



Does going to university in a different country affect your mental health? A Japanese international university surveyed its students in 2018 and published a study the following year that was approved by several ethical and regulatory boards.

The study found that international students have a higher risk of mental health difficulties than the general population, and that social connectedness (belonging to a social group) and acculturative stress (stress associated with joining a new culture) are predictive of depression.

We will explore the `students` data using PostgreSQL to find out if we would come to a similar conclusion for international students and see if the length of stay is a contributing factor.

Here is a data description of the columns you may find helpful.

Field Name	Description
<code>inter_dom</code>	Types of students (international or domestic)
<code>japanese_cate</code>	Japanese language proficiency
<code>english_cate</code>	English language proficiency
<code>academic</code>	Current academic level (undergraduate or graduate)
<code>age</code>	Current age of student
<code>stay</code>	Current length of stay in years
<code>todep</code>	Total score of depression (PHQ-9 test)
<code>tosc</code>	Total score of social connectedness (SCS test)
<code>toas</code>	Total score of acculturative stress (ASIIS test)

Projects Data DataFrame as students

```
-- Viewing the whole dataset to gain a preliminary understanding of the data
SELECT *
FROM students;
```

...	↑↓	i..	↑↓	...	↑↓	...	↑↓	...	↑↓	...	↑↓	...	↑↓	s.	...	↑↓	...	↑↓	japane...	...	↑↓	...	↑↓	engli...	...
44	inter	SEA	Female	unnaer	19	2	1	1	Short	5	Average	5	High	5	High	5	High	5	High	3	Average	4	High	5	High
45	Inter	SA	Male	Under	18	1	1	1	Short	2	Low	4	High	4	High	4	High	4	High	3	Average	4	High	5	High
46	Inter	SEA	Female	Under	20	2	3	3	Medium	3	Average	3	High	3	High	3	High	3	High	3	Average	4	High	5	High
47	Inter	SEA	Female	Under	23	4	3	3	Medium	3	Average	3	High	3	High	3	High	3	High	3	Average	4	High	5	High
48	Inter	Others	Male	Under	23	4	2	2	Medium	2	Low	5	High	5	High	5	High	5	High	5	High	4	High	5	High
49	Inter	Others	Female	Under	20	2	2	2	Medium	3	Average	5	High	5	High	5	High	5	High	5	High	4	High	5	High
50	Inter	Others	Female	Under	20	2	2	2	Medium	2	Low	5	High	5	High	5	High	5	High	5	High	4	High	5	High
51	Inter	EA	Male	Under	21	3	3	3	Medium	2	Low	4	High	4	High	4	High	4	High	4	High	3	Average	4	High
52	Inter	Others	Female	Under	23	4	3	3	Medium	4	High	5	High	5	High	5	High	5	High	5	High	4	High	5	High
53	Inter	SEA	Male	Under	20	2	3	3	Medium	3	Average	5	High	5	High	5	High	5	High	5	High	4	High	5	High
54	Inter	EA	Female	Under	23	4	4	4	Long	3	Average	3	High	3	High	3	High	3	High	3	Average	4	High	5	High
55	Inter	SEA	Female	Under	20	2	2	2	Medium	4	High	4	High	4	High	4	High	4	High	4	High	3	Average	4	High
56	Inter	SA	Male	Under	21	3	1	1	Short	3	Average	4	High	4	High	4	High	4	High	4	High	3	Average	4	High
57	Inter	SEA	Female	Under	21	3	3	3	Medium	1	Low	5	High	5	High	5	High	5	High	5	High	4	High	5	High
58	Inter	SEA	Female	Under	19	2	2	2	Medium	2	Low	4	High	4	High	4	High	4	High	4	High	3	Average	4	High
59	Inter	SEA	Male	Under	18	1	1	1	Short	2	Low	4	High	4	High	4	High	4	High	4	High	3	Average	4	High

Rows: 286

Demography

Q-1 (a) What is the distribution of international students by region?

Using a combination of GROUP BY and by filtering the data for international students by using the WHERE clause tells us that most of the students come from the South East Asia region. This could be due to the fact is is the nearest region from Japan. Surprisingly, there are two international students from Japan as well!

Projects Data DataFrame as df1

```
SELECT
    region,
    COUNT(age) AS students_from_region
FROM students
WHERE Inter_dom = 'Inter'
GROUP BY region
ORDER BY students_from_region DESC;
```

index	...	region	...	students_from_region
	0	SEA		
	1	EA		
	2	SA		
	3	Others		
	4	JAP		

Rows: 5

Q-1 (b) What is the distribution of international students by region in terms of percentage?

More than half of students studying in Japan come from the region of South East Asia. EA follows SEA by 23.8%.

