

Predicting Victorian electricity demand in a pre and post-COVID landscape

Introduction

On the 16th of March 2020, Premier Daniel Andrews announced a state of emergency across Victoria, and with stay at home restrictions resulting in a high proportion of the population transitioning to a working (and learning) from home environment. A hidden cost affecting many is the changing demand of household and business electricity usage. The stay-at-home restrictions mentioned above have resulted in lower business and higher residential electricity usage, shifting the costs from employer to employee. By building models to predict electricity demand and comparing 2020 performance to 2019 performance, the true effect of COVID-19 on demand can be observed, and implications can be discussed. If models perform poorly on 2020 data, then it could be inferred that COVID has impacted electricity demand, and data could be examined to see if models underestimate (meaning more power was used due to COVID) or overestimate (meaning less power was used due to COVID) demand. If the model performs as expected, then demand would not have been impacted by the pandemic, meaning that despite a large proportion of the workforce working from home, electricity demand as a whole remained the same. This would confirm the hypothesis that businesses would be saving on electricity usage, with individual consumption covering the shortfall.

Data Sources

AEMO

The main set of data for the project was sourced from AEMO. A python script was created to loop data requests over all months for all years, and merge the results into a single file, in order to streamline the data obtaining process. The AEMO data reported both total demand (in MWh) and Recommended Retail Price (in AUD) in set half hour intervals, for each state.

Weather

As advised, electricity consumption's key drivers included weather. Historical weather data including daily min/max temperatures (in degrees Celsius), and daily global solar exposure (in MJ/square metre) was obtained using the bomrang library in R. Each of the three fields had to be queried individually, and results were merged in python. Weather data was obtained from the Melbourne (Olympic Park) weather station, as it was listed as the main weather station for Victoria weather related information by AEMO.

Seasonality

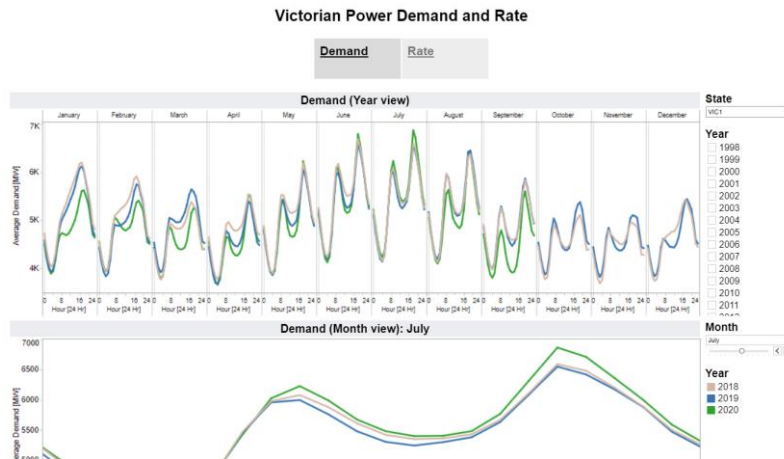
Another key driver of electricity consumption is business cycles – usage may differ on weekends, public holidays, and school holidays. A dataset of public holidays was obtained from data.vic.gov.au, and a dataset of school holidays from victoriaschoolholidays.com.au. The school holidays were manually inserted into a csv with year, start_date and end_date for all school holidays from 2014 onwards. These two datasets were then loaded into python, and a dim_date.csv dataset was created by feeding in every date from 01/01/2014 until 31/12/2020, and comparing it to both datasets to see if the date was a public holiday, a school holiday, or a regular day. Public holidays took precedence, so if a particular date was both a school holiday **and** a public holiday, it was classified as a public holiday. By using the datetime module, each date was also assigned a day of the week in both integer form (0-6) and string form ("Mon", "Tue", "Wed", etc).

Predicting Victorian electricity demand in a pre and post-COVID landscape

Exploratory Data Analysis

AEMO

By loading the AEMO dataset into Tableau, a dashboard could be created which allowed for easy comparison of both demand and rates across different states, years, and months.



