Transaction Fraud Detection

May 5th, 2022

Sumeet Duddagi

Jinglei Liu

Suixin Wang

Wenjing Xia

Yuqi Zhang

EXECUTIVE SUMMARY

Business Problem and the Purpose of the Project

Credit card transaction fraud occurs when the fraudster stoles the card or information of the card to make unauthorized purchases or obtain cash. The most common types of credit card transaction fraud are Card-not-present (CNP) fraud, account takeover fraud, and counterfeit and skimming fraud. Now the CNP fraud dominates due to the evolution of digital transactions. The direct losses caused by credit card transaction fraud are significant. According to the Merchant Risk Council, in 2020, the amount of credit card fraud in the US was around $11 billion. And the actual cost of this $11 billion was estimated to be $39.6 billion using the LexisNexis Fraud Multiplier™. And according to Nilson Report, in 2027, the estimated gross card fraud will be 6.1 cents per $100 in total volume. The good news is that there are several fraud tools to reduce transaction fraud, such as the address verification service, card verification number, and fraud scoring model. This project aims to build fraud models using supervised machine learning algorithms to detect credit card transaction fraud.

Methods of Solving the Problem

The original dataset contains 96,753 rows and 10 columns of credit card transaction records across the year 2006. Records have been labeled to indicate whether the transaction is fraudulent. To build our models, we first developed a Data Quality Report (DQR) to check the data quality and did data cleaning by removing and filling some data. Next, during the feature engineering step, we created 1,310 variables such as days since variables, frequency variables, amount variables, and velocity change variables. Next, we selected the top 20 variables using filter and wrapper combinedly (feature selection). Before experimenting with various non-linear models, we performed baseline modeling (logistic regression). After that, we trained and tuned the models with Single Decision Tree, Random Forest, Light Gradient Boosting Machine, Extreme Gradient Boosting, and Neural Network. We assessed model performance via fraud detection rate (FDR) on training, testing and out-of-time (OOT) data and chose the optimal model with the best performance. We also tried the Cascade model to improve the best model.

Results

Finally, we got our best model, which is Neural Network with 15 variables and hidden layer sizes equal to (20,20). The training, testing and OOT FDR at 3% are 73.6%, 73.2% and 62.6%, respectively. The result indicates that our best model can eliminate 62.6% of the fraud by declining 3% of the transactions. The expected annual saving for credit card transaction fraud using this model is $1.27 million. Please see the Results section for a complete description of our best model's training, testing, and OOT data performance for more detailed information.

TABLE OF CONTENTS

DESCRIPTION OF DATA

The data for this research project are credit card transactions from a government agency located in Tennessee, U.S.A. It covers transactions from Jan. 1, 2006 to Dec. 31, 2006. The data has 10 fields and 96,753 records, with 1,059 of them labeled as fraud. (Full DQR is in Appendix)

1. Summary Statistics Table

Table 1. Summary of Date Fields

Field Name	% Populated	Min	Max	Most Common Value
Date	100.00%	2006-01-01	2006-12-31	2006-02-28

Table 2. Summary of Categorical Fields

Field Name	% Populated	# Unique Values	Most Common Value
Recnum	100.00%	96,753	Unique for each record
Cardnum	100.00%	1,645	5142148452
Merchnum	96.51%	13,091	930090121224
Merch description	100.00%	13,126	GSA-FSS-ADV
Merch state	98.76%	227	TN
Merch zip	95.19%	4,567	38118
Transtype	100.00%	4	P
Fraud	100.00%	2	0

Table 3. Summary of Numerical Fields

Field Name	% Populated	Min	Max	Mean	Stdev	% Zero
Amount	100.00%	0.01	3,102,045.53	427.89	10,006.14	0.00%

2. Distributions for fields

2.1 Date

Figure below shows the number of transactions each month. We noticed the general upward trend through September, followed by a sharp drop in October. The monthly transactions are fewer in the last quarter of the year compared with other quarters. This is due to the government fiscal year which starts on October 1st, and people tend to be more cautious with their money in the first few months of the new fiscal year.

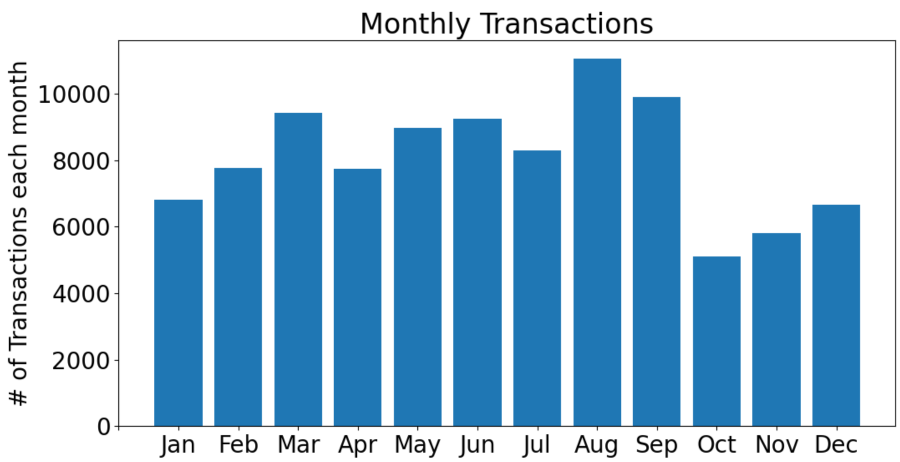

Figure 1. Monthly Transactions

2.2 Amount

This field is a numerical variable showing the amount spent for each transaction. It ranges from 0.01 to 3,102,045.53 with high variability. The maximum value 3,102,045.53 is an outlier. The zoom-in graph below covers 99% of the records in this field and indicates that the data is skewed to the right with a drop after 2,500.

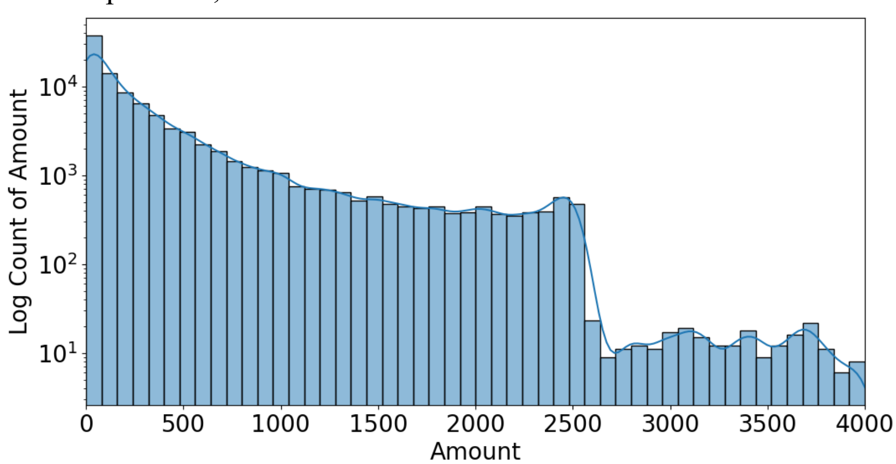

Figure 2. Amount For Each Transaction

2.3 Merch state

This field shows the state that merchants are in. The most frequent state is Tennessee which is about 12.4% of the total transactions. It makes sense because that is where the facility is located.

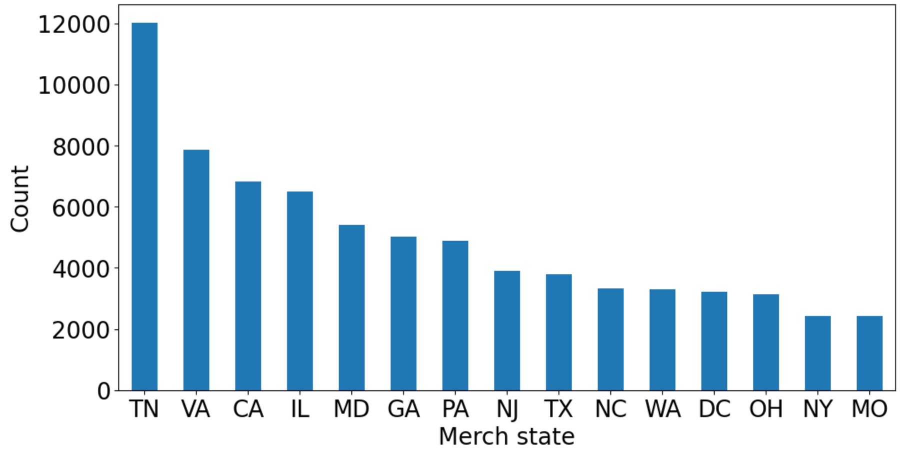

Figure 3. Top 15 Most Frequent States

2.4 Fraud

This field is a categorical variable showing the fraud label. The value of 0 represents non-fraud and 1 represents fraud. There are no missing values and 2 unique values. Most of the records are labeled as non-fraud. 1,059 transactions are labeled as fraud.

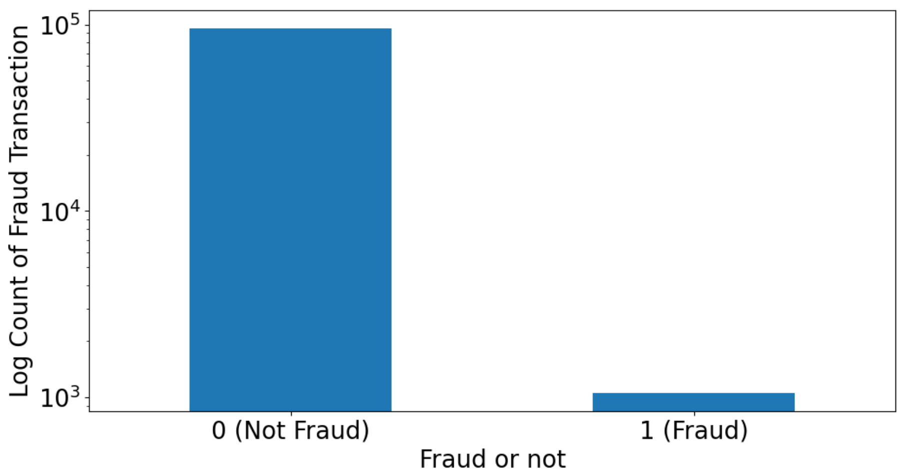

Figure 4. Count of Fraud

DATA CLEANING

1. Identify null values and outliers

- We do not have any missing values in our dataset.
- For the field Amount, only one record exceeds $100,000. So we treat this unusual, not-fraud transaction as an outlier and remove it from our dataset.

2. Identify exclusions

We only want to conduct our further analysis on purchase transactions, so we kept only records with "P" as their Transtype.

3. Fill in missing values

- For Merchnum column, 3,606 records do not have Merchnum values. We fill in missing values with Merchnum that has the same Merch description. Because Merch description is closely related to Merchnum. Then we assign "unknown" for the rest null values.
- For Merch state column, 1,195 records do not have Merch state values. We first fill in missing values with Merch state that has the same Merch zip. Because each Merch zip corresponds to one Merch state. So after further adding some known pairs of zip code and state, filling in with Merch zip will significantly reduce null values. We then fill Merch state by mapping with Merchnum and Merch description. Last, we assign "unknown" for the rest null values.
- For Merch zip column, 4,656 records do not have Merch zip values. We fill in missing values by mapping with Merchnum and Merch description. Then we assign "unknown" for the rest null values.

CANDIDATE VARIABLES

After cleaning the data, we ended up with 96,397 records with no missing values. Starting with 7 basic transaction information, we created 1,310 candidate variables via the following steps:

1. Add secondary information as fields

- dow: the day of week of the transaction
- zip3: the first 3 digits of the zip code
- month: the month of the transaction
- merchdesc_short: a shortened version of the merchant description, delete # and the number in the Merch description field (ex. OFFICE DEPOT #191 –> OFFICE DEPOT)

The entities included 2 fields from the original dataset, card number and merchant number, and combinations of the fields.

The full list of 15 entities: [Cardnum, Merchnum, card_merch, card_zip, card_state, merch_zip, merch_state, card_dow, merch_dow, card_merchdesc, card_ zip3, card_merch_month, card_state_month, card_zip_month, card_merchdesc_short].

2. Variables Created

Then we created variables based on the above entities. The formulas and logic for creating the variables are listed below. For the full list of all variables, please refer to the appendix.

2.1 Risk table variables

The likelihood of fraud for that day of the week and for that merchant state. The two variables, dow_risk and state_risk, show that there is a high risk of fraud on weekends and in Indiana.

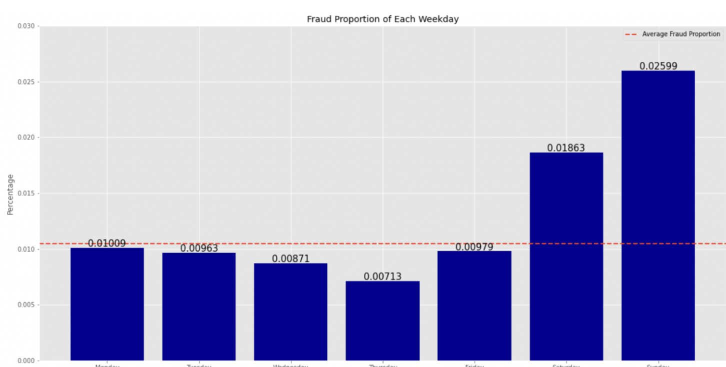

Figure 5. Fraud Proportion of Each Weekday

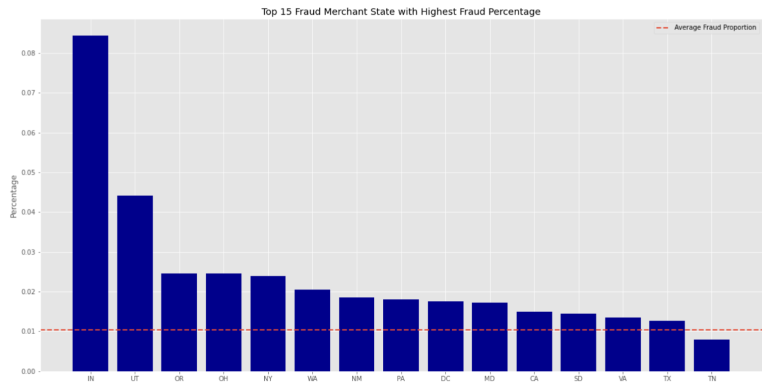

Figure 6. Top 15 Merchant State with the Highest Fraud Percentage

2.2 Days since variables

The current date minus the date of the most recent transaction with the same entities. If the entities were never seen before, the default value was 365. These variables could see if the card was used at the merchant or location not used before, a signal of potential fraud.

Examples of variables: Cardnum_day_since, Merchnum_day_since, card_merch_day_since, card_zip_day_since, card_state_day_since, etc.

2.3 Frequency variables

Count the number of transactions with the same entities over the past 0, 1, 3, 7, 14, 30, 90 days. The mode of fraud that the variables relate to is the burst of activity using the card at the same merchant or geography.

Examples of variables: Cardnum_count_0, Cardnum_count_1, Cardnum_count_3, Cardnum_count_7, Cardnum_count_14, Cardnum_count_30, etc.

2.4 Amount variables

The average, maximum, median, total, actual divided by average, actual divided by maximum, actual divided by median, actual divided by total amount and standard deviation at the same entities over the past 0, 1, 3, 7, 14, 30, 90 days. These variables would catch the larger than normal purchase amount on the card or at the merchant.

Examples of variables: Cardnum_avg_0, Cardnum_max_0, Cardnum_med_0, Cardnum_total_0, Cardnum_actual/avg_0, Cardnum_actual/max_0, etc.

2.5 Velocity change variables

$$\frac{\text{number of transactions with the same entities over the past 0, 1 day}}{\text{average daily number of transactions with the same entities over the past 7, 14, 30, 90 days}}$$

$$\frac{\text{amount of transactions with the same entities over the past 0, 1 day}}{\text{average daily amount of transactions with the same entities over the past 7, 14, 30, 90 days}}$$

The mode of fraud that the variables relate to is the burst of activity using the card and abnormal perchance amount at the same merchant or geography.

