

DATA QUALITY REPORT FOR APPLICATIONS DATA

Description

Dataset Name: Applications Data

Dataset Purpose: assessments on product application data

Data Source: N/A

Time Period: 2016-01-01 to 2016-12-31

Number of Fields: #10

Number of Records: 1,000,000

Summary Table

Categorical Fields:

Feature	# of records	% populated	Unique Values	Type	Most Common Field Value
record	1000000	100	1000000	object	1
date	1000000	100	365	datetime64[ns]	2016-08-16 00:00:00
ssn	1000000	100	835819	object	999999999
firstname	1000000	100	78136	object	EAMSTRMT
lastname	1000000	100	177001	object	ERJSAXA
address	1000000	100	828774	object	123 MAIN ST
zip5	1000000	100	26370	object	68138
dob	1000000	100	42673	datetime64[ns]	1907-06-26 00:00:00
homephone	1000000	100	28244	object	999999999
fraud_label	1000000	100	2	category	0

Field 1:

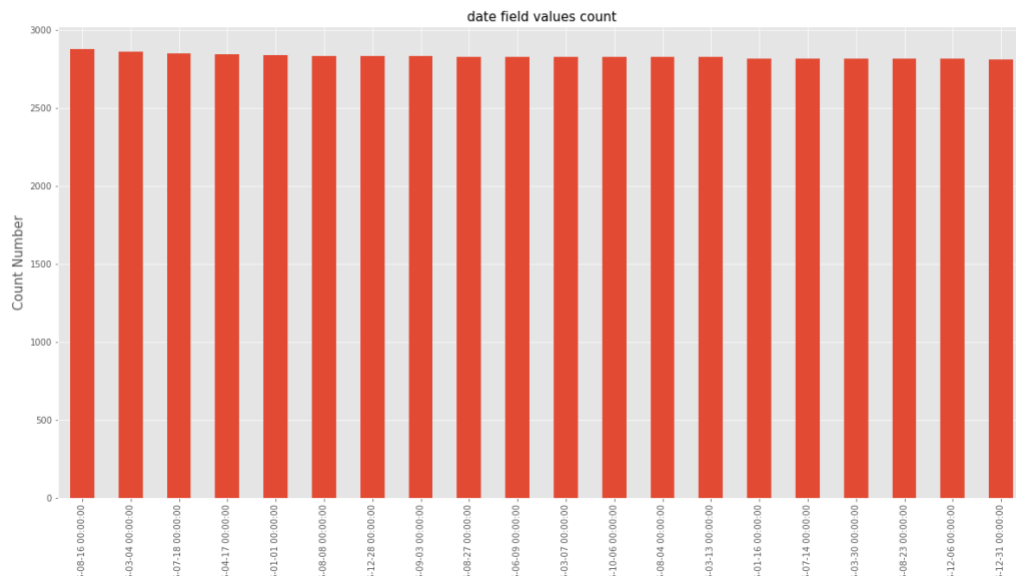
Name: record

Description: Unique identifier of each entry in the data.

Field 2:

Name: date

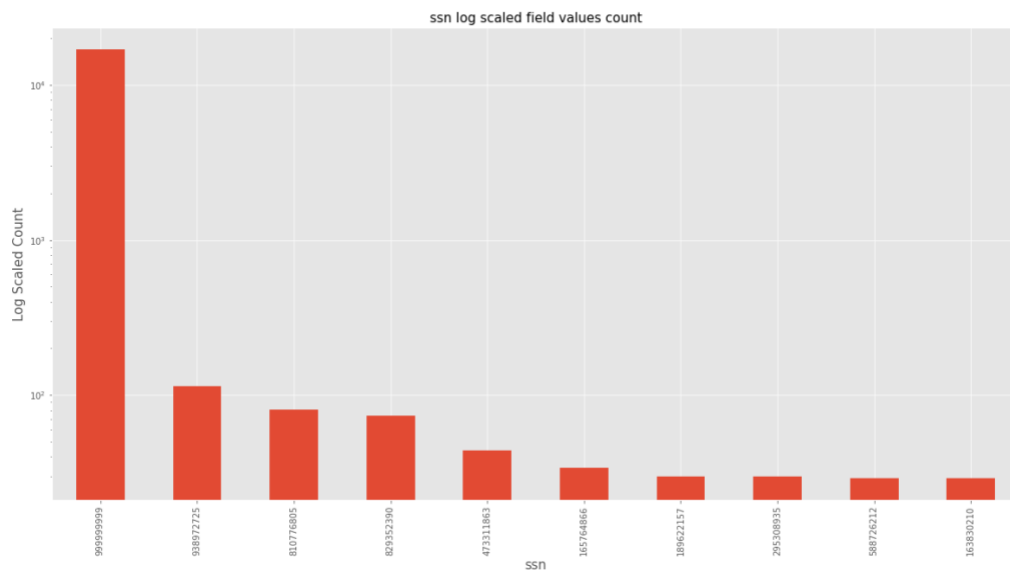
Description: Date of application.

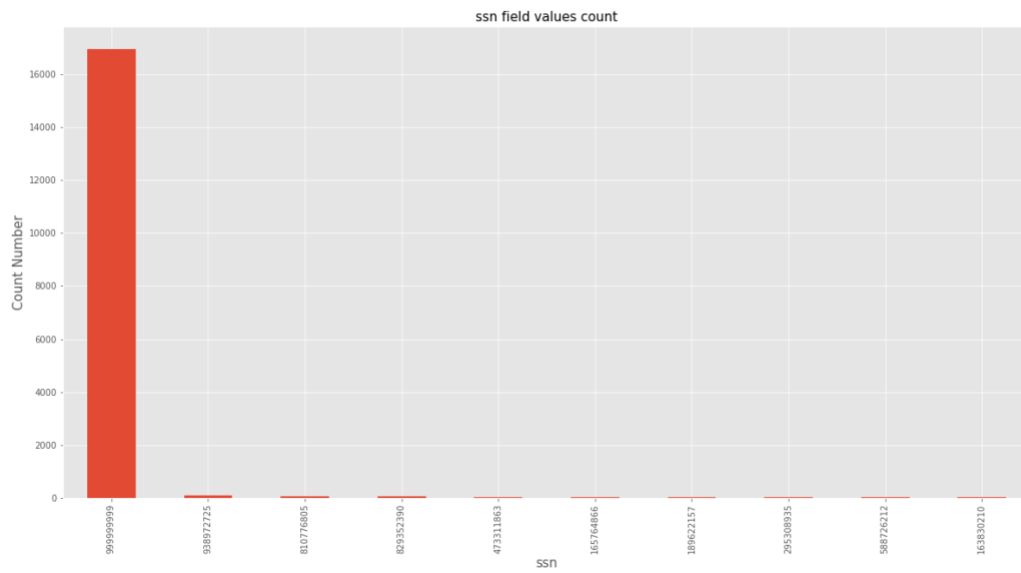


Field 3:

Name: ssn

Description: Social Security Number of applicants.

