

Data Quality Report

1. Data Description:

The data is about **Credit card application** fraud. The data covers the year 2017 data and has 1,000,000 records.

2. Summary Tables:

a) Numerical Table

Field Name	% Populated	Min	Max	Mean	St Dev	% Zero
Record	100.00	1	1,000,000	500,000	288,675.278933	0.00
Date	100.00	2017-01-01	2017-12-31	-	-	0.00
SSN	100.00	36	1,000,000,000	508,341,500	293,464,900	0.00
Zip5	100.00	2	99,999	50,073.333	28,915.9	0.00
Dob	100.00	1900-01-01	2016-10-31	-	-	0.00
Home phone	100.00	593,799	10,000,000,000	5,368,716,000	3,082,161,000	0.00

b) Categorical Table:

Field Name	% Populated	# Unique Values	Most Common Value
First Name	100.00	78136	EAMSTRMT
Last Name	100.00	177001	ERJSAXA
Address	100.00	828774	123 MAIN ST
Fraud Label	100.00	2	0

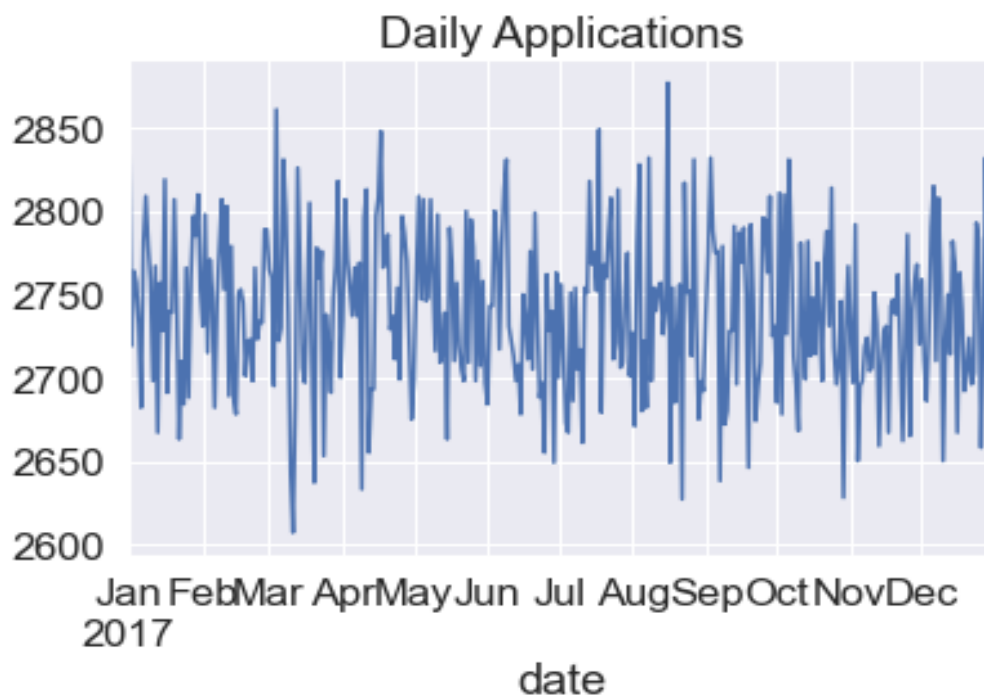
3. Visualization of Each Field

a) Field Name: Record

Description: Record Number field. Numbered integers for each record ranging from 1 to 1000000.

