financial_services_model_explainability

Creation Date: Thursday, October 30, 2025, 03:30:50 PM

Author: chris.parrish@sas.com

Contents

financial_services_model_explainability

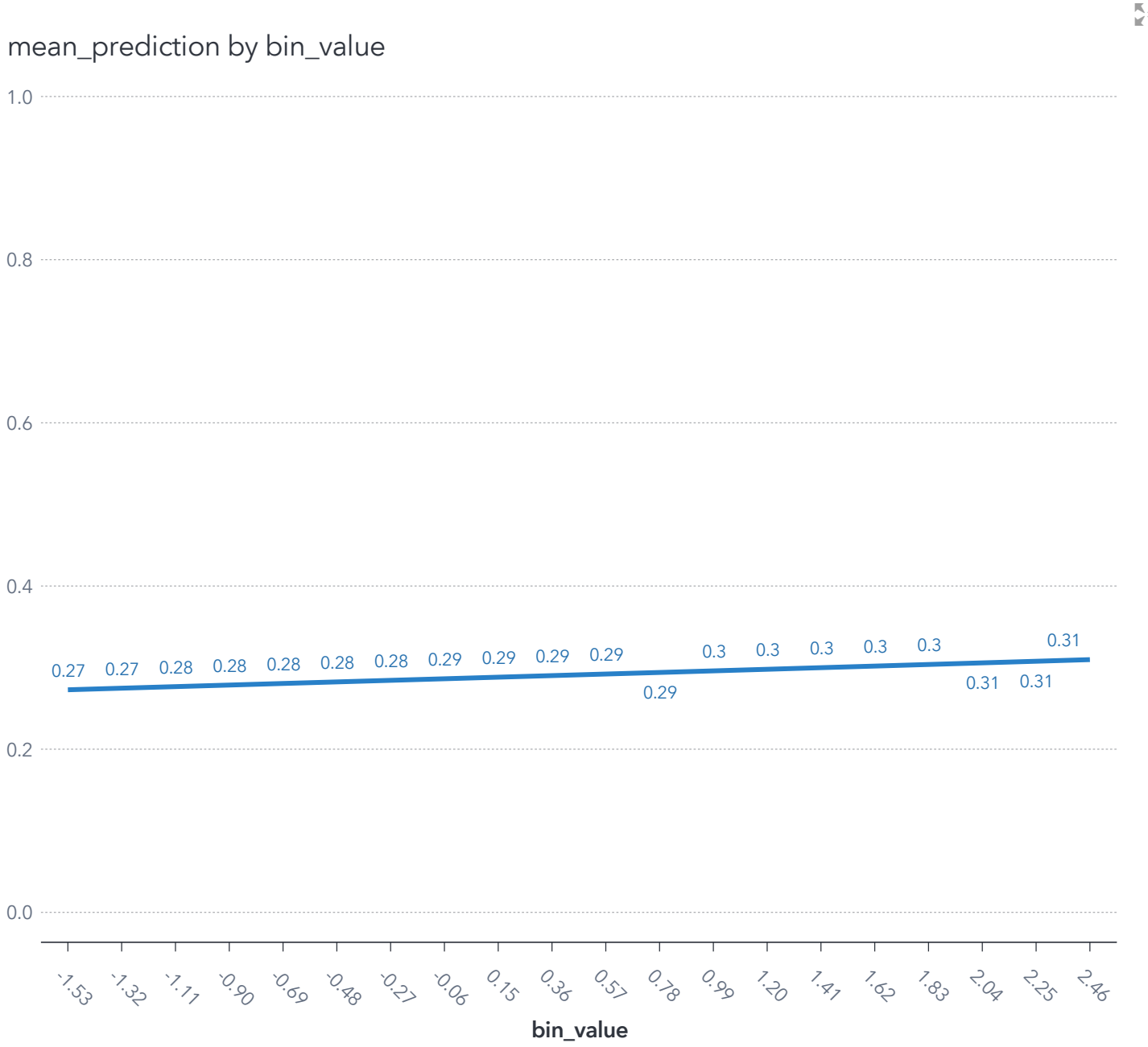
partial_dependence_r_logit

The partial dependency plot shows how the change in the average prediction changes as the input variable changes.

Thus, PD plots show how changes to the values of a model input affect the model's predictions.

PD plots can also visualize how a potentially sensitive variable can impact the average target event value.

