Wearable Products Case Study

Age, Income and Sex Impact on Customers' Choise and
Product Performance Over Dark Skin

Report prepared for MINGAR by Datanaly.Inc

2022-04-07

Contents

Executive summary

To gain market share, Mingar company needs to analysis who is the new customers and the difference between the traditional and new customers who afford Advance 2 products.By exploring the data analysis, income, sex and income have significant impact on types of customers. The new customers are people who are older female with lower income than traditional.

The social media team of Mingar company clearly indicates there are many complaints that the device are performing poorly for dark skin customers, especially the sleep score. They take this event seriously and would like to investigate whether the device are actually performance poorly for dark-skin customers.

By exploratory the data analysis, we notice the the customers' skin and device be the important factors can influence the sleep score and the device performance. The flag can represent the device performance.

There are some key findings:

- Most of new customers are females and it takes about 60%. New customer median and mean incomes are close which is about 66000 dollars. The mean age for new customers is 47.952 which is close to median age (47).

- Keep sex and income fix, the log odd of being a new male customer compared to female customer increased by 0.1919%. It represent there is no significant difference between new and traditional customers.

- Keep age and income fix, when a consumer's age increase one unit, the log odd of being a new customer increased by 0.0197%. It represent there is no significant difference between new and traditional customers on age.

- Keep sex and age fix, the when a consumer's income increase one unit, the log odd of being a new customer decreased by 0.0004%. It means there is no huge difference on income for new and traditional customers.

- For dark-skin customers with device Active Alpha, the expect value of flag occurrences per duration is 0.033.

- For light skin customers with device Active Alpha, the mean of flag occurrences per duration is 0.003.

- The expected value of flag occurrences per duration for dark-skin customers with Run 7 Plus smaller than dark-skin customers with Run 875 device 0.008118.

- The customers' skin is the significant factor that can affect the flag occurrences per sleep duration, and the device be another factor affact the flag occurrences.

Table 1:summary table for customers

customer	numbers on sex for female	numbers on sex for male	IQR income	IQR age
new customer	6110	4326	17613	31
traditional customer	4971	3416	20152	21

This table summaries the basic information of new and traditional customers. There are 6100 new female customers and 4326 male new customers which represent we have more female customers about 2000 people. In addition, the IQR income (17613) for new customers is smaller than traditional customers' (20152). As for the IQR age for traditional customers (21) is smaller than new customers 10 years old.
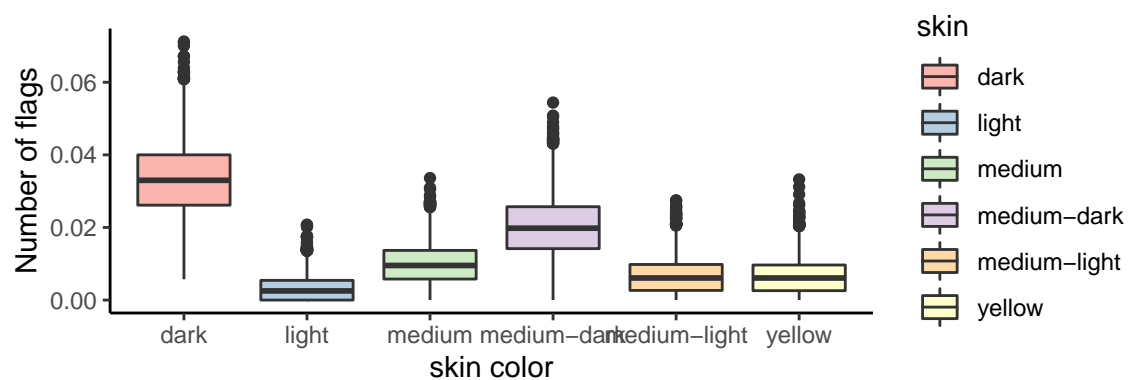
Figure 1: The distribution of flags on customers skin

The box plots indicate the distribution of median occurrences of flags on different customers' skin. We can observe the significant difference between dark skin customers and other skin customers. The median flag occurrences of devices of dark skin customers are 0,03 times which is the highest, and the lowest median flag occurrences are 0.002 times which is the light skin customers' devices.

Technical report

Introduction

Mingar company is now committed to developing wearable electronic products since the wearable market shows a growing trend. This company used to develop GPS units for military personnel and outdoor recreation. In order to growing the share of market and compete with Bitfit, we need to study who are the new customers and what is the main difference between new and traditional customers. Based on this question, knowing customers' basic information is necessary and we also need to connect the produces and device name with customers so that we have the overall understanding. After that, for the first question which define the new customer, we should filter the new customer to make plots and tables. Then, we have result on define the new customers. To compare the new and traditional customer, the performances of products and customers' basic information are necessary to study. So, we use the generalized linear model (glm) to analysis which factors have the significant impact for different types of customers. Lastly, select the each factors to study the importance for customers.

For the second question, this company meets some products issue that many customers complain their device performance poorly for dark-skin customers. And they would like to investigate whether their devices show this issue. Our company will use provided data to analyze this and aim to find any other factors that can affect the device's performance. We exploratory the data analysis and observe any related factors that can affect data visualization's flag occurrences. Then, we use the related factors we conclude from the data visualization to construct models. We will set several models to compare and then determine the most appropriate model that can help us investigate whether their device performs poorly for dark-skin customers.

Research questions

- Who are new customers? And what is the difference between new and traditional customers in company Mingar?

- Are devices performing poorly for dark skin customers and are there any other factors can effect the performance of devices?

Method

Data wrangling

In this study, for the first question, we want to study who is a new customer and the difference between new and traditional customers based on the products of Mingar. So, under seven data we have, put them together by the common variable of any two data sets. In addition, since the aim company is Mingar, for the device data, we only filter the information about Mingar and rename some variables so that in the following study, it can help us to analyze. Moreover, since we know the date of birth of each customer, mutate a variable to calculate the age. Also, based on the standard code modifiers of emoji, we can know the type of skins of customers. Furthermore, we need to study the new and traditional customers and define the new customer when the line is advanced and Active. For the sex, we only choose the sex male and female. last, delete the unnecessary column such as pronouns, date of birth, and code of emoji and we call the final data set. Therefore, in this data, we can know the basic information of customers which sex, age, and income. Also, the name and performance of the equipment purchased by the customer are recorded such as Heart rate sensor and GPS.

The full customers' sleep data is created by joining the customer information data and partial users' sleep data. These two data contain the postcode of customers that we can join them by these variables. One of the research aims is to discover whether the skin can affect the flags occurrences of devices. We can determine the customers' skin depending on the emoji they use. Therefore, we divided customers skin into 6 levels: dark, medium, medium-light, light, medium-dark, yellow. The flag represents the device occurs errors in customers' sleep duration, which might be data missing, data quality issues, or other device issues. If we would like to research the device's performance for customers, we need to calculate the number of flag occurrences in sleep duration. Since each customer's sleep duration is different, it is inaccurate to compare the flags of devices. Thus, we need to calculate the number of flag occurrences per sleep duration that will be more accurate. Moreover, we strictly follow the procedures to protect human rights and dignity, thus we use the virtual postcode (Customers' id) to represent the customers' postcode. The postcode correlated to several customers' id, and we decided to keep the customers' id with the most population that can mostly represent the proportion of the population in that postcode area. Finally, the customers' sleep data contains the customers' id, device name, date of sleep session started, the duration of sleep in minutes, flag occurrences, customers' sex, income and age.

Analysis of customers and products

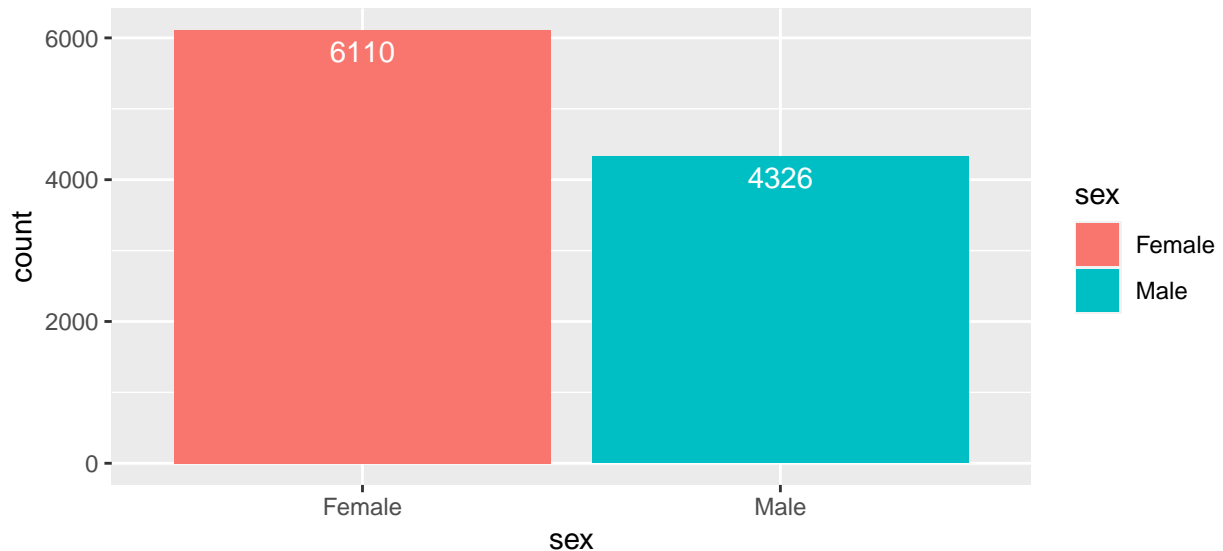### Characteristics of new customers

Figure 2: New Customers Sex

The bar plot shows that female new customers are 6110 which is 45% of all new customers. And male customers are 4326 which is about 41% of all new customers. And there is close to a 20% difference, which may be one of the important factors for the products. The consequence of this situation may be because females have more time for exercising and more females pay more attention to their status and health.

