BCG Gamma Technical Challenge PowerCo_v5

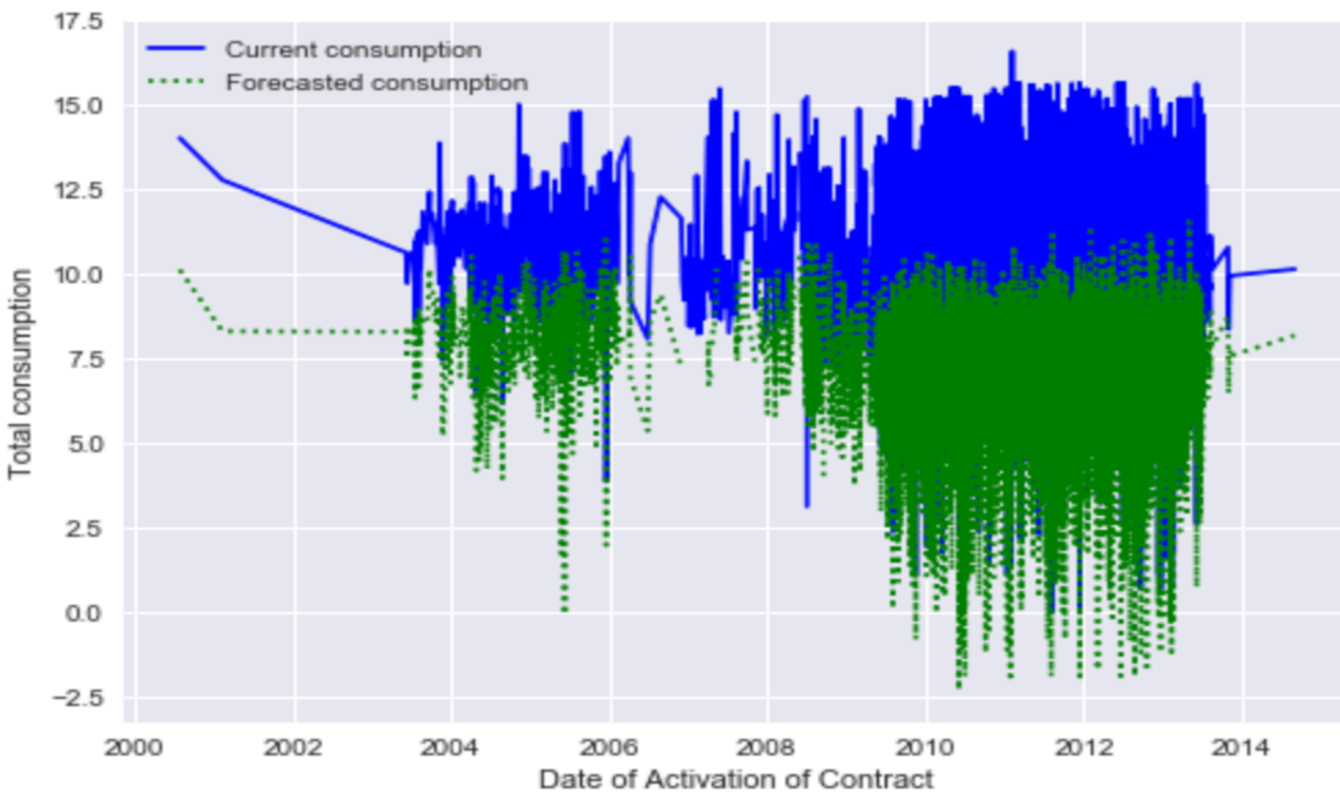
Samarth Goenka

Agenda

1. Scenario
2. Approach
3. Findings
4. Future implications
5. Appendix

Scenario – Decreasing forecasted consumption

With actual consumption following a similar trend to forecasted consumption, a recognized churn issue in the future means lower actual future consumption



So how do we find customers most likely to churn and prevent them from defecting to competitors?

