DSO 562 Project 3

USC Marshall
School of Business

Fraud Detection on Credit Card Transaction Data

UNDER THE GUIDANCE OF
Professor Stephen Coggeshall

5/03/2020

JENNY SHANG |
JENNY WANG |
LINGROU WANG |
RORY WANG |

Table of Contents

Part I. Executive Summary

This report is a summary of an analysis on credit card transaction data. The dataset contains 96,753 records with 10 attributes of basic information on each transaction. The objective of the report is to detect fraudulent events using supervised machine learning algorithms, including logistic regression, neural networks, random forests and boosted trees.

After data exploration and data cleaning, a total of 326 expert variables were built through data manipulation under expert's suggestion and linking of variables. Then the importance of each variable using KS distance and fraud detection rate at 3% of records was evaluated and ranked. The most influential variables were selected through recursive feature elimination method in combination with two special variables. To help alleviate unbalanced data, the analysis team adopted a 3% cutoff and SMOTE (synthetic minority oversampling technique) method to help simulate more minority cases. By splitting the data into training, testing and out-of-time dataset, the models are trained using training dataset, tested on the testing dataset, and validated on out-of-time dataset using fraud detection rate (FDR). For each algorithm, parameter tuning was adopted in order to deduce the optimal combination of variables to achieve the best results possible.

Overall, the best model is a random forest with 50 trees and a max depth of 20, using which we achieved a fraud detection rate of 58.16% on out-of-time data in the top 3% highest scored records.

	FDR @ 3%		
Model	**Training**	**Testing**	**OOT**
Logistic Regression	70.78%	69.30%	40.99%
Random Forest	100.00%	89.02%	58.16%
Neural Net	83.37%	78.28%	55.47%
Adaptive Boosting	92.92%	85.34%	55.31%

Table 1.1 Overview of Models

Based on estimates provided by domain experts, the analysis team calculated the maximum annual savings would be around $1.3 Million with the model developed, suggests a cutoff fraud detection percentage at 6%.

Part II. Description of Data

Overview of Data

The "card transactions" dataset contains a series of credit card transaction information. This dataset is obtained from a government agency in the State of Tennessee and tracks transactions made by employees at the agency on their expenditures. Of these transactions, some are labeled as fraudulent transactions that warrant further investigation.

In this dataset, the values for all fields are considered to be realistic. In summary, this dataset contains a total of 10 fields and 96,753 records. All records are generated with dates in 2010 in chronological order, from January 1, 2010 to December 31, 2010. The timing of the transaction is important to keep in mind when constructing supervised fraud detection models.

A data quality report is constructed with details in Appendix 1. The report begins with a summary of the variables in the dataset, follows with analysis of each of the variables, and concludes with additional notes on the dataset. A snippet of important features from the data quality report is included in the sections below.

Summary Tables

The dataset contains a unique identifier, "Recnum" for each transaction. A date field tracks the year, month, and date of the transaction in the format "mm/dd/yy". The dataset has one numerical field, which is the transaction amount of each record.

Field Name	# Records with Value	% Populated	# Unique Values	# Records with Value Zero	Mean	Standard Deviation	Min	Max
Amount	96753	100%	34909	0	427.89	10006.14	0.01	3102045.53

Table 2.1 - Amount Summary Table

The remaining fields are categorical fields related specifically to the credit card used in each transaction and information on the merchant. Information about the merchant may be required for management approval purposes. Given that not 100% of the "Merchnum", "Merch state" and "Merch zip" fields are populated, these fields are presumed to be optional.

Field Name	# Records with Value	% Populated	# Unique Values	# Records with Value Zero	Most Common Value
Recnum	96753	100%	96753	0	N/A
Cardnum	96753	100%	1645	0	514214452
Date	96753	100%	365	0	2/28/10
Merchnum	93378	96.51%	13091	231	930090121224
Merch description	96753	100%	13126	0	GSA-FSS-ADV
Merch state	95558	98.76%	227	0	TN
Merch zip	92097	95.19%	4567	0	38118
Transtype	96753	100%	4	0	P
Fraud	96753	100%	2	95694	0

Table 2.2 – Categorical Variables Summary Table

Visualizations for Variables

Figure 2.1 looks at the distribution of the transaction amount for all the records. From the graph, there is a noticeable sharp decrease in amounts greater than approx. $2,500. A possible explanation of this phenomenon may be that additional approvals are required to expense amounts beyond $2,500. Alternatively, this may be a spending limit for the majority of the employees.

Interestingly, there is also a small rise in the number of transactions right before the drop. Upon further investigation comparing the fraudulent and non-fraudulent records, it is evident that the number of fraudulent records also increases in this range. This aligns with the assumption that further approval is required for expenditures above $2,500 as fraudsters try to maximize the amount of benefit that can be attained without approval.

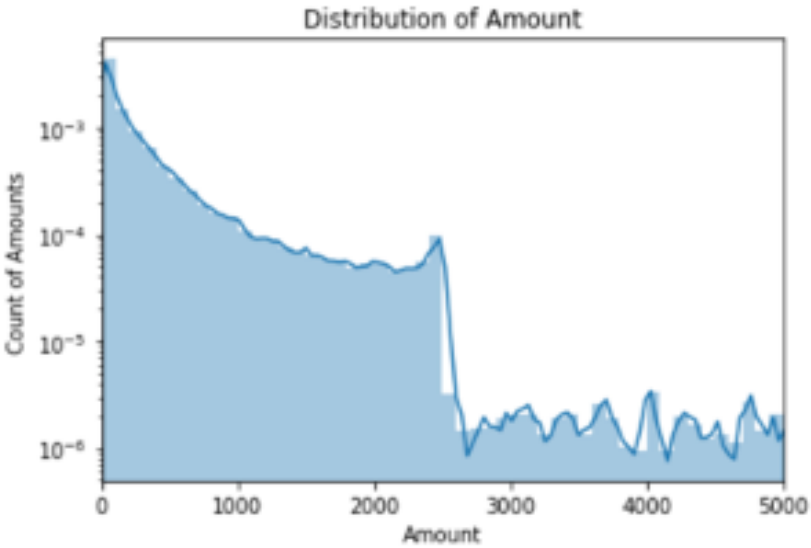


Figure 2.1 – “Amount” Distribution Graph

Figure 2.2 – Number of Fraud and Non-fraud Transactions

After examining the transactions across time, a noticeable pattern is that there are more transactions happening in September on a per day basis. This coincides with the fiscal year of government agencies where a new budget is set in October, and employees tend to spend whatever unspent budget they have at the end of fiscal year.

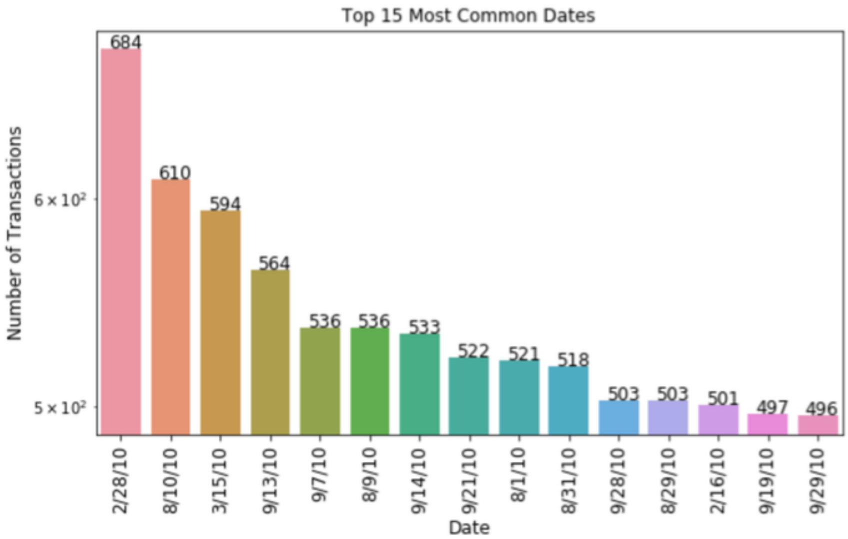


Figure 2.3 – Top 15 Days with Most Number of Transactions

Looking at the percentage of fraudulent and non-fraudulent transactions per day, it is also clear that more fraud activities occur in the latter end of the fiscal year around August and September. The variation on a day-to-day basis in Figure 2.4.1 is due to weekly cyclicality, where there are no employee transactions over the weekend.

When plotted out on a weekly basis, the percentage of non-fraudulent transactions each week is at a steady rate of about 2% of total non-fraudulent transaction (as indicated by the green line); whereas the percentage of fraudulent transactions per day has ups and downs with big spikes in April-May and August-September (as indicated by the red line). As mentioned previously, the fiscal cycle may be a plausible explanation for the spike in August-September. The rise in fraudulent activities in April-May may be due to tax reasons.

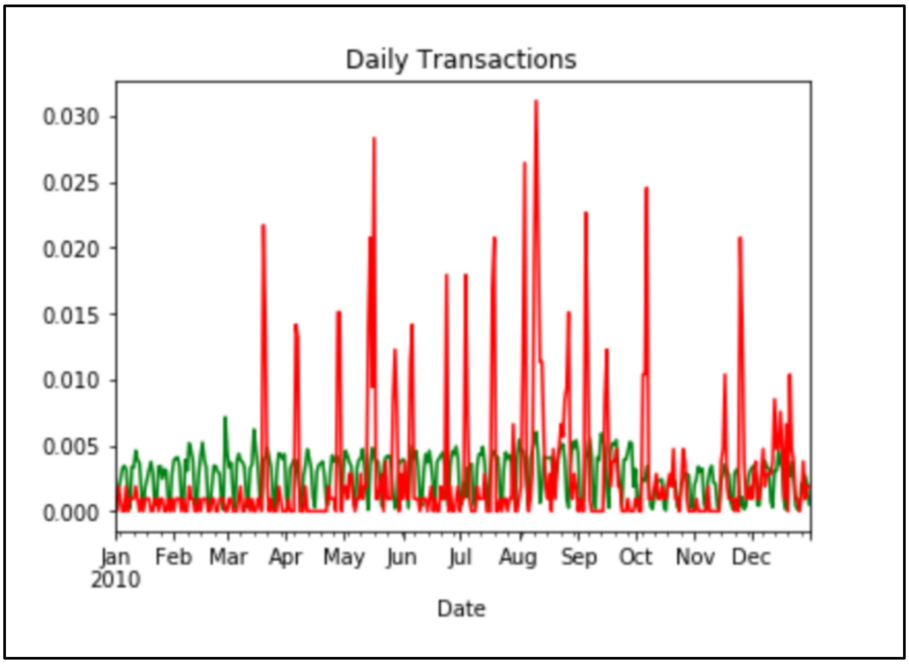


Figure 2.4.1 – Breakdown of Fraudulent vs. Non-Fraudulent Transactions by Day

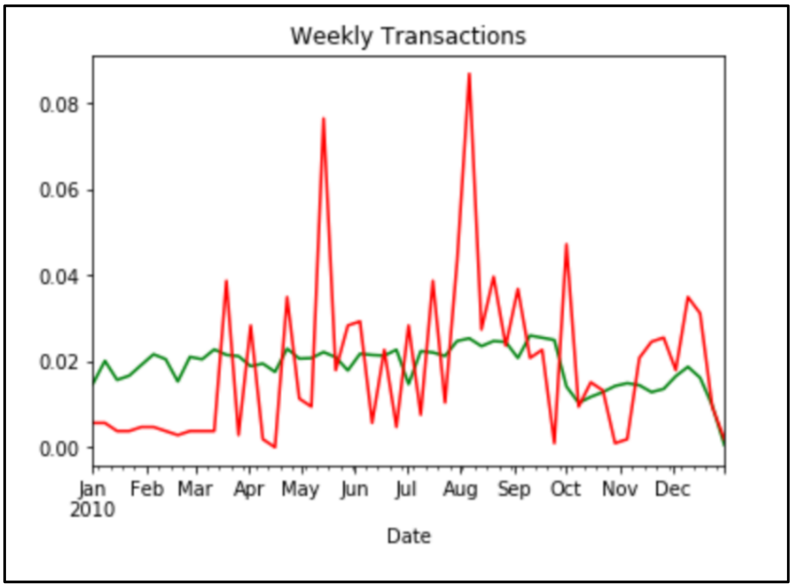


Figure 2.4.2 – Breakdown of Fraudulent vs. Non-Fraudulent Transactions by Week

More details regarding the dataset are addressed in the Data Quality Report in Appendix 1.

Part III. Data Cleaning

Exclusions

For the purposes of this report, only transactions of type P were used. Transactions of type A, D, and Y were removed from the dataset which all totaled to 355 records. As this isn't a significant amount of records to remove, these transaction types were not replaced or altered in any way and were simply cut from the dataset. This left the dataset with 96,398 records.

Outliers

One outlier was spotted in the data, that had an amount of $3,102,045.53, which was much higher than all of the other transaction amounts. This is an unusual transaction amount that may have arisen from an error in currency, leading to an outlier in the amount variable. To prevent this record from contaminating the fraud score calculation, it was removed from the data set, leaving a total of 96,397 records left.

Missing Data

This dataset contained missing fields in the merchant number, merchant state, and merchant zip variables. Instead of removing all of these records completely, the missing variables were filled in following the method outlined below.

First, the missing merchant numbers were filled in with the most common value from similar merchant descriptions. The reasoning assumed that the merchant number would correspond well with the merchant description, as similar descriptions were likely to have a similar merchant number. If no matching merchant description was available, it was filled in with its corresponding negative record number. This was done to prevent incorrect linkage.

After filling in all of the missing merchant numbers, the merchant numbers were used to fill in the merchant state variables, following the same logic as above where records with the same merchant number were replaced with the most common merchant state. This was based on the assumption that for one merchant number it was most likely from the same state. If no matching merchant number is available, it was filled in with the negative record number.

Likewise, after filling in all of the missing merchant state values, missing merchant zip codes were filled in with similar merchant states and card numbers. The reason for using the card number was that it was assumed that employee purchases tend to be repetitive. For example, an employee in a specific role will make repeated purchases from the same batch of merchants. Any remaining missing Merch zip variables were filled in with negative their corresponding record number. Additionally, there were also merchant states that were Canadian provinces, so we gave them a 3-digit zip code for each of the provinces.

Part IV. Variable Creation

Credit card transaction fraud typically involves a fraudster using someone else's credit card or credit account to make purchases not authorized by the actual card holder. In the context of the current dataset, credit card fraud can also be a result of government employees making non-work-related purchases to benefit from personal gains.

Before building the expert variables, three additional linking entities were created:

1) "Cardnum" linking with "Merch zip"
2) "Cardnum" linking with "Merchnum"
3) "Cardnum" linking with "Merch state"

The linking entities in addition to the original "Cardnum" and "Merchnum" fields help in identifying each individual record and distinguish the fraud records from non-fraud records. For each linking, it could be considered as a slightly more unique identifier in addition to the existing 10 fields available for each record.

While building the expert variables, it is important to keep in mind the types of credit card transaction fraud that this dataset exhibits. The most common kind of fraud that could be taking place is identity fraud, where a fraudster outside of the government agency steals the credit card information of an employee to make fraudulent purchases.

However, it is also likely for a government employee to make purchases in large quantities for personal gain. Under this type of fraud, frequent appearances of the same card number within short periods of time were expected. The second type of fraud that could occur is synthetic identity fraud, where the fraudster makes up random information. This may be reflected in records with the same card number but different merchant numbers/merchant zip code/merchant state.

