Name: Joshua Chang

Data Quality Report

1. **Data Description**
   This dataset is about Applications Data. The data is about credit card applications spanning from January 1, 2017 to December 31, 2017. There are 10 fields with 1,000,000 records.d

2. **Summary Table:**

   Numerical Table:

Field Name	% Populated	Min	Max	Mean	Standard Deviation	% Zero
date	100%	2017-01-01	2017-12-31	N/A	N/A	0

   Categorical Table:

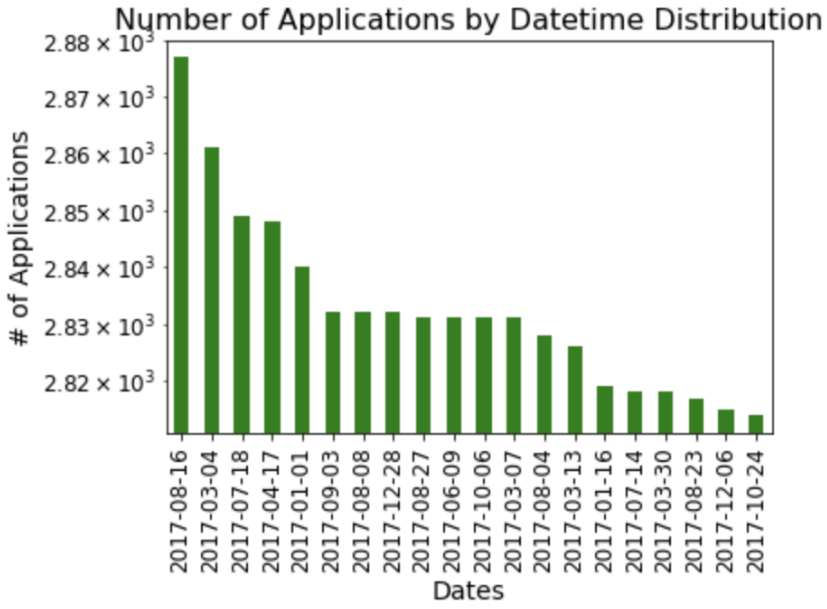
Field Name	% Populated	# of unique values	Most Common Field Value
records	100%	1,000,000	N/A
ssn	100%	835,819	999999999
firstname	100%	78,136	EAMSTRMT
lastname	100%	177,001	ERJSAXA
address	100%	828,774	123 MAIN ST
zip5	100%	26,370	68138
dob	100%	42,673	1907-06-26
homephone	100%	28,244	(999) 999-9999
fraud_label	100%	2	0

3. **Visualization of Each Field**
   a. Field Name: records

      Description: Record number field. These are ordinal unique positive integers for each record from 1 to 1,000,000.

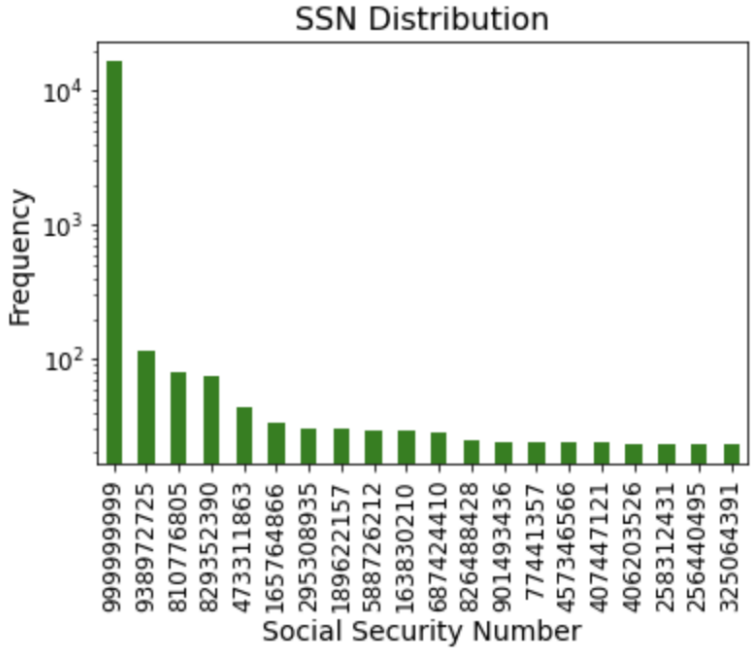
   b. Field Name: date

Description: Date(time) field. The distribution shows the top 20 most frequent dates in which applications were filed. The most common date to appear is 2017-08-16, the count of which is 2,877 applications.



