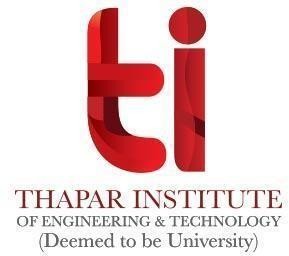
**Data Science Lab Report**

**THAPAR INSTITUTE OF ENGINEERING & TECHNOLOGY**



**BAN GLORE SMART BUILDING ANALYSIS**

# Data Visualization and Preprocessing

**Submitted To: Submitted By:**

Dr Divisha Singh Vaishnavi Tripathi-8024320118

Sonam-8024320111

Siddhanth-8024320104

Siddharth -8024320105

**DATA SCIENCES**

**BANGLORE SMART BUILDING ANALYSIS**

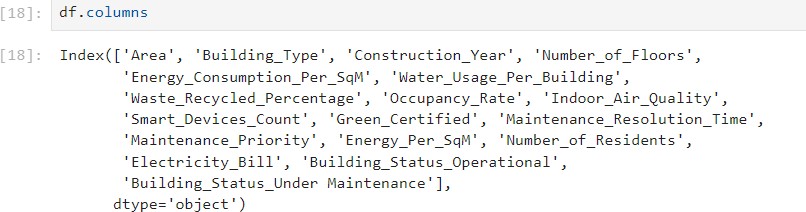
**About Dataset**

This dataset was meticulously compiled by aggregating data from multiple open sources available on government websites. The columns have been integrated to form a comprehensive dataset that provides insights into various aspects of smart city management. Each row represents a building within Bangalore, with data collected across different city areas, building types, and years of construction.

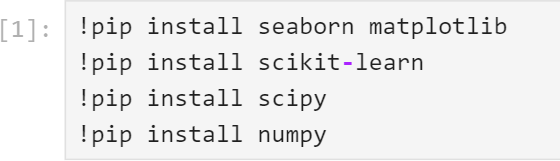
**Key Features**

|  |  |
| --- | --- |
| **Feature** | **Description** |
| city | Specifies the city, here focused on Bangalore. |
| area | Indicates the specific area within the city, such as Yelahanka, Whitefield, etc |
| Building\_type | Categorizes the building as Residential, Commercial, Institutional, or Industrial. |
| Construction\_ye ar | The year the building was constructed |
| Number\_of\_floo rs | The total number of floors in the building |
| Energy\_consum ption\_per\_sqM | The energy consumed per square meter of the building. |
| Water\_usage\_p er\_building | Total water usage in the building. |
| Water\_recycled\_ percentage | Percentage of waste recycled from the total waste generated. |
| Occupancy rate | The occupancy rate of the building. |
| Indoor\_air\_quali ty | Measures the air quality within the building. |

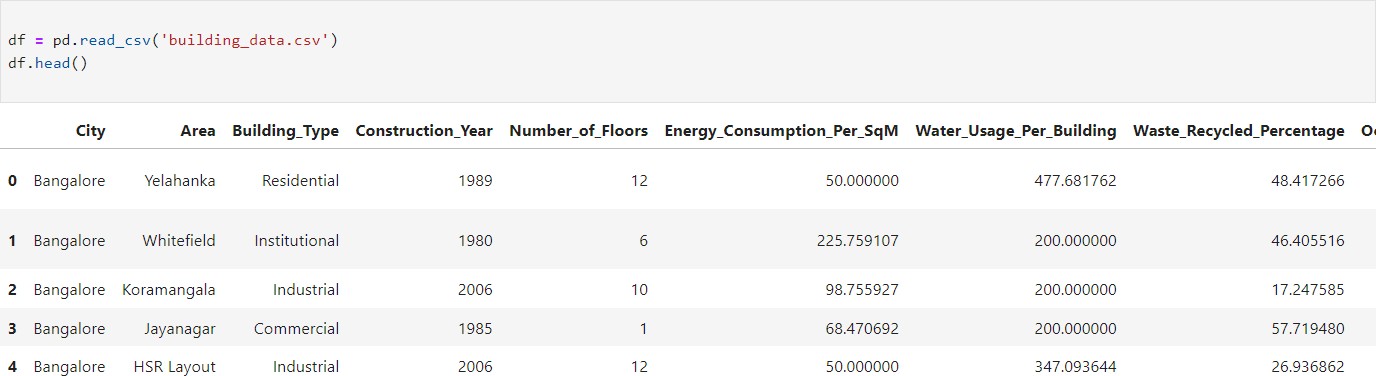
**We aim to predict energy consumption. COLUMNS CONSIDERED:**

