

Assignment 2: Data Science Project

June 1, 2023

Due Date

Sunday 13th August, 11:59PM

Special consideration for extensions

Subject to approval by the coordinator, submit the form from the participant guide on Canvas to the coordinator together with evidence.

Late submissions

Late submissions will only be accepted for up to 5 days after the due date with 10% late penalty for each day missed. Therefore, submitting 5 days after the due date will incur 50% late penalty.

Hypothetical Scenario and Objective

It has often been observed that energy consumption tends to be at its highest on days with hotter temperatures. As a team of data scientists, you have been tasked to ensure data is ready for analysis by performing data cleaning. After that, your goal is to develop a model that predicts the maximum daily energy use and pricing based on weather data. The hope is that this model can be used to aid energy companies understand and plan for future usage.

The following two datasets are provided:

- **weather.csv** contains key weather indicators, such as minimum and maximum temperatures for the city of Melbourne for each day between November 2022 and April 2023. This data has been extracted from the Bureau of Meteorology and collated into a single file for your convenience.
- **price_and_demand.csv** contains energy price and demand figures for the state of Victoria at a half an hour interval between November 2022 and April 2023. This data has been extracted from the Australian Energy Market Operator, and collated into a single file for your convenience.

Working in a team of 3-4 students, you should:

- Perform data cleaning on the provided data.

- Develop a model which predicts the maximum daily energy usage based on the provided data. This could either be a classification or a regression model.
- Evaluate the effectiveness of your model.
- Write a report about your exploratory analysis supported with tables and figures.
- Present a short research analysis of current trends which could be linked to your results and limitations.

It is your responsibility to ensure that all team members provide equal contributions to your project.

Assessment

Your Report

Your report should be written in a Word document (.pdf or .docx). It must display a word count and should be no more than 2500 words in length. The word count limit excludes any figures, tables, and appendices. Anything written over the word limit will not be assessed. Your report should include the following information:

1. What data cleaning methods have you applied? Why have you chosen these methods over other alternatives? Give examples to support your chosen methods.
2. Explain the process of building your model. How did you select the features included in the model? How does your model work?
3. How effective is your model? How have you evaluated this?
4. What insights about weather and daily energy usage can you draw from your analysis? Discuss any significant results.
5. What are the limitations of your results? Present a short research analysis of current trends which are not necessarily about weather and energy, but these trends could potentially be associated with your results and limitations.

Your report should also include the following:

1. Make effective use of visualisations to support your arguments.
2. A section for references to list any resources you have used, especially for the research component of this project.

Git repository

All of the code you develop as part of this project should be stored in a GitHub repository. Ensure that this is set as a private repository since this is an assessment.

Only one member of your group should create a GitHub repository, the other group members should be added to the same GitHub repository. This will ensure that all group members are able to collaborate on the same codebase.

README file

A README file must be included in your git repository with sufficient information on how to run the code to get the results, including the `random_state` values if required. You must ensure that this README file contains the full names of each member of your group. Important note: To get full marks, assessors must be able to replicate the models based on your explanations, provided code, and instructions provided in this README file.

Submission Instructions

Your final report must be uploaded via Canvas by the due date. All of your code files, and any other supporting files used, should be placed in a .zip archive and uploaded via Canvas by the due date. It is essential that any numerical results or visualisations used in the final report can be reproduced by running your code. You must also include a link to your GitHub repository.

Your report, code files and any other supporting documentation must also be pushed to your git repository.

Academic Honesty

You are expected to follow the academic honesty guidelines on the University website <https://academichonesty.unimelb.edu.au>

| Table 1: Marking Rubric | | |
|------------------------------------|---|--------------|
| Requirement | Description | Maximum Mark |
| Data cleaning and visualisation | What data cleaning methods have you applied? Why have you chosen these methods over other alternatives? Give examples to support your chosen methods. | 12 |
| Building a Model | Explain the process of building your model. How does your model work? Important note: To get full marks on this criteria, assessors must be able to replicate the models based on your explanations and provided code. | 6 |
| Model effectiveness and evaluation | How effective is your model? How have you evaluated this? Important note: To get full marks on this criteria, assessors must be able to replicate any calculations, visualisations etc. which you used to evaluate your models using the code you provided. | 6 |
| Data Analysis | What insights about weather and daily energy usage can you draw from your analysis? Discuss any significant results. | 6 |
| Limitations and Research Component | What are the limitations of your results? Present a short research and discussion on current trends which are not necessarily about weather and energy, but these trends could potentially be associated with your results and limitations. | 10 |
| Overall Maximum Mark | | 40 |