

Summary

- Overview of problem and hypothesis
- Overview of data
- Data Models
- Practical Applications
- Real-World Implementation

Overview of Problem and Hypothesis

Problem

No available metrics/standards for counselors working with an at-risk population through online chat

- How much time to spend on chat?
- What are keywords to watch out for?

Hypothesis

I believe that we can use data models to create chat metrics and identify patterns for at-risk users

- Use **Logistic Regression** to determine ideal chat time
- Use **Naive Bayes** to classify text to determine at-risk users

Overview of Data

Data

1. Pre-Chat Survey
2. Chat Logs
3. Post-Chat Survey
4. Counselor Evaluation

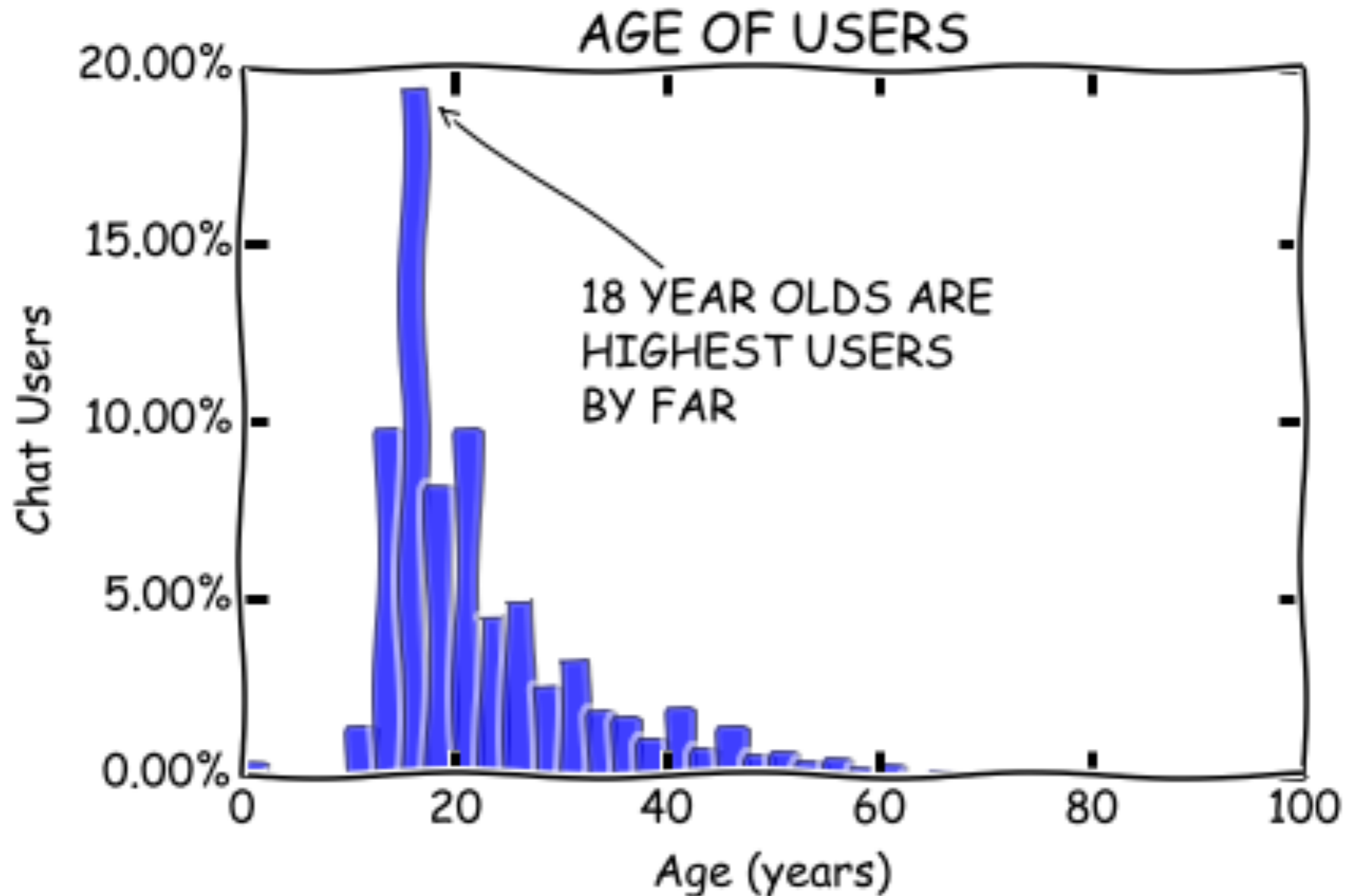
Overview of Data

Pre-Chat Survey

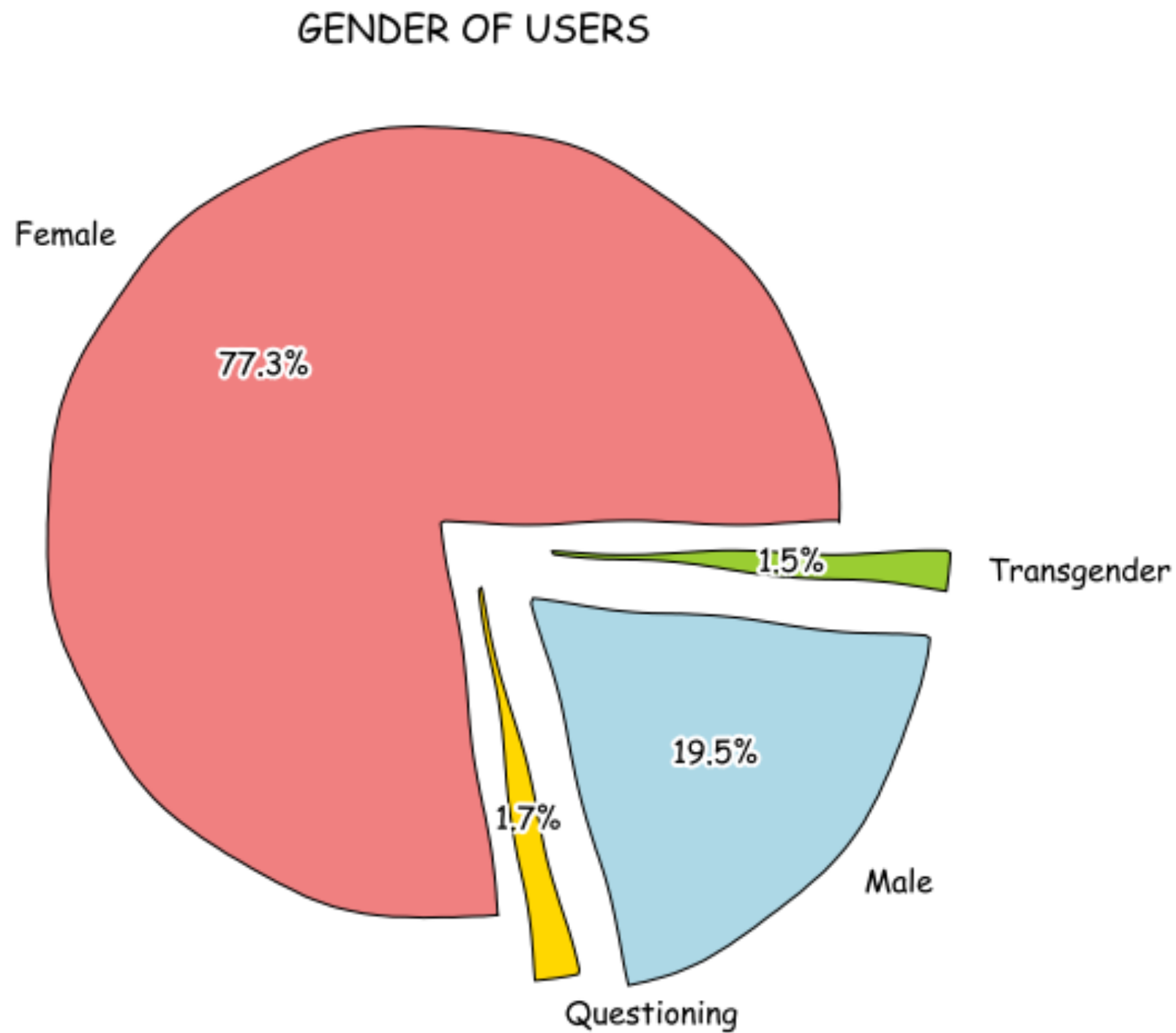
- **Age** ($n=76,712$)
Free-form text: Example: 23
- **Gender** ($n=76,712$)
Drop-down: Male, Female, Transgender, Questioning
- **What are you concerned about today?** ($n=70,602$)
Drop-down: Depression, Anxiety, Eating Disorder, Bullying, etc.
- **Do you have thoughts of suicide?** ($n=71,073$)
Drop-down: Yes - Current (within the past 24 hours)
Yes - Recent Past (within the past few days)
No
- **On a scale of 1-5: How upset are you?** ($n=76,428$)
Drop-down: 1 = I'm doing OK
2 = A little upset
3 = Moderately upset
4 = Very upset
5 = Extremely upset

