ENERGY DATA SCIENCE

Data Science Lifecycle

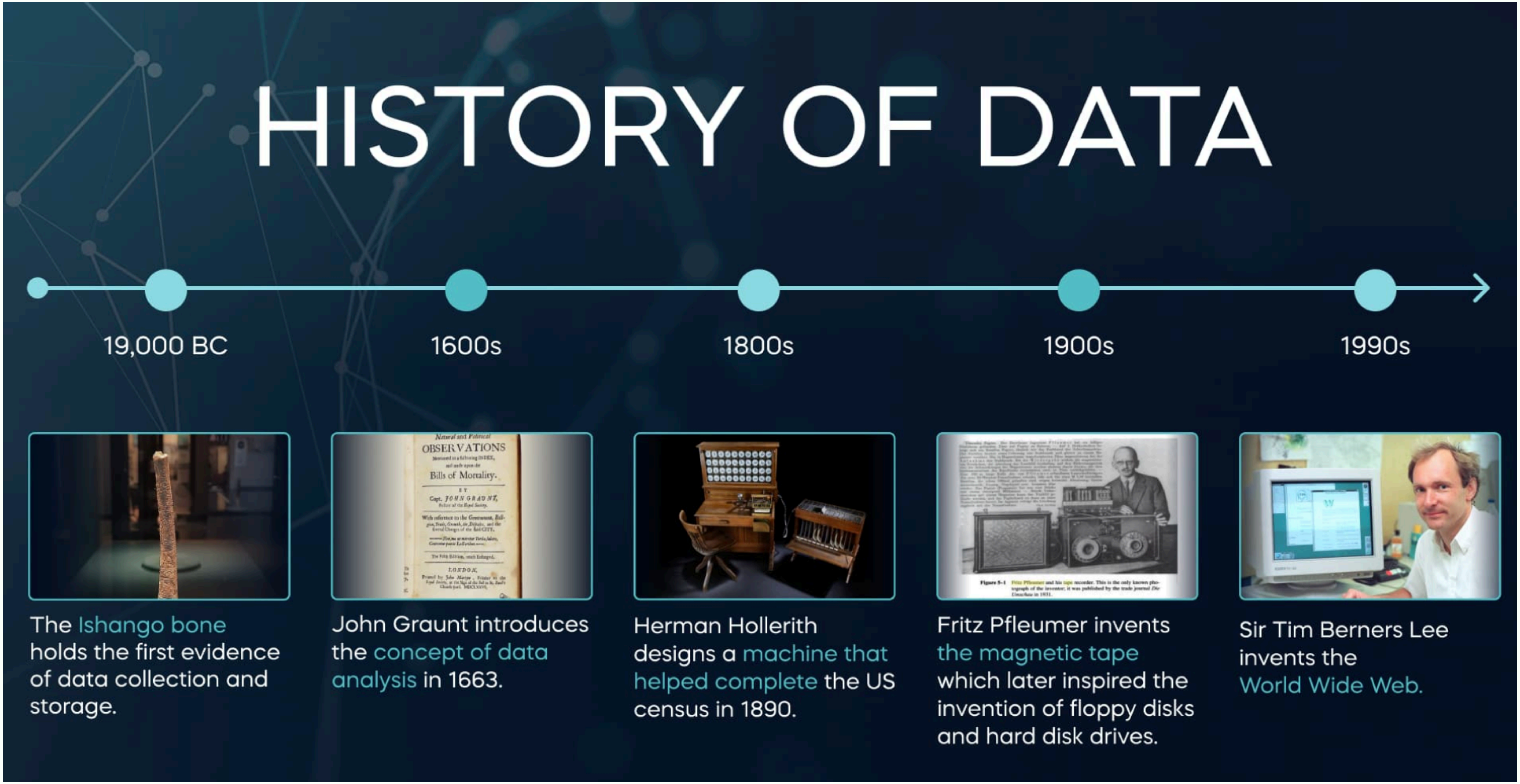
Prof. Juri Belikov

Department of Software Science
Tallinn University of Technology
juri.belikov@taltech.ee

PREVIOUSLY IN COURSE …

Key takeaways:

- Energy and information
- Digital transformation
- Smart grids
- Use cases

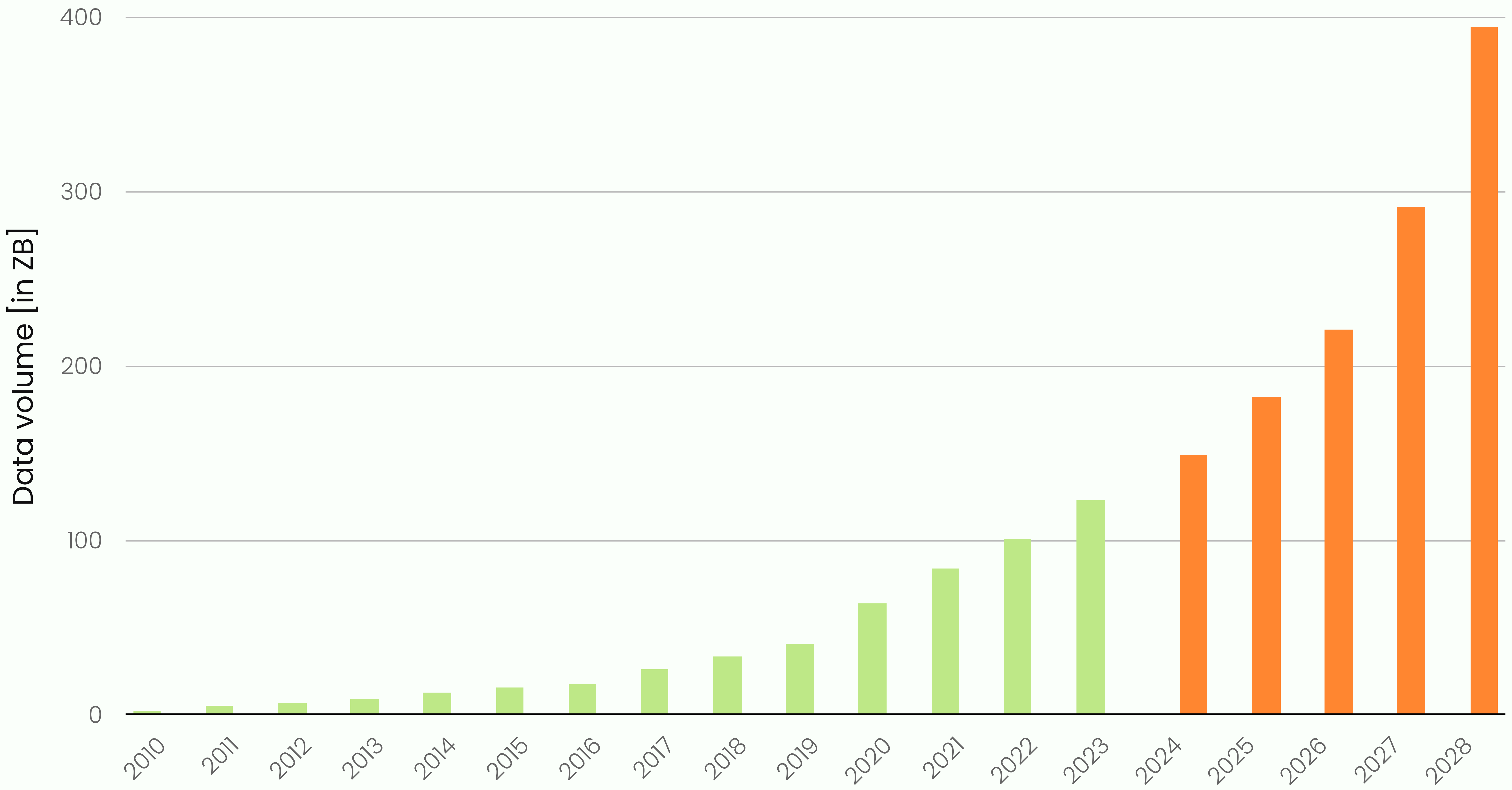

HISTORY OF DATA
19,000 BC
1600s
1800s
1900s
1990s
The Ishango bone holds the first evidence of data collection and storage.
John Graunt introduces the concept of data analysis in 1663.
Herman Hollerith designs a machine that helped complete the US census in 1890.
Fritz Pfleumer invents the magnetic tape which later inspired the invention of floppy disks and hard disk drives.
Sir Tim Berners Lee invents the World Wide Web.

More than 50 years ago, John Tukey called for a reformation of academic statistics. In "The Future of Data Analysis," he pointed to the existence of an as-yet unrecognized *science*, whose subject of interest was learning from data, or "data analysis." Ten to 20 years ago, John Chambers, Jeff Wu, Bill Cleveland, and Leo

John W. Tukey (1962) The Future of Data Analysis. Ann. Math. Statist. 33(1): 1-67.

D. Donoho. 50 Years of Data Science. *Journal of Computational and Graphical Statistics*, *26*(4), 745–766, 2017.

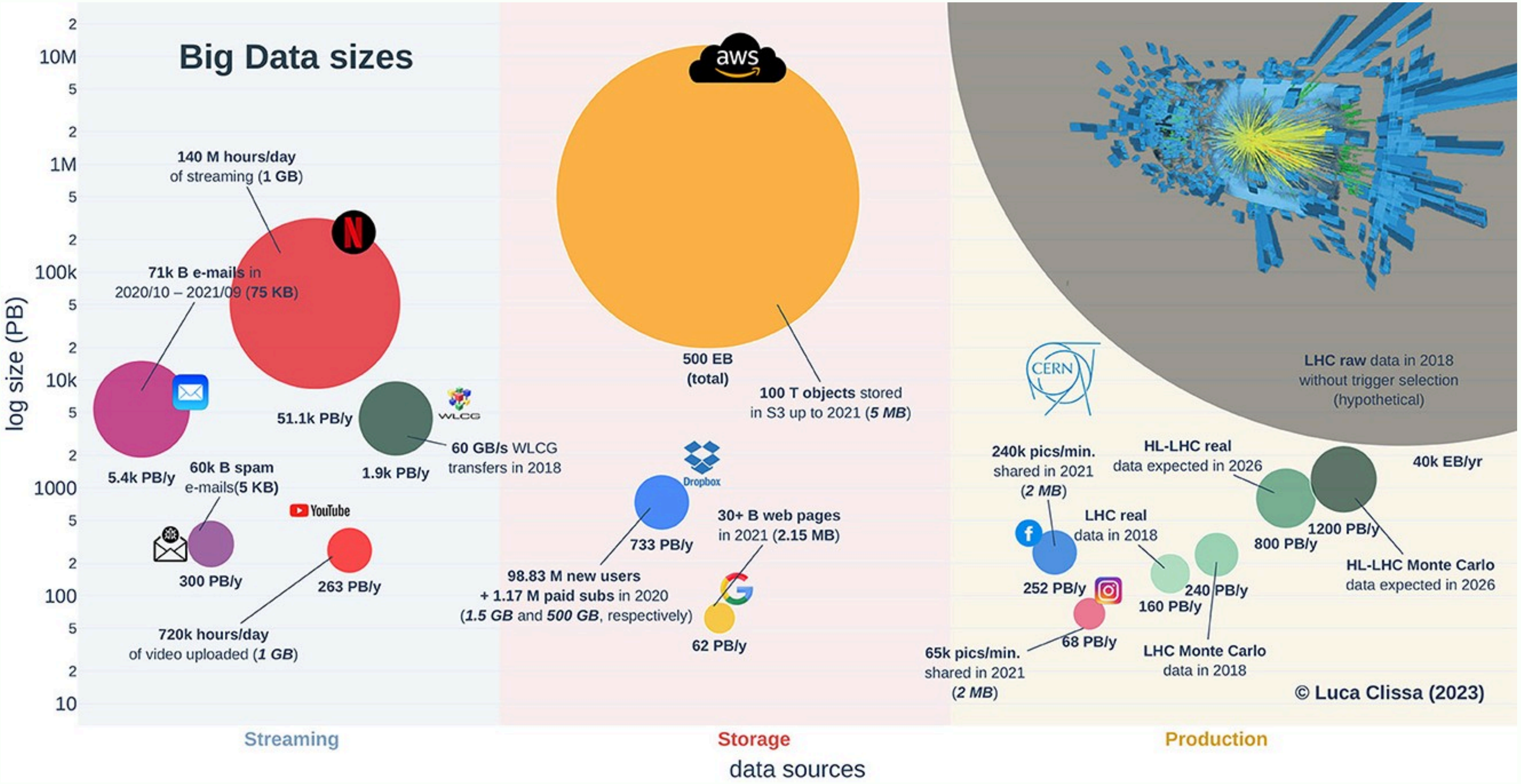
WORLD DATA IN NUMBERS



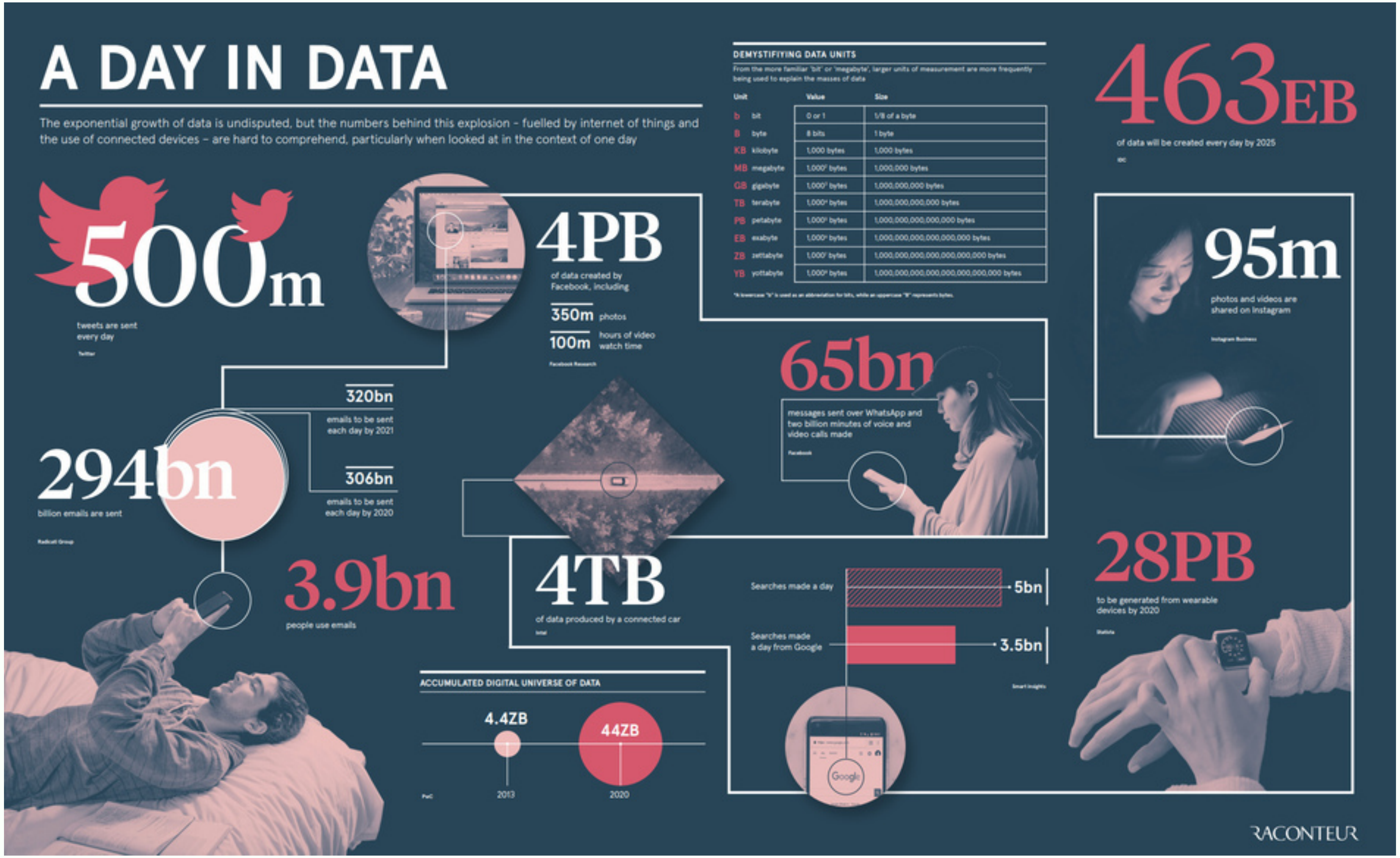
Unit	Value
bit	1
byte	1 bit
kilobyte (kB)	10^3 bytes
megabyte (MB)	10^6 bytes
gigabyte (GB)	10^9 bytes
terabyte (TB)	10^{12} bytes
petabyte (PB)	10^{15} bytes
exabyte (EB)	10^{18} bytes
zetabyte (ZB)	10^{21} bytes
yottabyte (YB)	10^{24} bytes

[Accessed June 9, 2025], https://www.statista.com/statistics/871513/worldwide-data-created/

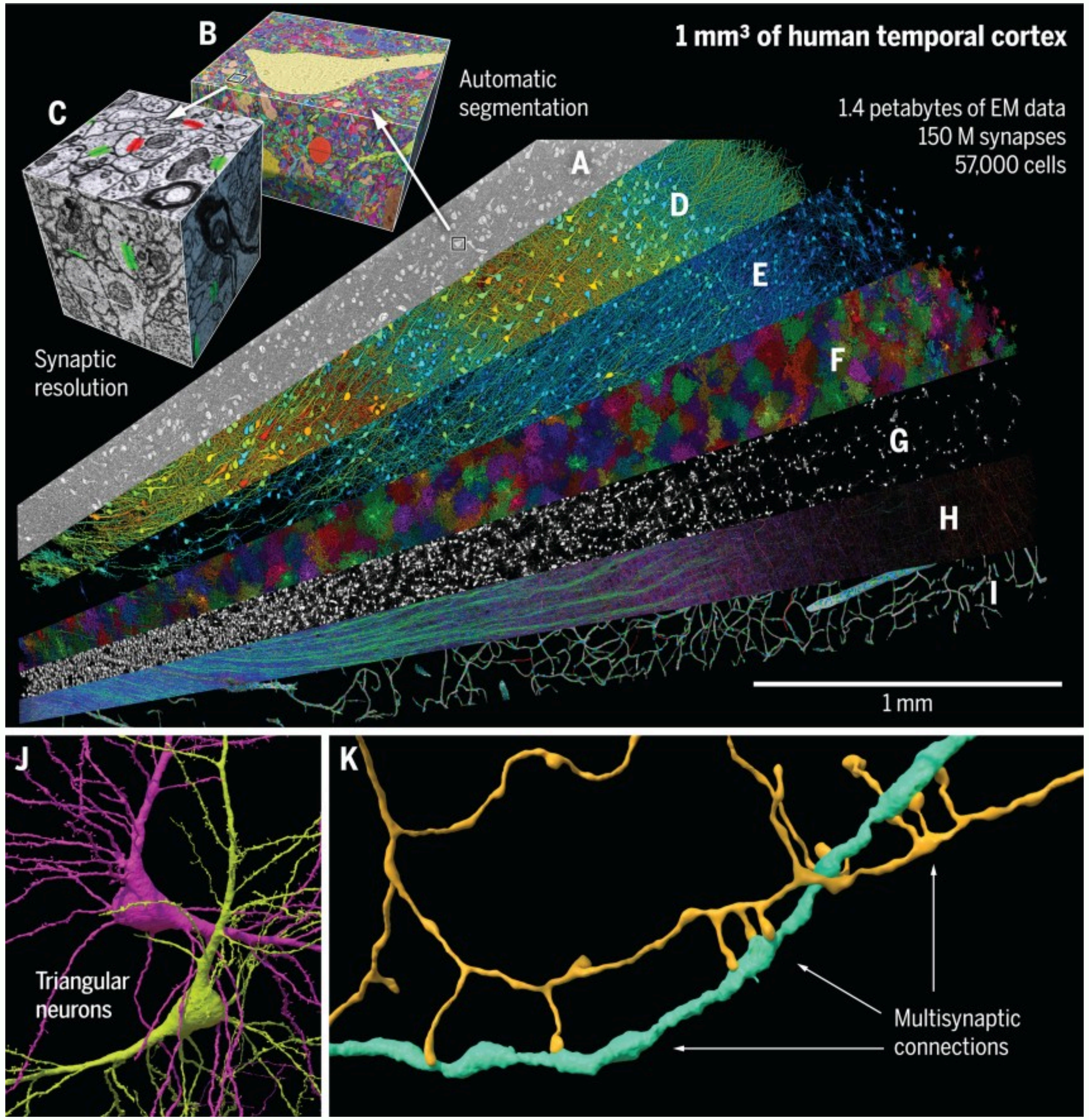
GETTING PERSPECTIVE



GETTING PERSPECTIVE (2)



3D HUMAN BRAIN MAP

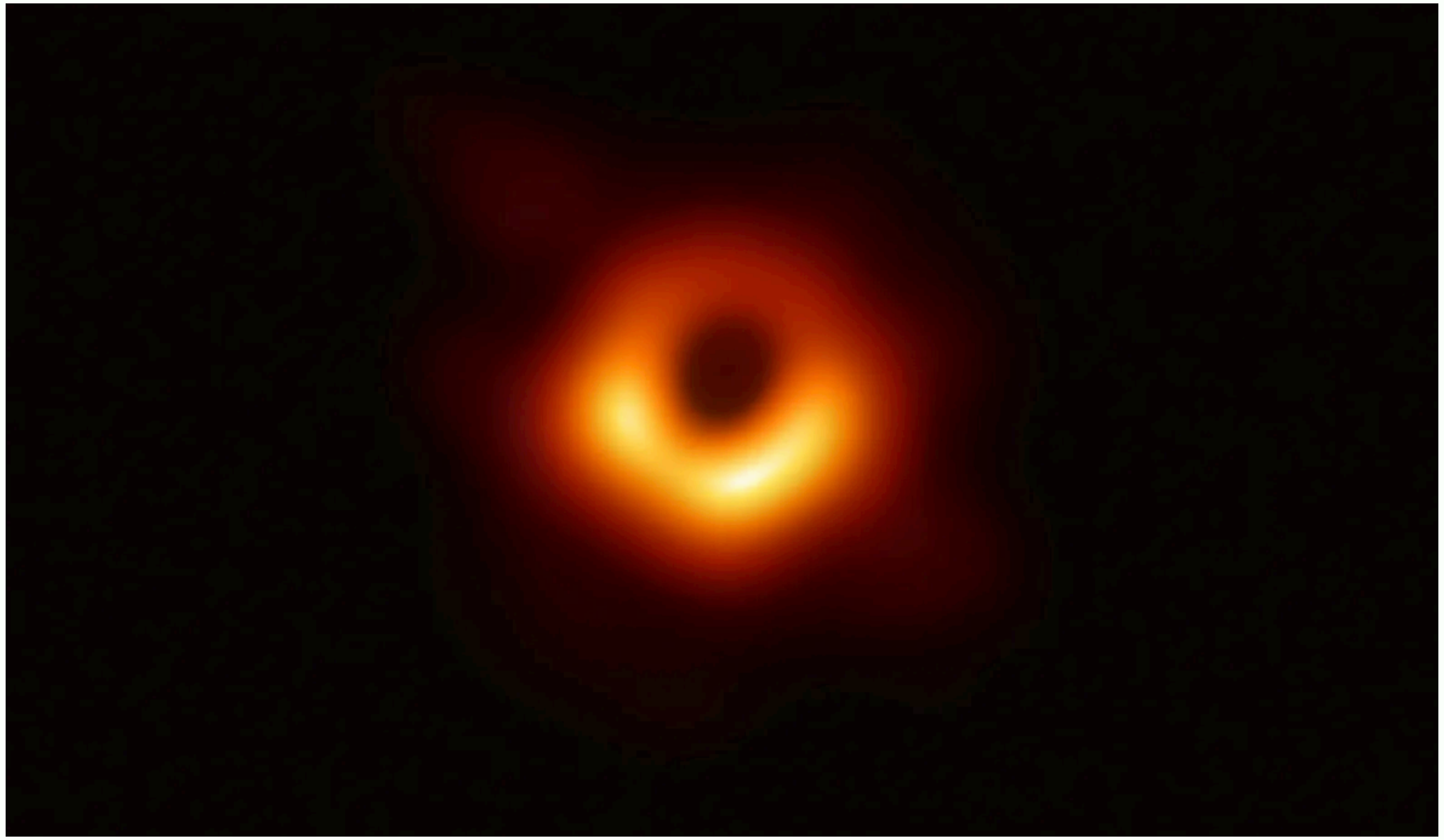


[Accessed June 9, 2025], https://www.science.org/doi/10.1126/science.adk4858
https://newatlas.com/biology/3d-brain-map-connectome-google-harvard/

BLACK HOLE

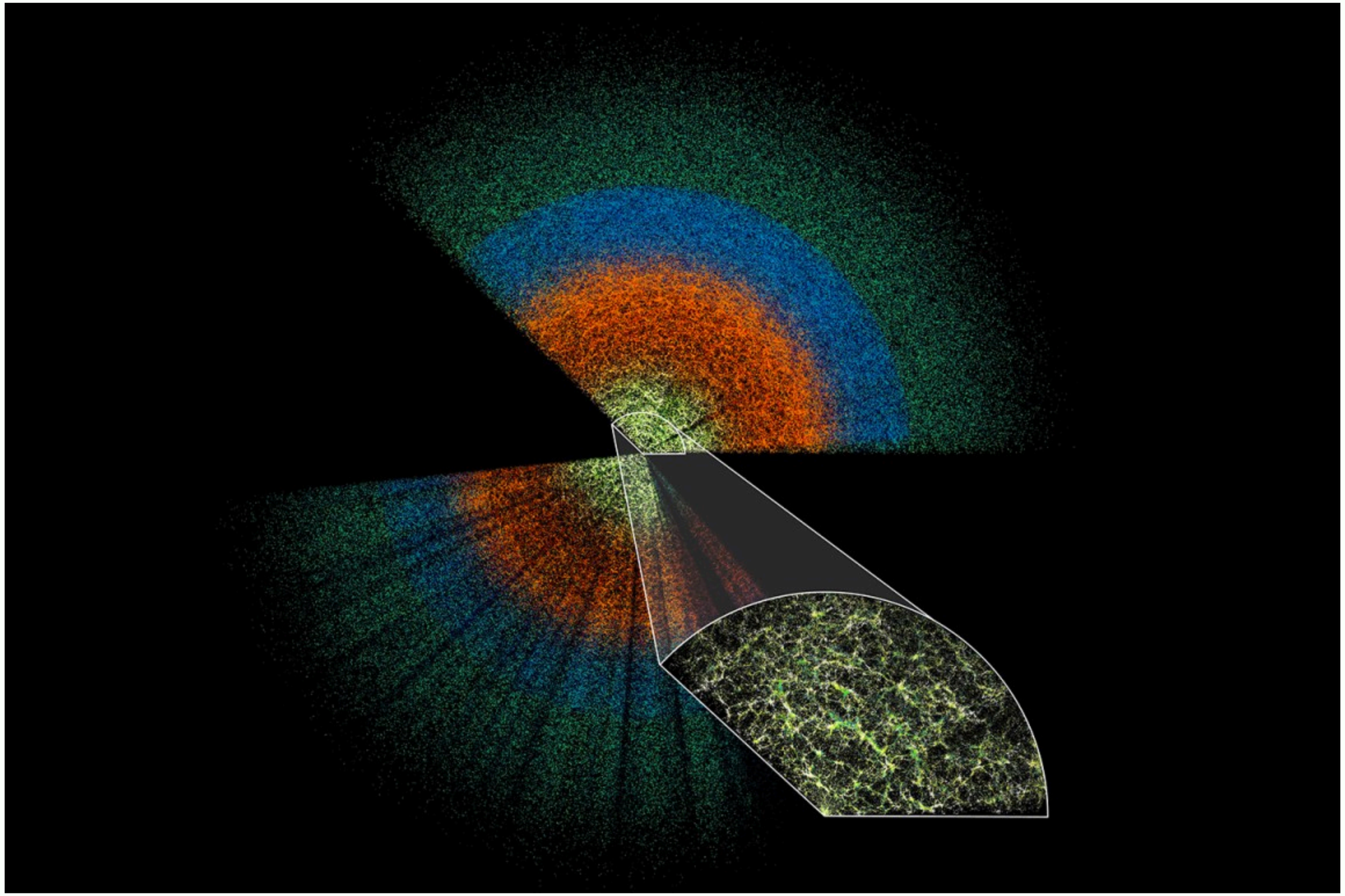
Home > Extreme

It Took Half a Ton of Hard Drives to Store the Black Hole Image Data



[Accessed June 9, 2025], https://www.extremetech.com/extreme/289423-it-took-half-a-ton-of-hard-drives-to-store-eht-black-hole-image-data
https://science.nasa.gov/resource/first-image-of-a-black-hole/

3D MAP OF THE UNIVERSE



Key Takeaways

- The Dark Energy Spectroscopic Instrument collaboration has publicly released the first 13 months of data from its main survey – a treasure trove that will help other researchers investigate big questions in astrophysics.
- Although DESI's Data Release 1 is only a fraction of what the experiment will capture, the 270-terabyte dataset holds a vast amount of information, including precise distances to millions of galaxies.
- DESI's data release contains more than twice as many unique objects outside our galaxy as in all previous 3D spectroscopic surveys combined.

[Accessed June 9, 2025], https://newscenter.lbl.gov/2025/03/19/desi-opens-access-to-the-largest-3d-map-of-the-universe-yet/

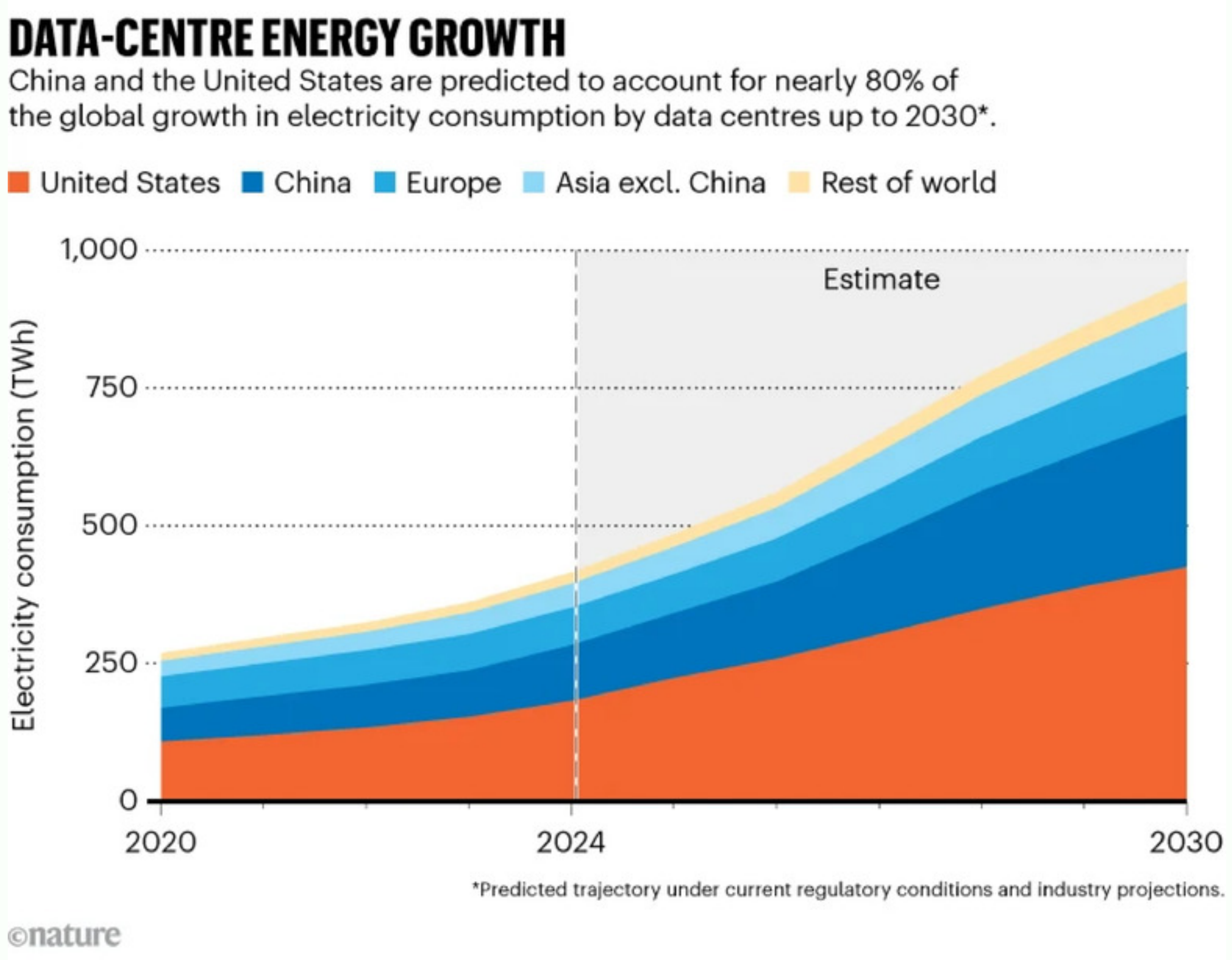
DATA CENTRES

APRIL 10, 2025 | 2 MIN READ

Data Centers Will Use Twice as Much Energy by 2030—Driven by AI

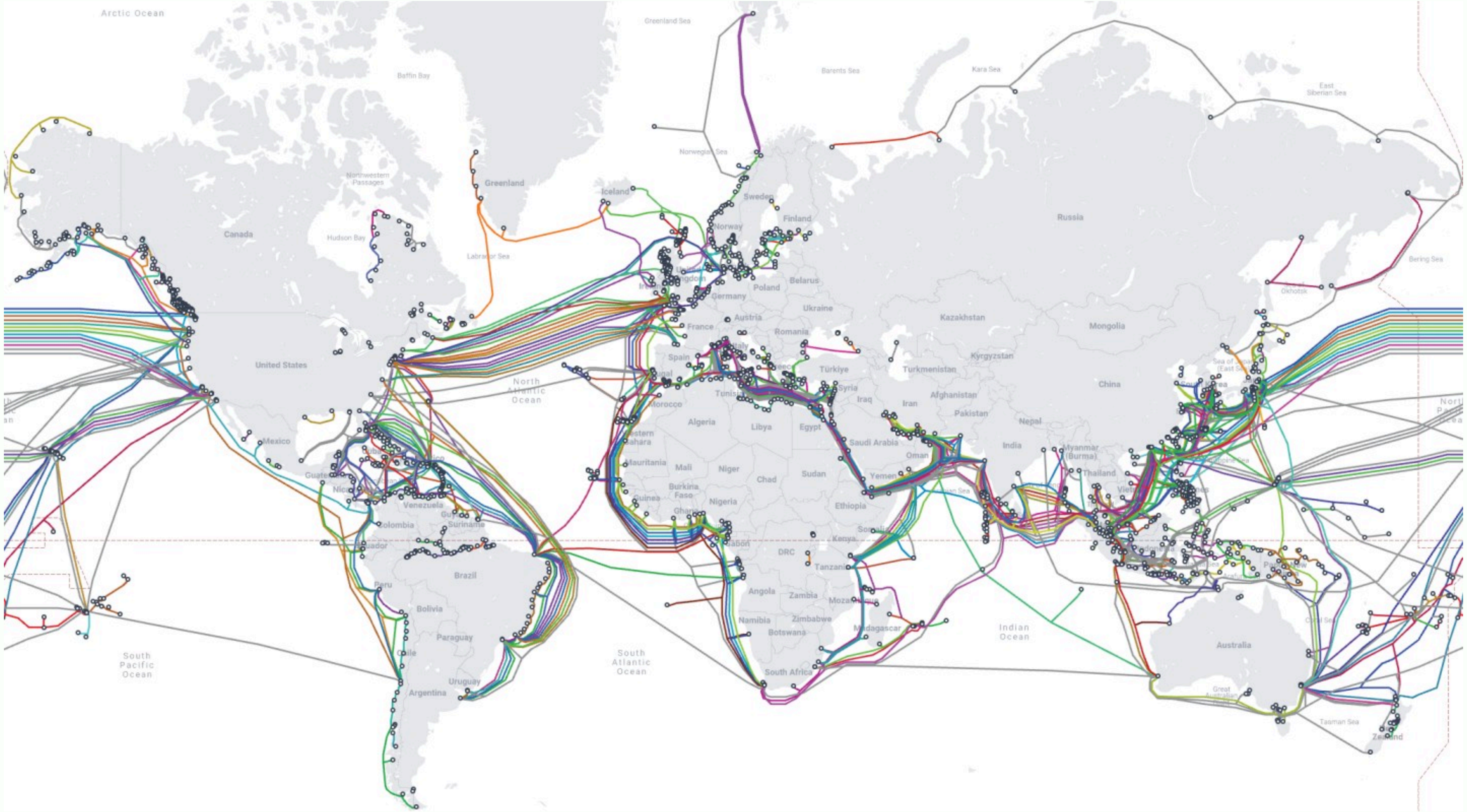
Data centers accounted for about 1.5 percent of global electricity consumption in 2024, an amount expected to double by 2030 because of AI use

BY SOPHIA CHEN & NATURE MAGAZINE



[Accessed June 9, 2025], https://www.scientificamerican.com/article/ai-will-drive-doubling-of-data-center-energy-demand-by-2030/

SUBMARINE CABLE MAP



[Accessed June 9, 2025], https://www.submarinecablemap.com/


DATA ARE THE NEW
GOLD NUGGETS

Data Scientist: The Sexiest Job of the 21st Century

Meet the people who can coax treasure out of messy, unstructured data. by Thomas H. Davenport and DJ Patil

From the Magazine (October 2012)

[Accessed June 11, 2025], https://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century
https://hbr.org/2022/07/is-data-scientist-still-the-sexiest-job-of-the-21st-century

Data Scientist: The Sexiest Job of the 21st Century

Meet the people who can coax treasure out of messy, unstructured data. by Thomas H. Davenport and DJ Patil

From the Magazine (October 2012)

Is Data Scientist Still the Sexiest Job of the 21st Century?

by Thomas H. Davenport and DJ Patil

July 15, 2022

[Accessed June 11, 2025], https://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century
https://hbr.org/2022/07/is-data-scientist-still-the-sexiest-job-of-the-21st-century

WHAT / HOW / WHY?

WHAT? Data science is the interdisciplinary field that combines:

HOW?
- ✓ mathematics and statistics,
- ✓ specialised programming,
- ✓ advanced analytics,
- ✓ artificial intelligence (AI) and machine learning.

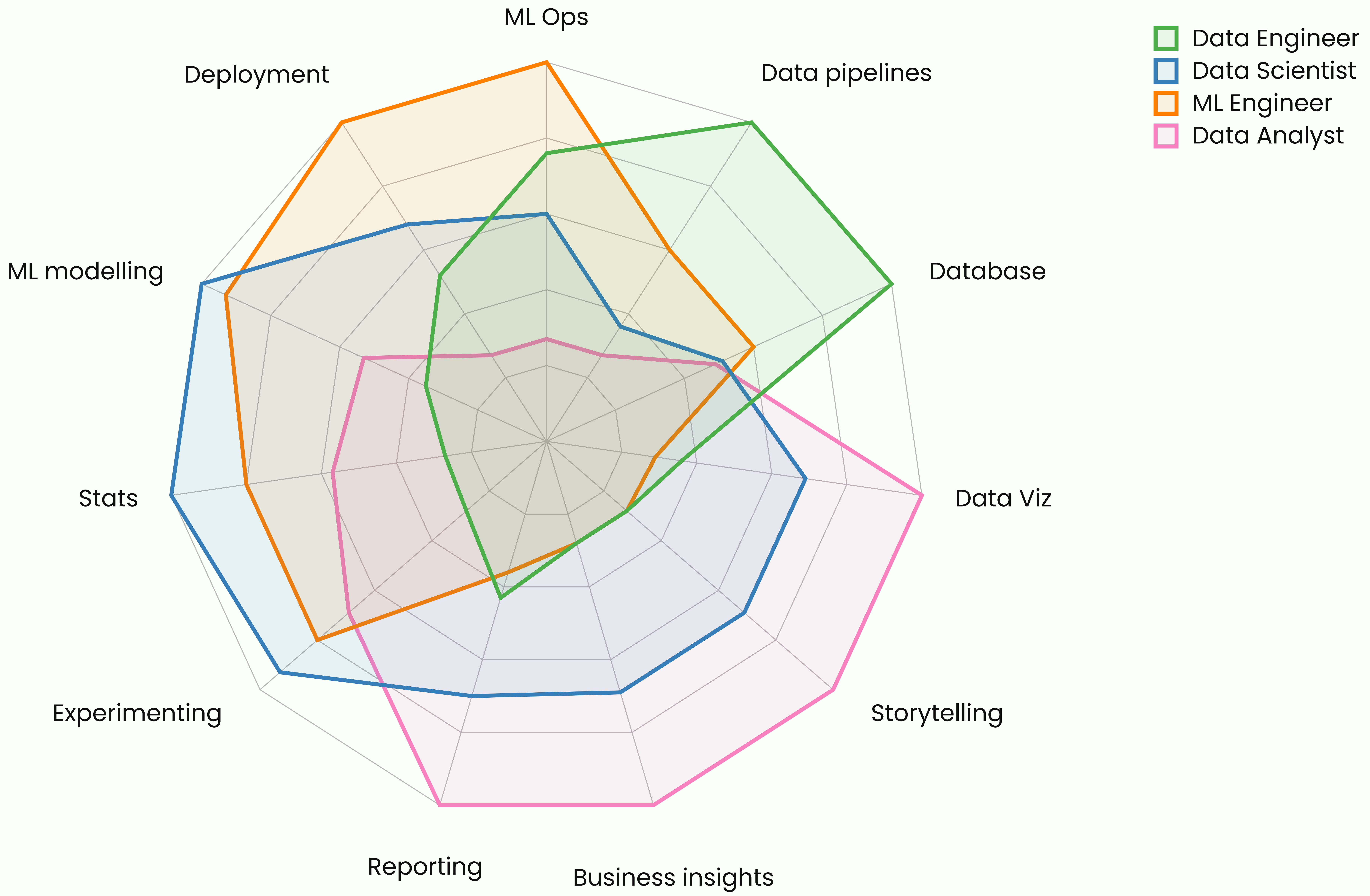
WHY? To extract insights and knowledge from data.

With specific subject matter expertise to uncover actionable insights hidden in an organisation’s data.

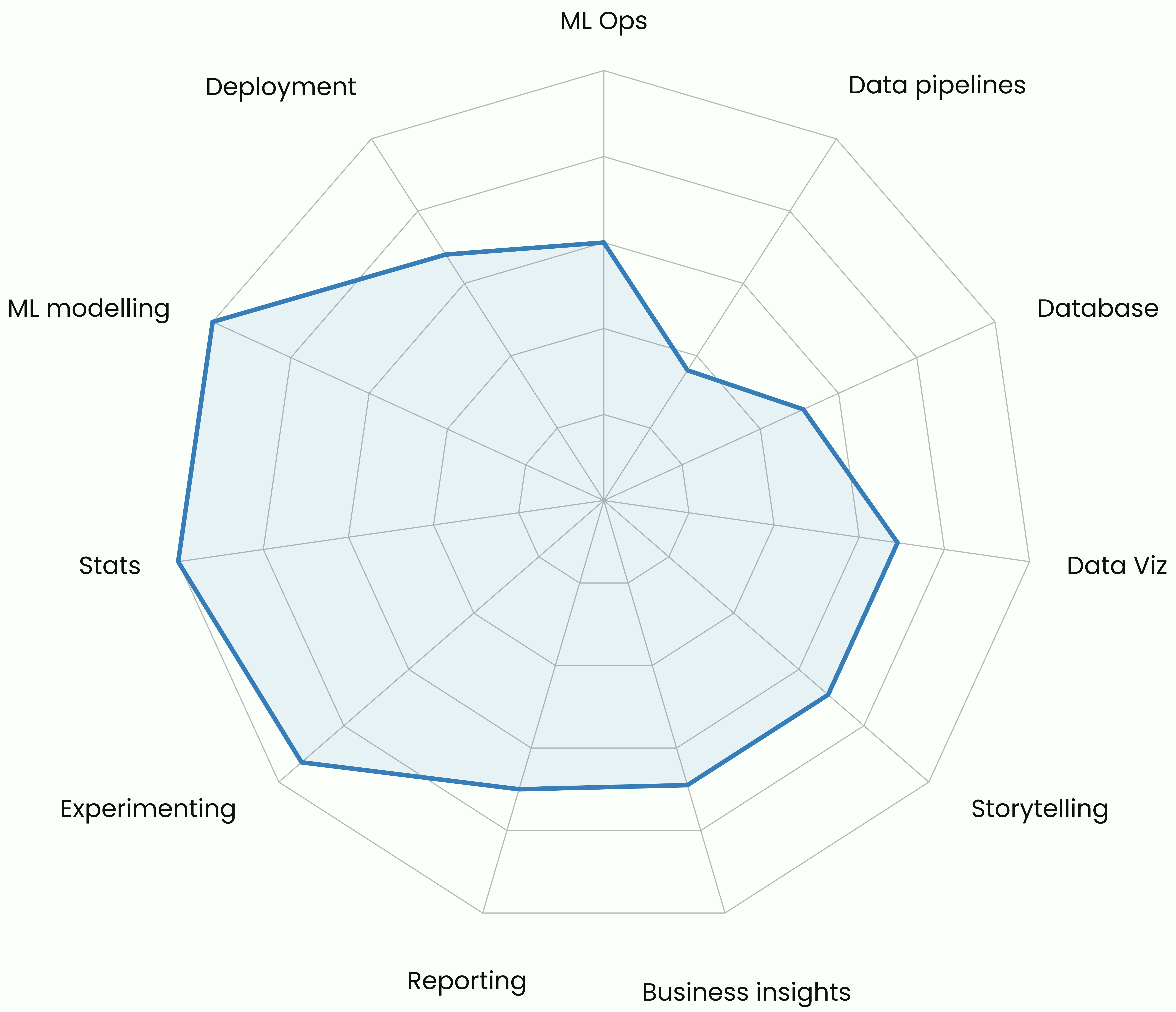
These insights can be used to guide decision making and strategic planning.

