Consultancy Reflections: Applying Big Data Analysis Approaches in the Real-world

MILLIGAN, Gregor

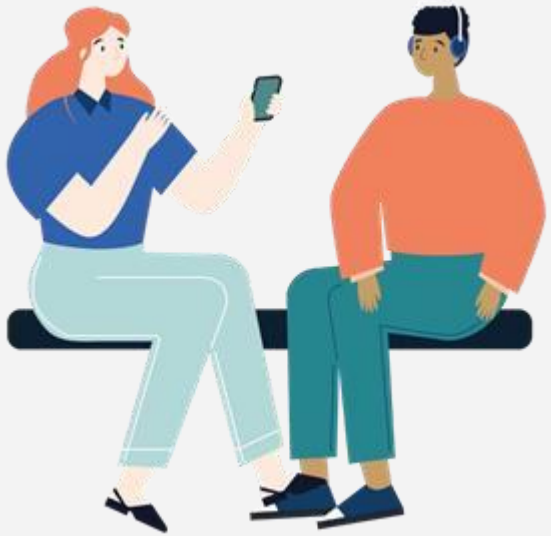
University of Nottingham
UK | CHINA | MALAYSIA

Big Data Business Projects Slides

Gregor Milligan
gregor.milligan@nottingham.ac.uk

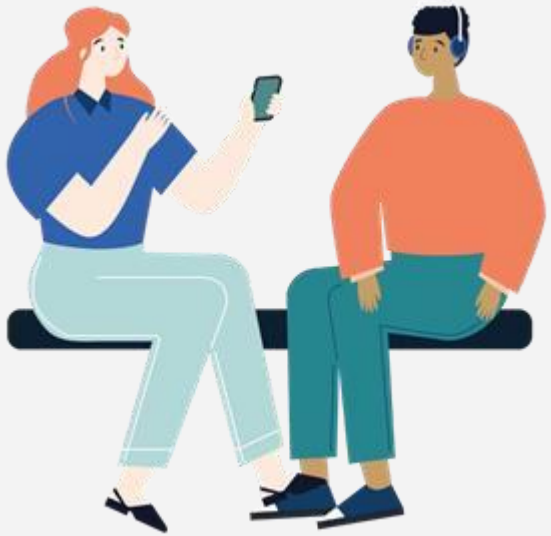
Kooth Projects

PhD Placement → Consultancy



Kooth Overview

- Kooth is a digital counselling and support platform for children and young people (CYP) aged between 11-25 years old Qwell is a digital counselling and support platform for Adults over 18