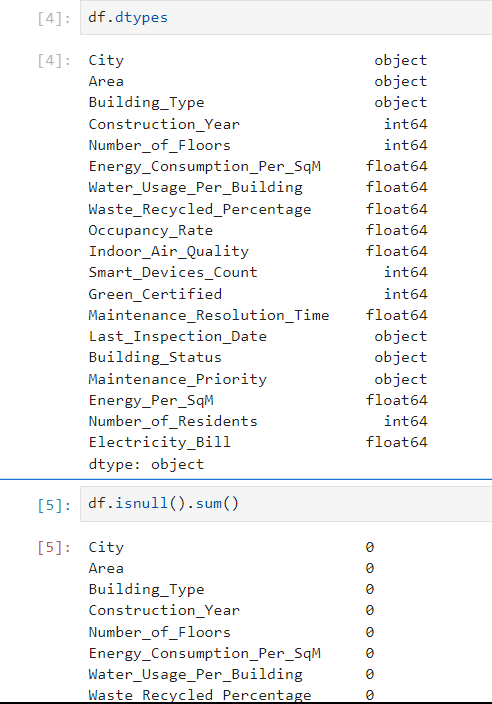


**IMPORT LIBRARIES And DEPENDENCIES**



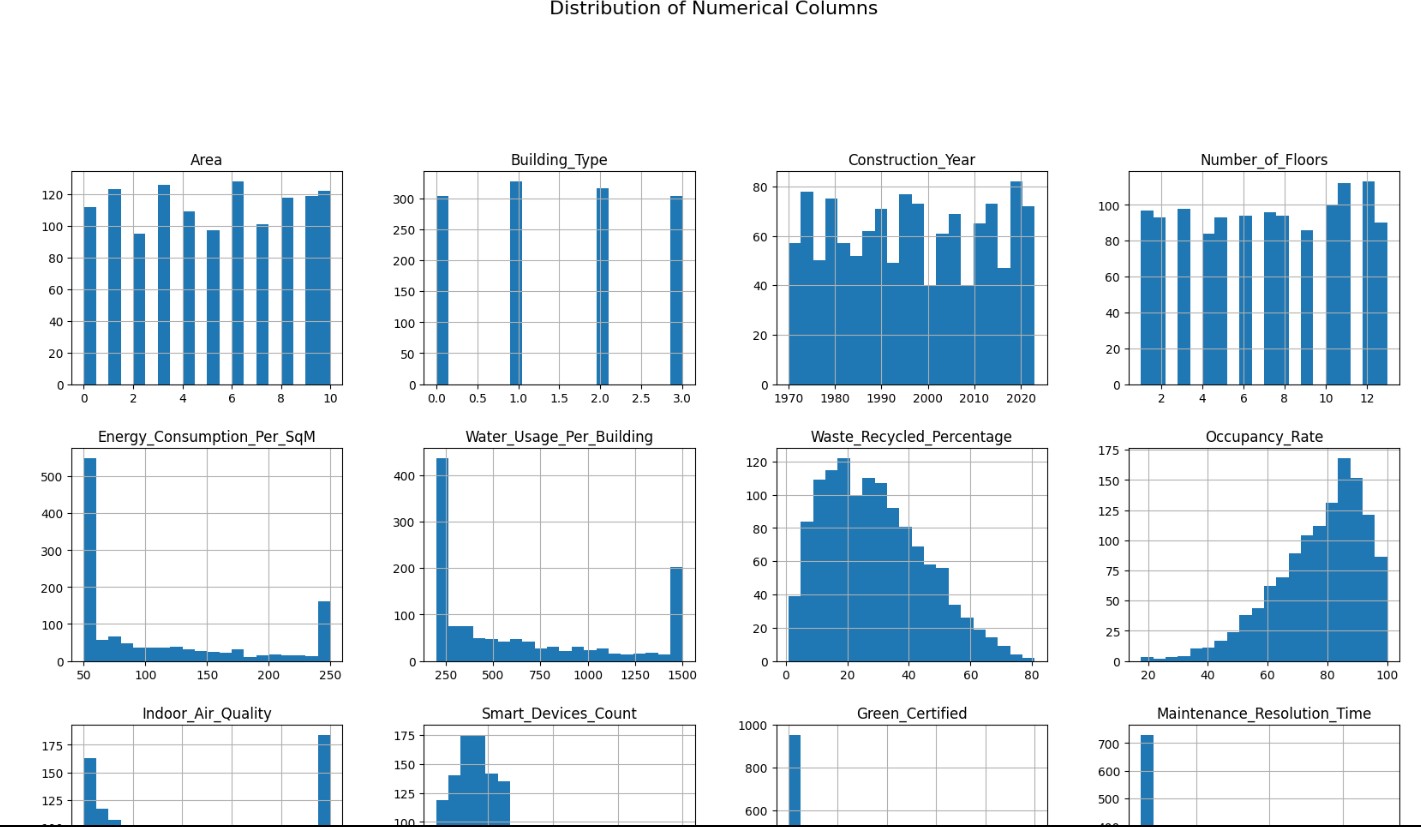
**DATA UNDERSTANDING**