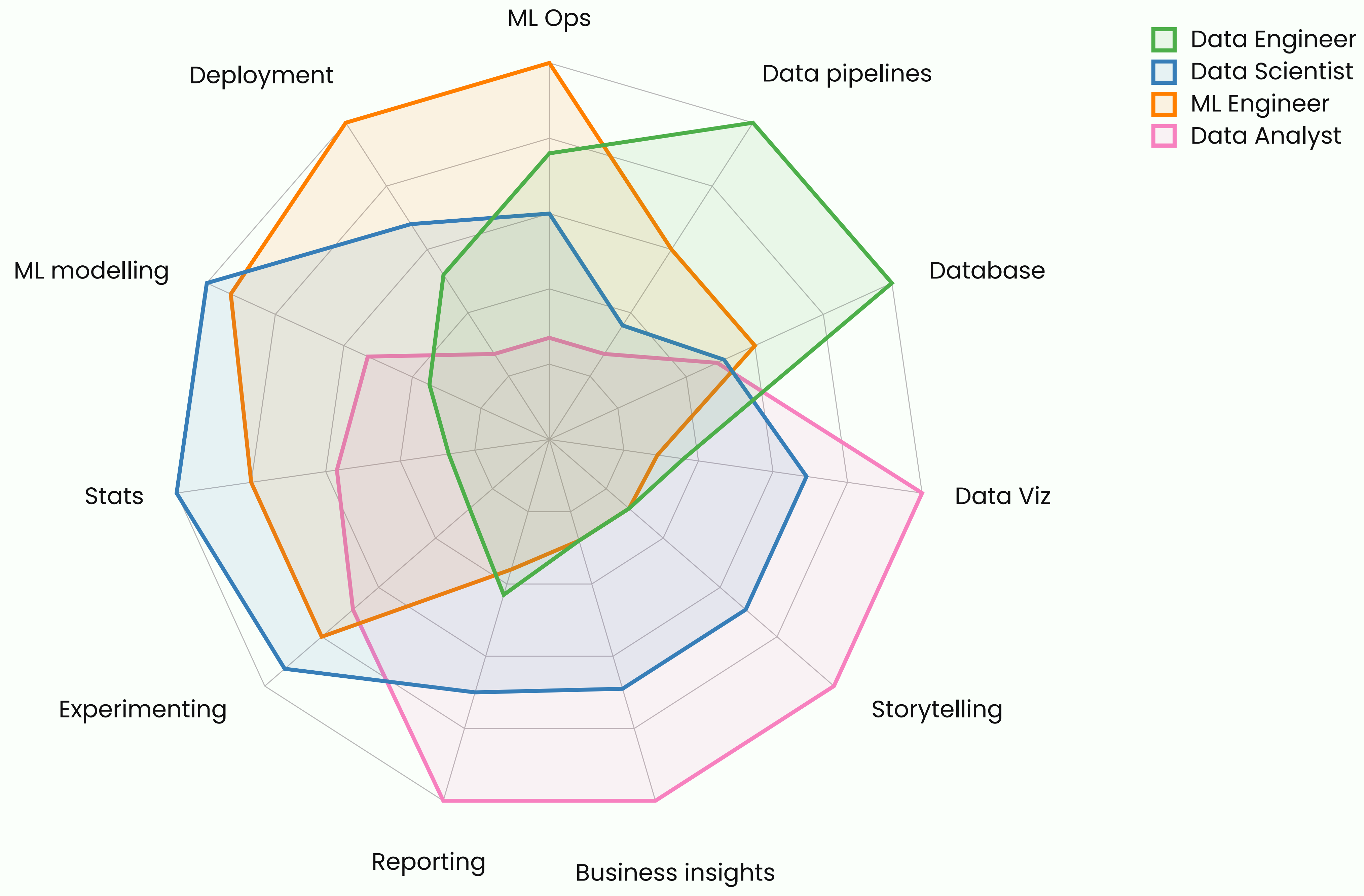
by IBM

[Accessed June 11, 2025], https://www.ibm.com/topics/data-science



[Accessed June 18, 2025], https://4techviews.net/distinguishing-data-roles-analysts-scientists-engineers-and-ml-engineers/



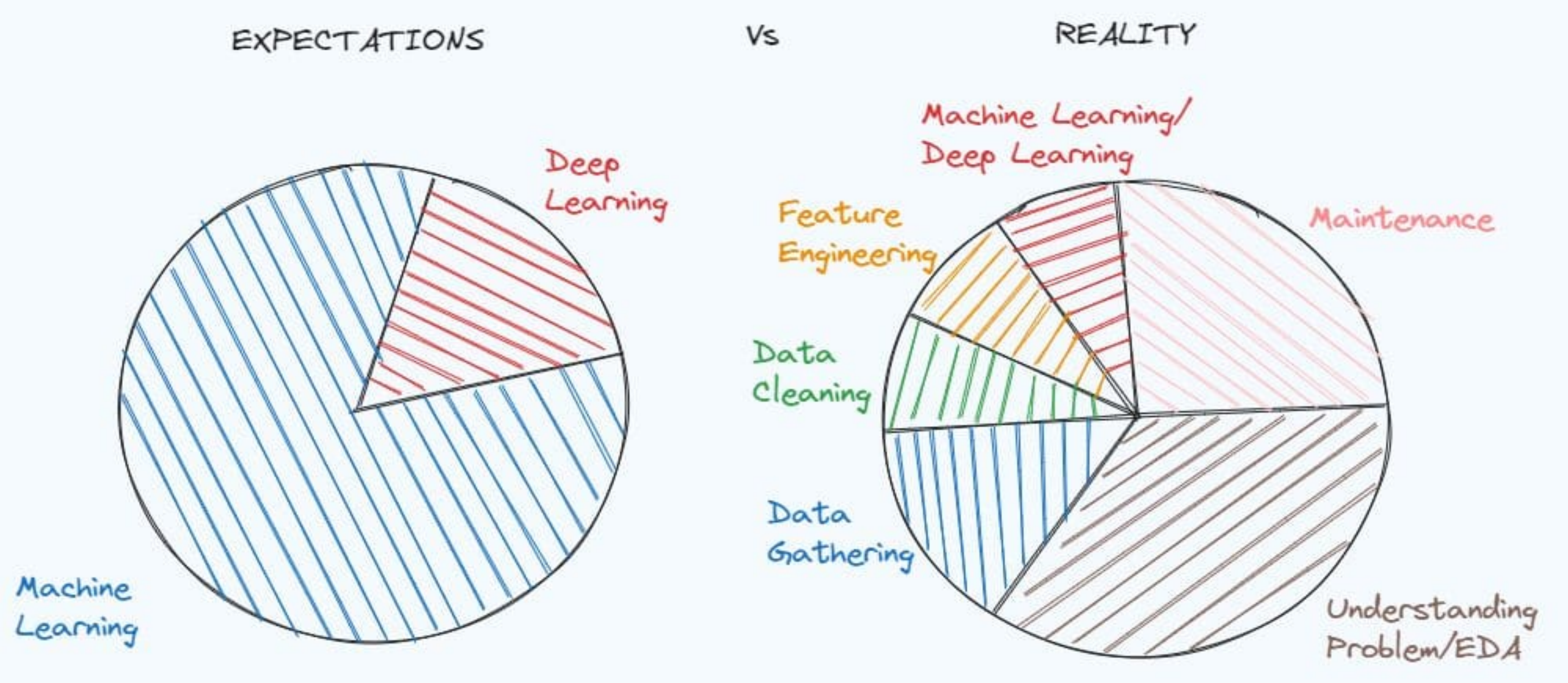


[Accessed June 18, 2025], https://4techviews.net/distinguishing-data-roles-analysts-scientists-engineers-and-ml-engineers/

DATA SCIENCE IN ENERGY INDUSTRY

Illustrative applications:

- Asset degradation and failure prediction, anomaly detection
- Recommendations - process and operations, inspection, maintenance
- Classification and clustering - safety incidents, failure modes, similar service
- Forecasting - production, demand, economical analysis
- Process optimisation

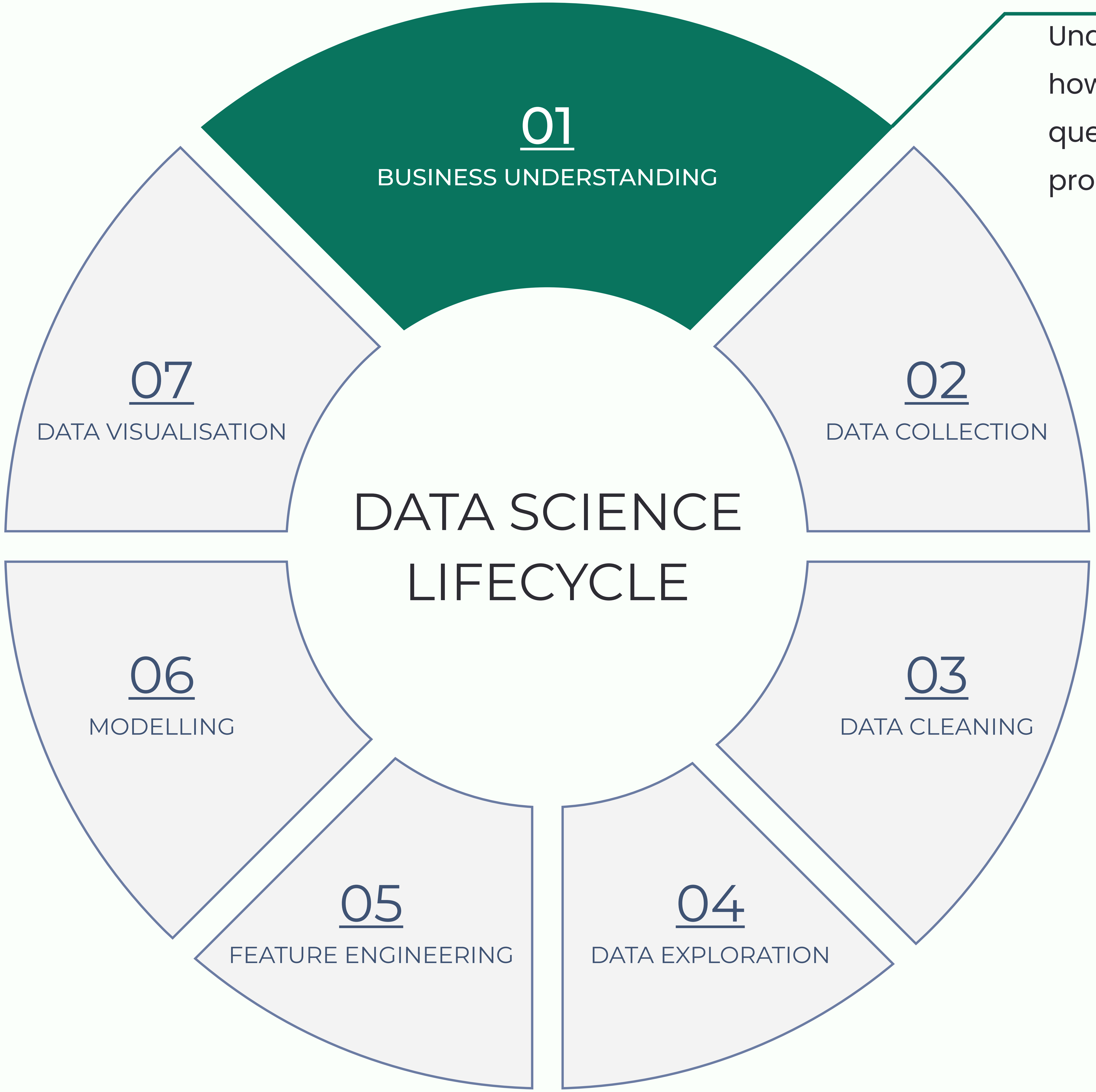


[Accessed June 11, 2025], https://www.kdnuggets.com/2023/08/7-steps-mastering-data-cleaning-preprocessing-techniques.html

Step by step

Data Science Lifecycle

with illustrative example



Understand the business use case and how it can be improved. Ask relevant questions and define objectives for the problem that needs to be solved.

Adopted from https://nearlearn.com/blog/machine-learning-data-science-life-cycle-whats-the-difference/

01

BUSINESS UNDERSTANDING

Understand the business use case and how it can be improved. Ask relevant questions and define objectives for the problem that needs to be solved.

> *At this stage, we need to ask: How to solve a problem?*

With this question, Kidlin's law comes into the game. Kidlin's law is a problem-solving theory that says: *"If you write the problem down clearly, then the matter is half solved."* It may sound as a simple sentence at first, but it has a significant depth in meaning.

[Accessed June 11, 2025], https://medium.com/@hhuseyincosgun/solve-any-problem-by-kidlins-law-4cb5ab67fd7

BUSINESS UNDERSTANDING

Consider a problem of developing a day-ahead electricity demand forecast model.

NORDPOOL DAY-AHEAD MARKET

Daily at 12:00 CET, all purchase and sell orders are aggregated into two curves for each delivery hour.

Optimisation algorithm finds the cheapest set of power plants (supply) to cover the expected demand for all 24h (next day):

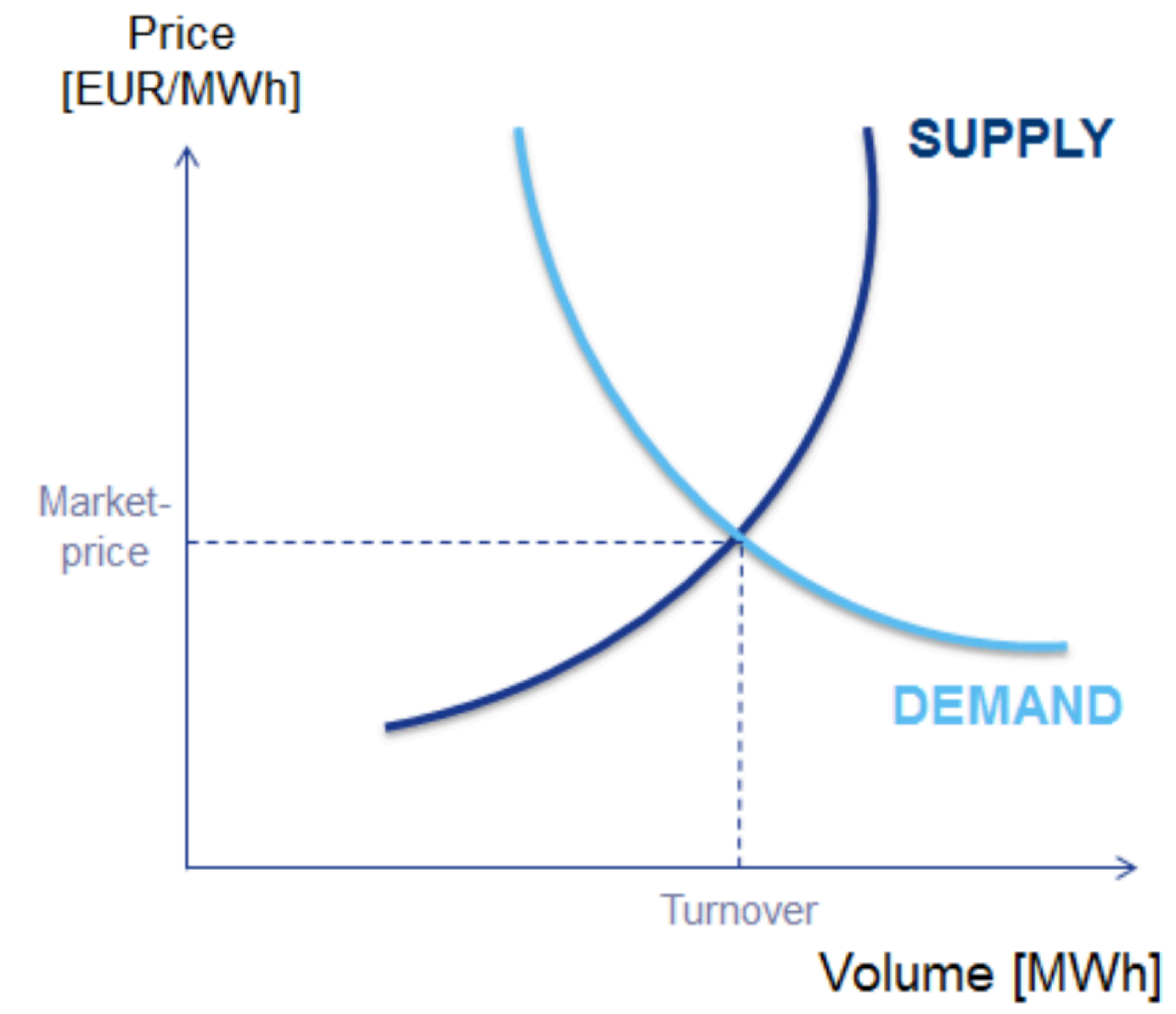
$$\min S = P_1 Q_1 + \cdots + P_n Q_n, \\ \text{s.t.}\, S = D$$

S is supply

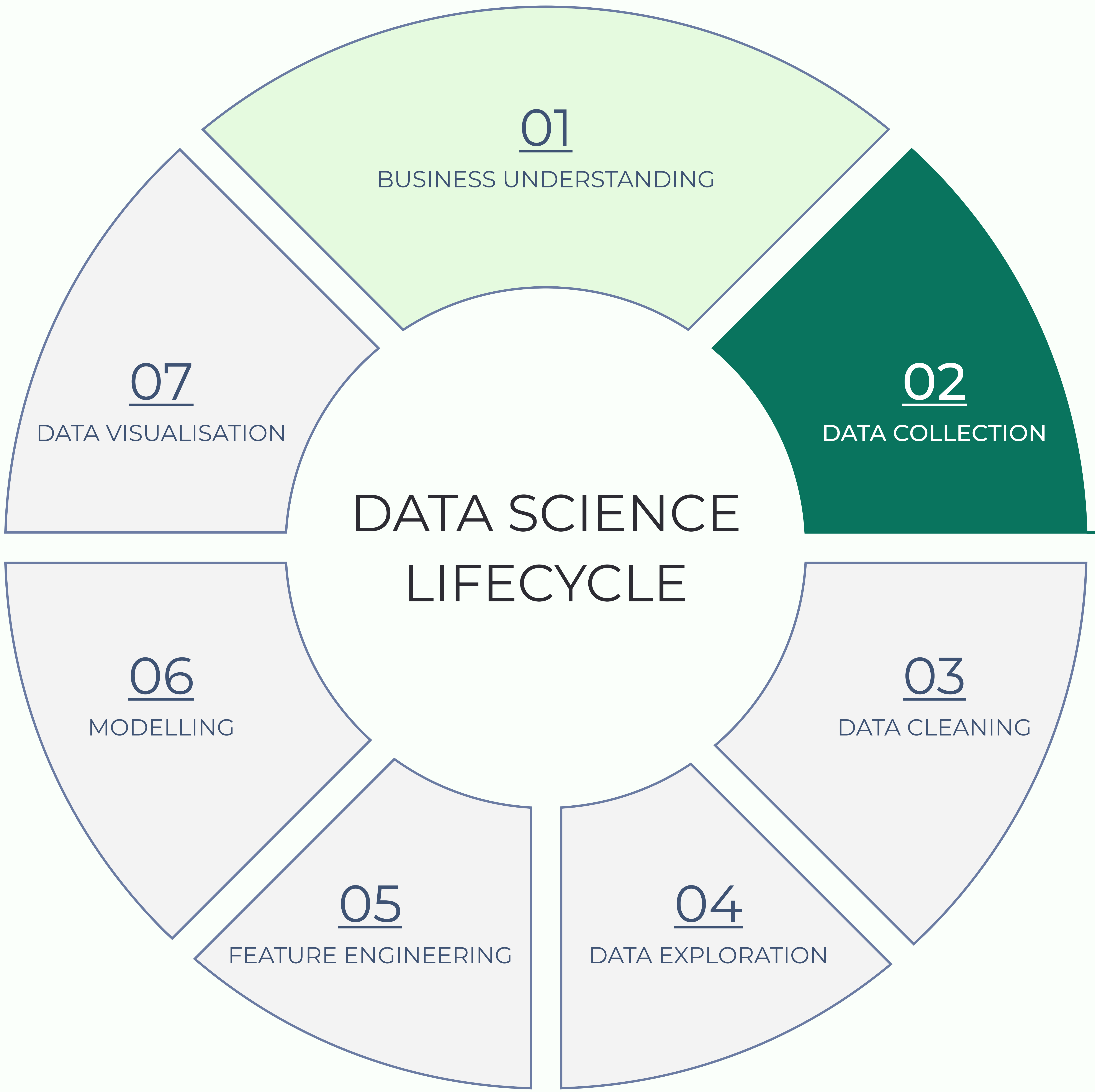
D is demand

P is power plants' production price

Q quantify that the power plan can produce at the offered price

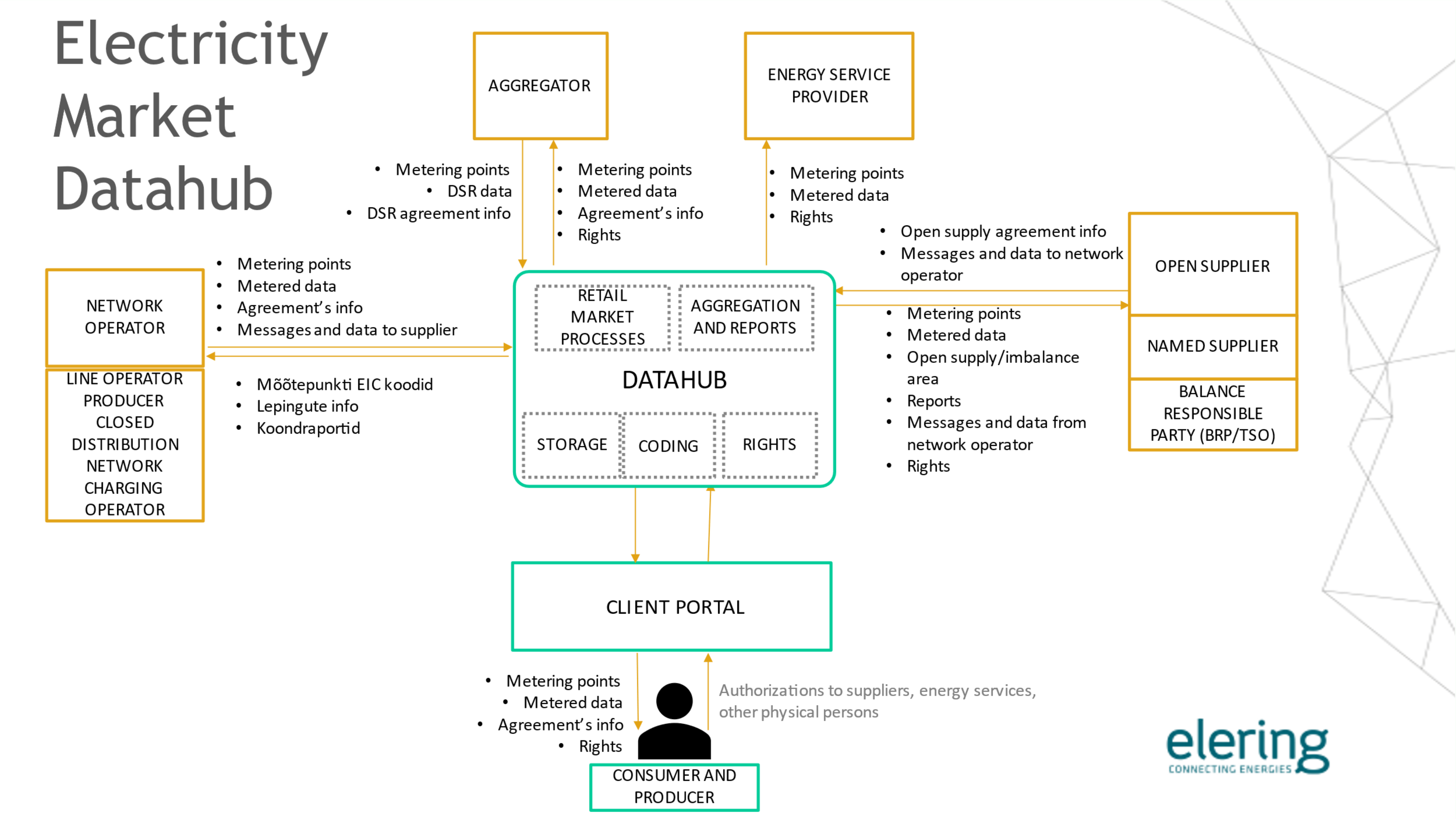


[Accessed June 11, 2025], https://www.nordpoolgroup.com/globalassets/download-center/day-ahead/the-market-setup-in-the-baltics.pdf



Collect the data necessary for the project.

ELERING DATAHUB



[Accessed June 11, 2025], https://elering.ee/en/energy-market
https://github.com/Elering/estfeed-datahub-docs/blob/main/eng/01.1-datahub-description.md

ELECTRICITY DATA

Sample of electricity production/demand data (in MWh)

Timestamp (UTC)	Date (Estonia time)	Demand	Production	Planned demand	Planned production
1672524000	01.01.2023 00:00	798,2	543,1	830,6	549,3
1672527600	01.01.2023 01:00	793,4	552,8	816,3	556,8
1672531200	01.01.2023 02:00	776,5	546,5	848,6	528,3
1672534800	01.01.2023 03:00	757	545,3	814,2	516,5
1672538400	01.01.2023 04:00	743,7	503,1	832,1	497,2
1672542000	01.01.2023 05:00	737,6	483,5	818	478,4
1672545600	01.01.2023 06:00	749,4	455,6	834,2	457

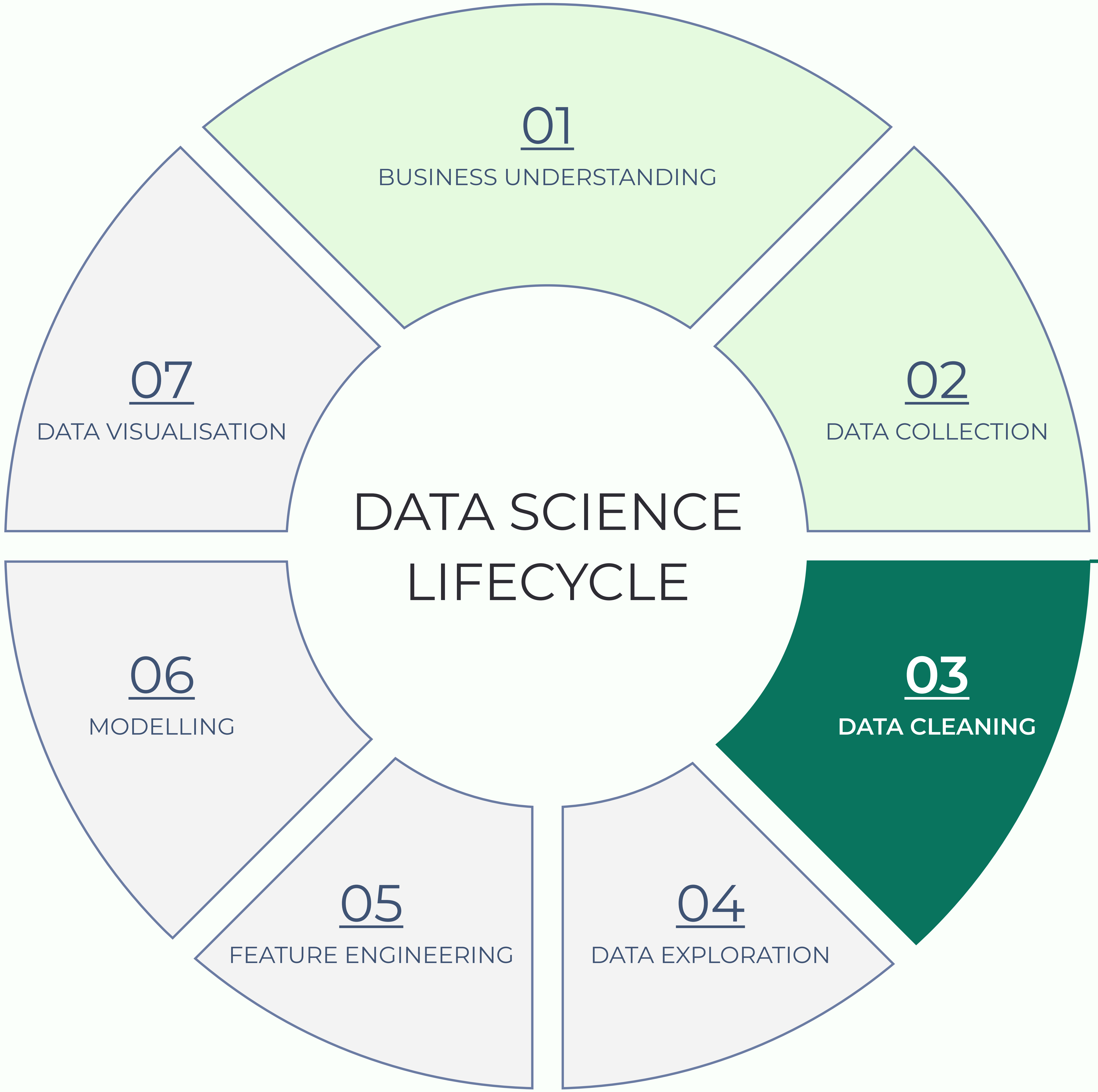
https://dashboard.elering.ee/en/system/with-plan/production-consumption

WEATHER DATA

Sample of weather related data

Local time in Tallinn (airport)	T	Po	P	Pa	U	DD	Ff	ff10	ff3	N	WW	W1	W2	Tn	Tx	Cl	Nh	H	Cm	Ch	VV	Td	RRR	tR
01.01.2023 00:00	5.8	744.1	747.1	-0.3	91	Wind blowing from the south-west	4	8	9	100%	Precipitation	Showers or intermittent precipitation	No significant weather observed	3.2	6.0		100%	200-300			35.0	4.5	0.1	3
01.01.2023 01:00	5.8	744.0	747.1	-0.3	88	Wind blowing from the west-southwest	5		9	60%	No significant weather observed	Showers or intermittent precipitation	No significant weather observed	3.6	6.0		60%	300-600			35.0	4.0	No precipitation	3
01.01.2023 02:00	5.2	744.4	747.5	0.3	91	Wind blowing from the west-southwest	4		9	90 or more, but not 100%	State of sky on the whole unchanged	Shower(s)	Cloud covering more than 1/2 of the sky throughout the appropriate period	4.1	6.0	Stratus nebulosus or Stratus fractus other than of bad weather, or both	90 or more, but not 100%	300-600	No Altocumulus, Altostratus or Nimbostratus	No Cirrus, Cirrocumulus or Cirrostratus	35.0	3.8	No precipitation	3
01.01.2023 03:00	4.8	744.8	747.8	0.7	89	Wind blowing from the west	4		9	60%	No significant weather observed.	No significant weather observed.	No significant weather observed.	4.5	6.0		10% or less, but not 0	300-600			35.0	3.1	No precipitation	3
01.01.2023 04:00	4.1	745.3	748.3	1.3	91	Wind blowing from the west	4		9	90 or more, but not 100%	No significant weather observed.	No significant weather observed.	No significant weather observed.	4.1	6.0		90 or more, but not 100%	2500 or more, or no clouds			35.0	2.7	No precipitation	3
01.01.2023 05:00	3.3	745.6	748.6	1.2	93	Wind blowing from the west-southwest	3		9	60%	State of sky on the whole unchanged.	Cloud covering more than 1/2 of the sky throughout the appropriate period.	Cloud covering more than 1/2 of the sky throughout the appropriate period.	3.3	6.0	No Stratocumulus, Stratus, Cumulus or Cumulonimbus	40%.	2500 or more, or no clouds	Altocumulus translucidus at a single level	Cirrocumulus alone, or Cirrocumulus accompanied by Cirrus or Cirrostratus or both, but Cirrocumulus is predominant	35.0	2.2	No precipitation	3

https://rp5.ru/Weather_archive_in_Tallinn_(airport)



Remove data that does not belong in your dataset. Fix inconsistencies within the data and handle missing values.

DATA QUALITY

Variables may be:

- **Qualitative** variables are those that express a qualitative attribute such as cloudiness, hair colour, gender, etc. They are also referred to as categorical variables.
- **Quantitative** variables are those variables that are measured in terms of numbers.
  - **Discrete** variables can take only certain values.
  - **Continuous** variables can take any value within the range of the scale.

DATA CLEANING

Sample of weather related data

Local time in Tallinn (airport)	T	Po	U	DD	N	Tn	Tx	Nh	Cm	Ch
01.01.2023 00:00	5.8	744.1	91	Wind blowing from the south-west	100%	3.2	6.0	100%		
01.01.2023 01:00	5.8	744.0	88	Wind blowing from the west-southwest	60%	3.6	6.0	60%		
01.01.2023 02:00	5.2	744.4	91	Wind blowing from the west-southwest	90 or more, but not 100%	4.1	6.0	90 or more, but not 100%	No Altocumulus, Altostratus or Nimbostratus	No Cirrus, Cirrocumulus or Cirrostratus
01.01.2023 03:00	4.8	744.8	89	Wind blowing from the west	60%	4.5	6.0	10% or less, but not 0		
01.01.2023 04:00	4.1	745.3	91	Wind blowing from the west	90 or more, but not 100%	4.1	6.0	90 or more, but not 100%		
01.01.2023 05:00	3.3	745.6	93	Wind blowing from the west-southwest	60%	3.3	6.0	40%.	Altocumulus translucidus at a single level	Cirrocumulus alone, or Cirrocumulus accompanied by Cirrus or Cirrostratus or both, but Cirrocumulus is predominant

DATA CLEANING

Sample of weather related data

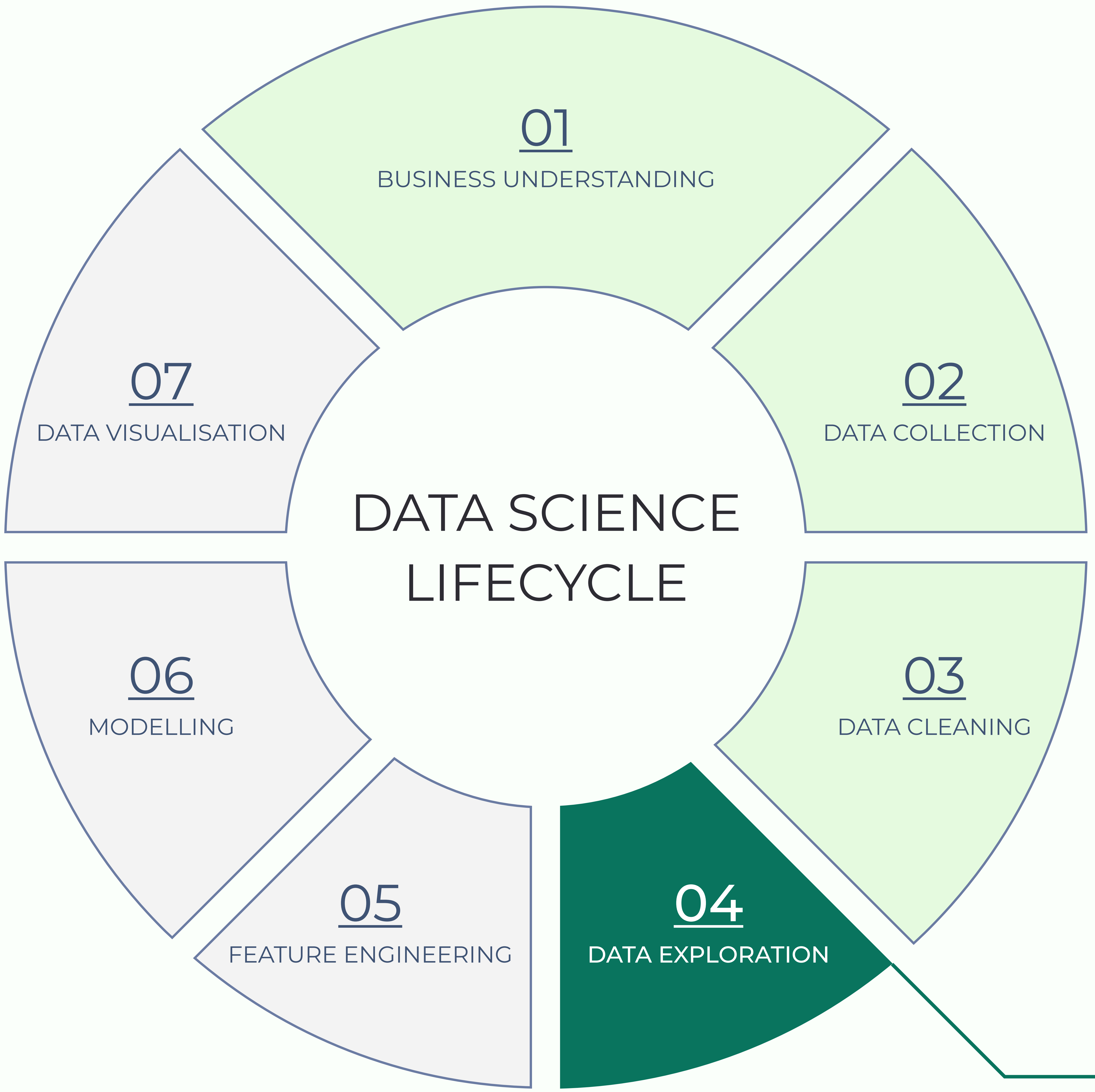
Local time in Tallinn (airport)	T	Po	U	DD	N	Tn	Tx	Nh	Cm	Ch
01.01.2023 00:00	5.8	744.1	91	Wind blowing from the south-west	100%	3.2	6.0	100%		
01.01.2023 01:00	5.8	744.0	88	Wind blowing from the west-southwest	60%	3.6	6.0	60%		
01.01.2023 02:00	5.2	744.4	91	Wind blowing from the west-southwest	90 or more, but not 100%	4.1	6.0	90 or more, but not 100%	No Altocumulus, Altostratus or Nimbostratus	No Cirrus, Cirrocumulus or Cirrostratus
01.01.2023 03:00	4.8	744.8	89	Wind blowing from the west	60%	4.5	6.0	10% or less, but not 0		
01.01.2023 04:00	4.1	745.3	91	Wind blowing from the west	90 or more, but not 100%	4.1	6.0	90 or more, but not 100%		
01.01.2023 05:00	3.3	745.6	93	Wind blowing from the west-southwest	60%	3.3	6.0	40%.	Altocumulus translucidus at a single level	Cirrocumulus alone, or Cirrocumulus accompanied by Cirrus or Cirrostratus or both, but Cirrocumulus is predominant

Still there may remain some anomalies that cannot be easily seen from the table, and further exploration is needed.

HOW TO CLEAN YOUR DATA

While the techniques used for data cleaning may vary according to the types of data, you can still follow these basic steps:

1. Remove duplicates
2. Remove irrelevant observations
3. Fix structural errors
4. Handle unwanted outliers
5. Convert and unify data types
6. Clear formatting
7. Handle missing data

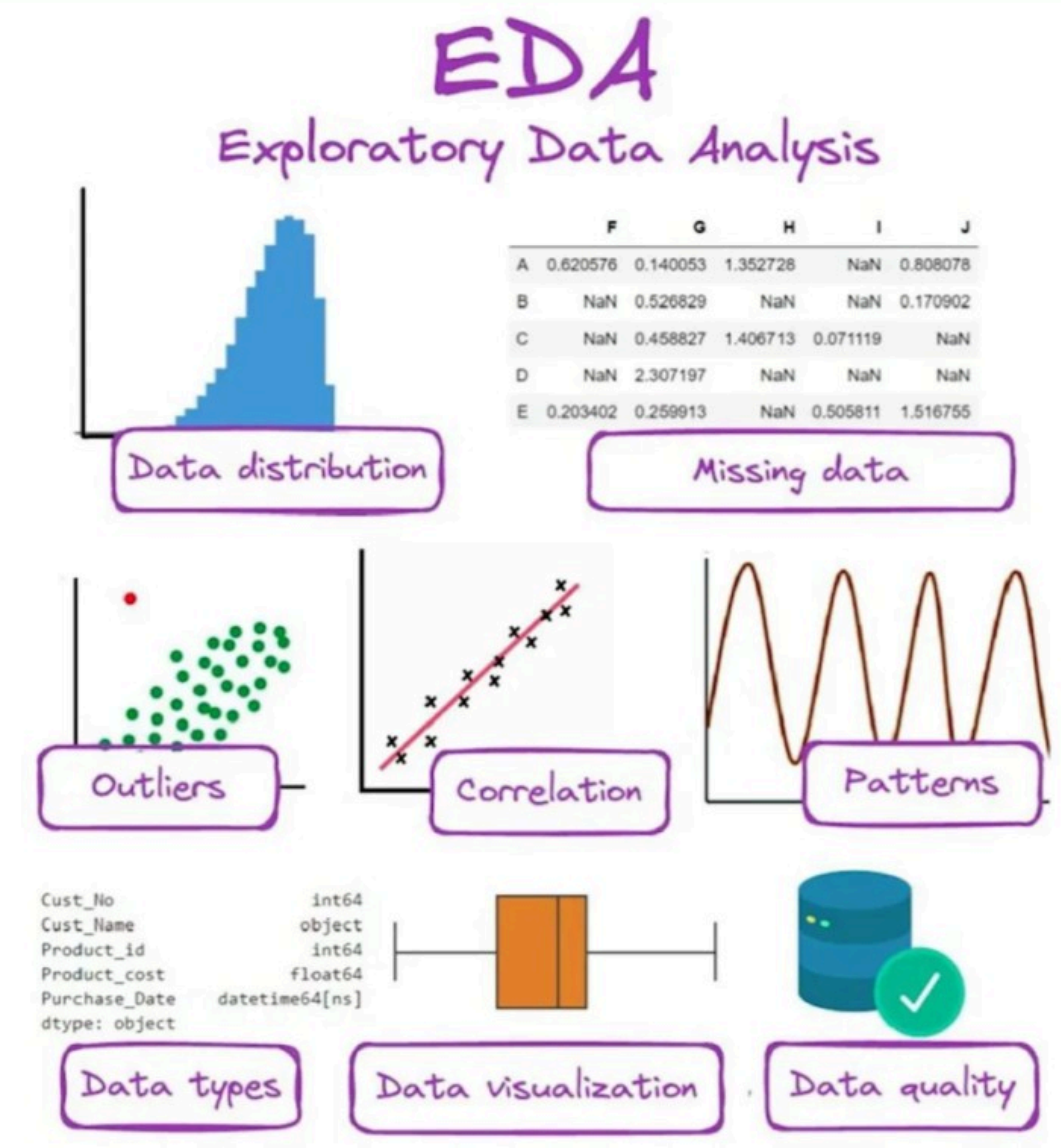


Explore and visualise data to uncover insights from the data and form hypotheses.

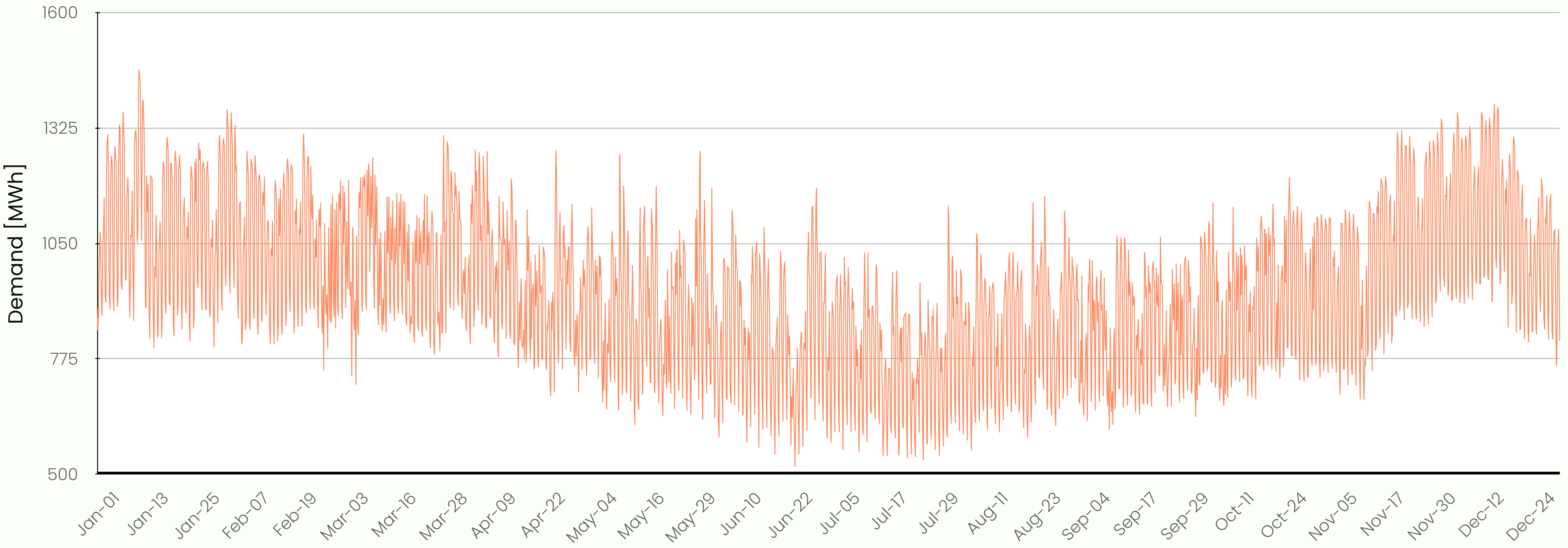
04

DATA EXPLORATION

Explore and visualise data to uncover insights from the data and form hypotheses.