- Anonymous at point of entry

Kooth Overview

kooth

Drop-in/Scheduled Chats with practitioners

Message Boards

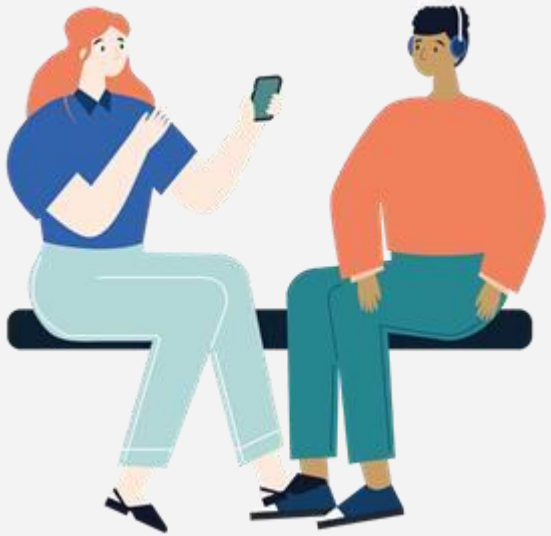
Personal Journal

Kooth Projects

Moderation Automation

Session Wants and Needs

Forum Post Analysis



Kooth Takeaways

More time chatting

More understanding the problem

Less time coding

Tech Stack

kooth

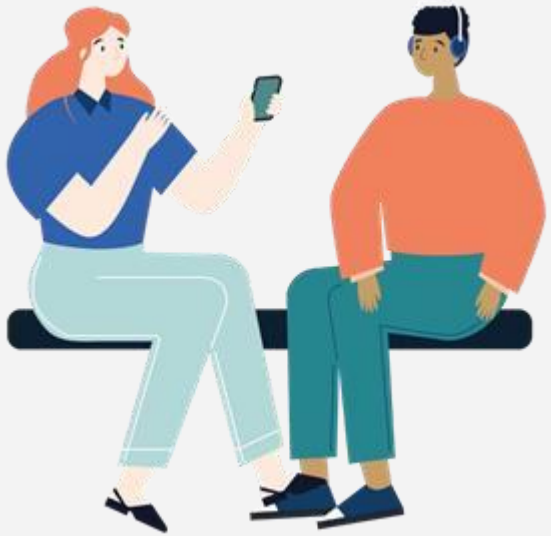
Python

Tableau

Power BI

SQL

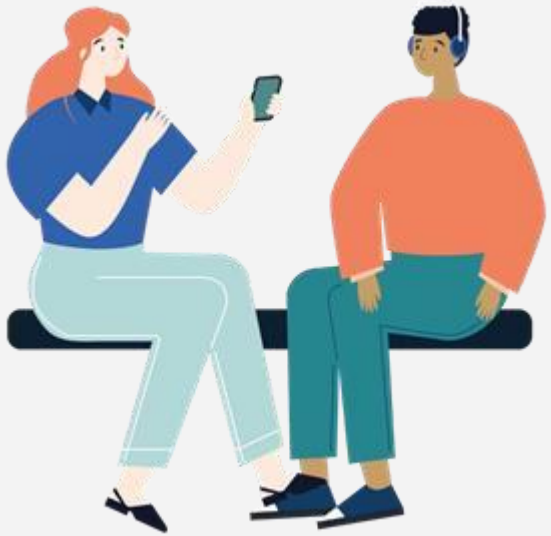
Project Management Tools