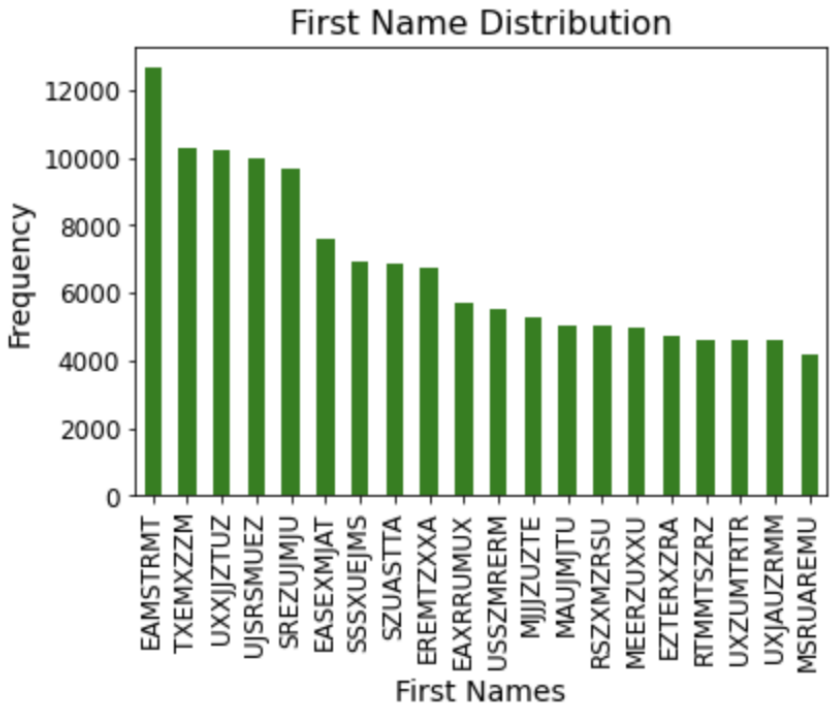
c. Field Name: ssn

Description: Social Security Number, or SSN, field. The distribution displays the top 20 social security numbers used for the applications filed. The most SSN used is 999999999, with the count being 16,935.



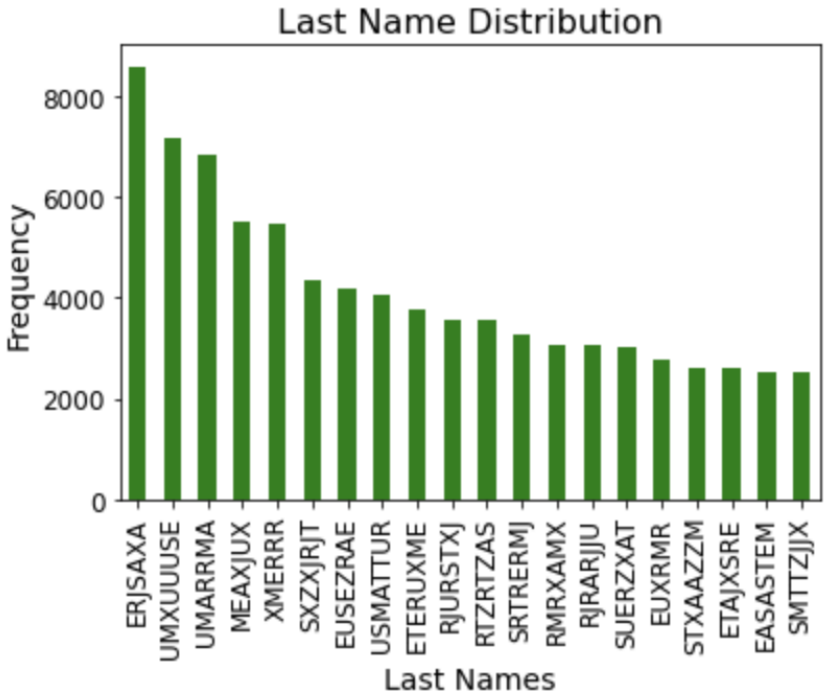
d. Field Name: firstname

Description: First name field. The distribution is the top 20 field values of first names. The most common first name is EAMSTRMT, the count of which is 12,658.



