Data Source

The dataset comes from Virtual Internships, which conducts web-based simulations for university students to replicate real-world professional scenarios. Specifically, it involves interns at Nephrotex, a fictional biomedical engineering firm, where teams develop a prototype device for kidney failure patients. Their tasks include research, understanding stakeholder requirements, prototyping, testing, and design justification. Teams use an online chat tool for communication and receive mentor support.

Principal Component Analysis (PCA)

PCA was used to reduce data dimensions, with 3 principal components explaining over 95% of the variance. Scatter plots of principal components revealed weak patterns and structures, indicating low association with mean outcome scores. A heatmap showed that total messages had a strong positive relationship with PC1, implying that group interaction improves outcomes. Total mentor interactions positively influenced PC3, while total asking questions negatively impacted PC2. This leads us to believe that groups ask more questions because they aren’t fully equipped with the knowledge for the task at hand, explaining the negative influence on outcome score.

Multiple Linear Regression (MLR)

We conducted MLR on a group-level dataset with 7 variables to predict the mean outcome score. Both raw and normalized datasets indicated non-linear relationships, leading to poor model performance. The analysis of coefficients showed that the most influential variables were group total performance parameters, asking questions, and total customer consultant requests. Multiplying the coefficient values by their standard deviation for each feature confirmed their importance. Groups aiming to meet customer consultant requests and performance parameters achieved higher scores, while those asking more questions scored lower, as mentioned from PCA.

PCA with logistic regression

Upon applying logistic regression, the model correctly predicted 60% of the cases, performing below average. The classification report indicates poor performance for classes 2.0 and 3.0, but relatively better performance for class 4.0.