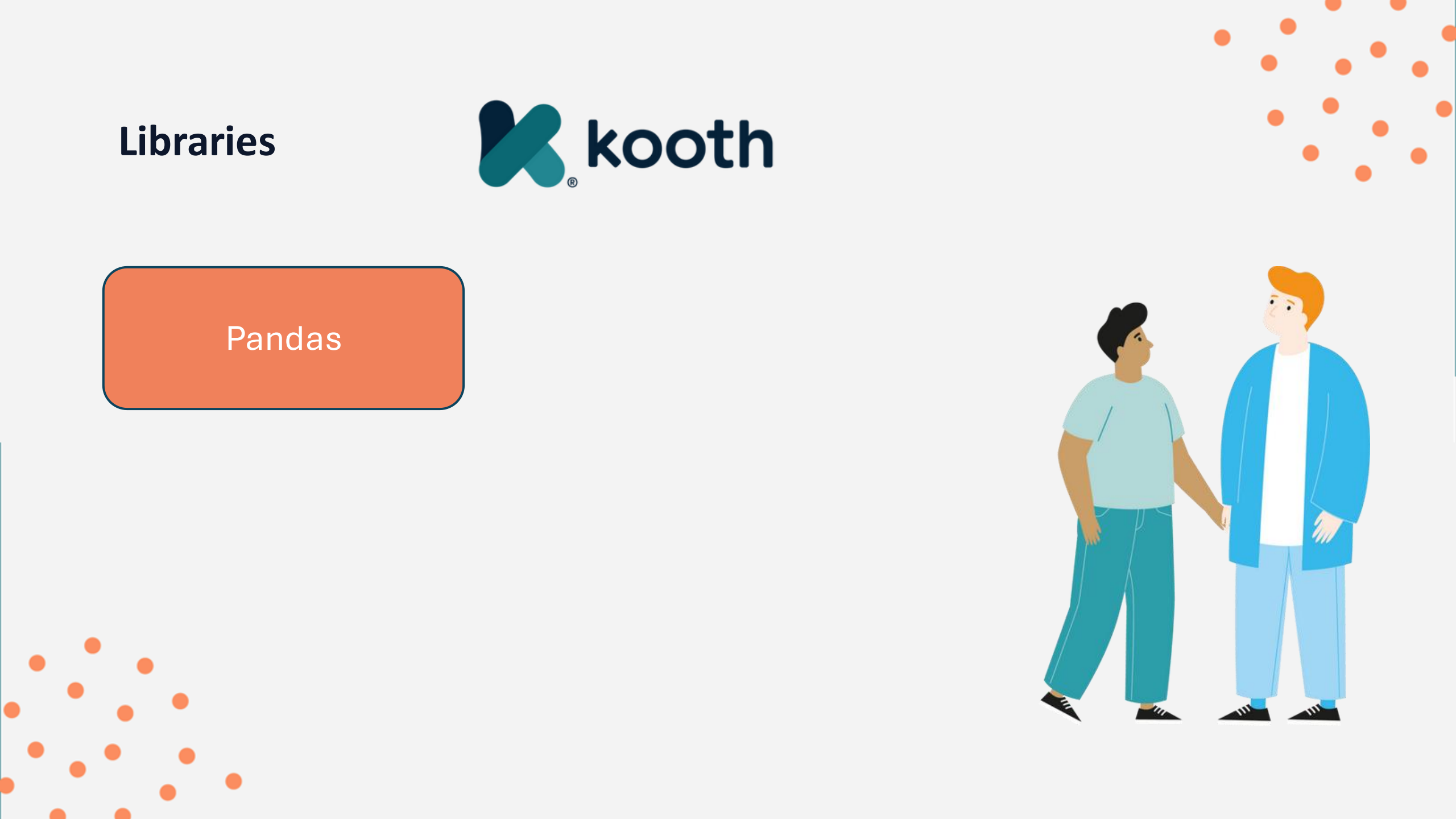
Tech Stack

kooth

GIT HUB



Libraries



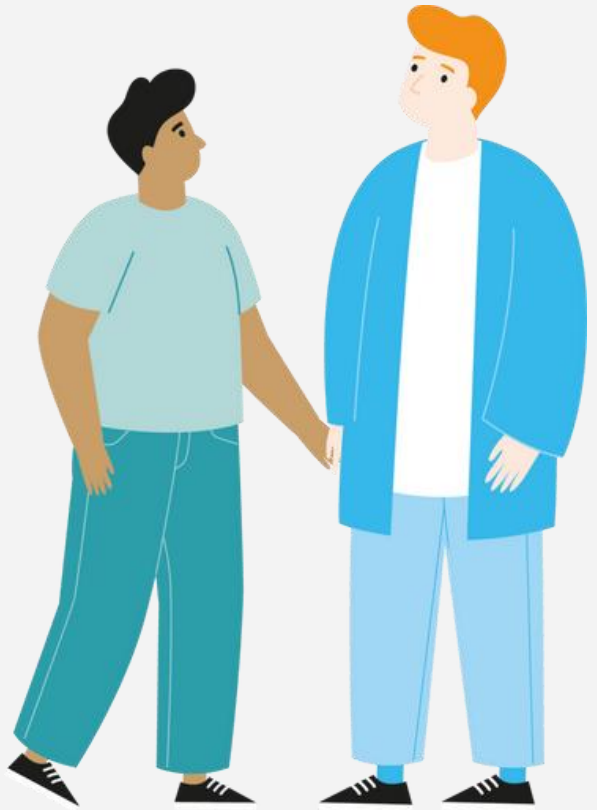
Pandas

Libraries

kooth

Pandas

Seaborn/Matplot lib



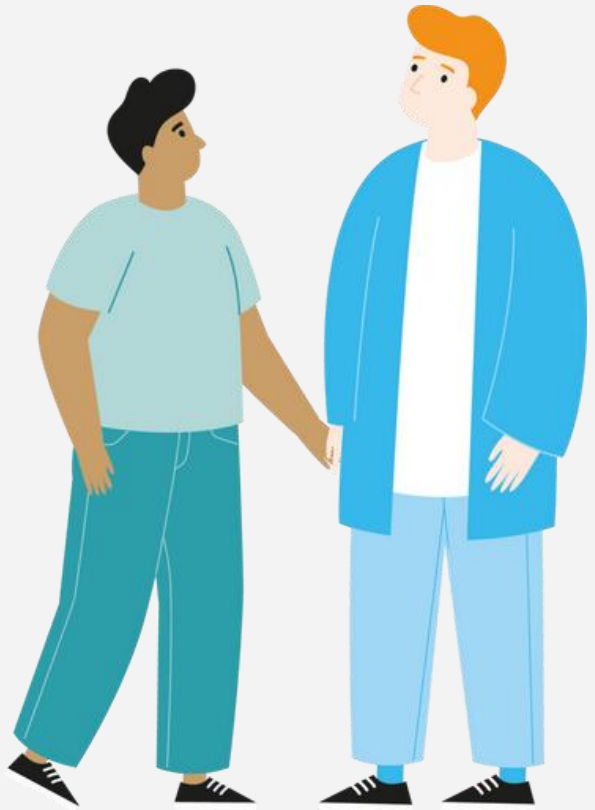
Libraries

kooth

Pandas

Seaborn/Matplotlib

sklearn



Libraries

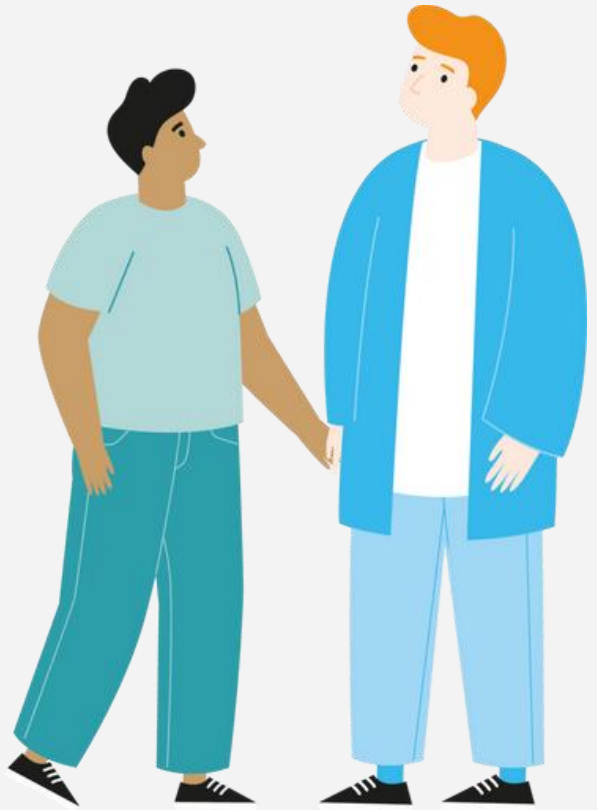
kooth

Pandas

Seaborn/Matplotlib

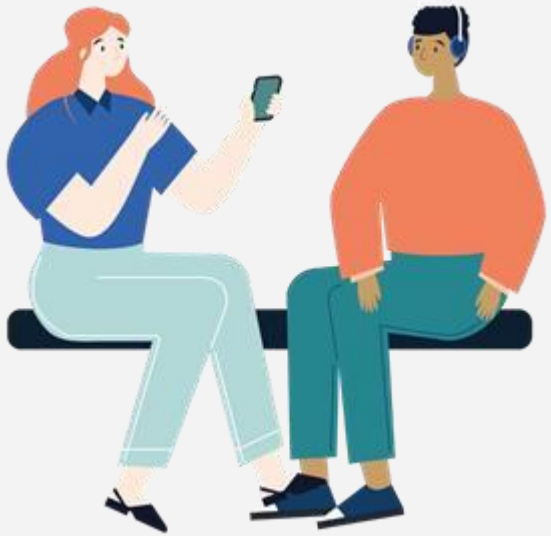
sklearn

Transformer



Roles Within kooth

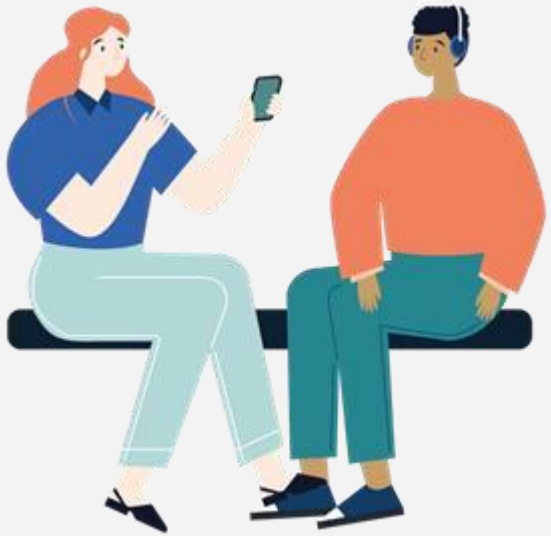
Data Analyst



Roles Within kooth

Data Analyst

