### **Understanding GCash Customers’ Reviews and Ratings**

By The Breakfast Club (DECSC 131)

1. **CONTEXT**

**Company Background**

GCash is one of the companies under Mynt, a FinTech startup partnership between Globe Telecom, Ayala Corporation, and Ant Financial. Mynt’s five key services revolve around payments, remittances, loans, business solutions, and platforms ([Mynt, n.d.](https://www.mynt.xyz/about-us/#:~:text=GCash,GCash%20American%20Express%20Virtual%20Pay)). GCash’s services also revolve around these solutions. Since its establishment in 2004, it has effectively been providing its mobile money services to its customers. The company is able to give Globe’s mobile phone subscribers, both local and international, access to a cashless and cardless transaction. Customers are entitled to online money remittance, donations, disbursement of salaries, and payment of bills, products, and services ([Unsworth, 2010](https://www.mobilepaymentsworld.com/case-study-gcash/)). Nowadays, especially in the context of the pandemic, people have been using GCash to facilitate online transactions so they would no longer have to go out to pay for their needs and wants.

Currently, the company, specifically its marketing team, drives its decisions based on qualitative insights generated from reading user reviews about their app. Most of their evidence is anecdotal and they are in need of a structured way of looking at their data to create more informed decisions when improving their service.

**Dataset Background**

For this project, the two datasets provided contain both quantitative and qualitative information about the GCash app reviews of Apple and Android phone users from March 2012 to April 2021. Data from both sets involved ratings, reviews, usernames, and date posted. To give more context, users have reviewed the app based on its features and usability. They would give a rating of 1 to 5 depending on how satisfied they are with the app and provide comments (reviews) to justify their rating.

**The Problem Being Tackled**

Given that the datasets contain information about users’ reviews and ratings, the group could use the insights generated from these to drive different strategies in improving the app’s services to keep existing users and even entice more users to use the app. They would have to look at the app’s strong points which the company can use in targeting potential customers. Furthermore, they would also need to analyze the reviews to pinpoint the app’s weak points and identify what features the company should improve and focus on for their succeeding updates on the app.

1. **SOLUTION**

**Brief Description of the Solution**

1. **Brief Summary:** A dashboard that shows the frequency of relevant words and proportions of topics from GCash customer reviews over a period of time.
2. **Visualization:** The frequency of words is visually represented by a line graph (x = time and y = count) and the distribution of topics is represented by a pie chart (colors = topic). Two pie charts are provided to represent topics from one-star ratings and five-star ratings.
3. **Interactive Elements:** The user may opt to adjust the slider to change the date range, which filters the data being visualized in the dashboard. There is also an interactive table containing the most frequently used words per rating within the period specified by the slider. To check the usage trend of a specific word, one may either select a word from the table or type in the search bar. Two search bars are provided to allow for comparing trends between two different words.

**Methodology**

1. **Process**

Setting up the data for use required1) combining the two datasets, 2) setting all words to lowercase, 3) removing English and Filipino stop words and non-word characters, and 4) tokenization to set the appropriate data format for the analytical tools to be used.

Afterwards, a text analytics approach utilizing word frequency and topic analysis was applied. Frequency distribution from the Natural Language Toolkit (NLTK) was used for the former, while Latent Dirichlet Allocation (LDA) was used for the latter. In addition to this, part of speech (POS) tagging was used to generate the data to be used by the interactive table. Shaping this data required ranking the word frequencies per rating, filtering out words that were not nouns, verbs, or adjectives, and removing the top 100 common words across all ratings.

For the topic modelling, special cleaning was made to remove meaningless words that impeded on the coherence of the topic models. Next, the combined dataset had to be split into two: one dataset containing data with rating = 1, and the other dataset containing data with rating = 5. Afterwards, coherence scores were generated from a list of LDA models across a range of two to ten topics (interval = 2). This was done to find the optimal number of topics for both datasets. In the end, the group decided from the coherence score graphs that four topics for both topic models would suffice. Then, looking at the top 10 words for each topic, the team had to create topic labels to make sense of them. To create the pie charts, the probability that each review belonged to a certain topic was summed up.

Finally, tying everything together is Dash, which required HTML and some CSS elements to ground the dashboard. The link between interactive elements (ie. table, slider, and search bars) and visualization elements (ie. line graph and pie charts) were also made possible through Dash, coupled with custom-made functions that made the process more efficient.

1. **Rationale**

The group merged the two datasets provided into a single dataset to generate meaningful insights from all reviews. Since GCash is targeting both Apple and Android gadget users, it would be beneficial for them to focus on the whole picture and analyze it in one go instead of just looking at one or the other as the features and app updates available for both users will generally be the same. The merged dataset may provide more insights for the company as it widens their perspective and this may help them form better strategic decisions that will address the concerns of both users.

The text analytics approach was deemed appropriate because the datasets involved user reviews. Since reviews vary from person to person, language processing techniques such as **stopword removal** and **tokenization** were used to structure the data. Meanwhile, **frequency distribution** was used to explore the words that appeared most frequently and to identify patterns among the reviews made by users across the different ratings. Lastly, the construction of reviews may vary from one user to another but may actually point out the same idea so the group used **topic modelling** to understand similarities among the words found in one-star and five-star reviews, while classifying them into relevant topics that the company can better interpret than singular words. This is in line with the aim of this project to generate insights that would help the company in marketing the strong points of the app to potential users as well as in addressing their weak points to improve their service.

