CREDIT CARD APPLICATION FRAUD DETECTION

Ge Zeng, Xiaoying (Shawn) Zou, Xinyue Yu, Zihan (Johnson) Ling, Zihang (Charlie) Li

March 24th, 2022

Content

1. Executive Summary

This project focuses on helping banks identify credit card application frauds. After utilizing the personal identifying information, we designed several supervised machine learning models and selected the best-performing model to help the bank identify the risk, detect the fraud, and mitigate fraudulent activity in real time.

The original dataset contains personal identifying information to identify fraudulent actions. There are 8 columns and a fraud label, with 1,000,000 records. We started by exploring the data and writing the data quality report to provide a basic characteristic of the data. With a general idea of our dataset, we created 1,063 new variables through feature engineering which contains 1,035 numerical features, and selected 30 important features as our final features through the feature selection process. By reserving the most recent two months of data as the out of time sample and randomly splitting the training and testing data in the proportion of 3:1, we constructed logistic regression, single decision tree, random forest, boosted tree, neural network, and adaptive boosting models using these 30 features, experimented with different hyperparameters, and found out the best performer, a boosted tree model. Finally, we evaluated its performance on the Training, Testing, and OOT dataset. Our boosted tree model detected 56.27% of fraud in the Training data, 54.53% of fraud in the Testing data, and 53.6% of fraud in the OOT data by looking at the top 3% of the corresponding dataset.

2. Description of Data

2.1 Summary statistics table

Table 1. *Numeric Fields Summary*

Field Name	% Populated	Min	Max	Mean	Stdev	*% Zero
date	100	01-01-2016	12-31-2016	-	-	0
dob	100	19000101	20161031	-	-	0

**% Zero: only including record whose value is 0.*

Table 2. *Categorical Fields Summary*

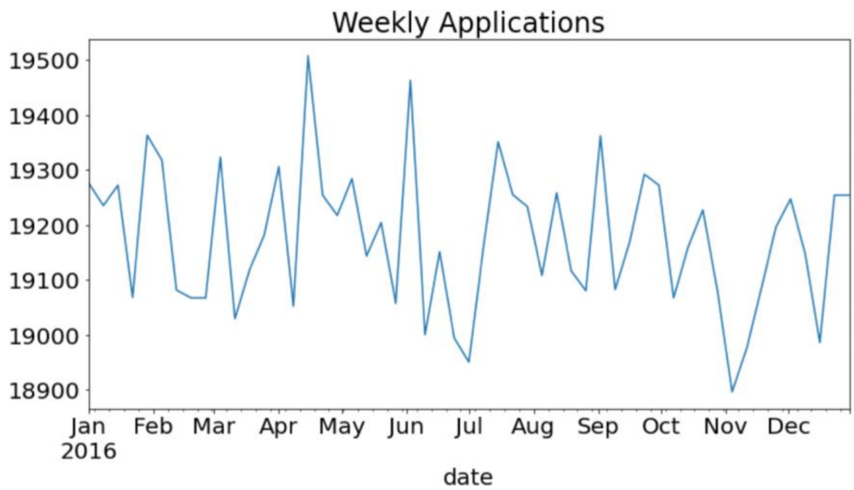
Field Name	% Populated	*Unique Values	Most Common Value
record	100	1,000,000	-
ssn	100	835,819	999999999
firstname	100	78,136	EAMSTRMT
lastname	100	177,001	ERJSAXA
address	100	828,774	123 MAIN ST
zip5	100	26,370	68138
homephone	100	28,244	9999999999
fraud_label	100	2	0

**Unique Values: does not include Nan.*

2.2 Field Distribution

2.2.1 date

Figure 1. *Number of Weekly Applications. (Set daily count of 02-26 as that of 02-19 and daily count of 12-30 as that of 12-23)*



2.2.2 fraud_label

Figure 2. *Distribution of fraud_label. (fraud_laebl_0 : fraud_laebl_1 = 985,607 : 14,393)*

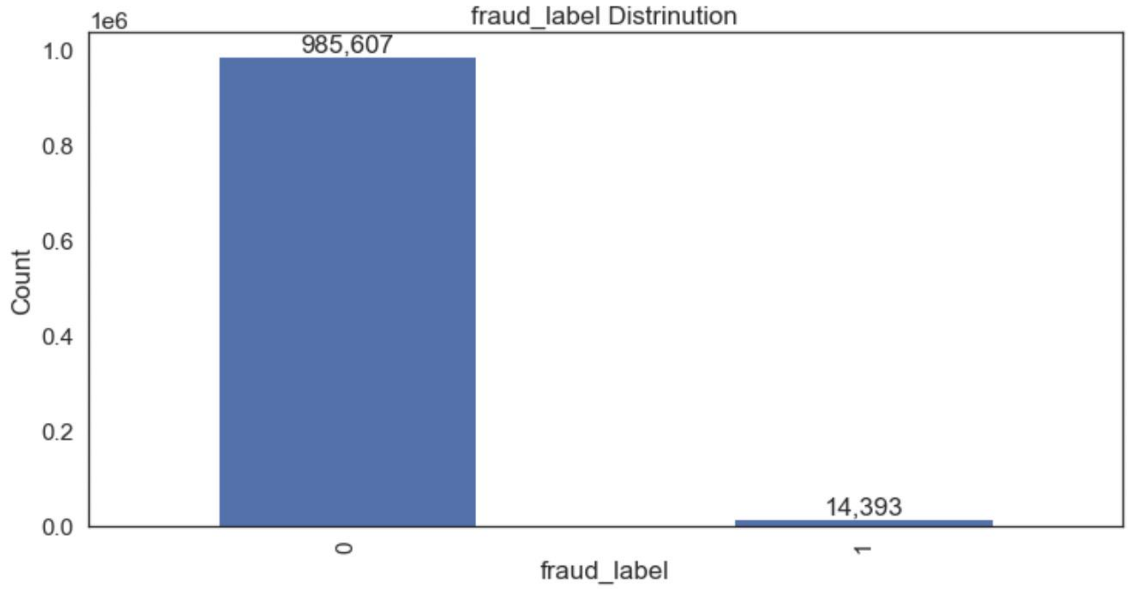
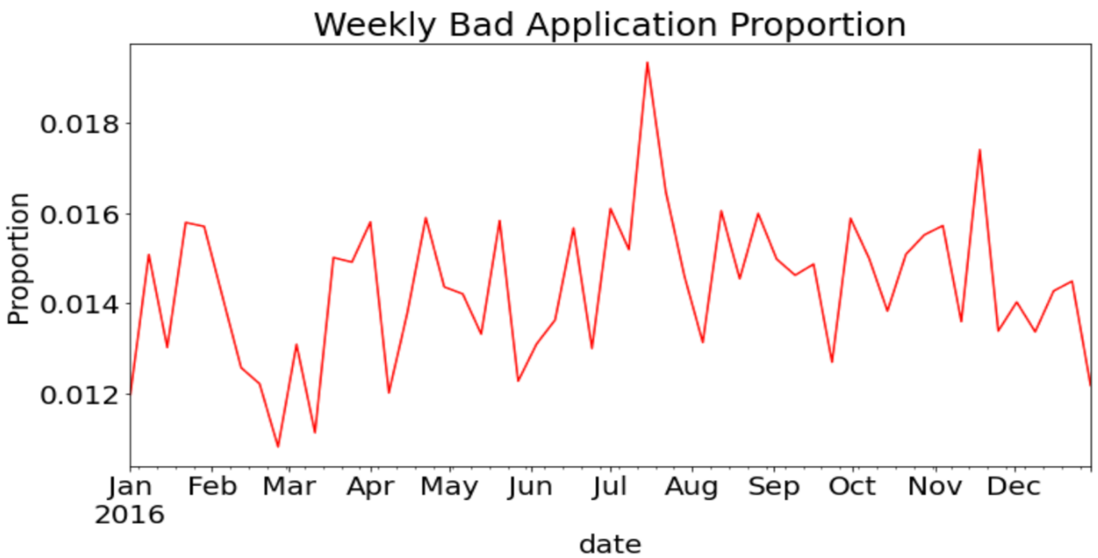


Figure 3. *Proportion Distribution of Weekly Bad Application*

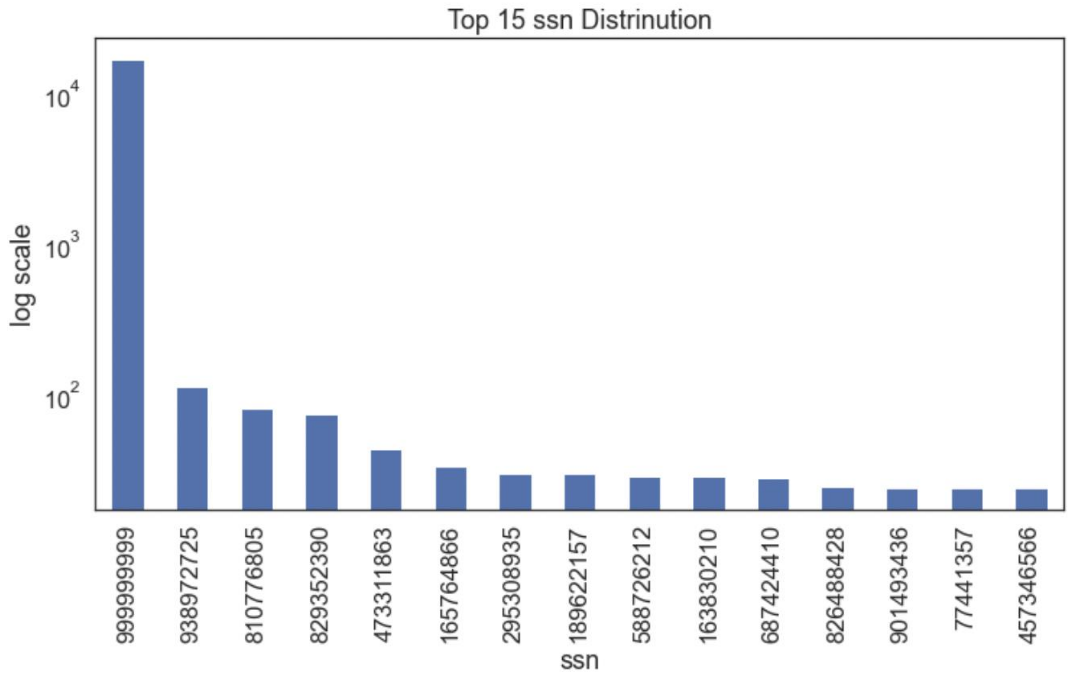
Bad (red): fraud_label=1

Proportion: (weekly count of bad applications) / (weekly count of total applications)



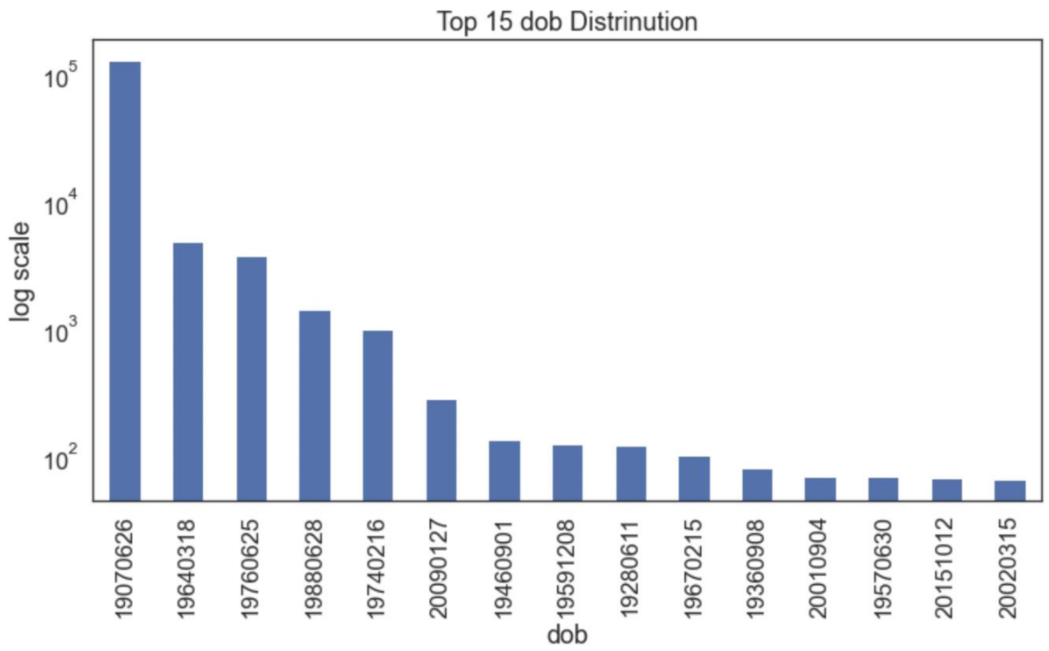
2.2.3 ssn

Figure 4. *Distribution of ssn. The value of y-axis is log(count).*



2.2.4 dob

Figure 5. *Distribution of dob. The value of y-axis is log(count).*



3. Data Cleaning

3.1 Filling in the incomplete digits

We assume that ‘ssn’ field is 9 digits, ‘zip5’ is 5 digits, ‘homephone’ is 10 digits, and ‘dob’ is 8 digits. So we filled in 0 to complete the number of the digits in these fields if the data less than our predefined digits.

3.2 Fixing frivolous values

3.2.1 ssn

After investigation, we found that the data “999999999” in ssn field is not real. As we don’t want them to be linked with others, we replaced the ssn which is “999999999” with the negative record number. This will ensure that these new values won’t link to any other record since the values are unique for each record and won’t accidentally overlap with a real value.

3.2.2 address

The data “123 MAIN ST” in the address field is not real. We added the record number after the address which is “123 MAIN ST” to ensure that these new values won’t link to any other record.

3.2.3 dob

The data “19070626” in dob field is frivolous. As we don’t want them to be linked with others, we replaced the dob which is “19070626” with the negative record number. This will ensure that these new values won’t link to any other record since the values are unique for each record and won’t accidentally overlap with a real value.

3.2.4 homephone

The data “9999999999” in dob field is frivolous. We found that they were automatically filling in the empty field values. As we don’t want them to be linked with others, we replaced the dob which is “9999999999” with the negative record number. This will ensure that these new values won’t link to any other record since the values are unique for each record and won’t accidentally overlap with a real value.

4. Candidate Variables

We created 1035 new numerical variables in total. Below is our process to build new variables:

4.1 Creating new entities by using existing entities.

We created 24 new entities and built variables associated with them to increase efficacy in fraud detection.

Table 3. *24 New Entities Created by Using Existing Entities*

New Entity Name	Formulating Process	Example
name	firstname + lastname	JAMESSMITH
fulladdress	address + zip5	324 MAINST54321
name_dob	name + dob	JAMESSMITH19800101
name_fulladdress	name + fulladdress	JAMESSMITH324 MAIN ST54321
name_homephone	name + homephone	JAMESSMITH8886661234
fulladdress_dob	fulladdress + dob	324 MAIN ST5432119800101
fulladdress_homephone	fulladdress + homephone	324 MAIN ST543218886661234
dob_homephone	dob + homephone	198001018886661234
name_homephone_dob	name + homephone + dob	JAMESSMITH888666123419800101
name_fulladdress_dob	name + fulladdress + dob	JAMESSMITH324 MAIN ST5432119800101
ssn_firstname	ssn + firstname	999888777JAMES
ssn_lastname	ssn + lastname	999888777SMITH
ssn_address	ssn + address	999888777324 MAIN ST
ssn_zip5	ssn + zip5	99988877754321
ssn_dob	ssn + dob	99988877719800101
ssn_homephone	ssn + homephone	9998887778886661234
ssn_name	ssn + name	999888777JAMESSMITH
ssn_fulladdress	ssn + fulladdress	999888777324 MAIN ST54321
ssn_name_dob	ssn + name + dob	999888777JAMESSMITH19800101
ssn_name_fulladdress	ssn + name + fulladdress	999888777JAMESSMITH324 MAIN ST54321
ssn_name_homephone	ssn + name + homephone	999888777JAMESSMITH8886661234
ssn_fulladdress_dob	ssn + fulladdress + dob	999888777324 MAIN ST5432119800101
ssn_fulladdress_homephone	ssn + fulladdress + homephone	999888777324 MAIN ST543218886661234
ssn_dob_homephone	ssn + dob + homephone	999888777198001018886661234

4.2 Combining the old entities and new entities.

We formulated two attribute lists of our interest and we built new variables based on these two lists.

Table 4. *Two Attributed Lists Used to Create New Features*

	Length	Attributes Included
List 1	27	'ssn', 'dob', 'homephone', 'name', 'fulladdress', 'name_dob', 'name_fulladdress', 'name_homephone', 'fulladdress_dob', 'fulladdress_homephone', 'dob_homephone', 'name_homephone_dob', 'name_fulladdress_dob', 'ssn_firstname', 'ssn_lastname', 'ssn_address', 'ssn_zip5', 'ssn_dob', 'ssn_homephone', 'ssn_name', 'ssn_fulladdress', 'ssn_name_dob', 'ssn_name_fulladdress', 'ssn_name_homephone', 'ssn_fulladdress_dob', 'ssn_fulladdress_homephone', 'ssn_dob_homephone
List 2	10	'ssn', 'name_dob', 'name_fulladdress', 'name_homephone', 'fulladdress_dob', 'dob_homephone', 'ssn_dob', 'ssn_homephone', 'ssn_name', 'ssn_fulladdress'

4.3 Building new variables.

We built 5 different types of variables using both list 1 and list 2. For a full table of 1035 new variables built, please refer to the appendix.

