

# DATA QUALITY REPORT

## 1. Dataset Overview

**Dataset Name:** Applications Data

**Description:** This dataset contains the information of the people who filled their applications for a product. It contains fields like Social security number, name, address, phone no., and date of birth. It also contains a fraud label field which tells whether the application is good or bad.

**Time Period:** 1 January 2016 – 31 December 2016

**No. of Fields:** 10

**No. of Records:** 1,000,000

**Size of Dataset file:** 83 MB

## 2. Summary Table

For this dataset, all the fields are categorical which provides a summarised data analysis of all the fields.

Field Name	# records that have a value	% populated	# unique values	Most common value
RECORD	1000000	100.0%	1000000	NA
DATE	1000000	100.0%	365	2016-08-16
SSN	1000000	100.0%	835819	999999999
FIRST NAME	1000000	100.0%	78136	EAMSTRMT
LAST NAME	1000000	100.0%	177001	ERJSAXA
ADDRESS	1000000	100.0%	828774	123 MAIN ST
ZIP5	1000000	100.0%	26370	68138
DOB	1000000	100.0%	42673	1907-06-26
HOMEPHONE	1000000	100.0%	28244	9999999999
FRAUD LABEL	1000000	100.0%	2	'0'

### 3. DATA FIELD EXPLORATION

#### 3.1. FIELD 1: RECORD

**DESCRIPTION:** A categorical field containing unique integer for each record from 1 to 1,000,000

#### 3.2. FIELD 2: DATE

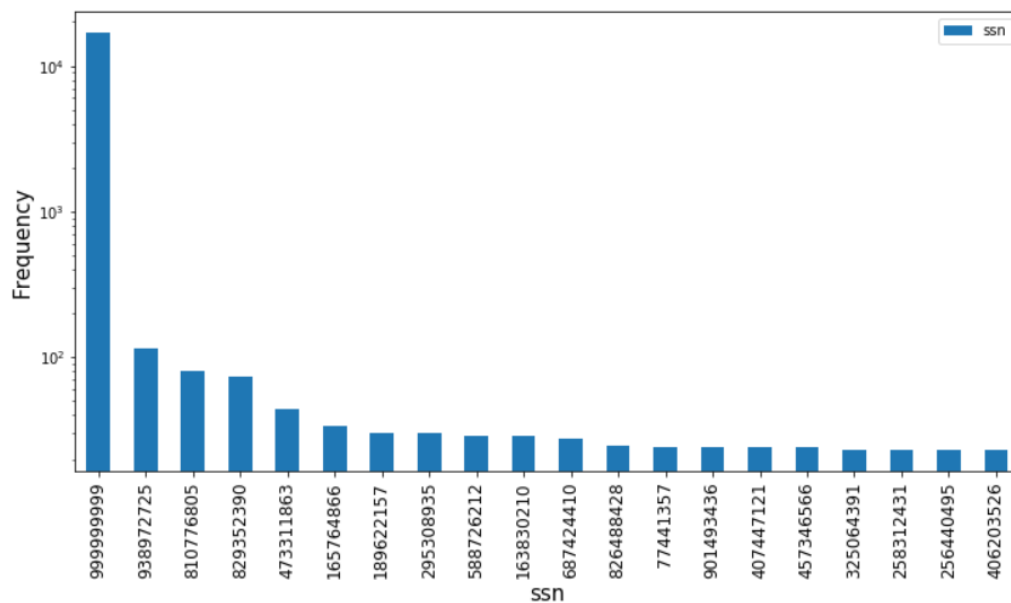
**DESCRIPTION:** A categorical field containing the date of application when it was filed.

Date	Count
20160816	2877
20160304	2861
20160718	2849
20160417	2848
20160101	2840
20161228	2832
20160903	2832
20160808	2832
20160827	2832
20160609	2831

