Pytown Capstone Energy Use Case

Predicting energy generation

PyLadies Amsterdam Bringing ML Models into Production Bootcamp
July 7th, 2021

The team

Mentor: Matthijs

Hassan Salam

Data Scientist

Chemnitz, DE

linkedin.com/in/hassansalamb/

- Expertise in predictive data modelling and deploying data-driven services
- Delivering insights

Dânia Meira

Data scientist

Berlin, DE

linkedin.com/in/meiradania

- Worked mostly on ML models, deployment only on-premise

Approach - End-to-end Batch Naive Forecast

Data preprocessing: aggregate daily: energy load -> consumption ; wind and solar predictions -> generation

Model training: train naive forecast - average consumption of the same day of the week, considering N weeks before

Model evaluation: backtest with a sliding window approach, optimize for MAPE

Model deployment: deploy the best performing model as an Azure Machine Learning Batch pipeline Model

Model post-processing: compare energy consumption forecast with energy generation from wind and solar predictions, classify the energy consumption per day (normal, middle, low charge), persist results to Azure Blob Storage (to be used by Power BI).

Model monitoring: design a basic technical and model specific monitoring

Challenges - Batch pipeline on the cloud

Challenge 1

Data Registering

- Uploaded but not registered

Challenge 2

Run pipeline on the cloud

Understand the interface

- Input
  - Tabular data
  - Index of DF
- Output
  - List or DF
  - Length

Challenge 3

Connect data to Power BI

From cloud storage to local environment

Avg consumption on the same day of the week

	dayofweek	load_pred_mw
data_index_		
2020-01-01 00:00:00+00:00	2	NaN
2020-01-02 00:00:00+00:00	3	NaN
2020-01-03 00:00:00+00:00	4	NaN
2020-01-04 00:00:00+00:00	5	NaN
2020-01-05 00:00:00+00:00	6	NaN
2020-01-06 00:00:00+00:00	0	NaN
2020-01-07 00:00:00+00:00	1	NaN
2020-01-08 00:00:00+00:00	2	NaN
2020-01-09 00:00:00+00:00	3	NaN
2020-01-10 00:00:00+00:00	4	NaN
2020-01-11 00:00:00+00:00	5	NaN
2020-01-12 00:00:00+00:00	6	NaN
2020-01-13 00:00:00+00:00	0	NaN
2020-01-14 00:00:00+00:00	1	NaN
2020-01-15 00:00:00+00:00	2	10178.149906
2020-01-16 00:00:00+00:00	3	10708.442896
2020-01-17 00:00:00+00:00	4	10806.161266
2020-01-18 00:00:00+00:00	5	10006.685330
2020-01-19 00:00:00+00:00	6	9754.771028
2020-01-20 00:00:00+00:00	0	11005.889968

