

# Relatório Sobre Análise de Dados do Titanic

Uma imagem com transporte, embarcação, Arquitetura naval, navio

Descrição gerada automaticamente

Atec - TPSICAS1123 - 5417

Relatório realizado por: João Pedro Fernandes Silva, Ricardo Lopes da Conceição e Diogo Baptista Louro

Formador: Nelson Alexandre Santos

Índice

[Introdução 2](#_Toc184262567)

[Estrutura dos dados 2](#_Toc184262568)

[Leitura e Exploração dos Dados 3](#_Toc184262569)

[Limpeza e Pré-Processamento 4](#_Toc184262570)

[Análise e Manipulação dos Dados 5](#_Toc184262571)

[Conclusão 10](#_Toc184262572)

Introdução

Neste relatório analisámos um conjunto de dados utilizando Python e bibliotecas como Pandas para manipulação e análise de dados e Matplotlib para visualizações gráficas. O conjunto de dados fornecido contém informações sobre os passageiros como a idade, sexo, classe de embarque e se sobreviveram ou não.

O objetivo deste projeto é realizar uma análise com uma exploração inicial dos dados e apresentação de insights relevantes. O relatório inclui etapas de limpeza e pré-processamento dos dados, criação de métricas adicionais e visualizações gráficas.

Estrutura dos dados

Os dados fornecidos contêm 12 colunas sobre as características dos passageiros e a sobrevivência dos mesmos. Analisando os dados temos o seguinte:

* PassengerId: identificador único;
* Survived: 1 (sobreviveu) ou 0 (não sobreviveu);
* Pclass: classe do passageiro (1ª, 2ª ou 3ª classe);
* Name: nome do passgeiro;
* Sex: género do passageiro;
* Age: idade dos passageiros;
* SibSp: número de irmãos e esposo/a do passageiro;
* Parch: números de pais e filhos do passageiro;
* Ticket: número do bilhete do passageiro;
* Fare: taxa cobrada por passageiro;
* Cabin: número da cabine do passageiro;
* Embarked: porto de embarque C, Q ou S.

O dataset foi carregado com pd.read\_csv. Os métodos .head(), .tail(), .describe() e .info() foram usados para analizar a estrutura de dados, identificar valores nulos e compreender o formato das colunas.

Leitura e Exploração dos Dados

Uma imagem com texto, captura de ecrã, Tipo de letra

Descrição gerada automaticamenteOs daddos foram carregados com pd.read\_csv() e inspecionados utilizando .head, .tail, .describe e .info respetivamente, obtendo os seguintes resultados:

Figura 1 - .head, .tail e .desscribe

Uma imagem com texto, captura de ecrã, Tipo de letra, design

Descrição gerada automaticamente

Figura 2 - .info

Limpeza e Pré-Processamento

Para tratar dos valores nulos e ajustar os dados na coluna Age os valores ausentes foram preenchidos com a mediana das idades, a coluna Cabin foi descartada na análise inicial devido à alta quantidade de valores ausentes e na coluna Embarked os valores ausentes foram preenchidos com o valor mais frequente.

Uma imagem com texto, captura de ecrã, software

Descrição gerada automaticamente

Depois criamos uma coluna chamada *Idade\_Milissegundos*, onde as idades foram convertidas para milissegundos.

Uma imagem com texto, captura de ecrã, Tipo de letra

Descrição gerada automaticamente

Figura 3 – idades em milisegundos

Análise e Manipulação dos Dados

Realizamos algumas análises com as funções do Pandas, como *groupby* para calcular a taxa de sobrevivência por classe e género. Foi possível observar que a classe não interferiu na taxa de sobrevivência. Relativamente ao género, as mulheres apresentaram uma taxa de sobrevivência absoluta com nenhum homem a sobreviver.

Uma imagem com texto, captura de ecrã, Tipo de letra, número

Descrição gerada automaticamente

Figura 4 – taxa de sobrevivência por classe e género

De seguida analisámos qual a idade média de um sobrevivente o que nos fez concluir que a probabilidaddde de sobreviver é mais favorável para pessoas mais jovens.

Uma imagem com texto, Tipo de letra, captura de ecrã

Descrição gerada automaticamente

Figura 5 – média da idade dos sobreviventes

Analizando as idades com a tarifa e sobrevivência podemos observar que a tarifa não teve impacto na taxa de sobrevivência pois maior parte dos sobreviventes pagaram uma tarifa Fare menor que 100.

Uma imagem com texto, captura de ecrã, Gráfico, diagrama

Descrição gerada automaticamente

Gráfico 1 – Idade vs tarifa, com os sobreviventes assinalados com pontos azuis.

Ao analisarmos o gráfico obtido da taxa de sobrevivência por porto de embarque concluímos que os passageiros que embarcaram em Queenstown tiveram uma maior probabilidade de sobreviver seguido do porto de Cherbourg e por fim os passageiros de Southampton tiveram a menor probabilidade de sobrevivência.

Uma imagem com texto, captura de ecrã, diagrama, Gráfico

Descrição gerada automaticamente

Gráfico 2 – Taxa de Sobrevivência por Porto de Embarque.

Uma imagem com file, Gráfico, diagrama, texto

Descrição gerada automaticamenteNo gráfico seguinte podemos ter uma visão mais clara sobre a classe e sexo dos sobreviventes onde imediatamente podemos concluir que relativamente às mulheres houve um número muito baixos de sobreviventes de segunda classe e um número mais alto para mulheres de terceira classe.

Gráfico 3 – Distribuição de Sobreviventes por Classe e Sexo.

Uma imagem com file, Gráfico, diagrama, captura de ecrã

Descrição gerada automaticamenteNo histograma seguinte podemos visualizar melhor a distribuição de idades sobre a taxa de sobrevivência

Gráfico 4 – Distribuição de Idades.

Neste histograma podemos visualizar a distribuição das tarifas sobre a taxa de sobreviventes.

Uma imagem com texto, captura de ecrã, Gráfico, file

Descrição gerada automaticamente

Gráfico 5 – Distribuição de Tarifas.

E neste gráfico podemos observar melhor a diferença do número de passageiros que sobreviveram e o número de passageiros que não sobreviveram em que claramente houve muitos mais passageiros que não sobreviveram.

Uma imagem com captura de ecrã, texto, diagrama, file

Descrição gerada automaticamente

Gráfico 6 – Distribuição de sobrevivência.

Uma imagem com texto, captura de ecrã, diagrama, file

Descrição gerada automaticamenteUma análise adicional que fizemos foi explorar qual a influência do tamanho da família na taxa de sobrevivência e obtemos o seguinte gráfico que mostra que um número de membros familiares de 4 e 6 apresentam uma taxa de sobrevivência relativamente mais elevada que passageiros com 1, 5 e 7 membros familiares com os restantes valores por volta de 50% de probabilidade de sobrevivência.

Gráfico 7 – Taxa de sobrevivência por tamanho da família.

# Conclusão

Neste relatório analisámos os dados sobre os passageiros do Titanic e desta forma conseguimos identificar padrões que influenciaram a sobrevivência ao naufrágio de 1912. Nos dados analizados encontramos claras evidências da política "mulheres e crianças primeiro” com taxas de sobrevivência mais altas para mulheres e passageiros mais jovens. Os passageiros da 3ª e 1ª classe tiveram mais sobreviventes que 2ª classe. Relativamente a tarifas não houve um impacto o sendo que os que tiveram tarifas inferiores a 100 tinham maior taxa de sobrevivência.