Problem Statement (Hypothesis Formation)

How can tech companies identify which groups of employees are at highest risk for mental health challenges post-pandemic, predict their likelihood of disclosure, and design targeted wellness interventions—while safeguarding privacy and overcoming potential stigma—given that 62% of surveyed tech workers report experiencing mental health issues, yet only 41% say they would disclose them to their employer?

Context

The COVID-19 pandemic reshaped workplace norms in the tech industry, increasing remote and hybrid work but also amplifying stressors and feelings of isolation. According to the OSMI Mental Health in Tech Survey, **62% of respondents report having a mental health condition**, but disclosure rates remain low (**41%**). Tech employees often face high job demands and long hours, making them vulnerable to burnout and other challenges. While **over 70%** of companies offer some form of wellness benefit, HR leaders lack data-driven insight into which groups are most at risk and which factors influence disclosure. Without this, wellness programs risk being too generic, missing high-need groups.

Criteria for Success

- Build a disclosure prediction model with **≥75% accuracy**.
- Identify **3–5 actionable employee personas** based on risk and disclosure likelihood.
- Deliver interactive dashboards enabling HR teams to explore patterns by demographics, benefits, and workplace culture.

Scope of Solution Space

- Analyzing the **OSMI Mental Health in Tech Survey** (~1,100 rows × 120+ columns).
- Examining stigma perception, benefits offered, remote work impact, and role-related stressors.
- Applying supervised classification (e.g., Random Forest, Transformers) and unsupervised clustering.
- Providing anonymized, aggregated insights for HR and DEI leaders.

Constraints within Solution Space

- Small sample size may affect generalizability.
- Self-reported survey data introduces potential bias.
- Privacy requirements limit reporting granularity.
- Potential resistance to adopting recommendations due to stigma concerns.

Stakeholders to Provide Key Insight

- HR Directors (Policy & Benefits)
- DEI Leaders (Diversity & Inclusion Strategy)
- Wellness Officers / EAP Managers
- Data Privacy and Compliance Officers
- Internal mental health advocacy groups

Key Data Sources

- **OSMI Mental Health in Tech Survey** (Kaggle) – demographics, workplace culture, benefits access, mental health history.
- Internal HR metrics – role type, tenure, department size (aggregated).
- Academic and industry research on mental health in remote/hybrid tech environments.

PART 2

1. What is the problem you want to solve?

Many tech companies have expanded wellness programs post-pandemic, but **62%** of employees report experiencing mental health issues while only **41%** would disclose them to their employer. Without knowing **who is most at risk** and **what influences disclosure**, HR and DEI leaders risk implementing generic, less effective interventions.

2. Who is your client and why do they care about this problem?

Client: HR, DEI leaders, and wellness program managers in tech companies.
Why they care:

- They need to **allocate resources effectively** toward the groups most at risk.
- They want to **reduce stigma** and improve disclosure rates so employees can get support earlier.
- They must **justify wellness program budgets** with data-driven targeting.
  What they will decide based on the analysis:
- Which employee personas to prioritize for wellness interventions.

- What benefit offerings and communication strategies are most effective.
- How to track improvements in disclosure over time.

3. What data are you using? How will you acquire the data?

- **Primary source:** OSMI Mental Health in Tech Survey – Kaggle (~1,100 rows × 120+ columns).
- Data includes demographics, workplace culture, benefit access, and self-reported mental health history.
- **Acquisition:** Download from Kaggle via web or API.
- **Preparation:** Normalize categorical values (e.g., gender), encode survey responses, handle missing data, and create composite indexes (stigma index, benefits index).

4. Briefly outline how you'll solve this problem.

- **EDA:** Identify correlations between stigma perception, benefit access, work arrangement, and disclosure.
- **Modeling:**
    - Supervised classification (Random Forest, XGBoost, Transformer encoder-decoder with attention) to predict disclosure likelihood.
    - Unsupervised clustering (KMeans, hierarchical) to segment respondents into personas.
- **Optimization:** Use KerasTuner for hyperparameter tuning in neural models.
- **Visualization:** Build interactive dashboards (Tableau or Plotly Dash) for HR decision-making.
- **Recommendations:** Translate patterns into targeted, actionable policies.

5. What are your deliverables?

- **Code:** Clean, documented Jupyter notebooks for data cleaning, EDA, modeling, and visualization.

- **Report:** Written summary of findings, including personas, risk factors, and recommendations.

- **Slide deck:** Executive-friendly presentation with visual highlights and policy implications.

- **Dashboard:** Interactive tool allowing HR teams to explore anonymized insights.