Kingdom of Saudi Arabia
Ministry of Education
Umm Al-Qura University
031

المملكة العربية السعودية
وزارة التعليم
جامعة أم القرى
٠٣١

Data Engineering

Project (Student Health & Data Processing)

Students Name:

Layan Al-Joudi 444000810 (2)

Table of Contents

Table of Figures

1. Phase 1: Relational Database (SQL)

Design Overview:

We designed a normalized relational schema consisting of five connected tables:

- Students: ID, Age, Gender
- Health_Metrics: Heart rate, blood pressure
- Stress_Levels: Biosensor and self-reported stress
- Lifestyle: Physical activity, sleep quality, mood
- Study_And_Project: Study hours, project hours, risk level

We used Python and SQLite to:

- Create the schema and insert data from a CSV file
- Perform CRUD operations (Insert, Read, Update, Partial Delete)
- Apply indexes on common fields (e.g., Mood, Study_Hours)
- Execute optimized SQL queries (e.g., top 5 heart rates, students with poor sleep)

2. Phase 2: NoSQL Database (TinyDB)

Design Overview:

To demonstrate flexibility in data modeling, we converted part of our dataset (Student info + Study_And_Project) into a JSON format. We used **TinyDB**, a lightweight NoSQL document database.

What we did:

- Converted selected fields into a list of JSON documents
- Stored the data in tinydb_students.json
- Ran NoSQL queries to extract meaningful information

3. Phase 3: Stream Processing (PySpark)

Design Overview:

We simulated real-time data stream processing using PySpark, loading the same dataset (student_health_data.csv) as a simulated stream.

Processing Logic:

- **Filter 1:** Students with Systolic BP ≥ 130
- **Filter 2:** Students with Biosensor Stress > 7
- **Filter 3:** Students studying more than 40 hours

4. Diagrams and Screenshots

1. Phase 1

```
# Create tables
cursor.execute('''
CREATE TABLE IF NOT EXISTS Students (
    Student_ID INTEGER PRIMARY KEY AUTOINCREMENT,
    Age INTEGER,
    Gender TEXT
)
''')

cursor.execute('''
CREATE TABLE IF NOT EXISTS Health_Metrics (
    Health_Metrics_ID INTEGER PRIMARY KEY AUTOINCREMENT,
    Student_ID INTEGER,
    Heart_Rate REAL,
    Blood_Pressure_Systolic REAL,
    Blood_Pressure_Diastolic REAL,
    FOREIGN KEY (Student_ID) REFERENCES Students(Student_ID)
)
''')

cursor.execute('''
CREATE TABLE IF NOT EXISTS Stress_Levels (
    Stress_Levels_ID INTEGER PRIMARY KEY AUTOINCREMENT,
    Student_ID INTEGER,
    Stress_Level_Biosensor REAL,
    Stress_Level_Self_Report REAL,
    FOREIGN KEY (Student_ID) REFERENCES Students(Student_ID)
)
''')
```

Figure 1 Table creation using SQLite

```
Lifestyle before update: (40, 40, 'Moderate', 'Moderate', 'Stressed')

Lifestyle after update (changed Physical_Activity, Sleep_Quality, Mood): (40, 40, 'Moderate', 'Good', 'Calm')

Health_Metrics before update: (25, 25, 62.947892636702186, 160.1716481045759, 78.58198375003913)

Health_Metrics after update (changed Heart_Rate, Blood_Pressure_Systolic): (25, 25, 82.0, 125.0, 78.58198375003913)

Before partial delete: (205, 205, 64.97744283805876, 132.65058933235588, 84.03300390245282)

After partial delete (cleared Blood_Pressure_Diastolic): (205, 205, 64.97744283805876, 132.65058933235588, None)

Study_And_Project before delete: (700, 700, 30.04910558563553, 15.720909957544643, 'Moderate')

Study_And_Project after delete (cleared Project_Hours, Health_Risk_Level): (700, 700, 30.04910558563553, None, None)

✅ CRUD operations completed.
```

Figure 2 CRUD operation Update and Delete example

```
Query 2: Students with 'Poor' sleep quality
(4, 'Poor')
(5, 'Poor')
(16, 'Poor')
(25, 'Poor')
(29, 'Poor')
(30, 'Poor')
(33, 'Poor')
(43, 'Poor')
(49, 'Poor')
(53, 'Poor')
(59, 'Poor')
(61, 'Poor')
(71, 'Poor')
```

