

## 03 - Pandas (Pt2)

### Docupedia Export

Author:Lima Queila (CtP/ETS)  
Date:27-Jan-2026 17:04

## Table of Contents

<b>1 TREINAMENTO DE PYTHON (PANDAS) - Parte 2</b>	<b>3</b>
1.1 Vamos analisar as informações sobre os passageiros	4
1.1.1 Vamos focar a nossa análise para detectar padrões que vão nos dizer quem sobreviveu ou não!	4
1.1.2 Será que tem alguma relação com a taxa de sobrevivencia se ele estava com a família?	6
1.1.2.1 Primeiro, qual a porcentagem geral de pessoas com parentes a bordo?	6
1.1.3 Será que tem alguma relação do "sexo" com a taxa de sobrevivência?	8
1.1.3.1 Primeiro qual a porcentagem geral de pessoas por sexo?	8
1.1.4 Começaria acreditando que existe uma forte relação de sobrevivência ou morte!	9
1.2 Vamos verificar se a classe a cabine (Pclass) também influencia na sobrevivência ou morte?	9
1.3 Agora vamos mudar um pouco, será que tem alguma relação entre a taxa de sobrevivência e a idade?	12
1.3.1 Agora vamos reaplicar o nosso filtro	13
1.4 Birkenhead Drill	14
1.4.1 Crie um filtro que aplique essa regra :)	15
1.5 Apagando colunas	16
1.5.1 Alterando o nome da coluna	17

# 1

## TREINAMENTO DE PYTHON (PANDAS) - Parte 2



Vamos continuar usando o **Pandas** para trabalhar nossos dados.

Vamos abrir o dataset que foi salvo no final da aula anterior.

```
import pandas as pd  
titanic = pd.read_csv("data/titanic_1_aula.csv")  
titanic
```

PassengerId	Survived	Pclass	Name			Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked	Relatives
0	1	0	Braund, Mr. Owen Harris			male	22.0	1	0	A/5 21171	7.2500	Nan	S	1
1	2	1	Cumings, Mrs. John Bradley (Florence Briggs Th... Heikkinen, Miss. Laina			female	38.0	1	0	PC 17599 STON/O2. 3101282	71.2833 7.9250	C85	C	1
2	3	1	Heikkinen, Miss. Laina			female	26.0	0	0			Nan	S	0
3	4	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)			female	35.0	1	0	113803	53.1000	C123	S	1
4	5	0	Allen, Mr. William Henry			male	35.0	0	0	373450	8.0500	Nan	S	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
886	887	0	Montvila, Rev. Juozas			male	27.0	0	0	211536	13.0000	Nan	S	0
887	888	1	Graham, Miss. Margaret Edith			female	19.0	0	0	112053	30.0000	B42	S	0
888	889	0	Johnston, Miss. Catherine Helen "Carrie"			female	Nan	1	2	W.C. 6607	23.4500	Nan	S	3
889	890	1	Behr, Mr. Karl Howell			male	26.0	0	0	111369	30.0000	C148	C	0
890	891	0	Dooley, Mr. Patrick			male	32.0	0	0	370376	7.7500	Nan	Q	0

1.1

## Vamos analisar as informações sobre os passageiros

Vamos ver quais passageiros sobreviveram e entender o porque disso.

**1.1.1 Vamos focar a nossa análise para detectar padrões que vão nos dizer quem sobreviveu ou não!**

Pra isso, primeiro vamos os dados brutos sobre esse dado

```
titanic.Survived[titanic["Survived"] == 0].count()
```

549

```
titanic.Survived[titanic["Survived"] == 1].count()
```

342

```
def calculate_percentage(val, total):
    """Calculates the percentage of a value over a total"""

    percent = float(val / total)
    beautiful_percent = ("%.2f" % (percent * 100)) + "%"
    return beautiful_percent
```

```
total = titanic.Survived.count()
nao_sobreviveu = titanic.Survived[titanic["Survived"] == 0].count()
sobreviveu = titanic.Survived[titanic["Survived"] == 1].count()
print(total)
print(nao_sobreviveu)
print(sobreviveu)
```

