Increasing Engagement via Data-driven Content

**Organization Name:** OneQuesh

**Business Advisor**

Name: Thomas Andrews

Role: President

Contact: [thomas@onequesh.com](mailto:thomas@onequesh.com)

**Purpose of the project**

The main issue you want to address, the broad and possibly long-term results you want to achieve with this project.

How can we use data to create the best questions to increase retention and interaction on our social media platform?

**Objectives**

Any measurable, specific outcomes, marked as milestones if they can be used as such. You can also mark them as “critical”, “good to have”, “stretch/long-shot”

* 2022-2023 Question Strategy to Increase Engagement
* Conclusions About User Base and High Performing Questions
* Questions That Raise Engagement and Interest
* Data Driven Questions Per Business

**Significance**

Why is this project significant for your organization? (Monetary/non-monetary)

We’re looking to create content that will increase our user base and activity.

**Data**

* Brief description of the data needed
* Availability of the data (is it already acquired and accessible?)

**What Are We Gonna Do Before Friday?**

* Map the possible direction of where we're gonna go <- If ‘this’ gonna go? / Do you have this kind of data? / Are you down with providing xxx etc.

11/19 - Initial Introductory Capstone Meeting

* Private vs public users
  + No data form private users
* Need to sign NDA still so not much info to share
* Download the app to get it out
* Platforms
  + Instagram
  + Website - onequesh.com
  + App

12/02 - Preparation for Meeting

Ideas / Proposal

* Do NLP on questions and categorize questions for profiling
* Quantify engagement based on responses for each question
* Find external sources for interesting questions to increase engagement

Features of the app

* Mood tracker and daily journaling
* Question page
* Rate the questions after answering
* Add people’s answer to collection
* Log in streak
* Log in cumulative

Questions to ask

* How many downloads does the app have?
* How many subscribers?
  + 134 users
* What was the peak use of users / time spent?
* Where are the question sources?
  + How many new questions do you have in-store?
  + How many questions/days has the app been running for?
    - 85 questions
* 917 responses

12/03 - Post-NDA Capstone Meeting

**Goals/objectives of the meeting:** To propose potential ideas and learn more about OneQuesh’s data after signing the NDA

* Turn cameras on
* Pick 1 person representing the team
* Share your idea to Thomas
* Ask questions

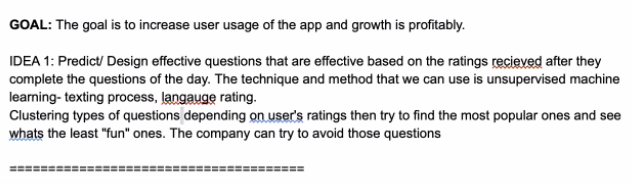
How to send emails

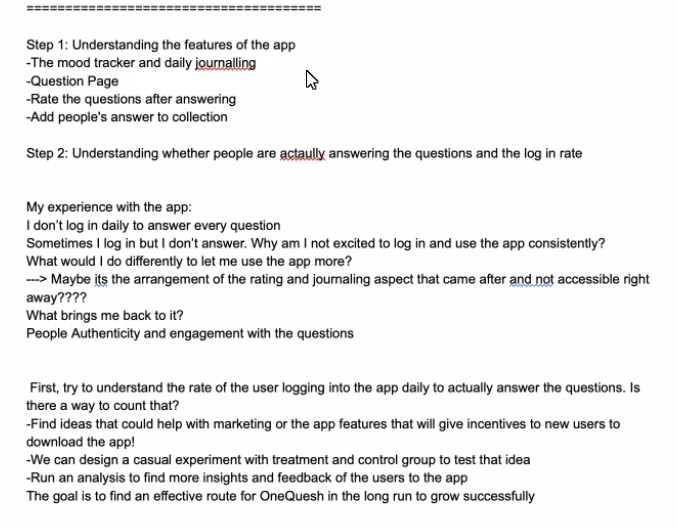
* Subject: OneQuesh - Team B3:

Have a rotating scribe

* Objectives/agenda of the meeting
* Action items after the meeting

Geech’s Pitch:





Alex’s Pitch:

* Predict the questions that are the most effective
  + Increase user engagement
    - Most responses

Data

* Accelerator program to grow scalability
  + Have ideas for app features
  + Equity free
  + Geared towards entrepreneurship
  + Impressed by their engagement rate
* Have back-end data
  + Questions
  + Responses
  + Sponsors
  + Avg Rating
  + Inspired By
  + Date
  + Quesh

