MENTAL WELL-BEING AMONG UNIVERSITY STUDENTS
Empowering Minds, Nurturing Well-being: From the Classroom to the Community
Name: Aditya Sharma
Student ID: 301558544
Program: Master's in Professional Computer Science in Big Data
Course: CMPT 629 : GRADUATE CAPSTONE PROJECT

Motivation and Background

University students encounter unique mental health challenges, influenced by academic pressures, social dynamics, and global events like the COVID-19 pandemic. This project aims to analyze mental health indicators among university students, particularly focusing on physical exercise, sleep duration, and volunteering activities.

Objectives of the Study:

To investigate demographic correlations with mental health outcomes, examine how academic and lifestyle factors affect student well-being, and develop recommendations for improving mental health support within academic communities.

Survey Form

The data for this study, titled "University Student Mental Health," was collected between September 22 and October 30, 2020, by Jennifer Reeves and Theone Paterson at the University of Victoria, Canada. Comprising 1192 cases and 147 variables, it includes comprehensive information on student demographics, lifestyle habits, and mental health assessments using various scales like CATS, DASS-21, GAD-2, PHQ-2, and PSS.

StartDate	EndDate	Duration (in seconds)	Eligibility	Age	Ethnicity	Ethnicity_text	Sex	Gender	Year_credits	Year_calendar	Program	Part_time	Degree	Disability	Living	Province	International	Emp
2020-09-22 13:57	2020-09-22 14:05	461	1	18	2		2	2	1	1	7	2	1	2	2	5	2	

Problem Statement:
Ques.1) Physical Exercise
Does Increased Engagement in Physical Exercise Causally Impact Mental well-being Among University Students?

Ques.2) Sleep Duration
What is the Causal Effect of Changes in Sleep Duration on Mental Health Outcomes?

Ques. 3) Volunteering
How Does Volunteering Causally Affect Perceived Stress Levels Among students, Controlling for Other Factors.

These questions delve into pivotal aspects of student life: physical activity, sleep, and volunteering, which are fundamental to their mental well-being. Exploring these areas is crucial as they directly influence stress, anxiety, and overall health, reflecting my commitment to improving mental health awareness and support within the educational ecosystem. These topics not only resonate with my personal interests but also address broader public health concerns, making them a significant and relevant choice for research.

Data Science Pipeline:
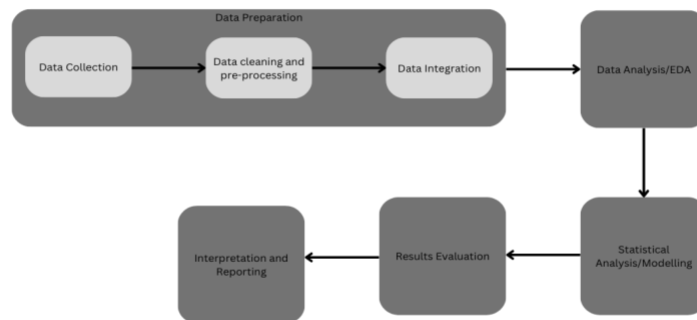
Methodology
Data Preparation:
- **Data Collection:**

For this study, I utilized a comprehensive dataset from the University of Victoria, titled "University Student Mental Health," which includes detailed student responses on mental health indicators collected via online surveys. This dataset encompasses demographic details, mental health assessments, and lifestyle factors such as physical activity, sleep patterns, and volunteering habits. Data collection occurred between September 22 and October 30, 2020, capturing insights from a broad cross-section of the student population across Canada.

- **Data Cleaning:**

Given the extensive nature of the data, meticulous cleaning was imperative to ensure its accuracy and relevance. I removed duplicate entries and handled missing values to maintain data integrity. Further, data normalization was conducted to standardize the diverse formats of student responses, especially in categorical variables like ethnicity, program of study, and lifestyle habits. This step was crucial to harmonize the data, facilitating seamless analysis and comparison.

- **Data Integration/Data Wrangling:**

The integration process was pivotal in consolidating data segments from various survey sections into a cohesive dataset. By employing exact matching techniques, particularly for categorical variables, I ensured that all data elements were correctly aligned and integrated. This effort resulted in a unified and comprehensive dataset, ready for in-depth analysis, capturing the multifaceted dimensions of student mental health and lifestyle practices.

Exploratory Data Analysis:

In my in-depth study on university student mental health indicators, Exploratory Data Analysis (EDA) was pivotal due to the dataset's vastness, encompassing numerous variables. With Tableau, I navigated through a substantial amount of data featuring a wide array of columns representing various aspects of student life. These included demographics, academic details, lifestyle choices, and comprehensive mental health scales.

