PREDICTIVE ANALYTICS: EARLY WARNING DELINQUENCY SCORECARD

CASE STUDY: PREDICTIVE MODELING | **Telecommunications - Prepaid & Postpaid**

Objective

Identify the Risk behavior of customers who are likely to default within the next six month period of coming into Collections. Analyze using a segmented approach by forming sub populations within the customer universe.

Background and Challenges

One of the top telecom services provider of U.S. facing challenges in collection of payments with high default rates. Customers are defaulting leading to high delinquent payments that needs to be taken care by the service provider. Many even opt for disconnection while raking huge bills that will never be paid. Bills that intentionally never meant to be paid. The company decides to get control over the bankrupt customers and at the same time disrupting the losses accumulating month on month.

The company had a large portfolio with following portfolio:

- 16 million accounts,
- 450+ Attributes
- Bad Rate in population = 7.5%
- Consumer & Business Accounts

"Extremely pleased with the numbers. Thanks for the hard work."

- Director Telecommunications Consulting Company, USA

How to create Collection Models

- **Spend time on Data. Feel the customers.**
- **Innovate variables**
- **Choose variables ruthlessly**

Our Approach

Project Scope: Data Dictionary, Data Audit - Bivariate Analysis, Features Creation, Sub-Population Analysis, Feature Selection, Model Results and Validation, Final Implementation Reports. 50% of the data was used to build the model while remaining was preserved for model validation.

Method: Regression, Cramer's V.

Analyses Performed: Univariate, Bivariate, Binning.

Methodology: The project begins with building a clear understanding of the customer attributes. This is done

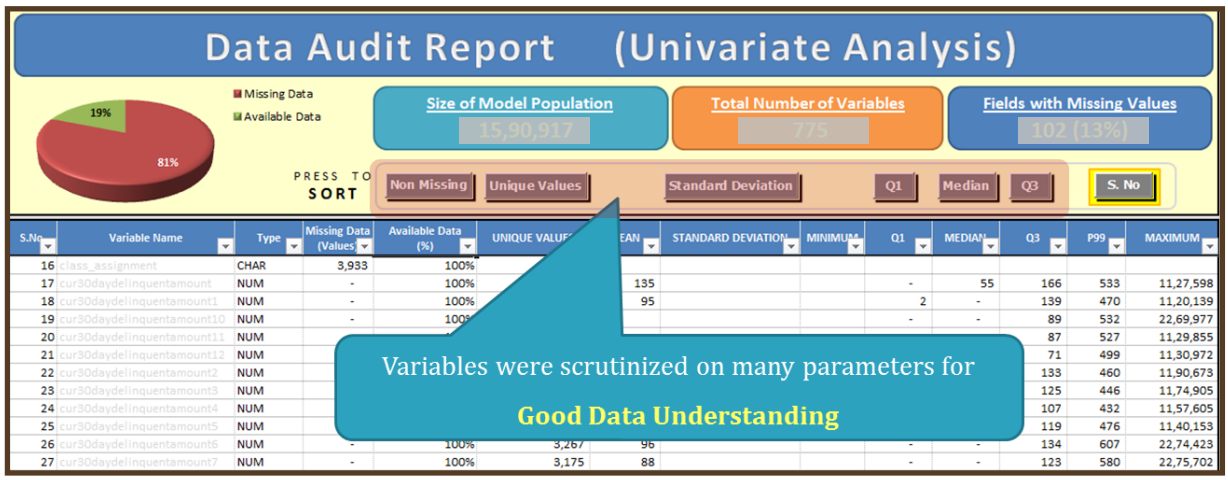


Illustration 1: A Dashboard to visually identify powerful characteristics of the variables.

