Application (Identity) Fraud Analysis

Team: Adreeta Guha
Harsh Bhuta
Karan Saxena
Naman Jain
Samanyu Raju Kakarlapudi

Table of Contents

Executive Summary……………………………………………………………………………………3

Description of Data……………………………………………………………………………………4

Data Cleaning……………………………………………………………………………………………7

Candidate Variables…………………………………………………………………………….………...8

Feature Selection Process………………………………………………………………………….14

Model Algorithms……………………………………………………………………………………..17

Results…………………………………………………………………………………………………….25

Conclusion……………………………………………………………………………………………..26

Appendix……………………………………………………………………………………………..27

Executive Summary:

The dataset provided contains credit card application fraud data. The goal was to identify synthetic identity fraud among a million personal information records. The dataset had 1,000,000 records and 10 data fields which contained several important information about the applicants such as the first name, last name, date of birth, social security number (SSN), address, and phone number. Our objective was to create a supervised fraud detection model with a high fraud detection rate.

We performed exploratory data analysis on each of the 10 fields which were identified as categorical. Summary statistics of each field in terms of % populated, Unique values, and most common values were described. This helped us identify the data better and then during the data cleaning process, we were able to identify the frivolous and missing values and treat them at the start to avoid faulty predictions.

To get a better insight of the data we engineered 1948 candidate variables which would help us identify fraudulent behaviors. The original categorical fields were concatenated to create different field combinations. Then for each combination, we calculated three more variable groups: the velocity candidate variables, the days-since candidate variables, and the relative velocity candidate variables.

Feature selection was then performed using select filter and wrapper methods to identify the top 30 candidate variables as a part of minimizing dimensionality. In the filtering process, Kolmogorov-Smirnov (KS) and fraud detection rate (FDR) at 3% were used to eliminate variables. For our wrapper method, we used recursive feature elimination with cross-validation. Once the final variables were chosen, the data was divided in three sections: training, testing, and out-of-time. The out-of-time section represented the last two months of data. In addition, the first two weeks of the data was omitted to get the most accurate results possible as there was little to no data prior to each datapoint.

To find the best model, the variables were tested in several different models. First, since we had a supervised binary classification problem, logistic regression was used as a base model to predict fraud. Logistic regression caught 48.3% of fraud at a 3% FDR. Next, nonlinear models such as random forest, boosted trees, and neural network were used to get better results than the base model. Parameter tuning was also performed on each model to get better results. Overall, our best model was random forest trees which caught 53.6% of the fraudulent records in the testing set and 50.5 % of the fraudulent records in the validation set at a 3% FDR after parameter tuning.

Description of the Data:

Dataset Name: Applications Data

Dataset Description: The dataset is a synthetic dataset originally created for academic organizations that were conducting research in collaboration with ID Analytics (https://www.idanalytics.com/). Each record in the dataset represents information on an application filed for a credit card. The dataset also contains a field representing the date on which the credit card application was made, as well as Personal Identity Information fields like SSN, name, address, phone number, date of birth, zip code and the fraud label which tells us that the application is fraudulent if the value is 1 and not fraudulent if the value is 0.

Fields: 10

Records: 1,000,000

Time Period: 1st January 2016 - 31st December 2016

Summary Table:

Field Name	% Populated	# Unique Values	Most Common Value
record	100	1,000,000	NA
date	100	365	8/16/2016
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
dob	100	42,673	6/26/1907
homephone	100	28,244	9999999999
fraud_label	100	2	0

Table 1: Summary Table

From the above fields, we determined that ‘ssn’, ‘address’, homephone and ‘dob’ are the most important fields in the given dataset. Few of the relevant depictions amongst the critical data fields are provided below.

1. **ssn:**
   The ssn field refers to the social security number of the applicants. There are 835,819 unique values. There are 16,935 records with the most common value being 999999999. The chart below indicates the top 15 most common SSNs used by applicants.

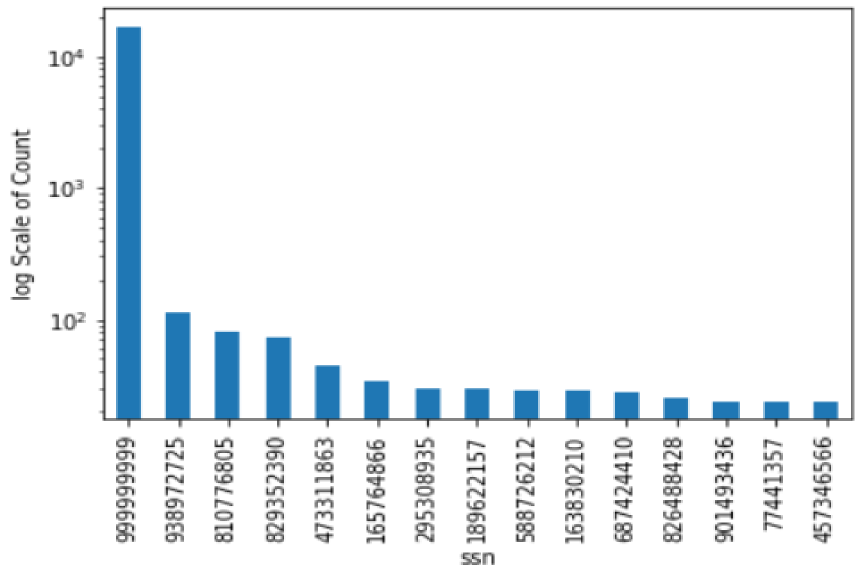


Fig 1. Bar plot displaying distribution of field 'ssn'

2. **address:**
   The “address” field refers to the street address of the applicants. There are 828,774 unique addresses, out of which 123 MAIN ST is the most common which is repeated 1079 times. The chart below indicates the top 15 most common addresses used by applicants.

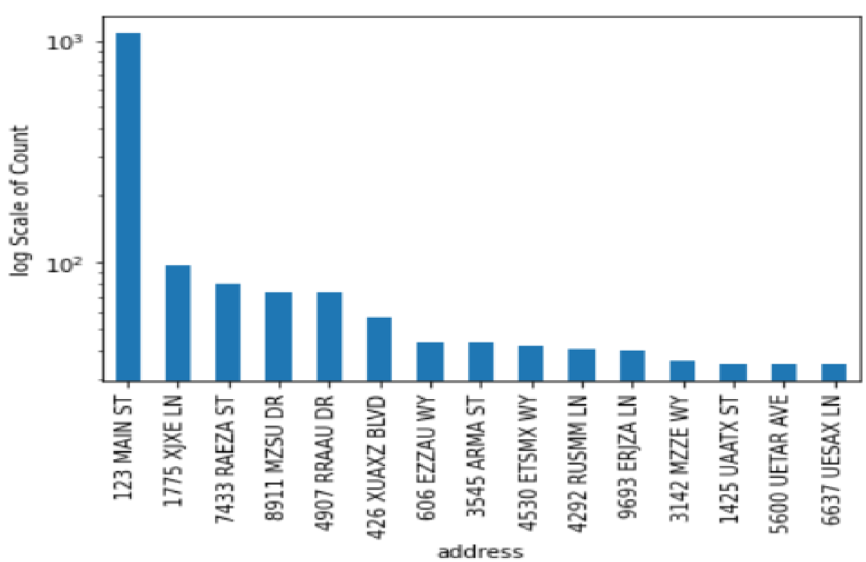


Fig 2. Bar plot displaying distribution of addresses

3. **Homephone:**
The field “homephone” refers to the phone number of the applicant. There are total of 28,244 unique phone numbers among 1 million credit card applications. The most common phone number being ‘9999999999’.

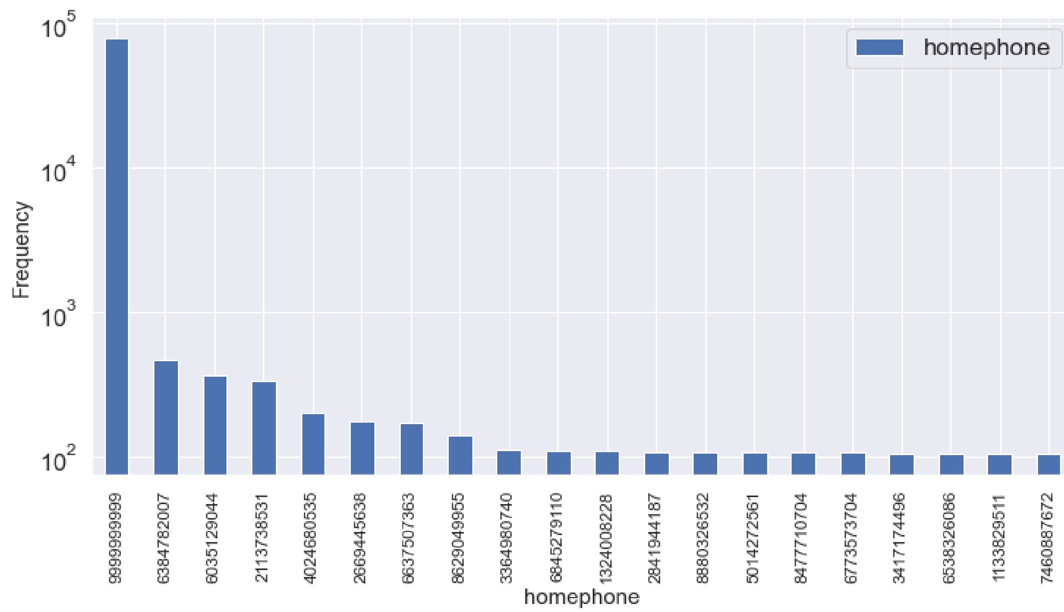


Fig 3. Bar Plot showing distribution of homephone

4. **Dob:**
The field “dob” refers to the date of birth of the applicant. There are a total of 42,673 unique date of births among 1 million credit card applications. The most common date of birth is ‘1907-06-26’.

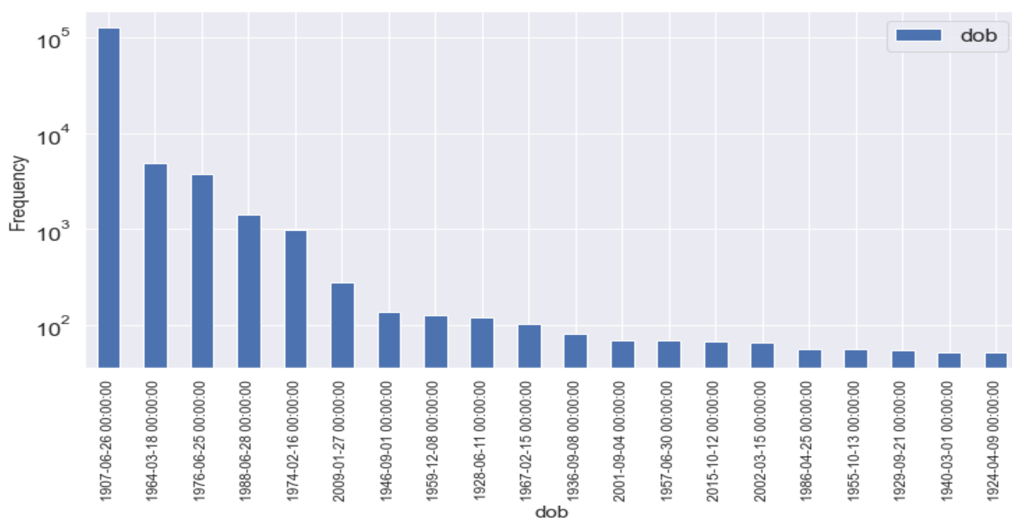


Fig 4. Bar Plot showing distribution of dob

Data Cleaning:

- **Missing field values:**
  The applications dataset did not have any missing values in any of its fields. All fields were 100% populated. Therefore, no action was required.

- **Frivolous Values:**
  We saw that all the data fields were 100% populated; however, we found that four of the data fields contained frivolous values.

Data Field	Frivolous Value	Number of Records
ssn	999999999	16,935
address	123 MAIN ST	1,079
dob	19070626	126,568
homephone	9999999999	78,512

Table 2: Frivolous Values

Having frivolous values will give misleading results as they will generate a high-count value for those records, generating a false alarm for fraudulent activity. Hence, we replaced these values by a unique value which was the negative of the corresponding record number. For instance, if application record number 105 had a value of “999999999” in the “ssn” data field and a value of “123 MAIN ST” in the “address” data field, then the number -105 was used to replace the frivolous values in both the “ssn” and “address” data fields.

- **Leading Zeros:**
  Some fields in the data set were supposed to have a specific number of digits i.e., it should have a fixed length. For example, each value in zip5 should have 5 digits. However, some of the values in the zip5 field have 4 digits. Similarly, some of the values in SSN have less than 9 digits, DOB has less than 8 digits and Homephone has less than 10 digits. To deal with this issue we came up with an idea of left padded ‘0’ (zeroes) to make the value equal to the definite fixed length. For example, if the value for the field zip5 was “1034”, we padded one zero and made it “01034”.

Candidate Variables

For candidate predictor variables were created for more in-depth analysis and insight and to better capture fraud applications. Two types of candidate variables were built:

- velocity in which applications were seen and the
- number of days since the last time a field was seen in an application.

Prior to creating candidate variables for our analysis and models, additional categorical variables were created.

a. Categorical variables

These categorical variables would be used to create candidate predictor variables. These categorical variables were created by using business knowledge and different combinations of existing fields were made, that fraudsters would most likely use in applications. For our project, we created 18 categorical variables as shown in the table below.

	Variable Name	Description
1	name	First name + Last name
2	fulladdress	Address + Zip code
3	name_dob	Name + Date of Birth
4	name_fulladdress	Name + Full Address
5	name_homephone	Name + Phone number
6	fulladdress_dob	Full Address + Date of Birth
7	fulladdress_homephone	Full Address + Phone number
8	dob_homephone	Date of Birth + Phone number
9	homephone_name_dob	Phone number + Name + Date of Birth
10	ssn_firstname	SSN + First name
11	ssn_lastname	SSN + Last name
12	ssn_address	SSN + address
13	ssn_zip5	SSN + zip5
14	ssn_dob	SSN + Date of Birth
15	ssn_homephone	SSN + Phone number
16	ssn_name	SSN + Name
17	ssn_fulladdress	SSN + Full Address
18	ssn_name_dob	SSN + Name + Date of Birth
19	ssn_name_fulladdress	SSN + Name + Full Address
20	ssn_name_homephone	SSN + Name + Phone Number
21	ssn_fulladdress_dob	SSN + Full Address + Date of Birth
22	ssn_fulladdress_homephone	SSN + Full Address + Phone Number
23	ssn_dob_homephone	SSN + Date of Birth + Phone Number
24	ssn_homephone_name_dob	SSN + Phone Number + Name + Date of Birth

25	ssn_fulladdress_name_dob	SSN + Full Address + Name + Date of Birth
26	name_zip5	Name + Zip Code
27	name_address	Name + Address
28	firstname_dob	First Name + Date of Birth
29	firstname_address	First Name + Address
30	firstname_zip5	First Name + Zip Code
31	firstname_homephone	First Name + Phone Number
32	firstname_fulladdress	First Name + Full Address
33	lastname_dob	Last Name + Date of Birth
34	lastname_address	Last Name + Address
35	lastname_zip5	Last Name + Zip Code
36	lastname_homephone	Last Name + Phone Number
37	lastname_fulladdress	Last Name + Full Address

Table 3: Categorical Variables

b. Velocity Candidate Variables

Velocity refers to the speed or frequency at which entities (normal variables + categorical variables) were seen in the dataset for a particular application record. The speed at which these applications appear in our dataset is a way of detecting and identifying potentially fraudulent applications. A higher value of velocity would indicate a greater likelihood of a fraudulent application. For our candidate variables, we look at a velocity over 0, 1, 3, 7, 14, and 30 days. 132 velocity candidate variables were created using the formula below.

Velocity = # of records with the same entity over the last n days

n = {0, 1, 3, 7, 14, 30} days

Entity in the above formula refers to the normal variables + various categorical variables created previously (Refer Table 2). The complete list of velocity candidate variables can be found in Appendix B.

	Variable Name		Variable Name
1	ssn_count_0	13	dob_count_0
2	ssn_count_1	14	dob_count_1
3	ssn_count_3	15	dob_count_3
4	ssn_count_7	16	dob_count_7

5	ssn_count_14	17	dob_count_14
6	ssn_count_30	18	dob_count_30
7	address_count_0	19	homephone_count_0
8	address_count_1	20	homephone_count_1
9	address_count_3	21	homephone_count_3
10	address_count_7	22	homephone_count_7
11	address_count_14	23	homephone_count_14
12	address_count_30	24	homephone_count_30

Table 4: Velocity Candidate Variables

c. Days Since Candidate Variables

Days since variable refers to the number of days since the last time a 'similar' entity was seen for a particular application record. Days since returns a whole number for the number of days since last seen. If the entity occurs more than once on the same date, then 1 is returned in the days since the field of that entity for that record. If the value is small but greater than zero, it means that there is a higher likelihood of fraud. 22 candidate variables were created using the equation found below.

Days since = # of days since the entity was last seen

Entity in the above formula refers to the various categorical variables created previously.

