

## ✓ Project: Analyzing Students' Mental Health

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.

Explore the students data using PostgreSQL to find out if you 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
inter_dom	Types of students (international or domestic)
japanese_cate	Japanese language proficiency
english_cate	English language proficiency
academic	Current academic level (undergraduate or graduate)
age	Current age of student
stay	Current length of stay in years
todep	Total score of depression (PHQ-9 test)
tosc	Total score of social connectedness (SCS test)
toas	Total score of acculturative stress (ASISS test)

```
import pandas as pd
import sqlite3

# Load the dataset (adjust the path if your dataset name is different)
df = pd.read_csv('/kaggle/input/dataset/students.csv')

# Create an in-memory SQLite database
conn = sqlite3.connect(':memory:')

# Load DataFrame into SQLite
df.to_sql('students', conn, index=False)

# View the first few rows (similar to SELECT * FROM students LIMIT 10;)
pd.read_sql_query('SELECT * FROM students LIMIT 10;', conn)
```

	inter_dom	region	gender	academic	age	age_cate	stay	stay_cate	japanese	japanese_cate	...	friends_bi	parents_bi	relative_bi
0	Inter	SEA	Male	Grad	24.0	4.0	5.0	Long	3.0	Average	...	Yes	Yes	Nc
1	Inter	SEA	Male	Grad	28.0	5.0	1.0	Short	4.0	High	...	Yes	Yes	Nc
2	Inter	SEA	Male	Grad	25.0	4.0	6.0	Long	4.0	High	...	No	No	Nc
3	Inter	EA	Female	Grad	29.0	5.0	1.0	Short	2.0	Low	...	Yes	Yes	Yes
4	Inter	EA	Female	Grad	28.0	5.0	1.0	Short	1.0	Low	...	Yes	Yes	Nc
5	Inter	SEA	Male	Grad	24.0	4.0	6.0	Long	3.0	Average	...	Yes	No	Nc
6	Inter	SA	Male	Grad	23.0	4.0	1.0	Short	3.0	Average	...	Yes	Yes	Nc
7	Inter	SEA	Female	Grad	30.0	5.0	2.0	Medium	1.0	Low	...	No	No	Nc
8	Inter	SEA	Female	Grad	25.0	4.0	4.0	Long	4.0	High	...	No	No	Nc
9	Inter	Others	Male	Grad	31.0	5.0	2.0	Medium	1.0	Low	...	No	No	Nc

10 rows × 50 columns

Find the number of international students and their average scores by length of stay, in descending order of length of stay.

```
query = """
SELECT stay,
       COUNT(*) 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;
"""
pd.read_sql_query(query, conn)

```

	stay	count_int	average_phq	average_scs	average_as
0	10.0	1	13.00	32.00	50.00
1	8.0	1	10.00	44.00	65.00
2	7.0	1	4.00	48.00	45.00
3	6.0	3	6.00	38.00	58.67
4	5.0	1	0.00	34.00	91.00
5	4.0	14	8.57	33.93	87.71
6	3.0	46	9.09	37.13	78.00
7	2.0	39	8.28	37.08	77.67
8	1.0	95	7.48	38.11	72.80

**Check if Japanese language proficiency affects mental health. For each language level, count students and average scores.**

```

query = """
SELECT japanese_cate,
       COUNT(*) AS count_students,
       ROUND(AVG(todep), 2) AS avg_depression,
       ROUND(AVG(tosc), 2) AS avg_connectedness,
       ROUND(AVG(toas), 2) AS avg_stress
FROM students
WHERE inter_dom = 'Inter'
GROUP BY japanese_cate
ORDER BY avg_depression DESC;
"""
pd.read_sql_query(query, conn)

```

	japanese_cate	count_students	avg_depression	avg_connectedness	avg_stress
0	Average	85	8.38	37.56	75.45
1	Low	91	7.91	36.99	76.65
2	High	25	7.40	38.48	72.00

**Compare undergraduates vs graduates among international students. See if academic level influences depression, connectedness, or stress.**

```

query = """
SELECT academic,
       COUNT(*) AS count_students,
       ROUND(AVG(todep), 2) AS avg_depression,
       ROUND(AVG(tosc), 2) AS avg_connectedness,
       ROUND(AVG(toas), 2) AS avg_stress
FROM students
WHERE inter_dom = 'Inter'
GROUP BY academic;
"""
pd.read_sql_query(query, conn)

```

	academic	count_students	avg_depression	avg_connectedness	avg_stress
0	Grad	20	4.95	40.90	75.80
1	Under	181	8.39	37.03	75.54


```

import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# Load the dataset (adjust path if needed)