Figure 3 Student with poor sleep quality

2. Phase 2

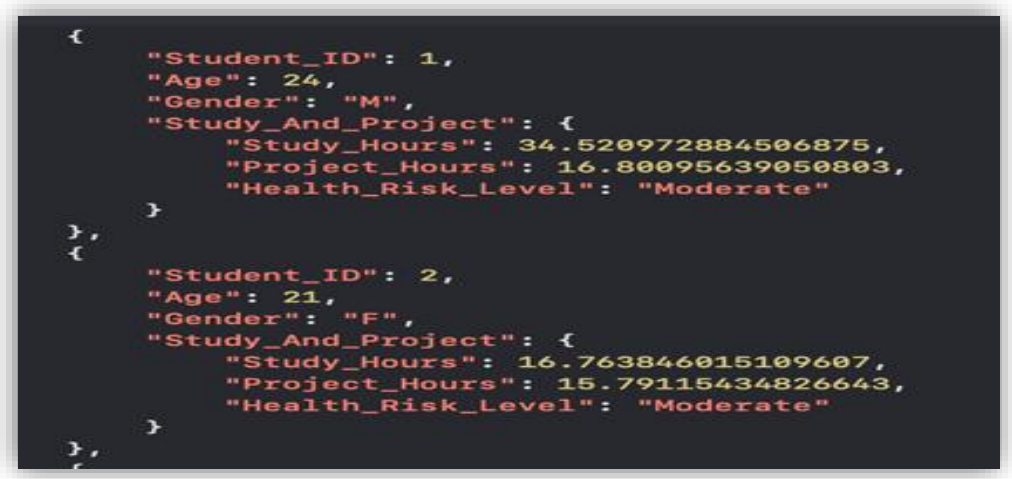


Figure 4 Sample of JSON document after transformation

Figure 5 TinyDB query result Count students with risk = Moderate

3. Phase 3

Student_ID	Age	Gender	Blood_Pressure_Systolic	Stress_Level_Biosensor	Study_Hours	Health_Risk_Level
1	24	M	135.0	3.13	34.52	Moderate
2	21	F	110.77	8.2	16.76	High
3	22	M	109.37	6.78	44.2	Moderate
4	24	M	125.14	9.0	21.77	High
5	20	M	140.51	7.26	41.96	High

Figure 6 Sample student records

Students with high stress (biosensor > 7):

