**Data Science Lab  
Project Proposal Report**

**Lour Hakim – 323038091  
Shahar Rashdi – 314948415**

**Mental Health in U.S. High School Students:   
Exploring Risk Factors and Economic Context through YRBS and GDP Data**

**Background:** [92 words]

Mental health issues among U.S. adolescents have increased significantly in recent years, with a growing number of high school students reporting sadness, hopelessness, or suicidal thoughts. These outcomes are influenced by a range of behavioral, demographic, and environmental factors, often overlapping in complex ways. While many studies have explored individual risk factors such as bullying or sleep deprivation, fewer have examined how these interact or how broader socioeconomic context affects youth mental health. Understanding these relationships can support more effective intervention strategies and policy design targeting the well-being of vulnerable student populations.

**Problem:** [145 words]

Many existing studies focus on individual factors such as bullying or sleep when analyzing adolescent mental health. However, real-life risk factors often overlap. This fragmentation makes it difficult to understand the combined effects of behavioral, demographic, and environmental risks. Furthermore, few studies integrate broader economic context into mental health analysis. While state-level economic indicators like GDP are routinely used in adult health and policy research, their impact on youth mental health remains underexplored. The rich, large-scale YRBS dataset remains underutilized in this regard. Our project addresses these gaps by combining student-level behavioral and demographic data with state-level GDP. This allows us to explore not only individual predictors, but also how economic context may intensify or mitigate psychological risk. In particular, we aim to examine how mental health outcomes such as sadness or suicidal thoughts vary across states and how economic disparities might shape youth vulnerability.

**Aim & Research Questions:** [105 words]

This project aims to explore the relationship between mental health and a combination of personal risk factors and economic context among U.S. high school students in 2023. Our research questions are:

1. Which behavioral, demographic, and environmental factors are most strongly associated with poor mental health among U.S. high school students in 2023?
2. Can a predictive model that combines student-level risk factors with state-level GDP accurately identify U.S. high school students at high risk of mental health issues in 2023?
3. Is the likelihood of reporting suicidal thoughts or feelings of hopelessness higher among students in U.S. states with lower GDP per capita (2023)?

**Feasibility:** [177 words]

We use two publicly available datasets from 2023: the Youth Risk Behavior Survey (YRBS), which includes over 100 variables on student behaviors, health outcomes, demographics, and academic performance; and state-level GDP data from the U.S. Bureau of Economic Analysis. The YRBS dataset is available in Access and ASCII formats, while the GDP dataset is provided in Excel. Both are well-structured, well-documented.  
The YRBS sample includes thousands of high school students across the U.S., offering sufficient volume and diversity for multivariate analysis. We plan to begin with data cleaning and EDA, including descriptive statistics, handling of missing values, and visualizations. We then plan to develop classification models including logistic regression, decision trees, random forests, and support vector machines (SVM).  
One potential shortcoming is the presence of missing or inconsistent responses in the YRBS dataset, which could affect model performance. This will be addressed using imputation techniques and careful variable selection. Additionally, while GDP adds valuable economic context, it reflects macro-level conditions and cannot capture individual-level socioeconomic factors. We will interpret the findings accordingly and avoid overstating causal relationships.

**Data Sources and Accessibility**

<https://www.cdc.gov/yrbs/data/index.html>

Navigate to: Combined High School > 2023 > Access > District (ZIP)

<https://www.bea.gov/data/gdp/gdp-county-metro-and-other-areas>