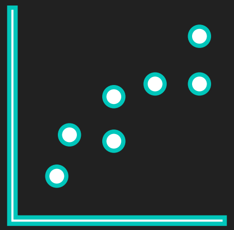
Electricity Consumption Forecasting and Pattern Recognition



AGENDA

- Motivation
- Background
- Literature Review
- Methodology
- Experiments and Results
- Discussion
- Future Scope for Improvement

Motivation

- California targets 60 percent renewable energy by 2030 and commits to a 100 percent zero-carbon energy supply by 2045
- Foreseeing the electricity requirements will enable precedent accommodations of transition to a sustainable future
- PG&E open data program

Background

AIM: To be able to forecast electricity consumption for various zip codes in California

- Look for various exogenous factors influencing the consumption pattern like seasons, weekends, weekdays, electricity prices, etc.
- Among different sectors of electricity consumption like Residential, Commercial, and Agricultural, we target the Residential sector
- Apply clustering to bring together various zip codes of similar electricity consumption patterns
- Employ Time Series Analysis to forecast at a time-step of a month over univariate, bivariate, and exogenous variables

Literature Review

Authors	Title	Data	Approach	Results/ Observation
González-Briones et al., (2019)	Machine learning models for electricity consumption forecasting: A Review	one-year data set of a shoe store	SVR, LR with Day of the week, Day of the year, Week, Weekend, Prev. day electricity consumption	85.7% accuracy
Alam et al., (2021)	Analyzing energy consumption patterns of an educational building through data mining	Building Management System Data from 64 submeters of the building	K-means Clustering	Energy consumption rate depends on the climate and days of the week (weekdays or weekend).
Syed et al., (2021)	Deep Learning-Based Short-Term Load Forecasting Approach in Smart Grid With Clustering and Consumption Pattern Recognition	Electric transformers-based data	k-Medoid Clustering + DNN	RMSE, MAPE, training time, and testing time with an improved 44 % training time

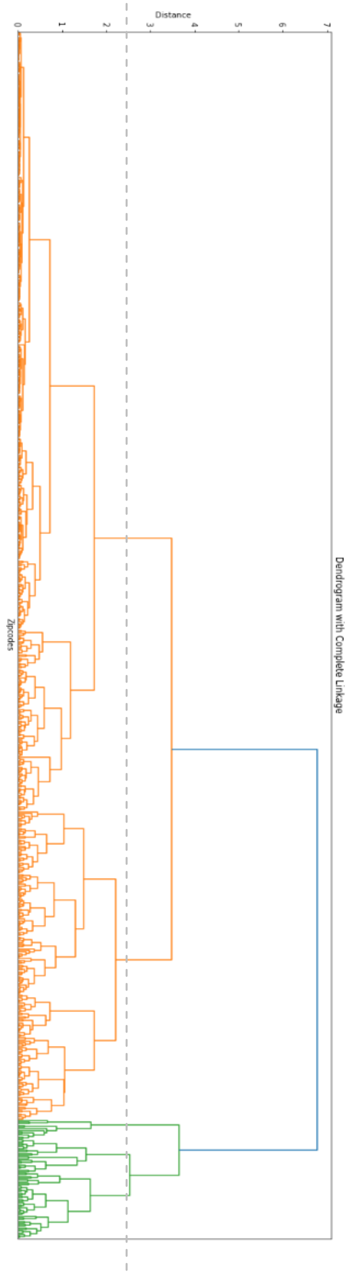
Datasets and Analysis

- PG&E Electricity consumption for 812 zip codes from 2013 to 2022 for every month

