

Fraud Profiling

Contents

Overview	1
Fraud/Charge-off Definitions	1
Fraud Volume	2
Monthly Counts	3
Monthly Fraud Rate	6
Cost of Fraud	7
Monthly Cost	7
Cost by charge-off type	11
Account Severity	12
Time Till Fraud/Charge-off Act	14
Marketing Channels and Fraud	16
Paid Media	17
Negative Account Balance Data by Fraud/Risk Group	19
Recomentations	20
Data Quality	20
Time to Account Termination	20
Data Breaches and Scripted Attacks on the Easiest Targets	20
TO DO	20
August 08, 2018	

Overview

Fraud/Charge-off Definitions

- **NSF Behavior:** *Non-Sufficient Funds or overdrafts often related to debit preauthorization 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 provisional funds from a bank or in person deposit before funds are returned to the issuing institution.*
- **Other** *All charge-off reasons that do not fit in the above categories.*

Note: For the purpose of this document, Opening Deposit Chargeback and Bank by Mail/In Person have been grouped in with Other in Other/Misc due to low cell counts.

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

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

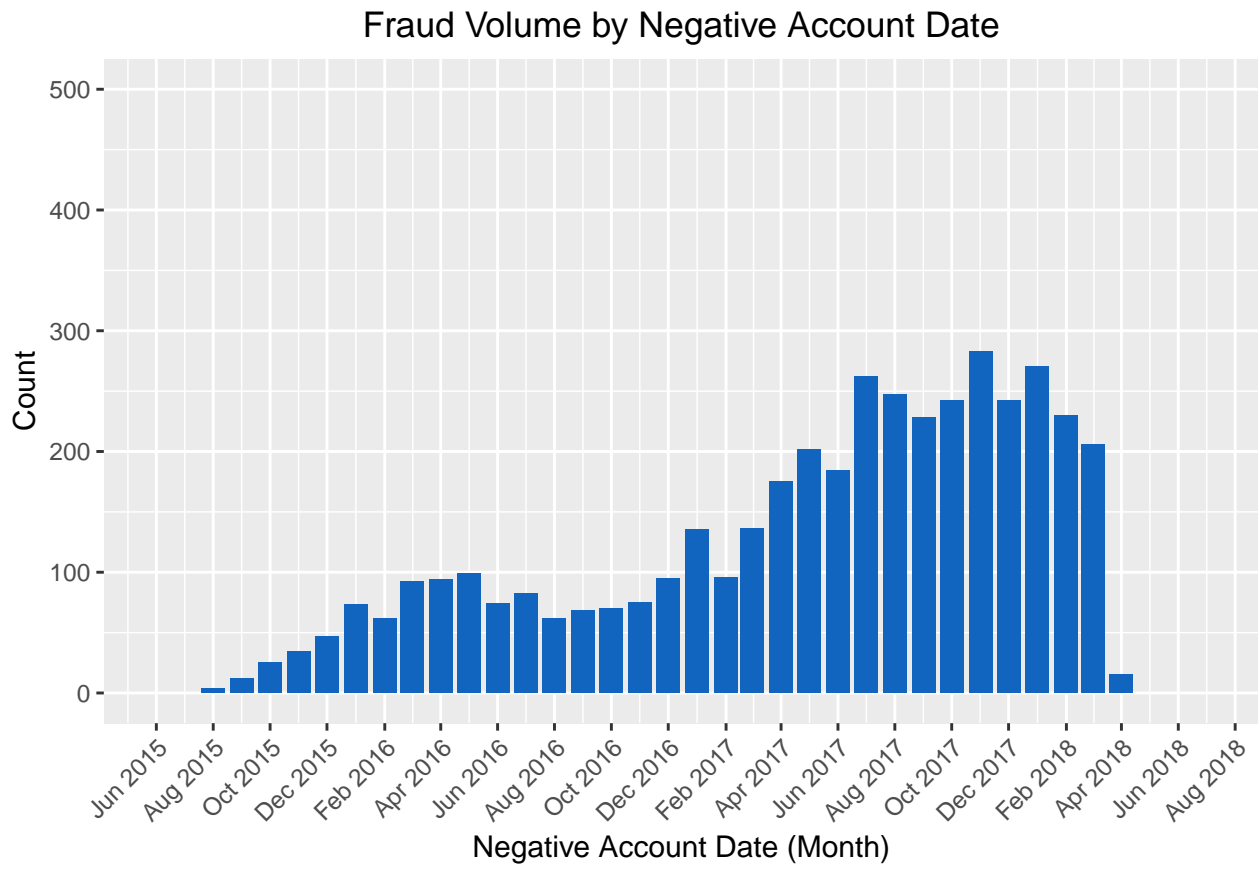
chargeoff_reason	count	rate
NSF Behavior	2494	59.07%
Mobile Dep Chargeback	413	9.78%
Ext Tran Chargeback	1207	28.59%
Other/Misc	19	0.45%
ATM Dep Chargeback	89	2.11%

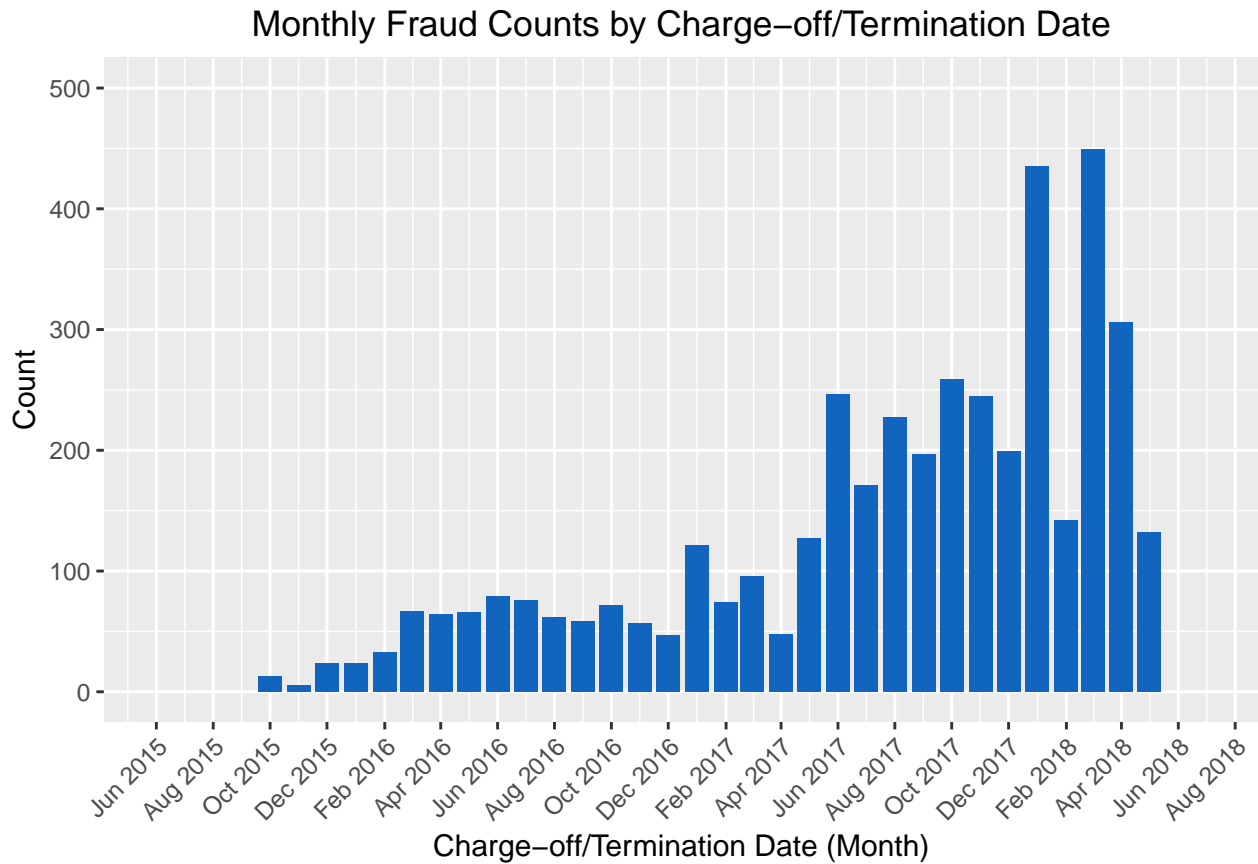
Fraud Volume

Four different methods of determining monthly fraud trends are presented below. They are calculated by the following:

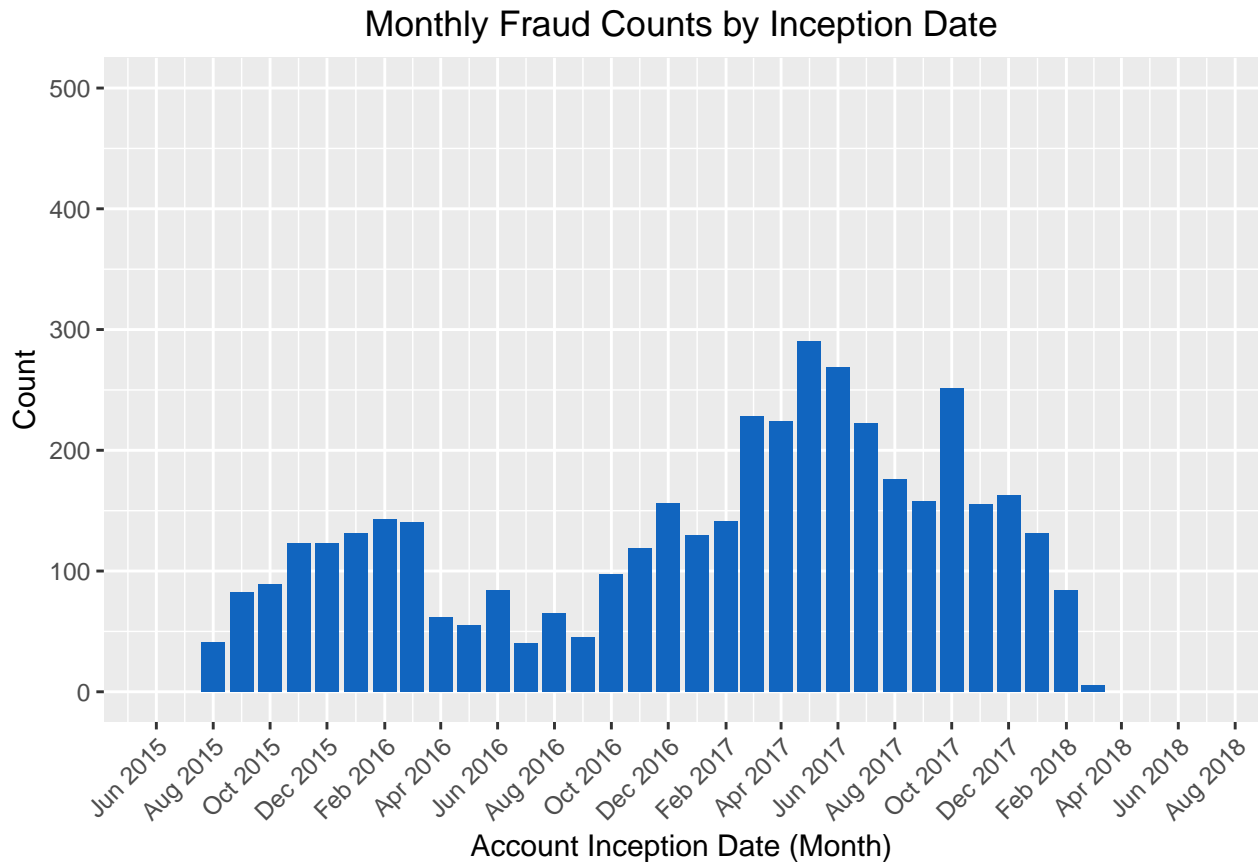
- **Negative Account Date:** The date at which an account becomes negative and stays negative for at least three consecutive days. This is in attempt to determine that date that the fraudulent activity occurred.
- **Charge-off/Termination Date:** The date that an account was charged-off, normally corresponding with an account termination.
- **Inception Date:** The date that the account was created.
- **Fraud Rate:** The proportion of fraudulent accounts over total accounts, calculated by any of the above date definitions.

Monthly Counts



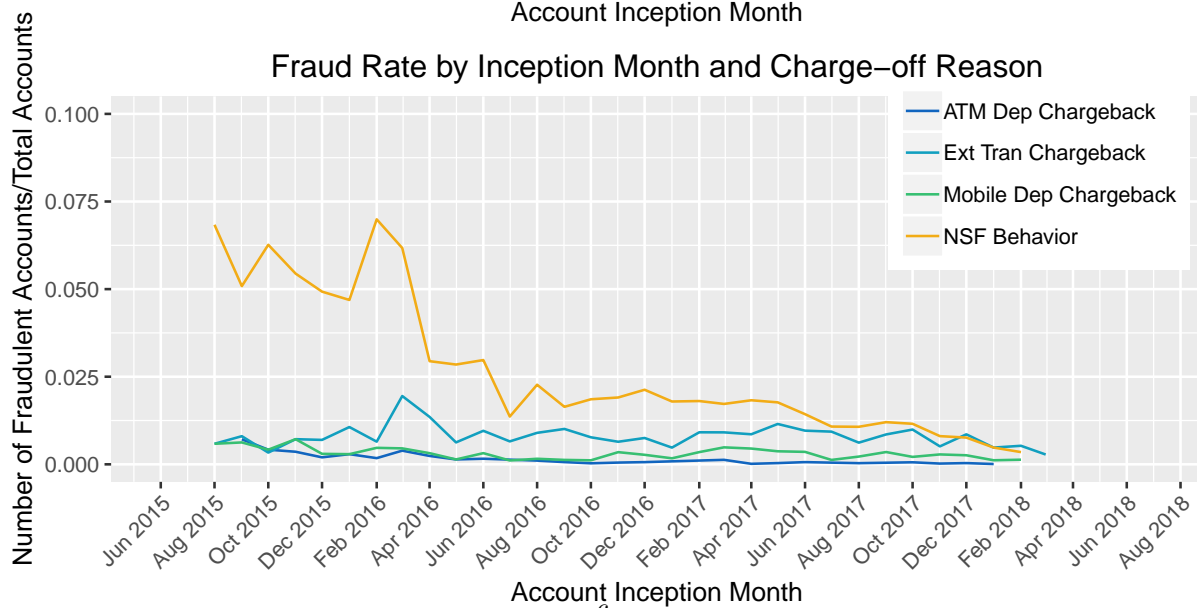
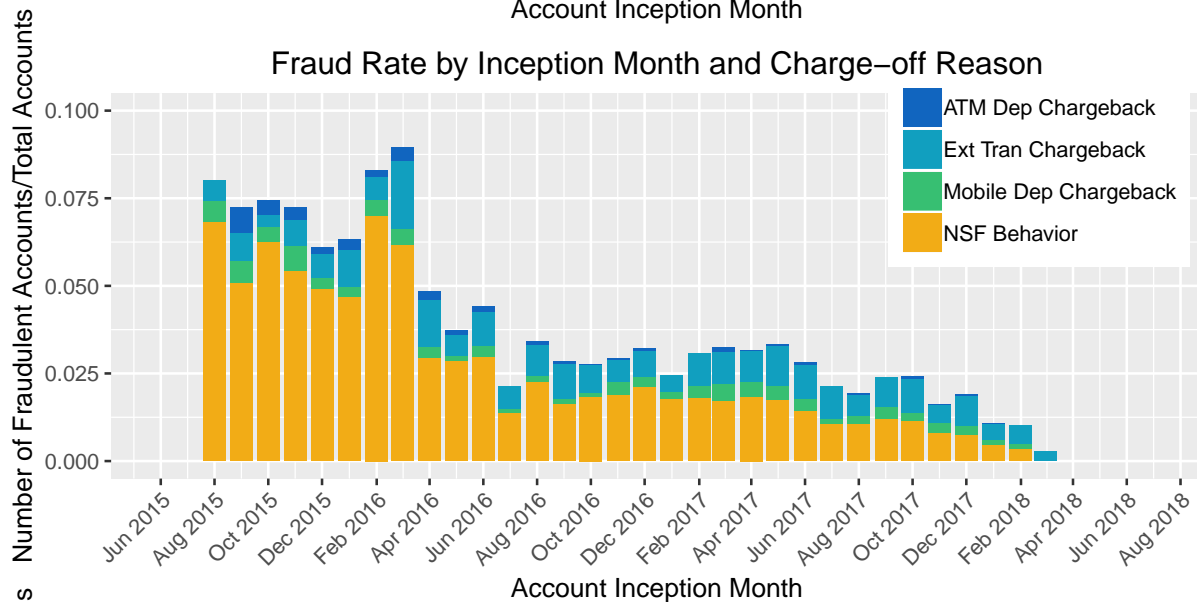
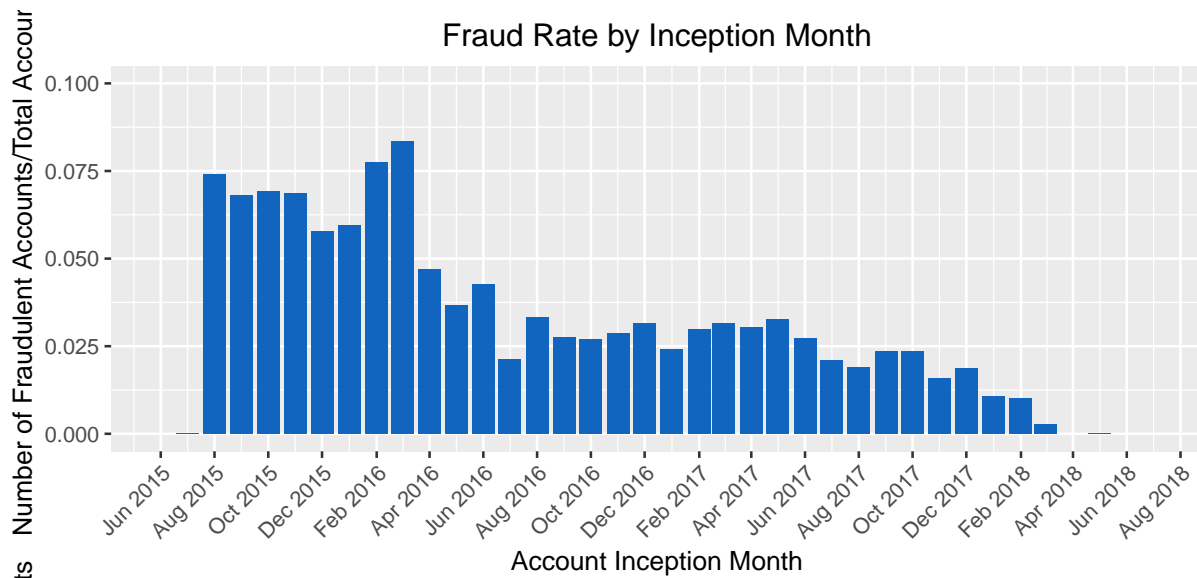


