Energy Efficiency of Residential Buildings as a Function of Building Characteristics

SCM 512 - Applied Business Analytics

University of Michigan - Flint

Barrie, Rhys - Dataset Overview (50%), Initial Observations, Cluster Analysis (50%), Regression Analysis (30%)

Blackwell, Brian - No Contributions

Bootz, Amanda - Problem Definition, Project Goals, Dataset Overview (50%), Cluster Analysis (50%), Literature Comparison

Boyapalli, Prathibha - Regression Analysis (70%), Managerial insights, Lessons Learned

Background and Objectives

Problem Definition

Our project aims to understand how a residential building's design characteristics impacts its energy efficiency, measured via the heating and cooling loads of the space. The heating and cooling loads are a measure of the energy that is required to maintain a comfortable temperature in an indoor space. We will evaluate a variety of design characteristics, including a building's size, orientation, and glass coverage, in order to determine which characteristics are most relevant.

The results of this analysis have several real-world applications. For example, by an architect who is designing a residential building and is attempting to improve energy efficiency. By understanding how design characteristics impact a residential building's energy efficiency, an architect or builder would be able to make data-driven design decisions. As there continues to be an increased focus on energy efficiency, this would also be data that consumers would be interested in reviewing when designing a new residential building or looking at home renovations.

Project Goals

We will achieve several objectives over the course of this investigation. First, we will identify overall trends and classes of building styles that correspond with particular ranges of heating and cooling loads. Second, we will develop a predictive model in order to estimate the heating and cooling loads based on a particular building's design characteristics. Third, we will compare our findings to existing literature in this field, and identify opportunities for further research. Finally, we will draw actionable conclusions that can be easily referenced by decision makers involved in a building's design, such as which design characteristics are most impactful on a building's energy efficiency, and which styles and classes of buildings are the most energy efficient.

1. Which design characteristics have the most significant impact on heating and cooling load?

2. What is the most energy efficient residential building design?
3. What impact do each of the design characteristics have on heating and cooling load?
4. How can optimizing features like glazing area, building compactness, and orientation help reduce a building's overall energy consumption?
5. What kinds of trade-offs exist between different design elements, such as balancing the amount of glazing area with wall insulation, and how do these trade-offs impact energy efficiency?

Methodology

Dataset Overview

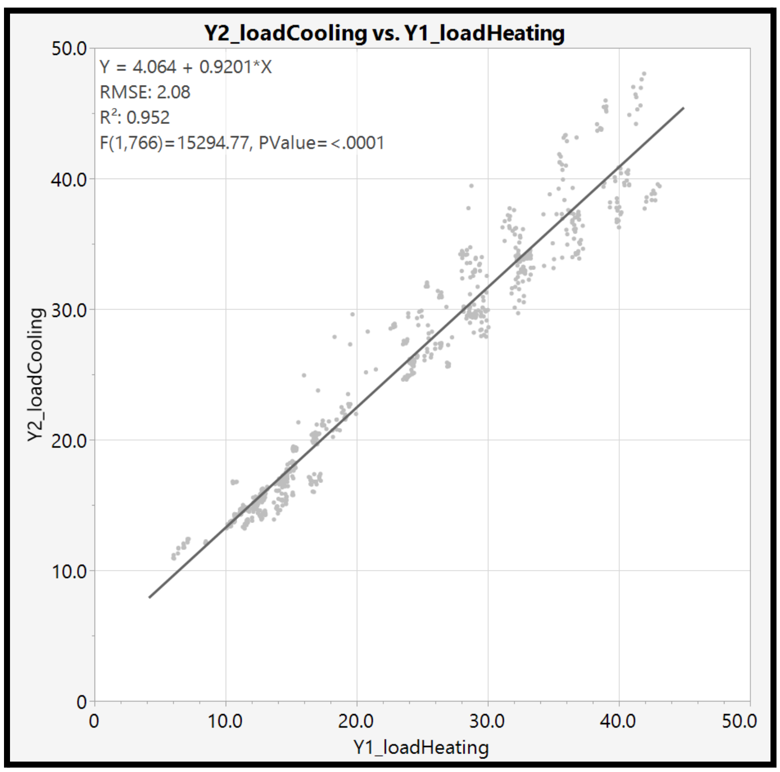
The dataset we are analyzing comes from the study "Accurate quantitative estimation of energy performance of residential buildings using statistical machine learning tools" by Tsanas and Xifara. The dataset contains measurements of various building design parameters as input variables, as well as the heating and cooling loads of the resulting buildings as output variables.

The residential buildings that were used in the creation of this dataset were all simulated using a software called Ecotect. All of the simulated buildings have the same volume of 771.75 cubic meters, but they have varying dimensions and surface areas. The dataset also assumes that the residential buildings are located in Athens, Greece, with an occupancy of seven people, sedentary activity, and the use of common construction materials. This helps to ensure that all other variables, outside of the input variables, were constant, and would not influence the output variables.

The dataset provides values for the following input (X) and output (Y) variables:

ID	Variable Name	Type	Notes
X1	Relative Compactness	Continuous Ratio (skew 0.50)	
X2	Surface Area	Continuous Ratio (skew -0.12)	
X3	Wall Area	Continuous Ratio (skew 0.53)	
X4	Roof Area	Continuous Ratio (skew -0.16)	
X5	Overall Height	Continuous Ratio (skew 0.00)	Represents single story (3.5 meter) and two-story (7 meter) buildings

X6	Orientation	Nominal	2 = North 3 = East 4 = South 5 = West
X7	Glazing Area	Continuous Ratio (skew -0.06)	Measured as percentage of the floor area - 0%, 10%, 25%, and 40%
X8	Glazing Area Distribution	Nominal	0 = no glazing 1 = uniform distribution (25% per side) 2 = 55% north, 15% other 3 = 55% east, 15% other 4 = 55% south, 15% other 5 = 55% west, 15% other
Y1	Heating Load	Continuous Ratio (skew 0.36)	
Y2	Cooling Load	Continuous Ratio (skew 0.40)	



Initial Observations

Upon initial review of the dataset, we discovered the following preliminary observations:

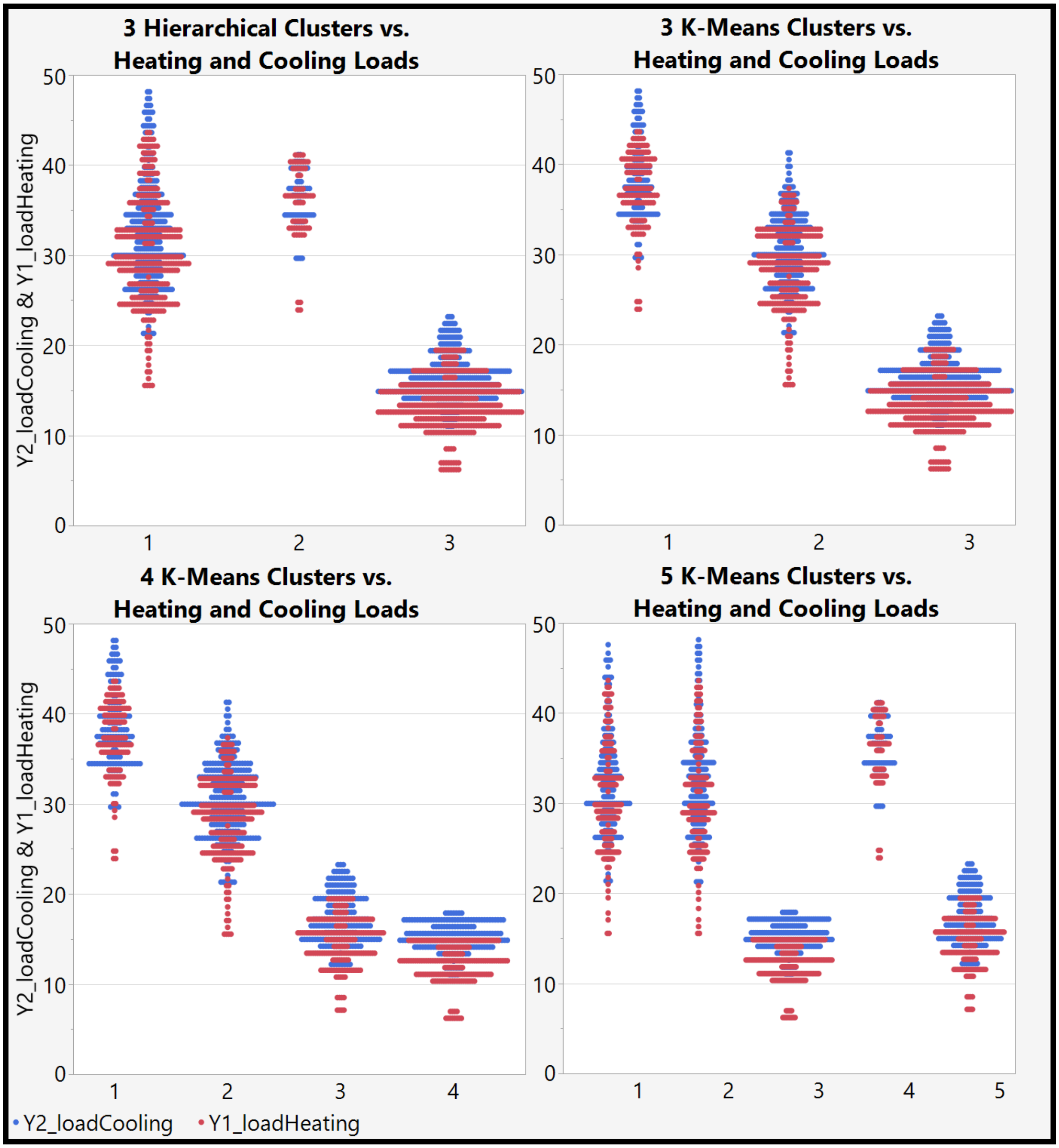
1. Heating and cooling loads are strongly correlated, and don't appear to diverge for any of the independent variables. Cooling loads are generally higher than heating loads, which aligns with the assumption that all simulated structures are located in Greece, with hot summers and mild winters.
2. Orientation of the building and orientation of the glazing do not appear to have significant differences in heating and cooling loads.
3. Taller homes have significantly higher heating and cooling loads for the same building volume.
4. Contradicting that, larger footprint homes appear to have lower heating and cooling loads.

