# Title: Sales cycle

Objective:

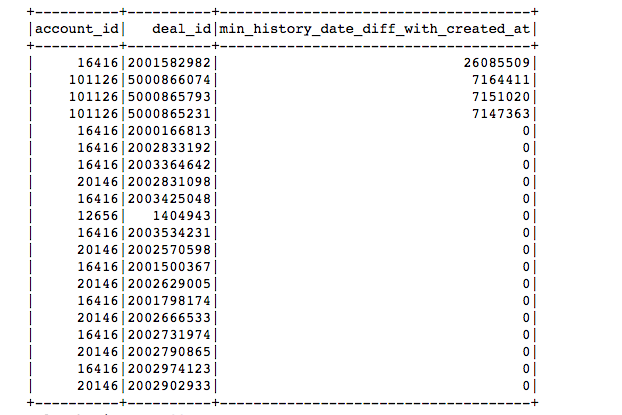
1. Calculate length of sales from open to close\_win/close\_lost
2. Plot distribution of cycle length
3. Calculate length from custom stages
4. Plot distribution from custom stages

[Bharathi Balasubramaniam](https://confluence.freshworks.com/display/~bharathi.balasubramaniam)

Start date - 5/9/19

Observations:

* Created\_at and updated at in stage\_histories is exactly the same for all entries
* account\_id = 66489 has multiple entries(~200) in stage\_histories for a particular deal, stage combination. This means that a deal comes back to the same stage multiple times. We could ignore this account for data exploration.
* 'first\_assigned\_at' column in 'Deals' table represent the timestamp when an owner is assigned to a particular deal. So, this would be the start of the sales cylce and not 'created\_at' which is deal creation timestamp
* TODO: Later, we should consider the first time an email was sent/call was made regarding a deal as the start of the sales\_cycle
* What do we do with deals with a very small sales\_cycle Eg: 1,2 days?
* Min date for a particular deal in stage\_histories table is same as created\_at in deals table in most cases.



* There are cases where closed\_date in deals table is greater than today's date(6th sep 2019)

+----------+----------+--------------+-----------+-------------------+-------------------+  
|account\_id| deal\_id|expected\_close|closed\_date| created\_at| first\_assigned\_at|  
+----------+----------+--------------+-----------+-------------------+-------------------+  
| 45578|2003318383| null| 2019-10-07|2019-05-20 21:11:22|2019-05-20 21:11:23|  
| 16416|2003187926| 2019-06-20| 2019-09-21|2019-03-21 10:26:23|2019-03-21 10:26:24|  
| 118761|5000687579| 2018-11-30| 2019-11-29|2019-03-13 01:31:28|2019-03-13 01:31:29|  
| 118761|5000940734| 2020-04-10| 2020-04-01|2019-04-10 23:04:32|2019-04-10 23:04:33|  
| 118761|5001003787| 2019-11-01| 2019-11-01|2018-11-01 00:00:00| null|  
| 118761|5001003788| 2019-11-01| 2019-11-01|2018-11-01 00:00:00| null|  
| 118761|5001003789| 2019-11-01| 2019-11-01|2018-11-01 00:00:00| null|  
| 118761|5001003790| 2019-11-01| 2019-11-01|2018-11-01 00:00:00| null|  
| 118761|5001003791| 2019-11-01| 2019-11-01|2018-11-01 00:00:00| null|  
| 118761|5001003792| 2019-11-01| 2019-11-01|2018-11-01 00:00:00| null|  
| 118761|5001003793| 2019-11-01| 2019-11-01|2018-11-01 00:00:00| null|  
| 118761|5001003794| 2019-11-01| 2019-11-01|2018-11-01 00:00:00| null|  
| 118761|5001003795| 2019-11-01| 2019-11-01|2018-11-01 00:00:00| null|  
| 118761|5001003796| 2019-11-01| 2019-11-01|2018-11-01 00:00:00| null|  
| 118761|5001003797| 2019-11-01| 2019-11-01|2018-11-01 00:00:00|2019-05-10 06:29:40|  
| 118761|5001003798| 2019-11-01| 2019-11-01|2018-11-01 00:00:00| null|  
| 118761|5001003799| 2019-11-01| 2019-11-01|2018-11-01 00:00:00| null|  
| 118761|5001003800| 2019-11-01| 2019-11-01|2018-11-01 00:00:00| null|  
| 118761|5001003801| 2019-11-01| 2019-11-01|2018-11-01 00:00:00| null|  
| 118761|5001003802| 2019-11-01| 2019-11-01|2018-11-01 00:00:00| null|  
+----------+----------+--------------+-----------+-------------------+-------------------+

