MCO1

GROUP E

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

INTRODUCTION

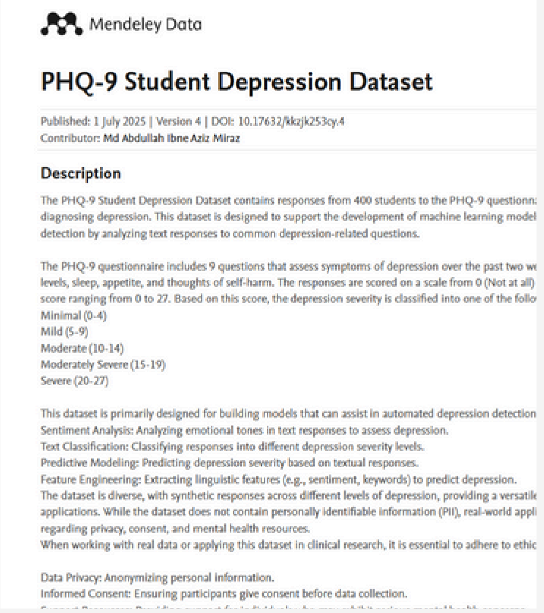
With tensions of all kinds rising in almost every corner of the world, coupled with significant strides in psychology and neuroscience, the topic of Mental Health has been ubiquitous within the global public discourse.

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

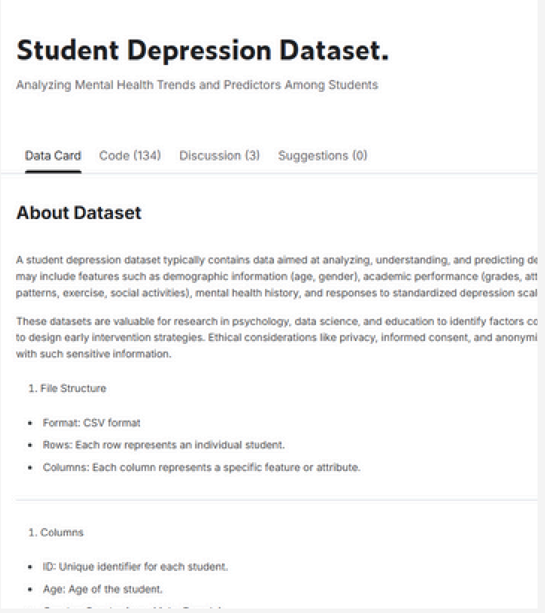
Reports from agencies and experts worldwide consistently reach a common conclusion: students are struggling.

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.

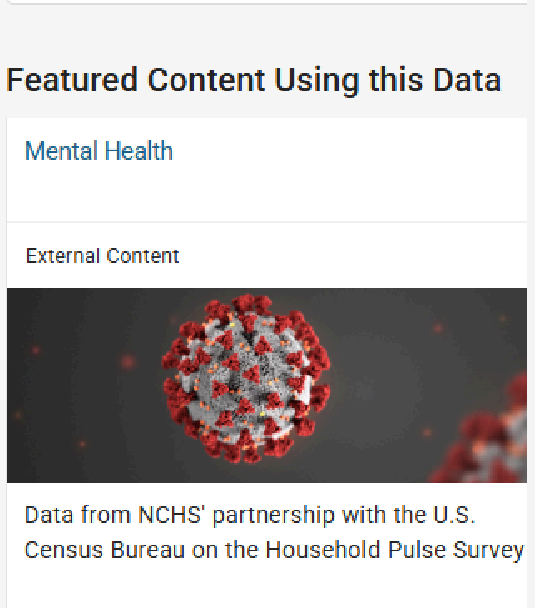
DATASETS



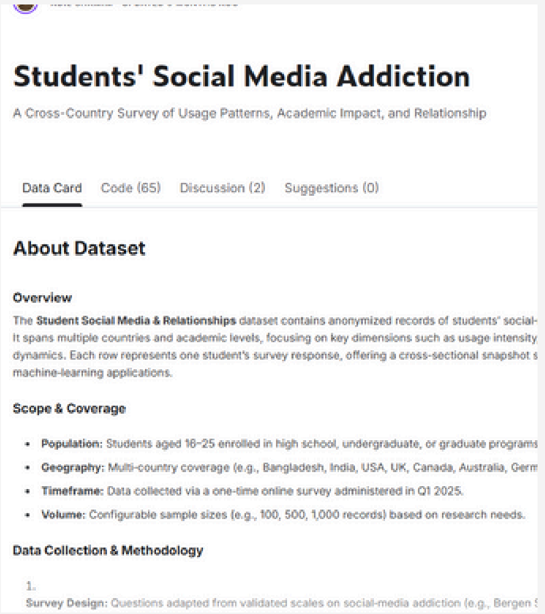
Contains 14 columns on age, gender, mental health questions, PHQ-9 scores, and depression level.



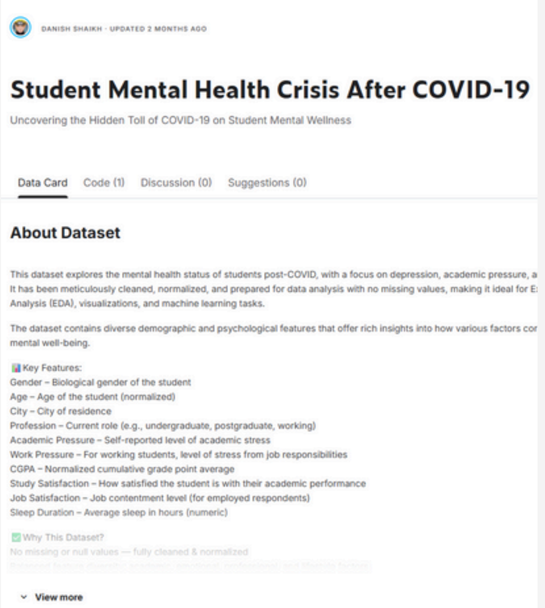
Has 18 columns on age, gender, geography, academic performance, sleep quality, and financial status.