The above graph shows the discrepancy between when an account commits the fraudulent activity versus when the account is actually closed and charged-off, especially noticed by the sporadic spikes in termination date versus negative account date. Previously, much of the account maintenance and fraud mitigation was performed by Radius.



There is a noticeable dip in fraud based on account inception (origination). Further investigation into on-boarding and/or marketing trends may shed further light, however comparing the rate of fraudulent account origination verse total account origination (as in the following graph) indicates that there was actually a more steady decrease in the rate of fraudulent account origination than the previous graph suggests (see figure below).

Monthly Fraud Rate



Aside from February and March of 2016, the rate of fraudulent accounts created has been going down when compared to total account origination. However, most (if not all) of the charge-off/fraud data labeling has been heavily relied on by Radius. Spot checking a number of account who went negative and never became positive has indicated that there may be upwards to an additional 30% of accounts that should have been charged-off or labeled as such, but currently have not been. This is likely due to the processes and procedures dictated by Radius.

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
Other/Misc	\$27,176.78	1.66%	\$1,430.36	\$0.03	\$106.20	\$174.77	\$1,627.70	\$8,733.81
ATM Dep Chargeback	\$122,241.57	7.48%	\$1,373.50	\$17.04	\$313.54	\$710.09	\$1,343.58	\$12,021.23

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

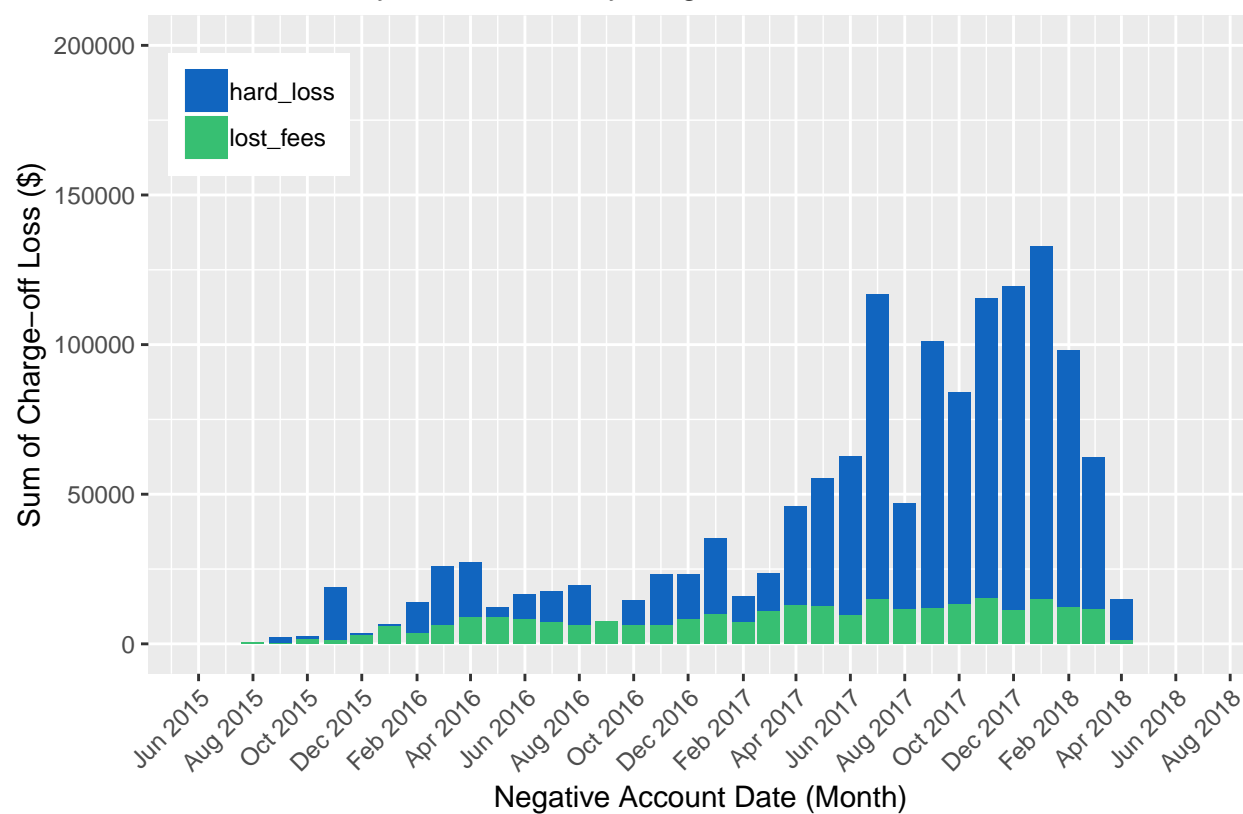
chargeoff_reason	sum	percent	mean	min	q25	q50	q75	max
NSF Behavior	\$110,526.34	8.09%	\$44.32	\$0.00	\$0.85	\$15.25	\$33.58	\$8,075.58
Mobile Dep Chargeback	\$366,152.04	26.81%	\$886.57	\$0.00	\$103.47	\$481.14	\$1,249.51	\$7,903.67
Ext Tran Chargeback	\$748,748.50	54.82%	\$620.34	\$0.00	\$25.09	\$120.06	\$465.01	\$45,695.79
Other/Misc	\$25,940.21	1.90%	\$1,365.27	\$0.00	\$14.49	\$146.77	\$1,542.70	\$8,733.81
ATM Dep Chargeback	\$114,470.69	8.38%	\$1,286.19	\$0.00	\$223.54	\$620.09	\$1,272.10	\$11,931.23

