Questrom School of Business
MSBA Capstone Project
Spring 2022

OneQuesh: How To generate Good Questions & Attract Public Attention

Team B3
Members:
Hanyu Chen
Yipeng Guo
Luke Hong
Ting-Ann Lu
Yuyan Ma

Table of Contents

Executive Summary

Business Analytics, known as a set of disciplines and technologies for solving business problems using data driven analysis, statistical models, and other quantitative methods, involves an iterative, methodical exploration of an organization's data, with an emphasis on statistical analysis, to drive decision-making. The capstone project is an essential practical application for candidates majoring in this field, allowing students to utilize their skills to the real business world.

OneQuesh is a social media app created by a start-up company that differentiates itself from the standard, mainstream social media apps in the market today. OneQuesh does not encourage people to post selfies and stories or like other people's photos. In fact, they encourage people to answer one question per day. The questions can include a variety of topics including personal interests, dreams, memories, activities, etc. Users can rate and comment on questions, reflecting varied interest and popularity for each question. As an analytical team, we proposed and aimed to find 'good' questions for OneQuesh to increase engagement.

'Good' questions can be defined in multiple ways. Our hypothesis is that they have more responses containing more characters and more positive responses, they can generate more attention for OneQuesh to make it a more successful company. Moreover, by analyzing the rate of new users joining the app, we can correlate the increase in users with advertising events with the growing user base.

In collaboration with the business advisors at OneQuesh, we have a couple of detailed datasets containing question sets and users' information sets. The question set contains every day's questions, responses, ratings towards questions and some specific comments regarding users' feeling towards the questions they have been asked. For the users' information set, we have status, join-in date, date of birth, and whether the profile is complete. With these datasets, we can analyze and find insights and trends to create 'good' questions.

OneQuesh's business format varies from other social media companies like Twitter or Facebook. Their data are all question related, making the textual data one of the most important components. So, we started utilizing text analysis skills, including sentiment analysis to check users' emotion, character count to check whether people tend to talk more (this means the question can generate interest). Then we conducted clustering analysis to label those questions and find the most popular target group. With these results, we can find similar characteristics of 'good' questions. We also checked with OneQuesh's business advisors, using our list of curated questions, asking which ones that align with their business goals and potentially be used for practical use. This made us validate our hypothesis towards 'good' questions and adjust our methodology of filtering said questions.

The OneQuesh team preferred questions with the following features - thought provoking, sharing common memories / nostalgia, non-triggering (or neutral), and personal preference questions. People may feel offended when being asked questions worded in a certain way that may be triggering. However, they always share commonalities when sharing something about a day that everybody can relate to. Meanwhile, those creative questions that make people think about their life goals and dreams also have higher popularity. For example, "what will you choose if you must abandon one among love and friendship?" People love to talk about themselves, once the questions curate their desire for expression - and it is our job to find the features those questions have and generate ones like them.

1. Problem Statement

1.1 Company Background

OneQuesh is a social media app created by a start-up company and asks probing questions resulting in real human interactions. The platform differs from traditional social media apps (Facebook, Twitter) by asking one question per day for its users to answer. The questions include a variety of topics such as personal interests, dreams, activities, etc.

1.2 Proposal

The goal of our business analysis is to help OneQuesh increase user engagement by analyzing and creating effective questions and identifying marketing activities that could attract more new users.

Creating Engaging Questions:

We identified popular questions through exploratory data analysis and Natural Language Processing text analysis to help achieve our goal.

- First, we created 100 questions prior to receiving the dataset as an unbiased set of questions.
- Then, we met with the business advisors of OneQuesh to examine the 100 questions and to align with their company expectations and goals.
- Afterwards, we explored and analyzed what kind of questions gain **higher popularity** ratings using data analytics.
- Apply what we have learned from the questions reviews and use the algorithmic model we have created, to create the ultimate 100 questions.

Identifying effective marketing activities:

We identified effective marketing activities through finding the relationship between the frequency of the networking activities and the number of new users joined in the same period.

