Exploring How Different Front Line Workers Utilize CommCare

A Collaboration between Dimagi, Inc. and Olin College of Engineering

Dimagi, Inc.
Neal Lesh
Rashmi Dayalu

Olin College of Engineering
Cynthia Chen | yun-hsin.chen@students.olin.edu
Liani Lye | liani.lye@students.olin.edu
Faculty Advisor | Allen Downey | allen.downey@olin.edu

Introduction and Data Overview

Dimagi, founded in 2001, is a privately held social enterprise headquartered in Cambridge, Massachusetts, USA. Dimagi's mission is to deliver open and innovative technology to low-resource settings. We, Olin collaborators Cynthia Chen and Liani Lye, worked to analyze CommCare user behavior. CommCare, Dimagi's flagship product, "a solution for community health and extension workers, also known as Front-Line Workers (FLWs), that provides [them with] case management, data collection, and data management."

Through Google Analytics, Dimagi has amassed data on how each FLW (also referred to as "user") interacts with CommCare. Insight on how FLWs subscribed to various Dimagi service programs engage with CommCare will help Dimagi support organizations to better implement CommCare and increase user engagement. By understanding when and how FLWs use CommCare, we hope to give Dimagi the information they need to increase FLW effectiveness.

The dataset we work with has 16,099 rows. Each row in the dataset represents a month of activity for a distinct user. This means that multiple rows can be associated with a single user. The columns, of which there are 31, contain metadata about the monthly activity: number of visits, location, visit time duration, etc.

Legal and Ethical Concerns

We signed a Research Data Use Agreement with Dimagi to define data confidentiality and usage boundaries. The data are owned by Dimagi and individual users must remain anonymous. To maintain confidentiality, all data and code are stored on a private GitHub repository.

The data include information about each FLW's monthly activity. The data are already anonymized; the only identifiers are a user's identification number, the country in which the FLW is working, and the healthcare field (subsector) the FLW is working in. This information could conceivably be used to identify a specific FLW; however, the data only contain user actions on CommCare, meaning the confidentiality of patients' health information is maintained, as per HIPPA, the federal Health Insurance Portability and Accountability Act.

Question Definition

Users represented in the dataset fall in one of five types of customer plans (or "service types"): Basic, Custom, Plus, Full, and Self. These customer plans vary by payment costs and feature support.

Service	Basic	Plus	Full
Cost	Free	$25,000	$100,000
Included Software Plan (See Plans)	Community Software Plan (FREE)	12 months of the Pro Software Plan	12 months of the Advanced Software Plan
Project Design & Deployment			
Application Creation	–	✔	✔
On-site application iteration and refinement by Field Engineer	–	✔	✔
Creation of Multimedia Content	–	✔	✔
Program Management			
Active Data Management Support (12 months)	–	Available at additional Cost - $15,000	✔
Software Development for New Features and Reports (12 months)	–	Available at additional Cost - $25,000 per Block	1 Block included
Fully managed program implementation by Dimagi (12 months)	–	–	✔
In-Kind Dimagi Management Team Support for strategy, evaluation, and scale-up (12 months)	–	–	✔
Support & Maintenance			
Access to Community Support	✔	✔	✔
Direct e-mail support from Dimagi (12 months)	–	✔	✔
Direct phone support from Dimagi (12 months)	–	Available through Field Engineer Support	✔
Additional Field Engineer Support	–	Available at additional Cost - $15,000 Per Block	2 Blocks included

FIGURE 1: Dimagi service breakdown from the CommCareHQ website (February 2015).

Basic users use CommCare free of charge. Plus and Full users pay $25,000 and $100,000 annually, respectively (FIGURE 1). Custom users are charged based on the a suite of features they choose to be supported. Self users start their own CommCare projects without any dedicated support from Dimagi. The dataset also contain 378 users with missing service information. These users are omitted from our analysis.

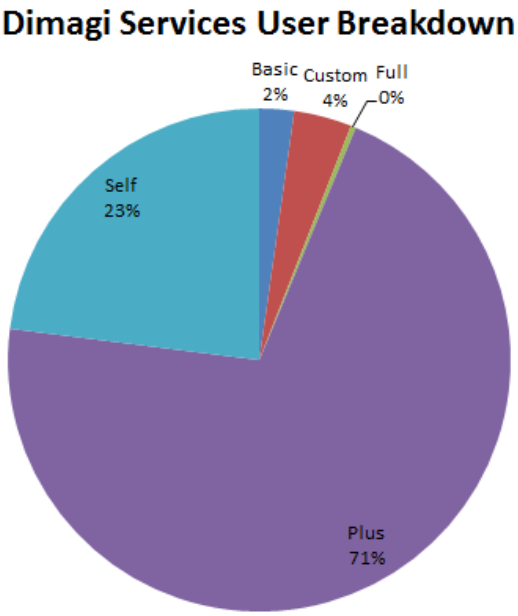


FIGURE 2: Dimagi Services User Breakdown.

In the 16,099 row dataset, there are 2,804 unique users. FIGURE 2 shows the user breakdown among the different Dimagi service types. There is a large percentage of Plus users, especially in comparison to the Basic and Custom service users.

We explore the characteristics of different the customer types.

