Supervised Credit Card Transaction Fraud Project

Julian Espinoza Martinez

Date: June 11, 2021

Rady School of Management

Table of Contents

Executive Summary

There are two main types of financial fraud that pose enormous threats to the industry if left unchecked. The first is application fraud - fraud that occurs during the account *opening* process - and the second is transaction fraud - fraud that occurs during the account *usage* process. Much to its name, transaction fraud occurs when fraudulent activity is present at the time of the transaction. This fraud can come in the form of stolen or lost credit cards, compromised accounts, and information, and even data breaches from the hacking of information systems. In 2020, transaction fraud totaled just under $35 billion worldwide, about 6.8 cents for every dollar spent, a number that is expected to continue to rise to roughly $40 billion by 2027. In 2016, 47% of customers in the US alone reported experiencing credit card fraud within the past five years.

This report explores the process of building a real-time fraud algorithm that can be used to identify fraudulent transactions. The goal of this project was to build a supervised machine learning algorithm that could effectively label a transaction as fraud at the time of sale, minimizing the costs associated with transaction fraud for the potential client. The report will outline this process, which we have divided into the following sections:

1. Data Description - an exploration of the data used to construct the models
2. Data Cleaning - the process of correcting or excluding inaccurate, omitted, or corrupted values and records within the data
3. Feature Creation - building unique, expert variables using link analysis and grouping
4. Feature Selection - selecting and ranking the most impactful features
5. Modeling - exploring and tuning various supervised machine learning algorithms
6. Results and Analysis - detailed description of final model results and recommendations for future implementation of these real-time fraud detection models

The team explored various algorithms, including Logistic Regression, Random Forest, Gradient Boosted Trees, and Neural Network. The selection of the best algorithm included turning the parameters of each model and recording the performance of the models based on their Fraud Detection Rates (FDR) at a 3% rejection rate.

The best performing model was a Random Forest Classifier (bootstrap=True, num_estimators=500, max_depth=none, max_features=20, min_sample_leaf=15, min_sample_split=2, criterion=gini). The average FDR@3% scores achieved by this model on the training, testing, and OOT data were 98.5%, 87.5%%, and 65.0%, respectively. More detailed statistics of the model's performance on the top 20% of the data shows that the model performs well on the OOT data.

We conclude with a discussion regarding future improvements we could make to our process ensuring a more effective fraud detection algorithm, as well as the future of fraud tools as a whole.

1 Data Description

1.1 File Description

The “card transactions.csv” is a record of credit card transactions for an unknown United States Government organization filed over 12 months. The file stores 96,753 records of miscellaneous credit card transactions for organization employees. It contains 10 fields, with “Amount” being the sole numerical field, with all nine others being categorical. A detailed description of the fields can be found in the appendix. The dataset comes from the U.S. Government and the organization the transactions are sourced from is classified. The time period of the records ranges from January 1, 2010, to December 31, 2010.

Table 1.1.1 File Description

Dataset Name	Card Transactions
Dataset Purpose	Credit card transactions released by the U.S Government to classify fraud
Data Source	Unknown U.S. Government Organization
Time Period	January 1, 2010 – December 31, 2010
# of Fields	10
# of Records	96,753

1.2 Summary Statistics Table

In the dataset only “Amount” is a numeric field, all other fields are categorical. The field “Fraud” was added to the original file inputs used to classify the fraudulence of the entry with a binary value of 1 or 0. A record classified as 1 would meet the characteristics of a fraudulent transaction, while a record classification of 0 would not signify fraudulence. “Transtype” contains four values, being ‘P’, ‘D’, ‘A’, and ‘Y’. We believe ‘D’ signifies delayed capture, ‘A’ and ‘Y’ are unknown classifications, but for our purposes, we will only be using the ‘P’ value, which signifies purchase. The “Amount” distribution below is exhibited as a log scale of count. Additionally, there was one extreme outlier discovered in the “Amount” field, “Recnum” 52715, with a transaction charge of $3,102,045.53. There were 11,775 values under the field “Merch description” that began with the string “FEDEX”. We assume this is because the employees of this organization made frequent transactions at a FedEx facility related to their work. Lastly, there was also a significant decrease in the frequency of transactions in the last four months of the year, which we attribute to the end and beginning of a new yearly budget cycle.

Table 1.2.1: Summary Statistics of Categorical Fields

Column Name	# of Records	% Populated	# Unique Values	Most Common Field
Recnum	96753	100.00	96753	N/A
Cardnum	96753	100.00	1645	5142148452
Date	96753	100.00	365	2/28/10
Merchnum	93378	96.51	13091	930090121224
Merch description	96753	100.00	13126	GSA-FSS-ADV
Merch state	95558	98.76	227	TN
Merch zip	92097	95.19	4567	38118.0
Transtype	96753	100.00	4	P
Fraud	96753	100.00	2	0

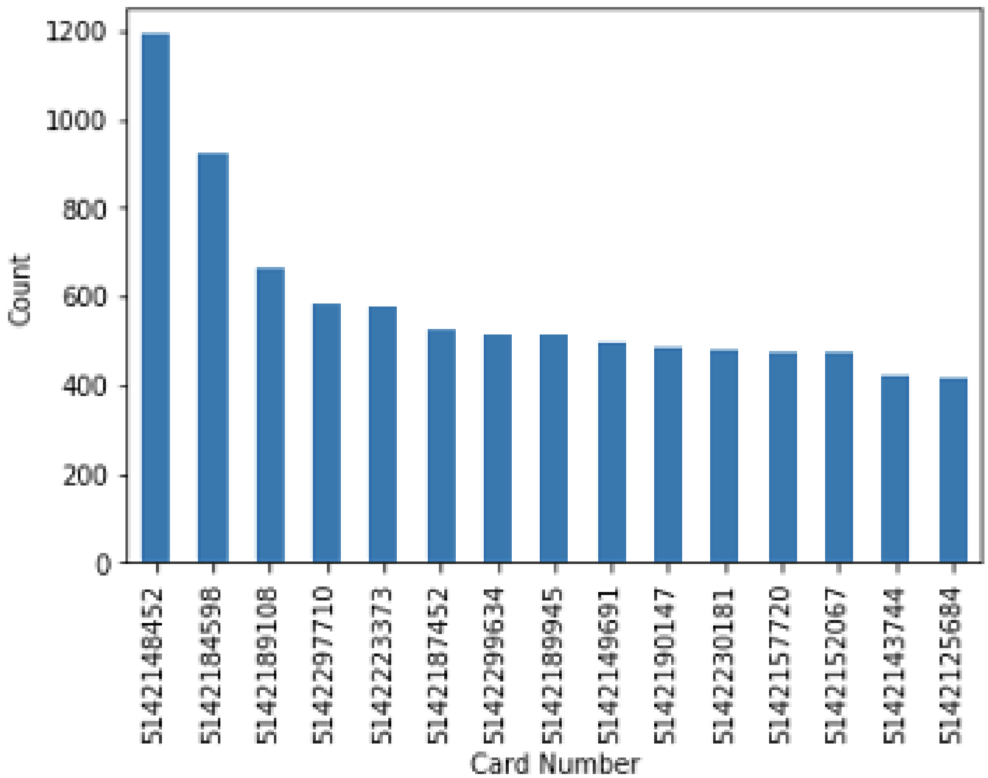
Table 1.2.2 Summary Statistics of Numeric Fields

Column Name	# of Records	% Populated	# Unique Values	# of Zeros	Mean	Std. Dev.	Min.	Max.
Amount	96753	100.00	34909	0	427.89	10006.14	0.01	3102045.53

1.3 Field Examples

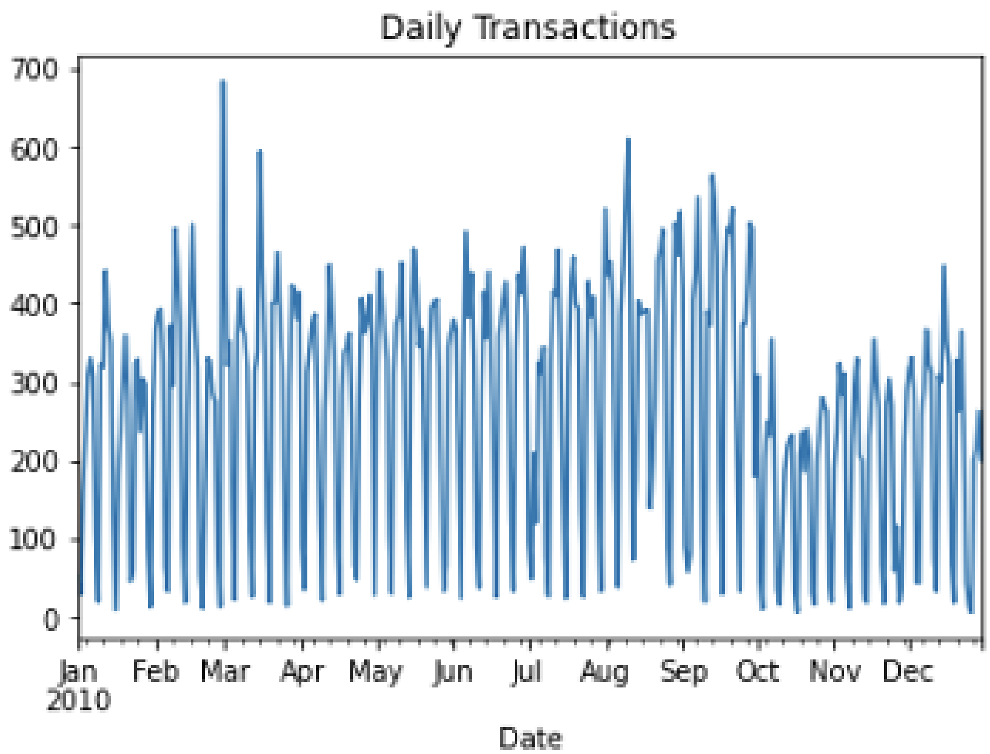
1.3.1 Field “Cardnum”

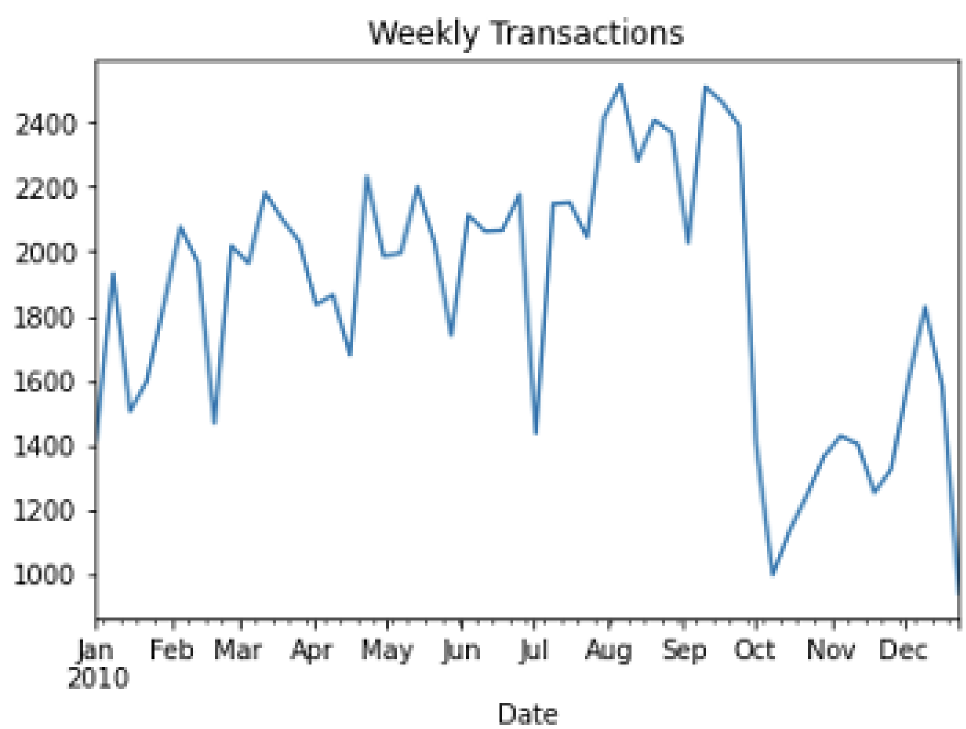
Description: Card number.

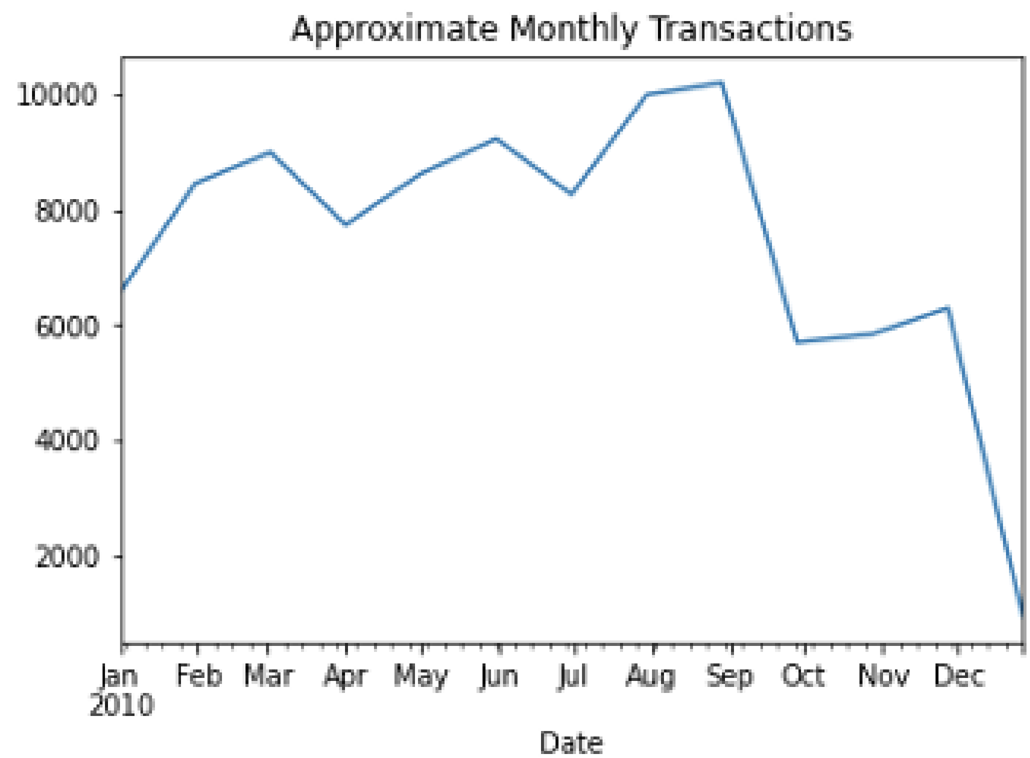


1.3.2 Field “Date”

Description: Date of transaction.

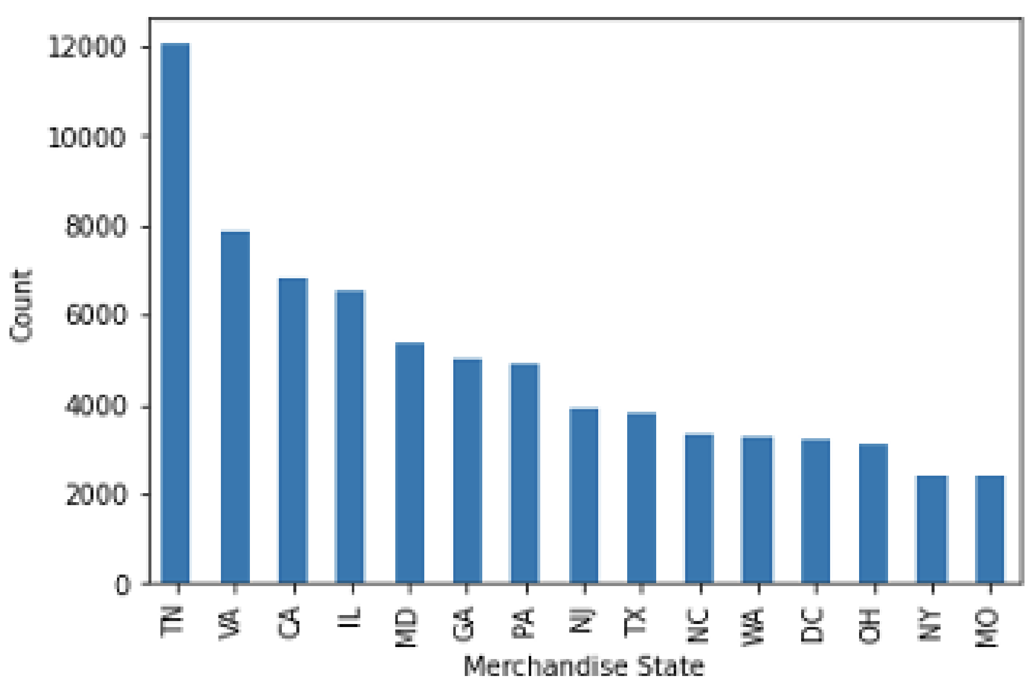






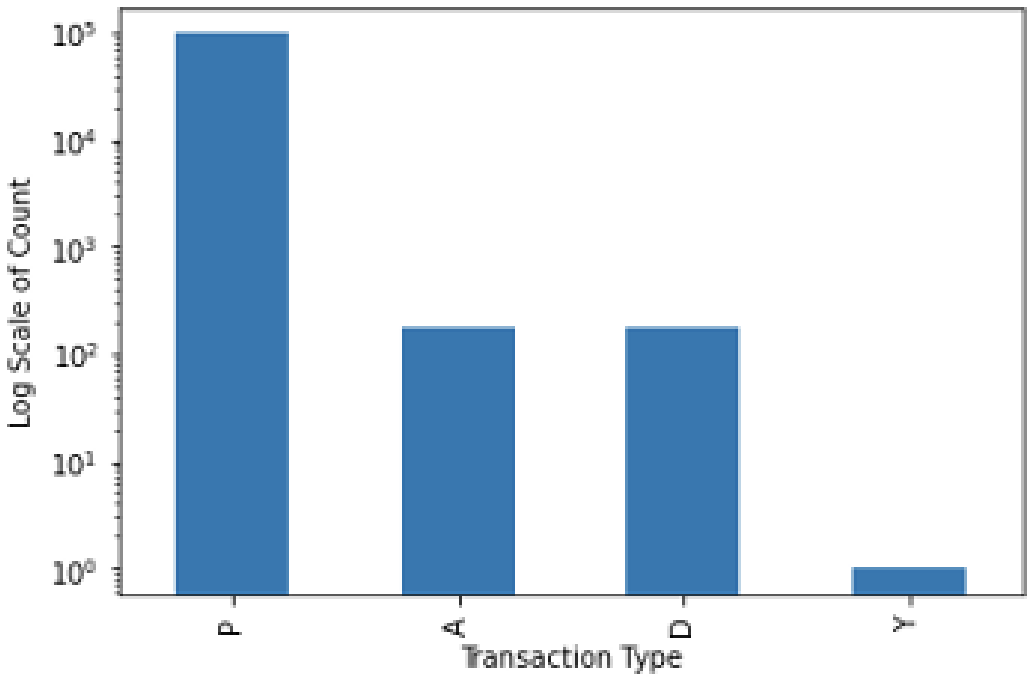
1.3.3 Field “Merch state”

Description: State or country of transaction.



