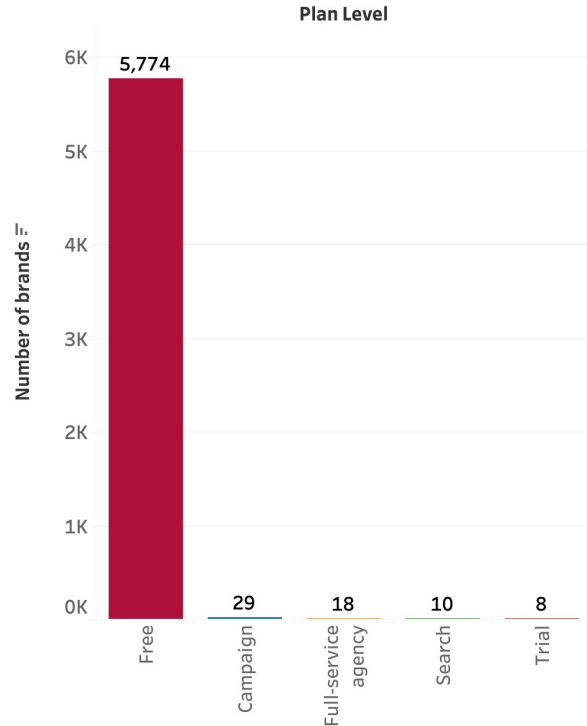# Analytics Data Challenge

Deliverable 1
[Link to Tableau visualizations](#)

Distribution of subscription plans


Distribution of different plan levels

Plan Level

- Dataset has 5 different plan levels:
  - 0 - Free, 1 - Trial, 2 - Search, 3 - Campaign, 9 - Full-service Agency
- Dataset is dominated with brands still using **Free** plan - plan Level **0.**
- Among the minority, brands use **Campaign** (plan level **3**) followed by **Full-service agency** (plan level **9**).
- Dataset is highly skewed, with ~ **98.8%** of the brands in the **Free** plan.

How many brands upgraded to a paid subscription? (plan level > 0)

Distribution of brands with free plan and paid subscription
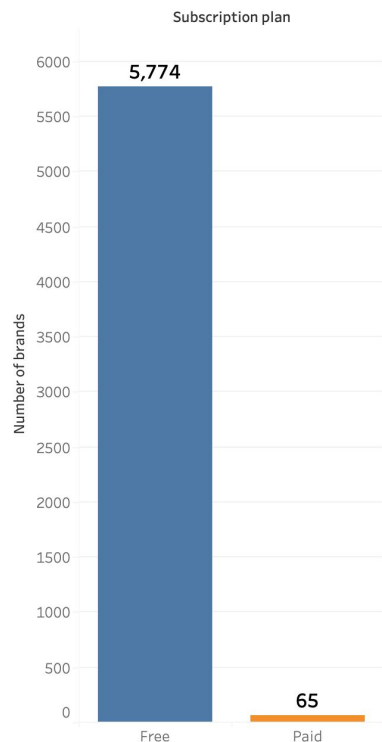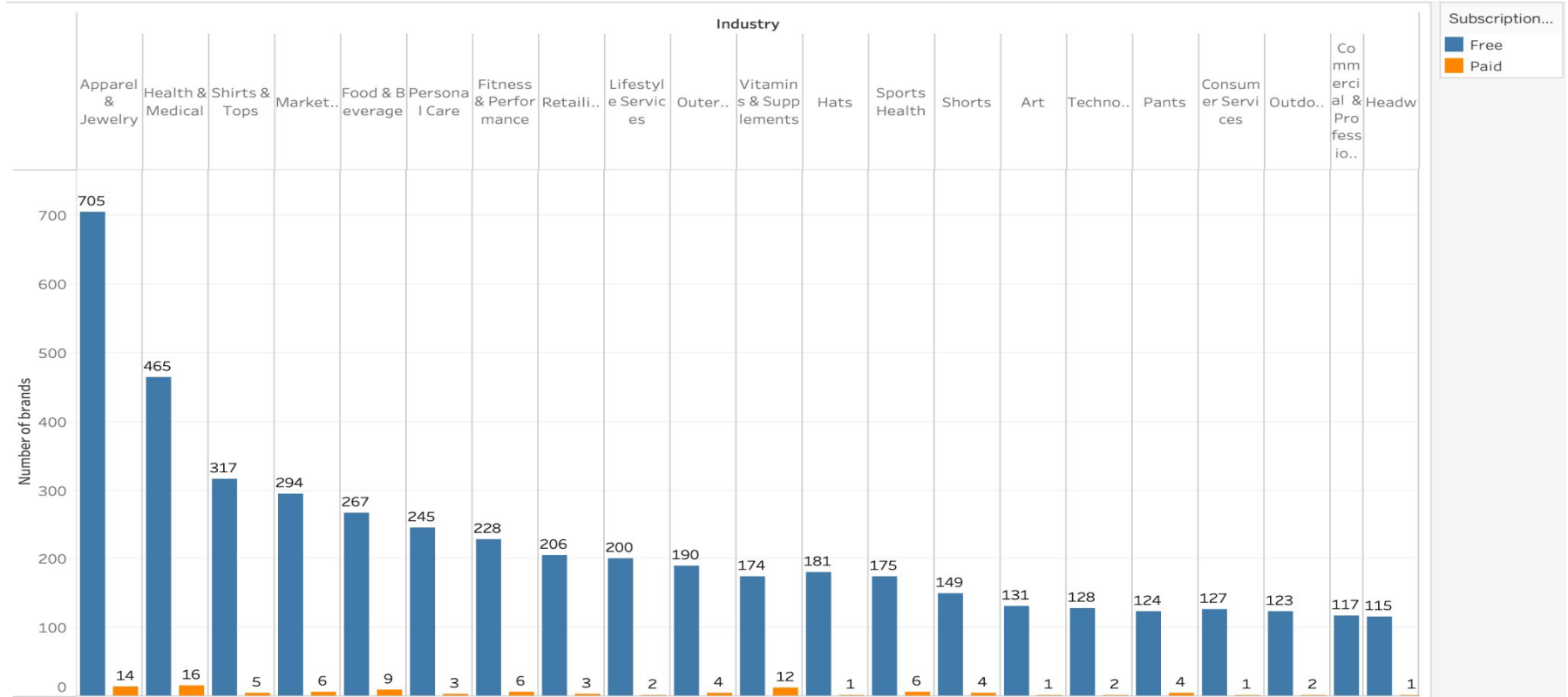
Subscription plan

