

Fraud Profiling

MVP

Contents

Set Up	1
Query	1
Helper functions	3
Fraud/Charge-off Definitions	3
Total Volume	3
Volume Trends (Monthly)	3
Cost of Fraud	6
Loss Trends (Monthly)	7
Cost by charge-off type	10
Time Till Fraud/Charge-off Act	11
Marketing Channel	13
TO DO	14
August 06, 2018	

Set Up

```
library(ggplot2)
library(xtable)
library(scales)
```

Query

```
df <- web_db("select
  a.user_id,
  --f.user_email,
  a.is_charged_off,
  a.chargeoff_reason,
  a.total_amount_charged_off,
  a.hard_loss_charge_off,
  a.fees_charged_off,
  a.application_created_datetime,
  --a.funding_bank_account_id,
  --a.funding_amount,
  a.inception_date,
  --a.depository_id,
  --d.balance_date,
  --f.lead_source,
  --f.utm_source,
```

```

--f.utm_content,
--f.initial_product_selected,
f.utm_campaign,
f.utm_medium,
f.utm_source_clean,
f.utm_channel_type,
f.utm_channel_grouping,
coalesce(f.utm_channel_grouping, 'Unattributed') as marketing_channel_group,
f.is_signed_up,
f.first_account_type,
f.total_accounts_opened,
d.balance_date as negative_bal_date,
h.termination_date,
d.balance_date::date - a.inception_date::date as days_to_negative_bal,
h.termination_date::date - a.inception_date::date as days_to_termination,
case when d.balance_date is null then 0::boolean else 1::boolean end neg_balance_ever,
case when f.utm_source in ('ph', 'fb', 'adwords', 'atez', 'dailykos', 'nw1',
    'inboxdollars', 'instagram') then f.utm_source
    when f.utm_source in ('facebook', 'fb-unbounced', 'facebook.com') then 'fb'
    when f.utm_source is null then 'none'
    else 'utm_other' end marketing_channel,
case when a.chargeoff_reason is null then 'Not Fraud'
    when a.chargeoff_reason in ('ATM Dep Chargeback', 'Bank by Mail / In Person',
    'Mobile Dep Chargeback', 'Early Detected Fraud',
    'Opening Dep Chargeback', 'Other') then
        'ATM Chargeback/Early Detected/Mobile Dep Chargeback/Other'
    else a.chargeoff_reason end fraud_type
from public.dt_accounts a
left join (
    with neg_bal as (
        select *,
            max(current_balance) over (order by
                balance_date rows between 2 preceding and current row) as max_balance
        from public.depositary_balance
        where current_balance < 0
        order by balance_date
    )
    select depository_id, min(balance_date) - integer '3' as balance_date
    from neg_bal
    where max_balance < 0
    group by 1
    ) d on a.unique_account_id = concat('d',d.depository_id)
left join public.dt_users f on a.user_id = f.user_id
left join public.dt_fraud_users_with_info h on f.external_user_id = h.external_user_id
where f.total_accounts_opened > 0
and a.account_type = 'Summit'
and f.first_account_type = 'Summit'
order by a.user_id;")

df <- data.table(df)

```

Helper functions

```
round0 <- function(x) format(round(x,0), nsmall = 0)
round2 <- function(x) format(round(x, 2), nsmall = 2)
percent_round2 <- function(x) paste0(round2(100*x), "%")
make_money <- function(x) paste("$",format(round(x, 2), nsmall = 2, big.mark=","),sep="")
my_colors <- c("#1165bf", "#119fbf", "#37bf72", "#f2ac16", "#d94514", "#b6babf")
```

Fraud/Charge-off Definitions

- **NSF Behavior:** *Non-Sufficient Funds or overdrafts often related to debit preauthorizations or debit hold issues.*
- **Mobile Deposit Chargeback:** *Mobile deposit returned by issuing bank, often leading to overdrawn accounts when customers spend the provisional credits.*
- **External Transfer Chargeback** *Customers spending provisionally credited funds transferred into their account before the funds are returned to the sending bank.*
- **Opening Deposit Chargeback** *Deposit used to open account is returned to issuing bank.*
- **ATM Deposit Chargeback** *Spending provisionally credited funds from an ATM deposit prior to the funds being returned to the issuing bank.*
- **Bank by Mail/In OPerson** *Spending provisionary funds from a bank or in person deposit before funds are returned to the issuing institution.*
- **Other**

Total Volume

Table 1: Overall count and cost of fraud (to date).