Student_ID	Stress_Level_Biosensor
2	8.2
4	9.0
5	7.26

Figure 7 Filtered students with high stress

4. Dashboard

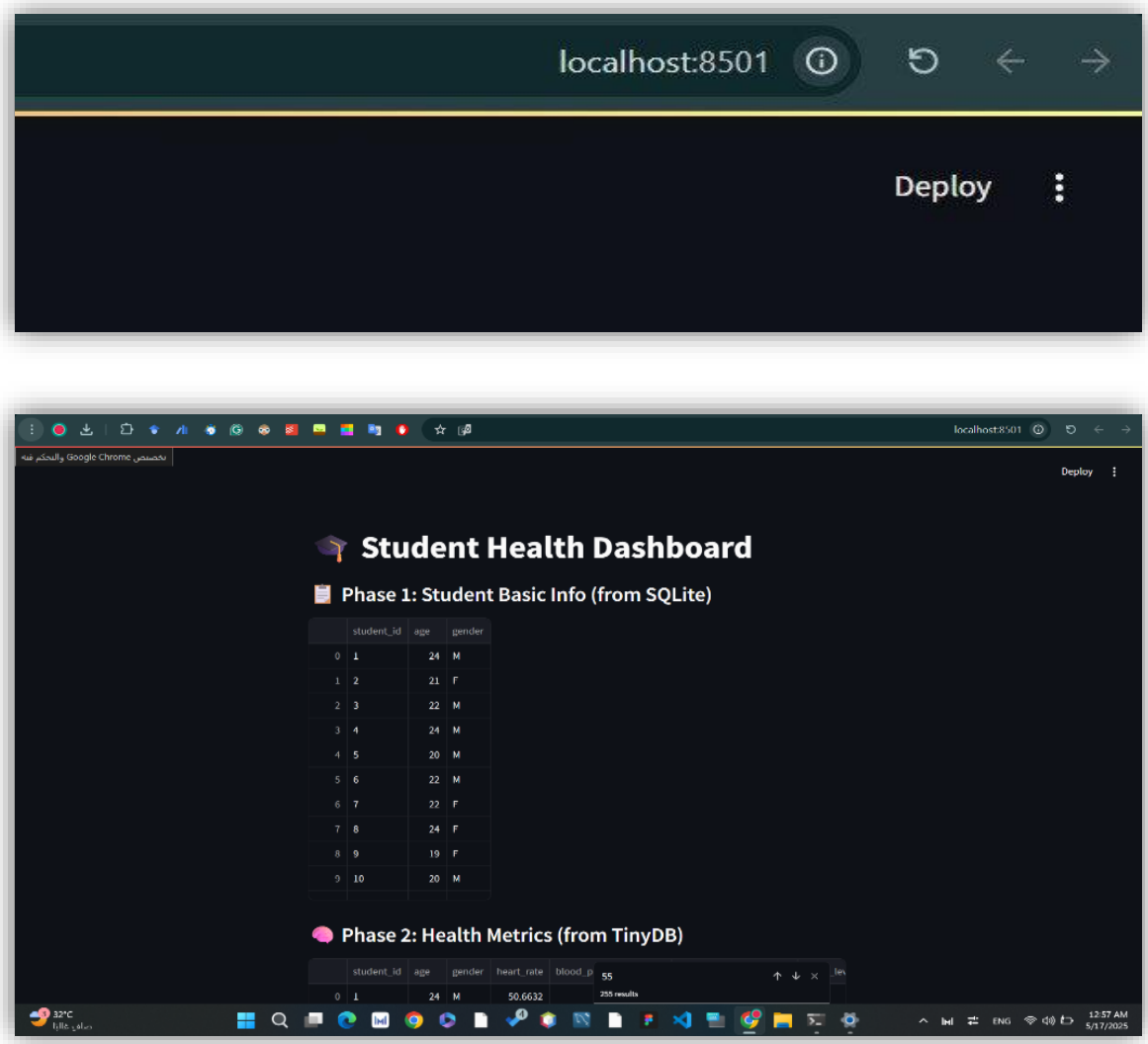
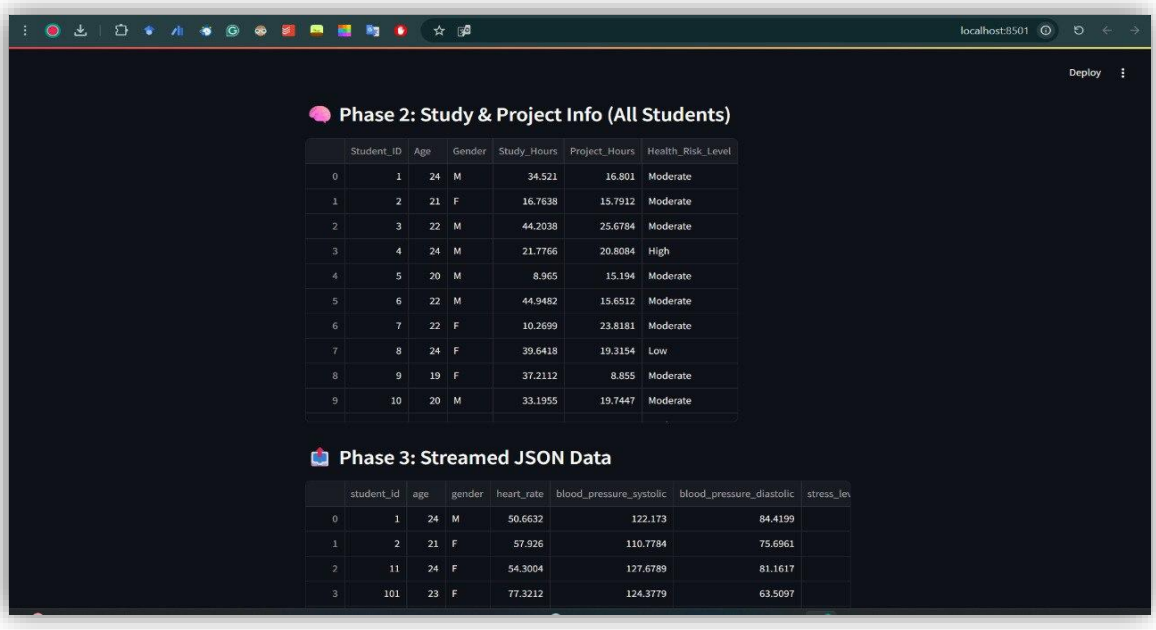


