

COMP20008 2021 SM2 Workshop Week 6 Experimental design

The Aurin repository contains a large number of datasets that data scientists can use to help answer important questions in society. The following exercise illustrates a possible simple scenario and is designed to get you thinking about how you might use the Aurin urban data repository that is hosted by the University.

Suppose our question is ***Are we building enough green spaces in Victoria to ensure a healthy population?***

- **Question 1:** Who would be interested in an answer to this question and why?
- Log in to the Aurin portal <https://aurin.org.au>
- Select Victoria as your region of interest.
- Browse through the available datasets and see what data is available.
- Add the dataset “2015 Local Government Area (LGA) Statistical Profiles”. You should select all the attributes to include. This dataset includes information about number of people reporting high blood pressure across different regions in the State. We will use this as a measure of people’s health.
- Add the dataset “ LGA Visit to green space (once per week)”. You should select all the attributes to include. This dataset contains information about number of people who visit local green space each week, across different regions in the State.
- Download each of these datasets as a CSV file.
- **Question 2:** What feature would you use to join these datasets together?
- **Question 3:** Describe two different techniques you could use to help identify a relationship between the visits to green space and reports of high blood pressure.
- **Question 4:** Describe how you could use the data to make a prediction about people’s overall health based on the information available. Describe the steps you might use to evaluate your prediction.
- **Question 5:** What challenges do you think might arise in studying these questions?
- **Question 6:** What are the next steps you might take when studying these questions?