

American Express Campus Challenge

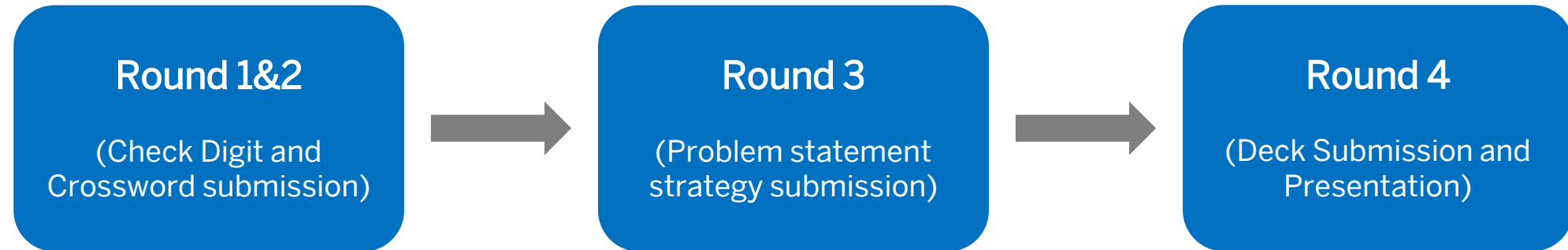
Strategy Problem Statement: Formulate efficient strategy to identify Fraudsters

A person wearing a dark hoodie is sitting at a desk, looking at a laptop. A large, light blue question mark is superimposed over their face. The background is a dark blue wall with a pattern of falling numbers, similar to the 'Matrix' effect. The overall lighting is dim, with a blue tint.

Background

- Did you know that credit card fraud results in billions of dollars in losses annually?
- Fraud is a deliberate act of deception intended to secure unfair or unlawful gain. It involves misrepresenting or concealing information, leading to financial or personal loss.
- We encourage all participants to don their detective hats and devise strategy to identify fraudsters as part of 2024 Amex Campus Challenge.

Stages of Competition



Round 1 & 2

Round 1:

Guessing the 15th digit – Guess the 15th digit for a 14-digit account number.

A credit card company or any financial organization uses the last digit of an account number to validate the authenticity of the account number. This is done using certain logic and is one of the basic checks done to detect errors in the account number.

Evaluation Process :

Anyone guessing the correct 15th digit sequence, will unlock the next around. All participants with correct guesses will have the same score for this round.

Round 2:

Crossword Puzzle – Solve the crossword puzzle related to terminologies used in finance industry & correctly guess the highlighted words in the crossword.

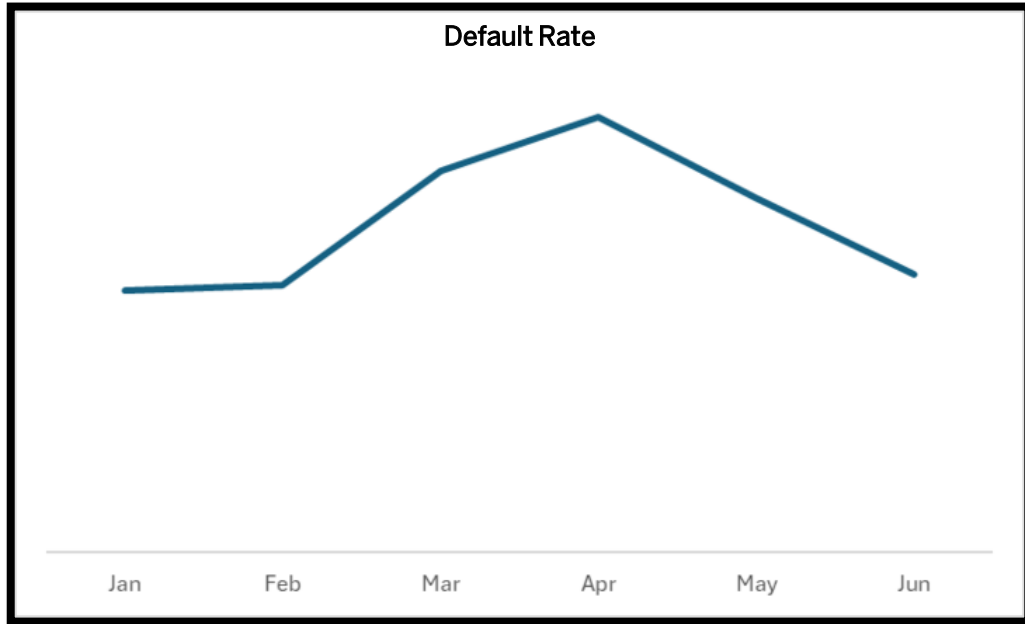
Evaluation Process :