For example, if the plot includes a protected characteristic (or its proxy), and the average prediction decreases as the input value increases, it could imply that the higher the value of the protected characteristic then the lower the probability of the event. This could be problematic if the event includes a desired outcome for the customer and the input variable measures the increasing extent of the customer's protected characteristic (or its proxy).



shapley_local_r_logit

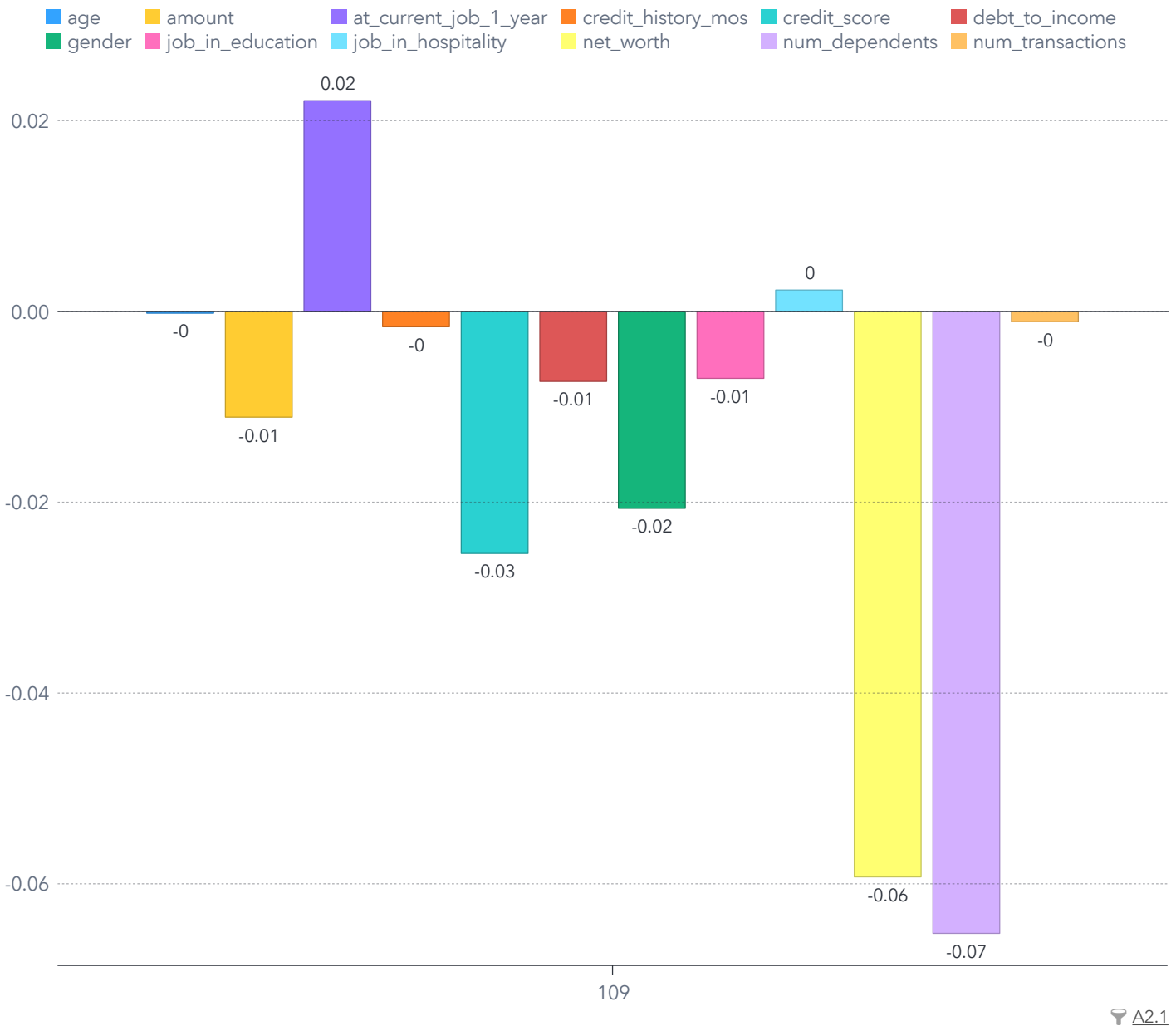
SHAPley values are a measure of variable importance for a single observation.

They are an additive measure of importance, the sum of which equals the predicted output from the model that is being explaining.