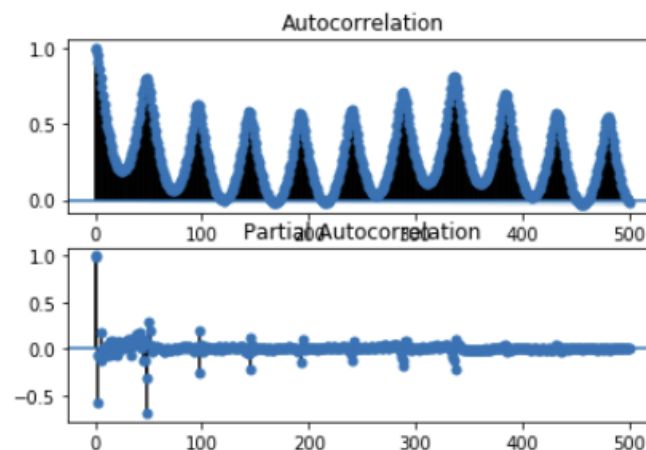
In order to narrow the scope of the project, only Victorian data was considered. This was due to the author's residence, and Victoria's comparatively high levels of restriction.

Observationally, demand appeared to be trending downwards since 2017, and demand during off-peak was down from 2019. However, the 5pm spike in June and July were higher in 2020 than 2019, which could have been attributed to higher working-from-home proportions resulting in higher heater costs during the evening (when people would normally be leaving work) during winter.

Partial/Autocorrelation graphs

As a time series model (LSTM) was being considered, an appropriate number of previous recorded data relevant in predicting current data needed to be determined – was the power demand at 9 am on Tuesday correlated with the power demand at 8 am, or at 9 am on the previous Monday?

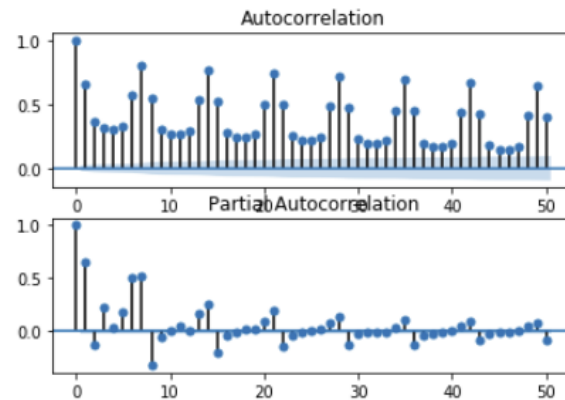
Autocorrelation and Partial Autocorrelation graphs were created in order to examine these trends. First, the entire series of Total Demand was observed:



Clear peaks of correlation occurred at intervals of 48, which corresponded to the amount of data entries in a single day. This inferred that demand at a certain time was strongly correlated with demand at the same time in previous days.

Predicting Victorian electricity demand in a pre and post-COVID landscape

In order to see how far back this correlation went, a subset of data was obtained for Total Demand at 9:30 AM across the dataset, and the above graphs were examined again:



It was apparent that cycles of length 7 were now present, indicating that demand on any day of the week at a certain time was correlated with demand at the same time, on the same day of the week in the previous week.

Using the above graphs, it was determined that the previous 7 days of data was a reasonable amount of data to use when building time series models.

Modelling

Methodology

Although AEMO data went back to 1998, only data from 2014 onwards was used due to the obtainable public holiday data. Models were developed on <2019 data, and tested on 2019 and 2020. The assumption was that if the model performs similarly on 2019 and 2020, covid had no effect on electricity usage prediction. Of the <2019 training data, 80% was used to train models, and 20% was used to validate them and choose the best performing models for the test data.

One Hot Encoding was used to create feature sets for Weekday and Seasonality, so that they could be appropriately used in models. For the Neural Networks, Demand, Year, Month, Day, minimum temperature, maximum temperature, and solar were all scaled in order to provide appropriate inputs before training.

For model comparison, Mean Absolute Percent Error (MAPE) was the main metric considered due to the inherent variability in demand across the years – as demand was generally lower in earlier years, Mean Square Error could similarly have been lower, but ideally absolute percentage would be more standard (if models were well built).

Linear

Intended as a baseline model, an additive linear model was trained using `sklearn.LinearRegression()` in order to build a quick first model.

Long Short Term Memory

Instead of creating one single LSTM model for all times, 48 different models were created using TensorFlow, with each focused a single half hour period of the day. This reduced computational resource usage and allowed LSTMs which considered 7 days of previous data to be created.