Table containing first 10 values of Date

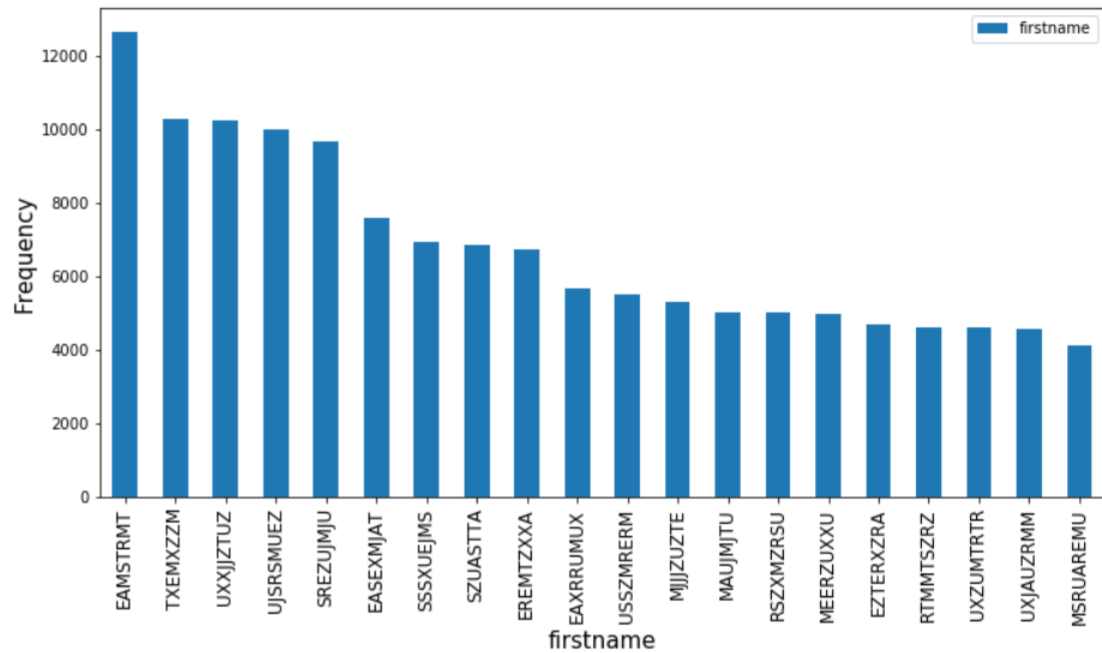
#### 3.3. FIELD 3: SSN

**DESCRIPTION:** A categorical field containing the Social Security Number of each applicant.



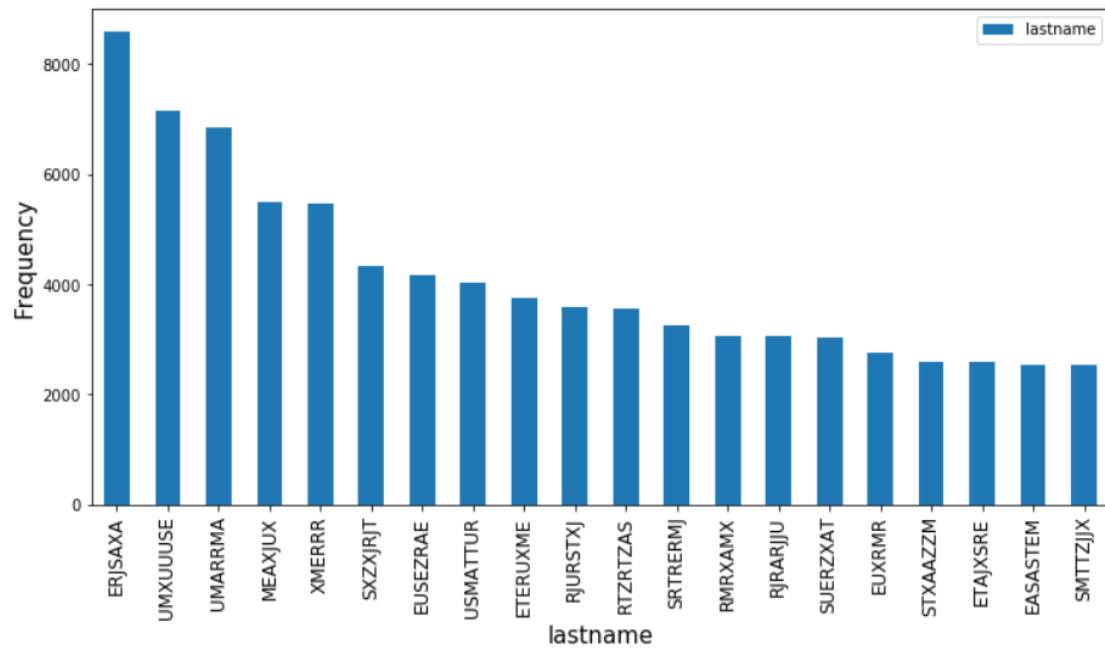
### 3.4. FIELD 4: FIRST NAME

**DESCRIPTION:** A categorical field containing the first name of each applicant.



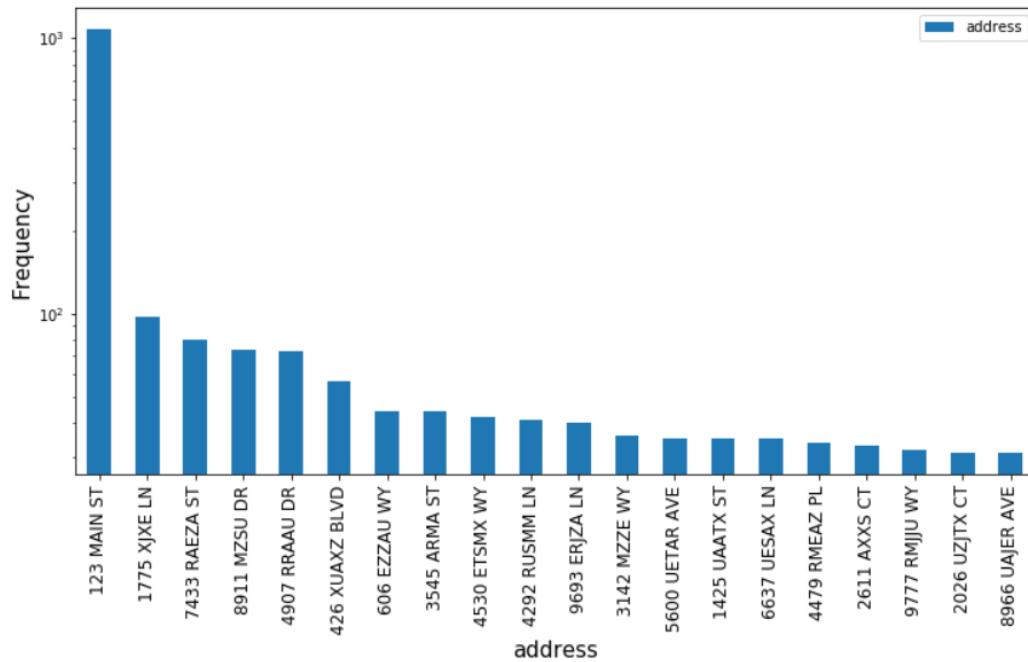
