Mike Truong
Naila Hajiyeva
Naima Sagar
Michael Kobrosky
Rahul Ranganath
Varunika Selvakumar
Ahad Kesaria

Energy Efficiency Modeling Predictions Based on Building Characteristics

Introduction

In recent years, there has been a growing interest in the energy performance of buildings (EFB) as nations around the world look to reduce their energy wastage and carbon footprints. Furthermore, spearheaded by the European Union, many nations have now implemented laws which require all buildings to meet a minimum energy efficiency requirement (EC, 2002). EFB has become an even more pertinent topic following the Russian invasion of Ukraine in February 2022. Indeed, the exorbitant cost to heat or cool one's home amidst rising fuel prices has been cited as one of the main contributing factors to the current global cost of living crisis (IMF, 2022)

Over the past few decades, building energy consumption has been steadily increasing worldwide. Most of the energy use in buildings can be accounted for by heating, ventilation and air conditioning (HVAC) systems used to control their indoor climates (Pérez-Lombard, 2008). Therefore one way in which nations could look to substantially reduce their carbon footprints is by designing more energy efficient buildings with improved energy conservation properties. This paper thus explores which design features of buildings are most relevant in increasing their heating and cooling efficiencies.

Data

In order to determine which building attributes are most important for energy performance, we will be assessing 8 different building features on cooling load, heating load and overall load. These 8 features are namely relative compactness, surface area, wall area, roof area, overall height, orientation, glazing area, and glazing area distribution.

The relative compactness is the building's surface area to volume ratio measured against a baseline of 1. The building's surface area, wall area and roof area are measured in square meters and its overall height in meters. The orientation of the building refers to the cardinal direction in which it faces (north, south, east or west) and its glazing area is expressed as a percentage of the floor area (0%, 10%, 25% or 40%). The glazing area distribution refers to one of six different

Mike Truong
Naila Hajiyeva
Naima Sagar
Michael Kobrosky
Rahul Ranganath
Varunika Selvakumar
Ahad Kesaria

scenarios: (1) uniform: with 25% glazing on each side, (2) north: 55% on the north side and 15% on each of the other sides, (3) east: 55% on the east side and 15% on each of the other sides, (4) south: 55% on the south side and 15% on each of the other sides, (5) west: 55% on the west side and 15% on each of the other sides and (6) no glazing. Finally, the heating load measures the amount of heat energy that would need to be added to a space whilst the cooling load measures the amount of heat energy that would need to be removed from a space to maintain the temperature in an acceptable range. The overall load is simply the sum of the cooling and heating loads and all 3 output variables are measured in British Thermal Units (BTU).

A table summary of all the 8 input and 3 output variables is provided below. It also includes the number of distinct values of each variable.

Mathematical Representation	Input or Output Variable	Number of Possible Values
X1	relative compactness	12
X2	surface area	12
X3	wall area	7
X4	roof area	4
X5	overall height	2
X6	orientation	4
X7	glazing area	4
X8	glazing area distribution	6
Y1	heating load	586
Y2	cooling load	636
Y3	overall load	786

The data set that we will be using to conduct the study is based on a sample of 768 simulated residential buildings, each with a unique combination of the 8 building attributes. All of the buildings have the same volume, 771.75 m^3 and are assumed to have been built with the same materials. The simulation also assumes that the residential buildings are in Athens, Greece with seven inhabitants that are primarily sedentary (70W). The internal design conditions were set to “clothing: 0.6 clo, humidity: 60%, air speed: 0.30 m/s, lighting level: 300 Lux.” In addition, the

Mike Truong
Naila Hajiyeva
Naima Sagar
Michael Kobrosky
Rahul Ranganath
Varunika Selvakumar
Ahad Kesaria

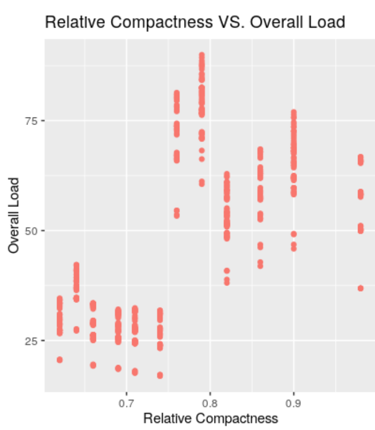
buildings are assumed to have 95% efficient mixed mode climate conditioning systems. The thermostat was set in the range of 19–24 °C, with 15–20 h of operation on weekdays and 10–20 hours on weekends (Tsanas, 2012).