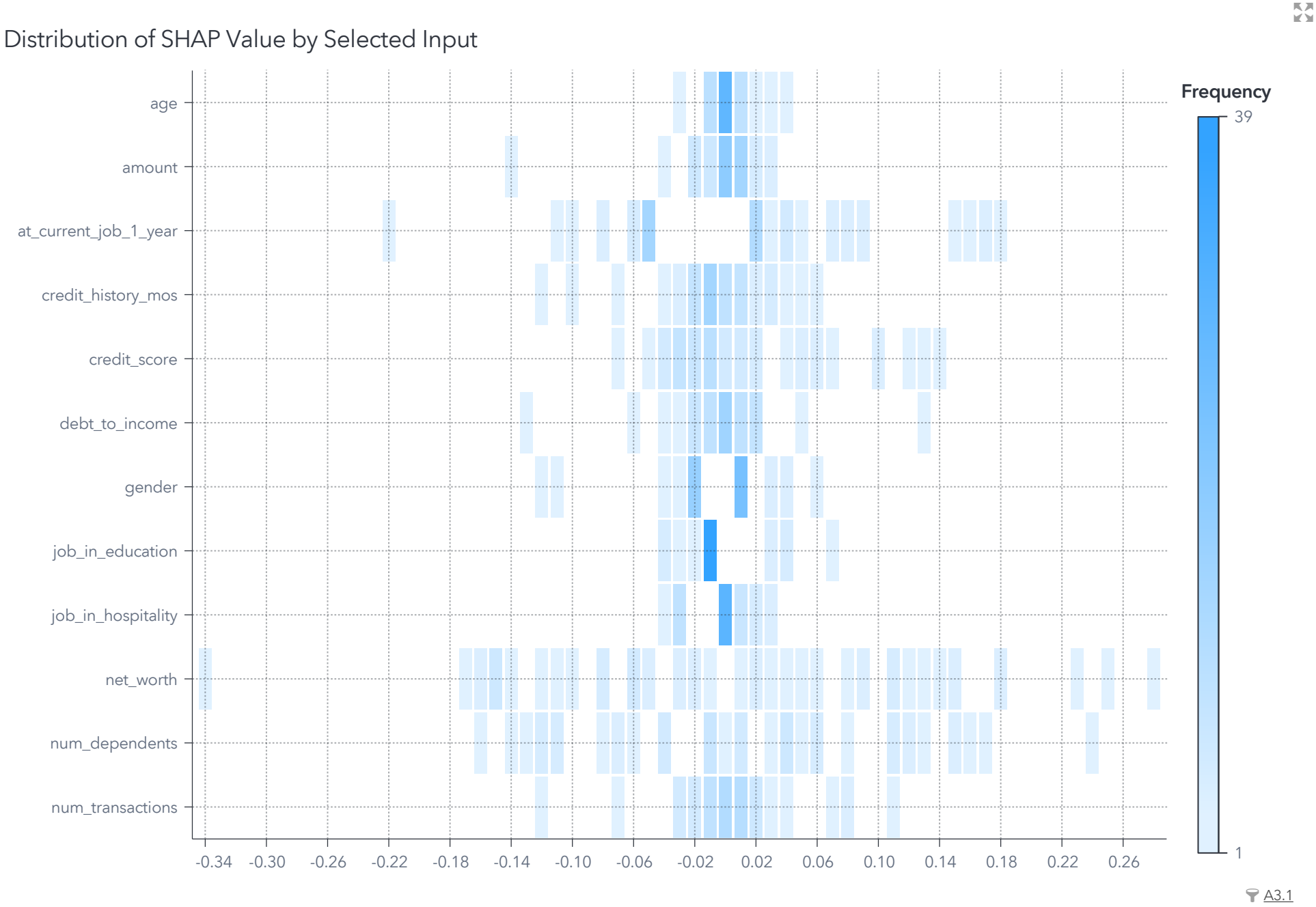
Charting HyperSHAP values can help visualize how a potentially sensitive variable can impact the predicted probability of an individual customer.

For example, if the largest (absolute) value in the SHAP Values chart is -0.32, then it decreases the predicted probability of the event by 0.32.

