Data Collection and Preprocessing Phase

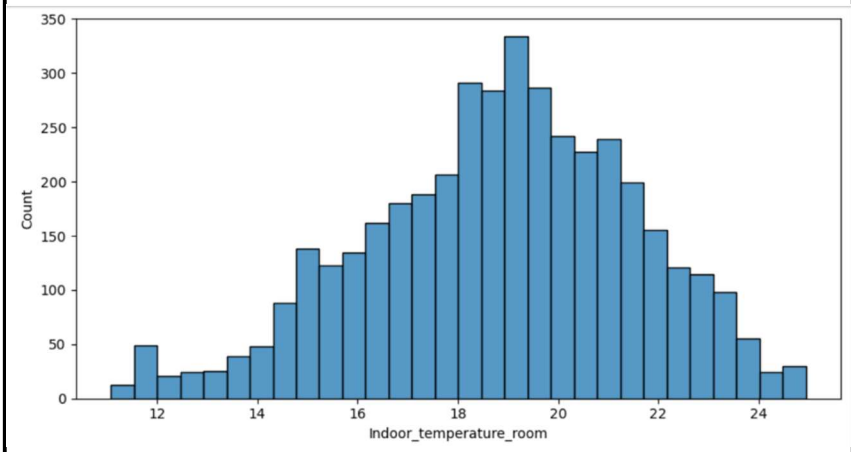
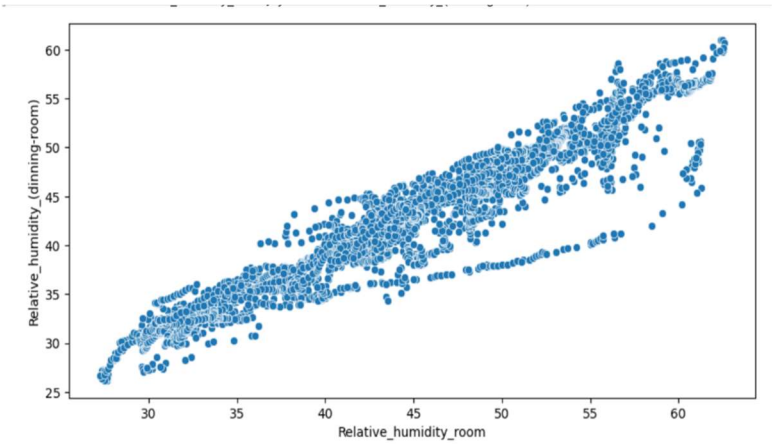
Date	18 June 2024
Team ID	739991
Project Title	Smart Home Temperature
Maximum Marks	6 Marks

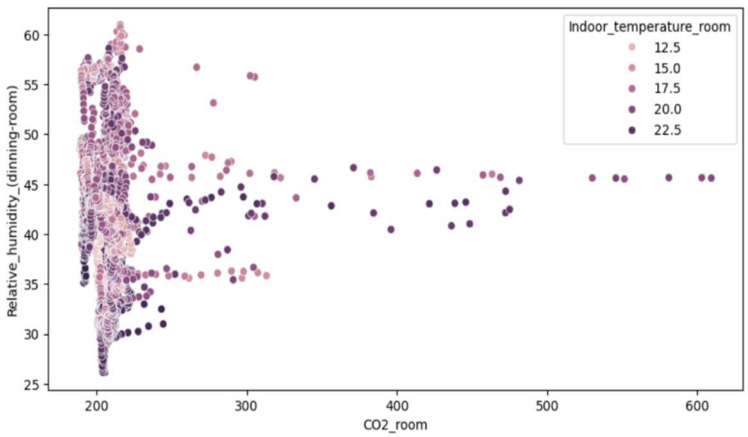
Data Exploration and Preprocessing Report

