

Assignment 2: Data Science Project

March 16, 2023

Due Date

Sunday 16th April, 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 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_data.csv` contains key weather indicators, such as minimum and maximum temperatures for the city of Melbourne for each day between January and August 2021. This data has been extracted from the Bureau of Meteorology and collated into a single file for your convenience and modified slightly for the purposes of this assignment.
- `price_demand_data.csv` contains energy price and demand figures for the state of Victoria for each half hour period between January and August 2021. This data has been extracted from the Australian Energy Market Operator, and modified slightly for the purposes of this assignment.

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.
- Do some research to perform a short analysis on current weather and energy price data in Melbourne and how this could be used to improve this project.

You will also need to evaluate the effectiveness of the model. It is your responsibility to ensure that all team members provide equal contributions to your project.

Assessment

Your Report

Your report should be no more than 2500 words in length excluding figures and tables. 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? Important note: To get full marks on this criteria, assessors must be able to replicate the models based on your explanations, provided code, and instructions in your README file.
3. 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 and instructions in your README file.
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 analysis on current weather and energy price data in Melbourne which could be used to improve this project in the future. For example, you can assess the availability, completeness, and/or up-to-dateness of the data (i.e., data currency). These are only suggestions and should not limit the discussion points in your research.

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. A link to create your git repository has been provided on Canvas.

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 names and student login IDs of each member of your group.

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 analysis on current weather and energy price data in Melbourne which could be used to improve this project in the future.	10
Overall Maximum Mark		40