Data Collection and Preprocessing Phase

Date	7 June 2024
Team ID	739871
Project Title	Smart Home Temperature Prediction using Machine Learning
Maximum Marks	6 Marks

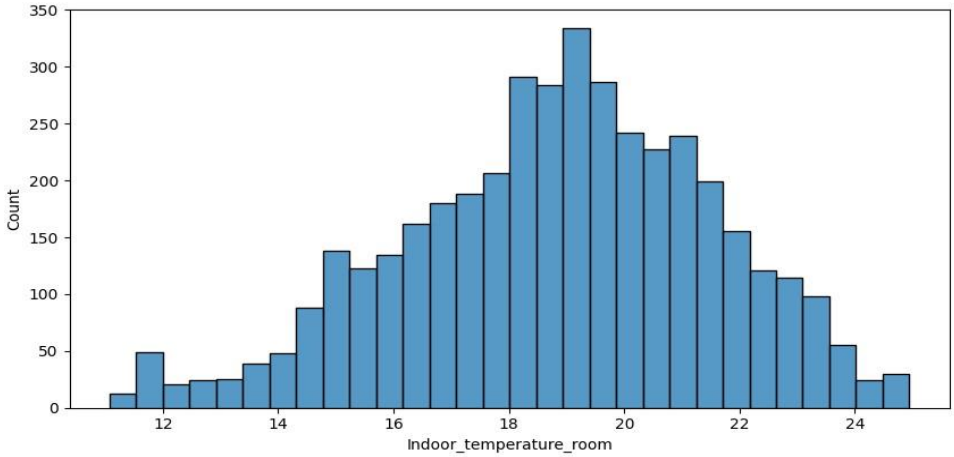
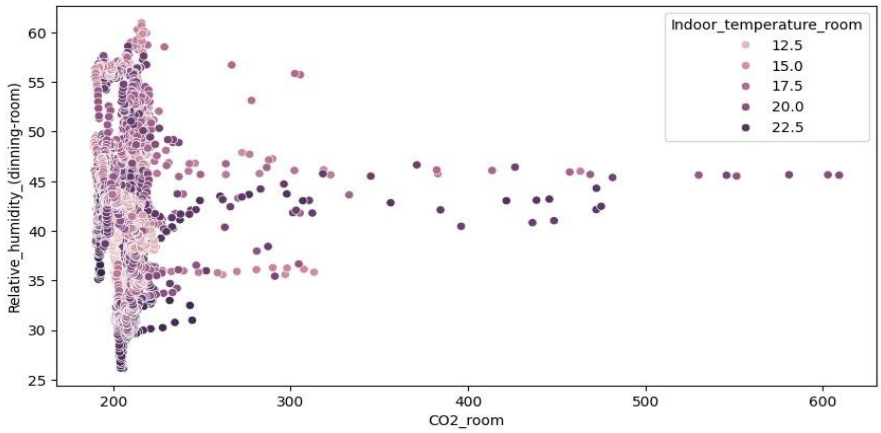
Data Exploration and Preprocessing Report

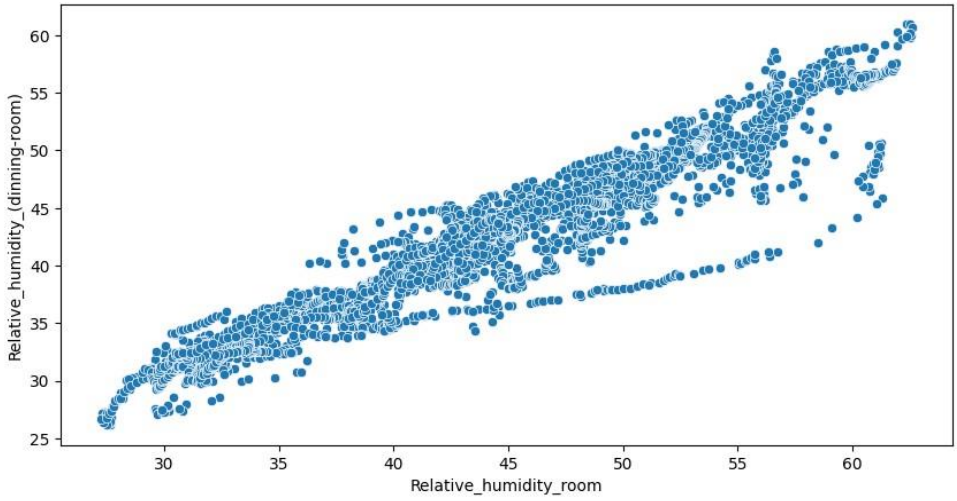
Dataset variables will be statistically analysed to identify patterns and outliers, with Python employed for preprocessing tasks like normalization and feature engineering. Data cleaning will address missing values and outliers, ensuring quality for subsequent analysis and modeling, and forming a strong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

Section	Description

Data Overview	Dimension: 4137 rows × 18 columns Descriptive statistics: 