4.3.1 Days Since

- Attribute list used: List 1
- Formula:

$$d_{since} = d_{new} - d_{old}$$

- Logic: Difference in the number of days since we last saw this entity. This type of variable is useful because we can find out the interval between the appearance of the same entity and easily identify abnormal applications if an entity shows up too often.
- Example: For the name JAMESSMITH, if we see this on application date 07/01/2016 and it appears again on 07/05/2016, then Days Since for row with ‘07/05/2016’ equals 4.
- Number of new variables created: 27

4.3.2 Velocity

- Attribute list used: List 1
- Formula:

$$velocity = \frac{\text{count of entity E in the past n days}}{n} \text{ for } n \in \{0, 1, 3, 7, 14, 30\}$$

- Logic: The number of times that we see the same entity appear for a given time period (from 0 to 30 days). It measures the frequency of certain entity being used. So an application might be considered fraudulent if a single entity appears too often in past applications.

- Example: For the name JAMESSMITH, if we see this name 6 times in the past 30 days, then Velocity is 6/30 = 0.2.
- Number of new variables created: 162

4.3.3 Relative Velocity

- Attribute list used: List 1
- Formula:

$$\text{relative velocity}_{(x,y)} = \frac{\text{count of entity E in the past x days}}{\text{count of entity E in the past y days}} \text{ for } x \in \{0, 1\}, \text{ for } y \in \{3, 7, 14, 30\}$$

- Logic: The number of times that we see this entity appears in the last (0, 1) day divided by the number of times we see this entity appears in the last (3,7,14,30) days. This type of variables measures whether we see a surge in application with the same entity in a relatively short period of time.
- Example: For the name JAMESSMITH, if we see this entity 6 times in the past 1 day and 7 times in the past 30 days then relative velocity is 0.83.
- Number of new variables created: 216

4.3.4 Unique Count

- Attribute list used: List 2
- Logic: The unique number of entity B that we see in the past (0,1,3,7,14,30) days for a particular entity A. This will help us to identify whether some applicants have been manipulating entities partially during a given timeframe.
- Example: For the ssn 999888777, if in the past 30 days we see this ssn appears in 5 applications with 3 unique name_dob entities (‘JAMESSMITH19800101’, ‘SMITHJAMES19800101’, ‘JAMESMITH19300201’) then the unique count is 3.
- Number of new variables created: 540

4.3.5 Days Since for entity in a particular field

- Attribute list used: List 2
- Logic: The number of days that we see entity B repeatedly for a particular entity A. This variable helps us to check the abnormality of the application timeline for related entities a and b when they both appear on applications.
- Example: For the ssn 999888777, if we saw the same name_dob entity ‘JAMESSMITH19800101’ on 06/01/2016 and 06/06/2016 then new variable for this 06/06/2016’s application entry is 5.
- Number of new variables created: 90

5. Feature Selection Process

In the feature selection process, we firstly applied KS as a filter to select the top 100 related variables. In addition, we used a wrapper with Random Forest Classifier to select the top 30 variables. The table below is the top 30 variables ordered by multivariate importance.

Figure 6. *Feature Selection Workflow*

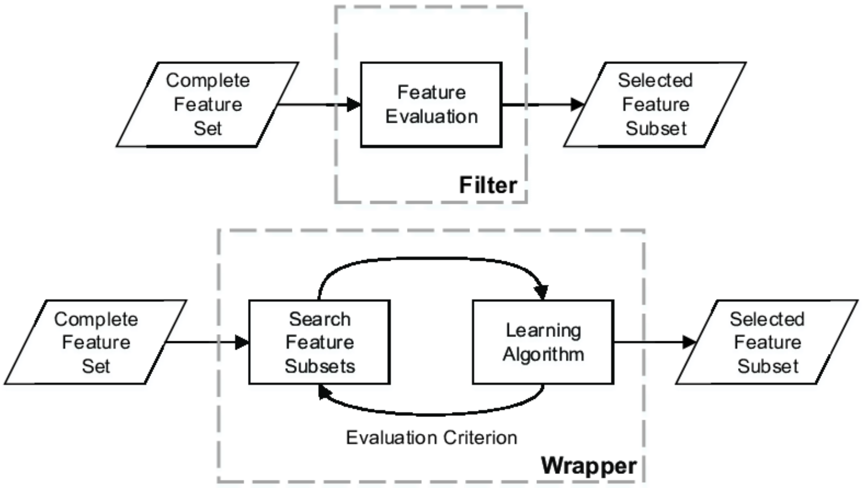


Table 5. *Top 30 Variables and Variable Descriptions*

No.	Variable	KS Score	Descriptions
1	fulladdress_count_30	0.3320	Number of times that fulladdress appeared in the past 30 days
2	fulladdress_homephone_count_30	0.2290	Number of times that fulladdress_homephone appeared in the past 30 days
3	ssn_dob_count_30	0.2285	Number of times that ssn_dob appeared in the past 30 days
4	name_dob_count_30	0.2276	Number of times that name_dob appeared in the past 30 days
5	ssn_count_30	0.2270	Number of times that ssn appeared in the past 30 days
6	ssn_name_dob_count_30	0.2262	Relative velocity of ssn_name_dob over the past 30 days
7	ssn_lastname_count_30	0.2260	Number of times that ssn_lastname appeared in the past 30 days
8	name_dob_count_14	0.2153	Number of times that name_dob appeared in the past 14 days
9	ssn_dob_count_14	0.2149	Number of times that ssn_dob appeared in the past 14 days

No.	Variable	KS Score	Descriptions
10	ssn_count_14	0.2144	Number of times that ssn number appeared in the past 14 days
11	ssn_firstname_count_14	0.2138	Number of times that ssn_firstname appeared in the past 14 days
12	ssn_name_dob_count_14	0.2135	Number of times that ssn_name_dob appeared in the past 14 days
13	ssn_lastname_count_14	0.2134	Number of times that ssn_lastname appeared in the past 14 days
14	ssn_dob_count_0_by_30	0.2077	Relative velocity of ssn_dob over the past 30 days
15	name_dob_count_0_by_30	0.2070	Relative velocity of name_dob over the past 30 days
16	ssn_count_0_by_30	0.2063	Relative velocity of ssn over the past 30 days
17	ssn_name_dob_count_0_by_30	0.2055	Relative velocity of ssn_name_dob over the past 30 days
18	fulladdress_homephone_count_7	0.1998	Number of times that fulladdress_homephone appeared in the past 7 days
19	homephone_count_3	0.1949	Number of times that homephone number appeared in the past 3 days
20	name_dob_count_0_by_14	0.1948	Relative velocity of name_dob over the past 14 days
21	ssn_dob_count_0_by_14	0.1942	Relative velocity of ssn_dob over the past 14 days
22	name_dob_count_7	0.1941	Number of times that name_dob appeared in the past 7 days
23	ssn_count_0_by_14	0.1938	Relative velocity of ssn over the past 14 days
24	ssn_firstname_count_0_by_14	0.1932	Relative velocity of ssn_firstname over the past 14 days
25	ssn_dob_count_7	0.1931	Number of times that ssn_dob appeared in the past 7 days
26	ssn_count_7	0.1930	Number of times that ssn appeared in the past 7 days
27	ssn_name_dob_count_0_by_14	0.1929	Relative velocity of ssn_name_dob over the past 14 days
28	ssn_lastname_count_0_by_14	0.1928	Relative velocity of ssn_lastname over the past 14 days
29	ssn_firstname_count_7	0.1927	Number of times that ssn_firstname appeared in the past 7 days
30	ssn_name_count_0_by_14	0.1924	Relative velocity of ssn_name over the past 14 days

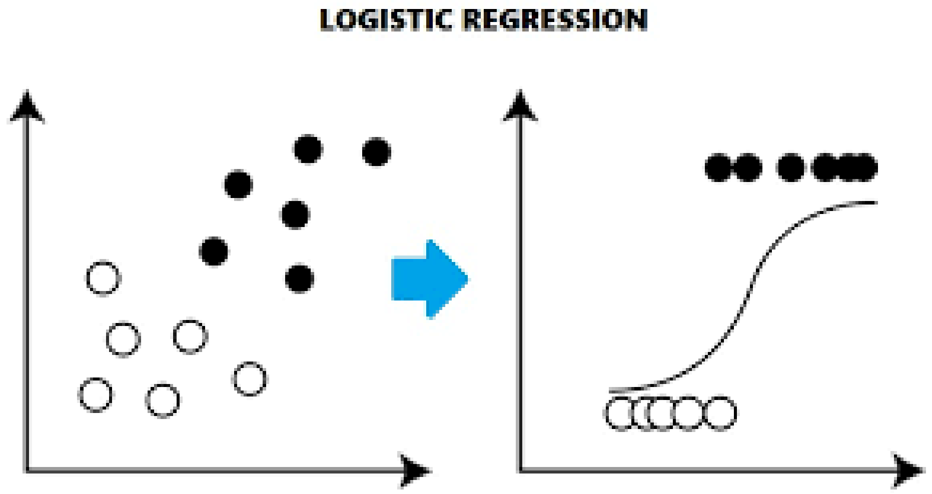
6. Model Algorithms

To select the most effective model, we compared results among 7 different models including logistic regression, single decision tree, random forest, boosted tree, neural network, and adaptive boosting models. We also adjusted critical hyperparameters in different models to achieve the best performance of each model. We used records before 11/01/2016 as our training and testing dataset and randomly split the training and testing data in the proportion of 3:1. Then we used records of the last two months (11/01/2016-12/31/2016) as out-of-time (OOT) data. To measure the results, we used 3% of the population in the training dataset, testing dataset, and OOT dataset respectively.

6.1 Logistic Regression

The logistic regression model is used to predict the class of individuals based on one or multiple predictor variables. It is a supervised algorithm that learns a linear relationship from the given dataset and then introduces non-linearity through the Sigmoid function. In our case, the model is used to model the binary outcome to predict whether a record is a fraud or not.

Figure 7. *Illustration of Logistic Regression*



Hyperparameters:

- penalty: It specifies the norm of the penalty. It has three values, ‘l1’, ‘l2’, ‘elasticnet’ and ‘none’, and the default value of penalty is ‘l2’.
- solver: It indicates the algorithm to use in the optimization problem. It has five values, ‘newton-cg’, ‘lbfgs’, ‘liblinear’, ‘sag’ and ‘saga’. The default value is ‘lbfgs’.

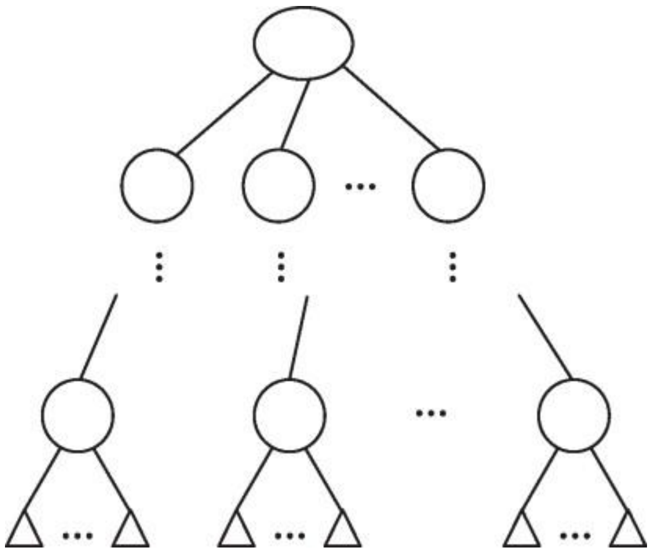
Table 6. ***Hyperparameter Tuning of Logistic Regression***

Model		Parameters			Avg FDR at 3%		
	Iteration	**Variables**	**penalty**	**solver**	**trn**	**tst**	**oot**
Logistic Regression	1	20	none	lbfgs	0.535	0.544	0.518
	2	20	l2	lbfgs	0.539	0.533	0.519
	3	25	none	lbfgs	0.539	0.544	0.522
	4	25	l2	lbfgs	0.536	0.548	0.520
	5	30	none	lbfgs	0.545	0.541	0.525
	6	30	l2	lbfgs	0.541	0.533	0.519
	7	20	l1	saga	0.538	0.543	0.520
	8	20	l2	saga	0.539	0.540	0.519
	9	25	l1	saga	0.538	0.538	0.519
	10	25	l2	saga	0.540	0.537	0.520
	11	30	none	saga	0.539	0.536	0.520
	12	30	l2	saga	0.541	0.536	0.520

6.2 Single Decision Tree

The decision tree model is a supervised machine learning algorithm that can be used for classification and regression. A decision tree is a flowchart resembling a tree structure where each internal node denotes a sub-classifier on an attribute, each branch represents an outcome of the classification, and each leaf node (terminal node) holds a class label. In our case, we use the decision tree to classify a record as fraud or not a fraud

Figure 8. *Illustration of Single Decision Tree*



Hyperparameters:

- max_depth: It represents the maximum depth of the tree. If its value is None, nodes are expanded until all leaves are pure, or until all leaves contain less than min_samples_split samples. The max_depth input should be integer and the default value of max_depth is None.
- splitter: It denotes the strategy to choose the split at each node. It has two values, 'best' and 'random', which means choosing the best split and choosing the best random split respectively. The default value of the splitter is 'best'.

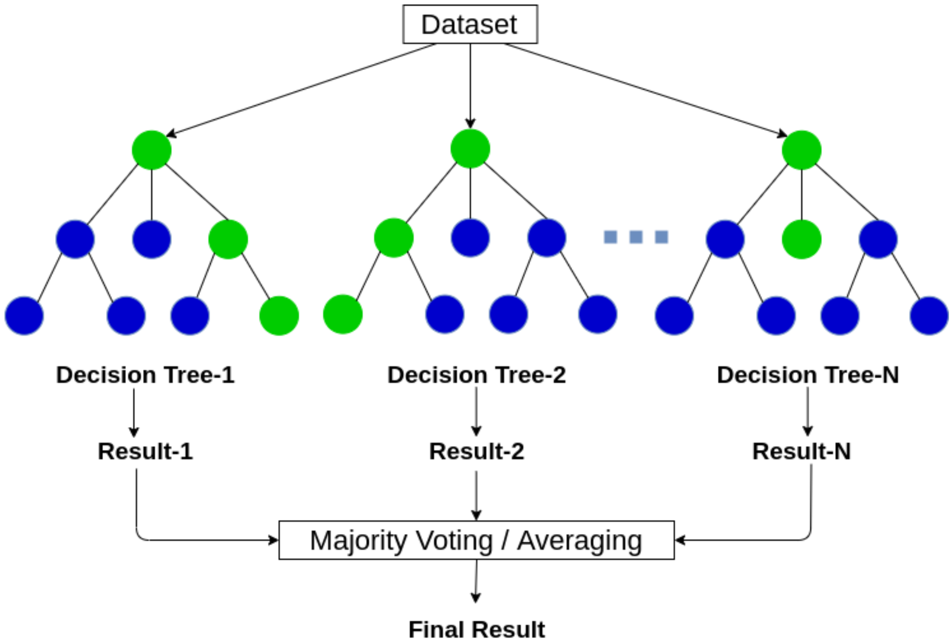
Table 7. ***Hyperparameter Tuning of Single Decision Tree***

Model		Parameters			Avg FDR at 3%		
	Iteration	Variables	max_depth	splitter	trn	tst	oot
Single Decision Tree	1	20	None	random	0.560	0.543	0.530
	2	20	None	best	0.560	0.544	0.529
	3	25	20	random	0.560	0.548	0.531
	4	25	20	best	0.558	0.551	0.528
	5	30	30	random	0.557	0.552	0.528
	6	30	30	best	0.559	0.548	0.530

6.3 Random Forest

The random forest is a classification algorithm consisting of many relatively strong and deep decisions trees. It combines ensemble techniques and training randomness when building each individual tree to create an uncorrelated forest of trees which predicts by averaging or majority voting and outperforms predictions of any individual tree in the forest. In our case, we use it to classify a record as fraud or not a fraud.

Figure 9. *Illustration of Single Decision Tree*



Hyperparameters:

- n_estimators: It denotes the number of trees in the forest. The input should be an integer and the default value of n_estimator is 100.
- max_depth: It represents the maximum depth of the tree. If the value is None, nodes are expanded until all leaves are pure or until all leaves contain less than min_samples_split samples. The input of max_depth should be an integer and the default value of max_depth is 100.
- min_samples_leaf: It represents the minimum number of samples required at a leaf node. The default value of max_depth is 1.
- min_samples_split: It represents the minimum number of samples required to split an internal node. The default value of max_depth is 2.

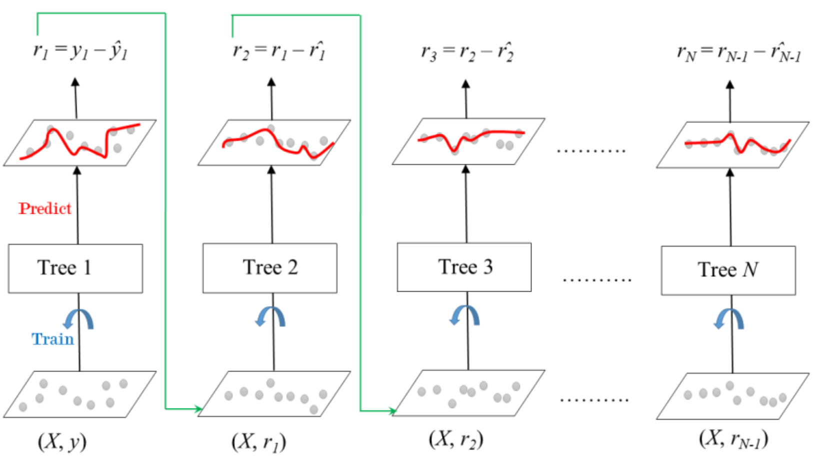
Table 8. ***Hyperparameter Tuning of Random Forest***

Model		**Parameters**					**Avg FDR at 3%**		
	Iteration	**Variables**	**n_estimators**	**max_depth**	**min_sample_leaf**	**min_samples_split**	**trn**	**tst**	**oot**
Random Forest	1	20	10	10	1	2	0.552	0.567	0.534
	2	25	50	20	1	300	0.560	0.550	0.535
	3	30	100	20	30	300	0.552	0.568	0.537
	4	30	150	30	30	500	0.555	0.549	0.531

6.4 Boosted Tree

The boosted tree model is a supervised learning classification and regression model consisting of many relatively weak and shallow trees. These trees are built sequentially to train on the residual errors of the current sum, each adding more correction. Boosting is a way to train a series of weak models to form a strong model. A gradient boosted tree model is built in a stage-wise fashion and allows optimization of an arbitrary differentiable loss function.

