Out[1]: Click here to toggle on/off the raw code.

1. Introduction

[illegible]

1.1 Intent

[illegible]

1.2 Objectives

[illegible]

1.3 Impact

[illegible]

2. Table of Contents

3. Data Overview

3.1 About the Data

[illegible] each energy type is in MW and the time-series contains hourly info, the number of each cell represent MWh.

Data source can be found here.

3.2 Preview Data

```
Data shape is: (35070, 21)
```

date_time	forecast_solar_day_ahead	forecast_wind_onshore_day_ahead	generation_biomass	generation_fossil_brown_coal/lignite	generation_fossil_gas	generat
2015-01-01 00:00:00	17.0	6436.0	447.0	329.0	4844.0	
2015-01-01 01:00:00	16.0	5856.0	449.0	328.0	5196.0	
2015-01-01 02:00:00	8.0	5454.0	448.0	323.0	4857.0	
2015-01-01 03:00:00	2.0	5151.0	438.0	254.0	4314.0	
2015-01-01 04:00:00	9.0	4861.0	428.0	187.0	4130.0	

3.3 Data Overview - Metrics

Data Description

Out[4]:

	count	mean	std	min	25%	50%	75%	max
forecast_solar_day_ahead	35070.0	1438.825121	1677.661532	0.000000	69.000	576.0000	2635.000	5836.0000
forecast_wind_onshore_day_ahead	35070.0	5471.372512	3176.148983	237.000000	2979.000	4855.5000	7353.000	17430.0000
generation_biomass	35070.0	383.536128	85.348006	0.000000	333.000	367.0000	433.000	592.0000
generation_fossil_brown_coal/lignite	35070.0	448.060251	354.603125	0.000000	0.000	509.0000	757.000	999.0000
generation_fossil_gas	35070.0	5622.474309	2201.444741	0.000000	4126.000	4969.0000	6428.750	20034.0000
generation_fossil_hard_coal	35070.0	4256.296179	1961.968024	0.000000	2527.000	4474.0000	5839.000	8359.0000
generation_fossil_oil	35070.0	298.335358	52.518153	0.000000	263.000	300.0000	330.000	449.0000
generation_hydro_pumped_storage_consumption	35070.0	475.867237	792.594472	0.000000	0.000	68.0000	617.000	4523.0000
generation_hydro_run-of-river_and_poundage	35070.0	972.117536	400.740520	0.000000	637.000	906.0000	1250.000	2000.0000
generation_hydro_water_reservoir	35070.0	2605.122241	1835.141359	0.000000	1077.250	2165.0000	3757.000	9728.0000
generation_nuclear	35070.0	6263.475278	840.224889	0.000000	5759.000	6564.0000	7025.000	7117.0000
generation_other	35070.0	60.226461	20.238237	0.000000	53.000	57.0000	80.000	106.0000
generation_other_renewable	35070.0	85.634674	14.076581	0.000000	73.000	88.0000	97.000	119.0000
generation_solar	35070.0	1432.564271	1679.964650	0.000000	71.000	616.0000	2578.000	5792.0000
generation_waste	35070.0	269.422327	50.214081	0.000000	240.000	279.0000	310.000	357.0000
generation_wind_onshore	35070.0	5465.275592	3213.583611	0.000000	2933.000	4849.5000	7400.500	17436.0000
price_actual	35070.0	57.883170	14.203079	9.330000	49.350	58.0200	68.010	116.8000
price_day_ahead	35070.0	49.873242	14.618080	2.060000	41.490	50.5200	60.530	101.9900
temp	35070.0	61.753049	13.069317	29.825352	51.638	60.5732	71.483	97.0888
total_load_actual	35070.0	28696.145709	4576.849067	18041.000000	24806.000	28901.0000	32194.000	41015.0000
total_load_forecast	35070.0	28710.702965	4595.004597	18105.000000	24792.250	28905.0000	32263.000	41390.0000