Data Exploration: Service Plans User Characterization

User Characterization Based On Country

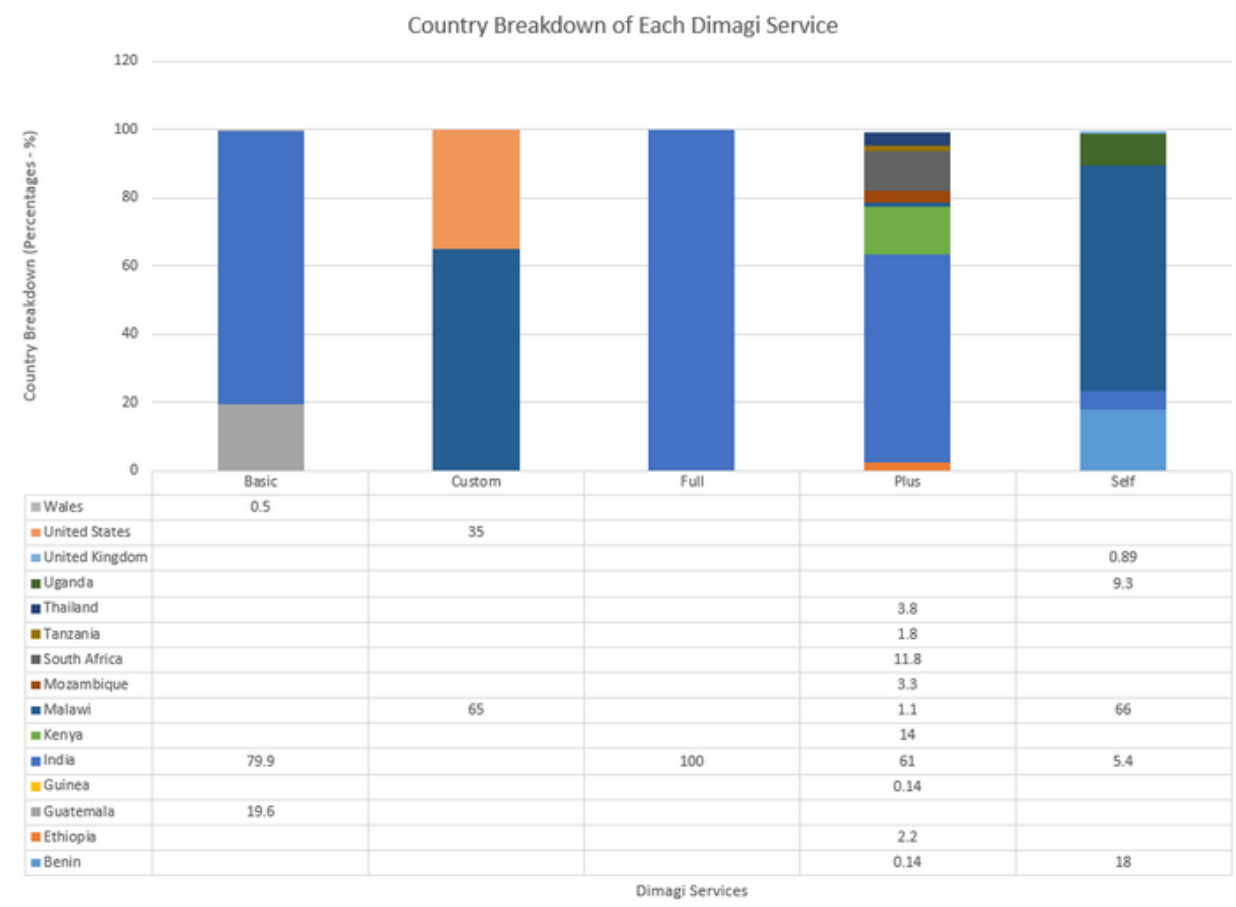


FIGURE 3: Country Breakdown of Each Dimagi Service.

With the specific breakdown of monthly records per county show that India and Malawi are the two countries with the greatest number of data points. After looking through all of the possibilities for user behavior comparison, we find that the distribution of the number of active days per month is the closest for both datasets.