- Out of 5839 unique brands, only ~ **0.01% (65)** of brands have upgraded to a paid subscription

Industries favourable to free plan and paid subscription

- 325 unique industries in the dataset covering all plan levels
- 65 brands with paid subscription belong to these 65 industries -
  - 'Apparel & Jewelry', 'Shorts' 'Sleepwear & Loungewear', 'Masks', 'Shirts & Tops' 'Outerwear', 'Hats', 'Retailing', 'Internet & Direct Marketing Retail', 'Health & Medical', 'Vitamins & Supplements', 'Food & Beverage', 'Lifestyle Services', 'Snacks', 'Fast Food', 'Energy Drinks', 'Marketing', 'Advertising', 'Sports Drinks','Fitness & Performance', 'Jewelry Stores', 'Food Distributors', 'Food Retail', 'Distributors', 'Toy Stores', 'Heath & Vision Care Stores', 'Jewelry Sets', 'Packaged Foods & Meats', 'Real Estate', 'Real Estate Development ', 'Diversified Financials', 'Consumer Finance', 'Consumer Durables', 'Outdoors', 'Technology', 'Suits', 'Pants', 'Sports Health', 'Personal Care', 'Skin Care', 'Female Care', 'Hair Care', 'Cosmetics', 'Wine', 'Champaigne', 'Headwear', 'Footwear','Education Services', 'Consumer Services', 'Casinos & Gaming', 'Art' 'Pets', 'Pet Food', 'Pet Clothing', 'Dog Training', 'Animal Shelter', 'Animal Adoption', 'Apparel', 'Dresses', 'Skirts', 'Broadcasting & Media Production' 'Software & Services', 'Investment Banking & Brokerage', 'Energy Bars' and 'Uniforms'
- 5774 free plan brands belong to all the 325 industries
- Top 20 industries favourable to free plan
  - 'Apparel & Jewelry', 'Health & Medical', 'Shirts & Tops', 'Marketing', 'Food & Beverage', 'Personal Care', 'Fitness & Performance', 'Retailing', 'Lifestyle Services', 'Outerwear', 'Hats', 'Sports Health', 'Vitamins & Supplements', 'Shorts', 'Art', 'Technology', 'Consumer Services', 'Pants', 'Outdoors', 'Commercial & Professional Services'
- Top 20 industries favourable to paid subscription
  - 'Health & Medical', 'Apparel & Jewelry', 'Vitamins & Supplements', 'Food & Beverage', 'Marketing', 'Sports Health', 'Fitness & Performance', 'Shirts & Tops', 'Outerwear', 'Pants', 'Shorts', 'Personal Care', 'Advertising', 'Footwear', 'Retailing', 'Skin Care', 'Technology', 'Consumer Finance', 'Diversified Financials', 'Energy Drinks'

# Compare top 20 industries favourable to free plan vs paid subscription

Comparing the top 20 industries favourable to free plan vs paid subscription

- Of 65 industries that upgraded to a paid plan, **Health and Medical (16), Apparel and Jewelry (14), and Vitamins and Supplements (12)** are the top 3 industries with most number of brands moving to a paid plan.
- **Health and Medical (465), Apparel and Jewelry (705)** are also the top 2 industries with most number of brands in the free plan.
- Most of the top 20 industries in the free plan did not upgrade to paid subscription.
- It is interesting to note that **Retailing, Hats, Art, Technology, Consumer Services, and Commercial and Professional services** (less related to sports) did not switch to paid subscription.
- **Health and Medical, Apparel and Jewelry, Vitamins and Supplements, Sports Health, Shirts and Tops, Food and Beverage, Pants, Shorts, Outerwear, Personal care and Sports Health** related to sports are favourable to subscription.

