**DATA QUALITY REPORT**

**BY**

**SUMEET G DUDDAGI**

1. **HIGH LEVEL DESCRIPTION**

The given data is a dataset of Product application data. It consists of 1,000,000 records and 10 columns.

Dataset Name: Product Application Data.

Dataset Description: Application for credit card.

Columns: RECORD, DATA, SSN, FIRSTNAME, LASTNAME, ADDRESS, ZIP5(ZIPCODE), DOB (DATE OF BIRTH), HOMEPHONE, FRAUD\_LABEL.

1. **TWO FIELD SUMMARY TABLES**

**NUMERICAL FIELDS**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Field Name** | **% Populated** | **Min** | **Max** | **Mean** | **Stdev** | **% zero** |
| **DATE** | 100 | 2016-01-01 | 2016-12-31 | - | - | 0 |
| **DOB** | 100 | 1900-01-01 | 2016-10-31 | - | - | 0 |

**CATEGORICAL FIELDS**

|  |  |  |  |
| --- | --- | --- | --- |
| **Field Name** | **% Populated** | **# Unique Values** | **Most Common Value** |
| **RECORD** | 100 | 1,000,000 | - |
| **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 |
| **HOMEPHONE** | 100 | 28,244 | 9999999999 |
| **FRAUD\_LABEL** | 100 | 2 | 0 |

**3. FIELD DESCRIPTIONS**

RECORD

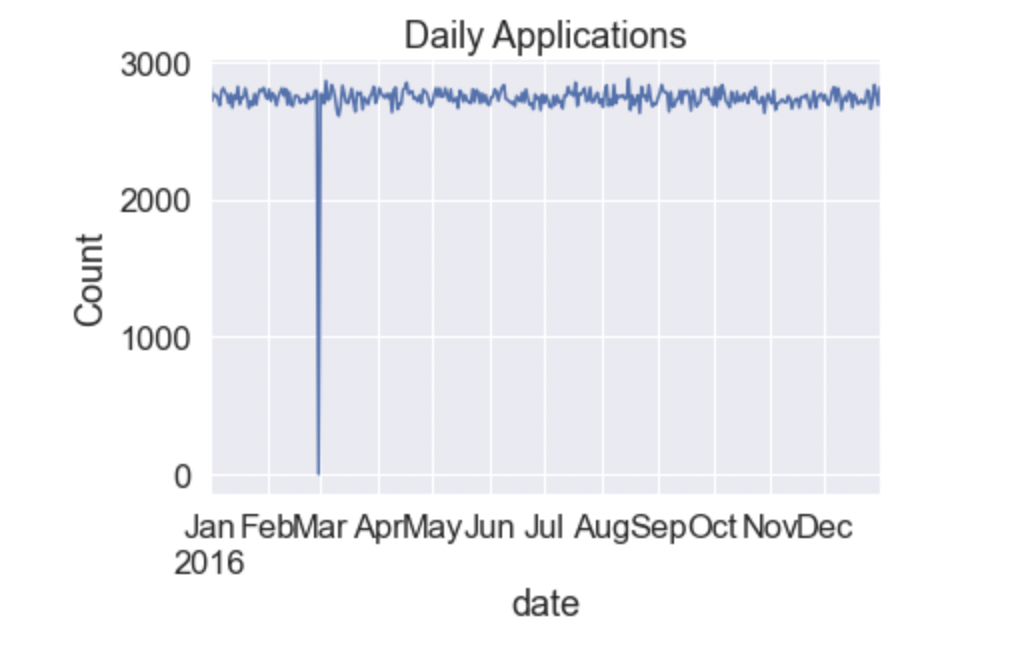
Column Description: This gives a serial number or a unique number for each application. This field value is unique to each row.