Anyone guessing the right word, will unlock the next around. All participants with correct guesses will have the same score for this round.

Submission for Round 1 and Round 2 will be combined. No cap on number of submissions for Round 1 & 2.

Round 3

Formulate optimal strategy to identify fraudulent individuals with similar profiles.



- There is a surge in fraud losses for certain months.
- Surge in fraud losses is caused due to few fraudsters who have exploited gaps in the acquisition system. These fraudsters have shown a similar pattern in they have defaulted.
- Participants need to devise an optimal strategy to identify **ONLY those fraudsters who are responsible for the surge in the default rate**, with minimal impact to good customers.
- Data shared includes variables & details from onboarding, spend & default indicator.
- Independent or derived variables can be leveraged in the strategy

Evaluation Criteria: Hit Rate * Coverage * Feature Score

- *Hit rate : Number of fraudsters captured by the strategy/Total unique IDs captured by the strategy*
- *Coverage : Number of fraudsters captured by the strategy/Total fraudsters in the data*
- *Feature Score: Score based on number of variables used in the strategy**

*If number of variables used in the strategy between 1 to 5 then feature score is 1; between 6 to 8 then feature score is 0.9; between 9 to 11 then feature score is 0.8; between 12 to 14 then feature score is 0.7; are more than 15 then feature score is 0.5;

Data Details

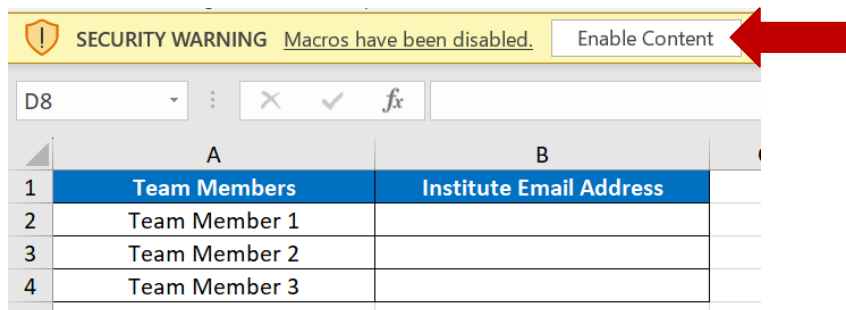
Data

Round	Theme	File Name	Desc	Data shared	Location/Name
Round 1	Check Digit	r1_r2_data.xlsx	Find the check digit of the account numbers shared	10 Account numbers	<i>To be downloaded from the unstop website (available to all registered candidates)</i>
Round 2	Crossword		Fill the crossword related to credit industry based on hints shared	15 words	
Round 3	Fraud	r3_data.xlsx	Create a strategy to find the reason for increased default rate	62484 x 61 (Rows x Columns)	<i>To be downloaded from the unstop website (available only after completion of R1 and R2)</i>

- Next round will only be unlocked after successfully solving current round – i.e., Round 2 will be unlocked after solving Round 1. Round 3 will be unlocked after solving Round 2.
- For Round 3, Modelling solution is not allowed; derived variables, if used, should not be created using any modelling algorithm.

Round 1 & 2 – Input Sheet Guidelines

- After opening the excel file of Round 1 and 2, a security warning pops up that macros are disabled. Please enable macros, by clicking “Enable Content” to continue further.