Overview of Data

Pre-Chat Survey

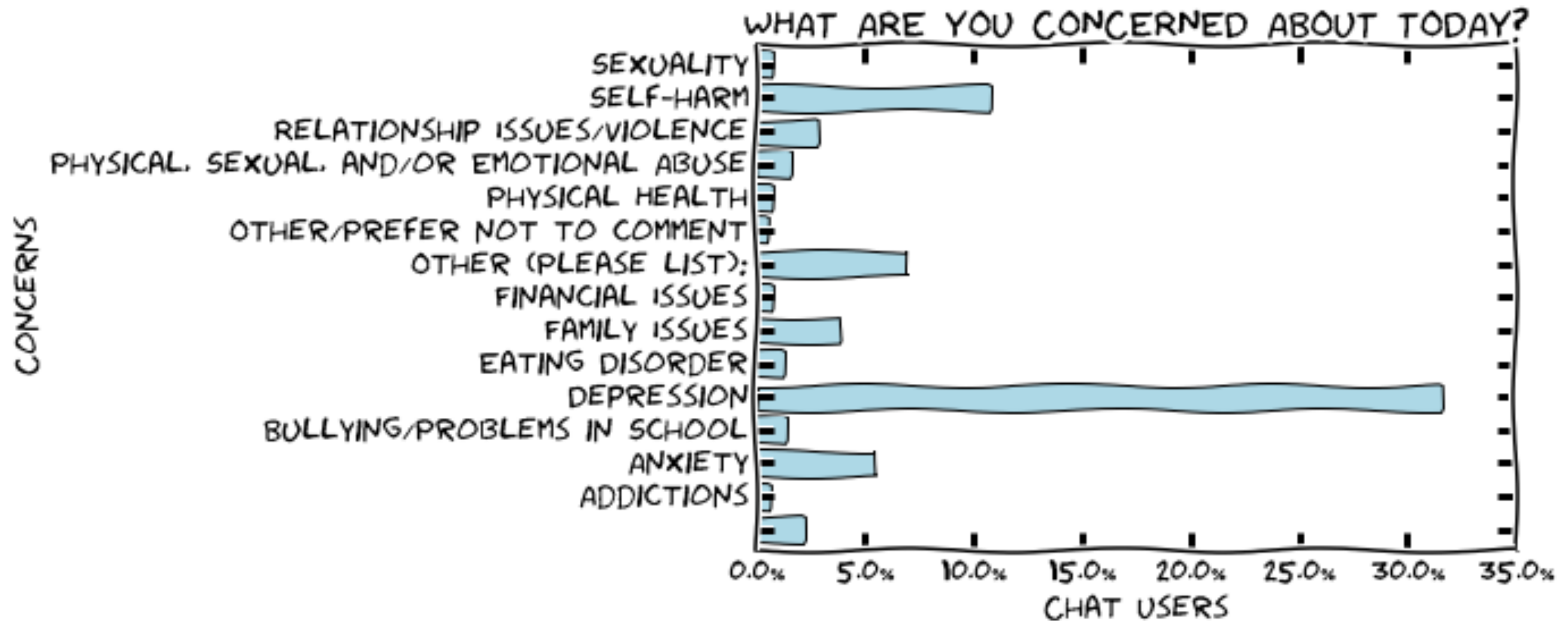


Pre-Chat Survey



Overview of Data

Pre-Chat Survey



Overview of Data

Chat Logs

- **Start and End Time** ($n=33,612$)

Example: 1/10/2014 10:32:38 PM, 1/10/2014 11:04:21 PM