	Variable Name		Variable Name
1	ssn_days_since	12	dob_homephone_days_since
2	address_days_since	13	homephone_name_dob_days_since
3	dob_days_since	14	ssn_firstname_days_since
4	homephone_days_since	15	ssn_lastname_days_since
5	name_days_since	16	ssn_address_days_since
6	fulladdress_days_since	17	ssn_zip5_days_since

7	name_dob_days_since	18	ssn_dob_days_since
8	name_fulladdress_days_since	19	ssn_homephone_days_since
9	name_homephone_days_since	20	ssn_name_days_since
10	fulladdress_dob_days_since	21	ssn_fulladdress_days_since
11	fulladdress_homephone_days_since	22	ssn_name_dob_days_since

Table 5: Days Since Candidate Variables

d. Relative Velocity Candidate Variables

Relative velocity refers to the speed at which an entity is seen in the dataset for a particular application record over a short period of time (0 - 1 days) in relation to how often the same entity is seen over a longer period (3 – 30 days). The speed at which these applications come through in a shorter timeframe versus a longer time frame is a way to detect and identify potentially fraudulent applications. A higher value of relative velocity would indicate a greater likelihood of a fraudulent application. For our model, we look at a relative velocity over 3, 7, 14, and 30 days. 176 candidate variables were created using the equation given below.

$$\textit{Rel Velocity} = \frac{\textit{\# of applications with entity seen in the recent past x days}}{\textit{\# of applications with the same entity seen in the recent past n days}}$$

$$x = \{0,1\} \text{ days}$$

$$n = \{3, 7, 14, 30\} \text{ days}$$

Entity in the above formula refers to the various categorical variables created previously. The complete list of relative velocity candidate variables can be found in Appendix B.

	Variable Name		Variable Name
1	ssn_count_0_by_3	12	address_count_0_by_30
2	ssn_count_0_by_7	13	address_count_1_by_3
3	ssn_count_0_by_14	14	address_count_1_by_7
4	ssn_count_0_by_30	15	address_count_1_by_14

5	ssn_count_1_by_3	16	address_count_1_by_30
6	ssn_count_1_by_7	17	dob_count_0_by_3
7	ssn_count_1_by_14	18	dob_count_0_by_7
8	ssn_count_1_by_30	19	dob_count_0_by_14
9	address_count_0_by_3	20	dob_count_0_by_30
10	address_count_0_by_7	21	dob_count_1_by_3
11	address_count_0_by_14	22	dob_count_1_by_7

Table 6: Relative Velocity Candidate Variables

e. Combination Unique Velocity Candidate Variables

Combination unique velocity refers to the speed at which one entity is seen in relation with another unique entity over a period of time. A higher value of velocity would indicate a greater likelihood of a fraudulent application. For our candidate variables, we look at a velocity over 1, 3, 7, 14, 30, and 60 days. 1440 candidate variables were created using the equation given below.

Combination Unique Velocity = # of Unique Entity2 used against Entity1 over the past n days

n = {1, 3, 7, 14, 30, 60}

Entity in the above formula refers to the various categorical variables created previously. Some of the combination unique velocity variables can be found in the table below. Complete list of combination unique velocity candidate variables can be found in Appendix B.

	Variable Name		Variable Name
1	ssn_unique_count_for_address_1	12	ssn_unique_count_for_dob_60
2	ssn_unique_count_for_address_3	13	ssn_unique_count_for_homephone_1
3	ssn_unique_count_for_address_7	14	ssn_unique_count_for_homephone_3
4	ssn_unique_count_for_address_14	15	ssn_unique_count_for_homephone_7
5	ssn_unique_count_for_address_30	16	ssn_unique_count_for_homephone_14

6	ssn_unique_count_for_address_60	17	ssn_unique_count_for_homephone_30
7	ssn_unique_count_for_dob_1	18	ssn_unique_count_for_homephone_60
8	ssn_unique_count_for_dob_3	19	ssn_unique_count_for_name_1
9	ssn_unique_count_for_dob_7	20	ssn_unique_count_for_name_3
10	ssn_unique_count_for_dob_14	21	ssn_unique_count_for_name_7
11	ssn_unique_count_for_dob_30	22	ssn_unique_count_for_name_14

Table 7: Combination Unique Velocity Candidate Variables

Feature Selection Process:

Feature selection is the process of reducing the number of input variables when developing a predictive model. It is desirable to reduce the number of input variables to both reduce the computational cost of modeling and to improve the performance of the model.
Statistical-based feature selection methods involve evaluating the relationship between each input variable and the target variable using statistics and selecting those input variables that have the strongest relationship with the target variable.

Also, features that were either highly correlated with others or not significant to perform an accurate prediction were ignored. Additionally, feature selection enhances model performance.

1. Filter Method: Filter method is generally used as a preprocessing step. The selection of features is independent of any machine learning algorithms. Instead, features are selected on the basis of their scores in various statistical tests for their correlation with the outcome variable. The correlation is a subjective term here.

Fig 5: Filtering

Filter Methods For our analysis, we performed feature selection using the Kolmogorov-Smirnov distance and the Fraud Detection Rate.

a. Kolmogorov–Smirnov (KS) Distance
We used univariate KS which works well for feature selection for binary classification. This is a filter method. For each variable, it plots separate distributions for the two populations, good and bad. With more separation between the distributions, the more important the variable is. For this analysis, we used the KS distance metric by calculating the univariate KS value as a filtering method to aid in determining which features provide a better separation between the “fraud_label” values of 1 and 0. Meaning, for each numerical candidate variable, we generated the distribution of the two classes (1 and 0) based on the dependent variable ("fraud_label” data field). Subsequently, we measured the KS distance between the distributions of the two classes for each of the numerical candidate variables. More formally, we then rank ordered the KS distance value in descending order for each of the numerical candidate variables and used this ranking to evaluate the importance of each variable.

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

b. Fraud Detection Rate (FDR) at 3%

In The second metric we used for filtering the features was the Fraud Detection Rate (FDR). As a rule, the FDR is the percentage of all the frauds that are caught up till a specific end point. With regards to this investigation in our project, we utilized an we used a cutoff threshold of 3% and calculated the univariate FDR for each numerical candidate variable. The FDR is set in stone by first arranging the numerical candidate variables in descending order, and afterward figuring the frauds in the top 3%. We then, at that point, relegated a position for every one of the numerical candidate variables and involved this positioning as a way to assess the significance of every factor.

c. Filter Method Aggregated Results
In our investigation, we acquired the univariate KS distance and the univariate FDR at 3% for each numerical candidate variables and utilized their normal position to fill in as a last score on the significance of each component. In this manner, we kept the top 33% positioned numerical candidate variables, reducing the total number of features from 634 to 217.

2. Wrapper Method: In wrapper methods, we try to use a subset of features and train a model using them. Based on the inferences that we draw from the previous model, we decide to add or remove features from your subset. The problem is essentially reduced to a search problem.

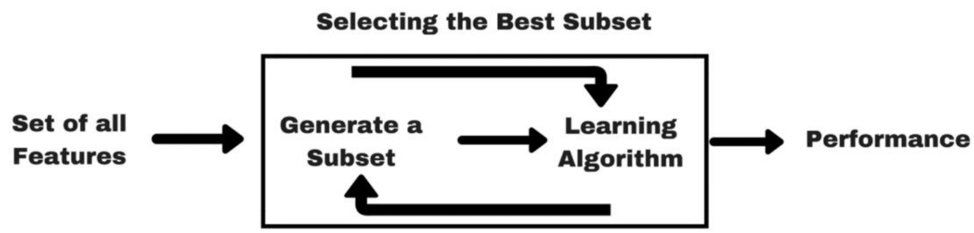


Fig 6: Wrapping

Sorted List of Variables from Forward Selection:

Rank	Variable Name	Avg_ Score
1	fulladdress_count_30	0.524986940
2	ssn_count_30	0.519502002
3	fulladdressunique_count_fordob_homephone_60	0.521330315
4	address_count_30	0.524203378
5	fulladdressunique_count_forssn_name_fulladdress_7	0.524725753
6	fulladdressunique_count_forssn_60	0.524725753
7	firstname_dob_count_30	0.524812816
8	fulladdressunique_count_forname_fulladdress_30	0.525857566

9	fulladdressunique_count_forssn_name_dob_60	0.525857566
10	fulladdressunique_count_forssn_lastname_30	0.525770503
11	fulladdressunique_count_forssn_zip5_30	0.525857566
12	fulladdress_count_0_by_30	0.525857566
13	fulladdressunique_count_forname_dob_7	0.525335191
14	fulladdressunique_count_forssn_7	0.525248128
15	fulladdressunique_count_fordob_homephone_7	0.525770503
16	fulladdressunique_count_fordob_homephone_30	0.525248128
17	fulladdressunique_count_forssn_fulladdress_30	0.525509316
18	fulladdressunique_count_forname_dob_30	0.525683441
19	fulladdressunique_count_forname_dob_60	0.525161066
20	fulladdressunique_count_forssn_name_fulladdress_30	0.525335191
21	fulladdressunique_count_forname_fulladdress_7	0.525509316
22	fulladdress_count_7	0.525683441
23	ssn_dob_count_30	0.525683441
24	lastname_dob_count_30	0.525248128
25	fulladdressunique_count_forname_dob_14	0.525161066
26	fulladdressunique_count_forssn_lastname_14	0.525248128
27	fulladdressunique_count_forssn_name_60	0.525074003
28	fulladdressunique_count_fordob_homephone_14	0.525248128
29	fulladdressunique_count_forname_fulladdress_1	0.525509316
30	fulladdressunique_count_forfulladdress_dob_60	0.524986941

Table 7: Top 30 Candidate Variables

Model Algorithms:

After selecting the top 30 best variables, with each having a wrapper score above 0.5. We start with a logistic regression to get a base line model and then test Decision Tree, Random Forest, Boosted Tree and Neural Network models with varying hyperparameters to choose the best models by comparing the Fraud Detection Rate (FDR) at 3% for the train, test and out of time (OOT) datasets. The following are the details of each test conducted.

Logistics Regression:
The logistic regression is one of the most popular classification algorithms. In logistic regression, a linear output is converted into a probability between 0 & 1 using the sigmoid function.

$$S(x) = \frac{1}{1+e^{-x}} = \frac{e^x}{e^x+1}$$

In the equation above, X is the set of predictor features and b is the corresponding vector of weights. Computing S(x) above produces a probability that indicates if an observation should be classified as “1” (if the calculated probability is at least 0.5), and “0” otherwise. It’s an S-shaped curve that can take a real-valued number and map it into a value between 0 and 1, but never

exactly at those limits.

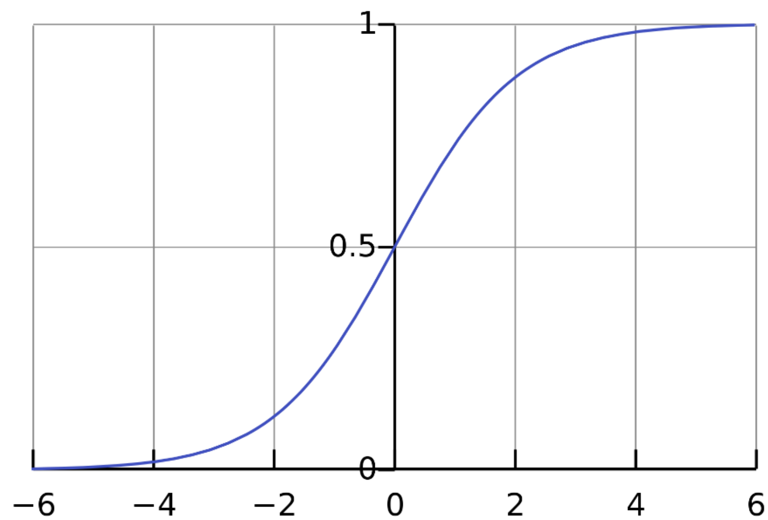


Fig 7. Logistic regression

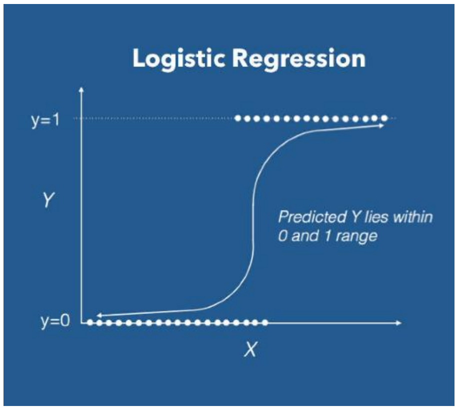


Fig 8. Sample Logistic regression curve

Below is an example logistic regression equation:

$$y = e^{\wedge}(b_0 + b_1{*}x) / (1 + e^{\wedge}(b_0 + b_1{*}x))$$

Where y is the predicted output, b_0 is the bias or intercept term and b_1 is the coefficient for the single input value (x). Each column in your input data has an associated b coefficient (a constant real value) that must be learned from your training data. The intercept term controls the location of the midpoint of the curve and b_1 controls the raise of the curve.

More simply put, here is how the Logistic Regression equation for Machine Learning looks like:

$$\text{logit}(p) = \ln(p/(1-p)) = h_0+h_1X1+h_2X_2+h_3X_3\ldots.+h_kX_k$$

p = probability of the occurrence of the feature

x_1, x_2 ,…., x_k = set of input features

h_1, h_2,….h_k = parametric values to be estimated in the Logistic Regression equation.

Syntax:

sklearn.linear_model.LogisticRegression()

For this project's fraud analysis, five versions of logistic regression were created by changing the solver, penalty and c hyperparameters and training the model with the identified 30 best variables. To understand the used hyperparameters further:

- Solver: This parameter represents which algorithm to use in the optimization problem.
  - liblinear − It is a good choice for small datasets. It also handles L1 penalty. For multiclass problems, it is limited to one-versus-rest schemes.
  - lbfgs − For multiclass problems, it handles multinomial loss. It also handles only L2 penalty.

  Default is 'lbfgs'.
- Penalty: Penalized logistic regression imposes a penalty to the logistic model for having too many variables. This results in shrinking the coefficients of the less contributive variables toward zero. This is also known as regularization. L1 is therefore useful for feature selection, as we can drop any variables associated with coefficients that go to zero. L2, on the other hand, is useful when you have collinear/codependent features. Default is 'L2'
- C: It represents the inverse of regularization strength, which must always be a positive float. Smaller values specify stronger regularization.

The results of the logistic regression are shown below.

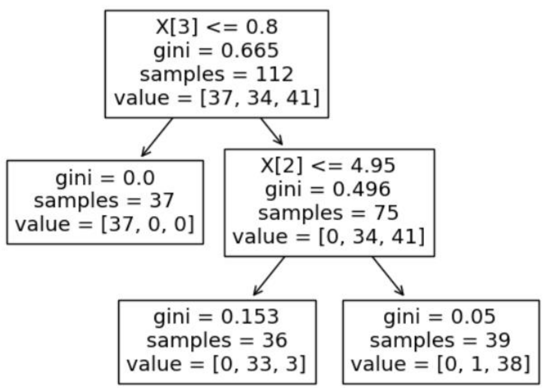
Model	Parameters			Average FDR at 3%		
	solver	penalty	C	Train	Test	OOT
Logistics Regression	liblinear	l1	1	0.497	0.491	0.483
	lbfgs	l2	1	0.491	0.493	0.479
	lbfgs	l2	0.1	0.49	0.492	0.477
	liblinear	l2	1	0.496	0.505	0.483
	lbfgs	l2	10	0.501	0.495	0.483

Table 8: Logistic Regression Model Results

The best results were given when solver = liblinear, penalty = l2, c = 1.

Decision Tree:

In decision analysis, a decision tree can be used to represent decisions and decision making visually and explicitly. The main goal of Decision Trees is to create a model predicting target variable value by learning simple decision rules deduced from the data features. Decision trees have two main entities; one is root node, where the data splits, and other is decision nodes or leaves, where we got final output. In three-dimensional view, decision trees approximate the surface into y = f(x) with steps or platforms. These steps form boxes and each box contains the average of the dependent variable y for its range.



by scikit-learn.org

Fig 9. Sample Decision Tree

The decision trees decide the cut point of these boxes by measuring the impurity of the resulting boxes to calculate the goodness of the candidate. Common measures of impurity are variance, Gini index and Entropy. Best cut point has the lowest impurity.

Syntax:

sklearn.tree.DecisionTreeClassifier()

For this project's fraud analysis, four versions of logistic regression were created by changing the max_depth, min_sample_leaf and min_samples_split hyperparameters and training the model with the identified 30 best variables. To understand the used hyperparameters further:

- max_depth: The maximum depth of the tree. If None, then nodes are expanded until all leaves are pure or until all leaves contain less than min_samples_split samples. Default is None

- min_sample_leaf: The minimum number of samples required to be at a leaf node. A split point at any depth will only be considered if it leaves at least min_samples_leaf training samples in each of the left and right branches. This may have the effect of smoothing the model, especially in regression. Default is 1.
- min_samples_split: The minimum number of samples required to split an internal node. Default is 2.

The results of the Decision Trees are shown below.