Due to time and computing resource constraints, the LSTMs were not batch tuned. Hyperparameters were chosen by a trial-and-error method, with a fully connected LSTM layer matching the input size and 3 fully connected hidden layers of size [6,4,2] being chosen after some experimentation. The learning rate was set to 0.001, after smaller rates resulted in significantly longer training times. Similarly, a batch size of 64 was used, and early stopping was implemented in order to reduce training time. The Adam optimiser was used for training, and the relu activation function was used for all hidden layers. In order to ensure lowest validation error, the epochs used was set to 500 (though early stopping meant that none of the models reached this).

Deep Neural Networks

Similarly to the LSTMs, Neural Networks were built using TensorFlow. To shorten optimisation time, epochs was set to 200 and early stopping was implemented, with the patience was decreased.

As only a single Neural Network needed to be trained for each trial, batch tuning could be more efficiently performed. Grid optimisation was performed, and it was found that the best performing model had hyperparameters:

- Hidden layers: [8,4]
- Learning Rate: 0.001
- Batch size: 128

Result analysis

Model	MAPE
Linear	0.1060
DNN (best)	0.0595
LSTM (average)	0.0657

The best performing model was the batch-tuned DNN, with a validation MAPE of 5.95%. This best model was then trained again (with an extended patience in to get the lowest error) and used to predict results for each year of data:

Year	MAPE
2014	0.0543
2015	0.0506
2016	0.0576
2017	0.0635
2018	0.0604
2019	0.0728
2020	0.0834

While the model has a worse performance on 2020, an examination of the median residuals show the model underestimates both 2019 and 2020 data by a similar amount (2019 median residual: -38.29, 2020 median residual: -33.05).

Discussion

The final model performs similarly on both 2019 and 2020 data, with slightly lower accuracy on the 2020 data. An examination of residuals indicates the model's tendency to underestimate both 2019 and 2020 data by a similar amount. This seems to indicate that when taking into account weather and seasonality, the COVID-19 pandemic has not affected prediction performance much.

This has implications for industry, business, individuals, and government.

- From an electricity industry viewpoint, these results are important for adjusting modelling for the remainder of 2020. In industry, modelling is used in order to predict demand and supply power accordingly. Such modelling is used to ensure that enough power is produced to match demand, but not so much lest the grid be oversupplied. If the prediction model became significantly more inadequate, then future modelling with additional considerations (taking into account COVID restriction level, for example) could be required.
- As a business, it could make sense to incorporate working from home expenses into salary packages in order to transfer savings from less office electricity use to employees. As many businesses and organisations have committed to more flexible work environments post-COVID, this could be a feature of working from home contracts in the near future.
- For individuals, it encourages them to monitor their electricity use carefully. If salary is unaffected when working from home, expenses may increase due to increased utility costs, which may need to be factored in to budgeting. A prudent individual may even build in factors such as transport in to their budgets, to examine if the transport and tax savings outweigh the working from home expenses.
- From a governmental policy point of view, it leads to a re-examining the current working from home tax guidelines – in the 2019-2020 and 2020-2021 fiscal years, the ATO introduced a “Shortcut method” of claiming working from home deductions on tax, which equated to 80 cents for each hour working from home. This varied to the existing 52 cents per hour deduction, though differed in that it did not require employees to have a dedicated work area. If more offices post-COVID adopt a more flexible working from home arrangement, then perhaps a restructure of the working from home tax break may be required.

Limitations

Due to computing resource and personal time limitations, models created were not as optimised as they could be. More epochs for tuning DNN hyperparameters would have been beneficial, and the LSTM models could have utilised a similar batch-tuning method. As models trained were set from random seeds, k-fold cross validation could have been used in order to more robustly compare models, though this would have increased training times by a factor of k.

Future directions

If pursuing this avenue more, it would be useful to have demand data at a more granular location level. By examining demand in smaller areas (such as Local Government Areas) and comparing it to various demographic features of that area, perhaps further insight could be obtained. By examining the occupation area proportions of each area and combining it with the pandemic lockdown level of that area, the proportion of the area population working or learning from home could be estimated, which could be a useful feature to include in prediction models. Finally, in order to provide more benefit for costing, integrating rates predictions and by extension total cost predictions could be useful to provide more use to consumers.

Marc Nguyen, marc.n.224@gmail.com

Predicting Victorian electricity demand in a pre and post-COVID landscape

Links

Tableau workbook available at:

<https://public.tableau.com/profile/marc.nguyen#!/vizhome/PowerDemandandRatesinAustralia1999-2020/PowerDashboard>

Code available at: <https://github.com/MarcNguyen22/MelbourneDatathon2020>