THE CURIOSITY CUP 2024

A Global SAS® Student Competition

Electric Consumption Patterns at Bundoora Campus – La Trobe University in 2018-2021

Hao Nguyen, Chan Tang, and Nhan Pham, La Trobe University

The Penguins Team

ABSTRACT

This paper aims to analyze the average electric consumption patterns at La Trobe University's Bundoora campus before and during the COVID-19 pandemic (2018-2021). The paper utilizes visualization techniques to explore significant patterns in average electric consumption. Additionally, models are developed to illustrate the factors that have the most significant impact on average electric consumption, all employing SAS Viya for Learners for analysis. Through this comprehensive analysis, insights into electric consumption trends and their determinants are provided, which can inform future energy management strategies at La Trobe University and environmental sustainability.

INTRODUCTION

Environmental issues have become a paramount concern for numerous organizations, the imperative to address climate change and mitigate environmental degradation intensifies. Previous studies have examined the nexus between energy consumption and environmental issues (Raza et al., 2019). Many organizations are actively striving to contribute to achieving significant milestones aimed at reducing the negative effects on the environment. La Trobe University has taken proactive measures through the establishment of the La Trobe Energy Analytics Platform (LEAP). This advanced technology platform serves as a comprehensive tool for energy analysis, actively monitoring consumption patterns and performance in up to 50 buildings across the university's campuses. By implementing such innovative solutions, La Trobe University is not only contributing to broader environmental goals but also positioning itself to achieve its Net Zero Emissions target by 2029 (La Trobe University, 2024), showcasing a concrete and impactful commitment to sustainable practices within the higher education sector.

DATA AND METHODOLOGY

The study utilizes the open-source dataset provided by La Trobe University (UNICON), aligning with the institution's commitment to achieving Net Zero Carbon Emissions by 2029. This dataset encompasses data on utilities, electricity, gas, and water consumption starting from January 1st, 2018, and is accessible in CSV format via the following link (https://github.com/CDAC-lab/UNICON).

Pre-processing steps involve utilizing Python and Excel due to constraints related to file size when using SAS Viya for Learners. The focus centers on retrieving data related to Bundoora campus. Subsequently, calculations are performed to derive key metrics, such as the average and total electricity consumption categorized by date.

The study's emphasis extends to four distinct building categories by their room sizes: library, office, teaching, and mixed use. It is based on the recognition of a correlation between campus average energy consumption and building size (Malakoutian et al., 2021). The data imported into SAS Viya for Learners is simplified and comprises 20 variables, with certain variables showing imbalanced distributions. Notably, "is_holiday" records a value of zero (0) in 94.9% of cases, "is_semester" in 80.9%, and "is_exam" in 92%. These zero values signify periods such as lockdowns and transitions to online teaching and studying during the COVID-19 pandemic.

EXPLORATORY ANALYSIS

SIGNIFICANT ELECTRIC CONSUMPTION PATTERNS

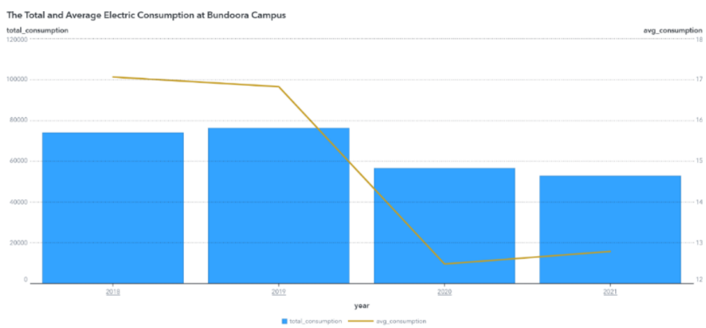

Figure 1. The total and average energy consumption

In 2020, electricity consumption significantly dropped at Bundoora campus due to COVID-19 lockdowns, with factors like the closure of non-essential buildings, including La Trobe Sport facilities, playing a key role. The data also reflects changes in human behaviour, such as La Trobe staff adopting work-from-home and students transitioning to online learning (OAKMAN et al., 2022). We hypothesized a strong correlation between temperature and electric consumption (Zhang, 2019). However, upon conducting a correlation analysis, we obtained a correlation coefficient (r) of 0.18, indicating a weak correlation between these variables. Additionally, a correlation matrix has been visualized to illustrate the relationships between temperature, electric consumption, and other relevant variables.

