

Relatório de Análise Exploratória de Dados

Titanic Dataset

1. Introdução.....	2
2. Objetivo.....	2
2.1 Objetivos específicos.....	2
3. Base de Dados.....	3
5. Metodologia.....	3
5.1 Carregamento e Exploração Inicial.....	4
5.2 Verificação de dados ausentes (sem tratamento aplicado).....	4
5.3 Análise exploratória de dados.....	5
6. Resultados e Discussões.....	6
6.1 Sobrevivência geral.....	6
6.2 Sobrevivência por sexo.....	6
6.3 Sobrevivência por classe social.....	6
6.4 Sobrevivência por idade.....	6
6.5 Crianças e Adultos.....	6
6.6 Classe social + Crianças.....	7
7. Conclusão.....	7

1. Introdução

O trabalho tem como objetivo realizar uma análise exploratória de dados utilizando o [conjunto de dados do Titanic](#). A base contém informações reais sobre passageiros do navio Titanic, que afundou em 1912 após colidir com um iceberg, resultando em milhares de mortes. A partir dos dados disponíveis, é possível investigar quais fatores podem ter influenciado a sobrevivência dos passageiros, como sexo, idade e classe social. Para isso, foram utilizadas técnicas de manipulação e análise de dados com a linguagem Python, além da criação de gráficos para facilitar a interpretação das informações.

Esse tipo de análise é importante pois permite compreender padrões e tendências presentes no conjunto de dados, auxiliando na tomada de decisões e servindo como base para etapas mais avançadas, como a preparação dos dados e a aplicação de modelos de previsão.

2. Objetivo

Realizar uma análise exploratória de dados utilizando o dataset do Titanic, com o intuito de identificar padrões e fatores que influenciaram a sobrevivência dos passageiros.

2.1 Objetivos específicos

- Analisar a taxa geral de sobrevivência dos passageiros
- Verificar a relação entre sobrevivência e sexo dos passageiros
- Avaliar a influência da classe social (Pclass) na sobrevivência
- Analisar a distribuição de idades e sua relação com a sobrevivência
- Criar uma variável para diferenciar crianças e adultos e comparar suas taxas de sobrevivência
- Gerar gráficos para facilitar a visualização e interpretação dos resultados obtidos

3. Base de Dados

Nesse projeto foi utilizado o conjunto de dados Titanic Dataset, disponibilizado na plataforma Kaggle, amplamente utilizado em estudos introdutórios de Análise de Dados e Machine Learning. O dataset contém informações sobre passageiros que estavam a bordo do navio Titanic, incluindo características pessoais e socioeconômicas. A base possui 891 registros, representando passageiro, e 12 colunas, contendo variáveis relevantes para análise, como sexo, idade, classe do bilhete e status de sobrevivência.

As principais colunas utilizadas durante a análise forma:

- Survived: indica se o passageiro sobreviveu (1) ou não (0)
- Pclass: classe do passageiro (1ª, 2ª ou 3ª classe)
- Sex: sexo do passageiro (male/female)
- Age: idade do passageiro
- Fare: valor pago na passagem
- Embarked: porto de embarque

4. Ferramentas utilizadas

Para a realização deste projeto, foram utilizadas ferramentas e bibliotecas voltadas para análise e visualização de dados, sendo elas:

- Python: linguagem principal utilizada para a manipulação e análise dos dados.
- Jupyter Notebook: ambiente utilizado para desenvolver o código e documentar o processo de análise
- Pandas: bibliotecas utilizada para leitura do dataset, manipulação de tabelas e cálculos estatísticos
- Matplotlib: biblioteca utilizada para criação de gráficos e visualizações dos resultados

Essas ferramentas permitiram realizar a análise exploratória, organizar informações importantes e gerar gráficos para facilitar a interpretação dos dados

5. Metodologia

Nesta seção, descrevemos de forma ordenada as etapas realizadas durante o desenvolvimento deste projeto, desde o carregamento dos dados até a geração de análises e visualizações.

5.1 Carregamento e Exploração Inicial

O conjunto de dados utilizados foi carregado diretamente do arquivo CSV disponibilizado no repositório do projeto. A leitura foi realizada utilizando a biblioteca Pandas, que permitiu transformar os dados em um Data Frame para análise.