- **Segment Text** ($n=2,520,542$)

Each segment is a message sent

Example: "Welcome to chat, my name is Brittany. How can I help today?"

- **Segment Type**

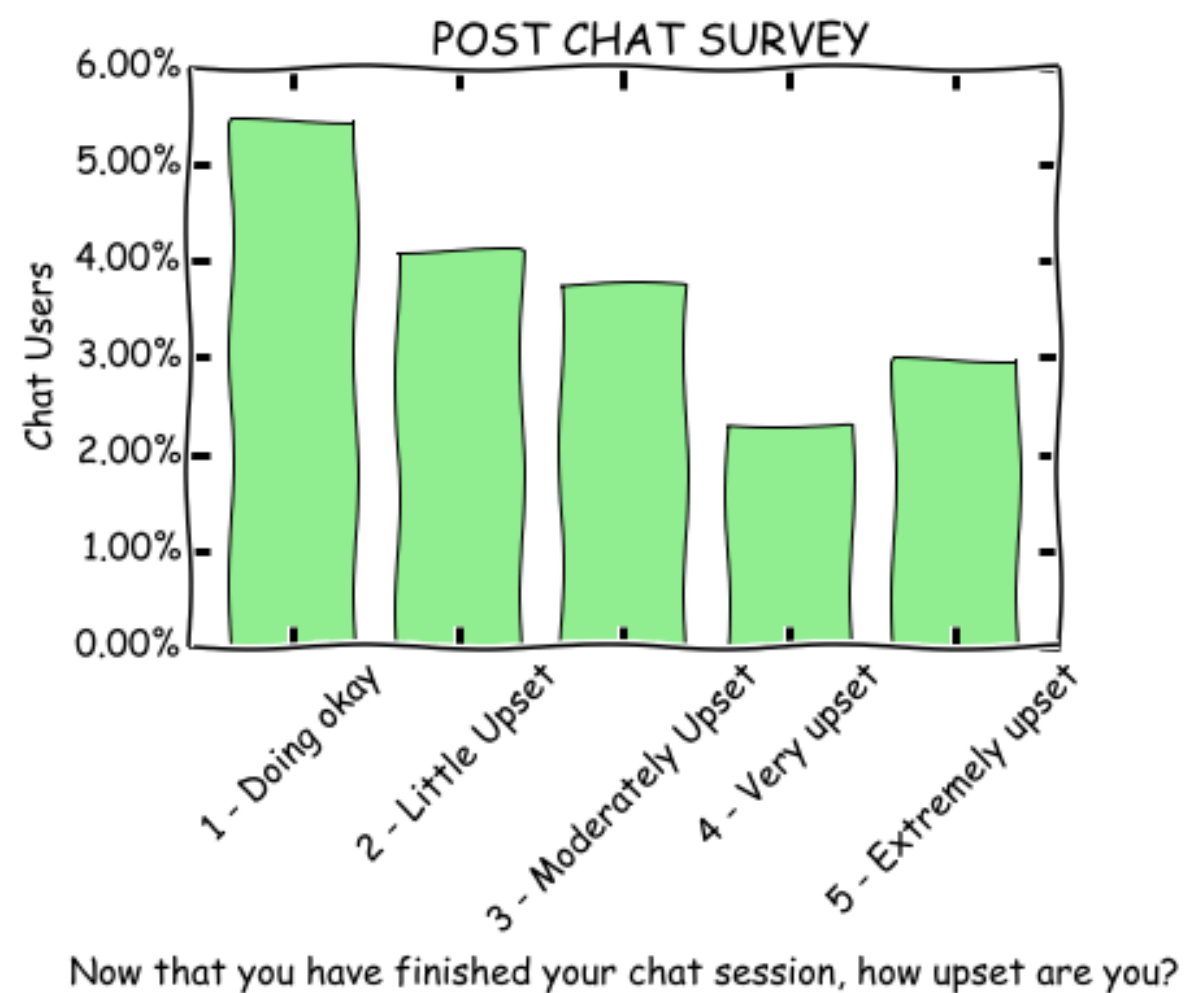
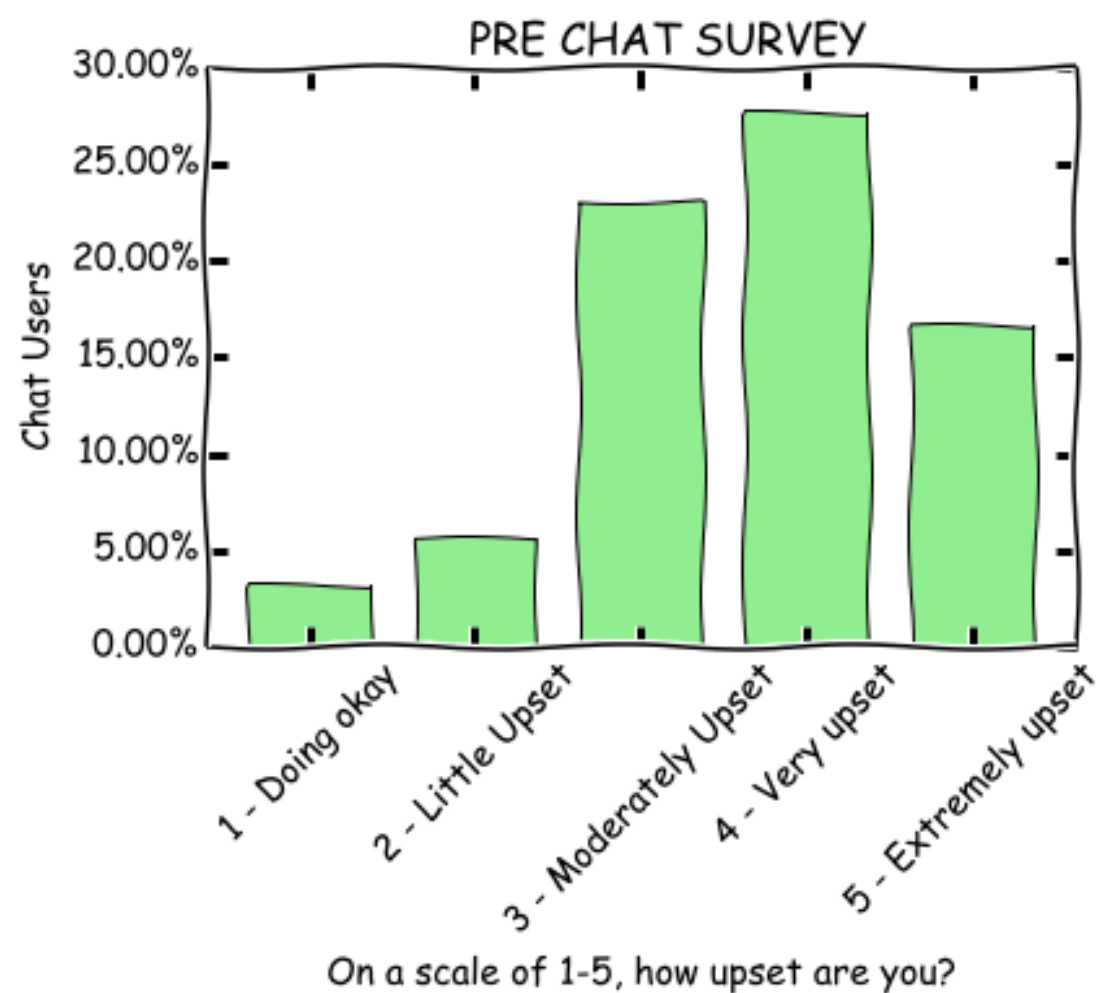
This describes who typed the segment