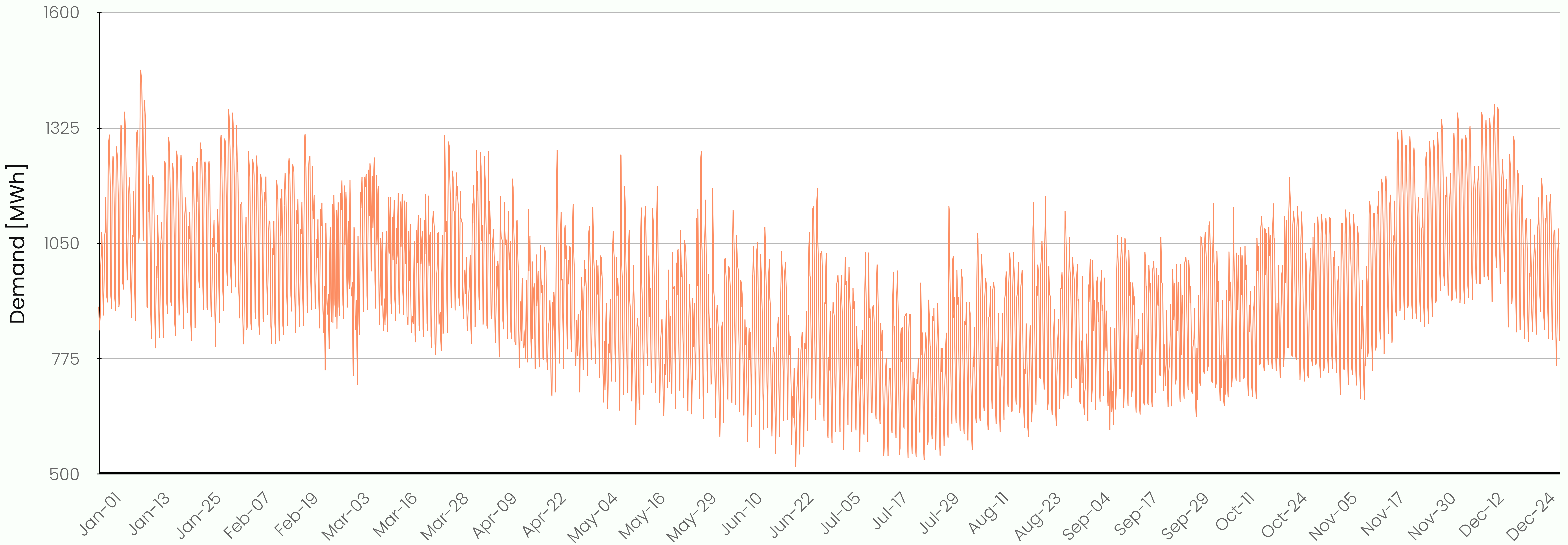


ELECTRICITY DEMAND IN ESTONIA, 2022



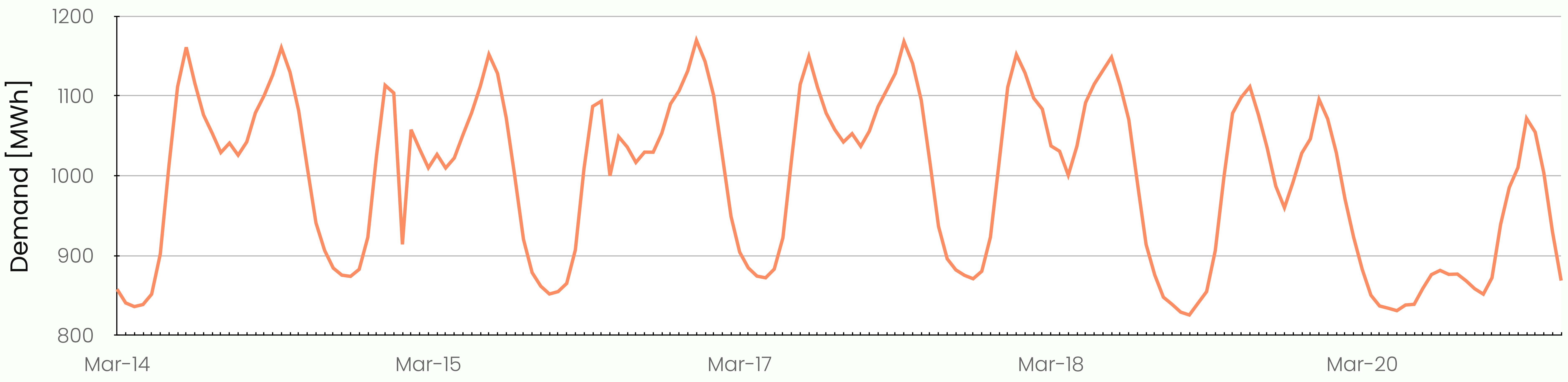
Any observations?

ELECTRICITY DEMAND IN ESTONIA, 2022

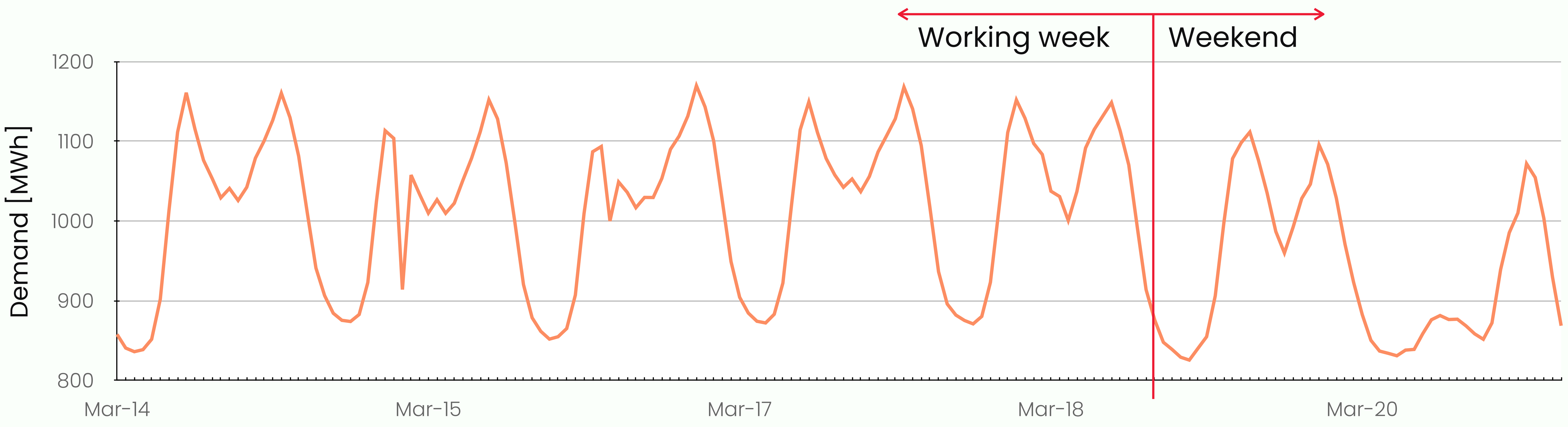


Any observations? Seasonality!

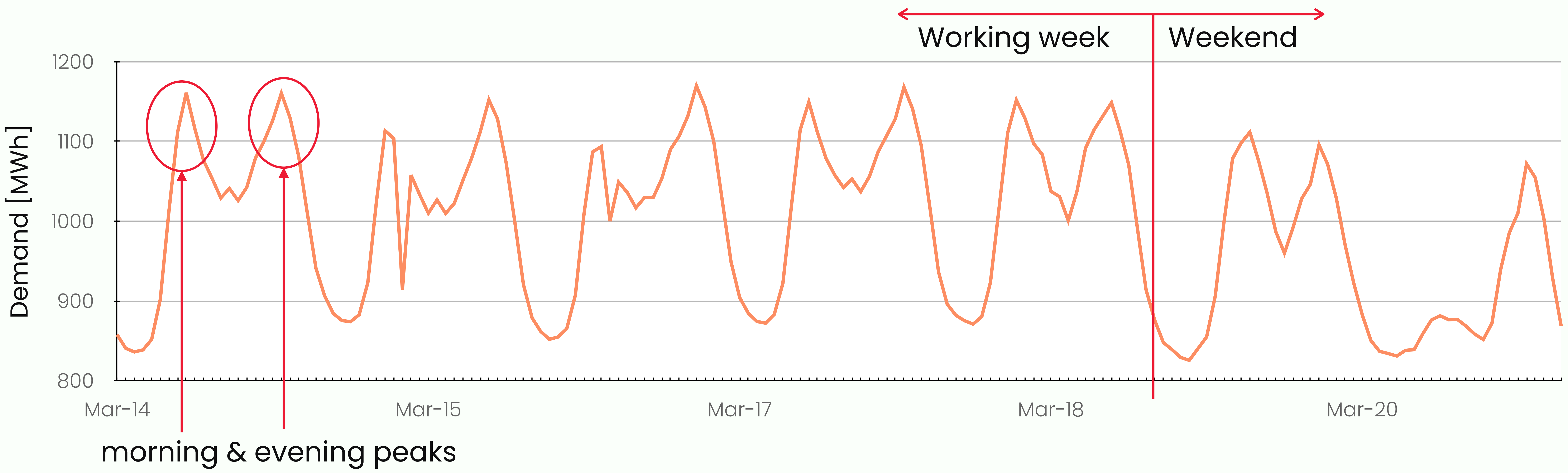
IT'S ALL ABOUT PATTERNS



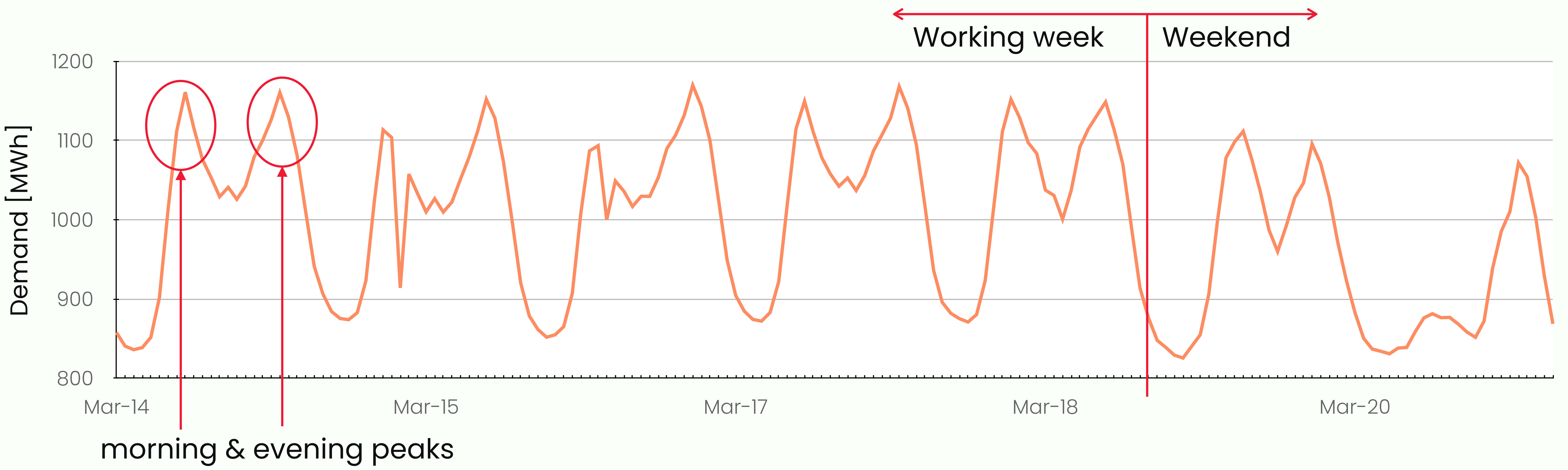
IT'S ALL ABOUT PATTERNS

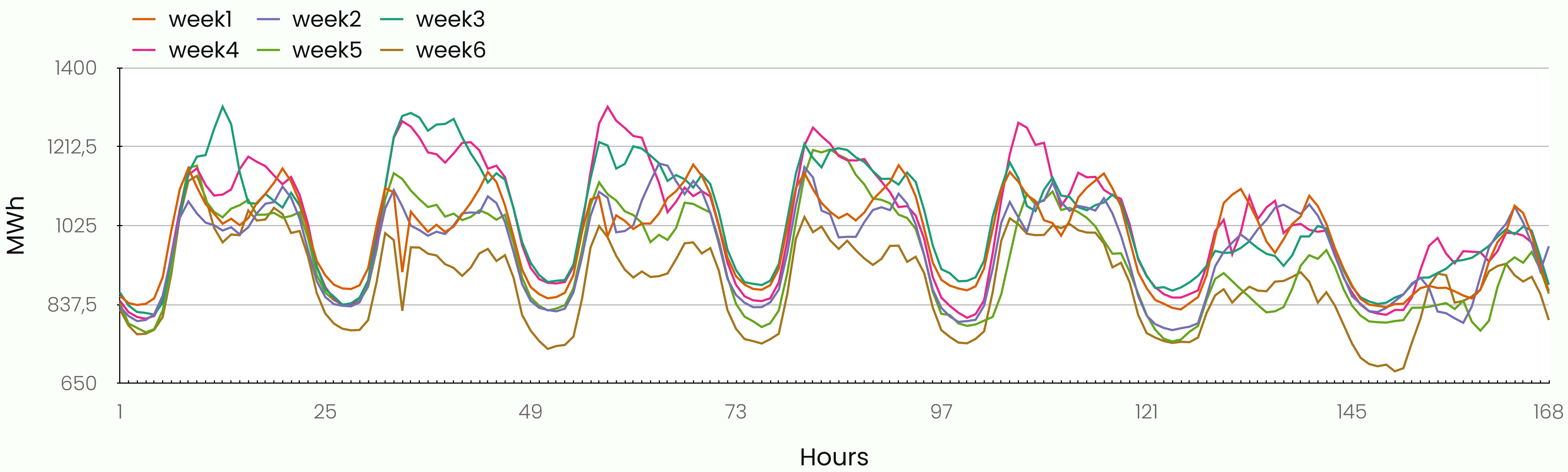


IT'S ALL ABOUT PATTERNS



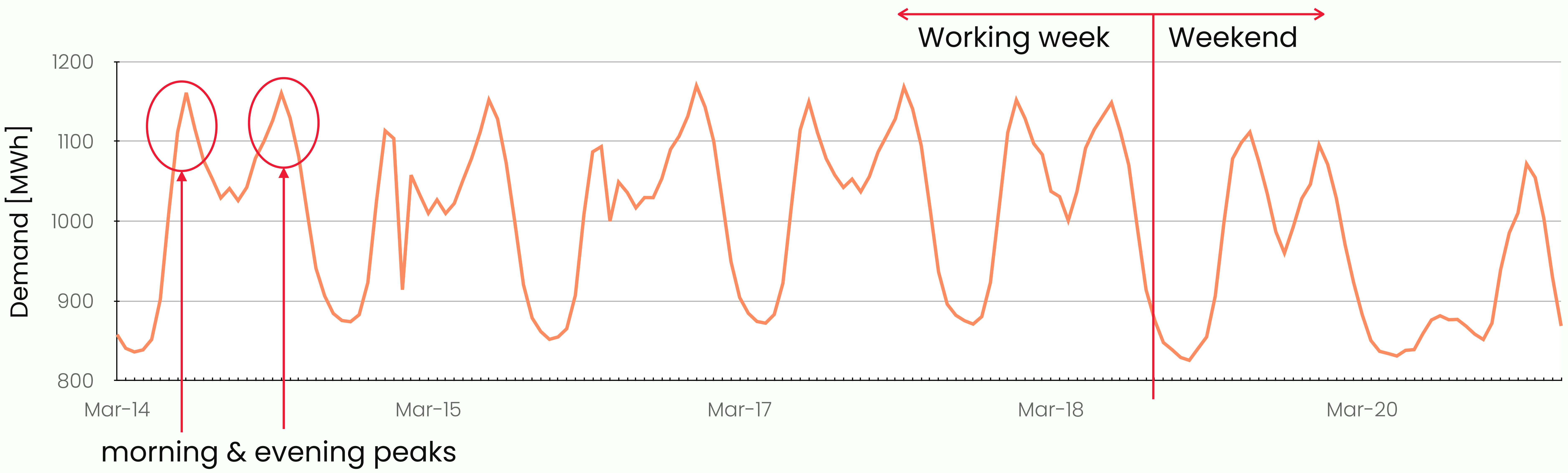
IT'S ALL ABOUT PATTERNS



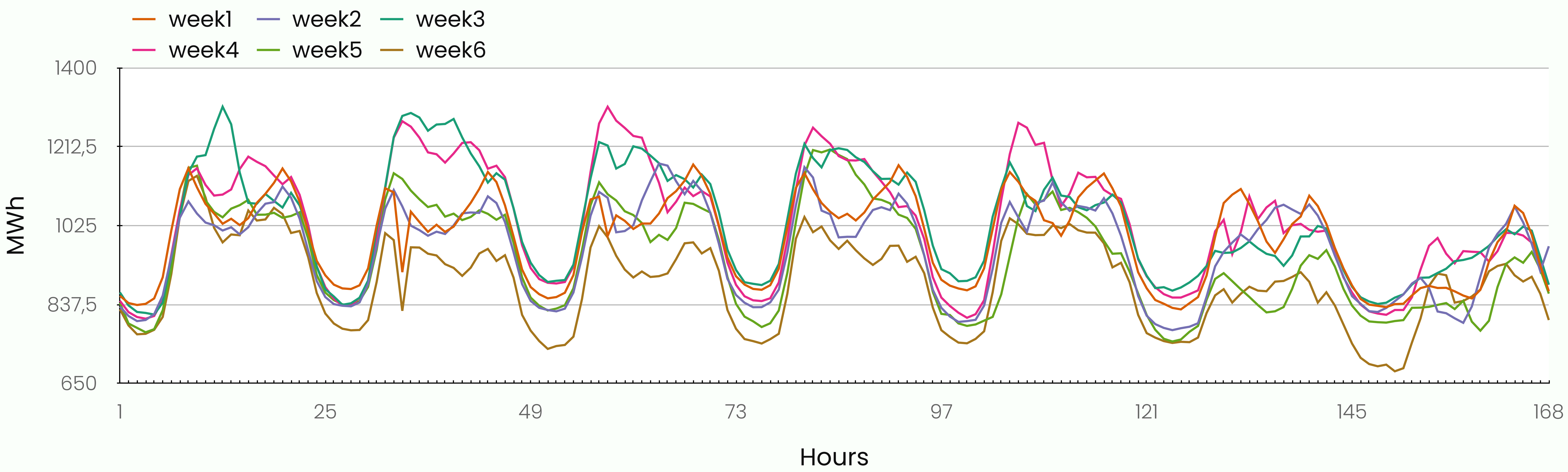


Let's plot several weeks

IT'S ALL ABOUT PATTERNS



Let's plot several weeks



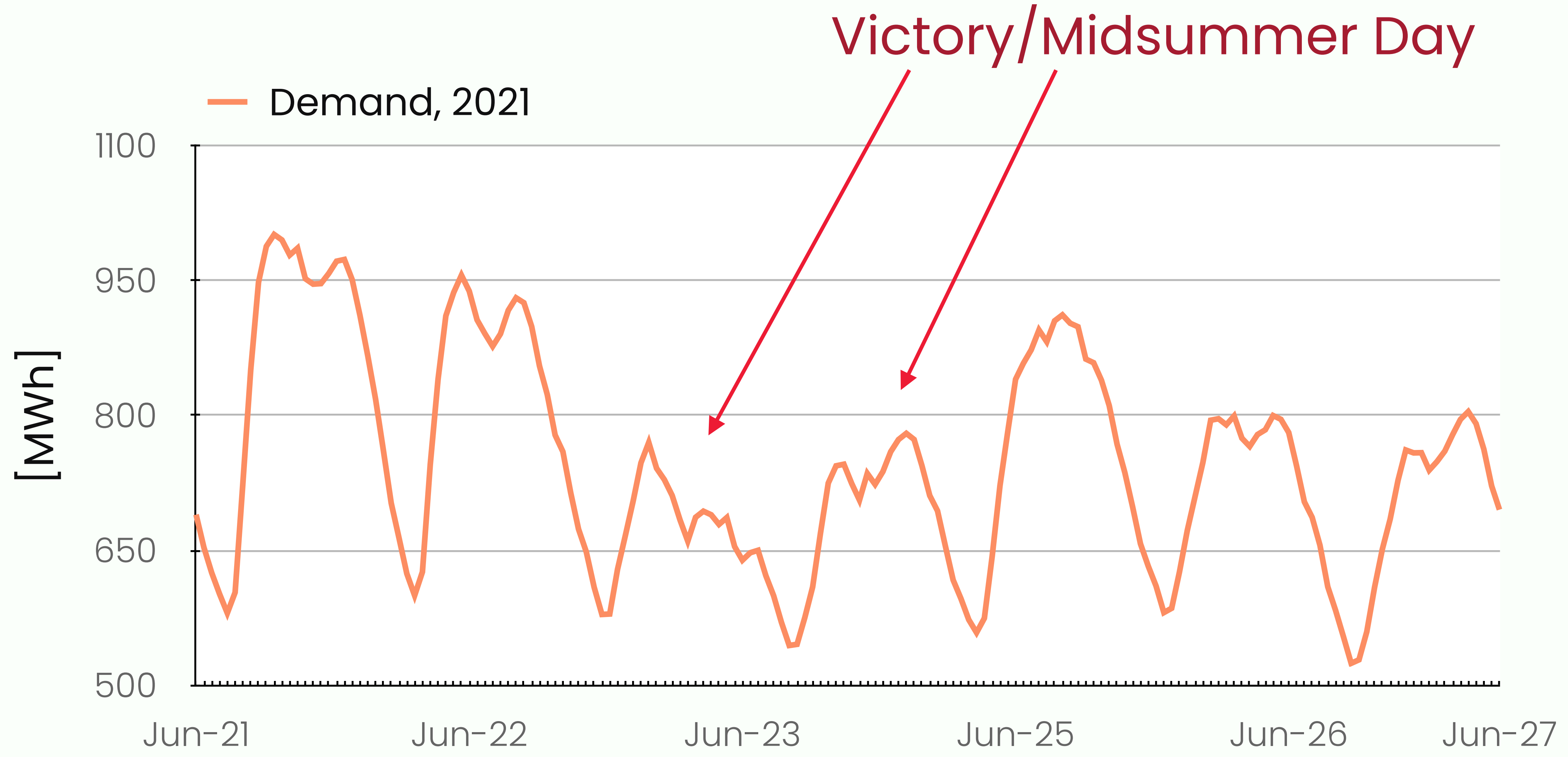
What can go wrong?

HOLIDAYS


December 19-25, 2022
December 26-January 1, 22/23
1400
1225
1050
875
700
[MWh]
1
36
71
106
141
168
Christmas

HOLIDAYS

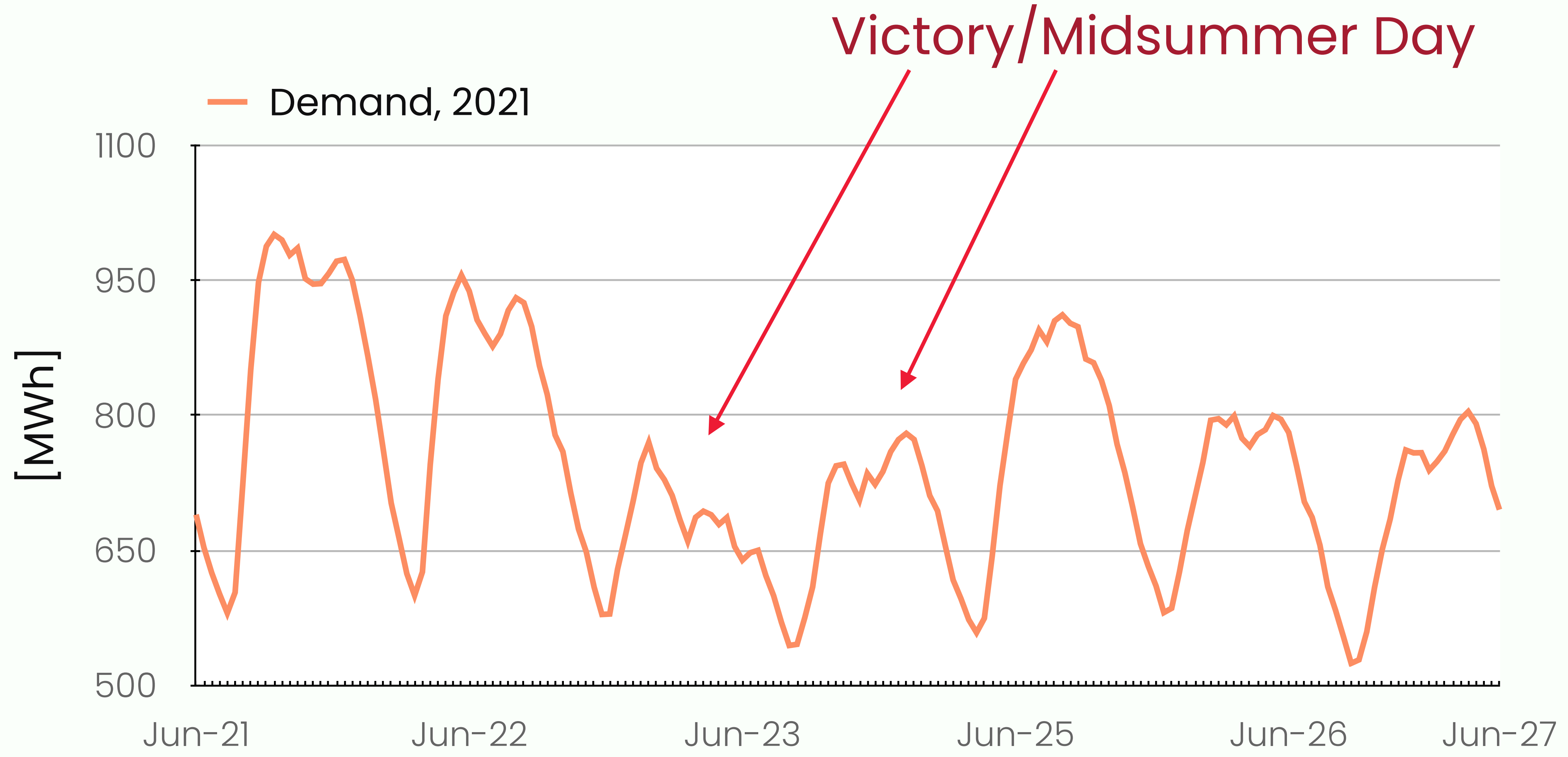
December 19-25, 2022
December 26-January 1, 22/23
1400
1225
1050
875
700
[MWh]
1
36
71
106
141
168
Christmas



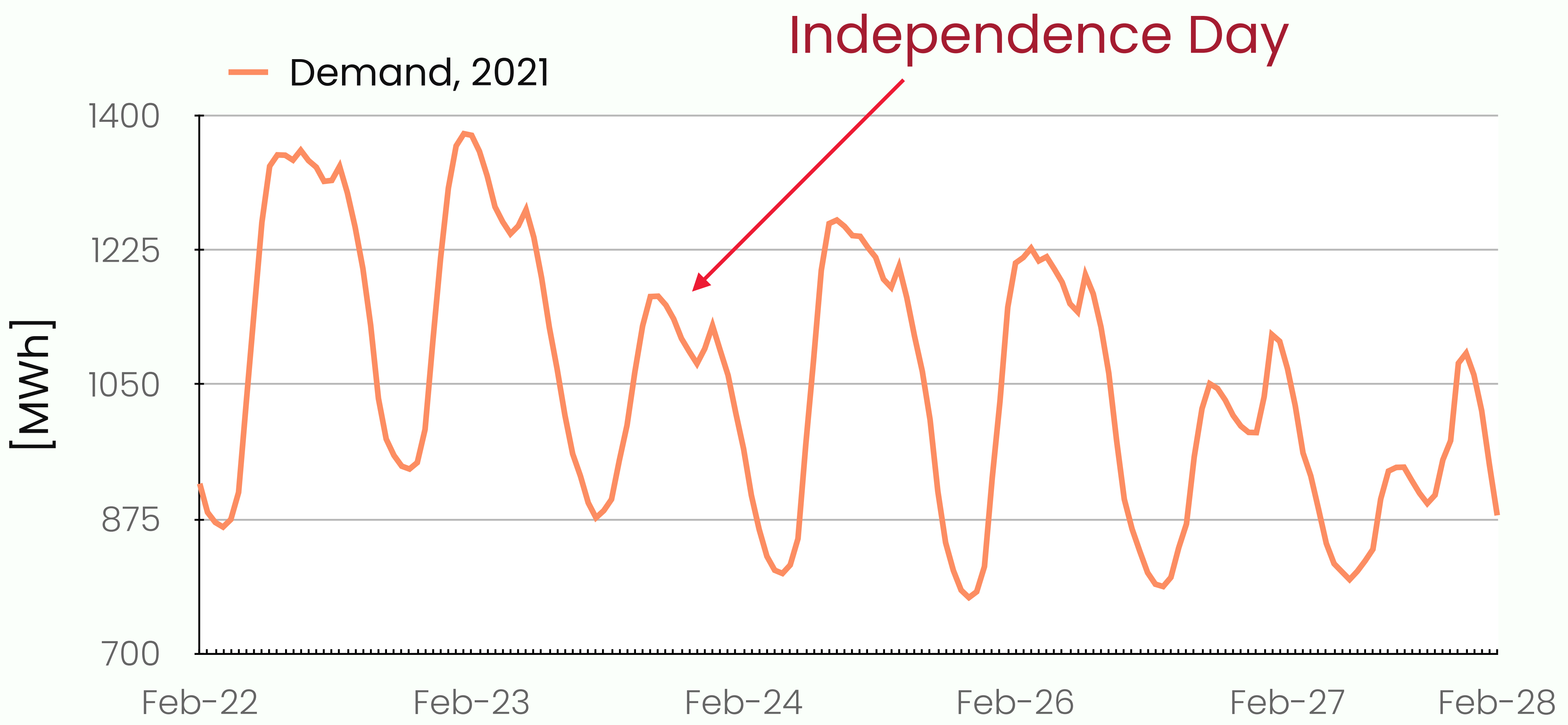
Victory/Midsummer Day
Demand, 2021
1100
950
800
650
500
[MWh]
Jun-21
Jun-22
Jun-23
Jun-25
Jun-26
Jun-27

HOLIDAYS

December 19-25, 2022
December 26-January 1, 22/23
1400
1225
1050
875
700
[MWh]
1
36
71
106
141
168
Christmas

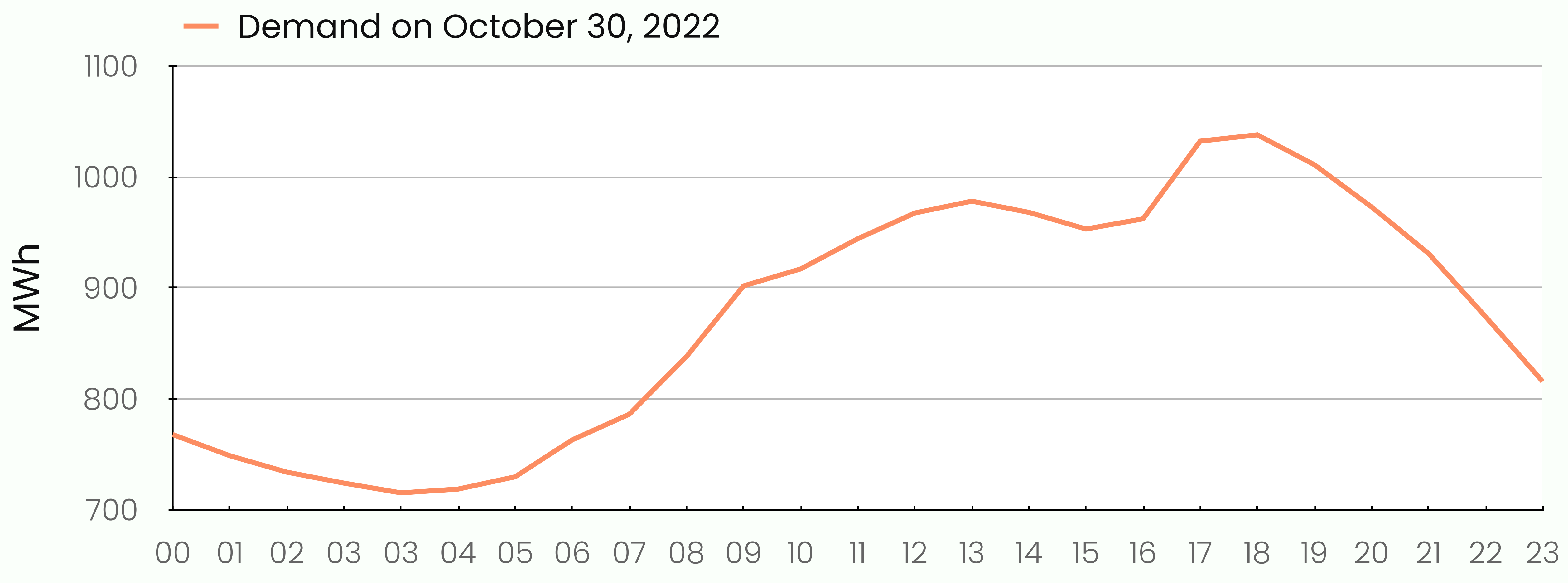


Victory/Midsummer Day
Demand, 2021
1100
950
800
650
500
[MWh]
Jun-21
Jun-22
Jun-23
Jun-25
Jun-26
Jun-27

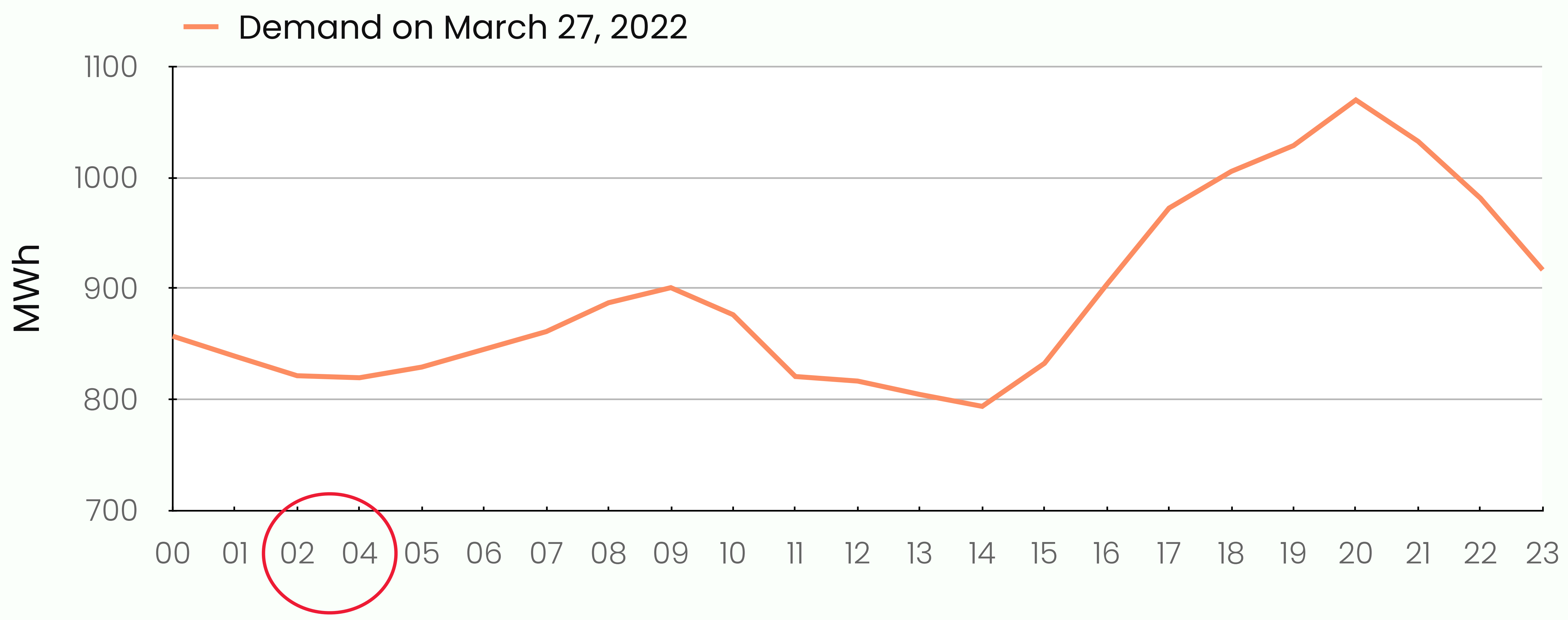


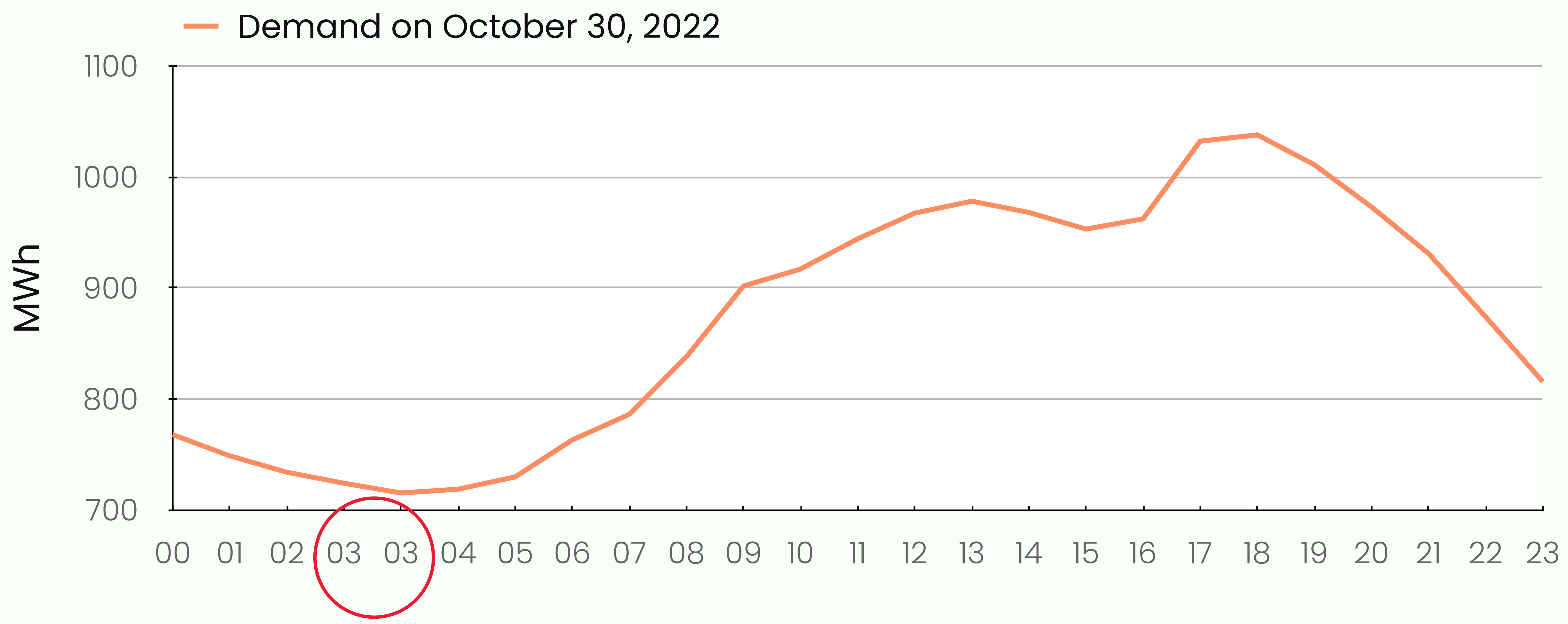
Independence Day
Demand, 2021
1400
1225
1050
875
700
[MWh]
Feb-22
Feb-23
Feb-24
Feb-26
Feb-27
Feb-28

DAYLIGHT SAVING TIME

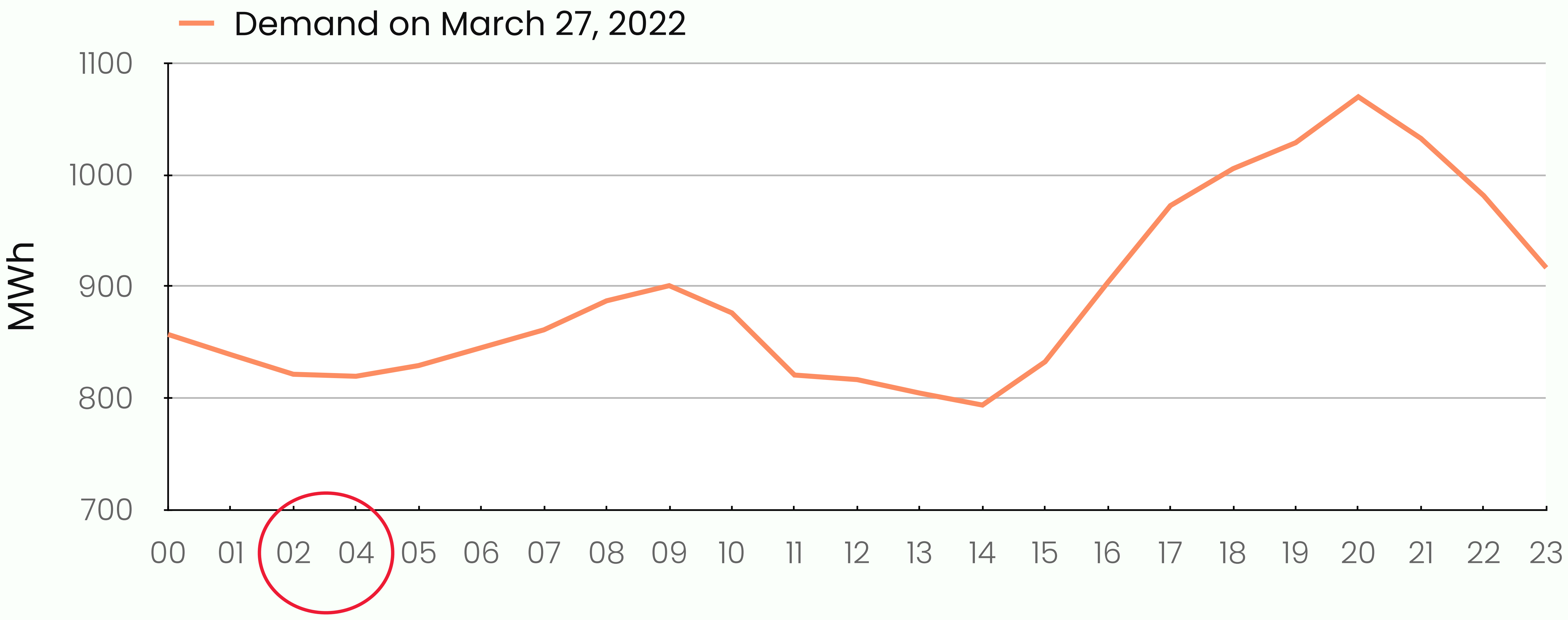


DAYLIGHT SAVING TIME



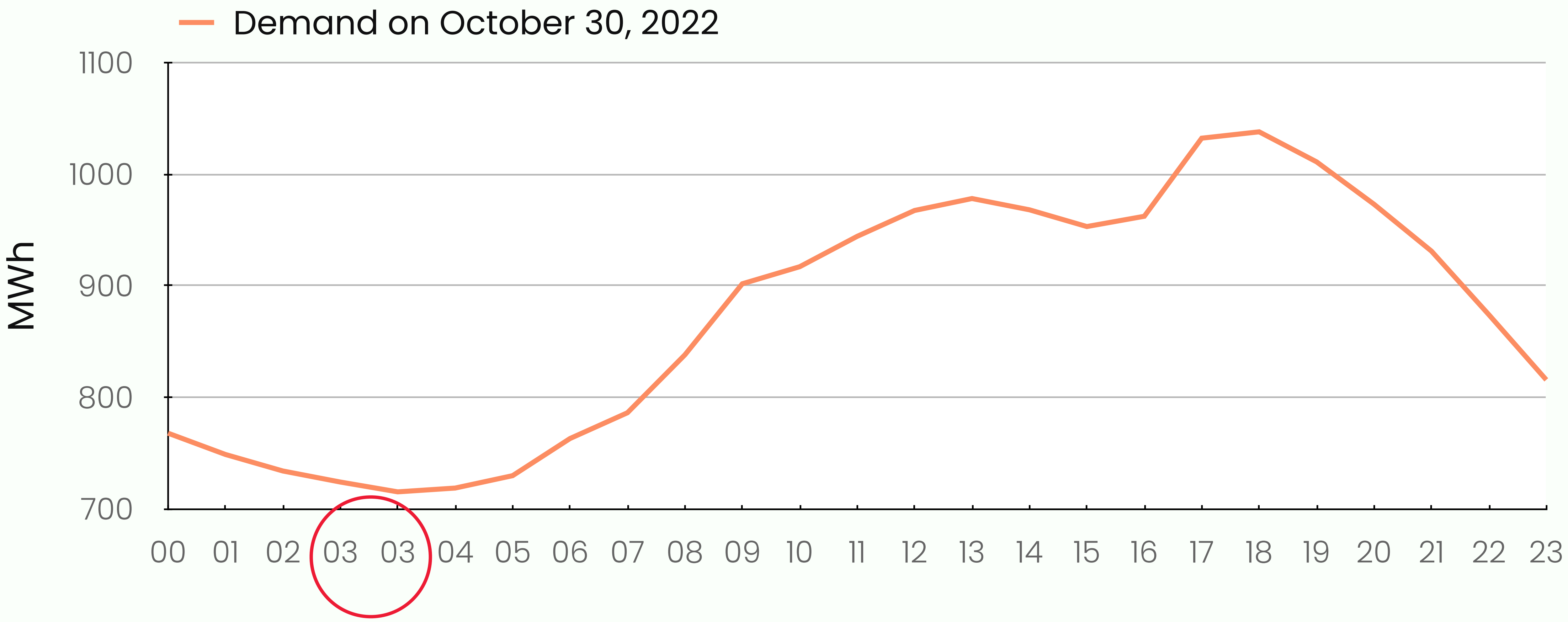


DAYLIGHT SAVING TIME



Sample of electricity production/demand data

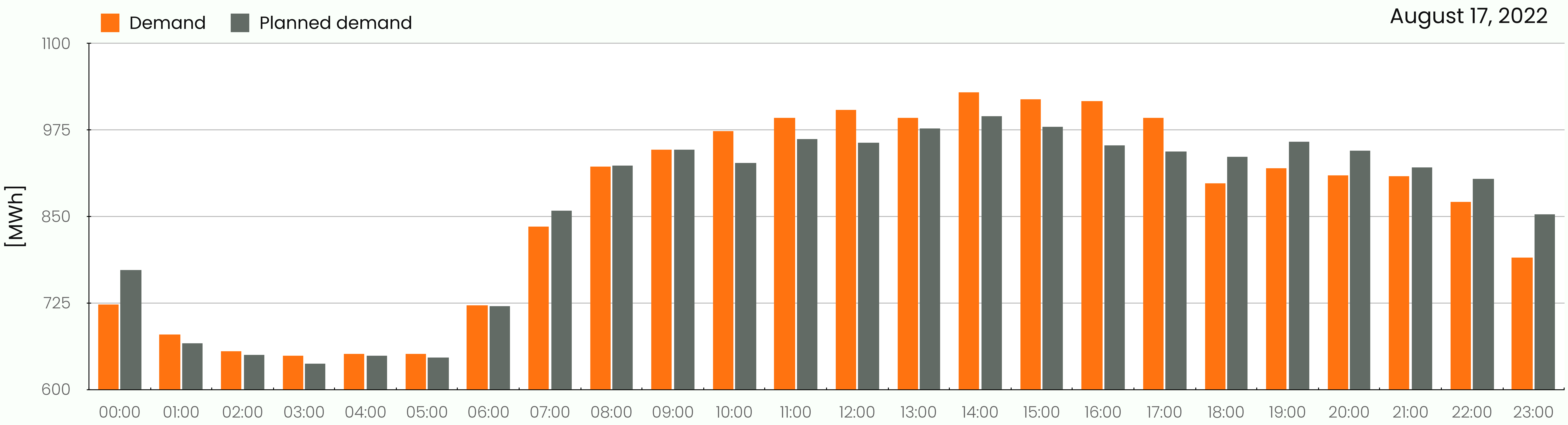
Timestamp (UTC)	Date (Estonia time)	Consumption	Production	Planned consumption	Planned production
1648332000	27.03.2022 00:00	857	554,1	903,8	570,3
1648335600	27.03.2022 01:00	838,9	574,4	913,5	569,5
1648339200	27.03.2022 02:00	821,3	528,9	882,4	537,5
1648342800	27.03.2022 04:00	819,5	547	943,7	593,2
1648346400	27.03.2022 05:00	829,1	512,3	881,2	536
1648350000	27.03.2022 06:00	845,1	520,3	915,4	530