5. There appears to be some interesting clustering occurring for the compactness and surface area variables.

Results and Analysis

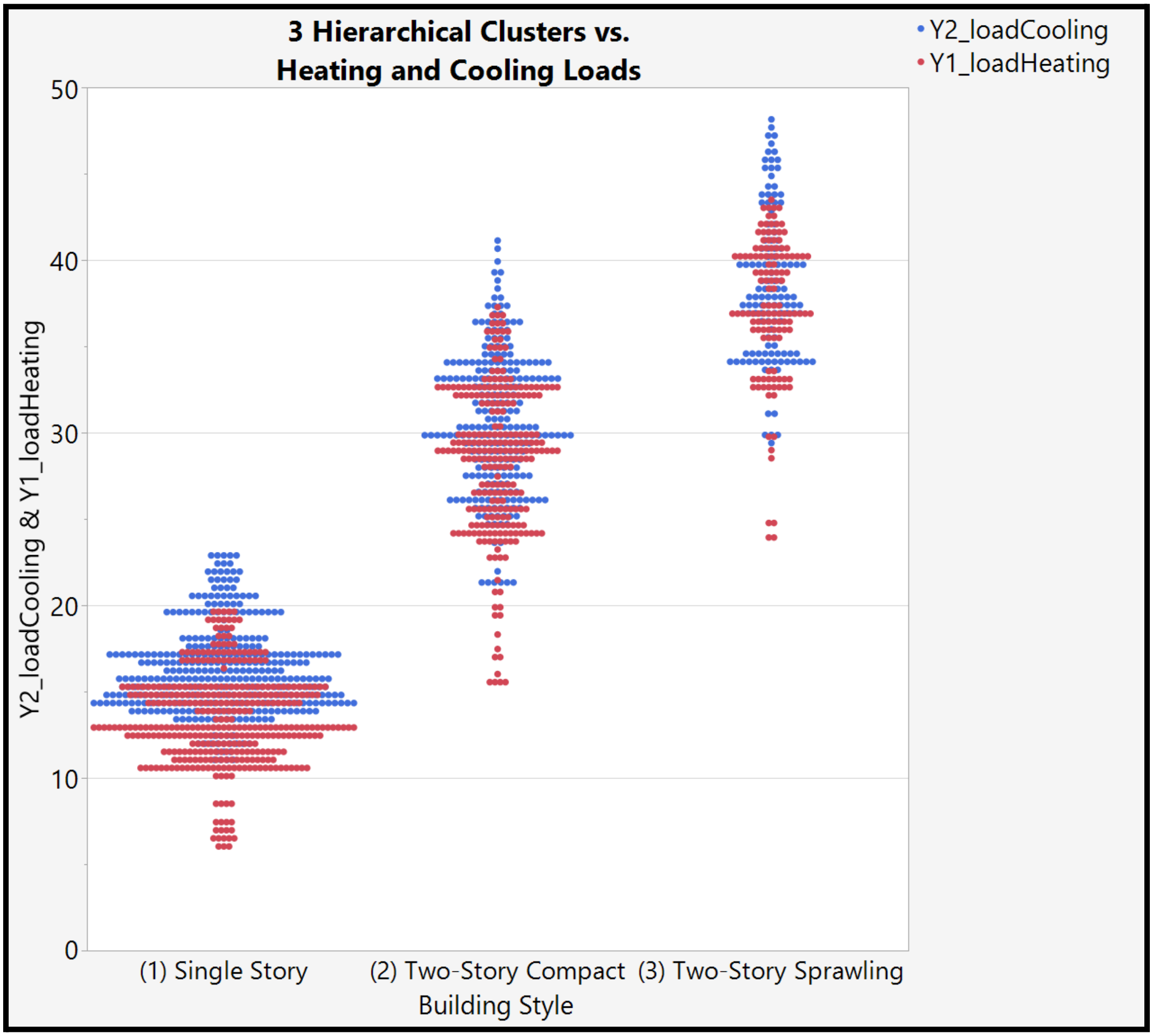
Cluster Analysis Model

For our Cluster Analysis, we wanted to analyze the x-variables against the heating and cooling loads to determine which variable had the most distinct impact on the loads. We reviewed the results for hierarchical clustering and k-means clusters to compare the differences between 3, 4, and 5 clusters to determine which one was the most effective to analyze our data. The charts below show our results for each of these clustering models, based on the number of clusters.



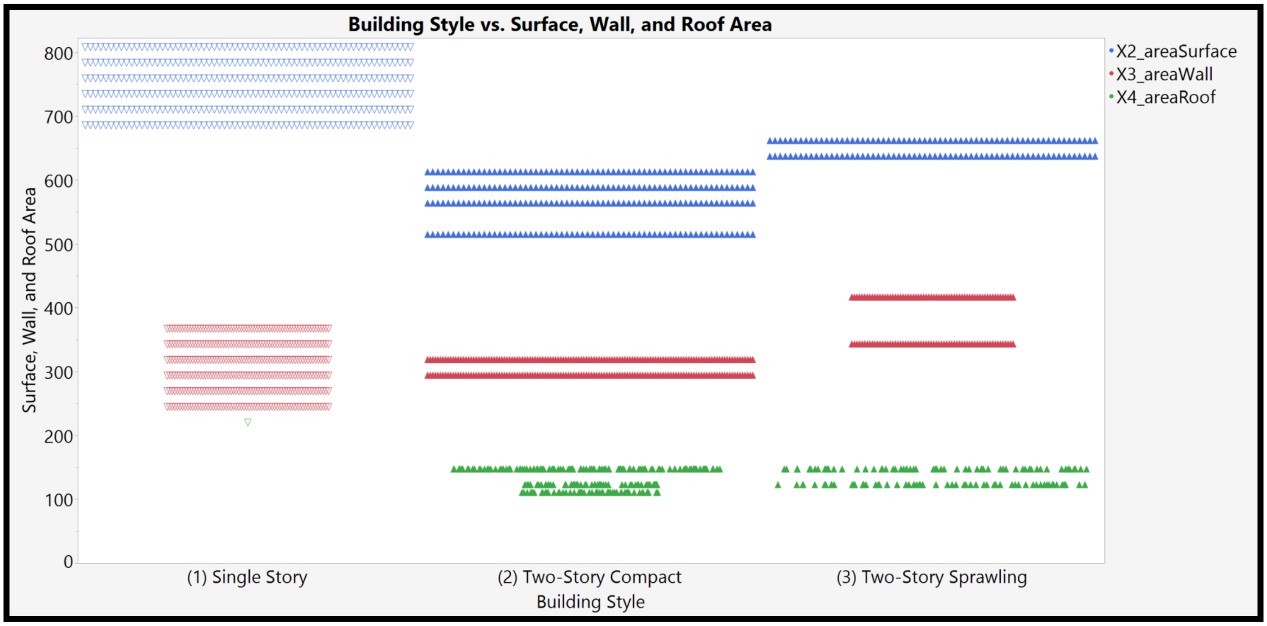
Hierarchical clustering appeared to show 3 clusters was ideal, but did not have significantly distinct heating/cooling loads for 2 of the 3 clusters. We then performed k-means cluster analysis with 3, 4, and 5 clusters, and plotted heating/cooling loads. This was then compared against the results for the hierarchical clustering.

