**CAR Data Analysis**

Aim: To identify patterns of attrited customers and prevent attrition

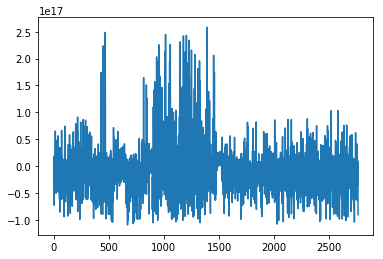
Who are attrited customers?

Attrited Customers: Customers whose agreement ended and they didn’t renew it. Records with agreement end date taken till Aug 2018 for a 3-month buffer window till Oct 2018. The total number of attrited customers comes out to be 6599. We will use this for our further analysis.

Hypothesis: Attrited customers would have less interaction or gradually decrease interaction. To be validated using support case data.

Observation:

Difference between their last support request date and agreement end date.



We can see that the difference is both positive and negative, which means customers are raising support requests both before and after their agreement end date. The ratio of difference for before:after is 2:1, which is quite significant.

count 2762

mean -115 days +09:47:03.523533

std 513 days 17:34:56.422907

min -1266 days +07:48:26

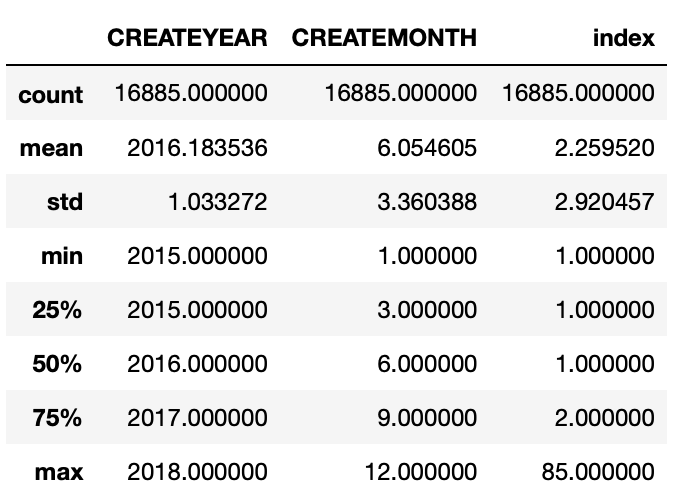
25% -376 days +16:00:32.250000

50% -150 days +17:12:46

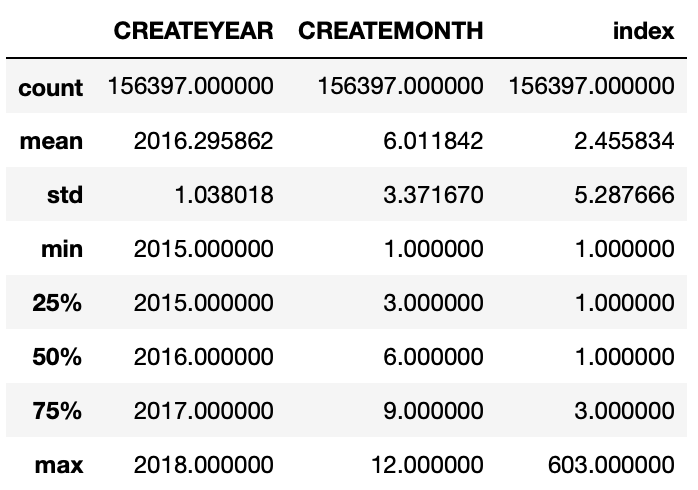
75% 83 days 06:13:21.750000

max 2988 days 15:55:25

Above are some stats about the difference, and we see that the median is about 5 months, which means the last support request raised was 5 months before the end agreement date of an attrited customer. The variation in the above distribution is very high. Trying to identify why the variation is so high I looked into no. of cases being raised by an individual customer on a monthly basis. Thus we can see that the median is only 1 case per customer per month with very close 1st and 3rd quartiles. Since this is a B2B scenario this might be normal, and that might be the reason for such high variation.



Now further comparing this with non-attrited customers



We see that it’s pretty much similar to the distribution of attrited customers except for the maximum.

The case analysis raises a few questions:

1. Why do we have support request date after the agreement end date?

One possible reason could be there are other non-renewable products which the customer is using. Which raises another question.

1. If a customer is still using some other product can we consider him attrited?
2. One more related observation was that there are a few customers that have support cases but no sales order. What is the reason behind this? In case the data is missing that could be another reason for having support cases after product end date.

