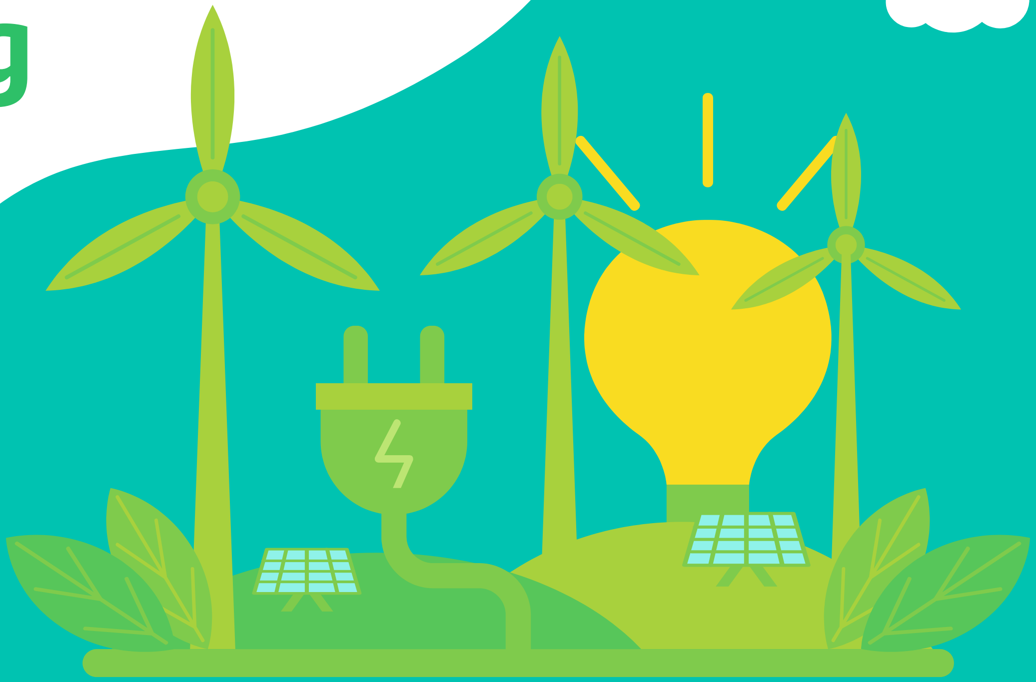
Power Consumption Forecasting

Eva Burns
05/25/2023



Agenda

01 | Problem statement

02 | Assumptions & Hypotheses

03 | Data properties & EDA

04 | Data Processing & Transformations

05 | Feature Engineering

06 | Proposed Approaches

07 | Results & Learnings

08 | Future work

Problem Statement

- A **power consumption** study in 2017 was done at Tétouan, Morocco by the Supervisory Control and Data Acquisition System (SCADA) of Amendis
- The energy distribution network of Tétouan is powered by 3 Zone stations: Quads, Smir and Boussafou
- The purpose of this project is to study and **forecast** power consumption in Tétouan
    - I will only focus on forecasting Zone 1's (Quads) power consumption
    - I will be forecasting 12/11/2017 – 12/31/2017

Data Description

The data was found on Kaggle and was collected every ten minutes for exactly one year (1/1/2017 0:00 - 12/30/2017 23:50) with nine columns:

- **Datetime**: Time window of ten minutes
- **Temperature**: Weather Temperature
- **Humidity**: Weather Humidity
- **WindSpeed**: Wind Speed
- **GeneralDiffuseFlows**: “Diffuse flow” is a catchall term to describe low-temperature (< 0.2° to ~ 100°C) fluids that slowly discharge through sulfide mounds, fractured lava flows, and assemblages of bacterial mats and macrofauna
- **Diffuse Flows**
- **PowerConsumption_Zone1**: This is what will be forecasted
- **PowerConsumption_Zone2**
- **PowerConsumption_Zone3**