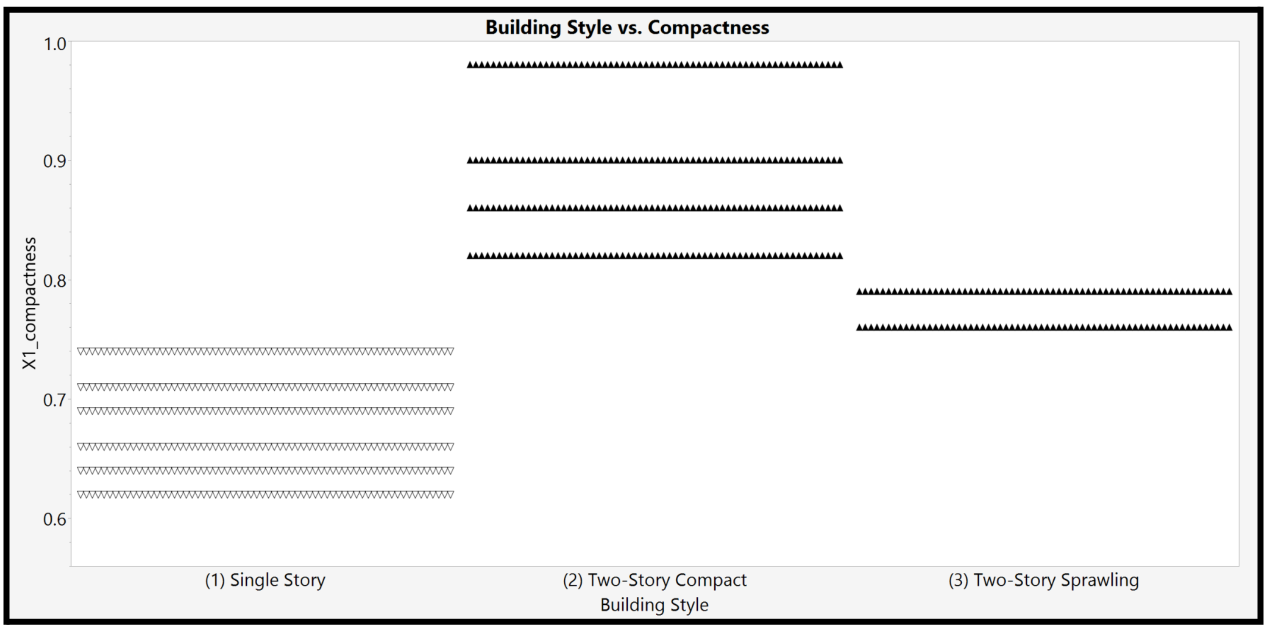
Based on these results, the 3 clusters k-means had clear distinction between low, medium, and high loads. We chose this model to move forward with our analysis, since it had such a clear distinction between the three clusters.



We identified three distinct clusters, representing low-, medium-, and high-load buildings. The clusters corresponded with the following building styles:

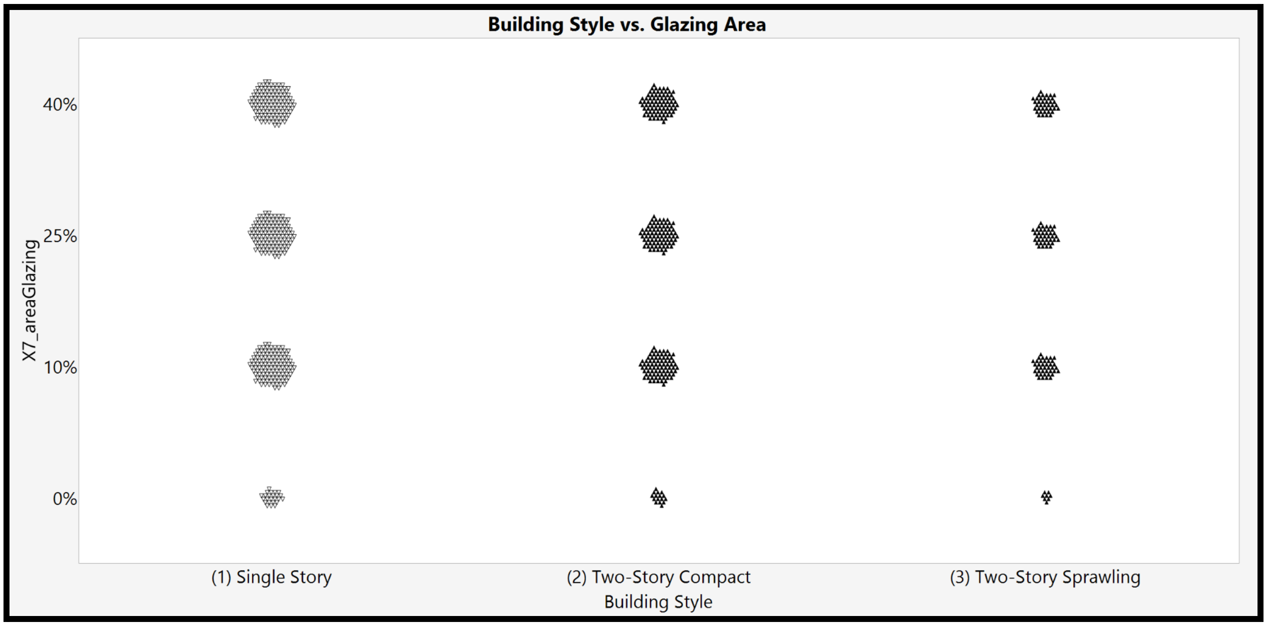
- Cluster 1 = single story buildings, resulting in low heating/cooling loads
- Cluster 2 = compact two-story buildings, resulting in medium heating/cooling loads
- Cluster 3 = less-compact two-story buildings, resulting in high heating/cooling loads

