**Topic Title:**

Multiple Linear Regression Analysis on DLSL College Students’ Current Lifestyle Choices towards their Overall Grade Performance

**Topic Background:**

The recent pandemic situation caused a significant impact towards every DLSL student’s lifestyle, resulting to a change of daily habits and routines that were tied to their student life. The abrupt change caused students to partition their daily activities in various ways and timelines to fit their current study at home and quarantine situation. Considering that each student now has different dedicated lifestyles due to the pandemic, it could be an interesting viewpoint to see whether the currently adapted lifestyles of students influence their overall performance in school. From such, our group aims to investigate into variables such as the daily average time they currently dedicate to games, exercise, and schoolwork and identify whether it has a significant impact towards their overall average grade performance in DLSL. By identifying the relationships between these variables, the study would be able to provide insights towards what activities or routines should students dedicate into more to improve their school performance while they live a home-based or hybrid lifestyle.

**Respondents of the Project:**

The respondents chosen for data gathering would be limited to college students of De La Salle Lipa to get a more specific perspective into the current lifestyle and school performance of students at a higher and more mature level, as they have more control into what they want to do in a daily basis. There are no limitations, however, to the year level and course taken by the respondents. The number of respondents will be a minimum of 30 to 40 to ensure that the data would be capable of supporting a minimum acceptable amount of margin of error of 15% (*Budiu and Moran, 2021*).

**Data Information and Gathering Methodology:**

The data to be gathered will be separated between two sections. First section asks for the student’s demographic information, which entails their gender, age, and course taken. The second section refers to questions in relation to their overall average grade from last semester, as well as details regarding their daily activities as a borderless education student. The overall average grade will serve as the dependent variable, while the other lifestyle variables would serve as the independent variables. The demographic data would also be compared to assess whether they have an influence towards these variables. The lifestyle activities that would be asked would be the following:

* Daily average time using computer/mobile devices
* Daily average playtime with games
* Number of games actively playing recently
* Daily average time spent in schoolwork
* Average meal eaten in a day
* Daily average time spent outside of the house
* Daily average time spent being active in social media
* Daily average time spent exercising

Alongside the following questions, the student respondents are also asked to provide questions they have in relation to the provided data, which would be answered in the exploratory data analysis (EDA) process of the project. Five questions provided by the respondents are:

1. Is there a better way to delegate my time wisely [to improve my overall grade performance]?
2. How significant is the relationship between social media hours and one's overall average grades?
3. How high is the correlation between hours spent playing video games to average grade compared to the correlation between hours spent in social media and average grade?
4. Is there a relationship between time spent with games with performance in grades?
5. Which factor is highest/that influences the most when it comes to overall average grades of a student?

The data from the survey will be both discrete (questions with multiple choices) and continuous (grade and time related responses) variables to be used in the EDA and ML model. The data from these variables will undergo data preprocessing if needed, mainly through the replacement of unfit answers using the replace method into more appropriate and usable formats. If unusable or null values are provided, the entry will be removed from the dataset.

For the machine learning model, the algorithm to be used will be a multiple linear regression model, as there will be more than one explanatory or independent variable to be used to predict the outcome or dependent variable, which would be the overall grade performance of the DLSL college students.

The following data will be gathered through a Google Form survey, which can be disseminated through social media.

The google form link containing the responses for the survey is as follows: https://docs.google.com/forms/d/1yy8zI4WU6byMO9kvnfUKyP7Vj6ZPBIxCLQ33psRxoHM/edit

**References:**

* Budiu, R. & Moran, K. (2021). How Many Participants for Quantitative Usability Studies: A Summary of Sample-Size Recommendations. Retrieved from: https://www.nngroup.com/articles/summary-quant-sample-sizes/