In-depth analysis of newest device users and issues with flag detection during sleep

Analysis showing Active Alpha and Advance 2 devices have wider range of age and more users with lower meidan incomes. Evidence supporting the cliam that there is a direct relationship between darker skin tones and increased quanlity flags recorded during sleep.

Report prepared for MINGAR by Confidence Consulting

2022-04-11

Contents

Executive summary

Mingar's goal is to produce high-quality and affordable devices for the everyday user. By using customer-feedback and device analytics, we are actively working to produce the most advanced technology in fitness tracker devices, while also providing flexible options for our users, both in terms of cost and functionality. One of our main goals with the newly released Active Alpha and Advance 2 devices was to reach a greater customer base; specifically, customers that were looking for the same quality products in the Active and Advance line, without the costly price.

The current report presents our new customer demographics, based on the release of these two devices. Additionally, to improve the quality of our devices and following customer complaints regarding sleep tracking, this report assesses the effect of customer characteristics, specifically skin tone, on the frequency of quality flags reported by the device during a single night of sleep. These flags may be signaled by missing or unusual data.

Are new users, specifically those that bought the Active Alpha and Advance 2 devices, different from old users?

- When looking at the average household median income, that of new customers was lower than old customers' income by $4136.
- Two sample t-tests were performed on the average household income of new and old customer groups. This test assesses whether two averages are actually different from one another, or the observed difference in the average is simply due to chance. We concluded that there is in fact a significant difference between both incomes.
- Additionally, the interquartile range, that is, the middle 50%, of users' age is wider in new users than old users. Specifically, this middle ranges from 30 to 65 for new users, and from 33 to 56 for old users.

Are devices recording more quality flags during sleep for darker skinned users, according to their selected emoji, than for lighter skinned users?

- Using a linear model we created, that took the users' emoji skin tone as variables to predict the average number of quality flags, had an estimated coefficient of 2.398. This coefficient means the flag count for users with dark skin tone is $0.36 * e^{1.8933}$ times that of medium-dark skin tone users, and $766 * e^{-5.7725}$ times that of light skin tone. This model controlled for any differences in the age of the population (Fig. 1, 2).
- This scatterplot also shows that not only are users' with darker skin signalling more quality flags recordings from their devices, but these increase at a faster rate than lighter skinned users, as customers sleep for longer periods (Fig 1.)
- The average flags recorded per night of sleep by each users' device is 4.94, regardless of the indicated emoji skin tone.

- The average sleep time recorded per night per user, also regardless of emoji skin tone, is 368.8 minutes which converts to approximately six hours per night.

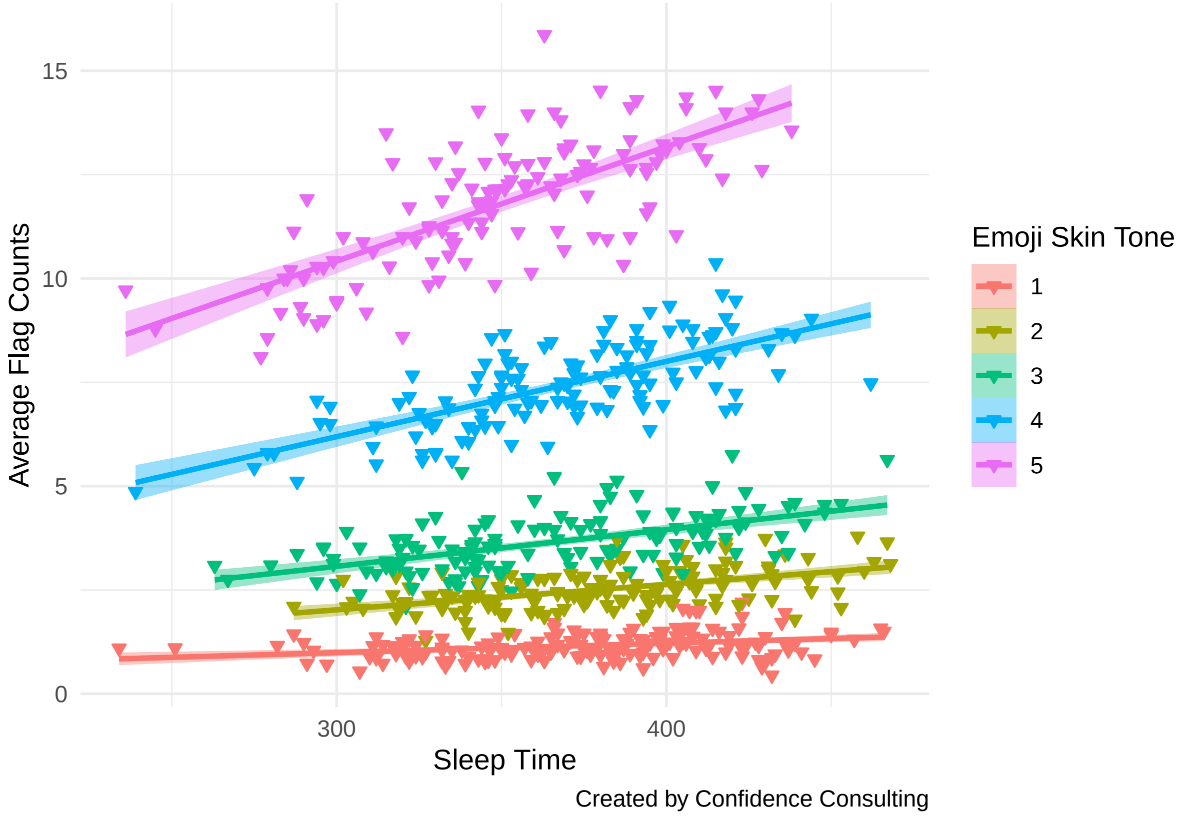

Figure 1: Scatterplot showing the distribution of averaged flag counts per customer based on their average sleep time, in minutes, per night