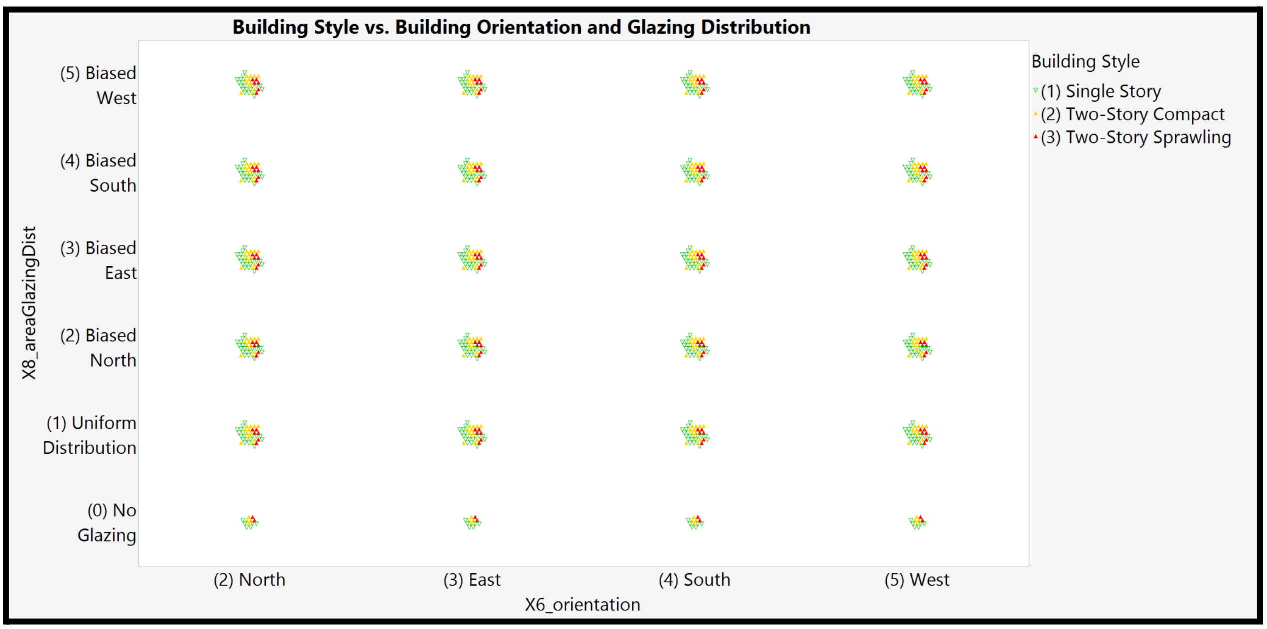




In the building areas graph above, all of the two-story buildings in cluster 2 have a smaller surface area and wall area than the two-story buildings in cluster 3, supporting the assertion that cluster 2 consists of two-story buildings with a more condensed and compact floor plan. Given that all of the buildings in this dataset have the same volume, buildings in cluster 2 may have been constructed with higher ceilings or in a more condensed floor plan, whereas buildings in cluster 3 may have lower ceilings or a more elongated / sprawling floor plan. Single-story buildings in cluster 1 have a larger footprint than the other two clusters, given that the single-story buildings spread outwards, rather than upwards.

These area measurements are reflected in the compactness graph below, where the single story buildings in cluster 1 are less compact than the two-story buildings in clusters 2 and 3. In addition, the two-story buildings with a smaller footprint in cluster 2, universally have a higher compactness rating than the two-story buildings in cluster 3.





As depicted in the glazing area and building/glazing direction graphs above, there does not appear to be any variation between clusters with regard to building orientation, glazing area, nor glazing distribution.

Regression Analysis Model

Heating Load

A multiple regression was carried out to investigate whether any of the 8 measurements of residential building characteristics could significantly predict the heating load of a building.

- Building orientation did not significantly predict increased heating load for a residential building, t(767df) = 0.0707, p = 0.9756. The variable was removed from the model.

- Surface area is redundant with the roof area and wall area variables, due to the relationship between the variables being X2_areaSurface = X3_areaWall + 2*X4_areaRoof. The variable was zeroed and removed from the model.

The results of the regression with the remaining 6 building characteristics indicated that the model explained 92.31% of the variance and that the model is a significant predictor of heading load, F(10df, 757df) = 921.1594, $p < 0.0001$.

- Glazing area significantly predicts increased heating load for a residential building, t(767df) = 19.78, $p < 0.0001$.
- Overall height significantly predicts increased heating load for a residential building, t(767df) = 12.93, $p < 0.0001$.
- Glazing distribution significantly predicts heating load for a residential building, t(767df) = 17.55, $p < 0.0001$.
- Relative compactness significantly predicts decreased heating load for a residential building, t(767df) = -6.60, $p < 0.0001$.
- Roof area significantly predicts decreased heating load for a residential building, t(767df) = -5.36, $p < 0.0001$.
- Wall area significantly predicts decreased heating load for a residential building, t(767df) = -2.17, $p = 0.0301$.

The predictive formula for heating loads is as follows:

$$Y_1\ (Heating\ Load) = 84.775 + 16.848 * X_7\ (Glazing\ Area) + 4.170 * X_5\ (Height)$$

$$- 3.620 * X_8[0]\ (No\ Glazing) + 0.908 * X_8[1]\ (Uniform\ Glazing)$$

$$+ 0.816 * X_8[2]\ (Glazing\ Biased\ North) + 0.563 * X_8[3]\ (Glazing\ Biased\ East)$$

$$+ 0.769 * X_8[4]\ (Glazing\ Biased\ South) - 64.773 * X_1\ (Compactness)$$

$$- 0.027 * X_4\ (Roof\ Area) - 0.175 * X_4\ (Wall\ Area)$$

<u>Cooling Load</u>

A multiple regression was carried out to investigate whether any of the 8 measurements of residential building characteristics could significantly predict the cooling load of a building.

- Building orientation did not significantly predict cooling load for a residential building, t(767df) = 1.4047, p = .2402. The variable was removed from the model.
- Surface area is redundant with the roof area and wall area variables, due to the relationship between the variables being X2_areaSurface = X3_areaWall + 2*X4_areaRoof. The variable was zeroed and removed from the model.

- Glazing distribution significantly predicts cooling load for a residential building, t(767df) = 3.08, p = 0.0092. However, the orientation of the glazing does not significantly predict cooling load for a residential building, only the uniform distribution or lack of glazing. Therefore, this variable is redundant with glazing area, and was removed from the model.
    - X_8[0] (No Glazing), t(767df) = -3.56, p = 0.0004.
    - X_8[1] (Uniform Glazing), t(767df) = 2.30, p = 0.0217.
    - X_8[2] (Biased North), t(767df) = 1.58, p = 0.1149.
    - X_8[3] (Biased East), t(767df) = 0.24, p = 0.8069.
    - X_8[4] (Biased South), t(767df) = 1.65, p = 0.0992.