Data Engineer

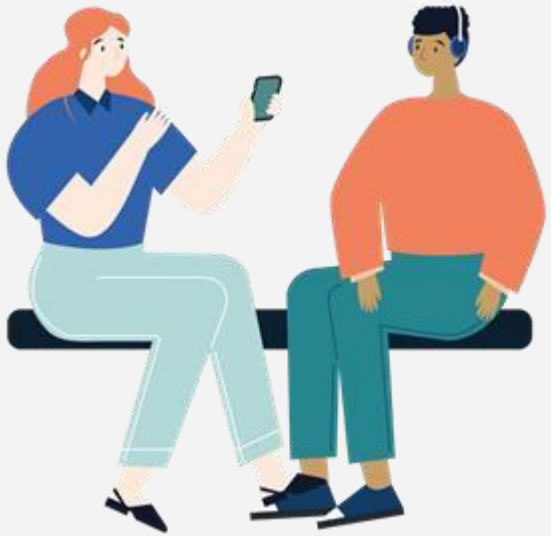


Roles Within kooth

Data Analyst

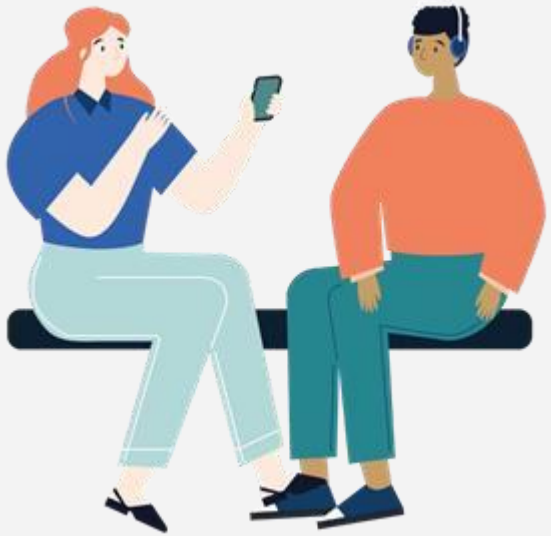
Data Engineer

Data Scientist



Roles Within kooth

- Data Analyst
- Data Engineer
- Data Scientist
- Research Lead



Kooth Takeaways

kooth

Real World Data Is Never Clean

Assisted moderation

Assisted moderation aims to **improve responsiveness** to service users and **efficiency of moderator workflows** by **automating detection** of low risk posts and **enriching post data**.

