

# Relatório do exercício proposto

## Dataset

Utilizei o LibreOffice Calc para abrir e interpretar o dataset que me foi fornecido. Após ver a dataset, fiquei com dificuldades em interpretar os mesmos, visto que não há nenhuma informação relacionado com dias, ou pelo menos eu não consegui interpretar dessa forma. Como tal, tive de arrajar forma de alterar o dataset de forma a que se pudessem verificar algumas informações por dia.

## Transformação do Dataset

Para alterar o dataset utilizei a linguagem de programação Python. A minha ideia foi olhar para o tempo das viagens, e quando a soma das viagens ultrapassasse 7h (horário normal de trabalho é 8h, mas pelo menos uma é de almoço), então toda essa informação seria referente a um dia.

## Knime

Utilizei o software Knime para calcular a previsão de custos e do número de viagens, pois já trabalhei com este software anteriormente. Fiz um cross validation para treinar melhor o modelo e utilizei o algoritmo Gradient Boosting Machine para fazer a previsão, tanto das viagens, como do valor total a receber. Os resultados foram os seguintes:

- Número de viagens que se irá fazer no dia seguinte:

Row ID	I trips	I passen...	I trip_tim...	I trip_dis...	D fare_a...	D Predicti...
27885	?	?	?	?	?	40.63

- Valor total a receber no dia seguinte:

Row ID	I trips	I passen...	I trip_tim...	I trip_dis...	D fare_a...	D Predicti...
27885	?	?	?	?	?	433.774