Figure 10. *Illustration of Boosted Tree*



Hyperparameters:

- n_estimators: It denotes the number of trees in the forest to fit. The default value is 100.
- max_depth: It denotes the maximum depth of the tree. If the value is None, nodes are expanded until all leaves are pure or until all leaves contain less than min_samples_split samples. When max_depth is less than or equal to 0, it means that the depth has no limit. The default value is -1.
- learning_rate: It represents boosting learning rate. To prevent overfitting the dataset, we can use a smaller learning rate to prevent overfitting and improve model performance. A learning rate in the range of 0.1 to 0.3 usually gives good results. The default learning rate is 0.1.

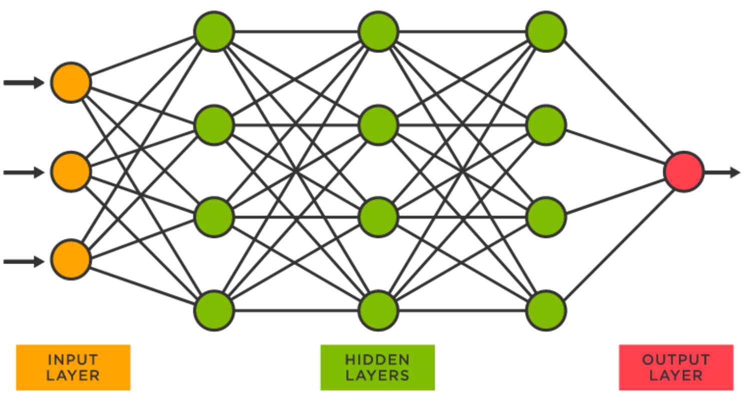
Table 9. ***Hyperparameter Tuning of Boosted Tree***

Model		**Parameters**				**Avg FDR at 3%**		
	Iteration	**Variables**	**n_estimators**	**max_ depth**	**learning_ rate**	**trn**	**tst**	**oot**
Random Forest	1	20	100	3	0.1	0.559	0.555	0.536
	2	20	200	3	0.01	0.557	0.552	0.535
	3	25	300	3	0.1	0.557	0.559	0.537
	4	25	500	5	0.01	0.556	0.557	0.537
	5	30	500	5	0.1	0.557	0.559	0.536
	6	30	1000	5	0.01	0.559	0.555	0.536

6.5 Neural Network

Neural network algorithm is inspired by the biological neural networks in brains. It consists of an input layer, some hidden layers, and an output layer with nodes resembling the neurons in the brain. Each node transmits signals to nodes in the next layer and the next nodes process signals and decides whether to release signals depending on whether the aggregate level reaches the threshold. Each node receives weighted signals from nodes in previous layers and performs a transformation based on the linear combination of signals received. The algorithm adjusts weights through backpropagation and the records are passed through many times until weights reach local optimum.

Figure 11. *Illustration of Neural Network*



Hyperparameters:

- hidden_layer_sizes: It denotes the ith element represents the number of neurons in the ith hidden layer. The default value is (100,), which means that the model has 1 hidden layer with 100 hidden neurons.
- n_layers_: It represents the number of layers in the neural network model.
- learning_rate: It schedules weight updates with three values, 'constant', 'invscaling' and 'adaptive'. The default value is 'constant'.
- activation: It indicates the activation function for the hidden layer. It has four values, 'identity', 'logistic', 'tanh' and 'relu'. The default value is 'relu'.

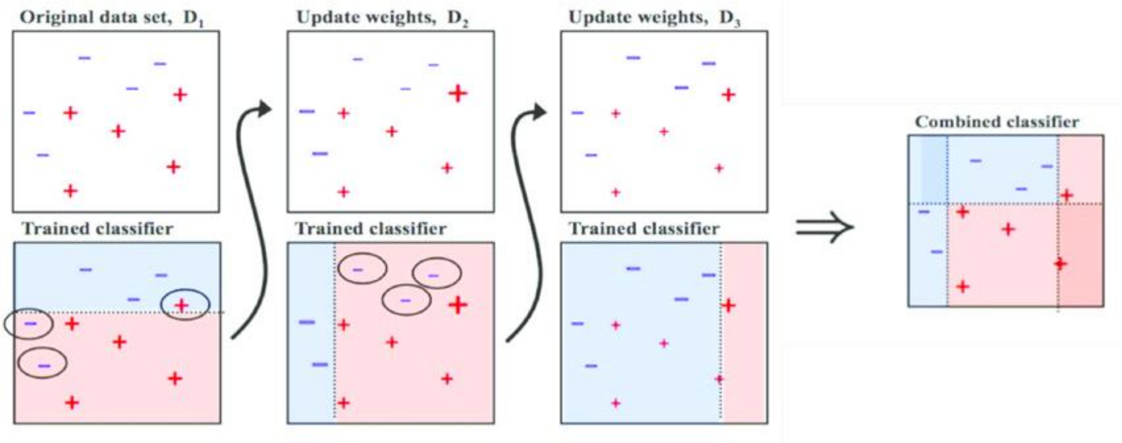
Table 10. *Hyperparameter Tuning of Neural Network*

Model		Parameters					Avg FDR at 3%		
	Iteration	Variables	hidden_layer_sizes	n_layers_	learning_rate	activation	trn	tst	oot
Neural Network	1	20	5	1	Constant	Relu	0.550	0.55	0.527
	2	20	5	1	Adaptive	Logistic	0.551	0.543	0.528
	3	25	10	1	Constant	Relu	0.553	0.562	0.535
	4	25	10	2	Adaptive	Logistic	0.554	0.548	0.531
	5	30	20	2	Constant	Relu	0.559	0.549	0.533
	6	30	20	2	adaptive	Logistic	0.552	0.551	0.532

6.6 Adaboost

Adaptive boosting is a boosted algorithm that increases the weight on misclassified records so that the training would focus more on these misclassified records in the next iteration. It is an iterative approach to learn from past mistakes of weak learners and gradually converge to a stronger learner. In each stage, adaptive boosting learns the relative 'hardness' of classifying each sample so that later trees pay more attention to harder examples.

Figure 12. *Illustration of Neural Network*



Hyperparameters:

- n_estimators: It denotes the maximum number of estimators at which boosting is terminated. The default value is 50.
- learning_rate: It represents the weight applied to each classifier at each boosting iteration. A higher learning rate increases the contribution of each classifier. There is trade-off between n_estimators and learning rate. The higher the learning rate, the smaller the estimators when the algorithm stops. The default value is 1.
- algorithm: It has 2 values, 'SAMME' and 'SAMME.R'. 'SAMME' uses real boosting algorithm while 'SAMME.R' uses discrete boosting algorithm. 'SAMME.R' usually converges faster with fewer iterations.

Table 11. ***Hyperparameter Tuning of Adaboost***

Model		**Parameters**				**Avg FDR at 3%**		
	Iteration	**Variables**	**n_ estimators**	**learning_ rate**	**algorithm**	**trn**	**tst**	**oot**
Adaboost	1	20	50	1	SAMMER.R	0.550	0.560	0.531
	2	20	50	0.1	SAMMER	0.540	0.529	0.518
	3	25	50	1	SAMMER	0.531	0.539	0.518
	4	25	100	1	SAMMER.R	0.549	0.541	0.527
	5	30	100	0.1	SAMMER.R	0.542	0.543	0.524
	6	30	200	1	SAMMER.R	0.548	0.543	0.526

Table 12. ***Hyperparameter Tuning of All Models***

Model		Parameters					Avg FDR at 3%		
Logistic Regression	Iteration	Variables	penalty		solver		trn	tst	oot
	1	20	none		lbfgs		0.535	0.544	0.518
	2	20	l2		lbfgs		0.539	0.533	0.519
	3	25	none		lbfgs		0.539	0.544	0.522
	4	25	l2		lbfgs		0.536	0.548	0.520
	5	30	none		lbfgs		0.545	0.541	0.525
	6	30	l2		lbfgs		0.541	0.533	0.519
	7	20	l1		saga		0.538	0.543	0.520
	8	20	l2		saga		0.539	0.540	0.519
	9	25	l1		saga		0.538	0.538	0.519
	10	25	l2		saga		0.540	0.537	0.520
	11	30	none		saga		0.539	0.536	0.520
	12	30	l2		saga		0.541	0.536	0.520
Single Decision Tree	Iteration	Variables	max_depth		splitter		Avg FDR at 3%		
	1	20	None		random		0.560	0.543	0.530
	2	20	None		best		0.560	0.544	0.529
	3	25	20		random		0.560	0.548	0.531
	4	25	20		best		0.558	0.551	0.528
	5	30	30		random		0.557	0.552	0.528
	6	30	30		best		0.559	0.548	0.530
Random Forest	Iteration	Variables	n_estimators	max_depth	min_samples_leaf	min_samples_split	Avg FDR at 3%		
	1	20	10	10	1	2	0.552	0.567	0.534
	2	25	50	20	1	300	0.560	0.550	0.535
	3	30	100	20	30	300	0.552	0.568	0.537
	4	30	150	30	30	500	0.555	0.549	0.531
Boosted Tree	Iteration	Variables	n_estimators	max_depth	learning_rate		Avg FDR at 3%		
	1	20	100	3	0.1		0.559	0.555	0.536
	2	20	200	3	0.01		0.557	0.552	0.535
	3	25	300	3	0.1		0.557	0.559	0.537
	4	25	500	5	0.01		0.556	0.557	0.537
	5	30	500	5	0.1		0.557	0.559	0.536
	6	30	1000	5	0.01		0.559	0.555	0.536
Neural Network	Iteration	Variables	hidden_layer_sizes	n_layers_	learning_rate	activation	Avg FDR at 3%		
	1	20	5	1	constant	relu	0.550	0.555	0.527
	2	20	5	1	adaptive	logistic	0.551	0.543	0.528
	3	25	10	1	constant	relu	0.553	0.562	0.535
	4	25	10	2	adaptive	logistic	0.554	0.548	0.531
	5	30	20	2	constant	relu	0.559	0.549	0.533
	6	30	20	2	adaptive	logistic	0.552	0.551	0.532
Adaboost	Iteration	Variables	n_estimators		learning_rate	algorithm	Avg FDR at 3%		
	1	20	50		1	SAMME.R	0.550	0.560	0.531
	2	20	50		0.1	SAMME	0.540	0.529	0.518
	3	25	50		1	SAMME	0.531	0.539	0.518
	4	25	100		1	SAMME.R	0.549	0.541	0.527
	5	30	100		0.1	SAMME.R	0.542	0.543	0.524
	6	30	200		1	SAMME.R	0.548	0.543	0.526

7. Results

After building 7 different machine learning models, implementing hyperparameter tuning with each model, and comparing model performance on the training dataset, testing dataset, and out-of-time dataset, we found out that the boosted tree model with 25 variables, the maximum depth of 3 layers, the learning rate of 0.1 and 300 trees is our best performer. The model achieves the fraud detection rate of 56.27% on training data, 54.53% on testing data, and 53.6% on out-of-time data in the top 3% population. We now take a closer look at the model performance on top 20% records of training data, testing data, and out-of-time data.

7.1 Training Data

Figure 13. *Boosted Tree Model Performance on Top 20% Records of Training Data*

Training													
# Records			# Goods			# Bads			Fraud Rate				
625130			616017			9113			0.014577762				
	Bin Statistics					Cumulative Statistics							
Population Bins %	# Records	# Goods	# Bads	% Goods	% Bads	Total # Records	Cumulative Goods	Cumulative Bads	% Goods	% Bads (FDR)	KS	FPR
1	6251	1429	4822	22.86%	77.14%	6251	1429	4822	0.23%	52.91%	52.68%	0.30
2	6251	6033	218	96.51%	3.49%	12503	7463	5040	1.21%	55.31%	54.09%	1.48
3	6251	6163	88	98.59%	1.41%	18754	13626	5128	2.21%	56.27%	54.06%	2.66
4	6251	6204	47	99.25%	0.75%	25005	19830	5175	3.22%	56.79%	53.57%	3.83
5	6251	6219	32	99.49%	0.51%	31257	26050	5207	4.23%	57.14%	52.91%	5.00
6	6251	6186	65	98.96%	1.04%	37508	32236	5272	5.23%	57.85%	52.62%	6.11
7	6251	6191	60	99.04%	0.96%	43759	38427	5332	6.24%	58.51%	52.27%	7.21
8	6251	6188	63	98.99%	1.01%	50010	44615	5395	7.24%	59.20%	51.96%	8.27
9	6251	6195	56	99.10%	0.90%	56262	50811	5451	8.25%	59.82%	51.57%	9.32
10	6251	6218	33	99.47%	0.53%	62513	57029	5484	9.26%	60.18%	50.92%	10.40
11	6251	6217	34	99.46%	0.54%	68764	63246	5518	10.27%	60.55%	50.28%	11.46
12	6251	6201	50	99.20%	0.80%	75016	69448	5568	11.27%	61.10%	49.83%	12.47
13	6251	6209	42	99.33%	0.67%	81267	75657	5610	12.28%	61.56%	49.28%	13.49
14	6251	6218	33	99.47%	0.53%	87518	81875	5643	13.29%	61.92%	48.63%	14.51
15	6251	6208	43	99.31%	0.69%	93770	88084	5686	14.30%	62.39%	48.10%	15.49
16	6251	6205	46	99.26%	0.74%	100021	94289	5732	15.31%	62.90%	47.59%	16.45
17	6251	6217	34	99.46%	0.54%	106272	100506	5766	16.32%	63.27%	46.96%	17.43
18	6251	6207	44	99.30%	0.70%	112523	106713	5810	17.32%	63.76%	46.43%	18.37
19	6251	6201	50	99.20%	0.80%	118775	112915	5860	18.33%	64.30%	45.97%	19.27
20	6251	6197	54	99.14%	0.86%	125026	119112	5914	19.34%	64.90%	45.56%	20.14

7.2 Testing Data

Figure 14. *Boosted Tree Model Performance on Top 20% Records of Testing Data*

Testing													
# Records			# Goods			# Bads			Fraud Rate				
208377			205483			2894			0.013888306				
	Bin Statistics					Cumulative Statistics							
Population Bins %	# Records	# Goods	# Bads	% Goods	% Bads	Total # Records	Cumulative Goods	Cumulative Bads	% Goods	% Bads (FDR)	KS	FPR
1	2084	594	1490	28.49%	71.51%	2084	594	1490	0.29%	51.49%	51.20%	0.40
2	2084	2021	63	96.98%	3.02%	4168	2615	1553	1.27%	53.66%	52.39%	1.68
3	2084	2059	25	98.80%	1.20%	6251	4673	1578	2.27%	54.53%	52.25%	2.96
4	2084	2064	20	99.04%	0.96%	8335	6737	1598	3.28%	55.22%	51.94%	4.22
5	2084	2064	20	99.04%	0.96%	10419	8801	1618	4.28%	55.91%	51.63%	5.44
6	2084	2070	14	99.33%	0.67%	12503	10871	1632	5.29%	56.39%	51.10%	6.66
7	2084	2068	16	99.23%	0.77%	14586	12938	1648	6.30%	56.95%	50.65%	7.85
8	2084	2063	21	98.99%	1.01%	16670	15001	1669	7.30%	57.67%	50.37%	8.99
9	2084	2070	14	99.33%	0.67%	18754	17071	1683	8.31%	58.15%	49.85%	10.14
10	2084	2066	18	99.14%	0.86%	20838	19137	1701	9.31%	58.78%	49.46%	11.25
11	2084	2069	15	99.28%	0.72%	22921	21205	1716	10.32%	59.30%	48.98%	12.36
12	2084	2073	11	99.47%	0.53%	25005	23278	1727	11.33%	59.68%	48.35%	13.48
13	2084	2074	10	99.52%	0.48%	27089	25352	1737	12.34%	60.02%	47.68%	14.60
14	2084	2070	14	99.33%	0.67%	29173	27422	1751	13.35%	60.50%	47.16%	15.66
15	2084	2069	15	99.28%	0.72%	31257	29491	1766	14.35%	61.02%	46.67%	16.70
16	2084	2063	21	98.99%	1.01%	33340	31553	1787	15.36%	61.75%	46.39%	17.66
17	2084	2074	10	99.52%	0.48%	35424	33627	1797	16.36%	62.09%	45.73%	18.71
18	2084	2069	15	99.28%	0.72%	37508	35696	1812	17.37%	62.61%	45.24%	19.70
19	2084	2078	6	99.71%	0.29%	39592	37774	1818	18.38%	62.82%	44.44%	20.78
20	2084	2066	18	99.14%	0.86%	41675	39839	1836	19.39%	63.44%	44.05%	21.70

7.3 Out of Time Data

Figure 15. *Boosted Tree Model Performance on Top 20% Records of Out of Time Data*