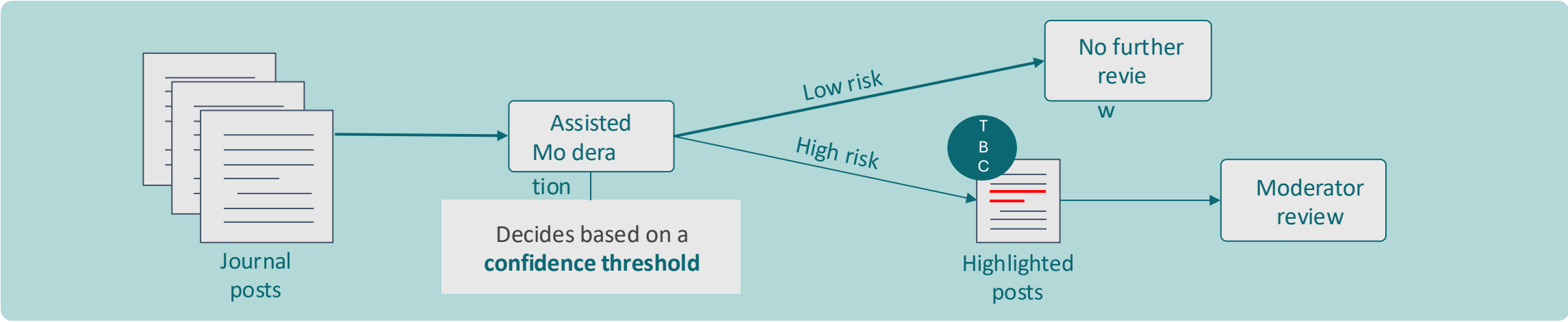
Moderation Automation

Process of detecting risk level of posts and highlight high risk phrases within the text

Journal Assisted Moderation Tool

Journal Assisted Moderation

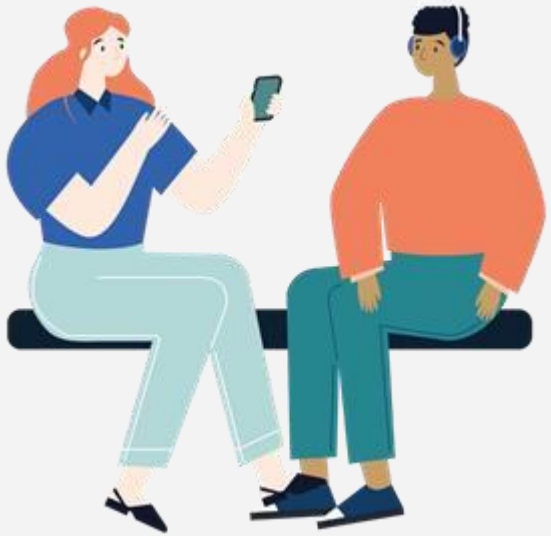
1. **Estimates** whether journal posts need further **moderator review**
2. **Highlights high risk phrases** within journal text



Assisted moderation

The following functionalities are being explored within Data Science:

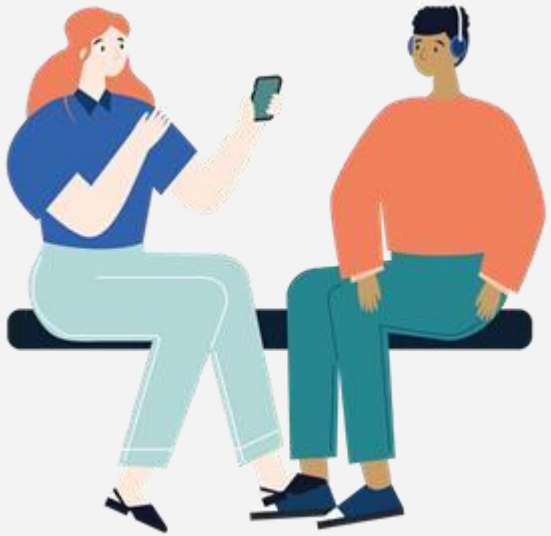
- **Identification of low risk posts** that do not need further review from a moderator
- **Highlighting high risk phrases** in posts to assist with long form moderation



Moderation Automation

Assisted moderation – The Bigger Picture

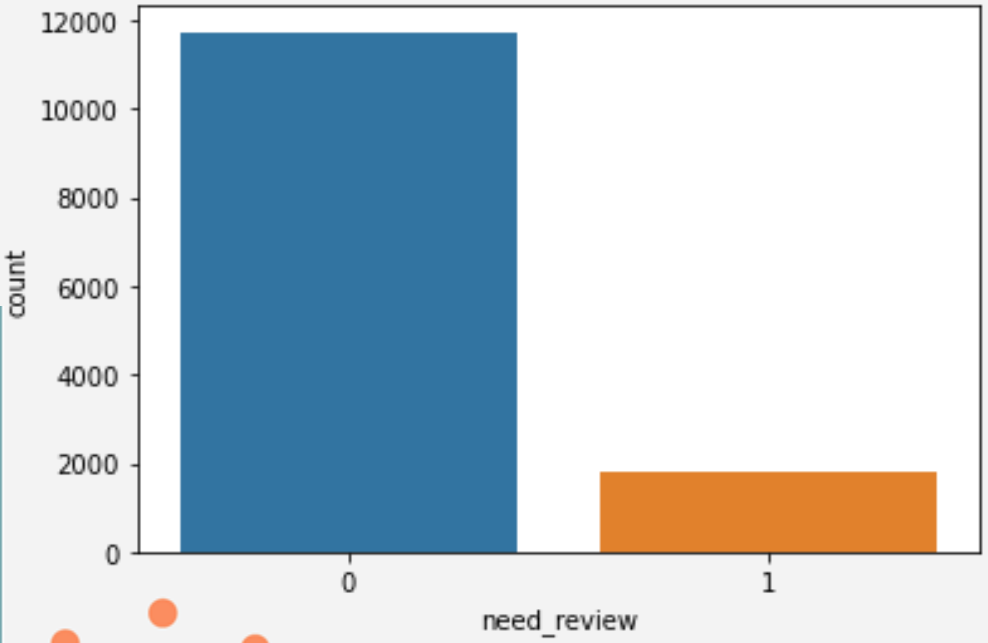
- A tool for **managing moderation workflows**. This
- tool will **improve usability**, **prioritise posts**, **provide context to posts** and **improve tracking** of moderation activity for auditing and escalation.



