Building Energy Efficiency : Capstone Project

Non Technical Audience

Presented By : Harshal Parikh

` Program : Masters in Data Analytics

Program Mentor : Jared Knepper

Introduction and Background

Introduction

Name : Harshal Parikh | Course: Masters in Data Analytics | Today's Date: 22nd February 2022

Student ID: 007091749 |Contact : hparikh@wgu.edu | Started: August 1st 2021

Background

Prior to this degree, I have about 5+ years of experience working in the oil and gas space along with a bachelors and masters in petroleum engineering. During my experience I encountered a lot of data and had the urge to develop a better understanding with data which I was able to obtain by pursuing a masters in data analytics.

Today my capstone project Building Energy Efficiency looks at ways to make buildings more energy efficient.

Introduction and Background

- Dataset has a total of 10 columns with 768 rows of which 8 attributes help determine Heating and Cooling Load of the building.

	Relative Compactness	Surface Area	Wall Area	Roof Area	Overall Height	Orientation	Glazing Area	Glazing Area Distribution	Heating Load	Cooling Load
0	0.98	514.5	294.0	110.25	7.0	2	0.0	0	15.55	21.33
1	0.98	514.5	294.0	110.25	7.0	3	0.0	0	15.55	21.33
2	0.98	514.5	294.0	110.25	7.0	4	0.0	0	15.55	21.33
3	0.98	514.5	294.0	110.25	7.0	5	0.0	0	15.55	21.33
4	0.90	563.5	318.5	122.50	7.0	2	0.0	0	20.84	28.28

- Relative Compactness, Surface Area, Wall Area, Roof Area, Overall Height, Glazing Area, Glazing Area Distribution, Heating Load and Cooling load.
- Variable used for PCA and regression analysis is Orientation

Statement of Problem and hypothesis

Statement of Problem

- "Could we identify the principal components that have the greatest impact on the Building Energy Efficiency?"

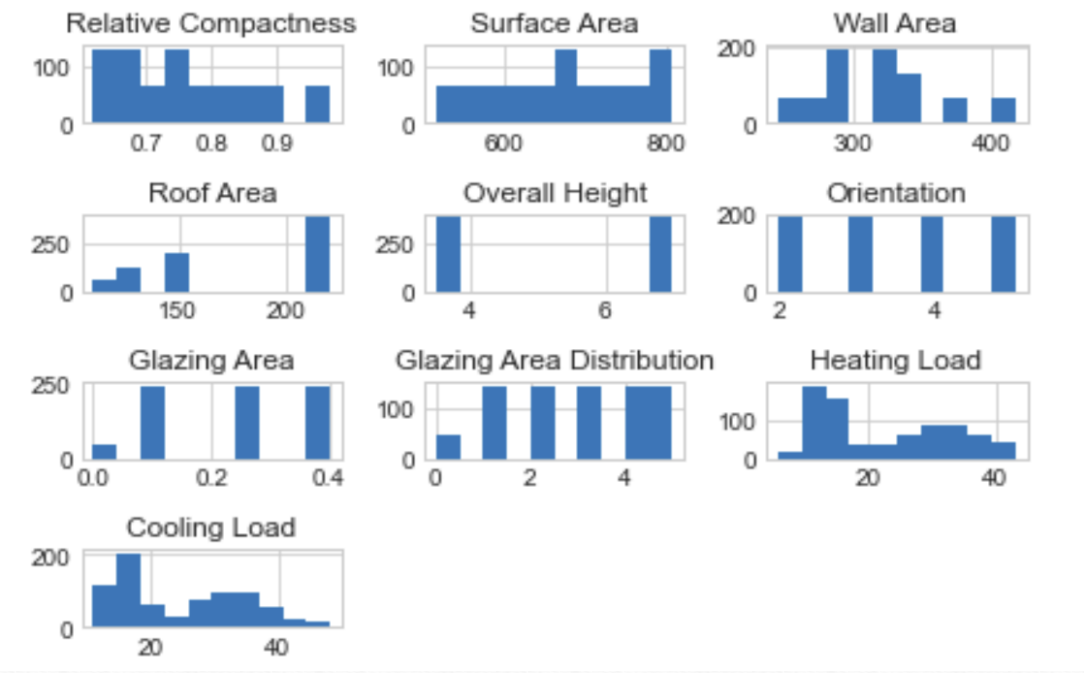
Context

- Bioclimatic architecture, high performing building envelopes and high-performance controlled ventilation.
- Characteristics of the buildings and spaces (including the occupancy and activity level)

Hypothesis

- The orientation of a building statistically significantly impacts the Heating and cooling of residential buildings.