Model	Parameters			Average FDR at 3%		
	max_depth	min_sample_leaf	min_samples_split	Train	Test	OOT
Decision Tree	10	1	2	0.522	0.515	0.495
	20	60	300	0.524	0.511	0.502
	20	60	320	0.525	0.514	0.5
	25	80	600	0.516	0.533	0.498

Table 9: Decision Tree model results

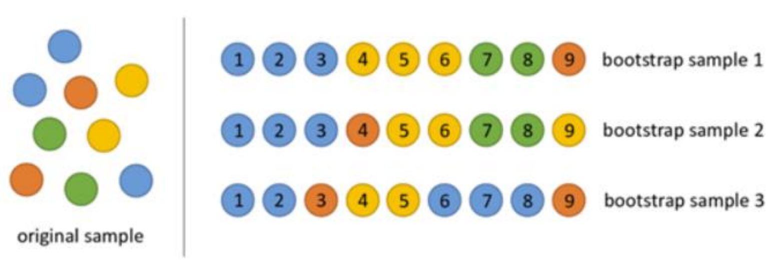
The best results were given when max_depth = 25, min_sample_leaf = 80, min_samples_split = 600.

Random Forest

Random forest is an ensemble of many decision trees. Random forests are built using a method called bagging in which each decision trees are used as parallel estimators. If used for a classification problem, the result is based on average prediction from each decision tree. For regression, the prediction of a leaf node is the mean value of the target values in that leaf. Random forest regression takes mean value of the results from decision trees.

Random forests reduce the risk of overfitting and accuracy is much higher than a single decision tree. Furthermore, decision trees in a random forest run in parallel so that the time does not become a bottleneck.

The success of a random forest highly depends on using uncorrelated decision trees. If we use same or very similar trees, overall result will not be much different than the result of a single decision tree. Random forests achieve to have uncorrelated decision trees by bootstrapping and feature randomness.



Random Forests follow the same basic principle as that of Decision Trees, however, Random Forests train multiple Decision Trees by a randomly chosen subset of variables or records for each tree and each split of the tree. It then gets the prediction from each of them and finally selects the best solution by means of voting (for classification type data) or by averaging (for regression type data) . It can be used for both classification as well as regression tasks.

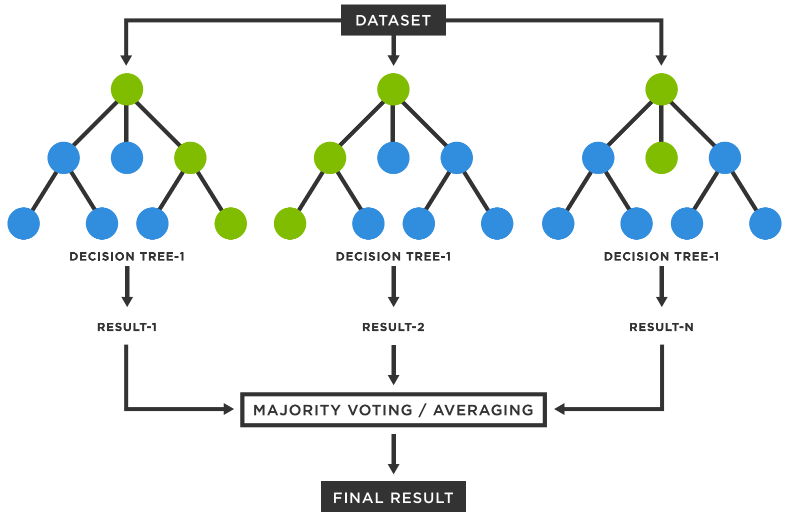


Fig 10. Sample working principle of Random Forest Model

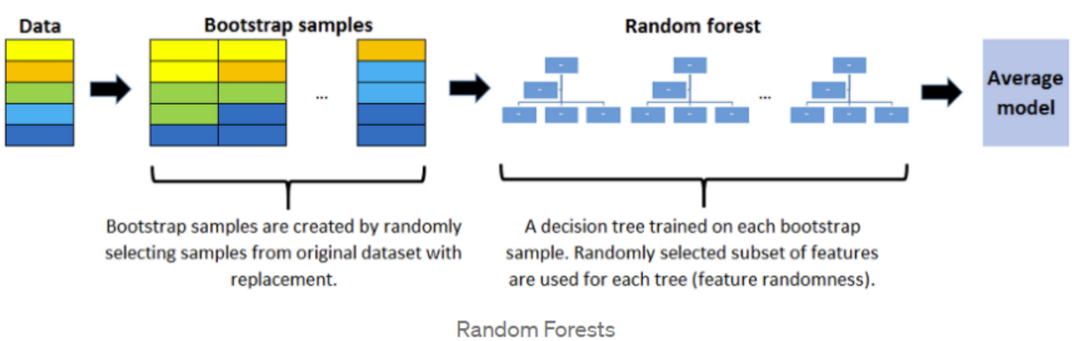


Fig 11. Random Forest Model

Syntax:

sklearn.ensemble.RandomForestClassifier()

For this project's fraud analysis, four versions of Random Forests were created by changing the n_estimators, max_depth, max_features, min_samples_leaf, and min_samples_split hyperparameters and training the model with the identified 30 best variables. To understand the used hyperparameters further:

- n_estimators: The number of trees in the forest. default=100
- max_depth: The maximum depth of the tree. If None, then nodes are expanded until all leaves are pure or until all leaves contain less than min_samples_split samples. default=None
- max_features: The number of features to consider when looking for the best split. default="auto"

- min_samples_leaf: The minimum number of samples required to be at a leaf node. A split point at any depth will only be considered if it leaves at least min_samples_leaf training samples in each of the left and right branches. This may have the effect of smoothing the model, especially in regression. Default is 1.
- min_samples_split: The minimum number of samples required to split an internal node. Default is 2.

The results of the Random Forest models are shown below.

Model	Parameters					Average FDR at 3%		
	n_estimators	max_depth	max_features	min_samples_leaf	min_samples_split	Train	Test	OOT
Random Forest	100	10	5	1	2	0.522	0.52	0.501
	50	20	5	30	500	0.52	0.523	0.502
	50	20	5	30	300	0.521	0.515	0.502
	75	10	8	30	100	0.524	0.513	0.5

Table 10: Random Forest model results

The best results were given when n_estimators = 50, max_depth = 20, max_features = 5, min_samples_leaf = 30 and min_samples_split = 500.

Boosted Tree

Gradient boosting algorithm sequentially combines weak learners in way that each new learner fits to the residuals from the previous step so that the model improves. The final model aggregates the results from each step and a strong learner is achieved. Gradient boosted decision trees algorithm uses decision trees as week learners. A loss function is used to detect the residuals. For instance, mean squared error (MSE) can be used for a regression task and logarithmic loss (log loss) can be used for classification tasks. It is worth noting that existing trees in the model do not change when a new tree is added. The added decision tree fits the residuals from the current model. The steps are as follows:

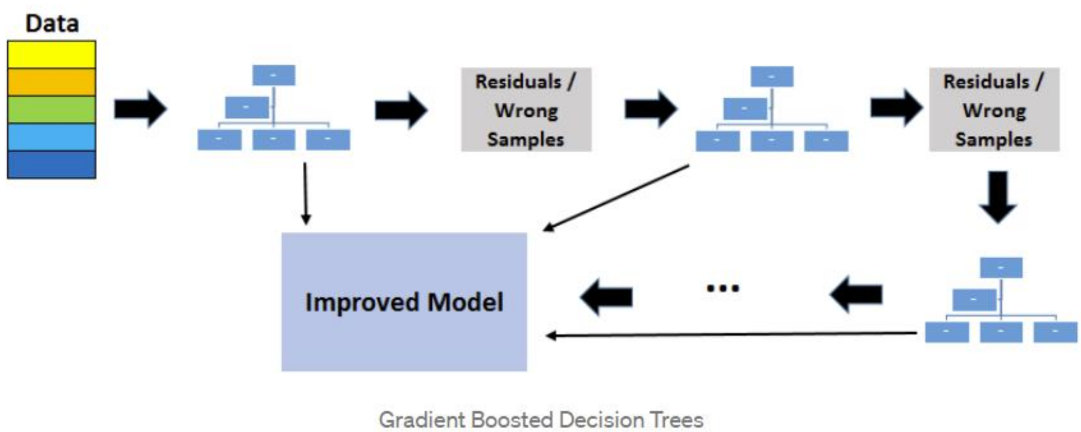


Fig 12. Sample working principle of Boosted Tree Model

Model	Parameters				Average FDR at 3%		
	learning_rate	n_estimators	max_depth	num_leaves	Train	Test	OOT
Boosted Tree	0.1	100	3	10	0.525	0.512	0.502
	0.01	800	5	30	0.52	0.524	0.502
	0.001	4000	4	30	0.518	0.52	0.499
	0.05	200	5	40	0.519	0.528	0.503

Table 11: Boosted Tree model results

Learning rate and n_estimators

Hyperparameters are key parts of learning algorithms which effect the performance and accuracy of a model. Learning rate and n_estimators are two critical hyperparameters for gradient boosting decision trees. Learning rate, denoted as α, simply means how fast the model learns. Each tree added modifies the overall model. The magnitude of the modification is controlled by learning rate. The steps of gradient boosted decision tree algorithms with learning rate introduced:

- $f_1(x) \approx y$
- The residual is $y - \alpha f_1(x)$
- $f_2(x) \approx y - \alpha f_1(x)$
- The residual is $y - \alpha f_1(x) - \alpha f_2(x)$
- $f_3(x) \approx y - \alpha f_1(x) - \alpha f_2(x)$

Gradient boosted decision tree algorithm with learning rate (α)

The lower the learning rate, the slower the model learns. The advantage of slower learning rate is that the model becomes more robust and generalized. In statistical learning, models that learn slowly perform better. However, learning slowly comes at a cost. It takes more time to train the model which brings us to the other significant hyperparameter. n_estimator is the number of trees used in the model. If the learning rate is low, we need more trees to train the model. However, we need to be very careful at selecting the number of trees. It creates a high risk of overfitting to use too many trees.

Neural Network

Neural network function in a way like biological neurons. They take in various inputs (at input layers), weight these inputs, and then combine the weighted inputs through a linear combination (much like linear regression). If the combined weighted output is past some thresholds set by an activation function, the output is then set out to other layers. The base unit is generally referred to as a perceptron. Perceptron’s are combined to form neural networks, which is why they are also called as multi-layer perceptron’s (MLPs).

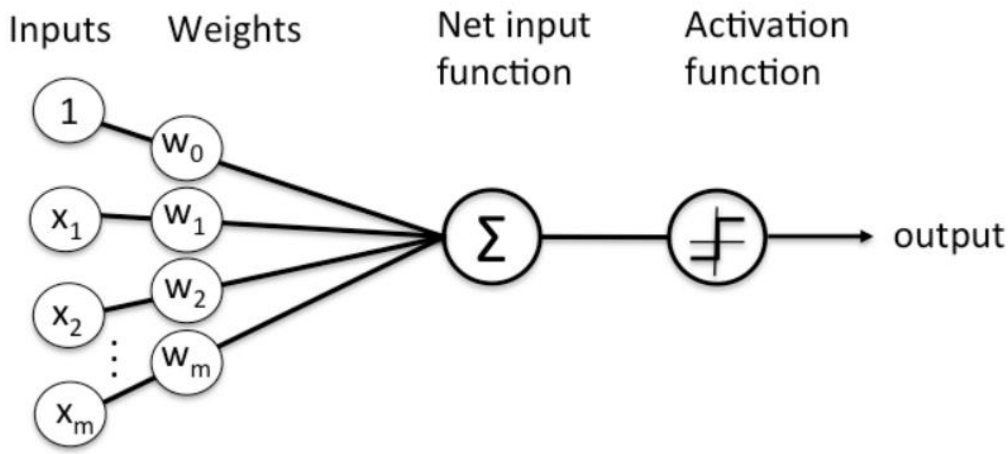


Fig 13. Sample working principle of Random Forest Model

The process of receiving inputs and generating an output continues until an output layer is reached. This is generally done in a forward manner, meaning that layer's process incoming data in a sequential forward way. The layers of neurons that are not input or output layers are called the hidden layers. Hidden layers allow for a specific transformation of the data within each layer. Each hidden layer can be specialized to produce a particular output.
The learning process for neural networks is called backpropagation. This technique modifies the weights of neural network iteratively through the calculation of deltas between predicted and expected outputs. After this calculation, the weights are updated backwards through earlier layers via stochastic gradient descent. The process continues until the weights that minimize the loss function is found.

Model	Parameters					Average FDR at 3%		
	activation	max_iter	learning_rate	learning_rate_init	alpha	Train	Test	OOT
Neural Network	relu	200	constant	0.005	0.0001	0.523	0.515	0.5
	relu	500	adaptive	0.01	0.0001	0.52	0.52	0.499
	logistic	100	constant	0.001	0.001	0.519	0.509	0.498
	tanh	200	invscaling	0.005	0.0003	0.511	0.534	0.498

Table 12: Neural Network model results

Final Model and Results:

The below diagram gives us the results from all models that were tested to predict fraud

Model	Parameters					Average FDR at 3%		
Logistics Regression	**solver**	**penalty**	**C**			**Train**	**Test**	**OOT**
	liblinear	l1	1			0.497	0.491	0.483
	lbfgs	l2	1			0.491	0.493	0.479
	lbfgs	l2	0.1			0.49	0.492	0.477
	liblinear	l2	1			0.496	0.505	0.483
	lbfgs	l2	10			0.501	0.495	0.483
Decision Tree	**max_depth**	**min_sample_leaf**	**min_samples_split**			**Train**	**Test**	**OOT**
	10	1	2			0.522	0.515	0.495
	20	60	300			0.524	0.511	0.502
	20	60	320			0.525	0.514	0.5
	25	80	600			0.516	0.533	0.498
Random Forest	**n_estimators**	**max_depth**	**max_features**	**min_samples_leaf**	**min_samples_split**	**Train**	**Test**	**OOT**
	100	10	5	1	2	0.522	0.52	0.501
	300	200	7	30	500	0.52	0.536	0.505
	50	20	5	30	300	0.521	0.515	0.502
	75	10	8	30	100	0.524	0.513	0.5
Boosted Tree	**learning_rate**	**n_estimators**	**max_depth**	**num_leaves**		**Train**	**Test**	**OOT**
	0.1	100	3	10		0.525	0.512	0.502
	0.01	800	5	30		0.52	0.524	0.502
	0.001	4000	4	30		0.518	0.52	0.499
	0.05	200	5	40		0.519	0.528	0.503
Neural Network	**activation**	**max_iter**	**learning_rate**	**learning_rate_init**	**alpha**	**Train**	**Test**	**OOT**
	relu	200	constant	0.005	0.0001	0.523	0.515	0.5
	relu	500	adaptive	0.01	0.0001	0.52	0.52	0.499
	logistic	100	constant	0.001	0.001	0.519	0.509	0.498
	tanh	200	invscaling	0.005	0.0003	0.511	0.534	0.498

Table 13: Final Summary Of Models

Final Model: Random Forest

After comparing the results between logistics regression, boosted trees, random forest, and a neural network, we determined that Random Forest performed the best. Random Forest outperformed other models for both testing and out of time validation datasets with 53.6% and 50.5% respectively.

The chosen hyperparameters are:

- No. of variables: 20
- n_estimators: 300
- max_depth: 200
- max_features: 7
- min_sample_leaf: 30min_samples_split: 500

Conclusion:

Identified synthetic identity fraud in credit card applications among a million personal information records. First, exploratory data analysis was performed on the dataset which helped us identify the missing and frivolous variables. Data cleaning was performed as a next step to impute the missing values and all frivolous variables were updated to match their record numbers. Then, over 1984 expert candidate variables were created to identify fraudulent behavior.

Feature selection was performed using (filter and wrapper methods) to pick the 30 best variables of multivariate importance. The top 30 variables were then selected used to build models such as logistic regression, random forest, neural networks, boosted trees to detect fraudulent applications. Our best model was a Random Forest with a 53.6% (FDR at the 3%) level on test data and 50.5% (FDR at the 3%) level on the OOT data.

For future work if we are allotted additional time and computer resources, there are a few items that we would investigate further in future iterations of this process. First, we would consult subject matter experts for making more effective variables in the variable building stage, that may include enhanced business knowledge too. Second, we would have consulted subject matter expert to employ non-linear techniques for selecting more relevant features for our model calculation. Third, we would spend more time evaluating each of the models with additional hyperparameters and additional values for each of the parameters.

Appendix A

Data Quality Report

1. Dataset Overview

Dataset Name- Applications Data

Dataset Description – Dataset contains information of the people who filled the application for a particular product. This dataset contains fields like SSN, name, address and phone number and fraud label-which tells us whether the application is fraudulent or not.

Total Fields – 10

Total Records – 1,000,000

Time Period - 1st January 2016 – 31st December 2016

2. Summary

For this dataset all the fields are categorical and below table gives the summary description of all the 10 records.