print(deals.filter((col("closed\_date")>(lit("2019-09-06")))).count())

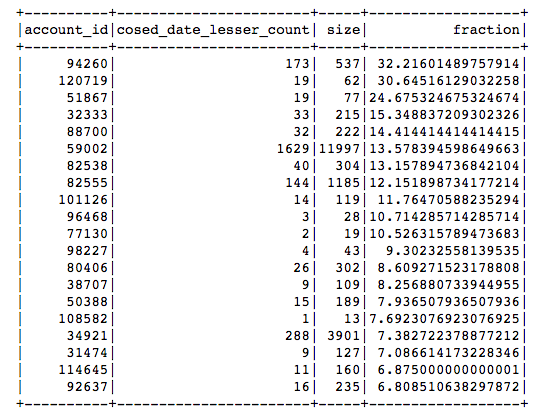
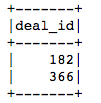
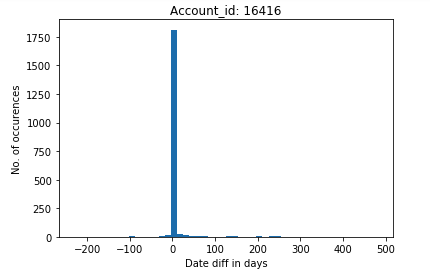
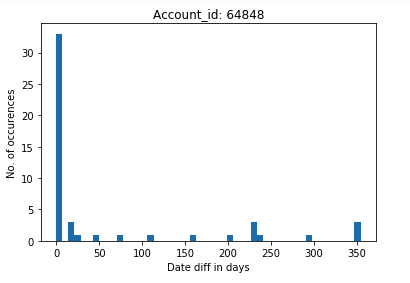
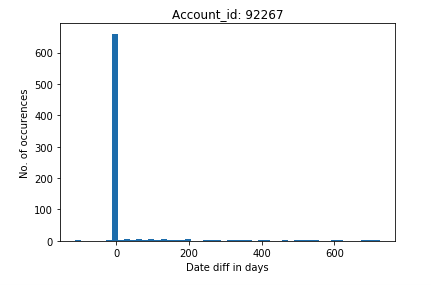
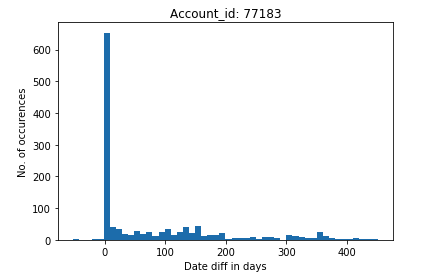
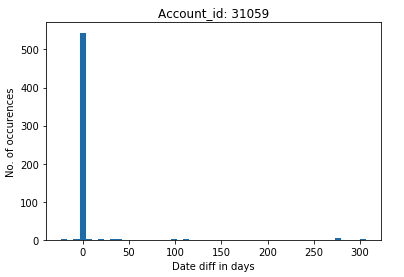