Data Collection Timeline/Info

* Launched October 5th
* 13 downloads sept
* 52 downloads oct
* 27 downloads nov
* Article released from the daily press increased downloads yesterday
* Double digit responses / 3 rating is baseline
* Can request data for certain features from the appCreating another data point to gain more information
  + Email to Thomas and Alana and they will ask development

Previous team findings

* Previous questions categories
  + This or that
  + What is?
  + Hypothetical
  + Alana will send the 4th category in the email

Dec 3. 2021 Meeting

A3’s idea:

1. Goal: increase the downloading rate
2. What you see right now and how to improve

* Questions that people interested?
* User journey and user’s experience

1. Goal: Understanding whether people are answering the questions and the log in rate

* Text analysis

Quesh Listing: inspired by, Total Resp, average rating (per user) concern if under 3

Question by month (what questions popular with different month)

* The app Launched on Oct 5th

Dataset we can use:

* Post info daily
* User’s info
* Report listing
* Content listing
* Emoji for the calendar (development?) keep private

**Action Items:**

* Think of A/B testing ideas (Thomas is very interested in this)
* Think of how to do NLP on the Quesh and responses
* Fun exercise: Challenge of creating a month of questions

01/07 - Capstone Meeting with FA (Elgar)

Hypothesis

What are the agents in this domain?

What are the concepts?

* Questions
  + Sources
    - OneQuesh proprietary
    - Users
    - Business Partners
  + Sentiment analysis
* Users
  + User data
    - Ask what attributes of data is available
  + What users answer which questions?
  + **Can you anonymize the user data?**
* Responses
  + Sentiment analysis
    - Are responses consistent
    - What questions get responses from users
* Environment
  + App or Instagram

Midterm presentation (6 weeks ish)

* Have something to show
* “These are the 3 main datasets we are talking about”
* Auxiliary datasets
* Have EDA with visualizations

Effective Communication

* Send FA summary of what we need from BA for approval (By Wednesday)
* Then set up meeting with BA
  + Send an agenda
    - Minimum 3 points
  + Short description of the meeting
  + \*1 day before the meeting\*

**Action Items:**

* Send Elgar summary of what we need from BA by wednesday
* Reach out to BA
  + Email BA and get data sources
  + Set up meeting to initiate conservation of our ideas
* Analyze the data after we receive it
* Think of A/B testing ideas after analyzing data

01/10 - Internal Team Meeting

Question Categories:

* This or that
* What is?
* Hypothetical
* **????**

Assign Roles

* PMO: Caroline
* Scrum Master: Luke
* Scribe: Mona
* Meeting Organizer: Ann
* Time Keeper / Proofreader: Hanyu

**Deliverables:**

Summary for Meeting

* Discuss receiving the data from the dashboard and best practices
* Discuss potential additional data sources (user data anonymized)
* Confirm questions categories from previous teams
* Confirm which questions are from what sources
  + OneQuesh proprietary
  + Users
  + Business Partners

Agenda for BA Meeting

* Discuss data sources and best practices
* Discuss initial ideas and expectations

To do

* Send initial email to Elgar (FA) about what we need
* Confirm bi-weekly meeting time
  + Thursdays
    - Except this week after 5:30
    - Recurring 3:30-6:30
* Set up meeting with the BA and FA on zoom
  + Include agenda
  + Subject: OneQuesh - Team B3:

02/02 - Internal Team Meeting

Brainstorm

* EDA
  + Graphs and tables that tell the story of our data
  + Categorize questions
    - Supervised machine learning
      * Label by hand and categorize new questions
        + New questions

Get potential new questions from Thomas

* + - * + Or break data into test and training data

Low data points so not the best method

* + - Unsupervised machine learning
      * Clustering and seeing which questions we can find
* Text machine learning model for generating questions with best engagement

What we need from OneQuesh

* User data (anonymized)
* Response data for each question
* Latest OQ - Question Data dataset including January
* App download data by day or month
* Past questions before the app or question bank of examples
  + Proprietary questions may be confidential or sensitive information\*

Action items:

* Send email to Elgar with meeting agenda: Caroline
* EDA
  + Word cloud - Caroline
  + Distribution for avg\_rating & total response - Ann, Hanyu
  + Label questions by hand - Luke
  + Clustering using unsupervised machine learning - Mona ]
* Supervised machine learning after labeling
* Book room: Mona

Question Categories:

* This or that: 1
* What is?: 2
* Hypothetical: 3
* New categories
  + Ask thomas about 4th category
  + Reflective: 4??

