Whirlpool

Reach Consulting Group

April 15th, 2021

rcg

Client:
● Whirlpool | ReNEWW House

Consultants:
Reach Consulting Group
- Tej Seth - Project Manager
- Jason Weisenfeld - Client Liaison
- Dani Lazarus - Analyst
- Logan Roach - Analyst
- Oliver Schaaf - Analyst
- Zachary Marmet - Analyst
- Jennani Jayaram - Advisor

Duration:

● Initial meeting: March 2nd, 2021
● Final Presentation: April 15th, 2021

About Reach Consulting Group:

Reach Consulting Group is a student-run organization providing advisory services in technology, strategy, finance, and management. We are affiliated with the University of Michigan's School of Information and have a strong emphasis on technology consulting.

Reach exists to both train and familiarize students to solve critical issues that real companies experience, while providing free service for companies. Our unique approach is to tackle problems with a combination of advanced business analysis and technical expertise in many different fields. We provide a broad range of services and solutions to help organizations facilitate change, achieve their vision and optimize performance and productivity.

Project Objective:

Based on the data that was provided and some of the areas Andrew Batek thought were worth looking into, we decided on three main objectives to analyze within the ReNEWW house:

1. Using the air quality data, we will attempt to predict the different variables that change air quality within the house and suggest where new renewable houses can be built based on these factors.
2. Using the temperature and humidity data, we will provide analysis that shows the variables that make up surpluses and droughts for water usage.
3. Using the electrical data, we can analyze the different occurrences instead and outside the house

that affect the watt usage.

The result of these objectives is represented in the form of a presentation and a deliverable.

Project Procedure:

1. Reach Consulting Group met with Andrew Batek at an initial meeting where he gave a presentation on Whirlpool's ReNEWW house.
2. Andrew provided the datasets to the group, including a spreadsheet with explanations of each dataset.
3. The group started with Early Data Analysis (EDA) of each dataset and then split into groups and took on a goal within the dataset.
4. Each mini-group of two people started their analysis on their dataset. This included cleaning the data, filtering out outliers, and taking notes of weird behavior inside each set.
5. Based on the weird behavior found, the group had a check-in with Andrew halfway through the project and he explained how the different keys operated in more detail.
6. With new information, each mini-group went back to work and started working on their graphs, tables and overall analysis.
7. The team put together a final report by compiling, visualizing, and assessing the resulting analyses.

Methods:

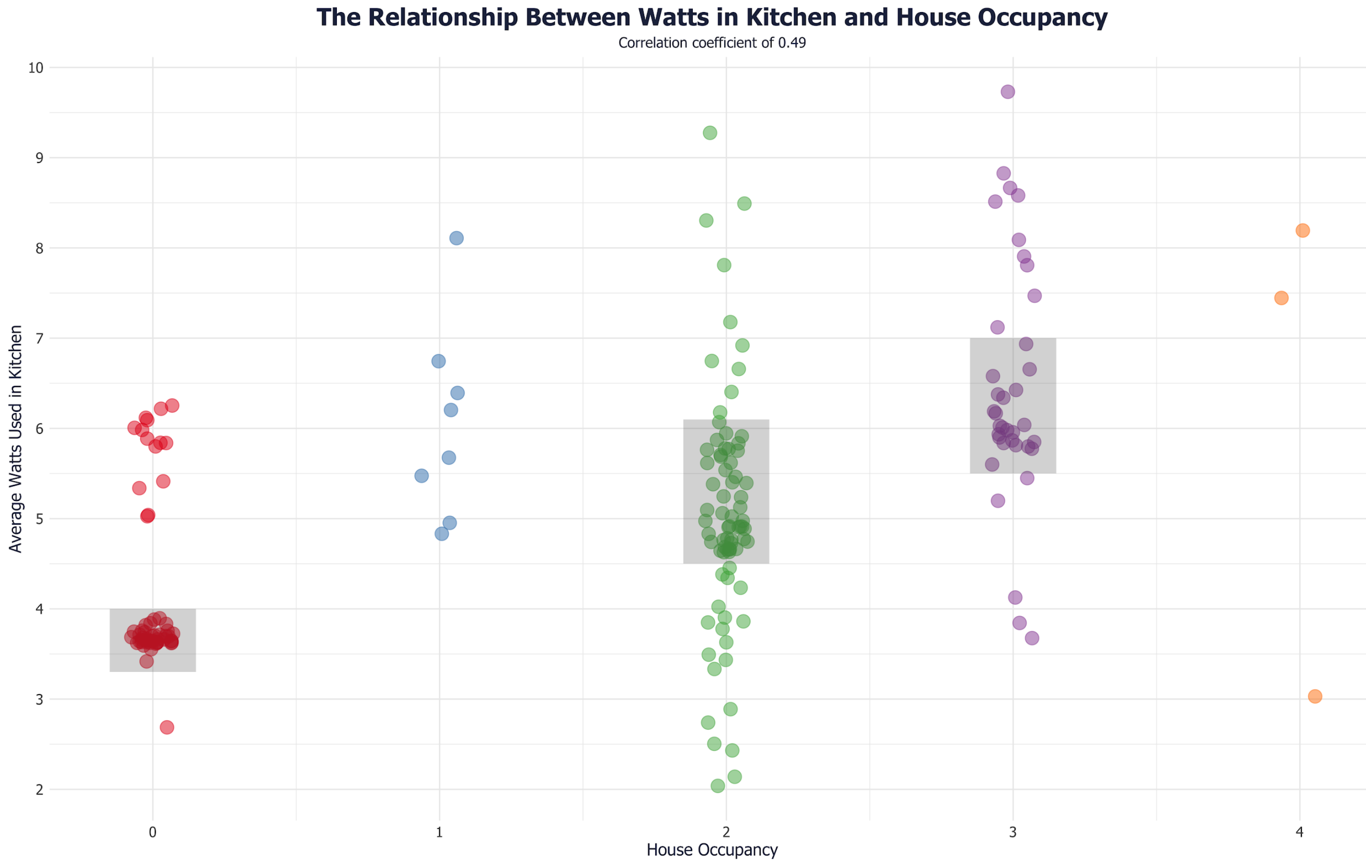
The data was uploaded to a shared Google Drive by Andrew and then forked over to our own Google Drive where we could organize the CSV files. Each member of the group downloaded the CSV files they needed and uploaded them into R. The data was analyzed using the tidyverse and dplyr packages, plotted using the ggplot2 package and made into tables using the gt package. Linear regression was also used on R to analyze different factors that contributed to the variables we were looking at.

After those steps were completed, the findings were compiled into a general area and recommendations were able to be made based on those findings. The rest of the analysis was put into this report and the presentation done on April 15th, 2021.

