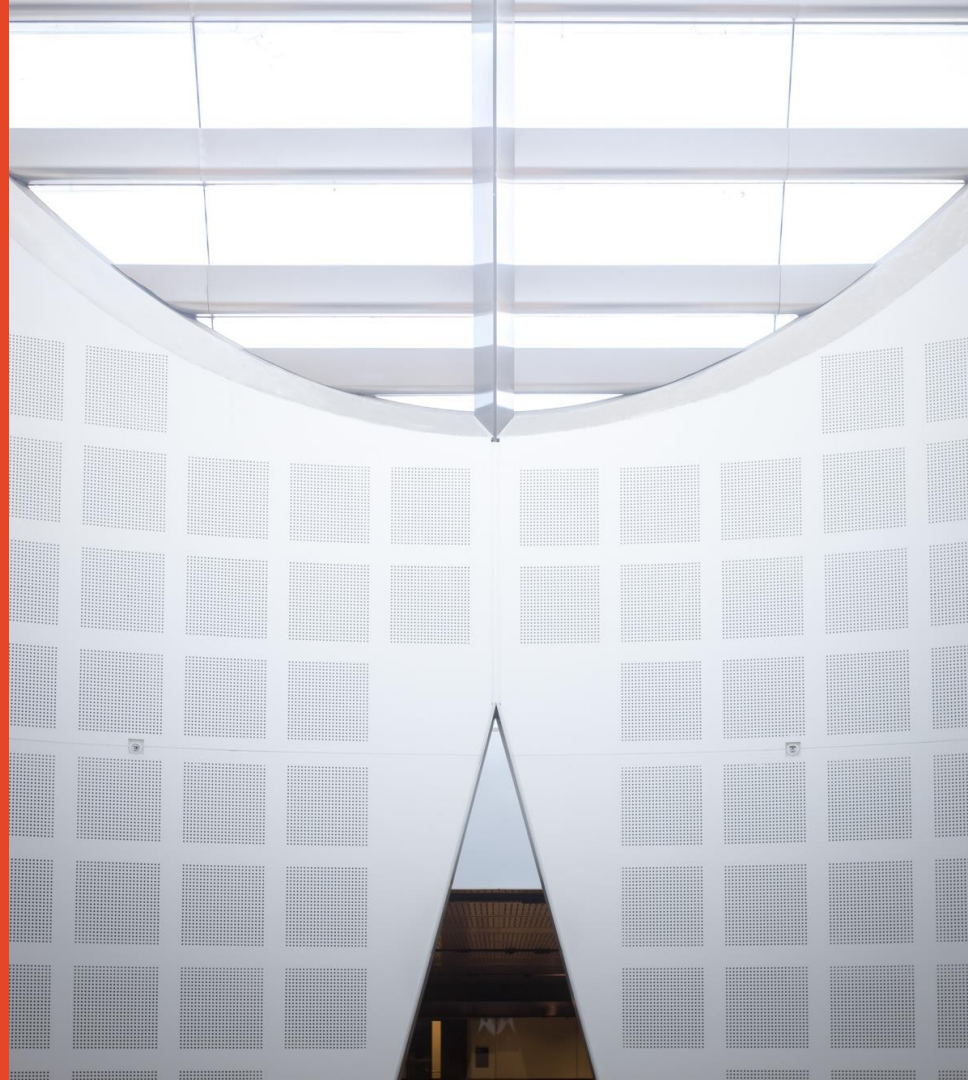


OCMP5310: Principles of Data Science

Week 3 Live Session

Presented by
Daniela Rivas



REVIEW OF LAST WEEK

Week 2

- Creating/querying databases with SQL.
- Analysing/summarising data from multiple Tables (joining) with SQL.

QUESTIONS?

WEEK 3: SESSION ACTIVITY

Week 3

- Hypothesis Testing.
- Model Evaluation.

Hypothesis Testing and Evaluation

Activity

- In Canvas, go to:
 - Exercise: Hypothesis Testing - Comparing Layouts.
- Download Jupyter Notebook:
 - `hypothesis_testing_and_evaluation.ipynb`

Association Rules

Activity

- In Canvas, go to:
 - Exercise: Association Rule Mining.
- Download Jupyter Notebook:
 - association rules.ipynb
- Download CSV file:
 - Groceries.csv

Project Stage 1

Project Stage 1 (10% overall mark)

- **Due:** Week 3, 7 May 2023, 23:59, Sydney Time.
- **Tasks:**
 - Selecting a dataset.
 - Defining the problem and project requirements.
 - Acquiring and loading the dataset into either a database or a Jupyter notebook.
 - Data cleaning.
- **Submissions:**
 - 2-page report.
 - Code.
- More details [here](#).

Report

- 2-page report (not counting title page and references or appendix) that describes the problem, proposed approach and dataset, and data cleaning process.
 - **Problem:** Describe the problem from a general perspective, highlighting the business/research need. List the research question(s) you will answer in Stages 2 and 3 of the project.
 - **Approach:** Describe the approach you will take to solving the problem and any requirements. This is your plan for Stage 3.
 - **Data:** Describe the data from a general perspective e.g. source, size, fields of interest. How did you acquire the data? Describe any data preparation steps e.g. transformation, sampling, cleaning.

Code

- Your code used to ingest and clean your dataset.
- Format: Jupyter Notebook (.ipynb), python script (.py) or similar.

QUESTIONS?