DSO 562 Project 2

Finding Anomalies in Application Data

Team 5

Badawy, Yasmine

Chuang, Hungli

Gupta, Varun

Hu, Yiting

Lin, Qiongqiong

Rawat, Akash

Srivastava, Astha

Yu, Shui

Table of Contents

Executive Summary

The purpose of this project is to identify anomalies among product application records by building a supervised fraud model and find out potential fraudulent applications by applying supervised machine learning algorithms.

The "Product Application" file is a credit card application dataset which includes the information of applicants. The document is mainly used to record the identity information from applicants. It contains one million records and nine fields on credit card applications. The records were assessed in 2016 and within the United States.

The report begins with a detailed description of the dataset. The full data quality report is included as an appendix of this report. The dataset includes nine fields and one million records. Five of nine fields are numeric fields and four are categorical.

Followed by the description of data and distribution, the report explains how data was being cleaned and replaced outliers with normal data. We also created additional 282 variables and utilized them in our model. After creating data, we calculated univariate KS and univariate FDR at 3% and sorted the variables by both of these measures and provided the two rank ordered lists. We removed about half the variables and then used a wrapper method to reduce to about 20 variables by stepwise logistic regression. Then selecting our best models and finalizing by applying a regularization method. After reducing dimensions, we used records before 11/1/2016 as training and testing data and fit our model to make predictions on the records after 11/1/2016.

We then used supervised algorithms including a logistic regression, a random forest, neural networks and Gradient Boosting methods to detect fraud in the application dataset provided. Lastly, we will create a threshold for the top 7 percent of applications to be rejected based on our fraud scoring model to optimize the balance between rejecting legitimate applications and accepting fraudulent ones.

Description of Data

Overall Description

The dataset contained the information of product application across 2016. It also contained a label for fraud identification, which enabled us to train supervised learning algorithms to identify fraud records. There were altogether 10 categorical fields and 1,000,000 records in the dataset. There were nine categorical variables with the 'record' variable uniquely defining each row. Following is a summary table of the categorical variables -

Variable	Number of records with value	% populated	# unique values	missing values	Most common value (MCV)	Frequency of MCV
date	1,000,000	100%	365	0	20160816	2,877
ssn	1,000,000	100%	835,819	0	999999999	16,935
firstname	1,000,000	100%	78,136	0	EAMSTRMT	12,658
lastname	1,000,000	100%	177,001	0	ERJSAXA	8,580
address	1,000,000	100%	828,774	0	123 MAIN ST	1,079
zip5	1,000,000	100%	26,370	0	68138	823
dob	1,000,000	100%	42,673	0	19070626	126,568
homephone	1,000,000	100%	28,244	0	9999999999	78,512
fraud_label	1,000,000	100%	2	0	0	985,607

Table 1.1 Summary of Categorical Variables

Description of Variables

date (Categorical, datetime)

This was the date of each application made in 2016. There were 365 unique values for this field with no missing/null values. Following is a distribution of the top 10 categories –

date	count
2016-08-16	2877
2016-03-04	2861
2016-07-18	2849
2016-04-17	2848
2016-01-01	2840
2016-09-03	2832
2016-08-08	2832
2016-12-28	2832
2016-08-27	2831
2016-10-06	2831

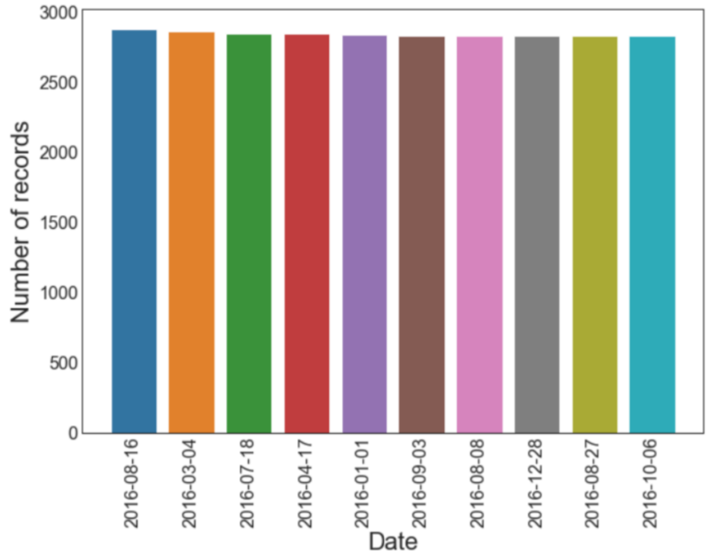

Figure 1.1 Distribution of ‘date’ variable

ssn (Categorical, 9-digit code)

This categorical variable defined the social security number of the applicant for each record/row. There were 835,819 unique values for this field with no missing/null values. Following is a distribution of the top 10 categories –

ssn	count
999999999	16935
938972725	114
810776805	81
829352390	74
473311863	44
165764866	34
295308935	30
189622157	30
163830210	29
588726212	29

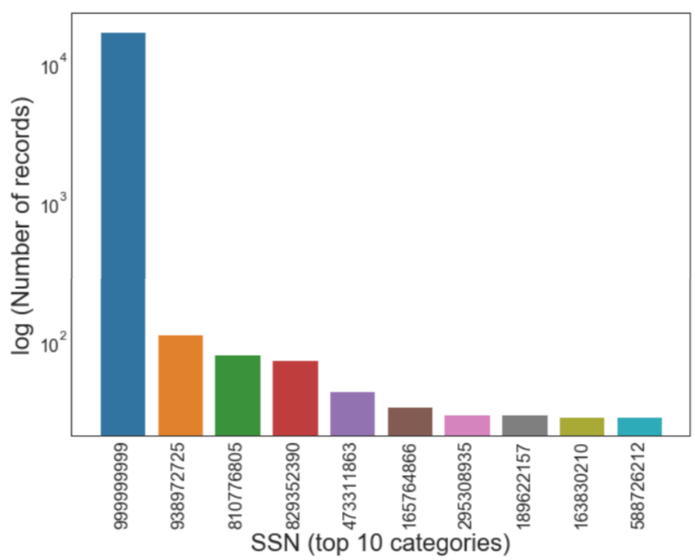

Fig 1.2 Categorical distribution of ‘ssn’ variable

We observed that ~17,000 values have SSN as ‘999999999’. This value could have been used to fill in missing values or where the SSN of the applicant was not available.

firstname (Categorical, string)

This categorical variable defined the first name of the applicant for each record/row. There were 78,136 unique values for this field with no missing/null values. Following is a distribution of the top 10 categories –

first name	count
EAMSTRMT	12658
TXEMXZZM	10297
UXXJJZTUZ	10235
UJSRSMUEZ	9994
SREZUJMJU	9688
EASEXMJAT	7576
SSSXUEJMS	6923
SZUASTTA	6878
EREMTZXXA	6717
EAXRRUMUX	5686

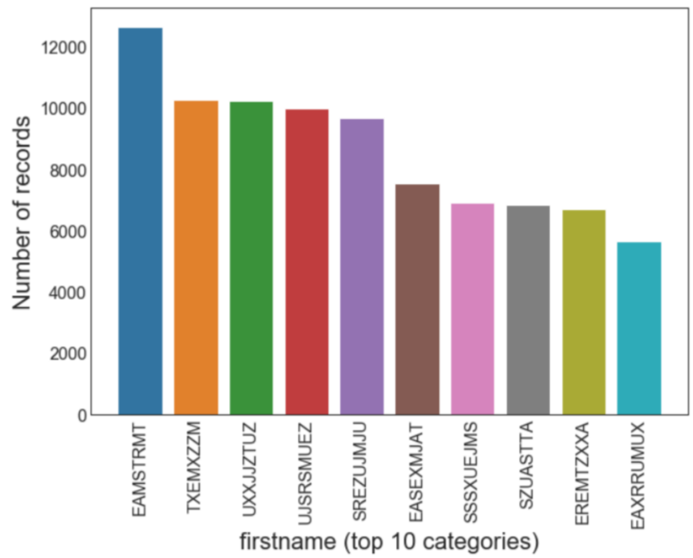

Fig 1.3 Categorical distribution of 'firstname' variable

address (Categorical, string)

This categorical variable defined the address of the applicant for each record/row. There were 828,774 unique values for this field with no missing/null values. Following is a distribution of the top 10 categories –

address	count
123 MAIN ST	1079
1775 XJXE LN	97
7433 RAEZA ST	80
8911 MZSU DR	74
4907 RRAAU DR	73
426 XUAXZ BLVD	57
3545 ARMA ST	44
606 EZZAU WY	44
4530 ETSMX WY	42
4292 RUSMM LN	41

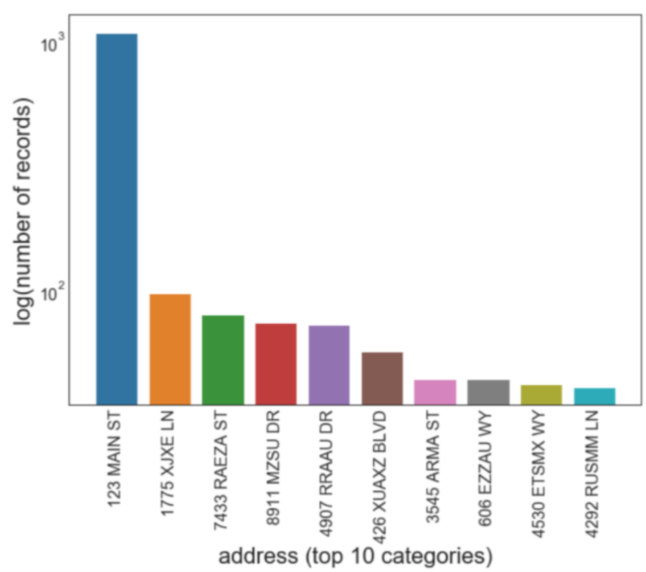

Fig 1.4 Categorical distribution of 'address' variable

zip5 (Categorical, 5-digit code)

This categorical variable defined the 5-digit zip code of the applicant for each record/row. There were 26,370 unique values for this field with no missing/null values. Following is a distribution of the top 10 categories –

zip5	count
68138	823
90042	514
89835	489
35227	478
14931	459
86500	438
12700	436
1362	434
59695	432
52317	432

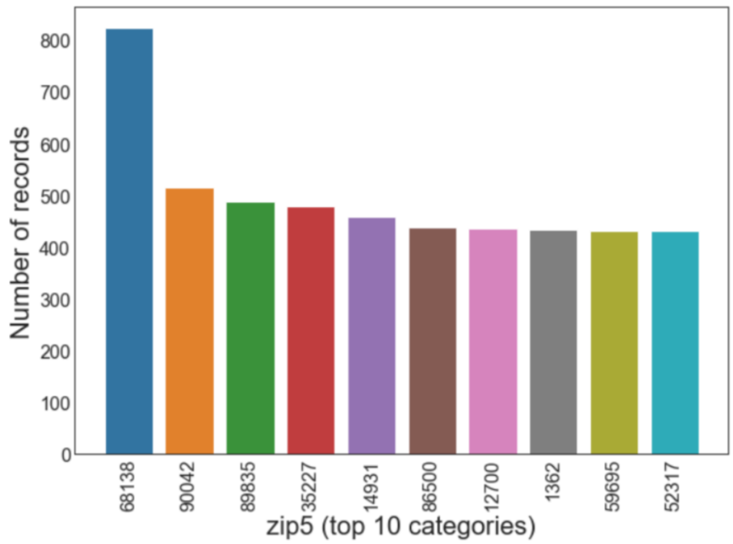

Fig 1.5 Categorical distribution of 'zip5' variable

dob (Categorical, datetime)

This categorical variable defined the date of birth of the applicant for each record/row. There were 42,673 unique values for this field with no missing/null values. Following is a distribution of the top 10 categories –

dob	count
1907-06-26	126568
1964-03-18	4818
1976-06-25	3723
1988-06-28	1404
1974-02-16	980
2009-01-27	280
1946-09-01	135
1959-12-08	126
1928-06-11	120
1967-02-15	102

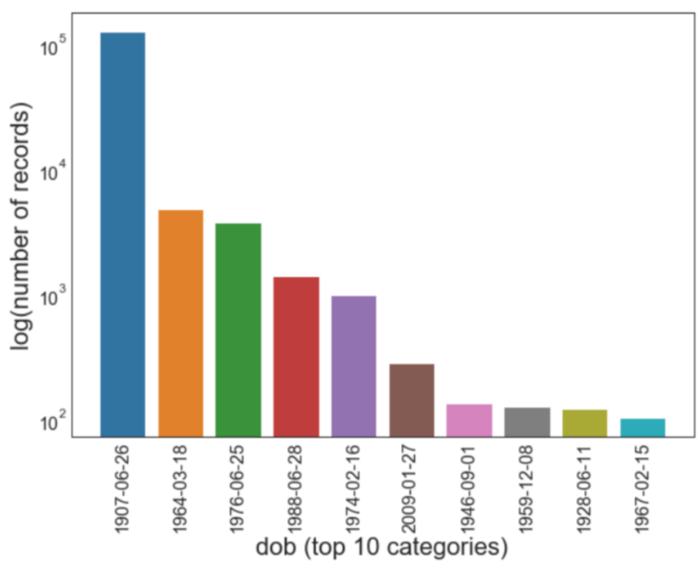

Fig 1.6 Categorical distribution of 'dob' variable

homephone (Categorical, 10-digit code)

This categorical variable defined the homephone of the applicant for each record/row. There were 28,244 unique values for this field with no missing/null values. Following is a distribution of the top 10 categories –

homephone	count
9999999999	78512
6384782007	466
6035129044	360
2113738531	331
4024680535	198
2669445638	172
6637507363	169
8629049955	139
3364980740	110
6845279110	108

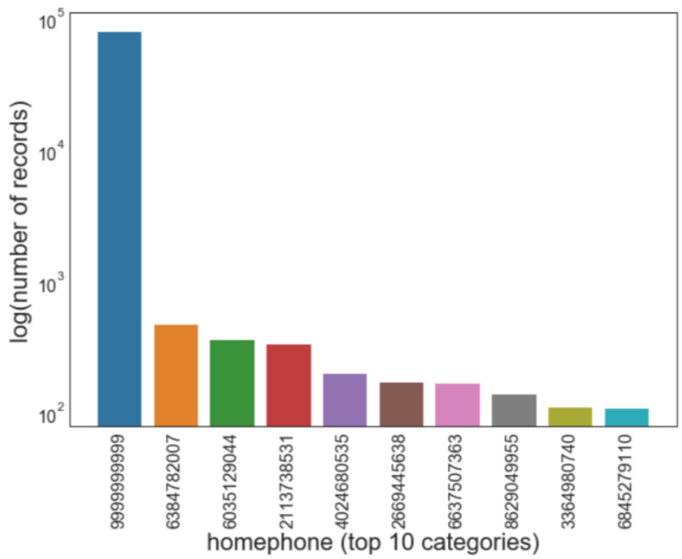

Fig 1.7 Categorical distribution of 'homephone' variable

Fraud_label (Categorical, 0 or 1)

This categorical variable indicated if the record/applicant is fraud or not. There were 2 unique values for this field with no missing/null values. Following is the distribution of the two categories –

fraud_label	count
0	985607
1	14393

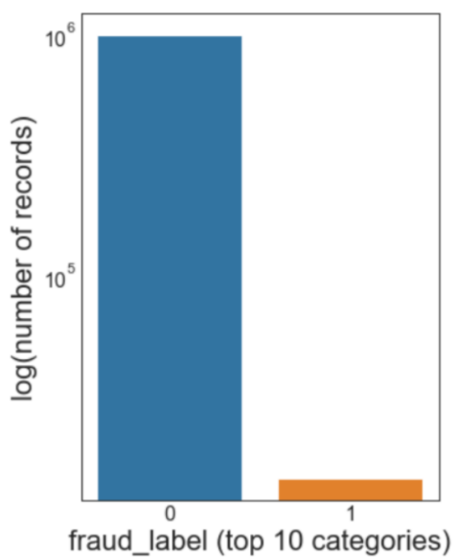

Fig 1.8 Categorical distribution of 'fraud_label' variable

Data Cleaning

Fix Frivolous Values

There were frivolous values in the dataset, which could have been used to fill in missing values or where the information of the applicant was not available. Following is the summary of the frivolous values –

Variable	Frivolous Value
ssn	999999999
address	123 MAIN ST
dob	19070626
homephone	9999999999

Table 2.1 Summary of Frivolous Values

To fix the frivolous values in 'ssn', 'dob' and 'homephone', we replaced these values using zeros followed by record number. To fix the frivolous values in 'address', we replaced these values using record number.