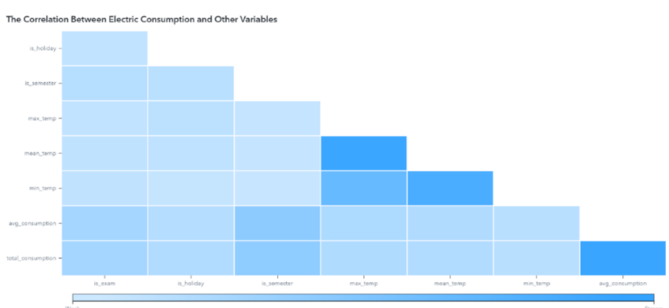

Figure 2. The correlation between electric consumption and other variables

In the years leading up to the COVID-19 pandemic, particularly in 2018-2019, a distinct recurring pattern emerges in electric consumption. Notably, spikes in usage occur in March, May, August, and October. Analysis indicates that specific university events and academic periods play a significant role in these fluctuations (Sharma & Kumar, 2017). Orientation events (March, August): Elevated electric consumption during March and August is likely linked to orientation events. These activities, involving presentations and gatherings, necessitate increased electricity usage for lighting, audio-visual equipment, and other purposes (Gui et al., 2020). Exam preparation period (May, October): The heightened consumption in May and October is associated with the exam preparation period. During this time, increased student presence in teaching spaces and libraries, along with greater use of electronic devices and lighting, contributes to a surge in overall electricity usage.

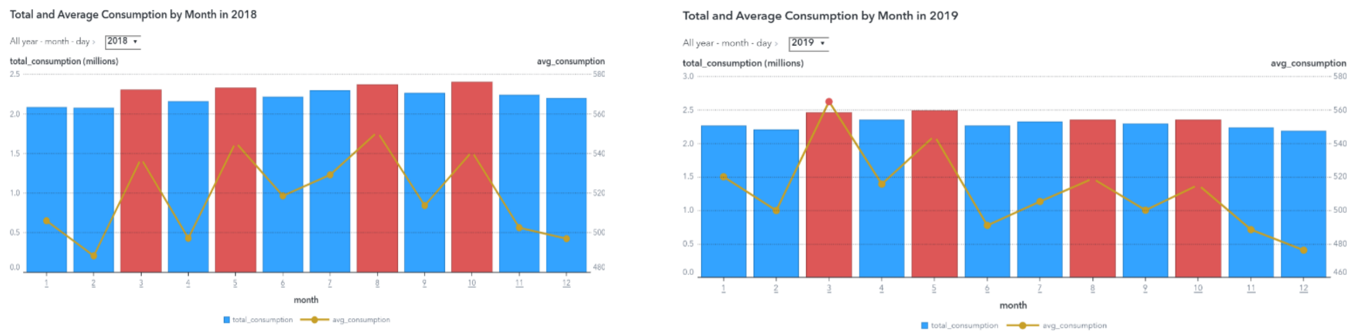

Figure 3. Total and average electric consumption by month in 2018 and 2019

The 2020 COVID-19 pandemic resulted in a significant decrease in average electric consumption in library, office, and mixed-use buildings. Conversely, an uptick in electric consumption was noted in teaching buildings during this period. This increase may be attributed to the shift to online teaching, video conferencing, and the use of digital learning tools. Seasonal changes, particularly heightened heating requirements in colder months, could also be a factor contributing to the observed variations in energy consumption (Li et al., 2020).

