SafeToNet

Sample Reporting

SafeToNet Risk Message & Behaviour Analytics

July Report: 06/07/2020 – 02/08/2020

Exploratory Analysis: 09/03/2020– 01/11/2020

© SafeToNet, Nov 2020 (updated report)

1. Key Monthly Insights (July 2020 – Aug 2020)

1.0 This report is a sample report of the data generated by the SafeToNet safeguarding solution. All data is anonymized through our privacy by-design approach.

1.1 Despite an initial increase in messaging activity in the first week of July, the past month has seen an overall continued decline of messaging activity levels across the entire userbase (see 3.1).

***Commentary**: This trend may continue to be associated with ongoing relaxation of lockdown measures. In addition, it is noted that children in most DE and UK schools entered their summer holidays during the month of July. A combination of both factors may be related to the steady decline of activity, as children spend increased time in face-to-face meetings with peers.*

1.2 Again, overall messaging activity is declining at a faster rate than messages that carry risk, with levels of risk messaging remaining fairly consistent over the month of July. However, exceptions are noted with both an increase of risk detected at the start of the data range, and a decrease observed across the weekend prior to the start of summer holidays (see 2.2).

1.3 There was a strong positive correlation observed between cyberbullying and sexting messaging rates in the past month (r = 0.982) meaning that increases in the use of aggressive language were correlated with increases in sexting.

1.4 Girls in both the UK and Germany consistently show the highest levels of messaging (see 3.2). Girls over the age of 10 are particularly at risk (see 5.2 & 6.2).

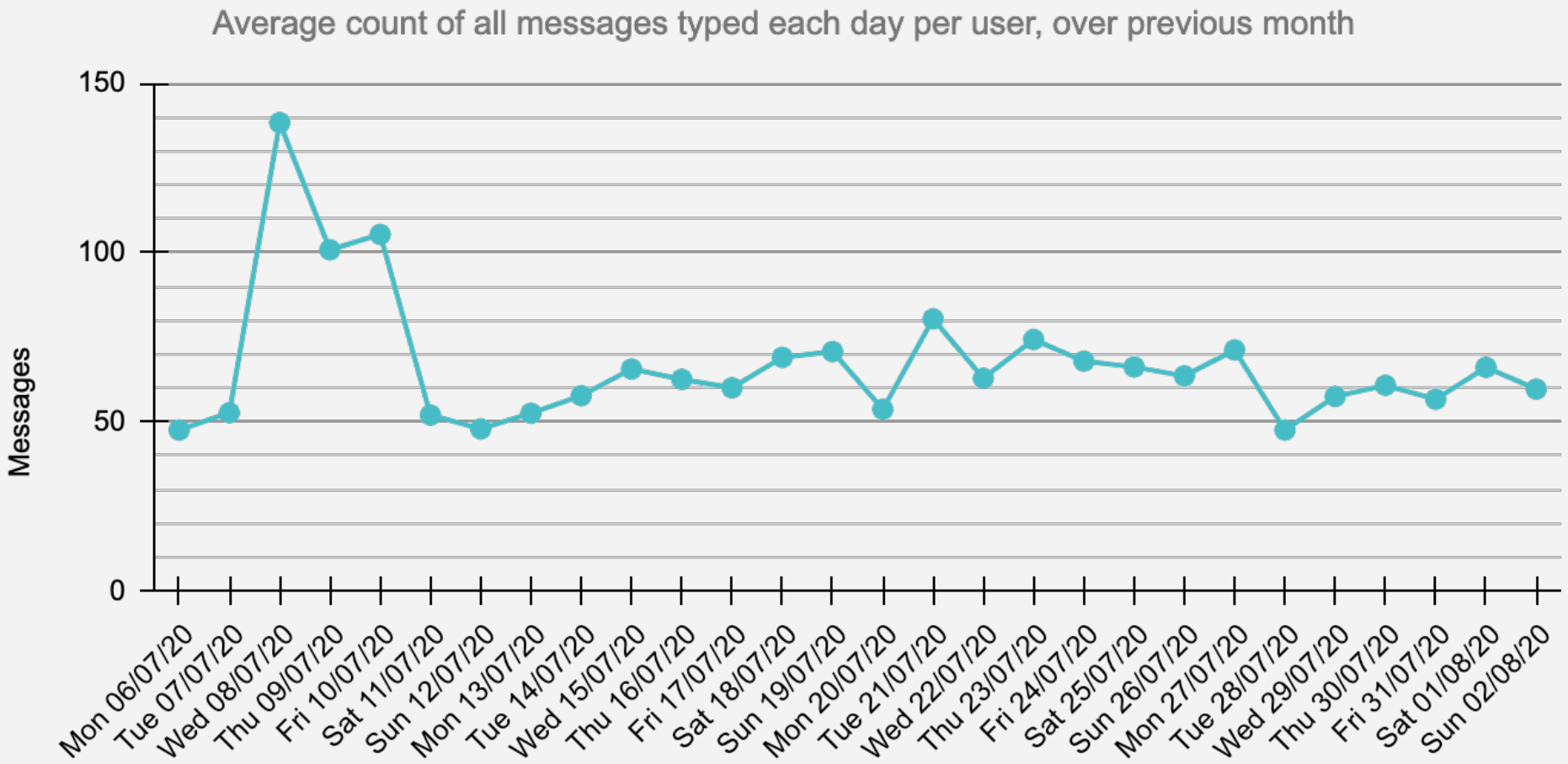
1.5 All types of messages are being sent throughout all times of day and days of the week. However during the first full week of July when children were still in online lessons, the majority of all types of messaging occurred in the evening (see 3.4).

***Commentary:** In the buildup to school holidays, children become more distracted from their online classes and spend more time on their devices during the day instead of waiting until the end of the school day.*

1.6 Cyberbullying tends to start at lower levels at the beginning of the week and increases steadily throughout, with Thursdays as the most common day for levels to peak (see 5.4). Cyberbullying on Fridays, Saturdays and Sundays takes place at a consistent rate. Sexting occurs at a consistent rate throughout days of the week, with highest levels observed during the daytime (see 6.4).

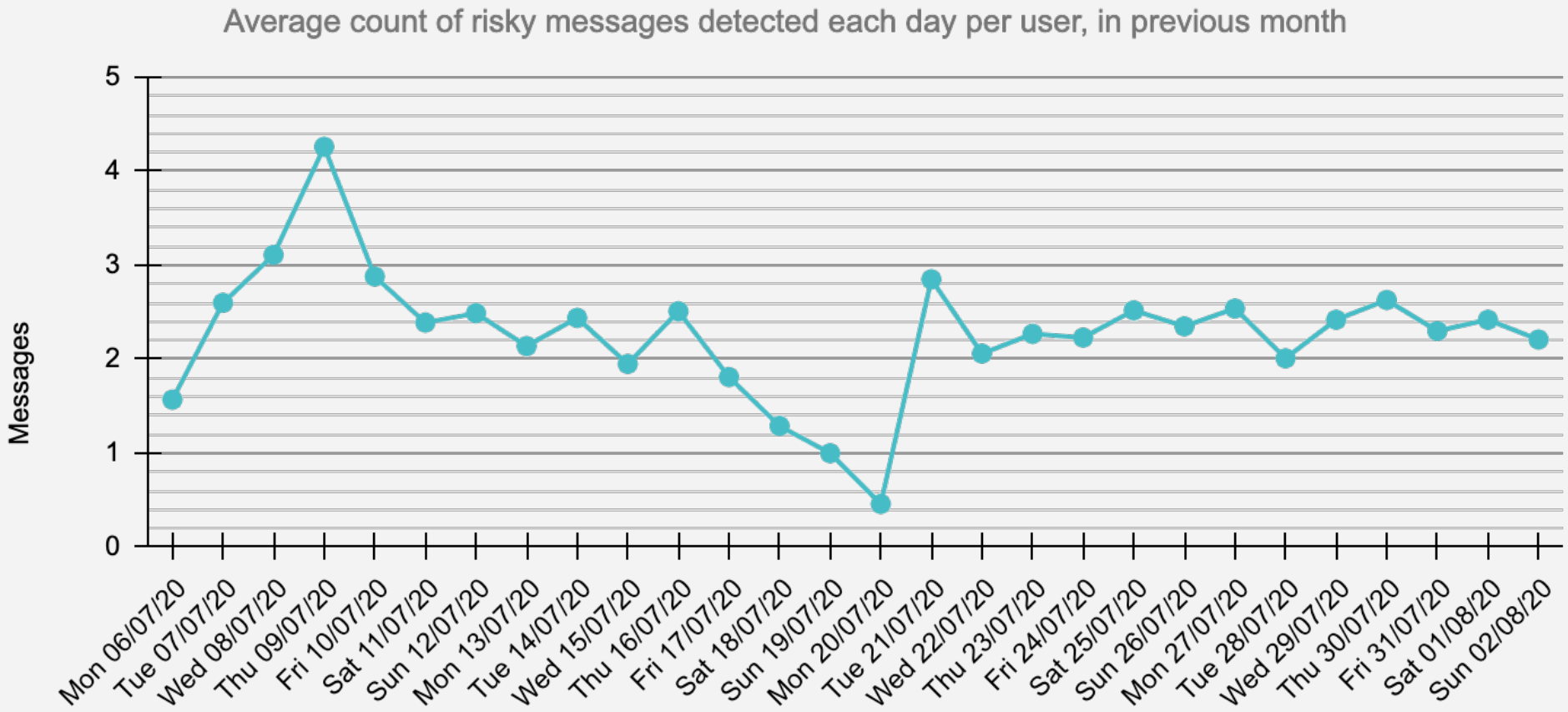
1.7 When looking at levels of risk on an app basis, Instagram and Snapchat moved to the top of the list this month as users typed more risky messages on these social media apps than any other apps on average. However, WhatsApp had the highest percentage of users who used the app to type risky messages compared to users of all other apps (see 7.1). Therefore it appears that users typed risky messages at a higher intensity level on Snapchat and Instagram, while WhatsApp was used by a larger proportion of its userbase to send risky messages, albeit at a lower intensity.

2.1 Up close - Daily trends 1: activity



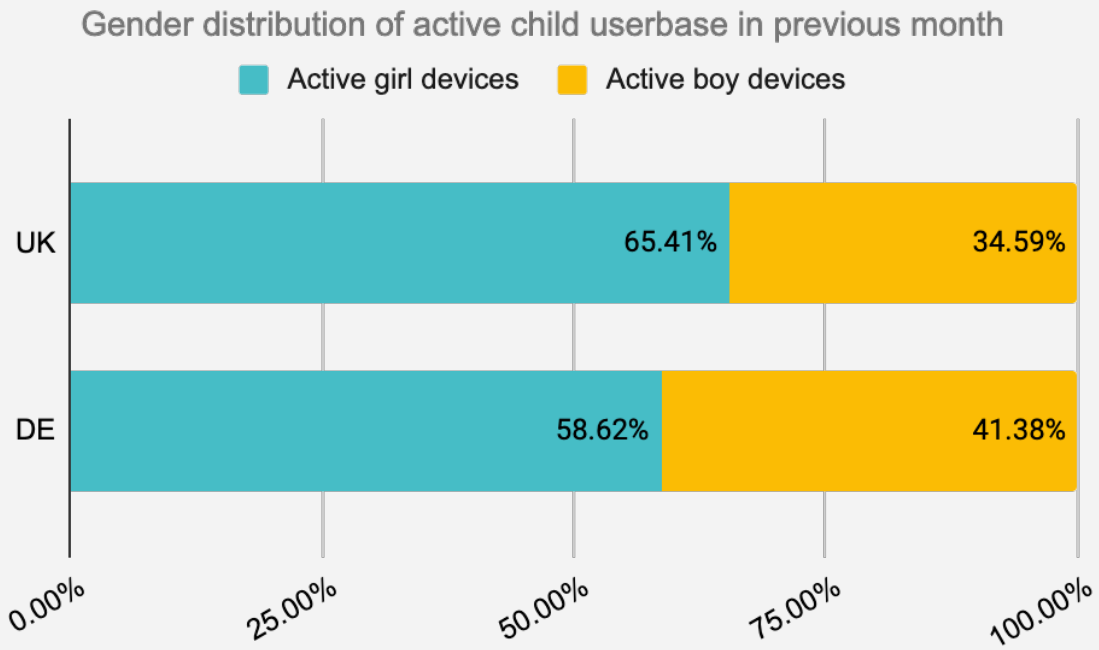
Note: Non-risk messaging levels rose sharply on Wednesday 8th July, and remained at an elevated level for the following few days. During this time, intensity levels of risk messaging also increased. Despite this anomaly at the start of July, overall messaging levels remained fairly consistent throughout the remainder of the month, with overall activity showing a decrease when compared to the previous month.