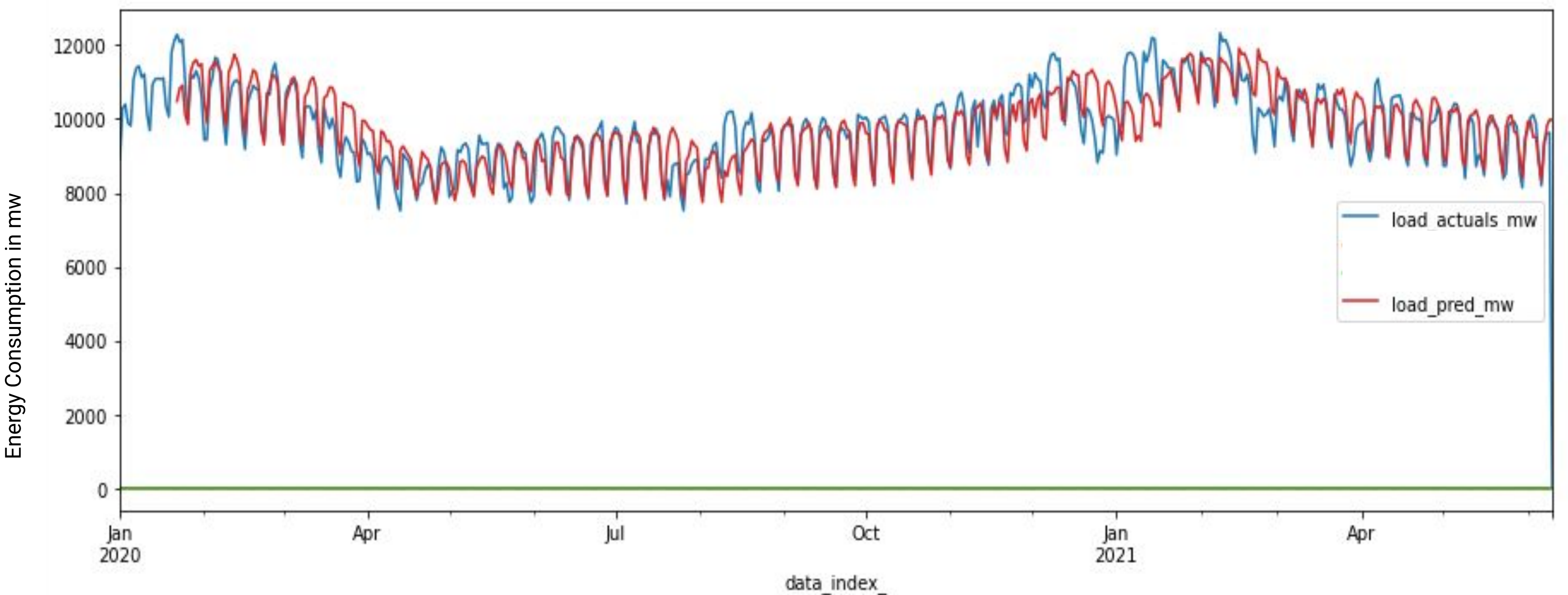
Example looking at previous 2 weeks to calculate average

Little difference in MAPE, not worth missing a week's prediction

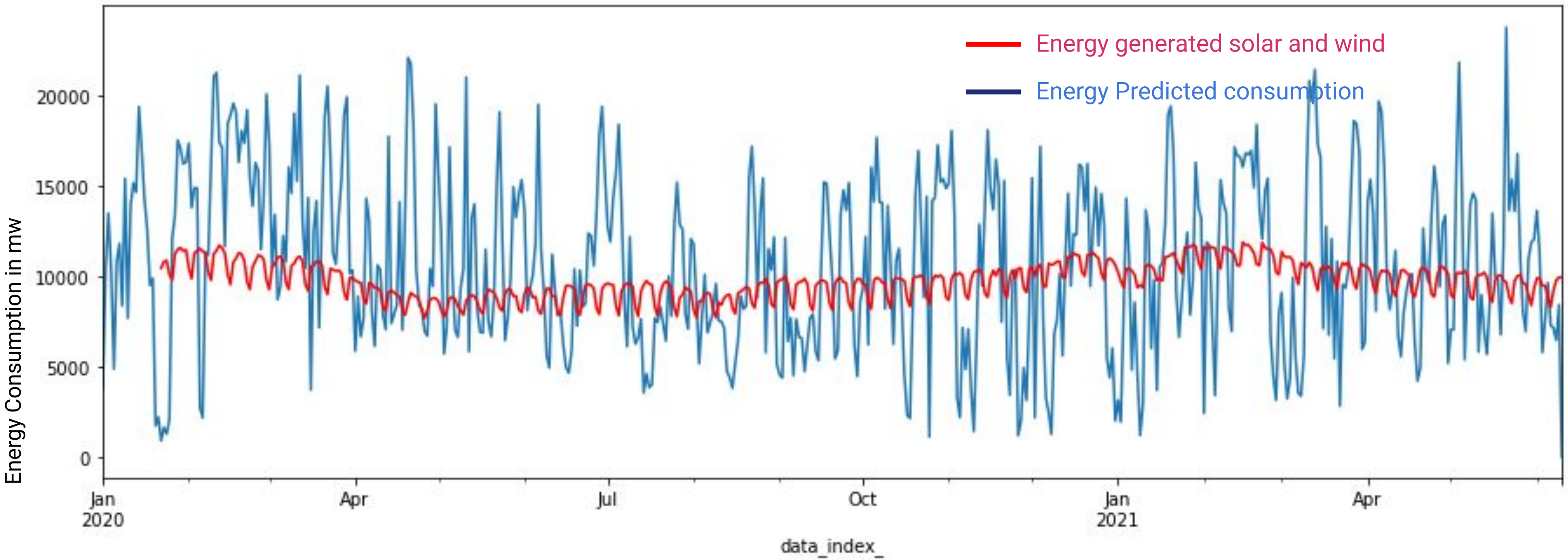
```
mape(df_load_daily, df_load_daily.load_actuals_mw)
✓ 0.6s
[(0, 0.0),
 (1, 99.96141440155532),
 (2, 26.580908781460245),
 (3, 26.520473692265178),
 (4, 26.52331795756313),
 (5, 25.881142816848723),
 (6, 26.413458601280254),
 (7, 26.851940956691394),
 (8, 27.059453655979333),
 (9, 27.433265149640484),
 (10, 27.725994962708295),
 (11, 27.85679854083038),
 (12, 28.04545630557903),
 (13, 28.29138621914685),
 (14, 28.347424431763706),
 (15, 28.402398387159145),
 (16, 28.2670147641685)]
```