Field Name	% Populated	# Unique Values	Most Common Value
record	100%	10,00,000	NA
date	100%	365	2016-08-16
ssn	100%	8,35,819	999999999
firstname	100%	78,136	EAMSTRMT
lastname	100%	1,77,001	ERJSAXA
address	100%	8,28,774	123 MAIN ST
zip5	100%	26,370	68138
dob	100%	42,673	1907-06-26
homephone	100%	28,244	9999999999
fraud_label	100%	2	0

3. Field Exploration

Field 1 – record

Description – A categorical field containing unique number for each record

Field 2 – date

Description – A categorical field containing date when the application was filled

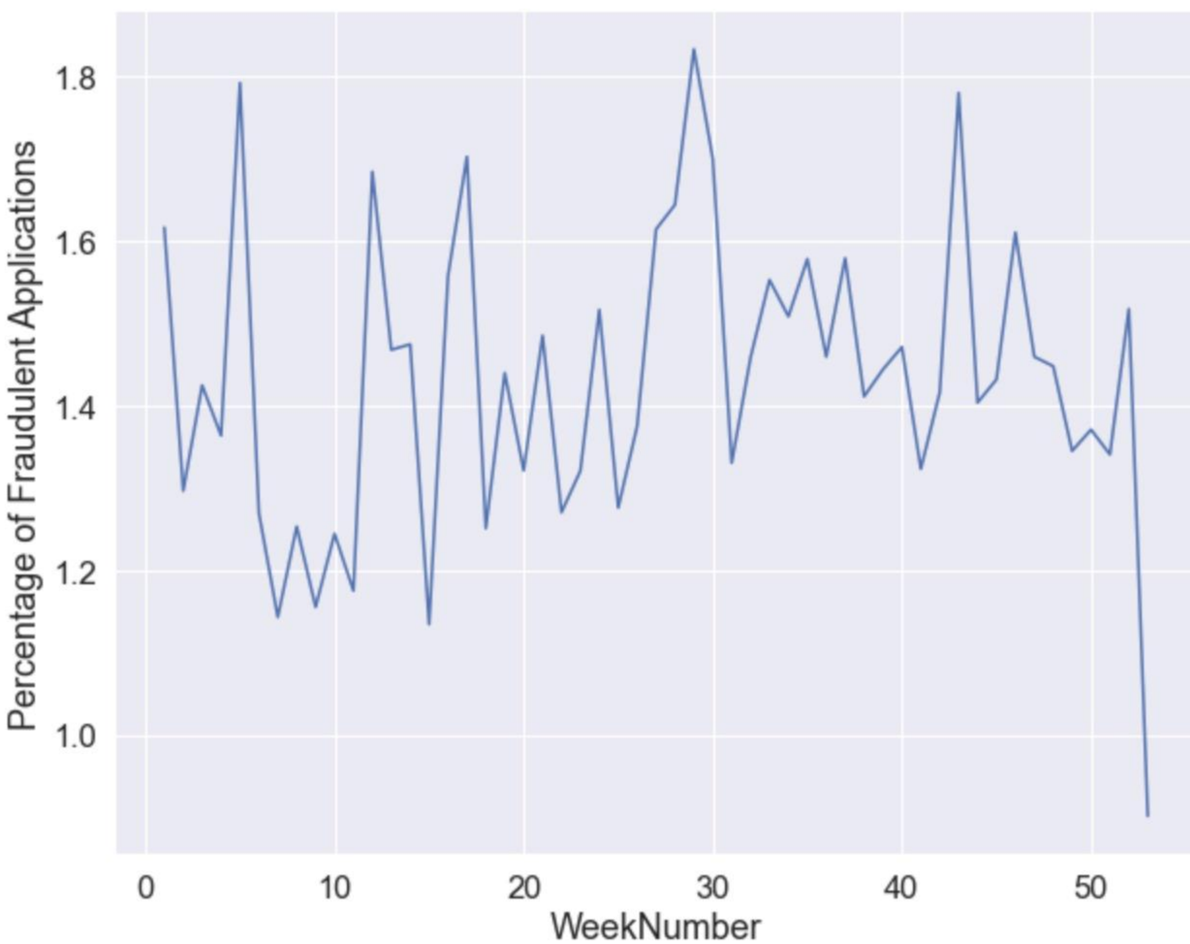


Fig 1 – Weekly Percentage of Fraud Applications

Field 3- ssn

Description- Field containing the social security number of each applicant

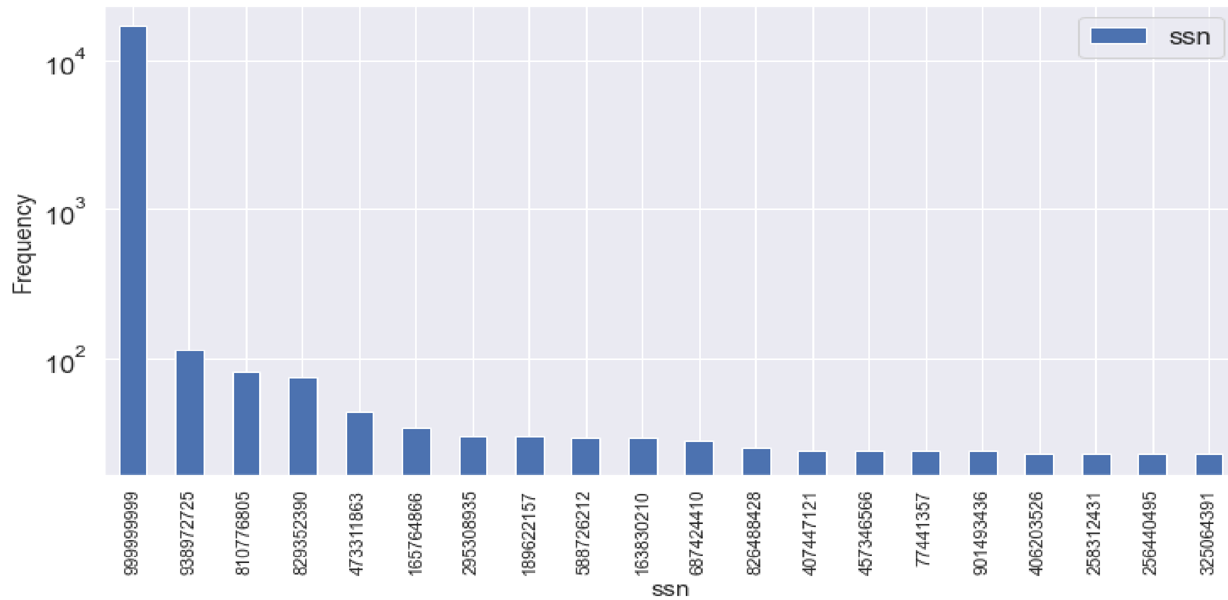


Fig 2 – Bar Plot showing distribution of ssn

Field 4 – firstname

Description – Field containing First Name of each applicant

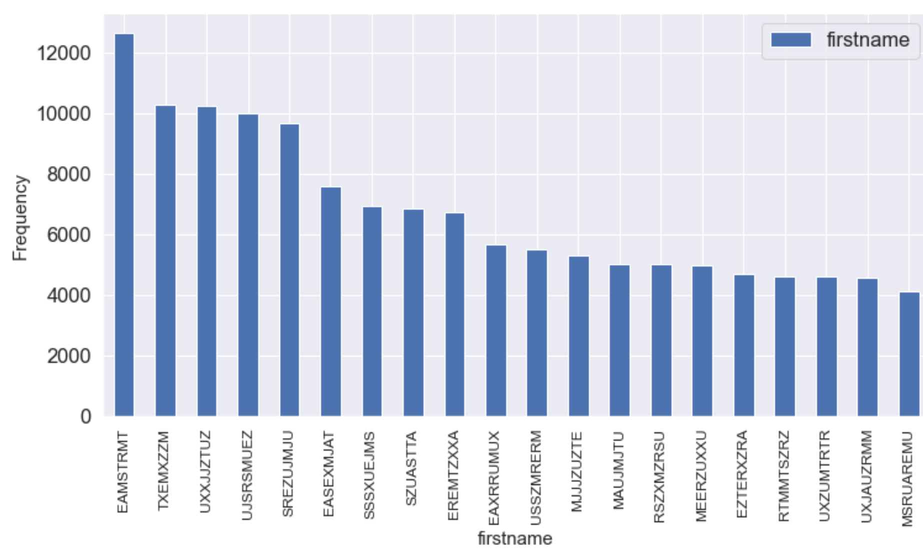


Fig 3 – Bar Plot showing distribution of firstname

Field 5 – lastname

Description – Field containing last name of each applicant

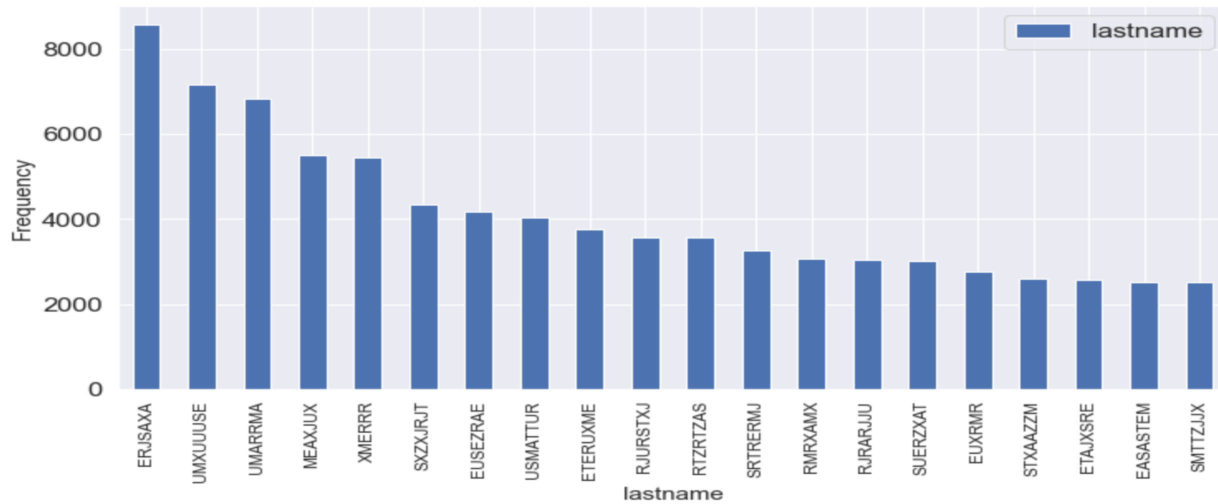


Fig 4 – Bar Plot showing distribution of lastname

Field 6 – address

Description – A field which describes the address of each applicant

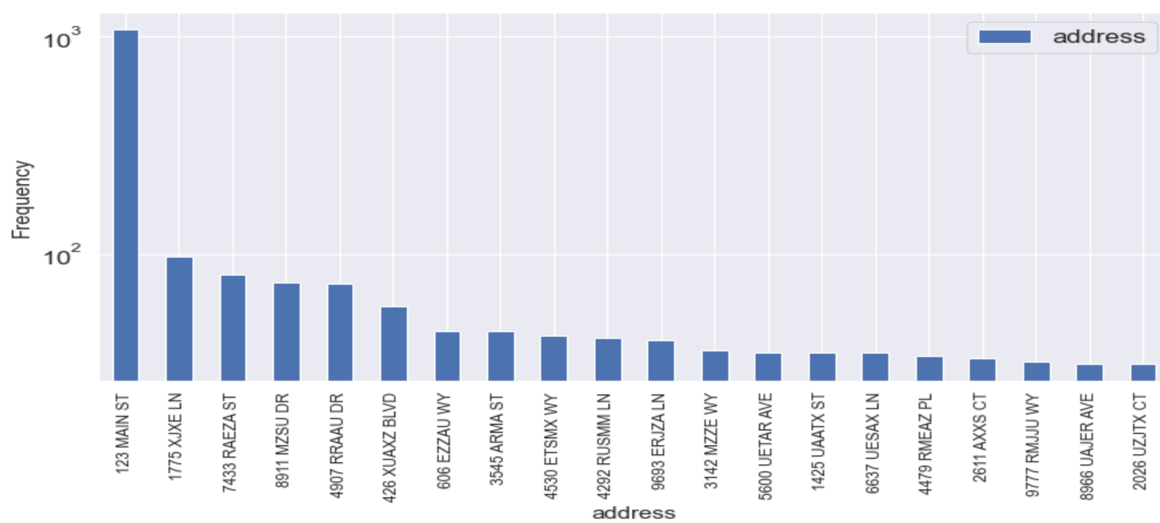


Fig 5 – Bar Plot showing distribution of address

Field 7 – zip5

Description – Field containing 5 digits of Zip code

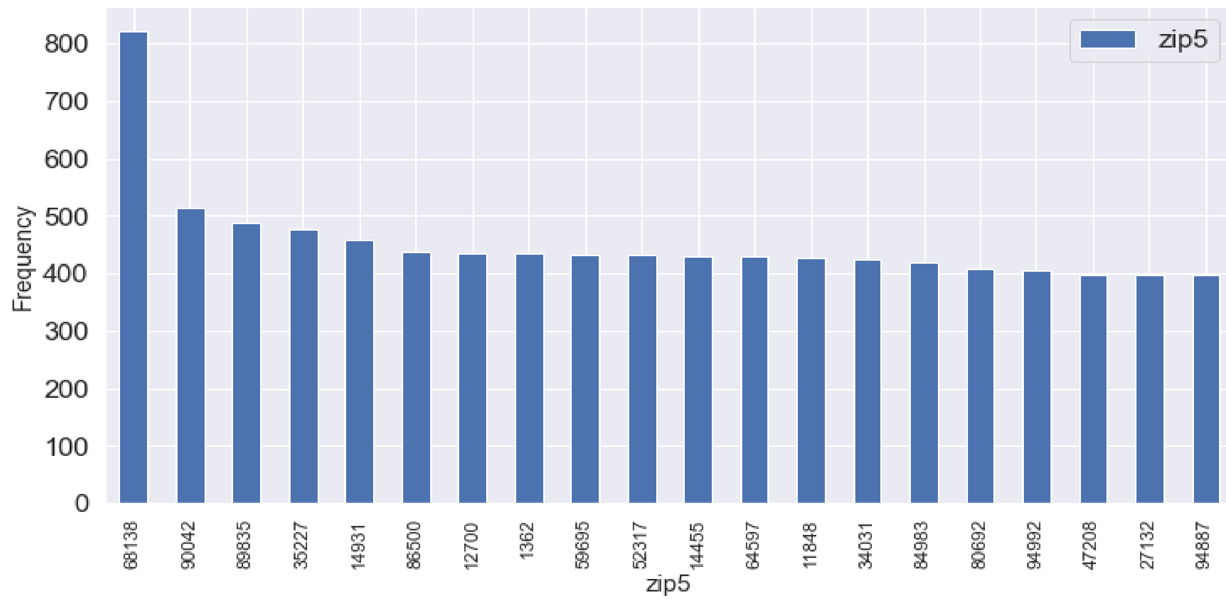


Fig 6 – Bar Plot showing distribution of zip5

Field 8 – dob

Description – Field which describes the date of birth of an applicant.

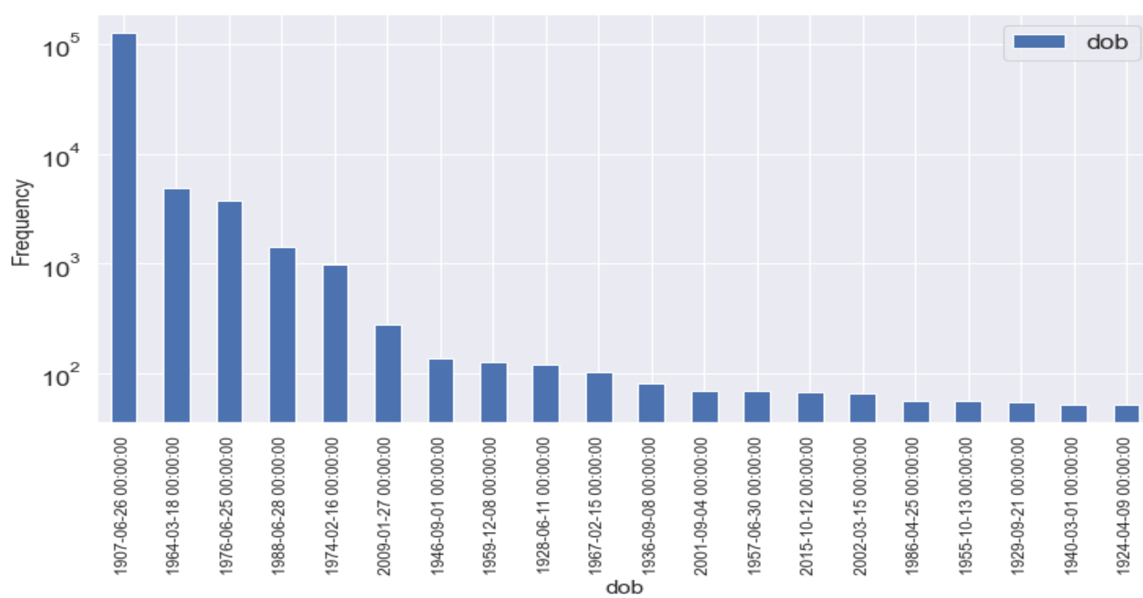


Fig 7 – Bar Plot showing distribution of dob

Field 9 – homephone

Description – Field which contains home phone number of applicant

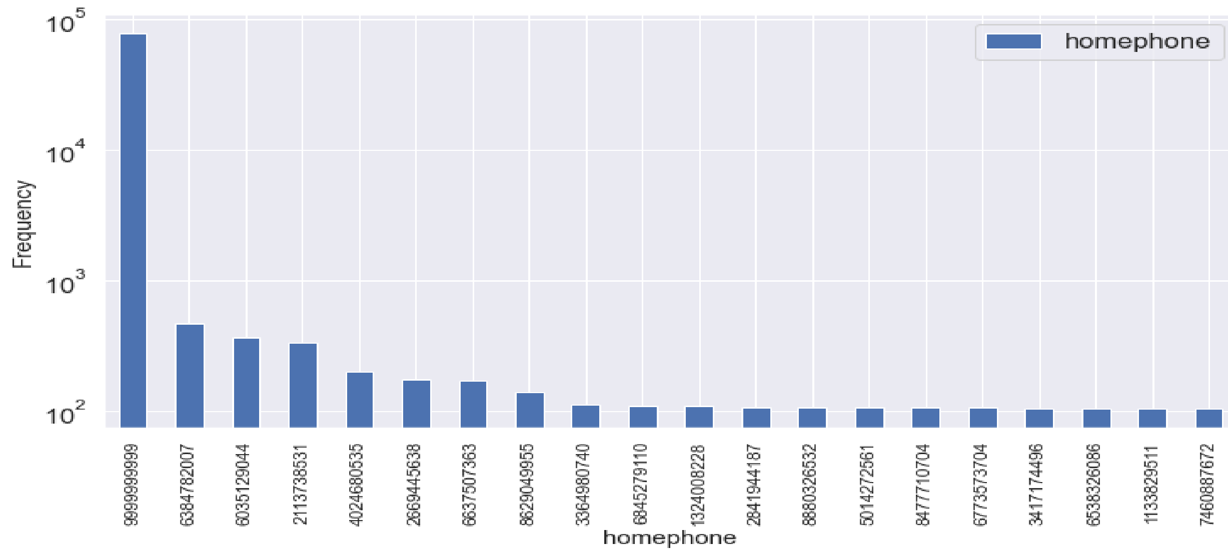


Fig 8 – Bar Plot showing distribution of homephone

Field 10 – fraud_label

Description – Field containing 2 categories. 0 indicates a good application and 1 indicates a bad application.