The dashboard was chosen as the data product and is intended for the interactive use of the company’s marketing and product development teams. It was designed to display the appropriate visualizations that contain the relevant information they would need to make decisions regarding what features to improve on and what angle they can use to market the app. The clients may search for specific words to determine usage trends, hopefully providing them with insights on whether they are able to address previous issues or if there are recurring issues that need to be solved. They may also opt to choose certain periods they want to explore either on the line graph or pie chart and it will display the most frequently used words or common topics within the specified period which could be helpful in their preparations for succeeding app updates or marketing campaigns.

1. **Contextualization**

By applying word frequency and topic modeling on reviews, the group was able to create a dashboard that can be used to determine what specific areas GCash must improve on to satisfy their customers. We identify the dashboard to be relevant in two different areas: product development and marketing.

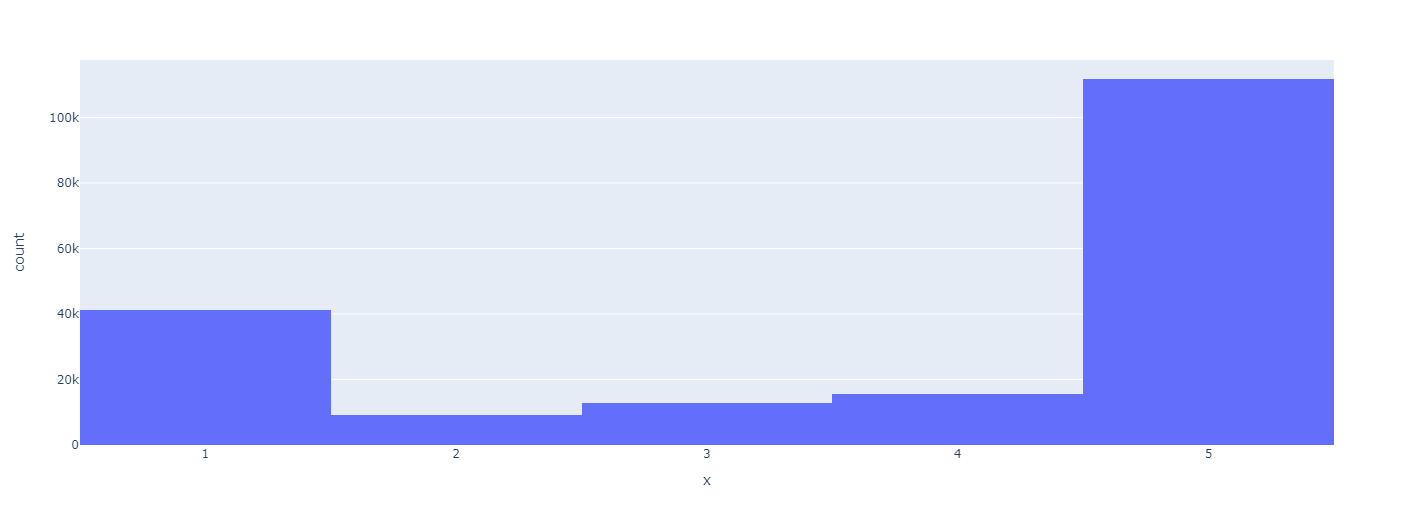
Product development is a crucial component for GCash because it provides numerous services to its customers. It is even more crucial today due to the pandemic, where more and more people are making digital transactions to avoid close contact ([Endo, 2020](https://asia.nikkei.com/Business/Companies/Digital-payment-grows-in-Philippines-amid-COVID-19-fears)). GCash needs to effectively update its app so users will continue to have a seamless and convenient experience. With this, the dashboard can give the product team a direction with regards to which app features they must improve on and what bugs they have to fix by looking at the word and topic trends.

On the other hand, marketing is also crucial as this is how the company could gather more customers to grow their business. Marketing requires GCash to know about their selling point from the perceptions of its customers. Therefore, by looking at word usage and topic trends of five-star app reviews, the marketing team can come up with advertisements and campaigns that would highlight the perceived strengths of the app. These marketing campaigns would prove effective because the messages conveyed are based on actual customer experiences.