- Identified the period of the time that the newest users joined by doing exploratory data analysis of the new user's data.
- Matched with the marketing activities that have been done during that period of time.
- Also, we identified the majority of users' age groups to do better targeting marketing activities.

1.3 Hypothesis

We conducted our analysis based on the following:

- Hypothesis 1 - Popular questions have **more responses, more characters, and more positive responses** (higher sentiment score). These are so-called 'good' questions.
- Hypothesis 2 - By clustering questions using k-means or hierarchical clustering, we can profile questions into groups to find common traits of 'good' questions.
- Hypothesis 3 - We assume that marketing activities will coincide with more new users joining the app than usual.
- Hypothesis 4 - We assume the majority of the users are in the 20-30-year-old age group.
- Hypothesis 5 - We assume by analyzing 'good' questions, we can find trends and common characteristics to train a model to predict the performance of future questions.

2. Data Overview

In collaboration with the business advisors from OneQuesh, we received detailed datasets containing:

- Full OneQuesh Rating Report
    - In this dataset, we have every user id, questions, dates, responses, ratings, and response comments with 2443 rows and 6 columns in total.
        - Each row is the data for every response of each user, resulting in multiple rows per question to record different responses every day.
        - Five month of data that dates from 2021-09-19 to 2022-02-20.
        - Question rating data that represent users' evaluations of the question.
        - 76 additional user comments to share their thoughts, opinions, and feelings about the questions.
- Users' information:
    - In this dataset, we have: User status, join-in date, date of birth, whether the profile is complete, and whether the users are active or not.

Feature engineering:

We created additional variables and data to help us with our analysis detailed below:

- Sentiment Score Data
    - This dataset has questions, dates, total responses, average ratings, and sentiment scores for each row with 98 rows and 5 columns in total.
- Character Counts Data
    - This dataset has questions, dates, total responses, average ratings, and character counts for each row with 98 rows and 5 columns in total.
- Question Types Data
    - This dataset has questions, dates, total responses, average ratings, and question types for each row with 98 rows and 5 columns in total.
    - We manually assigned the question types provided by OneQuesh under three types of questions: This or that, What, and Hypothetical.
- Merged Data
    - This dataset has questions, dates, total responses, average ratings, sentiment scores, character count, and clustering groups with 98 rows and 6 columns in total.

3. Methodology

Based on the data we received and the subset data we created, we explored what creates an effective question. The methods that we used in the analysis are listed as follows. One specific thing to mention is that because OneQuesh's business format is quite different and the datasets are all question related, the methodology we conducted are mostly text related.

3.1 100 Questions

We generated 100 new unbiased questions to have a test dataset to see the performance of our models.

3.2 Sentiment Analysis

To evaluate the emotional tone of a response, we calculated the average sentiment score of all the answers for each question. The sentiment score is represented on a scale of -1 to 1. The low end of the scale indicates negative responses and the high end of the scale indicates positive responses with zero being neutral.
In addition, we also calculated the sentiment scores for the questions to analyze whether the questions are neutral or not since we believe that questions should be as objective and neutral as possible. By creating sentiment scores to quantify the emotional tone of texts, we can explore the relationship between average rating. We also compared the sentiment scores of the questions that OneQuesh already asked with the 100 questions that we created before.

3.3 Character Count Analysis

Users' engagement is crucial for a social media platform, and character count of a question can be an important factor to determine 'good' questions. Using the average character length of a question or response, we explored if there is a higher response rate.

3.4 Clustering (H-Cluster and K-Means)

After we generated columns for the sentiment score, and character count, we clustered the questions based on those parameters. Using a hierarchical clustering model, we categorized the question pool into 4 groups: a great performing one, two moderate performing ones, and an underperforming one. Then, we found the common features of what separates a great performing question from an underperforming question. By clustering the questions, we can get a profile of the types of questions OneQuesh deploys in their app and their numerical features to assess the performance of ratings and engagement.

