credit_report

Creation Date: Friday, June 2, 2023 10:57:55 AM

Author: Chris Parrish

13 Pages

Contents

credit_report

Contents

Contents

This analytics project creates visualizations and models derived from synthetically generated cross-sectional, credit report-type information using the many automated featues of SAS Viya's business intelligence application, Visual Analytics.

The synthetically generated data were saved as a .csv file and imported under the 'Manage Data' tab in SAS Viya into an in-memory table in SAS Cloud Analytic Services called 'credit_report'.

The 'credit_report' table can also be viewed in the 'Discover Information Assets' app within SAS Viya, which provides automated data quality and column analysis as well as senstive data detection (e.g., PII).

The **'credit_report' table** is the primary data source for **this 'credit_report' visualization report**. Each of the visualizations in this report can be saved as a template to use in other reports and for other users.

Additionally, all of these visualizations can be customized and enhanced to support the requirements of end-users.

The 'credit_report' analysis starts with single column (univariate) evaluation where both the internval (**'univariate_analysis_int'**) and categorical (**'univariate_analysis_cat'**) columns in the table are charted to visualize scale and distribution.

Next, the 'campaign_response' and 'campaign_performance' tabs evaluate and chart statistics associated with peer group campaigns.

The **'campaign_response'** visualization is a heat map of responses by campaign product type and peer id (generic bank1, bank2, etc.).

The **'campaign_performance'** visualization shows how accounts that responded (or not) to a targeted campaign by a peer group performed (defined generally as 'bad_credit' and 'not_bad_credit').

The next set of tabs delve deeper into the analytics associated with this data table with the eventual goal of creating a scorecard.

In the **'correlation_matrix'** tab, pairwise correlations are automatically generated and more highly correlated features are shown in a darker hue. The column 'customer_event', which is a bad credit/good credit indicator, is shown in the first position on the x-axis so that users can quickly understand the relative correlations between this variable and others within the table.

The **'selected_cross_tab'** visualization shows some cross tabulation metrics associated with customer demographics, including the percentage of bad/good credits by employment status, martial status, and homeownership.

SAS Viya includes other automated analysis that can help shape the direction of scorecard development and weight-of-evidence investigation.

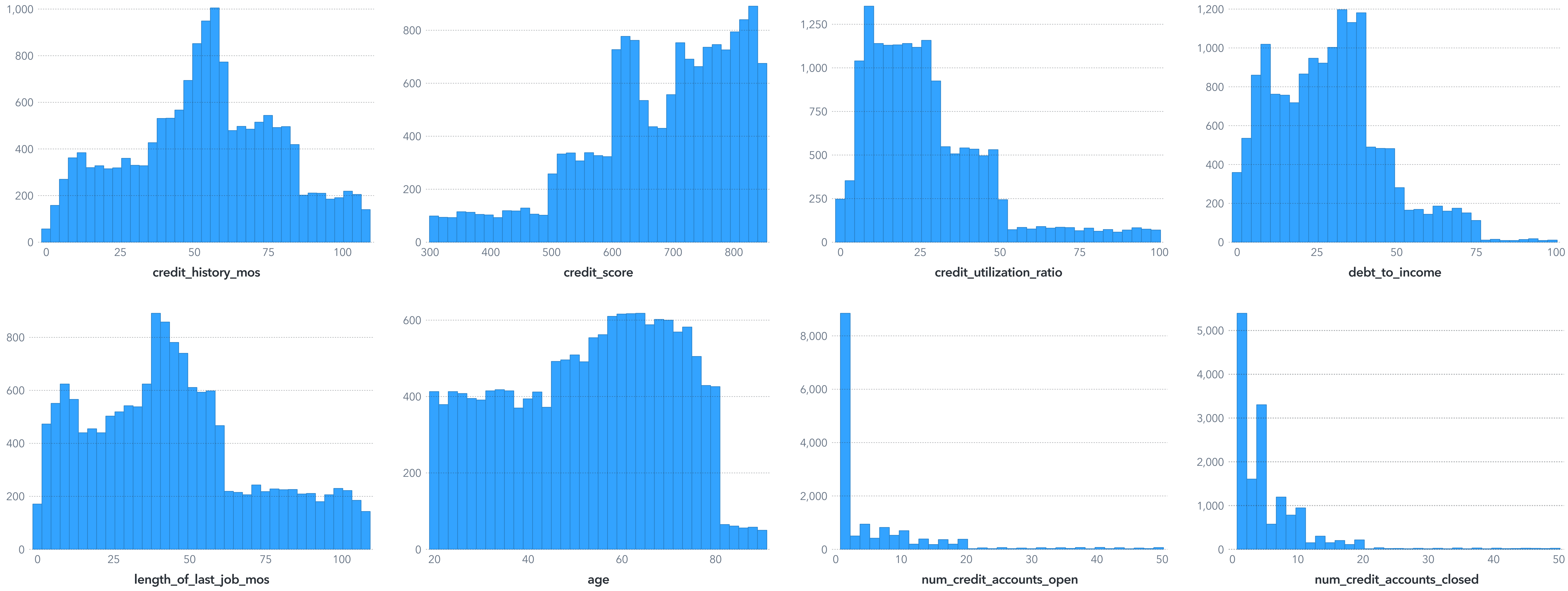
In the **'automated_explanation'** tab, an algorithm is run that seeks to identify the potential important variables associated with an outcome, which in this use case is whether the account is a "good" credit or "bad" credit (customer_event indicator).

The **'automated_prediction'** tab takes the analysis a step further by evaluating a set of machine learning algorithms that seek to optimize model accuracy and selects the best model. Interactive features allow the end-user to score new accounts or perform what-if analysis in a point-and-click format.