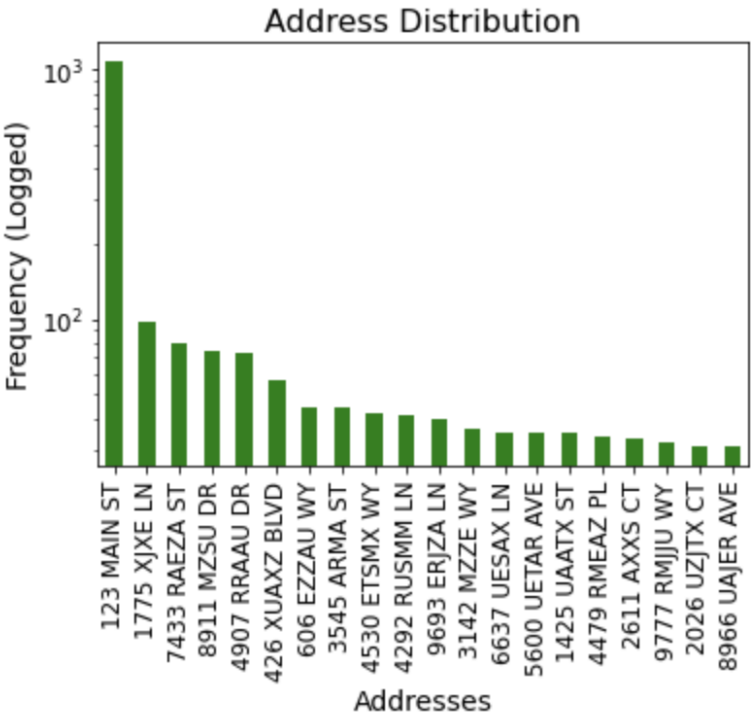
e. Field Name: lastname

Description: Last name field. The distribution is the top 20 field values of last names. The most common last name is ERJSAXA, where the count is 8,580.



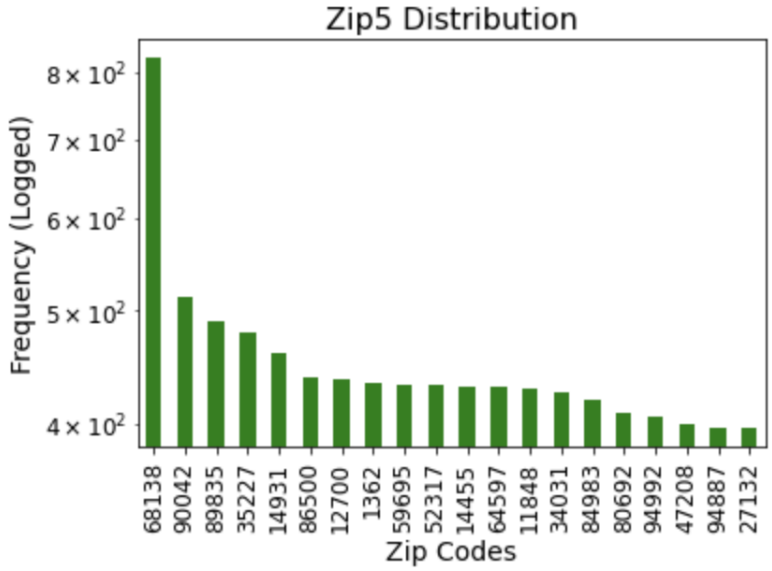
f. Field Name: address

Description: Address field. The distribution is the top 20 field values of addresses. The values of the frequency of appearances of each address was logged to better visualize the frequency distribution. The most common address is 123 MAIN ST, where the count is 1,079.



