PALS: Energy Data Analysis

Dylan Cathcart, Thitna Gruga, Do Yun Kim, Iskander Lou, Dagmawi Solomon, and Zhaojie Yin

Table of Content

Overview

The PG Parks and Recreation department is responsible for operating and maintaining facilities at local and regional parks providing recreational services and programs. This enables members of the PG community to be engaged, entertained, learn new skills, and feel inclusive within their community. Their main goal, according to their website, is to "preserve, enhance, and protect [county's] open spaces to enrich the quality of life for present and future generations in a safe and secure environment." The PG Parks and Recreation department is a very successful organization that provides quality services, having won six national gold medals for excellence in the past years.

Providing recreational services requires capital investments into building facilities and maintaining parks throughout the Prince George's County area resulting in the use of vast amounts of energy and resources to accomplish these goals. It is our job to help the department manage its expenditures and energy usage while still enabling them to fulfill their mission and commitment to the County. We will accomplish this by analyzing their energy dataset and providing information driven guidance on the future sustainability of facilities.

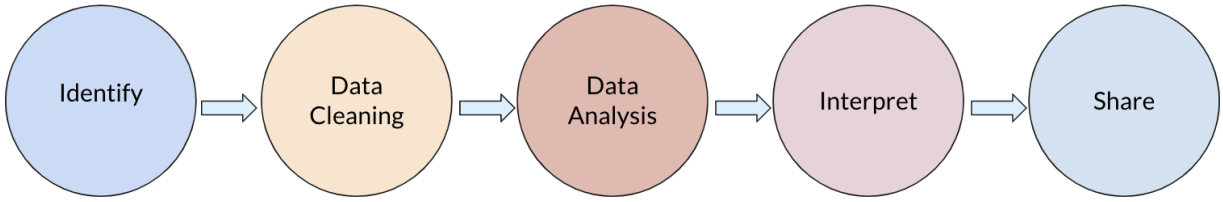
Throughout our data analysis and visualization process, we will be referring to the Energy Use Intensity (EUI) metric, which represents the energy efficiency of a building. EUI is calculated as the energy output of a building divided by area in square feet of the building. There are two types of EUI: source and site. Site EUI is the total energy consumption of a building. Source EUI is the total energy consumption of a building in addition to the energy generated at the power plants and any losses in the delivery of a building. Another abbreviation that we will refer to is EPA, Environmental Protection Agency. We will use the agency's recommended EUI to compare the energy efficiency of the PG county buildings to the national recommended level.

Project Pipeline

We decided to make a pipeline that reflects all the methodology in approaching this project. For technical projects, it is critical to have a project framework because there are a plethora of steps. By identifying and breaking down the project, we were able to have more clear flow of the project and have more organized procedures.

The figure below is the project pipeline that consists of five phases of this project.

Project Pipeline



Identify the Project

The first phase of the project was Identify. In this phase, we aimed to understand the project by researching and interacting with the client. As it is paramount to understand the aim and objectives of the project from the client, we asked many questions to clarify the goal. Although our team had experiences with data analysis with different fields, it was the first time for all the members to analyze data of energy. We prioritized having a solid understanding about energy consumption and measurements by looking at the United States Environmental Protection Agency(EPA).

In this phase, we elaborated on the project by asking many questions. These were the questionnaires that we have created to guide the project to have the right deliverables.

1. What are the metrics/measurements that our client wants to look at?
2. What is the goal and objective of the project?
3. How often does a client want to have a meeting?
4. What are our given resources for the project?
5. What deliverables do our client want from the project?
6. Are there any preferable tools/softwares for the project?
7. What do we have to know about “energy”?
8. What kind of facilities are we going to analyze?
9. What are the factors that cause a high amount of native use of the energy?

