

**School of InfoComm Technology**

**Data Science Capstone Project (DSCP)**

Year 3 (2023/24), Semester 5

**SCHOOL OF INFOCOMM TECHNOLOGY**

Diploma in Data Science

**ASSIGNMENT 3**

**Project Documentation**

|  |  |
| --- | --- |
| **Tutorial Group:** | **T01** |

|  |  |  |
| --- | --- | --- |
| **Student Number** | **Student Name** | **User Stories # & Short Description** |
|  | **Lim Wee Liang Kelven** | **Energy Consumption**   1. **As FM, I would like to know which months in 2022 had high energy consumption so that I can forecast 2023's contracted capacity.** 2. **As FM, I want to know how much I can reduce 2023's contracted capacity so that I can save money.** 3. **As FM, I want to forecast 2023's energy usage so that I know if and when the campus will reach 90% contracted capacity.** 4. **As FM, I want to predict 2024's contracted capacity so that I can save money.**   **Canteen Crowd**   1. **As FM, I want to know which day has the greatest number of people so that solutions can be adopted to prevent overcrowding.** 2. **As FM, I want to know which day has the least number of people so that canteen vendors can plan how much food to prepare to reduce food wastage.** 3. **As FM, I want to know which canteen is the most popular so that solutions could be implemented to spread out the number of people across all canteens.** 4. **As FM, I want to know which canteen is the least popular so that solutions could be implemented to attract students there.** |

Contents

[Introduction 5](#_Toc142121684)

[Energy Consumption 5](#_Toc142121685)

[Canteen Crowd 7](#_Toc142121686)

[Methodology 9](#_Toc142121687)

[Energy Consumption 9](#_Toc142121688)

[Energy Consumption 9](#_Toc142121689)

[Tariff Prices 20](#_Toc142121690)

[Contracted Capacity 23](#_Toc142121691)

[Canteen Crowd 32](#_Toc142121692)

[Canteen Crowd 32](#_Toc142121693)

[Canteen Coordinates 49](#_Toc142121694)

[Assumptions and Limitations 50](#_Toc142121695)

[Results 51](#_Toc142121696)

[Energy Consumption 51](#_Toc142121697)

[Energy Consumption Overview (ECO) 51](#_Toc142121698)

[Energy Pricing (GST Inclusive) (EP) 59](#_Toc142121699)

[Canteen Crowd (CC) 68](#_Toc142121700)

[Discussion 78](#_Toc142121701)

[Energy Consumption 78](#_Toc142121702)

[Question 1: I would like to know which months in 2022 had high energy consumption so that I can forecast 2023's contracted capacity. 78](#_Toc142121703)

[Question 2: I want to know how much I can reduce 2023's contracted capacity so that I can save money. 79](#_Toc142121704)

[Question 3: I want to forecast 2023's energy usage so that I know if and when the campus will reach 90% contracted capacity. 80](#_Toc142121705)

[Question 4: I want to predict 2024's contracted capacity so that I can save money. 80](#_Toc142121706)

[Canteen Crowd 81](#_Toc142121707)

[Question 1: I want to know which day has the greatest number of people so that solutions can be adopted to prevent overcrowding. 81](#_Toc142121708)

[Question 2: I want to know which day has the least number of people so that canteen vendors can plan how much food to prepare to reduce food wastage. 82](#_Toc142121709)

[Question 3: I want to know which canteen is the most popular so that solutions could be implemented to spread out the number of people across all canteens. 82](#_Toc142121710)

[Question 4: I want to know which canteen is the least popular so that solutions could be implemented to attract students there. 84](#_Toc142121711)

[Assumptions and Limitations 85](#_Toc142121712)

[Areas for Improvement 85](#_Toc142121713)

[Reflection 86](#_Toc142121714)

[Q1. Describe your experience working on the project, including any challenges faced and lessons learnt 86](#_Toc142121715)

[Q2. Reflect on your personal growth and development in the areas of data analytics, visualisation, and project management 87](#_Toc142121716)

[Q3. Identify any areas for improvement in your skills or knowledge 87](#_Toc142121717)

[Q4. Discuss the impact that the project had on your understanding of the topic and its relevance to real-world problems 88](#_Toc142121718)

[Correlation to other Datasets 88](#_Toc142121719)

[Canteen Crowd and Energy Consumption 88](#_Toc142121720)

[Canteen Crowd and Toilet Utilisation 89](#_Toc142121721)

[Conclusion 89](#_Toc142121722)

[Energy Consumption 90](#_Toc142121723)

[Canteen Crowd 90](#_Toc142121724)

[Professional Certification 91](#_Toc142121725)

[Bibliography 91](#_Toc142121726)

# Introduction

Facilities management is an essential aspect of any organisation as it encompasses the management of buildings, equipment, and other physical assets. Effective facility maintenance and operations are critical in ensuring the safety, comfort, and productivity of employees and customers. This capstone project aims to use data analytics to optimise facility maintenance and operations by identifying trends, patterns, and potential issues, and providing insights to help facilities managers make data-driven decisions.

This report will analyse energy consumption data and canteen occupancy data. The energy consumption data will be obtained from utility bills or energy management systems, while the canteen occupancy data will be obtained from access control systems or occupancy sensors.

## Energy Consumption

The energy consumption on campus is a pressing concern, especially with the increasing demand for energy in recent years. Chillers mainly consume energy during the day, data centres consume energy all day, and streetlights at night. In addition, 10 chiller plants on the campus consume 40 to 50% of the energy.

However, serious intervention is needed when the energy consumption reaches 90% of the contracted capacity or the campus risks facing double energy bills. Thus, there is an urgent need to reduce energy consumption and promote sustainable energy practices to ensure the campus's long-term sustainability.

The questions this project aims to answer are:

1. I would like to know which months in 2022 had high energy consumption so that I can forecast 2023's contracted capacity.
2. I want to know how much energy I can reduce 2023's contracted capacity so that I can save money.
3. I want to forecast 2023's energy usage so that I know if and when the campus will reach 90% contracted capacity.
4. I want to predict 2024's contracted capacity so that I can save money.

The datasets provided for this problem are Excel files from November 2019 to March 2023, which amounts to 41 files.

A group of numbers and symbols

Description automatically generated

These 41 files will herein be dubbed as “Energy Consumption” for convenience. Each file contains 2 columns: **Date** and **Energy\_Consumed (W)**. The amount of energy consumed is recorded every 30 minutes.

A screenshot of a data table

Description automatically generated

In addition to Energy Consumption, 2 more Excel files were created using Alteryx. These files are called “Contracted Capacity” and “Tariff Prices”.

Contracted Capacity contains the contracted capacities from previous years. There are 2 columns: **date** and **contracted\_capacity (kW)**.

A screenshot of a data

Description automatically generated

The contracted capacitiesfor 2019 to 2022 were taken from the challenge statement PDF and the data dictionary. I assumed that the contracted capacity for each year stated in the files was the same as the capacity in each month of that year. For example, if 2021’s contracted capacity was 7800 kW, every month in 2021 has the same contracted capacity of 7800 kW.

Tariff Prices contain the historical contracted capacity tariffs or rates from October 2014 to July 2023. This information was gathered from SP Group (SP Group, 2023).

