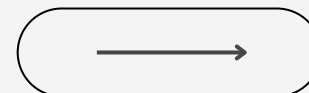


MCO1

GROUP E

Sean Gabriel M. Cardeno
Jeff Uriel V. Fonte
Aurelio Rodolfo F. Garcia
Raine Margaux A. Siongco



INTRODUCTION

With tensions of all kinds rising in almost every corner of the world, coupled with significant strides in psychology and neuroscience, the topic of Mental Health has been ubiquitous within the global public discourse.


In particular, as higher education has become more competitive and rigorous, the mental health of students has received notable attention.

Reports from agencies and experts worldwide consistently reach a common conclusion: students are struggling.

With all this in mind, the project aimed to thoroughly examine the available data from various sources and perform various queries to gain a better understanding of this phenomenon.



DATASETS



Mendeley Data

PHQ-9 Student Depression Dataset

Published: 1 July 2025 | Version 4 | DOI: 10.17632/kkzjk253cy4

Contributor: Md Abdullah ibne Aziz Miraz

Description

The PHQ-9 Student Depression Dataset contains responses from 400 students to the PHQ-9 questionn diagnosing depression. This dataset is designed to support the development of machine learning model detection by analyzing text responses to common depression-related questions.

The PHQ-9 questionnaire includes 9 questions that assess symptoms of depression over the past two we levels, sleep, appetite, and thoughts of self-harm. The responses are scored on a scale from 0 (Not at all) score ranging from 0 to 27. Based on this score, the depression severity is classified into one of the follo Minimal (0-4)
Mild (5-9)
Moderate (10-14)
Moderately Severe (15-19)
Severe (20-27)

This dataset is primarily designed for building models that can assist in automated depression detection Sentiment Analysis: Analyzing emotional tones in text responses to assess depression.
Text Classification: Classifying responses into different depression severity levels.
Predictive Modeling: Predicting depression severity based on textual responses.
Feature Engineering: Extracting linguistic features (e.g., sentiment, keywords) to predict depression. The dataset is diverse, with synthetic responses across different levels of depression, providing a versatile applications. While the dataset does not contain personally identifiable information (PII), real-world appli regarding privacy, consent, and mental health resources.
When working with real data or applying this dataset in clinical research, it is essential to adhere to ethic

Data Privacy: Anonymizing personal information.
Informed Consent: Ensuring participants give consent before data collection.

Contains 14 columns on age, gender, mental health questions, PHQ-9 scores, and depression level.

Student Depression Dataset.

Analyzing Mental Health Trends and Predictors Among Students

Data Card

Code (134)

Discussion (3)

Suggestions (0)

About Dataset

A student depression dataset typically contains data aimed at analyzing, understanding, and predicting de may include features such as demographic information (age, gender), academic performance (grades, att patterns, exercise, social activities), mental health history, and responses to standardized depression scal

These datasets are valuable for research in psychology, data science, and education to identify factors cc to design early intervention strategies. Ethical considerations like privacy, informed consent, and anonymi with such sensitive information.

1. File Structure

- Format: CSV format
- Rows: Each row represents an individual student.
- Columns: Each column represents a specific feature or attribute.

1. Columns

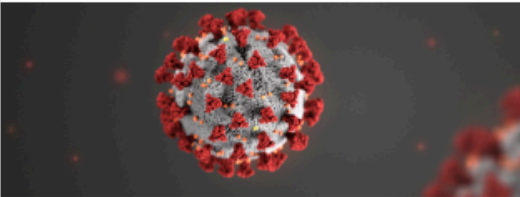
- ID: Unique identifier for each student.
- Age: Age of the student.

Has 18 columns on age, gender, geography, academic performance, sleep quality, and financial status.

Featured Content Using this Data

Mental Health

External Content



Data from NCHS' partnership with the U.S. Census Bureau on the Household Pulse Survey

Includes 14 columns on group, US states, time periods, confidence levels, and their quartile range.

Students' Social Media Addiction

A Cross-Country Survey of Usage Patterns, Academic Impact, and Relationship

Data Card

Code (65)

Discussion (2)

Suggestions (0)

About Dataset

OverviewThe Student Social Media & Relationships dataset contains anonymized records of students' social- It spans multiple countries and academic levels, focusing on key dimensions such as usage intensity dynamics. Each row represents one student's survey response, offering a cross-sectional snapshot s machine-learning applications.

Scope & Coverage

- Population: Students aged 16-25 enrolled in high school, undergraduate, or graduate programs
- Geography: Multi-country coverage (e.g., Bangladesh, India, USA, UK, Canada, Australia, Germ
- Timeframe: Data collected via a one-time online survey administered in Q1 2025.
- Volume: Configurable sample sizes (e.g., 100, 500, 1,000 records) based on research needs.

Data Collection & Methodology1. Survey Design: Questions adapted from validated scales on social-media addiction (e.g., Bergen t

Has 14 columns on age, gender, country, social media visits, sleep hours, mental health scores, relationship status, academic performance, and addicted scores.

Student Mental Health Crisis After COVID-19

Uncovering the Hidden Toll of COVID-19 on Student Mental Wellness

Data Card

Code (1)

Discussion (0)

Suggestions (0)

About Dataset

This dataset explores the mental health status of students post-COVID, with a focus on depression, academic pressure, a It has been meticulously cleaned, normalized, and prepared for data analysis with no missing values, making it ideal for E Analysis (EDA), visualizations, and machine learning tasks.

The dataset contains diverse demographic and psychological features that offer rich insights into how various factors cor mental well-being.

Key Features:
Gender – Biological gender of the student
Age – Age of the student (normalized)
City – City of residence
Profession – Current role (e.g., undergraduate, postgraduate, working)
Academic Pressure – Self-reported level of academic stress
Work Pressure – For working students, level of stress from job responsibilities
CGPA – Normalized cumulative grade point average
Study Satisfaction – How satisfied the student is with their academic performance
Job Satisfaction – Job contentment level (for employed respondents)
Sleep Duration – Average sleep in hours (numeric)

Why This Dataset?
No missing or null values — Fully cleaned & normalized
Comprehensive features — Includes both structured and unstructured features

View more

Contains 17 columns on age, gender, academic performance, work life, sleep hours, diet, mental health questions, and depression scores.

