ONBOARDING FUNNEL ANALYSIS CASE STUDY

Company: Company X

Author: Chigozirim Nwakanma

Date: 30th of March, 2022

My approach to the case study

1. **Read and understand the task:**
   This is the part where I get to take a quick skim over the whole case study; main objective, dataset provided, specific questions to be answered, etc. Every case study has a story behind it, so this aspect helps to paint that picture for me, which will help formulate my thinking as I go forward.
   It is important to follow the instructions first before going the extra mile. In this case study, I almost missed where I was asked to analyse data for only specific months data in the brief.
2. **Prepping the data:**
   Now that I've taken a good look and have clarity of the data provided. I get to clean the data provided, in this case 2 spreadsheets, using T-SQL on MS SQL Server as my tool of choice. At this stage, I get to inspect the data more closely for any errors, inconsistencies and ambiguities. If I can, I make corrections and adjustments, otherwise, I request for more clarity from the data source. In this case, leaving comments and recommendations.
3. **Deep dive:**
   Analyzing the data using SQL to answer the case-specific questions regarding the funnel stages. Also, noting any adjustments, assumptions and omissions made during the analyses. Creating visualizations in MS Excel, to explain the insights gained from analyses.
4. **Recommendations:**
   Based on insights gained, I give recommendations on how to achieve the case objective or drive results.

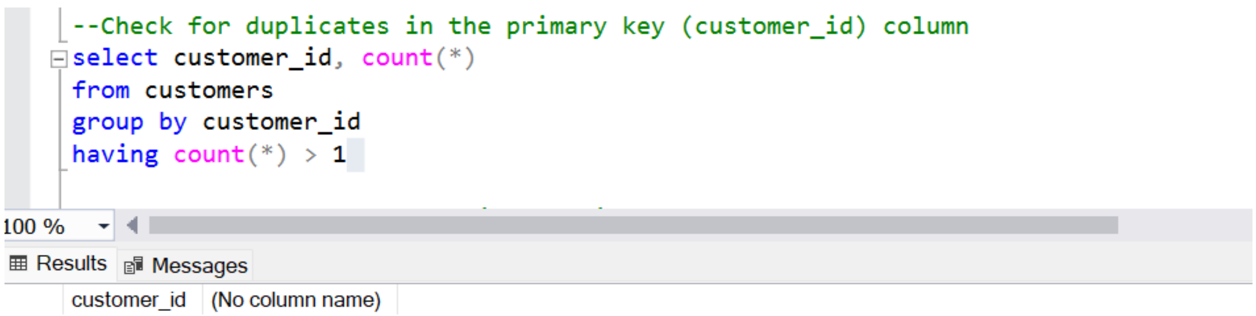
Sales and Marketing have requested reporting on the intake funnel for January, April, July, and September.

1. Looking at the records for this month, are there any issues with the data set?

a. If issues are found, how should these be explained to Sales and Marketing leadership?

Answer - Yes, some of the issues discovered within the dataset include:

- There are rows with NULL values in the *tos_accepted_date* column while having date entries in the *first_purchase_date* column. This makes it impossible to calculate the difference, hence figuring out if the different stages of the funnel were a success or not.
- There are about 27 rows where the *first_purchase_date* occurs before the *tos_acceptance_date*, which results in a negative difference value. This should not be possible as first purchase can't be made without first accepting the ToS.
- There is no clear indicator as to non-compliance with ToS, or how to tell if and when the terms were re-accepted.
- Although relatively trivial, the dataset would be a lot easier to use if the entries are all aligned uniformly (right, left or central). This helps avoid complications during file import into SQL.



Checking for duplicates in the primary key

b. Do you have any suggestions to decrease the chance of records being incompatible, or additional fields that could be useful?

