BI REPORTING CASE STUDY SUMMARY

Project Name		Brief Description	Key analyses
PROPENSITY TO PAY MODELING	**COLLECTION SERVICES COMPANY**	Partnered with the client to build a **"propensity to pay" model to help them identify high value accounts,** scientifically allocate agent resources to those accounts and enable the company to improve its Credit Collection Performance	• Poisson regression analysis

CASE STUDY: PREDICTIVE MODELING FOR A COLLECTION SERVICES COMPANY

ABOUT THE CLIENT

Propensity To Pay Modeling For A Collection Services Client

SITUATION

- Company was a **regional debt collection agency** that worked on third-party consumer portfolios
- The company **wanted to improve its credit collection performance** and **scientifically allocate its agent resources** to the high value account portfolios and customers
- Merilytics partnered with the company **to build a "propensity to pay" model** that would rank the portfolios and customers within the portfolios, on a monthly basis
- This propensity-to-pay ranking model would **enable the company to prioritize its resources** and work those accounts with higher intensity

VALUE ADDITION

- **Shortlisted accounts from the portfolio** for propensity to pay modeling, after factoring for data noise and shorter tenure
- **Defined** a set of **independent variables** across account characteristics, payment data and 'treatment' data and **dependent variable** as 'number of payments'
- **Ran a "matching" algorithm** on historically treated (contact made by any person or machine) and non-treated account sets, to remove treatment bias
- **Conducted a zero-inflated Poisson regression** analysis on the matched data set, and ranked accounts in descending order of payment likelihood

IMPACT

- Critical **variables in determining the payment likelihood** were identified
- Enabled the company to achieve **a ~10% lift in payments**, and by treating fewer accounts

DEFINED A SET OF INDEPENDENT VARIABLES ACROSS ACCOUNT CHARACTERISTICS, PAYMENTS AND 'TREATMENT' DATA AND DEPENDENT VARIABLE AS '# OF PAYMENTS

merilytics AN ACCORDION COMPANY

Snapshot of information captured for an account with tenure 'N'[1]

Account characteristics (static variables)

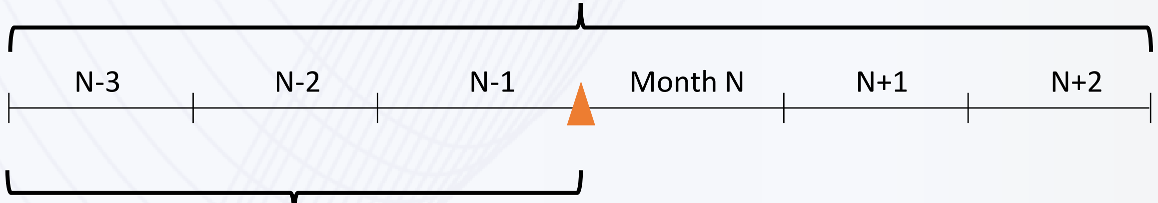
Account level attributes
- Date of Record
- Age
- Income Score
- Other Score
- # of Phones
- Placed Balance
- Charge Off Date
- Charge Off Reason
- Contract Date
- Type of Account/Card
- 'Settlement in full" related details

Account level metrics
- Tenure – time since Date of Record
- Age
- Income Score
- # of Phones
- Placed Balance
- Time since charge off
- Time since contract date as of tenure period

Payment activity attributes
- Mode of Payment
- Payment Amount
- Time of Payment
- Previous Balance
- Current Balance

Payment activity derived metrics
- Time since last payment
- # of payments made in 6 months (N-3 to N+2)
- % amount paid in 6 months - as a % of placed balance

Payment data: for the account during 6 months (attributable to the answered calls made in last 3 mo.)



Call data (treatment) – total answered calls in past 3 mo.

Call activity attributes
- Mode of contact of calls
- Duration of calls
- Outcome type of calls
- Time of calls

Call activity metrics
- # of calls made in past 3 months (Landline, mobile, other)
- # of calls made in past 3 months - by outcome category