Examples of variables: Cardnum_count_0_by_7, Cardnum_total_0_by_7, Cardnum_count_0_by_14, Cardnum_total_0_by_14, Cardnum_count_0_by_30, Cardnum_total_0_by_30, etc.

2.6 Duplicate transaction variable

Count the number of transactions at each merchant with the same amount. The variable, Merchant_with_same_amount_count, could catch recurring charges.

2.7 Benford's Law variables

Use Benford's Law to rank the card and merchant by unusualness. After removing all the transactions from FedEx, the transactions are grouped by card and merchant. For each group, we looked at the distribution of the first digit of the purchase amount and measure the unusualness by using a smoothing formula: $U^* = 1 + (\frac{U-1}{1+e^{-t}})$ and $t = (n - n_{mid})/c$ where $c = 3$ and $n_{mid} = 15$.

Benford's_Law_cardnum_U* and Benford's_Law_merchnum_U* variables were created based on the assumption that a cardholder or a merchant is "making up" fraudulent transactions.

FEATURE SELECTION PROCESS

The total number of candidate variables is 1,310. The model would be complex and the machines' computing power would be burned out if including all the variables. Feature selection is necessary for reducing substantial computational costs and overfitting. It chooses the features that contribute most to the prediction variables and eliminates irrelevant features which may decrease the accuracy of the models. Fewer features also make the model more interpretable.

There are several feature selection methods, such as Exhaustive Search, Filter, Wrapper and Embedded methods. The Embedded method was used for this application fraud project by combining Filter and Wrapper methods. Below are the detailed steps:

1. Filter

- No matter what models to choose, if the variables fail to differentiate the distribution of the events, then there are still wrong predictions. So, the Filter method is needed for finding the most relevant variables. There are many filter methods, such as Chi-square Test, Fisher Score and Kolmogorov–Smirnov (KS) test. For this binary classification problem, Univariate KS was picked to filter the variables. For each candidate variable, it performed the two-sample KS test for goodness of fit, one sample for non-frauds (good) and one sample for frauds (bad). It compared the distributions and calculated univariate KS which measured how separate these two curves are. The better the variable for separating, the more important the variable and thus higher the univariate KS

- Before the actual filtering process, the data of the last two months and the first two weeks were removed. The first one was used as out-of-time data and the other was removed because the data in the first two months are not sufficient for measuring the relevancy of the variables

- A random variable and actual fraud label were added to make sure the accuracy of the method. Based on the KS score, the number of variables was limited to 100. The limitation of this step is that it did not eliminate multicollinearity. Then the Wrapper method was used to reduce the multicollinearity problem and continue to simplify the model down to 20 variables

2. Wrapper

- There are three Wrapper methods: Forward Selection, Backward Selection and General Stepwise Selection. For this binary classification problem, Forward Selection was picked to see the order of multivariate importance. A one-variable Random Forest model was started and then the variable which yielded the highest Fraud Detection Rate (FDR) was kept. Then a second variable was added to find the best combination with the highest FDR. The step was repeated 20 times to get the final 20 variables

3. Top 20 Variables with Their Descriptions and Univariate KS

Table 4. Top 20 Variables

Order	Variable Name	Description	KS Score
1	card_zip_total_3	The total amount at that card in that zip code over the past 3 days	0.678
2	card_state_max_30	The maximum amount at that card in that state over the past 30 days	0.598
3	card_zip_total_14	The total amount at that card in that zip code over the past 14 days	0.672
4	card_merch_total_0	The total amount at that card at that merchant over the past 0 day	0.611
5	card_state_total_0	The total amount at that card in that state over the past 0 day	0.612
6	card_zip_total_0	The total amount at that card in that zip code over the past 0 day	0.610
7	merch_zip_avg_0	The average amount at that merchant in that zip code over the past 0 day	0.580
8	merch_dow_avg_3	The average amount at that merchant on that day of week over the past 3 days	0.580
9	Merchnum_avg_0	The average amount at that merchant over the past 0 day	0.580
10	merch_dow_avg_1	The average amount at that merchant on that day of week over the past 1 day	0.580
11	merch_dow_avg_0	The average amount at that merchant on that day of week over the past 0 day	0.580
12	merch_zip_avg_1	The average amount at that merchant in that zip code over the past 1 day	0.575
13	merch_state_avg_0	The average amount at that merchant in that state over the past 0 day	0.582
14	Merchnum_avg_1	The average amount at that merchant over the past 1 day	0.573
15	card_merch_avg_30	The average amount at that card in that merchant over the past 30 days	0.593
16	card_merch_max_30	The maximum amount at that card in that merchant over the past 30 days	0.650
17	card_zip_avg_1	The average amount at that card in that zip code over the past 1 day	0.582
18	card_merch_avg_3	The average amount at that card in that merchant over the past 3 days	0.590
19	merch_state_avg_1	The average amount at that merchant in that state over the past 1 day	0.576
20	card_zip_avg_0	The average amount at that card in that zip code over the past 0 day	0.573

CASE STUDY OF TIME DEPENDENCE

To demonstrate how transaction counts affect the fraud score, we did a case study using a specific card number and a merchant number. It takes time for Fraud Scores to discover that the transaction flow is becoming unusual. For example, for the card number 5142160778, we can see that there was only one transaction on Dec. 18, so the fraud score was lower, whereas on Dec. 19, there were four transactions, so the fraud score increased slightly, and on Dec. 21, there were a total of six transactions, which is why the fraud score increased dramatically, and we can see a similar pattern for merchant number 4353000719908 based on transaction counts on the right picture as well.

Card Number = 514216077

Merchant number = 4353000719908

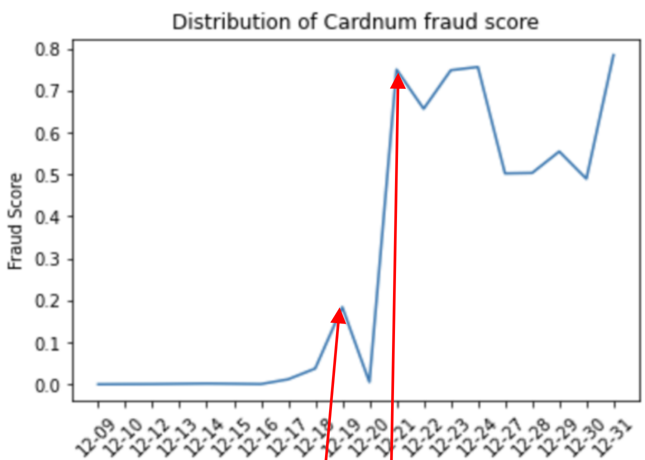

Figure 7. Date vs Fraud Scores (Cardnum)

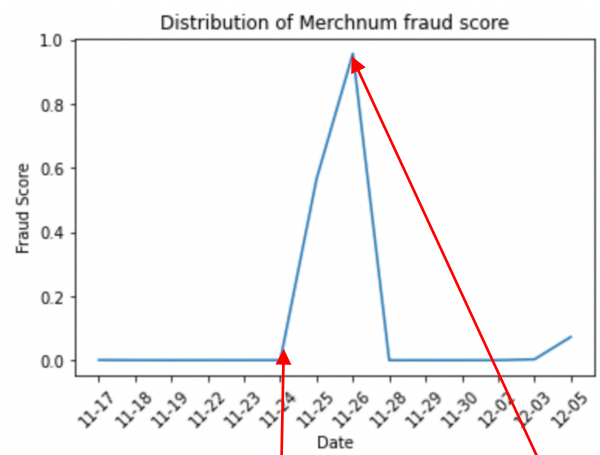

Figure 8. Date vs Fraud scores (Merchnum)

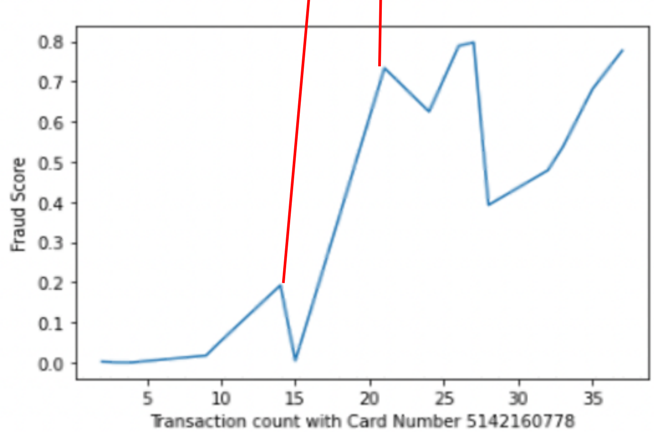

Figure 9. Transaction counts vs Fraud Scores (Cardnum)

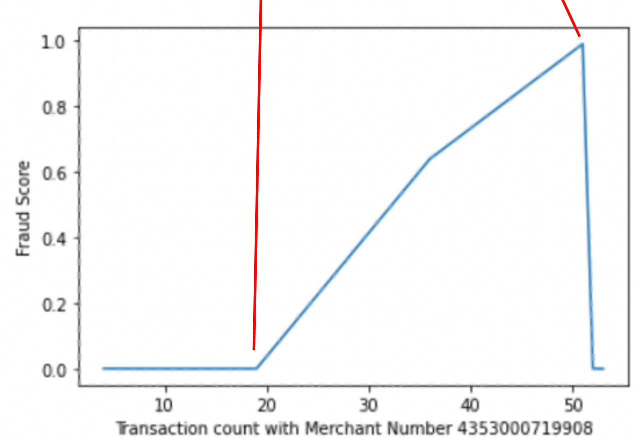

Figure 10. Transaction counts vs Fraud Scores (Merchnum)

Dec 18: 1 transaction
Dec 19: 4 transactions
Dec 21: 6 transactions

Nov 24: 1 transaction
Nov 25: 17 transactions
Nov 26: 15 transactions
Nov 28: 1 transaction

MODEL ALGORITHMS

1. Splitting of Data

Table 5. Details of splitting

Training	Testing	OOT
61.2%	26.2%	12.6%
70% of the first 10 months	30 % of the first 10 months	Last 2 months

Furthermore, we will go over the findings and how we went about the modeling approach.

To begin, we separated the data into three categories: 61 percent training, 26 percent testing, and 12 percent Out of time (OOT) data. Specifically, we used the last two months as OOT, and the first ten months of data were split into 70 percent training and 30 percent testing.

For modelling process, we began by experimenting with a linear model such as logistic regression. Later, we went on to use other complex models such as Decision Trees, Random forest, Light Gradient Boosting Machine, Extreme Gradient Boosting, Neural Network and Cascade model (Neural Network). Finally, we tuned hyperparameters until we identified the simplest model that could give us the best Fraud Detection Rate while rejecting as few transactions as possible. For thorough evaluation of the model outcomes on train, testing and OOT data, the Fraud Detection Rate was taken as a mean of 10 model results.

We will go through each of the models we used and the performance we could achieve in the next section, model tuning.

2. Model Tuning

2.1 Logistic Regression

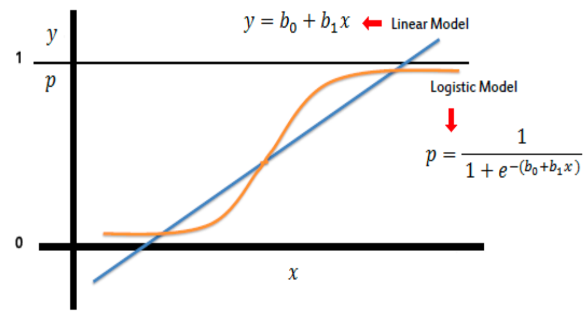

Figure 11. Logistic Regression

Logistic Regression is a process of modeling the probability of a discrete outcome. To map predicted values to probabilities, it uses the sigmoid function. The parameters we tuned are:

- penalty: specify the norm of the penalty
- C: inverse of regularization strength; must be a positive float
- solver: algorithm to use in the optimization problem
- l1-ratio: the Elastic-Net mixing parameter, with 0 <= l1_ratio <= 1

We used a range of hyperparameters as demonstrated below, but the Logistic regression model did not perform well since the best mean OOT performance was just 36.9% in the table below.

Therefore, we went on to use try other models, which could give us better results.

Table 6. Logistic Regression Result

Iteration	# Variables	penalty	C	solver	l1_ratio	Train	Test	OOT
1 (default)	20	l2	1	lbfgs	None	60.3	58.7	35.7
2	20	l2	1	liblinear	None	59.4	60.2	35.6
3	20	l2	0.1	liblinear	None	60.0	61.1	35.3
4	20	elasticnet	1	saga	0.2	60.1	58.5	35.6
5	20	elasticnet	1	saga	0.8	59.7	60.3	35.3
6	20	elasticnet	0.1	saga	0.2	59.8	59.7	35.6
7	20	l1	0.1	saga	None	60.2	59.1	34.6
8	15	l2	0.1	liblinear	None	60.5	60.4	34.8
9	15	l2	1	liblinear	None	59.5	59.6	31.0
10	10	l2	1	liblinear	None	63.7	64.3	36.9
11	5	l2	1	liblinear	None	64.8	63.0	36.7

2.2 Single Decision Tree

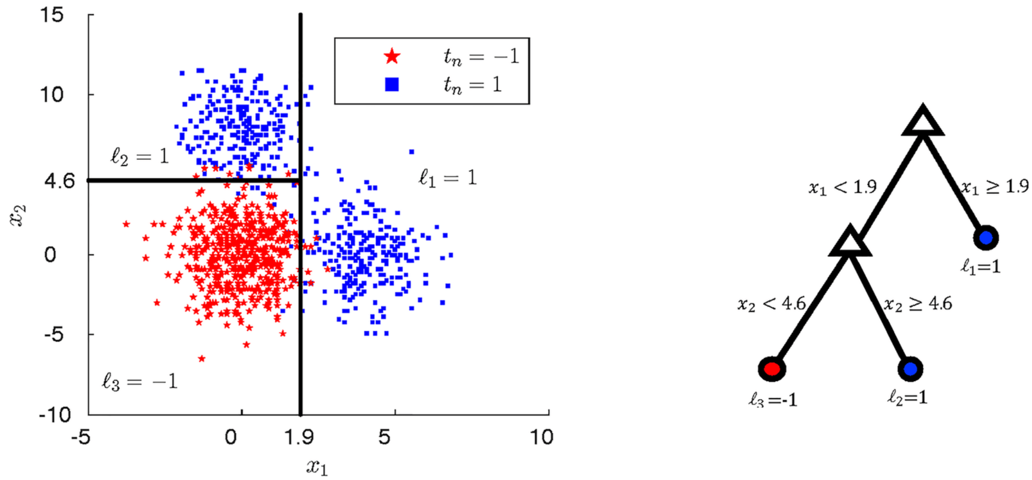

Figure 12. Decision Tree

A decision tree is a type of supervised machine learning used to categorize or make predictions based on how a previous set of questions were answered. The decision tree is fragile and unstable as with slightly different data, the model will choose different cut points that give different-looking trees. Also, they are prone to overfitting as they look at the training data and divide it into boxes which is easy to isolate regions unique to that data set.