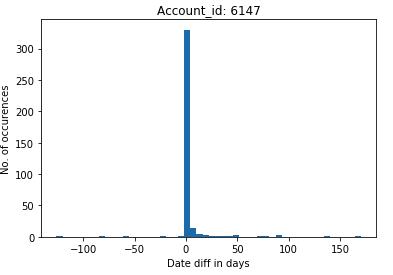
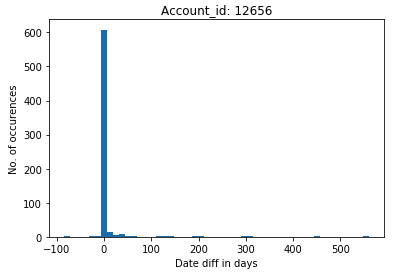
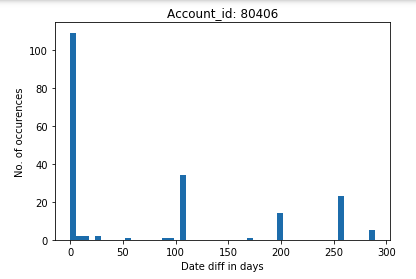
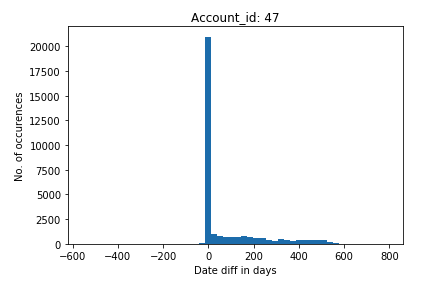
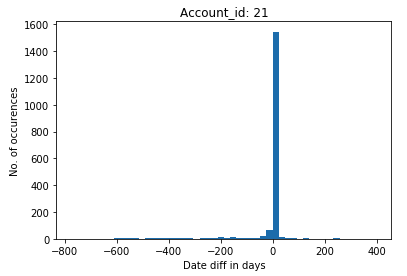
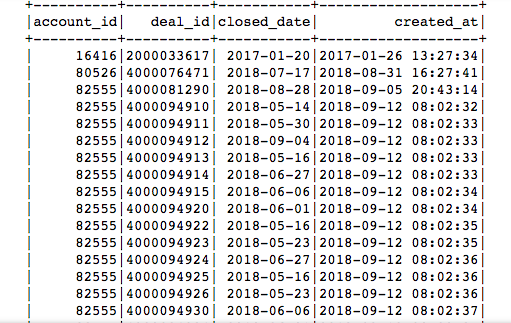
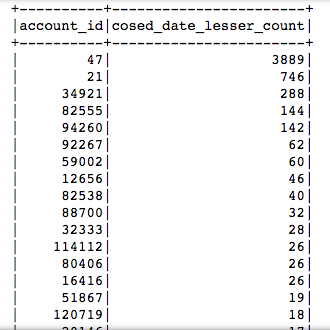
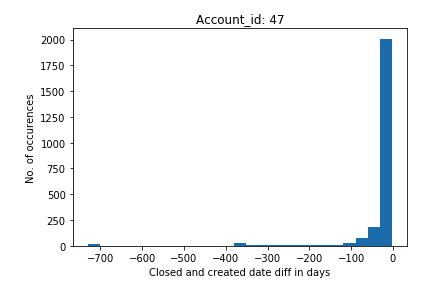
114

