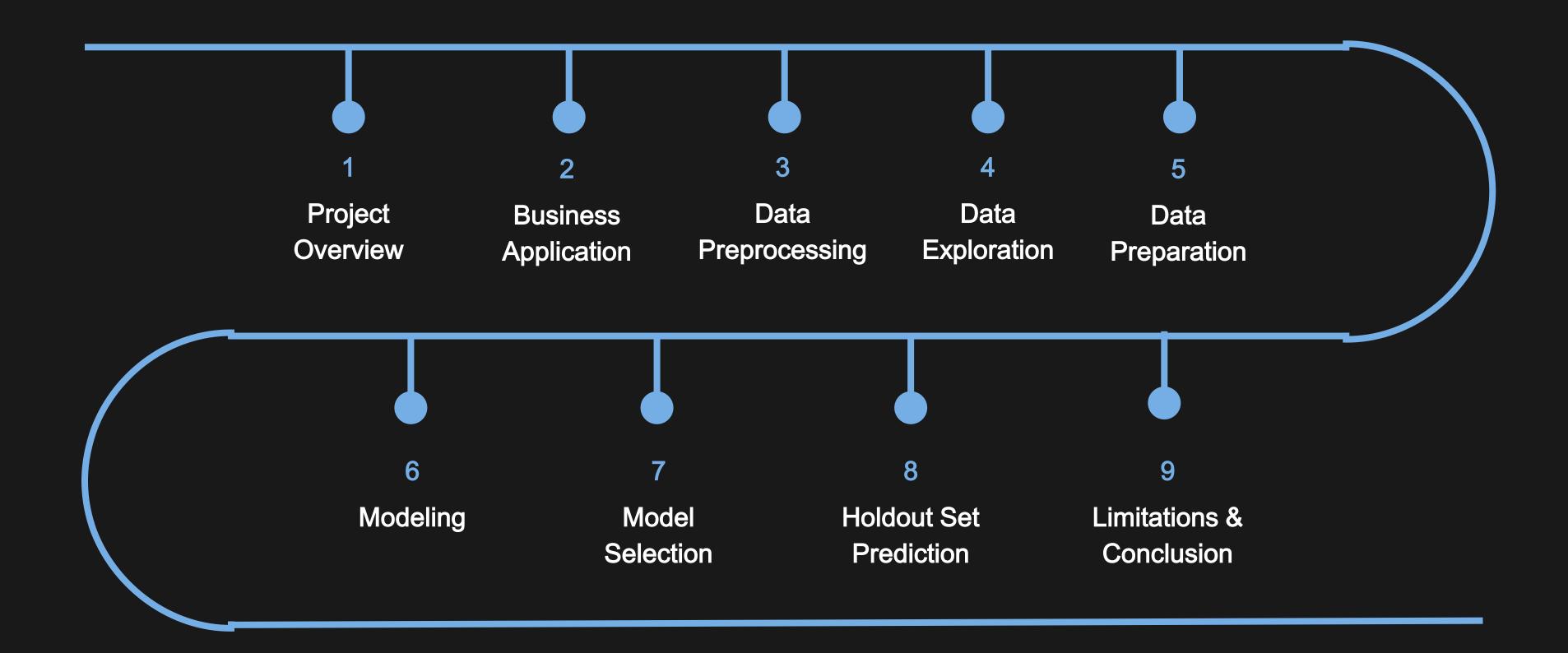
Residential Electricity Usage Prediction



DATA SCIENCE LIFECYCLE





This project focuses on forecasting household electricity usage. The approach combines feature engineering, time series modeling, and an ensemble machine learning strategy to deliver accurate and actionable predictions.

BUSINESS APPLICATION

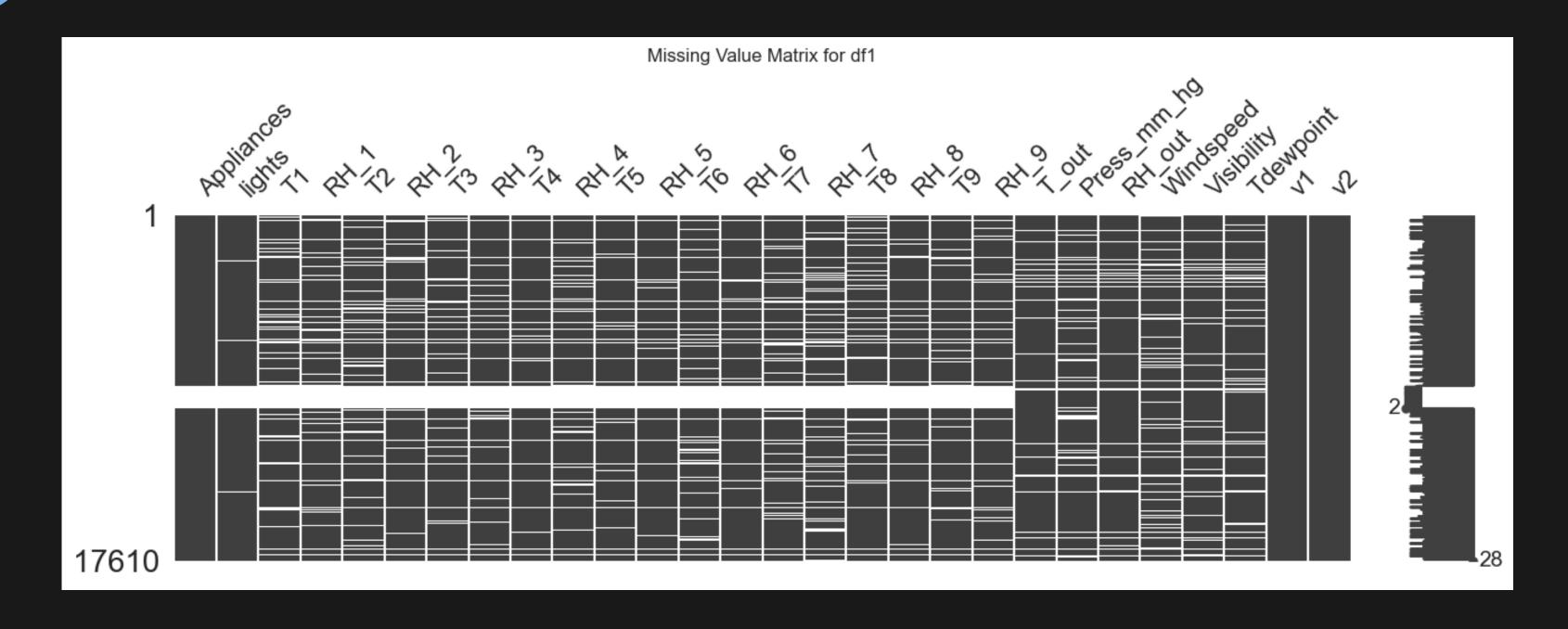
Business Applications:

- Demand Response & Grid Optimization
- Investment Risk Assessment
- Renewable Energy Integration
- Dynamic Pricing
- Personalized Energy Report
- Policy & Planning

Risks of Applying Machine Learning to Time Series Data:

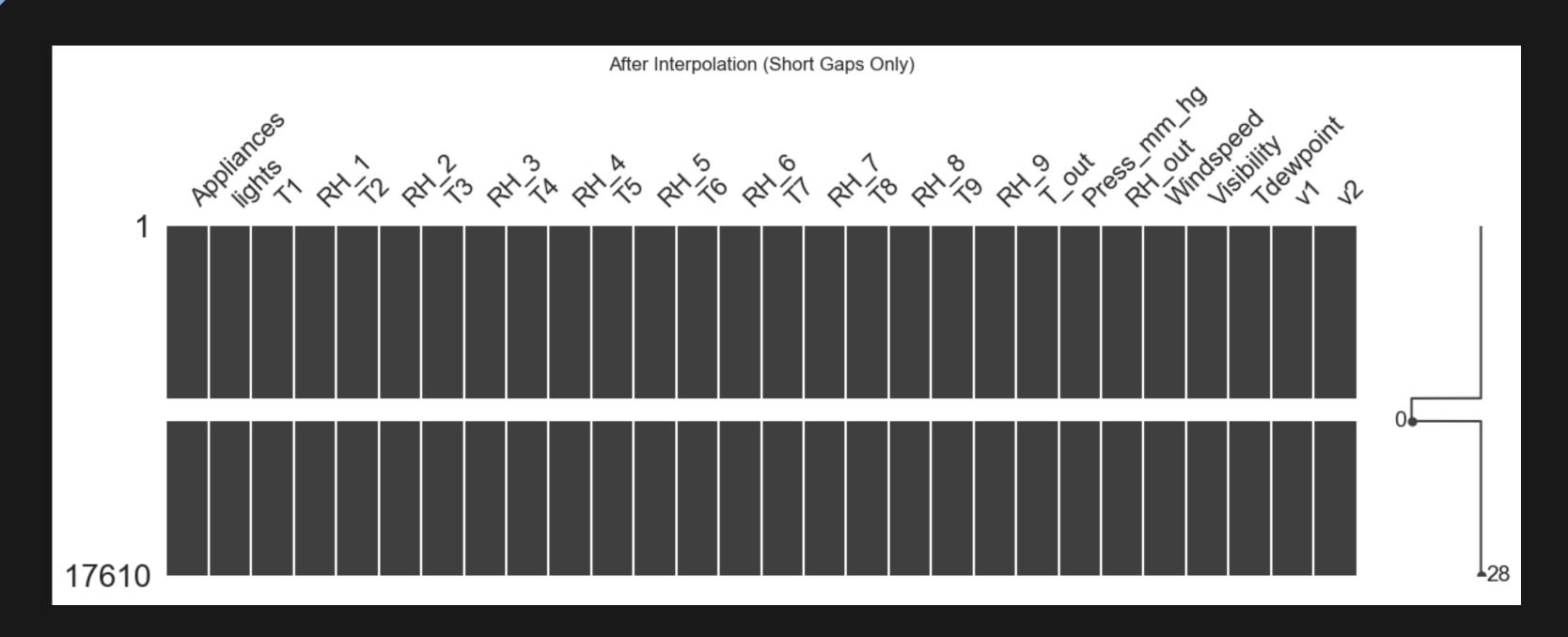
- Temporal Leakage
- Error Accumulation
- Overfitting

DATA PREPROCESSING



- Mark March 12–19 as fully missing to prevent inaccurate lags
- Interpolate short gaps (≤ 1 hour) using linear interpolation.

