Module: COMP1844 Information Analysis and Visualisation	Coursework
Contribution: 100% of the grade	Coursework submission: A single PDF document submitted on Moodle
Module leader: Konstantin Kapinchev	Due date:
Approximate time to complete the coursework: 50 hours	
Learning outcomes: 1. Identify and discuss fundamental concepts related to information analysis and visualisation 2. Demonstrate an understanding of different types of information visualisation and identify appropriate types of visualisation for various types of data 3. Apply analytical and visualisation tools and techniques to obtain insight from datasets	

Plagiarism is presenting somebody else's work as one's own. It includes copying information directly from online resources or books without referencing the material, submitting joint coursework as an individual effort; copying another student's coursework, purchasing coursework from someone else and submitting it as own work. Suspected plagiarism will be investigated and if found to have occurred will be dealt with according to the procedures set by the University. All material copied or amended from any source must be referenced correctly according to specific reference style.

Courseworks will be submitted for electronic plagiarism checks.

Coursework Submission Requirements:

- An electronic copy of the coursework should be uploaded on Moodle before the deadline

- The last uploaded version will be the one that is marked

- The format of the coursework is a single PDF document containing all answers including:
    - Python source code
    - Generated images
    - Obtained numerical results

- The limit of the file size is 100 MB

- The PDF document should be virus-free, not protected by a password or corrupted, otherwise it will be treated as not submitted

- Grade and feedback will be available on Moodle

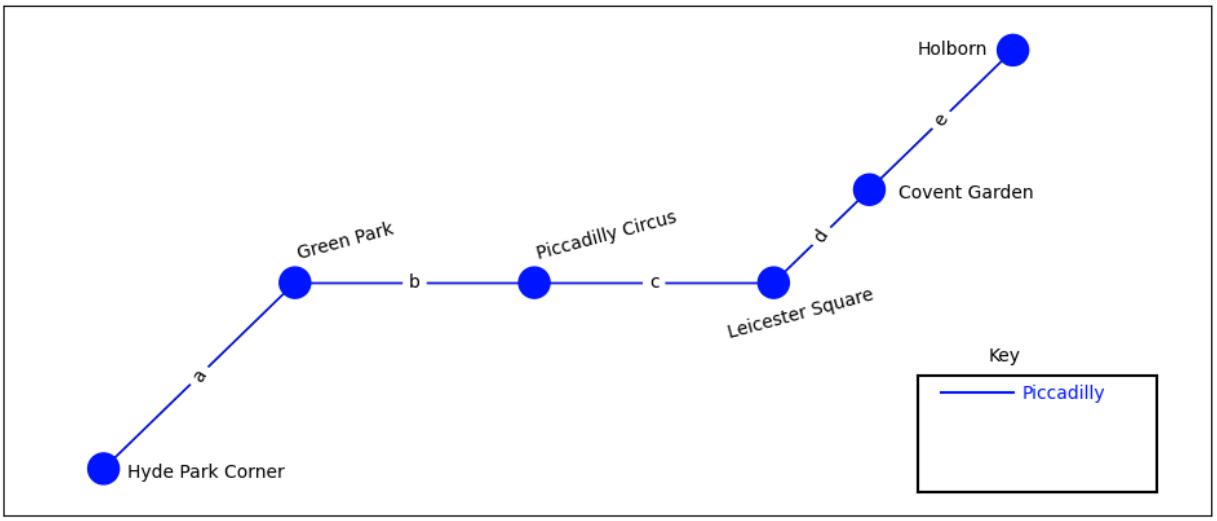
In order to be marked, the source code needs to meet the following requirements:

1. **To be selectable as text**
2. **To be error free (except errors due to indentation)**
3. **To generate the required visualisation and results**

The University website has details of the current Coursework Regulations, including details of penalties for late submission, procedures for Extenuating Circumstances, and penalties for Assessment Offences. See https://www.gre.ac.uk/student-services/regulations-and-policies for details.

Coursework Specification:

Consider the following image, generated by a **Python program**, which represents part of the Piccadilly line, which is part of the public transport network of London:



Task 1

By using **graph data structure**, develop a **Python program**, which generates the above image. Replace the edge labels, which are currently letters, with the actual distances between the stations, in kilometers or miles. Use online map services, such as Apple Maps or Google Maps, to approximate the distances.

Task 2

In its current state, the map shows only one line with six stations. Expand the **Python program** by adding more **lines** and more **stations**. Follow a consistent colour scheme. The resulting map is expected to have **minimum four lines** with **minimum six stations** on each line, including interchange stations. The generated graph is expected to be **connected**, that is, there is always at least one route between any pair of stations. The generated graph is expected to represent accurately a section of the actual map of the transport network. **A complete map of the transport network is not expected**. Complete the **Key** according to the generated map. By following the same style as shown in the image above, add distances between the stations.

Task 3

Extract the following data from the graph generated in **Task 2**:

Total length of the transport network
The average distance between the stations
The standard deviation of the distances between the stations

The coursework document is expected to have the following Content:

1. **Source code for Task 1**
2. **Image generated by code for Task 1**
3. **Source code for Task 2**
4. **Image generated by code for Task 2**
5. **Source code for Task 3**
6. **Data generated by code for Task 3**

In order to be marked, the Python program is expected to use only the following four Python libraries:

1. **NumPy**
2. **pandas**
3. **NetworkX**
4. **Matplotlib**

While it is not necessary for all suggested libraries to be used, no other libraries are expected to be added.

The following are the suggested first lines of the Python program:

```
import numpy as np
import pandas as pd
import networkx as nx
import matplotlib.pyplot as plt
```

Marking Scheme

Task	Achieved well	Partly achieved	Poorly achieved or not achieved
Task 1 (20%)			
Generated map is equivalent to the image provided	10-8	7-3	2-0
Accurate distances are added as edge labels	10-8	7-3	2-0
Task 2 (60%)			
Visualised graph consists of minimum 4 lines with minimum 6 stations on each line	15-11	10-5	4-0
The graph is connected	5-4	3-2	1-0
Consistent colour scheme is utilised	5-4	3-2	1-0
The generated graph represents accurately section of an actual map of the transport network	10-8	7-3	2-0
Accurate Key is visualised	10-8	7-3	2-0
Edge labels are utilised	10-8	7-3	2-0
Accurate distances are provided	5-4	3-2	1-0
Task 3 (20%)			
Total length	5-4	3-2	1-0
Average distance	5-4	3-2	1-0
Standard deviation	10-8	7-3	2-0