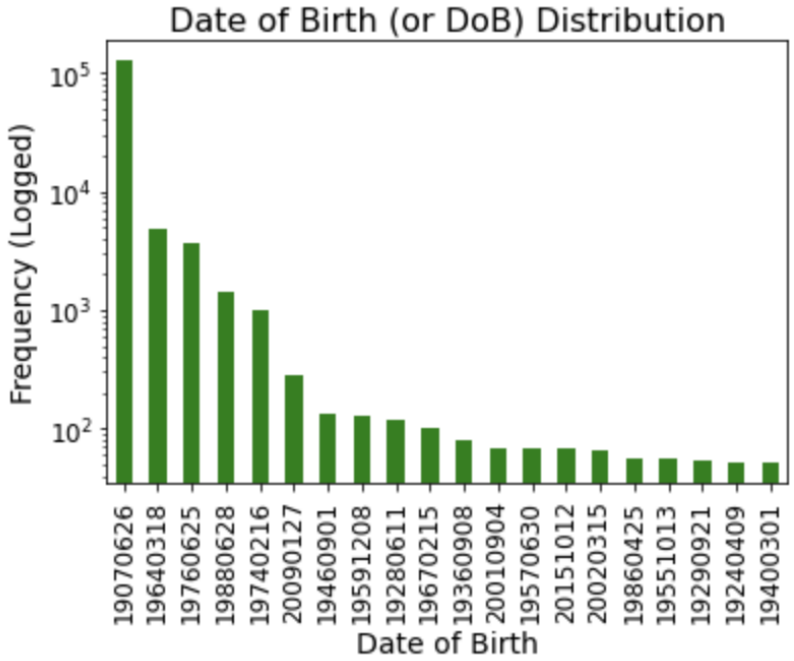
g. Field Name: zip5

Description: Zip5, or zip code, field. The distribution is the top 20 field values of zip codes. The values of the frequency of appearances of each zip code was logged to better visualize the frequency distribution. The most common zip code is 68138, with a count of 823.



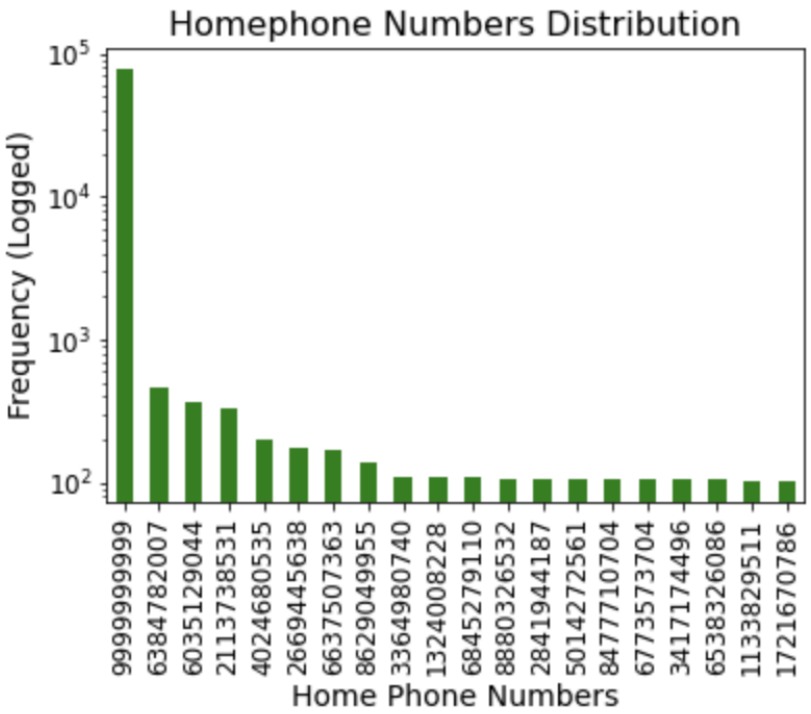
h. Field Name: dob

Description: Date of Birth field. The distribution is the top 20 field values of dates of birth. The values of the frequency of appearances of each date of birth was logged to better visualize the frequency distribution. The most common date of birth is 1907-06-26, where the count is 126,568.



i. Field Name: homephone

Description: Home Phone field. The distribution is the top 20 field values of home phone numbers. The values of the frequency of appearances of each date of birth was logged to better visualize the frequency distribution. The most common date of birth is 9999999999, where the count is 126,568.



j. Field Name: fraud_label

Description: Fraud label field. The distribution displays the number of applications that were (1) and were not (0) marked as fraud. The most common value is 0, or not fraud, where the count is 985,607.