4. Consumer Profile Analysis - Modeling Consumer Demand Patterns

4.1 Methodology

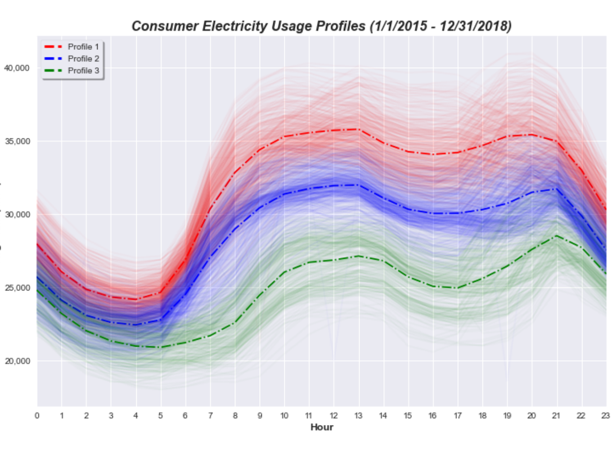
Used KMeans clustering algorithm to model different patterns of consumer behavior pertaining to energy demand. Silhouette scores used to select number of clusters(3). Hourly demand will be grouped by cluster and plotted along with mean value of cluster. Clustering algorithm validated by the t-Distributed Stochastic Neighbor Embedding (t-SNE) technique.

4.2 Results

Looking at the results of our KMeans clustering algorithm, we were able to model 3 valid and unique patterns of energy usage with varying intensities of demand throughout each hour of the day.

Some of the similarities between each profile are not all too surprising. Energy demand tends to decrease in the evening beginning around 8-9 pm and steadily decreases as the night goes on. We see rises in demand right around when most folks are waking up in the morning, though profile 1 has the slowest gradual increase of the 3(this may be indicative of consumer behavior on the weekend/holidays).

Out[5]:

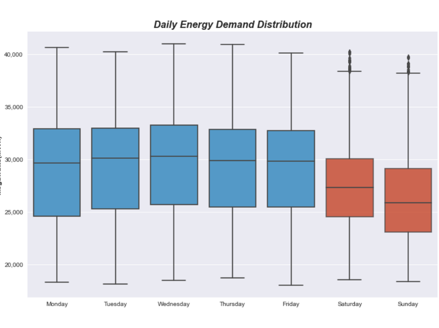


Insights into Consumer Profiles

Profile 1 Profile 1 appears to be indicative of consumer behavior of consumers that stay at home during the week. These people could be stay-at-home parents, working from home, or our elderly consumers. The curves of Profile 1 are very similar to Profile 3, which appears to represent energy usage of working consumers during the week. The matching behavior with profile 3 in the morning could indicate consumers getting ready for the day or preparing children for school

Profile 2 Profile 2 appears to be indicative of consumer behavior on weekends /holidays. This curve has the slowest rise in demand in the mornings relative to the other 2 profiles. This appears to represent indicate consumers are sleeping in without the pressure of getting ready for work. Refer to the "Daily Energy Demand Distribution" boxplot below to see that energy demand is quite a bit lower on Saturdays and Sundays as compared to the other days of the week.

Out[6]:



Most folks will travel to visit others and thus homes tend to be emptier on weekends which would explain the lower demand. Additionally, the leading contributor to the overall decrease in demand would be that most large office buildings are closed. Peak demand is right around 9pm, generally a good time for socializing and attending entertainment events.

Profile 3 Profile 3 appears to be indicative of energy demand during working days. Energy demand would be highest during these times since office building require immense amount of energy (and sometimes stay on all-night). We also see that profile 3 on average sees the decline for energy demand declining in the evening occurring first, indicating that these consumers are turning in early for night to prepare for work the next day.

4.3 Decision Recommendations

Consumer/Employer Energy Conservation Rebate Program

Based on the consumer profiles we were able to create with KMeans clustering, we should attempt to lower the energy demand of Profile 3 to try and match Profile 1. Some possible rebate strategies include:

- Providing large office space building and warehouses with energy saving lights and motion sensors to shut-off any lights not in use, especially at night after the workday is completed. Often times, lights will remain on while no one is at the office.
- Due to COVID 19, a lot of consumers are working from home. Though we do not have that data, it would stand to reason that energy demand would be significantly lower. If that is the case, then it is worth investigating a way to create a program that incentivizes employers to allow employees to work from home at least one day a week.

