Customer Churn Analysis – Cell2Cell

Kyle Suess

UMD

Marketing Analytics

Part 1 – The Final GLM Model

```
Coefficients:
              Estimate Std. Error z value Pr(>|z|)    
(Intercept)  2.811e+00  6.366e-01   4.415 1.01e-05 ***
MOU         -3.093e-04  3.457e-05  -8.948  < 2e-16 ***
RECCHRGE    -1.829e-03  5.564e-04  -3.287 0.001014 ** 
OVERAGE      1.310e-03  1.432e-04   9.143  < 2e-16 ***
ROAM         9.732e-03  1.890e-03   5.149 2.62e-07 ***
CHANGEM     -4.924e-04  5.291e-05  -9.306  < 2e-16 ***
CHANGER      2.356e-03  3.666e-04   6.429 1.29e-10 ***
DROPVCE      7.326e-03  1.502e-03   4.878 1.07e-06 ***
MONTHS      -2.370e-02  2.189e-03 -10.827  < 2e-16 ***
UNIQSUBS     1.926e-01  2.016e-02   9.551  < 2e-16 ***
ACTVSUBS    -2.251e-01  2.772e-02  -8.119 4.71e-16 ***
PHONES       5.310e-02  1.245e-02   4.264 2.01e-05 ***
EQPDAYS      1.313e-03  6.736e-05  19.488  < 2e-16 ***
CUSTOMER    -2.499e-06  5.768e-07  -4.333 1.47e-05 ***
AGE1        -4.066e-03  5.720e-04  -7.109 1.17e-12 ***
CHILDREN1    1.000e-01  2.634e-02   3.796 0.000147 ***
CREDITRTG2   5.363e-02  3.539e-02   1.515 0.129696    
CREDITRTG3   7.711e-02  3.849e-02   2.003 0.045142 *  
CREDITRTG4  -1.269e-01  4.528e-02  -2.804 0.005054 ** 
CREDITRTG5  -3.137e-01  4.512e-02  -6.954 3.56e-12 ***
CREDITRTG6  -1.453e-02  8.451e-02  -0.172 0.863483    
CREDITRTG7  -4.391e-02  6.139e-02  -0.715 0.474485    
REFURB1      2.423e-01  3.127e-02   7.747 9.44e-15 ***
WEBCAP1     -1.611e-01  3.750e-02  -4.297 1.73e-05 ***
MAILRES1    -1.380e-01  2.612e-02  -5.282 1.28e-07 ***
SETPRC       9.300e-04  2.343e-04   3.969 7.22e-05 ***
RETCALL      7.412e-01  5.764e-02  12.858  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Model:
CHURN = MOU + RECCHRGE + OVERAGE + ROAM + CHANGEM + CHANGER + DROPVCE + MONTHS + UNIQSUBS + ACTVSUBS + PHONES + EQPDAYS + CUSTOMER + AGE1 + CHILDREN + CREDITRTG* + REFURB + WEBCAP + MAILRES + SETPRC + RETCALL

AIC: 52429; D.O.F.: 28940 (Null);
Deviance: 53980 (Null), 52370 (Resid.)

