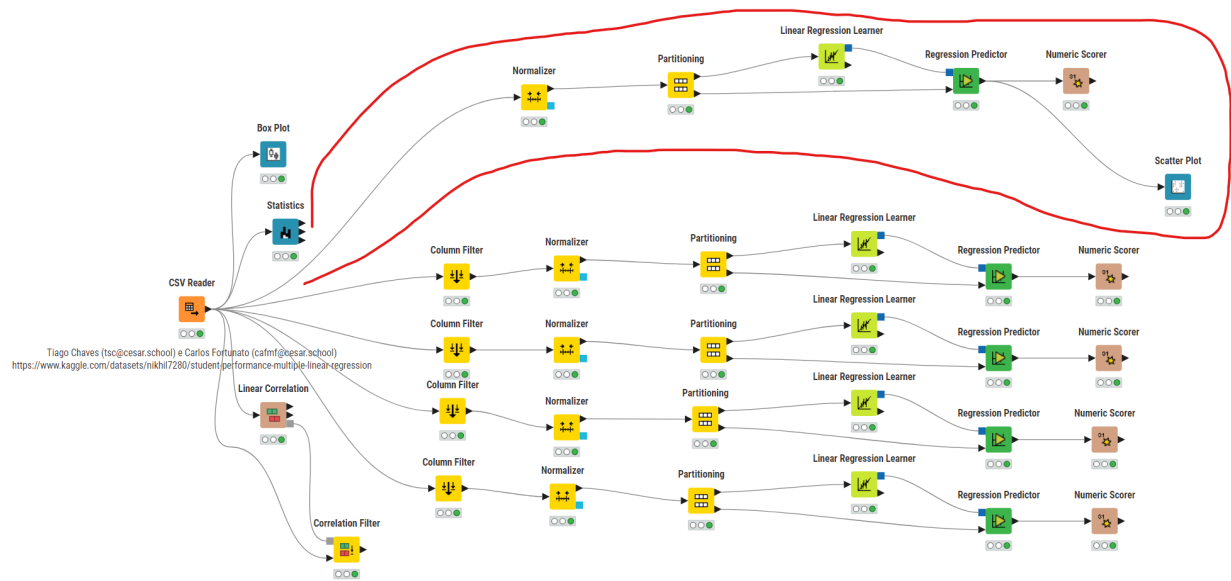
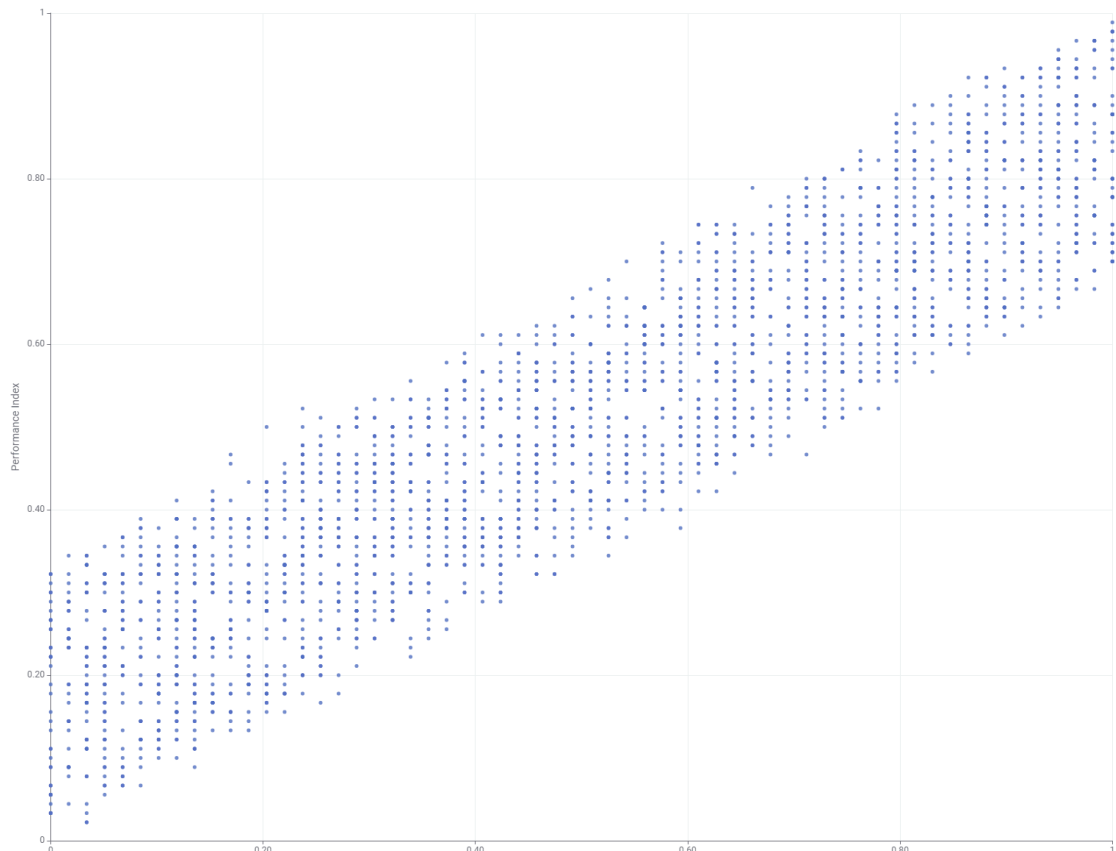


Tiago Chaves (tsc@cesar.school) e Carlos Fortunato (cafmf@cesar.school)
<https://www.kaggle.com/datasets/nikhil7280/student-performance-multiple-linear-regression>



Scatter Plot



No desenvolvimento do fluxo de trabalho para prever o Performance Index dos estudantes, utilizamos o dataset que contém variáveis como Hours Studied, Previous Scores, Extracurricular Activities, Sleep Hours e Sample Question Papers Practiced. Inicialmente, foi realizada uma análise exploratória para identificar a correlação entre as variáveis e a variável alvo. A coluna Previous Scores destacou-se como a mais correlacionada com o índice de desempenho, enquanto outras variáveis apresentaram correlações menores. Para avaliar o impacto dessas variáveis, foram implementados diversos fluxos no KNIME, testando tanto a inclusão quanto a exclusão de colunas preditoras no modelo. Surpreendentemente, os melhores resultados foram obtidos ao incluir todas as variáveis no treinamento, mesmo com algumas apresentando baixa correlação individual. O modelo final, utilizando todas as colunas, apresentou métricas robustas, como R^2 de 0.988, Mean Absolute Error (MAE) de 0.0179, e Root Mean Squared Error (RMSE) de 0.0223, indicando excelente precisão na previsão do desempenho estudantil.