Data Cleaning

We received the utility data in excel format with thousand sheets for 275 facilities. Data cleaning aims to support and enhance the data analysis process and generate accurate conclusions to present the current issue hidden under the datasets. We utilized Python (Pandas) and Openrefines as our main techniques to empower our data cleaning process. As the data cleaning process runes, we identified numerous data quality issues and divided them into five categories, mainly as below:

Missing Data:

- **Issue description:** Many rows lack native use data, which might mislead EUI calculation. Cost data lost might result in inaccurate EUI leading financial loss calculation. Most importantly, gas-relevant data are lacking for most years, which might cause extreme EUI data situations.

Days	Native Use	Cost	Demand	Common Use
30		83.43		
30		474.55		
29		331.37		
13		222.76		
28		871.59		
27		222.76		

- **Solution:** Remove all the rows without native use and cost records

Facilities Names standardization and redundant facility data:

- **Issue description:** We aim to analyze the energy efficiency of each building rather than the specific center nested inside the building. There are many records named with sub-centers. Besides, the project focuses on the six particular buildings and aim to highlight the energy efficacy. Any building out of the six buildings' scope and columns such as account code coincided as redundant information. Those data will increase the complexity of the analysis process and data misusage.

B	Southern Tech&Rec Complex/Athletic Center
B	Southern Tech&Rec Complex/Athletic Center
B	Southern Tech&Rec Complex/Athletic Center
B	Southern Tech&Rec Complex/Athletic Center
B	Southern Tech&Rec Complex/Athletic Center
B	Southern Tech&Rec Complex/Athletic Center
B	Southern Tech&Rec Complex/Athletic Center
B	Southern Tech & Rec Complex / Wellness Center
B	Southern Tech & Rec Complex / Wellness Center
B	Southern Tech & Rec Complex / Wellness Center
B	Southern Tech & Rec Complex / Wellness Center

- **Solution:** Unify all naming to the building level and remove redundant records/columns.

:

Inaccurate data:

- **Issue description:** We found that Unusual data records appeared in rows in the data inspection in the data inspection. It can be summarized as Negative Cost, Cost without utility, and .usage without cost. Those records are considered outliers that affect the accuracy of the conclusions.

Days	Native Use	Cost
33	348	379.66
28	299	362.33
28	270	325.48
30	226	264.01
33	222	273.37
28	181	202.2
29	132	155.86
32	0	134.61

Days	Native Use	Cost	Demand	Common Use	ay Amoun
33	86	0		8.349482	0

Days	Native Use	Cost	Demand	Common Use
191	-9	0		-9

- **Solution:** Remove all the rows without the issues described above.

Data Analysis

We use formulas to calculate EUI for gas and electricity to present energy efficiency. Below are the formulas:

- Electric: Native Use/ Facility Square Foots*3.412
- Gas: Native Use/ Facility Square Foots*99.9761

