Final Project Report

2023-12-09

Atharva Pore, Atharv Bagde, Jagadeesh Chitturi and Sarfaraz

Statement of goals

The primary goal of the dataset is to analyize the power consumption in Tetouan City. Also to find the relation between the temp and humidity factors on the consumption of power, and also to analyze the power consumption depending on the hour of the day, day, month and season.

Correlation analysis is proposed to quantify and visualize relationships between weather parameters and power consumption in each distribution zone. These insights aim to enhance understanding of how temperature, humidity, and wind speed impact energy usage.

To explore the inclusion of exogenous variable in time series analysis, also to develop a model using loess and GAM to predict power consumption of tetouan city based on climate or weather factors : temperature, humidity, wind speed. The analysis has valuable applications, particularly in areas such as predicting power consumption trends. Additionally, it has made a meaningful impact on global initiatives, exemplified by its role in supporting the UN Climate Conference in 2021, held in Paris.

Data Description

1. The dataset comprises 52,417 instances recorded at 10-minute intervals throughout the year 2017. Each instance includes six features: Date Time, Temperature, Humidity, Wind Speed, General Diffuse Flows, and Diffuse Flows.
2. The dataset is notable for its meticulous maintenance, guaranteeing data integrity and the absence of missing values. The features capture essential environmental parameters, such as temperature, humidity, and wind speed, while the diffuse flows components imply aspects related to the dispersion of substances.
3. With its detailed temporal resolution and comprehensive coverage, this dataset offers a robust foundation for analyzing and understanding the dynamic variations in environmental conditions over the specified timeframe.

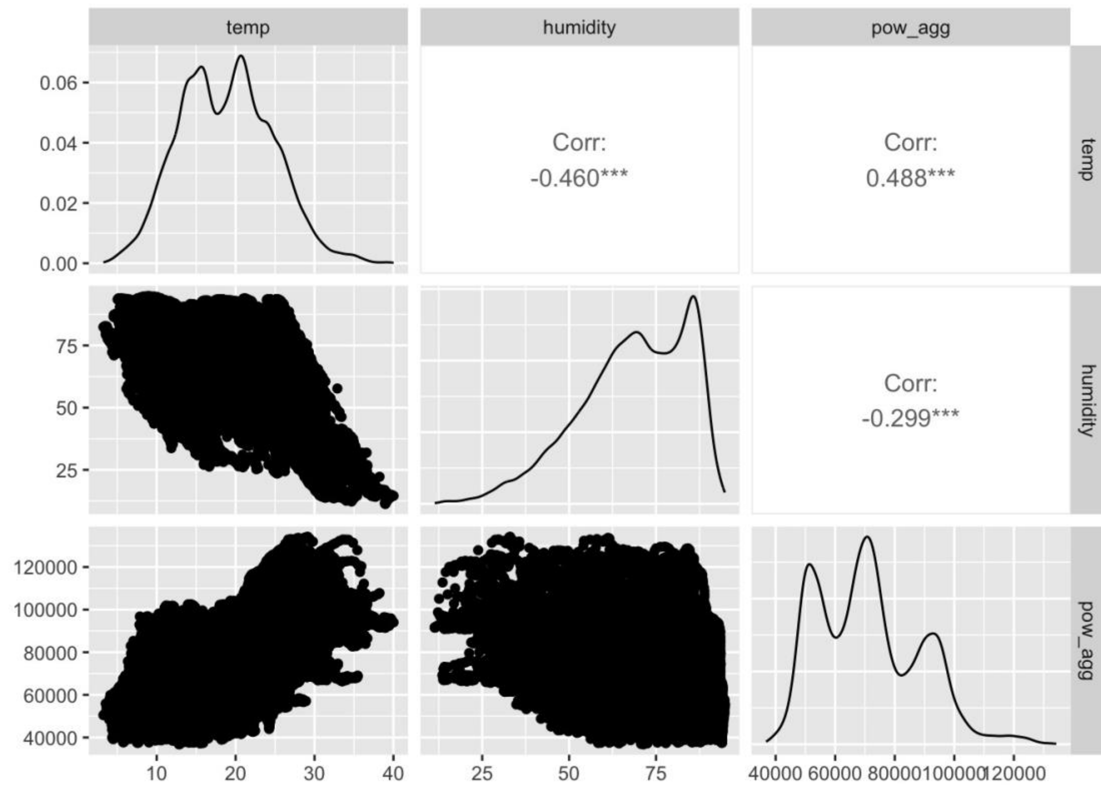
Insights

1. We found these insights based on an initial analysis of the data. The plots for those are given in the appendix.
2. The zone 1 has the highest power consumption followed by zone 2 and zone 3. We found some interesting trends for zone 3 power consumption in both monthly and hourly data, which suggests that zone 3 supplies power to mostly tourist

destinations of Tetuoan, which includes a large portion of the coastal area in Tetuoan.

