

CHURN

- **Dataset: CHURN**
- **OBJECTIVE:** Find out why customer has moved to other providers, determine what type of the customers should we send promotional materials to make increase their chance to stay with the company.

| | |
|--------------|---------------------|
| Tianyi Chang | changty1994@gwu.edu |
| Elvis Leng | eleng@gwu.edu |
| Jiafeng Xu | jiafengx@gwu.edu |
| Jingbo Zhang | bonnie_jz@gwu.edu |
| Shuwen Wang | shuwen_wang@gwu.edu |
| Xinran Tao | xinrantao@gwu.edu |



CHURN - Executive Summary

- The business problem and the proposed solution



01

Customer is Goal 14.5% > 5%

02

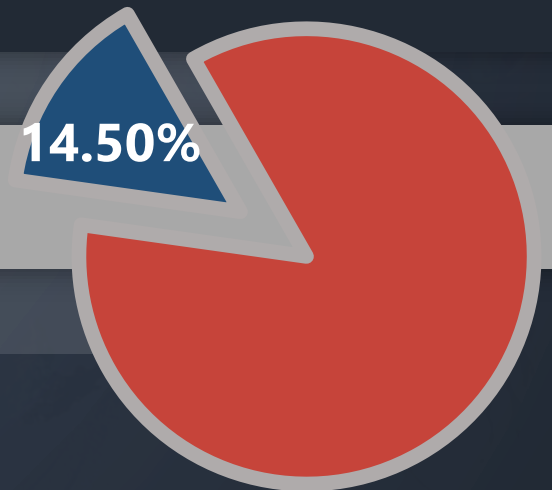
Models ---> Patterns ---> Target Group

03

Marketing Solutions

04

Churn Rate 14.5 ↓



CHURN - Introduction

George Washington University, School of Business
Introduction to Business Analytics

- The problem, why is it important, and the role of analytics in solving the problem