Student Performance & Behavior Dataset

Students Grading Analysis

Data Card

Code (131)

Discussion (9)

Suggestions (0)

About Dataset

Student Performance & Behavior Dataset

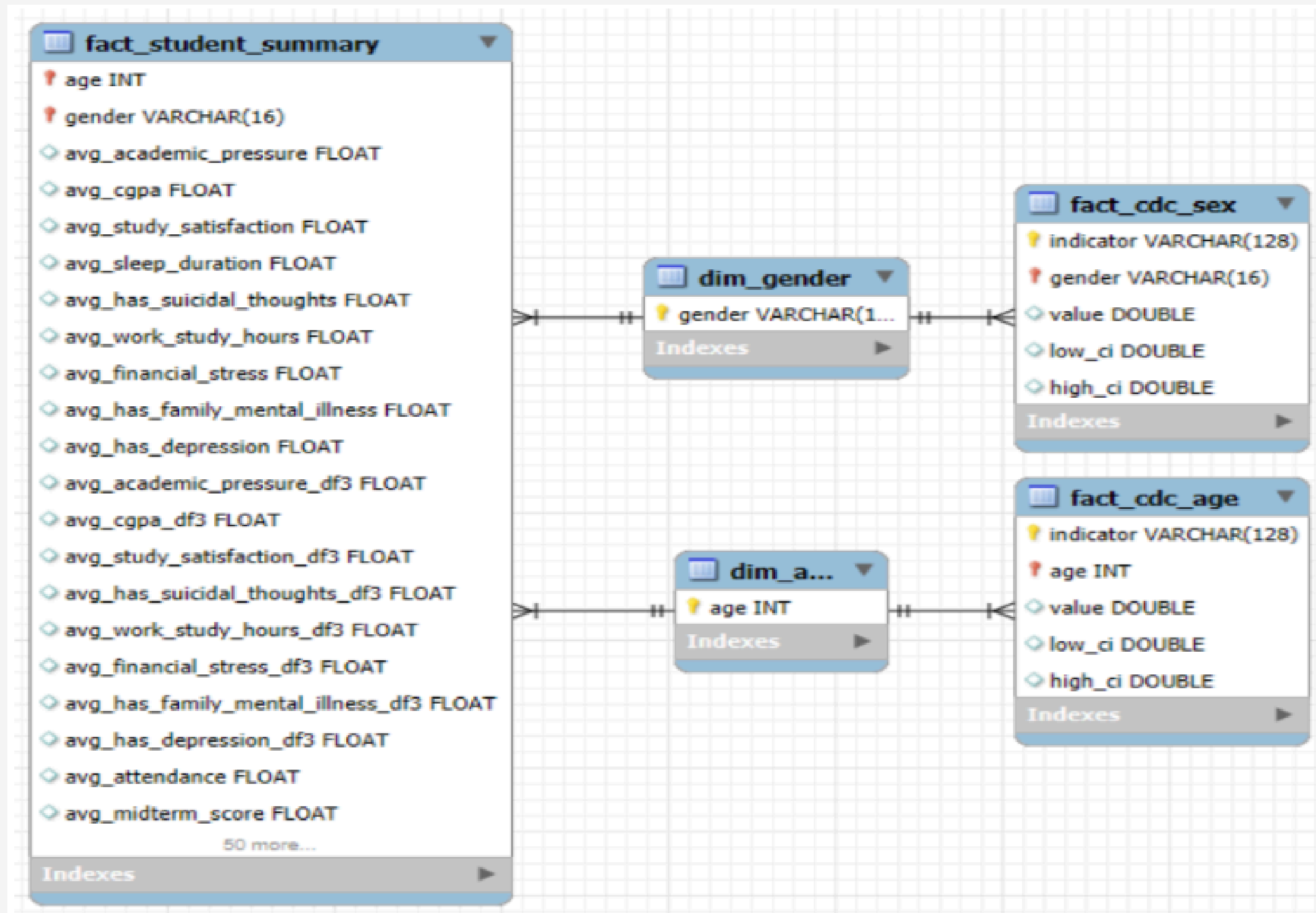
This dataset is real data of 5,000 records collected from a private learning provider. The dataset includes key attributes necessary for exploring patterns, correlations, and insights related to acad