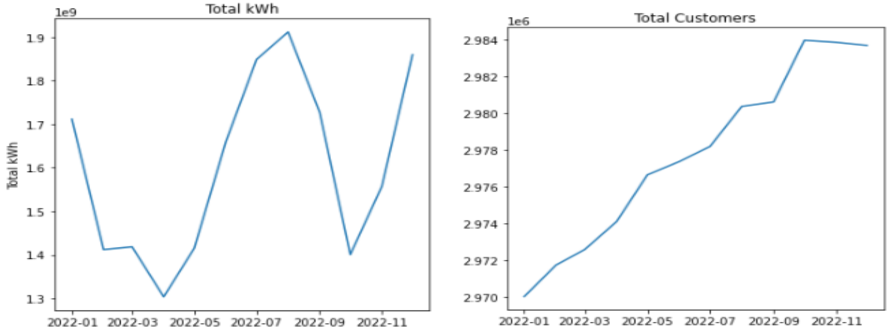
	Zip Code	Month	Year	CustomerClass	TotalCustomers	TotalkWh
0	93201	1	2013	Elec- Residential	307	243356
1	93201	2	2013	Elec- Residential	307	181203
2	93201	3	2013	Elec- Residential	303	154701
3	93201	4	2013	Elec- Residential	305	148667

- Exogenous Factors:
    - Electricity price over every year
    - Number of weekdays in a month
    - Number of weekends in a month
    - Season ['Spring', 'Fall', 'Summer','Winter']