- Counselor
- User

Overview of Data

Post-Chat Survey

- Did you find this chat service helpful? ($n=16,473$)
Drop-down: Yes, No
- Now that you have finished your chat session, how upset are you? ($n=18,640$)
Drop-down: Scale of 1 (Okay) to 5 (Extremely Upset)



Counselor Evaluation

- Was suicidal ideation present? ($n=9,843$)
Drop-down: Yes
No
- Was user at imminent risk for suicide? ($n=9,742$)
Drop-down: Yes
No
- Was emergency rescue dispatched? ($n=9,753$)
Drop-down: Yes
No

Question 1: How much time to spend per chat?

Fields

- Time Spent on Chat *(in minutes)*

Source: Chat Logs

Calculated from 'End Time' - 'Start Time'

- Did you find this chat service helpful? *(Yes / No)*

Source: Post-Chat Survey

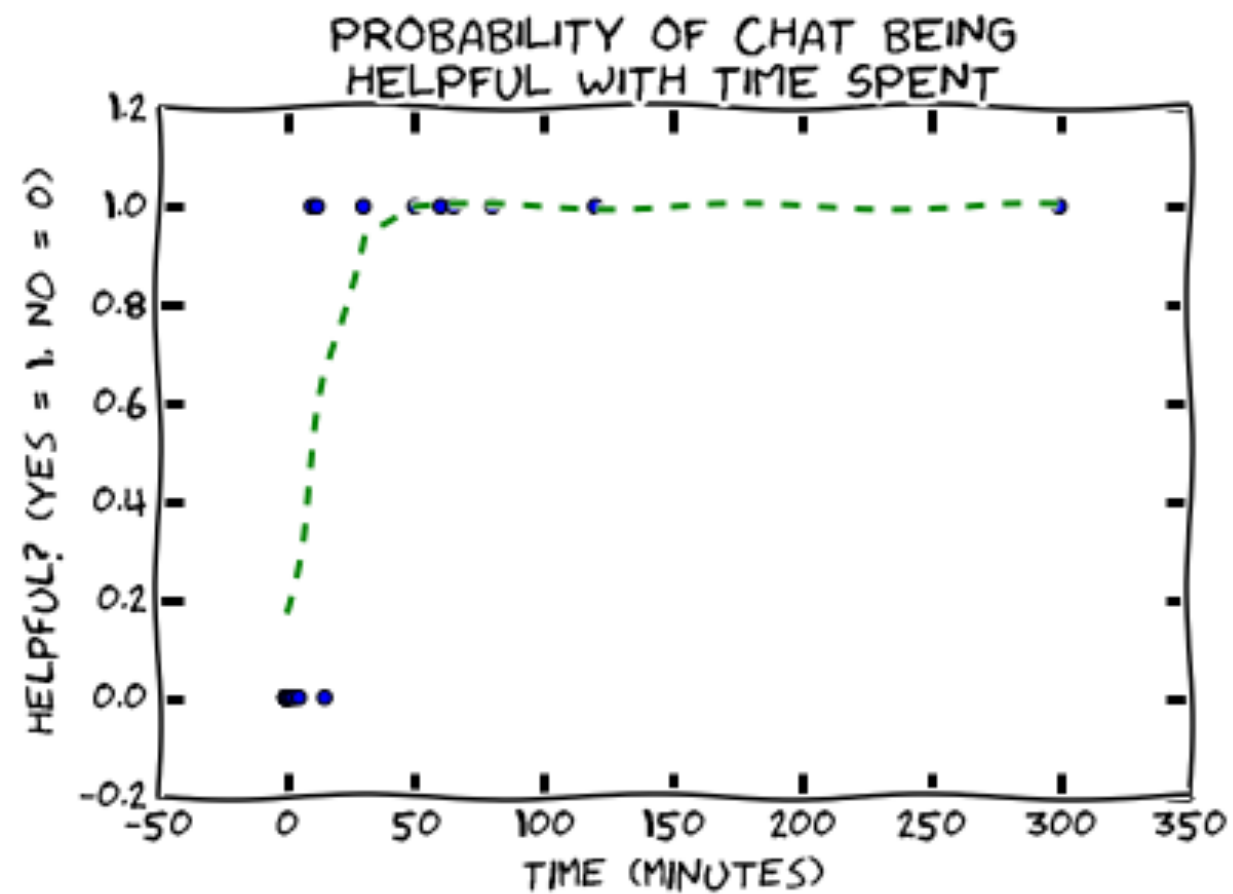
Data Model

Logistic Regression - I'm expecting that chats will be more helpful given more time.

My hypothesis is that the data will fit a sigmoid function.