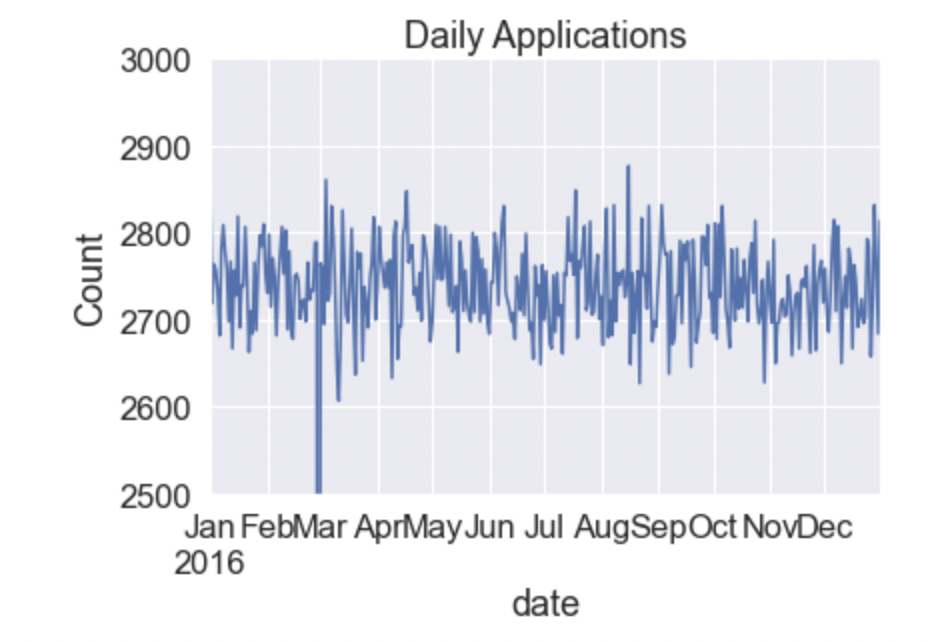
Type: Categorical

DATE

Column Description: This gives the date for each application. This field value has no missing value.

Type: Numerical

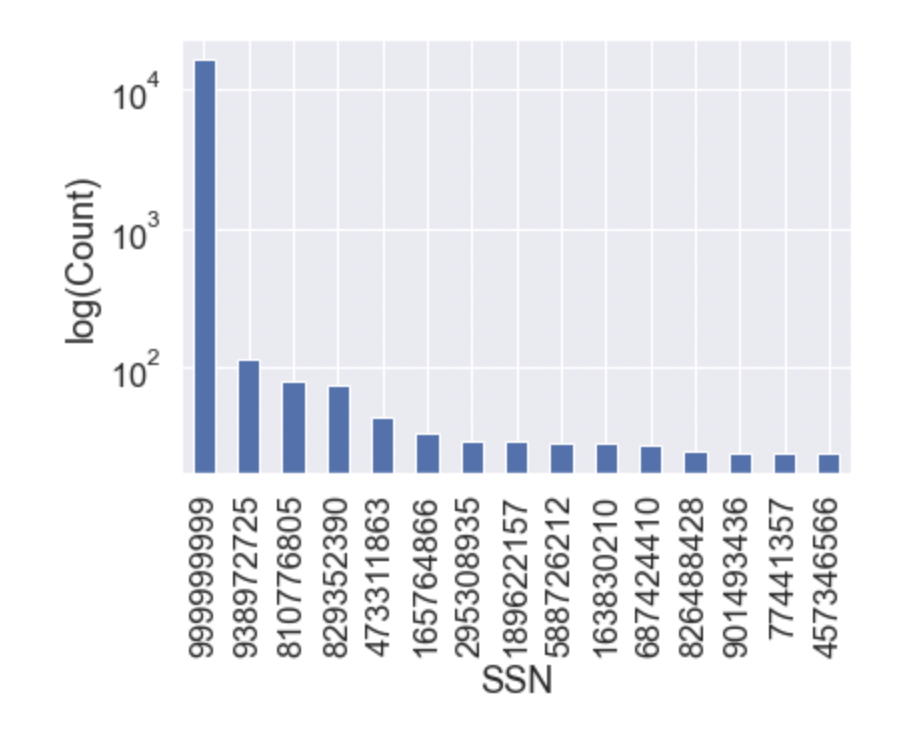




SSN

Column Description: It provides the Social Security Number of the users.