**Customer
Retention**



**Introduction to
the Dataset**



**Problems to
Solve**

CHURN - Methods

- The data, exploratory analysis, and the modeling approach or approaches used

Data Subset



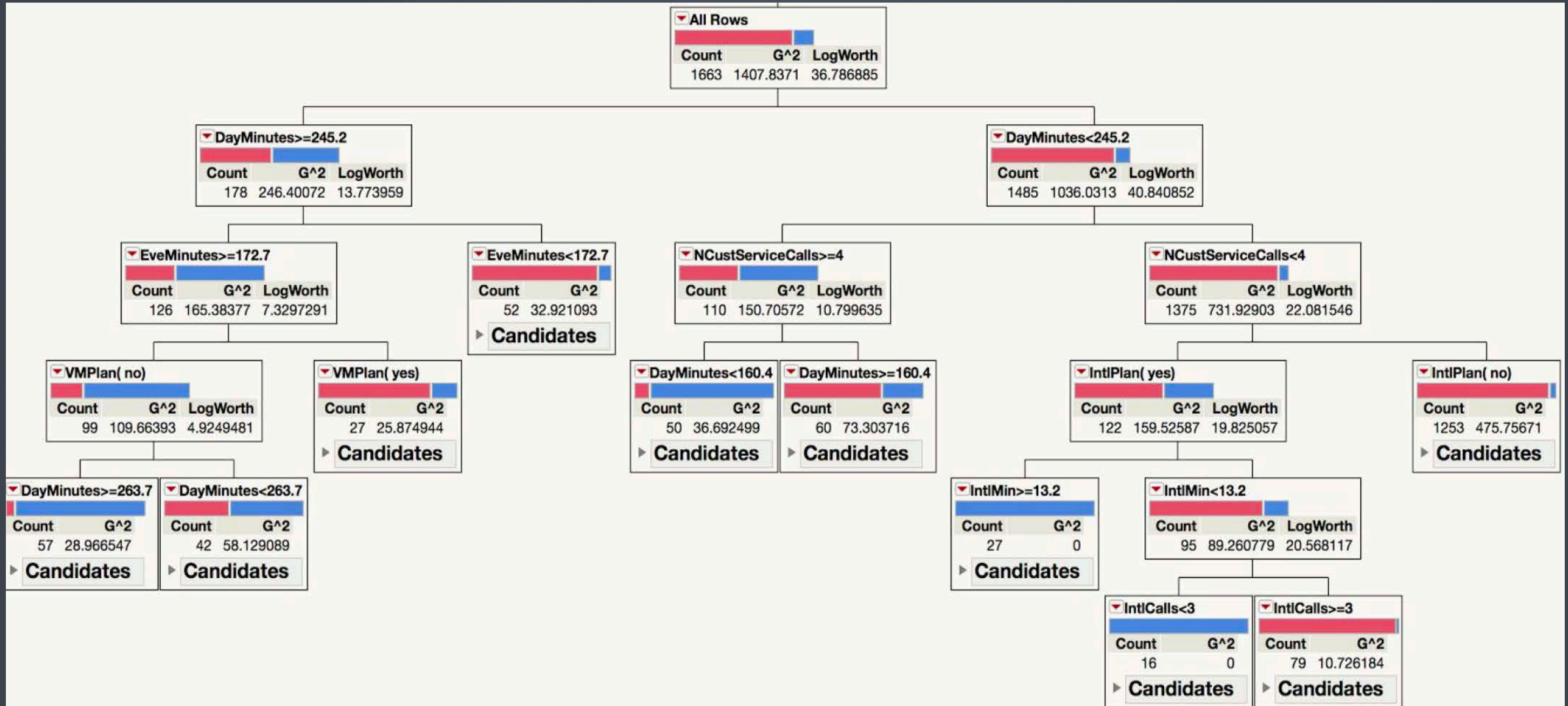
Missing Value



Mosaic Plot



CHURN – Decision Tree



CHURN - Results

- The selected model and model performance

Column Contributions

| Term | Number of Splits | G^2 | Portion |
|-------------------|------------------|------------|---------|
| DayMinutes | 3 | 188.68281 | 0.2835 |
| NCustServiceCalls | 1 | 153.396571 | 0.2305 |
| IntlPlan | 1 | 96.6464529 | 0.1452 |
| IntlCalls | 1 | 78.534595 | 0.1180 |
| IntlMin | 1 | 70.2650914 | 0.1056 |
| EveMinutes | 1 | 48.095861 | 0.0723 |
| VMPlan | 1 | 29.8448974 | 0.0448 |
| State | 0 | 0 | 0.0000 |
| AcctLength | 0 | 0 | 0.0000 |

| Measure | Training | Validation | Test |
|------------------------|----------|------------|--------|
| Misclassification Rate | 0.0710 | 0.0678 | 0.0712 |

Confusion Matrix

| | | Test | |
|--------|-------|-----------------|------|
| Actual | Churn | Predicted Count | |
| | | False | True |
| False | | 703 | 22 |
| True | | 38 | 80 |

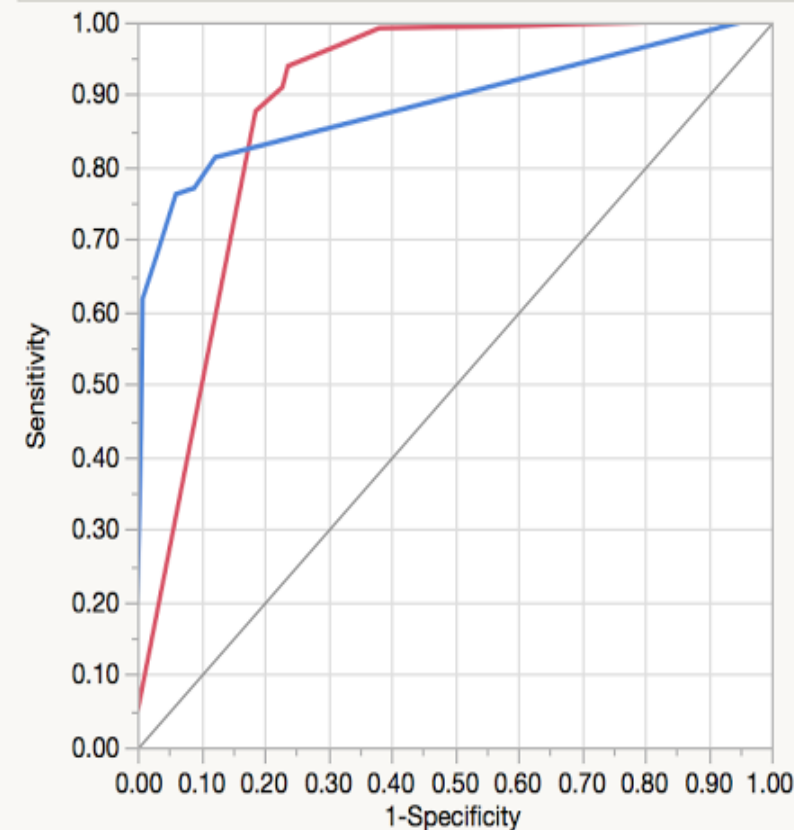
Misclassification Rate

$$= (FP+FN)/total=0.07$$

$$FPR=22/(22+703)= 0.031$$

$$FNR=38/(38+80)= 0.322$$

Receiver Operating Characteristic on Test Data



| Churn | Area |
|-------|--------|
| False | 0.8877 |
| True | 0.8877 |

CHURN - Results

- The selected model and model performance

| Test | | | |
|--------------|----------------|------|--|
| Actual Churn | Decision Count | | |
| | False | True | |
| False | 660 | 65 | |
| True | 27 | 91 | |

| Actual Churn | Decision Rate | | |
|--------------|---------------|-------|--|
| | False | True | |
| False | 0.910 | 0.090 | |
| True | 0.229 | 0.771 | |

| Misclassification Rate | |
|------------------------|--|
| 0.1091 | |



PROFIT MATRIX

| | TRUE | FALSE | CONVERSION RATE 80% |
|-------|------|--------|------------------------|
| TRUE | 4.40 | -10.00 | |
| FALSE | 8.00 | 10.00 | |

ORIGINAL

| | TRUE | FALSE | PROFIT 6070.00 |
|-------|------|-------|-------------------|
| TRUE | 0 | 118 | |
| FALSE | 0 | 725 | |

DECISION TREE(prob=0.5)

