

Abstract

Community-driven online question and answer forums (CQA), platforms that allow users to ask and answer questions, contain an extensive amount of crowd-sourced knowledge. Thousands of questions and answers are posted by users everyday. One such example of a CQA is iFixit’s *Answers* forum. This forum features user-asked questions related specifically to device repair, which are answered by both repair experts and everyday users. A reliable measure of the health of these CQAs is question response time. Fast response times enhance user engagement and satisfaction and increase web traffic. This paper presents a survival analysis of the time until a question receives its first answer. A Cox proportional hazards model was developed to predict the survival probability of a question, the probability that a question receives an answer before a certain time. Though several predictors were significant, the model’s predictive accuracy was low ($R^2 = 0.15$). Significant predictors included the device category of the question (questions pertaining to Apple products received answers faster than others (HR = 2.55, 95% CI = (2.32, 2.80))), factors related to the question’s title (e.g., whether or not it was phrased as a question (HR = 1.29, 95% CI = (1.21, 1.38))), and the day of the week the question was posted (questions posted over the weekend received answers slower than those posted on a weekday (HR = 0.88, 95% CI = (8.20, 9.34))). Future studies can investigate if factors identified as significant in this analysis can be generalized to other CQAs.

1. Introduction

Community-driven online question and answer forums (CQA) are becoming widely-used sources of information. These online platforms feature thousands of user-posted questions and answers and can receive millions of visits every month. The CQA analyzed in this paper is iFixit’s *Answers* forum. Founded in 2003, iFixit’s mission is to equip users with the knowledge and tools necessary to repair their broken devices, as part of an effort to save money and reduce electronic waste. This company provides over 30,000 free online repair guides and sells the specialized tools and parts needed for such repairs.

As not all possible repairs are covered in the published guides and users may have additional questions related to existing guides, iFixit’s *Answers* forum is an important resource for users. This platform features questions pertaining to over 9,000 devices, ranging from jammed zippers to shattered phone screens and faulty air conditioners, with over 100,000 solutions. As thousands of users rely on this forum for information, it is important that users receive timely answers. Fast response times enhance user experience and thus increase web traffic, which is valuable to the reputation and longevity of iFixit’s *Answers* forum. Analysis of response times can reveal factors that affect how quickly questions receive answers, which can lead to suggestions for how users can ask better questions to minimize response times and for how the forum design can be improved.

However, analysis and prediction of response times on CQAs has not been

thoroughly investigated. There is need for further analysis of response times in these forums, as the majority of existing research focuses on assessing and predicting question and answer quality. This paper presents a survival analysis on the time until a question receives its first answer on iFixit’s *Answers* forum, in order to determine factors significantly related to answer time and to predict the survival probability of a question.

2. Related Work

With the recent increase in the popularity and use of CQAs, these data-rich platforms have been the subject of a multitude of studies. Regarding analysis of response times, [?] developed a classification model using both textual and non-textual features to estimate response times for questions posted on *Stack Overflow*. [?] used parametric methods and proposed models to predict response times based on exponential distributions. [?] instead focused exclusively on unanswered questions and developed a taxonomy for classifying such questions on *Stack Overflow*, and created a model to predict how long a question would remain unanswered. However, all research mentioned restricted analysis to either questions for which response times were observed or to questions that remained unanswered. The research presented in this paper uses the framework of survival analysis, a method not yet applied to CQAs, to allow for analysis of both answered and unanswered questions in a unified framework.

The majority of existing research has been focused on predicting question and answer quality. [?] developed a classification algorithm to assess the quality of posts on *Nabble.com* using primarily textual features of questions. [?] also determined that textual features, i.e word count, was the most accurate predictor of Wikipedia article quality (Wikipedia articles can represent the same kind of user-generated content featured on CQAs). [?] instead found that non-textual features, like revision and comment count on an answer or the number of points a user has, were the most useful indicators of answer quality. A number of other studies have also developed classification algorithms using both textual and non-textual features with the similar goal of predicting question or answer quality [?] [?] [?] [?]. [?] took analysis further and presented models to predict the long term value of questions as well as the answers it receives. The present analysis seeks to utilize both textual and non-textual features of questions on iFixit’s *Answers* forum to investigate if factors indicating that a question is of high quality, can also lead to faster response times. (fix this section)