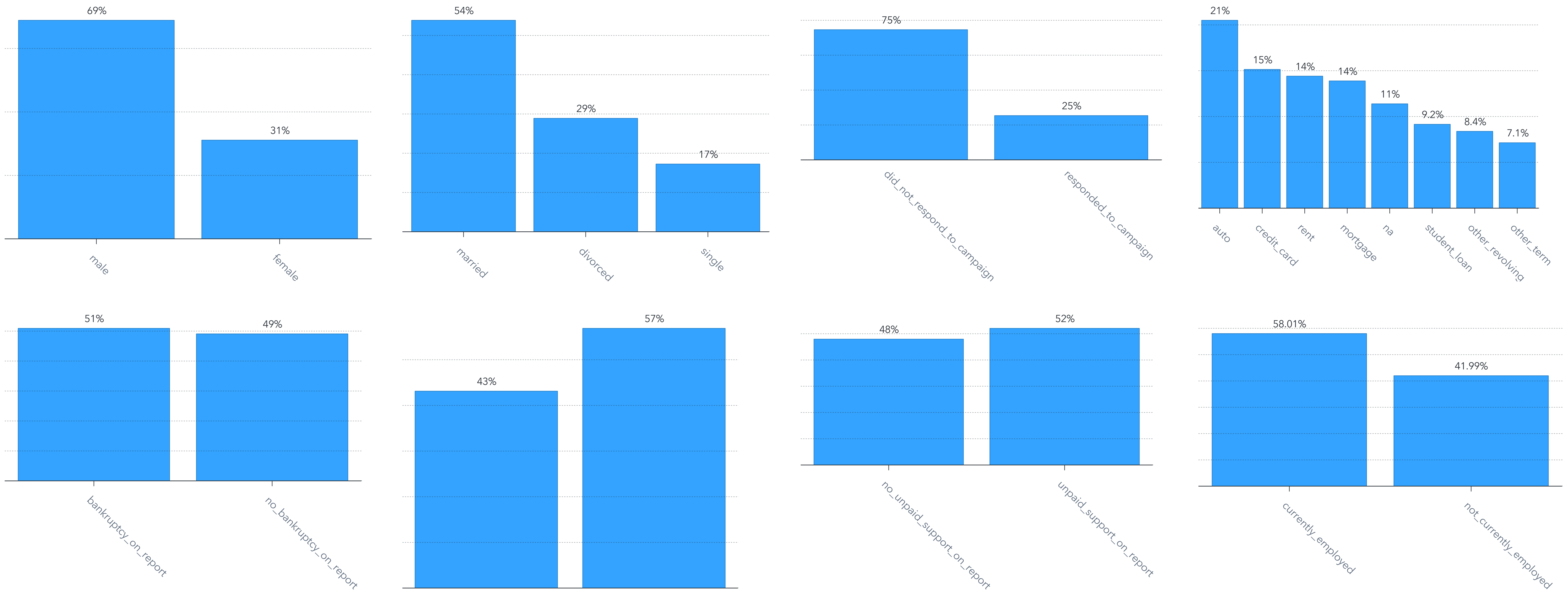
As shown in the **'decision_tree'** tab, interactive decision trees can also be developed, which allows end users to introduce manual changes to the decision tree based on domain knowledge or other business rules.

The last two tabs include functionality to create a custom score card in SAS Viya. In this use case, weight-of-evidence variable transformations were completed in the 'Prepare Data' application in SAS Viya using a custom code transformation. Once the data were transformed, the table with the new features were added to the data supporting this report.