count	total_loss	hard_loss	fees
4222	\$1,635,016.70	\$1,365,837.78	\$269,178.93

Table 2: Total counts by charge-off category

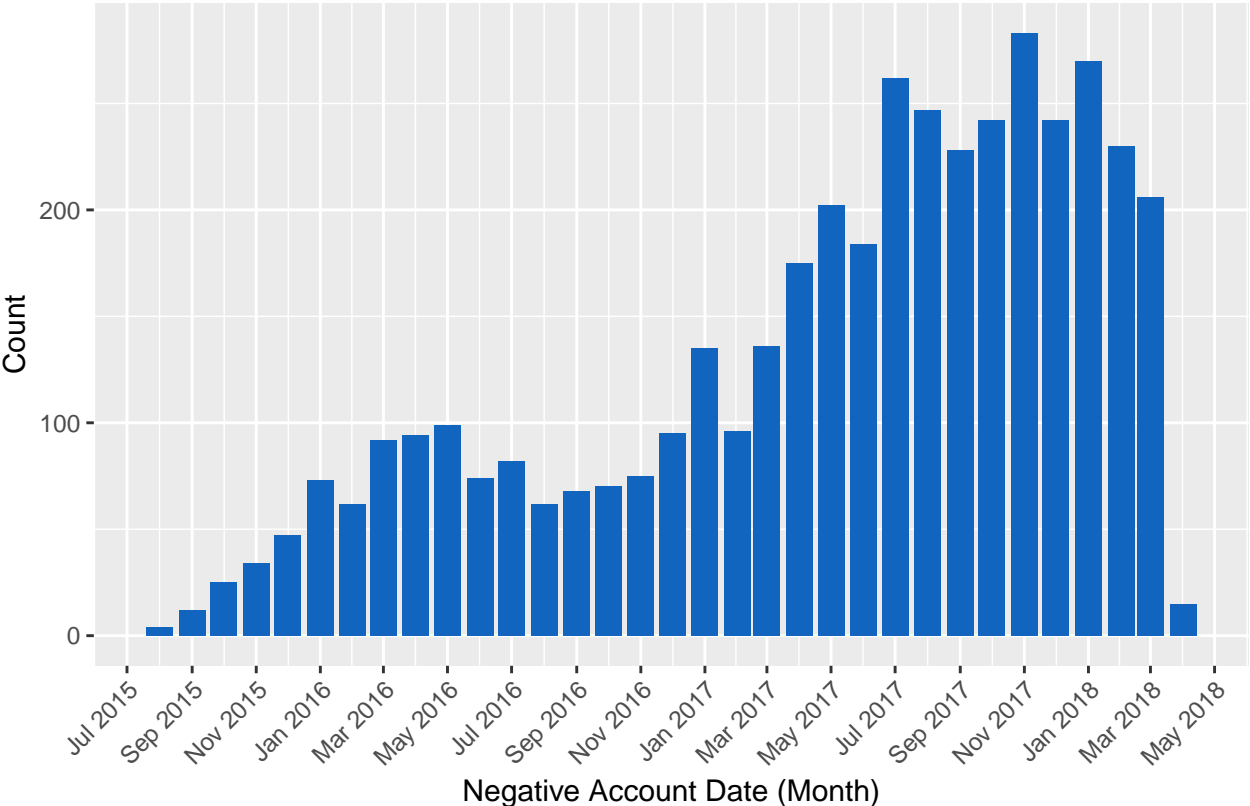
chargeoff_reason	count	rate
NSF Behavior	2494	59.07%
Mobile Dep Chargeback	413	9.78%
Ext Tran Chargeback	1207	28.59%
Opening Dep Chargeback	10	0.24%
ATM Dep Chargeback	89	2.11%
Bank by Mail / In Person	5	0.12%
Other	4	0.09%

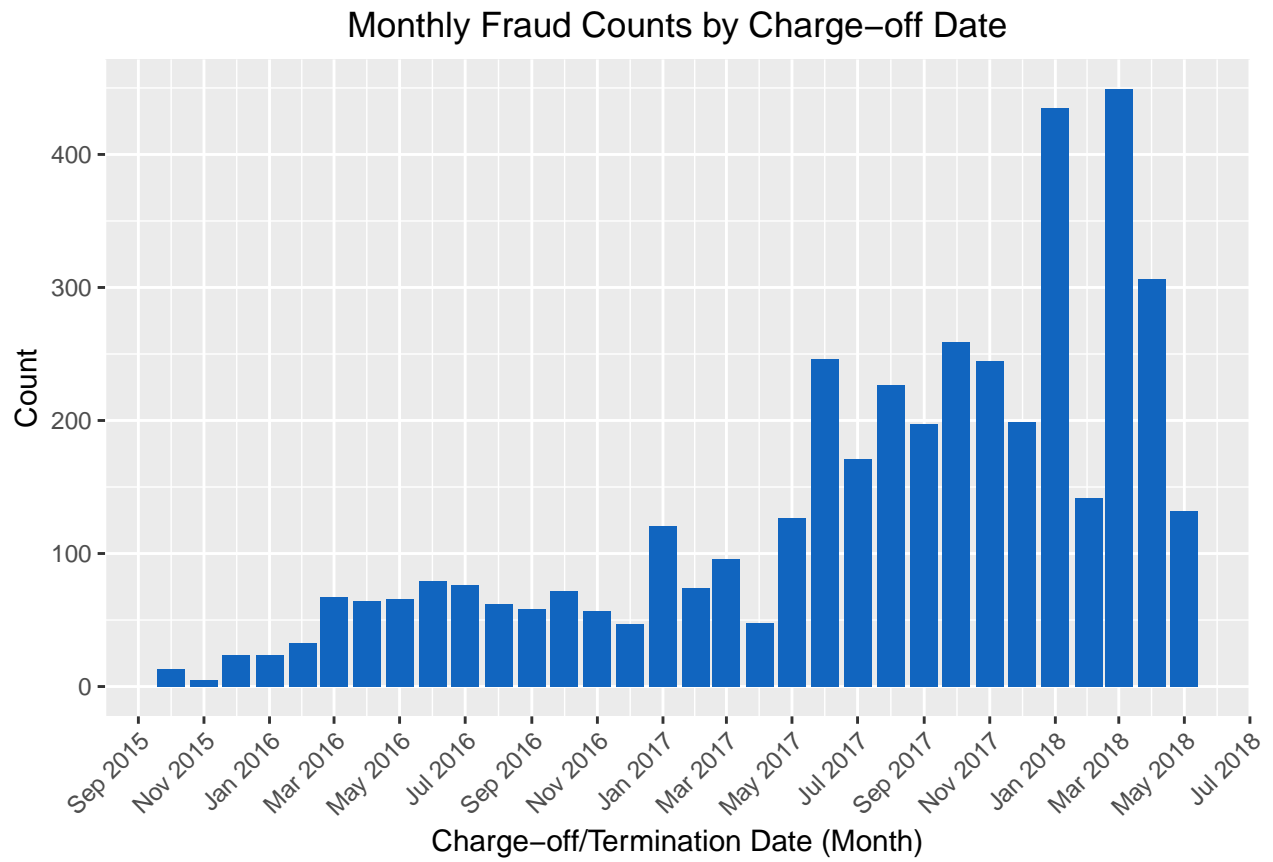
The most common type of charge-off is NSF Behavior, which accounts for 59.07% of all charge-off types.