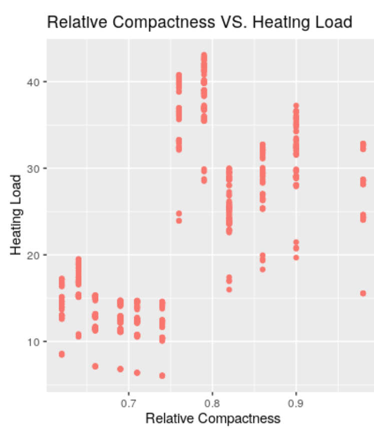
Finally, it should be noted that since the data set was generated via targeted simulations, we did not have to perform any data cleaning operations on the sample. The y3 variable (overall load) was not in the original data set but generated by adding the heating and cooling load efficiencies together.

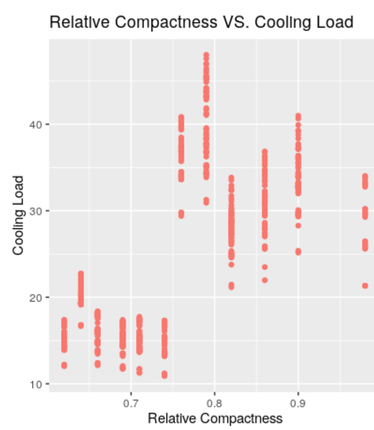
Exploratory Analysis

Based on initial plots of each of the explanatory variables against each of the output variables, we have generated the following hypotheses:

1. As the relative compactness of a building increases, cooling load, heating load and overall load will increase as well, resulting in a positive regression coefficient in all three cases.

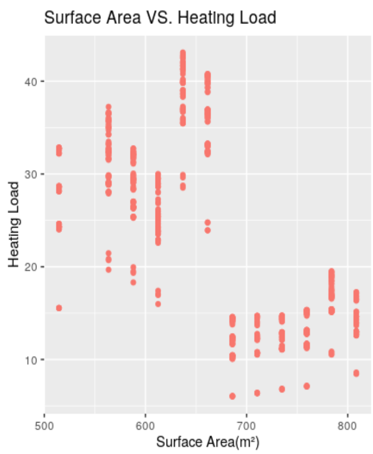


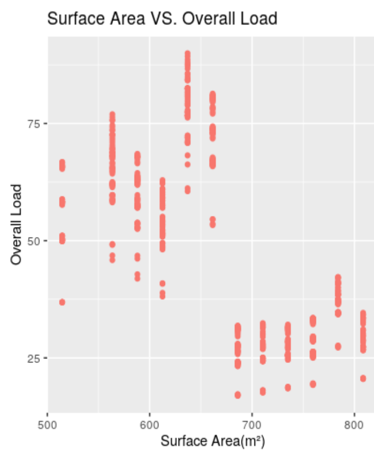


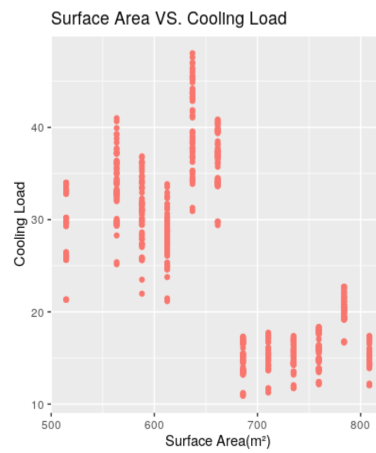


Mike Truong
Naila Hajiyeva
Naima Sagar
Michael Kobrosky
Rahul Ranganath
Varunika Selvakumar
Ahad Kesaria

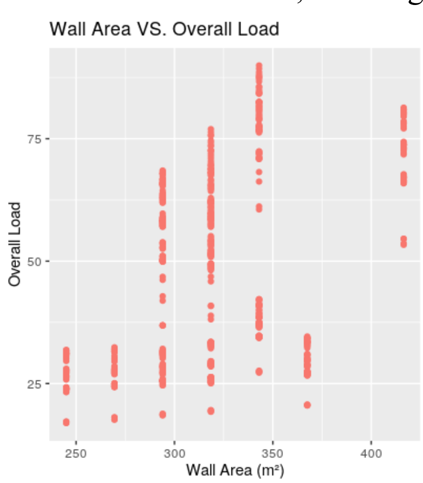
2. As the surface area of a building increases, cooling load, heating load and overall load will decrease, resulting in a negative regression coefficient in all three cases.