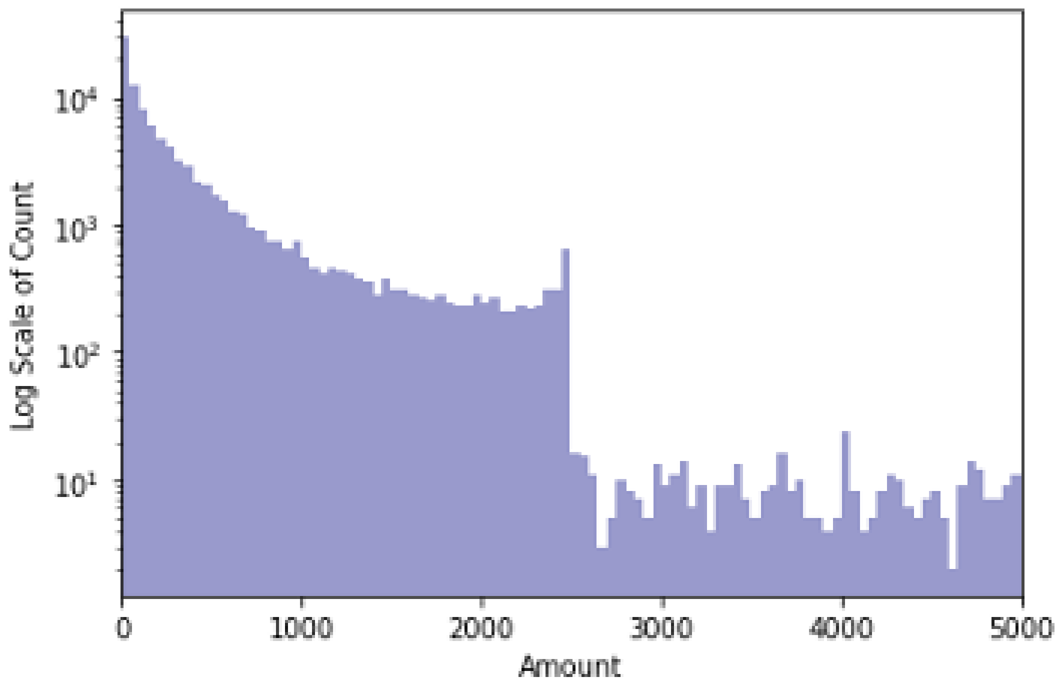
1.3.4 Field "Transtype"

Description: Classification of transaction type.



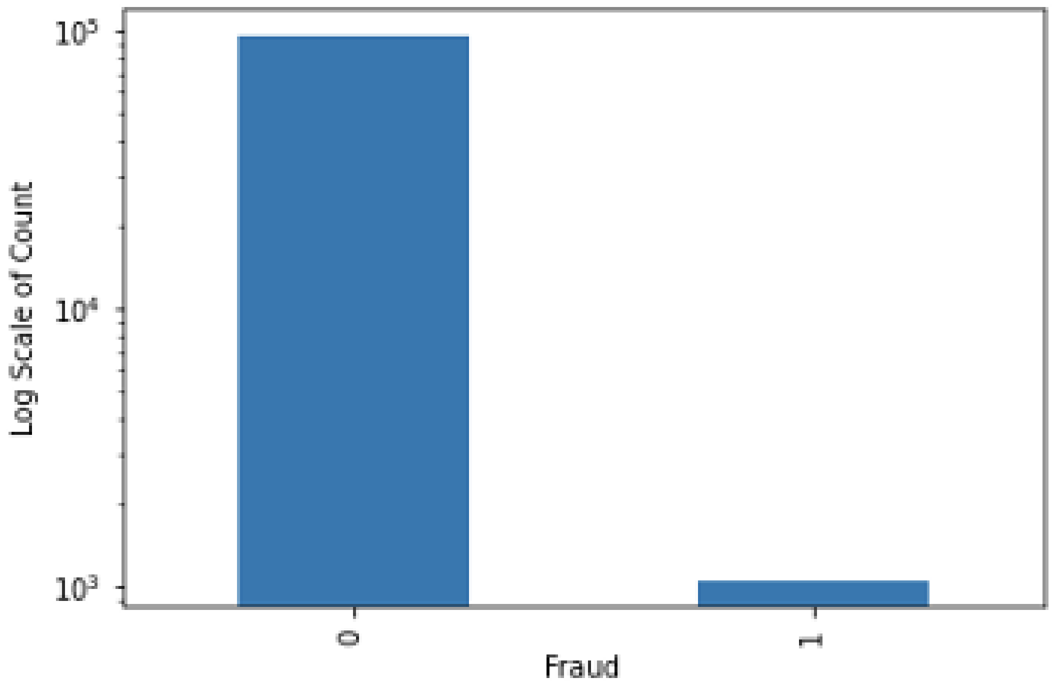
1.3.5 Field "Amount"

Description: Cost of transaction.
Exclude outliers > 5000
Data in the histogram is 99.68% populated.



1.3.6 Field “Fraud”

Description: Binary classification of whether the record appears fraudulent: 1 = Fraud; 0 = No Fraud.



2 Data Cleaning

Before candidate variable creation, the unclean data in the dataset needed to either be omitted or filled strategically. Missing fields, outliers, and unnecessary values need to be filled or removed so they do not negatively impact variable creation. As mentioned before, one record had an extreme outlier value in the “Amount” field. Additionally, we are unsure of three values in “Transtype” and they do not serve any purpose to us in classifying the fraudulence of a purchase. Also, there are missing values in “Merchnum”, “Merch state”, and “Merch zip” that we filled using intuitive techniques, filling with “Unknown” as a last resort. The details of our data cleaning processes are described below:

2.1 Removing Unnecessary Records

We initially removed “Recnum” 52715 with an outlier value in “Amount” of $3,102,045.53. Next, we removed all records with value types other than “P” in “Transtype”, in total there were 96,397 records of value ‘P’, 181 records of value ‘A’, 173 records of value ‘D’, and 1 record of value ‘Y’.

2.2 Filling in Missing “Merchnum” Values

The field “Merchnum” had 3,374 missing values. We first attempted to replace all null values with the most frequent value of “Merchnum” from records that shared the equal value of “Merch description”. If this technique was not applicable, we attempted to fill the value with the most frequent “Merchnum” of records that had an equal value of “Merch zip”, then “Merch state”. We filled in the remaining missing values with “Unknown”.

2.3 Filling in Missing “Merch state” Values

The field “Merch state” had 1,194 missing values. Initially, for records that did not have a “Merch state” value, but had a “Merch zip” value we created a dictionary of states relative to their United States zip codes to fill these instances. The records that had a Puerto Rican zip code, were filled with “PR”. The remaining records were filled in with the most frequent value of “Merch description”. If this was not applicable, we filled in the field with the most frequent value of “Merchnum”. The remaining missing values were replaced with “Unknown”.

2.4 Filling in Missing “Merch zip” Values

The field “Merch zip” had 4,655 missing values. We initially filled these with the most frequent “Merch zip” value of the records that had the equal value of “Merchnum”, then “Merch description”, and lastly “Merch state”. The remaining missing values were replaced with “Unknown”.

3 Candidate Variables

A total of 733 variables were created based on the existing fields in the data set. Using link analysis and grouping of the existing fields, we were able to construct expert variables allowing us to more effectively quantify the characteristics of fraudulent transactions.

When discussing the feature creation you will see us referencing particular *entities*. These entities are simply fields of the data which we are using to group and link the records. We used 11 entities in the feature creation process:

Cardnum - card number
Merchnum - merchant number
Card_merchnum - card number + merchant number
Card_merchdesc - card number + merchant description
Card_state - card number + state
Card_zip - card number + zip
Merchnum_state - merchant number + state
Merchnum_zip - merchant number + zip
Card_merchnum_state - card number + merchant number + state
Card_merchnum_zip - card number + merchant number + zip
Card_merchdesc_zip - card number + merchant description + zip

These entities allow us to group by their fields and analyze the resulting records. For example, the third entity, card_merchnum, identifies all transactions with a particular card at a particular merchant. This allows one to analyze historical data of specific user tendencies, such as how often or how much a card owner spends at a certain merchant or in a certain state. The value of these entities will be more clear throughout the discussion of how the variables are created. Table 3.0 provides a brief overview of the different types of expert variables we created. A list of all 733 created variables, and their summary statistics, can be found in the appendix.

Table 3.0: Summary of 733 Candidate Variables

Category of Variable	Description	Number of Variables
Amount	Measures of amount for particular entities over different periods.	560
Frequency	Number of transactions with an entity over different periods.	70
Days-Since	Days since the most recent transaction with the same entities.	11
Velocity Change	Measures frequency change by looking at	90

	recent activity compared to historic activity.	
Target Encoded	Average fraud likelihood of each weekday and state.	2

3.1 Amount Variables

One sign of transaction fraud is unique transaction amounts - amounts that are not representative of a user's transaction history. Oftentimes, when transaction fraud is committed, transaction amounts do not reflect past behavior. This can take the form of large transactions, big purchases before the fraud gets identified, or a series of smaller transactions, hoping to go undetected by the card owner.

These amount variables are centered around the transaction amount and seek to identify unusual transaction amounts based on past behavior. The variables were constructed using the following logic: Measures of the transaction amount, by/at an entity, over a period of time. Figure 3.1 shows this in detail.

Figure 3.1: Amount Variable Construction

A large value in an amount variable, depending on the time period, means the user has a history of making large purchases with/at a particular entity, whereas a smaller value can indicate the user does not have a history of large transactions in this situation.

3.2 Frequency Variables

Another characteristic of fraudulent transactions is the location at which the card is used. When fraudsters use stolen information to make transactions, they often do so at merchants or in areas that might not be within the card owner’s history.

These frequency variables track the frequency of the card usage and will show unusual usage patterns. The variables were created using this logic and a visual representation can be seen in figure 3.2: Number of transactions with an entity over a period of time.

Figure 3.2: Frequency Variable Construction

Number of transactions with this ...		… over the past ...	
	Card Merchant (num) Card at this merchant (num) Card at this merchant (desc) Card in this state Card in this zip Merchant (num) in this state Merchant (num) in this zip Card at this merchant (num) in this state Card at this merchant (num) in this zip		0 days 1 day 3 days 7 days 14 days 30 days 180 days

Large values in the frequency variables mean that there is a history of transactions with/at the particular entity in the time period.

3.3 Days-Since Variables

The recency of purchases can be indicative of fraudulent behavior. Fraudulent transactions are likely to come in quicker secession of each other. This means multiple recent transactions involving the same entity may be a signal of fraud. For example, recent, recurring purchases in a new state or zip may indicate a lost or stolen credit card.

The days-since variables measure the time since a transaction has been made by/at an entity. These variables track the purchase history of a card (i.e., when is the last time this card was used in a transaction/with this merchant/in this state, etc.). They were calculated by subtracting the most recent transaction with the same entity from the date of the current transaction. A visual representation can be seen in Figure 3.3.

Figure 3.3: Days-Since Variable Construction

Current data minus the date of the most recent transaction with the same...	Card Merchant (num) Card at this merchant (num) Card at this merchant (desc) Card in this state Card in this zip Merchant (num) in this state Merchant (num) in this zip Card at this merchant (num) in this state Card at this merchant (num) in this zip Card at this merchant (desc) in this zip

Small values of this variable mean recent activity with the respective entity. Large values indicated infrequent transactions with the entity. If a transaction is the first of its kind (first purchase at a certain merchant or in a certain state), we have input the value 365.

3.4 Velocity Change Variables

One of the more prominent characteristics of fraudulent behavior is an increase in the number of transactions over a short period of time. When fraudsters act, they often act quickly, making numerous transactions over a short duration before they get shut down. A rapid increase in credit card usage is often linked to fraudulent activity.

These velocity change variables seek to identify burst (rapid) buying behavior associated with stolen credit cards. To create these variables we used the following logic: The number of transactions with the same entity over a time period, divided by the average number of transactions with the same entity over a longer time period. A visual representation of this logic can be seen in figure 3.4.

Figure 3.4: Velocity Change Variable Construction

Number	… of transactions with the same …	Card Merchant (num) Card at this merchant (num) Card at this merchant (desc) Card in this state Card in this zip Merchant (num) in this state Merchant (num) in this zip Card at this merchant (num) in this state Card at this merchant (num) in this zip	… over the past ...	0 1 3	...days
Average daily...	Number	… of transactions with the same …	Same entity as above	… over the past ...	7 days 14 days 30 days

A large value of this variable indicated an increase in recent transactions and should be closely evaluated as a fraud.

3.5 Target Encoded Variables

We also created two target encoded variables, measuring the probability of fraud given the weekday on and state in which the transaction occurred. Target encoding variables, in general, entails assigning a particular value (in our case the average) relating the dependent variable to each level of a categorical variable. This will make more sense as we explain how we constructed our target encoded variables.

Often in a dataset, the number of records in the various levels of categorical variables differs and can be small for certain levels. When a categorical level has a small number of records, the assigned value (in our case the average) may be a misleading and poor estimation of the level's true value. To combat cases when we do have levels with limited records we can apply a logistic smoothing function, ensuring all levels of the categorical variable are closer to the *true* value. For this case we used the following logistic smoothing function:

$$value = Y_{low} + \frac{Y_{high} - Y_{low}}{1 + e^{1(n - n_{mid})/c}}$$

Where Y_{high} is the data wide average, Y_{low} is the categorical average, n is the number of records, n_{mid} is 20, and c is 4.

For this report, we created target encoded variables for weekdays and states. These variables return the average probability of fraud for each weekday and each state. These variables will allow one to see the tendencies of fraudulent activity and provide insight into the behaviors of fraudsters.

3.5.1 Weekday Risk

The average probabilities of fraud by weekdays are summarized in table 3.5.1.

Table 3.5.1.1: Average Probability of Fraud by Weekday

Weekday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
Average Fraud Probability	0.0086	0.0067	0.0098	0.0160	0.0400	0.0111	0.0086

3.5.2 State Risk

The average probabilities of fraud by states are summarized in table 3.5.2.2.

Table 3.5.2.2: Average Probability of Fraud by Top 10 States

Merchant State	Average Fraud Probability
Utah	0.0568
Oregon	0.0276
New York	0.0270
Washington D.C.	0.0216
Maryland	0.0208
Pennsylvania	0.0197
Ohio	0.0182
California	0.0167
Virginia	0.0138
Texas	0.0137

4 Feature Selection Process

4.1 Why apply feature selection?

Feature selection is a very important step when presented with data of high dimensionality. Models usually malfunction because data becomes sparse very quickly, all points become outliers and there is a need for exponentially more data to see true nonlinearities, rather than noise. Other problems that can arise are higher training time, increased risk of overfitting, and higher difficulty of model interpretation. For these reasons, it is convenient to reduce dimensionality by selecting only the appropriate features.

4.2 Data Preprocessing

Before feature selection we had to preprocess the data:

4.2.1 Add in additional test variables:

We added two additional test variables, the fraud label, and a random number to compare the performance of our created features. The fraud label should be the best performing variable since it essentially tells if a transaction is a fraud or not, and the random number should perform poorly given that it is just a random number.

4.2.2 Feature Scaling

Feature scaling is necessary for feature selection and modeling. For this project, we z-scaled our features before doing any feature selection.

The result of z-scaling is that features will be rescaled so that they will have the properties of a standard normal distribution with a mean of zero and a standard deviation of 1. Scaling features so that they follow a standard normal distribution is important if we are comparing measurements that have different units, but it is also a general requirement of some machine learning algorithms which we will create further along this project.

4.2.3 Drop the First Two Weeks Data

The first two weeks' data was dropped to avoid bias and get the most accurate results since there was little to no data before each data point.

4.2.4 Dataset Split

After removing the first two weeks' data, the remaining data was further split into two parts: modeling data and out-of-time (OOT) validation data, that is, the last four months' data. We left aside the OOT data and only used the modeling data as input for the feature selection process to avoid overfitting. After the feature selection is done, the modeling data will be further split into training and testing data for our models.

4.3 Filtering

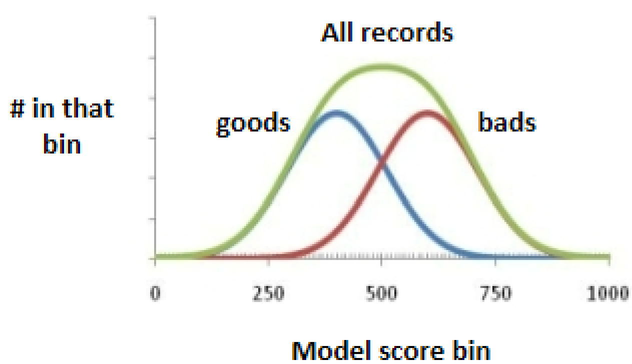
4.3.1 Why use filters?

Filter methods use statistical methods to evaluate the importance of the independent variables (features) for predicting the dependent variable (target). The goal of the filters is to score the importance of each variable by itself (univariate) to predict y (dependent variable). For this project, we used two methods, Kolmogorov-Smirnov and Fraud Detection Rate, and subsequently an average of both to find the most important variables for our models.

4.3.2 Kolmogorov-Smirnov (KS)

4.3.2.1 Why is Kolmogorov-Smirnov a good filter?

Kolmogorov-Smirnov is a robust measure of how well two distributions are separated (goods vs bads). For each candidate feature, we plot the goods and bads separately and observe the curves. The further apart the curves the better the variable fat sorting the goods from the bads, and as a result, the more important the variable is to our model.

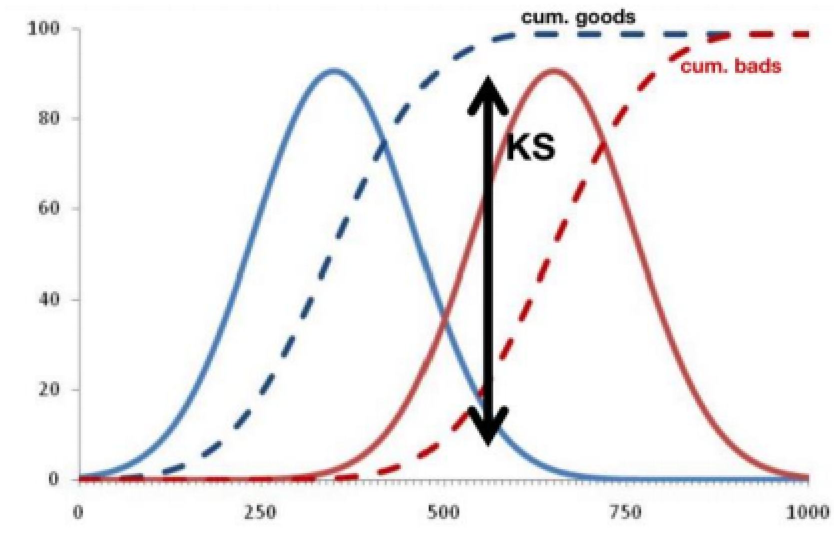


4.3.2.2 How does Kolmogorov-Smirnov work?

For each candidate feature, two distributions of fraud (bad) and non-fraud (good) records are built. Then, the Kolmogorov-Smirnov score is calculated based on the following formula:

$$KS = \max_x \int_{x_{min}}^{x} [P_{\text{goods}} - P_{\text{bads}}]\, dx$$

The formula presented above can be explained as follows: After plotting the frauds (bads) and non-frauds (goods) of a candidate feature, we add them up to form a cumulative distribution. The Kolmogorov-Smirnov score is the maximum of the difference between the cumulative distributions.



4.3.3 Fraud Detection Rate (FDR)

4.3.3.1 Why is FDR a good filter?

The Fraud Detection Rate tells us what percentage of all the frauds are caught at a particular examination cutoff location. It is very common in business applications and more robust and meaningful than the false positive measures of goodness.

4.3.3.2 How does FDR work?

FDR shows what percentage of all frauds are caught at a particular examination cutoff location. In this case, we will set the examination cutoff location at 3%. This means that for example if the FDR is 80% at 3% then the feature catches 80% of all the frauds in 3% of the population. We calculated the FDR of each feature by scoring all the records using each candidate feature, then sorting the records by the score in descending order, after that, we examined the top 3% subpopulation and counted the number of frauds caught, and divided that number by the total number of frauds in the dataset. The resulting percentage is the Fraud Detection Rate at a 3% examination cutoff location.

4.3.4 Kolmogorov-Smirnov and Fraud Detection Rate Combination

Each candidate feature was assigned a KS and an FDR score which we then used to rank them according to their performance following the logic that a higher KS and FDR score is better. We proceeded to take the average of these two rankings:

We then used this average rank of scores to sort the candidate variables in descending order and selected the top 80 features to be further analyzed by a wrapper method.

4.3.5 Wrapper Method

A wrapper method follows a greedy search approach by evaluating all possible combinations of features against the evaluation criterion. The technique we used for our wrapper method was forward selection, which starts with a null model then fits the model with each feature at a time, selecting the feature with the minimum p-value. The model is then fitted with two features, three, and so on, using the previously selected feature(s), trying combinations of previously selected features with other remaining features until we reach 30 selected features with a p-value of individual features less than the significance level.