Data Models

What I thought would happen

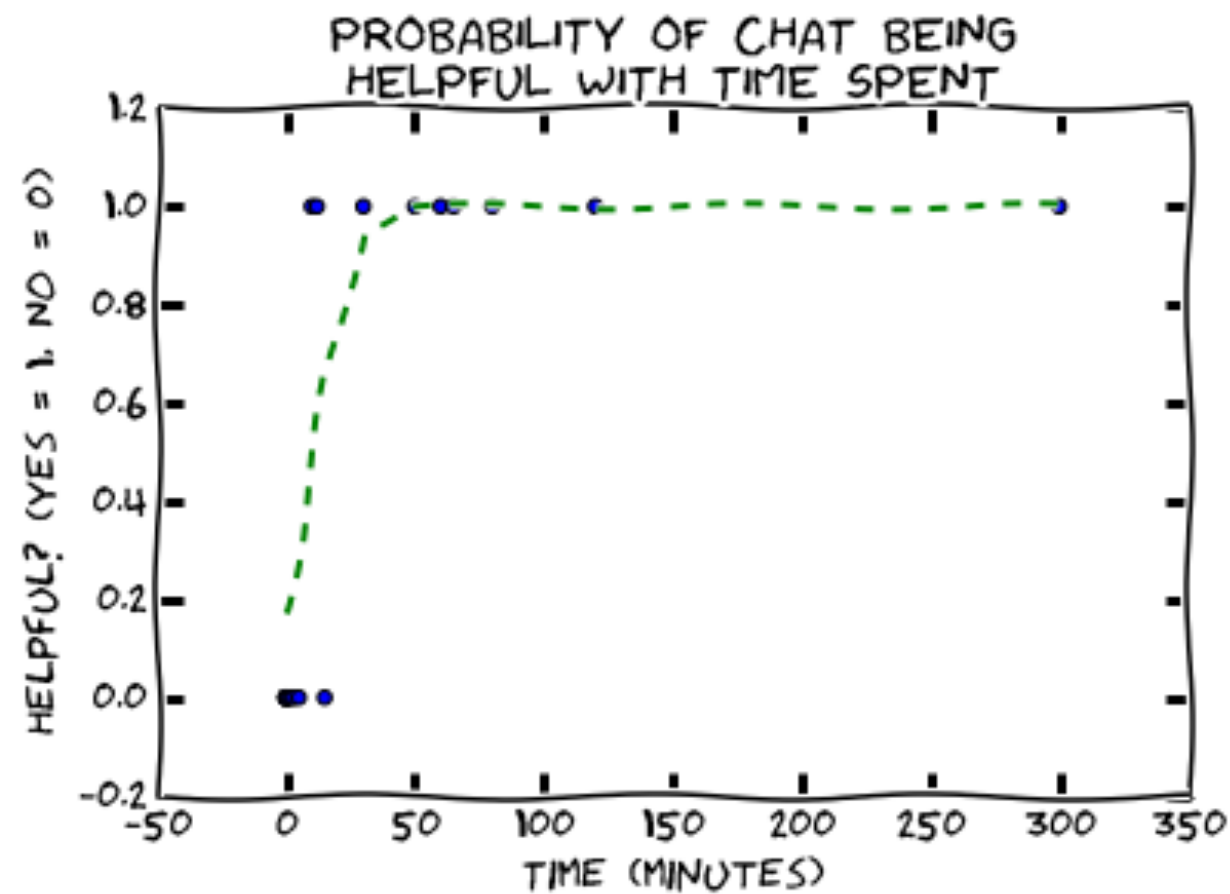


Green Dashed Line = Predicted Probability

Blue Mark = One Chat Session

Data Models

What I thought would happen