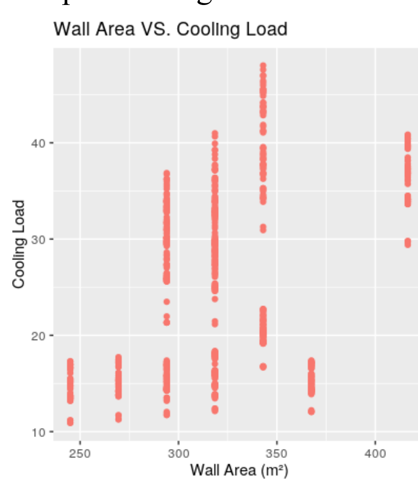


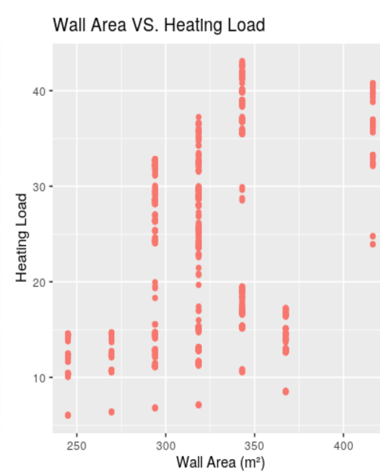




3. As the wall area of a building increases, cooling load, heating load and overall load will increase as well, resulting in a positive regression coefficient in all three cases.

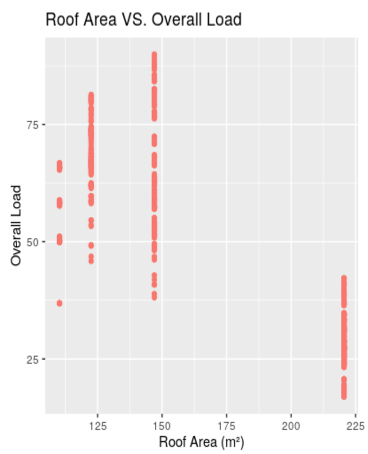


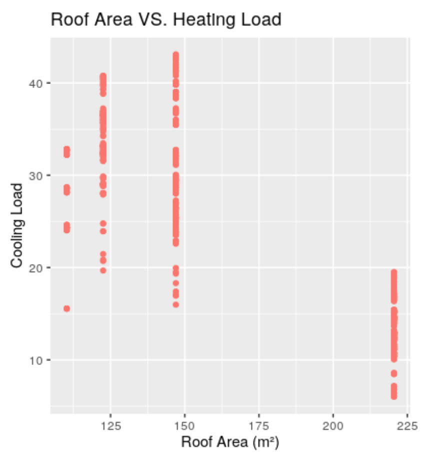


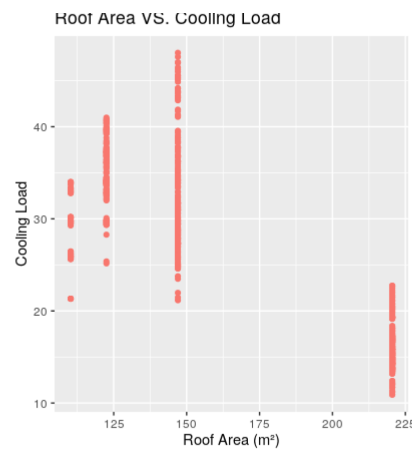


Mike Truong
Naila Hajiyeva
Naima Sagar
Michael Kobrosky
Rahul Ranganath
Varunika Selvakumar
Ahad Kesaria

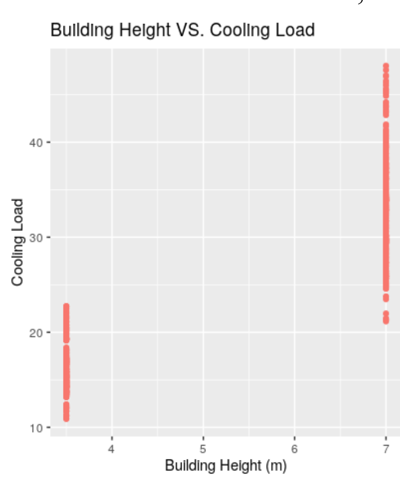
4. As the roof area of a building increases, cooling load, heating load and overall load will decrease, resulting in a negative regression coefficient in all three cases.

