Class 5b: Lift and Gains

Professor Blake McShane

MKTG 482: Customer Analytics
Kellogg School of Management

How do we assess a model's performance and compare it to that of different models?

MODEL PERFORMANCE MEASURES

General approach

- We are generally interested in models that predict or classify
- Use model to rank/score customers
- Calculate improvement in response over no targeting

We use the Tuango RFM case to illustrate how to calculate the Lift for the RFM prediction model

RFM DECILE SUMMARY

In this example: average response probability by rfm cell is the score

```
tuango_final %>%
  mutate(score_decile = ntile(-resp_rate_sq,10)) %>%
  group_by(score_decile) %>%
  summarise(num_cust=n(), num_buyers=sum(buyer))
```

Reverse scale so that decile 1 has best customers

```
## # A tibble: 10 x 3
##    score_decile num_cust num_buyers
##           <int>    <int>      <int>
##  1            1     1394        144
##  2            2     1394         84
##  3            3     1394         66
##  4            4     1394         51
##  5            5     1394         38
##  6            6     1394         21
##  7            7     1394         10
##  8            8     1394         16
##  9            9     1394          2
## 10           10     1393          0
```

These raw numbers can be used for the Lift calculations in Excel

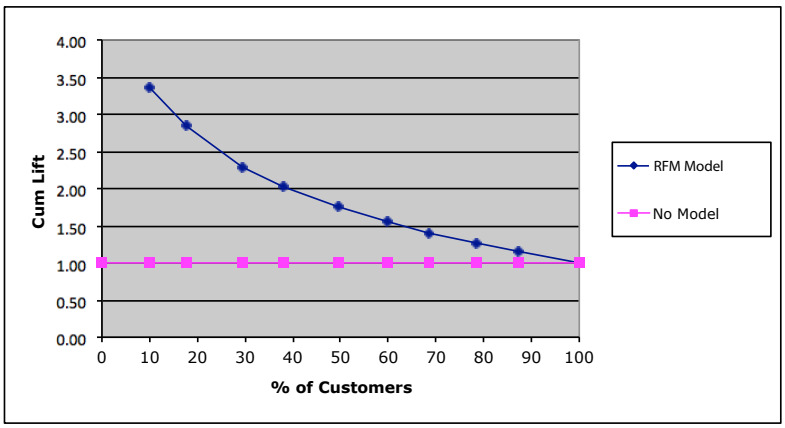
LIFT CALCULATIONS (RECENCY MODEL)

Score Decile	# Customers	Cum # Customers	Cum % Customers	# Buyers	Cum # Buyers	Response Rate	Lift	Cum Resp. Rate	Cum Lift
1	1394	1394	10.0%	144	144	10.3%	3.33	10.3%	3.33
2	1394	2788	20.0%	84	228	6.0%	1.94	8.2%	2.64
3	1394	4182	30.0%	66	294	4.7%	1.53	7.0%	2.27
4	1394	5576	40.0%	51	345	3.7%	1.18	6.2%	2.00
5	1394	6970	50.0%	38	383	2.7%	0.88	5.5%	1.77
6	1394	8364	60.0%	21	404	1.5%	0.49	4.8%	1.56
7	1394	9758	70.0%	10	414	0.7%	0.23	4.2%	1.37
8	1394	11152	80.0%	16	430	1.1%	0.37	3.9%	1.24
9	1394	12546	90.0%	2	432	0.1%	0.05	3.4%	1.11
10	1393	13939	100.0%	0	432	0.0%	0.00	3.1%	1.00
Total	13939			432		3.1%			

- Cumulative # customers: the number of total customers up to and including that decile
- Cumulative % customers: the percent of total customers up to and including that decile
- Cumulative # Buyers: the number of buyers up to and including that decile
- Response Rate: the actual response rate for each decile, computed by the number of buyers divided by the number of customers for each decile
- Lift: (response rate for each decile) / (overall response rate)
- Cumulative Response Rate: cumulative # buyers / cumulative # customers
- Cum(ulative) Lift: (cumulative response rate) / (overall response rate)

The Lift indicates the model's ability to beat the "no model"

CUMULATIVE LIFT CHARTS



- Lift for top decile=3.33: Targeting only these customers we expect to yield 3.33 times the number of buyer than if we did not target
- Note: Lift is relative index, e.g. 3.33 could refer to 4% or 40% response rate

