Data Science for Design: Assignments 2-4

Overview

- Assignment 2: Weighting: 30%, LOs: *Program*, *Data*, Due: 15th Nov
- Assignment 3: Weighting: 30%, LOs: *Communicate*, Due: 5th Dec
- Assignment 4: Weighting: 10%, LOs: *Professional*, Due: 5th Dec

Assignment 2

Weighting: 30%, LOs: *Program*, *Data*, Due: 15th Nov

Brief: carry out a structured analysis of a portion of the data provided by your data holder. This should be structured as follows:

- **Describe**: tell us about the data that you have. We expect to see:
    - what is the general type of the data (tabular, network, geographical, textual etc.),
    - how large and complex is it (rows/columns, size, variation, structure)
    - What fields and data types are present (max/min, levels for categorical values).
    - Links between this data and other data (e.g. foreign keys, unique ids)
    - Summary statistics about the data - how many people, what time frame, field averages etc.
    - How does the data relate to the questions that the data owner has discussed with you?
- **Explore**: carry out a deeper exploration of the data. This includes looking at individual fields/variables to see the distribution of values they take (e.g. evenly distributed, bell curves, bi-modal) or how they are distributed in time. It also includes relationships between variables in your dataset: are there correlations? In which direction? Complex curves? We would expect to see roughly:
    - 4-5 exploratory visualisations, presented in a readable form, with an explanation about what you have found
    - 1-2 relationships between variables analysed
    - Ideas about trends, outliers, clusters
    - Reference to statistics, i.e. a sense of which relationships are significant, and what claims you can back up.
- **Reflect and Hypothesise:** What do you think might be behind the relationships and distributions that you have found? How does your data relate to the world? In short, what hypotheses came up as you explored the data, and how would you go about testing them? We would expect to see:
    - A 200 word reflection on the data
    - 3-5 hypotheses with an explanation of:
        - What your hypothesis is
        - why you think that (referring to your graphs)

        - How you could go about testing it (note that this may include: further studies, asking the data owner, using some advanced algorithms, checking each data point individually, etc..)

    NOTE: You don't have to actually test the hypotheses - you just have to demonstrate that you have plausibly thought about how you would test them.

Submission: this should be submitted as in individual Jupyter notebook, with all of the output present, which we can run. Some notes:

- If you have installed nonstandard packages, document what they are and how to install them.
- You only need to submit your notebook through Learn - we will organise the data separately as some of you have large/private databases.
- Make sure you have saved your notebook *with the output*, so we can immediately see your graphs etc.
- Make sure you document which files are used at which point in the analysis.
- Plagiarism is treated extremely seriously! You may be working on the same data, but we would expect to see different reports. In particular, the explorations should not be the same between members of the group. Discussion is absolutely fine and encouraged - wholesale copying is not.

Assignment 3

Weighting: 30%, LOs: *Communicate*, Due: 5th Dec, Presentation: 30th Nov

Brief: Communicate about the context and implications of your data. This is an open brief, graded against a single learning outcome: Communicate. You can take whatever approach is appropriate and use whatever techniques are available to your group. However, credit is given based on how well your design communicates to an audience, not on technical prowess. You will have a chance to present this in Week 11 (30th Nov) shortly before handing in. Any medium is permissible, subject to space/time/technical/resource constraints - interactives, websets, data comics, infographics, AR/VR, physicalisations, dance/theatre work etc.

Submission: This is a Group submission - only one person from each group needs to submit. You should submit:

- Slides from your presentation

EITHER:

- A 3000 word PDF report (if you are presenting scientific knowledge to a scientific audience). This should contain appropriate figures and statistics to support communication to a critical, data literate audience.

OR:

- A 1000 word PDF report and some form of documentation of your creation. The creation could be a 2 minute video, or a link to a live interactive. The report should discuss the context and findings, and is used to back up all of the information not present in the interactive piece. Links to the video are fine if it cannot be uploaded to Learn.

Assignment 4

Weighting: 10%, LOs: *Professional*, Due: 5th Dec
Brief: demonstrate competence in the software systems that are in general use, and an ability to think critically about collaboration processes. This is a highly structured assignment, however it should be largely fulfilled by your existing practise. We expect to see:

- Engagement with GitHub (5% overall), with:
    - > 10 commits, with good comments. Aim for lots of small, incremental commits over a period of time.
    - at least one issue raised, properly tagged
    - at least one issue closed
    - at least one useful comment on an issue
- A reflection on the collaborative process (5% overall), using a template that will be provided, covering roughly:
    - What did you need to know about the data before you started?
    - How much did you communicate with the data holder? Was it enough?
    - What did you need that you did not get?
    - What worked really well?

Submission: a text document covering the reflection, and a link to your Git profile where we can see the collaboration activities that you have engaged in.