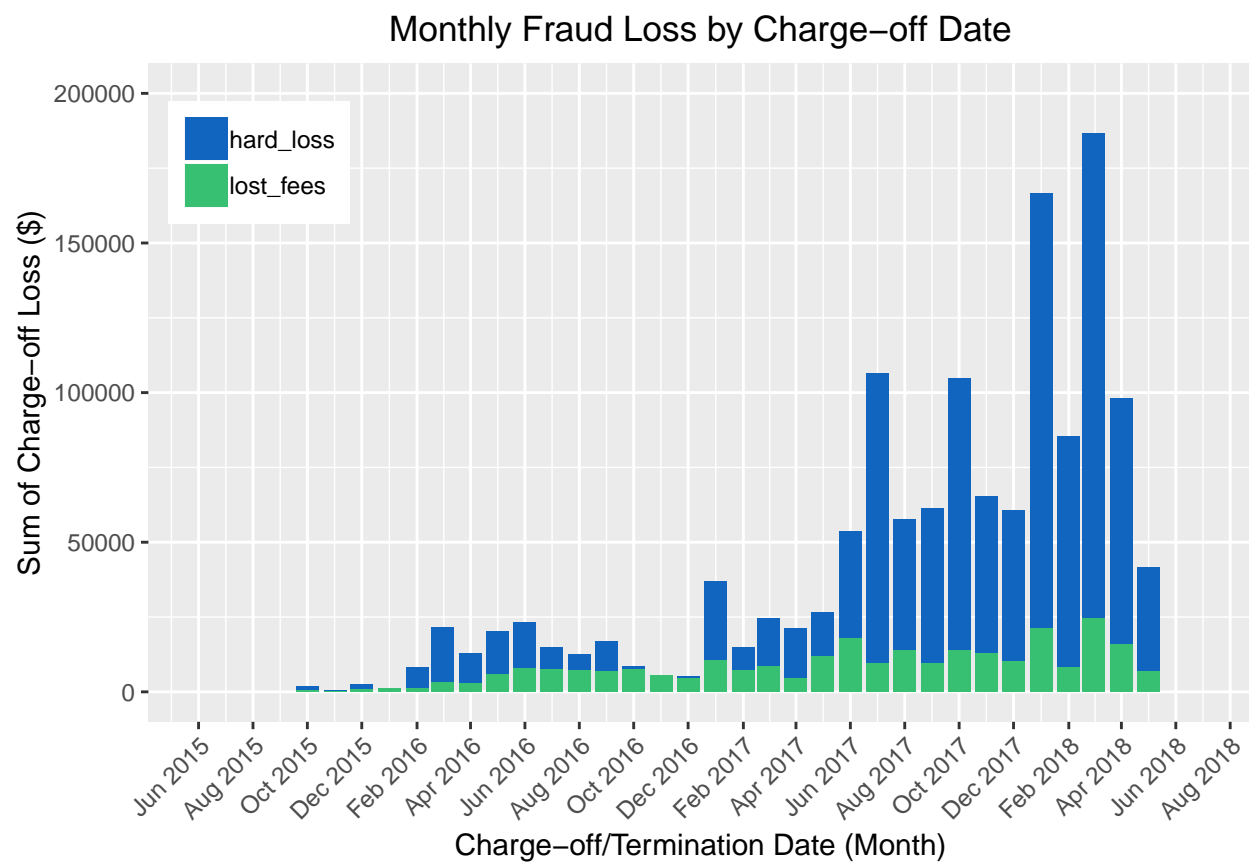
As displayed in the tables above, much of the total losses related to NSF Behavior are not actually realized. While NSF Behavior makes up the majority of volume of charge-offs, it is in fact, one of the least costly, especially when compared to the cost of some of the other charge-off reasons as shown below.

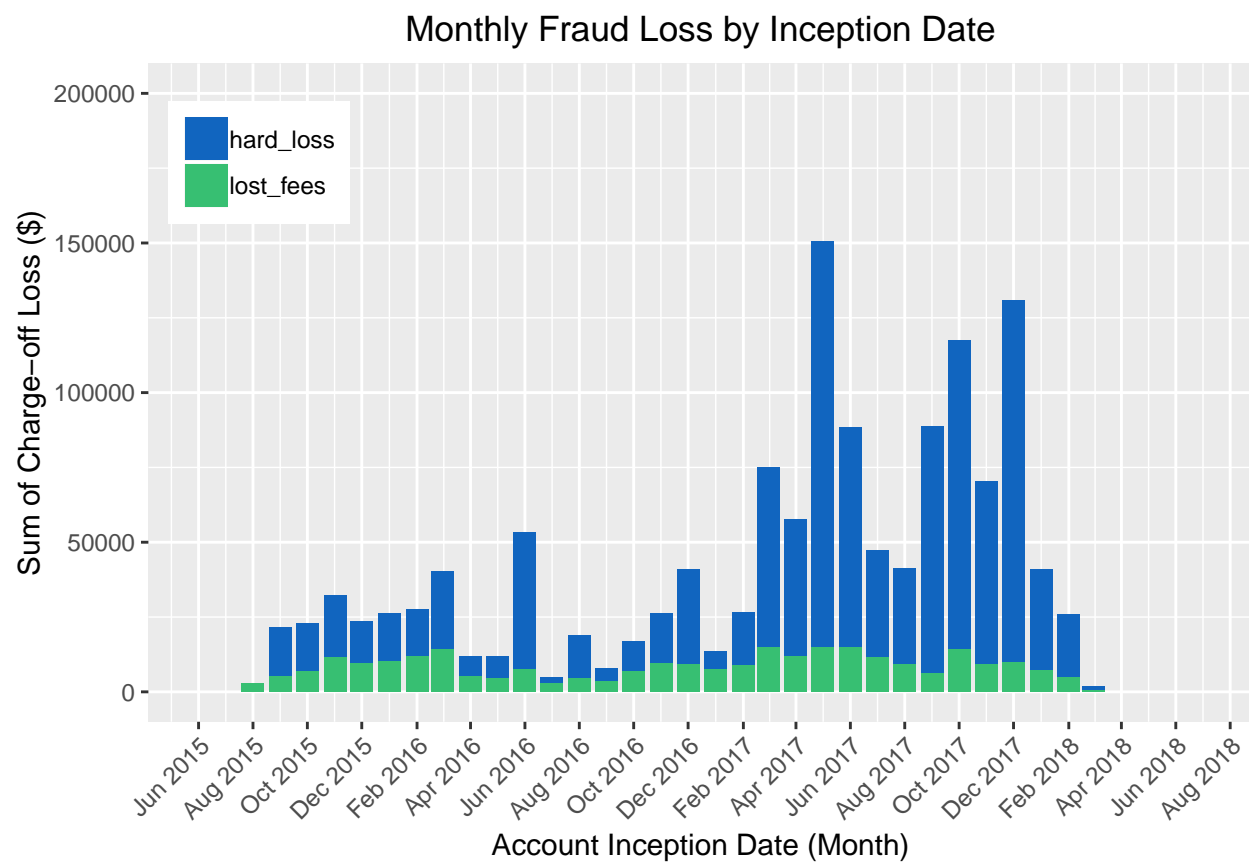
Monthly Cost

- **Total Loss:** The total amount charge-off amount for an account (hard loss + lost fees).
- **Hard Loss:** The actual amount of money lost from a charge-off (realized loss).
- **Lost Fees:** Fees that Aspiration was owed, but never collected (Deposited Item Reversal Fees, Check Reversal Fees, Over Draft Fees, etc.).