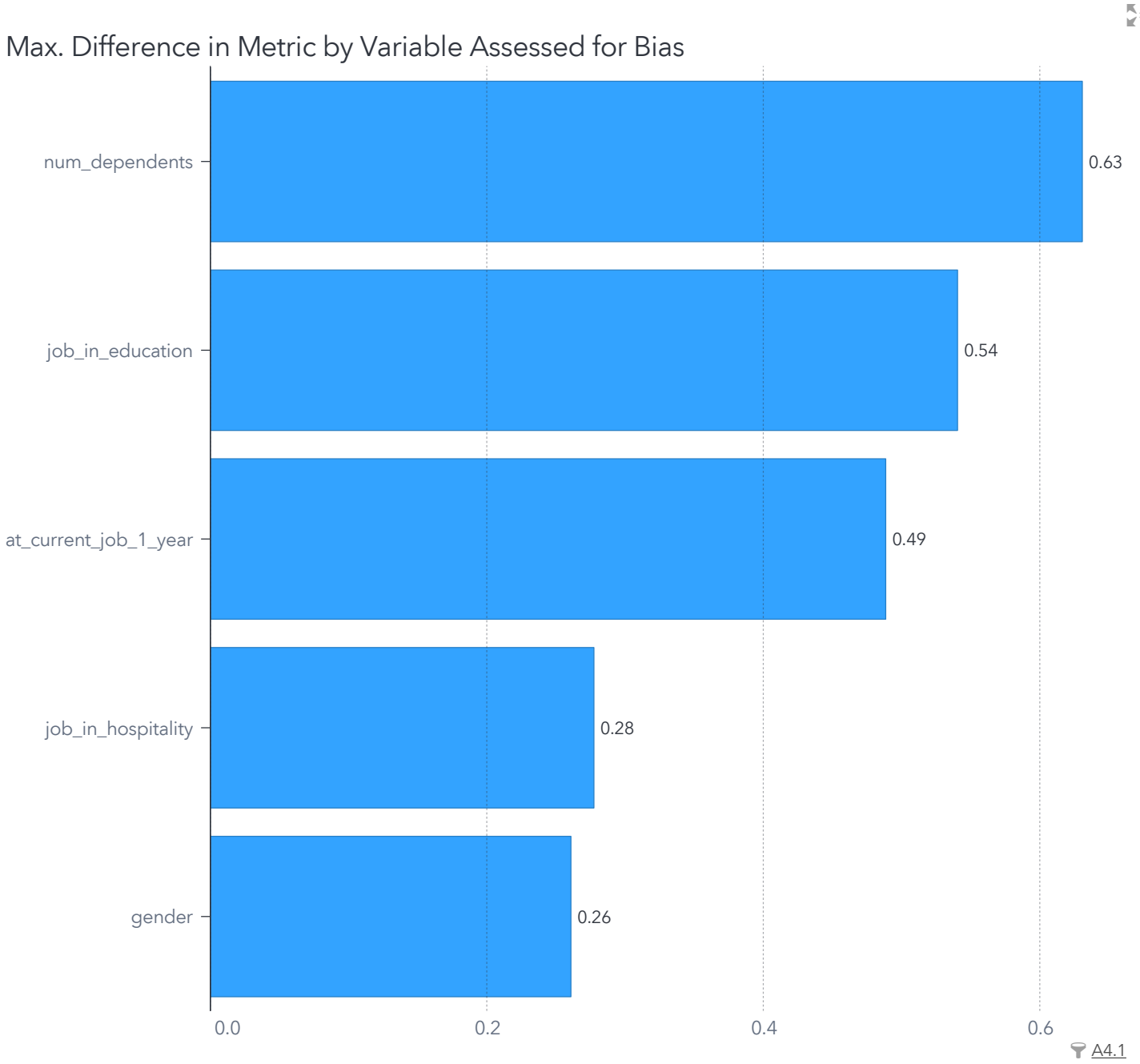
And if this variable was a sensitive variable (e.g., a protected characteristic or its proxy), then it could imply discriminatory behavior if the customer was declined a product based on this sensitive variable.



shapley_dist_r_logit



The maximum differences plot shows the differences in the selected metric in the drop-down box, which may indicate the extent of bias between the base and comparison values of each variable. The base and comparison values can be viewed by hovering over each variable.



