Data Quality Report

Cheng Chen

3/4/2017

Part 1: File Description

File name: Applications Dataset

Source: Stephen Coggeshall, ID Analytics

Number of records: 100,000

Number of fields: 9 fields, 1 index, 2 date, 6 categorial

Time frame: 01/01/2015 - 12/31/2015

Part 2: List of info for each field

Field 1: record
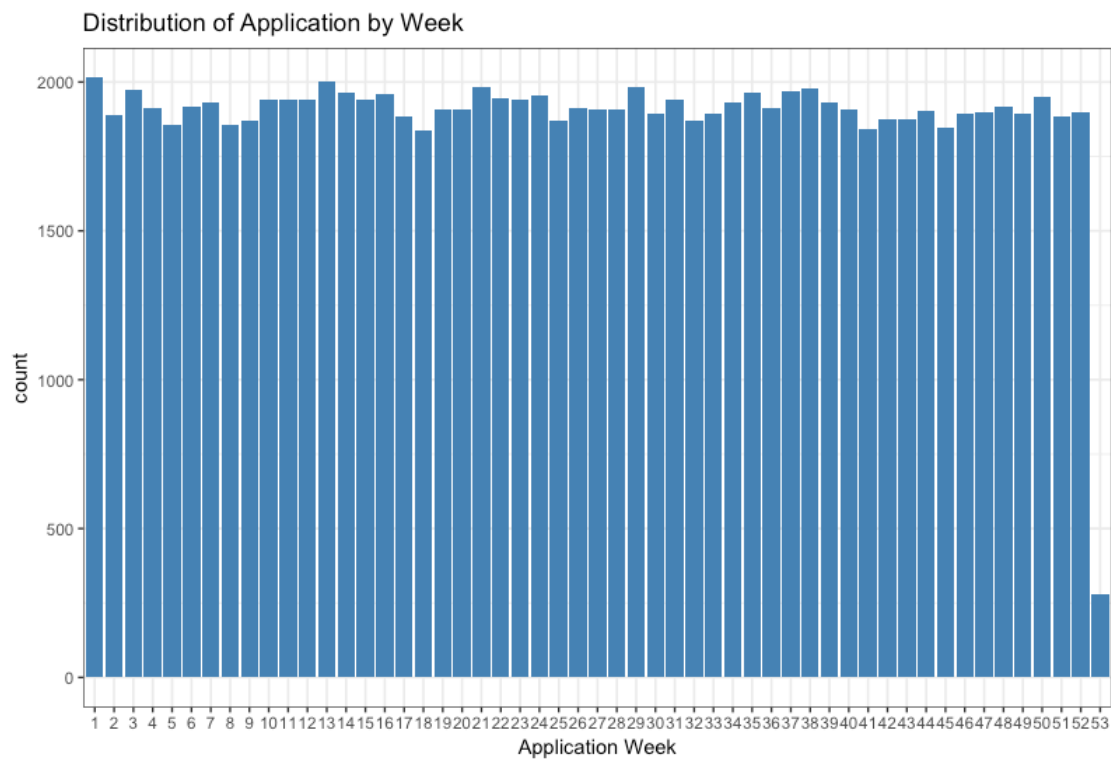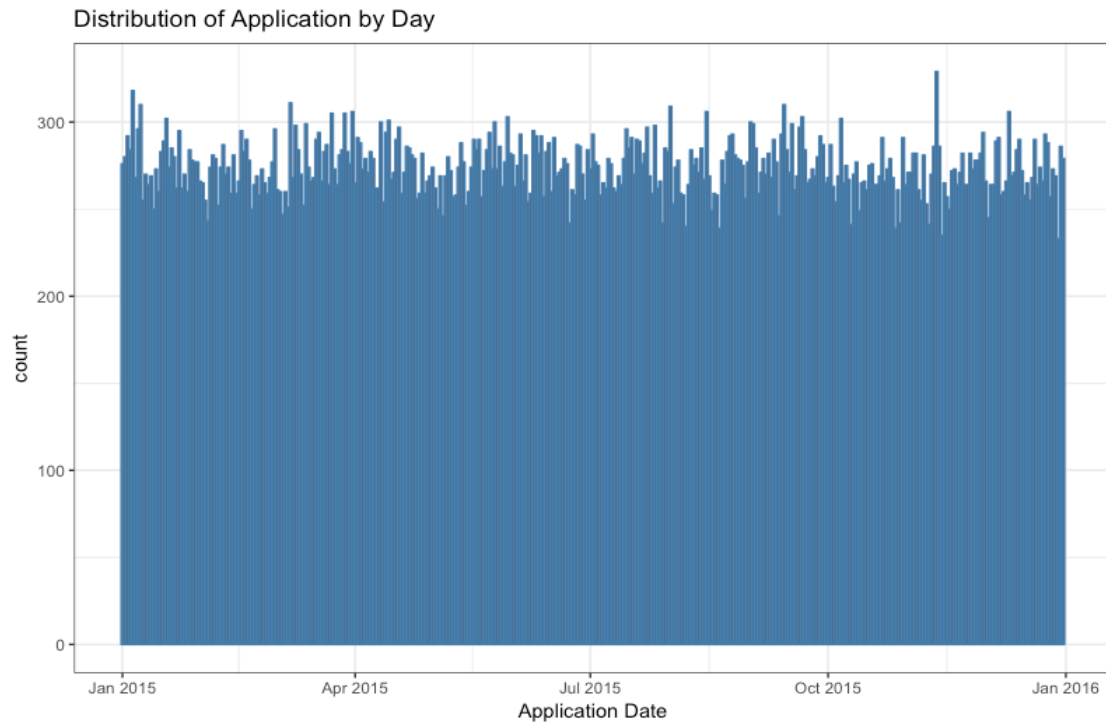
Description: Index of record