Predictors	Coefficient Estimates (Intercepts)
Light Skin Tone (1)	-5.7724
Medium-Light Skin Tone (2)	0.7726
Medium Skin Tone (3)	1.1868
Medium-Dark Skin Tone (4)	1.8933
Dark Skin Tone (5)	2.3958

Figure 2: Estimates of the coefficients in the generalised linear mixed model describing the number of quality flags detected according to emoji skin tone.

Technical report

Introduction

For this analysis, Confidence Consulting was hired by Mingar to help determine and act upon two of their concerns, both of which will be further elaborated in the following sections. The first of these involved familiarizing ourselves with new Mingar users, and investigating whether these users are different from older users. More specifically, it aimed to determine who was purchasing the Active Alpha and Advance 2 and whether these customers had also bought older models. This would help determine whether their marketing strategies worked and that new customers were being brought in. Mingar's second question was in regards to the performance of their product for users with darker skin tones, specifically on sleep scores. There had been a trend of complaints on poor performance for darker skin users, and we were hoping to investigate and discover the potential cause of this issue. In order to do this, the users' chosen emoji skin tones were used to represent their actual skin tone and thus help investigate the relationship between that and the frequency of flags detected during sleep. The data and analysis performed were used to provide insight for Mingar regarding their marketing and product consideration, as well as to evaluate their ethical practices.

Research questions

- Are new customers different than previous, traditional customers?
  - New customers are described as customers who have joined since the release of the new "Active" and "Advance" products.
  - Customers would be described as different if they are significantly different from traditional customers in any of the following parameters: age, household income, sex or skin color.
- Is sleep tracking significantly worse for darker-skinned users?
  - Sleep tracking can be measured by the amount of quality flags detected on average during sleep session for a single customer.
  - No information was acquired regarding race or ethnicity, and so customer skin-color will be gathered using the user's emoji skin-color.

Methods

Investigating differences in buyers' demographic characteristics for new and old products offered by Mingar

In order to determine demographic differences between buyers of the old and the newer 'ActiveandAdvance' models, additional data cleaning was performed. The data wrangling started with the original customer dataset which contained 16 variables and 56871 entries. The first step was the creation of a new variable, new_device, that grouped each customer into "New" or "Old" based on what device they had bought. After creating this variable, non-demographic measuring variables and rows containing NA values were removed in order to simplify the dataset and prepare it for analysis. The resulting dataset contained 42086 entries and the following variables:

- Age
- Sex
- Emoji_skin_tone
- Hhld_median_inc
- new_product

After cleaning the dataset, a summary table of each group was created in order to determine some general information about each group.

Variables	Min	Max	1st Qu.	3rd Qu.	Median	Mean
Age	17.00	92.00	30.00	65.00	48.00	48.18
Household Median Income	41880	195570	60592	85981	65829	70533
Average Frequency of Flags	0.000	25.000	1.000	6.000	3.000	4.298
Average Total Sleep Time	230.0	501.0	332.0	400.0	366.0	365.5

Figure 3: General information summary table for Active Alpha or Advance 2 device (new) users

Variables	Min	Max	1st Qu.	3rd Qu.	Median	Mean
Age	17.00	92.00	33.00	56.00	45.00	45.15
Household Median Income	41880	195570	65829	87225	68402	74669
Average Frequency of Flags	0.000	26.000	1.000	6.000	3.000	4.012

Variables	Min	Max	1st Qu.	3rd Qu.	Median	Mean
Average Total Sleep Time	222.0	496.0	349.0	405.0	377.0	376.3

Figure 4: General information summary table for other device (old) users

Variable Counts	New Users	Old Users
Female	13638	19199
Male	10279	12355
Intersex	326	565
NA's	162	347

Figure 5: Summary table comparing sex counts between Active Alpha or Advance 2 device (new) users and other device (old) users

Variable Counts	New Users	Old Users
Dark Skin Tone	3278	3786
Medium-Dark Skin Tone	3642	3634
Medium Skin Tone	3579	4381
Medium-Light Skin Tone	3302	6052
Light Skin Tone	3760	7009
NA's	6844	7604

Figure 6: Summary table comparing different emoji skin tones counts between Active Alpha or Advance 2 device (new) users and other device (old) users

After creating these summary tables, various tests were performed in order to determine differences between old and new device buyers. The first of these tests was a two-sample t-test. A two-sample t-test is a test that aims to determine whether any differences in means between two independent groups exist. In this case, this test was used to determine any differences in buyers' age and median household income. For the test, our hypotheses were as follows:

- H0: There exists no difference in income/age between new and old device buyers
- Ha: There exists a difference in income/age between new and old device buyers

Before conducting the t-test, however, it was important to see that the data satisfied the assumptions needed to perform the test. In order to do this, a qq plot, kernel plot and f-test were created and conducted respectively. Unfortunately, the normality and variance assumptions were not satisfied, however, the independence, continuous and random assumptions were. Despite the unsatisfied assumptions, due to the large size of the dataset, we felt comfortable continuing with the two-sample t-test. Prior to performing the test, a visualization with box plots of the median household income, both with (top) and without (bottom) outliers, according to whether they were new or old buyers was created.

