



**POWERLYTICS**

POWERFUL DATA, SMARTER DECISIONS

## *Consumer Default Modeling Case Study*

### **Challenge**

Improve the accuracy of consumer credit risk evaluation at the individual or portfolio level by more successfully identifying the probability that a consumer will default.

### **Solution**

Leveraging their proprietary consumer database that provides a comprehensive and anonymized financial view of over 150 million households (200 million adults) in the U.S. and a loan-level customer data file of funded loans and their performance, Powerlytics developed a consumer default score.

First, the customer loan data (to include default indicator) was merged with the Powerlytics data by the borrower's geographic location (zip+4, zip code, etc.), year of loan origination and other relevant characteristics to create a dataset to be used for probability of default analysis. Next, a set of Powerlytics variables that best predict default in the data was determined using various model selection techniques that include stepwise models (forward selection and backward elimination) and lasso regressions.

Variables relating to an individual's total federal tax due amount as a percentage of household income, adjusted gross income and disposable income were among the set of Powerlytics variables that were most predictive of default. Also, variables that relate to long-term savings in retirement accounts and pensions and annuities were significant predictors. Social security benefits amount (taxable) as a percentage of disposable income and adjusted gross income also contributed towards predicting default. Interestingly, various income metrics (e.g. business income) of a zip+4 in which a borrower resides relative to the broader geography (zip code, state, etc.) were among significant predictors as well.

After arriving at the selected subset of Powerlytics variables, binomial logit regressions are run with the binary dependent default variable and different combinations of client variables and chosen set of Powerlytics variables as independent variables to determine the model that best predicts default.

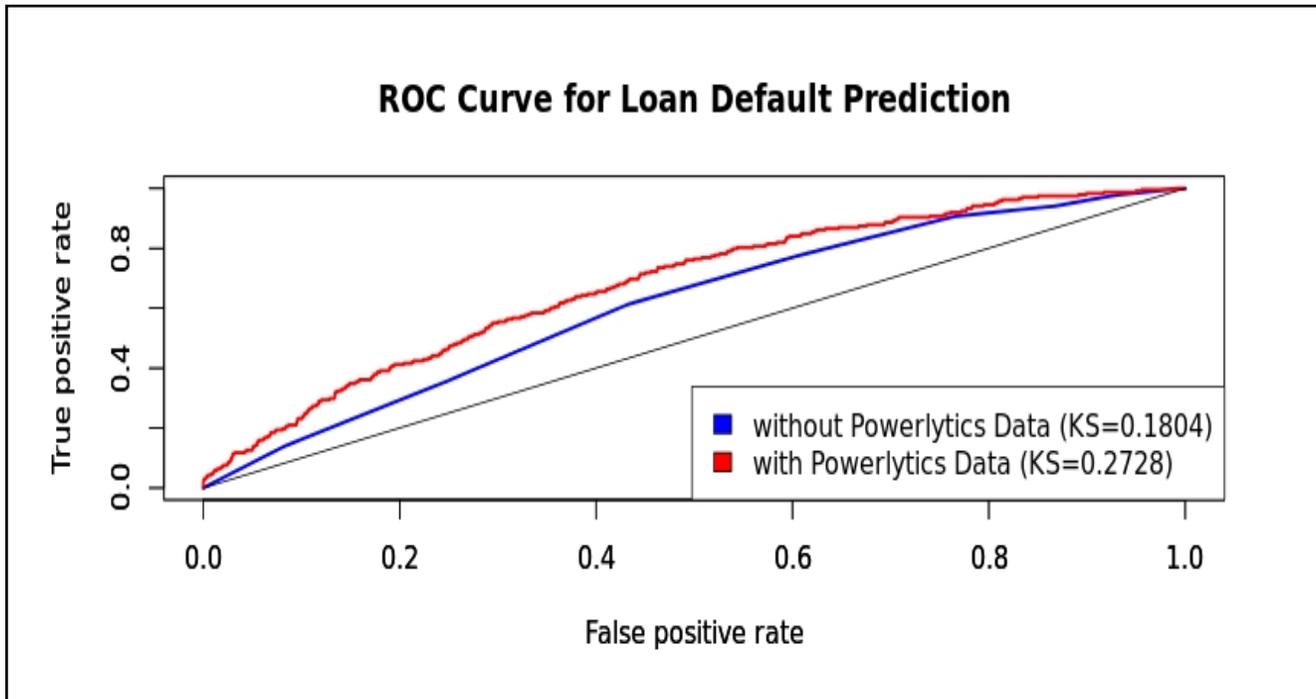
### **Results**

KS scores (quality of the results) from different specifications are compared to analyze the lift that Powerlytics data is able to generate when combined with loan-level data. In these client samples, combining Powerlytics data with loan term data, such as the customer's credit score (FICO®), APR, loan amount, loan term, etc., improved KS score from ~0.18 to ~0.27, an increase of ~50% (9 percentage points). Please refer to the ROC curve diagram on the next page for a depiction of the KS improvement associated with combining loan-level data with Powerlytics data.



**POWERLYTICS**

POWERFUL DATA, SMARTER DECISIONS



Powerlytics Default Models help in accurately predicting credit risk of borrowers applying for loans by combining applicants' demographic & financial information (self-reported/from credit bureaus) with Powerlytics' comprehensive database covering a wide range of income and tax fields. In addition, Powerlytics Default Models can help in accurately measuring changes in risk profile of loan portfolio overtime. The time-series of income and tax related variables in Powerlytics Consumer data can detect and signal a change in credit risk associated with a client's portfolio in a timely fashion, thereby enabling improved investment decision-making. The potential savings from anticipating probable defaults and correcting them can be quite large.