Answer – Some additional columns that could be added for clarity and more efficient analysis include:

- A column containing just the month, extracted from the *signup_date*.
- 2 new columns to record "*tos_acceptance_date – signup_date*" and "*first_purchase_date – tos_acceptance_date*" to help identify if each stage is successful or not.
- An additional column that indicates when terms of service was violated and the date of violation.

```sql
-- Creating new columns to count the difference in days from one step to the next
-- Also a column to state if the transition was a success or not
select *
from
(Select customer_id, customer_name, signup_date, signup_source, referrable_id, customer_state,
customer_type, tier, tos_accepted_date, first_purchase_date, DATEDIFF(day, signup_date, tos_accepted_date) as signup_diff,
DATEDIFF(day, tos_accepted_date, first_purchase_date) as diffbetweentos_purchase,

case
    when DATEDIFF(day, tos_accepted_date, first_purchase_date) <= 30 and DATEDIFF(day, signup_date, tos_accepted_date) <= 7   then 'Y'
    else 'N'
end as Success_measure
from
(select customer_id, customer_name,
TRY_CAST(signup_date as date) as signup_date,
signup_source, referrable_id,
customer_state,customer_type ,tier,
TRY_CAST(tos_accepted_date as date) as tos_accepted_date,
TRY_CAST(first_purchase_date as date) as first_purchase_date
from customers)a
                    )b
where tos_accepted_date is not null
```

00 %

Results Messages

	customer_id	customer_name	signup_date	signup_source	referrable_id	customer_state	customer_type	tier	tos_accepted_date	first_purchase_date	signup_diff	diffbetweentos_purchase	Success_measure
1	1	Jerimiah Navarro	2020-01-01	trade_show	NULL	NM	enterprise	2	2020-01-02	2020-02-06	1	35	N
2	2	Harley Mckinney	2020-01-01	referral_partner	3	FL	enterprise	1	2020-01-02	NULL	1	NULL	N
3	3	Jessie Meadows	2020-01-01	trade_show	NULL	CA	enterprise	2	2020-01-02	2020-01-21	1	19	Y
4	23	Elise Villarreal	2020-01-02	trade_show	NULL	VA	individual	2	2020-01-03	2020-01-24	1	21	Y
5	24	Isabella Callahan	2020-01-02	trade_show	NULL	CA	enterprise	1	2020-01-02	NULL	0	NULL	N
6	34	Jeremy Drake	2020-01-03	referral_partner	3	OR	individual	2	2020-01-06	2020-01-07	3	1	Y

Creating new columns to measure funnel stage success

2. Create visualizations for each of these months, illustrating how many customers were successfully onboarded.

a. Note any adjustments/omissions that you need to make to build these visualizations

- Negative differences were excluded, as those imply that first purchase was made before ToS was accepted
- Rows with NULL date entries were exempted
- Only successful transitions were taken into account for the monthly analysis chart.

