A Work Project, presented as part of the requirements for the Award of a Master Degree in Economics from the NOVA – School of Business and Economics.

HOW CONSUMERS ACQUIRE FINANCIAL LITERACY? AN INFORMATION ACQUISITION MODEL OF HUMAN CAPITAL.

EDWARD SOUTHERINGTON: 33039

A Project carried out on the Master in Economics Program, under the supervision of:

Carlos Santos

3/1/2020

Abstract:

This paper describes the Reddit online platform as a medium for financial literacy acquisition. Findings indicate the platform is a combination of smaller learning environments, within which consumers progress from learner to educator, and between which consumers progress, seemingly as their financial literacy accumulates. Within individual discussions five aspects of users' comments are found to be correlated with whether the consumer seeking financial literacy responds to and thanks this user. Leading to the conclusion that time, comment position within the discussion, Reddit metrics, text, and user experience are correlated with whether a consumer obtains financial literacy from a comment.

Keywords:

Financial Literacy Reddit

This work used infrastructure and resources funded by Fundação para a Ciência e a Tecnologia (UID/ECO/00124/2013, UID/ECO/00124/2019 and Social Sciences DataLab, Project 22209), POR Lisboa (LISBOA-01-0145-FEDER-007722 and Social Sciences DataLab, Project 22209) and POR Norte (Social Sciences DataLab, Project 22209).

1. Introduction

The developed world is experiencing record levels of personal debt. The population in the developed world lacks the financial literacy levels to optimize the opportunities facilitated by this debt, and there is an absence of affordable education interventions to address this growing imbalance. This results in increasing pressure on regulators to balance the trade-off between protecting financially illiterate consumers from opportunistic individuals and firms, and facilitating efficient capital allocation.

The world wide web is a relatively unresearched and untapped medium to expand consumer's financial literacy as a cost-effective means of education. The spread of online social media may be harnessed to provide timely, specific and ultimately efficient financial literacy education to the public.

One of the fastest growing areas of the internet is social news platforms, a pool of data that becomes increasingly more useful for analysis as the content contributed to it grows. In this paper the biggest topic driven social media platform, Reddit, is analysed. Whilst the focus of Reddit is not to provide education for the publics benefit, within it are environments in which a consumer can acquire specific personalized advice from the community, advice which is generally perceived as independent and usually based on other users own personal experiences rather than professional experiences. This learning environment contrasts to the generic information provided by the formal education sector and the personalized but incentivized financial advice that financial advisers provide for varying commission and other incentives. Within Reddit the specific topic of credit card literacy is analysed, credit cards are the financial tool where poor decision making by illiterate individuals has arguably the greatest impact per dollar of debt.

To better illustrate and understand the potential of this platform as a financial literacy learning environment it is described in two ways. Firstly, the way in which the community generally interacts in regards to credit card discussion on the platform, and secondly by describing factors within these consumer driven discussions that influence individual consumers proposed learning acquisition.

2. Literature Review

Financial literacy is commonly defined as "peoples' ability to process economic information and make informed decisions about financial planning, wealth accumulation, debt and pensions" (Lusardi and Mitchell, 2014). The acquisition of financial literacy and the growth of this form of human capital is important to society due to the positive relationship between financial literacy and financial outcomes (Benartzi and Thaler, 2007).

Nearly all financial literacy studies in the developed world found the surveyed population had overall concerningly low financial literacy. Financially vulnerable demographics in the community commonly perform the worst, including the elderly, woman, certain minorities and people with lower incomes and wealth (Lusardi, 2012).

This deficiency along with a deficiency in maths skills, has been found accountable for costlier decision-making behaviours. Studies on retirement and pension decisions, simple saving decisions and credit card decisions have found financial literacy levels impact an individual's decision-making behaviour (Lusardi and Mitchell, 2011; Mottola, 2013). Indicating this societal deficiency is leading to inefficient use of capital within the economy. The unequal distribution of financial literacy also indicates that this societal deficiency is exacerbating wealth inequality (Lusardi and Mitchell, 2014).

Findings indicate there is a disconnect between an individual's objective financial literacy level and their subjective financial literacy level. In Lusardi and Tufano (2015) it was concluded that overconfidence, specifically the elderly's overconfidence in their financial literacy and decision-making ability is a concern, and may partially explain the high incidence of financial frauds perpetrated against them.

In light of this widespread societal deficiency, it is not surprising that theoretical models predict societies would benefit from cost effective financial education interventions. Particularly if targeted at those with the lowest levels of financial literacy. In effect, promoting the type of decision making which would improve the allocation of capital and potentially alleviate some effects of inequality. Lusardi and Mitchell (2014) theoretical model predicts consumers would benefit from acquiring financial knowledge early in life even if they made no new investments thereafter. Finding the least educated consumers wellbeing would improve by the greatest percentage.

Empirical findings to date have not conclusively supported these models or the continuation of formal institutional intervention. The factor by which financial literacy depreciates has been found to be significantly larger than the theoretical models predict, with large interventions resulting in negligible effects after 20 months (Fernandes et all, 2013). The findings indicate formal education interventions are not only monetarily expensive, in some cases in the billions of dollars, the opportunity cost to consumers of supplanting other valuable activities that result in larger knowledge gains in seemingly comparable domains is even more costly (Fernandes et all, 2013).

Online financial literacy information has its own specific concern, uneven demographic uptake. Chin and Williams (2019) found that their financial literacy online information was predominantly taken up by first time buyers, the elderly, and consumers more objectively

knowledgeable. This last finding is consistent with “feeling of knowing” theories, and indicates these platforms, without intervention, may predominantly grow the human capital of those with the greatest financial literacy, arguably those with the lowest need. Without adaptations being made to these platforms they may also fail to address consumers overconfidence concerns, with the research finding subjectively knowledgeable individuals had low financial literacy uptake (Chin and Williams, 2019).

3. Methodology

In this paper four specific subreddits within Reddit are described; *personalfinance*, *churning*, *CreditCards* and *CRedit*. A subreddit is a board devoted to a specific topic within Reddit. All posts and comments are made within an associated subreddit in a treelike structure, an example of a Reddit discussion is illustrated in Appendix A. These four subreddits are the subreddits with the largest user membership with significant discussion relating to credit cards. Membership is not a requirement to post on a subreddit.

Each of the four subreddits, has a unique purpose, and as a result generally differing standards of financial literacy:

- ***personalfinance (pf)*** is the broadest of the subreddits with discussion in practice encompassing anything that is personal finance related with a minority of posts being about credit cards. Those that are, usually require low levels of credit card literacy.
- ***churning (ch)*** is the most specific subreddit with discussion predominantly on short to medium term strategies to profit from credit card promotions, rather than as a debt finance tool. Discussions usually involve the highest level of credit card literacy due to the complexity of owning numerous credit cards and utilising them for specific benefits.

- ***CRedit (CR)*** is a subreddit focused on individuals credit score. Credit card usage is very influential in credit score in the USA which results in a significant proportion of the discussion being credit card related. Discussion usually requires low credit card literacy.
- ***CreditCards (CC)*** is a subreddit focus on credit cards in general, but in practice, due to the existence of the other subreddits the predominant discussion involves medium to long term credit card strategy. Usually involving a low to moderate credit card literacy.

In this paper the Reddit credit card learning environment is analysed in two parts. Firstly, the analysis of the users, including overall trends of the subreddits, interaction levels of individuals, progression within subreddits and between subreddits. Secondly, the identification of aspects of the discussions between users that result in the original poster explicitly interacting in a manner consistent with credit card literacy acquisition.

3.1. Data

The data used in this descriptive analysis comes from fh-bigquery compiled by Reddit user /u/Stuck_In_the_Matrix, this is an unofficial database of Reddit discussions, no official data of this type is published. Two separate databases are analysed from this Big Query platform; the first contains the comments data from 2005 to 5/2019, the second contains the post data from 12/2015 to 8/2019. Comments are the submissions of users responding to an initial post on a subreddit dashboard. The majority of the analysis will use a combined data set of the two from 1/1/2016 to 30/5/2019.

A restructuring of *churning* in 2016, resulted in users' posts being predominantly posted inside administratively created daily posts, this resulted in these posts being classified as

comments in the dataset. After the initial analysis this data is restructured to reflect this reality. In most of this analysis only discussions stemming from posts containing questions are analysed. For three subreddits, this is done by restricting the discussions to those in which the post grammatically included a question. For *churning* the discussions were restricted to those posted inside administratively created daily question threads, which theoretically contain all user questions since 2016. The discussions on the broad subreddits; *CRedit* and *personalfinance* were also restricted to discussions stemming from posts containing text synonymous to credit cards.

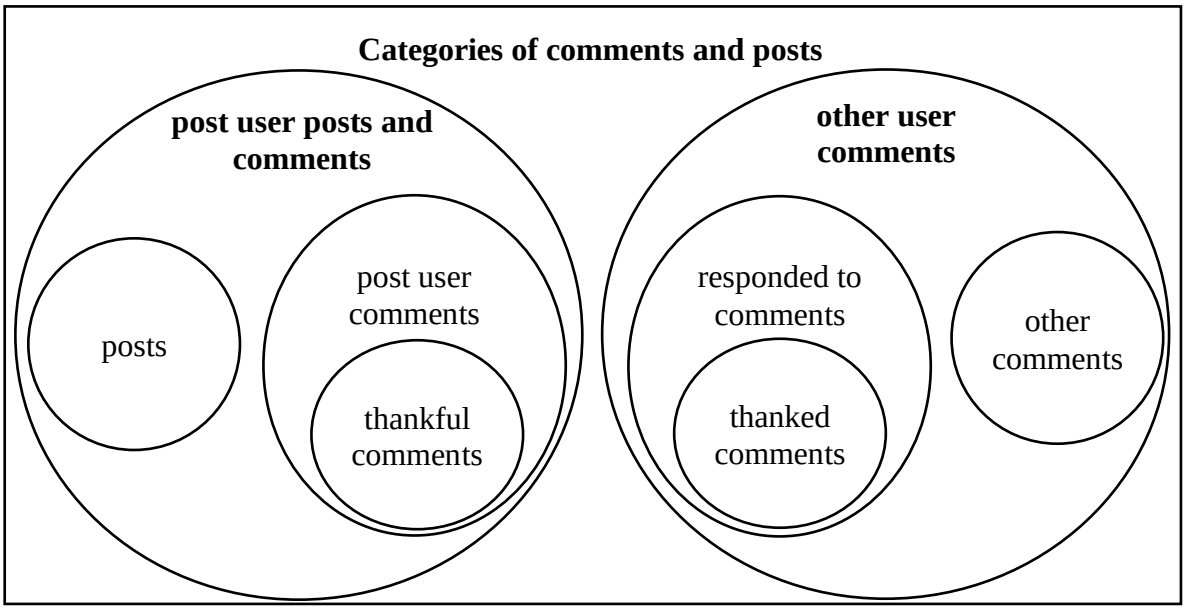
To isolate the post user interactions with the highest likelihood of involving financial literacy acquisition two aspects of the post users' own explicit interaction are identified within the discussion. The first is if the post user responds to a comment in the discussion, the second is if within this response the post user explicitly expresses gratitude. As discussions are restricted to those stemming from questions, comments are initially assumed to be responding to these questions. If the discussion branches off from this initial post, it is assumed further responses off that branch are responding to this branched off comment, which although on the same topic contains its own questions, answers and clarifications to be addressed by the community in their comments, as illustrated in Appendix A. The post user's explicit choice to respond to some of these comments whilst not responding to others, is an indication they believe it's a better use of their time to do so, it is assumed that this is primarily due to having an interest and potentially learning from these comments. The post user's explicit choice to express gratitude in these responses is assumed to be in return for being provided with useful financial information from which they learn.

Based on this understanding, comments are categorised into five groups;

- 'post user comments' - comments made by the post user,

- ‘thankful comments’ – ‘post user comments’ containing a version of thank you,
- ‘responded to comments’ - comments made by any user other than the post user which the post user then explicitly responds to,
- ‘thanked comments’ – ‘responded to comments’ which in the response are thanked.
- ‘other comments’ - comments made by any other user which are not responded to,

Figure 1. Venn Diagram of comments and post categories



3.2. The credit card literacy learning environment

To analyse how consumers, obtain financial literacy through Reddit, the environment and the user base is first described. This is done by first identifying the overall trend of three key metrics; posts, comment and users across the four sub learning environments year on year.