3. During this initial analysis, we also found that the variables in this dataset are more or less dependant on each other and we will subsequently study their relationship below in a detailed manner

Plotting the Features (temperature and humidity) against Target (aggregated power) and also each other

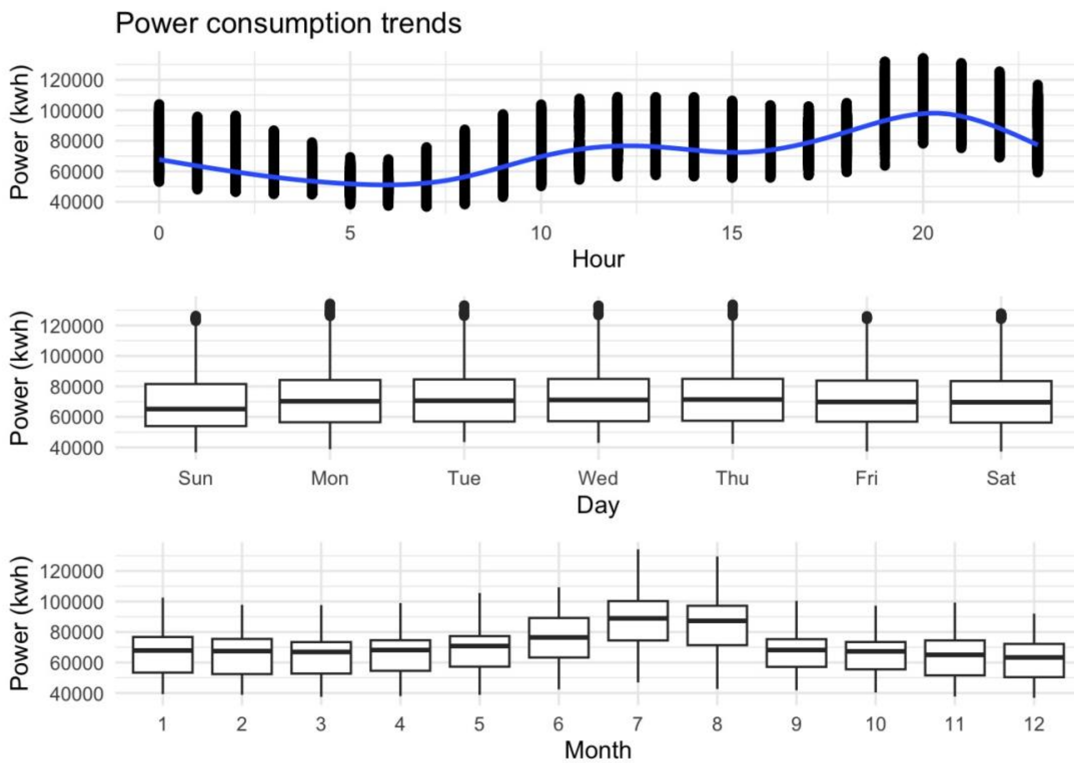


temp_pow_hum

1. The above plot shows that the power consumption shows a positive linear relationship with Temperature, correlation coefficient = +0.488, and a negative linear relationship with Humidity, correlation coefficient = -0.299.
2. Also, the distribution of the Power Consumption is trimodal, which is expected as it represents the power consumption aggregated across all 3 zones. 3. The temperature variable has a bimodal distribution where the 2 peaks occur in different seasons, the first one corresponding to Winter & Spring while the second to majorly Summer & Autumn with minor contributions from Spring.
3. The humidity variable also seems to follow a bimodal distribution where the first peak occurs during the afternoon & evening period and winter months while the second peak occurs during night & morning and especially summer months March & April.