3. Materials

The data analyzed contained 8,025 questions posted from April 8, 2017 to July 7, 2017 (the date the data was downloaded). Variables in the data included: device name and category, title, text, tags, whether or not the user was a member of iFixit’s site for less than one day before the question was posted, date and

| |
|---|
| Device category the question pertains to. Categories include: AndroidOther Phone, Apple Products, Camera, Electronics, Game Console, Home, Other, PC, Tablet, Vehicle |
| If the question was posted on the weekend or on a weekday |
| Whether or not the question’s text contains any end punctuation marks (. ? !) |
| Whether or not the question’s text is in all lower case |
| Whether or not the question’s title contains at least one word that is considered “frequently used” among answered questions. See Appendix for a complete list of these terms. |
| Whether or not the question’s title contains at least one word that is considered “frequently used” among unanswered question. See Appendix for a complete list of these terms. |
| Whether or not the question’s title ends in a question mark |
| Whether or not the user edited or added information to the question’s text sometime after posting it. |
| Whether or not the user made an effort to solve the problem prior to asking the question. |

Table 1: Categorical variables derived

| |
|--|
| Average number of characters in each question’s tags |
| Average tag “score” for all of a question’s tags. A tag score is defined as the proportion of times a tag appears in all of the data. Questions without tags were assigned a score of 0. |
| Number of characters in the question’s text |
| Number of characters in the user-defined device name |
| Ratio of the number of line breaks to the number of characters in the question’s text |

Table 2: Continuous variables derived

time the question was posted, and date and time the first answer was received. Variables derived can be found in Table 1 and 2. See Appendix for more details on variable derivation.

4. Methods

Questions analyzed were restricted to those posted in English. The time until event variable used in survival analysis was defined as the time since posting until a question receives its first answer. For questions that did not receive an answer by the download date, time until event values were defined as the time since posting to the time the data was downloaded. Such questions were considered right-censored, meaning that exact answer times for these questions are greater than the observed download time (questions may receive answers after the download date) [?].

Survival was defined as the event that a question did not receive an answer beyond a certain time, t . Estimates of survival probability were generated with the Kaplan-Meier method, which adjusts to the presence of right-censored, or unanswered questions [?]. From these estimates, survival curves were constructed to examine the survival experience of questions. Mean, median and other percentiles of survival times were also generated.

As the probability distribution for response times is unknown, a nonparametric Cox proportional hazards model was developed to predict the survival probability of questions [?]. To build the model, five-fold cross-validation was used [?]. Univariate analysis, performed on one training set, was used to identify variables to include in the final predictive Cox model [?]. Each predictor was entered into univariate Cox models and strength of association with response times was assessed. Those with partial likelihood ratio test p-values of less than 0.001 were included in the final model. Continuous predictors were entered into univariate Cox models both with and without square root and log transformations to determine the form of the predictor to include in the final model. Univariate analysis was again performed on each continuous predictor to investigate the use of splines. Each predictor, with the transformation found necessary in the previous analysis, was fit to three separate Cox models with restricted cubic splines of three, four, and five knots. AIC statistics were used to determine whether or not to include splines, as well as the optimum number of knots [?]. All predictors found to be significant, along with transformations and splines found to be necessary, were included in the model for cross-validation. No variable selection was used after identifying significant predictors in univariate analysis, as [?] determined that reducing the number of predictors in the model decreased predictive accuracy more so than including all predictors.

In each fold of cross-validation, the full model was built on the training set and used to generate predicted hazard ratios on the corresponding test set. To assess prediction performance, predicted hazard ratios were entered into separate Cox models as the single quantitative predictor with response times as the survival time. The resulting Nagelkerke's R-square statistic, concordance statis-

tic, Somers' Dxy , partial likelihood ratio and p-value, were assessed as indicators of the model's performance [?]. Significant results from this model would indicate high predictive accuracy. The partial likelihood ratio assessed is a ratio of the log partial likelihood function evaluated at the parameter estimates, equivalent to a goodness of fit measure of the model with predictors, and the log partial likelihood function for the null model without predictors and only the baseline hazard function, equivalent to a goodness of fit measure for the null model. The p-value for this statistic was calculated from the χ^2 distribution [?]. Concordance statistics and Somers' Dxy are a measure of the model's discriminative ability. Concordance is defined as the probability that for any two randomly chosen questions, the question with the shorter response time also has the higher predicted hazard ratio. Concordance statistics close to 1 indicate high discriminative ability, while statistics close to 0.5 indicate discordance, or random predictions. Somers' Dxy is the difference between the model's concordance statistic and discordance, 0.5. A Somers' Dxy statistic of 0 indicates random predictions, while a statistic equal to 1 indicates perfect predictions [?]. These metrics were computed for each iteration on every training and test set. Averages of metrics across training and test sets were evaluated. The final model was then fit to the full data and the same metrics were computed and compared.