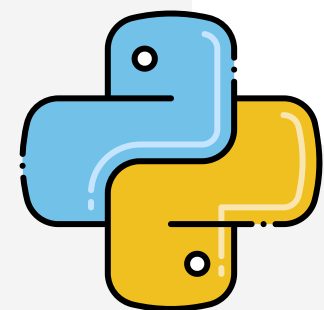
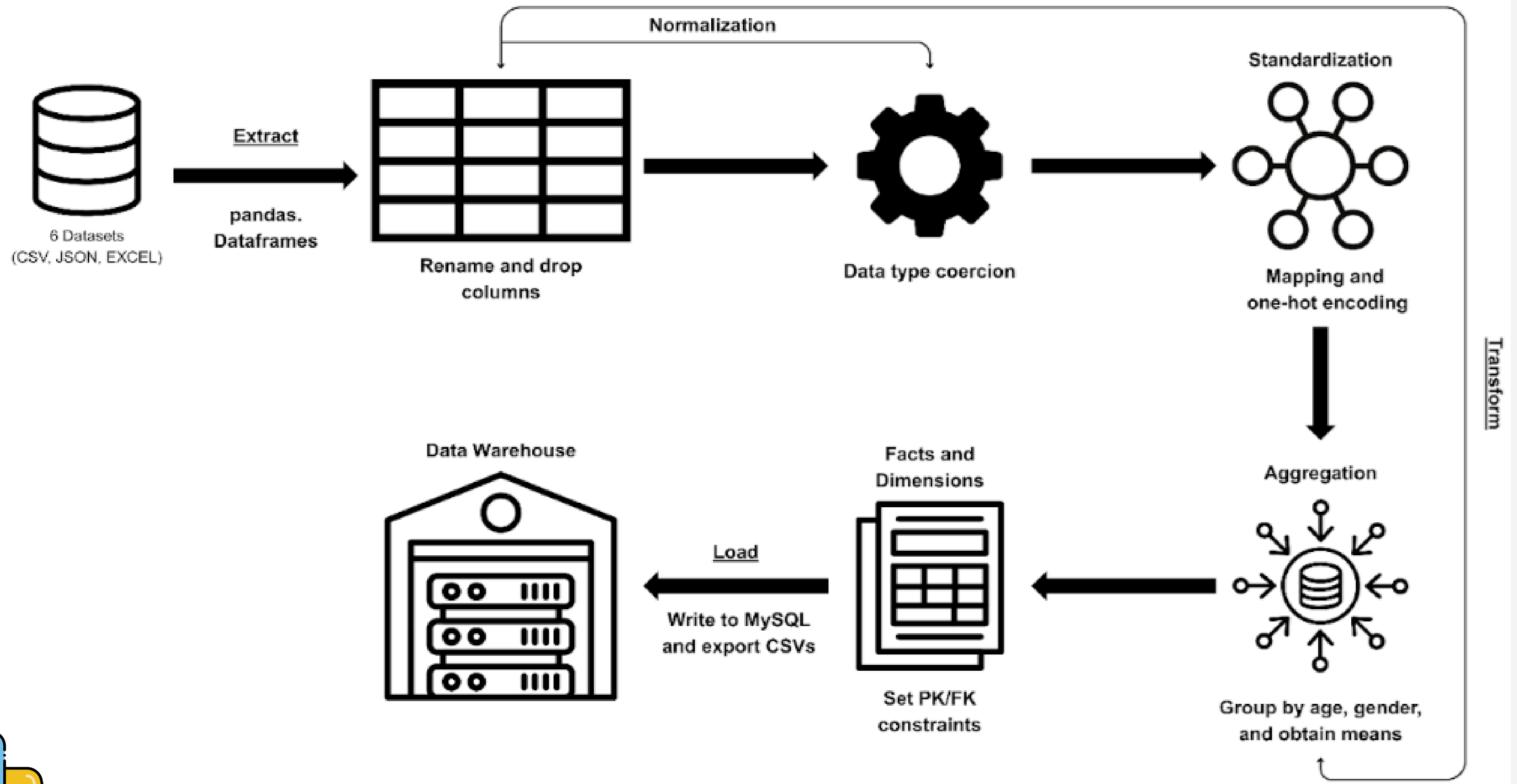
Columns:

- Student_ID: Unique identifier for each student.
- First_Name: Student's first name.
- Last_Name: Student's last name.
- Email: Contact email (can be anonymized).
- Gender: Male, Female, Other.
- Age: The age of the student.
- Department: Student's department (e.g., CS, Engineering, Business).
- Attendance (%): Attendance percentage (0-100%).
- Midterm_Score: Midterm exam score (out of 100).
- Final_Score: Final exam score (out of 100).

Has 23 columns on gender, age, academic performance, income level, sleep quality, and sleep hours.

SCHEMA





ETL WORKFLOW

OLAP Queries

NAME OF COMPANY

Rimberio co

Database used: MySQL

OLAP Operations Implemented:

- **ROLL-UP:** Summarizes data at multiple levels of aggregation (e.g., average stress by age group and gender).
- **DRILL-DOWN:** Provides more detailed insights by breaking down summarized data (e.g., viewing results from ALL to an age group).
- **SLICE:** Filters the data for a specific dimension (e.g., gender = 'Female').
- **DICE:** Selects subsets of data across multiple dimensions (e.g., specific age groups and stress levels).

Sample Queries

Example 1: Correlation between Sleep and Stress (ROLLUP)

```
SELECT
    gender,
    CASE
        WHEN age BETWEEN 18 AND 20 THEN '18-20'
        WHEN age BETWEEN 21 AND 23 THEN '21-23'
        ELSE '24+'
    END AS age_group,
    ROUND(AVG(avg_sleep_duration), 2) AS avg_sleep,
    ROUND(AVG(avg_stress_level), 2) AS avg_stress
FROM student_data
GROUP BY ROLLUP(gender, age_group);
```

Example 2: Slice and Dice

```
SELECT
    age_group,
    ROUND(AVG(avg_cgpa), 2) AS avg_cgpa
FROM student_data
WHERE gender = 'Female' AND age_group IN ('18-20', '21-23')
GROUP BY age_group;
```


OLAP Queries

Filters are all set to Gender = ALL, Age Group = ALL

OLAP Filters (Slice / Dice)