Secondly, using the data restricted to credit card questions, user contribution statistics are analyzed to identify the different types of users and their contribution of all comments, responded to comments and thanked comments.

Thirdly, user's progression within subreddits and between subreddits is analysed to identify the broader longer-term financial literacy acquisition trends. The former analyzed by comparing users' who both commented on their own posts and commented on other users' posts and then comparing the time between those two different types of comments, eq.1. The later was similarly done by comparing the time of comments that are posted on a specific subreddit, either on the users own post or on another user's post, with the timing of any comments made on other subreddits, either on their own posts or on others. The difference between these times indicating a general trend of users from one type of comment to the other over time. In each analysis, three comment pairs are compared; post user comments and other user comments, post user comments and responded to comments, thankful comments and thanked comments. The results an indication users transition between poster and commenter, between responder and responded to and between thankful and thanked.

$$\frac{\sum_{u=1}^{N}(\overline{comment\_time}_a - \overline{comment\_time}_b)_u}{N}$$

where: $\overline{comment\_time}_a$ = the mean time of subreddit a comments (utc)
$\overline{comment\_time}_b$ = the mean time of subreddit b comments(utc)
u = user who comments on both subreddit a and subreddit b
N= population of users who comment on both subreddit a and b

eq.1 – The average difference between the mean comment time of type a comments and type b comments, of all users who make a type a and b comment

3.3. The credit card literacy learning discussions

To identifying aspects of discussions within Reddit that lead to financial literacy acquisition, the comments that the community provide to a post user seeking advice are analysed and compared to see how the post user explicitly interacts with those comments. For this analyses the differentiation between all other comments, responded to comments and thanked

comments is again considered. Three separate dependent binary variables are generated based on these categorisations:

- ***responded_to*** – 1 if the comment is responded to and 0 if it is not.
- ***thanked*** – 1 if the comment is thanked and 0 if it is not.
- ***thanked_given*** – 1 if the comment is thanked given it is responded to and 0 if it is not thanked given it is responded to.

Five independent characteristics of these comments are then evaluated to identify correlations between these characteristics and these three independent variables. They relate to time, comment position, internal Reddit metrics, comment text, and user experience. From each category one or more specific independent variables are considered.

Time of comment – the length of time it takes a user to reply to a post user's comment is time the post user could be learning from other comments or other sources of information without the possibility of learning from this comment. Theoretically as this time increases the post users learning need from the comment decreases, and the likelihood of it being responded to and thanked decreases. ***delay_time*** specifically represents the time difference between the quickest response time to the post users comment and the response time of the comment being analysed.

Comment position – the positioning of a comment within the broader discussion determines the level of learning that the post user has already acquired within this post discussion and how far removed the comment is from the post user's comments, therefore theoretically how directly the comment addresses the post users comment. ***depth*** is a discrete variable representing the depth of the comment with respect to the initial post, in Appendix A indicated by the number of vertical lines before each comment text. ***immediate_post_depth*** is

a discrete variable indicating the depth of comment with respect to the most immediate post user comment. In effect, the difference between the depth of the comment analysed and the depth of the post users comment it is in response to.

Reddit score metrics – Reddit provides users with a method of voting in favour (upvote) and against (downvote) other user's comments and posts, from this a score is calculated which is the community's explicit valuation of the comment or post. ***post_score*** is the score of the post the comment is in response to and ***score*** is the score of the comment. These metrics generally measure both the societal value of the comment and the societal interest in a given post.

Text – the content and learning potential of a comment and therefore the likelihood a post user responds to and thanks another user's comment is dependent on the specific text of those discussions. In this model both the text of the comment analysed and the post users comment that this comment is responding to is analysed. ***comment_word_count*** is the number of words in the comment text, this represents the magnitude of the content and potentially the perceived amount of time expended by the user on that comment. ***direct_word_count*** is the number of words in the post users comment that the user's comment is responding to**,** an indication of the magnitude of content to be addressed by the comment analysed. ***word_count_comparison*** is the factor by which the comment word count is larger than the word count of the comment it is responding to (*comment_word_count* / *direct_word_count*). ***comment_question*** is a binary variable, 1 if the comment text contains a question, 0 if it does not. A question theoretically induces a response from the post user but not inherently one of gratitude. ***text_similarity*** is the cosine similarity of the comment text and the text of the post users comment, calculated using sklearn-metrics.pairwise.cosine_similarity library in python

(Pedregosa et al, 2011) . A user's utilisation of similar words to the post users own vocabulary is an indicate how well the post users' comment is addressed.

$$K(comment\_text, immediate\_text) = \frac{comment\_text \times immediate\_text^{\mathrm{T}}}{\|comment\_text\| \times \|immediate\_text\|}$$

where: comment_text= the comment text
immediate_text = the text of the post user text the comment is in response to
T = transpose of the vector

eq.2 – cosine similarity of the text of the comment and the text of the post users comment the comment is in response to

User experience – different users have different levels of knowledge and different levels of ability to communicate this knowledge effectively. ***comments_made_on_subreddit*** is the total number of comments the user has made within the subreddit that the comment is made in, an indication of the level of experience the user has. Theoretically an increase in experience would be correlated with an increase in their comment's financial literacy value.

The relationships between the variables were preliminarily analysed with preliminary findings. The results in Appendix B - Appendix D indicate low correlation between the dependent variables ***score***, ***word_count_comparison*** and ***user_subreddit_comments*** with the independent variables in some data sets. The results in Appendix E - Appendix F indicate minor multicollinearity concerns between certain independent variables. Correlation of 0.5 between ***comment_word_count*** and ***text_similarity*** in both the full data set and the full data set containing only comments that are responded to. Correlation of 0.7 between ***depth*** and ***immediate_post_depth*** in the full data set. This correlation potentially influencing the significance of these variables in the regression results. Histograms depicted in Appendix G - Appendix H of the independent variables indicate skewed distributions. Decision Trees illustrated in Appendix I - Appendix K indicate ***immediate_post_depth***, ***delay_time,***

comment_question, post_score and ***comment_word_count*** are useful independent variables in differentiating between comments that are thanked or responded to and those that are not.

Following the preliminary analysis, a series of logit regression models are analysed. For each of the dependent variables; ***responded_to***, ***thanked*** and ***thanked_given***, three different unconditional logistic regressions are reported. The three different regression include different dummies to account for different conditions. The dummies are:

Year – to control for long term trends which may affect the propensity for a user to respond or thank comments. This control variable is included in all regressions.

User – controls for individual characteristics of the user writing the comments, isolating the within effects. User dummies are included for the second of the three sets of logistic regressions.

Post User – controls for individual characteristics of the post user responding and thanking the comments, isolating the within effects. User dummies are included for the third of the three sets of logistic regressions.

User and Post User dummies do not represent characteristics of individuals that are definitively fixed over time, individuals' financial literacy and propensity to respond and thank changes over time. As this model is purely descriptive concerns as to constant heterogeneity of these two specifications has been relaxed and year dummy variable will control for some of this long-term user and post user changes.

Regression 1) Logit regression including the independent variables listed above, a constant, and year dummies, eq.3.

$$\ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 score + \beta_2 post\_score + \beta_3 commen\_word\_count + \beta_4 comment\_question + \beta_5 immediate\_post\_depth + \beta_6 depth + \beta_7 immediate\_word\_count + \beta_8 text\_similarity + \beta_9 delay\_time + \beta_{10} word\_count\_comparison + \beta_{11} total\_subreddit\_comments + \beta_{12} year\_2016 + \beta_{13} year\_2017 + \beta_{14} year\_2018$$

where: p = the probability of the independent variable occurring

eq.3 – Regression 1– Logit Regression including independent variables and year dummies

Regression 2) Same Logit regression as eq.3 with the inclusion of dummies for every user. Data restricted to users who comment more than once.

Regression 3) Same Logit regression as eq.3 with the inclusion of dummies for every post user. Data restricted to post users who comment more than once.

All three models are run on ten different data sets to identify if findings are consistent across the subreddits. The first data set is the full data set including all four subreddit comments, the second data set is the full data set with the independent variables winsorized at 5% to "robustify" the impact of extreme observations in skewed independent variables. Four regressions for each of the individual subreddit data sets. Four regressions for each of the individual subreddit data sets winsorized at 5%.

4. Results

4.1. The credit card literacy learning environment

Credit card related content on the platform Reddit has grown significantly since its inception, indicating that the platform has become an increasingly more useful source of financial literacy acquisition. Participation on all four subreddits has grown significantly since they were created; *CRedit* and *personalfinance* in 2009, *CreditCards* in 2011 and *churning* in 2012. Whilst this data includes discussions not necessarily related to credit cards, the data does indicate growing credit card learning environments albeit at a decreasing rate. All subreddits received less than 100 comments in their first year and respectively received 13197, 35673, 846182, 2346004 comments in 2018, with growth in some years exceeding 1000%, reported in Appendix L. The results however indicate there was significant decline in the number of posts, comments and users making these posts and comment in *churning* after the restructure of the subreddit in 2015. It is not clear the reason for this, it could indicate a more efficiently structured subreddit, saturation of information already posted on the subreddit or a decrease in the perceived value of the information provided on this subreddit. Later findings in Table 3 however indicate it is not due to users migrating to the other subreddits, which continue to grow almost every year.

4.1.1. User contribution within subreddits

The number of comments per person is positively skewed and the learning experience of hundreds of thousands of users rely on a relatively small percentage of users for that knowledge. Of the 218,216 users in the dataset, reported in Appendix M, a majority of users in three of the subreddits and 46% of churning users interact with only one post, reported in Appendix N. At the other end of the comment distribution, across the four subreddits,

between 3% and 7% of users contribute half of comments and 1% to 2% of users contributing half of the thanked comments, reported in Appendix O. Indicating that whilst a majority of users use the platform for one specific learning instance as the need to obtain financial literacy arises. A majority of the financial literacy provided to the community is contributed by a small percentage of users. The top 1% contributors contributing over a third of all comments on other users posts on all four subreddits, report in Appendix P.

4.1.2. User contribution between subreddits

Users who comment on multiple posts tend to stick to subreddits they are familiar with creating unique user environments. Users who contribute on multiple posts usually comment within the same subreddit multiple times rather than across multiple subreddits. The statistics in Table 1 illustrate that across all subreddits multi-post users contribute a larger percentage of comments than multi-subreddit users to a given subreddit. The difference between the two is most significant in the broadest and largest subreddit *personalfinance*. Only 3.3% of *personalfinance* users interact with other subreddits compared to the 30% who interact multiple times within that subreddit.

Table 1: Multi-subreddit user and Multi-post user comparison

Users that comment on multiple subreddits					
	Percentage of users	Percentage of all posts commented on	Percentage of all comments	Percentage of other users' posts commented on	Percentage of other user comments
pf	3.3%	51%	12%	49%	13%
ch	32%	66%	48%	63%	50%
CC	37%	97%	69%	97%	78%
CR	37%	94%	61%	93%	69%
Users that comment on multiple posts					
	Percentage of users	Percentage of all posts commented on	Percentage of all comments	Percentage of other users' posts commented on	Percentage of other user comments
pf	30%	99%	72%	99%	82%
ch	54%	97%	94%	99.5%	99%
CC	37%	99%	85%	99%	94%
CR	31%	99%	74%	99%	87%

This user base of only 3.3% however still interacts with 49% of other users posts, indicating how influential this small group is to this environment. In the smaller, more credit card specific subreddits, *CreditCards* and *CRedit,* the user base is significantly more fluid across subreddits. In *CRedit* more users interact with other subreddits (37%) than multiple posts within *CRedit* (31%).

4.1.3. Progression within subreddits

Across three of the four subreddits the timing of the comments by users who interact with multiple posts indicates users learn on their own posts and then go on to provide content that others then explicitly interact with in a manner consistent with learning. Across *personalfinance*, *CreditCards* and *CRedit* this progression is statistically and numerically significant with the time difference between the two types of interactions on average between 1 and 2 months (30-60 days), Table 2. Even though users who interact with multiple posts is a minority of users it is a significant finding given they contribute 82% - 94% of all other user comments in these subreddits, Table 1. Results in Table 2 show both *churning* and the full data set indicate the reverse relationship, albeit with a day difference of 15 days or less, the former is partially explainable by the statistics in Appendix M. *churning* users contribute more comments to more posts than other users, they are therefore involved in significantly more learning in a more complex area of credit card literacy, the result potentially indicating a more complex ongoing learning story. The combined data set includes users who change between subreddits, this progression is analysed further in Table 3.