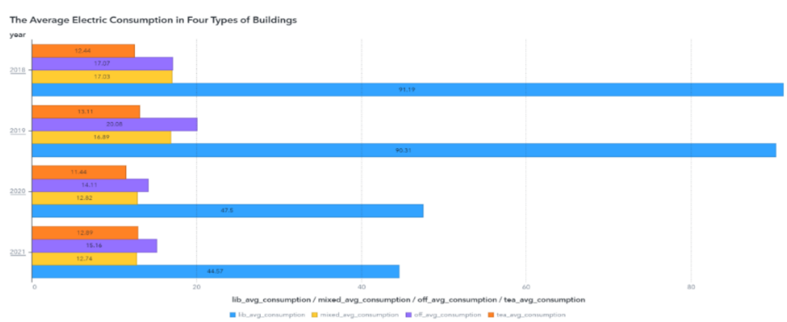

Figure 4. The average electric consumption in four types of buildings

MODELS

We used SAS Viya for Learners to create various models for identifying factors affecting average electricity consumption at the Bundoora campus. The seven models include Stepwise Linear Regression, Neural Network, Forward Linear Regression, Gradient Boosting, Decision Tree, Ensemble (post-processing), and Random Forest, targeting the variable "avg_consumption.". Refer to Figure 5 for the modeling pipeline.

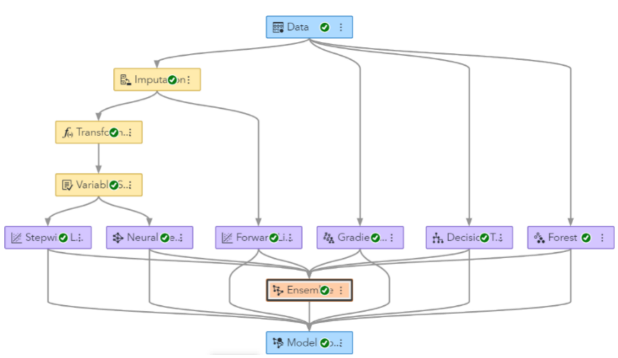

Figure 5. Modelling pipeline

Data needs to go under Imputation, Transformation, and Variable Selection nodes before fitting into training models. For the Imputation node, missing values are replaced by mean interval inputs and count for class inputs with cutoff percentage of 50%. Transformation node is used to replace skewed values by the square root interval inputs. Next, the Variable Selection node is connected to the Transformation node to determine eligible variables based on correlation statistics (cutoff variance beyond 1.00) and split the dataset into 30% validation, 10% test and 60% training. In Figure 6 (Appendix), only four factors significantly impact average consumption. The diagram reveals a single hidden layer with a total of 71 nodes, including 20 input nodes, 1 output node, and 50 hidden nodes. Notably, the most influential nodes are closely linked to SQRT_mixed_avg_consumption and SQRT_lib_avg_consumption, indicating that energy consumption is primarily influenced by mixed-use buildings and the library. Interestingly, among the four years studied (2018, 2019, 2020, and 2021), only 2020 appears as a significant feature. There are two Linear Regression models operated with different methods, one is Stepwise Regression, and the other is Forward Regression. These two models have soaring R-square values (beyond 98%), suggesting very good-to-fit models yet potentially raise a problem involving overfitting. The Decision Tree and Random Forest models are also included in the provided pipeline. Their ASEs, which are 0.12 and 0.21 respectively. Both models perform comparably in terms of error, with no clear superiority of one over the other.

According to Figure 7 (Appendix), the Gradient Boosting model yields an ASE of 0.0839 and an RMSE of 0.2897 for the test partition. Additionally, the five most crucial factors identified by the model are mixed_avg_consumption, lib_avg_consumption, off_avg_consumption, month, and tea_avg_consumption.

RESULTS

According to Figure 8, the worst performance accounts for the Neural Network model (ASE = 1.01) while the best performance goes to the Gradient Boosting model (ASE = 0.0839). The lower the ASE, the better the model, as it indicates how closely a regression line aligns with a set of data points. This suggests that the Gradient Boosting model demonstrates lower errors.