- Maximum number of participants in a team are 3. A team can have either 1 or 2 participants as well.
- All the members of the team must mandatorily fill their email addresses.
- Only official institute email address must be filled in the "Institute Email Address" column.
- Please make sure the email addresses entered are correct before proceeding further.
- After email addresses are filled, click on the "CLICK HERE TO UNLOCK ROUND 1" button to unlock Round 1.
- Once email address are filled and Round 1 is unlocked, do not change/edit the email addresses.
- If you have entered incorrect email addresses, correct them and start from the beginning by clicking on the "CLICK HERE TO UNLOCK ROUND 1" BUTTON again.

Sample Data - Round 1

Account numbers

14 Digits	Guess the 15th Digit
XXXXXXXXXXXXXX	1
XXXXXXXXXXXXXX	4
XXXXXXXXXXXXXX	7
XXXXXXXXXXXXXX	.
XXXXXXXXXXXXXX	.
XXXXXXXXXXXXXX	.
XXXXXXXXXXXXXX	.
XXXXXXXXXXXXXX	.
XXXXXXXXXXXXXX	.
XXXXXXXXXXXXXX	.
NUMBER PASSWORD	147.....
WORD PASSWORD	

Participants need to fill the check digit of the Account number shared here

This will automatically concatenate the digits entered and shows here

This will automatically populate once 15th digit is entered for all 10 Account numbers

Sample Data - Round 2

	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
2														P2				
3									L3									
4									K4									
5																		
6									I6									
7																		
8									I8								S8	
9												N9						
10																		
11		D11			G11						M11		O11					
12	C12																	
13																		
14	C14																	
15																		
16																		
17																		
18																		
19											M19							

Across:

- L3 - Hint 1
- I6 - Hint 2
- N9 - Hint 3
- C12 - Hint 4
- C14 - Hint 5
- M19 - Hint 6

Down:

- P2 - Hint 1
- K4 - Hint 2
- I8 - Hint 3
- S8 - Hint 4
- D11 - Hint 5
- G11 - Hint 6
- M11 - Hint 7
- O11 - Hint 8

Unscramble the letters highlighted in yellow to get the password for the next round

S N S I U E B S

Letters highlighted in yellow

B U S I N E S S

Unscrambled word – to be used as password for Round 3

Round 1 Guidelines

- The 15th digit could only be an integer value between 0 and 9 (inclusive of 0 and 9).
- Please enter the 15th digit in the column "Guess the 15th Digit" against the Account Number in the column "14 Digits".
- The 15th digit need not necessarily be a unique digit for the 10 Accounts shared.
- All the 15th digits should be entered to unlock next round.
- The Number Password will automatically concatenate and show you the value.
- A random Word Password is generated, once you fill the 15th digit of all ten Accounts.
- Once all 15th digits of all 10 Accounts are entered, click on the "CLICK HERE TO UNLOCK ROUND 2" button, to unlock Round 2.
- Round 2 will only be unlocked if you have entered the correct 15th digit for the 10 Accounts shared.
- You will be asked to submit both the Number Password and the Word Password for evaluation.
- Fill details asked in submission template and upload on the website in CSV format only.
- Only submit one row in the submission csv file.
- Participants are required to download and follow exact templates of the submission files from Unstop website.
- Submission file for Round 1 and Round 2 is combined and can't be made separately.
- Amex will thoroughly evaluate the solution to ensure integrity & guarding against any misuse or gaming.

Round 2 Guidelines

- The words in the crossword are related to common terminologies used in Credit Card Industry.
- Hints to fill the Crossword are provided in that sheet itself.
- The cell in which hint cell number is mentioned is also part of the word.
- Cells highlighted in yellow in the Crossword needs to be unscrambled to get a word relevant to the Credit Card Industry
- All the cells highlighted in yellow are to be used to form the unscrambled word.
- You will be asked to submit the unscrambled word as well as the completed crossword for evaluation.
- The unscrambled word will also act as a password for Round 3, the file for which will be shared separately.
- Fill details asked in submission template and upload on the website in CSV format only.
- Only submit one row in the submission csv file.
- Participants are required to download and follow exact templates of the submission files from Unstop website.
- Submission file for Round 1 and Round 2 is combined and can't be made separately.
- Amex will thoroughly evaluate the solution to ensure integrity & guarding against any misuse or gaming.