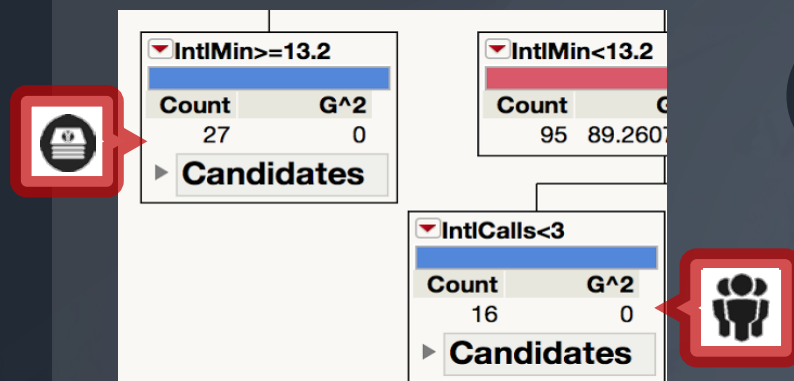
| | TRUE | FALSE | PROFIT 7178.00 | INCREASING RATE 18% |
|-------|------|-------|-------------------|------------------------|
| TRUE | 80 | 38 | | |
| FALSE | 22 | 703 | | |

DECISION TREE(best profit)

| | TRUE | FALSE | PROFIT 7250.40 | INCREASING RATE 19% |
|-------|------|-------|-------------------|------------------------|
| TRUE | 91 | 27 | | |
| FALSE | 65 | 660 | | |

CHURN – Target Group

- Link to business goal fit and key management insights



01

have **international plan**, international minute ≥ 13.2

02

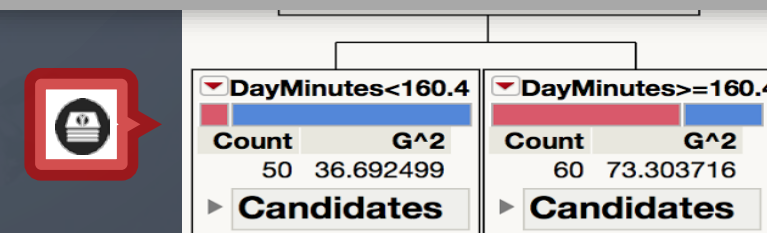
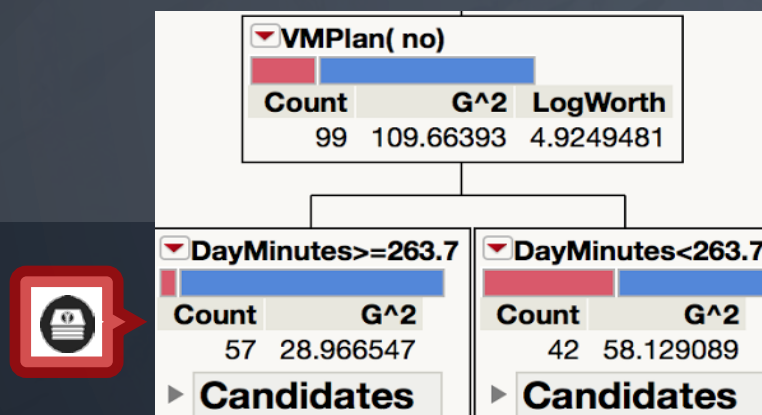
international minute < 13.2 , international call < 3

03
04

day minute/evening minute **no voicemail plan**

05

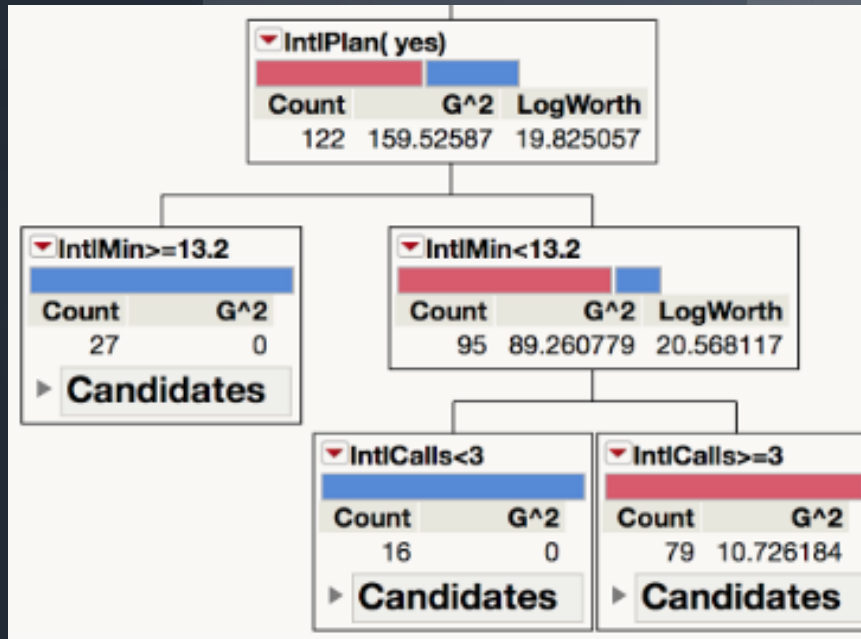
day minute < 160.4 , customer service call ≥ 4