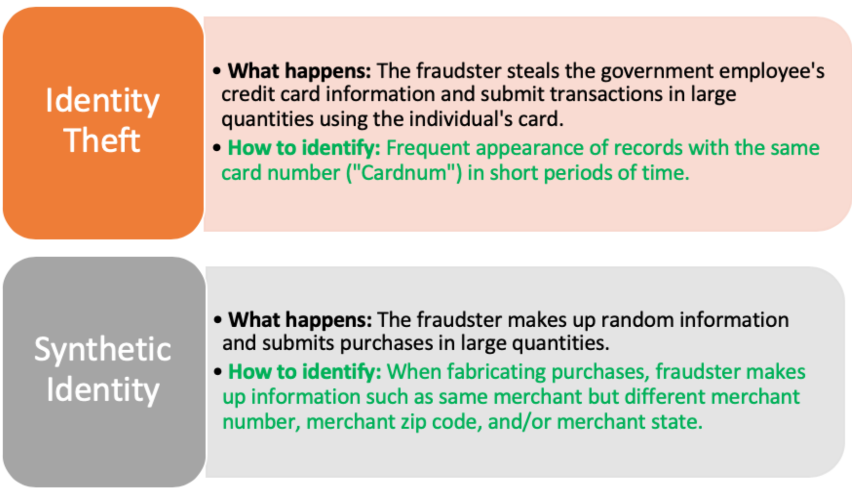


Figure 4.1 – Types of Fraud

After creating the linking entities, there are a total of 326 expert created based on the five entities, which include velocity variables, relative velocity variables, amount variables, days since last seen variables, Benford's Law variables, and one risk variable. These variables are used to correctly identify the above two types of fraud.

Velocity Variables

Velocity variables reflect the number of occurrences for each value seen over the past certain number of days. A total of 30 velocity variables were created based on different combinations of entities and number of day counts. As an example, for each record's "Cardnum", we can calculate how many times this particular card number was used to transact in the previous 7 days. All five entities were considered and the number of occurrences for each value across the five entities in the previous 0, 1, 3, 7, 14, and 30 days are calculated. It is important to consider the sequence of the records when creating the velocity variables.

Since the data represents a time series and it is technically impossible to know information from the future, the number of occurrences only factors in prior records with the same value. For example, if a particular card number appeared three times in the same day, the 0-day velocity variable would have different values of 1, 2, and 3 because at the time of first occurrence, the latter two transactions have not taken place. A summary of velocity variables created is shown in Figure 4.2.

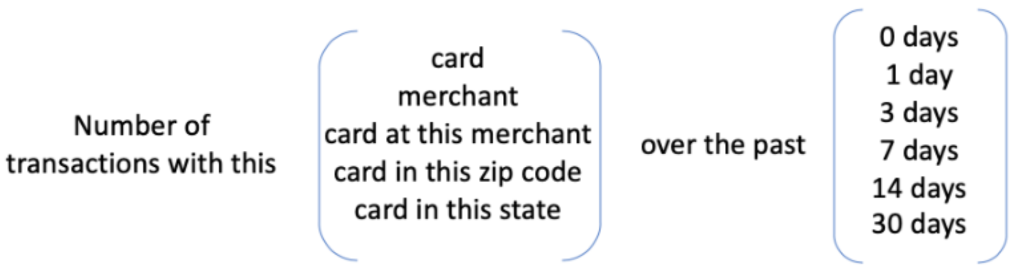


Figure 4.2 – Velocity Variables

Relative Velocity Variables

Relative velocity variables are calculated by taking the ratio of the previously calculated velocity variables. The numerator is the velocity variable for short durations, such as 0, 1 or 3 days; whereas the denominator is the same velocity variable but for longer periods, such as 7, 14, or 30 days. The relative velocity variables are good indicators of fraudulent credit card where fraudsters make multiple purchases in a short time frame. A total of 48 relative velocity variables are created and a visual is available in Figure 4.3.

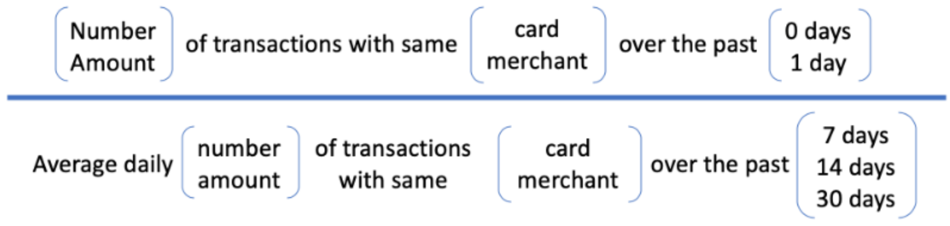


Figure 4.3 – Relative Velocity Variables

Amount Variables

Amount variables look at eight various measures of the amount by each of the entities over the past number of days. The amount measures we utilized are:

- Average amount
- Maximum amount
- Median amount
- Total amount
- Actual amount divided by average amount
- Actual amount divided by maximum amount
- Actual amount divided by median amount
- Actual amount divided by total amount

For example, for a particular merchant number, we can calculate the average amount of transactions associated with this merchant number over the past 7 days. The amount variables are useful in identifying transactions of abnormal amounts. As an example, in the case where a particularly large amount appeared, it would skew the average and total amount upward. A total of 240 amount variables are created across the five entities as shown in Figure 4.4.

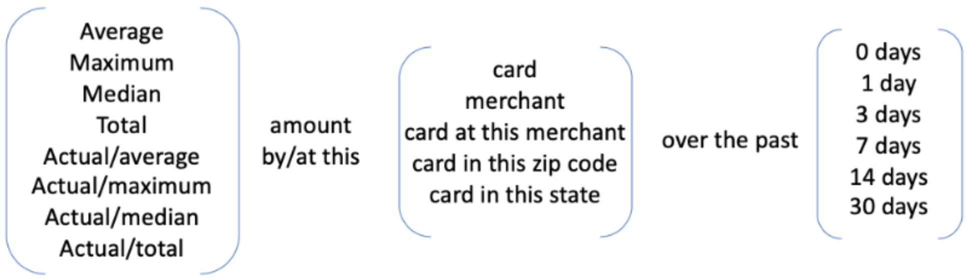


Figure 4.4 – Amount Variables

Days Since Variables

Days since variables calculate the number of days since a particular value was last seen, and it is computed for all five entities. The calculation first identifies the dates when a particular entity value appeared and computes the difference between each of the record dates and the previous date. This type of variable is a good indicator of fraudulent activities where a fraudster continues to make purchases across time. Similar to the velocity variables, the chronological occurrence of each record is taken into account. For values that appear multiple times in the same day, the latter occurrences except for the first will all have a value of 0. A summary of the day since variables is available in Figure 4.5.

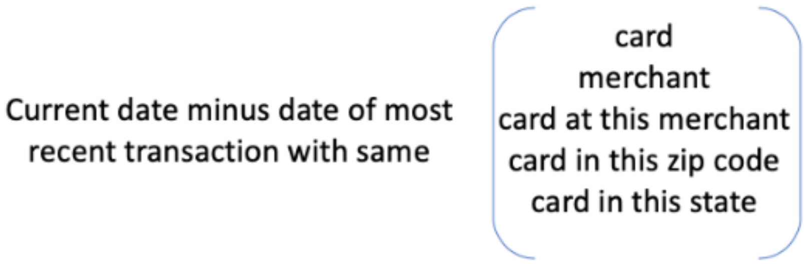


Figure 4.5 – Day Since Variables

Naturally, the higher the value, the lower the risk of it being a fraud transaction, which contrasts the behavior of other variables. Therefore, a transformation of using 366 minus the original days since value was adopted.

<u>Benford's Law Variables</u>

Benford's Law describes the fact that the distribution of the first digit of many measurements (with unit of account) is similar to an exponential logarithmic distribution. Similarly, Benford's law variables look at the relative proportion of the distribution of the first digit of the transaction amounts. Benford's Law is useful in fraud detection when a potential fraudster makes up a large number of transactions. When a single person makes up a lot of numbers, it distorts the expected distribution of the first digit numbers, which violates Benford's Law and thus flags the individual as abnormal.

For this dataset, two Benford's Law variables are constructed to rank the different "Cardnum" (card number) and "Merchnum" (merchant number) fields. By taking the ratio of the proportion of occurrences of small and large numbers for the first digit in transaction amount, an abnormality score to rank abnormality and spot potential fraud was constructed. The Benford's Law variables are useful in this case for determining fraud because as previously shown, there is a spike in transactions at about the $2000 mark, which will distort the first digit distributions to be different from the normal distribution.

Risk Variables

Risk variables are used to assign a value to categorical variables. In this case, a risk variable is created based on the day of week each transaction is created. For each day of week, the average fraud label value is computed for all records that occur on that particular day of week. This value then becomes the record's risk variable value. Note that the averages are only computed using training data, which are records that occurred before November 1, 2010.

In addition, a smoothing formula is applied such that the value is less variant and less affected by values calculated based on statistically insufficient observations in each category. As the below Figure 4.6 shows, there are on average more fraud transactions happening on Friday than other days of the week.

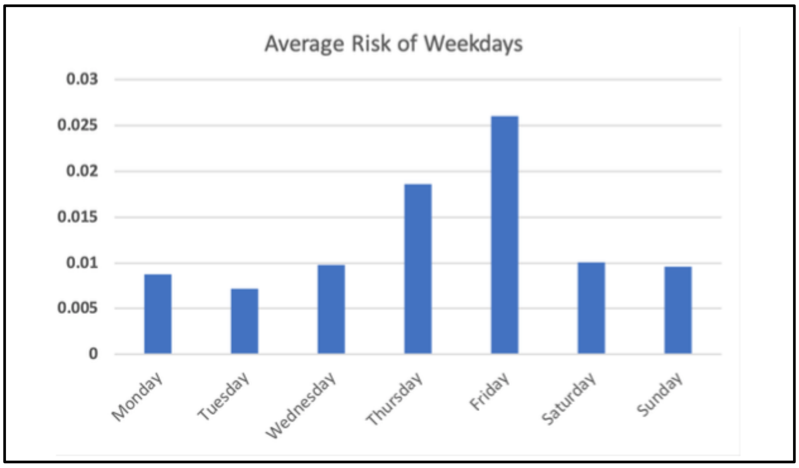


Figure 4.6 – Average Risk by Weekdays

A list of all 326 variables is included in Appendix 2.

Part V. Feature Selection Process

After creating the 326 expert variables, two additional variables were added as a check to ensure the appropriateness of the feature selection process: 1) actual fraud label, and 2) an independent random variable. If the feature selection process is accurate, the fraud label would be the top variable to be considered since it is the best variable for indicating fraudulent records, while the random variable should be relatively low in consideration.

Then, all variables are standardized by Z-scaling before feature selection to make sure all variables are on a comparable scale. With the 326 expert variables created and standardized, the next step is to choose the top variables that explain the majority of the data variance to reduce dimensionality. This step benefits algorithm development in later stages to enable models to be trained faster, reduces the likelihood of overfitting, and tries to make the balance between complexity of variables and the total variance of data explained by the variables. Two measures are adopted and all variables are ranked accordingly:

1. **Kolmogorov-Smirnov (KS)**, a measure of how well two distributions are separated (in this case, fraud and regular transactions). The higher the KS, the more separate the two distributions and the better the variable is as a predictor.

2. **Fraud detection rate (FDR)**, which is the relative percentage of all fraud identified at a specific cutoff location. In this case, a 3% cutoff was suggested by the subject matter expert, which means that the top 3% of rows after sorting the data by predicted probability, will be used to calculate the fraud detection rate of that variable.

A detailed ranking of the variables can be found in Appendix 3. The rankings served as a primary filter for the feature selection process. Feature ranking with recursive feature elimination and cross-validated selection of the best number of features using logistic regression was used to narrow down what variables to include for model building.

The top 80 variables with the highest average KS and FDR ranking are used for further feature selection. A wrapper method called recursive feature elimination is used to find the best subset of variables to build algorithms. This is a process of repeatedly creating models and keeping the best or the worst performing feature in each iteration. Then it builds the next model with the features left until all the features are exhausted. Finally, it will rank the features based on the order of their elimination. This selection process using the top ranked 80 variables is initially used as input for a desired top 30 variables.

However, the initial selected variables were far more than 30 variables, and with multiple selection processes, each selection yields slightly different results. This was due to the fact that among the 326 variables created, some variables are highly correlated (for example, zip code with a card number seen in the past 3 days and state name with the same card number in the past 3 days). Each time when there were highly correlated variables, the selection algorithm randomly selects one variable out of the correlated group. Depending on what prior variables were chosen, it gave different results, which would result in a drastically different variable list.

After the initial feature selection, the top 50 variables with the highest ranking were chosen. However, many variables were identified with the same importance, having the same ranking as others. In this case, it would be hard to tell from the 50 variables, which ones are more important than others.

Therefore, an additional recursive feature elimination was again used to find among the best previously identified 50 variables, and tried to seek the specific ranking of importance of each variable.

After running recursive feature elimination two times, the final result seemed to be heavily focused on amount related variables. With only amounts variable, the model might not be able to flag the fraud incidents that violated the Benford's law. Thus, the analysis team decided to add in two Benford's Variables in addition to the top 33 features that were selected, which made the final features list of 35 variables in total:

Variable	Feature_Name
1	CardMerchanttot_AMT3
2	CardStatetot_AMT1
3	CardStatetot_AMT7
4	CardZiptot_AMT3
5	CardZiptot_AMT7
6	CardMerchanttot_AMT1
7	CardZiptot_AMT14
8	CardMerchanttot_AMT14
9	CardMerchanttot_AMT30
10	CardZiptot_AMT1
11	CardStatetot_AMT14
12	CardMerchantmax_AMT14
13	CardStatemax_AMT7
14	CardMerchanttot_AMT0
15	CardZipmax_AMT14
16	CardZiptot_AMT0
17	CardStatemax_AMT14
18	Cardnumtot_AMT3
19	Merchnumtot_AMT1
20	CardZipmax_AMT3
21	Cardnumtot_AMT7
22	CardStatetot_AMT30
23	CardMerchantmax_AMT1
24	CardStatemax_AMT30
25	CardZipmax_AMT1
26	CardStatemax_AMT1
27	Merchnummax_AMT0
28	Merchnumtot_AMT3
29	Cardnumtot_AMT1
30	CardStatemax_AMT0
31	Cardnumtot_AMT0
32	Merchnumtot_AMT7
33	Cardnummax_AMT7
34	Merchustar
35	CardUstar

Table 5.1 Features List

The figure below shows that the majority of the data variance can be captured by 25-35 variables. Even if more variables were added, not much additional variance was explained. Therefore, the feature selection process effectively reduced dimensionality and only chose the ones that are vital.

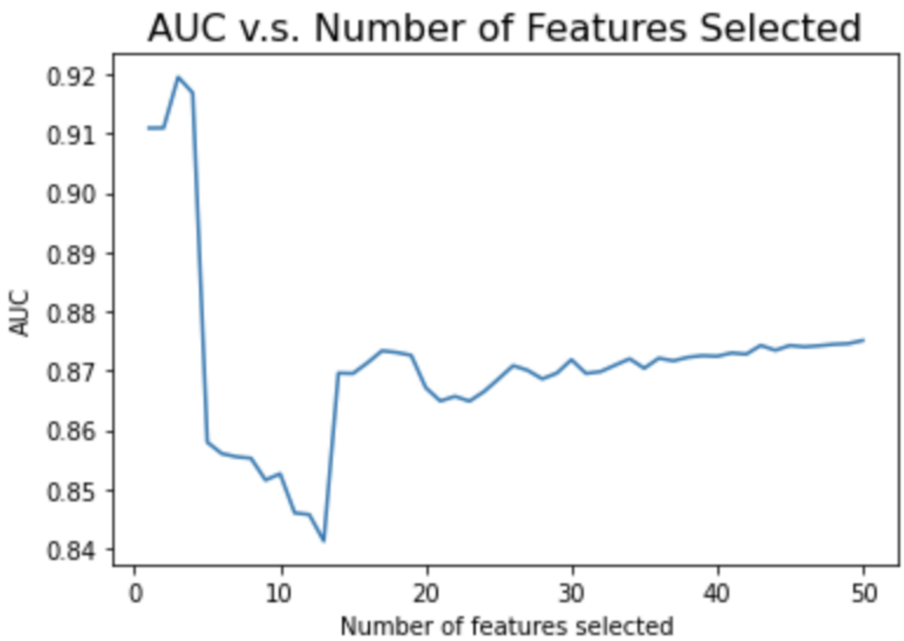


Figure 5.1 – AUC vs. Number of Features Selected

Methods to Alleviate Unbalanced Data

Fraudulent transactions tend to be a rare event within the dataset, since only 1.098% of the transactions are labeled fraudulent. Therefore, regular model development might seem to be overly lenient when classifying fraud transactions to achieve a higher accuracy score. To help alleviate this issue, two specific methods are adopted:

1. **Fraud Detection Rate (FDR) at 3%**
   The model was used to predict the probability of a record to be fraudulent instead of giving out either 0 or 1 predictions (to classify if the records are fraud or not). The probability list was used to rank the data, from the records with the highest probability to the lowest. A suggested cutoff point was provided by the domain expert at 3%, and a count of fraud transactions before the cutoff was used to calculate as the percentage of fraud caught over total fraud in the dataset.