DATA PREPROCESSING



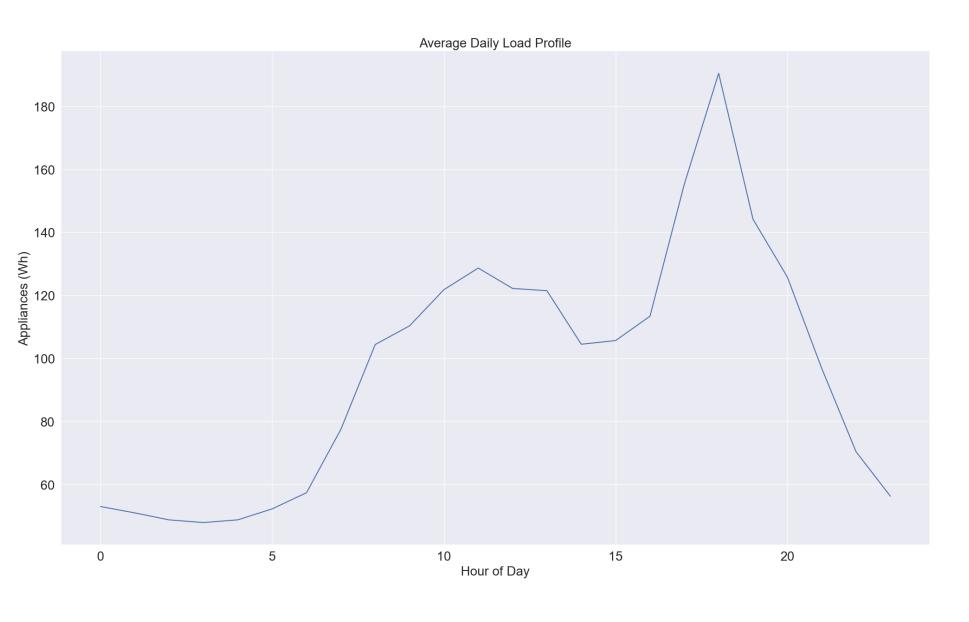
DATA EXPLORATION

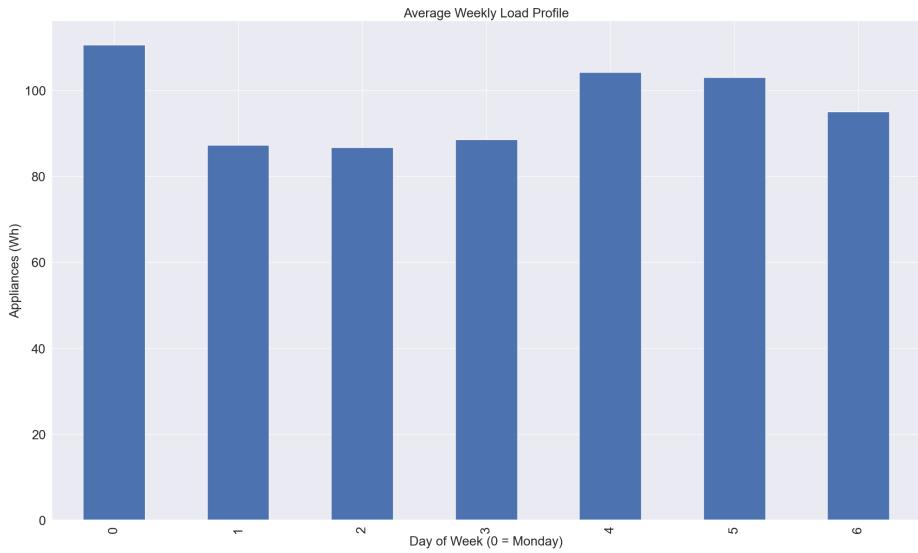
							(Correlat	ion He	atmap:	Appliar	nces vs	. Top V	ariables	S						
Appliances	1.00	-0.04	0.09	-0.04	0.04	0.01	-0.05	-0.06	-0.10	-0.06	-0.12	0.04	0.09	0.07	0.02	0.08	0.03	-0.02	0.07	0.07	0.22
Press_mm_hg	-0.04	1.00	-0.23	-0.23	-0.15	-0.08	-0.03	-0.16	-0.13	-0.11	-0.08	-0.17	-0.11	-0.20	-0.10	-0.12	-0.25	-0.23	-0.11	-0.24	-0.02
RH_1	0.09	-0.23	1.00	0.80	0.84	0.31	0.35	0.78	0.70	0.74	0.29	0.07	0.18	0.12	0.00	0.22	-0.11	-0.00	0.26	0.26	0.15
RH_2	-0.04	-0.23	0.80	1.00	0.66	0.25	0.43	0.69	0.65	0.65	0.58	-0.07	-0.24	0.07	-0.11	-0.05	-0.08	-0.01	-0.00	0.12	0.08
RH_3	0.04	-0.15	0.84	0.66	1.00	0.38	0.58	0.83	0.82	0.84	0.35	-0.06	0.10	-0.10	-0.18	0.02	-0.32	-0.27	0.07	0.30	0.16
RH_5	0.01	-0.08	0.31	0.25	0.38	1.00	0.28	0.32	0.36	0.27	0.20	-0.01	0.03	-0.09	-0.08	-0.11	-0.07	-0.15	-0.08	0.08	0.16
RH_6	-0.05	-0.03	0.35	0.43	0.58	0.28	1.00	0.45	0.56	0.45	0.73	-0.56	-0.51	-0.63	-0.67	-0.64	-0.64	-0.73	-0.61	0.11	0.17
RH_7	-0.06	-0.16	0.78	0.69	0.83	0.32	0.45	1.00	0.88	0.86	0.40	0.07	0.16	0.05	-0.04	0.18	-0.19	-0.07	0.23	0.27	0.08
RH_8	-0.10	-0.13	0.70	0.65	0.82	0.36	0.56	0.88	1.00	0.85	0.51	-0.07	0.01	-0.06	-0.16	0.00	-0.25	-0.20	0.05	0.25	0.04
RH_9	-0.06	-0.11	0.74	0.65	0.84	0.27	0.45	0.86	0.85	1.00	0.35	0.03	0.12	0.05	-0.09	0.14	-0.21	-0.09	0.18	0.30	0.02
RH_out	-0.12	-0.08	0.29	0.58	0.35	0.20	0.73	0.40	0.51	0.35	1.00	-0.37	-0.53	-0.32	-0.42	-0.60	-0.32	-0.37	-0.60	-0.15	0.10
T1	0.04	-0.17	0.07	-0.07	-0.06	-0.01	-0.56	0.07	-0.07	0.03	-0.37	1.00	0.83	0.87	0.84	0.61	0.80	0.80	0.64	-0.07	0.00
T2	0.09	-0.11	0.18	-0.24	0.10	0.03	-0.51	0.16	0.01	0.12	-0.53	0.83	1.00	0.68	0.72	0.75	0.54	0.62	0.75	0.07	0.02
T3	0.07	-0.20	0.12	0.07	-0.10	-0.09	-0.63	0.05	-0.06	0.05	-0.32	0.87	0.68	1.00	0.82	0.63	0.80	0.89	0.65	-0.07	-0.09
T4	0.02	-0.10	0.00	-0.11	-0.18	-0.08	-0.67	-0.04	-0.16	-0.09	-0.42	0.84	0.72	0.82	1.00	0.60	0.76	0.85	0.62	-0.18	0.02
T6	0.08	-0.12	0.22	-0.05	0.02	-0.11	-0.64	0.18	0.00	0.14	-0.60	0.61	0.75	0.63	0.60	1.00	0.44	0.63	0.97	0.21	-0.07
Т8				-0.08									0.54			0.44		0.86		-0.22	
Т9				-0.01											0.85			1.00		-0.16	
T_out				-0.00										0.65				0.64		0.23	
Windspeed				0.12																1.00	
lights				0.12																	1.00
ligitis		-0.02	0.15	0.00				0.00													
	Appliances	ess_mm_hg	RH_1	RH_2	RH_3	RH_5	RH_6	RH_7	RH_8	RH_9	RH_out	T	72	T3	T	T6	T8	T9	T_out	Windspeed	lights