To illustrate the method to fix frivolous values more clearly, the following table shows that if record *p* has frivolous values in the above fields, the values will be substituted to be –

Variable	Fixed Value
ssn	0000000-p
address	p RECORD
dob	000000-p
homephone	00000000-p

Table 2.2 Example of Fixed Values

Candidate Variables

Combine Related Variables

After fixing the frivolous variables, we combined the related variables to be used as our expert variables/attributes. For example, as firstname and lastname were related for the same applicant, we combined these two variables to create a new variable called 'name'.

Fields like name-DOB (combination of 'firstname', 'lastname', and 'dob' fields) can be a really good unique identifier of a person rather than only using these entities individually. Also, there can be several similar addresses, but they can be located at completely different locations, so it's important to attach 'Zip code' with an address value to make it a unique address identifier ('addr').

After adding these combined variables, we had 26 variables altogether. Following is the information of related variables we created –

Variables	Combination
ssn	-
address	-
dob	-
homephone	-
name	lastname, firstname
addr	address, zip5
name-dob	lastname, firstname, dob
name-addr	lastname, firstname, address, zip5
name-homephone	lastname, firstname, homephone
dob-addr	dob, address, zip5
dob-homephone	dob, homephone
addr-homephone	address, homephone
name-dob-addr	lastname, firstname, dob, address, zip5
name-dob-homephone	lastname, firstname, dob, homephone
name-addr-homephone	lastname, firstname, address, zip5, homephone
dob-addr-homephone	dob, address, zip5, homephone
name-dob-addr-homephone	lastname, firstname, dob, address, zip5, homephone
ssn-firstname	ssn, firstname
ssn-lastname	ssn, lastname
ssn-address	ssn, address
ssn-zip5	ssn, zip5
ssn-dob	ssn, dob
ssn-homephone	ssn, homephone
ssn-name	ssn, firstname, lastname
ssn-addr	ssn, address, zip5
ssn-name-dob	ssn, firstname, lastname, dob

Table 2.3 Variables Including the Combined Variables

Create Variables Across Time

After combining all the related variables, for each entites and combination group, we created the days since, velocity and relative velocity variables to make the model more robust and invariant to seasonality.

a) Days since variables - To create days-since variable, we calculated the number of days since we last saw a specific combination group or entity.
For each variable, we created 1 'Days since' variable, so overall, we created 26 'Days since' variables. For example, 'diff_date.ssn_fulladdress' indicates how many days since an application has been filed with a unique combination of SSN and full address
b) Velocity variables: As for velocity variables, we first created a timeframe called lags = [0, 1, 3, 7, 14, 30], and calculated the number of records with the same combination group we saw in past lags day, which represented the frequency of seeing same entity or combination group over past lags day.
For each of the variables listed above, we created 6 variables (one for each timestamp). So overall, we created 156 velocity variables. For example, 'fulladdresshomephone14' means number of applications filed with a combination of full address and homephone in the last 14 days.
c) Relative Velocity variables Lastly, we created the relative velocity variables using the formula of number of applications with a specific group we saw in the recent past divided by number of applications with the same group we saw in past lags days.

After creating the velocity variables and relative velocity variables, we had a total of 288 expert variables. The list of the variables can be found in the appendix.

Feature Selection Process

Univariate Filter using KS and FDR

Before doing KS and FDR, we standardized our candidate variables using Z-scaling. For each of our candidate variables, we calculated Kolmogorov–Smirnov (KS) score and fraud detection rate individually. Both the KS score and the FDR rate will help us determine how well candidate variables individually predict fraud, allowing us to rank order the variables in terms of usefulness for our models.

The KS score is a filter method that helps determine how well a candidate variable separates the goods from the bads, or in this case, the frauds and the not frauds. For each variable, we will use the formula below to calculate a KS score and rank order the variables by the score.

$$KS = \max_{x} \int_{x_{min}}^{x} \left[P_{\text{goods}} - P_{\text{bads}}\right] dx$$

$$KS = \max_{x} \sum_{x_{min}}^{x} \left[P_{\text{goods}} - P_{\text{bads}}\right]$$

The FDR for each variable be determined at a 3% level. It’s the value representingthe % of all frauds caught at a particular examination cutoff. For each variable, we will determine what percent of frauds are captured by the top 3% of the variable and rank order as such.

First, we divided the whole dataset into training, test and out of time sets. We set the records between ‘2016-01-14’ and ‘2016-11-01’ to be the training and test set, and set the records on and after ‘2016-11-01’ to be the out of time set.

Then, we calculated the Kolmogorov-Smirnov(KS) and Fraud Detection Rate(FDR) of each variable, and ranked them by KS and FDR respectively. After that, to select top ranked variables, we calculated the average rank of each variable and selected the top 100 variables with the highest average rank. Following table shows the top 10 variables –

Field	KS	FDR	KS Rank	FDR Rank	Average Rank
fraud_label	1.0	1.0	292.0	292.0	292.0
addr_lag30_count	0.332032	0.354953	290.0	291.0	290.5
address_lag30_count	0.332724	0.353299	291.0	290.0	290.5
addr_#days_since	0.323542	0.349381	288.0	289.0	288.5
address_#days_since	0.324626	0.348075	289.0	288.0	288.5
address_lag14_count	0.322252	0.345812	287.0	287.0	287.0
addr_lag14_count	0.321755	0.342329	286.0	286.0	286.0
address_lag7_count	0.301444	0.320999	285.0	285.0	285.0
addr_lag7_count	0.301367	0.319954	284.0	284.0	284.0
address_lag3_count	0.278445	0.299059	282.0	283.0	282.5

Table 3.1 KS and FDR of All Expert Variables (Top 10)

The full table of the KS and FDR rank can be found in the appendix. Next, we used a wrapper method to continue our feature selection process.

Recursive Feature Elimination and Cross-validated selection

The wrapper method we chose was the recursive feature elimination and cross-validated selection. Recursive Feature Elimination(RFE) is a feature selection method that fits a model and removes the weakest feature until the specified number of features is reached. Features are ranked by the model's coefficients or feature importances attribute, followed by recursive elimination of a small number of features per loop. Cross validation is combined to select the best parameters for the RFE.

This method was be implemented using the RFECV function in the Scikit-learn package in Python. For the parameters, we used logistic regression as the estimater, with the settings "step" set to 1 and we set the 'Cross Validation' count as 3 which essentially splits the data into 3 parts and choose 1 part as test and other two as the training data.

Based on this, we finally got a list of 20 variables on which we built our below models. The 20 variables are as follows - ['addr_lag30_count', 'address_lag14_count', 'addr_lag14_count', 'address_lag7_count', 'addr-homephone_lag30_count', 'name-dob_lag30_count', 'ssn-name_lag30_count', 'ssn-lastname_lag14_count', 'ssn-name_lag14_count', 'ssn-name-dob_lag7_count', 'ssn-name_lag7_count', 'address_lag0_count', 'addr_lag0_count', 'addr-homephone_lag3_count', 'ssn_lag3_count', 'ssn-firstname_lag3_count', 'ssn-dob_lag3_count', 'ssn-name_lag3_count', 'name_lag3_count', 'homephone_lag0_count']

Model Algorithms

Logistic Regression

A multiple logistic regression employs multiple variables to predict the likelihood of the target variable. Using least squares method the model optimizes the coeffecients for each of the predictor variables.

We made use of the logistic regression model using different combinations of our identified 20 wrapper variables. Although we used the wrapper to identify the top 20 variables, we also needed to use a different tool to identify smaller combinations of variables that would perform best.

We used recursive feature elimination to find the most effective, smaller, combinations of variables to try models of sizes 15-20. The RFE recursively removes attributes and builds a model on the attributes that remain and computes which combinations of attributes contribute the most to predicting the target. After running the RFE, we identified the smaller combinations have used them to predict fraud.

Our model's top performance occurred with a combination of size 20. The model's fraud detection rate at 3% threshold was 50.78% for training, 50.14% for testing and 48.36% for the holdout sample. This model would serve as our baseline for to improve upon with more advanced algorithms.

Random Forest:

In random forests, when building these decision trees, each time a split in a tree is considered, a random sample of predictors is chosen as split candidates from the full set of predictors. The number of predictors considered at each split is approximately equal to the square root of the total number of predictors.

In other words, in building a random forest, at each split in the tree, the algorithm is not allowed to consider most of the available predictors. Random forests considers a subset of predictors and this helps to reduce the effect of highly correlated predictors. On a long run, this will help to reduce variance when we take average of predicted values.

We used the RandomForestClassifier package from the library sklearn to make the Random Forest model on our reduced set of variables. We varied the number of estimators i.e. no. of trees and then we trained our model on training data. Then we predicted the probability of Fraud over training, test and OOT (validation data).

Our model's top performance occurred with the number of estimators as 300. The model's fraud detection rate at 3% threshold was 54.88% for training, 54.10% for testing and 52.77% for the holdout sample. Our Random Forest model was our top performing model, boasting an OOT accuracy of 52.77%.

Gradient Boosted Trees:

Boosted trees is another approach for improving the predictions resulting from a decision tree. Boosting can be applied to many statistical models for regression and classification. In boosting, trees are grown sequentially, with each tree grown using information from previously grown trees. Each tree is fit on a modified version of the original data set, with each boost learning slowly. This approach is different than fitting a single large decision tree to the data, which results in fitting the data hard and potentially overfitting.

Given the current model, we fit the decision tree to the residuals from the model. That is, we fit a tree using the current residuals, rather than the outcome Y, as the response. We then added this new decision tree into the fitted function in order to update the residuals. By fitting small trees to the residuals, we slowly improved. In general, statistical learning approaches that learn slowly tend to perform well. In boosting, the construction of each tree depends strongly on the trees that have already been grown. In summary, the boosted trees approach combines many simple models in a linear fashion, creating a series of weak learners. The linear combinations of all the simple models create a strong learner.

Our model's top performance occurred with the number of estimators as 200 and max depth as 2. The model's fraud detection rate at 3% threshold was 54.60% for training, 53.89% for testing and 52.26% for the holdout sample.

Neural Net:

Neural Net is a type of machine learning designed to recognize patterns. The neural net was inspired by the biological neural networks that constitutes animal brains. The typical neural net consists of an input layer, some number of hidden layers and an output layer. A neural net with more than one hidden layer is a deep learning neural net. Deep learning is a neural net architecture. With deep learning, the computer trains itself to process and learn from data instead of teaching computers to process and learn from data (which is how machine learning works).

Each node in the hidden layer receives weighted signals from all the nodes in the incoming layer and does a transformation on this linear combination of signals. The transform/activation function can be one of a number of functions, for example a logistic function (sigmoid). To obtain a more robust understanding of the model's performance, we trained the network six times, tuning a combination of various parameters into it for each run.

Our model's top performance occurred with two hidden layers of sizes (32, 64) and 50 iterations. The model's fraud detection rate at 3% threshold was 50.72% for training, 54.03% for testing and 52.05% for the holdout sample.

FDR (Train, Test and OOT) of different models:

Model	Parameter				Average FDR		
Logistic Regression	Total Variables	Number of variables selected			Train	Test	OOT
1	20	15			48.61%	47.66%	46.44%
2	20	16			48.57%	47.74%	46.40%
3	20	17			48.85%	48.05%	46.73%
4	20	18			48.85%	48.08%	46.77%
5	20	19			48.88%	48.05%	46.70%
6	20	20			50.78%	50.14%	48.36%
Random Forest	Number of Variables	Number of trees					
1	20	200			55.08%	53.96%	52.64%
2	20	300			54.95%	54.10%	52.72%

3	20	400			54.88%	54.10%	52.77%
4	20	500			55.14%	54.07%	52.14%
Gradient Boosted Tree	Number of Variables	Number of trees	Max Depth	Learning rate			
1	20	100	2	0.1	54.56%	53.72%	51.68%
2	20	200	2	0.1	54.60%	53.89%	52.26%
3	20	400	2	0.1	54.69%	54.03%	51.93%
4	20	400	5	0.1	54.99%	54.10%	51.93%
4	20	500	2	0.1	54.62%	54.10%	51.93%
Neural Network	Total Variables	No. of hidden layers	No. of neurons per layer	No. of epochs			
1	20	2	(64;128)	50	49.83%	53.75%	51.92%
2	20	1	48	50	44.88%	53.18%	51.29%
3	20	2	(48;96)	50	51.55%	53.75%	51.5%
4	20	1	32	40	41.08%	52.72%	50.16%
5	20	2	(32;64)	50	50.72%	54.03%	52.05%
6	20	1	24	40	47.48%	53.57%	51.46%

Results

Our best performing algorithm is Random Forest model and we have generated cumulative Good, Bads, % Good, % Bad (FDR), KS and FPR for all three populations (training, testing, and Validation (OOT), and the fraud savings plot. We have listed the top 20 batches for each set of data. The complete list can be found in the appendix.

1) Training Data

Training	# Records	# Goods	# Bads	Fraud Rate
	596,247	587,587	8,660	1.45%

	Bin Statistics					Cumulative Statistics						
Population Bin %	Records	Goods	Bads	% Goods	% Bads	Total # Records	Cumulative Goods	Cumulative Bads	% Goods	% Bads	KS	FPR
0	5963	1395	4568	23.39%	76.61%	5963	1395	4568	0.24%	52.75%	52.51%	0.31
1	5963	5828	135	97.74%	2.26%	11926	7223	4703	1.23%	54.31%	53.08%	1.54
2	5963	5913	50	99.16%	0.84%	17889	13136	4753	2.24%	54.88%	52.65%	2.76
3	5963	5883	80	98.66%	1.34%	23852	19019	4833	3.24%	55.81%	52.57%	3.94
4	5963	5899	64	98.93%	1.07%	29815	24918	4897	4.24%	56.55%	52.31%	5.09
5	5963	5896	67	98.88%	1.12%	35778	30814	4964	5.24%	57.32%	52.08%	6.21
6	5963	5895	68	98.86%	1.14%	41741	36709	5032	6.25%	58.11%	51.86%	7.30
7	5963	5917	46	99.23%	0.77%	47704	42626	5078	7.25%	58.64%	51.38%	8.39
8	5963	5929	34	99.43%	0.57%	53667	48555	5112	8.26%	59.03%	50.77%	9.50
9	5963	5925	38	99.36%	0.64%	59630	54480	5150	9.27%	59.47%	50.20%	10.58
10	5963	5912	51	99.14%	0.86%	65593	60392	5201	10.28%	60.06%	49.78%	11.61
11	5963	5930	33	99.45%	0.55%	71556	66322	5234	11.29%	60.44%	49.15%	12.67
12	5963	5915	48	99.20%	0.80%	77519	72237	5282	12.29%	60.99%	48.70%	13.68
13	5963	5918	45	99.25%	0.75%	83482	78155	5327	13.30%	61.51%	48.21%	14.67
14	5963	5938	25	99.58%	0.42%	89445	84093	5352	14.31%	61.80%	47.49%	15.71
15	5963	5929	34	99.43%	0.57%	95408	90022	5386	15.32%	62.19%	46.87%	16.71
16	5963	5924	39	99.35%	0.65%	101371	95946	5425	16.33%	62.64%	46.32%	17.69
17	5963	5925	38	99.36%	0.64%	107334	101871	5463	17.34%	63.08%	45.75%	18.65
18	5963	5930	33	99.45%	0.55%	113297	107801	5496	18.35%	63.46%	45.12%	19.61
19	5963	5932	31	99.48%	0.52%	119260	113733	5527	19.36%	63.82%	44.47%	20.58
20	5963	5926	37	99.38%	0.62%	125223	119659	5564	20.36%	64.25%	43.88%	21.51