univariate_analysis_int

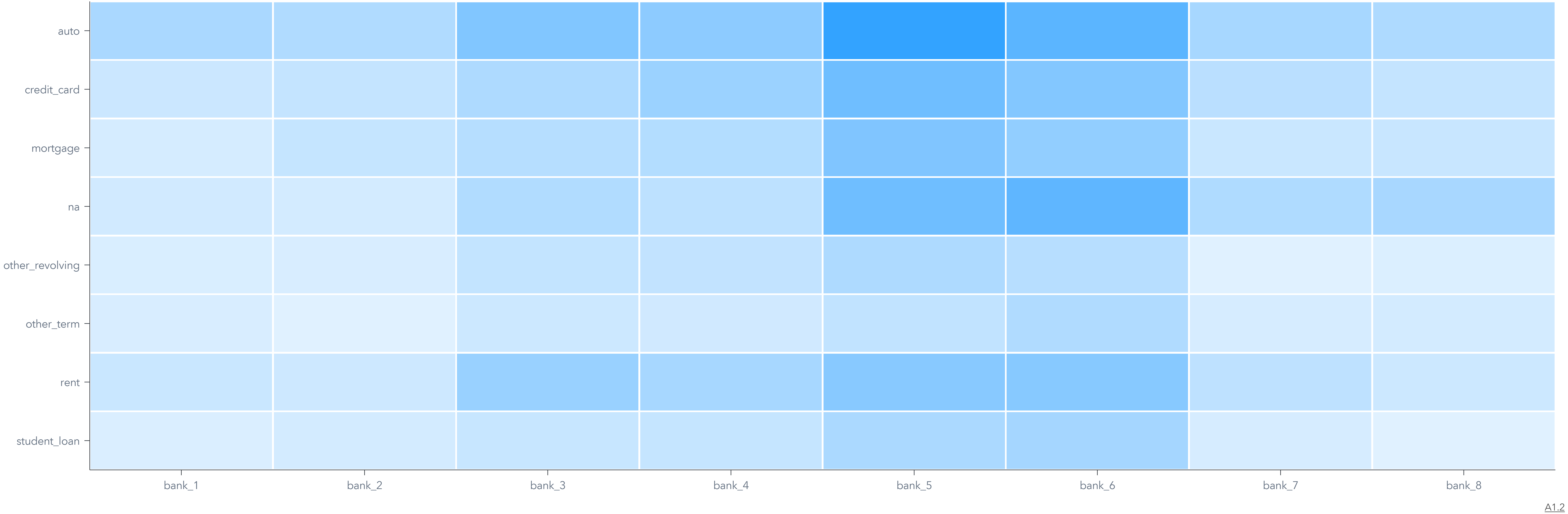


univariate_analysis_cat



campaign_response

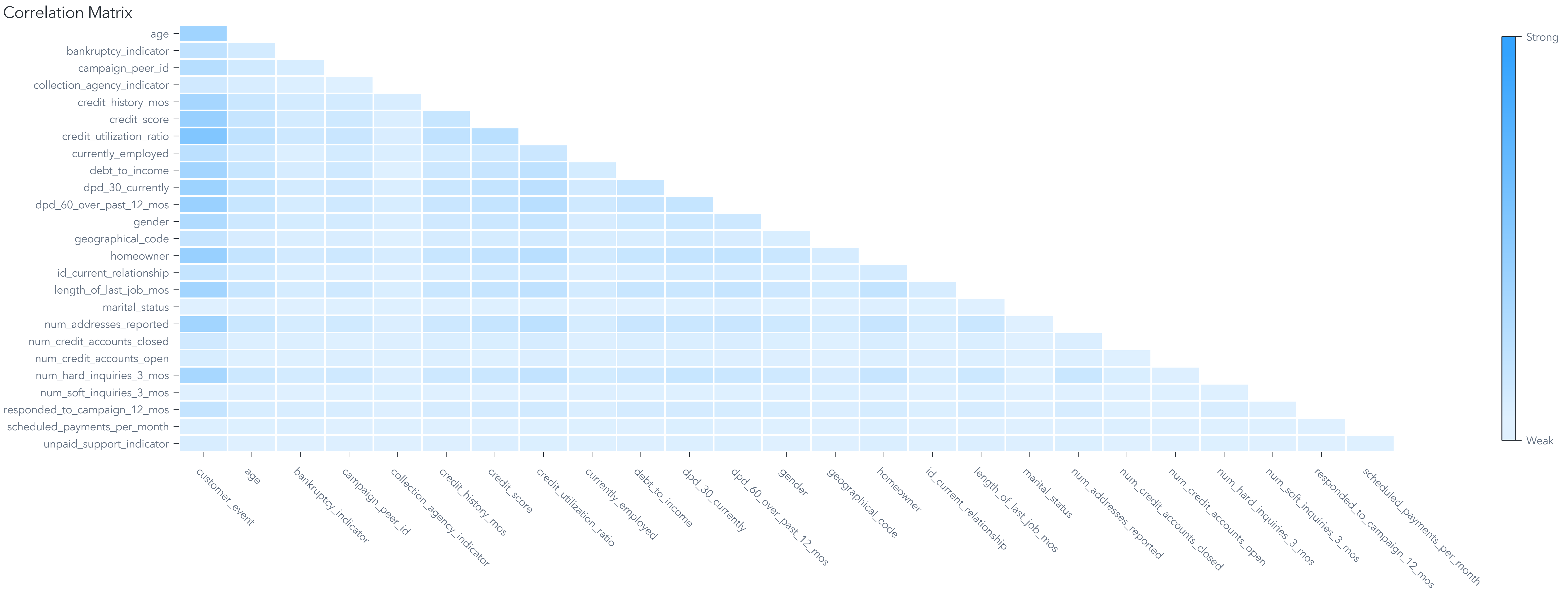
Heat Map of Response to Campaigns by Peer ID and Account Type



Darker Blocks Indicate Customers Responded in Higher Numbers to Targeted Campaigns by Peer (Bank) Group campaign_performance

	customer_event_char	bad_credit		not_bad_credit	
responded_to_campaign_12_mos_char	**campaign_peer_id_char**	**Frequency**	**Frequency Percent**	**Frequency**	**Frequency Percent**
did_not_respond_to_campaign	**bank_1**	163	1.06%	857	5.58%
	bank_2	144	0.94%	868	5.65%
	bank_3	292	1.90%	1,297	8.45%
	bank_4	281	1.83%	1,250	8.14%
	bank_5	605	3.94%	1,754	11.43%
	bank_6	566	3.69%	1,674	10.90%
	bank_7	426	2.78%	400	2.61%
	bank_8	473	3.08%	404	2.63%
responded_to_campaign	**bank_1**	70	0.46%	216	1.41%
	bank_2	78	0.51%	219	1.43%
	bank_3	164	1.07%	345	2.25%
	bank_4	162	1.06%	331	2.16%
	bank_5	372	2.42%	459	2.99%
	bank_6	344	2.24%	443	2.89%
	bank_7	243	1.58%	116	0.76%
	bank_8	223	1.45%	112	0.73%