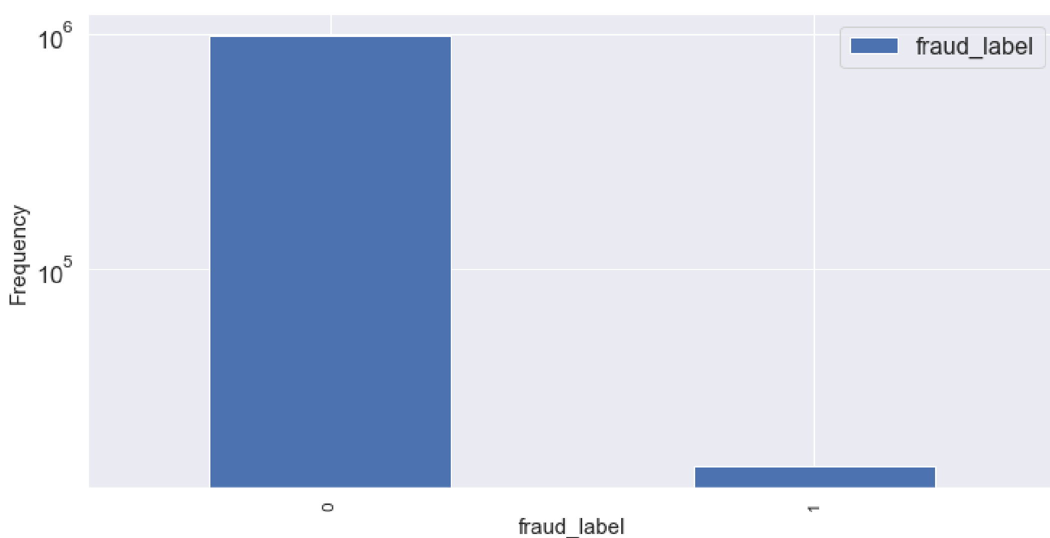


Fig 9 – Bar Plot showing distribution of fraud_label

Appendix B

record	ssnunique_count_forssn_name_dob_3
date	ssnunique_count_forssn_name_dob_7
ssn	ssnunique_count_forssn_name_dob_14
firstname	ssnunique_count_forssn_name_dob_30
lastname	ssnunique_count_forssn_name_dob_60
address	ssnunique_count_forssn_name_fulladdress_1
zip5	ssnunique_count_forssn_name_fulladdress_3
dob	ssnunique_count_forssn_name_fulladdress_7
homephone	ssnunique_count_forssn_name_fulladdress_14
fraud_label	ssnunique_count_forssn_name_fulladdress_30
dow	ssnunique_count_forssn_name_fulladdress_60
dow_risk	fulladdressunique_count_forssn_1
name	fulladdressunique_count_forssn_3
fulladdress	fulladdressunique_count_forssn_7
name_dob	fulladdressunique_count_forssn_14
name_fulladdress	fulladdressunique_count_forssn_30
name_homephone	fulladdressunique_count_forssn_60
fulladdress_dob	fulladdressunique_count_forname_dob_1
fulladdress_homephone	fulladdressunique_count_forname_dob_3
dob_homephone	fulladdressunique_count_forname_dob_7
homephone_name_dob	fulladdressunique_count_forname_dob_14
fulladdress_name_dob	fulladdressunique_count_forname_dob_30
name_zip5	fulladdressunique_count_forname_dob_60
name_address	fulladdressunique_count_forname_fulladdress_1
firstname_dob	fulladdressunique_count_forname_fulladdress_3
firstname_address	fulladdressunique_count_forname_fulladdress_7
firstname_zip5	fulladdressunique_count_forname_fulladdress_14
firstname_homephone	fulladdressunique_count_forname_fulladdress_30
firstname_fulladdress	fulladdressunique_count_forname_fulladdress_60
lastname_dob	fulladdressunique_count_forfulladdress_dob_1
lastname_address	fulladdressunique_count_forfulladdress_dob_3
lastname_zip5	fulladdressunique_count_forfulladdress_dob_7
lastname_homephone	fulladdressunique_count_forfulladdress_dob_14
lastname_fulladdress	fulladdressunique_count_forfulladdress_dob_30
ssn_firstname	fulladdressunique_count_forfulladdress_dob_60
ssn_lastname	fulladdressunique_count_fordob_homephone_1
ssn_address	fulladdressunique_count_fordob_homephone_3
ssn_zip5	fulladdressunique_count_fordob_homephone_7
ssn_dob	fulladdressunique_count_fordob_homephone_14
ssn_homephone	fulladdressunique_count_fordob_homephone_30
ssn_name	fulladdressunique_count_fordob_homephone_60

ssn_fulladdress	fulladdressunique_count_forssn_lastname_1
ssn_name_dob	fulladdressunique_count_forssn_lastname_3
ssn_name_fulladdress	fulladdressunique_count_forssn_lastname_7
ssn_name_homephone	fulladdressunique_count_forssn_lastname_14
ssn_fulladdress_dob	fulladdressunique_count_forssn_lastname_30
ssn_fulladdress_homephone	fulladdressunique_count_forssn_lastname_60
ssn_dob_homephone	fulladdressunique_count_forssn_zip5_1
ssn_homephone_name_dob	fulladdressunique_count_forssn_zip5_3
ssn_fulladdress_name_dob	fulladdressunique_count_forssn_zip5_7
ssn_day_since	fulladdressunique_count_forssn_zip5_14
ssn_count_0	fulladdressunique_count_forssn_zip5_30
ssn_count_1	fulladdressunique_count_forssn_zip5_60
ssn_count_3	fulladdressunique_count_forssn_name_1
ssn_count_7	fulladdressunique_count_forssn_name_3
ssn_count_14	fulladdressunique_count_forssn_name_7
ssn_count_30	fulladdressunique_count_forssn_name_14
address_day_since	fulladdressunique_count_forssn_name_30
address_count_0	fulladdressunique_count_forssn_name_60
address_count_1	fulladdressunique_count_forssn_fulladdress_1
address_count_3	fulladdressunique_count_forssn_fulladdress_3
address_count_7	fulladdressunique_count_forssn_fulladdress_7
address_count_14	fulladdressunique_count_forssn_fulladdress_14
address_count_30	fulladdressunique_count_forssn_fulladdress_30
dob_day_since	fulladdressunique_count_forssn_fulladdress_60
dob_count_0	fulladdressunique_count_forssn_name_dob_1
dob_count_1	fulladdressunique_count_forssn_name_dob_3
dob_count_3	fulladdressunique_count_forssn_name_dob_7
dob_count_7	fulladdressunique_count_forssn_name_dob_14
dob_count_14	fulladdressunique_count_forssn_name_dob_30
dob_count_30	fulladdressunique_count_forssn_name_dob_60
homephone_day_since	fulladdressunique_count_forssn_name_fulladdress_1
homephone_count_0	fulladdressunique_count_forssn_name_fulladdress_3
homephone_count_1	fulladdressunique_count_forssn_name_fulladdress_7
homephone_count_3	fulladdressunique_count_forssn_name_fulladdress_14
homephone_count_7	fulladdressunique_count_forssn_name_fulladdress_30
homephone_count_14	fulladdressunique_count_forssn_name_fulladdress_60
homephone_count_30	name_dobunique_count_forssn_1
name_day_since	name_dobunique_count_forssn_3
name_count_0	name_dobunique_count_forssn_7
name_count_1	name_dobunique_count_forssn_14
name_count_3	name_dobunique_count_forssn_30
name_count_7	name_dobunique_count_forssn_60
name_count_14	name_dobunique_count_forfulladdress_1

name_count_30	name_dobunique_count_forfulladdress_3
fulladdress_day_since	name_dobunique_count_forfulladdress_7
fulladdress_count_0	name_dobunique_count_forfulladdress_14
fulladdress_count_1	name_dobunique_count_forfulladdress_30
fulladdress_count_3	name_dobunique_count_forfulladdress_60
fulladdress_count_7	name_dobunique_count_forname_fulladdress_1
fulladdress_count_14	name_dobunique_count_forname_fulladdress_3
fulladdress_count_30	name_dobunique_count_forname_fulladdress_7
name_dob_day_since	name_dobunique_count_forname_fulladdress_14
name_dob_count_0	name_dobunique_count_forname_fulladdress_30
name_dob_count_1	name_dobunique_count_forname_fulladdress_60
name_dob_count_3	name_dobunique_count_forfulladdress_dob_1
name_dob_count_7	name_dobunique_count_forfulladdress_dob_3
name_dob_count_14	name_dobunique_count_forfulladdress_dob_7
name_dob_count_30	name_dobunique_count_forfulladdress_dob_14
name_fulladdress_day_since	name_dobunique_count_forfulladdress_dob_30
name_fulladdress_count_0	name_dobunique_count_forfulladdress_dob_60
name_fulladdress_count_1	name_dobunique_count_fordob_homephone_1
name_fulladdress_count_3	name_dobunique_count_fordob_homephone_3
name_fulladdress_count_7	name_dobunique_count_fordob_homephone_7
name_fulladdress_count_14	name_dobunique_count_fordob_homephone_14
name_fulladdress_count_30	name_dobunique_count_fordob_homephone_30
name_homephone_day_since	name_dobunique_count_fordob_homephone_60
name_homephone_count_0	name_dobunique_count_forssn_lastname_1
name_homephone_count_1	name_dobunique_count_forssn_lastname_3
name_homephone_count_3	name_dobunique_count_forssn_lastname_7
name_homephone_count_7	name_dobunique_count_forssn_lastname_14
name_homephone_count_14	name_dobunique_count_forssn_lastname_30
name_homephone_count_30	name_dobunique_count_forssn_lastname_60
fulladdress_dob_day_since	name_dobunique_count_forssn_zip5_1
fulladdress_dob_count_0	name_dobunique_count_forssn_zip5_3
fulladdress_dob_count_1	name_dobunique_count_forssn_zip5_7
fulladdress_dob_count_3	name_dobunique_count_forssn_zip5_14
fulladdress_dob_count_7	name_dobunique_count_forssn_zip5_30
fulladdress_dob_count_14	name_dobunique_count_forssn_zip5_60
fulladdress_dob_count_30	name_dobunique_count_forssn_name_1
fulladdress_homephone_day_since	name_dobunique_count_forssn_name_3
fulladdress_homephone_count_0	name_dobunique_count_forssn_name_7
fulladdress_homephone_count_1	name_dobunique_count_forssn_name_14
fulladdress_homephone_count_3	name_dobunique_count_forssn_name_30
fulladdress_homephone_count_7	name_dobunique_count_forssn_name_60
fulladdress_homephone_count_14	name_dobunique_count_forssn_fulladdress_1
fulladdress_homephone_count_30	name_dobunique_count_forssn_fulladdress_3

dob_homephone_day_since	name_dobunique_count_forssn_fulladdress_7
dob_homephone_count_0	name_dobunique_count_forssn_fulladdress_14
dob_homephone_count_1	name_dobunique_count_forssn_fulladdress_30
dob_homephone_count_3	name_dobunique_count_forssn_fulladdress_60
dob_homephone_count_7	name_dobunique_count_forssn_name_dob_1
dob_homephone_count_14	name_dobunique_count_forssn_name_dob_3
dob_homephone_count_30	name_dobunique_count_forssn_name_dob_7
homephone_name_dob_day_since	name_dobunique_count_forssn_name_dob_14
homephone_name_dob_count_0	name_dobunique_count_forssn_name_dob_30
homephone_name_dob_count_1	name_dobunique_count_forssn_name_dob_60
homephone_name_dob_count_3	name_dobunique_count_forssn_name_fulladdress_1
homephone_name_dob_count_7	name_dobunique_count_forssn_name_fulladdress_3
homephone_name_dob_count_14	name_dobunique_count_forssn_name_fulladdress_7
homephone_name_dob_count_30	name_dobunique_count_forssn_name_fulladdress_14
fulladdress_name_dob_day_since	name_dobunique_count_forssn_name_fulladdress_30
fulladdress_name_dob_count_0	name_dobunique_count_forssn_name_fulladdress_60
fulladdress_name_dob_count_1	name_fulladdressunique_count_forssn_1
fulladdress_name_dob_count_3	name_fulladdressunique_count_forssn_3
fulladdress_name_dob_count_7	name_fulladdressunique_count_forssn_7
fulladdress_name_dob_count_14	name_fulladdressunique_count_forssn_14
fulladdress_name_dob_count_30	name_fulladdressunique_count_forssn_30
name_zip5_day_since	name_fulladdressunique_count_forssn_60
name_zip5_count_0	name_fulladdressunique_count_forfulladdress_1
name_zip5_count_1	name_fulladdressunique_count_forfulladdress_3
name_zip5_count_3	name_fulladdressunique_count_forfulladdress_7
name_zip5_count_7	name_fulladdressunique_count_forfulladdress_14
name_zip5_count_14	name_fulladdressunique_count_forfulladdress_30
name_zip5_count_30	name_fulladdressunique_count_forfulladdress_60
name_address_day_since	name_fulladdressunique_count_forname_dob_1
name_address_count_0	name_fulladdressunique_count_forname_dob_3
name_address_count_1	name_fulladdressunique_count_forname_dob_7
name_address_count_3	name_fulladdressunique_count_forname_dob_14
name_address_count_7	name_fulladdressunique_count_forname_dob_30
name_address_count_14	name_fulladdressunique_count_forname_dob_60
name_address_count_30	name_fulladdressunique_count_forfulladdress_dob_1
firstname_dob_day_since	name_fulladdressunique_count_forfulladdress_dob_3
firstname_dob_count_0	name_fulladdressunique_count_forfulladdress_dob_7
firstname_dob_count_1	name_fulladdressunique_count_forfulladdress_dob_14
firstname_dob_count_3	name_fulladdressunique_count_forfulladdress_dob_30
firstname_dob_count_7	name_fulladdressunique_count_forfulladdress_dob_60
firstname_dob_count_14	name_fulladdressunique_count_fordob_homephone_1
firstname_dob_count_30	name_fulladdressunique_count_fordob_homephone_3
firstname_address_day_since	name_fulladdressunique_count_fordob_homephone_7