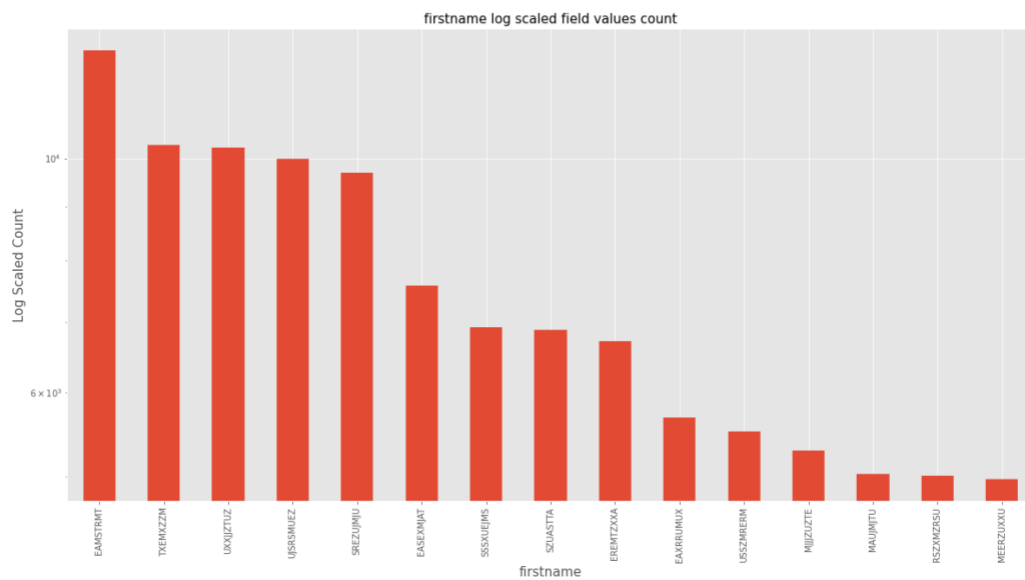


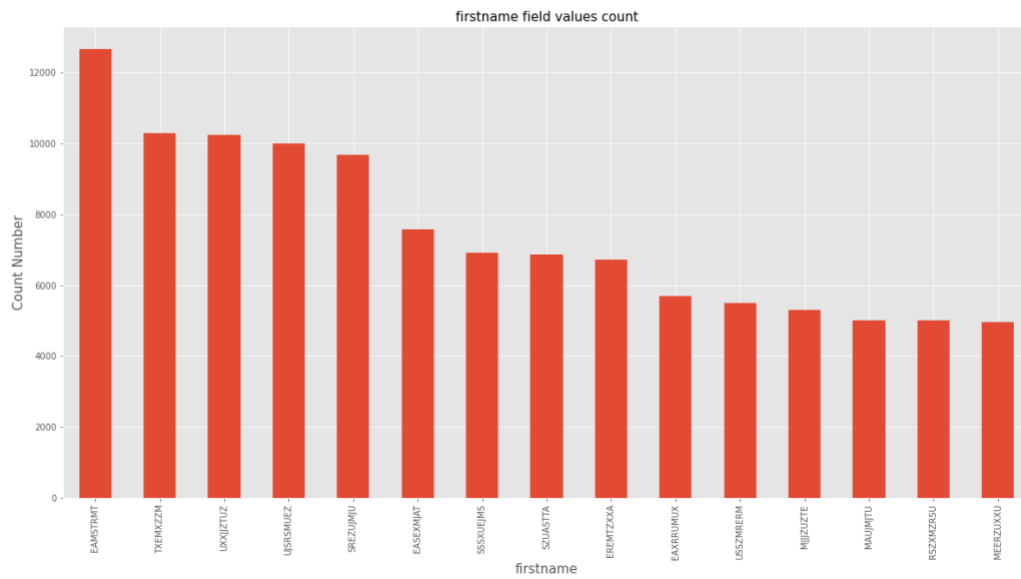


Field 4:

Name: firstname

Description: First name of applicants.