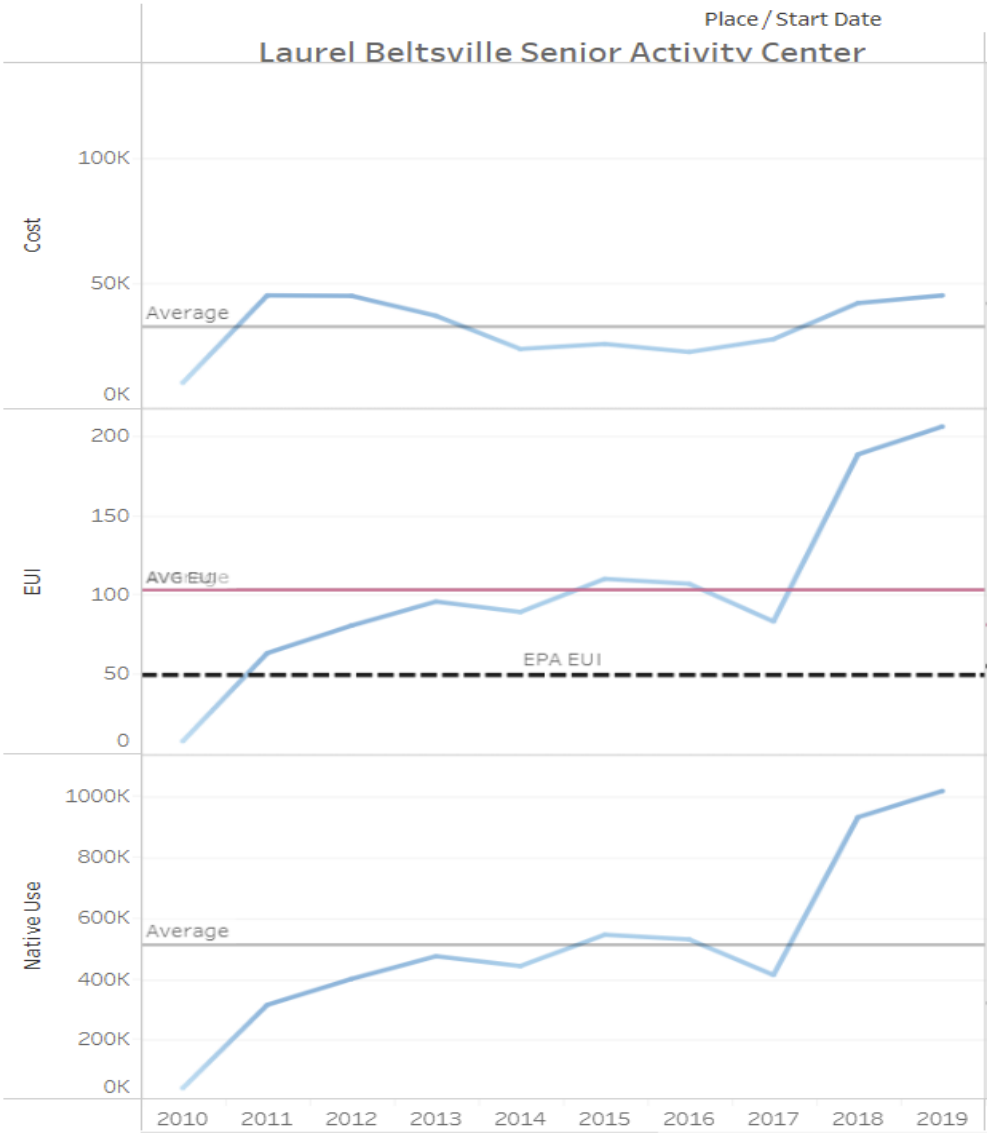
In terms of illustrating cost-saving prediction, we generate a formula to fulfill the calculation as below:

- Cost * (Actual Facility EUI - EPA recommended EUI) / Actual Facility EUI

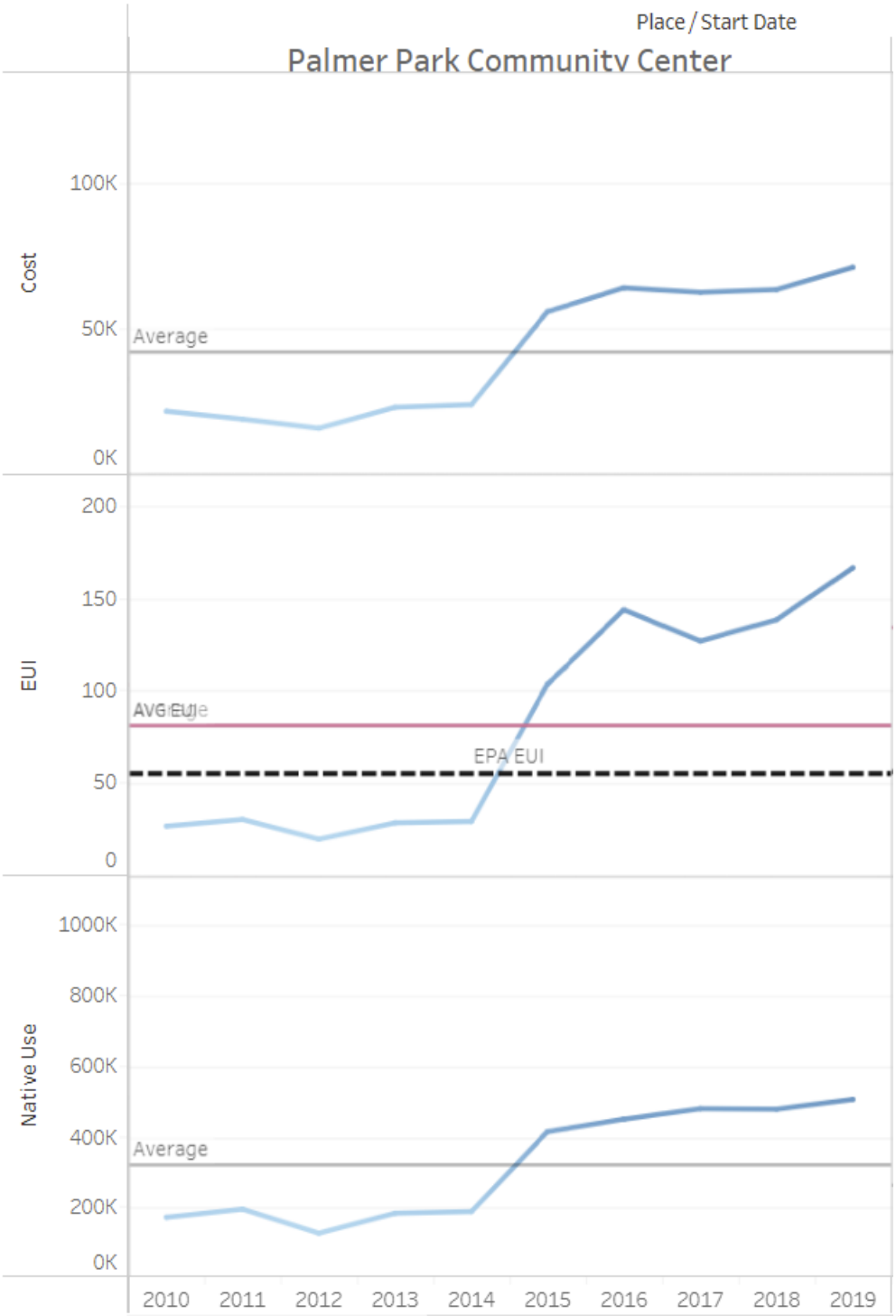
Data Visualizations

Link to Tableau Public:
https://public.tableau.com/app/profile/dylan.cathcart/viz/Book2_16517785490590/EUIANALYSIS