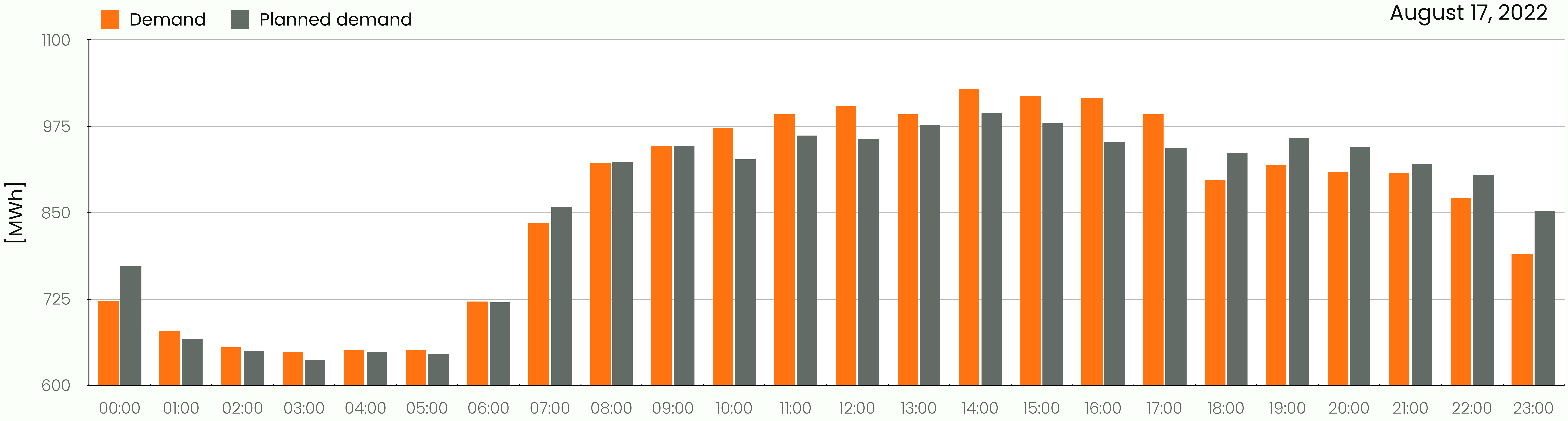
Sample of electricity production/demand data

Timestamp (UTC)	Date (Estonia time)	Consumption	Production	Planned consumption	Planned production
1667077200	30.10.2022 00:00	768	641,2	747,8	638,2
1667080800	30.10.2022 01:00	748,9	640,2	732,4	635,2
1667084400	30.10.2022 02:00	734,1	631	720,6	623,5
1667088000	30.10.2022 03:00	724,2	625,3	708,8	626,3
1667091600	30.10.2022 03:00	715,4	620,9	695,4	614,8
1667095200	30.10.2022 04:00	718,8	624,4	673,2	619,2

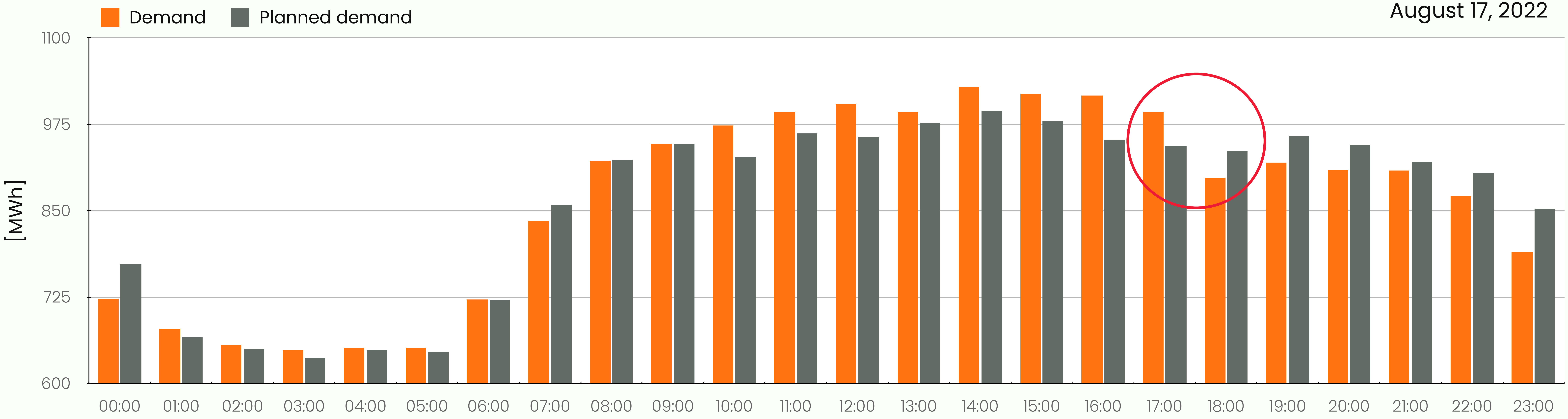
ELECTRICITY PRICES AND MASS MEDIA



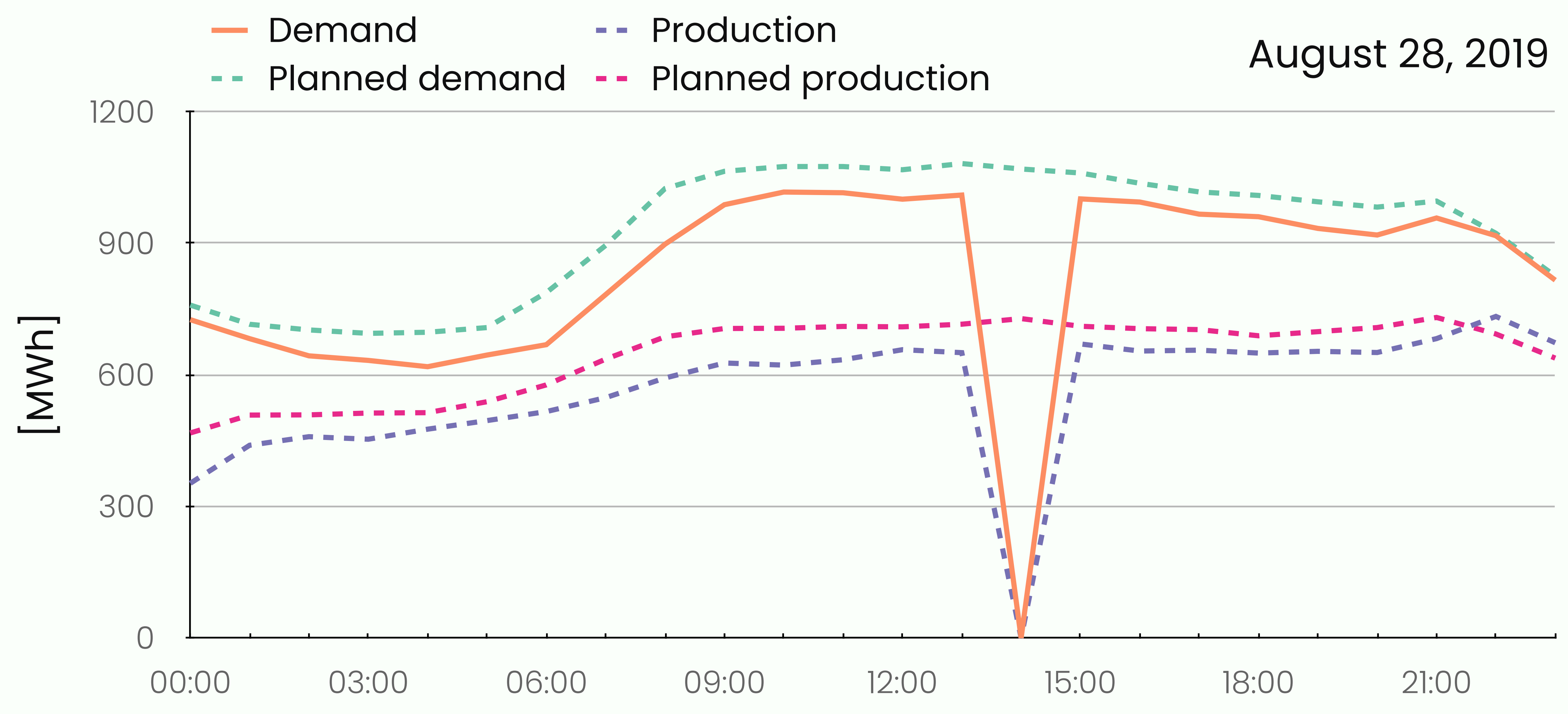
ELECTRICITY PRICES AND MASS MEDIA

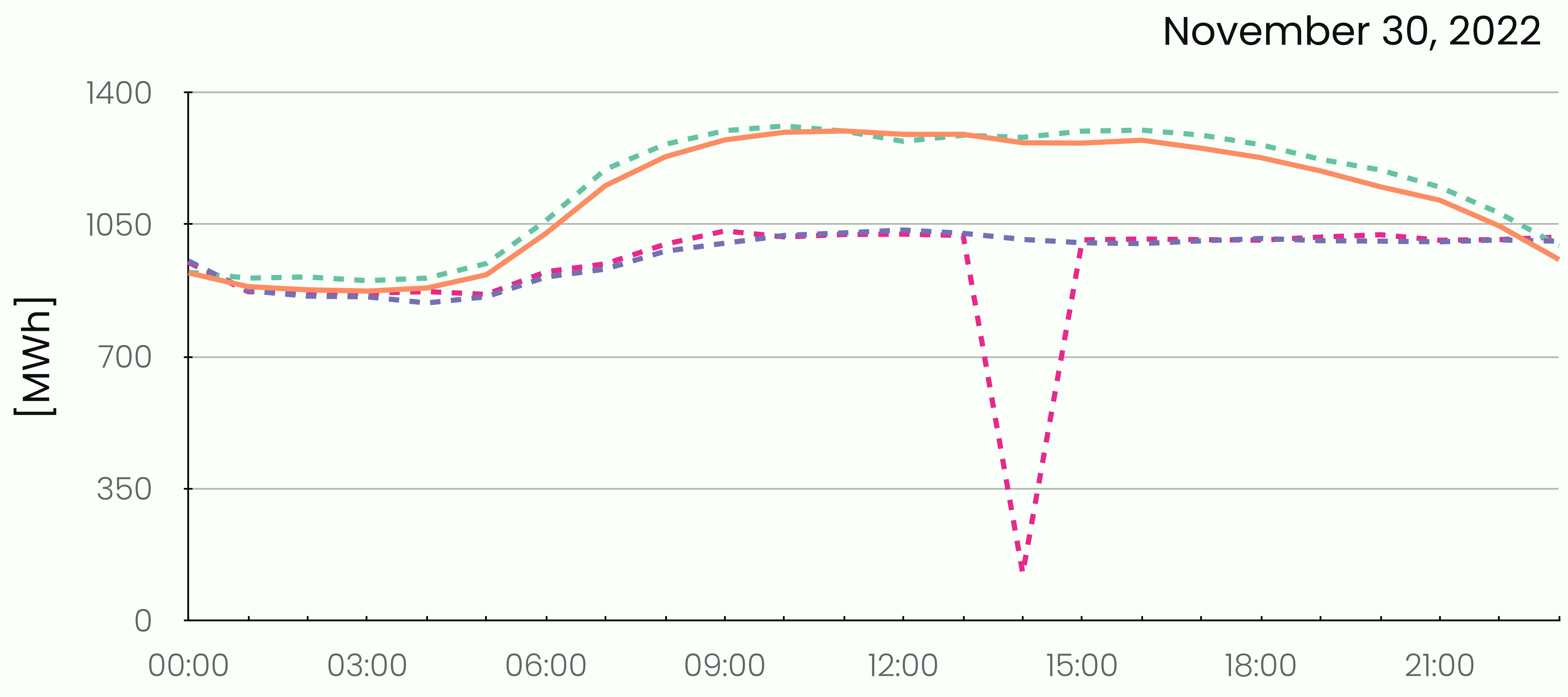


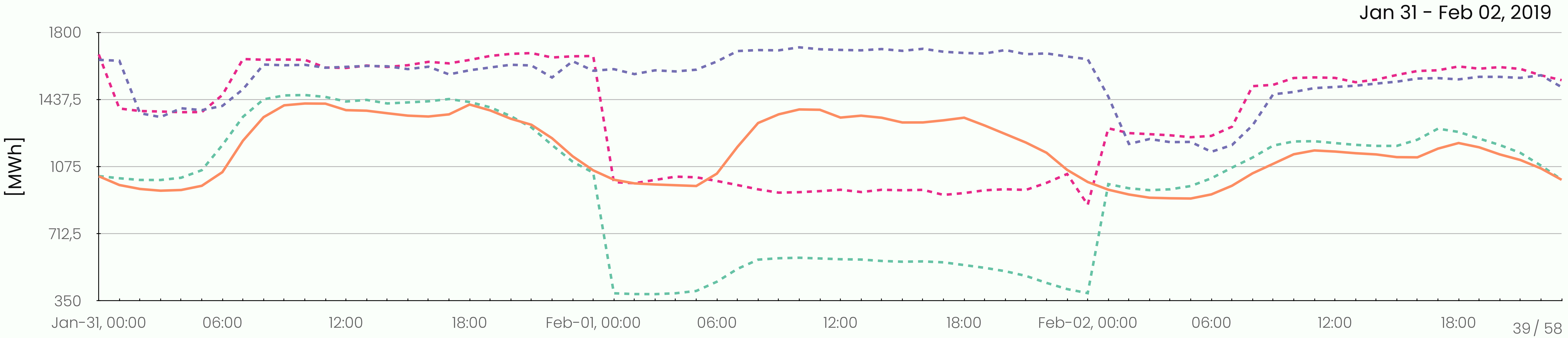
ELECTRICITY PRICES AND MASS MEDIA

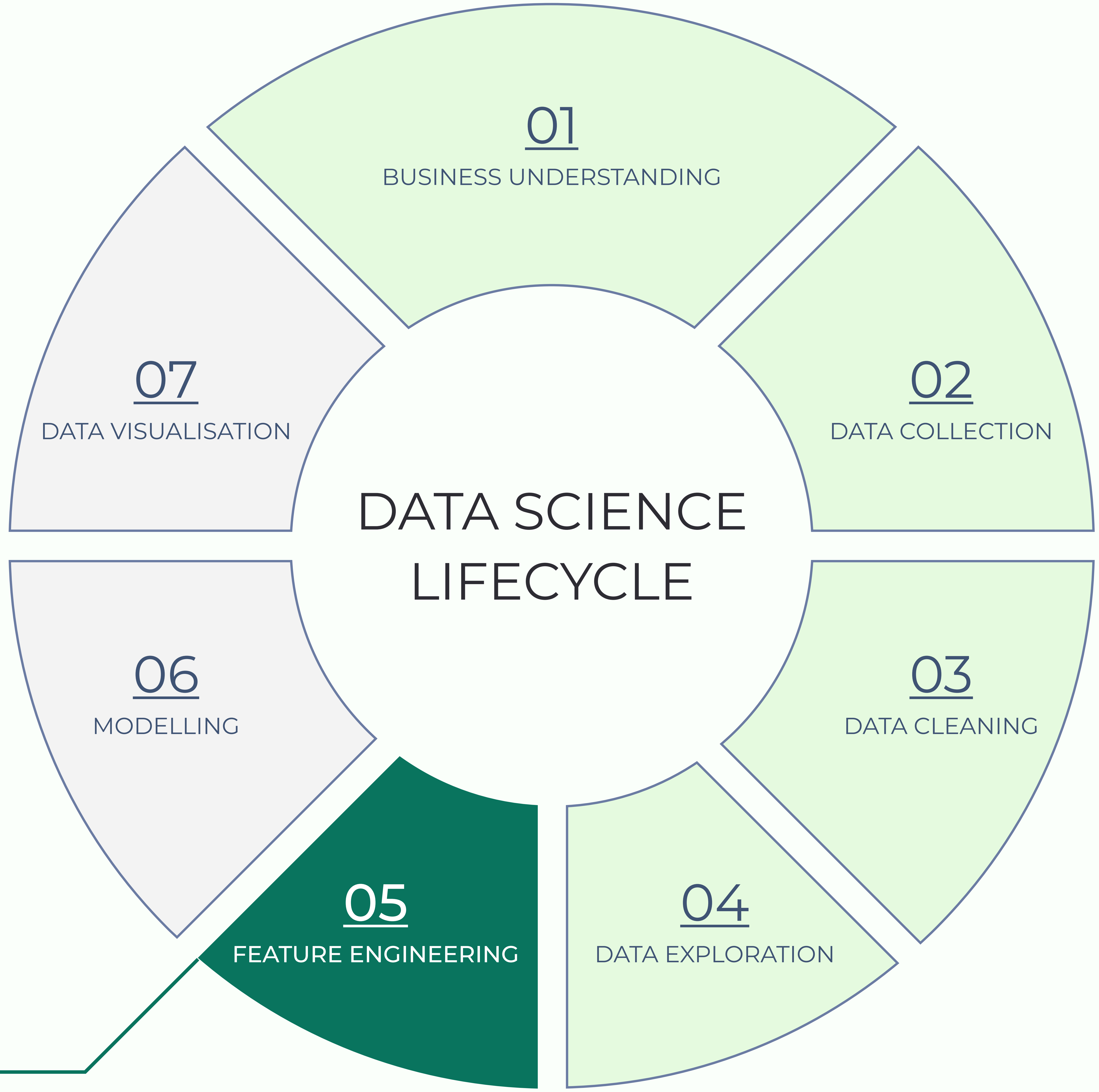


SUSPICIOUS DATA/OUTLIERS









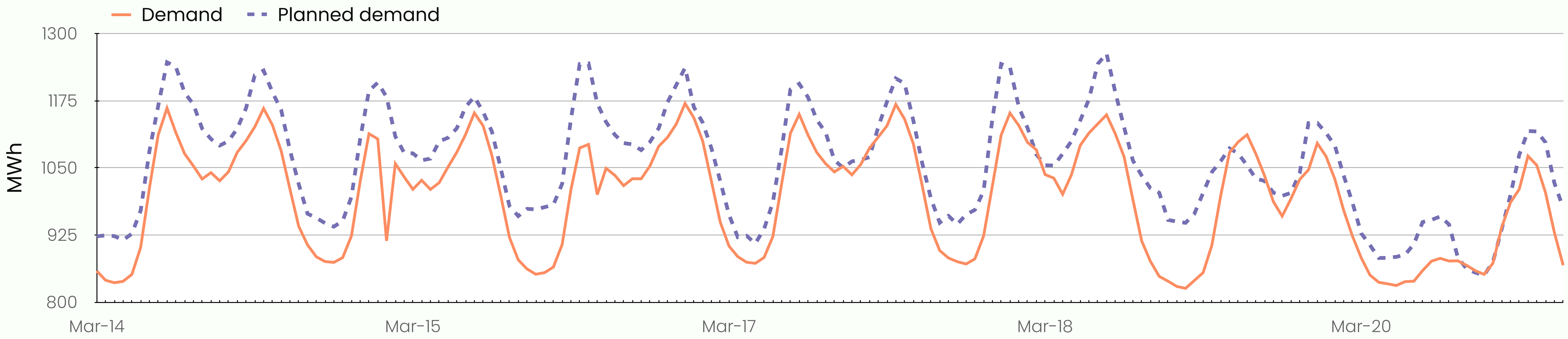
Select important features and construct more meaningful ones.

SUMMARY OF FEATURE ENGINEERING

Feature engineering involves the following main steps:

1. Data exploration and understanding to identify relevant features and potential issues.
2. Data preprocessing, including handling missing values, scaling, normalisation, and encoding categorical features.
3. Feature selection to identify the most relevant features for the model.
4. Feature transformation to create new features or derive more meaningful features from existing ones.

WHAT FEATURES CAN BE RELEVANT?



A shot list on possible features:

- Previous electricity consumption: a day, a week ago
- Weather: temperature, irradiation, wind direction/speed, humidity
- Day length: sunset/sunrise
- Weekday
- Hour of the day
- Electricity price

GENERATE NEW FEATURES

Temperature:

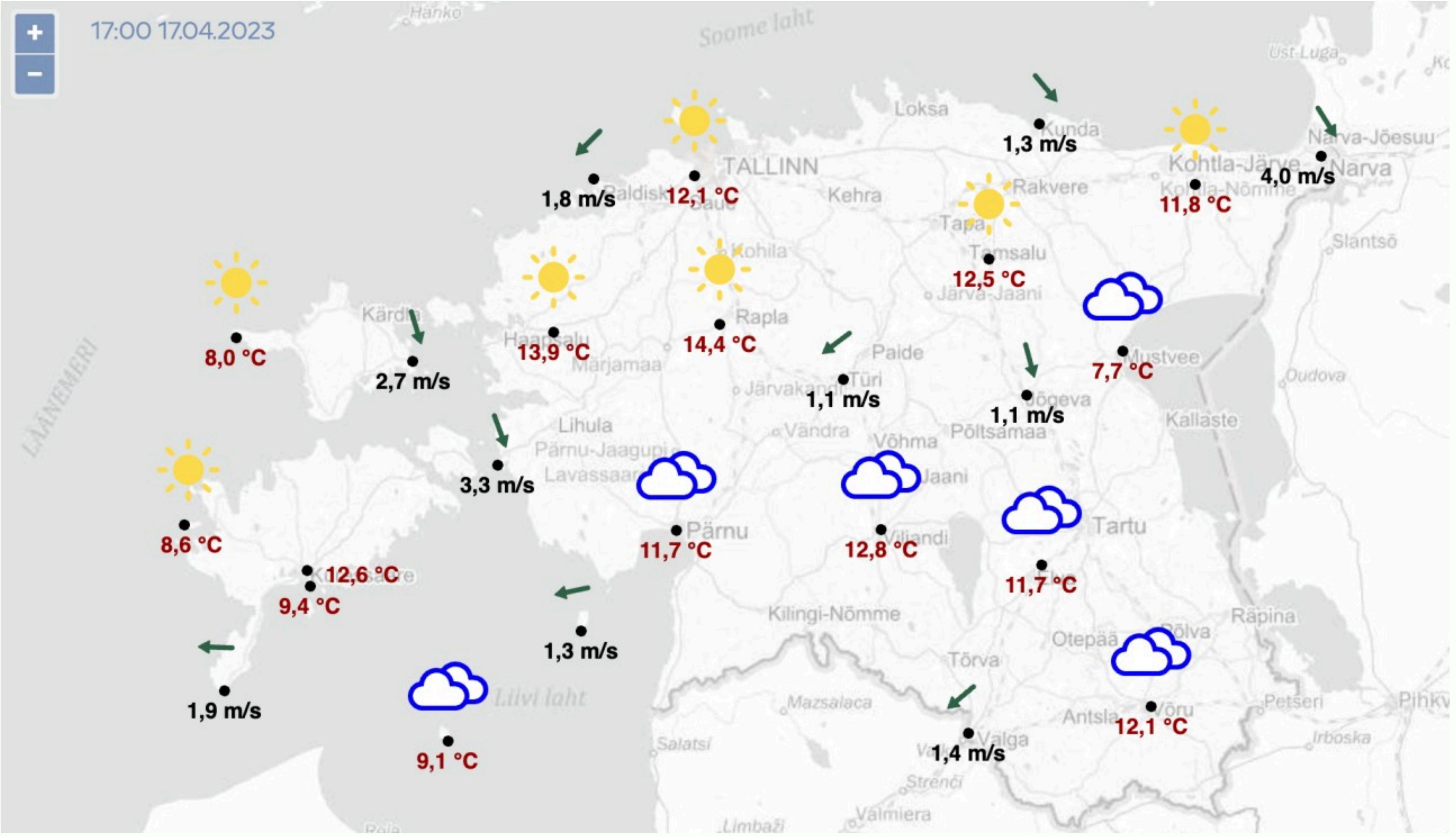
- average
- average + difference
- weighted average (population, correlation)
- heating/cooling degree-day

Wind:

- windchill (temperature + wind)
- FeelsLike (RealFeel):

  temperature + wind + humidity + pressure

Clouds + Sun

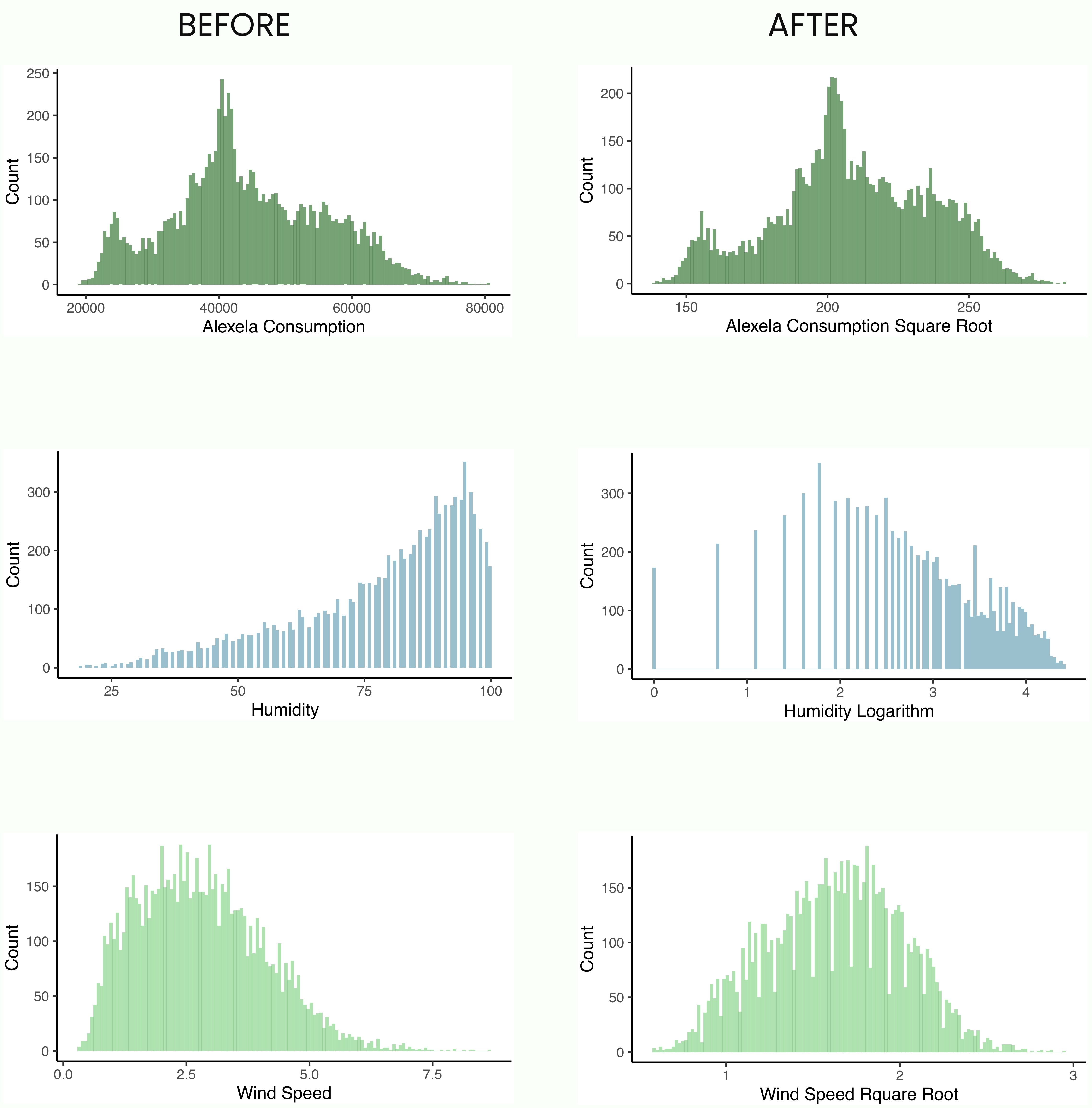


DATA TRANSFORMATION

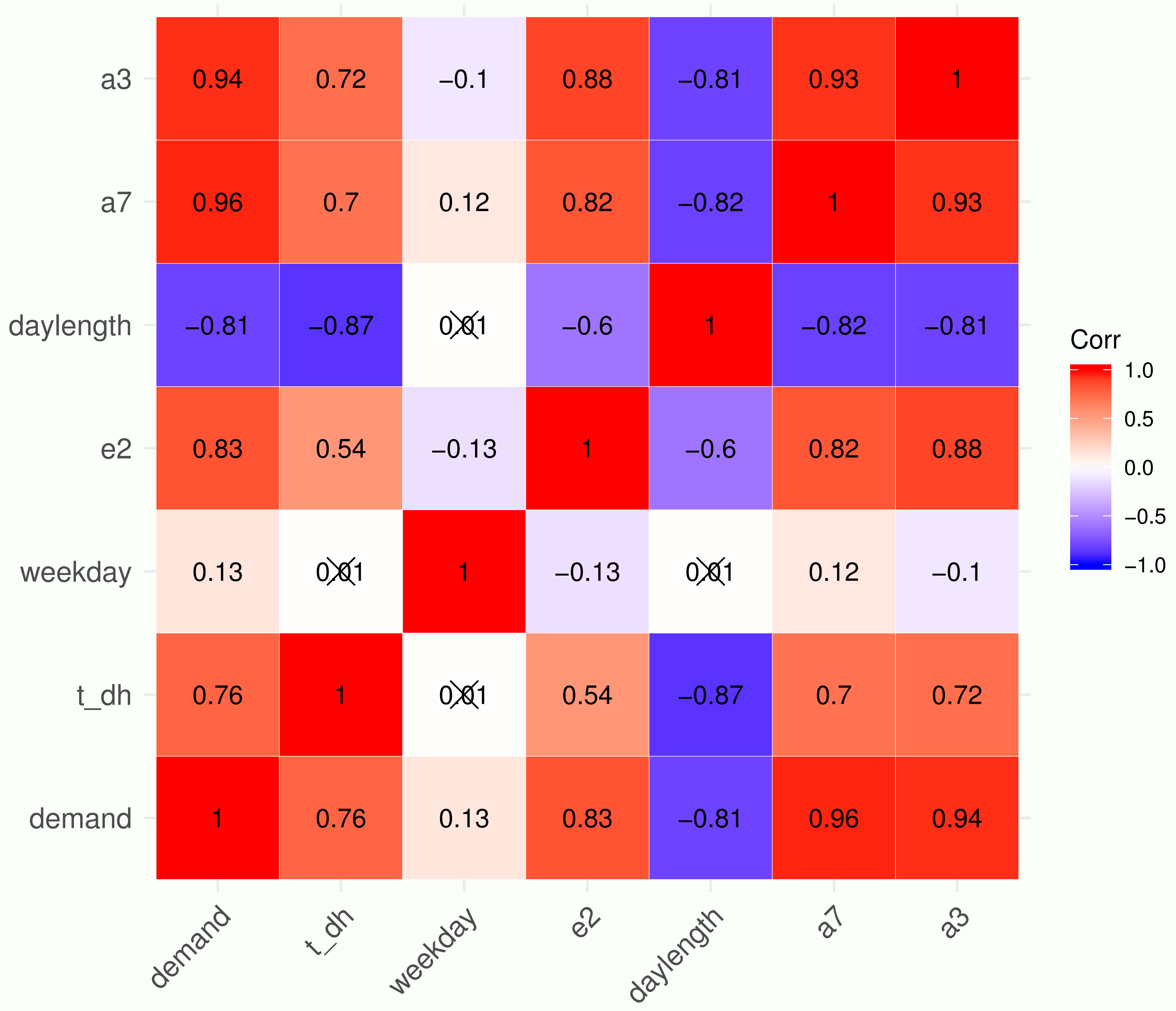
Data transformation is the process of converting raw data into a format or structure more appropriate for model building and data exploration.

Popular methods include, but not limited to:

- Log transform
- Square transform
- Square root transform
- Box Cox transform
- Scaling: min-max, standardisation, robust scaler, decimal

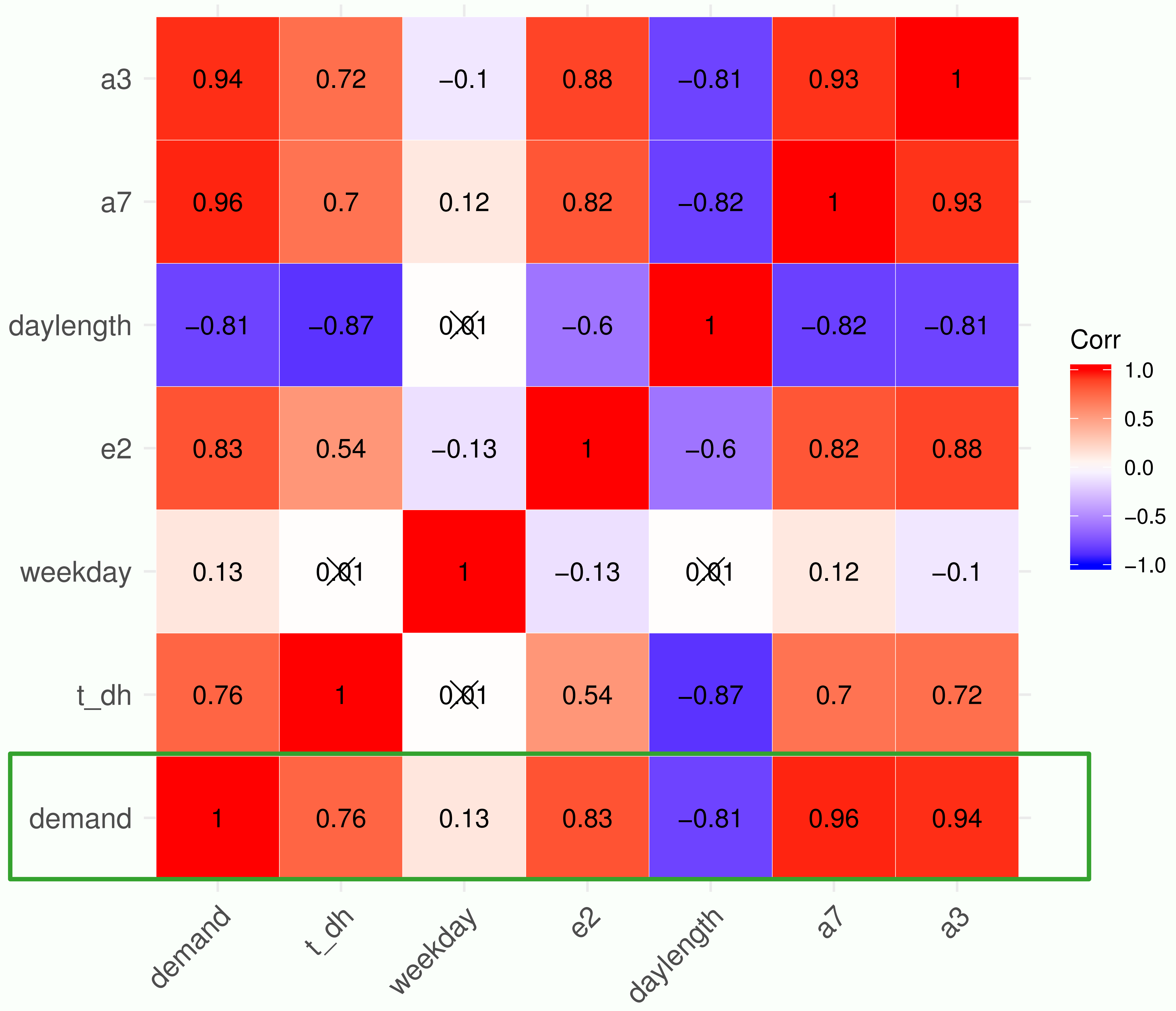


CHECK WHICH FEATURES ARE IMPORTANT



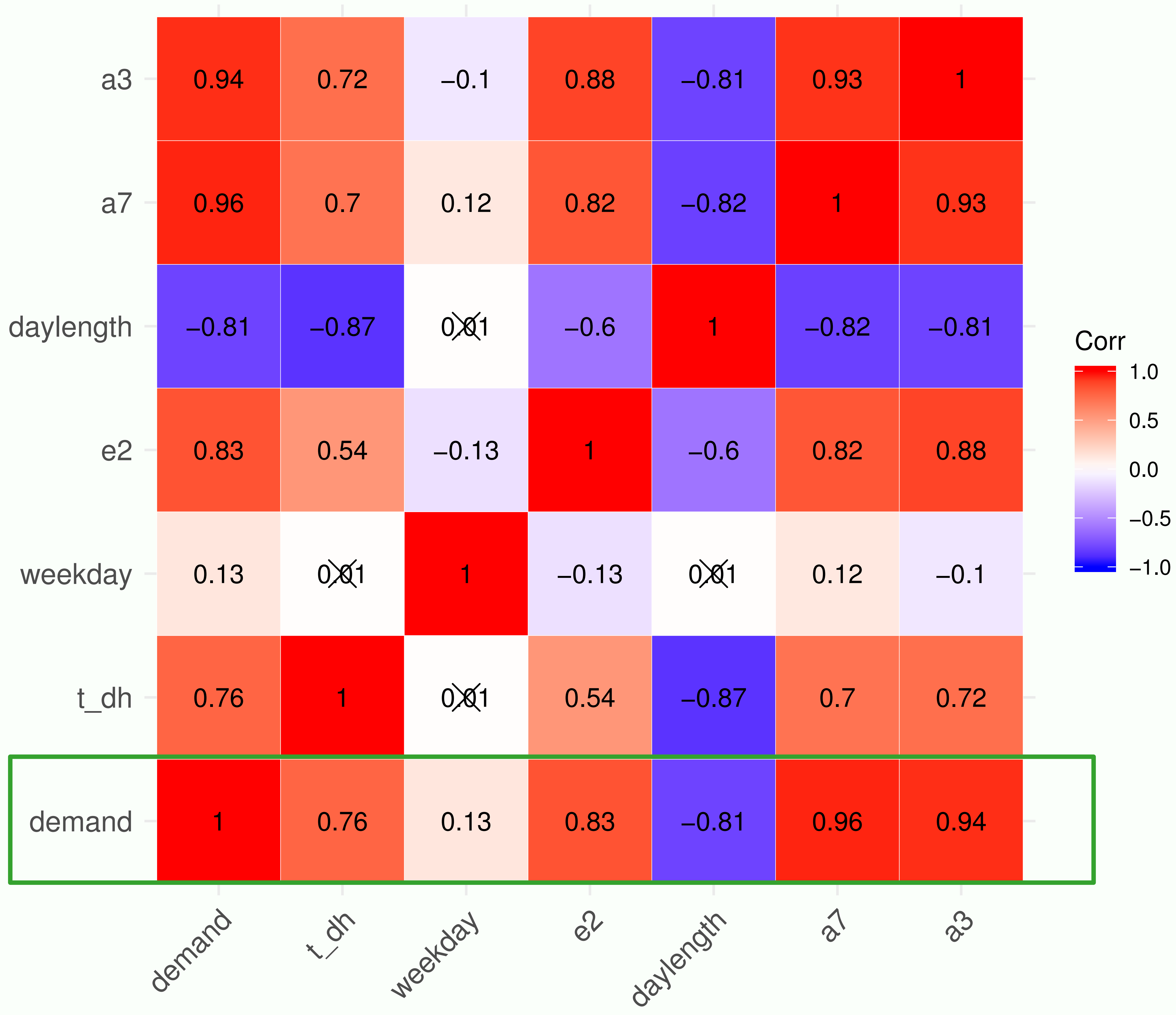
M. Spichakova, J. Belikov, K. Nõu, E. Petlenkov. Feature engineering for short-term forecast of energy consumption. *IEEE PES Innovative Smart Grid Technologies Europe: Bucharest, Romania*, 2019.
M. Sinimaa, M. Spichakova, J. Belikov, E. Petlenkov. Feature engineering of weather data for short-term energy consumption forecast. *IEEE Madrid PowerTech: Madrid, Spain*, 2021.

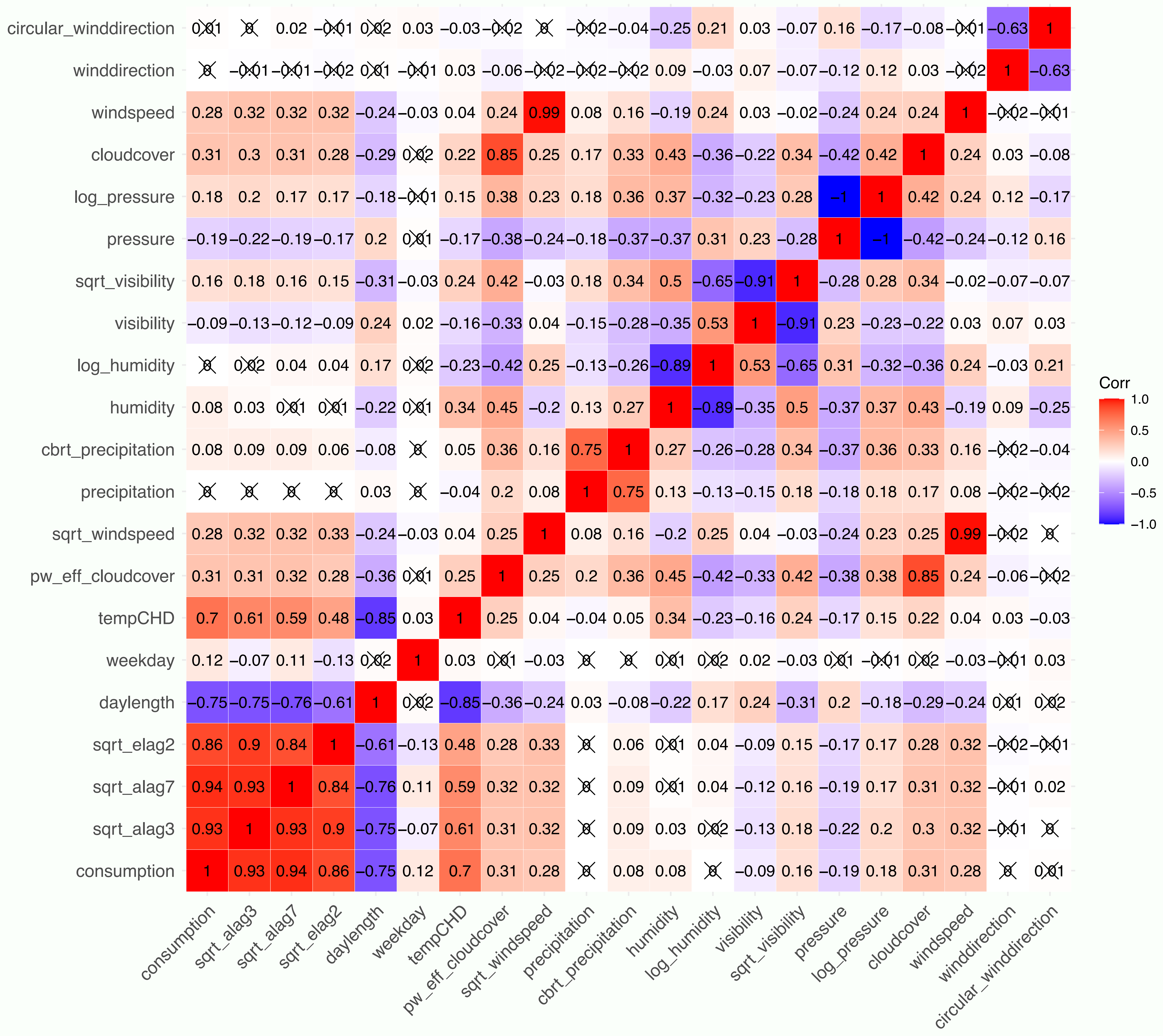
CHECK WHICH FEATURES ARE IMPORTANT



M. Spichakova, J. Belikov, K. Nõu, E. Petlenkov. Feature engineering for short-term forecast of energy consumption. *IEEE PES Innovative Smart Grid Technologies Europe: Bucharest, Romania*, 2019.
M. Sinimaa, M. Spichakova, J. Belikov, E. Petlenkov. Feature engineering of weather data for short-term energy consumption forecast. *IEEE Madrid PowerTech: Madrid, Spain*, 2021.