Table 2: Number of days later users comment on others posts compared to on their own

Type of comment	*pf*	*ch*	*CC*	*CR*	*all 4*
All	48***	-15***	30***	36***	0
Responded to & responding	36***	-12***	37***	64***	-12***
Thanked and thankful	46***	-10***	41***	57***	-14***

*** statistically significant at 0.01 significance

4.1.4. Progression between subreddits

Table 3: Given users comment on both A and B, the number of days later B comments are than A comments

		A							
		other user's ***CR*** post	other user's ***CC*** post	other user's ***ch*** post	other user's ***pf*** post	own ***CR*** post	own ***CC*** post	own ***ch*** post	own ***pf*** post
B	other user's ***CR*** post	0 0 0	-17 -18 15	-80*** -112* -66	23** 51*** 57*	26 57** -17	-11 -4 -14	-170*** -153*** -154***	89*** 96*** 30
	other user's ***CC*** post	17** 18 -15	0 0 0	-45*** -41** -25	90*** 84*** 96***	37 80* 3	28*** 43*** 37***	-44*** -41*** -33*	158*** 181*** 235***
	other user's ***ch*** post	80*** 112* 66	45*** 41** 25	0 0 0	153*** 165*** 193***	172** 183* 110	107*** 136*** 122*	-14*** -1 2	305*** 315*** 354***
	other user's ***pf*** post	-23** -51*** -57*	-90*** -84*** -96***	-153*** -165*** -193***	0 0 0	50** 43 71	-54*** -78*** -85***	-173*** -178*** -190***	65*** 55*** 47*
	own ***CR*** post	-26 -57** 17	-37 -80* -3	-172*** -183*** -110***	-50** -43 -71	0 0 0	-39* -39* -41**	-167*** -168*** -115**	-5 -5 -20
	own ***CC*** post	11 4 14	-28*** -43*** -37***	-107*** -136*** -122***	54*** 78*** 85***	39* 39* 41**	0 0 0	-65*** -65*** -62***	87*** 87*** 73***
	own ***ch*** post	170*** 153*** 154***	44*** 41*** 33*	14 1 -2***	173*** 178*** 190***	167*** 168*** 115*	65*** 65*** 62***	0 0 0	246*** 246*** 253***
	own ***pf*** post	-89*** -96*** -30	-158*** -181*** 235***	-305*** -315*** -354***	-65*** -55*** -47*	5 5 20	-87*** -87*** -73***	-246*** -246*** -253***	0 0 0

*** statistically significant at 0.01 significance
** statistically significant at 0.05 significance
* statistically significant at 0.1 significance

All comment comparison
responded to & responding
thanked & thankful

Users who contribute across multiple subreddits tend to progress to subreddits that this paper subjectively assessed as more complex as human capital accumulates. The progression of the 3% to 37% of users who comment on multiple subreddits is a significant aspect for learning

on this platform given they contribute between 13% and 78% of comments on other users posts. The results in Table 3 indicate consistent progression exists in all three categories; all comments, responded to and responding comments, thanked and thankful comments, across multiple subreddits. Nearly all *churning* related results indicate a statistically significant progression to *churning* from other subreddits, with the average progression predominantly over 100 days later. This indicates that although *churning* has a shrinking user base, reported in Appendix L, users are still predominantly progressing towards this subreddit not away from it. In Table 3 progression to the *CreditCards* from *personalfinance* is also evident. This progression to *CreditCards* unlike that of churning coincides with it being the fastest growing user base over the years analysed, reported in Appendix L. *CRedit* statistical significance is low across many of the comparisons partially due to the lower number of users.

4.2. Information acquisition model

The results of the logit regressions in Table 4 indicate all five comment characteristics; time, comment position, Reddit metrics, comment text and user experience are correlated with the log likelihood of a post user responding to and thanking comments. These results are generally supported by the complete set of logistic regressions reported in Appendix Q – Appendix Y. Table 4 presents the results of the, three different regressions on the three different independent variables using the full data set.

Time - The results in Table 4 indicate the time financial literacy is conveyed to a consumer is significantly correlated with whether the comment is then learnt from. Eight of the nine coefficients of the ***delay_time*** variable are negative at 0.01 significance, as time increases the log likelihood of being responded to or thanked decreases. This result is also consistent

across nearly all the individual subreddit data regressions in Appendix Q – Appendix Y with respect to the independent variables ***responded_to*** and ***thanked***.

Table 4: Logit regression results on the full data set - sign and significance of coefficients

		Regression 1	Regression 2	Regression 3
Time Variable				
delay_time	*responded_to*	-***	-***	-***
	thanked	-***	-***	-***
	thanked_given	-***	-***	-
Comment Position Variables				
depth	*responded_to*	+***	+***	- ***
	thanked	+***	+***	- ***
	thanked_given	-***	-***	+***
immediate_post_depth	*responded_to*	-***	-***	-***
	thanked	-***	-***	-***
	thanked_given	-***	-***	-***
Reddit Metrics Variables				
score	*responded_to*	+***	+***	+***
	thanked	+***	+***	+***
	thanked_given	+***	+	+***
post_score	*responded_to*	-***	-***	-***
	thanked	-***	-***	-***
	thanked_given	-***	-***	-***
Text Variables				
comment_word_count	*responded_to*	-	+***	+***
	thanked	+***	+***	+***
	thanked_given	+***	+***	+***
immediate_word_ count	*responded_to*	+***	+***	+***
	thanked	+***	+***	+
	thanked_given	-***	-**	-*
word_count_ comparison	*responded_to*	-***	+	+
	thanked	-***	-***	+*
	thanked_given	-***	-***	-
comment_question	*responded_to*	+***	+***	+***
	thanked	-***	-***	-***
	thanked_given	-***	-***	-***
text_similarity	*responded_to*	+***	+***	+***
	thanked	+***	+***	+***
	thanked_given	+***	+***	+***
User Variable				
total_subreddit_ comments	*responded_to*	+***	NA	+
	thanked	+***	NA	+***
	thanked_given	+***	NA	+**

*** - statistically significant at 0.01 significance
** - statistically significant at 0.05 significance
* - statistically significant at 0.1 significance

The ***thanked_given*** finding however is not consistent with the findings in Appendix Q – Appendix Y indicating that if the comment comes too late, given it is responded to is not necessarily significant on whether it is thanked across certain subreddits.

Comment Position - The findings in Table 4 indicate a comments position within the discussion is significant correlated with whether the post user learns from it. The ***depth*** variable results indicate that in the regression if the post user characteristics are not controlled for, the deeper the comment is made within a discussion the more likely a comment is responded to and thanked. The coefficient of ***thanked_given*** regressions is however negative indicating these results are increasingly more likely to be responded to but not thanked. In the Regression 3, results with post user dummies, the results are the reverse indicating the previous results may be primarily due to post user characteristics. Comments with high ***depth*** disproportionately provided by users who respond to and thank more than the average post user. These results are generally consistent with Appendix Q – Appendix Y results with only the smallest data set *CRedit* being inconsistent. The ***immediate_post_depth*** results in Table 4 indicate the less direct the response is to a post users comment within the discussion structure the less likely it is to be thanked and responded to. These results are generally consistent across the Appendix Q – Appendix Y results with the exception of the winsorized data where too much of the variance in this variable is removed.

Reddit Metrics - The findings in Table 4 indicate Reddit's internal metrics for scoring are correlated with the likelihood a comment is being learnt from. ***score*** is positive across eight of the nine regressions indicating as score increases the likelihood of the post user responding or thanking the comment increases. ***post_score*** coefficient results are the opposite as ***post_score*** increases the likelihood of responding or thanking the comment decreases.

post_score is usually correlated with the number of comments on a post, this result therefore may rather reflect that as the number of comments increases the chance that any one comment is responded to or thanked decreases. These results are generally supported in the Appendix Q – Appendix Y results although the results of the ***thanked_given,*** Regression 2, are only significant in three of the nine regressions indicating the finding may be subreddit specific.

Text - The text variables in Table 4 indicate the specific content of a comment is correlated with the log likelihood of a post user learning from a comment. The longer the comment, the more similar the words used in the comment are correlated with an increased likelihood of comment getting responded to and thanked. If the comment contains a question this is correlated with an increase in the chance of a response but correlated with a decrease in the likelihood of being thanked. These findings are generally consistent in the subreddit specific regressions in Appendix Q – Appendix Y. The ***word_count_comparison*** results of Table 4 and the Appendix Q – Appendix Y are however relatively inconsistent with respect to ***responded_to*** and the results in Regression 3. This result may be partially due to multicollinearity.

User - The user variable in Table 4 and supported in the Appendix Q – Appendix Y results indicate more experienced users comments are positively correlated with the log likelihood of a post user learning from their comments. ***total_subreddit_comments***, is positively correlated across all metrics. This variable was not included in Regression 2 due to lack of variation in this variable in these regressions.

5. Further Research

This paper provides a broad introduction to the financial literacy learning opportunity that is facilitated by social news platforms. In this paper users' explicit posts and comments are analysed. This paper does not consider the millions of views that these platforms receive and the thousands of interactions with respect to user voting habits which are not captured by the metrics in this data. These two areas offer unresearched areas that can be further researched.

This research tested general and basic elements of comments, identifying their relationship with the dependent variables. Numerous other variables within the text can be analysed with a more comprehensive text analysis, such as those implemented by multinational technology companies. Further analysis of how post users respond may facilitate a more accurate means of quantifying financial literacy acquisition. Alternatively, if a metric on a similar platform in which a user could explicitly make a subjective assessment of the value they believe each comment has provided them was included, this would enable more accurate analysis and machine learning. User data may also offer demographic data that could be researched to understand which demographics are obtaining financial literacy.

6. Conclusion

This paper shows the platform Reddit supports at least four different credit card literacy learning environments, contributed by unique user communities, which users progress between over time. The majority of content developed is contributed by users who contribute multiple times, these users providing a key financial literacy distribution role. We find that these users tend to progress from poster to commenter, responder to responded to and thankful to thanked. Users also progress from the subreddits with the least complex financial

discussion to subreddits with more complex financial discussion, finding that this progression can be on average over 100 days later in some specific subreddits. Within these discussions aspects of a given comment lend that comment to be more or less likely to be thanked and responded to, comments which seem to provide the greatest content for other users' financial literacy acquisition. The findings indicate it is the quickest responses, the longest responses, the most direct response, utilising the most similar wording as the original post users own wording that tend to be both responded to and thanked. These comments tend to receive the highest score and be from users more experienced in providing comments on the subreddit, further reinforcing the importance of these high interaction users. Whether the user asks a question or not seems to indicate whether a comment is purely responded to or thanked. Given a comment is responded to, a user with more experience on the subreddit is more likely to be responded to whilst if the comment is significantly larger than the text which it is responding to the reverse is true.

The Reddit platform has provided answers to hundreds of thousands of posts. Through further research and investment these platforms could facilitate a more optimal level of financial literacy education which other interventions have failed to provide. Whilst a social media platform is unlikely to become the panacea to this societal deficiency in the short term, if harnessed effectively utilising the existing content and trends it may provide an efficient means of providing consumer specific education on a significant and influential scale that as a part of a broader strategy may be an efficient allocation of state capital.

Reference List

Benartzi, Shlomo, and Richard Thaler. 2007. "Heuristics and biases in retirement savings behavior." *Journal of Economic perspectives* 21, no. 3. 81-104.

Chin, Alycia, and Alanna K. Williams. 2019. "Take-Up of Financial Education: Demographic Characteristics and Prior Knowledge." Journal of Public Policy & Marketing, 0743915619858928.

Fernandes, Daniel, John G. Lynch Jr, and Richard G. Netemeyer. 2004. "Financial literacy, financial education, and downstream financial behaviors." *Management Science* 60, no. 8, 1861-1883.

Lusardi, Annamaria. 2012. *Numeracy, financial literacy, and financial decision-making*. No. w17821. National Bureau of Economic Research.

Lusardi, Annamaria, and Olivia S. Mitchell. 2011b. “Financial Literacy and Retirement Planning in the United States.” *Journal of Pension Economics and Finance* 10 (4): 509–25.

Lusardi, Annamaria, and Olivia S. Mitchell.2014. "The economic importance of financial literacy: Theory and evidence." *Journal of economic literature* 52, no. 1, 5-44.

