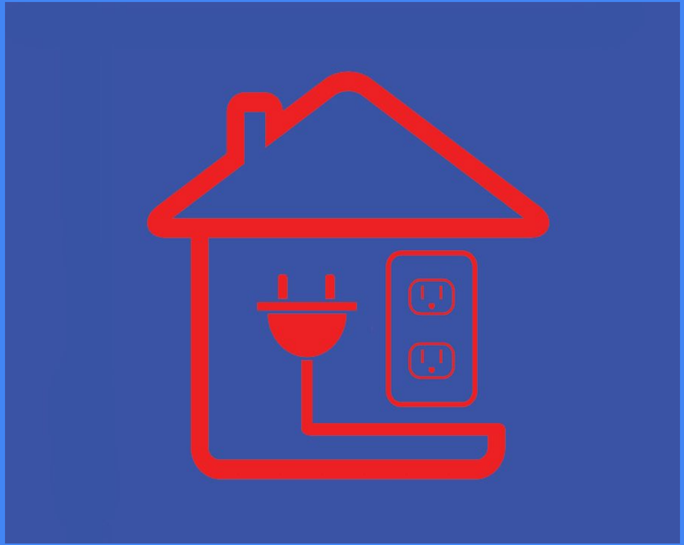
London Smart Meter Analysis

Energy Load Forecasting using Neural Networks and Household classification based on consumption patterns

By: Robbie Davison



Background

The way we produce and consume energy is constantly changing

- Roughly 90 million smart meters have been installed in America and almost 15 million in the UK
- Smart meters provide more regular updates on energy consumption than regular meters
- Can this data be used to create better load forecast models
- Can it be used to better understand individual house/business consumption patterns and inform business decisions?

Other potential applications of more accurate electricity consumption data:

- Individualized dynamic pricing during certain hours to discourage overconsumption
- Early Alzheimer’s detection
- Advanced modeling can help improve production strategy, eg which power plants to start up and when

Problem Statement

Given a households electricity consumption profile, can you:

- Untangle meaningful consumption patterns of houses with varying economic backgrounds?
- Predict load for the population and individual groups? If so which models are best?

The Data

- Kaggle dataset containing:
- Daily electricity consumption metrics for 5,500 individual homes in London
- Daily weather statistics eg. max temperature, humidity, cloud cover etc.
- UK Holidays
- The ACORN Classification for each home

What are ACORNs?

A Classification Of Residential Neighborhoods

acorn

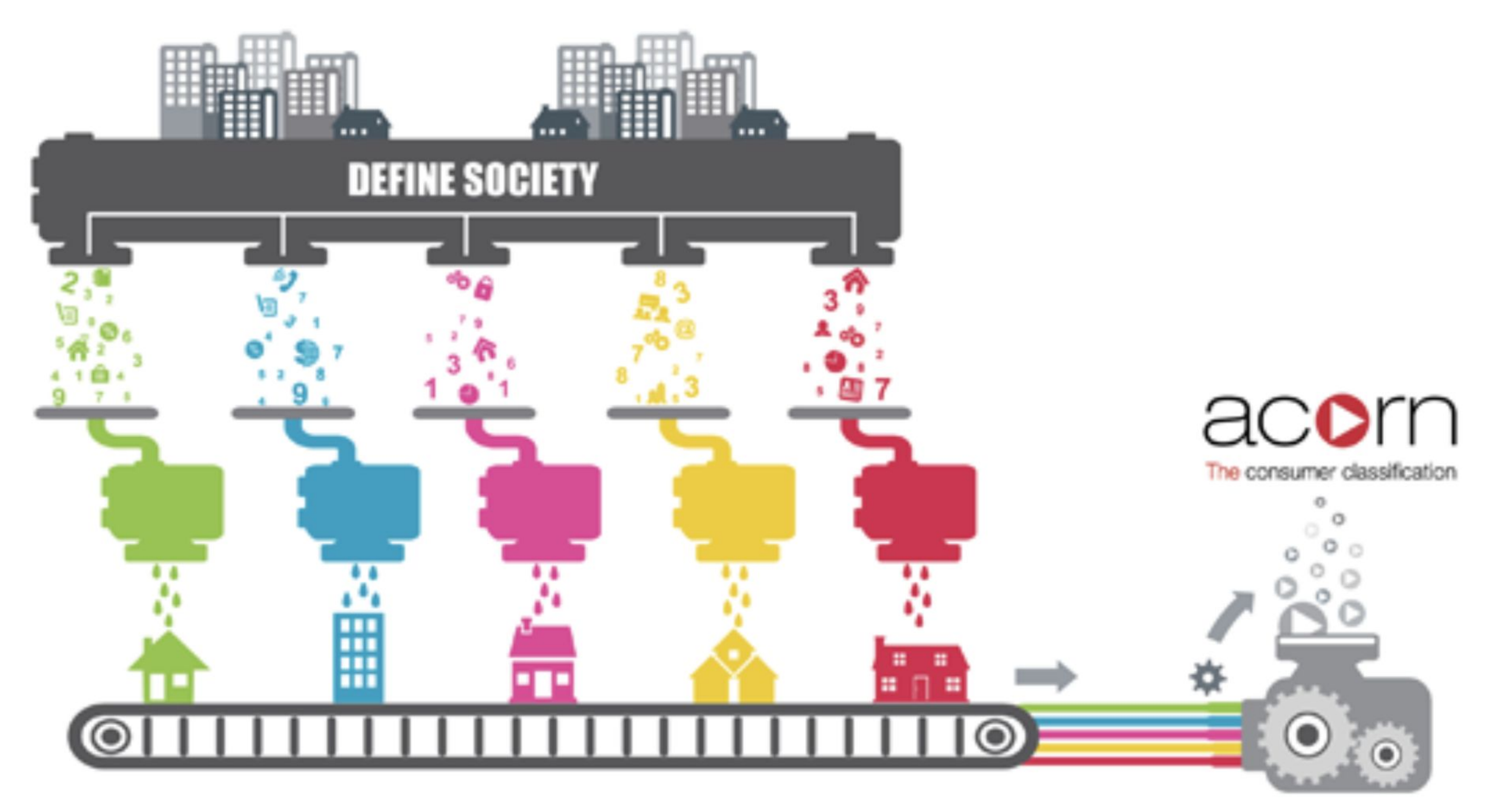
TRY ACORN NOW | WHAT IS ACORN? | WHAT CAN ACOR