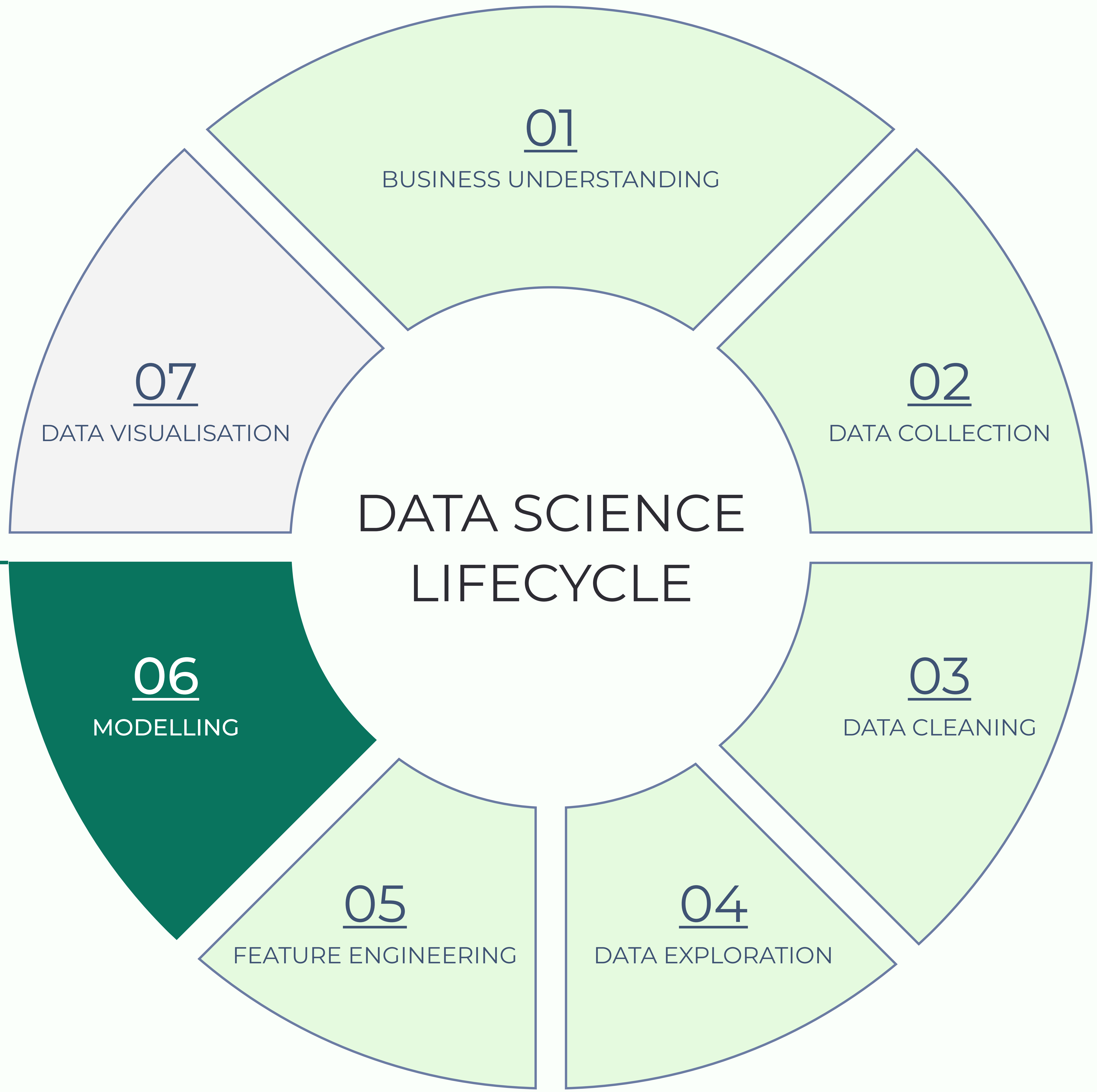
CHECK WHICH FEATURES ARE IMPORTANT



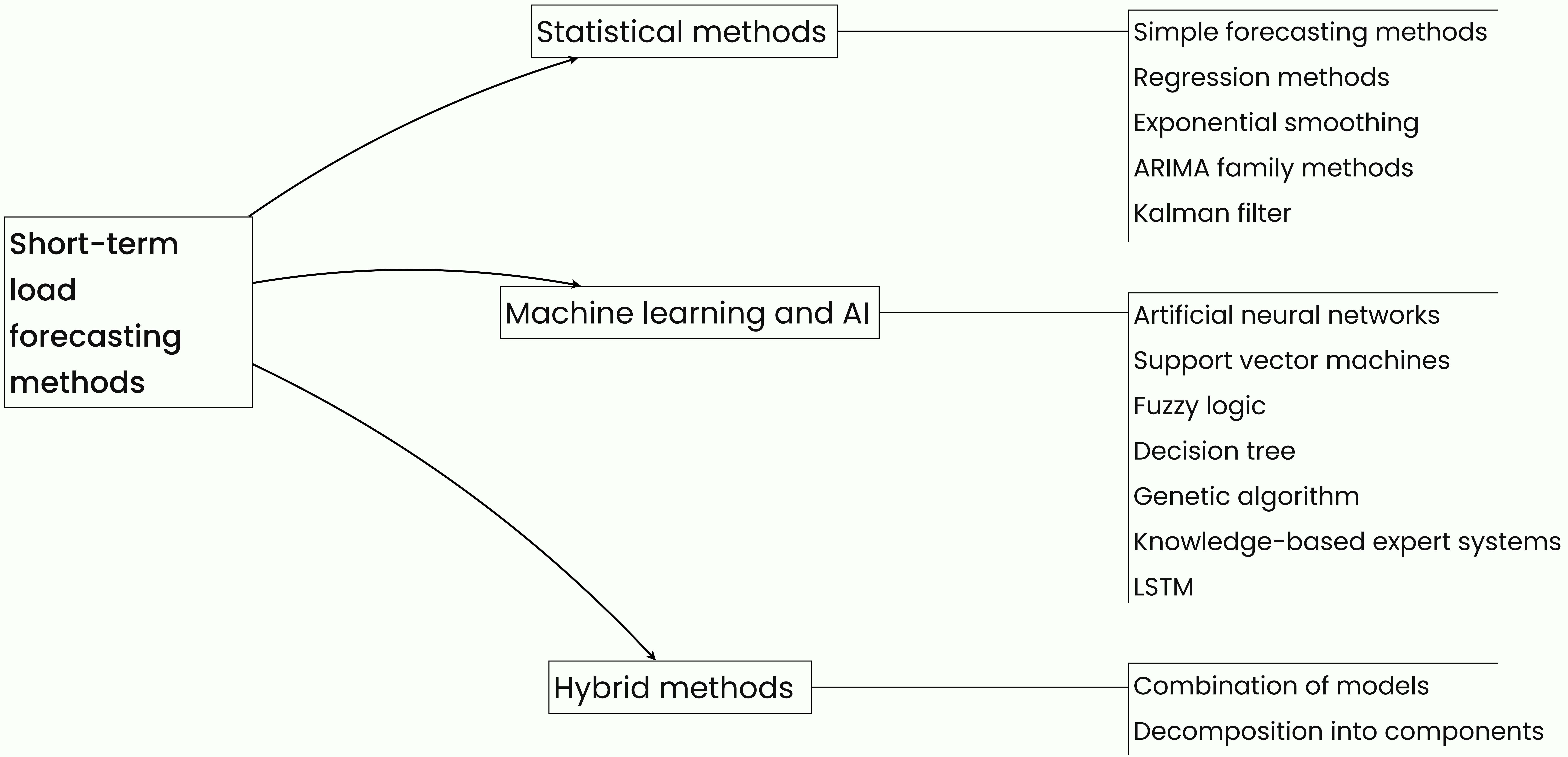


M. Spichakova, J. Belikov, K. Nõu, E. Petlenkov. Feature engineering for short-term forecast of energy consumption. *IEEE PES Innovative Smart Grid Technologies Europe: Bucharest, Romania*, 2019.
M. Sinimaa, M. Spichakova, J. Belikov, E. Petlenkov. Feature engineering of weather data for short-term energy consumption forecast. *IEEE Madrid PowerTech: Madrid, Spain*, 2021.

Train ML models, evaluate performance, and make predictions.



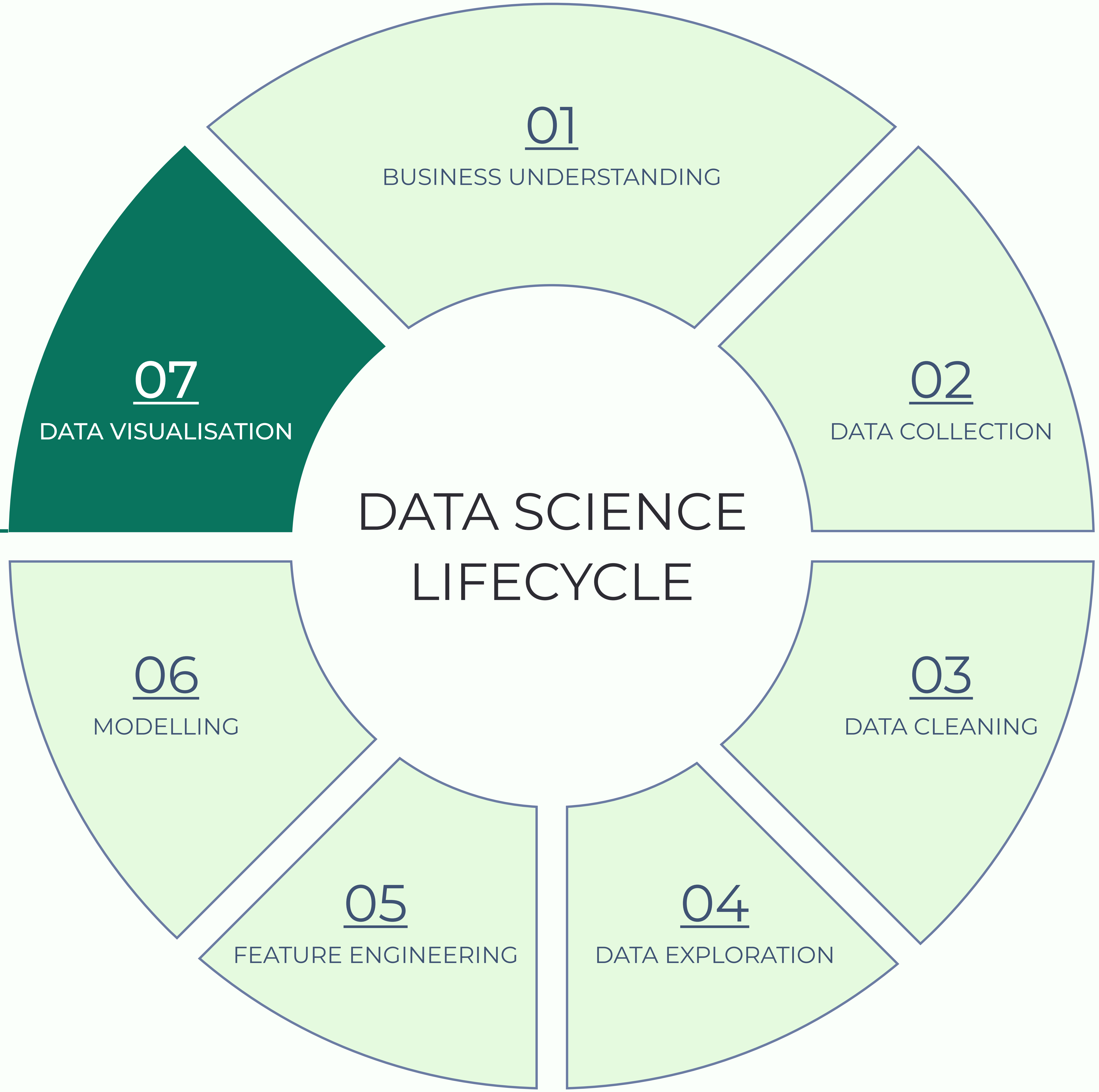
OVERVIEW OF THE MODELLING TECHNIQUES



MODELLING

We will skip this step for now and will return to it in the upcoming lectures.

Communicate the findings with key people using plots and interactive visualisations.



COMMUNICATE YOUR RESULTS

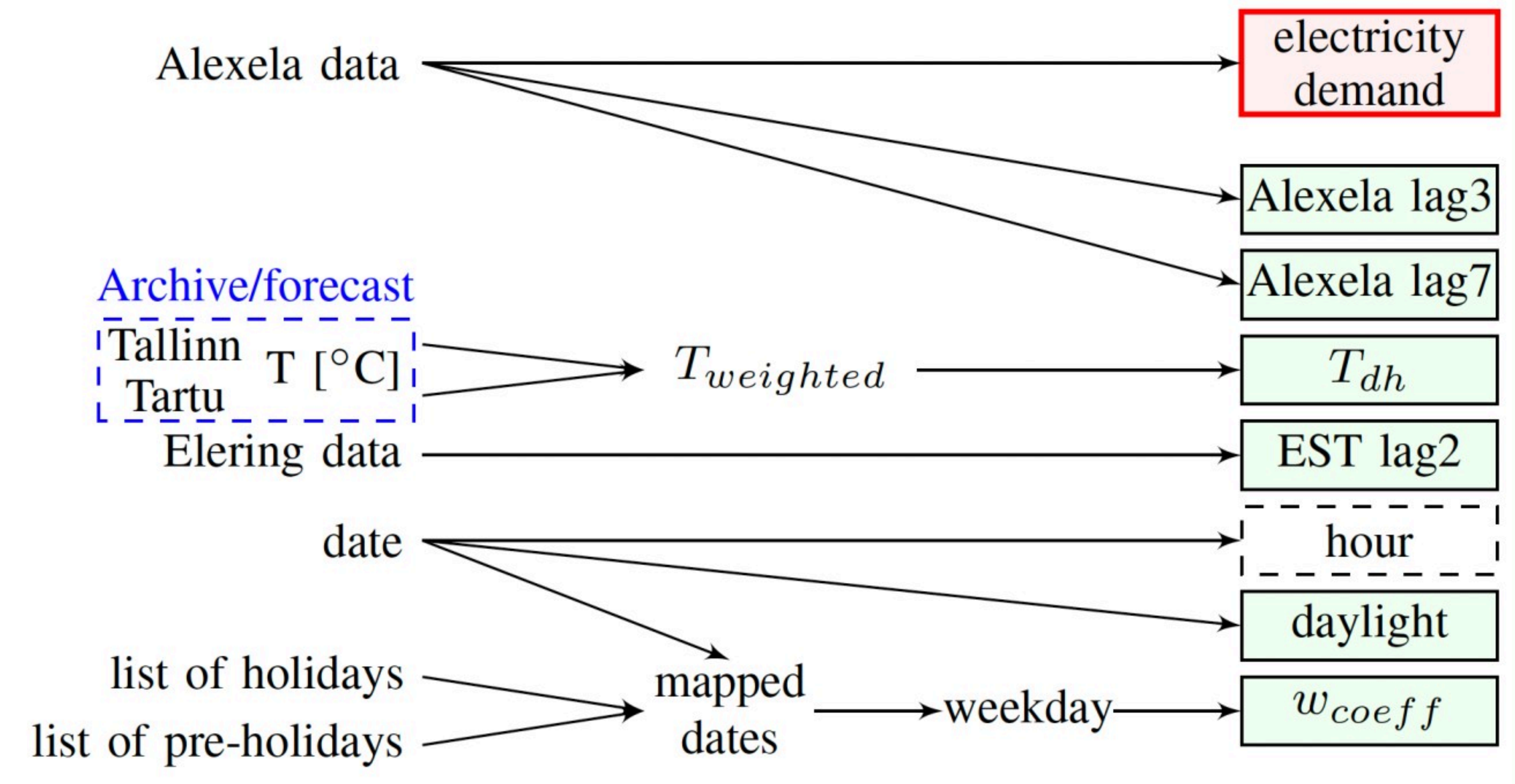


Fig. 5: Relations between different factors revealed by the feature analysis.

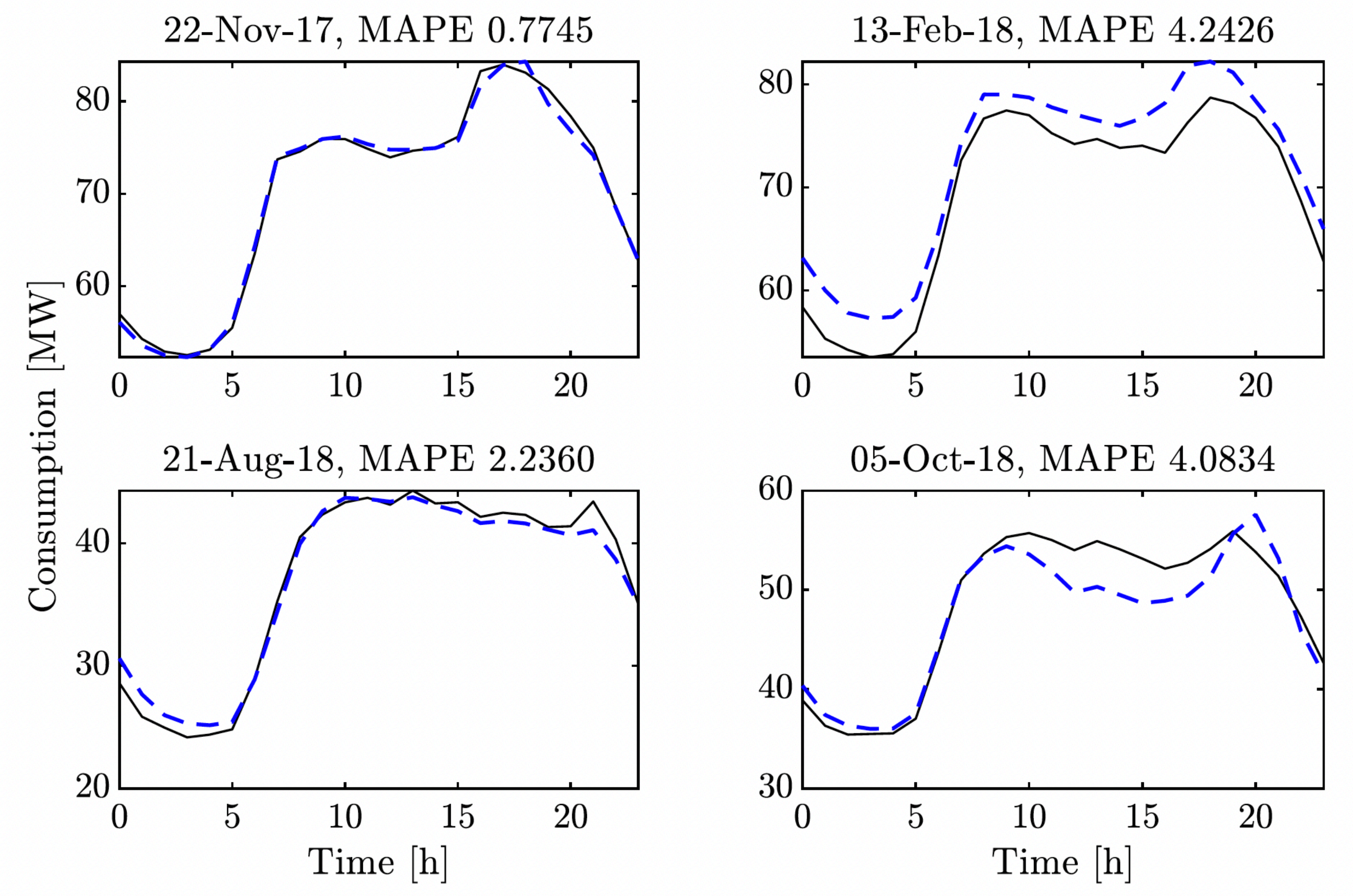


Fig. 8: Real ('—') and predicted ('- - -') electricity consumption for several representative days within the test set.

TABLE II: Validation Results: Monthly MAPE

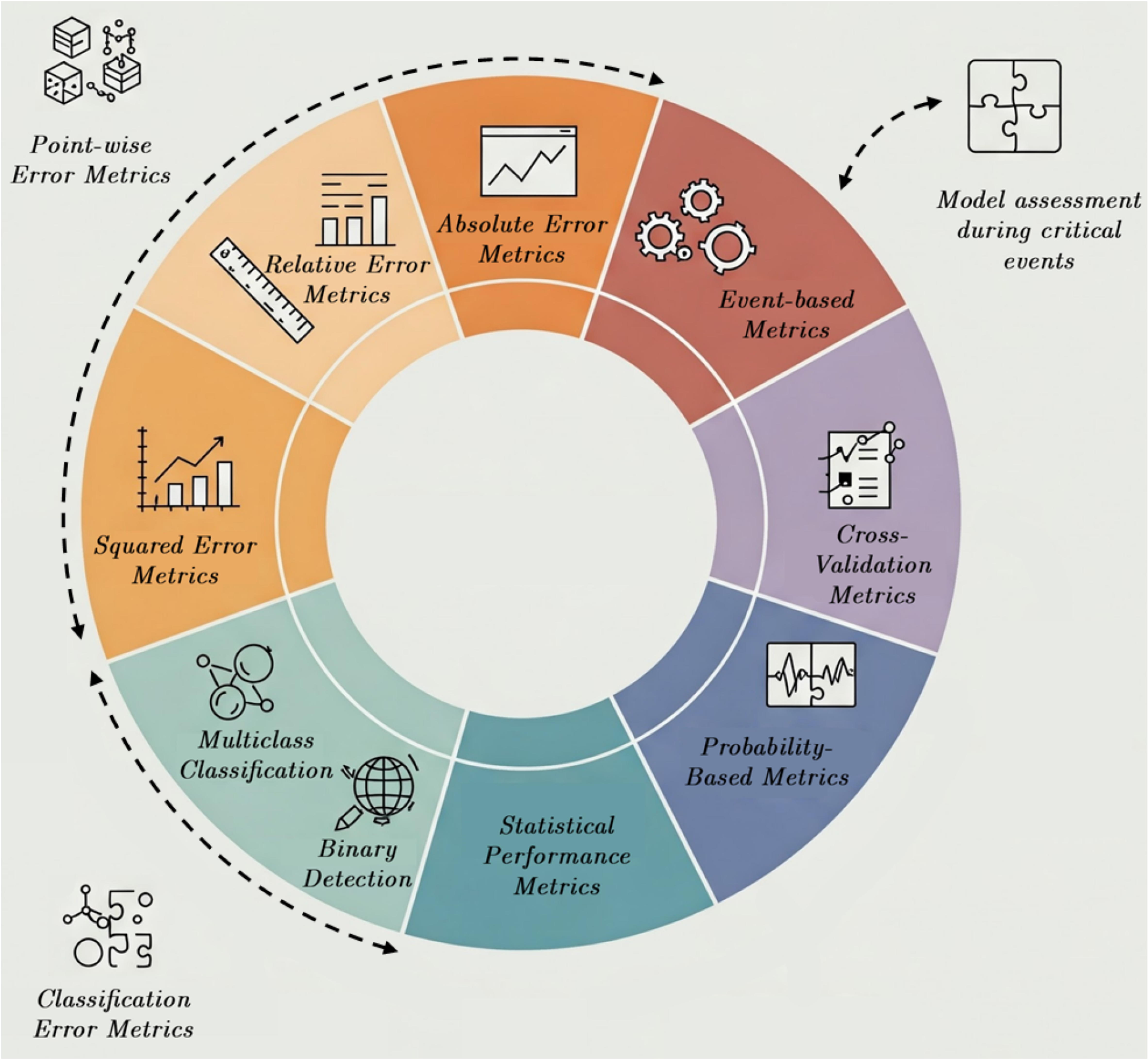
Year/Month		Proposed	Var1	Var2	Var3	Var4
2017	November	1.959	2.589	2.562	2.578	3.791
	December	2.358	3.39	3.104	3.39	4.009
2018	January	2.829	4.561	4.782	4.561	7.183
	February	4.412	4.654	4.444	4.654	5.808
	March	4.217	5.046	5.236	5.046	8.779
	April	2.73	3.823	3.725	3.951	8.563
	May	3.235	4.066	4.065	4.722	4.453
	June	4.001	4.376	4.635	4.474	4.246
	July	3.779	4.282	4.509	4.177	4.317
	August	2.949	3.735	4.081	3.95	3.661
	September	3.968	4.498	4.71	4.497	4.964
	October	3.629	4.172	4.43	4.663	7.744
Total		**3.339**	**4.099**	**4.19**	**4.222**	**5.626**

[a]Var1 has all 6 features, except $w_{tln} = w_{trt} = 0.5$ in T_{dh}.
[b]Var2 has 5 features: $a3$, $a7$, $e2$, T_{dh}, and daylight.
[c]Var3 has all 6 features, but T_{dh} is replaced by $(T_{tln} + T_{trt})/2$.
[d]Var4 has 2 features: $a7$ and T_{dh}.

M. Spichakova, J. Belikov, K. Nõu, E. Petlenkov. Feature engineering for short-term forecast of energy consumption. *IEEE PES Innovative Smart Grid Technologies Europe*: *Bucharest, Romania*, 2019.

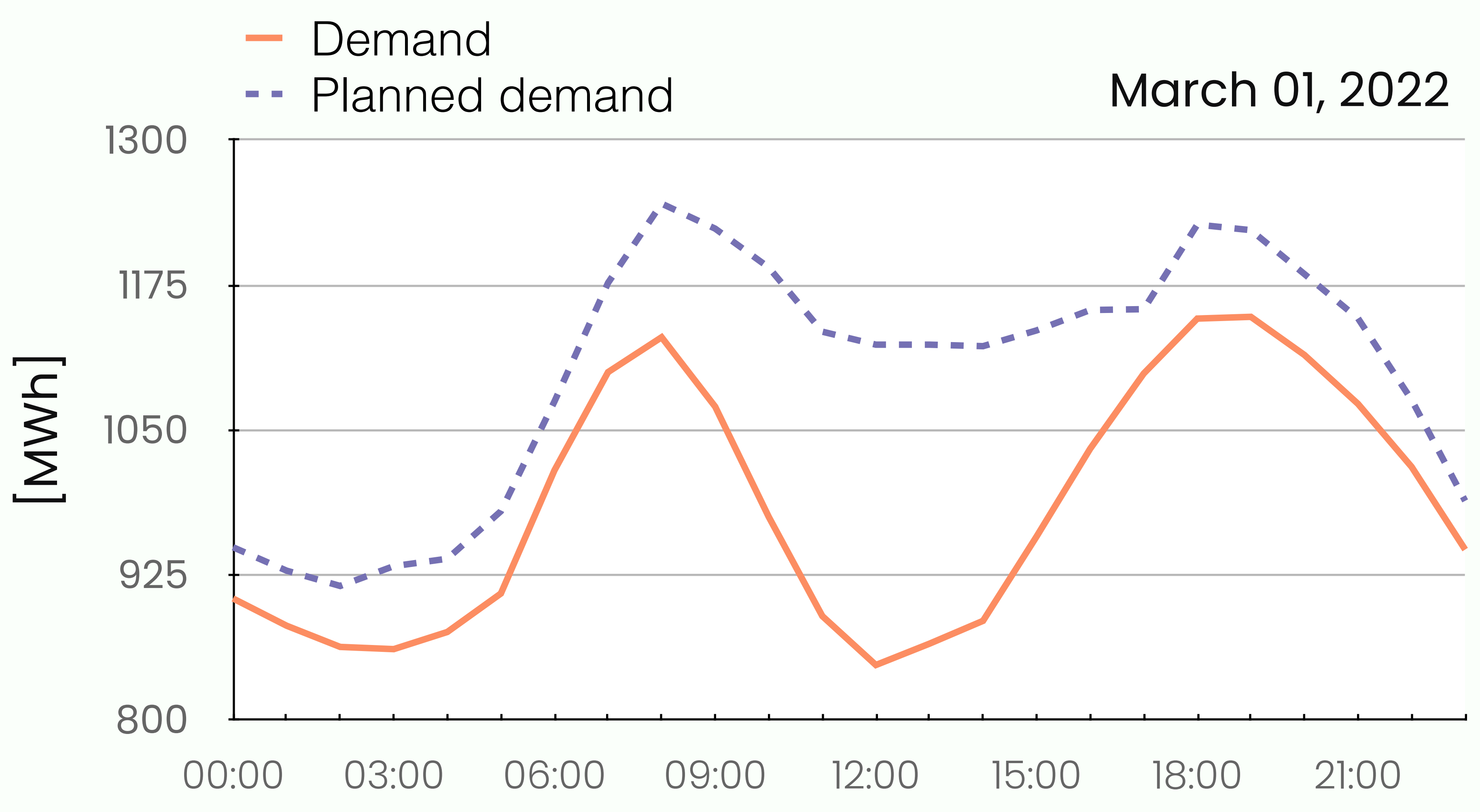
Model Validation

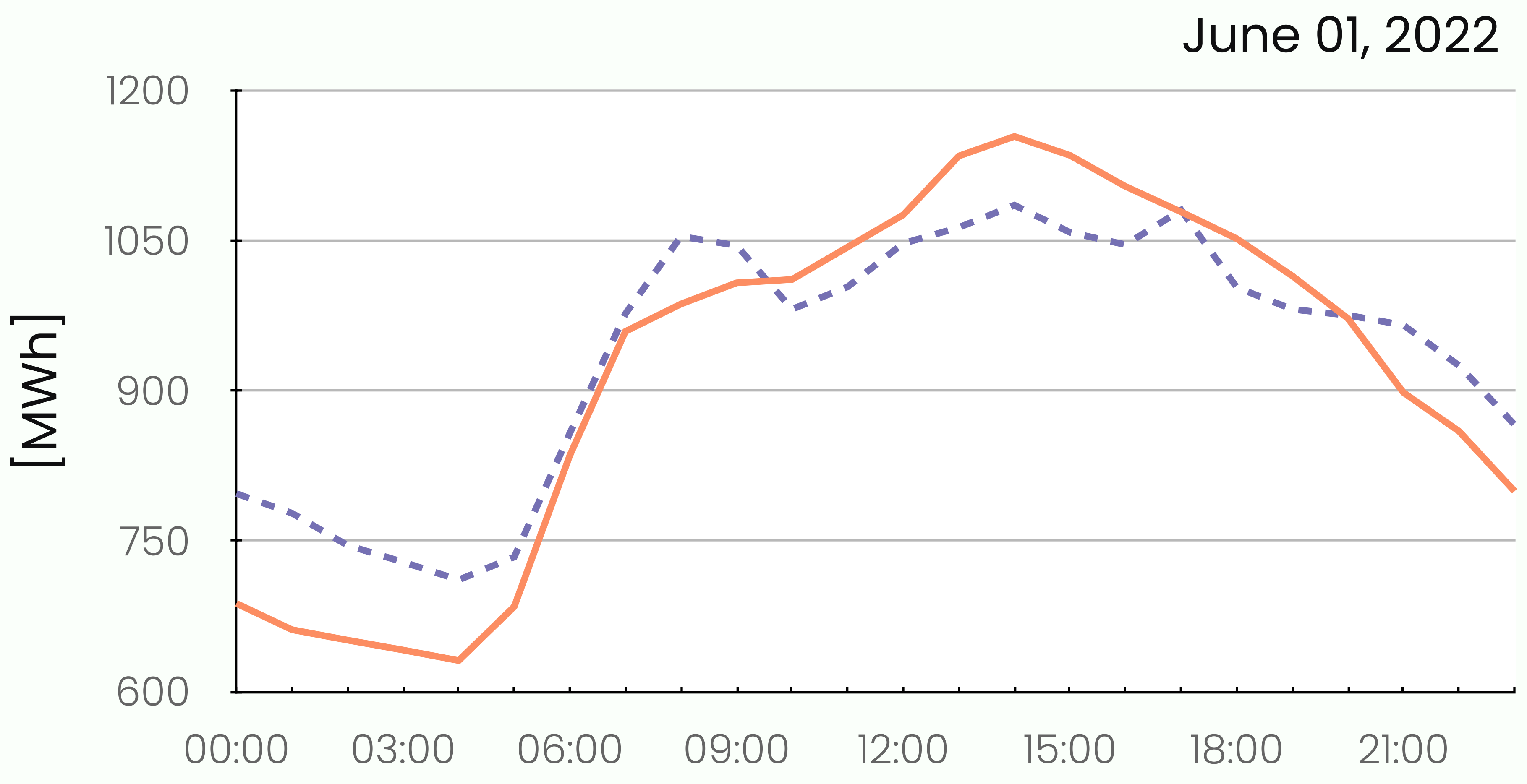
PERFORMANCE METRICS

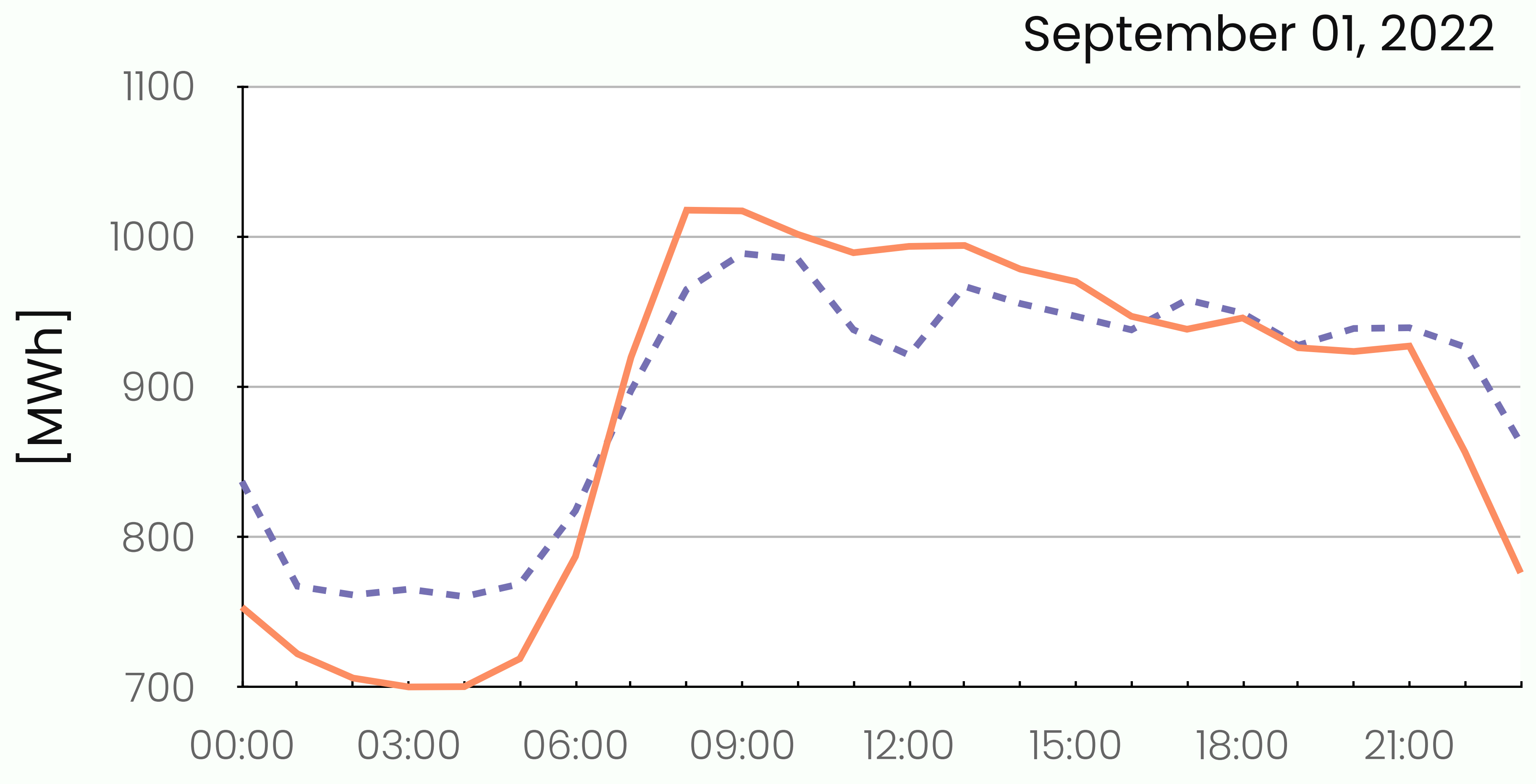


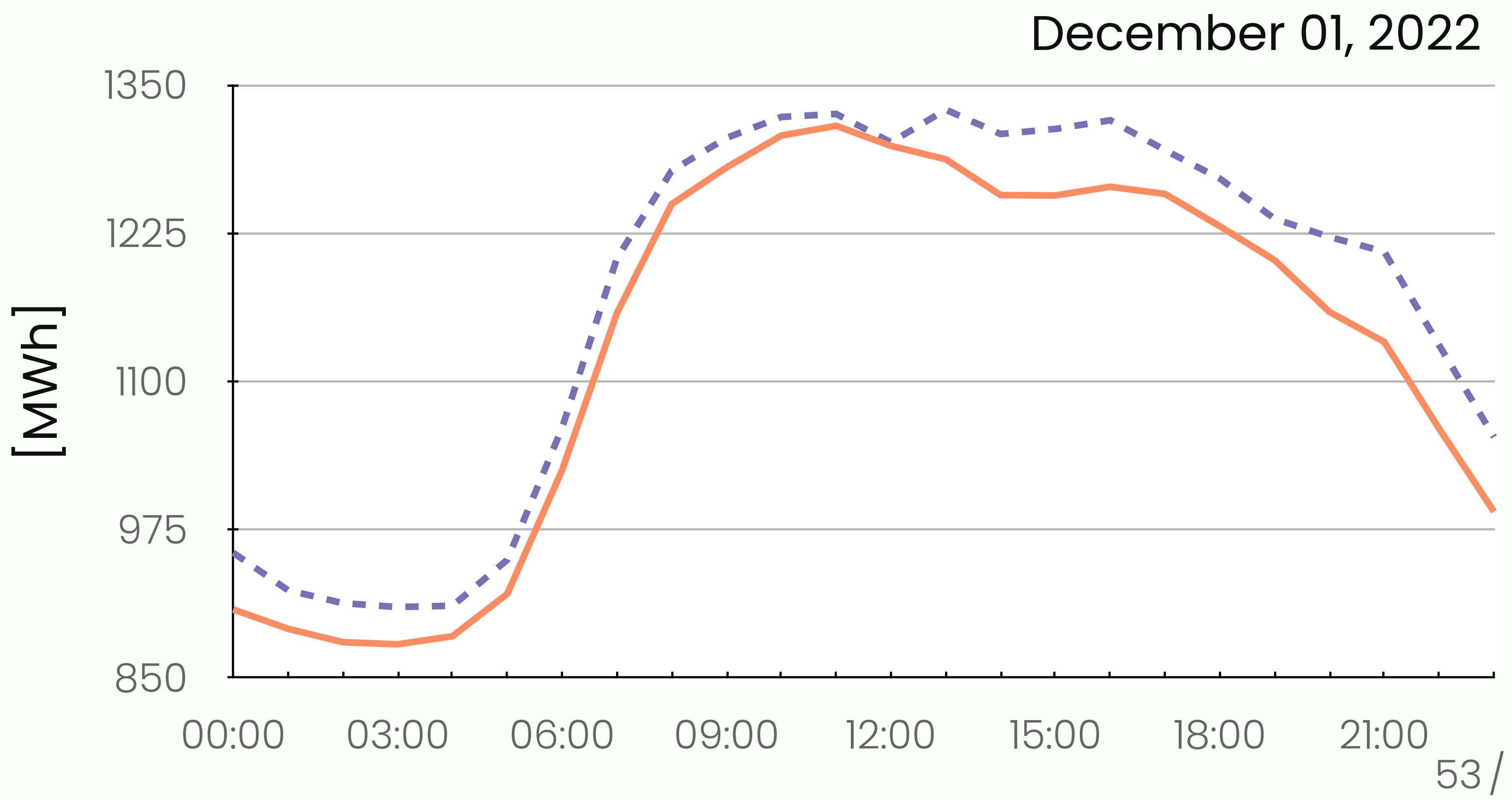
E. Ginzburg-Ganz, E: D. Horodi, O. Shadafny, U. Savir, R. Machlev, Y. Levron. Statistical foundations of Generative AI for optimal control problems in power systems: Comprehensive review and future directions. *Energies* 2025, 18, 2461.

