Introduction

This data analysis project aims to develop my skills and show them to possible employers. The focus is to show how simple, yet, effective skills combined with a good analytical thought can help us understand reality better to make data driven decisions.

I will analyze a dataset of a study conducted with international students to investigate whether going to university in a different country affects their mental health. The study was conducted by a Japanese international university in 2018 and was approved by several ethical and regulatory boards.

I will clean and manipulate the data using SQL to prepare it for analysis. Then, I will use statistical analysis in spreadsheets to understand the correlation between some variables, such as:

| `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) |

‘suicide’ ------------------ If the student has commited suicide |  
‘religion’ ---------------------------------------if the student had a religion |

Between each step made, I will log here the sql queries made to process data to use and interpret in the spreadsheet, with the conclusions of the analysis.

What are we going to test correlations to give us insights:

total depression score within international students by years living outside home country

To understand better how studying in a foreign country affects the happiness in general, i'll make a comparison between the mean of the total score of depression within international students grouped by years staying out of their home country.   
  
If there is a clear difference between the different samples, i will conduct a test to see how the years staying in the country affects the social connectedness.

Correlation between social connectedness and stress

To analyze the correlation between social connectedness and stress, I will use Pearson's correlation.

Length of stay in the foreign country and correlation with total score of depression

To analyze the correlation between length of stay in the foreign country and total score of depression.

Religion and suicide

To analyze the correlation between religion and suicide, i will use the chi-squared test to verify if there is a strict correlation between both.

Religion and stress

To analyze the correlation between religion and stress, I will compare the mean stress of those who have religion and those who do not. If the mean stress is higher for those who do not have religion, this would indicate that there is a negative correlation between the two variables. In other words, the lower the religiosity, the higher the risk of stress.

Total depression score correlation with stress and social connectedness  
  
With this analysis we want to check which is the most influential to deteriorate mental health, the lack of social activity or the amount of stress accumulated. It will be made a Pearson correlation between stress and todep and between social connectedness and todep, after that we will compare both.

CALCULATING THE CORRELATION BETWEEN RELIGION AND STRESS  
  
I could easily calculate this with SQL using the following query:  
  
SELECT religion, ROUND(AVG(toas), 0) as avg\_stress

FROM students

WHERE religion is not null

GROUP BY religion;

But since my intention is to learn and use the spreadsheets with SQL, this was the query made:   
  
  
SELECT religion, ROUND(toas, 0) as stress

FROM students

WHERE religion IS NOT NULL

conclusion of analysis  
  
The group with religious students had an average of toas of 71,31, while the non religious group had a score of 72,93.

Despite the existing differences, they are not significant enough for our analysis to indicate any correlation between the stress level and religiosity.

The difference between religious and non religious stress average score is just 1,62, the amplitude of the toas of all students is 106 and the standard deviation is 22.6. The difference is small in comparison to the amplitude and the standard deviation

This hypothesis may not have been validated through these data; however, through a more qualitative analysis, there may be another response to the test. Additionally, it would also be important to filter among religious students which of them are practitioners of their religions, which ones attend social centers associated with them, among other data that could influence things like specifying the religions.

CALCULATING THE CORRELATON BETWEEN RELIGION AND SUICIDE

Here i had two options, i could group my query by religion and count the numbers of suicide in each group. But having a chi square test is the objective and is a better way to find if there is any correlation between both.   
  
So my SQL query was this one:  
  
SELECT religion, suicide

FROM 'students.csv'

WHERE religion IS NOT NULL

After analyzing the chi square, we can assume that there is no direct correlation between religion and suicides. Our chi square total was of 0,03 a very small amount and our critical value was of 3,841.   
  
Since chi square total value is less then the critical value, we can affirm that there is no real correlation between suicide and religion

Correlation between social connectedness and stress

Since we will be conducting this test with all students, the SQL query was a very simple one.

SELECT tosc, toas

FROM students;

The Pearson Correlation was a value of -0,55 and the covariance was also negative. This indicates that the variables have a correlation where when one increases the other decreases.   
  
This way we can establish that with a better social connectedness the stress can significantly reduce. With a more qualitative and in depth analysis we could seek deeper the question which influences the other more.   
  
There is a possibility where stress influences negatively social connectedness or that a person that has a social life more active suffers less from stress.

total depression score within international students by years living outside home country

The sql queries used to gather data were:   
  
SELECT ROUND(AVG(todep),0) AS avg\_depression\_score, stay

FROM 'students.csv'

WHERE todep IS NOT null AND inter\_dom IS NOT null

GROUP BY stay

ORDER BY stay DESC;

SELECT ROUND(AVG(todep),0) AS avg\_total\_depression

FROM 'students.csv'

The Pearson correlation between the years of stay and the todep is around 0.27. This means that it exists a small correlation between both variables, but not enough to point out a direct connection. To avoid hasty decisions, it would be important to deepen the analysis through additional metrics and conduct a more qualitative study.

Total depression score correlation with stress and social connectedness

The sql queries used to gather data were:

SELECT ROUND(todep,0), ROUND(toas,0), ROUND(tosc)

FROM students.csv

WHERE todep IS NOT null AND toas IS NOT null AND tosc IS NOT null;

The results were:

The Pearson correlation of total depression score and stress = 0,3940945338

The Pearson correlation of total depression score and social connectedness = -0,5517952837

Both have a good correlation.When is stress is higher, the todep tends to increase, when the social score is smaller, the todep tends to increase.

FINAL CONCLUSION

In our analysis, we tested two hypotheses, including the relationship between religion and suicide and sadness, in an attempt to understand if it had any impact on these scores. Apparently, our findings were negative. I do not completely rule it out, as qualitative studies could be conducted to delve deeper into this analysis, differentiating religious students between practitioners and non-practitioners and among various religions.

What stood out the most was the Pearson correlation between accumulated stress and the TOSC (Total Occupational Stress Checklist), which was nearly identical to the correlation between the total depression score and the TOSC. This suggests that, among the areas we examined, the one that most significantly influences people's happiness is the level of contact and social life they have.

From this, we can conclude that among the students in our sample, those who have a more active social life experience less stress and greater happiness.