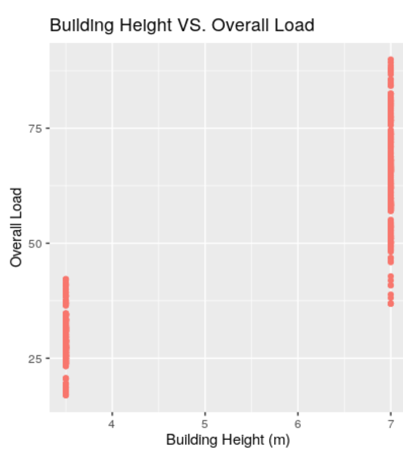


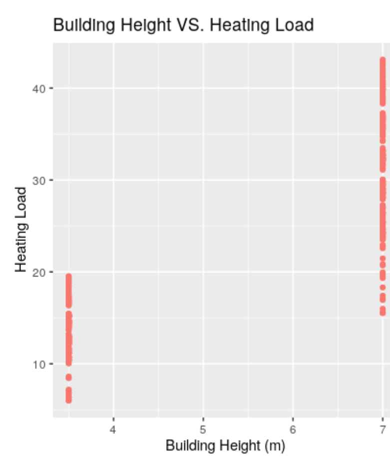




5. As the overall height of a building increases, cooling load, heating load and overall load will increase as well, resulting in a positive regression coefficient in all three cases.

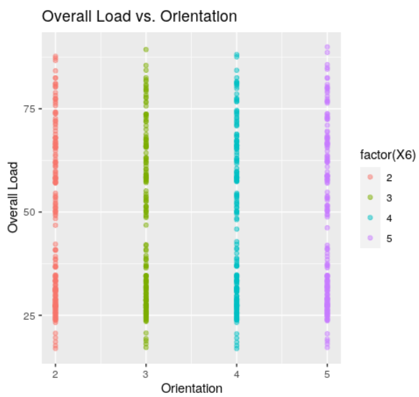


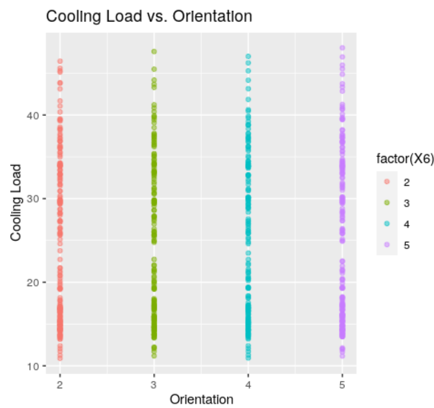


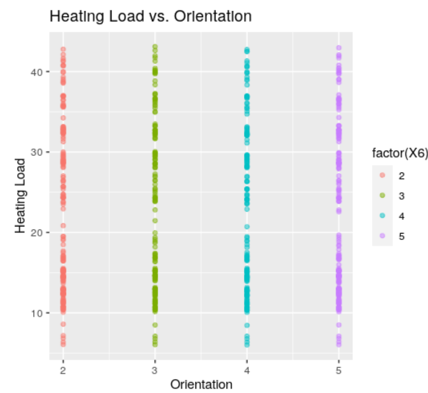


Mike Truong
Naila Hajiyeva
Naima Sagar
Michael Kobrosky
Rahul Ranganath
Varunika Selvakumar
Ahad Kesaria

6. The orientation of the building will have little effect on cooling load, heating load and overall load. Thus the regression coefficient will be close to 0 for each orientation in all three cases.

