**Project Scoping Document**

## **Name**

**Ismail Saleh**

## **Business Problem**

Many university students suffer from mental health problems. The health department at a certain university aims to locate students experiencing depression or stress to provide suitable care. This includes hiring enough psychiatrists and therapists and purchasing sufficient medications.

## **Business Impact**

This data could help outline and define the costs needed to treat students while also helping to predict and identify students who might be suffering from mental health issues. This will enable the university to provide care and support to these students in their journey toward better mental health.

## **Dataset(s)**

**Dataset 1:**

University Students' Mental Health “Malaysia”

**Strengths:**

The dataset provides sixteen key attributes about students’ academic life. In addition, the data is clean and ready to use.

**Weaknesses:**

The dataset consists of only 1000 rows, which are too few and need to use additional data to generalize the findings with higher confidence.

**Dataset 2:**

University Students Mental Health “Bangladesh”

**Strengths:**

With thirty-nine attributes, the dataset is rich in information about students and their health that may help us analyze their behavior and pin down the factors contributing to their mental health problems.

**Weaknesses:**

The dataset has only 1977 rows which might not be enough to generalize the findings with much confidence.

## **Methods**

* **Variables and comparisons**
* The first dataset consists of **16** variables:

1. **Timestamp**: the date the survey was taken.
2. **Gender**: either ‘male’ or ‘female’.
3. **Age**: between 18 and 25.
4. **Course**: Categorical variable, the major of the student.
5. **YearOfStudy**: Categorical variable, from ‘Year 1’ to ‘Year 4’.
6. **CGPA**: Cumulative Grade Point Average, ranging from 0.00 to 4.00.
7. **Depression**: Binary; 1: has depression, 0: doesn’t have depression.
8. **Anxiety**: Binary; 1: has anxiety, 0: doesn’t have anxiety.
9. **PanicAttack**: Binary; 1: has panic attacks, 0: doesn’t have panic attacks.
10. **SpecialistTreatment**: Binary; 1: has had a specialist treatment, 0: hasn’t had specialist treatment.
11. **SymptomFrequency\_Last7Days**: Count of days the student experienced any of the symptoms 0-7.
12. **HasMentalHealthSupport**: Binary; 1: has support, 0: doesn’t have support.
13. **SleepQuality**: The student is asked to rate the sleep quality from 1-5.
14. **StudyStressLevel**: The student is asked to rate the sleep quality from 1-5.
15. **StudyHoursPerWeek**: Number estimate of studying hours per week.
16. **AcademicEngagement**: The student is asked to rate the sleep quality from 1-5.

* The second dataset contains **39** variables, primarily composed of scaled questions related to stress, anxiety, and depression.