Number of unique values: 100,000, start from 1, interval is 1, no missing value
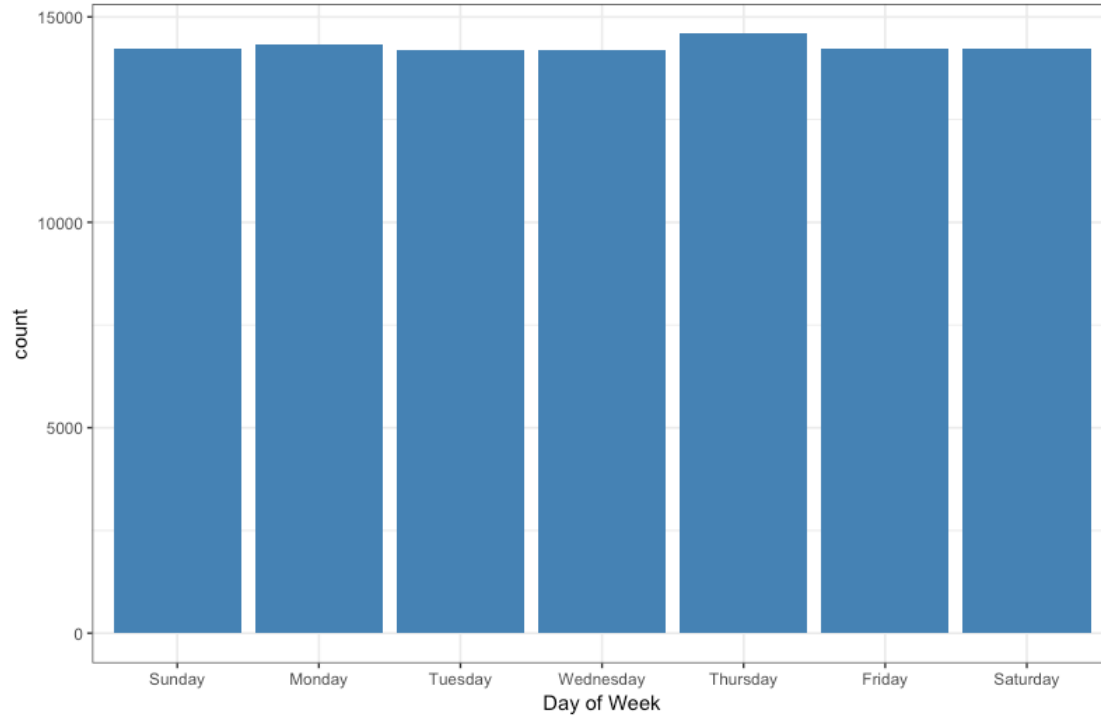
Field 2: date

Description: date of each application record, categorical with metric
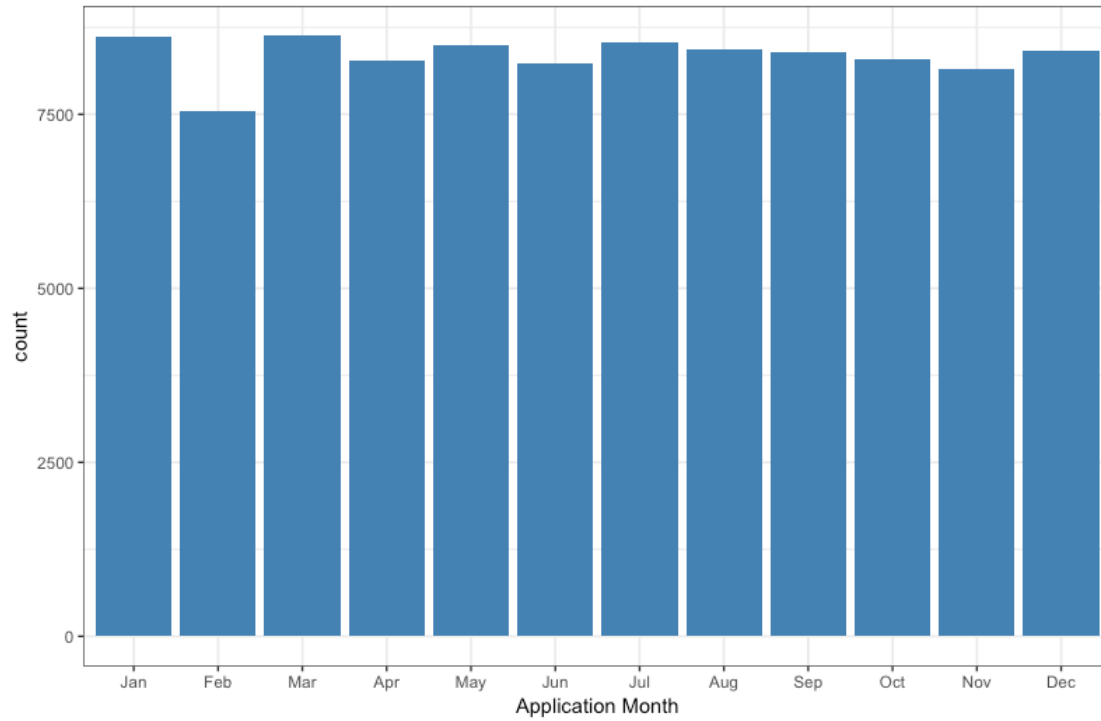
Percent of Populated: 100%, no missing values