OOT													
# Records			# Goods			# Bads			Fraud Rate				
166493			164107			2386			0.014330933				
	Bin Statistics					Cumulative Statistics							
Population Bins %	# Records	# Goods	# Bads	% Goods	% Bads	Total # Records	Cumulative Goods	Cumulative Bads	% Goods	% Bads (FDR)	KS	FPR
1	1665	455	1210	27.32%	72.68%	1665	455	1210	0.28%	50.71%	50.44%	0.38
2	1665	1612	53	96.82%	3.18%	3330	2067	1263	1.26%	52.93%	51.67%	1.64
3	1665	1649	16	99.04%	0.96%	4995	3716	1279	2.26%	53.60%	51.34%	2.91
4	1665	1655	10	99.40%	0.60%	6660	5371	1289	3.27%	54.02%	50.75%	4.17
5	1665	1644	21	98.74%	1.26%	8325	7015	1310	4.27%	54.90%	50.63%	5.35
6	1665	1649	16	99.04%	0.96%	9990	8664	1326	5.28%	55.57%	50.29%	6.53
7	1665	1649	16	99.04%	0.96%	11655	10313	1342	6.28%	56.24%	49.96%	7.68
8	1665	1649	16	99.04%	0.96%	13319	11961	1358	7.29%	56.92%	49.63%	8.81
9	1665	1650	15	99.10%	0.90%	14984	13611	1373	8.29%	57.54%	49.25%	9.91
10	1665	1649	16	99.04%	0.96%	16649	15260	1389	9.30%	58.21%	48.92%	10.99
11	1665	1656	9	99.46%	0.54%	18314	16916	1398	10.31%	58.59%	48.28%	12.10
12	1665	1651	14	99.16%	0.84%	19979	18567	1412	11.31%	59.18%	47.86%	13.15
13	1665	1656	9	99.46%	0.54%	21644	20223	1421	12.32%	59.56%	47.23%	14.23
14	1665	1642	23	98.62%	1.38%	23309	21865	1444	13.32%	60.52%	47.20%	15.14
15	1665	1653	12	99.28%	0.72%	24974	23518	1456	14.33%	61.02%	46.69%	16.15
16	1665	1656	9	99.46%	0.54%	26639	25174	1465	15.34%	61.40%	46.06%	17.18
17	1665	1652	13	99.22%	0.78%	28304	26826	1478	16.35%	61.94%	45.60%	18.15
18	1665	1650	15	99.10%	0.90%	29969	28476	1493	17.35%	62.57%	45.22%	19.07
19	1665	1650	15	99.10%	0.90%	31634	30126	1508	18.36%	63.20%	44.84%	19.98
20	1665	1651	14	99.16%	0.84%	33299	31777	1522	19.36%	63.79%	44.43%	20.88

8. Conclusions

To identify fraud in the 2016 application data, we investigated and cleaned the dataset with 1,000,000 records and 8 fields. Before building machine learning models, we generated 1,035 variables based on our domain knowledge. We used filter and wrapper in the feature selection stage and attained the top 30 most relevant variables. In the model constructing process, we kept the data in the last two months of 2016 as out-of-time (OOT) data and randomly split 75% of remaining data as the training data and 25% as the testing data. We developed 7 different models including logistic regression, decision tree, random forest, boosted tree, neural network and adaboost and experimented with different hyperparameters to find the best performer.

We found that boosted tree model with 25 variables, the maximum depth of 3 layers, learning rate of 0.1 and 300 trees is our best performer in detecting fraud in this dataset. Our boosted tree model detected 56.27% of all frauds in top 3% training data, 54.53% of all frauds in top 3% of testing data, and 53.6% of all frauds in top 3% of out-of-time data.

By constructing a supervised model based on a synthetic dataset of applications, we have successfully completed several important procedures of fraud analytics including data preparation, feature engineering, feature selection and model building. For each step, there is still room for improvements for us to better generalize our model for out of time data prediction. In the data gathering stage, we will have a greater scope to detect frauds if we have more demographic information such as gender, income, and employment information. In the data preparation stage, we can investigate more suspicious and erroneous records. For example, some records have some date of birth values which is after the application date. In feature engineering stage, we can try to create more new entities and variables associated with ssn since most of our effective variables are built upon ssn in this dataset. Finally, although we have used FDR at top 3% as our main evaluation metric for this project, we can explore more metrics to adapt to different business objectives.

9. Reference

1. *Lightgbm.lgbmclassifier*. lightgbm.LGBMClassifier - LightGBM 3.3.2.99 documentation. (n.d.). Retrieved March 24, 2022, from https://lightgbm.readthedocs.io/en/latest/pythonapi/lightgbm.LGBMClassifier.html

2. *Sklearn.ensemble.adaboostclassifier*. scikit. (n.d.). Retrieved March 24, 2022, from https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.AdaBoostClassifier.html?highlight=classifier#sklearn.ensemble.AdaBoostClassifier

3. *Sklearn.ensemble.randomforestclassifier*. scikit. (n.d.). Retrieved March 24, 2022, from https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html?highlight=classifier#sklearn.ensemble.RandomForestClassifier

4. *Sklearn.linear_model.logisticregression*. scikit. (n.d.). Retrieved March 24, 2022, from https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html?highlight=logistic#sklearn.linear_model.LogisticRegression

5. *Sklearn.neural_network.MLPClassifier*. scikit. (n.d.). Retrieved March 24, 2022, from https://scikit-learn.org/stable/modules/generated/sklearn.neural_network.MLPClassifier.html?highlight=mlpclassifier#sklearn.neural_network.MLPClassifier

6. *Sklearn.tree.decisiontreeclassifier*. scikit. (n.d.). Retrieved March 24, 2022, from https://scikit-learn.org/stable/modules/generated/sklearn.tree.DecisionTreeClassifier.html?highlight=decisiontreeclassifier#sklearn.tree.DecisionTreeClassifier

10. Appendix

10.1 Data Quality Report on Product Application Data

10.1.1 File Description

The dataset includes 1,000,000 records of product applications and 10 fields for each record. It covers records from 1/1/2016 to 12/31/2016. 8 fields are personal identifying information except the ‘record’ and ‘fraud_label’ fields. The dataset is synthesized from a few billion real U.S. credit card applications over the past 10 years.

10.1.2 Summary statistics table

Table 13 (Table 1)

Numeric Fields Summary

Field Name	% Populated	Min	Max	Mean	Stdev	*% Zero
date	100	01-01-2016	12-31-2016	-	-	0
dob	100	19000101	20161031	-	-	0

**% Zero: only including record whose value is 0.*

Table 14 (Table 2)

Categorical Fields Summary

Field Name	% Populated	*Unique Values	Most Common Value
record	100	1,000,000	-
ssn	100	835,819	999999999
firstname	100	78,136	EAMSTRMT
lastname	100	177,001	ERJSAXA
address	100	828,774	123 MAIN ST
zip5	100	26,370	68138
homephone	100	28,244	9999999999
fraud_label	100	2	0

**Unique Values: does not include Nan.*

10.1.3 Fields Distribution

- **date**

Figure 16

Daily application. (Set daily count of 02-29 as daily count of 02-28)

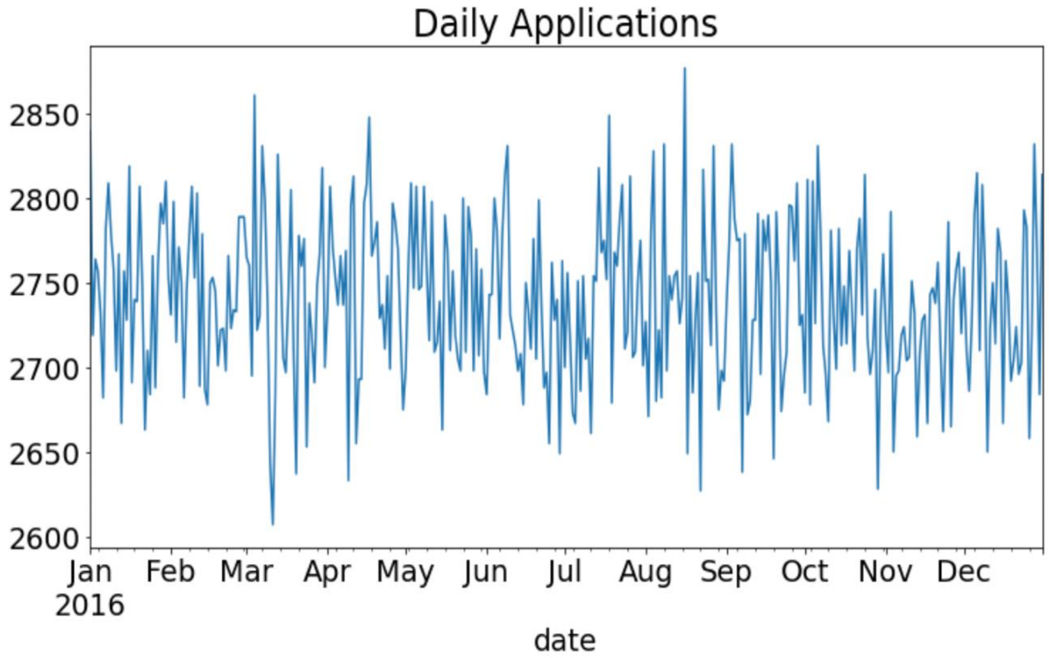
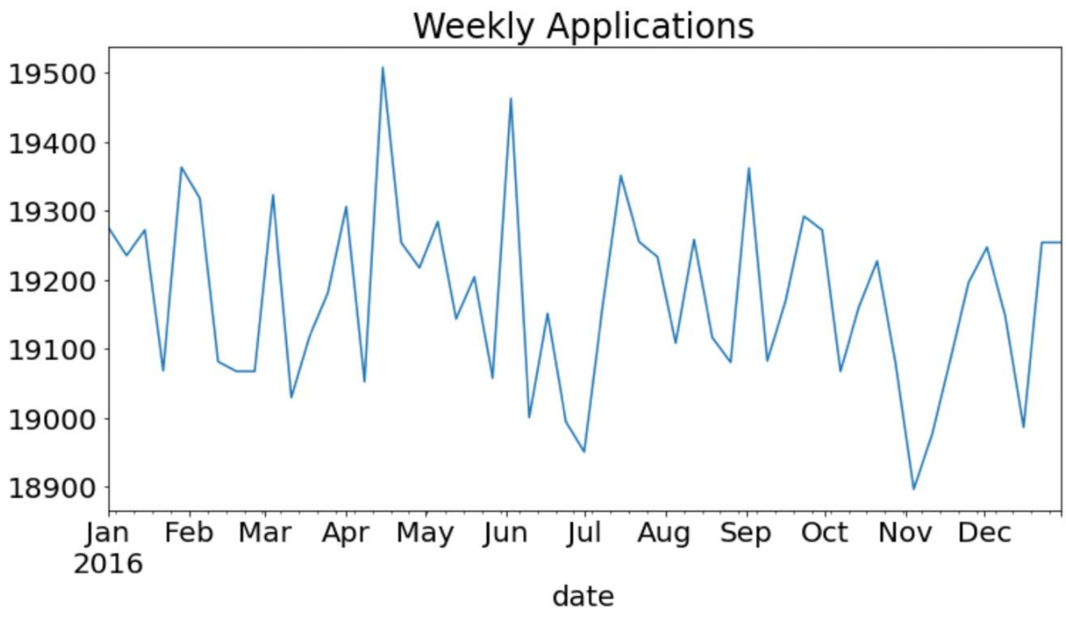


Figure 17 (Figure 1)

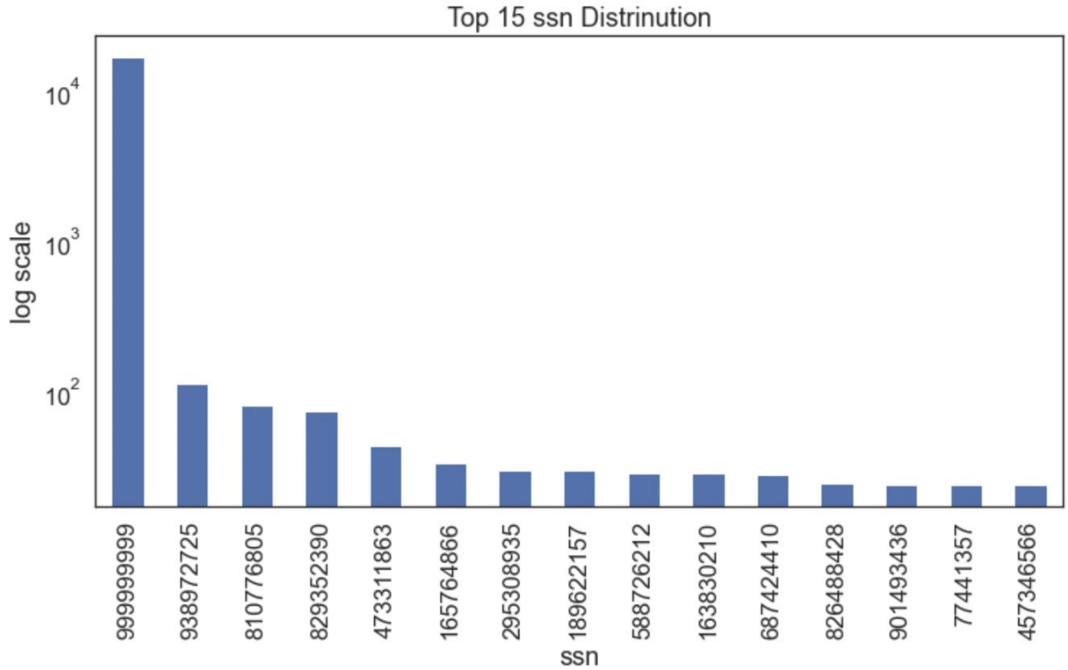
Weekly Application. (Set daily count of 02-26 as that of 02-19 and daily count of 12-30 as that of 12-23)



- **ssn (Figure 4)**

Figure 18

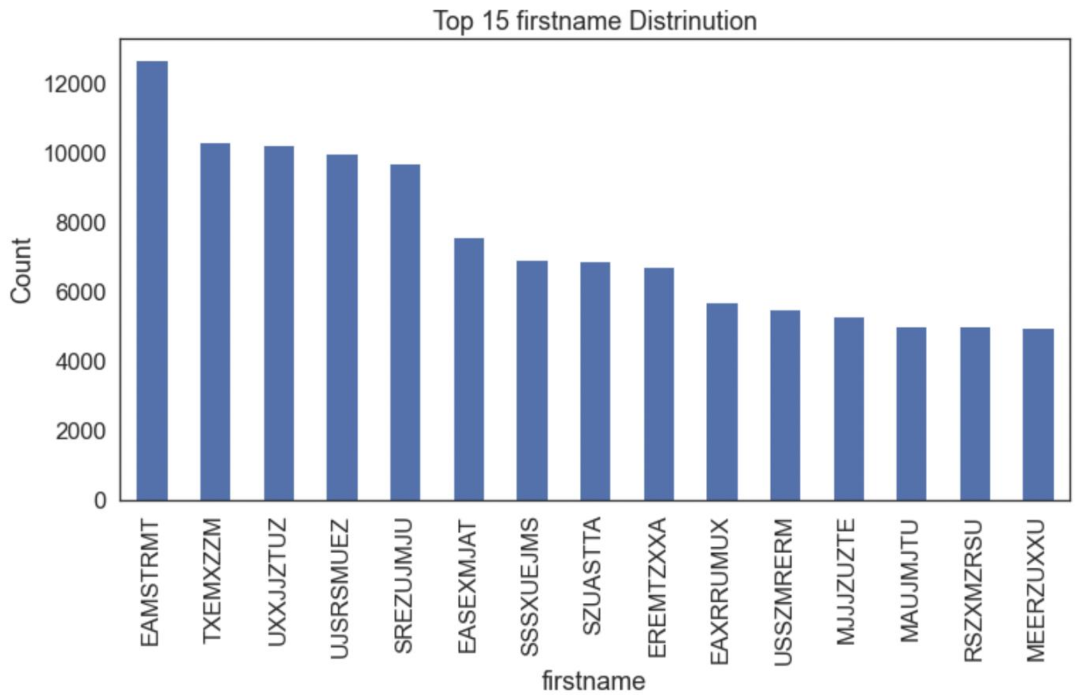
Distribution of ssn. The value of y-axis is log(count).