The results of the regression with the remaining 5 building characteristics indicated that the model explained 88.68% of the variance and that the model is a significant predictor of cooling load, F(5df, 762df) = 1202.957, p < .0001.

- Glazing area significantly predicts increased cooling load for a residential building, t(767df) = 17.08, p < 0.0001.
- Overall height significantly predicts increased cooling load for a residential building, t(767df) = 11.62, p < 0.0001.
- Relative compactness significantly predicts decreased cooling load for a residential building, t(767df) = -6.31, p < 0.0001.
- Roof area significantly predicts decreased cooling load for a residential building, t(767df) = –4.74, p < 0.0001.
- Wall area significantly predicts decreased cooling load for a residential building, t(767df) = –3.13, p = 0.0018.

The predictive formula for cooling loads is as follows:

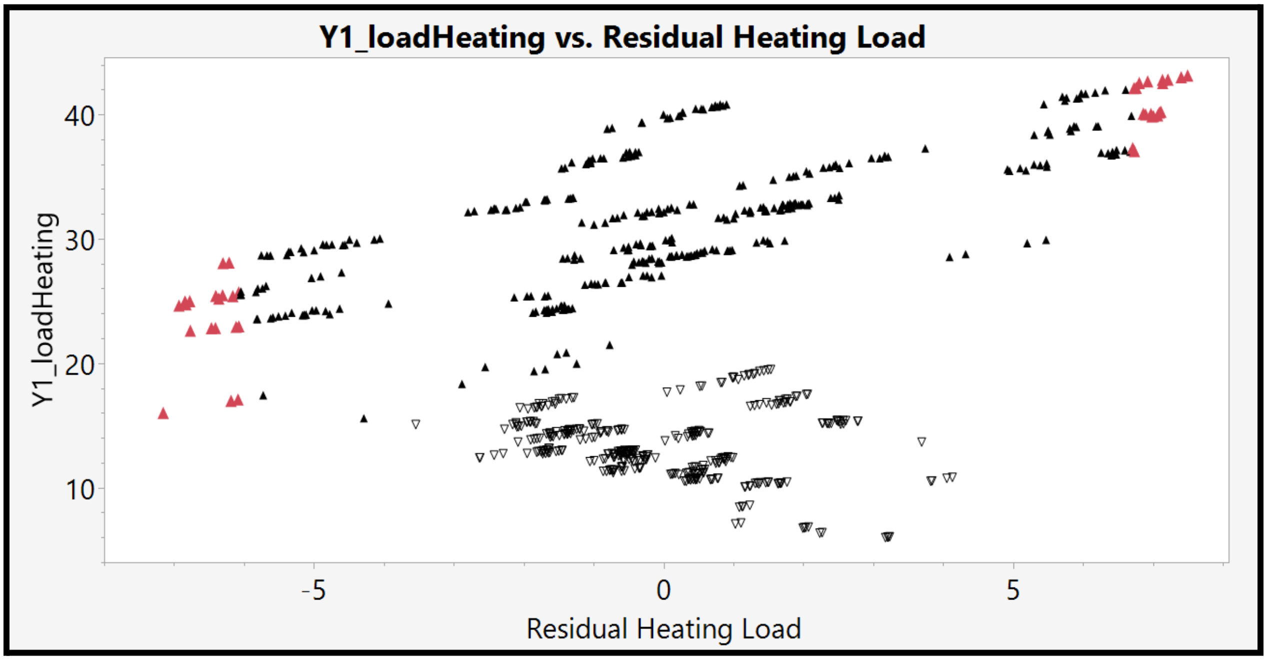
$$Y_2\,(Cooling\ Load) = 97.931 + 13.253 * X_7\,(Glazing\ Area) + 4.284 * X_5\,(Height)$$

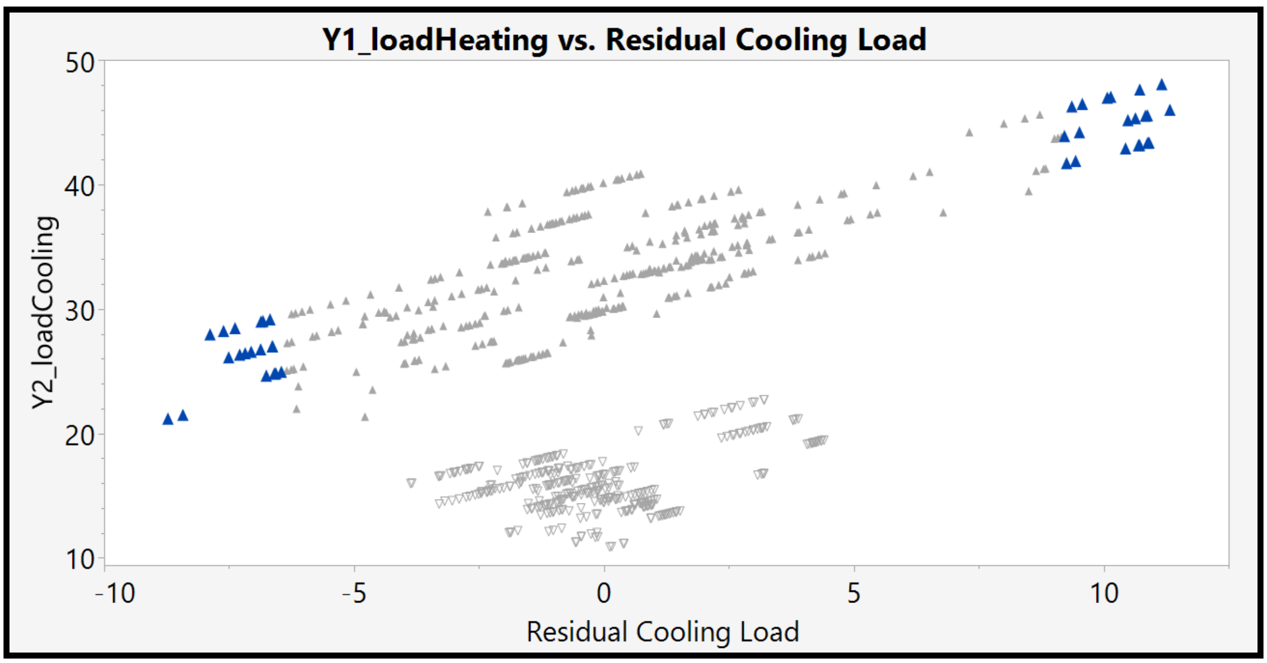
$$- 70.788 * X_1\,(Compactness) - 0.176 * X_4\,(Roof\ Area) - 0.044 * X_4\,(Wall\ Area)$$

<u>Overestimated and Underestimated Data Points</u>

The 20 most overestimated and underestimated buildings for both heating and cooling loads are all two-story buildings. There are no other discernible similarities between the outliers, except for those variables that are inherently correlated with two-story buildings – they are relatively uniformly distributed with regards to glazing area, orientation, etc.