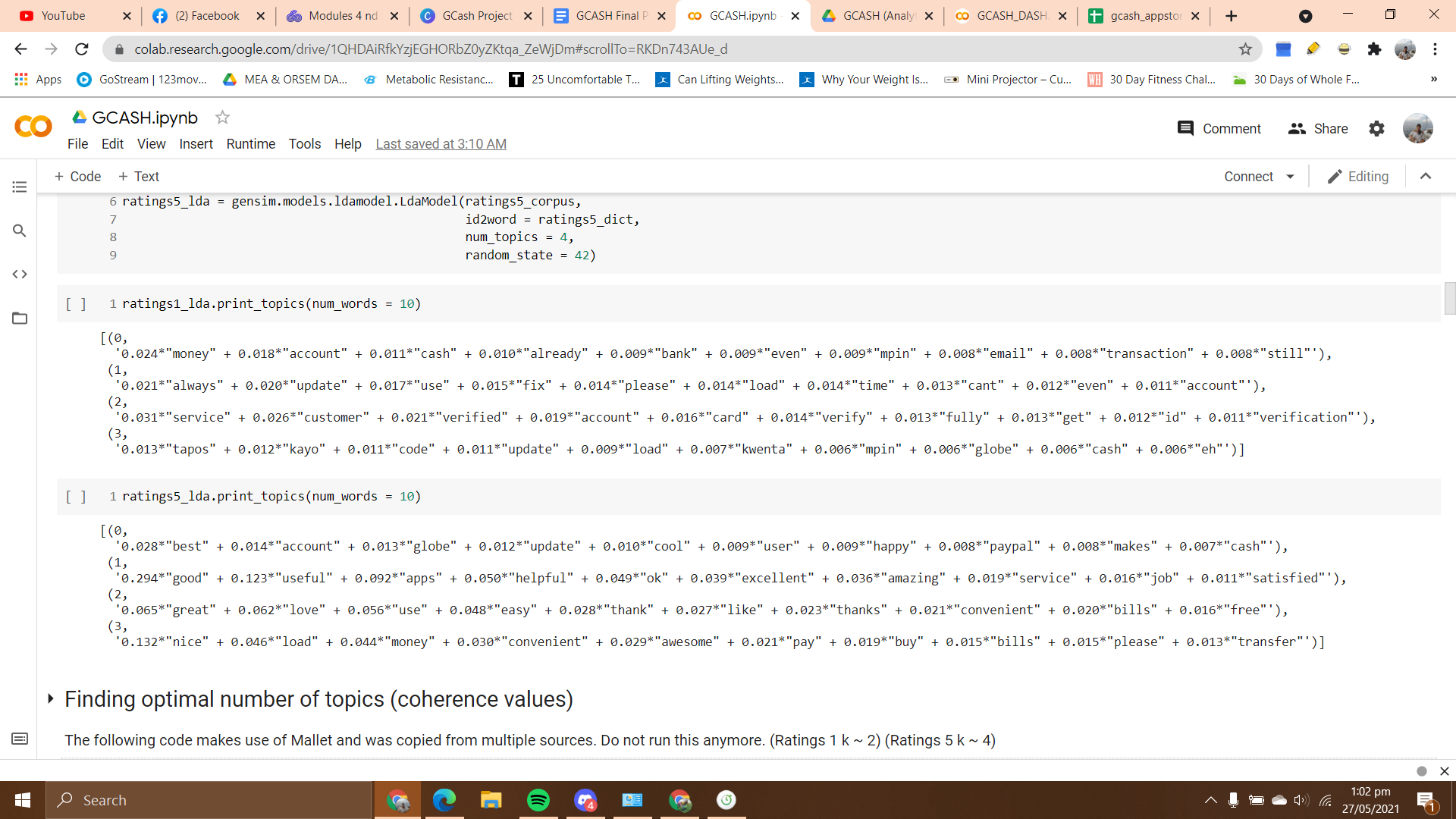
1. **IMPACT**

**Key Insights**

Distribution of ratings

****

Looking at the histogram of ratings, the distribution is concentrated at rating = 1 and rating = 5. Calculating the proportions, 58.7% of reviews rated the app a 5, while 21.6% rated the app a 1.

Topics generated from LDA

| **Topic** | **Rating 1** | **Rating 5** |
| --- | --- | --- |
| **1** | Customer Service | Easy and Fast Transactions |
| **2** | Service Reliability | App Usability |
| **3** | Log-in/Verification | Services Offered |
| **4** | Customer Dissatisfaction | Customer Satisfaction |

Topic Title Explanations

**Rating 1**

* *Customer Service* refers to how fast or often does GCash respond to customers’ messages
* *Service Reliability* refers to transaction errors/bugs
* *Log-in/Verification* refers to bugs when logging in, hassle verification process, mpin errors
* *Customer Dissatisfaction* refers to general discontent (bad descriptions about the app)

**Rating 5**

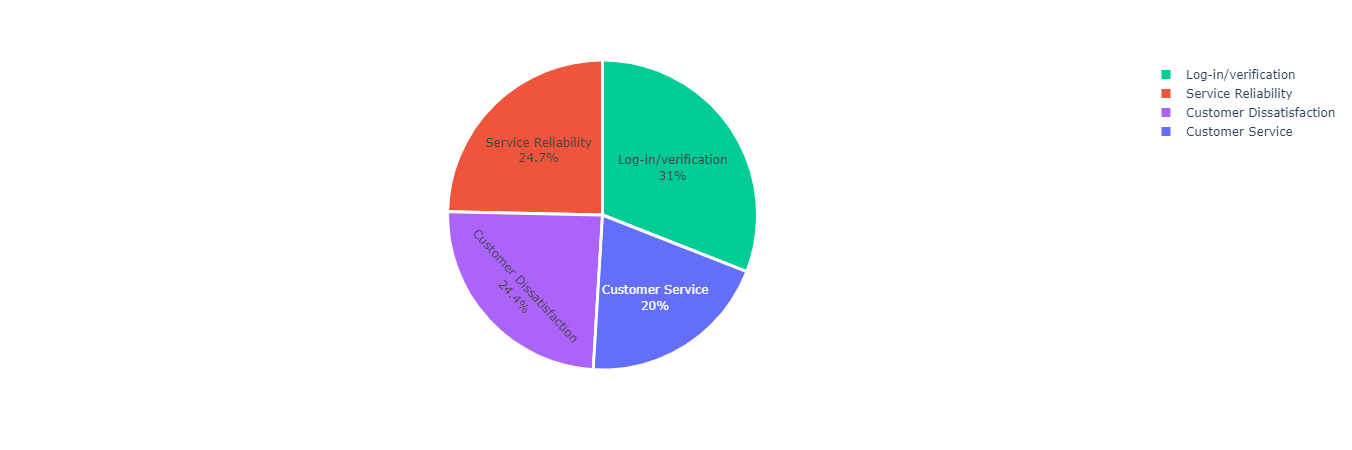
* *Easy & Fast Transactions* refers no errors encountered during transactions
* *App Usability* refers to overall user experience
* *Services Offered* refers to diverse services customers are able to use (the customers are able to use a lot of other services such as loans, investing, saving money)
* *Customer Satisfaction* refers to praises from customers (good descriptions about the app)