Approach – Develop an algorithmic model that can predict customers most likely to churn

Model	Decision Tree	Logistic Regression	Bagging Classifier	Random Forest	Adaptive Boosting
Reason	Easy to implement; sets a good benchmark	Gives a smooth fit when working with continuous data	Averages out variances and biases of estimators, which is important due to high variance in data	Many randomized decision trees implemented together	Accounts for anomalies and outliers in the dataset by weighing classifiers based on performance
Time efficiency	1.34 s	4.60 s	1.38 s	5.91 s	9.93 s
Accuracy on training data (RoC)	0.63	0.59	0.93	0.96	1.00

Voting Classifier

Combines all 3 to give the majority classification

Margins and date of activation are the top 3 important factors in our model

Feature importance:

	Feature	Importance
16	margin_net_pow_ele	0.195327
15	margin_gross_pow_ele	0.129410
3	date_activ	0.088858
35	origin_up_lxidpiddsbxsbosboudacockeimpuepw	0.072867
2	cons_last_month	0.068439
6	date_renewal	0.065548
10	forecast_meter_rent_12m	0.064408
24	channel_sales_foosdfpfkusacimwkcsosbicdxkicaua	0.039183
18	net_margin	0.036204
0	cons_12m	0.034778

- Seems to have a high price elasticity, due to high importance to net margin and gross margin. This clearly follows the SME head's view of customer defects to competitors
- Contracts activated towards the end of the dataset timeline are more likely to result in customers churning. This may be due to changes in market landscape, such as increase in competition due to liberalization of European energy market
- The electricity campaign represented by lxidpiddsbxsbosboudacockeimpuepw has not performed well. We could look into issues related to transparency of campaign details, and levels of customer engagement
- Last month's consumption is also highly weighted with low consumption more likely to result in customers defecting. This implies that customers are gradually decreasing consumption before defecting

A closer look at sales channels tells us something important!

Feature importance:

	Feature	Importance
16	margin_net_pow_ele	0.195327
15	margin_gross_pow_ele	0.129410
3	date_activ	0.088858
35	origin_up_lxidpiddsbxsbosboudacockeimpuepw	0.072867
2	cons_last_month	0.068439
6	date_renewal	0.065548
10	forecast_meter_rent_12m	0.064408
24	channel_sales_foosdfpfkusacimwkcsosbicdxkicaua	0.039183
18	net_margin	0.036204
0	cons_12m	0.034778
19	num_years_antig	0.034065
11	forecast_price_energy_p1	0.030318
20	pow_max	0.022780
33	origin_up_kamkkxfxxuwbdslkwifmmcsiusiuosws	0.018741
14	imp_cons	0.018218
4	date_end	0.018001
5	date_modif_prod	0.015139
7	forecast_cons_12m	0.013133
8	forecast_cons_year	0.010551
12	forecast_price_energy_p2	0.008491
1	cons_gas_12m	0.008057
17	nb_prod_act	0.003887
25	channel_sales_lmkebamcaaclubfxadlmueccxoimlema	0.002034
13	forecast_price_pow_p1	0.001564
30	has_gas_t	0.000000
36	origin_up_unknown	0.000000
34	origin_up_ldkssxwpmemidmecebumciepifcamkci	0.000000
9	forecast_discount_energy	0.000000
32	origin_up_ewxeelcelemmiwuafmddpobolfuxioce	0.000000
31	origin_up_aabpopmuoobccoxasfsksebxoxffdcxs	0.000000
27	channel_sales_unknown	0.000000
29	has_gas_f	0.000000
28	channel_sales_usilxuppasemubllopkaafesmlibmsdf	0.000000
26	channel_sales_sddiedcslfslkckwlfkdpoeeailfpeds	0.000000
23	channel_sales_fixdbufsefwooaasfcxdxadsiekoceaa	0.000000
22	channel_sales_ewpakwlliwisiwduibdlfmalxowmwpci	0.000000
21	channel_sales_epumfxlbckeskwekxbiuasklxalciiuu	0.000000
37	origin_up_usapbepcfoloekilkwsdiboslwaxobdp	0.000000