UNDERSTANDING ACCURACY

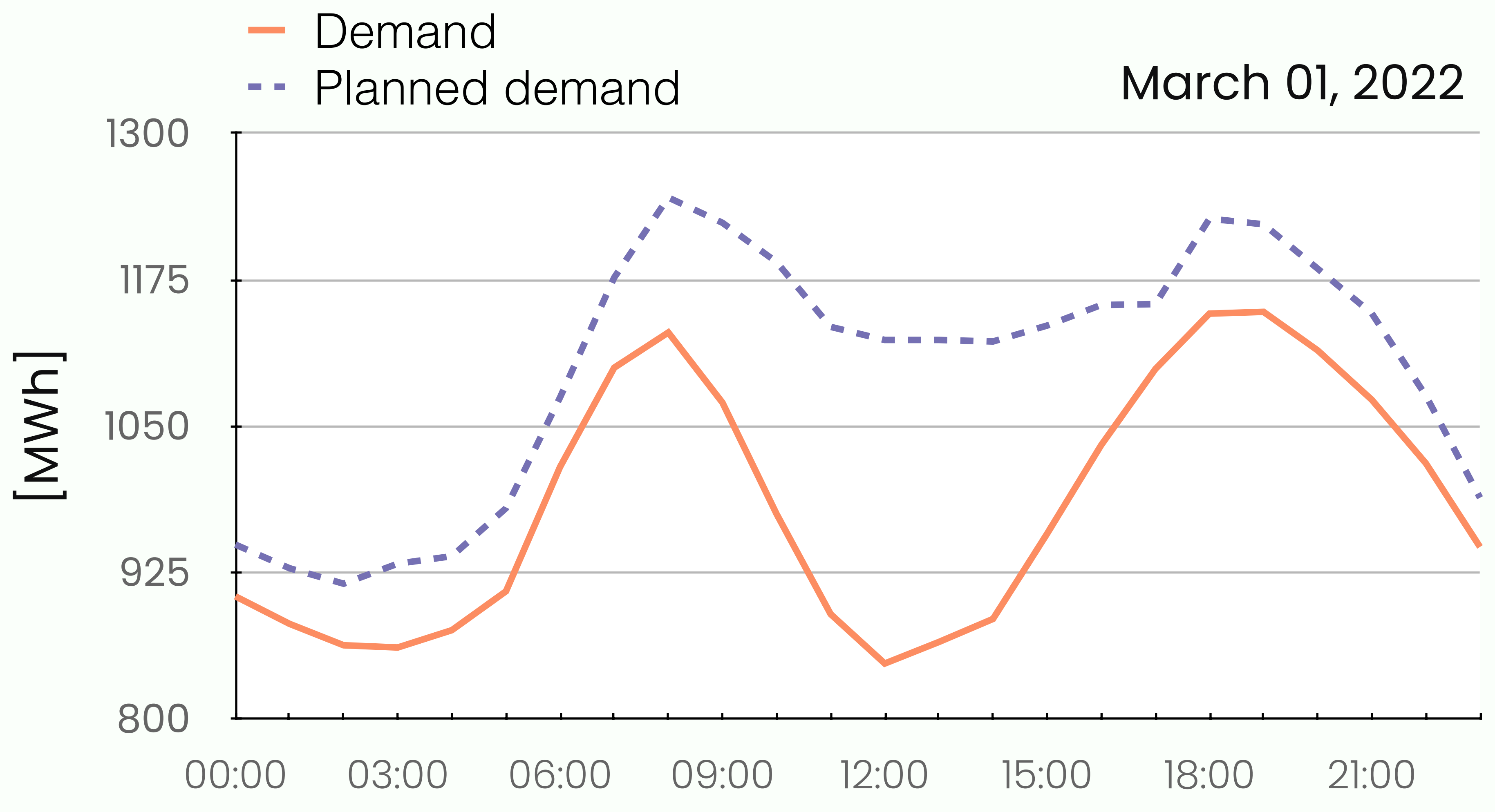


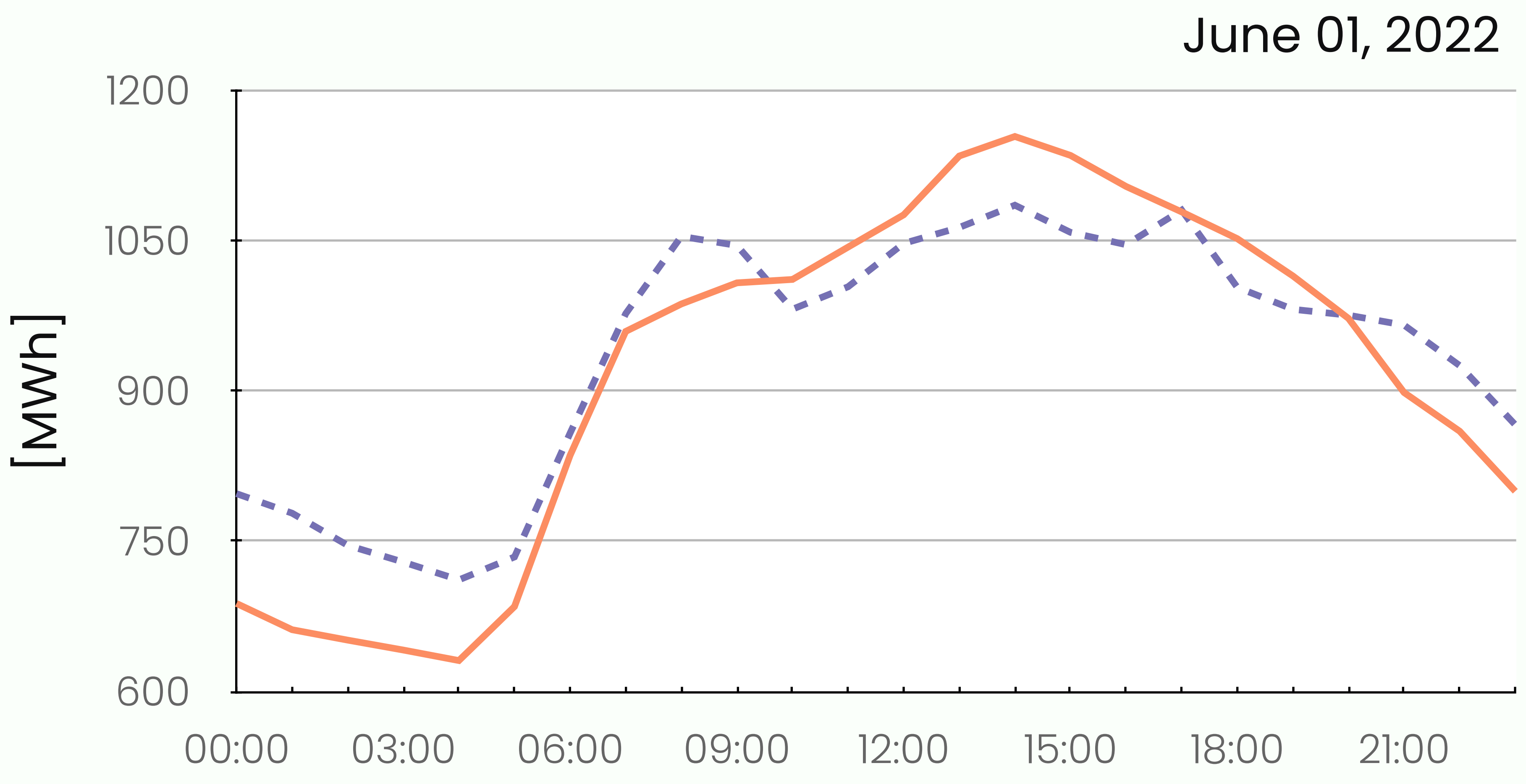




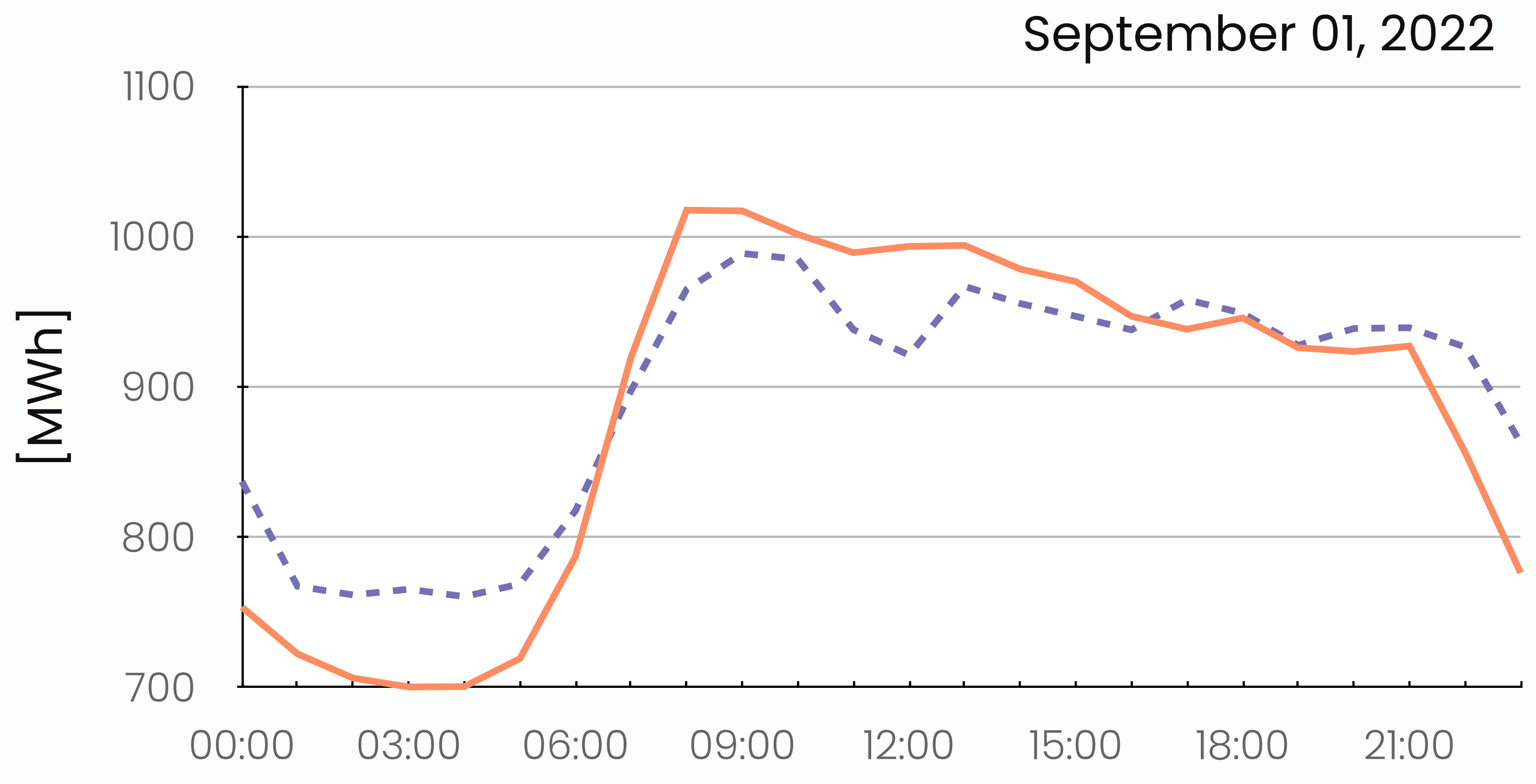


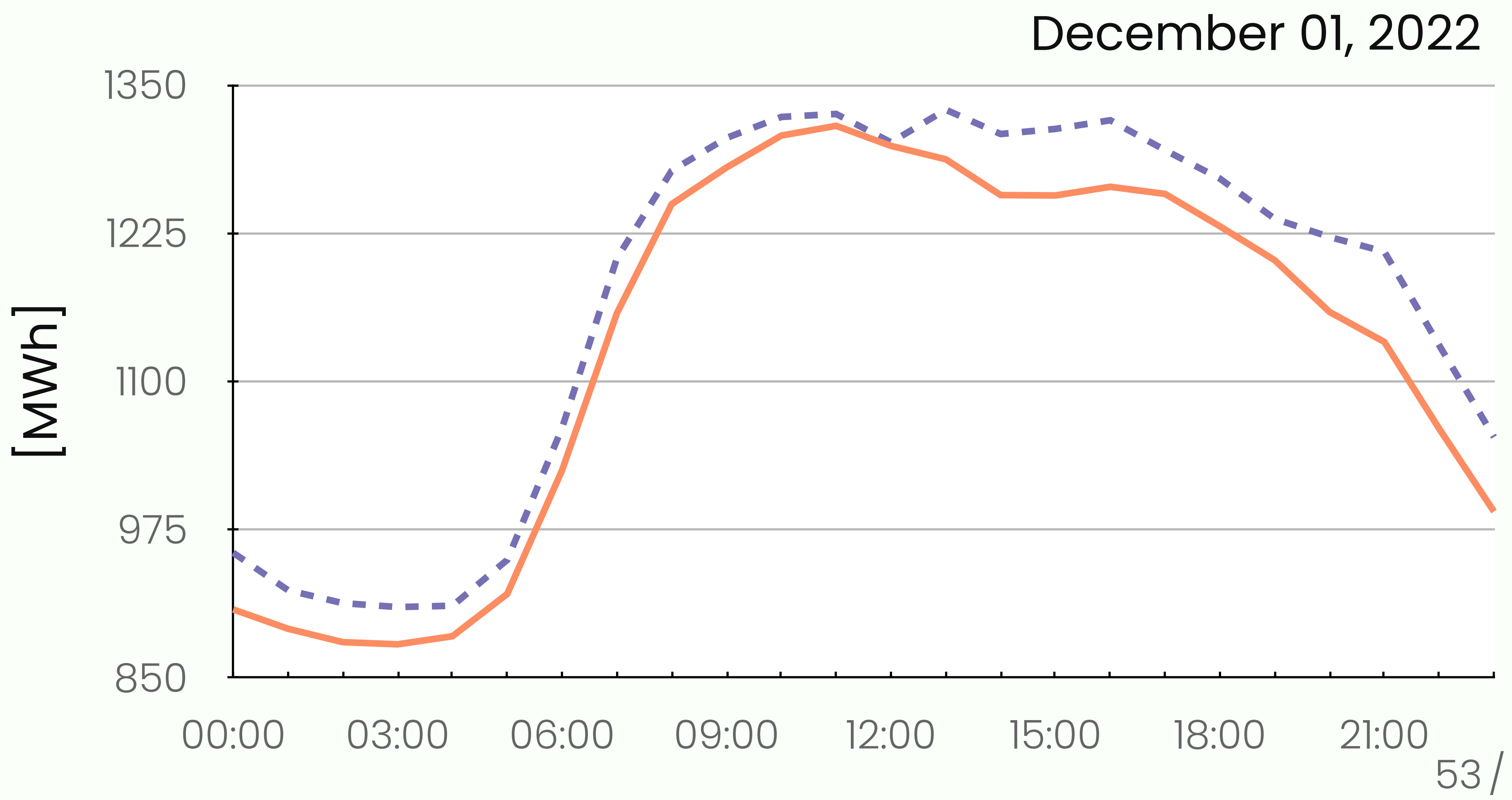
UNDERSTANDING ACCURACY





How to understand which prediction result is better?





COMMON METRICS

Metrics	Formula
Mean squared error (MSE)	$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i^{obs} - y_i^{pred})^2$
Root mean squared error (RMSE)	$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i^{obs} - y_i^{pred})^2}$
Mean absolute error (MAE)	$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left\lvert y_i^{obs} - y_i^{pred}\right\rvert$
Mean absolute percentage error (MAPE)	$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left\lvert\frac{y_i^{obs} - y_i^{pred}}{y_i^{obs}}\right\rvert \times 100\,\%$

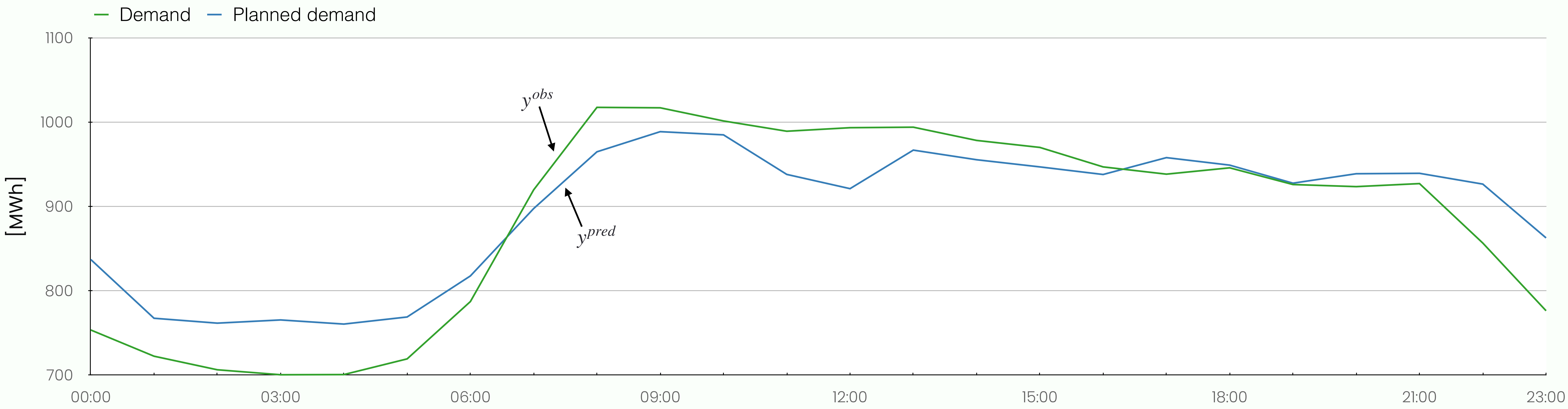
y_i^{obs} is the observed value

y_i^{pred} is the predicted value

i is the ith data point

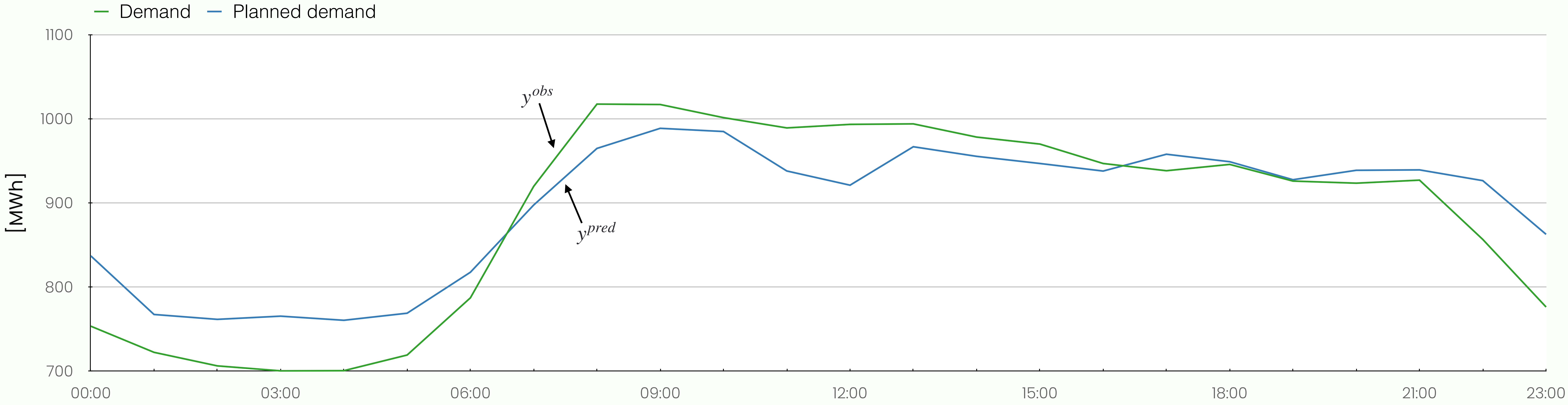
n is the total number of data points

ONE-DAY EXAMPLE (01.09.2022)



$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i^{obs} - y_i^{pred}}{y_i^{obs}}\right| \times 100\,\%$$

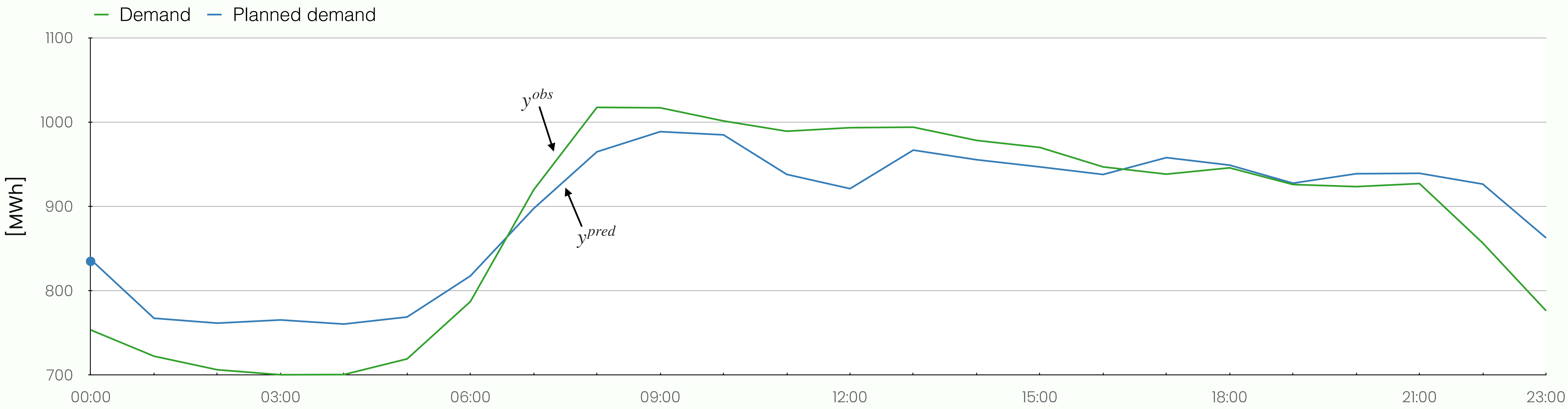
ONE-DAY EXAMPLE (01.09.2022)



i 1

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i^{obs} - y_i^{pred}}{y_i^{obs}}\right| \times 100\,\%$$

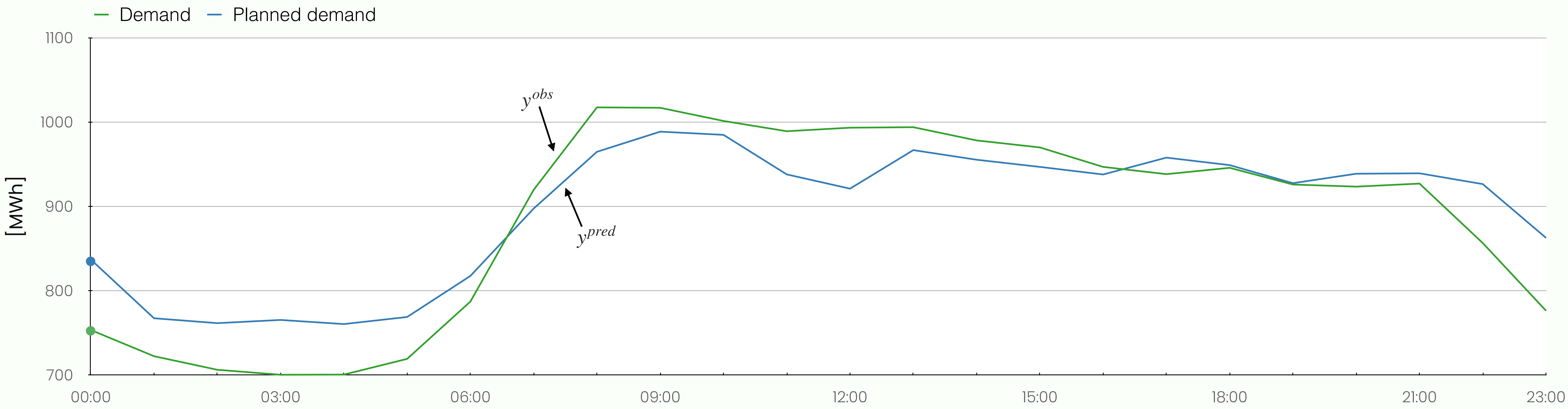
ONE-DAY EXAMPLE (01.09.2022)



i	1
y_i^{obs}	753.5

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i^{obs} - y_i^{pred}}{y_i^{obs}}\right| \times 100\,\%$$

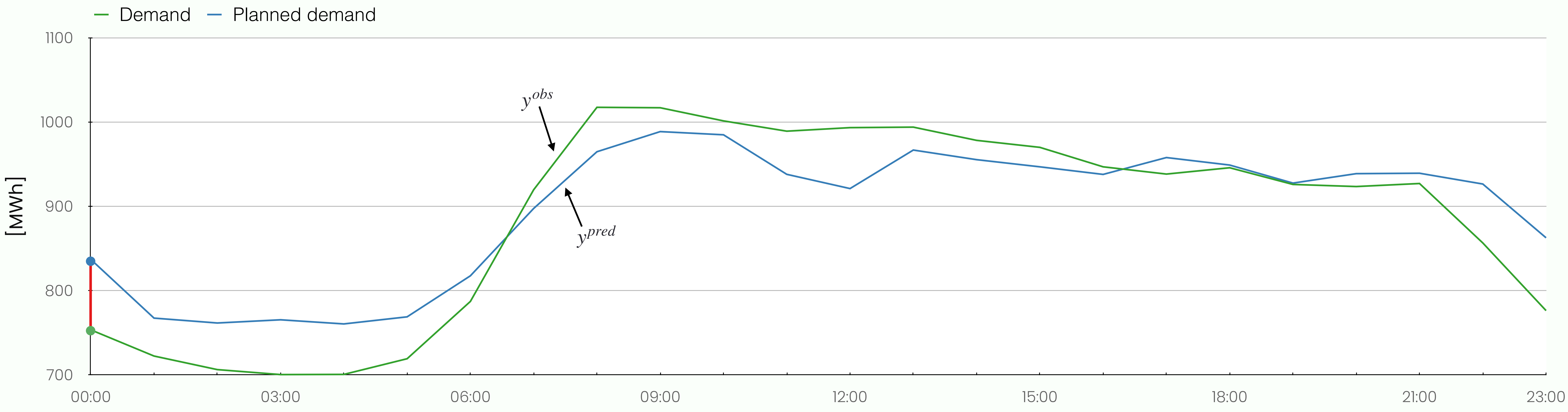
ONE-DAY EXAMPLE (01.09.2022)



i	1
y_i^{obs}	753.5
y_i^{pred}	837.1

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i^{obs} - y_i^{pred}}{y_i^{obs}}\right| \times 100\,\%$$

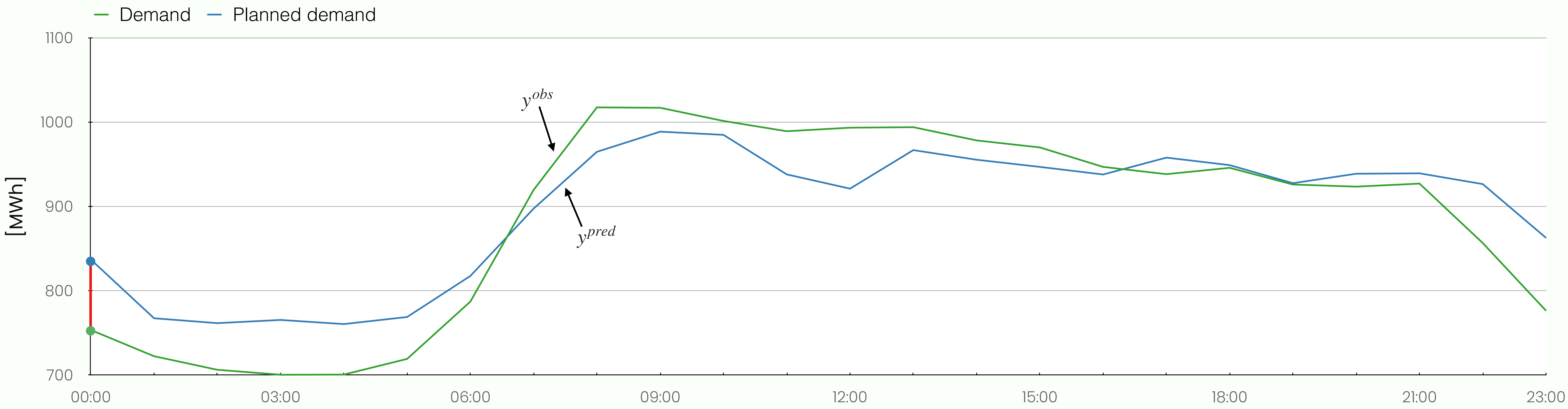
ONE-DAY EXAMPLE (01.09.2022)



i	1
y_i^{obs}	753.5
y_i^{pred}	837.1
$y_i^{obs} - y_i^{pred}$	−83.6

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i^{obs} - y_i^{pred}}{y_i^{obs}}\right| \times 100\,\%$$

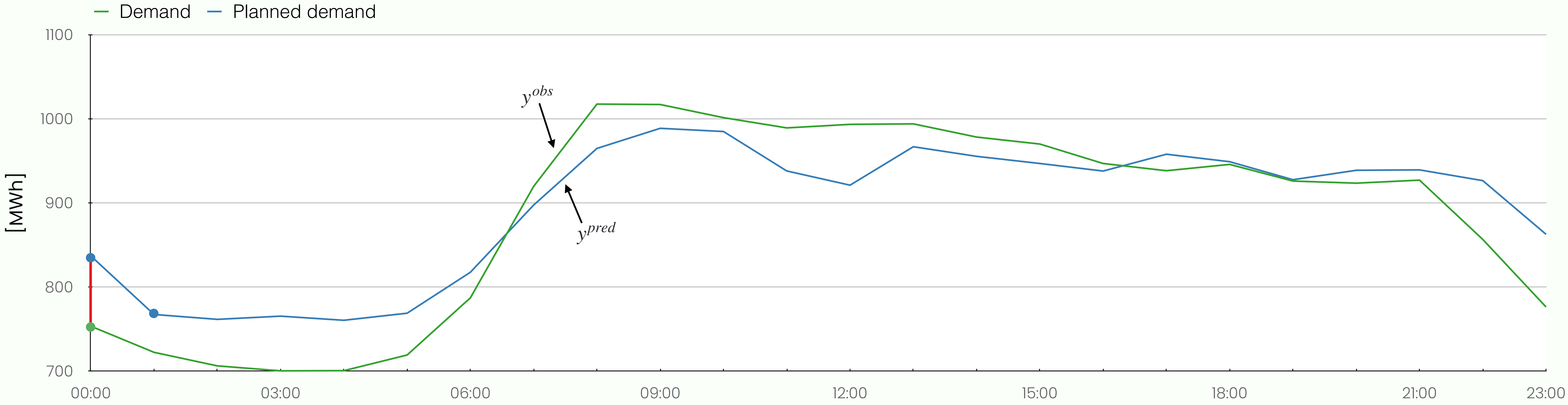
ONE-DAY EXAMPLE (01.09.2022)



i	1	2
y_i^{obs}	753.5	
y_i^{pred}	837.1	
$y_i^{obs} - y_i^{pred}$	−83.6	

$$\mathrm{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i^{obs} - y_i^{pred}}{y_i^{obs}}\right| \times 100\,\%$$

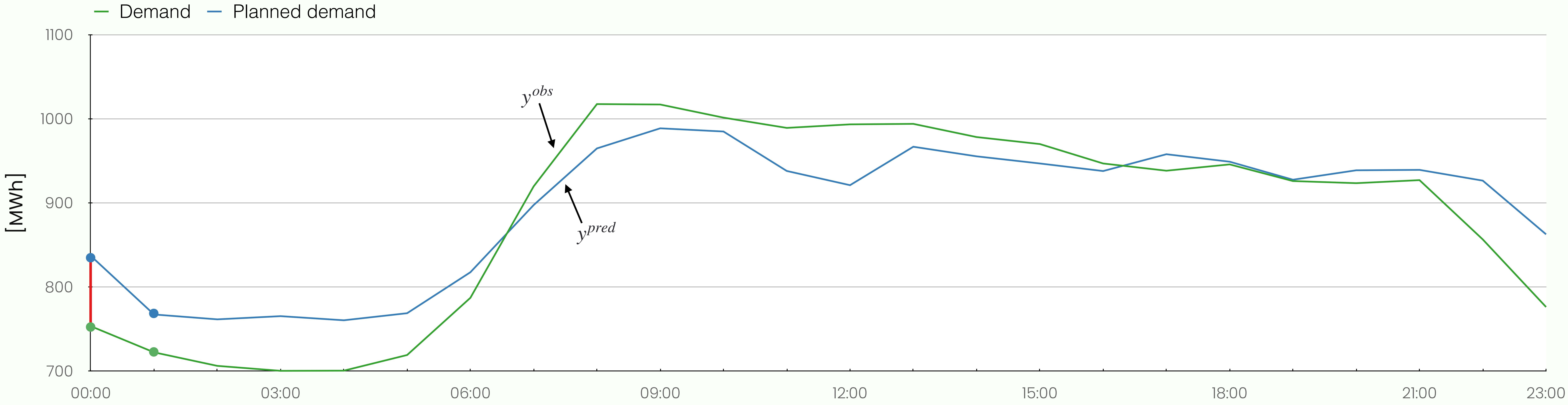
ONE-DAY EXAMPLE (01.09.2022)



i	1	2
y_i^{obs}	753.5	722.4
y_i^{pred}	837.1	
$y_i^{obs} - y_i^{pred}$	−83.6	

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i^{obs} - y_i^{pred}}{y_i^{obs}}\right| \times 100\,\%$$

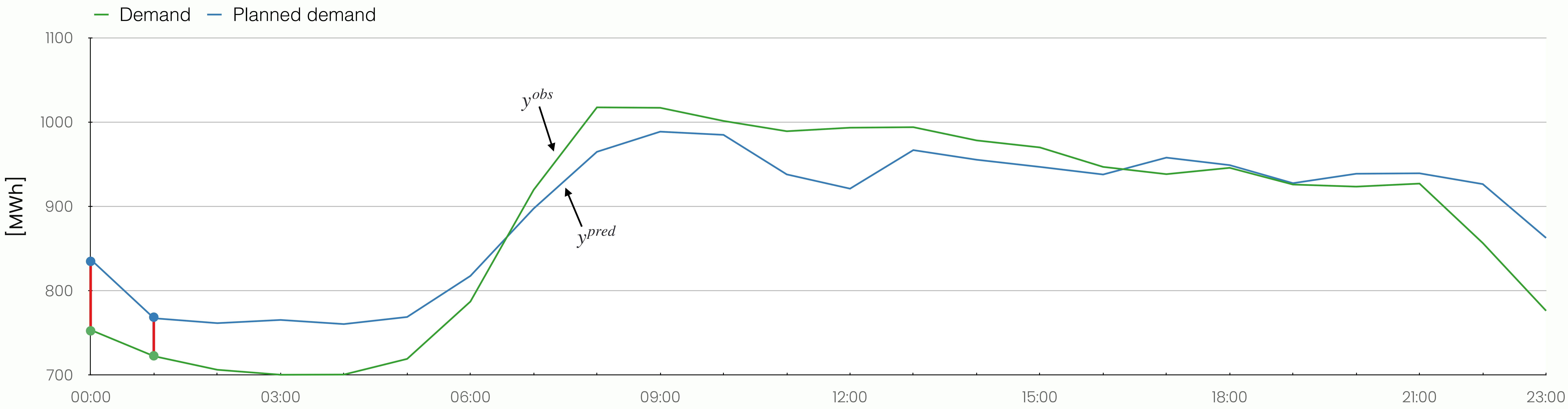
ONE-DAY EXAMPLE (01.09.2022)



i	1	2
y_i^{obs}	753.5	722.4
y_i^{pred}	837.1	767.4
$y_i^{obs} - y_i^{pred}$	-83.6	

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i^{obs} - y_i^{pred}}{y_i^{obs}}\right| \times 100\,\%$$

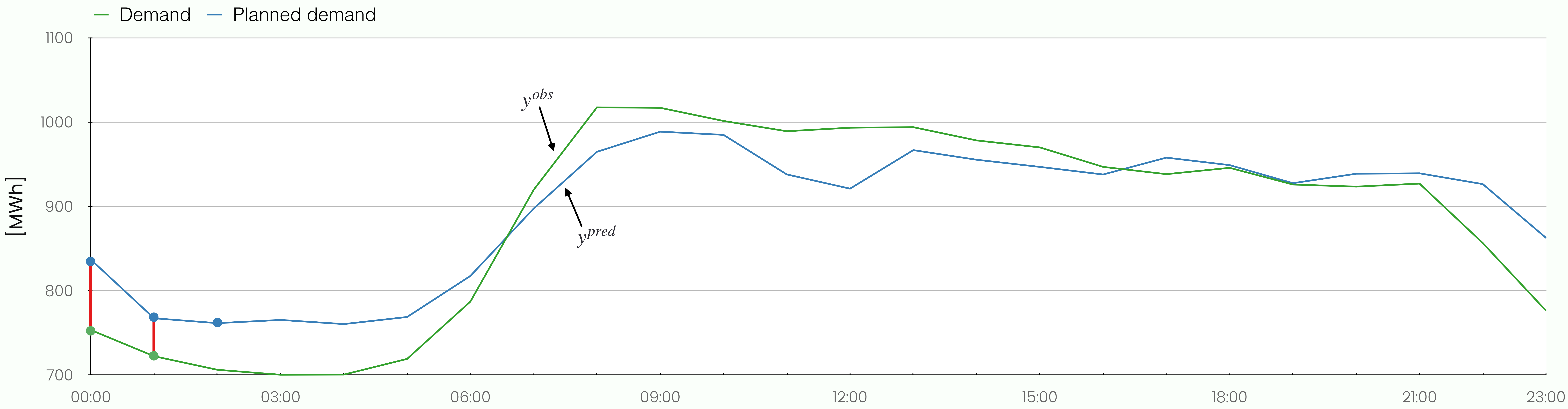
ONE-DAY EXAMPLE (01.09.2022)



i	1	2
y_i^{obs}	753.5	722.4
y_i^{pred}	837.1	767.4
$y_i^{obs} - y_i^{pred}$	-83.6	-45

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i^{obs} - y_i^{pred}}{y_i^{obs}}\right| \times 100\,\%$$

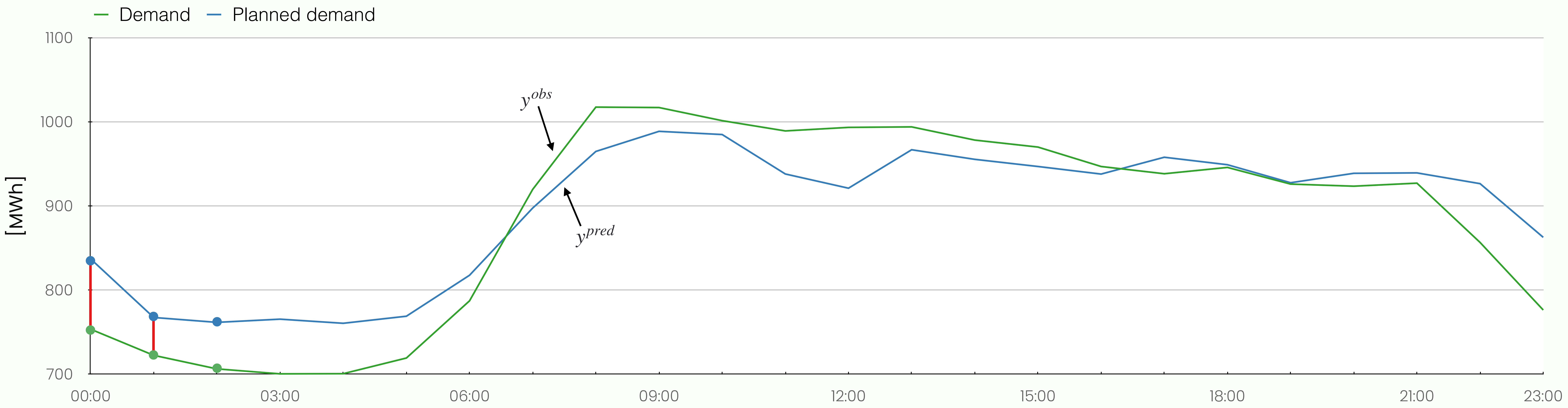
ONE-DAY EXAMPLE (01.09.2022)



i	1	2	3
y_i^{obs}	753.5	722.4	706.3
y_i^{pred}	837.1	767.4	
$y_i^{obs} - y_i^{pred}$	-83.6	-45	

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n} \left| \frac{y_i^{obs} - y_i^{pred}}{y_i^{obs}} \right| \times 100\,\%$$

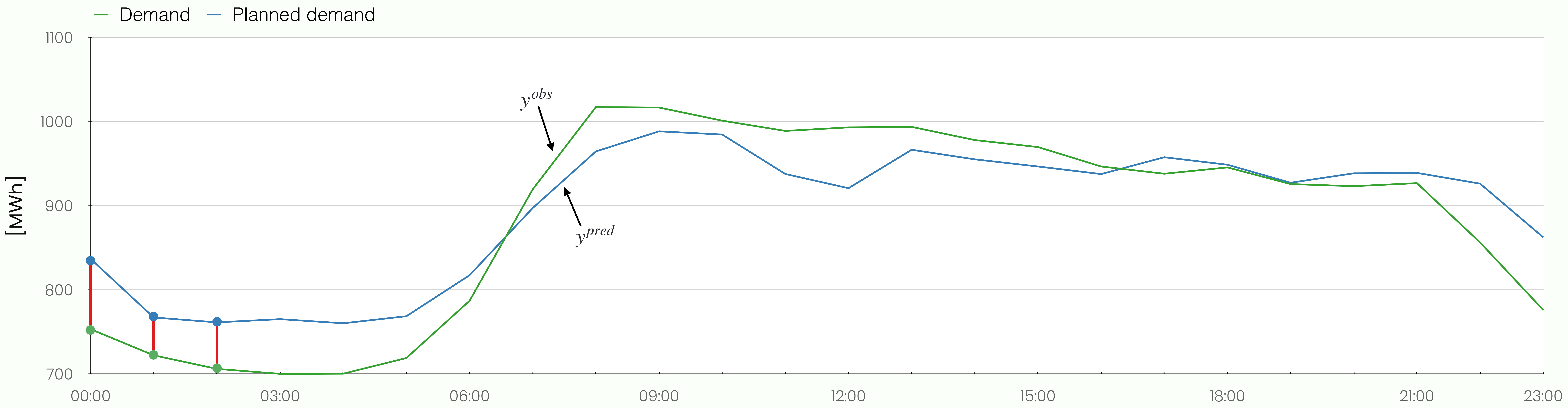
ONE-DAY EXAMPLE (01.09.2022)



i	1	2	3
y_i^{obs}	753.5	722.4	706.3
y_i^{pred}	837.1	767.4	761.6
$y_i^{obs} - y_i^{pred}$	-83.6	-45	

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i^{obs} - y_i^{pred}}{y_i^{obs}}\right| \times 100\,\%$$

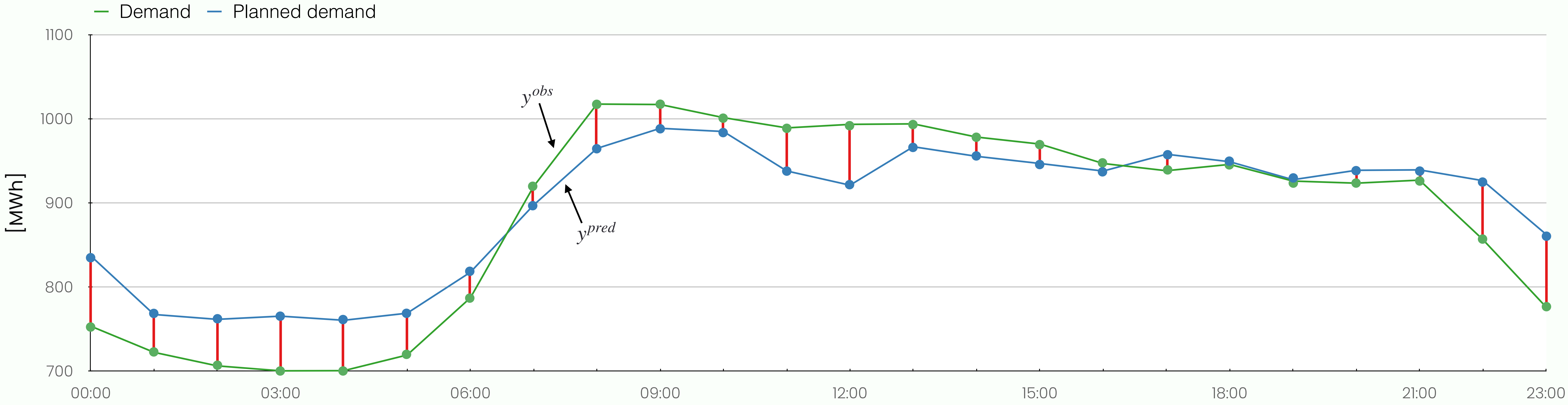
ONE-DAY EXAMPLE (01.09.2022)



i	1	2	3
y_i^{obs}	753.5	722.4	706.3
y_i^{pred}	837.1	767.4	761.6
$y_i^{obs} - y_i^{pred}$	-83.6	-45	-55.3

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i^{obs} - y_i^{pred}}{y_i^{obs}}\right| \times 100\,\%$$

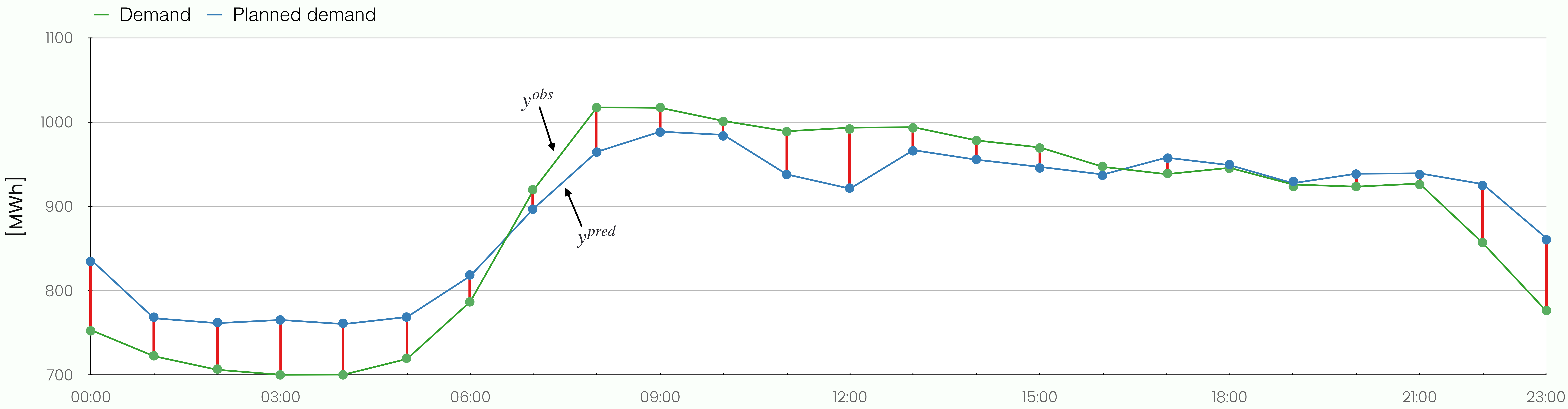
ONE-DAY EXAMPLE (01.09.2022)



i	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
y_i^{obs}	753.5	722.4	706.3	700.4	700.6	719.2	787.2	919.7	1017.5	1017	1001.4	989.2	993.4	994	978.3	970	946.9	938.3	945.8	926	923.5	927.1	856.6	776.3
y_i^{pred}	837.1	767.4	761.6	765.4	760.5	768.9	817.6	897.5	964.8	988.7	984.9	938	921.1	966.8	955.4	946.9	937.9	957.9	949	927.6	938.8	939.3	926.5	862.7
$y_i^{obs} - y_i^{pred}$	-83.6	-45	-55.3	-65	-59.9	-49.7	-30.4	22.2	52.7	28.3	16.5	51.2	72.3	27.2	22.9	23.1	9	-19.6	-3.2	-1.6	-15.3	-12.2	-69.9	-86.4

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{y_i^{obs} - y_i^{pred}}{y_i^{obs}} \right| \times 100\,\%$$

ONE-DAY EXAMPLE (01.09.2022)

