SJM Data201 Project Report

An investigation on the impact of Hydro station energy output, due to river flow, on retails electricity costs.

Data sources

The main source accessed to retrieve information related to the hydro stations was the Electrical Market Information website (https://www.emi.ea.govt.nz/). Within, three sub addresses were accessed:

1. /Wholesale/Datasets/Generation/Generation_MD
2. /Wholesale/Datasets/Generation/Generation_fleet/Existing
3. /Wholesale/Reports/R_NSPL_DR?_si=v|3

Sub address 1 contains the generation output in kWh for every trading period since 1997. Each trading period occurs every 30 minutes, where the energy output of each power producing station documented is recorded. Sub address 2 contains the fleet information, where the general location, station name, node name and further comments are stored. Sud address 3 contains information related to the network supply nodes in the electricity network.

Regarding the hydro stations, the Electricity Market Information website was used as the source of information, because it is used by the Electricity Authority to publish data related to the New Zealand electricity industry. As an "independent crown entity" (Electricity Authority, n.d.), the Electricity Authority can be trusted as a reliable source of information, while allowing the adaptation and distribution of the non-restricted data it supplies (Electricity Authority, n.d.). However, the website must be cited when acquiring the data, which we have done.

The main source accessed to retrieve information related to the river flow was the NIWA Aquarius WebPortal (https://hydrowebportal.niwa.co.nz/), which was used to extract the following download URLs for the datasets:

1. https://hydrowebportal.niwa.co.nz/Export/BulkExport?DateRange=Custom&StartTime=2015-01-01%2000%3A00&EndTime=2020-12-31%2000%3A00&TimeZone=12&Calendar=CALENDARYEAR&Interval=Monthly&Step=1&ExportFormat=csv&TimeAligned=True&RoundData=False&IncludeGradeCodes=False&IncludeApprovalLevels=False&IncludeInterpolationTypes=False&Datasets[0].DatasetName=Discharge%20Value.NRWQN%40TK3&Datasets[0].Calculation=Aggregate&Datasets[0].UnitId=140
2. https://hydrowebportal.niwa.co.nz/Export/BulkExport?DateRange=Custom&StartTime=2015-01-01%2000%3A00&EndTime=2020-12-31%2000%3A00&TimeZone=12&Calendar=CALENDARYEAR&Interval=Monthly&Step=1&ExportFormat=csv&TimeAligned=True&RoundData=False&IncludeGradeCodes=False&IncludeApprovalLevels=False&IncludeInterpolationTypes=False&Datasets[0].DatasetName=Discharge.Master%4068526&Datasets[0].Calculation=Instantaneous&Datasets[0].UnitId=140
3. https://hydrowebportal.niwa.co.nz/Export/BulkExport?DateRange=Custom&StartTime=2015-01-01%2000%3A00&EndTime=2020-12-31%2000%3A00&TimeZone=12&Calendar=CALENDARYEAR&Interval=PointsAsRecorded&Step=1&ExportFormat=csv&TimeAligned=True&RoundData=False&IncludeGradeCodes=False&IncludeApprovalLevels=False&IncludeInterpolationTypes=False&Datasets[0].DatasetName=Discharge%20Value.NRWQN%40TK6&Datasets[0].Calculation=Instantaneous&Datasets[0].UnitId=140
4. https://hydrowebportal.niwa.co.nz/Export/BulkExport?DateRange=Custom&StartTime=2015-01-01%2000%3A00&EndTime=2020-12-31%2000%3A00&TimeZone=12&Calendar=CALENDARYEAR&Interval=PointsAsRecorded&Step=1&ExportFormat=csv&TimeAligned=True&RoundData=False&IncludeGradeCodes=False&IncludeApprovalLevels=False&IncludeInterpolationTypes=False&Datasets[0].DatasetName=Discharge%20Value.NRWQN%40AX2&Datasets[0].Calculation=Instantaneous&Datasets[0].UnitId=140
5. https://hydrowebportal.niwa.co.nz/Export/BulkExport?DateRange=Custom&StartTime=2015-01-01%2000%3A00&EndTime=2020-12-31%2000%3A00&TimeZone=12&Calendar=CALENDARYEAR&Interval=Monthly&Step=1&ExportFormat=csv&TimeAligned=True&RoundData=False&IncludeGradeCodes=False&IncludeApprovalLevels=False&IncludeInterpolationTypes=False&Datasets[0].DatasetName=Discharge.Master%4075207&Datasets[0].Calculation=Instantaneous&Datasets[0].UnitId=140
6. https://hydrowebportal.niwa.co.nz/Export/BulkExport?DateRange=Custom&StartTime=2015-01-01%2000%3A00&EndTime=2020-12-31%2000%3A00&TimeZone=12&Calendar=CALENDARYEAR&Interval=PointsAsRecorded&Step=1&ExportFormat=csv&TimeAligned=True&RoundData=False&IncludeGradeCodes=False&IncludeApprovalLevels=False&IncludeInterpolationTypes=False&Datasets[0].DatasetName=Discharge%20Value.NRWQN%40DN5&Datasets[0].Calculation=Instantaneous&Datasets[0].UnitId=140

Each dataset differs by a unique value for DatasetName, and some differ by the interval value, it was with these differences, I was able to modify each URL, to download the correct information. The difference in interval value was to accommodate the field values. Each dataset contained 2-3 columns:

the date at which the sample was taken, possibly the end of the sampling time (only on half the datasets), and the recorded value in cumecs (m^3/s).