- Some of the least popular industries in the free plan with less than 5 brands from these industries are -

    - 'Printers, Copiers & Fax Machines', 'Industrial Machinery', 'Storage & Wearhousing', 'Precious Metals & Minerals', 'Security & Alarm Services', 'Office REITs ', 'Electric Utilities', 'Multi-Utilities', 'Gas Utilities', 'Cutlery', 'Water Utilities', 'Steel', 'Equipment Repair Services', 'Industrial Maintenance Services', 'Forest Products', 'Animal Shelter', 'Industrial REITs', 'Integrated Oil & Gas', ''Heavy Electrical Equipment', 'Sashes', 'Oil & Gas Equipment & Services', 'Oil & Gas Storage & Transportation', 'Independent Power Producers & Energy Traders', 'Renewable Electricity ', 'Retail REITs', 'Construction Materials', 'Hand Muffs', 'Pet Clothing', 'Aluminum', 'Health Care Services', 'Diversified REITs', 'Coal & Consumable Fuels', 'Consumer Auctioneers', 'Electrical Components & Equipment', 'Tobacco Retail', 'Paper Packaging', 'Industrial Gases', 'Home Security', 'Metal & Glass Containers', 'Fertilizers & Agricultural Chemicals', 'Aerospace & Defense', 'Beverage', 'Diversified Metals & Mining', 'Specialty Chemicals', 'Funeral Services', 'Copper', 'Commodity Chemicals', 'Reinsurance', 'Uniform Rental', 'Industrial Auctioneers', 'Commercial Cleaning', 'Diversified Chemicals', 'Light Fruit Juices', 'Personal Products', 'Media', 'Printers', 'Software', 'Mobile Phone & Tablet Tripods & Monopods'

- Most of these industries are not immediately associated with sports which might indicate why only a few of them are even in the free plan.

Company size and different paid subscription levels


Company size and different paid subscription levels
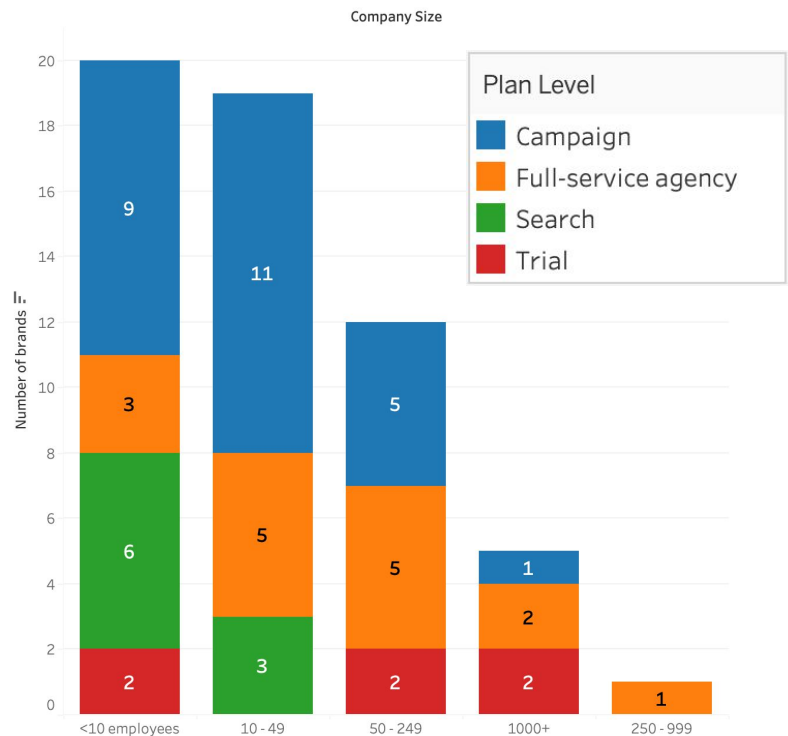
Observations:

● It is clear that the brands with less than 250 employees have subscribed to a paid plan more than the larger brands.

● Brands with employees between 10 - 250 have chosen the highest subscription level **9 (Full-service agency)** more than any other brands, closely followed by companies with less than 10 employees.

● Only the brands with **less than 50 employees** have the **Search (level 2)** plan.

● Only 1 brand with more than 250 employees has a paid plan - **Campaign (level 3)**.

Insights: It'd be easier to analyze why some subscription levels were chosen if the features of each plan were provided.

Company size and free plan

Company size and free plan

Company Size
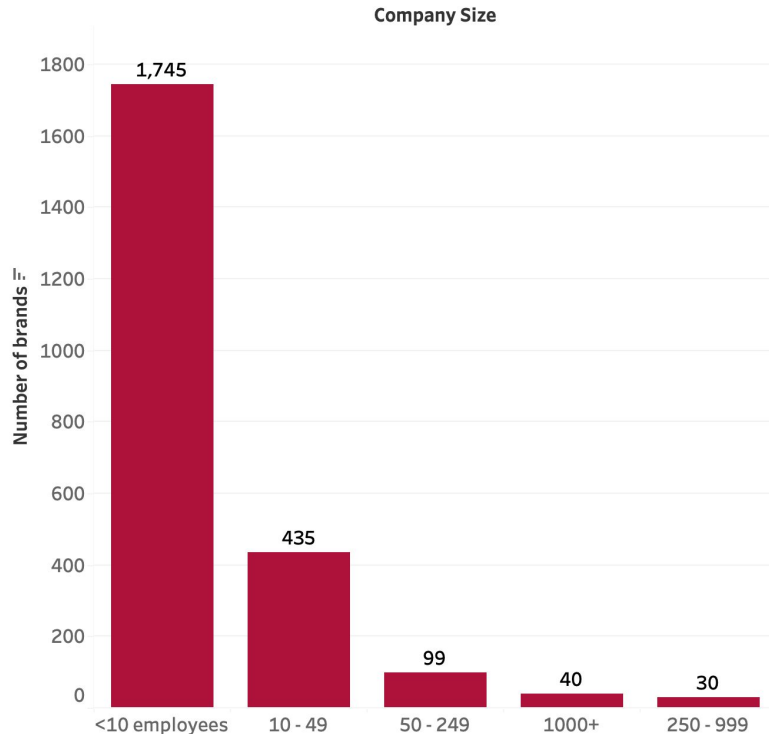
Observations:

- Most of the brands in the free plan (1745) have **less than 10 employees**, followed by brands with employees between 10 and 50 (435).

