# Queens Building Dataset

## Energy

A picture containing chart

Description automatically generated

Figure 1: Daily Energy Consumption

This is the heatmap depicting the daily electricity usage of the building with normal operating days. We can see that the usage shows a yearly gradual reduction in energy consumption and the largest jump occurs from 2020 due to the sudden outbreak of the coronavirus all the way till the 2021.

Chart

Description automatically generated

Figure 2: Hourly Energy Consumption

This is something

Chart

Description automatically generated

Figure 3: Monthly Energy Consumption

This is something

## Gas

A picture containing chart

Description automatically generated

This is something

Chart, bar chart

Description automatically generated

From the above figure we can see that gas reading is all centred around 0 and -1 with two outliers in 2019 where it spikes at 6am and drops down at 7am before it continues the normal pattern.

Chart, treemap chart

Description automatically generated

This is the monthly gas reading showing how much gas was consumed from January to December. From this we can see that the summer months (June, July and August) shows significant gas usage which is expected.

## Water

A picture containing chart

Description automatically generated

This is something

Chart

Description automatically generated

This is something

Chart

Description automatically generated

This is something

## Correlation

### Energy

### Gas

### Water

## Feature Importance

## Anomaly Detection

# Hugh Aston Building Dataset

## Energy

Chart

Description automatically generated

This is something

Chart

Description automatically generated

This is something

Chart, treemap chart

Description automatically generated

This is something

## Gas

A picture containing chart

Description automatically generated

Chart

Description automatically generated

Chart, treemap chart

Description automatically generated

## Water

Chart

Description automatically generated with medium confidence

Chart, treemap chart

Description automatically generated

Chart

Description automatically generated

## Correlation

## Feature Importance

## Anomaly Detection

# Gateway House Building Dataset

## Energy

A picture containing chart

Description automatically generated

This is something

Chart

Description automatically generated

This is something

Chart, treemap chart

Description automatically generated

This is something

## Gas

Chart

Description automatically generated

Chart, bar chart

Description automatically generated

Chart, treemap chart

Description automatically generated

## Water

Chart

Description automatically generated with medium confidence

Chart

Description automatically generated

Chart, treemap chart

Description automatically generated

## Correlation

## Feature Importance

## Anomaly Detection

# Weather Dataset

## Preprocessing

View Data

Remove null value

Encode Categorical features

Remove useless features

## Visualization

Correlation heatmap

## Feature Selection

Select highest correlated features

# Normalization

## MinMax Scaler

## Standard Scaler

## Robust Scaler

Datasets Analysis

**Queens Building**

From the correlation heatmap of all energy readings, we can see that temperature and dew point are highly correlated with each other while a negative correlation occurs between temperature and humidity.

Calculating the correlation between each feature and plotting a heatmap, gives the feature importance. An example of a negative correlation is seen between the humidity and temperature features. As the temperature rises and falls, the humidity does the opposite. High correlation is represented in a fig ?, between the temperature and dew point features being almost identical.

Heatmap pics

Negative Correlation

Positive Correlation

The next step was to use XGBoost to determine the contribution of each feature during regression on the reading(elec, water, and gas).

The most important features are (???) shown below. Table? shows the dataset feature descriptions such as the mean value, maximum, quantiles, and standard deviations. All values are of different ranges thereby making the predictive model be less accurate and requires further preprocessing.

Description screenshot

**Hugh Aston**

From the correlation heatmap of all energy readings, we can see that temperature and dew point are highly correlated with each other while a negative correlation occurs between temperature and humidity.

Calculating the correlation between each feature and plotting a heatmap, gives the feature importance. An example of a negative correlation is seen between the humidity and temperature features. As the temperature rises and falls, the humidity does the opposite. High correlation is represented in a fig ?, between the temperature and dew point features being almost identical.

Heatmap pics

Negative Correlation

Positive Correlation

The next step was to use XGBoost to determine the contribution of each feature during regression on the reading(elec, water, and gas).

The most important features are (???) shown below. Table? shows the dataset feature descriptions such as the mean value, maximum, quantiles, and standard deviations. All values are of different ranges thereby making the predictive model be less accurate and requires further preprocessing.

Description screenshot

**Gateway House**

From the correlation heatmap of all energy readings, we can see that temperature and dew point are highly correlated with each other while a negative correlation occurs between temperature and humidity.

Calculating the correlation between each feature and plotting a heatmap, gives the feature importance. An example of a negative correlation is seen between the humidity and temperature features. As the temperature rises and falls, the humidity does the opposite. High correlation is represented in a fig ?, between the temperature and dew point features being almost identical.

Heatmap pics

Negative Correlation

Positive Correlation

The next step was to use XGBoost to determine the contribution of each feature during regression on the reading(elec, water, and gas).

The most important features are (???) shown below. Table? shows the dataset feature descriptions such as the mean value, maximum, quantiles, and standard deviations. All values are of different ranges thereby making the predictive model be less accurate and requires further preprocessing.

Description screenshot