Number of Unique Values: 365, from 01/01/2015 to 12/31/2015

Distribution of Application by Day

Distribution of Application by Week

Distribution of Application by Day of Week
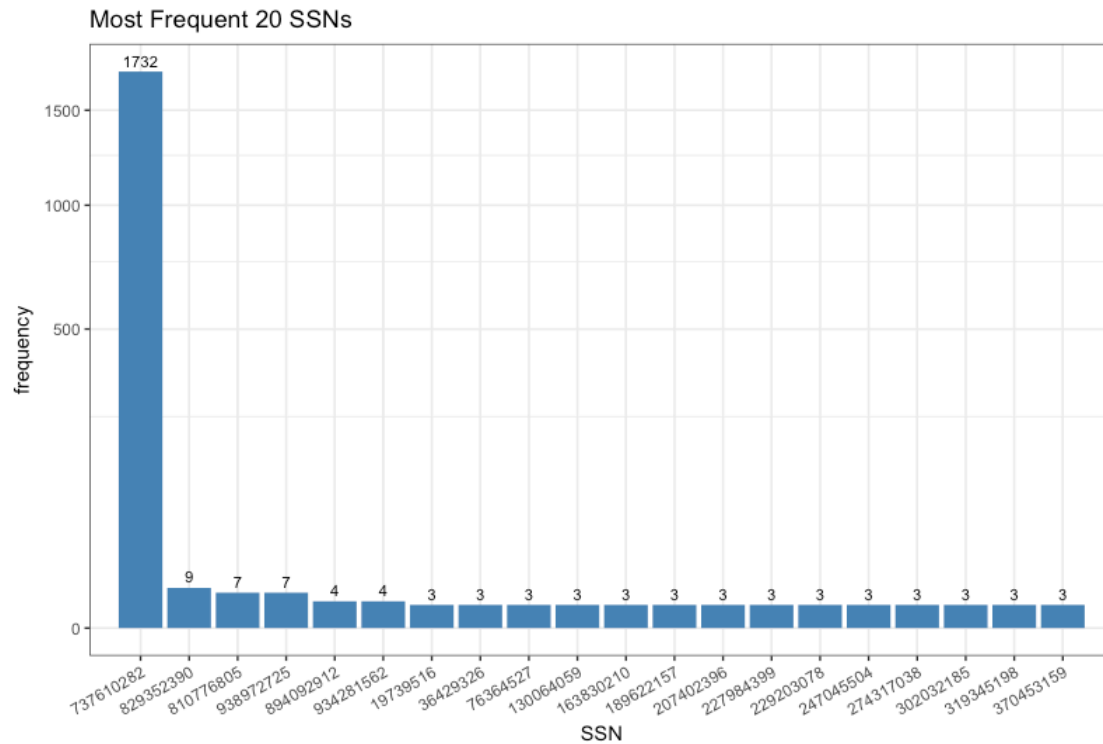

Distribution of Application by Month

Field 3: ssn

Description: ssn of each application record, nine-digit, categorical with no metric