Between heating and cooling loads, there are 6 buildings where *both* heating and cooling load were overestimated. These 6 buildings are all the same structure, simply oriented in different directions.

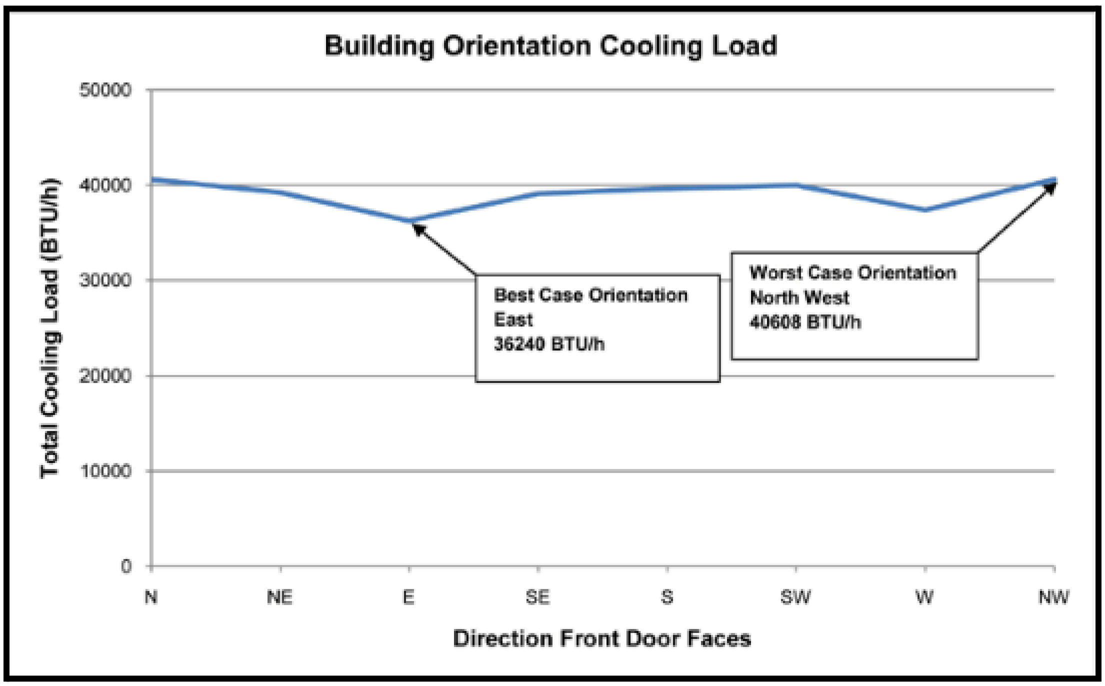




Literature Comparison

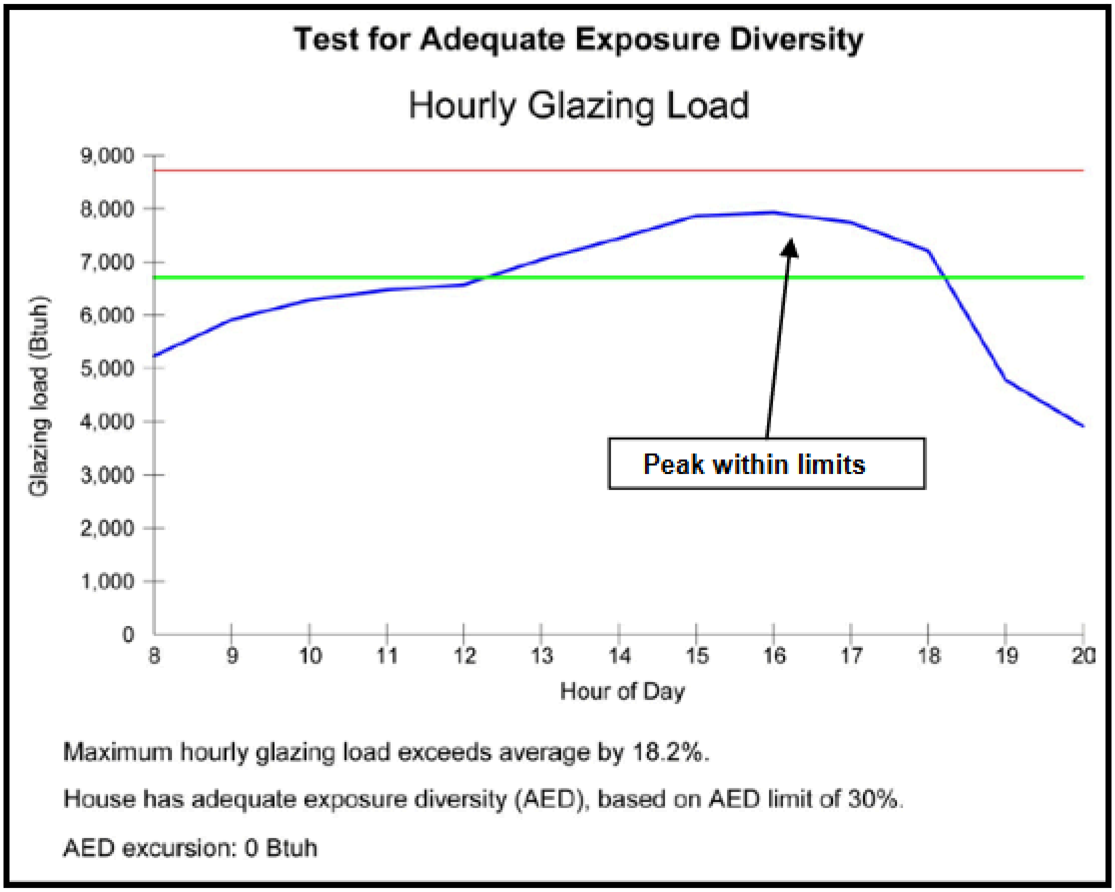
The U.S. Department of Energy published a report in June of 2011 titled “Strategy Guideline: Accurate Heating and Cooling Load Calculations”. We reviewed this report to compare our findings to what they published, as far as variables that impact heating and cooling loads and how this data is used within industry. According to this report, the four variables that impact heating and cooling loads within structures include the location of the building, specific design conditions within the interior of the building, the orientation, and how the building is constructed.