Based on the above analysis, we concluded that for this data analysis, the Zones(1, 2 or 3) did not contribute much to the discussion and we decided to have their summation so that we could study the Power(total) against the other variables.

power consumption trends

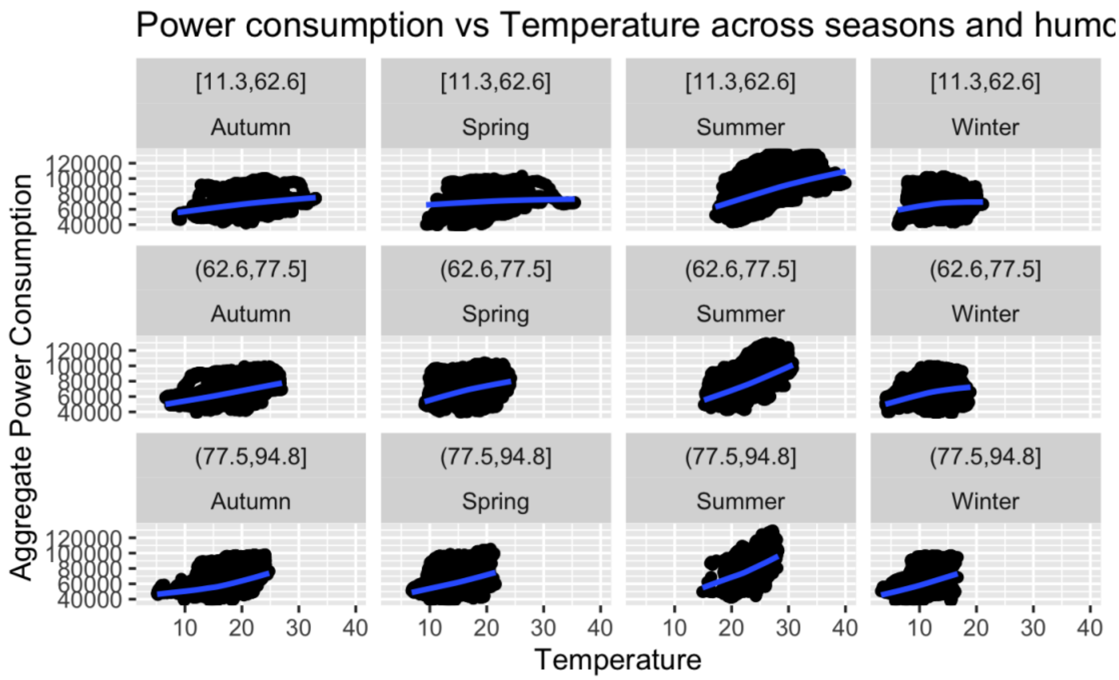


Power consumption trends

1. From the above plot, we can see that the power consumption exhibits variations with hour which shows daily trend in the power consumption while the monthly variations show annual trend.
2. The power consumed each day doesn't vary much with the day of the week which is evident from its boxplot above.

We will now understand the relationship between the temperature faceted across seasons in a year and their consequent effect on the power consumption.