Table 2: Numerical Summary of Household Median Income for New Customers

min	median	max	mean	sd	IQR
41880	65829	195570	68817.89	14463.61	17613

Based on the table above, the minim income for the new customers is about 41880 dollars. And the maxim income for the new customers is about 195570 dollars. This means the range of income is quite large. In addition, the mean for income is 68817.89 dollars. This represents that most people's income is less than 70000 dollars.

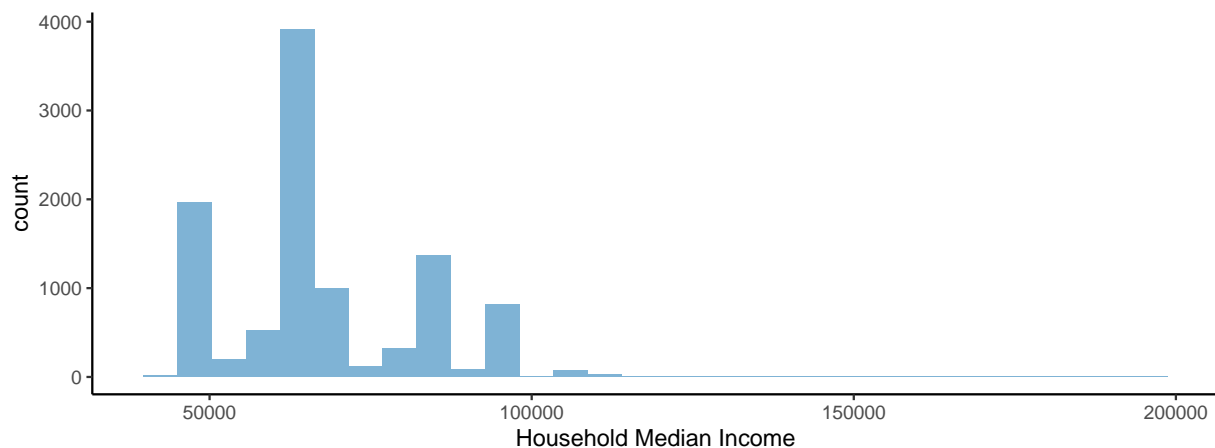For the more directly represent will show below:

Figure 3: New Customers' Household Median Income

The histogram shows that most new customers' income is between 40000 to 100000 dollars. There are about 2000 new customers' income reached close to 50000 dollars. And about 3900 new customers are 65000 dollars. In addition, only a few customers can earn the money more than 110000 dollars. The different income leads to different purchasing powers. It will directly influence the sales volume of the company products.

Table 3: Numerical Summary of Age

Variables	Mean	Median	IQR	Min	Max
age	47.952	47	31	18	92

The Table shows that the minim age for new customers is 18 and the eldest new customer is 92 years old. This means almost all the different ages people are our potential customers. And the mean age is about 47.952 years old. This means that about 47 years old people are paying more attention to their health and prefer to do more exercise.

Table 4: New Customer Devices Name with Proportion

Measure	Active	Active Alpha	Active HR	Advance	Advance 2
count	275.000	1734.000	294.000	2018.000	6115.000
percentage	0.026	0.166	0.028	0.193	0.586

The above Table 4 shows that although the company has 15 different types of products. The new customers are all only interested in five products, which are Active, Active Alpha, Active HR, Advance and Advance 2.In detail, there are 6115 new customers bought the product Advance 2 which takes about 58.6% of all the selling. And 2018 Advance products are sold to the new customers which are about 19.3% of all products. As for the last three products, Active Alpha takes 16.6% which is sold 1734. And the customers who bought the products Active and Active HR all less than 3% in total.
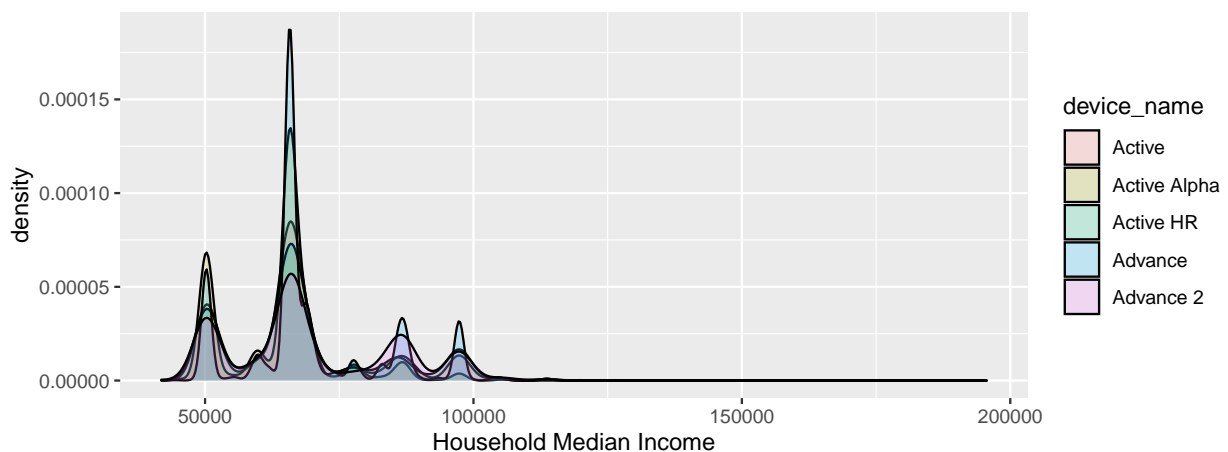

Figure 4: The Combination of New Customers' Household Median Income and Devices

According to the graph above, it shows that the new customers whose income is above 65000 dollars, most of them prefer to buy the Active product. However, the people whose income is less than 6500 dollars, they prefer to buy the Active Alpha product.

Table 5: The summary table for all the functions that customers' devices owned:

device_name	count for Yes	proportion for Yes	count for No	proportion for No
Pulse Oximiterr	0	0.00	10436	1.00
Heart Rate Sensor	10161	0.9736	275	0.0263
GPS	8133	0.7793	2303	0.2206
Sleep Tracking	9867	0.9463	569	0.0545
Smart Notifications	9867	0.9463	569	0.0545
Contactless Payments	9867	0.9463	569	0.0545

Based on Table 6 above, for the function Pulse Oximiterr, according to the previous summary table, all products do not have this function, thus, the proportion for this function is 0. In addition, the customer whose product has heart rate sensor, Sleep tracking, Smart notifications and Contactless payments functions make up about 95% of all the new customers. This means almost all the customers think these four functions bring convenience to them. As for GPS function, only 77.93% of new customers enjoy it. And 22.06% of new customers are not interested.
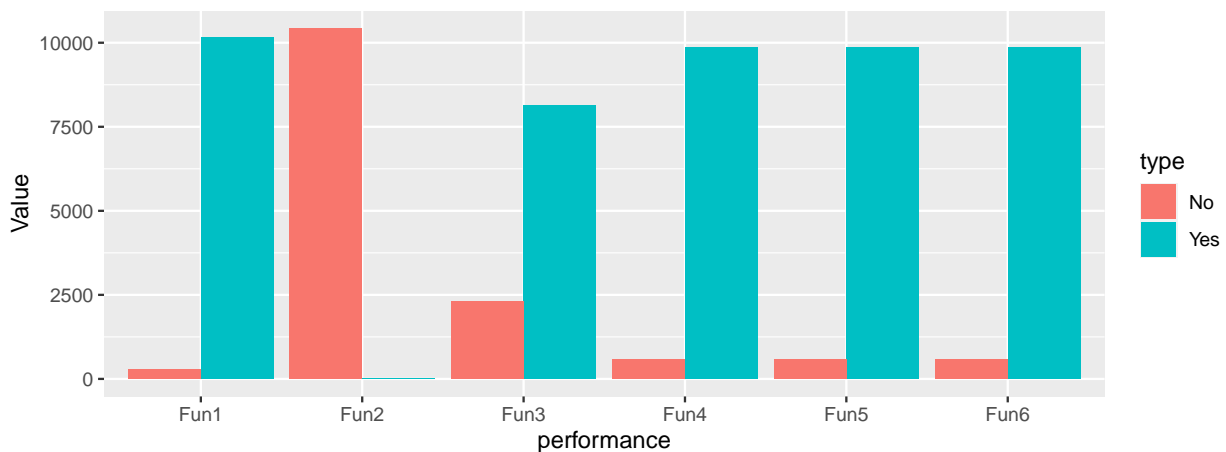

Figure 5: The Devices Functions for New Customers

For the bar plot, yes means the customers whose device has a certain function. No means the customer cannot enjoy the certain function because of the different products they bought. Fun1 represents as Heart Rate Sensor function. Fun2 represents Pulse Oximiterr. Fun3 represents GPS. Fun4 represents Sleep Tracking and Fun5 represents Smart Notifications. As for Fun6, it represents Contactless Payments. This graph shows that about 10000 customers' products have contactless payments, heart rate sensor, sleep track, and smart notifications functions. And no customers' product has the pulse oximiterr function. Based on the device function table, three devices have these four functions and no one product has pulse oximiterr function. As for the GPS, almost 8100 people's products have this function. And less than 2500 people don't have this function for their devices.

Conclusion: New Customer Strength

Based on the sex variable for new customers, most new customers are females. After exploring the numerical variable household median income, about 3900 new customers earn 65000 dollars. After connected with the production analysis table, most of them are afforded the product Advance 2 (145 dollars) with functions, Heart Rate Sensor, GPS, Sleep Tracking, Smart Notifications,

Contactless Payments and Recommended Retail Price.