UNDERSTANDING CONSUMERS AND COMMUNITIES

Knowing who and where your consumers are is crucial for effective targeting and involves intelligent customer analysis and consumer segmentation.

Acorn is used to understand consumers' lifestyle, behaviour and attitudes, together with the needs of communities and is important to both private sector and public service organisations. It is used to analyse customers, identify profitable prospects, evaluate local markets and focus on the specific needs of each catchment and neighbourhood.

- Acorn segments postcodes into 6 categories, 18 groups and 62 types.
- Able to add another layer of complexity and questioning to my models
- Types are further subdivided into 313 micro-segments for an extra level of precision for specialist analyses.


DEFINE SOCIETY
acorn
The consumer classification

What are ACORNs?

6 Categories

1. Affluent Achievers
2. Rising Prosperity
3. Comfortable Communities
4. Financially Stretched
5. Urban Adversity
6. Not Private Households

Acorn User Guide Introduction 4

		Types		
1	Affluent Achievers			10
A	Lavish Lifestyles			11
		1	Exclusive enclaves	12
		2	Metropolitan money	13
		3	Large house luxury	14
B	Executive Wealth			15
		4	Asset rich families	16
		5	Wealthy countryside commuters	17
		6	Financially comfortable families	18
		7	Affluent professionals	19
		8	Prosperous suburban families	20
		9	Well-off edge of towners	21
C	Mature Money			22
		10	Better-off villagers	23
		11	Settled suburbia, older people	24
		12	Retired and empty nesters	25
		13	Upmarket downsizers	26

		Types		
5	Urban Adversity			82
O	Young Hardship			83
		49	Young families in low cost private flats	84
		50	Struggling younger people in mixed tenure	85
		51	Young people in small, low cost terraces	86
P	Struggling Estates			87
		52	Poorer families, many children, terraced housing	88
		53	Low income terraces	89
		54	Multi-ethnic, purpose-built estates	90
		55	Deprived and ethnically diverse in flats	91
		56	Low income large families in social rented semis	92
Q	Difficult Circumstances			93
		57	Social rented flats, families and single parents	94
		58	Singles and young families, some receiving benefits	95
		59	Deprived areas and high-rise flats	96