Stepwise Binomial GLM with Logit Link

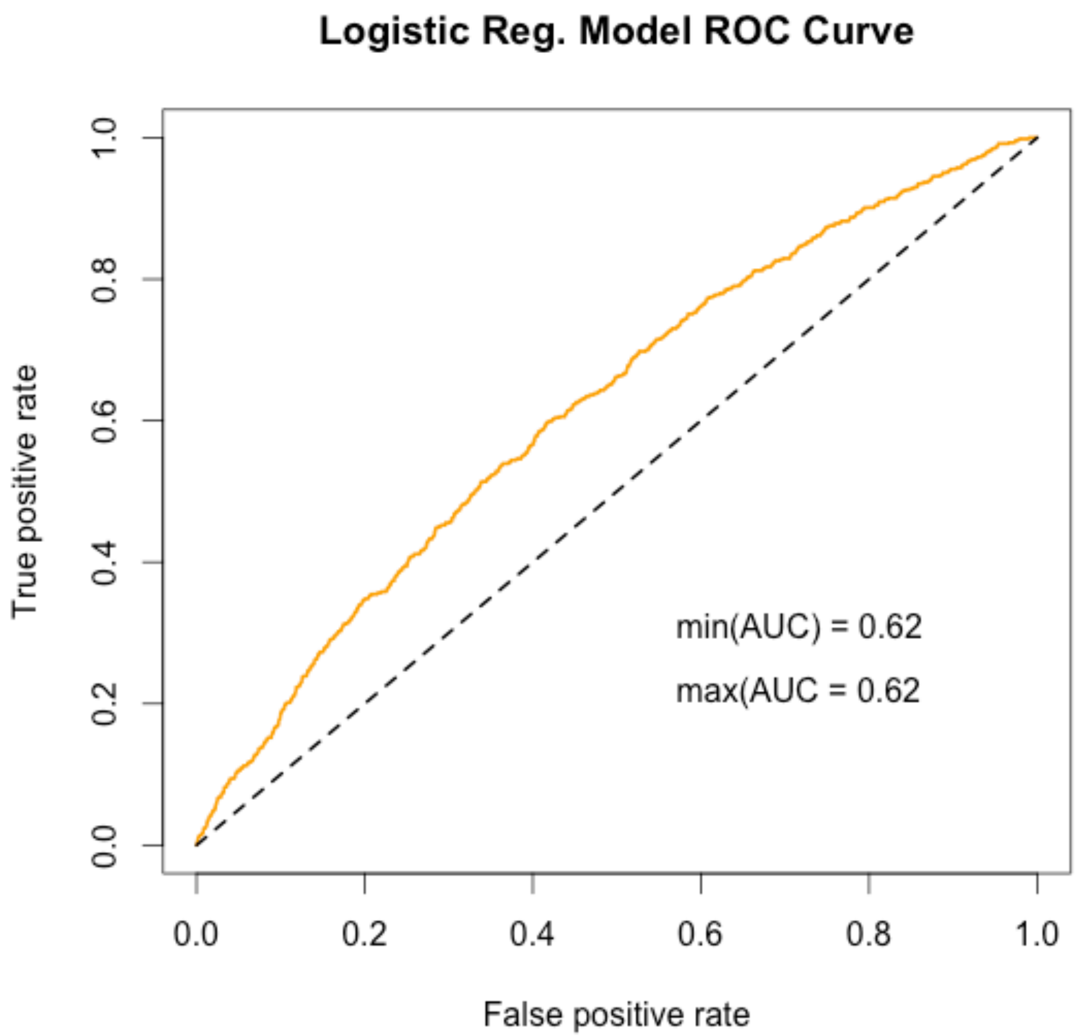
* = Created variable merging the 7 different credit score variables.