Summary of the Data Analysis Process



Prior to performing Data Analysis:

- Ensuring the libraries and packages have been imported and the csv file has been read
- Renaming the columns and performing steps to ensure the data has been cleaned.
- Cleaning the data would constitute: Checking for empty rows/columns, unique elements and Duplicates

```
{'X1' : 'Relative Compactness',
 'X2' : 'Surface Area',
 'X3' : 'Wall Area',
 'X4' : 'Roof Area',
 'X5' : 'Overall Height',
 'X6' : 'Orientation',
 'X7' : 'Glazing Area',
 'X8' : 'Glazing Area Distribution',
 'Y1' : 'Heating Load',
 'Y2' : 'Cooling Load'}, inplace = True)
```

Summary of the Data Analysis Process

The variable used for PCA, and regression analysis was Orientation. Summarizing the data analysis process is a multistep process.

- Defining the variables followed by standardizing the data

```
# Defining the x and y variables
X = data.drop(["Orientation"], axis=1)
y = data[["Orientation"]]
```

```
# Importing Standard Scaler from Scikit learn
from sklearn.preprocessing import StandardScaler

# Standardize the data
X_standardized = StandardScaler().fit_transform(X)
```

- Creating a covariance matrix and performing an eigen decomposition on the covariance matrix.

Summary of the Data Analysis Process

OLS Regression Results

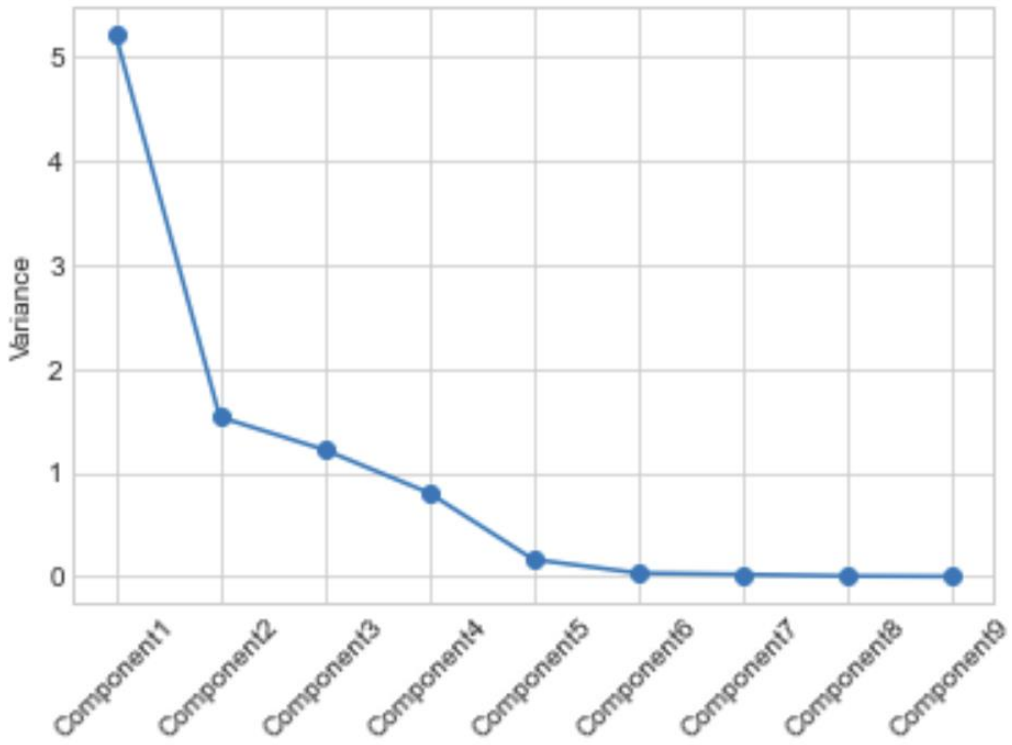
Dep. Variable:	Orientation	R-squared (uncentered):	0.908
Model:	OLS	Adj. R-squared (uncentered):	0.907
Method:	Least Squares	F-statistic:	937.8
Date:	Sun, 20 Feb 2022	Prob (F-statistic):	0.00
Time:	10:28:38	Log-Likelihood:	-1172.9
No. Observations:	768	AIC:	2362.
Df Residuals:	760	BIC:	2399.
Df Model:	8		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Relative Compactness	1.8766	0.732	2.565	0.011	0.440	3.313
Surface Area	0.0022	0.000	4.472	0.000	0.001	0.003
Wall Area	0.0004	0.002	0.261	0.794	-0.003	0.003
Roof Area	0.0009	0.001	0.922	0.357	-0.001	0.003
Overall Height	0.0131	0.113	0.116	0.908	-0.209	0.235
Glazing Area	0.1748	0.417	0.419	0.675	-0.644	0.994
Glazing Area Distribution	0.0071	0.027	0.262	0.793	-0.046	0.060
Heating Load	-0.0438	0.023	-1.929	0.054	-0.088	0.001
Cooling Load	0.0476	0.021	2.289	0.022	0.007	0.088