A grid of numbers and letters

Description automatically generated

Since Energy Consumption only spans from November 2019 to March 2023, the rates from October 2019 to January 2023 are taken. The school uses High Tension Large (HTL) Supplies, the prices are inclusive of GST and under the GST rate at their given times. The rates are taken from Contracted Capacity Charge ($/kW/month). Since the prices are given every quarter (*e.g., January, April, July, October*), I assumed that the months in between contain the same rate (*e.g., the rate in October 2019 is $9.52/kW/month, hence November and December will have the same rate*). There are 2 columns: **date** and **contracted\_capacity\_price** which is $/kW/month.

A screenshot of a table

Description automatically generated

## Canteen Crowd

The canteen has a problem with overcrowding during lunch hours, especially during school term time. In response, people counting sensors have been installed at all exit points in the canteen to monitor the number of people entering and leaving the canteens. With this data, the canteen capacity during given hours can be determined.

The questions this project aims to answer are:

1. How can I utilise this data to suggest to users the best time to go for lunch?
2. Can the data be used to plan the amount of food to prepare by the canteen operators?
3. I want to know which day has the greatest number of people so that possible solutions can be adopted to prevent overcrowding.
4. I want to know which day has the least number of people so that canteen vendors can plan how much food to prepare to reduce food wastage.
5. I want to know which canteen is the most popular so that solutions could be implemented to spread out the number of people across all canteens.
6. I want to know which canteen is the least popular so that solutions could be implemented to attract students there.

The dataset provided, herein dubbed as “Canteen Crowd”, for this problem is retrieved from a website using an API call (Ngee Ann Polytechnic, n.d.). To retrieve Canteen Crowd, an api\_key and a date must be given. Location can be ignored as all the canteen information will be retrieved by default.

A screenshot of a computer

Description automatically generated

After filling in a date, the following details will be given in JSON format: description, the date, the number of people per hour, peak hour (the hour with the most people), the total number of people, and the location name.

A screenshot of a computer

Description automatically generated

In addition to Canteen Crowd, 1 more Excel file was created using Alteryx. This file is called “Canteen Coordinates”.

A screenshot of a computer

Description automatically generated

Canteen Coordinates contain the coordinates of the 3 canteens in NP. There are 3 columns: **canteen\_location**, **long**, and **lat**. The coordinatesfor each canteen were not provided by the school. Hence, they were taken from an online website (MapTiler Team, n.d.).

# Methodology

## Energy Consumption

### Energy Consumption

#### Loading Data

A screen shot of a computer

Description automatically generated

The Input Data tool was used to read the raw data.



Since there are 41 files in the folder, I specified \*.xlsx at the end to read all Excel files in the folder.

A screenshot of a computer

Description automatically generated

#### Transformation

A screenshot of a computer

Description automatically generated

Examining the data, there were some null rows.

A screenshot of a computer

Description automatically generated

The Filter tool was used to remove the null rows.

A blue circle with white and black text

Description automatically generated

**Energy\_Consumed (kW)** was created from **Energy\_Consumed** using the Formula tool. Since energy consumed was measured in Watts (W), it is converted into kilo Watts (kW) by dividing by 1000. This is needed as the contracted capacity and tariffs are measured in kW.

A blue and green logo

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

The columns **Energy\_Consumed**, and **FileName** are filtered out and **Date** is renamed to **date** using the Select tool.

A blue circle with a check mark and green arrows

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a calculator

Description automatically generated

A new column, **hour**, is created and the timings for **date** are removed using the Formula tool.

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

To find the hourly consumption, the average is taken from 2 rows. For example, the average between 2019-11-01 00:00:00 and 2019-11-01 00:30:00 is taken to find the hourly consumption of midnight.

A screenshot of a phone number

Description automatically generated

To do this, the Summarize tool is used.

A screen shot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a calculator

Description automatically generated

The **date** column is converted into a proper datetime format using the DateTime tool to allow proper time series analysis later.

A computer screen shot of a diagram

Description automatically generated

A screenshot of a computer program

Description automatically generated

A screenshot of a computer

Description automatically generated

The **year**, **month**, and **DayOfWeek** are retrieved from the **Date\_Out** column using the Formula tool. year and month can link this Energy Consumption to Contracted Capacity and Tariff prices, while **DayOfWeek** is created for analysis later.

A blue and white circle with a blue circle with a white circle with a blue circle with a white circle with a blue circle with a white circle with a blue circle with a white circle with a

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a calendar

Description automatically generated

Since there is no standard deviation (SD) formula in the Formula tool, the Summarize tool is used to calculate the average and SD of the energy consumed.

A diagram of a graph

Description automatically generated

A screenshot of a computer

Description automatically generated



The SD and average are then added back to the original table using the Append Fields tool.

A diagram of a diagram

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

After appending, the energy consumed is transformed using a Z-score transformation. This new Z-score will be used with DayOfWeek for analysis later.

The hour of the day is checked if it was within peak hours. Using **year**, **month**, and whether it was during peak hours, the cents/kWh rate is found, and the electricity usage price is calculated.

All these calculations were performed using a Formula tool.

A screenshot of a phone

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

The pricing for **rate\_cents** was gathered from SP Group (SP Group, 2023).

A grid of numbers and letters

Description automatically generated

The kWh charge (cents/kWh) was taken under HTL Supplies. The prices are inclusive of GST and under the GST rate at their given times.

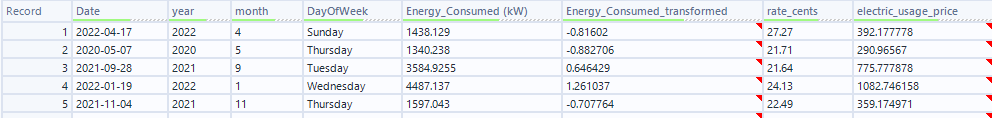
Lastly, the columns **date**, **hour**, **StdDev**, **Avg**, and **peak** are filtered out and **Date\_Out** is renamed to **Date** using the Select tool.

A diagram of a diagram

Description automatically generated

A screenshot of a computer

Description automatically generated



#### Data Output

A screenshot of a computer

Description automatically generated

The final dataset is then saved as an Excel file using the Output Data tool.

A screenshot of a computer

Description automatically generated

### Tariff Prices

#### Loading Data

A screenshot of a computer

Description automatically generated

The Text Input tool was used to create the data. As mentioned previously, this information was gathered from SP Group (SP Group, 2023).

A screenshot of a data

Description automatically generated

#### Transformation

A diagram of a transformation

Description automatically generated

The **date** column is converted into a proper DateTime using the DateTime tool to allow proper time series analysis later.

A green and white diagram with a clock and arrow

Description automatically generated

A screenshot of a computer program

Description automatically generated

A screenshot of a computer error

Description automatically generated A screenshot of a computer

Description automatically generated

The year and month are retrieved from **DateTime\_Out** using the Formula tool.

A blue and white graphic with arrows and a blue circle with a blue circle with a blue circle with a blue circle with a white circle with a blue circle with a blue circle with a white circle

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a calendar

Description automatically generated

Lastly, the column **date** is removed and **DateTime\_Out** is renamed to **date** using the Select tool.