Insights:

- This could be because the smaller companies might **not have much funding/budget** to switch to a paid version.
- They also might be trying the free version to **check out the features before upgrading**.

Company size and how active paid subscription vs free plan is

Company size and how active paid subscription vs free plan is
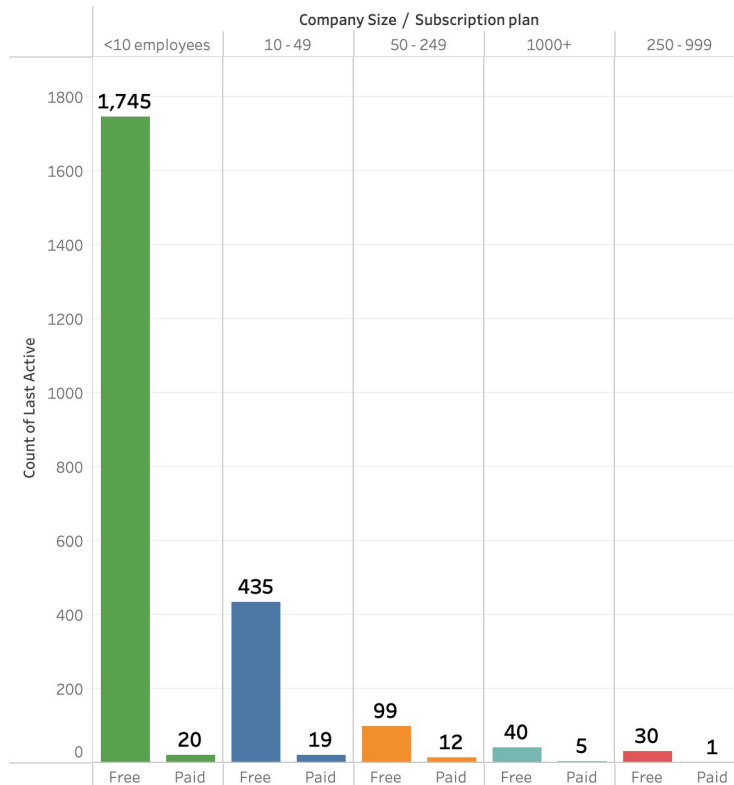
Observations:

- From the plot, it is clear that **brands with less than 50 employees** are highly active on the platform in both the free and paid plans.
- Brands with less than 10 employees on free trial being the most active of all.
- The activity of free and paid plan is consistent across company size, i.e., brands with less than 10 employees have the highest activity on both free and paid plan, followed by brands with employees between 10 and 50, with the lowest activity registered by brands with employees between 250 and 1000 on both free and paid plan.

Insights:

- It could be that the brands with fewer employees have l**ower budget** and would make the most of the features available with the free plan.
- As the brands with more employees were not very active during free plan, they might not have explored the platform well and were likely to **not switch to a subscription model**

Activity trend - paid vs free plan
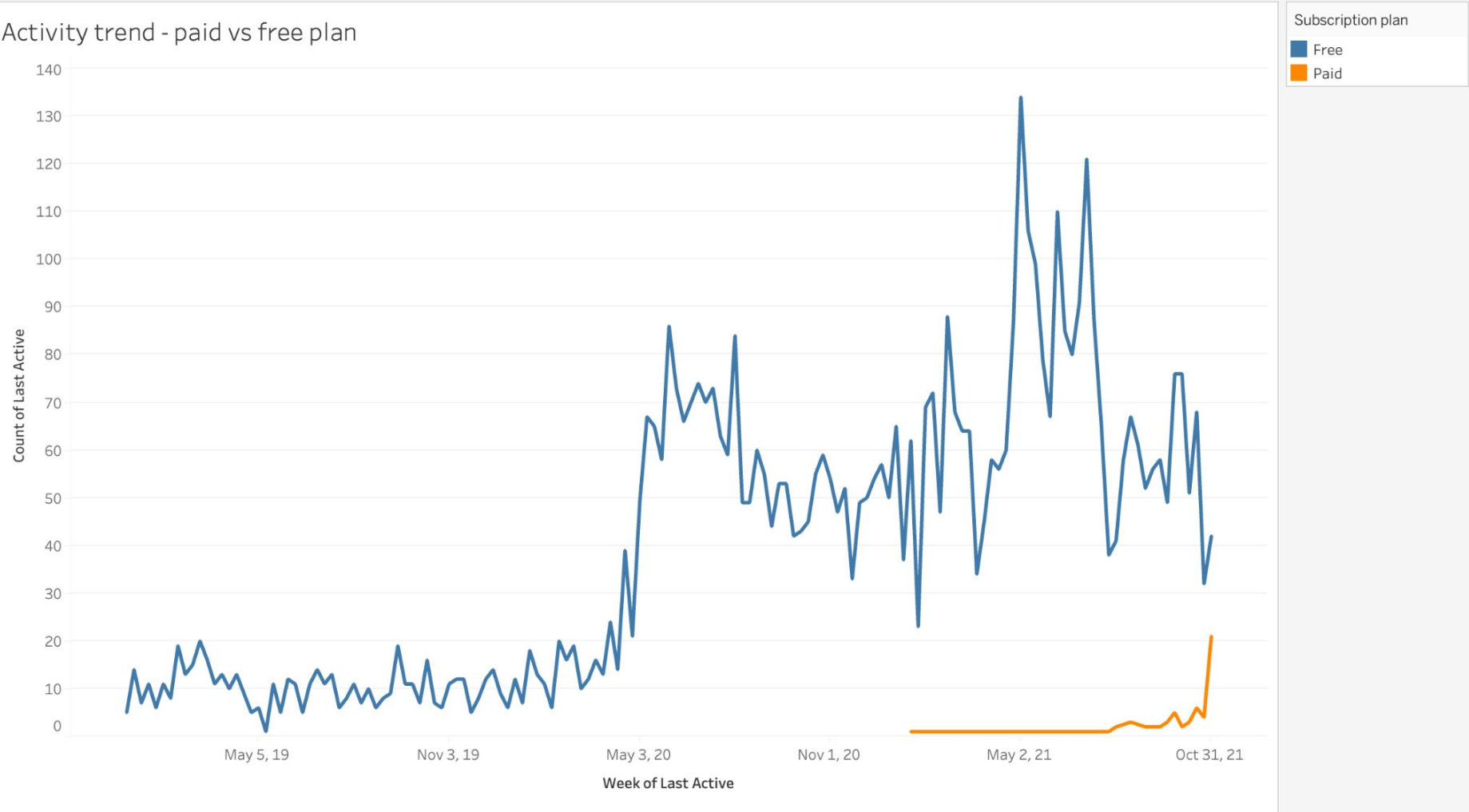
Activity trend - paid vs free plan