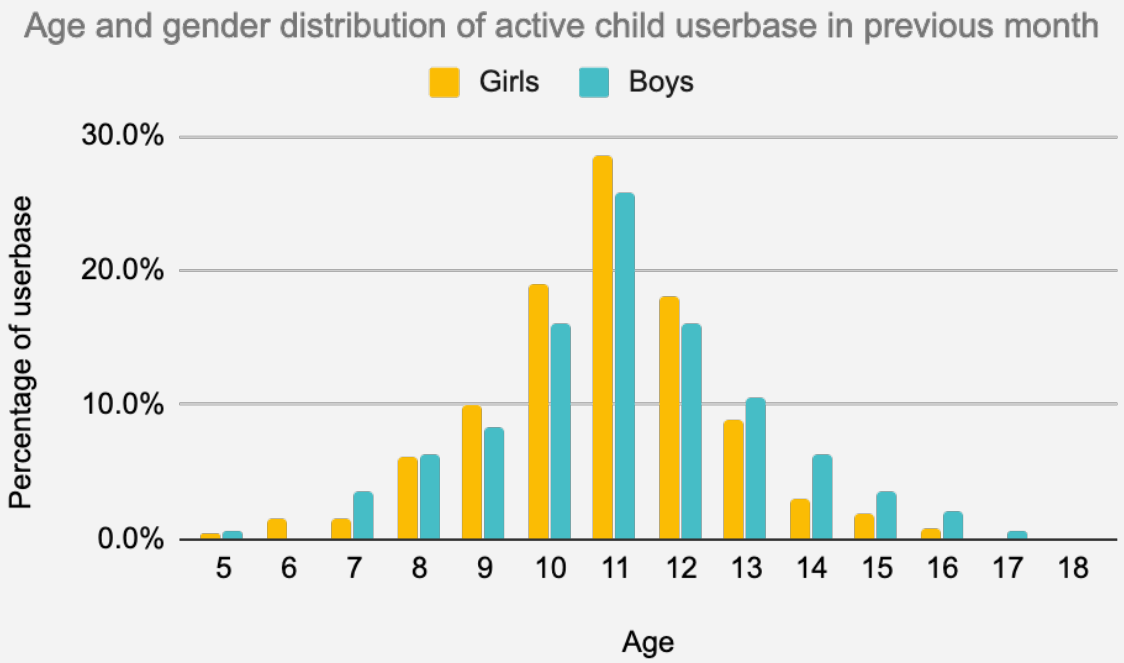
2.2 Up close - Daily trends 2: risk detection levels



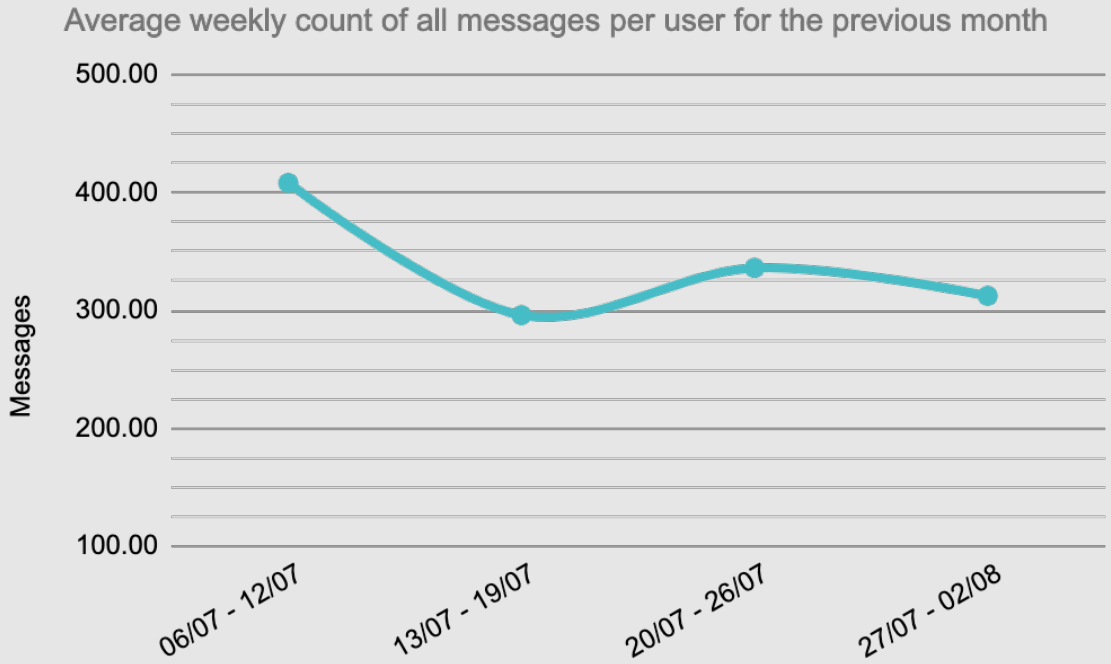
Note: Risk messaging (and non-risk messaging) was at it's lowest point on the 20th July as schools closed for summer holidays. Levels of all types of risk decreased on this day, followed by a sharp increase in cyberbullying and sexting the next day. These higher levels of risk detection remained consistent for the duration of the month. The peak day for risk messaging (predominantly as a result of increased cyberbullying messaging) occurred before holidays began on Thursday 9th July and a day after the sharp increase in overall activity levels. Risk messaging at this time occurred predominantly in the evening, in contrast to the rest of the data range.

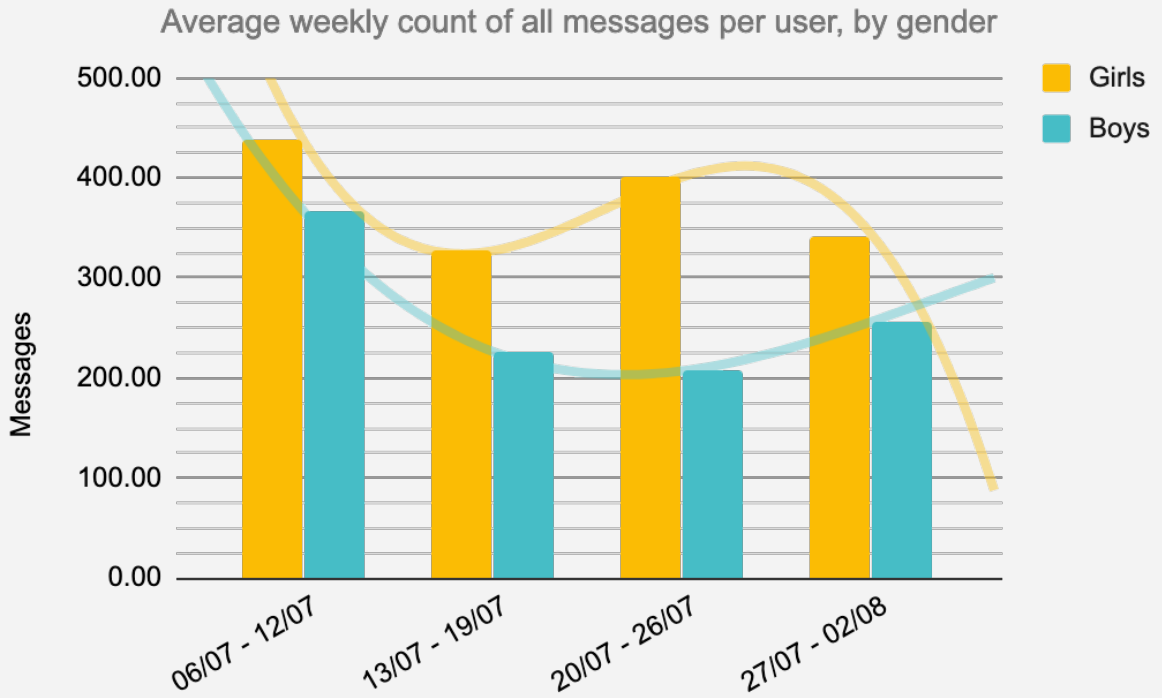
2.3 Overview demographics





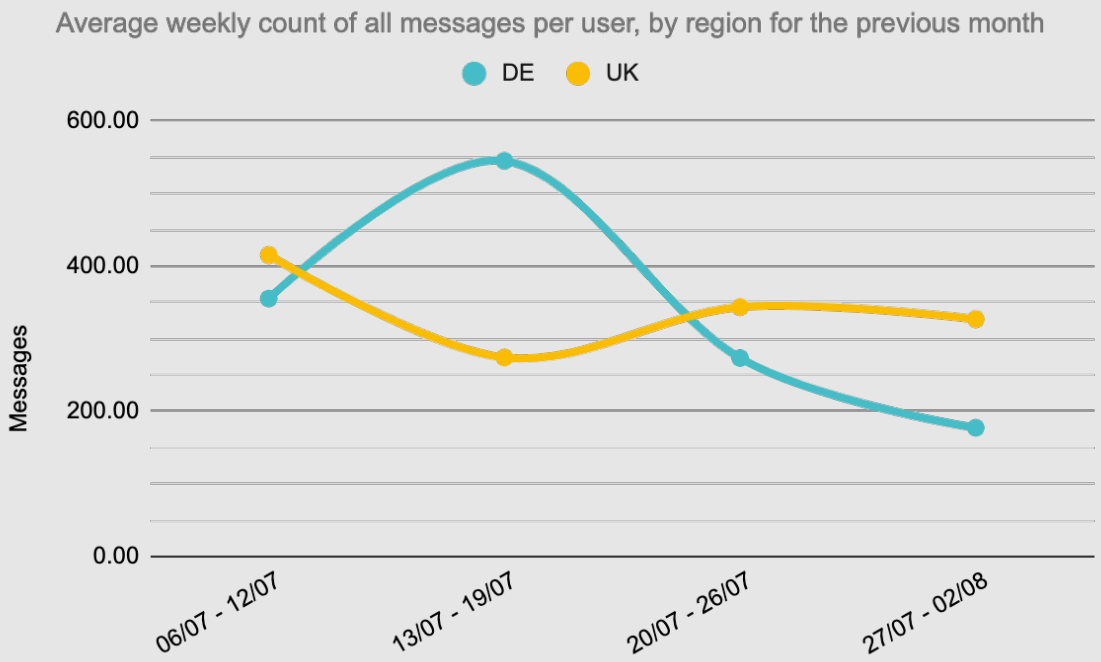
3.1 Messaging activity rates 1: Overview and gender

