



UDACITY

**PROGRAMA NANODEGREE
FUNDAMENTOS DE DATA SCIENCE I**

BRUNO EDUARDO FABOCI

1. CÓDIGO UTILIZADO

Como base para o desenvolvimento do projeto, foi utilizado o arquivo **titanic-data-6.csv**.

2. PERGUNTAS

Com base no conjunto de dados supracitado, foram feitas as seguintes perguntas:

- Quais foram os fatores que fizeram com que algumas pessoas fossem mais propensas a sobreviver?
- Mulheres tiveram uma taxa maior de sobrevivência em relação aos homens?
- Existe alguma relação entre as classes sociais e o número de sobreviventes?

3. PROCESSO DE ANÁLISE

A análise dos dados foi feita por meio de um Notebook Jupyter. Para responder cada uma das três perguntas foi criada uma função denominada “análise” para auxiliar a verificar os parâmetros descritos abaixo:

- **Avaliar os dados** – foi checado a quantidade de amostra e colunas do conjunto de dados, bem como os tipos dos dados de cada coluna. Também foi analisado se havia valores faltantes, linhas duplicadas e valores únicos.
- **Tratamento dos dados** – no tratamento dos dados foi feita a substituição dos rótulos das colunas, mantendo os nomes originais, porém convertendo os nomes para caixa baixa e eliminando underlines.

Foi constatado que na coluna “age” havia valores faltantes. Para resolver esse problema foi calculada a média dos valores de idade e esse valor foi utilizado para preenchimento das linhas em branco na coluna “age”.

As colunas que apresentavam valores irrelevantes ou que não ofereciam contribuição relevante para a análise foram descartadas. Essas colunas foram “cabin”, “ticket”, “fare” e “embarked”.

4. RESPONDENDO AS PERGUNTAS

Para obter as respostas pertinentes às perguntas elaboradas, foi feita a análise dos dados por agrupamento.

Foram utilizadas as funções “.groupby” e “.query” para fazer o agrupamento dos dados, e a função “.plot” para facilitar a visualização.

5. RESULTADOS

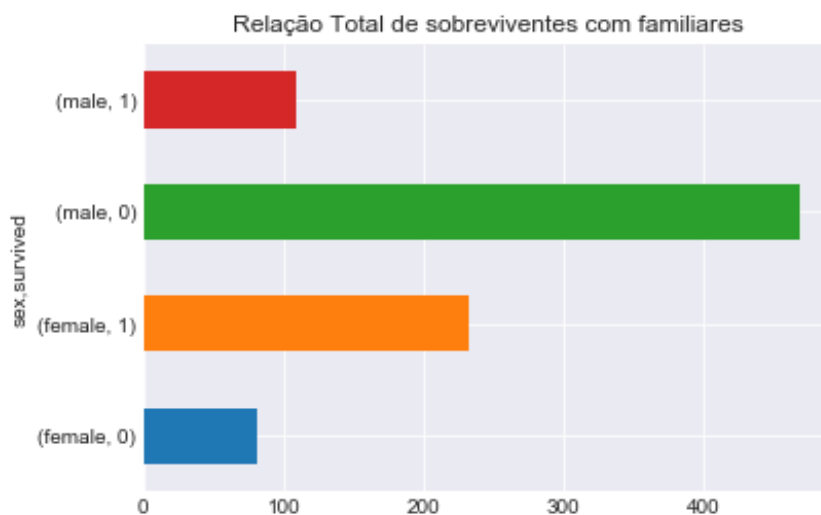
- Quais foram os fatores que fizeram com que algumas pessoas fossem mais propensas a sobreviver?

Um fator constatado foi que houve um número maior de sobreviventes que possuíam parente a bordo (filhos/irmãos).

Pela definição fornecida no site **kaggle** relacionada aos sobreviventes, 1 significa que o passageiro sobreviveu e 0 significa que não.