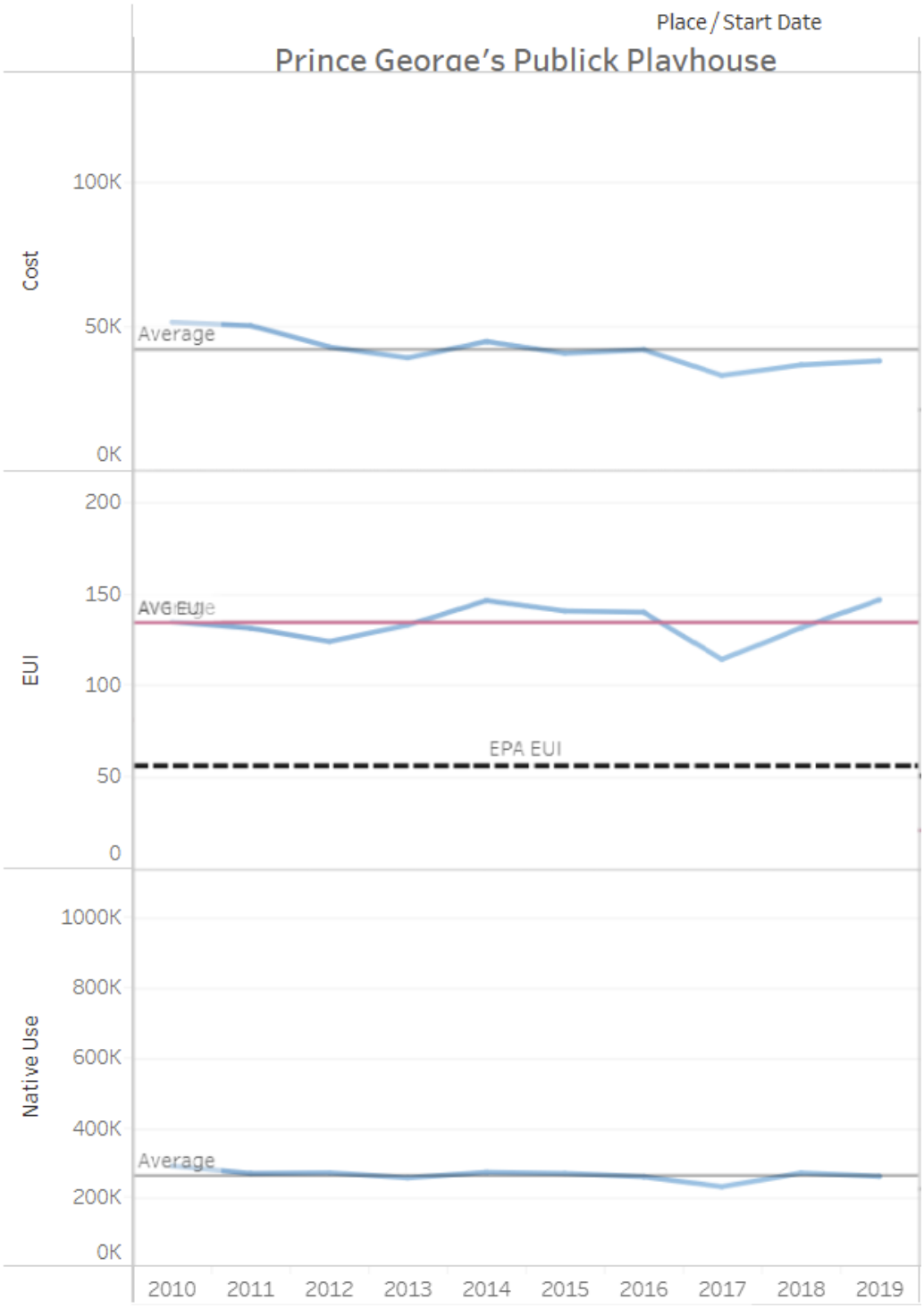
We used Tableau to visualize the data. Tableau is a great resource for creating graphics and charts to understand the data problems and also to take a closer look at the data. Here are some examples and explanations for some of our most telling visualizations. Here we will highlight some of our visualizations. In the Tableau workbook, there are more visualizations to explore.