02/03 - Internal Team Meeting Before Elgar

Meeting Agenda:

* Debrief our results so far
  + Data visualization & distribution analysis
  + Different clustering
* Introduce our next steps to implement
* Discuss possible dataset to gain & what can we do if have that

What we can do next :

* Sentimental score
* A/B testing ( Asking about how to do it)
* Question performance assessment.

02/03 - FA Meeting with Elgar

Action:

* **Send to Elgar**
  + Description of the project (Business proposal)
  + What kind of data we got
    - All the questions from Oct to Jan with number of total response and average score
  + Progress report — Checked
* **Data requested by Monday Feb 7** 
  + We are interesting in which which subject and we are required to gain….
  + Communicating with other team PMO
  + Data bank
* Generating 100 questions this week
* Meeting with Thomas Next week maybe Feb 10 Th
* Should send the progress report every week
* Set up requiring meeting every week - one time action
* Task assignment using project management tool (example: github)
  + Trello
* Refined the business proposal
* Read syllabus - suggestion

Note:

* Meeting with elgar every two weeks next should before Feb -17
* We have the latest data and we will get same data as other group
* We will receive the data that other team had
* We will see the evaluation report guideline for us soon
* For the progress report
  + Using the legend more for example ( require the data - red)
  + Initial initial analysis - more detail ( H clustering and word cloud )
* In the future send the agenda and progress report one day before
* If we do have coding issue we have someone can help
* midterm - march 14
* Final - May 2 2022
* Presentation should be ready one week before report due

02/10 - Internal Team Meeting

Tasks

* Generating 100 questions this week
  + 20 each
* Supervised machine learning
  + Expand the question dataset, label the question type(as variable) - Mona ✅
  + Word counts for each response <- average word counts for each question - Caro ✅ (file uploaded as cvs under data folder)
  + Do sentiment analysis of new responses data <- average sentimental score for each question - Ann & Hanyu ✅(already uploaded as csv file under data folder)
  + Do a new cluster - adding question type, avg word counts, avg sentimental score - Luke (maybe with k-means)
  + \*Assess what questions do well
  + \*Predict performance of question

Action Items:

* Do assigned tasks
* Meeting with BA?
  + Data request
    - Update the two datasets to latest
    - Question bank they have (to test our model)
  + Biweekly meeting with Prof.Elgar on next Thursday (Feb. 17th)
  + Two brief meetings with OneQuesh before Mar. 4thx

02/17 - Internal Team Meeting and FA

Future Ideas

* See the retention of users responding
* Do sentimental on questions

Presentation with BA

* 1 speaker
* It is ok to present only for 5 minutes
* Mention the importance of getting updated data consistently

Elgar’s feedback

* What are the
* Where come from
* More concise to tell the story
* Be very genetic - to make wanted impression
* Credit other teams
* Present to people who don't really understand
* For next step
  + Making hypothesis (eg, rating up, something up)
  + Asking about if we can get the question bank
  + Cool? With 100 questions? And really crazy too:))
  + What do u want