Lusardi, Annamaria, and Peter Tufano. 2015. "Debt literacy, financial experiences, and overindebtedness." *Journal of Pension Economics & Finance* 14, no. 4, 332-368.

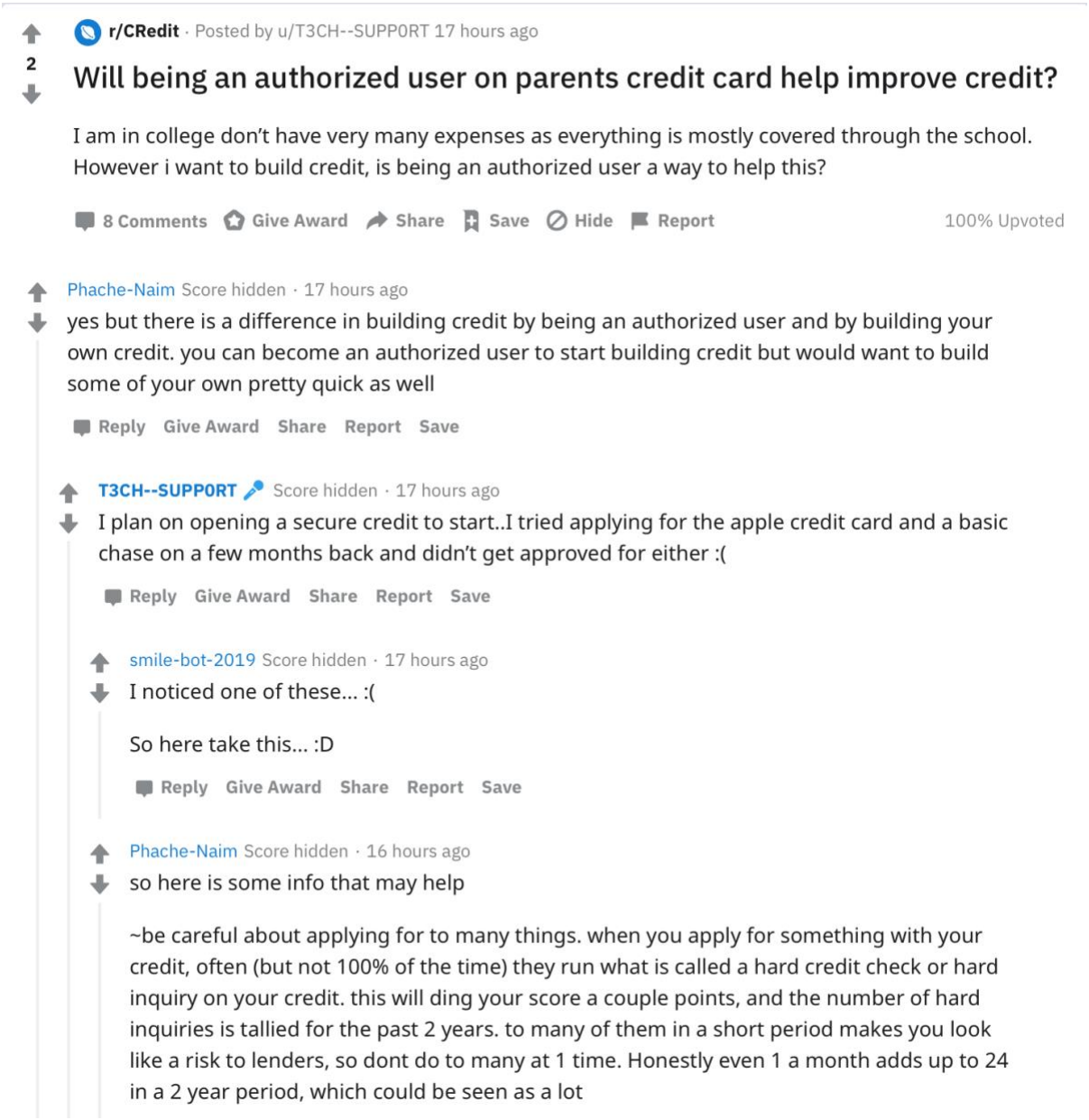
Mottola, Gary R. 2013."In our best interest: Women, financial literacy, and credit card behavior." *Numeracy* 6, no. 2, 1-15.

Pedregosa, Fabian, et al. 2011."Scikit-learn: Machine learning in Python." *Journal of machine learning research* 12, no, 2825-2830.

T3CH—SUPPORT, Phache-Naim, smile-bot-2019, LV-Knight, “Will being an authorized user on parents credit card help improve credit?” *Reddit.* Accessed December 31, 2019 //www.reddit.com/r/CRedit/comments/ehv9m2/will_being_an_authorized_user_on_parents_credit/.

Appendix

Appendix A: Example of a Reddit post



r/CRedit · Posted by u/T3CH--SUPP0RT 17 hours ago

2

Will being an authorized user on parents credit card help improve credit?

I am in college don't have very many expenses as everything is mostly covered through the school. However i want to build credit, is being an authorized user a way to help this?

8 Comments Give Award Share Save Hide Report 100% Upvoted

Phache-Naim Score hidden · 17 hours ago

yes but there is a difference in building credit by being an authorized user and by building your own credit. you can become an authorized user to start building credit but would want to build some of your own pretty quick as well

Reply Give Award Share Report Save

T3CH--SUPP0RT Score hidden · 17 hours ago

I plan on opening a secure credit to start..I tried applying for the apple credit card and a basic chase on a few months back and didn't get approved for either :(

Reply Give Award Share Report Save

smile-bot-2019 Score hidden · 17 hours ago

I noticed one of these... :(

So here take this... :D

Reply Give Award Share Report Save

Phache-Naim Score hidden · 16 hours ago

so here is some info that may help

~be careful about applying for to many things. when you apply for something with your credit, often (but not 100% of the time) they run what is called a hard credit check or hard inquiry on your credit. this will ding your score a couple points, and the number of hard inquiries is tallied for the past 2 years. to many of them in a short period makes you look like a risk to lenders, so dont do to many at 1 time. Honestly even 1 a month adds up to 24 in a 2 year period, which could be seen as a lot

~if you do become an authorized user on your parents credit, many credit card companies will backdate the credit history for you. in other words, if your parents opened the card 5 years ago, you will get all 5 years of credit history on your credit report; note that not all credit cards do this, but many do, especially most of the large credit card companies like discover or Capital one. once you have all of that history on your account, your score will jump up pretty quick. with a higher score you will likely qualify for better credit cards easier. you may not qualify for the best cards ever or a 10K limit, but you can probably get a regular, non secured credit card in your own name

~Credit is a long term game. dont worry to much about it right now. your main concern should be avoiding any negative info for now. negative info is so much easier to get on your credit than positive info.

Reply Give Award Share Report Save

T3CH--SUPP0RT Score hidden · 16 hours ago

sounds good! my parents are on the capital one for 10+ years so hopefully that carries over. After i get the discover secured card, I will not apply for a credit card for at least 8-10 months until I graduate to a real credit card

Reply Give Award Share Report Save

Phache-Naim Score hidden · 16 hours ago

one thing i forgot to mention: being an Authorized user (AU) comes with the good and bad. if they have 10 years of great history, that really helps. if they have 10 years of spotty history and missed payments, that also carries over. if they miss any payments or get sent to collections while you are an AU, that also shows on your credit. so typically getting a good history card is best, and you want to take yourself off from being an AU at the first sign of trouble to protect your credit

Discover does let you monitor your FICO score as well from transunion data for free, so thats a good place to see your score and how you are doing.

Reply Give Award Share Report Save

T3CH--SUPP0RT Score hidden · 16 hours ago

great thanks! I am pretty my parents have good credit history both are 750+ but I will find out when i get the discover card haha

Reply Give Award Share Report Save

LV-Knight Score hidden · 16 hours ago

I'm still an authorized user on a business AMEX card that I sold like 3 years ago! 🤣 I'm cool with it because the bills always get paid and it's helped my credit a lot.

Reply Give Award Share Report Save

(T3CH—SUPPORT et all, 2019)

Appendix B: Linear correlation with *responded_to*

Variable	All 4	personal-finance	churning	CRedit	Credit-Cards
thanked	0.497	0.500	0.468	0.540	0.492
comment_question	0.165	0.169	0.192	0.137	0.138
user_subreddit_comments	0.115	0.108	0.130	0.022	0.094
text_similarity	0.087	0.105	0.113	0.106	0.144
comment_word_count	0.072	0.086	0.108	0.079	0.103
year	0.032	0.014	0.022	0.017	-0.019
word_count_comparison	0.023	0.018	0.005	0.024	0.013
score	0.006	0.009	0.047	0.008	0.064
delay_time	-0.037	-0.035	-0.015	-0.085	-0.059
direct_word_count	-0.054	-0.040	0.098	0.028	0.057
depth	-0.116	-0.168	0.025	-0.091	-0.040
post_score	-0.150	-0.157	0.003	-0.089	-0.048
immediate_post_depth	-0.269	-0.274	-0.192	-0.214	-0.239

Appendix C: Linear correlation with *thanked*

Variable	All 4	personal-finance	churning	CRedit	Credit-Cards
responded_to	0.497	0.500	0.468	0.540	0.492
comment_word_count	0.131	0.145	0.115	0.158	0.140
text_similarity	0.104	0.117	0.108	0.139	0.134
user_subreddit_comments	0.059	0.053	0.065	0.048	0.051
year	0.032	0.014	0.022	0.017	-0.019
score	0.005	0.007	0.103	0.020	0.078
word_count_comparison	0.000	0.001	-0.028	0.002	-0.010
comment_question	-0.008	0.001	-0.040	-0.019	-0.026
direct_word_count	-0.021	-0.014	0.072	0.044	0.050
delay_time	-0.022	-0.021	-0.015	-0.047	-0.032
depth	-0.069	-0.095	-0.001	-0.063	-0.040
post_score	-0.088	-0.094	0.000	-0.065	-0.023
immediate_post_depth	-0.139	-0.142	-0.092	-0.118	-0.121

Appendix D: Linear correlation with *thanked_given*

Variable	All 4	personal-finance	churning	CRedit	Credit-Cards
comment_word_count	0.199	0.218	0.106	0.239	0.176
text_similarity	0.138	0.153	0.097	0.185	0.138
direct_word_count	0.016	0.017	0.043	0.067	0.048
year	0.032	0.014	0.022	0.017	-0.019
score	0.003	0.004	0.144	0.120	0.099
user_subreddit_comments	0.003	-0.001	0.007	0.081	0.009
delay_time	-0.013	-0.013	-0.020	-0.008	-0.014
word_count_comparison	-0.023	-0.015	-0.054	-0.019	-0.029
depth	-0.028	-0.033	-0.021	-0.038	-0.042
post_score	-0.050	-0.056	-0.003	-0.064	0.001
immediate_post_depth	-0.051	-0.056	-0.018	-0.013	-0.031
comment_question	-0.178	-0.173	-0.205	-0.182	-0.180