Part 1 – The Final GLM Model

Model (21 vars.):
CHURN = MOU + RECCHRGE + OVERAGE + ROAM + CHANGEM + CHANGER + DROPVCE + MONTHS + UNIQSUBS + ACTVSUBS + PHONES + EQPDAYS + CUSTOMER + AGE1 + CHILDREN + CREDITRTG + REFURB + WEBCAP + MAILRES + SETPRC + RETCALL

Essentially, explanatory variables for churn relate to: Mean monthly minutes use + Mean total recurring charge + Mean overage minutes of use + Mean number of roaming calls + % Change in minutes use + % Change in revenues + Mean number of dropped voice calls + Months in service + # Unique subscriptions + # Active subscriptions + # Handsets issued + Number of days of current equipment + Customer ID + Age of first HH member + Presence of children in HH + Credit rating + Handset is refurbished + Handset is web capable + Responds to mail offers + Handset price + Customer made call to retention team

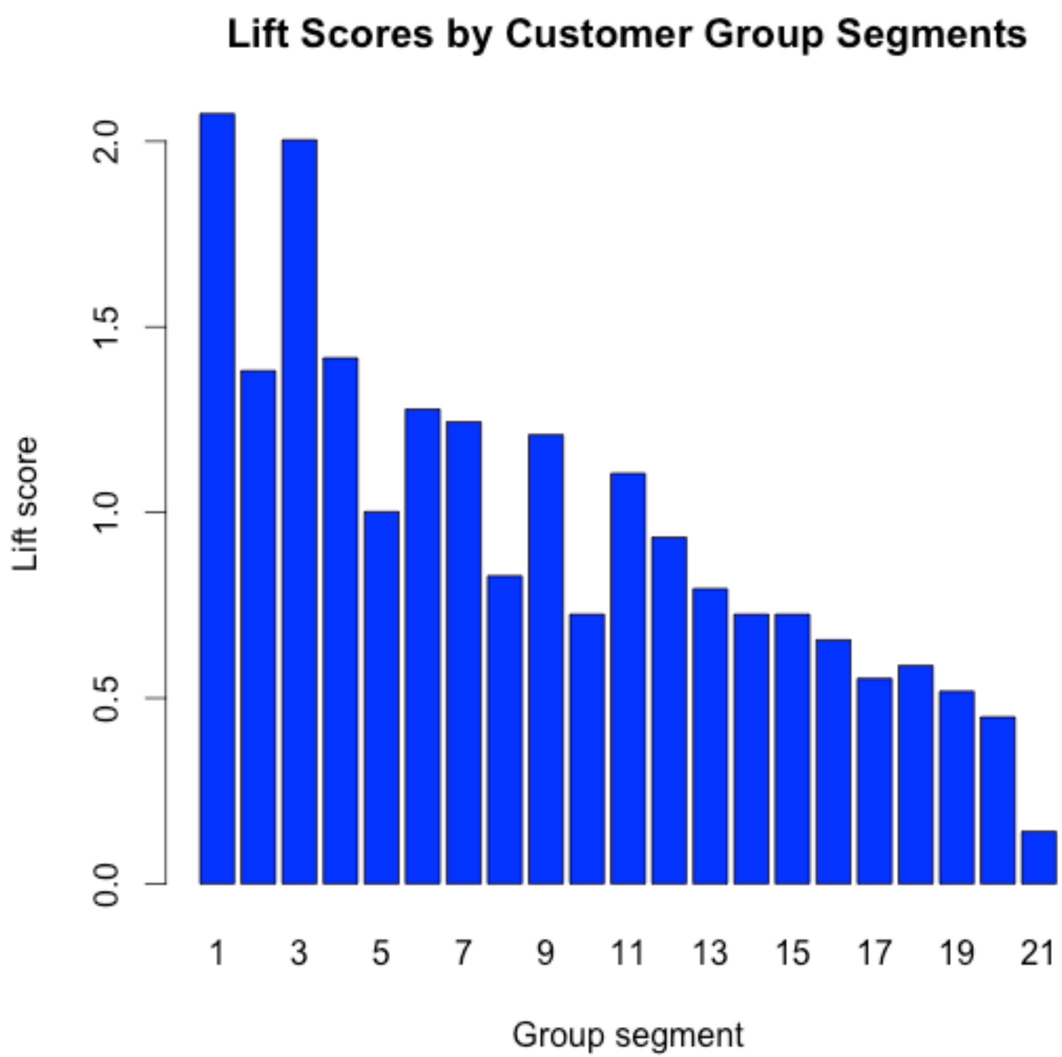
Part 1 – ROC Curve



Showing that the model is fairly accurate, but not amazing due to some inherent flaws present within our model estimation techniques (e.g. stepwise regression)

The AUC is also just "good"

Part 1 – Lift Chart



1^{st} Group of 1500 and 3^{rd} Group of 1500 have the highest lift scores among the 20 groups. These are the only two groups with lift above 1.75

Lift = Expected response from sample using our predictive model / Expected response from sample without using the model

Part 2 – Standardized Variable Estimates

```
Call:  glm(formula = CHURN ~ z.MOU + z.RECCHRGE + z.OVERAGE + z.ROAM +
    z.CHANGEM + z.CHANGER + z.DROPVCE + z.MONTHS + z.UNIQSUBS +
    z.ACTVSUBS + z.PHONES + z.EQPDAYS + z.CUSTOMER + z.AGE1 +
    c.CHILDREN + CREDITRTG + c.REFURB + c.WEBCAP + c.MAILRES +
    z.SETPRC + c.RETCALL, family = binomial(link = "logit"),
    data = Calibration)

Coefficients:
(Intercept)        z.MOU   z.RECCHRGE    z.OVERAGE       z.ROAM    z.CHANGEM    z.CHANGER    z.DROPVCE     z.MONTHS
    0.01218     -0.32384     -0.08643      0.25505      0.16035     -0.25435      0.18634      0.12983     -0.45331
 z.UNIQSUBS   z.ACTVSUBS     z.PHONES    z.EQPDAYS   z.CUSTOMER       z.AGE1   c.CHILDREN  CREDITRTG2  CREDITRTG3
    0.50785     -0.30507      0.14016      0.67005     -0.14337     -0.17937      0.10000      0.05363     0.07711