- **firstname**

Figure 19

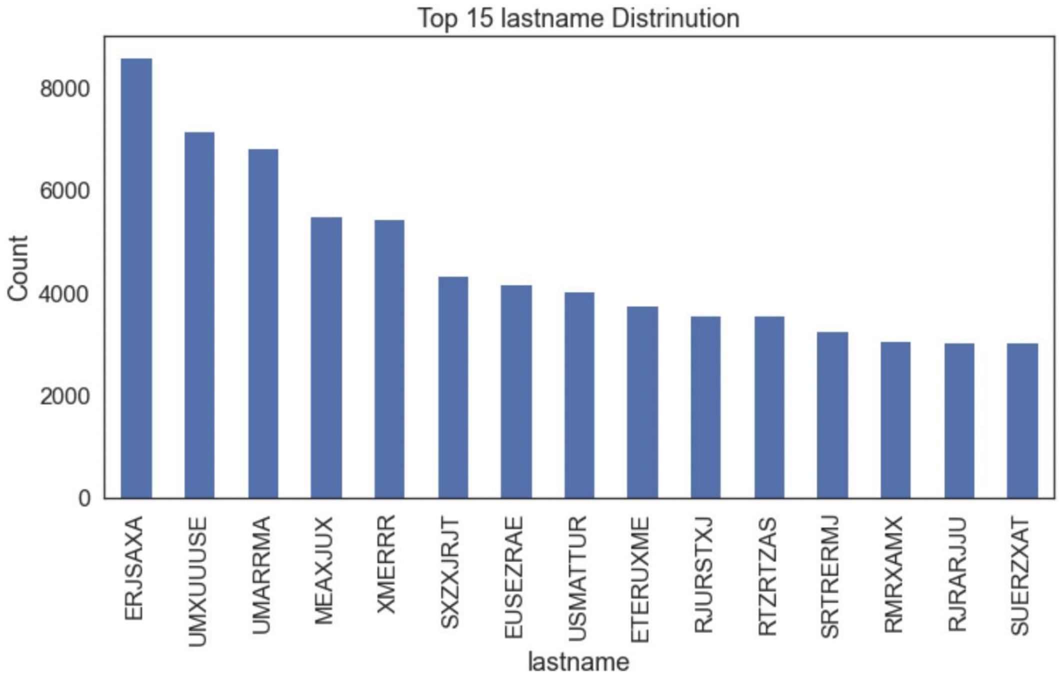
Distribution of Firstname



- **lastname**

Figure 20

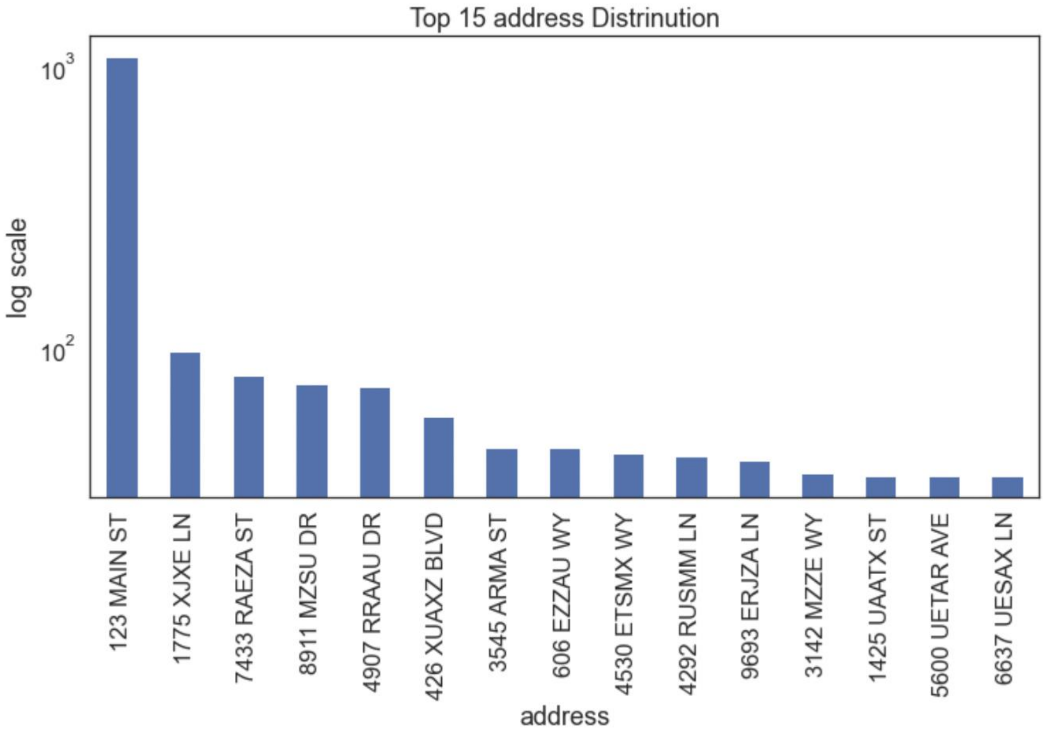
Distribution of Lastname



- **address**

Figure 21

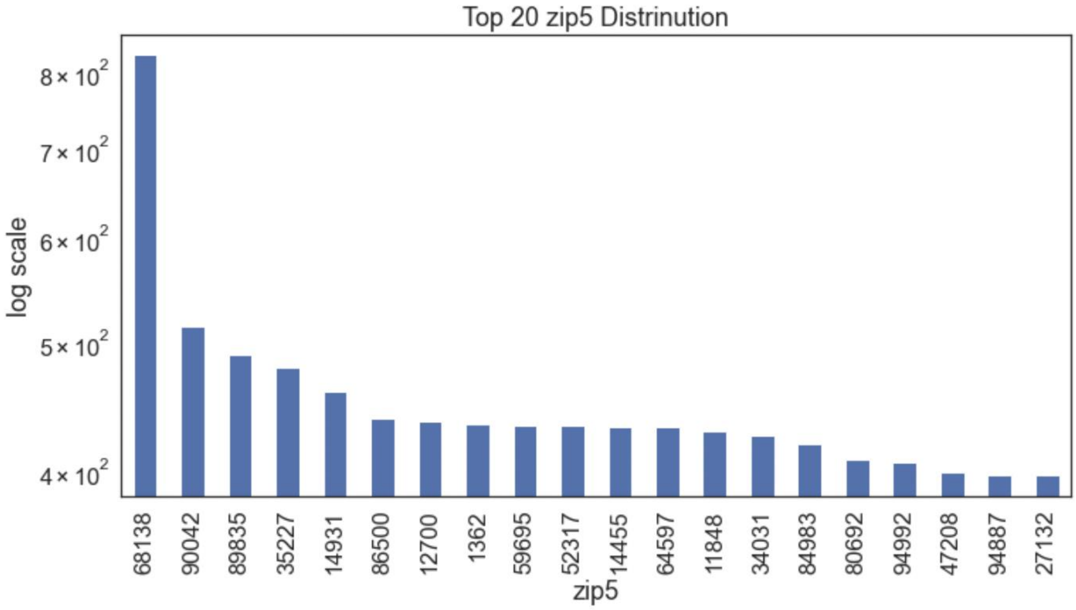
Distribution of Address. The value of y-axis is log(count).



- **zip5**

Figure 22

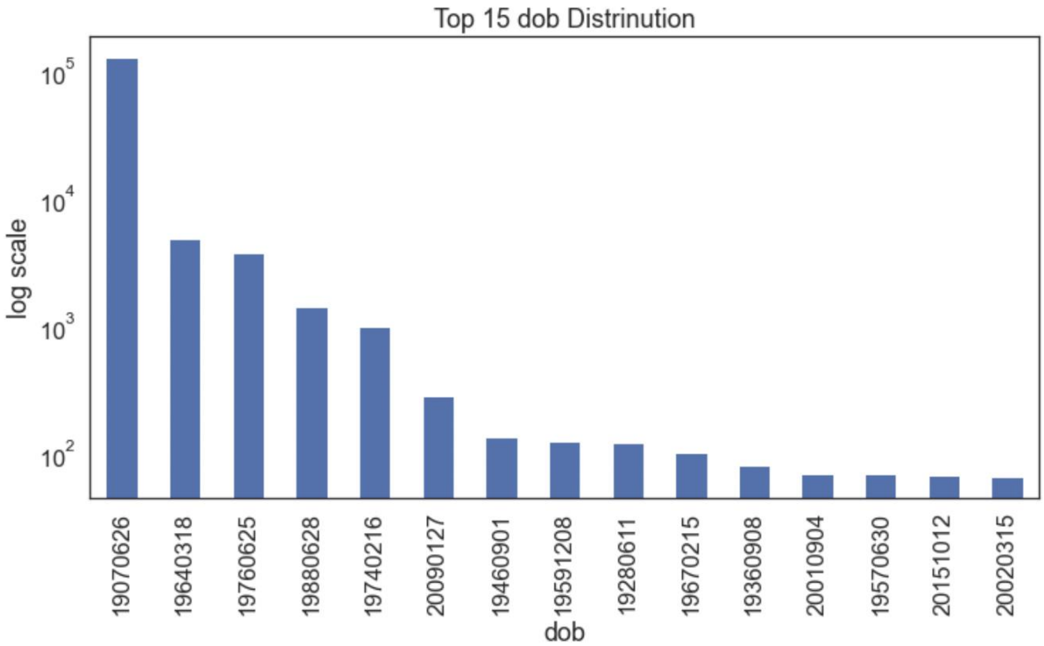
Distribution of zip5. The value of y-axis is log(count).



- **dob (Figure 5)**

Figure 23

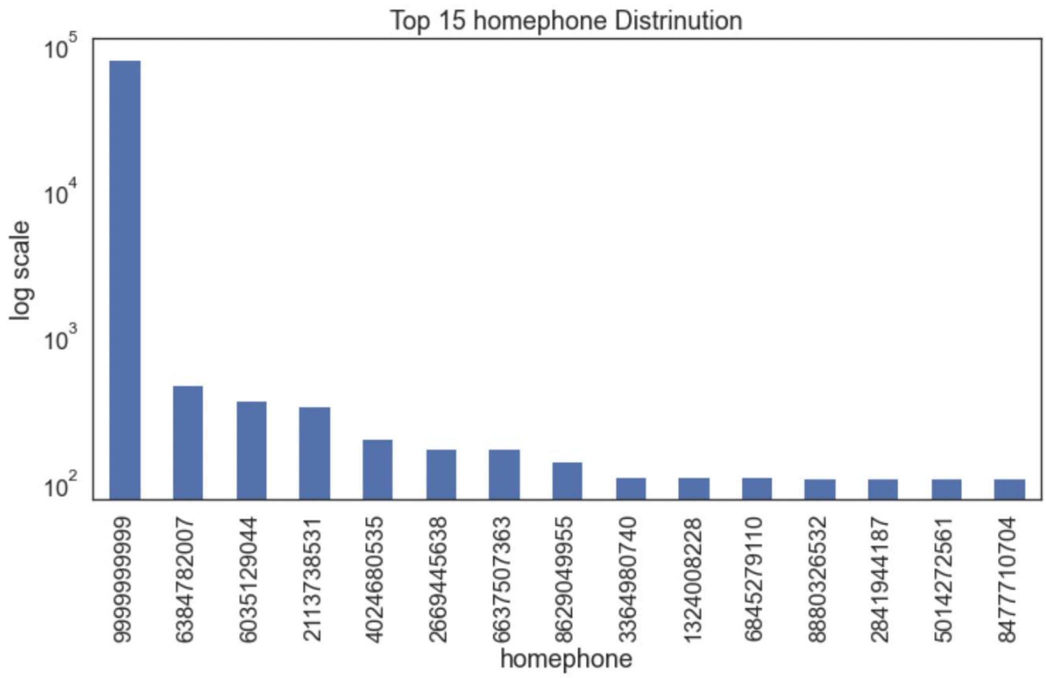
Distribution of dob. The value of y-axis is log(count).



- **homephone**

Figure 24

Distribution of homephone. The value of y-axis is log(count).



- **fraud_label**

Figure 25 (Figure 2)

Distribution of fraud_label. (fraud_laebl_0 : fraud_laebl_1 = 985,607 : 14,393)

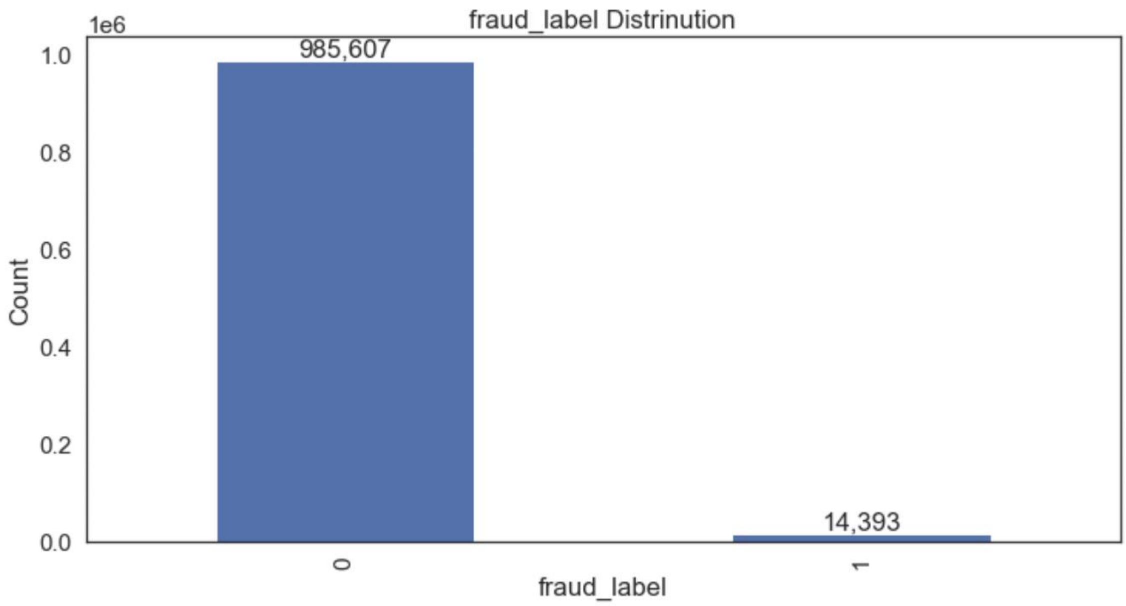


Figure 26

Bad Application Daily Proportion Distribution

Bad (red): fraud_label=1

Proportion: (daily count of bad applications) / (daily count of total applications)

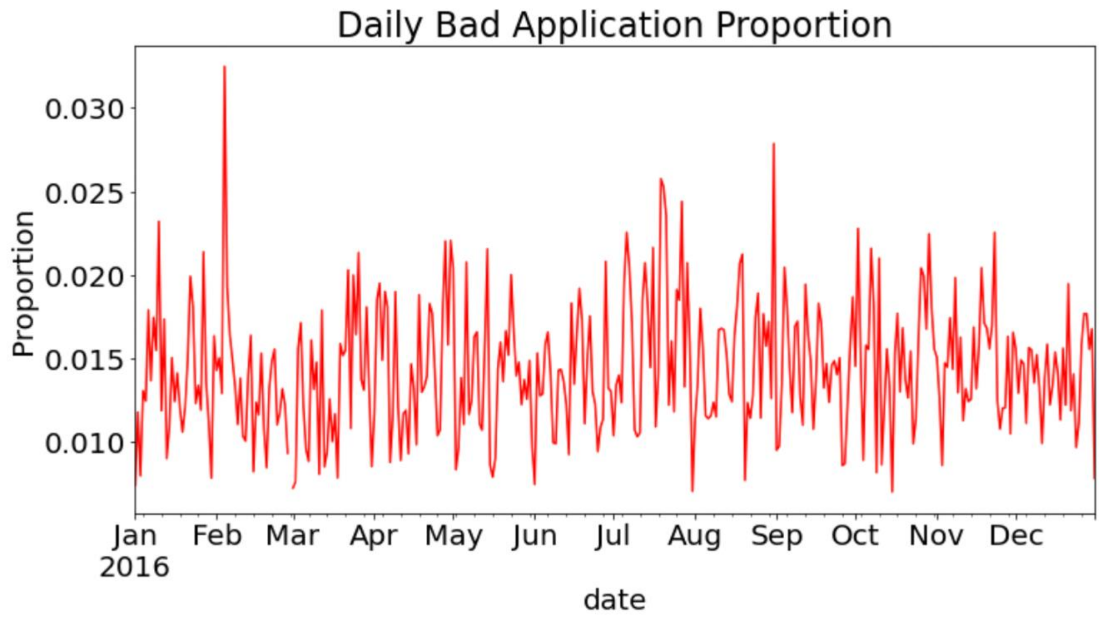


Figure 27 (Figure 3)

Bad Application Weekly Proportion Distribution

Bad (red): fraud_label=1

Proportion: (weekly count of bad applications) / (weekly count of total applications)

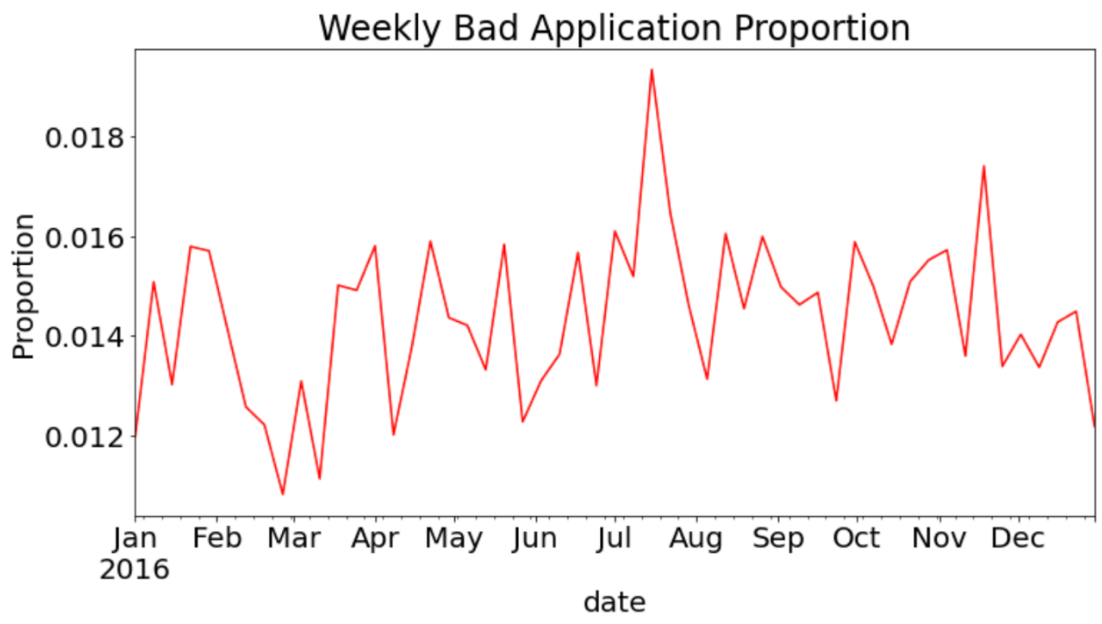


Figure 28

Good Application Daily Proportion Distribution

Good (green): fraud_label=0

Proportion: (daily count of good applications) / (daily count of total applications)