Figure 7: Boxplot visualing the median household income differences of users based on whether they purchased the Active Alpha or Advance 2 device (new), or any other devices (old) with outliers

Figure 8: Boxplot visualing the median household income differences of users based on whether they purchased the Active Alpha or Advance 2 device (new), or any other devices (old) without outliers

Figure 9: Boxplot visualing the age difference between users who the purchased the Active Alpha or Advance 2 device (new), or any other devices (old)

Through conducting the test, it was determined that the p-value in each case was below our threshold of p=0.05, indicating a rejection of the null hypothesis that no significant difference between income and age for each group exists. Through the boxplots that were previously created, it was also discovered that a greater number of lower income persons were buying the new devices compared to buyers of the older devices. Additionally, the newer devices seemed to be attracting both younger and older persons as the interquartile range for new device buyers was much larger than for old device buyers.

The second test performed to determine demographic differences was the chi-square test. The chi-square test is used to determine if two categorical variables from the same population have significant correlation between themselves. In this case, the relationship between skin tone or sex and whether they bought a new or old device was measured, meaning our hypotheses were as follows:

- H0: Skin tone/sex and new_device are independent
- Ha: Skin tone/sex and new_device are related

Before conducting the chi-square test, it was important to determine whether the data satisfied the required assumptions. The data satisfied the categorical and group assumptions allowing us to feel comfortable with proceeding. Once again, the resulting p-values for both tests was below 0.05, indicating the rejection of the null hypothesis that skin tone or sex and new_device are independent. In order to visualise these relationships, grouped bar charts were created.

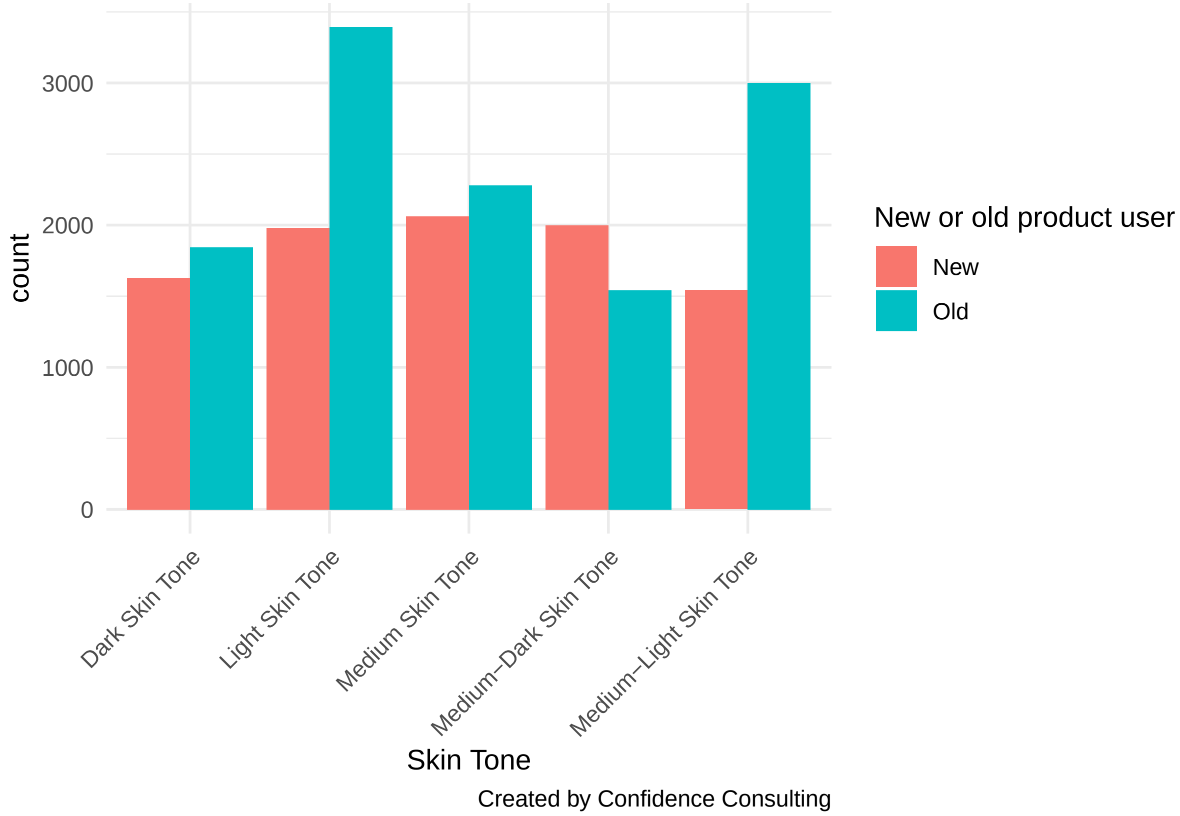

Figure 10: Bar graph comparing the skin tone counts differences between the Active Alpha or Advance 2 device (new) users and any other devices (old) users