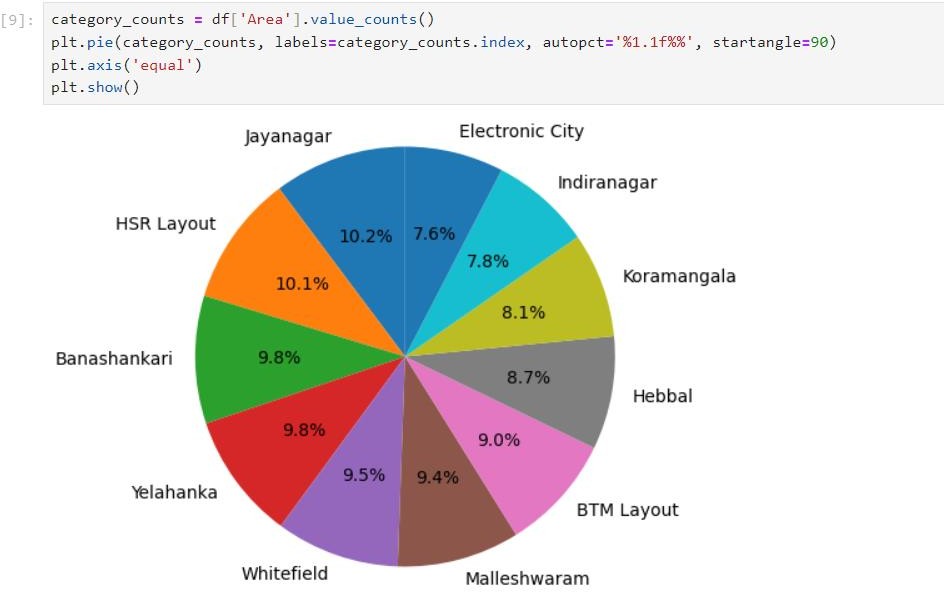
## 1.DATA VISUALIZATION

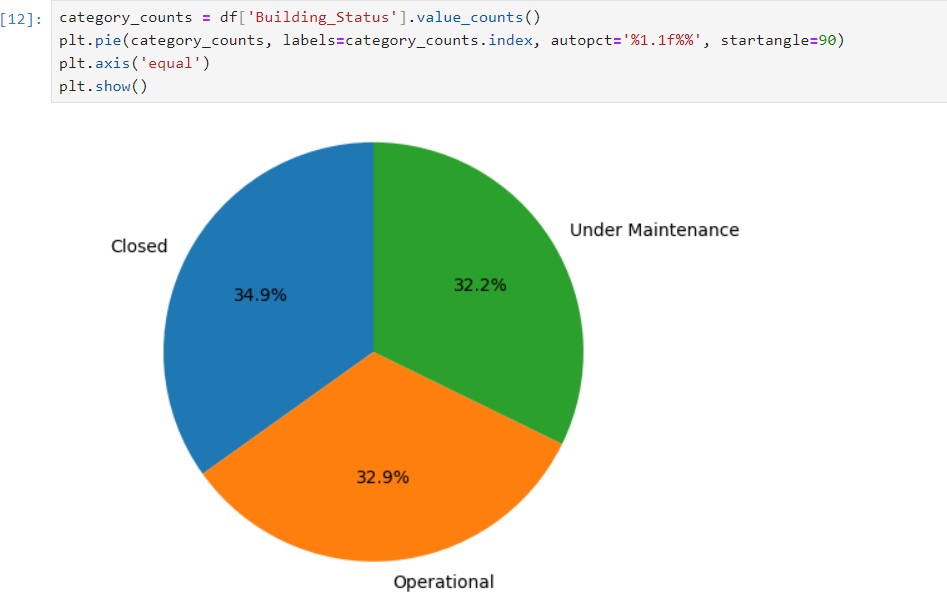
**HISTOGRAM :**