User Stories

**Product Development Team**

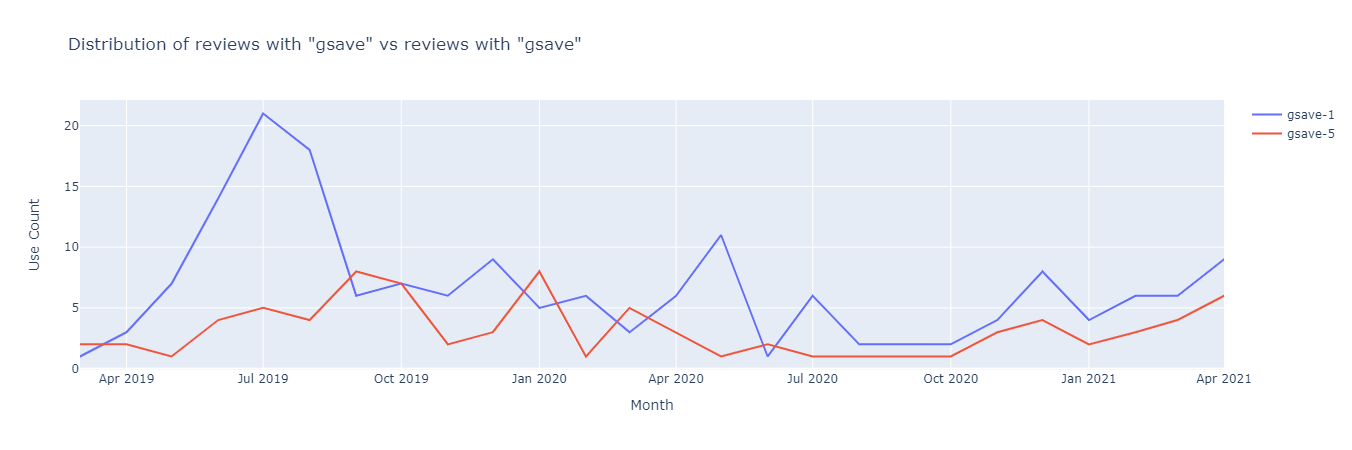
The Product Development Team can use the dashboard to see the prevalent topics that represent the issues faced by users who gave 1-star ratings and search certain words or features to generate insights on how that specific feature or service is performing.

* Log-in/Verification (all periods)



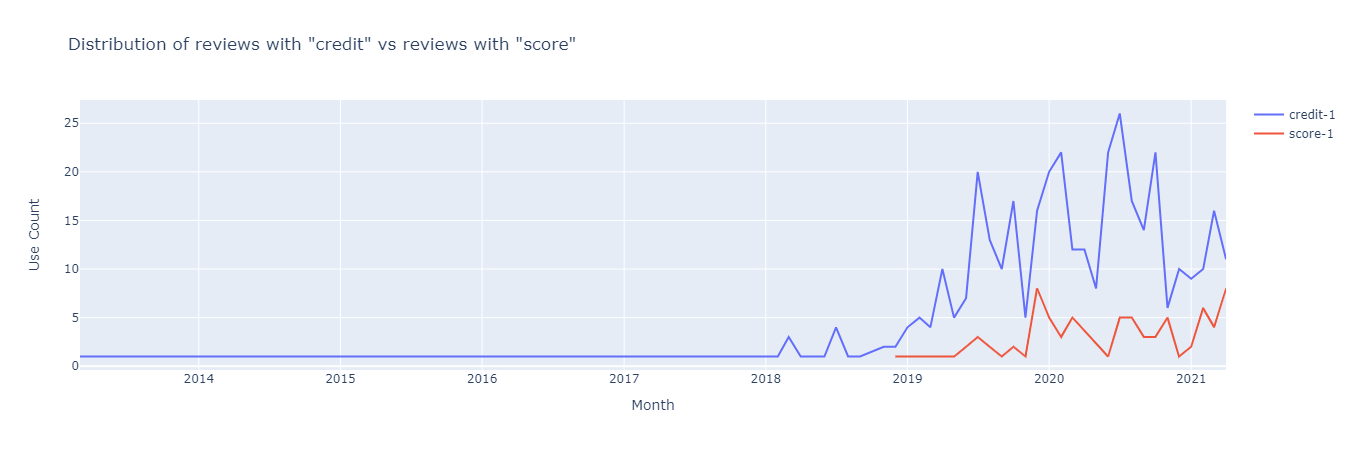
If we observe how topics in the pie chart progressed each year, we will see that a huge chunk of the issues faced by users during the first year was centered on Log-in/Verification (**34.5%**). From 2013 to 2018, GCash has focused its efforts on improving its services when it started to comprise the largest percentage of issues faced by users who rated 1 in their reviews. While GCash was effective over these years in reducing the percentage of users whose reviews are related to Service Reliability, issues regarding Log-in/Verification also started to increase over the years and has overtaken the former in 2019. Since then, Gcash has not been able to reduce this percentage and it currently remains at **31%**.

