

PROJETO DE ANÁLISE DE DADOS DO NAUFRÁGIO DO TITANIC

CONJUNTO DE DADOS UTILIZADO

O conjunto de dados utilizado foi o titanic-data-6.csv obtido no menu “2.detalhes do projeto”.

QUESTÃO LEVANTADA PARA A ANÁLISE DE DADOS

O processo de análise de dados tem como base um arquivo csv com registros de 891 passageiros abordo do Titanic. Sendo assim, o principal foco desta análise é investigar e descobrir se os passageiros da primeira classe, julgados como os mais ricos, tiveram prioridades para embarcar nos botes salva-vidas e se salvarem, Logo a questão levantada para o projeto é:

Os mais ricos tiveram prioridades para embarcar nos botes salva-vidas e se salvarem?

DESCRIÇÃO DA INVESTIGAÇÃO

Primeiramente, foi realizado um comparativo por classes sociais e gêneros de quantos passageiros sobreviveram identificando que o fator da classe social não era relevante para a sobrevivência.

Obtendo este retorno, através do mesmo raciocínio de obter os sobreviventes por classe e gênero, todo o conjunto foi segregado por faixa etária abstraindo ainda mais a informação para definir qual faixa etária estava mais propensa à sobrevivência.

LIMPEZA DE DADOS

Para dar início a análise de dados foi necessário realizar uma limpeza nos registros como por exemplo o campo “Age” onde haviam uns campos nulos que foram preenchidos com a média de idade de todos os passageiros

Além disso, para facilitar a visualização dos dados, foram removidas algumas colunas que não seriam relevantes para o processo de análise.

Houve também uma validação de registros duplicados que no caso não foi necessário aplicar a remoção de nenhum registro por não haver duplicidade.

CONCLUINDO E COMUNICANDO

Após o processo de limpeza e exploração, ficou um tanto quanto claro que em proporção, não houve uma prioridade de salvação por classes sociais e sim por gênero e por faixa etária onde idosos e crianças do sexo feminino tiveram uma diferença na quantidade de sobreviventes superior a de adultos do sexo masculino.

Abaixo encontra-se alguns gráficos que mostram o comparativo de sobreviventes por classe e gênero.