Best Model - MAPE 25.5

Naive Forecast showing predicted energy consumption vs actual energy consumed

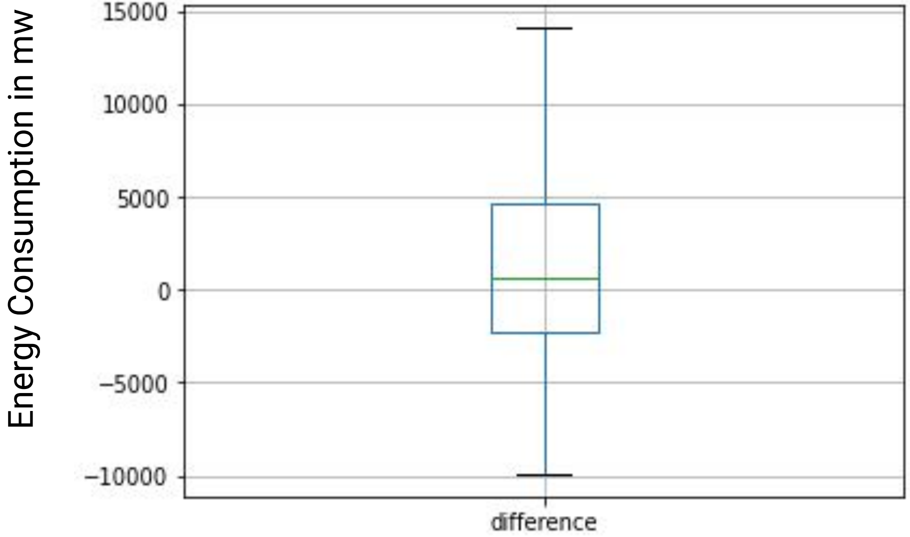


Difference in Predicted Energy Consumption and Total Energy Generated

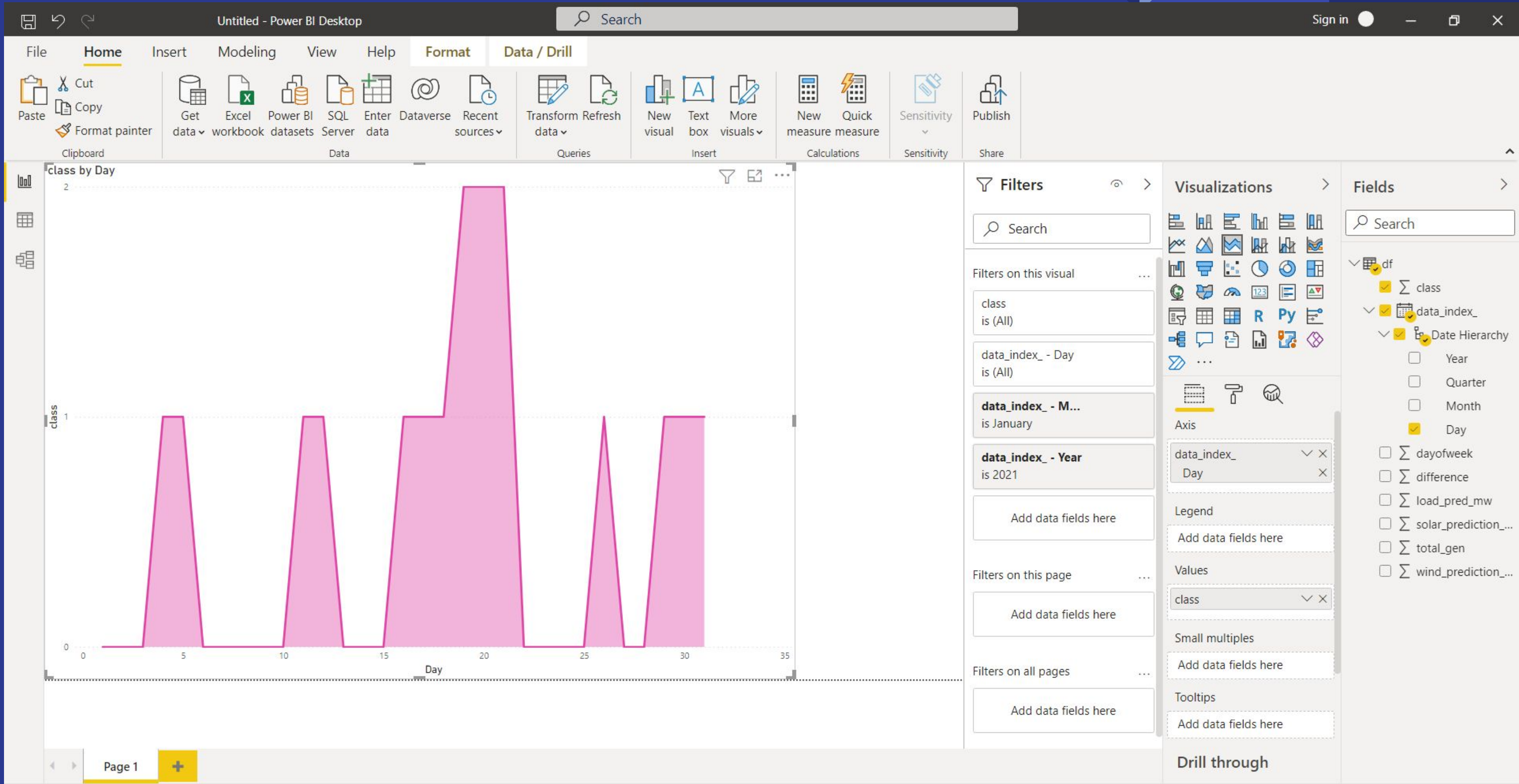


Classification

Difference <0 : Low charge
0 < Difference < 5000: Medium charge
Difference > 5000: Normal charge



Dashboard - Power BI



Open tasks

- Saving output in the desired blob storage
- Write tests and documentation
- Schedule the pipeline
- Implement near real-time inference