Preferences of new and existing customers in choosing products

Table 6: Numerical Summary of Sex for Traditional Customers

sex	n	perc
Female	4971	0.592703
Male	3416	0.407297

The Table shows the traditional customers' sex. And it represents that there are 4971 customers belonging females which takes 59.27% of all the traditional customers. And 3416 males customers which are 40.72% of all the traditional customers. This means that there are about 20% more females than males.
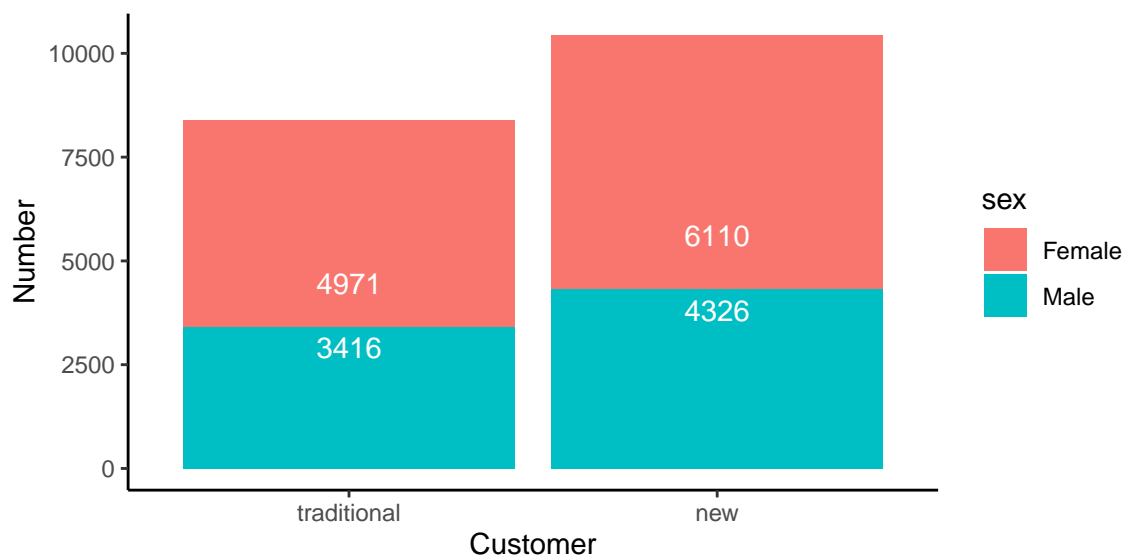
Figure 6: Barplot of Sex for New and Traditional Customer

From the bar plot above, the total number of new customers is much greater than the traditional customers. The total number of new customers is more than 10000. And the traditional customers is about 8700 in total. However, both have a similar ratio between males and females. The females all take about 60% of all their customers. And male customers take up about 40% of all their customers. There is a huge difference between female and male customers. Thus, the main customers' group for the company is always the females.

Table 7: Numerical Summary of Age For traditional Customers

Variables	Mean	Median	IQR	Min	Max
age	46.516	46	21	18	92
income	73181.986	65829	20152	41880	195570

This table shows the summary of the age and income of the traditional customers. The traditional customers' minim age is 18 and the maxim age is 92. The traditional customers' age range is quite large which means different ages of people will be the customers of the company. In addition, the mean age for the traditional customer is about 46 which is the same as the median. The income range for traditional people is between 41880 and 195570 dollars. This is a quite large difference between the two sides of people. In addition, the mean is 73181.99 dollars and is close to the median income of 65829 dollars. And different incomes bring different purchasing power. This may be the reason that affects the selling of products.
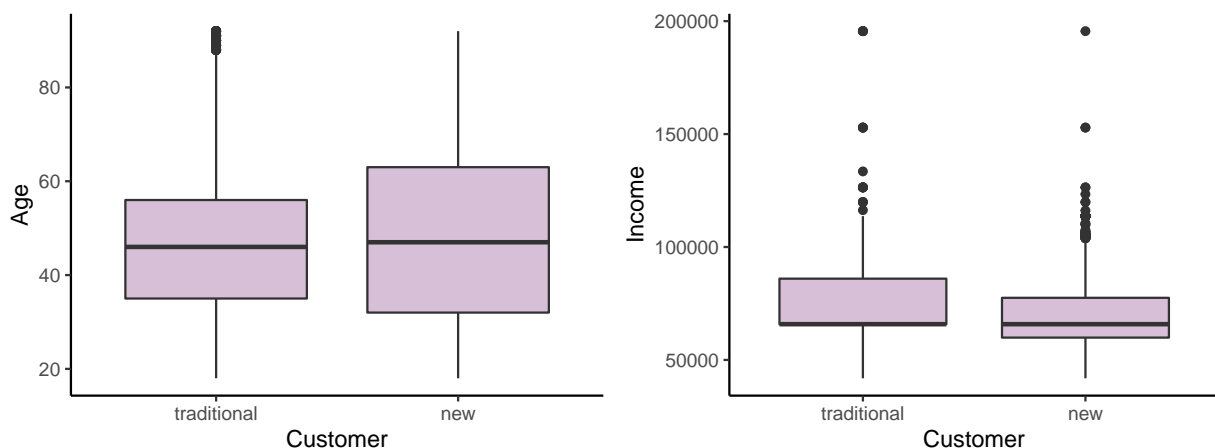
Figure 7: Boxplot of Age for New and Traditional Customers

The figure compares the numerical variable age and income between traditional and new customers. The first box plot shows that the age range for a traditional and new customers is similar. Both of them are between about 20 to 90 years old. And they also have a similar median age which is about 47. However, traditional customers, gather closer than new customers. Because for the main 50% of customers are between 35 and 57 years old. As for new customers, about from 32 to 63 years old takes 50% of total new customers. Which is a quite larger range of people in comparison to the traditional customers. The second box plot shows the two groups of people's incomes. They all have a similar range which is from about 45000 and 190000 dollars. However,

for traditional customers, the first quarter income is almost the same as the median which is about 63000 dollars. And the third quarter is about 87000 dollars. And new customer' first quarter is about 60000 dollars. And the third quarter is about 75000 dollars. This means that traditional customers may have stronger purchasing power compared to new customers. And household median income will be another important factor to distinguish between new and traditional customers.

Table 8: Numerical Summary of Population for Customers

Variables	Mean	Median	IQR	Min	Max
New Customer Population	1519423	1704694	2100085	3914	2731571
Traditional Customer Population	1478851	1239220	2026327	2803	2731571

This table shows the population of new and traditional customers. The minim population for new customers is 3914 which means that the new customer lived in a certain area based on the CSDuid, there are only 3914 people. And the maxim population for new customers is 2731571 which means that the maxim population for an area has 2731571 people. The areas are tracked by CSDuid.In addition, the mean for the new customer population is 1519423. As for the traditional customers population, the range for the population is between 2803 and 2731571. This means that the traditional customers are lived in an address that has 2803 people minim and 2731571 maxim based on the CSDuid classification.
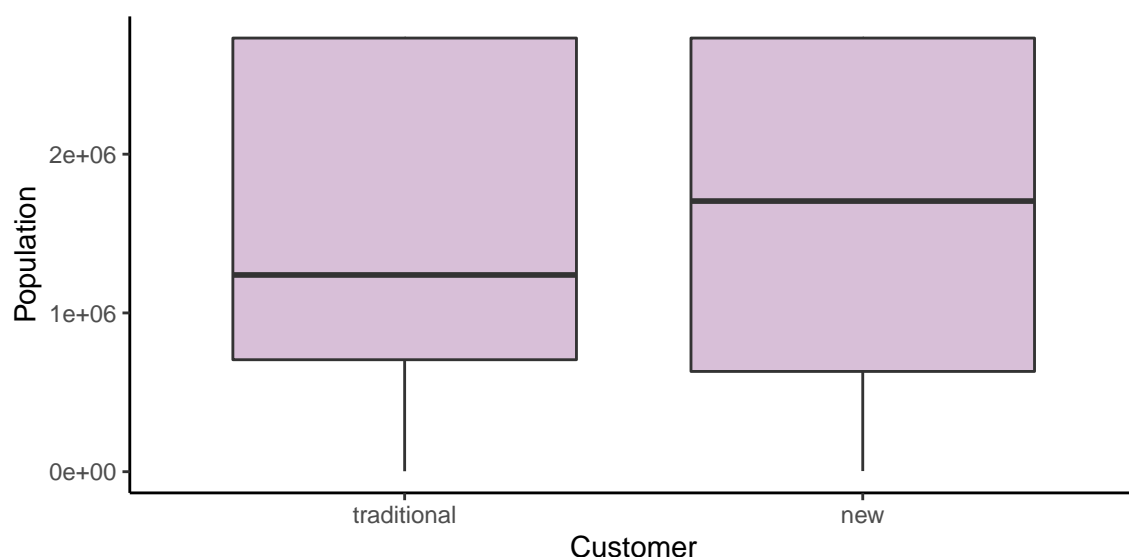

Figure 8: The Population for New And Tredictional Customers

From the figure, we know that the first and third quarter for the population is similar to new and traditional customers which are about 700000 and 2800000. However, the median between these two groups is different a lot. For traditional customers, the median is about 1200000. And the new customer's population median is about 1700000. The reason that may lead to this consequence is the CSDuid. The different CSDuid means the different addresses. And less population may represent the brownstone district. Thus, CSDuid may be one of the important factors for discriminating the different groups of customers.
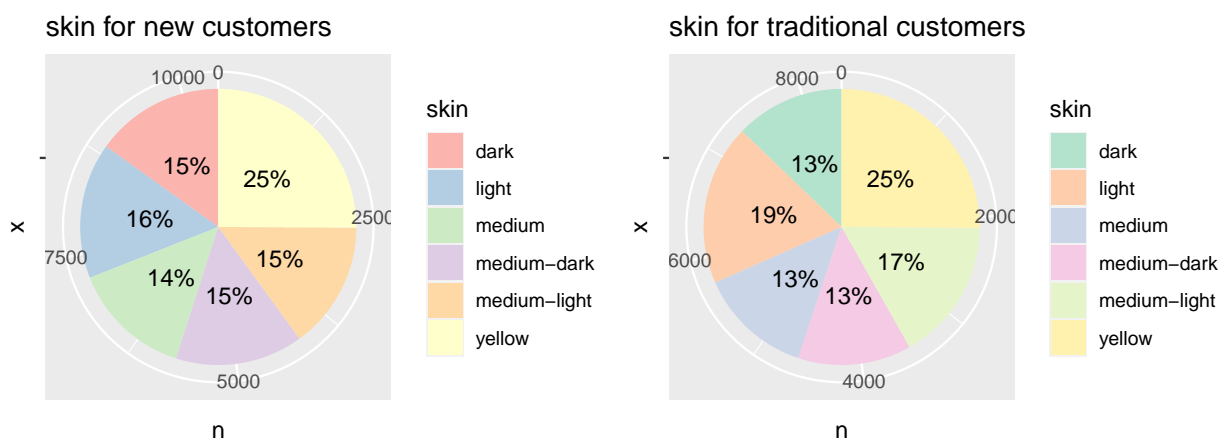

Figure 9: The Emoji Skin for New And Tredictional Customers Using

The figure shows the percentage of the different emojis skins that the customers use for chat. Different colors represent the different emojis' skins color. And almost all the skins used by the new customers make up 15%. In addition, there are 25% of new customers do not set up the emojis skins. As for the traditional customers, 19% of them use light emoji skins and 17% of traditional customers use medium-light skins. As for the rest three different skins color, all makeup about 13%. In addition, 25% of the traditional customers do not set up the skins. Since both groups of customers have 25% unsigned skins, this variable will not be considered in our model.