For our wrapper method, we used a random forest, utilizing five trees, over 80 iterations, and running all processors, as our model for the wrapper method for feature elimination. Random forests have a strong predictive performance, low overfitting, and easy interpretability. For these reasons, we saw this algorithm as a good fit to find the top 30 significant features. We used sequential feature selection with 2-fold cross-validation. The rank order of the 30 features selected by our wrapper is exhibited in table 4.3.5.1 below:

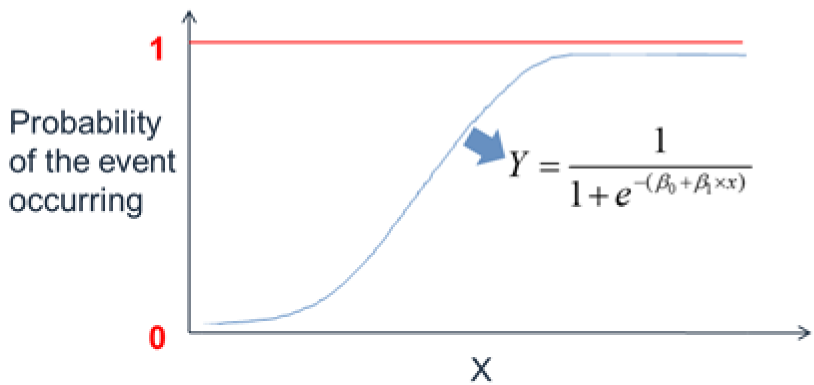
Table 4.3.5.1 Final Features for Fraud Algorithms

Rank	Variable Name
1	card_zip_total_7
2	card_merchnum_zip_total_7
3	card_state_total_7
4	card_merchdesc_total_7
5	card_merchdesc_total_14
6	card_state_total_1
7	card_merchnum_zip_total_14
8	card_state_total_14
9	card_merchnum_state_total_1
10	card_zip_total_30
11	card_merchnum_state_total_30
12	card_merchnum_total_30
13	card_merchnum_zip_total_30
14	card_zip_max_7
15	Cardnum_total_3
16	card_merchnum_total_0
17	card_merchnum_zip_max_14
18	card_state_total_30
19	Cardnum_total_7
20	card_merchdesc_max_3
21	card_zip_total_0
22	card_merchnum_max_3
23	card_state_max_30
24	merchnum_state_total_1
25	card_merchdesc_max_1
26	merchnum_zip_total_1
27	card_merchnum_zip_max_1
28	merchnum_state_max_0
29	merchnum_zip_total_0
30	card_merchdesc_max_180

5 Model Algorithms

5.1 Logistic Regression

We started our model building with a logistic regression to serve as a benchmark for what we could achieve with subsequent models. Logistic regression is one of the simplest and most popular machine learning algorithms for binary classification. The model takes several parameters called features or independent variables and a binary dependent variable or target that can take the values of either 0 or 1. The value “1” represents that an event happens, the value “0” represents the contrary. The following figure represents the basics of logistic regression.



Logistic regression models are simple and effective, easy to modify and interpret, therefore they are widely used in binary classification problems. In this credit card transaction fraud project, we used a logistic regression as a baseline to compare other models like a neural network, gradient boosted trees and random forest. We used the top 30 variables from our feature selection process as independent variables and a fraud label field as our target or dependent variable, which indicates if the transaction is fraudulent (1) or non-fraudulent (0). We then tuned the following parameters:

Name	Description
Penalty	Used to specify the norm used in the penalization. We used “l2”, “l1”, and “elasticnet”.
C	The inverse of regularization strength; must be a positive float. The smaller the value the stronger the regularization.
Solver	Refers to the optimization method to use to find the optimum of the objective function. We tested the methods “lbfgs”, “liblinear”, and “saga”.

L1 Ratio	A parameter in a [0,1] range weighting l1 vs l2 regularization.

The results of the logistic regression model are summarized in the following table:

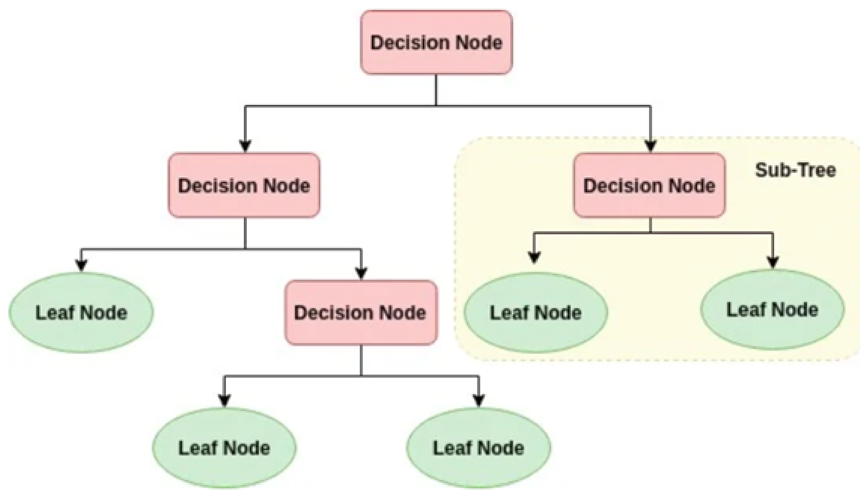
Table 5.1.1 Logistic Regression Results

Iteration	# of Variables	Penalty	C	Solver	L1 Ratio	Training	Testing	OOT
1	30	l2	10	lbfgs	none	0.684	0.691	0.537
2	30	l2	1	lbfgs	none	0.676	0.69	0.543
3	30	l1	1	liblinear	none	0.682	0.694	0.538
4	30	l2	1	liblinear	none	0.681	0.68	0.54
5	30	elasticnet	1	saga	0.5	0.653	0.648	0.533
6	30	elasticnet	1	saga	0.3	0.649	0.659	0.527
7	30	elasticnet	1	saga	0.7	0.65	0.657	0.538
8	20	l2	1	liblinear	none	0.619	0.619	0.468
9	20	l2	1	lbfgs	none	0.616	0.616	0.472
10	30	l2	0.1	lbfgs	none	0.664	0.677	0.544

5.2 Random Forest

Next, we used random forest, an ensemble supervised machine learning algorithm that fits several weak decision tree classifiers on various sub-samples of the dataset and applies to average to improve the predictive accuracy and to control over-fitting. Random forests are popularly applied to both data science competitions and practical problems. They are often accurate, do not require feature scaling, categorical feature encoding, and need little parameter tuning. They are also more interpretable than other complex models such as neural networks.

Two types of randomness are built into the trees. First, each tree is built on a random sample from the original data. Second, at each tree node, a subset of features is randomly selected to generate the best split. Random forests are powerful not only in classification and regression, but also for purposes such as outlier detection, clustering, and interpreting a data set. The figure below represents a decision tree that forms the foundations of this modeling technique:



In our case, we have leveraged the random forest model to perform binary classification, which is one of the strong use cases that random forests can handle with ease and efficiency. We fed the top 30 variables from our feature selection process as independent variables and the dependent variable “fraud label”, which determines whether a transaction is classified as fraudulent or not. We tuned the following parameters on the algorithm followed by its performance:

Name	Description
n_estimators	Number of decision trees it creates to predict and average out the prediction results
max_features	One of the features that induce randomness to the process and is the

	number of features that are considered for each tree
min_samples_leaf	The minimum number of samples required to be at a leaf node
min_samples_split	The minimum number of samples required to split an internal node

The results of the random forest model are summarized in the following table:

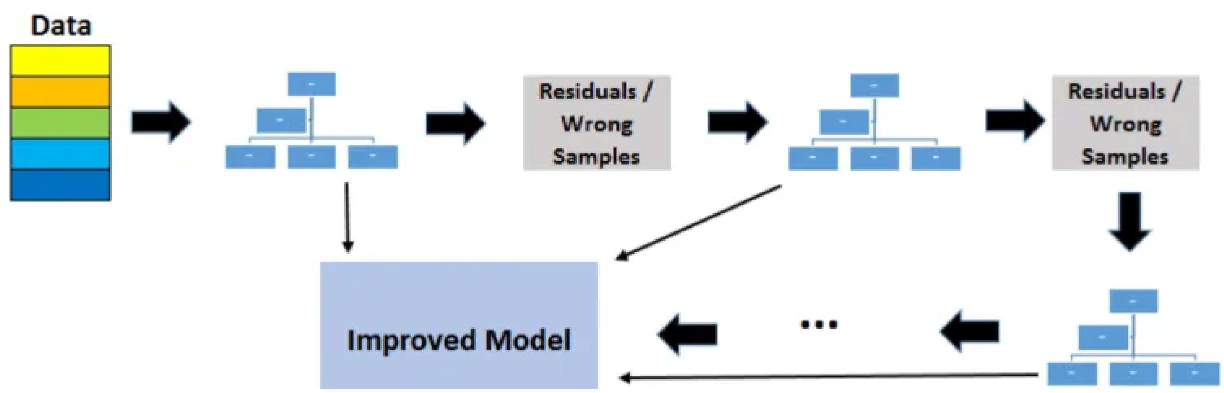
Table 5.2.1 Random Forest Model Results

Iteration	Bootstrap	# of Estimators	Max Depth	Max Features	Min Sample Leaf	Min Sample Split	Criterion	Training	Testing	OOT
1	TRUE	100	none	none	1	2	gini	1.000	0.897	0.611
2	TRUE	100	none	5	3	2	gini	1.000	0.867	0.625
3	TRUE	100	none	5	3	2	entropy	1.000	0.869	0.636
4	TRUE	200	none	5	3	2	entropy	1.000	0.874	0.632
5	TRUE	100	none	10	10	2	gini	1.000	0.870	0.638
6	TRUE	100	none	10	10	2	entropy	1.000	0.882	0.637
7	TRUE	50	none	10	30	500	entropy	1.000	0.792	0.606
8	TRUE	500	none	20	15	2	gini	0.985	0.875	0.650
9	TRUE	500	none	20	15	2	entropy	0.997	0.878	0.637

5.3 Gradient Boosted Trees

Gradient boosting is one of the most powerful techniques for building predictive models. In gradient boosting decision trees, we combine many weak learners to come up with one strong learner. The weak learners here are the individual decision trees. All the trees are connected in series and each tree attempts to minimize the error of the previous tree.
Due to this sequential connection, boosting algorithms are usually slow to learn (controllable by the developer using the learning rate parameter), but also highly accurate. In statistical learning, models that learn slowly perform better. The weak learners are fitted in such a way that each new learner fits into the residuals of the previous step so as the model improves. The final model adds up the result of each step and thus a stronger learner is eventually achieved.

The figure below explains the flow within the algorithm:



Below are the parameters we have tuned followed by the performance of the model:

Name	Description
learning_rate	The learning rate shrinks the contribution of each tree. There is a trade-off between learning_rate and n_estimators.
n_estimators	The number of boosting stages to perform.
max_depth	The maximum depth of the individual regression estimators.
max_features	The number of features to consider when looking for the best split
min_samples_leaf	The minimum number of samples required to be at a leaf node
min_samples_split	The minimum number of samples required to split an internal node

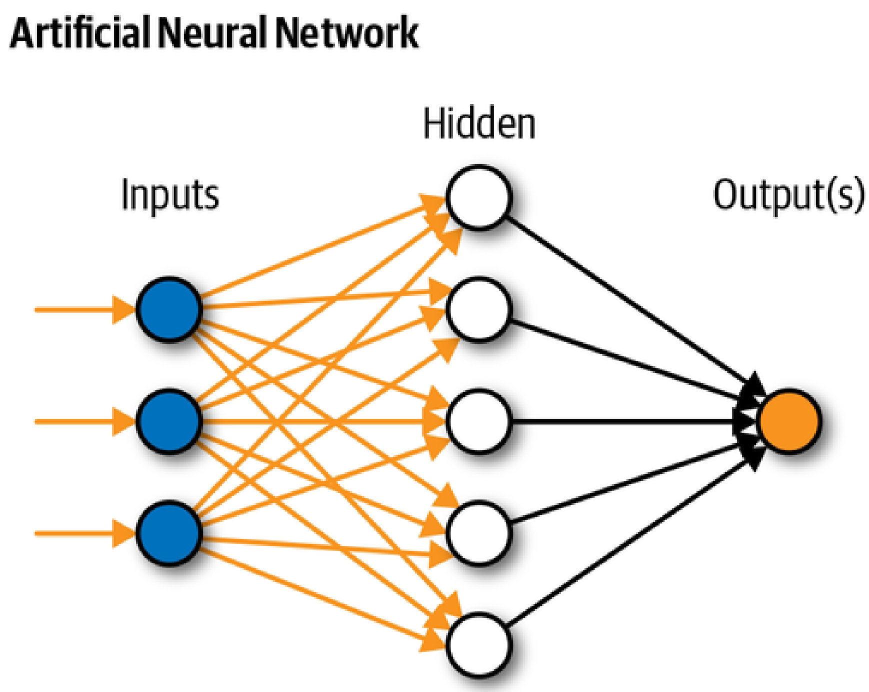
The results of the boosted trees model are summarized in the following table:

Table 5.3.1 Gradient Boosted Trees Model Results

Iteration	Learning Rate	# of Estimators	Max Depth	Max Features	Min Sample Leaf	Min Sample Split	SubSample	Training	Testing	OOT
1	0.1	100	3	none	1	2	1	0.893	0.839	0.608
2	0.1	500	3	5	1	2	1	0.984	0.88	0.637
3	0.1	500	5	5	10	20	1	0.890	0.773	0.540
4	0.1	1000	5	5	10	200	0.7	0.875	0.814	0.569
5	0.01	500	3	5	1	2	0.7	0.834	0.774	0.600
6	0.01	500	3	5	1	2	1	0.838	0.771	0.597
7	0.01	100	3	5	1	2	1	0.733	0.714	0.544
8	0.001	500	3	5	40	1000	0.7	0.711	0.697	0.553

5.4 Neural Network

Artificial neural networks (ANNs), or simply called neural networks (NNs), are one type of machine learning algorithm. The name neural network is derived from the biological architecture of the neural system, where signals flow from a neuron to another neuron multiple times. Just like a neuron system, NNs consist of an input layer, an output layer, and hidden layers that exist between the input and output layer. Each of these layers contains nodes that convey signal(information) to nodes in the next layer. Once a node receives a signal, signals are combined and activated by some computational methods such as Rectified Linear Activation (ReLU), Logistic (Sigmoid), or Hyperbolic Tangent (Tanh). A diagram of a neural network with one hidden layer is exhibited below:



Below are the parameters we have tuned followed by the performance of the model:

Name (sklearn)	Description
hidden_layer_size	The ith element represents the number of neurons in the ith hidden layer. We used combinations of 1-4 hidden layers that have 10 or 20 nodes per layer.
activation	Activation function for the hidden layer. We used 'relu'.
solver	The solver for weight optimization. We used 'lbfgs' and 'sgd'.
alpha	L2 penalty (regularization term) parameter. We used the default value of 0.0001.

The results of the neural network model are summarized in the following table:

Table 5.4.1 Neural Network Model Results

Iteration	Hidden Layers	Nodes Per Layer	Activation	Max Iterations	Alpha	Solver	Tolerance	Learning Rate	Learning Rate init	Training	Testing	OOT
1	1	20	relu	2000	0.0001	adam	0.00001	constant	0.01	0.882	0.834	0.629
2	1	10	relu	2000	0.0001	adam	0.00001	constant	0.01	0.826	0.796	0.596
3	2	10	relu	2000	0.0001	adam	0.00001	constant	0.01	0.867	0.811	0.61
4	3	10	relu	2000	0.0001	adam	0.00001	constant	0.01	0.853	0.795	0.644
5	4	10	relu	2000	0.0001	adam	0.00001	constant	0.01	0.873	0.817	0.639
6	3	10	relu	2000	0.0001	sgd	0.00001	constant	0.01	0.859	0.796	0.597
7	3	10	relu	2000	0.0001	sgd	0.00001	adaptive	0.01	0.867	0.8	0.622
8	2	10	relu	2000	0.0001	sgd	0.00001	adaptive	0.01	0.81	0.781	0.619
9	1	20	relu	2000	0.0001	sgd	0.00001	adaptive	0.01	0.794	0.784	0.634
10	1	10	relu	2000	0.0001	sgd	0.00001	adaptive	0.01	0.775	0.765	0.621

6 Results

6.1 Final Model

We determined that our Random Forest Model performed the best of our four initial training models. Although subject to overfitting, the model outperformed the four other models in fraud detection rates for the testing and out-of-time validation data. With an average FDR at a rejection rate of 3% at 86.7% for the testing set and 63.0% for the out-of-time validation set.

To select the optimal hyperparameters, we utilized GridSearchCV and our tuned selection of hyperparameters included:

bootstrap=True, n_estimators=500, max_depth=none, max_features=20, min_sample_leaf=15, min_sample_split=2, criterion=gini

Table 6.1.1 Key Statistics for Random Forest Final Model

Iteration	Bootstrap	# of Estimators	Max Depth	Max Features	Min Sample Leaf	Min Sample Split	Criterion	Training	Testing	OOT
1	TRUE	100	none	none	1	2	gini	1.000	0.897	0.611
2	TRUE	100	none	5	3	2	gini	1.000	0.867	0.625
3	TRUE	100	none	5	3	2	entropy	1.000	0.869	0.636
4	TRUE	200	none	5	3	2	entropy	1.000	0.874	0.632
5	TRUE	100	none	10	10	2	gini	1.000	0.870	0.638
6	TRUE	100	none	10	10	2	entropy	1.000	0.882	0.637
7	TRUE	50	none	10	30	500	entropy	1.000	0.792	0.606
8	TRUE	500	none	20	15	2	gini	0.985	0.875	0.650
9	TRUE	500	none	20	15	2	entropy	0.997	0.878	0.637

Table 6.1.3 Final Model Results Training Data

Training	# of Records	# of Goods	# of Bads	Fraud Rate								
	48,332	47,845	487	0.0102								
	Bin Statistics					Cumulative Statisitcs						
Population Bin %	# Records	# Goods	# Bads	% Goods	% Bads	Total # Records	Cumulative Goods	Cumulative Bads	% Goods	% Bads (FDR)	KS	FPR
1	483	114	369	23.6	76.4	483	114	369	0.2	75.8	75.5	0.3
2	484	415	69	85.7	14.3	967	529	438	1.1	89.9	88.8	1.2
3	483	439	44	90.9	9.1	1450	968	482	2.0	99.0	97.0	2.0
4	483	478	5	99.0	1.0	1933	1446	487	3.0	100	97.0	3.0
5	484	484	0	100	0	2417	1930	487	4.0	100	96.0	4.0
6	483	483	0	100	0	2900	2413	487	5.0	100	95.0	5.0
7	483	483	0	100	0	3383	2896	487	6.1	100	93.9	5.9
8	484	484	0	100	0	3867	3380	487	7.1	100	92.9	6.9
9	483	483	0	100	0	4350	3863	487	8.1	100	91.9	7.9
10	483	483	0	100	0	4833	4346	487	9.1	100	90.9	8.9
11	484	484	0	100	0	5317	4830	487	10.1	100	89.9	9.9
12	483	483	0	100	0	5800	5313	487	11.1	100	88.9	10.9
13	483	483	0	100	0	6283	5796	487	12.1	100	87.9	11.9
14	483	483	0	100	0	6766	6279	487	13.1	100	86.9	12.9
15	484	484	0	100	0	7250	6763	487	14.1	100	85.9	13.9
16	483	483	0	100	0	7733	7246	487	15.1	100	84.9	14.9
17	483	483	0	100	0	8216	7729	487	16.2	100	83.8	15.9
18	484	484	0	100	0	8700	8213	487	17.2	100	82.8	16.9
19	483	483	0	100	0	9183	8696	487	18.2	100	81.8	17.9
20	483	483	0	100	0	9666	9179	487	19.2	100	80.8	18.8