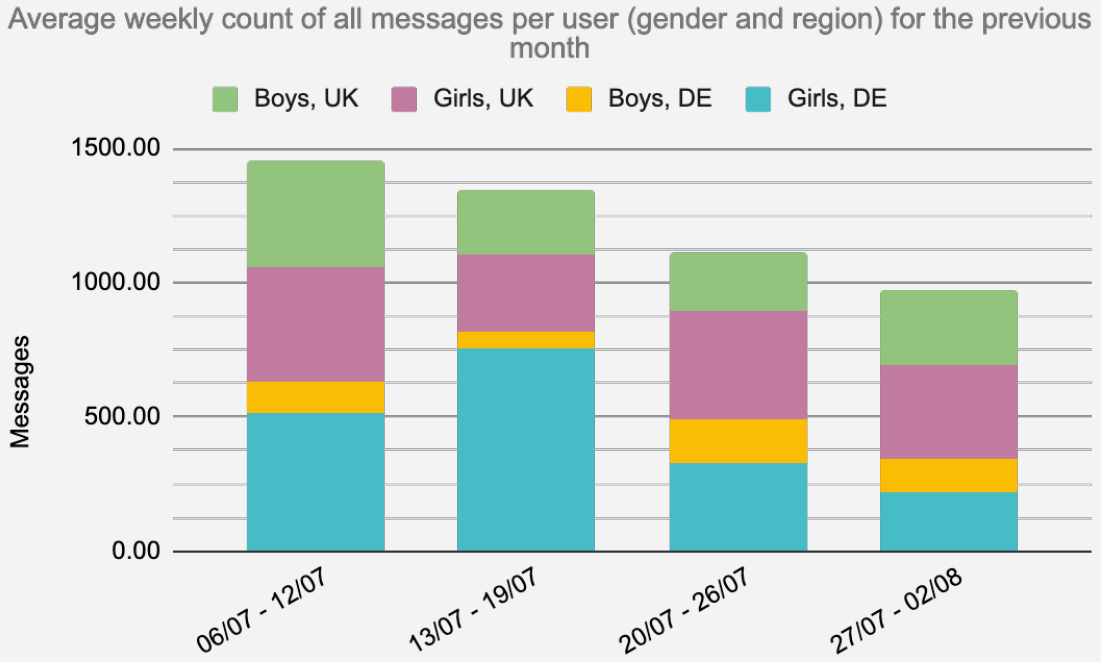




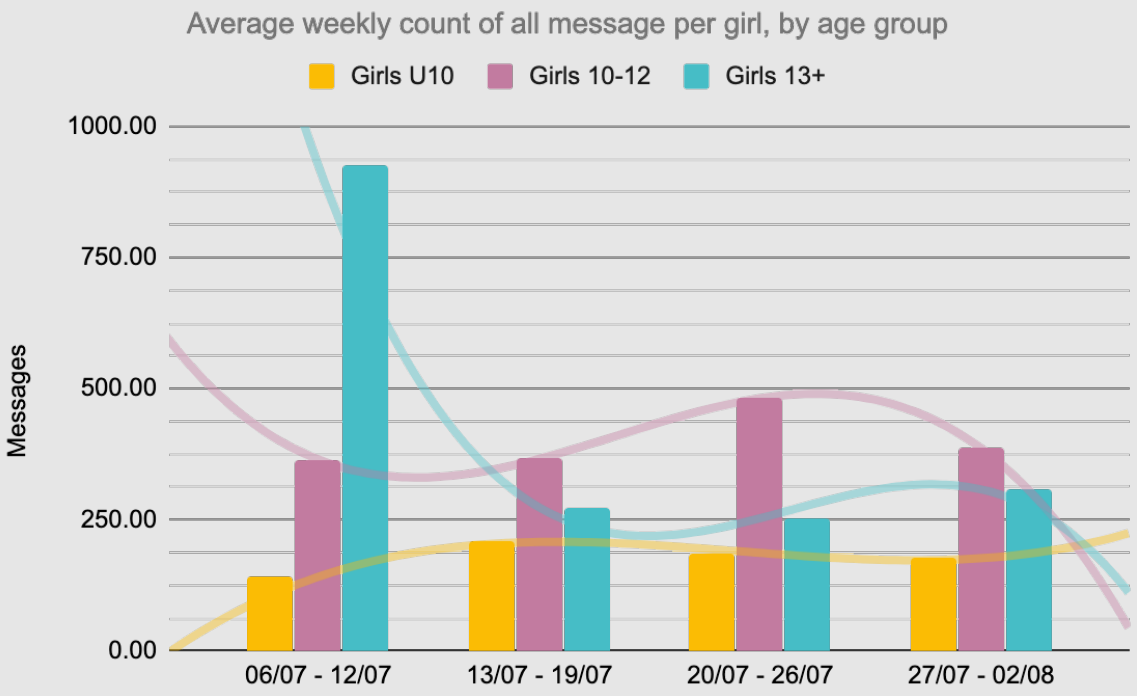
CONFIDENTIAL

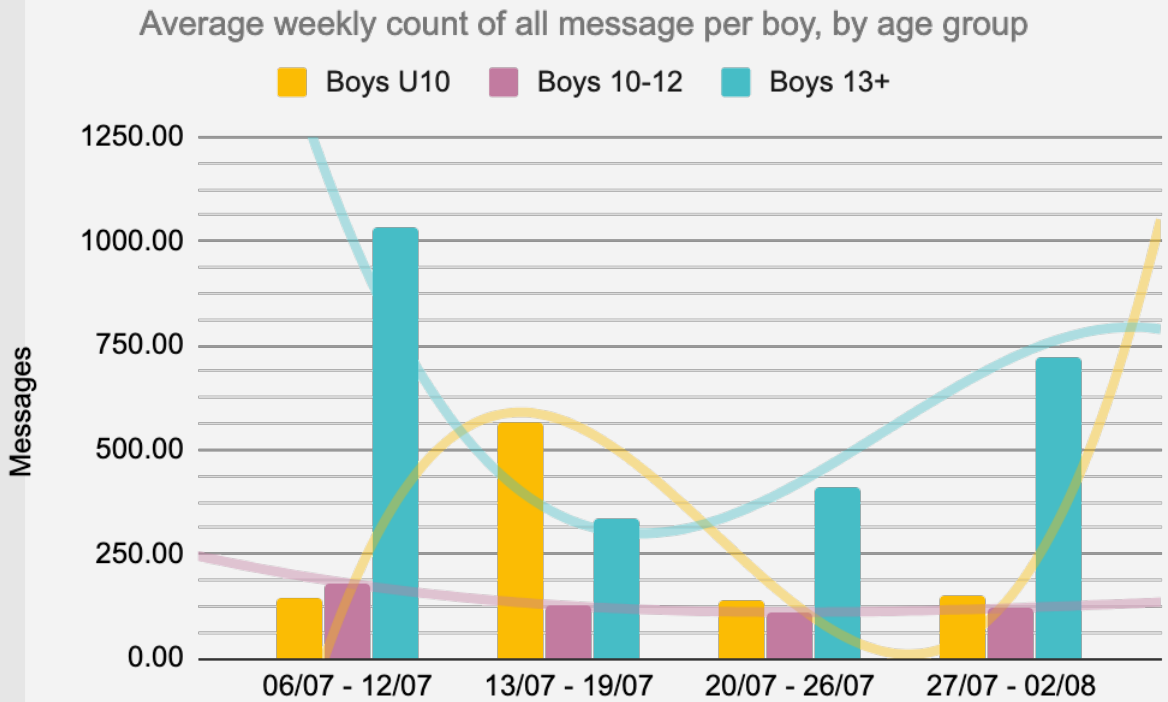
3.2 Messaging activity rates 2: Gender and region





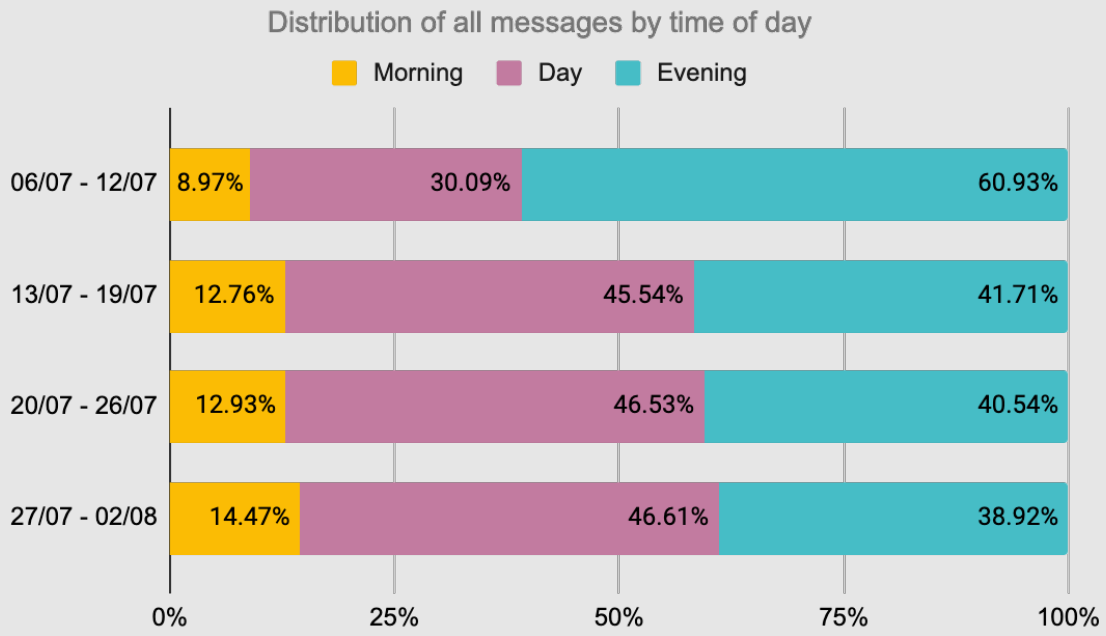
3.3 Messaging activity rates 3: Gender and Age

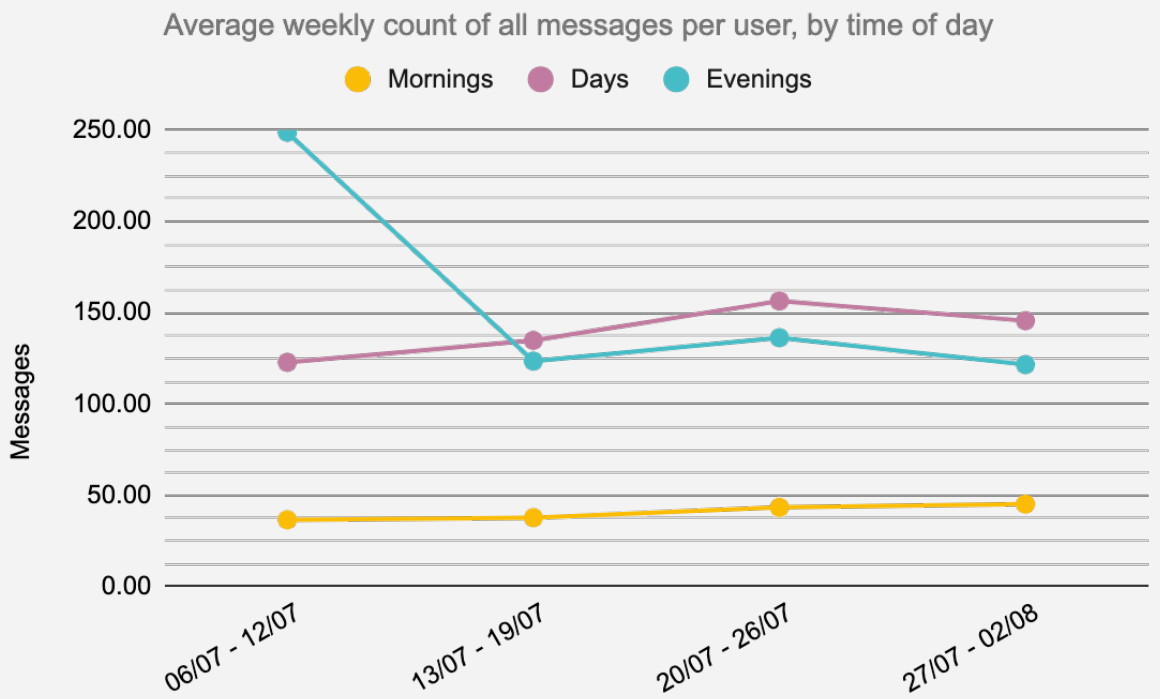




CONFIDENTIAL

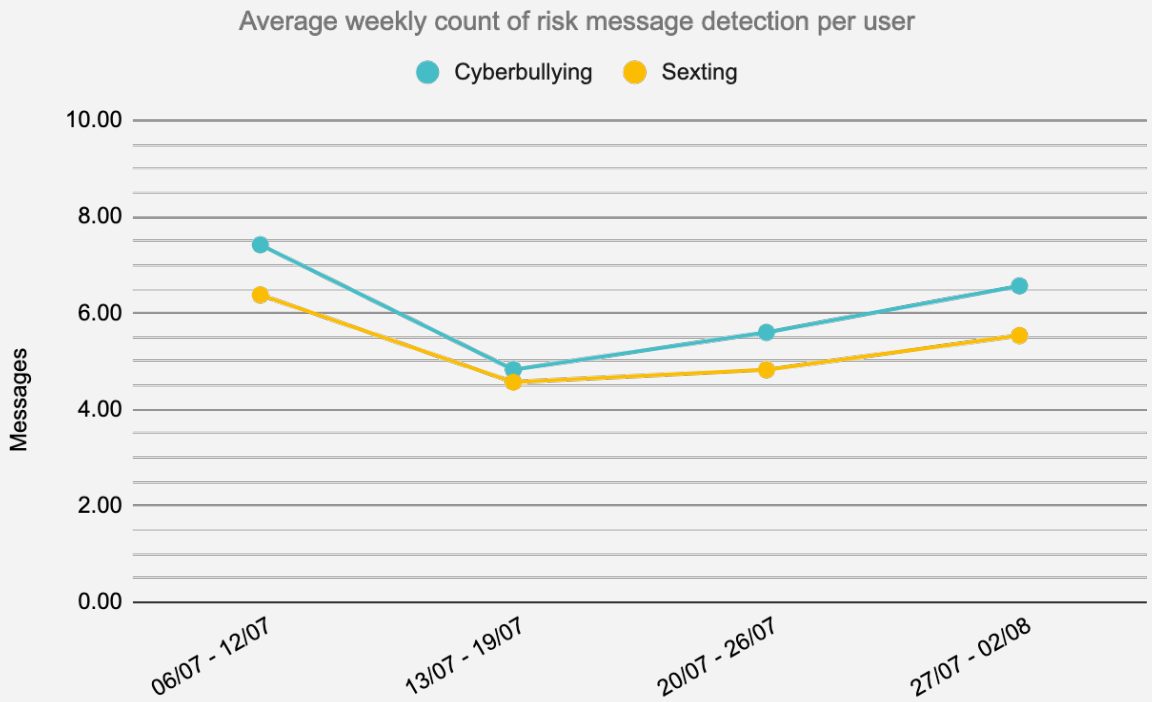
3.4 Messaging activity rates: Time of the day