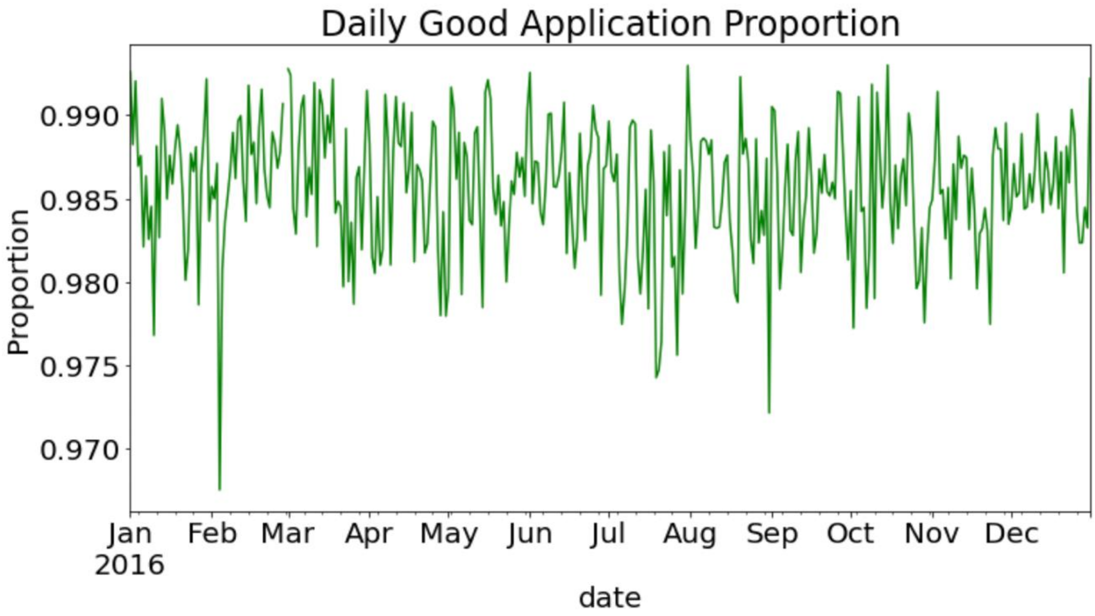
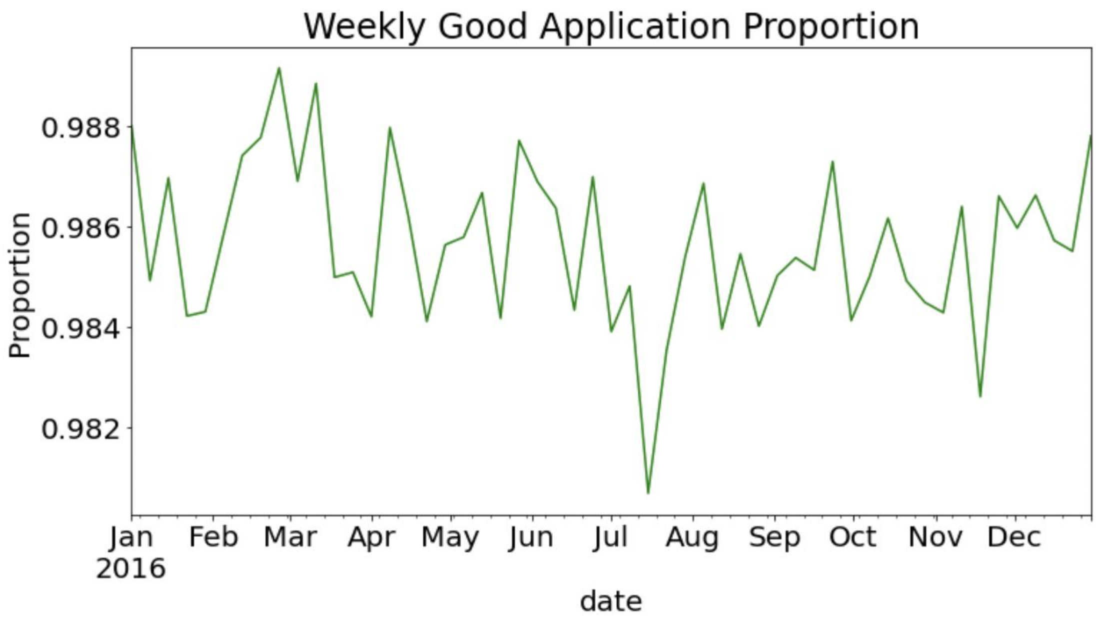


Figure 29

Good Application Weekly Proportion Distribution

Good (green): fraud_label=0

Proportion: (weekly count of good applications) / (weekly count of total applications)



10.2 All Created Variables

Table 15. *List of All Created Variables*

ssn_day_since	ssn_name_homephone_day_since	dob_homephone_unique_count_for_name_fulladdress_7
ssn_count_0	ssn_name_homephone_count_0	dob_homephone_unique_count_for_name_fulladdress_14
ssn_count_1	ssn_name_homephone_count_1	dob_homephone_unique_count_for_name_fulladdress_30
ssn_count_3	ssn_name_homephone_count_3	dob_homephone_unique_count_for_name_homephone_0
ssn_count_7	ssn_name_homephone_count_7	dob_homephone_unique_count_for_name_homephone_1
ssn_count_14	ssn_name_homephone_count_14	dob_homephone_unique_count_for_name_homephone_3
ssn_count_30	ssn_name_homephone_count_30	dob_homephone_unique_count_for_name_homephone_7
ssn_count_0_by_3	ssn_name_homephone_count_0_by_3	dob_homephone_unique_count_for_name_homephone_14
ssn_count_0_by_7	ssn_name_homephone_count_0_by_7	dob_homephone_unique_count_for_name_homephone_30
ssn_count_0_by_14	ssn_name_homephone_count_0_by_14	dob_homephone_unique_count_for_fulladdress_dob_0
ssn_count_0_by_30	ssn_name_homephone_count_0_by_30	dob_homephone_unique_count_for_fulladdress_dob_1
ssn_count_1_by_3	ssn_name_homephone_count_1_by_3	dob_homephone_unique_count_for_fulladdress_dob_3
ssn_count_1_by_7	ssn_name_homephone_count_1_by_7	dob_homephone_unique_count_for_fulladdress_dob_7
ssn_count_1_by_14	ssn_name_homephone_count_1_by_14	dob_homephone_unique_count_for_fulladdress_dob_14
ssn_count_1_by_30	ssn_name_homephone_count_1_by_30	dob_homephone_unique_count_for_fulladdress_dob_30
dob_day_since	ssn_fulladdress_dob_day_since	dob_homephone_unique_count_for_ssn_dob_0
dob_count_0	ssn_fulladdress_dob_count_0	dob_homephone_unique_count_for_ssn_dob_1
dob_count_1	ssn_fulladdress_dob_count_1	dob_homephone_unique_count_

		for_ssn_dob_3
dob_count_3	ssn_fulladdress_dob_count_3	dob_homephone_unique_count_for_ssn_dob_7
dob_count_7	ssn_fulladdress_dob_count_7	dob_homephone_unique_count_for_ssn_dob_14
dob_count_14	ssn_fulladdress_dob_count_14	dob_homephone_unique_count_for_ssn_dob_30
dob_count_30	ssn_fulladdress_dob_count_30	dob_homephone_unique_count_for_ssn_homephone_0
dob_count_0_by_3	ssn_fulladdress_dob_count_0_by_3	dob_homephone_unique_count_for_ssn_homephone_1
dob_count_0_by_7	ssn_fulladdress_dob_count_0_by_7	dob_homephone_unique_count_for_ssn_homephone_3
dob_count_0_by_14	ssn_fulladdress_dob_count_0_by_14	dob_homephone_unique_count_for_ssn_homephone_7
dob_count_0_by_30	ssn_fulladdress_dob_count_0_by_30	dob_homephone_unique_count_for_ssn_homephone_14
dob_count_1_by_3	ssn_fulladdress_dob_count_1_by_3	dob_homephone_unique_count_for_ssn_homephone_30
dob_count_1_by_7	ssn_fulladdress_dob_count_1_by_7	dob_homephone_unique_count_for_ssn_name_0
dob_count_1_by_14	ssn_fulladdress_dob_count_1_by_14	dob_homephone_unique_count_for_ssn_name_1
dob_count_1_by_30	ssn_fulladdress_dob_count_1_by_30	dob_homephone_unique_count_for_ssn_name_3
homephone_day_since	ssn_fulladdress_homephone_day_since	dob_homephone_unique_count_for_ssn_name_7
homephone_count_0	ssn_fulladdress_homephone_count_0	dob_homephone_unique_count_for_ssn_name_14
homephone_count_1	ssn_fulladdress_homephone_count_1	dob_homephone_unique_count_for_ssn_name_30
homephone_count_3	ssn_fulladdress_homephone_count_3	dob_homephone_unique_count_for_ssn_fulladdress_0
homephone_count_7	ssn_fulladdress_homephone_count_7	dob_homephone_unique_count_for_ssn_fulladdress_1
homephone_count_14	ssn_fulladdress_homephone_count_14	dob_homephone_unique_count_for_ssn_fulladdress_3

homephone_count_30	ssn_fulladdress_homephone_count_30	dob_homephone_unique_count_for_ssn_fulladdress_7
homephone_count_0_by_3	ssn_fulladdress_homephone_count_0_by_3	dob_homephone_unique_count_for_ssn_fulladdress_14
homephone_count_0_by_7	ssn_fulladdress_homephone_count_0_by_7	dob_homephone_unique_count_for_ssn_fulladdress_30
homephone_count_0_by_14	ssn_fulladdress_homephone_count_0_by_14	ssn_dob_unique_count_for_ssn_0
homephone_count_0_by_30	ssn_fulladdress_homephone_count_0_by_30	ssn_dob_unique_count_for_ssn_1
homephone_count_1_by_3	ssn_fulladdress_homephone_count_1_by_3	ssn_dob_unique_count_for_ssn_3
homephone_count_1_by_7	ssn_fulladdress_homephone_count_1_by_7	ssn_dob_unique_count_for_ssn_7
homephone_count_1_by_14	ssn_fulladdress_homephone_count_1_by_14	ssn_dob_unique_count_for_ssn_14
homephone_count_1_by_30	ssn_fulladdress_homephone_count_1_by_30	ssn_dob_unique_count_for_ssn_30
name_day_since	ssn_dob_homephone_day_since	ssn_dob_unique_count_for_name_dob_0
name_count_0	ssn_dob_homephone_count_0	ssn_dob_unique_count_for_name_dob_1
name_count_1	ssn_dob_homephone_count_1	ssn_dob_unique_count_for_name_dob_3
name_count_3	ssn_dob_homephone_count_3	ssn_dob_unique_count_for_name_dob_7
name_count_7	ssn_dob_homephone_count_7	ssn_dob_unique_count_for_name_dob_14
name_count_14	ssn_dob_homephone_count_14	ssn_dob_unique_count_for_name_dob_30
name_count_30	ssn_dob_homephone_count_30	ssn_dob_unique_count_for_name_fulladdress_0
name_count_0_by_3	ssn_dob_homephone_count_0_by_3	ssn_dob_unique_count_for_name_fulladdress_1
name_count_0_by_7	ssn_dob_homephone_count_0_by_7	ssn_dob_unique_count_for_name_fulladdress_3
name_count_0_by_14	ssn_dob_homephone_count_0_b	ssn_dob_unique_count_for_nam

	y_14	e_fulladdress_7
name_count_0_by_30	ssn_dob_homephone_count_0_by_30	ssn_dob_unique_count_for_name_fulladdress_14
name_count_1_by_3	ssn_dob_homephone_count_1_by_3	ssn_dob_unique_count_for_name_fulladdress_30
name_count_1_by_7	ssn_dob_homephone_count_1_by_7	ssn_dob_unique_count_for_name_homephone_0
name_count_1_by_14	ssn_dob_homephone_count_1_by_14	ssn_dob_unique_count_for_name_homephone_1
name_count_1_by_30	ssn_dob_homephone_count_1_by_30	ssn_dob_unique_count_for_name_homephone_3
fulladdress_day_since	ssn_unique_count_for_name_dob_0	ssn_dob_unique_count_for_name_homephone_7
fulladdress_count_0	ssn_unique_count_for_name_dob_1	ssn_dob_unique_count_for_name_homephone_14
fulladdress_count_1	ssn_unique_count_for_name_dob_3	ssn_dob_unique_count_for_name_homephone_30
fulladdress_count_3	ssn_unique_count_for_name_dob_7	ssn_dob_unique_count_for_fulladdress_dob_0
fulladdress_count_7	ssn_unique_count_for_name_dob_14	ssn_dob_unique_count_for_fulladdress_dob_1
fulladdress_count_14	ssn_unique_count_for_name_dob_30	ssn_dob_unique_count_for_fulladdress_dob_3
fulladdress_count_30	ssn_unique_count_for_name_fulladdress_0	ssn_dob_unique_count_for_fulladdress_dob_7
fulladdress_count_0_by_3	ssn_unique_count_for_name_fulladdress_1	ssn_dob_unique_count_for_fulladdress_dob_14
fulladdress_count_0_by_7	ssn_unique_count_for_name_fulladdress_3	ssn_dob_unique_count_for_fulladdress_dob_30
fulladdress_count_0_by_14	ssn_unique_count_for_name_fulladdress_7	ssn_dob_unique_count_for_dob_homephone_0
fulladdress_count_0_by_30	ssn_unique_count_for_name_fulladdress_14	ssn_dob_unique_count_for_dob_homephone_1
fulladdress_count_1_by_3	ssn_unique_count_for_name_fulladdress_30	ssn_dob_unique_count_for_dob_homephone_3
fulladdress_count_1_by_7	ssn_unique_count_for_name_homephone_0	ssn_dob_unique_count_for_dob_homephone_7

fulladdress_count_1_by_14	ssn_unique_count_for_name_homephone_1	ssn_dob_unique_count_for_dob_homephone_14
fulladdress_count_1_by_30	ssn_unique_count_for_name_homephone_3	ssn_dob_unique_count_for_dob_homephone_30
name_dob_day_since	ssn_unique_count_for_name_homephone_7	ssn_dob_unique_count_for_ssn_homephone_0
name_dob_count_0	ssn_unique_count_for_name_homephone_14	ssn_dob_unique_count_for_ssn_homephone_1
name_dob_count_1	ssn_unique_count_for_name_homephone_30	ssn_dob_unique_count_for_ssn_homephone_3
name_dob_count_3	ssn_unique_count_for_fulladdress_dob_0	ssn_dob_unique_count_for_ssn_homephone_7
name_dob_count_7	ssn_unique_count_for_fulladdress_dob_1	ssn_dob_unique_count_for_ssn_homephone_14
name_dob_count_14	ssn_unique_count_for_fulladdress_dob_3	ssn_dob_unique_count_for_ssn_homephone_30
name_dob_count_30	ssn_unique_count_for_fulladdress_dob_7	ssn_dob_unique_count_for_ssn_name_0
name_dob_count_0_by_3	ssn_unique_count_for_fulladdress_dob_14	ssn_dob_unique_count_for_ssn_name_1
name_dob_count_0_by_7	ssn_unique_count_for_fulladdress_dob_30	ssn_dob_unique_count_for_ssn_name_3
name_dob_count_0_by_14	ssn_unique_count_for_dob_homephone_0	ssn_dob_unique_count_for_ssn_name_7
name_dob_count_0_by_30	ssn_unique_count_for_dob_homephone_1	ssn_dob_unique_count_for_ssn_name_14
name_dob_count_1_by_3	ssn_unique_count_for_dob_homephone_3	ssn_dob_unique_count_for_ssn_name_30
name_dob_count_1_by_7	ssn_unique_count_for_dob_homephone_7	ssn_dob_unique_count_for_ssn_fulladdress_0
name_dob_count_1_by_14	ssn_unique_count_for_dob_homephone_14	ssn_dob_unique_count_for_ssn_fulladdress_1
name_dob_count_1_by_30	ssn_unique_count_for_dob_homephone_30	ssn_dob_unique_count_for_ssn_fulladdress_3
name_fulladdress_day_since	ssn_unique_count_for_ssn_dob_0	ssn_dob_unique_count_for_ssn_fulladdress_7
name_fulladdress_count_0	ssn_unique_count_for_ssn_dob_	ssn_dob_unique_count_for_ssn_