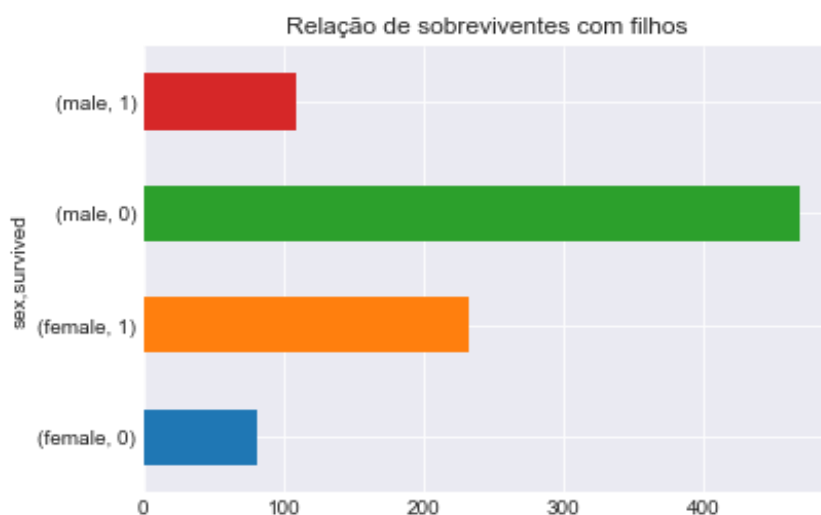
O gráfico abaixo ilustra a relação de homens e mulheres que possuíam familiares a bordo.

Podemos notar que o número de mulheres com familiares que sobreviveram é bem maior em relação ao número de homens.



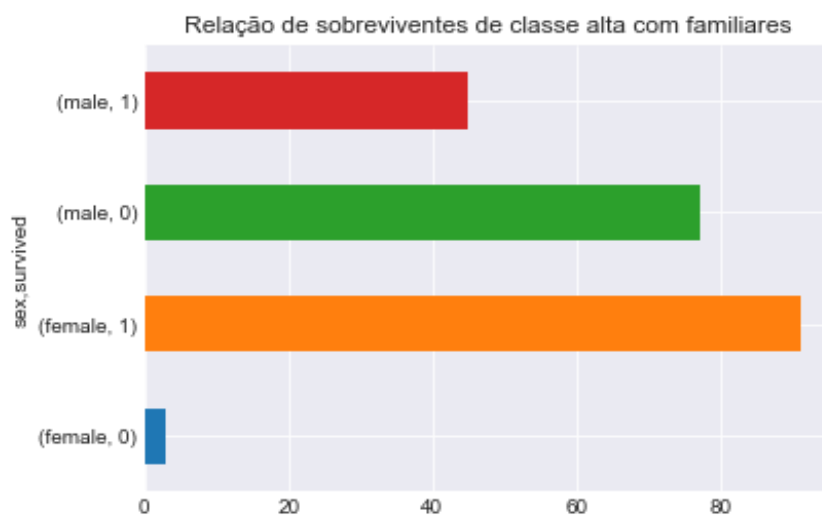
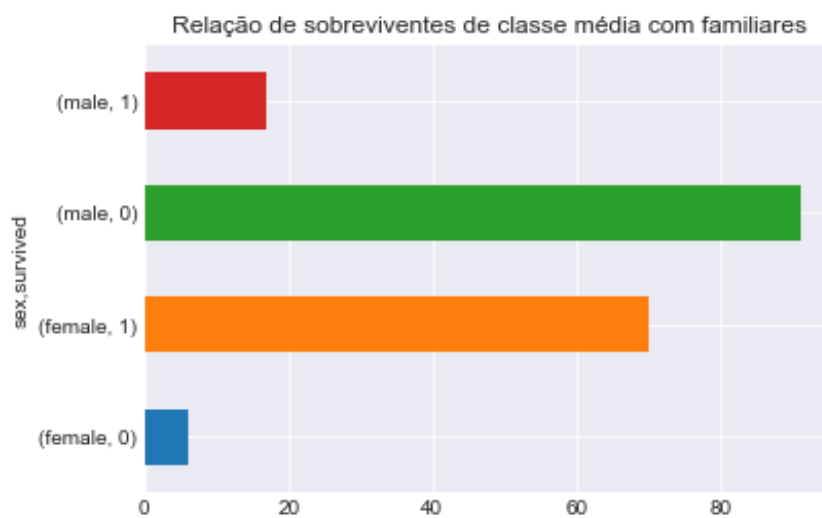
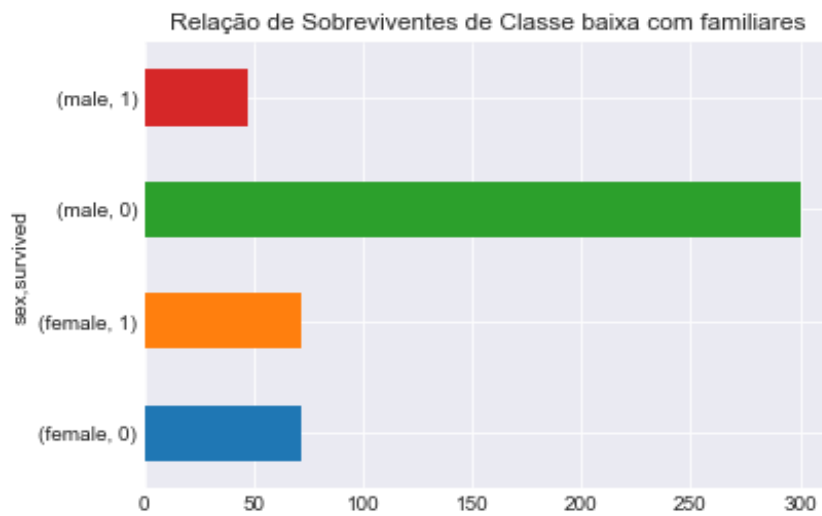
O mesmo efeito pode ser notado no gráfico abaixo, que apresenta a relação de homens e mulheres com filhos a bordo.