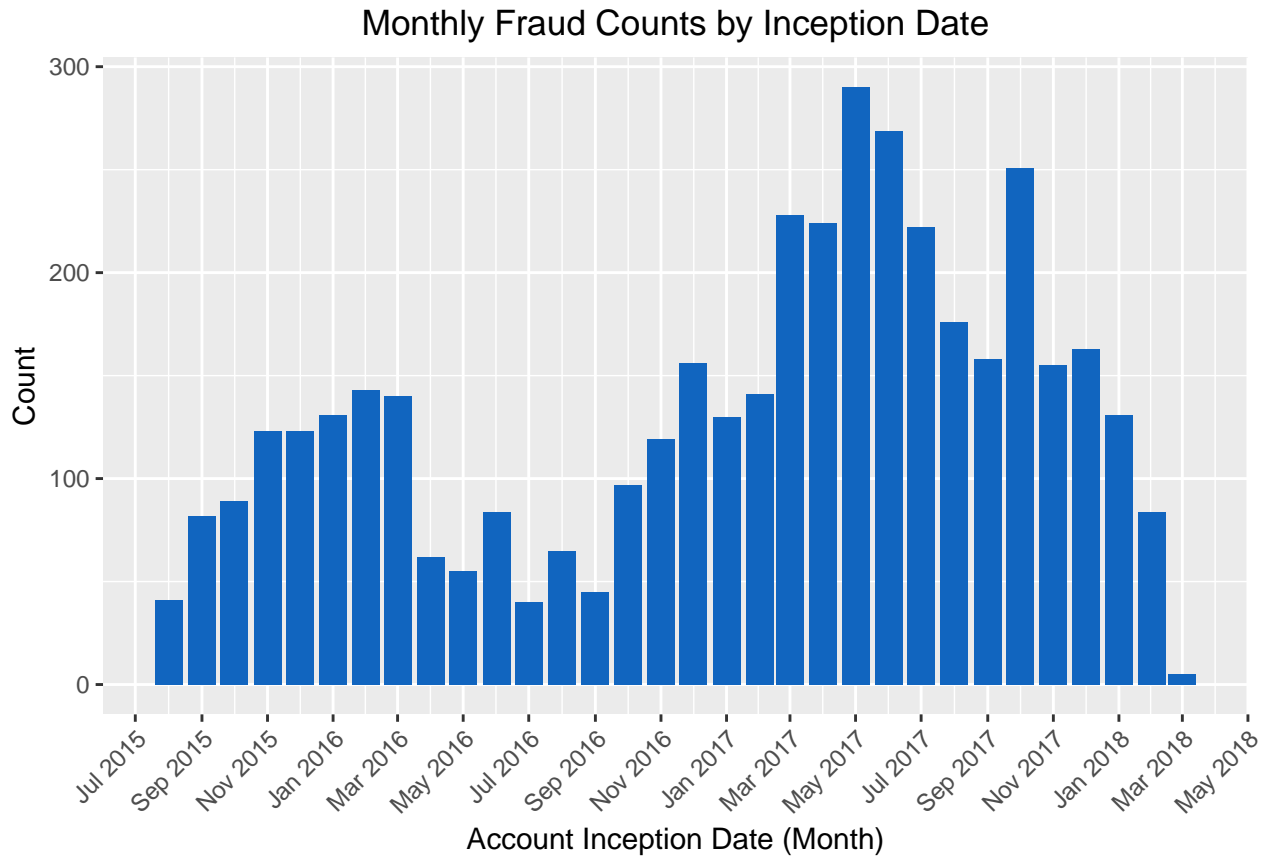
Volume Trends (Monthly)

Counts by account going negative

Monthly Fraud Counts by Negative Account Date







Cost of Fraud

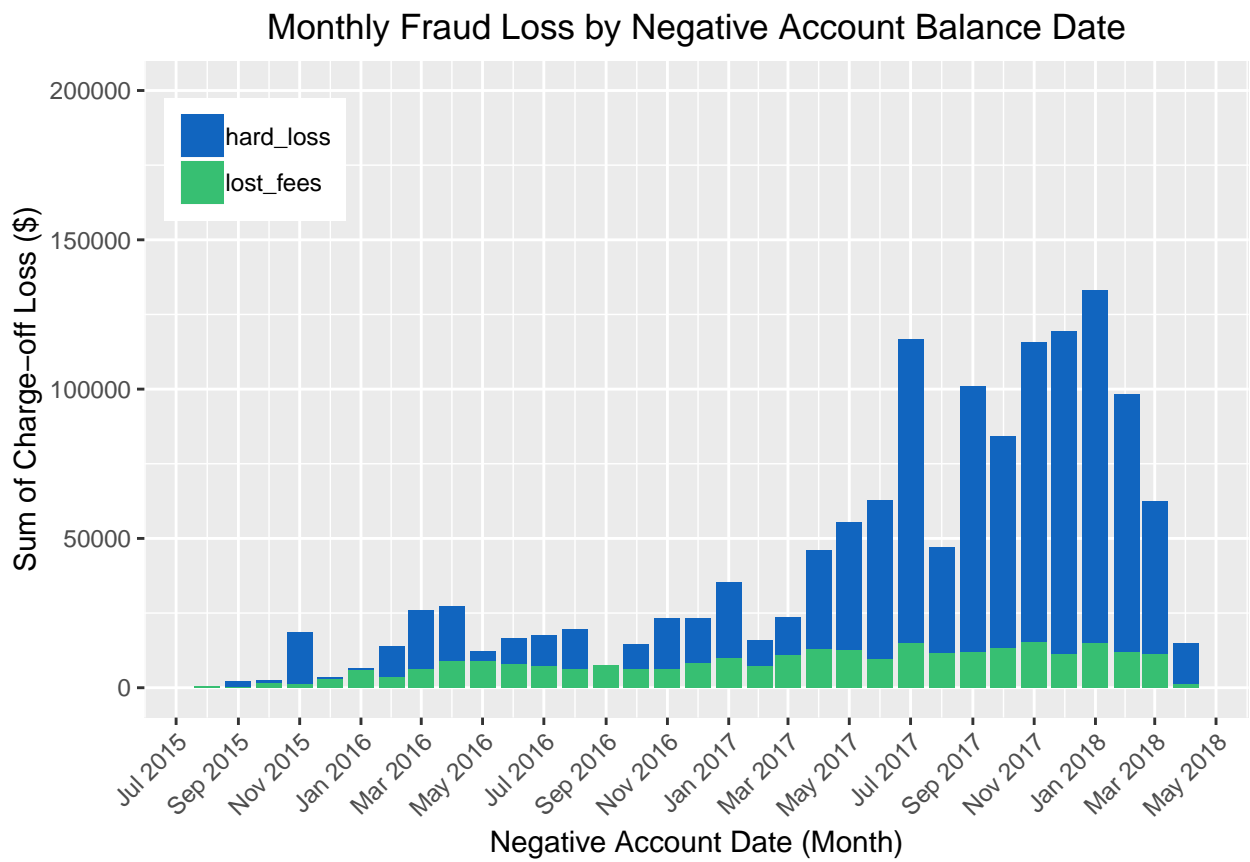
Table 3: Total charged-off losses (principal + realized)