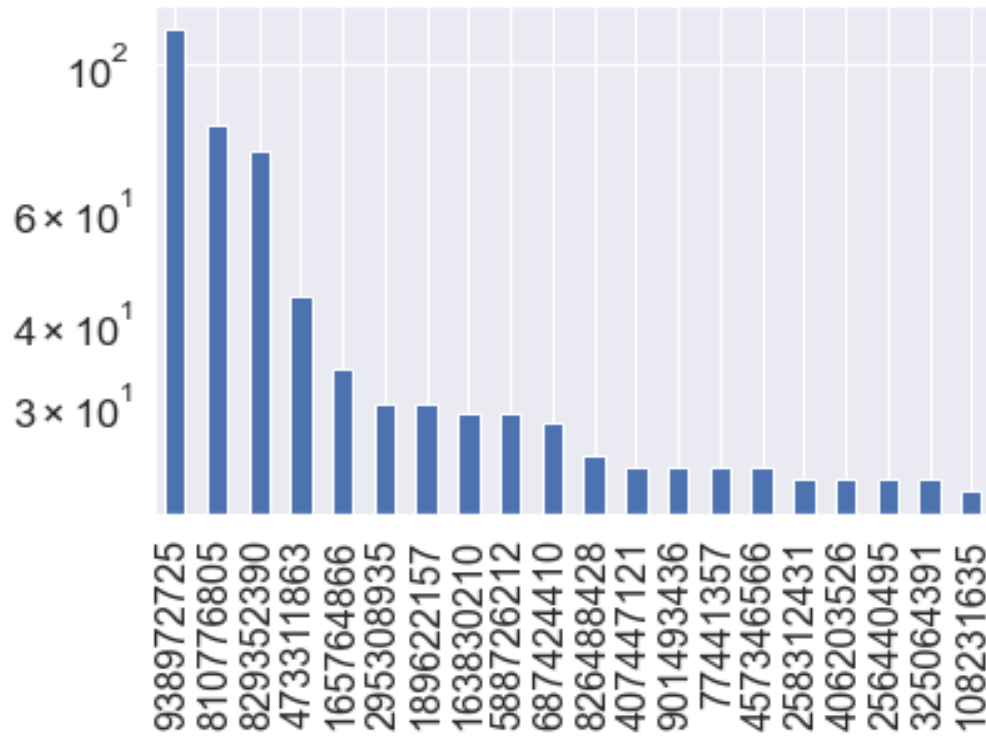
b) Field Name: Date

Description: Date Field. Daily transactions over the period of time ranging from 01-01-2017 to 12-31-2017.



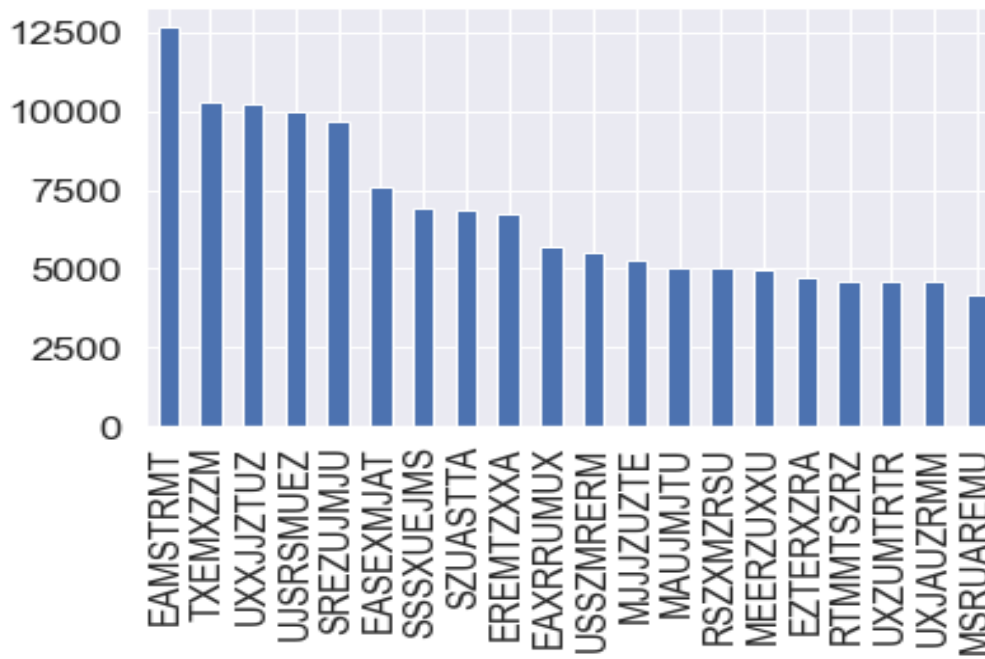
c) Field Name: SSN

Description: SSN field. Numbered integers for each record's SSN.