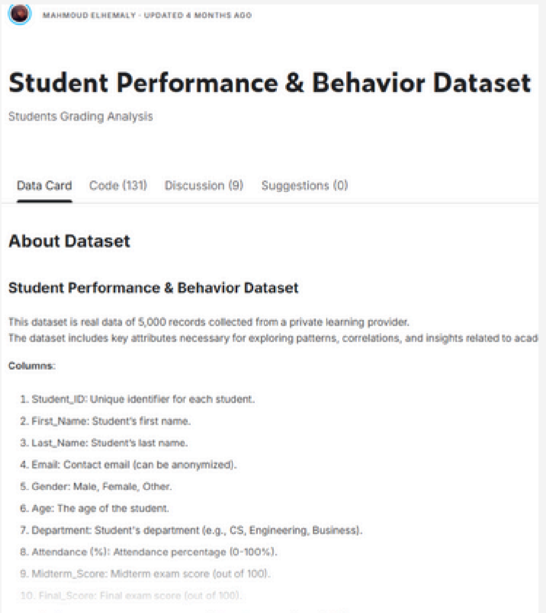
Includes 14 columns on group, US states, time periods, confidence levels, and their quartile range.



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

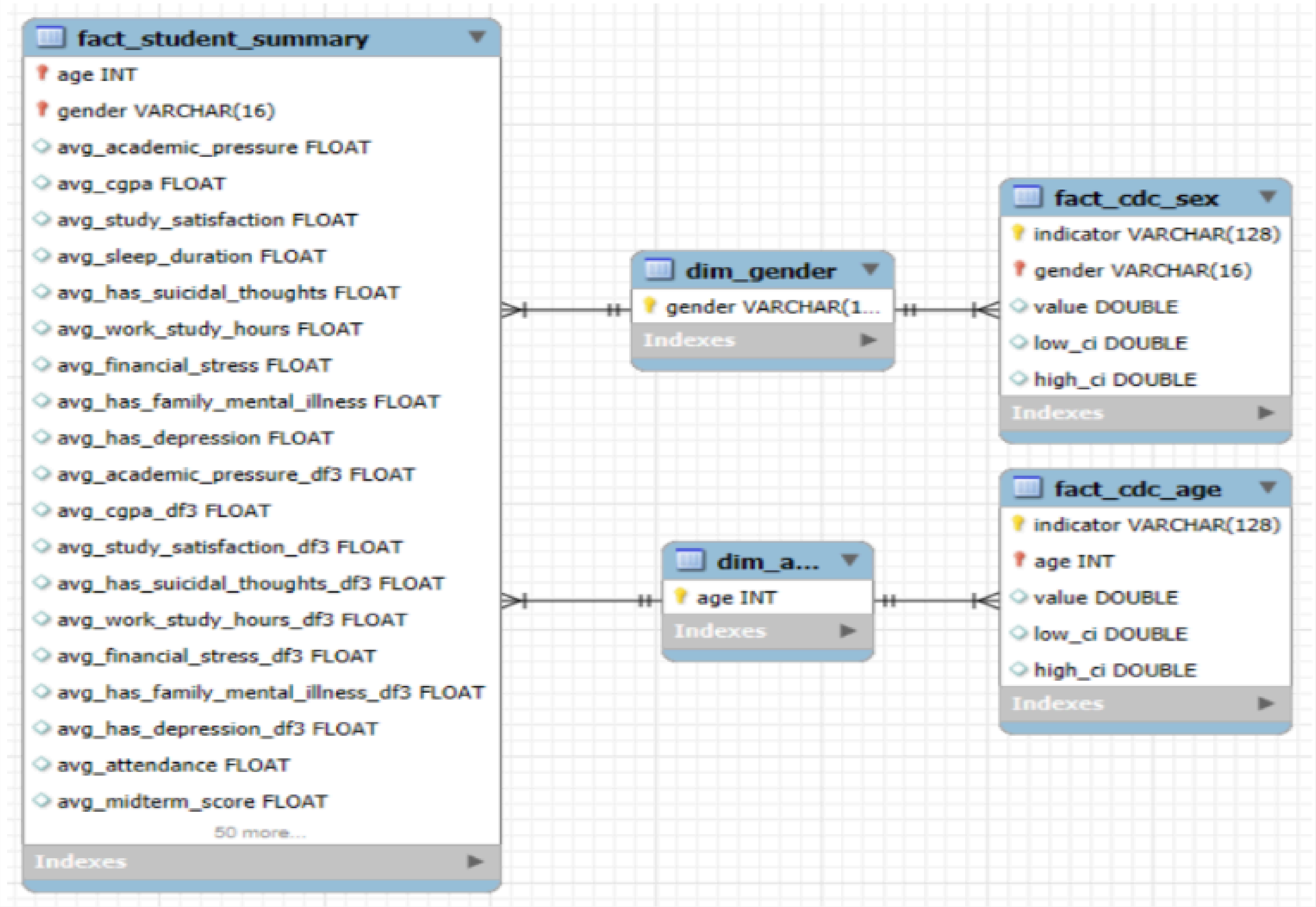


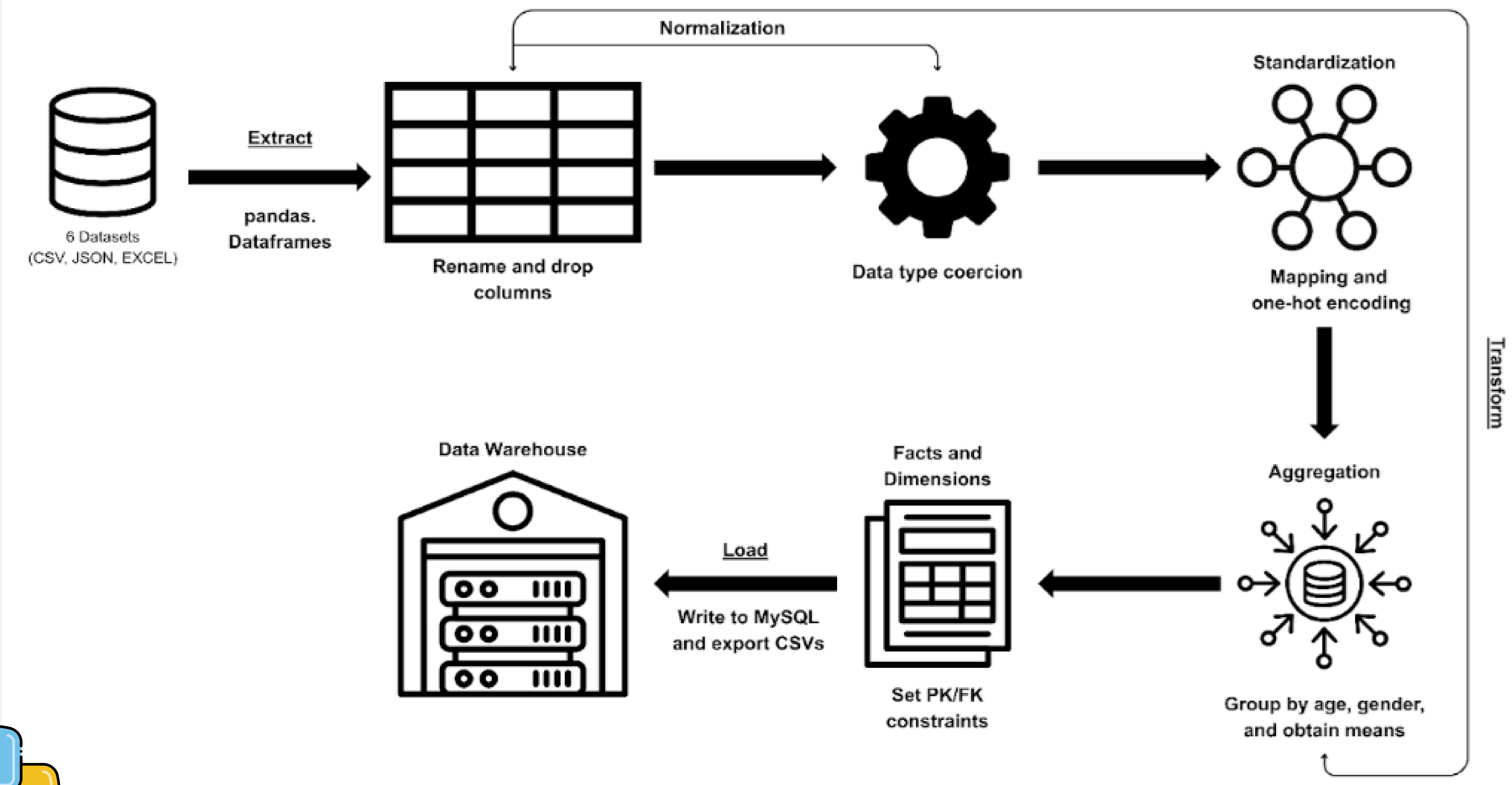
Contains 17 columns on age, gender, academic performance, work life, sleep hours, diet, mental health questions, and depression scores.



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

SCHEMA


fact_student_summary
age INT
gender VARCHAR(16)
avg_academic_pressure FLOAT
avg_cgpa FLOAT
avg_study_satisfaction FLOAT
avg_sleep_duration FLOAT
avg_has_suicidal_thoughts FLOAT
avg_work_study_hours FLOAT
avg_financial_stress FLOAT
avg_has_family_mental_illness FLOAT
avg_has_depression FLOAT
avg_academic_pressure_df3 FLOAT
avg_cgpa_df3 FLOAT
avg_study_satisfaction_df3 FLOAT
avg_has_suicidal_thoughts_df3 FLOAT
avg_work_study_hours_df3 FLOAT
avg_financial_stress_df3 FLOAT
avg_has_family_mental_illness_df3 FLOAT
avg_has_depression_df3 FLOAT
avg_attendance FLOAT
avg_midterm_score FLOAT
50 more...
Indexes
dim_gender
gender VARCHAR(1...
Indexes
fact_cdc_sex
indicator VARCHAR(128)
gender VARCHAR(16)
value DOUBLE
low_ci DOUBLE
high_ci DOUBLE
Indexes
dim_a...
age INT
Indexes
fact_cdc_age
indicator VARCHAR(128)
age INT
value DOUBLE
low_ci DOUBLE
high_ci DOUBLE
Indexes



ETL WORKFLOW

OLAP Queries

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)