* GSave (rating 1 vs. rating 5)



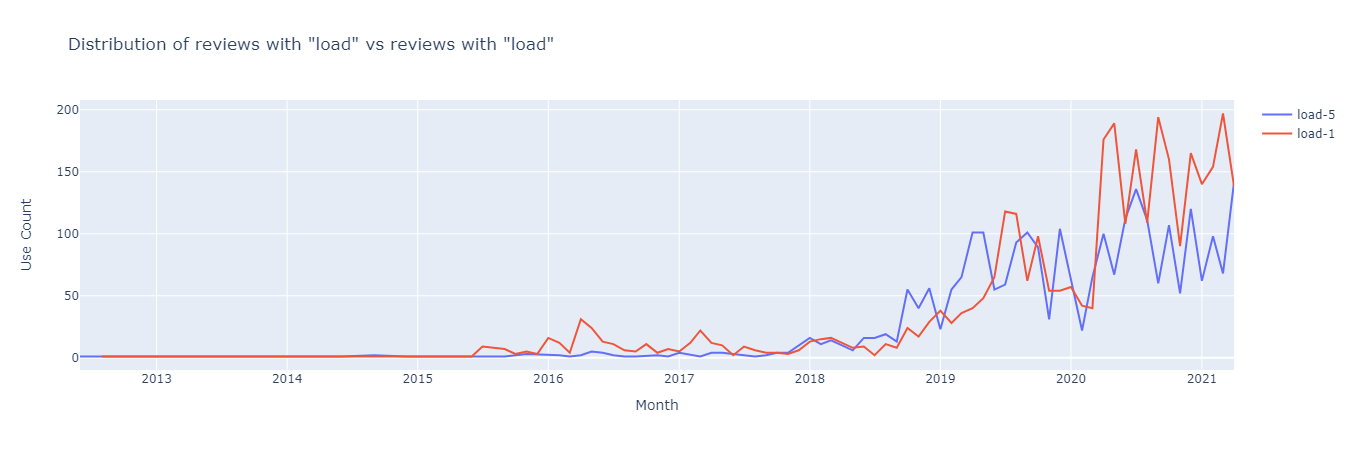
**GSave** was just recently launched in 2019 and this feature enables users to set up their own savings account in the app. Based on the comparison of its usage trend between ratings 1 and 5, more people have been generally dissatisfied with the feature especially during its first few months.

* Credit vs Score (rating 1)



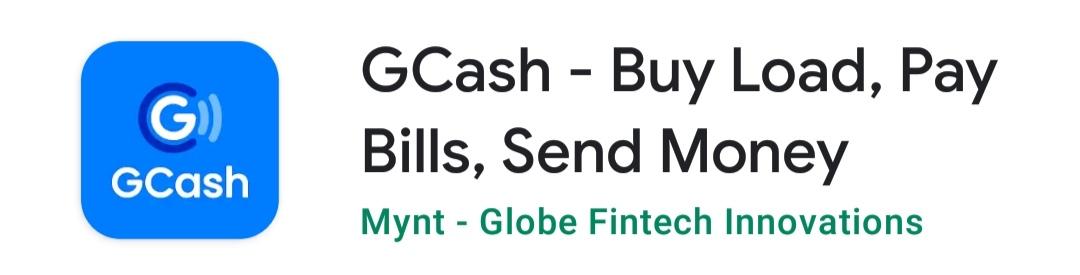
Another feature that the customers use is **GCredit**, which is available to customers who have a high score (those who have transacted regularly). Looking at the graph, both lines have similar behaviors; however, we can see that the word score only appeared in the ratings in late 2018 onward. This could mean that the feature was newly introduced. Also, based on the graph, we can say that there is a continuous use of the feature as depicted by the fluctuating behavior. This year, though, we can see that it has an increasing trend so far, which could imply that customers are having issues with it as this graph is based on rating 1 reviews.

* Load (rating 1 vs. rating 5)



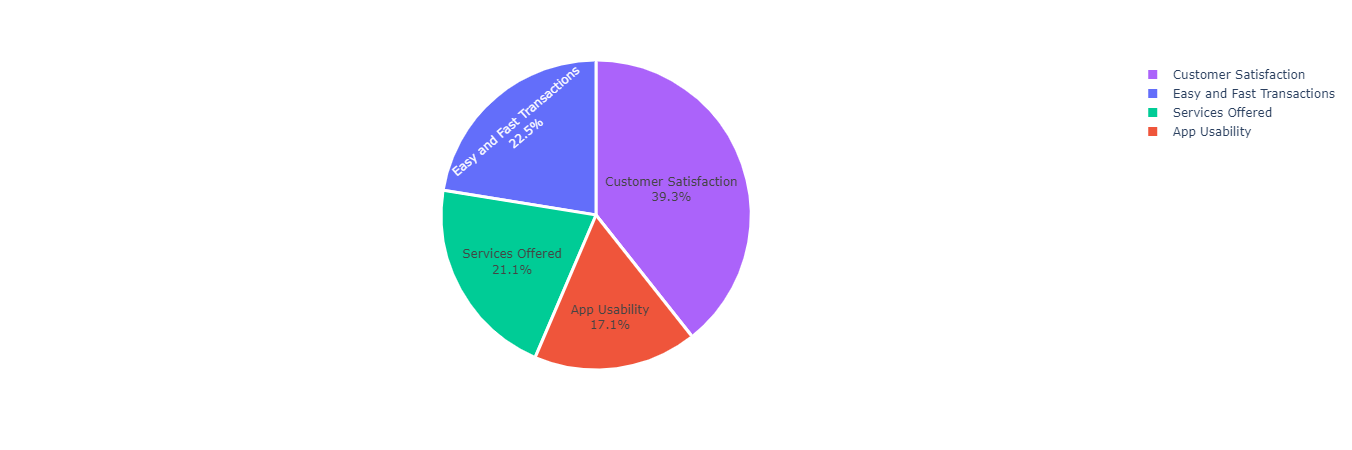
One of GCash’s main services offered is allowing users to buy load through the app. The most recent periods reveal that more people are becoming dissatisfied with this service as shown in the opposing trends between rating 1 and rating 5 reviews.