Após o carregamento, foram realizadas análises iniciais com o objetivo de compreender melhor a estrutura do conjunto de dados, como:

- Verificação da quantidade de linhas e colunas com **shape**

```
df.shape
✓ 0.0s
(891, 12)
```

- Visualização das primeiras linhas com head()

```
df.head()
✓ 0.0s
```

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	S

- Verificação de valores ausentes para identificar quais colunas precisavam de tratamento

Essa etapa foi essencial para entender quais variáveis poderiam impactar na sobrevivência e quais problemas de dados precisam ser resolvidos

5.2 Verificação de dados ausentes (sem tratamento aplicado)

Após a exploração inicial, foi realizada a limpeza e preparação dos dados para permitir análises mais confiáveis

As principais ações nessa etapa forma:

- Padronização de informações necessárias para análise
- Criação de variáveis auxiliares para facilitar comparações e agrupamentos

Além disso, foi criada uma variável categórica para classificar os passageiros em dois grupos

- Criança: idade menor que 18 anos
- Adulto: idade maior ou igual 18 anos

Essa divisão foi feita para observar se a idade influenciou diretamente a chance de sobrevivência

5.3 Análise exploratória de dados

Com os dados organizados, foi realizada a etapa de análise exploratória, com foco em identificar padrões de sobrevivência.

Foram analisadas as seguintes comparações:

- Sobrevivência geral

Foi calculada a proporção total de passageiros que sobreviveram e os que não sobreviveram

- Sobrevivência por sexo

Foi analisado o impacto de variáveis Sex, comparando a taxa de sobrevivência entre homens e mulheres

- Sobrevivência por idade

A idade foi analisada para entender se pessoas mais novas ou mais velhas tiveram mais novas ou mais velhas tiveram mais chances de sobreviver

- Sobrevivência por classe social

Foi feita uma análise de variáveis Pclass, observando como a classe social influenciou a sobrevivência

- Sobrevivência entre crianças e adultos

Após a criação do grupo Criança/Adulto, foi feita uma comparação direta para identificar diferenças na taxa de sobrevivência

- Cruzamentos de variáveis

Por fim, foram realizados cruzamentos para aprofundar a análise e encontrar padrões mais específicos, como:

Sexo + Criança

Classe social + Criança

Esses cruzamentos ajudaram a entender melhor o comportamento dos dados e identificar quais grupos foram mais favorecidos durante o desastre.

6. Resultados e Discussões

Nesta etapa foram realizadas análises exploratórias utilizando gráficos para visualizar padrões de sobrevivência entre os passageiros do Titanic. Os resultados foram organizados em tópicos para facilitar a interpretação

6.1 Sobrevivência geral

A análise geral mostra que a maioria dos passageiros não sobreviveu ao desastre.

6.2 Sobrevivência por sexo

Ao analisar a sobrevivência por sexo, foi possível perceber que mulheres apresentaram uma taxa de sobrevivência muito maior do que os homens. Esse resultado indica que houve prioridade no resgate para mulheres, seguindo o princípio conhecido como “mulheres e crianças primeiro”

6.3 Sobrevivência por classe social

A análise por classe social mostra que passageiros da 1 classe tiveram maior chance de sobrevivência, enquanto a 3 classe apresentou a menor taxa. Isso pode estar relacionado à localização das cabines, maior acesso aos botes salva-vidas e condições privilegiadas durante a evacuação.

6.4 Sobrevivência por idade

A distribuição de idade indica que a maioria dos passageiros estava na faixa adulta. Além disso, é possível observar que idades menores tendem a apresentar maior taxa de sobrevivência, sugerindo que crianças foram mais protegidas durante o resgate.

6.5 Crianças e Adultos

Comparando crianças e adultos, observa-se que crianças tiveram maior chance de sobrevivência. Esse resultado reforça o padrão de prioridade no salvamento, onde crianças foram colocadas em situação de vantagem em relação aos adultos.

6.6 Classe social + Crianças

Ao cruzar a classe social com o grupo de crianças, percebe-se que, mesmo havendo uma quantidade significativa de crianças na 3ª classe, a taxa de sobrevivência ainda foi inferior em comparação à 1ª classe. Isso mostra que **a classe social também influenciou diretamente a sobrevivência, mesmo entre crianças**, indicando desigualdade no acesso ao resgate.

7. Conclusão

Com base nas análises realizadas, foi possível concluir que a sobrevivência no Titanic não ocorreu de forma aleatória, mas foi fortemente influenciada por fatores sociais e demográficos.

Os resultados indicaram que:

- A maioria dos passageiros não sobreviveu ao desastre.
- **Mulheres tiveram uma taxa de sobrevivência muito superior** aos homens.
- Passageiros da **1ª classe tiveram mais chances de sobreviver**, enquanto a 3ª classe foi a mais afetada.
- A idade também teve impacto, com **crianças apresentando maior chance de sobrevivência**.
- Mesmo entre crianças, a **classe social ainda foi um fator determinante**, mostrando que passageiros em classes mais altas tiveram maior acesso às chances de resgate.

Assim, o projeto permitiu compreender melhor os padrões presentes no dataset e reforçou como a análise de dados pode revelar comportamentos e desigualdades presentes em eventos históricos.