2. **Synthetic Minority Oversampling (SMOTE)**
   SMOTE is a statistical method to help simulate more minority cases, in this situation, the fraudulent records, while it does not change the number of majority cases. After resampling with SMOTE, the dataset achieved a fraud transaction to regular transaction of 1:1, and is no longer imbalanced, and no information was lost during this process.

Part VI. Algorithm Development

<u>Model Performances</u>

After finalizing the set of variables to use, the next step is model development. Out of time data is excluded in the development stage, which are records that occurred after 10/31/2010, and serve as an out of sample model performance evaluation after the model selection process. The metric on selecting the best fraud detection model uses the fraud detection rate at 3%, which is how many fraud incidents the model can identify at a 3% cutoff.

An overview of a total of 4 algorithms are examined on the training data with a 70% training and 30% testing split. The model performance is assessed by averaging a total of 10 trials on each set of parameters, splitting the data differently. Below is an overview of the best outcomes from each model, and individual model effectiveness is available after the general overview.

Algorithm Name	Parameters	Training FDR at 3%	Testing FDR at 3%	OOT FDR at 3%
Logistic Regression (Baseline)	30 Variables out of 35 Variables	71.94%	70.94%	40.38%
Random Forest	35 Variables, Max_depth 20, 50 trees	100%	89.02%	58.16%
Neural Net	35 Variables, 3 layers with 10 nodes each, Epochs =50, learning rate = 0.1	86.76%	80.69%	53.52%
Gradient Boosting Trees	35 Variables, 1500 trees, depth=1	92.55%	86.12%	42.51%

Table 6.1 Overview of Algorithm Evaluation

a. <u>Logistic Regression</u>

The logistic regression model is used as a baseline model where it uses a logistic function to model the binary dependent variable, which is if the transaction is a fraud incident or not. Different subsets of selected variables applied to try to find the best performance result with minimum features. By building the regression model with the top 30 variables out of the pre-selected 35, the average model performance seems to do best compared with using all variables.

Model	Parameters		Average FDR (%) at 3%		
Logistic Regression	Total Variables	Parameters	Train	Test	OOT
1	30	l2 penalty, C=1	70.86%	69.41%	40.83%
2	30	l2 penalty, C=2.78	70.78%	69.30%	40.99%
3	35	l2 penalty, C=1	71.93%	70.94%	40.38%
4	35	l2 penalty, C=2.78	71.88%	70.84%	40.33%

Table 6.2 Logistic Regression Result

b. <u>Random Forest</u>

A random forest model is an ensemble model. This method builds many trees with some randomness involved, which randomizes the inputs (all features) for each tree development, and each tree is a strong learner that could serve as a full model to predict the output of whether the incident is fraud or not. Ensemble models like random forests have advantages on model robustness, stability, and generalization. The best performance of a random forest model achieved an out of time fraud detection rate at 3% of 58.16%, which is a relatively significant increase compared to the logistic model.

Random Forest	Total Variables	Parameters	Train	Test	OOT
1	30	Default	100.00%	82.87%	49.52%
2	30	20 Trees, Max Depth = 20, min_sample leaf =5	99.97%	83.56%	52.77%
3	30	25 Trees, Max Depth = 20, min_sample leaf =5	99.94%	83.39%	52.16%
4	30	30 Trees, Max Depth = 30, min_sample leaf = 5	99.99%	83.39%	52.56%
5	35	20 Trees, Max Depth = 20	99.81%	87.36%	55.81%
6	35	20 Trees, Max Depth = 20, min_sample leaf = 5	99.80%	87.47%	55.75%
7	35	50 Trees, Max Depth = 20, min_sample leaf = 5	99.99%	88.62%	57.99%
8	35	50 Trees, Max Depth = 20	100.00%	89.02%	58.16%

Table 6.3 Random Forest Result

c. Neural Net

A neural net model is a supervised machine learning algorithm. This method is a math function that maps inputs to an output with tunable parameters with vector and matrix multiplication with transfer functions. A typical neural net algorithm consists of an input layer, some number of hidden layers and an output layer. The input layer has all the independent variables (our features), and the output layer has the dependent variable identifying whether the incident is fraud.

Starting simple with only 1 layer and 1 node, the model did not seem to be effective. The best parameter chosen produces an out of time fraud detection rate at 3% cutoff of 55.47%.

Neural Net	Total Variables	Layers	Nodes	Epochs	Learning Rate	Train	Test	OOT
1	30	Default	Default	Default	Default	70.86%	69.41%	40.83%
2	35	1	1	30	0.001	47.02%	44.27%	23.52%
3	35	1	1	30	0.1	52.25%	48.54%	23.07%
4	35	1	6	30	0.001	82.01%	78.35%	51.01%
5	35	1	6	30	0.1	82.04%	77.88%	48.66%
6	35	3	10	50	0.01	78.11%	75.40%	51.12%
7	35	3	10	50	0.1	83.37%	78.28%	55.47%
8	35	3	15	50	0.01	80.33%	77.24%	53.18%
9	35	3	15	50	0.1	86.76%	80.69%	53.52%
10	30	10	8	20	0.001	80.82%	77.17%	52.97%
11	30	10	8	30	0.001	97.35%	77.69%	38.14%
12	30	20	18	30	0.001	87.25%	79.48%	55.05%

Table 6.4 Neural Net Result

d. Adaptive Boosting

Adaptive boosting is another ensemble method of supervised machine learning algorithm that develops models in sequence, where the latter model learns from the previous mistakes (classification errors, in this case). The methods build decision trees one by one and make predictions by a weighted average among all trees developed.

The best parameters from adaptive boosting provided a fraud detection rate at 3% of 55.31%, which was also better than the logistic regression baseline model.

Adaptive Boosting	Total Variables	Parameters	Train	Test	OOT
1	35	Default	77.30%	73.85%	45.66%
2	35	Trees = 20, Learning Rate = 1, Depth = 1	71.98%	69.55%	42.36%
3	35	Trees = 50, Learning Rate = 1, Depth = 1	78.93%	76.46%	42.07%
4	35	Trees = 50, Learning Rate = 0.1, Depth = 20	92.29%	85.34%	55.31%
5	35	Trees = 50, Learning Rate = 1, Depth = 20	98.99%	84.48%	43.58%
6	35	Trees = 200, Learning Rate = 1, Depth = 1	86.17%	82.62%	42.01%
7	35	Trees = 300, Learning Rate = 1, Depth = 1	88.03%	83.73%	42.40%
8	35	Trees = 400, Learning Rate = 1, Depth = 1	89.11%	84.81%	42.74%
9	35	Trees = 1500, Learning Rate = 0.1, Depth = 1	84.18%	81.27%	47.15%
10	35	Trees = 1500, Learning Rate =1, Depth = 1	92.55%	86.12%	42.51%

Table 6.5 Adaptive Boosting Result

Part VII. Results

In a strict algorithm development process, one should try not to use out of time data to make model selection. However, due to the lack of samples and unbalanced data, the analysis team found a significant discrepancy between testing data model effectiveness and out of time data model effectiveness. Therefore, the out of time fraud detection rate was used as the final selection metric for the best algorithm.

Since the random forest model with 50 trees and maximum depth of 20 yielded the highest performance of fraud detection rate at 3% cutoff at 58.16%, it was selected to be the final model. The probability cutoff using the random forest model is 0.36, which means for any row of data, if the predicted fraud score (probability of being a fraudulent incident), it will be flagged as fraud using the current algorithm. The statistics for training, testing, and out of time data is presented below.

Training - Model Performance													
Training		# of Records	56442	# of Bads	608		# of Goods		55834		Fraud Rate	0.01077212	
	Bin Statistics						Cumulative Statistics						
Population Bin %	Bin Size	Fraud_Caught	# of Goods	# of Bads	% of Goods	% of Bads	Total # of Record	Cumulative Goods	Cumulative Bads	% of Goods	% of Bads (FDR)	KS	FDR
1	565	552	13	552	2.30%	97.70%	6370	13	552	0.02%	90.79%	90.77	0.02
2	565	55	510	55	90.27%	9.73%	12740	523	607	0.94%	99.84%	98.90	0.86
3	565	1	564	1	99.82%	0.18%	19110	1087	608	1.95%	100.00%	98.05	1.79
4	565	0	565	0	100.00%	0.00%	25480	1652	608	2.96%	100.00%	97.04	2.72
5	565	0	565	0	100.00%	0.00%	31850	2217	608	3.97%	100.00%	96.03	3.65
6	565	0	565	0	100.00%	0.00%	38220	2782	608	4.98%	100.00%	95.02	4.58
7	565	0	565	0	100.00%	0.00%	44590	3347	608	5.99%	100.00%	94.01	5.50
8	565	0	565	0	100.00%	0.00%	50960	3912	608	7.01%	100.00%	92.99	6.43
9	565	0	565	0	100.00%	0.00%	57330	4477	608	8.02%	100.00%	91.98	7.36
10	565	0	565	0	100.00%	0.00%	63700	5042	608	9.03%	100.00%	90.97	8.29
11	565	0	565	0	100.00%	0.00%	70070	5607	608	10.04%	100.00%	89.96	9.22
12	565	0	565	0	100.00%	0.00%	76440	6172	608	11.05%	100.00%	88.95	10.15
13	565	0	565	0	100.00%	0.00%	82810	6737	608	12.07%	100.00%	87.93	11.08
14	565	0	565	0	100.00%	0.00%	89180	7302	608	13.08%	100.00%	86.92	12.01
15	565	0	565	0	100.00%	0.00%	95550	7867	608	14.09%	100.00%	85.91	12.94
16	565	0	565	0	100.00%	0.00%	101920	8432	608	15.10%	100.00%	84.90	13.87
17	565	0	565	0	100.00%	0.00%	108290	8997	608	16.11%	100.00%	83.89	14.80
18	565	0	565	0	100.00%	0.00%	114660	9562	608	17.13%	100.00%	82.87	15.73
19	565	0	565	0	100.00%	0.00%	121030	10127	608	18.14%	100.00%	81.86	16.66
20	565	0	565	0	100.00%	0.00%	127400	10692	608	19.15%	100.00%	80.85	17.59

Figure 7.1 Overall Training Statistics

Testing - Model Performance													
Testing		# of Records	24190	# of Bads	260		# of Goods		23930		Fraud Rate	0.01074824	
	Bin Statistics						Cumulative Statistics						
Population Bin %	Bin Size	Fraud_Caught	# of Goods	# of Bads	% of Goods	% of Bads	Total # of Record	Cumulative Goods	Cumulative Bads	% of Goods	% of Bads (FDR)	KS	FDR
1	242	184	58	184	23.97%	76.03%	1590	58	184	0.24%	70.77%	70.53	0.32
2	242	36	206	36	85.12%	14.88%	3180	264	220	1.10%	84.62%	83.51	1.20
3	242	7	235	7	97.11%	2.89%	4770	499	227	2.09%	87.31%	85.22	2.20
4	242	7	235	7	97.11%	2.89%	6360	734	234	3.07%	90.00%	86.93	3.14
5	242	10	232	10	95.87%	4.13%	7950	966	244	4.04%	93.85%	89.81	3.96
6	242	2	240	2	99.17%	0.83%	9540	1206	246	5.04%	94.62%	89.58	4.90
7	242	3	239	3	98.76%	1.24%	11130	1445	249	6.04%	95.77%	89.73	5.80
8	242	1	241	1	99.59%	0.41%	12720	1686	250	7.05%	96.15%	89.11	6.74
9	242	1	241	1	99.59%	0.41%	14310	1927	251	8.05%	96.54%	88.49	7.68
10	242	1	241	1	99.59%	0.41%	15900	2168	252	9.06%	96.92%	87.86	8.60
11	242	1	241	1	99.59%	0.41%	17490	2409	253	10.07%	97.31%	87.24	9.52
12	242	1	241	1	99.59%	0.41%	19080	2650	254	11.07%	97.69%	86.62	10.43
13	242	0	242	0	100.00%	0.00%	20670	2892	254	12.09%	97.69%	85.61	11.39
14	242	0	242	0	100.00%	0.00%	22260	3134	254	13.10%	97.69%	84.60	12.34
15	242	0	242	0	100.00%	0.00%	23850	3376	254	14.11%	97.69%	83.58	13.29
16	242	0	242	0	100.00%	0.00%	25440	3618	254	15.12%	97.69%	82.57	14.24
17	242	0	242	0	100.00%	0.00%	27030	3860	254	16.13%	97.69%	81.56	15.20
18	242	1	241	1	99.59%	0.41%	28620	4101	255	17.14%	98.08%	80.94	16.08
19	242	0	242	0	100.00%	0.00%	30210	4343	255	18.15%	98.08%	79.93	17.03
20	242	0	242	0	100.00%	0.00%	31800	4585	255	19.16%	98.08%	78.92	17.98

Figure 7.2 Overall Testing Statistics

Out of Time		# of Records	12427		# of Bads	179		# of Goods	12248		Fraud Rate	0.01440412	
	Bin Statistics						Cumulative Statistics						
Population Bin %	Bin Size	Fraud_Caught	# of Goods	# of Bads	% of Goods	% of Bads	Total # of Record	Cumulative Goods	Cumulative Bads	% of Goods	% of Bads (FDR)	KS	FPR
1	124	55	69	55	55.65%	44.35%	124	69	55	0.56%	30.73%	30.16	1.25
2	124	39	85	39	68.55%	31.45%	248	154	94	1.26%	52.51%	51.26	1.64
3	124	12	112	12	90.32%	9.68%	372	266	106	2.17%	59.22%	57.05	2.51
4	124	7	117	7	94.35%	5.65%	496	383	113	3.13%	63.13%	60.00	3.39
5	124	7	117	7	94.35%	5.65%	620	500	120	4.08%	67.04%	62.96	4.17
6	124	4	120	4	96.77%	3.23%	744	620	124	5.06%	69.27%	64.21	5.00
7	124	1	123	1	99.19%	0.81%	868	743	125	6.07%	69.83%	63.77	5.94
8	124	0	124	0	100.00%	0.00%	992	867	125	7.08%	69.83%	62.75	6.94
9	124	1	123	1	99.19%	0.81%	1116	990	126	8.08%	70.39%	62.31	7.86
10	124	1	123	1	99.19%	0.81%	1240	1113	127	9.09%	70.95%	61.86	8.76
11	124	3	121	3	97.58%	2.42%	1364	1234	130	10.08%	72.63%	62.55	9.49
12	124	3	121	3	97.58%	2.42%	1488	1355	133	11.06%	74.30%	63.24	10.19
13	124	5	119	5	95.97%	4.03%	1612	1474	138	12.03%	77.09%	65.06	10.68
14	124	3	121	3	97.58%	2.42%	1736	1595	141	13.02%	78.77%	65.75	11.31
15	124	4	120	4	96.77%	3.23%	1860	1715	145	14.00%	81.01%	67.00	11.83
16	124	4	120	4	96.77%	3.23%	1984	1835	149	14.98%	83.24%	68.26	12.32
17	124	0	124	0	100.00%	0.00%	2108	1959	149	15.99%	83.24%	67.25	13.15
18	124	1	123	1	99.19%	0.81%	2232	2082	150	17.00%	83.80%	66.80	13.88
19	124	2	122	2	98.39%	1.61%	2356	2204	152	17.99%	84.92%	66.92	14.50
20	124	1	123	1	99.19%	0.81%	2480	2327	153	19.00%	85.47%	66.48	15.21

Figure 7.3 Overall Out of Time Statistics

After the overview of the algorithm performance, here are some examples of fraud identifications.