Data Cleaning and EDA

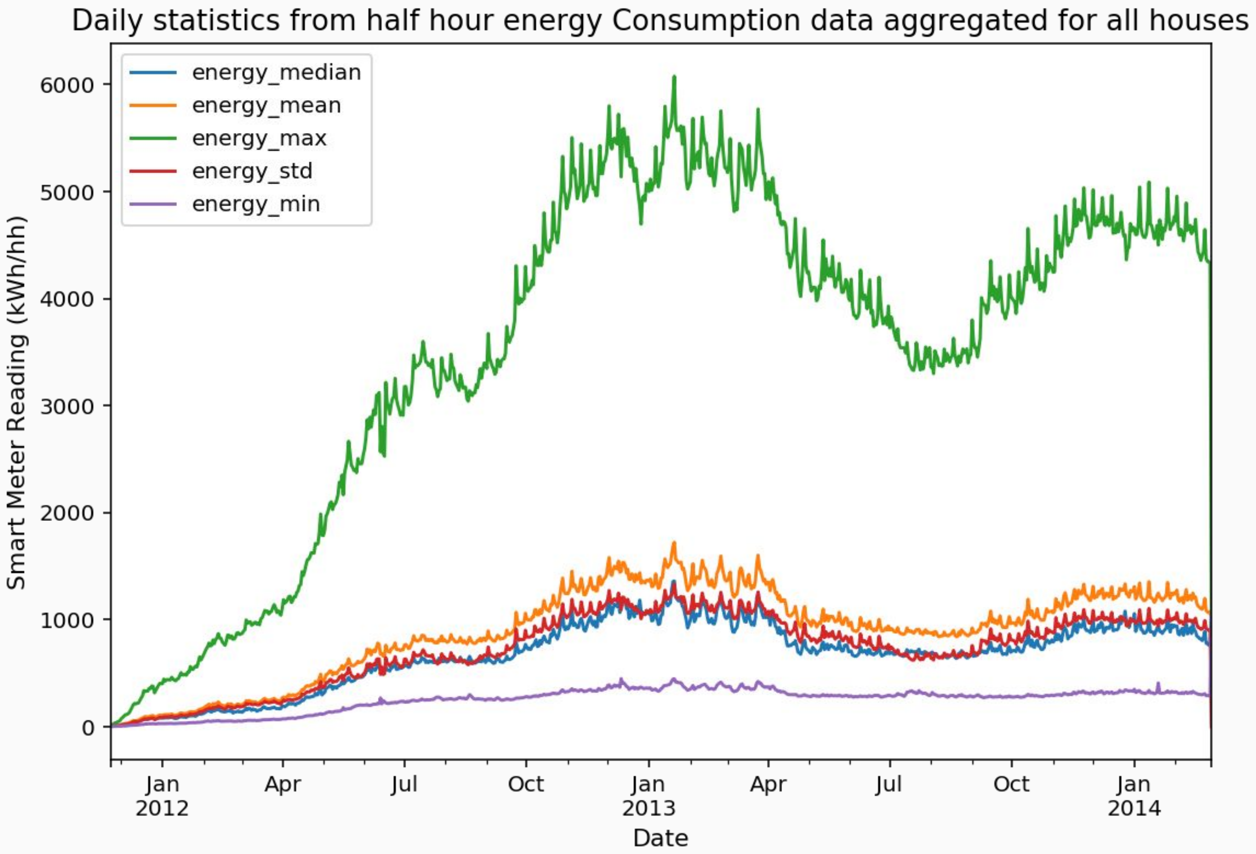
- Lots of preprocessing already completed
- Combined electricity consumption, weather, holiday and acorn columns
- For the first 10 months, not all houses had data recorded, so I had to drop those days

	energy_median	energy_max	energy_sum	energy_min	temperatureMax	temperatureLow	windSpeed	cloudCover	humidity	visibility	holiday
day											
2011-11-23	2.97	9.84	90.39	1.11	10.36	8.24	2.04	0.36	0.93	8.06	0
2011-11-24	4.84	19.05	213.41	2.02	12.93	9.71	4.04	0.41	0.89	10.64	0
2011-11-25	5.69	23.25	303.99	2.28	13.03	7.01	5.02	0.48	0.79	12.38	0