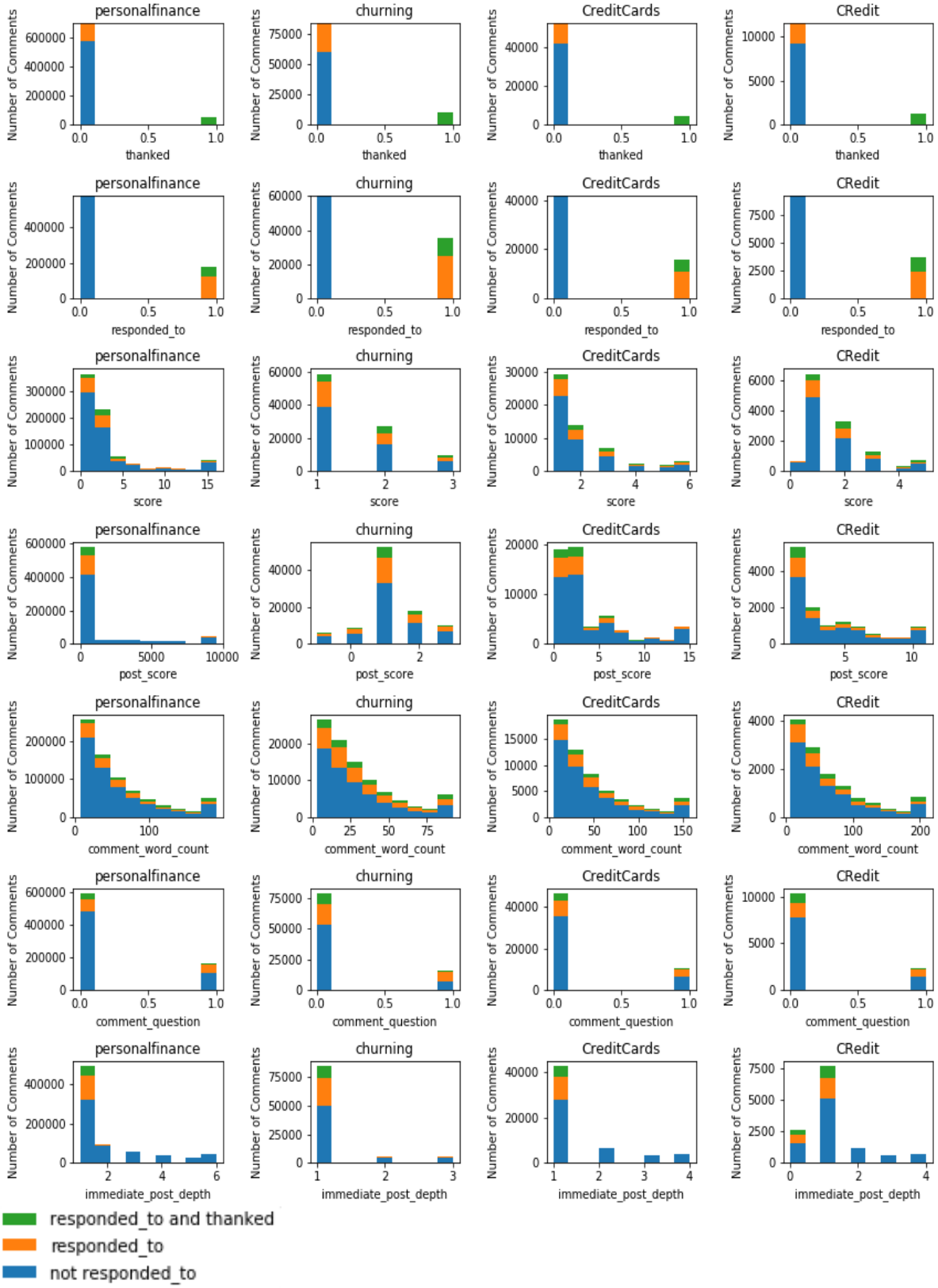
Appendix E: Correlation between independent variables, all four subreddits

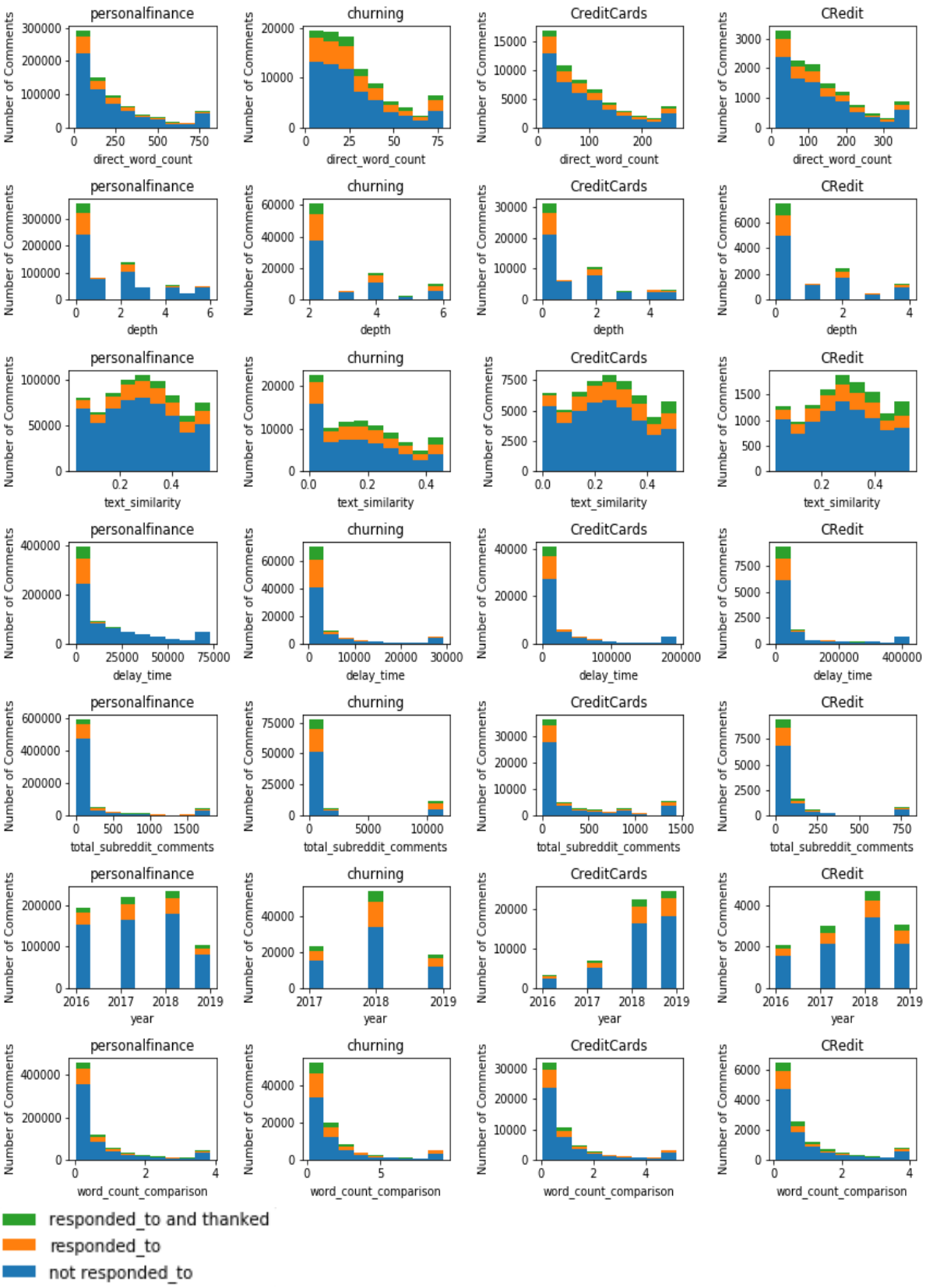
	score	post_score	comment_word_count	comment_question	immediate_post_depth	submitter_word_count	depth	text_similarity	delay_time	user_subreddit_comments	year	word_count_comparison
score	1.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
post_score	0.1	1.0	0.0	0.0	0.2	0.3	0.1	0.0	0.0	-0.1	0.0	0.0
comment_word_count	0.0	0.0	1.0	0.1	-0.1	0.1	-0.1	0.5	0.0	0.0	-0.1	0.1
comment_question	0.0	0.0	0.1	1.0	0.0	0.0	-0.1	0.1	0.0	0.0	0.0	0.0
immediate_post_depth	0.0	0.2	-0.1	0.0	1.0	0.1	0.7	-0.1	0.0	-0.1	0.0	0.0
direct_word_count	0.0	0.3	0.1	0.0	0.1	1.0	-0.1	0.3	0.0	-0.1	-0.1	-0.1
depth	0.0	0.1	-0.1	-0.1	0.7	-0.1	1.0	-0.2	0.0	0.1	0.0	0.1
text_similarity	0.0	0.0	0.5	0.1	-0.1	0.3	-0.2	1.0	0.0	0.0	-0.1	-0.1
delay_time	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0
user_subreddit_comments	0.0	-0.1	0.0	0.0	-0.1	-0.1	0.1	0.0	0.0	1.0	0.0	0.0
year	0.0	0.0	-0.1	0.0	0.0	-0.1	0.0	-0.1	0.0	0.0	1.0	0.0
word_count_comparison	0.0	0.0	0.1	0.0	0.0	-0.1	0.1	-0.1	0.0	0.0	0.0	1.0

Appendix F: Correlation between independent variables given *responded_to, all four subreddits*

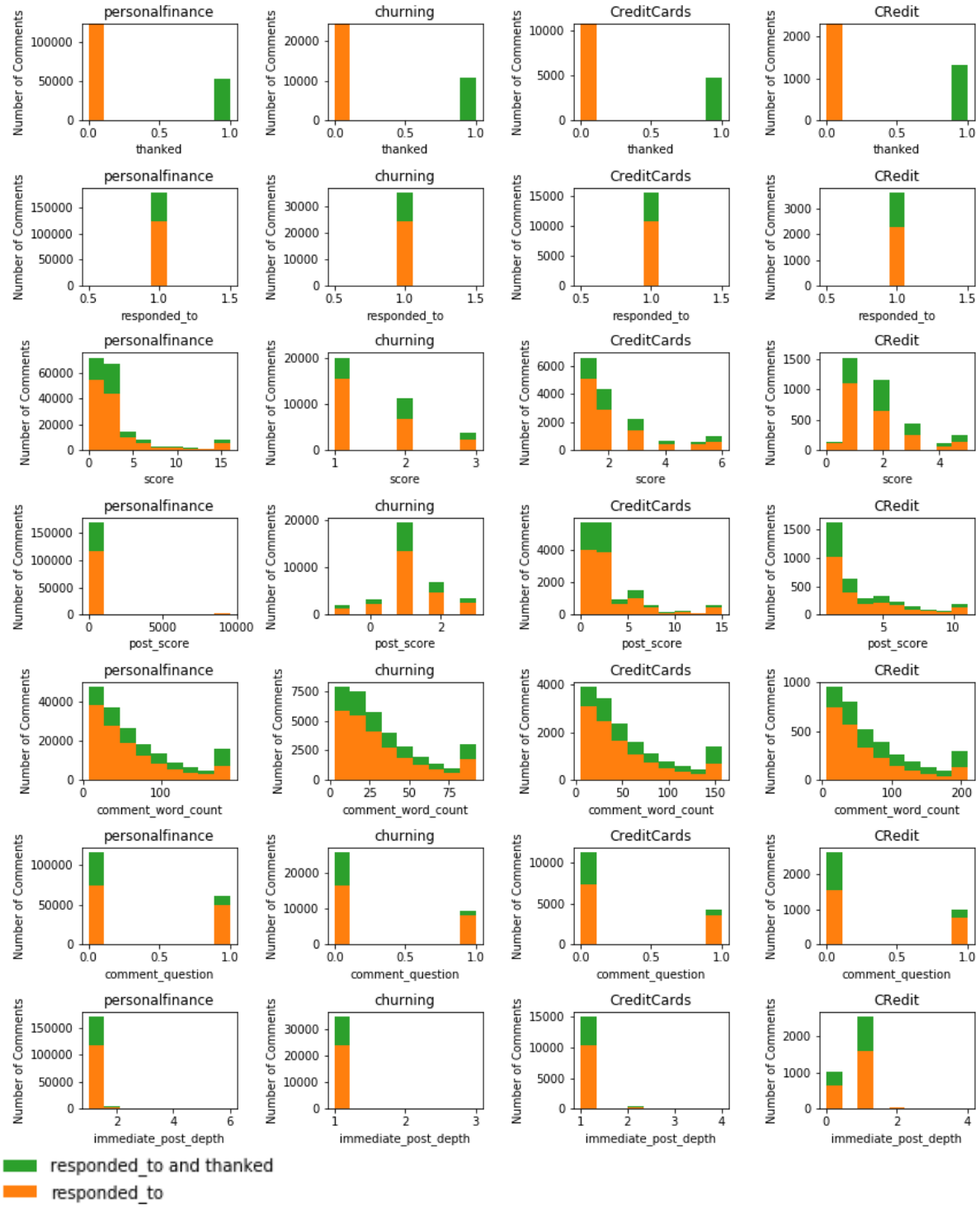
	score	post_score	comment_word_count	comment_question	immediate_post_depth	submitter_word_count	depth	text_similarity	delay_time	user_subreddit_comments	year	word_count_comparison
score	1.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
post_score	0.1	1.0	0.0	0.0	0.2	0.2	0.0	0.0	0.0	0.0	0.0	0.0
comment_word_count	0.0	0.0	1.0	0.1	0.0	0.2	-0.1	0.5	0.0	-0.1	-0.1	0.1
comment_question	0.0	0.0	0.1	1.0	0.0	0.1	-0.1	0.1	0.0	0.0	0.0	0.0
immediate_post_depth	0.0	0.2	0.0	0.0	1.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0
direct_word_count	0.0	0.2	0.2	0.1	0.0	1.0	-0.3	0.3	0.0	-0.1	-0.1	-0.1
depth	0.0	0.0	-0.1	-0.1	0.1	-0.3	1.0	-0.2	0.0	0.2	0.1	0.1
text_similarity	0.0	0.0	0.5	0.1	0.0	0.3	-0.2	1.0	0.0	0.0	-0.1	-0.2
delay_time	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0
user_subreddit_comments	0.0	0.0	-0.1	0.0	0.0	-0.1	0.2	0.0	0.0	1.0	0.1	0.0
year	0.0	0.0	-0.1	0.0	0.0	-0.1	0.1	-0.1	0.0	0.1	1.0	0.0
word_count_comparison	0.0	0.0	0.1	0.0	0.0	-0.1	0.1	-0.2	0.0	0.0	0.0	1.0