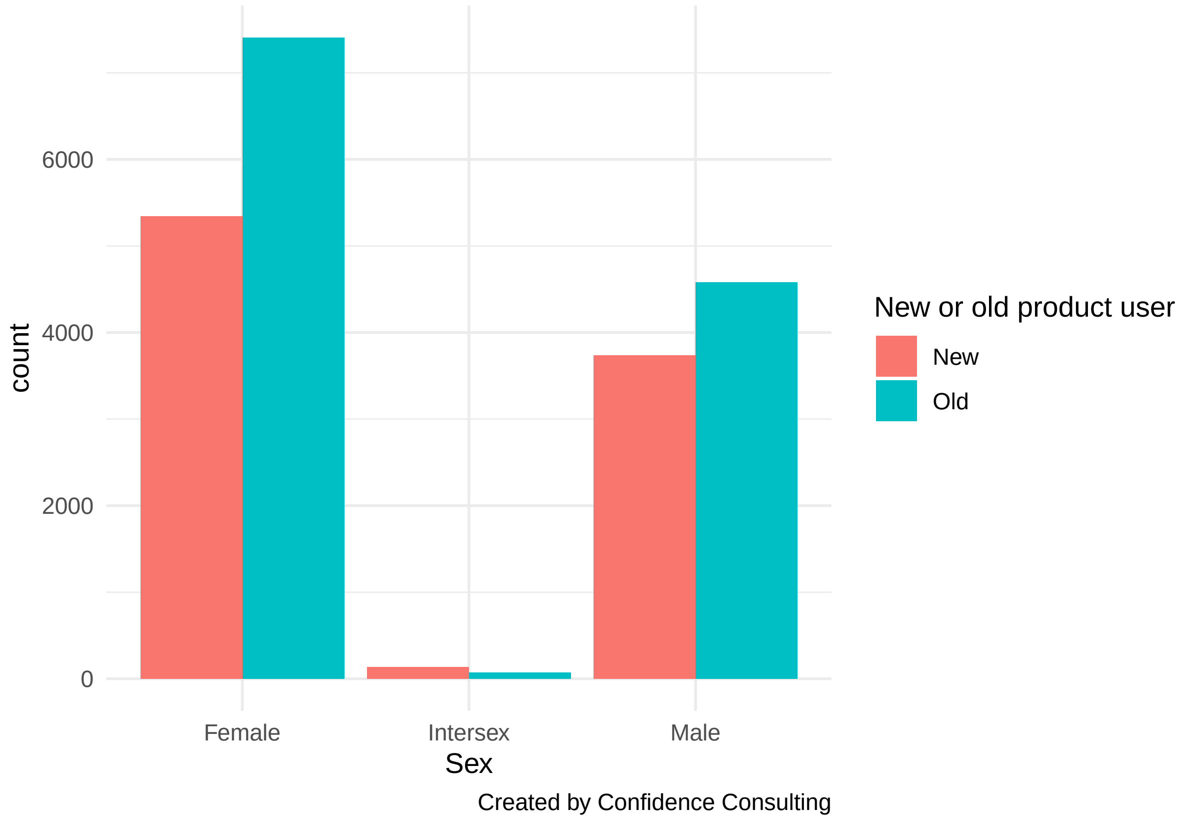

Figure 11: Bar graph comparing the sex between the Active Alpha or Advance 2 device (new) users, or any other devices (old) users

For sex, we found it was dependent on what group you are in as both males and females bought less new devices than old. This, however, could be a result of less people in total so far buying the new devices and such more data on new device buyers should be collected. It was also discovered that females were still more likely than males to buy Mingar products. In the case of intersex individuals, there is not enough data to make a strong conclusion regarding differences between new and old device buyers. When looking at the results for skin tone, we can see there is a much more uniform distribution of new device buyers than previously. Overall, this indicates the marketing performed by Mingar has resulted in a greater diversity of customers for their new devices.

Investigating the relationship between Mingar users' skin tone, based off of their chosen emoji, and the frequency of flags detected during sleep

In order to investigate the relationship between Mingar users' skin tone and the frequency of flags detected during sleep, additional cleaning was done on the newly merged customer data.

This involved removing any observations containing missing sleep data, such as NA inputs for flags recorded. Furthermore, the assumption that the chosen emoji skin tone represented the user's skin tone was made. Therefore, any observations containing NA for the emoji skin tone were removed as otherwise we would not be able to model our response variable, the average number of flags per user, according to the user's skin tone. In order to model the response, the levels in the Emoji_skin_tone variable were reordered to include the first level, 1, as the lightest skin tone and counted up to level 5, the darkest skin tone. Additionally, to better manage the amount of data we received, two variables were modified to calculate the average values for a single user. Duration, the number of minutes of sleep every night, and flags, the number of quality flags counted every night, were simplified to include one these values averaged for every user. All of these changes were included in a new dataset called cust_sleep, which included both of these variables as well as the cust_id and date. This dataset contained a total of 20622, as opposed to the 56871 observations from the initial dataset. An initial visualization of the average flag counts according to skin color is included below. An additional visualization of the same response with the inclusion of age as a predictor is also presented. This plot provides some evidence suggesting that skin color is a greater predictor than age, since there is no bias of flags for any age group. Instead, there are simply less older individuals buying these devices, and as such represent less data.