correlation _matrix



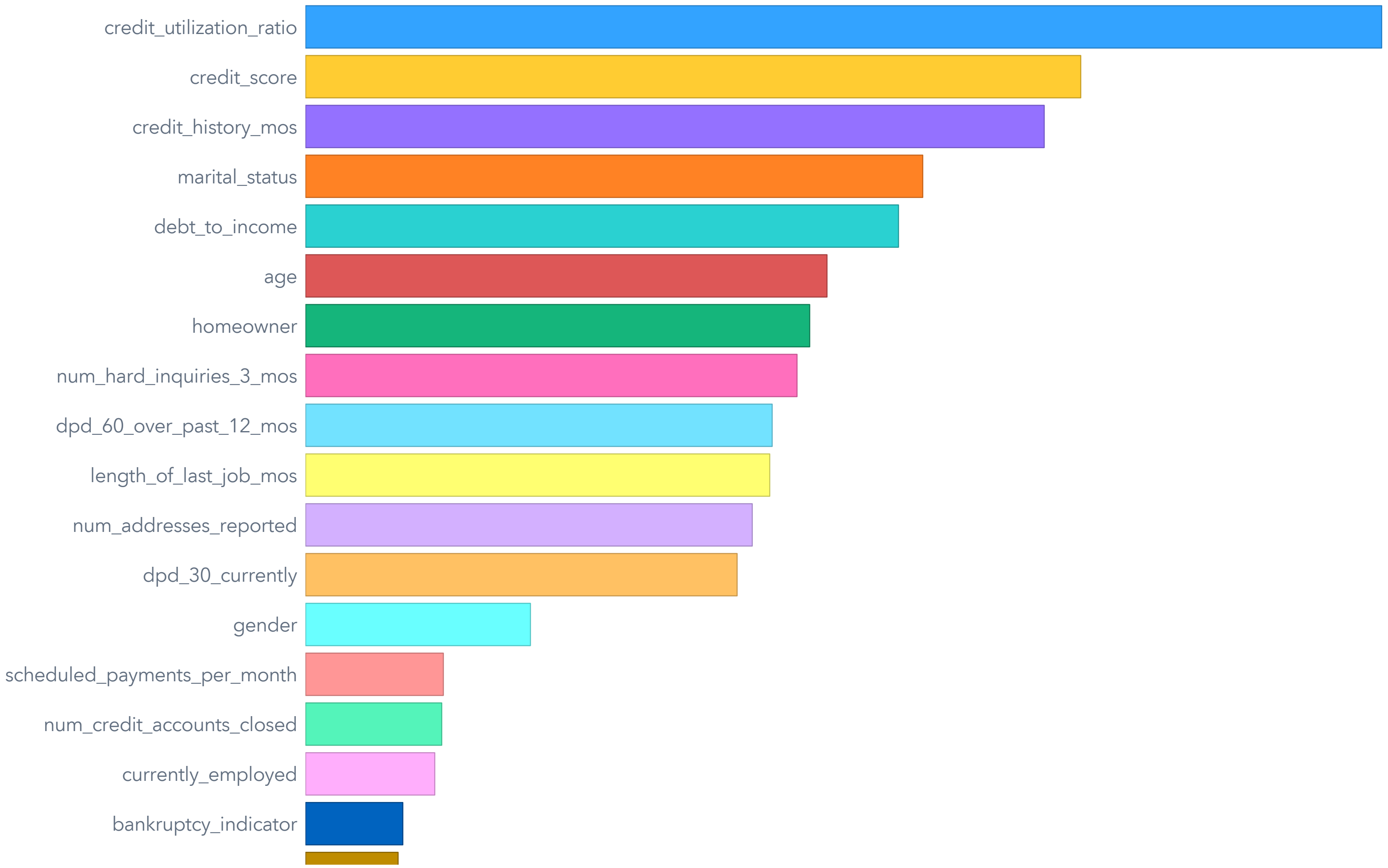
selected_cross_tab

		customer_event_char	bad_credit	not_bad_credit
currently_employed_char	marital_status_char	homeowner_char	Frequency Percent	Frequency Percent
currently_employed	divorced	homeowner	0.53%	5.27%
		not_homeowner	5.26%	4.49%
	married	homeowner	0.24%	16.73%
		not_homeowner	2.27%	14.42%
	single	homeowner	0.50%	2.47%
		not_homeowner	3.87%	1.97%
not_currently_employed	divorced	homeowner	0.74%	2.79%
		not_homeowner	7.35%	2.46%
	married	homeowner	0.31%	9.22%