User Characterization Based On Subsector

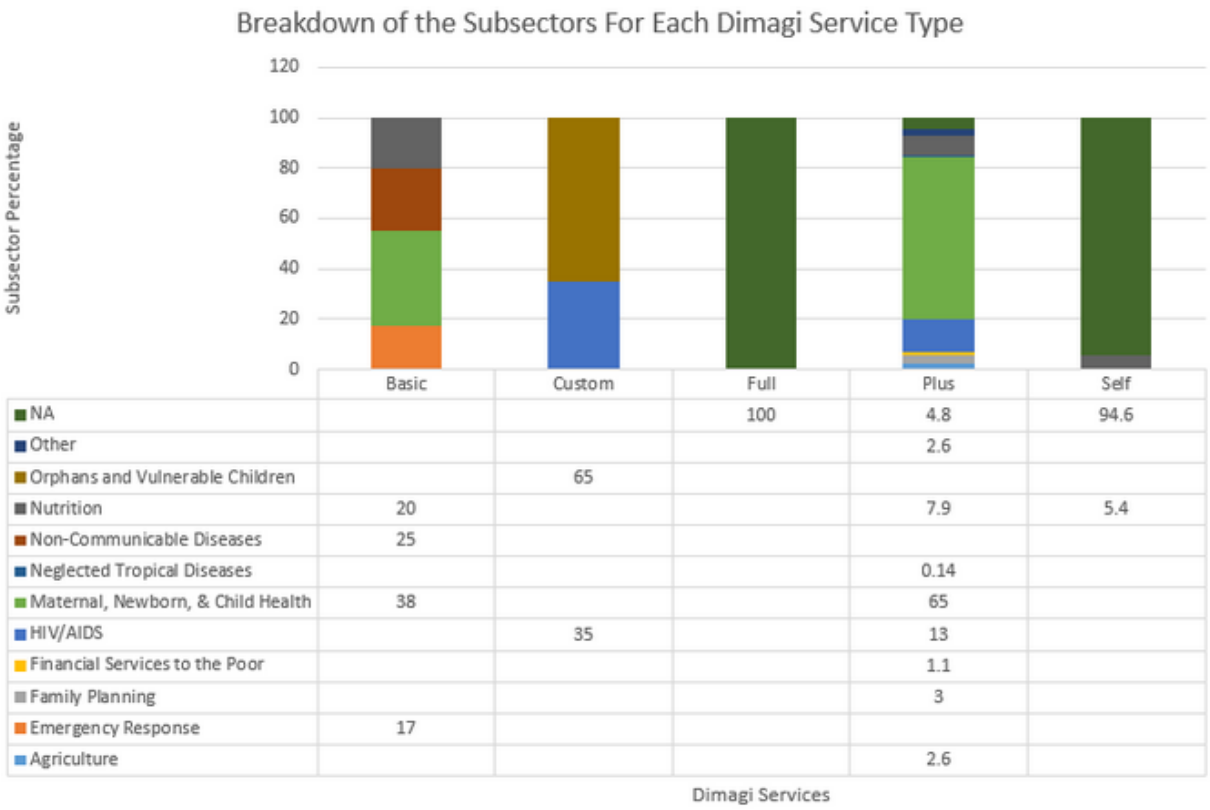


FIGURE 4: Subsector Breakdown of Each Dimagi Service.

Points of interest: What are the characteristics of users from different subsectors? Do users from particular subsectors have key behaviors that can be capitalized upon?

User Characterization Based on Device Type

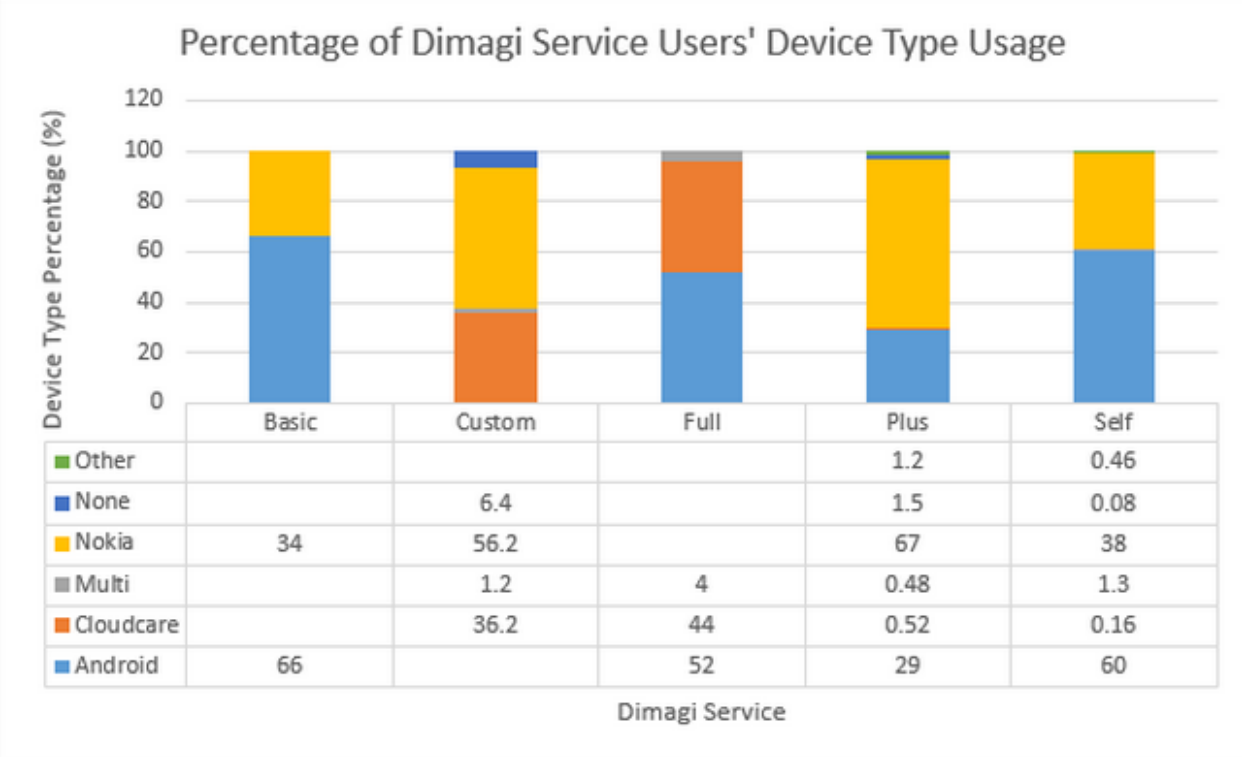


FIGURE 5: Device Breakdown of each Service Type.

