

Report of Home Energy Feedback Usage in IDEAL Project

Introduction

Martin Pullinger introduced the IDEAL project of home energy feedback usage in data fair. Martin and his team were running a large scale experimental study to test new approaches to gathering home energy use data and providing householders with feedback about their patterns of energy use and energy saving advice. Feedback is delivered to participating householders via an app. Over the course of participation they have developed and released various new features within the app for providing new visualisations of home energy use data. The data provided by data holder are two csv files and show the user information and app usage separately. The aim of this project is helping the data holder to improve the app by analysing the data. How app usage varies over time and how this is influenced by the characteristics of the participants was explored and the results demonstrated in final presentation was aimed at the audience, Martin and his team.

Process of the Project

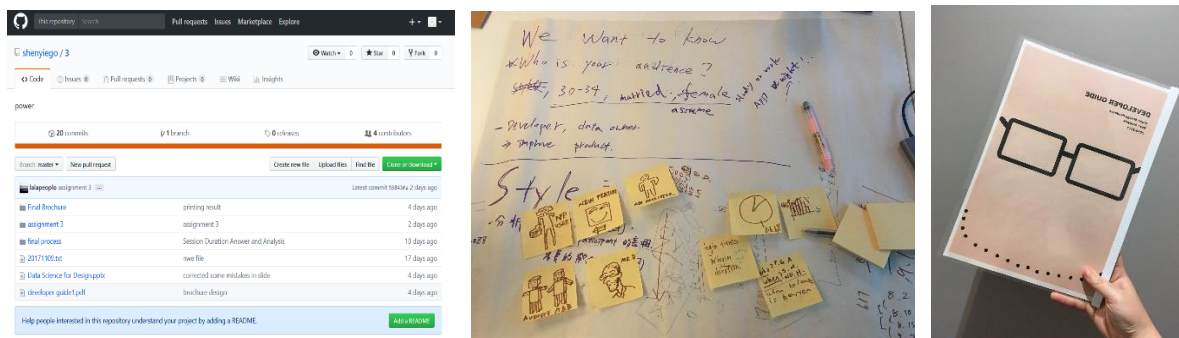


Figure 1 Exchange file on GitHub, Brainstorms and the Final Presentation Brochure (Left to Right)

At the beginning, each member in the group assigned a task and analysed data from different aspect by Python individually. Then three members exchanged codes by GitHub and had a meeting to discuss the analysis results. The group faced many difficulties of writing code because all of three members in the group are from MA background. The members in the group seek for help from the reaching assistant Kai and finally solved all issues. The group used brainstorm to generated idea for final presentation and discussed it with tutor. The group finally decided to design a guide brochure which summarise key findings from data analysis. Considering the audience of the presentation are a group of professional app developers, the content and layout should tend to commercial report and a guide brochure was the most suitable approach to give the most valuable information to them. Each member in the

group assigned a task to contribute to the final presentation. Jerry Lin was in charge of sorting out the visualisation and analysis of the relationships between new features released time and participants' log in time. Hsin Chieh Tsai was responsible for sorting out the visualisation and analysis of participants' log in time and log in duration. Yi Shen was responsible for sorting out the visualisation and analysis of the gender and age of participants. And took charge of the art design and the layout of the final presentation brochure. Jerry Lin and Hsin Chieh Tsai designed the presentation slides. The group met up and practiced the presentation for several times and finally presented to the targeted audience, data owner, and the tutors and rest of students in the course.