Table 6.1.4 Final Model Results Testing Data

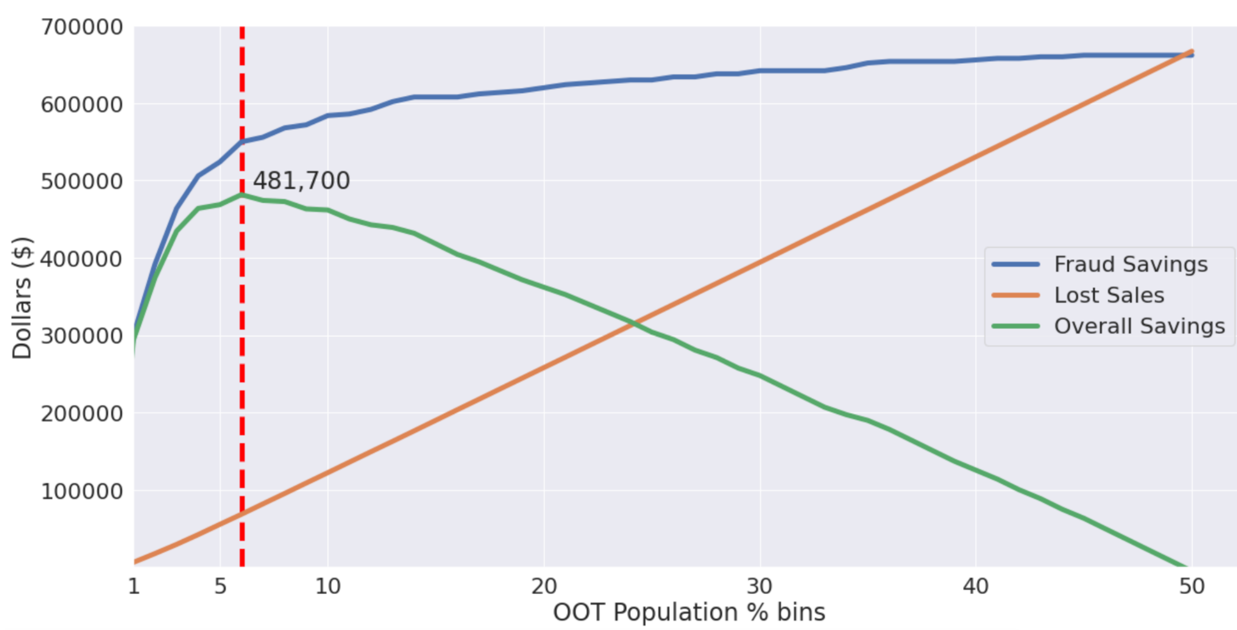
Testing	# of Records	# of Goods	# of Bads	Fraud Rate								
	20,714	20,498	216	0.0105								
	Bin Statistics					Cumulative Statisitcs						
Population Bin %	# Records	# Goods	# Bads	% Goods	% Bads	Total # Records	Cumulative Goods	Cumulative Bads	% Goods	% Bads (FDR)	KS	FPR
1	207	64	143	30.9	69.1	207	64	143	0.3	66.2	65.9	0.4
2	207	170	37	82.1	17.9	414	234	180	1.1	83.3	82.2	1.3
3	207	195	12	94.2	5.8	621	429	192	2.1	88.9	86.8	2.2
4	208	204	4	98.1	1.9	829	633	196	3.1	90.7	87.7	3.2
5	207	204	3	98.6	1.4	1036	837	199	4.1	92.1	88.0	4.2
6	207	206	1	99.5	0.5	1243	1043	200	5.1	92.6	87.5	5.2
7	207	205	2	99.0	1.0	1450	1248	202	6.1	93.5	87.4	6.2
8	207	205	2	99.0	1.0	1657	1453	204	7.1	94.4	87.4	7.1
9	207	205	2	99.0	1.0	1864	1658	206	8.1	95.4	87.3	8.0
10	207	206	1	99.5	0.5	2071	1864	207	9.1	95.8	86.7	9.0
11	208	207	1	99.5	0.5	2279	2071	208	10.1	96.3	86.2	10.0
12	207	207	0	100.0	0.0	2486	2278	208	11.1	96.3	85.2	11.0
13	207	207	0	100.0	0.0	2693	2485	208	12.1	96.3	84.2	11.9
14	207	206	1	99.5	0.5	2900	2691	209	13.1	96.8	83.6	12.9
15	207	207	0	100.0	0.0	3107	2898	209	14.1	96.8	82.6	13.9
16	207	206	1	99.5	0.5	3314	3104	210	15.1	97.2	82.1	14.8
17	207	207	0	100.0	0.0	3521	3311	210	16.2	97.2	81.1	15.8
18	208	208	0	100.0	0.0	3729	3519	210	17.2	97.2	80.1	16.8
19	207	207	0	100.0	0.0	3936	3726	210	18.2	97.2	79.0	17.7
20	207	207	0	100.0	0.0	4143	3933	210	19.2	97.2	78.0	18.7

Table 6.1.5 Final Model Results OOT Data

Out Of Time	# of Records	# of Goods	# of Bads	Fraud Rate								
	27,315	26,995	356	0.0132								
	Bin Statistics					Cumulative Statisitcs						
Population Bin %	# Records	# Goods	# Bads	% Goods	% Bads	Total # Records	Cumulative Goods	Cumulative Bads	% Goods	% Bads (FDR)	KS	FPR
1	274	123	151	44.9	55.1	207	64	143	0.3	66.2	65.9	0.4
2	273	228	45	83.5	16.5	414	234	180	1.1	83.3	82.2	1.3
3	274	238	36	86.9	13.1	621	429	192	2.1	88.9	86.8	2.2
4	273	252	21	92.3	7.7	829	633	196	3.1	90.7	87.7	3.2
5	274	265	9	96.7	3.3	1036	837	199	4.1	92.1	88.0	4.2
6	273	260	13	95.2	4.8	1243	1043	200	5.1	92.6	87.5	5.2
7	274	271	3	98.9	1.1	1450	1248	202	6.1	93.5	87.4	6.2
8	273	267	6	97.8	2.2	1657	1453	204	7.1	94.4	87.4	7.1
9	274	272	2	99.3	0.7	1864	1658	206	8.1	95.4	87.3	8.0
10	273	267	6	97.8	2.2	2071	1864	207	9.1	95.8	86.7	9.0
11	274	273	1	99.6	0.4	2279	2071	208	10.1	96.3	86.2	10.0
12	273	270	3	98.9	1.1	2486	2278	208	11.1	96.3	85.2	11.0
13	274	269	5	98.2	1.8	2693	2485	208	12.1	96.3	84.2	11.9
14	273	270	3	98.9	1.1	2900	2691	209	13.1	96.8	83.6	12.9
15	274	274	0	100.0	0.0	3107	2898	209	14.1	96.8	82.6	13.9
16	273	273	0	100.0	0.0	3314	3104	210	15.1	97.2	82.1	14.8
17	274	272	2	99.3	0.7	3521	3311	210	16.2	97.2	81.1	15.8
18	273	272	1	99.6	0.4	3729	3519	210	17.2	97.2	80.1	16.8
19	274	273	1	99.6	0.4	3936	3726	210	18.2	97.2	79.0	17.7
20	273	271	2	99.3	0.7	4143	3933	210	19.2	97.2	78.0	18.7

The following figure shows the suggested cutoff recommendation as well as the expected annual savings when using our final model.

Figure 6.1.5 Fraud Savings Calculation

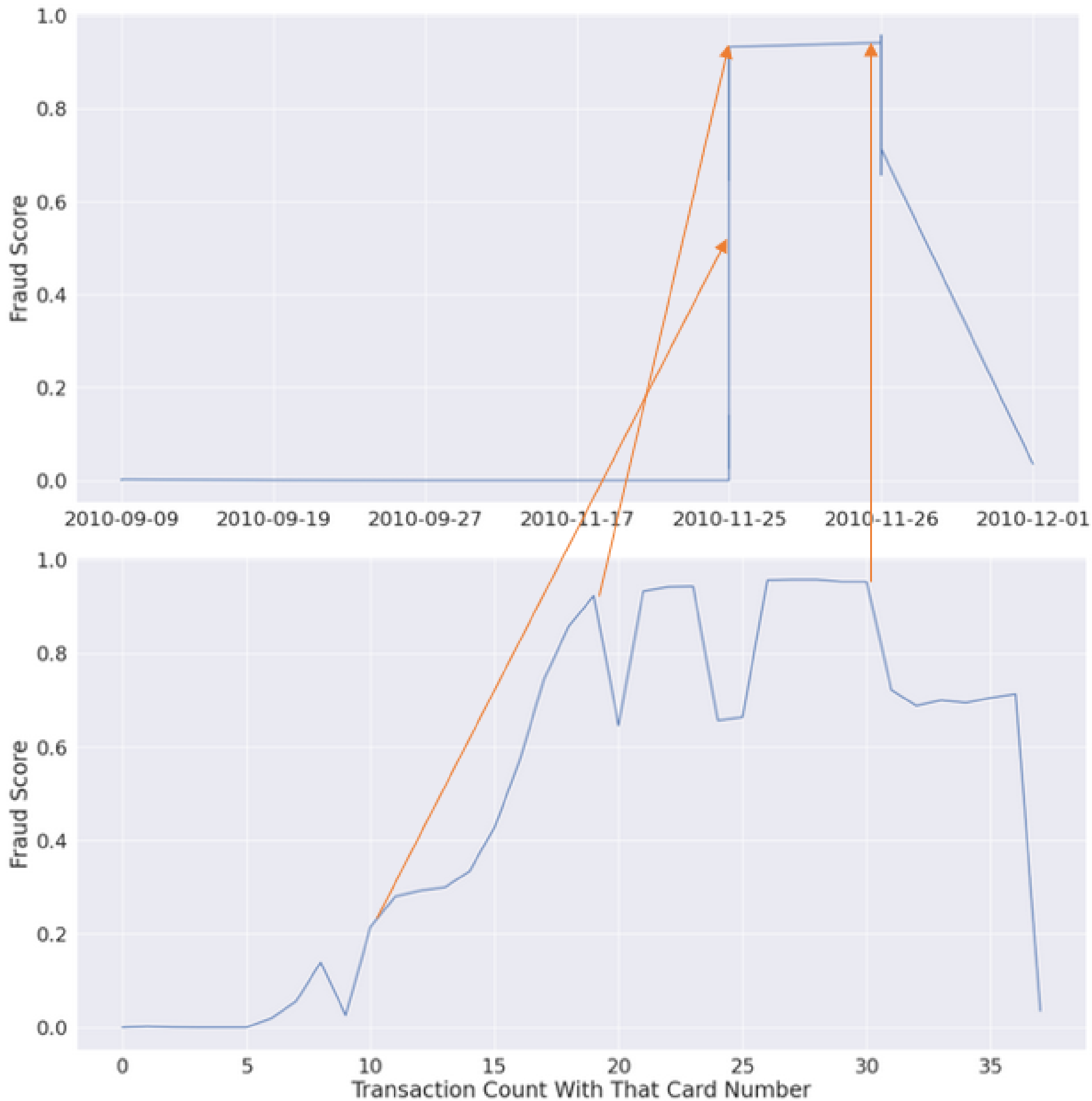


Score cutoff recommendation: 6%

Expected annual savings when using our final model: $481,700

6.2 Card Number and Merchant Fraud Examples as Calculated by our Model

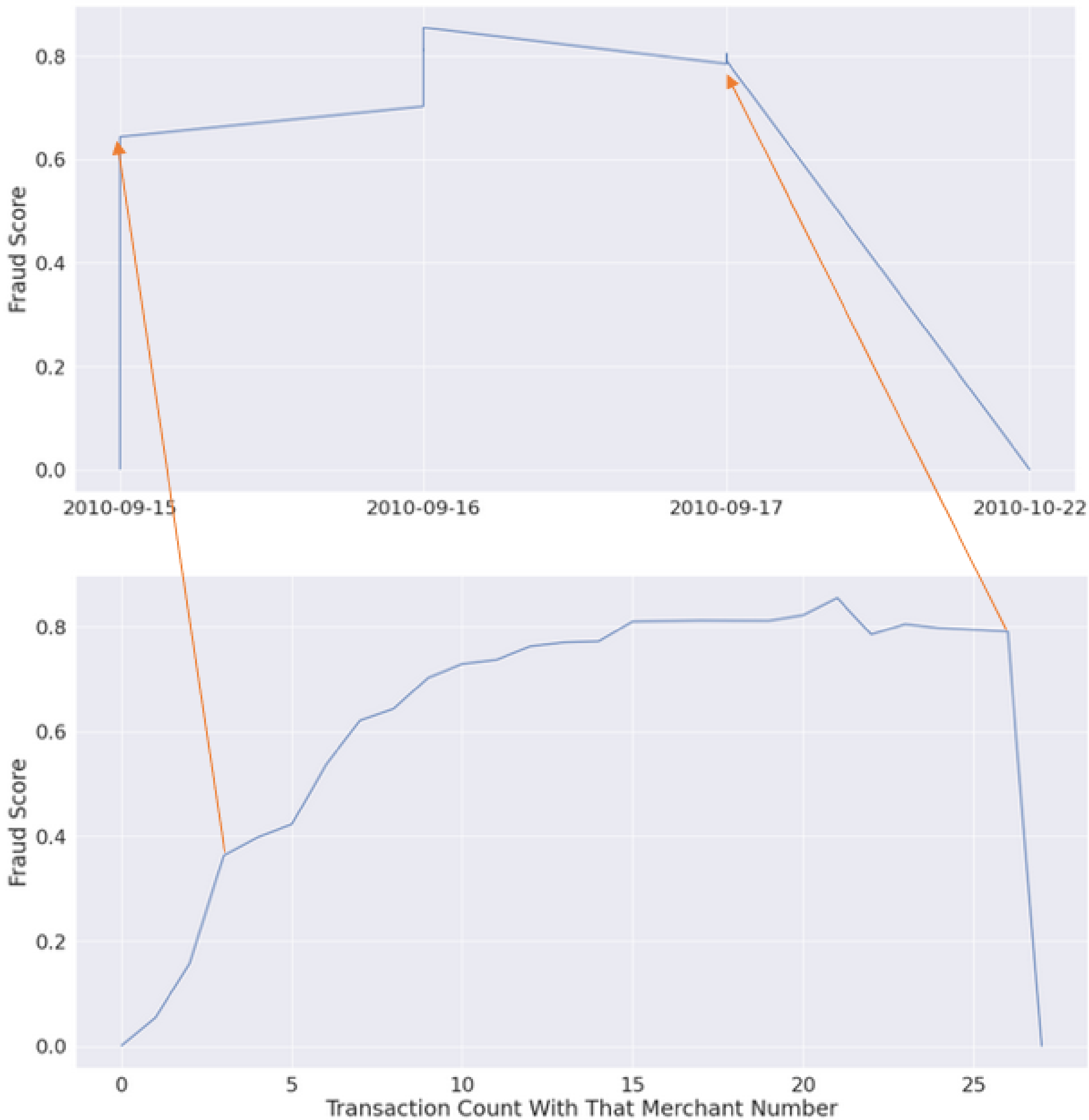
Figure 6.2.1 Card Number #5142235211 Fraud Example



Interpretation:

Card number 5142235211 has 36 recorded transactions over 83 days (2 months and 22 days). 11 transactions occurred between the days 11/25 and 11/26. Almost a third of all transactions registered to this card number occur across two days, making the fraud score grow rapidly.

Figure 6.2.2 Merchant Number #6070095870009 Fraud Example



Interpretation:
Merchant number 6070095870009 has 26 recorded transactions over 37 days. 23 transactions occurred between the days 09/15 and 09/17. Almost two-thirds of all transactions registered to this merchant number occur across three days, making the fraud score to grow rapidly.

7 Summary and Conclusions

In this project, we created a machine learning model to identify fraudulent credit card applications. We first conducted Exploratory Data Analysis (EDA) to create a Data Quality Report (DQR). We then cleaned our data according to the findings in the DQR. After that, we created 733 features based on our research. Subsequently, we reduced the number of features in our data to 80 features using the Kolmogorov-Smirnov (KS) filter and Recursive Feature Elimination (RFE). Next, we built a series of models using four different machine learning algorithms: Logistic Regression, Random Forest, Gradient Boosted Trees, and Neural Network; tuning the hyperparameters of each one of these algorithms. Finally, even though all the tree-based models are a little overfit, we decided to use Random Forest, to be the algorithm of our final model based on the accuracy of the testing dataset. Our final model achieved a Fraud Detection Rate of 86.7% for the testing set and 63.0% for the out-of-time validation set at 3% of the population.

Every project has a time constraint. Given additional time and computational resources, we have few additional suggestions of improvement that we would recommend for future attempts to accurately classify fraud in credit card transaction data. In the feature selection process, we recommend using a smaller filter to allow testing of more features against the wrapper method. Additionally, we would recommend testing two other wrapper methods, backward elimination, and bi-directional elimination. Having three sets of wrapper selected features to run through fraud algorithms would give us a better idea of important features and their roles in our models. With additional time, we would have liked to test a larger variety and combinations of values of hyperparameters in tuning our models in hopes to limit the overfitting that occurred in many of our fraud algorithms. We believe that in the future we would limit overfitting and increase our accuracy through testing an ensemble and stacking the optimal models. Most importantly for the accuracy of the algorithms, given more years of credit card transaction data, we would have had a broader set of data to train, test, and validate our models, marginally improving our predictive accuracy.

8 Appendix

8.1 Data Quality Report

Description:

Dataset Name: Card Transactions

Dataset Purpose: Record of credit card transactions of a U.S. government organization over 12 months.

Data Source: U.S. Government

Time Period: January 1, 2010 – December 31, 2010

Number of Fields: 10

Number of Records: 96,753

Summary Table:

Numeric Fields:

Column Name	# of Records	% Populated	# Unique Values	# of Zeros	Mean	Std. Dev.	Min.	Max.
Amount	96753	100.00	34909	0	427.89	10006.14	0.01	3102045.53

Categorical Fields:

Column Name	# of Records	% Populated	# Unique Values	Most Common Field
Recnum	96753	100.00	96753	N/A
Cardnum	96753	100.00	1645	5142148452
Date	96753	100.00	365	2/28/10
Merchnum	93378	96.51	13091	930090121224
Merch description	96753	100.00	13126	GSA-FSS-ADV
Merch state	95558	98.76	227	TN
Merch zip	92097	95.19	4567	38118.0

Transtype	**96753**	**100.00**	**4**	**P**
Fraud	**96753**	**100.00**	**2**	**0**

Data Field Exploration:

<u>Field 1</u>

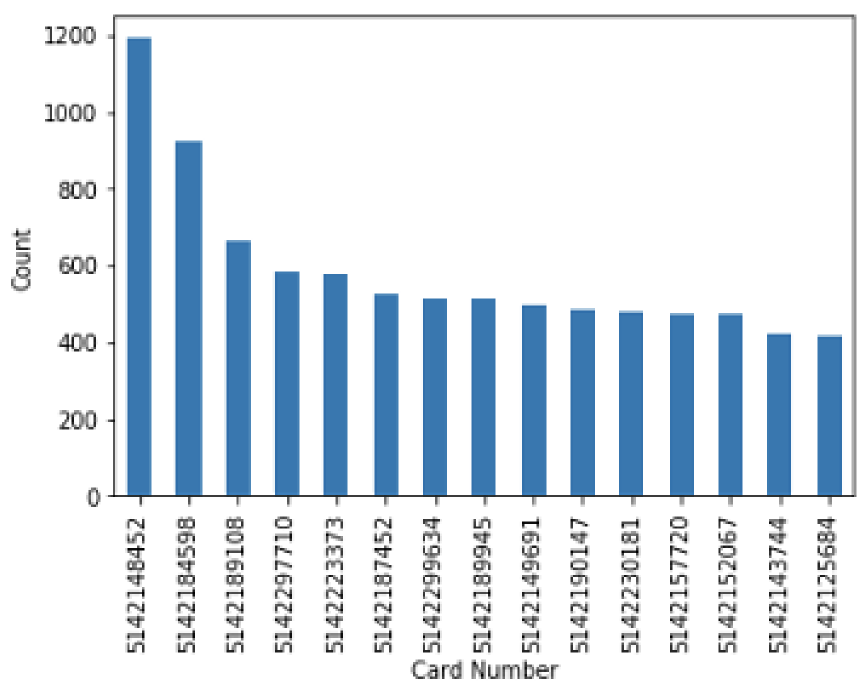
Name: Recnum

Description: Identifier for each data entry.