The parameters we tuned are:

- criterion: the function to measure the quality of a split, supported criteria are “gini” for the Gini impurity and “entropy” for the information gain
- splitter: the strategy used to choose the split at each node, supported strategies are “best” to choose the best split and “random” to choose the best random split
- max_depth: the maximum depth of the tree. If None, then nodes are expanded until all leaves are pure or until all leaves contain less than min_samples_split samples
- min_samples_split: the minimum number of samples required to split an internal node
- min_samples_leaf: the minimum number of samples required to be at a leaf node. A split point at any depth will only be considered if it leaves at least min_samples_leaf training samples in each of the left and right branches

Table 7. Single Decision Tree Result

Iteration	# Variables	criterion	splitter	max_depth	min_samples_split	min_samples_leaf	Train	Test	OOT
1 (default)	20	gini	best	None	2	1	100	57.9	26.5
2	20	gini	random	None	2	1	100	53.8	27.9
3	20	gini	random	20	20	20	80.9	71.0	40.7
4	20	gini	best	20	20	20	92.2	74.6	35.6
5	20	entropy	random	20	20	20	81.2	71.9	36.1
6	20	entropy	best	20	20	20	94.0	74.4	36.5
7	20	entropy	best	20	200	60	80.7	71.8	50.3
8	15	entropy	best	20	200	60	81.1	73.2	51.6
9	10	entropy	best	20	200	60	79.2	73.4	53.1
10	5	entropy	best	20	200	60	79.7	73.4	52.6

2.3 Random Forest

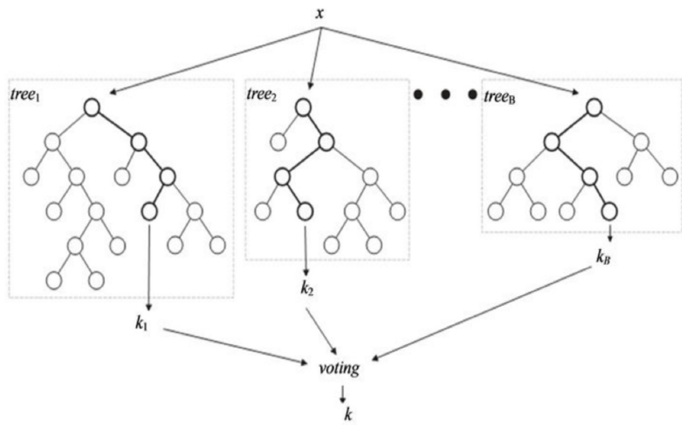

Figure 13. Random Forest

Random forest is a collection of many strong trees, and each tree has some randomness associated with it. Each individual tree in the random forest spits out a class prediction and the class with the most votes becomes our model's prediction. The fundamental concept behind random forest is "the wisdom of crowds". A large number of relatively uncorrelated models (trees) operating as a committee will outperform any of the individual constituent models. With a random forest model, our chances of making correct predictions increase with the number of uncorrelated trees in our model.

The parameters we tuned are:

- n_estimators: the number of trees in the forest
- criterion: The function to measure the quality of a split. Supported criteria are "gini" for the Gini impurity and "entropy" for the information gain
- max_features: the number of features to consider when looking for the best split. If "auto", then max_features=sqrt(n_features)
- max_depth: The maximum depth of the tree. If None, then nodes are expanded until all leaves are pure or until all leaves contain less than min_samples_split samples
- min_samples_split: the minimum number of samples required to split an internal node.
- min_samples_leaf: the minimum number of samples required to be at a leaf node. A split point at any depth will only be considered if it leaves at least min_samples_leaf training samples in each of the left and right branches

Random Forest performed better than Decision Tree with the hyperparameters shown below.
Best mean OOT performance is 56.8%.

Table 8. Random Forest Result

Iteration	# Variables	n_estimators	criterion	max_feature	max_depth	min_samples_split	min_samples_leaf	Train	Test	OOT
1	20	10	gini	'auto'	5	2	1	85.8	75.3	45.3
2	20	100	gini	'auto'	None	2	1	100.0	78.0	47.7
3	20	50	gini	None	20	100	30	86.8	78.9	55.9
4	20	50	entropy	None	20	100	30	88.7	79.6	54.6
5	20	50	gini	None	None	100	30	86.8	78.8	56
6	20	50	gini	None	None	300	50	79.9	76.6	56.2
7	20	50	gini	None	None	400	50	79.2	74.5	55.9
8	20	200	entropy	'auto'	20	2	1	100.0	78.6	50.8
9	20	200	gini	'auto'	20	2	1	100.0	78.6	51.1
10	20	200	gini	'auto'	20	10	2	100.0	80.8	53.4
11	20	200	gini	log2'	20	10	2	100.0	79.0	51.0
12	20	200	gini	None	20	10	2	100.0	81.0	51.6
13	20	100	gini	None	30	100	30	78.5	75.2	56.8
14	15	100	gini	None	30	100	30	86.6	79.2	55.9
15	15	100	gini	None	None	100	30	85.8	78.3	56.5
16	15	100	gini	None	None	100	50	83.9	77.1	56.0
17	15	100	gini	None	None	100	100	78.5	73.9	55.7
18	10	100	gini	None	None	100	30	77.2	73.8	56.1
19	5	100	gini	None	None	100	100	77.4	73.5	56.2

2.4 Light Gradient Boosting Machine

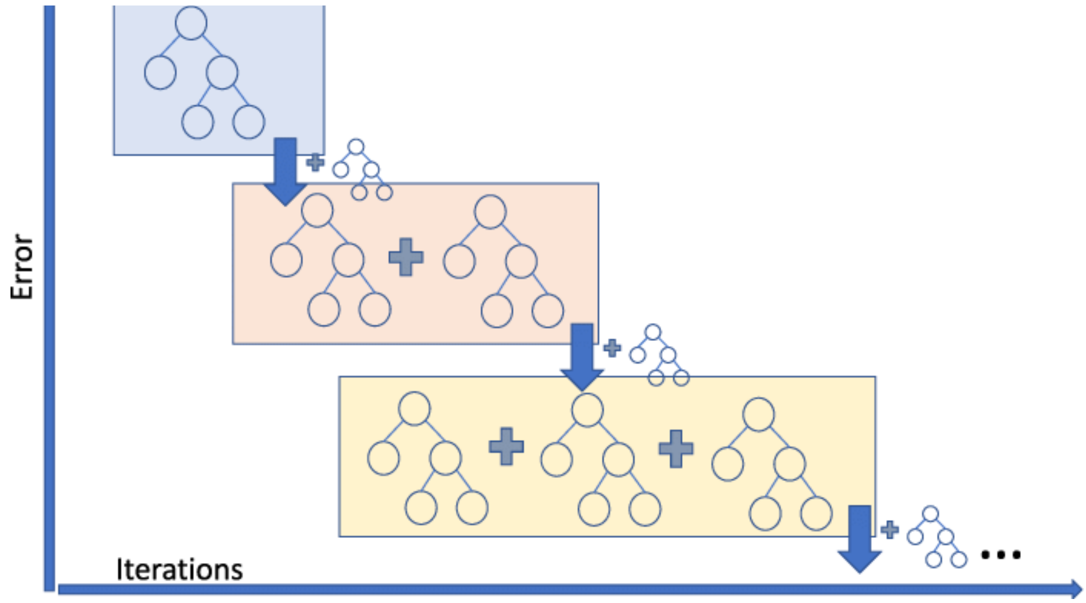

Figure 14. Light Gradient Boosting Machine

We tried two types of boosted trees (LightGBM and XGBoost). LightGBM grows trees vertically while XGBoost grows trees horizontally, meaning that LightGBM grows trees leaf-wise while XGBoost grows level-wise. It will choose the leaf with max delta loss to grow.

The parameters we tuned are:

- num_leaves: maximum number of leaves of a tree
- max_depth: maximum number of nodes of a tree
- learning rate: how fast the model learns. The lower the learning rate, the more precise the model
- n_estimators: number of boosted trees to fit
- min_child_samples: minimum number of data needed in a child (leaf)
- subsample: subsample ratio of the training instance

LightGBM performed similar to Random Forest with the hyperparameters shown below. Best mean OOT performance on average is 56.1%.

Table 9. Light Gradient Boosting Machine Result

Iteration	# Variables	num_leaves	max_depth	learning_rate	n_estimators	min_child_samples	subsample	Train	Test	OOT
1 (default)	20	31	-1	0.1	100	20	1	99.4	80.3	41.7
2	20	31	3	0.1	100	20	1	86.5	78.7	50.5
3	20	31	5	0.1	100	20	1	96.4	81.3	41.6
4	20	50	3	0.1	50	20	1	91.8	79.3	46.7
5	20	100	5	0.01	100	20	1	81.2	77.4	51.2
6	20	50	5	0.01	100	20	1	81.7	76.9	53.2
7	20	50	4	0.01	50	20	1	75.5	73.1	55.2
8	20	50	4	0.01	50	20	0.7	75.8	72.9	54.0
9	20	50	4	0.01	50	30	1	75.3	74.7	55.7
10	20	50	5	0.01	1000	60	1	89.5	81.6	51.7
11	15	50	4	0.01	50	20	1	75.2	73.8	55.3
12	15	50	5	0.01	50	20	1	77.8	75	54.3
13	15	50	4	0.01	100	20	1	77.1	75.7	55.0
14	15	100	4	0.01	100	20	1	77.7	75.9	56.1
15	10	100	4	0.01	100	20	1	76.4	74.4	55.0
16	5	100	4	0.01	100	20	1	77.6	73.7	55.4

2.5 Extreme Gradient Boosting

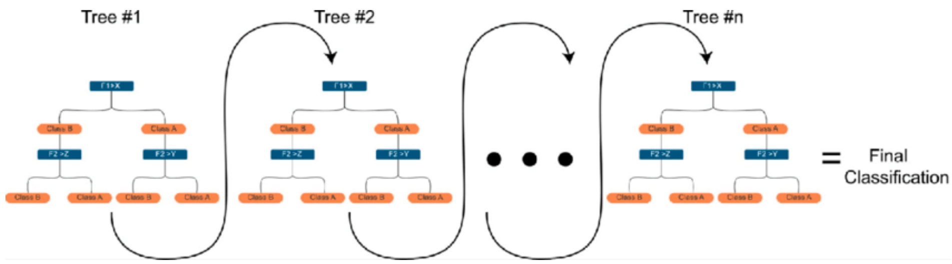

Figure 15. Extreme Gradient Boosting

XGBoost is a sequential ensemble learning technique, similar to LightGBM, in which the model's performance improves over iterations, with the exception that the splits are level-wise.

The parameters we tuned are:

- n_estimators: number of boosting rounds
- max_depth: maximum tree depth for base learners.
- max_leaves: aximum number of leaves; 0 indicates no limit
- learning_rate: boosting learning rate
- min_child_weight: minimum sum of instance weight(hessian) needed in a child

XGBoost performed similar to Random forest and LightGBM with the hyperparameters shown below. Best mean OOT performance is 56.3%.

Table 10. Extreme Gradient Boosting Result

Iteration	# Variables	n_estimators	max_depth	max_leaves	learning_rate	min_child_weight	Train	Test	OOT
1	20	50	3	0	0.01	5	66.3	65.2	42.2
2	20	100	3	0	0.01	5	70.9	66.7	48.2
3	20	100	5	0	0.01	5	74.3	72.7	55.0
4	20	100	5	0	0.2	5	93.0	82.4	41.1
5	20	300	5	0	0.01	5	77.9	74.4	55.8
6	20	500	5	0	0.01	5	81.8	76.4	55.4
7	15	300	5	0	0.01	5	77.7	75.6	52.8
8	15	300	5	0	0.01	7	77.0	74.2	55.6
9	15	500	5	0	0.01	3	83.5	78.3	55.5
10	15	500	5	0	0.01	5	81.9	78.1	56.3
11	10	300	5	0	0.01	7	76.8	73.8	56.1
12	10	500	5	0	0.01	7	80.5	77.1	54.7
13	5	300	5	0	0.01	7	76.6	75.5	54.2

2.6 Neural Network

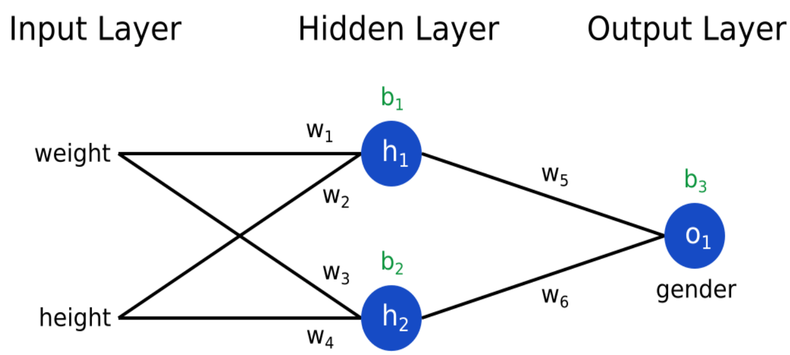

Figure 16. Neural Network

Finally, we tried Neural Network as the final algorithm. Neural Networks are the functional unit of Deep Learning and are known to mimic the behavior of the human brain. The input data is processed through different layers of artificial neurons stacked together to produce the desired output. We used Multi-layer Perceptron classifier for this purpose.

The parameters we tuned are:

- hidden_layer_sizes: the i^{th} element represents the number of neurons in the i^{th} hidden layer
- activation: activation function for the hidden layer
- solver: the solver for weight optimization
- alpha: L2 penalty (regularization term) parameter
- learning_rate: learning rate schedule for weight updates. Only used when solver = 'sgd'
- learning_rate_init: the initial learning rate used. It controls the step-size in updating the weights, only used when solver = 'sgd' or 'adam'
- max_iter: maximum number of iterations
- shuffle: whether to shuffle samples in each iteration. Only used when solver = 'sgd' or 'adam'

Neural Network gave us the highest performance out of all the models tried with the hyperparameters shown in the table below. Best mean OOT performance is 59.5%.

Table 11. Neural Network results

Iteration	# Variables	hidden_layer_sizes	activation	solver	alpha	learning_rate	learning_rate_init	max_iter	shuffle	Train	Test	OOT
1	20	-1	relu	adam	0.0001	constant	0.001	200	TRUE	53.9	53.8	31.7
2	20	-3	relu	adam	0.0001	constant	0.001	200	TRUE	62.4	64.9	43.4
3	20	-5	relu	adam	0.0001	constant	0.001	200	TRUE	65.1	65.1	47.2
4	20	(10, 10)	relu	adam	0.0001	constant	0.001	200	TRUE	74.2	71.5	52.8
5	20	(20, 20)	relu	adam	0.0001	constant	0.001	200	TRUE	77.8	74.4	52.2
6	20	(30, 30)	relu	adam	0.0001	constant	0.001	200	TRUE	80.8	74.5	47.6
7	15	(20, 20)	relu	adam	0.0001	constant	0.001	200	TRUE	75.8	74.0	54.0
8	15	(20, 20)	identity	adam	0.0001	constant	0.001	200	TRUE	60.8	61.3	39.9
9	15	(20, 20)	logistic	adam	0.0001	constant	0.001	200	TRUE	69.4	69.8	53.0
10	15	(20, 20)	tanh	adam	0.0001	constant	0.001	200	TRUE	75.8	73.0	58.2
11	15	(20, 20)	tanh	sgd	0.0001	adaptive	0.001	200	TRUE	64.1	64.2	36.4
12	15	(20, 20)	tanh	sgd	0.0001	invscaling	0.001	200	TRUE	44.5	43.9	22.9
13	15	(20, 20)	tanh	lbfgs	0.0001	constant	0.001	1000	TRUE	88.0	77.0	23.5
14	15	(20, 20)	tanh	adam	0.00001	constant	0.001	200	TRUE	75.6	72.9	59.5
15	15	(20, 20)	tanh	adam	0.0000001	constant	0.001	200	TRUE	75.0	72.9	57.8
16	10	(20, 20)	tanh	adam	0.0001	constant	0.001	200	TRUE	71.3	69.8	50.7
17	10	(30, 30)	tanh	adam	0.0001	constant	0.001	200	TRUE	70.3	71.5	50.8
18	5	(20, 20)	tanh	adam	0.0001	constant	0.001	200	TRUE	70.4	70.3	50.5

2.6 Cascade Model

Further, we performed cascading on our final model, which is Neural Network to check whether we can improve OOT performance. The predicted scores came from the Neural Network model that achieved the highest OOT performance. The table below shows the summary statistics of the scores in training data.

Table 12. Summary Statistics of Predicted Score

	Predicted Score
count	59,010
mean	0.010879
std	0.068620
min	0.000004
25%	0.000218
50%	0.000371
75%	0.004450
max	0.999919

Firstly, we filtered out extreme scores in training data i.e., predicted score >= 0.9. This could let the model focus on records that were hard to figure out. We could obtain OOT FDR between 50% to 60% as shown in the table below.