Model Comparison

Champion	Name	Algorithm Name	Average Squared Error	Root Average Squared Error
	Gradient Boosting	Gradient Boosting	0.0839	0.2897
	Forward Linear Regression	Linear Regression	0.1011	0.3180
	Ensemble	Ensemble	0.1026	0.3203
	Stepwise Linear Regression	Linear Regression	0.1044	0.3231
	Forest	Forest	0.1200	0.3465
	Decision Tree	Decision Tree	0.2153	0.4640
	Neural Network	Neural Network	1.0100	1.0050

Figure 8. Model comparison evaluation metrics

Factors including mixed_avg_consumption, lib_avg_consumption, off_avg_consumption, month, and tea_avg_consumption have the most significant influence on the model's average electric consumption, suggesting that these factors play crucial roles in determining the overall electricity usage at the Bundoora campus. It is understandable that when comparing 8 building categories in terms of gross floor area, we have identified the 4 buildings with the largest room sizes: office, mixed-use, teaching, and library. Additionally, there is a relationship between the gross floor area and electric consumption; generally, the larger the room size, the greater the electricity usage tends to be (Malakoutian et al., 2021).

LIMITATIONS

The limitations of our research stem from the absence of data post-COVID-19. As our study primarily focuses on the period preceding and during the pandemic, we lack information to accurately analyze electricity usage after the situation returns to normalcy. The dynamics of energy consumption may undergo significant changes as activities resume, and the university implements more energy efficiency projects.

FUTURE STUDY

In future studies, it is recommended to incorporate data and analysis from La Trobe University's energy efficiency projects, including the implementation of LED lighting and the installation of solar panels, which can greatly aid in estimating La Trobe's efficiency in reducing electric consumption and advancing towards its commitment to achieving Net Zero Emissions by 2029. Furthermore, although our current analysis has compared and identified the Gradient Boosting model as the most effective, it is acknowledged that there may be room for improvement. For instance, one potential avenue for improvement could involve refining the model by systematically removing variables to assess their impact on reducing the RMSE. This will allow for a more nuanced understanding of the model's performance and may lead to further optimizations in predictive accuracy.

CONCLUSION

In summary, this paper analyzes electric consumption patterns at La Trobe University's Bundoora campus before and during the COVID-19 pandemic (2018-2021). Utilizing visualization techniques and SAS Viya for Learners, it identifies significant trends and factors influencing consumption. The insights gained can inform future energy management strategies, promoting environmental sustainability. While the study has limitations, such as a lack of post-COVID-19 data, it suggests potential avenues for further research when teaching-learning activities return to normal. Overall, this research lays the groundwork for informed decision-making in energy management at La Trobe University.

REFERENCES

La Trobe University. (2024). Our journey so far. https://www.latrobe.edu.au/sustainability/net-zero/our-journey-so-far

Li, Y., Wang, W., Wang, Y., Xin, Y., He, T., & Zhao, G. (2020). A review of studies involving the effects of climate change on the energy consumption for building heating and cooling. International Journal of Environmental Research and Public Health, 18(1), 1–18. https://www.mdpi.com/1660-4601/18/1/40

Malakoutian, M. A., Malakoutian, Y., Mostafapoor, P., & Amjadi, M. (2021). Correlation Between Energy Consumption and Building Size. ENG Transactions, 2, 1-6.

Oakman, J., Kinsman, N., Graham, M., Stuckey, R., & Weale, V. (2022). Strategies to manage working from home during the pandemic: the employee experience. Industrial Health, 60(4), 319–333. https://www.jstage.jst.go.jp/article/indhealth/60/4/60_2022-0042/_article

Raza, S. A., Shah, N., & Sharif, A. (2019). Time frequency relationship between energy consumption, economic growth, and environmental degradation in the United States: Evidence from transportation sector. Energy, 173, 706-720.

Sharma, A., & Kumar, H. (2019). Seasonal variation on electricity consumption of academic building-a case study. Journal of Thermal Energy System, 2(2), 1-6.