3.5 Analysis in Comments

By analyzing the 76 specific comments by users on the questions, we further explored qualitatively what makes a question considered "good".

3.6 Predict Average Ratings Based on New 100 Questions

By using character counts, sentiment score of questions and sentiment score of responses as parameters, we trained our model and tested the accuracy by using our 100 questions.

However, our predictions were very inaccurate and our algorithm could not be used for practical use. Due to only having a few features as well as the lack of observations of data to train on, our idea for the prediction algorithm was scrapped.

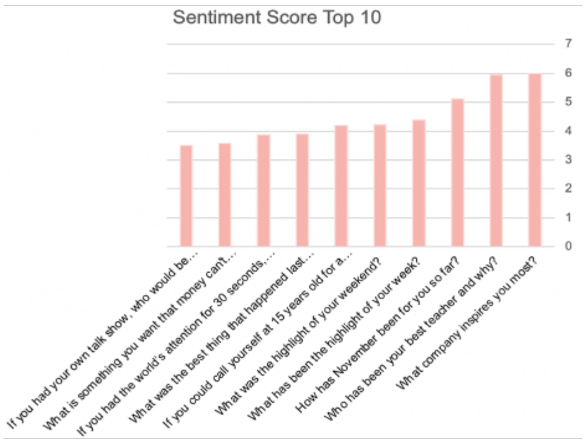
3.7 Using Feedback from OneQuesh to Improve Question Analysis

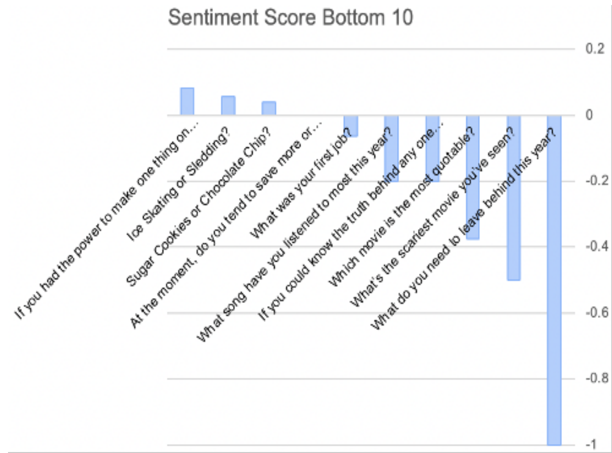
To get an insight of how OneQuesh evaluates questions, we received feedback about the 100 questions that we created. As a result, questions which contained the following qualitative characteristics below are considered a 'good' question:

- Thought provoking
- Nostalgic
- Non-triggering / neutral worded questions
- Personal preferences

4. Results

4.1 Sentiment Analysis

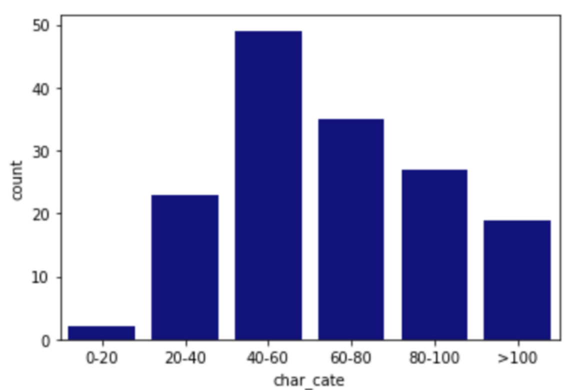


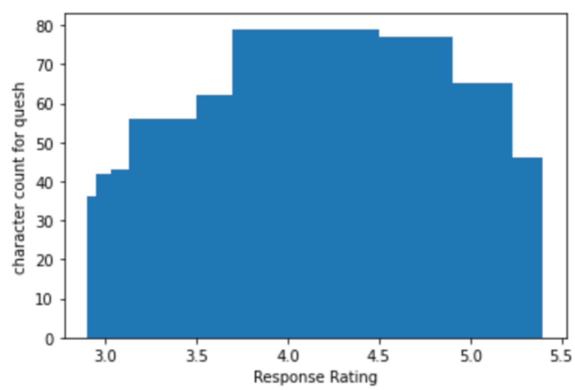


We calculated the average sentiment score for all the answers to each question and got a ranking to have some initial insights into the performance of the questions. We see that questions with higher scores of answers usually contained some positive words such as “inspire”, “best”, or “highlight”. On the contrary, the questions with lower scores contained negative words such as “leave”, “scariest”, or “spend”.