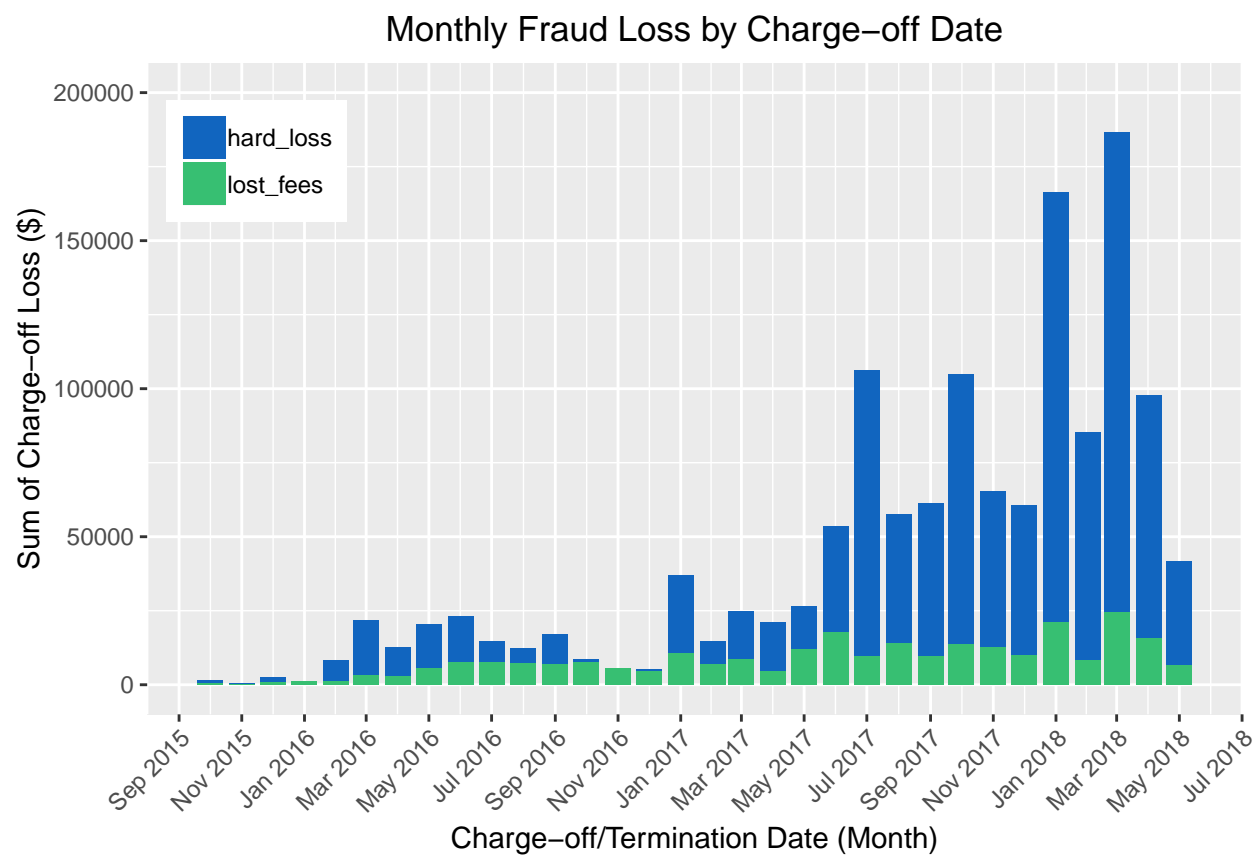
chargeoff_reason	sum	percent	mean	min	q25	q50	q75	max
NSF Behavior	\$237,003.91	14.50%	\$95.03	\$0.01	\$15.92	\$38.45	\$118.99	\$8,165.58
Mobile Dep Chargeback	\$400,062.11	24.47%	\$968.67	\$0.07	\$193.60	\$579.98	\$1,329.51	\$8,001.67
Ext Tran Chargeback	\$848,532.33	51.90%	\$703.01	\$0.01	\$115.02	\$216.12	\$550.95	\$45,765.79
Opening Dep Chargeback	\$4,760.12	0.29%	\$476.01	\$0.03	\$31.25	\$106.29	\$216.96	\$2,935.87
ATM Dep Chargeback	\$122,241.57	7.48%	\$1,373.50	\$17.04	\$313.54	\$710.09	\$1,343.58	\$12,021.23
Bank by Mail / In Person	\$11,965.69	0.73%	\$2,393.14	\$110.82	\$174.77	\$816.13	\$2,130.16	\$8,733.81
Other	\$10,450.97	0.64%	\$2,612.74	\$139.56	\$480.92	\$2,613.51	\$4,745.33	\$5,084.40