Regarding the river flow data, all of the data was obtained holding to the standards said in the NIWA public data delivery strategy, which I found there to be no substantial requirements, as said in the public data delivery strategy, “*encouraging free access to valuable public data”* (NIWA et al., 2020). NIWAs data can be trusted as they are a recognized crown research institute of New Zealand.

https://teamwork.niwa.co.nz/display/NEDA/NIWA+Environmental+Data+Access+through+standards+based+systems

The main source accessed to retrieve information related to the retail electricity cost was the MINISTRY OF BUSINESS, INNOVATION & ENPLOYMENT Web Portal (https://www.mbie.govt.nz/). Within, the sub addresses were accessed:

1. https://www.mbie.govt.nz/building-and-energy/energy-and-natural-resources/energy-statistics-and-modelling/energy-statistics/energy-prices/electricity-cost-and-price-monitoring/
2. https://www.mbie.govt.nz/assets/Data-Files/Energy/nz-energy-quarterly-and-energy-in-nz/qsdep-report-15-nov-2020.xlsx (Excel file, sheet 3)

This data contains retails electricity costs which are surveyed mid-point of each quarter (15 February, 15 May, 15 August and 15 November each year from 2003 to 2020) for around 40 towns and cities across New Zealand. The unit of electricity costs value is denoted in cents per kWh (c/kWh).

Target audience

Hydro stations provide the majority of the electricity needed in New Zealand (MBIE, 2021). For this reason, our project investigated the effects hydro station generation output would have on the retail electricity costs in the South Island, which would in turn be affected by river flow. Although we could not make a correlation between river flow and retail electricity costs, the data gathered could be of use to:

- Residential homeowners or buyers.
- Hydro station designers.
- Water conservancy engineers, power engineers and actuaries of energy companies.
- Companies which operate on and around the rivers and hydro stations.

Our data would be of use to residential homeowners or buyers as the data contains the prices, and change in price of power over 5 years, and could be used as a guide or to estimate how the cost will change for years to come. The data would be of use to hydro station designers as the data contains information about river flow, and the power output related to river flow for each of the three rivers identified, which could be used to maximize efficiency of future stations. The data would be of use to water conservancy engineers, power engineers and actuaries of energy companies because of the links between water flow and power, and the links between power output and cost in cents per kWh at the residential level. The data would be of use to companies which operate on and around the rivers and hydro stations as water flow levels may affect them, since hydro stations may affect people who provide recreational services on the rivers, as you cannot navigate through hydro stations.

Difficulties

The code to web scrape the hydro station data involved reading multiple csv documents. The URLs were passed to 'read_csv()' via the 'map_dfr()' function, from the 'purrr' package. However, this was found to be too slow. Instead, 'future_map_dfr()' was used, from the 'furrr' package, which allowed parallelization of the 'map_dfr()' function, dramatically decreasing the total time for the process to occur.

The code to web scrape the town names of the south island also proved to be difficult. This was due to the town names being separated alphabetically when scraped into R. To separate them into one single list, with the help of Thomas Li, we initialized a blank list, then iterated through each alphabetically list, then indexed into the big alphabetical list at each index 'i' to obtain the town names at each alphabetical character, then extracted the towns as a vector and append each town to the blank list.

When locating the river data, there was quite a bit of difficulty locating the rivers. Jensen provided Matt with NZTM coordinates, which proved to not be a reliable way of locating the rivers, so instead, Matt manually searched the gen code in google maps, in hope to locate the hydro station, which in turn would reveal the river name to search in the Aquarius WebPortal. This took a long time but was our only reliable way of locating the river name which is required to obtain the data sets for river flow.

The merging processes between the final data frames proved difficult. A common key needed to be identified between the data frames, and in the case between the river flow data frame and the hydro station data frame, the key needed to be created before the merging process could begin. The creation of the key proved to be the most work, as the year and month needed to be extracted from the dates of each row in both data frames. The data within the river flow data frame contained flow data by month, whereas the hydro station generation output data frame needed to be averaged by month. Once the averaging was done, both data frames could be merged. Refer to the relational data model, stored in the repository to view a more detailed format of the merging process.

While wrangling the electricity costs dataset, the conversion was successful when the dates were converted from characters to the date data type using the as.date function, but in the process of renaming the column names, the column defining the region name was deleted, making it a little difficult to solve the problem.