02/17 - External Team Meeting with BA

<https://www.bu.edu/articles/2022/onequesh-taking-toxicity-out-of-social-media/>

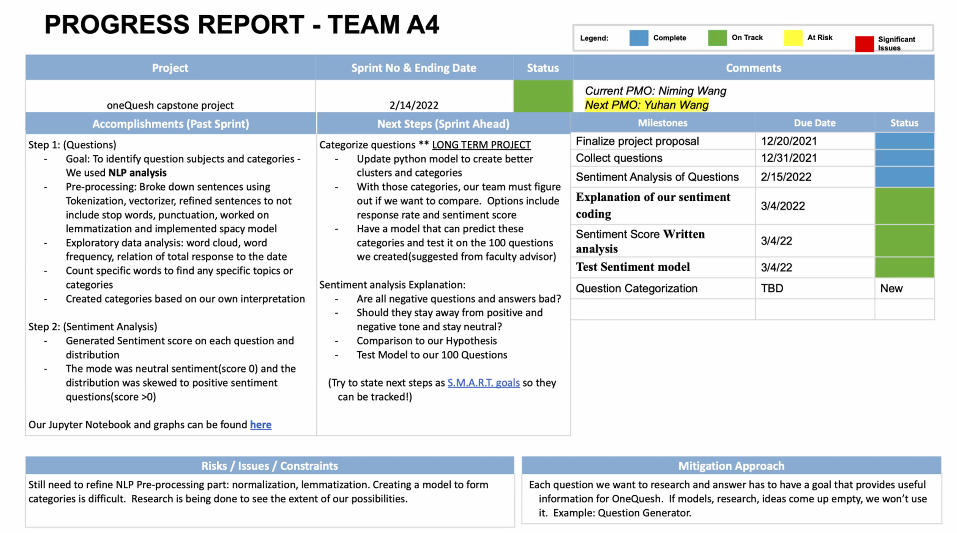
Thomas Andrews & Alana Ginnard

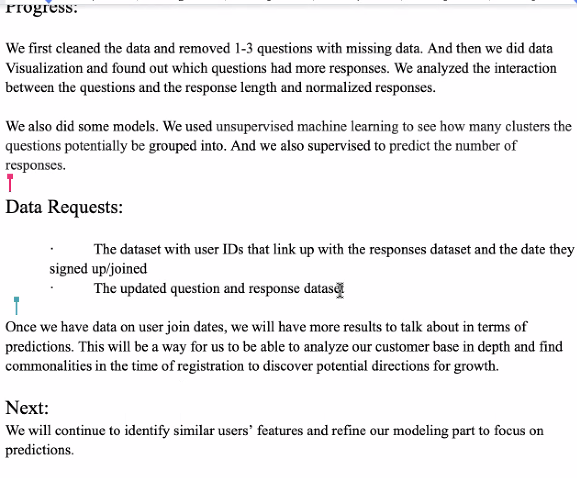
Data we can get:

* Question Bank
* Updated dataset of Questions and Response

Things to think about

* Whether users use personal account
* How to generate the proporate questions





Request we have,

* Cc alana with emails ([alana@OneQuesh.com](mailto:alana@OneQuesh.com))
* Give OneQuesh the 100 questions we generated
* If we have new analysis, send them over in the following 2 weeks

Afterwards reflect meeting with Elgar,

* Hypothesis testing of ourselves
* Give credits to other teams
* Think about other great ideas and try to find the relations of our project
* Repeat ourselves to remind our audience & answer some technic questions
* Give examples of the good questions we think

Action to take,

* Send Thomas the 100 questions (cc alana)
* Prof would push forward with the data request (by end of Feb) <- Thank you so much!
* Get this month’s updated data by today or tomorrow. (great! ✅

02/22 - Internal Team Meeting

What do we have for now:

* Target / Data
* Hypos
* Quesh type, sentiment score, character count <- update them into most recent
* H-Cluster & K-Means <- update
* Recommendations
  + Read through the comments against questions
  + Neutral questions
* Users’ Insights
  + EDA
  + Frequent users’ info <- insights

Actions to take:

* Email <- what is ‘inactive’ <- Caro ✅
* Label the question type + Total Response <- Caro Sep + Mona Jan + Ann Feb ✅
* Sentiment against answers (average <- Ann & Hanyu ✅
* Character count (average <- Caro ✅
* Average rating ✅& Cluster <- Luke
* Users stuff <- Mona
* Comments <- Mona
* Slides

02/28 - Internal Team Meeting 👏

Slides design (10 min)

* Introduction (background, proposal, data we have) - Caro ✅
* EDA (users’ info + response analysis & features towards responses) - Mona
* Sentiment score (same as following) - Ann ✅
* Character count (introduce that tool, explain the results) - Hanyu
* K-means & H-cluster (visualization graphics) - Luke
* Next steps (prediction model etc) - Luke

03/01 - External Team Meeting With Elgar

**Slides**

* List what we did and which are useful, which are not
* Shorten the length of our presentation
* Lovely graphics :) <- but does that colorful bars mean anything? We can strengthen the important points (+change the color maybe
* X & y labels (+ labels
* Line for continuous & bar for discrete (+ adjust some charts
* top 20 list (+ maybe delete the chart, only conclude those top20 in common
* Add a summary of the data findings
* List what did we try and what did we find
* List what will we try in the future