Table 9: device name proportion

iDOL	Run	Run7	Run7Plus	Run875	Run875X	RunBE	RunHYYH	RunLeader	RunON
155.000	3	43.000	71.000	626.000	77.000	3696.000	1	2	3713.000
0.018	0	0.005	0.008	0.075	0.009	0.441	0	0	0.443

The tables show all the products that the traditional customers purchased before. And it shows

that products Run ON and Run BE are the most popular products. For Run ON products, 3713 traditional customers bought which takes 44.3% of all sold products. And 3696 products Run BE are sold which is about 44.1%. As for the others, they all sold about 1% to 2%. Thus, Run ON and Run BE will be the most representative products for traditional customers.

For the comparing on the device between new and traditional customers are below:
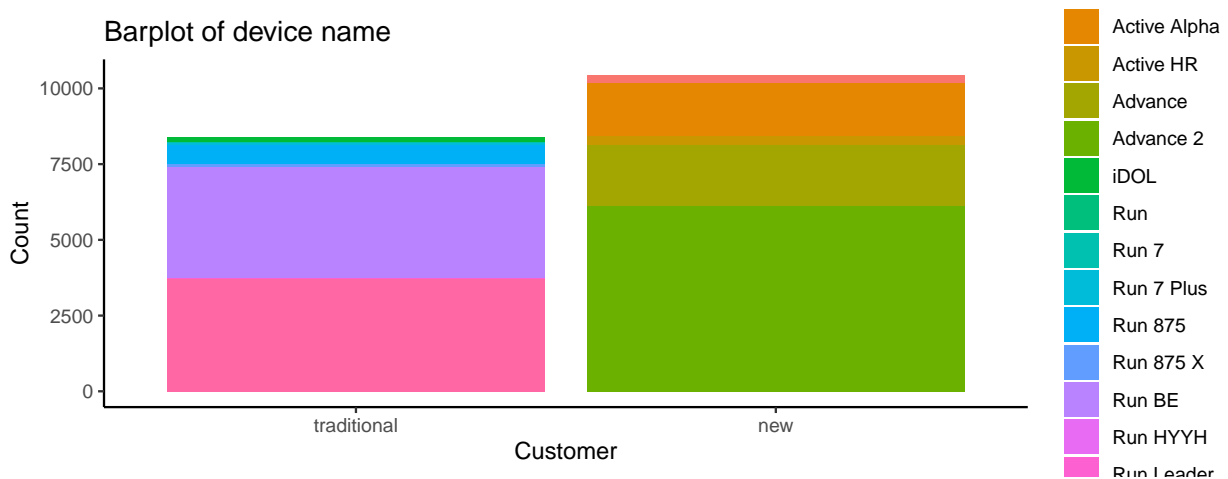
Figure 10: Device Comparison Between New and Traditional Customer

This figure shows that the traditional and new customers all bought the different types of products. New customers bought more than 10000 products in total and traditional customers bought about 8000 products in total. The different color represents the different products. And this graph shows that both groups of customers bought the different products without any intersection. And for traditional customers, products Run ON and Run BE make up the most areas which are closed to 7500. As for new customers, iDOL and Advance are more popular and the total number of both products is almost equal to the total number of all the traditional products purchased. The different products have different functions and recommended prices, these may be the key for customers choosing the different products.

Figure 11: The Products' Functions for New And Tredictional Customers

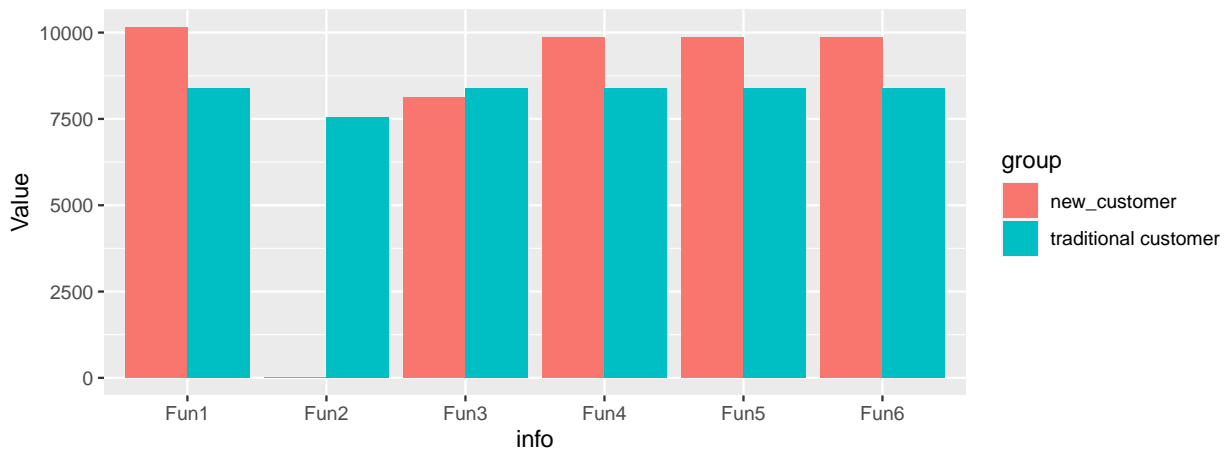The figure shows the all functions that new customers and traditional customers enjoyed in their devices. Fun1 represents as Heart Rate Sensor function. Fun2 represents Pulse Oximiterr. Fun3 represents GPS. Fun4 represents Sleep Tracking and Fun5 represents Smart Notifications. As for Fun6, it represents Contactless Payments. This bar plot shows that more new customers' products have Heart Rate Sensor, Sleep Tracking, Smart Notifications, and Contactless Payments functions than traditional customers. However, Pulse Oximiterr function is only used by traditional customers. As for function GPS, both group customers have a similar number which is about 8100. Thus, the product's functions will be another important factor to represent the different groups.

Model set up

Based on the plot and graphs analysis above, age, sex and income are important variables since they have different. It is because for age, although they have the same range from 18 to 92, the first quantile and third quantile are quite different. The new customer has a larger interquartile range (IQR) which means the age distribution is quite dispersive. So, it is meaningful to study. For the sex, based on the plot, the number of females increases from 4971 to 6110 which is greater than the change of males. Therefore, sex is an important factor in the basic model. For the income, the new and traditional customers have a different interquartile range (IQR) which means the concentration of income for customers is different. Therefore, in the basic model, sex, age, and income are the three important factors with the response variable being the type of customers (new and traditional).

In this study, we use the generalized linear model (GLM) because the response variable is binary and there are two types of customers. In addition, the impact of random effect is very small

which means we can ignore it. So, in this study, the random effect is unnecessary. Based on this model, it can show the correlation between response variables and factors. In the following study, we set up 8 models. Add one more variable through the 1 to 8 model. Then, use the likelihood ratio test to test each model and get the best model among them.

Create the Model:

Table 10: The Likelihood ratio test

Model	Pr(>Chisq)	Better model
1 & 2	< 2.2e-16 ***	model 2
2 & 3	0.0481 *	model 3
3 & 4	0.0113 *	model 4
4 & 5	< 2.2e-16 ***	model 5
5 & 6	< 2.2e-16 ***	model 6
6 & 7	< 2.2e-16 ***	model 7
7 & 8	1	model 7

From model 1 to model 7, each model adds one more variable and the added variable is the performance of the device name. Then, use the likelihood ratio test to compare models. We set H0 as the mean of added variables is zero. The Ha means the mean value is not 0. So, if the p-value is greater than 0.05, then, we have no evidence against the null hypothesis which means there is no difference between the full model and reduce model. So, under this condition, we choose the reduced model. If the p-value is less than 0.05, then we have strong evidence against the null hypothesis which means the mean value is not equal to 0. So, the full model is better than reduce model.

By comparing two models using the likelihood ratio test, we have the conclusion that model 7 is the best among the 8 models. It is because by comparing models, the p-value is always less than 0.05 except for the comparison between model 7 and model 8. The p-value of Models 8 and 7 is greater than 0.05 which means the reduced model (model 7) is better. Therefore, model 7 is the best and it is the final model we choose. There are nine variables in model 7 which are sex, age, income, smart notifications, heart rate sensor, contactless payments, GPS, pulse oximiterr, and recommended retail price. Based on the final model, there are three most important variables we would like to study in model7 which are sex, age, and income.

Summary table of final model

Table 11: Summary table of final model (Model 7)

term	estimate
Intercept	2.317e-01
sexFemale	1.917e-03
age	1.967e-04
household median inceom	-4.046e-06
Smart NotificationsNo	-2.027e+02
Heart Rate SensorNo	-3.432e+01
Contactless PaymentsNo	1.856e+02
GPSNo	-3.713e+01
Pulse OximiterrNo	-8.225e+01
Recommended Retail Price	-8.574e-01

$$log(\pi_i/(1-\pi_i)) = \beta_0 + \beta_1 sex_{female_i} + \beta_2 age_i + \beta_3 income_i + \beta_{4\,Smart\,notifications\,No_i} + \beta_{5\,Heart\,rate\,sensor\,No_i}$$

$$+ \beta_{6\,Contactless\,payments\,No_i} + \beta_{7\,GPS\,NO_i} + \beta_{8\,Pulse\,oximiterr_i} + \beta_{9\,Recommended\,retail\,price_i}$$

As we want age and income equals 0 are meaningful, modify age to age minus mean age. And modify income to income minus mean income. If age is 0 means age at mean.