Omnibus:	2874.973	Durbin-Watson:	2.389
Prob(Omnibus):	0.000	Jarque-Bera (JB):	57.195
Skew:	-0.019	Prob(JB):	3.80e-13
Kurtosis:	1.664	Cond. No.	4.48e+16

Notes:
[1] R^2 is computed without centering (uncentered) since the model does not contain a constant.
[2] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[3] The smallest eigenvalue is 2.27e-25. This might indicate that there are strong multicollinearity problems or that the design matrix is singular.

Performing an OLS regression and developing an Initial Estimated Regression Equation



Performing an PCA. Listing the descending sorted eigenvalues, fitting a standardized matrix of features, and printing the explained variance ratio.

Outline of Findings

- Orientation having a major impact on cooling and heating load of the building was not plausible
- Overall height correlates a 0.89 and 0.9 to the heating and cooling load of the building.

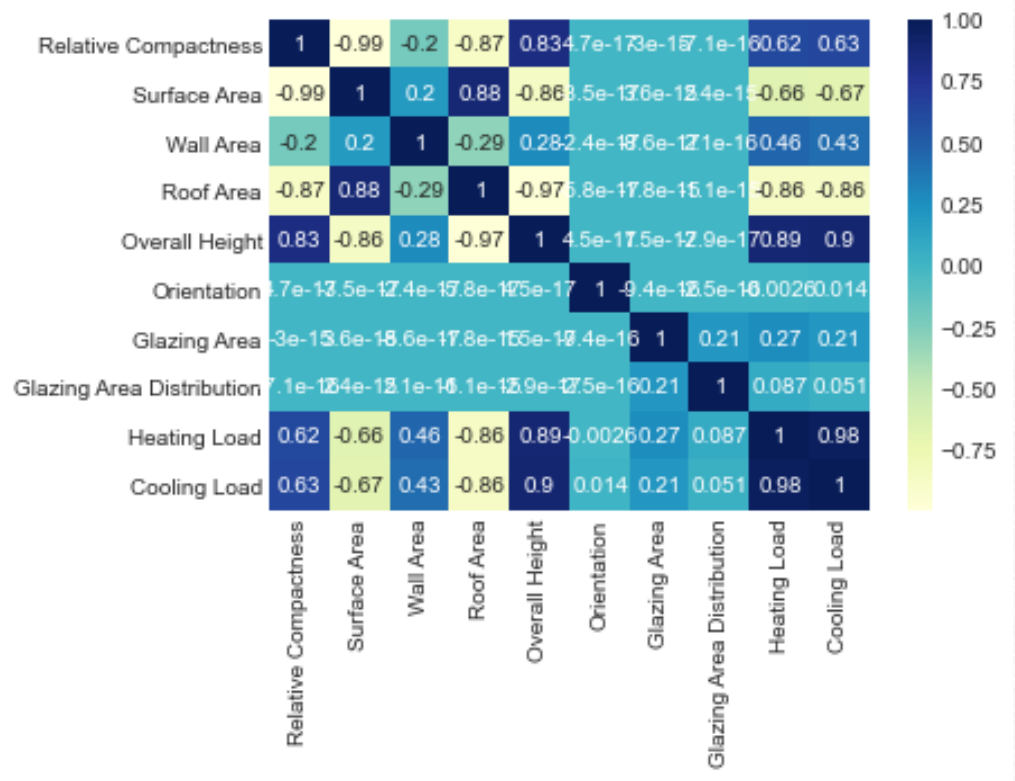
	Standard Deviation
PC1	5.222880e+00
PC2	1.533373e+00
PC3	1.218894e+00
PC4	8.047728e-01
PC5	1.631503e-01
PC6	3.309927e-02
PC7	1.947696e-02
PC8	4.353926e-03
PC9	1.365120e-30

