

PROBLEMA: ANÁLISE DE SOBREVIVÊNCIA NO TITANIC (4ptos)

CONTEXTO DO PROBLEMA

No dia 15 de abril de 1912, o RMS Titanic, considerado "inafundável", colidiu com um iceberg e afundou em sua viagem inaugural de Southampton para Nova York. Do total de 2.224 passageiros e tripulantes, apenas 705 sobreviveram, resultando em uma das maiores tragédias marítimas da história.

Problema ou Problema de Negócio:

Como cientista de dados contratado pela comissão de investigação do acidente, você foi encarregado de analisar os padrões de sobrevivência para responder à pergunta fundamental:

“QUAIS TIPOS DE PESSOAS TÊM A MAIOR PROBABILIDADE DE SOBREVIVER AO NAUFRÁGIO DO TITANIC?”

Objetivo Principal:

Identificar e caracterizar os perfis demográficos e socioeconômicos que tiveram maiores taxas de sobrevivência durante o naufrágio.

Objetivos Específicos:

- Análise Demográfica: Investigar a relação entre gênero, idade e sobrevivência
- Análise Socioeconômica: Examinar o impacto da classe social na probabilidade de sobrevivência
- Análise Familiar: Verificar como a composição familiar influenciou as chances de sobrevivência
- Análise Geográfica: Explorar padrões relacionados ao porto de embarque

CRITÉRIOS DE AVALIAÇÃO

Pré-processamento (30%)

- Tratamento adequado de missing values
- Criação de features relevantes
- Transformações apropriadas das variáveis
- Documentação das decisões de pré-processamento.

Visualização (30%)

- Qualidade e clareza das visualizações
- Diversidade de tipos de gráficos
- Interpretação correta dos padrões visuais

Análise e Insights (40%)

- Resposta clara à pergunta central
- Identificação de perfis de sobrevivência
- Fundamentação estatística das conclusões
- Recomendações baseadas em evidências

Nota: Dentro da visualização que você fizer você deve ter obrigatoriamente o seguinte:

VISUALIZAÇÕES OBRIGATÓRIAS

1. ANÁLISE DEMOGRÁFICO

- Gráfico de barras: Taxa de sobrevivência por gênero
- Histograma: Distribuição de idades por sobrevivência
- Boxplot: Idade vs Sobrevivência por classe social

2. ANÁLISE SOCIOECONÔMICA

- Gráfico de barras: Taxa de sobrevivência por classe
- Scatter plot: Tarifa vs Idade (colorido por sobrevivência)
- Heatmap: Matriz de correlação entre variáveis.

3. ANÁLISE FAMILIAR E RELACIONAL

- Gráfico de barras: Taxa de sobrevivência por porto de embarque
- Mapa de calor: Distribuição geográfica dos passageiros

Observação: deve enviar um arquivo do **Jupyter Notebook** (**envios em outro formato não serão avaliados**) com documentação completa do seu trabalho, bem como a interpretação de cada análise e gráfico, conclusões e referências utilizadas.

Deve apresentar seu trabalho em uma apresentação de 25 e 30 minutos.

Dicionário de Dados

PassengerId : Identificador único do passageiro

Survived: Sobreviveu (1) ou não (0) - VARIÁVEL TARGET

Pclass: Classe do ticket (1 = 1^a classe, 2 = 2^a classe, 3 = 3^a classe)

Name : Nome completo do passageiro

Sex: Gênero (male/female)

Age: Idade em anos

SibSp: Número de irmãos/cônjuges a bordo

Parch: Número de pais/filhos a bordo

Ticket: Número do ticket

Fare: Tarifa paga

Cabin: Número da cabine

Embarked: Porto de embarque (C = Cherbourg, Q = Queenstown, S = Southampton)