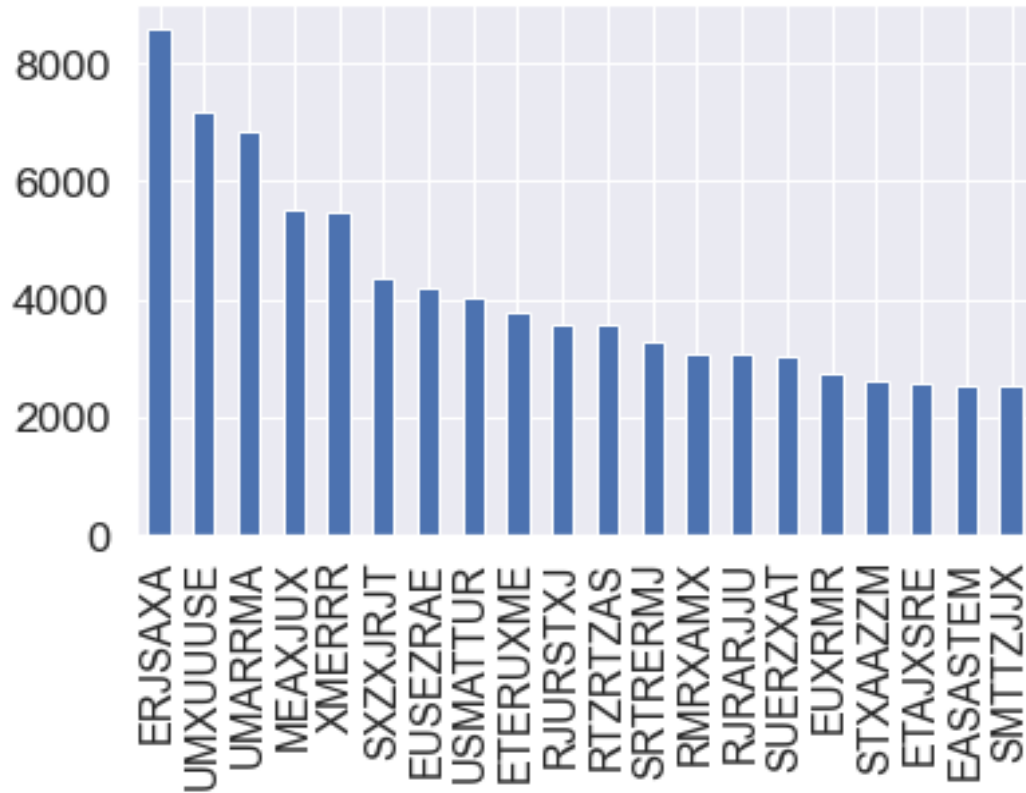
d) Field Name: First Name

Description: First Name field. First Names of all the records.