		not_homeowner	3.03%	7.64%
	single	homeowner	0.61%	1.47%
		not_homeowner	5.31%	1.06%

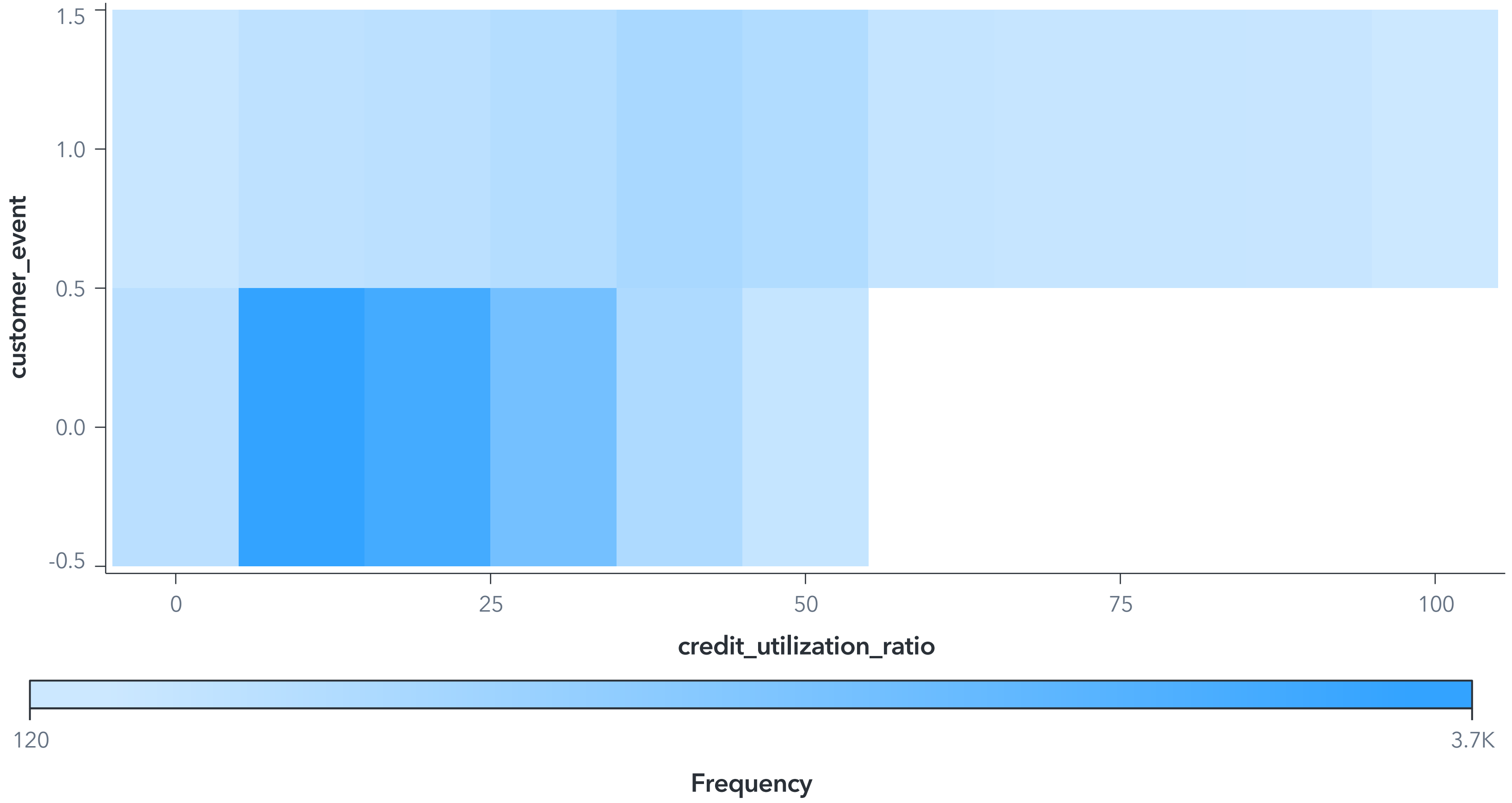
What are the characteristics of Response?

customer_event ranges from 0 to 1. Average customer_event is 0.3. Most cases (12K of 15K) have a customer_event between 0 and 1. credit_utilization_ratio best differentiates the highest (top 10%) and the lowest (bottom 10%) customer_event cases. The three most related factors are credit_utilization_ratio, credit_score, and credit_history_mos.

