

Date Quality Report

1. Data Description

The data is a synthetic application on account opening data made from an analysis on a few billion real U.S. applications over about 10 years. There are 10 fields and 1000000 records. The time the data covers is starting from 2017/01/01 to 2017/12/31. The data contains Personal Identifying Information includes name, SSN, date of birth, address, phone number, etc.

2. Summary Tables

- Numerical Table

Field name	% Populated	Min	Max	Mean	Stdev	% Zero
Date	100%	2017-01-01	2017-12-31	/	/	0.00

- Categorical Data

Field name	% Populated	# Unique Values	Most Common Value
Record	100%	1000000	N/A
Firstname	100%	78136	EAMSTRMT
Lastname	100%	177001	ERJSAXA
Address	100%	829852	1775 XJXE LN
ssn	100%	852745	938972725
Zip5	100%	26370	68138
dob	100%	169240	19640318
homephone	100%	106755	6384782007
Fraud_label	100%	2	0

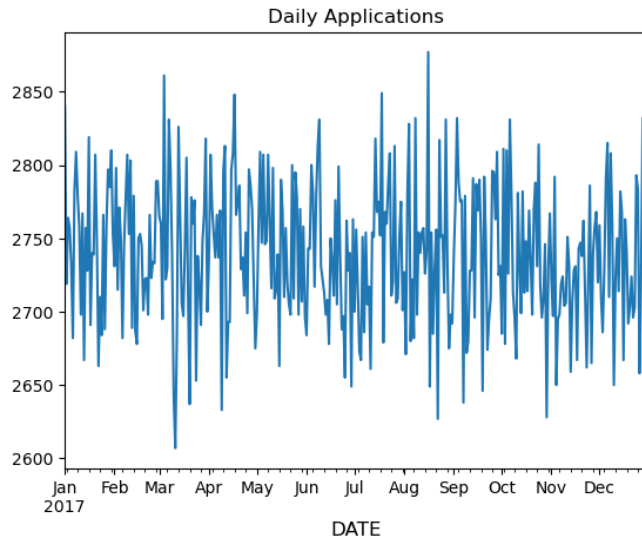
3. Visualization of Each Field

(1) Field name: Record

Description: Record number, ordinal unique positive integer for each record, from 1 to 1,000,000.

(2) Field name: Date

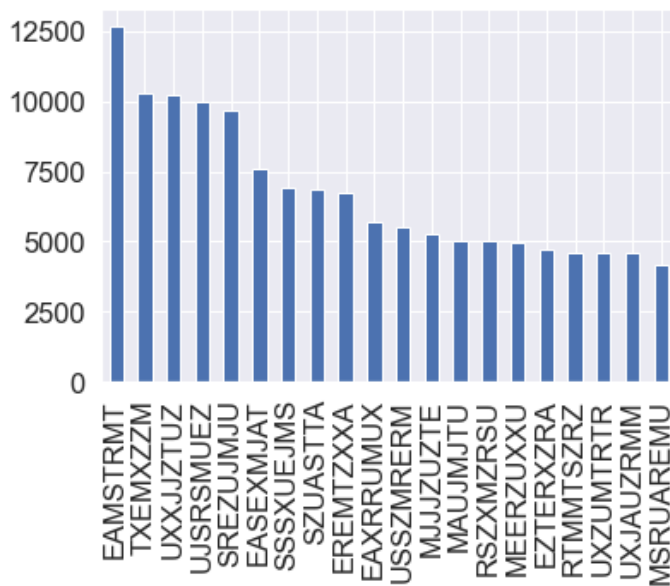
Description: Datetime field. The daily application distribution across time