1 Tenure N represents 'N^{th}' month of the account since the Date of Record

'MATCHING' OF ACCOUNTS WAS DONE TO REMOVE EFFECT OF 'TREATMENT' BIAS

merilytics
AN ACCORDION COMPANY

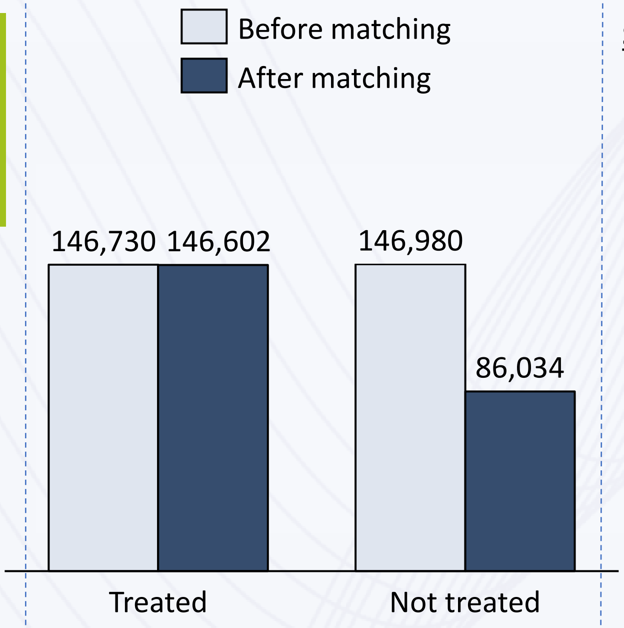
Shortlisted attributes that could have impacted treatment are used for matching	Matching algorithm is applied on ~293K accounts, resulting in a matched sample of ~233k accounts	…. to generate set of accounts with similar characteristics, as evident by the reduced Standardized Difference score[1]

Treatment variables:
- 0 (no answered call in past 3 months) or
- 1 (at least 1 answered call in past 3 months)

Co-variates selected:
- Tenure of account
- Age of customer
- Time since charge off
- Income Score
- Time since last payment

Matching algorithm:

Method: Logit regression
Distance: Nearest neighbor method
Significance level[2]: 0.1



Standardized Difference score[2]

1 Standardized Difference signifies the extent of lack of overlap in distribution curves of the metric, between treated and non-treated sets of accounts
2 Significance level indicates the extent of robustness of the predictive model for propensity to treat. E.g., significance level of 0.10 indicates there that there is less than 10% probability that the results from the regression are not representative of the true relationship between covariates and dependent variable

ZERO INFLATED POISSON REGRESSION' WAS RUN ON THE MATCHED ACCOUNTS TO DEVELOP A PROPENSITY MODEL THAT WAS USED TO RANK ALL ACCOUNTS

merilytics AN ACCORDION COMPANY

Selected attributes that affect the propensity to pay, defined the dependent variable…	…ran 'Zero-inflated Poisson regression model' on the matched sample of accounts - 233K	…generated the ranking of accounts by applying the model on 407K accounts
Dependent Variable • # of payments made in 6 months (N-3 to N+2 months) **Independent Co-variates:** • **Account Level Attributes** — Tenure — Age of customer — Income Score — Number of phone numbers — Time since charge-off • **Call Activity Attributes:** # of successful calls made in past 3 months • **Payment Activity Attributes**: Time since last payment • **Matching Attribute**: Derived P-score[1] from matching step	**Regression on matched sample of data** • Matched sample of accounts – 232,636 • **% of accounts that made zero payments in 6 month period – 95.58%** • **Regression technique: Two stage – zero inflated model** o Binomial with logit link – Predicts whether a person is likely to pay or not o Poisson with log link – Predicts the number of payments that a person is likely to make • **Significance level cut-off[2]:** 0.05	• Applied regression model on 407K accounts to predict number of payments • Accounts are ranked based on the estimated likelihood of number of payments over a 6 month period *Rank 1* *Rank 2* *Rank 3* . . . . . . *Rank 407,335*

1 P-score is the outcome variable in the matching step (Step 1) and signifies the propensity to treat (i.e., being contacted and answered by a person or machine)
2 Significance level indicates the extent of robustness of the predictive model. E.g., significance level of 0.05 indicates there that there is less than 5% probability that the results from the predictive model are not representative of the true relationship between covariates and dependent variable

ACCOUNTS FROM THE PORTFOLIO WERE SHORTLISTED ACCOUNTING FOR DATA NOISE AND SHORTER TENURE

Account funnel for the portfolio:

427K accounts: Active at some point in history (ensured that only one historic snapshot of each account is captured)

Excluded accounts with:
- **Placed Balance <= 0**
- **Current Balance <=0**
- **Account life <=30 days**
- **Age = (>90, <18)**

407K accounts: Accounts used for 'Final ranking'

Excluded accounts with
- **Tenure of 1 month as of the selected month**
- **Age = 'n/a'**
- **Income score = 'n/a'**

293K accounts: Accounts used for 'Matching' to remove treatment bias and for Regression equation

Reasoning

- **Negative/zero placed or current balance indicates data noise**
- **If total account life is <=30 days, there is not enough treatment history**

- **For each snapshot, if current tenure is too low, we would not have sufficient history to run the matching algorithm, for that snapshot, for that account**
- **Missing age and income attributes could distort the matching model**