

Data Collection and Preprocessing Phase

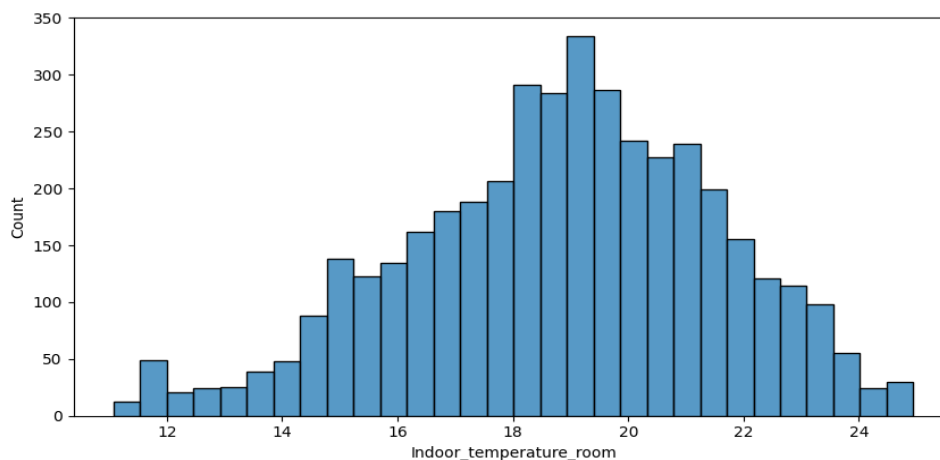
Date	7 June 2024
Team ID	739670
Project Title	Smart Home Temperature Prediction using Machine Learning
Maximum Marks	6 Marks

Data Exploration and Preprocessing Report

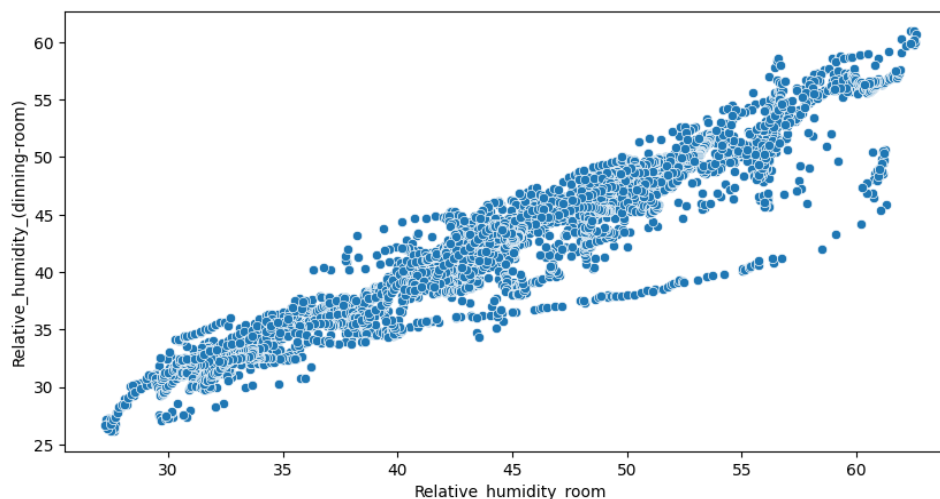
Dataset variables will be statistically analysed to identify patterns and outliers, with Python employed for preprocessing tasks like normalization and feature engineering. Data cleaning will address missing values and outliers, ensuring quality for subsequent analysis and modeling, and forming a strong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

Section	Description																																																																																
Data Overview	<div><div><div>Dimension:</div><div>4137 rows × 18 columns</div><div>Descriptive statistics:</div></div><table><tr><th></th><th>Date</th><th>Time</th><th>CO2 (dinning-room)</th><th>CO2_room</th><th>Relative humidity (dinning-room)</th><th>Relative_humidity_room</th><th>Lighting (dinning-room)</th><th>Lighting_rec</th></tr><tr><td>0</td><td>13-03-12</td><td>11:45</td><td>216.560</td><td>221.920</td><td>39.9125</td><td>42.4150</td><td>81.6650</td><td>113.52</td></tr><tr><td>1</td><td>13-03-12</td><td>12:00</td><td>219.947</td><td>220.363</td><td>39.9267</td><td>42.2453</td><td>81.7413</td><td>113.60</td></tr><tr><td>2</td><td>13-03-12</td><td>12:15</td><td>219.403</td><td>218.933</td><td>39.7720</td><td>42.2267</td><td>81.4240</td><td>113.60</td></tr><tr><td>3</td><td>13-03-12</td><td>12:30</td><td>218.613</td><td>217.045</td><td>39.7760</td><td>42.0987</td><td>81.5013</td><td>113.34</td></tr><tr><td>4</td><td>13-03-12</td><td>12:45</td><td>217.714</td><td>216.080</td><td>39.7757</td><td>42.0686</td><td>81.4657</td><td>113.03</td></tr><tr><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td></tr><tr><td>32</td><td>02-05-12</td><td>06:30</td><td>199.424</td><td>201.963</td><td>43.0160</td><td>44.9813</td><td>21.8500</td><td>24.34</td></tr><tr><td>33</td><td>02-05-12</td><td>06:45</td><td>199.200</td><td>202.091</td><td>43.1920</td><td>44.9413</td><td>21.1653</td><td>30.96</td></tr></table></div>		Date	Time	CO2 (dinning-room)	CO2_room	Relative humidity (dinning-room)	Relative_humidity_room	Lighting (dinning-room)	Lighting_rec	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	02-05-12	06:45	199.200	202.091	43.1920	44.9413	21.1653	30.96
		Date	Time	CO2 (dinning-room)	CO2_room	Relative humidity (dinning-room)	Relative_humidity_room	Lighting (dinning-room)	Lighting_rec																																																																								
	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	02-05-12	06:45	199.200	202.091	43.1920	44.9413	21.1653	30.96																																																																								