Appendix

A1.1 mean_prediction by bin_value

Filters: variable = 'age'

A2.1 SHAP Values

Filters: account_id = 109

A3.1 Distribution of SHAP Value by Selected Input

Filters: (Variable IN { 'age'; 'amount'; 'at_current_job_1_year'; 'credit_history_mos'; 'credit_score'; 'debt_to_income'; 'gender'; 'job_in_education'; 'job_in_hospitality'; 'net_worth'; 'num_dependents'; 'num_transactions' }) OR Variable MISSING

A4.1 Max. Difference in Metric by Variable Assessed for Bias

Filters: MaxDifferences.MetricLabel = 'Average Prediction for Event'

Line - bin_value 1 Results

bin_value ▲	mean_prediction
-1.53	0.2728085054
-1.32	0.2747420775
-1.11	0.2766729224
-0.90	0.2786017647
-0.69	0.2805293279
-0.48	0.2824563327
-0.27	0.2843834952
-0.06	0.2863115258
0.15	0.2882411271
0.36	0.2901729933
0.57	0.2921078088
0.78	0.294046247
0.99	0.2959889695
1.20	0.297936625
1.41	0.2998898491
1.62	0.3018492633
1.83	0.3038154743
2.04	0.3057890741
2.25	0.3077706395
2.46	0.3097607315

Bar - account_id 1 Results

account_id ▲	age ▼	amount	at_current_job_1_year	credit_history_mos	credit_score	debt_to_income	gender	job_in_education	job_in_hospitality	net_worth	num_dependents	num_transactions
109	-0.000189501	-0.011083561	0.0220927107	-0.001607643	-0.025362389	-0.007333794	-0.020639228	-0.007015594	0.0022398988	-0.05927832	-0.06519584	-0.001076165