deals.filter((col("closed\_date")>(lit("2019-09-06")))).groupby('account\_id').count().sort(desc('count')).show()

+----------+-----+  
|account\_id|count|  
+----------+-----+  
| 118761| 100|  
| 94260| 2|  
| 30456| 2|  
| 92637| 2|  
| 45578| 2|  
| 70978| 1|  
| 38707| 1|  
| 88700| 1|  
| 51151| 1|  
| 92267| 1|  
| 16416| 1|  
+----------+-----+  
deals.groupby('account\_id').count().filter(col('account\_id')==118761).show()

+----------+-----+  
|account\_id|count|  
+----------+-----+  
| 118761| 534|  
+----------+-----+  
100/534 are incorrect, so removing this account from our analysis  
  
There are cases where created\_at in deals table is greater than closed date. These deals might have been closed before they migrated to Freshsales crm.  
This data is not useful for us.

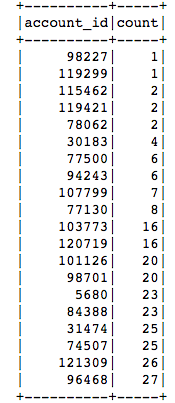
print(deals.filter((col("closed\_date")<(col("created\_date")))).count())

2859  
Accounts with have more records of closed\_date < created\_at  
  
Deals present in deals table but not in stage\_histories table  
  
Histogram of (max date in stage\_histories - closed date)  
Abnormal accounts:  
  
  
  
  
  
After removing deals which have created\_at or closed\_date older than 2017, we still have deals(~5829) where closed\_date is older than created\_at  
  
  
~1485 of the above deals of account 47 are just 1 day apart(between created and closed\_date)  
Distribution of the remaining ~2404 deals for account 47. TODO: These deals have to be analysed  
  
General filters on deals data:  
1. deal is not deleted  
2. deals.filter((col('created\_date')> (lit("2017-01-01"))))  
3. Deals which are created after a quarter of onboarding of the account  
  
Sales cycle specific filters:  
1. deal closed\_date is not null (deal is closed)  
2. deals.filter((col('closed\_date')> (lit("2017-01-01"))))  
2. deals.filter((col("closed\_date")>(col("created\_date")))& (col("closed\_date")<(lit("2019-09-07"))))  
3. deal amount is greater than 0

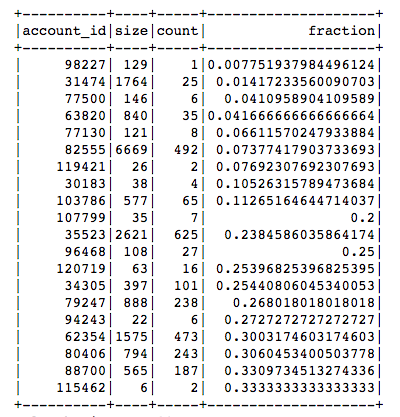
(On top of the above filters when a filter to fetch only deals with amount > 0 is added, deal count decreases from 93506 to 82957. So it is fine to add this filter to the list of sales\_cycle filters.

On top of the above filters when a filter to fetch only deals which are created after a quarter of onboarding is added, deal count decreases from 82956 to 73170. So it is fine to add this filter to the list of general filters.)

Count of closed\_date not null deals(after applying general filters):



Fraction of closed\_date is not null deals:

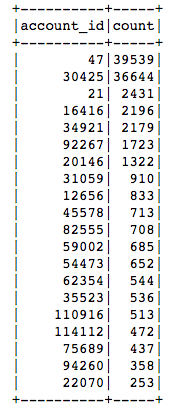


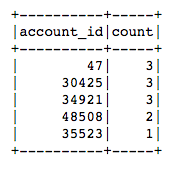
Following accounts should be removed because doesn't have either 5% closed deals or 20 closed deals (after applying general filters):

98227, 119299, 115462, 119421, 78062, 30183, 77500, 94243, 107799, 77130, 103773, 120719, 101126, 98701, 31474, 63820

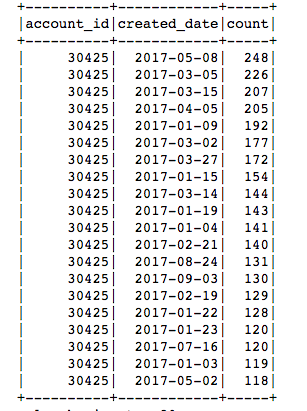
Conditions to decide if we can enable Revenue forecasting for a particular account(after applying general filters):

* Should have 20% closed deals or 20 closed deals
* x deals or y% of deals closed after 2017
* x deals or y% of deals with closed\_date > created\_date
* x deals or y% of deals with closed\_date < today
* x deals or y% of deals with deal amount > 0
* x deals or y% of deals with first\_assigned\_at not null
* x deals or y% of deals with a single stage visited not more than z times

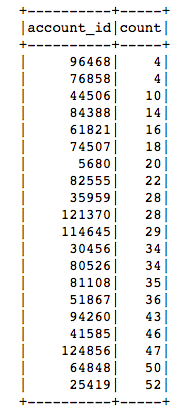
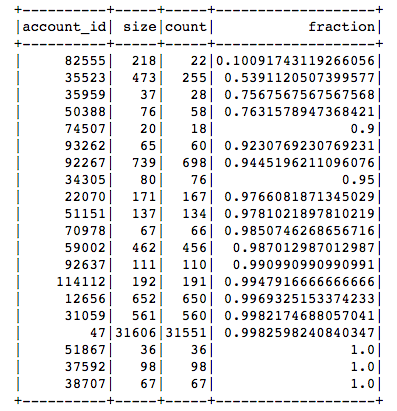
All further analysis is after these filters  
Account size after the filters  
  