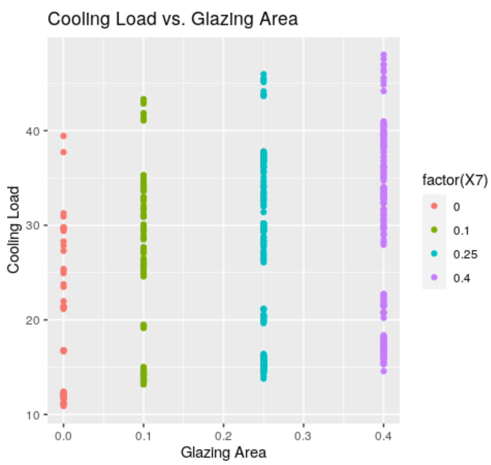


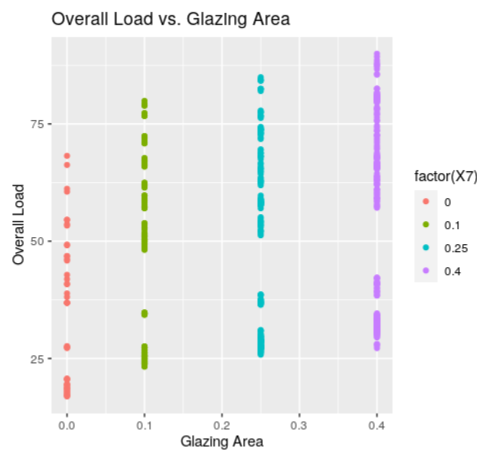


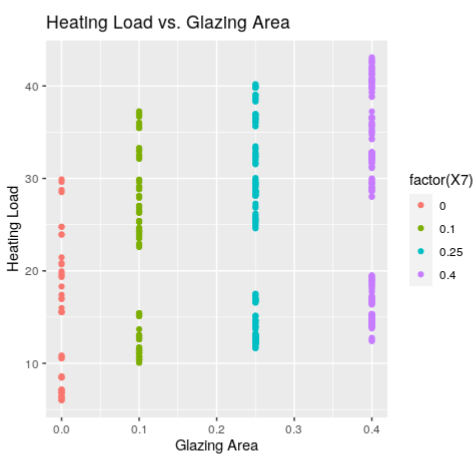


Mike Truong
Naila Hajiyeva
Naima Sagar
Michael Kobrosky
Rahul Ranganath
Varunika Selvakumar
Ahad Kesaria

7. As the glazing area of a building increases, cooling load, heating load and overall load will increase slightly as well, resulting in a positive regression coefficient in all three cases.

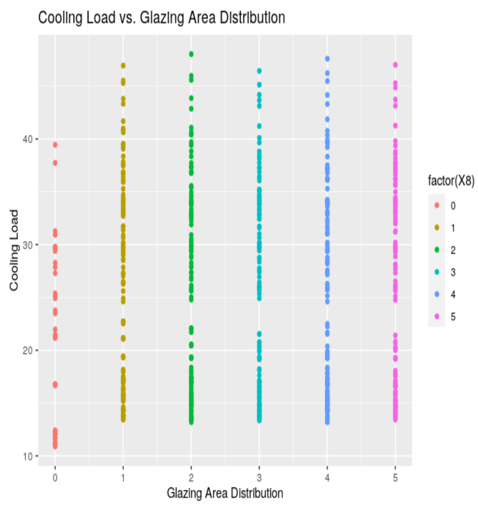


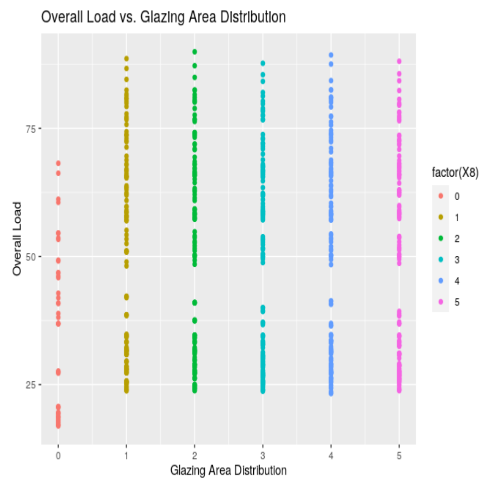


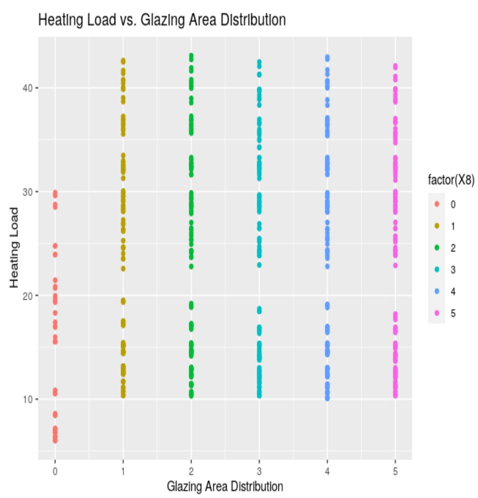


Mike Truong
Naila Hajiyeva
Naima Sagar
Michael Kobrosky
Rahul Ranganath
Varunika Selvakumar
Ahad Kesaria

8. The glazing area distribution in a building will have little effect on its cooling load, heating load and overall load. Thus the regression coefficient will be close to 0 for each distribution model in all three cases.