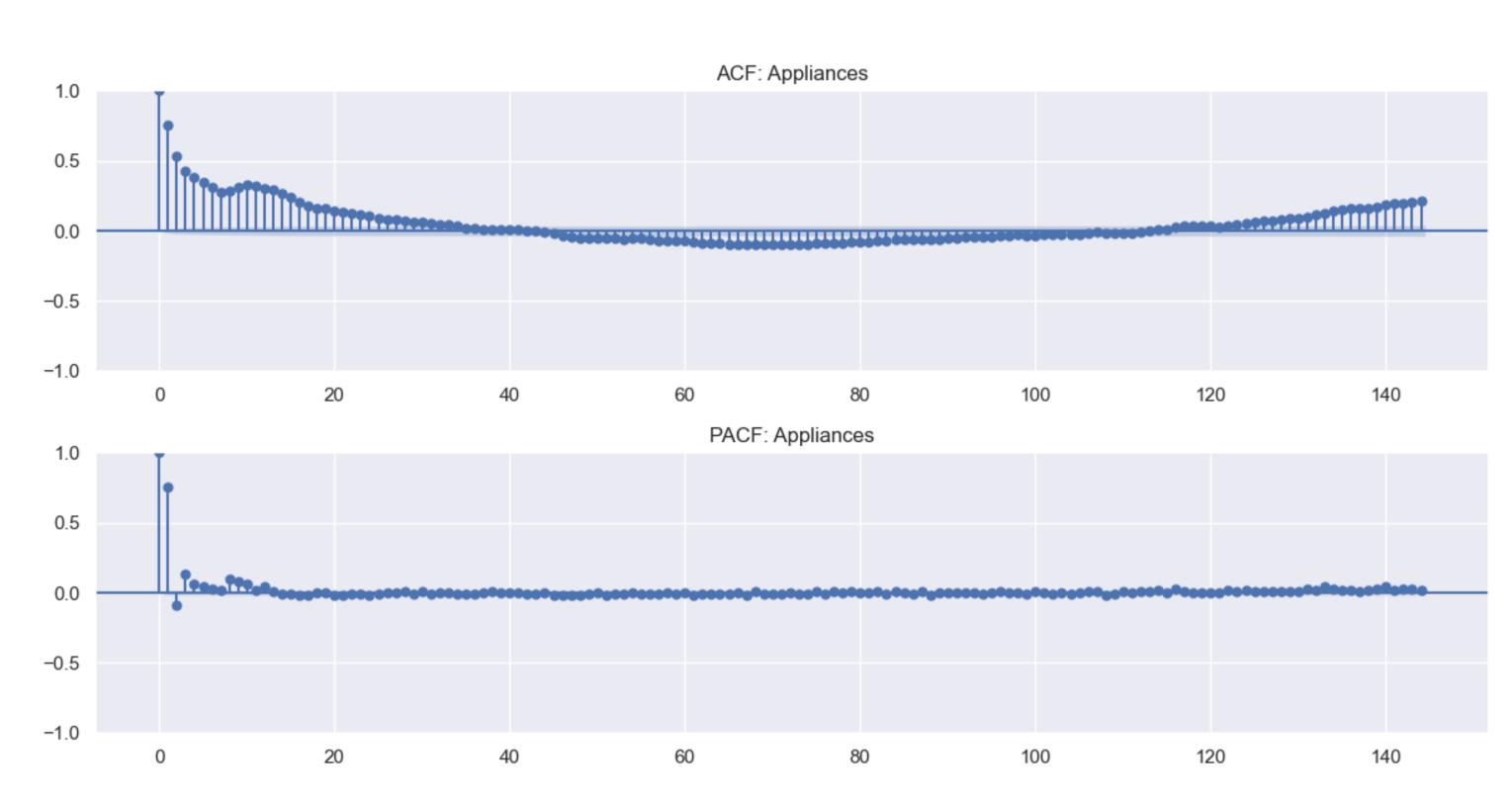
- 0.2

- 0.0

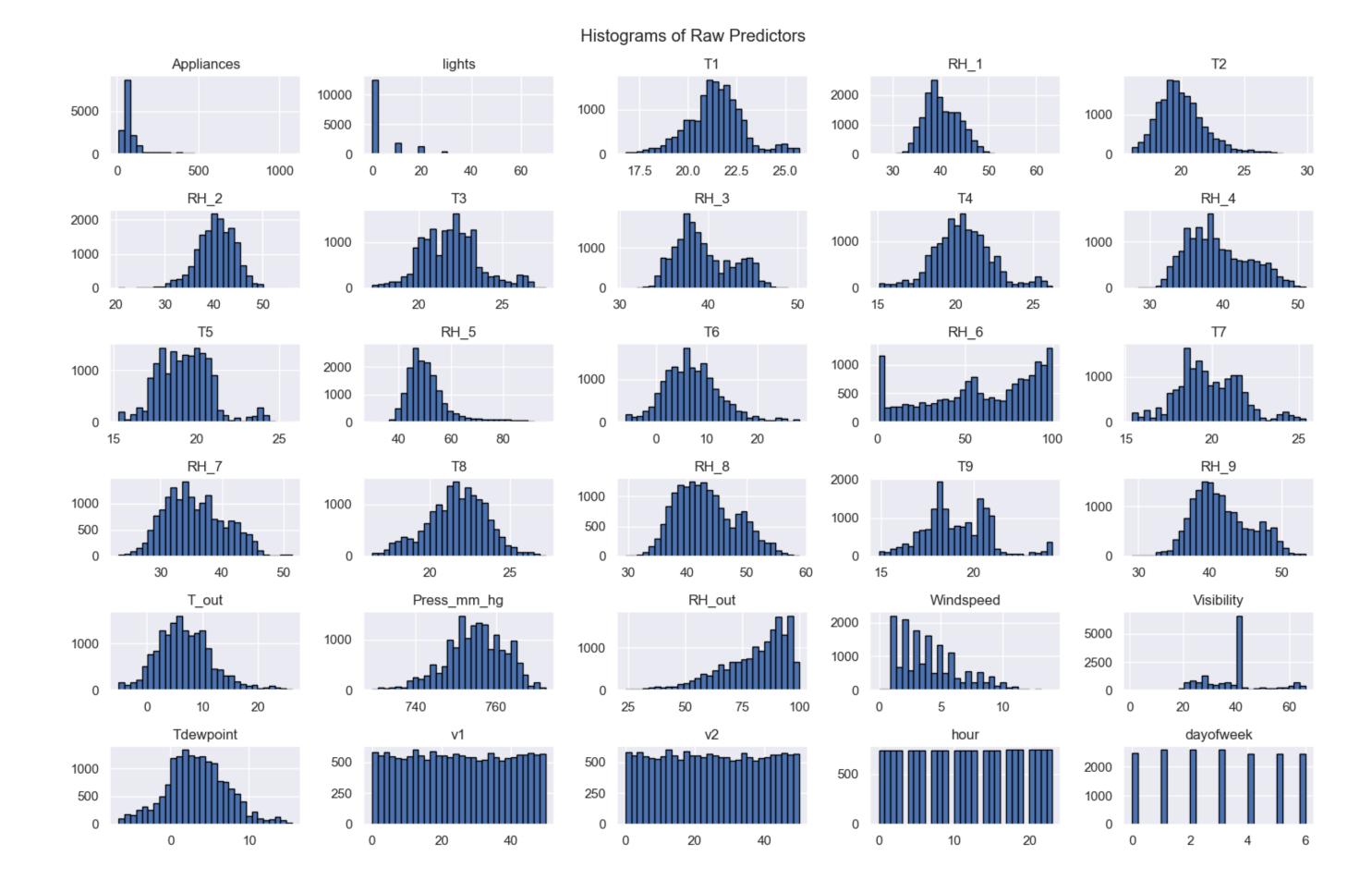
DATA EXPLORATION



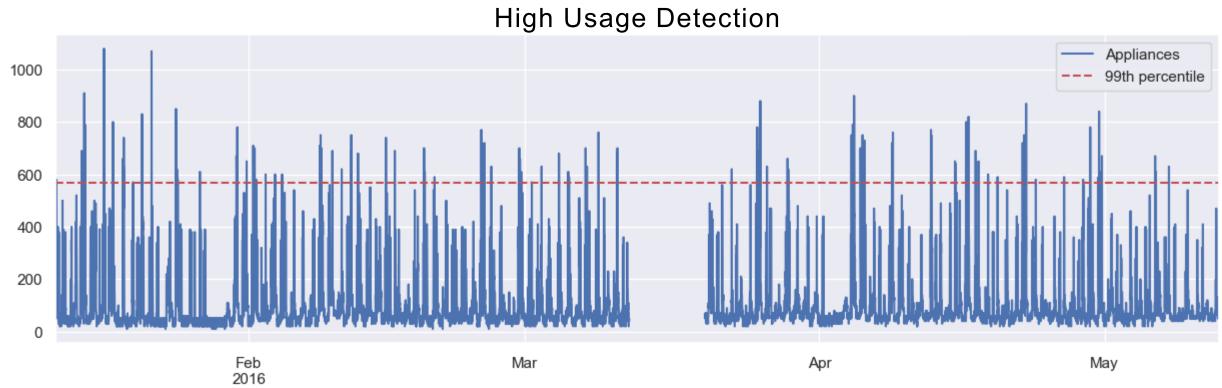


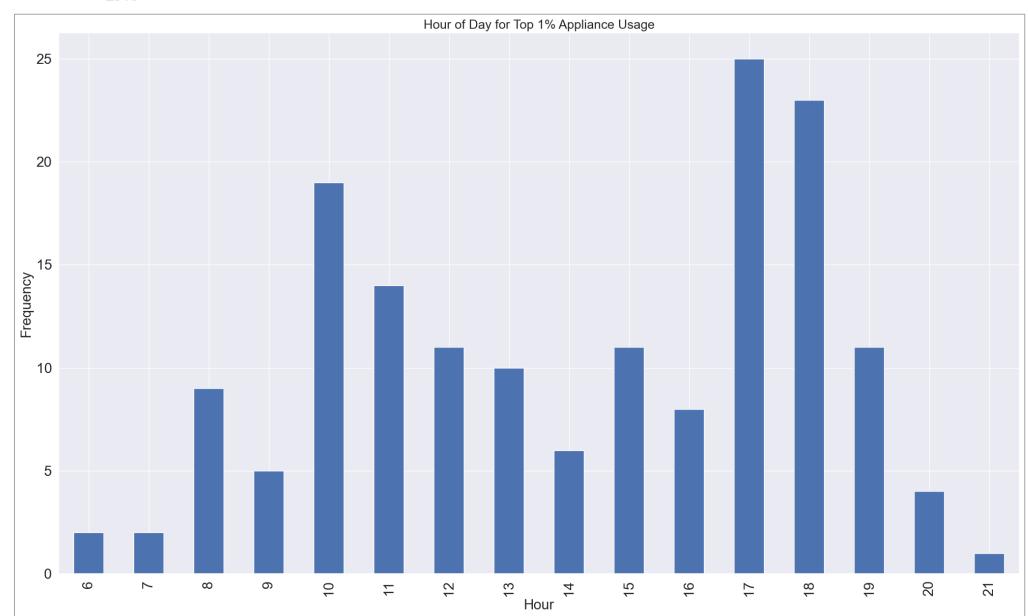


DATA EXPLORATION



DATA EXPLORATION - OUTLIER





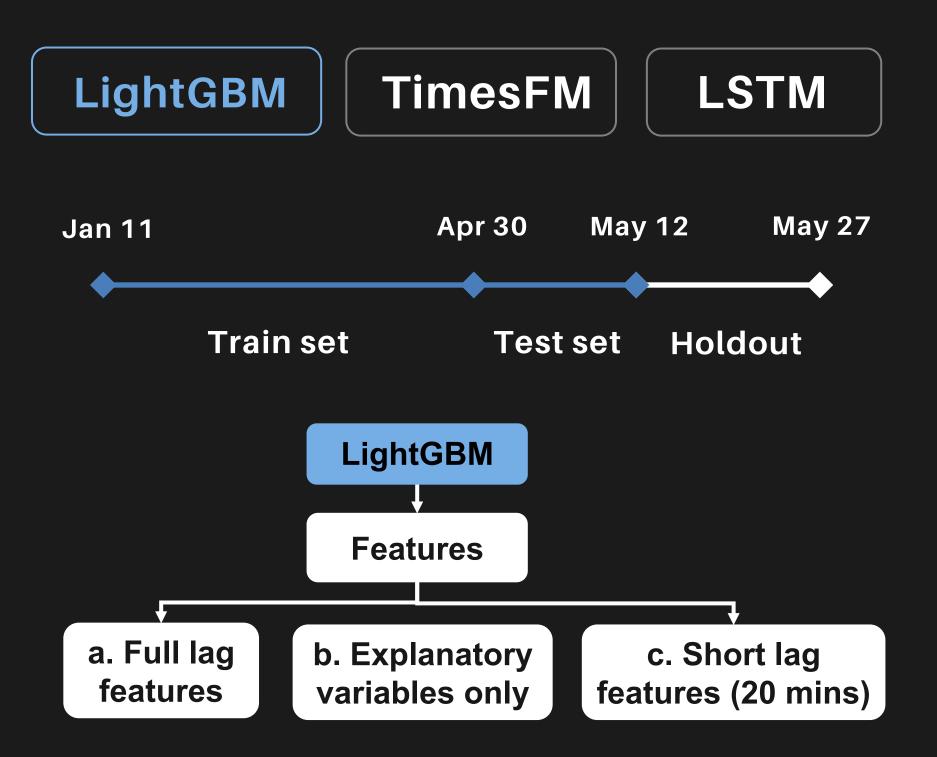
DATA PREPARATION

KEY FEATURES

- External variables (temperature, humidity, windspeed, etc.)
- Time-based features (hour, day of week, weekend indicator, peak hour indicator, hour_sin, hour_cos)
- Interaction features (e.g., temperature × hour, windspeed × weekend)
- Rolling and lag features on key predictors and target