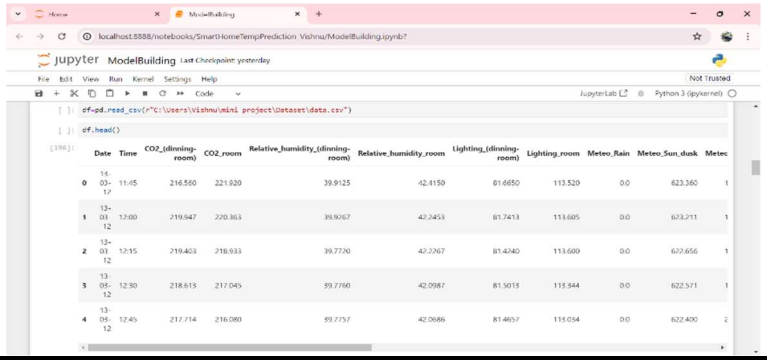
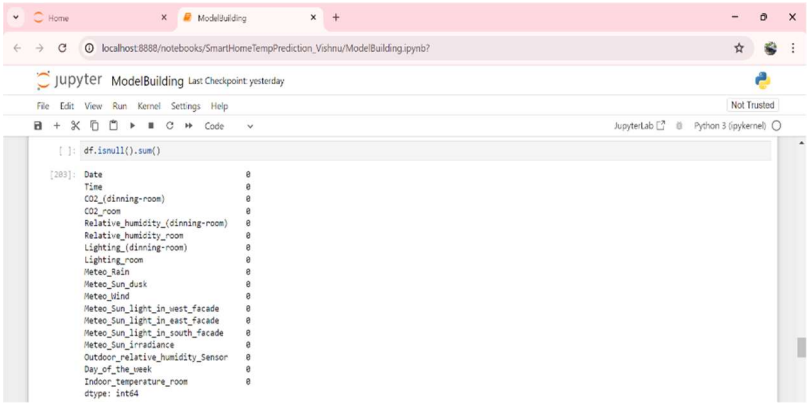
Dataset variables will be statistically analyzed to identify patterns and outliers, with Python employed for preprocessing tasks like normalization and feature engineering. Data cleaning will address missing values and outliers, ensuring quality for subsequent analysis and modeling, and forming a strong foundation for insights and predictions.

Section	Description
Data Overview	Dimension: 4137rows × 18 columns Descriptive statistics:
Univariate Analysis	

	CO2_(dinning-room)	CO2_room	Relative_humidity_(dinning-room)	Relative_humidity_room	Lighting_(dinning-room)	Lighting_room	Meteo_Rain	Meteo_Sun_dusk
count	4137.000000	4137.000000	4137.000000	4137.000000	4137.000000	4137.000000	4137.000000	4137.000000
mean	206.599835	209.611623	42.389879	44.546069	28.970248	42.335496	0.038756	335.094312
std	22.763114	24.183477	7.215405	8.297436	25.684356	42.602571	0.187128	304.513038
min	187.339000	188.907000	26.173300	27.256000	10.740000	11.328000	0.000000	0.606667
25%	200.228000	201.707000	36.088000	38.446700	11.540700	13.509300	0.000000	0.650000
50%	205.131000	208.907000	42.776000	44.802700	14.126700	22.085300	0.000000	612.821000
75%	210.016000	212.331000	47.584000	50.301300	40.034700	55.064000	0.000000	619.712000
max	594.389000	609.237000	60.957300	62.594700	111.797000	162.965000	1.000000	625.003000

	
Bivariate Analysis	

Multivariate Analysis	

Outliers and Anomalies	-
Data Preprocessing Code Screenshots	
Loading Data	
Handling Missing Data	

Data Transformation	`from sklearn.preprocessing import StandardScaler`
	`sc=StandardScaler()`
	`x_train_scaled= sc.fit_transform(x_train)`
	`x_test_scaled= sc.transform(x_test)`

Feature Engineering	Attached the codes in final submission.
Save Processed Data	-