Sample Submission File 1 – Round 1&2

Template:

team_name	team_member1_email	team_member2_email	team_member3_email	number_password	word_password	crossword_password

Example 1:

team_name	team_member1_email	team_member2_email	team_member3_email	number_password	word_password	crossword_password
Amexians	karan_zzz@iitz.ac.in	arjun_pandav@iitz.ac.in	bheesm_guru@iitz.ac.in	1228890465	HELLO	COMMERCE

Example 2:

team_name	team_member1_email	team_member2_email	team_member3_email	number_password	word_password	crossword_password
Lone Wolf	ram.sri@iim.in			8439994713	DUMMY	PASSWORD

The submissions should follow the naming convention: *r1_r2_submission_<team-name>.csv*

Sample Submission File 2 – Round 2

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD
1																														
2																														
3																														
4																														
5																														
6																														
7																														
8																														
9																														
10																														
11																														
12																														
13																														
14																														
15																														
16																														
17																														
18																														
19																														
20																														
21																														
22																														
23																														
24																														

Words can go across or down.

ACROSS

L3. hint.....

I6. hint.....

N9. hint.....

C12. hint.....

C14. hint.....

M19. hint.....

DOWN

P2. hint.....

K4. hint.....

I8. hint.....

S8. hint.....

D11. hint.....

G11. hint.....

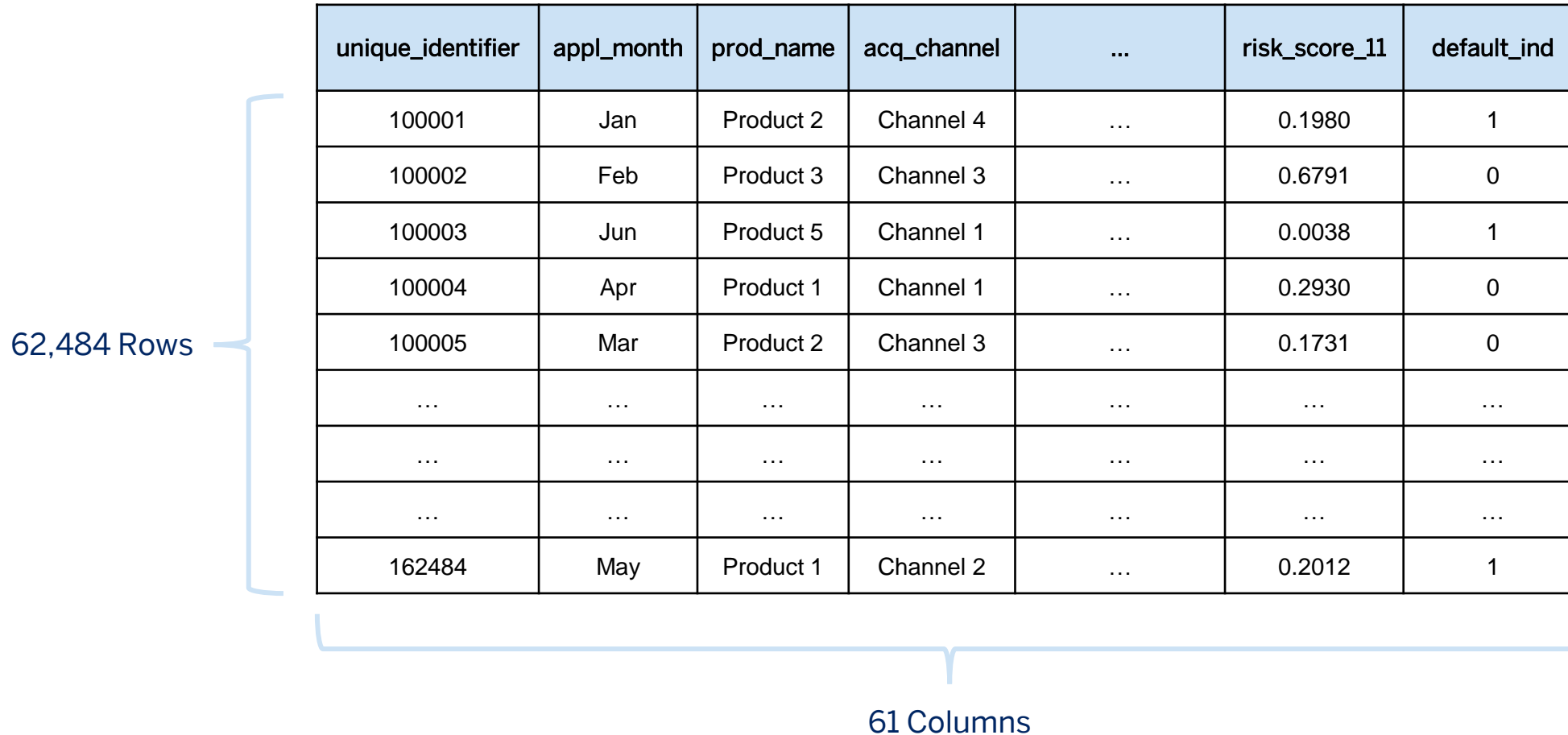
M11. hint.....