**Marketing Team**



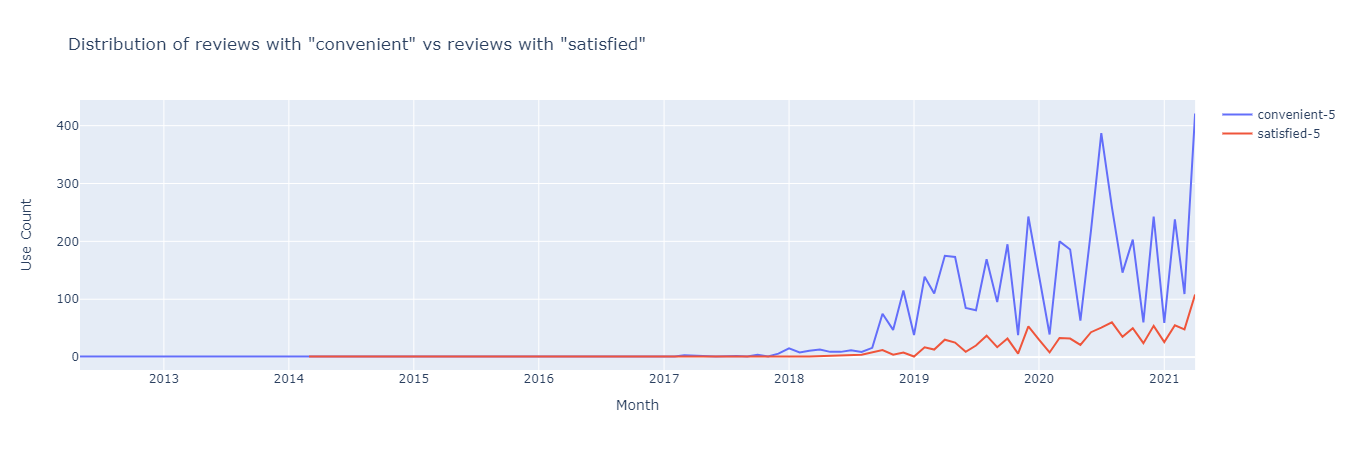
GCash currently markets and associates itself best with services such as buying load, paying bills, and sending money through the app. According to its description in the playstore, users will be able to *“enjoy fast and easy payments”* in this *safe* and *secure* mobile wallet. The marketing team may look at the top words and also search for keywords to check how users generally find the app and its services. They can also see whether their marketing efforts are effective in promoting recently launched features.

* General perception of users (all periods)



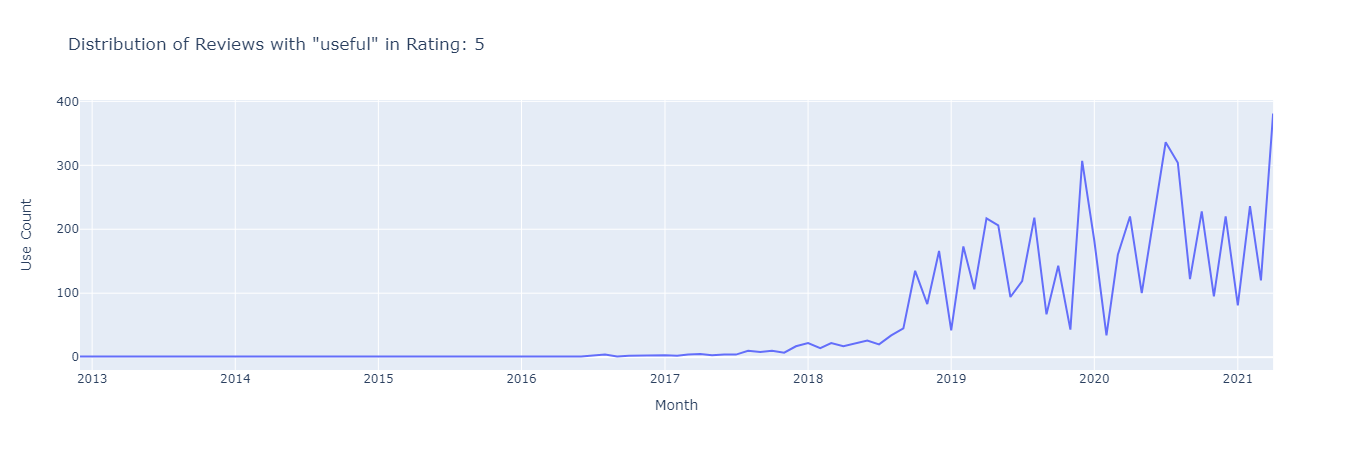
Based on the pie chart, reviews made by users who gave 5-star ratings across all time periods are centered on expressing their satisfaction with the app. The next topics, *Easy and fast transactions* as well as *Services Offered* follow next, having similar percentages at around 20%