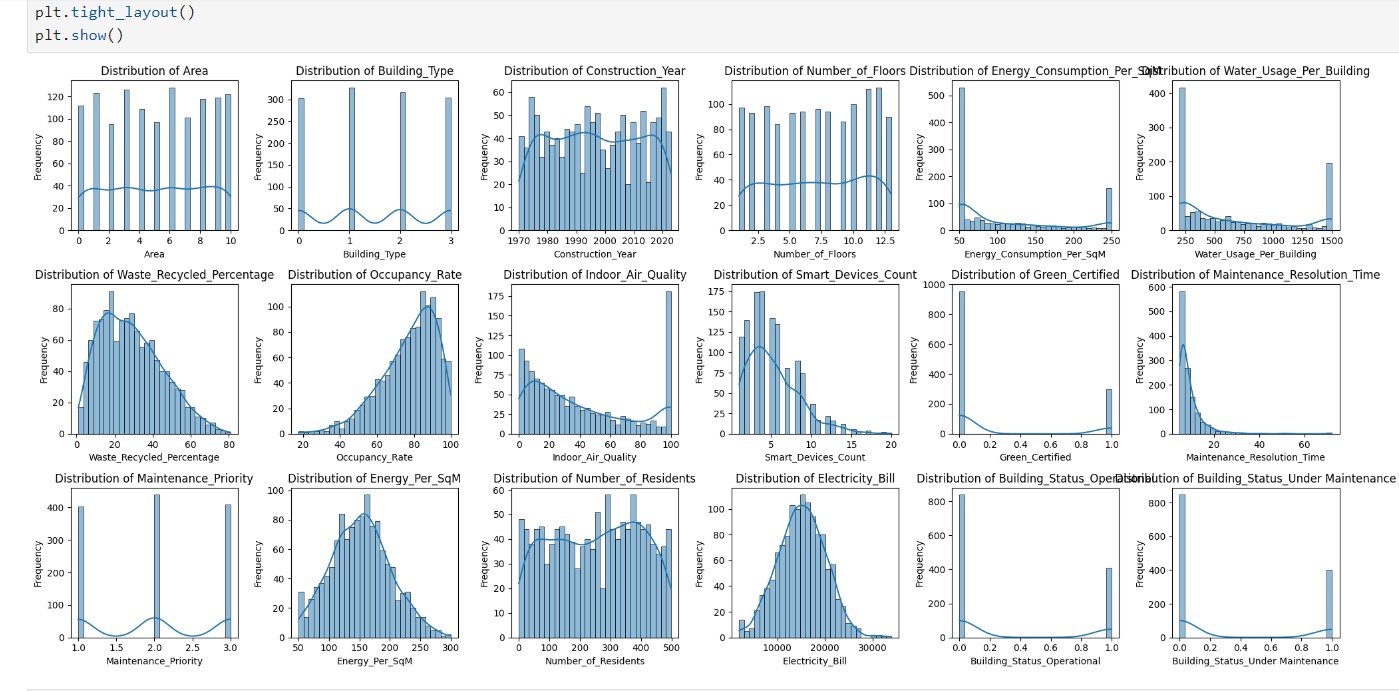


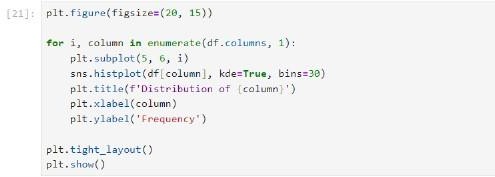
**PIE CHARTS :**



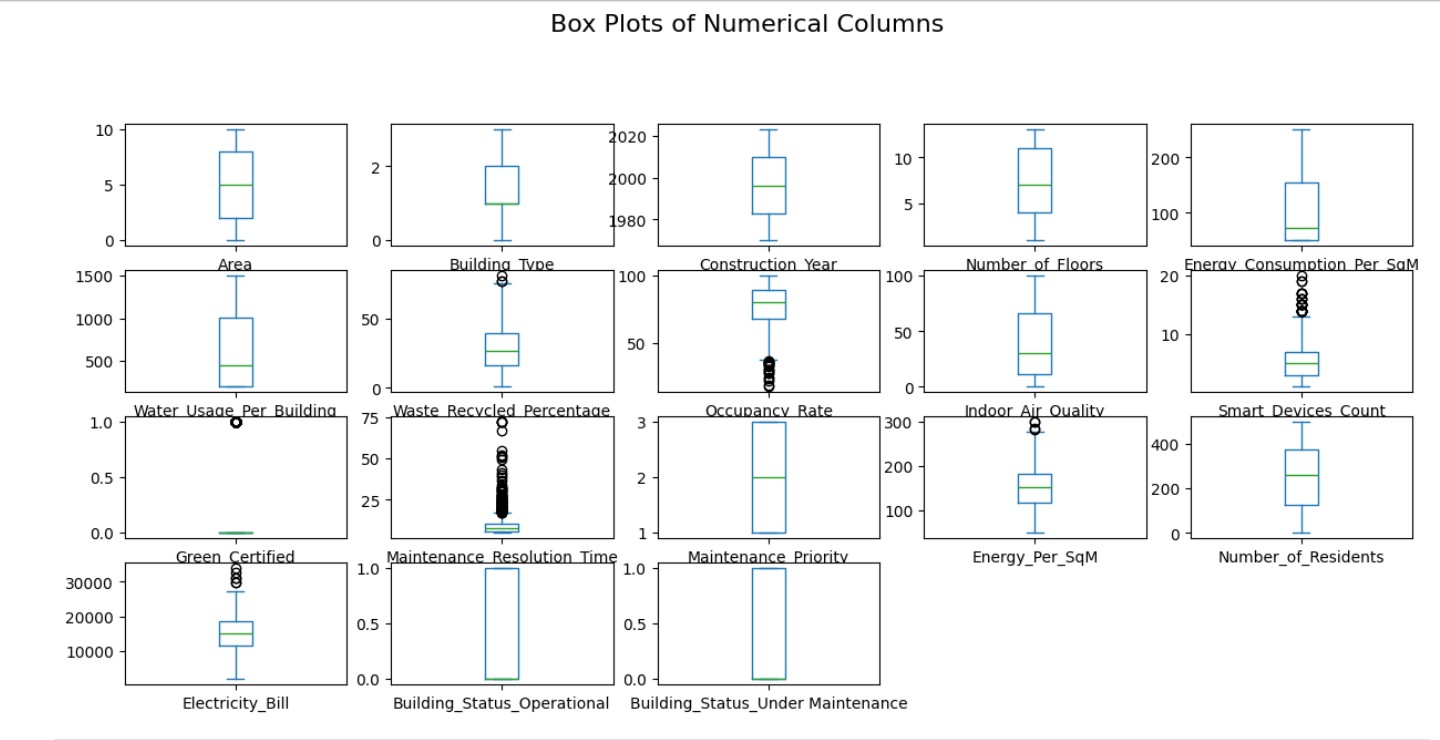