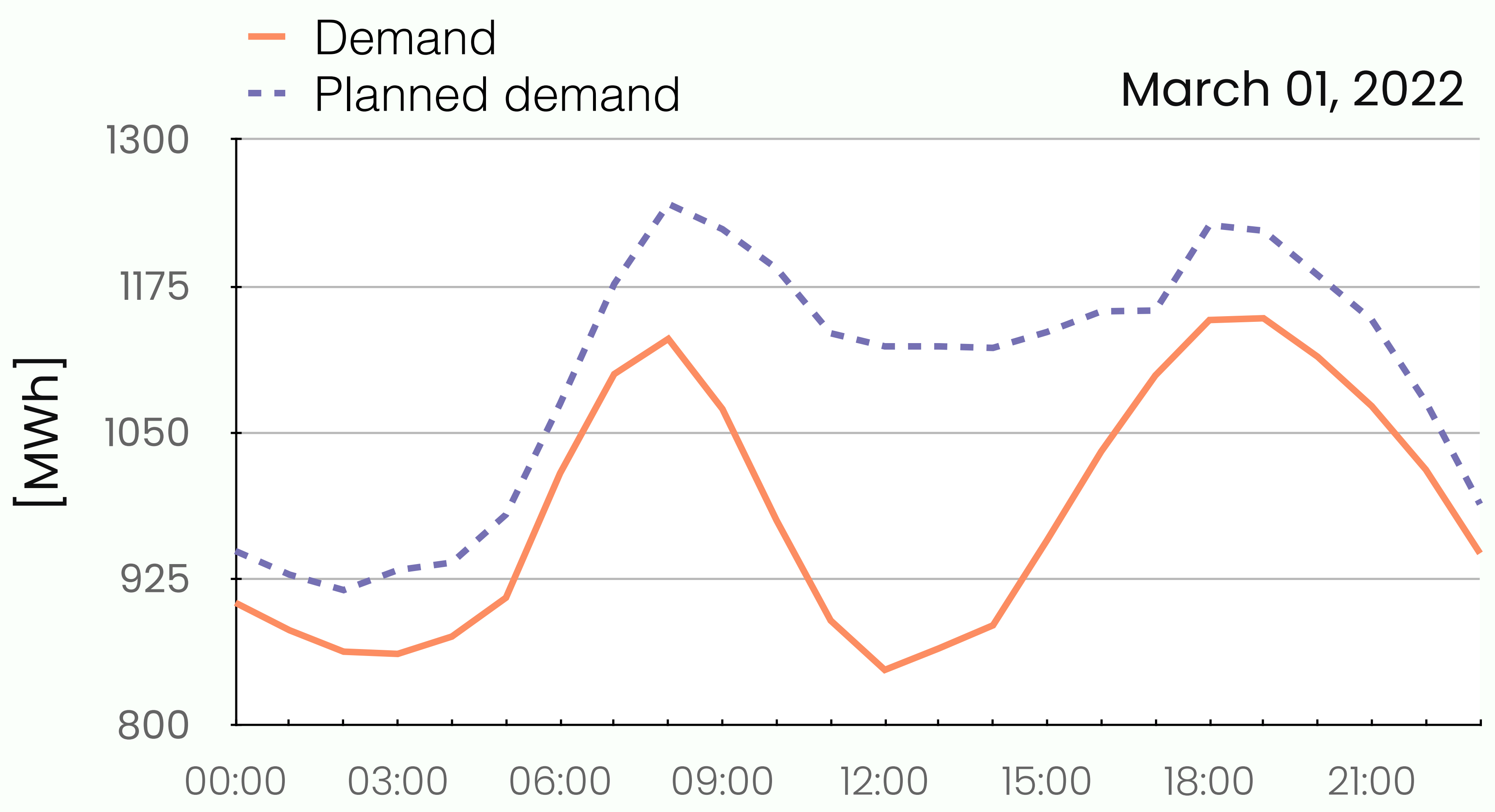


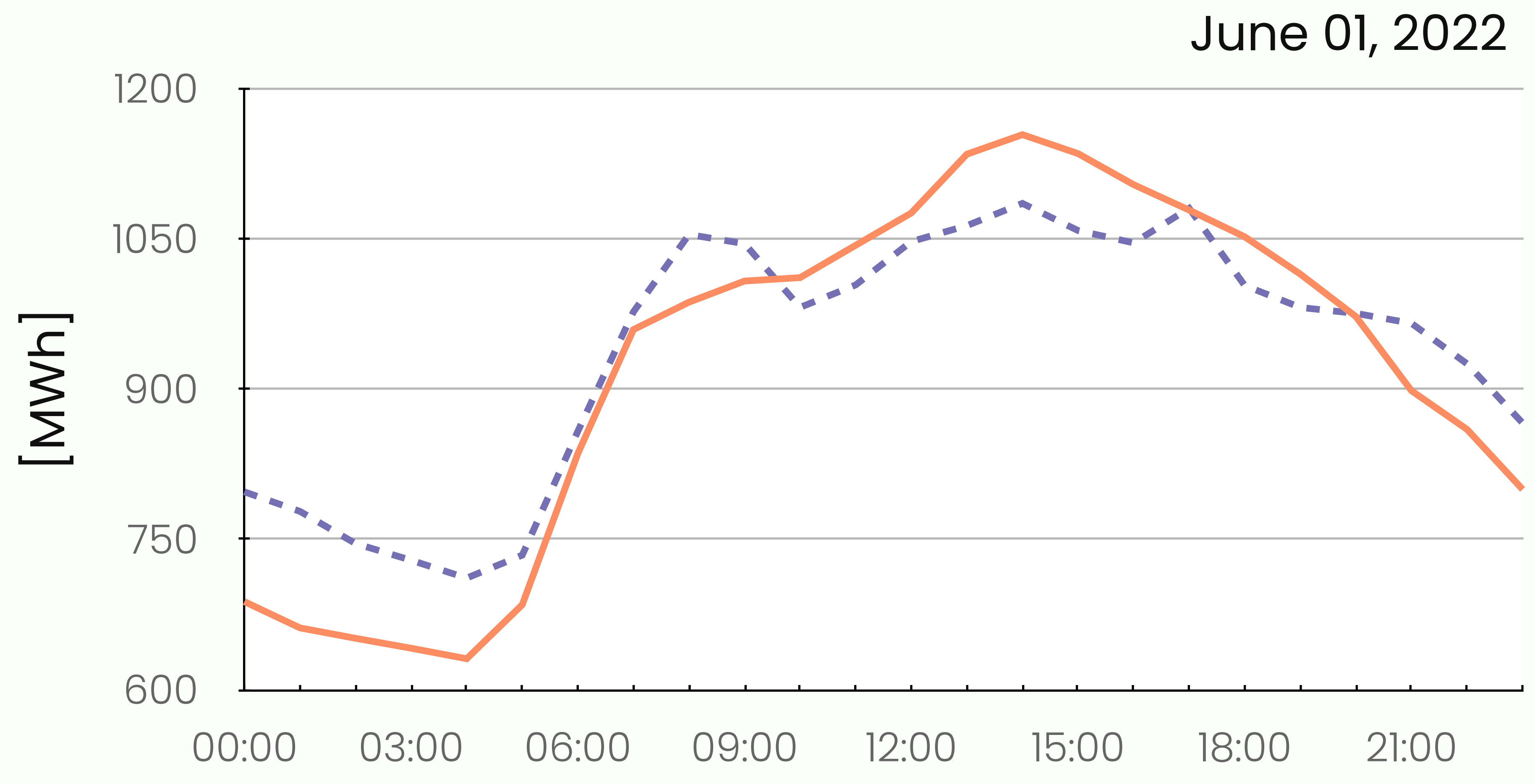
i	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
y_i^{obs}	753.5	722.4	706.3	700.4	700.6	719.2	787.2	919.7	1017.5	1017	1001.4	989.2	993.4	994	978.3	970	946.9	938.3	945.8	926	923.5	927.1	856.6	776.3
y_i^{pred}	837.1	767.4	761.6	765.4	760.5	768.9	817.6	897.5	964.8	988.7	984.9	938	921.1	966.8	955.4	946.9	937.9	957.9	949	927.6	938.8	939.3	926.5	862.7
$y_i^{obs} - y_i^{pred}$	-83.6	-45	-55.3	-65	-59.9	-49.7	-30.4	22.2	52.7	28.3	16.5	51.2	72.3	27.2	22.9	23.1	9	-19.6	-3.2	-1.6	-15.3	-12.2	-69.9	-86.4

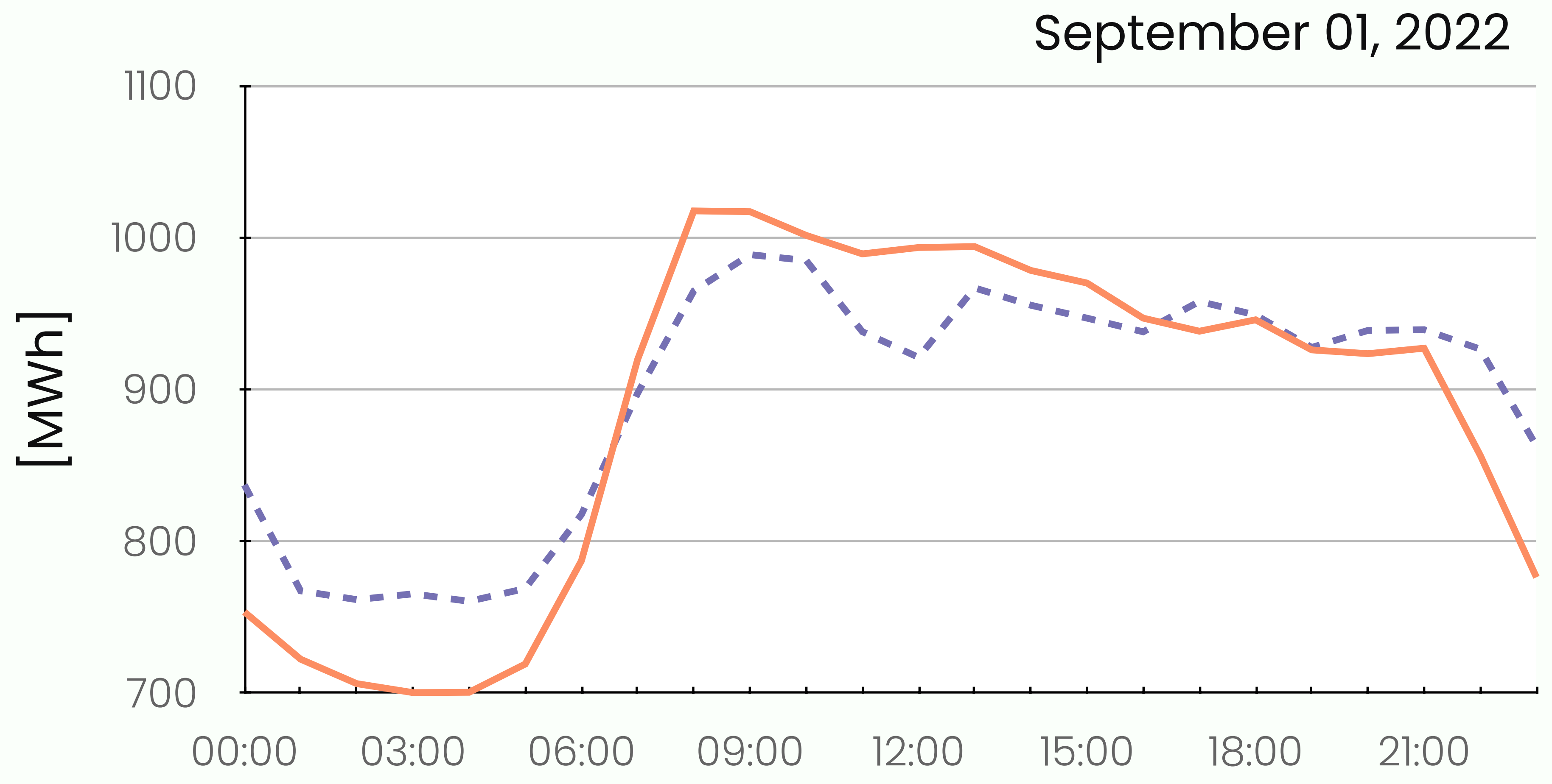
$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i^{obs} - y_i^{pred}}{y_i^{obs}}\right| \times 100\,\% \xrightarrow{n=24} \text{MAPE} = \frac{1}{24}\left(\frac{83.6}{753.5} + \frac{45}{722.4} + \frac{55.3}{706.3} + \cdots + \frac{12.2}{927.1} + \frac{69.9}{856.6} + \frac{86.4}{776.3}\right) \times 100\,\%$$

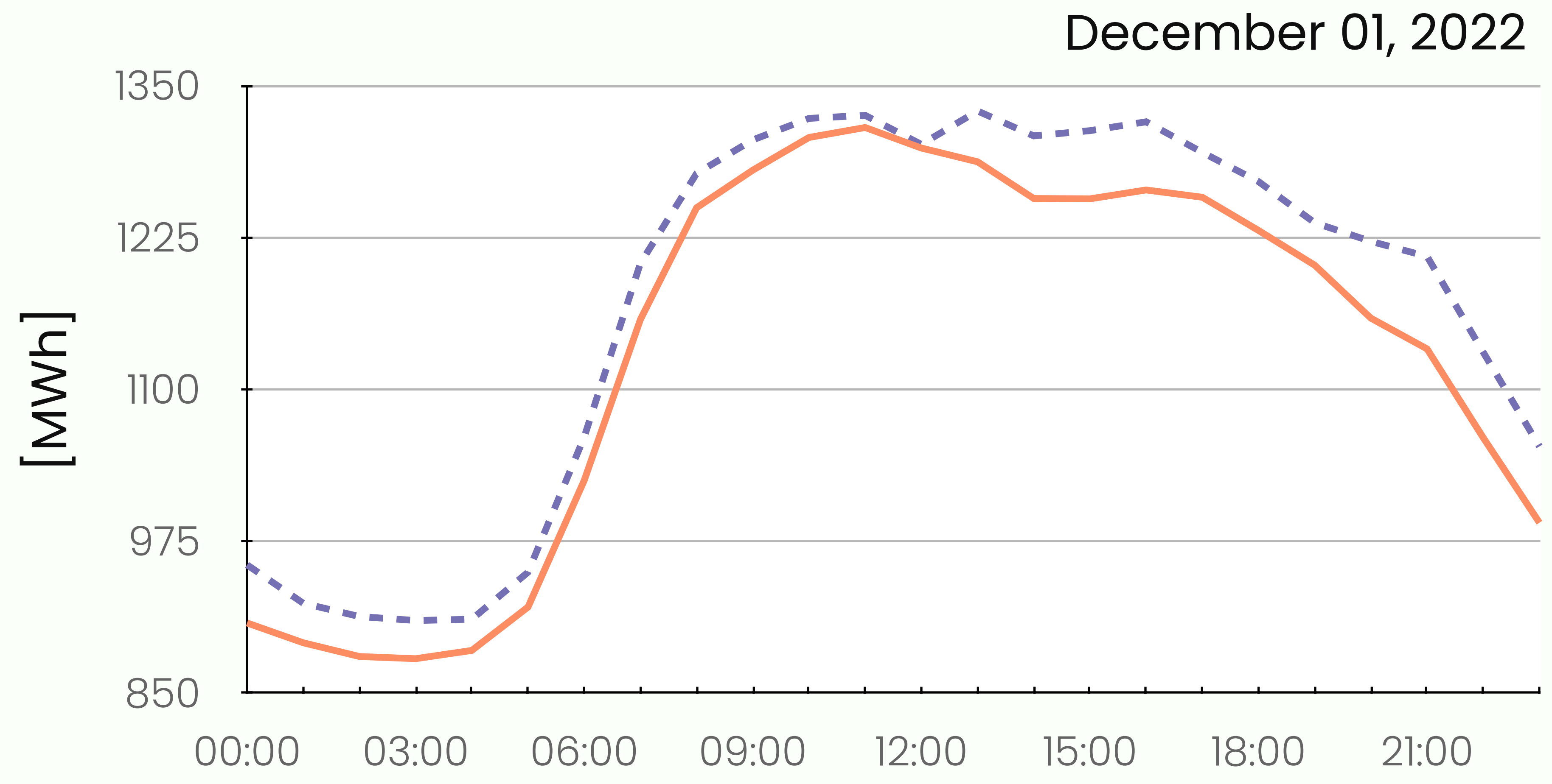
$$= \frac{1}{24}\left(11.09\% + 6.22\% + 7.83\% + \cdots + 1.32\% + 8.16\% + 11.13\,\%\right)$$

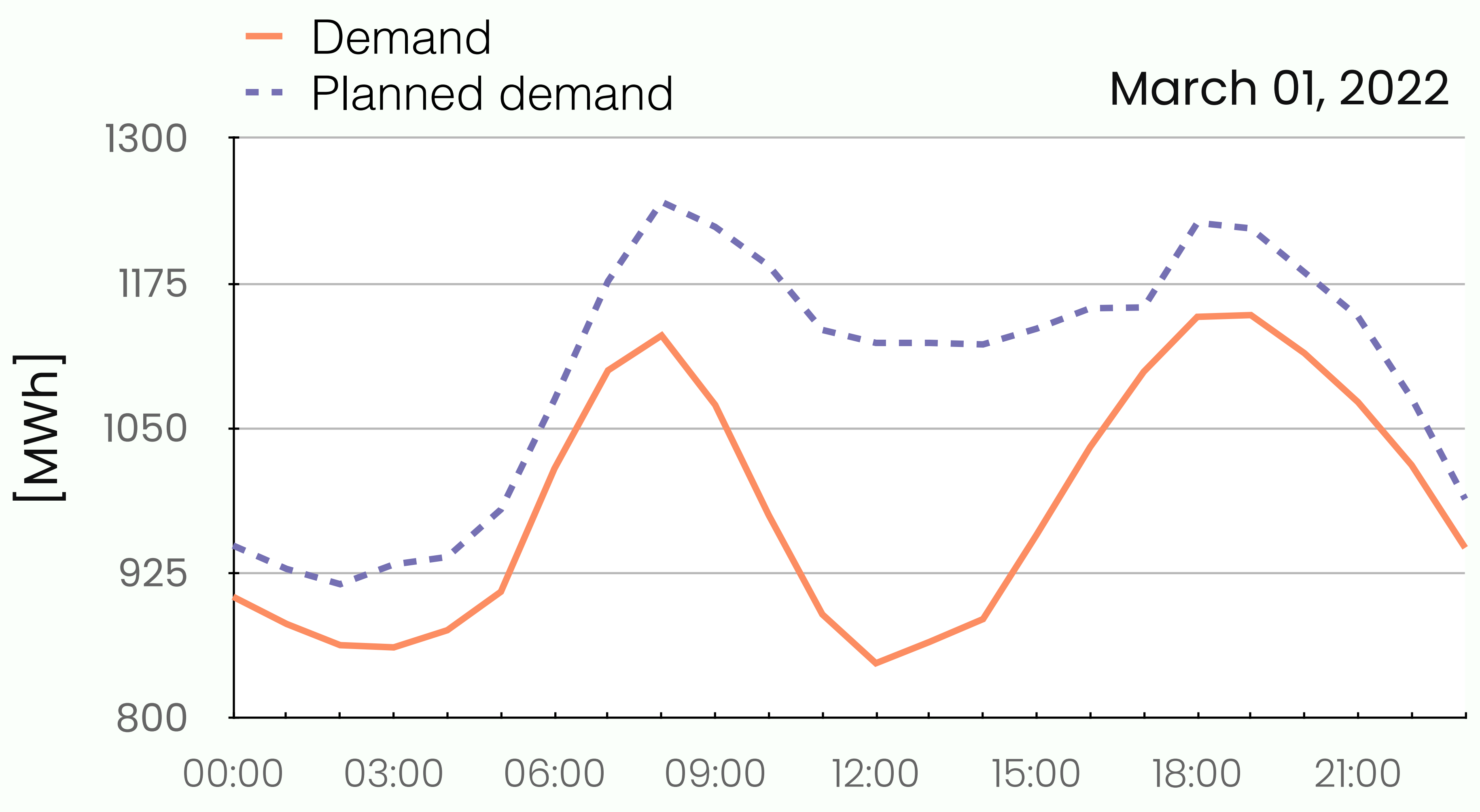
$$\approx 4.65\,\%$$


Demand
Planned demand
March 01, 2022
1300
1175
1050
925
800
[MWh]
00:00
03:00
06:00
09:00
12:00
15:00
18:00
21:00


June 01, 2022
1200
1050
900
750
600
[MWh]
00:00
03:00
06:00
09:00
12:00
15:00
18:00
21:00

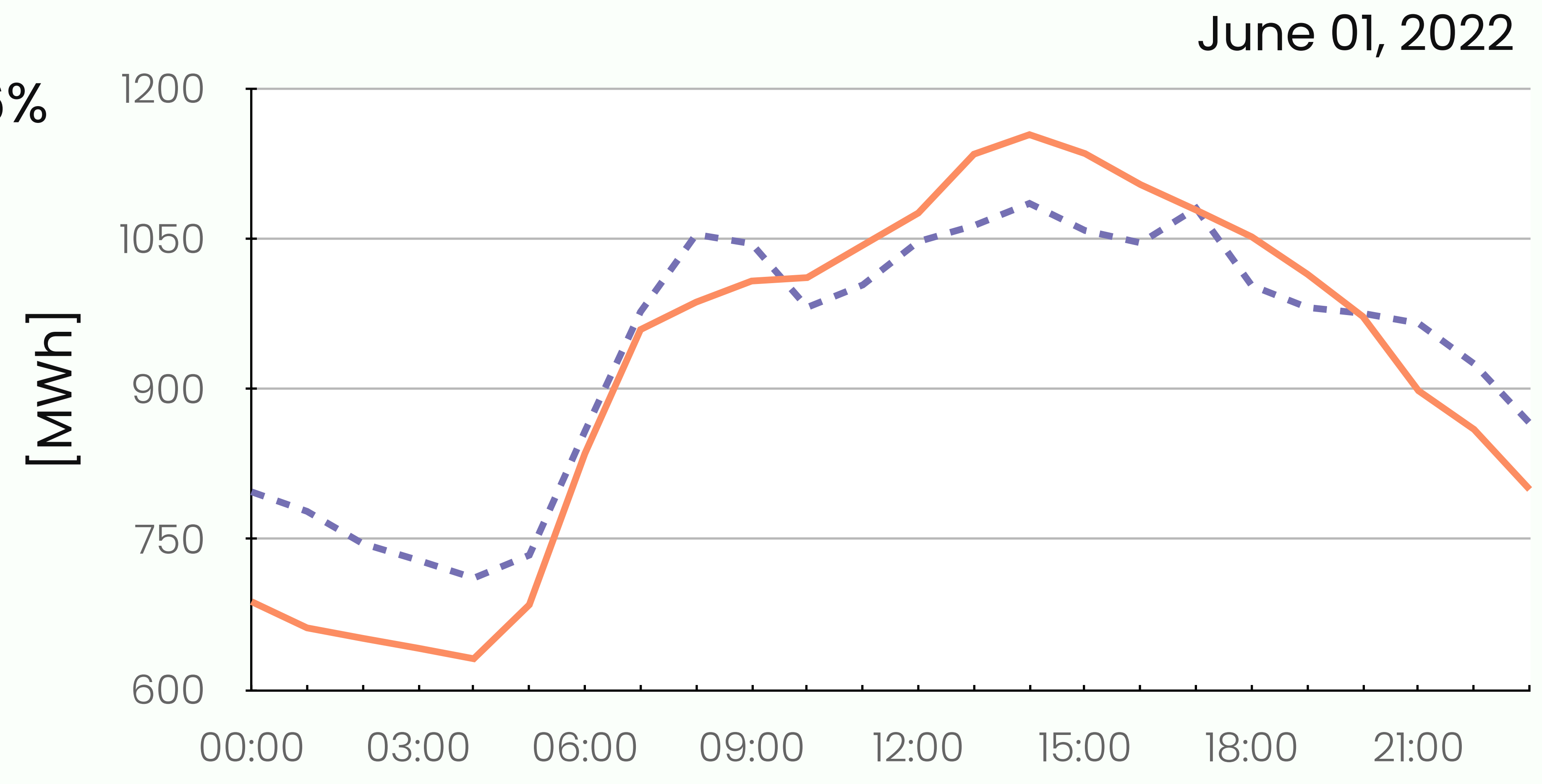

September 01, 2022
1100
1000
900
800
700
[MWh]
00:00
03:00
06:00
09:00
12:00
15:00
18:00
21:00

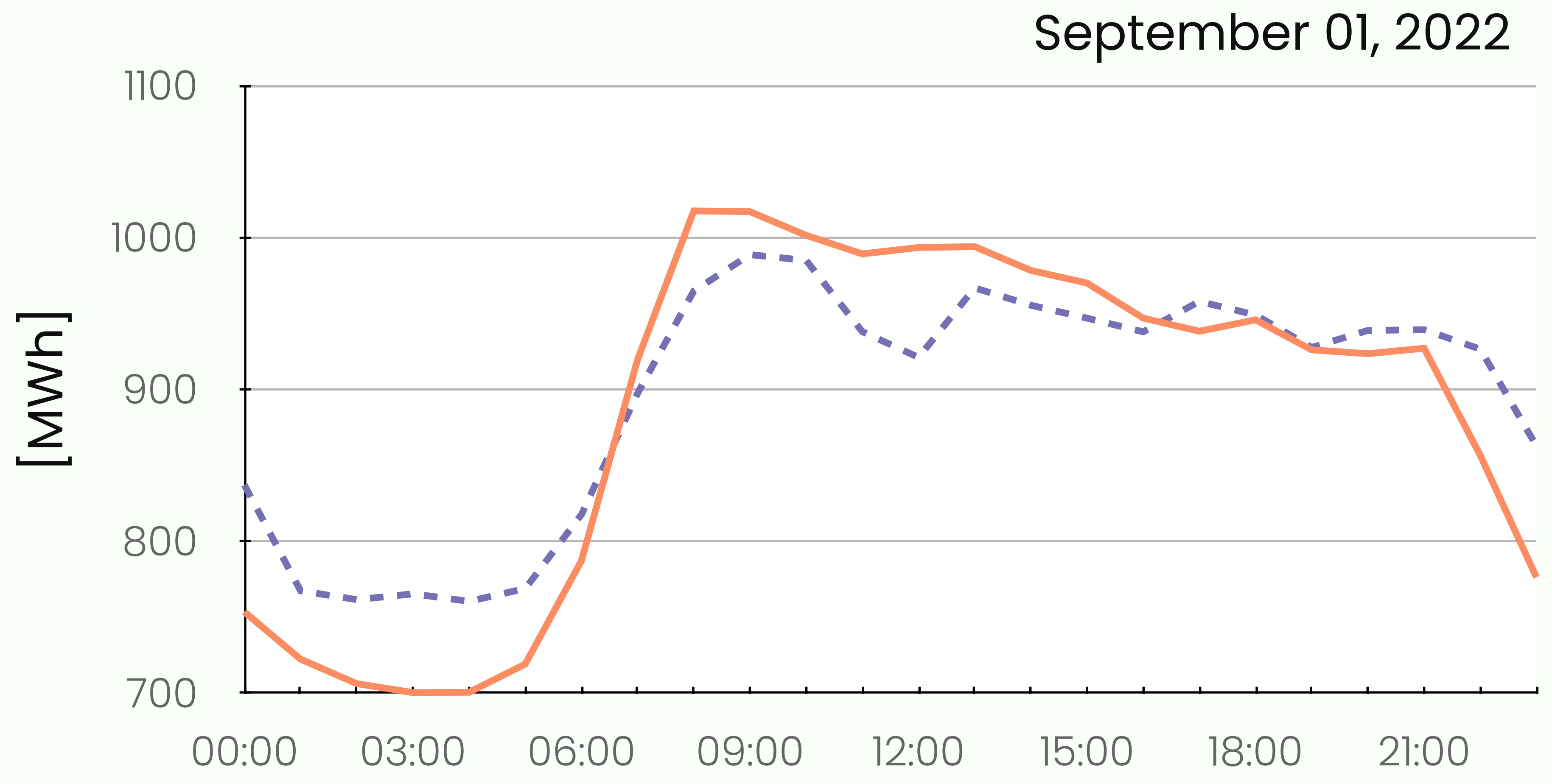

December 01, 2022
1350
1225
1100
975
850
[MWh]
00:00
03:00
06:00
09:00
12:00
15:00
18:00
21:00