Table 13. Cascade Model Results 1

Iteration	# Variables	hidden_ layer_sizes	activation	solver	alpha	Learning _rate	Learning _rate_init	max_iter	shuffle	Train	Test	OOT
1	15	(20, 20)	tanh	adam	0.00001	constant	0.001	200	TRUE	68.5	70.8	57.3
2	15	(20, 20)	tanh	adam	0.0001	constant	0.001	200	TRUE	68.1	70.1	58.4
3	15	(30, 30)	tanh	adam	0.0001	constant	0.001	200	TRUE	72.3	70.5	50.8

Similarly, we filtered out intermediate scores i.e., predicted scores between 0.0003 and 0.004. The model focused on easier ones this time. The FDR in OOT declined as shown in the table below.

Table 14. Cascade Model results 2

Iteration	# Variables	hidden_ layer_sizes	activation	solver	alpha	Learning _rate	learning_ rate_init	max_iter	shuffle	Train	Test	OOT
1	15	(20, 20)	tanh	adam	0.00001	constant	0.001	200	TRUE	12.9	43.6	29.3
2	15	(20, 20)	tanh	adam	0.0001	constant	0.001	200	TRUE	13.2	43.0	30.7

As a result, cascading did not improve our model's performance, and we ended up using the Neural Network as our final model.

3. Final Model Description

Finally, after experimenting with various models and modifying hyperparameters, we found that Neural Network provided the greatest model performance.

Best hyperparameters of Neural Network model is as shown below.

Table 15. Best model hyperparameters

# Variables	hidden_ layer_sizes	activation	solver	alpha	learning_rate	learning_rate_init	max_iter	shuffle
15	(20, 20)	tanh	adam	0.00001	constant	0.001	200	TRUE

Through these hyperparameters, we could obtain the Fraud Detection rate at a 3% decline rate as shown below.

Table 16. Best model FDR at 3% of transaction

Training FDR	Testing FDR	OOT FDR
75.6%	72.9%	59.5%

The Neural Network model offered us 59.5 percent FDR at 3% on OOT without overfitting the data, as shown in the table above. As a result, we chose Neural Network as our final model with these hyperparameters.

To further understand the effectiveness of our neural network model, we created an FDR table across population bins to generate individual and cumulative statistics for goods and bads, as well as their accompanying Fraud detection rate, KS, and false-positive rate, as displayed in the following pages. This table was constructed for training, testing, and OOT data.

4. FDR across Population Bins

In this part, we would like to see what percent of fraud can be detected when declining different percentage transaction over train, test and OOT dataset. We constructed the below tables in the following steps:

- Binned transaction records via predicted probabilities of fraud decreasingly (i.e., 1st bin has the highest probability)
- Calculated the number and the percentage of records, non-fraud transaction, fraud transaction respectively
- Calculated the cumulative number of records, non-fraud transaction, fraud transaction
- Got percentage of cumulative good and of cumulative bad (FDR at n% of transactions). We used FDR as measure of performance
- Calculated KS (= % of bad - % of good) and FPR (= cumulative good / cumulative bad)

4.1 Training

The table below showcases the results our final model predicted for training data.

Table 17. Training FDR table across population bins

Training	# Records			# Goods			# Bads			Fraud Rate		
	59010			58384			626			0.0106		
	Bin Statistics					Cumulative statistics						
Population bin %	**# Records**	**# Goods**	**# Bads**	**% Goods**	**% Bads**	**Total # of records**	**Cumulative Goods**	**Cumulative Bads**	**% Cumulative Goods**	**% Bads (FDR)**	**KS**	**FPR**
1	590	237	353	40.2	59.8	590	237	353	0.4	56.4	56.0	0.7
2	590	507	83	85.9	14.1	1180	744	436	1.3	69.6	68.4	1.7
3	590	565	25	95.8	4.2	1770	1309	461	2.2	73.6	71.4	2.8
4	590	566	24	95.9	4.1	2360	1875	485	3.2	77.5	74.3	3.9
5	590	577	13	97.8	2.2	2950	2452	498	4.2	79.6	75.4	4.9
6	591	575	16	97.3	2.7	3541	3027	514	5.2	82.1	76.9	5.9
7	590	576	14	97.6	2.4	4131	3603	528	6.2	84.3	78.2	6.8
8	590	588	2	99.7	0.3	4721	4191	530	7.2	84.7	77.5	7.9
9	590	585	5	99.2	0.8	5311	4776	535	8.2	85.5	77.3	8.9
10	590	583	7	98.8	1.2	5901	5359	542	9.2	86.6	77.4	9.9
11	590	588	2	99.7	0.3	6491	5947	544	10.2	86.9	76.7	10.9
12	590	588	2	99.7	0.3	7081	6535	546	11.2	87.2	76.0	12.0
13	590	585	5	99.2	0.8	7671	7120	551	12.2	88.0	75.8	12.9
14	590	587	3	99.5	0.5	8261	7707	554	13.2	88.5	75.3	13.9
15	591	589	2	99.7	0.3	8852	8296	556	14.2	88.8	74.6	14.9

4.2 Testing

The table below showcases the results our final model predicted for testing data.

Table 18. Test FDR table across population bins

Testing	# Records			# Goods			# Bads			Fraud Rate		
	25290			25036			254			0.010		
	Bin Statistics					Cumulative statistics						
Population bin(%)	# Records	# Goods	# Bads	% Goods	% Bads	Total # of records	Cumulative Goods	Cumulative Bads	% Cumulative Goods	% Bads (FDR)	KS	FPR
1	253	108	145	42.7	57.3	253	108	145	0.4	57.1	56.7	0.7
2	253	224	29	88.5	11.5	506	332	174	1.3	68.5	67.2	1.9
3	253	241	12	95.3	4.7	759	573	186	2.3	73.2	70.9	3.1
4	253	245	8	96.8	3.2	1012	818	194	3.3	76.4	73.1	4.2
5	252	249	3	98.8	1.2	1264	1067	197	4.3	77.6	73.3	5.4
6	253	249	4	98.4	1.6	1517	1316	201	5.3	79.1	73.9	6.5
7	253	250	3	98.8	1.2	1770	1566	204	6.3	80.3	74.1	7.7
8	253	251	2	99.2	0.8	2023	1817	206	7.3	81.1	73.8	8.8
9	253	251	2	99.2	0.8	2276	2068	208	8.3	81.9	73.6	9.9
10	253	252	1	99.6	0.4	2529	2320	209	9.3	82.3	73.0	11.1
11	253	253	0	100.0	0.0	2782	2573	209	10.3	82.3	72.0	12.3
12	253	253	0	100.0	0.0	3035	2826	209	11.3	82.3	71.0	13.5
13	253	253	0	100.0	0.0	3288	3079	209	12.3	82.3	70.0	14.7
14	253	252	1	99.6	0.4	3541	3331	210	13.3	82.7	69.4	15.9
15	253	253	0	100.0	0.0	3794	3584	210	14.3	82.7	68.4	17.1

4.3 OOT

The table below showcases the results our final model predicted for OOT data.

Table 19. OOT FDR table across population bins

OOT	# Records		# Goods			# Bads			Fraud Rate		
	12097		11918			179			0.015		
	Bin Statistics				Cumulative statistics						
Population bin(%)	# Records	# Goods	# Bads	% Goods	% Bads	Total # of records	Cumulative Goods	Cumulative Bads	% Cumulative Goods	% Bads (FDR)	KS	FPR
1	121	51	70	42.1	57.9	121	51	70	0.4	39.1	38.7	0.7
2	121	90	31	74.4	25.6	242	141	101	1.2	56.4	55.2	1.4
3	121	110	11	90.9	9.1	363	251	112	2.1	62.6	60.5	2.2
4	121	119	2	98.3	1.7	484	370	114	3.1	63.7	60.6	3.2
5	121	120	1	99.2	0.8	605	490	115	4.1	64.2	60.1	4.3
6	121	120	1	99.2	0.8	726	610	116	5.1	64.8	59.7	5.3
7	121	120	1	99.2	0.8	847	730	117	6.1	65.4	59.2	6.2
8	121	119	2	98.3	1.7	968	849	119	7.1	66.5	59.4	7.1
9	121	118	3	97.5	2.5	1089	967	122	8.1	68.2	60.0	7.9
10	121	120	1	99.2	0.8	1210	1087	123	9.1	68.7	59.6	8.8
11	121	119	2	98.3	1.7	1331	1206	125	10.1	69.8	59.7	9.6
12	121	120	1	99.2	0.8	1452	1326	126	11.1	70.4	59.3	10.5
13	121	119	2	98.3	1.7	1573	1445	128	12.1	71.5	59.4	11.3
14	121	118	3	97.5	2.5	1694	1563	131	13.1	73.2	60.1	11.9
15	121	119	2	98.3	1.7	1815	1682	133	14.1	74.3	60.2	12.6

FINANCIAL CUTOFF AND IMPLICATIONS

To showcase our results to business, we made some assumptions as below:

- $2,000 gain for every fraud that's caught
- $50 loss for every false positive our model predicted

We calculated fraud savings (the blue line), lost sales (the orange line), and overall savings (the green line) and plotted a graph based on the above assumption and out-of-sample result. The result is shown below.

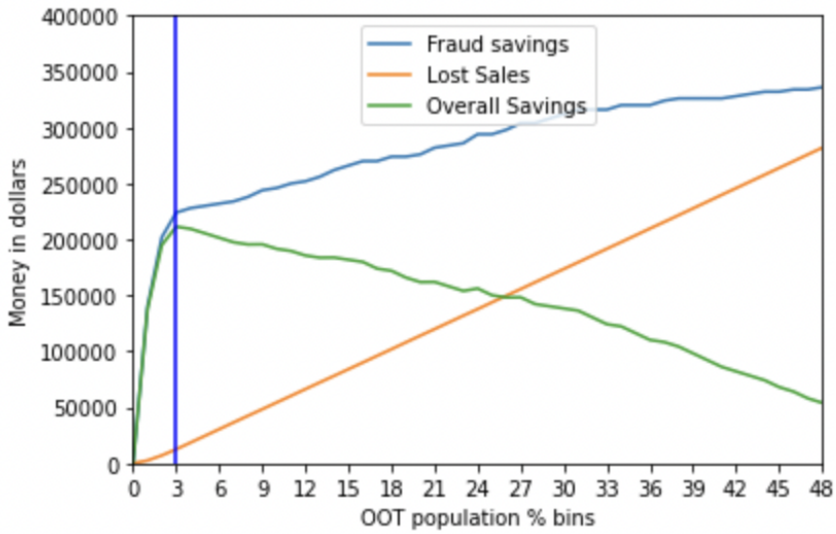

Figure 17. Financial Estimation

By keeping in mind to maximize the annual savings while rejecting as few customers as possible, our recommendation is to set a 3% cutoff. This cutoff company could save around 1.27 million dollars annually by rejecting 3% of transactions.

CONCLUSION AND DISCUSSION

The original dataset has 96,753 records and 10 features and is about transactions across 2006. In summary, we did the following steps:

- Identified null values and abnormalities needed further examination
    - 2 columns contain abnormalities
    - 4 merchant-related features with missing values
- Cleaned dataset based on EDA, getting dataset with 96,397 records
    - Removed 356 abnormal records
    - Fill missing values by mapping with other relevant merchant features. If not able to map, assign ‘unknown’
- Generated 1,310 features in total via feature engineering
- Split train (70% of first 10 months), test (30% of first 10 months) and OOT (last 2 months) dataset
- Use filter and wrapper sequentially to select the top 20 most important predictors
- Tune 6 types of models with different feature sizes and hyperparameters
- Compared FDR at 3% of transaction on OOT data, getting the final model

We measured our final model’s goodness via FDR at 3%. On train, test and OOT data set, our model gives FDR at 3% of transaction of 73.6%, 73.2% and 62.6%, respectively. Notice that the performance on train and test sets are close, implying no overfitting. The below table shows parameters we have tuned for the model. The model uses 15 features and has 20 layers, which is a relatively simple one.

Table 20. Tuning Parameter of the Final Model

# Variables	**hidden_ layer_sizes**	**activation**	**solver**	**alpha**	**learning_rate**	**learning_rate_init**	**max_iter**	**shuffle**
15	(20, 20)	tanh	adam	0.00001	constant	0.001	200	TRUE

Overall, our final model is a good one. With out-of-sample FDR at 3% of transactions, it indicates that the model can eliminate 62.6% of all the frauds while declining only 3% of the transactions. According to our estimation, the company can save 1.27 million dollars per year by implementing this model.

Many different feature selection and modeling methodologies have been left for the future due to time limit. Future work concerns the deeper analysis of entity patterns, new features to try different, more sophisticated models. Specifically, our model can be improved in the following ways:

Due to time and resource limits, our work still has room for improvement. Specifically, we identify 3 future directions:

- In data cleaning process, we overlooked the problem of imbalance dataset. In future, we can consider resampling methods, such as Synthetic Minority Oversampling Technique (SMOTE), to tackle this problem
- In feature engineering process, we are not 100% sure that we include most of features with industrial significance, and we may miss some important fields. We should turn to expertise for further discussion
- In model tuning process, though tried complex, dedicated grids, we aren't likely to reach the sweet spot yet. We can try more sophisticated grids or more advanced model to get a better performance.

APPENDIX

1. Data Quality Report (DQR)

1) File description

The file documents credit card transaction data of 96,753 records in 2006. The data has been labelled and contains a column indicating whether the transaction is identity fraud.

2) Summary statistics table

Table 1 provides main summary of all 10 fields.

Table 1. Summary of Fields

Date Fields						
Field Name	**% Populated**	**Min**		**Max**	**Most Common Value**	
Date	100.00%	2006-01-01		2006-12-31	2006-02-28	
Categorical Fields						
Field Name	**% Populated**	**# Unique Values**			**Most Common Value**	
Recnum	100.00%	96,753			Unique for each record	
Cardnum	100.00%	1,645			5142148452	
Merchnum	96.51%	13,091			930090121224	
Merch description	100.00%	13,126			GSA-FSS-ADV	
Merch state	98.76%	227			TN	
Merch zip	95.19%	4,567			38118	
Transtype	100.00%	4			P	
Fraud	100.00%	2			0	
Numerical Fields						
Field Name	**% Populated**	**Min**	**Max**	**Mean**	**Stdev**	**% Zero**
Amount	100.00%	0.01	3,102,045.53	427.89	10,006.14	0.00%

3) Distributions for fields

a) Recnum

The column serves as a unique identifier of each record.

b) Fraud & Date

i) Daily Trend

The *Date* column marks the date of record is generated, ranging from 2006-01-01 to 2006-12-31. Figure 1 shows the number of records received daily. Notice that not every day it has transaction record.

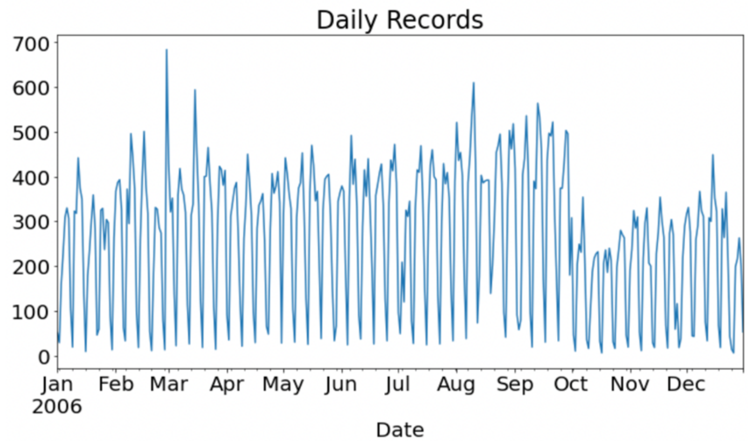

Figure 1. Daily Records

ii) Weekly Trend

Figure 3 shows the proportion of records per week. The sharp drop in late September is because a new fiscal year starts. During most of week, the number of applications weekly flatulates between 1,500 and 2,500.

