you will be required to explain what data processing algorithms you’ve implemented and demonstrate how they work. Teams are advised to prepare some additional support materials (e.g., slides) for your final demonstration. ALL STUDENTS MUST BE PRESENT AT THE DEMO. Exceptions need to be agreed prior to the demo with the module leader.

**Task 2) Individually:**

**Once you have implemented the core data processing and algorithmic work as a group, you should create a feature branch in git that contains a visualisation of the data.** This visualisation should be written in Python and make use of existing Python packages (e.g., Matplotlib, NetworkX). Your visualisation should make up part of the group presentation.

This is *not an exhaustive specification*: to aid analysis, you can consult the module leader or lab instructors regarding specification queries, or to solicit feedback on development (e.g., designs, requirements, prototypes, etc). Beyond this, you can make (and document) assumptions based on your research and what you believe feasible in the allowed time.

In addition to the lab sessions, teams are expected to have regular meetings (either face-to-face or online using on Teams/Zoom) to discuss their progress and coordinate efforts. **Each team is responsible for organizing and setting the agenda for these meetings.**

IMPORTANT: Your entire team takes responsibility for submitting the group report that will also contain the link to the git repo. As such, it is strongly recommended that your team devise, agree and use a submission protocol where EVERY MEMBER checks the submitted work (on the Canvas submission section) and ensures it is correct and as expected. **Ensure this is not left to a single team member!**

Conversely, **you** take responsibility for your own individual contribution (data visualisation part). As such, non-submission of this component will be considered a non-submission for you in this assessment.

**What you must hand in**

You should prepare the following assessed components:

**Task 1) As a group you should submit**

* Project Report document that includes the following:
  + - Discussion of the open data considered by the group and the rational for the final selection
    - Assessment of the potential insights that might be extracted from the data set and the interfaces needed to interact with your application (requirements)
    - Algorithm exploration with a description of any optimisations made.
    - Design that identifies how the application will process the data and interact with the user.
    - Description of the implementations.
    - Testing.
    - 10 Project meeting minutes.
    - Evaluation.
  + This report must be provided as a single technical report in PDF or Word DOC/DOCX format.
* Source code via a private source code repository (e.g., GitHub) setup and shared with module leader (guest access is not sufficient). **The source code must include a readme.txt file that clearly explains how to run your code**

**Task 2) As an individual you should submit**

Individual report identifying the visualisations that you have created, and the libraries/approach used. This must be provided as a single Word DOC/DOCX file.

* Individual data visualisation code should be provided in the same repository, in a separate folder and or in a separate feature branch clearly identified by your name e.g. NAME-visualisation-feature. You should have a section in the main readme.txt identifying how to run your visualisation.

**Recommended literature:**

Havill, Jessen. **Discovering Computer Science: Interdisciplinary Problems, Principles, and Python Programming**. Boca Raton, Florida; CRC Press, 2016.

Gries, Paul et al. **Practical Programming: An Introduction to Computer Science Using Python 3.6.** Raleigh, North Carolina: The Pragmatic Bookshelf, 2017.

Kopec, David. **Classic Computer Science Problems in Python.** Manning Publications, 2019.

**Online tutorials:**

Git book: <https://git-scm.com/book/en/v2>

Matplotlib: <https://matplotlib.org/stable/index.html>

NetworkX: <https://networkx.org/documentation/stable/tutorial.html>

**Marking Scheme/Assessment Criteria**

Your final mark will be determined via two tasks: 1) teamwork exercise (this will be determined by the total group mark and weighted by the sum of your weekly contributions) and 2) your individual contribution as detailed in the following table.

|  |  |  |
| --- | --- | --- |
| **Task** | **Assessment Criteria** | **% weighting** |
| 1 a) | **Requirements and design** | **15** |
| 1 b) | **Project management and usage of a source control repository** | **15** |
| 1 c) | **Group Implementation and testing** | **40** |
| 1 d) | **Demonstration** | **10** |
|  | **Total group marks** | **80** |
| 2 | **Data visualisation (marked individually)** | **20** |
|  | **Total marks** | **100** |

In addition, you are issued a more detailed assessment rubric / grading grid, which also will be used for the assessment.

**Personal/Extenuating Circumstances:** If something serious happens that means that you will not be able to complete this assignment, you need to contact the module leader as soon as possible. There are a number of things that can be done to help, but we can only arrange this if you tell us.

*To ensure that the system is not abused, you will need to provide some evidence of the problem.*

More guidance, including on the Personal Circumstances process, is available in the appropriate Academic Framework document, available here:

<https://www.ljmu.ac.uk/about-us/public-information/student-regulations/guidance-policy-and-process>

**Any coursework submitted late without the prior agreement of the module leader will receive 0 marks.**

**Academic Misconduct:** The University defines Academic Misconduct as ‘any case of deliberate, premeditated cheating, collusion, plagiarism or falsification of information, in an attempt to deceive and gain an unfair advantage in assessment’*. This includes attempting to gain marks as part of a team without making a contribution*.

The Faculty takes Academic Misconduct very seriously and any suspected cases will be investigated through the University’s standard policy. If you are found guilty, you may be expelled from the University with no award.

<https://www.ljmu.ac.uk/about-us/public-information/student-regulations/academic-misconduct>

**It is your responsibility to ensure that you understand what constitutes Academic Misconduct and to ensure that you do not break the rules. If you are unclear about what is required, please ask.**