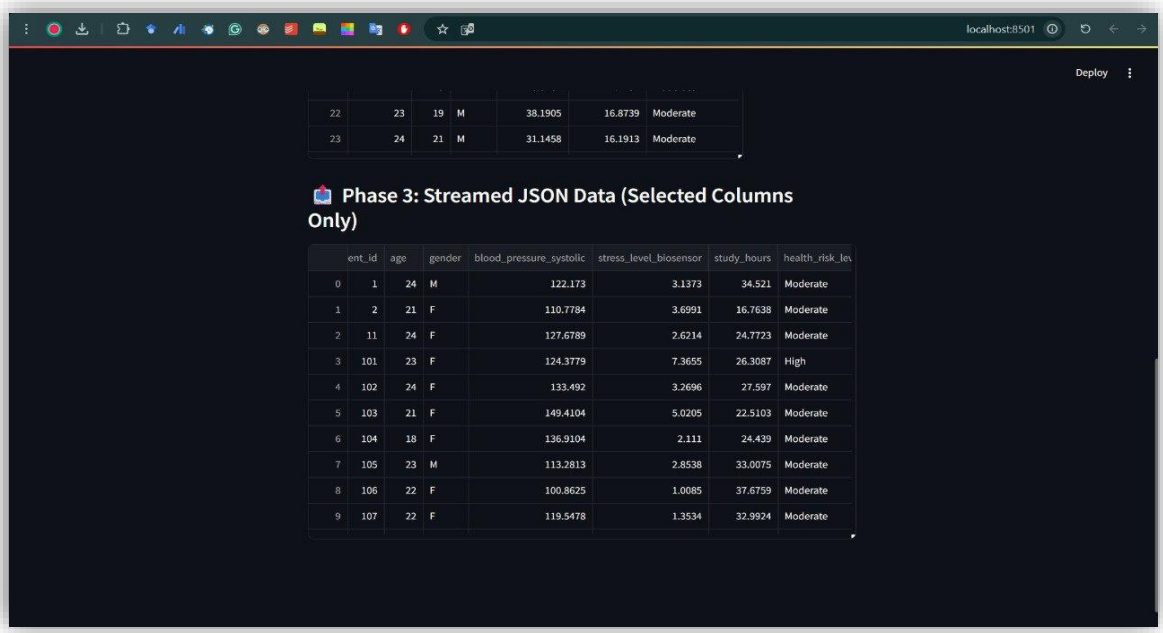
Figure 8 Phase1 Dashboard



	Student_ID	Age	Gender	Study_Hours	Project_Hours	Health_Risk_Level
0	1	24	M	34.521	16.801	Moderate
1	2	21	F	16.7638	15.7912	Moderate
2	3	22	M	44.2038	25.6784	Moderate
3	4	24	M	21.7766	20.8084	High
4	5	20	M	8.965	15.194	Moderate
5	6	22	M	44.9482	15.6512	Moderate
6	7	22	F	10.2699	23.8181	Moderate
7	8	24	F	39.6418	19.3154	Low
8	9	19	F	37.2112	8.855	Moderate
9	10	20	M	33.1955	19.7447	Moderate

	student_id	age	gender	heart_rate	blood_pressure_systolic	blood_pressure_diastolic	stress_lev
0	1	24	M	50.6632	122.173	84.4199	
1	2	21	F	57.926	110.7784	75.6961	
2	11	24	F	54.3004	127.6789	81.1617	
3	101	23	F	77.3212	124.3779	63.5097	