Correlations between scaled Schoenfeld residuals, differences between observed and expected predictor values for questions that received answers, and a function of time, were examined to assess the proportional hazards assumption that the effect of predictors on hazard does not depend on time [?]. Significant correlations would indicate that a predictor has violated this assumption.

To obtain predicted survival probabilities using the model, the *predictSurvProb* function was used from the *pec* package. This function uses the Cox model and a vector of times to compute predicted survival probabilities.

5. Results and Discussion

Of 8,025 questions in the full data set, 7,760 were in English (97% of the full data). Of those questions, 4,951 (63.8%) received an answer by the download date. The shortest response time was 0.5 hours. The longest was 2,159.02 hours (89.96 days). Figure 1 shows the distribution of response times for all questions analyzed.

Figure 2 shows the Kaplan-Meier estimated survival probability curve for all questions in the data. The curve indicates that if a question does not receive an answer within the first 100 hours after it has been posted, the likelihood of it receiving an answer in the future is low. The Kaplan-Meier estimated mean survival time, or the average time until a question received its first answer was 775.75 hours, or 32.32 days. The median survival time, the time at which 50% of the questions in the data received an answer, was 9.16 hours. Table 3 provides additional percentiles of survival time.

Each training set contained 6,208 questions, and each test set contained 1,552

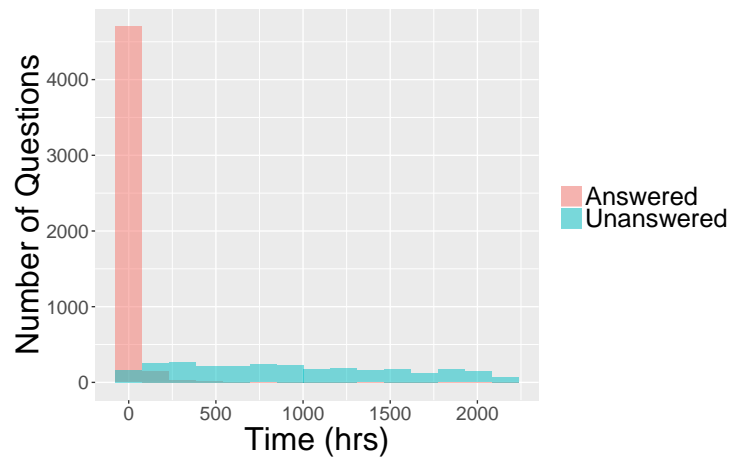


Figure 1: Distribution of response times

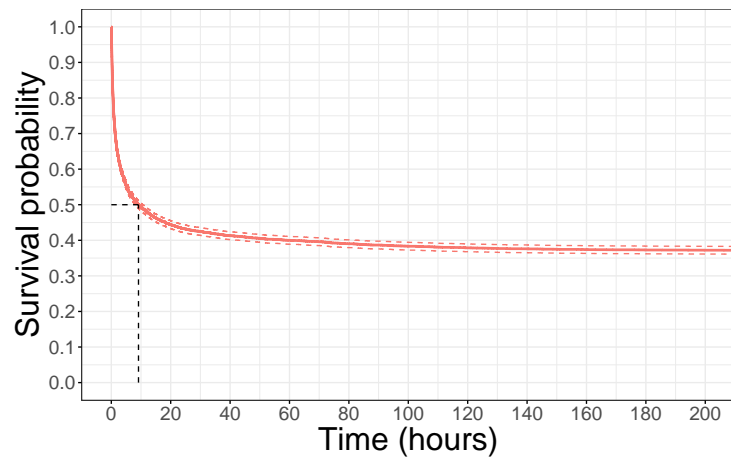


Figure 2: Kaplan-Meier curve for all questions

| Percent Answered | Time (hrs) |
|------------------|------------|
| 25 | 0.88 |
| 50 | 9.16 |
| 55 | 18.38 |
| 58 | 33.71 |
| 60 | 59.47 |
| 64 | 683.19 |

Table 3: Kaplan-Meier estimated quantiles of survival time

questions. Results of univariate analysis performed on one training set are shown in Table 4. For continuous predictors that contained zeros, one was added to each observation before applying log transformations. All categorical predictors and all continuous predictors marked with an asterisk (*) were entered into the full model.