To calculate gains we again begin with the raw numbers from the RFM model

RFM DECILE SUMMARY

In this example: average response probability by rfm cell is the score

```
tuango_final %>%
  mutate(score_decile = ntile(-resp_rate_sq,10)) %>%
  group_by(score_decile) %>%
  summarise(num_cust=n(), num_buyers=sum(buyer))
```

Reverse scale so that decile 1 has best customers

```
## # A tibble: 10 x 3
##    score_decile num_cust num_buyers
##           <int>    <int>      <int>
##  1            1     1394        144
##  2            2     1394         84
##  3            3     1394         66
##  4            4     1394         51
##  5            5     1394         38
##  6            6     1394         21
##  7            7     1394         10
##  8            8     1394         16
##  9            9     1394          2
## 10           10     1393          0
```

We make the Gains calculations in Excel

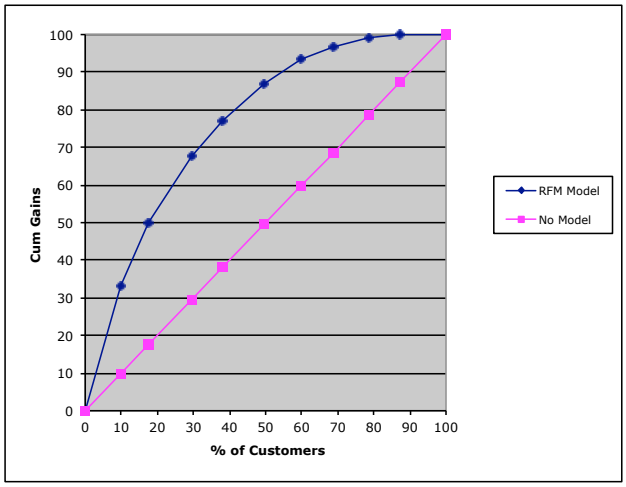
GAINS CALCULATIONS

Score Decile	# Customers	Cum # Customers	Cum % Customers	# Buyers	Cum # Buyers	Cum Gains
1	1394	1394	10.0%	144	144	33.3%
2	1394	2788	20.0%	84	228	52.8%
3	1394	4182	30.0%	66	294	68.1%
4	1394	5576	40.0%	51	345	79.9%
5	1394	6970	50.0%	38	383	88.7%
6	1394	8364	60.0%	21	404	93.5%
7	1394	9758	70.0%	10	414	95.8%
8	1394	11152	80.0%	16	430	99.5%
9	1394	12546	90.0%	2	432	100.0%
10	1393	13939	100.0%	0	432	100.0%
Total	13939			432		

- Gains: the proportion of responders (i.e. buyers) in each decile
- Cum(ulative) Gains: the proportion of responders/buyers up to and including the decile, or simply the sum of the gains up to that decile.

The Gains chart reveals what proportion of responders we can expect to gain from targeting a specific percent of customers using the model

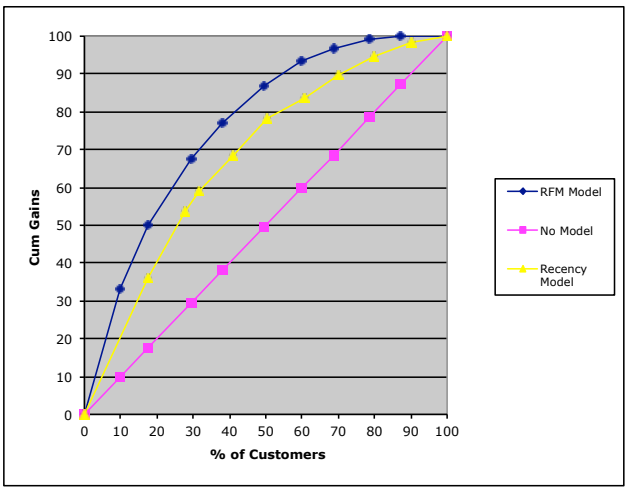
CUMULATIVE GAINS CHARTS



- By using the recency model to target customers, we can gain 33.3% of buyers by targeting 10% of the customers
- We can gain 76.7% of buyers by targeting 38% of the customers

Lift and Gains can also be used to compare two different models

EXAMPLE: RECENCY VS. FULL SEQUENTIAL N-TILE RFM MODEL



- Area under the curve (AUC) is a metric of model performance