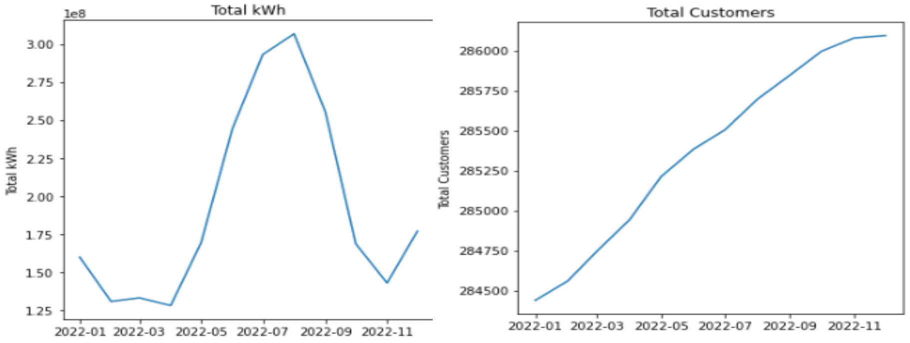
Methodology – Initial Data View



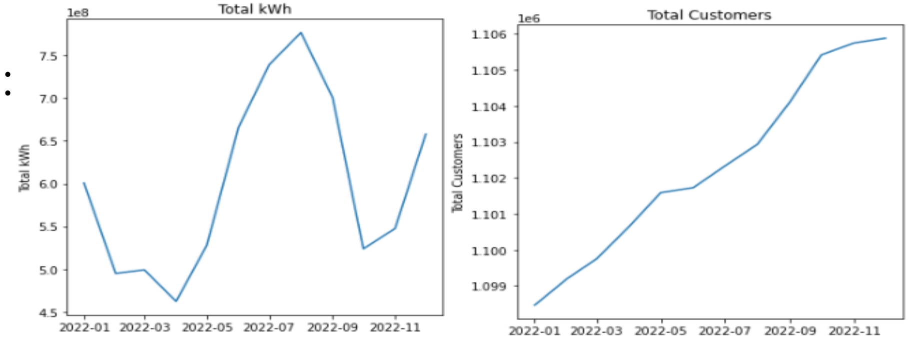
Cluster 1:
High density zips: 659



Cluster 2:
Low dense zips: 12



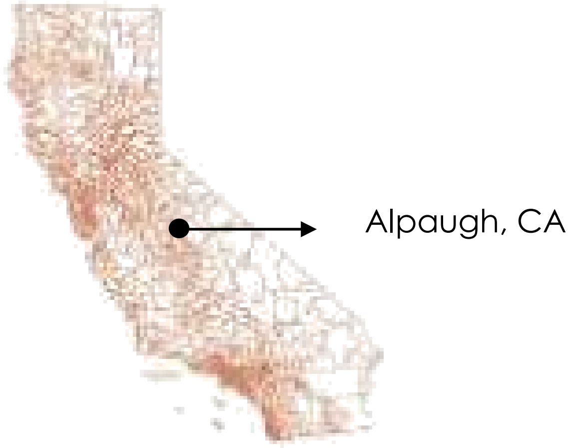
Cluster 3:
Medium dense zips: 60



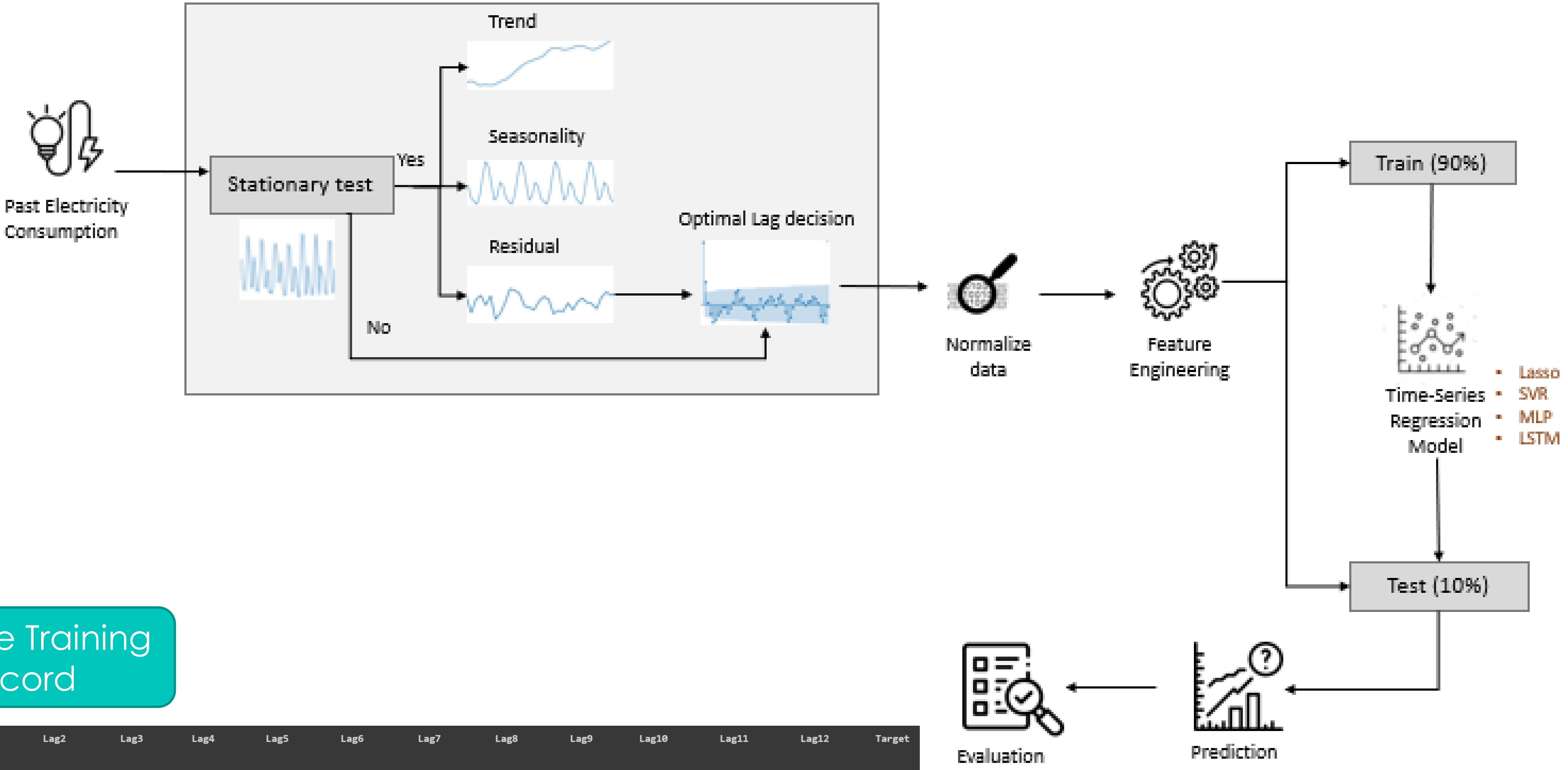
Silhouette Coefficient: 0.54 (Moderately good clustering)