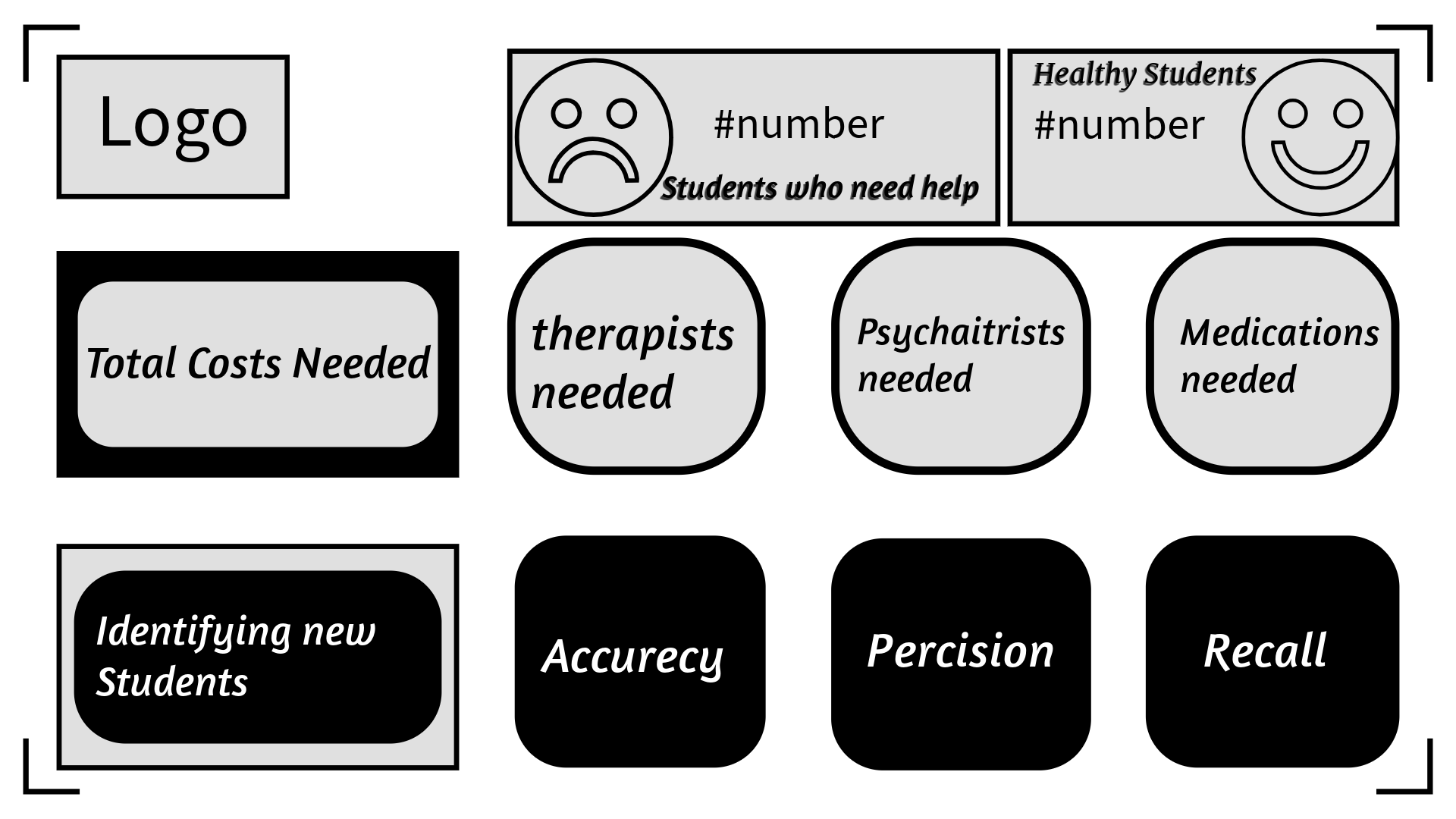
The columns are as follows:

1. **Age:** Five categories; “below 18”, “18-22”, “23-26”, “27-30”, “above 30”.
2. **Gender:** either ‘male’ or ‘female’.
3. **University:** name of the university.
4. **Department:** name of the department, e.g. “Engineering”.
5. **Academic Year:** Categorical variable, “FirstYear” to “FourthYear”.
6. **Current CGPA:** Categorical variable, from “below 2.5” to “3.80 – 4.00”.
7. **Scholarship:** Binary; “Yes” or “No”.
8. **Anxiety value:** Accumulative sum of 7 anxiety related numerical questions where the student answers an integer between 0-3 for each question.
9. **Anxiety label:** Categorical according to the anxiety value: Minimal, Mild, Moderate, Severe.
10. **Stress value:** Accumulative sum of 10 stress related numerical questions where the student answers an integer between 0-4 for each question.
11. **Stress label:** Categorical according to the stress value: Low, Moderate, High.
12. **Depression value:** Accumulative sum of 9 depression related numerical questions where the student answers an integer between 0-3 for each question.
13. **Depression label:** Categorical according to the depression value: No depression, Minimal, Mild, Moderate, Moderately severe, Severe.

* Expected relationships between variables:
* Sleep quality and anxiety.
* CGPA and stress level
* Year of study and depression

## **Dashboard**

Initial Dashboard Sketch:



## **Milestones**

* Identify characteristics of variables
* Find additional datasets
* Clean data
* Create visualizations
* Show relationships between variables
* Create dashboard
* Write final report

## **Timeline**

|  |  |
| --- | --- |
| Week | Tasks |
| Week 1 | * Explore initial dataset, and find additional datasets (if needed) * Clean data * Identify characteristics of variables * Envision final product * Feature Engineering |
| Week 2 | * Create visualizations * Show relationships between variables * Tabulation and Pivot Tables |
| Week 3 | * Create dashboard * Write final report |

Add additional weeks as necessary.