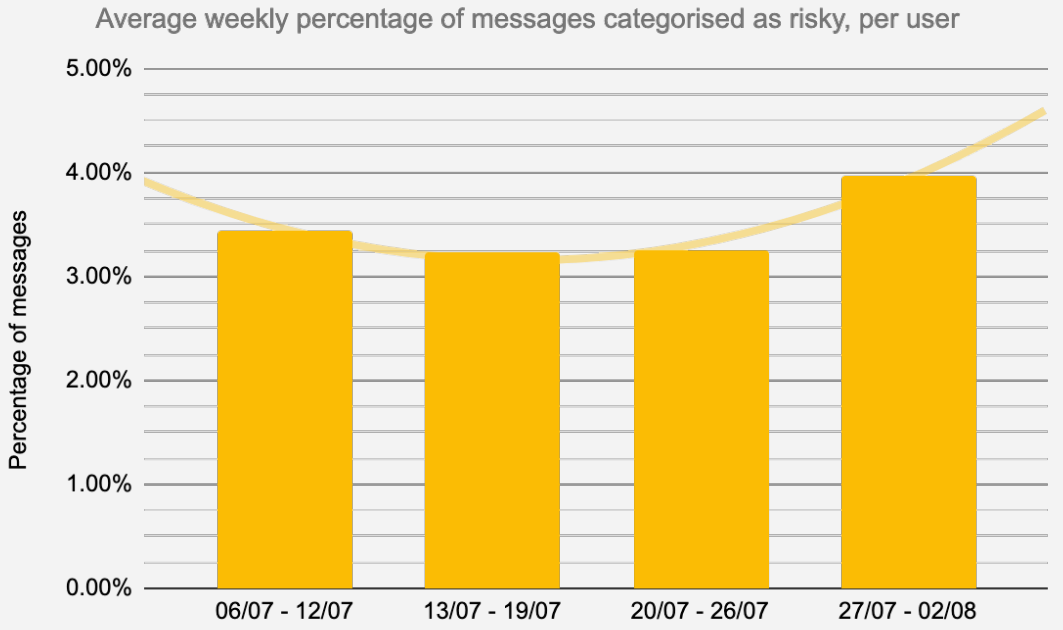




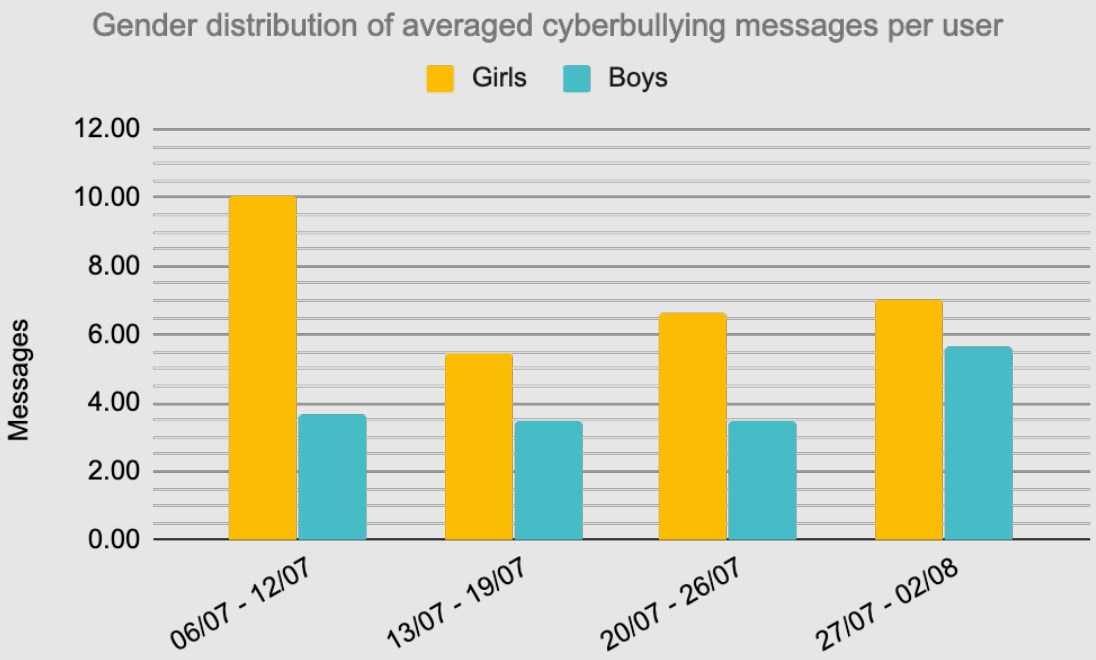
CONFIDENTIAL

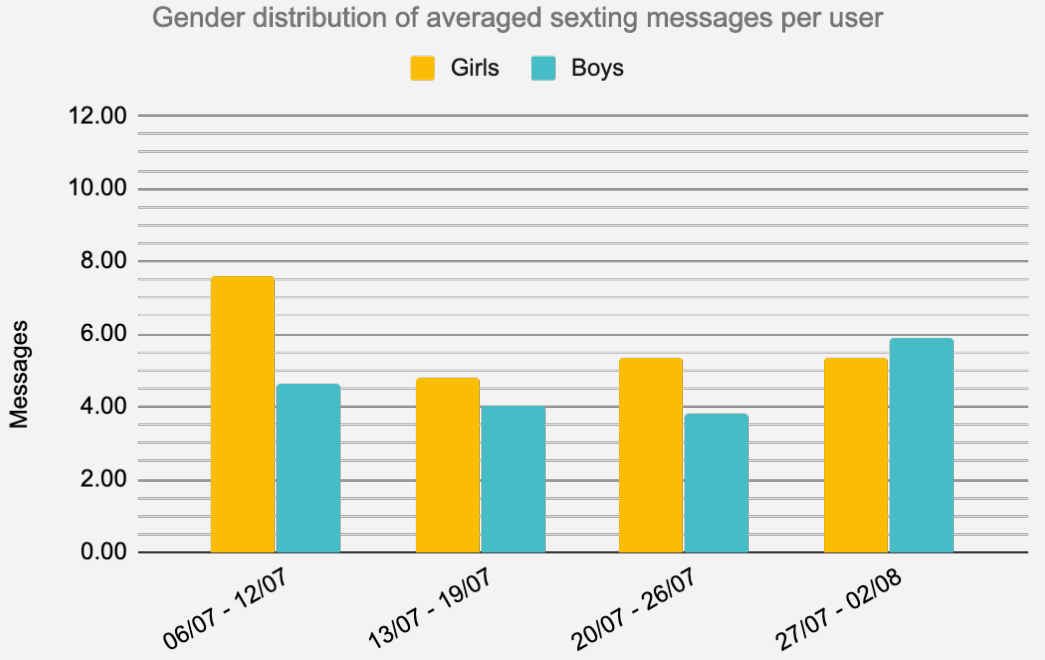
4.1 Risk detection levels: Overview



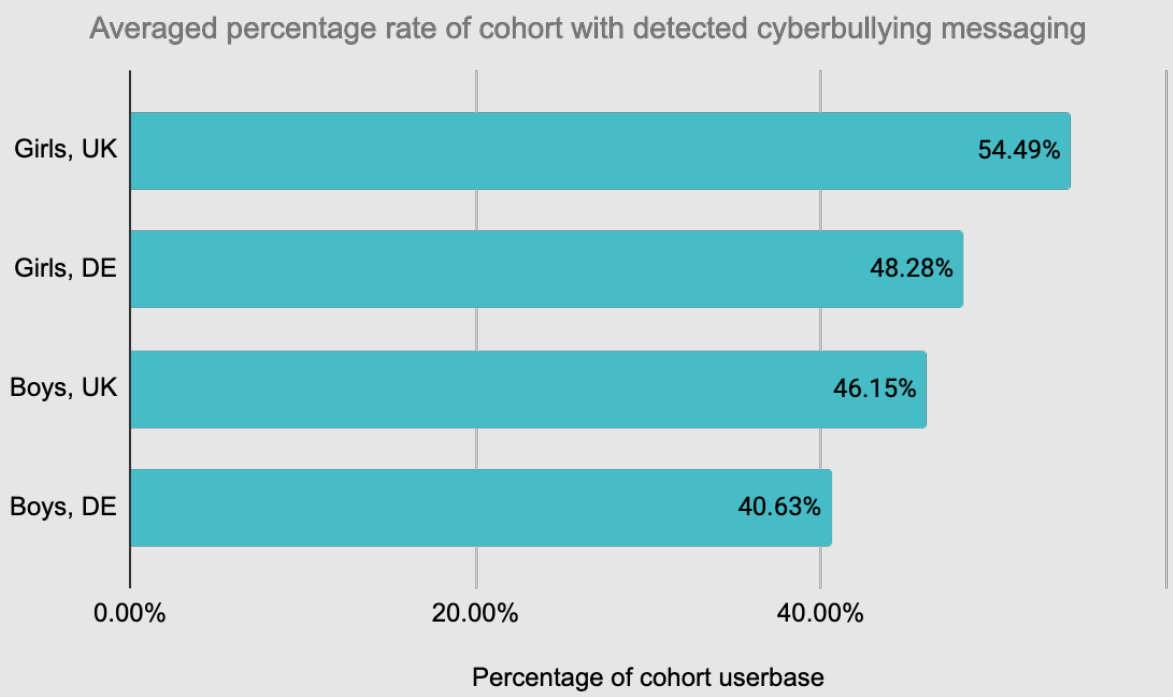


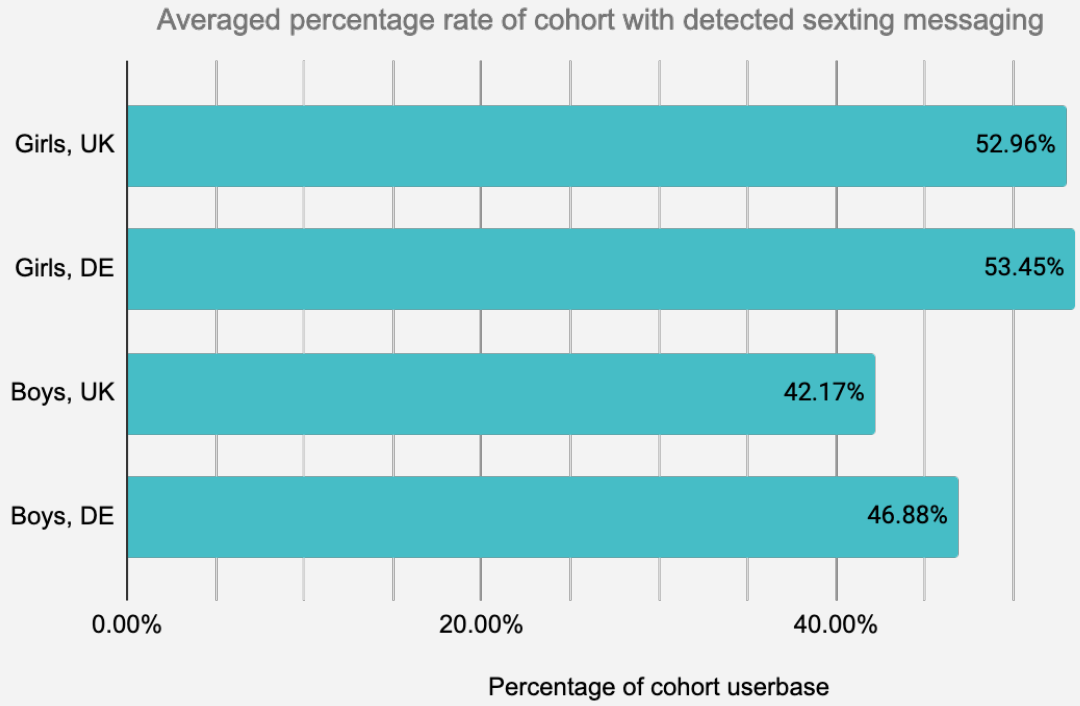
4.2 Risk Detection levels: gender distribution by type