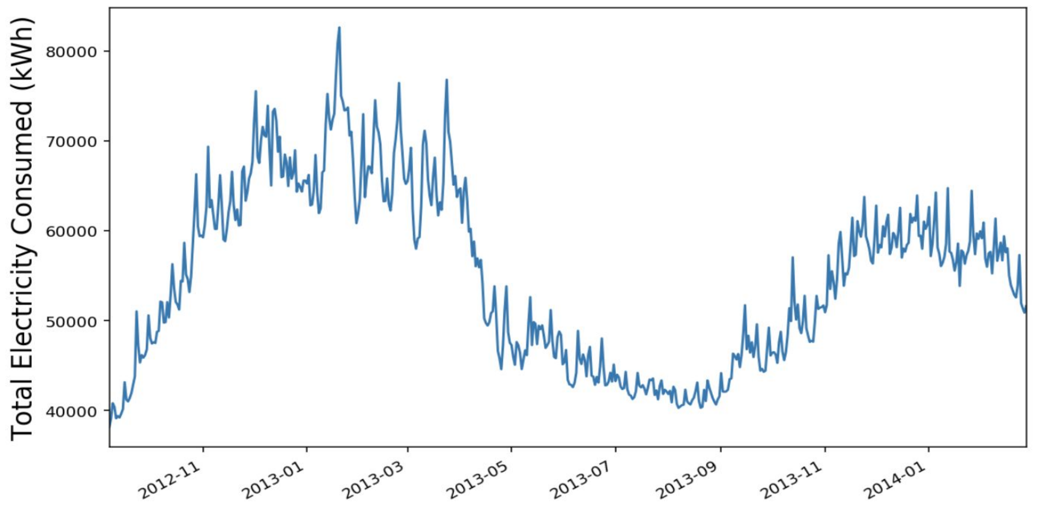
EDA

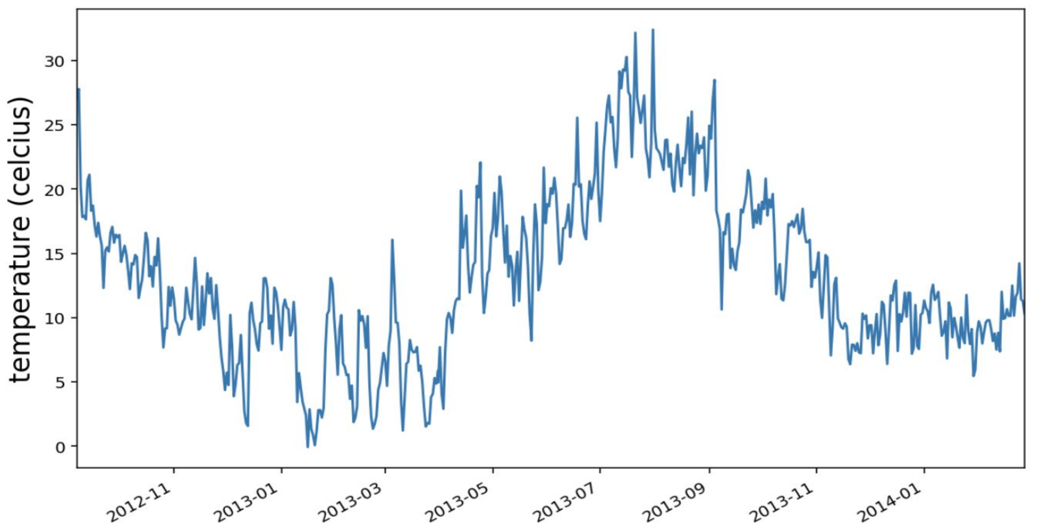
Daily consumption summed across all households

- Seasonal
- Stationary
- Sudden shocks
- Limited modeling to data after October 2012