Novamente o número de mulheres sobreviventes é bastante superior em relação aos homens.



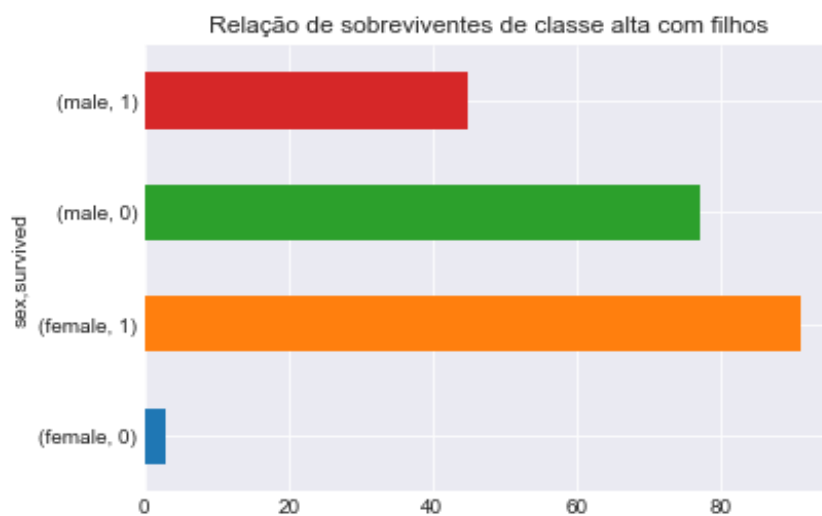
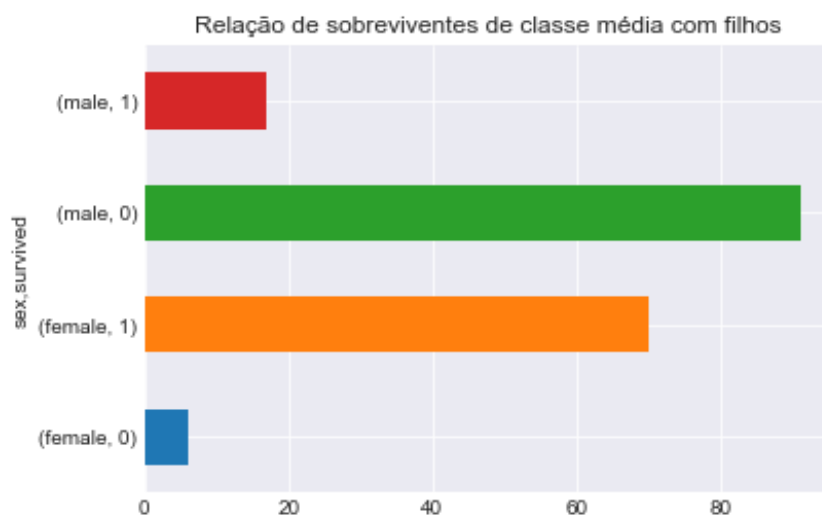
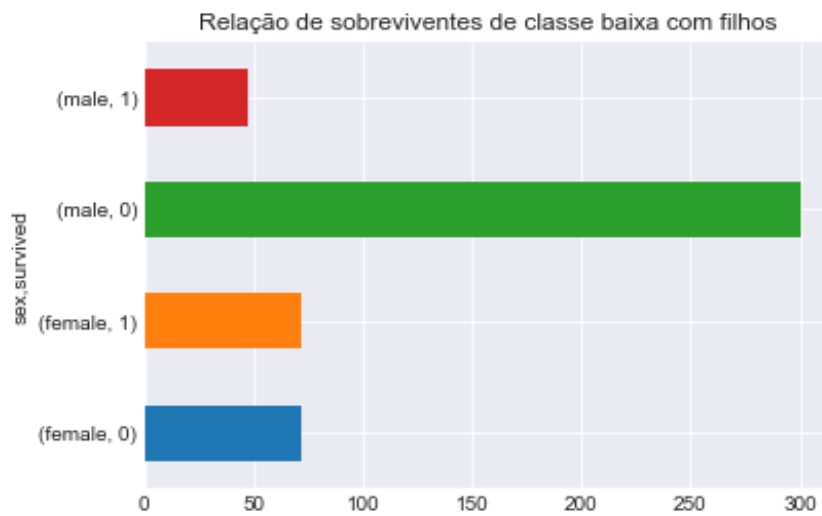
Os gráficos abaixo ilustram a mesma análise feita acima, no entanto, nestes casos as pessoas foram agrupadas em classes sócias.