Assumptions & Hypotheses

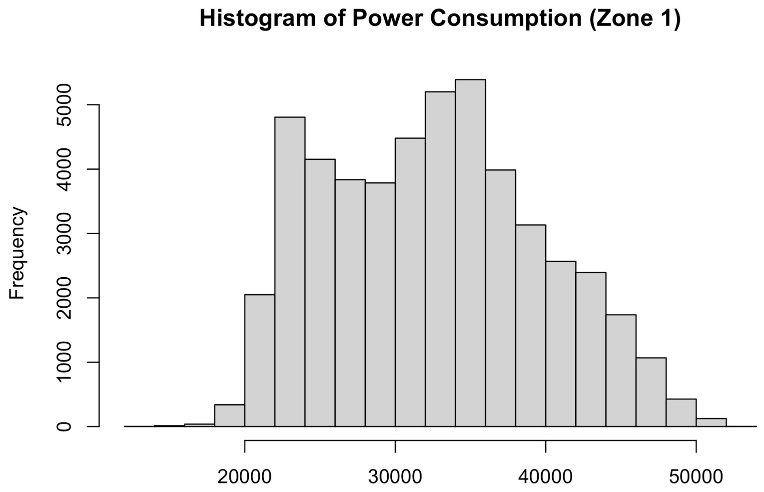
Assumption: The data was actually collected every ten minutes. Also, that the sensors were calibrated correctly throughout the year.

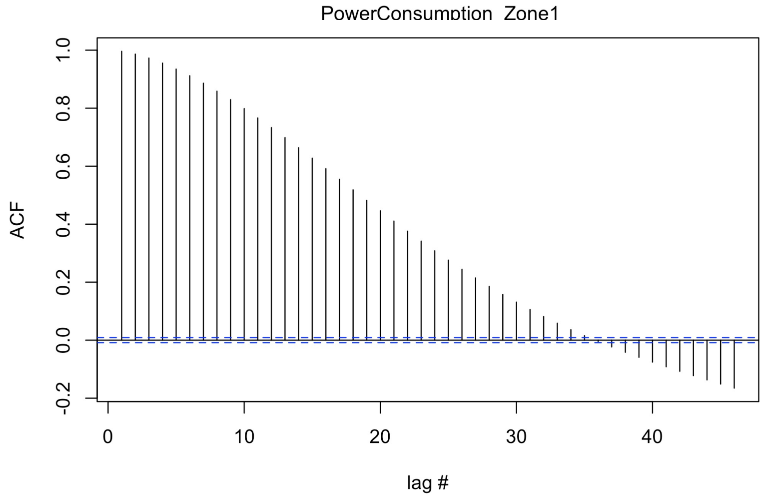
Hypothesis: a model with a **seasonal** component will be most appropriate.

- Possibly be two seasonalities: **daily** and **yearly**.

    - Throughout the day, power consumption changes. For example, when you are sleeping, power consumption is lower because you are not using appliances or lights around your house.

    - I also think power consumption changes yearly as weather changes during the seasons, affecting AC/heating consumption. However, since this data only covers 2017, yearly seasonality will not be able to be used effectively