Probability of sales channels for churned customers:

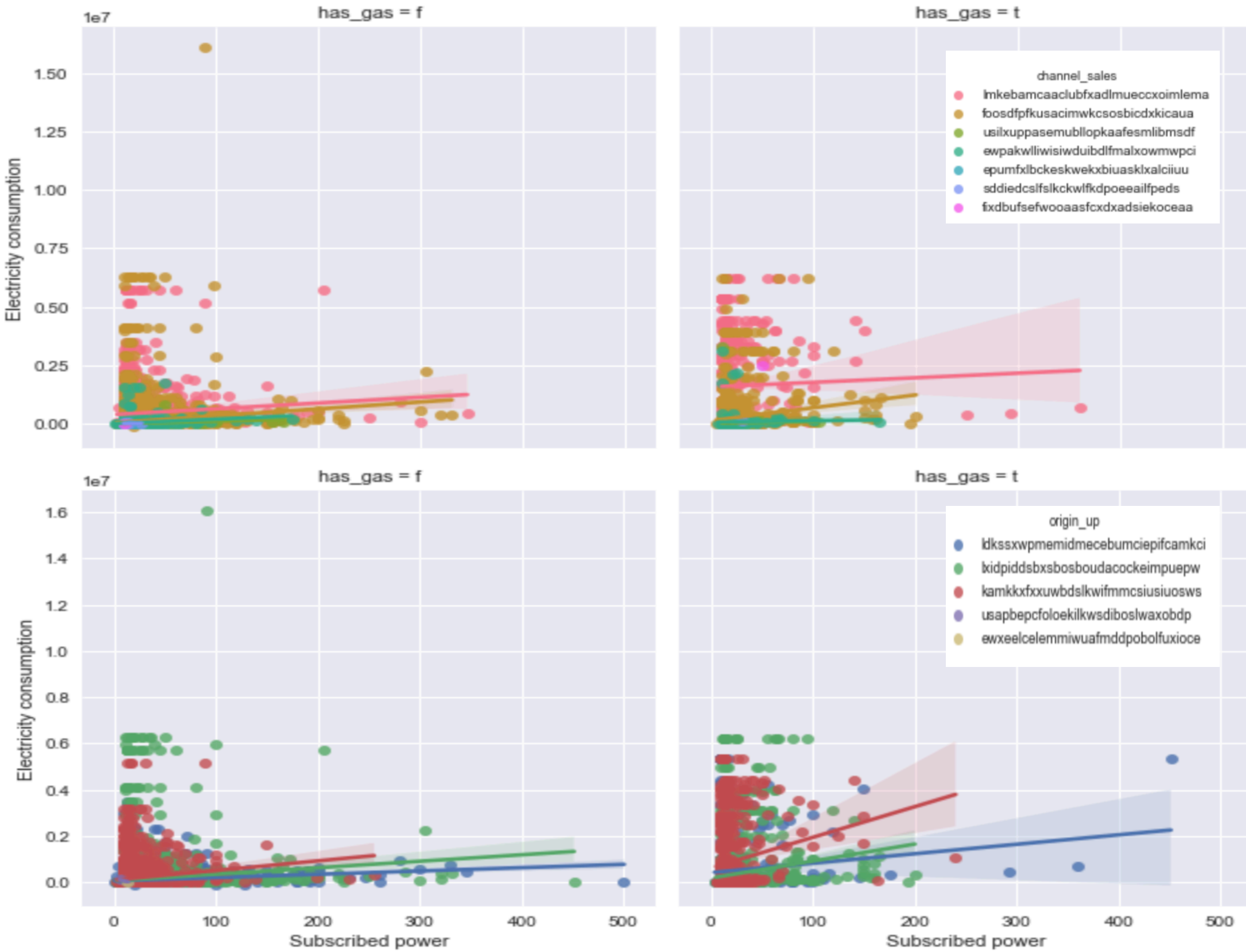
foosdfpfkusacimwkcsosbicdxkicaua	0.9
ewpakwlliwisiwduibdlfmalxowmwpci	0.1

9 out of 10 customers are consuming through this channel!