MAPE1 = 12%

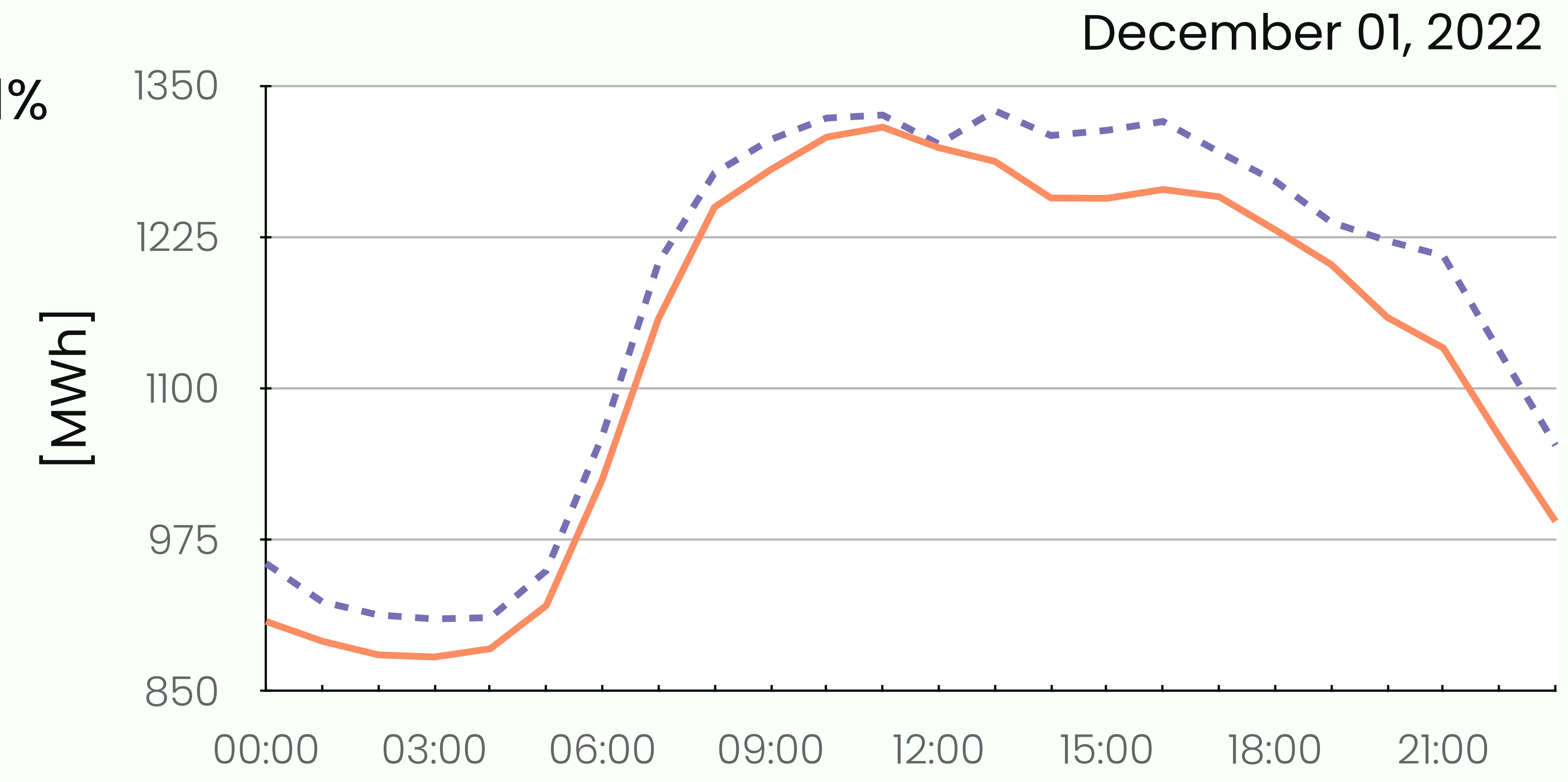
MAPE2 = 6.76%

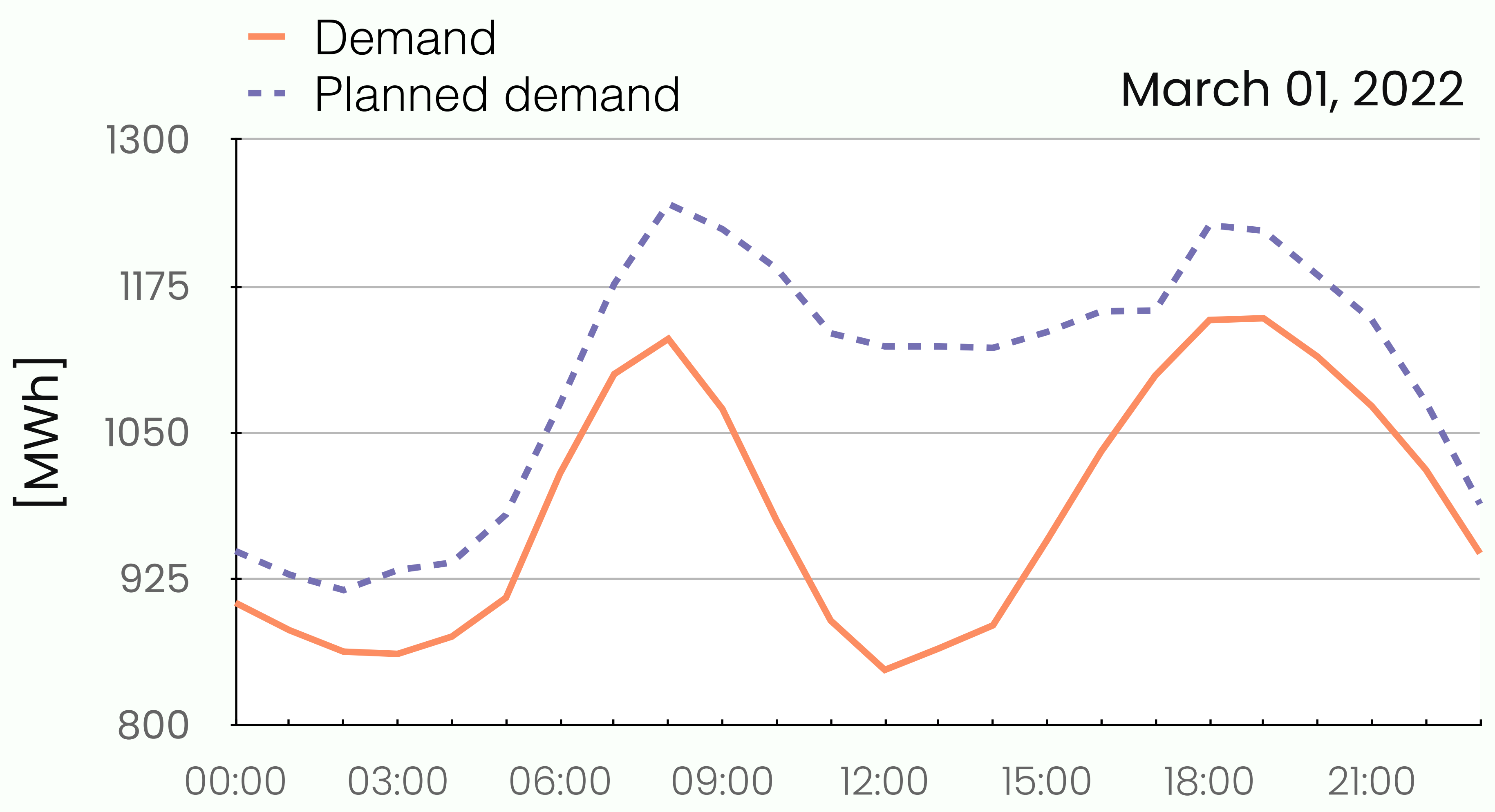




MAPE3 = 4.65%

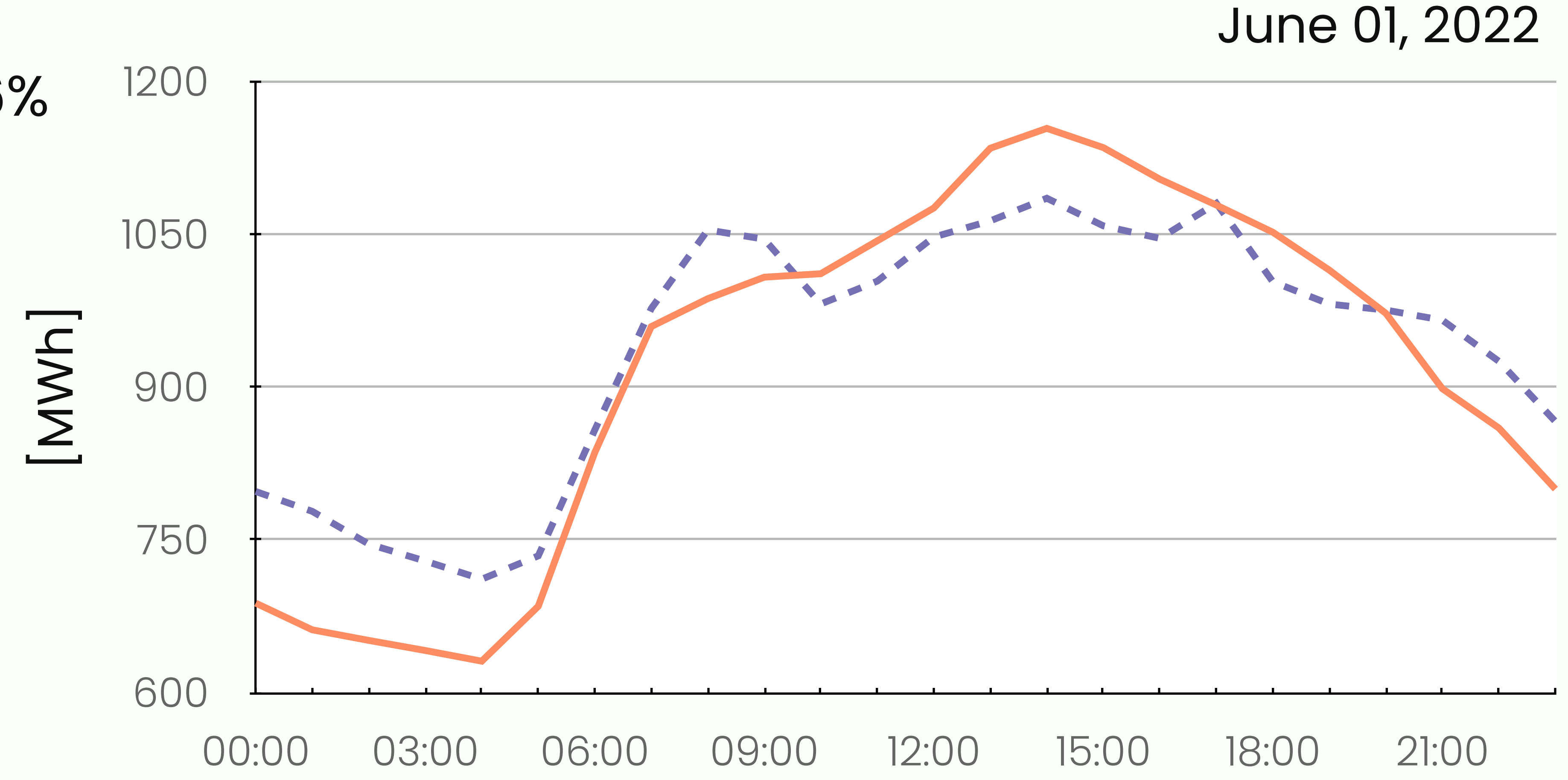
MAPE4 = 3.61%



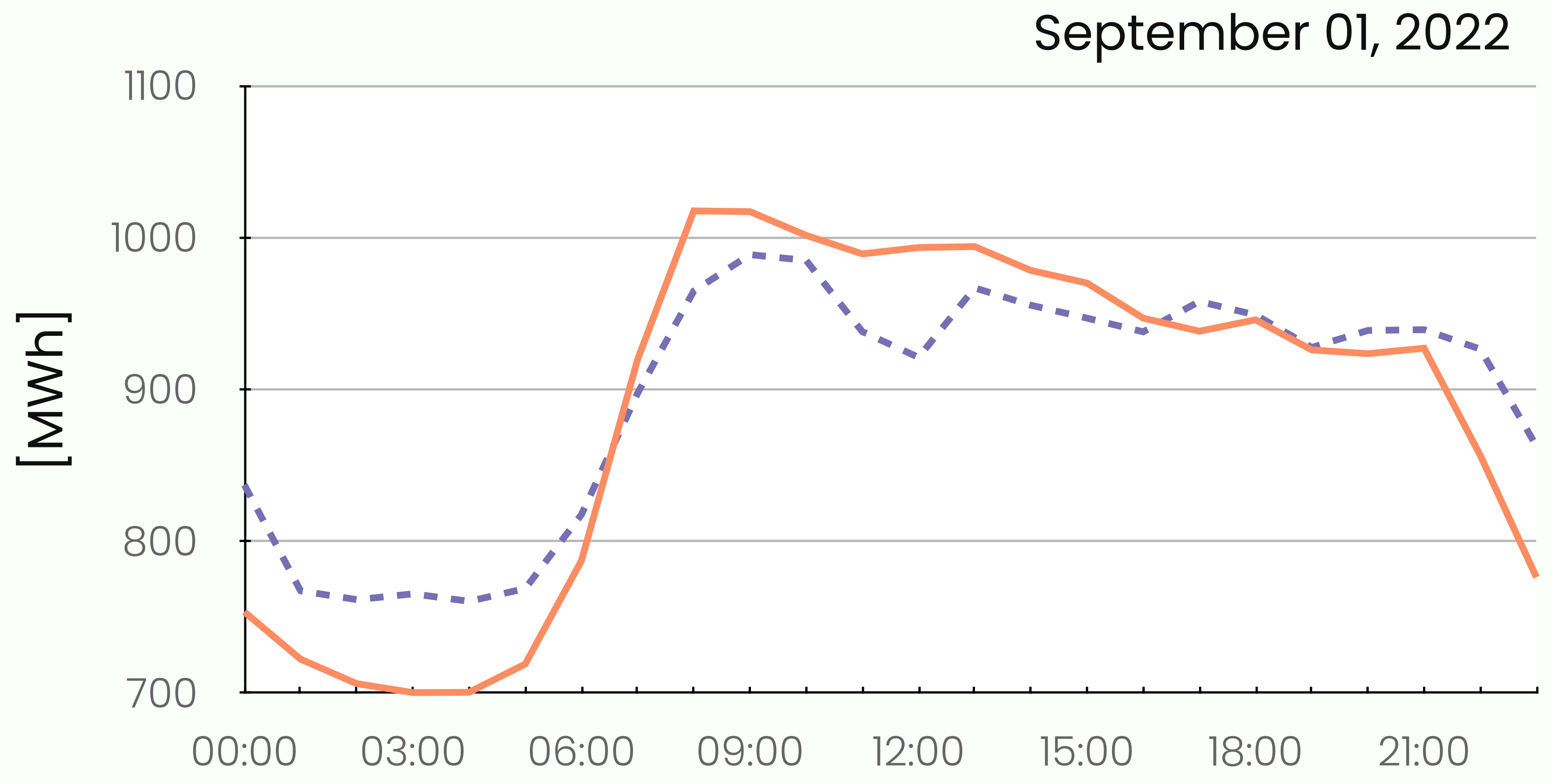


MAPE1 = 12%

MAPE2 = 6.76%

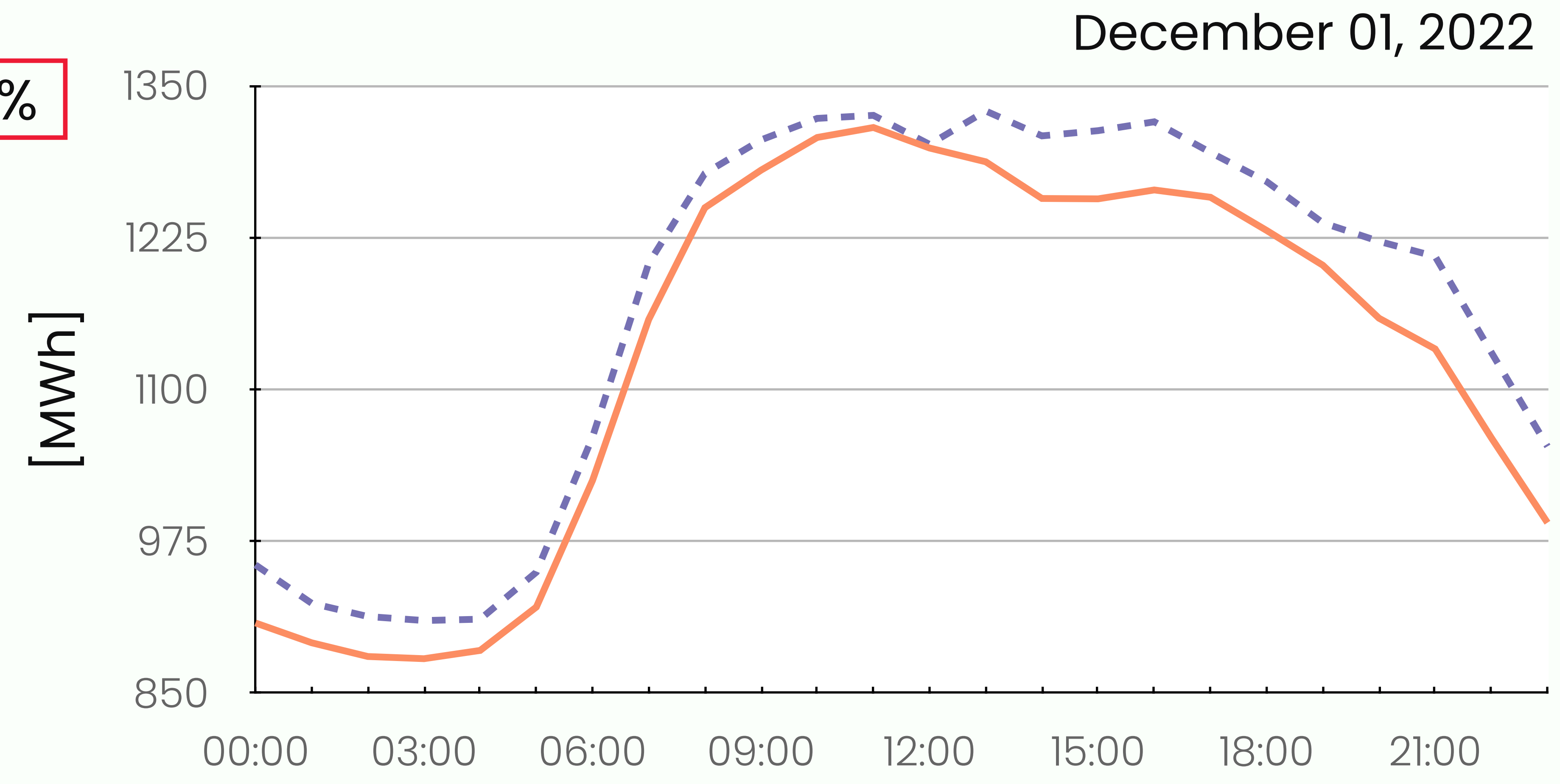


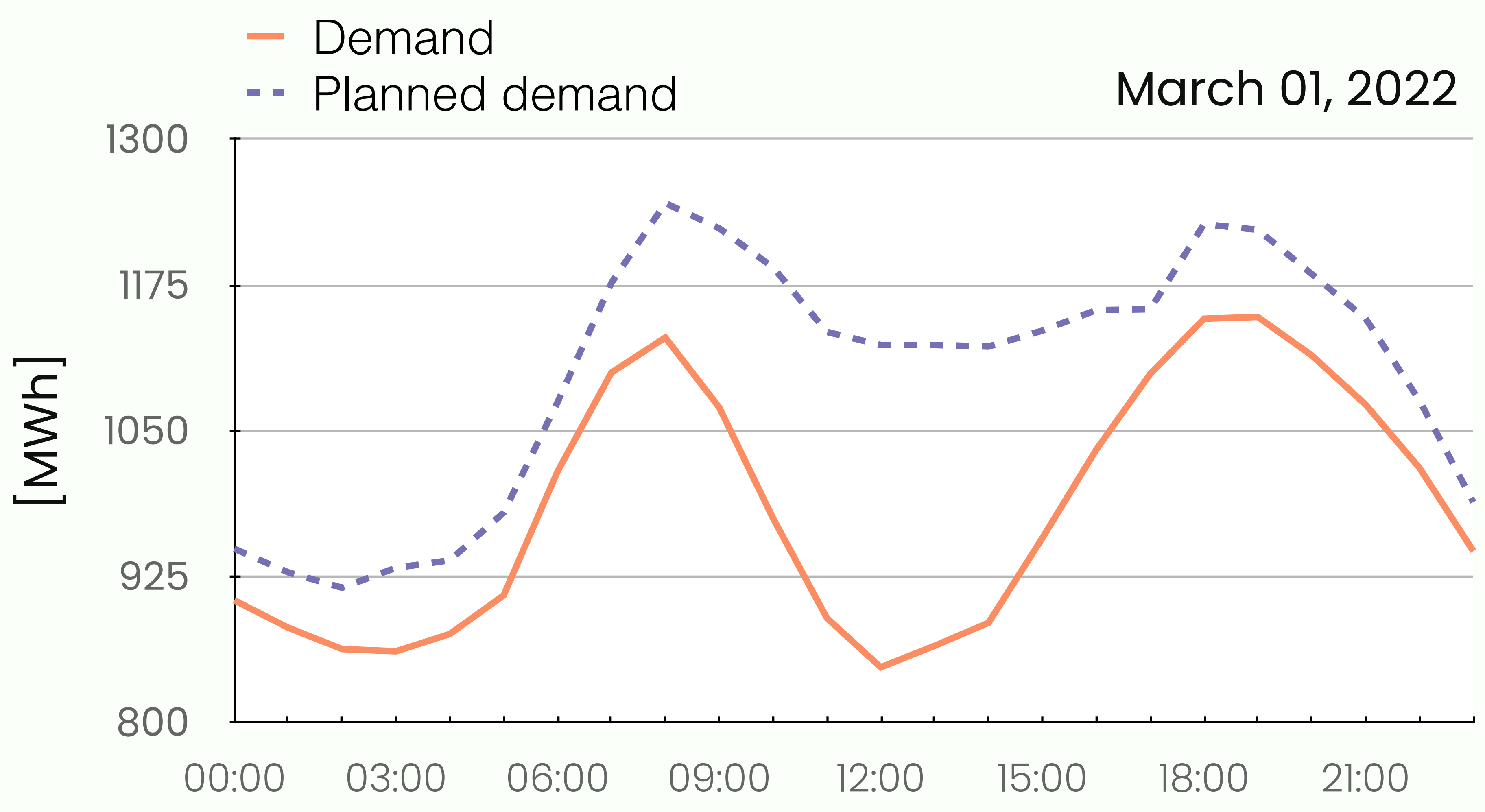
We have a clear winner - MAPE4



MAPE3 = 4.65%

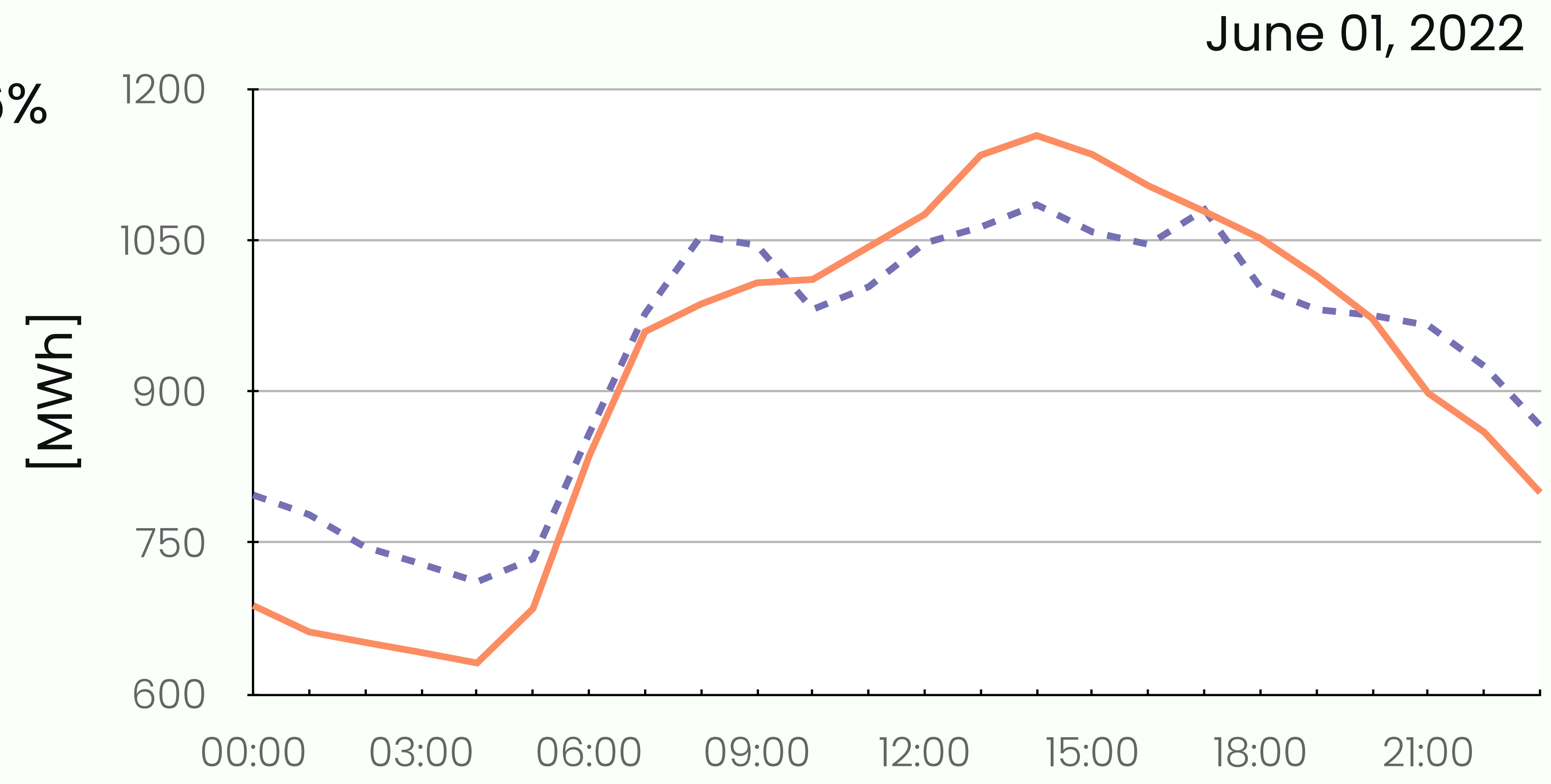
$\boxed{\text{MAPE4 = 3.61\%}}$



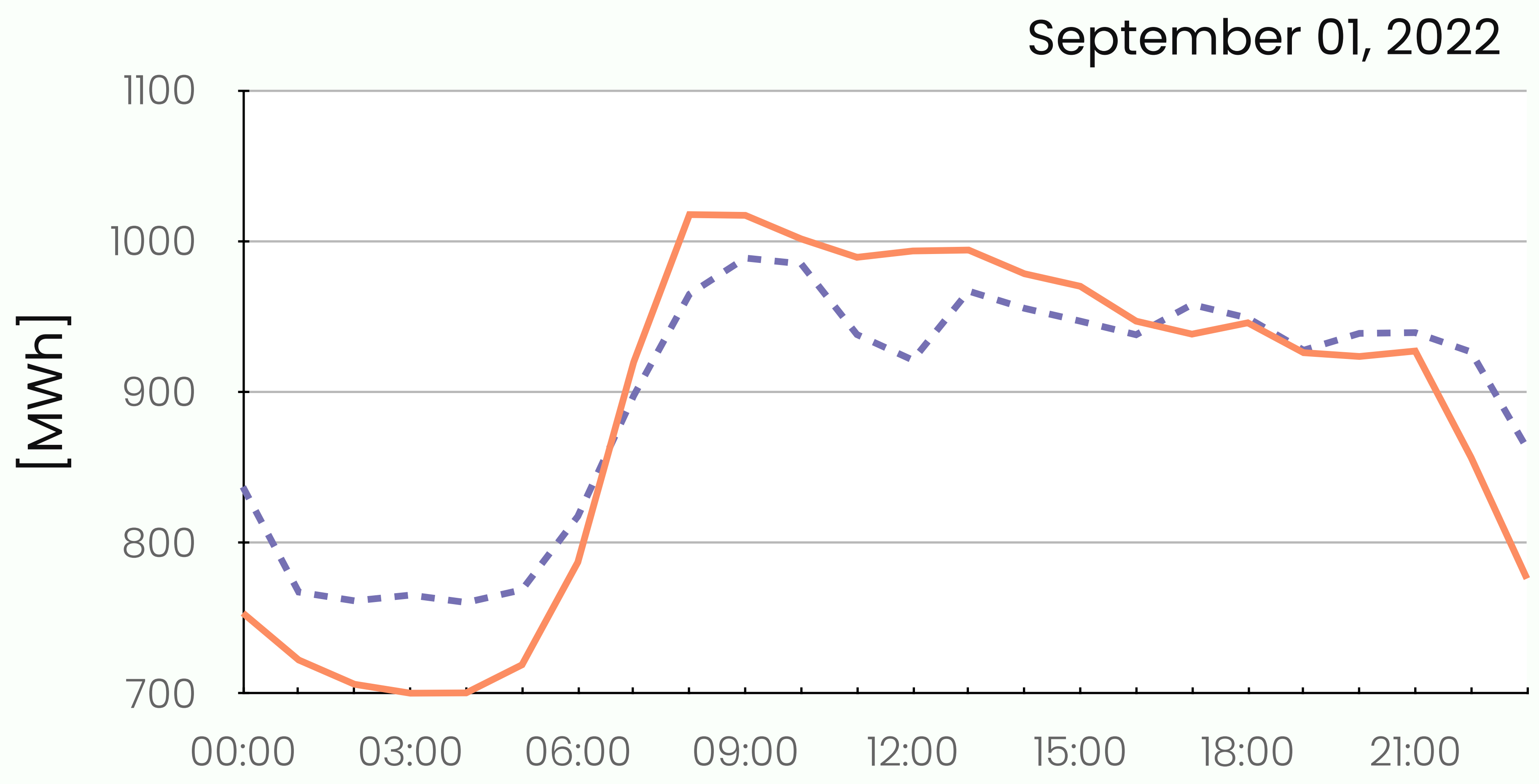


MAPE1 = 12%

MAPE2 = 6.76%



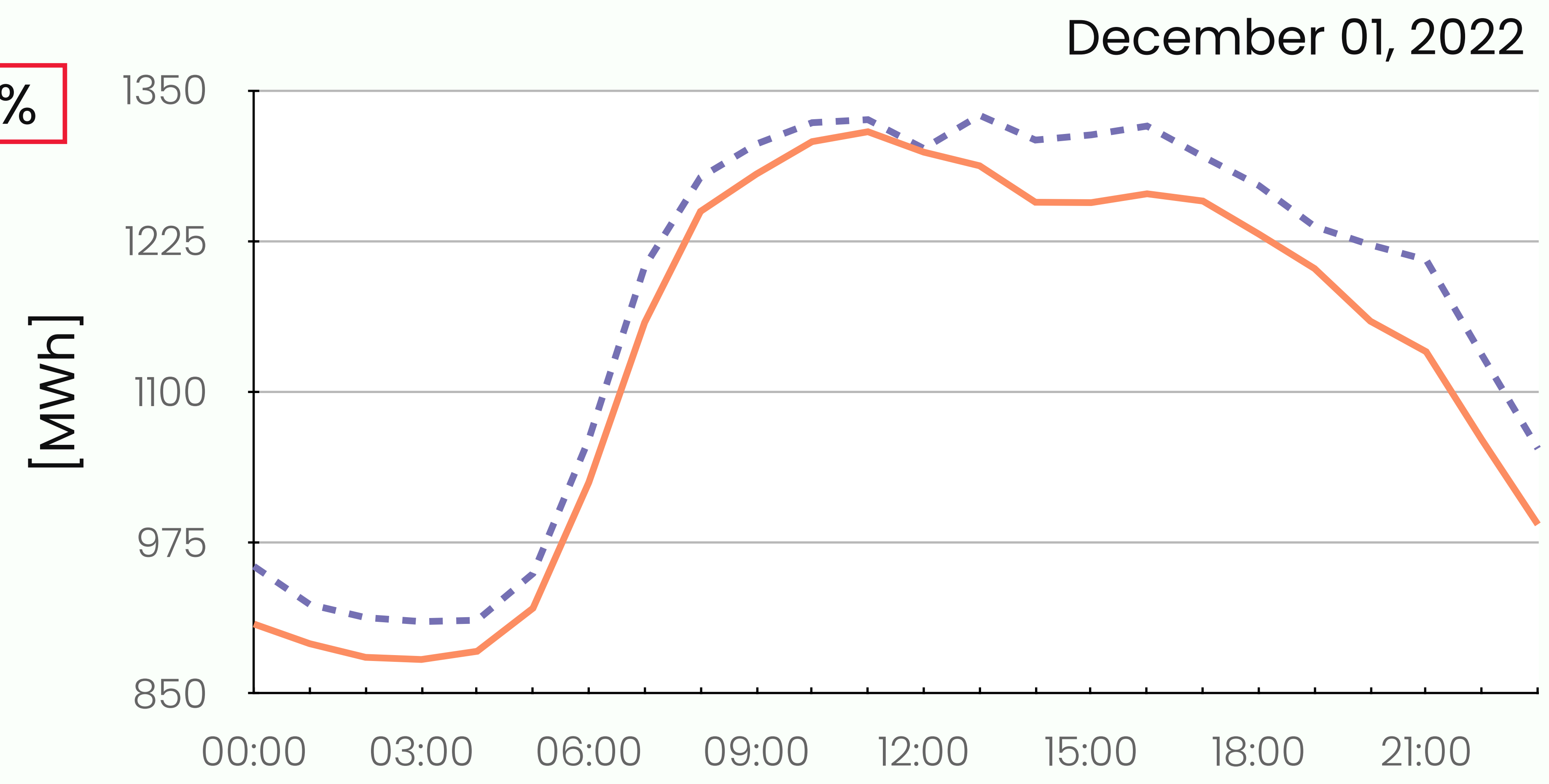
We have a clear winner - MAPE4



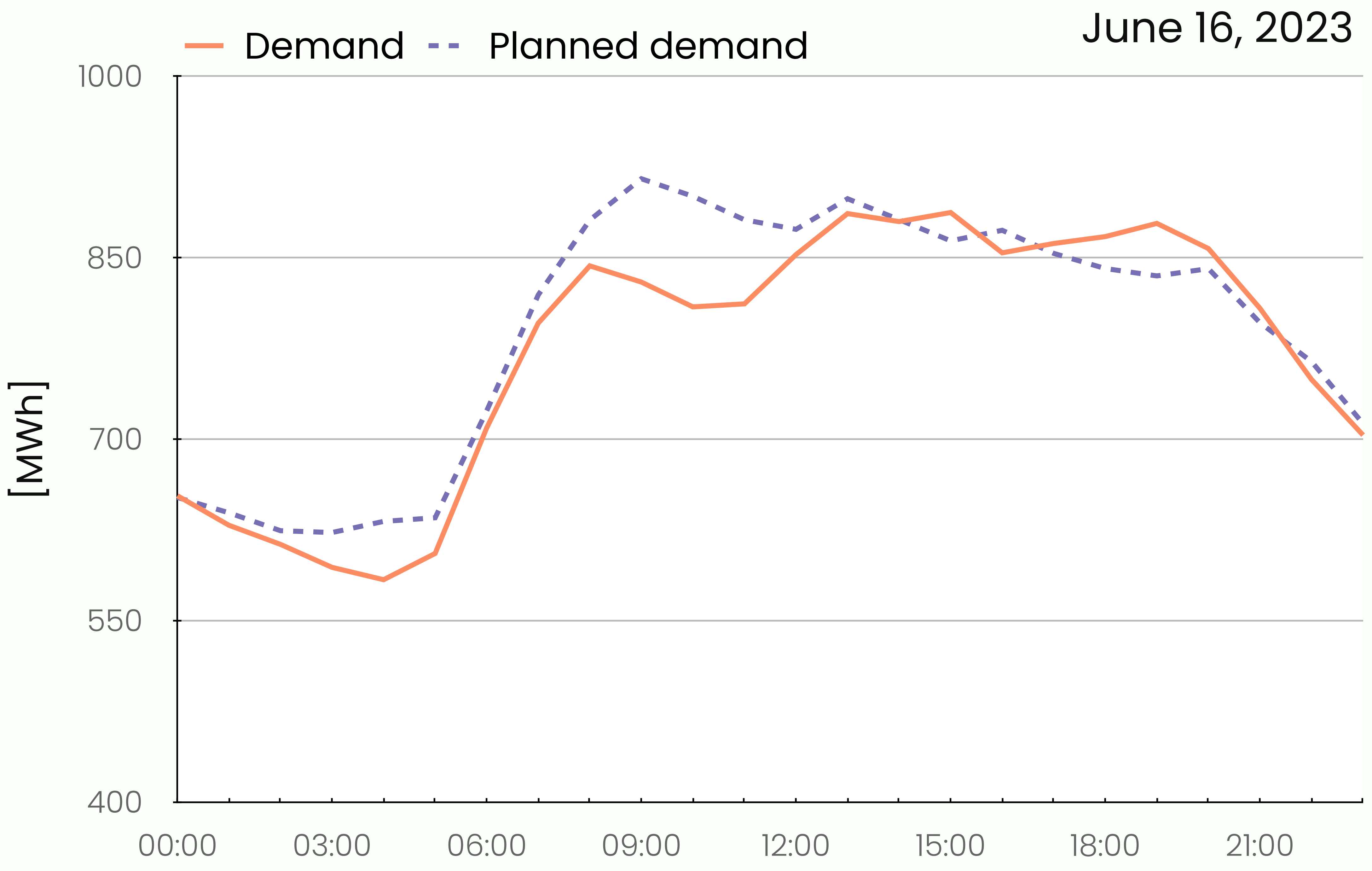
MAPE3 = 4.65%

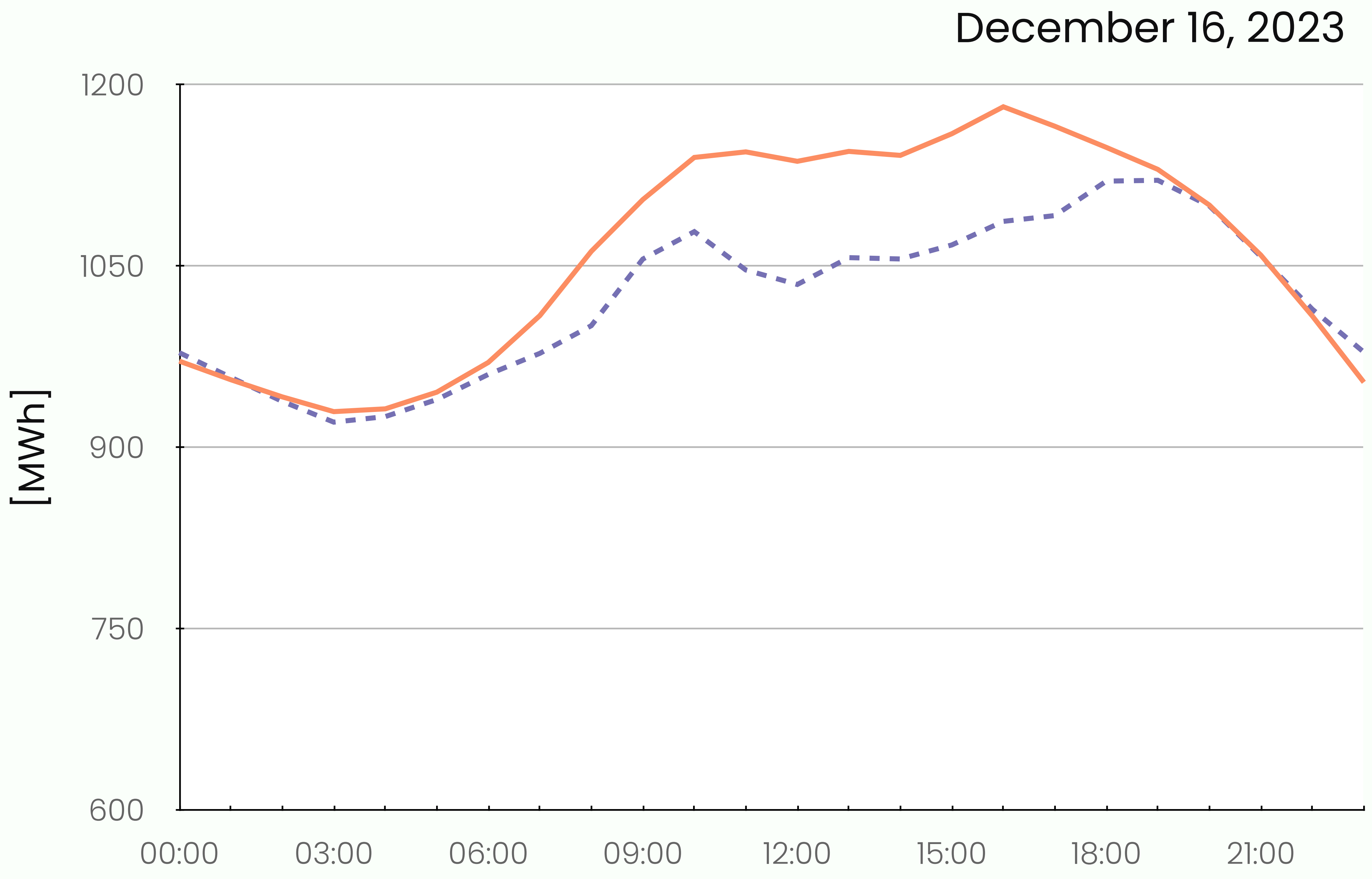
MAPE4 = 3.61%

How about others?

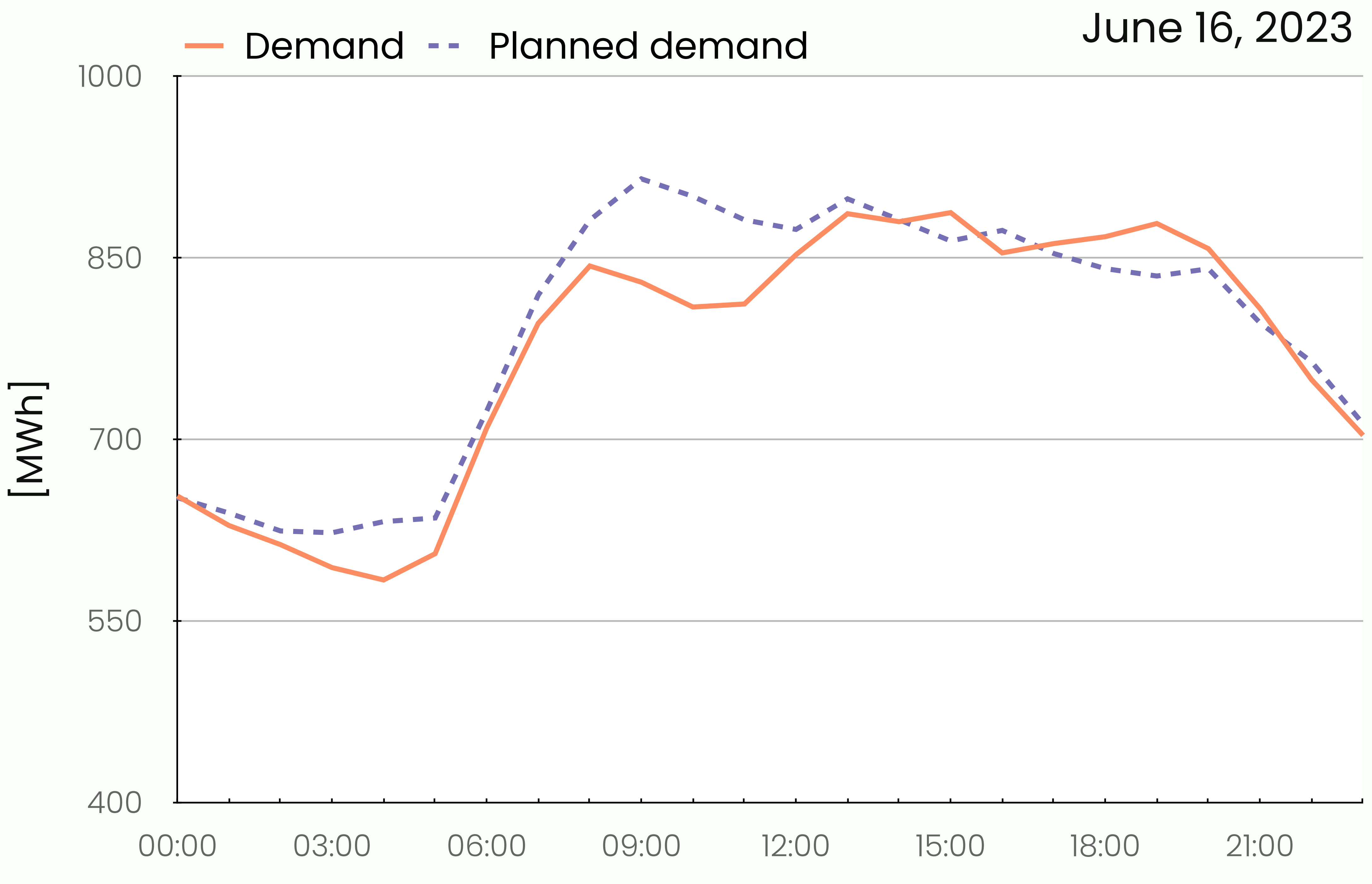


FINAL REMARKS

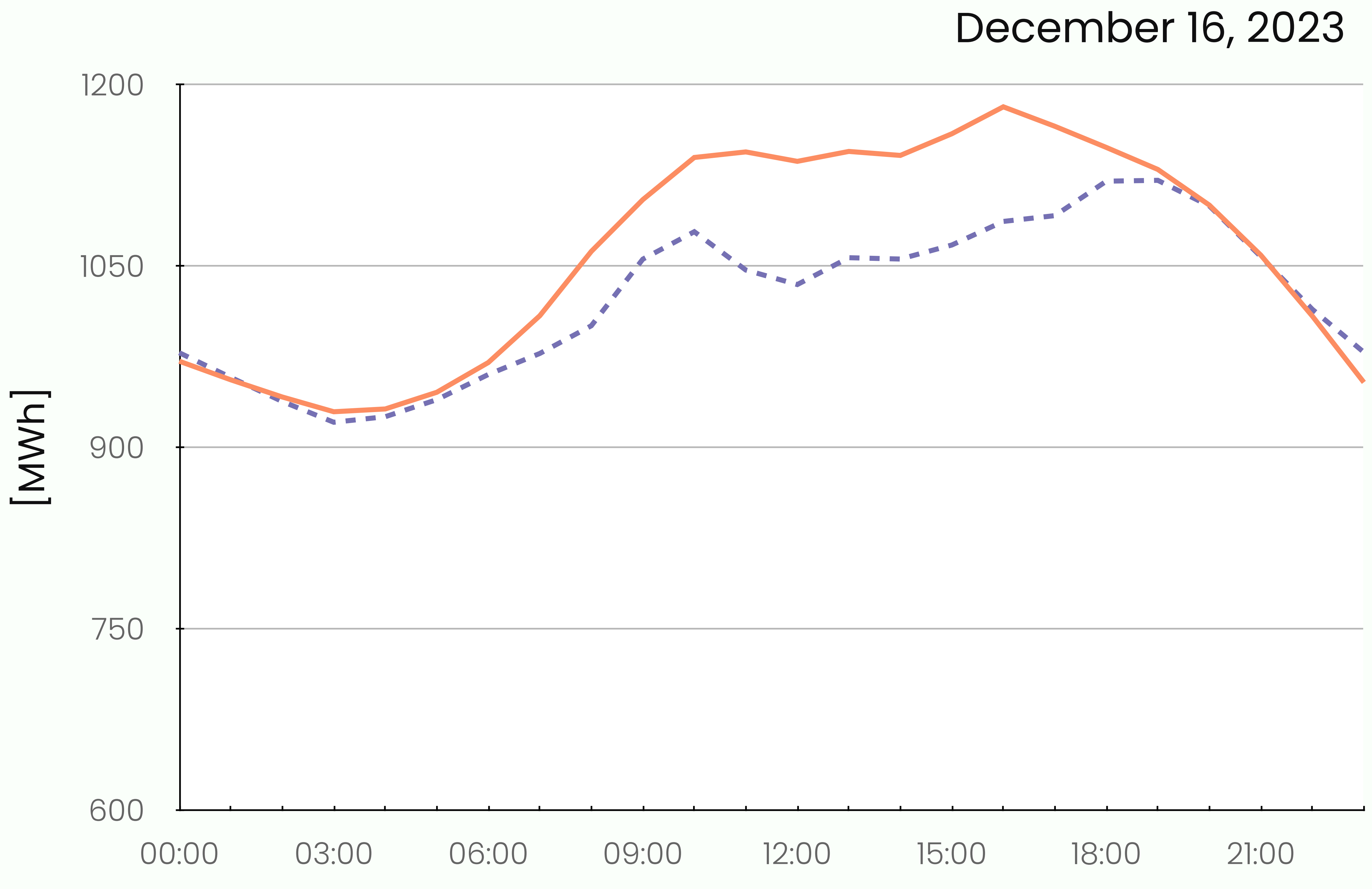




FINAL REMARKS

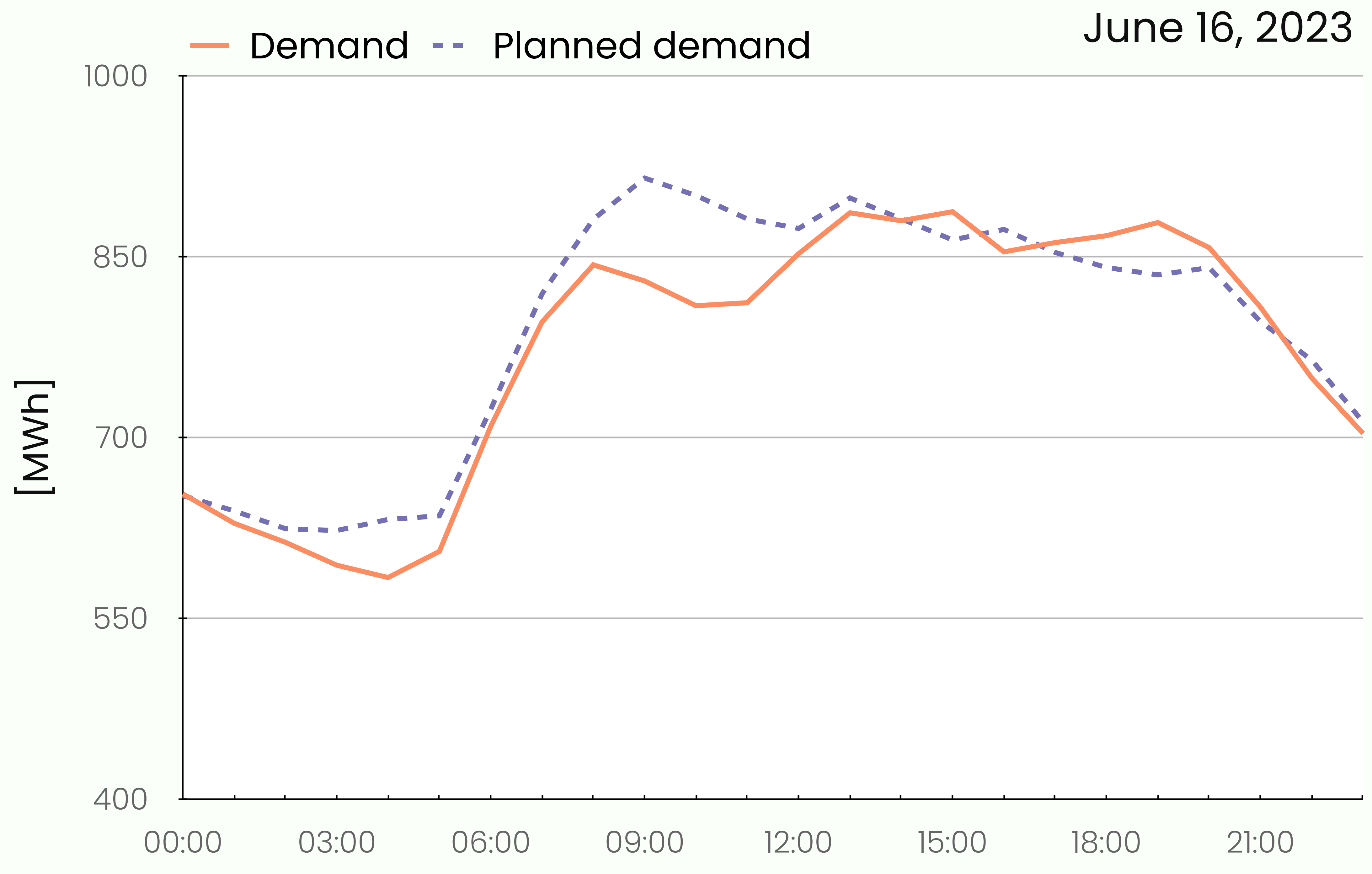


MAPE = 3.57%



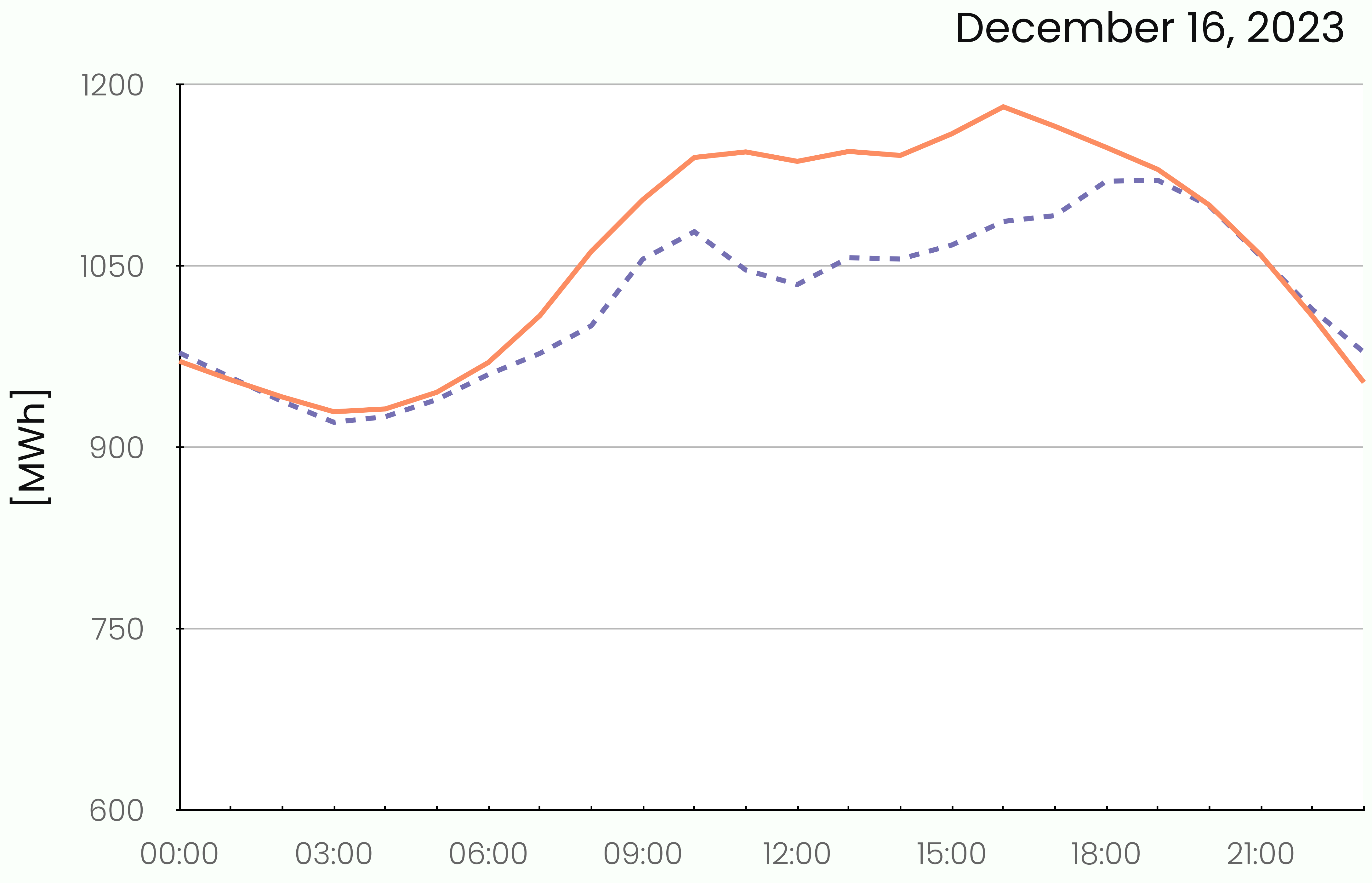
MAPE = 3.54%

FINAL REMARKS



MAPE = 3.57%

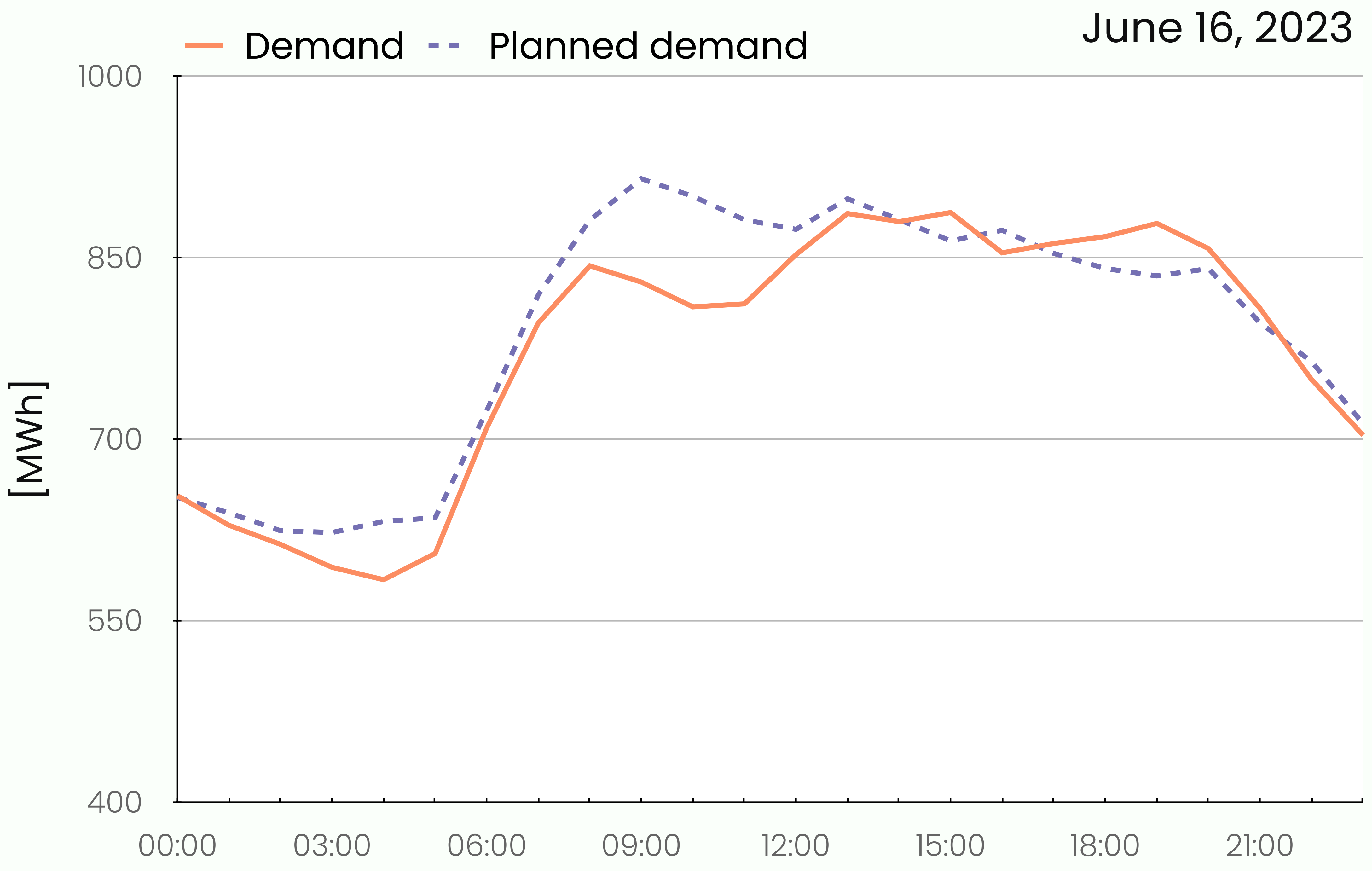
sMAPE = 0.87%



MAPE = 3.54%

sMAPE = 0.91%

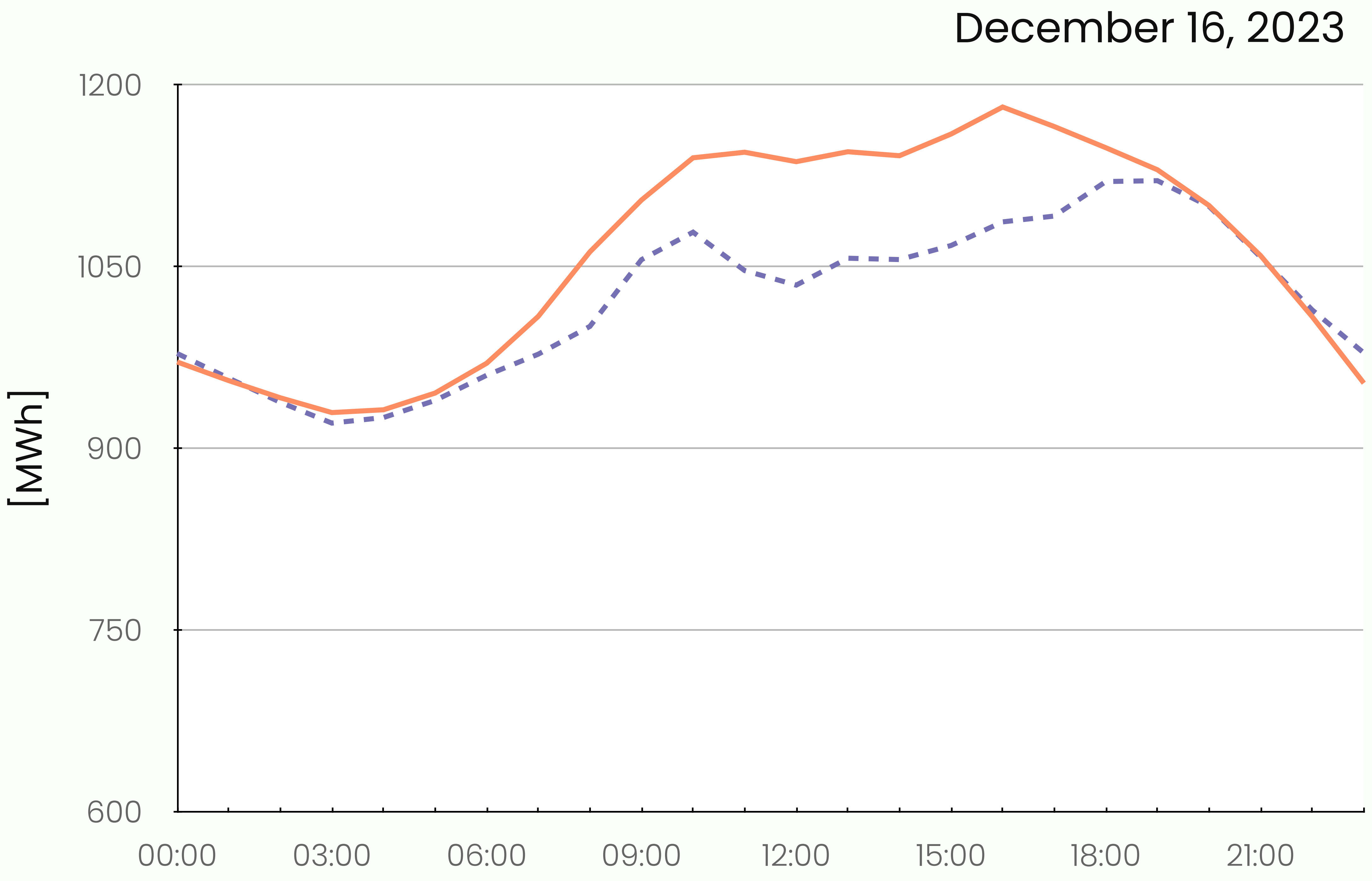
FINAL REMARKS



MAPE = 3.57%

sMAPE = 0.87%

R^2 = 0.91



MAPE = 3.54%

sMAPE = 0.91%

R^2 = 0.84

Thank you!

Questions?