Type: Categorical



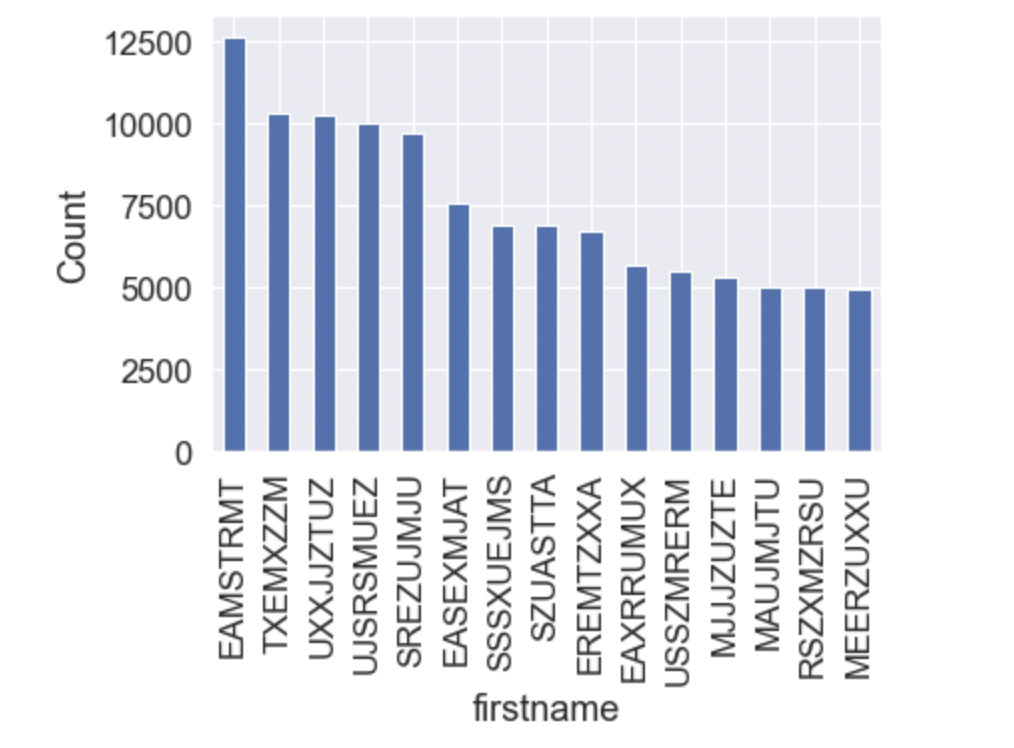
FIRSTNAME

Column Description: Gives the first name of the users.

Type: Categorical

No. of Unique Entries: 78,136

Most Common Value: EAMSTRMT



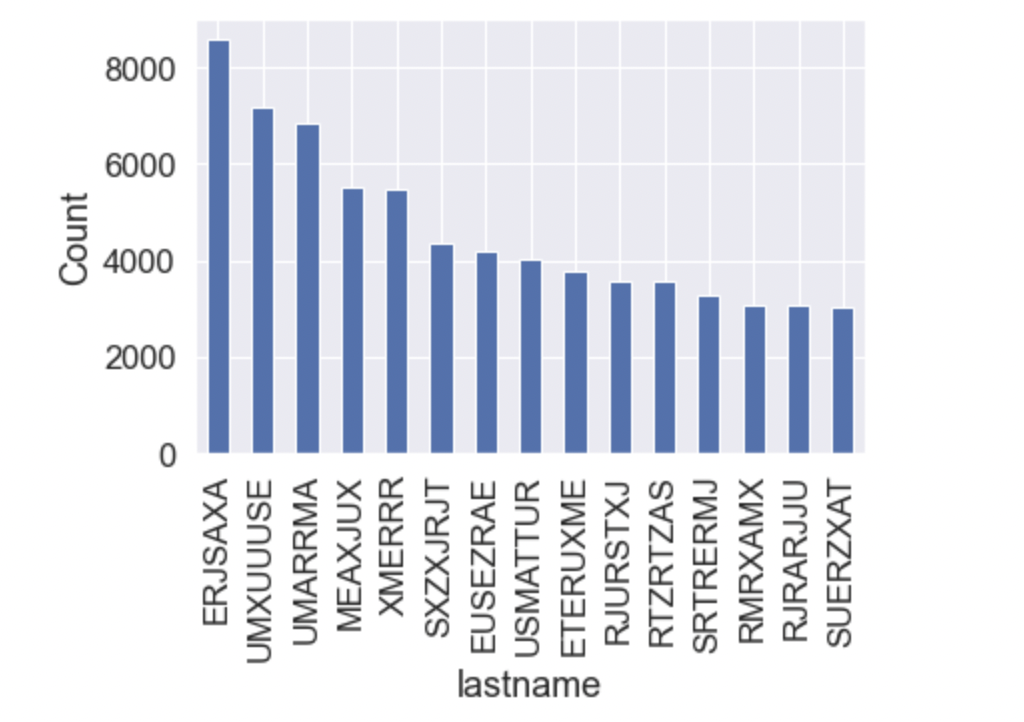
LASTNAME

Column Description: It shows the last name of the users who filles out the application.