- **Employment Distribution of Students:**
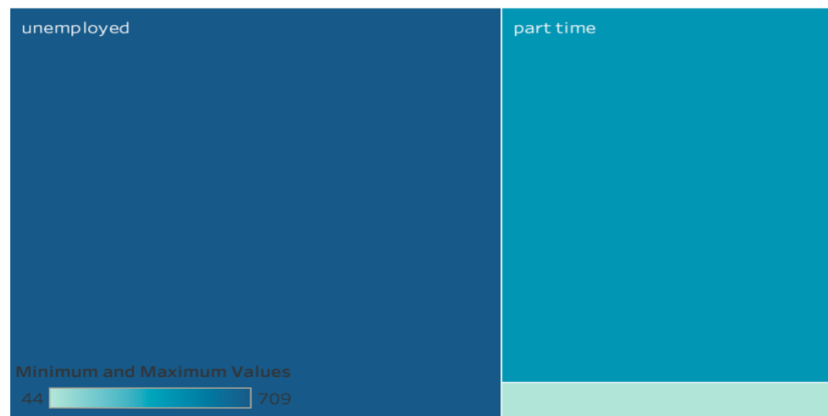
The bar chart meticulously illustrated employment status across the student body. This visualization was particularly telling, shedding light on the potential impact of employment on student mental health and academic engagement.

- **Common Mental Issues:**

Utilizing a word cloud allowed me to emphasize the most frequently reported mental health concerns, such as Anxiety and Depression, standing out due to their size and prominence.

- **Courses Taken by Students:**

Courses Taken up by students

The stream graph offered a dynamic visualization of course enrollments over the years. It showcased trends and patterns in academic paths, which could be linked to stress and mental health trajectories among students.
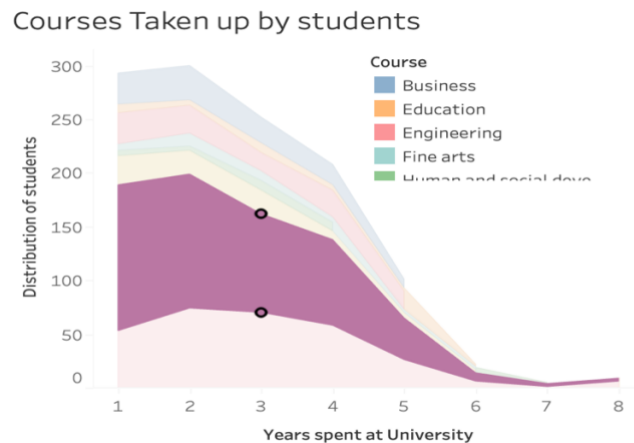These visual representations, drawn from a dataset rich in detail, provided a clearer understanding of the intricate relationships between various factors and student mental health. They underscored the importance of a nuanced approach to student support systems and paved the way for targeted interventions.

Statistical Analysis/Model:

In the data science pipeline for addressing the research questions, I employed several models tailored to the specific nature of each query. **For question 1**, regarding the impact of physical exercise on mental well-being, I utilized regression analysis to explore the causal relationship and quantify the strength of this impact. This approach helped in understanding how variations in physical activity levels influence mental health metrics among university students.

For question 2, which examines the effect of sleep duration changes on mental health, Propensity Score Matching (PSM) was used to create comparable groups based on sleep patterns. This method allowed for an accurate causal inference analysis by matching individuals with similar characteristics but different sleep habits, thereby isolating the effect of sleep duration on mental health outcomes.

In question 3, addressing the role of volunteering in stress management, I employed a Mixed Linear Model (MLM) to analyze the data. This model accounted for the time-variant nature of stress levels and the non-linear effects of volunteering, providing a nuanced view of how volunteering activities, coupled with other life factors, impact perceived stress levels among students.

Results Evaluation and interpretation:

To assess the performance of the models used in this study, I applied various evaluation metrics relevant to each research question.

For question 1, which explored the impact of physical exercise on mental well-being, I used the OLS regression analysis for evaluating the impact of moderate exercise on the total CAMS-R score, a measure of mindfulness, provided valuable insights.
Moderate Exercise: The analysis indicates that an increased frequency of moderate exercise is associated with enhanced mindfulness and potentially improved mental well-being among university students.
**Diagnosis Impact:** The higher mindfulness levels observed in students with mental health diagnoses highlight the crucial need for customized support and intervention strategies for this group.
Gender Considerations: The gender disparities observed shed light on unique challenges and experiences that influence mental well-being, stressing the importance of gender-sensitive support services.