(3) Field name: Firstname

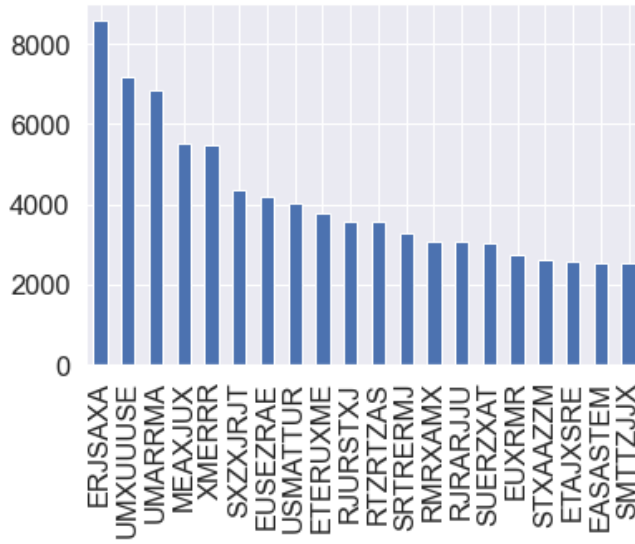
Description: The top 20 field values of the Applicant First Name.

Most common Applicant First Name is EAMSTRMT, and the count is 12658



(4) Field name: Lastname

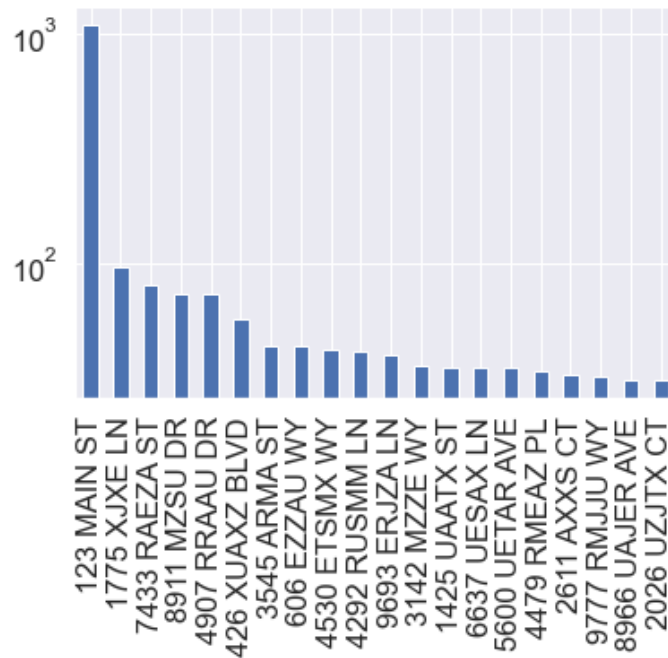
Description: The top 20 field values of the Applicant Last Name.



Most common Applicant Last Name is ERJSAXA, and the count is 8580

(5) Field name: Address

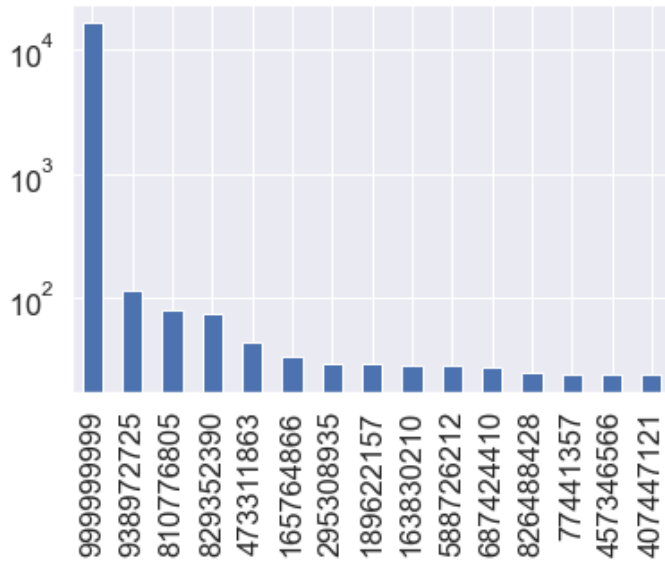
Description: The top 20 field values of the Applicant Address



