Summarize Feedback and next steps

Feedback

- Normalize, select and squeeze features out of existing datasets (Features which make sense)
- Too many time features (with moving averages) can cause overfitting
- Describe why features + hyperparameters have been chosen
- "Formular" for energy prices (Influence e.g. weather, CO2 tax etc.)
- Explanation of data integration for the features was missing (Integration was done by datetime)
- More Statistical understanding is needed (which distribution)
- (Holistic view, features are the same)
- Logical background for story telling
    - Where did we start
      → Blank model with no features, we chose the features etc.
    - How did we get to where we are
    - Explain why the features have been chosen
    - Why did we choose 2 different models
        - XGBoost: Blank model as a baseline
        - Prophet: Time based series model for possible better results

Summarize of our storyline

- "Formular" for energy prices (Influences: Gas + Oil prices, Energy mix, Weather, CO2 tax)
- Scraping Data (Energy price, Weather, Energy mix data [daily] → later hourly for a better resolution)
- Baseline: XGboost → we wanted to improve up on this benchmark / baseline
- Using Prophet (Time based series model)
- Both models were blank at the beginning (no features, no hyperparameter)
  → Again Benchmark / baseline to improve from this
- Finding lower and upper bound for Gaussians distributions – (2 or 3 σ from μ)
- Use RMSE for validation
- Adding Hyperparameter
- Hyperparameter grid search
- Adding features
- Moving averages

Next steps in general:

- Research for new metrics (Move not just averages – Min and Max)
- Make some hypotheses and try to contradict them
- Consider how to combine models

Next steps till next Thursday:

- Validation with more errors functions
- Extract more features out of datasets
- Find more validate features
- Prophet: Optimize model and test with different error functions
- XGboost:
    - Normalize and select features
    - Optimize model and test with different error functions