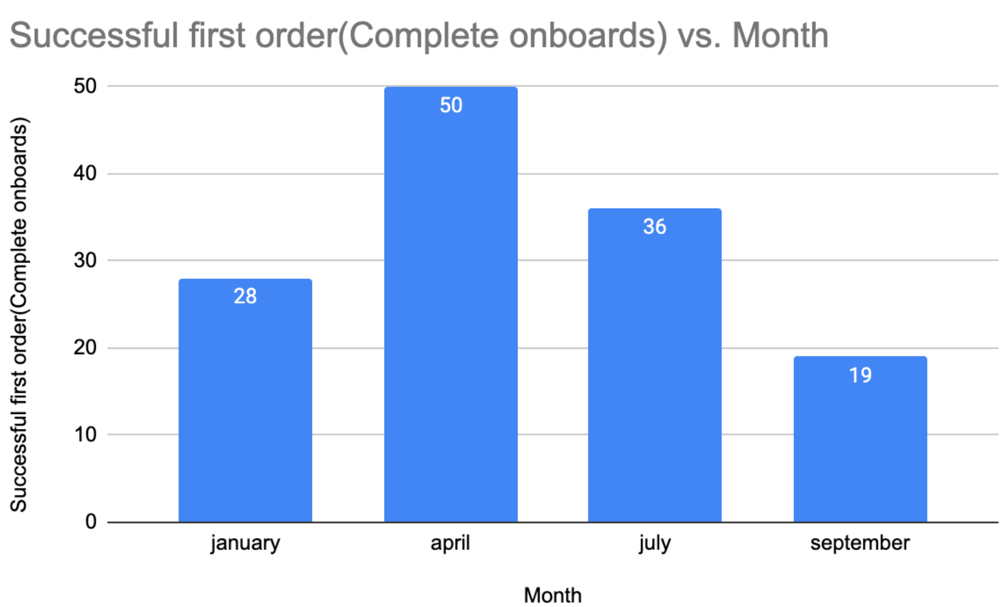


Figure 1. Successful onboardings for specified months

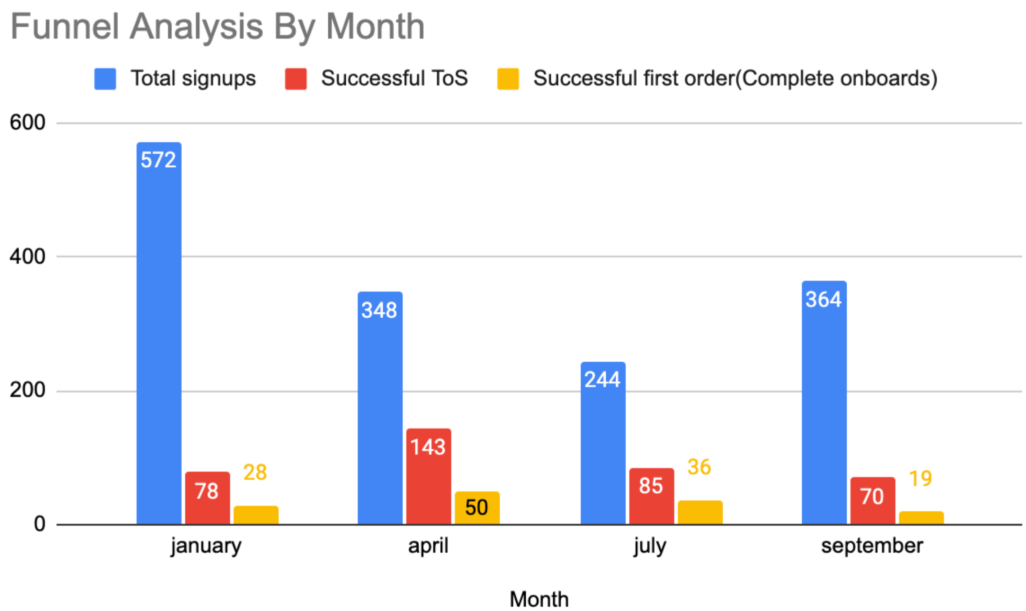


Figure 2. Success level of each funnel stage for specified months

Additional Tasks

Sales and Marketing have been using this metric for a number of years, and the business has gone through numerous changes. After reviewing your visualization, they return with the following questions. You've been asked to take a cursory look, and if anything jumps out at you to share your early findings to kick-start a project with the Data Science team.

1. Do the conversion timelines for the metrics still make sense?
   Generally, when a company makes numerous changes, those changes should be reflected on most, if not all facets of the business. In this case, it would be beneficial to adjust the timelines used to measure these metrics, to ensure consistency across the board.
   For instance, in table 1 below, it is seen that the highest number of complete onboardings happen with tier 2 customers, despite the tier 1 customers being more heavily targeted. This is one area where the company may want to invest more into campaigns that target tier 2 customers, in order to maintain and potentially boost that sector even more.

Tiers	Complete onboardings
1	89
2	132
3	6

Table 1. Number of complete onboardings by Tier

2. How long does it typically take for a 2020 customer to move from each step to the next?

	All	Partner referrals	Customer referrals	Others
Avg no of days before ToS acceptance (days)	8.5	5	57	4
Avg no of days from ToS to first purchase(days)	43	45	11	43

Table 2. Time taken for a 2020 customer to progress to next stages

3. Do referred customers have any different behavior than the rest?
Referred customers tend to behave differently, depending on what kind of referral. From the table above, it is seen that partner referrals tend to be more effective in getting the customers to accept the ToS more quickly and in almost half the time it takes in general. This could be based on the level of trust perceived by the customer, when dealing with a known company partner, as opposed to an online ad posting.

However, customer referrals result in faster first purchase compared to partner referrals and other signup sources, and this could be as a result of potential customers being able to get a first-hand experience of the product from the existing customer who referred them.