	Date	Time	CO2_(dinning-room)	CO2_room	Relative_humidity_(dinning-room)	Relative_humidity_room	Lighting_(dinning-room)	Lighting_roc
0	13-03-12	11:45	216.560	221.920	39.9125	42.4150	81.6650	113.52
1	13-03-12	12:00	219.947	220.363	39.9267	42.2453	81.7413	113.60
2	13-03-12	12:15	219.403	218.933	39.7720	42.2267	81.4240	113.60
3	13-03-12	12:30	218.613	217.045	39.7760	42.0987	81.5013	113.34
4	13-03-12	12:45	217.714	216.080	39.7757	42.0686	81.4657	113.03
...	...	...	...	...	...	...	...	
32	02-05-12	06:30	199.424	201.963	43.0160	44.9813	21.8500	24.34
33	02-05-	06:45	199.200	202.091	43.1920	44.9413	21.1653	30.96

Univariate Analysis	
Multivariate Analysis	 -

Bivariate Analysis	

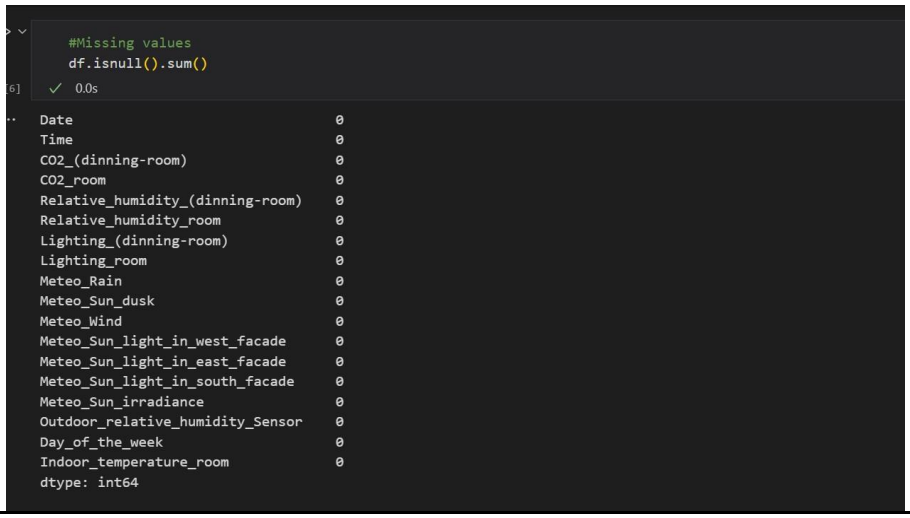
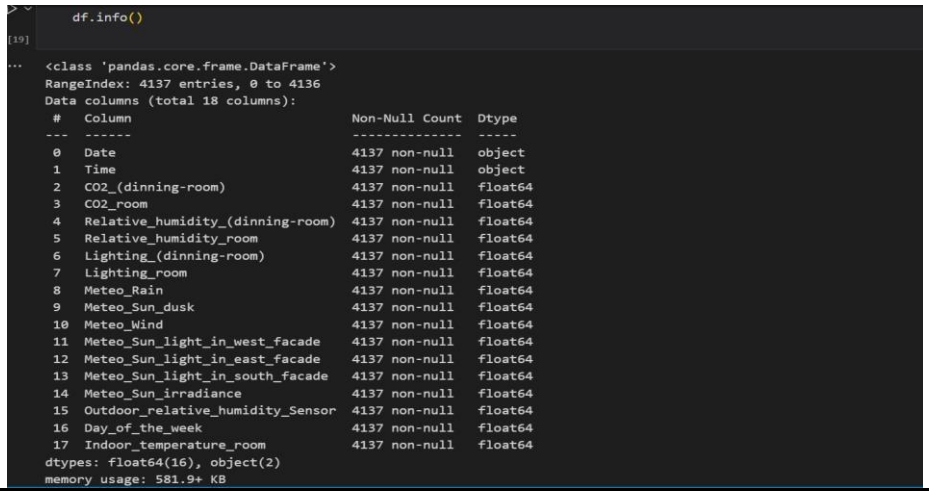
Data Preprocessing Code Screenshots

Loading Data

```
[10] df = pd.read_csv(r"C:\Users\nikky\Downloads\smarthome\data.csv")
                                                                    Python

[11] df.head()
                                                                    Python
```

	Date	Time	CO2 (dinning-room)	CO2 room	Relative humidity (dinning-room)	Relative humidity room	Lighting (dinning-room)	Lighting rc
0	13-03-12	11:45	216.560	221.920	39.9125	42.4150	81.6650	113
1	13-03-12	12:00	219.947	220.363	39.9267	42.2453	81.7413	113
2	13-03-12	12:15	219.403	218.933	39.7720	42.2267	81.4240	113
3	13-03-12	12:30	218.613	217.045	39.7760	42.0987	81.5013	113
4	13-03-12	12:45	217.714	216.080	39.7757	42.0686	81.4657	113

Handling Missing Data	
Handling Categorical Values	

Scaling the data	Attached the codes in final submission.
Splitting data into test and train	Attached the codes in final submission.