 CREDITRTG4   CREDITRTG5   CREDITRTG6   CREDITRTG7     c.REFURB     c.WEBCAP    c.MAILRES     z.SETPRC    c.RETCALL
   -0.12694     -0.31372     -0.01453     -0.04391      0.24226     -0.16113     -0.13795      0.10516      0.74117

Degrees of Freedom: 38940 Total (i.e. Null);  38914 Residual
Null Deviance:      53980
Residual Deviance: 52370        AIC: 52430
```

arm package in R used to standardize reg. coeffs

Top 10 Most influential variables: *RETCALL, EQPDAYS, UNIQSUBS, MONTHS, MOU, CREDITRTG5 (Low), ACTVSUBS, OVERAGE, CHANGEM, REFURB*

Future thought… can utilize random forests for variable importance (varImp)

Part 2 – Making It Actionable

Top 10 Most influential variables: *RETCALL, EQPDAYS, UNIQSUBS, MONTHS, MOU, CREDITRTG5 (Low Credit Score), ACTVSUBS, OVERAGE, CHANGEM, REFURB*

Actionable?
RETCALL: Yes. Help the customer sign a contract as easily as possible.
EQPDAYS: Yes. The firm can manage how long a customer must (or should) hold on to current equipment for through contracts.
UNIQSUBS: Maybe. Number of unique subscriptions is dependent on the customer, but there can be ways to better manage subscriptions.
ACTVSUBS: Yes. Similar to UNIQSUBS, a subscription incentive model could be used or subscriptions could be merged to one account.
MONTHS: Yes. Encourage members to stay through incentives.
CREDITRTG5: Maybe. Tradeoffs to accepting high-risk customers.
MOU: Yes. Encourage users to talk more through plans/offers.
OVERAGE: *Yes. Can better manage overage settings for customers.*
CHANGEM: *Yes. Encourage people to talk more each month.*
REFURB: *Yes. Instead of refurbished, offer other "new" or "used" phones.*

Part 2 - Recommendations

PHONE PLAN/ACCOUNT RELATED:

- Subscription management is very important. Monetary incentives are not required to manage subscriptions, but offering plan discounted rates to merge unique subscriptions together, or combine multiple active subscriptions to one billing account for ease (and since people will less likely quit the acct.)
- Encourage people to stay longer through incentives. As a two-year contract is closing, offer a “special” monthly rate for re-signing and develop a loyalty program for discounts on new smartphones and early access to new phone release upgrades (e.g. offer loyalty members early pre-ordering on new iPhones).