firstname_address_count_0	name_fulladdressunique_count_fordob_homephone_14
firstname_address_count_1	name_fulladdressunique_count_fordob_homephone_30
firstname_address_count_3	name_fulladdressunique_count_fordob_homephone_60
firstname_address_count_7	name_fulladdressunique_count_forssn_lastname_1
firstname_address_count_14	name_fulladdressunique_count_forssn_lastname_3
firstname_address_count_30	name_fulladdressunique_count_forssn_lastname_7
firstname_zip5_day_since	name_fulladdressunique_count_forssn_lastname_14
firstname_zip5_count_0	name_fulladdressunique_count_forssn_lastname_30
firstname_zip5_count_1	name_fulladdressunique_count_forssn_lastname_60
firstname_zip5_count_3	name_fulladdressunique_count_forssn_zip5_1
firstname_zip5_count_7	name_fulladdressunique_count_forssn_zip5_3
firstname_zip5_count_14	name_fulladdressunique_count_forssn_zip5_7
firstname_zip5_count_30	name_fulladdressunique_count_forssn_zip5_14
firstname_homephone_day_since	name_fulladdressunique_count_forssn_zip5_30
firstname_homephone_count_0	name_fulladdressunique_count_forssn_zip5_60
firstname_homephone_count_1	name_fulladdressunique_count_forssn_name_1
firstname_homephone_count_3	name_fulladdressunique_count_forssn_name_3
firstname_homephone_count_7	name_fulladdressunique_count_forssn_name_7
firstname_homephone_count_14	name_fulladdressunique_count_forssn_name_14
firstname_homephone_count_30	name_fulladdressunique_count_forssn_name_30
firstname_fulladdress_day_since	name_fulladdressunique_count_forssn_name_60
firstname_fulladdress_count_0	name_fulladdressunique_count_forssn_fulladdress_1
firstname_fulladdress_count_1	name_fulladdressunique_count_forssn_fulladdress_3
firstname_fulladdress_count_3	name_fulladdressunique_count_forssn_fulladdress_7
firstname_fulladdress_count_7	name_fulladdressunique_count_forssn_fulladdress_14
firstname_fulladdress_count_14	name_fulladdressunique_count_forssn_fulladdress_30
firstname_fulladdress_count_30	name_fulladdressunique_count_forssn_fulladdress_60
lastname_dob_day_since	name_fulladdressunique_count_forssn_name_dob_1
lastname_dob_count_0	name_fulladdressunique_count_forssn_name_dob_3
lastname_dob_count_1	name_fulladdressunique_count_forssn_name_dob_7
lastname_dob_count_3	name_fulladdressunique_count_forssn_name_dob_14
lastname_dob_count_7	name_fulladdressunique_count_forssn_name_dob_30
lastname_dob_count_14	name_fulladdressunique_count_forssn_name_dob_60
lastname_dob_count_30	name_fulladdressunique_count_forssn_name_fulladdress_1
lastname_address_day_since	name_fulladdressunique_count_forssn_name_fulladdress_3
lastname_address_count_0	name_fulladdressunique_count_forssn_name_fulladdress_7
lastname_address_count_1	name_fulladdressunique_count_forssn_name_fulladdress_14
lastname_address_count_3	name_fulladdressunique_count_forssn_name_fulladdress_30
lastname_address_count_7	name_fulladdressunique_count_forssn_name_fulladdress_60
lastname_address_count_14	fulladdress_dobunique_count_forssn_1
lastname_address_count_30	fulladdress_dobunique_count_forssn_3
lastname_zip5_day_since	fulladdress_dobunique_count_forssn_7
lastname_zip5_count_0	fulladdress_dobunique_count_forssn_14

lastname_zip5_count_1	fulladdress_dobunique_count_forssn_30
lastname_zip5_count_3	fulladdress_dobunique_count_forssn_60
lastname_zip5_count_7	fulladdress_dobunique_count_forfulladdress_1
lastname_zip5_count_14	fulladdress_dobunique_count_forfulladdress_3
lastname_zip5_count_30	fulladdress_dobunique_count_forfulladdress_7
lastname_homephone_day_since	fulladdress_dobunique_count_forfulladdress_14
lastname_homephone_count_0	fulladdress_dobunique_count_forfulladdress_30
lastname_homephone_count_1	fulladdress_dobunique_count_forfulladdress_60
lastname_homephone_count_3	fulladdress_dobunique_count_forname_dob_1
lastname_homephone_count_7	fulladdress_dobunique_count_forname_dob_3
lastname_homephone_count_14	fulladdress_dobunique_count_forname_dob_7
lastname_homephone_count_30	fulladdress_dobunique_count_forname_dob_14
lastname_fulladdress_day_since	fulladdress_dobunique_count_forname_dob_30
lastname_fulladdress_count_0	fulladdress_dobunique_count_forname_dob_60
lastname_fulladdress_count_1	fulladdress_dobunique_count_forname_fulladdress_1
lastname_fulladdress_count_3	fulladdress_dobunique_count_forname_fulladdress_3
lastname_fulladdress_count_7	fulladdress_dobunique_count_forname_fulladdress_7
lastname_fulladdress_count_14	fulladdress_dobunique_count_forname_fulladdress_14
lastname_fulladdress_count_30	fulladdress_dobunique_count_forname_fulladdress_30
ssn_firstname_day_since	fulladdress_dobunique_count_forname_fulladdress_60
ssn_firstname_count_0	fulladdress_dobunique_count_fordob_homephone_1
ssn_firstname_count_1	fulladdress_dobunique_count_fordob_homephone_3
ssn_firstname_count_3	fulladdress_dobunique_count_fordob_homephone_7
ssn_firstname_count_7	fulladdress_dobunique_count_fordob_homephone_14
ssn_firstname_count_14	fulladdress_dobunique_count_fordob_homephone_30
ssn_firstname_count_30	fulladdress_dobunique_count_fordob_homephone_60
ssn_lastname_day_since	fulladdress_dobunique_count_forssn_lastname_1
ssn_lastname_count_0	fulladdress_dobunique_count_forssn_lastname_3
ssn_lastname_count_1	fulladdress_dobunique_count_forssn_lastname_7
ssn_lastname_count_3	fulladdress_dobunique_count_forssn_lastname_14
ssn_lastname_count_7	fulladdress_dobunique_count_forssn_lastname_30
ssn_lastname_count_14	fulladdress_dobunique_count_forssn_lastname_60
ssn_lastname_count_30	fulladdress_dobunique_count_forssn_zip5_1
ssn_address_day_since	fulladdress_dobunique_count_forssn_zip5_3
ssn_address_count_0	fulladdress_dobunique_count_forssn_zip5_7
ssn_address_count_1	fulladdress_dobunique_count_forssn_zip5_14
ssn_address_count_3	fulladdress_dobunique_count_forssn_zip5_30
ssn_address_count_7	fulladdress_dobunique_count_forssn_zip5_60
ssn_address_count_14	fulladdress_dobunique_count_forssn_name_1
ssn_address_count_30	fulladdress_dobunique_count_forssn_name_3
ssn_zip5_day_since	fulladdress_dobunique_count_forssn_name_7
ssn_zip5_count_0	fulladdress_dobunique_count_forssn_name_14
ssn_zip5_count_1	fulladdress_dobunique_count_forssn_name_30

ssn_zip5_count_3	fulladdress_dobunique_count_forssn_name_60
ssn_zip5_count_7	fulladdress_dobunique_count_forssn_fulladdress_1
ssn_zip5_count_14	fulladdress_dobunique_count_forssn_fulladdress_3
ssn_zip5_count_30	fulladdress_dobunique_count_forssn_fulladdress_7
ssn_dob_day_since	fulladdress_dobunique_count_forssn_fulladdress_14
ssn_dob_count_0	fulladdress_dobunique_count_forssn_fulladdress_30
ssn_dob_count_1	fulladdress_dobunique_count_forssn_fulladdress_60
ssn_dob_count_3	fulladdress_dobunique_count_forssn_name_dob_1
ssn_dob_count_7	fulladdress_dobunique_count_forssn_name_dob_3
ssn_dob_count_14	fulladdress_dobunique_count_forssn_name_dob_7
ssn_dob_count_30	fulladdress_dobunique_count_forssn_name_dob_14
ssn_homephone_day_since	fulladdress_dobunique_count_forssn_name_dob_30
ssn_homephone_count_0	fulladdress_dobunique_count_forssn_name_dob_60
ssn_homephone_count_1	fulladdress_dobunique_count_forssn_name_fulladdress_1
ssn_homephone_count_3	fulladdress_dobunique_count_forssn_name_fulladdress_3
ssn_homephone_count_7	fulladdress_dobunique_count_forssn_name_fulladdress_7
ssn_homephone_count_14	fulladdress_dobunique_count_forssn_name_fulladdress_14
ssn_homephone_count_30	fulladdress_dobunique_count_forssn_name_fulladdress_30
ssn_name_day_since	fulladdress_dobunique_count_forssn_name_fulladdress_60
ssn_name_count_0	dob_homephoneunique_count_forssn_1
ssn_name_count_1	dob_homephoneunique_count_forssn_3
ssn_name_count_3	dob_homephoneunique_count_forssn_7
ssn_name_count_7	dob_homephoneunique_count_forssn_14
ssn_name_count_14	dob_homephoneunique_count_forssn_30
ssn_name_count_30	dob_homephoneunique_count_forssn_60
ssn_fulladdress_day_since	dob_homephoneunique_count_forfulladdress_1
ssn_fulladdress_count_0	dob_homephoneunique_count_forfulladdress_3
ssn_fulladdress_count_1	dob_homephoneunique_count_forfulladdress_7
ssn_fulladdress_count_3	dob_homephoneunique_count_forfulladdress_14
ssn_fulladdress_count_7	dob_homephoneunique_count_forfulladdress_30
ssn_fulladdress_count_14	dob_homephoneunique_count_forfulladdress_60
ssn_fulladdress_count_30	dob_homephoneunique_count_forname_dob_1
ssn_name_dob_day_since	dob_homephoneunique_count_forname_dob_3
ssn_name_dob_count_0	dob_homephoneunique_count_forname_dob_7
ssn_name_dob_count_1	dob_homephoneunique_count_forname_dob_14
ssn_name_dob_count_3	dob_homephoneunique_count_forname_dob_30
ssn_name_dob_count_7	dob_homephoneunique_count_forname_dob_60
ssn_name_dob_count_14	dob_homephoneunique_count_forname_fulladdress_1
ssn_name_dob_count_30	dob_homephoneunique_count_forname_fulladdress_3
ssn_name_fulladdress_day_since	dob_homephoneunique_count_forname_fulladdress_7
ssn_name_fulladdress_count_0	dob_homephoneunique_count_forname_fulladdress_14
ssn_name_fulladdress_count_1	dob_homephoneunique_count_forname_fulladdress_30
ssn_name_fulladdress_count_3	dob_homephoneunique_count_forname_fulladdress_60

ssn_name_fulladdress_count_7	dob_homephoneunique_count_forfulladdress_dob_1
ssn_name_fulladdress_count_14	dob_homephoneunique_count_forfulladdress_dob_3
ssn_name_fulladdress_count_30	dob_homephoneunique_count_forfulladdress_dob_7
ssn_name_homephone_day_since	dob_homephoneunique_count_forfulladdress_dob_14
ssn_name_homephone_count_0	dob_homephoneunique_count_forfulladdress_dob_30
ssn_name_homephone_count_1	dob_homephoneunique_count_forfulladdress_dob_60
ssn_name_homephone_count_3	dob_homephoneunique_count_forssn_lastname_1
ssn_name_homephone_count_7	dob_homephoneunique_count_forssn_lastname_3
ssn_name_homephone_count_14	dob_homephoneunique_count_forssn_lastname_7
ssn_name_homephone_count_30	dob_homephoneunique_count_forssn_lastname_14
ssn_fulladdress_dob_day_since	dob_homephoneunique_count_forssn_lastname_30
ssn_fulladdress_dob_count_0	dob_homephoneunique_count_forssn_lastname_60
ssn_fulladdress_dob_count_1	dob_homephoneunique_count_forssn_zip5_1
ssn_fulladdress_dob_count_3	dob_homephoneunique_count_forssn_zip5_3
ssn_fulladdress_dob_count_7	dob_homephoneunique_count_forssn_zip5_7
ssn_fulladdress_dob_count_14	dob_homephoneunique_count_forssn_zip5_14
ssn_fulladdress_dob_count_30	dob_homephoneunique_count_forssn_zip5_30
ssn_fulladdress_homephone_day_since	dob_homephoneunique_count_forssn_zip5_60
ssn_fulladdress_homephone_count_0	dob_homephoneunique_count_forssn_name_1
ssn_fulladdress_homephone_count_1	dob_homephoneunique_count_forssn_name_3
ssn_fulladdress_homephone_count_3	dob_homephoneunique_count_forssn_name_7
ssn_fulladdress_homephone_count_7	dob_homephoneunique_count_forssn_name_14
ssn_fulladdress_homephone_count_14	dob_homephoneunique_count_forssn_name_30
ssn_fulladdress_homephone_count_30	dob_homephoneunique_count_forssn_name_60
ssn_dob_homephone_day_since	dob_homephoneunique_count_forssn_fulladdress_1
ssn_dob_homephone_count_0	dob_homephoneunique_count_forssn_fulladdress_3
ssn_dob_homephone_count_1	dob_homephoneunique_count_forssn_fulladdress_7
ssn_dob_homephone_count_3	dob_homephoneunique_count_forssn_fulladdress_14
ssn_dob_homephone_count_7	dob_homephoneunique_count_forssn_fulladdress_30
ssn_dob_homephone_count_14	dob_homephoneunique_count_forssn_fulladdress_60
ssn_dob_homephone_count_30	dob_homephoneunique_count_forssn_name_dob_1
ssn_homephone_name_dob_day_since	dob_homephoneunique_count_forssn_name_dob_3
ssn_homephone_name_dob_count_0	dob_homephoneunique_count_forssn_name_dob_7
ssn_homephone_name_dob_count_1	dob_homephoneunique_count_forssn_name_dob_14
ssn_homephone_name_dob_count_3	dob_homephoneunique_count_forssn_name_dob_30
ssn_homephone_name_dob_count_7	dob_homephoneunique_count_forssn_name_dob_60
ssn_homephone_name_dob_count_14	dob_homephoneunique_count_forssn_name_fulladdress_1
ssn_homephone_name_dob_count_30	dob_homephoneunique_count_forssn_name_fulladdress_3
ssn_fulladdress_name_dob_day_since	dob_homephoneunique_count_forssn_name_fulladdress_7
ssn_fulladdress_name_dob_count_0	dob_homephoneunique_count_forssn_name_fulladdress_14
ssn_fulladdress_name_dob_count_1	dob_homephoneunique_count_forssn_name_fulladdress_30
ssn_fulladdress_name_dob_count_3	dob_homephoneunique_count_forssn_name_fulladdress_60
ssn_fulladdress_name_dob_count_7	ssn_lastnameunique_count_forssn_1

ssn_fulladdress_name_dob_count_14	ssn_lastnameunique_count_forssn_3
ssn_fulladdress_name_dob_count_30	ssn_lastnameunique_count_forssn_7
ssn_count_0_by_3	ssn_lastnameunique_count_forssn_14
ssn_count_0_by_7	ssn_lastnameunique_count_forssn_30
ssn_count_0_by_14	ssn_lastnameunique_count_forssn_60
ssn_count_0_by_30	ssn_lastnameunique_count_forfulladdress_1
ssn_count_1_by_3	ssn_lastnameunique_count_forfulladdress_3
ssn_count_1_by_7	ssn_lastnameunique_count_forfulladdress_7
ssn_count_1_by_14	ssn_lastnameunique_count_forfulladdress_14
ssn_count_1_by_30	ssn_lastnameunique_count_forfulladdress_30
address_count_0_by_3	ssn_lastnameunique_count_forfulladdress_60
address_count_0_by_7	ssn_lastnameunique_count_forname_dob_1
address_count_0_by_14	ssn_lastnameunique_count_forname_dob_3
address_count_0_by_30	ssn_lastnameunique_count_forname_dob_7
address_count_1_by_3	ssn_lastnameunique_count_forname_dob_14
address_count_1_by_7	ssn_lastnameunique_count_forname_dob_30
address_count_1_by_14	ssn_lastnameunique_count_forname_dob_60
address_count_1_by_30	ssn_lastnameunique_count_forname_fulladdress_1
dob_count_0_by_3	ssn_lastnameunique_count_forname_fulladdress_3
dob_count_0_by_7	ssn_lastnameunique_count_forname_fulladdress_7
dob_count_0_by_14	ssn_lastnameunique_count_forname_fulladdress_14
dob_count_0_by_30	ssn_lastnameunique_count_forname_fulladdress_30
dob_count_1_by_3	ssn_lastnameunique_count_forname_fulladdress_60
dob_count_1_by_7	ssn_lastnameunique_count_forfulladdress_dob_1
dob_count_1_by_14	ssn_lastnameunique_count_forfulladdress_dob_3
dob_count_1_by_30	ssn_lastnameunique_count_forfulladdress_dob_7
homephone_count_0_by_3	ssn_lastnameunique_count_forfulladdress_dob_14
homephone_count_0_by_7	ssn_lastnameunique_count_forfulladdress_dob_30
homephone_count_0_by_14	ssn_lastnameunique_count_forfulladdress_dob_60
homephone_count_0_by_30	ssn_lastnameunique_count_fordob_homephone_1
homephone_count_1_by_3	ssn_lastnameunique_count_fordob_homephone_3
homephone_count_1_by_7	ssn_lastnameunique_count_fordob_homephone_7
homephone_count_1_by_14	ssn_lastnameunique_count_fordob_homephone_14
homephone_count_1_by_30	ssn_lastnameunique_count_fordob_homephone_30
name_count_0_by_3	ssn_lastnameunique_count_fordob_homephone_60
name_count_0_by_7	ssn_lastnameunique_count_forssn_zip5_1
name_count_0_by_14	ssn_lastnameunique_count_forssn_zip5_3
name_count_0_by_30	ssn_lastnameunique_count_forssn_zip5_7
name_count_1_by_3	ssn_lastnameunique_count_forssn_zip5_14
name_count_1_by_7	ssn_lastnameunique_count_forssn_zip5_30
name_count_1_by_14	ssn_lastnameunique_count_forssn_zip5_60
name_count_1_by_30	ssn_lastnameunique_count_forssn_name_1
fulladdress_count_0_by_3	ssn_lastnameunique_count_forssn_name_3