<u>Modeling</u>:

With the starting 8 independent variables and 2 dependent variables, we ran a couple of Python-based regressor algorithms against our data to determine the importance of each independent variable on the outcome of the 2 dependent variables. For this scenario, we utilized

Mike Truong
Naila Hajiyeva
Naima Sagar
Michael Kobrosky
Rahul Ranganath
Varunika Selvakumar
Ahad Kesaria

4 different regressor algorithms on our dataset - regular gradient boosting, AdaBoost, XGBoost, and Random Forest - and monitored the resulting feature importance estimation of each building feature on the heating and cooling loads. Each regressor was implemented exactly 1000 times to output 1000 different lists of feature importance percentage scores in order to minimize the effects of random number generators and random datapoint choice across all algorithms, but mostly to mitigate the impact of such effects on AdaBoost and XGBoost. The 1000 scores were then averaged across the 1000 trials to create a Mean Feature Importance Percentage Score list of this dataset produced by that specific regressor. The same process is replicated across all 4 regressors, with regular gradient boosting and AdaBoost each requiring to be split into 1-dimensional dependent variable input runs of heating and cooling load separately, and thus resulting in 6 different Feature Importance Percentage Score lists. These score lists were then ranked by descending order (i.e. from most to least important), and their corresponding feature names were extracted to create a cumulative comparison table between regressors. The table ranking goes as such - 0 is the most important feature, and 7 is the least.

Feature Importance Across of Dataset Across Four Different Regressor Algorithms, Six Different Regression Models

	Random Forest	RegularGradientBoosting on Heating	RegularGradientBoosting on Cooling	AdaBoost on Heating	AdaBoost on Cooling	XGBoost
0	Relative Compactness	Surface Area	Surface Area	Relative Compactness	Relative Compactness	Relative Compactness
1	Surface Area	Roof Area	Roof Area	Surface Area	Surface Area	Glazing Area
2	Overall Height	Relative Compactness	Relative Compactness	Overall Height	Overall Height	Roof Area
3	Roof Area	Overall Height	Overall Height	Glazing Area	Glazing Area	Wall Area
4	Glazing Area	Glazing Area	Glazing Area	Roof Area	Roof Area	Glazing Area Distribution
5	Wall Area	Wall Area	Wall Area	Wall Area	Wall Area	Orientation
6	Glazing Area Distribution	Glazing Area Distribution	Glazing Area Distribution	Glazing Area Distribution	Glazing Area Distribution	Surface Area
7	Orientation	Orientation	Orientation	Orientation	Orientation	Overall Height