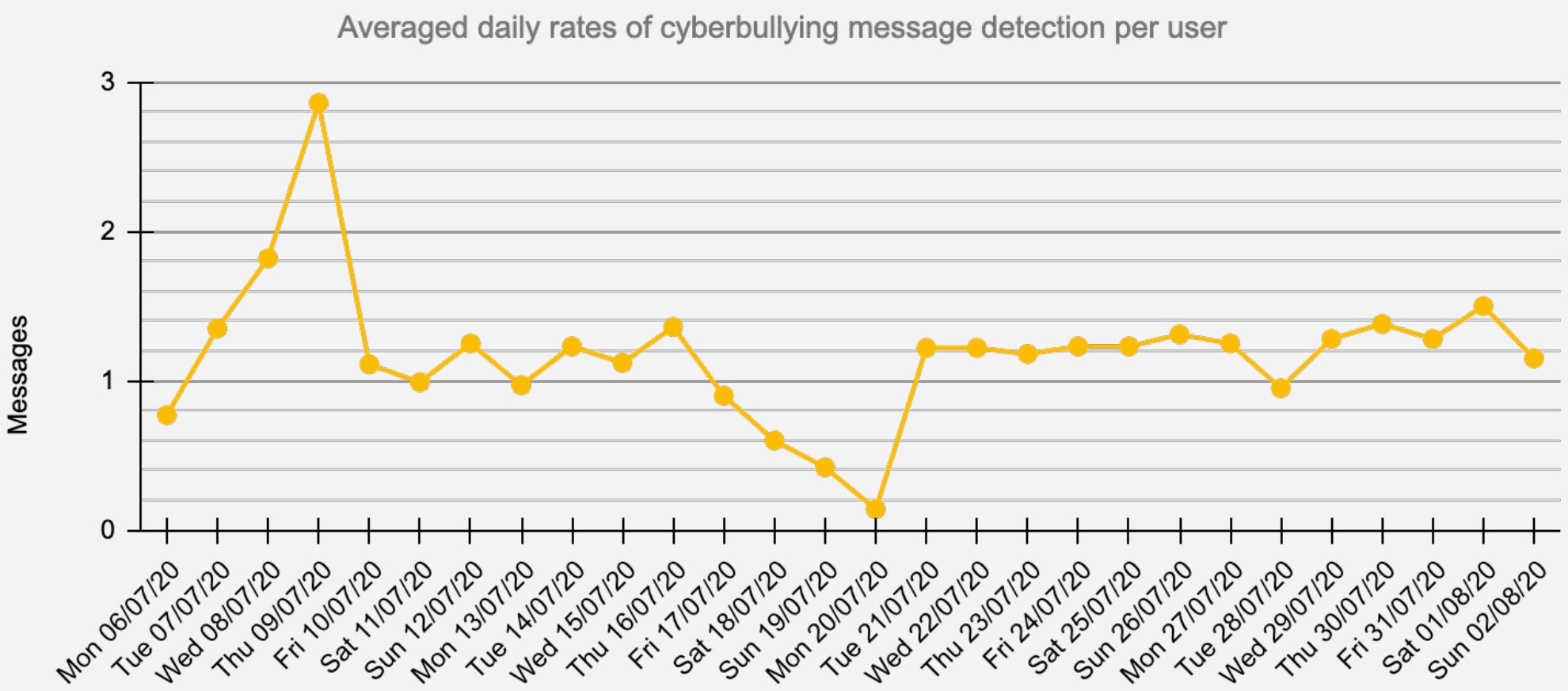


4.3 Risk detection: Prevalence by cohort analysis



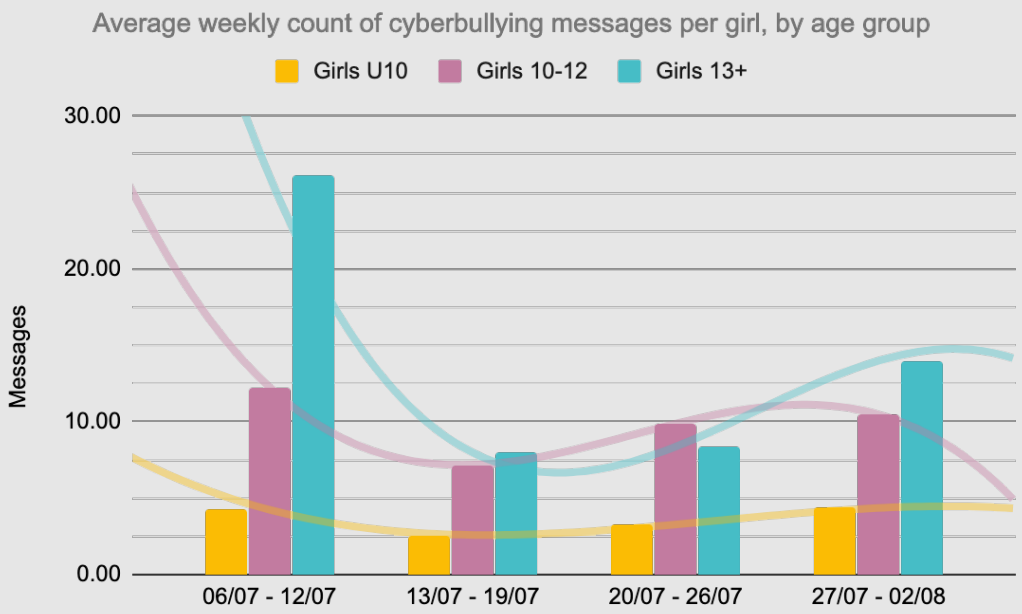


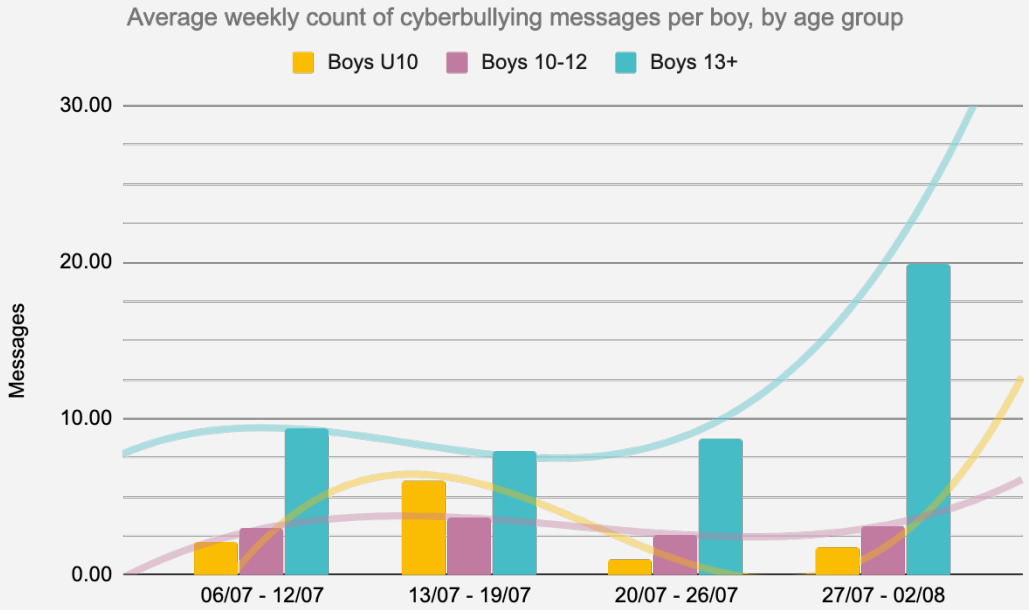
5.1 Cyberbullying up close: Daily detection rates



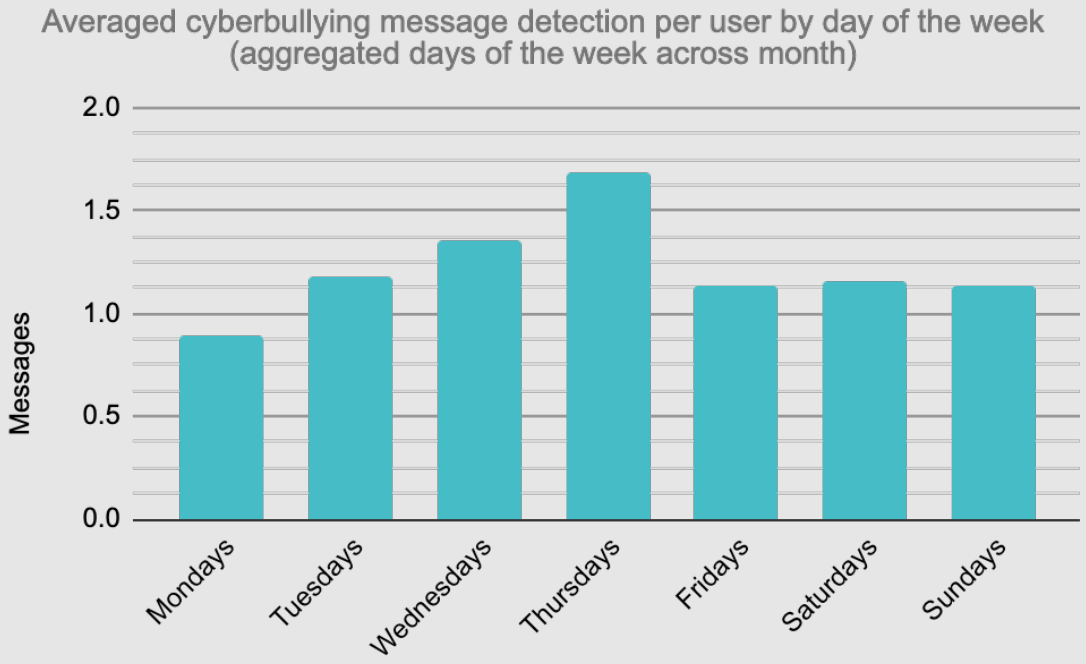
Note: Cyberbullying tends to start at lower levels around the beginning of the week and rises throughout the week, with highest levels observed on Thursdays during term-time. Cyberbullying increased dramatically on 9th July followed by an immediate decrease the day after. This increase appears to be attributed to girls, with aggressive messaging observed to a high degree on Snapchat and WhatsApp. Cyberbullying decreased steadily in the build-up to the summer holiday, falling to it's lowest point on the 20th July, with a large majority of users displaying little to no aggressive language. However, levels increased and have remained consistent for the remainder of the month.