<u>Field 2</u>

Name: Cardnum

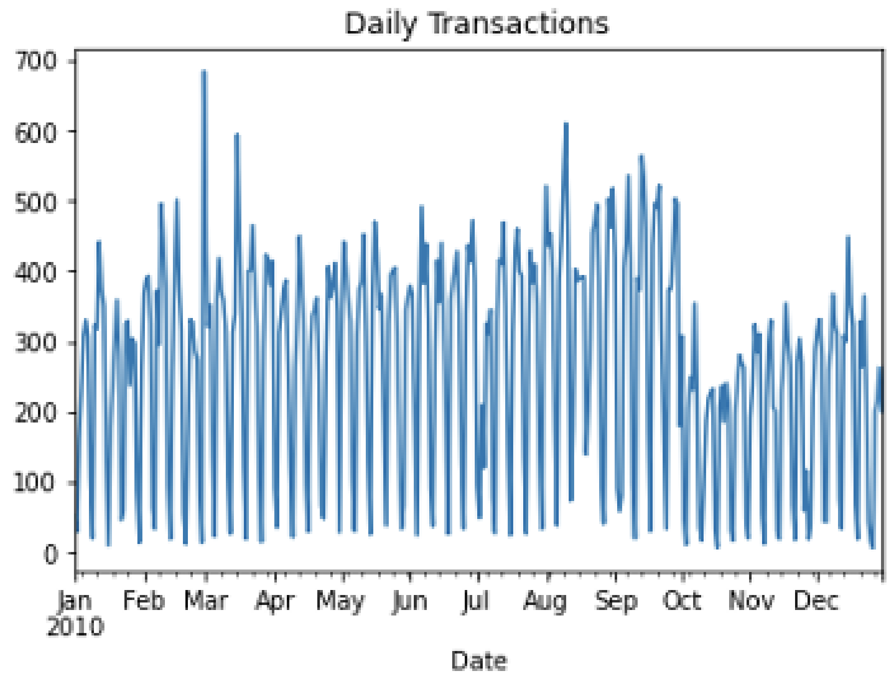
Description: Card number.

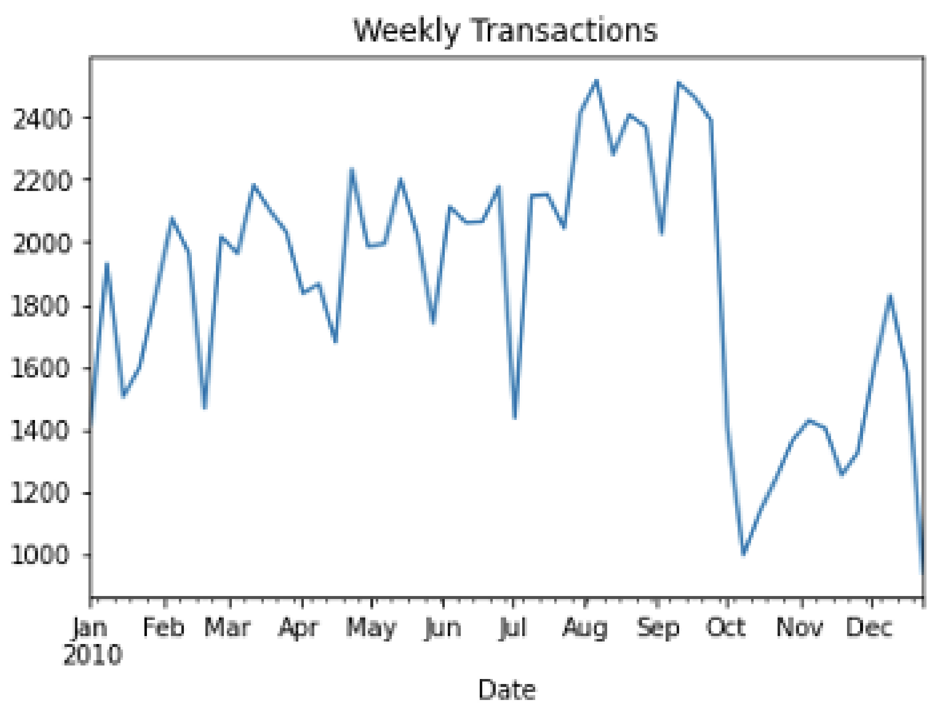


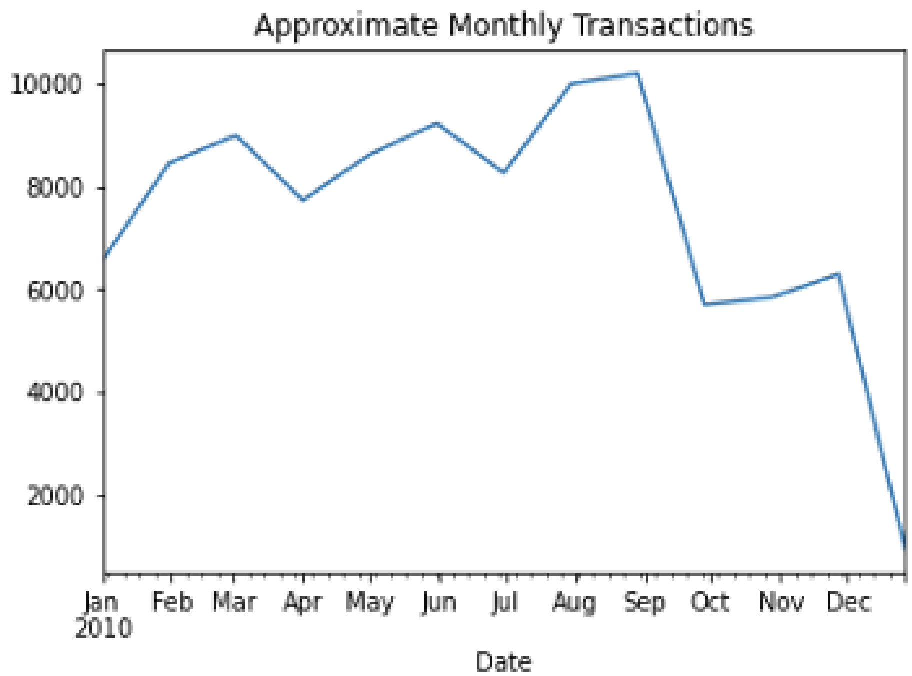
<u>Field 3</u>

Name: Date

Description: Date of transaction.



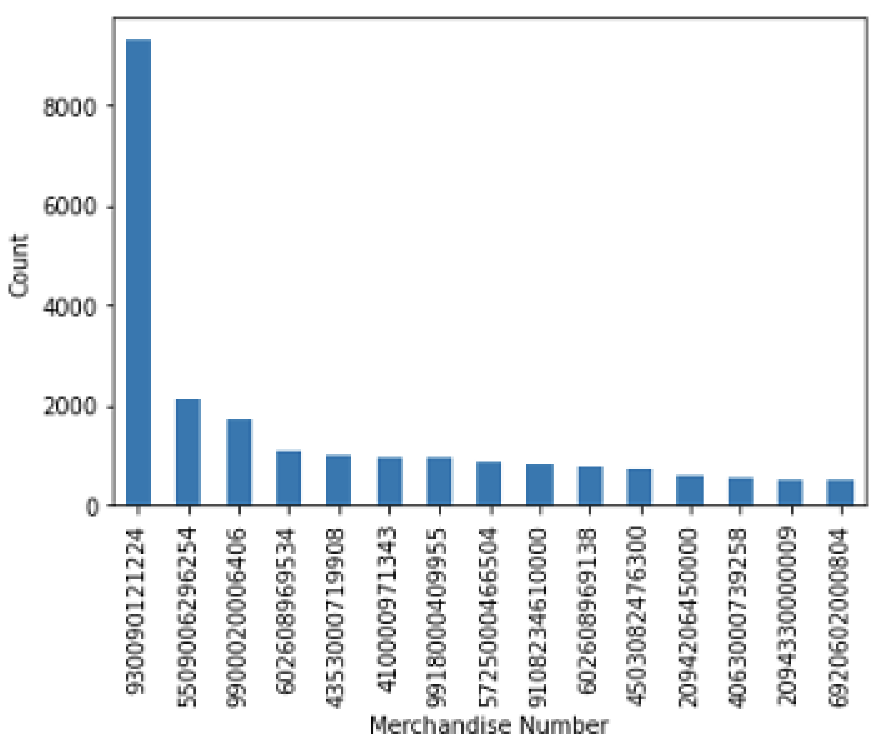




Field 4

Name: Merchnum

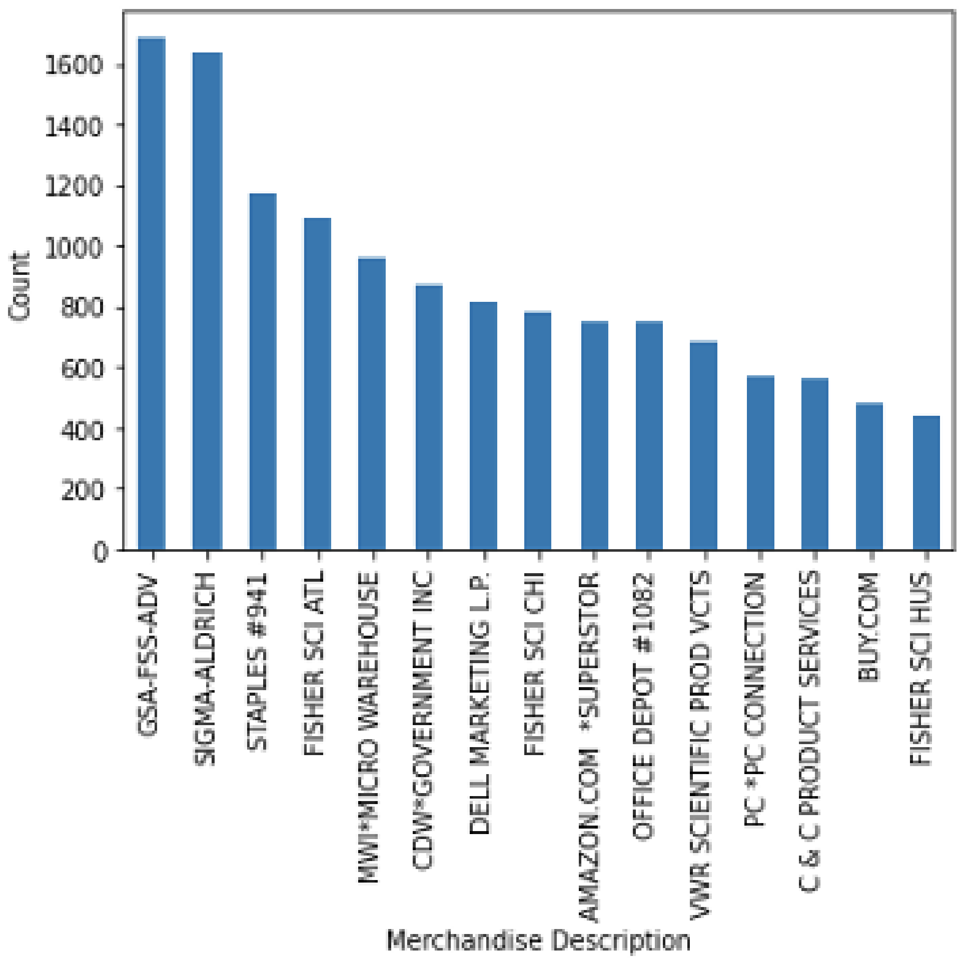
Description: Identifier for merchandise item.



Field 5

Name: Merch description.

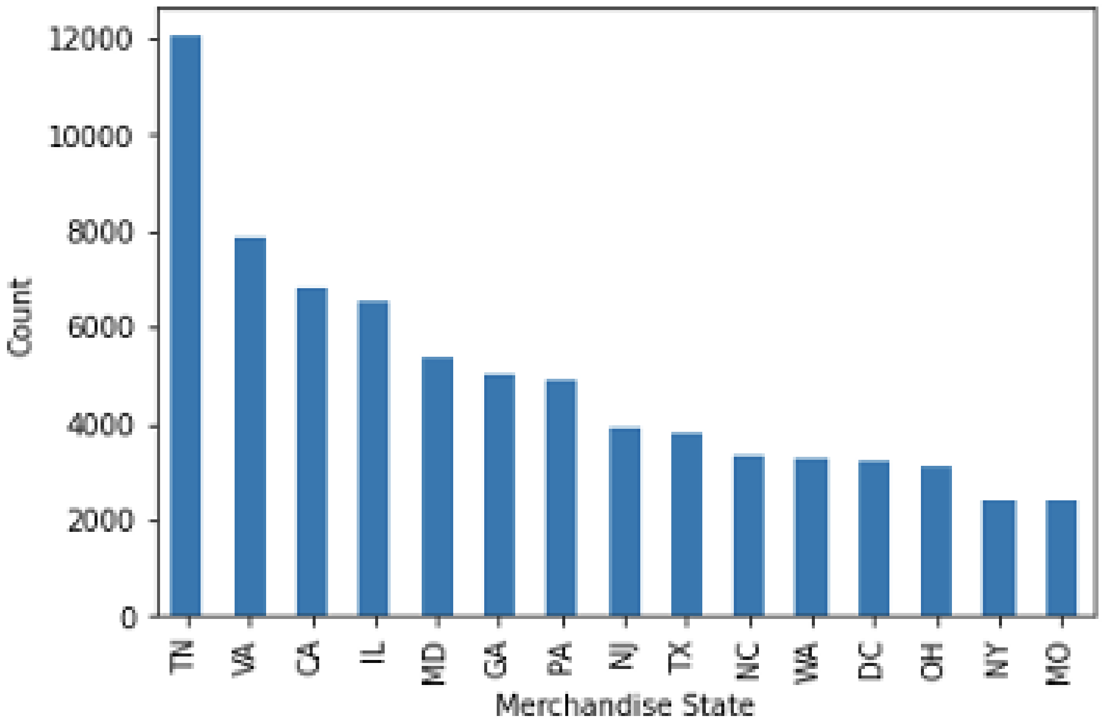
Description: Name and location identifier of the business of purchase.



<u>Field 6</u>

Name: Merch state

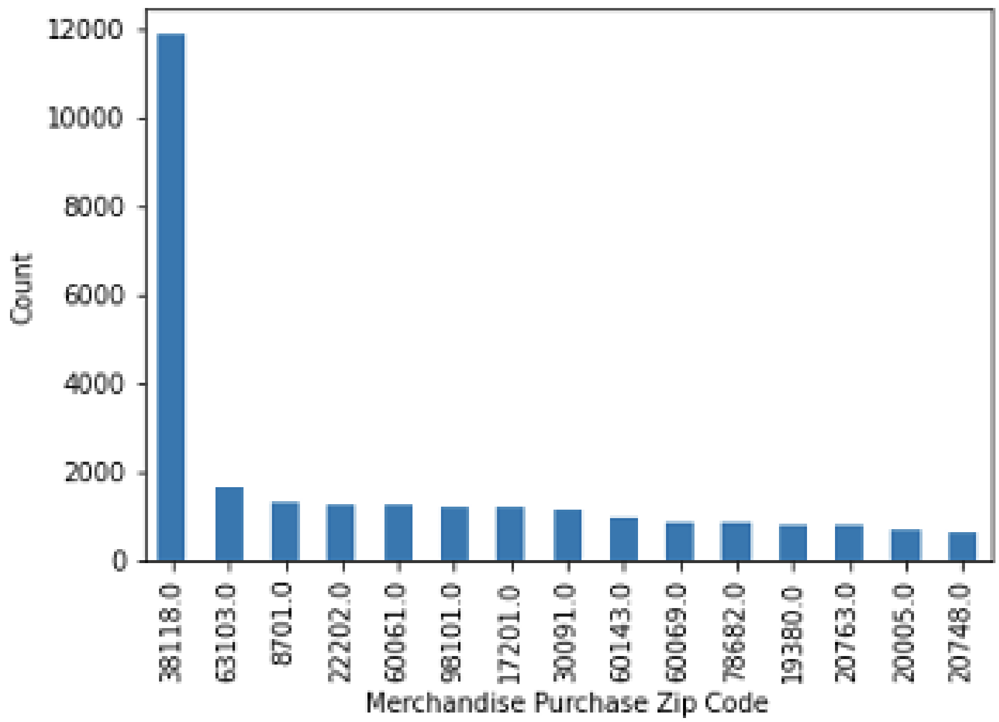
Description: State or country of transaction.



<u>Field 7</u>

Name: Merch zip

Description: Zipcode in which a transaction was made.

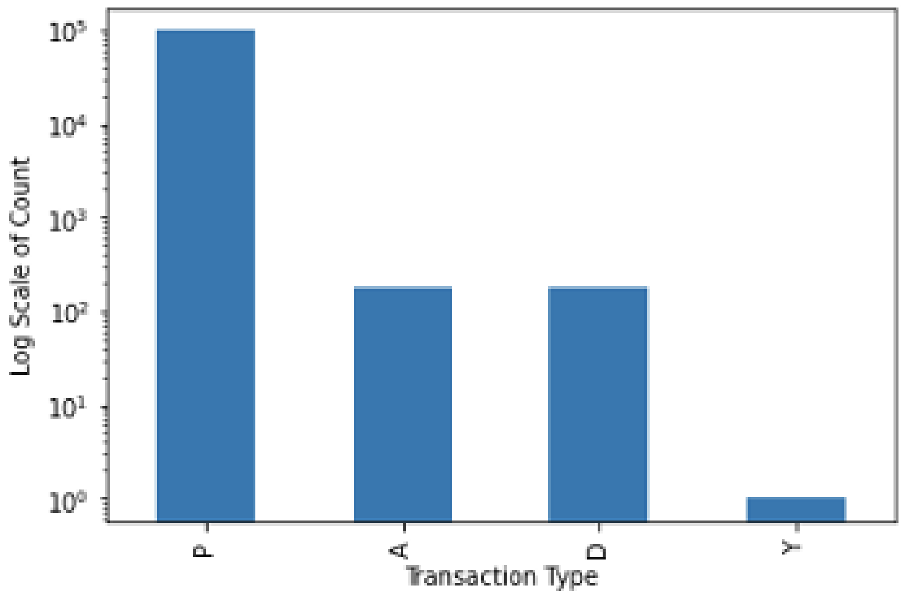


<u>Field 8</u>

Name: Transtype

Description: Classification of transaction type.

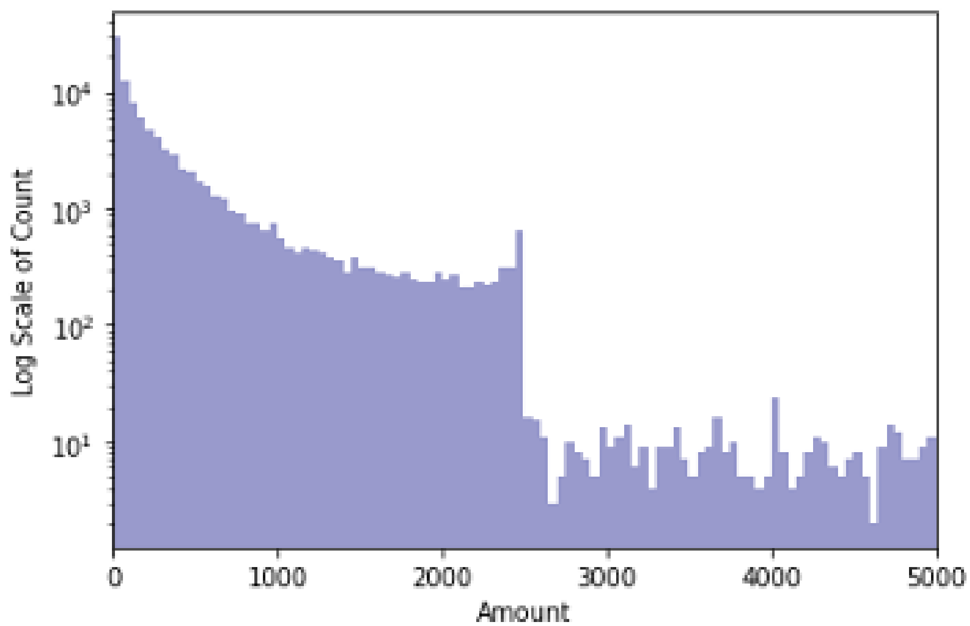
‘P’	‘D’	‘A’	‘Y’
Purchase	**Delayed Capture**	**Unknown**	**Unknown**



<u>Field 9</u>

Name: Amount

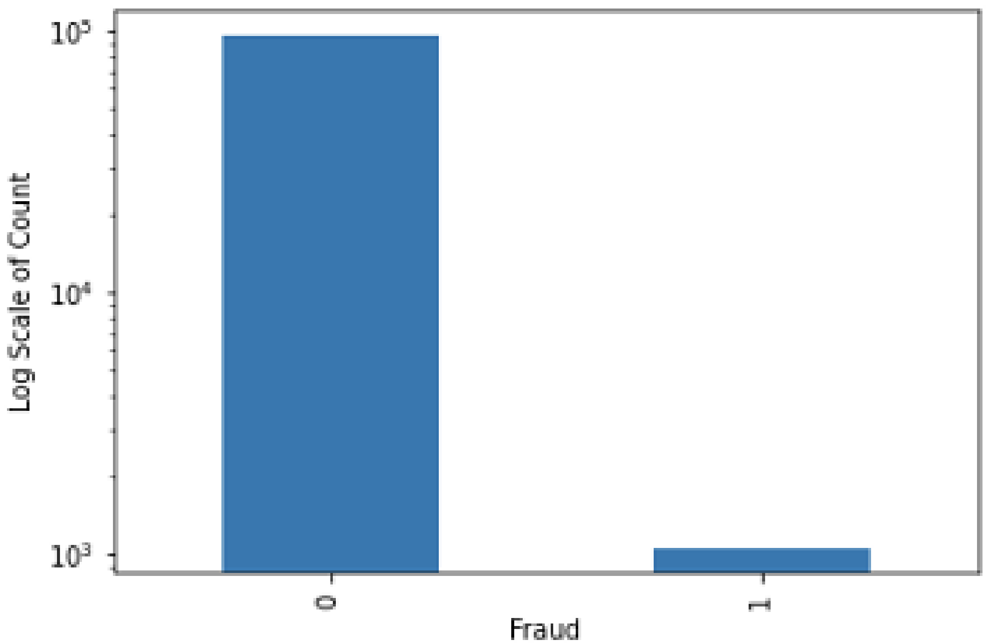
Description: Cost of transaction.
Exclude outliers > 5000
Data in histogram is 99.68% populated.