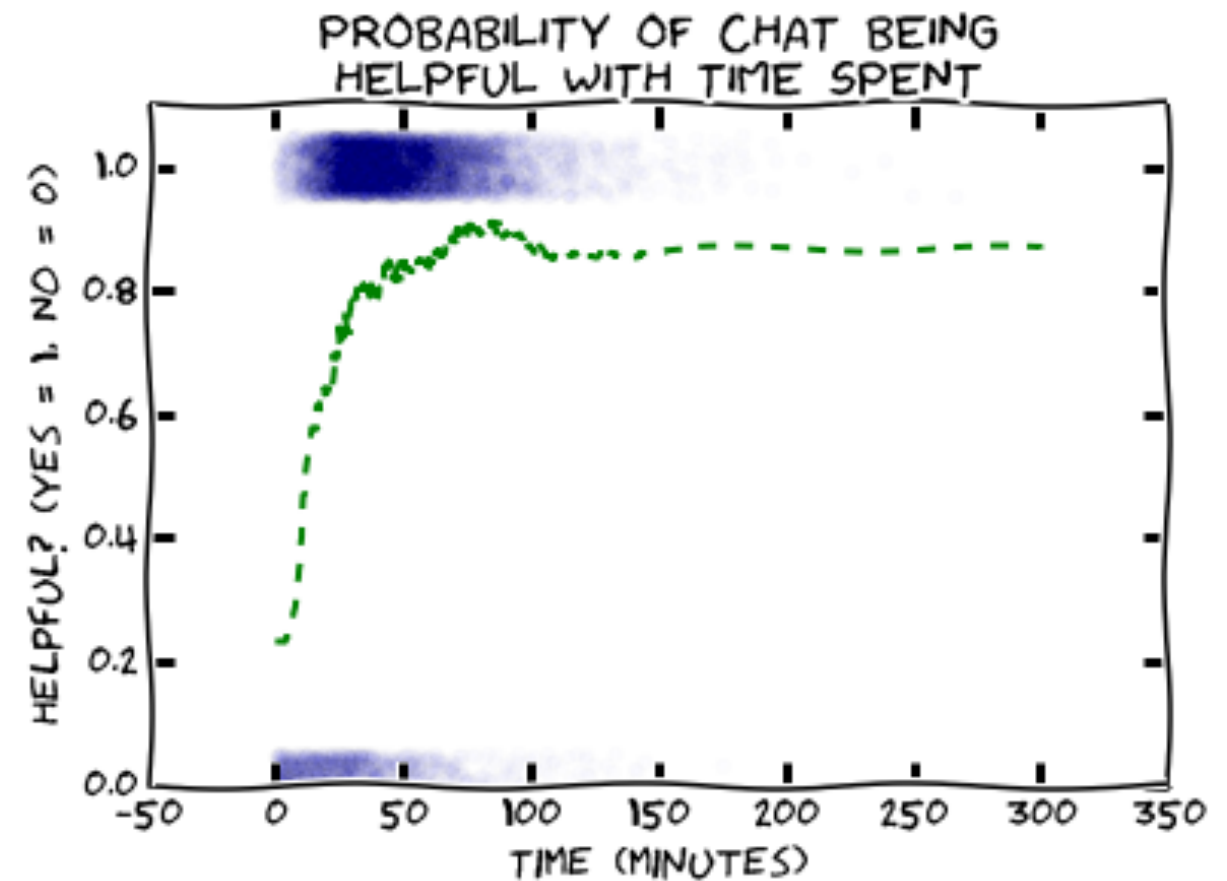


Green Dashed Line = Predicted Probability

Blue Mark = One Chat Session

What really happened

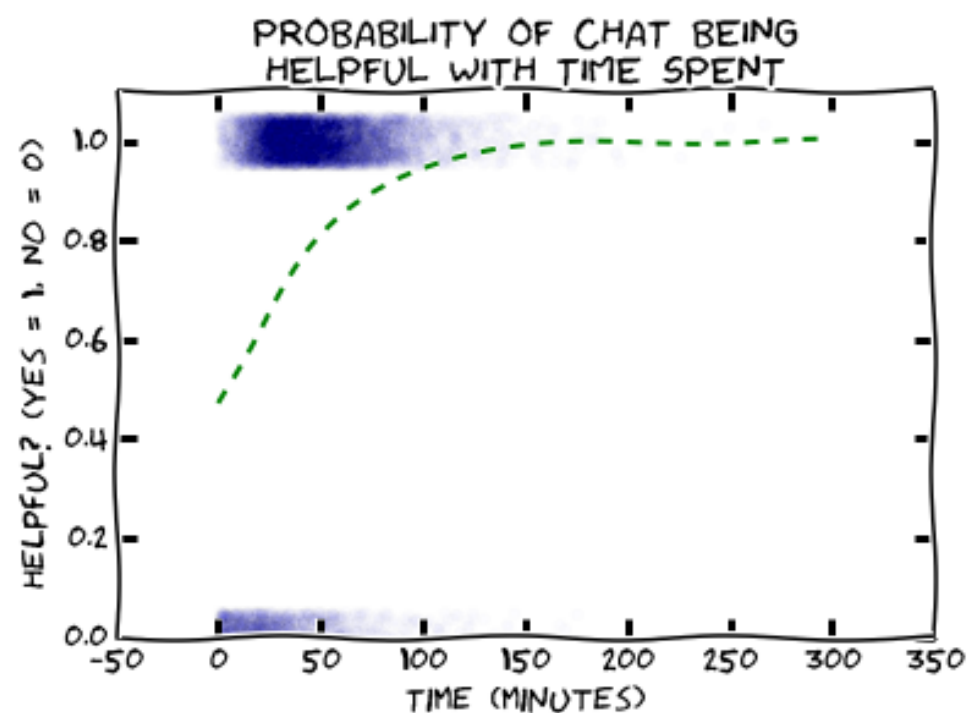
Jitter added to visualize clumped data better