Type: Categorical

No. of Unique Entries: 177,001

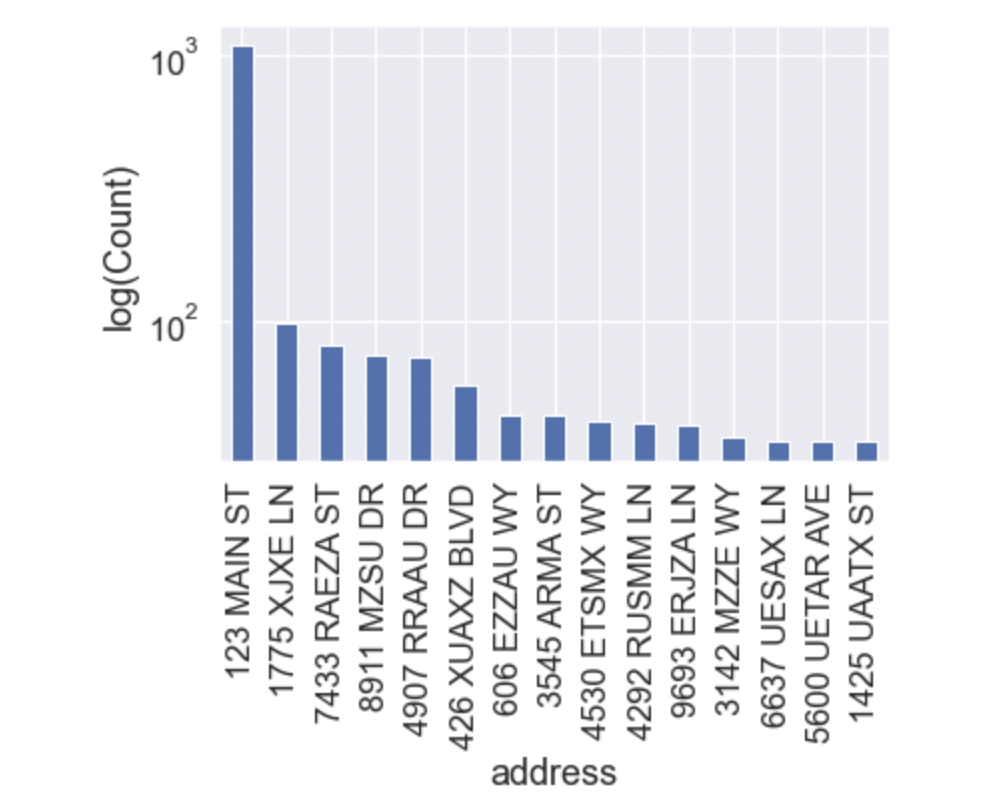
Most Common Value: ERJSAXA



ADDRESS

Column Description: This column gives the address.

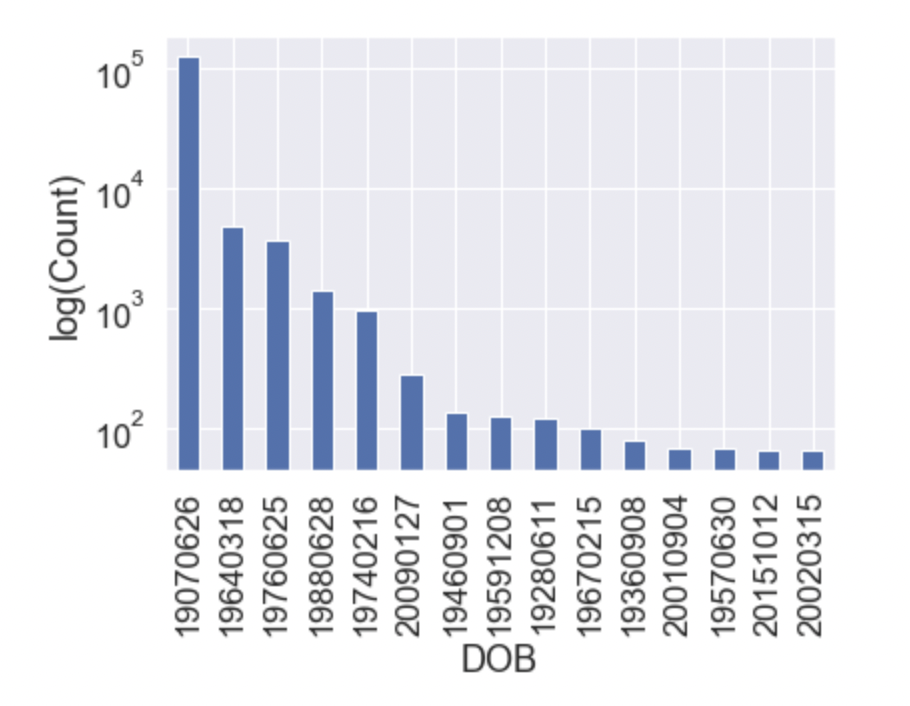
Type: Categorical.



DOB

Column Description: It gives the Date of Birth.