However, the negative sentiment score does not necessarily mean the question is not effective and there is still a potential for users to reply. Therefore, we still needed to move forward to the advanced analysis to investigate the relationships between the sentiment scores and the clustering model.

4.2 Character Count Analysis



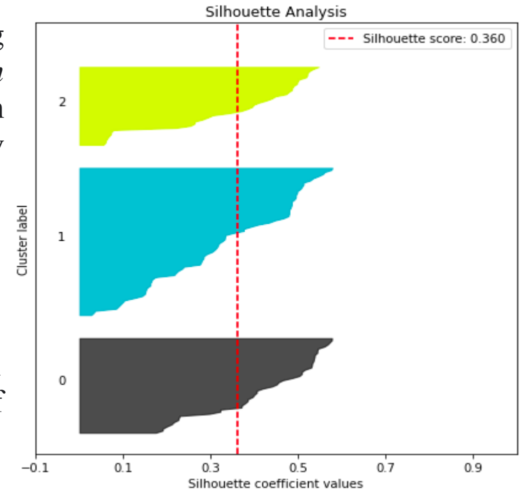


We calculated the average character count of the responses for every question. Then we got some distribution graphics regarding the results. Most of the response character counts were around 40-60 characters. Subscribers who answered questions tended to answer in full sentences instead of responding in a couple of words.

Most of the questions have around 70-80 characters and have a response rating of 4.0 to 4.5. The highest response rating has a character count around 4.

Also, popular questions tend to have higher character count and gain higher average ratings. Character count and average rating have positive relationships.