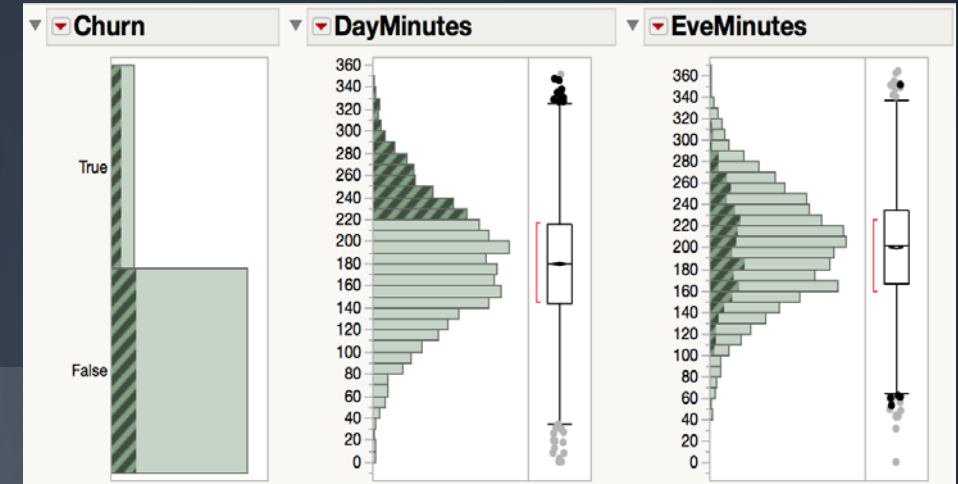
CHURN - Conclusion

- Link to business goal fit and key management insights

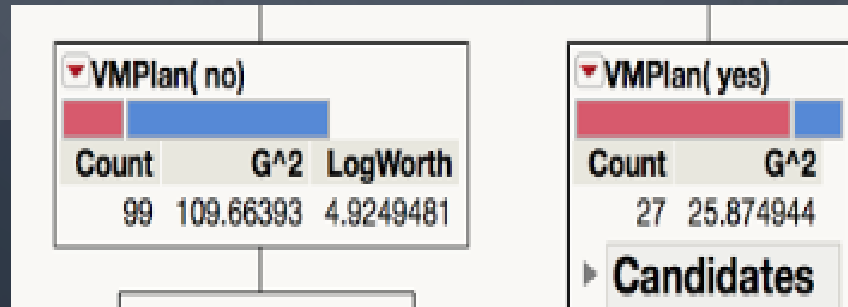
International Plans



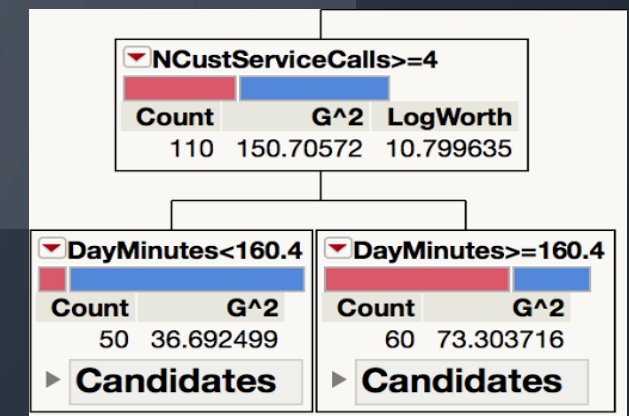
DayCall Plans



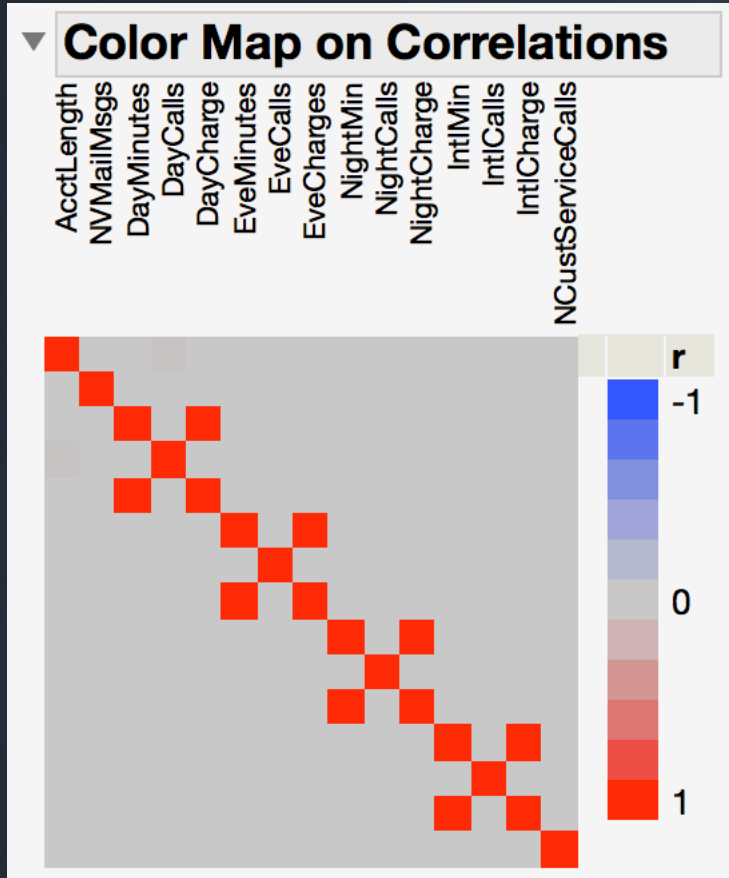