* Convenient vs. Satisfied (rating 5)



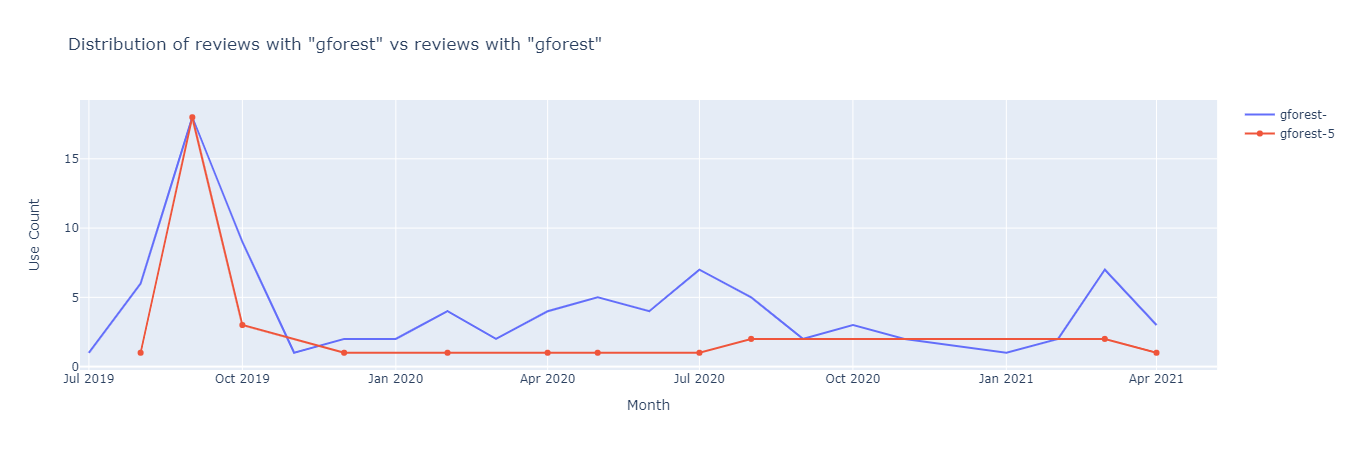
The words *convenient* and *satisfied* are two of the top words in rating 5 reviews. Looking at the graph, the word *convenient* has been increasing over the years with some fluctuations. On the other hand, reviews with the word *satisfied* are increasing as well but at a lesser rate. It does not appear much in the reviews, not exceeding 100 customers who use it throughout the years. The marketing team could use this angle to prove that the app could satisfy the consumers’ needs through the different features that they could use.

* Useful (rating 5 reviews)



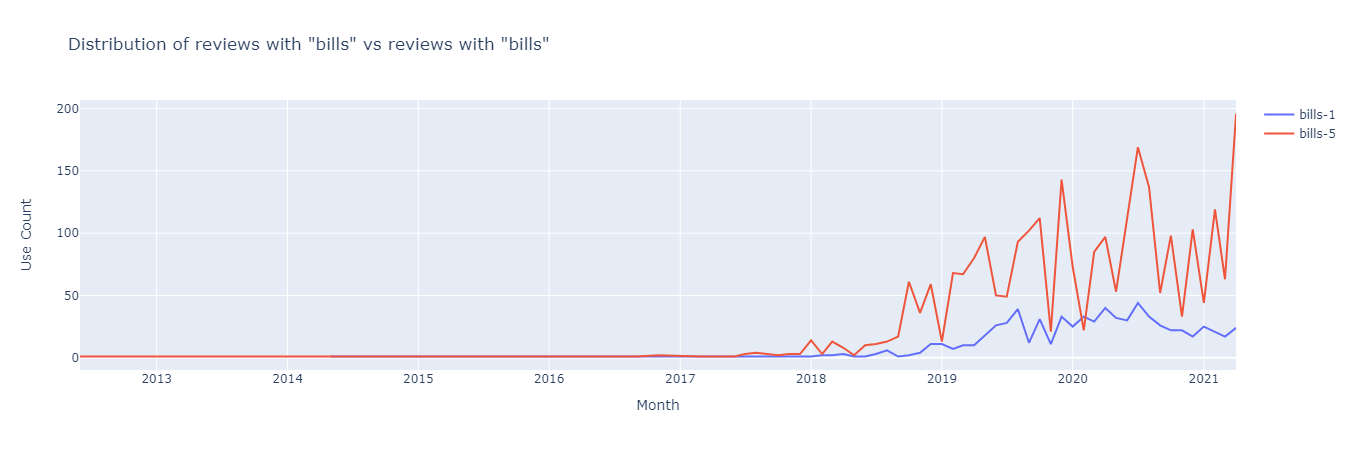
Looking at the graph, we can see that the word useful has been increasingly used by consumers throughout the years, with a sharp increase in late 2018 when online shopping and digital transactions have become more popular. We can infer that people are finding the app and its features useful based on this graph. Although there are fluctuations, probably because of the bad reviews or errors the consumers encounter, the app’s usability has still been commended by the customers.