891

549

342

```
print(calculate_percentage(sobreviveu, total))
```

38.38%

```
print(calculate_percentage(nao_sobreviveu, total))
```

61.62%

Descobrimos que mais gente morreu :(

## 1.1.2 Será que tem alguma relação com a taxa de sobrevivencia se ele estava com a família?

### 1.1.2.1 Primeiro, qual a porcentagem geral de pessoas com parentes a bordo?

```
titanic.columns
```

```
Index(['PassengerId', 'Survived', 'Pclass', 'Name', 'Sex', 'Age', 'SibSp',
       'Parch', 'Ticket', 'Fare', 'Cabin', 'Embarked', 'Relatives'],
      dtype='object')
```

```
sem_familia = titanic.Survived[titanic["Relatives"] == 0].count()
com_familia = titanic.Survived[titanic["Relatives"] > 0].count()

print(calculate_percentage(sem_familia, total))
print(calculate_percentage(com_familia, total))
```

60.27%

39.73%

Vamos começar pelos que morreram!

```
mortos = titanic[titanic["Survived"] == 0]
mortos_sem_familia = mortos.Survived[mortos["Relatives"] == 0].count()
mortos_sem_familia
```

374

```
mortos_com_familia = mortos.Survived[mortos["Relatives"] > 0].count()
```

```
mortos_com_familia
```

175

```
print("Não sobreviveu que NÃO tem família no Barco", calculate_percentage(mortos_sem_familia, nao_sobreviveu))
print("Não sobreviveu que TEM família no Barco", calculate_percentage(mortos_com_familia, nao_sobreviveu))
```

Não sobreviveu que NÃO tem família no Barco 68.12%  
Não sobreviveu que TEM família no Barco 31.88%

**Maioria dos mortos não tinha família no barco!**

**Vamos ver os vivos para ver se essa informação é relevante!**

```
vivos = titanic[titanic["Survived"] == 1]
vivos_sem_familia = vivos.Survived[vivos["Relatives"] == 0].count()
vivos_sem_familia
```

1

```
vivos_com_familia = vivos.Survived[vivos["Relatives"] > 0].count()
vivos_com_familia
```

179

```
print("Sobrevivente que NÃO tem família no Barco", calculate_percentage(vivos_sem_familia, sobreviveu))
print("Sobrevivente que TEM família no Barco", calculate_percentage(vivos_com_familia, sobreviveu))
```

Sobrevivente que NÃO tem família no Barco 47.66%
Sobrevivente que TEM família no Barco 52.34%

Não começaria acreditando que isso contém uma forte relação para definir sobrevivência mas sobre os mortos teve um belo impacto!

## 1.1.3 Será que tem alguma relação do "sexo" com a taxa de sobrevivência?

### 1.1.3.1 Primeiro qual a porcentagem geral de pessoas por sexo?

```
male = titanic.Survived[titanic["Sex"] == "male"].count()
female = titanic.Survived[titanic["Sex"] == "female"].count()

print(calculate_percentage(male, total))
print(calculate_percentage(female, total))
```

64.76%

35.24%

```
mortos_homem = mortos.Survived[mortos["Sex"] == "male"].count()
mortos_homem
```

468

```
mortos_mulher = mortos.Survived[mortos["Sex"] == "female"].count()
mortos_mulher
```

81

```
print("Não sobreviveu do sexo Masculino", calculate_percentage(mortos_homem, nao_sobreviveu))
print("Não sobreviveu do sexo Feminino", calculate_percentage(mortos_mulher, nao_sobreviveu))
```

Não sobreviveu do sexo Masculino 85.25%

Não sobreviveu do sexo Feminino 14.75%

```
vivos_homem = vivos.Survived[vivos["Sex"] == "male"].count()  
vivos_homem
```

109

```
vivos_mulher = vivos.Survived[vivos["Sex"] == "female"].count()  
vivos_mulher
```

233

```
print("Sobreviveu do sexo Masculino", calculate_percentage(vivos_homem, sobreviveu))  
print("Sobreviveu do sexo Feminino", calculate_percentage(vivos_mulher, sobreviveu))
```