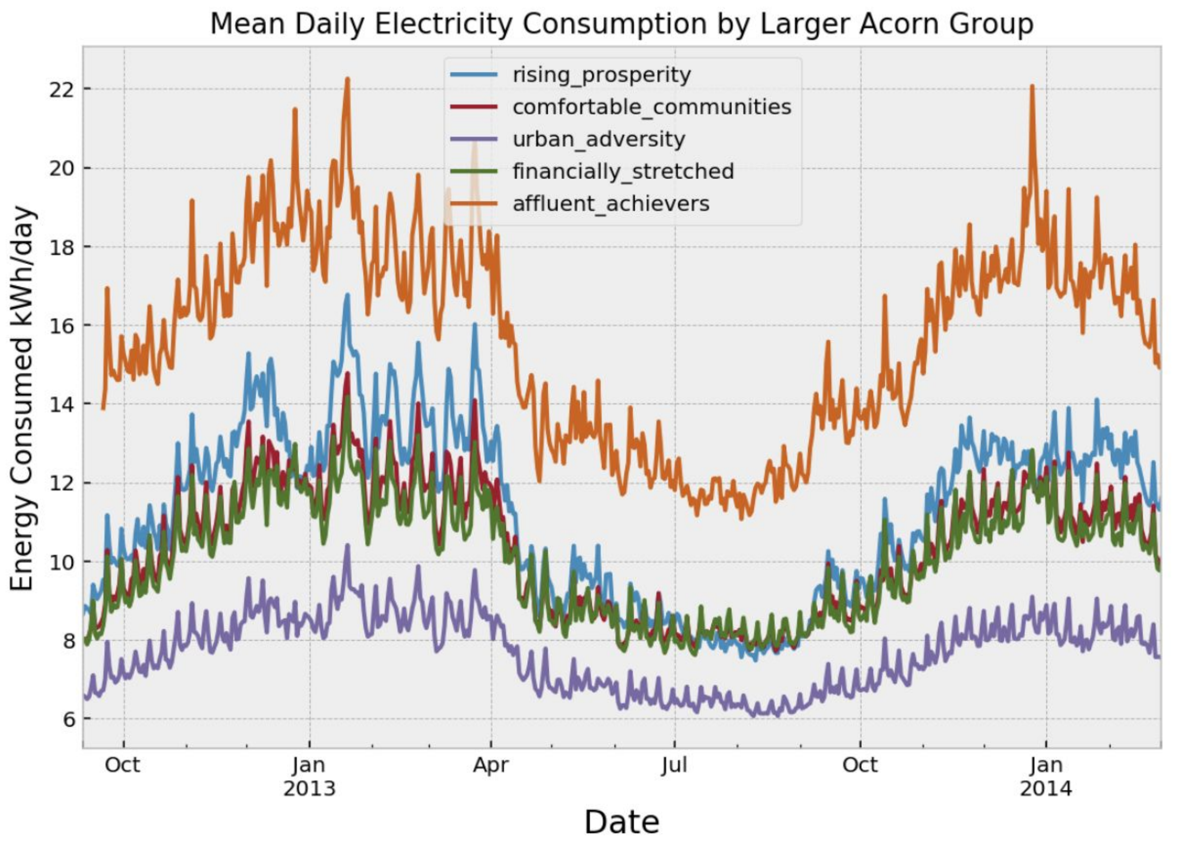
Consumption and Temperature are inversely related as expected





Consumption Decreases with ACORN

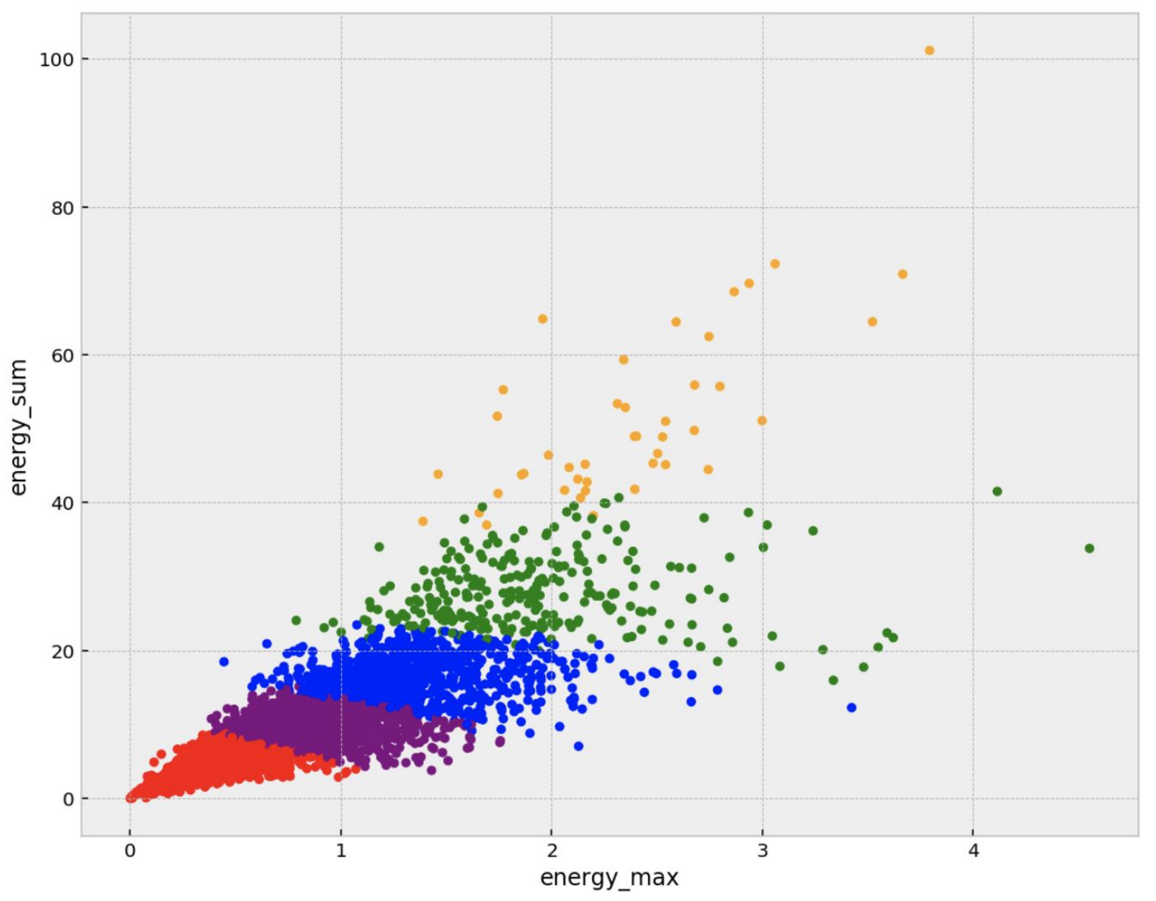
big_acorn	Avg Consumption (kWh/day)
affluent_achievers	15.33
rising_prosperity	10.85
comfortable_communities	10.01
financially_stretched	9.87
urban_adversity	7.55