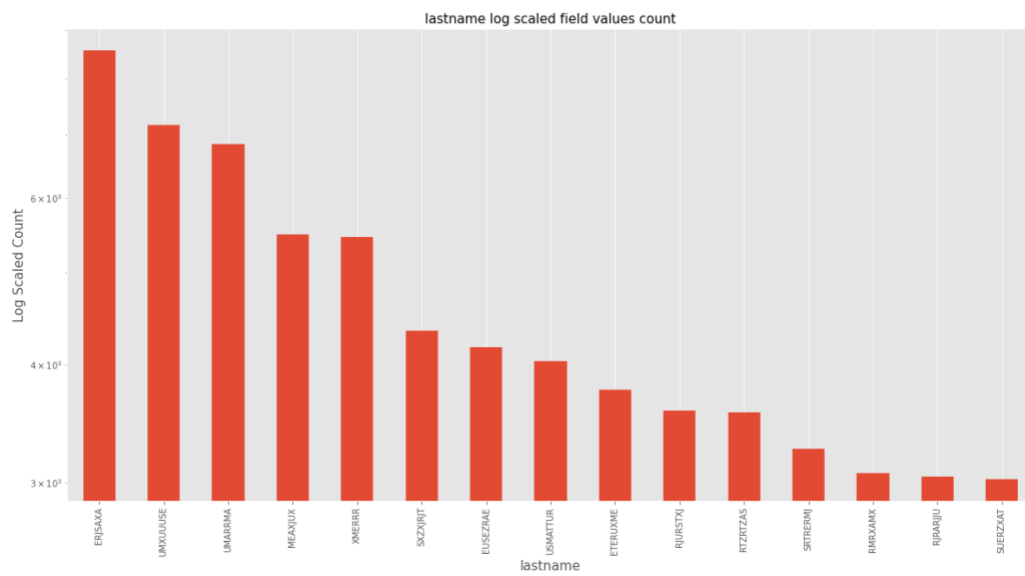


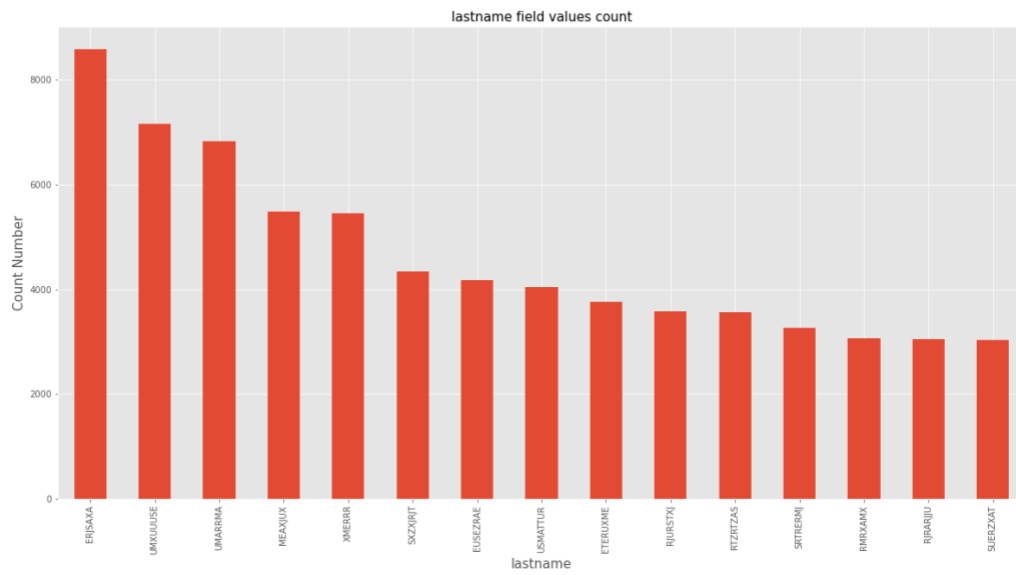


Field 5:

Name: lastname

Description: Last name of applicants.