Sobreviveu do sexo Masculino 31.87%

Sobreviveu do sexo Feminino 68.13%

#### 1.1.4      Começaria acreditando que existe uma forte relação de sobrevivência ou morte!

## 1.2 Vamos verificar se a classe a cabine (Pclass) também influencia na sobrevivência ou morte?

```
primeira = titanic.Survived[titanic["Pclass"] == 1].count()  
segunda = titanic.Survived[titanic["Pclass"] == 2].count()  
terceira = titanic.Survived[titanic["Pclass"] == 3].count()  
print(calculate_percentage(primeira, total))  
print(calculate_percentage(segunda, total))  
print(calculate_percentage(terceira, total))
```

24.24%  
20.65%  
55.11%

```
mortos_primeira = mortos.Survived[mortos["Pclass"] == 1].count()
mortos_primeira
```

80

```
mortos_segunda = mortos.Survived[mortos["Pclass"] == 2].count()
mortos_segunda
```

97

```
mortos_terceira = mortos.Survived[mortos["Pclass"] == 3].count()
mortos_terceira
```

372

```
print("Não sobreviveu da primeira classe", calculate_percentage(mortos_primeira,nao_sobreviveu))
print("Não sobreviveu do segunda classe", calculate_percentage(mortos_segunda,nao_sobreviveu))
print("Não sobreviveu do terceira classe", calculate_percentage(mortos_terceira,nao_sobreviveu))
```

Não sobreviveu da primeira classe 14.57%  
Não sobreviveu do segunda classe 17.67%  
Não sobreviveu do terceira classe 67.76%

```
vivos_primeira = vivos.Survived[vivos["Pclass"] == 1].count()  
vivos_primeira
```

136

```
vivos_segunda = vivos.Survived[vivos["Pclass"] == 2].count()  
vivos_segunda
```

87

```
vivos_terceira = vivos.Survived[vivos["Pclass"] == 3].count()  
vivos_terceira
```

119

```
print("Sobreviveu da primeira classe", calculate_percentage(vivos_primeira, sobreviveu))  
print("Sobreviveu da segunda classe", calculate_percentage(vivos_segunda, sobreviveu))  
print("Sobreviveu da terceira classe", calculate_percentage(vivos_terceira, sobreviveu))
```

Sobreviveu da primeira classe 39.77%

Sobreviveu da segunda classe 25.44%

Sobreviveu da terceira classe 34.80%

## 1.3 Agora vamos mudar um pouco, será que tem alguma relação entre a taxa de sobrevivência e a idade?

Precisamos de dados de faixa etária e familiares para resolver ele. Vamos pegar a faixa etária primeiro

```
# Transformando dados com apply

# Aplicando uma função aos dados para transformar a idade em faixas etárias.
# O resultado do apply é uma Série.

def faixa_etaria(linhas):
    idade = linhas["Age"]
    if idade < 12:
        return "criança"
    elif idade >=12 and idade < 18:
        return "adolescente"
    elif idade >=18 and idade < 65:
        return "adulto"
    elif idade >= 65:
        return "idoso"
    else:
        return "nada"

# Criamos a coluna "faixa_etaria" onde atribuímos o valor da Série
# resultante do apply
titanic["faixa_etaria"] = titanic.apply(faixa_etaria, axis=1)
```

```
titanic.head(10)
```

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked	Relatives	faixa_etaria
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S	1	adulto
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th... Th...	female	38.0	1	0	PC 17599	71.2833	C85	C	1	adulto
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S	0	adulto
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S	1	adulto
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	S	0	adulto
5	6	0	3	Moran, Mr. James	male	NaN	0	0	330877	8.4583	NaN	Q	0	nada
6	7	0	1	McCarthy, Mr. Timothy J	male	54.0	0	0	17463	51.8625	E46	S	0	adulto
7	8	0	3	Palsson, Master. Gosta Leonard	male	2.0	3	1	349909	21.0750	NaN	S	4	criança
8	9	1	3	Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)	female	27.0	0	2	347742	11.1333	NaN	S	2	adulto
9	10	1	2	Nasser, Mrs. Nicholas (Adele Achem)	female	14.0	1	0	237736	30.0708	NaN	C	1	adolescente