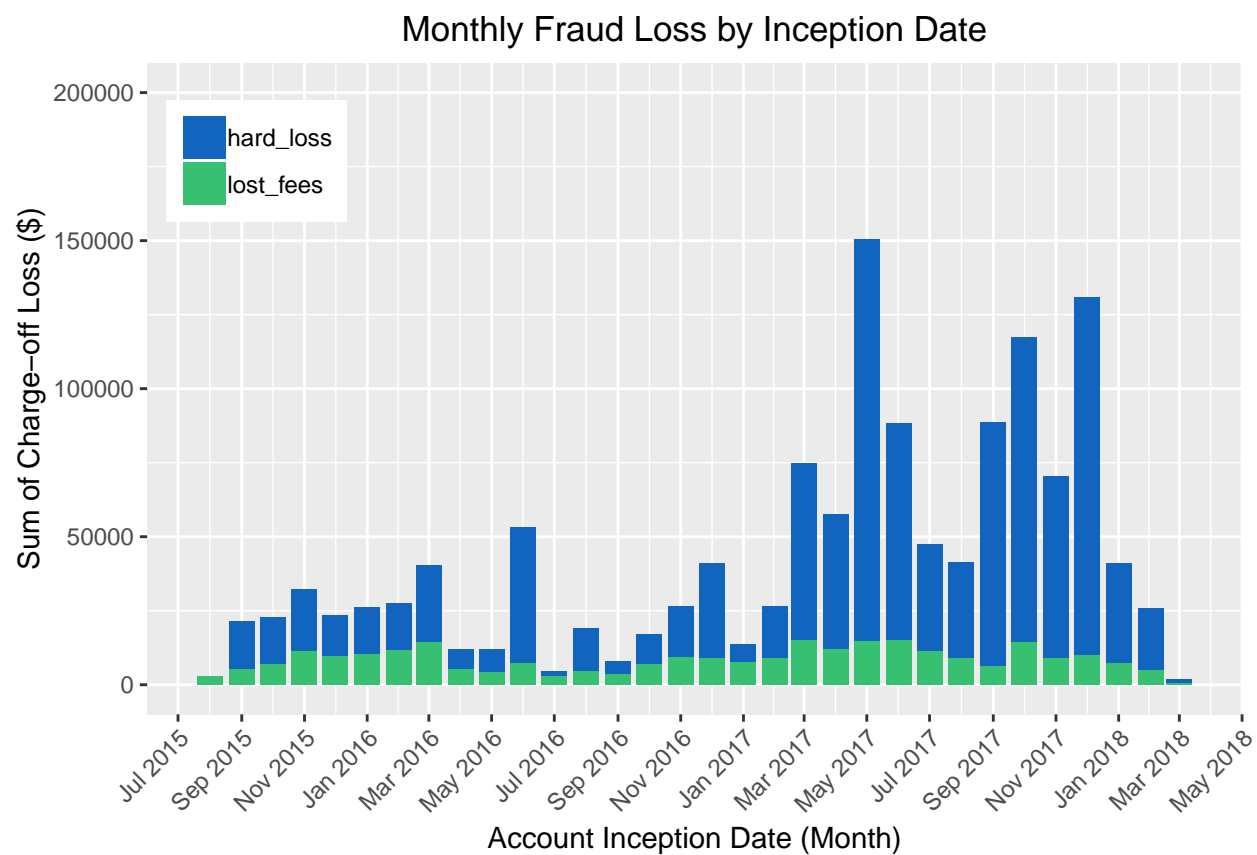
Table 4: Total charged-off losses (realized only)

chargeoff_reason	sum	percent	mean	min	q25	q50	q75	max
NSF Behavior	\$110,526.34	8.09%	\$44.32	\$0.00	\$0.85	\$15.25	\$33.58	\$8,075.58
Mobile Dep Chargeback	\$366,152.04	26.81%	\$886.57	\$0.00	\$103.47	\$481.14	\$1,249.51	\$7,903.67
Ext Tran Chargeback	\$748,748.50	54.82%	\$620.34	\$0.00	\$25.09	\$120.06	\$465.01	\$45,695.79
Opening Dep Chargeback	\$4,120.37	0.30%	\$412.04	\$0.00	\$7.44	\$19.02	\$118.45	\$2,845.87
ATM Dep Chargeback	\$114,470.69	8.38%	\$1,286.19	\$0.00	\$223.54	\$620.09	\$1,272.10	\$11,931.23
Bank by Mail / In Person	\$11,646.87	0.85%	\$2,329.37	\$0.00	\$146.77	\$726.13	\$2,040.16	\$8,733.81
Other	\$10,172.97	0.74%	\$2,543.24	\$139.56	\$409.67	\$2,522.01	\$4,655.58	\$4,989.40