Points of interest: What are the characteristics of users who access CommCare through different devices? Is CommCare effectively supported across all platforms? Are customers of a particular service type more or less likely to access CommCare through a particular platform?

Service Plan Users Behavior Activity

For each service type, we compare the amount of time users are actively using CommCare to the amount of time the user is paying for the CommCare support. Ideally, the percentage of months a user is active should be as high as possible. Some users who only use CommCare for a fraction of the months that they are registered for.

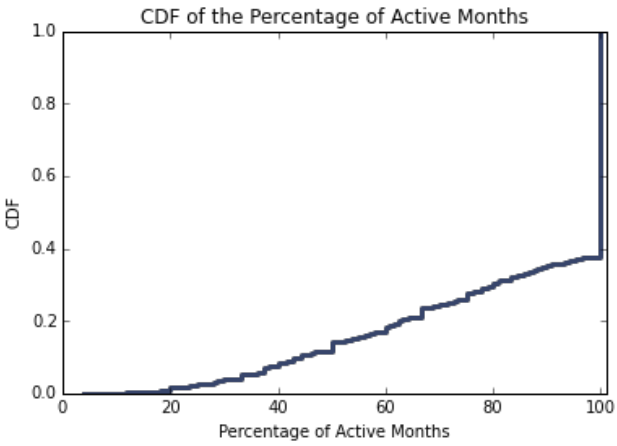
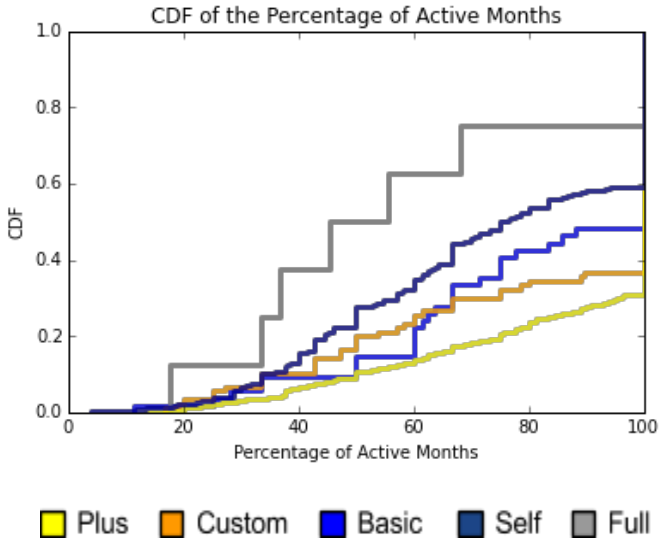


FIGURE 6: CDF of Percentage of Active Months

FIGURE 3 is a CDF of the percentage that users are actively using CommCare. 62.5% of all users used CommCare for 100% of the time they were registered and/or paid for. There is an uniform distribution of users among the percentage of active months that is below 100%.



.FIGURE 7: CDF of active month percentage by customer plan.

Service Type	%age of Users that Fully Utilized Plans
Plus	69
Custom	63
Basic	52
Self	41
Full	25

FIGURE 8: Percentage of users in each plan who have fully utilized their plans, using the service 100% of the time that they registered for.

As per FIGURE 8, all of the service plans have an even distribution of frontline workers using the service for the percentages of time below 100% of active months. However, the percentage of users in all of the service plans skyrocket at different CDF values when it gets to 100% of active months. These different CDF values indicate that there are higher percentage of users in particular service plans who utilize the plans for a higher amount of time. FIGURE 8 shows that Plus users have the largest proportion of users using CommCare 100% of the time that they are registered for it, followed by Custom users, Basic users, Self users, and Full users.

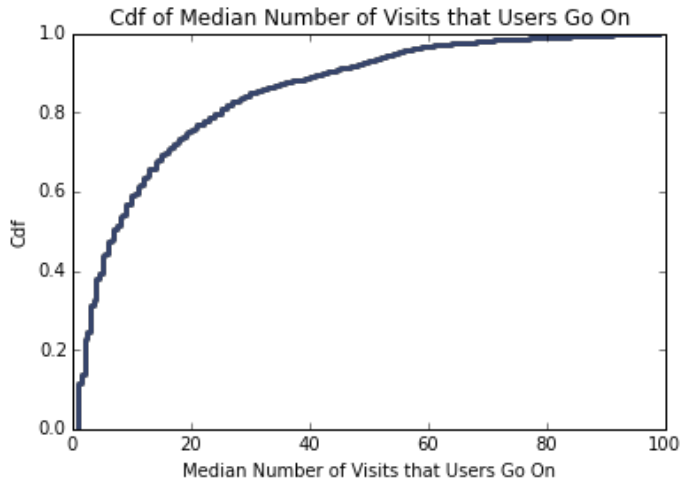


FIGURE 9: CDF of maximum number of visits for all users.