Most of the sales channels seem to not play a role in determining the churning of a customer.

However, the channel represented by <u>foosdfpfkusacimwkcsosbicdxkicaua</u> is significantly important in resulting in the customer defecting

Why higher power subscription incentives will not necessarily translate into higher consumption?



The correlation between overall subscribed power and electricity consumption:

Pearson correlation = 0.102
Kendall correlation = 0.282

This shows a weak positive correlation, and one that is not strong enough to consider resources spent to increase subscribed power in order to drive consumption

These graphs show the correlations broken down by sales channels and electricity campaigns

Does the 20% discount that keeps customers from defecting help PowerCo overall?

```
12 month consumption for churned customers:

40       342593
256       54072
1471     154251
1628      86232
1973      53819
2206      48667
2673      86232
3449      54072
3490     154251
3964     154251
```

```
Profit loss from churning customers:
$ 105100.70968491897

Profit recovered from churning customers with 20% discount:
$ 84080.56774793519
```

```
Forecasted 12 month consumption for churned customers:

40       1169.39
256       326.69
1471     1333.69
1628      302.46
1973      251.04
2206      273.93
2673      317.18
3449      292.13
3490     1160.26
3964        0.00
```

The model suggests that the 20% discount definitely makes sense for the forecasted time period!

```
Percentage difference:

40      -99.658665
256     -99.395824
1471    -99.135377
1628    -99.649249
1973    -99.533548
2206    -99.437134
2673    -99.632178
3449    -99.459739
3490    -99.247810
3964   -100.000000
```

However, what are its implications for the future?

What other analysis can be done to give PowerCo a concrete justified suggestion?

Realistically, PowerCo will not be able to go back to the original price after offering the 20% discount to the churned customers. Moreover, through word-of-mouth the customers that are not churning would threaten to defect if their prices were not discounted. Hence, we would have to recalculate the profits if the 20% discount was offered to all customers.

I would also look at the average time between date of activation and renewal of contracts for both churned and non-churned customers. Through this, we can infer a timeline of customer engagement, which can be used alongside proven campaigns.

We can assess which time periods the churning customers seem to be consuming their electricity in and apply a price discount for those periods only. Unlike a 20% discount on all prices, this method allows PowerCo to retain the churners and offer a 20% discount on that time period to other customers as well.

Although my classifiers worked well, I would use more classifiers, such as Naïve Bayes to look at the correlation between training features in an attempt to get a better accuracy on the training data. I can potentially develop a neural network to provide more efficient and accurate results.

Appendix

Assumptions

We cannot make inferences from columns that have a lot of empty cells/ missing values but we can generalize results from columns that do have a low proportion of empty cells

For columns, we do not know whether an empty cell implies missing value or nothing worth mentioning. Hence, we will fill them with "unknown" in the case of categorical data such as sales channel, and the mean value in the case of numerical data such as net margin.

We have no prior knowledge of which feature (factor) is more likely to affect the probability of a customer churning

We will assume that the training data is a good representation of the testing data, and therefore split our training data into random groups of training and testing data to test the accuracy of our model on at least the training data.

For the pricing data, we will assume that empty cells are the same as 0, and that a 0 means that there was no consumption in that time period. Further, we will assume that time time periods are three, 8 hour periods with period 1 representing the period of maximum consumption for each customer and not any one particular period. We will also assume that the prices are in cents and are monthly prices, with the prices for power being variable prices, and prices for energy being fixed prices.

We will then average the average monthly variable price to get the average annual variable price and sum the sum of monthly fixed prices to get the total fixed price.
Revenue = Annual var price * Consumption for 12 months + Total fixed price