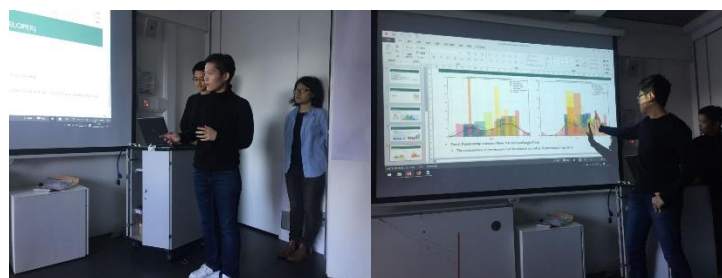


Figure 2 Final Presentation: Hsin Chieh Tsai, Yi Shen, Jerry Lin

Analysing Data

The persondetails.csv file shows the released time of new features including WEEK_VIEW, MINUTE_VIEW, ENHANCED_VIEW, COMPARE_CHART, DUAL_CHART, and HEATING_INTERVENTION for each person. This csv file also shows the age and gender of participants and if they are primary participants. The primary participant is the person in a household who signed the household up to participate in the IDEAL project they are marked with a '1' in the primary participant column.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
personid	participation_start	WEEK_VIEW	MINUTE_VIEW	ENHANCED_VIEW	COMPARE_CHART	DUAL_CHART	HEATING_INTERVENTION	primaryparticipant	age	gender	have_email_address				
2	0	765	2017-07-14 11:27	2017-07-14 11:27	2017-07-14 11:27	2017-07-14 11:27	2017-07-14 11:27		0:55-59	Male	FALSE				
3	1	990	2017-09-28	2017-10-15 23:30	2017-10-15 23:30	2017-10-15 23:30	2017-10-15 23:30		0:15-19	Male	FALSE				
4	2	673	2017-01-30	2017-03-15 15:30	2017-03-15 15:30	2017-04-07 14:44	2017-05-29 11:18	2017-07-03 11:00	0:45-49	Male	FALSE				
5	3	906	2017-08-31	2017-10-15 23:30	2017-10-15 23:30	2017-10-15 23:30	2017-10-15 23:30		0:35-39	Female	FALSE				
6	4	729	2017-04-18	2017-05-15 08:49					0:45-49	Male	FALSE				
7	5	796	2017-07-04	2017-09-11 14:32	2017-09-11 14:32	2017-09-11 14:32	2017-09-11 14:32		0:35-39	Female	FALSE				
8	6	711	2017-03-07	2017-04-14 11:38					0:45-49	Female	FALSE				
9	7	997	2017-09-29	2017-10-15 23:30					0:35-39	Female	FALSE				
10	8	741	2017-06-13	2017-07-14 11:38					1:30-34	Male	TRUE				
11	9	841	2017-07-25	2017-08-14 17:09	2017-08-14 17:09	2017-08-14 17:09	2017-08-14 17:09		0:25-29	Male	FALSE				
12	10	749	2017-06-14	2017-07-14 11:26					1:75-79	Female	TRUE				
13	11	617	2016-08-16	2017-03-15 15:30	2017-04-07 14:44	2017-05-29 11:18	2017-07-03 11:00		1:25-29	Female	TRUE				
14	12	900	2017-08-29	2017-09-15 09:04					0:25-29	Female	FALSE				
15	13	667	2017-01-24	2017-03-15 15:30					1:20-24	Female	TRUE				
16	14	633	2016-11-09	2017-03-13 16:39	2017-03-08 14:09	2017-08-03 14:45	2017-06-13 09:47		1:45-49	Male	TRUE				
17	15	863	2017-08-03	2017-10-06 13:04	2017-10-06 13:04	2017-10-06 13:04	2017-10-06 13:04	2017-03-08 14:10	1:65-69	Female	TRUE				
18	16	752	2017-06-15	2017-07-14 11:27	2017-07-14 11:27	2017-07-14 11:27	2017-07-14 11:27		1:50-54	Male	TRUE				
19	17	715	2017-03-09	2017-05-15 08:48	2017-05-15 08:48	2017-05-15 08:48	2017-05-29 11:18	2017-07-03 11:00	1:50-54	Female	TRUE				
20	18	834	2017-07-19	2017-08-14 17:09					1:25-29	Female	TRUE				
21	19	928	2017-09-20	2017-10-15 23:30	2017-10-15 23:30	2017-10-15 23:30	2017-10-15 23:30		1:50-54	Female	FALSE				
22	20	847	2017-07-26	2017-09-15 09:04	2017-09-15 09:04	2017-09-15 09:04	2017-09-15 09:04		1:70-74	Female	TRUE				
23	21	782	2017-06-28	2017-08-14 17:09	2017-08-14 17:09	2017-08-14 17:09	2017-08-14 17:09		1:65-69	Female	TRUE				
24	22	770	2017-06-21	2017-07-14 11:26					0:25-29	Male	FALSE				
25	23	908	2017-09-01	2017-10-15 23:30	2017-10-15 23:30	2017-10-15 23:30	2017-10-15 23:30		0:50-54	Male	FALSE				
26	24	640	2016-11-24	2017-05-15 08:48	2017-05-15 08:48	2017-05-15 08:48	2017-05-29 11:18	2017-07-03 11:00	1:50-54	Female	TRUE				
27	25	650	2017-01-10	2017-03-15 15:30	2017-03-15 15:30	2017-04-07 14:44	2017-05-29 11:18	2017-07-03 11:00	0:30-34	Male	TRUE				
28	26	734	2017-04-13	2017-05-15 08:49					1:45-49	Male	TRUE				
29	27	895	2017-08-25	2017-09-15 09:04	2017-09-15 09:04	2017-09-15 09:04	2017-09-15 09:04		1:40-44	Female	TRUE				

