# Data Collection and Preprocessing Phase
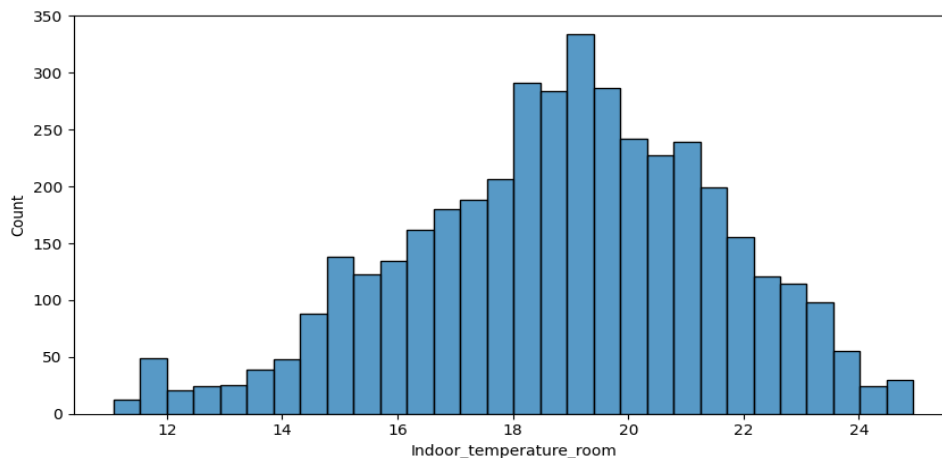
| | |
|---|---|
| Date | 7 June 2024 |
| Team ID | 740117 |
| Project Title | Smart Home Temperature Prediction using Machine Learning |
| Maximum Marks | 6 Marks |

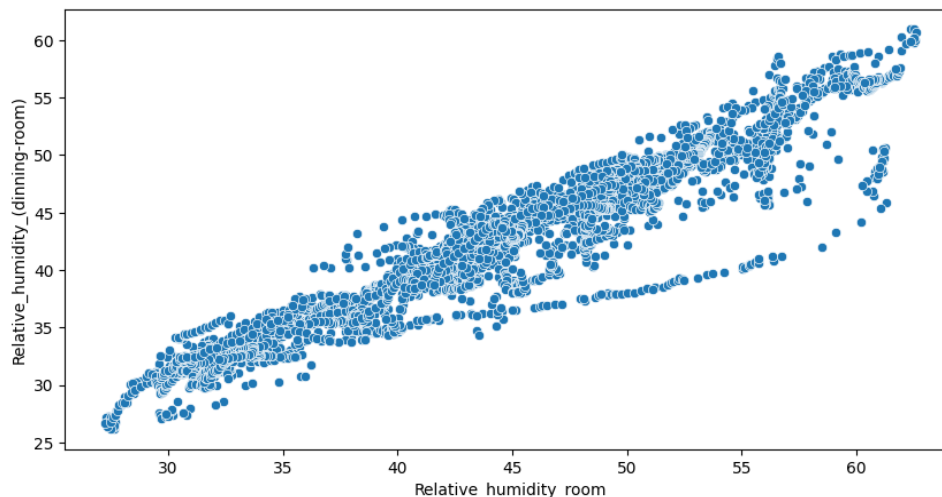**Data Exploration and Preprocessing Report**

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:<br>4137 rows × 18 columns<br>Descriptive statistics:<br> |

| | |
|---|---|
| Univariate Analysis |  |
| Bivariate Analysis |  |
| Multivariate Analysis |  |
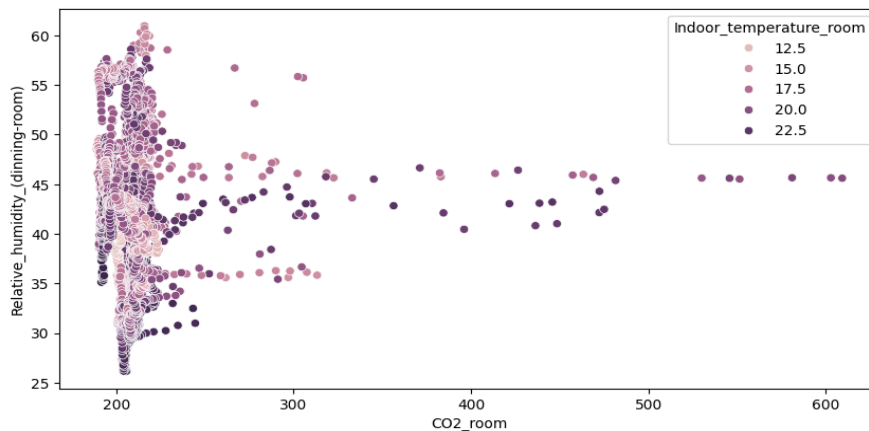
## Data Preprocessing Code Screenshots

| | |
|---|---|
| **Loading Data** | ```python
df = pd.read_csv(r"C:\Users\nikky\Downloads\smarthome\data.csv")
```<br><br>```python
df.head()
```<br><br>See table below |

**Loading Data output:**

| | 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**

```python
#Missing values
df.isnull().sum()
```
✓ 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**

```python
df.info()
```

```
<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. |