A screenshot of a computer

Description automatically generated

**A screenshot of a computer

Description automatically generated**

A screenshot of a table

Description automatically generated

#### Data Output

A screenshot of a computer

Description automatically generated

The final dataset is then saved as an Excel file using the Output Data tool.

A screenshot of a computer

Description automatically generated

### Contracted Capacity

#### Loading Data

A screenshot of a computer

Description automatically generated

The Text Input tool was used to create the data. As mentioned previously, this information was gathered from the challenge statement and data dictionary provided.

A screenshot of a computer

Description automatically generated

#### Transformation

A screenshot of a computer

Description automatically generated

The **date** column is converted into a proper DateTime using the DateTime tool to allow proper time series analysis later.

A green and white diagram with a clock and arrow

Description automatically generated

A screenshot of a computer program

Description automatically generated

A screenshot of a computer error

Description automatically generated

A screenshot of a computer

Description automatically generated

The year and month are retrieved using the **DateTime\_Out** column and the 90% capacity is calculated using the Formula tool.

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a data sheet

Description automatically generated

Using the cleaned Energy Consumption from before, the highest consumption by year and month is found using the Summarize tool.

A screen shot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a data

Description automatically generated

The Max\_Energy\_Consumed (kW) column is joined back to Contracted Capacity using the Join tool based on **year** and **month**. **date**, **year**, and **month** from the right input are redundant and filtered out.

A diagram of a diagram

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

The **predicted\_capacity (kW)** is calculated assuming **Max\_Energy\_Consumed (kW)** was 90% of max capacity. If **predicted\_capacity (kW)** is more than the contracted capacity, the value would be set to the contracted capacity. **reduced\_capacity** is calculated using the difference between the original and the predicted using the Formula tool.

A white paper with black text

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

The contracted capacity price from the cleaned Tariff Prices is added to Contracted Capacity using the Join tool based on year and month. **date**, **year**, and **month** from the right input are redundant and filtered out.

A diagram of a diagram

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

The amount saved from reducing the capacity is calculated using the Formula tool.

A diagram of a chemical reaction

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

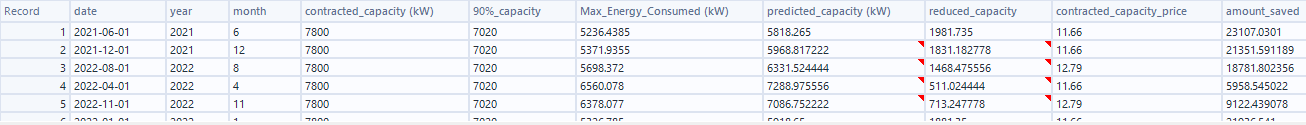
Lastly, the column **DateTime\_Out** is renamed to **date** using the Select tool.

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated



#### Data Output

A screenshot of a computer

Description automatically generated

The final dataset is then saved as an Excel file using the Output Data tool.

A screenshot of a computer

Description automatically generated

## Canteen Crowd

### Canteen Crowd

#### Input Year and Month Number

A diagram of a diagram

Description automatically generated

5 columns were created using the Text Input tool: **year**, **month**, **date (y/m/d)**, **url**, and **api\_key**.

A computer screen shot of a computer

Description automatically generated



The user can change the **year** and **month** of when to retrieve data. The **date** column will be used later. The **url** column is the website link which the data will be retrieved from. The **api\_key** column is the API key of the website.

The Select tool was used to change the data types of **year**, **month,** and **date** columns to V\_WStrings.

A diagram of a data flow

Description automatically generated

A screenshot of a computer

Description automatically generated

#### Create Date and Download Data from the Web

A screenshot of a computer

Description automatically generated

Using the year and month value provided, 30 days’ worth of data are downloaded. Firstly, the date is created using the Formula tool.

A close-up of a diagram

Description automatically generated

A screenshot of a computer program

Description automatically generated

The ‘1’ at the end signifies the date of the month. Only 28 days are used on account for February having 28 days.

Next, the URL is created using the Formula tool. The URL combines the API key and the date created.

A blue and white symbol with green arrows

Description automatically generated

A screenshot of a computer

Description automatically generated

Using the URL, the data is downloaded from the web using the Download tool.

A diagram of a data transfer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated



After downloading the data, the Select tool is used to filter out all columns except for **DownloadData** and **\*Unknown**.

A blue circle with white dots and a check mark

Description automatically generated

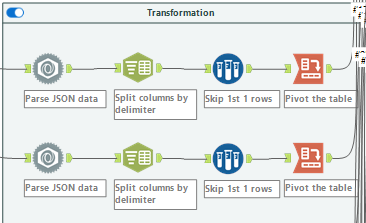
A screenshot of a computer

Description automatically generated



These 4 steps: creating the date, creating the URL, downloading the data, and filtering columns, are repeated for 28 days a month.

#### Transformation



Since the data is in JSON format, the JSON Parse tool is used to unpack the data.

A diagram of a data processing

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

The data is now in key-value pairs (JSON\_Name and JSON\_ValueString). The names have redundant information in them such as “data” and “reports”. To separate them, the Text to Columns tool is used.

A green and blue line with a hexagon and a white rectangle with black text

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

The first row contains null values and is removed using the Sample tool to skip the first row.

A diagram of test tubes

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

Finally, the table is pivoted using the Cross Tab tool.

A close-up of a diagram

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

These 4 steps: parsing, splitting, skipping, and pivoting, are repeated for 28 days of the month.

#### Summarise Data

A screenshot of a computer screen

Description automatically generated

The data is combined using the Union tool.

A purple and white logo with green arrows

Description automatically generated

A screenshot of a computer

Description automatically generated

Next, columns **3**, **location\_id**, **peak\_hour**, and **total** are removed, **location\_name** is renamed to **canteen\_location**, and the data types for **h0 to h23** are changed to Int64 using the Select tool. For this analysis, **peak\_hour** was deemed irrelevant.

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

The names in **canteen\_location** contain the zone number, block number, floor level number, and door. For this analysis, only the block number is important. Thus, the values are renamed using the Formula tool.

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

Finally, **h0 to h23** are summarised and grouped by **canteen\_location** and **date** using the Summarize tool.

A close-up of a logo

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

#### Secondary Transformation

A diagram of a diagram

Description automatically generated

Firstly, the hours and number of people are pivoted using the Transpose tool.

A computer screen shot of a diagram

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

Next, the columns **Name** and **Value** are renamed to **hour** and **number\_of\_people** respectively, and **hour**’s data type is changed using the Select tool.

A close-up of a diagram

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

The seating capacities of each canteen were shared by the product owner via Microsoft Teams. This was added to the workflow using the Text Input tool.

A close-up of a computer screen

Description automatically generated

A screenshot of a computer

Description automatically generated

This capacity information is then joined to the original table using the Join tool. The **canteen\_location** from the right input is redundant and filtered out.

A diagram of a diagram

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

The **dayofweek** is found using the Formula tool.

A blue and white logo with arrows

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a phone

Description automatically generated

Since school is closed during the weekends, Saturdays and Sundays are filtered out using the Filter tool.

