**Project Documentation: Mental Health Analysis of College Students Using MySQL**

**Project Overview:**

This project aims to analyze the mental health and lifestyle patterns of college students using **MySQL**. The primary goal was to explore correlations between academic performance, lifestyle habits, and mental health indicators such as anxiety, stress, and resilience. The project demonstrates **data management, cleaning, analysis, and insight generation** in a structured relational database.

**1. Dataset Overview**

The dataset consists of four main tables collected from college students:

1. **Students Table (students)**
   * student\_id, name, gender, age, year\_of\_study, department
   * Contains demographic information of students.
2. **Lifestyle Table (lifestyle)**
   * life\_id, student\_id, sleep\_hours, exercise\_hours, screen\_time\_hours, social\_media\_hours, study\_hours
   * Captures lifestyle patterns and daily habits.
3. **Academic Performance Table (academicperformance)**
   * perf\_id, student\_id, cgpa, attendance\_percent, academic\_stress
   * Provides academic metrics, including stress levels.
4. **Mental Health Table (mentalhealth)**
   * mh\_id, student\_id, anxiety\_score, stress\_level, resilience\_score, covid\_related\_anxiety, survey\_date
   * Captures mental health indicators and survey responses.
5. **Support System Table (supportsystem)**
   * support\_id, student\_id, family\_support, friends\_support, mentor\_support
   * Shows perceived support from family, friends, and mentors.

**2. Tools and Techniques Used**

* **Database:** MySQL 8.0+
* **Environment:** MySQL Workbench
* **Techniques Applied:**
  + Data cleaning: Handling NULLs, dropping unnecessary columns
  + Joins: Inner Join, Left Join
  + Aggregations: AVG(), COUNT()
  + Conditional columns using CASE WHEN
  + View creation for reusable queries (CREATE VIEW)
  + Categorization of continuous variables into groups
  + Correlation calculation using CORR()
  + Risk assessment and flagging of at-risk students

**3. Key Steps and Queries Implemented**

**3.1 Data Exploration**

* **Checked tables, record counts, and NULL values:**

SHOW TABLES;

SELECT COUNT(\*) FROM mentalhealth;

SELECT \* FROM lifestyle WHERE sleep\_hours IS NULL;

* **Observed basic statistics:**

SELECT student\_id, sleep\_hours, 'poor sleep quality' AS sleep\_comment

FROM lifestyle

WHERE sleep\_hours < 6;

**3.2 Combining Tables for Insights**

* **Academic performance + Mental health**

SELECT a.student\_id, a.cgpa, a.academic\_stress, m.anxiety\_score, m.resilience\_score

FROM academicperformance a

JOIN mentalhealth m ON a.student\_id = m.student\_id;

* **Flagging At-Risk Students**

CREATE OR REPLACE VIEW at\_risk\_students AS

SELECT a.student\_id, a.cgpa, a.academic\_stress, l.sleep\_hours, l.exercise\_hours,

CASE

WHEN a.cgpa < 6.5 OR a.academic\_stress='High' OR l.sleep\_hours < 6 OR l.exercise\_hours < 1

THEN 'At Risk' ELSE 'Normal' END AS risk\_status

FROM academicperformance a

JOIN lifestyle l ON a.student\_id = l.student\_id;

Querying the view:

SELECT \* FROM at\_risk\_students WHERE risk\_status = 'At Risk';

**3.3 Lifestyle Analysis**

* **Average sleep and study patterns by gender**

SELECT s.gender, COUNT(\*) AS student\_count,

AVG(l.sleep\_hours) AS avg\_sleep\_hours,

AVG(l.exercise\_hours) AS avg\_exercise\_hours,

AVG(l.screen\_time\_hours) AS avg\_screen\_time\_hours,

AVG(l.social\_media\_hours) AS avg\_social\_media\_hours,

AVG(l.study\_hours) AS avg\_study\_hours

FROM students s

JOIN lifestyle l ON s.student\_id = l.student\_id

GROUP BY s.gender;

* **Exercise vs Anxiety**

SELECT CASE

WHEN l.exercise\_hours >= 5 THEN 'High Exercise'

WHEN l.exercise\_hours BETWEEN 2 AND 4.9 THEN 'Moderate Exercise'

ELSE 'Low Exercise'

END AS exercise\_level,

AVG(m.anxiety\_score) AS avg\_anxiety,

COUNT(\*) AS student\_count

FROM mentalhealth m

JOIN lifestyle l ON m.student\_id = l.student\_id

GROUP BY exercise\_level;

* **Screen time vs Study Hours**

SELECT CASE

WHEN l.screen\_time\_hours >= 5 THEN 'High Screen Time'

WHEN l.screen\_time\_hours BETWEEN 2 AND 4.9 THEN 'Moderate Screen Time'

ELSE 'Low Screen Time'

END AS screen\_time\_level,

AVG(l.study\_hours) AS avg\_study\_hours,

COUNT(\*) AS student\_count

FROM students s

JOIN lifestyle l ON s.student\_id = l.student\_id

GROUP BY screen\_time\_level;

**3.4 Academic Performance and Mentor Support**

SELECT s.mentor\_support, AVG(a.cgpa) AS avg\_cgpa, COUNT(\*) AS student\_count

FROM academicperformance a

JOIN supportsystem s ON a.student\_id = s.student\_id

GROUP BY s.mentor\_support;

* Observed trends in **mentor support vs CGPA** to explore correlations.

**4. Insights Generated**

* Students with **less than 6 hours of sleep** were flagged as **poor sleep quality**, enabling early interventions.
* **High academic stress and high anxiety** students were flagged as **at-risk**, helping identify priority students.
* **Exercise hours** negatively correlated with **anxiety scores**, indicating consistent physical activity reduces anxiety.
* **Screen time** impacts study hours, with high screen time students showing lower study engagement.
* **Mentor support** is positively associated with higher CGPA, highlighting the role of guidance in academic success.

**5. Key Learnings and Skills Demonstrated**

* Hands-on experience in **relational database design and querying**.
* Data cleaning, handling NULLs, and dropping unnecessary columns.
* Using **JOINs, GROUP BY, CASE WHEN, views, and aggregate functions** to extract insights.
* Combining academic, lifestyle, and mental health data to **perform multi-dimensional analysis**.
* Ability to generate **actionable insights** for student mental health interventions.

**6. Conclusion**

This project demonstrates **how SQL can be used to analyze complex, multi-table datasets** and extract meaningful insights. The analysis of college students’ mental health, lifestyle, and academic performance highlights patterns that can guide **counseling, mentoring, and wellness programs**.

By implementing **views, aggregated queries, and conditional logic**, the project showcases **practical data analysis skills** applicable in educational and health data analytics.