Data Properties

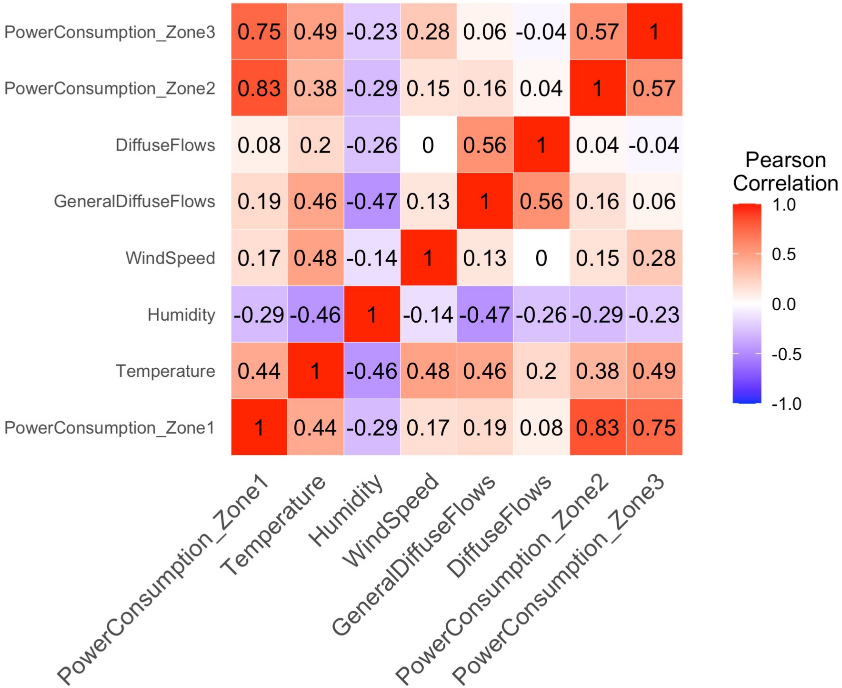
ADF test p-value: 0.01

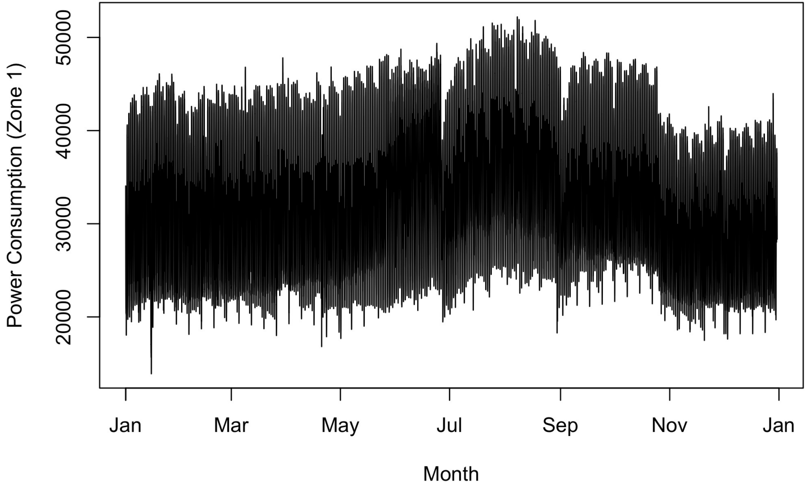
KPSS test p-value: 0.01

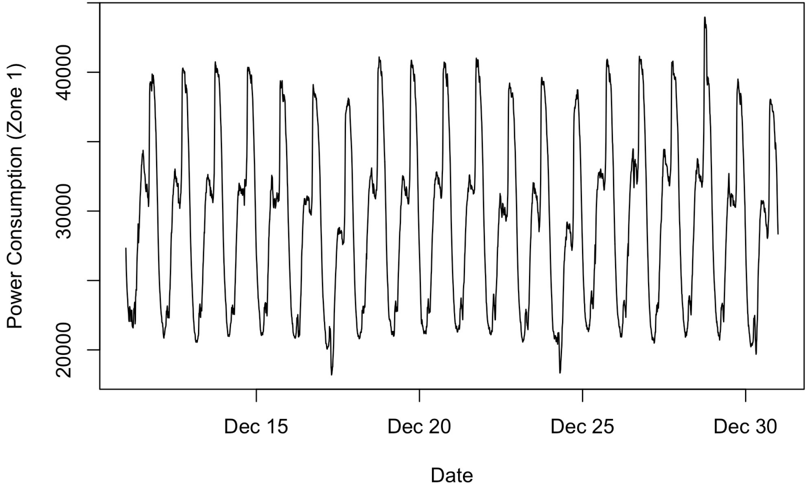
While the ADF test gives a low p-value concluding that the data is stationary, the ACF plot and KPSS indicate the data is **non-stationary** since the KPSS p-value is low and the ACF plot does not die down quickly