Temp effect across months

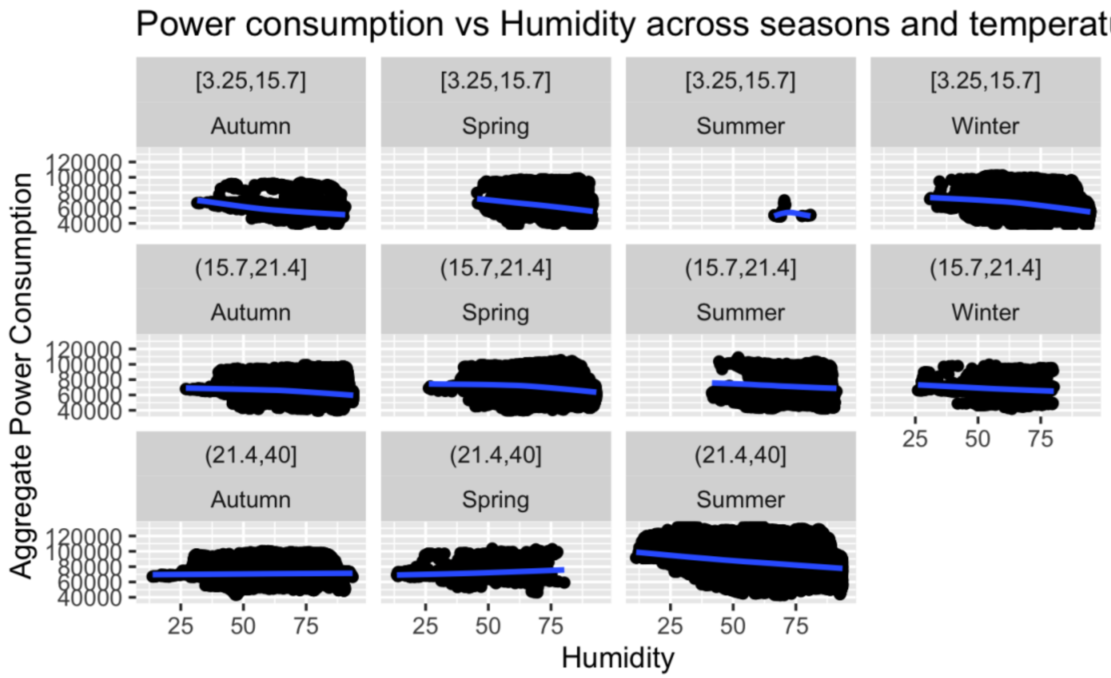


temp effect across months

Observations:

1. The above plot shows that the rate of power consumption varies with the temperature along with the season and humidity.
2. During high humidity days, the rate is higher than that during the low humidity days, where the power consumption shows little variation with temperature.
3. Moreover, the rate is significantly higher during summer days as compared to the other seasons.
4. This is obviously counter-intuitive because humidity is negatively correlated with power consumption so the temperature effect could be playing a significant role here.

Now, we will now understand the relationship between the temperature faceted across seasons in a year and their consequent effect on the power consumption.