Percent of Populated: 100%, no missing values

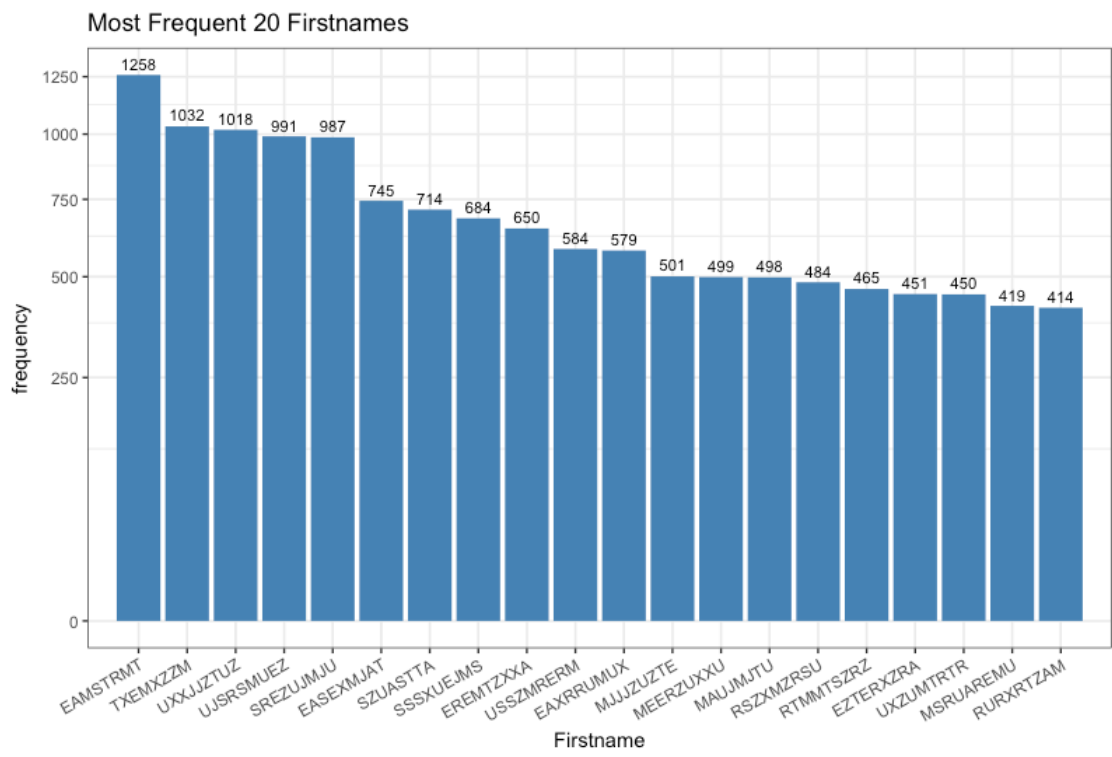Number of Unique Values: 96535 (Max: 999993079, Min:2503)

Most Frequent 20 SSNs

Field 4: firstname

Description: firstname of each application record, categorical with no metric

Percent of Populated: 100%, no missing values

Number of unique values: 16576
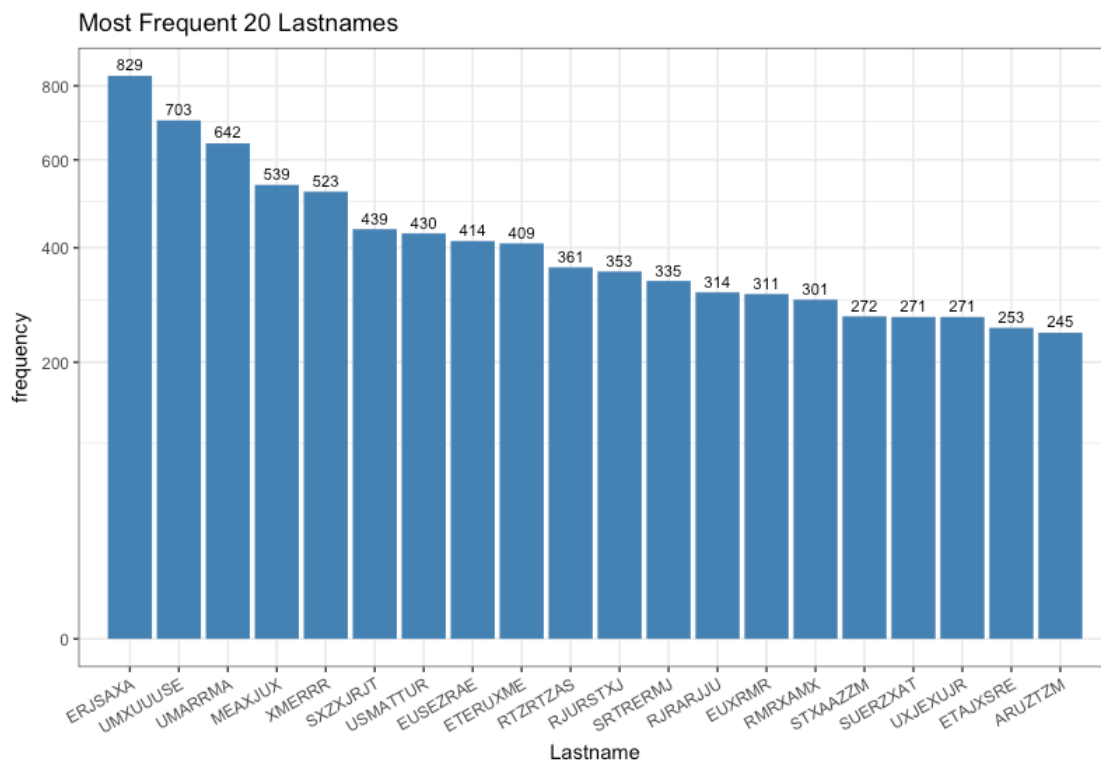
Most Frequent 20 Firstnames

