This project was part of my quest to apply what I had learnt about the R statistical programming language in a real-world context. The analysis explored the relationship between Victorian school electricity consumption, and factors like school type (primary, secondary etc), total floor area, building age, enrolment numbers, geography, climate zone, and solar photovoltaic (PV) system use. Regression modelling (code to be added soon) was also conducted to predict annual consumption based on these factors. However, special attention was given to comparing schools using solar PV systems with those that did not, with the aim of providing performance baselines for a state-government energy efficiency program.

What is presented here is not intended to be a polished report on the research I conducted. Rather, it is meant to show how I integrated R coding with my thought processes in analysing the data. My data analysis approach was based on a five-step framework of ‘import-review-clean-explore-model’. However, real-world data is messy, and often you are not able to fully define the scope of a project (especially if this is the first one of its kind). For example, in this project the ‘clean’ and ‘explore’ stages were part of an iterative process, as initial exploratory data analysis revealed outliers demanding deeper investigation. Where possible, I updated any incorrect observations with the correct data, but others had to be removed altogether as they could not be remedied. Other outliers could not be proven as spurious, so these remained untouched. Therefore, the relation between the ‘clean’ and ‘explore’ stages was cyclical rather than sequential.

I have made every effort has been made to preserve the commercial-in-confidence status of the consumption and cost data and the anonymity of schools. As a result, any elements of code, plots, or other output that risk breaching this confidentiality have been censored. Nonetheless, I have supplemented the code and plots with comments to share a general overview of the insights I uncovered over this learning journey.