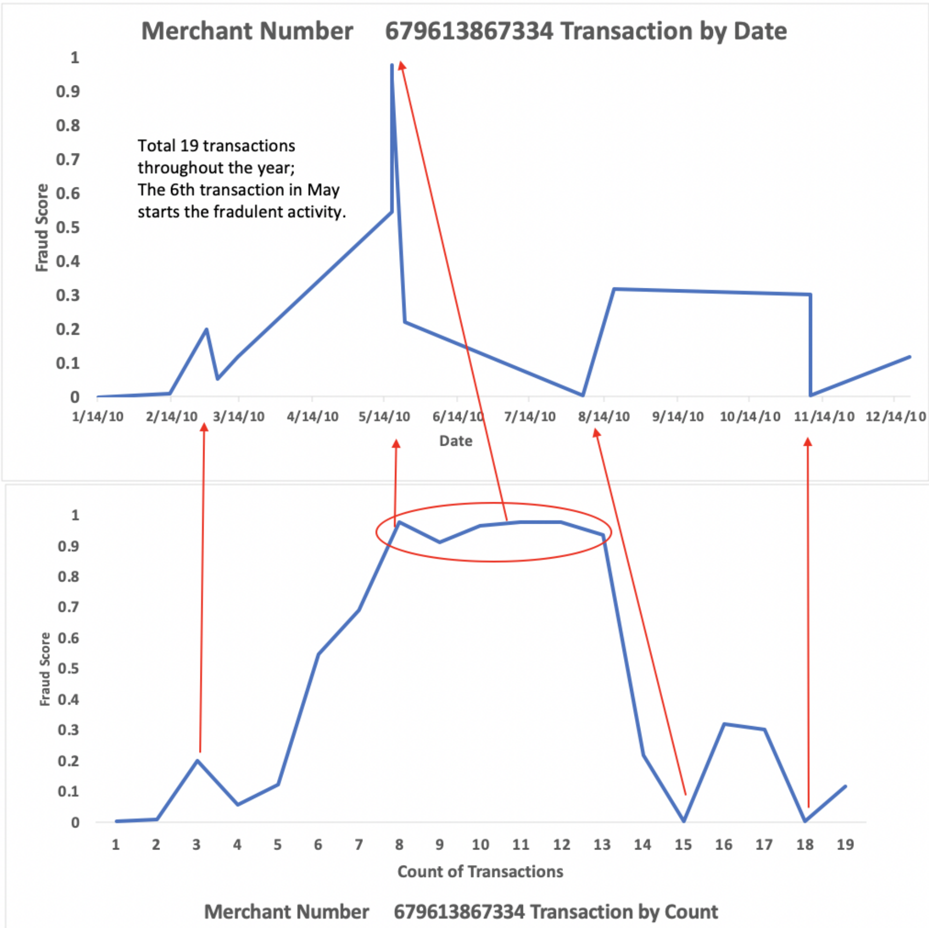


Figure 7.4 Merchant Transaction Example

The above Figure 7.4 is an example of the suspicious transactions identified by the algorithm. This merchant has a total of 19 transactions out of the year. It starts with some normal transactions in January, and then from the 6th transaction starting in May, frequent fraudulent transactions show up and the fraud score increases sharply. Then starting in June, the merchant does not seem to have any more red-flag fraud charges, and from an investigation prospective, it might be one specific customer purchase at this merchant. The next step suggested for graphs like this is to identify the most frequently used card number within the high-risk period, which is from May to June, which this algorithm also would be able to flag.

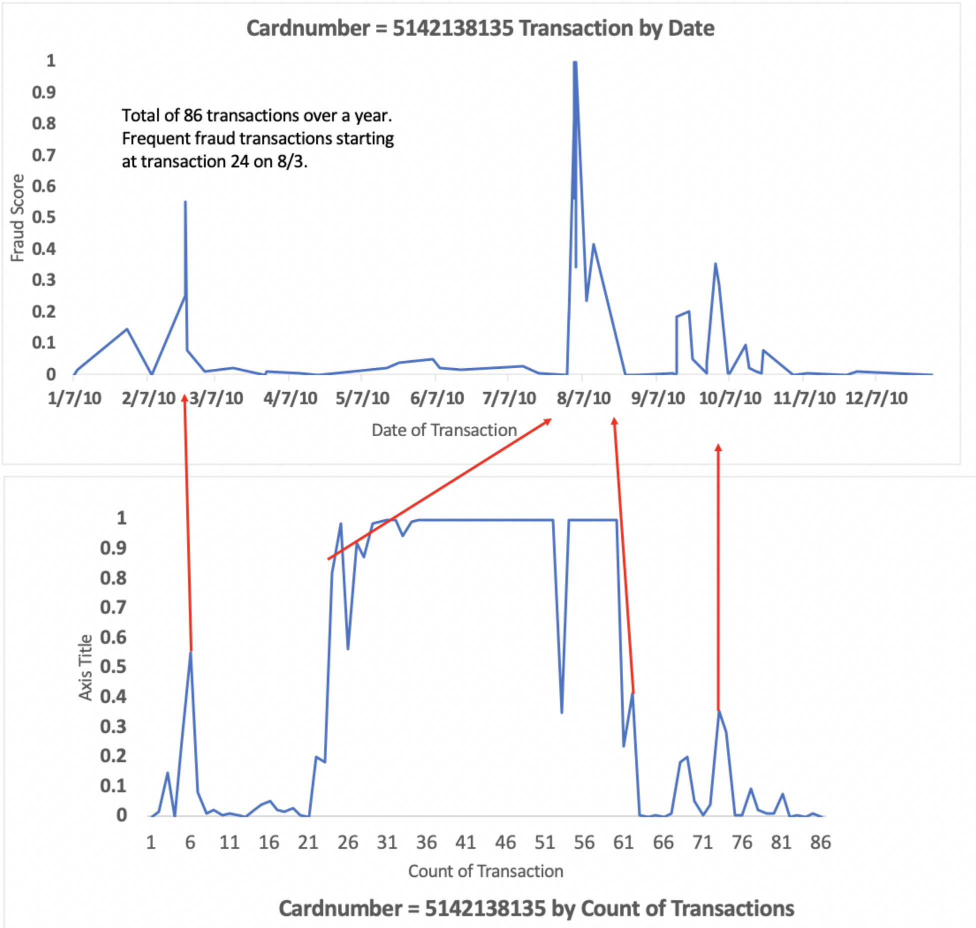


Figure 7.5 Card Number Fraud Example

The graph above shows how the algorithm flags suspicious transactions with credit card numbers. This specific credit card makes 86 transactions in the year, and it’s fairly normal at the beginning and because there are less transactions, the fraud score is relatively low at 0.5-0.6. But in August the fraud score sharply increases and the count of transactions spiked, resulting in a lot of fraud transactions being identified. This large bunch of transactions made the fraud score peak and from the 22nd transaction to almost 50th, the score is around 1, indicating the model suggested high risk or fraudulent transactions on this card. Then at roughly 61, the transactions went back to normal and fraud scores became low.

Last but not least, it is critical to quantify the possible benefit this algorithm can bring. The analysis team uses the assumption that for each fraud incident correctly identified, the client will gain $2,000, and for incidents identified incorrectly, the client loses $50. According to the current algorithm, based on the out of time model performance, the maximum gain for the client will result in $217,000 every two months with a recommended fraud detection rate at 6% instead of 3%. Converting the two-month savings to an annual base will be $1.3 million.

A detailed graph showing how the saving progresses with cost and benefit analysis is available below. The total savings increases as cutoff points move from the beginning, but then decreases because less fraud is caught and now clients will sacrifice more good customers to identify an additional fraud incident.

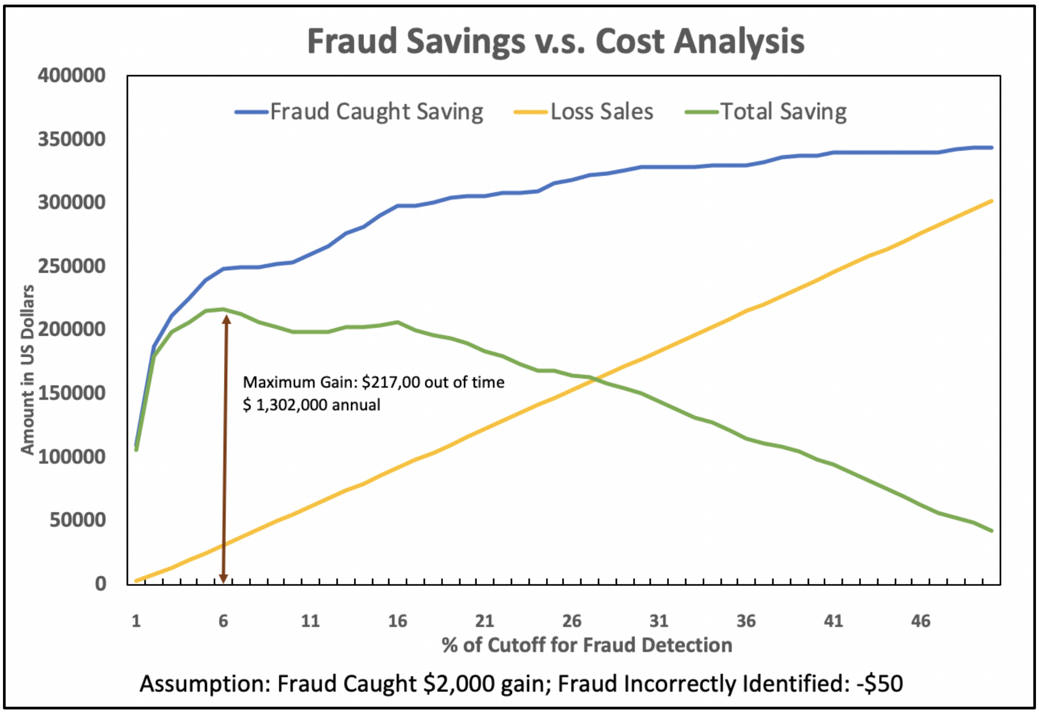


Figure 7.6 Fraud Savings vs. Cost Analysis

Part VIII. Conclusion

In conclusion, the random forest model with 50 trees and maximum depths at 20 was chosen as the final model as it yields the highest performance of fraud detection rate at 3% cutoff of 58.16%. This final model was reached through first conducting an analysis of the data, performing data cleaning, creating variables, selecting features, building various models and analyzing their performance, then finally using that analysis to select the final model.

For data cleaning, one outlier with a large transaction amount was removed, as were all transactions that were not of type P. Missing variables in merchant state, merchant zip, and merchant number were also filled in according to the mode value of corresponding groups. With the data cleaned, variables were then created to help build models in later steps. Before creating variables, three linking identities were first formed to aid variable creation. From those three linking identities and the Cardnum and Merchnum variables, 326 expert variables were formed. This included velocity variables, relative velocity variables, amount variables, days since variables, Benford's law variables, and a risk variable.

After creating 326 variables, feature selection was then performed. A filter method that used a fraud detection rate and the Kolmogorov-Smirnov test then narrowed down to the top 80 variables. Then recursive feature selection was executed through using different results were used in each run. This also included stepwise regression with the two additional Benford's law variables. From these feature selection methods, a final list of the top 35 variables was then created.

Before running various models and analyzing their performance, SMOTE (synthetic minority oversampling technique) was applied onto the data to handle the imbalance between good and bad fraud scores. Additionally, for model development the Z-scaled data was capped at a maximum of ten.

Four total algorithms were applied to the data: logistic regression, random forest, neural net, and adaptive boosting. The best performance of the logistic regression model was achieved using all the variables and an L2 penalty of 2.78, resulting in an out of time fraud detection rate at 3% of 40.99%. The best performance of random forest achieved an out of time fraud detection rate at 3% of 58.16% when using 50 trees with a max depth of 20, which is a relatively significant increase compared with the logistic model. In comparison, the best neural net model chosen produces an out of time fraud detection rate at 3% cutoff of 55.47% when the model had three layers, ten nodes, 50 epochs, and a learning rate of 0.1. Finally, the adaptive boosting model had the best out of time fraud detection rate at 3% of 55.31% when it used 50 trees with a depth of 20 and learning rate of 0.1.

From these models, the random forest model with 50 trees with a max depth of 20 was selected as the final model as it produced the highest out of time fraud detection rate at 3% with 58.16%. Additionally, the probability cutoff using the random forest model is 0.36.

For future steps, it will be beneficial to try more models such as support vector machine (SVM) models, which might be helpful in fraud classification. Also, more expert variables could be built if the team could have a discussion with domain experts, as there are some expert variables that can only be created if one is or knows an expert on the subject. The current estimated savings are based on assumptions of fixed cost and benefit per fraud identification or misidentification. Thus, a cost benefit analysis tool that allows clients to change assumptions about benefits and cost to see how that would affect the recommended cutoff will add additional value.

Appendix

USC Marshall
School of Business

Fraud Detection on Credit Card Transaction Data

UNDER THE GUIDANCE OF

Professor Stephen Coggeshall

5/03/2020

JENNY SHANG |

JENNY WANG |

LINGROU WANG |

RORY WANG |

Table of Contents

Part IX. Data Quality Report

Description of Data

This report provides a background analysis of the “credit card transactions data” dataset, including summary statistics and visualizations of each of the fields contained in the dataset. The “credit card transactions data” is a dataset consisting of the time, place and other basic information of credit card transactions made by a government department in Tennessee. The label of whether a record is fraudulent or not helps in developing a supervised fraud detection model, which will be built in future stages. The data is provided and slightly modified by Professor Coggeshall for academic purpose only.

In this dataset, all the fraudulent records are generated by Professor Coggeshall based on real-life fraud cases scenario. In summary, this dataset contains a total of 10 fields and 96,753 records. All records are generated with dates in 2010, from January 1, 2010 to December 31, 2010.