Findings:

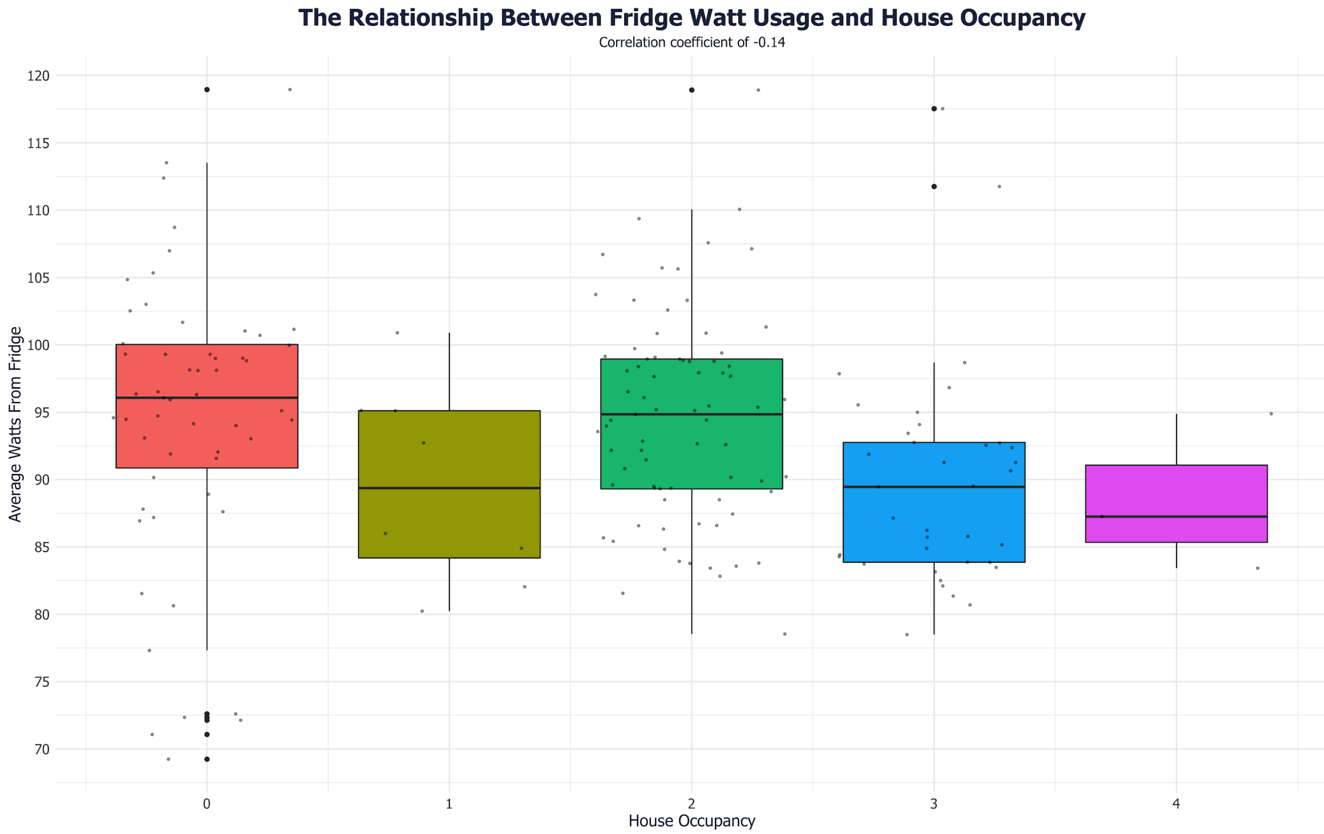
Electrical Energy Usage

One of the variables that our group analyzed was energy consumption inside the ReNEWW house and how the number of watts compared to a variety of factors. The first graph looks at watt usage in specifically the kitchen of the house compared to house occupancy:



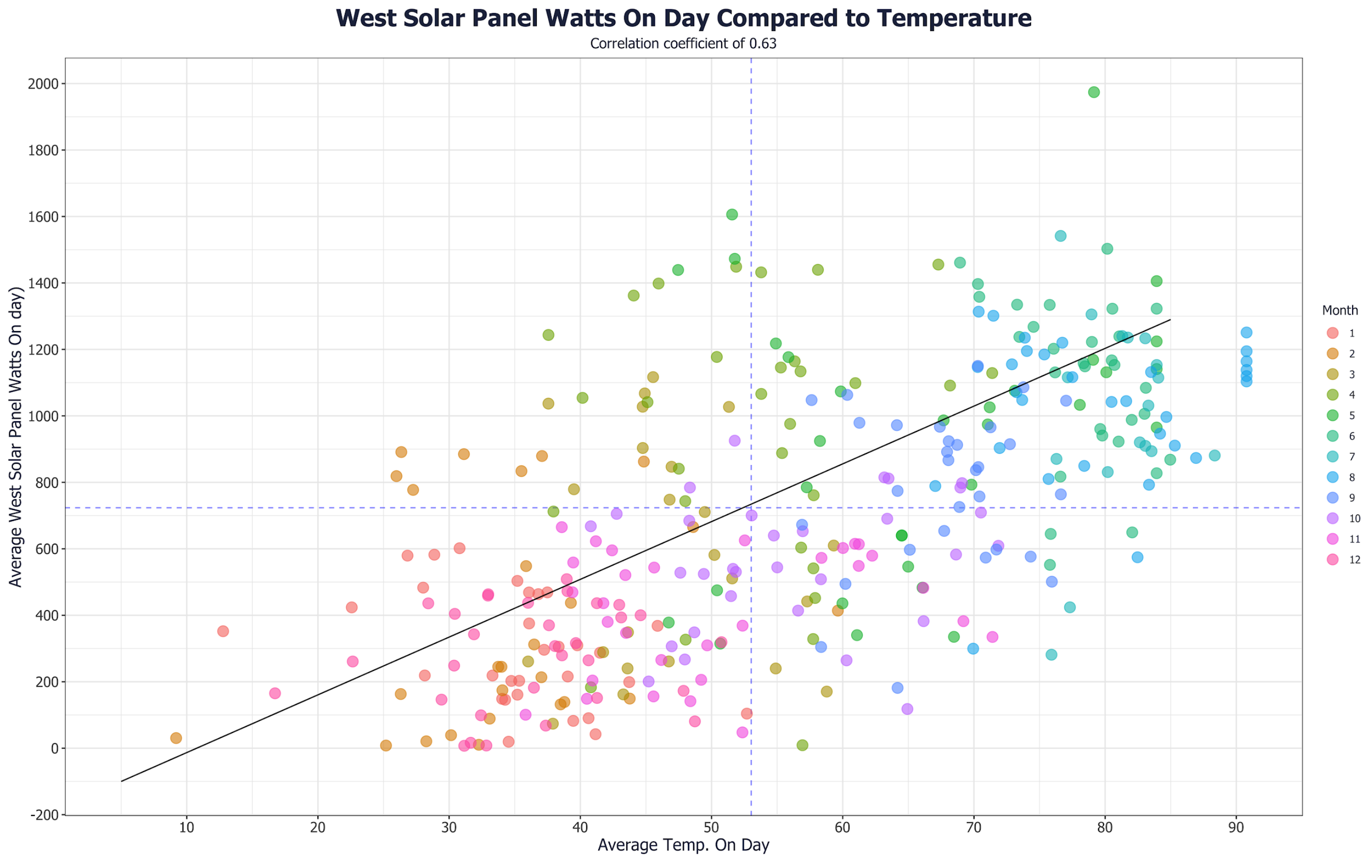
When looking at the relationship between the average watts used in the kitchen on each day and the house occupancy, we found that there was a strong correlation between the occupancy increasing and average kitchen watts also increasing. The shaded boxes represent where watt usage was the most dense for each house occupancy. We can see a density of about 3.75 watts when zero people were in the house, around 5.50 watts when two people were and all the way up to 6.50 watts when there were three people.