Aside from coding difficulties, pushing changes to GitHub proved to be difficult as well. It was often the case when two team members would edit the same document at the same time. One team member would push their changes to the main repository. However, once the other team member tried pushing their change, a merge conflict would occur. Since Jupyter notebooks were being edit, it was difficult to resolve the merge conflicts, as GitHub would write to the notebooks to identify the merge conflicts. This would corrupt the notebooks as Jupyter lab would no longer identify the notebooks as Json files. To avoid this issue, the work was split into four notebooks, one for each topic of data being web scraped, and one for the South Island town name web scrape. The finished code was then combined into a single notebook to produce the final product.

Techniques

The entire project was programmed within the R language, with a heavy use of the tidyverse collection library and presented in a Jupyter notebook.

Instead of using multiple for loops to iterate over each row, such as to convert to a date type with ‘as.Date()’; ‘map_dfr()’ or ‘future_map_dfr()’ were used instead. This allowed the use of custom functions, such as in the case where part of the data across a column needed to be extracted into another column (for example, when creating the key to merge the river flow with the hydro station generation output data frame). A custom function was created to extract a subset of the data, which was then further processed, and the result was returned. The ability to map a function to all the rows of a data frame significantly decreased the total lines of code for the project.

Custom functions were also used to produce plots from the final data frames. In the example of the electricity cost by town, a town name could be passed as an argument. The function would select the river flow for the chosen town only and return a plot of the selected data.

To acquire the URL links to web scrape the data, the html code from the relevant pages were extracted. The URL links were identified by filtering for relevant attributes (e.g., ‘href’) and tags (e.g., ‘a’). These attributes were identified by searching within the pages themselves through a browser inspector in the console and inspecting the source code, except for the river data which was acquired from the download URL at the bottom of the webpage of the Aquarius WebPortal and was manipulated for each river and timeframe.

Only the data from 2015 to 2020 was used, which are data required from retail electrical cost, and retail average electrical cost data from only South Island were extracted.

In addition, the character was converted to a date string using the as.date function, and the converted values were reset in the columns name.

Achievements and failures

"Hydro station energy production by river flow with rainfall" was the topic that we chose first. However, we could not find sufficient rain fall data over the long term, for the locations of the hydro stations. As a result, the topic was changed to "Hydro station energy production by river flow with electricity cost".

In order to identify which river each hydro station is situated on; an attempt was made to use the NZTM coordinates stored within the network supply node data set. A data frame was formed containing an NZTM coordinate for each hydro station name identified. This was achieved by merging the hydro station fleet data frame with the network supply node data frame by the node name: a 3-letter code followed by 4 digits identifying a node within the network. However, when the coordinates were used to locate each power station, the 3 checked ended up pointing to the office building (assumed) rather than the actual hydro station, so the coordinates were not a reliable way to locate the river, thus we manually searched for the hydro station and then located the name of the river the station is situated on, and due to this, we were only able to locate 13 hydro stations situated on rivers, with only 11 of those hydro stations river containing 2015-2020 data points.

When the river names had been identified, we referred to the NIWA Aquarius portal to obtain data sets on each river, this caused issues, as some rivers ended up having only field measures of flow, rather than continuous river flow measurements like we expected, this was resolved by setting the interval for some, to all points as recorded, and others to summarize by month.

With the 11 river datasets, only 3 hydro stations containing the generation output data were identified, on three different rivers. This further reduced the breadth of our data, meaning only three hydro stations generation output could be compared with three different rivers. As a result, a correlation could not be made between the effect of river flow on the retail electricity costs in the South Island. Instead, the effects of changes in hydro station generation output on retail electricity costs were investigated, as well as the changes in hydro station generation output due to river flow, for the three hydro stations identified.

Even with these failures, we were still able to get good results. There are some obvious links between river flow and hydro station power output, with the link being as river flow increases, the power output for hydro stations located on the river increases, and link between cost of power and hydro station output, with the link being as hydro station output increases, the cost of residential power decreases, but due to cost of residential power relying on a lot of other factors, we expected not a huge impact.

Report Bibliography

Electricity Authority. (n.d.-a). *Electricity Authority - EMI*. Retrieved October 30, 2021, from https://www.emi.ea.govt.nz/LegalInformation

Electricity Authority. (n.d.-b). *Who we are — Electricity Authority*. Retrieved October 30, 2021, from https://www.ea.govt.nz/about-us/who-we-are/

MBIE. (2021, September 10). *Electricity statistics | Ministry of Business, Innovation & Employment*. https://www.mbie.govt.nz/building-and-energy/energy-and-natural-resources/energy-statistics-and-modelling/energy-statistics/electricity-statistics/

NIWA, Schmidt, J., & Brown, D. (2020). *NIWA Environmental Data Access through standards based systems - NIWA Environmental Data Access - Teamwork*. https://teamwork.niwa.co.nz/display/NEDA/NIWA+Environmental+Data+Access+through+standards+based+systems