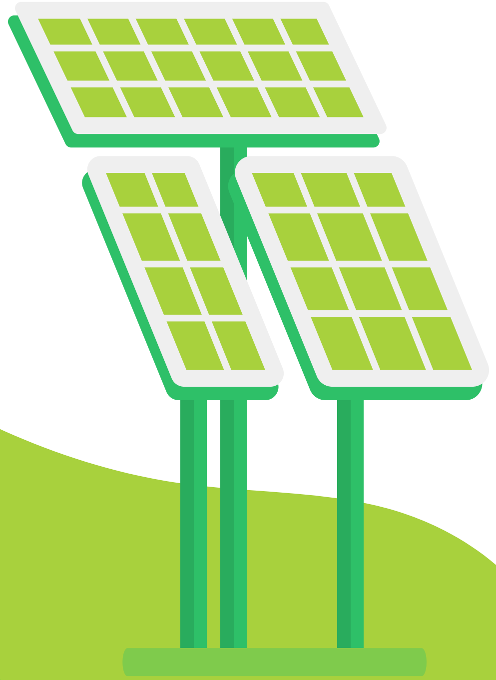
Exploratory Data Analysis

The power consumption in Zones 2 and 3 are the most correlated with power consumption in Zones 1









Data Processing

Outliers/Anomalies

Checked using `tsoutliers`

No outliers found, so no data needs to be replaced

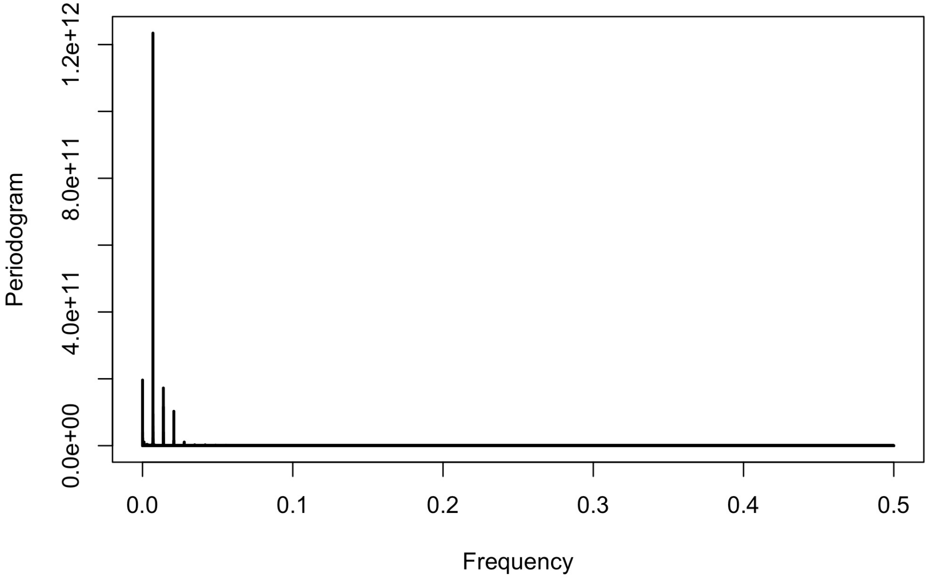
Missing Values

The data is full and collected at regular intervals with **no null** values

No cleaning is required

Feature Engineering

- A frequency **periodogram** was applied to the power consumption (Zone 1) to determine if/what the seasonal period is
- Seasonality with the highest frequency is 144.092
- Indicates there is **daily seasonality** since the data is collected in 10-minute increments
    - 1440 minutes in a day, which divided by 10 minutes is 144