- i represents the ith customer' age, sex and income.

- β_0 is 2.317e-01. It means the log odd of being a new customer for a 47 years old male with income is 70762.

- β_1 means keep sex and income fix, the log odd of being a new male customer compared to female customer increased by 0.1919%.

- β_2 means keep age and income fix, when a consumer's age increase one unit, the log odd of being a new customer increased by 0.0197%.

- β_3 means keep sex and age fix, the when a consumer's income increase one unit, the log odd of being a new customer decreased by 0.0004%.

Table 12: This model is about the sex for new and traditional customers

Coefficients	Estimate	Std. Error	Pr(>
Intercept	46.5160	0.1841	< 2e-16 ***
new_customernew	1.4356	0.2473	6.51e-09 ***

$$Age = \beta_0 + \beta_{1NewCustomer}$$

-Age: The customers' age.

-β_0: The expected value of age for traditional customers which is 46.5160 years old.

-β_1: The expected value of age that new customers are greater than traditional customers which is 0.1841 years old.

New customers' age is greater than the traditional customer 1.4356 years old.

Table 13: This model is about the age for new and traditional customers

Coefficients	Estimate	Std. Error	Pr(>
Intercept	73182.0	159.8	<2e-16 ***
new_customer	-4364.1	214.6	<2e-16 ***

$$HouseholdMedianIncome = \beta_0 + \beta_{1NewCustomer}$$

-Household Median Income: The customers median income in the certain ares that identified by postcode.

-β_0: The expected value of household median income for traditional customers which is 73182.0 dollars.

-β_1: The expected value of household median income that new customers is less than traditional customers' which is less than 4364.1 dollars.

New Customers' household median income is less than 4364.1 dollar comparing to the traditional customers' household median income.

The performance of devices when worn by customers

Table 14: The occurrences of flag per duration for each skin

Skin of custmers	Number of flag occurrences per duration
dark	0.0334151
light	0.0030657
medium	0.0099211
medium-dark	0.0202156
medium-light	0.0066490
yellow	0.0065309

Table summarizes the flag occurrences per duration on different customers' skin. According to the table, dark skin customers' devices possible performance 0.033 flags per sleep duration and almost 0.02 flag occurrences of medium-dark skin customers. However, the light-skin users' device will only perform a 0.003 flag per sleep duration and 0.066 flag occurrences for medium-light skin customers. By comparison, customers with dark skin have the highest flags occurrences per sleep duration, and customers with light skin show the lowest flag occurrences per sleep duration. There is a significant difference and almost a tenfold difference between dark skin users' devices and light skin users' devices. Thus, we can consider that the devices may perform poorly for dark skin customers, but we cannot rule out any other factors that can affect device performance also.
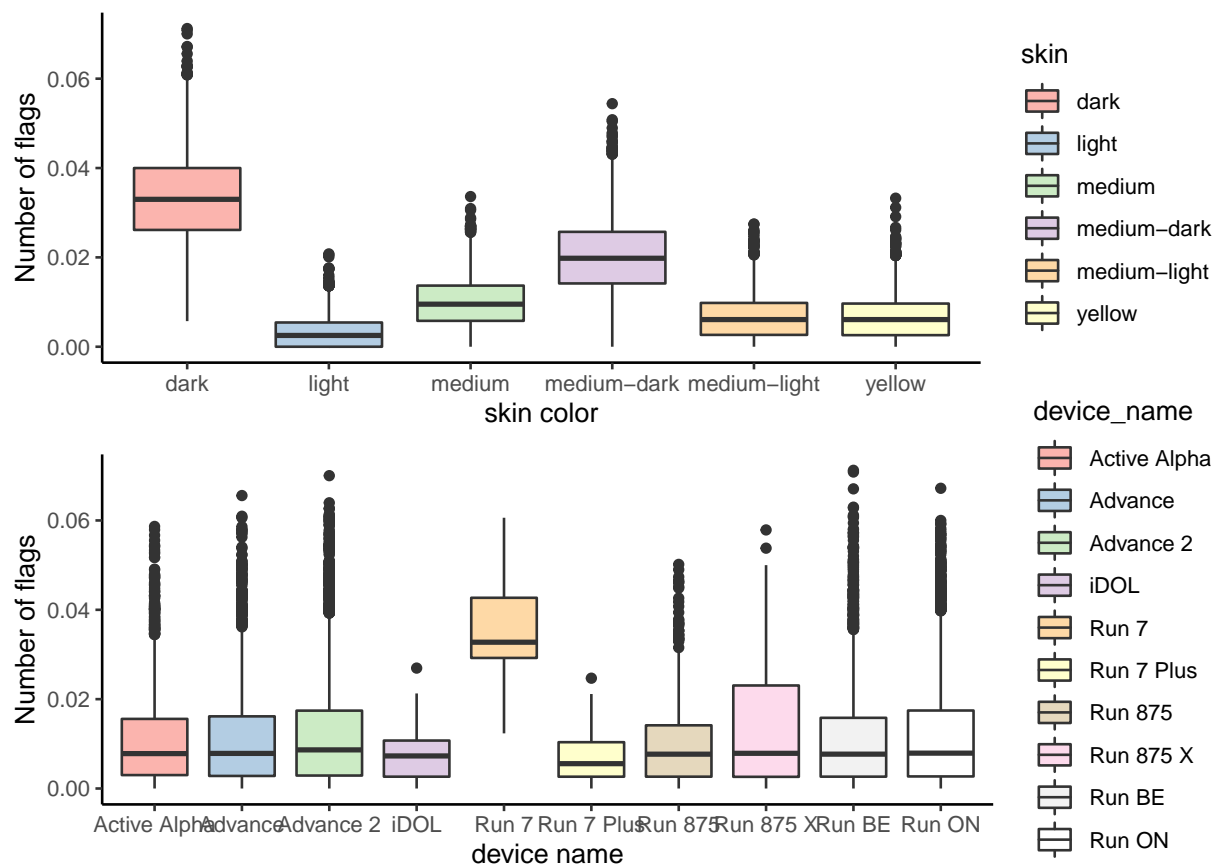
Figure 12: The distribution of flags on customers skin and devices

The top box plots indicate the distribution of median occurrences of flags on different customers' skin. We can observe the significant difference between dark skin customers and other skin customers. The median flag occurrences of devices of dark skin customers are 0,03 times which is the highest, and the lowest median flag occurrences are 0.002 times which is the light skin customers' devices. By comparing the interquartile ranges, the interquartile range for dark skin customers is more significant than others, especially the light skin customers interquartile range. For dark skin, the larger the interquartile range the flag occurrences data dispersed widely and shows the symmetric distribution. By contrast, the interquartile range of light-skin customers is smaller, representing the flag occurrences data for light-skin customers' devices less dispersed and showing the right-skewed distribution. Most of the flag occurrences are less dispersed and show the right-skewed distribution. Thus, the box plot indicates the approximate data distribution and can reflect the difference in device performance for different skin customers.

These plots at the bottom figures show the distribution of flag occurrences of different devices. Obviously, the device of Run 7 shows the highest median flag occurrences, and other devices show similar median flag occurrences. The Run 7 device shows round 0.035 flag occurrences

per sleep duration, and other devices show round 0.008 flag occurrences per sleep duration. Moreover, flag data of Run 875 device disperse more than other devices' flag occurrences data since the interquartile range is larger than other devices. For the Run 7 device, the data distribute symmetric, and other devices flag data showing right-skewed distribution. Thus, we can assume the different device shows different flag occurrences, and the Run 7 device may be performing poorly (shows flag occurrences highly) for customers.
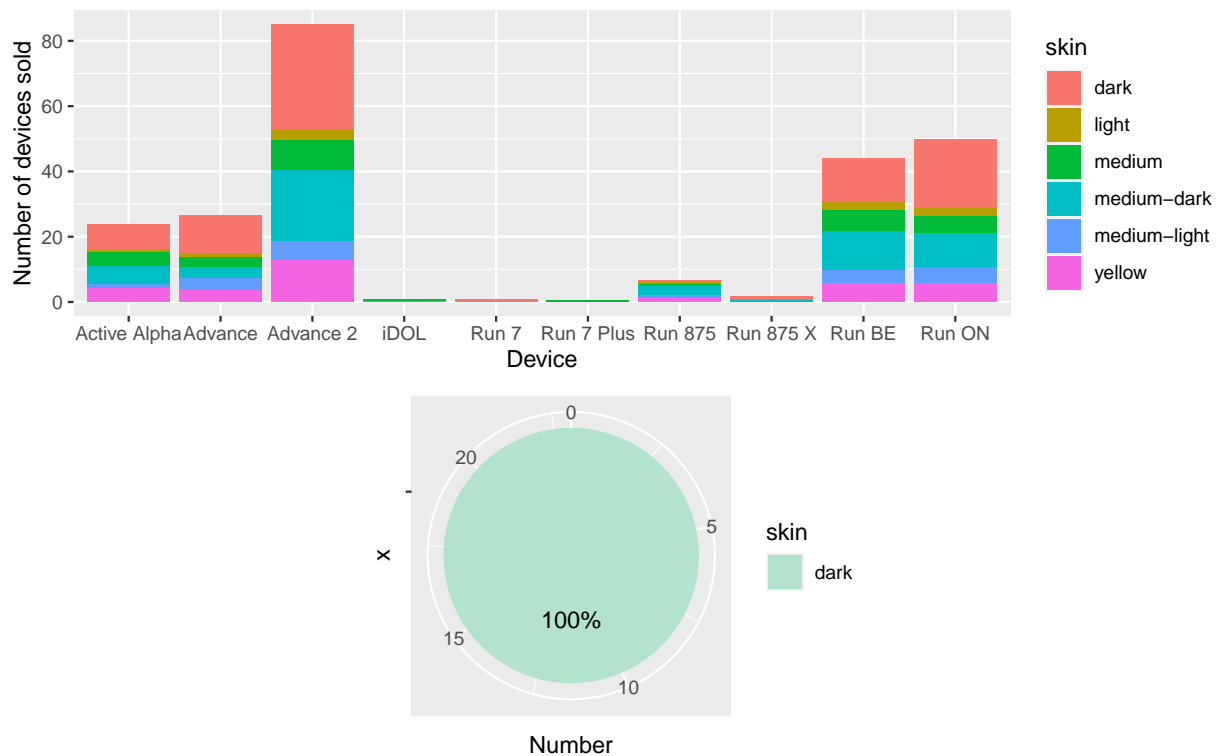
Figure 13: The distribution of customers skin on different devices