Figure 9 Phase2 Dashboard



22	23	19	M	38.1905	16.8739	Moderate	
23	24	21	M	31.1458	16.1913	Moderate	

	ent_id	age	gender	blood_pressure_systolic	stress_level_biosensor	study_hours	health_risk_lev
0	1	24	M	122.173	3.1373	34.521	Moderate
1	2	21	F	110.7784	3.6991	16.7638	Moderate
2	11	24	F	127.6789	2.6214	24.7723	Moderate
3	101	23	F	124.3779	7.3655	26.3087	High
4	102	24	F	133.492	3.2696	27.597	Moderate
5	103	21	F	149.4104	5.0205	22.5103	Moderate
6	104	18	F	136.9104	2.111	24.439	Moderate
7	105	23	M	113.2813	2.8538	33.0075	Moderate
8	106	22	F	100.8625	1.0085	37.6759	Moderate
9	107	22	F	119.5478	1.3534	32.9924	Moderate

Figure 10 Phase3 Dashboard

5. Data From Excel

Student_ID	Age	Gender	Heart_Rate	Blood_Pressure_Systolic	Blood_Pressure_Diastolic	Stress_Level_Biosensor	Stress_Level_Self_Report	Physical_Activity	Sleep_Quality	Mood	Study_Hours	Project_Hours	Health_Risk_Level
1	24	M	50.6632169889244	122.17301496986300	84.41986006395030	3.137349746524460	9.028669159842590	High	Moderate	Happy	34.520972884506900	16.80095639050800	Moderate
2	21	F	57.926041683536300	110.77840702264600	75.69614523602980	3.699078336950830	5.819697244773670	Moderate	Good	Stressed	16.763846015109600	15.791154348266400	Moderate
3	22	M	59.29421920322830	109.37567306097500	83.80381448216310	6.785155634556780	5.892360174201630	Low	Moderate	Happy	44.2037984802835	25.67843704706690	Moderate
4	24	M	76.82623236457150	125.14222743603900	78.0915868135665	6.408509360839470	6.884001306605410	High	Poor	Happy	21.776645286472800	20.80839115661620	High
5	20	M	68.34276946853390	107.51559175322700	80.67493681247280	7.264718754103350	4.483450107413980	Moderate	Poor	Happy	8.964999082744990	15.19404500227400	Moderate
6	22	M	61.74415156982540	90.0	84.45086471353400	4.2625178607836200	6.825000550580700	Moderate	Good	Happy	44.948228998923300	15.651195419247700	Moderate
7	22	F	93.09722755406190	106.67778156408900	76.49981545408860	8.415978866919800	1.4029154981963800	Moderate	Moderate	Happy	10.26985000015160	23.81809603875400	Moderate
8	24	F	63.36103086453330	115.39678276968900	70.03950213998230	2.836788950401840	1.486428646147540	Low	Good	Neutral	39.641787733834300	19.315438930026200	Low
9	19	F	81.67102914383640	142.82097746747900	74.6768773407021	5.221366727663390	5.115073776575900	Moderate	Moderate	Neutral	37.211195208758500	8.854985697527770	Moderate
10	20	M	86.21108557787720	112.63593825193000	91.95139179833440	8.208448493872870	5.946600709954310	Moderate	Moderate	Happy	33.195541012198800	19.744741114035500	Moderate
11	24	F	54.30039662342660	127.67886136543900	81.1617308407655	2.621427829593570	8.990358763022370	High	Moderate	Happy	24.772250517834400	21.214005410299700	Moderate
12	20	M	70.73962964073570	117.47537649056900	71.55101546589530	1.9887680236670500	1.9419457051710700	Moderate	Good	Stressed	35.02797294525580	23.088585378878900	Low
13	20	F	64.82466349091780	138.39607715805600	84.01560608387720	8.953155709191910	8.403831344745000	Moderate	Good	Neutral	40.48484873704550	12.992541591982600	High