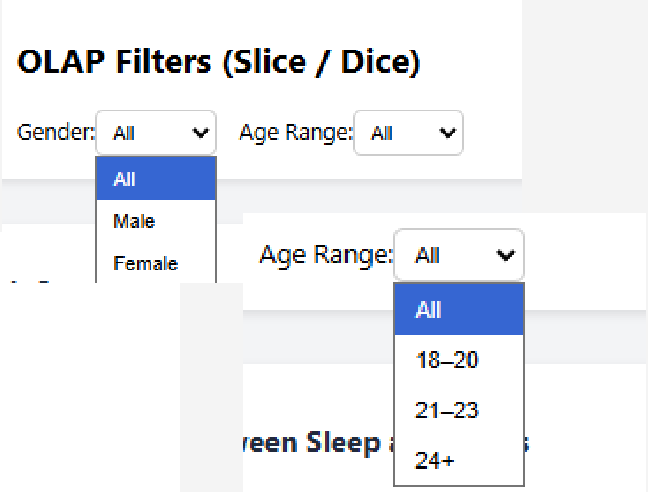
```
SELECT
    gender,
    CASE
        WHEN age BETWEEN 18 AND 20 THEN '18–20'
        WHEN age BETWEEN 21 AND 23 THEN '21–23'
        ELSE '24+'
    END AS age_group,
    ROUND(AVG(avg_sleep_duration), 2) AS avg_sleep,
    ROUND(AVG(avg_stress_level), 2) AS avg_stress
FROM student_data