Parameters considered:
Static: Past Electricity consumption, total PG&E customers
(*Method*: Agglomerative Complete Linkage)

Dynamic: Past Electricity consumption, total PG&E customers, electricity price, etc.
(*Method*: Dynamic Time Warping)



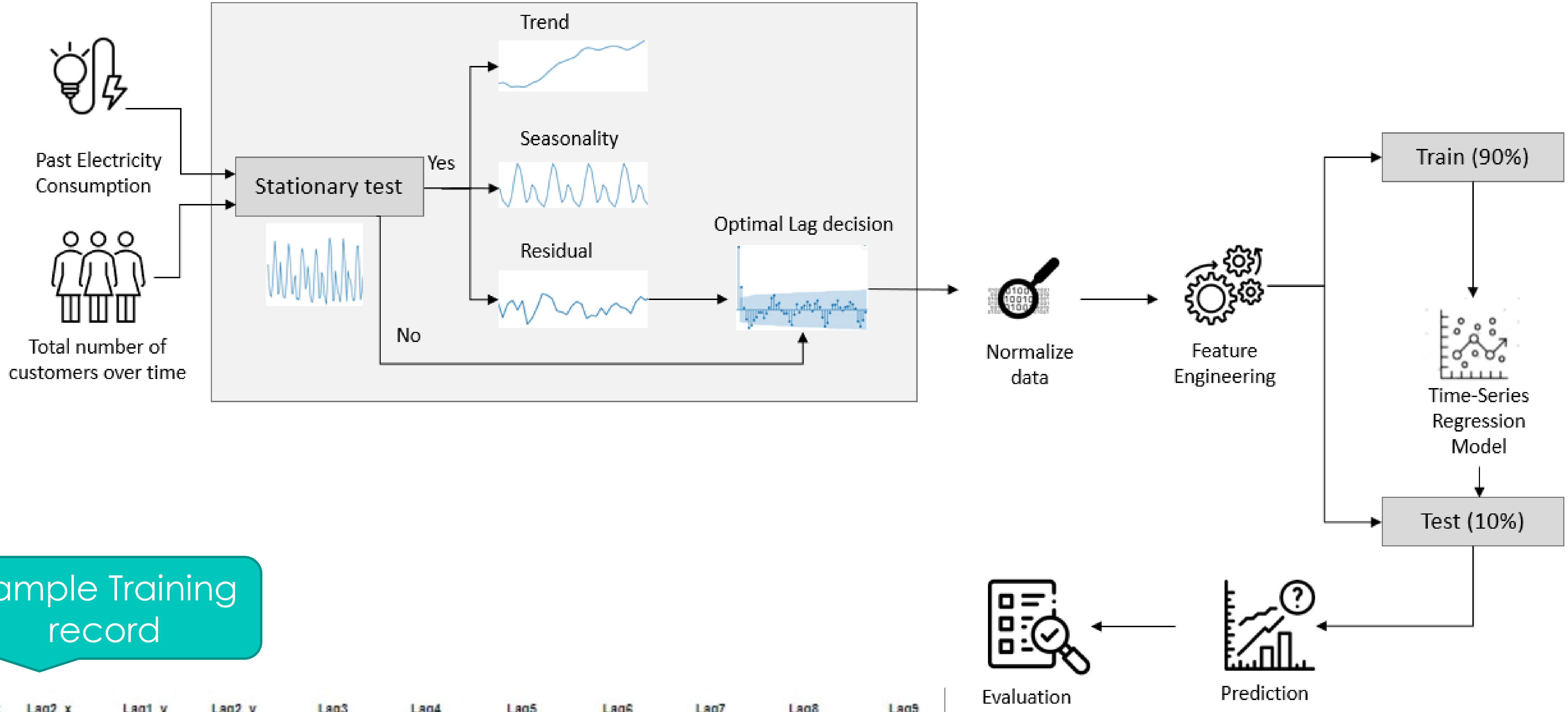
Methodology – Univariate Modeling



Sample Training record

	Lag1	Lag2	Lag3	Lag4	Lag5	Lag6	Lag7	Lag8	Lag9	Lag10	Lag11	Lag12	Target
Date													
2014-01-01	-7159.406636	10488.773920	9254.908179	41.741512	-3037.263117	-6534.091821	25710.685957	-2987.045525	3248.074846	-4520.781636	-27040.277006	-5053.402006	-4340.193673
2014-02-01	-4340.193673	-7159.406636	10488.773920	9254.908179	41.741512	-3037.263117	-6534.091821	25710.685957	-2987.045525	3248.074846	-4520.781636	-27040.277006	-12817.735340