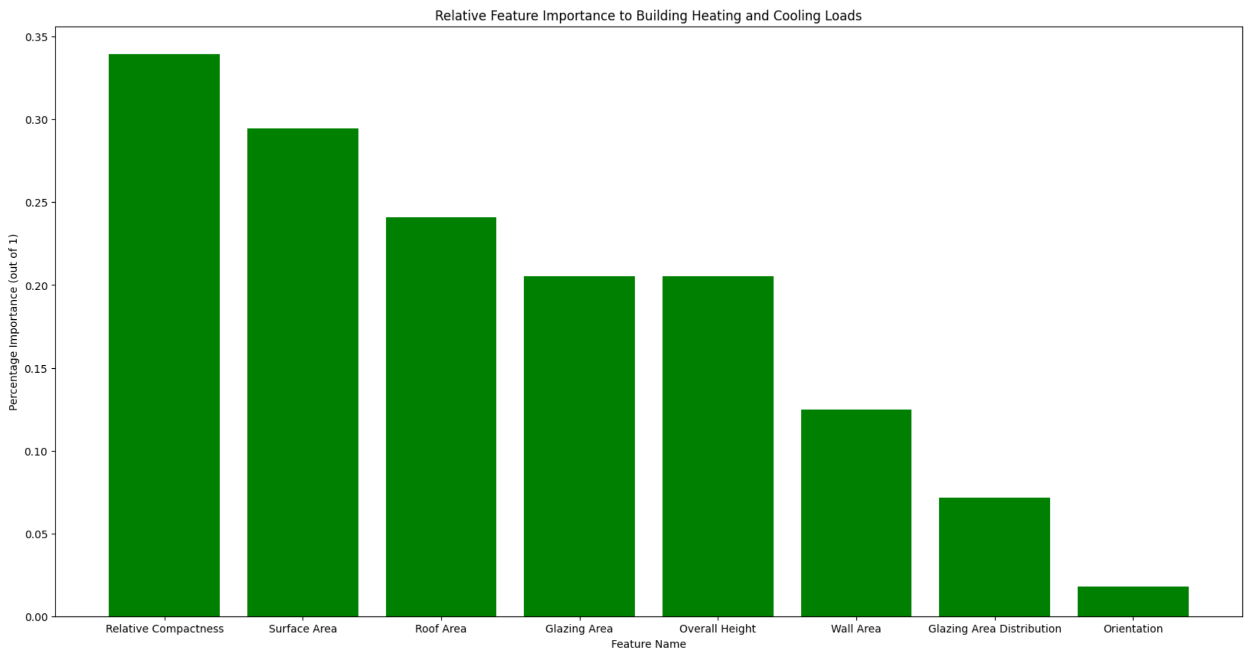
After compiling the ranked score feature list, it is not initially clear as to which order of importance each feature has when we consider the cumulative results from all 6 regression models. Thus an additional ranking step is necessary to completely interpret the cumulative feature importance ranking. We assigned weights to each individual feature name by assigning weighted scores for them as such (see matrix below):

```
[[7, 7, 7, 7, 7, 7],
 [6, 6, 6, 6, 6, 6],
 [5, 5, 5, 5, 5, 5],
 [4, 4, 4, 4, 4, 4],
 [3, 3, 3, 3, 3, 3],
 [2, 2, 2, 2, 2, 2],
 [1, 1, 1, 1, 1, 1]
```

Mike Truong
Naila Hajiyeva
Naima Sagar
Michael Kobrosky
Rahul Ranganath
Varunika Selvakumar
Ahad Kesaria

For the final round of score compilation and ranking, we summed up the scores across each feature name to see how much each feature scores and divided them against 112 (the total score available in the table), and ranked them on a final descending order list. The resulting ranking list is as such:

{'Relative Compactness': 0.3392857142857143, 'Surface Area': 0.29464285714285715, 'Roof Area': 0.24107142857142858, 'Glazing Area': 0.20535714285714285, 'Overall Height': 0.20535714285714285, 'Wall Area': 0.125, 'Glazing Area Distribution': 0.07142857142857142, 'Orientation': 0.017857142857142856}



Relative Compactness is by far the most important factor when it comes to effect on heating and cooling loads of buildings, followed not so closely by Surface Area, then Roof Area, then Glazing Area and Overall Height at a tie, then Wall Area, then Glazing Area Distribution, and then finally Orientation with little to no effect on building heating and cooling loads. Thus, in order to minimize heating and cooling loads (i.e. maximize energy efficiency) of a given building, one would most likely want to prioritize minimizing its relative compactness. This makes intuitive sense, as well, since the less compact a building is, the less empty area there inside the building to have to heat up or cool down.

Mike Truong
Naila Hajiyeva
Naima Sagar
Michael Kobrosky
Rahul Ranganath
Varunika Selvakumar
Ahad Kesaria

<u>Discussion</u>:

The findings of this analysis indicate that relative compactness and surface area to be the most significant contributors to a building's energy efficiency, with relative compactness ranking most important with a weight score of approximately 0.339 (calculated from the four regression algorithms) and surface area ranking second most important with a weight score of approximately 0.294. This indicates that the greatest effect on energy efficiency is from lower building compactness and lower surface areas free a critical point. Roof area is also a significant contributor with a weight score of approximately 0.241 and a negative relationship, indicating a larger roof area is correlated with a lower overall load. Glazing area and overall height have the same impact on overall load, with its weight being approximately 0.205 and higher glazing areas and overall height associated with higher overall loads. Wall area, glazing distribution and orientation have the least effect on energy efficiency with respective approximate weights of 0.125, 0.071, and 0.017, indicating these are not significant factors in energy efficient building design.