O11. hint.....

Unscramble the letters highlighted in yellow to get the password for the next round

The submissions should follow the naming convention: *r1_r2_crossword_<team-name>.xlsx*

Sample Data - Round 3



62,484 Rows

unique_identifier	apl_month	prod_name	acq_channel	...	risk_score_11	default_ind
100001	Jan	Product 2	Channel 4	...	0.1980	1
100002	Feb	Product 3	Channel 3	...	0.6791	0
100003	Jun	Product 5	Channel 1	...	0.0038	1
100004	Apr	Product 1	Channel 1	...	0.2930	0
100005	Mar	Product 2	Channel 3	...	0.1731	0
...
...
...
162484	May	Product 1	Channel 2	...	0.2012	1

61 Columns

The strategy that teams will create should capture the segment of defaulters that is responsible for spike in the default rate. However, strategy created will be evaluated on the complete data.

Sample Data – Round 3

#	Variable Name	Description
1	unique_identifier	Unique Identifier
2	appl_month	Application Month
3	prod_name	Product Name
4	acq_channel	Acquisition Channel
5	state_code	State Code
6	bureau_src	Bureau Source
7	risk_score_1	Risk score assigned based on profitability
8	bureau_score	Bureau score
9	limit	Assigned credit limit
10	income	Income of applicant
11	old_open_trd_ms	Months since the Oldest Open Trade
12	no_open_trd	Total Open Trades
13	no_trd	Number of trades
14	debt_cap	Debt Capacity
15	basic_max_trd_tnr	Maximum individual trade tenure
16	basic_no_trd	Individual Trade Count
17	risk_score_2	Risk score assigned based on probability of default
18	trust_identity	Positive personal identification indicator
19	risk_score_3	Risk score assigned based on probability of fraud
20	no_inq	Number of trade Inquiries
21	risk_score_4	Risk score assigned based on the applicant's email
22	income_incons	Income inconsistency
23	bureau_mismatch	Application bureau data inconsistency score; Higher value means higher mismatches
24	risk_score_5	Risk score assigned based on the applicant's phone number
25	risk_score_6	Risk score assigned based on the applicant's address
26	addr_mismatch	Address mismatch indicator
27	basic_old_open_trd_ms	Number of months since the oldest basic trade has been opened
28	addr_ds	Days since applicant is staying at this address
29	dl_id	Sources confirming the applicant's driver license
30	addr_changes	Indicates how often applicant has changed addresses