Voicemail Plans



Customer Service



CHURN - Appendix

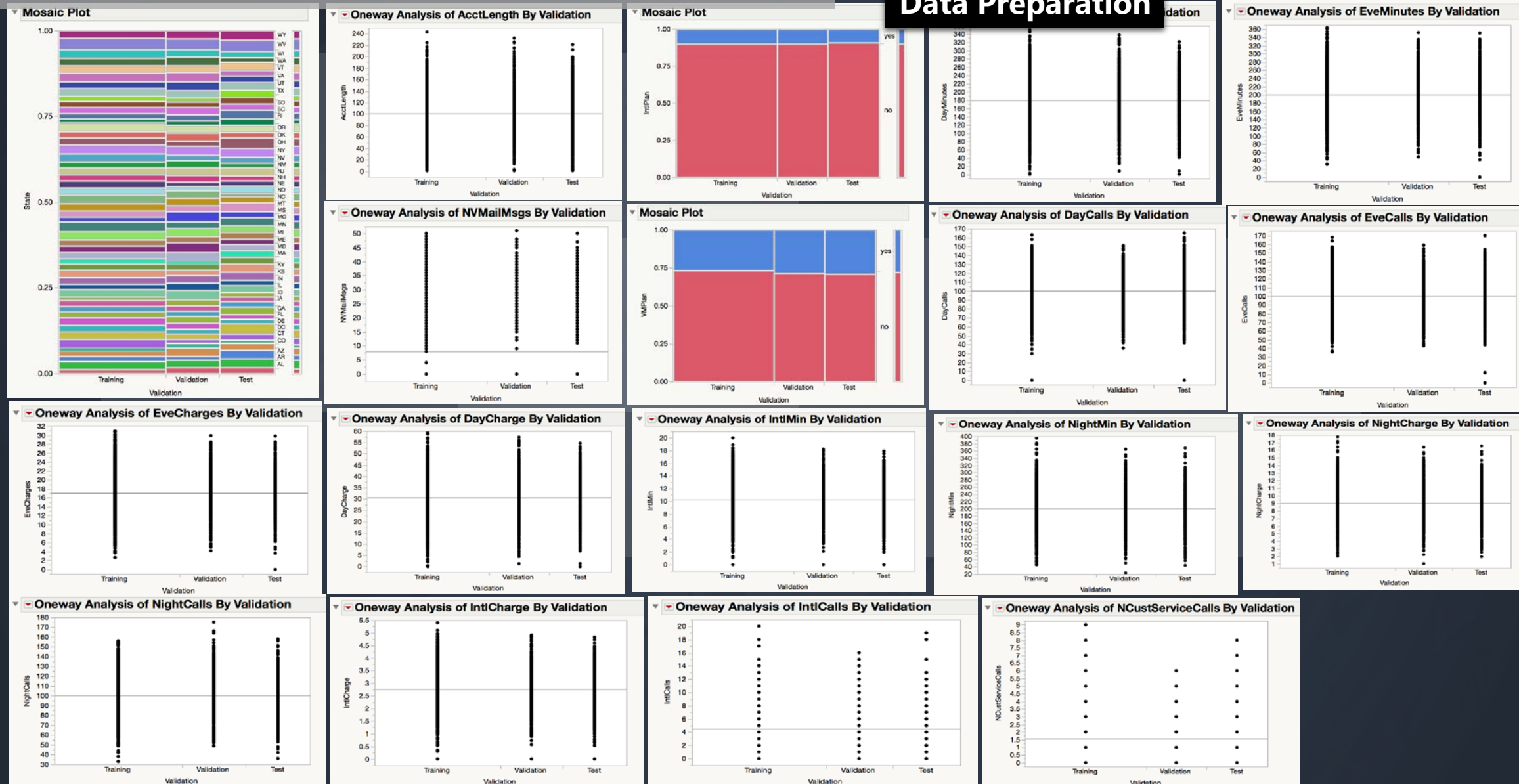


- 1. no missing value
- 2. all variables almost normally distributed
- 3. no significant outliers
- 4. no linear relationship(except for time and charge)