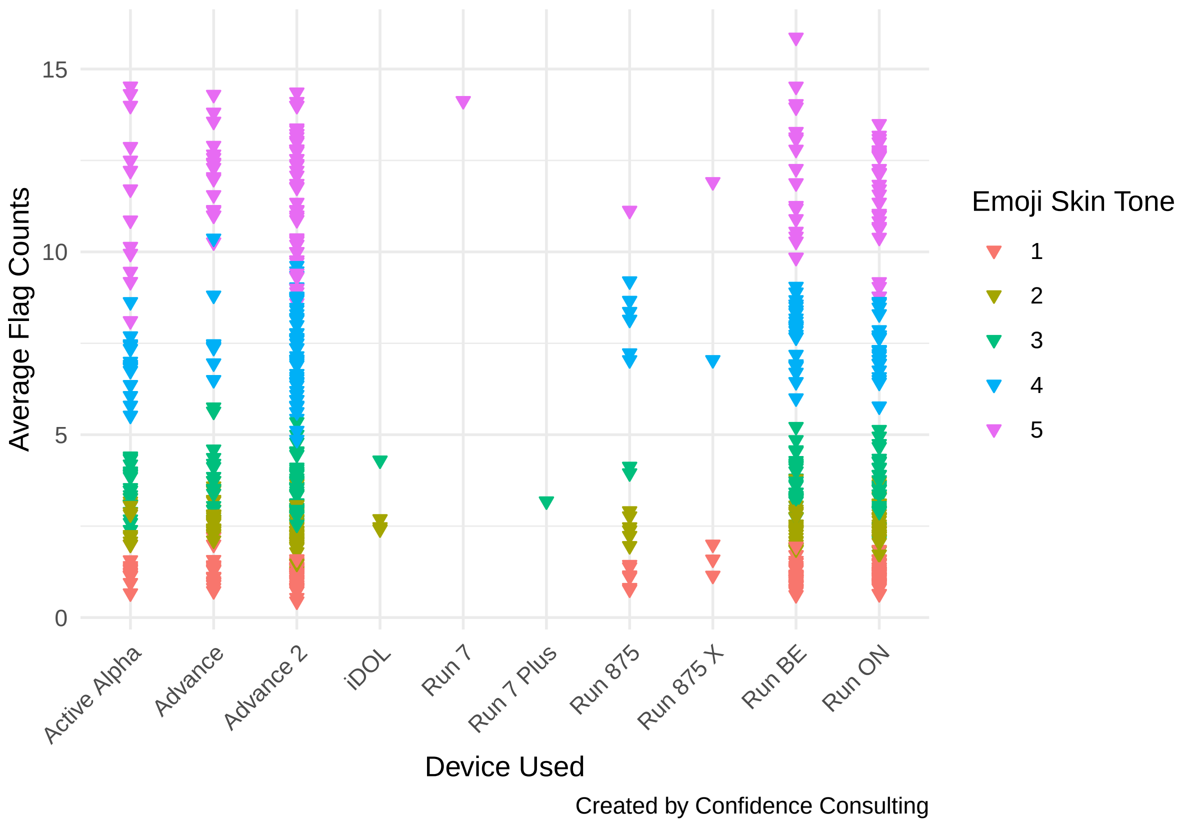

Figure 12: Scatterplot visualizing the distribution of average flag counts per customer based on the devices used, represented by their emoji skin tone colour

