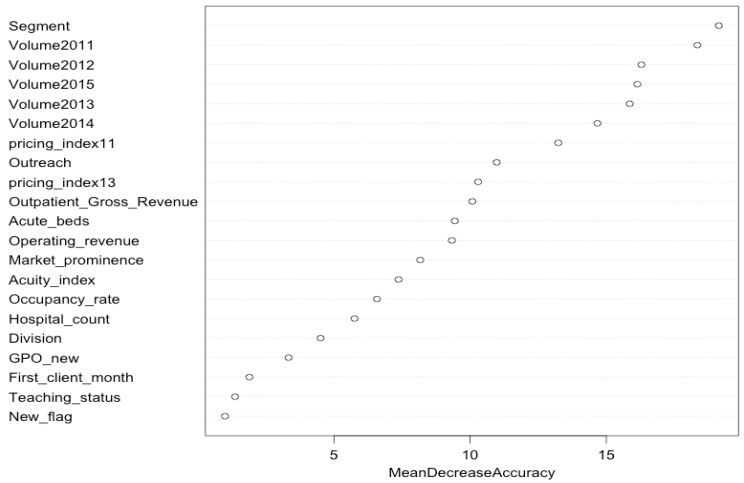
Lost-Contract Client Characteristics Analysis

Midwest Undergraduate Data Analytics Competition
Winona State University | www.MUDAC.org

Active

Lost Contract

Random Forest



Important Variables

- Volume 2011-2015
- Segment
- Pricing_index11 & 13
- Outpatient Gross Revenue
- Outreach
- Operating Revenue

Methods

Analytical Method

- Random Forest Model
  - ✓ Black-box algorithm
  - ✓ Importance plot
- Ridge Regression:
- Accuracy: 96.59%
- Other methods: logistic regression

Active Clients with Highest Probabilities to Drop

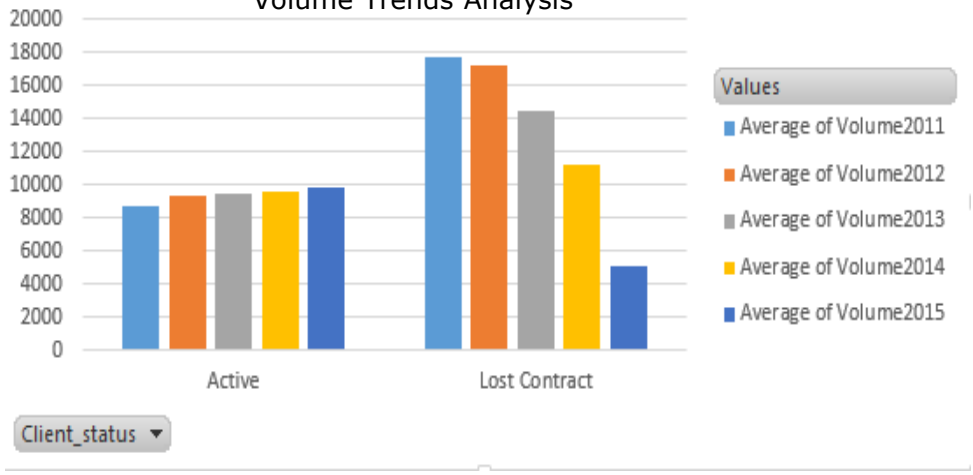
- Client ID:2488 3798 4742 1573 4811 1873 3995 4584 3022 681

Random Forest model plus ridge regression generates the best result.