```

```
df = pd.read_csv('/kaggle/input/dataset/students.csv')
# Define df_inter for international students
df_inter = df[df['inter_dom'] == 'Inter']
# Quick preview (optional)
df.head(10)
```



	inter_dom	region	gender	academic	age	age_cate	stay	stay_cate	japanese	japanese_cate	...	friends_bi	parents_bi	relative_bi
0	Inter	SEA	Male	Grad	24.0	4.0	5.0	Long	3.0	Average	...	Yes	Yes	No
1	Inter	SEA	Male	Grad	28.0	5.0	1.0	Short	4.0	High	...	Yes	Yes	No
2	Inter	SEA	Male	Grad	25.0	4.0	6.0	Long	4.0	High	...	No	No	No
3	Inter	EA	Female	Grad	29.0	5.0	1.0	Short	2.0	Low	...	Yes	Yes	Yes
4	Inter	EA	Female	Grad	28.0	5.0	1.0	Short	1.0	Low	...	Yes	Yes	No
5	Inter	SEA	Male	Grad	24.0	4.0	6.0	Long	3.0	Average	...	Yes	No	No
6	Inter	SA	Male	Grad	23.0	4.0	1.0	Short	3.0	Average	...	Yes	Yes	No
7	Inter	SEA	Female	Grad	30.0	5.0	2.0	Medium	1.0	Low	...	No	No	No
8	Inter	SEA	Female	Grad	25.0	4.0	4.0	Long	4.0	High	...	No	No	No
9	Inter	Others	Male	Grad	31.0	5.0	2.0	Medium	1.0	Low	...	No	No	No