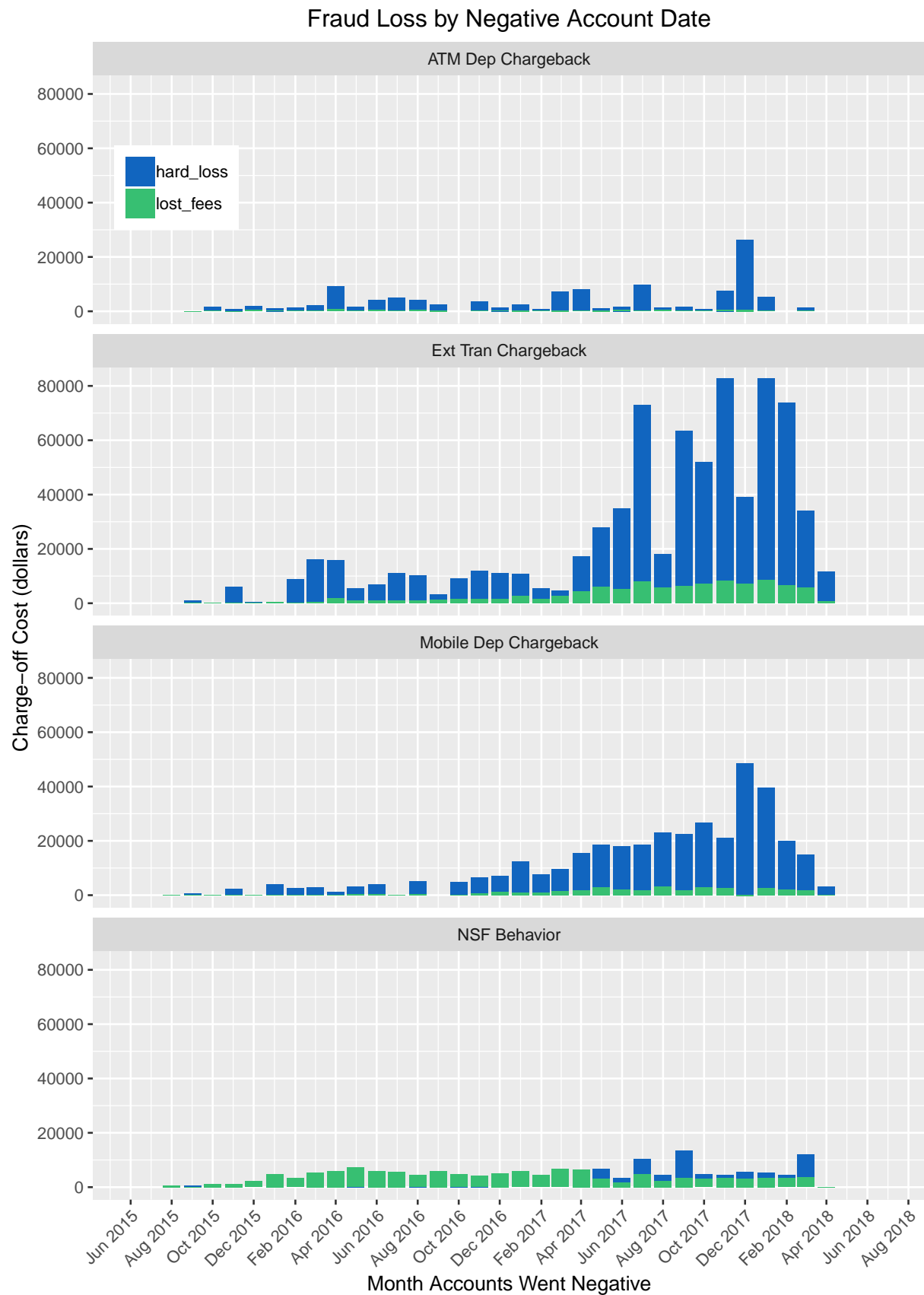
Monthly Fraud Loss by Negative Account Balance Date







Cost by charge-off type

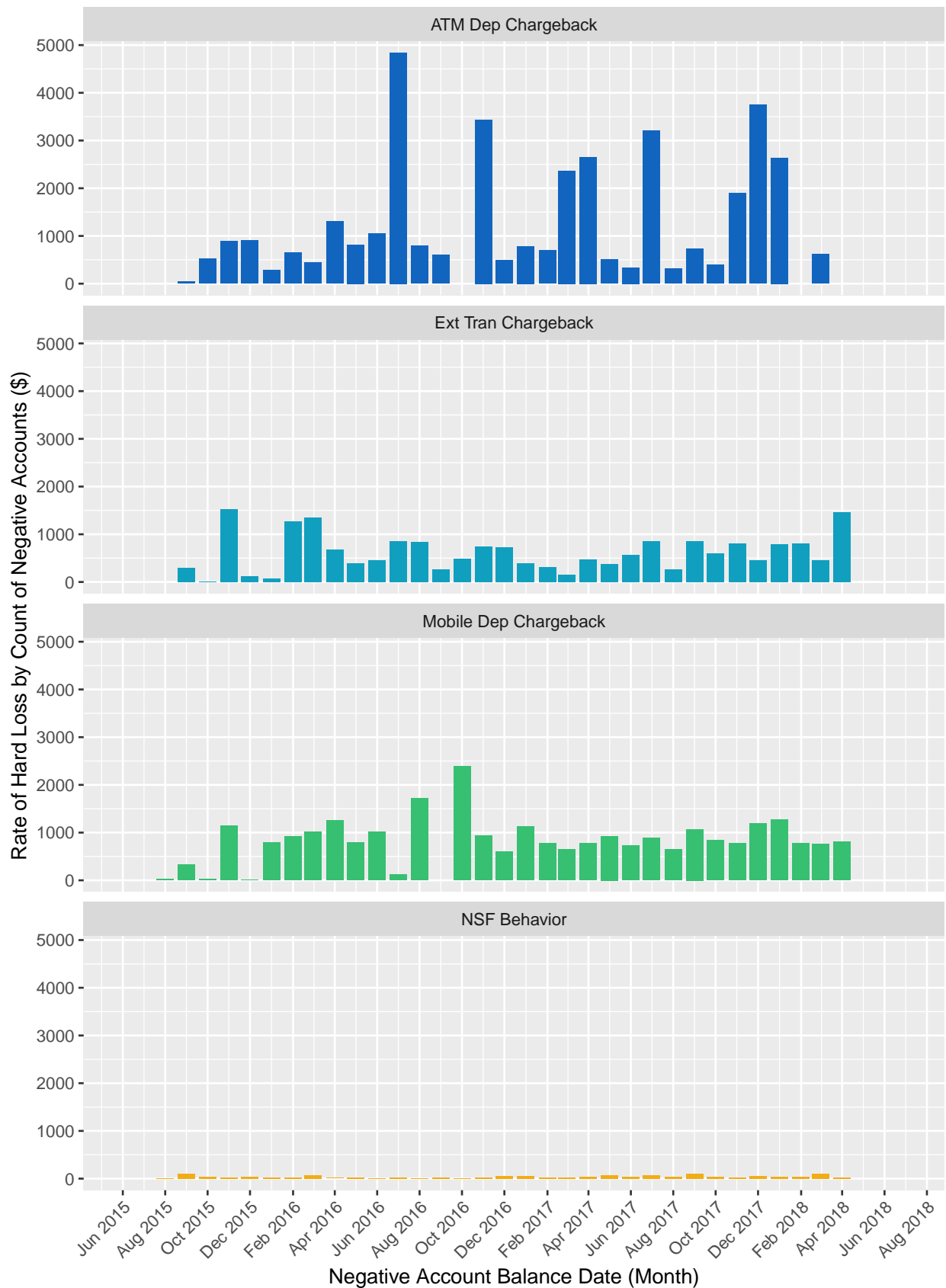


Note: Bank by Mail / In Person, Opening Dep Chargeback, and Other charge-off reasons remove due to low counts.