14	22	F	64.01075603440660	137.3564457819590	84.68315584939140	4.086366964832760	2.890426035286300	Moderate	Moderate	Neutral	20.14189503524530	12.627747522655200	Low
15	21	F	75.55376816806910	137.8426265785990	85.1999907579781	7.154831124181290	4.546752836902900	High	Moderate	Neutral	39.763736729794800	10.290924892783300	Moderate
16	20	M	63.77129503576720	93.99024705873970	70.22162799628710	1.2281374383754500	3.849421913446720	High	Poor	Happy	39.308745111937240	16.54063732168070	High
17	23	F	65.0403095367353	123.10086070867500	73.17609005714110	7.210421995685050	4.601397455788590	Moderate	Good	Happy	32.930510909215600	18.664337912278500	Moderate
18	22	F	68.31885104374710	131.3840790018290	73.23170783612990	3.844911803922740	2.35445948265327	Moderate	Moderate	Happy	24.620251219188900	12.433354066612500	Low
19	19	M	66.26600731623430	125.72307843884000	76.39352087933810	1.5016676822869800	5.532130669632000	Moderate	Good	Stressed	39.354573648848780	14.290923511654200	Moderate
20	21	F	58.26429792987880	129.54894488732700	85.38919185143740	6.488314853933050	8.214783308482310	Moderate	Good	Stressed	35.083129561625200	4.7314110219046800	Moderate
21	23	M	77.95906504847470	120.87526506428200	82.96184804731010	3.167234344658560	4.973806093410070	Low	Moderate	Neutral	19.840403368054300	22.54724525201330	Low
22	23	F	80.48530938507600	108.8282368991020	83.75216957157160	4.48126517846692	5.415522183083790	Moderate	Moderate	Happy	26.326441912317500	7.175118142499520	Moderate
23	19	M	73.43618589535220	94.54353922520620	93.54019142371520	1.7252273105786200	8.638908267531830	Low	Good	Happy	38.190549384207000	16.87389876096400	Moderate
24	21	M	66.23447348977730	110.31645857027300	86.8889093249163	2.324469259854510	7.354523031742280	High	Good	Happy	31.145806171008400	16.19127416686170	Moderate
25	22	F	62.947892636702200	160.1716481045760	78.58198375003910	3.6817951194625900	5.414089676678060	Moderate	Poor	Neutral	40.859828944453300	15.229039842658000	Moderate
26	18	M	62.96379937073550	112.71523235864000	68.16114823947500	1.8078355050414100	3.319260833702620	High	Moderate	Happy	26.583533043516000	15.343984391698400	Low
27	21	M	93.57376365032710	90.60914660863880	72.8545299152913	9.197436382176840	8.9643407216047	Moderate	Moderate	Stressed	20.001047077519000	9.941094121385830	High
28	19	M	79.37883301566640	130.98974041093600	77.38294932321770	1.0544233886558100	6.193183615215030	Moderate	Moderate	Happy	51.936026155610070	13.359673979027500	Moderate
29	23	M	61.598899214166600	132.83190013024000	88.29129827336110	5.354741969778800	2.9903284909962400	Moderate	Poor	Stressed	13.191049172535100	19.844753261260900	Moderate
30	22	M	71.6469782743714	139.30986585228200	79.72668065893910	2.109227123192960	3.4220777737420900	Low	Poor	Neutral	25.001604025983400	21.32577750218320	Low
31	21	M	66.37464974900640	127.0646775556020	73.43804045111560	3.7691118053015400	8.228197800426210	Moderate	Good	Neutral	32.797480006999880	12.895495307125300	Moderate