Gender: All Age Range: All

All

Male

Female

Age Range: All

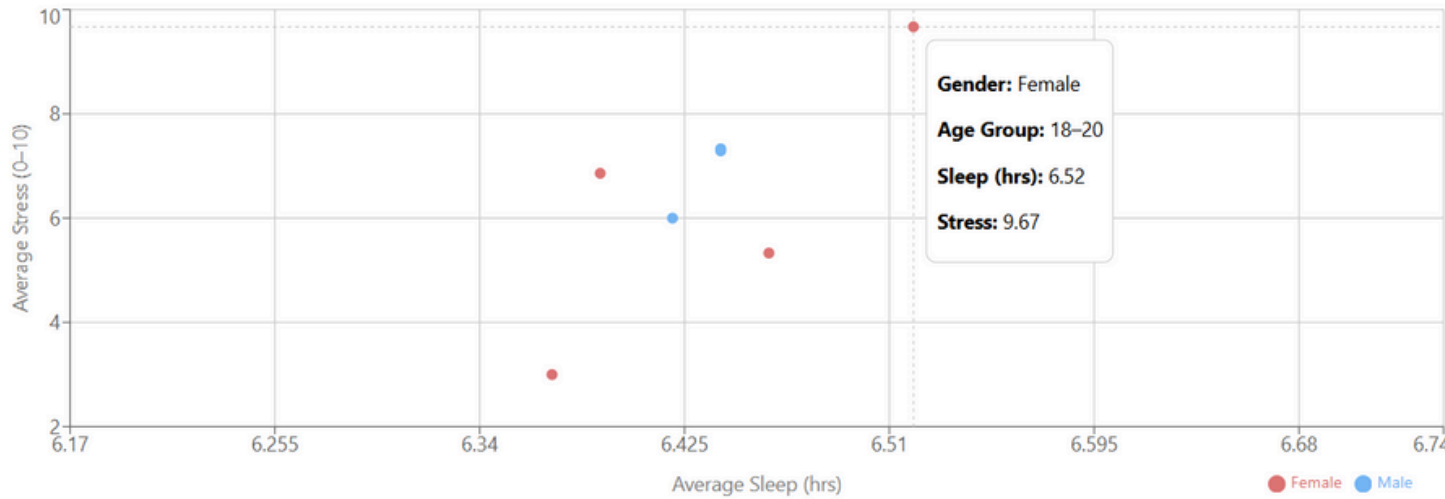
All

18-20

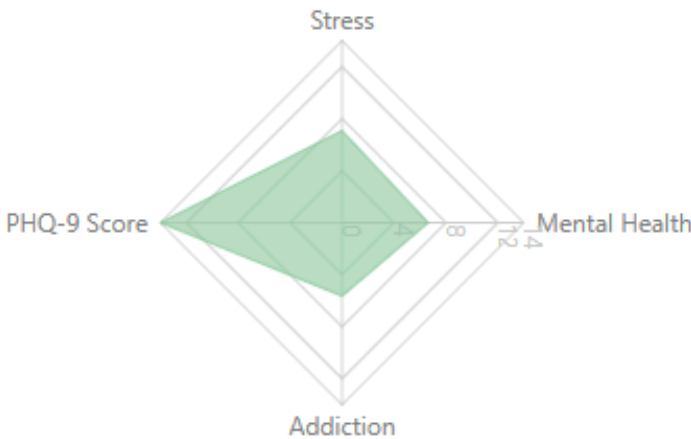
21-23

24+

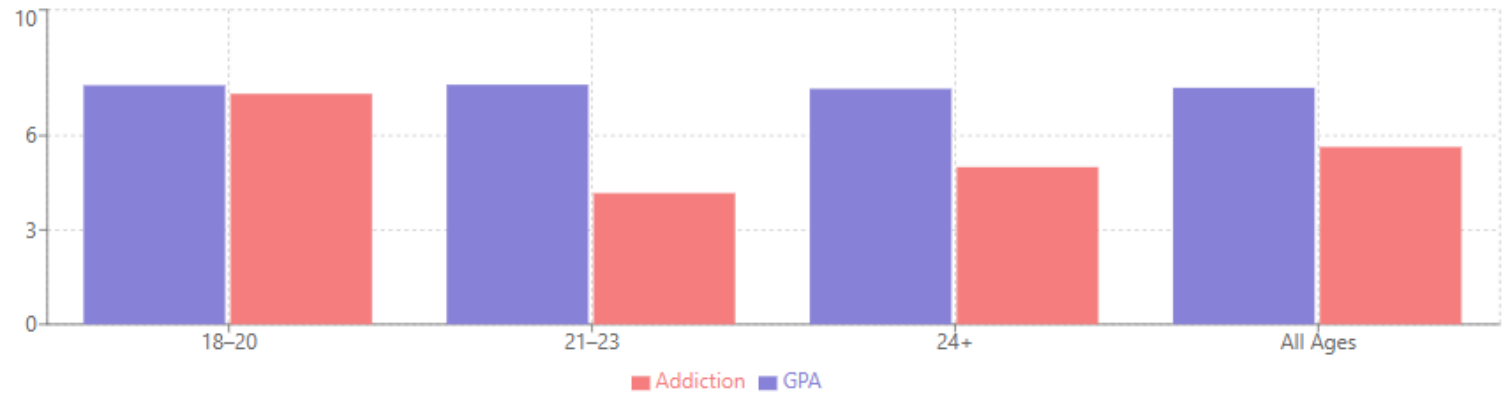
1. Correlation between Sleep and Stress



2. Mental Risk Indicators



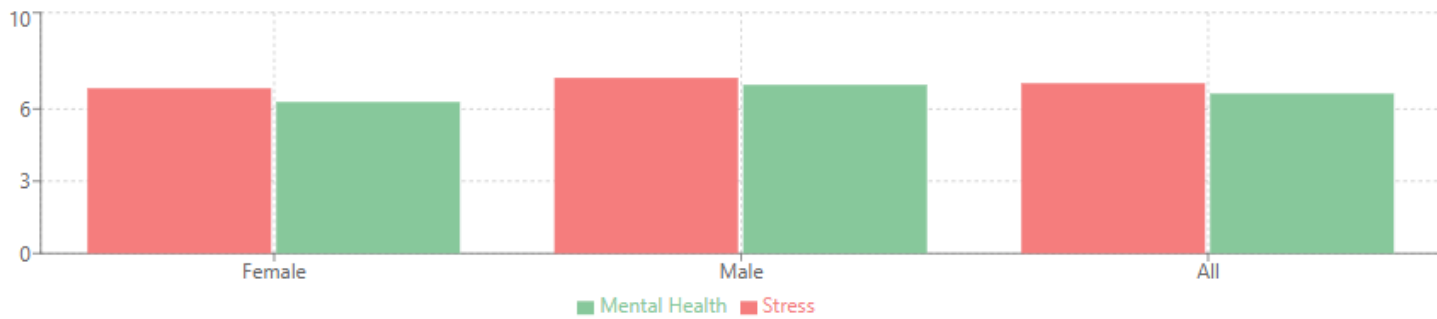
3. Social Media Addiction vs Academic Performance