Account Severity

Similar to the Fraud Rate, Account Severity refers to rate of hard loss by number of accounts per the given charge-off reason.

Account Severity (Hard Loss)



Time Till Fraud/Charge-off Act

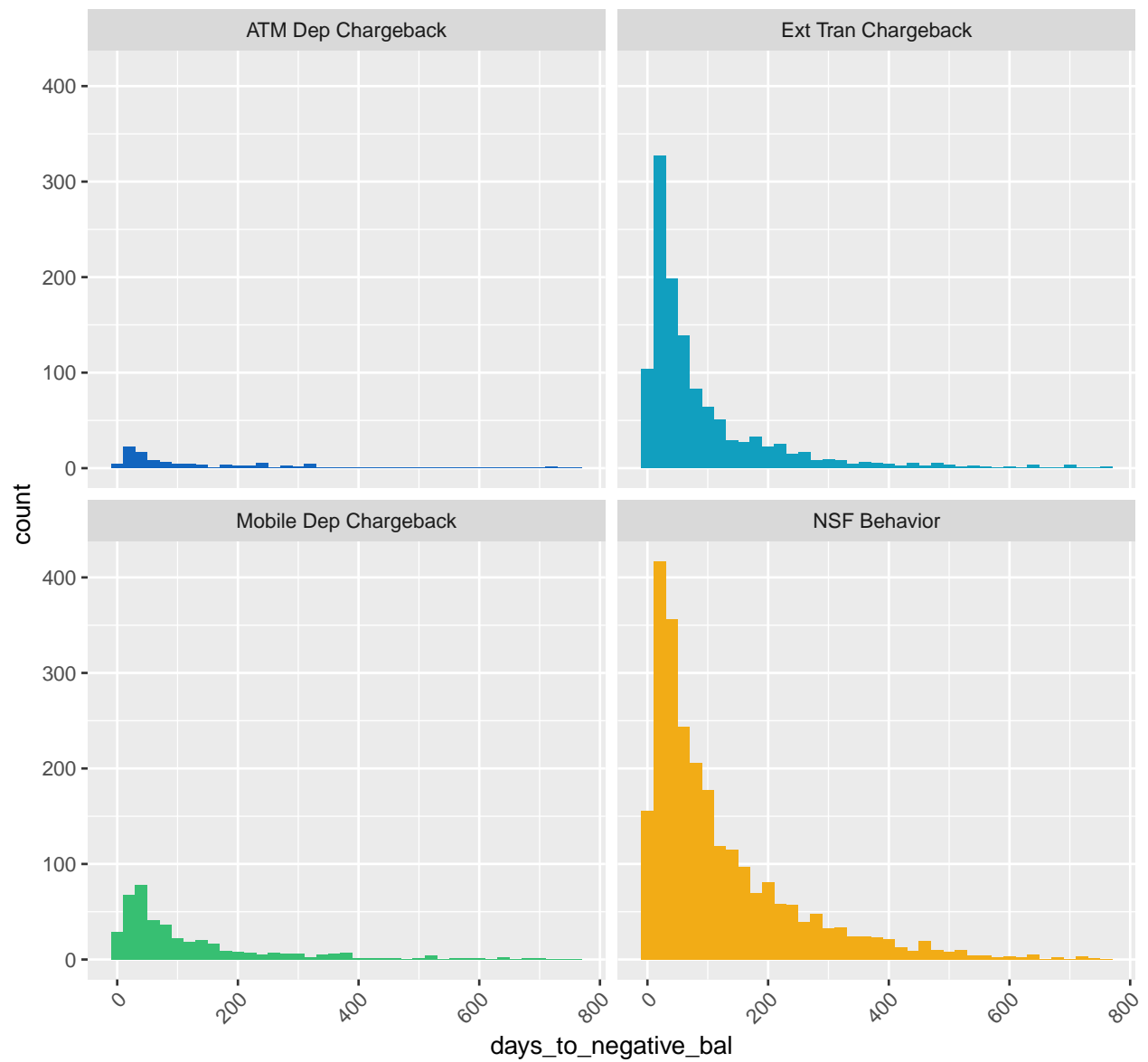
Time between account creation and first day an account goes negative and stays negative for at least three days in a row.

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
Other/Misc	56.42	0	7.50	38.00	82.50	257
ATM Dep Chargeback	101.06	5	25.75	52.50	140.25	712

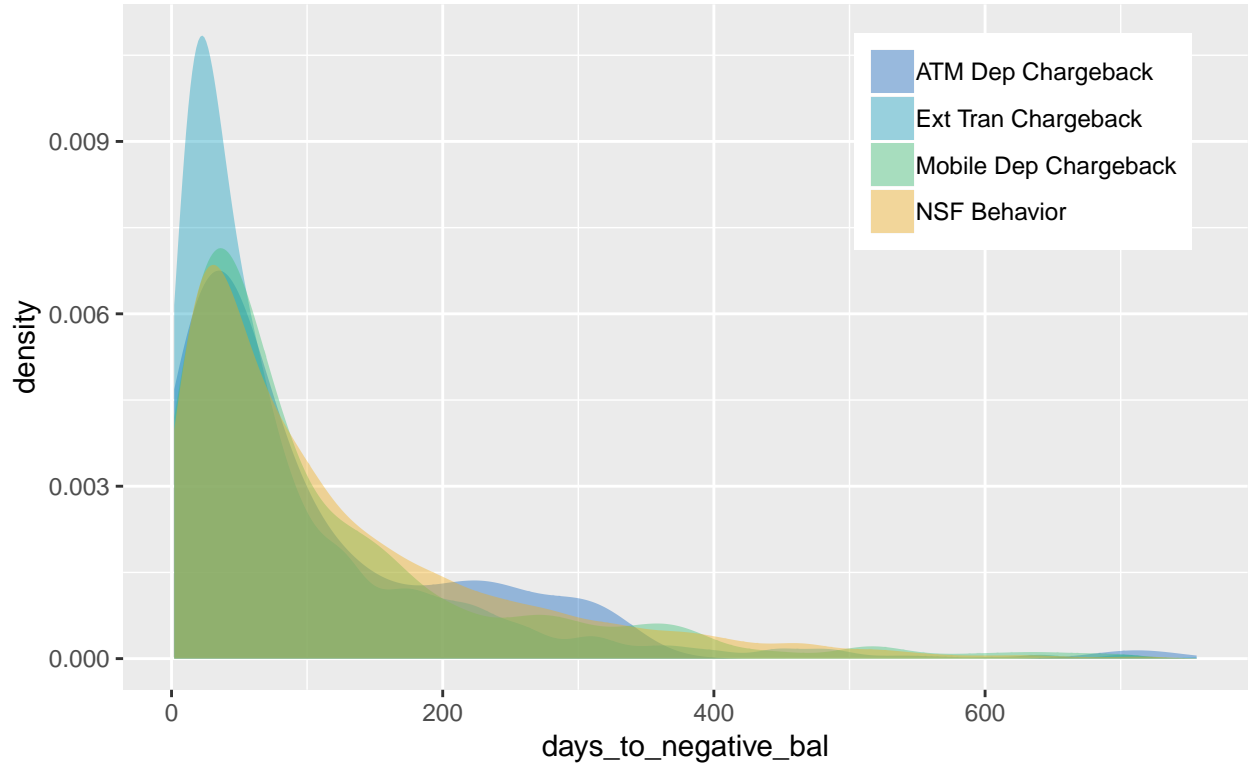
Both External Transfer Charge-backs and NSF Behavior has the larges volume as well as more accounts that have a longer time till committing fraud, though it seems mostly proportional to the other types of charge-backs, mostly differing in volume.