GROUP BY ROLLUP(gender,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

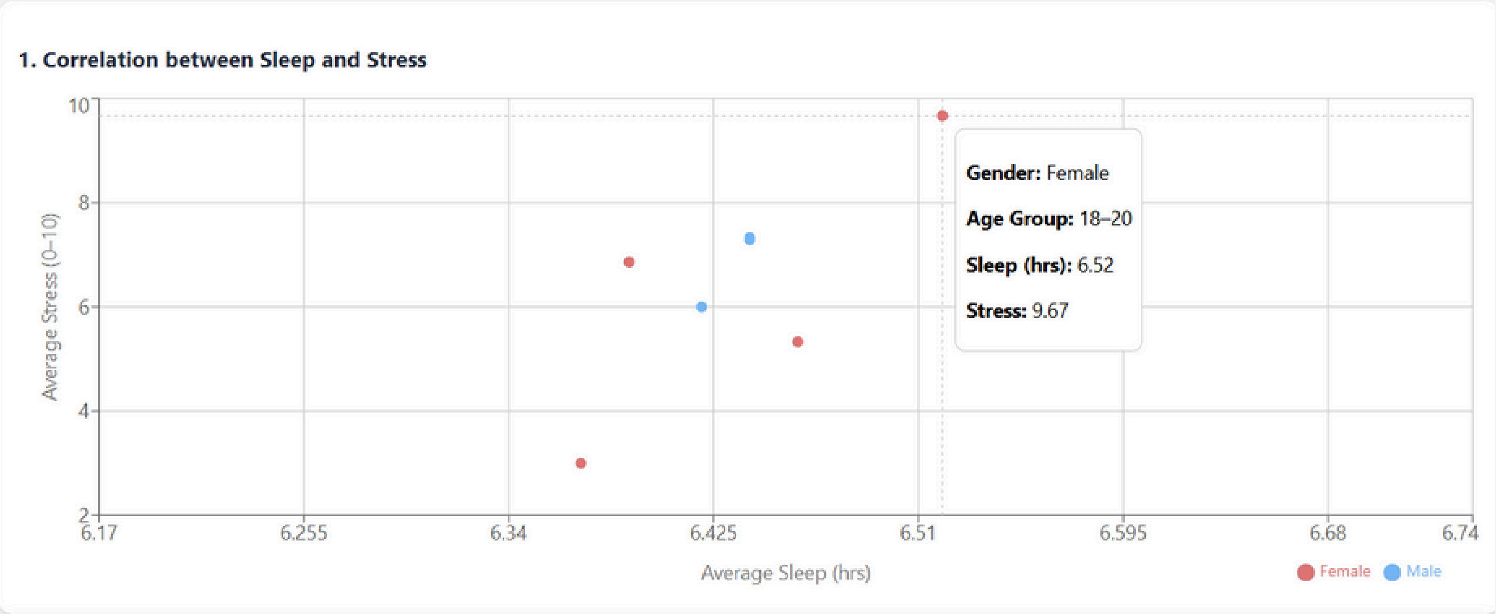
Example 2: Slice and Dice

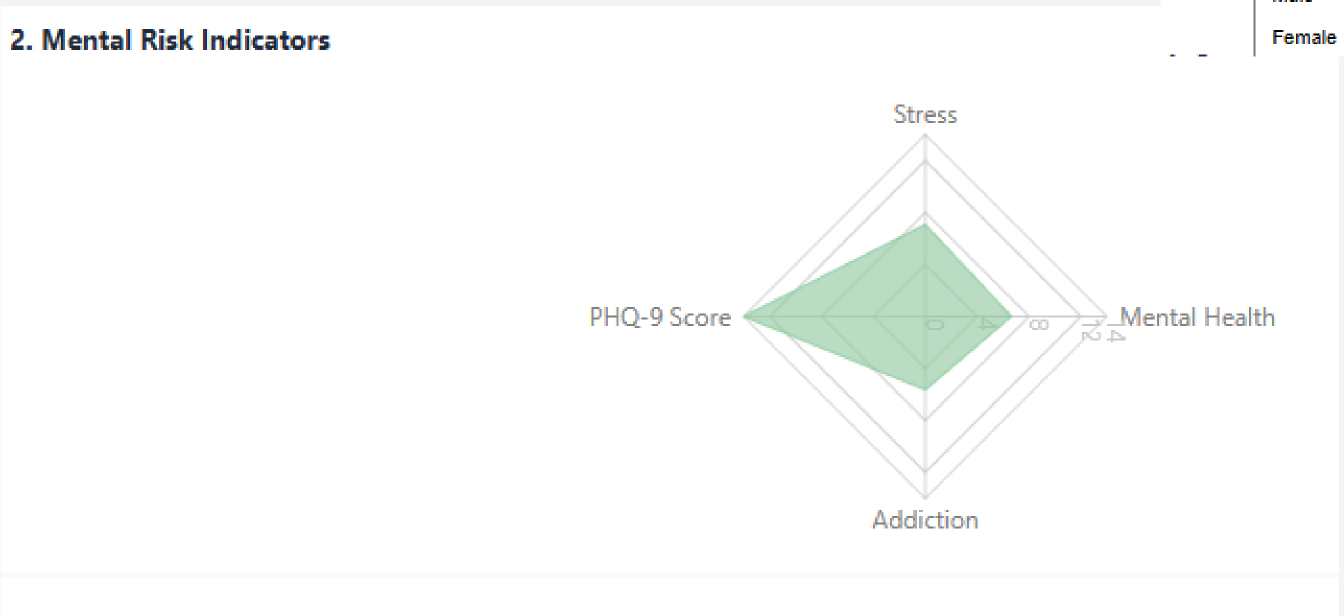
```
SELECT
    age_group,
    ROUND(AVG(avg_cgpa), 2) AS avg_cgpa
FROM student_data
WHERE gender = 'Female' AND age_group IN ('18–20', '21–23')
GROUP BY age_group;
```

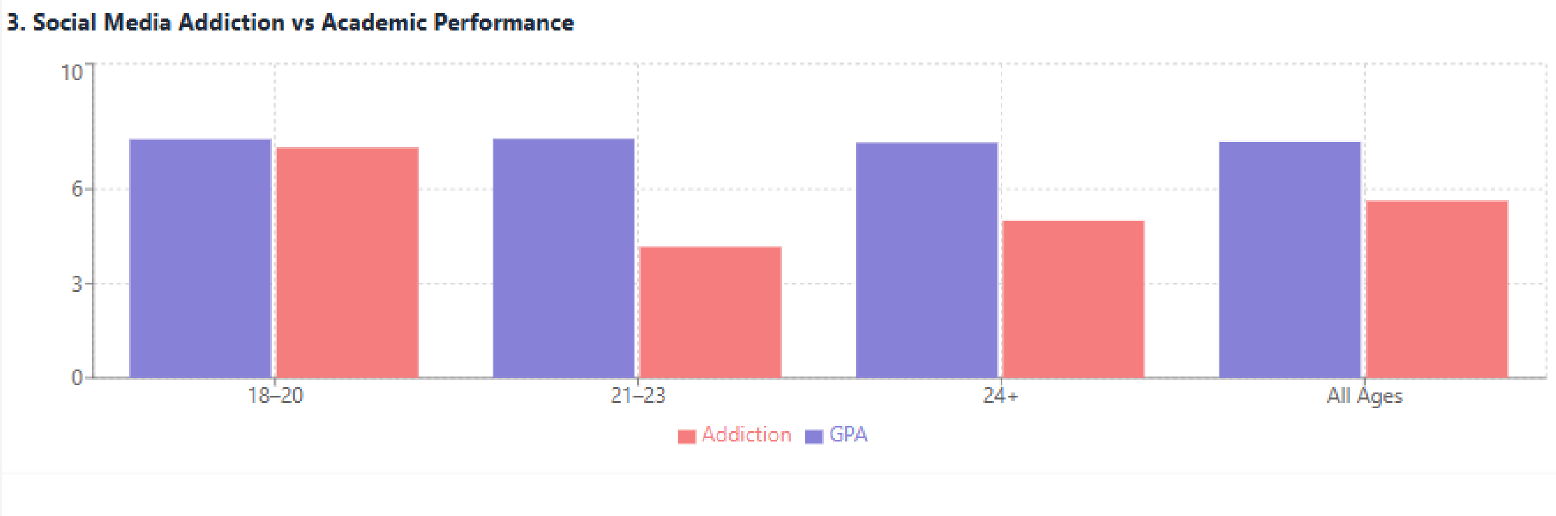