Heat - ShapleyValue 1 ShapleyValue, Variable

ShapleyValue (lower)	ShapleyValue (upper)	Variable	Frequency
-0.035	-0.025	age	1
0.035	0.045	age	1
0.025	0.035	age	1
-0.005	0.005	age	29
0.015	0.025	age	1
-0.015	-0.005	age	9
0.005	0.015	age	8
-0.005	0.005	amount	19
-0.045	-0.035	amount	1
0.015	0.025	amount	3
-0.025	-0.015	amount	5
0.025	0.035	amount	2
0.005	0.015	amount	14
-0.015	-0.005	amount	5
-0.145	-0.135	amount	1
0.065	0.075	at_current_job_1_year	2
0.155	0.165	at_current_job_1_year	1
0.075	0.085	at_current_job_1_year	2
0.165	0.175	at_current_job_1_year	2
0.035	0.045	at_current_job_1_year	4
0.145	0.155	at_current_job_1_year	1
0.025	0.035	at_current_job_1_year	1
-0.055	-0.045	at_current_job_1_year	14
0.175	0.185	at_current_job_1_year	1
0.085	0.095	at_current_job_1_year	2
-0.065	-0.055	at_current_job_1_year	3
-0.105	-0.095	at_current_job_1_year	1
-0.085	-0.075	at_current_job_1_year	2
-0.115	-0.105	at_current_job_1_year	1
0.045	0.055	at_current_job_1_year	1
-0.225	-0.215	at_current_job_1_year	1
0.015	0.025	at_current_job_1_year	11
0.055	0.065	credit_history_mos	1
0.035	0.045	credit_history_mos	1
0.025	0.035	credit_history_mos	3
0.015	0.025	credit_history_mos	2
0.045	0.055	credit_history_mos	1
-0.015	-0.005	credit_history_mos	14
-0.005	0.005	credit_history_mos	8
-0.025	-0.015	credit_history_mos	7
-0.035	-0.025	credit_history_mos	2
-0.075	-0.065	credit_history_mos	1
-0.105	-0.095	credit_history_mos	1
-0.125	-0.115	credit_history_mos	1
-0.045	-0.035	credit_history_mos	1
0.005	0.015	credit_history_mos	7
0.135	0.145	credit_score	1
0.095	0.105	credit_score	2
0.055	0.065	credit_score	2
0.035	0.045	credit_score	1
0.015	0.025	credit_score	3
0.005	0.015	credit_score	4
0.045	0.055	credit_score	2
-0.005	0.005	credit_score	4
-0.015	-0.005	credit_score	9
0.065	0.075	credit_score	1
-0.025	-0.015	credit_score	6
-0.035	-0.025	credit_score	7
0.125	0.135	credit_score	1
-0.045	-0.035	credit_score	4
-0.055	-0.045	credit_score	1
0.115	0.125	credit_score	1
-0.075	-0.065	credit_score	1
-0.035	-0.025	debt_to_income	1
0.015	0.025	debt_to_income	7
0.005	0.015	debt_to_income	8
-0.005	0.005	debt_to_income	15
-0.015	-0.005	debt_to_income	8
-0.025	-0.015	debt_to_income	6
0.045	0.055	debt_to_income	1
0.125	0.135	debt_to_income	1
-0.045	-0.035	debt_to_income	1
-0.065	-0.055	debt_to_income	1
-0.135	-0.125	debt_to_income	1
0.055	0.065	gender	1
0.035	0.045	gender	2
0.025	0.035	gender	2
0.005	0.015	gender	24
-0.025	-0.015	gender	17
-0.035	-0.025	gender	1
-0.045	-0.035	gender	1
-0.115	-0.105	gender	1
-0.125	-0.115	gender	1
0.065	0.075	job_in_education	1
0.035	0.045	job_in_education	3
0.025	0.035	job_in_education	2
-0.015	-0.005	job_in_education	39
-0.025	-0.015	job_in_education	1
-0.035	-0.025	job_in_education	1
-0.045	-0.035	job_in_education	3
0.025	0.035	job_in_hospitality	2
0.015	0.025	job_in_hospitality	3
0.005	0.015	job_in_hospitality	6
-0.005	0.005	job_in_hospitality	30
-0.035	-0.025	job_in_hospitality	8
-0.045	-0.035	job_in_hospitality	1
0.245	0.255	net_worth	1
0.225	0.235	net_worth	1
0.175	0.185	net_worth	2
0.135	0.145	net_worth	1
0.125	0.135	net_worth	1
0.115	0.125	net_worth	1
0.085	0.095	net_worth	2
0.075	0.085	net_worth	1
0.055	0.065	net_worth	1
0.045	0.055	net_worth	1
0.035	0.045	net_worth	1
0.025	0.035	net_worth	1
0.015	0.025	net_worth	2
0.005	0.015	net_worth	1
-0.015	-0.005	net_worth	1
-0.025	-0.015	net_worth	2
-0.035	-0.025	net_worth	2
-0.055	-0.045	net_worth	2
-0.065	-0.055	net_worth	4
-0.085	-0.075	net_worth	3
-0.105	-0.095	net_worth	1
0.105	0.115	net_worth	3
-0.115	-0.105	net_worth	1
-0.125	-0.115	net_worth	1
-0.145	-0.135	net_worth	1
0.145	0.155	net_worth	2
-0.155	-0.145	net_worth	5
-0.165	-0.155	net_worth	2
-0.175	-0.165	net_worth	1
0.275	0.285	net_worth	1
-0.345	-0.335	net_worth	1
0.165	0.175	num_dependents	1
0.155	0.165	num_dependents	1
0.145	0.155	num_dependents	2
0.125	0.135	num_dependents	1
0.115	0.125	num_dependents	2
0.105	0.115	num_dependents	3
0.075	0.085	num_dependents	1
0.045	0.055	num_dependents	1
0.035	0.045	num_dependents	5
0.025	0.035	num_dependents	1
0.005	0.015	num_dependents	4
-0.005	0.005	num_dependents	1
-0.015	-0.005	num_dependents	6
-0.045	-0.035	num_dependents	4
0.055	0.065	num_dependents	3
-0.065	-0.055	num_dependents	1
-0.075	-0.065	num_dependents	2
-0.085	-0.075	num_dependents	1
-0.115	-0.105	num_dependents	3
-0.125	-0.115	num_dependents	3
-0.135	-0.125	num_dependents	1
-0.145	-0.135	num_dependents	1
0.235	0.245	num_dependents	1
-0.165	-0.155	num_dependents	1
0.105	0.115	num_transactions	1
0.075	0.085	num_transactions	2
0.065	0.075	num_transactions	1
0.035	0.045	num_transactions	1
0.025	0.035	num_transactions	1
0.015	0.025	num_transactions	5
0.005	0.015	num_transactions	10
-0.005	0.005	num_transactions	11
-0.015	-0.005	num_transactions	8
-0.025	-0.015	num_transactions	4
-0.035	-0.025	num_transactions	4
-0.075	-0.065	num_transactions	1
-0.125	-0.115	num_transactions	1

Bar - bias_var 1 Results

bias_var	MaxDifferences.Value
num_dependents	0.6307643342
job_in_education	0.5405194138
at_current_job_1_year	0.4885545495
job_in_hospitality	0.2775080545
gender	0.2608617321