Moderation Automation

Real Life Challenges and Experience

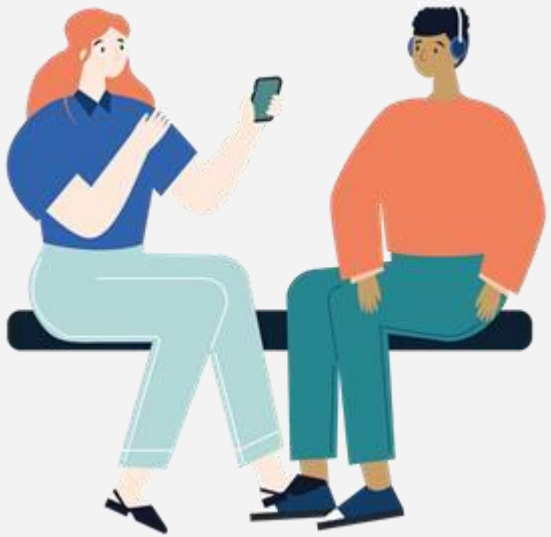
Data Imbalance



Model Exploration

Next Steps for AM

Work with the data engineering and product team to trial the RoBERTa model in production **Spend time with the moderation team to understand their Expectations**



counselling & psychotherapy research
Linking Research with Practice
bacp | counselling changes lives

ORIGINAL ARTICLE | Open Access

Developing a single-session outcome measure using natural language processing on digital mental health transcripts

Gregor Milligan, Aynsley Bernard, Liz Dowthwaite, Elvira Perez Vallejos, Jamie Davis, Louisa Salhi, James Goulding

First published: 30 May 2024 | https://doi.org/10.1002/capr.12766

counselling & psychotherapy research
Linking Research with Practice
bacp | counselling changes lives

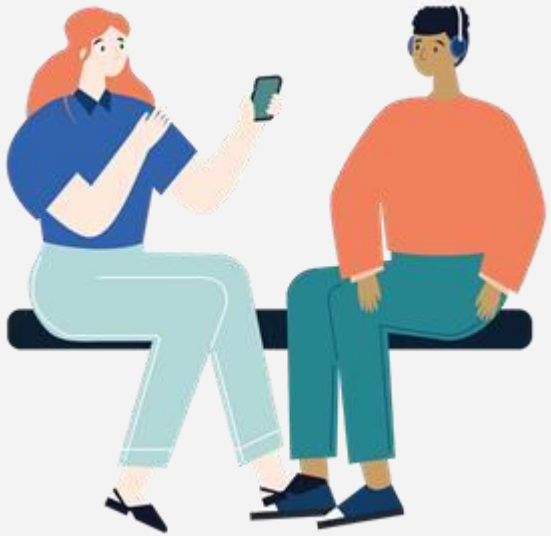
ORIGINAL ARTICLE | Open Access

Developing a single-session outcome measure using natural language processing on digital mental health transcripts

Gregor Milligan, Aynsley Bernard, Liz Dowthwaite, Elvira Perez Vallejos, Jamie Davis, Louisa Salhi, James Goulding

First published: 30 May 2024 | https://doi.org/10.1002/capr.12766

People with lived experience



counselling & psychotherapy research
Linking Research with Practice
bacp | counselling changes lives

ORIGINAL ARTICLE | Open Access

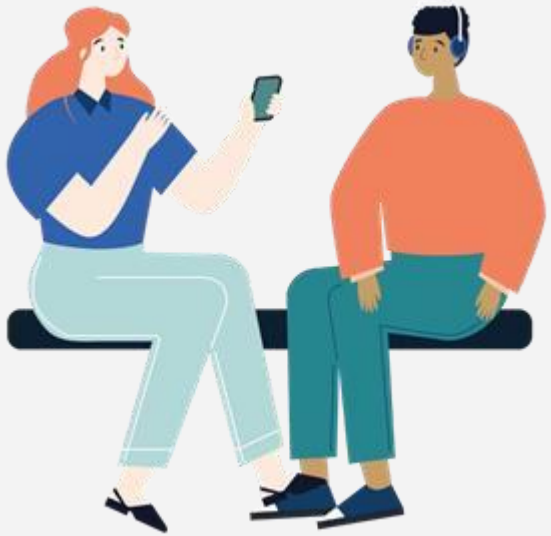
Developing a single-session outcome measure using natural language processing on digital mental health transcripts

Gregor Milligan, Aynsley Bernard, Liz Dowthwaite, Elvira Perez Vallejos, Jamie Davis, Louisa Salhi, James Goulding

First published: 30 May 2024 | https://doi.org/10.1002/capr.12766

People with lived experience

Clinicians/Practitioners



counselling & psychotherapy research
Linking Research with Practice
bacp | counselling changes lives