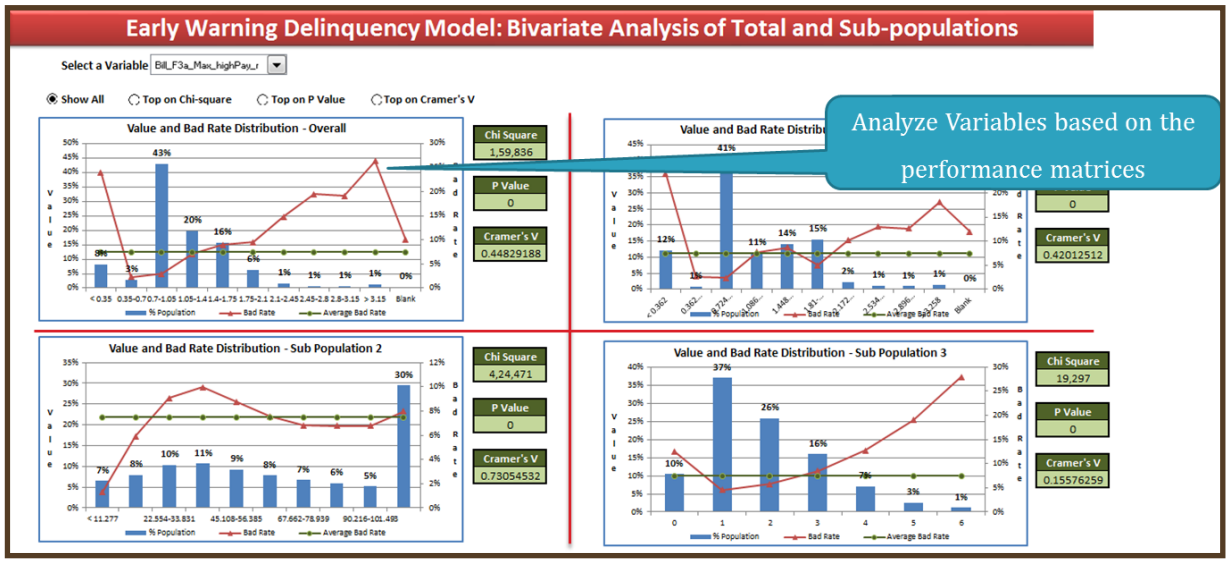


Illustration 2: A tool to compare attributes of customers with the target rate on a decile based layout.

Our Approach

Methodology (contd..): The variables were cleaned, transformed, and binned to derive the maximum value out of the logistic regression - which was the main technique to model bad rate or delinquency rate

- Features Creation
- Sub-Population Analysis
- Feature Selection
- Model Results and Validation
- Final Implementation Reports

New Variables created using business rules and Data Dispersion

Clusters considered with varying **Bad Rates** and **Population within the Segments**

High Performing Model Variables were selected after due diligence – Remove Colinearity

Model Results extensively studied and **Variables** Readjusted for Maximum **Performance**

Results Presented to Client and Model attributes integrated into **Strata**

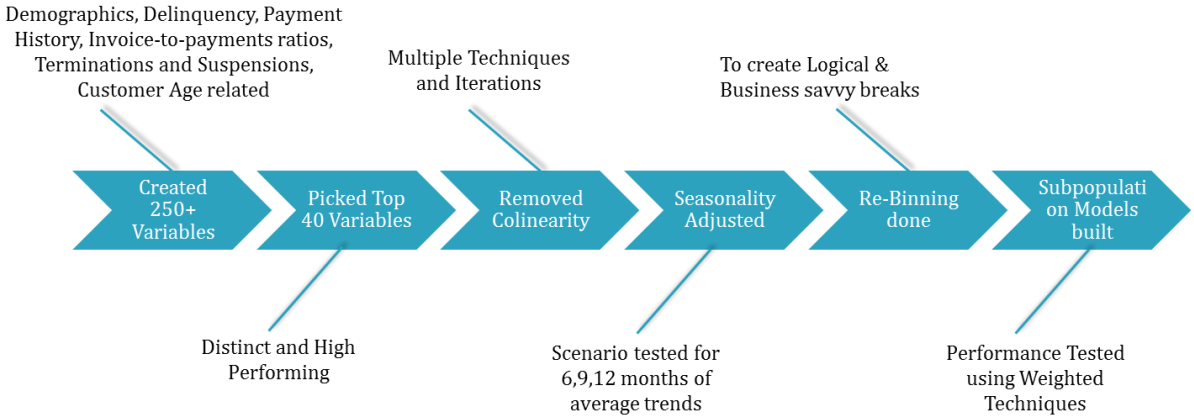


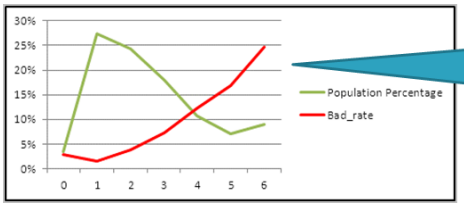
Illustration 3: Overview of our methodology followed for modeling Bad Rate.