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

Lastly, the number of **seats\_left**, **seats\_left\_%**, and **excess\_people** are calculated using the Formula tool. **seats\_left\_%** is used to find the remaining capacity for each canteen, and **excess\_people** is the number of people who enter when the canteen is full.

A screenshot of a phone

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

#### Data Output

A screenshot of a computer

Description automatically generated

The final dataset is then saved as an Excel file using the Output Data tool. A screenshot of a computer

Description automatically generated

### Canteen Coordinates

#### Loading Data

A screenshot of a computer

Description automatically generated

The Text Input tool was used to create the data. As mentioned previously, the coordinatesfor each canteen were not provided by the school. Hence, they were taken from an online website (MapTiler Team, n.d.).

A screenshot of a table

Description automatically generated

#### Data Output

A screenshot of a computer

Description automatically generated

The final dataset is then saved as an Excel file using the Output Data tool.

A screenshot of a computer

Description automatically generated

## Assumptions and Limitations

When loading the data for Energy Consumption, 2.07% of the data were null. It is assumed that this 2% is insignificant to the overall accuracy of the data.

There was also a limitation in identifying which files the null values came from. Even though the **FileName** column shows which file it is, one must still sift through rows of data manually to find the missing rows.

For Energy Consumption, it is assumed that the contracted capacities for each year stated were the same for each month of that year. For example, if 2021’s contracted capacity was 7800 kW, every month in 2021 has the same contracted capacity of 7800 kW.

The contracted capacity was not given for 2023. Hence, when calculating **reduced\_capacity** and **amount\_saved**, they are 0.

For Canteen Crowd, it is assumed that the canteen zones, door numbers, and peak hours were irrelevant. It also assumed that the school and canteens remain closed during the weekends.

Loading the data using the way shown makes the workflow impractical. Copying and pasting 28 copies of the steps is laborious. The use of a macro might be more efficient as it is more automated and cleaner-looking. Reducing the number of steps to 4 steps instead of 28x4 minimises the chance of human error and makes the transformation phase runs faster.

# Results

## Energy Consumption

### Energy Consumption Overview (ECO)

A screenshot of a computer

Description automatically generated

**Year and month slicer**



The year and month slicers are grouped in a box. They allow the user to filter the charts by year and month. The box was added using the Insert tab 🡪 Shapes.

A screenshot of a computer

Description automatically generated

The year slicer was created using **Year** from the Date Hierarchy from **energy\_consumption**.

A screenshot of a computer

Description automatically generated

The month slicer was created using **Month** from the Date Hierarchy from **energy\_consumption**.

A screenshot of a computer

Description automatically generated

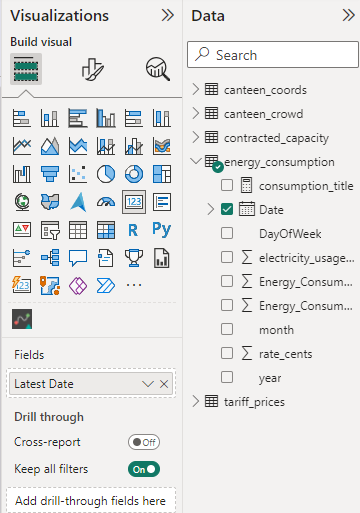
**Highest Consumption card**

A close up of a sign

Description automatically generated

This card shows the date when the highest consumption occurred. It is affected by the year and month slicers. It shows that in December 2022, 1 December had the highest energy consumption.

This chart was created using **Date** from **energy\_consumption**.



A filter was applied to show the top 1 date with the highest energy consumption.

A screenshot of a computer

Description automatically generated

Knowing the date may allow users to reference past events, such as a celebration occurring which resulted in high energy consumption.

**Average Daily Consumption clustered bar chart**

A screenshot of a computer

Description automatically generated

This bar chart shows the average daily energy consumption excluding weekends. It is affected by the year and month slicers. It is observed that in December 2022, Thursday had the highest average consumption at 3.1k kW. Whereas Monday had the lowest at 2.7k kW.

This chart was created using **DayOfWeek** and the average of **Energy\_Consumed (kW)** from **energy\_consumption**.

A screenshot of a computer

Description automatically generated

This chart allows users to know which days are the most and least busy. This can enable more efficient resource and lesson planning.

**Highest Consumption Reached Gauge**

A screen shot of a graph

Description automatically generated

This gauge shows the energy consumption of the date with the highest consumption. The max value is the contracted capacity of the given year, and the red line near the right end represents 90% of the capacity. It is affected by the year and month slicers. It shows that for Thursday, 1 December 2022, it consumed 5.83k kW of electricity.

This chart was created using the averages of **90%\_capacity** and **contracted\_capacity (kW)** from **contracted\_capacity** and the max of **Energy\_Consumed (kW)** from **energy\_consumption**.

A screenshot of a computer

Description automatically generated

This chart helps visualise how close the school was to reaching 90% capacity. Since it shows how dates with the highest consumption are not always near the capacity limit, it helps users make better-informed decisions about the school’s performance.

**Average Monthly and Daily Consumption Heatmap**

A screenshot of a graph

Description automatically generated

This heatmap shows the average energy consumption based on month and day of week, green squares represent low consumption, white represents moderate consumption, and red represents high consumption (green (low) 🡪 white (medium) 🡪 red (high)). The scale of the values has been reduced via a Z-score transformation (). To identify which months or days had high consumption, the number of red squares is counted per row or column respectively. It is affected by the year slicer. This shows that May had high energy consumption on Wednesdays. This could be due to a new semester starting. Conversely, February, March, and September had low consumption because of semester breaks.

This chart was created using **Month** from the Date Hierarchy, **DayOfWeek**, and **Energy\_Consumed\_transformed** from **energy\_consumption**.

A screenshot of a computer

Description automatically generated

A filter was also used to filter out Saturday and Sunday from DayOfWeek.

A screenshot of a computer

Description automatically generated

This heatmap is used to find how months and days of the week affect consumption levels. An important thing to remember is this chart compares consumption levels with each month and day of the week and does not explicitly show the actual consumption levels. Nevertheless, it helps identify which months or days have high energy consumption.

**Daily Consumption line chart**

A graph on a screen

Description automatically generated

This line chart is a time-series analysis of the daily highest energy consumption. It is affected by the year and month slicers. It is observed that the start of December had high consumption, which decreases to the end of the month. The dips in each cycle are Saturdays and Sundays.

This chart was created using **Date** and **Energy\_Consumed (kW)** from **energy\_consumption**.

A screenshot of a computer

Description automatically generated

This line chart shows weekly cycles/trends and how energy consumption changes throughout the month. This line chart also helps identify outliers. For example, 26 December was a Monday, but it was also a public holiday, hence it had low consumption.

### Energy Pricing (GST Inclusive) (EP)

A screenshot of a computer

Description automatically generated

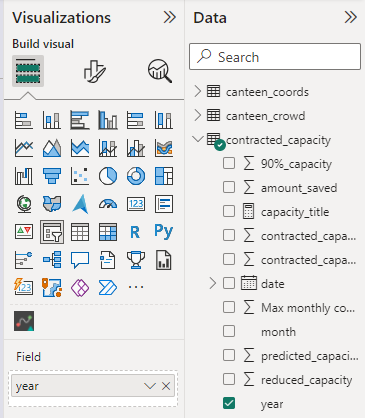
**Year slicer**

A screenshot of a computer

Description automatically generated

The year slicer allows the user to filter the charts by year. This allows users to filter the visuals by year.

