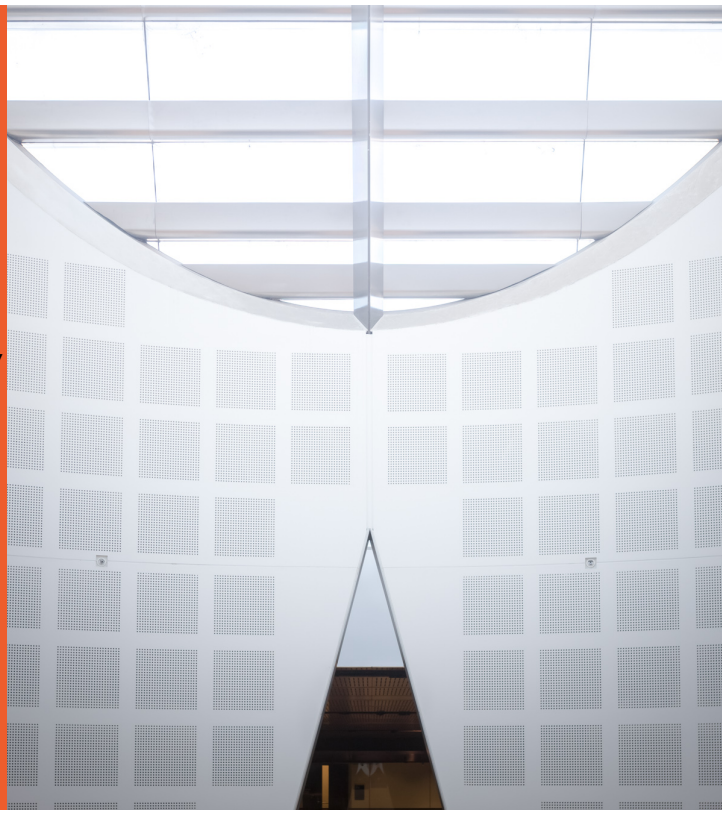
DATA2001 – Data Science, Big Data, and Data Diversity
Week 10: Assignment Overview

Presented by
A/Prof Uwe Roehm
School of Computer Science

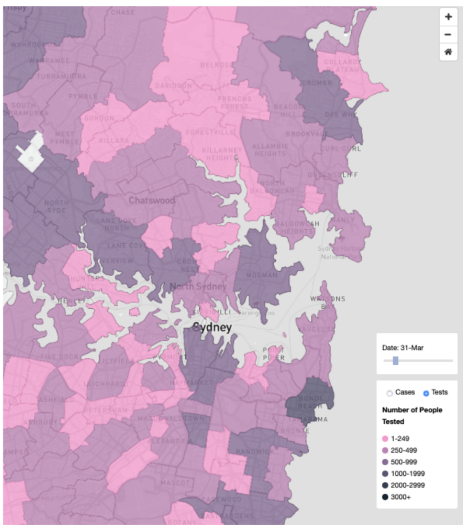
THE UNIVERSITY OF SYDNEY



Practical Assignment: Viral Vulnerability Analysis

- Assignment to be published this week (Canvas: Modules –> Assignment)
  - **Worth 20% pf the final grade in DATA2001/DATA2901**
  - **Due in tutorial of Week 12**

- Main idea:
  - Calculate a 'risk' or 'vulnerability score' per suburb in Greater Sydney wrt. viruses
    - Based on ABS data about population and availability of health services
  - Visualise and correlate with Covid-19 data



Practical Assignment: Viral Vulnerabiliy Analysis

- Goal: Practical experience with data variety, data analysis, and presentation
  - Technologies as covered in this course: Python, Jupyter notebooks, and SQL
- Three tasks:
  - Data import and integration
    - We provided census data, COVID-19 stats, and data about health-services
    - Needs to be combined, eg. via spatial join
    - Feel free to extend with own datasets
    - **Milestone 1: Integration of provided datasets to be ready in Week 11 tutes**
  - Vulnerability Analysis
    - Computation of 'vulnerability' score; example formula given
    - When adding other datasets, feel free to adjust formula
    - Correlation of your score with COVID-19 tests and confirmed cases
  - Documentation and (brief) Report
  - Additional tasks/options on web scraping and ML for teams in advanced stream

Provided Datasets (cf. Canvas)

- ABS Data
  - Census data on *neighbourhoods* (SA2-level areas) in Greater Sydney
    such as population, land area, number of dwellings, gender and age distribution
- Health Services in NSW
  - Location and some states (e.g. average available hospital beds)
- COVID-19 Data
  - Conducted tests and confirmed cases by postcode to check for correlation with
- Note that SA2-level data from the ABS does not always match suburbs,
  and that the health services have a GPS location, but not the neighbourhoods
  - cf. tutorial this week on how to retrieve boundary data for neighbourhoods too
- Adding more datasets from your side is explicitly encouraged.
  - Try different types and forms, not just CSV…

Assignment Rules

- Groupwork
  - teams of 2 (unless odd-size class or other good reasons)
  - All team members should be in the same tutorial class
- Deliverables see handout, page 3
- Due on Friday of Week 12
  - Submission page and marking rubric will be published in Canvas
  - Only one member per team needs to submit for the whole group;
    she should submit both a ZIP archive under "Viral Vulnerability Analysis Assignment"
    and also the PDF of your report in the separate "TurnItIn Dropbox – Viral Vulnerability Analysis"
  - Late submissions: -20% of achieved mark per day late
- Demo in Week 12
  - There will be a short demo via Zoom to tutors
  - Individual grades can be scaled based on participation in project or demo