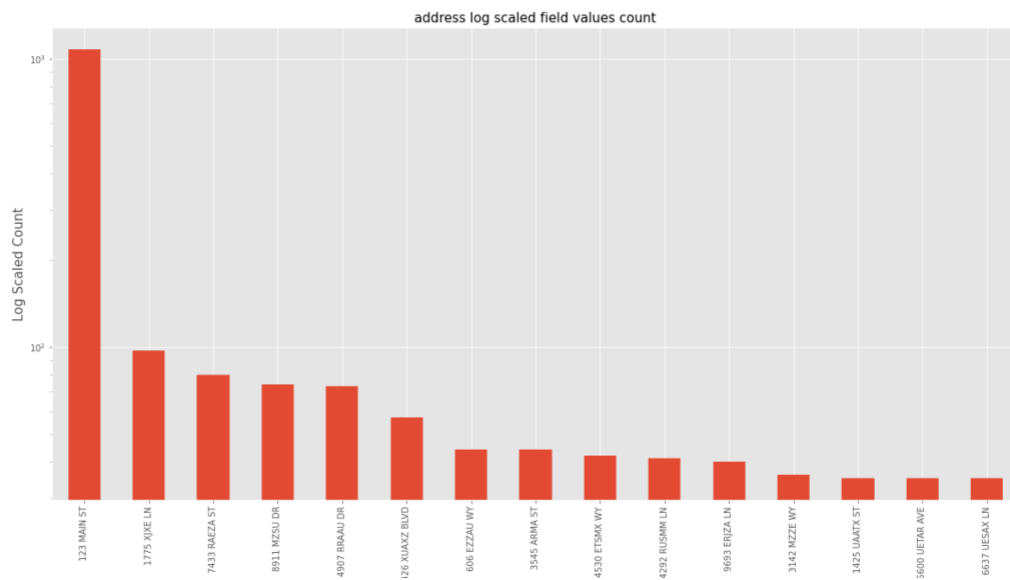


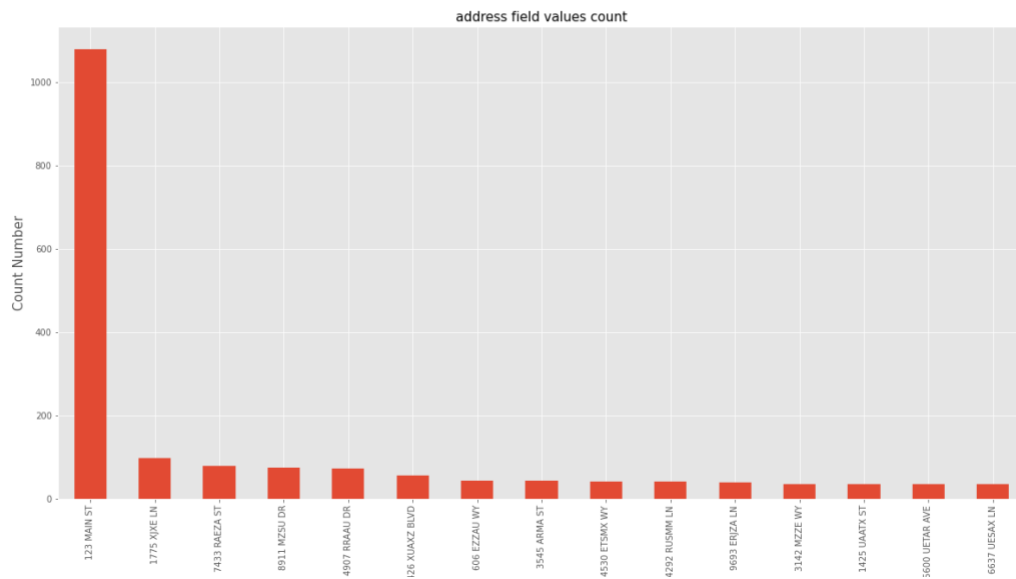


Field 6:

Name: address

Description: Addresses of applicants.