Our Approach

Features Creation & Selection: Creating statistically significant variables, we are able to get meaningful variables that are also hugely correlating to the bad rate.

Variable Definition:

Number of times the current 30 day delinquent amount > twice the payment amount in any of the last 6 months



Defaulters have aggregated into smaller and focused groups – as **the Population % goes down, the bad rates go up**

Variable	Variable_ Category	Unique Values of Variable	Population	Population Percentage	No. of Bads	Bad_rate	Chi-Square	Cramer's V
delq_f1009_pay_grt2	0	7	27,272	3%	776	2.85%	55,700	0.2646
delq_f1009_pay_grt2	1	7	218,318	27%	3,545	1.62%	55,700	0.2646
delq_f1009_pay_grt2	2	7	193,487	24%	7,382	3.82%	55,700	0.2646
delq_f1009_pay_grt2	3	7	143,053	18%	10,549	7.37%	55,700	0.2646
delq_f1009_pay_grt2	4	7	84,911	11%	10,388	12.23%	55,700	0.2646
delq_f1009_pay_grt2	5	7	56,195	7%	9,450	16.82%	55,700	0.2646
delq_f1009_pay_grt2	6	7	72,106	9%	17,782	24.66%	55,700	0.2646

Results: Excellent results were obtained for all four sub-populations. Top Deciles were able to capture almost 80% of the bad customers and validation data nicely with the training data.

Cons wFiOS

Results of Individual Model - Subpop A

Decile	Total	Bad	Good	Bad Rate	Cumulative Capture Rate
1	4990	13	4977	0.26%	0.5%
2	4990	20	4970	0.40%	1.3%
3	4990	39	4951	0.78%	2.9%
4	4990	55	4935	1.10%	5.1%
5	4991	85	4906	1.70%	8.6%
6	4990	152	4838	3.05%	14.7%
7	4990	217	4773	4.35%	23.5%
8	4990	302	4688	6.05%	35.7%
9	4990	502	4488	10.06%	56.0%
10	4991	1089	3902	21.82%	100.0%
Total	49902	2474	47428	4.96%	100.00%

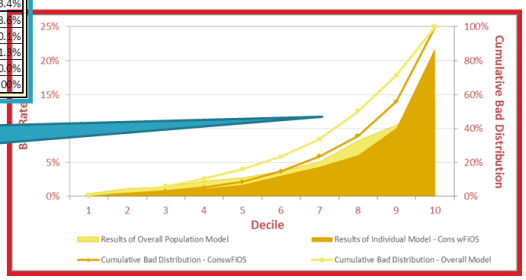
Results of Overall Population Model

Decile	Total	Bad	Good	Bad Rate	Cumulative Capture Rate
1	4990	19	4971	0.38%	0.8%
2	4990	58	4932	1.16%	3.1%
3	4990	69	4921	1.38%	5.9%
4	4990	112	4878	2.24%	10.4%
5	4991	136	4855	2.72%	15.9%
6	4990	185	4805	3.71%	23.4%
7	4990	253	4737	5.07%	33.[illegible]%
8	4990	408	4582	8.18%	50.[illegible]%
9	4990	525	4465	10.52%	71.[illegible]%
10	4991	709	4282	14.21%	100.[illegible]%
Total	49902	2474	47428	4.96%	100.0[illegible]%

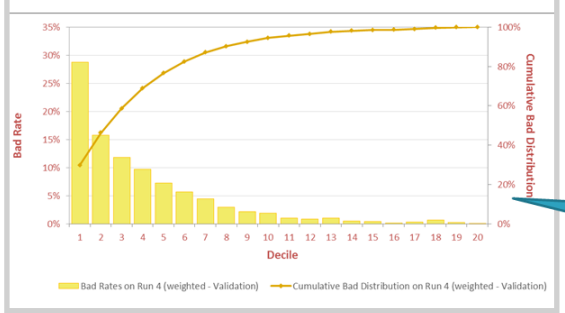
The Subpop A model performs better than the model built on overall population. **K-S lift of 49 vs. 39**

Top Decile Captures **45%** of Bad Customers

Top 3 Deciles Captures **78%** of Bad Customers



Validation Data Set KS = 55



The Model Validation conducted on 50% of the Data performs very well.