Assignment 2: COMP20008 Data Science Project

November 14, 2022

Due Date

Sunday 11th December, 11:59PM

Objectives

The objectives of this assignment are

- To gain experience in practical data wrangling with real-world datasets.
- To practice using a variety of data analysis algorithms and determining suitable approaches to a real-world problem.
- To build e-portfolio as a data scientist in training.
- To gain experience in written communication of analysis and results on a topic related to data science.

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 data scientist, you have been tasked with developing models that predicts the maximum daily energy use and pricing based on weather data. The hope is that these models can be used to predict likely energy demands based on a weather forecast, which can help energy companies understand plan for future usage, and help businesses plan when to conduct energy-intensive operations.

The following two datasets are provided:

- `weather_data.csv` contains key weather indicators, such as minimum and maximum temperatres for the city of Melbourne for each day between January and August 2021. This data has been extracted as-is from the Bureau of meteorology and collated into a single file for your convenience.
- `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 build two models:

- A model which predicts the total daily energy usage based on the provided weather data.
- A model which predicts the maximum daily price category based on the provided weather data.

You will also need to evaluate the effectivness of the model.

Assessment

Your Report

Your report should be no more than 2000 words in length excluding figures and tables. Your report should include the following information:

1. What wrangling and aggregation methods have you applied? Why have you chosen these methods over other alternatives?
2. How have you gone about building your models and how do your models work?
3. How effective are your models? How have you evaluated this?
4. What insights can you draw from your analysis? For example, which input variables are most valuable for predicting energy usage/price?
5. Why are your results significant and valuable?
6. What are the limitations of your results and how can the project be improved for future?

Your report should make effective use of visualisations to support your argument.

Git repository

All of the code you develop as part of this project should be stored in a GitHub repository

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.

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 in your report.

Your report, code files and any other supporting documentation must also be pushed to your git repository. You must ensure that the README file within your git repository contains the names of each member of your group.

Academic Honesty

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

A project discussion forum has also been created on the subject LMS. Please use this in the first instance if you have questions, since it will allow discussion and responses to be seen by everyone.