The figure contains a bar plot of devices and a pie chart of device Run 7. This bar plot shows that dark-skin customers occupied more on most devices than other skin customers. By comparing the distribution of skin customers on each device, we found that the Advance 2 has the highest sales and dark-skin and medium-dark customers occupied the most significant proportion, around 30% respectively. Moreover, there are some devices sold less. For the iDoL device, medium-dark customers prefer to buy this product, and it seems like only dark-skin customers buy the Run 7 device. Moreover, we know that Run 7 shows the largest median flag occurrences per sleep duration from the above box plot and devices performing poorly for dark-skin customers. Thus, we would like to know whether only dark-skin customers purchase the Run 7 device by the pie chart. From the pie chart, we know that it is true that only dark-skin customers purchase the Run 7 device.

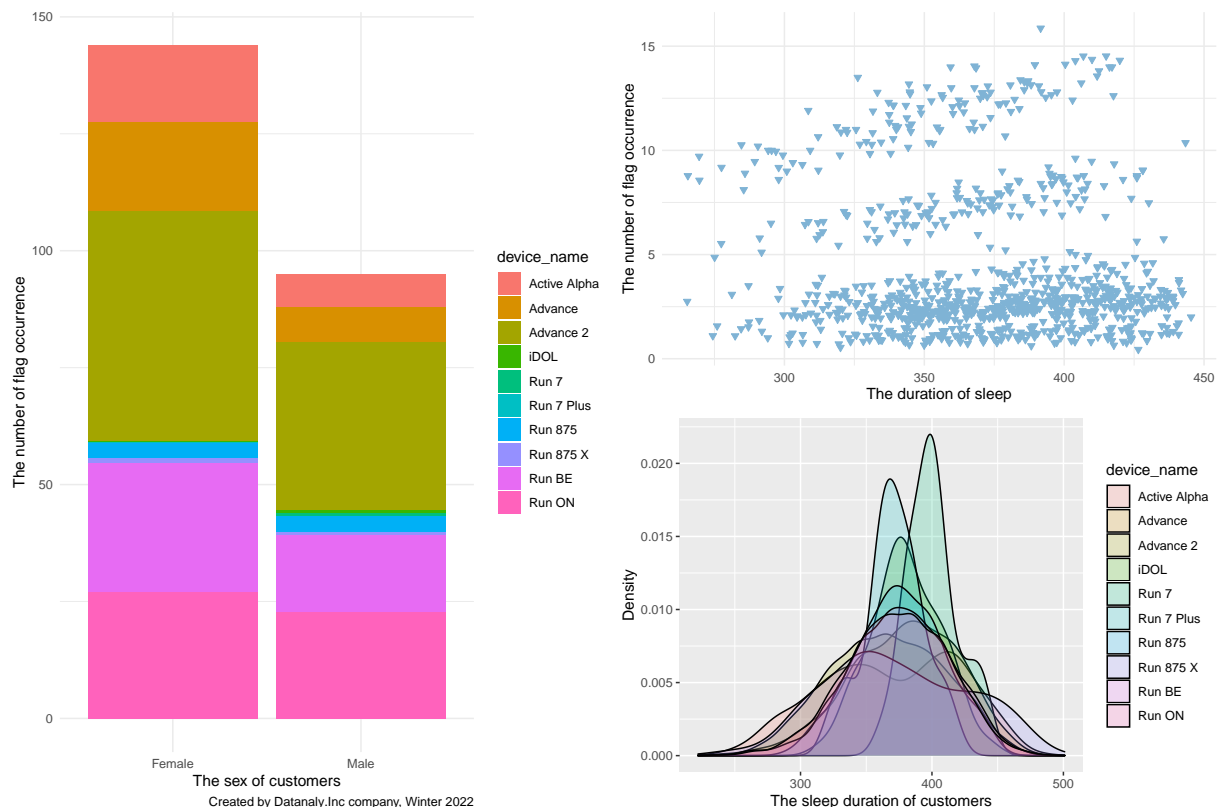**Figure 14:** The distribution of sleep duration of devices and flag occurrences on customers' sex

The left side bar plot shows the flag occurrences on customers' sex and contains the device occupation. It is obviously that the female customers' devices show more flag occurrences than male customers' devices. For female customers' devices, almost 140 flags occurred, and the Advance 2 device occupied more flag occurrences than other devices. For male customers, there are around 95 flags appeared, and the Advance 2 device also occupied more flag occurrences. By this comparison, we can assume that sex can be one of the factors that can influence flag occurrence and prove that different devices show different numbers of flag occurrences. The devices can also be one of the factors that can affect the flag occurrence.

The bottom right density plot indicates the density of the flag on each device. It is noticed that the peak of density for all device are around 380 minutes, and all devices are unimodal and contain only one peak, and show symmetric shape. We can observe that the Run 7 and Run 7 plus device show smaller variation of duration data, and other devices show a large variation in customers' sleep duration.

The upper right plot distributes the mean sleep duration and the flag occurrences in different areas. The postcode represents the customers' living area. We can observe a positive increasing trend that the number of flag occurrences will increase with the increase in sleep duration. It

is noticed that customers living in different areas show similar sleep duration. In general, the customers who live in the same postcode represent they may have similar incomes and skin. Thus, we can assume the postcode can also be one factor that can affect the flag occurrences.

Model set up

We can use the model to discover any factors that can influence the device performance. We would like to research what factors can affect the number of flag occurrences. Thus, we set the flag occurrences per sleep duration as the response variables and use the Poisson distribution. Moreover, we would like to include the customers' skin, sex, device as the factors and we use customers' id represent the virtual postcode since the postcode are customers' privacy and too specific. We consider the virtual postcode are not independent variables. Then, we set the customers' id (virtual postcode) as random intercept and the response variable is not follows normal distribution. Thus, we need to use the **Poisson Generalized Linear Mixed Model**.

We set several models that contains different factors and they are nested with other models. In order to select the most appropriate model, we use likelihood ratio test to compare these models and then determined the final model. The likelihood ratio test can help use determined the "best" model between the two nested models. We set 5 models that model 1 be the simplest model and add factors for the following models one by one that can include all factors may will effect flag occurrences we discussed above . For all models, we set the number of flag occurrences as the response variable. For model 1, we only include the customers' skin as the fixed effect and the virtual postcode as the random effect. Then, we add the device as the fixed effect for model 2 since we observed above that the flag occurrences shows different in different device. For model 3, we include the product age as fixed effect which is total time of sleep session of device. For model 4, we include the customers' sex as fixed effect since we observed the difference of flag occurrences between female and male customers.

Table 15: Summary table of likelihood ratio test

Model comparison	Result of likelihood ratio test	Best model
Model 1 & Model 2	p-value: 0.005675 (< 0.05)	Model 2
Model 2 & Model 3	p-value: 0.5407 (< 0.05)	Model 2
Model 2 & Model 4	p-value: 0.6827 (> 0.05)	Model 2

Then table shows the result of model comparison. Since likelihood ratio test only can compare two nested model, we only can compare two models step by step. By compare the model 1 and model 2, the p-value is 0.01029 which is smaller than 0.05. Thus, the model 2 is more appropriate which means the device is an important factor that can affect the flag occurrence and different device occurs different number of flag and. We keep compare model 2 and model 3, but we found the appropriate model still be model 2. Thus, we consider the product age is not the important factor that can influence the device performance. Finally, we compare the model 2 and model 4 since model 2 is the best model in previous comparison. And then, we found that the model 2 still be the most appropriate model and the customers' sex also is not the important factor that can affect the device performance. Thus, the model 3 is our final model. The final model contains fixed effects (customers' skin, device) and random effect (virtual postcode).

Table 16: Summary table of final model

Factors	Estimate	Std. Error	Pr(>
Intercept	-3.423645	0.011916	< 2e-16 ***
skin light	-2.392389	0.016511	< 2e-16 ***
skin medium	-1.209982	0.011308	< 2e-16 ***
skin medium-dark	-0.501080	0.009082	< 2e-16 ***
skin medium-light	-1.616847	0.012790	< 2e-16 ***
skin yellow	-1.630688	0.010630	< 2e-16 ***
device: Advance	0.043088	0.014833	0.00367 **
device: Advance 2	0.016230	0.012257	0.18546
device: iDOL	0.069193	0.063690	0.27730
device: Run 7	0.079534	0.056736	0.16097
device: Run 7 Plus	-0.112884	0.092448	0.22207
device: Run 875	0.051073	0.023005	0.02642 *
device: Run 875 X	0.129918	0.043507	0.00283 **
device: Run BE	0.028928	0.013394	0.03078 *
device: Run ON	0.031711	0.013182	0.01615 *

Final model

$$log(\frac{E(flags)}{duration}) = \beta_0 + \beta_1 skin : light + \beta_2 skin : medium + \beta_3 skin : medium\_dark + \beta_4 skin : medium\_light +$$

$$\beta_5 skin : yellow + \beta_6 device : Advance + \beta_7 device : Advance2 + \beta_8 device : iDOL + \beta_9 device : Run7 +$$

$$\beta_{10} device : Run7Plus + \beta_{11} device : Run875 + \beta_{12} device : Run875X +$$

$$\beta_{13} device : RunBE + \beta_{14} device : RunON + U_i$$

- $\frac{flags}{duration}$: The number of flag occurrences per duration.

