Content Page

1. Pre-Project Plan
2. Monitor
3. Introduction
4. Data Cleaning
5. Exploratory Data Analysis

1. Further Insights
2. Data Modeling
3. Conclusion
4. Reflection
5. References (if any)

## Pre-Project Plan

|  |
| --- |
| **Goal Setting** |
| I aim to complete my project by 24*/12/2021.*  I shall take initiative to find out the information needed.  I shall check the project rubric to ensure all items are done before submission. |

My data set is Appliances Energy – March and May.

My preliminary questions that I will answer from my data set:

1. Which day of the month do we have the lowest average temperature?
2. Which day of the month do we have the most ideal average relative humidity (30-50%)?
3. What is the relationship between relative humidity and temperature?
4. Is there a correlation between relative humidity and dew point temperature? A low dew-point temperature of 16 is ideal so that peoples’ perspiration would be at its fastest rate.
5. How does pressure affect windspeed? Are there other factors which strongly affect windspeed?

**Monitor**

|  |  |  |  |
| --- | --- | --- | --- |
| **Task/Milestone** | **By When** | **Actual Completed Date** | **Comment**  (on-time/delay/early) |
| Download the data.  Understand the rows and columns. | **21 Nov 2021** | **12 Dec 2021** | **Delay** |
| Background research of delivery mode, function of weather. | **28 Nov 2021** |  |  |
| Perform data cleaning. | **3 Dec 2021** |  |  |
| Perform data transformation. | **8 Dec 2021** |  |  |
| Exploratory Data Analysis | **17 Dec 2021** |  |  |
| Submit Report 1 | **24 Dec 2021** |  |  |
| Answer my preliminary questions | **9 Jan 2022** |  |  |
| Data modeling | **10 Jan 2022** |  |  |
| Final report conclusion and reflection | **14 Jan 2022** |  |  |
| Create Dashboard | **17 Jan 2022** |  |  |

1. **Introduction**

Questions

Provide some background information. Refer to the question description. Include some other research that are related to the topic. Could be why the data was collected? Or how the data was collected? Or what can we learn from the data? Or any past analysis done before?

Include your research references in section 10 of the report.

Give an overview of the data structure.

Describe your scenario (can include google map or diagram to illustrate)

Describe the data variables that was collected.

List total number of observations and variables

Some research on the background

Your question related to the data

The data set is at 10 min for about 4.5 months. The house temperature and humidityconditions were monitored with a ZigBee wireless sensor network. Each wireless nodetransmitted the temperature and humidity conditions around 3.3 min. Then, the wirelessdata was averaged for 10 minutes periods. The energy data was logged every 10 minuteswith m-bus energy meters. Weather from the nearest airport weather station (ChievresAirport, Belgium) was downloaded from a public data set from Reliable Prognosis (rp5.ru),

and merged with the experimental data sets using the date and time column.

This data contains information on the temperature, pressure, and relative humidity for the months of March and May respectively. For any weather data, knowing the geographical location would be useful as well so that we would know what season that region is experiencing. We would be assessing the data based on the populace satisfaction ie. to find weather conditions that people would prefer. We would also predict the future weather conditions and check how accurate our results are to the actual data.

There are 8317 observations (rows of data) and 27 variables (columns headers).

|  |  |  |  |
| --- | --- | --- | --- |
| Name of Variable | Data Type | Minimum Value | Maximum Value |
| date | Categorical | 1/3/2016 00:00 | 31/5/2016 23:50 |
| Appliances | Numeric | 20 | 880 |
| lights | Numeric | 0 | 50 |
| T1 | Numeric | 18.6 | 26.26 |
| RH\_1 | Numeric | 27.02333333 | 54.66666667 |
| T2 | Numeric | 16.2 | 29.85666667 |
| RH\_2 | Numeric | 20.46333333 | 54.09 |
| T3 | Numeric | 19.1 | 29.236 |
| RH\_3 | Numeric | 28.76666667 | 47.69333333 |
| T4 | Numeric | 17.29 | 26.2 |
| RH\_4 | Numeric | 28.716 | 50.76 |
| T5 | Numeric | 17.23 | 25.795 |
| RH\_5 | Numeric | 29.815 | 90.0 |
| T6 | Numeric | -3.59 | 28.29 |
| RH\_6 | Numeric | 1.0 | 86.96666667 |
| T7 | Numeric | 17.39 | 26.0 |
| RH\_7 | Numeric | 23.39 | 48.29 |
| T8 | Numeric | 19.0 | 27.23 |
| RH\_8 | Numeric | 29.79 | 55.0 |
| T9 | Numeric | 17.26 | 24.5 |
| RH\_9 | Numeric | 29.85 | 51.536 |
| T\_out | Numeric | -3.0 | 26.1 |
| Press\_mm\_hg | Numeric | 738.8 | 769.9 |
| RH\_out | Numeric | 24.0 | 100.0 |
| Windspeed | Numeric | 0.0 | 13.0 |
| Visibility | Numeric | 1.0 | 65.0 |
| Tdewpoint | Numeric | -3.5 | 15.5 |

1. **Data Cleaning**

*Questions:*

State if there is any cleaning needed?

How do you perform the cleaning?

How do you know if your data is clean?

Any reason for missing value?

* Combine the two data sets by using the Concatenate Node.
* Remove unwanted columns to keep data simple by using the Column Filter Node.
* Convert string data to integer data by using the String to Number Node.
* Look for missing values by using the Missing Value Column Filter Node.

1. Exploratory Data Analysis

Questions:

Perform statistical analysis of all the numerical variables.

Perform data transformation if needed.

Perform data mining to gain further insights. For example, plot a pie-chart or histogram to find out more.

Show and explain box plots (if any)

Show contingency table (if any)

Describe the percentage of probability.

Describe your findings.

Each sub-section can be analysis of each variable. You can choose to analyse all variables or pick a few important ones to discuss (due to time constraint)

1. Further Insights

1. Data Modeling
2. Conclusion
3. Reflection
4. References