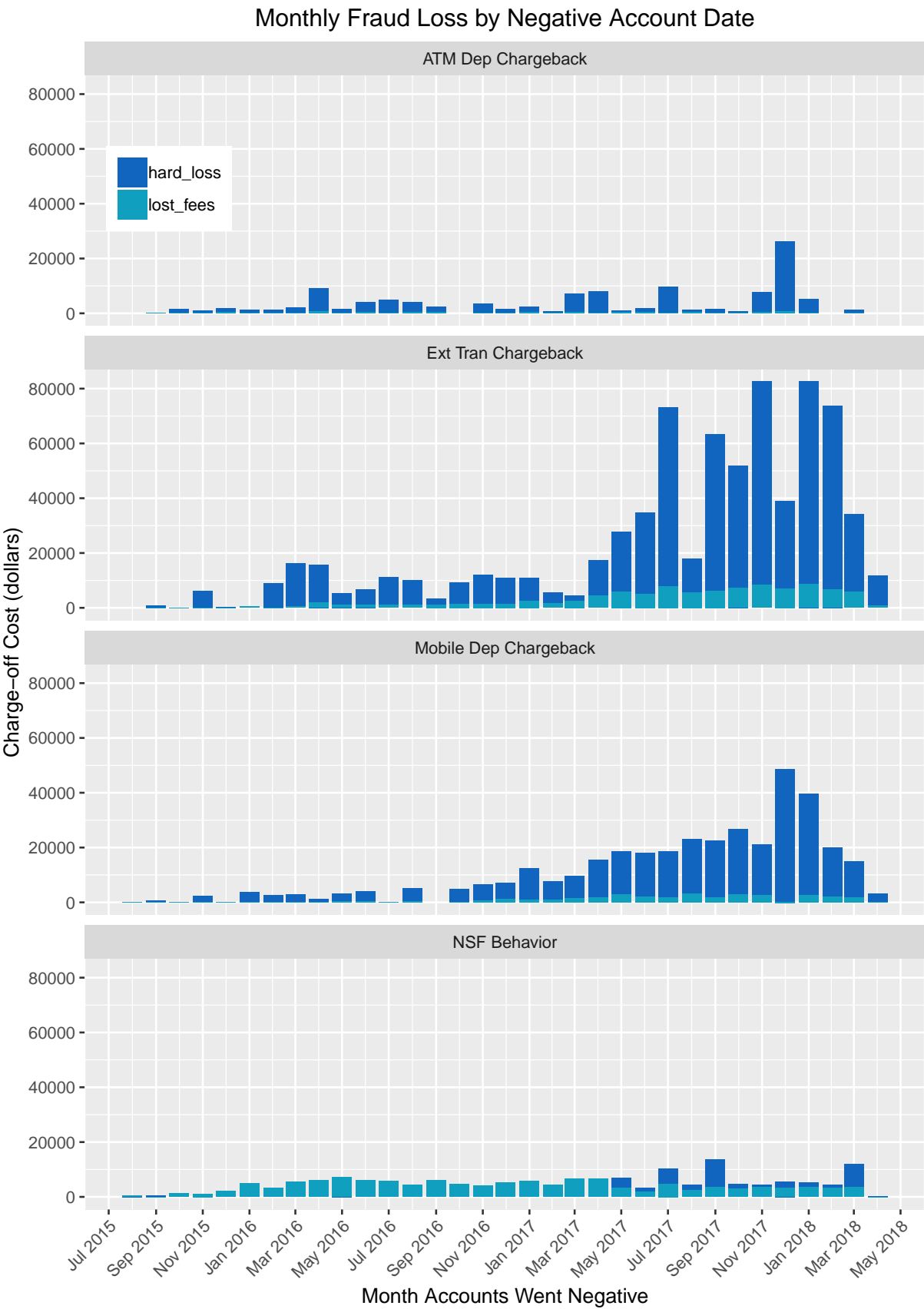
Loss Trends (Monthly)







Cost by charge-off type



Note: Bank by Mail / In Person, Opening Dep Chargeback, and Other charge-off reasons remove due to low counts.