  
No of deals(amount > 0) created on a particular date(Might mean that they were imported into CRM)

* No. of instances where an account has more than 100 deals created on the same day  
  

Need to understand the industry, their integration with freshsales and their sales process for the following:

* The above odd cases(lot of deals created on same date) seem to have correct created\_at, first\_assigned and closed\_date because created\_at < first\_assigned\_at < closed\_date/ max\_date in stage\_histories.  
  They dont have a corresponding lead stage.
* 30425 seems to be creating bulk of deals on the same day. created\_at < first\_assigned\_at < closed\_date/ max\_date in stage\_histories. But many of those deals have same created\_date, same first\_assigned and same closed\_date.  
  
* Accounts with a high fraction of first\_assigned\_at null(first\_assigned\_at null will mean that no owner was assigned to the deal)

          96468,76858,44506,84388,61821,74507,5680,82555

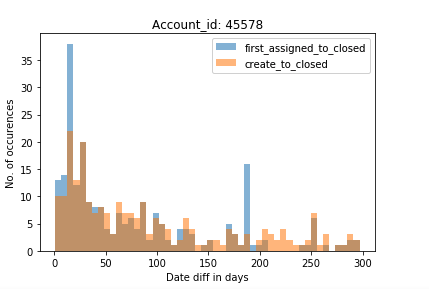
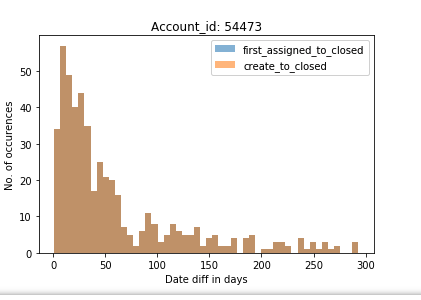
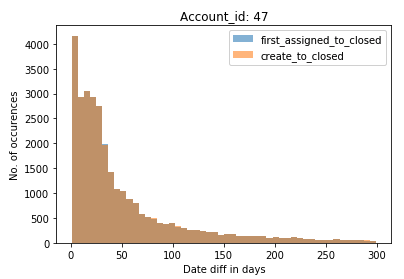
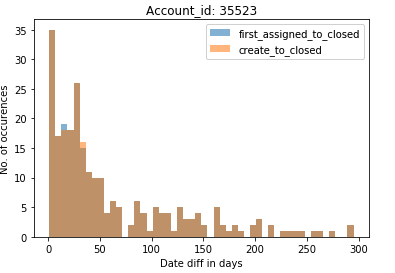
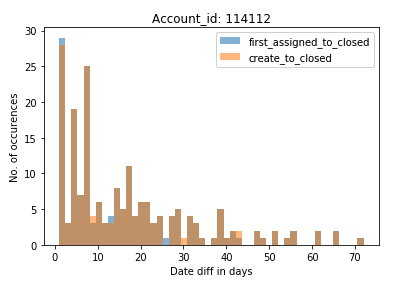
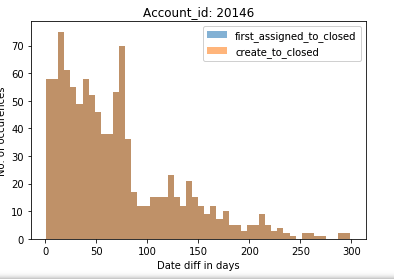
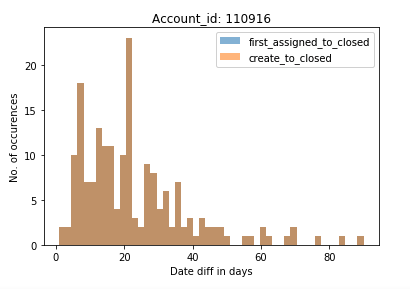
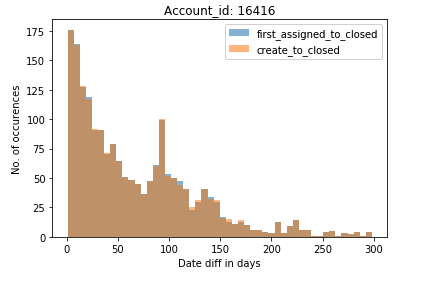
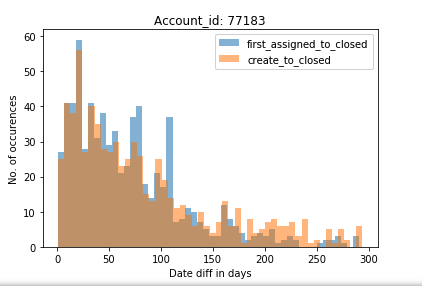
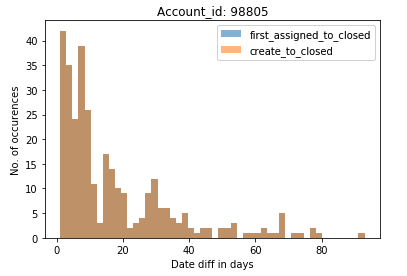
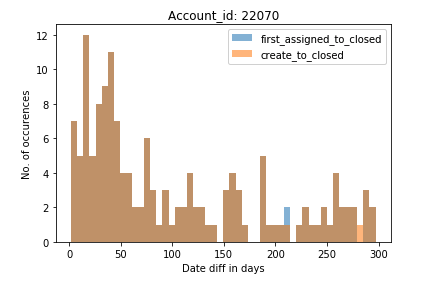
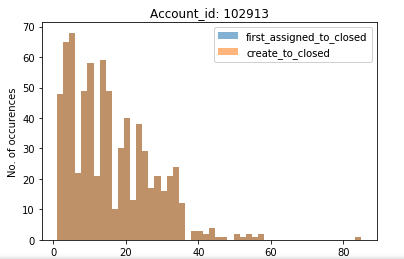
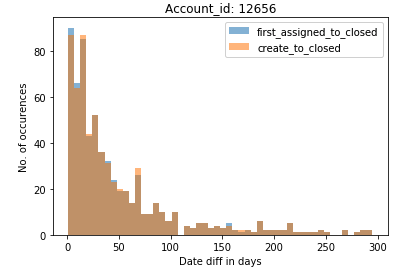
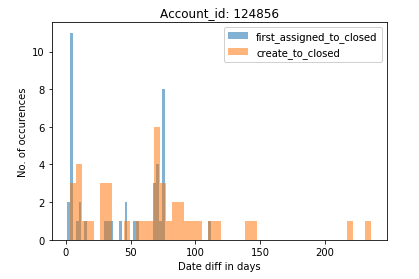
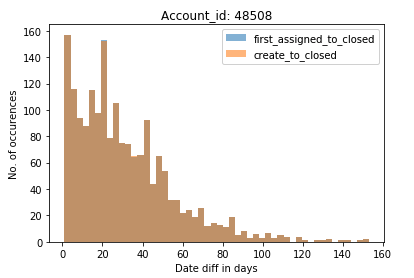
*   
  

Plot of sales cycle (first\_assigned\_at to closed\_date and created\_at to closed\_date)