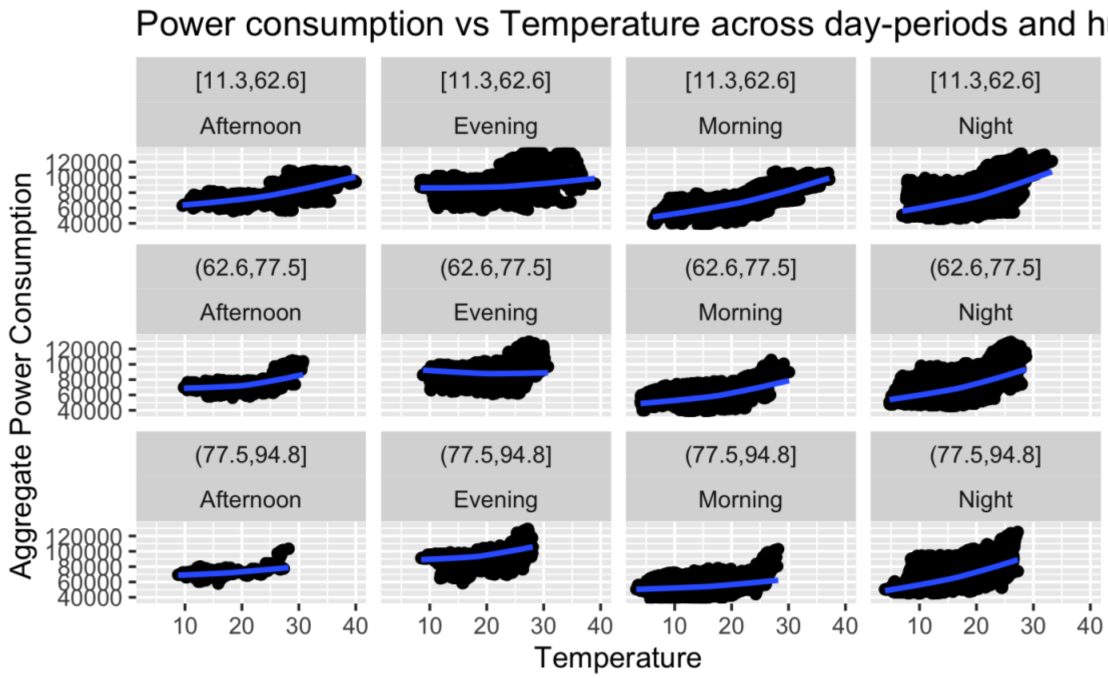


humidity effect across months

1. This plot shows that the power consumption decreases with the increase in humidity as expected from the negative coefficient of correlation between them. 2. However, the temperature plays an important role in changing the rate of change of power consumption w.r.t the humidity as the rate shrinks towards zero with increase in temperature during Autumn & Spring season while during the summer it shows decline in power consumption with increase in humidity which is normal.
2. An interesting observation is very little data for low temperature days in Summer season and no data for high temperature days in Winter season which shows that such days hardly occur during these seasons and so we should not make any predictions for data in this region, which we'll avoid in the modelling phase, as the available data is not enough to recognize the trend or pattern.
3. In the summer there is very low data for prediction

We concluded in our earlier analysis, that the daily aggregated data consistently remained stable, resulting in its exclusion from subsequent analysis due to its unchanging nature. Thus, we will directly be looking at the relationship between the temperature and the humidity variables across the hours in a day.

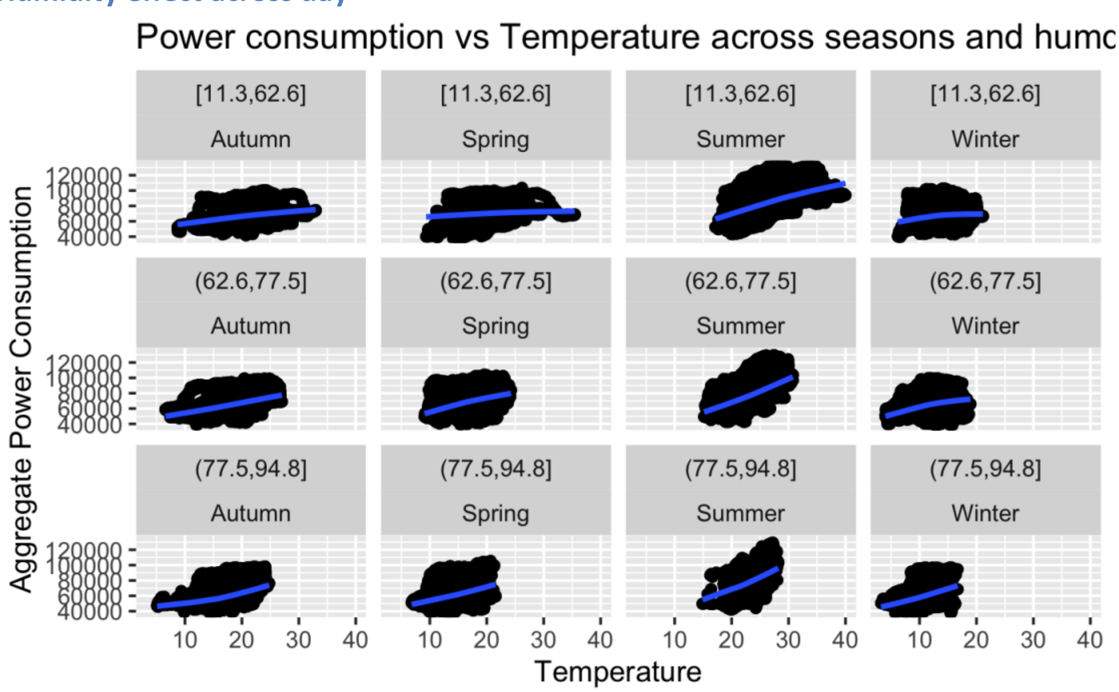
Temp effect across day



Temp effect across day

The plot shows that the power consumption generally increases with increase in temperature with high rate of increase for most periods of day and humidity excluding the afternoons and mornings with high humidity where the rate is very low.