What factors are most related to customer_event?



What is the relationship between customer_event and credit_utilization_ratio?



customer_event may have a strong positive relationship with credit_utilization_ratio. It appears to be a linear relationship. For every 1 increase in credit_utilization_ratio, customer_event increases by 0.013. Average credit_utilization_ratio is 27, and it ranges from 0 to 100.

Prediction of "Bad Credit" Based on the Most Important Factors

What values for the most important factors should be used to predict?

credit_utilization_ratio

23

credit_score

697

credit_history_mos

54

marital_status_char

married

homeowner_char

homeowner	not_homeowner

age

54

dpd_60_over_past_12_mos_char

has_been_60_dpd_last_12_mos	has_not_been_60_dpd_last_12_mos

dpd_30_currently_char

currently_30_dpd	not_currently_30_dpd

gender_char

male

currently_employed_char

What is the prediction for customer_event?

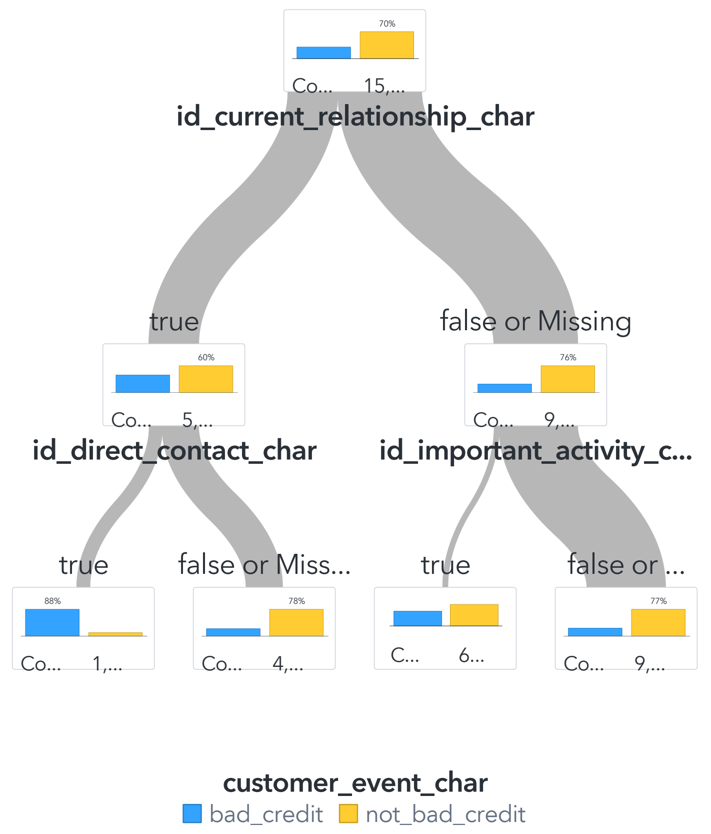
0.0382468553

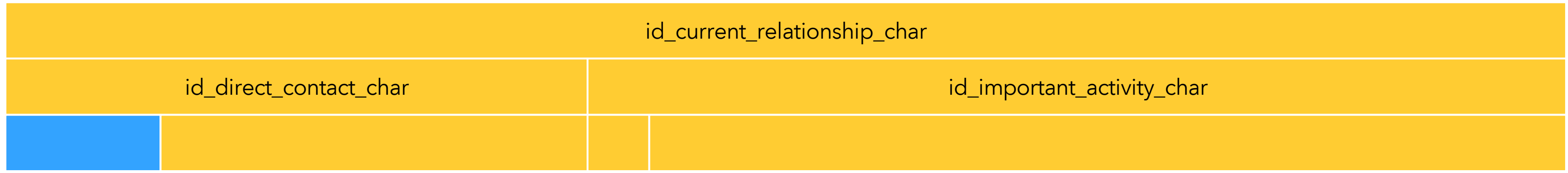
The predicted customer_event for this case is 87.25% lower than the observed average customer_event of 0.3. Most observations (70%) have a lower customer_event than this predicted case. The prediction is based on an automatically selected Gradient Boosting model.

decision_tree

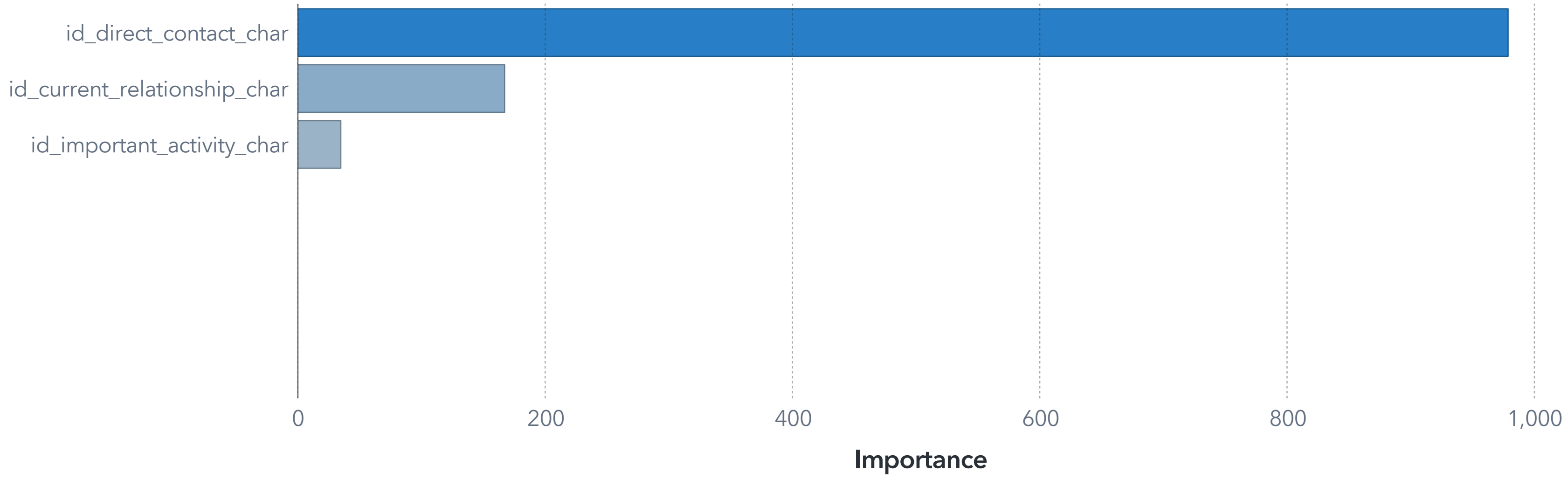
Decision Tree **customer_event_char** Event: **not_bad_credit** Fit: **KS (Youden) 0.2950** Observations: **15,351 of 15,351**