**DISTRIBUTION PLOTS:**





**BOX PLOT:**





## 2.DATA PREPROCESSING

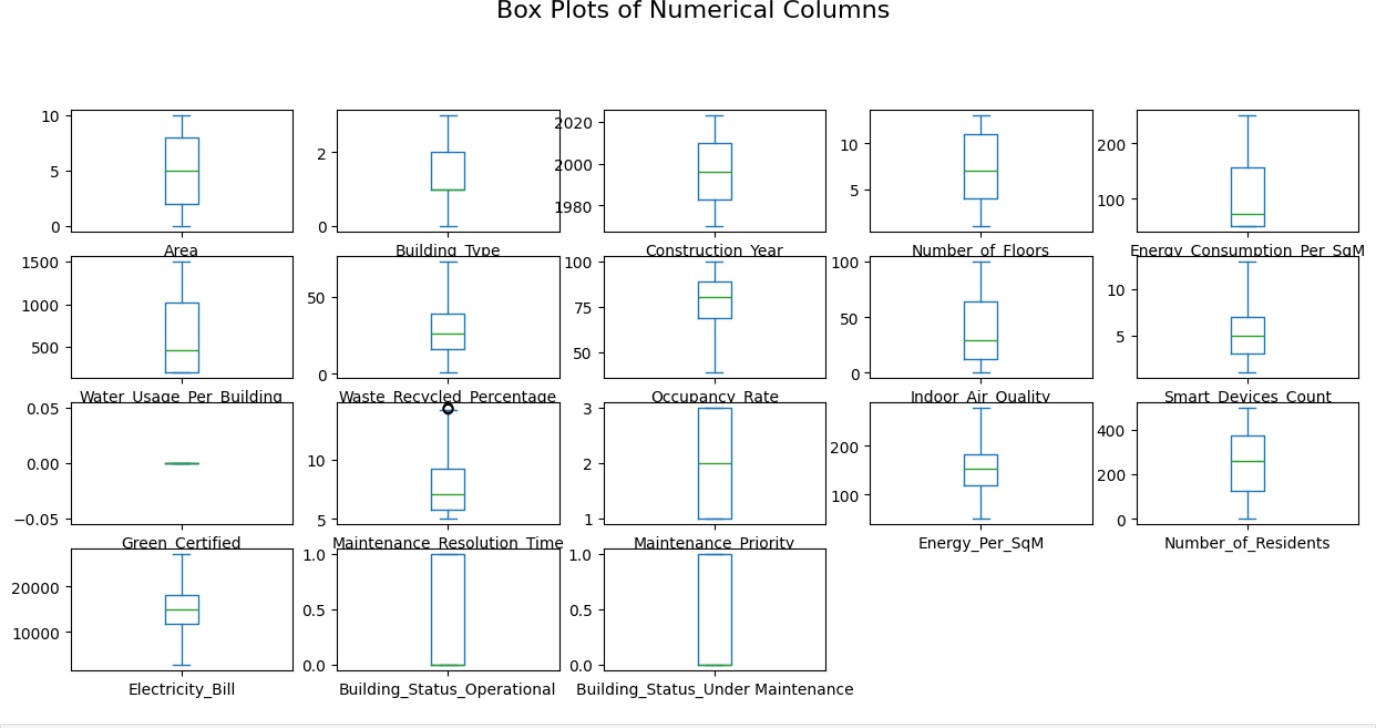
## Converting into numerical data (encoding):

## 

## 

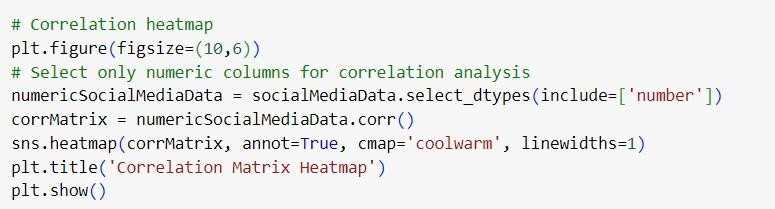
## We performed one hot and label encoding to convert our categorical data into numberical data.