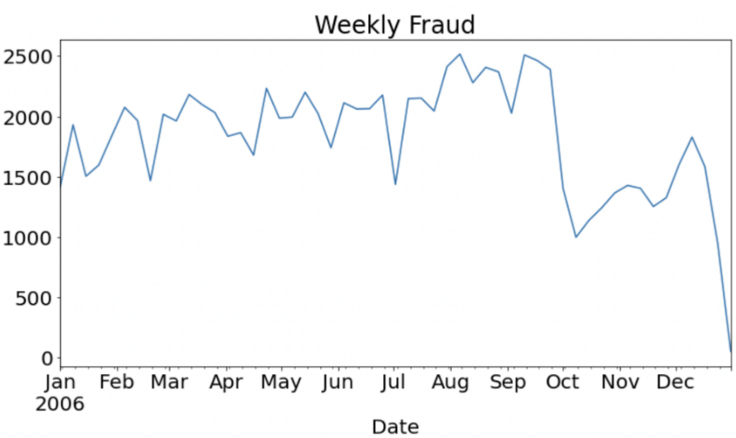

Figure 3. Weekly Records

Figure 4 shows the proportion of records weekly grouped by *Fraud*. The number of not fraud records is more stable than that of fraud records. The first week in the August has the highest number of fraud records, reaching 92. For most of week, the number of not fraud records fluctuates around 1,800 – 2,300.

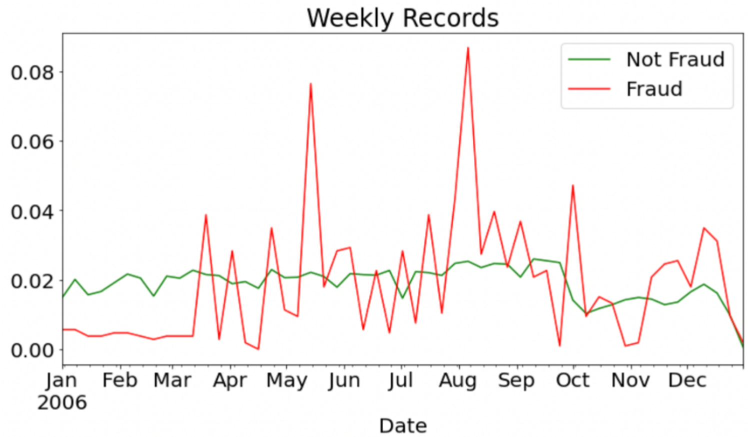

Figure 4. Weekly Records Grouped By *Fraud*

Resampling the data to month level, the trend becomes more stable, while certain patterns hold. There is still a sharp drop in October, and the trend of not fraud records is more stable than that of fraud records.

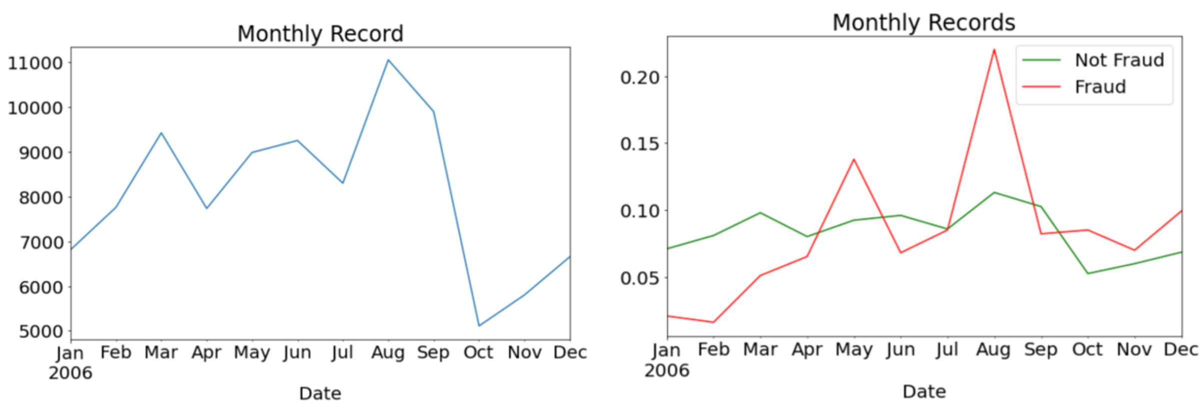

Figure 5. Monthly Records (Left: all data; Right: Group by *Fraud*)

c) Cardnum

The column marks the card number with respect to a transaction. In total, it has 1,645 unique levels. As shown in the figure 5, the most common card number is 5142148452, occurring 1192 times.

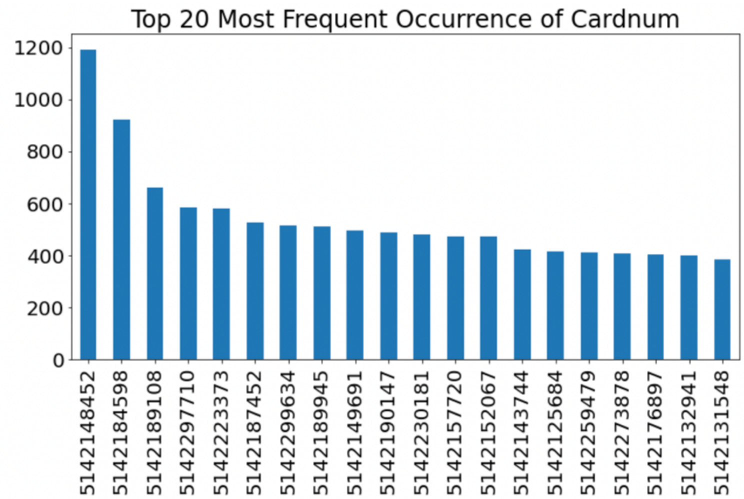

Figure 6. Top 20 Most Frequent Occurrence of *Cardnum*

d) Merchnum

The column marks the merchant number with respect to a record. It has 13,092 unique values, indicating that we have 13,092 different merchants in the data. Figure 6 shows the top 20 most frequent occurrence of *Merchnum*. 9300-9012-1224 is the most frequent one, occurring 9,310 times. Notice that this column contains 3.49% of missing values.

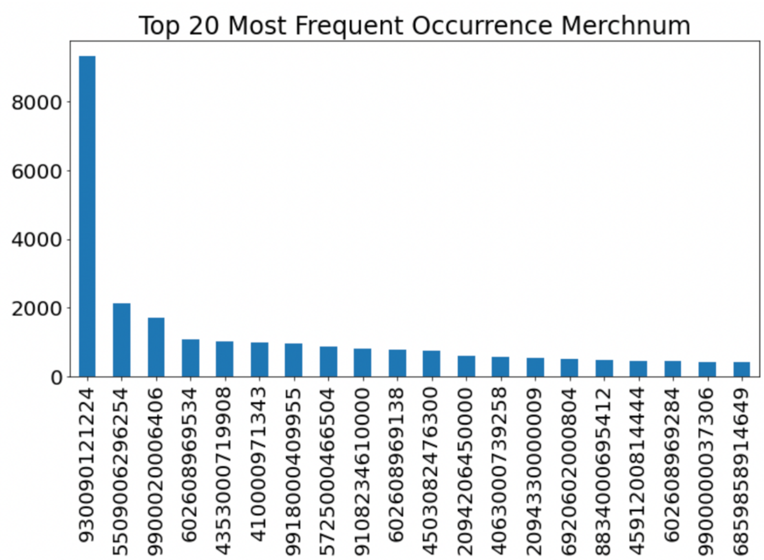

Figure 7. Top 20 Most Frequent Occurrence *Merchnum*

e) Merch description

The column marks the merchant description with respect to a record. It contains 13,126 unique values, a bit more than that of *Merchnum*. The most frequent one is GSA-FSS-ADV, occurring 1,688 times.

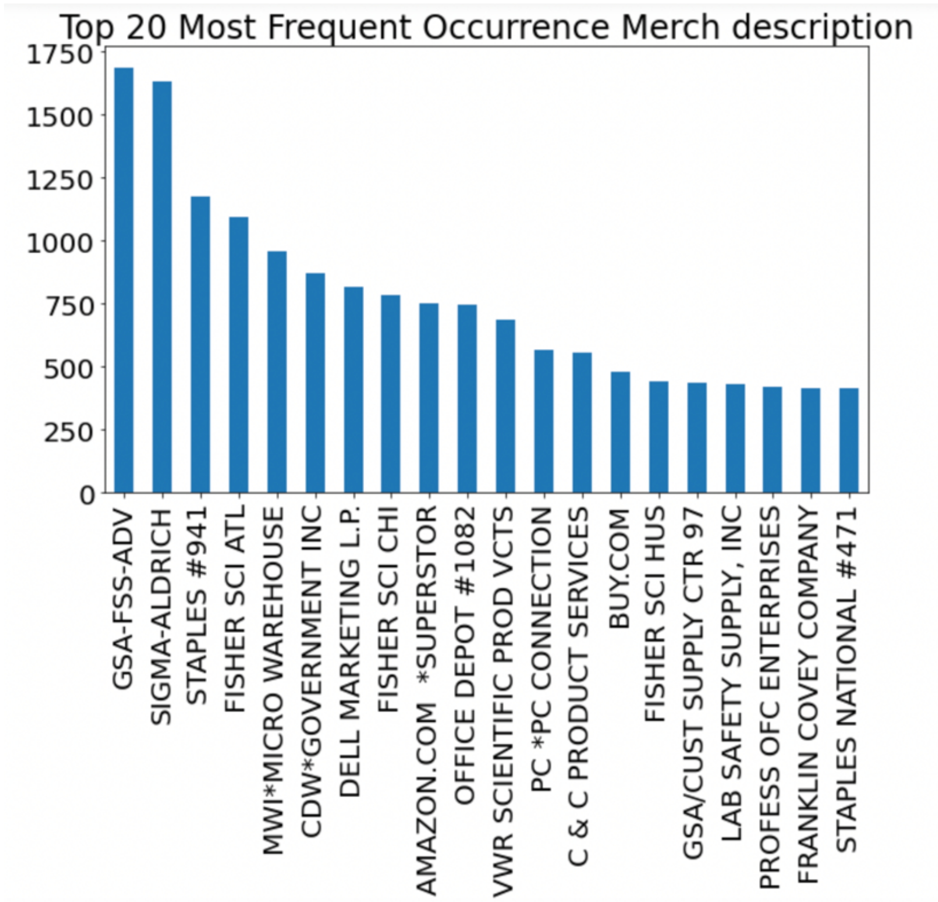

Figure 8. Top 20 Most Frequent Occurrence *Merch description*

f) Merch state

The column marks the state a merchant in with respect to a record. Figure 8 shows top 20 most frequent occurrence of *Merch state*. Many of merchants in the dataset per record are from Tennessee (TN), occurring 11,868 times.

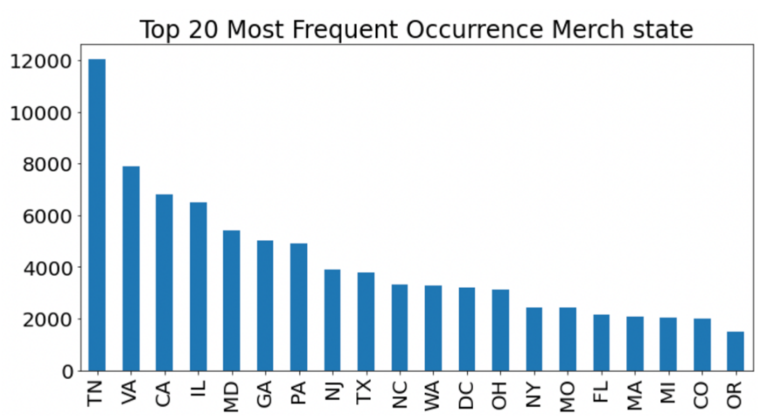

Figure 9. Top 20 Most Frequent Occurrence *Merch state*

g) Merch zip

The column marks the zip code of a merchant with respect to a record. Figure 9 shows top 20 most frequent occurrence of *Merch zip*. The most frequent zip code of merchant is 38118, in Memphis, Tennessee, which is consistent with our previous finding. Notice that the column contains 4.81% of missing values.

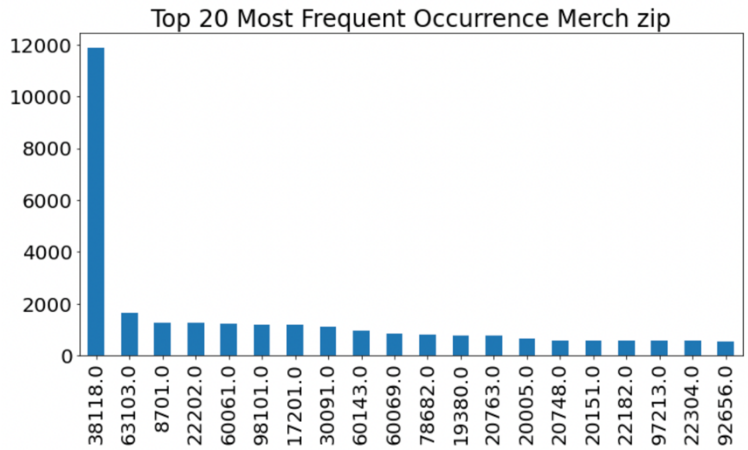

Figure 10. Top 20 Most Frequent Occurrence *Merch zip*

h) Transtype

The column marks transaction type of a record. There are four types of records – P, A, D and Y. The majority of transaction is type P, accounting for 96,398 records. Notice that there is only one record has type Y, which needs further inspect.

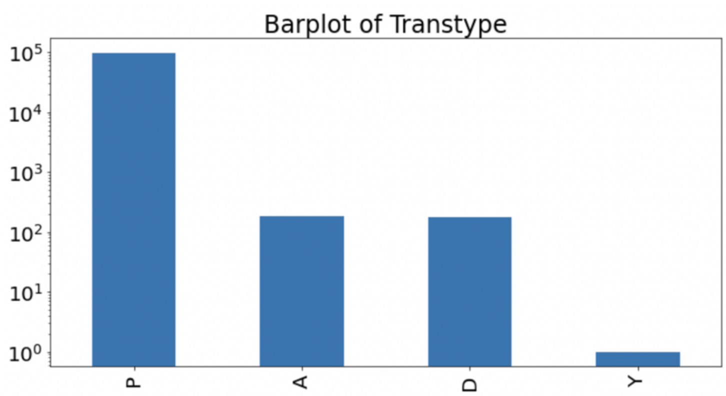

Figure 11. Bar plot of *Transtype*

i) Amount

The column marks the amount spent with respect to a record. Figure 11 shows that there is an outlier on the right-hand side. The record number of this extreme transaction is 52715, spending 3,102,045.53 dollars. After excluding the outlier, the *Amount* is right skewed.

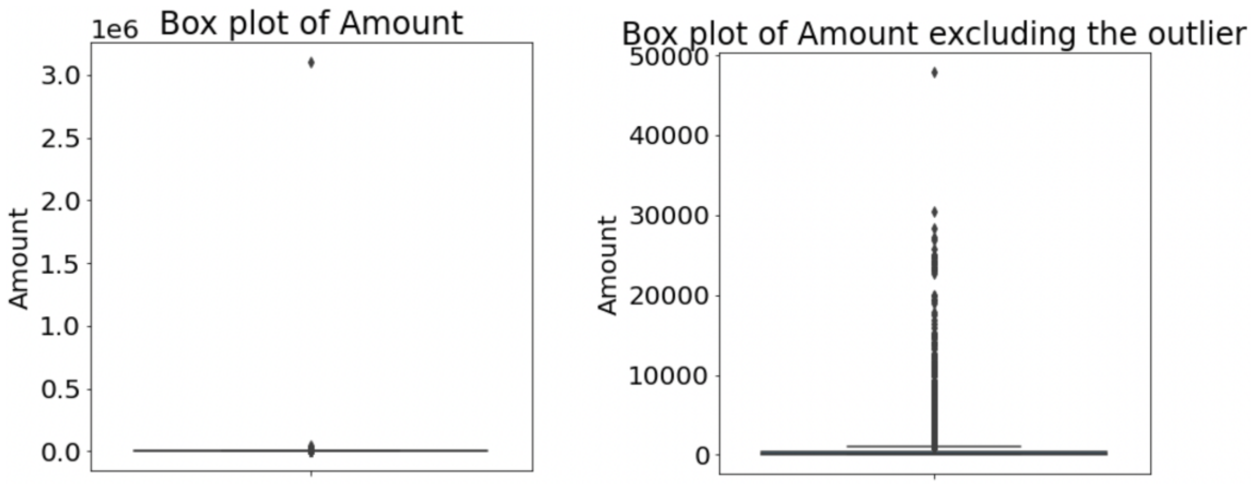

Figure 12. Box plot of *Amount* (Left: all data; Right: excluding the outlier)

To capture the distribution of the majority of data, I zoomed in to records spending less than 3,000 dollars and taking up to 99.33% of data. Figure 12 shows that the distribution of the majority of the data.