The results indicate that the most efficient building designs in Athens will have a low relative compactness with a low surface area after a critical point, higher roof area, smaller glazing areas, and lower overall height. All of these factors suggest that residential buildings that are designed in shapes that are smaller and shorter, such as domes and cubes, are more likely to keep hot or cool air in when number of inhabitants, hours of energy usage, and type of heating and cooling systems are controlled.

Although simulated building aspects provide a relatively accurate prediction of the trends in actual data, we cannot say with a guarantee that it accurately represents actual data in the real world. A solution to this limitation could be to obtain a random sample of residential buildings and measure their building attributes. Despite financial limitations, this would provide a more accurate representation of the data and allow for a better production of the effects of building attributes on the loads in their respective geographic location.

In terms of modeling, it was hard for us to cluster our model because there are too many conflicting variables that account for all of the overall heating and cooling load. It is not possible

Mike Truong
Naila Hajiyeva
Naima Sagar
Michael Kobrosky
Rahul Ranganath
Varunika Selvakumar
Ahad Kesaria

to cluster it because there's too many conflicting variables that account for all the overall heating and cooling load. There's too many compact variables so it's hard to get a cluster for it.

Conclusion:

The data indicates that relative compactness and surface area of a building contribute most to its overall load. Accounting for inhabitants, temperature setting range, hours of operation, and HVAC systems, the best way for nations to reduce their carbon footprint is to make buildings less compact and to minimize the surface area after a critical point. By creating less compact buildings with lower surface areas, the heating and cooling systems would require less energy since there is less surface area for heat/cool air to escape. The data also indicates that roof area plays a significant role, with lower roof areas indicating a larger heating and cooling load, which suggests that larger roof sizes should be used for more energy efficient buildings. Variables that contribute least to energy efficiency are orientation and glazing area distribution, indicating these factors should not be considered as significant when designing energy efficient buildings.

Future studies on the energy efficiency of buildings should take measurements and record data from real residential buildings in Athens and compare it with the findings from these simulated residential buildings. Overlapping similarities can be targeted to find factors that play a significant role in both simulated and existing residential buildings. Another future study that can be done on energy efficiency is to compare the factors contributing energy usage in commercial buildings as opposed to residential buildings, since single and multi-family homes are often designed differently from commercial buildings and require different energy needs. Lastly, research on energy efficiency will vary geographically since climate and lifestyles contribute to residential usage, so it may be beneficial to compare data from buildings of different countries.

Acknowledgment:

To begin the overall project, Naima and Ahad worked on the introduction and the data portion of the project, where they discussed the importance of the project and assessed how the data could be applied to the situation. After that, Rahul and Michael worked on the exploratory portion of the project, where they investigated the hypothesis and created graphs and assessed whether or not the hypothesis was supported or not. The modeling portion of the project was performed by Naila and Mike, where they ran multiple regression algorithms to investigate which variables

Mike Truong
Naila Hajiyeva
Naima Sagar
Michael Kobrosky
Rahul Ranganath
Varunika Selvakumar
Ahad Kesaria

were important in determining the heating and cooling load of the project. The discussion portion of the project was a collective effort but mostly done by Mike and Naila since they performed the regression analysis of the project. The limitations aspect of the discussion was done by Ahad and Rahul. The conclusion was wrapped up by Varunika as she investigated which ones were important and drew conclusions based on the regression analysis.

TEAM MEMBER	**PERCENT CONTRIBUTION (%)**
Ahad Kesaria	100
Michael Kobrosky	100
Mike Truong	100
Naima Sagar	100
Naila Hajiyeva	100
Rahul Ranganath	100
Varunika Selvakumar	100

Bibliography:

European Commission, Directive 2002/91/EC of the European Parliament and of the Council of 16th December 2002 on the energy performance of buildings, Official journal of the European Communities, L1/65–L1/71, 04/01/2003.

International Monetary Fund. 2022. “World Economic Outlook: Countering the Cost-of-Living Crisis”. Washington, DC, October.

Pérez-Lombard, Luis, José Ortiz, and Christine Pout. "A review on building energy consumption information." *Energy and buildings* 40.3 (2008): 394-398.

Tsanas, Athanasios, and Angeliki Xifara. "Accurate quantitative estimation of energy performance of residential buildings using statistical machine learning tools." *Energy and buildings* 49 (2012): 560-567.