**Análise Avançada dos Dados do Titanic e Criação de Cubos**

**Contexto**

Você recebeu um dataset contendo informações sobre os passageiros do Titanic. Sua tarefa é analisar os dados de maneira aprofundada, integrando fontes externas, criando um cubo OLAP complexo, e realizando análises preditivas. O objetivo é praticar a modelagem dimensional avançada, a integração de dados e o uso de técnicas de Business Intelligence para insights mais profundos.

**Descrição dos Campos do Dataset do Titanic**

1. PassengerId:
   * Descrição: Identificador único para cada passageiro.
   * Tipo: Numérico (inteiro)
   * Exemplo: 892
2. Survived:
   * Descrição: Indicador de sobrevivência do passageiro (0 = Não sobreviveu, 1 = Sobreviveu).
   * Tipo: Numérico (inteiro)
   * Exemplo: 1
3. Pclass:
   * Descrição: Classe de ticket do passageiro (1 = Primeira classe, 2 = Segunda classe, 3 = Terceira classe).
   * Tipo: Numérico (inteiro)
   * Exemplo: 3
4. Name:
   * Descrição: Nome do passageiro.
   * Tipo: Texto (string)
   * Exemplo: "Kelly, Mr. James"
5. Sex:
   * Descrição: Gênero do passageiro.
   * Tipo: Texto (string)
   * Exemplo: "male"
6. Age:
   * Descrição: Idade do passageiro em anos.
   * Tipo: Numérico (decimal)
   * Exemplo: 47.0
7. SibSp:
   * Descrição: Número de irmãos/cônjuges a bordo com o passageiro.
   * Tipo: Numérico (inteiro)
   * Exemplo: 1
8. Parch:
   * Descrição: Número de pais/filhos a bordo com o passageiro.
   * Tipo: Numérico (inteiro)
   * Exemplo: 0
9. Ticket:
   * Descrição: Número do ticket do passageiro.
   * Tipo: Texto (string)
   * Exemplo: "330911"
10. Fare:

* Descrição: Tarifa paga pelo passageiro.
* Tipo: Numérico (decimal)
* Exemplo: 78292.0

1. Cabin:

* Descrição: Número da cabine do passageiro.
* Tipo: Texto (string)
* Exemplo: "C85"

1. Embarked:

* Descrição: Porto de embarque do passageiro (C = Cherbourg, Q = Queenstown, S = Southampton).
* Tipo: Texto (string)
* Exemplo: "S"

**Passos**

**Passo 1: Exploração Inicial dos Dados**

1. Carregue o dataset e visualize as primeiras linhas.
2. Verifique se há valores ausentes e trate-os adequadamente.
3. Explore as estatísticas descritivas dos dados, como média, mediana, desvio padrão, etc.

**Passo 2: Limpeza e Transformação dos Dados**

1. Substitua valores ausentes de acordo com a estratégia escolhida (média, mediana, valor fixo, etc.).
2. Crie novas colunas a partir das existentes, como:
   * Faixa etária (crianças, adolescentes, adultos, idosos)
   * Família a bordo (com base em SibSp e Parch)
3. Converta colunas categóricas em numéricas, se necessário, para facilitar a análise.
4. Normalize ou padronize as colunas de valores numéricos.

**Passo 3: Análise Exploratória de Dados (EDA)**

1. Crie gráficos para visualizar a distribuição dos dados, como:
   * Histograma da idade
   * Gráfico de barras para classe (Pclass)
   * Gráfico de barras para sobrevivência (Survived)
2. Analise a correlação entre variáveis, como a relação entre a classe do passageiro e a taxa de sobrevivência.
3. Use técnicas de visualização avançadas, como gráficos de violino, gráficos de densidade e mapas de calor.

**Passo 4: Integração de Fontes Externas**

1. Encontre e integre dados externos relevantes, como informações meteorológicas da noite do naufrágio, detalhes sobre a tripulação ou dados socioeconômicos da época.
2. Combine esses dados com o dataset do Titanic para enriquecer a análise.

**Passo 5: Modelagem Dimensional e Criação de Cubos**

1. Identifique as dimensões e fatos para a criação do cubo. Possíveis dimensões:
   * Dimensão Tempo: Ano, Mês, Dia
   * Dimensão Passageiro: Idade, Sexo, Classe, Embarque
   * Dimensão Ticket: Número do Ticket, Tarifa
   * Dimensão Externa: Dados meteorológicos, Socioeconômicos
2. Crie um modelo dimensional (esquema estrela ou floco de neve) baseado nas dimensões identificadas.
3. Use uma ferramenta de BI (como Microsoft Power BI, Tableau, etc.) para criar um cubo OLAP e carregue os dados.
4. Considere a otimização do desempenho do cubo, como a criação de agregados e índices.

**O seu cubo precisa responder:**

* + Qual a taxa de sobrevivência por classe?
  + Qual a distribuição de idade entre os passageiros que sobreviveram e os que não sobreviveram?
  + Como a tarifa do ticket influencia a sobrevivência?
  + Qual o impacto das condições meteorológicas na sobrevivência?
  + Crie relatórios e dashboards que visualizam essas análises.
  + Otimize as consultas OLAP para melhorar o desempenho.

**Passo 7: Análise Preditiva**

1. Utilize técnicas de machine learning para criar um modelo preditivo que estime a probabilidade de sobrevivência dos passageiros.
   * Divida os dados em conjunto de treino e teste.
   * Utilize algoritmos como regressão logística, árvores de decisão ou redes neurais.
   * Avalie o modelo usando métricas apropriadas (acurácia, precisão, recall, F1-score).