**Report**

* Grammar check ✅
* Structure is good ✅
* No fellow words ✅
* What is good questions are (we should state) ✅

<- ‘Good’ questions <- we still need more detailed clarification

* hypo <- making onequesh a more successful company / add a little bit explanation ✅

<- these 200 (for example) questions would be our suggestions for onequesh to post ✅

* Links to repositories
* Background <- what is express true self <- we can just express in an easy way
* Adding some summary after the analysis

**To do**

* Ask Thomas whether we could invite Midas to listen
* Generate meeting link
* Send the modified report & slides on late thurs or early fri

03/21 - Internal Team Meeting

* 100 questions - find the neutral ones (Ann & Hanyu)
  + The OneQuesh’s questions’ sentiment score (average
  + Calculate our questions’ sentiment & select ‘neutral’ ones
* Predict average rating (Luke)
* Users’ data (Mona)
* Work on the popular questions (Caro)

03/21 - External Team Meeting with Elgar

* Midterm
  + Maximum 5 slides for 15 mins presentation
  + Not expected the people are familiar with the data analysis
  + Only put the important slides
  + Conclusion: way too much information should connect it to short version
  + Another 100 questions ??????
  + Conversations with OneQuesh
    - Confirming hypothesis
    - How does Onequesh will use for our results
    - How do you clusfiled with questions
    - Who should we talk to if we want to validate our classification (expert?)
    - Does our clusters means anything to you
    - We want to do good to you

03/31 - External Team Meeting with Elgar

Requirements of Final

* Report: One page summary, what we learned, what is our findings, what are our demerits or what we don't have for what reason
* Presentation:
  + 20-25 minutes
  + dig deeper in the hypothesis,

<- Not everything, 2-3 major findings, 2-4 graphs and purely describe them

<- Introduce our project to everybody not familiar with us

* Posters: abt 14 posters hanging, attract ppls attention
  + More concise, more summarize
  + 3-5 graphs on there, easily readable
  + Do not include to0 much ‘technical words’ - be straight forward
* Preparations
  + Submit before May 1st, wait for recommendations
  + May 10th - presentation, before that to do the modification.
  + Counting backwards, pick up major findings, make reports more precise (point out what we are doing and what are our findings), well-organized, make your graph very clear (do not mess up with colors, label the axis etc.)
* Email To OneQuesh ✅
  + Questions - any feedback? Why do they like that?
  + Present popular questions - what are our definitions, any feedback based on our findings (selection policy)?
  + Grammar check for the email - ‘What else we can dig into and make contributions to’ / we would like to discuss (2-3 bullet points, including our technical finds, to be more precise) / list some interesting areas, leave them to pick (give abt 4-5, let them choose at least 2)
  + <- ‘We’ve picked out some popular questions, have you tested our questions and what are the differences or similarities, are the algorithms biased or not etc’
  + Double blinded study - ask for popular questions out of our questions (be very direct, and compare that with our popular questions)
  + False positive, false negative <- to confirm our hypothesis is good or not <- think of what is a good level of overlapping
  + If not overlap good, find out what are the gaps between ours and theirs, then make some suggestions

04/25 - Internal Meeting - Final Submission

To Do List

* Report
* Poster
* Presentation Slides

Report

* 6-10 pages
* + 1 page summary (for promotional purposes) ✅
* Outline
  + Company background <- leave it that way
  + Proposal <- extend
  + Hypothesis <- dig deeper (3 + 1) - **1 page - Mona ✅**
    - + new users - marketing activity
    - + new user groups
  + Data (detailed description) - **1 page - Ann ✅**
    - Columns’ description
    - Data size
    - etc.
  + Methodology - **1.5 pages - Hanyu ✅**
    - Brief description of every terminology
    - By utilizing it, what do we expect to get etc.
    - Using feedback from BA to improve question analysis
      * Thought provoking
      * Nostalgic / memories
      * Non-triggering / neutral worded questions
      * Mundane/everyday life
  + Results - **1.5 ~ 2 pages - Everybody**
    - **Questions: Sentiment / Charatercount / H-cluster & k-means**
    - **Users: xxx**
  + Conclusion - **half page - Caro**
  + **Recheck & Modification - Luke**

Presentation slides

* May 1st deadline
* 20 minutes
* 4 slides per person guideline

Poster <- May 11th

* Large poster sizing (36” x 48” horizontal)
* Need permission for using OneQuesh name or logo
* May 1st deadline