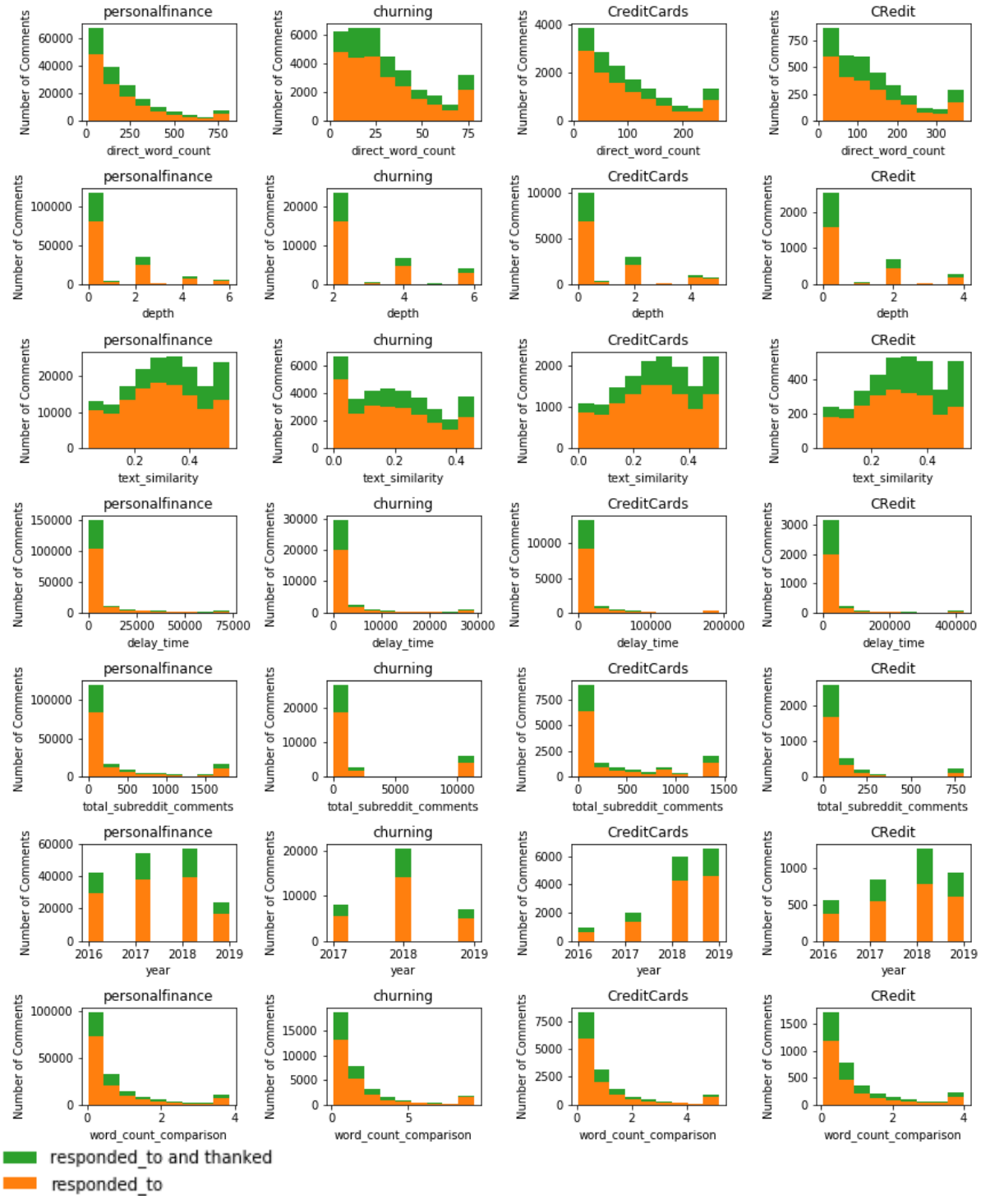
Appendix G: Histograms of variables across all four subreddits, colours indicating comment category - Data has been winsorized at 5% for visualisation due to long tails



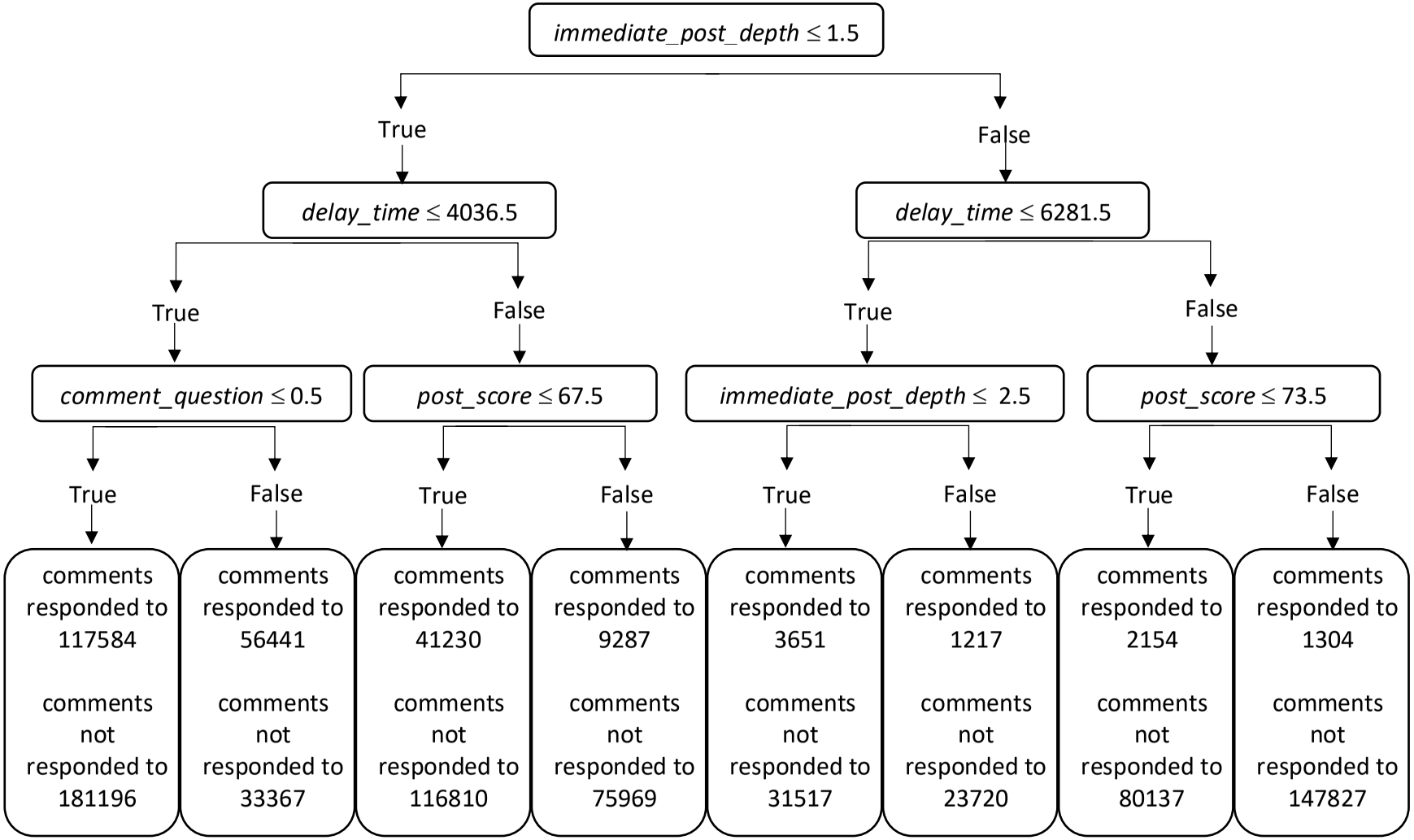

personalfinance
churning
CreditCards
CRedit
Number of Comments
direct_word_count
depth
text_similarity
delay_time
total_subreddit_comments
year
word_count_comparison
responded_to and thanked
responded_to
not responded_to

Appendix H: Histograms of variables across all four subreddits - *responded_to* data - Data has been winsorized at 5% for visualisation due to long tails

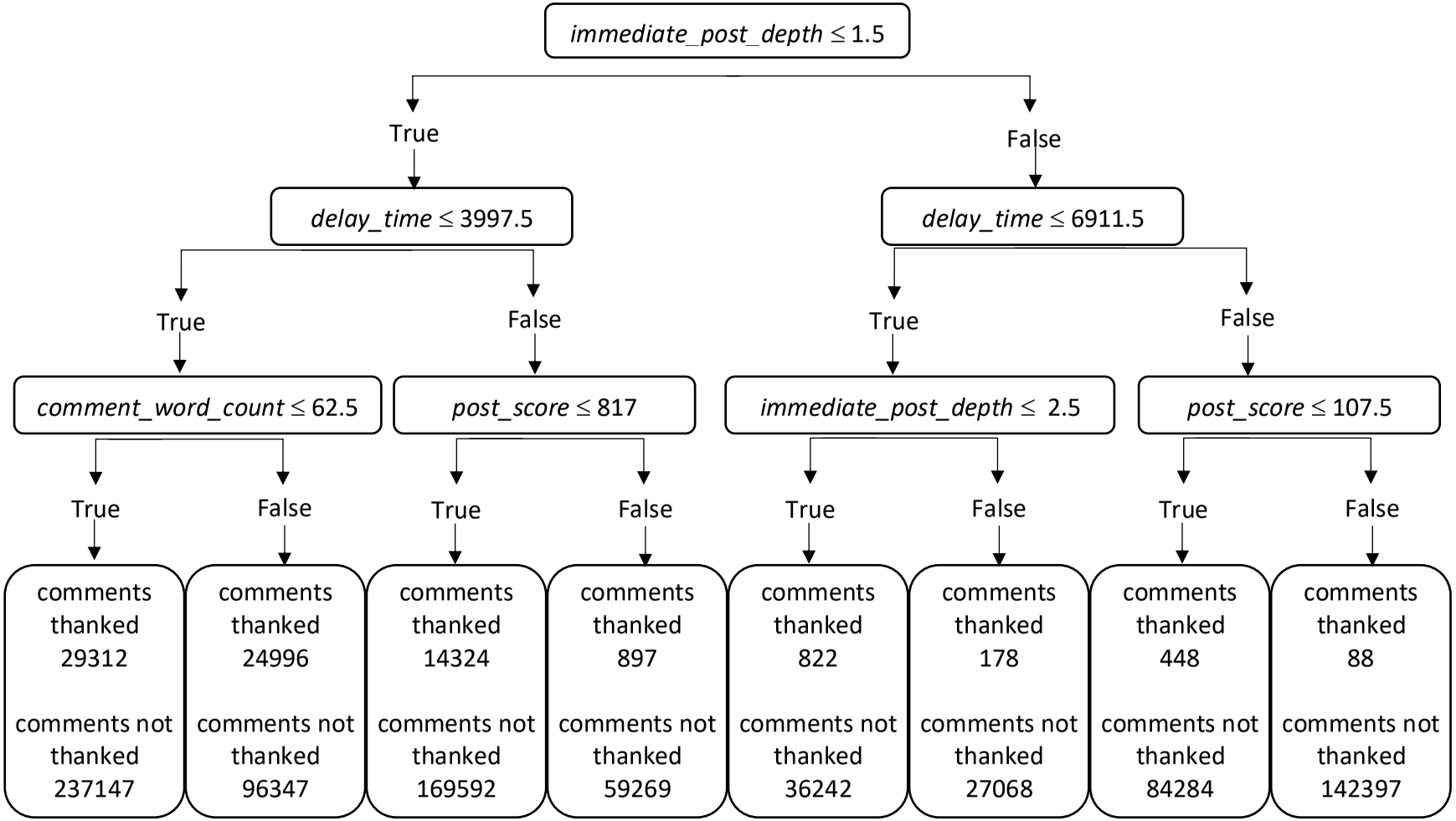



personalfinance
churning
CreditCards
CREdit
Number of Comments
direct_word_count
depth
text_similarity
delay_time
total_subreddit_comments
year
word_count_comparison
responded_to and thanked
responded_to

