New Customer Profile Table

A new customer profile table will be created to contain customers that have at least one transaction in transaction table.
Also, mean_amount, std_amount, will be calculated using data from original transaction table.

This new table is created for engineering features in transaction table. It is not intended to be used for model training.

CUSTOMER_ID		Done
x_customer_id	Same as original	y
y_customer_id	Same as original	y
mean_amount		y
std_amount		y
mean_nb_tx_per_day	Same as original	y
available_terminals	Same as original	y
nb_terminals	Same as original	y
mean_terminal_distance		y
std_terminal_distance		y

New Transaction Table

This will be the table for model training/testing.

Existing columns in original transaction table

TRANSACTION_ID	
TX_DATETIME	
CUSTOMER_ID	
TERMINAL_ID	
TX_AMOUNT	
TX_FRAUD	
TX_FRAUD_SCENARIO	For debugging only

Derived/Engineered features (making use of new customer profile table)

TX_TIME_DAYS (already exists)	Number of days since dataset starts day	int	Given	Y
TX_TIME_SECONDS (already exists)	Number of seconds since dataset starts day	int	given	Y
TX_LAST_DATETIME	Last transaction datetime of same customer	dt	Fiona	Y
TX_LAST_DAYS	Number of days between TX_DATETIME and TX_LAST_DATETIME	int	Fiona	Y
TX_LAST_HOURS	Number of hours between TX_DATETIME and TX_LAST_DATETIME	float	Fiona	Y
TX_TIME_HOUR_BIN_0	1 if TX_DATETIME between 00h - 03h	int	Dale	Y
TX_TIME_HOUR_BIN_1	1 if TX_DATETIME between 04h - 07h	int	Dale	Y
TX_TIME_HOUR_BIN_2	1 if TX_DATETIME between 08h - 11h	int	Dale	Y
TX_TIME_HOUR_BIN_3	1 if TX_DATETIME between 12h - 15h	int	dale	Y
TX_TIME_HOUR_BIN_4	1 if TX_DATETIME between 16h - 19h	int	Dale	Y
TX_TIME_HOUR_BIN_5	1 if TX_DATETIME between 20h - 23h	int	Dale	Y
AMOUNT_Z_SCORE		float	Ralph	Y
CUSTOMER_ID_AVG_AMOUNT_1DAY_WINDOW	Average spending amount in the last n day(s), for n in {1,7,30}, for customer ID	float	Dale	

CUSTOMER_ID_AVG_AMOUNT_7DAY_WINDOW			Dale	
CUSTOMER_ID_AVG_AMOUNT_30DAY_WINDOW			Dale	
CUSTOMER_ID_AVG_AMOUNT_2REC	Average spending amount in the last n records for n in {2,10}, for customer ID	float	Dale	
CUSTOMER_ID_AVG_AMOUNT_10REC			Dale	
CUSTOMER_TERMINAL_DISTANCE_Z_SCORE		float	Ralph	
TERMINAL_ID_RISK_2DAY_WINDOW	Average number of frauds on the terminal in the last n+d day(s), for n in {1,7,30} and d=7. The parameter d is called delay	float	Dale	
TERMINAL_ID_RISK_7DAY_WINDOW			Dale	
ONE_DOLLAR_FOUND	A one dollar (or very small amount, say $0.5-1.5) transaction found on last 24 hours	1 or 0	Queenie	Y
SUM_TX_AMOUNT_CUSOMTER_ID_SAME_TERMINAL_SAME_DAY	Total amount spent by the customer in the same terminal ID on the same day	float	Fiona	
nb_TX_CUSOMTER_ID_SAME_TERMINAL_SAME_DAY	No. of transactions of each customer in the same terminal in the same day	int	Fiona	

NOTE

It may happen that some customers in the original transaction table have no genuine transaction (because all genuine transactions have been turned to fraudulent by the simulator). All transactions of these customers will be removed from transaction table (because z-score cannot be calculated).