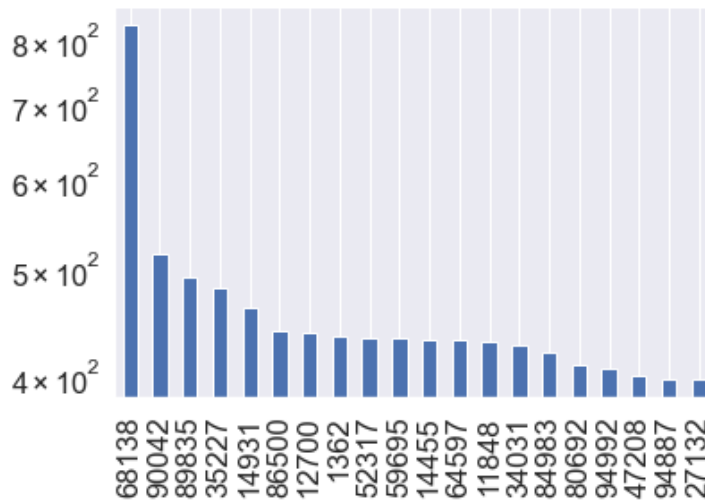
Most common Applicant Address is 123 MAIN ST, and the count is 1079

(6) Field name: ssn

Description: The top 15 field values of the Applicant SSN, and the count is 16935

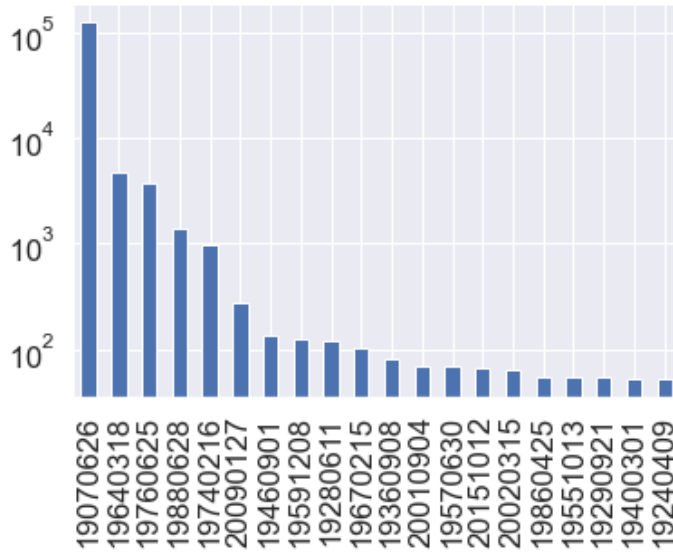


- (7) Field name: Zip5
Description: The top 20 field values of the Applicant Zipcode



Most common Applicant Zipcode is 68138, and the count is 823

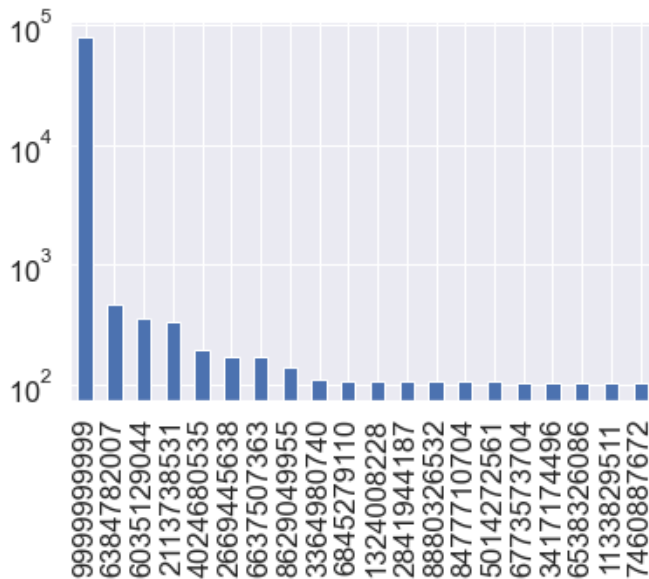
- (8) Field name: dob
Description: The top 20 field values of the Applicant Date of Birth



Most common Applicant Date of Birth is 1907-06-26, and the count is 126568

(9) Field name: homephone

Description: The top 20 field values of the Applicant Homephone



Most Common Homephone is 9999999999, and the count is 7851

(10) Field name: Fraud_label

Description: Fraud = 0 (non-fraud label), Fraud = 1 (fraud label)

The count of non-fraud is 985607, fraud is 14393