Initial multiple regression model

Y* = 1.8766***Relative Compactness** + 0.0022***Surface Area** + 0.0004***Wall Area** + 0.0009***Roof Area** + 0.0131***Overall Height** + 0.1748***Glazing Area** + 0.0071***Glazing Area Distribution** -0.0438* **Heating Load** + 0.0476***Cooling Load**

*Where Y is Orientation

Outline of Findings: Heatmap



The seaborn heatmap gives the following correlation findings

With respect to the heating load

- Heating Load and Relative Compactness is 0.62
- Heating Load and Wall Area is 0.46
- Heating Load and Overall Height is 0.89
- Heating Load and Orientation is 0.0026

With respect to the Cooling Load

- Cooling load and Relative Compactness is 0.63
- Cooling load and Wall Area is 0.43
- Cooling load and Overall Height is 0.9
- Cooling load and Orientation is 0.014

Limitations - Techniques

Dataset Limitations

- Public Dataset – Out of Date (UCI Machine Learning Repository)
- Mosaic Effect

PCA Limitations

- Independent variables – becomes less interpretable
- Data standardization is must before PCA
- Information loss

OLS Regression Limitations

If you fit a linear model to a data that is non-linearly related, the model will be incorrect and hence unreliable. When you use the model for extrapolation, you are likely to get erroneous results

Summary of Proposed Action

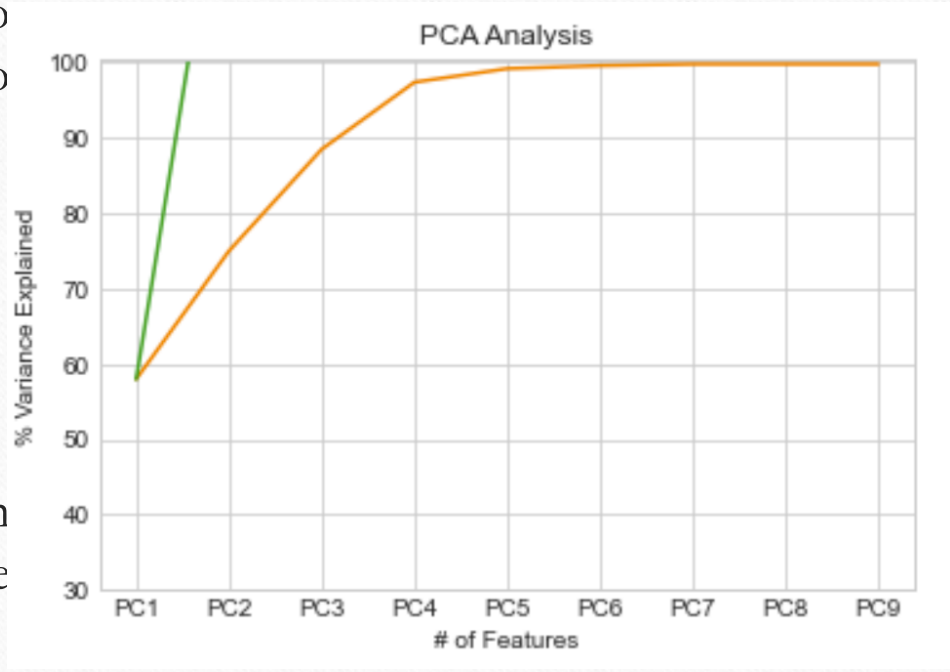
- Obtain more relevant and up-to-date information along with conducting Customer surveys and looking for ways to obtain additional data to improve stakeholder insight.
- Perform an analysis to see if additional parameters like glazing area, glazing area distribution, Wall Area, Roof Area and Surface Area prove to be statistically significant.
- Since time series analysis provides a specific way of analyzing the sequence of data points which were gathered over an interval of time that would be an additional proposed action.

Expected Benefits of the Study

Principal Component Analysis has been widely used to reduce the dimensionality of the data. We were able to identify

- 2 components give us more than 70% of the variance.
- 4 Components were able to give 95% of the variance
- 5 components give 100% of the variance of the data.

The study would also give stakeholders an understanding on what component has the highest impact in improving the energy efficiency of the building



References

1. What are Pros and Cons of PCA. Retrieved from https://www.i2tutorials.com/what-are-the-pros-and-cons-of-the-pca

2. Time Series Analysis: Definitions, Types, Techniques and when it is used. Retrieved from https://www.tableau.com/learn/articles/time-series analysis#:~:text=Time%20series%20analysis%20is%20a,data%20points%20intermittently%20or%20randomly

Panopto Recording

https://wgu.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=387afaea-09bd-4961-83f5-ae440177c2b7

Thank You for your Time