Humidity effect across day



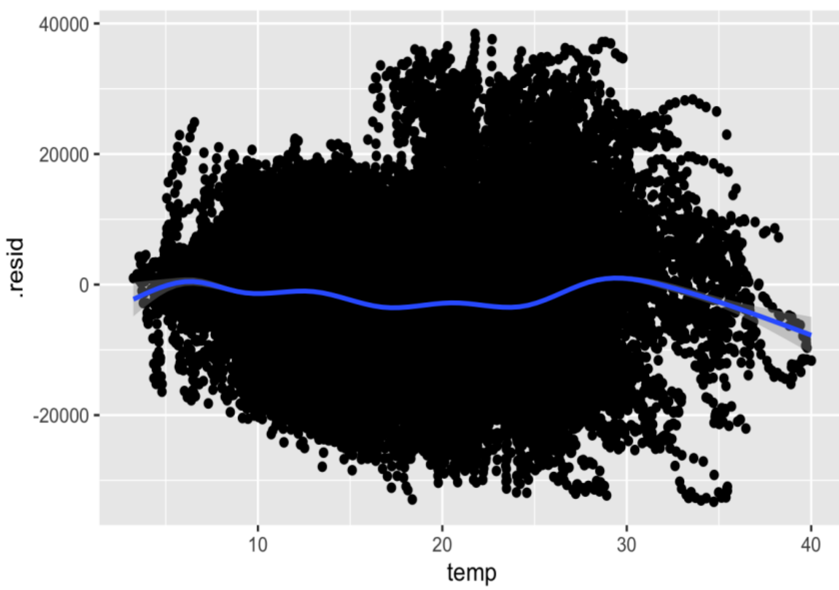
humidity_effect_across_the_day

1. From the above plot, we can see that the rate of change of power consumption with humidity is very low for most periods of day and most humidity range except for mornings and nights where the usual decline in power consumption with increase in humidity is seen.
2. Also, during the evenings with medium temperature the power consumption shows rise with increase in humidity, which can be further investigated.

Modelling:

While starting the modelling part, we initially thought about using a time series model, but we quickly realized that minute or hour time did not play as big of a part in this part as the other variables such as seasons, temperature and humidity did. For this reason, we restricted the modelling to loess, as the data was too big for a linear model and lesser complex than what is required for an additive model. We added the predictor variables of hour, month, temperature and humidity and kept the total power consumed variable as the response variable.

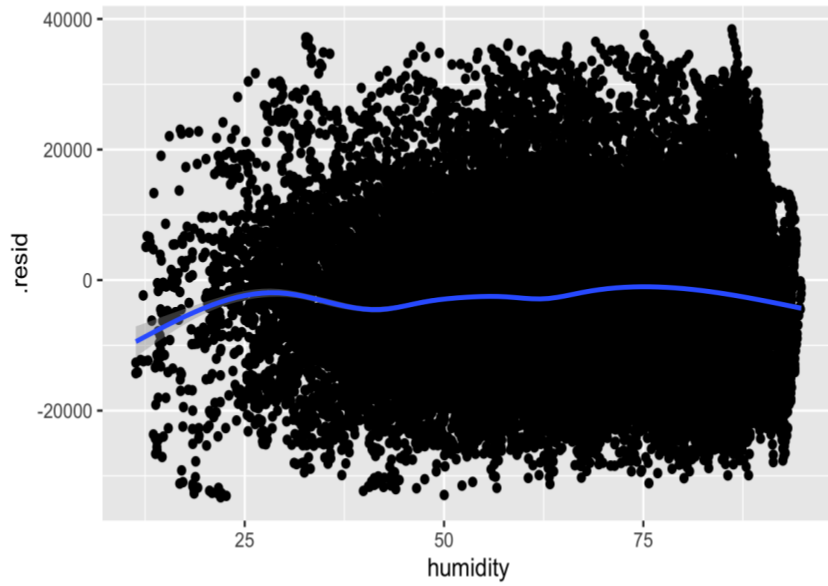
Resid vs Temp:



resid vs temp

1. As we can see from this plot, the residuals are pretty close to the 0 line, which suggests no sudden changes or abnormalities in the data.
2. We can also see that towards, the end the data gets a little uneven, which suggests that there might be some sudden changes in the data after it gets really warm.