Both of these ideas would require testing before investing any considerable resources into deploying, however it would be very beneficial to the energy grid without having to request most of our consumers to do anything to preserve energy.

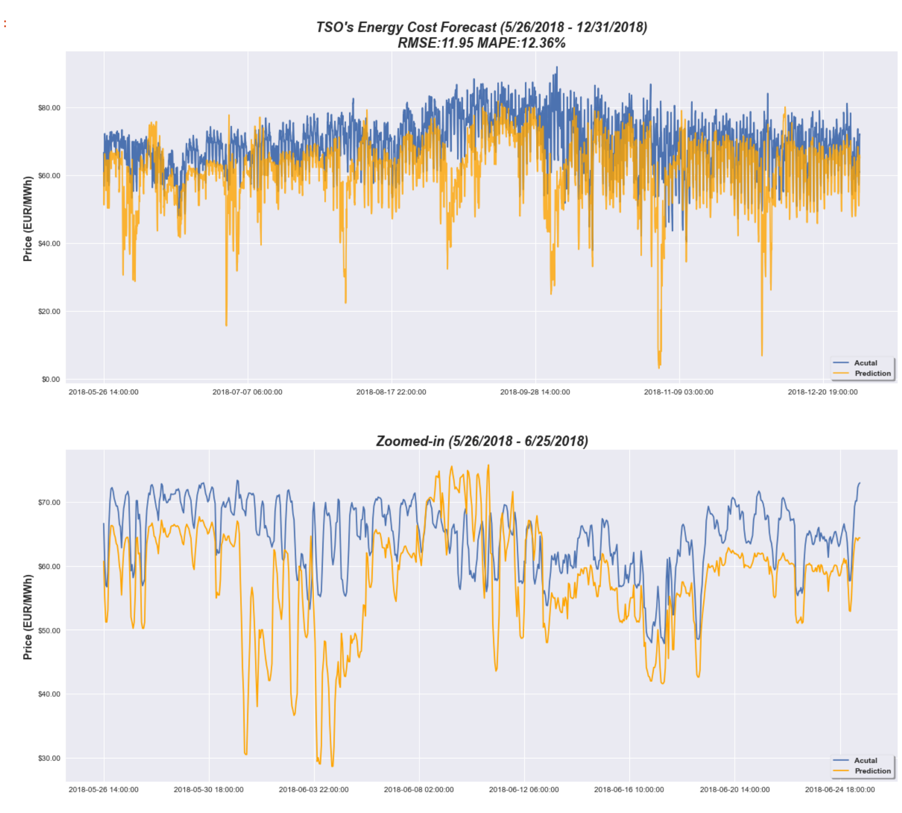
5. Multivariate Time Series Analysis - Forecasting Energy Production costs

5.1 Methodology

Developed and optimized a multivariate time series XGBRegressor to better predict and forecast energy productions costs. Model incorporates demand and energy production variables available in data set and surpasses predictions made by the TSO. Also conducted a cost-benefit analysis to evaluate total revenue saved by using XGBRegressor as opposed to the TSO model.

5.2 Original TSO Forecast

Out[7]:



The primary issue of the TSO model is that it tends to underestimate peak points of hourly production cost. This underestimation leads to inaccurate projections when determining future company performance.

5.3 Multivariate XGBRegressor Forecast

The XGBRegressor easily outperformed the TSO model. The most important component of this model is that it is able to capture the increasing variance of energy demand, particularly in the second half of 2018.

5.4 Decision Recommendations

Projected Savings with new Cost Model

The error metrics of the original TSO model forecasts from 5/26/2018 - 12/31/2018 are as follows:

- **Total Error**: $41,595.38
- **Average Daily Error**: $7.90
- **Average Hourly Error**: $7.91

The error metrics of the XGBRegressor model forecasts from 5/26/2018 - 12/31/2018 are as follows:

- **Total Error**: $134.00
- **Average Daily Error**: $.02
- **Average Hourly Error**: $.03

Using the new cost model stands to save the energy production company $41,461.38 over a roughly 6-month period. If we extrapolate the model out, we can provide a savings of \$82,922.76 per year.