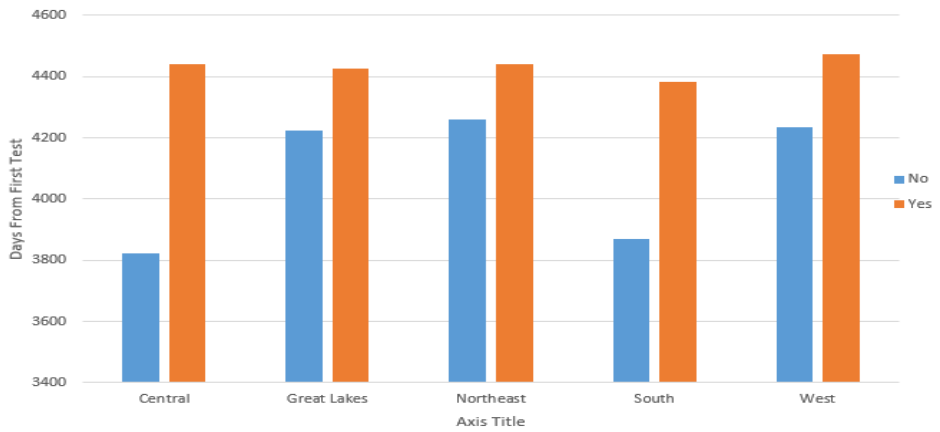
Lost-Contract Client Insights I

Midwest Undergraduate Data Analytics Competition
Winona State University | www.MUDAC.org

Volume Trends Analysis



Business Relation Analysis



Business Trends

Key Characteristics of Lost-contract Client	Explanation
➢ Volumes are constantly decreasing ➢ Old Clients in Central and South are more likely to be lost ➢ A lost-contract client is more likely can to be from Northeast (Next Slide)	Variable *First_Client_month is* modified as the number of days from now. The average values of this variable for Lost-contract clients are significant higher in Central and South. It indicates that the clients which have longer business relation with Mayo are more likely to quit.

Data indicates that the volume is an useful predictor when tracking relationship with clients. Further analysis about client retention in Central and South should be implemented to avoid further lose with old customers.

Lost-Contract Client Insights II

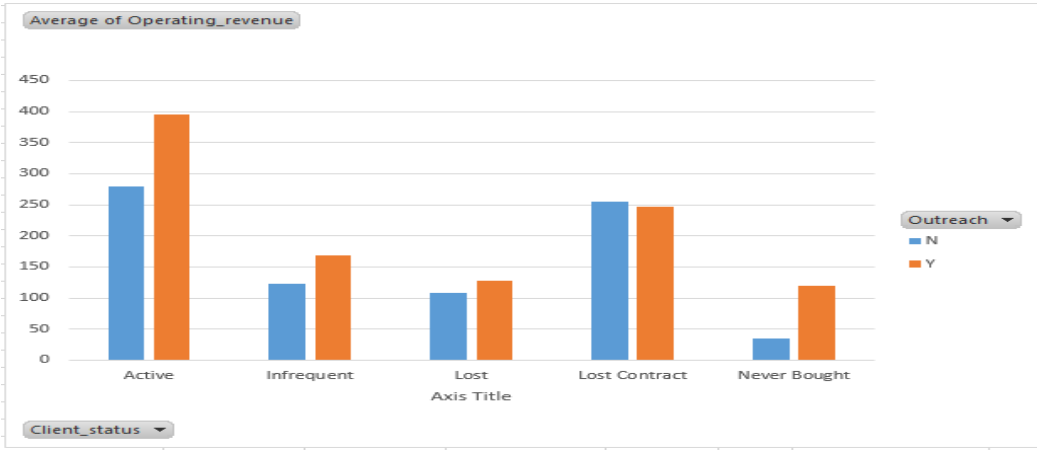
Lost-Contract Rate

Lost-Contract/Active Client rate of each division is calculated

8.33%

The rate is much higher in Northeast, compared with an average of 3.53% among other divisions

Outreach-Operating Income Analysis



Comparison

Operating Income

Average operation income for clients who do outreach is lower only in lost-contract clients group

Further analysis on clients in northeast should be conducted to understand the high quit rate. Industry standards about operating income could be set so that if a client who does outreach, its operating income could be compared with benchmarks to predict.