Our analysis didn't review building location, as it was defined in this report through latitude and elevation, so we could not easily compare our findings with this data. The orientation was the first variable that we could compare against the findings in this literature. Through our data analysis, orientation which was defined as north, east, south, and west did not strongly predict heating and cooling loads. This variable was actually removed from our models. While reviewing the report published by the U.S. Department of Energy, it was noted that orientation did have an impact on the amount of heat a house would increase by, depending on other factors such as window placement on particular faces of the structure. Their findings, as shown below, show that the total cooling load fluctuated about 12% from its highest to lowest point. This is interesting, as we compared to our findings, since we didn't find orientation to be a strong predictor. There is little fluctuation, which would substantiate our findings.



Burdick, 2011. Figure 10 - Building Orientation Cooling load

It was also interesting when we compared our findings against the data published by the U.S. Department of Energy report when it came to glazing. The report provided data from a building that was studied for its hourly glazing load, to show any variations in the glazing load, by each hour of the day. Per the data published, the glazing load fluctuated by 100% throughout the course of the day. We showed that glazing distribution significantly impacts the heating and cooling load. This data supports our findings, since there seems to be a significant impact on the glazing load, by each hour of the day. The findings from the U.S. Department of Energy have been shown in the chart below.



Burdick, 2011. Figure 11 - Test for Adequate Exposure Diversity

The other variables that were outlined in this report had more to do with the construction materials, rather than surface area, which were the variables that were analyzed in our studies. Glazing and orientation were the two we were easily able to compare between the two reports.

Conclusions

Managerial Insights

- What are the most significant factors influencing heating and cooling loads?

  According to regression calculations on residential buildings, the building's total height and glazing area have the most significant impact on heating and cooling loads. ~~Because of its vast surface area, which impacts heat intake during higher temperatures and loss during lower ones, wall area specifically affects energy consumption.~~ Another important consideration is glazing area since, depending on their size and quality, windows can permit significant heat gain in the summer and significant heat loss in the winter. Another important factor is the building's overall height; taller structures usually have higher thermal loads because of increased surface area exposure and possible stack effects. In order to improve building energy efficiency, these findings highlight the need of incorporating energy-efficient design elements like high insulation, well-placed windows, and cutting-edge glazing technologies.

These factors are essential for reducing operating costs, enhancing customer satisfaction, and accomplishing environmental goals in building design.

- How does the glazing area affect the cooling load?

  Building cooling loads are significantly impacted by the glazing area, mainly because of its role in solar heat gain. Because larger windows let in more sunlight, indoor temperatures rise, increasing the need for air conditioning equipment to keep the room suitable. If the windows lack energy-efficient features like double glazing or low-emissivity coatings, which may reduce heat transfer, this effect is amplified. Strategic architectural features, such as shade-producing extensions or the use of tinted and specifically coated glass, can help overcome these difficulties by lowering the cooling load, increasing energy efficiency, and enhancing building comfort in general.

- What impact does overall height have on the heating load?

  Due to a number of interconnected factors, the building's overall height significantly affects its heating load. First of all, taller structures lose more heat because they have more surface area exposed to the outside. Furthermore, heating a taller building uses more energy to heat its bigger volume of air, which increases the strain on heating systems. Particularly in multi-story structures, the stack effect, which causes warm air to rise and leave at higher levels while drawing cold air at lower levels, makes the need for heating substantially greater. A greater amount of heat loss can also result from taller structures' upper levels being more exposed to lower temperatures and colder winds.

Lessons Learned

Throughout this project, we gained a deeper understanding of how residential building design impacts energy efficiency, particularly heating and cooling loads. We realized that certain variables, such as glazing area, overall height, and compactness, play critical roles in determining energy consumption. It was fascinating to see how small design changes could significantly influence the energy requirements of a building.

One key insight was that taller buildings, despite having the same volume as single-story buildings, face higher heating and cooling loads due to their larger exposed surface area. Similarly, the importance of glazing distribution and compactness highlighted how strategic design choices could optimize energy efficiency without compromising on usability or aesthetics.

We also encountered some challenges along the way. For instance, interpreting how variables like orientation and glazing distribution fit into the overall model required careful analysis since their impact was less straightforward than we initially expected. Working through these challenges reinforced the importance of thorough data cleaning, proper model selection, and cross-referencing findings with existing literature to validate our conclusions.

Overall, this project taught us the value of approaching design decisions with a data-driven mindset, emphasizing sustainability and efficiency. It was a reminder that even in complex systems, actionable insights can be extracted to create tangible, real-world benefits.

Citations

Burdick, Arlan. *Strategy Guideline: Accurate heating and cooling load calculations*. No. NREL/SR-5500-51603; DOE/GO-102011-3304. National Renewable Energy Lab.(NREL), Golden, CO (United States), 2011. https://www.nrel.gov/docs/fy11osti/51603.pdf.

Tsanas, Athanasios, and Angeliki Xifara. *Accurate quantitative estimation of energy performance of residential buildings using Statistical Machine Learning Tools*. Energy and Buildings, vol. 49, June 2012, pp. 560–567, https://doi.org/10.1016/j.enbuild.2012.03.003.