Table 5 shows the results of determining the optimal number of knots for each continuous predictor. Values of knots marked with an asterisk (*) were included in the final model.

Average performance metrics for test and training sets and in Table 6. Partial likelihood ratio statistics and p-values indicate that the model as a whole is significantly associated with response time. However, its R^2 statistic and discrimination indexes are considerably low. Metrics did not change significantly from training to test sets, indicating that the model was not overfit. Metrics for the final model’s performance on the full data, also found in Table 6, are consistent with metrics found in cross-validation and indicate low predictive accuracy.

Assessing the proportional hazards assumption indicated that several levels of the device categorization variable, specifically Apple Product, Camera, Game Console, and Other, violated the assumption.

Interpretations of the coefficients highlighted in Table 7:

- The estimated hazard of receiving an answer is 154.6% (95% CI (131.9%, 179.5%)) higher for questions pertaining to Apple products than the hazard for questions about Android and Other phones, controlling for all other predictors.
- The estimated hazard of receiving an answer for questions posted on the weekend is 12.5% (95% CI (6.6%, 18.0%)) lower than the hazard for questions posted on a weekday, controlling for all other predictors.
- The estimated hazard of receiving an answer is 24.5% (95% CI (19.3%, 29.3%)) lower for questions with titles that contain at least one word considered to be frequently-used among unanswered questions, than the hazard for questions with titles that do not, controlling for all other predictors.

| Continuous Predictors | Transformation | p-value |
|---|----------------|--------------|
| Average frequency score of a question's tags | Untransformed | 1.270620e-08 |
| | Log + 1 | 1.186521e-08 |
| | *Square root | 1.464184e-11 |
| Average number of characters in each question's tags | Untransformed | 2.611529e-02 |
| | *Log + 1 | 6.018542e-05 |
| | Square root | 2.999917e-04 |
| Number of characters in a question's text | Untransformed | 6.947221e-12 |
| | *Log | 0.000000e+00 |
| | Square root | 1.110223e-16 |
| Number of characters in the user-defined device title | Untransformed | 6.530431e-03 |
| | *Log + 1 | 1.912977e-05 |
| | Square root | 2.291424e-04 |
| Ratio of the number of line breaks to the number of characters in a question's text | Untransformed | 1.398819e-07 |
| | Log + 1 | 1.092952e-07 |
| | *Square root | 1.148792e-11 |
| Categorical Predictors | | p-value |
| *Device category | | 0.000000e+00 |
| *Weekend or Weekday | | 4.123729e-04 |
| *Whether or not the question's text contained at least one end punctuation mark | | 2.535394e-09 |
| *Whether or not the question's text is in all lower case | | 1.540030e-07 |
| *Whether or not the question's title contains terms considered to be frequently used among answered questions | | 7.639738e-05 |
| *Whether or not the question's title contains terms considered to be frequently used among unanswered questions | | 0.000000e+00 |
| *Whether or not the question's title ended in a questionmark | | 1.142308e-12 |
| *Whether or not the user edited or added information to the question's text sometime after posting it | | 2.547222e-07 |
| *Whether or not the user had been a member for less than one day before the question was posted | | 0.000000e+00 |
| *Whether or not the user made an effort to solve the problem on their own, prior to asking the question | | 3.164815e-05 |

Table 4: Univariate analysis results

| Predictor | K | AIC |
|---|----|----------|
| Log transformation of the number of characters in a question's text | 0 | 65862.83 |
| | *5 | 65862.02 |
| | 4 | 65863.35 |
| | 3 | 65862.08 |
| Square root transformation of the ratio of the number of line breaks to the number of characters in a question's text | 0 | 65890.28 |
| | 5 | 65884.70 |
| | 4 | 65882.98 |
| | *3 | 65881.93 |
| Square root transformation of the average frequency "score" of a question's tags | *0 | 65890.75 |
| | 5 | 65891.69 |
| | 4 | 65891.75 |
| | 3 | 65892.35 |
| Log transformation of the number of characters in the user-defined device title + 1 | 0 | 65918.06 |
| | *5 | 65847.34 |
| | 4 | 65880.89 |
| | 3 | 65898.00 |
| Log transformation of the average number of characters in a question's tags + 1 | 0 | 65920.24 |
| | 5 | 65911.20 |
| | *4 | 65910.81 |
| | 3 | 65914.98 |