This chart was created using **year** from **contracted\_capacity**.



A filter is applied to remove 2024, blanks, and 0.

A screenshot of a computer

Description automatically generated

**Contracted Capacity card**

A close up of numbers

Description automatically generated

This card shows the contracted capacity of the given year. It is affected by the year slicer. It is observed that the contracted capacity for 2022 is 7800 kW.

This chart was created using the max of **contracted\_capacity (kW)** from **contracted\_capacity**.

A screenshot of a computer

Description automatically generated

**HTL Contracted Price line chart**

A graph of a price

Description automatically generated

This line chart is a time-series analysis of the HTL contracted capacity price rate ($/kW). It is unaffected by the year slicer. It is observed that the price increases over time. In 2020, it was $9.52/kW, this increased to $12.91/kW in 2023.

This chart was created using **contracted\_capacity\_price** and **date** from **tariff\_prices**.

A calendar and date with text

Description automatically generated

This chart shows the forecasted prices and helps users create more accurate calculations for predicted costs.

**Contracted Capacity Possible Savings matrix**

A screenshot of a graph

Description automatically generated

This matrix shows the monthly consumption, predicted capacity, reduced capacity, and amount saved each month. It is unaffected by the year slicer. It shows the highest Monthly Consumption, Predicted Capacity, Reduced Capacity, and Amount Saved. The calculation for these columns can be found in the Methodology section.

This chart was created using **Month** from the Date Hierarchy, **Max monthly consumption** (renamed to Monthly Consumption), **predicted\_capacity (kW)**, **reduced\_capacity**, and **amount\_saved**.

A screenshot of a computer

Description automatically generated

This matrix is mainly used to show how much capacity could be reduced in 2022 and the resulting savings. The totals for Reduced Capacity and Amount Saved are shown since they are the most important features. In 2023, if the contracted capacity is adjusted each month, it could lead to a yearly savings of $200,770.

**Monthly Electricity Bill table**

A screenshot of a computer

Description automatically generated

(^ 2022)

This matrix shows the monthly electricity bill. It is unaffected by the year slicer. The calculation for this column can be found in the Methodology section. The months with the highest electricity bills are highlighted in red. It is observed that May, July, and November have the highest bills and the total electric bill in 2022 was $6,442,444.

This matrix was created using **Month** from the Date Hierarchy and **electric\_usage\_price** from **energy\_consumption**.

A screenshot of a computer

Description automatically generated

A screenshot of a computer screen

Description automatically generated

(^ 2021)

The matrix above is from 2021. It is observed that 2022 paid ~$1.8m more on bills than 2021. This could be attributed to safe distancing measures easing up allowing more students to attend school.

**Current Capacity and Predicted Capacity area chart**

A screenshot of a computer

Description automatically generated

This area chart is a time-series analysis of daily energy usage (white), current contracted capacity (blue), and the predicted capacity (orange). It is affected by the year slicer. Using the predicted capacity, it is observed that there are months that reach 90% capacity, such as January and November.

This chart is a combination of 2 overlapping line charts since their x-axis could not be synced.

A screen shot of a computer

Description automatically generated

(^ Daily energy usage and current contracted capacity)

This chart was created using **contracted\_capacity (kW)** from **contracted\_capacity**, and **Date** and **Energy\_Consumed (kW)** from **energy\_consumption**.

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A brown and blue background

Description automatically generated

(^ Predicted capacity)

This chart was created using **date** and **predicted\_capacity (kW)** from **contracted\_capacity**.

A screenshot of a computer

Description automatically generated

Additionally, the background is set to transparent.

A screenshot of a computer

Description automatically generated

This chart helps visualise the reduced capacity shown in an earlier table. Instead of having a constant capacity (blue), the contracted capacity can be adjusted monthly (orange) to just meet energy usage which saves money. This is a similar concept to cloud computing.

## Canteen Crowd (CC)

A screenshot of a computer

Description automatically generated

**Date, hour, and canteen slicers**



The date, hour, and canteen slicers are grouped in a box. They allow the user to filter the charts by date, hour, and canteen. The box was added using the Insert tab 🡪 Shapes like Energy Consumption Overview.

The date slicer was created using **date** from **canteen\_crowd**.

A screenshot of a computer

Description automatically generated

The hour slicer was created using **hour** from **canteen\_crowd**.

A screenshot of a computer

Description automatically generated

Additionally, it was filtered to only show 7 to 18.

A screenshot of a computer

Description automatically generated

The canteen slicer was created using **canteen\_location** from **canteen\_crowd**.

A screenshot of a computer

Description automatically generated

**Average Number of Patrons During Noon clustered column chart**

A graph of different colored bars

Description automatically generated

This column chart shows the average number of patrons during noon across all canteens. It is affected by the date and canteen slicers. It is observed that B51 has the greatest number of people across all days, followed by B73, then B22.

A graph of a number of patrons

Description automatically generated

If the chart is filtered to B22, It is observed that it has the greatest number of patrons on Tuesdays and the least on Wednesdays.

This chart was created using **canteen\_location**, **dayofweek**, and **number\_of\_people** from **canteen\_crowd**.

A screenshot of a computer

Description automatically generated

This chart is unaffected by the hour slicer as it is filtered to only show values at noon.

A screenshot of a computer

Description automatically generated

This chart helps users identify the most and least busy days for each canteen and enables better schedule planning.

**Average % of Excess Patrons During Noon 100% stacked column chart**

A graph of a number of people

Description automatically generated

(^ B22)

This 100% stacked column chart shows the average percentage of excess patrons during noon across all canteens. It is affected by the date and canteen slicers. “Excess patrons” is defined as the number of people who enter after the canteen reaches full capacity. It is observed that since the canteen at B22 has a high number of patrons on Tuesdays, 12% of patrons during noon could not find seats. Conversely, since Wednesdays have a low number of patrons, all of them were able to find seats, hence no excess.

This chart was created using **dayofweek**, **excess\_people**, and **number\_of\_people** from **canteen\_crowd**.

A screenshot of a computer

Description automatically generated

This chart is unaffected by the hour slicer as it is filtered to only show values at noon.

A screenshot of a computer

Description automatically generated

Unlike the earlier column chart, this chart highlights the number of people who could not find seats.

A graph of a number of patterns

Description automatically generated

(^ B73)

In B73, it is observed that there is a greater proportion of excess patrons across all days than in B22. This could imply more people go to B73 than B22, leading to overcrowding.

**Canteen Popularity map**

A screenshot of a computer screen

Description automatically generated

This map shows the number of patrons at each canteen. It is affected by the date and hour slicers. The number of people is represented by bubble size. It is observed that at 11 am, B51 has the greatest number of patrons while B22 has the least. This could be due to more students studying near B51, B51 having more popular food stalls, or B51 having more seats. This map also shows us that despite being uphill and the furthest away from the school entrance, many people travel to B51.

This chart was created using **lat** and **long** from **canteen\_coords**, and **number\_of\_people** from **canteen\_crowd**.