Active Clients Analysis

Identify significant attributes to movements in volume based on random forest

Significant drops	Set two new variables: yearly increase amount yearly increase rate	Define significant criterion: > 100 > 0.3	Extract top 5 important variables from random forest model using varImpPlot() for each year
Significant increases	Set two new variables: yearly decrease amount yearly decrease rate	Define significant criterion: < (-100) < (-0.3)	Extract top 5 important variables from random forest model using varImpPlot() for each year

Key important variable for analysis

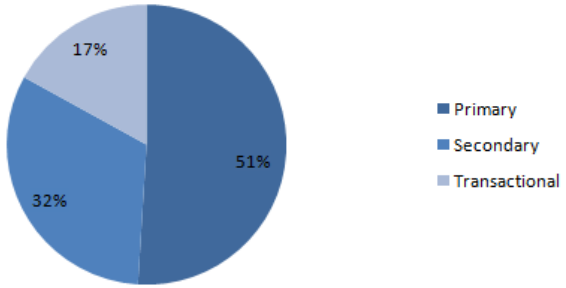
1	Pricing_index rate of change (2011 – 2013)
2	Segment
3	Acuity_index

Model indicates that among Active Clients, pricing_index change, segment and acuity_index are highly correlated to significant movements in volume

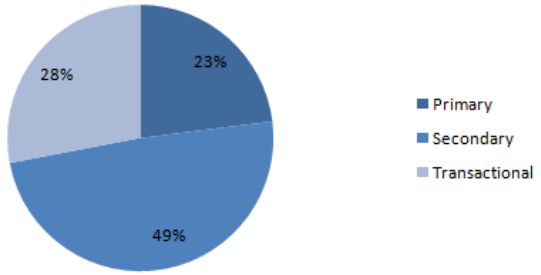
Active Clients Insights

Segments Distribution

Increase: Pie Chart of Segments

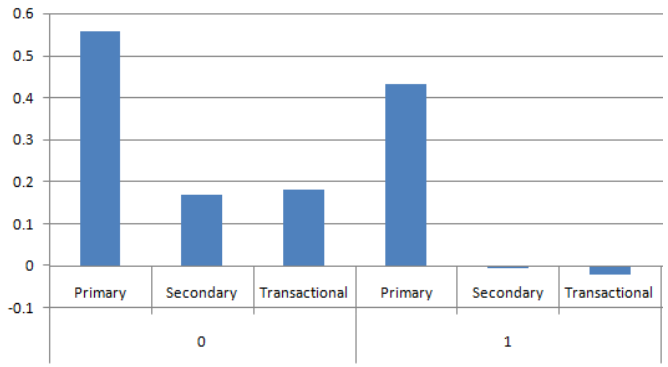


Decrease: Pie Chart of Segments

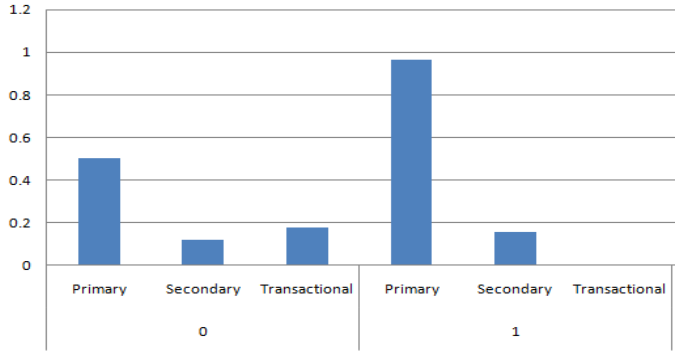


Pricing_Index Distribution

Increase: Average Pricing_Index Change Rate



Decrease: Average Pricing_Index Change Rate



Key Characteristics

- 51% significant increase active clients fall into Primary segment
- For Secondary and Transaction segment, decrease rate in pricing_index leads to significant increase

- 49% significant decrease active clients fall into secondary segment
- For Primary segment, increase rate in pricing_index leads to significant decrease