Part 1: Categorization

K-Means Clustering

- Each dot corresponds to an individual house
- Plot of K=5 clusters had poor separation of classes
- Silhouette score = 0.38



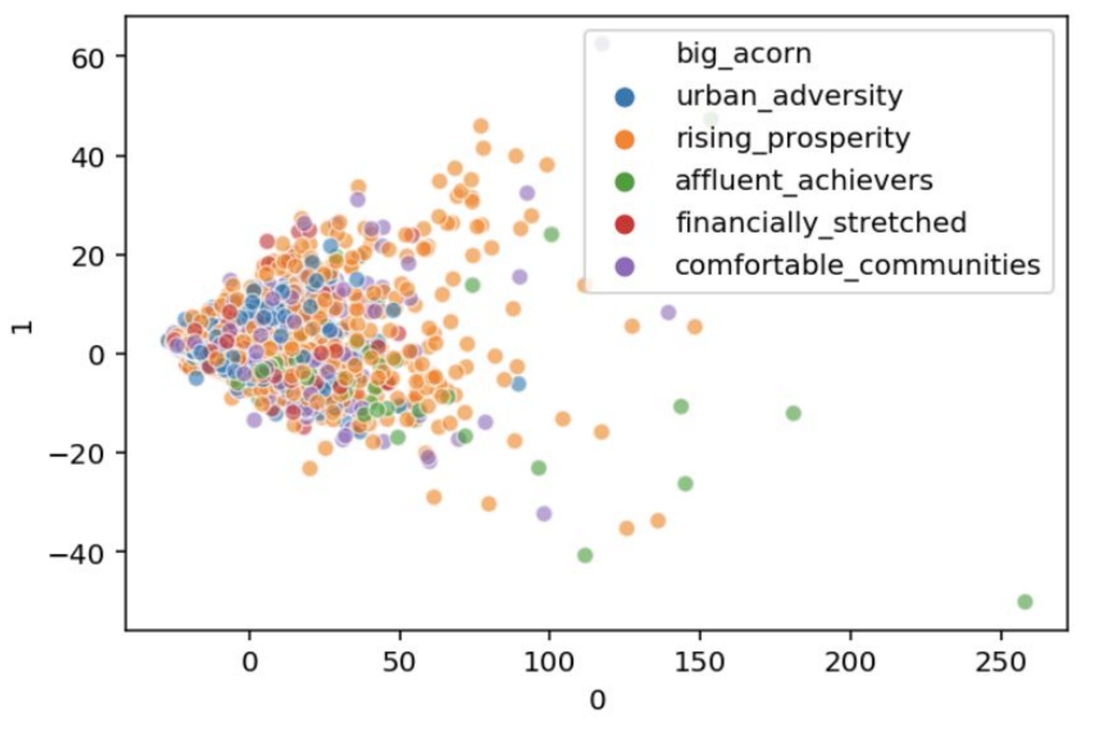
Part 1: Categorization
PCA

- Rearrange data in pivot table so that each row is a house and the columns are days
- Run pca on this to see if you can isolate relevant clusters
- Then plot the first two components to look for separation

day	2012-09-11	2012-09-12	2012-09-13	2012-09-14	2012-09-15	2012-09-16	2012-09-17	2012-09-18	2012-09-19	2012-09-20	2012-09-21	2012-09-22	2012-09-23
LCLid													
MAC000003	11.929	12.834	13.488	14.136	11.713	11.577	12.333	12.954	13.775	13.921	13.727	12.473	12.404
MAC000004	1.499	1.650	1.475	1.518	1.560	1.581	1.613	1.547	1.604	1.714	1.456	1.530	1.719
MAC000005	1.917	1.808	1.780	1.762	1.759	2.853	4.197	2.988	4.259	2.730	5.044	4.306	4.281
MAC000006	2.413	2.485	2.971	2.621	2.161	2.412	2.353	2.570	1.958	2.617	2.406	2.389	2.745
MAC000013	4.919	5.047	4.968	6.269	5.874	5.116	4.715	4.961	4.908	5.475	7.717	6.422	5.229