Tree

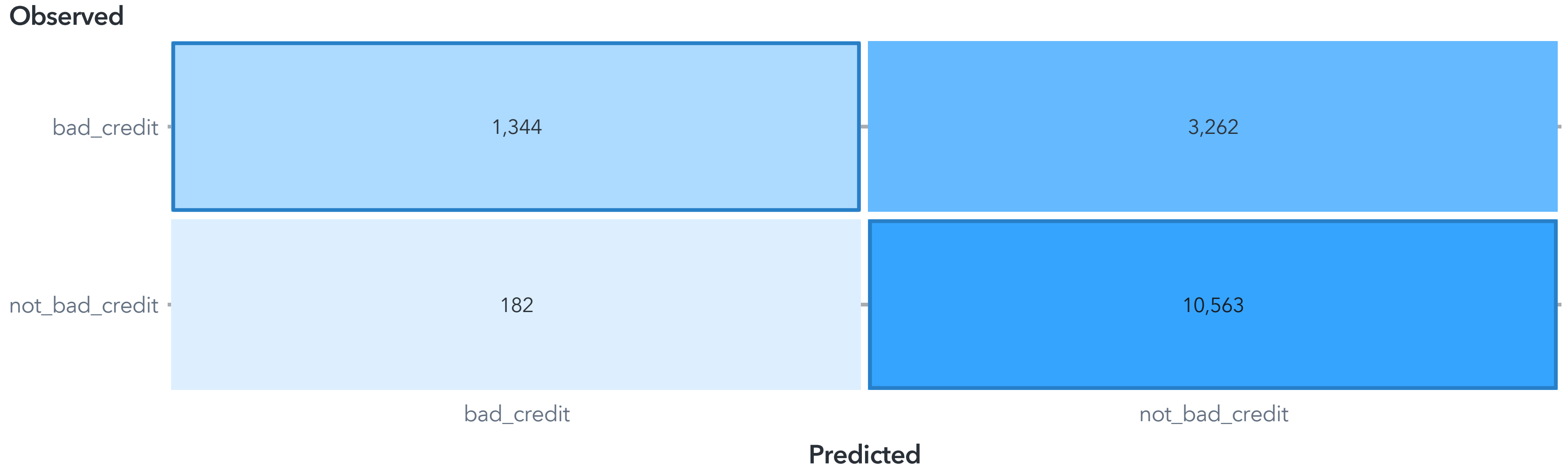




Variable Importance



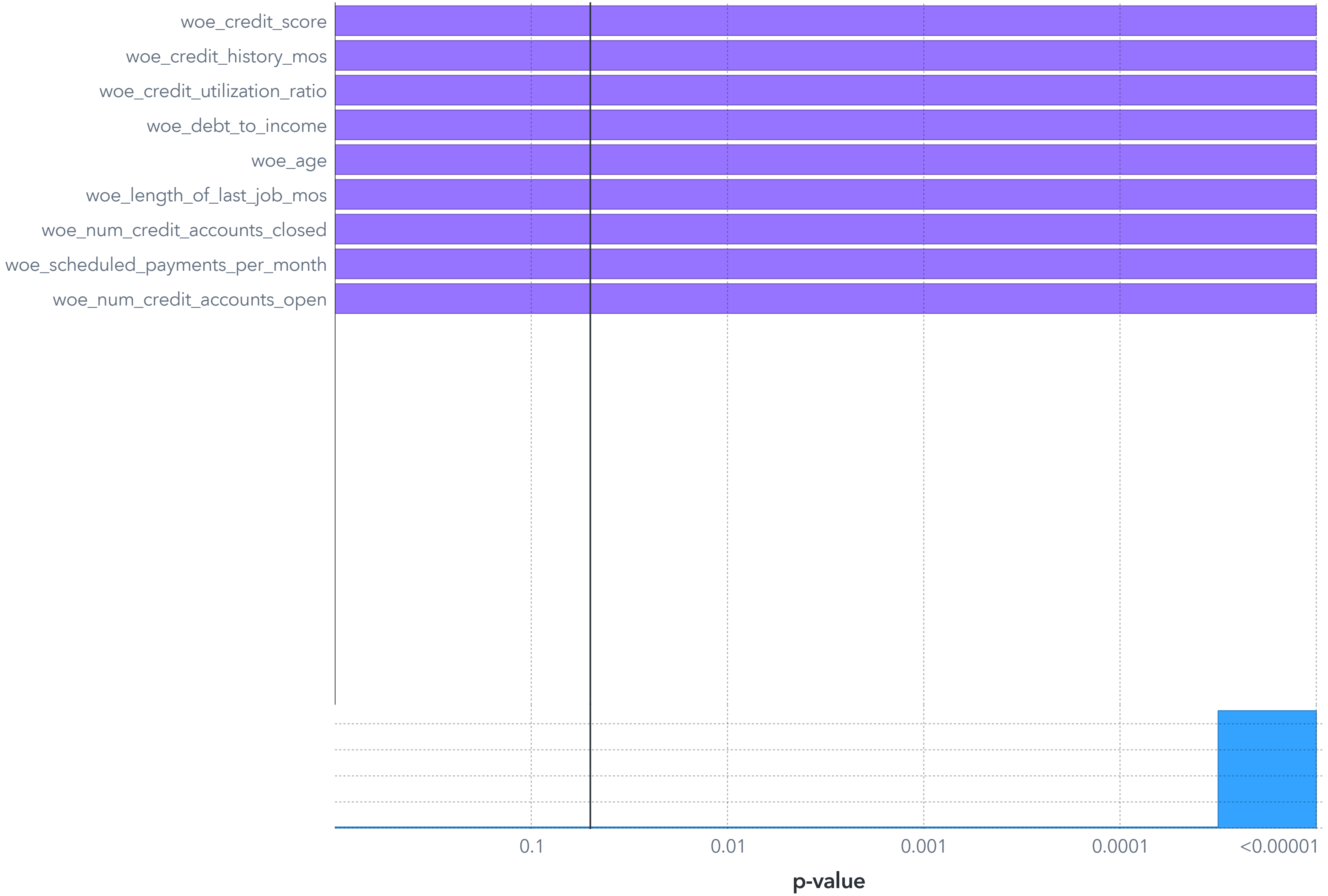
Confusion Matrix



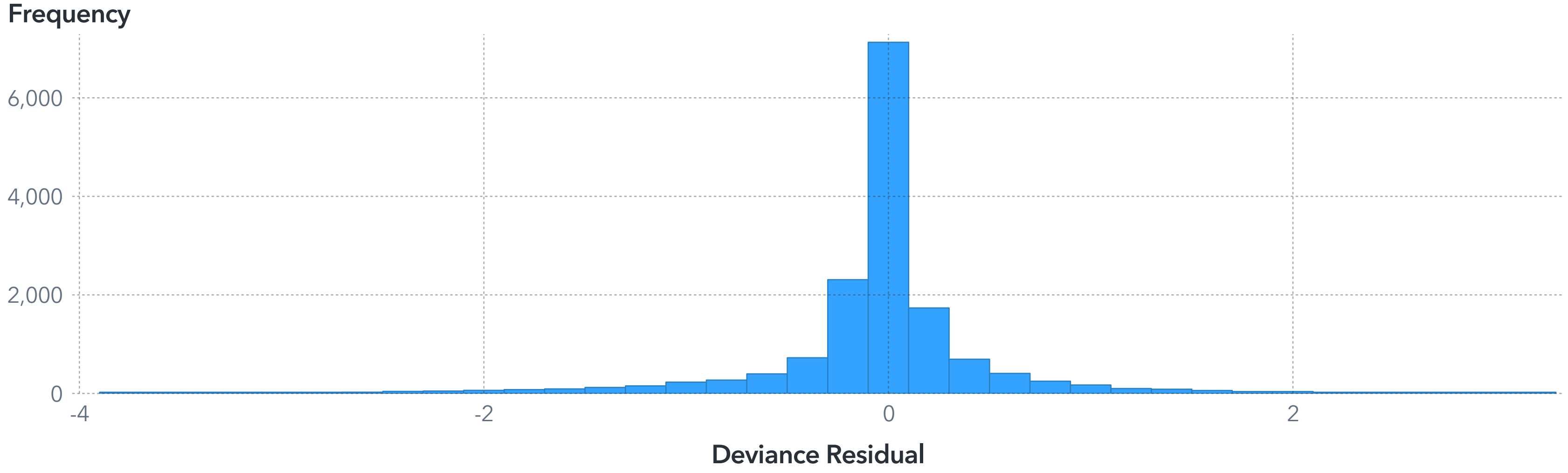
logistic_regression

Logistic Regression **customer_event_char** Event: **not_bad_credit** Fit: **KS (Youden) 0.8610** Observations: **15,351 of 15,351**