2) Test Data

Test	# Records	# Goods	# Bads	Fraud Rate
	198,749	195,923	2,826	1.42%

	Bin Statistics					Cumulative Statistics						
Population Bin %	Records	Goods	Bads	% Goods	% Bads	Total # Records	Cumulative Goods	Cumulative Bads	% Goods	% Bads	KS	FPR
0	1988	520	1468	26.16%	73.84%	1988	520	1468	0.27%	51.95%	51.68%	0.35
1	1988	1948	40	97.99%	2.01%	3976	2468	1508	1.26%	53.36%	52.10%	1.64
2	1988	1967	21	98.94%	1.06%	5964	4435	1529	2.26%	54.10%	51.84%	2.90

3	1988	1971	17	99.14%	0.86%	7952	6406	1546	3.27%	54.71%	51.44%	4.14
4	1988	1971	17	99.14%	0.86%	9940	8377	1563	4.28%	55.31%	51.03%	5.36
5	1988	1968	20	98.99%	1.01%	11928	10345	1583	5.28%	56.02%	50.74%	6.54
6	1988	1969	19	99.04%	0.96%	13916	12314	1602	6.29%	56.69%	50.40%	7.69
7	1988	1979	9	99.55%	0.45%	15904	14293	1611	7.30%	57.01%	49.71%	8.87
8	1988	1972	16	99.20%	0.80%	17892	16265	1627	8.30%	57.57%	49.27%	10.00
9	1988	1974	14	99.30%	0.70%	19880	18239	1641	9.31%	58.07%	48.76%	11.11
10	1988	1976	12	99.40%	0.60%	21868	20215	1653	10.32%	58.49%	48.17%	12.23
11	1988	1976	12	99.40%	0.60%	23856	22191	1665	11.33%	58.92%	47.59%	13.33
12	1988	1969	19	99.04%	0.96%	25844	24160	1684	12.33%	59.59%	47.26%	14.35
13	1988	1975	13	99.35%	0.65%	27832	26135	1697	13.34%	60.05%	46.71%	15.40
14	1988	1970	18	99.09%	0.91%	29820	28105	1715	14.34%	60.69%	46.34%	16.39
15	1988	1974	14	99.30%	0.70%	31808	30079	1729	15.35%	61.18%	45.83%	17.40
16	1988	1983	5	99.75%	0.25%	33796	32062	1734	16.36%	61.36%	44.99%	18.49
17	1988	1979	9	99.55%	0.45%	35784	34041	1743	17.37%	61.68%	44.30%	19.53
18	1988	1973	15	99.25%	0.75%	37772	36014	1758	18.38%	62.21%	43.83%	20.49
19	1988	1983	5	99.75%	0.25%	39760	37997	1763	19.39%	62.38%	42.99%	21.55
20	1988	1973	15	99.25%	0.75%	41748	39970	1778	20.40%	62.92%	42.51%	22.48

3) Validation Data

Validation	# Records	# Goods	# Bads	Fraud Rate
	166,493	164,107	2,386	1.43%

	Bin Statistics					Cumulative Statistics						
Population Bin %	Records	Goods	Bads	% Goods	% Bads	Total # Records	Cumulative Goods	Cumulative Bads	% Goods	% Bads	KS	FPR
0	1,665	547	1,118	32.85%	67.15%	1,665	547	1,118	0.33%	46.86%	46.52%	0.49
1	1,665	1,654	11	99.34%	0.66%	3,330	2,201	1,129	1.34%	47.32%	45.98%	1.95
2	1,665	1,650	15	99.10%	0.90%	4,995	3,851	1,144	2.35%	47.95%	45.60%	3.37
3	1,665	1,647	18	98.92%	1.08%	6,660	5,498	1,162	3.35%	48.70%	45.35%	4.73
4	1,665	1,652	13	99.22%	0.78%	8,325	7,150	1,175	4.36%	49.25%	44.89%	6.09
5	1,665	1,656	9	99.46%	0.54%	9,990	8,806	1,184	5.37%	49.62%	44.26%	7.44
6	1,665	1,651	14	99.16%	0.84%	11,655	10,457	1,198	6.37%	50.21%	43.84%	8.73
7	1,665	1,651	14	99.16%	0.84%	13,320	12,108	1,212	7.38%	50.80%	43.42%	9.99
8	1,665	1,659	6	99.64%	0.36%	14,985	13,767	1,218	8.39%	51.05%	42.66%	11.30
9	1,665	1,650	15	99.10%	0.90%	16,650	15,417	1,233	9.39%	51.68%	42.28%	12.50
10	1,665	1,650	15	99.10%	0.90%	18,315	17,067	1,248	10.40%	52.31%	41.91%	13.68
11	1,665	1,655	10	99.40%	0.60%	19,980	18,722	1,258	11.41%	52.72%	41.32%	14.88

12	1,665	1,653	12	99.28%	0.72%	21,645	20,375	1,270	12.42%	53.23%	40.81%	16.04
13	1,665	1,653	12	99.28%	0.72%	23,310	22,028	1,282	13.42%	53.73%	40.31%	17.18
14	1,665	1,659	6	99.64%	0.36%	24,975	23,687	1,288	14.43%	53.98%	39.55%	18.39
15	1,665	1,655	10	99.40%	0.60%	26,640	25,342	1,298	15.44%	54.40%	38.96%	19.52
16	1,665	1,656	9	99.46%	0.54%	28,305	26,998	1,307	16.45%	54.78%	38.33%	20.66
17	1,665	1,645	20	98.80%	1.20%	29,970	28,643	1,327	17.45%	55.62%	38.16%	21.58
18	1,665	1,649	16	99.04%	0.96%	31,635	30,292	1,343	18.46%	56.29%	37.83%	22.56
19	1,665	1,653	12	99.28%	0.72%	33,300	31,945	1,355	19.47%	56.79%	37.32%	23.58
20	1,665	1,651	14	99.16%	0.84%	34,965	33,596	1,369	20.47%	57.38%	36.90%	24.54

Conclusions

Conclusions

Application fraud is one of the most common identity frauds. Falsified or stolen personal information is used to apply for cards, accounts, etc. In this report, we have examined the dataset to draw the following conclusion.

Comparing all the above models, we can conclude that RandomForest performed the best. The FDR on training dataset is 54.88%, 54.10% on test set and 52.77% on the validation dataset. We used supervised algorithms including logistic regression, RandomForest, GradientBoostedTrees and NeuralNets.

Potential Improvements

We trained our models by training, testing and validating with the original dataset, which had only 1.4% of potential fraudulent records. In our perspective, weighting a dataset can improve the model accuracy. Also, as fraud datafiles are imbalanced, we can choose to scramble the goods or unscramble the bads to increase the model accuracy.

Gains in FDR can be achieved with the addition of external datasets related to our potential applicants. For example, more legitimate data from a cell phone company containing accurate name and phone number combinations could make it much easier to identify algorithmically whether or not someone is using falsified information in their application. Similarly, a collection of addresses and the last name of the owner could potentially lead to greater accuracy if utilized correctly. Adding additional variables or information related to the interactions between variables in the dataset could potentially help increase FDR in the future.

Appendix

Description of Variables

date (Categorical, datetime)

This was the date of each application made in 2016. There were 365 unique values for this field with no missing/null values. Following is a distribution of the top 10 categories –

date	count
2016-08-16	2877
2016-03-04	2861
2016-07-18	2849
2016-04-17	2848
2016-01-01	2840
2016-09-03	2832
2016-08-08	2832
2016-12-28	2832
2016-08-27	2831
2016-10-06	2831

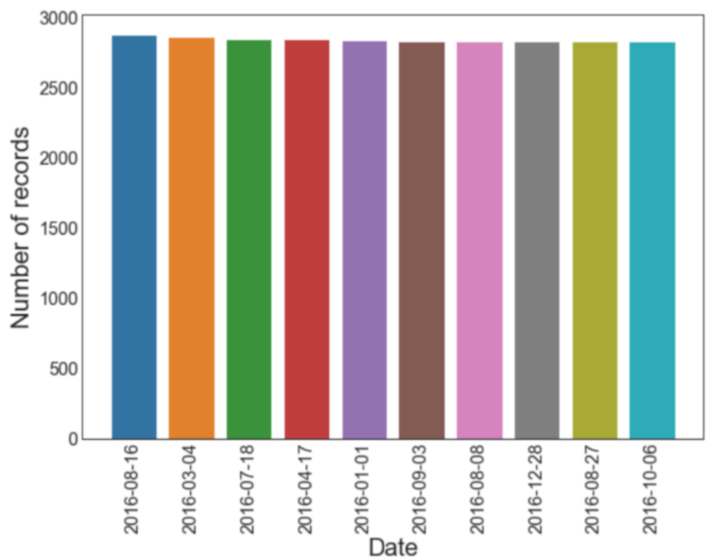

Figure 2.1 Distribution of 'date' variable

ssn (Categorical, 9-digit code)

This categorical variable defined the social security number of the applicant for each record/row. There were 835,819 unique values for this field with no missing/null values. Following is a distribution of the top 10 categories –

ssn	count
999999999	16935
938972725	114
810776805	81
829352390	74
473311863	44
165764866	34
295308935	30
189622157	30
163830210	29
588726212	29

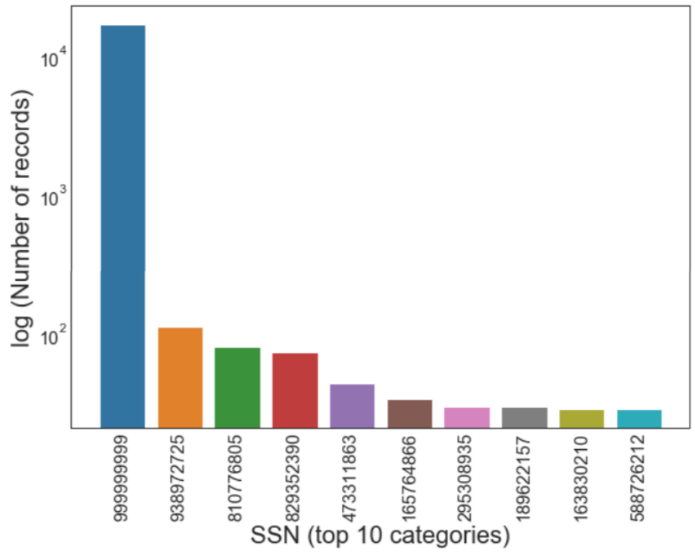

Fig 1.2 Categorical distribution of 'ssn' variable

We observed that ~17,000 values have SSN as '999999999'. This value could have been used to fill in missing values or where the SSN of the applicant was not available.

firstname (Categorical, string)

This categorical variable defined the first name of the applicant for each record/row. There were 78,136 unique values for this field with no missing/null values. Following is a distribution of the top 10 categories –

first name	count
EAMSTRMT	12658
TXEMXZZM	10297
UXXJJZTUZ	10235
UJSRSMUEZ	9994
SREZUJMJU	9688
EASEXMJAT	7576
SSSXUEJMS	6923
SZUASTTA	6878
EREMTZXXA	6717
EAXRRUMUX	5686

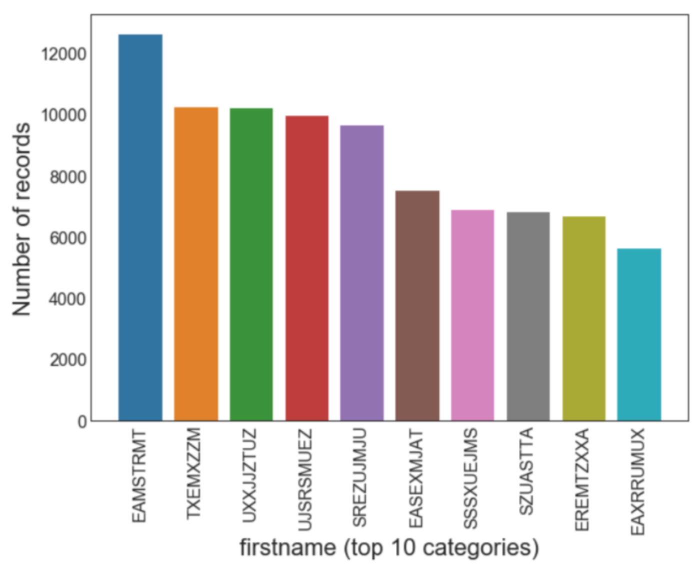

Fig 1.3 Categorical distribution of 'firstname' variable

lastname (Categorical, string)

This categorical variable defined the last name of the applicant for each record/row. There were 177,001 unique values for this field with no missing/null values. Following is a distribution of the top 10 categories –

lastname	count
ERJSAXA	8580
UMXUUUSE	7156
UMARRMA	6832
MEAXJUX	5492
XMERRR	5451
SXZXJRJT	4340
EUSEZRAE	4173
USMATTUR	4036
ETERUXME	3762
RJURSTXJ	3575

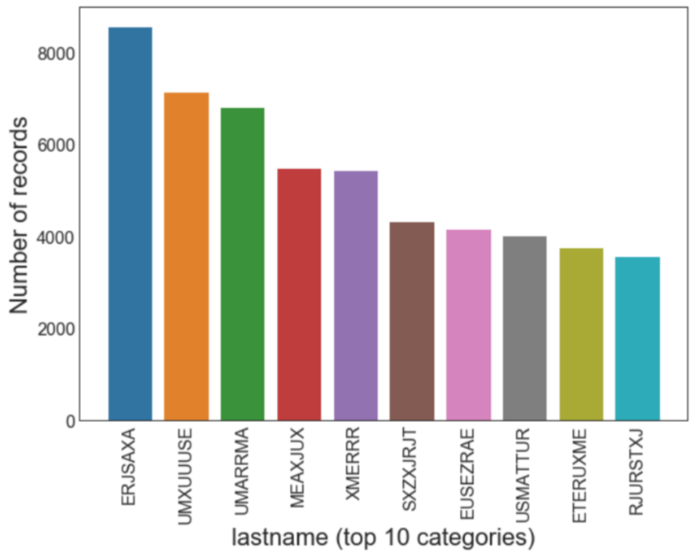

Fig 1.4 Categorical distribution of ‘lastname’ variable

address (Categorical, string)

This categorical variable defined the address of the applicant for each record/row. There were 828,774 unique values for this field with no missing/null values. Following is a distribution of the top 10 categories –

address	count
123 MAIN ST	1079
1775 XJXE LN	97
7433 RAEZA ST	80
8911 MZSU DR	74
4907 RRAAU DR	73
426 XUAXZ BLVD	57
3545 ARMA ST	44
606 EZZAU WY	44
4530 ETSMX WY	42
4292 RUSMM LN	41

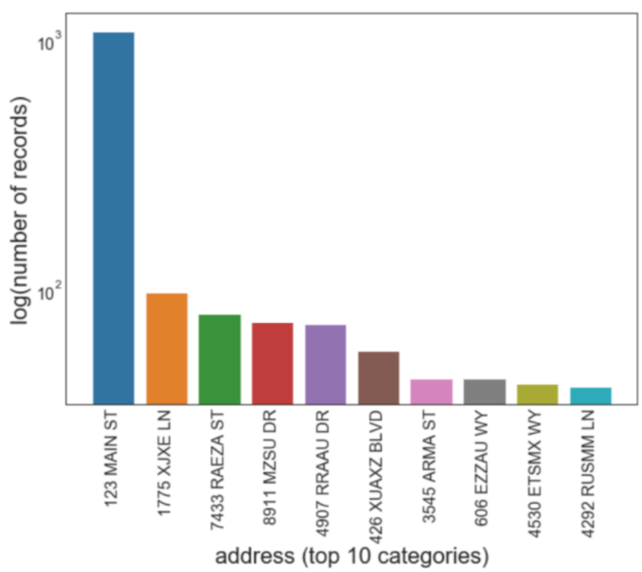

Fig 1.5 Categorical distribution of ‘address’ variable

zip5 (Categorical, 5-digit code)

This categorical variable defined the 5-digit zip code of the applicant for each record/row. There were 26,370 unique values for this field with no missing/null values. Following is a distribution of the top 10 categories –

zip5	count
68138	823
90042	514
89835	489
35227	478
14931	459
86500	438
12700	436
1362	434
59695	432
52317	432