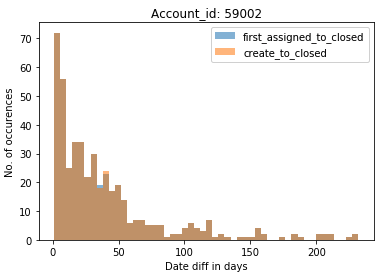
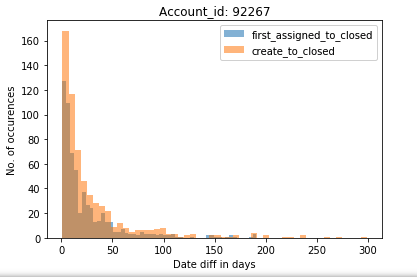
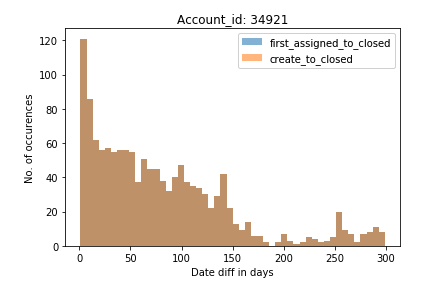
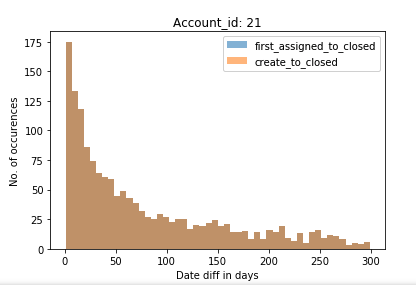
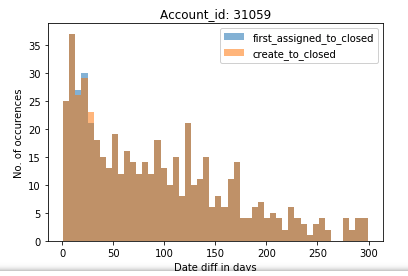
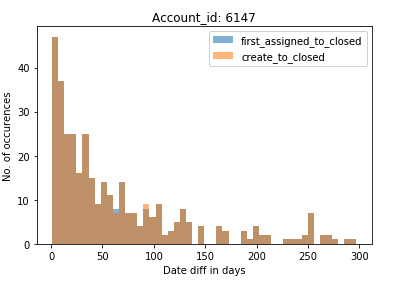
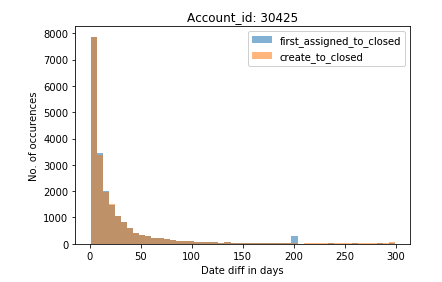
Following additional filters are applied for these graphs:

* sales\_cycle is > 0 and < 300
* first\_assigned\_at is not null

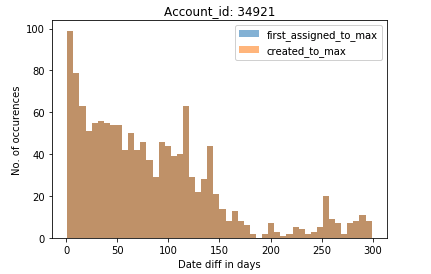
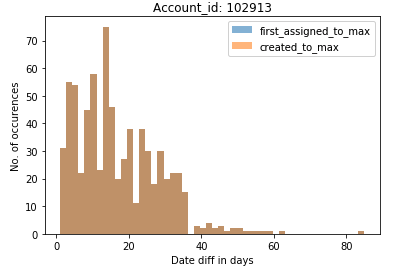
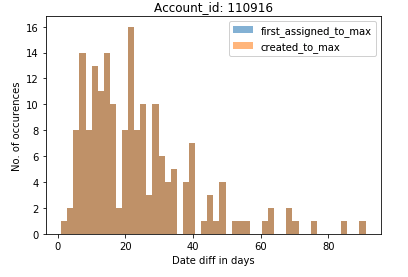
Good accounts:



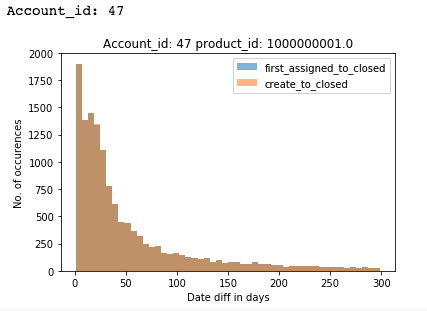
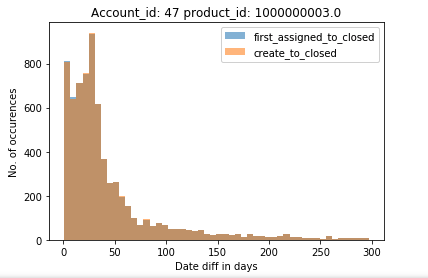
Accounts where the usage of crm has to be understood(peak is close to 0):