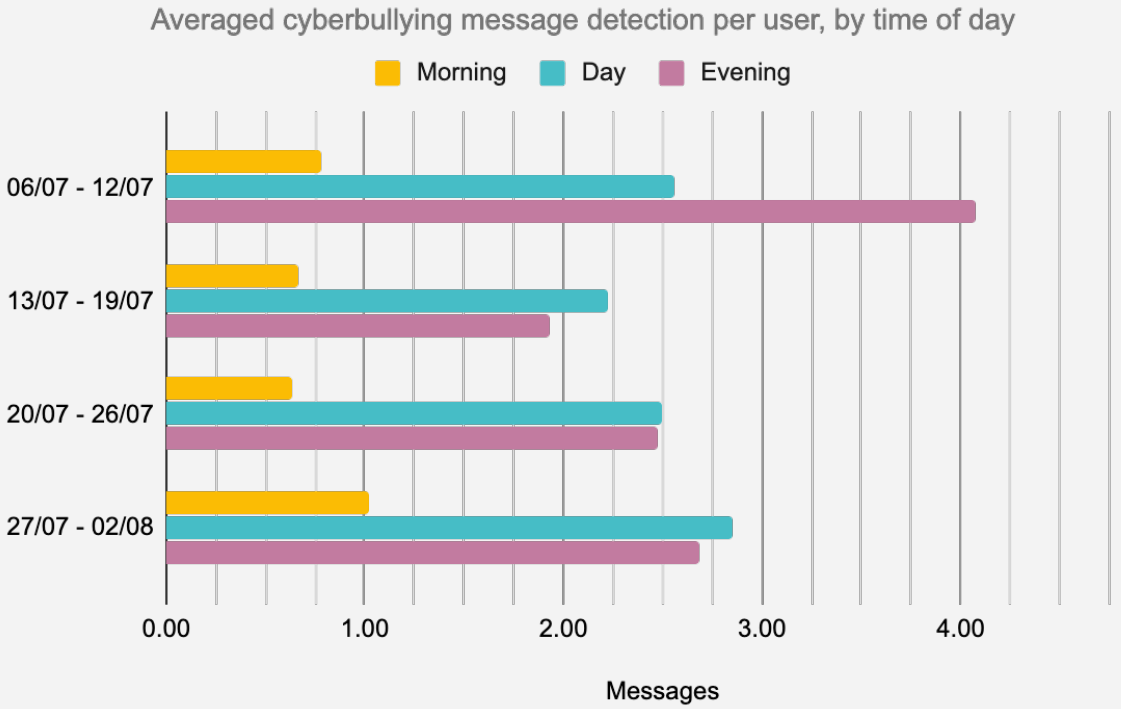
5.2 Cyberbullying: Gender and age



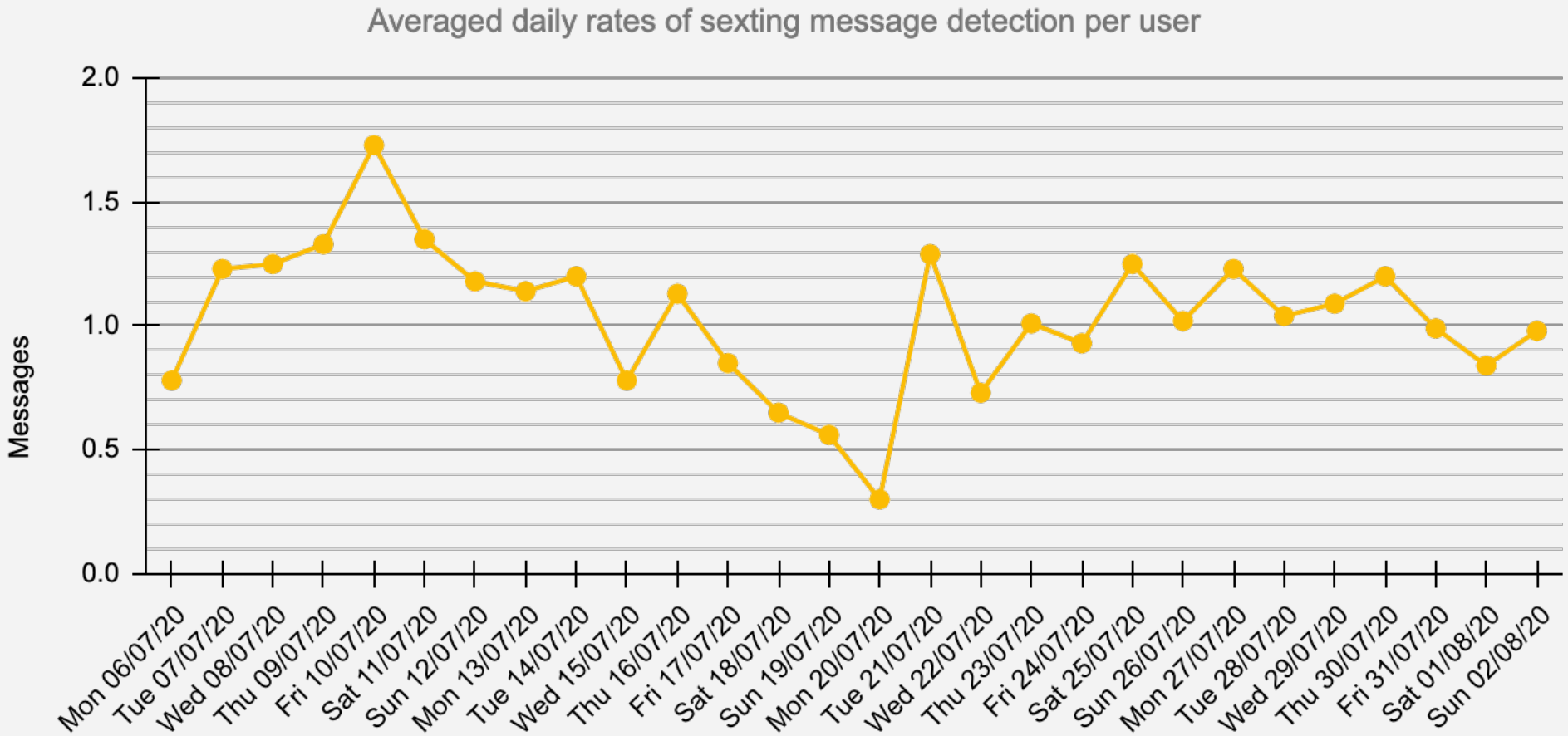


5.3 Cyberbullying: Days of the week and time of the day



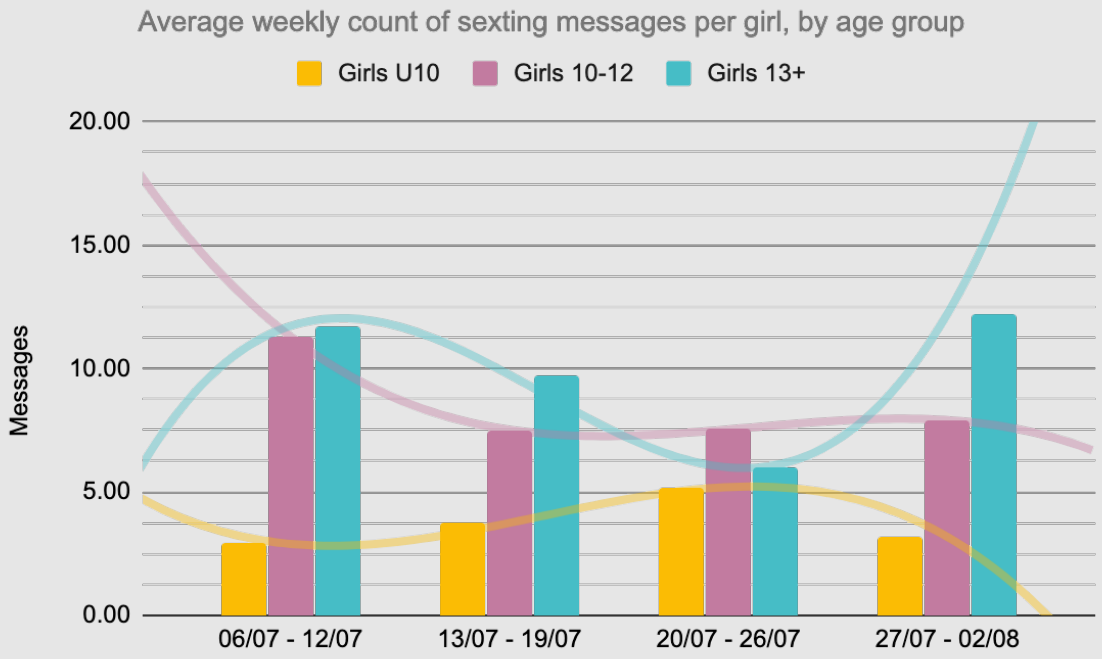


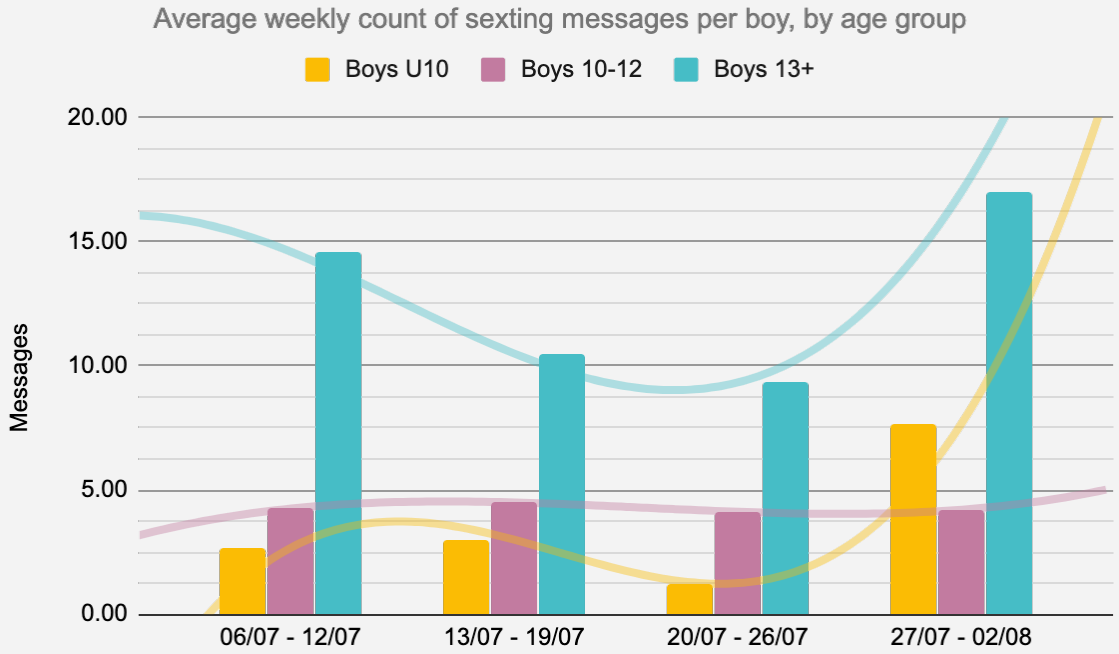
6.1 Sexting up close: Daily detection rates



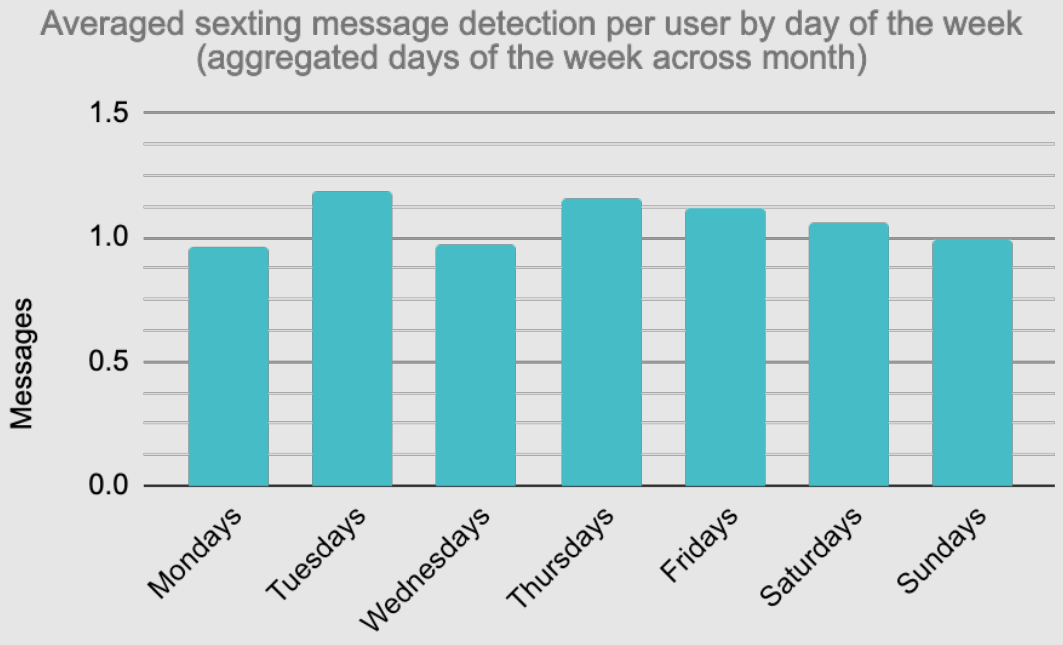
Note: Overall levels of sexting message detection continue to be higher than those observed prior to lockdown. As levels of neutral and cyberbullying messaging increased around the start of the month, sexting also increased with a high concentration of sexting messages observed on Tuesday 7th and Friday 10th July (in the evenings on WhatsApp). While levels of sext messaging increased around the start of the school holidays, overall rates throughout the remainder of the month remain lower than at the start of the range and previous months. During this time children were in online classes and under tighter lockdown restrictions, and less likely to meet peers face-to-face.

