# Yapper User/Team Analysis

August 2023

Candidate: Brett Feeley
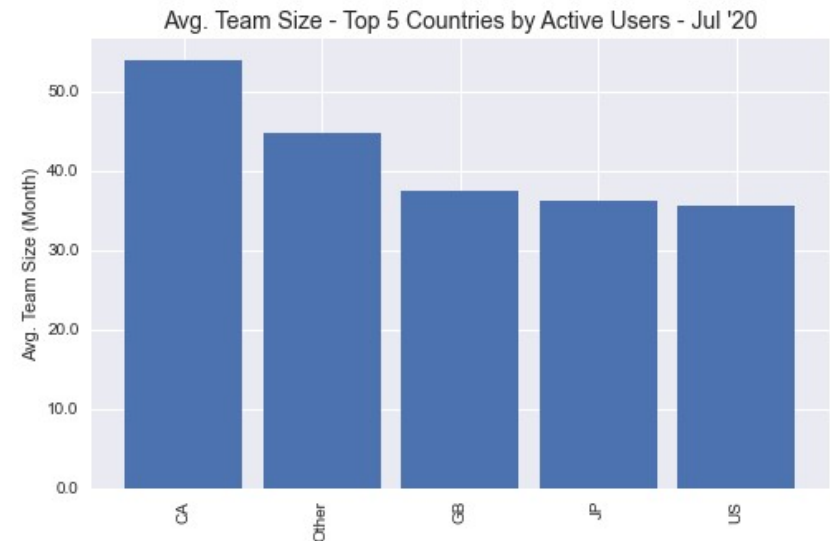
1: Industries with Most Active Users

- Methodology: Active user counts were totaled by industry over the 7 days from 7/25/2020 (Sat) to 7/31/2020 (Fri) to have a complete week of data. This total was then divided by 7 to give an average daily active user figure for each industry.

- The most active user bases by industry ranked by most to least daily active users (DAU) are:
  - **#1) Technology – 48% of DAU**
  - **#2) Professional, Scientific, and Technical Services – 19% of DAU**
  - **#3) Education (Non-Commercial) – 7% of DAU**
  - **#4) Retail Trade – 5% of DAU**

- Together, the above industries represent close to 80% of all users.
  - The rest of the industries present in the analysis represent <5% of daily active users each

Daily Active Users by Industry (7/25-7/31 Avg.)
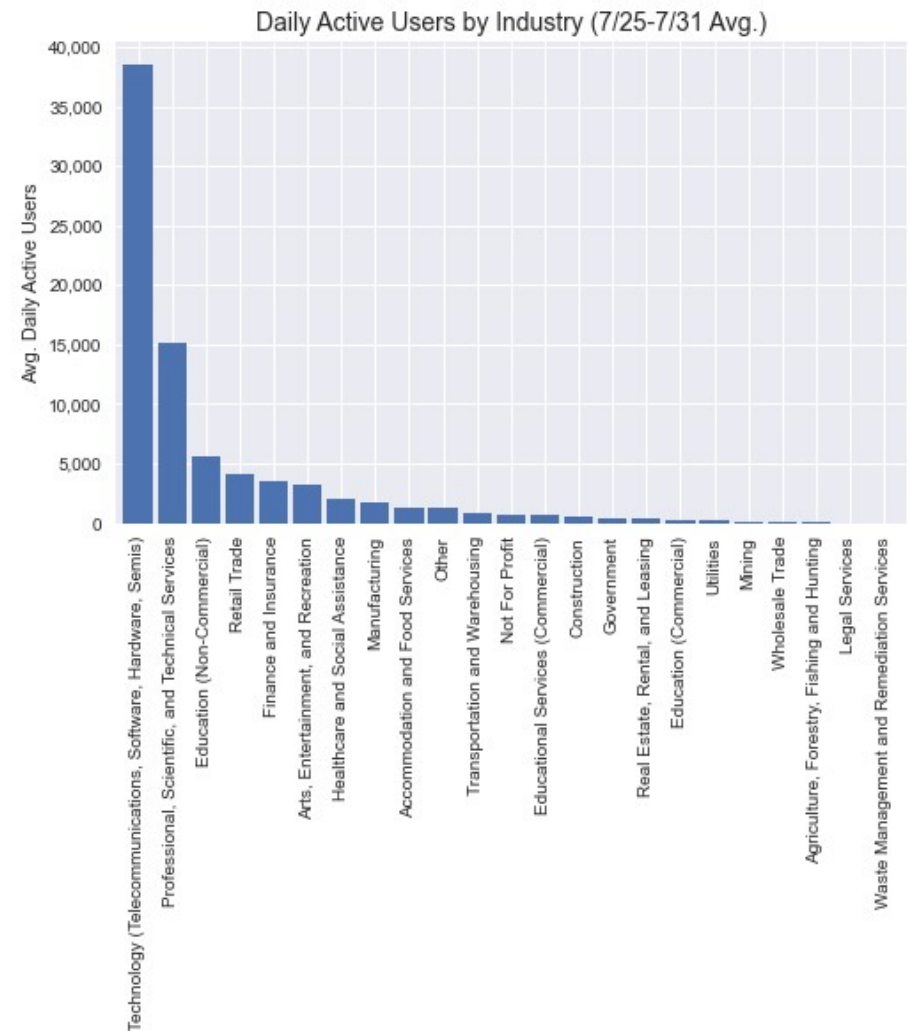
2: Average Team Size of Top 5 Countries (by Active Users in July 2020)