Projects Data DataFrame as df2

```
SELECT
    region,
    COUNT(age) AS students_from_region,
    ROUND(COUNT(age) * 100.0 /
        (SELECT COUNT(age)
        FROM students
        WHERE Inter_dom = 'Inter'), 2) AS students_from_region_perc
FROM students
WHERE Inter_dom = 'Inter'
GROUP BY region
ORDER BY students_from_region DESC;
```

index	... ↑↓	region	... ↑↓	students_from_region	... ↑↓	students_from_region_perc
0		SEA				122
1		EA				48
2		SA				18
3		Others				11
4		JAP				2

Rows: 5

Q-2 What is the distribution of international and domestic students by gender?

The gender ratio, regardless of the student type is more or less the same (6:3).

Projects Data DataFrame as df3

```
SELECT
    Inter_dom,
    COUNT(CASE WHEN gender = 'Female' THEN 1 ELSE NULL END) AS female_students,
    ROUND(AVG(CASE WHEN gender = 'Female' THEN 1 ELSE 0 END), 2) AS female_students_perc,
    COUNT(CASE WHEN gender = 'Male' THEN 1 ELSE NULL END) AS male_students,
    ROUND(AVG(CASE WHEN gender = 'Male' THEN 1 ELSE 0 END), 2) AS male_students_perc
FROM students
WHERE Inter_dom IN ('Inter', 'Dom')
GROUP BY Inter_dom
```

in... ... ↑↓	inter_d... ... ↑↓	female_students	... ↑↓	female_students_perc	... ↑↓	male_students	... ↑↓	male_students_pe
0	Inter			128		0.64		73
1	Dom			42		0.63		25

Rows: 2

Q-3 What is the distribution of international vs. domestic students by academic level?

This question is crucial in determining the specific area which needs special targeted care and attention when it comes to mental health needs. The analysis indicates that 90 percent of students study undergraduate courses and our analysis will concentrate on their mental health levels.

Projects Data DataFrame as df4

```
SELECT
    Inter_dom,
    COUNT(CASE WHEN academic = 'Grad' THEN 1 ELSE NULL END) AS grad_students,
    ROUND(AVG(CASE WHEN academic = 'Grad' THEN 1 ELSE 0 END), 3) AS grad_students_perc,
    COUNT(CASE WHEN academic = 'Under' THEN 1 ELSE NULL END) AS undergrad_students,
    ROUND(AVG(CASE WHEN academic = 'Under' THEN 1 ELSE 0 END), 3) AS undergrad_students_perc
FROM students
WHERE Inter_dom != ''
GROUP BY Inter_dom;
```

i... ... ↑↓	inter... ... ↑↓	grad_students	... ↑↓	grad_students_perc	... ↑↓	undergrad_students	... ↑↓	undergrad_students_perc
0	Inter			20		0.1		181
1	Dom			1		0.015		66

Rows: 2

Q-4 (a) What is the average age of students categorized by their academic level (undergraduate/graduate)?

This question comes into play to understand the age distribution of the students in question.

Projects Data DataFrame as df5

```
SELECT
    academic AS course,
    ROUND(AVG(age), 2) AS avg_age
FROM students
WHERE Inter_dom != ''
GROUP BY academic;
```

index	...	↑↓	course	...	↑↓	avg_age
			0 Under			
			1 Grad			

Rows: 2

The average age of undergraduate students is 20.3 years, while graduate students average 27.67 years

Q-4 (b) What is the average age of students by language proficiency category?

Projects Data DataFrame as df6

```
SELECT
    japanese_cate AS jap_proficiency,
    ROUND(AVG(age), 2) AS avg_age
FROM students
WHERE Inter_dom != ''
GROUP BY japanese_cate;
```

index	...	↑↓	jap_proficiency	...	↑↓	avg_age
			0 Average			
			1 High			
			2 Low			

Rows: 3

Q-5 What is the relationship between the length of stay and language proficiency among international students?

Projects Data DataFrame as df7

```
SELECT
    japanese_cate,
    ROUND(AVG(stay), 2) AS avg_stay_duration
FROM students
WHERE inter_dom = 'Inter'
GROUP BY japanese_cate
ORDER BY avg_stay_duration DESC;
```

index	...	↑↓	japanese_cate	...	↑↓	avg_stay_duration
			0 High			
			1 Average			
			2 Low			

Rows: 3

As expected, students who stay longer in Japan demonstrate a higher proficiency in Japanese language. This information can be useful later in determining if a high proficiency in the Japanese language shows a lower level of mental health issues.

Mental Health Analysis

Now, we'll expand our scope of analysis and examine the mental health status among students. This analysis is significantly useful because indicators like Japanese proficiency and stay duration can impact mental health levels in students.