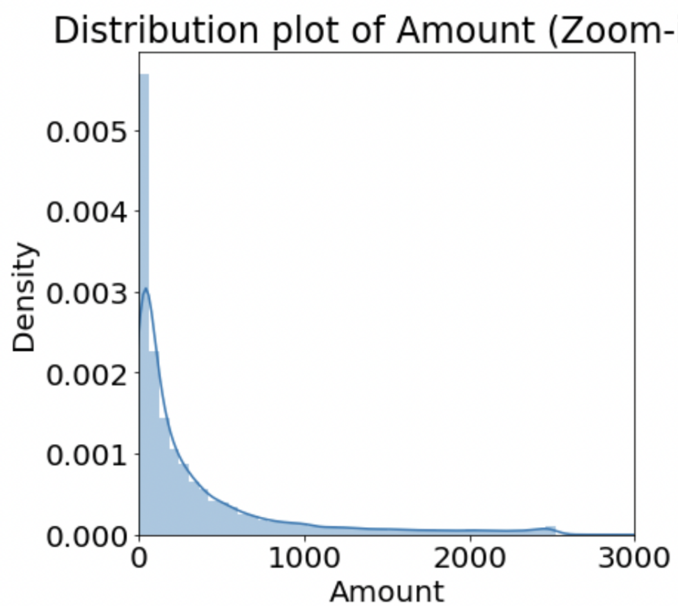

Figure 13. Distribution plot of *Amount (Zoom-in)*

2. Full List of Variables

No.	Name
1	dow_risk
2	state_risk
3	Cardnum_day_since
4	Merchnum_day_since
5	card_merch_day_since
6	card_zip_day_since
7	card_state_day_since
8	merch_zip_day_since
9	merch_state_day_since
10	card_dow_day_since
11	merch_dow_day_since
12	card_merchdesc_day_since
13	card_ zip3_day_since
14	card_merch_month_day_since
15	card_state_month_day_since
16	card_zip_month_day_since
17	card_merchdesc_short_day_since
18	Cardnum_count_0
19	Cardnum_count_1
20	Cardnum_count_3
21	Cardnum_count_7
22	Cardnum_count_14
23	Cardnum_count_30
24	Cardnum_count_90
25	Merchnum_count_0
26	Merchnum_count_1
27	Merchnum_count_3
28	Merchnum_count_7
29	Merchnum_count_14
30	Merchnum_count_30
31	Merchnum_count_90
32	card_merch_count_0
33	card_merch_count_1
34	card_merch_count_3
35	card_merch_count_7
36	card_merch_count_14
37	card_merch_count_30
38	card_merch_count_90
39	card_zip_count_0

40	card_zip_count_1
41	card_zip_count_3
42	card_zip_count_7
43	card_zip_count_14
44	card_zip_count_30
45	card_zip_count_90
46	card_state_count_0
47	card_state_count_1
48	card_state_count_3
49	card_state_count_7
50	card_state_count_14
51	card_state_count_30
52	card_state_count_90
53	merch_zip_count_0
54	merch_zip_count_1
55	merch_zip_count_3
56	merch_zip_count_7
57	merch_zip_count_14
58	merch_zip_count_30
59	merch_zip_count_90
60	merch_state_count_0
61	merch_state_count_1
62	merch_state_count_3
63	merch_state_count_7
64	merch_state_count_14
65	merch_state_count_30
66	merch_state_count_90
67	card_dow_count_0
68	card_dow_count_1
69	card_dow_count_3
70	card_dow_count_7
71	card_dow_count_14
72	card_dow_count_30
73	card_dow_count_90
74	merch_dow_count_0
75	merch_dow_count_1
76	merch_dow_count_3
77	merch_dow_count_7
78	merch_dow_count_14
79	merch_dow_count_30
80	merch_dow_count_90

81	card_merchdesc_count_0
82	card_merchdesc_count_1
83	card_merchdesc_count_3
84	card_merchdesc_count_7
85	card_merchdesc_count_14
86	card_merchdesc_count_30
87	card_merchdesc_count_90
88	card_ zip3_count_0
89	card_ zip3_count_1
90	card_ zip3_count_3
91	card_ zip3_count_7
92	card_ zip3_count_14
93	card_ zip3_count_30
94	card_ zip3_count_90
95	card_merch_month_count_0
96	card_merch_month_count_1
97	card_merch_month_count_3
98	card_merch_month_count_7
99	card_merch_month_count_14
100	card_merch_month_count_30
101	card_merch_month_count_90
102	card_state_month_count_0
103	card_state_month_count_1
104	card_state_month_count_3
105	card_state_month_count_7
106	card_state_month_count_14
107	card_state_month_count_30
108	card_state_month_count_90
109	card_zip_month_count_0
110	card_zip_month_count_1
111	card_zip_month_count_3
112	card_zip_month_count_7
113	card_zip_month_count_14
114	card_zip_month_count_30
115	card_zip_month_count_90
116	card_merchdesc_short_count_0
117	card_merchdesc_short_count_1
118	card_merchdesc_short_count_3
119	card_merchdesc_short_count_7
120	card_merchdesc_short_count_14
121	card_merchdesc_short_count_30

122	card_merchdesc_short_count_90
123	Cardnum_avg_0
124	Cardnum_max_0
125	Cardnum_med_0
126	Cardnum_std_0
127	Cardnum_total_0
128	Cardnum_actual/avg_0
129	Cardnum_actual/max_0
130	Cardnum_actual/med_0
131	Cardnum_actual/toal_0
132	Cardnum_avg_1
133	Cardnum_max_1
134	Cardnum_med_1
135	Cardnum_std_1
136	Cardnum_total_1
137	Cardnum_actual/avg_1
138	Cardnum_actual/max_1
139	Cardnum_actual/med_1
140	Cardnum_actual/toal_1
141	Cardnum_avg_3
142	Cardnum_max_3
143	Cardnum_med_3
144	Cardnum_std_3
145	Cardnum_total_3
146	Cardnum_actual/avg_3
147	Cardnum_actual/max_3
148	Cardnum_actual/med_3
149	Cardnum_actual/toal_3
150	Cardnum_avg_7
151	Cardnum_max_7
152	Cardnum_med_7
153	Cardnum_std_7
154	Cardnum_total_7
155	Cardnum_actual/avg_7
156	Cardnum_actual/max_7
157	Cardnum_actual/med_7
158	Cardnum_actual/toal_7
159	Cardnum_avg_14
160	Cardnum_max_14
161	Cardnum_med_14
162	Cardnum_std_14

163	Cardnum_total_14
164	Cardnum_actual/avg_14
165	Cardnum_actual/max_14
166	Cardnum_actual/med_14
167	Cardnum_actual/toal_14
168	Cardnum_avg_30
169	Cardnum_max_30
170	Cardnum_med_30
171	Cardnum_std_30
172	Cardnum_total_30
173	Cardnum_actual/avg_30
174	Cardnum_actual/max_30
175	Cardnum_actual/med_30
176	Cardnum_actual/toal_30
177	Cardnum_avg_90
178	Cardnum_max_90
179	Cardnum_med_90
180	Cardnum_std_90
181	Cardnum_total_90
182	Cardnum_actual/avg_90
183	Cardnum_actual/max_90
184	Cardnum_actual/med_90
185	Cardnum_actual/toal_90
186	Merchnum_avg_0
187	Merchnum_max_0
188	Merchnum_med_0
189	Merchnum_std_0
190	Merchnum_total_0
191	Merchnum_actual/avg_0
192	Merchnum_actual/max_0
193	Merchnum_actual/med_0
194	Merchnum_actual/toal_0
195	Merchnum_avg_1
196	Merchnum_max_1
197	Merchnum_med_1
198	Merchnum_std_1
199	Merchnum_total_1
200	Merchnum_actual/avg_1
201	Merchnum_actual/max_1
202	Merchnum_actual/med_1
203	Merchnum_actual/toal_1

204	Merchnum_avg_3
205	Merchnum_max_3
206	Merchnum_med_3
207	Merchnum_std_3
208	Merchnum_total_3
209	Merchnum_actual/avg_3
210	Merchnum_actual/max_3
211	Merchnum_actual/med_3
212	Merchnum_actual/toal_3
213	Merchnum_avg_7
214	Merchnum_max_7
215	Merchnum_med_7
216	Merchnum_std_7
217	Merchnum_total_7
218	Merchnum_actual/avg_7
219	Merchnum_actual/max_7
220	Merchnum_actual/med_7
221	Merchnum_actual/toal_7
222	Merchnum_avg_14
223	Merchnum_max_14
224	Merchnum_med_14
225	Merchnum_std_14
226	Merchnum_total_14
227	Merchnum_actual/avg_14
228	Merchnum_actual/max_14
229	Merchnum_actual/med_14
230	Merchnum_actual/toal_14
231	Merchnum_avg_30
232	Merchnum_max_30
233	Merchnum_med_30
234	Merchnum_std_30
235	Merchnum_total_30
236	Merchnum_actual/avg_30
237	Merchnum_actual/max_30
238	Merchnum_actual/med_30
239	Merchnum_actual/toal_30
240	Merchnum_avg_90
241	Merchnum_max_90
242	Merchnum_med_90
243	Merchnum_std_90
244	Merchnum_total_90

245	Merchnum_actual/avg_90
246	Merchnum_actual/max_90
247	Merchnum_actual/med_90
248	Merchnum_actual/toal_90
249	card_merch_avg_0
250	card_merch_max_0
251	card_merch_med_0
252	card_merch_std_0
253	card_merch_total_0
254	card_merch_actual/avg_0
255	card_merch_actual/max_0
256	card_merch_actual/med_0
257	card_merch_actual/toal_0
258	card_merch_avg_1
259	card_merch_max_1
260	card_merch_med_1
261	card_merch_std_1
262	card_merch_total_1
263	card_merch_actual/avg_1
264	card_merch_actual/max_1
265	card_merch_actual/med_1
266	card_merch_actual/toal_1
267	card_merch_avg_3
268	card_merch_max_3
269	card_merch_med_3
270	card_merch_std_3
271	card_merch_total_3
272	card_merch_actual/avg_3
273	card_merch_actual/max_3
274	card_merch_actual/med_3
275	card_merch_actual/toal_3
276	card_merch_avg_7
277	card_merch_max_7
278	card_merch_med_7
279	card_merch_std_7
280	card_merch_total_7
281	card_merch_actual/avg_7
282	card_merch_actual/max_7
283	card_merch_actual/med_7
284	card_merch_actual/toal_7
285	card_merch_avg_14

286	card_merch_max_14
287	card_merch_med_14
288	card_merch_std_14
289	card_merch_total_14
290	card_merch_actual/avg_14
291	card_merch_actual/max_14
292	card_merch_actual/med_14
293	card_merch_actual/toal_14
294	card_merch_avg_30
295	card_merch_max_30
296	card_merch_med_30
297	card_merch_std_30
298	card_merch_total_30
299	card_merch_actual/avg_30
300	card_merch_actual/max_30
301	card_merch_actual/med_30
302	card_merch_actual/toal_30
303	card_merch_avg_90
304	card_merch_max_90
305	card_merch_med_90
306	card_merch_std_90
307	card_merch_total_90
308	card_merch_actual/avg_90
309	card_merch_actual/max_90
310	card_merch_actual/med_90
311	card_merch_actual/toal_90
312	card_zip_avg_0
313	card_zip_max_0
314	card_zip_med_0
315	card_zip_std_0
316	card_zip_total_0
317	card_zip_actual/avg_0
318	card_zip_actual/max_0
319	card_zip_actual/med_0
320	card_zip_actual/toal_0
321	card_zip_avg_1
322	card_zip_max_1
323	card_zip_med_1
324	card_zip_std_1
325	card_zip_total_1
326	card_zip_actual/avg_1

327	card_zip_actual/max_1
328	card_zip_actual/med_1
329	card_zip_actual/toal_1
330	card_zip_avg_3
331	card_zip_max_3
332	card_zip_med_3
333	card_zip_std_3
334	card_zip_total_3
335	card_zip_actual/avg_3
336	card_zip_actual/max_3
337	card_zip_actual/med_3
338	card_zip_actual/toal_3
339	card_zip_avg_7
340	card_zip_max_7
341	card_zip_med_7
342	card_zip_std_7
343	card_zip_total_7
344	card_zip_actual/avg_7
345	card_zip_actual/max_7
346	card_zip_actual/med_7
347	card_zip_actual/toal_7
348	card_zip_avg_14
349	card_zip_max_14
350	card_zip_med_14
351	card_zip_std_14
352	card_zip_total_14
353	card_zip_actual/avg_14
354	card_zip_actual/max_14
355	card_zip_actual/med_14
356	card_zip_actual/toal_14
357	card_zip_avg_30
358	card_zip_max_30
359	card_zip_med_30
360	card_zip_std_30
361	card_zip_total_30
362	card_zip_actual/avg_30
363	card_zip_actual/max_30
364	card_zip_actual/med_30
365	card_zip_actual/toal_30
366	card_zip_avg_90
367	card_zip_max_90

368	card_zip_med_90
369	card_zip_std_90
370	card_zip_total_90
371	card_zip_actual/avg_90
372	card_zip_actual/max_90
373	card_zip_actual/med_90
374	card_zip_actual/toal_90
375	card_state_avg_0
376	card_state_max_0
377	card_state_med_0
378	card_state_std_0
379	card_state_total_0
380	card_state_actual/avg_0
381	card_state_actual/max_0
382	card_state_actual/med_0
383	card_state_actual/toal_0
384	card_state_avg_1
385	card_state_max_1
386	card_state_med_1
387	card_state_std_1
388	card_state_total_1
389	card_state_actual/avg_1
390	card_state_actual/max_1
391	card_state_actual/med_1
392	card_state_actual/toal_1
393	card_state_avg_3
394	card_state_max_3
395	card_state_med_3
396	card_state_std_3
397	card_state_total_3
398	card_state_actual/avg_3
399	card_state_actual/max_3
400	card_state_actual/med_3
401	card_state_actual/toal_3
402	card_state_avg_7
403	card_state_max_7
404	card_state_med_7
405	card_state_std_7
406	card_state_total_7
407	card_state_actual/avg_7
408	card_state_actual/max_7

409	card_state_actual/med_7
410	card_state_actual/toal_7
411	card_state_avg_14
412	card_state_max_14
413	card_state_med_14
414	card_state_std_14
415	card_state_total_14
416	card_state_actual/avg_14
417	card_state_actual/max_14
418	card_state_actual/med_14
419	card_state_actual/toal_14
420	card_state_avg_30
421	card_state_max_30
422	card_state_med_30
423	card_state_std_30
424	card_state_total_30
425	card_state_actual/avg_30
426	card_state_actual/max_30
427	card_state_actual/med_30
428	card_state_actual/toal_30
429	card_state_avg_90
430	card_state_max_90
431	card_state_med_90
432	card_state_std_90
433	card_state_total_90
434	card_state_actual/avg_90
435	card_state_actual/max_90
436	card_state_actual/med_90
437	card_state_actual/toal_90
438	merch_zip_avg_0
439	merch_zip_max_0
440	merch_zip_med_0
441	merch_zip_std_0
442	merch_zip_total_0
443	merch_zip_actual/avg_0
444	merch_zip_actual/max_0
445	merch_zip_actual/med_0
446	merch_zip_actual/toal_0
447	merch_zip_avg_1
448	merch_zip_max_1
449	merch_zip_med_1