Field 5: lastname

Description: lastname of each application record, categorical with no metric

Percent of Populated: 100%, no missing values

Number of unique values: 36312
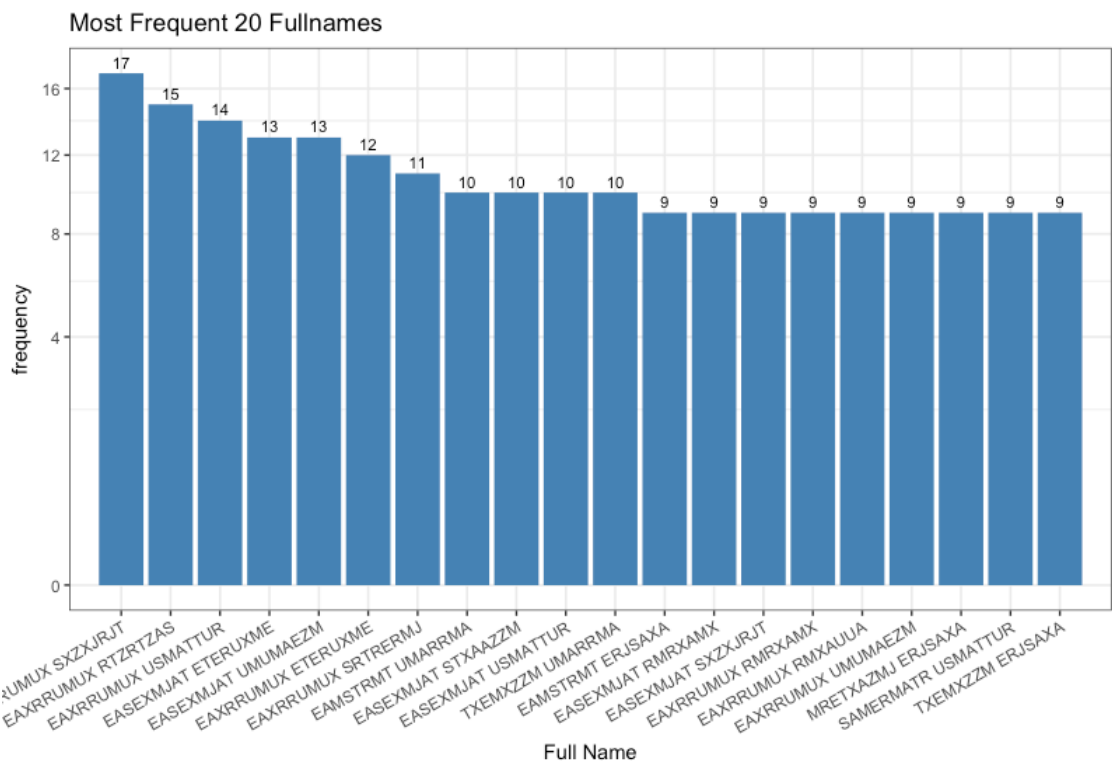

Most Frequent 20 Lastnames

Combined Fullname

Description: Fullname of each application record, categorical with no metric