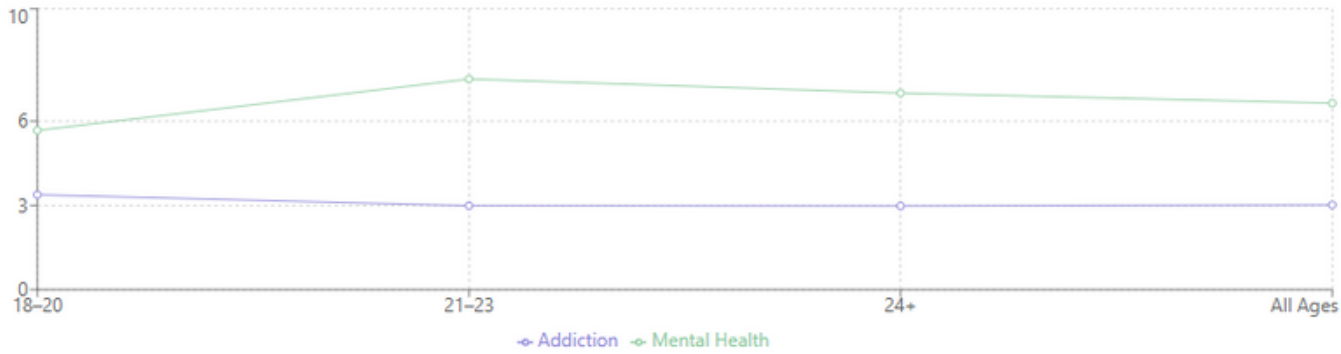
4. Gender & Age Group Comparison of Stress and Mental Health

(Note: Filters do not apply to this section)