	1	fulladdress_14
name_fulladdress_count_1	ssn_unique_count_for_ssn_dob_3	ssn_dob_unique_count_for_ssn_fulladdress_30
name_fulladdress_count_3	ssn_unique_count_for_ssn_dob_7	ssn_homephone_unique_count_for_ssn_0
name_fulladdress_count_7	ssn_unique_count_for_ssn_dob_14	ssn_homephone_unique_count_for_ssn_1
name_fulladdress_count_14	ssn_unique_count_for_ssn_dob_30	ssn_homephone_unique_count_for_ssn_3
name_fulladdress_count_30	ssn_unique_count_for_ssn_homephone_0	ssn_homephone_unique_count_for_ssn_7
name_fulladdress_count_0_by_3	ssn_unique_count_for_ssn_homephone_1	ssn_homephone_unique_count_for_ssn_14
name_fulladdress_count_0_by_7	ssn_unique_count_for_ssn_homephone_3	ssn_homephone_unique_count_for_ssn_30
name_fulladdress_count_0_by_14	ssn_unique_count_for_ssn_homephone_7	ssn_homephone_unique_count_for_name_dob_0
name_fulladdress_count_0_by_30	ssn_unique_count_for_ssn_homephone_14	ssn_homephone_unique_count_for_name_dob_1
name_fulladdress_count_1_by_3	ssn_unique_count_for_ssn_homephone_30	ssn_homephone_unique_count_for_name_dob_3
name_fulladdress_count_1_by_7	ssn_unique_count_for_ssn_name_0	ssn_homephone_unique_count_for_name_dob_7
name_fulladdress_count_1_by_14	ssn_unique_count_for_ssn_name_1	ssn_homephone_unique_count_for_name_dob_14
name_fulladdress_count_1_by_30	ssn_unique_count_for_ssn_name_3	ssn_homephone_unique_count_for_name_dob_30
name_homephone_day_since	ssn_unique_count_for_ssn_name_7	ssn_homephone_unique_count_for_name_fulladdress_0
name_homephone_count_0	ssn_unique_count_for_ssn_name_14	ssn_homephone_unique_count_for_name_fulladdress_1
name_homephone_count_1	ssn_unique_count_for_ssn_name_30	ssn_homephone_unique_count_for_name_fulladdress_3
name_homephone_count_3	ssn_unique_count_for_ssn_fulladdress_0	ssn_homephone_unique_count_for_name_fulladdress_7
name_homephone_count_7	ssn_unique_count_for_ssn_fulladdress_1	ssn_homephone_unique_count_for_name_fulladdress_14

name_homephone_count_14	ssn_unique_count_for_ssn_fulladdress_3	ssn_homephone_unique_count_for_name_fulladdress_30
name_homephone_count_30	ssn_unique_count_for_ssn_fulladdress_7	ssn_homephone_unique_count_for_name_homephone_0
name_homephone_count_0_by_3	ssn_unique_count_for_ssn_fulladdress_14	ssn_homephone_unique_count_for_name_homephone_1
name_homephone_count_0_by_7	ssn_unique_count_for_ssn_fulladdress_30	ssn_homephone_unique_count_for_name_homephone_3
name_homephone_count_0_by_14	name_dob_unique_count_for_ssn_0	ssn_homephone_unique_count_for_name_homephone_7
name_homephone_count_0_by_30	name_dob_unique_count_for_ssn_1	ssn_homephone_unique_count_for_name_homephone_14
name_homephone_count_1_by_3	name_dob_unique_count_for_ssn_3	ssn_homephone_unique_count_for_name_homephone_30
name_homephone_count_1_by_7	name_dob_unique_count_for_ssn_7	ssn_homephone_unique_count_for_fulladdress_dob_0
name_homephone_count_1_by_14	name_dob_unique_count_for_ssn_14	ssn_homephone_unique_count_for_fulladdress_dob_1
name_homephone_count_1_by_30	name_dob_unique_count_for_ssn_30	ssn_homephone_unique_count_for_fulladdress_dob_3
fulladdress_dob_day_since	name_dob_unique_count_for_name_fulladdress_0	ssn_homephone_unique_count_for_fulladdress_dob_7
fulladdress_dob_count_0	name_dob_unique_count_for_name_fulladdress_1	ssn_homephone_unique_count_for_fulladdress_dob_14
fulladdress_dob_count_1	name_dob_unique_count_for_name_fulladdress_3	ssn_homephone_unique_count_for_fulladdress_dob_30
fulladdress_dob_count_3	name_dob_unique_count_for_name_fulladdress_7	ssn_homephone_unique_count_for_dob_homephone_0
fulladdress_dob_count_7	name_dob_unique_count_for_name_fulladdress_14	ssn_homephone_unique_count_for_dob_homephone_1
fulladdress_dob_count_14	name_dob_unique_count_for_name_fulladdress_30	ssn_homephone_unique_count_for_dob_homephone_3
fulladdress_dob_count_30	name_dob_unique_count_for_name_homephone_0	ssn_homephone_unique_count_for_dob_homephone_7
fulladdress_dob_count_0_by_3	name_dob_unique_count_for_name_homephone_1	ssn_homephone_unique_count_for_dob_homephone_14
fulladdress_dob_count_0_by_7	name_dob_unique_count_for_na	ssn_homephone_unique_count_f

	me_homephone_3	or_dob_homephone_30
fulladdress_dob_count_0_by_14	name_dob_unique_count_for_name_homephone_7	ssn_homephone_unique_count_for_ssn_dob_0
fulladdress_dob_count_0_by_30	name_dob_unique_count_for_name_homephone_14	ssn_homephone_unique_count_for_ssn_dob_1
fulladdress_dob_count_1_by_3	name_dob_unique_count_for_name_homephone_30	ssn_homephone_unique_count_for_ssn_dob_3
fulladdress_dob_count_1_by_7	name_dob_unique_count_for_fulladdress_dob_0	ssn_homephone_unique_count_for_ssn_dob_7
fulladdress_dob_count_1_by_14	name_dob_unique_count_for_fulladdress_dob_1	ssn_homephone_unique_count_for_ssn_dob_14
fulladdress_dob_count_1_by_30	name_dob_unique_count_for_fulladdress_dob_3	ssn_homephone_unique_count_for_ssn_dob_30
fulladdress_homephone_day_since	name_dob_unique_count_for_fulladdress_dob_7	ssn_homephone_unique_count_for_ssn_name_0
fulladdress_homephone_count_0	name_dob_unique_count_for_fulladdress_dob_14	ssn_homephone_unique_count_for_ssn_name_1
fulladdress_homephone_count_1	name_dob_unique_count_for_fulladdress_dob_30	ssn_homephone_unique_count_for_ssn_name_3
fulladdress_homephone_count_3	name_dob_unique_count_for_dob_homephone_0	ssn_homephone_unique_count_for_ssn_name_7
fulladdress_homephone_count_7	name_dob_unique_count_for_dob_homephone_1	ssn_homephone_unique_count_for_ssn_name_14
fulladdress_homephone_count_14	name_dob_unique_count_for_dob_homephone_3	ssn_homephone_unique_count_for_ssn_name_30
fulladdress_homephone_count_30	name_dob_unique_count_for_dob_homephone_7	ssn_homephone_unique_count_for_ssn_fulladdress_0
fulladdress_homephone_count_0_by_3	name_dob_unique_count_for_dob_homephone_14	ssn_homephone_unique_count_for_ssn_fulladdress_1
fulladdress_homephone_count_0_by_7	name_dob_unique_count_for_dob_homephone_30	ssn_homephone_unique_count_for_ssn_fulladdress_3
fulladdress_homephone_count_0_by_14	name_dob_unique_count_for_ssn_dob_0	ssn_homephone_unique_count_for_ssn_fulladdress_7
fulladdress_homephone_count_0_by_30	name_dob_unique_count_for_ssn_dob_1	ssn_homephone_unique_count_for_ssn_fulladdress_14
fulladdress_homephone_count_1_by_3	name_dob_unique_count_for_ssn_dob_3	ssn_homephone_unique_count_for_ssn_fulladdress_30

fulladdress_homephone_count_1_by_7	name_dob_unique_count_for_ssn_dob_7	ssn_name_unique_count_for_ssn_0
fulladdress_homephone_count_1_by_14	name_dob_unique_count_for_ssn_dob_14	ssn_name_unique_count_for_ssn_1
fulladdress_homephone_count_1_by_30	name_dob_unique_count_for_ssn_dob_30	ssn_name_unique_count_for_ssn_3
dob_homephone_day_since	name_dob_unique_count_for_ssn_homephone_0	ssn_name_unique_count_for_ssn_7
dob_homephone_count_0	name_dob_unique_count_for_ssn_homephone_1	ssn_name_unique_count_for_ssn_14
dob_homephone_count_1	name_dob_unique_count_for_ssn_homephone_3	ssn_name_unique_count_for_ssn_30
dob_homephone_count_3	name_dob_unique_count_for_ssn_homephone_7	ssn_name_unique_count_for_name_dob_0
dob_homephone_count_7	name_dob_unique_count_for_ssn_homephone_14	ssn_name_unique_count_for_name_dob_1
dob_homephone_count_14	name_dob_unique_count_for_ssn_homephone_30	ssn_name_unique_count_for_name_dob_3
dob_homephone_count_30	name_dob_unique_count_for_ssn_name_0	ssn_name_unique_count_for_name_dob_7
dob_homephone_count_0_by_3	name_dob_unique_count_for_ssn_name_1	ssn_name_unique_count_for_name_dob_14
dob_homephone_count_0_by_7	name_dob_unique_count_for_ssn_name_3	ssn_name_unique_count_for_name_dob_30
dob_homephone_count_0_by_14	name_dob_unique_count_for_ssn_name_7	ssn_name_unique_count_for_name_fulladdress_0
dob_homephone_count_0_by_30	name_dob_unique_count_for_ssn_name_14	ssn_name_unique_count_for_name_fulladdress_1
dob_homephone_count_1_by_3	name_dob_unique_count_for_ssn_name_30	ssn_name_unique_count_for_name_fulladdress_3
dob_homephone_count_1_by_7	name_dob_unique_count_for_ssn_fulladdress_0	ssn_name_unique_count_for_name_fulladdress_7
dob_homephone_count_1_by_14	name_dob_unique_count_for_ssn_fulladdress_1	ssn_name_unique_count_for_name_fulladdress_14
dob_homephone_count_1_by_30	name_dob_unique_count_for_ssn_fulladdress_3	ssn_name_unique_count_for_name_fulladdress_30
name_homephone_dob_day_sin	name_dob_unique_count_for_ss	ssn_name_unique_count_for_na

ce	n_fulladdress_7	me_homephone_0
name_homephone_dob_count_0	name_dob_unique_count_for_ssn_fulladdress_14	ssn_name_unique_count_for_name_homephone_1
name_homephone_dob_count_1	name_dob_unique_count_for_ssn_fulladdress_30	ssn_name_unique_count_for_name_homephone_3
name_homephone_dob_count_3	name_fulladdress_unique_count_for_ssn_0	ssn_name_unique_count_for_name_homephone_7
name_homephone_dob_count_7	name_fulladdress_unique_count_for_ssn_1	ssn_name_unique_count_for_name_homephone_14
name_homephone_dob_count_14	name_fulladdress_unique_count_for_ssn_3	ssn_name_unique_count_for_name_homephone_30
name_homephone_dob_count_30	name_fulladdress_unique_count_for_ssn_7	ssn_name_unique_count_for_fulladdress_dob_0
name_homephone_dob_count_0_by_3	name_fulladdress_unique_count_for_ssn_14	ssn_name_unique_count_for_fulladdress_dob_1
name_homephone_dob_count_0_by_7	name_fulladdress_unique_count_for_ssn_30	ssn_name_unique_count_for_fulladdress_dob_3
name_homephone_dob_count_0_by_14	name_fulladdress_unique_count_for_name_dob_0	ssn_name_unique_count_for_fulladdress_dob_7
name_homephone_dob_count_0_by_30	name_fulladdress_unique_count_for_name_dob_1	ssn_name_unique_count_for_fulladdress_dob_14
name_homephone_dob_count_1_by_3	name_fulladdress_unique_count_for_name_dob_3	ssn_name_unique_count_for_fulladdress_dob_30
name_homephone_dob_count_1_by_7	name_fulladdress_unique_count_for_name_dob_7	ssn_name_unique_count_for_dob_homephone_0
name_homephone_dob_count_1_by_14	name_fulladdress_unique_count_for_name_dob_14	ssn_name_unique_count_for_dob_homephone_1
name_homephone_dob_count_1_by_30	name_fulladdress_unique_count_for_name_dob_30	ssn_name_unique_count_for_dob_homephone_3
name_fulladdress_dob_day_since	name_fulladdress_unique_count_for_name_homephone_0	ssn_name_unique_count_for_dob_homephone_7
name_fulladdress_dob_count_0	name_fulladdress_unique_count_for_name_homephone_1	ssn_name_unique_count_for_dob_homephone_14
name_fulladdress_dob_count_1	name_fulladdress_unique_count_for_name_homephone_3	ssn_name_unique_count_for_dob_homephone_30
name_fulladdress_dob_count_3	name_fulladdress_unique_count_for_name_homephone_7	ssn_name_unique_count_for_ssn_dob_0

name_fulladdress_dob_count_7	name_fulladdress_unique_count_for_name_homephone_14	ssn_name_unique_count_for_ssn_dob_1
name_fulladdress_dob_count_14	name_fulladdress_unique_count_for_name_homephone_30	ssn_name_unique_count_for_ssn_dob_3
name_fulladdress_dob_count_30	name_fulladdress_unique_count_for_fulladdress_dob_0	ssn_name_unique_count_for_ssn_dob_7
name_fulladdress_dob_count_0_by_3	name_fulladdress_unique_count_for_fulladdress_dob_1	ssn_name_unique_count_for_ssn_dob_14
name_fulladdress_dob_count_0_by_7	name_fulladdress_unique_count_for_fulladdress_dob_3	ssn_name_unique_count_for_ssn_dob_30
name_fulladdress_dob_count_0_by_14	name_fulladdress_unique_count_for_fulladdress_dob_7	ssn_name_unique_count_for_ssn_homephone_0
name_fulladdress_dob_count_0_by_30	name_fulladdress_unique_count_for_fulladdress_dob_14	ssn_name_unique_count_for_ssn_homephone_1
name_fulladdress_dob_count_1_by_3	name_fulladdress_unique_count_for_fulladdress_dob_30	ssn_name_unique_count_for_ssn_homephone_3
name_fulladdress_dob_count_1_by_7	name_fulladdress_unique_count_for_dob_homephone_0	ssn_name_unique_count_for_ssn_homephone_7
name_fulladdress_dob_count_1_by_14	name_fulladdress_unique_count_for_dob_homephone_1	ssn_name_unique_count_for_ssn_homephone_14
name_fulladdress_dob_count_1_by_30	name_fulladdress_unique_count_for_dob_homephone_3	ssn_name_unique_count_for_ssn_homephone_30
ssn_firstname_day_since	name_fulladdress_unique_count_for_dob_homephone_7	ssn_name_unique_count_for_ssn_fulladdress_0
ssn_firstname_count_0	name_fulladdress_unique_count_for_dob_homephone_14	ssn_name_unique_count_for_ssn_fulladdress_1
ssn_firstname_count_1	name_fulladdress_unique_count_for_dob_homephone_30	ssn_name_unique_count_for_ssn_fulladdress_3
ssn_firstname_count_3	name_fulladdress_unique_count_for_ssn_dob_0	ssn_name_unique_count_for_ssn_fulladdress_7
ssn_firstname_count_7	name_fulladdress_unique_count_for_ssn_dob_1	ssn_name_unique_count_for_ssn_fulladdress_14
ssn_firstname_count_14	name_fulladdress_unique_count_for_ssn_dob_3	ssn_name_unique_count_for_ssn_fulladdress_30
ssn_firstname_count_30	name_fulladdress_unique_count_for_ssn_dob_7	ssn_fulladdress_unique_count_for_ssn_0
ssn_firstname_count_0_by_3	name_fulladdress_unique_count	ssn_fulladdress_unique_count_f

	_for_ssn_dob_14	or_ssn_1
ssn_firstname_count_0_by_7	name_fulladdress_unique_count_for_ssn_dob_30	ssn_fulladdress_unique_count_for_ssn_3
ssn_firstname_count_0_by_14	name_fulladdress_unique_count_for_ssn_homephone_0	ssn_fulladdress_unique_count_for_ssn_7
ssn_firstname_count_0_by_30	name_fulladdress_unique_count_for_ssn_homephone_1	ssn_fulladdress_unique_count_for_ssn_14
ssn_firstname_count_1_by_3	name_fulladdress_unique_count_for_ssn_homephone_3	ssn_fulladdress_unique_count_for_ssn_30
ssn_firstname_count_1_by_7	name_fulladdress_unique_count_for_ssn_homephone_7	ssn_fulladdress_unique_count_for_name_dob_0
ssn_firstname_count_1_by_14	name_fulladdress_unique_count_for_ssn_homephone_14	ssn_fulladdress_unique_count_for_name_dob_1
ssn_firstname_count_1_by_30	name_fulladdress_unique_count_for_ssn_homephone_30	ssn_fulladdress_unique_count_for_name_dob_3
ssn_lastname_day_since	name_fulladdress_unique_count_for_ssn_name_0	ssn_fulladdress_unique_count_for_name_dob_7
ssn_lastname_count_0	name_fulladdress_unique_count_for_ssn_name_1	ssn_fulladdress_unique_count_for_name_dob_14
ssn_lastname_count_1	name_fulladdress_unique_count_for_ssn_name_3	ssn_fulladdress_unique_count_for_name_dob_30
ssn_lastname_count_3	name_fulladdress_unique_count_for_ssn_name_7	ssn_fulladdress_unique_count_for_name_fulladdress_0
ssn_lastname_count_7	name_fulladdress_unique_count_for_ssn_name_14	ssn_fulladdress_unique_count_for_name_fulladdress_1
ssn_lastname_count_14	name_fulladdress_unique_count_for_ssn_name_30	ssn_fulladdress_unique_count_for_name_fulladdress_3
ssn_lastname_count_30	name_fulladdress_unique_count_for_ssn_fulladdress_0	ssn_fulladdress_unique_count_for_name_fulladdress_7
ssn_lastname_count_0_by_3	name_fulladdress_unique_count_for_ssn_fulladdress_1	ssn_fulladdress_unique_count_for_name_fulladdress_14
ssn_lastname_count_0_by_7	name_fulladdress_unique_count_for_ssn_fulladdress_3	ssn_fulladdress_unique_count_for_name_fulladdress_30
ssn_lastname_count_0_by_14	name_fulladdress_unique_count_for_ssn_fulladdress_7	ssn_fulladdress_unique_count_for_name_homephone_0
ssn_lastname_count_0_by_30	name_fulladdress_unique_count_for_ssn_fulladdress_14	ssn_fulladdress_unique_count_for_name_homephone_1

ssn_lastname_count_1_by_3	name_fulladdress_unique_count_for_ssn_fulladdress_30	ssn_fulladdress_unique_count_for_name_homephone_3
ssn_lastname_count_1_by_7	name_homephone_unique_count_for_ssn_0	ssn_fulladdress_unique_count_for_name_homephone_7
ssn_lastname_count_1_by_14	name_homephone_unique_count_for_ssn_1	ssn_fulladdress_unique_count_for_name_homephone_14
ssn_lastname_count_1_by_30	name_homephone_unique_count_for_ssn_3	ssn_fulladdress_unique_count_for_name_homephone_30
ssn_address_day_since	name_homephone_unique_count_for_ssn_7	ssn_fulladdress_unique_count_for_fulladdress_dob_0
ssn_address_count_0	name_homephone_unique_count_for_ssn_14	ssn_fulladdress_unique_count_for_fulladdress_dob_1
ssn_address_count_1	name_homephone_unique_count_for_ssn_30	ssn_fulladdress_unique_count_for_fulladdress_dob_3
ssn_address_count_3	name_homephone_unique_count_for_name_dob_0	ssn_fulladdress_unique_count_for_fulladdress_dob_7
ssn_address_count_7	name_homephone_unique_count_for_name_dob_1	ssn_fulladdress_unique_count_for_fulladdress_dob_14
ssn_address_count_14	name_homephone_unique_count_for_name_dob_3	ssn_fulladdress_unique_count_for_fulladdress_dob_30
ssn_address_count_30	name_homephone_unique_count_for_name_dob_7	ssn_fulladdress_unique_count_for_dob_homephone_0
ssn_address_count_0_by_3	name_homephone_unique_count_for_name_dob_14	ssn_fulladdress_unique_count_for_dob_homephone_1
ssn_address_count_0_by_7	name_homephone_unique_count_for_name_dob_30	ssn_fulladdress_unique_count_for_dob_homephone_3
ssn_address_count_0_by_14	name_homephone_unique_count_for_name_fulladdress_0	ssn_fulladdress_unique_count_for_dob_homephone_7
ssn_address_count_0_by_30	name_homephone_unique_count_for_name_fulladdress_1	ssn_fulladdress_unique_count_for_dob_homephone_14
ssn_address_count_1_by_3	name_homephone_unique_count_for_name_fulladdress_3	ssn_fulladdress_unique_count_for_dob_homephone_30
ssn_address_count_1_by_7	name_homephone_unique_count_for_name_fulladdress_7	ssn_fulladdress_unique_count_for_ssn_dob_0
ssn_address_count_1_by_14	name_homephone_unique_count_for_name_fulladdress_14	ssn_fulladdress_unique_count_for_ssn_dob_1
ssn_address_count_1_by_30	name_homephone_unique_count	ssn_fulladdress_unique_count_f