PCA:
Plot of the first two components

- Again poor separation
- Would have liked to see more separated clusters
- If clusters were separated, we might assume energy use time series distinguish acorns

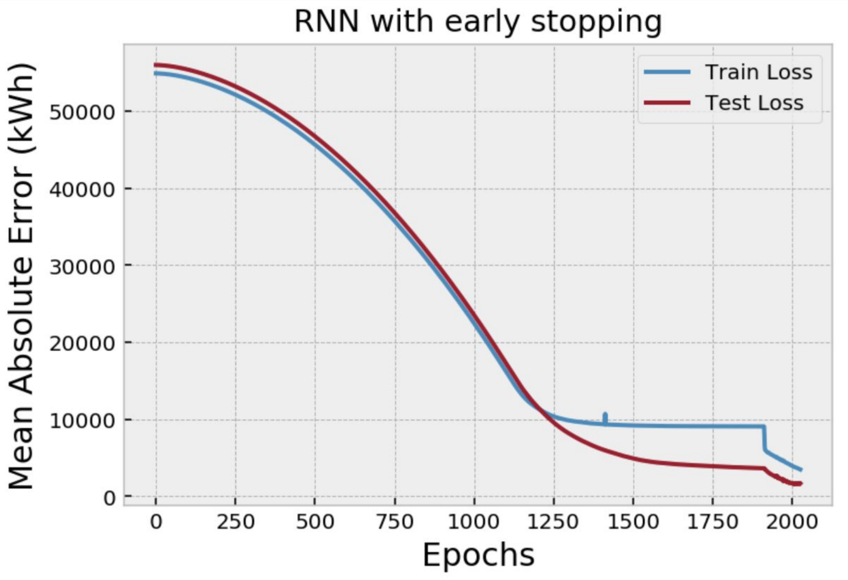


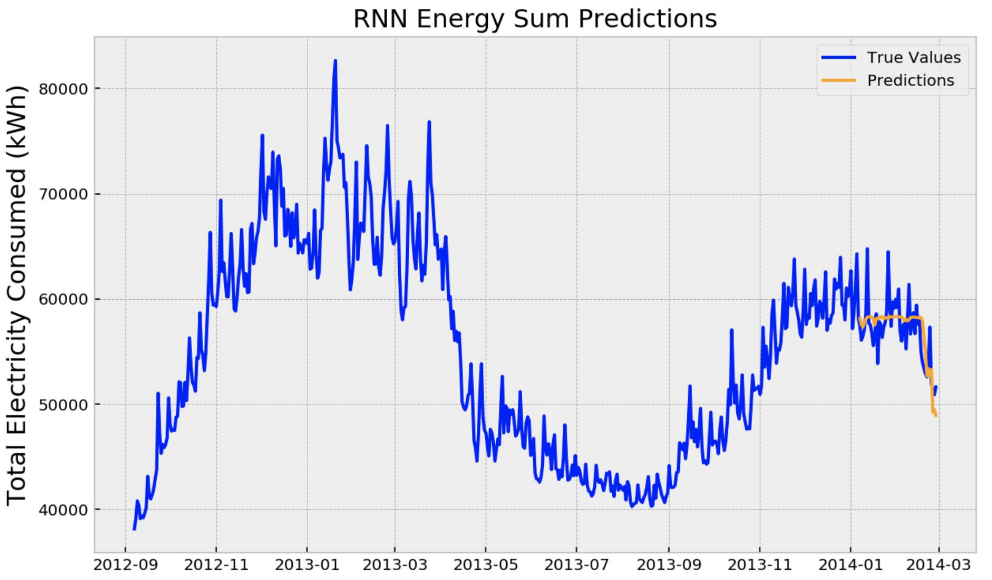
Part 2: Load Forecasting
Recurrent Neural Network (RNN)

- RNN’s are good at time series forecasting
- They can handle multivariate series so they can incorporate the metadata like weather and holidays
- They don’t provide any insight into uncertainty
- They are a black box model that can be hard to interpret

Part 2: Load Forecasting

Recurrent Neural Network (RNN)





- Best model had a time lag of 3 days and batch size of 16
- Achieved an RMSE of **2,188** kWh
- RNN couldn’t account for sudden spikes - didn’t stray far from the mean