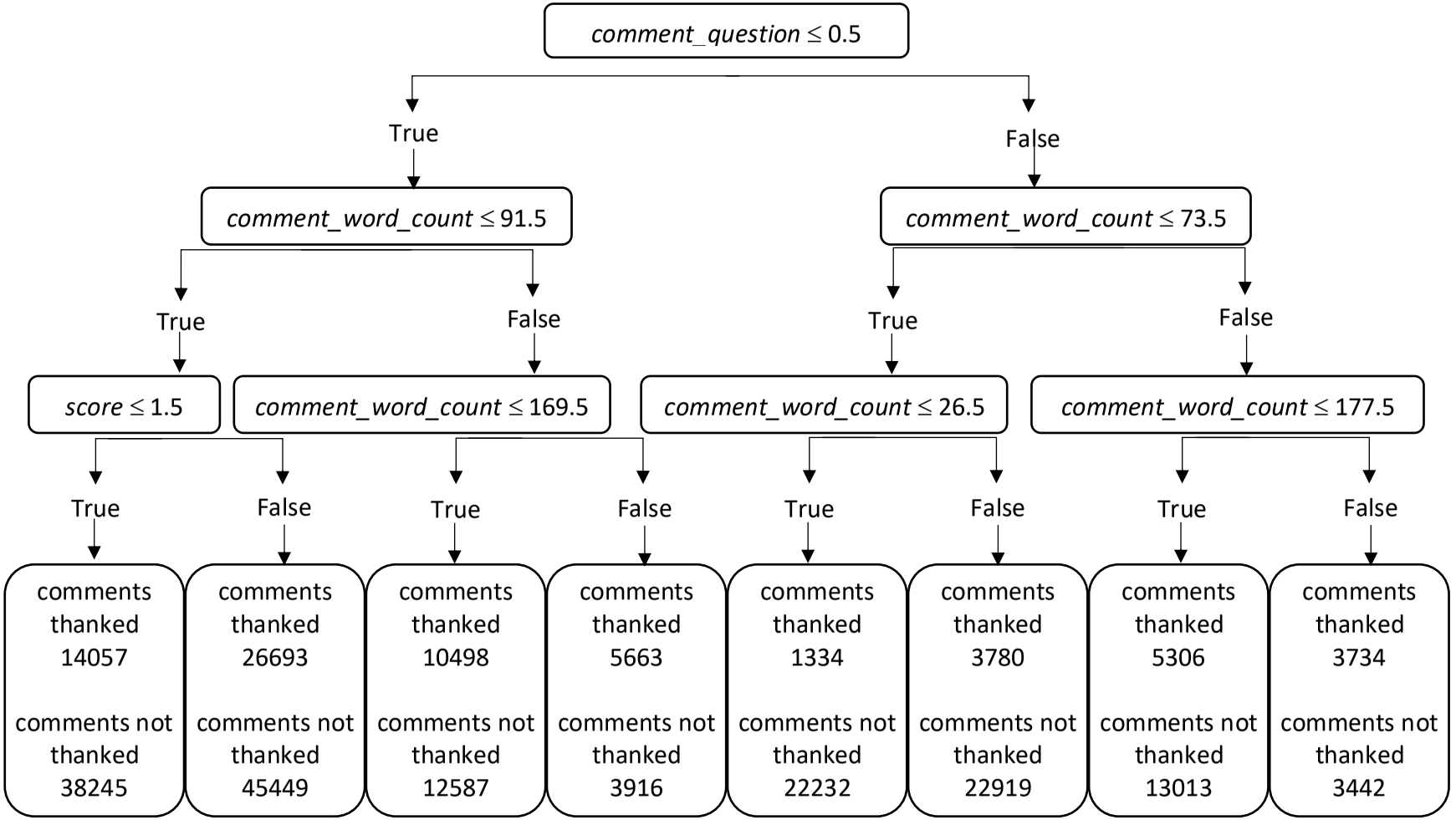
Appendix I: Decision Tree – *responded_to*



Appendix J: Decision Tree – *thanked*



Appendix K: Decision Tree – *thanked_given*



Appendix L: Yearly number of posts, comments and users across the four subreddits

		posts			comments			users		
subreddit	year	total	absolute change	percentage change (%)	total	absolute change	percentage change (%)	total	absolute change	percentage change (%)
CR	2009	1	NaN	NaN	1	NaN	NaN	1	NaN	NaN
	2010	4	3	300	4	3	300	3	2	200
	2011	4	0	0	4	0	0.	2	-1	-33
	2012	25	21	525	54	50	1250	27	25	1250
	2013	78	53	212	174	120	222	61	34	126
	2014	895	817	1047	4192	4018	2309	857	796	1305
	2015	819	-76	-8	4202	10	0.2	927	70	8
	2016	927	108	13	5162	960	23	1125	198	21
	2017	1345	418	45	7772	2610	51	1548	423	38
	2018	1911	566	42	13197	5425	70	2348	800	52
	2019	1274	1146*	60*	8499	7200*	55*	1542	1352*	58
CC	2011	1	NaN	NaN	1	NaN	NaN	1	NaN	NaN
	2012	1	0	0	1	0	0	1	0	0
	2013	7	6	600	16	15	1500	10	9	900
	2014	134	127	1814	389	373	2331	125	115	1150
	2015	596	462	345	2766	2377	611	582	457	366
	2016	1134	538	90	5164	2398	87	1118	536	92
	2017	1994	860	76	11347	6183	120	1997	879	79
	2018	4636	2642	133	35673	24326	214	4432	2435	122
	2019	4474	6101*	132*	40080	60519*	170	4354	6017*	136*
ch	2012	18	NaN	NaN	93	NaN	NaN	28	NaN	NaN
	2013	993	975	5417	8962	8869	9537	1207	1179	4211
	2014	4467	3474	350	55219	46257	516	4644	3437	285
	2015	11265	6798	152	203140	147921	268	10044	5400	117
	2016	9079	-2186	-19	496647	293507	144	18513	8469	84
	2017	5129	-3950	-44	953357	456710	92	23449	4936	27
	2018	2734	-2395	-47	846182	-107175	-11	22205	-1244	-5
	2019	889	-600*	-22*	298343	-130158*	-15*	12286	7281*	33*
pf	2009	11	NaN	NaN	11	NaN	NaN	7	NaN	NaN
	2010	102	91	827	259	248	2255	89	82	1171
	2011	3550	3448	3380	48670	48411	18692	5386	5297	5952
	2012	10792	7242	204	182847	134177	276	16615	11229	208
	2013	24600	13808	128	416805	233958	128	38117	21502	129
	2014	57617	33017	134	956272	539467	129	115692	77575	204
	2015	106041	48424	84	1460845	504573	53	189510	73818	64
	2016	125148	19107	18	1736217	275372	19	232653	43143	23
	2017	148006	22858	18	2142353	406136	23	275862	43209	19
	2018	146046	-1960	-1	2346004	203651	10	302596	26734	10
	2019	64990	9930*	7*	973710	-9100*	-0.4*	153991	66982*	22*

***Projected based on utilising 5 months data, projection simple multiplication (12/5)**

Appendix M: Basic user interaction statistics

	Users	Posts	Post interactions per user	Comments	Comments per user
pf	199,658	75,064	2.96	945,156	4.73
ch	14,369	68,057	9.59	226,547	15.76
CC	9,040	9,973	5.18	73,694	8.15
CR	3,534	2,488	3.03	16,753	4.74
all 4	218,216	155,582	3.61	1,262,150	5.78

Appendix N: Interaction statistics of users who only interact with one post

	Users	Percentage of total users	Percentage of total comments	Percentage of non-post user comments	Percentage of responded to comments	Percentage of thanked comments
pf	140,386	70%	28%	18%	11%	10%
ch	6,541	46%	6%	1%	1%	1
CC	5,671	63%	15%	6%	4%	4%
CR	2,425	69%	26%	13%	11%	9%

Appendix O: The fewest users that contribute 50% of comments

Type of comment	*pf*	*ch*	*CC*	*CR*
All comments	4%	4%	3%	7%
Non-post user comments	3%	2%	2%	3%
Responded to	1%	1%	1%	3%
Thanked	1%	1%	1%	2%

Appendix P: Interaction statistics of the users who contribute the top 1% of comments

	Users	Percentage of total users	Percentage of total comments	Percentage of non-post user comments	Percentage of responded to comments	Percentage of thanked comments
pf	1991	1%	32%	39%	54%	55%
ch	144	1%	31%	49%	54%	55%
CC	90	1%	36%	46%	52%	54%
CR	35	1%	26%	34%	37%	41%

Appendix Q: Logit regressions- independent variables and year dummies

Dependent Variable: *responded_to*

	All 4		personal-finance		churning		CreditCards		CRedit	
winsorised at 5%	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
const	+	+	+***	+***	+***	+***	+***	+***	+	+
comment_question	+***	+***	+***	+***	+***	+***	+***	+***	+***	+***
comment_word_count	-	+***	-***	+***	+***	+***	+	+***	+	+**
delay_time	-***	-***	-***	-***	-	-***	-***	-***	-***	-***
depth	+***	+***	+***	+***	+***	+***	+***	+***	-***	-***
immediate_post_depth	-***	-***	-***	-***	-***	-***	-***	-***	-***	-***
post_score	-***	-***	-***	-***	+**	+***	-***	-***	-***	-***
score	+***	+***	+***	+***	+***	+***	+***	+***	+***	+***
Immediate_word_count	+***	+***	+***	+***	+***	+***	+***	+***	+***	+***
total_subreddit_ comments	+***	+***	+***	+***	+***	+***	+***	+***	-	-
text_similarity	+***	+***	+***	+***	+***	+***	+***	+***	+***	+
word_count_comparison	+***	-***	+**	-***	+	+***	+	-***	-	-***
year_2016	-***	-***	-***	-***	-***	-**	+	+**	-**	-
year_2017	-	+*	+*	+***	-***	-***	-	+	-	+
year_2018	+	+	+***	+***	-**	-**	-***	-***	-	-
Pseudo R-square	0.177	0.213	0.187	0.229	0.110	0.120	0.147	0.166	0.123	0.142
No. Observations	923411		757838		95545		57233		12795	

*** - statistically significant at 0.01 significance
** - statistically significant at 0.05 significance
* - statistically significant at 0.1 significance
Pseudo R-square - McFadden's pseudo-R-squared.

Appendix R: Logit regressions- independent variables and year dummies

Dependent Variable: *thanked*

	All 4		personal-finance		churning		CreditCards		CRedit	
winsorised at 5%	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
const	-	-	-***	-***	-***	-***	-***	-***	-***	-***
comment_question	-***	-***	-***	-***	-***	-***	-***	-***	-***	-***
comment_word_count	+***	+***	+***	+***	+***	+***	+***	+***	+***	+***
delay_time	-***	-***	-***	-***	-**	-**	-***	-***	-***	-***
depth	+***	+***	+***	+***	+***	+***	+***	+***	-***	-***
immediate_post_depth	-***	-***	-***	-***	-***	-***	-***	-***	-***	-***
post_score	-***	-***	-***	-***	-	-	-***	-***	-***	-***
score	+***	+***	+***	+***	+***	+***	+***	+***	+***	+***
immediate _word_count	+***	+***	+***	+***	+**	+	+***	+***	+***	+*
total_subreddit_ comments	+***	+***	+***	+***	+***	+***	+***	+***	+***	+***
text_similarity	+***	+***	+***	+***	+***	+***	+***	+***	+***	+
word_count_comparison	-***	-***	-***	-***	-***	-***	-***	-***	-**	-***
year_2016	-***	-***	-***	-***	-	-	+***	+***	-	-
year_2017	-	-	-	-	-*	-*	+***	+**	-	-
year_2018	-	-	+	-	+	+	-***	-***	-	-
Pseudo R-square	0.131	0.155	0.146	0.171	0.069	0.080	0.103	0.121	0.104	0.131
No. Observations	923411		757838		95545		95545		12795	

*** - statistically significant at 0.01 significance
** - statistically significant at 0.05 significance
* - statistically significant at 0.1 significance
Pseudo R-square - McFadden's pseudo-R-squared.