For the above accounts plot where sales cycle(first\_asssigned\_at to max\_date and created\_at to max\_date) differs from above:



For 47, Freshsales and Freshdesk have different sales\_cycles:



Sales cycle prediction for different accounts:

* Following accounts should be removed because doesn't have either 5% closed deals or 20 closed deals (after applying general filters):  
  98227, 119299, 115462, 119421, 78062, 30183, 77500, 94243, 107799, 77130, 103773, 120719, 101126, 98701, 31474, 63820
* Accounts with a high fraction of first\_assigned\_at null(first\_assigned\_at null will mean that no owner was assigned to the deal):  
  96468, 76858, 44506, 84388, 61821, 74507, 5680, 82555
* account\_id = 66489 has multiple entries(~200) in stage\_histories for a particular deal, stage combination. This means that a deal comes back to the same stage multiple times.
* account\_id = 118761, has 100/534 where closed date is in the future
* Sales cycle is > 300 days or negative:  
  89865, 62354, 110485, 41753, 108582, 80550, 124583, 77094, 30630, 112432, 28860, 105028, 121413, 72783, 40668, 121309, 57576, 116844, 108400, 94065, 124404
* No clear peak (and less than 20 deals in peak):  
  41585, 61821, 51867, 44506, 96468, 76858, 38707, 22070, 35959, 94260, 5680, 80526, 84388, 30456, 103786, 64848
* Less than 20 deals even at a peak sales\_cycle(could predict sales\_cycle if more deals are added):  
  70978, 124856, 25419, 37592, 115686, 121370, 34305, 88700, 114645, 74507, 93262, 81108, 4420, 94752
* Peak at 0 (could predict sales\_cycle if the reason for peak at 0 is identified ):  
  6147, 30425, 59002, 114112, 34921, 79075, 52241, 92267, 32333, 50388, 51151, 80406
* The following accounts have a non-zero sales cycle if different fields are considered:
  + 48508 first\_assigned to max/closed date
  + 92637 first assigned to max date
* Clear Sales\_cycle:  
  12656, 102913, 31059, 98805, 82538, 45578, 77183, 79247, 75689, 110916, 20146, 21, 35523, 47, 16416, 54473, 82555

Source analysis for won deals on the accounts with 0 as peak(after removing sales\_cycle > 300 and < 0):

Acc 34921 doesnt have any deal starting from lead stage and almost all of them have source as null

acc 6147(71% of closed deals are won) → source NaN → 85% of won deals belong to this source(328/385) → 51%(170/328) are of sales\_cycle 0

                                                                   → max no. of non zero sales\_cycle also belong to NaN source

acc 48508 (72% of closed deals are won) → source NaN → 68% of won deals belong to this source(2614/3837) → 91%(2389/2614) are of sales\_cycle 0

                                                                        → source 'Cold calling'(2000594863.0) → 15% of won deals belong to this source(602/3837) → 83%(502/602) are of sales\_cycle 0

                                                                        → source 'Direct'(2000471127.0) → 13% of won deals belong to this source(505/3837) → 50%(255/505) are of sales\_cycle 0

                                                                        → max. no. of non-zero sales\_cycle belong to 'Direct' source. But NaN also has similar no. of deals that are non-zero sales\_cycle

acc 30425(2% of closed deals are won) → source NaN → 44% of won deals belong to this source (245/555) → 33%(83/245) are of sales\_cycle 0. Not many deals in any another sales\_cycle

                                                                    → source 'Organic Search'(2000233245) → 37% of won deals belong to this source (207/555) → sales\_cycle is very inconsistent

                                                                    → max. no. of non-zero sales\_cycle belong to 'Organic Search' source. But NaN also has similar no. of deals that are non-zero sales\_cycle