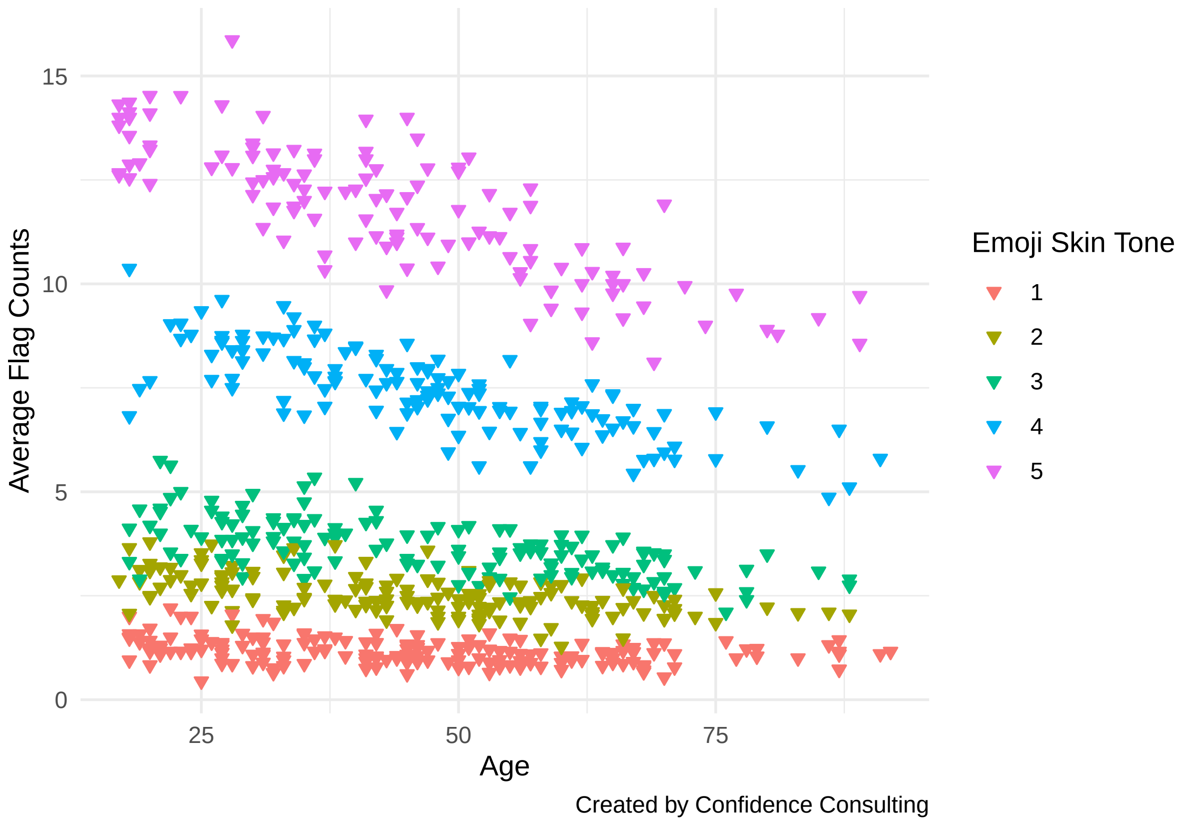

Figure 13: Scatterplot showing the average flag counts per customer based on the user's age, represented with their emoji skin tones

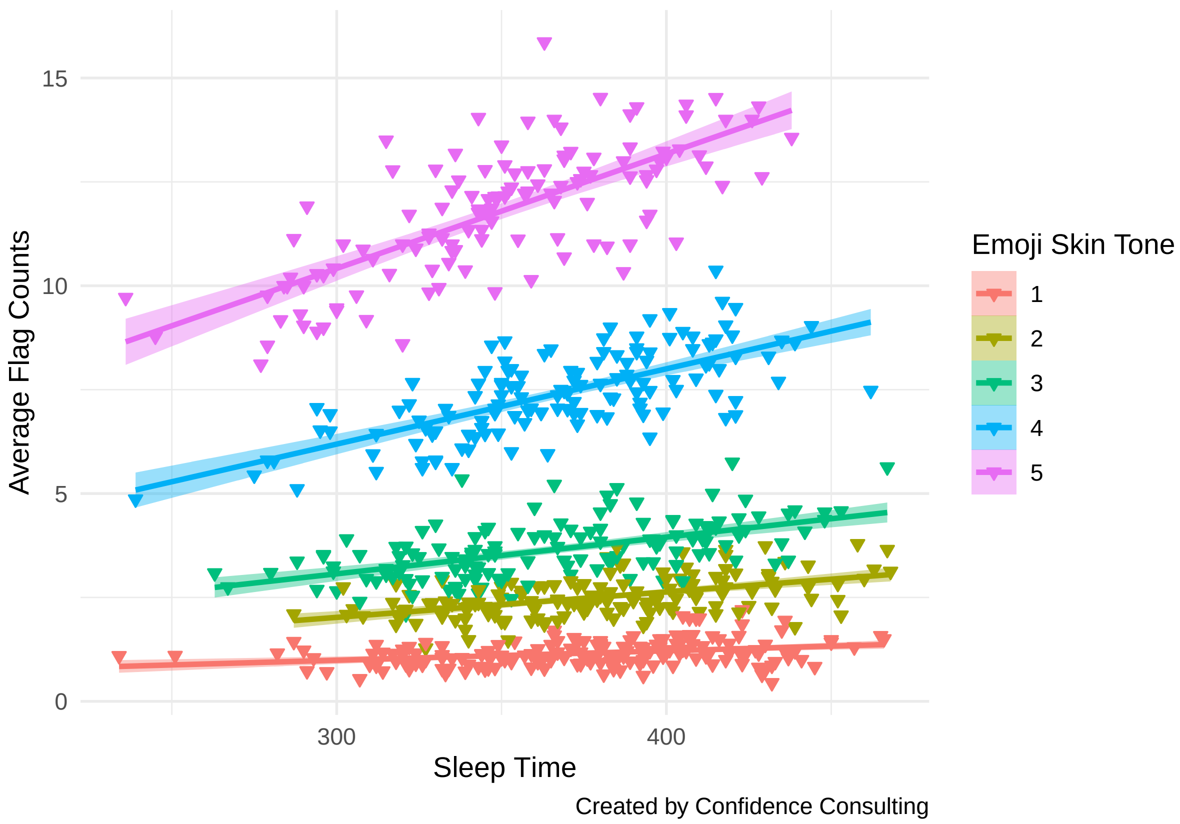

Figure 14: Scatterplot showing the distribution of averaged flag counts per customer based on their average sleep time, in minutes, per night

Variables	Min	Max	1st Qu	3rd Qu	Median	Mean
Average Frequency Flag Counts	0.4444	15.8636	1.9444	7.6500	3.3333	4.9408
Avergae Total Sleep Time	234.0	468.0	339.0	402.0	370.0	368.8

Figure 15: Summary table showing general information for average frequency of flag counts and total sleep time

After cleaning the data, the predictors were fitted into a generalised linear mixed model. The initial model had the Emoji_skin_tone as a predictor and cust_id as a random effect, with an offset of sleep time. Two predictors, Age and device_name, were then added one at a time. Using a likelihood ratio test, the model with the predictor Emoji_skin_tone was the best fitting model.