Unique customers with support cases: 25911

Unique customers with salesorder: 29856

Unique customers with both salesorder and support cases: 16978

Hypothesis:

If a customer has bought a smaller number of line-items he is more likely to attrite.

Observations:

Checking the distribution of count of line items for attrited vs non-attrited customers for 90% quantile. Removed 10% as they were skewing the visualizations.

Attrited:

count 5986.000000

mean 4.358503

std 3.946439

min 1.000000

25% 2.000000

50% 3.000000

75% 6.000000

max 18.000000

Name: ID, dtype: float6

Non-Attrited:

count 8711.000000

mean 6.086098

std 5.102187

min 1.000000

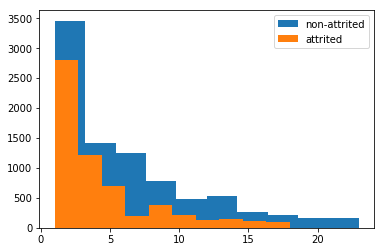
25% 2.000000

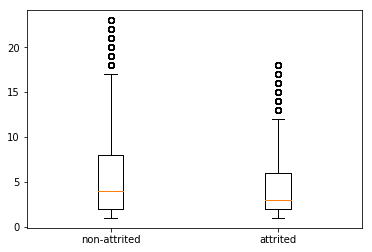
50% 4.000000

75% 8.000000

max 23.000000

Name: ID, dtype: float64





Although the non-attrited class has higher numbers but the distribution mostly looks similar to attrited class, thus this also doesn’t seem to have any distinguishing pattern.

Hypothesis:

Customers who have lower quantities or more refunds seem to attrite more.

Note: Again we have removed the top 10% quantile for better visualization, although logically we shouldn’t.

Observation

Attrited:

count 5939.000000

mean 558.568109

std 803.584739

min -13001.000000

25% 60.000000

50% 224.000000

75% 701.000000

max 3908.000000

Name: QUANTITY\_\_C, dtype: float64

Non-Attrited:

count 8687.000000

mean 1808.417981

std 2458.282334

min -5162.000000

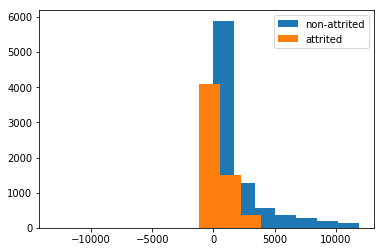
25% 210.000000

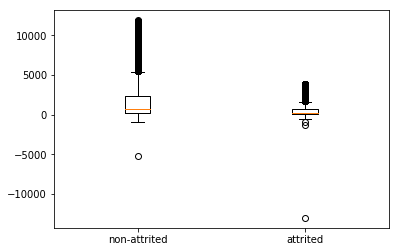
50% 752.000000

75% 2300.500000

max 11894.000000

Name: QUANTITY\_\_C, dtype: float64





We can see the difference attrited customers seem to have more refunds and purchase in lower quantities, hence there is slight distinguishing feature in volumes. However, the employee count of customer might have an impact on quantity, but again it proves that customers with higher employee count are less likely to attrite.

Hypothesis:

Customers with longer relationship tend to attrite less

Relationship: Difference between min(agreement\_start\_date) and max(agreement end date)

Note: Again we have removed the top 10% quantile for better visualization, although logically we shouldn’t.

Observations

Attrited:

count 5951

mean 641 days 05:18:26.435893

std 434 days 13:28:11.782322

min 0 days 00:00:00

25% 364 days 00:00:00

50% 454 days 00:00:00

75% 963 days 12:00:00

max 1826 days 00:00:00

Name: relationship, dtype: object

Non-Attrited:

count 8729

mean 781 days 16:00:26.394776

std 417 days 03:59:29.417867

min -1 days +00:00:00

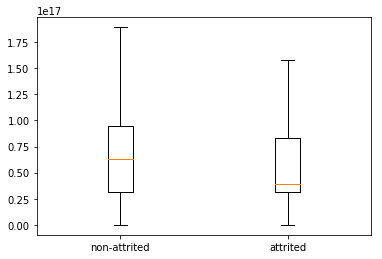
25% 364 days 00:00:00

50% 730 days 00:00:00

75% 1095 days 00:00:00

max 2190 days 00:00:00

Name: relationship, dtype: object



We can see that distribution of relationship for attrited and non-attrited customers is slightly different, however they are pretty much within the same range, hence doesn’t give us much content to discriminate.