Data Preparation

CHURN - Appendix

Data Preparation



CHURN - Appendix

Logistic Model

Step History

| Step | Parameter | Action | ChiSquare | L-R "Sig Prob" | RSquare | p | AICc | BIC | RSquare Validation |
|------|--|----------|-----------|-------------------|---------|----|---------|---------|-----------------------|
| 1 | IntlPlan{ no- yes} | Entered | 86.03989 | 0.0000 | 0.0611 | 2 | 1325.8 | 1336.63 | 0.0649 |
| 2 | NCustServiceCalls | Entered | 71.53292 | 0.0000 | 0.1119 | 3 | 1256.28 | 1272.51 | 0.1208 |
| 3 | DayCharge | Entered | 75.82018 | 0.0000 | 0.1658 | 4 | 1182.47 | 1204.11 | 0.1902 |
| 4 | State{NH&NC&CT&MA&PA&MN&NY&KS&AR-MT&ME&NV&WA&MS&MI&SC&MD&TX&NJ&CA} | Entered | 30.76574 | 0.0000 | 0.1876 | 6 | 1155.73 | 1188.18 | 0.2092 |
| 5 | EveCharges | Entered | 14.30701 | 0.0002 | 0.1978 | 7 | 1143.44 | 1181.29 | 0.2295 |
| 6 | VMPlan{ yes- no} | Entered | 12.10321 | 0.0005 | 0.2064 | 8 | 1133.36 | 1176.6 | 0.2445 |
| 7 | IntlCharge | Entered | 9.875194 | 0.0017 | 0.2134 | 9 | 1125.5 | 1174.14 | 0.2483 |
| 8 | NightMin | Entered | 6.675634 | 0.0098 | 0.2182 | 10 | 1120.85 | 1174.88 | 0.2545 |
| 9 | State{UT-OR} | Entered | 12.10104 | 0.0005 | 0.2267 | 15 | 1118.91 | 1199.86 | 0.2347 |
| 10 | DayCalls | Entered | 3.899092 | 0.0483 | 0.2295 | 16 | 1117.05 | 1203.38 | 0.2295 |
| 11 | State{ID-IN&FL&OH} | Entered | 4.94998 | 0.0261 | 0.2330 | 18 | 1116.18 | 1213.26 | 0.2233 |
| 12 | IntlCalls | Entered | 3.073316 | 0.0796 | 0.2352 | 19 | 1115.16 | 1217.61 | 0.2314 |
| 13 | IntlMin | Entered | 2.070011 | 0.1502 | 0.2367 | 20 | 1115.14 | 1222.95 | 0.2267 |
| 14 | EveCalls | Entered | 1.737124 | 0.1875 | 0.2379 | 21 | 1115.45 | 1228.63 | 0.2215 |
| 15 | State{VT-MO} | Entered | 2.278775 | 0.1312 | 0.2395 | 23 | 1117.28 | 1241.18 | 0.2155 |
| 16 | NVMailMsgs | Entered | 0.807024 | 0.3690 | 0.2401 | 24 | 1118.53 | 1247.79 | 0.2173 |
| 17 | AcctLength | Entered | 0.546421 | 0.4598 | 0.2405 | 25 | 1120.05 | 1254.66 | 0.2167 |
| 18 | DayMinutes | Entered | 0.350401 | 0.5539 | 0.2407 | 26 | 1121.76 | 1261.73 | 0.2160 |
| 19 | Best | Specific | . | . | 0.2182 | 10 | 1120.85 | 1174.88 | 0.2545 |

Parameter Estimates

| Term | Estimate | Std Error | ChiSquare | Prob>ChiSq |
|-------------------|------------|-----------|-----------|------------|
| Intercept | -7.8164664 | 0.7121348 | 120.47 | <.0001* |
| IntlPlan[no] | -0.9854719 | 0.1004958 | 96.16 | <.0001* |
| VMPlan[no] | 0.35364128 | 0.0969591 | 13.30 | 0.0003* |
| DayCharge | 0.07466578 | 0.0087593 | 72.66 | <.0001* |
| EveCharges | 0.08026481 | 0.0184009 | 19.03 | <.0001* |
| NightMin | 0.00395829 | 0.0015618 | 6.42 | 0.0113* |
| IntlCharge | 0.35544816 | 0.1056853 | 11.31 | 0.0008* |
| NCustServiceCalls | 0.51217281 | 0.0550193 | 86.66 | <.0001* |

For log odds of True/ False

Fit Details

| Measure | Training | Validation | Test | Definition |
|------------------------|----------|------------|--------|--|
| Entropy RSquare | 0.4702 | 0.4727 | 0.4913 | 1-Loglike(model)/Loglike(0) |
| Generalized RSquare | 0.5750 | 0.5727 | 0.5914 | $(1-(L(0)/L(model))^{2/n})/(1-L(0)^{2/n})$ |
| Mean -Log p | 0.2242 | 0.2128 | 0.2060 | $\sum -\log(p_{ij})/n$ |
| RMSE | 0.2427 | 0.2364 | 0.2319 | $\sqrt{\sum (y_{ij}-p_{ij})^2/n}$ |
| Mean Abs Dev | 0.1195 | 0.1203 | 0.1166 | $\sum y_{ij}-p_{ij} /n$ |
| Misclassification Rate | 0.0710 | 0.0678 | 0.0712 | $\sum (p_{ij} \neq p_{Max})/n$ |
| N | 1663 | 826 | 843 | n |