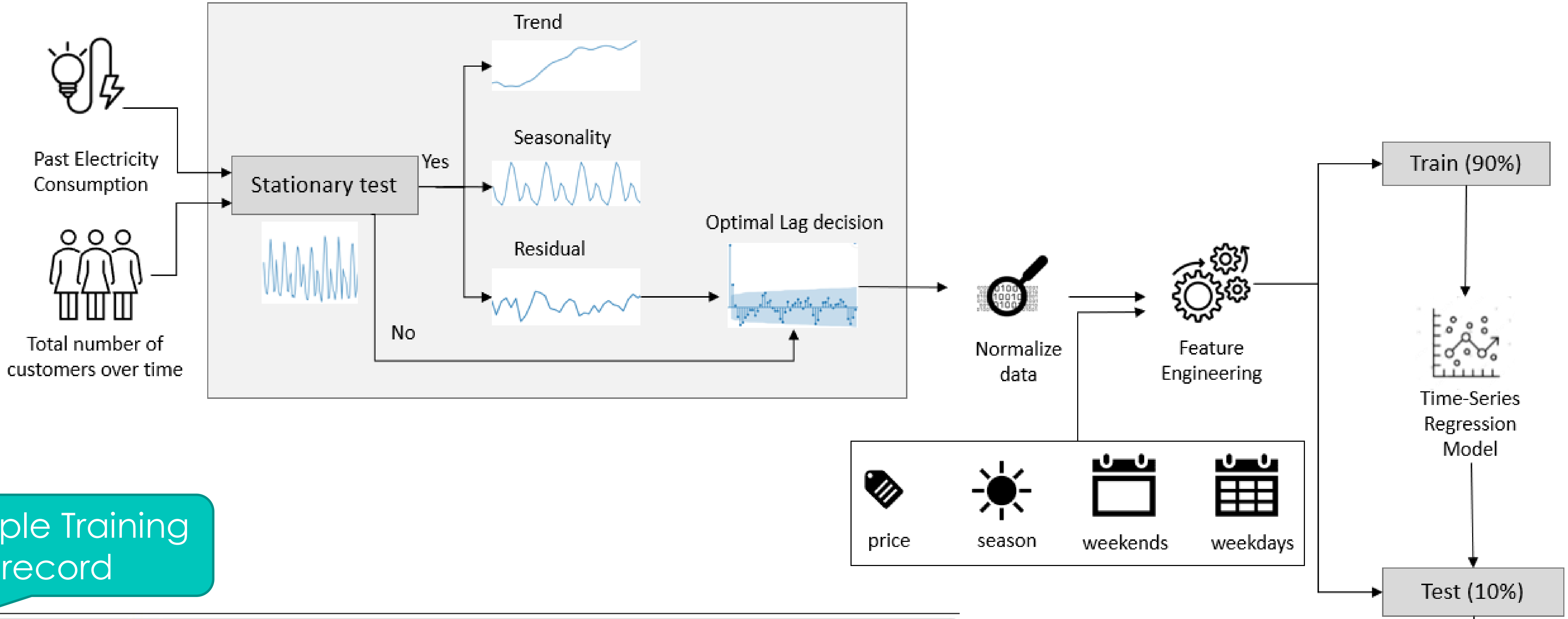
Methodology – Bivariate Modeling

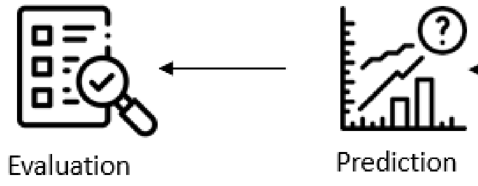


Sample Training record

	Lag1_x	Lag2_x	Lag1_y	Lag2_y	Lag3	Lag4	Lag5	Lag6	Lag7	Lag8	Lag9
datetime											
2013-11-01	-0.308642	-0.683642	9254.908179	41.741512	-3037.263117	-6534.091821	25710.685957	-2987.045525	3248.074846	-4520.781636	-27040.277006
2013-12-01	-0.998457	-0.308642	10488.773920	9254.908179	41.741512	-3037.263117	-6534.091821	25710.685957	-2987.045525	3248.074846	-4520.781636

Methodology – Modeling with Exogenous factors





Sample Training record

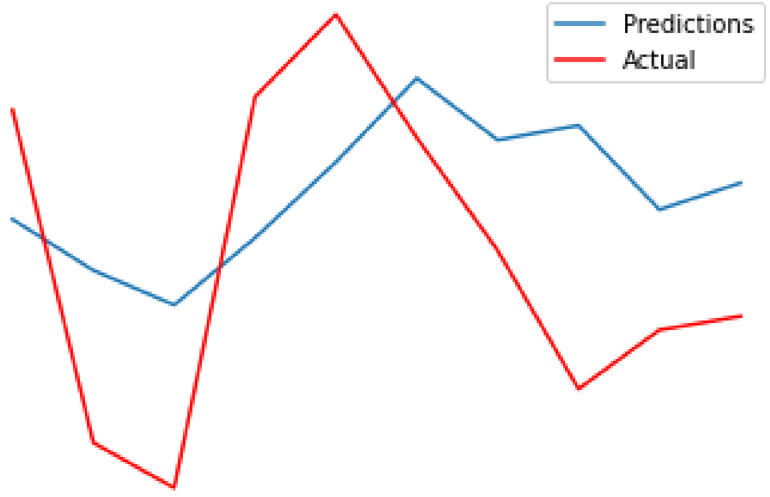
```
1  lagged_features_exo.head(2)
```

Lag4	Lag5	Lag6	Lag7	Lag8	Lag9	Lag10	price_updated	weekends	Fall	Spring	Summer	Target
34.091821	25710.685957	-2987.045525	3248.074846	-4520.781636	-27040.277006	-5053.402006	0.17	9	1	0	0	10488.773920
37.263117	-6534.091821	25710.685957	-2987.045525	3248.074846	-4520.781636	-27040.277006	0.17	9	0	0	0	-7159.406636

Experimental Results

		Lasso	SVR	MLP
Univariate	**RMSE**	0.938	1.004	1.25
	MAPE	127.6	145.77	371.00
Bivariate	**RMSE**	0.9393	0.983	1.017
	MAPE	127.1	134.18	329.42
Exogenous	**RMSE**	0.9759	0.97	1.2
	MAPE	114.6	115.88	252.41

Visual Inspection for 1 model



- Having multiple zip codes gives rise to multiple time series instances. To try a model to generalize over all the zip codes, LSTM is implemented.

Parameters	**Observation**
Learning Rate	1x10^-6
Epochs	200
MSE	0.00059
RMSE	0.02445
MAE	0.01

Discussion

- Time Series is a complex task especially because there are a lot of exogenous variables which for a fact directly impact the electricity consumption. But a lack of common suitable variables to be able to bring both together current tasks don't establish full potential. Observed this with climate data as well as EV vehicle registration data

- Handling less sized data involves appropriate cross validation techniques to ensure the robustness or results obtained specifically for the data with high amount of volatility

Future Scope for Improvement

Dynamic Clustering	Household level prediction	Smaller Time-Step Prediction	Multi-step prediction	Stacked-Prediction Model
Adding exogenous factors which change with time, like climate, can better cluster zip codes	Can be extended to household-level prediction if we have IoT-based data and enable personalized predictions	Moving from the current month's time-step prediction to a smaller time-step like day or hour level	To Predict multiple time-steps into the future	It involves using the predictions of multiple models as inputs to a new model. Works better than ensemble in time series analysis

THANK YOU

Appendix-Clustering