<u>Field 10</u>

Name: Fraud

Description: Binary classification of whether a record is fraudulent: 1 = fraud, 0 = no fraud.



8.2 Full List of Candidate Variables

	count	mean	std	min	max
Recnum	96397	48365.48	27945.00	1.00	96753.00
Amount	96397	395.86	832.33	0.01	47900.00
Fraud	96397	0.01	0.10	0.00	1.00
Cardnum_day_since	96397	10.74	48.22	0.00	365.00
Merchnum_day_since	96397	60.45	124.82	0.00	365.00
card_merchnum_day_since	96397	165.58	170.04	0.00	365.00
Cardnum_count_0	96397	2.47	6.00	1.00	146.00
Cardnum_avg_0	96397	393.56	726.85	0.01	28392.84
Cardnum_max_0	96397	498.21	1030.96	0.01	47900.00
Cardnum_med_0	96397	381.35	718.73	0.01	28392.84
Cardnum_total_0	96397	741.65	3431.45	0.01	218301.83
Cardnum_actual/avg_0	96397	1.00	0.44	0.00	23.79
Cardnum_actual/max_0	96397	0.88	0.28	0.00	1.00
Cardnum_actual/med_0	96397	1.41	9.90	0.00	657.89
Cardnum_actual/toal_0	96397	0.77	0.35	0.00	1.00
Cardnum_count_1	96397	3.37	7.94	1.00	177.00
Cardnum_avg_1	96397	395.45	675.83	0.01	28392.84
Cardnum_max_1	96397	610.87	1212.86	0.01	47900.00
Cardnum_med_1	96397	364.22	658.72	0.01	28392.84
Cardnum_total_1	96397	1110.05	5669.43	0.01	307468.06
Cardnum_actual/avg_1	96397	1.00	0.64	0.00	23.79
Cardnum_actual/max_1	96397	0.77	0.36	0.00	1.00
Cardnum_actual/med_1	96397	1.83	13.35	0.00	674.76
Cardnum_actual/toal_1	96397	0.64	0.39	0.00	1.00
Cardnum_count_3	96397	4.79	11.45	1.00	251.00
Cardnum_avg_3	96397	395.85	629.37	0.01	28392.84
Cardnum_max_3	96397	739.79	1367.58	0.01	47900.00
Cardnum_med_3	96397	341.36	592.65	0.01	28392.84
Cardnum_total_3	96397	1512.93	6115.51	0.01	310843.06
Cardnum_actual/avg_3	96397	1.00	0.84	0.00	38.00
Cardnum_actual/max_3	96397	0.67	0.40	0.00	1.00
Cardnum_actual/med_3	96397	2.34	17.12	0.00	1570.86
Cardnum_actual/toal_3	96397	0.51	0.40	0.00	1.00
Cardnum_count_7	96397	7.63	16.61	1.00	369.00
Cardnum_avg_7	96397	397.19	560.08	0.14	25500.00

Cardnum_max_7	96397	960.43	1603.13	0.14	47900.00
Cardnum_med_7	96397	307.16	501.73	0.14	25500.00
Cardnum_total_7	96397	2384.04	7158.50	0.14	312616.06
Cardnum_actual/avg_7	96397	0.99	1.08	0.00	59.87
Cardnum_actual/max_7	96397	0.54	0.41	0.00	1.00
Cardnum_actual/med_7	96397	2.99	27.09	0.00	5747.54
Cardnum_actual/toal_7	96397	0.36	0.37	0.00	1.00
Cardnum_count_14	96397	11.80	20.72	1.00	380.00
Cardnum_avg_14	96397	396.99	522.92	0.14	25500.00
Cardnum_max_14	96397	1188.64	1829.50	0.14	47900.00
Cardnum_med_14	96397	279.02	456.12	0.14	25500.00
Cardnum_total_14	96397	3768.18	9421.92	0.14	313995.06
Cardnum_actual/avg_14	96397	1.00	1.27	0.00	71.33
Cardnum_actual/max_14	96397	0.44	0.40	0.00	1.00
Cardnum_actual/med_14	96397	3.47	29.25	0.00	6145.64
Cardnum_actual/toal_14	96397	0.25	0.32	0.00	1.00
Cardnum_count_30	96397	20.36	30.91	1.00	426.00
Cardnum_avg_30	96397	396.58	479.34	0.17	25500.00
Cardnum_max_30	96397	1482.17	2076.88	0.17	47900.00
Cardnum_med_30	96397	250.94	402.43	0.17	25500.00
Cardnum_total_30	96397	6675.65	14591.23	0.17	353997.29
Cardnum_actual/avg_30	96397	1.01	1.61	0.00	137.99
Cardnum_actual/max_30	96397	0.34	0.37	0.00	1.00
Cardnum_actual/med_30	96397	3.95	32.61	0.00	6288.56
Cardnum_actual/toal_30	96397	0.16	0.25	0.00	1.00
Cardnum_count_180	96397	72.14	89.69	1.00	765.00
Cardnum_avg_180	96397	388.94	386.53	0.17	9150.00
Cardnum_max_180	96397	2112.84	2547.87	0.17	47900.00
Cardnum_med_180	96397	209.45	268.83	0.17	9150.00
Cardnum_total_180	96397	25872.39	51814.89	0.17	877817.47
Cardnum_actual/avg_180	96397	1.04	2.01	0.00	173.07
Cardnum_actual/max_180	96397	0.24	0.31	0.00	1.00
Cardnum_actual/med_180	96397	4.82	41.55	0.00	6247.05
Cardnum_actual/toal_180	96397	0.07	0.17	0.00	1.00
Merchnum_count_0	96397	6.82	18.99	1.00	260.00
Merchnum_avg_0	96397	395.41	759.77	0.01	28392.84
Merchnum_max_0	96397	503.99	1011.19	0.01	47900.00
Merchnum_med_0	96397	381.15	750.17	0.01	28392.84
Merchnum_total_0	96397	785.36	2851.61	0.01	217467.18

Merchnum_actual/avg_0	96397	1.00	0.63	0.00	37.96
Merchnum_actual/max_0	96397	0.81	0.34	0.00	1.00
Merchnum_actual/med_0	96397	1.26	3.46	0.00	405.45
Merchnum_actual/toal_0	96397	0.72	0.39	0.00	1.00
Merchnum_count_1	96397	11.60	31.62	1.00	327.00
Merchnum_avg_1	96397	397.31	748.61	0.01	28392.84
Merchnum_max_1	96397	589.87	1195.43	0.01	47900.00
Merchnum_med_1	96397	371.13	737.27	0.01	28392.84
Merchnum_total_1	96397	1139.30	4386.47	0.01	306633.41
Merchnum_actual/avg_1	96397	1.00	0.77	0.00	43.43
Merchnum_actual/max_1	96397	0.74	0.38	0.00	1.00
Merchnum_actual/med_1	96397	1.38	4.07	0.00	405.45
Merchnum_actual/toal_1	96397	0.63	0.42	0.00	1.00
Merchnum_count_3	96397	21.07	55.18	1.00	466.00
Merchnum_avg_3	96397	397.47	727.71	0.01	28392.84
Merchnum_max_3	96397	677.12	1273.37	0.01	47900.00
Merchnum_med_3	96397	361.47	715.84	0.01	28392.84
Merchnum_total_3	96397	1609.92	5194.94	0.01	307302.58
Merchnum_actual/avg_3	96397	1.00	0.93	0.00	64.09
Merchnum_actual/max_3	96397	0.69	0.41	0.00	1.00
Merchnum_actual/med_3	96397	1.47	4.87	0.00	467.83
Merchnum_actual/toal_3	96397	0.57	0.43	0.00	1.00
Merchnum_count_7	96397	42.25	106.14	1.00	762.00
Merchnum_avg_7	96397	396.78	691.00	0.01	28392.84
Merchnum_max_7	96397	834.31	1470.21	0.01	47900.00
Merchnum_med_7	96397	346.25	679.25	0.01	28392.84
Merchnum_total_7	96397	2680.67	6411.41	0.01	313984.55
Merchnum_actual/avg_7	96397	1.00	1.10	0.00	82.46
Merchnum_actual/max_7	96397	0.61	0.42	0.00	1.00
Merchnum_actual/med_7	96397	1.57	4.77	0.00	473.98
Merchnum_actual/toal_7	96397	0.48	0.43	0.00	1.00
Merchnum_count_14	96397	76.23	191.92	1.00	1091.00
Merchnum_avg_14	96397	398.26	664.44	0.01	28392.84
Merchnum_max_14	96397	1009.00	1773.78	0.01	47900.00
Merchnum_med_14	96397	335.70	646.18	0.01	28392.84
Merchnum_total_14	96397	4438.78	8948.99	0.01	319334.68
Merchnum_actual/avg_14	96397	1.00	1.27	0.00	133.53
Merchnum_actual/max_14	96397	0.55	0.43	0.00	1.00
Merchnum_actual/med_14	96397	1.62	4.72	0.00	473.98

Merchnum_actual/toal_14	96397	0.41	0.42	0.00	1.00
Merchnum_count_30	96397	146.95	376.40	1.00	1828.00
Merchnum_avg_30	96397	397.71	623.61	0.01	28392.84
Merchnum_max_30	96397	1266.09	2309.17	0.01	47900.00
Merchnum_med_30	96397	321.78	592.68	0.01	28392.84
Merchnum_total_30	96397	8155.63	14475.98	0.01	320373.00
Merchnum_actual/avg_30	96397	1.00	1.45	0.00	172.64
Merchnum_actual/max_30	96397	0.49	0.43	0.00	1.00
Merchnum_actual/med_30	96397	1.68	4.55	0.00	481.59
Merchnum_actual/toal_30	96397	0.34	0.40	0.00	1.00
Merchnum_count_180	96397	520.11	1520.38	1.00	7852.00
Merchnum_avg_180	96397	391.47	553.95	0.01	28392.84
Merchnum_max_180	96397	2077.20	4434.99	0.01	47900.00
Merchnum_med_180	96397	296.39	522.96	0.01	28392.84
Merchnum_total_180	96397	31096.40	59357.03	0.01	413457.84
Merchnum_actual/avg_180	96397	1.02	1.74	0.00	177.99
Merchnum_actual/max_180	96397	0.39	0.41	0.00	1.00
Merchnum_actual/med_180	96397	1.84	5.13	0.00	473.98
Merchnum_actual/toal_180	96397	0.23	0.36	0.00	1.00
card_merchnum_count_0	96397	2.10	5.91	1.00	145.00
card_merchnum_avg_0	96397	395.81	796.95	0.01	28392.84
card_merchnum_max_0	96397	421.28	935.81	0.01	47900.00
card_merchnum_med_0	96397	393.47	790.66	0.01	28392.84
card_merchnum_total_0	96397	528.81	2621.86	0.01	217467.18
card_merchnum_actual/avg_0	96397	1.00	0.22	0.00	20.24
card_merchnum_actual/max_0	96397	0.96	0.17	0.00	1.00
card_merchnum_actual/med_0	96397	1.03	0.68	0.00	100.00
card_merchnum_actual/toal_0	96397	0.88	0.27	0.00	1.00
card_merchnum_count_1	96397	2.42	7.59	1.00	177.00
card_merchnum_avg_1	96397	397.20	799.53	0.01	28392.84
card_merchnum_max_1	96397	432.52	1010.27	0.01	47900.00
card_merchnum_med_1	96397	394.45	794.04	0.01	28392.84
card_merchnum_total_1	96397	599.36	4020.30	0.01	306633.41
card_merchnum_actual/avg_1	96397	1.00	0.25	0.00	20.24
card_merchnum_actual/max_1	96397	0.94	0.19	0.00	1.00
card_merchnum_actual/med_1	96397	1.03	0.71	0.00	71.11
card_merchnum_actual/toal_1	96397	0.86	0.30	0.00	1.00
card_merchnum_count_3	96397	3.03	10.98	1.00	248.00
card_merchnum_avg_3	96397	398.12	797.32	0.01	28392.84

card_merchnum_max_3	96397	441.33	1014.58	0.01	47900.00
card_merchnum_med_3	96397	394.69	792.15	0.01	28392.84
card_merchnum_total_3	96397	630.90	4062.89	0.01	306633.41
card_merchnum_actual/avg_3	96397	0.99	0.30	0.00	20.24
card_merchnum_actual/max_3	96397	0.92	0.22	0.00	1.00
card_merchnum_actual/med_3	96397	1.05	1.50	0.00	301.10
card_merchnum_actual/toal_3	96397	0.83	0.32	0.00	1.00
card_merchnum_count_7	96397	4.05	15.65	1.00	358.00
card_merchnum_avg_7	96397	399.87	792.83	0.01	28392.84
card_merchnum_max_7	96397	458.87	1022.82	0.01	47900.00
card_merchnum_med_7	96397	394.72	788.12	0.01	28392.84
card_merchnum_total_7	96397	689.92	4103.73	0.01	306633.41
card_merchnum_actual/avg_7	96397	0.99	0.36	0.00	20.24
card_merchnum_actual/max_7	96397	0.89	0.26	0.00	1.00
card_merchnum_actual/med_7	96397	1.08	2.20	0.00	442.87
card_merchnum_actual/toal_7	96397	0.78	0.36	0.00	1.00
card_merchnum_count_14	96397	5.34	19.04	1.00	369.00
card_merchnum_avg_14	96397	401.85	789.99	0.01	28392.84
card_merchnum_max_14	96397	480.16	1043.59	0.01	47900.00
card_merchnum_med_14	96397	394.71	786.15	0.01	28392.84
card_merchnum_total_14	96397	770.54	4168.44	0.01	306633.41
card_merchnum_actual/avg_14	96397	0.99	0.43	0.00	23.11
card_merchnum_actual/max_14	96397	0.85	0.29	0.00	1.00
card_merchnum_actual/med_14	96397	1.11	2.28	0.00	400.00
card_merchnum_actual/toal_14	96397	0.73	0.38	0.00	1.00
card_merchnum_count_30	96397	7.73	27.43	1.00	409.00
card_merchnum_avg_30	96397	403.95	784.29	0.01	28392.84
card_merchnum_max_30	96397	512.69	1066.25	0.01	47900.00
card_merchnum_med_30	96397	393.32	783.56	0.01	28392.84
card_merchnum_total_30	96397	924.13	4299.65	0.01	306633.41
card_merchnum_actual/avg_30	96397	0.98	0.51	0.00	25.03
card_merchnum_actual/max_30	96397	0.81	0.33	0.00	1.00
card_merchnum_actual/med_30	96397	1.15	2.28	0.00	397.86
card_merchnum_actual/toal_30	96397	0.68	0.40	0.00	1.00
card_merchnum_count_180	96397	15.77	52.95	1.00	606.00
card_merchnum_avg_180	96397	404.75	771.93	0.01	28392.84
card_merchnum_max_180	96397	607.95	1151.66	0.01	47900.00
card_merchnum_med_180	96397	382.36	767.26	0.01	28392.84
card_merchnum_total_180	96397	1664.55	5524.75	0.01	306633.41

card_merchnum_actual/avg_180	96397	0.99	0.71	0.00	71.83
card_merchnum_actual/max_180	96397	0.73	0.37	0.00	1.00
card_merchnum_actual/med_180	96397	1.26	2.70	0.00	405.45
card_merchnum_actual/toal_180	96397	0.56	0.42	0.00	1.00
Cardnum_count_0_by_7	96397	0.07	0.05	0.00	0.14
Cardnum_count_0_by_14	96397	0.02	0.02	0.00	0.07
Cardnum_count_0_by_30	96397	0.01	0.01	0.00	0.03

Cardnum_count_1_by_7	96397	0.08	0.04	0.00	0.14
Cardnum_count_1_by_14	96397	0.03	0.02	0.00	0.07
Cardnum_count_1_by_30	96397	0.01	0.01	0.00	0.03
Cardnum_count_3_by_7	96397	0.10	0.04	0.00	0.14
Cardnum_count_3_by_14	96397	0.04	0.02	0.00	0.07
Cardnum_count_3_by_30	96397	0.01	0.01	0.00	0.03
Merchnum_count_0_by_7	96397	0.08	0.05	0.00	0.14
Merchnum_count_0_by_14	96397	0.03	0.03	0.00	0.07
Merchnum_count_0_by_30	96397	0.01	0.01	0.00	0.03
Merchnum_count_1_by_7	96397	0.09	0.05	0.00	0.14
Merchnum_count_1_by_14	96397	0.04	0.03	0.00	0.07
Merchnum_count_1_by_30	96397	0.01	0.01	0.00	0.03
Merchnum_count_3_by_7	96397	0.11	0.04	0.00	0.14
Merchnum_count_3_by_14	96397	0.04	0.02	0.00	0.07
Merchnum_count_3_by_30	96397	0.02	0.01	0.00	0.03
card_merchnum_count_0_by_7	96397	0.12	0.04	0.00	0.14
card_merchnum_count_0_by_14	96397	0.06	0.02	0.00	0.07
card_merchnum_count_0_by_30	96397	0.02	0.01	0.00	0.03
card_merchnum_count_1_by_7	96397	0.13	0.03	0.00	0.14
card_merchnum_count_1_by_14	96397	0.06	0.02	0.00	0.07
card_merchnum_count_1_by_30	96397	0.03	0.01	0.00	0.03
card_merchnum_count_3_by_7	96397	0.13	0.03	0.00	0.14
card_merchnum_count_3_by_14	96397	0.06	0.02	0.00	0.07
card_merchnum_count_3_by_30	96397	0.03	0.01	0.00	0.03
card_merchdesc_day_since	96397	177.02	170.75	0.00	365.00
card_state_day_since	96397	77.00	132.30	0.00	365.00
card_zip_day_since	96397	143.56	164.89	0.00	365.00
card_merchdesc_count_0	96397	1.70	3.16	1.00	85.00
card_merchdesc_avg_0	96397	395.69	795.36	0.01	28392.84
card_merchdesc_max_0	96397	421.19	935.85	0.01	47900.00
card_merchdesc_med_0	96397	393.35	788.97	0.01	28392.84

