

# Trabalho Em Grupo

## Parte 2



ADM1007 - TURMA U


---

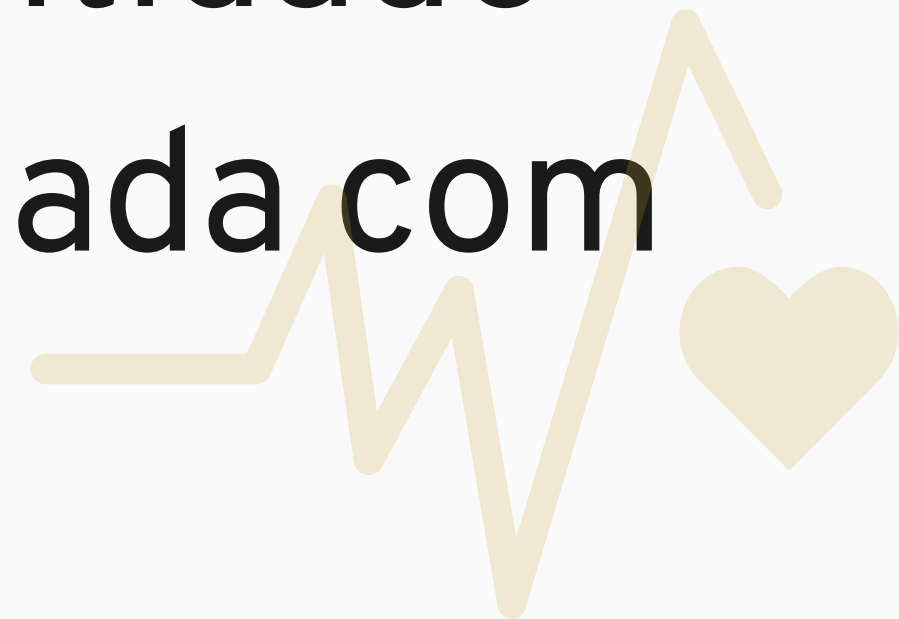
BRENDA FARIAS, NAIOME ALVES



- Objetivo: estudar a área da saúde, para identificar os municípios brasileiros mais adequados para a abertura de novos negócios nessa área.
- Desafio: muitos gaps na área da saúde.



- Optamos por analisar os dados sobre  crianças que nasceram com algum tipo de anomalia, para identificar quais variáveis estão afetando a ocorrência desse fenômeno.
- Por exemplo: a falta de uma quantidade mínima de médicos está relacionada com esse fenômeno?



- Nesses municípios, pode-se encontrar oportunidades para abertura de clínicas que ofereçam serviços essenciais de pré-natal para mulheres grávidas.



# pré-processamento & análise exploratória



- Foi utilizado dados no Tabnet - DATASUS e IBGE com dados sobre população, área territorial, idh, mortalidade infantil, quantidade de hospitais, ub's, e diversos profissionais da saúde por município.





- Foi realizado a limpeza dos arquivos csv, eliminando cabeçalhos e rodapés e combinando mais de um banco de dados; houve tratamento de caracteres não numerais e a codificação para latin-1.




- a *Análise 1* dataframe nos apresenta 5570 linhas correspondente a quantidade de municípios brasileiros, distribuída em 31 variáveis.
- O dataframe não possui dados nulos e os tipos de dados são todos numéricos, exceto o nome dos municípios.



- Foi plotado o gráfico de Person para descobrir as correlações, o gráfico apresentou que quanto maior a estimativa populacional, maior a quantidade de profissionais da saúde no município.





- A partir da análise exploratória,   
verificamos a relação entre a  
disponibilidade de profissionais da saúde  
nos municípios com a porcentagem de  
recém nascidos com algum tipo de  
anomalia. Adicionamos uma nova tabela  
do Data SUS com os dados sobre  
nascimento nos municípios.



- Foi criada uma nova coluna ANOM (crianças que nasceram com algum tipo de anomalia), e elas foram calculadas com o total de nascimentos - a quantidade de crianças sem anomalia sendo igual nascidos com anomalia em porcentagem.



- Para descobrirmos quais são as possíveis variáveis que podem estar interferindo na quantidade de nascidos vivos com anomalia, utilizaremos modelos de predição por aprendizado de máquina.



- Esse é um problema de classificação e como análise inicial utilizaremos modelos baseados em construção de árvores de decisão: Decision Tree e Random Forest.





- Foi decidido separar o dataset em 10 partes para que cada vez que o algoritmo rodar ele tente acertar em 10%.



- O modelo está errando em média 1,30 unidades em relação aos valores verdadeiros, com um erro quadrático médio de 2,86 e um RMSE de 1,69.



- A média das pontuações de validação cruzada é 1.0, que é a pontuação máxima possível para este caso. O desvio padrão é 0, indicando que todas as pontuações de validação cruzada são exatamente iguais.