Figure 3 First CSV file of Users' Information

By analysing the data in the file. The group found that all of participants received the

WEEK_VIEW in six new features. The percentage of participants received MINUTE_VIEW, ENHANCED_VIEW, COMPARE_CHART and DUAL_CHART is same, which is 56%. Only 4% participants received the feature HEATING_INTERVENTION. Martin introduced that some people only receive WEEK_VIEW because they were in the control group and developers often released new features around the 14th of each month to participants to these participants who are not in the control group.

The data also showed that there were 222 participants (the invalid data including people who under age of 15 and those who did not have a gender were removed by the group from the data frame as the data holder introduced only occupants over age of 15 were allowed to have an account) used the app and female participants which took 52% are more than male which took 48%. Moreover, participants aged 30-34 occupies the largest number of all participants.

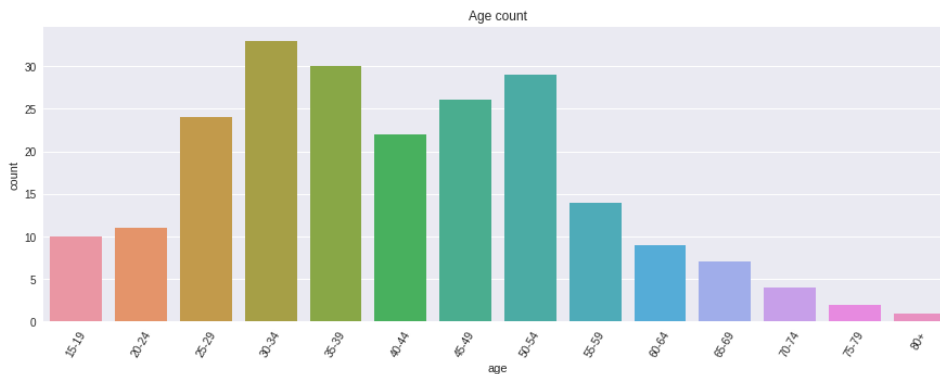


Figure 4 Count of Different Age Group of All Participants

The second csv file, sessionlogs.csv, shows the time of users log in to system and the session duration (the initial login until the last interaction with the app).The data owner introduced that the 0 seconds long represents a user logged in and do not touch anything and if participants do not interact with app more than five minutes, the app would end the record automatically.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y
1	time	personid	session_duration																					
2	0	10082016	11:57	617	0	days	00:04:48																	
3	1	10082016	22:14	617	0	days	00:01:42																	
4	2	10082016	23:43	617	0	days	00:05:16																	
5	3	11482016	19:39	617	0	days	00:02:06																	
6	4	14482016	17:12	617	0	days	00:02:39																	
7	5	15082016	22:20	617	0	days	00:02:45																	
8	6	17082016	11:57	617	0	days	00:00:00																	
9	7	17082016	12:04	617	0	days	00:00:00																	
10	8	17082016	12:34	618	0	days	00:00:00																	