- If a customer makes the first move towards re-signing a contract, be very receptive to the customer and make the process very easy (5 min. or less) over the phone or online.

Part 2 - Recommendations

PLAN MINUTES RELATED:

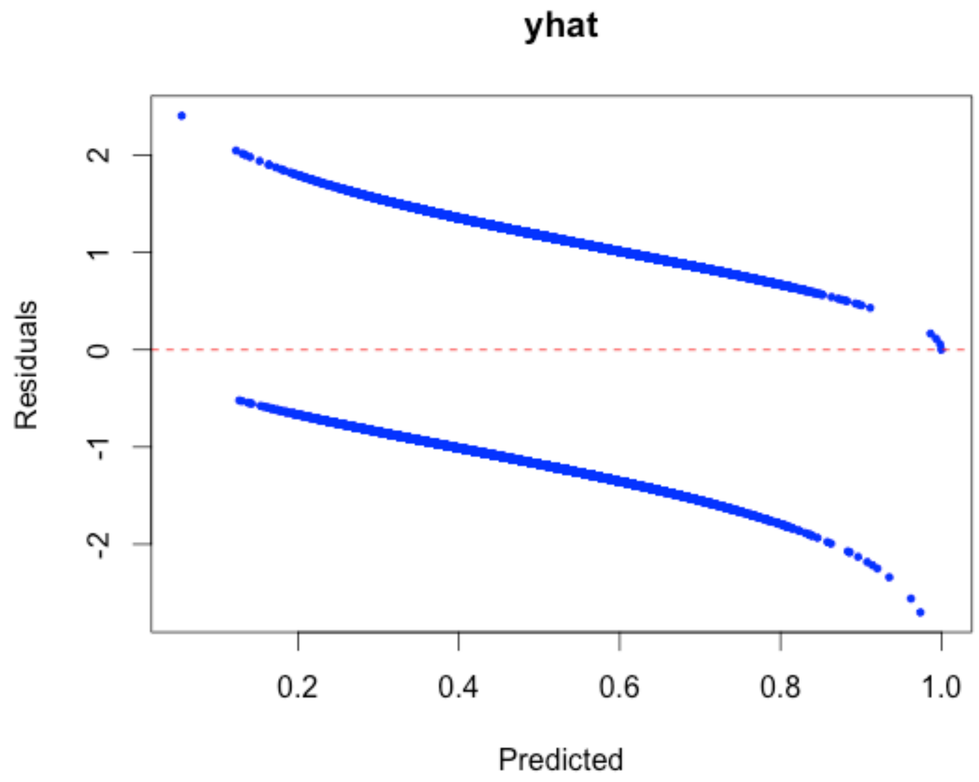
- The primary goal should be to get consumers to utilize their phone plans but not go over their allotted minutes and regretfully pay extra. One incentive could be no overage charges for talk/ text. Bundling an unlimited talk or unlimited text add-on feature to plans could also suffice.

PHONE RELATED:

- The firm should provide consumers with options to trade-in a phone for a discount on a new phone or a comparable used phone as opposed to a refurbished phone. Perhaps "refurbished" has a negative connotation with users and would incline them to switch carriers.