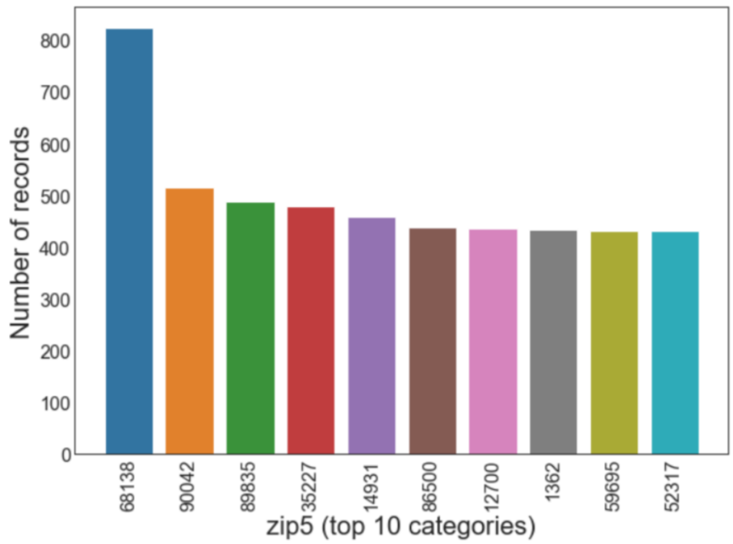

Fig 1.6 Categorical distribution of 'zip5' variable

dob (Categorical, datetime)

This categorical variable defined the date of birth of the applicant for each record/row. There were 42,673 unique values for this field with no missing/null values. Following is a distribution of the top 10 categories –

dob	count
1907-06-26	126568
1964-03-18	4818
1976-06-25	3723
1988-06-28	1404
1974-02-16	980
2009-01-27	280
1946-09-01	135
1959-12-08	126
1928-06-11	120
1967-02-15	102

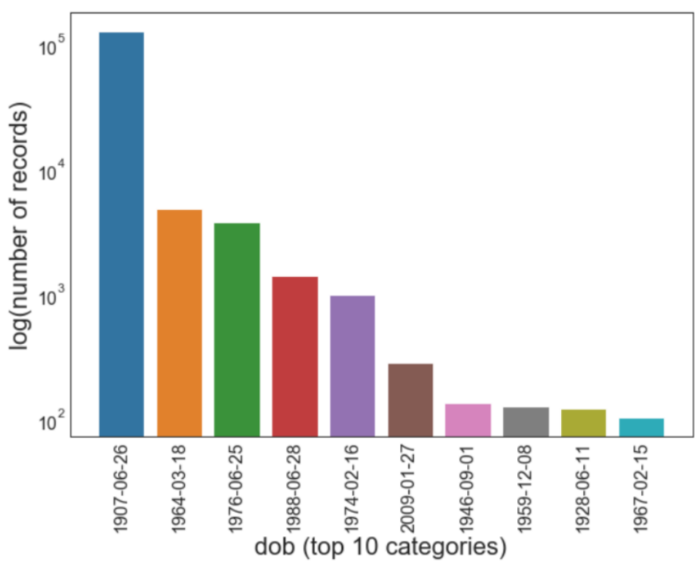

Fig 1.7 Categorical distribution of 'dob' variable

homephone (Categorical, 10-digit code)

This categorical variable defined the homephone of the applicant for each record/row. There were 28,244 unique values for this field with no missing/null values. Following is a distribution of the top 10 categories –

homephone	count
9999999999	78512
6384782007	466
6035129044	360
2113738531	331
4024680535	198
2669445638	172
6637507363	169
8629049955	139
3364980740	110
6845279110	108

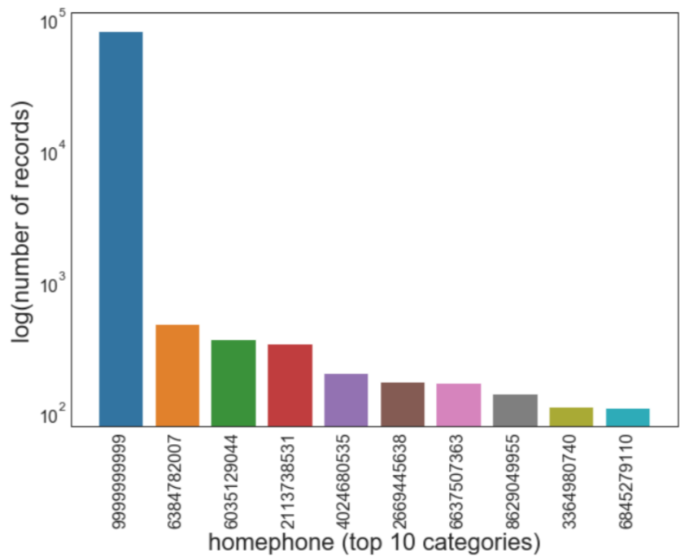

Fig 1.8 Categorical distribution of 'homephone' variable

Fraud_label (Categorical, 0 or 1)

This categorical variable indicated if the record/applicant is fraud or not. There were 2 unique values for this field with no missing/null values. Following is the distribution of the two categories –

fraud_label	count
0	985607
1	14393

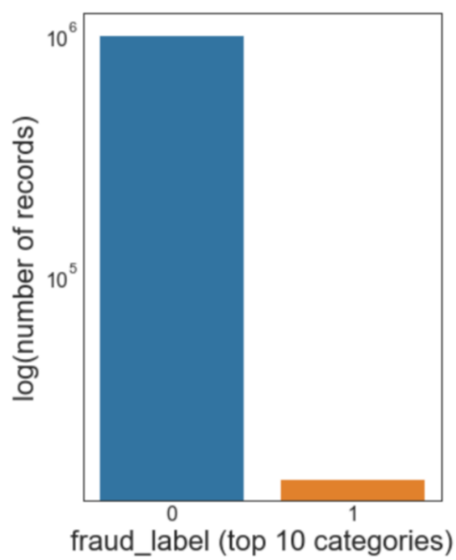

Fig 1.9 Categorical distribution of 'fraud_label' variable

All Expert Variables

1	record	145	name-dob-addr_lag1_lag30_avg
2	date	146	name-dob-homephone_#days_since
3	ssn_#days_since	147	name-dob-homephone_lag0_count
4	ssn_lag0_count	148	name-dob-homephone_lag1_count
5	ssn_lag1_count	149	name-dob-homephone_lag3_count
6	ssn_lag3_count	150	name-dob-homephone_lag7_count
7	ssn_lag7_count	151	name-dob-homephone_lag14_count
8	ssn_lag14_count	152	name-dob-homephone_lag30_count
9	ssn_lag30_count	153	name-dob-homephone_lag1_lag3_avg
10	ssn_lag1_lag3_avg	154	name-dob-homephone_lag1_lag7_avg
11	ssn_lag1_lag7_avg	155	name-dob-homephone_lag1_lag14_avg
12	ssn_lag1_lag14_avg	156	name-dob-homephone_lag1_lag30_avg
13	ssn_lag1_lag30_avg	157	name-addr-homephone_#days_since
14	address_#days_since	158	name-addr-homephone_lag0_count
15	address_lag0_count	159	name-addr-homephone_lag1_count
16	address_lag1_count	160	name-addr-homephone_lag3_count
17	address_lag3_count	161	name-addr-homephone_lag7_count
18	address_lag7_count	162	name-addr-homephone_lag14_count
19	address_lag14_count	163	name-addr-homephone_lag30_count
20	address_lag30_count	164	name-addr-homephone_lag1_lag3_avg
21	address_lag1_lag3_avg	165	name-addr-homephone_lag1_lag7_avg
22	address_lag1_lag7_avg	166	name-addr-homephone_lag1_lag14_avg
23	address_lag1_lag14_avg	167	name-addr-homephone_lag1_lag30_avg
24	address_lag1_lag30_avg	168	dob-addr-homephone_#days_since
25	dob_#days_since	169	dob-addr-homephone_lag0_count
26	dob_lag0_count	170	dob-addr-homephone_lag1_count
27	dob_lag1_count	171	dob-addr-homephone_lag3_count
28	dob_lag3_count	172	dob-addr-homephone_lag7_count
29	dob_lag7_count	173	dob-addr-homephone_lag14_count
30	dob_lag14_count	174	dob-addr-homephone_lag30_count
31	dob_lag30_count	175	dob-addr-homephone_lag1_lag3_avg
32	dob_lag1_lag3_avg	176	dob-addr-homephone_lag1_lag7_avg
33	dob_lag1_lag7_avg	177	dob-addr-homephone_lag1_lag14_avg
34	dob_lag1_lag14_avg	178	dob-addr-homephone_lag1_lag30_avg
35	dob_lag1_lag30_avg	179	name-dob-addr-homephone_#days_since
36	homephone_#days_since	180	name-dob-addr-homephone_lag0_count
37	homephone_lag0_count	181	name-dob-addr-homephone_lag1_count
38	homephone_lag1_count	182	name-dob-addr-homephone_lag3_count
39	homephone_lag3_count	183	name-dob-addr-homephone_lag7_count
40	homephone_lag7_count	184	name-dob-addr-homephone_lag14_count
41	homephone_lag14_count	185	name-dob-addr-homephone_lag30_count
42	homephone_lag30_count	186	name-dob-addr-homephone_lag1_lag3_avg
43	homephone_lag1_lag3_avg	187	name-dob-addr-homephone_lag1_lag7_avg

44	homephone_lag1_lag7_avg	188	name-dob-addr-homephone_lag1_lag14_avg
45	homephone_lag1_lag14_avg	189	name-dob-addr-homephone_lag1_lag30_avg
46	homephone_lag1_lag30_avg	190	ssn-firstname_#days_since
47	name_#days_since	191	ssn-firstname_lag0_count
48	name_lag0_count	192	ssn-firstname_lag1_count
49	name_lag1_count	193	ssn-firstname_lag3_count
50	name_lag3_count	194	ssn-firstname_lag7_count
51	name_lag7_count	195	ssn-firstname_lag14_count
52	name_lag14_count	196	ssn-firstname_lag30_count
53	name_lag30_count	197	ssn-firstname_lag1_lag3_avg
54	name_lag1_lag3_avg	198	ssn-firstname_lag1_lag7_avg
55	name_lag1_lag7_avg	199	ssn-firstname_lag1_lag14_avg
56	name_lag1_lag14_avg	200	ssn-firstname_lag1_lag30_avg
57	name_lag1_lag30_avg	201	ssn-lastname_#days_since
58	addr_#days_since	202	ssn-lastname_lag0_count
59	addr_lag0_count	203	ssn-lastname_lag1_count
60	addr_lag1_count	204	ssn-lastname_lag3_count
61	addr_lag3_count	205	ssn-lastname_lag7_count
62	addr_lag7_count	206	ssn-lastname_lag14_count
63	addr_lag14_count	207	ssn-lastname_lag30_count
64	addr_lag30_count	208	ssn-lastname_lag1_lag3_avg
65	addr_lag1_lag3_avg	209	ssn-lastname_lag1_lag7_avg
66	addr_lag1_lag7_avg	210	ssn-lastname_lag1_lag14_avg
67	addr_lag1_lag14_avg	211	ssn-lastname_lag1_lag30_avg
68	addr_lag1_lag30_avg	212	ssn-address_#days_since
69	name-dob_#days_since	213	ssn-address_lag0_count
70	name-dob_lag0_count	214	ssn-address_lag1_count
71	name-dob_lag1_count	215	ssn-address_lag3_count
72	name-dob_lag3_count	216	ssn-address_lag7_count
73	name-dob_lag7_count	217	ssn-address_lag14_count
74	name-dob_lag14_count	218	ssn-address_lag30_count
75	name-dob_lag30_count	219	ssn-address_lag1_lag3_avg
76	name-dob_lag1_lag3_avg	220	ssn-address_lag1_lag7_avg
77	name-dob_lag1_lag7_avg	221	ssn-address_lag1_lag14_avg
78	name-dob_lag1_lag14_avg	222	ssn-address_lag1_lag30_avg
79	name-dob_lag1_lag30_avg	223	ssn-zip5_#days_since
80	name-addr_#days_since	224	ssn-zip5_lag0_count
81	name-addr_lag0_count	225	ssn-zip5_lag1_count
82	name-addr_lag1_count	226	ssn-zip5_lag3_count
83	name-addr_lag3_count	227	ssn-zip5_lag7_count
84	name-addr_lag7_count	228	ssn-zip5_lag14_count
85	name-addr_lag14_count	229	ssn-zip5_lag30_count
86	name-addr_lag30_count	230	ssn-zip5_lag1_lag3_avg
87	name-addr_lag1_lag3_avg	231	ssn-zip5_lag1_lag7_avg

88	name-addr_lag1_lag7_avg	232	ssn-zip5_lag1_lag14_avg
89	name-addr_lag1_lag14_avg	233	ssn-zip5_lag1_lag30_avg
90	name-addr_lag1_lag30_avg	234	ssn-dob_#days_since
91	name-homephone_#days_since	235	ssn-dob_lag0_count
92	name-homephone_lag0_count	236	ssn-dob_lag1_count
93	name-homephone_lag1_count	237	ssn-dob_lag3_count
94	name-homephone_lag3_count	238	ssn-dob_lag7_count
95	name-homephone_lag7_count	239	ssn-dob_lag14_count
96	name-homephone_lag14_count	240	ssn-dob_lag30_count
97	name-homephone_lag30_count	241	ssn-dob_lag1_lag3_avg
98	name-homephone_lag1_lag3_avg	242	ssn-dob_lag1_lag7_avg
99	name-homephone_lag1_lag7_avg	243	ssn-dob_lag1_lag14_avg
100	name-homephone_lag1_lag14_avg	244	ssn-dob_lag1_lag30_avg
101	name-homephone_lag1_lag30_avg	245	ssn-homephone_#days_since
102	dob-addr_#days_since	246	ssn-homephone_lag0_count
103	dob-addr_lag0_count	247	ssn-homephone_lag1_count
104	dob-addr_lag1_count	248	ssn-homephone_lag3_count
105	dob-addr_lag3_count	249	ssn-homephone_lag7_count
106	dob-addr_lag7_count	250	ssn-homephone_lag14_count
107	dob-addr_lag14_count	251	ssn-homephone_lag30_count
108	dob-addr_lag30_count	252	ssn-homephone_lag1_lag3_avg
109	dob-addr_lag1_lag3_avg	253	ssn-homephone_lag1_lag7_avg
110	dob-addr_lag1_lag7_avg	254	ssn-homephone_lag1_lag14_avg
111	dob-addr_lag1_lag14_avg	255	ssn-homephone_lag1_lag30_avg
112	dob-addr_lag1_lag30_avg	256	ssn-name_#days_since
113	dob-homephone_#days_since	257	ssn-name_lag0_count
114	dob-homephone_lag0_count	258	ssn-name_lag1_count
115	dob-homephone_lag1_count	259	ssn-name_lag3_count
116	dob-homephone_lag3_count	260	ssn-name_lag7_count
117	dob-homephone_lag7_count	261	ssn-name_lag14_count
118	dob-homephone_lag14_count	262	ssn-name_lag30_count
119	dob-homephone_lag30_count	263	ssn-name_lag1_lag3_avg
120	dob-homephone_lag1_lag3_avg	264	ssn-name_lag1_lag7_avg
121	dob-homephone_lag1_lag7_avg	265	ssn-name_lag1_lag14_avg
122	dob-homephone_lag1_lag14_avg	266	ssn-name_lag1_lag30_avg
123	dob-homephone_lag1_lag30_avg	267	ssn-addr_#days_since
124	addr-homephone_#days_since	268	ssn-addr_lag0_count
125	addr-homephone_lag0_count	269	ssn-addr_lag1_count
126	addr-homephone_lag1_count	270	ssn-addr_lag3_count
127	addr-homephone_lag3_count	271	ssn-addr_lag7_count
128	addr-homephone_lag7_count	272	ssn-addr_lag14_count
129	addr-homephone_lag14_count	273	ssn-addr_lag30_count
130	addr-homephone_lag30_count	274	ssn-addr_lag1_lag3_avg
131	addr-homephone_lag1_lag3_avg	275	ssn-addr_lag1_lag7_avg