fulladdress_count_0_by_7	ssn_lastnameunique_count_forssn_name_7
fulladdress_count_0_by_14	ssn_lastnameunique_count_forssn_name_14
fulladdress_count_0_by_30	ssn_lastnameunique_count_forssn_name_30
fulladdress_count_1_by_3	ssn_lastnameunique_count_forssn_name_60
fulladdress_count_1_by_7	ssn_lastnameunique_count_forssn_fulladdress_1
fulladdress_count_1_by_14	ssn_lastnameunique_count_forssn_fulladdress_3
fulladdress_count_1_by_30	ssn_lastnameunique_count_forssn_fulladdress_7
name_dob_count_0_by_3	ssn_lastnameunique_count_forssn_fulladdress_14
name_dob_count_0_by_7	ssn_lastnameunique_count_forssn_fulladdress_30
name_dob_count_0_by_14	ssn_lastnameunique_count_forssn_fulladdress_60
name_dob_count_0_by_30	ssn_lastnameunique_count_forssn_name_dob_1
name_dob_count_1_by_3	ssn_lastnameunique_count_forssn_name_dob_3
name_dob_count_1_by_7	ssn_lastnameunique_count_forssn_name_dob_7
name_dob_count_1_by_14	ssn_lastnameunique_count_forssn_name_dob_14
name_dob_count_1_by_30	ssn_lastnameunique_count_forssn_name_dob_30
name_fulladdress_count_0_by_3	ssn_lastnameunique_count_forssn_name_dob_60
name_fulladdress_count_0_by_7	ssn_lastnameunique_count_forssn_name_fulladdress_1
name_fulladdress_count_0_by_14	ssn_lastnameunique_count_forssn_name_fulladdress_3
name_fulladdress_count_0_by_30	ssn_lastnameunique_count_forssn_name_fulladdress_7
name_fulladdress_count_1_by_3	ssn_lastnameunique_count_forssn_name_fulladdress_14
name_fulladdress_count_1_by_7	ssn_lastnameunique_count_forssn_name_fulladdress_30
name_fulladdress_count_1_by_14	ssn_lastnameunique_count_forssn_name_fulladdress_60
name_fulladdress_count_1_by_30	ssn_zip5unique_count_forssn_1
name_homephone_count_0_by_3	ssn_zip5unique_count_forssn_3
name_homephone_count_0_by_7	ssn_zip5unique_count_forssn_7
name_homephone_count_0_by_14	ssn_zip5unique_count_forssn_14
name_homephone_count_0_by_30	ssn_zip5unique_count_forssn_30
name_homephone_count_1_by_3	ssn_zip5unique_count_forssn_60
name_homephone_count_1_by_7	ssn_zip5unique_count_forfulladdress_1
name_homephone_count_1_by_14	ssn_zip5unique_count_forfulladdress_3
name_homephone_count_1_by_30	ssn_zip5unique_count_forfulladdress_7
fulladdress_dob_count_0_by_3	ssn_zip5unique_count_forfulladdress_14
fulladdress_dob_count_0_by_7	ssn_zip5unique_count_forfulladdress_30
fulladdress_dob_count_0_by_14	ssn_zip5unique_count_forfulladdress_60
fulladdress_dob_count_0_by_30	ssn_zip5unique_count_forname_dob_1
fulladdress_dob_count_1_by_3	ssn_zip5unique_count_forname_dob_3
fulladdress_dob_count_1_by_7	ssn_zip5unique_count_forname_dob_7
fulladdress_dob_count_1_by_14	ssn_zip5unique_count_forname_dob_14
fulladdress_dob_count_1_by_30	ssn_zip5unique_count_forname_dob_30
fulladdress_homephone_count_0_by_3	ssn_zip5unique_count_forname_dob_60
fulladdress_homephone_count_0_by_7	ssn_zip5unique_count_forname_fulladdress_1
fulladdress_homephone_count_0_by_14	ssn_zip5unique_count_forname_fulladdress_3
fulladdress_homephone_count_0_by_30	ssn_zip5unique_count_forname_fulladdress_7

fulladdress_homephone_count_1_by_3	ssn_zip5unique_count_forname_fulladdress_14
fulladdress_homephone_count_1_by_7	ssn_zip5unique_count_forname_fulladdress_30
fulladdress_homephone_count_1_by_14	ssn_zip5unique_count_forname_fulladdress_60
fulladdress_homephone_count_1_by_30	ssn_zip5unique_count_forfulladdress_dob_1
dob_homephone_count_0_by_3	ssn_zip5unique_count_forfulladdress_dob_3
dob_homephone_count_0_by_7	ssn_zip5unique_count_forfulladdress_dob_7
dob_homephone_count_0_by_14	ssn_zip5unique_count_forfulladdress_dob_14
dob_homephone_count_0_by_30	ssn_zip5unique_count_forfulladdress_dob_30
dob_homephone_count_1_by_3	ssn_zip5unique_count_forfulladdress_dob_60
dob_homephone_count_1_by_7	ssn_zip5unique_count_fordob_homephone_1
dob_homephone_count_1_by_14	ssn_zip5unique_count_fordob_homephone_3
dob_homephone_count_1_by_30	ssn_zip5unique_count_fordob_homephone_7
homephone_name_dob_count_0_by_3	ssn_zip5unique_count_fordob_homephone_14
homephone_name_dob_count_0_by_7	ssn_zip5unique_count_fordob_homephone_30
homephone_name_dob_count_0_by_14	ssn_zip5unique_count_fordob_homephone_60
homephone_name_dob_count_0_by_30	ssn_zip5unique_count_forssn_lastname_1
homephone_name_dob_count_1_by_3	ssn_zip5unique_count_forssn_lastname_3
homephone_name_dob_count_1_by_7	ssn_zip5unique_count_forssn_lastname_7
homephone_name_dob_count_1_by_14	ssn_zip5unique_count_forssn_lastname_14
homephone_name_dob_count_1_by_30	ssn_zip5unique_count_forssn_lastname_30
fulladdress_name_dob_count_0_by_3	ssn_zip5unique_count_forssn_lastname_60
fulladdress_name_dob_count_0_by_7	ssn_zip5unique_count_forssn_name_1
fulladdress_name_dob_count_0_by_14	ssn_zip5unique_count_forssn_name_3
fulladdress_name_dob_count_0_by_30	ssn_zip5unique_count_forssn_name_7
fulladdress_name_dob_count_1_by_3	ssn_zip5unique_count_forssn_name_14
fulladdress_name_dob_count_1_by_7	ssn_zip5unique_count_forssn_name_30
fulladdress_name_dob_count_1_by_14	ssn_zip5unique_count_forssn_name_60
fulladdress_name_dob_count_1_by_30	ssn_zip5unique_count_forssn_fulladdress_1
name_zip5_count_0_by_3	ssn_zip5unique_count_forssn_fulladdress_3
name_zip5_count_0_by_7	ssn_zip5unique_count_forssn_fulladdress_7
name_zip5_count_0_by_14	ssn_zip5unique_count_forssn_fulladdress_14
name_zip5_count_0_by_30	ssn_zip5unique_count_forssn_fulladdress_30
name_zip5_count_1_by_3	ssn_zip5unique_count_forssn_fulladdress_60
name_zip5_count_1_by_7	ssn_zip5unique_count_forssn_name_dob_1
name_zip5_count_1_by_14	ssn_zip5unique_count_forssn_name_dob_3
name_zip5_count_1_by_30	ssn_zip5unique_count_forssn_name_dob_7
name_address_count_0_by_3	ssn_zip5unique_count_forssn_name_dob_14
name_address_count_0_by_7	ssn_zip5unique_count_forssn_name_dob_30
name_address_count_0_by_14	ssn_zip5unique_count_forssn_name_dob_60
name_address_count_0_by_30	ssn_zip5unique_count_forssn_name_fulladdress_1
name_address_count_1_by_3	ssn_zip5unique_count_forssn_name_fulladdress_3
name_address_count_1_by_7	ssn_zip5unique_count_forssn_name_fulladdress_7
name_address_count_1_by_14	ssn_zip5unique_count_forssn_name_fulladdress_14

name_address_count_1_by_30	ssn_zip5unique_count_forssn_name_fulladdress_30
firstname_dob_count_0_by_3	ssn_zip5unique_count_forssn_name_fulladdress_60
firstname_dob_count_0_by_7	ssn_nameunique_count_forssn_1
firstname_dob_count_0_by_14	ssn_nameunique_count_forssn_3
firstname_dob_count_0_by_30	ssn_nameunique_count_forssn_7
firstname_dob_count_1_by_3	ssn_nameunique_count_forssn_14
firstname_dob_count_1_by_7	ssn_nameunique_count_forssn_30
firstname_dob_count_1_by_14	ssn_nameunique_count_forssn_60
firstname_dob_count_1_by_30	ssn_nameunique_count_forfulladdress_1
firstname_address_count_0_by_3	ssn_nameunique_count_forfulladdress_3
firstname_address_count_0_by_7	ssn_nameunique_count_forfulladdress_7
firstname_address_count_0_by_14	ssn_nameunique_count_forfulladdress_14
firstname_address_count_0_by_30	ssn_nameunique_count_forfulladdress_30
firstname_address_count_1_by_3	ssn_nameunique_count_forfulladdress_60
firstname_address_count_1_by_7	ssn_nameunique_count_forname_dob_1
firstname_address_count_1_by_14	ssn_nameunique_count_forname_dob_3
firstname_address_count_1_by_30	ssn_nameunique_count_forname_dob_7
firstname_zip5_count_0_by_3	ssn_nameunique_count_forname_dob_14
firstname_zip5_count_0_by_7	ssn_nameunique_count_forname_dob_30
firstname_zip5_count_0_by_14	ssn_nameunique_count_forname_dob_60
firstname_zip5_count_0_by_30	ssn_nameunique_count_forname_fulladdress_1
firstname_zip5_count_1_by_3	ssn_nameunique_count_forname_fulladdress_3
firstname_zip5_count_1_by_7	ssn_nameunique_count_forname_fulladdress_7
firstname_zip5_count_1_by_14	ssn_nameunique_count_forname_fulladdress_14
firstname_zip5_count_1_by_30	ssn_nameunique_count_forname_fulladdress_30
firstname_homephone_count_0_by_3	ssn_nameunique_count_forname_fulladdress_60
firstname_homephone_count_0_by_7	ssn_nameunique_count_forfulladdress_dob_1
firstname_homephone_count_0_by_14	ssn_nameunique_count_forfulladdress_dob_3
firstname_homephone_count_0_by_30	ssn_nameunique_count_forfulladdress_dob_7
firstname_homephone_count_1_by_3	ssn_nameunique_count_forfulladdress_dob_14
firstname_homephone_count_1_by_7	ssn_nameunique_count_forfulladdress_dob_30
firstname_homephone_count_1_by_14	ssn_nameunique_count_forfulladdress_dob_60
firstname_homephone_count_1_by_30	ssn_nameunique_count_fordob_homephone_1
firstname_fulladdress_count_0_by_3	ssn_nameunique_count_fordob_homephone_3
firstname_fulladdress_count_0_by_7	ssn_nameunique_count_fordob_homephone_7
firstname_fulladdress_count_0_by_14	ssn_nameunique_count_fordob_homephone_14
firstname_fulladdress_count_0_by_30	ssn_nameunique_count_fordob_homephone_30
firstname_fulladdress_count_1_by_3	ssn_nameunique_count_fordob_homephone_60
firstname_fulladdress_count_1_by_7	ssn_nameunique_count_forssn_lastname_1
firstname_fulladdress_count_1_by_14	ssn_nameunique_count_forssn_lastname_3
firstname_fulladdress_count_1_by_30	ssn_nameunique_count_forssn_lastname_7
lastname_dob_count_0_by_3	ssn_nameunique_count_forssn_lastname_14
lastname_dob_count_0_by_7	ssn_nameunique_count_forssn_lastname_30

lastname_dob_count_0_by_14	ssn_nameunique_count_forssn_lastname_60
lastname_dob_count_0_by_30	ssn_nameunique_count_forssn_zip5_1
lastname_dob_count_1_by_3	ssn_nameunique_count_forssn_zip5_3
lastname_dob_count_1_by_7	ssn_nameunique_count_forssn_zip5_7
lastname_dob_count_1_by_14	ssn_nameunique_count_forssn_zip5_14
lastname_dob_count_1_by_30	ssn_nameunique_count_forssn_zip5_30
lastname_address_count_0_by_3	ssn_nameunique_count_forssn_zip5_60
lastname_address_count_0_by_7	ssn_nameunique_count_forssn_fulladdress_1
lastname_address_count_0_by_14	ssn_nameunique_count_forssn_fulladdress_3
lastname_address_count_0_by_30	ssn_nameunique_count_forssn_fulladdress_7
lastname_address_count_1_by_3	ssn_nameunique_count_forssn_fulladdress_14
lastname_address_count_1_by_7	ssn_nameunique_count_forssn_fulladdress_30
lastname_address_count_1_by_14	ssn_nameunique_count_forssn_fulladdress_60
lastname_address_count_1_by_30	ssn_nameunique_count_forssn_name_dob_1
lastname_zip5_count_0_by_3	ssn_nameunique_count_forssn_name_dob_3
lastname_zip5_count_0_by_7	ssn_nameunique_count_forssn_name_dob_7
lastname_zip5_count_0_by_14	ssn_nameunique_count_forssn_name_dob_14
lastname_zip5_count_0_by_30	ssn_nameunique_count_forssn_name_dob_30
lastname_zip5_count_1_by_3	ssn_nameunique_count_forssn_name_dob_60
lastname_zip5_count_1_by_7	ssn_nameunique_count_forssn_name_fulladdress_1
lastname_zip5_count_1_by_14	ssn_nameunique_count_forssn_name_fulladdress_3
lastname_zip5_count_1_by_30	ssn_nameunique_count_forssn_name_fulladdress_7
lastname_homephone_count_0_by_3	ssn_nameunique_count_forssn_name_fulladdress_14
lastname_homephone_count_0_by_7	ssn_nameunique_count_forssn_name_fulladdress_30
lastname_homephone_count_0_by_14	ssn_nameunique_count_forssn_name_fulladdress_60
lastname_homephone_count_0_by_30	ssn_fulladdressunique_count_forssn_1
lastname_homephone_count_1_by_3	ssn_fulladdressunique_count_forssn_3
lastname_homephone_count_1_by_7	ssn_fulladdressunique_count_forssn_7
lastname_homephone_count_1_by_14	ssn_fulladdressunique_count_forssn_14
lastname_homephone_count_1_by_30	ssn_fulladdressunique_count_forssn_30
lastname_fulladdress_count_0_by_3	ssn_fulladdressunique_count_forssn_60
lastname_fulladdress_count_0_by_7	ssn_fulladdressunique_count_forfulladdress_1
lastname_fulladdress_count_0_by_14	ssn_fulladdressunique_count_forfulladdress_3
lastname_fulladdress_count_0_by_30	ssn_fulladdressunique_count_forfulladdress_7
lastname_fulladdress_count_1_by_3	ssn_fulladdressunique_count_forfulladdress_14
lastname_fulladdress_count_1_by_7	ssn_fulladdressunique_count_forfulladdress_30
lastname_fulladdress_count_1_by_14	ssn_fulladdressunique_count_forfulladdress_60
lastname_fulladdress_count_1_by_30	ssn_fulladdressunique_count_forname_dob_1
ssn_firstname_count_0_by_3	ssn_fulladdressunique_count_forname_dob_3
ssn_firstname_count_0_by_7	ssn_fulladdressunique_count_forname_dob_7
ssn_firstname_count_0_by_14	ssn_fulladdressunique_count_forname_dob_14
ssn_firstname_count_0_by_30	ssn_fulladdressunique_count_forname_dob_30
ssn_firstname_count_1_by_3	ssn_fulladdressunique_count_forname_dob_60