- There is an increase in activity from May 2020, and there is a similar trend in May 2021 by brands with free plan
- The paid plan brands started activity from 2021. There was no activity before that
- The activity of brands with free plan plummeted during October 2021 while the activity of subscribed brands peaked
- The subscribed brands were not very active during the first half of 2021
- The data shows the activity during COVID period and the trends right might not be the same with less restrictions

What's the limitation of the data set? What additional data points do you think that can be added to this study? Why?

Limitations:

- The data does not include most recent data points, which might help understand current trends and behaviours
- The data shows the activity during **COVID** period and the trends right might not be the same with less restrictions
- Free plan users are over-represented.

Additional data points:

- The features offered by each subscription plan could add more light on to why some brands upgraded to a subscription.
- User feedback and surveys can help improve the platform and understand the reason for low retention
- Information on how brands discovered the company (e.g., referral, online search, social media) could help understand the most effective marketing channels for user acquisition.
- More detailed geographic data, such as regional or city-level information, can aid in identifying potential market opportunities or regional preferences.

Deliverable 2

1. What is the ratio of brands that published campaigns vs. did not publish any campaign within each plan level?

Published vs Unpublished campaigns within each plan Level
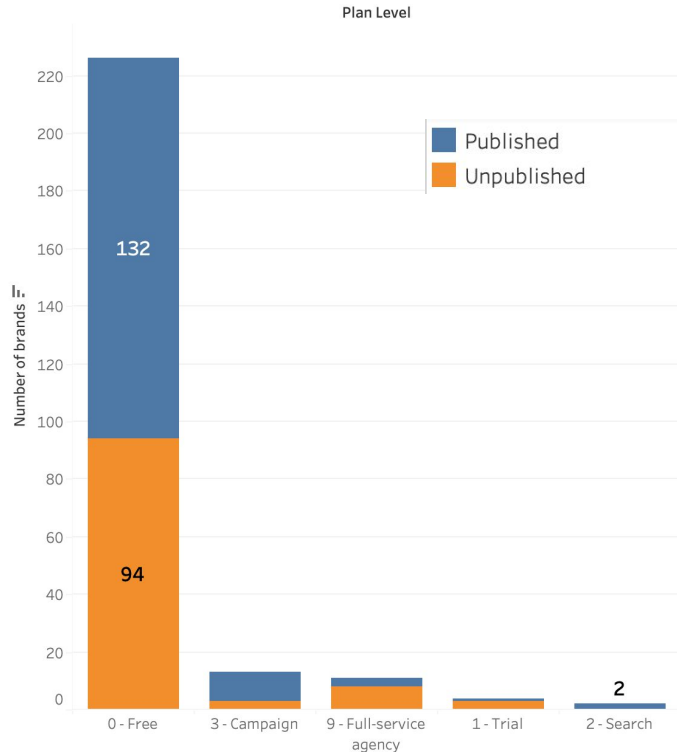
Calculation:

- To calculate the ratio, firstly, the campaigns data (left) and users data (right) were left joined on `user_id` column
- **Denominator** - The number of brands that did not publish any campaign was calculated by grouping `planLevel` and counting the number of `isUnpublished` unpublished campaigns
- **Numerator** - The number of brands that published campaigns was calculated by subtracting the number of brands that did not publish from the total number of brands.

1. What is the ratio of brands that published campaigns vs. did not publish any campaign within each plan level?

- The ratios within each plan level

Table 1	
planLevel	**Ratio**
0 - Free	1.404255
1 - Trial	0.33333
2 - Search	0
3 - Campaign	3.33333
9 - Full-service agency	0.37500

- The ratio of search plan (2) is 0 because all the brands within this plan (only 2 brands) published a campaign
- Free plan (0) and Campaign plan (3) have ratios greater than 1 because they have more number of brands that published campaigns.
- It is interesting to see that more number of campaigns were published during free plan.