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 Channels and Fraud

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

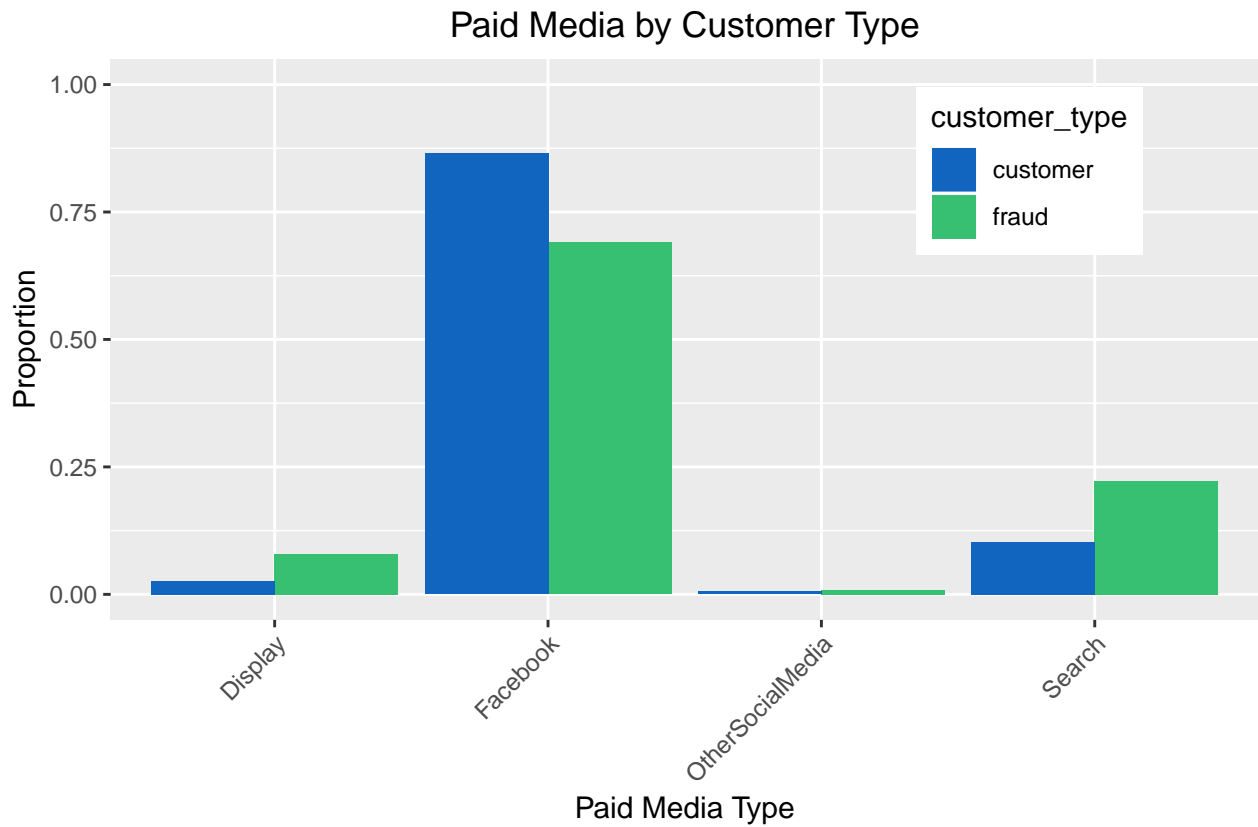
	Affiliates	Organic	PaidMedia	Unattributed
customer	74362 (48.64%)	353 (0.23%)	24619 (16.1%)	53542 (35.02%)
fraud/charge-off	1340 (31.75%)	9 (0.21%)	1421 (33.67%)	1451 (34.38%)

Table 7: Marketing channel group by charged-off reason

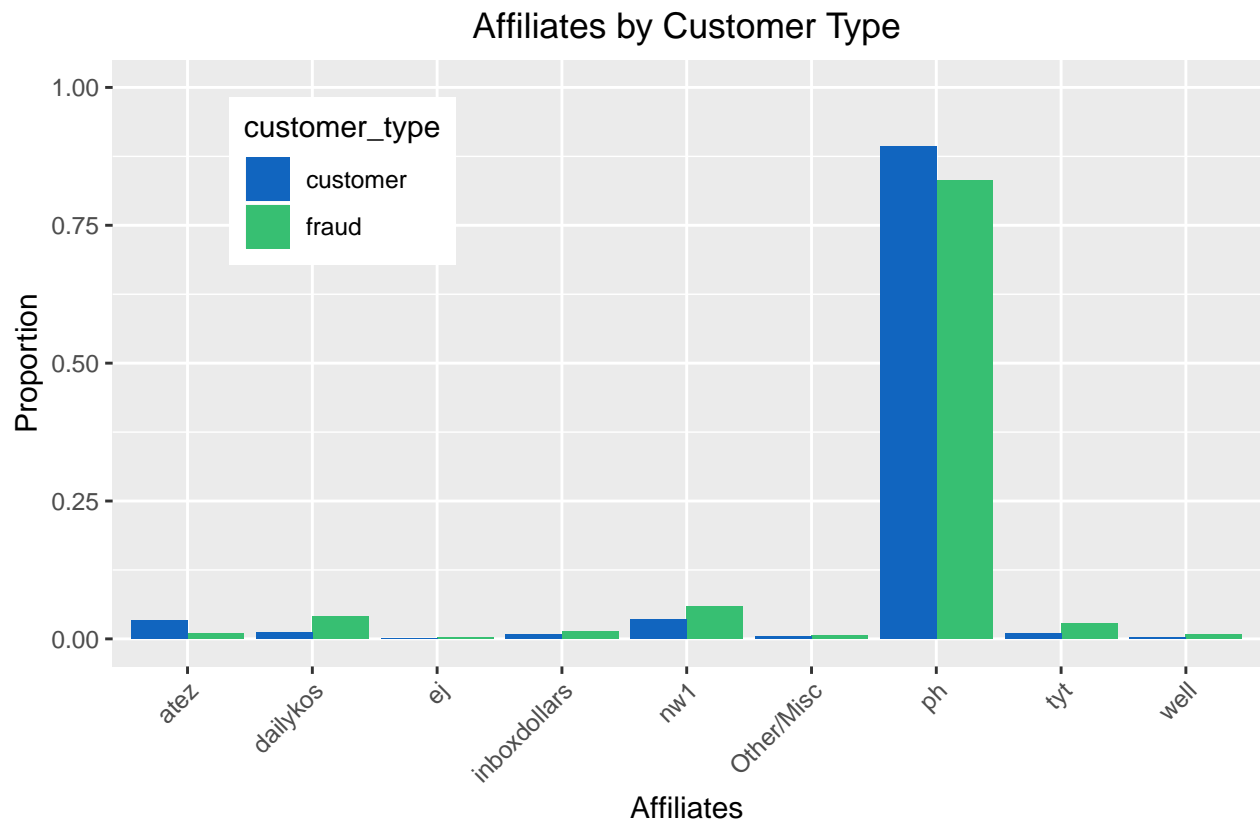
	Affiliates	Organic	PaidMedia	Unattributed
ATM Dep Chargeback	13 (14.61%)	1 (1.12%)	46 (51.69%)	29 (32.58%)
Ext Tran Chargeback	420 (34.8%)	2 (0.17%)	313 (25.93%)	472 (39.11%)
Mobile Dep Chargeback	110 (26.63%)	0 (0%)	143 (34.62%)	160 (38.74%)
NSF Behavior	797 (31.97%)	6 (0.24%)	908 (36.42%)	782 (31.37%)
Other/Misc	0 (0%)	0 (0%)	11 (57.89%)	8 (42.11%)

Both Paid Media and Affiliates seem to have a higher rate among fraudulent and charged-off accounts.