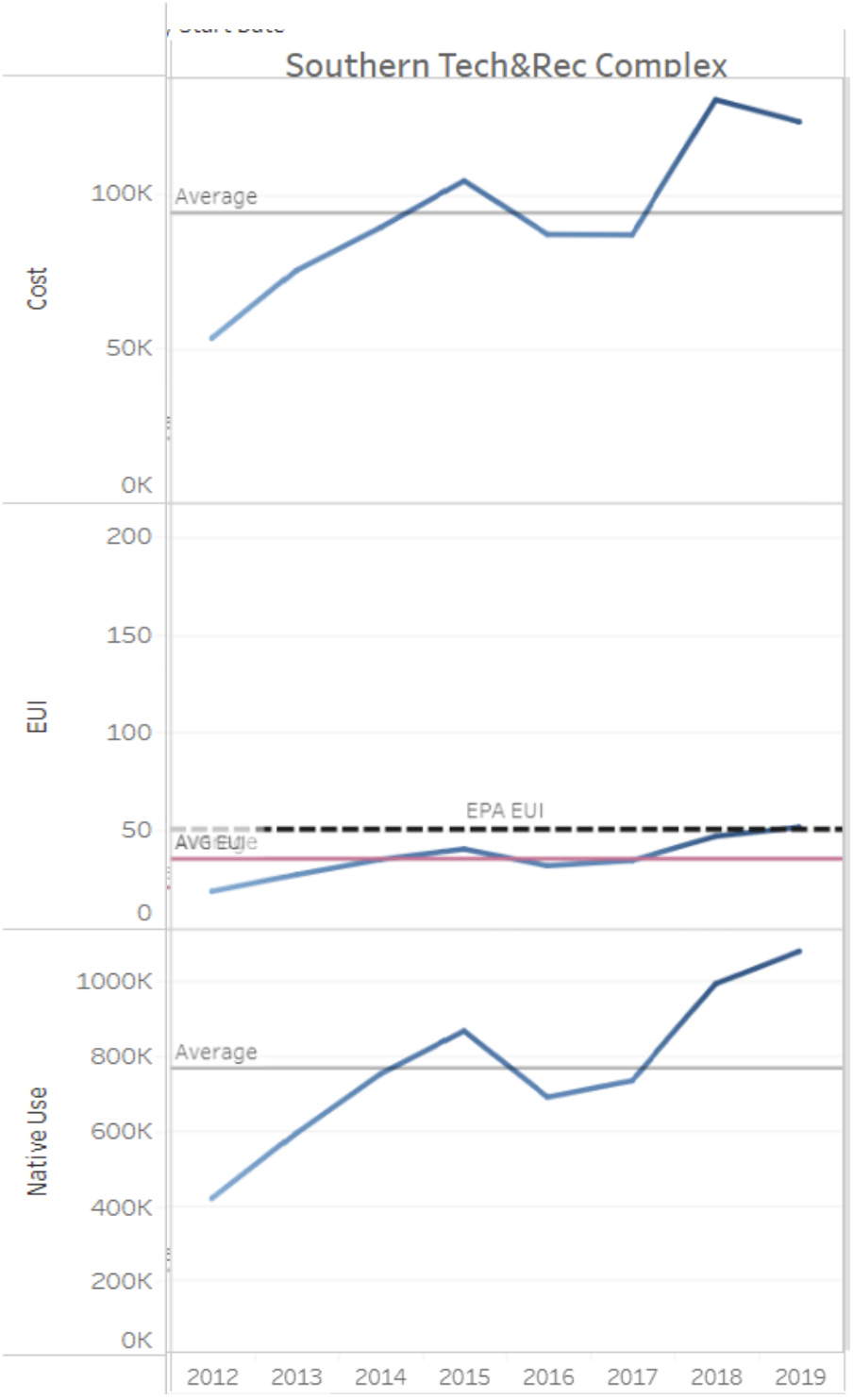
- This graph for Laurel Beltsville Senior Activity Center shows the cost, native use, and EUI from 2010-2019 (These years are adjustable on the interactive tableau webpage). The solid lines on the graph show the average over the years, and the dashed line on the EUI graph shows the EPA recommended EUI.



- This graph for the Palmer Park Community Center is formatted the same as the previous graph. While the cost and native use are strongly correlated, the EUI increases at a much higher rate from 2014-2016. The building was running efficiently for quite some time, it has recently become much less efficient for the standards set in 2019.



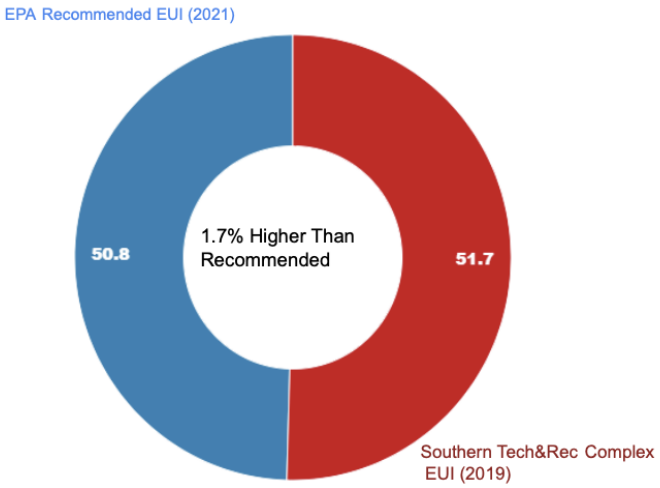
- This graph for the Publick Playhouse is formatted the same as the previous graph. The data for this facility has not changed much from 2010-2019. Although it is not an energy efficient building, changes may be difficult to implement due to the age of the building.