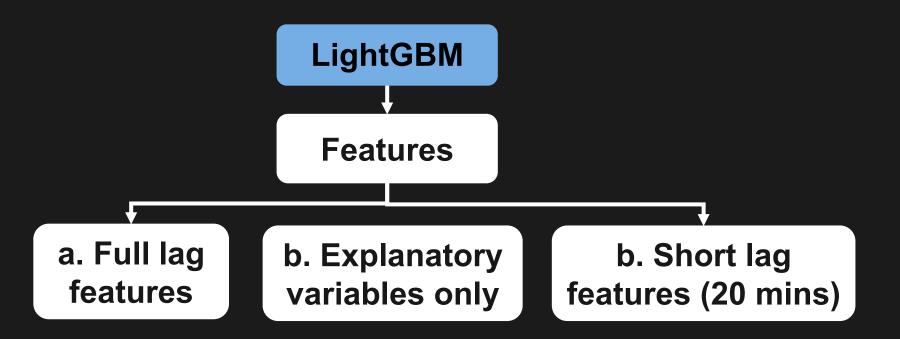
	FEATURE TYPE	TIME WINDOWS			
Appliances	Lag features (to capture autoregressive effects)	1 to 144 steps, i.e., 10 mins to 1 day			
	Rolling statistics (shifted to avoid data leakage)	over 1-hour and 3- hour windows			
(T1, RH_1,	Lag features	1, 6 steps, i.e., 10- 60 mins			
T_out, RH_out etc.)	Rolling statistics	over 1-hour and 3- hour windows			

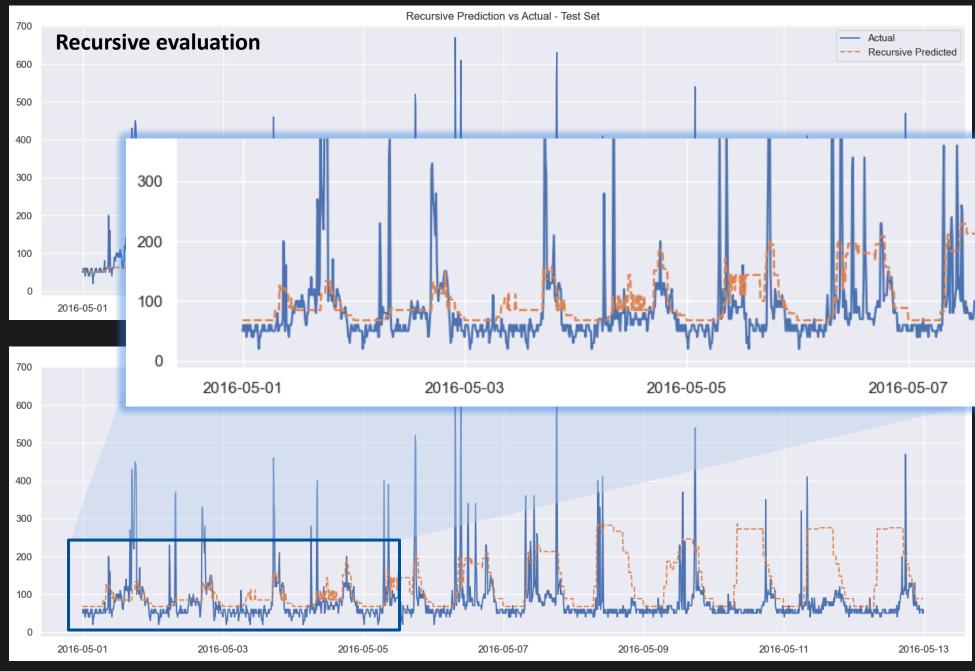
6 MODELLING

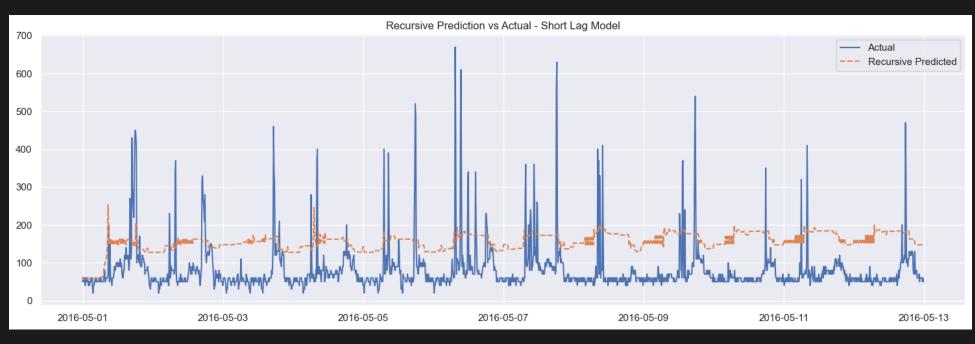


6 MODELLING

	Test MAE	Test RMSE
a. Full lag features	41.195	72.869
b. Explanatory variables only	65.805	100.258
b. Short lag features (20 mins)	89.43	98.62