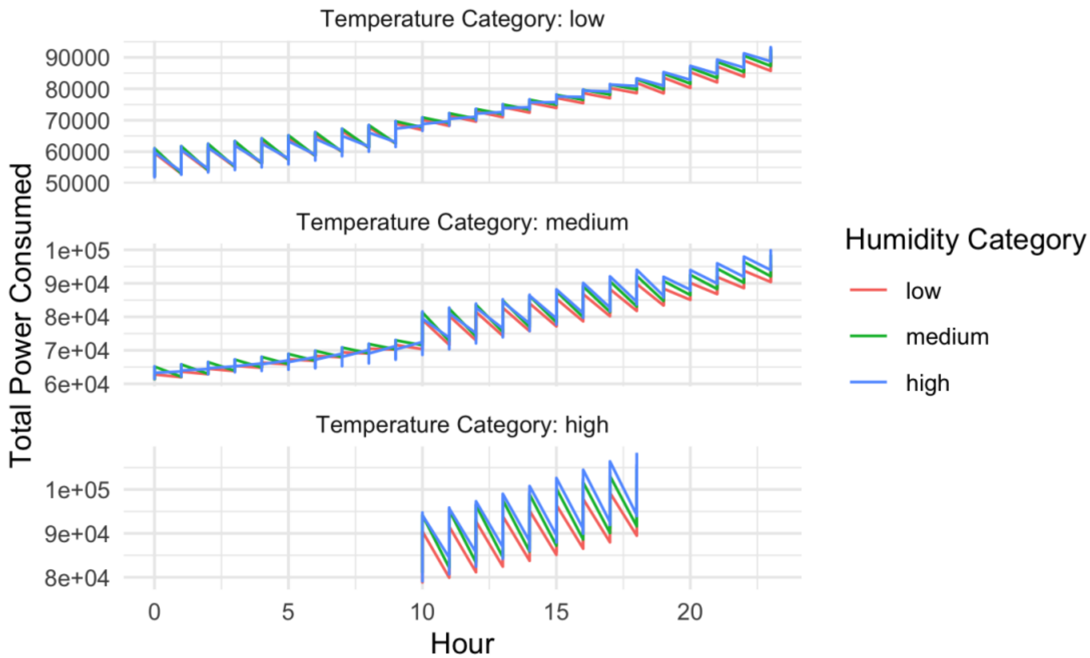
Resid vs Humid:



resid vs humid

1. Similar to the temperature, the residual line is pretty smooth, except for the start, where it increases rapidly.
2. However, it does become smooth and close to zero till the end, which suggests very less or no abnormalities in the humidity data.

Predicted Data trend:



Predicted data trend

1. The above plots are done using the predicted data. We used the loess model. We also needed to have the testing data in a particular way so had to pay special attention to the way we expanded the grid.
2. As we can see from the above plot, the trends of the humidity and the temperature are exactly opposite. When the temperature is high, the humidity is low and vice versa.
3. Moreover, when the power consumption increases, the humidity generally is low, which makes sense as the power consumption is generally high when the temperature is high.
4. Also, the trend of the power consumption is also the most when the temperature is at the maximum. The only thing which was counter intuitive in this case was that the power consumption did not get lesser at nights when in reality it does.

Conclusions:

After analyzing the data, we have found some interesting observations in the variation of power consumption with respect to the environment variables such as temperature and humidity. We can conclude that the power consumption, irrespective of the zones, is the maximum in the summer months and lesser in the winter months. The power consumption peaks in August, which makes sense as August in Tetouan is the hottest month of the year which leads to usage of appliances such as air conditioners and fans in residential areas. While, in the industrial areas, the surrounding temperatures affect the machines and the cooling down consumes a lot of power in summer. Whereas, in the winters, the average temperature is around 10, which does not require any home appliances as well as any cooldown machinery. As we have studies from the above plots, the trends of humidity and temperature are inversely proportional to each other, meaning that when the temperature is high during the days, the humidity is low. The wind speed does not play a big role in this context as it usually remains constant throughout. Our model understands these relations between the environmental factors and the power consumption, but is unable to grasp the intracies of the trends between power used and the time of the day, which leads us to a future scope in time series analysis. The limitation of the dataset resides in the fact that no geographical or any physical distinction of the zones is given as a metadata, which becomes an obstacle in explaining the process of power consumption variations in all three different zones.

Future Scope

1. We did not find any relevant information about the zones and what areas they comprise of exactly in our research. Thus one of our future tasks would be to have an even in depth research to find more about the zones so as to have better models and conclusions.
2. With intriguing time-based data in our dataset, our future focus is on time series analysis. By delving into temporal patterns, we aim to improve the depth of our insights and refine our modeling approaches.