

# **The following paragraph summarizes the proposed case study's approaches and methodology, emphasizing exploration**

To analyze the dataset and understand the key factors influencing student dropout rates, I started with an exploratory data analysis (EDA) approach using R programming. The first step involved applying descriptive statistics to gain a comprehensive overview of the dataset, focusing on outliers and missing values. I used summary functions to examine the mean, median, and mode for each column, enabling a clear understanding of the data's central tendencies and distributions. To visualize the data, I plotted box plots and histogram plots for each column, allowing me to identify any irregularities and the overall data spread.

Given the categorical nature of some columns, I employed encoding techniques to convert these categories into numerical values, facilitating further analysis. Once the data was numerically encoded, I created a heatmap to visualize the correlations among the variables. This heatmap provided insights into which columns were most related to the target variable, indicating whether a student had dropped out, enrolled, or graduated. After pinpointing the most important columns, I examined their individual correlations with the target variable, leading to more specific insights into their relationships with student retention and dropout rates.

To summarize the findings, I documented the observations from the heatmap analysis and described the correlations and their potential implications for student outcomes. At the end of the analysis, I provided a summary of the key insights derived from the exploration and pre-processing steps.

Throughout the analysis, I utilized several R libraries to streamline the process, including `'dplyr'` for data manipulation, `'ggplot2'` for data visualization, `'readr'` for reading data files, and `'tidyr'` for tidying the dataset. These libraries were instrumental in efficiently conducting the exploratory and pre-processing tasks and generating the visualizations required for deeper analysis.