Summary of Fields

a. **Numerical Fields**

Data Field	Num Records w/ a value	Percent Populated	Num Unique Values	Num Records w/ a value of Zero	Mean	Std Dev	Min	Max
Amount	96,753	100%	34,909	0	427.89	10,006	0.01	3,102,046

b. **Categorical Fields**

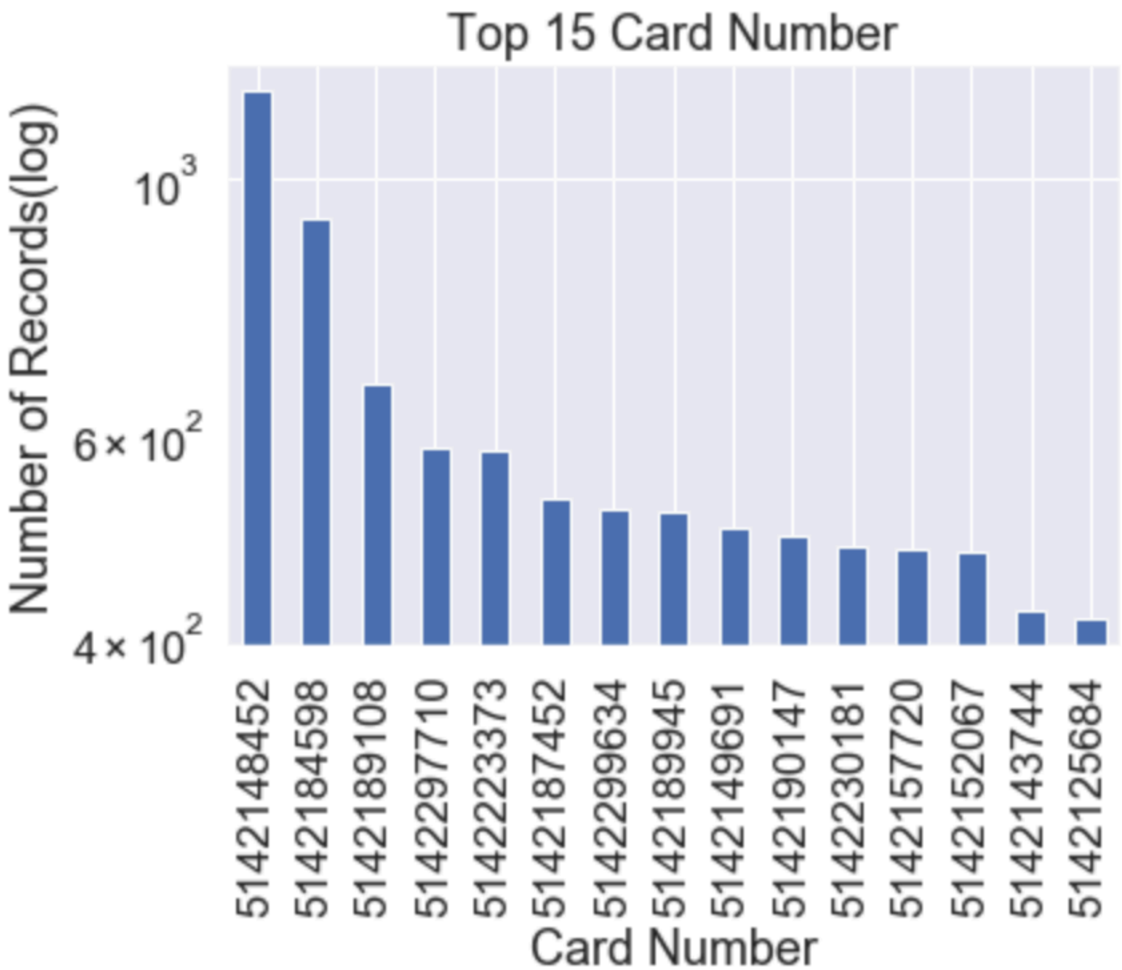
Data Field	Num Records w/ a value	Percent Populated	Num Unique Values	Most Common Value
Recnum	96,753	100%	96,753	N/A
Cardnum	96,753	100%	1,645	5142148452
Date	96,753	100%	365	2010-02-28
Merchnum	93,378	96.5%	13,092	930090121224
Merch description	96,753	100%	13,126	GSA-FSS-ADV
Merch state	95,558	98.8%	228	TN
Merch zip	92,097	95.1%	4,568	38118
Transtype	96,753	100%	4	P
Fraud	96,753	100%	2	0

Description of Fields

a. **Field Name**: Recnum
   Description: A categorical data field containing an integer representing the unique record number identifier from 1 to 96,753. All records in the dataset contain a record number.

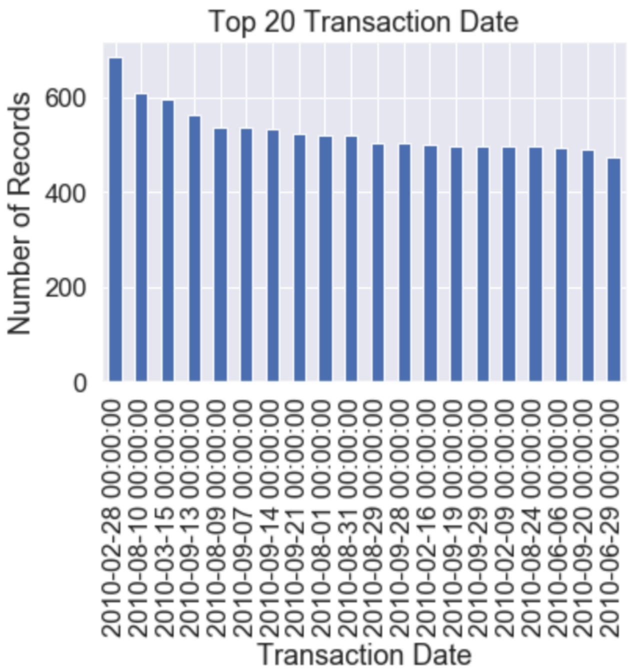
b. **Field Name**: Cardnum
   Description: A categorical data field containing a string of number representing the credit card number making the transaction. All records in the dataset contain a credit card number.

   The bar chart below provides the top 15 credit card number with the most records in the dataset.

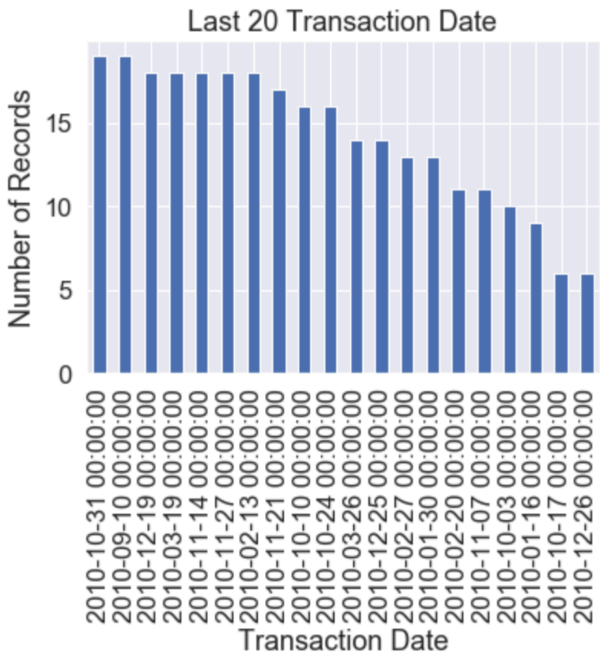


c. **Field Name**: Date
Description: A categorical data field containing year, month and day representing the date the transaction taken place, from 2010-01-01 to 2010-12-31. All records in the dataset contain a date.

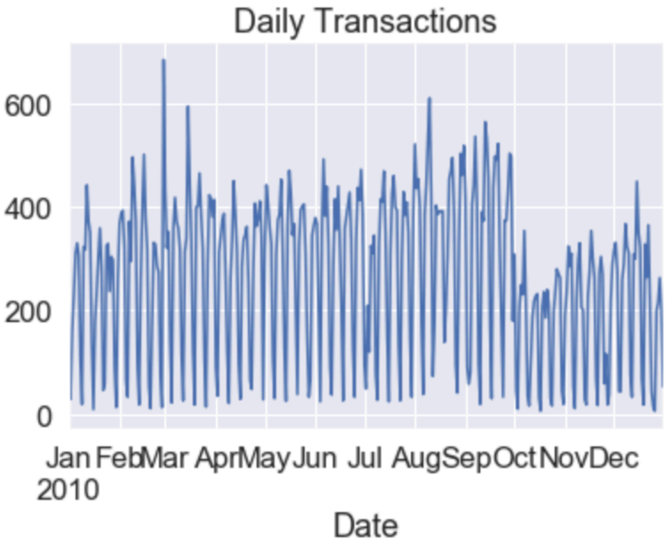
The bar chart below provides the top 20 date with the most records in the dataset.



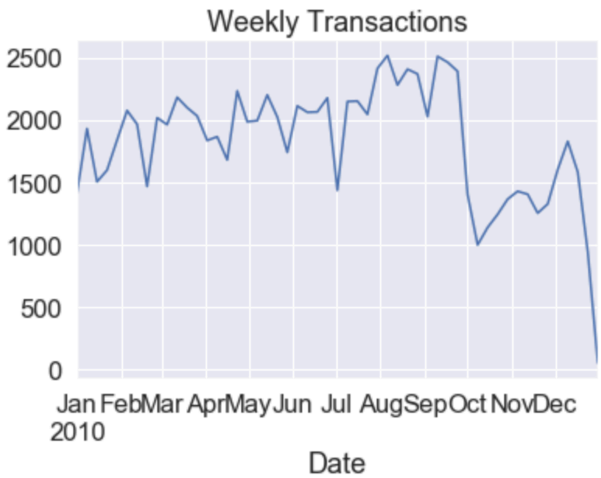
The bar chart below provides the last 20 date with the least records in the dataset.

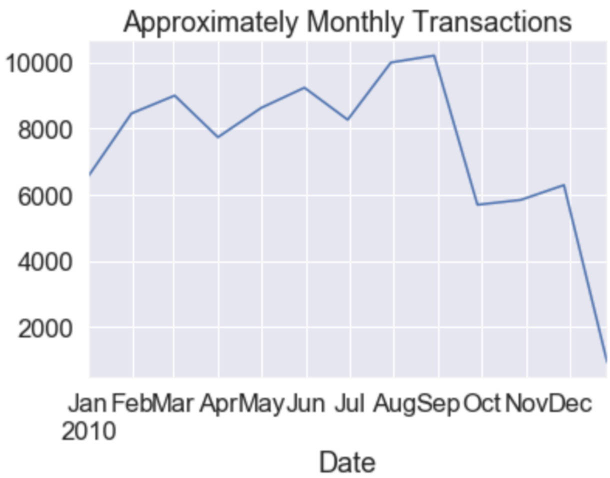


The line plot below provides the number of transactions happen everyday in 2010. We can clearly see the difference between weekday records and weekend records in this dataset.



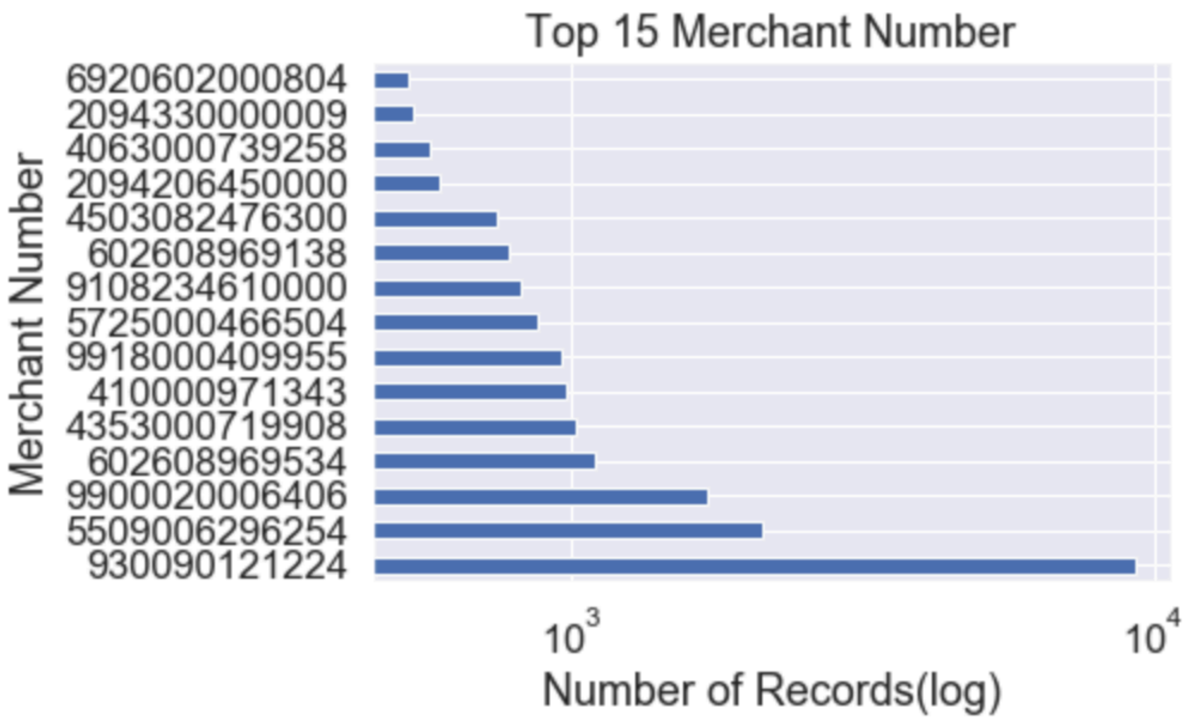
Two line-plots below provide the number of transactions happen weekly and monthly in 2010. We can clearly see the seasonal pattern of transactions made by the government department.





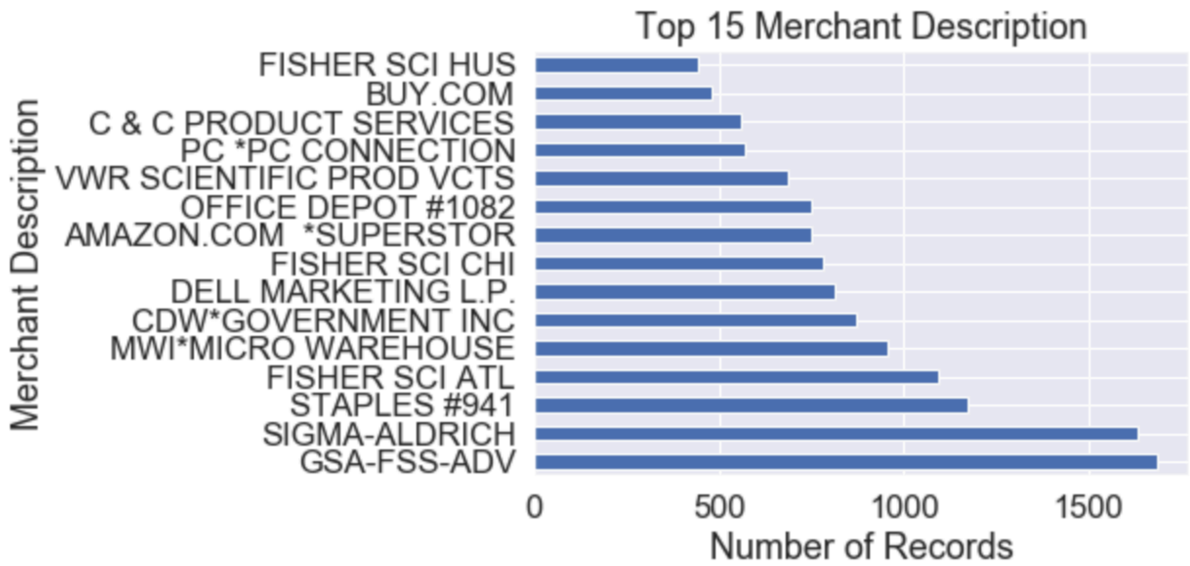
d. **Field Name**: Merchnum
Description: A categorical data field containing the number representing the merchant bought in transactions. Only 96.5% records in the dataset contain a merchant number.

The bar chart below provides the top 15 merchant numbers with the most records in the dataset.



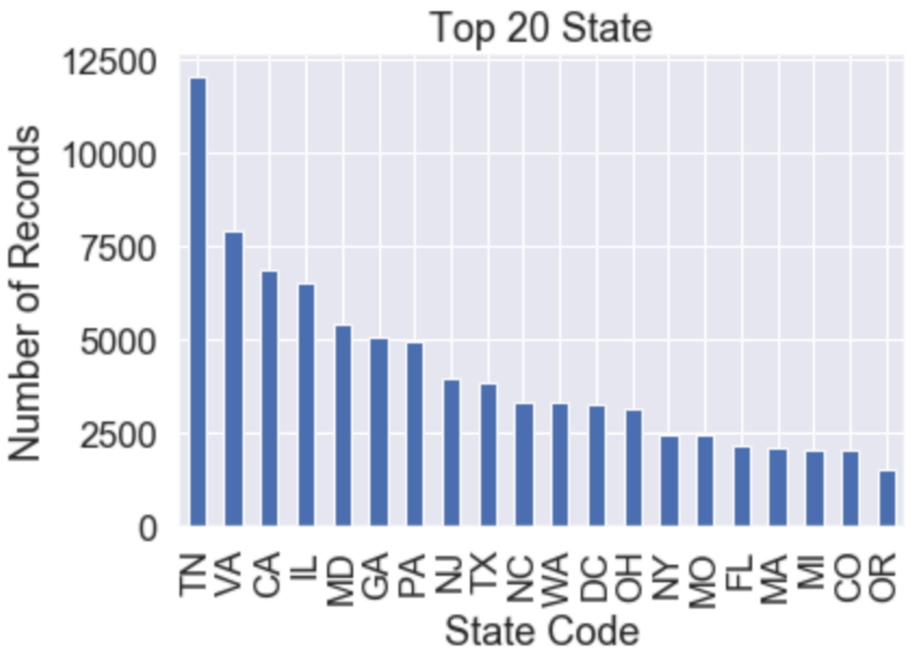
e. **Field Name**: Merch description
Description: A categorical data field containing the description of the merchant bought in the transaction. All records in the dataset contain a merchant description.