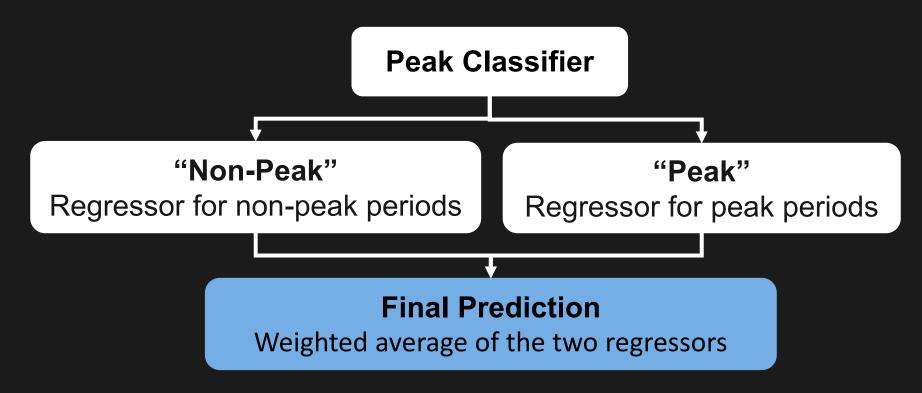




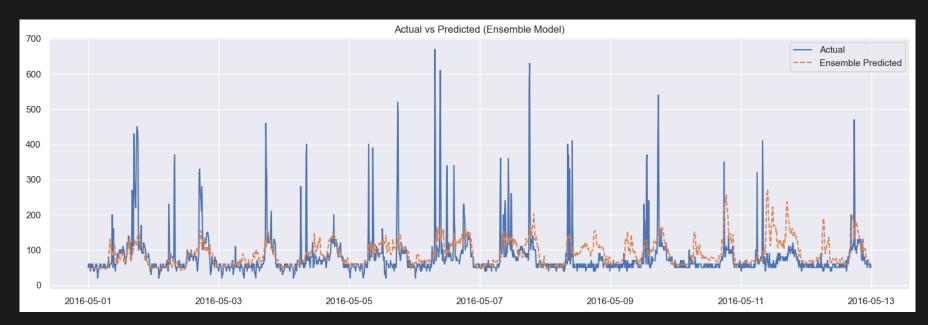


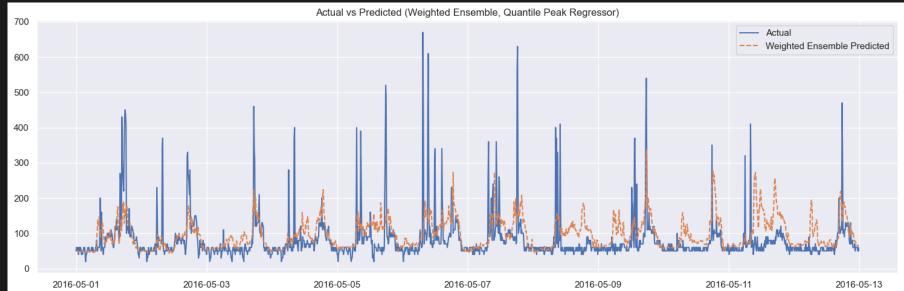
MODEL SELECTION

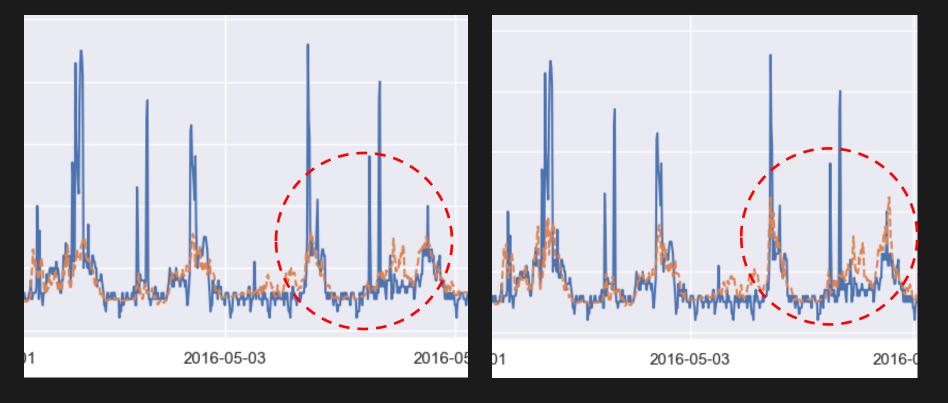
Ensemble approach (only on explanatory variables)



	Test MAE	Test RMSE
a. Default	34.42	63.50
b. Quantile Regression	37.34	64.22
b. Higher quantile (alpha=0.98)	39.91	64.86