### 3.5. FIELD 5: LAST NAME

**DESCRIPTION:** A categorical field containing the last name of each applicant.



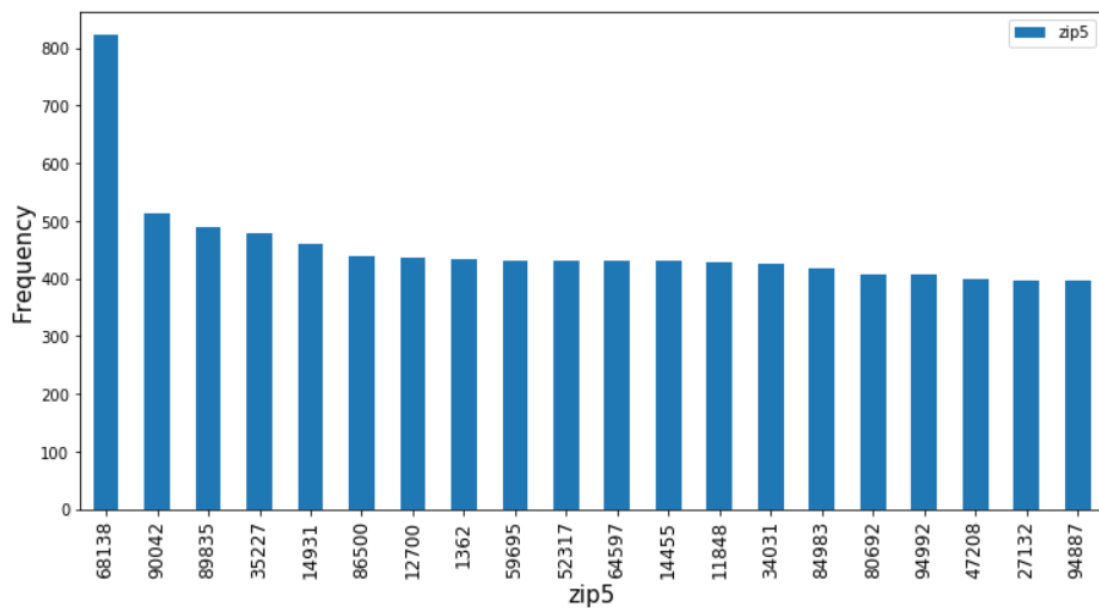
### 3.6. FIELD 6: ADDRESS

**DESCRIPTION:** A categorical field containing the address of each applicant.



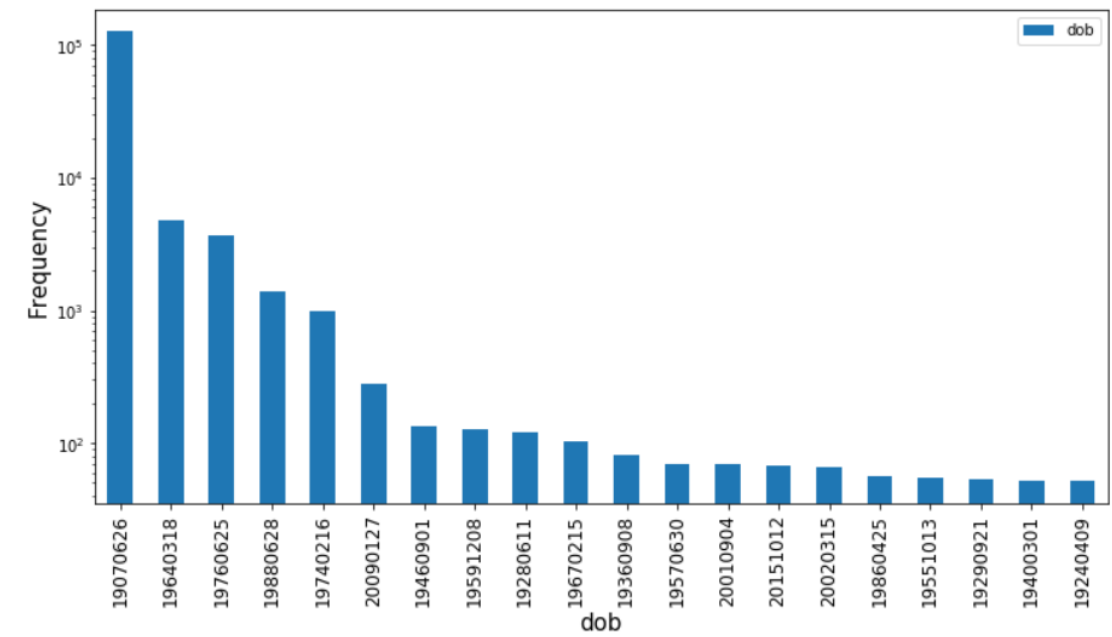
### 3.7. FIELD 7: ZIP5

**DESCRIPTION:** A categorical field containing all the five digits of the ZIP code.



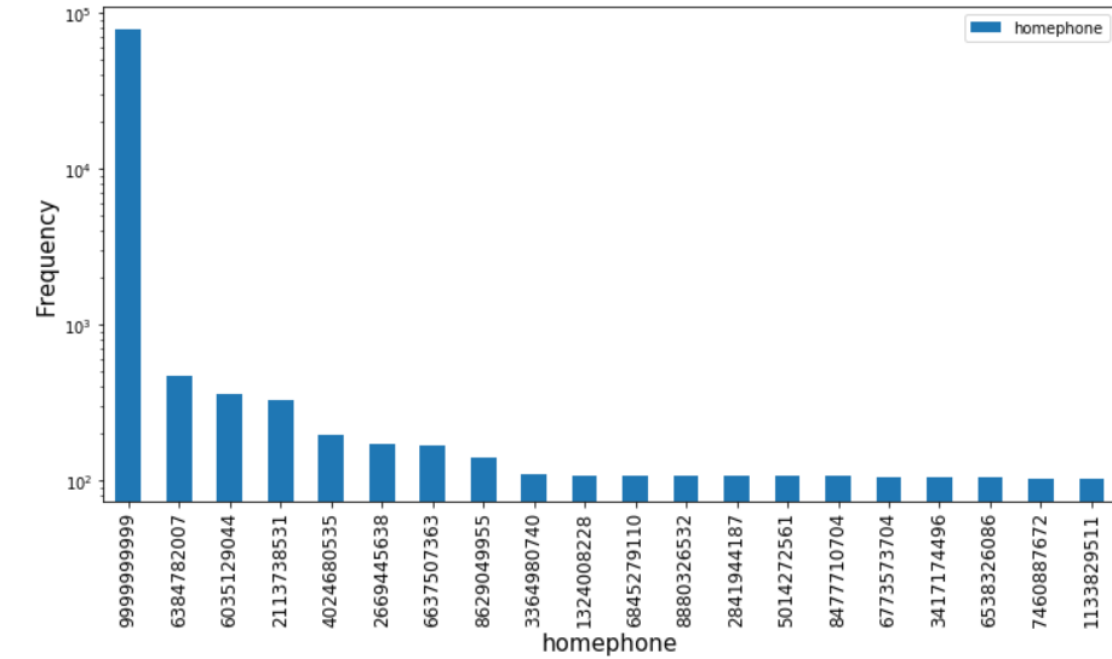
### 3.8. FIELD 8: DOB

**DESCRIPTION:** A categorical field containing the date of birth of each applicant.



### 3.9. FIELD 9: HOMEPHONE

**DESCRIPTION:** A categorical field containing the phone number of their homes.



### 3.10. FIELD 10: FRAUD LABEL

**DESCRIPTION:** A categorical field containing two categories:

0 – Good application

1 – Bad application