- This graph for the Publick Playhouse is formatted the same as the previous graph. Although this facility is very expensive, and uses lots of energy, it is a facility that only surpassed the EPA recommended EUI in 2019.

- **EPA recommended EUI and used EUI across four facilities**

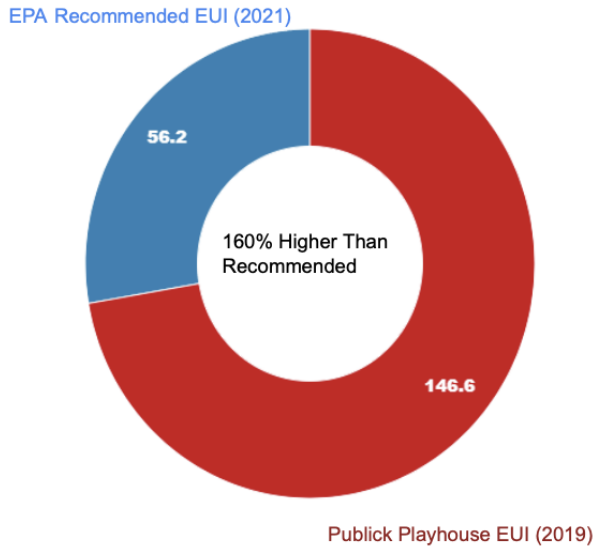
  - **Southern Regional Tech and Rec Complex**



Cost in 2019: $123,495
Expected Savings: $2,188

- This pie chart illustrates the EUI of Southern Regional Tech and Rec Complex and the EPA recommended EUI. As shown the EUI is 1.7% higher than the EPA recommended EUI and this percentage is the least among all the other percentages of buildings.

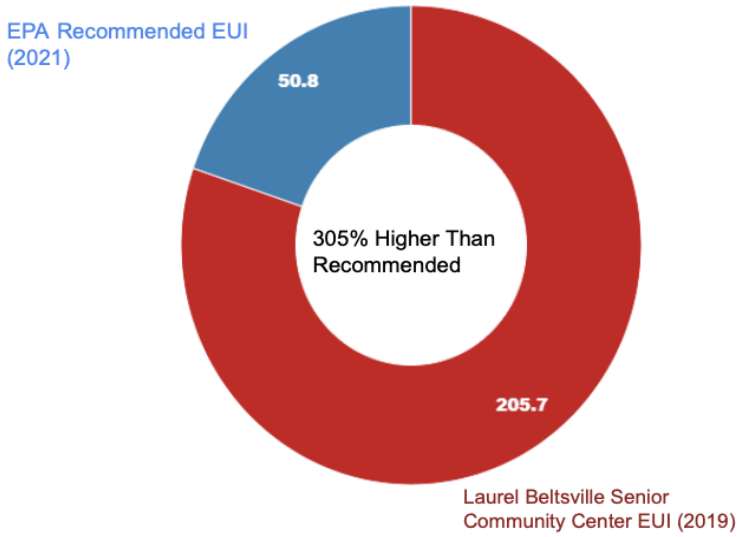
  - **Publick Playhouse**



Cost in 2019: $37,758
Expected Savings: $23,283

- The EUI for Publick playhouse is 160% higher than the EPA recommended.