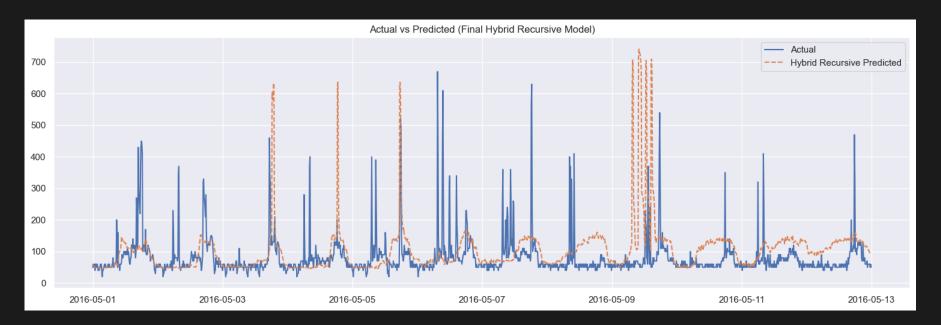


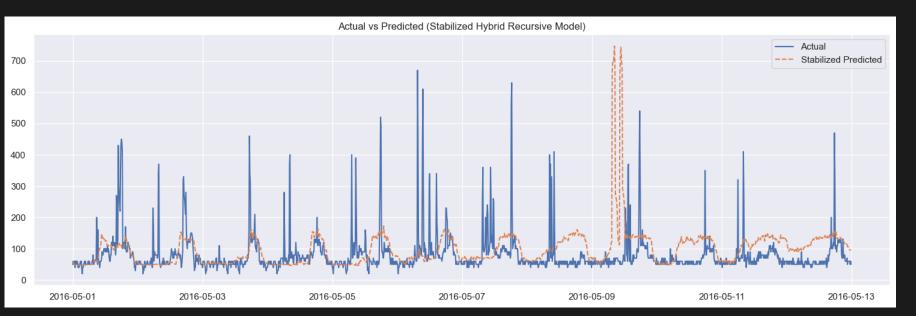


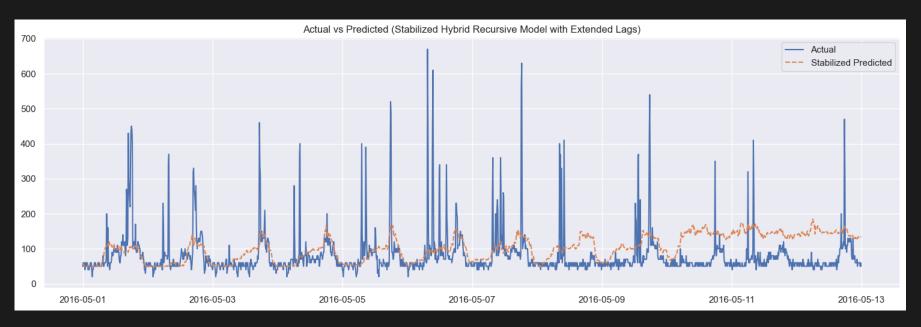


MODEL SELECTION

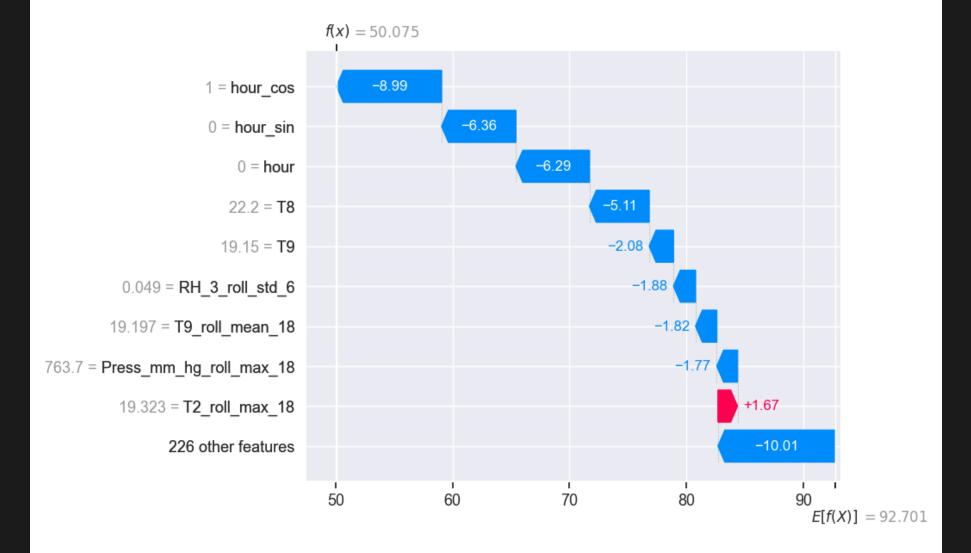
Trials: Appliance lags added back



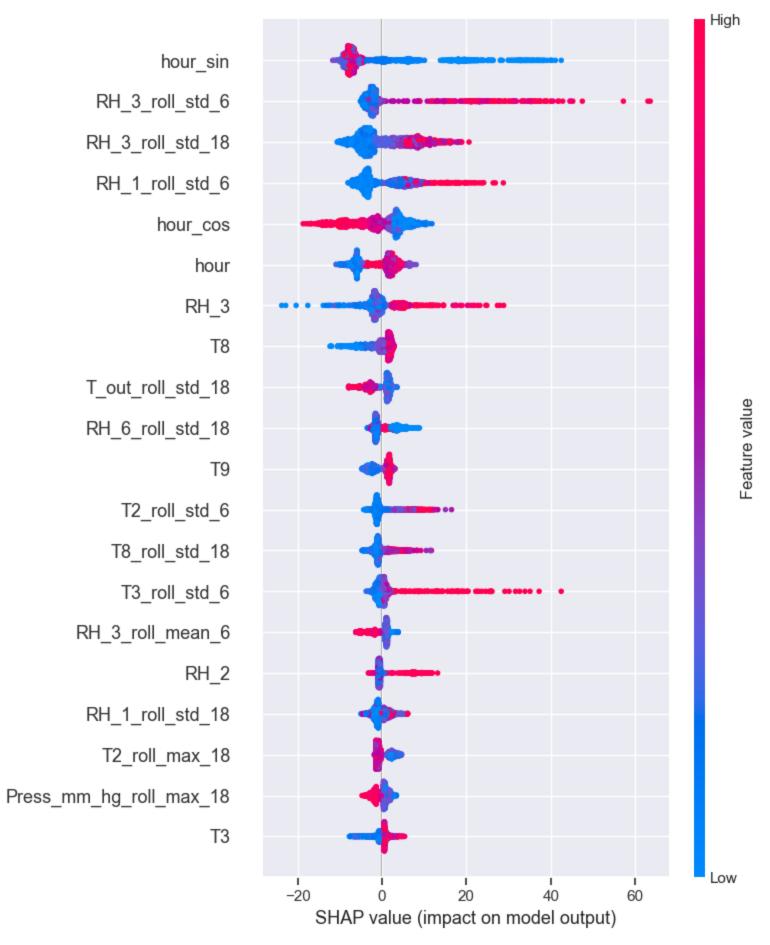


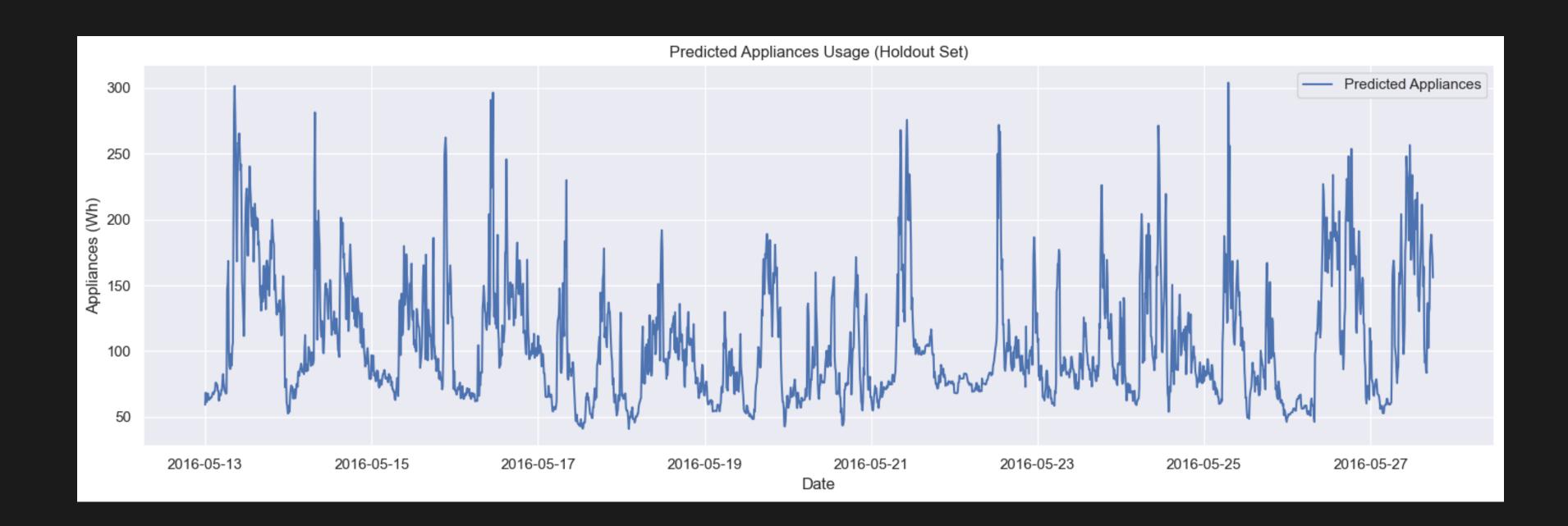


SHAP Waterfall Plot



SHAP Summary Plot (Beeswarm)





9 LIMITATIONS & CONCLUSION

- For high-demand periods where actual usage exceeded 500-700 Wh, model predictions remained significantly lower.
- Post-prediction adjustment could be done
- Alternative architectures (e.g., sequence models further improvement on TimesFM and LSTM)

KEY TAKEAWAYS:

- Recursive evaluation is critical for honest assessment of time series models using lagged targets.
- Ensemble approaches (with peak-aware logic) can improve both accuracy and robustness.
- Consistent feature engineering and careful handling of missing data are essential for reliable deployment.

Thank you