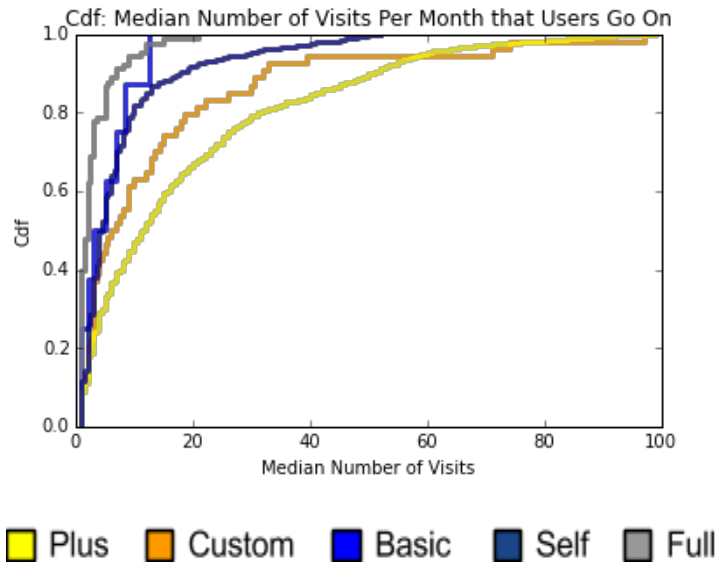


FIGURE 10: CDF of a user's maximum number of visits by customer plan.

FIGURE 9 plots the the number of users that go on a certain number of visits. Half of the users go on seven visits or less. FIGURE 10 is a breakdown of the median number of visits per month that users from each customer plan go on. The order of customer types who go on the most visits: Basic, Plus, Self, Custom, and Full. Self and Full users have similar usage behaviors; however, because there are very few Full users and therefore fewer datapoints, the line is not as smooth. On this note, because the number of Plus users eclipses the number of users from any other service type, the CDF line for Plus users is smoother. For Basic users, the slope decreases when they reach 77%.

Users Behavior Over Time

The dataset spans from August 2010 to December 2014. FIGURE 11 shows the number of each user type from December 2011 to December 2014. The two users from 2010 are omitted because we assume, due to the low count and stretch of time, that they are beta testing accounts.

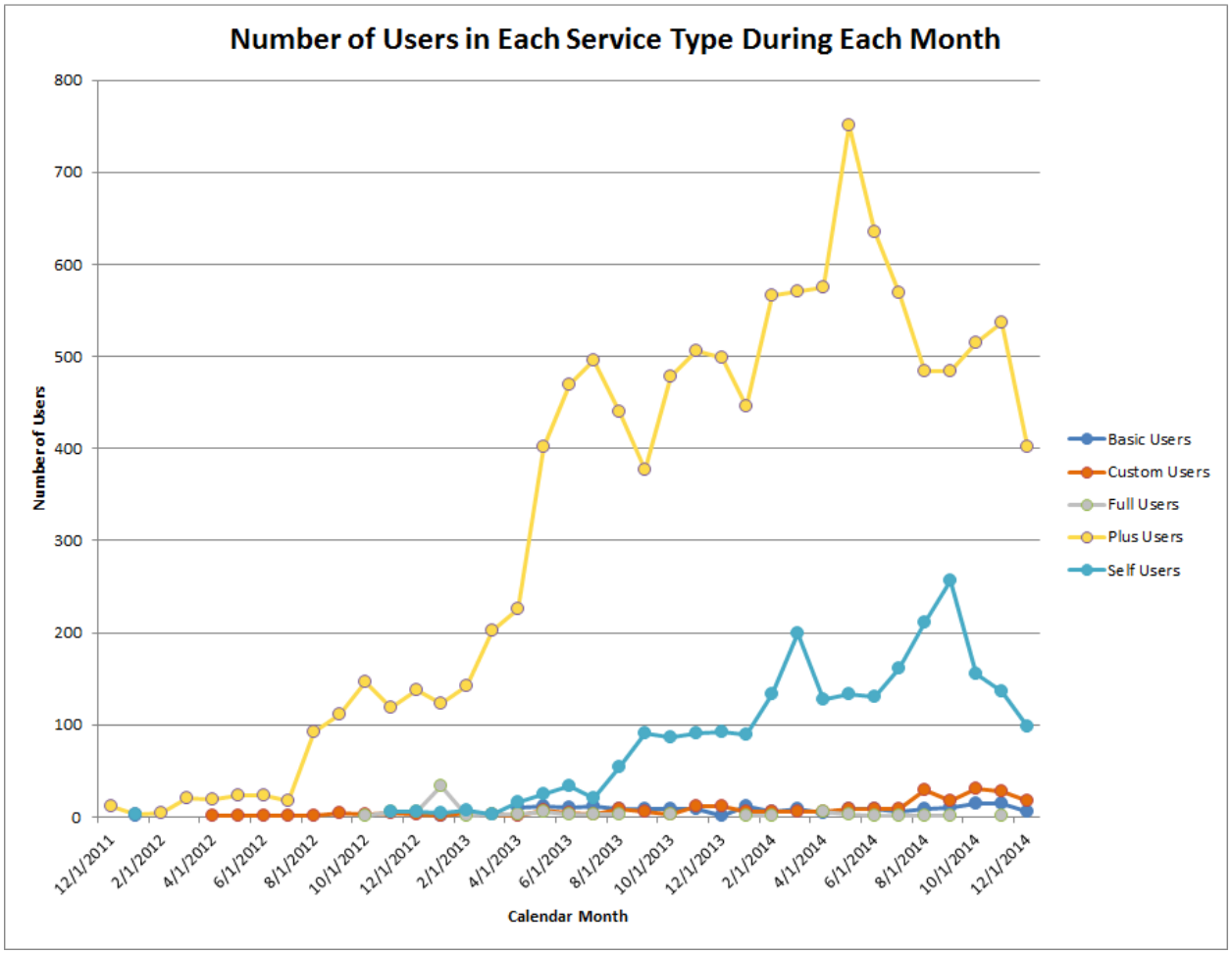


FIGURE 11: Number of users per service type over time.