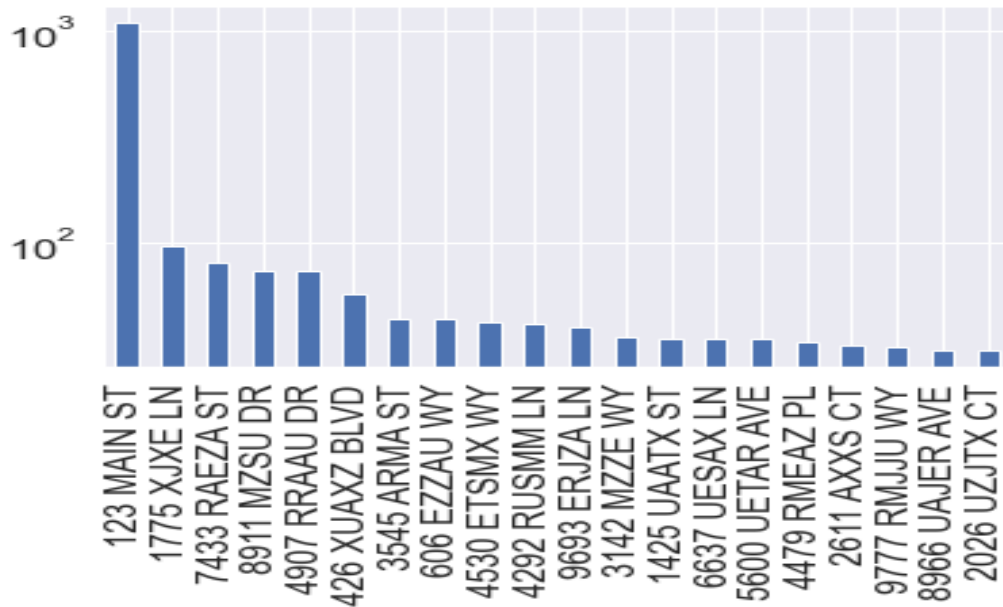
e) Field Name: Last Name

Description: Last Name field. Last Names of all the records.



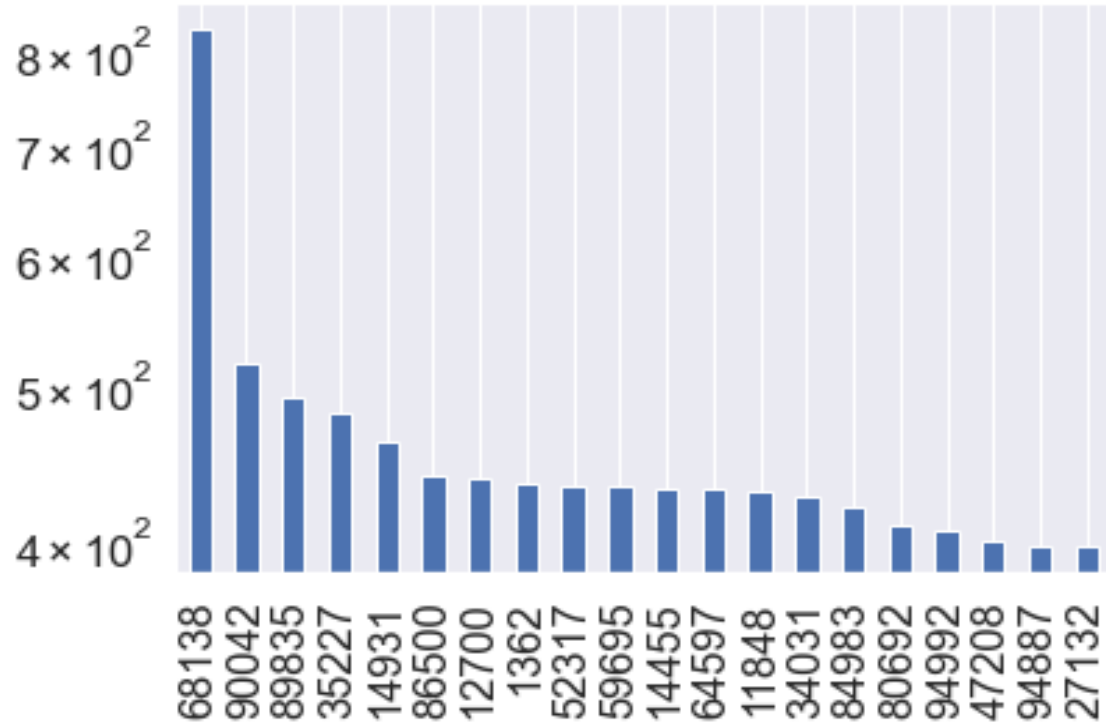
f) Field Name: Address

Description: Address field. Address of each record or transaction.