	_for_name_fulladdress_30	or_ssn_dob_3
ssn_zip5_day_since	name_homephone_unique_count_for_fulladdress_dob_0	ssn_fulladdress_unique_count_for_ssn_dob_7
ssn_zip5_count_0	name_homephone_unique_count_for_fulladdress_dob_1	ssn_fulladdress_unique_count_for_ssn_dob_14
ssn_zip5_count_1	name_homephone_unique_count_for_fulladdress_dob_3	ssn_fulladdress_unique_count_for_ssn_dob_30
ssn_zip5_count_3	name_homephone_unique_count_for_fulladdress_dob_7	ssn_fulladdress_unique_count_for_ssn_homephone_0
ssn_zip5_count_7	name_homephone_unique_count_for_fulladdress_dob_14	ssn_fulladdress_unique_count_for_ssn_homephone_1
ssn_zip5_count_14	name_homephone_unique_count_for_fulladdress_dob_30	ssn_fulladdress_unique_count_for_ssn_homephone_3
ssn_zip5_count_30	name_homephone_unique_count_for_dob_homephone_0	ssn_fulladdress_unique_count_for_ssn_homephone_7
ssn_zip5_count_0_by_3	name_homephone_unique_count_for_dob_homephone_1	ssn_fulladdress_unique_count_for_ssn_homephone_14
ssn_zip5_count_0_by_7	name_homephone_unique_count_for_dob_homephone_3	ssn_fulladdress_unique_count_for_ssn_homephone_30
ssn_zip5_count_0_by_14	name_homephone_unique_count_for_dob_homephone_7	ssn_fulladdress_unique_count_for_ssn_name_0
ssn_zip5_count_0_by_30	name_homephone_unique_count_for_dob_homephone_14	ssn_fulladdress_unique_count_for_ssn_name_1
ssn_zip5_count_1_by_3	name_homephone_unique_count_for_dob_homephone_30	ssn_fulladdress_unique_count_for_ssn_name_3
ssn_zip5_count_1_by_7	name_homephone_unique_count_for_ssn_dob_0	ssn_fulladdress_unique_count_for_ssn_name_7
ssn_zip5_count_1_by_14	name_homephone_unique_count_for_ssn_dob_1	ssn_fulladdress_unique_count_for_ssn_name_14
ssn_zip5_count_1_by_30	name_homephone_unique_count_for_ssn_dob_3	ssn_fulladdress_unique_count_for_ssn_name_30
ssn_dob_day_since	name_homephone_unique_count_for_ssn_dob_7	ssn_for_name_dob_day_since
ssn_dob_count_0	name_homephone_unique_count_for_ssn_dob_14	ssn_for_name_fulladdress_day_since
ssn_dob_count_1	name_homephone_unique_count_for_ssn_dob_30	ssn_for_name_homephone_day_since

ssn_dob_count_3	name_homephone_unique_count_for_ssn_homephone_0	ssn_for_fulladdress_dob_day_since
ssn_dob_count_7	name_homephone_unique_count_for_ssn_homephone_1	ssn_for_dob_homephone_day_since
ssn_dob_count_14	name_homephone_unique_count_for_ssn_homephone_3	ssn_for_ssn_dob_day_since
ssn_dob_count_30	name_homephone_unique_count_for_ssn_homephone_7	ssn_for_ssn_homephone_day_since
ssn_dob_count_0_by_3	name_homephone_unique_count_for_ssn_homephone_14	ssn_for_ssn_name_day_since
ssn_dob_count_0_by_7	name_homephone_unique_count_for_ssn_homephone_30	ssn_for_ssn_fulladdress_day_since
ssn_dob_count_0_by_14	name_homephone_unique_count_for_ssn_name_0	name_dob_for_ssn_day_since
ssn_dob_count_0_by_30	name_homephone_unique_count_for_ssn_name_1	name_dob_for_name_fulladdress_day_since
ssn_dob_count_1_by_3	name_homephone_unique_count_for_ssn_name_3	name_dob_for_name_homephone_day_since
ssn_dob_count_1_by_7	name_homephone_unique_count_for_ssn_name_7	name_dob_for_fulladdress_dob_day_since
ssn_dob_count_1_by_14	name_homephone_unique_count_for_ssn_name_14	name_dob_for_dob_homephone_day_since
ssn_dob_count_1_by_30	name_homephone_unique_count_for_ssn_name_30	name_dob_for_ssn_dob_day_since
ssn_homephone_day_since	name_homephone_unique_count_for_ssn_fulladdress_0	name_dob_for_ssn_homephone_day_since
ssn_homephone_count_0	name_homephone_unique_count_for_ssn_fulladdress_1	name_dob_for_ssn_name_day_since
ssn_homephone_count_1	name_homephone_unique_count_for_ssn_fulladdress_3	name_dob_for_ssn_fulladdress_day_since
ssn_homephone_count_3	name_homephone_unique_count_for_ssn_fulladdress_7	name_fulladdress_for_ssn_day_since
ssn_homephone_count_7	name_homephone_unique_count_for_ssn_fulladdress_14	name_fulladdress_for_name_dob_day_since
ssn_homephone_count_14	name_homephone_unique_count_for_ssn_fulladdress_30	name_fulladdress_for_name_homephone_day_since
ssn_homephone_count_30	fulladdress_dob_unique_count_f	name_fulladdress_for_fulladdres

	or_ssn_0	s_dob_day_since
ssn_homephone_count_0_by_3	fulladdress_dob_unique_count_for_ssn_1	name_fulladdress_for_dob_homephone_day_since
ssn_homephone_count_0_by_7	fulladdress_dob_unique_count_for_ssn_3	name_fulladdress_for_ssn_dob_day_since
ssn_homephone_count_0_by_14	fulladdress_dob_unique_count_for_ssn_7	name_fulladdress_for_ssn_homephone_day_since
ssn_homephone_count_0_by_30	fulladdress_dob_unique_count_for_ssn_14	name_fulladdress_for_ssn_name_day_since
ssn_homephone_count_1_by_3	fulladdress_dob_unique_count_for_ssn_30	name_fulladdress_for_ssn_fulladdress_day_since
ssn_homephone_count_1_by_7	fulladdress_dob_unique_count_for_name_dob_0	name_homephone_for_ssn_day_since
ssn_homephone_count_1_by_14	fulladdress_dob_unique_count_for_name_dob_1	name_homephone_for_name_dob_day_since
ssn_homephone_count_1_by_30	fulladdress_dob_unique_count_for_name_dob_3	name_homephone_for_name_fulladdress_day_since
ssn_name_day_since	fulladdress_dob_unique_count_for_name_dob_7	name_homephone_for_fulladdress_dob_day_since
ssn_name_count_0	fulladdress_dob_unique_count_for_name_dob_14	name_homephone_for_dob_homephone_day_since
ssn_name_count_1	fulladdress_dob_unique_count_for_name_dob_30	name_homephone_for_ssn_dob_day_since
ssn_name_count_3	fulladdress_dob_unique_count_for_name_fulladdress_0	name_homephone_for_ssn_homephone_day_since
ssn_name_count_7	fulladdress_dob_unique_count_for_name_fulladdress_1	name_homephone_for_ssn_name_day_since
ssn_name_count_14	fulladdress_dob_unique_count_for_name_fulladdress_3	name_homephone_for_ssn_fulladdress_day_since
ssn_name_count_30	fulladdress_dob_unique_count_for_name_fulladdress_7	fulladdress_dob_for_ssn_day_since
ssn_name_count_0_by_3	fulladdress_dob_unique_count_for_name_fulladdress_14	fulladdress_dob_for_name_dob_day_since
ssn_name_count_0_by_7	fulladdress_dob_unique_count_for_name_fulladdress_30	fulladdress_dob_for_name_fulladdress_day_since
ssn_name_count_0_by_14	fulladdress_dob_unique_count_for_name_homephone_0	fulladdress_dob_for_name_homephone_day_since

ssn_name_count_0_by_30	fulladdress_dob_unique_count_for_name_homephone_1	fulladdress_dob_for_dob_homephone_day_since
ssn_name_count_1_by_3	fulladdress_dob_unique_count_for_name_homephone_3	fulladdress_dob_for_ssn_dob_day_since
ssn_name_count_1_by_7	fulladdress_dob_unique_count_for_name_homephone_7	fulladdress_dob_for_ssn_homephone_day_since
ssn_name_count_1_by_14	fulladdress_dob_unique_count_for_name_homephone_14	fulladdress_dob_for_ssn_name_day_since
ssn_name_count_1_by_30	fulladdress_dob_unique_count_for_name_homephone_30	fulladdress_dob_for_ssn_fulladdress_day_since
ssn_fulladdress_day_since	fulladdress_dob_unique_count_for_dob_homephone_0	dob_homephone_for_ssn_day_since
ssn_fulladdress_count_0	fulladdress_dob_unique_count_for_dob_homephone_1	dob_homephone_for_name_dob_day_since
ssn_fulladdress_count_1	fulladdress_dob_unique_count_for_dob_homephone_3	dob_homephone_for_name_fulladdress_day_since
ssn_fulladdress_count_3	fulladdress_dob_unique_count_for_dob_homephone_7	dob_homephone_for_name_homephone_day_since
ssn_fulladdress_count_7	fulladdress_dob_unique_count_for_dob_homephone_14	dob_homephone_for_fulladdress_dob_day_since
ssn_fulladdress_count_14	fulladdress_dob_unique_count_for_dob_homephone_30	dob_homephone_for_ssn_dob_day_since
ssn_fulladdress_count_30	fulladdress_dob_unique_count_for_ssn_dob_0	dob_homephone_for_ssn_homephone_day_since
ssn_fulladdress_count_0_by_3	fulladdress_dob_unique_count_for_ssn_dob_1	dob_homephone_for_ssn_name_day_since
ssn_fulladdress_count_0_by_7	fulladdress_dob_unique_count_for_ssn_dob_3	dob_homephone_for_ssn_fulladdress_day_since
ssn_fulladdress_count_0_by_14	fulladdress_dob_unique_count_for_ssn_dob_7	ssn_dob_for_ssn_day_since
ssn_fulladdress_count_0_by_30	fulladdress_dob_unique_count_for_ssn_dob_14	ssn_dob_for_name_dob_day_since
ssn_fulladdress_count_1_by_3	fulladdress_dob_unique_count_for_ssn_dob_30	ssn_dob_for_name_fulladdress_day_since
ssn_fulladdress_count_1_by_7	fulladdress_dob_unique_count_for_ssn_homephone_0	ssn_dob_for_name_homephone_day_since
ssn_fulladdress_count_1_by_14	fulladdress_dob_unique_count_f	ssn_dob_for_fulladdress_dob_da

	or_ssn_homephone_1	y_since
ssn_fulladdress_count_1_by_30	fulladdress_dob_unique_count_for_ssn_homephone_3	ssn_dob_for_dob_homephone_day_since
ssn_name_dob_day_since	fulladdress_dob_unique_count_for_ssn_homephone_7	ssn_dob_for_ssn_homephone_day_since
ssn_name_dob_count_0	fulladdress_dob_unique_count_for_ssn_homephone_14	ssn_dob_for_ssn_name_day_since
ssn_name_dob_count_1	fulladdress_dob_unique_count_for_ssn_homephone_30	ssn_dob_for_ssn_fulladdress_day_since
ssn_name_dob_count_3	fulladdress_dob_unique_count_for_ssn_name_0	ssn_homephone_for_ssn_day_since
ssn_name_dob_count_7	fulladdress_dob_unique_count_for_ssn_name_1	ssn_homephone_for_name_dob_day_since
ssn_name_dob_count_14	fulladdress_dob_unique_count_for_ssn_name_3	ssn_homephone_for_name_fulladdress_day_since
ssn_name_dob_count_30	fulladdress_dob_unique_count_for_ssn_name_7	ssn_homephone_for_name_homephone_day_since
ssn_name_dob_count_0_by_3	fulladdress_dob_unique_count_for_ssn_name_14	ssn_homephone_for_fulladdress_dob_day_since
ssn_name_dob_count_0_by_7	fulladdress_dob_unique_count_for_ssn_name_30	ssn_homephone_for_dob_homephone_day_since
ssn_name_dob_count_0_by_14	fulladdress_dob_unique_count_for_ssn_fulladdress_0	ssn_homephone_for_ssn_dob_day_since
ssn_name_dob_count_0_by_30	fulladdress_dob_unique_count_for_ssn_fulladdress_1	ssn_homephone_for_ssn_name_day_since
ssn_name_dob_count_1_by_3	fulladdress_dob_unique_count_for_ssn_fulladdress_3	ssn_homephone_for_ssn_fulladdress_day_since
ssn_name_dob_count_1_by_7	fulladdress_dob_unique_count_for_ssn_fulladdress_7	ssn_name_for_ssn_day_since
ssn_name_dob_count_1_by_14	fulladdress_dob_unique_count_for_ssn_fulladdress_14	ssn_name_for_name_dob_day_since
ssn_name_dob_count_1_by_30	fulladdress_dob_unique_count_for_ssn_fulladdress_30	ssn_name_for_name_fulladdress_day_since
ssn_name_fulladdress_day_since	dob_homephone_unique_count_for_ssn_0	ssn_name_for_name_homephone_day_since
ssn_name_fulladdress_count_0	dob_homephone_unique_count_for_ssn_1	ssn_name_for_fulladdress_dob_day_since

ssn_name_fulladdress_count_1	dob_homephone_unique_count_for_ssn_3	ssn_name_for_dob_homephone_day_since
ssn_name_fulladdress_count_3	dob_homephone_unique_count_for_ssn_7	ssn_name_for_ssn_dob_day_since
ssn_name_fulladdress_count_7	dob_homephone_unique_count_for_ssn_14	ssn_name_for_ssn_homephone_day_since
ssn_name_fulladdress_count_14	dob_homephone_unique_count_for_ssn_30	ssn_name_for_ssn_fulladdress_day_since
ssn_name_fulladdress_count_30	dob_homephone_unique_count_for_name_dob_0	ssn_fulladdress_for_ssn_day_since
ssn_name_fulladdress_count_0_by_3	dob_homephone_unique_count_for_name_dob_1	ssn_fulladdress_for_name_dob_day_since
ssn_name_fulladdress_count_0_by_7	dob_homephone_unique_count_for_name_dob_3	ssn_fulladdress_for_name_fulladdress_day_since
ssn_name_fulladdress_count_0_by_14	dob_homephone_unique_count_for_name_dob_7	ssn_fulladdress_for_name_homephone_day_since
ssn_name_fulladdress_count_0_by_30	dob_homephone_unique_count_for_name_dob_14	ssn_fulladdress_for_fulladdress_dob_day_since
ssn_name_fulladdress_count_1_by_3	dob_homephone_unique_count_for_name_dob_30	ssn_fulladdress_for_dob_homephone_day_since
ssn_name_fulladdress_count_1_by_7	dob_homephone_unique_count_for_name_fulladdress_0	ssn_fulladdress_for_ssn_dob_day_since
ssn_name_fulladdress_count_1_by_14	dob_homephone_unique_count_for_name_fulladdress_1	ssn_fulladdress_for_ssn_homephone_day_since
ssn_name_fulladdress_count_1_by_30	dob_homephone_unique_count_for_name_fulladdress_3	ssn_fulladdress_for_ssn_name_day_since