450	merch_zip_std_1
451	merch_zip_total_1
452	merch_zip_actual/avg_1
453	merch_zip_actual/max_1
454	merch_zip_actual/med_1
455	merch_zip_actual/toal_1
456	merch_zip_avg_3
457	merch_zip_max_3
458	merch_zip_med_3
459	merch_zip_std_3
460	merch_zip_total_3
461	merch_zip_actual/avg_3
462	merch_zip_actual/max_3
463	merch_zip_actual/med_3
464	merch_zip_actual/toal_3
465	merch_zip_avg_7
466	merch_zip_max_7
467	merch_zip_med_7
468	merch_zip_std_7
469	merch_zip_total_7
470	merch_zip_actual/avg_7
471	merch_zip_actual/max_7
472	merch_zip_actual/med_7
473	merch_zip_actual/toal_7
474	merch_zip_avg_14
475	merch_zip_max_14
476	merch_zip_med_14
477	merch_zip_std_14
478	merch_zip_total_14
479	merch_zip_actual/avg_14
480	merch_zip_actual/max_14
481	merch_zip_actual/med_14
482	merch_zip_actual/toal_14
483	merch_zip_avg_30
484	merch_zip_max_30
485	merch_zip_med_30
486	merch_zip_std_30
487	merch_zip_total_30
488	merch_zip_actual/avg_30
489	merch_zip_actual/max_30
490	merch_zip_actual/med_30

491	merch_zip_actual/toal_30
492	merch_zip_avg_90
493	merch_zip_max_90
494	merch_zip_med_90
495	merch_zip_std_90
496	merch_zip_total_90
497	merch_zip_actual/avg_90
498	merch_zip_actual/max_90
499	merch_zip_actual/med_90
500	merch_zip_actual/toal_90
501	merch_state_avg_0
502	merch_state_max_0
503	merch_state_med_0
504	merch_state_std_0
505	merch_state_total_0
506	merch_state_actual/avg_0
507	merch_state_actual/max_0
508	merch_state_actual/med_0
509	merch_state_actual/toal_0
510	merch_state_avg_1
511	merch_state_max_1
512	merch_state_med_1
513	merch_state_std_1
514	merch_state_total_1
515	merch_state_actual/avg_1
516	merch_state_actual/max_1
517	merch_state_actual/med_1
518	merch_state_actual/toal_1
519	merch_state_avg_3
520	merch_state_max_3
521	merch_state_med_3
522	merch_state_std_3
523	merch_state_total_3
524	merch_state_actual/avg_3
525	merch_state_actual/max_3
526	merch_state_actual/med_3
527	merch_state_actual/toal_3
528	merch_state_avg_7
529	merch_state_max_7
530	merch_state_med_7
531	merch_state_std_7

532	merch_state_total_7
533	merch_state_actual/avg_7
534	merch_state_actual/max_7
535	merch_state_actual/med_7
536	merch_state_actual/toal_7
537	merch_state_avg_14
538	merch_state_max_14
539	merch_state_med_14
540	merch_state_std_14
541	merch_state_total_14
542	merch_state_actual/avg_14
543	merch_state_actual/max_14
544	merch_state_actual/med_14
545	merch_state_actual/toal_14
546	merch_state_avg_30
547	merch_state_max_30
548	merch_state_med_30
549	merch_state_std_30
550	merch_state_total_30
551	merch_state_actual/avg_30
552	merch_state_actual/max_30
553	merch_state_actual/med_30
554	merch_state_actual/toal_30
555	merch_state_avg_90
556	merch_state_max_90
557	merch_state_med_90
558	merch_state_std_90
559	merch_state_total_90
560	merch_state_actual/avg_90
561	merch_state_actual/max_90
562	merch_state_actual/med_90
563	merch_state_actual/toal_90
564	card_dow_avg_0
565	card_dow_max_0
566	card_dow_med_0
567	card_dow_std_0
568	card_dow_total_0
569	card_dow_actual/avg_0
570	card_dow_actual/max_0
571	card_dow_actual/med_0
572	card_dow_actual/toal_0

573	card_dow_avg_1
574	card_dow_max_1
575	card_dow_med_1
576	card_dow_std_1
577	card_dow_total_1
578	card_dow_actual/avg_1
579	card_dow_actual/max_1
580	card_dow_actual/med_1
581	card_dow_actual/toal_1
582	card_dow_avg_3
583	card_dow_max_3
584	card_dow_med_3
585	card_dow_std_3
586	card_dow_total_3
587	card_dow_actual/avg_3
588	card_dow_actual/max_3
589	card_dow_actual/med_3
590	card_dow_actual/toal_3
591	card_dow_avg_7
592	card_dow_max_7
593	card_dow_med_7
594	card_dow_std_7
595	card_dow_total_7
596	card_dow_actual/avg_7
597	card_dow_actual/max_7
598	card_dow_actual/med_7
599	card_dow_actual/toal_7
600	card_dow_avg_14
601	card_dow_max_14
602	card_dow_med_14
603	card_dow_std_14
604	card_dow_total_14
605	card_dow_actual/avg_14
606	card_dow_actual/max_14
607	card_dow_actual/med_14
608	card_dow_actual/toal_14
609	card_dow_avg_30
610	card_dow_max_30
611	card_dow_med_30
612	card_dow_std_30
613	card_dow_total_30

614	card_dow_actual/avg_30
615	card_dow_actual/max_30
616	card_dow_actual/med_30
617	card_dow_actual/toal_30
618	card_dow_avg_90
619	card_dow_max_90
620	card_dow_med_90
621	card_dow_std_90
622	card_dow_total_90
623	card_dow_actual/avg_90
624	card_dow_actual/max_90
625	card_dow_actual/med_90
626	card_dow_actual/toal_90
627	merch_dow_avg_0
628	merch_dow_max_0
629	merch_dow_med_0
630	merch_dow_std_0
631	merch_dow_total_0
632	merch_dow_actual/avg_0
633	merch_dow_actual/max_0
634	merch_dow_actual/med_0
635	merch_dow_actual/toal_0
636	merch_dow_avg_1
637	merch_dow_max_1
638	merch_dow_med_1
639	merch_dow_std_1
640	merch_dow_total_1
641	merch_dow_actual/avg_1
642	merch_dow_actual/max_1
643	merch_dow_actual/med_1
644	merch_dow_actual/toal_1
645	merch_dow_avg_3
646	merch_dow_max_3
647	merch_dow_med_3
648	merch_dow_std_3
649	merch_dow_total_3
650	merch_dow_actual/avg_3
651	merch_dow_actual/max_3
652	merch_dow_actual/med_3
653	merch_dow_actual/toal_3
654	merch_dow_avg_7

655	merch_dow_max_7
656	merch_dow_med_7
657	merch_dow_std_7
658	merch_dow_total_7
659	merch_dow_actual/avg_7
660	merch_dow_actual/max_7
661	merch_dow_actual/med_7
662	merch_dow_actual/toal_7
663	merch_dow_avg_14
664	merch_dow_max_14
665	merch_dow_med_14
666	merch_dow_std_14
667	merch_dow_total_14
668	merch_dow_actual/avg_14
669	merch_dow_actual/max_14
670	merch_dow_actual/med_14
671	merch_dow_actual/toal_14
672	merch_dow_avg_30
673	merch_dow_max_30
674	merch_dow_med_30
675	merch_dow_std_30
676	merch_dow_total_30
677	merch_dow_actual/avg_30
678	merch_dow_actual/max_30
679	merch_dow_actual/med_30
680	merch_dow_actual/toal_30
681	merch_dow_avg_90
682	merch_dow_max_90
683	merch_dow_med_90
684	merch_dow_std_90
685	merch_dow_total_90
686	merch_dow_actual/avg_90
687	merch_dow_actual/max_90
688	merch_dow_actual/med_90
689	merch_dow_actual/toal_90
690	card_merchdesc_avg_0
691	card_merchdesc_max_0
692	card_merchdesc_med_0
693	card_merchdesc_std_0
694	card_merchdesc_total_0
695	card_merchdesc_actual/avg_0

696	card_merchdesc_actual/max_0
697	card_merchdesc_actual/med_0
698	card_merchdesc_actual/toal_0
699	card_merchdesc_avg_1
700	card_merchdesc_max_1
701	card_merchdesc_med_1
702	card_merchdesc_std_1
703	card_merchdesc_total_1
704	card_merchdesc_actual/avg_1
705	card_merchdesc_actual/max_1
706	card_merchdesc_actual/med_1
707	card_merchdesc_actual/toal_1
708	card_merchdesc_avg_3
709	card_merchdesc_max_3
710	card_merchdesc_med_3
711	card_merchdesc_std_3
712	card_merchdesc_total_3
713	card_merchdesc_actual/avg_3
714	card_merchdesc_actual/max_3
715	card_merchdesc_actual/med_3
716	card_merchdesc_actual/toal_3
717	card_merchdesc_avg_7
718	card_merchdesc_max_7
719	card_merchdesc_med_7
720	card_merchdesc_std_7
721	card_merchdesc_total_7
722	card_merchdesc_actual/avg_7
723	card_merchdesc_actual/max_7
724	card_merchdesc_actual/med_7
725	card_merchdesc_actual/toal_7
726	card_merchdesc_avg_14
727	card_merchdesc_max_14
728	card_merchdesc_med_14
729	card_merchdesc_std_14
730	card_merchdesc_total_14
731	card_merchdesc_actual/avg_14
732	card_merchdesc_actual/max_14
733	card_merchdesc_actual/med_14
734	card_merchdesc_actual/toal_14
735	card_merchdesc_avg_30
736	card_merchdesc_max_30

737	card_merchdesc_med_30
738	card_merchdesc_std_30
739	card_merchdesc_total_30
740	card_merchdesc_actual/avg_30
741	card_merchdesc_actual/max_30
742	card_merchdesc_actual/med_30
743	card_merchdesc_actual/toal_30
744	card_merchdesc_avg_90
745	card_merchdesc_max_90
746	card_merchdesc_med_90
747	card_merchdesc_std_90
748	card_merchdesc_total_90
749	card_merchdesc_actual/avg_90
750	card_merchdesc_actual/max_90
751	card_merchdesc_actual/med_90
752	card_merchdesc_actual/toal_90
753	card_ zip3_avg_0
754	card_ zip3_max_0
755	card_ zip3_med_0
756	card_ zip3_std_0
757	card_ zip3_total_0
758	card_ zip3_actual/avg_0
759	card_ zip3_actual/max_0
760	card_ zip3_actual/med_0
761	card_ zip3_actual/toal_0
762	card_ zip3_avg_1
763	card_ zip3_max_1
764	card_ zip3_med_1
765	card_ zip3_std_1
766	card_ zip3_total_1
767	card_ zip3_actual/avg_1
768	card_ zip3_actual/max_1
769	card_ zip3_actual/med_1
770	card_ zip3_actual/toal_1
771	card_ zip3_avg_3
772	card_ zip3_max_3
773	card_ zip3_med_3
774	card_ zip3_std_3
775	card_ zip3_total_3
776	card_ zip3_actual/avg_3
777	card_ zip3_actual/max_3

778	card_ zip3_actual/med_3
779	card_ zip3_actual/toal_3
780	card_ zip3_avg_7
781	card_ zip3_max_7
782	card_ zip3_med_7
783	card_ zip3_std_7
784	card_ zip3_total_7
785	card_ zip3_actual/avg_7
786	card_ zip3_actual/max_7
787	card_ zip3_actual/med_7
788	card_ zip3_actual/toal_7
789	card_ zip3_avg_14
790	card_ zip3_max_14
791	card_ zip3_med_14
792	card_ zip3_std_14
793	card_ zip3_total_14
794	card_ zip3_actual/avg_14
795	card_ zip3_actual/max_14
796	card_ zip3_actual/med_14
797	card_ zip3_actual/toal_14
798	card_ zip3_avg_30
799	card_ zip3_max_30
800	card_ zip3_med_30
801	card_ zip3_std_30
802	card_ zip3_total_30
803	card_ zip3_actual/avg_30
804	card_ zip3_actual/max_30
805	card_ zip3_actual/med_30
806	card_ zip3_actual/toal_30
807	card_ zip3_avg_90
808	card_ zip3_max_90
809	card_ zip3_med_90
810	card_ zip3_std_90
811	card_ zip3_total_90
812	card_ zip3_actual/avg_90
813	card_ zip3_actual/max_90
814	card_ zip3_actual/med_90
815	card_ zip3_actual/toal_90
816	card_merch_month_avg_0
817	card_merch_month_max_0
818	card_merch_month_med_0

819	card_merch_month_std_0
820	card_merch_month_total_0
821	card_merch_month_actual/avg_0
822	card_merch_month_actual/max_0
823	card_merch_month_actual/med_0
824	card_merch_month_actual/toal_0
825	card_merch_month_avg_1
826	card_merch_month_max_1
827	card_merch_month_med_1
828	card_merch_month_std_1
829	card_merch_month_total_1
830	card_merch_month_actual/avg_1
831	card_merch_month_actual/max_1
832	card_merch_month_actual/med_1
833	card_merch_month_actual/toal_1
834	card_merch_month_avg_3
835	card_merch_month_max_3
836	card_merch_month_med_3
837	card_merch_month_std_3
838	card_merch_month_total_3
839	card_merch_month_actual/avg_3
840	card_merch_month_actual/max_3
841	card_merch_month_actual/med_3
842	card_merch_month_actual/toal_3
843	card_merch_month_avg_7
844	card_merch_month_max_7
845	card_merch_month_med_7
846	card_merch_month_std_7
847	card_merch_month_total_7
848	card_merch_month_actual/avg_7
849	card_merch_month_actual/max_7
850	card_merch_month_actual/med_7
851	card_merch_month_actual/toal_7
852	card_merch_month_avg_14
853	card_merch_month_max_14
854	card_merch_month_med_14
855	card_merch_month_std_14
856	card_merch_month_total_14
857	card_merch_month_actual/avg_14
858	card_merch_month_actual/max_14
859	card_merch_month_actual/med_14

860	card_merch_month_actual/toal_14
861	card_merch_month_avg_30
862	card_merch_month_max_30
863	card_merch_month_med_30
864	card_merch_month_std_30
865	card_merch_month_total_30
866	card_merch_month_actual/avg_30
867	card_merch_month_actual/max_30
868	card_merch_month_actual/med_30
869	card_merch_month_actual/toal_30
870	card_merch_month_avg_90
871	card_merch_month_max_90
872	card_merch_month_med_90
873	card_merch_month_std_90
874	card_merch_month_total_90
875	card_merch_month_actual/avg_90
876	card_merch_month_actual/max_90
877	card_merch_month_actual/med_90
878	card_merch_month_actual/toal_90
879	card_state_month_avg_0
880	card_state_month_max_0
881	card_state_month_med_0
882	card_state_month_std_0
883	card_state_month_total_0
884	card_state_month_actual/avg_0
885	card_state_month_actual/max_0
886	card_state_month_actual/med_0
887	card_state_month_actual/toal_0
888	card_state_month_avg_1
889	card_state_month_max_1
890	card_state_month_med_1
891	card_state_month_std_1
892	card_state_month_total_1
893	card_state_month_actual/avg_1
894	card_state_month_actual/max_1
895	card_state_month_actual/med_1
896	card_state_month_actual/toal_1
897	card_state_month_avg_3
898	card_state_month_max_3
899	card_state_month_med_3
900	card_state_month_std_3