Table 5: Determining the optimal k number of splines for each predictor

| | HR | LR | p-value | R^2 | Dxy | Concordance |
|----------------------|----------|----------|---------|-----------|-----------|-------------|
| Training Sets | 2.027030 | 937.3877 | 0.0000 | 0.1401511 | 0.2683734 | 0.6341867 |
| Test Sets | 1.986844 | 220.8309 | 0.0000 | 0.1400518 | 0.2582790 | 0.6291395 |
| Full Data | 2.028512 | 1165.441 | 0.0000 | 0.1394545 | 0.2675286 | 0.6337643 |

Table 6: Performance Metrics

| Variable | | Coefficient (SE) | 95% CI | p-value |
|--|---------------|------------------|--------|--------------|
| Device Category | Apple Product | 0.9346 (0.0476) | | 0.000000e+00 |
| | Camera | -0.3199 (0.0476) | | |
| | Electronics | -0.0138 (0.0904) | | |
| | Game Console | 0.2458 (0.0784) | | |
| | Home | 0.3368 (0.0697) | | |
| | Other | -0.1304 (0.0555) | | |
| | PC | 0.2818 (0.0597) | | |
| | Tablet | -0.1566 (0.0815) | | |
| | Vehicle | 0.4056 (0.0695) | | |
| Weekend or weekday | | -0.1334 (0.0331) | | 5.525114e-05 |
| Whether or not the question's text contained at least one end punctuation mark | | 0.0247 (0.0499) | | 6.204270e-01 |
| Whether or not the question's text is in all lower case | | -0.1770 (0.0645) | | 6.065638e-03 |
| Whether or not the question's title contains terms considered to be frequently used among answered questions | | 0.0475 (0.0422) | | 2.604130e-01 |
| Whether or not the question's title contains terms considered to be frequently used among unanswered questions | | -0.2808 (0.0340) | | 1.110223e-16 |
| Whether or not the question's title ended in a questionmark | | 0.2571 (0.0332) | | 9.880985e-15 |
| Whether or not the user edited or added information to the question's text sometime after posting it | | 0.2983 (0.0857) | | 5.027830e-04 |
| Whether or not the user had been a member for less than one day before the question was posted | | -0.1069 (0.0364) | | 3.300423e-03 |
| Whether or not the user made an effort to solve the problem on their own, prior to asking the question | | -0.0724 (0.0362) | | 4.534803e-02 |
| Square root of the average frequency score of a question's tags | | 2.2609 (0.7196) | | 1.679666e-03 |
| Square root of the average tag length | | -0.0814 (0.0309) | | |
| | | 0.5945 (0.2661) | | |
| | | -1.5879 (0.7231) | | |
| Log of text length | | -0.0619 (0.1856) | | |
| | | -0.5052 (0.8452) | | |
| | | 2.9957 (4.1706) | | |
| | | -4.4755 (6.8975) | | |
| Log of device title length | | -0.0488 (0.0714) | | |
| | | -0.4348 (0.9724) | | |
| | | 1.2208 (2.3149) | | |
| | | -1.3199 (2.3255) | | |
| Square root of the ratio of the number of line breaks to text length | | -0.0076 (0.4776) | | |
| | | 0.4345 (0.7265) | | |

Table 7: Coefficients for predictors in the final model

| | |
|-------------------------|----------|
| LR | 1266.073 |
| p-value | 0 |
| R^2 | 0.151 |
| Dxy | 0.268 |

Table 8: Final model statistics

See Table 7 for the full list of model coefficients. Table 8 displays statistics for the final model.

Results and coefficients of the final model reveal how users can ask a question in order to decrease response time. Findings indicate that users should phrase the title in the form of a question, use correct grammar (e.g. punctuation and capitalization), include specific and concise tags, and post the question on a weekday.