A screenshot of a computer

Description automatically generated

This map helps visualise the number of people and the concentration of people.

**Average Number of Patrons per Hour line and clustered column chart**

A graph with numbers and text

Description automatically generated

This column chart shows the distribution of patrons across all canteens per hour. It is affected by the date, hour, and canteen slicers. The red line represents the seating capacity of the chosen canteen. It is observed that B22 seems to reach full capacity at 11 am and stays at full capacity for 3 hours until 2 pm. There also seem to be patrons during the early morning and evening.

This chart is created using **hour**, **number\_of\_people**, and **seating\_capacity** from **canteen\_crowd**.

A screenshot of a computer

Description automatically generated

This chart allows users to understand when rush hour starts and ends for each canteen which enables them to plan more effectively.

**Canteen Fullness matrix**

A screenshot of a computer

Description automatically generated

This matrix shows the percentage of seats left and the number of excess people at each canteen. It is affected by the date and hour slicers. The calculation for these columns can be found in the Methodology section. “Number of Excess People” is defined as the number of people who enter after the canteen reaches full capacity, or % Seats Left reaches 0%.

This chart was created using **canteen\_location**, **excess\_people**, and **seats\_left\_%** from **canteen\_crowd**.

A screenshot of a computer

Description automatically generated

Conditional formatting is used for % Seats Left. A measure is created called “condition\_colour”.

condition\_colour = SWITCH (

    TRUE (),

    SELECTEDVALUE(canteen\_crowd[seats\_left\_%]) >= 0.3, "#287E8F",

    SELECTEDVALUE(canteen\_crowd[seats\_left\_%]) < 0.3 && SELECTEDVALUE(canteen\_crowd[seats\_left\_%]) > 0, "#F6B53D",

    "#E85E76"

)

It returns green if the % Seats Left is more than 30%, orange if it is less than 30%, and red if 0%.

This conditional formatting was applied through Format your visual 🡪 Cell elements 🡪 Background color.

A screenshot of a cell phone

Description automatically generated

Format style = Field value

Apply to = Values only

What field should… = condition\_colour

A screenshot of a computer

Description automatically generated

This matrix accompanies the column chart previously and shows the number of people exceeding the capacity.

# Discussion

For easier reference, Energy Consumption Overview will be referred to as **ECO**, Energy Pricing (GST Inclusive) as **EP**, and Canteen Crowd as **CC**.

## Energy Consumption

### Question 1: I would like to know which months in 2022 had high energy consumption so that I can forecast 2023's contracted capacity.

A screenshot of a graph

Description automatically generated

Regarding the heatmap in ECO, in 2022, it is observed that April, May, July, and November had higher consumptions than other months. However, even though they had high consumption, they did not exceed 90% of the contracted capacity. This can be seen using the gauge chart in ECO. The values for the different months are changed using the month slicer.

|  |  |
| --- | --- |
| **April**  A screen shot of a computer  Description automatically generated | **May**  A screen shot of a computer  Description automatically generated |
| **July**  A screen shot of a device  Description automatically generated | **November**  A screen shot of a graph  Description automatically generated |

Since these 4 months did not exceed 90% capacity even though they had high consumption, we can forecast 2023’s contracted capacity to be lower. If May is assumed to be the highest across the entire year, its predicted capacity is around 7.5k kW (6.72 / 0.9 ≈ 7.5). Hence, 2023’s contracted capacity could be reduced by about 0.3 kW.

### Question 2: I want to know how much I can reduce 2023's contracted capacity so that I can save money.

A screenshot of a computer

Description automatically generated

Regarding the area chart in EP, instead of having a fixed capacity each month across the year, the contracted capacity can be reviewed and adjusted monthly to just meet energy demand.

A screenshot of a graph

Description automatically generated

Next, regarding the matrix for possible savings in EP, the adjustments each month could lead to a total reduction of 16,435 kW in contracted capacity amounting to a yearly savings of $200,770.

### Question 3: I want to forecast 2023's energy usage so that I know if and when the campus will reach 90% contracted capacity.

A screenshot of a computer

Description automatically generated

Regarding the area chart in EP, it is observed that certain months reach close to the predicted capacity, namely February, May, and November. Thus, the school may hit 90% capacity during February, May, and November assuming 2023’s consumption trend is the same as 2022’s.

### Question 4: I want to predict 2024's contracted capacity so that I can save money.

So far, it is not possible to accurately predict 2024’s contracted capacity due to a lack of information about 2023’s consumption trend. Despite this, it is possible to make some assumptions. If we assume the school becomes more energy efficient each year, it can further reduce its contracted capacity. If the school and especially students put in sufficient effort, this could be reduced by an additional 100 to 200 kW per month.

|  |  |  |
| --- | --- | --- |
| Month | 2023 predicted capacity (kW) (rounded to the nearest hundred) | 2024 predicted capacity (kW) (rounded to the nearest hundred) |
| Jan | 5900 | 5600 – 5700 |
| Feb | 5500 | 5400 – 5300 |
| Mar | 5200 | 5100 – 5000 |
| Apr | 7300 | 7000 – 7100 |
| May | 7500 | 7400 – 7300 |
| Jun | 7000 | 6900 – 6800 |
| Jul | 7000 | 6900 – 6800 |
| Aug | 6300 | 6200 – 6100 |
| Sep | 5300 | 5200 – 5100 |
| Oct | 6500 | 6400 – 6300 |
| Nov | 7000 | 6900 – 6800 |
| Dec | 6500 | 6400 – 6300 |

Thus, at the current rate of $14.52/kW of July 2023 – October 2023, the school could save at least $17,424 to $34,848 (GST inclusive).

## Canteen Crowd

### Question 1: I want to know which day has the greatest number of people so that solutions can be adopted to prevent overcrowding.

A graph of different colored bars

Description automatically generated

Regarding the column chart in CC, it is observed that Thursdays have the greatest number of people. A possible solution is to schedule more lessons on Mondays and Wednesdays. This would not only reduce overcrowding, but it can also reduce overloading of energy consumption, and water metering. Since the lunchtime rush is more spread out across the week, it reduces the workload of canteen vendors and school cleaners.

### Question 2: I want to know which day has the least number of people so that canteen vendors can plan how much food to prepare to reduce food wastage.

Using the same column chart above, Wednesdays have the least number of people. This reduction is partly due to students from the School of ICT having online asynchronous classes on Wednesdays. Food vendors could reduce the amount of food to prepare, and the number of cleaners could also be reduced.

### Question 3: I want to know which canteen is the most popular so that solutions could be implemented to spread out the number of people across all canteens.

For this question, the date and time will be set to Thurs 25 May 2023, at noon.

The first way of evaluating popularity is using plain numbers.

A screenshot of a map

Description automatically generated

Regarding the map chart in CC, B73 has the greatest number of people at 1498, followed by B51 with 1426, and B27 with 1009.

|  |  |
| --- | --- |
| Average number of patrons per hour (Thurs 25 May 2023) | |
| B51  A graph with numbers and text  Description automatically generated | B73  A screen shot of a graph  Description automatically generated |

Regarding the above line column charts from CC, it is observed that B73 had a bigger spike during noon as compared to B51 while the distribution of patrons in B51 is more spread out. Hence, it is implied that B73 is the most popular since it had the greatest number of patrons in an hour.

|  |  |
| --- | --- |
| Average number of patrons during noon (Thurs 25 May 2023) | |
| B51 | B73 |

However, when comparing the 2 column charts from CC, B51 seems to have 300 more people on average than B73 on Thursdays at noon. Hence from this pair of charts, it can be implied that B51 is the most popular.