132	addr-homephone_lag1_lag7_avg	276	ssn-addr_lag1_lag14_avg
133	addr-homephone_lag1_lag14_avg	277	ssn-addr_lag1_lag30_avg
134	addr-homephone_lag1_lag30_avg	278	ssn-name-dob_#days_since
135	name-dob-addr_#days_since	279	ssn-name-dob_lag0_count
136	name-dob-addr_lag0_count	280	ssn-name-dob_lag1_count
137	name-dob-addr_lag1_count	281	ssn-name-dob_lag3_count
138	name-dob-addr_lag3_count	282	ssn-name-dob_lag7_count
139	name-dob-addr_lag7_count	283	ssn-name-dob_lag14_count
140	name-dob-addr_lag14_count	284	ssn-name-dob_lag30_count
141	name-dob-addr_lag30_count	285	ssn-name-dob_lag1_lag3_avg
142	name-dob-addr_lag1_lag3_avg	286	ssn-name-dob_lag1_lag7_avg
143	name-dob-addr_lag1_lag7_avg	287	ssn-name-dob_lag1_lag14_avg
144	name-dob-addr_lag1_lag14_avg	288	ssn-name-dob_lag1_lag30_avg

Table A.1 All Expert Variables

Expert Variables Ranked by KS and FDR

Field	KS	FDR	KS Rank	FDR Rank	Average Rank
fraud_label	1.0	1.0	292.0	292.0	292.0
addr_lag30_count	0.332032	0.354954	290.0	291.0	290.5
address_lag30_count	0.332725	0.353300	291.0	290.0	290.5
addr_#days_since	0.323543	0.349382	288.0	289.0	288.5
address_#days_since	0.324627	0.348076	289.0	288.0	288.5
address_lag14_count	0.322252	0.345812	287.0	287.0	287.0
addr_lag14_count	0.321756	0.342330	286.0	286.0	286.0
address_lag7_count	0.301445	0.320999	285.0	285.0	285.0
addr_lag7_count	0.301368	0.319955	284.0	284.0	284.0
address_lag3_count	0.278445	0.299060	282.0	283.0	282.5
addr_lag3_count	0.278488	0.297493	283.0	282.0	282.5
address_lag1_count	0.249267	0.268936	281.0	281.0	281.0
addr_lag1_count	0.249083	0.267717	280.0	280.0	280.0
addr-homephone_lag30_count	0.228954	0.255703	279.0	279.0	279.0
ssn-dob_lag30_count	0.228512	0.254745	278.0	278.0	278.0
name-dob_lag30_count	0.227623	0.254223	277.0	277.0	277.0
ssn_lag30_count	0.227027	0.253526	276.0	276.0	276.0
ssn-name-dob_lag30_count	0.226202	0.252394	275.0	275.0	275.0
ssn-firstname_lag30_count	0.226099	0.252307	273.0	274.0	273.5
ssn-lastname_lag30_count	0.226009	0.252133	272.0	273.0	272.5
ssn-name_lag30_count	0.224987	0.251698	271.0	272.0	271.5
addr-homephone_#days_since	0.226167	0.248302	274.0	271.0	272.5
addr-homephone_lag14_count	0.218906	0.245255	267.0	270.0	268.5
ssn-dob_#days_since	0.219637	0.243340	270.0	269.0	269.5
name-dob_#days_since	0.219290	0.243253	268.0	268.0	268.0
ssn_#days_since	0.218524	0.242643	266.0	267.0	266.5

ssn-firstname_#days_since	0.217755	0.241773	265.0	266.0	265.5
name-dob_lag14_count	0.215317	0.241598	260.0	265.0	262.5
ssn-name-dob_#days_since	0.217635	0.241424	264.0	264.0	264.0
ssn-lastname_#days_since	0.217486	0.241163	263.0	263.0	263.0
ssn-name_#days_since	0.216700	0.240815	261.0	262.0	261.5
ssn-dob_lag14_count	0.214858	0.240293	259.0	261.0	260.0
ssn_lag14_count	0.214434	0.240205	258.0	260.0	259.0
ssn-name-dob_lag14_count	0.213518	0.240031	255.0	259.0	257.0
name_lag30_count	0.213916	0.239770	257.0	258.0	257.5
ssn-lastname_lag14_count	0.213396	0.239596	254.0	257.0	255.5
ssn-firstname_lag14_count	0.213822	0.239335	256.0	256.0	256.0
ssn-name_lag14_count	0.213007	0.238551	253.0	255.0	254.0
address_lag1_lag14_avg	0.210771	0.237419	252.0	254.0	253.0
addr_lag1_lag14_avg	0.209092	0.235243	251.0	253.0	252.0
addr-homephone_lag7_count	0.199751	0.225318	248.0	252.0	250.0
name-dob_lag7_count	0.194062	0.220094	245.0	251.0	248.0
ssn-dob_lag7_count	0.193128	0.219223	244.0	250.0	247.0
ssn_lag7_count	0.193036	0.218614	243.0	249.0	246.0
ssn-name-dob_lag7_count	0.192461	0.218527	240.0	248.0	244.0
ssn-firstname_lag7_count	0.192673	0.218440	242.0	247.0	244.5
ssn-lastname_lag7_count	0.192597	0.218353	241.0	246.0	243.5
ssn-name_lag7_count	0.192358	0.218092	239.0	245.0	242.0
name_lag14_count	0.204487	0.210865	249.0	244.0	246.5
name_#days_since	0.205259	0.210169	250.0	243.0	246.5
address_lag1_lag7_avg	0.185147	0.209908	233.0	242.0	237.5
addr_lag1_lag7_avg	0.185152	0.209734	234.0	241.0	237.5
name_lag7_count	0.188519	0.209647	237.0	240.0	238.5
address_lag0_count	0.186847	0.208428	236.0	239.0	237.5
homephone_lag7_count	0.194198	0.208254	246.0	238.0	242.0
addr_lag0_count	0.186815	0.208166	235.0	237.0	236.0
addr-homephone_lag3_count	0.179292	0.205206	230.0	236.0	233.0
homephone_lag3_count	0.194923	0.204771	247.0	235.0	241.0
homephone_lag14_count	0.189357	0.201811	238.0	234.0	236.0
ssn-firstname_lag3_count	0.172088	0.199721	224.0	233.0	228.5
name-dob_lag3_count	0.172657	0.198851	226.0	232.0	229.0
ssn_lag3_count	0.172102	0.198328	225.0	231.0	228.0
name_lag3_count	0.169738	0.198241	219.0	229.5	224.25
ssn-dob_lag3_count	0.172059	0.198241	223.0	229.5	226.25
ssn-lastname_lag3_count	0.171914	0.198067	221.0	227.5	224.25
ssn-name_lag3_count	0.171928	0.198067	222.0	227.5	224.75
ssn-name-dob_lag3_count	0.171814	0.197980	220.0	226.0	223.0
homephone_lag1_count	0.179189	0.194846	228.5	225.0	226.75
ssn-dob_lag1_lag30_avg	0.162006	0.190667	218.0	224.0	221.0
name-dob_lag1_lag30_avg	0.161293	0.190232	216.0	223.0	219.5

ssn_lag1_lag30_avg	0.160498	0.188055	214.0	222.0	218.0
addr-homephone_lag1_lag30_avg	0.160162	0.187968	213.0	221.0	217.0
ssn-firstname_lag1_lag30_avg	0.159554	0.187271	211.0	220.0	215.5
ssn-lastname_lag1_lag30_avg	0.159553	0.187097	210.0	219.0	214.5
ssn-name-dob_lag1_lag30_avg	0.159774	0.186836	212.0	218.0	215.0
ssn-name_lag1_lag30_avg	0.158524	0.185617	209.0	217.0	213.0
addr-homephone_lag1_count	0.157627	0.184224	207.0	216.0	211.5
addr_lag1_lag30_avg	0.217349	0.183615	262.0	215.0	238.5
ssn-dob_lag1_lag14_avg	0.150097	0.179610	204.0	214.0	209.0
addr-homephone_lag1_lag14_avg	0.150730	0.178739	205.0	213.0	209.0
ssn-lastname_lag1_lag14_avg	0.148687	0.178652	196.0	212.0	204.0
name-dob_lag1_lag14_avg	0.150733	0.177956	206.0	211.0	208.5
ssn_lag1_lag14_avg	0.149656	0.177869	203.0	210.0	206.5
name_lag1_count	0.148064	0.177346	190.0	209.0	199.5
ssn-name_lag1_lag14_avg	0.148291	0.177172	191.0	208.0	199.5
ssn-firstname_lag1_lag14_avg	0.149025	0.176998	201.0	207.0	204.0
name-dob_lag1_count	0.148751	0.175605	199.0	206.0	202.5
ssn-dob_lag1_count	0.148578	0.175518	193.0	204.5	198.75
ssn-name-dob_lag1_lag14_avg	0.148835	0.175518	200.0	204.5	202.25
ssn-name-dob_lag1_count	0.148501	0.175431	192.0	202.0	197.0
ssn_lag1_count	0.148704	0.175431	198.0	202.0	200.0
ssn-firstname_lag1_count	0.148700	0.175431	197.0	202.0	199.5
ssn-name_lag1_count	0.148620	0.175344	195.0	199.5	197.25
ssn-lastname_lag1_count	0.148613	0.175344	194.0	199.5	196.75
address_lag1_lag30_avg	0.219441	0.173516	269.0	198.0	233.5
homephone_lag30_count	0.180430	0.172993	232.0	197.0	214.5
dob_lag7_count	0.161562	0.171687	217.0	196.0	206.5
dob_lag14_count	0.173890	0.170904	227.0	195.0	211.0
dob_lag3_count	0.157717	0.170207	208.0	194.0	201.0
dob_lag30_count	0.180134	0.169424	231.0	193.0	212.0
addr_lag1_lag3_avg	0.138655	0.166202	186.0	192.0	189.0
address_lag1_lag3_avg	0.138528	0.166115	185.0	191.0	188.0
addr-homephone_lag1_lag7_avg	0.128012	0.155581	184.0	190.0	187.0
name-dob_lag1_lag7_avg	0.125393	0.153927	183.0	189.0	186.0
ssn_lag1_lag7_avg	0.124270	0.153056	181.0	188.0	184.5
ssn-dob_lag1_lag7_avg	0.124369	0.152969	182.0	187.0	184.5
ssn-name_lag1_lag7_avg	0.123644	0.152882	177.0	186.0	181.5
ssn-firstname_lag1_lag7_avg	0.123878	0.152795	179.0	184.5	181.75
ssn-lastname_lag1_lag7_avg	0.123890	0.152795	180.0	184.5	182.25
ssn-name-dob_lag1_lag7_avg	0.123779	0.152446	178.0	183.0	180.5
dob_lag1_count	0.142622	0.150792	188.0	182.0	185.0
homephone_lag0_count	0.149143	0.149573	202.0	181.0	191.5
name_lag1_lag7_avg	0.121350	0.148964	176.0	180.0	178.0
addr-homephone_lag0_count	0.115963	0.143914	173.0	179.0	176.0

name_lag1_lag30_avg	0.143573	0.135382	189.0	178.0	183.5
name_lag0_count	0.107146	0.135121	172.0	177.0	174.5
ssn_lag0_count	0.107093	0.134947	171.0	176.0	173.5
ssn-firstname_lag0_count	0.107039	0.134860	170.0	175.0	172.5
name-dob_lag0_count	0.106993	0.134773	168.5	172.5	170.5
ssn-lastname_lag0_count	0.106952	0.134773	166.0	172.5	169.25
ssn-name_lag0_count	0.106953	0.134773	167.0	172.5	169.75
ssn-dob_lag0_count	0.106993	0.134773	168.5	172.5	170.5
ssn-name-dob_lag0_count	0.106910	0.134686	165.0	170.0	167.5
dob_lag0_count	0.102196	0.124935	163.0	169.0	166.0
addr-homephone_lag1_lag3_avg	0.091400	0.120059	159.0	168.0	163.5
dob_#days_since	0.160599	0.118927	215.0	167.0	191.0
name-dob_lag1_lag3_avg	0.089132	0.118840	158.0	166.0	162.0
ssn_lag1_lag3_avg	0.088679	0.118492	157.0	165.0	161.0
ssn-firstname_lag1_lag3_avg	0.088614	0.118405	155.0	164.0	159.5
ssn-lastname_lag1_lag3_avg	0.088527	0.118318	153.0	162.0	157.5
ssn-name_lag1_lag3_avg	0.088535	0.118318	154.0	162.0	158.0
ssn-dob_lag1_lag3_avg	0.088620	0.118318	156.0	162.0	159.0
ssn-name-dob_lag1_lag3_avg	0.088452	0.118144	152.0	160.0	156.0
name_lag1_lag3_avg	0.087846	0.117273	151.0	159.0	155.0
name_lag1_lag14_avg	0.141516	0.111440	187.0	158.0	172.5
homephone_#days_since	0.179189	0.109612	228.5	157.0	192.75
dob-homephone_lag30_count	0.065561	0.089326	144.0	156.0	150.0
name-dob-homephone_lag1_lag30_avg	0.060966	0.089239	130.0	155.0	142.5
dob-homephone_lag1_lag30_avg	0.061037	0.088456	131.0	154.0	142.5
ssn-homephone_lag1_lag30_avg	0.060486	0.088107	124.0	153.0	138.5
ssn-homephone_lag30_count	0.064923	0.088020	137.0	152.0	144.5
name-homephone_lag1_lag30_avg	0.061272	0.087759	133.0	151.0	142.0
dob-addr_lag1_lag30_avg	0.062788	0.087585	134.0	150.0	142.0
name-dob-addr_lag1_lag30_avg	0.062800	0.087237	135.0	148.5	141.75
name-dob-addr_lag30_count	0.067328	0.087237	148.0	148.5	148.25
dob-addr_lag30_count	0.067315	0.087063	147.0	147.0	147.0
ssn-address_lag30_count	0.065268	0.086975	140.0	146.0	143.0
name-addr_lag1_lag30_avg	0.063093	0.086801	136.0	145.0	140.5
name-dob-homephone_lag30_count	0.065493	0.086453	143.0	144.0	143.5
name-dob-addr-homephone_lag1_lag30_avg	0.060799	0.086366	129.0	143.0	136.0
ssn-addr_lag1_lag30_avg	0.060660	0.085757	126.0	142.0	134.0
ssn-address_lag1_lag30_avg	0.060746	0.085582	127.0	140.0	133.5
dob-addr-homephone_lag1_lag30_avg	0.060795	0.085582	128.0	140.0	134.0
name-homephone_lag30_count	0.065798	0.085582	146.0	140.0	143.0
dob-addr-homephone_lag30_count	0.065323	0.085495	141.0	138.0	139.5
ssn-zip5_lag1_lag30_avg	0.060655	0.085060	125.0	137.0	131.0
name-dob-addr-homephone_lag30_count	0.065327	0.084973	142.0	136.0	139.0