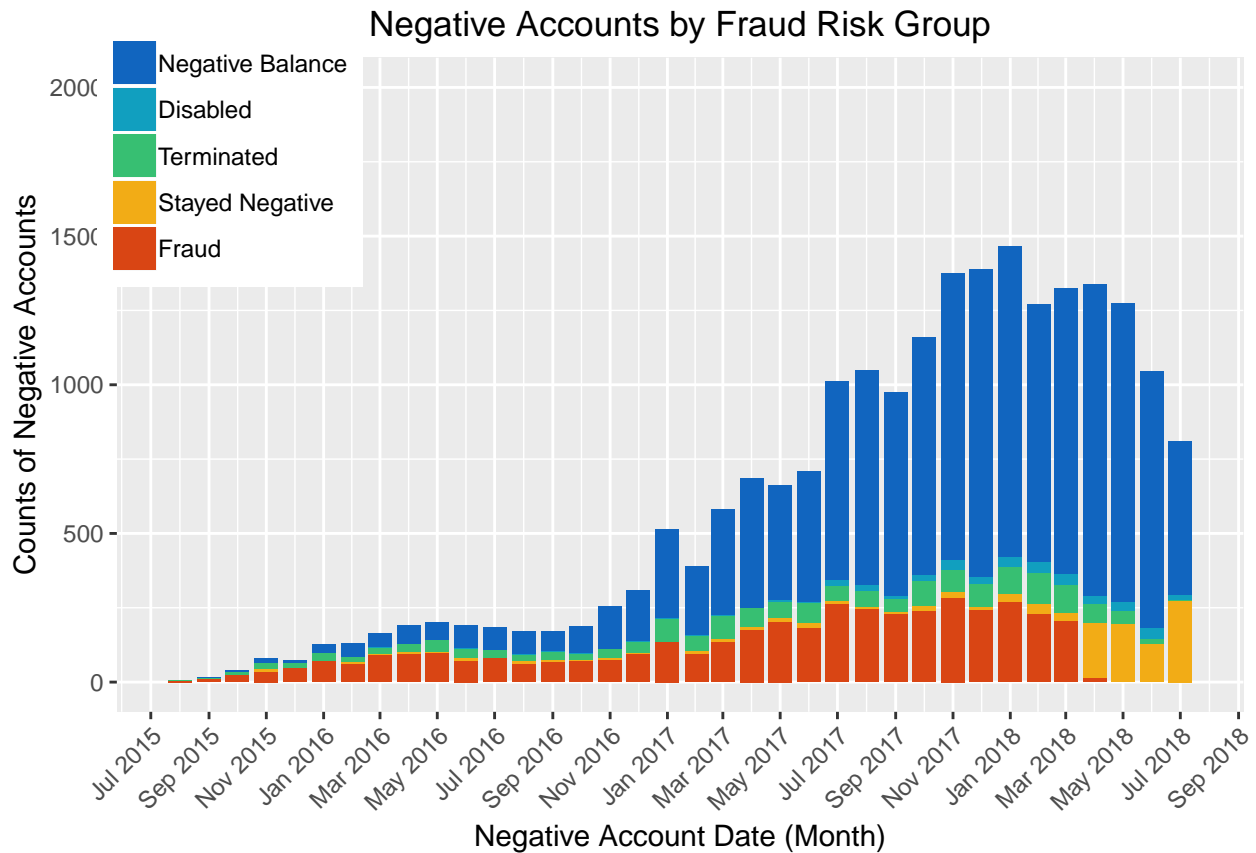
Paid Media



The proportion of Search results is higher among fraudulent/charged-off accounts, which could be a result of users seeking out institutions like Aspiration for more malicious behaviors.



Negative Account Balance Data by Fraud/Risk Group



```
# Top states
df_states <- df[, .N, by = .(state, is_charged_off)]
df_states[, total := sum(N), by = is_charged_off][, rate := round(N/total, 4), by = is_charged_off]
df_states[order(-rate)][1:10]
```

##	state	is_charged_off	N	total	rate
## 1:	California	FALSE	24569	152959	0.1606
## 2:	California	TRUE	649	4222	0.1537
## 3:	Texas	TRUE	400	4222	0.0947
## 4:	Texas	FALSE	11822	152959	0.0773
## 5:	Florida	TRUE	322	4222	0.0763
## 6:	New York	FALSE	11165	152959	0.0730
## 7:	Florida	FALSE	10594	152959	0.0693
## 8:	New York	TRUE	209	4222	0.0495
## 9:	Georgia	TRUE	192	4222	0.0455
## 10:	Illinois	FALSE	5881	152959	0.0384

```
# Customer age
df[, .(mean_customer_age = round(mean(customer_age), 2),
      median_customer_age = median(customer_age)), by = .(is_charged_off)]
```

##	is_charged_off	mean_customer_age	median_customer_age
## 1:	FALSE	NA	NA
## 2:	TRUE	35.87	34

```

# days to first summit debit card use
df[, .(mean_days_till_first_debit = round(mean(days_to_first_summit_debit_card_use, na.rm= TRUE),2),
      median_days_till_first_debit = median(days_to_first_summit_debit_card_use, na.rm=TRUE)), by = .(is_charged_off)]

##      is_charged_off mean_days_till_first_debit median_days_till_first_debit
## 1:      FALSE      85.26      41
## 2:      TRUE      37.37      19

# Inital funding
df[, .(mean_initial_funding_amount = round(mean(first_account_funding_amount, na.rm= TRUE),2),
      median_initial_funding_amount = median(first_account_funding_amount, na.rm=TRUE)), by = .(is_charged_off)]

##      is_charged_off mean_initial_funding_amount
## 1:      FALSE      794.14
## 2:      TRUE      146.43
##      median_initial_funding_amount
## 1:      20
## 2:      10

```

Recomentations

Data Quality

Thus far, much, if not all, of the fraud reporting and labeling has been dictated by Radius. Numerous accounts have been found that display fraudulent behaviors, but have not been reported as fraudulent by Radius to Aspiration. Review of such accounts as well as the continued work in the Ops department to take control over the process and procedures around mitigation of risk and fraud, including labeling, will improve the ability to monitor and model fraud.

Furthermore, Radius has not passed on information regarding 1st or 3rd party fraud verses synthetic fraud. Since these groups tend to have very different qualities, it would be wise to differentiate between them going forward, so as best to target the different types of fraud while optimizing the customer experience for good customers.

Time to Account Termination

The average time between an account going negative and it being charged-off ranged between 56 and 120 days. While radius has a standard of charging off accounts after 30 days of being over drafts, the data shows that they have not always stuck to their deadline. Shortening the time to account termination for fraudulent accounts could prevent further loss.

Data Breaches and Scripted Attacks on the Easiest Targets

From anecdotal experience, some FinTechs saw a large spike in scripted 3rd Party ID Theft following the data breaches in May 2016 from Myspace and LinkedIn.

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): Data seems to be in web_db.user_browser_fingerprint table, but doesn't list actual browser names, just coded variables. Waiting for hear back on what the dummy labels correspond to.
- 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:~~ This turns out to be harder said than done, but working on it
- ~~Look at chargeoff types that have min of \$0~~ This is due mostly to fees charged for OD accounts. Should we include in report?
- ~~Remove low count of charge off groups~~
- ~~Update fraud count graphs to be on same x and y axis~~
- ~~Next steps section to include difference between negative account date and charge off date as something~~
- ~~Look at percentage of accounts that went negative but werent charged off as a comparision/validation~~
 -In process, confirming new data - Figure on page 20 shows much larger amount of negative accounts that were not charged off. Would it be worth while looking at actual transaction patterns for fraud? i.e., accounts that deposited a check and then spent the money or withdrew it from an ATM and then had the check return?
- ~~Show graph of % of fraud inception~~
- If I were to make a recommendation what would it be? Filter all the charts like that
- ~~Collapse into 5 grouops, call it other/Mise~~
- ~~trend of fraud by inception month/overall fraud by, what is industry~~ Industry has been reported as ~10%, though some FinTechs have been reportedly closer to 35%. Current Aspiration rates seem very low, perhaps due to not actively on-boarding? Or failure to correctly label all fraud due to Radius?
- ~~2 fraud rates: fraud by counts and fraud by dollars~~
- ~~Note about no way to distinguish 1st vs 3 vs syntheistic fraud~~
- ~~Account severity: loss of 18K in External transfer and it was 10 people to see if it really increasing or~~
- ~~Add graph of counts by inception date~~
- Why do some people wait so long? After 200 days - compare customers who regularly engage account who then commit fraud. Is it sleeper accounts? This needs more evaluation using a better metric
- ~~call out data mislabeling~~
- ~~Look at affiliate fraud for marketing like done for paid media~~

- ~~Move script and all coding to the end for an appendix~~

Fraud Section, Cost Section,