10 rows × 50 columns

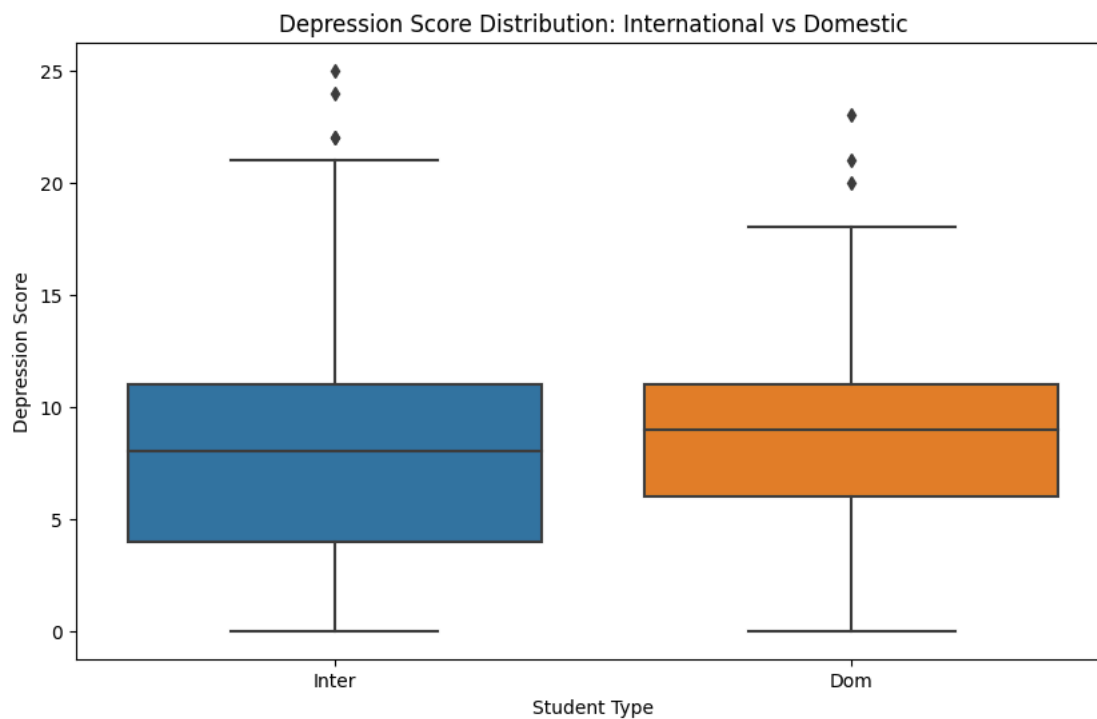
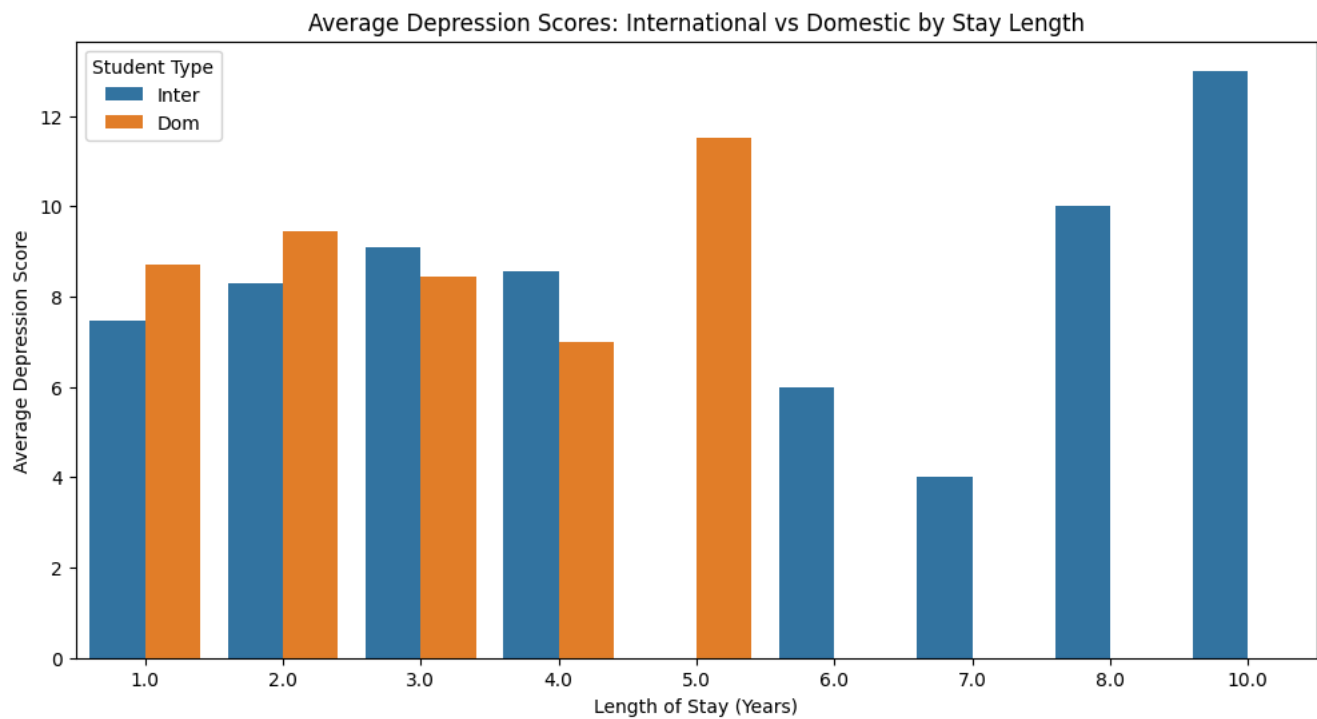
**Compare averages across stay lengths and visualize differences in scores between student types.**

```
# Group by stay and inter_dom
df_compare = df.groupby(['stay', 'inter_dom']).agg(
    count_students=('inter_dom', 'size'),
    avg_depression=('todep', 'mean'),
    avg_connectedness=('tosc', 'mean'),
    avg_stress=('toas', 'mean')
).round(2).sort_values(['stay', 'inter_dom'], ascending=[False, True]).reset_index()

# Plot: Grouped bar for depression
plt.figure(figsize=(12, 6))
sns.barplot(data=df_compare, x='stay', y='avg_depression', hue='inter_dom')
plt.title('Average Depression Scores: International vs Domestic by Stay Length')
plt.xlabel('Length of Stay (Years)')
plt.ylabel('Average Depression Score')
plt.legend(title='Student Type')
plt.show()

# Box plot for overall depression distribution
plt.figure(figsize=(10, 6))
sns.boxplot(data=df, x='inter_dom', y='todep')
plt.title('Depression Score Distribution: International vs Domestic')
plt.xlabel('Student Type')
plt.ylabel('Depression Score')
plt.show()

# Repeat for connectedness and stress if needed (change y='avg_connectedness' or y='avg_stress')
```



Explore how proficiency levels affect scores for international students and visualize the patterns.

```
# Group by japanese_cate for internationals
df_lang = df_inter.groupby('japanese_cate').agg(
    count_students=('japanese_cate', 'size'),
    avg_depression=('todep', 'mean'),
    avg_connectedness=('tosc', 'mean'),
    avg_stress=('toas', 'mean')
).round(2).sort_values('avg_depression', ascending=False).reset_index()

# Plot: Bar chart
df_lang.plot(x='japanese_cate', y=['avg_depression', 'avg_connectedness', 'avg_stress'], kind='bar', figsize=(8, 5))
plt.title('Mental Health Scores by Japanese Proficiency Level (International Students)')
plt.xlabel('Proficiency Level')
plt.ylabel('Average Score')
```

```
plt.legend()  
plt.show()
```