11	9	17082016	22:26	617	0	days	00:00:31																	
12	10	20082016	13:09	617	0	days	00:00:00																	
13	11	20082016	13:09	618	0	days	00:00:42																	
14	12	21082016	09:55	617	0	days	00:02:27																	
15	13	23082016	18:20	617	0	days	00:00:00																	
16	14	23082016	18:20	618	0	days	00:00:56																	
17	15	24082016	11:49	619	0	days	00:03:07																	
18	16	24082016	18:28	619	0	days	00:02:47																	
19	17	24082016	19:57	619	0	days	00:01:53																	
20	18	24082016	20:20	619	0	days	00:02:28																	
21	19	25082016	19:07	619	0	days	00:00:11																	
22	20	26082016	20:44	618	0	days	00:00:14																	
23	21	27082016	01:45	618	0	days	00:00:47																	
24	22	27082016	11:38	618	0	days	00:00:15																	
25	23	27082016	23:44	619	0	days	00:00:09																	
26	24	27082016	22:41	618	0	days	00:01:51																	
27	25	28082016	22:21	619	0	days	00:00:08																	
28	26	29082016	22:09	619	0	days	00:00:09																	
29	27	30082016	21:28	619	0	days	00:01:58																	

Figure 5 Second CSV File of App Usage

The result of analysis of the file indicated the most participant log in duration was below 500 second (8 minute 20 second) and there were only very few people logged in over 2000 second (33 minute 20 second). Because the information in session duration column was written in a way which was unable to calculate or plot in python so the group convert all the times into seconds.

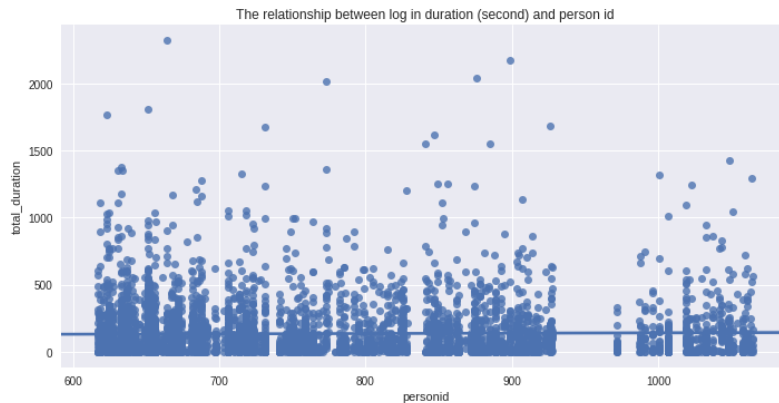


Figure 6 Login Duration of Each ID

The login time can be counted by person ID in sessionlogs.csv, the user largest logged in times 265, the average log in time is 24, and the smallest log in time is 1. For the information of the peak of login, first, the group sorted the data by year because this app started to record data from August 2016 to October 2017 which crossed two different year. According to the statistics, participants tend to login to system in the after summer. The bar chart also showed that the frequency of people's logins did not drop off overtime as the their logins reached the maximum in October 2017