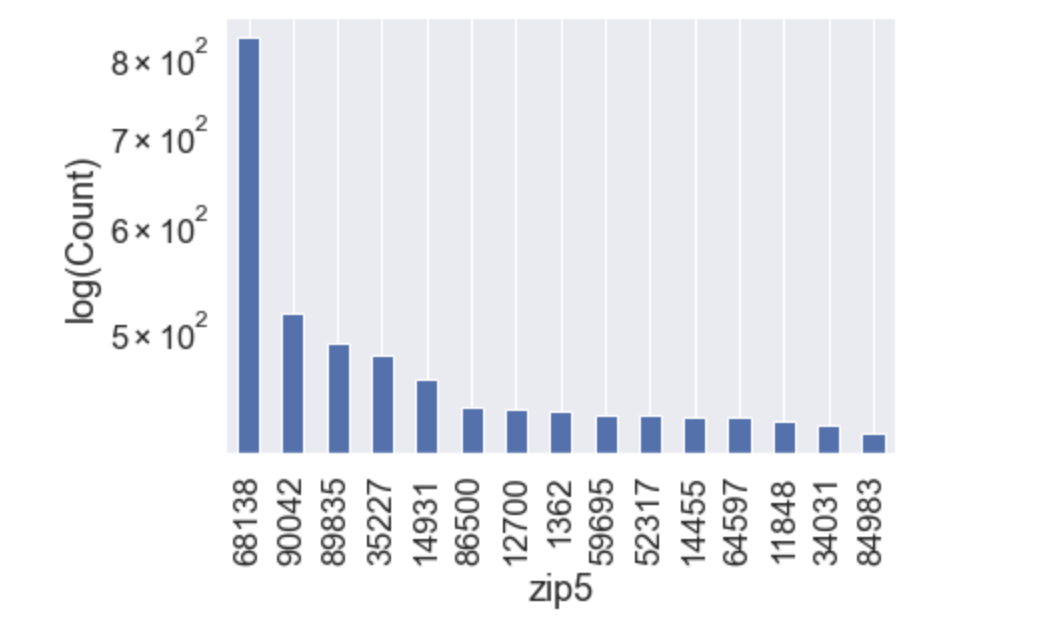
Type: Numerical



ZIP5

Column Description: It gives the zip-code of the users who filled out the application.

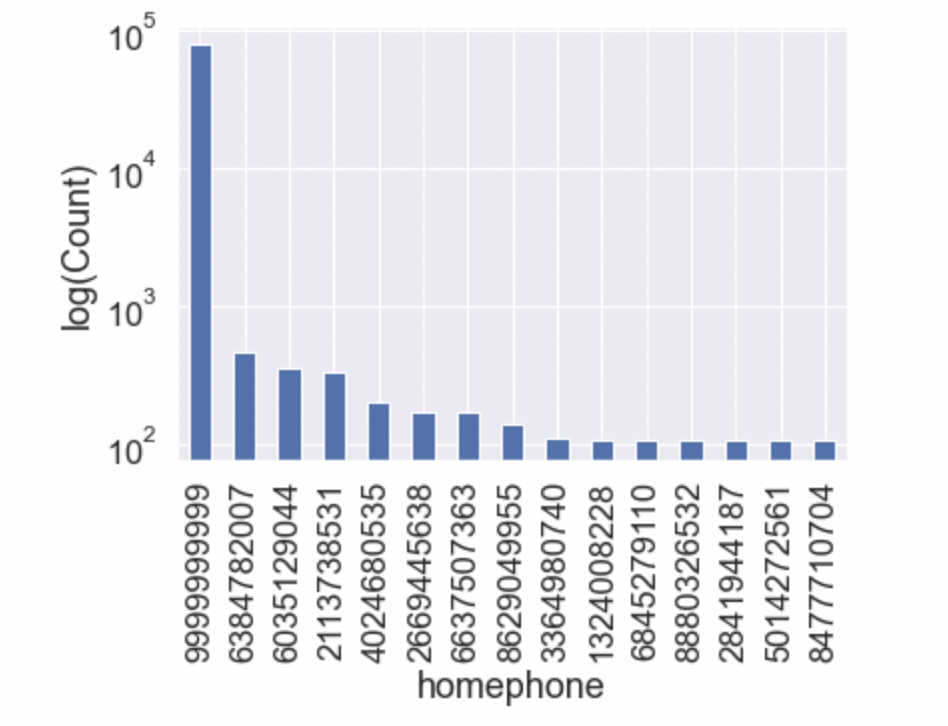
Type: Categorical



HOMEPHONE

Column Description: It gives the home contact number of the users who filled out the application.

Type: Categorical



FRAUDLABEL

Column Description: It gives the fraud label for the records in the dataset.

0: non-fraud labelled

1: Fraud-labelled

Type: Categorical