OLAP Queries

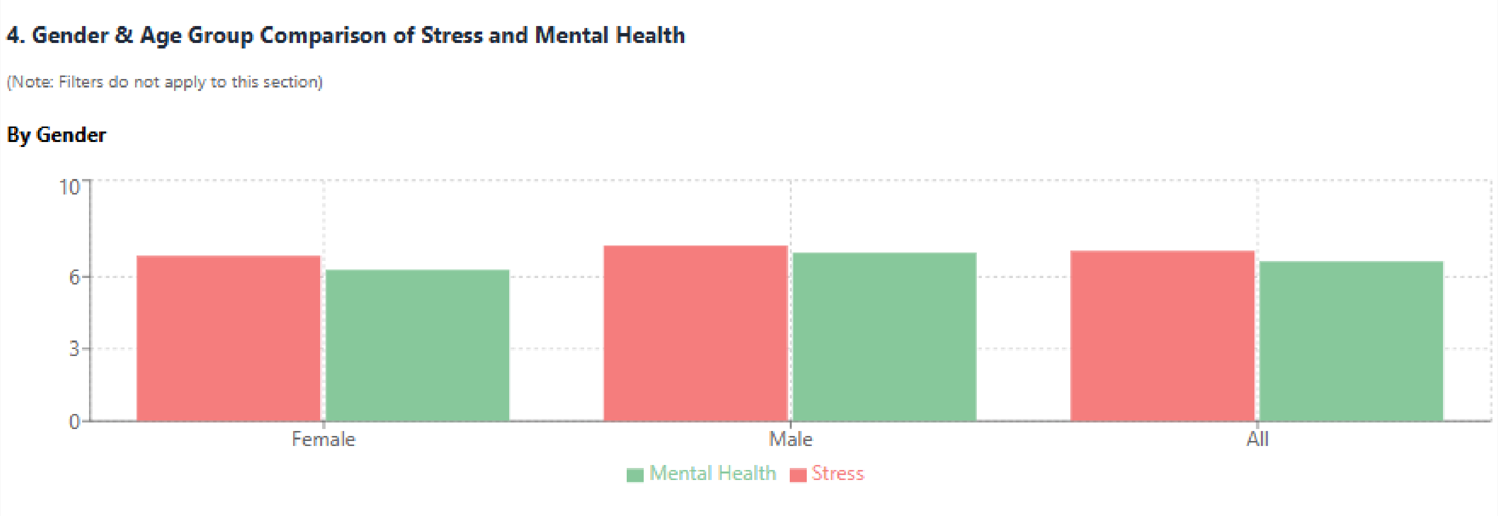
Filters are all set to Gender = ALL, Age Group = ALL

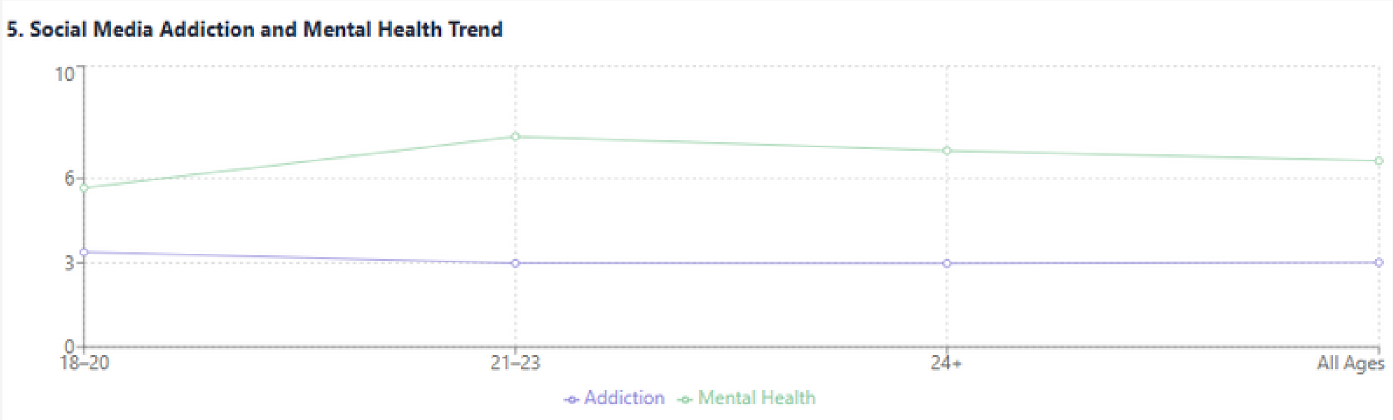




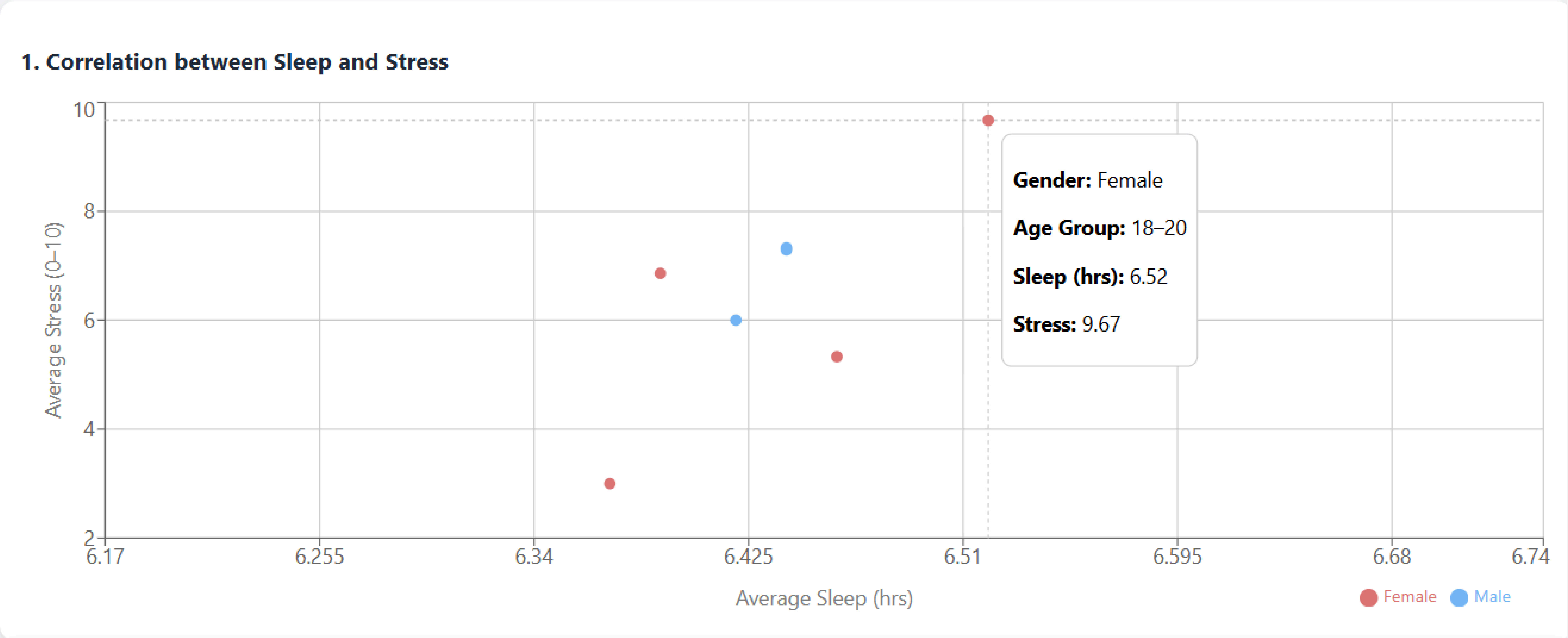








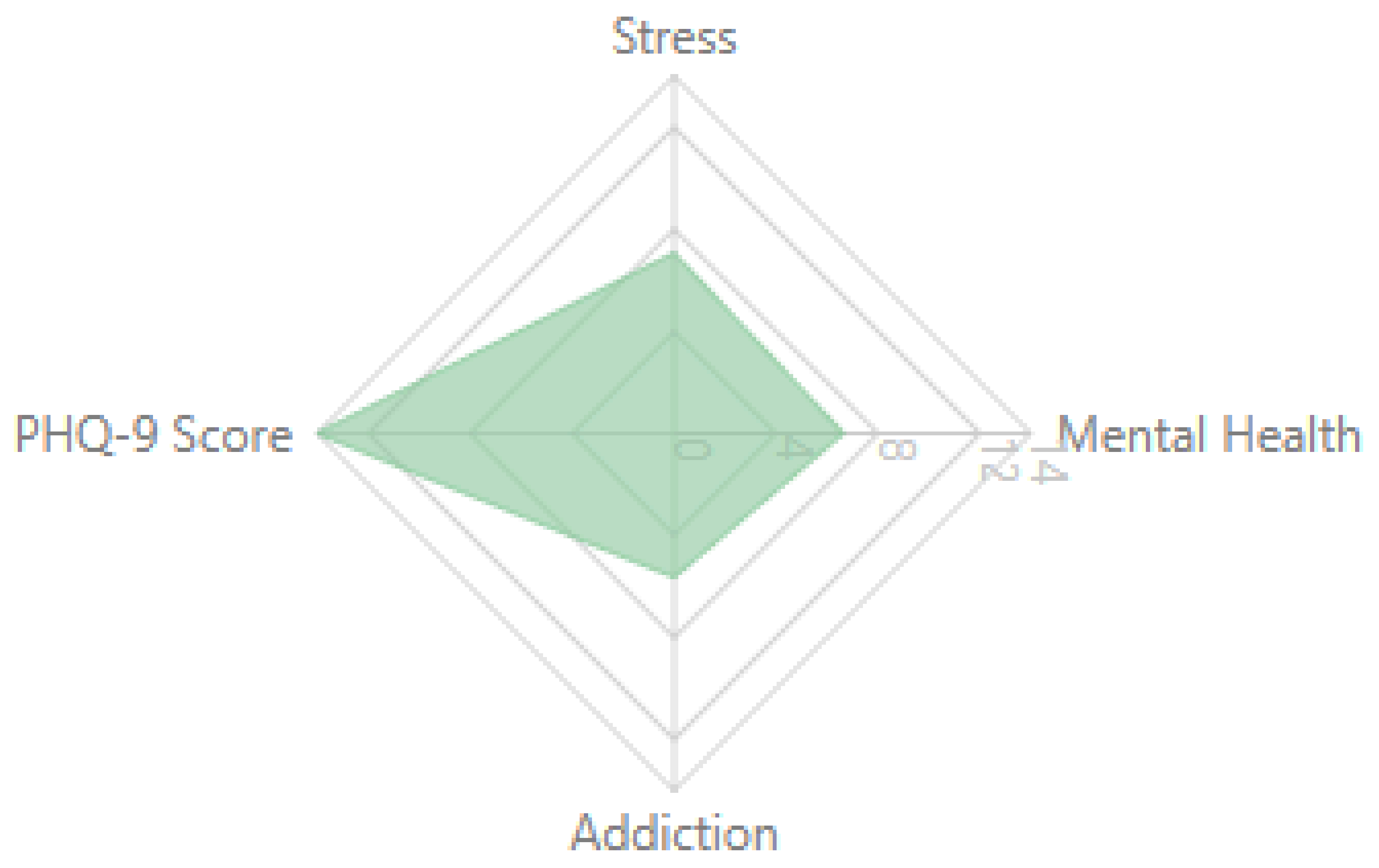
Query 1: Correlation Between Sleep and Stress



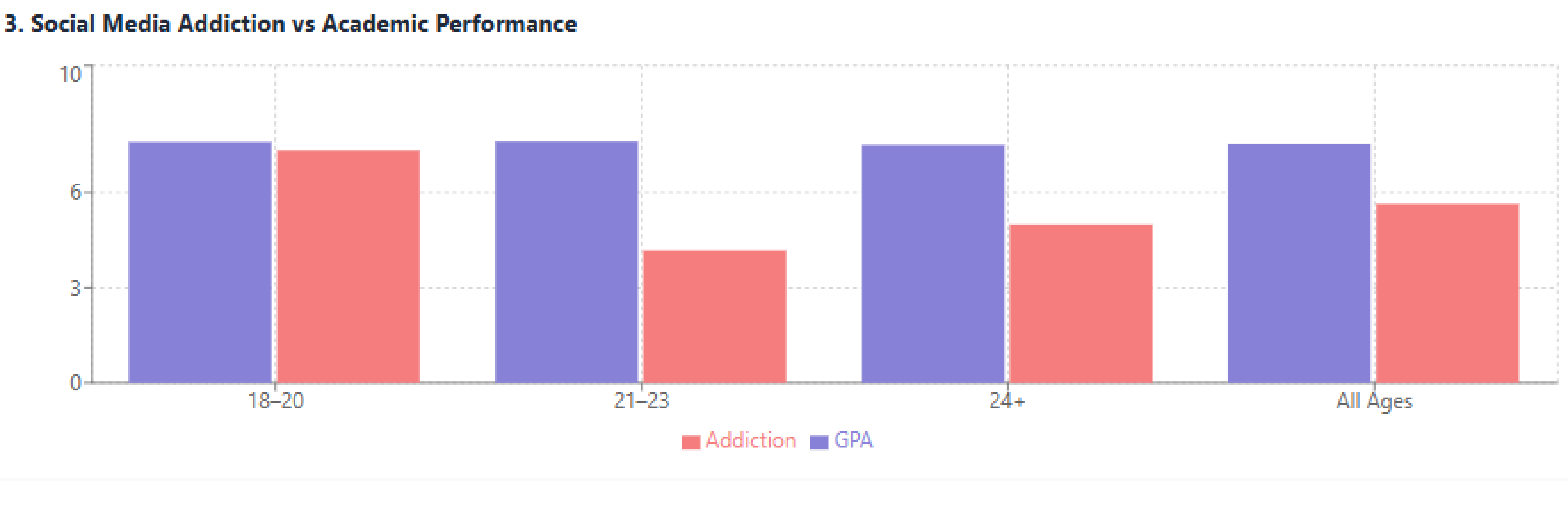