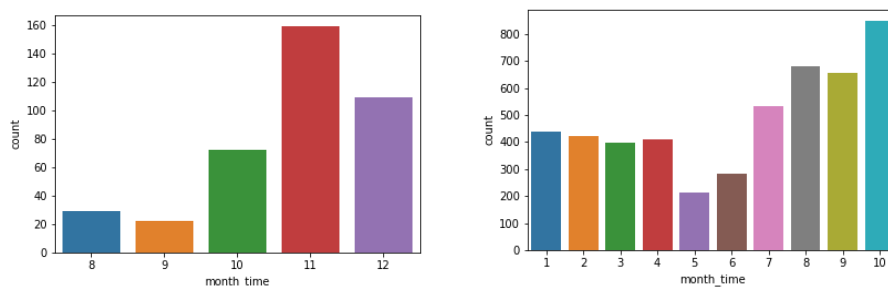


Figure 7 Count of Logins in each Month in 2016(Left) and 2017(Right)

After analysing two file saperately, the group brought the result of two file together to explore the relationship between the users' log in time and new features released time. The figure 7 below demonstrated that the peak of participants' login months (red graph) was similar to that of release month of most new features.

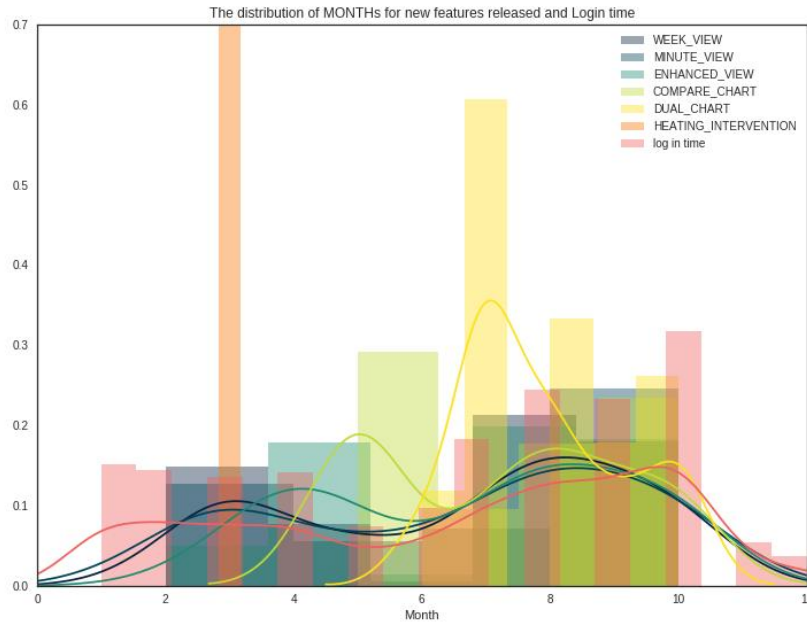


Figure 8 Relationship between New Features Released Mouth and Participants Logins' Month

Figure 8 showed that every features except HEATING_INTERVENTION has two peaks of releasing time on Monday and Friday. Comparing to the participants' login weekday, it showed that the log in weekday are influenced by the new feature released weekday because the distributions of each weekday for log in time are quite similar.

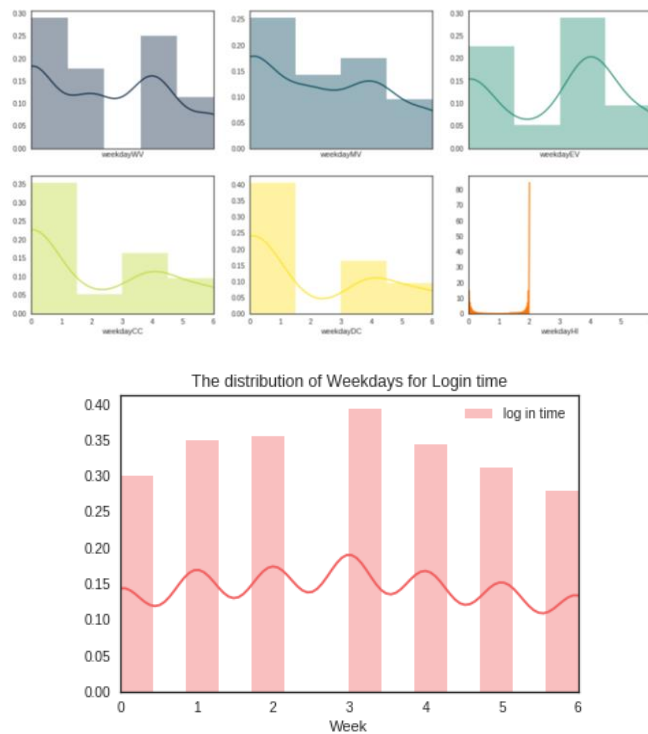


Figure 9 Relationship between New Features Released Weekday and Participants Logins' Weekday