Although the fridge is located in the kitchen, there was a different effect when compared to house occupancy:



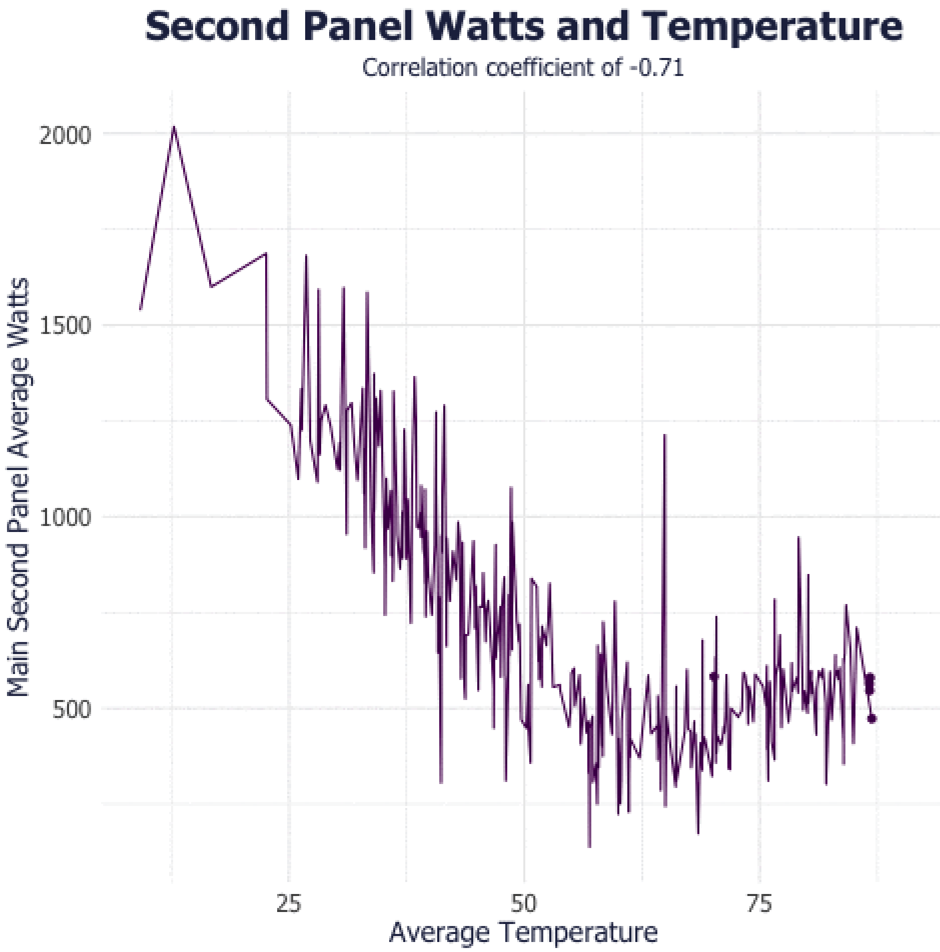
One of the most interesting findings was that there was no relationship between house occupancy and watts from the fridge. In most of our other findings, house occupancy was a big factor for watt usage but when it came to the fridge, there was actually a slight negative correlation. Because of that, we do not believe the occupants in the house have to be as conscious of fridge energy usage as of other energy-use appliances.

While it is important for members inside the ReNEWW house to understand areas where they use energy, it's always important to understand which factors contribute to energy coming into the house:

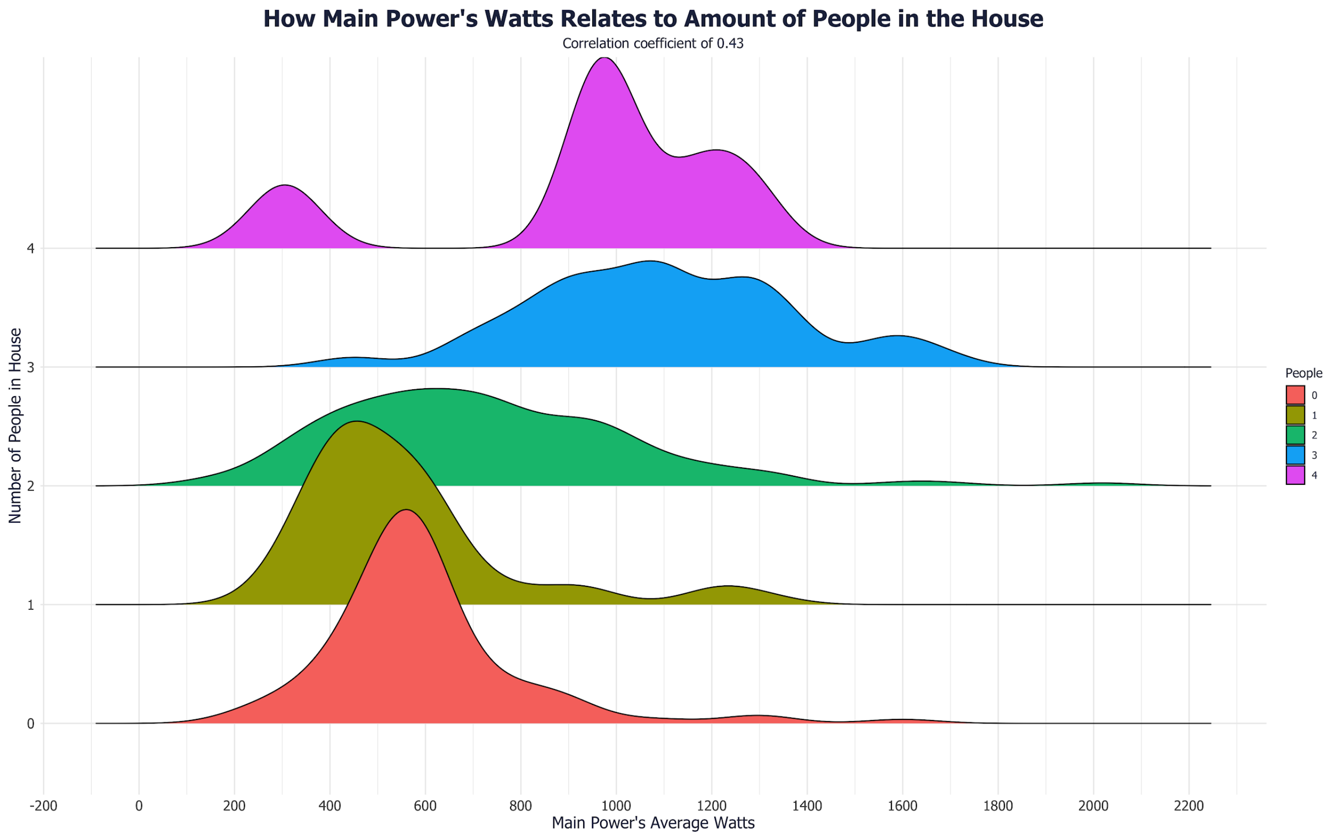


Taking a look into how the average temperature of a day affects the watts on a solar panel, we find a very strong correlation. By binning by month, we can see that warmer months like September and August provided the most watts for the solar panels with temperatures in the 70's and 80's. Colder months like December and January had temperatures in the 30's and didn't provide as much energy for the solar panels.

The solar panels weren't the only thing temperature had an impact on. We also found that temperature impacted the main second panel usage:



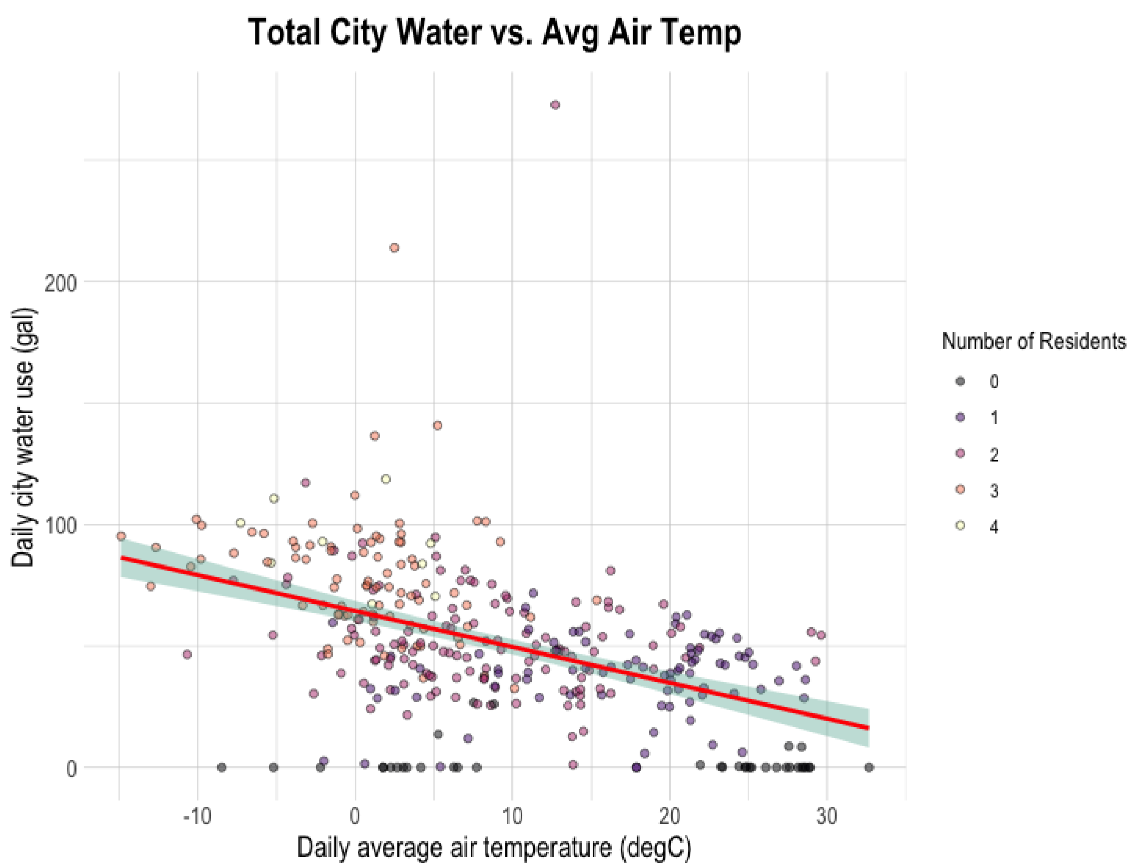
There was a strong negative relationship: as the temperature increased, the average watts decreased. However, it should also be noted that we found days with higher temperatures tend to have had fewer people in the house so that could have played a factor as well. When looking closer, house occupancy had a huge influence on the main power strip's wattage count:



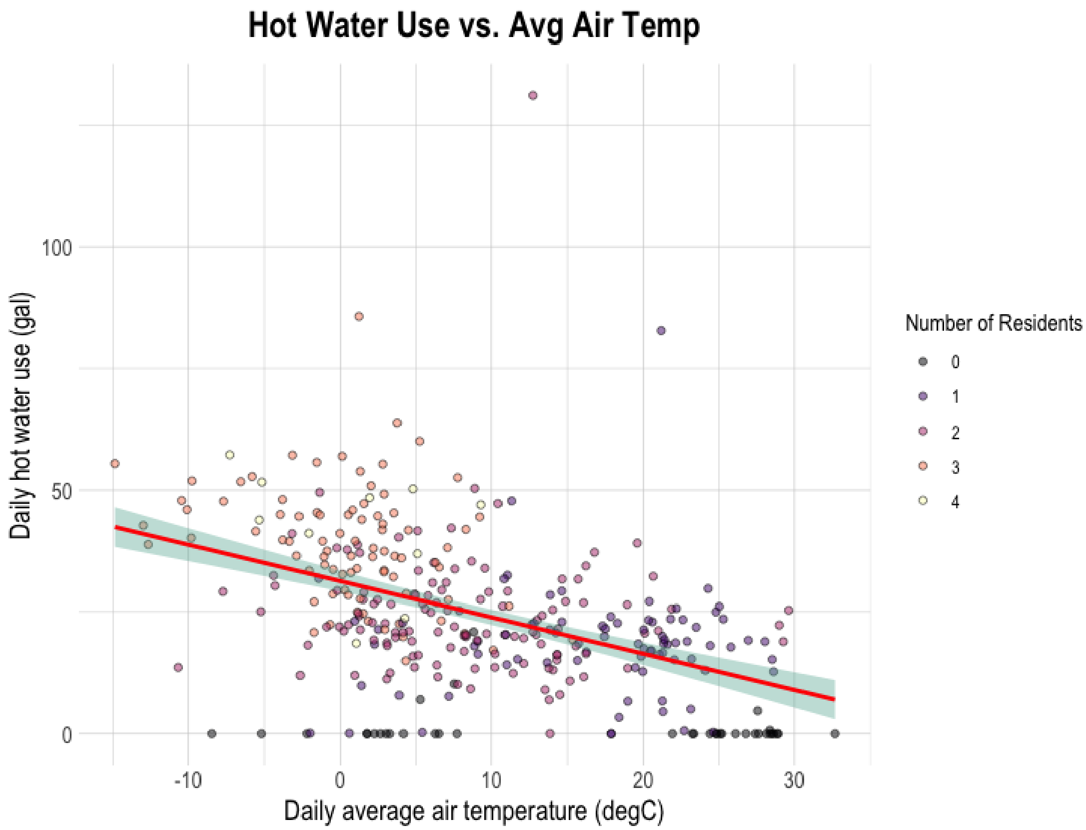
In these ridges with house occupancy being on the y-axis and main power's average watts being on the x-axis, it is clear that house occupancy affects how much energy the house is using. This confirms our hypothesis. It was really interesting to find that when there's 0 and 1 people in the house, the range of outcomes is much smaller than when they are 2, 3, and 4 people. More people in the house seem to contribute more to variance and to average energy usage.

Water Flow and Water Usage

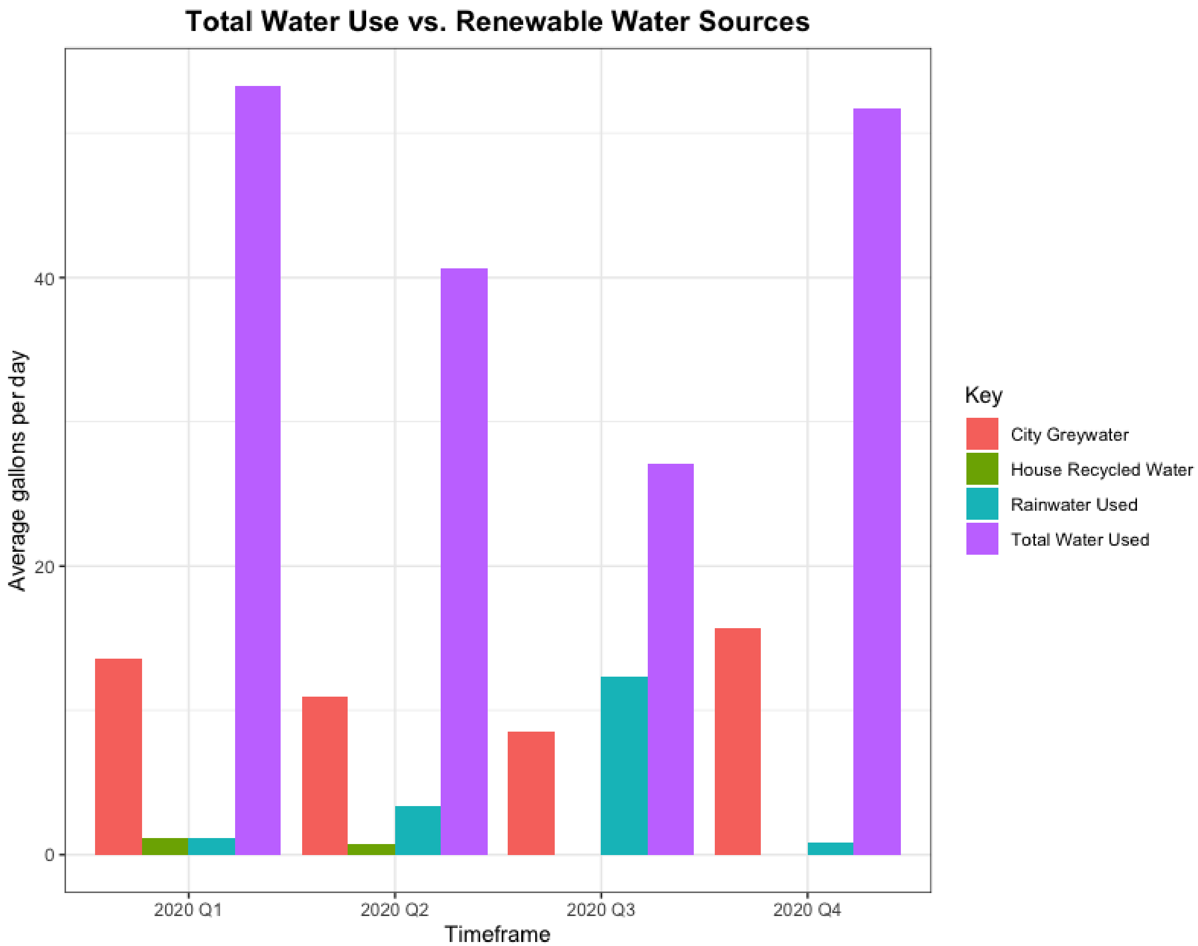
There was additional ReNEWW house information that was gathered from the waterflow, temperature, and humidity spreadsheets. Based on the findings we were able to put comparisons into graph form:



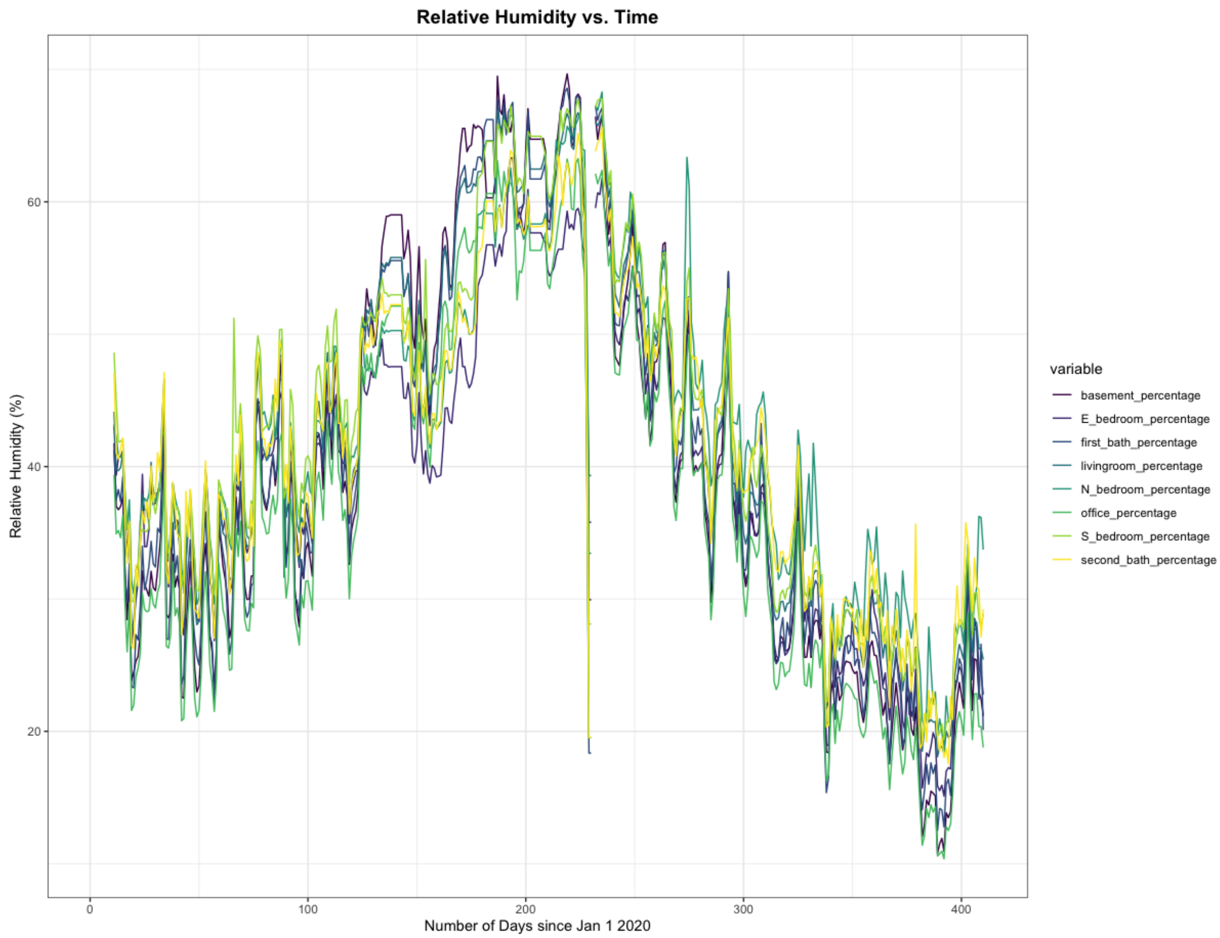
The first comparison we drew was between daily overall water use (measured by intake from the city) and average air temperature. We expected that water use would increase in the summer months, but were surprised to find a strong negative relationship between usage and air temperature instead. However, we then adjusted for a third variable—house occupancy—and found that the number of residents was much higher during fall, winter, and spring. Suspecting that the house occupancy may be skewing the results, we ran a multiple linear regression model with daily city water use as the response and house occupancy and daily average air temperature as possible predictors. The variable for house occupancy had a significant p-value less than $2.0*10^{-16}$ while the variable for outdoor air temperature had a non-significant p-value of 0.4973. As a result, we concluded that total water use is highly dependent on the number of residents and not other external factors. We additionally concluded that the large seasonal adjustment in the number of residents is an important factor to keep in mind when drawing other comparisons moving forward.



We then drew a second comparison using outdoor air temperature, but this time we focused only on hot water use. We measured this using the WL_1000 variable, which accounts for the sum of all hot water sources in the house by tracking the cold water inlet to the left solar water heat tank. Due to the results in the previous graph, we adjusted for house occupancy once again. We found another negative relationship as expected, but the trendline this time was slightly steeper. After running the multiple linear regression, we found that the number of residents still had a strongly significant p-value, but the daily hot water use had a significant p-value as well. This meant that residents are more likely to use increasing amounts of hot water as the temperature outside gets colder.



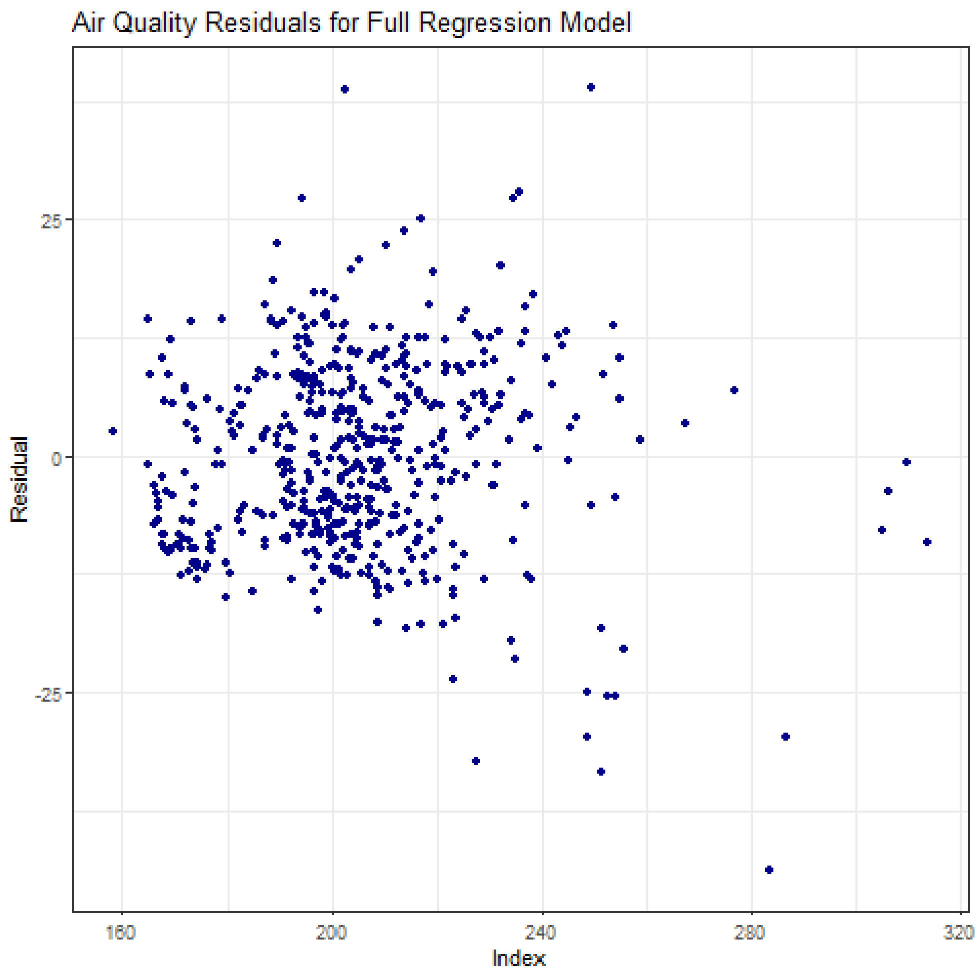
Next, we wanted to look at the different types of water split up by quarter in 2020. We were able to use some of the aggregated keys such as WL_1000 to make a larger comparison. We found that there was a lot of potential for improvement in recycling systems. Additionally, house recycled water is the most underutilized. Furthermore, in Q3 and Q4 of 2020, house recycled water seemed sparsely working and therefore the use is near zero. It should again be noted that the change in total water use from quarter to quarter is mostly due to the house occupancy and not necessarily the quarter itself.