The bar chart below provides the top 15 merchant descriptions with the most records in the dataset.



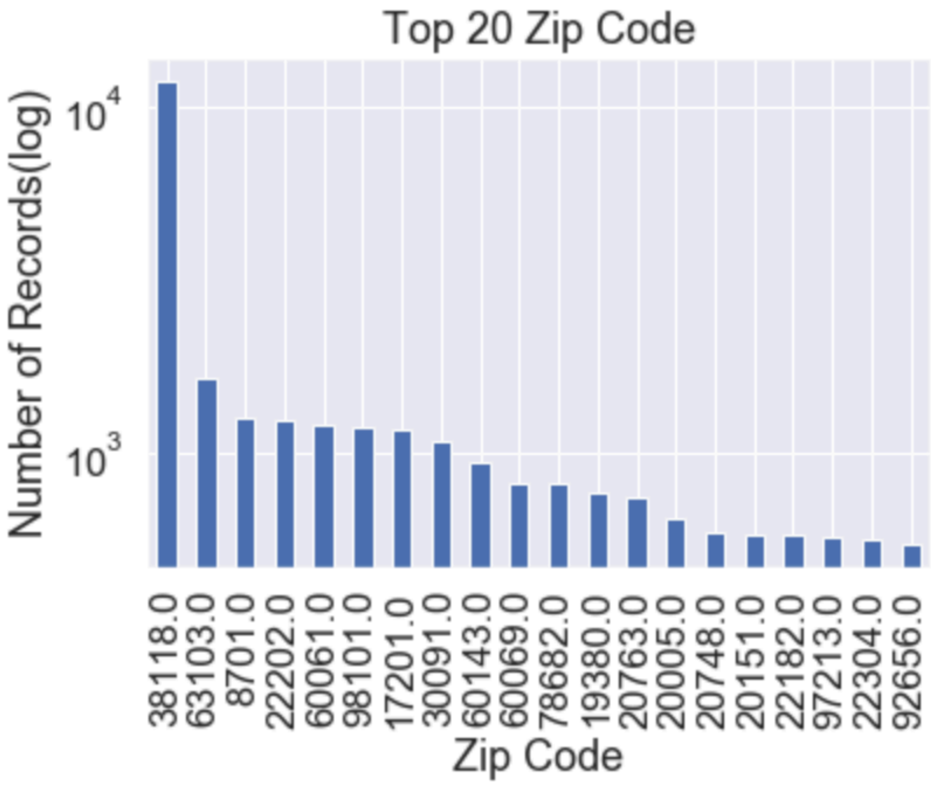
f. **Field Name**: Merch state
Description: A categorical data field containing the state code representing the state where the transaction was taken place. Only 98.8% records in the dataset contain a merchant state data.

The bar chart below provides the top 20 states with the most records in the dataset.



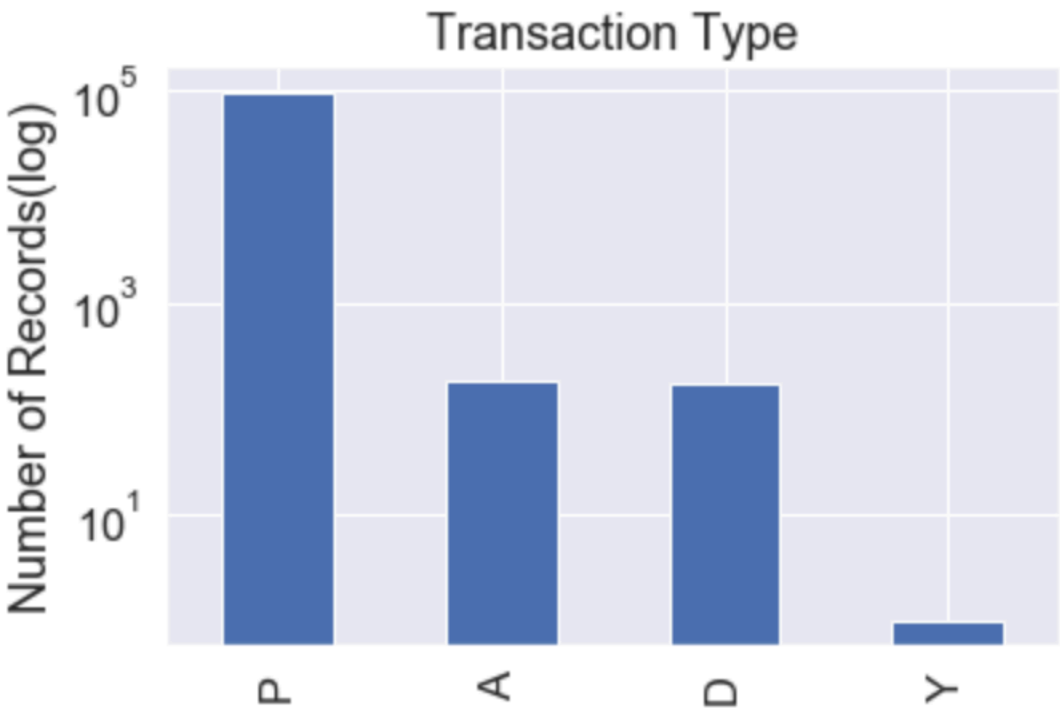
g. **Field Name**: Merch zip
Description: A categorical data field containing the 5 digits zip code representing the place where the transaction was taken place. Only 95.1% records in the dataset contain a merchant zip code.

The bar chart below provides the top 20 five digits zip codes with the most records in the dataset.



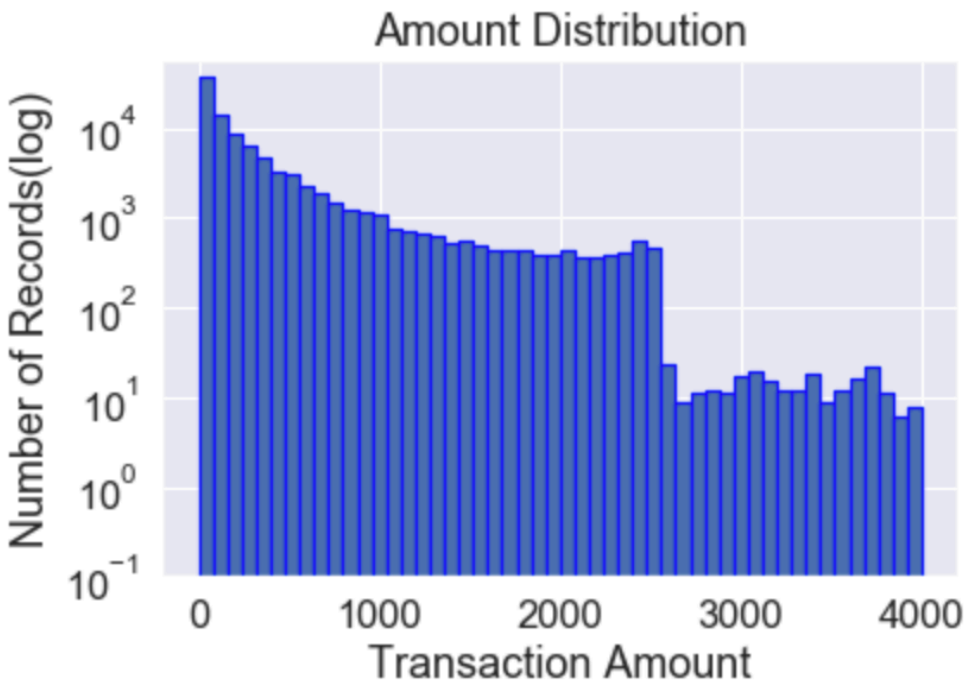
h. **Field Name**: Transtype
Description: A categorical data field containing a character representing the type of transaction. All records in the dataset contain a transaction type.

The bar chart below provides the number of records of each transaction type in the dataset.



i. **Field Name**: Amount
Description: A numerical data field containing the amount of the transaction. All records in the dataset contain the data of the transaction amount.

The plot below provides the distribution of all amount paid in transactions. We can see a clear drop around $2,500.



j. **Field Name**: fraud_label
Description: A categorical data field containing 0 or 1 representing the possibility of that certain application being a potential fraud. All records in the dataset contain a number.

The bar chart below provides the number of 0 and 1 in the dataset.

Part X. List of Variables Created

Variable Name	Variable Name
CardMerchanttot_AMT7	CardMerchantactmed_AMT3
CardZiptot_AMT7	CardStateactmed_AMT3
CardZiptot_AMT3	CardMerchantactmed_AMT1
CardMerchanttot_AMT14	CardZipactmed_AMT1
CardMerchanttot_AMT3	CardZipacttot_AMT7
CardStatetot_AMT3	CardMerchantactavg_AMT1
CardStatetot_AMT7	CardZip_days_since_last_seen
CardZiptot_AMT14	Cardnumactmax_AMT30
CardStatetot_AMT1	Cardnum_1_amount_Cardnum_7_amount_ave
CardMerchanttot_AMT1	CardStateactavg_AMT30
CardMerchanttot_AMT30	Cardnum_1_amount_Merchnum_7_amount_ave
CardZiptot_AMT1	CardMerchantactavg_AMT3
CardStatetot_AMT14	Cardnumacttot_AMT0
CardZiptot_AMT30	CardStateactavg_AMT14
CardStatetot_AMT0	CardMerchantactavg_AMT0
CardZipmax_AMT14	CardStateactmed_AMT7
CardMerchanttot_AMT0	CardStateacttot_AMT7
CardMerchantmax_AMT14	CardStateactavg_AMT7
CardStatemax_AMT7	CardStateactavg_AMT3
CardZiptot_AMT0	CardZipactmed_AMT3
CardZipmax_AMT3	CardMerchantactavg_AMT7
CardZipmax_AMT7	CardZipactmax_AMT7
CardMerchantmax_AMT30	CardMerchantactmed_AMT0
CardStatemax_AMT14	Cardnum_0_amount_Cardnum_30_amount_ave
Cardnumtot_AMT3	Cardnum_0_amount_Merchnum_30_amount_ave
CardStatemax_AMT3	Merchnumactavg_AMT0
CardMerchantmax_AMT3	CardStateactmed_AMT1
CardMerchantmax_AMT7	Merchustar
CardZipmax_AMT30	CardMerchantacttot_AMT14
CardMerchantavg_AMT3	CardZipactavg_AMT30
CardZipavg_AMT3	CardStateacttot_AMT14
Merchnummax_AMT7	Cardnumactmax_AMT0
CardStateavg_AMT0	Merchnum_1_amount_Merchnum_14_amount_ave
CardZipavg_AMT0	Merchnum_1_amount_Cardnum_14_amount_ave
CardZipavg_AMT7	dayofweek_risk
CardStateavg_AMT1	Cardnumactmed_AMT1
Cardnumavg_AMT0	Merchnum_1_amount_Merchnum_30_amount_ave
CardStateavg_AMT14	Merchnum_1_amount_Cardnum_30_amount_ave
CardZipavg_AMT1	Cardnum_count3_Date
CardMerchantavg_AMT30	Cardnumactmax_AMT14
CardMerchantavg_AMT14	Merchnum_0_amount_Cardnum_30_amount_ave
CardStateavg_AMT30	Merchnum_0_amount_Merchnum_30_amount_ave
CardZipavg_AMT14	Cardnumactmax_AMT1
CardMerchantavg_AMT7	Cardnumactavg_AMT0

CardMerchantavg_AMT1	CardState_count3_Date
Cardnummed_AMT1	CardZipactmed_AMT30
Cardnumavg_AMT7	Cardnumacttot_AMT3
Merchnumavg_AMT1	CardState_count1_Date
Cardnumavg_AMT14	Cardnum_count1_Date
Merchnummax_AMT14	CardMerchant_count3_Date
CardMerchantmed_AMT0	CardState_count14_Date
CardStatemed_AMT0	CardZip_count3_Date
Cardnummed_AMT0	CardMerchant_count14_Date
CardZipmed_AMT0	Cardnumactmax_AMT3
Cardnumavg_AMT30	CardState_count7_Date
Cardnummed_AMT3	CardMerchantactmax_AMT30
Cardnummax_AMT30	CardMerchant_count7_Date
Cardnumtot_AMT30	CardMerchant_count1_Date
CardStatemed_AMT1	CardZip_count14_Date
CardZipmed_AMT1	Cardnum_0_amount_Merchnum_14_amount_ave
CardZipmed_AMT3	Cardnum_0_amount_Cardnum_14_amount_ave
CardMerchantmed_AMT1	CardZip_count1_Date
CardMerchantmed_AMT30	Cardnum_0_amount_Merchnum_7_amount_ave
CardMerchantmed_AMT3	Cardnum_0_amount_Cardnum_7_amount_ave
Cardnummed_AMT14	Merchnumacttot_AMT3
CardStatemed_AMT30	Merchnum_count7_Date
CardStatemed_AMT3	Cardnum_count7_Date
Cardnummed_AMT7	Merchnum_count3_Date
Merchnumavg_AMT3	CardZip_count7_Date
CardMerchantmed_AMT14	CardStateactmax_AMT14
CardStatemed_AMT7	CardState_count30_Date
CardStatemed_AMT14	Merchnumacttot_AMT1
CardZipmed_AMT7	CardMerchant_count30_Date

CardMerchantmed_AMT7	CardZipacttot_AMT30
Cardnummed_AMT30	Merchnumactmax_AMT30
CardZipmed_AMT30	Cardnumacttot_AMT30
Merchnummed_AMT0	Merchnumactmax_AMT14
CardZipmed_AMT14	CardZip_count30_Date
Merchnumavg_AMT7	Cardnum_count14_Date
Merchnumavg_AMT14	Cardnum_count0_Date
Merchnummed_AMT1	CardState_count0_Date
Merchnummed_AMT3	CardMerchant_count0_Date
Merchnumtot_AMT14	Merchnumacttot_AMT0
Merchnummax_AMT30	Merchnum_1_amount_Cardnum_7_amount_ave
Merchnumavg_AMT30	Merchnum_1_amount_Merchnum_7_amount_ave
Merchnummed_AMT7	CardZip_count0_Date
Merchnummed_AMT14	Merchnum_count1_Date
Merchnumactavg_AMT30	Merchnumacttot_AMT7
Merchnummed_AMT30	Merchnum_count0_Date
Merchnumtot_AMT30	Cardnumacttot_AMT7
Merchnumactmed_AMT30	CardZipactmax_AMT30