card_merchdesc_total_0	96397	526.06	2618.50	0.01	217467.18
card_merchdesc_actual/avg_0	96397	1.00	0.18	0.00	7.64
card_merchdesc_actual/max_0	96397	0.96	0.15	0.00	1.00
card_merchdesc_actual/med_0	96397	1.02	0.60	0.00	100.00
card_merchdesc_actual/toal_0	96397	0.90	0.26	0.00	1.00
card_merchdesc_count_1	96397	1.81	3.41	1.00	85.00
card_merchdesc_avg_1	96397	397.21	798.08	0.01	28392.84
card_merchdesc_max_1	96397	432.63	1010.33	0.01	47900.00
card_merchdesc_med_1	96397	394.44	792.43	0.01	28392.84
card_merchdesc_total_1	96397	595.01	4010.95	0.01	306633.41
card_merchdesc_actual/avg_1	96397	1.00	0.22	0.00	7.64
card_merchdesc_actual/max_1	96397	0.95	0.18	0.00	1.00
card_merchdesc_actual/med_1	96397	1.02	0.55	0.00	50.91
card_merchdesc_actual/toal_1	96397	0.88	0.28	0.00	1.00
card_merchdesc_count_3	96397	1.92	3.57	1.00	85.00
card_merchdesc_avg_3	96397	398.10	795.23	0.01	28392.84
card_merchdesc_max_3	96397	441.72	1014.99	0.01	47900.00
card_merchdesc_med_3	96397	394.58	789.94	0.01	28392.84
card_merchdesc_total_3	96397	625.05	4054.24	0.01	306633.41
card_merchdesc_actual/avg_3	96397	0.99	0.25	0.00	7.64
card_merchdesc_actual/max_3	96397	0.94	0.20	0.00	1.00
card_merchdesc_actual/med_3	96397	1.03	1.17	0.00	301.10
card_merchdesc_actual/toal_3	96397	0.85	0.30	0.00	1.00
card_merchdesc_count_7	96397	2.10	3.84	1.00	90.00
card_merchdesc_avg_7	96397	399.91	790.52	0.01	28392.84
card_merchdesc_max_7	96397	459.60	1024.16	0.01	47900.00
card_merchdesc_med_7	96397	394.66	785.79	0.01	28392.84
card_merchdesc_total_7	96397	681.79	4097.68	0.01	306633.41
card_merchdesc_actual/avg_7	96397	0.99	0.30	0.00	7.97
card_merchdesc_actual/max_7	96397	0.91	0.24	0.00	1.00
card_merchdesc_actual/med_7	96397	1.05	1.96	0.00	442.87
card_merchdesc_actual/toal_7	96397	0.81	0.33	0.00	1.00
card_merchdesc_count_14	96397	2.30	4.11	1.00	97.00
card_merchdesc_avg_14	96397	402.02	788.09	0.01	28392.84
card_merchdesc_max_14	96397	481.42	1046.77	0.01	47900.00
card_merchdesc_med_14	96397	394.58	783.26	0.01	28392.84
card_merchdesc_total_14	96397	759.46	4167.47	0.01	306633.41
card_merchdesc_actual/avg_14	96397	0.98	0.34	0.00	8.94
card_merchdesc_actual/max_14	96397	0.89	0.26	0.00	1.00

card_merchdesc_actual/med_14	96397	1.07	2.10	0.00	400.00
card_merchdesc_actual/toal_14	96397	0.77	0.35	0.00	1.00
card_merchdesc_count_30	96397	2.70	4.77	1.00	99.00
card_merchdesc_avg_30	96397	404.20	780.77	0.01	28392.84
card_merchdesc_max_30	96397	514.94	1074.89	0.01	47900.00
card_merchdesc_med_30	96397	393.05	778.59	0.01	28392.84
card_merchdesc_total_30	96397	908.66	4314.58	0.01	306633.41
card_merchdesc_actual/avg_30	96397	0.98	0.40	0.00	11.64
card_merchdesc_actual/max_30	96397	0.85	0.30	0.00	1.00
card_merchdesc_actual/med_30	96397	1.09	1.64	0.00	225.07
card_merchdesc_actual/toal_30	96397	0.72	0.37	0.00	1.00
card_merchdesc_count_180	96397	4.85	11.32	1.00	229.00
card_merchdesc_avg_180	96397	405.26	769.52	0.01	28392.84
card_merchdesc_max_180	96397	614.20	1167.45	0.01	47900.00
card_merchdesc_med_180	96397	382.01	764.63	0.01	28392.84
card_merchdesc_total_180	96397	1683.03	5738.65	0.01	306633.41
card_merchdesc_actual/avg_180	96397	0.98	0.52	0.00	14.31
card_merchdesc_actual/max_180	96397	0.78	0.35	0.00	1.00
card_merchdesc_actual/med_180	96397	1.18	1.99	0.00	225.07
card_merchdesc_actual/toal_180	96397	0.60	0.41	0.00	1.00
card_state_count_0	96397	2.17	5.94	1.00	146.00
card_state_avg_0	96397	395.31	787.18	0.01	28392.84
card_state_max_0	96397	432.19	944.23	0.01	47900.00
card_state_med_0	96397	391.98	780.88	0.01	28392.84
card_state_total_0	96397	553.96	2640.51	0.01	217467.18
card_state_actual/avg_0	96397	1.00	0.26	0.00	20.24
card_state_actual/max_0	96397	0.94	0.19	0.00	1.00
card_state_actual/med_0	96397	1.05	1.49	0.00	234.79
card_state_actual/toal_0	96397	0.86	0.29	0.00	1.00
card_state_count_1	96397	2.59	7.82	1.00	177.00
card_state_avg_1	96397	396.84	784.45	0.01	28392.84
card_state_max_1	96397	457.41	1030.03	0.01	47900.00
card_state_med_1	96397	391.24	778.77	0.01	28392.84
card_state_total_1	96397	658.51	4053.10	0.01	306633.41
card_state_actual/avg_1	96397	1.00	0.33	0.00	20.17
card_state_actual/max_1	96397	0.91	0.24	0.00	1.00
card_state_actual/med_1	96397	1.08	1.68	0.00	231.59
card_state_actual/toal_1	96397	0.81	0.33	0.00	1.00
card_state_count_3	96397	3.33	11.28	1.00	251.00

card_state_avg_3	96397	397.52	770.86	0.01	28392.84
card_state_max_3	96397	485.58	1055.50	0.01	47900.00
card_state_med_3	96397	387.87	764.23	0.01	28392.84
card_state_total_3	96397	737.52	4119.57	0.01	306633.41
card_state_actual/avg_3	96397	0.99	0.40	0.00	20.17
card_state_actual/max_3	96397	0.87	0.28	0.00	1.00
card_state_actual/med_3	96397	1.13	2.40	0.00	301.10
card_state_actual/toal_3	96397	0.76	0.36	0.00	1.00
card_state_count_7	96397	4.67	16.29	1.00	369.00
card_state_avg_7	96397	399.64	755.05	0.01	28392.84
card_state_max_7	96397	540.12	1121.85	0.01	47900.00
card_state_med_7	96397	381.30	744.63	0.01	28392.84
card_state_total_7	96397	901.74	4244.99	0.01	306633.41
card_state_actual/avg_7	96397	0.99	0.52	0.00	32.76
card_state_actual/max_7	96397	0.80	0.33	0.00	1.00
card_state_actual/med_7	96397	1.23	3.70	0.00	442.87
card_state_actual/toal_7	96397	0.67	0.39	0.00	1.00
card_state_count_14	96397	6.45	19.95	1.00	380.00
card_state_avg_14	96397	401.48	734.47	0.01	28392.84
card_state_max_14	96397	604.53	1190.46	0.01	47900.00
card_state_med_14	96397	373.11	726.08	0.01	28392.84
card_state_total_14	96397	1149.62	4505.68	0.01	306633.41
card_state_actual/avg_14	96397	0.99	0.63	0.00	32.76
card_state_actual/max_14	96397	0.74	0.37	0.00	1.00
card_state_actual/med_14	96397	1.34	4.21	0.00	416.25
card_state_actual/toal_14	96397	0.59	0.41	0.00	1.00
card_state_count_30	96397	9.88	28.91	1.00	425.00
card_state_avg_30	96397	402.69	692.18	0.01	28392.84
card_state_max_30	96397	708.86	1298.44	0.01	47900.00
card_state_med_30	96397	358.42	677.89	0.01	28392.84
card_state_total_30	96397	1664.80	5178.68	0.01	306633.41
card_state_actual/avg_30	96397	0.99	0.75	0.00	32.76
card_state_actual/max_30	96397	0.66	0.39	0.00	1.00
card_state_actual/med_30	96397	1.48	4.91	0.00	452.86
card_state_actual/toal_30	96397	0.49	0.41	0.00	1.00
card_state_count_180	96397	26.90	69.85	1.00	756.00
card_state_avg_180	96397	399.41	625.16	0.17	28392.84
card_state_max_180	96397	1019.44	1621.16	0.17	47900.00
card_state_med_180	96397	318.80	586.27	0.17	28392.84

card_state_total_180	96397	4842.42	12110.80	0.17	306633.41
card_state_actual/avg_180	96397	1.01	1.09	0.00	72.13
card_state_actual/max_180	96397	0.51	0.41	0.00	1.00
card_state_actual/med_180	96397	1.76	5.42	0.00	490.77
card_state_actual/toal_180	96397	0.31	0.37	0.00	1.00
card_zip_count_0	96397	2.12	5.94	1.00	146.00
card_zip_avg_0	96397	395.66	794.79	0.01	28392.84
card_zip_max_0	96397	422.72	936.57	0.01	47900.00
card_zip_med_0	96397	393.14	788.41	0.01	28392.84
card_zip_total_0	96397	531.72	2623.00	0.01	217467.18
card_zip_actual/avg_0	96397	1.00	0.23	0.00	20.24
card_zip_actual/max_0	96397	0.95	0.17	0.00	1.00
card_zip_actual/med_0	96397	1.03	1.08	0.00	234.79
card_zip_actual/toal_0	96397	0.88	0.28	0.00	1.00
card_zip_count_1	96397	2.48	7.82	1.00	177.00
card_zip_avg_1	96397	397.22	797.21	0.01	28392.84
card_zip_max_1	96397	435.42	1012.03	0.01	47900.00
card_zip_med_1	96397	394.19	791.65	0.01	28392.84
card_zip_total_1	96397	605.63	4022.54	0.01	306633.41
card_zip_actual/avg_1	96397	1.00	0.28	0.00	20.17
card_zip_actual/max_1	96397	0.93	0.21	0.00	1.00
card_zip_actual/med_1	96397	1.04	1.14	0.00	231.59
card_zip_actual/toal_1	96397	0.85	0.31	0.00	1.00
card_zip_count_3	96397	3.13	11.28	1.00	251.00
card_zip_avg_3	96397	398.12	794.03	0.01	28392.84
card_zip_max_3	96397	446.31	1018.20	0.01	47900.00
card_zip_med_3	96397	394.12	788.91	0.01	28392.84
card_zip_total_3	96397	641.80	4066.67	0.01	306633.41
card_zip_actual/avg_3	96397	0.99	0.32	0.00	20.17
card_zip_actual/max_3	96397	0.91	0.24	0.00	1.00
card_zip_actual/med_3	96397	1.07	1.85	0.00	301.10
card_zip_actual/toal_3	96397	0.81	0.33	0.00	1.00
card_zip_count_7	96397	4.25	16.31	1.00	369.00
card_zip_avg_7	96397	400.03	787.90	0.01	28392.84
card_zip_max_7	96397	467.72	1029.23	0.01	47900.00
card_zip_med_7	96397	393.80	783.35	0.01	28392.84
card_zip_total_7	96397	710.68	4112.29	0.01	306633.41
card_zip_actual/avg_7	96397	0.99	0.40	0.00	32.76
card_zip_actual/max_7	96397	0.87	0.28	0.00	1.00

card_zip_actual/med_7	96397	1.11	2.84	0.00	442.87
card_zip_actual/toal_7	96397	0.76	0.37	0.00	1.00
card_zip_count_14	96397	5.70	19.98	1.00	380.00
card_zip_avg_14	96397	402.28	785.62	0.01	28392.84
card_zip_max_14	96397	493.82	1055.42	0.01	47900.00
card_zip_med_14	96397	393.24	780.80	0.01	28392.84
card_zip_total_14	96397	805.60	4186.71	0.01	306633.41
card_zip_actual/avg_14	96397	0.98	0.48	0.00	32.76
card_zip_actual/max_14	96397	0.83	0.32	0.00	1.00
card_zip_actual/med_14	96397	1.15	2.99	0.00	400.00
card_zip_actual/toal_14	96397	0.71	0.39	0.00	1.00
card_zip_count_30	96397	8.41	28.95	1.00	425.00
card_zip_avg_30	96397	404.37	776.68	0.01	28392.84
card_zip_max_30	96397	534.56	1086.90	0.01	47900.00
card_zip_med_30	96397	390.50	774.87	0.01	28392.84
card_zip_total_30	96397	988.95	4344.95	0.01	306633.41
card_zip_actual/avg_30	96397	0.98	0.57	0.00	32.76
card_zip_actual/max_30	96397	0.78	0.35	0.00	1.00
card_zip_actual/med_30	96397	1.23	7.93	0.00	2248.70
card_zip_actual/toal_30	96397	0.64	0.41	0.00	1.00
card_zip_count_180	96397	20.61	68.98	1.00	756.00
card_zip_avg_180	96397	405.70	762.98	0.01	28392.84
card_zip_max_180	96397	663.72	1190.22	0.01	47900.00
card_zip_med_180	96397	376.20	758.28	0.01	28392.84
card_zip_total_180	96397	1964.12	5885.68	0.01	306633.41
card_zip_actual/avg_180	96397	0.99	0.83	0.00	72.13
card_zip_actual/max_180	96397	0.67	0.39	0.00	1.00
card_zip_actual/med_180	96397	1.37	3.88	0.00	490.77
card_zip_actual/toal_180	96397	0.50	0.42	0.00	1.00
card_merchdesc_count_0_by_7	96397	0.13	0.03	0.00	0.14
card_merchdesc_count_0_by_14	96397	0.06	0.02	0.00	0.07
card_merchdesc_count_0_by_30	96397	0.03	0.01	0.00	0.03
card_merchdesc_count_1_by_7	96397	0.13	0.03	0.00	0.14
card_merchdesc_count_1_by_14	96397	0.06	0.02	0.00	0.07
card_merchdesc_count_1_by_30	96397	0.03	0.01	0.00	0.03
card_merchdesc_count_3_by_7	96397	0.14	0.02	0.00	0.14
card_merchdesc_count_3_by_14	96397	0.07	0.01	0.00	0.07
card_merchdesc_count_3_by_30	96397	0.03	0.01	0.00	0.03

card_state_count_0_by_7	96397	0.11	0.04	0.00	0.14
card_state_count_0_by_14	96397	0.05	0.02	0.00	0.07
card_state_count_0_by_30	96397	0.02	0.01	0.00	0.03
card_state_count_1_by_7	96397	0.12	0.04	0.00	0.14
card_state_count_1_by_14	96397	0.05	0.02	0.00	0.07
card_state_count_1_by_30	96397	0.02	0.01	0.00	0.03
card_state_count_3_by_7	96397	0.12	0.03	0.00	0.14
card_state_count_3_by_14	96397	0.05	0.02	0.00	0.07
card_state_count_3_by_30	96397	0.02	0.01	0.00	0.03
card_zip_count_0_by_7	96397	0.12	0.04	0.00	0.14
card_zip_count_0_by_14	96397	0.06	0.02	0.00	0.07
card_zip_count_0_by_30	96397	0.02	0.01	0.00	0.03
card_zip_count_1_by_7	96397	0.13	0.03	0.00	0.14
card_zip_count_1_by_14	96397	0.06	0.02	0.00	0.07
card_zip_count_1_by_30	96397	0.02	0.01	0.00	0.03
card_zip_count_3_by_7	96397	0.13	0.03	0.00	0.14
card_zip_count_3_by_14	96397	0.06	0.02	0.00	0.07
card_zip_count_3_by_30	96397	0.03	0.01	0.00	0.03
merchnum_state_day_since	96397	60.73	125.09	0.00	365.00
merchnum_state_count_0	96397	6.82	18.99	1.00	260.00
merchnum_state_avg_0	96397	395.41	760.10	0.01	28392.84
merchnum_state_max_0	96397	503.08	1010.60	0.01	47900.00
merchnum_state_med_0	96397	381.37	750.60	0.01	28392.84
merchnum_state_total_0	96397	782.61	2850.05	0.01	217467.18
merchnum_state_actual/avg_0	96397	1.00	0.63	0.00	37.96
merchnum_state_actual/max_0	96397	0.81	0.34	0.00	1.00
merchnum_state_actual/med_0	96397	1.26	3.51	0.00	405.45
merchnum_state_actual/toal_0	96397	0.72	0.39	0.00	1.00
merchnum_state_count_1	96397	11.58	31.62	1.00	327.00
merchnum_state_avg_1	96397	397.26	748.87	0.01	28392.84
merchnum_state_max_1	96397	587.42	1184.31	0.01	47900.00
merchnum_state_med_1	96397	371.39	737.60	0.01	28392.84
merchnum_state_total_1	96397	1131.48	4379.53	0.01	306633.41
merchnum_state_actual/avg_1	96397	1.00	0.77	0.00	43.43
merchnum_state_actual/max_1	96397	0.74	0.38	0.00	1.00
merchnum_state_actual/med_1	96397	1.38	4.11	0.00	405.45
merchnum_state_actual/toal_1	96397	0.63	0.42	0.00	1.00
merchnum_state_count_3	96397	21.04	55.18	1.00	466.00
merchnum_state_avg_3	96397	397.49	728.28	0.01	28392.84

merchnum_state_max_3	96397	673.59	1261.76	0.01	47900.00
merchnum_state_med_3	96397	361.75	716.20	0.01	28392.84
merchnum_state_total_3	96397	1594.83	5180.85	0.01	307302.58
merchnum_state_actual/avg_3	96397	1.00	0.93	0.00	64.09
merchnum_state_actual/max_3	96397	0.69	0.41	0.00	1.00
merchnum_state_actual/med_3	96397	1.47	4.88	0.00	467.83
merchnum_state_actual/toal_3	96397	0.57	0.43	0.00	1.00
merchnum_state_count_7	96397	42.18	106.16	1.00	762.00
merchnum_state_avg_7	96397	396.63	691.15	0.01	28392.84
merchnum_state_max_7	96397	827.26	1418.63	0.01	47900.00
merchnum_state_med_7	96397	346.37	679.41	0.01	28392.84
merchnum_state_total_7	96397	2645.38	6363.83	0.01	313984.55
merchnum_state_actual/avg_7	96397	1.00	1.10	0.00	82.46
merchnum_state_actual/max_7	96397	0.61	0.42	0.00	1.00
merchnum_state_actual/med_7	96397	1.56	4.65	0.00	473.98
merchnum_state_actual/toal_7	96397	0.48	0.43	0.00	1.00
merchnum_state_count_14	96397	76.09	191.95	1.00	1091.00
merchnum_state_avg_14	96397	398.33	664.92	0.01	28392.84
merchnum_state_max_14	96397	1000.23	1717.20	0.01	47900.00
merchnum_state_med_14	96397	335.97	646.46	0.01	28392.84
merchnum_state_total_14	96397	4371.48	8845.14	0.01	319334.68
merchnum_state_actual/avg_14	96397	1.00	1.27	0.00	133.53
merchnum_state_actual/max_14	96397	0.56	0.43	0.00	1.00
merchnum_state_actual/med_14	96397	1.63	4.97	0.00	490.71
merchnum_state_actual/toal_14	96397	0.41	0.42	0.00	1.00
merchnum_state_count_30	96397	146.67	376.45	1.00	1828.00
merchnum_state_avg_30	96397	397.85	624.03	0.01	28392.84
merchnum_state_max_30	96397	1253.08	2239.49	0.01	47900.00
merchnum_state_med_30	96397	322.07	593.01	0.01	28392.84
merchnum_state_total_30	96397	8013.59	14194.22	0.01	320373.00
merchnum_state_actual/avg_30	96397	1.00	1.45	0.00	172.64
merchnum_state_actual/max_30	96397	0.50	0.43	0.00	1.00
merchnum_state_actual/med_30	96397	1.68	4.56	0.00	481.59
merchnum_state_actual/toal_30	96397	0.34	0.40	0.00	1.00
merchnum_state_count_180	96397	518.79	1520.53	1.00	7852.00
merchnum_state_avg_180	96397	391.55	554.22	0.01	28392.84
merchnum_state_max_180	96397	2031.89	4217.36	0.01	47900.00
merchnum_state_med_180	96397	296.48	523.22	0.01	28392.84
merchnum_state_total_180	96397	30408.92	57432.18	0.01	413457.84