Q-1 How do mental health indicators vary among international students based on their length of stay?

Projects Data DataFrame as df8

```
SELECT stay,
       COUNT(inter_dom) AS count_int,
       ROUND(AVG(todep), 2) AS average_phq,
       ROUND(AVG(tosc), 2) AS average_scs,
       ROUND(AVG(toas), 2) AS average_as
  FROM students
 WHERE inter_dom = 'Inter'
 GROUP BY stay
 ORDER BY stay DESC
 LIMIT 9;
```

index	stay	count_int	average_phq	average_scs	average_as
0	10	1	13	32	
1	8	1	10	44	
2	7	1	4	48	
3	6	3	6	38	
4	5	1	0	34	
5	4	14	8.57	33.93	
6	3	46	9.09	37.13	
7	2	39	8.28	37.08	
8	1	95	7.48	38.11	

Rows: 9

An international student living in Japan for 4 years, on average, experienced rising levels of depression (PHQ-9) and reduced social connectedness (SCS). Students who continued to live for an elongated period of upto 10 years, showed further increase in depressive symptoms and a decline in social connectedness. This suggests that students studying in countries away from their homes have a higher need of feeling connected with their social surroundings but in Japan's case, data hints towards the non-fulfillment of this basic requirement. However, students living for longer durations in Japan have reported a lower level of acculturative stress (ASISS), potentially reflecting successful cultural adjustment over the period of time. Not many international students stay in Japan for more than three years, attributing to the fact that most students leave after obtaining an undergraduate degree.

Q-2 Is there a correlation between mental health scores by language proficiency level?

Projects Data DataFrame as df

```
SELECT japanese_cate,
       COUNT(inter_dom) AS count_int,
       ROUND(AVG(todep), 2) AS average_phq,
       ROUND(AVG(tosc), 2) AS average_scs,
       ROUND(AVG(toas), 2) AS average_as
  FROM students
 WHERE inter_dom = 'Inter'
 GROUP BY japanese_cate
 ORDER BY
 CASE
     WHEN japanese_cate = 'Low' THEN 0
     WHEN japanese_cate = 'Average' THEN 1
     WHEN japanese_cate = 'High' THEN 2
     ELSE NULL
 END;
```

index	japanese_cate	count_int	average_phq	average_scs	average_as
0	Low	91	7.91	36.99	
1	Average	85	8.38	37.56	
2	High	25	7.4	38.48	

Rows: 3

International students with an average proficiency in Japanese language, on average, exhibited higher levels of depressive symptoms. A high proficiency comes with a lower level of depression, but this level of 8.38 still fell within the range of mild depression. As an international student develops a higher level of proficiency of the Japanese language, mastering complex Kanji symbols and pronunciations, they experience less acculturative stress. This indicates that being able to speak the native language fluently plays a vital role in successfully adapting to the native culture.

Are older students more likely to report higher depression scores?

Projects Data DataFrame as df9

```
SELECT
CASE
WHEN age BETWEEN 18 AND 22 THEN '18-22'
WHEN age BETWEEN 23 AND 27 THEN '23-27'
WHEN age >= 28 THEN '28+'
ELSE 'Unknown'
END AS age_group,
ROUND(AVG(todep), 2) AS average_phq
FROM students
GROUP BY age_group
ORDER BY age_group;
```

index	age_group	average_phq
0	18-22	
1	23-27	
2	28+	
3	Unknown	

Rows: 4

At a PHQ-9 score of 8.47, younger students face higher levels of depression symptoms on average. Most students struggle with homesickness, being unable to adapt to an unexperienced culture, and lack of friends and emotional support. As they get older and reach the ages of 23-27, the average PHQ-9 score eases to 7.85, suggesting overcoming of homesickness and a better network of emotional belongingness, which lowers mental stress and depressive symptoms. This trend persists into the middle-ages and by this time, they typically develop stable coping mechanisms and a robust network of mental support systems.

Conclusion

The exploration of the `students` dataset using PostgreSQL revealed some interesting findings which suggested that there is a critical need to address mental health problem among young international students. The local government can focus on developing Japanese proficiency by introducing special crash courses on Japanese in their curriculum, targeted towards foreign students. They can make Japanese culture easier to integrate into by publishing informative videos, blogs and short-form content and making signs and instructions accessible to English-speaking students. These measures hold the potential to save Japan's falling birth rates as making the country's education international friendly will boost admissions and consequently, large scale immigration into Japan will help increase its birth rate. This project was essential to my understanding of grouping, filtering and aggregating functions of PostgreSQL and its usefulness in deriving useful insights from real world data.