CardStateacttot_AMT1	Merchnum_0_amount_Merchnum_14_amount_ave
Merchnumactavg_AMT14	Merchnum_0_amount_Cardnum_14_amount_ave
Merchnumactavg_AMT7	Merchnumactmax_AMT7
CardMerchantacttot_AMT1	CardStateactmax_AMT30
Merchnumactmed_AMT14	Merchnum_0_amount_Merchnum_7_amount_ave
CardMerchantacttot_AMT3	Merchnum_0_amount_Cardnum_7_amount_ave
CardZipacttot_AMT1	Merchnum_0_count_Merchnum_7_count
Merchnumactmed_AMT7	Merchnum_0_count_Cardnum_7_count
CardState_days_since_last_seen	Merchnum_count30_Date
CardZipacttot_AMT3	Merchnum_0_count_Cardnum_30_count
Cardnumactavg_AMT30	Merchnum_0_count_Merchnum_30_count
Cardnum_1_amount_Cardnum_30_amount_ave	Merchnum_1_count_Cardnum_30_count
Cardnum_1_amount_Merchnum_30_amount_ave	Merchnum_1_count_Merchnum_30_count
Merchnumactavg_AMT3	Merchnum_count14_Date
CardMerchant_days_since_last_seen	Cardnum_count30_Date
CardMerchantactmax_AMT1	Cardnum_0_count_Cardnum_7_count
Cardnumactmed_AMT30	Cardnum_0_count_Merchnum_7_count
CardStateacttot_AMT3	CardStateacttot_AMT30
CardStateacttot_AMT0	Merchnum_0_count_Merchnum_14_count
CardZipactmax_AMT1	Merchnum_0_count_Cardnum_14_count
CardMerchantacttot_AMT0	Cardnum_1_count_Cardnum_7_count
CardMerchantactmax_AMT3	Cardnum_1_count_Merchnum_7_count
Merchnumactmed_AMT3	Merchnumactmax_AMT1
CardZipacttot_AMT0	Merchnum_1_count_Merchnum_7_count
CardStateactmax_AMT1	Merchnum_1_count_Cardnum_7_count
CardZipactmax_AMT3	Cardnumactmax_AMT7
CardMerchantactmax_AMT0	Merchnumacttot_AMT30
Merchnumactavg_AMT1	Merchnumacttot_AMT14
CardZipactmax_AMT0	Merchnumactmax_AMT0
Cardnumactavg_AMT14	Merchnum_1_count_Cardnum_14_count
CardStateactmax_AMT0	Merchnum_1_count_Merchnum_14_count
Cardnumactmed_AMT14	Merchnumactmax_AMT3
CardMerchantacttot_AMT7	Cardnum_1_count_Merchnum_30_count
Cardnumactavg_AMT7	Cardnum_1_count_Cardnum_30_count
Cardnumactmed_AMT7	Cardnumacttot_AMT14
CardMerchantactmed_AMT7	Cardnum_1_count_Merchnum_14_count
CardZipactmed_AMT7	Cardnum_1_count_Cardnum_14_count
Merchnumactmed_AMT1	Cardnum_0_count_Merchnum_30_count
CardMerchantactmax_AMT7	Cardnum_0_count_Cardnum_30_count
CardStateactmax_AMT3	Cardnum_0_count_Merchnum_14_count
CardMerchantactmed_AMT14	Cardnum_0_count_Cardnum_14_count

Merchnumtot_AMT1	CardMerchantactavg_AMT14
Cardnumtot_AMT7	CardZipactavg_AMT0
Merchnumtot_AMT0	CardZipactavg_AMT1
CardStatetot_AMT30	CardStateactavg_AMT1
CardMerchantmax_AMT1	Merchnum_days_since_last_seen
CardZipmax_AMT1	CardZipactmed_AMT14
CardStatemax_AMT30	CardStateactmed_AMT14
CardStatemax_AMT1	CardMerchantactmed_AMT30
Merchnumtot_AMT3	CardZipactavg_AMT7
Merchnummax_AMT0	Cardnumactavg_AMT3
Cardnumtot_AMT1	CardMerchantactmax_AMT14
Cardnumtot_AMT0	CardZipactavg_AMT3
CardStatemax_AMT0	CardMerchantactavg_AMT30
CardZipmax_AMT0	Cardnum_1_amount_Merchnum_14_amount_ave
Merchnummax_AMT1	Cardnum_1_amount_Cardnum_14_amount_ave
CardMerchantmax_AMT0	Cardnumacttot_AMT1
Merchnummax_AMT3	CardZipactmed_AMT0
Cardnummax_AMT0	CardZipactavg_AMT14
Merchnumtot_AMT7	CardStateactavg_AMT0
Cardnummax_AMT7	CardStateactmed_AMT30
Cardnummax_AMT1	CardStateactmax_AMT7
Cardnummax_AMT3	Cardnum_days_since_last_seen
Cardnummax_AMT14	CardStateactmed_AMT0
CardStateavg_AMT7	Cardnumactmed_AMT3
Cardnumtot_AMT14	CardZipacttot_AMT14
CardStateavg_AMT3	CardMerchantacttot_AMT30
Cardnumavg_AMT1	CardUstar
Merchnumavg_AMT0	Merchnumactmed_AMT0
Cardnumavg_AMT3	CardZipactmax_AMT14
CardMerchantavg_AMT0	Cardnumactmed_AMT0
CardZipavg_AMT30	Cardnumactavg_AMT1

Part XI. List of the rankings of variables

VariableName	ks	FDR	rank_ks	rank_FDR	average_rank
Fraud	1	1	327	327	327
CardMerchanttot_AMT7	0.6833189	0.63364055	326	324.5	325.25
CardZiptot_AMT7	0.68234425	0.63364055	325	324.5	324.75
CardZiptot_AMT3	0.6761578	0.63824885	322	326	324
CardMerchanttot_AMT14	0.67801443	0.63248848	324	323	323.5
CardMerchanttot_AMT3	0.67697311	0.63133641	323	322	322.5
CardStatetot_AMT3	0.6755116	0.63018433	321	321	321
CardStatetot_AMT7	0.67081496	0.60023041	320	318	319
CardZiptot_AMT14	0.66823027	0.61866359	318	320	319
CardStatetot_AMT1	0.66035438	0.60599078	316	319	317.5
CardMerchanttot_AMT1	0.65989069	0.59907834	315	317	316
CardMerchanttot_AMT30	0.66192376	0.56221198	317	313	315
CardZiptot_AMT1	0.65840144	0.59677419	314	316	315
CardStatetot_AMT14	0.66991698	0.52419355	319	306	312.5
CardZiptot_AMT30	0.65428296	0.52764977	310	307	308.5
CardStatetot_AMT0	0.61165045	0.56336406	295	314	304.5
CardZipmax_AMT14	0.65513651	0.47695853	311	297.5	304.25
CardMerchanttot_AMT0	0.61258882	0.56105991	296	312	304
CardMerchantmax_AMT14	0.65572252	0.47465438	313	294	303.5
CardStatemax_AMT7	0.64757729	0.48963134	304	301.5	302.75
CardZiptot_AMT0	0.61016918	0.55760369	294	311	302.5
CardZipmax_AMT3	0.64841628	0.47695853	306	297.5	301.75
CardZipmax_AMT7	0.65534051	0.46658986	312	290	301
CardMerchantmax_AMT30	0.6520887	0.4735023	309	292	300.5
CardStatemax_AMT14	0.63107654	0.48847926	301	300	300.5
Cardnumtot_AMT3	0.60220924	0.55299539	289	310	299.5
CardStatemax_AMT3	0.64841877	0.4735023	307	292	299.5
CardMerchantmax_AMT3	0.64547674	0.47580645	303	295.5	299.25
CardMerchantmax_AMT7	0.65156497	0.46543779	308	289	298.5
CardZipmax_AMT30	0.6476149	0.4735023	305	292	298.5
Merchnumtot_AMT1	0.60964425	0.49078341	293	303	298
Cardnumtot_AMT7	0.60024515	0.51843318	287	305	296
Merchnumtot_AMT0	0.58327331	0.56682028	271	315	293
CardStatetot_AMT30	0.63631521	0.44585253	302	283.5	292.75
CardMerchantmax_AMT1	0.62131473	0.45737327	298	287	292.5
CardZipmax_AMT1	0.62382935	0.4562212	299	285.5	292.25
CardStatemax_AMT30	0.59871974	0.4781106	285	299	292
CardStatemax_AMT1	0.62612021	0.44239631	300	281	290.5
Merchnumtot_AMT3	0.61693014	0.42165899	297	277	287
Merchnummax_AMT0	0.60898689	0.44470046	292	282	287
Cardnumtot_AMT1	0.57702867	0.54493088	266	308	287
Cardnumtot_AMT0	0.57087978	0.55184332	259	309	284
CardStatemax_AMT0	0.60268304	0.41935484	290	276	283
CardZipmax_AMT0	0.60410013	0.41589862	291	274.5	282.75
Merchnummax_AMT1	0.59503305	0.44124424	284	280	282
CardMerchantmax_AMT0	0.60116514	0.41589862	288	274.5	281.25
Merchnummax_AMT3	0.58343086	0.44585253	272	283.5	277.75
Cardnummax_AMT0	0.58519187	0.42511521	273	278	275.5
Merchnumtot_AMT7	0.58925593	0.34562212	277	265	271

Cardnummax_AMT7	0.55881325	0.48963134	240	301.5	270.75
Cardnummax_AMT1	0.57039465	0.43087558	257	279	268
Cardnummax_AMT3	0.56131637	0.4562212	247	285.5	266.25
Cardnummax_AMT14	0.52539138	0.49884793	224	304	264
CardStateavg_AMT7	0.58730213	0.30875576	274	253.5	263.75
Cardnumtot_AMT14	0.54757173	0.47580645	231	295.5	263.25
CardStateavg_AMT3	0.58902992	0.30529954	276	250.5	263.25
Cardnumavg_AMT1	0.5718814	0.35483871	260	266	263
Merchnumavg_AMT0	0.58305284	0.30875576	270	253.5	261.75
Cardnumavg_AMT3	0.56996382	0.36059908	255	267	261
CardMerchantavg_AMT0	0.57372434	0.31912442	264	257.5	260.75
CardZipavg_AMT30	0.59942655	0.29262673	286	235.5	260.75
CardMerchantavg_AMT3	0.59011174	0.29723502	280	241	260.5
CardZipavg_AMT3	0.58925922	0.2983871	278	243	260.5
Merchnummax_AMT7	0.55098843	0.46313364	233	288	260.5
CardStateavg_AMT0	0.5726587	0.32373272	261	259.5	260.25
CardZipavg_AMT0	0.57337273	0.31912442	263	257.5	260.25
CardZipavg_AMT7	0.59105733	0.29493088	281	239	260
CardStateavg_AMT1	0.58260417	0.30529954	269	250.5	259.75
Cardnumavg_AMT0	0.57002137	0.32834101	256	262.5	259.25
CardStateavg_AMT14	0.57301852	0.3156682	262	256	259
CardZipavg_AMT1	0.57936506	0.30184332	268	248	258
CardMerchantavg_AMT30	0.59414524	0.29147465	283	231.5	257.25
CardMerchantavg_AMT14	0.58953105	0.29262673	279	235.5	257.25
CardStateavg_AMT30	0.56655342	0.32718894	253	261	257
CardZipavg_AMT14	0.59317267	0.29147465	282	231.5	256.75
CardMerchantavg_AMT7	0.58847639	0.2937788	275	238	256.5
CardMerchantavg_AMT1	0.57694045	0.29953917	265	245.5	255.25
Cardnummed_AMT1	0.55715819	0.33179724	238	264	251
Cardnumavg_AMT7	0.54713288	0.40322581	230	270	250
Merchnumavg_AMT1	0.57744159	0.29147465	267	231.5	249.25
Cardnumavg_AMT14	0.53120935	0.40437788	225	271.5	248.25
Merchnummax_AMT14	0.5244693	0.4078341	223	273	248
CardMerchantmed_AMT0	0.56360012	0.29953917	249	245.5	247.25
CardStatemed_AMT0	0.56106482	0.30299539	245	249	247
Cardnummed_AMT0	0.55777308	0.30875576	239	253.5	246.25
CardZipmed_AMT0	0.56257342	0.2983871	248	243	245.5
Cardnumavg_AMT30	0.50885625	0.38479263	222	269	245.5
Cardnummed_AMT3	0.54472907	0.32373272	229	259.5	244.25
Cardnummax_AMT30	0.50459218	0.38018433	220	268	244
Cardnumtot_AMT30	0.48735532	0.40437788	216	271.5	243.75
CardStatemed_AMT1	0.56012206	0.29608295	242	240	241
CardZipmed_AMT1	0.5612557	0.29262673	246	235.5	240.75
CardZipmed_AMT3	0.56897184	0.28801843	254	227.5	240.75
CardMerchantmed_AMT1	0.56412858	0.29147465	250	231.5	240.75
CardMerchantmed_AMT30	0.57047212	0.27764977	258	218.5	238.25
CardMerchantmed_AMT3	0.56566341	0.28686636	251	225.5	238.25
Cardnummed_AMT14	0.48292307	0.32834101	213	262.5	237.75
CardStatemed_AMT30	0.55001136	0.2983871	232	243	237.5
CardStatemed_AMT3	0.56028313	0.28686636	243	225.5	234.25
Cardnummed_AMT7	0.48335424	0.30875576	214	253.5	233.75
Merchnumavg_AMT3	0.56591802	0.27419355	252	215	233.5

CardMerchantmed_AMT14	0.5609309	0.27880184	244	220.5	232.25
CardStatemed_AMT7	0.55172494	0.29032258	235	229	232
CardStatemed_AMT14	0.53680327	0.29262673	227	235.5	231.25
CardZipmed_AMT7	0.554999	0.28110599	236	223.5	229.75
CardMerchantmed_AMT7	0.55514546	0.27995392	237	222	229.5
Cardnummed_AMT30	0.45377302	0.30069124	211	247	229
CardZipmed_AMT30	0.56008428	0.27534562	241	216	228.5
Merchnummed_AMT0	0.54114049	0.28801843	228	227.5	227.75
CardZipmed_AMT14	0.55114275	0.27764977	234	218.5	226.25
Merchnumavg_AMT7	0.53138984	0.2764977	226	217	221.5
Merchnumavg_AMT14	0.49642112	0.28110599	219	223.5	221.25
Merchnummed_AMT1	0.50475013	0.27880184	221	220.5	220.75
Merchnummed_AMT3	0.49056889	0.26958525	218	214	216
Merchnumtot_AMT14	0.48976549	0.24884793	217	210	213.5
Merchnummax_AMT30	0.48431069	0.25	215	211	213
Merchnumavg_AMT30	0.4559838	0.26036866	212	213	212.5
Merchnummed_AMT7	0.44987038	0.25230415	210	212	211
Merchnummed_AMT14	0.42061645	0.23387097	208	208	208
Merchnumactavg_AMT30	0.38031635	0.21082949	204	207	205.5
Merchnummed_AMT30	0.41493928	0.19815668	205	205	205
Merchnumtot_AMT30	0.4430305	0.13018433	209	200	204.5
Merchnumactmed_AMT30	0.35565824	0.24423963	198	209	203.5
CardStateacttot_AMT1	0.36035862	0.12557604	203	199	201
Merchnumactavg_AMT14	0.34493542	0.1578341	196	203	199.5
Merchnumactavg_AMT7	0.34415431	0.12096774	195	198	196.5
CardMerchantacttot_AMT1	0.32264633	0.15437788	191	202	196.5
Merchnumactmed_AMT14	0.31736358	0.20852535	184	206	195
CardMerchantacttot_AMT3	0.33097511	0.11635945	193	196	194.5
CardZipacttot_AMT1	0.3174889	0.14285714	185	201	193
Merchnumactmed_AMT7	0.30796841	0.16820276	180	204	192
CardState_days_since_last_seen	0.36023325	0.08640553	202	176.5	189.25
CardZipacttot_AMT3	0.31882203	0.09792627	187	187	187
Cardnumactavg_AMT30	0.35784123	0.08410138	199	172	185.5
Cardnum_1_amount_Cardnum_30_amount_ave	0.418193	0.07718894	206.5	159.5	183
Cardnum_1_amount_Merchnum_30_amount_ave	0.418193	0.07718894	206.5	159.5	183
Merchnumactavg_AMT3	0.29358006	0.10714286	171	192.5	181.75
CardMerchant_days_since_last_seen	0.3309375	0.08294931	192	169.5	180.75
CardMerchantactmax_AMT1	0.2941626	0.10138249	173	188.5	180.75
Cardnumactmed_AMT30	0.28154132	0.11175115	165	195	180
CardStateacttot_AMT3	0.35431851	0.07718894	197	159.5	178.25
CardStateacttot_AMT0	0.29168547	0.09447005	170	185	177.5
CardZipactmax_AMT1	0.29409916	0.08870968	172	179	175.5
CardMerchantacttot_AMT0	0.2707215	0.10714286	157	192.5	174.75
CardMerchantactmax_AMT3	0.30173724	0.08410138	177	172	174.5
Merchnumactmed_AMT3	0.25008395	0.11751152	149	197	173
CardZipacttot_AMT0	0.26429459	0.10368664	155	190	172.5
CardStateactmax_AMT1	0.31887704	0.07258065	188	149	168.5
CardZipactmax_AMT3	0.29611992	0.07718894	175	159.5	167.25
CardMerchantactmax_AMT0	0.23040285	0.11059908	138	194	166
Merchnumactavg_AMT1	0.25530973	0.08410138	152	172	162
CardZipactmax_AMT0	0.22817848	0.09677419	135	186	160.5
Cardnumactavg_AMT14	0.28498795	0.07373272	168	152.5	160.25
CardStateactmax_AMT0	0.24283595	0.08294931	146	169.5	157.75
Cardnumactmed_AMT14	0.25258008	0.07834101	151	163.5	157.25