2. How many brands continued to create campaigns after their first published one?

Calculation:

- Filter to include only published campaigns
- Sort the filtered data by 'user_id' and 'campaign_created' (date) columns in ascending order
- Group by 'user_id' and use transform to get the first published campaign for each group
- Filter to include only campaigns created after the first published one
- Count the number of brands that continued to create campaigns after their first published one

Another method:

- Filter to include only published campaigns
- Get the number of times each `user_id` appears in the data
- The number of times a user_id appears in a dataframe is going to be > 1, if the brands create campaigns after the first published one
- Find the user id's that appear more than once in the data

ANSWER - 5 brands continued to create campaigns after their first published one.

Table showing the user ID's and the campaigns published along with the budget available for each campaign

Table 2			
User ID	**Campaign Number**	**Total Budget**	**Budget Availability**
A	1	5120	Yes
A	2	5120	Yes
A	3	12620	Yes
A	4	5620	Yes
A	5	0	No
A	6	5120	Yes
A	7	15120	Yes
A	8	0	Yes
B	1	230	Yes

Table 2			
User ID	**Campaign Number**	**Total Budget**	**Budget Availability**
B	2	100120	Yes
C	1	missing	Yes
C	2	missing	Yes
C	3	missing	Yes
D	1	620	Yes
D	2	2020	Yes
E	1	100120	Yes
E	2	40000	Yes

Table 2 details

- These are the 5 users that published campaigns after the first one. The user ID's A, B, C, D, E map to

Table 3	
A	56fb496c7ed840380e1c1a8e
B	5efb7f52dc1fab678e8b8492
C	5f5692666881dc0d56f8d904
D	603c9e595557ba0d6a819a8c
E	60f73c0d69248c0d60fde6da

- **Campaign number** refers to the order of the campaigns published. 1 being the first one and so on
- **Total budget** refers to the total budget available for each campaign. Calculation of total budget is explained for question 3.
- **Budget availability** says if the budget is available for a campaign or not.
- The 5th campaign of User A is the only campaign that published without a budget available and the 8th campaign of User A published with 0 budget.

Insights: User D's campaigns and 1st campaign of User B are **published with very low budgets**. The information on the **type of campaigns** would help analyze the **relationship between the budget availability and campaigns published**.

3. Which brands have the highest paying campaigns?

Calculation:

- In the `payment` column of the data, there are 4 different budgets - cash, product, royalty and equity. Assuming that the brands with highest paying campaigns will have the highest total budget, which is the sum of cash, product, royalty and equity budgets, the brands and their campaigns with the highest budget were selected.
- Check the Jupyter notebook for more details

3. Which brands have the highest paying campaigns?

- Table 4 shows the **top 5** brands with highest paying campaigns and their budget

| Table 4			
Brand / User ID	**Campaign ID**	**Total Budget**	**Budget Availability**
5fac7d3e220e690d333e01ff	5fadff3d414c780d3fe7eef1	600020.0	Yes
5ebf5f5634cd0a0de29d782f	5ebf5ff7f6bc7e0de3140267	110011.1	Yes
60bfaf9d074fb50d59e60187	60bfb8700969130d5ad6c11f	105920.0	Yes
5f25e38d7e8a2b0dfe538335	5fd56919bdf6e10d694df52b	105000.0	Yes
5d8689aa01688c0dc6472965	5d868c3501688c0dc6472a42	104689.5	Yes

4. How long does it take for new brands to publish a campaign?

- To calculate that we need to find the difference between the date the brand was created and the date of the first campaign launched by the brand
- Check Jupyter notebook for more details on how it is calculated
- Table 5 below shows the brands and campaigns along with the time taken to publish their first campaign.
- The first 2 rows show the quickest ones to launch a campaign and the next 2 rows show the brands that took the longest time

Table 5			
Order	Brand	Campaign	Time taken to publish first campaign
Quickest 1	5f5bb1e4ee03f20dc9bf718a	5f5bb237f8d8a50d75192289	0 days 00:01:23
Quickest 2	5e9dba58ecb3ba0dd051efb7	5e9dbab2c24aa30dd12c9c47	0 days 00:01:29
Slowest 1	5d4afb92fa5c1f26f9205f9a	60cccad3efa4b20d86c31b21	681 days 00:07:30
Slowest 2	5ca9641a288a552580879ed5	5eaf1855dbe0330de1da424f	392 days 16:30:51