- Methodology: Using July 2020 data, get a distinct count of teams present by country for the month to find top 5 countries. After ranking the countries by team count, find the average team size within those countries by averaging active users.

country	countd_teams ▼	avg_team_sz	teams_cnt_rank
US	1130	35.6142	1
Other	341	44.7213	2
GB	135	37.5013	3
JP	126	36.0837	4
CA	92	53.8822	5

- **The United States has the most teams by count (1,130)**
- **Canada has the largest teams by average size within the top 5 countries by teams**
  - Some countries outside the top 5 have teams of 100+ people, and may be using the product inefficiently (e.g. RU)

- **There are many teams present in the 'Other' category** that would likely benefit from being broken out into component countries for further analysis

Avg. Team Size - Top 5 Countries by Active Users - Jul '20

3: Usage Trends

- Using a consistent cohort of teams active during the analysis period, a few conclusions are evident:
    1) **Daily active user growth is a function of growth in team size** (and not additional teams) as a consistent team count was used
    2) **Daily active users/messages seem to plummet during the holiday season** through January. Likely a result of worker PTO
    3) **Users and messages per user rose quickly following the pandemic**, but have started to revert after peaking in March
    4) **Daily active users have started to decline** toward pre-pandemic totals following the March increase

Avg. Team Size - Active Users

Messages Sent Last 7 Days (mm)

Daily Active Users

Further Investigation Needed – See Final Slide

4: Other Insights (including cleanups needed)

A number of anomalies were noticed during the exploration of the Yapper data. The issues and actions taken (if any) are noted below:

- <u>Teams were starting to be lost after 2/5/2020 at a steady pace</u>. 18 teams had to be removed from the analysis as they were not present for the entire analysis period – a consistent cohort was used for all assignment answers
- <u>Some teams had zero active users</u> for the entire analysis period and were removed from the analysis for all assignment answers
- <u>3/13/2020 is a missing date</u>, and could affect 7d message counts if date lookbacks were used instead of window/row lookbacks – no change made to data, but noted
- <u>Industries #11 and #12 were consolidated</u> as they look to be representing the same industry (Healthcare vs. Health Care) – #12 not in data and does not affect current analysis. Cleaned up map in my script
- <u>No industry mapping was present in the map for Industry #1,</u> but does not appear in usage data or affect current analysis – no change made
- <u>Team sizes in RU, IN, DE are very large for some reason</u> - may not be breaking up their teams well
- <u>The 'Other' country</u> has a very large team base and should be broken out into component countries if possible
- <u>Holidays seem to be causing issues</u> with the consistency of usage – no changes made to source data, but noted
- <u>Pandemic-era usage seems to be causing anomalous results</u> – a longer analysis period should be used

5: Further Points of Analysis Needed

1) **Find out why unique team counts started dropping off in early February**

   - Look at contract/renewal periods to see if these were natural ends or non-renewals
   - Use some quantitative measure/model of consumer health to figure out if these losses were predictable by collecting/creating additional usage insights about each team (e.g. health score etc.)

2) **Separate 'Other' country label into component countries**

   - Given that this 'country' represents a large portion of teams (340+), it would be good to know if it is one country or a group of smaller ones
   - If it is indeed one mislabeled country, then it is an important insight to know for customer support

3) **Investigate Orange Circled Areas on the 'Usage' Trends Slide**

   - March spike is most likely Covid, but may be a coincidence with a marketing campaign, other promotion, etc.
   - Find out whether Yapper did anything 'exogenous' to cause this spike in usage
   - Holiday periods seem to be showing large troughs which could be outages or natural outflow of use during holidays
   - Abrupt drop in usage noted around 4/23-4/24 '19. Further investigate this period for what could be causing the anomaly