#	Variable Name	Description
31	no_phn_to_addr	Total number of unique phone numbers associated with input address
32	no_nid_to_addr	Total number of unique National IDs associated with input address
33	no_uid_to_addr	Total number of unique identities associated with input address on multiple records
34	uid_addr_matches	Sources confirming the applicant's identity and current address
35	uid_nid_matches	Sources confirming the applicant's national ID
36	home_value	Home Value
37	no_addr_to_nid	Total number of unique addresses associated with national ID
38	risk_score_7	Risk score assigned based on the applicant's identity
39	no_trd_del	Major derog count
40	spend	Total amount spent
41	payments	Total amount of payments made
42	return_payments	Total amount of payments failed
43	attempt_txn	Total transactions attempted
44	decline_txn	Total transactions declined
45	merchant1	Merchant 1
46	merchant1_amt	The total amount spent at merchant 1
47	merchant1_ten	Tenure of merchant 1
48	merchant1_cat	Merchant 1 Category
49	merchant2	Merchant 2
50	merchant2_amt	The total amount spent at merchant 2
51	merchant2_ten	Tenure of merchant 2
52	merchant2_cat	Merchant 2 Category
53	merchant3	Merchant 3
54	merchant3_amt	The total amount spent at merchant 3
55	merchant3_ten	Tenure of merchant 3
56	merchant3_cat	Merchant 3 Category
57	risk_score_8	Risk score assigned based on probability of bankruptcy
58	risk_score_9	Closed loop risk
59	risk_score_10	Transactional risk
60	risk_score_11	Risk score assigned based on applicant's demographics
61	default_ind	1 indicates that applicant has defaulted, 0 indicates that applicant has not defaulted

Round 3 Guidelines

- Please use the unscrambled word in uppercase from Round 2 as password for opening the excel file for Round 3.
- Participants can use existing variables, create derived variables from existing ones and use them in their strategy
- Participants must not use these variables in their strategy – unique_identifier, appl_month, default_ind.
- Participants are not allowed to add new rows or alter the shared data.
- The final strategy that is created must be run on all unique_identifiers.
- no_of_vars in the submission file, is how many unique variables from the dataset shared that the team is using in their strategy.
- Participants need to submit 2 files in Round 3. 1 CSV file with unique_identifier, strategy hit indicator and no_of_vars. Other is an Excel workbook with the exact strategy used to attain max score. Derived variables, if used, are to be mentioned with their logic.
- Participants should try to create a strategy that is generalized, keeping in mind the evaluation criteria.
- Modelling solution is not allowed; derived variables, if used, should not be created using any modelling algorithm.
- Participants are required to download and follow exact templates of the submission files from Unstop website.
- Max submissions allowed per team in Round 3 are 10, leaderboard would be public and team would be rank ordered basis max final score attained.
- Amex will be thoroughly evaluating all solutions to ensure integrity & guarding against any misuse or gaming.

Sample Submission File 1 – Round 3

Template:

unique_identifier	strategy_hit	no_of_vars
100001		
100002		
100003		
100004		
...		
...		
...		
...		
...		
...		
162483		
162484		

Example 1:

unique_identifier	strategy_hit	no_of_vars
100001	1	9
100002	0	9
100003	1	9
100004	1	9
...
...
...
...
...
...
...
162483	0	9
162484	0	9

The submissions should follow the naming convention: *r3_submission_<team-name>.csv*

Sample Submission File 2 – Round 3

	A	B	C	D	E	F	G
1	Final Strategy						
2	(variable_1 >= 100 or derived_variable_2 > 100) and (variable_2 = 'true' or variable_3 = 1) and derived_variable_1 = 0.1 and (variable_4 = 'Y' or variable_5 <= 10)						
3							
4	Derived Variables						
5	#	Derived Variable	Derivation Logic	Description			
6	1	derived_variable_1	variable_6/(variable_7 + variable_8)	a small description that gives details of the derived_variable_1			
7	2	derived_variable_2	variable_9*variable_10	a small description that gives details of the derived_variable_2			
8	3	.	.	.			
9	4	.	.	.			
10							

- In the above example, no_of_vars used will be 10, as 10 unique variables from the Round 3 dataset shared is used in the strategy.
- The below table gives details about the derivation logic of the derived variable, if used, and a short description about that variable.

The submissions should follow the naming convention: *r3_strategy_<team-name>.xlsx*

Final Round Guidelines

- Top teams with combined evaluation score across Round 1, Round 2, Round 3, and who qualify Amex sanity checks, will be shortlisted for the Final Round.
- Shortlisted teams will also create a presentation detailing their approach including (but not limited to) Profiling Technique, Feature Innovation, Intuitiveness, Selection, Framework used, Presentation, QnA etc. They will be asked to present the same to a panel.

Top 3 teams will be selected as winners based on Round 3 score as well as scores from the presentation.



All the Best!!!