901	card_state_month_total_3
902	card_state_month_actual/avg_3
903	card_state_month_actual/max_3
904	card_state_month_actual/med_3
905	card_state_month_actual/toal_3
906	card_state_month_avg_7
907	card_state_month_max_7
908	card_state_month_med_7
909	card_state_month_std_7
910	card_state_month_total_7
911	card_state_month_actual/avg_7
912	card_state_month_actual/max_7
913	card_state_month_actual/med_7
914	card_state_month_actual/toal_7
915	card_state_month_avg_14
916	card_state_month_max_14
917	card_state_month_med_14
918	card_state_month_std_14
919	card_state_month_total_14
920	card_state_month_actual/avg_14
921	card_state_month_actual/max_14
922	card_state_month_actual/med_14
923	card_state_month_actual/toal_14
924	card_state_month_avg_30
925	card_state_month_max_30
926	card_state_month_med_30
927	card_state_month_std_30
928	card_state_month_total_30
929	card_state_month_actual/avg_30
930	card_state_month_actual/max_30
931	card_state_month_actual/med_30
932	card_state_month_actual/toal_30
933	card_state_month_avg_90
934	card_state_month_max_90
935	card_state_month_med_90
936	card_state_month_std_90
937	card_state_month_total_90
938	card_state_month_actual/avg_90
939	card_state_month_actual/max_90
940	card_state_month_actual/med_90
941	card_state_month_actual/toal_90

942	card_zip_month_avg_0
943	card_zip_month_max_0
944	card_zip_month_med_0
945	card_zip_month_std_0
946	card_zip_month_total_0
947	card_zip_month_actual/avg_0
948	card_zip_month_actual/max_0
949	card_zip_month_actual/med_0
950	card_zip_month_actual/toal_0
951	card_zip_month_avg_1
952	card_zip_month_max_1
953	card_zip_month_med_1
954	card_zip_month_std_1
955	card_zip_month_total_1
956	card_zip_month_actual/avg_1
957	card_zip_month_actual/max_1
958	card_zip_month_actual/med_1
959	card_zip_month_actual/toal_1
960	card_zip_month_avg_3
961	card_zip_month_max_3
962	card_zip_month_med_3
963	card_zip_month_std_3
964	card_zip_month_total_3
965	card_zip_month_actual/avg_3
966	card_zip_month_actual/max_3
967	card_zip_month_actual/med_3
968	card_zip_month_actual/toal_3
969	card_zip_month_avg_7
970	card_zip_month_max_7
971	card_zip_month_med_7
972	card_zip_month_std_7
973	card_zip_month_total_7
974	card_zip_month_actual/avg_7
975	card_zip_month_actual/max_7
976	card_zip_month_actual/med_7
977	card_zip_month_actual/toal_7
978	card_zip_month_avg_14
979	card_zip_month_max_14
980	card_zip_month_med_14
981	card_zip_month_std_14
982	card_zip_month_total_14

983	card_zip_month_actual/avg_14
984	card_zip_month_actual/max_14
985	card_zip_month_actual/med_14
986	card_zip_month_actual/toal_14
987	card_zip_month_avg_30
988	card_zip_month_max_30
989	card_zip_month_med_30
990	card_zip_month_std_30
991	card_zip_month_total_30
992	card_zip_month_actual/avg_30
993	card_zip_month_actual/max_30
994	card_zip_month_actual/med_30
995	card_zip_month_actual/toal_30
996	card_zip_month_avg_90
997	card_zip_month_max_90
998	card_zip_month_med_90
999	card_zip_month_std_90
1000	card_zip_month_total_90
1001	card_zip_month_actual/avg_90
1002	card_zip_month_actual/max_90
1003	card_zip_month_actual/med_90
1004	card_zip_month_actual/toal_90
1005	card_merchdesc_short_avg_0
1006	card_merchdesc_short_max_0
1007	card_merchdesc_short_med_0
1008	card_merchdesc_short_std_0
1009	card_merchdesc_short_total_0
1010	card_merchdesc_short_actual/avg_0
1011	card_merchdesc_short_actual/max_0
1012	card_merchdesc_short_actual/med_0
1013	card_merchdesc_short_actual/toal_0
1014	card_merchdesc_short_avg_1
1015	card_merchdesc_short_max_1
1016	card_merchdesc_short_med_1
1017	card_merchdesc_short_std_1
1018	card_merchdesc_short_total_1
1019	card_merchdesc_short_actual/avg_1
1020	card_merchdesc_short_actual/max_1
1021	card_merchdesc_short_actual/med_1
1022	card_merchdesc_short_actual/toal_1
1023	card_merchdesc_short_avg_3

1024	card_merchdesc_short_max_3
1025	card_merchdesc_short_med_3
1026	card_merchdesc_short_std_3
1027	card_merchdesc_short_total_3
1028	card_merchdesc_short_actual/avg_3
1029	card_merchdesc_short_actual/max_3
1030	card_merchdesc_short_actual/med_3
1031	card_merchdesc_short_actual/toal_3
1032	card_merchdesc_short_avg_7
1033	card_merchdesc_short_max_7
1034	card_merchdesc_short_med_7
1035	card_merchdesc_short_std_7
1036	card_merchdesc_short_total_7
1037	card_merchdesc_short_actual/avg_7
1038	card_merchdesc_short_actual/max_7
1039	card_merchdesc_short_actual/med_7
1040	card_merchdesc_short_actual/toal_7
1041	card_merchdesc_short_avg_14
1042	card_merchdesc_short_max_14
1043	card_merchdesc_short_med_14
1044	card_merchdesc_short_std_14
1045	card_merchdesc_short_total_14
1046	card_merchdesc_short_actual/avg_14
1047	card_merchdesc_short_actual/max_14
1048	card_merchdesc_short_actual/med_14
1049	card_merchdesc_short_actual/toal_14
1050	card_merchdesc_short_avg_30
1051	card_merchdesc_short_max_30
1052	card_merchdesc_short_med_30
1053	card_merchdesc_short_std_30
1054	card_merchdesc_short_total_30
1055	card_merchdesc_short_actual/avg_30
1056	card_merchdesc_short_actual/max_30
1057	card_merchdesc_short_actual/med_30
1058	card_merchdesc_short_actual/toal_30
1059	card_merchdesc_short_avg_90
1060	card_merchdesc_short_max_90
1061	card_merchdesc_short_med_90
1062	card_merchdesc_short_std_90
1063	card_merchdesc_short_total_90
1064	card_merchdesc_short_actual/avg_90

1065	card_merchdesc_short_actual/max_90
1066	card_merchdesc_short_actual/med_90
1067	card_merchdesc_short_actual/toal_90
1068	Cardnum_count_0_by_7
1069	Cardnum_total_0_by_7
1070	Cardnum_count_0_by_14
1071	Cardnum_total_0_by_14
1072	Cardnum_count_0_by_30
1073	Cardnum_total_0_by_30
1074	Cardnum_count_0_by_90
1075	Cardnum_total_0_by_90
1076	Cardnum_count_1_by_7
1077	Cardnum_total_1_by_7
1078	Cardnum_count_1_by_14
1079	Cardnum_total_1_by_14
1080	Cardnum_count_1_by_30
1081	Cardnum_total_1_by_30
1082	Cardnum_count_1_by_90
1083	Cardnum_total_1_by_90
1084	Merchnum_count_0_by_7
1085	Merchnum_total_0_by_7
1086	Merchnum_count_0_by_14
1087	Merchnum_total_0_by_14
1088	Merchnum_count_0_by_30
1089	Merchnum_total_0_by_30
1090	Merchnum_count_0_by_90
1091	Merchnum_total_0_by_90
1092	Merchnum_count_1_by_7
1093	Merchnum_total_1_by_7
1094	Merchnum_count_1_by_14
1095	Merchnum_total_1_by_14
1096	Merchnum_count_1_by_30
1097	Merchnum_total_1_by_30
1098	Merchnum_count_1_by_90
1099	Merchnum_total_1_by_90
1100	card_merch_count_0_by_7
1101	card_merch_total_0_by_7
1102	card_merch_count_0_by_14
1103	card_merch_total_0_by_14
1104	card_merch_count_0_by_30
1105	card_merch_total_0_by_30

1106	card_merch_count_0_by_90
1107	card_merch_total_0_by_90
1108	card_merch_count_1_by_7
1109	card_merch_total_1_by_7
1110	card_merch_count_1_by_14
1111	card_merch_total_1_by_14
1112	card_merch_count_1_by_30
1113	card_merch_total_1_by_30
1114	card_merch_count_1_by_90
1115	card_merch_total_1_by_90
1116	card_zip_count_0_by_7
1117	card_zip_total_0_by_7
1118	card_zip_count_0_by_14
1119	card_zip_total_0_by_14
1120	card_zip_count_0_by_30
1121	card_zip_total_0_by_30
1122	card_zip_count_0_by_90
1123	card_zip_total_0_by_90
1124	card_zip_count_1_by_7
1125	card_zip_total_1_by_7
1126	card_zip_count_1_by_14
1127	card_zip_total_1_by_14
1128	card_zip_count_1_by_30
1129	card_zip_total_1_by_30
1130	card_zip_count_1_by_90
1131	card_zip_total_1_by_90
1132	card_state_count_0_by_7
1133	card_state_total_0_by_7
1134	card_state_count_0_by_14
1135	card_state_total_0_by_14
1136	card_state_count_0_by_30
1137	card_state_total_0_by_30
1138	card_state_count_0_by_90
1139	card_state_total_0_by_90
1140	card_state_count_1_by_7
1141	card_state_total_1_by_7
1142	card_state_count_1_by_14
1143	card_state_total_1_by_14
1144	card_state_count_1_by_30
1145	card_state_total_1_by_30
1146	card_state_count_1_by_90

1147	card_state_total_1_by_90
1148	merch_zip_count_0_by_7
1149	merch_zip_total_0_by_7
1150	merch_zip_count_0_by_14
1151	merch_zip_total_0_by_14
1152	merch_zip_count_0_by_30
1153	merch_zip_total_0_by_30
1154	merch_zip_count_0_by_90
1155	merch_zip_total_0_by_90
1156	merch_zip_count_1_by_7
1157	merch_zip_total_1_by_7
1158	merch_zip_count_1_by_14
1159	merch_zip_total_1_by_14
1160	merch_zip_count_1_by_30
1161	merch_zip_total_1_by_30
1162	merch_zip_count_1_by_90
1163	merch_zip_total_1_by_90
1164	merch_state_count_0_by_7
1165	merch_state_total_0_by_7
1166	merch_state_count_0_by_14
1167	merch_state_total_0_by_14
1168	merch_state_count_0_by_30
1169	merch_state_total_0_by_30
1170	merch_state_count_0_by_90
1171	merch_state_total_0_by_90
1172	merch_state_count_1_by_7
1173	merch_state_total_1_by_7
1174	merch_state_count_1_by_14
1175	merch_state_total_1_by_14
1176	merch_state_count_1_by_30
1177	merch_state_total_1_by_30
1178	merch_state_count_1_by_90
1179	merch_state_total_1_by_90
1180	card_dow_count_0_by_7
1181	card_dow_total_0_by_7
1182	card_dow_count_0_by_14
1183	card_dow_total_0_by_14
1184	card_dow_count_0_by_30
1185	card_dow_total_0_by_30
1186	card_dow_count_0_by_90
1187	card_dow_total_0_by_90

1188	card_dow_count_1_by_7
1189	card_dow_total_1_by_7
1190	card_dow_count_1_by_14
1191	card_dow_total_1_by_14
1192	card_dow_count_1_by_30
1193	card_dow_total_1_by_30
1194	card_dow_count_1_by_90
1195	card_dow_total_1_by_90
1196	merch_dow_count_0_by_7
1197	merch_dow_total_0_by_7
1198	merch_dow_count_0_by_14
1199	merch_dow_total_0_by_14
1200	merch_dow_count_0_by_30
1201	merch_dow_total_0_by_30
1202	merch_dow_count_0_by_90
1203	merch_dow_total_0_by_90
1204	merch_dow_count_1_by_7
1205	merch_dow_total_1_by_7
1206	merch_dow_count_1_by_14
1207	merch_dow_total_1_by_14
1208	merch_dow_count_1_by_30
1209	merch_dow_total_1_by_30
1210	merch_dow_count_1_by_90
1211	merch_dow_total_1_by_90
1212	card_merchdesc_count_0_by_7
1213	card_merchdesc_total_0_by_7
1214	card_merchdesc_count_0_by_14
1215	card_merchdesc_total_0_by_14
1216	card_merchdesc_count_0_by_30
1217	card_merchdesc_total_0_by_30
1218	card_merchdesc_count_0_by_90
1219	card_merchdesc_total_0_by_90
1220	card_merchdesc_count_1_by_7
1221	card_merchdesc_total_1_by_7
1222	card_merchdesc_count_1_by_14
1223	card_merchdesc_total_1_by_14
1224	card_merchdesc_count_1_by_30
1225	card_merchdesc_total_1_by_30
1226	card_merchdesc_count_1_by_90
1227	card_merchdesc_total_1_by_90
1228	card_ zip3_count_0_by_7

1229	card_ zip3_total_0_by_7
1230	card_ zip3_count_0_by_14
1231	card_ zip3_total_0_by_14
1232	card_ zip3_count_0_by_30
1233	card_ zip3_total_0_by_30
1234	card_ zip3_count_0_by_90
1235	card_ zip3_total_0_by_90
1236	card_ zip3_count_1_by_7
1237	card_ zip3_total_1_by_7
1238	card_ zip3_count_1_by_14
1239	card_ zip3_total_1_by_14
1240	card_ zip3_count_1_by_30
1241	card_ zip3_total_1_by_30
1242	card_ zip3_count_1_by_90
1243	card_ zip3_total_1_by_90
1244	card_merch_month_count_0_by_7
1245	card_merch_month_total_0_by_7
1246	card_merch_month_count_0_by_14
1247	card_merch_month_total_0_by_14
1248	card_merch_month_count_0_by_30
1249	card_merch_month_total_0_by_30
1250	card_merch_month_count_0_by_90
1251	card_merch_month_total_0_by_90
1252	card_merch_month_count_1_by_7
1253	card_merch_month_total_1_by_7
1254	card_merch_month_count_1_by_14
1255	card_merch_month_total_1_by_14
1256	card_merch_month_count_1_by_30
1257	card_merch_month_total_1_by_30
1258	card_merch_month_count_1_by_90
1259	card_merch_month_total_1_by_90
1260	card_state_month_count_0_by_7
1261	card_state_month_total_0_by_7
1262	card_state_month_count_0_by_14
1263	card_state_month_total_0_by_14
1264	card_state_month_count_0_by_30
1265	card_state_month_total_0_by_30
1266	card_state_month_count_0_by_90
1267	card_state_month_total_0_by_90
1268	card_state_month_count_1_by_7
1269	card_state_month_total_1_by_7

1270	card_state_month_count_1_by_14
1271	card_state_month_total_1_by_14
1272	card_state_month_count_1_by_30
1273	card_state_month_total_1_by_30
1274	card_state_month_count_1_by_90
1275	card_state_month_total_1_by_90
1276	card_zip_month_count_0_by_7
1277	card_zip_month_total_0_by_7
1278	card_zip_month_count_0_by_14
1279	card_zip_month_total_0_by_14
1280	card_zip_month_count_0_by_30
1281	card_zip_month_total_0_by_30
1282	card_zip_month_count_0_by_90
1283	card_zip_month_total_0_by_90
1284	card_zip_month_count_1_by_7
1285	card_zip_month_total_1_by_7
1286	card_zip_month_count_1_by_14
1287	card_zip_month_total_1_by_14
1288	card_zip_month_count_1_by_30
1289	card_zip_month_total_1_by_30
1290	card_zip_month_count_1_by_90
1291	card_zip_month_total_1_by_90
1292	card_merchdesc_short_count_0_by_7
1293	card_merchdesc_short_total_0_by_7
1294	card_merchdesc_short_count_0_by_14
1295	card_merchdesc_short_total_0_by_14
1296	card_merchdesc_short_count_0_by_30
1297	card_merchdesc_short_total_0_by_30
1298	card_merchdesc_short_count_0_by_90
1299	card_merchdesc_short_total_0_by_90
1300	card_merchdesc_short_count_1_by_7
1301	card_merchdesc_short_total_1_by_7
1302	card_merchdesc_short_count_1_by_14
1303	card_merchdesc_short_total_1_by_14
1304	card_merchdesc_short_count_1_by_30
1305	card_merchdesc_short_total_1_by_30
1306	card_merchdesc_short_count_1_by_90
1307	card_merchdesc_short_total_1_by_90
1308	Merchant_with_same_amount_count
1309	Benford’s_Law_cardnum_U*
1310	Benford’s_Law_merchnum_U*