FIGURE 11 shows that the numbers of Plus and Self users have increased over time, but have recently declined in usership. The Plus users, in particular had its greatest peak around May 2014, and have been declining in number ever since. This figure highlights the fact that the number of Plus users eclipses all others.

In order to get a closer look at the size of users over time in the Basic, Custom, Full, Plus and Self user groups, FIGURE 12 specifically shows the time-series of the number of users in the Basic, Custom and Full user type categories, which were drastically overpowered by Plus and Self users in FIGURE 12.

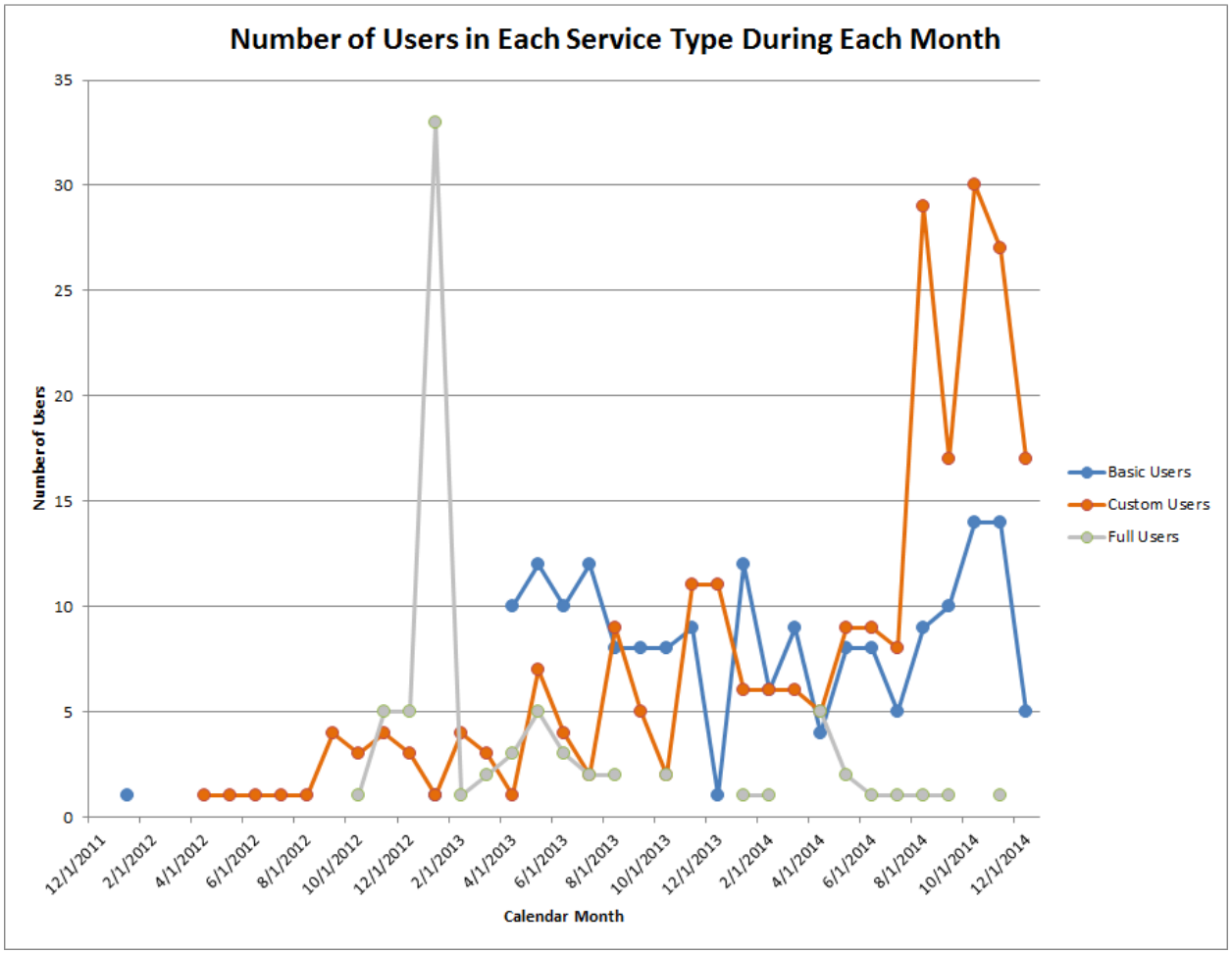


Figure 12: Zoomed-In Version of Figure 10 to see distribution of Basic, Custom, and Full Customers.

FIGURE 12 shows that there has been a more recent increase in the number of Custom users, while the number of Full users dramatically dropped after January of 2013 and has been reduced to no more than 5 users in the service type. In comparison to Custom and Full users, Basic user have been pretty consistently in the 5-15 user size range, except for December 2013, which shows a brief plummet in user size. Understanding why the number of Full users have been so low could be an interesting next step.

The influx in numbers of users in each category over time is very prevalent. As a result, we also want to see how the number of interactions that users go on change over time and different service plans. FIGURE 13 demonstrates how the number of interactions that the users go on differ over time and in the different service plans., the time of the year also affect the way that the FLW uses the service, specifically the number of interactions that they go on during those times? And do the different service types have similarities or difference in the number of interactions that the user has?