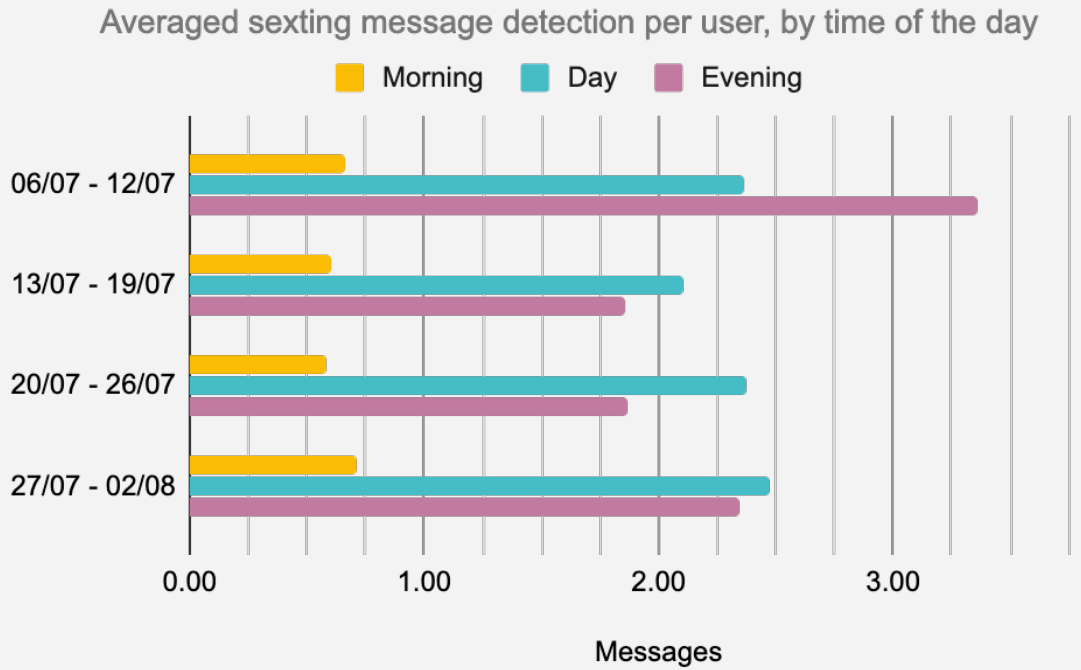
6.2 Sexting: Gender and age



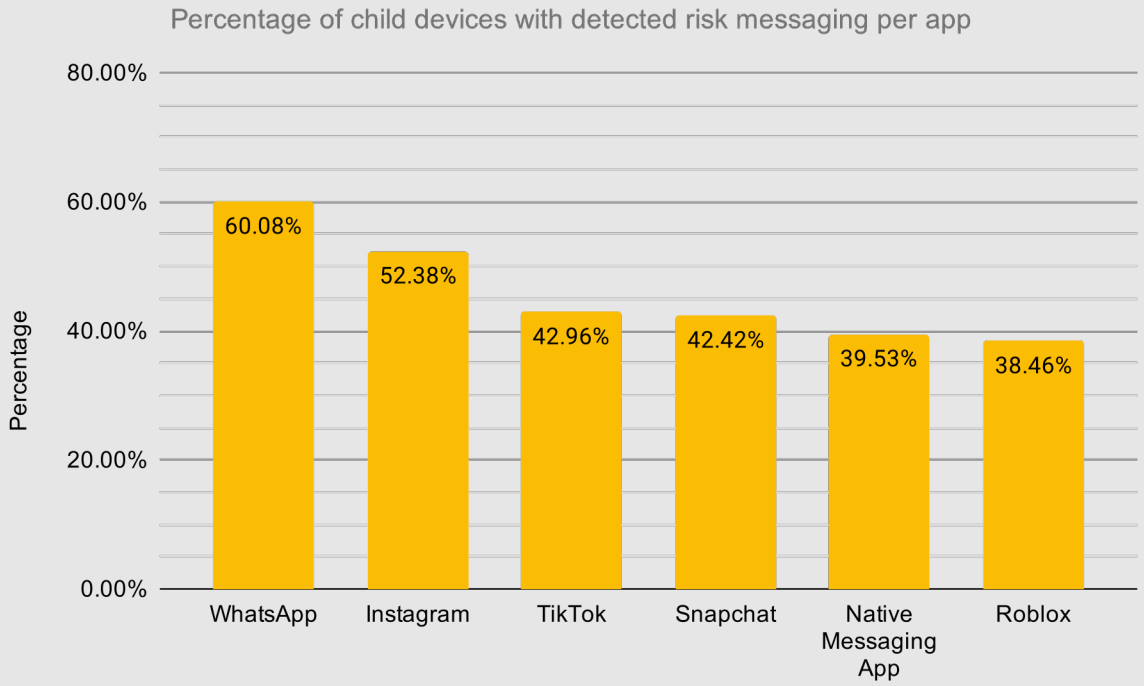


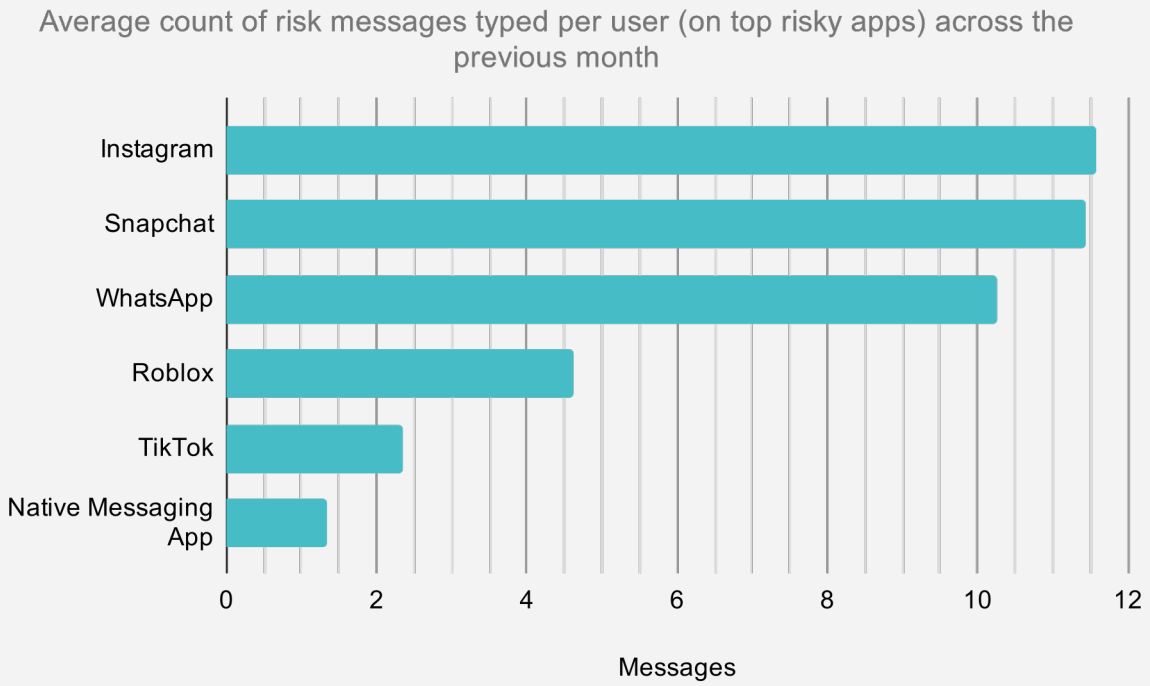
6.3 Sexting: Days of the week and time of the day





7. Apps overview: Risk activity levels and user averages





CONFIDENTIAL

8.1 Exploratory Analysis - behaviour change across COVID-19 lockdown and online schooling (extended range)

8.1.1 In the weeks leading up to the formal announcement of initial lockdown measures, messaging activity increased by 111.5%. Similarly levels of both cyberbullying and sexting detection rapidly increased. Rates of cyberbullying and aggression for instance, increased by 112% across the space of 1 week; while rates of sexting climbed by 228%.

8.1.2 During the first few weeks of lockdown (and with children in online schooling) messaging activity levels remained high. Messaging during the day-time also remained significantly higher (by 48.3%) when compared to levels observed while children were still physically in school. The elevated rates of risk message detection also remained stable for the first few weeks of lockdown with instances of risk messaging during the day-time higher than pre-lockdown figures.

8.1.3 As lockdown continued, there were signs that children began to pay more attention to online classes than they did during the initial lockdown excitement, as day-time messaging levels fall. Risk message detection levels also begin to decrease.

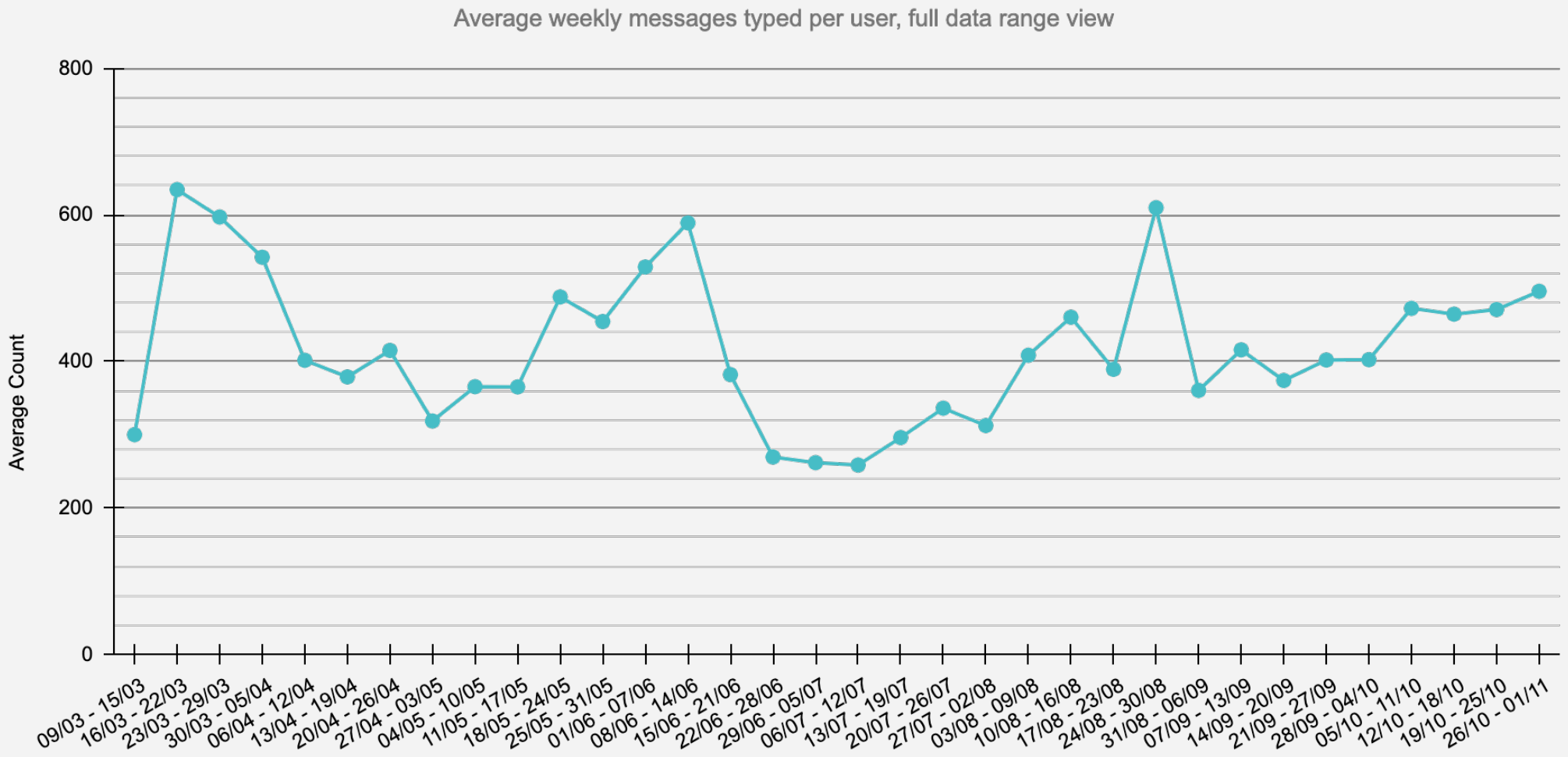
8.1.4 As lockdown measures began easing, we again observed a buzzing increase in messaging activity. In the build-up to heavier lifting of lockdown measures (18/05 - 30/05) averaged messaging activity rose by 25%. Additionally, we see a returned resurgence of risk messaging. For 5 weeks in a row during this stage of prolonged lockdown, the average user is sending more sexually inappropriate messages than aggressive messages.