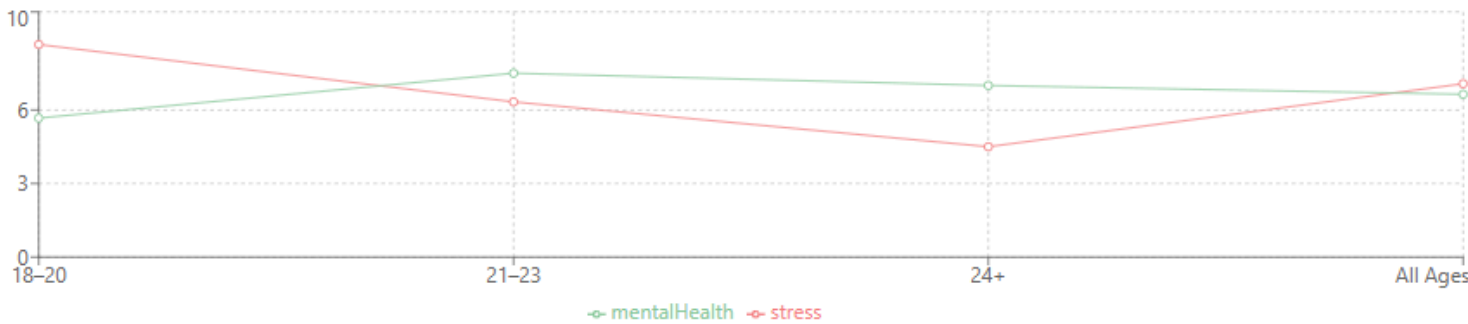
By Gender



5. Social Media Addiction and Mental Health Trend

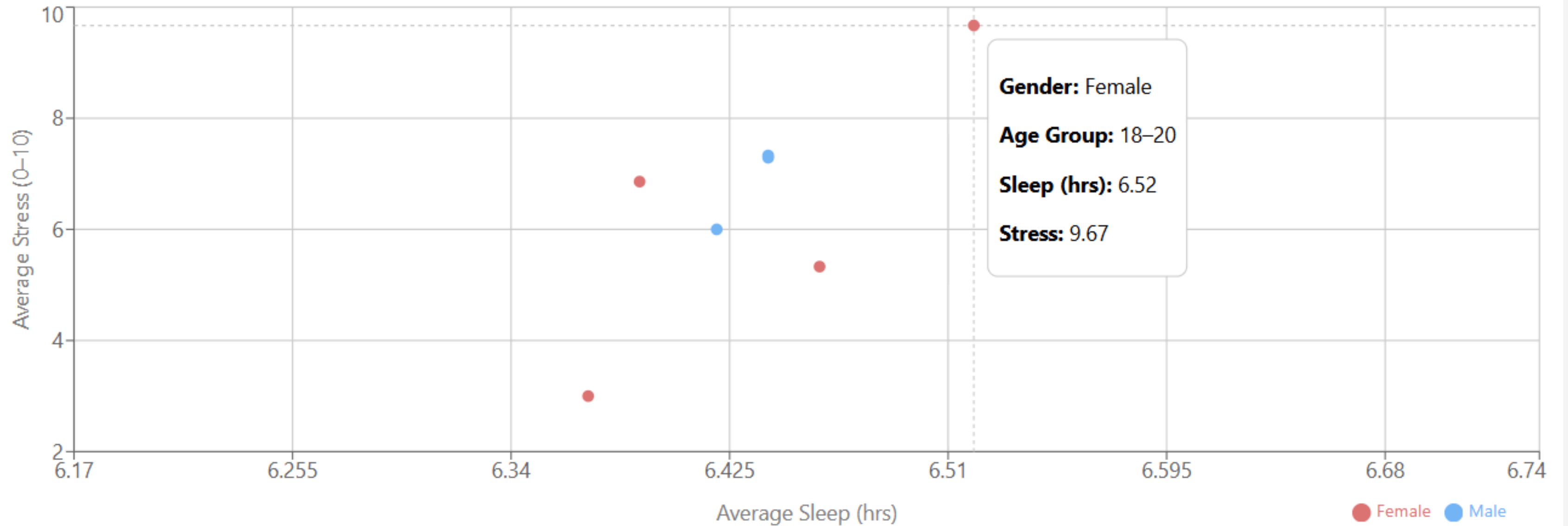


By Age Group



Query 1: Correlation Between Sleep and Stress

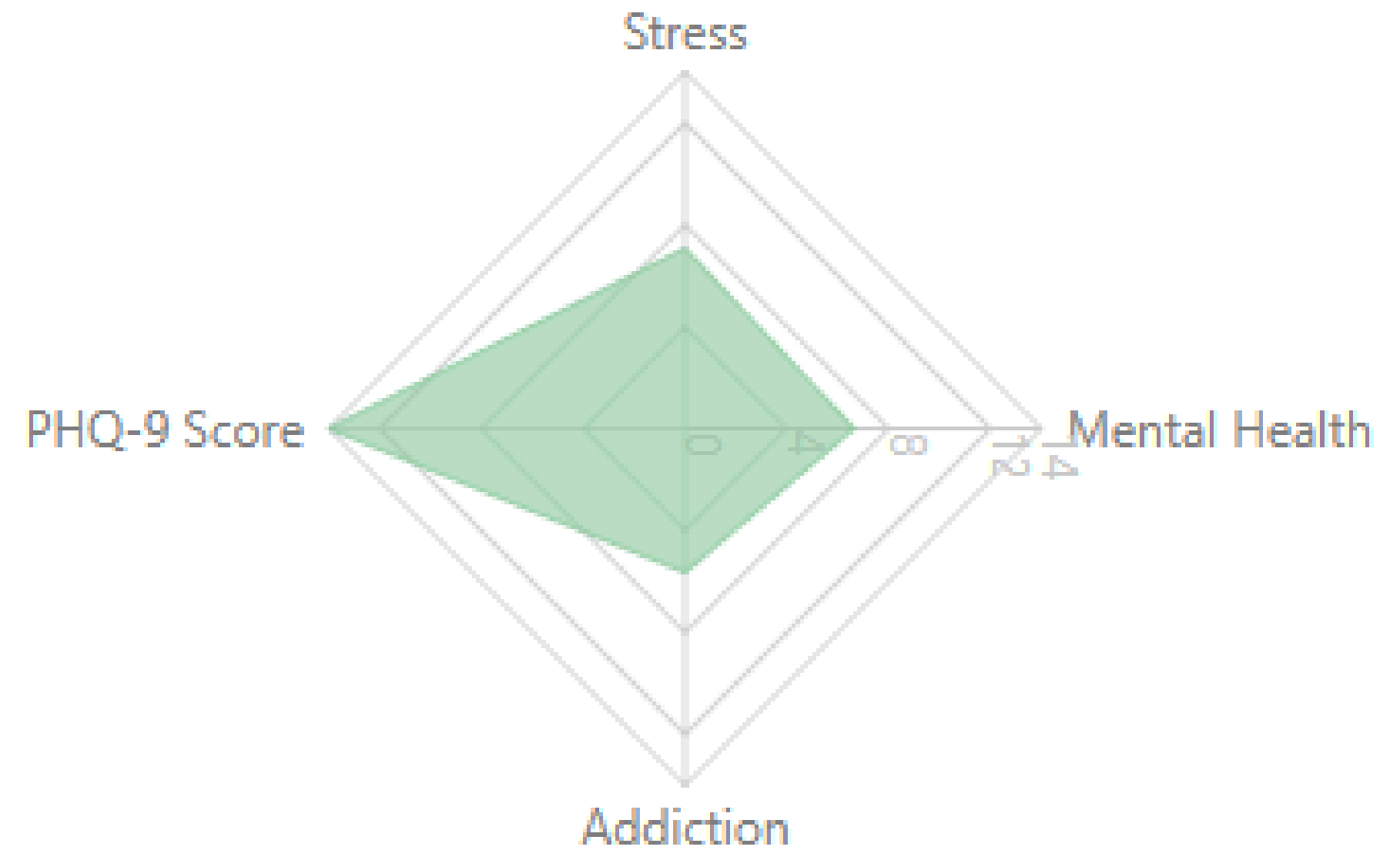
1. Correlation between Sleep and Stress



SORTED BY GENDER AGE GROUP

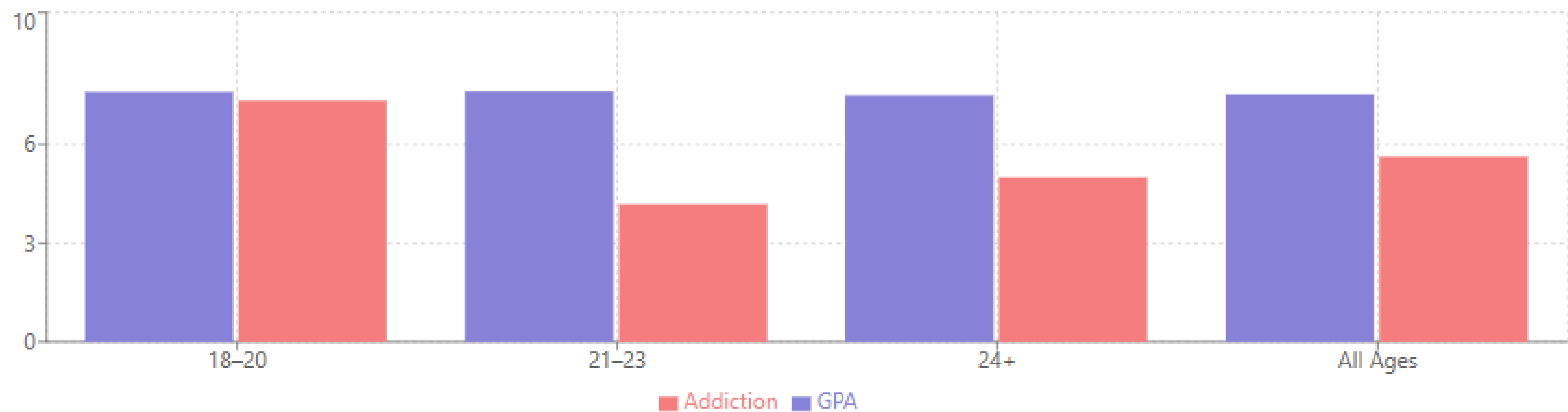
Query 2: Mental Risk Indicators

2. Mental Risk Indicators



Query 3: Social Media VS Academics

3. Social Media Addiction vs Academic Performance

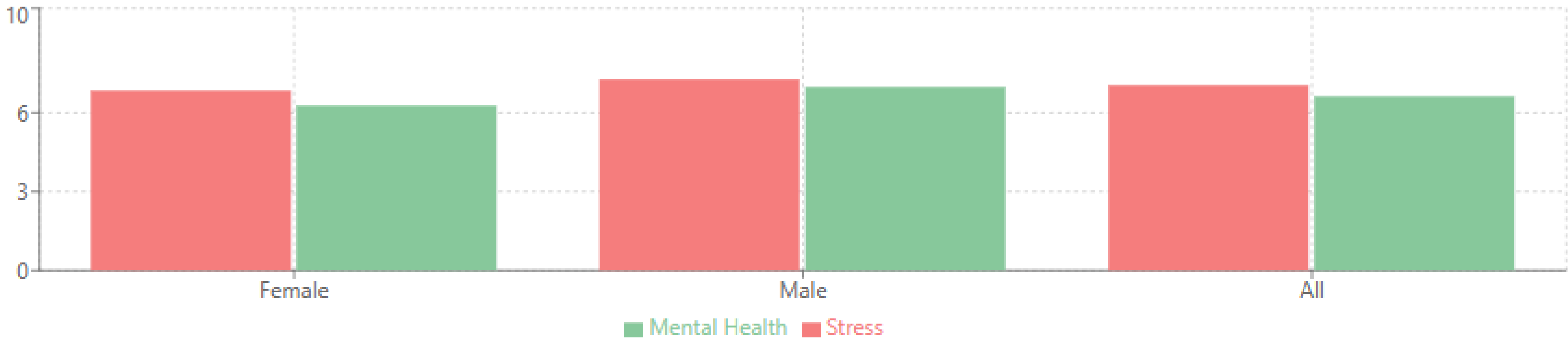


Query 4: Stress by Gender and Age Group

4. Gender & Age Group Comparison of Stress and Mental Health

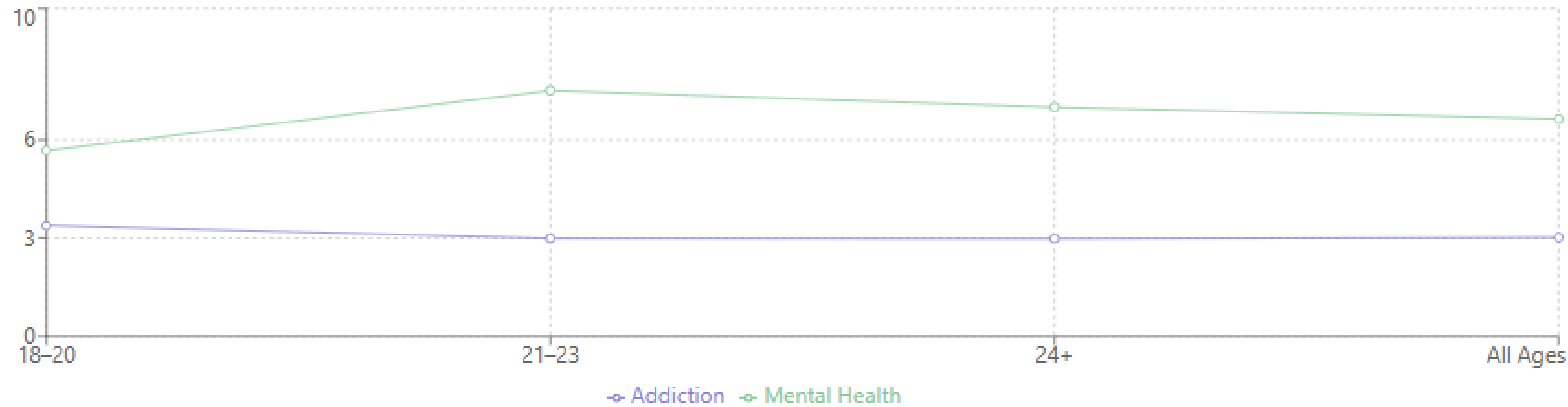
(Note: Filters do not apply to this section)

By Gender



Query 5: Social Media VS Mental Health

5. Social Media Addiction and Mental Health Trend



Query 6: Stress and Mental Health by Age Group



TESTING

```
C:\Users\nipot\Desktop\VSOCODE\STADMDB-MC01-GroupE\server>node functionTest.js

=== Testing with Gender=All, AgeRange=All ===
sleep-stress ✅ PASS – returned 9 rows
mental-health-indicators ✅ PASS – returned 4 rows
social-media-impact ✅ PASS – returned 4 rows
gender-age-comparison ✅ PASS – returned 1 rows
addiction-trend ✅ PASS – returned 4 rows

=== Testing with Gender=Female, AgeRange=18-20 ===