ORIGINAL ARTICLE | Open Access

Developing a single-session outcome measure using natural language processing on digital mental health transcripts

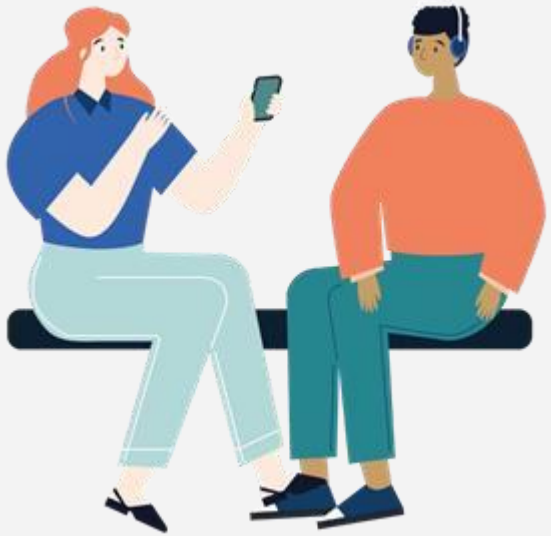
Gregor Milligan, Aynsley Bernard, Liz Dowthwaite, Elvira Perez Vallejos, Jamie Davis, Louisa Salhi, James Goulding

First published: 30 May 2024 | https://doi.org/10.1002/capr.12766

People with lived experience

Clinicians/Practitioners

Data Engineers



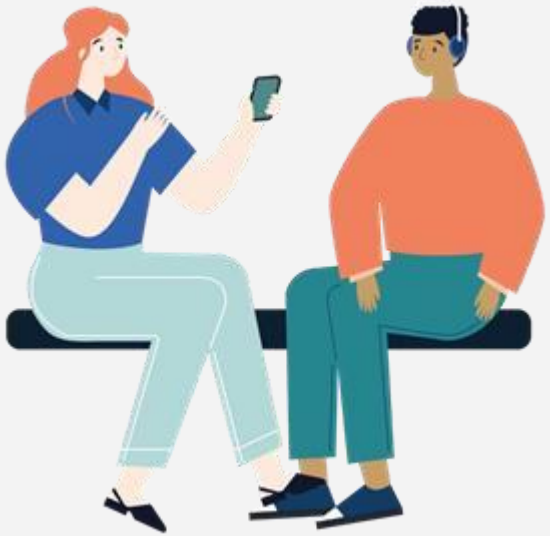
counselling & psychotherapy research
Linking Research with Practice
bacp | counselling changes lives

ORIGINAL ARTICLE | Open Access

Developing a single-session outcome measure using natural language processing on digital mental health transcripts

Gregor Milligan, Aynsley Bernard, Liz Dowthwaite, Elvira Perez Vallejos, Jamie Davis, Louisa Salhi, James Goulding

First published: 30 May 2024 | https://doi.org/10.1002/capr.12766



People with lived experience

Clinicians/Practitioners

Data Engineers

Researchers

Data

Kooth adult data from **May ‘19 - July ‘22**

Service users: **2909, chats: 7254**.

Ob ject iv es

1. **Explore** how NLP can be used to extract **outcomes** of users from **single sessions**

2. **Pinpoint** the **challenges** associated with detecting wants and needs of users through **NLP**

3. **Co-design Adult SWAN-OM** using NLP outputs with **service users, clinical and service delivery**

People with lived experience

Clinicians/Practitioners

Researchers

Data Engineers

Research Steps

- Gather transcript data and labelled outcomes

- **Supervised learning stage**: learn the **textual features** that predict good EOS outcomes, then **highlight different wants and needs**
- **Unsupervised learning stage**: filter text to only look at phrases that link to good EOS scores, then extract these topics to create an initial list of potential measure items
- Consult with **SUs and Experts** to determine if the items we have generated are **useful** when a SU is **engaging in a single session**?