Another variable that we looked into was relative humidity for every sensor in the house over the course of one year. In the graph above, relative humidity is shown for every day during the year 2020. We found that the kitchen sensor read inconsistent data multiple times throughout the year and had to be filtered out. It was reading extremely low humidity (roughly 0%) for about two months during the second quarter. We believe this might be a sensor error as it does not follow the general trend of all other rooms in the house. Once accounting for that, there was still a clear increase in humidity during the summer months.

Air Quality

Indoor air quality is a very important metric for modern homes, as fresh, unpolluted air inside a home is essential to protecting the health of inhabitants. Ensuring good air quality involves multiple ventilation and filtering processes, which in turn require reliable sources of power. For a home like the Whirlpool ReNEWW house, air quality is a very telling statistic for energy consumption, as more power will likely be used on days with worse air qualities. We explored this air quality metric, finding a solid method for predicting, as well as using it to help identify individual days to look into further.



The graph above shows the residual for air quality based on a full linear regression model. This model incorporates variables such as averages on indoor temperature, air pressure, CO_2 levels, relative humidity, TVOC, and PM2.5, and whether the day is in the winter or on a weekend as predictors. As seen above, it is a very serviceable model, minimizing the Actual - Predicted differentiable on either side.

Below, the coefficients and significance of each predictor can be seen. Further, a coefficient of determination of approximately 83% is achieved. This means 83% of the distribution of air quality can be explained by this multiple regression model, which is extremely promising.

```
Coefficients:
              Estimate Std. Error t value            Pr(>|t|)    
(Intercept)  51.444320  71.108229   0.723              0.4697    
mean_temp    -1.245060   0.169118  -7.362   0.000000000000704 ***
mean_airpres  0.163249   0.071073   2.297              0.0220 *  
mean_co2      0.078095   0.003309  23.602 < 0.0000000000000002 ***
mean_RH       0.231762   0.057090   4.060   0.000056648028992 ***
mean_TVOC     0.116924   0.004632  25.243 < 0.0000000000000002 ***
mean_PM2.5    1.207936   0.082167  14.701 < 0.0000000000000002 ***
isWeekend    -1.446443   0.973350  -1.486              0.1379    
isWinter      1.987249   1.203398   1.651              0.0993 .  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 10.18 on 525 degrees of freedom
  (13 observations deleted due to missingness)
Multiple R-squared:  0.8318,    Adjusted R-squared:  0.8293 
F-statistic: 324.6 on 8 and 525 DF,  p-value: < 0.00000000000000022
```

When tested on random data not included in the configuration of the model, the model performed with a maximum error of 15%. This is very good for a model with several variables and a degree of complexity where over-fitting could be an issue. Balancing complexity with performance is the priority when creating a model like this. Below is part of the dataset used for this model. When using all these variables, even when standardized and normalized, it presents a larger opportunity for variance to skew the resulting model. For that reason, identifying the more determinant predictors is valuable.

Air Quality Statistics

At Whirlpool RENEWW House

Date	Mean Temperature	Mean Air Pressure	Mean CO2	Mean Air Quality Value	Mean RH	Mean TVOC	Mean PM2.5	Weekend?	Winter?
	72.74	996.00	699.67	NA	37.31	57.18	4.07	NA	NA
1/1/2020	62.46	986.89	646.80	207.97	35.15	26.05	7.36	0	1
1/1/2021	70.00	990.27	712.00	193.12	30.16	19.50	1.00	0	1
1/10/2020	67.89	995.47	929.20	252.56	39.68	129.24	8.17	0	1
1/10/2021	65.77	1,004.43	791.00	201.28	26.57	84.50	1.00	1	0
1/11/2020	68.67	984.84	846.88	230.47	41.56	66.32	8.95	1	1
1/11/2021	72.60	1,003.67	766.50	198.52	26.60	33.50	1.25	0	0
1/12/2020	68.72	1,003.05	883.36	255.80	36.01	234.67	7.44	1	1
1/12/2021	68.80	998.50	768.50	195.46	25.80	60.00	1.17	0	0

The blue column above is the variable predicted by our model, air quality, and the green columns (Average Temperature, Air Pressure, Relative Humidity, and PM2.5) are the most relevant predictors.