The second way of evaluating popularity is with the seating capacity. B51’s seating capacity is the greatest at 1376 while B73’s is 780. Due to B51’s bigger seating capacity, more people can find seats.

|  |  |
| --- | --- |
| Average % of excess patrons during noon (Thurs 25 May 2023) | |
| B51 | B73 |

Regarding the above column charts from CC, B51 has a lower percentage of excess patrons during noon than B73 across the week. Even though Wednesdays have the least number of people, there are still excess patrons in B73.

A screenshot of a computer

Description automatically generated

Regarding the above matrix from CC, it is observed that on Thurs 25 May 2023 noon, B51 had 53 excess people/people who could not find seats while B73 had 718. Hence, B51 might be the most popular because it is more convenient to find seats.

Thus, both B51 and B73 could be the most popular depending on which evidence is given. A possible solution to reduce overcrowding is allowing patrons to order food ahead of time. This could be done using an app. Students can place orders ahead of time and choose whether to pay at the counter or pay using a card. This can be effective since most patrons know they may not find seats during noon, and they can simply grab and go.

### Question 4: I want to know which canteen is the least popular so that solutions could be implemented to attract students there.

For this question, the date and time will be set to Thurs 25 May 2023, at noon.

|  |  |
| --- | --- |
| B22  A graph of a number of patrons  Description automatically generated | A screenshot of a graph  Description automatically generated |
| A graph of a number of patrons per a number  Description automatically generated |  |

Regarding the earlier charts from CC, B22 is the lowest of all of them. B22 has 4.5 k patrons on average on Thursday noon, it has the least percentage of excess patrons and did not exceed the seating capacity.

A possible solution to attract more students to B22 is moving some stalls. Currently, B22 has several vacant lots as compared to the other canteens which are fully occupied with food stalls. Thus, moving 1 or 2 food stalls from either B51, B73, or both to B22 could spread out traffic across each canteen.

## Assumptions and Limitations

All results from ECO regarding 2023 assume that 2023’s consumption pattern remains the same as 2022’s. 2023’s consumption pattern will be different because the school had not fully relaxed its safe-distancing measures in certain months. Thus, the suggestions shown in this report can be wrong, so it is important to adapt them to fit the current context of the situation.

Without data before COVID-19, it is difficult to assess the school’s energy usage performance. For example, did energy usage increase/decrease? Regardless, more effort is needed to reduce electricity usage in 2023. The less electricity the school uses, the lower it can reduce its contracted capacity, leading to savings on both sides.

When calculating predicted capacity, some values were more than the original

There were 2 ideas when creating the map for CC. The first was to have each bubble be a pie chart to show the number of seats available and seats left as percentages to show the canteen occupancy rate. However, this was not possible as this pie chart feature only accepts binary values such as 0’s and 1’s or yeses and nos. The second idea was to have 2 sets of overlapping bubbles. This was also not possible as the map visual only allows 1 set of bubbles.

This analysis did not make use of door numbers. Perhaps a more elaborate and in-depth result could have been made using door numbers. For example, by understanding which doors are most frequently used, we can identify where patrons come from.

The solution of creating an online ordering app has some limitations. Firstly, security. Allowing patrons to store their credit/debit card information in the app poses a security risk. Secondly, costs. The school may have to invest in order-receiving devices like existing food delivery services and create the app itself.

The suggestions to move food stalls to attract patrons to less popular canteens may not be feasible because of the cost involved in hiring manpower and moving the equipment. Additionally, fewer students may study or have classes near B22, leading to the root problem of few patrons. Shifting where students study or have class is impractical as well due to the infrastructure and equipment already being in place.

## Areas for Improvement

In Methodology, for both Energy Consumption and Canteen Crowd, multiple datasets were created, transformed, and saved using the same workflow. This is inefficient because if a small change was made for one dataset, the calculations and transformations of the other datasets would run and waste time. A more efficient way is to create separate workflows for each dataset. If one dataset is needed for a calculation for another, a Text Input tool should be used.

For Energy Consumption, if it was the average hourly consumption was shown, relationships between it and Canteen Crowd could have been found.

When downloading the data for Canteen Crowd, a single macro might be more efficient than 28 sets of tools. The macro will significantly reduce the number of tools needed, make the workflow tidier, and run the workflow faster.

There is an online article about the Building and Construction Authority (BCA) Green Mark 2021 Certification Standards (Building and Construction Authority, 2023). One of the metrics described was Energy Use Intensity (EUI). EUI is defined as the total building's annual energy consumption over the gross floor area (GFA) of the building (kWh/m^2/year). If building name and their GFAs were given in the provided datasets, comparisons could be made to the BCA guide.

# Reflection

## Q1. Describe your experience working on the project, including any challenges faced and lessons learnt

In the beginning, the team had to ideate user stories/problems the product owner wants to answer. There were some difficulties because I believe we overthought and overcomplicated it.

When it came to delegating tasks, my team decided each of us will clean and visualise our datasets instead of having dedicated groups for cleaning and visualising since this was a small-scale project. However, in the real world, I believe the latter would be used.

Since Assignment 2, I have improved my dashboards. For Energy Consumption, I have added a heatmap, gauge, and energy pricing. For Canteen Crowd, I have added a map and new calculated columns. Throughout the weeks, I have changed my user stories, added new tasks, asked the product owner several questions, and used external websites and documentation to support my user stories.

When it came to calculating new columns and using external documents, it was challenging to integrate them into my existing Alteryx workflows. I had to manually create new data, make copies of my workflows and dashboard, and kept testing for the intended outcome. I also added annotations and comments to each tool in the workflow to ensure other users understand the work.

My communication with the team was growing weaker near the end due to fatigue and other concurrent assignments. Instead of checking every 2 or 3 days, it grew to every 1 or 2 weeks.

My team wanted to present something more, something that would push us from a B grade to an A. We wanted to use a website, but we found out the existing Power BI website works well enough. We also wanted a Telegram bot. We experimented with 3 different bots. One was coded through Python, another was Bot Penguin, and the third was Flow XO. We wanted the bot to be dynamic and updated along with the data. But since the data given to us was static, the bot had to be static as well. Now, users can select which user storey to view, and the bot will show screenshots of our dashboards.

We also wanted to use a different dashboard theme rather than the default one and standardise our designs and colour palette. Since I oversaw adding the dashboards to the website, everyone sent me their files and I redesigned them to match.

Overall, my challenges faced could be summed up as task management, time management, and communication.

Throughout this project, I have learnt continuous learning and development is important. Improvements are always to be made, and creating new approaches helps improve the scope, objective, and solution. Collaboration and communication are also important, not just within the team, but also with the product owner. The team needed to coordinate our efforts, and our solution must adapt to the product owners changing needs.