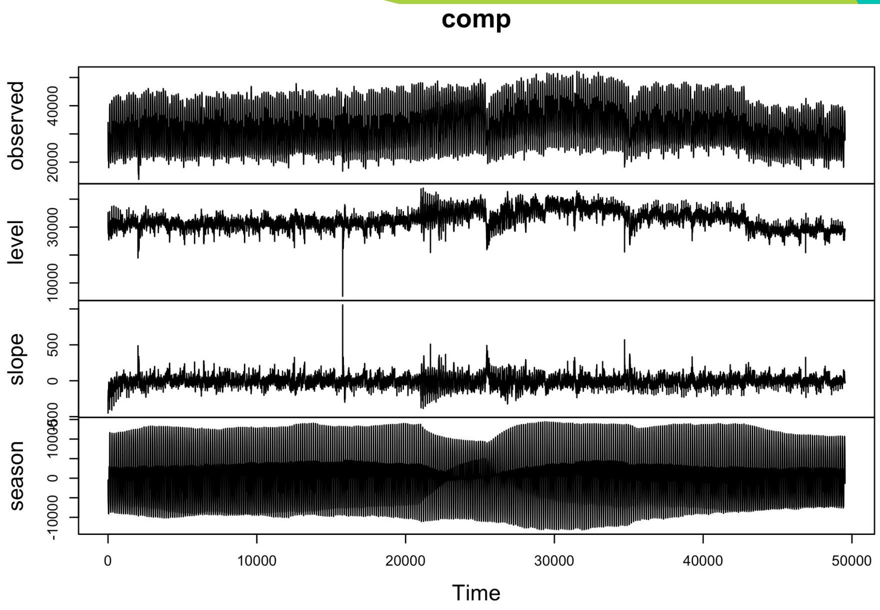


Proposed Approach - TBATS

Trigonometric regressors to model multiple-seasonalities
Box-Cox transformations
ARMA errors
Trend
Seasonality

However, it is a complex model, so there is risk of overfitting the train data, as well as the model being computationally expensive

I built the TBATS model using the train data (1/1/2017 00:00 - 12/10/2017 23:50) with a seasonal period of 144.



Proposed Approach - SARIMA

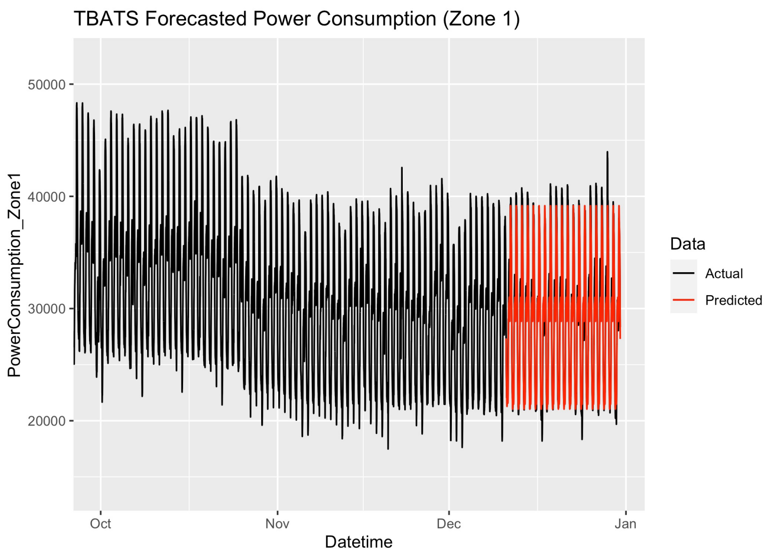
SARIMA also handles seasonality but is a simpler model. This will make the time to create the model faster, and may help with overfitting

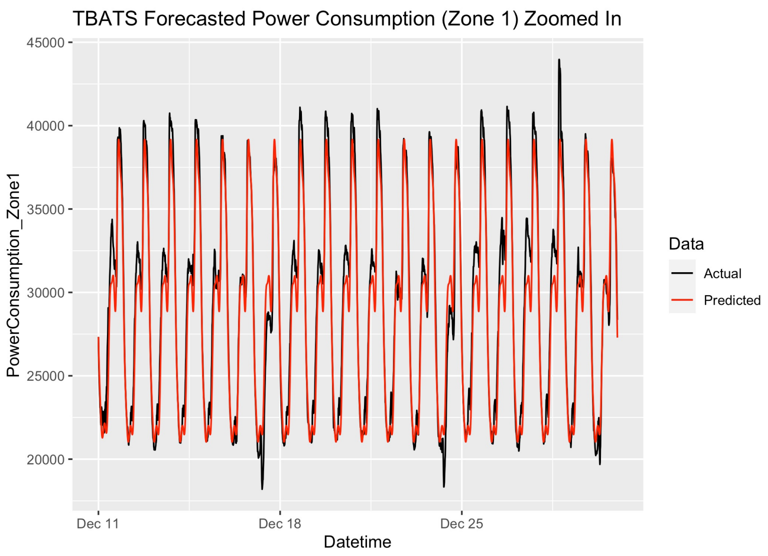
I built the SARIMA model using the train data (1/1/2017 00:00 - 12/10/2017 23:50) using `auto.arima` with a seasonal frequency of 144.