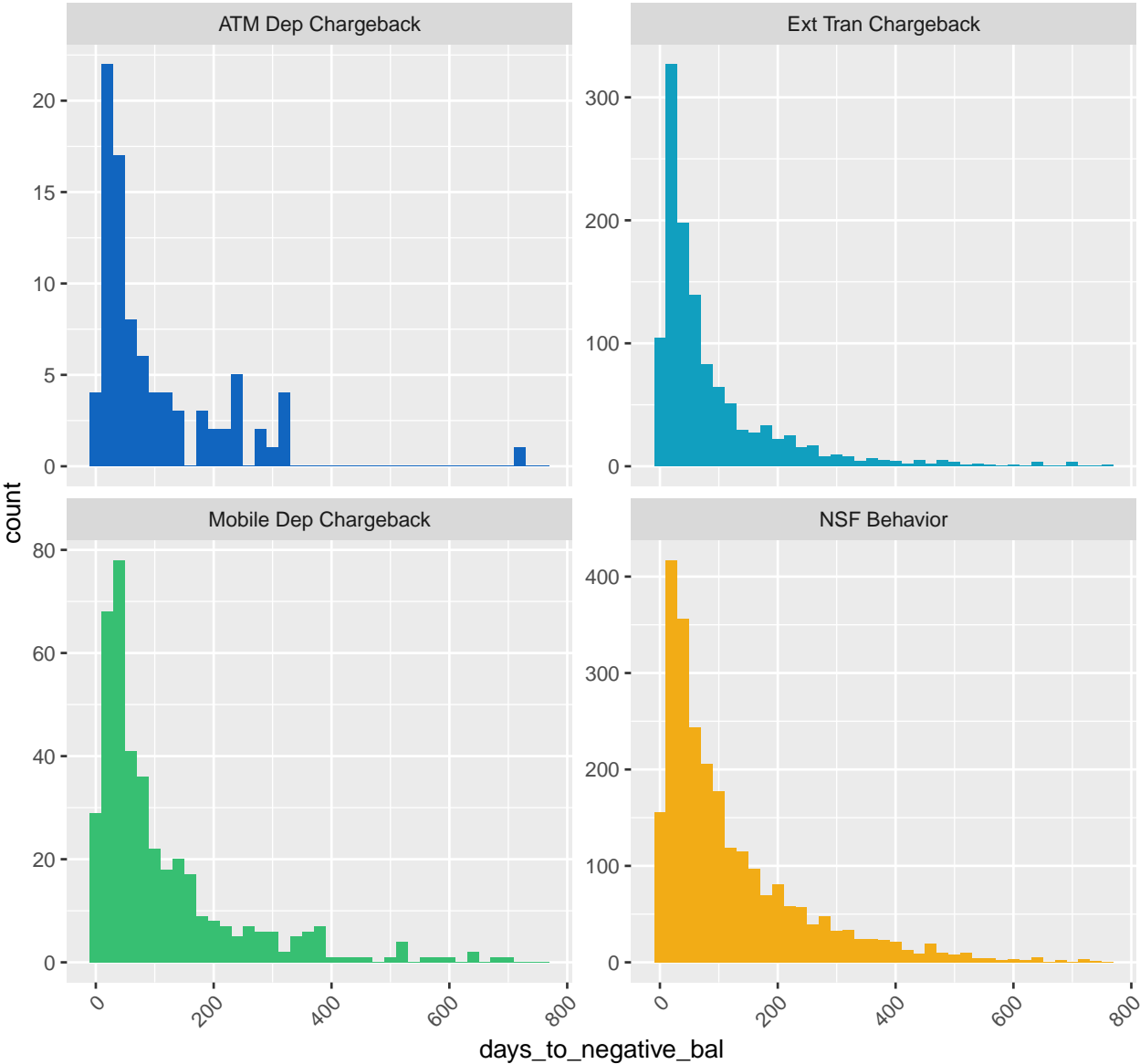
Time Till Fraud/Charge-off Act

Time between account creation and first day an account goes negative and stays negative for at least three days in a row.

Table 5: Days from account creation to negative balance

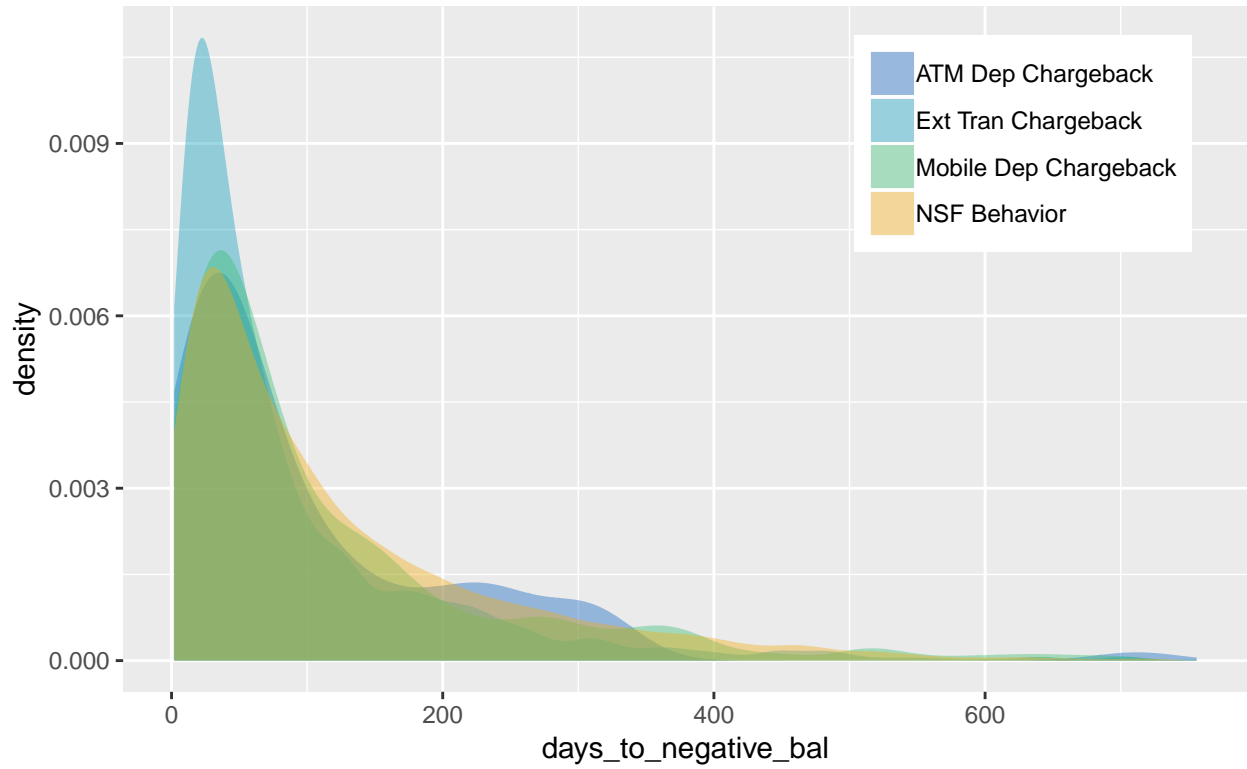
chargeoff_reason	mean	min	q25	q50	q75	max
NSF Behavior	120.47	3	32.25	78.00	166.00	731
Mobile Dep Chargeback	113.81	2	33.00	65.00	148.00	700
Ext Tran Chargeback	86.61	2	21.00	47.00	107.00	756
Opening Dep Chargeback	42.30	0	4.25	7.50	56.75	161
ATM Dep Chargeback	101.06	5	25.75	52.50	140.25	712
Bank by Mail / In Person	84.00	9	12.00	39.00	103.00	257
Other	57.25	9	30.75	47.50	74.00	125

Time to Negative Balance



Note: Bank by Mail / In Person, Opening Dep Chargeback, and Other charge-off reasons remove due to low counts.

Time to Negative Balance



Note: Bank by Mail / In Person, Opening Dep Chargeback, and Other charge-off reasons remove due to low counts.