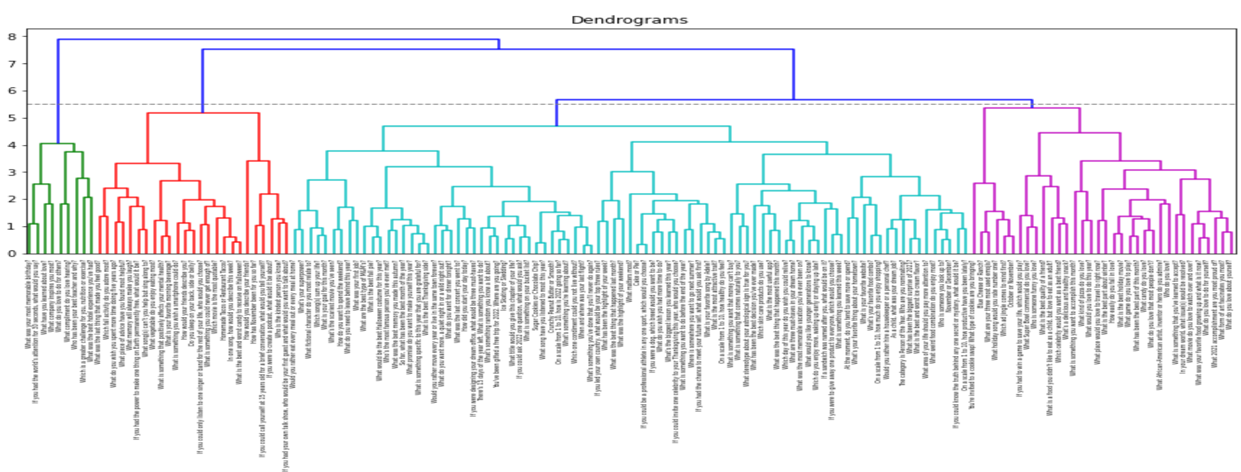
4.3 K-Means Clustering



K-Means clustering is an unsupervised machine learning method. It takes in numerical features and partitions *n* observations into *k* clusters in which each observation belongs to the cluster with the nearest mean calculated by squared Euclidean distance.
After we utilized K-Means clustering, we have 3 clusters as shown in the figure below. However, when explaining the results of how the questions were grouped, we don't have a great visualization of the process. K-means requires a pre-specified number of clusters and although 3 clusters has the highest prediction power, the number of clusters does not fit the purpose of the business problem well.

4.4 H-Clustering

H-Clustering (hierarchical clustering) is also referred to as agglomerative clustering and it is a "bottom-up" approach. It is an intuitive approach and lets us as analysts determine our cluster solution. We have 4 clusters based on this algorithm and it enables us to have a great vision of how the clusters have been done.
Specifically, we tried to conduct clustering analysis, aiming at finding the best way to segment those questions into different groups with different features. Since we have multiple features after our feature engineering process, we were able to dive deeper and check which group is the best fit for good questions. Again, a 'good' question has a higher total response, higher average rating, higher sentiment score and higher character count - characteristics of questions that proxy popularity. Judging from these two clustering methods, k-means does not achieve the accuracy requirement, so the measure of fit of this clustering algorithm is not practical for our business problem. That's why we decided to take H-clustering.

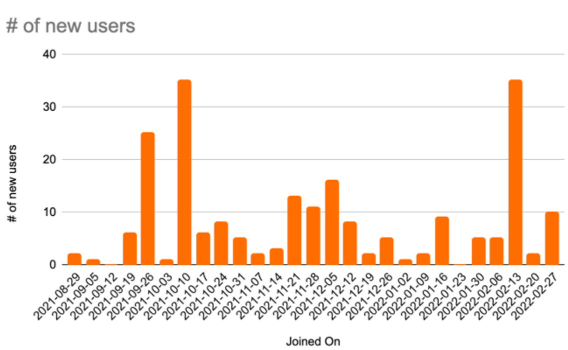


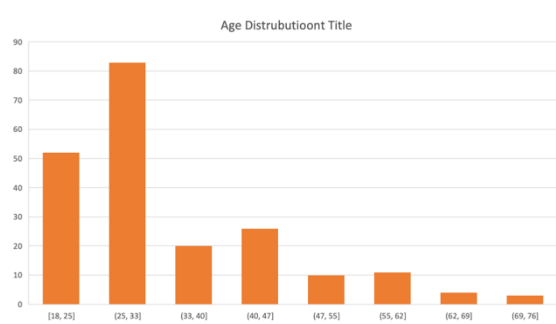
The corresponding results under each group for h-clustering is shown in the below figure. Cluster number 1 profiles as the best question group for our business problem.

Cluster	Avg_rating	Total respo	Sent_ans	Sent_quesh	Char_count
1	4.31	15.11	**5.37**	2.44	**139.53**
2	**4.60**	8.28	1.75	0.84	58.83
3	4.01	15.84	1.49	0.93	63.38
4	3.98	**21.21**	2.01	**3.06**	67.19

Cluster 1 has the second highest rating, second total response, highest sentiment score and character count. As for the average rating, there are also limitations. For instance, there is not a big sample size for those who rate the questions, leading to one's rating weighing more which makes the feature not as robust.
Cluster 2 has the highest rating, lowest total response; moderate character count and sentiment score.
Cluster 3 has just a moderate performance - it doesn't have some outstanding merits.
Cluster 4 has the lowest rating, highest total response; moderate character count and sentiment score.