ssn-addr_lag30_count	0.065183	0.084886	139.0	135.0	137.0
name-addr_lag30_count	0.067702	0.084799	149.0	134.0	141.5
name-addr-homephone_lag30_count	0.065634	0.084102	145.0	133.0	139.0
ssn-zip5_lag30_count	0.065178	0.083754	138.0	131.5	134.75
name-addr-homephone_lag1_lag30_avg	0.061108	0.083754	132.0	131.5	131.75
dob-homephone_#days_since	0.058542	0.082187	122.0	130.0	126.0
name-dob-addr_#days_since	0.056644	0.081403	118.0	129.0	123.5
ssn-homephone_#days_since	0.056909	0.081142	121.0	128.0	124.5
dob-addr_#days_since	0.056785	0.080620	119.0	126.5	122.75
name-dob-addr_lag14_count	0.048290	0.080620	107.0	126.5	116.75
name-dob-homephone_#days_since	0.056786	0.080446	120.0	125.0	122.5
ssn-address_#days_since	0.055952	0.080359	116.0	124.0	120.0
dob-addr-homephone_#days_since	0.054441	0.080098	114.0	122.5	118.25
name-addr_#days_since	0.055595	0.080098	115.0	122.5	118.75
name-homephone_#days_since	0.056043	0.079923	117.0	121.0	119.0
name-dob-homephone_lag14_count	0.047537	0.079662	103.0	120.0	111.5
name-addr-homephone_#days_since	0.053714	0.078792	110.0	119.0	114.5
name-homephone_lag14_count	0.047291	0.078705	102.0	118.0	110.0
ssn-addr_#days_since	0.054177	0.078617	112.0	116.0	114.0
name-dob-addr_lag1_lag14_avg	0.046238	0.078617	97.0	116.0	106.5
ssn-zip5_#days_since	0.054110	0.078617	111.0	116.0	113.5
name-dob-addr-homephone_#days_since	0.054380	0.078530	113.0	114.0	113.5
dob-addr-homephone_lag14_count	0.046196	0.078443	96.0	113.0	104.5
name-addr-homephone_lag14_count	0.045881	0.078095	94.0	112.0	103.0
dob-homephone_lag14_count	0.048824	0.078008	109.0	111.0	110.0
name-dob-homephone_lag1_lag14_avg	0.045456	0.077921	90.0	109.5	99.75
dob-addr_lag14_count	0.048568	0.077921	108.0	109.5	108.75
ssn-address_lag1_lag14_avg	0.045465	0.077834	91.0	107.5	99.25
name-addr_lag14_count	0.047797	0.077834	106.0	107.5	106.75
ssn-address_lag14_count	0.047640	0.077660	104.0	106.0	105.0
ssn-homephone_lag14_count	0.047782	0.077137	105.0	105.0	105.0
name-dob-addr-homephone_lag1_lag14_avg	0.044212	0.076702	85.0	103.5	94.25
ssn-zip5_lag1_lag14_avg	0.044272	0.076702	86.0	103.5	94.75
ssn-zip5_lag14_count	0.046447	0.076528	100.0	102.0	101.0
ssn-addr_lag14_count	0.046444	0.076441	99.0	101.0	100.0
dob-homephone_lag1_lag14_avg	0.046600	0.075744	101.0	100.0	100.5
dob-addr_lag1_lag14_avg	0.046438	0.075657	98.0	99.0	98.5
name-dob-addr-homephone_lag14_count	0.046126	0.075396	95.0	98.0	96.5
name-addr_lag1_lag14_avg	0.045719	0.075222	93.0	97.0	95.0
ssn-homephone_lag1_lag14_avg	0.045537	0.075048	92.0	96.0	94.0
name-homephone_lag1_lag14_avg	0.045140	0.074526	89.0	95.0	92.0
ssn-addr_lag1_lag14_avg	0.044351	0.073829	88.0	94.0	91.0

dob-addr-homephone_lag1_lag14_avg	0.044286	0.073655	87.0	93.0	90.0
name-addr-homephone_lag1_lag14_avg	0.043891	0.073220	84.0	92.0	88.0
dob-homephone_lag7_count	0.030275	0.061814	82.0	91.0	86.5
name-dob-addr_lag7_count	0.030192	0.061727	81.0	90.0	85.5
dob-addr_lag7_count	0.030331	0.061640	83.0	89.0	86.0
name-addr_lag7_count	0.030004	0.061379	80.0	88.0	84.0
ssn-address_lag7_count	0.029616	0.061205	77.0	87.0	82.0
name-dob-homephone_lag7_count	0.029689	0.061118	79.0	86.0	82.5
name-homephone_lag7_count	0.029624	0.060944	78.0	84.5	81.25
ssn-homephone_lag7_count	0.029562	0.060944	76.0	84.5	80.25
ssn-zip5_lag7_count	0.029064	0.060596	75.0	82.5	78.75
ssn-addr_lag7_count	0.029022	0.060596	72.0	82.5	77.25
dob-addr-homephone_lag7_count	0.029026	0.060508	73.0	81.0	77.0
name-dob-addr-homephone_lag7_count	0.029034	0.060421	74.0	79.5	76.75
name-addr-homephone_lag7_count	0.028929	0.060421	71.0	79.5	75.25
dob-addr_lag1_lag7_avg	0.026721	0.057984	68.0	78.0	73.0
name-dob-addr_lag1_lag7_avg	0.026659	0.057897	67.0	77.0	72.0
dob-homephone_lag1_lag7_avg	0.026485	0.057722	66.0	75.5	70.75
name-addr_lag1_lag7_avg	0.026447	0.057722	65.0	75.5	70.25
ssn-address_lag1_lag7_avg	0.025963	0.057374	63.0	74.0	68.5
ssn-homephone_lag1_lag7_avg	0.025838	0.057287	61.0	73.0	67.0
name-dob-homephone_lag1_lag7_avg	0.026042	0.057200	64.0	72.0	68.0
name-homephone_lag1_lag7_avg	0.025906	0.057113	62.0	71.0	66.5
dob-addr-homephone_lag1_lag7_avg	0.025549	0.057026	59.0	70.0	64.5
ssn-zip5_lag1_lag7_avg	0.025410	0.056765	57.0	68.0	62.5
name-dob-addr-homephone_lag1_lag7_avg	0.025552	0.056765	60.0	68.0	64.0
ssn-addr_lag1_lag7_avg	0.025449	0.056765	58.0	68.0	63.0
dob_lag1_lag7_avg	0.119221	0.056678	174.0	65.5	119.75
name-addr-homephone_lag1_lag7_avg	0.025372	0.056678	56.0	65.5	60.75
dob_lag1_lag30_avg	0.100089	0.056068	162.0	64.0	113.0
homephone_lag1_lag3_avg	0.096837	0.054849	161.0	63.0	112.0
dob_lag1_lag14_avg	0.119389	0.054501	175.0	62.0	118.5
dob_lag1_lag3_avg	0.092116	0.050061	160.0	61.0	110.5
homephone_lag1_lag7_avg	0.104764	0.047710	164.0	60.0	112.0
name-dob-addr_lag3_count	0.015159	0.047449	52.0	59.0	55.5
name-addr_lag3_count	0.015262	0.047362	54.0	57.5	55.75
dob-addr_lag3_count	0.015231	0.047362	53.0	57.5	55.25
dob-homephone_lag3_count	0.015056	0.047275	51.0	56.0	53.5
name-dob-homephone_lag3_count	0.014861	0.047188	48.0	55.0	51.5
ssn-address_lag3_count	0.014941	0.047101	50.0	54.0	52.0
ssn-zip5_lag3_count	0.014848	0.047014	47.0	52.0	49.5
ssn-addr_lag3_count	0.014776	0.047014	46.0	52.0	49.0
name-homephone_lag3_count	0.014903	0.047014	49.0	52.0	50.5

ssn-homephone_lag3_count	0.014739	0.046840	45.0	49.0	47.0
name-dob-addr-homephone_lag3_count	0.014537	0.046840	43.0	49.0	46.0
dob-addr-homephone_lag3_count	0.014532	0.046840	42.0	49.0	45.5
homephone_lag1_lag14_avg	0.075148	0.046753	150.0	47.0	98.5
name-addr-homephone_lag3_count	0.014593	0.046666	44.0	46.0	45.0
name-dob-addr_lag1_lag3_avg	0.010321	0.042399	39.0	45.0	42.0
name-addr_lag1_lag3_avg	0.010400	0.042312	40.0	43.5	41.75
dob-addr_lag1_lag3_avg	0.010315	0.042312	38.0	43.5	40.75
dob-homephone_lag1_lag3_avg	0.009960	0.042051	36.0	42.0	39.0
name-dob-homephone_lag1_lag3_avg	0.009907	0.041964	35.0	41.0	38.0
ssn-address_lag1_lag3_avg	0.009983	0.041877	37.0	40.0	38.5
dob-addr-homephone_lag1_lag3_avg	0.009749	0.041790	30.0	37.0	33.5
ssn-addr_lag1_lag3_avg	0.009899	0.041790	34.0	37.0	35.5
name-dob-addr-homephone_lag1_lag3_avg	0.009750	0.041790	31.0	37.0	34.0
name-homephone_lag1_lag3_avg	0.009878	0.041790	32.0	37.0	34.5
ssn-zip5_lag1_lag3_avg	0.009889	0.041790	33.0	37.0	35.0
ssn-homephone_lag1_lag3_avg	0.009709	0.041616	28.0	33.5	30.75
name-addr-homephone_lag1_lag3_avg	0.009730	0.041616	29.0	33.5	31.25
homephone_lag1_lag30_avg	0.059510	0.040484	123.0	32.0	77.5
dob-homephone_lag1_count	0.005791	0.038482	27.0	31.0	29.0
dob-addr_lag1_count	0.005700	0.038395	21.0	30.0	25.5
ssn-homephone_lag1_count	0.005726	0.038308	23.0	27.5	25.25
name-dob-homephone_lag1_count	0.005650	0.038308	19.0	27.5	23.25
name-dob-addr_lag1_count	0.005622	0.038308	18.0	27.5	22.75
name-homephone_lag1_count	0.005721	0.038308	22.0	27.5	24.75
name-dob-addr-homephone_lag1_count	0.005484	0.038220	16.0	23.5	19.75
ssn-zip5_lag1_count	0.005743	0.038220	26.0	23.5	24.75
ssn-address_lag1_count	0.005742	0.038220	25.0	23.5	24.25
name-addr_lag1_count	0.005733	0.038220	24.0	23.5	23.75
name-addr-homephone_lag1_count	0.005560	0.038133	17.0	20.0	18.5
dob-addr-homephone_lag1_count	0.005480	0.038133	15.0	20.0	17.5
ssn-addr_lag1_count	0.005661	0.038133	20.0	20.0	20.0
ssn-zip5_lag0_count	0.001857	0.034912	11.0	16.0	13.5
dob-homephone_lag0_count	0.001986	0.034912	13.0	16.0	14.5
name-addr_lag0_count	0.001850	0.034912	10.0	16.0	13.0
ssn-address_lag0_count	0.001858	0.034912	12.0	16.0	14.0
date	0.026996	0.034912	69.0	16.0	42.5
record	0.027045	0.034825	70.0	9.5	39.75
name-dob-addr-homephone_lag0_count	0.001826	0.034825	8.5	9.5	9.0
dob-addr-homephone_lag0_count	0.001826	0.034825	8.5	9.5	9.0
name-dob-homephone_lag0_count	0.001821	0.034825	7.0	9.5	8.25
dob-addr_lag0_count	0.001812	0.034825	5.0	9.5	7.25
name-addr-homephone_lag0_count	0.001791	0.034825	3.0	9.5	6.25

name-homephone_lag0_count	0.001782	0.034825	2.0	9.5	5.75
ssn-addr_lag0_count	0.001774	0.034825	1.0	9.5	5.25
name-dob-addr_lag0_count	0.001813	0.034738	6.0	4.5	5.25
ssn-homephone_lag0_count	0.001793	0.034738	4.0	4.5	4.25
weekday_risk	0.022088	0.033780	55.0	3.0	29.0
RANDOM	0.005414	0.030037	14.0	2.0	8.0
weekday	0.013294	0.028382	41.0	1.0	21.0

Table A.2 Expert Variables Ranked by KS and FDR

Training Data Statistics

Training	# Records	# Goods	# Bads	Fraud Rate
	596,247	587,587	8,660	1.45%

	Bin Statistics	**Cun Statistics**										
Pop Bin %	**Records**	**Goods**	**Bads**	**% Goods**	**% Bads**	**Total # Records**	**Cum Goods**	**Cum Bads**	**% Goods**	**% Bads**	**KS**	**FPR**
0	5963	1395	4568	23.39%	76.61%	5963	1395	4568	0.24%	52.75%	52.51%	0.31
1	5963	5828	135	97.74%	2.26%	11926	7223	4703	1.23%	54.31%	53.08%	1.54
2	5963	5913	50	99.16%	0.84%	17889	13136	4753	2.24%	54.88%	52.65%	2.76
3	5963	5883	80	98.66%	1.34%	23852	19019	4833	3.24%	55.81%	52.57%	3.94
4	5963	5899	64	98.93%	1.07%	29815	24918	4897	4.24%	56.55%	52.31%	5.09
5	5963	5896	67	98.88%	1.12%	35778	30814	4964	5.24%	57.32%	52.08%	6.21
6	5963	5895	68	98.86%	1.14%	41741	36709	5032	6.25%	58.11%	51.86%	7.30
7	5963	5917	46	99.23%	0.77%	47704	42626	5078	7.25%	58.64%	51.38%	8.39
8	5963	5929	34	99.43%	0.57%	53667	48555	5112	8.26%	59.03%	50.77%	9.50
9	5963	5925	38	99.36%	0.64%	59630	54480	5150	9.27%	59.47%	50.20%	10.58
10	5963	5912	51	99.14%	0.86%	65593	60392	5201	10.28%	60.06%	49.78%	11.61
11	5963	5930	33	99.45%	0.55%	71556	66322	5234	11.29%	60.44%	49.15%	12.67
12	5963	5915	48	99.20%	0.80%	77519	72237	5282	12.29%	60.99%	48.70%	13.68
13	5963	5918	45	99.25%	0.75%	83482	78155	5327	13.30%	61.51%	48.21%	14.67
14	5963	5938	25	99.58%	0.42%	89445	84093	5352	14.31%	61.80%	47.49%	15.71
15	5963	5929	34	99.43%	0.57%	95408	90022	5386	15.32%	62.19%	46.87%	16.71
16	5963	5924	39	99.35%	0.65%	101371	95946	5425	16.33%	62.64%	46.32%	17.69
17	5963	5925	38	99.36%	0.64%	107334	101871	5463	17.34%	63.08%	45.75%	18.65
18	5963	5930	33	99.45%	0.55%	113297	107801	5496	18.35%	63.46%	45.12%	19.61
19	5963	5932	31	99.48%	0.52%	119260	113733	5527	19.36%	63.82%	44.47%	20.58
20	5963	5926	37	99.38%	0.62%	125223	119659	5564	20.36%	64.25%	43.88%	21.51
21	5963	5921	42	99.30%	0.70%	131186	125580	5606	21.37%	64.73%	43.36%	22.40
22	5963	5919	44	99.26%	0.74%	137149	131499	5650	22.38%	65.24%	42.86%	23.27
23	5963	5930	33	99.45%	0.55%	143112	137429	5683	23.39%	65.62%	42.23%	24.18