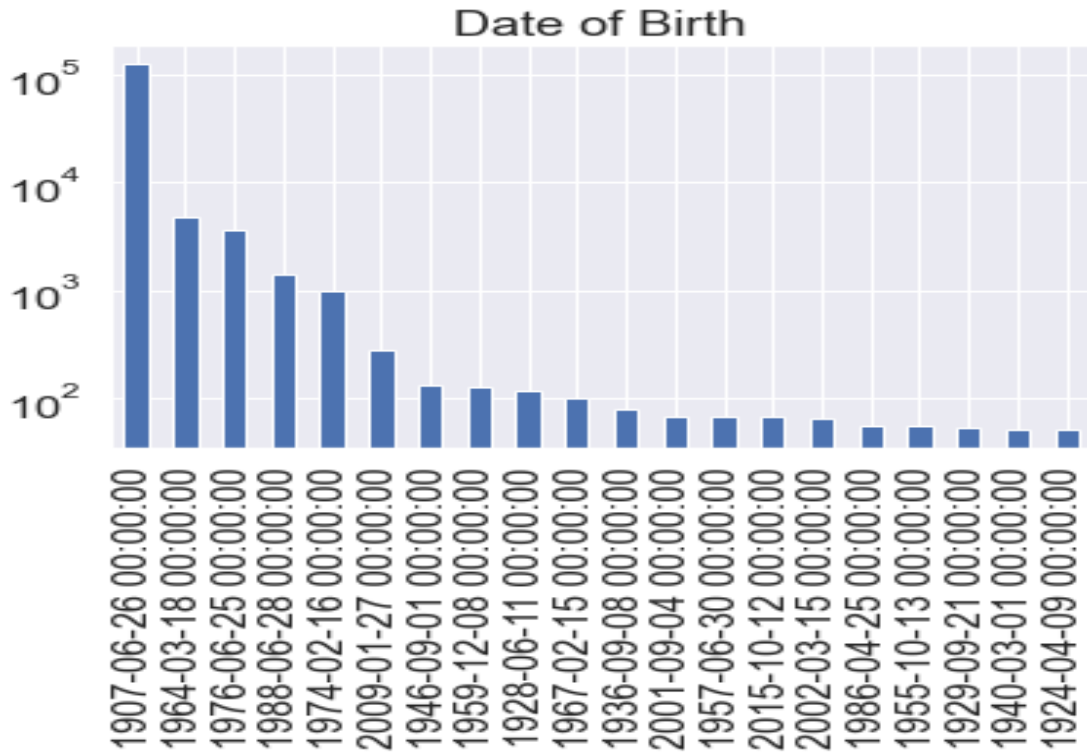
g) Field Name: zip5

Description: Zip code field. Zip code for each record or transaction.



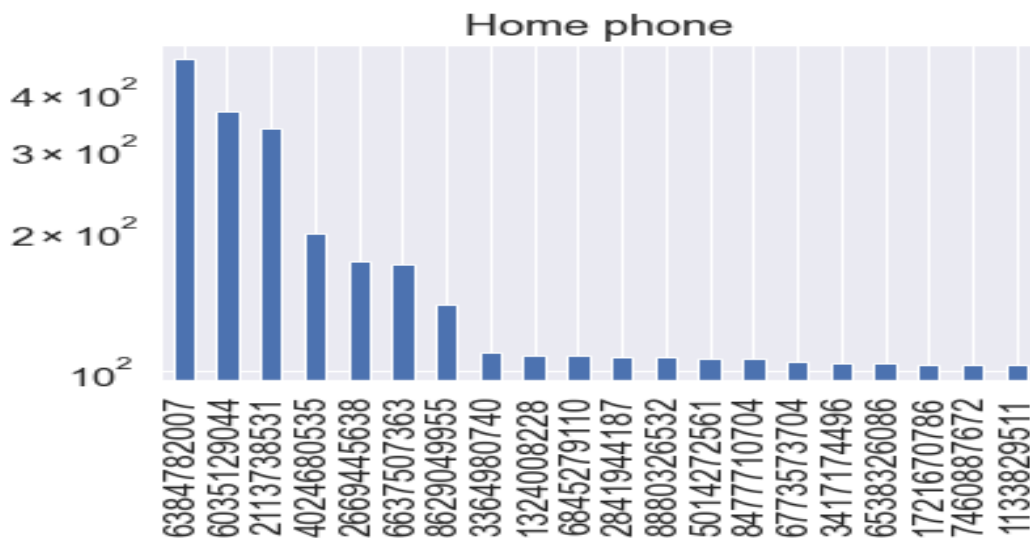
h) Field Name: dob

Description: Date of Birth field. Date of birth field of each customer's transaction.



i) Field Name: home phone

Description: Home phone field. Phone numbers of each record/transaction



j) Field Name: Fraud_label

Description: Fraud Label field. To check whether that transaction is a fraud or not.