Appendix S: Logit regressions- independent variables and year dummies

Dependent Variable: *thanked_given*

	All 4		personal-finance		churning		CreditCards		CRedit	
winsorised at 5%	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
const	-	-	-***	-***	-***	-***	-***	-***	-***	-**
comment_question	-***	-***	-***	-***	-***	-***	-***	-***	-***	-***
comment_word_count	+***	+***	+***	+***	+***	+***	+***	+***	+***	+***
delay_time	-***	-***	-***	-**	-	-	-	-	+	-
depth	-***	-***	-***	+	-***	-***	-***	+	-**	-**
immediate_post_depth	-***	-	-***	NA	-***	NA	-***	NA	-***	-***
post_score	-***	-***	-***	-***	-	-**	+	-***	-***	-***
score	+***	+***	+***	+***	+***	+***	+***	+***	+***	+***
immediate_word_count	-***	-***	-***	-***	-***	-***	-	-**	+	-
total_subreddit_ comments	+***	+***	+***	+***	+***	+***	+***	+*	+***	+***
text_similarity	+***	+***	+***	+***	+***	+***	+***	+*	+**	+
word_count_comparison	-***	-***	-***	-***	-***	-***	-***	-***	-***	-***
year_2016	-***	-***	-***	-***	+	+	+***	+***	-	-
year_2017	-	-	-*	-*	+	+	+***	+***	-*	-*
year_2018	-	-	-**	-**	+***	+***	-	-*	-	+
Pseudo R-square	0.077	0.077	0.082	0.085	0.073	0.078	0.071	0.073	0.104	0.108
No. Observations	232868		178337		35384		15517		3630	

*** - statistically significant at 0.01 significance
** - statistically significant at 0.05 significance
* - statistically significant at 0.1 significance
Pseudo R-square - McFadden's pseudo-R-squared.
NA - not enough variation in variable to winsorise at 5%

Appendix T: Logit regressions- independent variables, year and user dummies

Dependent Variable: *responded_to*

	All 4		personal-finance		churning		CreditCards		CRedit	
winsorised at 5%	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
score	+***	+*	+**	+**	+***	+***	+***	+***	+*	+***
post_score	-***	-***	-***	-***	+	+***	-***	-***	-***	-***
comment_word_count	+***	+***	+***	+***	+***	+***	+**	+**	+	+**
comment_question	+***	+***	+***	+***	+***	+***	+***	+***	+***	+***
immediate_post_depth	-***	-***	-***	-***	-***	-***	-***	-***	-***	-***
immediate_word_count	+***	+***	+***	+***	+***	+***	+***	+***	+***	+**
depth	+***	+***	+***	+***	+***	+***	+***	+***	-***	-***
text_similarity	+***	+***	+***	+***	+***	+***	+***	+***	+***	+*
delay_time	-***	-***	-***	-***	-	-***	-***	-***	-***	-***
word_count_comparison	+	-	+	+	+	+***	+	-***	-*	-***
year_2016	-***	-***	-***	-***	-***	-*	-	+	+	+
year_2017	-	-	+***	-	-***	-***	-	+	+	+
year_2018	-	-	+**	-	-***	-***	-*	-	+	+
Pseudo R-square	0.204	0.205	0.194	0.189	0.129	0.136	0.287	0.181	0.131	0.150
No. Observations	224914		91988		71673		31413		6076	

*** - statistically significant at 0.01 significance
** - statistically significant at 0.05 significance
* - statistically significant at 0.1 significance
Pseudo R-square - McFadden's pseudo-R-squared.

Appendix U: Logit regressions- independent variables, year and user dummies

Dependent Variable: *thanked*

	All 4		personal-finance		churning		CreditCards		CRedit	
winsorised at 5%	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
score	+***	+***	+***	+**	+***	+***	+***	+***	+**	+***
post_score	-***	-***	-***	-***	-	+	-***	-***	-***	-***
comment_word_count	+***	+***	+***	+***	+***	+***	+***	+***	+***	+***
comment_question	-***	-***	-***	-***	-***	-***	-***	-***	-**	-***
immediate_post_depth	-***	-***	-***	-***	-***	-***	-***	-***	-***	-***
immediate_word_count	+***	+***	+***	+***	+***	+***	+***	+***	+***	+
depth	+***	+***	+***	+***	+***	+***	+***	+***	-***	-**
text_similarity	+***	+***	+***	+***	+***	+***	+***	+***	+***	+
delay_time	-***	-***	-***	-***	-	-***	-***	-***	-***	-***
word_count_comparison	-***	-***	-***	-***	-***	-***	-***	-***	-	-***
year_2016	-	-	+	+	-	-	+**	+*	-	+
year_2017	-	-	+	+	-*	-**	+	+	-	+
year_2018	+	+	+	+	+	+	-*	-*	+	+
Pseudo R-square	0.153	0.153	0.149	0.151	0.0.89	0.095	0.124	0.133	0.121	0.146
No. Observations	201398		91988		71673		31661		6076	

*** - statistically significant at 0.01 significance
** - statistically significant at 0.05 significance
* - statistically significant at 0.1 significance
Pseudo R-square - McFadden's pseudo-R-squared.

Appendix V: Logit regressions- independent variables, year and user dummies

Dependent Variable: *thanked_given*

	All 4		personal-finance		churning		CreditCards		CRedit	
winsorised at 5%	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
score	+	+***	+	+*	+***	+***	+***	+***	+***	+***
post_score	-***	-***	-***	-*	-	-	-	-	-	-
comment_word_count	+***	+***	+***	+***	+***	+***	+***	+***	+***	+***
comment_question	-***	-***	-***	-***	-***	-***	-***	-***	-***	-***
immediate_post_depth	-***	NA	-***	NA	-***	-*	-***	-***	-***	-***
immediate_word_count	-**	-***	+	-**	-	-**	-	-***	+	-
depth	-***	+**	+	-***	-***	-	+	+	-*	-
text_similarity	+***	+***	+***	-***	+***	+***	+***	+*	+**	+
delay_time	-	-**	+	-***	-	-	-	-	-*	-
word_count_comparison	-***	-***	-***	-***	-***	-***	-***	-***	-	-***
year_2016	-**	-	-***	-***	-	-	+	+	-**	-*
year_2017	+	+	-	-	+	+	+	+	-*	-
year_2018	-	-***	-	-	+***	+***	-***	-**	-**	-*
Pseudo R-square	0.089	0.091	0.102	0.097	0.083	0.087	0.078	0.082	0.143	0.147
No. Observations	113386		89172		16796		6159		1259	

*** - statistically significant at 0.01 significance
** - statistically significant at 0.05 significance
* - statistically significant at 0.1 significance
Pseudo R-square - McFadden's pseudo-R-squared.
NA - not enough variation in variable to winsorise at 5%

Appendix W: Logit regressions- independent variables, year and post_user dummies

Dependent Variable: *responded_to*

	All 4		personal-finance		churning		CreditCards		CRedit	
winsorised at 5%	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
score	+***	+***	+***	+***	+***	+***	+***	+***	+***	+***
post_score	-***	-***	-***	-***	+***	+***	-***	-***	-***	-***
comment_word_count	+***	+***	+***	+***	+***	+***	+***	+***	+***	+***
comment_question	+***	+***	+***	+***	+***	+***	+***	+***	+***	+***
immediate_post_depth	-***	-***	-***	-***	-***	-***	-***	-***	-***	-***
Immediate_word_count	+***	+***	+	+***	+***	+***	+***	+***	+***	+***
depth	-***	-***	-***	-***	-***	-***	-***	+	-***	-***
text_similarity	+***	+***	+***	+***	+***	+***	+***	+***	+***	+**
delay_time	-***	-***	-***	-***	-	-***	-***	-***	-***	-***
word_count_comparison	+	+	+***	-***	+	+***	+***	+	+***	-
total_subreddit_ comments	+***	+***	+***	+***	+***	+***	+***	+***	+***	+
year_2016	-**	-**	-*	-	-	-	+***	+***	+***	-*
year_2017	-**	-	+	+	+***	-***	+***	+***	+***	-***
year_2018	-	+	+**	+*	-	-**	-	-	-	-***
Pseudo R-square	0.354	0.383	0.321	0.348	0.187	0.199	0.287	0.309	0.289	0.311
No. Observations	199190		91351		70958		31413		5468	

*** - statistically significant at 0.01 significance
** - statistically significant at 0.05 significance
* - statistically significant at 0.1 significance
Pseudo R-square - McFadden's pseudo-R-squared.

Appendix X: Logit regressions- independent variables, year and post_user dummies

Dependent Variable: *thanked*

	All 4		personal-finance		churning		CreditCards		CRedit	
winsorised at 5%	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
score	+***	+***	+***	+***	+***	+***	+***	+***	+***	+***
post_score	-***	-***	-	-	-***	-***	-***	-***	-***	-***
comment_word_count	+***	+***	+***	+***	+***	+***	+***	+***	+***	+***
comment_question	-***	-***	-***	-***	-***	-***	-***	-***	-***	-***
immediate_post_depth	-***	-***	-***	-***	-***	-***	-***	-***	-***	-***
immediate_word_count	+	+***	+***	+***	+	+**	+***	+***	+***	+***
depth	-***	-***	-*	+	-	-	-	+*	-	-***
text_similarity	+***	+***	+***	+***	+***	+***	+***	+	+***	+
delay_time	-***	-***	-***	-***	-**	-***	-***	-***	-***	-***
word_count_comparison	+*	-	+***	-	-	+	-**	-**	-**	-
total_subreddit_ comments	+***	+***	+***	+*	+***	+***	+***	+***	+***	+*
year_2016	-	-	-	-	+	+	+**	+*	+**	-
year_2017	-	+	+	+	-	-	+**	+*	+**	-
year_2018	+	+	+	+	+	+	+**	+*	+**	-
Pseudo R-square	0.296	0.324	0.378	0.407	0.186	0.196	0.296	0.312	0.265	0.287
No. Observations	199190		91351		70958		31413		5468	

*** - statistically significant at 0.01 significance
** - statistically significant at 0.05 significance
* - statistically significant at 0.1 significance
Pseudo R-square - McFadden's pseudo-R-squared.

Appendix Y: Logit regressions- independent variables, year and post_user dummies

Dependent Variable: *thanked_given*

	All 4		personal-finance		churning		CreditCards		CRedit	
winsorised at 5%	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
score	+***	+***	+***	+***	+***	+***	+***	+***	+***	+***
post_score	-***	-***	-***	-***	-***	-***	-***	-**	-	-
comment_word_count	+***	+***	+***	+***	+***	+***	+***	+***	+***	+***
comment_question	-***	-***	-***	-***	-***	-***	-***	-***	-***	-***
immediate_post_depth	-***	-*	-***	-***	-***	-***	-***	-	-***	-**
immediate_word_count	-*	-*	+***	+***	-***	-***	-**	-**	+	+
depth	+***	+***	+***	+***	+***	+***	+***	+***	+**	+*
text_similarity	+***	+***	+***	+*	+***	+***	+	-	+	-
delay_time	-	-**	-*	+	-	-	+	-**	+	+
word_count_comparison	-***	-***	-***	-***	-***	-***	-**	-***	-	-
total_subreddit_comments	+**	+***	+	+**	+***	+***	+*	+	+*	+*
year_2016	+	+	-	-	+	+	+	+	-	-
year_2017	-	-	-**	-*	+	+	+	+	+	+
year_2018	-	+	-**	-**	+**	+**	+*	+*	+	+
Pseudo R-square	0.248	0.248	0.259	0.263	0.197	0.197	0.229	0.228	0.235	0.235
No. Observations	106614		66384		28607		9747		1876	

*** - statistically significant at 0.01 significance
** - statistically significant at 0.05 significance
* - statistically significant at 0.1 significance
Pseudo R-square - McFadden's pseudo-R-squared.