24	5963	5930	33	99.45%	0.55%	149075	143359	5716	24.40%	66.00%	41.61%	25.08
25	5963	5921	42	99.30%	0.70%	155038	149280	5758	25.41%	66.49%	41.08%	25.93
26	5963	5912	51	99.14%	0.86%	161001	155192	5809	26.41%	67.08%	40.67%	26.72
27	5963	5929	34	99.43%	0.57%	166964	161121	5843	27.42%	67.47%	40.05%	27.58
28	5963	5924	39	99.35%	0.65%	172927	167045	5882	28.43%	67.92%	39.49%	28.40
29	5963	5929	34	99.43%	0.57%	178890	172974	5916	29.44%	68.31%	38.88%	29.24
30	5963	5922	41	99.31%	0.69%	184853	178896	5957	30.45%	68.79%	38.34%	30.03
31	5963	5930	33	99.45%	0.55%	190816	184826	5990	31.46%	69.17%	37.71%	30.86
32	5963	5919	44	99.26%	0.74%	196779	190745	6034	32.46%	69.68%	37.21%	31.61
33	5963	5915	48	99.20%	0.80%	202742	196660	6082	33.47%	70.23%	36.76%	32.33
34	5963	5927	36	99.40%	0.60%	208705	202587	6118	34.48%	70.65%	36.17%	33.11
35	5963	5926	37	99.38%	0.62%	214668	208513	6155	35.49%	71.07%	35.59%	33.88
36	5963	5918	45	99.25%	0.75%	220631	214431	6200	36.49%	71.59%	35.10%	34.59
37	5963	5917	46	99.23%	0.77%	226594	220348	6246	37.50%	72.12%	34.62%	35.28
38	5963	5925	38	99.36%	0.64%	232557	226273	6284	38.51%	72.56%	34.05%	36.01
39	5963	5934	29	99.51%	0.49%	238520	232207	6313	39.52%	72.90%	33.38%	36.78
40	5963	5925	38	99.36%	0.64%	244483	238132	6351	40.53%	73.34%	32.81%	37.50
41	5963	5914	49	99.18%	0.82%	250446	244046	6400	41.53%	73.90%	32.37%	38.13
42	5963	5907	56	99.06%	0.94%	256409	249953	6456	42.54%	74.55%	32.01%	38.72
43	5963	5917	46	99.23%	0.77%	262372	255870	6502	43.55%	75.08%	31.53%	39.35
44	5963	5917	46	99.23%	0.77%	268335	261787	6548	44.55%	75.61%	31.06%	39.98
45	5963	5930	33	99.45%	0.55%	274298	267717	6581	45.56%	75.99%	30.43%	40.68
46	5963	5924	39	99.35%	0.65%	280261	273641	6620	46.57%	76.44%	29.87%	41.34
47	5963	5936	27	99.55%	0.45%	286224	279577	6647	47.58%	76.76%	29.17%	42.06
48	5963	5924	39	99.35%	0.65%	292187	285501	6686	48.59%	77.21%	28.62%	42.70
49	5963	5918	45	99.25%	0.75%	298150	291419	6731	49.60%	77.73%	28.13%	43.30
50	5963	5919	44	99.26%	0.74%	304113	297338	6775	50.60%	78.23%	27.63%	43.89
51	5963	5926	37	99.38%	0.62%	310076	303264	6812	51.61%	78.66%	27.05%	44.52
52	5963	5917	46	99.23%	0.77%	316039	309181	6858	52.62%	79.19%	26.57%	45.08
53	5963	5913	50	99.16%	0.84%	322002	315094	6908	53.63%	79.77%	26.14%	45.61
54	5963	5926	37	99.38%	0.62%	327965	321020	6945	54.63%	80.20%	25.56%	46.22
55	5963	5915	48	99.20%	0.80%	333928	326935	6993	55.64%	80.75%	25.11%	46.75
56	5963	5919	44	99.26%	0.74%	339891	332854	7037	56.65%	81.26%	24.61%	47.30
57	5963	5922	41	99.31%	0.69%	345854	338776	7078	57.66%	81.73%	24.08%	47.86
58	5963	5927	36	99.40%	0.60%	351817	344703	7114	58.66%	82.15%	23.48%	48.45
59	5963	5924	39	99.35%	0.65%	357780	350627	7153	59.67%	82.60%	22.93%	49.02
60	5963	5918	45	99.25%	0.75%	363743	356545	7198	60.68%	83.12%	22.44%	49.53
61	5963	5928	35	99.41%	0.59%	369706	362473	7233	61.69%	83.52%	21.83%	50.11
62	5963	5926	37	99.38%	0.62%	375669	368399	7270	62.70%	83.95%	21.25%	50.67
63	5963	5929	34	99.43%	0.57%	381632	374328	7304	63.71%	84.34%	20.64%	51.25
64	5963	5916	47	99.21%	0.79%	387595	380244	7351	64.71%	84.88%	20.17%	51.73

65	5963	5927	36	99.40%	0.60%	393558	386171	7387	65.72%	85.30%	19.58%	52.28
66	5963	5924	39	99.35%	0.65%	399521	392095	7426	66.73%	85.75%	19.02%	52.80
67	5963	5920	43	99.28%	0.72%	405484	398015	7469	67.74%	86.25%	18.51%	53.29
68	5963	5922	41	99.31%	0.69%	411447	403937	7510	68.75%	86.72%	17.98%	53.79
69	5963	5928	35	99.41%	0.59%	417410	409865	7545	69.75%	87.12%	17.37%	54.32
70	5963	5921	42	99.30%	0.70%	423373	415786	7587	70.76%	87.61%	16.85%	54.80
71	5963	5929	34	99.43%	0.57%	429336	421715	7621	71.77%	88.00%	16.23%	55.34
72	5963	5936	27	99.55%	0.45%	435299	427651	7648	72.78%	88.31%	15.53%	55.92
73	5963	5930	33	99.45%	0.55%	441262	433581	7681	73.79%	88.70%	14.91%	56.45
74	5963	5930	33	99.45%	0.55%	447225	439511	7714	74.80%	89.08%	14.28%	56.98
75	5963	5941	22	99.63%	0.37%	453188	445452	7736	75.81%	89.33%	13.52%	57.58
76	5963	5924	39	99.35%	0.65%	459151	451376	7775	76.82%	89.78%	12.96%	58.05
77	5963	5919	44	99.26%	0.74%	465114	457295	7819	77.83%	90.29%	12.46%	58.49
78	5963	5933	30	99.50%	0.50%	471077	463228	7849	78.84%	90.64%	11.80%	59.02
79	5963	5922	41	99.31%	0.69%	477040	469150	7890	79.84%	91.11%	11.27%	59.46
80	5963	5919	44	99.26%	0.74%	483003	475069	7934	80.85%	91.62%	10.77%	59.88
81	5963	5923	40	99.33%	0.67%	488966	480992	7974	81.86%	92.08%	10.22%	60.32
82	5963	5922	41	99.31%	0.69%	494929	486914	8015	82.87%	92.55%	9.69%	60.75
83	5963	5923	40	99.33%	0.67%	500892	492837	8055	83.87%	93.01%	9.14%	61.18
84	5963	5923	40	99.33%	0.67%	506855	498760	8095	84.88%	93.48%	8.59%	61.61
85	5963	5922	41	99.31%	0.69%	512818	504682	8136	85.89%	93.95%	8.06%	62.03
86	5963	5917	46	99.23%	0.77%	518781	510599	8182	86.90%	94.48%	7.58%	62.41
87	5963	5919	44	99.26%	0.74%	524744	516518	8226	87.90%	94.99%	7.08%	62.79
88	5963	5918	45	99.25%	0.75%	530707	522436	8271	88.91%	95.51%	6.60%	63.16
89	5963	5923	40	99.33%	0.67%	536670	528359	8311	89.92%	95.97%	6.05%	63.57
90	5963	5924	39	99.35%	0.65%	542633	534283	8350	90.93%	96.42%	5.49%	63.99
91	5963	5922	41	99.31%	0.69%	548596	540205	8391	91.94%	96.89%	4.96%	64.38
92	5963	5926	37	99.38%	0.62%	554559	546131	8428	92.94%	97.32%	4.38%	64.80
93	5963	5922	41	99.31%	0.69%	560522	552053	8469	93.95%	97.79%	3.84%	65.19
94	5963	5927	36	99.40%	0.60%	566485	557980	8505	94.96%	98.21%	3.25%	65.61
95	5963	5913	50	99.16%	0.84%	572448	563893	8555	95.97%	98.79%	2.82%	65.91
96	5963	5930	33	99.45%	0.55%	578411	569823	8588	96.98%	99.17%	2.19%	66.35
97	5963	5930	33	99.45%	0.55%	584374	575753	8621	97.99%	99.55%	1.56%	66.78
98	5963	5938	25	99.58%	0.42%	590337	581691	8646	99.00%	99.84%	0.84%	67.28
99	5910	5896	14	99.76%	0.24%	596247	587587	8660	100.00%	#######	0.00%	67.85

Test Data Statistics

Test	# Records	# Goods	# Bads	Fraud Rate
	198,749	195,923	2,826	1.42%

	Bin Statistics	Cum Statistics										
Pop Bin %	Records	Goods	Bads	% Goods	% Bads	Total # Records	Cum Goods	Cum Bads	% Goods	% Bads	KS	FPR
0	1988	520	1468	26.16%	73.84%	1988	520	1468	0.27%	51.95%	51.68%	0.35
1	1988	1948	40	97.99%	2.01%	3976	2468	1508	1.26%	53.36%	52.10%	1.64
2	1988	1967	21	98.94%	1.06%	5964	4435	1529	2.26%	54.10%	51.84%	2.90
3	1988	1971	17	99.14%	0.86%	7952	6406	1546	3.27%	54.71%	51.44%	4.14
4	1988	1971	17	99.14%	0.86%	9940	8377	1563	4.28%	55.31%	51.03%	5.36
5	1988	1968	20	98.99%	1.01%	11928	10345	1583	5.28%	56.02%	50.74%	6.54
6	1988	1969	19	99.04%	0.96%	13916	12314	1602	6.29%	56.69%	50.40%	7.69
7	1988	1979	9	99.55%	0.45%	15904	14293	1611	7.30%	57.01%	49.71%	8.87
8	1988	1972	16	99.20%	0.80%	17892	16265	1627	8.30%	57.57%	49.27%	10.00
9	1988	1974	14	99.30%	0.70%	19880	18239	1641	9.31%	58.07%	48.76%	11.11
10	1988	1976	12	99.40%	0.60%	21868	20215	1653	10.32%	58.49%	48.17%	12.23
11	1988	1976	12	99.40%	0.60%	23856	22191	1665	11.33%	58.92%	47.59%	13.33
12	1988	1969	19	99.04%	0.96%	25844	24160	1684	12.33%	59.59%	47.26%	14.35
13	1988	1975	13	99.35%	0.65%	27832	26135	1697	13.34%	60.05%	46.71%	15.40
14	1988	1970	18	99.09%	0.91%	29820	28105	1715	14.34%	60.69%	46.34%	16.39
15	1988	1974	14	99.30%	0.70%	31808	30079	1729	15.35%	61.18%	45.83%	17.40
16	1988	1983	5	99.75%	0.25%	33796	32062	1734	16.36%	61.36%	44.99%	18.49
17	1988	1979	9	99.55%	0.45%	35784	34041	1743	17.37%	61.68%	44.30%	19.53
18	1988	1973	15	99.25%	0.75%	37772	36014	1758	18.38%	62.21%	43.83%	20.49
19	1988	1983	5	99.75%	0.25%	39760	37997	1763	19.39%	62.38%	42.99%	21.55
20	1988	1973	15	99.25%	0.75%	41748	39970	1778	20.40%	62.92%	42.51%	22.48
21	1988	1974	14	99.30%	0.70%	43736	41944	1792	21.41%	63.41%	42.00%	23.41
22	1988	1980	8	99.60%	0.40%	45724	43924	1800	22.42%	63.69%	41.28%	24.40
23	1988	1984	4	99.80%	0.20%	47712	45908	1804	23.43%	63.84%	40.40%	25.45
24	1988	1972	16	99.20%	0.80%	49700	47880	1820	24.44%	64.40%	39.96%	26.31
25	1988	1974	14	99.30%	0.70%	51688	49854	1834	25.45%	64.90%	39.45%	27.18
26	1988	1968	20	98.99%	1.01%	53676	51822	1854	26.45%	65.61%	39.15%	27.95
27	1988	1970	18	99.09%	0.91%	55664	53792	1872	27.46%	66.24%	38.79%	28.74
28	1988	1973	15	99.25%	0.75%	57652	55765	1887	28.46%	66.77%	38.31%	29.55
29	1988	1971	17	99.14%	0.86%	59640	57736	1904	29.47%	67.37%	37.91%	30.32
30	1988	1978	10	99.50%	0.50%	61628	59714	1914	30.48%	67.73%	37.25%	31.20

31	1988	1970	18	99.09%	0.91%	63616	61684	1932	31.48%	68.37%	36.88%	31.93
32	1988	1973	15	99.25%	0.75%	65604	63657	1947	32.49%	68.90%	36.41%	32.69
33	1988	1972	16	99.20%	0.80%	67592	65629	1963	33.50%	69.46%	35.96%	33.43
34	1988	1975	13	99.35%	0.65%	69580	67604	1976	34.51%	69.92%	35.42%	34.21
35	1988	1972	16	99.20%	0.80%	71568	69576	1992	35.51%	70.49%	34.98%	34.93
36	1988	1974	14	99.30%	0.70%	73556	71550	2006	36.52%	70.98%	34.46%	35.67
37	1988	1975	13	99.35%	0.65%	75544	73525	2019	37.53%	71.44%	33.92%	36.42
38	1988	1975	13	99.35%	0.65%	77532	75500	2032	38.54%	71.90%	33.37%	37.16
39	1988	1973	15	99.25%	0.75%	79520	77473	2047	39.54%	72.43%	32.89%	37.85
40	1988	1973	15	99.25%	0.75%	81508	79446	2062	40.55%	72.97%	32.42%	38.53
41	1988	1966	22	98.89%	1.11%	83496	81412	2084	41.55%	73.74%	32.19%	39.07
42	1988	1972	16	99.20%	0.80%	85484	83384	2100	42.56%	74.31%	31.75%	39.71
43	1988	1973	15	99.25%	0.75%	87472	85357	2115	43.57%	74.84%	31.27%	40.36
44	1988	1977	11	99.45%	0.55%	89460	87334	2126	44.58%	75.23%	30.65%	41.08
45	1988	1982	6	99.70%	0.30%	91448	89316	2132	45.59%	75.44%	29.86%	41.89
46	1988	1975	13	99.35%	0.65%	93436	91291	2145	46.60%	75.90%	29.31%	42.56
47	1988	1977	11	99.45%	0.55%	95424	93268	2156	47.60%	76.29%	28.69%	43.26
48	1988	1979	9	99.55%	0.45%	97412	95247	2165	48.61%	76.61%	28.00%	43.99
49	1988	1977	11	99.45%	0.55%	99400	97224	2176	49.62%	77.00%	27.38%	44.68
50	1988	1974	14	99.30%	0.70%	101388	99198	2190	50.63%	77.49%	26.86%	45.30
51	1988	1975	13	99.35%	0.65%	103376	101173	2203	51.64%	77.95%	26.32%	45.93
52	1988	1981	7	99.65%	0.35%	105364	103154	2210	52.65%	78.20%	25.55%	46.68
53	1988	1974	14	99.30%	0.70%	107352	105128	2224	53.66%	78.70%	25.04%	47.27
54	1988	1973	15	99.25%	0.75%	109340	107101	2239	54.66%	79.23%	24.56%	47.83
55	1988	1978	10	99.50%	0.50%	111328	109079	2249	55.67%	79.58%	23.91%	48.50
56	1988	1978	10	99.50%	0.50%	113316	111057	2259	56.68%	79.94%	23.25%	49.16
57	1988	1976	12	99.40%	0.60%	115304	113033	2271	57.69%	80.36%	22.67%	49.77
58	1988	1977	11	99.45%	0.55%	117292	115010	2282	58.70%	80.75%	22.05%	50.40
59	1988	1974	14	99.30%	0.70%	119280	116984	2296	59.71%	81.25%	21.54%	50.95
60	1988	1978	10	99.50%	0.50%	121268	118962	2306	60.72%	81.60%	20.88%	51.59
61	1988	1977	11	99.45%	0.55%	123256	120939	2317	61.73%	81.99%	20.26%	52.20
62	1988	1972	16	99.20%	0.80%	125244	122911	2333	62.73%	82.55%	19.82%	52.68
63	1988	1971	17	99.14%	0.86%	127232	124882	2350	63.74%	83.16%	19.42%	53.14
64	1988	1967	21	98.94%	1.06%	129220	126849	2371	64.74%	83.90%	19.16%	53.50
65	1988	1982	6	99.70%	0.30%	131208	128831	2377	65.76%	84.11%	18.36%	54.20
66	1988	1977	11	99.45%	0.55%	133196	130808	2388	66.77%	84.50%	17.74%	54.78
67	1988	1976	12	99.40%	0.60%	135184	132784	2400	67.77%	84.93%	17.15%	55.33
68	1988	1974	14	99.30%	0.70%	137172	134758	2414	68.78%	85.42%	16.64%	55.82
69	1988	1974	14	99.30%	0.70%	139160	136732	2428	69.79%	85.92%	16.13%	56.31
70	1988	1967	21	98.94%	1.06%	141148	138699	2449	70.79%	86.66%	15.87%	56.63
71	1988	1971	17	99.14%	0.86%	143136	140670	2466	71.80%	87.26%	15.46%	57.04