### 1.3.1 Agora vamos reaplicar o nosso filtro

```
mortos = titanic[titanic["Survived"] == 0]
vivos = titanic[titanic["Survived"] == 1]
```

```
mortos.Survived[titanic["faixa_etaria"] == "criança"].count()
```

24

```
vivos.Survived[titanic["faixa_etaria"] == "criança"].count()
```

38

```
mortos.Survived[titanic["faixa_etaria"] == "adolescente"].count()
```

28

```
vivos.Survived[titanic["faixa_etaria"] == "adolescente"].count()
```

23

```
mortos.Survived[titanic["faixa_etaria"] == "adulto"].count()
```

362

```
vivos.Survived[titanic["faixa_etaria"] == "adulto"].count()
```

228

## 1.4

## Birkenhead Drill

Em 1852, existia um código de conduta chamado **Birkenhead Drill**, que protocolava a ordem de resgate de uma situação de vida e morte. Esse código priorizava a vida das mulheres e crianças. Sabemos que o acidente do Titanic ocorreu em 1912, décadas depois da criação deste código.



THE WRECK OF THE "BIRKENHEAD," 1852.

"BIRKENHEAD," Troopship, 1,400 tons. Launched 1845. Heaviest Gun, 10 in. Pivot.

#### 1.4.1

#### Crie um filtro que aplique essa regra :)

```
filtro_mulher = titanic.Sex == 'female'  
filtro_crianca = titanic.faixa_etaria == 'criança'
```

```
mortos.Survived[filtro_mulher | filtro_crianca].count()
```

94

```
vivos.Survived[filtro_mulher | filtro_crianca].count()
```

252

## 1.5

## Apagando colunas

Nem todas as colunas são necessárias para realizar uma análise, a coluna `Name`, por exemplo, pode não ser muito relevante.

```
titanic.drop(["Name"], axis=1)
```

PassengerId	Survived	Pclass	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked	Relatives	faixa_etaria	
0	1	0	3	male	22.0	1	0	A/5 21171	7.2500	NaN	S	1	adulto
1	2	1	1	female	38.0	1	0	PC 17599	71.2833	C85	C	1	adulto
2	3	1	3	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S	0	adulto
3	4	1	1	female	35.0	1	0	113803	53.1000	C123	S	1	adulto
4	5	0	3	male	35.0	0	0	373450	8.0500	NaN	S	0	adulto
5	6	0	3	male	NaN	0	0	330877	8.4583	NaN	Q	0	nada
6	7	0	1	male	54.0	0	0	17463	51.8625	E46	S	0	adulto
7	8	0	3	male	2.0	3	1	349909	21.0750	NaN	S	4	criança
8	9	1	3	female	27.0	0	2	347742	11.1333	NaN	S	2	adulto
9	10	1	2	female	14.0	1	0	237736	30.0708	NaN	C	1	adolescente
10	11	1	3	female	4.0	1	1	PP 9549	16.7000	G6	S	2	criança
11	12	1	1	female	58.0	0	0	113783	26.5500	C103	S	0	adulto
12	13	0	3	male	20.0	0	0	A/5. 2151	8.0500	NaN	S	0	adulto

## 1.5.1

## Alterando o nome da coluna

```
titanic.rename(columns={'faixa_etaria': 'AgeRange'})  
titanic.head()
```

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

```
titanicx=titanic.drop_duplicates(subset="Embarked")  
titanicx
```

PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked	Relatives
0	1	0	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S	1
1	2	1	Cumings, Mrs. John Bradley (Florence Briggs Th... Th...	female	38.0	1	0	PC 17599	71.2833	C85	C	1
5	6	0	Moran, Mr. James	male	NaN	0	0	330877	8.4583	NaN	Q	0
61	62	1	Icard, Miss. Amelie	female	38.0	0	0	113572	80.0000	B28	NaN	0