* GForest (rating 5 vs. alln ratings)



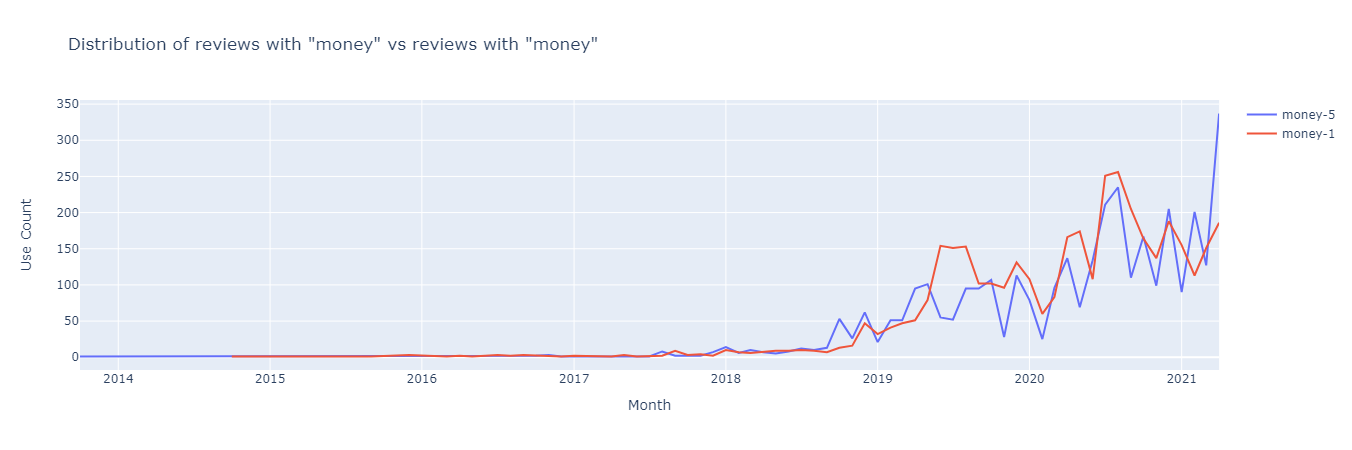
**GForest** is one of the newest features of GCash where users can plant and grow a tree in one of our Philippine forests by collecting “green energy” in the app (walking and making cashless transactions through the app). Based on the graph above, not so many users seem to find this feature relevant in their experience. GForest became only popular in the first 2 months after it was launched, where it reached 18 mentions in September 2019 and all of which are five star ratings. Since then, it has not received much feedback from users who barely mentioned the feature -- not exceeding 10 times until the latest period recorded.

* Bills (rating 1 vs. rating 5)



Based on this graph comparing the usage trend of *bills,* we see a stark difference between the mentions of the word in ratings 1 and 5. More users are generally satisfied with their ability to **pay bills** through Gcash, reaching an all-time high of 196 mentions in the latest period recorded (March-April 2021). On the other hand, the mentions of *bills* in rating 1 seems to be constant and has not exceeded 50 times across all periods (highest recorded is at 44). It seems that Gcash is continuously able to satisfy its users through this service.

* Money (rating 1 vs. rating 5)



For the longest time, more users have been dissatisfied with sending money through the app. In the most recent period, however, GCash was able to significantly increase user satisfaction with this service and achieved an all-time high of 337 mentions of the word among the 5-star reviews.

**Recommendations**

Based on these insights generated, the product development team can:

* Consider focusing its efforts on addressing user concerns regarding log-in and verification.
  + Based on the LDA, some of the words mentioned by users which are associated with this topic are: *mpin, verification, id,* etc. The team may look further into how they can ease the transaction process for users by fixing bugs that cause these system errors for them.
* Delve deeper into what specific issues users of recently launched features like GSave and GCredit might be facing in order to improve their overall experience and encourage current users to maximize these features.
* Similarly, since one of the main services provided by GCash is allowing users to *buy load*, this may be one of their priorities to fix as well in their next updates since based on the recent usage trends of the word, it shows that significantly more people have been dissatisfied with the service in the past year.

As for the marketing team:

* In their marketing campaigns, they may consider re-introducing recently launched features such as GForest as it shows that this has not been maximized by users who are probably not aware of this feature in the app.
* Continue to highlight how GCash is able to provide convenience to users through easy and fast transactions through services that allow them to pay bills and send money through the app.

**Decisions Improved**

From currently only being able to generate qualitative insights from user reviews and report changes in the app’s ratings, the structured data can now guide the marketing team’s decisions, specifically regarding what features they can highlight for their next marketing campaign based on quantitative evidence presented in the dashboard.

The company can also gauge the performance of the app’s services and features better and compare it across time periods. In particular, the product development team can have a better understanding of what issues need to be addressed and prioritized in developing succeeding updates on the app based on usage trends of words and classified topics of user insights.

1. **SUSTAINABILITY**

**Application to Long-term Business Processes**

In the long run, they could keep using this solution by updating the dataset and adding additional data (aka reviews). They could simply edit and run the code to generate information from the dashboard.

The company can continuously monitor their performance through the dashboard the group has created and provided. They could also look at how the consumers respond to the app’s features and updates and garner feedback and meaningful insights from there. By looking at certain time periods, they can also determine if certain business decisions such as marketing campaigns or feature updates had certain effects, may it be positive or negative.

For the product development team, they can use the solution to generate insights about the app’s usability and user experience. They could use the information gathered to check which features are performing well and which need to be improved further. As for the marketing team, they could use the rating 5 reviews to further market how the app could be used by Filipinos all over the world. They could also check how satisfied the customers are with GCash and use this to introduce how well the company is performing.