Zhang, C., Liao, H., & Mi, Z. (2019). Climate impacts: temperature and electricity consumption. Natural Hazards (Dordrecht), 99(3), 1259–1275. https://doi.org/10.1007/s11069-019-03653-w

APPENDIX

Variable	Mean	SE	Median	SD	Skewness	Minimum	Maximum	Count
Year	2019.50	0.03	2019.00	1.12	0.00	2018.00	2021.00	1458
month	6.52	0.09	7.00	3.45	-0.01	1.00	12.00	1458
day	15.73	0.23	16.00	8.80	0.01	1.00	31.00	1458
date	43829.78	11.04	43829.50	421.46	0.00	43101.00	44561.00	1458
avg_consumption	14.79	0.07	14.44	2.76	0.06	9.48	20.38	1458
total_consumption	64849.72	343.27	63780.40	13107.16	-0.10	9429.37	87313.59	1458
mean_temp	14.93	0.13	14.33	4.90	0.55	4.98	32.33	1458
max_temp	20.86	0.17	19.70	6.61	0.78	7.80	44.70	1458
min_temp	9.74	0.11	9.65	4.34	0.12	-2.60	26.80	1458
is_holiday (dummy)	0.05	0.01	0.00	0.22	4.10	0.00	1.00	1458
is_semester (dummy)	0.19	0.01	0.00	0.39	1.58	0.00	1.00	1458
is_exam (dummy)	0.08	0.01	0.00	0.27	3.11	0.00	1.00	1458
lib_avg_consumption	68.43	0.68	79.10	25.95	-0.28	19.68	115.26	1458
lib_total_consumption	6407.55	65.50	7220.75	2501.18	-0.30	23.00	10131.00	1458
mixed_avg_consumption	14.87	0.08	14.05	3.07	0.58	8.42	40.84	1458
mixed_total_consumption	32413.50	195.85	30664.42	7478.46	0.01	4544.98	46974.57	1458
off_avg_consumption	16.61	0.10	16.14	3.81	0.24	-1.12	26.37	1458
off_total_consumption	4683.30	29.65	4576.42	1132.26	0.13	-140.70	7405.80	1458
tea_avg_consumption	12.47	0.05	12.61	1.72	-0.19	5.56	16.94	1458
tea_total_consumption	6416.67	28.52	6378.39	1089.18	0.05	1023.50	10029.21	1458

Table 1. Summary statistics

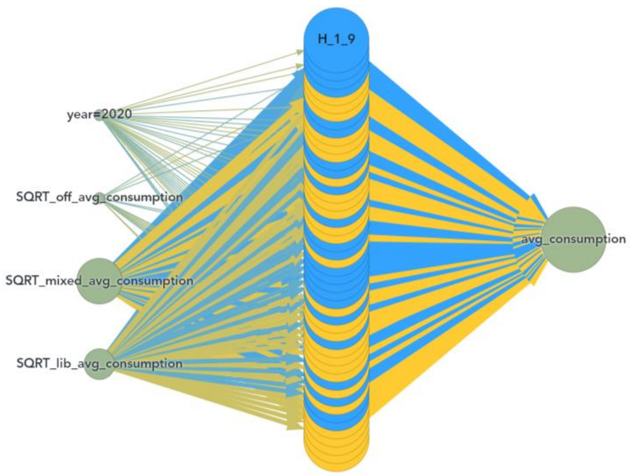

Figure 6. Neural Network diagram

Variable Importance			
Variable	Importance	Std Dev Importance	Relative Importance
mixed_avg_consumption	162.73	267.59	1.0000
lib_avg_consumption	3.9328	7.3727	0.0242
off_avg_consumption	3.7221	6.6060	0.0229
month	2.8303	2.5562	0.0174
tea_avg_consumption	1.0834	2.4618	0.0067
year	1.0644	2.7029	0.0065
mean_temp	0.4459	0.8929	0.0027
day	0.1826	0.1181	0.0011
is_semester	0.03030	0.4523	0.0002
is_holiday	0.01578	0.05810	0.0001
is_exam	0.01106	0.1102	0.0001

Figure 7. Gradient Boosting model evaluation metrics

CONTACT INFORMATION

Nhan Pham
La Trobe University
0435787900
21333993@students.latrobe.edu.au