For question 2, investigating the effect of sleep duration on mental health, the effectiveness of the Propensity Score Matching (PSM).
Robustness checks for the study affirmed the treatment effect's stability across a range of caliper widths, as evidenced by consistent T-statistics and P-values. The balance was maintained, confirming the comparability of groups. Additionally, Cohen's d value suggests a moderate effect size. The overlap and balance of propensity scores pre- and post-matching, alongside stratification, validated the treatment effect's robustness within similar propensity score strata.
The analysis suggests a tangible effect of sleep duration on mental health, reinforcing the importance of sleep hygiene in student health policies. The steadfast nature of the results across multiple robustness tests lends credibility to the findings, providing a solid foundation for advocating for policy interventions.
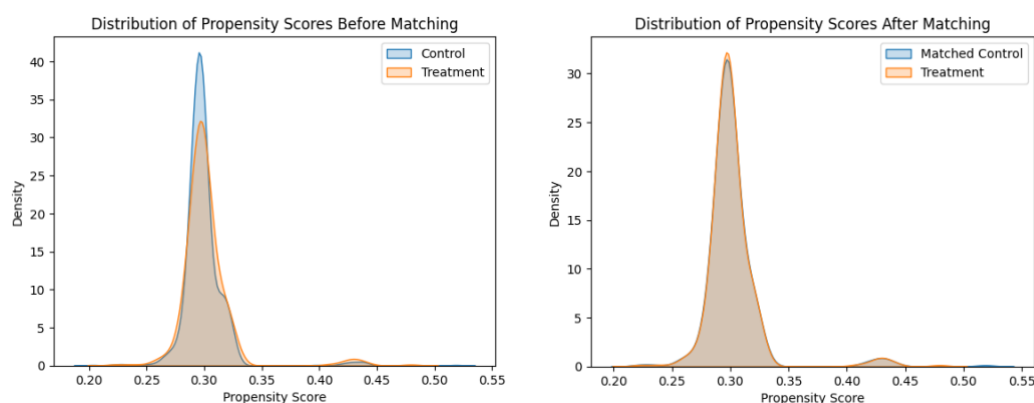
Distribution of propensity scores before and after matching

For question 3, concerning the impact of volunteering on stress levels, I utilized the Durbin-Watson statistic to evaluate the model's performance.

The role of volunteering in mental health may be varied and context-dependent. During COVID-19, increased stress is evident, but volunteering's effect is not distinguishable from other factors Findings contribute to more in depth understanding, important for university policy and program development. Future research should explore long-term effects and identify which aspects of volunteering contribute most to well-being..

The obtained Durbin-Watson statistic value is close to 2, which indicates a minimal level of autocorrelation in the residuals of the Mixed Linear Model (MLM) and suggests that the model's error terms are independent. This value is considered good, as it falls within the acceptable range (approximately 1.5 to 2.5), demonstrating the reliability of the regression analysis results.

Lessons Learnt:

In this project, I delved into the realm of causal analysis and propensity score matching (PSM), techniques that were initially new to me. My exploration into these areas, alongside other statistical methods, significantly bolstered my technical skills.

I discovered how lifestyle factors like physical activity, sleep, and volunteering profoundly influence student mental health. This endeavor not only reinforced the necessity for personalized mental health interventions within academic environments but also provided a hands-on experience in applying complex statistical techniques to real-world issues.

Through this process, I gained a comprehensive understanding of data's role in crafting evidence-based mental health strategies and truly enjoyed the journey of transforming raw data into actionable insights, enhancing my analytical acumen.

Future Work:

For the ongoing enhancement of this study, delving into longitudinal research to capture the evolving dynamics of student mental health is essential. Enlarging the dataset to include a more diverse range of variables could provide deeper insights into the impact of various lifestyle factors. Moreover, developing predictive models that can adapt to the emerging trends in mental health among university students would be advantageous. Such models could help ensure that academic institutions' support systems are adaptive and effective in promoting student well-being.

Summary:
The student mental health project aimed to explore the impact of lifestyle factors such as physical activity, sleep, and volunteering on student mental health. To achieve this, data was gathered from surveys and academic records, followed by rigorous data cleaning to ensure the reliability of the analysis. The study utilized regression analysis to uncover relationships between lifestyle factors and mental health outcomes, employing propensity score matching (PSM) to assess the causal effects of these factors. This methodological approach, supplemented by causal analysis, allowed for a nuanced understanding of how specific lifestyle choices influence student well-being.

Statistical techniques played a crucial role in the project's analysis phase. Various statistical tests, including t-tests and Cohen's d, were used to determine effect sizes and validate findings. Additionally, the PSM technique was calibrated using different caliper widths to test the sensitivity of the results, ensuring their robustness. The analytical process underscored the significance of precise data handling and sophisticated statistical methods in drawing evidence-based conclusions about mental health.

The project not only provided insights into the factors affecting student mental health but also enhanced my technical skills in statistical analysis, data cleaning, and causal inference. It highlighted the importance of personalized mental health interventions and the potential of lifestyle adjustments in improving student well-being. Overall, this endeavor was both a learning journey and a practical application of data science techniques, offering valuable perspectives for developing targeted support strategies in academic environments.