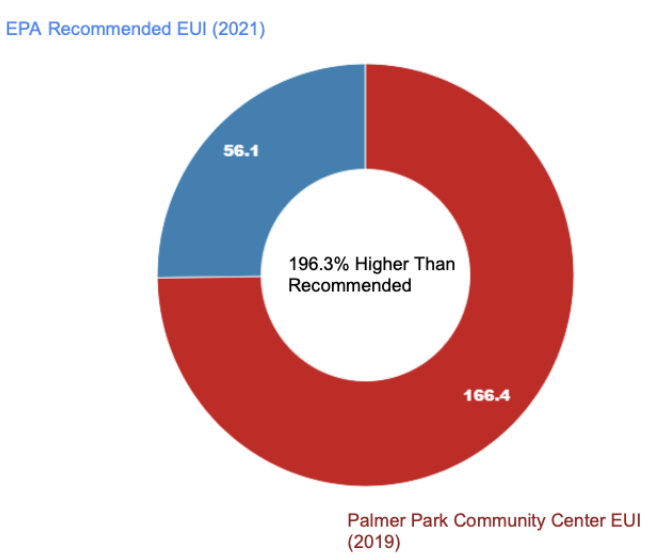
- **Laurel Beltsville Senior Community Center**



- This building EUI has the highest percentage compared to the EPA recommended among all the other buildings.

Cost in 2019: $44,916
Expected Savings: $33,675

- **Palmer Park Community Center**



- The EUI for Palmer Park Community Center is also higher than the recommended.. It is 193.3% higher than the EPA recommended.

Cost in 2019: $70,825
Expected Savings: $46,947

Conclusion

The EUI is an amazing and very effective tool we were able to utilize in this project. By calculating the average EUIs for the facilities that were operated by the Prince George's County Parks and Recreation Department and comparing it to the EUI that is recommended by the Environmental Protection Agency (EPA), we were able to determine how efficiently these buildings were running. What we found was that all of the facilities were operating at inefficient conditions, some much more than others. If we look at the Southern Regional Tech and Rec Complex, the average EUI for that facility was only slightly higher, by 1.7%, than the EPA recommended levels which is not terrible but certainly should be improved. However, when focusing on the other three facilities, we found that these facilities are operating with over 100% and in one case over 300% the recommended EPA levels which is seriously concerning. It's not only impacting the environment by utilizing tremendously more energy than necessary but it also is driving up the costs for operating and maintaining those facilities.

For example, analyzing the Southern Regional Tech and Rec Complex shows us that slightly above $2,000 can be saved by reducing the EUI levels to EPA recommended levels, which is important to accomplish but the savings for this facility are relatively low compared to the other facilities. If we observe the data for the Publick Playhouse and the Palmer Park Community Center, these facilities can save over $23,000 and over $46,000 by reducing their EUI levels to EPA recommended levels. The Laurel Beltsville Senior Community Center, which is over 300% the recommended level, can save over $33,000 which is approximately 75% of what it costs to operate and maintain that facility. What this data shows is that these facilities are unnecessarily utilizing a tremendous amount of energy and a significant amount of money can be saved by dedicating efforts to designing more energy efficient buildings in the future.

Solutions

Although devising detailed solutions was beyond the scope of our project and the time we had to complete it, we have proposed a few ideas that can be explored in the future. One of the ideas we came up with is ensuring buildings are not using their energy consuming equipment beyond what they intend to use it for. For example, we had observed that some of the facilities had adjacent buildings that both had parking lots right next to each other. We found that some of the street lamps for the parking lots of the Prince George's facilities are being shared with the parking lots of adjacent buildings. In the future, it is worth considering constructing facilities in a way that street lamps are not shared with parking lots of adjacent buildings for free.

Another solution we came up with is implementing a Building Energy Management System (BEMS) which essentially tracks energy usage of a building and conserves energy via autonomous monitoring. One way this system can reduce energy consumption is by deploying air conditioning to only areas of the building where there are occupants. For example, if most of

the employees leave the building by 5pm but there are still a few people who always stay a few hours longer, there is no need to have the air conditioning running in every area of the building. This type of smart monitoring could significantly reduce energy consumption and costs for the facilities.

We have also come across some issues when analyzing the data that impacts the reliability of the data and the accuracy of our calculations. There were some data that were missing and some that were inaccurate. If it is possible, it would be ideal to find the missing data and conduct the same calculations we did in order to have more accurate results regarding the EUI of the facilities. However, if retrieving the missing data is not possible, it is crucial that data is collected and documented more carefully and intentionally in the future to avoid these types of problems again.