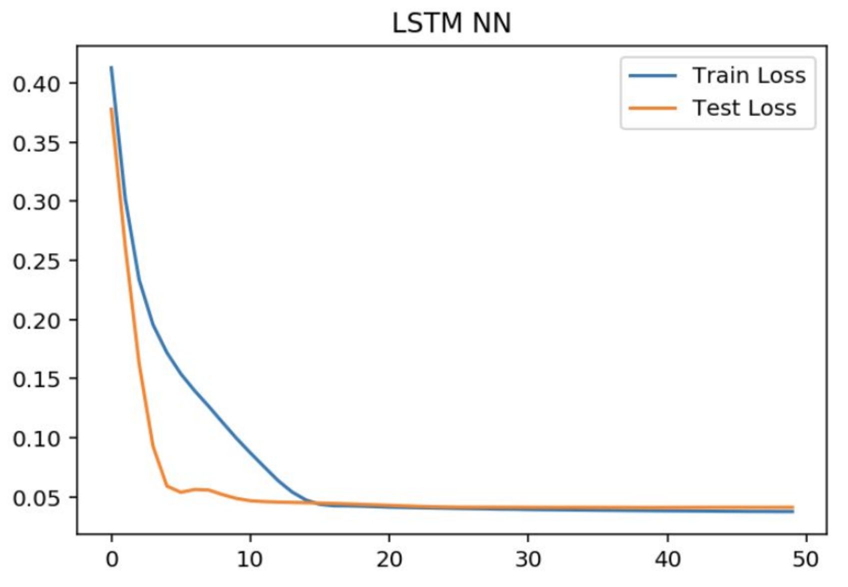
Part 2: Load Forecasting

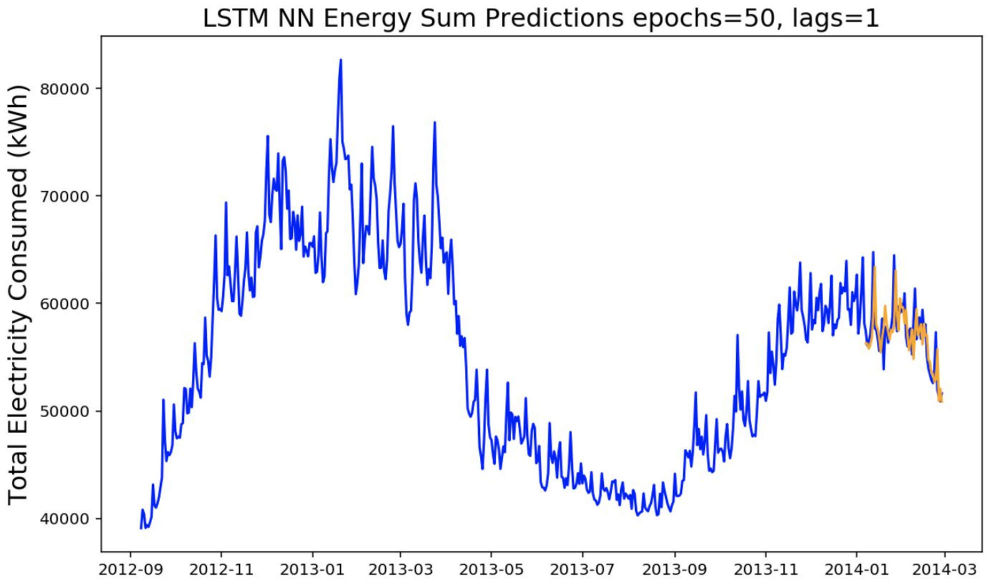
Long-Term Short-Term Neural Network (LSTM)

- LSTMs are well suited for time series prediction because they have feedback connections
- This should help the LSTM predict the sudden spikes that the RNN couldn’t
- Like RNNs, LSTMs can’t provide confidence intervals

Part 2: Load Forecasting

Long-Term Short-Term Neural Network (LSTM)





- Best model had a time lag of 1 days and batch size of 64
- Achieved an RMSE of **2,367** kWh
- Better at predicting sudden spikes than the RNN

Results and Takeaways

- ACORNs do not show a close relationship with clusters of houses formed using K Means clustering
- Only the wealthiest and poorest groups consumed a notably different amount of electricity
- RNN model had the best RMSE but the LSTM model was better at predicting sudden spikes

Next steps

- Try applying a Bayesian Neural Net or Gaussian process in order to account for confidence intervals
- Add columns that tell when a holiday or bad weather is expected the next day
- Experiment with feature engineering, eg, creating a column that accounts for both temperature and max consumption for a house when trying to cluster