## Q2. Reflect on your personal growth and development in the areas of data analytics, visualisation, and project management

This project has deepened aligned my thinking more as the product owner. I am better able to empathise with them and understand their needs and wants. And by using external datasets, I could ideate more features that concern the product owner.

This project has improved my grasp of Alteryx. I understand its tools better and how to clean my data more effectively. In Power BI, I have used new features and visuals untouched before. I made references to videos and online images of “good” dashboards and improved mine accordingly.

Even though this project introduced me to new people, I was able to establish teamwork and communication quickly. Even though this project mimics real-life projects, my “colleagues” have become good friends to remember. Even though this semester had few modules, the project has taught me to better organise my time and prioritise my tasks as all projects were more in-depth than past projects.

## Q3. Identify any areas for improvement in your skills or knowledge

To improve my data cleaning process, I could have learnt more about Alteryx macros and how to integrate them into my workflow. But alas, I had insufficient time. If more meet-up sessions with the product owner were organised, more important features could be created, and my dashboard could be more comprehensive and better answer the problems faced.

When redesigning the dashboards for consistency, I noticed areas for improvement and asked my teammates to change them. If more time was taken to analyse their dashboards at an earlier time, I could have noticed them and provided help earlier.

My hard and soft skills can always be improved. I could learn more about Alteryx and Power BI, and possibly integrate other software into my data analysis. Communication could have been improved as well. I have asked a few people from other groups how their dashboards looked, but I could have asked much more from my class and other classes.

## Q4. Discuss the impact that the project had on your understanding of the topic and its relevance to real-world problems

This project gave me a taste of how data analysis projects are meant to be run in the real world. If there were no other assignments, the weekly ceremonies (stand-up meetings, retrospectives, etc.) would be done daily and the whole project would be faster. Having the product owner come in and answer our questions was much appreciated. In past assignments, the “product owners”, or “instructions” were static. On the contrary, this project was more complex and dynamic as our product owner had dynamic and changing needs. Thus, my solution had to change several times to meet them.

# Correlation to other Datasets

## Canteen Crowd and Energy Consumption

|  |  |
| --- | --- |
| CC: Average Number of Patrons During Noon (25 May 2022)  A graph of a number of patrons  Description automatically generated | EC: Average Monthly and Daily Consumption (May 2022) |

In Canteen Crowd, there was a high number of patrons on Thursdays. For Energy Consumption, it was discovered that Thursdays in May have relatively high energy consumption. These 2 datasets could be correlated based on this relationship. However, on Wednesdays, even though it had the lowest number of patrons, it had high energy consumption. Hence, this contradicts the earlier relationship.

Apart from this, it is difficult to assess any relationship between these 2 datasets. Canteen Crowd stores data in hours and Energy Consumption in days. If Energy Consumption was cleaned and hourly consumption was shown, new relationships could have been found.

## Canteen Crowd and Toilet Utilisation

|  |  |
| --- | --- |
| CC: Average Number of Patrons During Noon (25 May 2022) | Toilet Frequency: Average Monthly and Daily Consumption (25 May 2022, 12 pm)  A screenshot of a computer screen  Description automatically generated |

In Canteen Crowd, there was a high number of patrons on Thursdays. For Toilet Frequency, it was discovered that the male toilet near the B73 canteen and the male toilet near the B22 canteen were highly utilised. B73 with 55 and B22 with 30.

|  |  |
| --- | --- |
| Toilet Utilisation: Average Monthly and Daily Consumption (25 May 2022, 11 am) | Toilet Utilisation: Average Monthly and Daily Consumption (25 May 2022, 1 pm) |

For 11 am, the male and female toilets near the B73 canteen were highly utilised with 45 and 41 people respectively.

For 1 pm, the male and female toilets near the B73 canteen were highly utilised with 41 and 32 people respectively.

From above, it seems that Canteen Crowd and Toilet Frequency are related only to B73. It could be possible that toilets from the other canteens are correlated as well but are not shown in the Top 10.

Since time series analysis was not integrated into Toilet Feedback, it is difficult to ascertain relationships between Canteen Crowd and Toilet Feedback.

# Conclusion

At the time of this report, it was August 2023. In hindsight, most if not all suggestions shown may be ineffective since it was assumed they could be implemented at the start of 2023.

## Energy Consumption

Without data before COVID-19, it is difficult to assess the school’s performance. For example, did energy usage increase/decrease? Regardless, more effort is needed to reduce electricity usage in 2023. The less electricity the school uses, the lower it can reduce its contracted capacity, leading to savings on the monthly utility bill and purchasing contracted capacity.

To summarise the answers to the research questions in this report:

1. April, May, July, and November are the months to look out for in 2023 since they had high consumption in 2022.
2. 16,435 kW in contracted capacity could be reduced in 2023, amounting to a yearly savings of $200,770.
3. Using the new lower, predicted capacity, the school may hit 90% capacity during February, May, and November assuming 2023’s consumption trend is the same as 2022’s.
4. If we assume the school becomes more energy efficient each year and students put in sufficient effort to save electricity, the school could reduce its contracted capacity by an additional 100 to 200 kW per month, leading to savings between $17,424 to $34,848.

## Canteen Crowd

No data was collected before April 2023 since the sensors were newly installed. The sensors are also 70% – 80% accurate. There are occasions when patrons enter and leave in groups, leading to miscounts.

To summarise the answers to the research questions in this report:

1. Since Thursdays have the greatest number of patrons, more lessons could be scheduled on Mondays and Wednesdays to reduce overcrowding, energy consumption and water metering overloading, the workload of canteen vendors and school cleaners.
2. Since Wednesdays have the least number of people, food vendors could reduce the amount of food to prepare, and the number of cleaners could also be reduced.
3. Since B51 and B73 are popular canteens, a food ordering app could be created to reduce the number of patrons eating there and reduce waiting time.
4. Since B22 is the least popular canteen, 1 or 2 stalls could be moved to B22 to disperse the number of patrons.

More lessons could be shifted to the least busy day to reduce chances of overloading not just energy consumption, but also water metering and canteen occupancy.

# Professional Certification

A certificate of achievement with blue and white design

Description automatically generated

# Bibliography

Building and Construction Authority. (2023, March 21). *Green Mark 2021 EE Technical Guide.* Retrieved from 20230321\_energy-technical-guide\_r2.pdf: https://www1.bca.gov.sg/docs/default-source/bca-awards-2020/20230321\_energy-technical-guide\_r2.pdf?sfvrsn=aaf60fd0\_0

MapTiler Team. (n.d.). *epsg.io*. Retrieved from epsg.io: https://epsg.io/map#srs=4326&x=0.000000&y=0.000000&z=1&layer=streets

Ngee Ann Polytechnic. (n.d.). *Ngee Ann Poly People Counter V1 API*. Retrieved from Ngee Ann Poly People Counter V1 API: https://xciot.com.sg/docs/

SP Group. (2023, July 1). *Historical Electricity Tariff Oct 18 (w GST).* Retrieved from SP Group: https://www.spgroup.com.sg/wcm/connect/spgrp/e0b9800a-c39b-4f41-8664-2bbe8e2ca826/Historical%2BElectricity%2BTariff\_Oct+18%28w+GST%29.xls?MOD=AJPERES