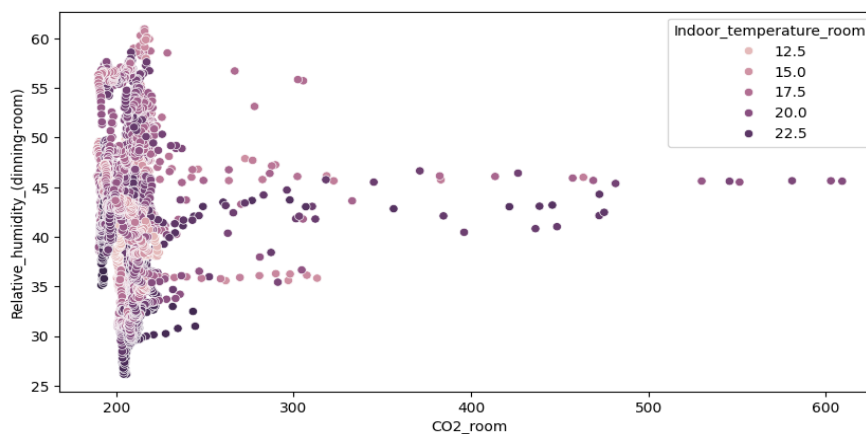
Univariate Analysis



Bivariate Analysis



Multivariate Analysis



Data Preprocessing Code Screenshots

Loading Data

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

df.head()
Python

...

```

	Date	Time	CO2_(dinning-room)	CO2_room	Relative_humidity_(dinning-room)	Relative_humidity_room	Lighting_(dinning-room)	Lighting_room
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

```
#Missing values
df.isnull().sum()
[6] ✓ 0.0s

Date      0
Time      0
CO2_(dinning-room)  0
CO2_room  0
Relative_humidity_(dinning-room)  0
Relative_humidity_room  0
Lighting_(dinning-room)  0
Lighting_room  0
Meteo_Rain  0
Meteo_Sun_dusk  0
Meteo_Wind  0
Meteo_Sun_light_in_west_facade  0
Meteo_Sun_light_in_east_facade  0
Meteo_Sun_light_in_south_facade  0
Meteo_Sun_irradiance  0
Outdoor_relative_humidity_Sensor  0
Day_of_the_week  0
Indoor_temperature_room  0
dtype: int64
```

Handling Categorical Values

```
df.info()
[19]

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4137 entries, 0 to 4136
Data columns (total 18 columns):
#   Column                                     Non-Null Count  Dtype
---  ---
0   Date                                     4137 non-null   object
1   Time                                     4137 non-null   object
2   CO2_(dinning-room)                     4137 non-null   float64
3   CO2_room                                4137 non-null   float64
4   Relative_humidity_(dinning-room)        4137 non-null   float64
5   Relative_humidity_room                  4137 non-null   float64
6   Lighting_(dinning-room)                 4137 non-null   float64
7   Lighting_room                           4137 non-null   float64
8   Meteo_Rain                              4137 non-null   float64
9   Meteo_Sun_dusk                          4137 non-null   float64
10  Meteo_Wind                              4137 non-null   float64
11  Meteo_Sun_light_in_west_facade           4137 non-null   float64
12  Meteo_Sun_light_in_east_facade           4137 non-null   float64
13  Meteo_Sun_light_in_south_facade          4137 non-null   float64
14  Meteo_Sun_irradiance                     4137 non-null   float64
15  Outdoor_relative_humidity_Sensor         4137 non-null   float64
16  Day_of_the_week                          4137 non-null   float64
17  Indoor_temperature_room                  4137 non-null   float64
dtypes: float64(16), object(2)
memory usage: 581.9+ KB
```

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