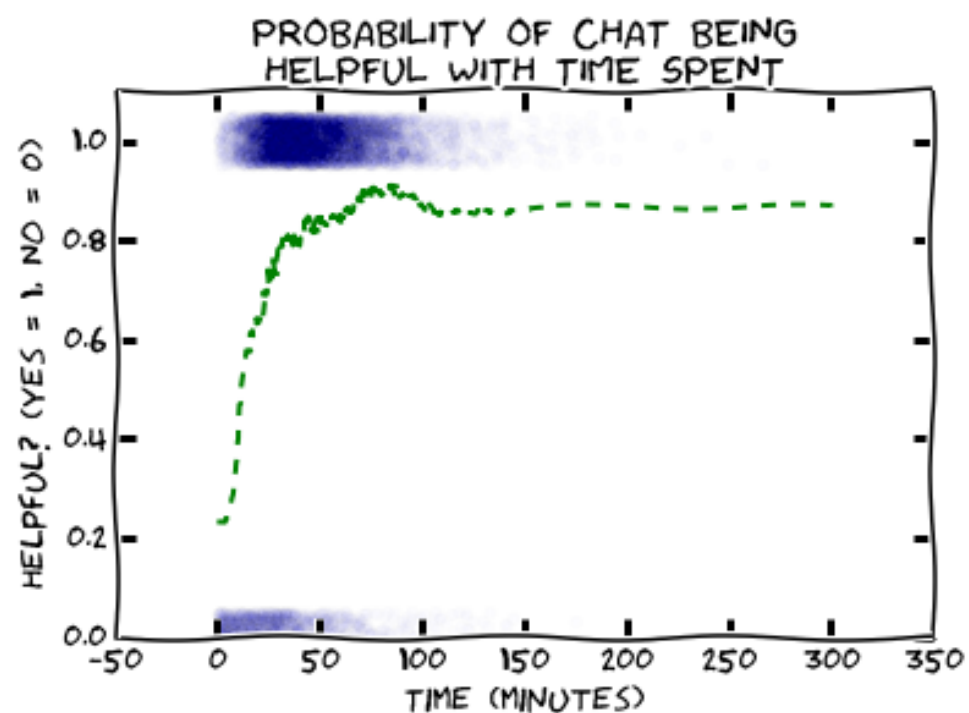
Data Models

Results of Different Models

Logistic Regression

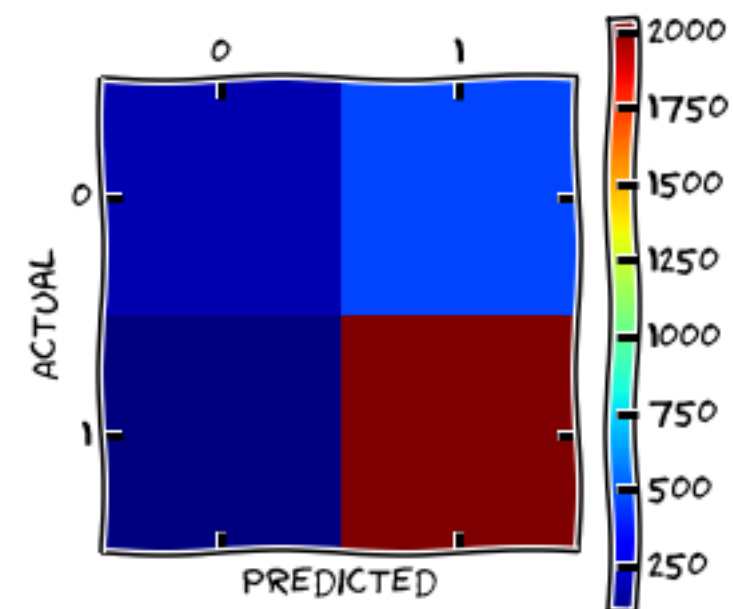
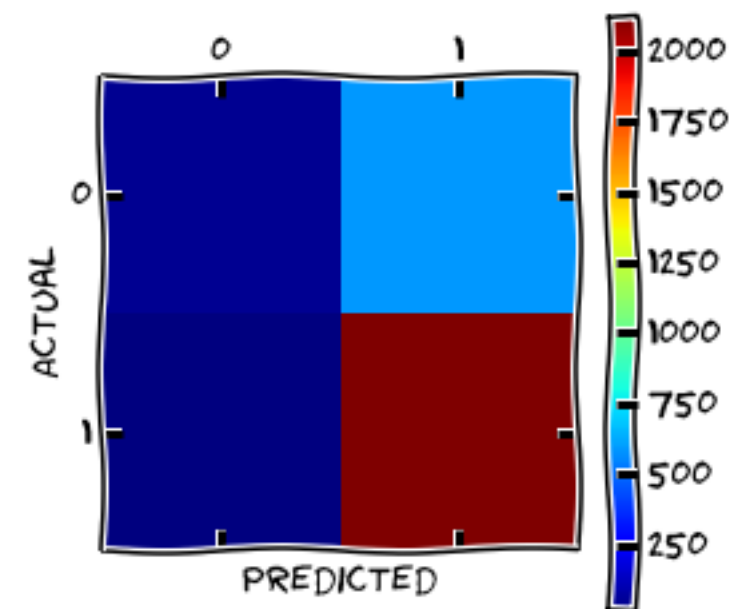