4. How long does it take for new brands to publish a campaign?

Distribution of the hours taken for new brands to publish a campaign
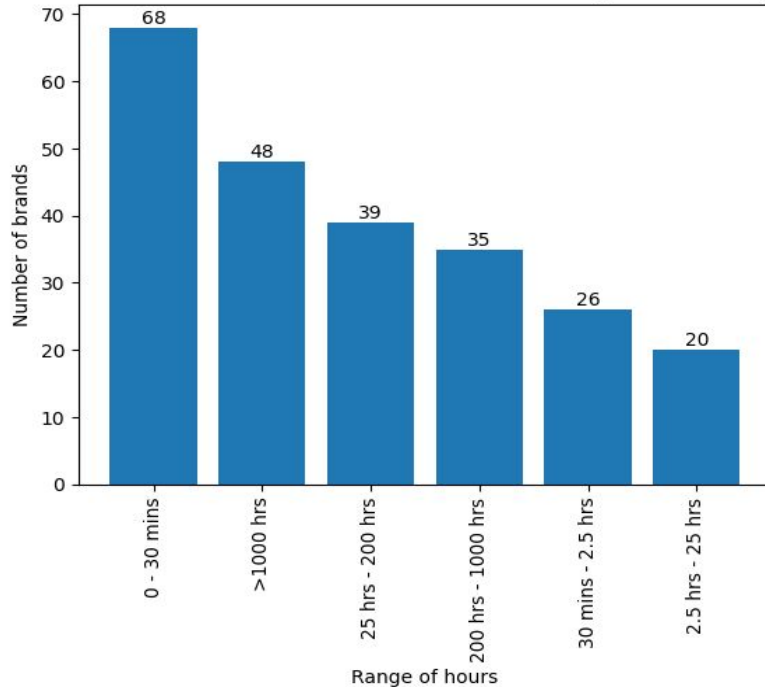
- For new brands to publish a campaign, it takes a few minutes to a few years. Few brands published campaigns within minutes with the fastest being 1 minute 23 seconds and a few brands took years to publish their first campaign, the slowest one taking 681 days.
- The bar plot shows the distribution of how long brands take to publish their first campaign.

Observations:

- More brands took less than 30 minutes from creating their profile to launch a campaign
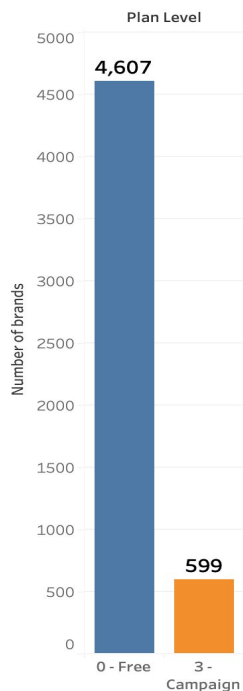- 48 brands took more than 1000 hrs (~42 days)

Insights: The details about the campaigns would help identify why the time taken to launch a campaign varies

Deliverable 3
Link to Tableau Visualizations

Understanding the distribution of plan Levels in the data

Plan Distribution

What trends do you see for plan level 0 and plan level 3?

- About **88.5% brands** in the data are in the free plan (level 0)
- This data is also **skewed** towards free plan
- Unlike other datasets in Task 1 and 2, this data only has brands with Campaign plan and no other paid plan

Analyzing the number of **views** made by each plan every month

Number of views created by each plan every month
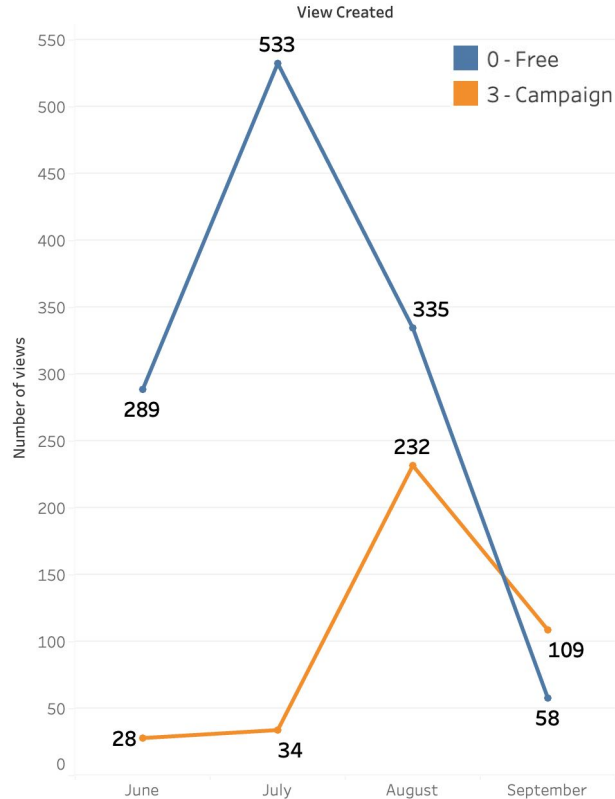
What differences would you highlight between the trends in search for plan levels?

- The number of views made by **campaign plan** have jumped significantly in August and the activity reduced in September.
- **Free plan** were most active during July and the views declined in the next months.

Insights:

- The reason for the activity between June and September could be due to the major sports leagues like **NBA, NHL** beginning after September and the brands were looking for players to advertise their product, to **attract a large audience**. (More details could be found if the sport associated with athletes was provided)
- The brands with the free plan might be more active during July to explore features to decide whether to upgrade their plan or not.