**After removing Outliers:**



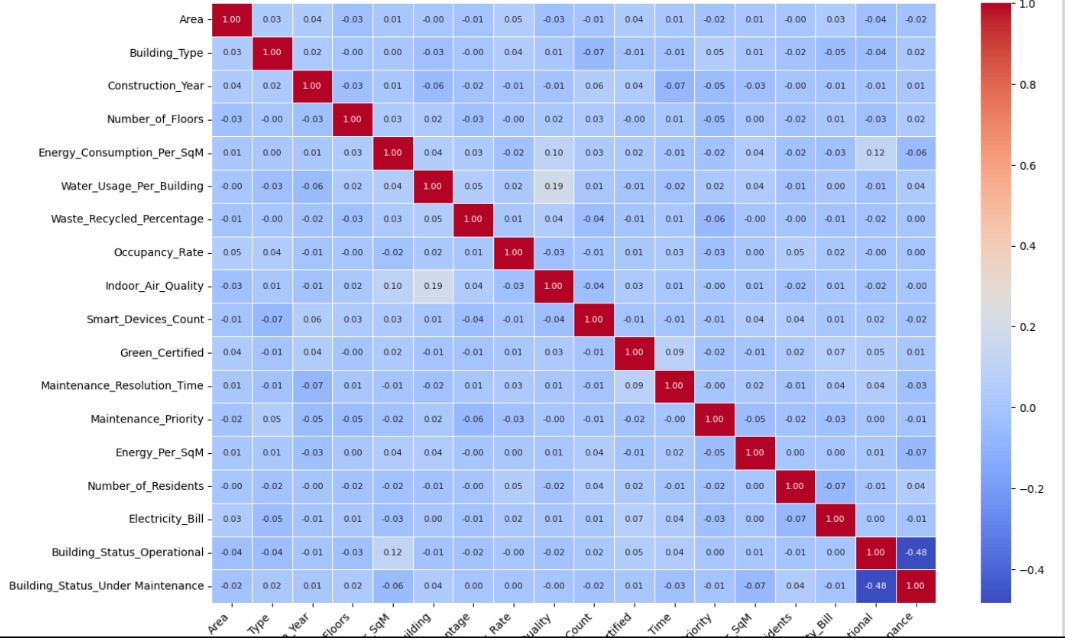


Lets check for **correlation** between different features: code:



CORRELATION

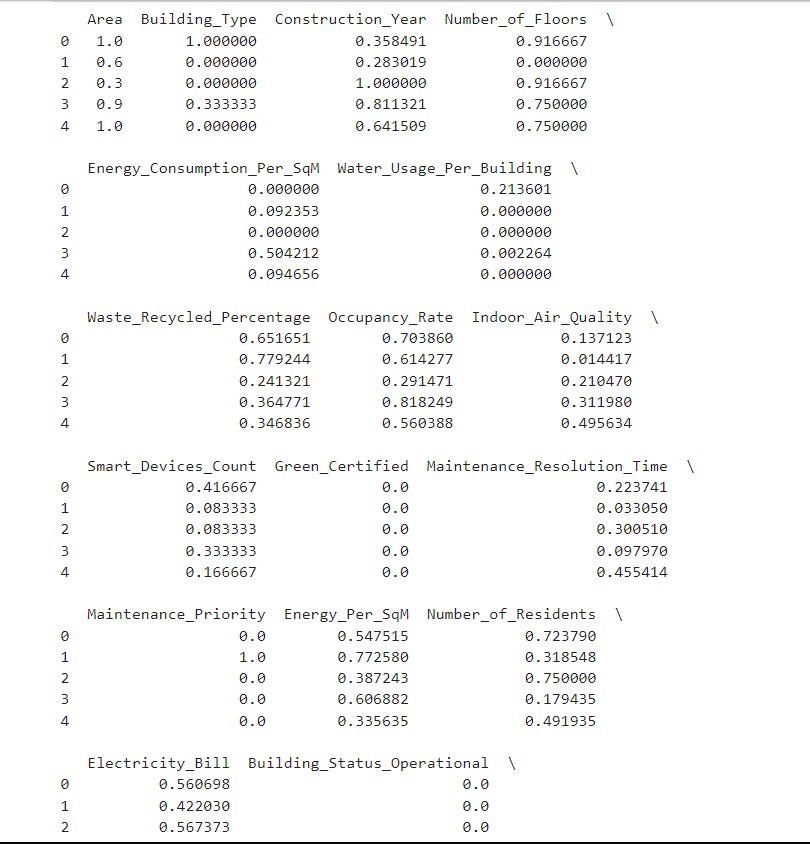
### BANHGLORE SMART BUILDING ANALYSIS



We can observe that among the list of columns, that we are considering Building\_status\_operational and Building\_status\_under\_maintainance are highly correlated (negative correlation), This make cause redundancy.We removed one.

# 3. Data NORMALIZATION





# 4. Model Training/Testing

# SVM(Support Vector Machine)

# 

# 

# Random Forest

# 

# 

# Linear Regression

# 

# 

**Conclusion**

The analysis of the Bangalore Smart Building dataset has provided meaningful insights into the energy consumption patterns and other critical attributes of buildings in the city. Through comprehensive data preprocessing, including encoding categorical variables, removing outliers, and normalizing data, we ensured the dataset was suitable for model training. Various machine learning algorithms, such as Support Vector Machines, Random Forest, and Linear Regression, were implemented, allowing us to evaluate their performance in predicting energy consumption.

Our findings underscore the importance of data-driven approaches in smart city management, particularly in optimizing energy use and improving sustainability measures. Future work could explore incorporating additional features like weather data or real-time energy usage metrics to enhance the model's predictive accuracy further. This project demonstrates the potential of data science in addressing urban challenges and contributes to building smarter and more efficient cities.