Percent of Populated: 100%, no missing values

Number of Unique Values: 93726
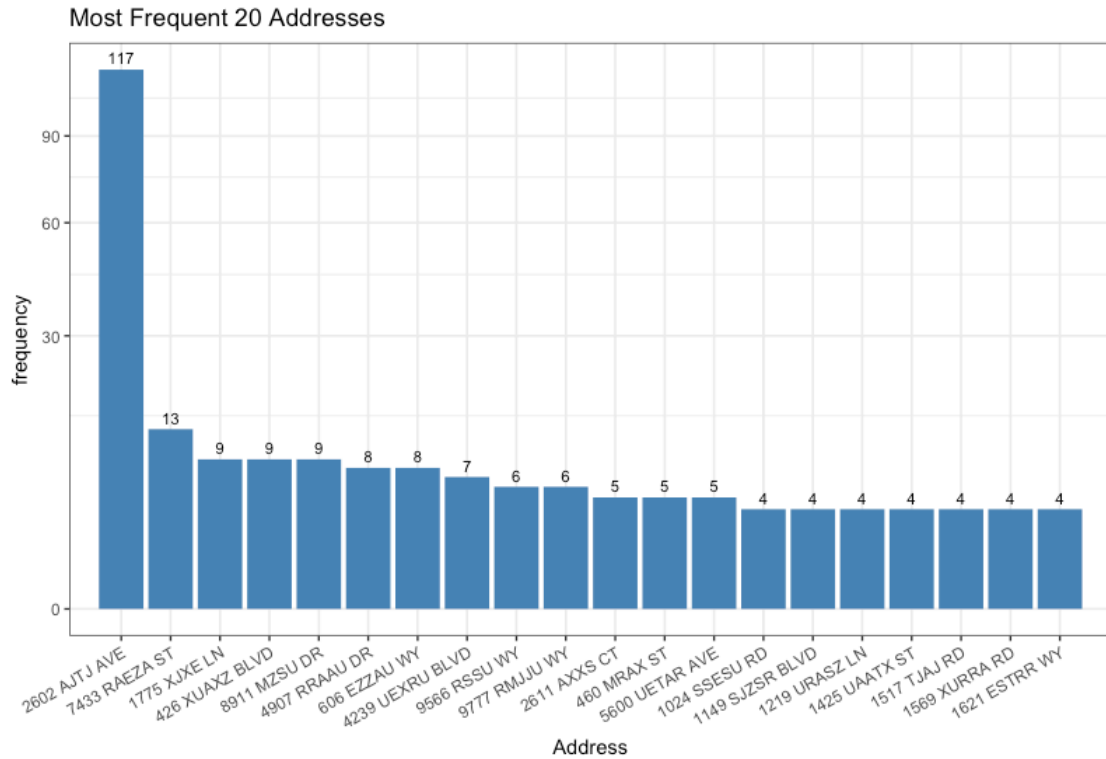

Most Frequent 20 Fullnames

Field 6: address

Description: address of each application record, categorical with no metric

Percent of Populated: 100%, no missing values

Number of Unique Values: 97563

Most Frequent address: 2602 AJTJ AVE