Aqui temos a relação de indivíduos que possuíam parentes a bordo.



É possível notar que o número de indivíduos de classe baixa que não sobreviveram é muito maior em relação às classes média e alta.

Abaixo temos a relação de indivíduos que possuíam filhos a bordo.

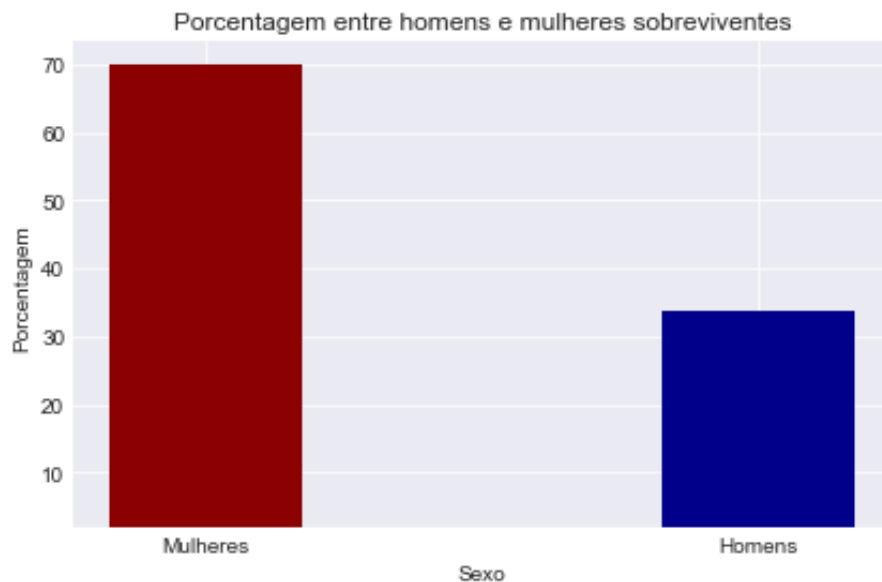


Novamente os indivíduos de classe social baixa apresentam um número muito maior de não-sobreviventes.

Além disso, podemos notar que o número de mulheres que sobreviveram é bem maior que o número de homens.

- **Mulheres tiveram uma taxa maior de sobrevivência em relação aos homens?**

De acordo com a análise realizada e com os dados já apresentados na questão anterior, as mulheres tiveram uma taxa de sobrevivência aproximada de 70%, enquanto os homens ficaram em torno de 30%, como ilustra o gráfico abaixo.

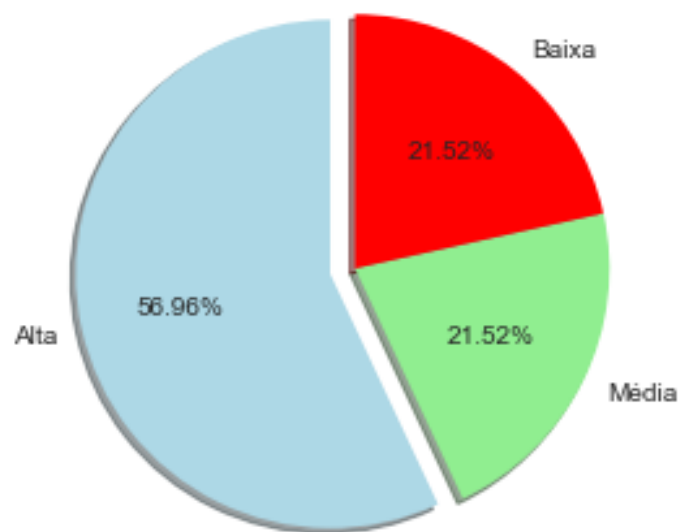


É possível concluir que sim, as mulheres tiveram uma taxa de sobrevivência maior em relação aos homens.