merchnum_state_actual/avg_180	96397	1.02	1.74	0.00	177.99
merchnum_state_actual/max_180	96397	0.39	0.41	0.00	1.00
merchnum_state_actual/med_180	96397	1.84	5.14	0.00	473.98
merchnum_state_actual/toal_180	96397	0.24	0.36	0.00	1.00
merchnum_state_count_0_by_7	96397	0.08	0.05	0.00	0.14
merchnum_state_count_0_by_14	96397	0.03	0.03	0.00	0.07
merchnum_state_count_0_by_30	96397	0.01	0.01	0.00	0.03
merchnum_state_count_1_by_7	96397	0.09	0.05	0.00	0.14
merchnum_state_count_1_by_14	96397	0.04	0.03	0.00	0.07
merchnum_state_count_1_by_30	96397	0.01	0.01	0.00	0.03
merchnum_state_count_3_by_7	96397	0.11	0.04	0.00	0.14
merchnum_state_count_3_by_14	96397	0.04	0.02	0.00	0.07
merchnum_state_count_3_by_30	96397	0.02	0.01	0.00	0.03
merchnum_zip_day_since	96397	61.92	126.23	0.00	365.00
merchnum_zip_count_0	96397	6.82	18.99	1.00	260.00
merchnum_zip_avg_0	96397	395.35	759.74	0.01	28392.84
merchnum_zip_max_0	96397	503.21	1010.17	0.01	47900.00
merchnum_zip_med_0	96397	381.15	750.11	0.01	28392.84
merchnum_zip_total_0	96397	782.39	2848.69	0.01	217467.18
merchnum_zip_actual/avg_0	96397	1.00	0.63	0.00	37.96
merchnum_zip_actual/max_0	96397	0.81	0.34	0.00	1.00
merchnum_zip_actual/med_0	96397	1.26	3.45	0.00	405.45
merchnum_zip_actual/toal_0	96397	0.72	0.39	0.00	1.00
merchnum_zip_count_1	96397	11.58	31.62	1.00	327.00
merchnum_zip_avg_1	96397	397.35	748.90	0.01	28392.84
merchnum_zip_max_1	96397	588.29	1194.19	0.01	47900.00
merchnum_zip_med_1	96397	371.42	737.53	0.01	28392.84
merchnum_zip_total_1	96397	1132.46	4379.76	0.01	306633.41
merchnum_zip_actual/avg_1	96397	1.00	0.77	0.00	43.43
merchnum_zip_actual/max_1	96397	0.74	0.38	0.00	1.00
merchnum_zip_actual/med_1	96397	1.37	4.06	0.00	405.45
merchnum_zip_actual/toal_1	96397	0.63	0.42	0.00	1.00
merchnum_zip_count_3	96397	21.04	55.18	1.00	466.00
merchnum_zip_avg_3	96397	397.44	728.08	0.01	28392.84
merchnum_zip_max_3	96397	674.26	1271.35	0.01	47900.00
merchnum_zip_med_3	96397	361.89	716.23	0.01	28392.84
merchnum_zip_total_3	96397	1599.18	5187.61	0.01	307302.58
merchnum_zip_actual/avg_3	96397	1.00	0.93	0.00	64.09
merchnum_zip_actual/max_3	96397	0.69	0.41	0.00	1.00

merchnum_zip_actual/med_3	96397	1.47	4.86	0.00	467.83
merchnum_zip_actual/toal_3	96397	0.57	0.43	0.00	1.00
merchnum_zip_count_7	96397	42.20	106.16	1.00	762.00
merchnum_zip_avg_7	96397	396.67	691.83	0.01	28392.84
merchnum_zip_max_7	96397	829.57	1468.22	0.01	47900.00
merchnum_zip_med_7	96397	346.82	680.16	0.01	28392.84
merchnum_zip_total_7	96397	2659.51	6397.51	0.01	313984.55
merchnum_zip_actual/avg_7	96397	1.00	1.10	0.00	82.46
merchnum_zip_actual/max_7	96397	0.61	0.42	0.00	1.00
merchnum_zip_actual/med_7	96397	1.56	4.74	0.00	473.98
merchnum_zip_actual/toal_7	96397	0.48	0.43	0.00	1.00
merchnum_zip_count_14	96397	76.13	191.95	1.00	1091.00
merchnum_zip_avg_14	96397	398.15	666.25	0.01	28392.84
merchnum_zip_max_14	96397	1001.57	1770.31	0.01	47900.00
merchnum_zip_med_14	96397	336.54	648.17	0.01	28392.84
merchnum_zip_total_14	96397	4400.08	8924.60	0.01	319334.68
merchnum_zip_actual/avg_14	96397	1.00	1.27	0.00	133.53
merchnum_zip_actual/max_14	96397	0.56	0.43	0.00	1.00
merchnum_zip_actual/med_14	96397	1.61	4.67	0.00	473.98
merchnum_zip_actual/toal_14	96397	0.41	0.42	0.00	1.00
merchnum_zip_count_30	96397	146.76	376.45	1.00	1828.00
merchnum_zip_avg_30	96397	397.60	626.04	0.01	28392.84
merchnum_zip_max_30	96397	1254.90	2303.69	0.01	47900.00
merchnum_zip_med_30	96397	322.94	595.44	0.01	28392.84
merchnum_zip_total_30	96397	8077.23	14420.81	0.01	320373.00
merchnum_zip_actual/avg_30	96397	1.00	1.45	0.00	172.64
merchnum_zip_actual/max_30	96397	0.50	0.43	0.00	1.00
merchnum_zip_actual/med_30	96397	1.67	4.51	0.00	481.59
merchnum_zip_actual/toal_30	96397	0.34	0.41	0.00	1.00
merchnum_zip_count_180	96397	519.17	1520.56	1.00	7852.00
merchnum_zip_avg_180	96397	391.50	555.93	0.01	28392.84
merchnum_zip_max_180	96397	2050.77	4403.53	0.01	47900.00
merchnum_zip_med_180	96397	297.79	524.64	0.01	28392.84
merchnum_zip_total_180	96397	30691.66	58935.72	0.01	413457.84
merchnum_zip_actual/avg_180	96397	1.02	1.73	0.00	177.99
merchnum_zip_actual/max_180	96397	0.40	0.41	0.00	1.00
merchnum_zip_actual/med_180	96397	1.83	5.10	0.00	473.98
merchnum_zip_actual/toal_180	96397	0.24	0.36	0.00	1.00
merchnum_zip_count_0_by_7	96397	0.08	0.05	0.00	0.14

merchnum_zip_count_0_by_14	96397	0.03	0.03	0.00	0.07
merchnum_zip_count_0_by_30	96397	0.01	0.01	0.00	0.03
merchnum_zip_count_1_by_7	96397	0.09	0.05	0.00	0.14
merchnum_zip_count_1_by_14	96397	0.04	0.03	0.00	0.07
merchnum_zip_count_1_by_30	96397	0.01	0.01	0.00	0.03
merchnum_zip_count_3_by_7	96397	0.11	0.04	0.00	0.14
merchnum_zip_count_3_by_14	96397	0.04	0.02	0.00	0.07
merchnum_zip_count_3_by_30	96397	0.02	0.01	0.00	0.03
card_merchnum_zip_day_since	96397	166.37	170.21	0.00	365.00
card_merchnum_zip_count_0	96397	2.10	5.91	1.00	145.00
card_merchnum_zip_avg_0	96397	395.81	796.97	0.01	28392.84
card_merchnum_zip_max_0	96397	421.25	935.81	0.01	47900.00
card_merchnum_zip_med_0	96397	393.46	790.65	0.01	28392.84
card_merchnum_zip_total_0	96397	528.24	2621.18	0.01	217467.18
card_merchnum_zip_actual/avg_0	96397	1.00	0.22	0.00	20.24
card_merchnum_zip_actual/max_0	96397	0.96	0.17	0.00	1.00
card_merchnum_zip_actual/med_0	96397	1.03	0.67	0.00	100.00
card_merchnum_zip_actual/toal_0	96397	0.88	0.27	0.00	1.00
card_merchnum_zip_count_1	96397	2.41	7.59	1.00	177.00

card_merchnum_zip_avg_1	96397	397.20	799.54	0.01	28392.84
card_merchnum_zip_max_1	96397	432.50	1010.27	0.01	47900.00
card_merchnum_zip_med_1	96397	394.44	794.03	0.01	28392.84
card_merchnum_zip_total_1	96397	598.59	4019.58	0.01	306633.41
card_merchnum_zip_actual/avg_1	96397	1.00	0.25	0.00	20.24
card_merchnum_zip_actual/max_1	96397	0.94	0.19	0.00	1.00
card_merchnum_zip_actual/med_1	96397	1.03	0.71	0.00	71.11
card_merchnum_zip_actual/toal_1	96397	0.86	0.30	0.00	1.00
card_merchnum_zip_count_3	96397	3.02	10.98	1.00	248.00
card_merchnum_zip_avg_3	96397	398.11	797.34	0.01	28392.84
card_merchnum_zip_max_3	96397	441.28	1014.57	0.01	47900.00
card_merchnum_zip_med_3	96397	394.69	792.16	0.01	28392.84
card_merchnum_zip_total_3	96397	630.06	4062.16	0.01	306633.41
card_merchnum_zip_actual/avg_3	96397	0.99	0.30	0.00	20.24
card_merchnum_zip_actual/max_3	96397	0.92	0.22	0.00	1.00

card_merchnum_zip_actual/med_3	96397	1.05	1.50	0.00	301.10
card_merchnum_zip_actual/toal_3	96397	0.83	0.32	0.00	1.00
card_merchnum_zip_count_7	96397	4.05	15.65	1.00	358.00
card_merchnum_zip_avg_7	96397	399.84	792.83	0.01	28392.84
card_merchnum_zip_max_7	96397	458.76	1022.77	0.01	47900.00
card_merchnum_zip_med_7	96397	394.69	788.11	0.01	28392.84
card_merchnum_zip_total_7	96397	688.95	4102.95	0.01	306633.41
card_merchnum_zip_actual/avg_7	96397	0.99	0.36	0.00	20.24
card_merchnum_zip_actual/max_7	96397	0.89	0.26	0.00	1.00
card_merchnum_zip_actual/med_7	96397	1.08	2.20	0.00	442.87
card_merchnum_zip_actual/toal_7	96397	0.78	0.36	0.00	1.00
card_merchnum_zip_count_14	96397	5.34	19.04	1.00	369.00
card_merchnum_zip_avg_14	96397	401.80	790.03	0.01	28392.84
card_merchnum_zip_max_14	96397	479.93	1043.49	0.01	47900.00
card_merchnum_zip_med_14	96397	394.68	786.19	0.01	28392.84
card_merchnum_zip_total_14	96397	769.25	4167.49	0.01	306633.41
card_merchnum_zip_actual/avg_14	96397	0.99	0.43	0.00	23.11
card_merchnum_zip_actual/max_14	96397	0.86	0.29	0.00	1.00
card_merchnum_zip_actual/med_14	96397	1.11	2.28	0.00	400.00
card_merchnum_zip_actual/toal_14	96397	0.74	0.38	0.00	1.00
card_merchnum_zip_count_30	96397	7.73	27.43	1.00	409.00
card_merchnum_zip_avg_30	96397	403.91	784.39	0.01	28392.84
card_merchnum_zip_max_30	96397	512.31	1066.09	0.01	47900.00
card_merchnum_zip_med_30	96397	393.32	783.63	0.01	28392.84
card_merchnum_zip_total_30	96397	922.37	4298.53	0.01	306633.41
card_merchnum_zip_actual/avg_30	96397	0.98	0.51	0.00	25.03
card_merchnum_zip_actual/max_30	96397	0.81	0.33	0.00	1.00
card_merchnum_zip_actual/med_30	96397	1.15	2.28	0.00	397.86
card_merchnum_zip_actual/toal_30	96397	0.68	0.40	0.00	1.00
card_merchnum_zip_count_180	96397	15.76	52.95	1.00	606.00
card_merchnum_zip_avg_180	96397	404.69	772.07	0.01	28392.84
card_merchnum_zip_max_180	96397	606.88	1151.28	0.01	47900.00
card_merchnum_zip_med_180	96397	382.40	767.26	0.01	28392.84
card_merchnum_zip_total_180	96397	1659.65	5523.08	0.01	306633.41
card_merchnum_zip_actual/avg_180	96397	0.99	0.70	0.00	71.83
card_merchnum_zip_actual/max_180	96397	0.73	0.37	0.00	1.00
card_merchnum_zip_actual/med_180	96397	1.26	2.70	0.00	405.45
card_merchnum_zip_actual/toal_180	96397	0.56	0.42	0.00	1.00
card_merchnum_zip_count_0_by_7	96397	0.12	0.04	0.00	0.14

card_merchnum_zip_count_0_by_14	96397	0.06	0.02	0.00	0.07
card_merchnum_zip_count_0_by_30	96397	0.02	0.01	0.00	0.03
card_merchnum_zip_count_1_by_7	96397	0.13	0.03	0.00	0.14
card_merchnum_zip_count_1_by_14	96397	0.06	0.02	0.00	0.07
card_merchnum_zip_count_1_by_30	96397	0.03	0.01	0.00	0.03
card_merchnum_zip_count_3_by_7	96397	0.13	0.03	0.00	0.14
card_merchnum_zip_count_3_by_14	96397	0.06	0.02	0.00	0.07
card_merchnum_zip_count_3_by_30	96397	0.03	0.01	0.00	0.03
card_merchnum_state_day_since	96397	165.83	170.10	0.00	365.00
card_merchnum_state_count_0	96397	2.10	5.91	1.00	145.00
card_merchnum_state_avg_0	96397	395.81	796.95	0.01	28392.84
card_merchnum_state_max_0	96397	421.28	935.81	0.01	47900.00
card_merchnum_state_med_0	96397	393.47	790.66	0.01	28392.84
card_merchnum_state_total_0	96397	528.79	2621.85	0.01	217467.18
card_merchnum_state_actual/avg_0	96397	1.00	0.22	0.00	20.24
card_merchnum_state_actual/max_0	96397	0.96	0.17	0.00	1.00
card_merchnum_state_actual/med_0	96397	1.03	0.68	0.00	100.00
card_merchnum_state_actual/toal_0	96397	0.88	0.27	0.00	1.00
card_merchnum_state_count_1	96397	2.42	7.59	1.00	177.00
card_merchnum_state_avg_1	96397	397.20	799.53	0.01	28392.84
card_merchnum_state_max_1	96397	432.51	1010.27	0.01	47900.00
card_merchnum_state_med_1	96397	394.45	794.04	0.01	28392.84
card_merchnum_state_total_1	96397	599.32	4020.29	0.01	306633.41
card_merchnum_state_actual/avg_1	96397	1.00	0.25	0.00	20.24
card_merchnum_state_actual/max_1	96397	0.94	0.19	0.00	1.00
card_merchnum_state_actual/med_1	96397	1.03	0.71	0.00	71.11
card_merchnum_state_actual/toal_1	96397	0.86	0.30	0.00	1.00
card_merchnum_state_count_3	96397	3.03	10.98	1.00	248.00
card_merchnum_state_avg_3	96397	398.13	797.33	0.01	28392.84
card_merchnum_state_max_3	96397	441.32	1014.58	0.01	47900.00
card_merchnum_state_med_3	96397	394.69	792.16	0.01	28392.84
card_merchnum_state_total_3	96397	630.83	4062.87	0.01	306633.41
card_merchnum_state_actual/avg_3	96397	0.99	0.30	0.00	20.24
card_merchnum_state_actual/max_3	96397	0.92	0.22	0.00	1.00
card_merchnum_state_actual/med_3	96397	1.05	1.50	0.00	301.10
card_merchnum_state_actual/toal_3	96397	0.83	0.32	0.00	1.00
card_merchnum_state_count_7	96397	4.05	15.65	1.00	358.00
card_merchnum_state_avg_7	96397	399.89	792.85	0.01	28392.84
card_merchnum_state_max_7	96397	458.85	1022.82	0.01	47900.00

card_merchnum_state_med_7	96397	394.73	788.14	0.01	28392.84
card_merchnum_state_total_7	96397	689.81	4103.71	0.01	306633.41
card_merchnum_state_actual/avg_7	96397	0.99	0.36	0.00	20.24
card_merchnum_state_actual/max_7	96397	0.89	0.26	0.00	1.00
card_merchnum_state_actual/med_7	96397	1.08	2.20	0.00	442.87
card_merchnum_state_actual/toal_7	96397	0.78	0.36	0.00	1.00
card_merchnum_state_count_14	96397	5.34	19.04	1.00	369.00
card_merchnum_state_avg_14	96397	401.89	790.02	0.01	28392.84
card_merchnum_state_max_14	96397	480.12	1043.59	0.01	47900.00
card_merchnum_state_med_14	96397	394.76	786.19	0.01	28392.84
card_merchnum_state_total_14	96397	770.35	4168.41	0.01	306633.41
card_merchnum_state_actual/avg_14	96397	0.99	0.43	0.00	23.11
card_merchnum_state_actual/max_14	96397	0.85	0.29	0.00	1.00
card_merchnum_state_actual/med_14	96397	1.11	2.28	0.00	400.00
card_merchnum_state_actual/toal_14	96397	0.73	0.38	0.00	1.00
card_merchnum_state_count_30	96397	7.73	27.43	1.00	409.00
card_merchnum_state_avg_30	96397	403.98	784.35	0.01	28392.84
card_merchnum_state_max_30	96397	512.57	1066.24	0.01	47900.00
card_merchnum_state_med_30	96397	393.36	783.61	0.01	28392.84
card_merchnum_state_total_30	96397	923.69	4299.58	0.01	306633.41
card_merchnum_state_actual/avg_30	96397	0.98	0.51	0.00	25.03
card_merchnum_state_actual/max_30	96397	0.81	0.33	0.00	1.00
card_merchnum_state_actual/med_30	96397	1.15	2.28	0.00	397.86
card_merchnum_state_actual/toal_30	96397	0.68	0.40	0.00	1.00
card_merchnum_state_count_180	96397	15.77	52.95	1.00	606.00
card_merchnum_state_avg_180	96397	404.80	772.06	0.01	28392.84
card_merchnum_state_max_180	96397	607.38	1151.03	0.01	47900.00
card_merchnum_state_med_180	96397	382.51	767.43	0.01	28392.84
card_merchnum_state_total_180	96397	1662.03	5523.10	0.01	306633.41
card_merchnum_state_actual/avg_180	96397	0.99	0.71	0.00	71.83
card_merchnum_state_actual/max_180	96397	0.73	0.37	0.00	1.00
card_merchnum_state_actual/med_180	96397	1.26	2.70	0.00	405.45
card_merchnum_state_actual/toal_180	96397	0.56	0.42	0.00	1.00
card_merchnum_state_count_0_by_7	96397	0.12	0.04	0.00	0.14

card_merchnum_state_count_0_by_14	96397	0.06	0.02	0.00	0.07
card_merchnum_state_count_0_by_30	96397	0.02	0.01	0.00	0.03
card_merchnum_state_count_1_by_7	96397	0.13	0.03	0.00	0.14
card_merchnum_state_count_1_by_14	96397	0.06	0.02	0.00	0.07
card_merchnum_state_count_1_by_30	96397	0.03	0.01	0.00	0.03
card_merchnum_state_count_3_by_7	96397	0.13	0.03	0.00	0.14
card_merchnum_state_count_3_by_14	96397	0.06	0.02	0.00	0.07
card_merchnum_state_count_3_by_30	96397	0.03	0.01	0.00	0.03

8.3 Top 30 Feature Variables After Filtering and Wrapping

Order	Variable Name

1	card_zip_total_7
2	card_merchnum_zip_total_7
3	card_state_total_7
4	card_merchdesc_total_7
5	card_merchdesc_total_14
6	card_state_total_1
7	card_merchnum_zip_total_14
8	card_state_total_14
9	card_merchnum_state_total_1
10	card_zip_total_30
11	card_merchnum_state_total_30
12	card_merchnum_total_30
13	card_merchnum_zip_total_30
14	card_zip_max_7
15	Cardnum_total_3
16	card_merchnum_total_0
17	card_merchnum_zip_max_14
18	card_state_total_30
19	Cardnum_total_7
20	card_merchdesc_max_3
21	card_zip_total_0
22	card_merchnum_max_3
23	card_state_max_30
24	merchnum_state_total_1
25	card_merchdesc_max_1
26	merchnum_zip_total_1
27	card_merchnum_zip_max_1
28	merchnum_state_max_0

29	merchnum_zip_total_0
30	card_merchdesc_max_180