32	18	F	75.83800376156290	132.68402271700400	72.67107114272770	9.334536038364390	1.3973100587871000	Low	Moderate	Neutral	39.86083891477710	7.1979023476154300	Moderate
33	18	M	75.43499102789340	147.60325702263000	79.30421791078770	2.4552509794658800	9.92882968325938	Moderate	Poor	Neutral	36.46449368542260	11.106630743514800	Moderate
34	20	M	80.83609653905470	108.81789761906800	88.899875622171	9.692490877464990	2.173165117433130	Moderate	Moderate	Neutral	53.112590049428800	24.26348547426440	Moderate
35	20	F	61.494435413378000	105.84223941864400	95.91019305775120	2.852149392904800	8.164616728572440	Moderate	Good	Neutral	28.824015473418700	11.404277550063900	Moderate
36	24	F	71.0473529863813	103.58120190670600	77.94231401970330	7.6740949762014700	4.552977850945150	Low	Good	Neutral	21.270167386494700	14.513981648701100	Moderate
37	19	F	52.51453015042470	116.01345620259000	79.55256819237400	5.920095384075810	5.429945649587200	Low	Moderate	Stressed	27.803039757835500	22.079561810099100	Moderate
38	21	F	57.426596995711600	141.50133629737800	67.86058293842620	8.960308548370830	8.925855971812	Moderate	Moderate	Neutral	26.800234706489800	12.365519973314500	High
39	21	F	78.84100836559180	127.92102551619400	99.39192569099470	5.985939299528180	4.716200706302920	High	Good	Neutral	29.317048527098600	14.172323051761600	Moderate
40	24	F	60.49316208594670	127.33037929063800	65.98997161843920	5.811032972406780	5.83744378305911	Moderate	Moderate	Stressed	20.175118771007200	18.97249145378160	Moderate
41	23	F	61.590912019521500	92.71038107002910	91.71662311780550	2.759122003315600	1.4436996161825300	Moderate	Moderate	Happy	37.84056038250360	23.559034337647500	Low
42	23	M	69.9172192544995	109.60219759145400	76.3466404698753	2.4960642584649300	1.0211272648070800	Moderate	Moderate	Happy	33.708302093362100	11.772240676525100	Low
43	24	M	76.9476103084654	92.46513550706950	74.00633506159680	9.248402360039630	3.203198857015530	Low	Poor	Neutral	17.361191709258900	21.960476220605900	Moderate
44	23	F	59.39173204165900	137.56868683982300	68.89996428114070	3.9505437896503100	7.125247836268910	High	Moderate	Neutral	39.680250758058000	13.592356450678400	Moderate
45	20	M	50.0	143.784697393286	104.10048649407100	4.348740294666060	2.0913840402322300	Moderate	Good	Stressed	21.885184630155000	10.570356472675300	Low
46	21	M	61.29776757856490	131.69350090860600	82.21491623278400	7.838412925855500	2.023857629249400	Low	Good	Neutral	18.171849476957100	16.247782151537800	Moderate
47	24	M	82.05628077952840	105.19820931390700	80.21280437104060	4.049411341805270	4.698645472323130	Moderate	Moderate	Neutral	30.873287039216500	9.212159711157020	Low
48	21	F	73.43135051016870	142.23033740423700	88.86732359554540	9.568562019751020	6.830533347474040	Low	Good	Neutral	12.421565506006900	17.360057157314400	Moderate
49	18	F	83.23874803787210	116.20117634962200	73.49167274370040	7.641349920989200	7.408647930231380	High	Poor	Happy	35.895030497772500	30.74824005567270	High