Regarding the model’s weak predictive accuracy, it may be explained by limitations in the data. Many users asking a question incorrectly specified the device name (e.g. “Turtle Beach Ear Force Xmy grandson chewed through the wire while he was playing it’s brand-new is there anyway I can have it fixed0 One”), or did not include a device name at all. Many users also incorrectly used the tagging system by including ambiguous and lengthy tags like “someone sat on it :(” or “help me please!!!!” (tags are generally a couple key words that allow the answering community to quickly discern the question’s topic). It is likely that these inconsistencies contributed to the final model’s low predictive power. This, along with the results of the final model, reveal some of the ways that the CQA can be structured to improve analysis and decrease response time. As findings indicate that questions with correctly defined tags and device names may lead to quicker response times, the CQA can provide a stricter framework for asking questions. Allowing users to enter any device name or tag leaves room for user error. Instead, the CQA can restrict the tags or devices that users can include to a drop-down list. The CQA can also include more tips to guide users asking questions. Establishing and implementing a more restrictive framework for asking questions can help users ask understandable and clear questions and in turn decrease response time, as well as create a set of consistent questions for improved analysis.

6. Conclusion

This study developed a Cox proportional hazards model to predict the probability that a question posted on iFixit’s *Answers* forum receives an answer before a certain time. Predictors found to be significant included: device category, if the question was posted on a weekend or a weekday, whether or not the question’s text contained at least one word considered frequently-used among unanswered questions, and whether or not the question’s title ends in a questionmark. While overall the model was significant, its predictive performance was considerably low.

7. Acknowledgement

This research was supported by the Bill and Linda Frost fund of the California Polytechnic State University of San Luis Obispo. Special thanks goes to iFixit for providing access to the CQA data and for the help and assistance in its analysis.

8. Appendix

8.1 Posted questions

| Answered | Device | Title | Date Posted |
|----------|----------------|---|-------------------------|
| Yes | iPhone 6 | iPhone water damage, touch screen issue | 2017-05-10 21:50:48 GMT |
| | New User | Text | Date of First Answer |
| | Yes | So I dropped my iPhone in water 4 days ago. Have done the whole rice thing and seen huge difference in it. However, only one side of my screen works and it is the one side which I need to unlock the phone. What would be the best way forward? | 2017-05-10 23:12:27 GMT |
| | | | Response Time (hr) |
| | | | 1.361 |
| Answered | Device | Title | Date Posted |
| No | android tablet | ccccaaaan you help me fix my touch screen | 2017-05-23 07:35:12 GMT |
| | New User | Text | Date of First Answer |
| | No | touch screen not worki | NA |
| | | | Response Time (hr) |
| | | | NA |

Table 9: Answered and unanswered questions

Table 9 provides an example of an answered and unanswered question. All fields in the table are shown as the users entered it.

8.2 Variable Derivation

8.2.1 Device Categorization

Original device categories defined by iFixit included: Apparel, Appliance, Camera, Car and Truck, Computer Hardware, Electronics, Game Console, Household, Mac, Media Player, PC, Phone, Skills, Tablet, Vehicle. Device titles were

parsed and certain categories were combined or separated to create a new device categorization variable to use in the Cox model. Final categories created included: Android/Other Phone, Apple Product, Camera, Electronics, Game Console, Home, Other, PC, Tablet, Vehicle. Under the original device categorization, 1,954 questions (25.2% of 7,760) were categorized incorrectly and indicated an NA for its category. Missing values were a result of users creating questions for devices not already in the iFixit’s database, or from the user incorrectly defining the device name. For questions with missing categories, key words were searched for in device titles to recategorize accordingly. All other questions with ambiguous device titles were categorized as Other.

8.2.2 Whether or not the question’s text contained at least one end punctuation mark

This variable was created to investigate if run-on sentences, sentences with no end punctuation, take longer to receive an answer.

8.2.3 Whether or not the question’s title contains terms considered to be frequently-used among answered/unanswered questions

These variables were created to investigate if questions pertaining to popular topics receive answers faster than questions that do not. Similarly, another variable was created to investigate if questions pertaining to unpopular topics receive answers slower than questions that do not. This variable was created based off of methods from [?] and [?].