Primary clients are more likely to significantly increase volume; Secondary and Transaction clients with lower pricing_index are more likely to increase volume

Secondary clients are more likely to significantly decrease volume; Primary clients with higher pricing_index are more likely to decrease volume

Predicted Client Status Analysis

Midwest Undergraduate Data Analytics Competition
Winona State University | www.MUDAC.org

Classification

- Supervised learning in classified data
- Linear Discriminant Analysis
- K-Nearest Neighbor

Result

- Classification accuracy
  - KNN: 70%
  - LDA: 75%

Predict Revenue Potential

- Eliminate outliers
- Box-cox transformation: log
- Linear regression with stepwise and ridge

Result: R-Squared

- Primary:0.69
- Secondary:0.59
- Transactional: 0.58
- Fit all data: 0.61

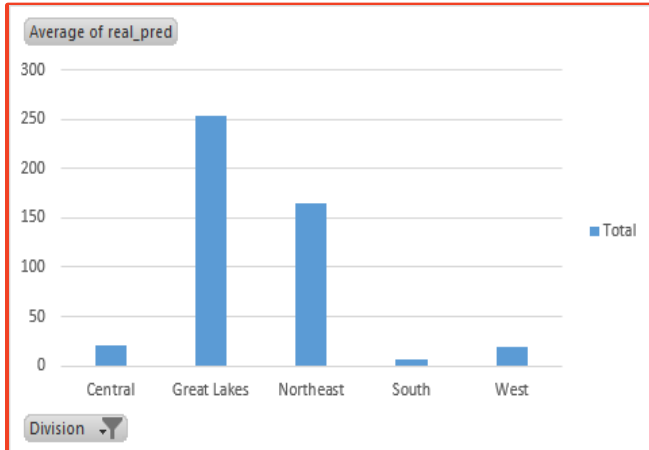
Classify data using K-Nearest Neighbor model and predict the revenue potential by modified linear regression.

Predicted Client Insight

Midwest Undergraduate Data Analytics Competition
Winona State University | www.MUDAC.org

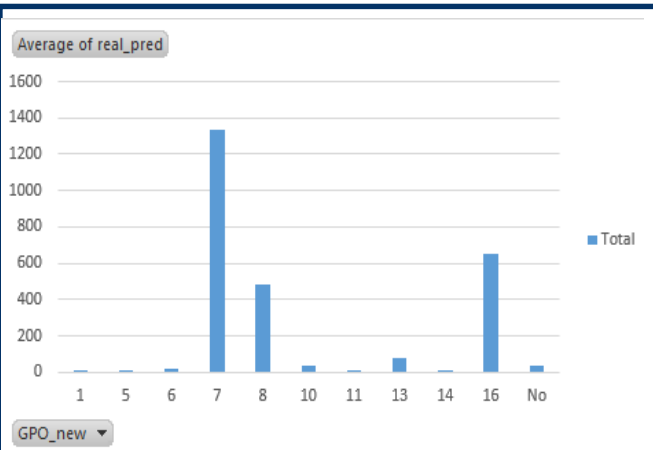
Analysis on Potential Clients' Division, Group Purchase Organization (GPO), and Client Status

Division



- Among predicted clients, Great Lakes and Northeast have the highest average volume
- Market strategies may be weighted along these two divisions

GPO



- Predicted clients in group 7,8,16 have high average volume
- Client relation may be researched among the three groups for potential business

Client Status

- Among the predicted clients, 66.28% are transactional, 33.42% are primary, and 0.2% are secondary.

- Although primary clients are not the largest group in potential clients, they have the largest total predicted volume.
- Mayo should focus more on clients if they are predicted as primary

Future client status and volume of never-bought clients are analyzed. Base on estimated average volume, Mayo should focus more on Great Lake and Northeastern division, GPO 7, 8 and 16 and clients with predicted status as primary.