Figure 11 Excel data before normalisation

6. ER Model: Chen's Notation

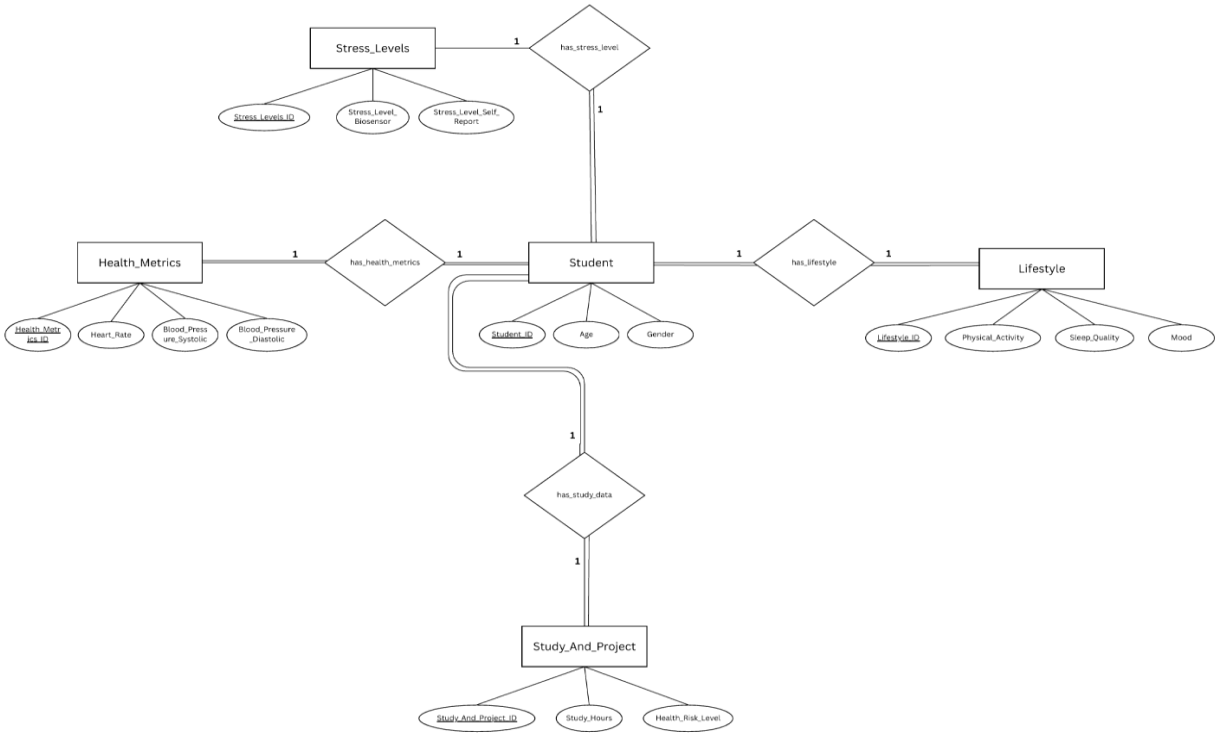


Figure 12 Chen's Notation

7. Normalisation

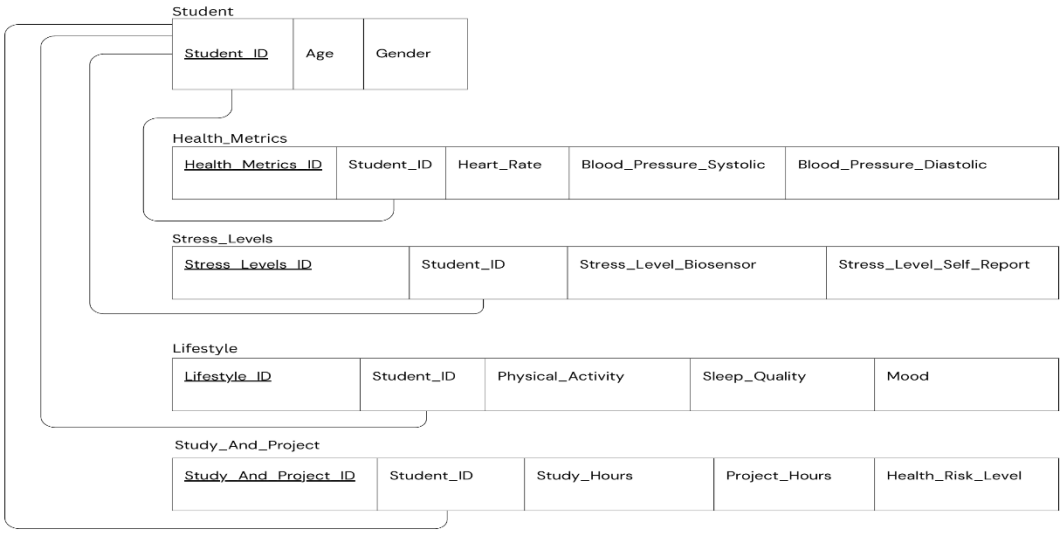


Figure 13 Table after normalisation with primary and foreign key

5. Lessons Learned

- How to build a full data pipeline from ingestion to analysis
- Difference between structured (SQL) and semi-structured (NoSQL) models
- Importance of indexing and normalized design
- Hands-on experience with PySpark stream filtering
- Real teamwork in dividing tasks across phases

6. Project Demo

We have prepared a live demo using Google Colab, which includes:

- **Table creation and SQL queries** using SQLite (**Phase 1**)
  Relational schema, data insertion, indexing, and analytical queries.
- **TinyDB JSON integration and search** (**Phase 2**)
  JSON conversion, document insertion, and NoSQL-style filtering.
- **Stream simulation and filters** using PySpark (**Phase 3**)
  Real-time-like processing to detect high blood pressure, stress, and heavy study loads.
- **Interactive dashboard** using Streamlit (**Phase 4**)
  Combines all data sources into one visual, user-friendly interface to explore and filter student health insights.