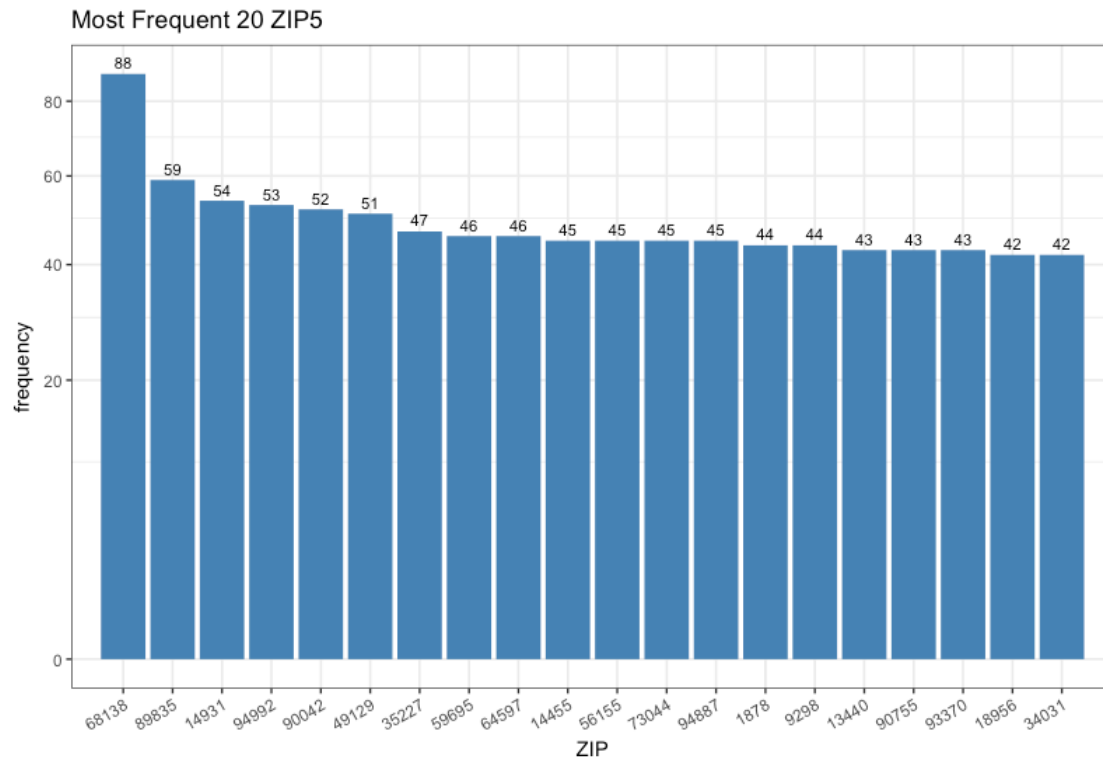

Most Frequent 20 Addresses

Field 7: ZIP5

Description: zip5 of each application record, categorical with no metric

Percent of Populated: 100%, no missing values

Number of Unique Values: 16547 (Max: 99999, Min: 2)
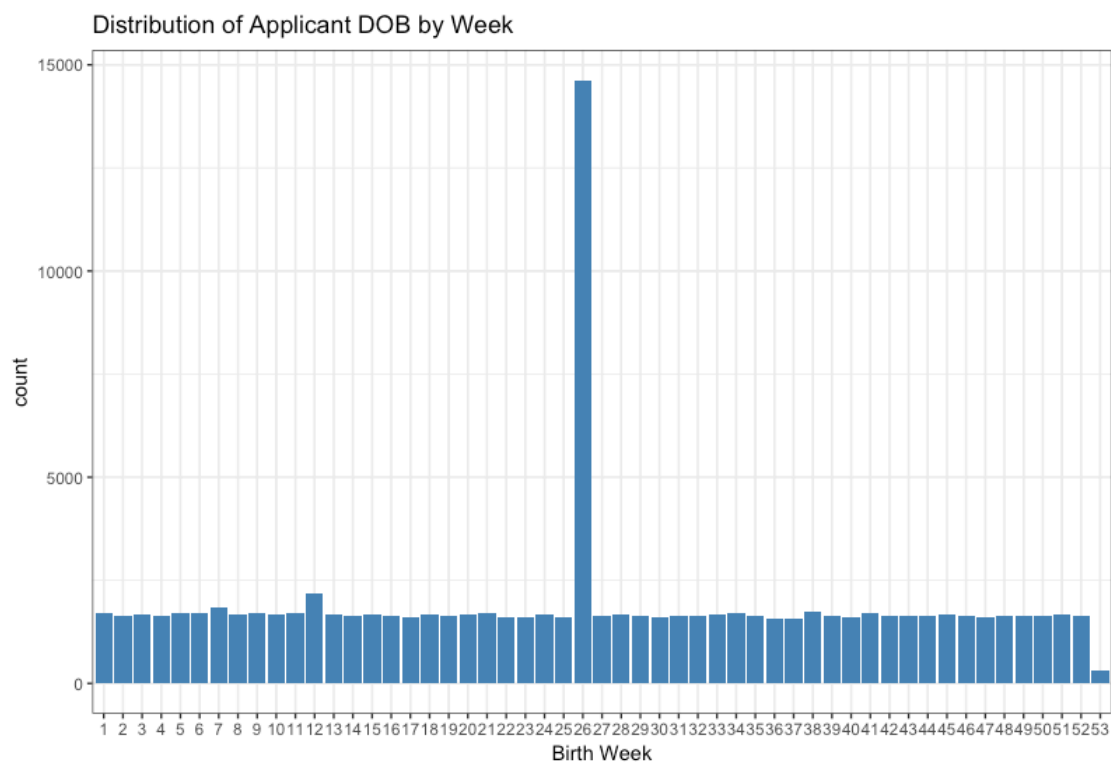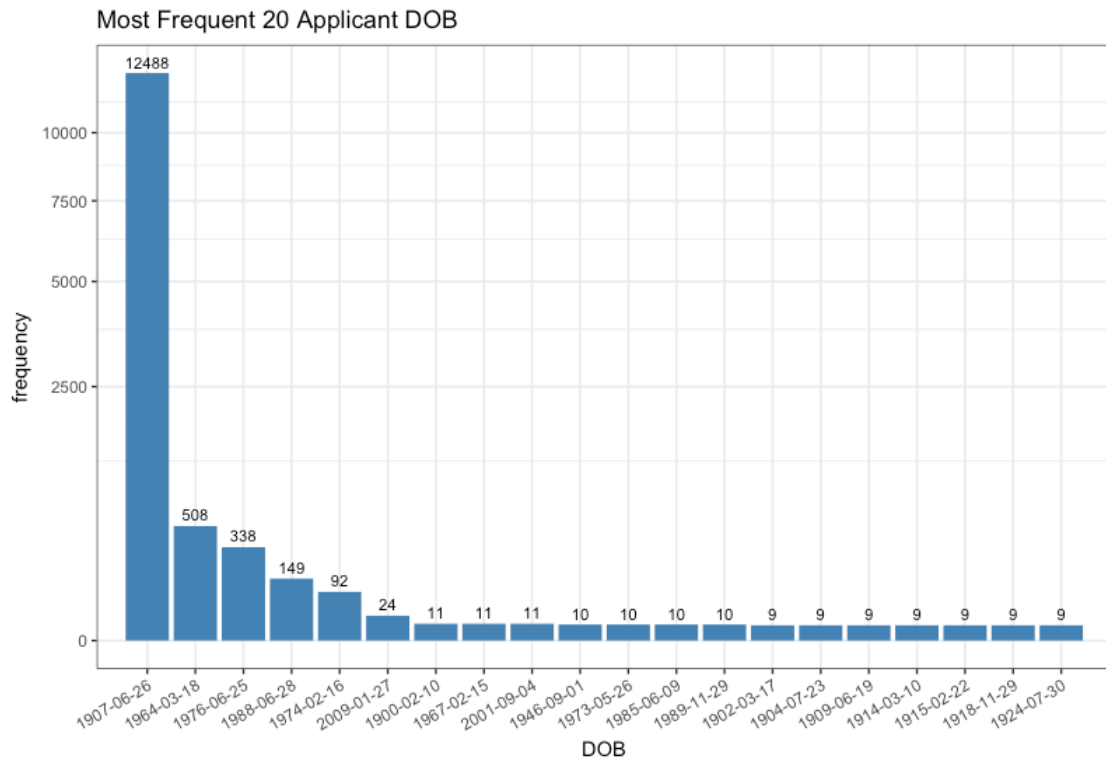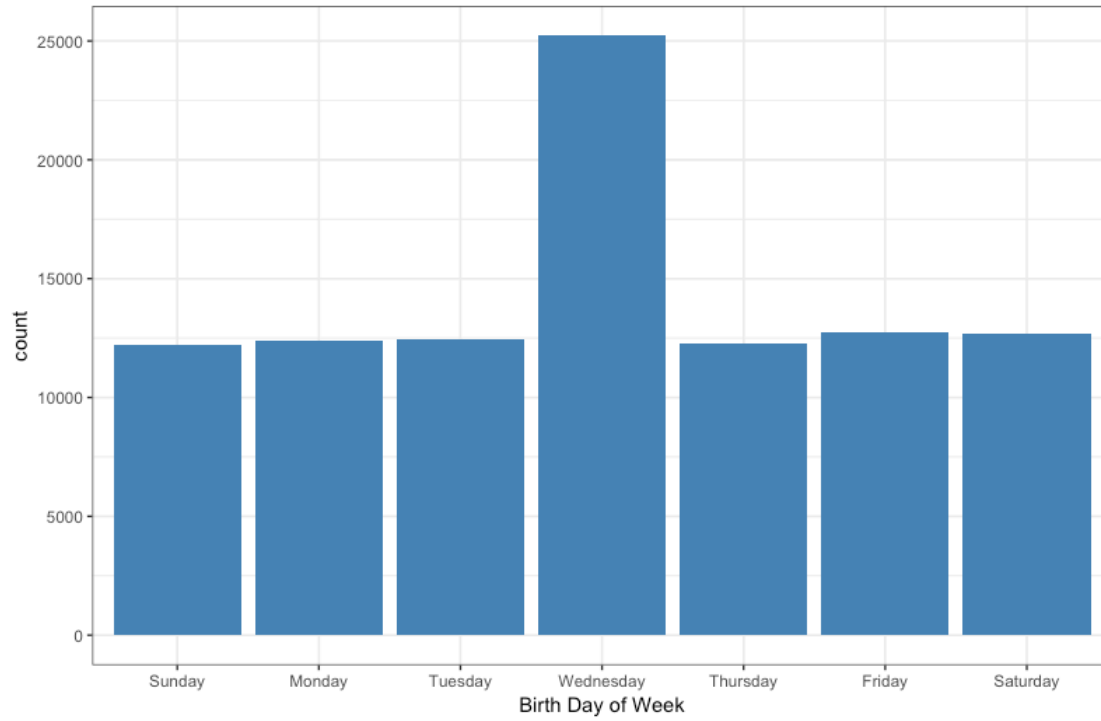