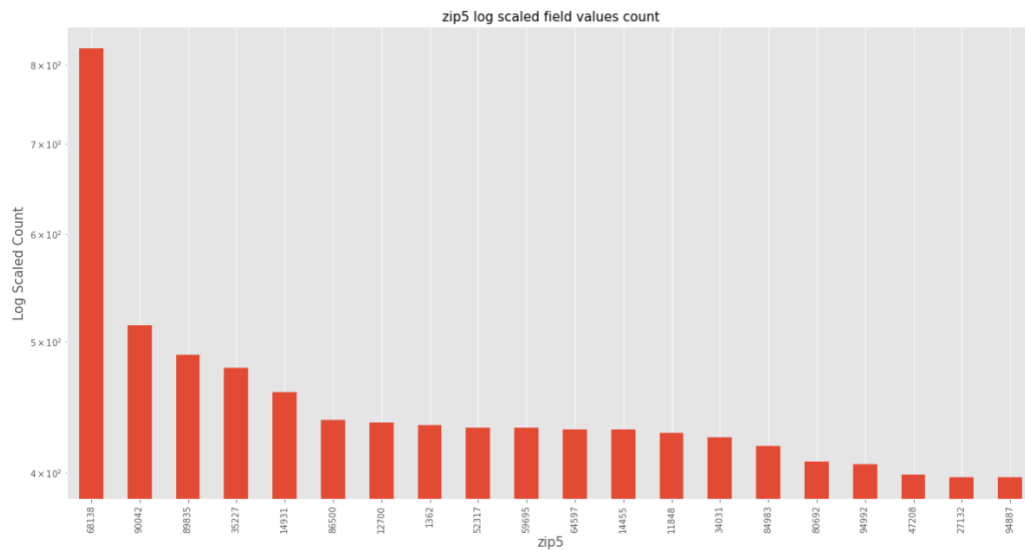


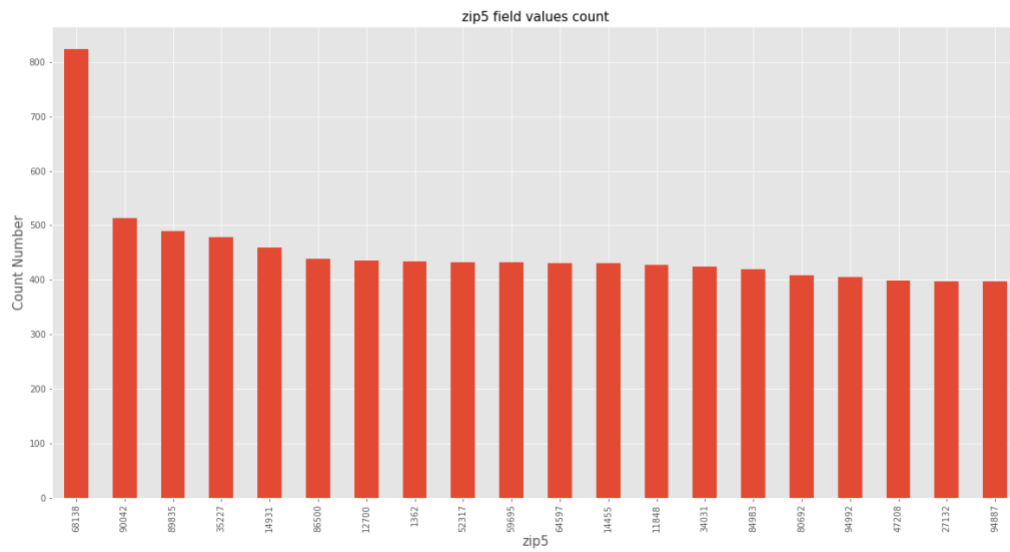


Field 7:

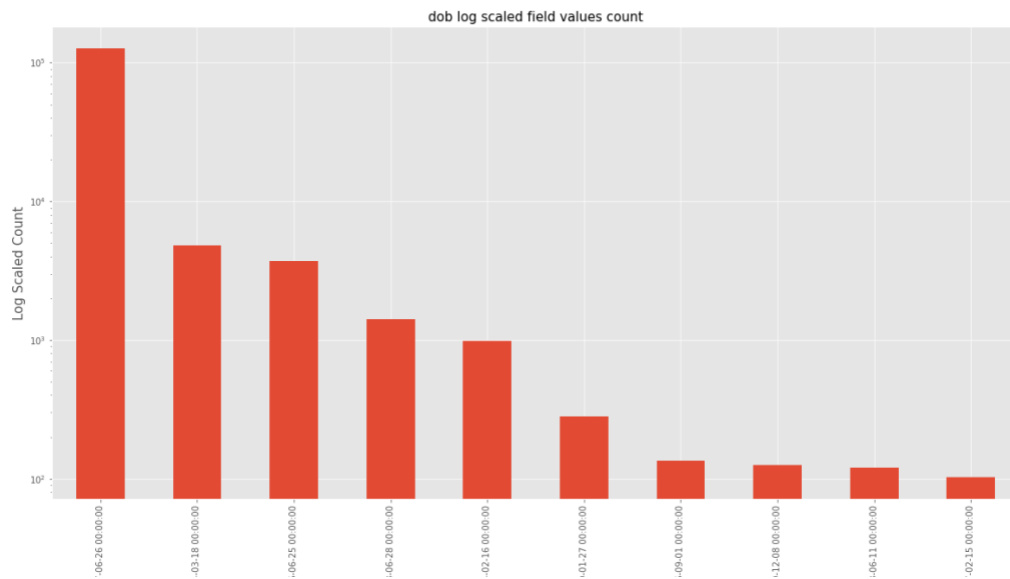
Name: zip5

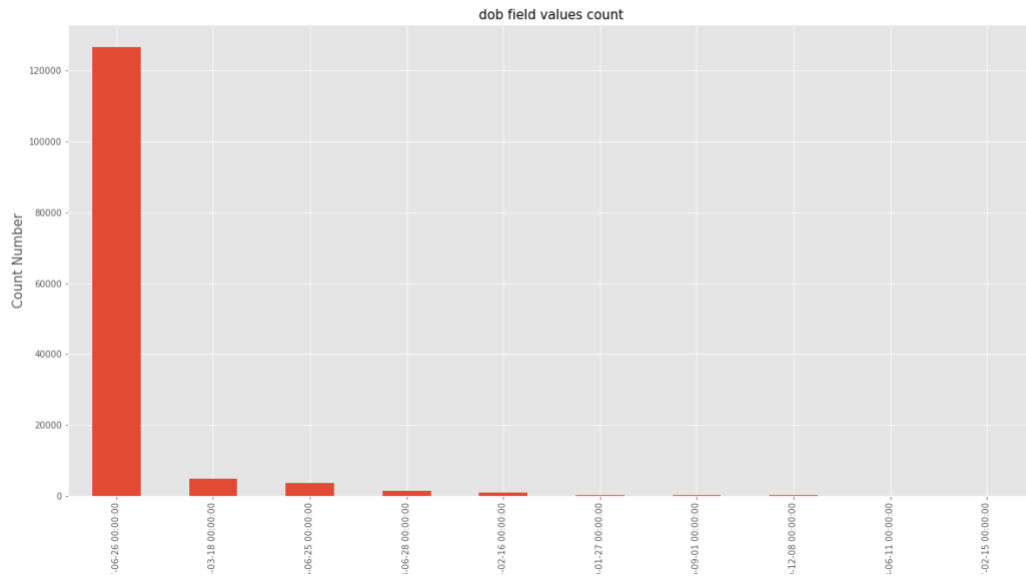
Description: 5-digit zip code of applicants.





Field 8:
Name: dob
Description: Date of Birth of applicants.