The second research question requires modeling relationships between multiple predictors in the data. However, some values of data are missing after tidying, which leaves a certain level of uncertainty and biases in the values of predictors. The purpose of choosing a linear mixed model for the question is this statistical model includes random effects that accounts for the variability in the data. The model uses maximum likelihood estimation that is statistically appropriate for a model where there are more than one parameter present.

After comparing the initial model, mod1, with the second model, which included the additional predictor Age, this resulted in a p-value of 0.7167. In other words, mod2, with predictors Emoji_skin_tone and Age, was not a better fitting model. An additional predictor device_name was then added into mod3, and when comparing this model with mod2, the likelihood ratio test result had a p-value of 0.2685, signifying that the mod1's fit is more statistically significant, making mod1 our final model.

From mod1, a linear mixed model between Emoji_skin_tone and cust_id, the intercept of fixed effects can be interpreted as the average flag for emoji skin tone 1. For a skin tone of 2, the average flag increased by approximately 1.02178; for skin tone of 3, the average flag increased by approximately 1.24224; for a skin tone of 4, the average flag increased by approximately 2.08761, and for skin tone for 5, the average flag increased by 2.55537. Thus, the data shows that the average flags increase as the skin tone gets darker. Furthermore, the proportion of variance in flag count explained by different random effects such as cust_id is around 0.03835. This means that the level of effects from the interaction are close to the mean and each other, therefore, there will not be significant changes to the projection of the predicted model.

Results

To investigate the differences in buyers' demographic characteristics for new and old products offered by Mingar, The null hypothesis was that there exists no difference in income/age between new and old device buyers. After running the t-test, the p-value was 2.2eˆ-16, which is below the threshold of 0.05, meaning the null hypothesis should be rejected. Secondly, a chi-squared test was used to test the null hypothesis of whether skin tone/sex and new_device are independent. The p-value for both was 2.2eˆ-16. They are also below 0.05, meaning the null hypothesis is also rejected.

For the next investigation of the relationship between Mingar users' skin tone, based off of their chosen emoji, and the frequency of flags detected during sleep, a generalised linear mixed model with a likelihood ratio test was used.

Discussion

After conducting research question 1, it was determined that the marketing conducted by Mingar for their new and more affordable models resulted in previously underrepresented portions of the population purchasing their product. Through various tests and plots, it was shown that lower income customers were buying the new Mingar products at higher rates than for the previous more expensive models. Age of customers was also affected by the marketing as both younger and older people were attracted to the new product. In the case of sex, it did not appear as if the marketing affected either male or female buying habits, as both bought proportionately less than previous models. Finally, Mingar's marketing campaign appeared to result in a more uniform distribution of buyers based on skin tone than the previous models, which were largely dominated by light and medium-light skin tones.

In regards to research question 2, the investigation of the relationship between the frequency of flag counts per sleep and the users' skin tone concluded that the average frequency of flags does increase as skin tone gets darker. Overall, the likelihood ratio test showed that the emoji skin tone have an effect on the average frequency of flag counts per sleep. Emoji skin tone does not exactly explain the increase of the average frequency of flags. This is something that Mingar should further investigate on to uphold their ethical promise standards.

Strengths and limitations

For the first question, in order to investigate new and old customers, we had to group customers according to which device they bought. Specifically, customers were categorised as "New" if they bought either the Active Alpha or Advance 2 devices, and "Old" if they bought another device. This isn't entirely accurate, as some devices, such as the Run ON, were released after the Active Alpha or Advance 2 devices. As such, these customers are not exactly old, and may have been affected by the advertisements for the Active or Advance lines initially, and then decided to opt for the Run line. Additionally, the data provided for the median household was also an estimate based on census and thus another assumption when investigating the relationship between a customer's income and whether there were a new or old customer. Finally, the fact that household income was not normally distributed or had equal variance for both groups may have resulted in a type 1 error in which the null hypothesis was incorrectly rejected. Keeping these assumptions in mind however, the average income for new and old users was statistically significant, and there was also a significantly wider range of ages for new users as compared to old users.

During our analysis for the second research question, in order to improve data visualisation and efficiency of code run-time, data was simplified to express the average number of flags and the average sleep time of a single user, rather than using the results from every single night of sleep for all users. Although this generalises user information, simplifying up to 100 nights of sleep and flags into one observation, this makes the data much more manageable and facilitates visualisation. Additionally, this allows each observation to be independent from one another, as there is now only one observation per user. Moreover, sex and gender were not included in this analysis, as we did not feel confident in drawing conclusions relating to user sex. Due to the fact that some customers used different pronouns than their specified sex, this would require us to either group them according to sex or gender, in which case we would either be mischaracterizing their biological sex, or misgendering them. A final limitation for the second question is the assumption that the emoji skin tone represents the users' skin colour. This may not be an accurate representation of the users' actual skin tone. As Mingar claims that they do not collect information on race or ethnicity, using the skin tone of the emoji is the closest option or substitute available. This is an efficient way to analyse such data as it is safe to assume that most users choose the emoji that represent their skin tone. However, this could possibly skew the data since it is still an assumption we are making.