4.5 Users' Data Findings





Looking at the timeline of users joining the app, we detected a peak in October 2021. Even though we don't have statistics on marketing activities, we wanted to look into what activity caused a surge in the number of new users. With collaboration with our business advisors, we had a discussion regarding our findings. As a result, OneQuesh engaged in a lot of networking activities throughout that month, which could explain the increase in users. The networking activities include in-person networking events, social media collaboration, and different types of virtual events. As a reference, OneQuesh's manager published an article in Boston university's alumni magazine on Feb 9th, 2022, which resulted in a significant increase in new sign-ups.
The findings support the hypothesis that marketing activities like networking can increase the number of new users.

As shown in the 2nd figure above, the bar graph represents the age distribution of the users, the majority of users are between the ages of 18 and 33. Our findings validated our previous hypothesis that the majority of users are between the ages of 20 and 30. By identifying our target age group and tailoring questions to this age group, an increase in engagement can possibly be achieved. OneQuesh may also network with more people in this age range, resulting in a considerable growth in the number of users.

4.6 Prediction Model

We built predictive models to predict the question type and what kind of questions have a higher popularity. For the first model, we utilized the NLP model to preprocess the text data and KNN model to predict the question type. From the left table below, we can see that our algorithm to predict the question types has 91% accuracy on average and is reliable enough. However, it is hard to implement the algorithm to new questions since we do not have the required features to input in the model, such as number of responses, sentiment scores, character counts, or average ratings.
As for our second model, it predicts average ratings for the 100 questions we created to analyze what kind of questions have a higher popularity. We utilized BERT as our base model and trained using the questions data provided by OneQuesh. However, the result turned out to be unreliable since we don't have enough data to make a convincing model. From the right table below, we can see that the majority of the questions have 4.4 average ratings and the minimum value is also higher 4.2, which is only 0.2 away from the maximum value. Hence, the model is not effective and useful to assess the popularity of questions based on a limited training dataset.

Question Type	precision
1	0.83
2	0.94
3	1
4	0.88
Avg accuracy	0.9125

	predict result
count	100
mean	4.38
std	0.054
min	4.2
25%	4.4
50%	4.4
75%	4.4
max	4.4

5. Summary and Recommendation

In summary, we used the existing question data and feature engineered variables (sentiment score, character counts) to group questions with hierarchical clustering. Cluster 1, the great performing one had the second most highest average rating, higher sentiment score in the answer, and the highest average character count per response. We recommend aiming to create questions that share similar characteristics to questions in this group.

As for the user data, we recommend targeting 18-33 and 33-47 age groups and tailor questions to them. We also recommend focusing on networking with these age groups to facilitate the growth of the user base.

6. Limitations and Challenges

Limitation in data gave us a challenge to make a predictive model. In our last phase, we tried to build a predictive model to predict what kind of questions would have a higher popularity. However, we were not able to obtain practical results since we don’t have enough data to make a convincing model.

Due to the privacy of the un-published questions from the question bank, we used our 100 questions to test the result of our model. If more questions were given, we can get more practical results.

7. Next Steps

Since we already have many insights towards the Quesh and the responses, we would be able to create an algorithmic model to predict those will-be popular questions when getting more data.

Additionally, we can add more features or metrics to evaluate the performance of questions. For instance, we can gather information about the users’ mood based on the emoji that they pick after answering questions, and combine it with the sentiment score of their response to make better predictions.

Further, we can add age as another factor to our analysis, and we can tailoring questions and see if this method gets a higher response rate.

Link to Github Repository

References

1. This BU Alumni Wants to Take The Toxicity Out of Social Media. Bostonia, Boston University's Alumni Magazine https://www.bu.edu/articles/2022/onequesh-taking-toxicity-out-of-social-media/

2. Social Media Analytics: The Complete Guide https://emplifi.io/resources/blog/social-media-analytics-the-complete-guide

3. How to Write a Market Analysis https://articles.bplans.com/how-to-write-a-market-analysis/

4. How to Start a Social Media Agency (Steps & Strategies) https://biz30.timedoctor.com/social-media-marketing-agency/

5. Emotional Triggers: What They Are & How to Understand Them? https://www.mindbodygreen.com/0-18348/what-are-emotional-triggers-why-you-need-to-understand-them.html