- $\hat{\beta}_0$: The expect value of flag occurrences per duration is 0.033 for the dark skin customers with device Active Alpha.

- $\hat{\beta}_1$: The mean of flag occurrences per duration for light skin customers with device Active Alpha is 0.091 times of the mean for the dark skin customers with device Active Alpha during the same duration period. Thus, the mean of the flag occurrences per duration for light skin customers with device Active Alpha is 0.003003.

- $\hat{\beta}_2$: The mean of flag occurrences per duration for medium skin customers with device Active Alpha is 0.298 times of the mean for the dark skin customers with device Active Alpha during the same duration period. Thus, the mean of the flag occurrences per duration for medium skin customers with device Active Alpha is 0.009834.

- $\hat{\beta}_3$: The mean of flag occurrences per duration for medium dark skin customers with device Active Alpha is 0.606 times of the mean for the dark skin customers with device Active Alpha during the same duration period. Thus, the mean of the flag occurrences per duration for medium dark skin customers with device Active Alpha is 0.019998.

- $\hat{\beta}_4$: The mean of flag occurrences per duration for median light skin customers with device Active Alpha is 0.199 times the mean for the dark skin customers with device Active Alpha during the same duration period. Thus, the mean of the flag occurrences per duration for median light skin customers with device Active Alpha is 0.006567.

- $\hat{\beta}_5$: The mean of flag occurrences per duration for yellow skin customers with device Active Alpha is 0.196 times the mean for the dark skin customers with device Active Alpha during the same duration period. Thus, the mean of the flag occurrences per duration for median light skin customers with device Active Alpha is 0.006468.

- $\hat{\beta}_6$: The mean of flag occurrences per duration for dark skin customers with device Advance is 1.044 times the mean for the dark skin customers with device Active Alpha during the same duration period. Thus, the mean of flag occurrences per duration for dark skin customers with device Advance is 0.0034

- $\hat{\beta}_7$: The mean of flag occurrences per duration for dark skin customers with device Advance 2 is 1.016 times the mean for the dark skin customers with device Active Alpha during the same duration period. The mean of flag occurrences per duration for dark skin customers with device Advance 2 is 0.0335.

- $\hat{\beta}_8$: The mean of flag occurrences per duration for dark skin customers with device iDOL is 1.072 times the mean for the dark skin customers with device Active Alpha during the same duration period. The mean of flag occurrences per duration for dark skin customers with device iDOL is 0.0353.

- $\hat{\beta}_9$: The mean of flag occurrences per duration for dark skin customers with device Run 7 is 1.083 times the mean for the dark skin customers with device Active Alpha during the same duration period. The mean of flag occurrences per duration for dark skin customers with device Run 7 is 0.0357.

- $\hat{\beta}_{10}$: The mean of flag occurrences per duration for dark skin customers with device Run 7 Plus is 0.893 times the mean for the dark skin customers with device Active Alpha during the same duration period. The mean of flag occurrences per duration for dark skin customers with device Run 7 Plus is 0.0295.

- $\hat{\beta}_{11}$: The mean of flag occurrences per duration for dark skin customers with device Run 875 is 1.052 times the mean for the dark skin customers with device Active Alpha during the same duration period. The mean of flag occurrences per duration for dark skin customers with device Run 875 is 0.0347.

- $\hat{\beta}_{12}$: The mean of flag occurrences per duration for dark skin customers with device Run 875 X is 1.139 times the mean for the dark skin customers with device Active Alpha during the same duration period. The mean of flag occurrences per duration for dark skin customers with device Run 875 X is 0.0376.

- $\hat{\beta}_{13}$: The mean of flag occurrences per duration for dark skin customers with device Run BE is 1.029 times the mean for the dark skin customers with device Active Alpha during the same duration period. The mean of flag occurrences per duration for dark skin customers with device Run BE is 0.0339

- $\hat{\beta}_{14}$: The mean of flag occurrences per duration for dark skin customers with device Run ON is 1.032 times the mean for the dark skin customers with device Active Alpha during the same duration period. The mean of flag occurrences per duration for dark skin customers with device Run ON is 0.034.

- U_i: The random effects follows $N(0, \sigma_u^2)$

Discussion

Based on the data analysis, we figure out that about 60% of new customers are females and new customers' income is 65000 dollars for 3900 new customers. And the most popular product is Advance 2 for new customers. On studying the differences between new and traditional customers, based on the table and plots, we find out that sex, age, and income are the most significant variables for new and traditional customers. Moreover, the customer preference for product performance is the main aim of studying. So, set up 8 generalized linear models (GLM) by adding the performance one by one. Using the likelihood ratio test gets the final model which is model 7. Based on the final model, we focus the study, sex, age and income. The other factors are used to better fit the model. The conclusion is the difference in age, sex, and income between new and traditional customers is not significant. If we keep sex and income fixed, the log odd of being a new male customer compared to female customer increased by 0.1919%. If we keep age and income fixed, when a consumer's age increase one unit, the log odd of being a new customer increases by 0.0197%. And if we keep sex and age fixed, the when a consumer's income increase one unit, the log odd of being a new customer decreased by 0.0004%.

By the research above, we find that, with the same duration and the same device, there are more flag occurrences as the color of skin darkens. Therefore, indeed, our device does not work well on dark skin customers. And in the process of data analysis, we found that the average flag occurrences of the device RUN 7 is significantly higher than that of other devices. But we know from the summary of the model that all types of devices have a similar effect on flag appearances. We also found in the pie chart that 100% of the customers who bought run7 were dark skin people. So it may because our device work poorly on dark skin, and the users of RUN 7 are all with dark skin, which causes the average flag occurrences of RUN 7 to be higher than other devices.

Strengths and limitations

During this research, for the first question, we analyze the new customers and the difference between new and traditional customers in choosing products successfully. Moreover, we also confirm that Mingar company has indeed made great progress in outdoor equipment. And the products also becoming popular, which has helped a lot in expanding the market share. For the second question, we not only answered the questions required by the client, we also conducted further exploration of the data, for example, we found that RUN 7 has more flag occurrences. Then we did research on the issues we found and gave some explanations for them. We believe in this way, we can better help our clients to find out the potential problems in their product.

On analyzing the characteristics of new customers, there is a lot of missing value for sex. When

we study whether sex is significant for new customers, the result may not very accurate since we can only use the known sex to study. So, it is a limitation of data. When we research whether the device is performing poorly for dark-skin customers, we use the emoji to determine the customers' skin. It might occur some misleading since some customers are not set on the skin color of emoji. Thus, we cannot misidentify these groups of customers' skin and we classify them as yellow-skin. This may cause our estimate model less accurate than the model.

Consultant information

Consultant profiles

Xiaoqing Chen. Xiaoqing Chen is the senior consultant with Eminence Analytics. She is specializes in model construction and data visualization. In 2023, she get Bachelor of Science, major in Statistics and Genetic application from Univeristy of Toronto.

Jiahao Li . Jiahao Li is a junior consultant with Eminence Analytics. He specialize in reproducible analysis and model analysis. Jiahao Li earned his Bachelor of Science, Majoring in Statistic, Mathematics and Economics from the University of Toronto in 2023.

Xinjing Guo. Xinjing Guo is a junior consultant with Eminence Analytics. She specialize in making up models under different situations. And Visualizing complex data makes it easy for readers to observe. She will majoring in Computer Science and Statistics from the University of Toronto in 2023.

Xinxue Guo. Xinxue Guo is a senior consultant with Eminence Analytics. He specializes in creating the model based on the different data. He earned Bachelor of Science, Majoring in Statistic, Mathematics and Computer Science from the University of Toronto in 2023.

Code of ethical conduct

- Our company take job responsibility and provide objective and reliable information on any professional review or evaluation procedures.

- The most important ethic for our company is that we comply with relevant privacy laws or privacy standards established by the SSC or other relevant bodies and we strictly follow the procedures to protect human rights and dignity.

- The statistical practitioners in our company are ethical that they will not misuse or condone the misuse of data and they are responsible to protect and respect human rights and animal subjects.

- Our company will remain objective and strive to avoid procedural or personal bias. We believe that building effective data-based information is critical to informed public opinion and policy.

References

1. Grolemund, G. (2014, July 16) *Introduction to R Markdown.* RStudio. https://rmarkdown. rstudio.com/articles_intro.html. (Last Accessed: October 12, 2021)

2. Dekking, F. M., et al. (2005) *A Modern Introduction to Probability and Statistics: Understanding why and how.* Springer Science & Business Media.

3. Allaire, J.J., et. el. *References: Introduction to R Markdown.* RStudio. https://rmarkdown.rstudio.com/docs/. (Last Accessed: October 12, 2021)

4. Douglas Bates, Martin Maechler, Ben Bolker, Steve Walker (2015). Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software, 67(1), 1-48. doi:10.18637/jss.v067.i01.

5. Hadley Wickham (2021). rvest: Easily Harvest (Scrape) Web Pages. R package version 1.0.2. https://CRAN.R-project.org/package=rvest

6. Dmytro Perepolkin (2019). polite: Be Nice on the Web. R package version 0.1.1. https://CRAN.R-project.org/package=polite

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

8. Hadley Wickham and Evan Miller (2021). haven: Import and Export 'SPSS', 'Stata' and 'SAS' Files. R package version 2.4.3. https://CRAN.R-project.org/package=haven

9. von Bergmann, J., Dmitry Shkolnik, and Aaron Jacobs (2021). cancensus: R package to access, retrieve, and work with Canadian Census data and geography. v0.4.2.