The Figure 10 demonstrated every features except HEATING INTERVENTION has three release peaks of released hours in a day. The first peak was around 9:00 to 11:00, the second was between 15:00 to 17:00, the third on was between 21:00 - 23:00. There were also three peaks of log in hours in a day but two of them were different that of most new function released hours.

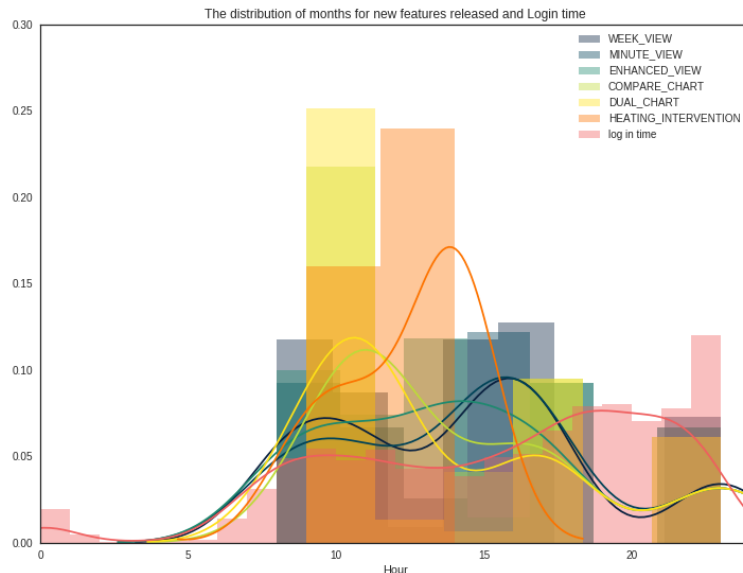


Figure 10 Relationship between New Features Released Hour and Participants Logins' Hour

How Could the Findings Help IDEAL Developers to Improve the Project

For app developers, the definition of target users is essential, the findings can help them to develop market strategy and improve the project arrangement. In persondetails.csv, the patterns of gender and age group of participants could reflect target their users' identity and preferences. Even the male participants were 8% less than female participant but the disparity is not too obvious. However, for the age group, it was obvious that participants mostly distributed in age group from 25 to 55. If the app developer would popularise their project, they might choose users aged 25 to 55 as their focus group. The analysis results also might help developers of IDEAL to adjust their new features released time. The relationship between participants' login month and new features released month showed that the concentrated released months of COMPARE_CHART and DUAL_CHART could be changed to March and April. The release weekdays did not need to be adjusted as they were not related to login date. The released hour of the new features in a day might need to be adjust to later ones to be closer to the peak of users' login hours. However, the data owner introduced that they sent email to notify participants when new features were released. There might be a period for participants to read the email and react

between new features released and they logged in. As a result, whether adjust the released hour in a day should be adjusted according to the specific situation developers known.

Conclusion

In conclusion, the achievement of this project is to help developers of IDEAL to define target users, improve the functions of the app and market strategy by understanding their feedback usage. The developer guide brochure communicated the groups' main findings about the app usage and the problems data owners most concerned. The brochure presented the data analysis in different aspect and gave suggestions of improving app. The key findings was successfully delivered to data owners via the final presentation and gained some positive comments from them. However, there were some defects in data analysis such as the frequency of logins of an individual was ignored. Further research should concentrate on several typical individual participants and the relationship between each new feature and the gender and age group of the participants who received it.