Most Frequent 20 ZIP5

Field 8: dob

Description: birth date of each applicant, categorical with metric

Percent of Populated: 100%, no missing values

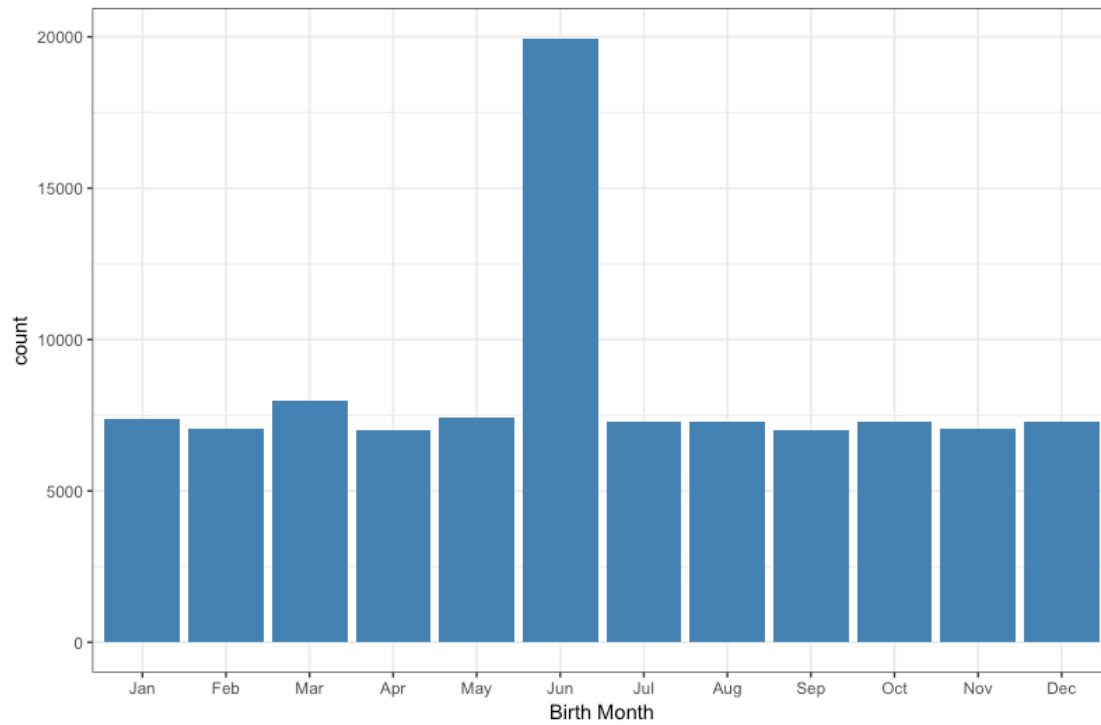Number of Unique Values: 36816, from 01/01/1900 to 10/31/2016

Most Frequent 20 Applicant DOB

Distribution of Applicant DOB by Week

Distribution of Applicant DOB by Day of Week

Distribution of Applicant DOB by Month

Field 9: homephone

Description: homephone of each application record, categorical with no metric

Percent of Populated: 100%, no missing values

Number of Unique Values: 22181 (Max: 9996906703, Min: 635392)

Most Frequent 20 homephone