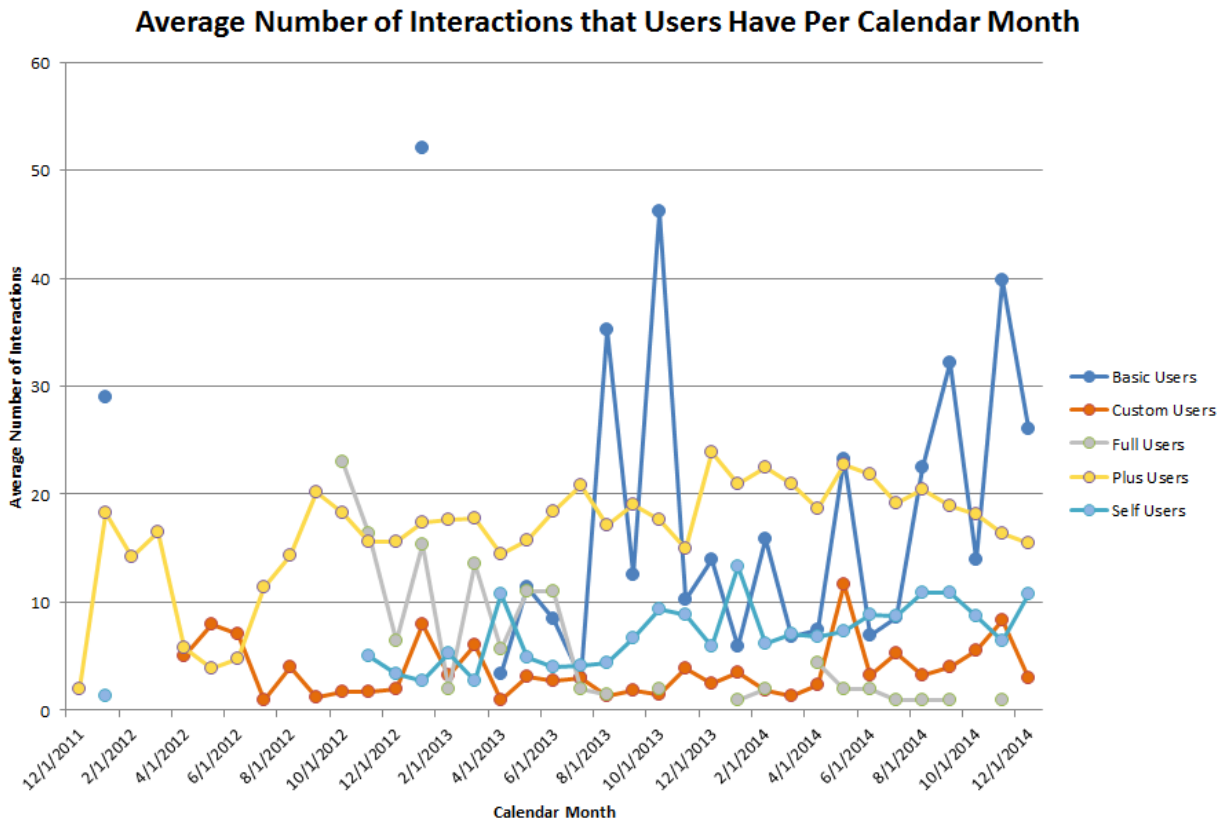


FIGURE 13: Average Number of Interactions per service type over time.

FIGURE 13 shows a time-series of how the average number of interactions (over all of the users who submit data that month) per month changes over time. The time series indicates Basic users tend to have more cases than any of the other user types. Plus users have the most consistent number of cases throughout all of the months, compared to all of the other user groups. Further analysis of the type of users that join the user group could reveal a lot more about the users.

As a result, we also examine if the number of visits that a user logged is the same as the number of interactions logged. An FLW's visits indicates the number of places that the user goes to, while the number of interactions is the number of patients they log. The reason the two characteristics are different is that during one visit, an FLW can have multiple interactions and work on more than one interaction in a single visit. Comparing the number of interactions to visits reveals whether different customer types have more interactions than visits or vice versa.