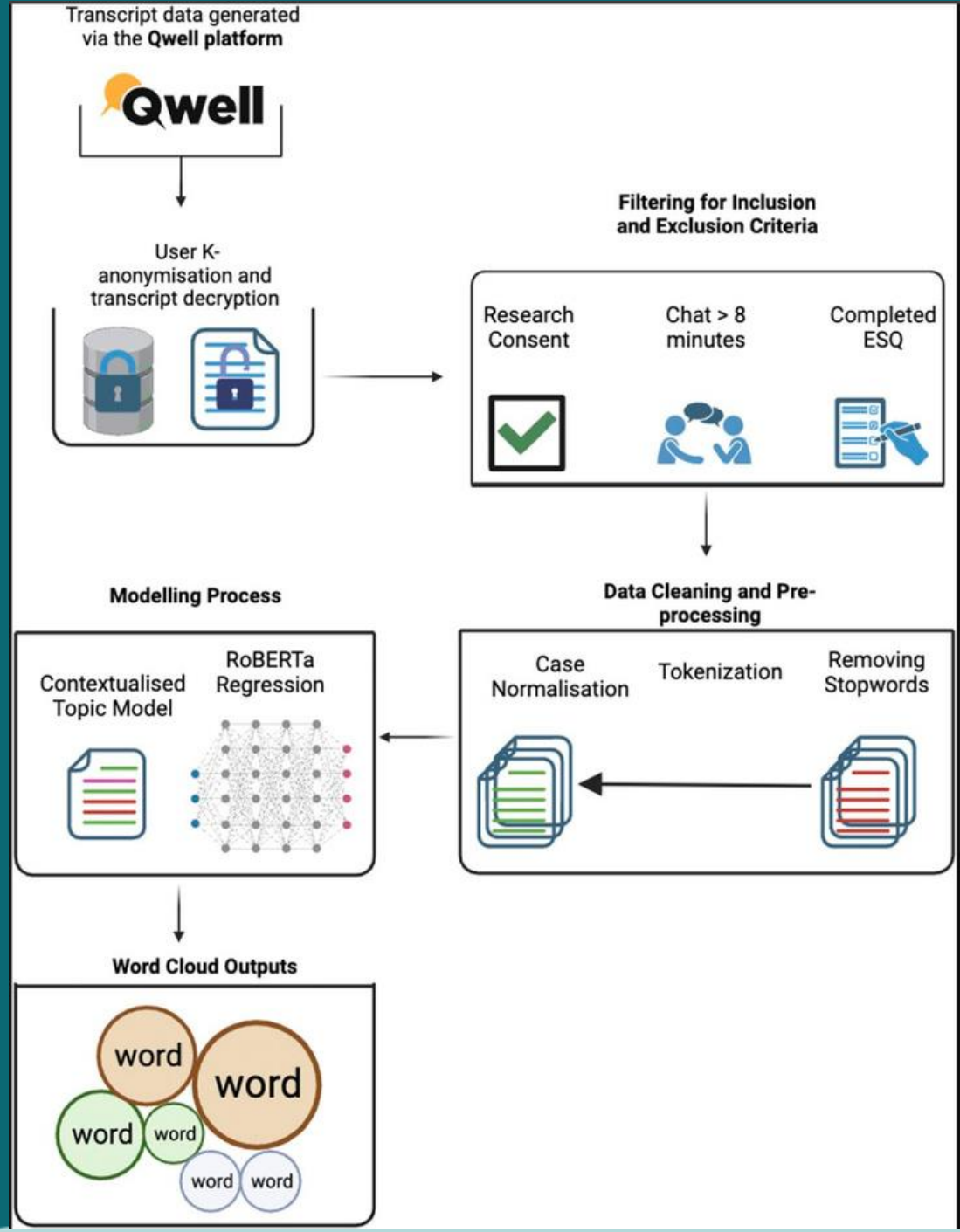
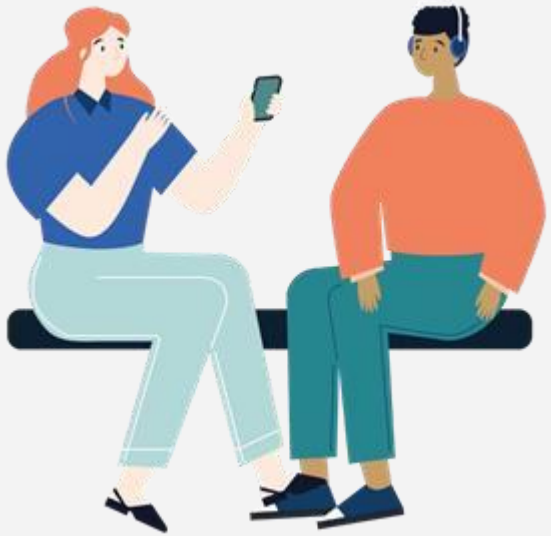


TABLE 3 Finalised items selected from the expert workshops.

Item #	Statement pre-chat	Statement post-chat	I-CVI relevance	I-CVI clarity
Item 1	Discuss and explore how to improve a specific relationship	I have discussed and explored how to improve a specific relationship	0.71	0.83
Item 2	Learn about coping strategies	I learned about coping strategies	1	1
Item 3	Feel heard or understood	I felt heard and understood	0.83	0.83
Item 4	Build my support system	I have found ways to build my support system	1	0.83
Item 5	Learn how to become more accepting of myself	I have learnt how to become more accepting of myself	0.83	1
Item 6	Help with grieving a loss	I had help with grieving a loss	1	1
Item 7	Understand how my physical and mental health could be linked	I understand how my physical and mental health could be linked	0.83	0.66
Item 8	Understand what my values are and how they could shape my actions	I now understand what my values are and how they could shape my actions	1	0.83
Item 9	Work through a specific problem	I worked through a specific problem	1	0.83
Item 10	Feel calmer	I feel calmer	0.86	0.83
Item 11	Talk about my story or my concerns with someone who is not judgemental	I have been able to talk about my story or my concerns with someone who is not judgemental	0.83	0.66
Item 12	Begin to understand unhelpful patterns of behaviour and how to change them	I have begun to understand unhelpful patterns of behaviour and how to change them	1	1
S-CVI			0.91	0.86

Real-world Contextualisation and Skills Learned

- Get access to some data
- Apply some NLP to those data



Responsible Research in the real world –

mental health data is sensitive – it is not just data **it is people**

Questions?