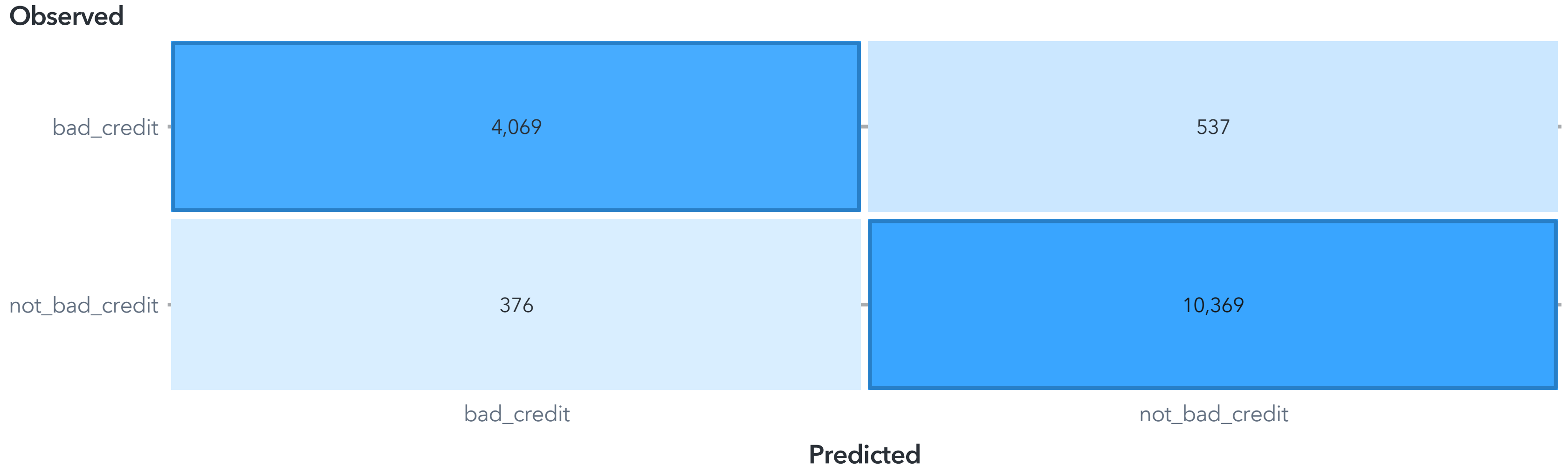
Fit Summary



Residual Plot



Confusion Matrix



scorecard

Target Score

600

500 800

Target Odds

30

1 100

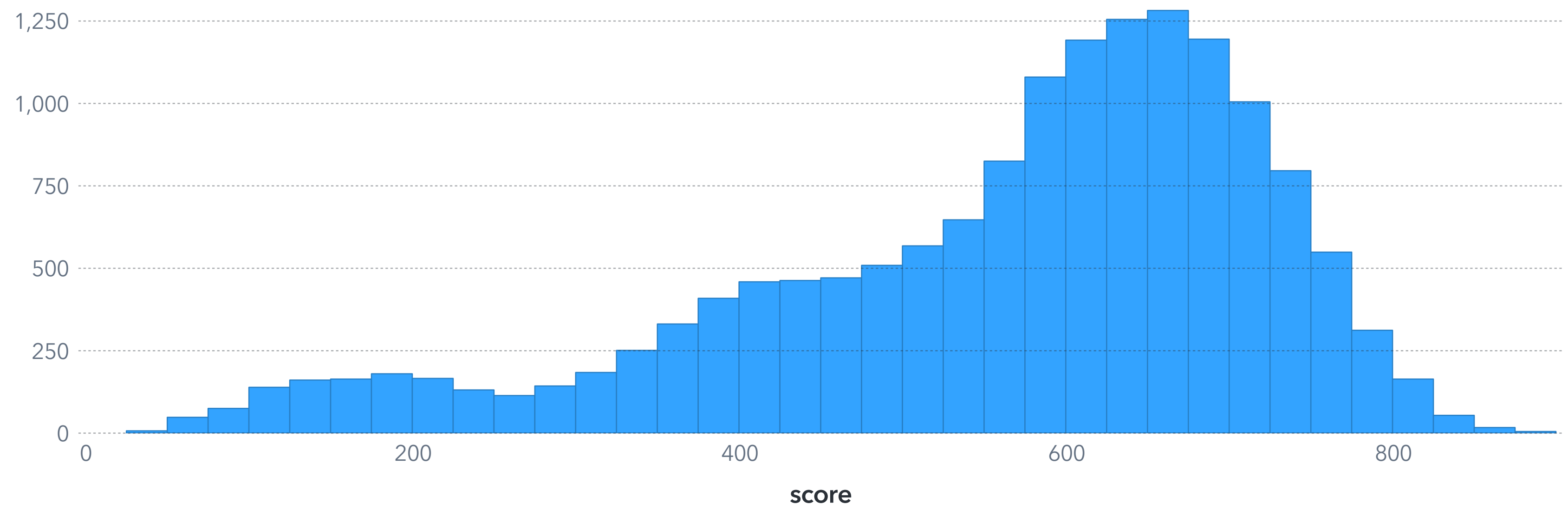
Points to Double the Odds

20

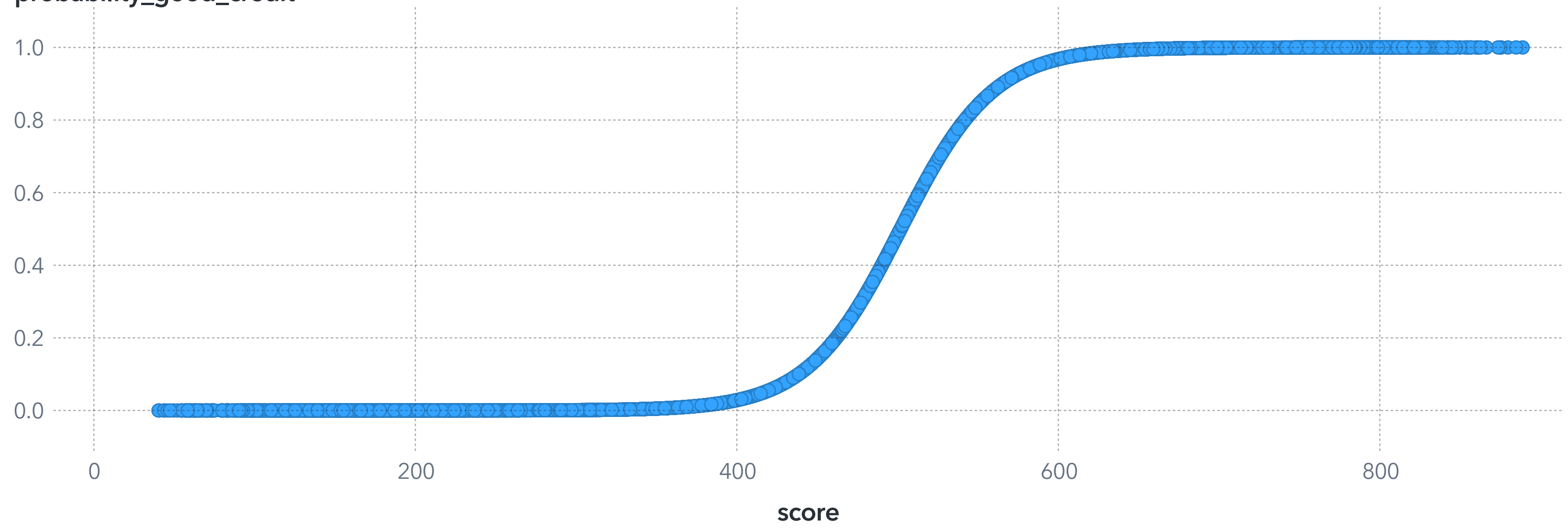
0 100

	account_id ▲	log_odds	odds	probability_good_credit	score
1	1	9.03	8,353.42	99.98803%	762
2	2	1.65	5.20	83.86772%	549
3	3	5.13	169.10	99.41210%	650
4	4	5.79	328.60	99.69660%	669
5	5	5.47	237.24	99.58025%	660
6	6	4.04	57.07	98.27807%	619
7	7	1.28	3.61	78.32004%	539
8	8	4.94	140.40	99.29280%	645
9	9	9.64	15,382.18	99.99350%	780
10	10	8.17	3,545.50	99.97180%	738
11	11	9.11	9,076.73	99.98898%	765

Distribution of Score



probability_good_credit



Appendix

A1.1 credit_report

Parameters:

target_score = 600
target_odds = 30
points_to_double_the_odds = 20

A1.2 Heat Map of Response to Campaigns by Peer ID and Account Type