Appendix – Model VIF and Residuals

```
> vif(regmodel)
              GVIF Df GVIF^(1/(2*Df))
MOU       2.911908  1        1.706431
RECCHRGE  1.566152  1        1.251460
OVERAGE   1.565333  1        1.251133
ROAM      1.021479  1        1.010682
CHANGEM   1.658880  1        1.287975
CHANGER   1.647443  1        1.283527
DROPVCE   1.585867  1        1.259312
MONTHS    3.986617  1        1.996651
UNIQSUBS  2.822542  1        1.680042
ACTVSUBS  2.802703  1        1.674127
PHONES    2.463902  1        1.569682
EQPDAYS   2.684469  1        1.638435
CUSTOMER  2.528056  1        1.589986
AGE1      1.486707  1        1.219306
CHILDREN  1.213816  1        1.101733
CREDITRTG 1.679303  6        1.044145
REFURB    1.125471  1        1.060882
WEBCAP    1.200454  1        1.095652
MAILRES   1.500627  1        1.225001
SETPRC    1.633705  1        1.278165
RETCALL   1.019650  1        1.009777
```



Appendix – Model Variable Odds Ratio with 95% Confidence Interval

```
> print(oddcal)
```

	OR	2.5 %	97.5 %
(Intercept)	16.6234477	4.7766876	57.9495541
MOU	0.9996907	0.9996229	0.9997584
RECCHRGE	0.9981730	0.9970842	0.9992614
OVERAGE	1.0013104	1.0010313	1.0015934
ROAM	1.0097794	1.0061651	1.0136212
CHANGEM	0.9995078	0.9994039	0.9996112
CHANGER	1.0023592	1.0016435	1.0030831
DROPVCE	1.0073524	1.0044001	1.0103314
MONTHS	0.9765746	0.9723838	0.9807653
UNIQSUBS	1.2123603	1.1655568	1.2614223
ACTVSUBS	0.7984455	0.7561436	0.8429576
PHONES	1.0545365	1.0291135	1.0806192
EQPDAYS	1.0013137	1.0011817	1.0014461
CUSTOMER	0.9999975	0.9999964	0.9999986
AGE1	0.9959424	0.9948262	0.9970592
CHILDREN1	1.1051664	1.0495625	1.1637357
CREDITRTG2	1.0550981	0.9843797	1.1308906
CREDITRTG3	1.0801625	1.0016754	1.1648157
CREDITRTG4	0.8807836	0.8059636	0.9625022
CREDITRTG5	0.7307231	0.6688486	0.7982445
CREDITRTG6	0.9855740	0.8350419	1.1631369
CREDITRTG7	0.9570433	0.8485306	1.0794170
REFURB1	1.2741316	1.1984144	1.3547205
WEBCAP1	0.8511814	0.7908061	0.9160435
MAILRES1	0.8711415	0.8276598	0.9168889