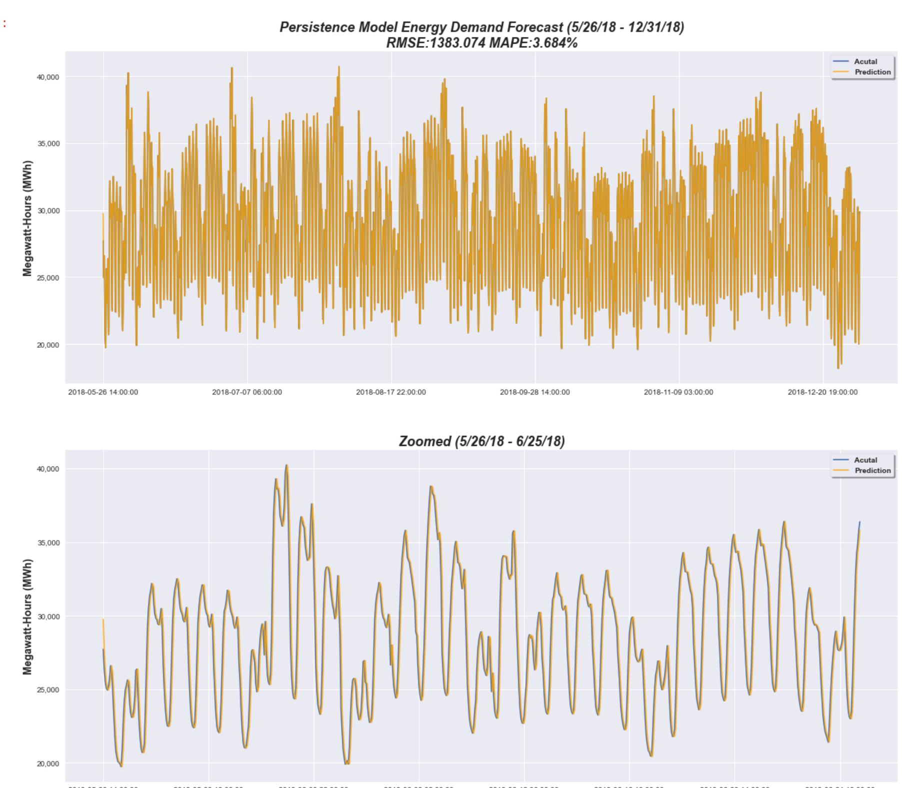
6. Univariate Time Series Analysis - Forecasting Consumer Demand

6.1 Methodology

Created a univariate Deep Learning (LSTM) model to better predict consumer demand. Model is univariate because there are no meaningful features related to consumer demand other than lagged time series features. Base model to compare LSTM performance to will be a persistence model.

6.2 Univariate Persistence Model

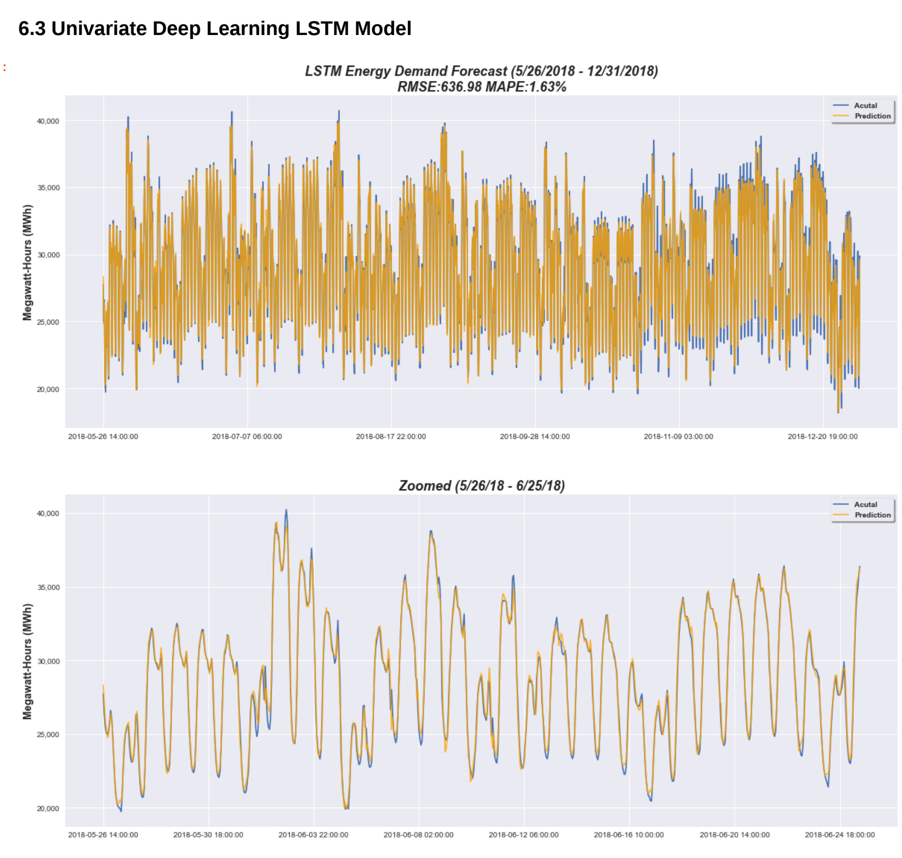
Out[9]:



A persistence model is generally a very good model when predicting hourly data. Each time step has a high correlation to the previous time step, thus the nice fit we see here. The key drawback to this type of model is that we are always an hour behind demand, leaving some room for improvement.

6.3 Univariate Deep Learning LSTM Model

Out[10]:




```
LSTM Model Summary:

Model: "sequential"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
lstm (LSTM)                  (None, 100, 64)           16896
_________________________________________________________________
dropout (Dropout)            (None, 100, 64)           0
_________________________________________________________________
lstm_1 (LSTM)                (None, 100, 64)           33024
_________________________________________________________________
dropout_1 (Dropout)          (None, 100, 64)           0
_________________________________________________________________
lstm_2 (LSTM)                (None, 100, 64)           33024
_________________________________________________________________
dropout_2 (Dropout)          (None, 100, 64)           0
_________________________________________________________________
lstm_3 (LSTM)                (None, 100, 64)           33024
_________________________________________________________________
dropout_3 (Dropout)          (None, 100, 64)           0
_________________________________________________________________
lstm_4 (LSTM)                (None, 64)                33024
_________________________________________________________________
dropout_4 (Dropout)          (None, 64)                0
_________________________________________________________________
dense (Dense)                (None, 1)                 65
=================================================================
Total params: 149,057
Trainable params: 149,057
Non-trainable params: 0
_________________________________________________________________
None
```

LSTM model was able to outperform persistence model by about 44.2% with the main benefit being that this model is able to generalize months into the future on unseen data.

6.4 Decision Recommendations

Improving energy production efficiency and ushering in the usage of more green energy production technology

An accurate demand model is critical for devising new strategies in energy production. Remember, the single limiting factor for energy production is that we cannot store power at scale and we would like to have just the right amount of energy ready at any given time to lower our expected production costs. Additionally, with this updated demand model we can also look for new ways to use more and more renewable energy sources and perhaps prioritize using those methods slightly more than in the past.

7. Conclusion

7.1 Key Takeaways

In summary, the findings of this analysis are as follows:

- We can leverage the results of our consumer behavior analysis to develop new consumer rebate programs to attempt to reduce wasteful energy usage from large office buildings and warehouses by investing in more efficient bulbs and light sensors. We can also look for ways to incentivize employers to allow employees to work from home a few days a month to also help lower energy demands
- By transitioning to the new XGBRegressor model to forecast energy production costs, the energy production company stands to save $82,922.76 annually.
- By transitioning to the new LSTM demand forecast model, we can begin strategizing new ways to incorporate renewable energy production techniques to meet our consumer demands, lower our costs and also lower our CO2 emissions and become an organization dedicated to a green future

7.2 Future Research

This was a very interesting analysis overall. Though the data did not contain too many features as I would have liked. If I were to conduct future research, I would like to collect the following pieces of information from customers:

- Unit type (apartment unit/house/office building, etc.)
- Address (latitude and longitude would also be helpful)
- Age and profession of consumer

These features would be helpful in better understanding patterns of behaviors in our initial consumer groups. We could also leverage geolocation data to better understand where in Spain there is more energy demand relative to other locations. This could provide clues as to additional strategies that can be used to help lower energy demand in Spain.