CardMerchantacttot_AMT7	0.30340073	0.06451613	179	131.5	155.25
Cardnumactavg_AMT7	0.27363817	0.07258065	161	149	155
Cardnumactmed_AMT7	0.2347853	0.07949309	143	165	154
CardMerchantactmed_AMT7	0.20623685	0.0921659	118	182.5	150.25
CardZipactmed_AMT7	0.20103326	0.09331797	111	184	147.5
Merchnumactmed_AMT1	0.21558269	0.08179724	126	168	147
CardMerchantactmax_AMT7	0.28036475	0.06336406	164	128.5	146.25
CardStateactmax_AMT3	0.3111182	0.05760369	181	109.5	145.25
CardMerchantactmed_AMT14	0.19021782	0.0921659	102	182.5	142.25
CardMerchantactmed_AMT3	0.19852621	0.08525346	107	174.5	140.75
CardStateactmed_AMT3	0.20853286	0.07718894	122	159.5	140.75
CardMerchantactmed_AMT1	0.19003029	0.08986175	101	180.5	140.75
CardZipactmed_AMT1	0.18732819	0.0875576	99	178	138.5
CardZipacttot_AMT7	0.29468204	0.05529954	174	102.5	138.25
CardMerchantactavg_AMT1	0.18646031	0.08640553	98	176.5	137.25
CardZip_days_since_last_seen	0.31878442	0.04723502	186	86.5	136.25
Cardnumactmax_AMT30	0.33720293	0.04262673	194	76.5	135.25
Cardnum_1_amount_Cardnum_7_amount_ave	0.23487266	0.06221198	144.5	124	134.25
CardStateactavg_AMT30	0.2173738	0.06912442	128	140.5	134.25
Cardnum_1_amount_Merchnum_7_amount_ave	0.23487266	0.06221198	144.5	124	134.25
CardMerchantactavg_AMT3	0.18629883	0.08525346	94	174.5	134.25
Cardnumacttot_AMT0	0.31654007	0.04608295	183	82	132.5
CardStateactavg_AMT14	0.21307852	0.06682028	124	137.5	130.75
CardMerchantactavg_AMT0	0.15920069	0.10599078	68	191	129.5
CardStateactmed_AMT7	0.22049793	0.06336406	130	128.5	129.25
CardStateacttot_AMT7	0.29951288	0.04608295	176	82	129
CardStateactavg_AMT7	0.20768132	0.06682028	120	137.5	128.75
CardStateactavg_AMT3	0.20536366	0.06912442	117	140.5	128.75
CardZipactmed_AMT3	0.1904484	0.07488479	103	154	128.5
CardMerchantactavg_AMT7	0.18390011	0.08064516	90	166.5	128.25
CardZipactmax_AMT7	0.27195757	0.05299539	160	96.5	128.25
CardMerchantactmed_AMT0	0.15333032	0.10138249	66	188.5	127.25
Cardnum_0_amount_Cardnum_30_amount_ave	0.32064481	0.03686636	189.5	63	126.25
Cardnum_0_amount_Merchnum_30_amount_ave	0.32064481	0.03686636	189.5	63	126.25
Merchnumactavg_AMT0	0.18132454	0.07834101	89	163.5	126.25
CardStateactmed_AMT1	0.19924197	0.07142857	108	144	126
Merchustar	0.31338722	0.03917051	182	68	125
CardMerchantacttot_AMT14	0.27125972	0.0483871	159	90	124.5
CardMerchantactavg_AMT14	0.18517871	0.07603687	93	155.5	124.25
CardZipactavg_AMT0	0.15765333	0.08986175	67	180.5	123.75
CardZipactavg_AMT1	0.18453817	0.07603687	91	155.5	123.25
CardStateactavg_AMT1	0.20266156	0.06451613	112	131.5	121.75
Merchnum_days_since_last_seen	0.2322184	0.05529954	141	102.5	121.75
CardZipactmed_AMT14	0.18636158	0.07142857	97	144	120.5
CardStateactmed_AMT14	0.20842788	0.05990783	121	118.5	119.75
CardMerchantactmed_AMT30	0.16770464	0.07718894	77	159.5	118.25
CardZipactavg_AMT7	0.18081393	0.07258065	86	149	117.5
Cardnumactavg_AMT3	0.21993175	0.05645161	129	106	117.5
CardMerchantactmax_AMT14	0.24886445	0.04723502	148	86.5	117.25
CardZipactavg_AMT3	0.18081353	0.07258065	85	149	117
CardMerchantactavg_AMT30	0.16967104	0.07373272	79	152.5	115.75
Cardnum_1_amount_Merchnum_14_amount_ave	0.35797336	0.02534562	200.5	30.5	115.5
Cardnum_1_amount_Cardnum_14_amount_ave	0.35797336	0.02534562	200.5	30.5	115.5
Cardnumacttot_AMT1	0.30304768	0.03341014	178	52	115

CardZipactmed_AMT0	0.15123531	0.08064516	62	166.5	114.25
CardZipactavg_AMT14	0.18048647	0.07142857	83	144	113.5
CardStateactavg_AMT0	0.16888953	0.07142857	78	144	111
CardStateactmed_AMT30	0.19839205	0.05875576	106	114.5	110.25
CardStateactmax_AMT7	0.26812322	0.03686636	156	63	109.5
Cardnum_days_since_last_seen	0.29122305	0.03225806	169	49.5	109.25
CardStateactmed_AMT0	0.15991928	0.07258065	69	149	109
Cardnumactmed_AMT3	0.20446111	0.05529954	115	102.5	108.75
CardZipacttot_AMT14	0.25020383	0.03686636	150	63	106.5
CardMerchantacttot_AMT30	0.2288021	0.04262673	136	76.5	106.25
CardUstar	0.19934192	0.05529954	109	102.5	105.75
Merchnumactmed_AMT0	0.15170363	0.07142857	63	144	103.5
CardZipactmax_AMT14	0.23050424	0.03917051	139	68	103.5
Cardnumactmed_AMT0	0.16328191	0.06336406	74	128.5	101.25
Cardnumactavg_AMT1	0.20387505	0.04723502	114	86.5	100.25
CardZipactavg_AMT30	0.16477422	0.06221198	76	124	100
CardStateacttot_AMT14	0.2288876	0.03686636	137	63	100
Cardnumactmax_AMT0	0.27959601	0.0264977	163	35	99
Merchnum_1_amount_Merchnum_14_amount_ave	0.26011336	0.02764977	153.5	38.5	96
Merchnum_1_amount_Cardnum_14_amount_ave	0.26011336	0.02764977	153.5	38.5	96
dayofweek_risk	0.13458643	0.06336406	60	128.5	94.25
Cardnumactmed_AMT1	0.19227776	0.04608295	104	82	93
Merchnum_1_amount_Merchnum_30_amount_ave	0.28175503	0.02073733	166.5	16.5	91.5
Merchnum_1_amount_Cardnum_30_amount_ave	0.28175503	0.02073733	166.5	16.5	91.5
Cardnum_count3_Date	0.07030365	0.0656682	49	134	91.5
Cardnumactmax_AMT14	0.24464561	0.02534562	147	30.5	88.75
Merchnum_0_amount_Cardnum_30_amount_ave	0.22301798	0.02995392	131.5	45.5	88.5
Merchnum_0_amount_Merchnum_30_amount_ave	0.22301798	0.02995392	131.5	45.5	88.5
Cardnumactmax_AMT1	0.27844081	0.01728111	162	14	88
Cardnumactavg_AMT0	0.17114052	0.05184332	82	93.5	87.75
CardState_count3_Date	0.07143227	0.06221198	50	124	87
CardZipactmed_AMT30	0.16133833	0.05529954	71	102.5	86.75
Cardnumacttot_AMT3	0.27115157	0.01612903	158	12.5	85.25
CardState_count1_Date	0.05496867	0.0656682	35	134	84.5
Cardnum_count1_Date	0.05062729	0.06682028	31	137.5	84.25
CardMerchant_count3_Date	0.06857545	0.06105991	46	121	83.5
CardState_count14_Date	0.06908514	0.05990783	48	118.5	83.25
CardZip_count3_Date	0.0649161	0.06221198	42	124	83
CardMerchant_count14_Date	0.06891147	0.05990783	47	118.5	82.75
Cardnumactmax_AMT3	0.21330078	0.02764977	125	38.5	81.75
CardState_count7_Date	0.07230183	0.05760369	52	109.5	80.75
CardMerchantactmax_AMT30	0.20349825	0.03110599	113	48	80.5
CardMerchant_count7_Date	0.07158248	0.05760369	51	109.5	80.25
CardMerchant_count1_Date	0.04491869	0.0656682	24	134	79
CardZip_count14_Date	0.06570824	0.05875576	43	114.5	78.75
Cardnum_0_amount_Merchnum_14_amount_ave	0.22571777	0.02304147	133.5	23.5	78.5
Cardnum_0_amount_Cardnum_14_amount_ave	0.22571777	0.02304147	133.5	23.5	78.5
CardZip_count1_Date	0.04205286	0.06682028	19	137.5	78.25
Cardnum_0_amount_Merchnum_7_amount_ave	0.13125824	0.05299539	58.5	96.5	77.5
Cardnum_0_amount_Cardnum_7_amount_ave	0.13125824	0.05299539	58.5	96.5	77.5
Merchnumacttot_AMT3	0.23225601	0.01497696	142	10.5	76.25
Merchnum_count7_Date	0.06367973	0.05760369	41	109.5	75.25
Cardnum_count7_Date	0.06133122	0.05760369	39	109.5	74.25
Merchnum_count3_Date	0.07770486	0.05184332	54	93.5	73.75

CardZip_count7_Date	0.06594569	0.05529954	44	102.5	73.25
CardStateactmax_AMT14	0.20479741	0.02534562	116	30.5	73.25
CardState_count30_Date	0.07745609	0.04953917	53	91.5	72.25
Merchnumacttot_AMT1	0.23159473	0.00921659	140	3.5	71.75
CardMerchant_count30_Date	0.06816855	0.05299539	45	96.5	70.75
CardZipacttot_AMT30	0.20722947	0.02073733	119	16.5	67.75
Merchnumactmax_AMT30	0.21684742	0.01267281	127	7	67
Cardnumacttot_AMT30	0.19959439	0.02304147	110	23.5	66.75
Merchnumactmax_AMT14	0.1846722	0.02764977	92	38.5	65.25
CardZip_count30_Date	0.06123866	0.04953917	38	91.5	64.75
Cardnum_count14_Date	0.04856654	0.05414747	29	99	64
Cardnum_count0_Date	0.04089402	0.05760369	18	109.5	63.75
CardState_count0_Date	0.03608543	0.05990783	9	118.5	63.75
CardMerchant_count0_Date	0.03871201	0.05875576	13	114.5	63.75
Merchnumacttot_AMT0	0.208651	0.00921659	123	3.5	63.25
Merchnum_1_amount_Cardnum_7_amount_ave	0.18631865	0.02534562	95.5	30.5	63
Merchnum_1_amount_Merchnum_7_amount_ave	0.18631865	0.02534562	95.5	30.5	63
CardZip_count0_Date	0.03658072	0.05875576	10	114.5	62.25
Merchnum_count1_Date	0.05717454	0.04723502	36	86.5	61.25
Merchnumacttot_AMT7	0.19503526	0.01497696	105	10.5	57.75
Merchnum_count0_Date	0.05347214	0.04493088	34	79	56.5
Cardnumacttot_AMT7	0.18959433	0.01612903	100	12.5	56.25
CardZipactmax_AMT30	0.18061675	0.02419355	84	27	55.5
Merchnum_0_amount_Merchnum_14_amount_ave	0.18091688	0.0218894	87.5	19.5	53.5
Merchnum_0_amount_Cardnum_14_amount_ave	0.18091688	0.0218894	87.5	19.5	53.5
Merchnumactmax_AMT7	0.15226144	0.02880184	64	42	53
CardStateactmax_AMT30	0.12087467	0.03225806	55	49.5	52.25
Merchnum_0_amount_Merchnum_7_amount_ave	0.17094178	0.02304147	80.5	23.5	52
Merchnum_0_amount_Cardnum_7_amount_ave	0.17094178	0.02304147	80.5	23.5	52
Merchnum_0_count_Merchnum_7_count	0.05222087	0.04032258	32.5	70.5	51.5
Merchnum_0_count_Cardnum_7_count	0.05222087	0.04032258	32.5	70.5	51.5
Merchnum_count30_Date	0.05809182	0.03801843	37	66	51.5
Merchnum_0_count_Cardnum_30_count	0.04640781	0.04147465	25.5	73.5	49.5
Merchnum_0_count_Merchnum_30_count	0.04640781	0.04147465	25.5	73.5	49.5
Merchnum_1_count_Cardnum_30_count	0.03755075	0.04723502	11.5	86.5	49
Merchnum_1_count_Merchnum_30_count	0.03755075	0.04723502	11.5	86.5	49
Merchnum_count14_Date	0.0618487	0.03571429	40	58	49
Cardnum_count30_Date	0.04859144	0.03917051	30	68	49
Cardnum_0_count_Cardnum_7_count	0.04383704	0.04147465	20.5	73.5	47
Cardnum_0_count_Merchnum_7_count	0.04383704	0.04147465	20.5	73.5	47
CardStateacttot_AMT30	0.15280718	0.02304147	65	23.5	44.25
Merchnum_0_count_Merchnum_14_count	0.04766775	0.03571429	27.5	58	42.75
Merchnum_0_count_Cardnum_14_count	0.04766775	0.03571429	27.5	58	42.75
Cardnum_1_count_Cardnum_7_count	0.02951322	0.04493088	5.5	79	42.25
Cardnum_1_count_Merchnum_7_count	0.02951322	0.04493088	5.5	79	42.25
Merchnumactmax_AMT1	0.16163726	0.01267281	72	7	39.5
Merchnum_1_count_Merchnum_7_count	0.04432667	0.03456221	22.5	54.5	38.5
Merchnum_1_count_Cardnum_7_count	0.04432667	0.03456221	22.5	54.5	38.5
Cardnumactmax_AMT7	0.16058894	0.01267281	70	7	38.5
Merchnumacttot_AMT30	0.16339896	0.00460829	75	2	38.5
Merchnumacttot_AMT14	0.16314025	0.00345622	73	1	37
Merchnumactmax_AMT0	0.12967413	0.02073733	57	16.5	36.75
Merchnum_1_count_Cardnum_14_count	0.03969718	0.03571429	14.5	58	36.25
Merchnum_1_count_Merchnum_14_count	0.03969718	0.03571429	14.5	58	36.25
Merchnumactmax_AMT3	0.14057611	0.01382488	61	9	35

Cardnum_1_count_Merchnum_30_count	0.04022078	0.03341014	16.5	52	34.25
Cardnum_1_count_Cardnum_30_count	0.04022078	0.03341014	16.5	52	34.25
Cardnumacttot_AMT14	0.1287327	0.01152074	56	5	30.5
Cardnum_1_count_Merchnum_14_count	0.03551225	0.02880184	7.5	42	24.75
Cardnum_1_count_Cardnum_14_count	0.03551225	0.02880184	7.5	42	24.75
Cardnum_0_count_Merchnum_30_count	0.02600644	0.02995392	3.5	45.5	24.5
Cardnum_0_count_Cardnum_30_count	0.02600644	0.02995392	3.5	45.5	24.5
Cardnum_0_count_Merchnum_14_count	0.02406881	0.0264977	1.5	35	18.25
Cardnum_0_count_Cardnum_14_count	0.02406881	0.0264977	1.5	35	18.25