**DATA ANALYTICS CASE STUDY**

***(For mindline.sg Data Analyst Role)***

**Instructions for Completing the Case Study**

This section assesses your technical understanding and communication ability across data analytics, visualization, and research methodologies.  
You are required to answer ALL questions.

You have 48 hours from the time you receive the case study to complete it.

**Submission Requirements**

Please submit the following:

1. A slide deck summarizing your responses (.pptx or .pdf)
2. The complete code used for any analysis (.py, .R, or .ipynb)

**Interview Format**

Points to note in the case studies:

If selected for the next round, you will be invited to a panel interview with members of the mindline.sg team.

1. Your deck will be circulated to the mindline.sg team for reading beforehand.
2. The mindline.sg team includes individuals from both **technical and non-technical backgrounds**. Your presentation should be clear and accessible to a diverse audience, while still demonstrating technical rigour.
3. You will be given **15 minutes to present your findings**, followed by a **15 minute Q&A session** along with **30 minutes of general/ technical interview**.
4. The mindline.sg team will not be available for further clarification before the presentation. If you find any of the questions coached in overly broad terms, you must make reasonable and justifiable assumptions to narrow things down by demonstrating understanding of the domain
5. The mindline.sg team is interested to hear what can be done to further improve the analysis, given more time/ resources/ etc., if the problem described in the scenario surfaces again in the future. Be creative and ambitious in your scope.

**TECHNICAL QUESTIONS**

Use the data provided for question 1 and 2. You are free to use any other publicly available data needed for your analysis. Question 3 tests you on experiment design.

**Question 1 – Data Storytelling**

The [*let’s* talk](https://letstalk.mindline.sg/) forum has grown as a peer support community for mental health since its launch.

Using forum data (e.g., number of posts, content of posts, replies, toxic content removed, user return rate), craft a data story on how the platform has evolved in user engagement since 2022. Propose a data-driven way to increase engagement of the platform.

Present your findings in a blog-style write-up or a series of annotated visualizations suitable for general public understanding illustrating the impact of let’s talk as a mental health forum.

**Question 2 – Predictive Analytics**

Using historical [*let’s talk*](https://letstalk.mindline.sg/)user engagement data, propose a model to predict which users are most likely to drop off (churn) within the next 30 days.

* What features would you engineer from the platform logs? (e.g., session frequency, feature usage, recency)
* Outline your modelling approach (e.g., logistic regression, gradient boosting) and how you would evaluate model performance

**Question 3 – Experiment Design**

[mindline.sg](https://mindline.sg/) wants to test a new AI feature where user will be guided by a conversational AI through the platform. The goal is to improve user engagement and satisfaction. You are tasked to:

* Design an experiment to evaluate its impact on user engagement
* Specify the primary and secondary metrics, the sample size estimation, and the statistical test(s) you would use
* Briefly explain how you would ensure the validity of the experiment
* Explain how your insights could be used for product improvements

**Data dictionary**

1. **Posts**
2. **User**
3. **Topics**

**Data**

