Administration
Course administration

Staff block (not visible to students)
Click on the link for the Informatics Assessment Calendar 22/23: LINK

Activities
Assignments
External tools
Forums
Lessons
Resources

King's Resources
Email
Student Services Online
King's Academic Skills for Learning
Disability Support
IT Services
Library Guides
LinkedIn Learning
Health & Wellbeing

Coursework 2 Instructions

This coursework asks you to analyse the road network and road events in the centre of a UK city, Leeds. The coursework consists of a series of connected tasks. In each task, you should consider and apply the ideas and techniques learnt in the module along with your technical skills. The tasks are not a set of options: all three tasks must be completed by each student.

Primary Data:

The road network of Leeds is available from OpenStreetMap, while datasets on road traffic accidents in Leeds over a few years are available from the following source:

https://data.gov.uk/dataset/6efe5505-941f-45bf-b576-4c1e09b579a1/road-traffic-accidents

Task A (spatial networks and planarity):

In this task, you construct and investigate the road network. You are asked to choose an area of roughly 1 square kilometre around the centre of Leeds for your analysis. You should look for an area where a significant number of the recorded road accidents occurred in the area, e.g. 300 or more total across multiple years, and show you have tested this in your report. You only need to consider roads used for driving, not walking paths or private roads (investigate the osmnx documentation for how you might do this filtering). Your report should give the coordinates of the area chosen. You should then answer the following questions:

1. What are the characteristics of this road network? Include, at least, the spatial diameter of the network, the average street length, node density, intersection density, and edge density.
2. What is the average circuitry of the network? What does this tell you about the efficiency of using roads in this area?
3. Is the network planar? Why/why not?

Task B (road accidents):

In this task, you should investigate the road accidents on the road network.

1. Plot the distribution of road accidents on your road network and visualise this. Aggregate across multiple years of accident data. You do not need to consider or represent when the accidents happened, only their location; but you are welcome to add information about time if you believe there is something interesting to show.
2. Investigate whether a high number of accidents on one road correlates with a high number on connecting roads.
3. Investigate whether accidents happen nearer to intersections or partway along roads. Consider this as asking at what fraction of the road length away from the nearest intersection do accidents typically occur.

For the above, we suggest using the spaghetti library as shown in your lessons. The third question will require investigating the API of the library to find the relevant functions to answer the question.

Task C (Voronoi diagrams):

Despite its accidents, the city of Leeds is ideal for organising marathons. The city major would like to organise a day of parallel, simultaneous marathons in different parts of the city. The major would like to maximise the participation of citizens by organising these marathons in diverse locations of the city, dividing the city into various areas (or "cells") so that every person can join a marathon that is close to their home. Within each of these cells, a path of exactly 42 Km is needed. Assuming that the mayor would like to organise N=10 simultaneous marathons:

1. Select the initial set of 10 cell seed points. For this, you can use several criteria, such as being far away from frequent accident roads, being close to public transport, being evenly spread, etc. (explain your choice in the report).
2. Visualise the cells yield by your selection of seed points in a Voronoi diagram
3. Find 2 or 3 cells for which you can find at least one path (or more, if possible) that is (a) exactly 42 Km long, and (b) finishes at the same point where it starts. Visualise both the cells and the found paths.
4. Try to extend the previous step to all cells. Can you find at least one such a path for every cell?
5. If for steps 3-4 there were cells with no such path, what different options could you consider to increase the number of cells that include such paths? (Hint: think about the number and location of seed points; the size of the area under consideration; etc.) Choose one of such options, repeat steps 3-4, and report the results you obtain, explaining your reasoning in the report.

Task D (TransE, PROV, PageRank):

The mayor's office is also interested in finding an efficient way of representing the provenance of important events in the road network of Leeds, such as accidents or marathons, and how this could be used for insights.

1. How would you represent such an event using the W3C PROV provenance data model standard? What would correspond to Agents, Entities, and Activities? Provide a diagram with an example.
2. If you used this model to generate a large provenance network, and you computed the PageRank algorithm for all the nodes of this network, then what could the output values tell you about the events in the city?
3. If you computed TransE embeddings for the previous large provenance network, what could these embeddings be used for from a practical point of view? What kind of problems could they help address?

Report*:*

As in Coursework 1, Tasks A, B and C require you document your reasoning, explained snippets of the code used in the analysis, and any informative visualisations that result. Task D will be text paragraphs with one or more references to sources used – it is not expected to involve any coding.

Different to Coursework 1, you should put the full code you have written in the appendix of your report. Additionally, you may include a GitHub link pointing at a repository with your code. It does not need to have accompanying explanation, as any explanation of the most important code snippets should be in the main report (but including useful comments in your code is good practice).

Remember to include your name and student number on the report!

Assessment. It is good to apply ideas and methods from any of the modules on your programme (Introduction to Urban Analytics, Telling Stories with Data, Spatial Data Analysis etc.) but the focus should be on networks and network data analysis.

The following criteria will be used to determine the mark for each submission:

· Demonstrated understanding of network data analysis concepts and how they can apply to the questions in the coursework tasks.

· Technical ability in using programming to tackle a data analytics problem, showing ability to research and apply data manipulation techniques as required for the problem.

· Creative thinking about the problems described in the coursework and specifically the network-related aspects.

· Clarity of explanation of what code does and why and what results mean, plus good use of visualisations and presentation.

· Succinctness of reporting, i.e. conveying a lot of substance clearly in a short amount of space. Note that marks will be deducted for exceeding the page limit.

Submission instructions.

Deadline. You have approximately 3.5 weeks between release of the coursework and submission. The strict deadline is **4pm (16:00) UK time on Tuesday 11 April**.

Size limits. The report should be at most 6 pages in length including references and figures. The appendix should list the full code you have written and this can be any length. The amount of space used per task may vary depending on how much you find to say on each, but expect each task to fill between 1 and 3 pages. Use font size 11 of larger for readability. Otherwise, there is no restriction on the format.

Submission format. The report should be submitted as a PDF document. You can submit the appendix within the same or a different PDF. Include your source code files in a separate compressed file, and ensure you compress all files (one or two PDFs for the report and appendix, and the source code) into a single file (ZIP, GZIP, etc.), and submit that single compressed file on KEATS.

Plagiarism, collusion and technical support. You are not allowed to submit anyone else's work as your own (plagiarism) or conduct the coursework in collaboration with other students (collusion). Both are serious matters of misconduct.

However, you are allowed and even encouraged to discuss specific technical problems you face with the coursework and how to solve them with the rest of the class via the KEATS discussion forum. For full guidance on what is acceptable to ask the class and what must be done individually see the 'Coursework questions and collusion' page in the Assessment section of the KEATS page and feel free to ask clarifying questions in the tutorial or discussion forum.

Last modified: Monday, 13 March 2023, 5:46 PM

◀ CW1 submission link (LATE SUBMISSIONS ONLY)
Jump to...
Reading club 1 instructions ▶