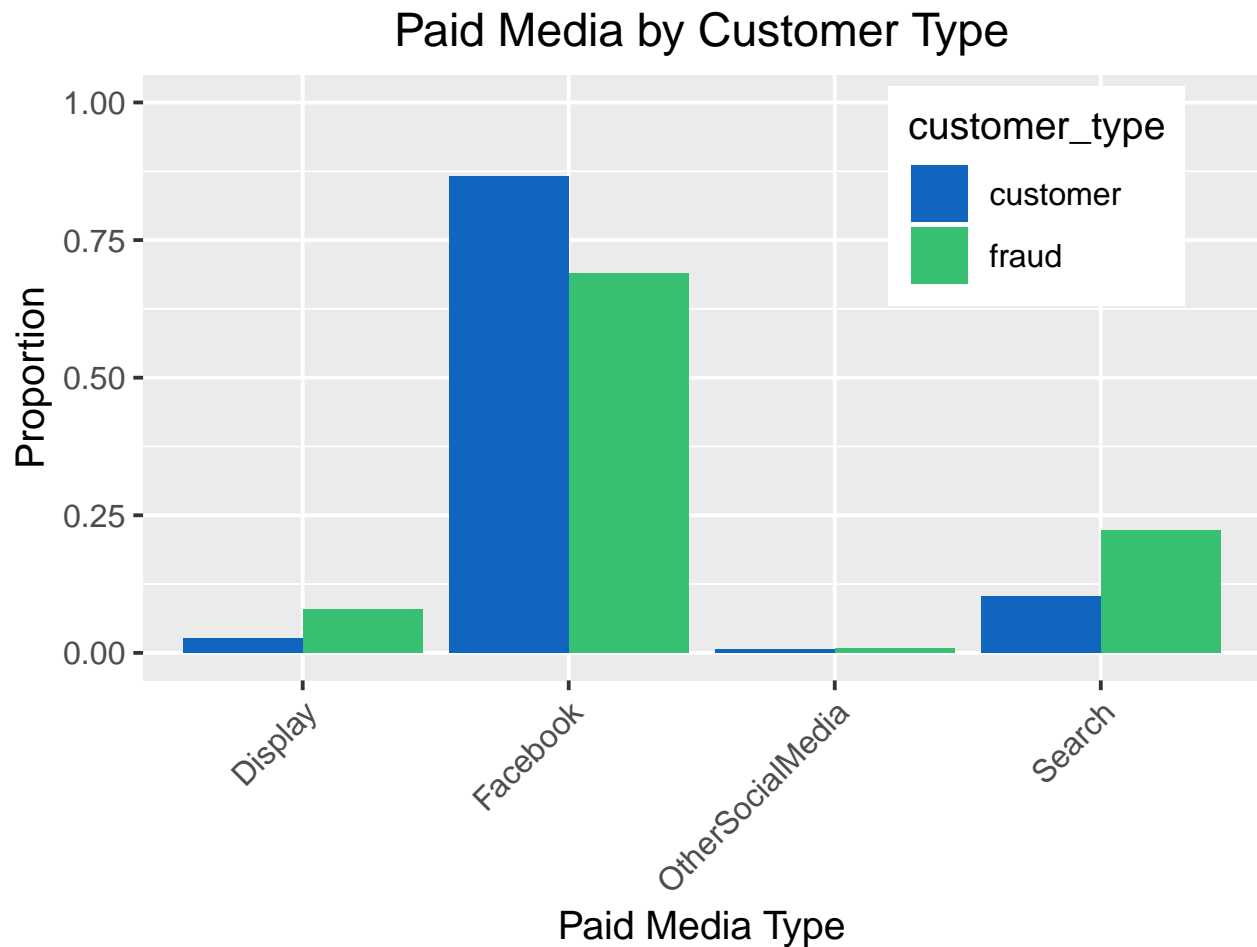
Marketing Channel

Table 6: Maketing channel group by customer vs fraud/charge-off

	Affiliates	Ignore	Organic	PaidMedia	Unattributed
customer	74362 (48.62%)	80 (0.05%)	353 (0.23%)	24619 (16.1%)	53542 (35%)
fraud/charge-off	1340 (31.74%)	1 (0.02%)	9 (0.21%)	1421 (33.66%)	1451 (34.37%)

Table 7: Marketing channel group by charged-off reason

	Affiliates	Organic	PaidMedia	Unattributed
ATM Dep Chargeback	13 (14.61%)	1 (1.12%)	46 (51.69%)	29 (32.58%)
Bank by Mail / In Person	0 (0%)	0 (0%)	4 (80%)	1 (20%)
Ext Tran Chargeback	420 (34.8%)	2 (0.17%)	313 (25.93%)	472 (39.11%)
Mobile Dep Chargeback	110 (26.63%)	0 (0%)	143 (34.62%)	160 (38.74%)
NSF Behavior	797 (31.97%)	6 (0.24%)	908 (36.42%)	782 (31.37%)
Opening Dep Chargeback	0 (0%)	0 (0%)	7 (70%)	3 (30%)
Other	0 (0%)	0 (0%)	0 (0%)	4 (100%)



TO DO

- Compare counts of fraud vs total accounts to see if rate of fraud is increasing or not
- Charge-off per account incepted
- ~~Cost by chargeoff reason in a stacked bar -- fees vs hard loss by fraud loss type (facet_wrap?)~~
- ~~Remove low counts from graphs and add footnote of their removal~~
- ~~Switch from histograms to density plots~~
- Look at edge cases, give some qualitative insight
- ~~Ignore ignore from marketing tables~~
- ~~Look at utm channel type to see what is causing the increase in PaidMedia~~
- Break down also by browser (chrome, etc)
- Distribution of accounts by states (counts of accounts vs fraud)
- Profiling fraud vs non with account origination data
- ~~Add monthly counts of attempted date (aka, negative balance).~~
- Break down fraud ratio by chrageoff type

- Look at chargeoff types that have min of \$0
- Remove low count of charge off groups