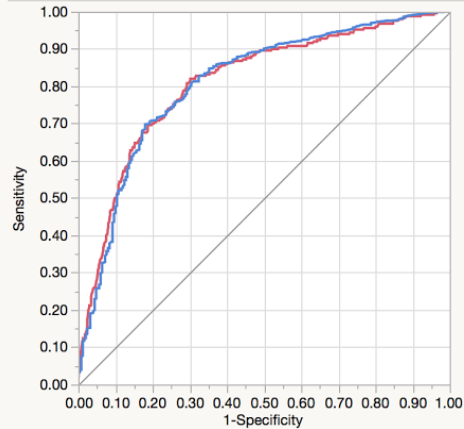
Confusion Matrix

| Training | | | Validation | | | Test | | |
|----------|-----------|-------|------------|-----------|-------|--------|-----------|-------|
| Actual | Predicted | | Actual | Predicted | | Actual | Predicted | |
| | True | False | | True | False | | True | False |
| Churn | True | False | Churn | True | False | Churn | True | False |
| True | 44 | 206 | True | 30 | 85 | True | 22 | 96 |
| False | 36 | 1377 | False | 19 | 692 | False | 20 | 705 |

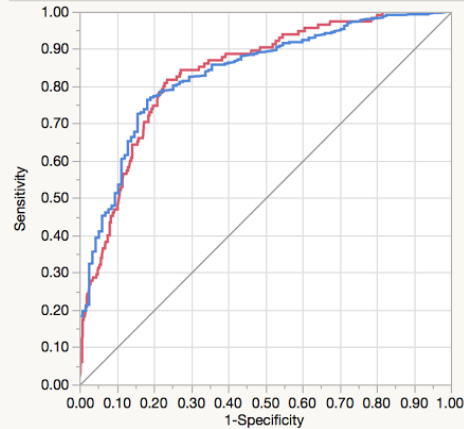
CHURN - Appendix

Logistic Model

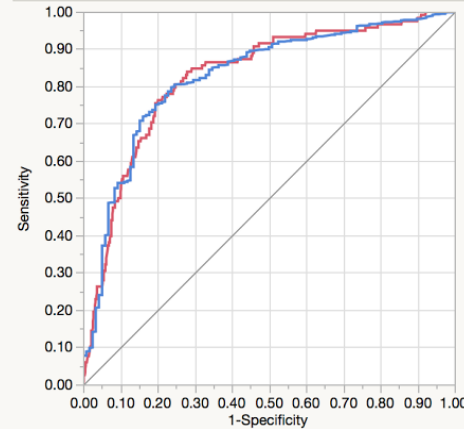
▼ Receiver Operating Characteristic



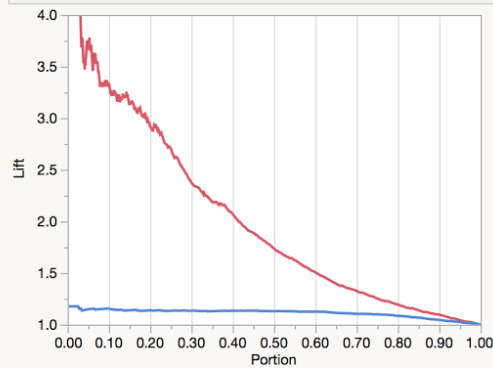
▼ Receiver Operating Characteristic on Validation Data



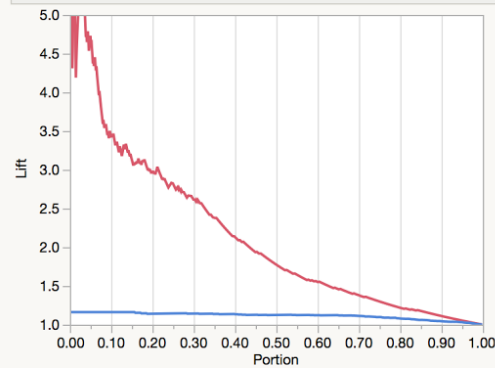
▼ Receiver Operating Characteristic on Test Data