Consultant information

Consultant profiles

Maria Fahim. Maria is the executive director at Confidence Consulting. She graduated from the University of Toronto with a Bachelor's in Physiology and Immunology, and a minor in Statistics in 2022. Before joining Confidence Consulting, Maria also completed her Master's in Biostatistics at the Dalla Lana School of Public Health.

Sunny (Wei-Han) Wang. Sunny is the junior analyst at Confidence Consulting. She graduated from University of Toronto with bachelor degrees in Mathematics and Statistics in 2023.

Emily Xu. Emily is a junior analyst at Confidence Consulting. She graduated from the University of Toronto with a Bachelor's in Economics and Statistics with a minor in Psychology in 2024. Her statistical knowledge and analytical skills are a great contribution to the team.

James Richards. James is an intern at Confidence Consulting. He graduated from the University of Toronto in 2023 with a Bachelor's in Statistics and Minors in Mathematics and Economics. While interning at Confidence Consulting, Jim is also completing his MBA at Harvard University.

Code of ethical conduct

Confidence Consulting employees adhere to a strict code of conduct outlined below

Promise of confidentiality. When working with sensitive data, in addition to signing non-disclosure agreement contracts with every new client, we have promised to protect the privacy of customers and anonymized any data that may violate this promise, as well as promise to not misuse or abuse any data. **Promise of accountability.** This includes taking responsibility in our work and making ethically-sound and valid decisions that are backed up by reason and relevant methodology. **Promise of professionalism.** Confidence Consulting employees behave in a professional and ethical manner when working with each other, our clients, and on projects involving members from multiple disciplines, **Promise of respect.** Respect towards all individuals is of the utmost importance at Confidence Consulting. We are dedicated to creating a welcoming and honest environment, where everyone feels welcome.

Appendix

Web scraping industry data on fitness tracker devices

Web scraping was used in this report to convert the data from eternal sources to a structured format that was easily accessible.

We set user agent to have our contact information and the purpose of our web-scraping so they are able to contact us if they were any issues. We respect any details provided in the robots.txt file and the crawl limit time and which agents are allowed to scrape. If the craw limit is 5 seconds, the site will only have fetching rate of 5 seconds.

We scraped the industry data we needed and converted it into html dataset that we use in the report. The same process of web scraping is used on creating emoji data that we use in the report.

Accessing Census data on median household income

First we set a folder for cache that includes raw data for census API. Next, we obtain all region at the 2016 census and save under the name *regions*. We then filter the *regions* data where the level is CSD (what does this mean) and save under a new name.

We retrieved a census API that fits our requirements and simplify the retrieved data to variables we needed, which include median household income, population, and census ID.

Accessing postcode conversion files

We followed the ethical considerations, as we accepted the license agreement to access this data, not publish online and only use information that was needed.

We opened the file containing postal code conversion information in rds file type. After accessing the raw data, we selected the variables needed and saved under a new dataset name. Lastly, we converted the dataset into a rds file type store under the name postcode.

References

Bates, Douglas, and Martin Maechler. 2021. *Matrix: Sparse and Dense Matrix Classes and Methods*.

Bates, Douglas, Martin Mächler, Ben Bolker, and Steve Walker. 2015. "Fitting Linear Mixed-Effects Models Using lme4." *Journal of Statistical Software* 67 (1): 1–48. https://doi.org/10.18637/jss.v067.i01.

Fox, John, and Sanford Weisberg. 2019. *An R Companion to Applied Regression*. Third. Thousand Oaks CA: Sage. https://socialsciences.mcmaster.ca/jfox/Books/Companion/.

Perepolkin, Dmytro. 2019. *Polite: Be Nice on the Web*. https://github.com/dmi3kno/polite.

R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org.

von Bergmann, Jens, Dmitry Shkolnik, and Aaron Jacobs. 2021. *Cancensus: R Package to Access, Retrieve, and Work with Canadian Census Data and Geography*. https://mountainmath.github.io/cancensus/.

Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org.

———. 2021. *Rvest: Easily Harvest (Scrape) Web Pages*.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Wickham, Hadley, Romain François, Lionel Henry, and Kirill Müller. 2021. *Dplyr: A Grammar of Data Manipulation*. https://CRAN.R-project.org/package=dplyr.

Wickham, Hadley, and Evan Miller. 2021. *Haven: Import and Export 'Spss', 'Stata' and 'Sas' Files*.

Xie, Yihui. 2021. *Knitr: A General-Purpose Package for Dynamic Reportgeneration in R*. https://yihui.org/knitr/.

Zeileis, Achim, and Torsten Hothorn. 2002. "Diagnostic Checking in Regression Relationships." *R News* 2 (3): 7–10. https://CRAN.R-project.org/doc/Rnews/.