8.1.5 As school holidays approach, messaging rates decreased. A switch in risk messaging patterns is also observed. In the week before school holidays begin and online classes end, cyberbullying messaging rates per average user overtake those of sexting message detection rates.

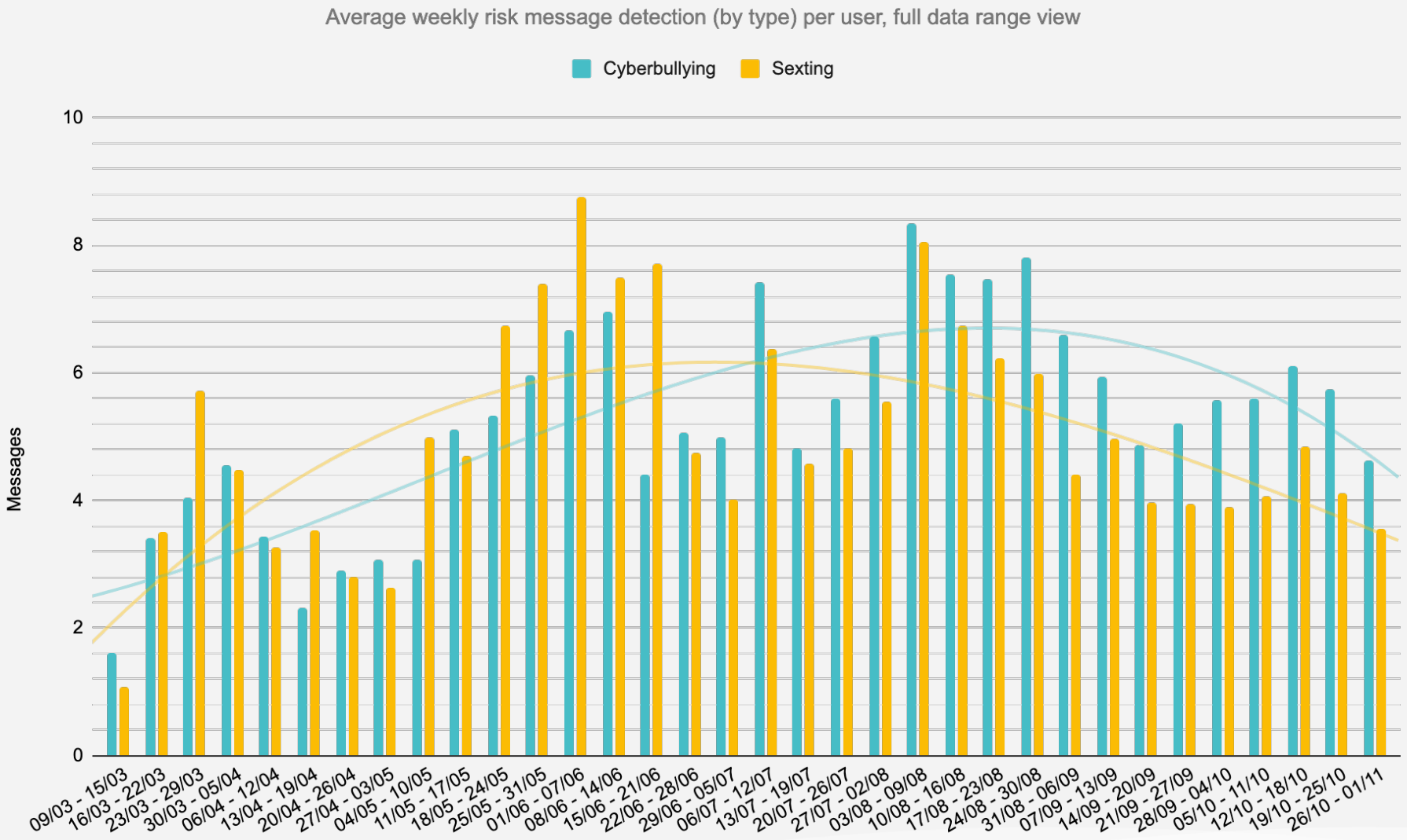
8.1.6 Throughout the holidays and as September approached (start of the new school term) messaging rates again increased. Messaging activity increases by 47.1% in the 2 weeks around Boris Johnson's announcement of a moral duty to have children back in schools when term starts in September.

8.1.7 As children physically returned to school in September, cyberbullying continues to be the dominant risk detected - in contrast to patterns observed during online schooling in lockdown. Further, As the second wave of COVID approached and rumours of a possible 2nd lockdown go around messaging levels further increased, mirroring the pattern observed in the build-up to the first lockdown. From 05/10 - 25/10 and with hospital admissions reaching their highest point since June and fatality rates increasing, children were sending around 61.9% more messages than they were during the majority of the school holidays.

8.2 In context: Messaging activity levels over time

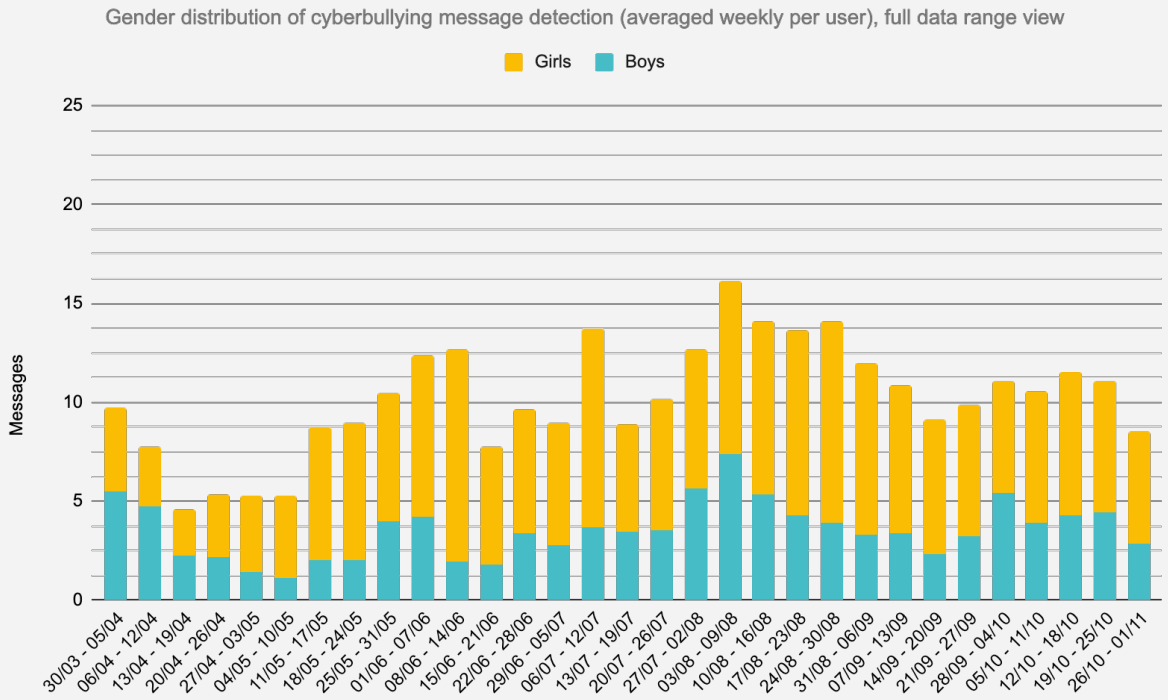


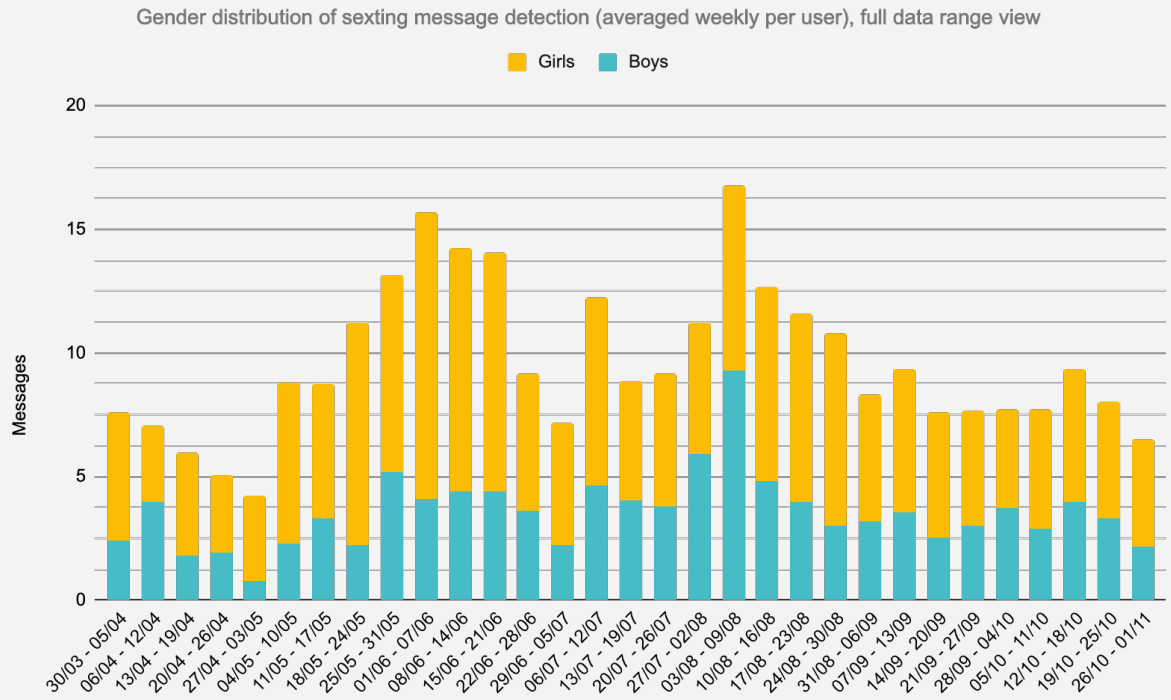
8.3 In context: Risk detection levels over time



8.4 In context: Risk message detection levels by region, over time

8.5 In context: Cyberbullying and sexting by gender, over time





9.1 Notes and Context

Data Collection

This report provides a snapshot of behavioural insights generated by the SafeToNet app which records when a child's message is filtered and prevented from being sent. Approximately 1m messages are analysed for inclusion into each month's report. Additional meta-data is collected including time of day, gender, age and more. Filtering occurs when the software detects cyberbullying, abuse, aggression and sexual connotation including sexting. Other behavioural patterns are recorded including a child's emotion.

Definitions:

Cyberbullying: SafeToNet's cyberbullying AI classifier detects forms of aggressive and abusive language. This includes but is not limited to: direct forms of aggression, bullying and name-calling; strong and explicit language or swearing of a threatening nature; indirect aggression and exclusion.

Sexting: SafeToNet's sexting AI classifier detects sexually explicit language used with intent, including sexual comments or compliments; overly suggestive or coercive language relating to sexual topics.

Dark Thoughts: SafeToNet's dark thoughts classifiers cover a range of themes associated with wellbeing and mental states, including anxiety, stress, fear, low self-esteem and victimisation.

9.2 Notes and Context

Statistical Significance and Methodology

This report does not test the statistical significance of the data. It is compiled using exploratory data analysis techniques. As more data is amassed and additional data patterns added to the report, SafeToNet will deploy more sophisticated techniques to provide statistical evidence of the data patterns. It is therefore not a scientifically produced report and should be referred to for trends and guidance only. This report takes a balanced sample of data and removes extremis. Data is cleaned and data noise removed where possible. As further data is collected we will apply forecast modeling to predict likely outcomes based upon trends. Furthermore hypotheses will be tested by collecting additional data to substantiate theories.

Independent Review

We are forming an independent data review panel of subject matter experts who will verify SafeToNet's data collection and analysis processes and also provide their own hypotheses of the collected data patterns.

Note:

SafeToNet does not see the child's messages.

Commercially sensitive data has been removed from this report to allow external distribution. This includes: source of users by business customer, number of users by customer, growth rate of users and other.

Report End

For further information please contact Emily Nicholass
enicholass@safetonet.com