sleep-stress ✅ PASS – returned 1 rows
mental-health-indicators ✅ PASS – returned 4 rows
social-media-impact ✅ PASS – returned 1 rows
gender-age-comparison ✅ PASS – returned 1 rows
addiction-trend ✅ PASS – returned 1 rows

=== Testing with Gender=Male, AgeRange=21-23 ===
sleep-stress ✅ PASS – returned 1 rows
mental-health-indicators ✅ PASS – returned 4 rows
social-media-impact ✅ PASS – returned 1 rows
gender-age-comparison ✅ PASS – returned 1 rows
social-media-impact ✅ PASS – returned 1 rows
social-media-impact ✅ PASS – returned 1 rows
gender-age-comparison ✅ PASS – returned 1 rows
social-media-impact ✅ PASS – returned 1 rows
social-media-impact ✅ PASS – returned 1 rows
gender-age-comparison ✅ PASS – returned 1 rows
addiction-trend ✅ PASS – returned 1 rows

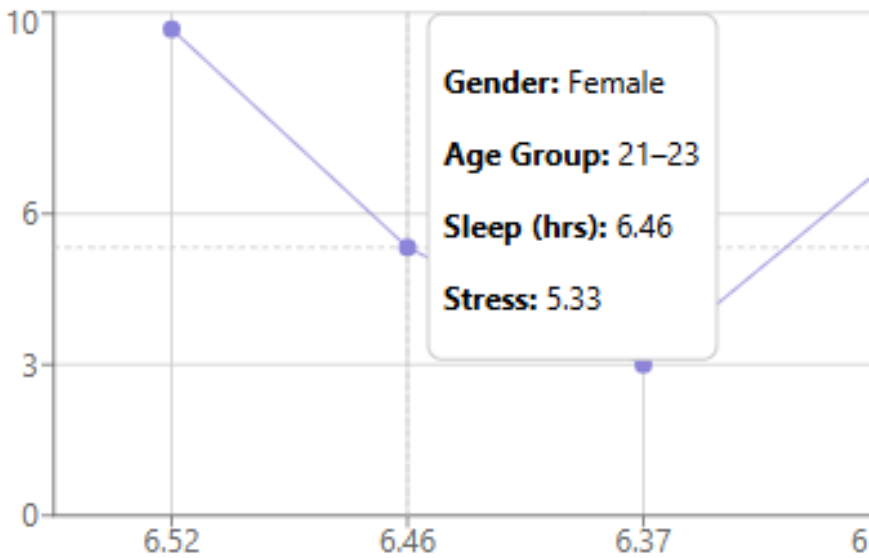
=== Testing with Gender=Female, AgeRange=24+ ===
sleep-stress ✅ PASS – returned 1 rows
mental-health-indicators ✅ PASS – returned 4 rows
social-media-impact ✅ PASS – returned 1 rows
social-media-impact ✅ PASS – returned 1 rows
gender-age-comparison ✅ PASS – returned 1 rows
addiction-trend ✅ PASS – returned 1 rows

=== Testing with Gender=Female, AgeRange=24+ ===
sleep-stress ✅ PASS – returned 1 rows
mental-health-indicators ✅ PASS – returned 4 rows
social-media-impact ✅ PASS – returned 1 rows
```