K Nearest Neighbor



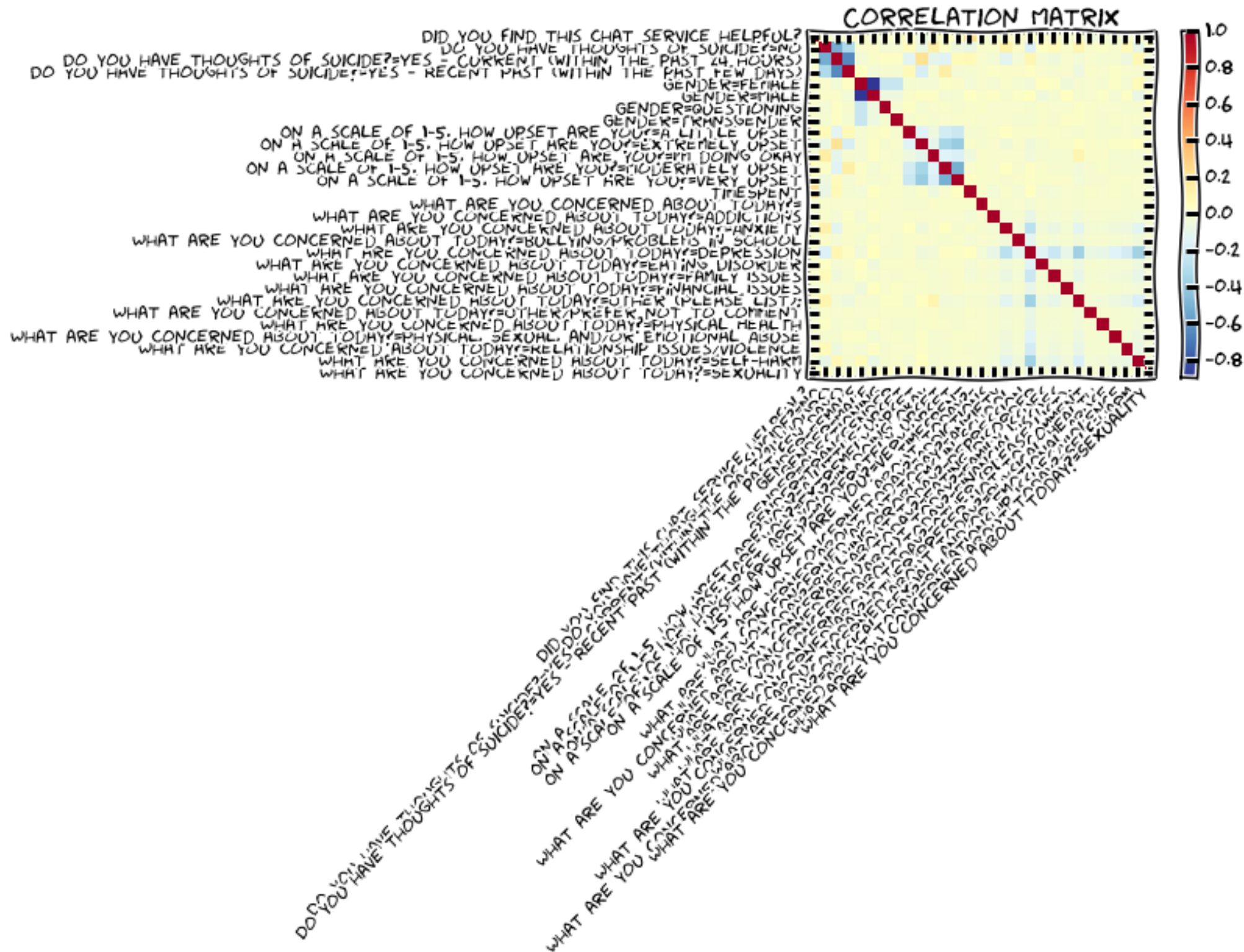
Confusion Matrix

Predicting if chat service was helpful



Data Models

Correlations



Question 2: What keywords to watch for?

Features

- Chat Segments *(text)*

Source: Chat Logs

Filtered only for User Segments (i.e. no Counselor messages)

Removed 'Stop Words' (e.g. 'a', 'the', 'that')

- Do you have thoughts of suicide? *(Yes / No)*

Source: Pre-Chat Survey

Combined Answers 'Yes - Current (within the past 24 hours)' with 'Yes - Recent Past (within past few days)'

Data Model

Multinomial Naive Bayes Classifier using TF-IDF Vectorizer

Word Patterns with suicidal ideation

unigram

afraid, anxiety, bad, believe, college, cut, dad, depressed, family, feeling, friend, hard, home, hospital, hurt, issues, job, kill, living, love, mom, money, pain, parents, pills, problem, sad, scared, school, sleep, sorry, talk, thank, therapist, thoughts, tired, wanted, wish, work, wrong, years

bigram

best friend, dont know, feel better, feel like, just need, just want, long time, need help, need talk, really bad, really hard, self-harm, suicidal thoughts, years ago

trigram

having hard time, having suicidal thoughts, just don't know, just feel like, just want stop

How is this useful?

Question

How much time to spend per chat?

Answer

We have an estimate of how long chats will take in order for the user to find chats as helpful

.....

Question

What are keywords to watch out for?

Answer

See previous slide

Real World Implementation

Difficulties

- Getting data in real-time from a variety of sources
 - Internal Server (Surveys, Chat Logs)
 - SurveyMonkey (Counselor Evaluations)
 - Data from other crisis centers (Surveys, Counselor Evaluations)