SORTED BY GENDER AGE GROUP

Query 2: Mental Risk Indicators

2. Mental Risk Indicators



Query 3: Social Media VS Academics

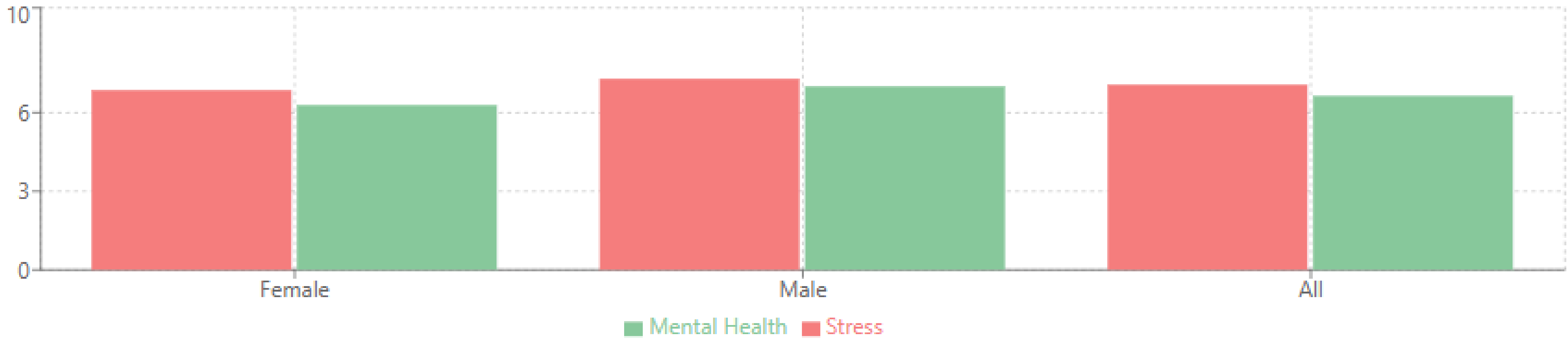


Query 4: Stress by Gender and Age Group

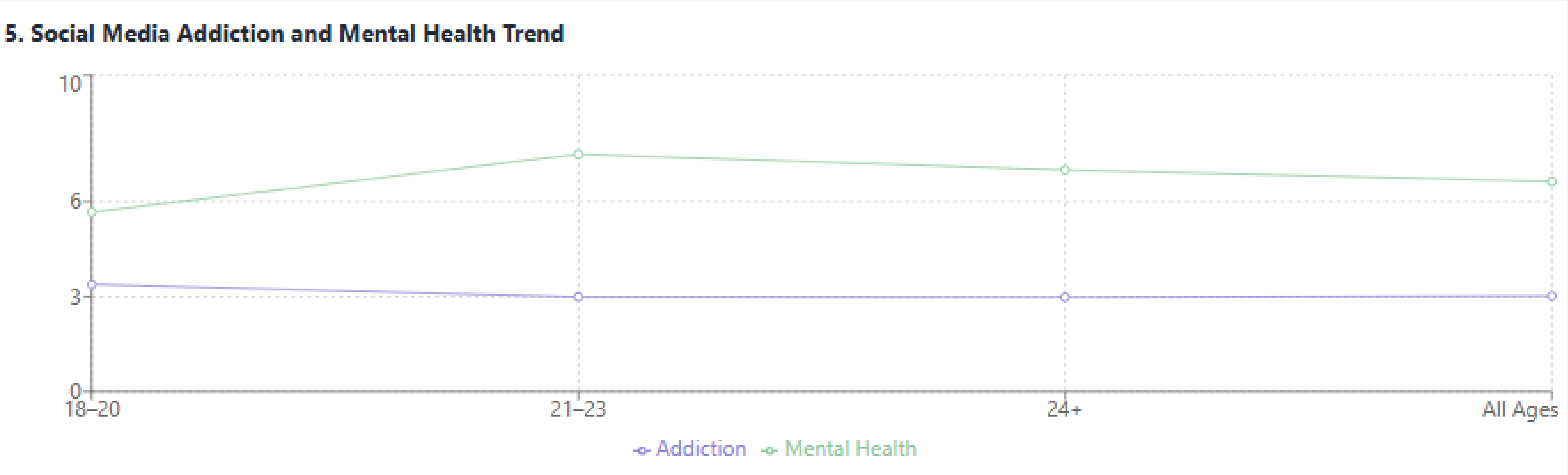
4. Gender & Age Group Comparison of Stress and Mental Health

(Note: Filters do not apply to this section)

By Gender