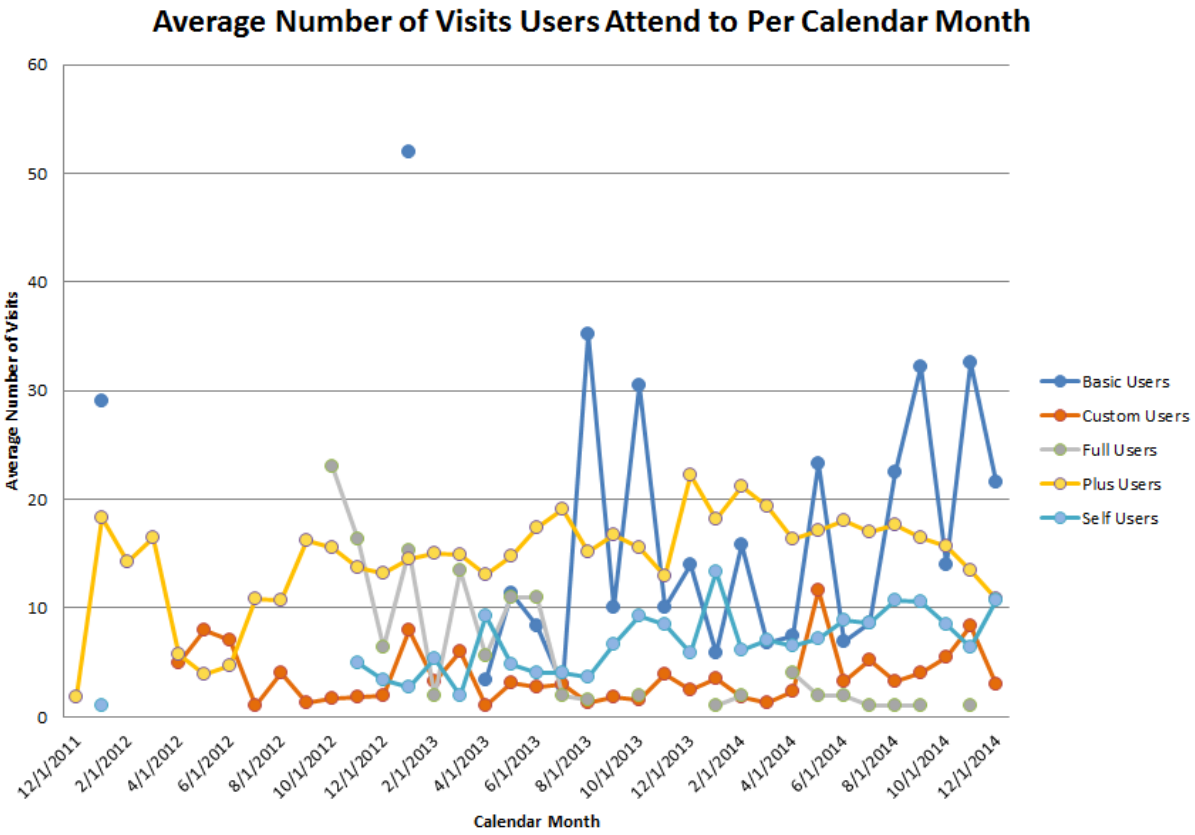


FIGURE 14: Average Number of visits per service type over time.

FIGURE 14 illustrates the average number (over all of the users during that month) of additional interactions that users in each service type go on per number of visit.

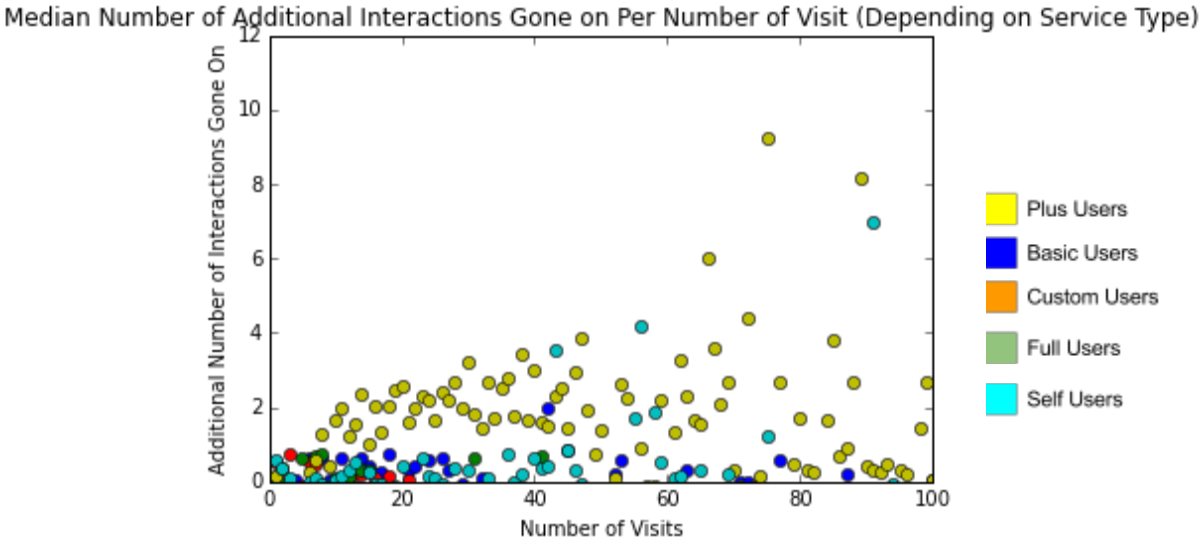


FIGURE 15: Median Number of Additional Interactions Gone on vs. Number of Visits.

As FIGURE 15 shows, Custom and Full users don't go on as many, if not zero, additional interactions per the number of visits they go on. In comparison, Self and, especially, Plus users tend to go on more additional interactions, which also increases as the number of visits that they go on is a larger value. This could demonstrate that Plus and Self Users are generally health workers who work at hospitals or health booths/centers, while the other user types go on individual cases and calls.

Conclusion and Future Directions

There are five types of CommCare users – Basic, Custom, Plus, Full, and Self – defined by the level of support they are paying Dimagi for. We characterize each customer category by overall activity, the number of visits per month, and the quantity users over time. We determine that the Plus category has the highest number of visits, additional interactions, and users.

Future avenues of exploration: FIGURE 16 is a time-series analysis of the seven users that have the most data (the most number of active months). The peaks at November and February 2013 are worth investigating. Perhaps those peaks are correlated to other characteristics.