ssn_firstname_count_1_by_7	ssn_fulladdressunique_count_forname_fulladdress_1
ssn_firstname_count_1_by_14	ssn_fulladdressunique_count_forname_fulladdress_3
ssn_firstname_count_1_by_30	ssn_fulladdressunique_count_forname_fulladdress_7
ssn_lastname_count_0_by_3	ssn_fulladdressunique_count_forname_fulladdress_14
ssn_lastname_count_0_by_7	ssn_fulladdressunique_count_forname_fulladdress_30
ssn_lastname_count_0_by_14	ssn_fulladdressunique_count_forname_fulladdress_60
ssn_lastname_count_0_by_30	ssn_fulladdressunique_count_forfulladdress_dob_1
ssn_lastname_count_1_by_3	ssn_fulladdressunique_count_forfulladdress_dob_3
ssn_lastname_count_1_by_7	ssn_fulladdressunique_count_forfulladdress_dob_7
ssn_lastname_count_1_by_14	ssn_fulladdressunique_count_forfulladdress_dob_14
ssn_lastname_count_1_by_30	ssn_fulladdressunique_count_forfulladdress_dob_30
ssn_address_count_0_by_3	ssn_fulladdressunique_count_forfulladdress_dob_60
ssn_address_count_0_by_7	ssn_fulladdressunique_count_fordob_homephone_1
ssn_address_count_0_by_14	ssn_fulladdressunique_count_fordob_homephone_3
ssn_address_count_0_by_30	ssn_fulladdressunique_count_fordob_homephone_7
ssn_address_count_1_by_3	ssn_fulladdressunique_count_fordob_homephone_14
ssn_address_count_1_by_7	ssn_fulladdressunique_count_fordob_homephone_30
ssn_address_count_1_by_14	ssn_fulladdressunique_count_fordob_homephone_60
ssn_address_count_1_by_30	ssn_fulladdressunique_count_forssn_lastname_1
ssn_zip5_count_0_by_3	ssn_fulladdressunique_count_forssn_lastname_3
ssn_zip5_count_0_by_7	ssn_fulladdressunique_count_forssn_lastname_7
ssn_zip5_count_0_by_14	ssn_fulladdressunique_count_forssn_lastname_14
ssn_zip5_count_0_by_30	ssn_fulladdressunique_count_forssn_lastname_30
ssn_zip5_count_1_by_3	ssn_fulladdressunique_count_forssn_lastname_60
ssn_zip5_count_1_by_7	ssn_fulladdressunique_count_forssn_zip5_1
ssn_zip5_count_1_by_14	ssn_fulladdressunique_count_forssn_zip5_3
ssn_zip5_count_1_by_30	ssn_fulladdressunique_count_forssn_zip5_7
ssn_dob_count_0_by_3	ssn_fulladdressunique_count_forssn_zip5_14
ssn_dob_count_0_by_7	ssn_fulladdressunique_count_forssn_zip5_30
ssn_dob_count_0_by_14	ssn_fulladdressunique_count_forssn_zip5_60
ssn_dob_count_0_by_30	ssn_fulladdressunique_count_forssn_name_1
ssn_dob_count_1_by_3	ssn_fulladdressunique_count_forssn_name_3
ssn_dob_count_1_by_7	ssn_fulladdressunique_count_forssn_name_7
ssn_dob_count_1_by_14	ssn_fulladdressunique_count_forssn_name_14
ssn_dob_count_1_by_30	ssn_fulladdressunique_count_forssn_name_30
ssn_homephone_count_0_by_3	ssn_fulladdressunique_count_forssn_name_60
ssn_homephone_count_0_by_7	ssn_fulladdressunique_count_forssn_name_dob_1
ssn_homephone_count_0_by_14	ssn_fulladdressunique_count_forssn_name_dob_3
ssn_homephone_count_0_by_30	ssn_fulladdressunique_count_forssn_name_dob_7
ssn_homephone_count_1_by_3	ssn_fulladdressunique_count_forssn_name_dob_14
ssn_homephone_count_1_by_7	ssn_fulladdressunique_count_forssn_name_dob_30
ssn_homephone_count_1_by_14	ssn_fulladdressunique_count_forssn_name_dob_60
ssn_homephone_count_1_by_30	ssn_fulladdressunique_count_forssn_name_fulladdress_1

ssn_name_count_0_by_3	ssn_fulladdressunique_count_forssn_name_fulladdress_3
ssn_name_count_0_by_7	ssn_fulladdressunique_count_forssn_name_fulladdress_7
ssn_name_count_0_by_14	ssn_fulladdressunique_count_forssn_name_fulladdress_14
ssn_name_count_0_by_30	ssn_fulladdressunique_count_forssn_name_fulladdress_30
ssn_name_count_1_by_3	ssn_fulladdressunique_count_forssn_name_fulladdress_60
ssn_name_count_1_by_7	ssn_name_dobunique_count_forssn_1
ssn_name_count_1_by_14	ssn_name_dobunique_count_forssn_3
ssn_name_count_1_by_30	ssn_name_dobunique_count_forssn_7
ssn_fulladdress_count_0_by_3	ssn_name_dobunique_count_forssn_14
ssn_fulladdress_count_0_by_7	ssn_name_dobunique_count_forssn_30
ssn_fulladdress_count_0_by_14	ssn_name_dobunique_count_forssn_60
ssn_fulladdress_count_0_by_30	ssn_name_dobunique_count_forfulladdress_1
ssn_fulladdress_count_1_by_3	ssn_name_dobunique_count_forfulladdress_3
ssn_fulladdress_count_1_by_7	ssn_name_dobunique_count_forfulladdress_7
ssn_fulladdress_count_1_by_14	ssn_name_dobunique_count_forfulladdress_14
ssn_fulladdress_count_1_by_30	ssn_name_dobunique_count_forfulladdress_30
ssn_name_dob_count_0_by_3	ssn_name_dobunique_count_forfulladdress_60
ssn_name_dob_count_0_by_7	ssn_name_dobunique_count_forname_dob_1
ssn_name_dob_count_0_by_14	ssn_name_dobunique_count_forname_dob_3
ssn_name_dob_count_0_by_30	ssn_name_dobunique_count_forname_dob_7
ssn_name_dob_count_1_by_3	ssn_name_dobunique_count_forname_dob_14
ssn_name_dob_count_1_by_7	ssn_name_dobunique_count_forname_dob_30
ssn_name_dob_count_1_by_14	ssn_name_dobunique_count_forname_dob_60
ssn_name_dob_count_1_by_30	ssn_name_dobunique_count_forname_fulladdress_1
ssn_name_fulladdress_count_0_by_3	ssn_name_dobunique_count_forname_fulladdress_3
ssn_name_fulladdress_count_0_by_7	ssn_name_dobunique_count_forname_fulladdress_7
ssn_name_fulladdress_count_0_by_14	ssn_name_dobunique_count_forname_fulladdress_14
ssn_name_fulladdress_count_0_by_30	ssn_name_dobunique_count_forname_fulladdress_30
ssn_name_fulladdress_count_1_by_3	ssn_name_dobunique_count_forname_fulladdress_60
ssn_name_fulladdress_count_1_by_7	ssn_name_dobunique_count_forfulladdress_dob_1
ssn_name_fulladdress_count_1_by_14	ssn_name_dobunique_count_forfulladdress_dob_3
ssn_name_fulladdress_count_1_by_30	ssn_name_dobunique_count_forfulladdress_dob_7
ssn_name_homephone_count_0_by_3	ssn_name_dobunique_count_forfulladdress_dob_14
ssn_name_homephone_count_0_by_7	ssn_name_dobunique_count_forfulladdress_dob_30
ssn_name_homephone_count_0_by_14	ssn_name_dobunique_count_forfulladdress_dob_60
ssn_name_homephone_count_0_by_30	ssn_name_dobunique_count_fordob_homephone_1
ssn_name_homephone_count_1_by_3	ssn_name_dobunique_count_fordob_homephone_3
ssn_name_homephone_count_1_by_7	ssn_name_dobunique_count_fordob_homephone_7
ssn_name_homephone_count_1_by_14	ssn_name_dobunique_count_fordob_homephone_14
ssn_name_homephone_count_1_by_30	ssn_name_dobunique_count_fordob_homephone_30
ssn_fulladdress_dob_count_0_by_3	ssn_name_dobunique_count_fordob_homephone_60
ssn_fulladdress_dob_count_0_by_7	ssn_name_dobunique_count_forssn_lastname_1
ssn_fulladdress_dob_count_0_by_14	ssn_name_dobunique_count_forssn_lastname_3

ssn_fulladdress_dob_count_0_by_30	ssn_name_dobunique_count_forssn_lastname_7
ssn_fulladdress_dob_count_1_by_3	ssn_name_dobunique_count_forssn_lastname_14
ssn_fulladdress_dob_count_1_by_7	ssn_name_dobunique_count_forssn_lastname_30
ssn_fulladdress_dob_count_1_by_14	ssn_name_dobunique_count_forssn_lastname_60
ssn_fulladdress_dob_count_1_by_30	ssn_name_dobunique_count_forssn_zip5_1
ssn_fulladdress_homephone_count_0_by_3	ssn_name_dobunique_count_forssn_zip5_3
ssn_fulladdress_homephone_count_0_by_7	ssn_name_dobunique_count_forssn_zip5_7
ssn_fulladdress_homephone_count_0_by_14	ssn_name_dobunique_count_forssn_zip5_14
ssn_fulladdress_homephone_count_0_by_30	ssn_name_dobunique_count_forssn_zip5_30
ssn_fulladdress_homephone_count_1_by_3	ssn_name_dobunique_count_forssn_zip5_60
ssn_fulladdress_homephone_count_1_by_7	ssn_name_dobunique_count_forssn_name_1
ssn_fulladdress_homephone_count_1_by_14	ssn_name_dobunique_count_forssn_name_3
ssn_fulladdress_homephone_count_1_by_30	ssn_name_dobunique_count_forssn_name_7
ssn_dob_homephone_count_0_by_3	ssn_name_dobunique_count_forssn_name_14
ssn_dob_homephone_count_0_by_7	ssn_name_dobunique_count_forssn_name_30
ssn_dob_homephone_count_0_by_14	ssn_name_dobunique_count_forssn_name_60
ssn_dob_homephone_count_0_by_30	ssn_name_dobunique_count_forssn_fulladdress_1
ssn_dob_homephone_count_1_by_3	ssn_name_dobunique_count_forssn_fulladdress_3
ssn_dob_homephone_count_1_by_7	ssn_name_dobunique_count_forssn_fulladdress_7
ssn_dob_homephone_count_1_by_14	ssn_name_dobunique_count_forssn_fulladdress_14
ssn_dob_homephone_count_1_by_30	ssn_name_dobunique_count_forssn_fulladdress_30
ssn_homephone_name_dob_count_0_by_3	ssn_name_dobunique_count_forssn_fulladdress_60
ssn_homephone_name_dob_count_0_by_7	ssn_name_dobunique_count_forssn_name_fulladdress_1
ssn_homephone_name_dob_count_0_by_14	ssn_name_dobunique_count_forssn_name_fulladdress_3
ssn_homephone_name_dob_count_0_by_30	ssn_name_dobunique_count_forssn_name_fulladdress_7
ssn_homephone_name_dob_count_1_by_3	ssn_name_dobunique_count_forssn_name_fulladdress_14
ssn_homephone_name_dob_count_1_by_7	ssn_name_dobunique_count_forssn_name_fulladdress_30
ssn_homephone_name_dob_count_1_by_14	ssn_name_dobunique_count_forssn_name_fulladdress_60
ssn_homephone_name_dob_count_1_by_30	ssn_name_fulladdressunique_count_forssn_1
ssn_fulladdress_name_dob_count_0_by_3	ssn_name_fulladdressunique_count_forssn_3
ssn_fulladdress_name_dob_count_0_by_7	ssn_name_fulladdressunique_count_forssn_7
ssn_fulladdress_name_dob_count_0_by_14	ssn_name_fulladdressunique_count_forssn_14
ssn_fulladdress_name_dob_count_0_by_30	ssn_name_fulladdressunique_count_forssn_30
ssn_fulladdress_name_dob_count_1_by_3	ssn_name_fulladdressunique_count_forssn_60
ssn_fulladdress_name_dob_count_1_by_7	ssn_name_fulladdressunique_count_forfulladdress_1
ssn_fulladdress_name_dob_count_1_by_14	ssn_name_fulladdressunique_count_forfulladdress_3
ssn_fulladdress_name_dob_count_1_by_30	ssn_name_fulladdressunique_count_forfulladdress_7
ssnunique_count_forfulladdress_1	ssn_name_fulladdressunique_count_forfulladdress_14
ssnunique_count_forfulladdress_3	ssn_name_fulladdressunique_count_forfulladdress_30
ssnunique_count_forfulladdress_7	ssn_name_fulladdressunique_count_forfulladdress_60
ssnunique_count_forfulladdress_14	ssn_name_fulladdressunique_count_forname_dob_1
ssnunique_count_forfulladdress_30	ssn_name_fulladdressunique_count_forname_dob_3
ssnunique_count_forfulladdress_60	ssn_name_fulladdressunique_count_forname_dob_7

ssnunique_count_forname_dob_1	ssn_name_fulladdressunique_count_forname_dob_14
ssnunique_count_forname_dob_3	ssn_name_fulladdressunique_count_forname_dob_30
ssnunique_count_forname_dob_7	ssn_name_fulladdressunique_count_forname_dob_60
ssnunique_count_forname_dob_14	ssn_name_fulladdressunique_count_forname_fulladdress_1
ssnunique_count_forname_dob_30	ssn_name_fulladdressunique_count_forname_fulladdress_3
ssnunique_count_forname_dob_60	ssn_name_fulladdressunique_count_forname_fulladdress_7
ssnunique_count_forname_fulladdress_1	ssn_name_fulladdressunique_count_forname_fulladdress_14
ssnunique_count_forname_fulladdress_3	ssn_name_fulladdressunique_count_forname_fulladdress_30
ssnunique_count_forname_fulladdress_7	ssn_name_fulladdressunique_count_forname_fulladdress_60
ssnunique_count_forname_fulladdress_14	ssn_name_fulladdressunique_count_forfulladdress_dob_1
ssnunique_count_forname_fulladdress_30	ssn_name_fulladdressunique_count_forfulladdress_dob_3
ssnunique_count_forname_fulladdress_60	ssn_name_fulladdressunique_count_forfulladdress_dob_7
ssnunique_count_forfulladdress_dob_1	ssn_name_fulladdressunique_count_forfulladdress_dob_14
ssnunique_count_forfulladdress_dob_3	ssn_name_fulladdressunique_count_forfulladdress_dob_30
ssnunique_count_forfulladdress_dob_7	ssn_name_fulladdressunique_count_forfulladdress_dob_60
ssnunique_count_forfulladdress_dob_14	ssn_name_fulladdressunique_count_fordob_homephone_1
ssnunique_count_forfulladdress_dob_30	ssn_name_fulladdressunique_count_fordob_homephone_3
ssnunique_count_forfulladdress_dob_60	ssn_name_fulladdressunique_count_fordob_homephone_7
ssnunique_count_fordob_homephone_1	ssn_name_fulladdressunique_count_fordob_homephone_14
ssnunique_count_fordob_homephone_3	ssn_name_fulladdressunique_count_fordob_homephone_30
ssnunique_count_fordob_homephone_7	ssn_name_fulladdressunique_count_fordob_homephone_60
ssnunique_count_fordob_homephone_14	ssn_name_fulladdressunique_count_forssn_lastname_1
ssnunique_count_fordob_homephone_30	ssn_name_fulladdressunique_count_forssn_lastname_3
ssnunique_count_fordob_homephone_60	ssn_name_fulladdressunique_count_forssn_lastname_7
ssnunique_count_forssn_lastname_1	ssn_name_fulladdressunique_count_forssn_lastname_14
ssnunique_count_forssn_lastname_3	ssn_name_fulladdressunique_count_forssn_lastname_30
ssnunique_count_forssn_lastname_7	ssn_name_fulladdressunique_count_forssn_lastname_60
ssnunique_count_forssn_lastname_14	ssn_name_fulladdressunique_count_forssn_zip5_1
ssnunique_count_forssn_lastname_30	ssn_name_fulladdressunique_count_forssn_zip5_3
ssnunique_count_forssn_lastname_60	ssn_name_fulladdressunique_count_forssn_zip5_7
ssnunique_count_forssn_zip5_1	ssn_name_fulladdressunique_count_forssn_zip5_14
ssnunique_count_forssn_zip5_3	ssn_name_fulladdressunique_count_forssn_zip5_30
ssnunique_count_forssn_zip5_7	ssn_name_fulladdressunique_count_forssn_zip5_60
ssnunique_count_forssn_zip5_14	ssn_name_fulladdressunique_count_forssn_name_1
ssnunique_count_forssn_zip5_30	ssn_name_fulladdressunique_count_forssn_name_3
ssnunique_count_forssn_zip5_60	ssn_name_fulladdressunique_count_forssn_name_7
ssnunique_count_forssn_name_1	ssn_name_fulladdressunique_count_forssn_name_14
ssnunique_count_forssn_name_3	ssn_name_fulladdressunique_count_forssn_name_30
ssnunique_count_forssn_name_7	ssn_name_fulladdressunique_count_forssn_name_60
ssnunique_count_forssn_name_14	ssn_name_fulladdressunique_count_forssn_fulladdress_1
ssnunique_count_forssn_name_30	ssn_name_fulladdressunique_count_forssn_fulladdress_3
ssnunique_count_forssn_name_60	ssn_name_fulladdressunique_count_forssn_fulladdress_7
ssnunique_count_forssn_fulladdress_1	ssn_name_fulladdressunique_count_forssn_fulladdress_14

ssnunique_count_forssn_fulladdress_3	ssn_name_fulladdressunique_count_forssn_fulladdress_30
ssnunique_count_forssn_fulladdress_7	ssn_name_fulladdressunique_count_forssn_fulladdress_60
ssnunique_count_forssn_fulladdress_14	ssn_name_fulladdressunique_count_forssn_name_dob_1
ssnunique_count_forssn_fulladdress_30	ssn_name_fulladdressunique_count_forssn_name_dob_3
ssnunique_count_forssn_fulladdress_60	ssn_name_fulladdressunique_count_forssn_name_dob_7
ssnunique_count_forssn_name_dob_1	ssn_name_fulladdressunique_count_forssn_name_dob_14