Model: **ARIMA(3,0,0)(0,1,0)[144]**

Results - TBATS

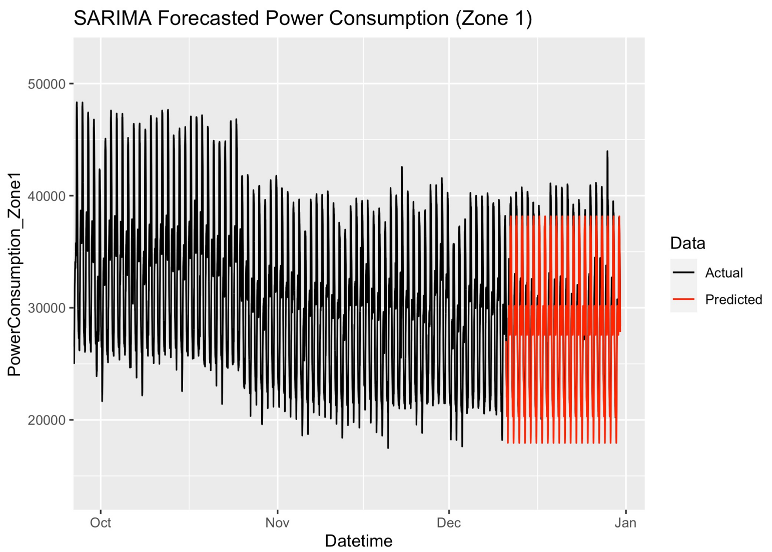
	ME	RMSE	MAE	MPE	MAPE	MASE
Train	0.474	410.923	260.04	-0.010	0.820	0.600
Test	674.98	1491.04	1156.175	1.957	3.909	2.666

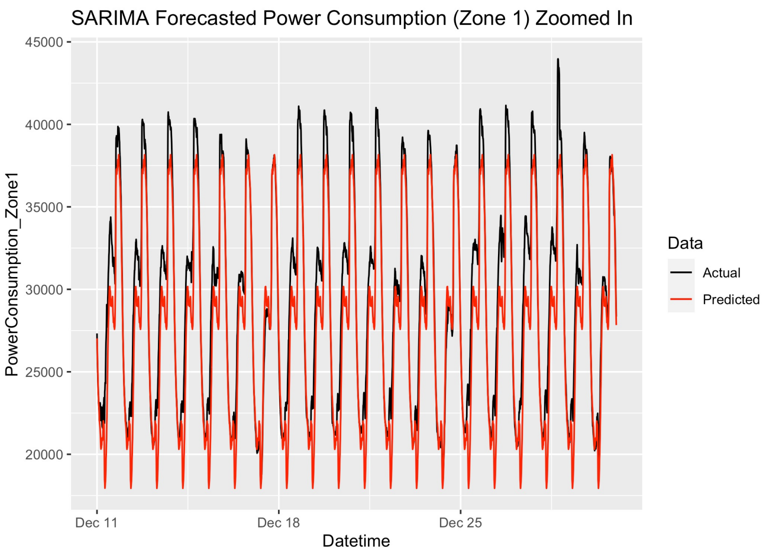




Results - SARIMA

	ME	RMSE	MAE	MPE	MAPE	MASE
Train	0.142	438.935	266.596	-0.011	0.852	0.615
Test	1638.84	2377.233	1802.43	5.495	6.144	4.156





Comparison of Models

TBATS performed better that SARIMA on all accuracy metrics for the test data

It should be noted that the accuracy of the training data on both models were about equal. This tells us that the SARIMA model overfit the training data.

Future Work

- The current dataset ignores one of the main capabilities of the TBATS model: multiple seasonality. I believe there is **yearly** seasonality to this data because of how weather changes throughout the year
- However, this data is only for 2017, so yearly seasonality cannot be used on this project. If more data from multiple years were used, the TBATS model could be even more accurate because it can handle both daily and yearly seasonality

In other words, in the future I believe that collecting more data for **multiple years** will yield an even more accurate **TBATS** model because it can be built on multiple seasonalities: daily and yearly.

Thank you!