10. H. Wickham. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2016.

11. Baptiste Auguie (2017). gridExtra: Miscellaneous Functions for "Grid" Graphics. R package version 2.3. https://CRAN.R-project.org/package=gridExtra

12. R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

13. Hadley Wickham, Romain Fran?ois, Lionel Henry and Kirill Müller (2022). dplyr: A Grammar of Data Manipulation. R package version 1.0.8. https://CRAN.R-project.org/package=dplyr

14. Hadley Wickham and Dana Seidel (2020). scales: Scale Functions for Visualization. R package version 1.1.1. https://CRAN.R-project.org/package=scales

15. H. Wickham. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2016.

16. Kamil Slowikowski (2021). ggrepel: Automatically Position Non-Overlapping Text Labels with 'ggplot2'. R package version 0.9.1. https://CRAN.R-project.org/package=ggrepel

17. GGPLOT2 point shapes. STHDA. (n.d.). Retrieved April 6, 2022, from http://www.sthda.com/english/wiki/ggplot2-point-shapes

18. Color chart. HTML Color Codes. (n.d.). Retrieved April 6, 2022, from https://htmlcolorcodes.com/color-chart/

19. GGPLOT2 - easy way to mix multiple graphs on the same page. STHDA. (n.d.). Retrieved April 6, 2022, from http://www.sthda.com/english/wiki/wiki.php?id_contents=7930

20. Yihui Xie, C. D. (2022, February 18). R markdown cookbook. 5.4 Control the size of plots/images. Retrieved April 6, 2022, from https://bookdown.org/yihui/rmarkdown-cookbook/figure-size.html

21. Mcleod, S. (2019, July 19). Box plots (also known as box and Whister plots). Box Plot | Simply Psychology. Retrieved April 6, 2022, from https://www.simplypsychology.org/boxplots.html

Appendix

Web scraping industry data on fitness tracker devices

First, I copy the url (https://fitnesstrackerinfohub.netlify.app/.) which get the web link from professor's hits. Then, I need to change add my university email address for the data using purpose. We do not change any numbers and data. And we do not show the private information of customers in our analysis.

```r
# These are the libraries I find useful for webscraping
library(tidyverse)
library(polite)
library(rvest)

#url link for the website
url <- "https://fitnesstrackerinfohub.netlify.app/."

# update informative user_agent details
target <- bow(url,
              user_agent = "xinxue.guo@mail.utoronto.ca for STA303/1002 project",
              force = TRUE)

# Any details provided in the robots text on crawl delays and
# which agents are allowed to scrape
target
```

```
## <polite session> https://fitnesstrackerinfohub.netlify.app/.
##      User-agent: xinxue.guo@mail.utoronto.ca for STA303/1002 project
##      robots.txt: 2 rules are defined for 2 bots
##    Crawl delay: 12 sec
##   The path is scrapable for this user-agent
```

```r
html <- scrape(target)

device_data <- html %>%
  html_elements("table") %>%
  html_table() %>%
  pluck(1) # added, in case you're getting a list format

head(device_data)
```

```
## # A tibble: 6 x 13
##   `Device name` Line     `Recommended retail pr~` `Battery life` `Water resitan~`
##   <chr>         <chr>                       <dbl> <chr>          <chr>
## 1 Run ON        Run                          350. Up to 21 days  Waterproof, 10 ~
## 2 Advance 2     Advance                      145  Up to 7 days   Resistant
## 3 Rush 5        Rush                         180. Up to 7 days   Waterproof
## 4 Active Alpha  Active                       100. Up to 7 days   Resistant
## 5 Run BE        Run                          300. Up to 14 days  Waterproof, 10 ~
## 6 Advance       Advance                      120  Up to 7 days   Resistant
## # ... with 8 more variables: `Heart rate sensor` <chr>, `Pulse oximiter` <chr>,
## #   GPS <chr>, `Sleep tracking` <chr>, `Smart notifications` <chr>,
## #   `Contactless payments` <chr>, Released <chr>, Brand <chr>
```

Accessing Census data on median household income

First, install the packages named cancensus. Then sign up for an API account by using my
Google mail address. After that, I can get a personal API key from my account. Copy and post
on the cancensus api key position. Finally, I have permission to access the data. We do not
change any numbers and data. And we do not show the private information of customers in our
analysis. We do not change any numbers and data. And we do not show the private information
of customers in our analysis.

```r
#install.packages("cancensus")
#the library that will use
library(cancensus)


#API key get based on the account information.
options(cancensus.api_key = "CensusMapper_23fa7456f28ecd8dd76ae32e6127278d",
        cancensus.cache_path = "cache") # this sets a folder for your cache



# get all regions as at the 2016 Census (2020 not up yet)
regions <- list_census_regions(dataset = "CA16")

regions_filtered <-  regions %>%
  filter(level == "CSD") %>% # Figure out what CSD means in Census data
  as_census_region_list()

# This can take a while
# We want to get household median income
```

```r
census_data_csd <- get_census(dataset='CA16', regions = regions_filtered,
                              vectors=c("v_CA16_2397"),
                              level='CSD', geo_format = "sf")

# Simplify to only needed variables
median_income <- census_data_csd %>%
  as_tibble() %>%
  select(CSDuid = GeoUID, contains("median"), Population) %>%
  mutate(CSDuid = parse_number(CSDuid)) %>%
  rename(hhld_median_inc = 2)
head(median_income)
```

```
## # A tibble: 6 x 3
##     CSDuid hhld_median_inc Population
##      <dbl>           <dbl>      <int>
## 1 1001101              NA         36
## 2 1001105           64640        150
## 3 1001113           53888        481
## 4 1001120              NA         66
## 5 1001124           66240       1625
## 6 1001126           61568        489
```

Accessing postcode conversion files

I get the Postal code conversion file web page by pressing the provided link. Then choose the certain year which I will use to explore. And in that certain year, I can choose the released month for downloading the data. I choose September 2021's sav data. After finishing downloading it, I put it into the data-row file for use. We do not change any numbers and data. And we do not show the private information of customers in our analysis.

```r
#install.packages("haven")
library(haven)
library(tidyverse)
dataset = read_sav("data-raw/pccfNat_fccpNat_082021sav.sav")
# from dataset, select two columns which are PC and CSDuid. In addition,  rename the
↪  PC to post code.
postcode <- dataset %>%
  select(PC, CSDuid) %>% rename(c("postcode" = "PC"))
```

```r
# show first 6 row about postcode data.
head(postcode)
```

```
## # A tibble: 6 x 2
##   postcode  CSDuid
##   <chr>     <dbl>
## 1 A0A1A0    1001144
## 2 A0A1B0    1001464
## 3 A0A1C0    1001557
## 4 A0A1C0    1001557
## 5 A0A1C0    1001557
## 6 A0A1C0    1001557
```

Table 17: Important variable

Variable	Variable type	Description
CSDuid	Character	The virtual postcode of customer
sex	Character	Biological sex
skin	Character	The customers skin
deviceName	Character	The device name of devices in this company
duration	Numeric	Duration, in minutes, of sleep session
flags	Numeric	Number of times there was a quality flag during the sleep session. Flags may occur due to missing data, or due to data being recorded but sufficiently unusual to suggest it may be a sensor error or other data quality issue
age	Numeric	The age for each customer
line	Numeric	Character
hholdmedianinc	Numeric	Customer's median income
HeartRateSensor	Character	Whether device name has heart rate sensor function
PulseOximiterr	Character	Whether device name has pulse oximiter function
GPS	Character	Whether device name has GPS function
SleepTracking	Character	Whether device name has sleep tracking function

Variable	Variable type	Description
SmartNotifications	Character	Whether device name has smart notifications function
ContactlessPayments	Character	Whether device name has contactless payments function
Brand	Character	Brand of products (Mingar & Bitfit)
newcustomer	Character	Define the customer (new & traditional)

Table 18:The summary table for devices that traditional customers bought

Device Name	HeartRate	PulseOximiter	GPS	SleepTrack	SmartNotes	ContactlessPayments	Price
iDOL	Yes	No	Yes	Yes	Yes	Yes	199.99
Run	Yes No	No	No	No	No		450.00
Run7	Yes	No	Yes	Yes	Yes	Yes	399.99
Run7P	Yes	Yes	Yes	Yes	Yes	Yes	435.00
Run875	Yes	No	Yes	Yes	Yes	Yes	350.00
Run875X	Yes	Yes	Yes	Yes	Yes	Yes	399.99
RunBE	Yes	Yes	Yes	Yes	Yes	Yes	299.99
RunH	Yes	No	Yes	No	Yes	No	420
RunLead	Yes	No	Yes	No	Yes	No	479.99
RunON	Yes	Yes	Yes	Yes	Yes	Yes	349.99

The table shows all the 10 different products that traditional customers purchased. With the comparison, the product Run 7Plus, Run 875X, Run BE, and Run ON have the full functions which are Heart Rate Sensor, Pulse Oximiterr, GPS, Sleep Tracking, and Smart Notifications, and Contactless Payments. However, the price for them is also higher than the other products which Run 7Plus is 435.00 dollars, Run 875X is 399.99 dollars, Run BE is 299.99 dollars and Run ON is 349.99 dollars. Based on the previous traditional product summary table, with the full functions and lower price, compared to the Run 7Plus and Run 875X, Run BE and Run ON are popular and are reasonable. As for the products Run and Run Leader, both do not have the full fix functions but the quite higher price. Thus, the traditional customers may not accept them.

Table 19: The summary table for products with functions and recommended price

Device Name	Heart Rate Sensor	Pulse Oximiterr	GPS	Sleep Tracking	Smart Notifications	Contactless Payments	Recommended Retail Price
Advance 2	Yes	No	Yes	Yes	Yes	Yes	145.00
Advance Alpha	Yes	No	No	Yes	Yes	Yes	99.99
Advance	Yes	No	Yes	Yes	Yes	Yes	120.00
Active	No	No	No	No	No	No	39.99
Active HR	Yes	No	No	No	No	No	79.99

Based on the table, all the devices do not have the Pulse Oximiterr function. Advance and Advance 2 are two products that have the most function of all products, which include Heart Rate Sensor, GPS, Sleep Tracking, Smart Notifications, Contactless Payments, and Recommended Retail Price. Since Advance 2 is the newer version product compared to Advance, the price for it is higher than Advance which is 145 dollars. And Advance recommended price is 120 dollars. Advance Alpha's price is 99.99 dollars and compared to the previous two products, it does not include a GPS function. As for Active HR, it only includes Heart Rate Sensor with a 79.99 recommended price. Active is the poorest one since it does not have any above functions and sold for 39.99 dollars. The functions and the price are all the important factors for customers to choose the products.