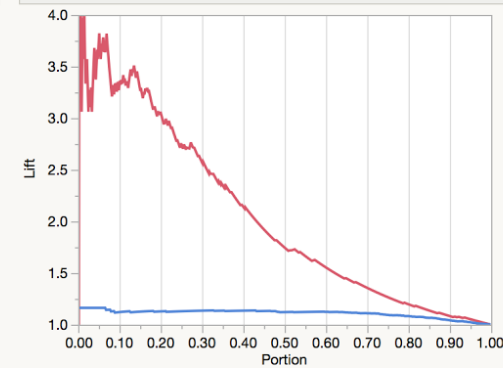
▼ Lift Curve



▼ Lift Curve on Validation Data

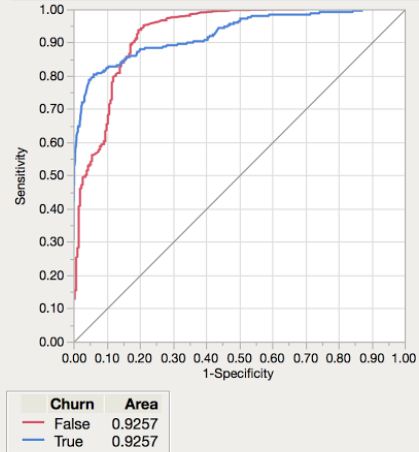


▼ Lift Curve on Test Data

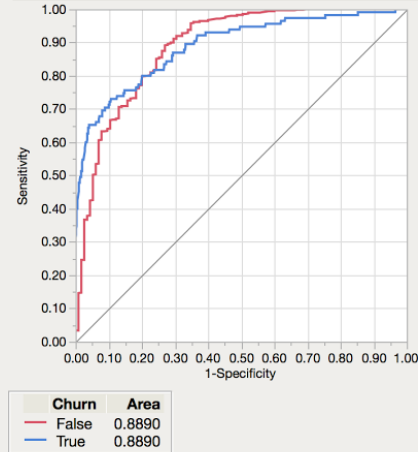


CHURN - Appendix

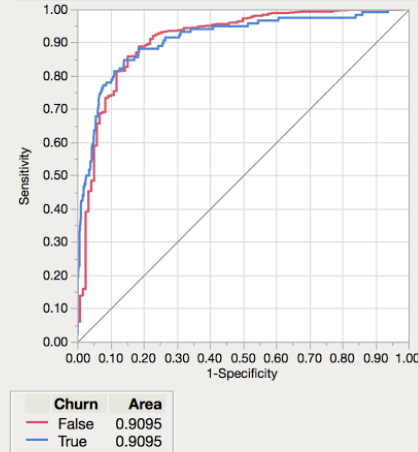
▼ Receiver Operating Characteristic



▼ Receiver Operating Characteristic on Validation Data



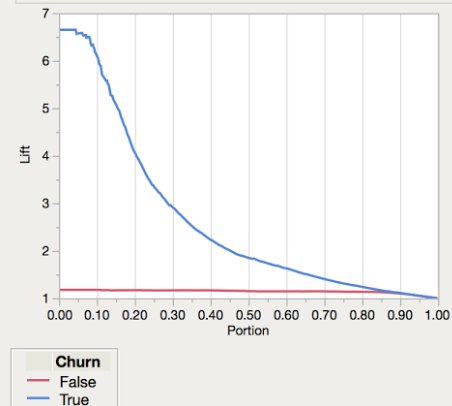
▼ Receiver Operating Characteristic on Test Data



Neural Network

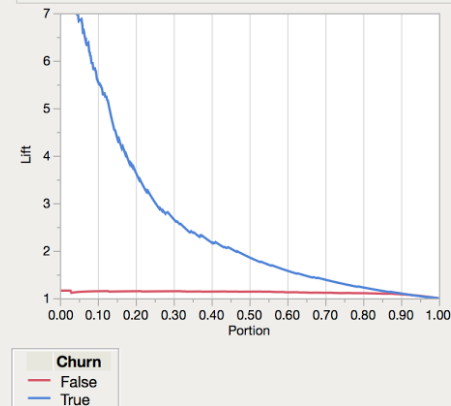
Training

▼ Lift Curve



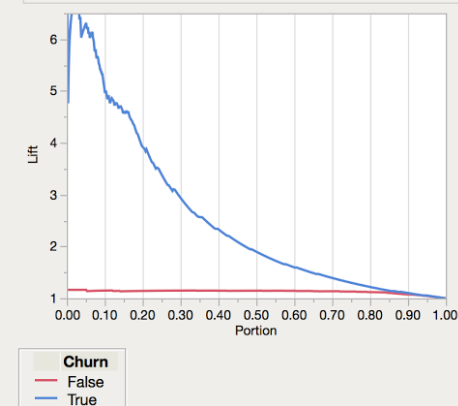
Validation

▼ Lift Curve on Validation Data



Test










▼ Lift Curve on Test Data



CHURN - Appendix

Model Comparison

▼ Measures of Fit for Churn

| Validation | Creator | .2 .4 .6 .8 | Entropy RSquare | Generalized RSquare | Mean -Log p | RMSE | Mean Abs Dev | Misclassification Rate | N |
|------------|----------------------|--|--------------------|------------------------|-------------|--------|-----------------|---------------------------|------|
| Training | Neural |  | 0.5062 | 0.6103 | 0.209 | 0.2377 | 0.1349 | 0.0661 | 1663 |
| Training | Partition |  | 0.4702 | 0.5750 | 0.2242 | 0.2427 | 0.1195 | 0.0710 | 1663 |
| Training | Fit Nominal Logistic |  | 0.1991 | 0.2717 | 0.339 | 0.3216 | 0.2058 | 0.1455 | 1663 |
| Validation | Neural |  | 0.4114 | 0.5102 | 0.2375 | 0.2574 | 0.1446 | 0.0823 | 826 |
| Validation | Partition |  | 0.4727 | 0.5727 | 0.2128 | 0.2364 | 0.1203 | 0.0678 | 826 |
| Validation | Fit Nominal Logistic |  | 0.2318 | 0.3081 | 0.31 | 0.3076 | 0.1962 | 0.1259 | 826 |
| Test | Neural |  | 0.4153 | 0.5145 | 0.2368 | 0.2608 | 0.1472 | 0.0925 | 843 |
| Test | Partition |  | 0.4913 | 0.5914 | 0.206 | 0.2319 | 0.1166 | 0.0712 | 843 |
| Test | Fit Nominal Logistic |  | 0.2033 | 0.2735 | 0.3226 | 0.3142 | 0.1980 | 0.1376 | 843 |