Analyzing the number of **searches** made by each plan every month

Number of searches by each plan every month
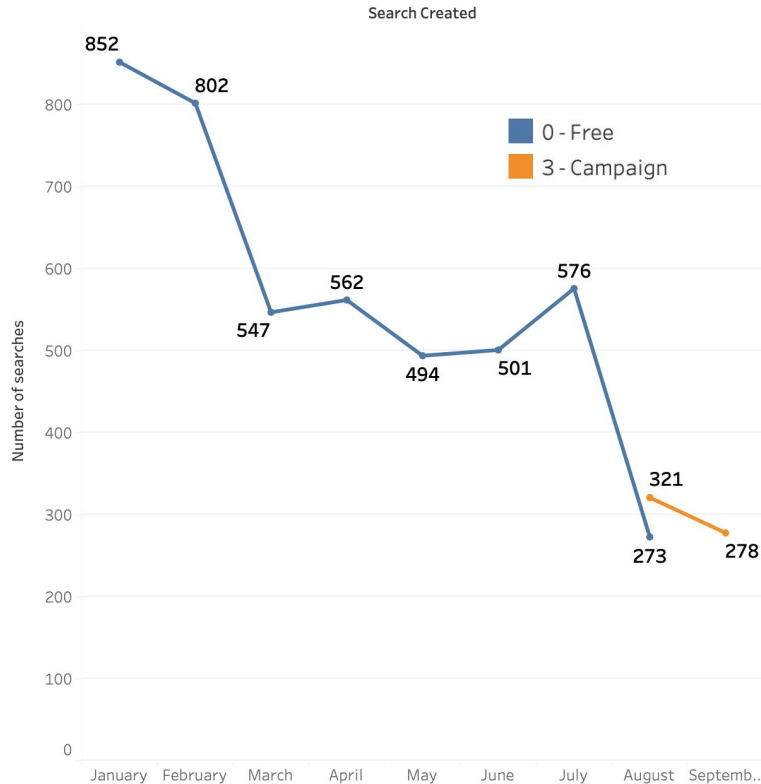
What differences would you highlight between the trends in search for plan levels?

- The number of searches made by **free plan** brands were high between January and July. The number dropped in August and with **0 searches in September**
- The brands with **Campaign plan** only started searches in August and September.

Insights:

- It looks like the **free plan** brands were more actively searching for the players that meet their needs in the first half of the year.
- The **reason for no activity by free plan brands** in September could be because they switched to **Campaign plan.**

Plot showing the number of brands that switched from plan 0 to plan 3 and the brands that remained in the same plan throughout
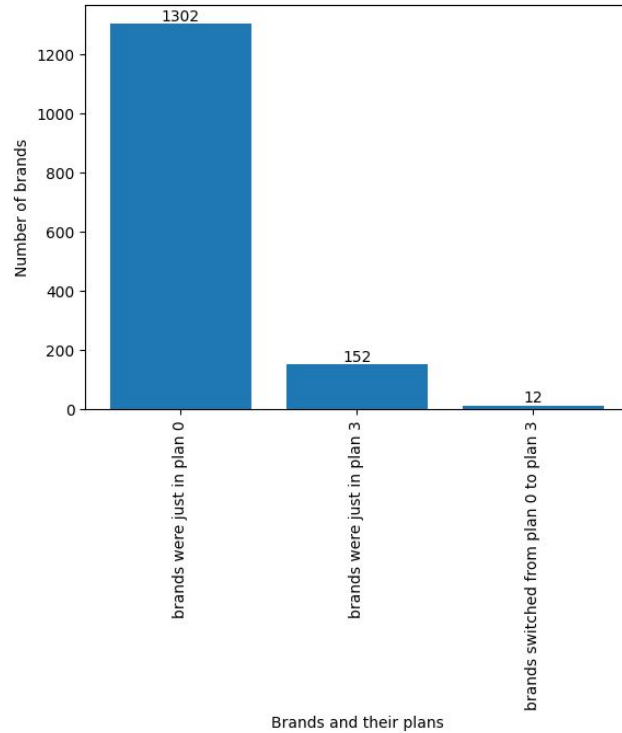
- Only 12 brands switched from Free to Campaign plan.
- As usual, the data is dominated by brands with free plan

Table showing the 12 users that switched to plan 3. Their first activity date in each plan and the time taken to switch to Campaign plan

	TABLE 6		
USER	First activity (plan 0)	First activity (plan 3)	Time taken to switch from plan 0 to 3
1	2022-08-16 13:49:16	2022-08-16 13:49:16	0 days 00:00:00
2	2022-06-02 00:28:35	2022-09-15 04:16:49	105 days 03:48:14
3	2022-04-05 01:55:49	2022-08-18 21:01:32	135 days 19:05:43
4	2022-08-12 18:16:11	2022-08-24 16:31:07	11 days 22:14:56
5	2022-01-06 07:50:12	2022-09-12 22:46:22	249 days 14:56:10
6	2022-03-17 07:47:17	2022-08-19 11:41:38	155 days 03:54:21

	TABLE 6		
USER	First activity (plan 0)	First activity (plan 3)	Time taken to switch from plan 0 to 3
7	2022-07-07 03:06:22	2022-09-13 17:57:17	68 days 14:50:55
8	2022-07-14 15:33:20	2022-08-29 10:36:19	45 days 19:02:59
9	2022-07-27 01:39:52	2022-09-12 19:00:00	47 days 17:20:08
10	2022-07-28 09:24:21	2022-08-23 11:49:14	26 days 02:24:53
11	2022-02-22 08:13:08	2022-08-31 21:39:25	190 days 13:26:17
12	2022-06-24 21:48:40	2022-09-08 18:52:33	75 days 21:03:53

Table 6 observations

- If you notice, the first activity of all the 12 users that switch from plan 0 to 3 is in either **August or September**; indicating that the users upgraded plans during peak months (Campaign plan is more active during these months)
- Some of the brands' first activity in their **free plan** was in the **first half of the year**. It could be that these brands were **actively searching and exploring** features during these early months before deciding to upgrade.