04/29 - Internal Meeting - Presentation / Poster

Poster / Presentation Breakdown / 3 slides per person

* Project Goal / Objective / Business Problem / Problem Statement -1 Luke
* EDA & Key Findings (3 ppl)
  + Sentiment & character count -1 Ann**✅**
  + Clusters -1 Caro
  + Users prospective - Mona 1
* Summary, Limitations & Challenges -hanyu
  + How hard that we would like to make a predictive model
  + Limited data
* Next Steps / Looking Ahead -1 Hanyu

**Reference of posters:**

Team 03, 06, **18**, 20

<https://drive.google.com/drive/folders/1y0xx-CHyUkB9cPcloNJZN6A6vaPC64sy?usp=sharing>

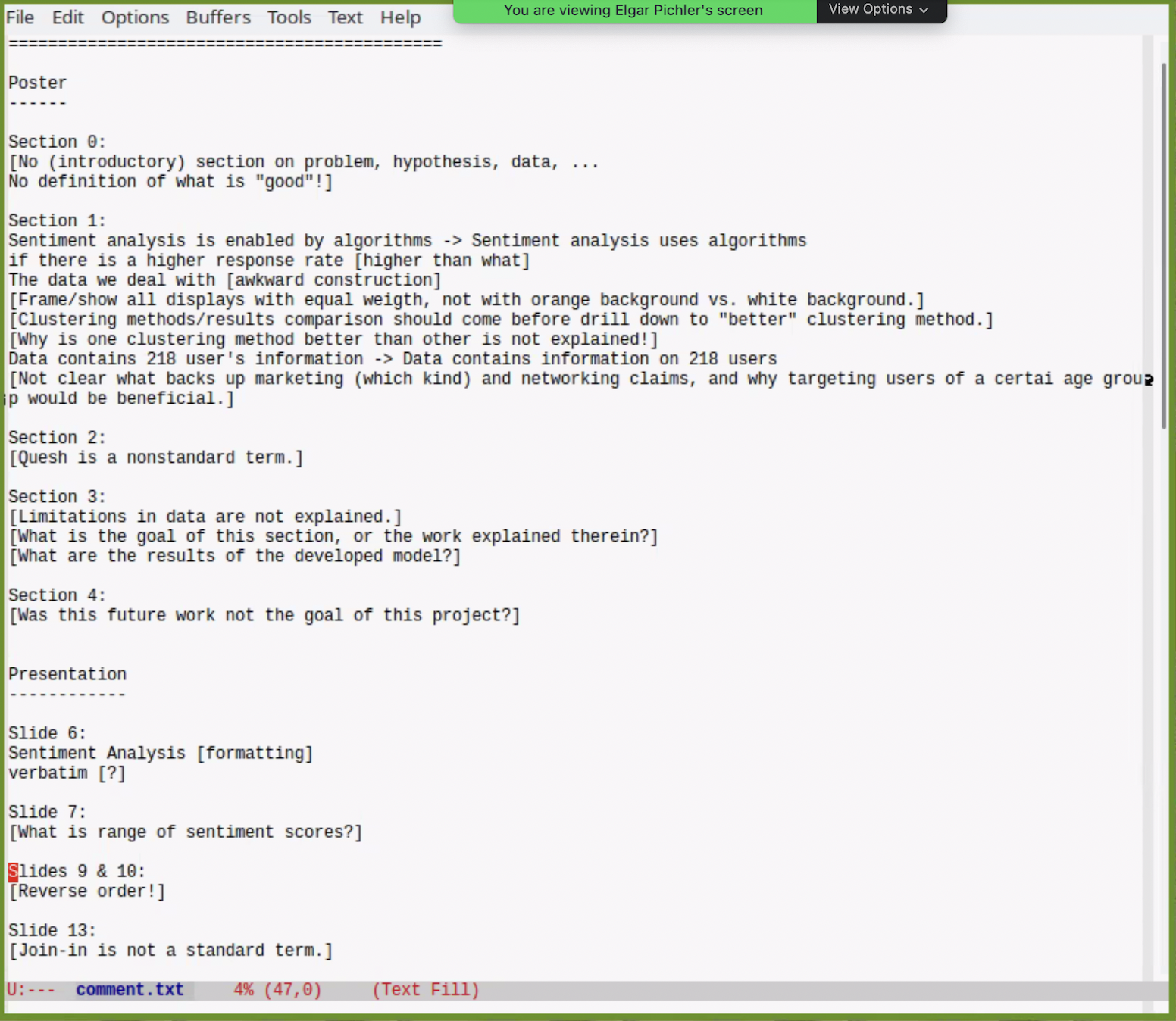
How we define good questions

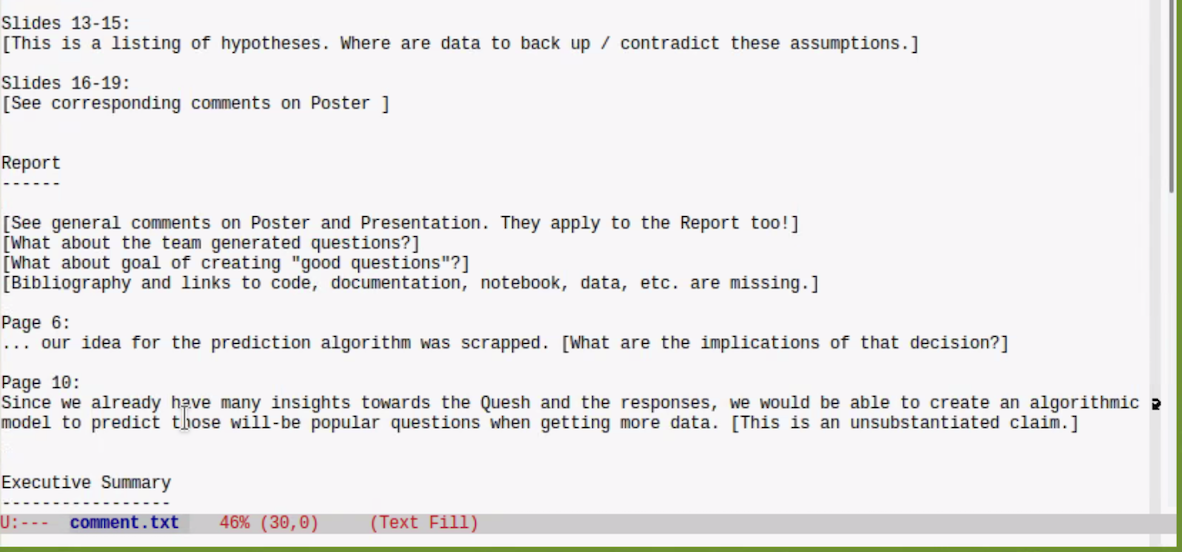
How we relate the peak of new users joining with advertising & marketing

**Sunday 10am**

05/05 - External Meeting With Elgar

* Problem to solve





* Comments (continued)
  + No company name (remove ‘OneQuesh’ from the reports, poster and slides) ✅
  + Slides <- LIKE THAT! Very clean <- no drastic change ✅
  + Poster <- a bit distracting
  + Missing a **link** to github repository <- Add that into the report caro ✅
  + Add some **references** (examples of related papers, internet etc.) caro✅
    - <https://www.bu.edu/articles/2022/onequesh-taking-toxicity-out-of-social-media/>
  + Add & reverse the deliverable about **clustering** (like **why k-means** did not apply) <- we tried clustering to achieve xx goals, and xxx, we think h-clustering gave us the best vision and here shows the results table. Caro ✅
  + **Address hypothesis** (first 3 are good, the following we just said that they did not work) mona✅
  + Correction about the user base statement <- ‘may enlarge the user base’
    - Maybe change the hypothesis? <- Hypo 4

Rewrite hypo5 ( consign? ) mona✅

* + If **networking events** happened, state them. (list the networking events and link them with increase of new users) mona✅
    - <https://www.bu.edu/articles/2022/onequesh-taking-toxicity-out-of-social-media/>
  + **Predictive machine** <- It does not give enough credit to what we did
    - Elaborate what we did and what we achieved
    - What we have concluded after having that & problem add as 4.6 ann ✅
  + Change the **terms into full name (replace or have some explanation about the terms) caro ✅**
  + **Slides (minor change among the same contents before)**
* To do
  + Generate an email to OneQuesh and ask whether they have looked through our package & Do we have permission to use company name and icon ✅

**Report & Slides Saturday 6pm**

**Poster**

**Sunday 6pm**