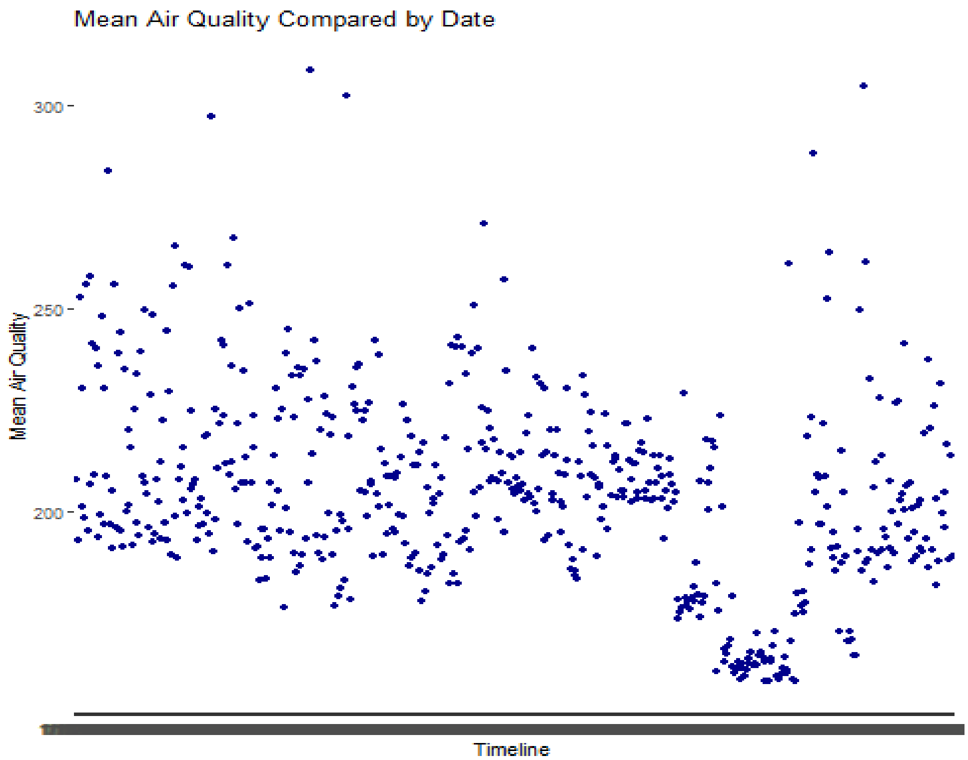
When applying LASSO (Least Absolute Shrinkage and Selection Operator) Regression to the model, these four were found to be the most important. LASSO, through a process of reducing coefficients of variables, shrunk the weight of every variable to zero except these four, both simplifying the model and making it more applicable, as well as revealing areas within the ReNEWW House that deserve more focus.

```
mean_temp       -1.22163699
mean_airpres     .
mean_co2         0.02418620
mean_RH         -0.07195566
mean_TVOC        .
mean_PM2.5       0.17261362
isWeekend        .
isWinter         .
```

Another valuable use of the LASSO model is applying it to dates. Each data point is a unique date, so the model's choice of which dates to reduce and which to assign a coefficient gives insight into outlier and anomalous data points. As seen below, certain individual dates require specific assistance for the model. How this can be interpreted is the variable "date1/2/2020" will be represented as $37.04425347*x_{1/2/2020}$ in the regression model. $x_{1/2/2020}$ will only equal one when the date is 1/2/2020, so the coefficient of 37.04425347 will only be added to predicted air quality on that date. For a date like 1/19/2021, no extra amount will be added or subtracted to air quality depending on if it is that date or not.

```
date1/15/2020    .
date1/15/2021    0.61235365
date1/16/2020    .
date1/16/2021    .
date1/17/2020    .
date1/17/2021    .
date1/18/2020    .
date1/19/2020    .
date1/19/2021    .
date1/2/2020    37.04425347
date1/2/2021     .
date1/20/2020    .
date1/20/2021    .
date1/21/2020    7.98635703
date1/21/2021    .
date1/22/2020    .
```

This holds little to no value in terms of a predictive model, but it does help highlight days that should be looked into for specific differences causing a severe difference in air quality. Depending on the day, month, and season, air quality will vary, but that can be accounted for in a model as such. Identifying a previously unknown factor and then accounting for it will only improve the day-to-day use and maintenance of the ReNEWW House, especially regarding energy consumption. As seen in this graph, there is a clear relationship between air quality and time of the year. The LASSO method above gives a clear and quantifiable method for identifying these deviants.



Recommendations:

Electrical Energy Usage

Based on our findings and analysis, we can give the following advice to the residents of the ReNEWW house in regards to their energy consumption:

- Residents should be aware of the house occupancy in regards to energy consumption in the kitchen. This, however, does not apply specifically to the fridge as the fridge's energy stays stagnant no matter how many people are in the house.
- When living in the ReNEWW house, it is important to be aware of when there will be energy surges and droughts from the solar panels based on temperature. If it isn't already displayed on the Magic Mirror, the temperature should be listed as well as a projected amount of watts taken in by the solar panels for the day.

- When it comes to the main power panel, we believe that it is again house occupancy that plays the biggest role so residents should be aware of that. Temperature seems to play some role however the results aren't conclusive

Water Flow and Water Usage

Using our analysis of the temperature, humidity, water flow, and water usage data, we are able to provide the following advice to the residents of the ReNEWW house and Whirlpool:

- Residents of the ReNEWW should be cognizant of their hot water usage during the winter months. There is a clear increase in usage and residents could be advised to take shorter showers or wash their clothes with cold water, rather than hot water.
- Whirlpool should look to install more greywater recycling systems in the ReNEWW house so that the residents can provide more renewable water for themselves. This would save the city greywater for other homes that don't have these built-in facilities or enough residents to generate sufficient greywater.
- Further investigations into humidity differences between rooms may cue Whirlpool to differences in insulation between rooms. Rooms with less functional insulation could be cause for energy being wasted to heat/cool the home. Additionally, Whirlpool should look into potential damages to their kitchen humidity sensor. There was a large quantity of inconsistent data that prevented us from performing an effective analysis of the kitchen humidity.

Air Quality

Air quality is an important measurement for the ReNEWW House, as it affects the health of the inhabitants and the amount of energy used within the house. After exploring this statistic, there are a few key points to focus on in regard to air quality:

- Indoor temperature, CO_2 levels, relative humidity, and PM2.5 levels are the most important factors in air quality. If air quality is an area that the Whirlpool team would like to focus on improving, these areas will be the most effective in doing so.
- When CO_2 and PM2.5 levels increase, it increases the air quality statistic. While indoor temperature and relative humidity increase, it decreases the air quality statistic. It is valuable to have found predictors that are directly and indirectly related to our target variable. Besides knowing what is good for air quality and what is bad for it, having variables that do both strengthens our model.
- Outliers can offer information about a certain statistic. Identifying days with very unexpected air qualities then identifying what factors contributed to that difference could locate previously unknown ways that energy is being wasted or overused.

Conclusion

After weeks of analyzing data sets and comparing findings with each other as a group, we were able to arrive at a plethora of suggestions for the residents of the ReNEWW house. We hope that our recommendations will be helpful as Whirlpool looks to develop more of these houses across the United States.

Reach Consulting Group has been grateful for the opportunity to provide Whirlpool with our findings and recommendations. Our members have gained a wide variety of skills and knowledge ranging from time management, teamwork, communicating with clients and understanding client needs, to utilizing RStudio to conduct educated analysis of large datasets. We appreciate all the resources Andrew and the rest of the team has provided Reach in this journey, as well as the exceptional feedback and support.