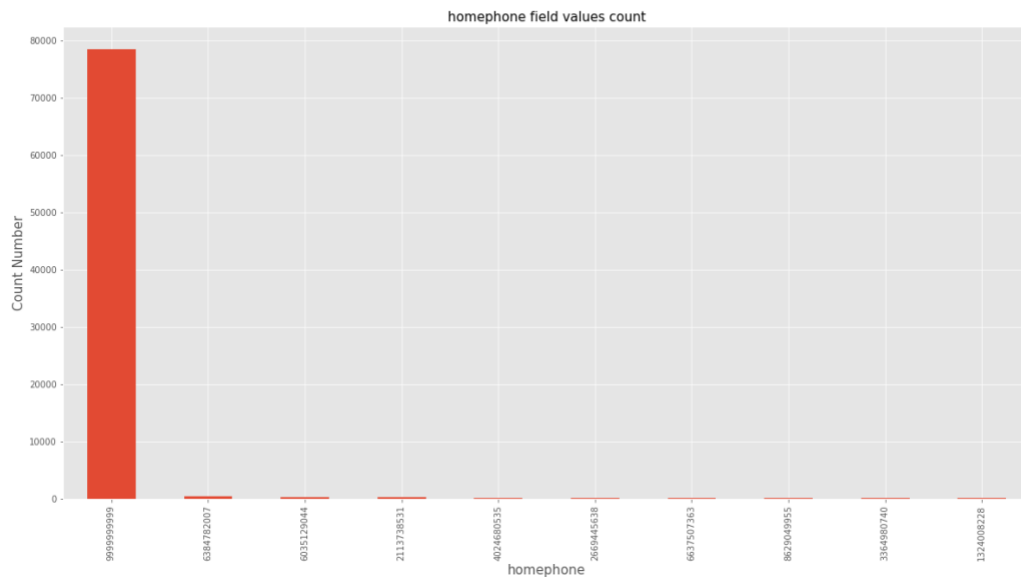


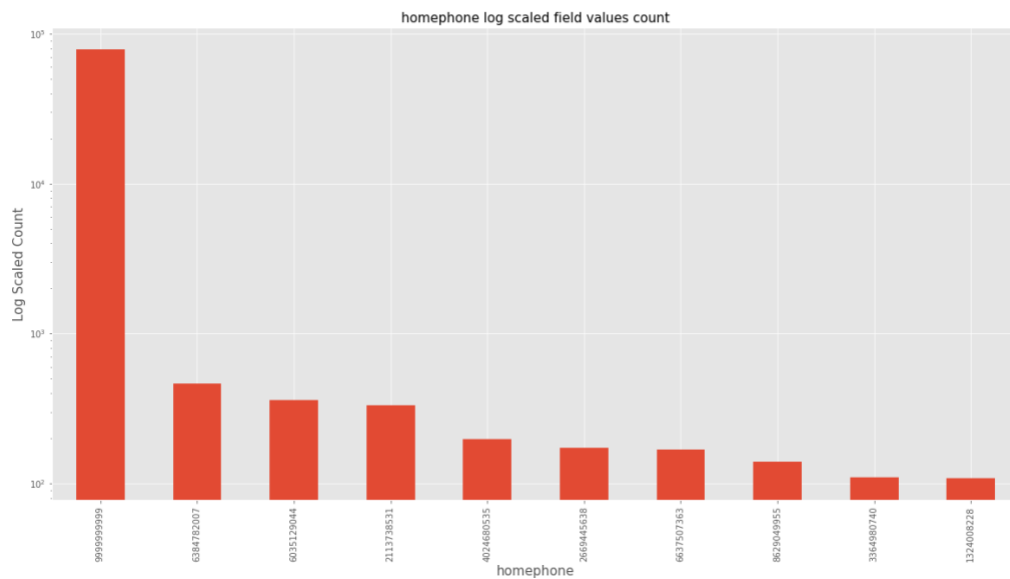


Field 9:

Name: homephone

Description: Home phone number of applicants.





Field 10:

Name: fraud label

Description: Whether the application was fraud or not.

'1' Indicates the application was fraudulent.

'0' indicates the application was non-fraudulent.