SETPRC	1.0009305	1.0004709	1.0013904
RETCALL	2.0983854	1.8753010	2.3508587

```
oddcal <- exp(cbind(OR = coef(regmodel), confint(regmodel)))
```

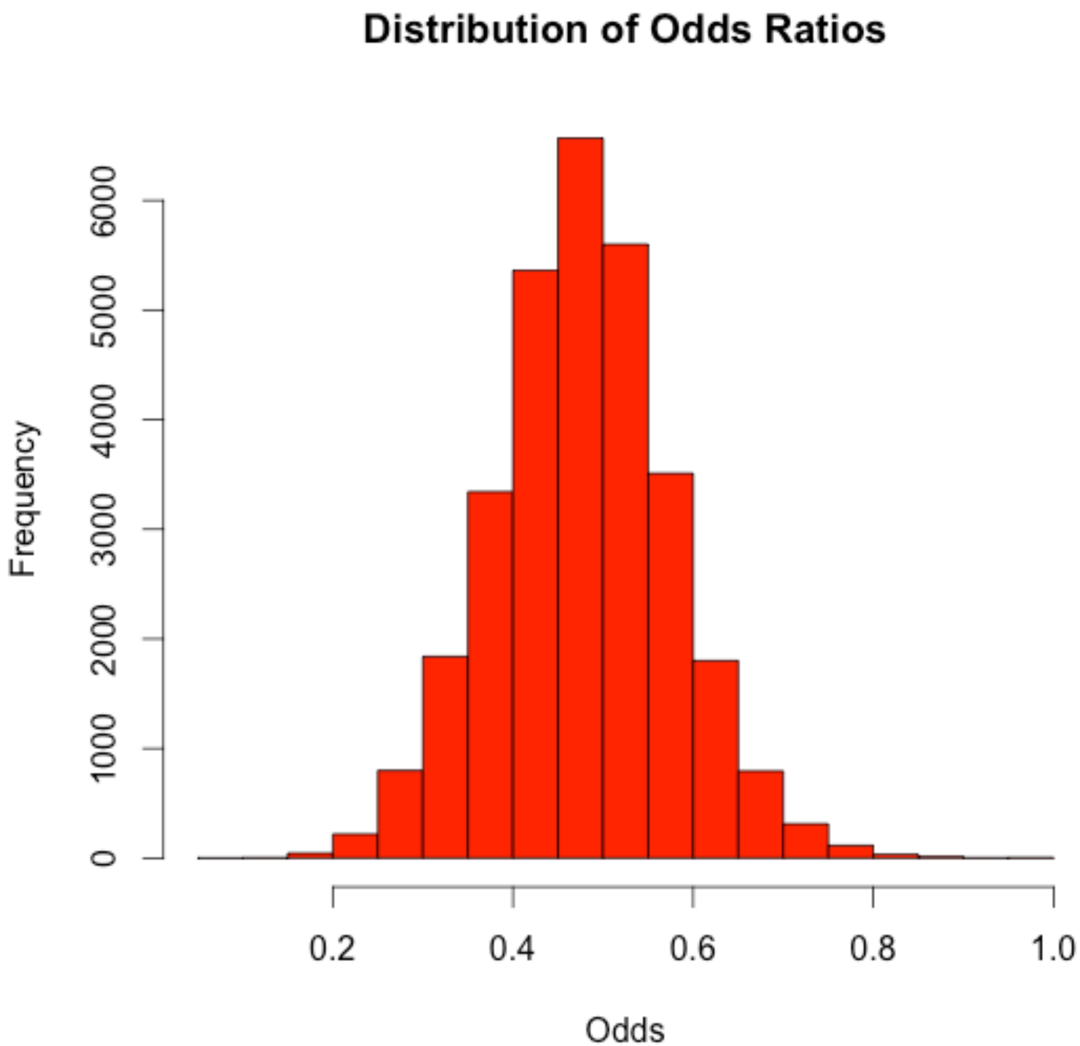
Odds-Ratio:

exp(b) =1 : indicates no change in odds of event occurring.

exp(b) >1: indicates increase in odds of event occurring.

exp(b) <1: indicates decrease in odds of event occurring.

Appendix – Odds Ratio Plot for Customer Churn



Appendix – Future Model Considerations

A feasible alternative to stepwise logistic reg. would be Bayesian Model Averaging (BMA). BMA accounts for the uncertainty inherent in the model selection process, which typical logistic reg. neglects. By averaging over many different competing models, BMA incorporates model uncertainty into conclusions about parameters and prediction.

```
#BMA model for the glm with logit link and factor.type=TRUE (models will contain either all or non of dummy vars)
output <- bic.glm(predictors, dfcal2$CHURN, data=dfcal2, glm.family="binomial", factor.type=TRUE)
summary(output)
imageplot.bma(output)
#posterior probabilities of each model
output$postprob
#the variables in the models
output$label
#probability a variable should be in the model
output$probne0
#bayesian model averaged means for each variable
output$postmean
#bayesian model averaged std. devs for each variable
output$postsd
#model by model estimates for confounding checks to see the associations of a var with independent and dependent vars.
output$mle
#standard error of each coefficient in model
output$se
```

Appendix - References

- http://www.stat.columbia.edu/~gelman/research/unpublished/standardizing.pdf
- http://rstudio-pubs-static.s3.amazonaws.com/2897_9220b21cfc0c43a396ff9abf122bb351.html
- http://www.nesug.org/proceedings/nesug07/sa/sa07.pdf
- http://web.stanford.edu/~hastie/TALKS/enet_talk.pdf
- http://rocr.bioinf.mpi-sb.mpg.de/
- http://cran.r-project.org/doc/contrib/Sharma-CreditScoring.pdf
- http://www.soc.iastate.edu/sapp/soc512LogisticNotes.pdf