Query 5: Social Media VS Mental Health



Query 6: Stress and Mental Health by Age Group

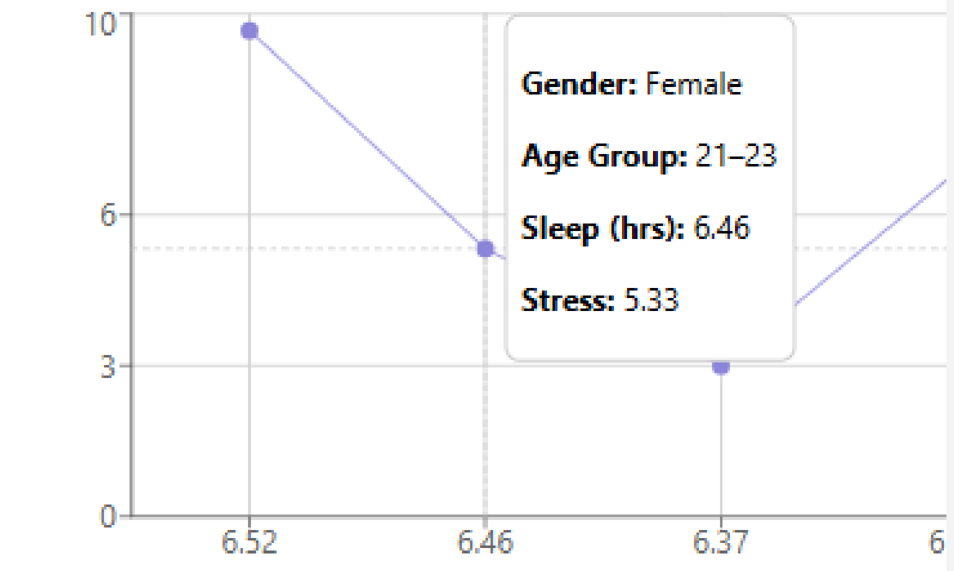
TESTING

```
C:\Users\nipot\Desktop\VSCODE\STADVDB-MCO1-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=
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(dr

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.