72	1988	1980	8	99.60%	0.40%	145124	142650	2474	72.81%	87.54%	14.74%	57.66
73	1988	1976	12	99.40%	0.60%	147112	144626	2486	73.82%	87.97%	14.15%	58.18
74	1988	1976	12	99.40%	0.60%	149100	146602	2498	74.83%	88.39%	13.57%	58.69
75	1988	1977	11	99.45%	0.55%	151088	148579	2509	75.84%	88.78%	12.95%	59.22
76	1988	1978	10	99.50%	0.50%	153076	150557	2519	76.84%	89.14%	12.29%	59.77
77	1988	1975	13	99.35%	0.65%	155064	152532	2532	77.85%	89.60%	11.74%	60.24
78	1988	1974	14	99.30%	0.70%	157052	154506	2546	78.86%	90.09%	11.23%	60.69
79	1988	1972	16	99.20%	0.80%	159040	156478	2562	79.87%	90.66%	10.79%	61.08
80	1988	1974	14	99.30%	0.70%	161028	158452	2576	80.87%	91.15%	10.28%	61.51
81	1988	1977	11	99.45%	0.55%	163016	160429	2587	81.88%	91.54%	9.66%	62.01
82	1988	1975	13	99.35%	0.65%	165004	162404	2600	82.89%	92.00%	9.11%	62.46
83	1988	1982	6	99.70%	0.30%	166992	164386	2606	83.90%	92.22%	8.31%	63.08
84	1988	1971	17	99.14%	0.86%	168980	166357	2623	84.91%	92.82%	7.91%	63.42
85	1988	1973	15	99.25%	0.75%	170968	168330	2638	85.92%	93.35%	7.43%	63.81
86	1988	1970	18	99.09%	0.91%	172956	170300	2656	86.92%	93.98%	7.06%	64.12
87	1988	1974	14	99.30%	0.70%	174944	172274	2670	87.93%	94.48%	6.55%	64.52
88	1988	1976	12	99.40%	0.60%	176932	174250	2682	88.94%	94.90%	5.97%	64.97
89	1988	1975	13	99.35%	0.65%	178920	176225	2695	89.95%	95.36%	5.42%	65.39
90	1988	1975	13	99.35%	0.65%	180908	178200	2708	90.95%	95.82%	4.87%	65.81
91	1988	1971	17	99.14%	0.86%	182896	180171	2725	91.96%	96.43%	4.47%	66.12
92	1988	1974	14	99.30%	0.70%	184884	182145	2739	92.97%	96.92%	3.95%	66.50
93	1988	1975	13	99.35%	0.65%	186872	184120	2752	93.98%	97.38%	3.41%	66.90
94	1988	1979	9	99.55%	0.45%	188860	186099	2761	94.99%	97.70%	2.71%	67.40
95	1988	1969	19	99.04%	0.96%	190848	188068	2780	95.99%	98.37%	2.38%	67.65
96	1988	1975	13	99.35%	0.65%	192836	190043	2793	97.00%	98.83%	1.83%	68.04
97	1988	1983	5	99.75%	0.25%	194824	192026	2798	98.01%	99.01%	1.00%	68.63
98	1988	1973	15	99.25%	0.75%	196812	193999	2813	99.02%	99.54%	0.52%	68.97
99	1937	1924	13	99.33%	0.67%	198749	195923	2826	100.00%	#######	0.00%	69.33

Validation Data Statistics

Validation	# Records	# Goods	# Bads	Fraud Rate
	166,493	164,107	2,386	1.43%

	Bin Statistics	Cumulative Statistics										
Pop Bin %	**Records**	**Goods**	**Bads**	**% Goods**	**% Bads**	**Total # Records**	**Cum Goods**	**Cum Bads**	**% Goods**	**% Bads**	**KS**	**FPR**
0	1,665	547	1,118	32.85%	67.15%	1,665	547	1,118	0.33%	46.86%	46.52%	0.49
1	1,665	1,654	11	99.34%	0.66%	3,330	2,201	1,129	1.34%	47.32%	45.98%	1.95
2	1,665	1,650	15	99.10%	0.90%	4,995	3,851	1,144	2.35%	47.95%	45.60%	3.37
3	1,665	1,647	18	98.92%	1.08%	6,660	5,498	1,162	3.35%	48.70%	45.35%	4.73
4	1,665	1,652	13	99.22%	0.78%	8,325	7,150	1,175	4.36%	49.25%	44.89%	6.09
5	1,665	1,656	9	99.46%	0.54%	9,990	8,806	1,184	5.37%	49.62%	44.26%	7.44
6	1,665	1,651	14	99.16%	0.84%	11,655	10,457	1,198	6.37%	50.21%	43.84%	8.73
7	1,665	1,651	14	99.16%	0.84%	13,320	12,108	1,212	7.38%	50.80%	43.42%	9.99
8	1,665	1,659	6	99.64%	0.36%	14,985	13,767	1,218	8.39%	51.05%	42.66%	11.30
9	1,665	1,650	15	99.10%	0.90%	16,650	15,417	1,233	9.39%	51.68%	42.28%	12.50
10	1,665	1,650	15	99.10%	0.90%	18,315	17,067	1,248	10.40%	52.31%	41.91%	13.68
11	1,665	1,655	10	99.40%	0.60%	19,980	18,722	1,258	11.41%	52.72%	41.32%	14.88
12	1,665	1,653	12	99.28%	0.72%	21,645	20,375	1,270	12.42%	53.23%	40.81%	16.04
13	1,665	1,653	12	99.28%	0.72%	23,310	22,028	1,282	13.42%	53.73%	40.31%	17.18
14	1,665	1,659	6	99.64%	0.36%	24,975	23,687	1,288	14.43%	53.98%	39.55%	18.39
15	1,665	1,655	10	99.40%	0.60%	26,640	25,342	1,298	15.44%	54.40%	38.96%	19.52
16	1,665	1,656	9	99.46%	0.54%	28,305	26,998	1,307	16.45%	54.78%	38.33%	20.66

17	1,665	1,645	20	98.80%	1.20%	29,970	28,643	1,327	17.45%	55.62%	38.16%	21.58
18	1,665	1,649	16	99.04%	0.96%	31,635	30,292	1,343	18.46%	56.29%	37.83%	22.56
19	1,665	1,653	12	99.28%	0.72%	33,300	31,945	1,355	19.47%	56.79%	37.32%	23.58
20	1,665	1,651	14	99.16%	0.84%	34,965	33,596	1,369	20.47%	57.38%	36.90%	24.54
21	1,665	1,651	14	99.16%	0.84%	36,630	35,247	1,383	21.48%	57.96%	36.49%	25.49
22	1,665	1,648	17	98.98%	1.02%	38,295	36,895	1,400	22.48%	58.68%	36.19%	26.35
23	1,665	1,647	18	98.92%	1.08%	39,960	38,542	1,418	23.49%	59.43%	35.94%	27.18
24	1,665	1,653	12	99.28%	0.72%	41,625	40,195	1,430	24.49%	59.93%	35.44%	28.11
25	1,665	1,654	11	99.34%	0.66%	43,290	41,849	1,441	25.50%	60.39%	34.89%	29.04
26	1,665	1,651	14	99.16%	0.84%	44,955	43,500	1,455	26.51%	60.98%	34.47%	29.90
27	1,665	1,640	25	98.50%	1.50%	46,620	45,140	1,480	27.51%	62.03%	34.52%	30.50
28	1,665	1,659	6	99.64%	0.36%	48,285	46,799	1,486	28.52%	62.28%	33.76%	31.49
29	1,665	1,654	11	99.34%	0.66%	49,950	48,453	1,497	29.53%	62.74%	33.22%	32.37
30	1,665	1,654	11	99.34%	0.66%	51,615	50,107	1,508	30.53%	63.20%	32.67%	33.23
31	1,665	1,655	10	99.40%	0.60%	53,280	51,762	1,518	31.54%	63.62%	32.08%	34.10
32	1,665	1,659	6	99.64%	0.36%	54,945	53,421	1,524	32.55%	63.87%	31.32%	35.05
33	1,665	1,655	10	99.40%	0.60%	56,610	55,076	1,534	33.56%	64.29%	30.73%	35.90
34	1,665	1,656	9	99.46%	0.54%	58,275	56,732	1,543	34.57%	64.67%	30.10%	36.77
35	1,665	1,648	17	98.98%	1.02%	59,940	58,380	1,560	35.57%	65.38%	29.81%	37.42
36	1,665	1,656	9	99.46%	0.54%	61,605	60,036	1,569	36.58%	65.76%	29.18%	38.26
37	1,665	1,655	10	99.40%	0.60%	63,270	61,691	1,579	37.59%	66.18%	28.59%	39.07
38	1,665	1,648	17	98.98%	1.02%	64,935	63,339	1,596	38.60%	66.89%	28.29%	39.69
39	1,665	1,656	9	99.46%	0.54%	66,600	64,995	1,605	39.61%	67.27%	27.66%	40.50

40	1,665	1,651	14	99.16%	0.84%	68,265	66,646	1,619	40.61%	67.85%	27.24%	41.16
41	1,665	1,654	11	99.34%	0.66%	69,930	68,300	1,630	41.62%	68.32%	26.70%	41.90
42	1,665	1,656	9	99.46%	0.54%	71,595	69,956	1,639	42.63%	68.69%	26.06%	42.68
43	1,665	1,652	13	99.22%	0.78%	73,260	71,608	1,652	43.63%	69.24%	25.60%	43.35
44	1,665	1,657	8	99.52%	0.48%	74,925	73,265	1,660	44.64%	69.57%	24.93%	44.14
45	1,665	1,660	5	99.70%	0.30%	76,590	74,925	1,665	45.66%	69.78%	24.13%	45.00
46	1,665	1,652	13	99.22%	0.78%	78,255	76,577	1,678	46.66%	70.33%	23.66%	45.64
47	1,665	1,652	13	99.22%	0.78%	79,920	78,229	1,691	47.67%	70.87%	23.20%	46.26
48	1,665	1,650	15	99.10%	0.90%	81,585	79,879	1,706	48.67%	71.50%	22.83%	46.82
49	1,665	1,653	12	99.28%	0.72%	83,250	81,532	1,718	49.68%	72.00%	22.32%	47.46
50	1,665	1,655	10	99.40%	0.60%	84,915	83,187	1,728	50.69%	72.42%	21.73%	48.14
51	1,665	1,652	13	99.22%	0.78%	86,580	84,839	1,741	51.70%	72.97%	21.27%	48.73
52	1,665	1,652	13	99.22%	0.78%	88,245	86,491	1,754	52.70%	73.51%	20.81%	49.31
53	1,665	1,655	10	99.40%	0.60%	89,910	88,146	1,764	53.71%	73.93%	20.22%	49.97
54	1,665	1,656	9	99.46%	0.54%	91,575	89,802	1,773	54.72%	74.31%	19.59%	50.65
55	1,665	1,651	14	99.16%	0.84%	93,240	91,453	1,787	55.73%	74.90%	19.17%	51.18
56	1,665	1,655	10	99.40%	0.60%	94,905	93,108	1,797	56.74%	75.31%	18.58%	51.81
57	1,665	1,649	16	99.04%	0.96%	96,570	94,757	1,813	57.74%	75.98%	18.24%	52.27
58	1,665	1,648	17	98.98%	1.02%	98,235	96,405	1,830	58.75%	76.70%	17.95%	52.68
59	1,665	1,650	15	99.10%	0.90%	99,900	98,055	1,845	59.75%	77.33%	17.58%	53.15
60	1,665	1,650	15	99.10%	0.90%	101,565	99,705	1,860	60.76%	77.95%	17.20%	53.60
61	1,665	1,652	13	99.22%	0.78%	103,230	101,357	1,873	61.76%	78.50%	16.74%	54.11
62	1,665	1,654	11	99.34%	0.66%	104,895	103,011	1,884	62.77%	78.96%	16.19%	54.68

63	1,665	1,656	9	99.46%	0.54%	106,560	104,667	1,893	63.78%	79.34%	15.56%	55.29
64	1,665	1,658	7	99.58%	0.42%	108,225	106,325	1,900	64.79%	79.63%	14.84%	55.96
65	1,665	1,648	17	98.98%	1.02%	109,890	107,973	1,917	65.79%	80.34%	14.55%	56.32
66	1,665	1,653	12	99.28%	0.72%	111,555	109,626	1,929	66.80%	80.85%	14.05%	56.83
67	1,665	1,655	10	99.40%	0.60%	113,220	111,281	1,939	67.81%	81.27%	13.46%	57.39
68	1,665	1,649	16	99.04%	0.96%	114,885	112,930	1,955	68.81%	81.94%	13.12%	57.76
69	1,665	1,647	18	98.92%	1.08%	116,550	114,577	1,973	69.82%	82.69%	12.87%	58.07
70	1,665	1,644	21	98.74%	1.26%	118,215	116,221	1,994	70.82%	83.57%	12.75%	58.29
71	1,665	1,654	11	99.34%	0.66%	119,880	117,875	2,005	71.83%	84.03%	12.20%	58.79
72	1,665	1,651	14	99.16%	0.84%	121,545	119,526	2,019	72.83%	84.62%	11.78%	59.20
73	1,665	1,644	21	98.74%	1.26%	123,210	121,170	2,040	73.84%	85.50%	11.66%	59.40
74	1,665	1,658	7	99.58%	0.42%	124,875	122,828	2,047	74.85%	85.79%	10.95%	60.00
75	1,665	1,644	21	98.74%	1.26%	126,540	124,472	2,068	75.85%	86.67%	10.82%	60.19
76	1,665	1,653	12	99.28%	0.72%	128,205	126,125	2,080	76.86%	87.18%	10.32%	60.64
77	1,665	1,652	13	99.22%	0.78%	129,870	127,777	2,093	77.86%	87.72%	9.86%	61.05
78	1,665	1,658	7	99.58%	0.42%	131,535	129,435	2,100	78.87%	88.01%	9.14%	61.64
79	1,665	1,649	16	99.04%	0.96%	133,200	131,084	2,116	79.88%	88.68%	8.81%	61.95
80	1,665	1,651	14	99.16%	0.84%	134,865	132,735	2,130	80.88%	89.27%	8.39%	62.32
81	1,665	1,654	11	99.34%	0.66%	136,530	134,389	2,141	81.89%	89.73%	7.84%	62.77
82	1,665	1,653	12	99.28%	0.72%	138,195	136,042	2,153	82.90%	90.23%	7.34%	63.19
83	1,665	1,656	9	99.46%	0.54%	139,860	137,698	2,162	83.91%	90.61%	6.70%	63.69
84	1,665	1,652	13	99.22%	0.78%	141,525	139,350	2,175	84.91%	91.16%	6.24%	64.07
85	1,665	1,654	11	99.34%	0.66%	143,190	141,004	2,186	85.92%	91.62%	5.70%	64.50

86	1,665	1,650	15	99.10%	0.90%	144,855	142,654	2,201	86.93%	92.25%	5.32%	64.81
87	1,665	1,657	8	99.52%	0.48%	146,520	144,311	2,209	87.94%	92.58%	4.64%	65.33
88	1,665	1,651	14	99.16%	0.84%	148,185	145,962	2,223	88.94%	93.17%	4.23%	65.66
89	1,665	1,648	17	98.98%	1.02%	149,850	147,610	2,240	89.95%	93.88%	3.93%	65.90
90	1,665	1,655	10	99.40%	0.60%	151,515	149,265	2,250	90.96%	94.30%	3.34%	66.34
91	1,665	1,656	9	99.46%	0.54%	153,180	150,921	2,259	91.96%	94.68%	2.71%	66.81
92	1,665	1,655	10	99.40%	0.60%	154,845	152,576	2,269	92.97%	95.10%	2.12%	67.24
93	1,665	1,648	17	98.98%	1.02%	156,510	154,224	2,286	93.98%	95.81%	1.83%	67.46
94	1,665	1,652	13	99.22%	0.78%	158,175	155,876	2,299	94.98%	96.35%	1.37%	67.80
95	1,665	1,647	18	98.92%	1.08%	159,840	157,523	2,317	95.99%	97.11%	1.12%	67.99
96	1,665	1,644	21	98.74%	1.26%	161,505	159,167	2,338	96.99%	97.99%	1.00%	68.08
97	1,665	1,649	16	99.04%	0.96%	163,170	160,816	2,354	97.99%	98.66%	0.66%	68.32
98	1,665	1,650	15	99.10%	0.90%	164,835	162,466	2,369	99.00%	99.29%	0.29%	68.58
99	1,658	1,641	17	98.97%	1.03%	166,493	164,107	2,386	100.00%	100.00%	0.00%	68.78