Testing API Endpoints

	gender	age_group	avg_sleep	avg_stress_level
►	Female	18–20	6.52	9.67
	Female	21–23	6.46	5.33
	Female	24+	6.37	3
	Female	NULL	6.39	6.86
	Male	18–20	6.54	7.67
	Male	21–23	6.44	7.33
	Male	24+	6.42	6
	Male	NULL	6.44	7.29
	NULL	NULL	6.42	7.07

1. Correlation between Sleep and Stress



Testing Frontend Using Filters

5.1 Functional Testing

Test Case	Query	Expected output	Actual Result	Status
Validate dataset extraction	Load all datasets	All loaded with no file errors	Loaded	Pass
Validate data and preprocess	Apply cleaning pipeline	Cleaned with renamed columns and no missing values	All cleaned	Pass
Validate MySQL load	Write fact and dim tables to olap_dashboard	Tables created with Primary Keys	Verified locally in Workbench	Pass
Test API endpoints	Run each endpoints	Returns JSON response with no SQL errors	Returns rows/data	Pass
Test Frontend using filters	Change gender/age filters	Dashboard dynamically updates graphs and summaries	Behavior matched manual SQL validation in Workbench	Pass

Test Script

QUERY PROCESSING, OPTIMIZATION

```
8. Postprocess fact_student_summary.

group_cols = ["age", "gender"]

# (A) Numeric non-onehot aggregation
numeric_summary = base.groupby(group_cols, as_index=False)[numeric_non_onehot].mean()

# (B) One-hot aggregation
if one_hot_cols:
    onehot_summary = base.groupby(group_cols, as_index=False)[one_hot_cols].mean()
    fact_student_summary = pd.merge(numeric_summary, onehot_summary, on=group_cols, how='outer')
else:
    fact_student_summary = numeric_summary

# Rename numeric aggregates
rename_map = {c: ("avg_" + c) for c in numeric_non_onehot}
fact_student_summary.rename(columns=rename_map, inplace=True)

# Keep valid ages only
fact_student_summary = fact_student_summary[
    fact_student_summary["age"].between(AGE_MIN, AGE_MAX)
].sort_values(["age", "gender"]).reset_index(drop=True)

# Rename aggregates to show they are group-level means
rename_map = {c: ('avg_' + c) for c in numeric_non_onehot}
fact_student_summary.rename(columns=rename_map, inplace=True)

# one-hot columns represent fractions/proportions
fact_student_summary = fact_student_summary.sort_values(['age', 'gender']).reset_index(drop=True)

print("Fact student summary shape:", fact_student_summary.shape)
print(fact_student_summary.head())

✓ 0.1s
```

Snippet of the ETL script creating and finalizing the fact_student_summary

Field	Type	Null	Key	Default	Extra
age	int	NO	PRI	NULL	
gender	varchar(16)	NO	PRI	NULL	
avg_academic_pressure	float	YES		NULL	
avg_cgpa	float	YES		NULL	
avg_study_satisfaction	float	YES		NULL	
avg_sleep_duration	float	YES		NULL	
avg_has_suicidal_thoughts	float	YES		NULL	
avg_work_study_hours	float	YES		NULL	
avg_financial_stress	float	YES		NULL	
avg_has_family_mental_illness	float	YES		NULL	
avg_has_depression	float	YES		NULL	
avg_academic_pressure_df3	float	YES		NULL	
avg_cgpa_df3	float	YES		NULL	
avg_study_satisfaction_df3	float	YES		NULL	
avg_has_suicidal_thoughts_df3	float	YES		NULL	
avg_work_study_hours_df3	float	YES		NULL	

Primary Keys, fact_student_summary Table

Query Execution Times				
0.00035450	SELECT	ROUND(AVG(avg_stress_level), 2) AS avg_stress,		
0.00045075	SELECT	CASE WHEN age BETWEEN 18 AND 20 THEN '1		
0.00059200	SELECT	gender, ROUND(AVG(avg_stress_level), 2) AS av		
0.00041950	SELECT	CASE WHEN age BETWEEN 18 AND 20 THEN '		
0.00047475	SELECT	CASE WHEN age BETWEEN 18 AND 20 THEN '1		
0.00047550	SELECT	gender, CASE WHEN age BETWEEN 18 AND '		

Query	Execution Time
Sleep-Stress	0.00059300
Mental Health Indicators	0.00040500
Social Media Impact	0.00042275
Gender-Age Comparison of Stress and Mental Health	0.00051650
	0.00045200
Social Media Addiction and Mental Health	0.00053050

Conclusions

The project workflow follows the instructions from the machine specifications, where each member is assigned a major part to distribute the overall workload.

The process of building and maintaining a data warehouse allows developers to analyze and combine datasets. It reveals significant factors, such as, in this case, mental health risk indicators amongst different student groups, such as PHQ-9 ratings, stress levels, and addiction scores.

The datasets used for the project, which had the psychiatric records of collectively thousands of individuals, would be extremely difficult to perform regular SQL Queries on.