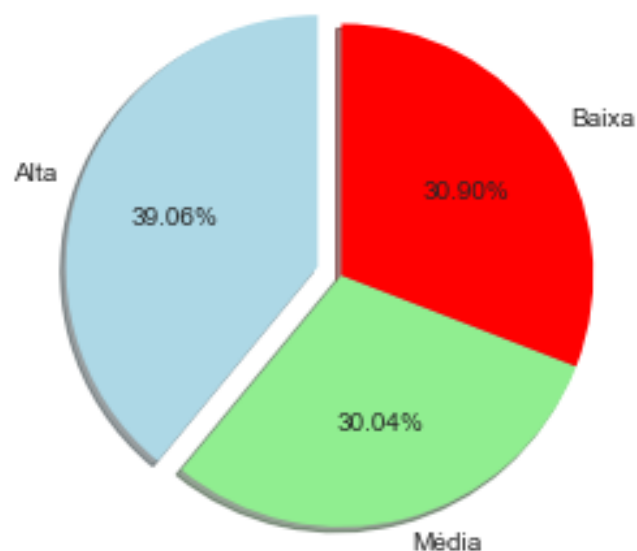
- **Existe alguma relação entre as classes sociais e o número de sobreviventes?**

De acordo com a análise realizada e com os dados apresentados nas questões anteriores, a taxa de sobrevivência foi maior entre homens e mulheres de classe social alta, como ilustram os gráficos abaixo.

Aqui é possível notar que, dentre os sobreviventes do sexo masculino, aproximadamente 57% eram de classe social alta, enquanto os indivíduos das classes média e baixa representam pouco mais de 21,5% cada.

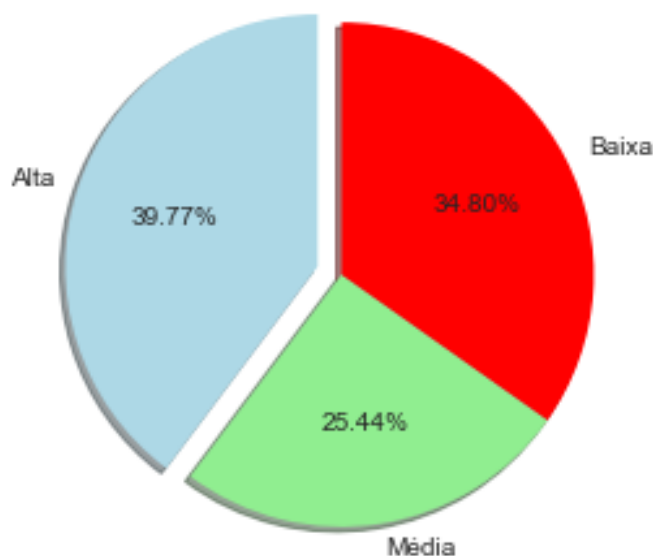


Entre os indivíduos do sexo feminino a diferença é um pouco menor, cerca de 39% dos sobreviventes foram de classe social alta, enquanto as classes média e baixa representam um valor aproximado de 30% cada.



No gráfico abaixo podemos ver a taxa de sobreviventes sem divisão entre homens e mulheres. Indivíduos de classe social alta somam 39% dos

sobreviventes, seguidos pelos indivíduos pertencentes à classe social baixa, com aproximadamente 35% dos sobreviventes. Por fim, os indivíduos de classe social média representam pouco mais de 25% dos sobreviventes.



6. CONCLUSÃO

Diante dos resultados apresentados, podemos concluir que pessoas de classe social alta tiveram prioridade no resgate, dado o elevado número de sobreviventes em relação às classes média e baixa.

Também concluímos que houveram muito mais sobreviventes mulheres que homens, independente da classe social. Isso provavelmente se deve ao fato de que mulheres e crianças tem prioridade em situações de resgate.

- **LIMITAÇÕES**

Uma dificuldade percebida foi a falta de dados em algumas colunas que continham informações relevantes.

Na coluna "age" (idade) dos 891 registros, 177 não continham informações sobre a idade dos passageiros. Com estas informações seria possível determinar se a idade dos indivíduos foi um fator que contribuiu ou não para a sobrevivência.

Da mesma forma, a coluna "cabin", que se refere à identificação da cabine dos passageiros, continha 687 dos 891 registros em branco, o que impossibilita determinar se a posição da cabine no navio foi um fator relevante na sobrevivência dos indivíduos.