To create the variables, the data was separated between answered and unanswered questions. For the data frame containing answered questions, text mining techniques were used to create a list of every word within each questions’ titles and the frequency, or proportion of times those words occurred among all answered questions’ titles. The same was performed for the data frame containing unanswered questions. “Frequently-used” words in answered questions were defined as those that appeared in more than 1% of all answered questions’ titles, and appeared in more answered questions than in unanswered questions. To determine the latter, a ratio of frequencies, or proportions of times a word occurs, was assessed. The ratio was calculated as the proportion of time a word occurred among answered questions, to the proportion of times a word occurred in unanswered questions. As an example, if “cracked” appeared in 2% of answered questions and 0.1% of unanswered questions, it would be considered “frequently-used” among answered questions as it occurs in more than 1% of answered questions’ titles and occurs 20 times more in answered question than in unanswered questions ($\text{ratio} = 0.02/0.001 = 20$). Similarly, “frequently-used” words in unanswered questions must occur in 1% or more of all unanswered question’s titles and occur in more unanswered questions than answered questions. As there was some overlap between “frequently-used” words in each list and the device categories, every word that matched a device name was removed from the lists. The resulting list for answered questions contained 111 words.

The list for unanswered questions contained 32 words. Lists can be found in Table 10

| | |
|---|---|
| Frequently-used terms among answered questions | screen, turn, working, replacement, power, work, replace, charging, charge, button, touch, black, turning, broken, start, stuck, new, lcd, upgrade, problem, change, port, replaced, card, open, boot, replacing, remove, reset, back, drive, error, cable, ssd, cracked, hard, one, dropped, logic, lines, white, keeps, pro, dead, now, front, damage, switch, parts, glass, still, charger, issue, sim, turns, digitizer, just, mode, model, backlight, usb, stopped, logo, starting, know, unresponsive, password, factory, call, use, damaged, find, sensor, possible, fixed, side, galaxy, data, ipod, problems, issue, slow, system, connector, without, overheating, code, ram, air, microphone, please, red, much, plugged, getting, booting, left, way, buy, plus, time, loose, lock, coming, got, shuts, says, install, key, door, top |
| Frequently-used terms among unanswered questions | sound, light, wifi, speaker, connect, picture, stay, noise, bluetooth, isnt, apps, going, rear, question, play, stop, service, take, hear, lights, showing, network, volume, come, keep, connection, flashing, shut, print, blue, buttons, edit |

Table 10: Frequently-used terms among answered and unanswered questions' titles, ordered by decreasing frequency

8.2.4 Whether or not the question's title ended in a question mark

This variable was created to investigate if questions with titles in the form of questions receive answers faster than those that do not Bhat2014.

8.2.5 Whether or not the user made an effort to solve the problem on their own, prior to asking the question

Logical variable indicating true if the asker included words in the question's text that indicate that prior effort was made to find the solution to the question before posting. Words searched for include: tried, searched, researched, tested, replaced, used, checked, investigated, considered, measured, attempted, inspected, fitted. This variable was created based off of findings and methods from [?] [?].

8.2.6 Average frequency “score” of a question’s tags

This variable was created to investigate if including popular or widely-used tags in a question lead to faster response times. This variable is the average frequency, or proportion of times a question’s tags appear in all of the data set. If a question has a higher average, than at least one of it’s tags are frequently used. This variable was created based off of findings from [?].

8.2.7 Average number of characters in a question’s tags

This variable was created to capture when a user correctly or incorrectly used the tagging system. For example, question 390989 included the tag tag: “i need to repair the headset because i can not find the bluetooth”, which is 64 characters. Question 410254 includes tags “sound”, “sound driver” and “speaker” and has an average of 8 characters per tag. It is hypothesized that questions with correctly defined tags, as with the latter example, have faster response times.

8.2.8 Number of characters in the user-defined device title

This variable was included to capture when a user incorrectly defines the device name. For example, the device variable for question 390271 is, “Turtle Beach Ear Force Xmy grandson chewed through the wire while we was playing it’s brand-new is there anyway I can have it fixed”, and is 136 characters long. 11 questions in the data also did not include any device title. This variable was created to investigate if questions with incorrectly defined device names make it difficult for the answering community to discern the question’s topic, and thus have longer response times or remain unanswered.

8.2.9 Ratio of the number of line breaks to the number of characters in a question’s text

This variable was based on the hypothesis that questions that include line breaks in the text are easier to read than questions that don’t include any, and thus have shorter response times.