

Faça o resumo estatístico por gênero (coluna Sex), contendo a média de idade, o número de sobreviventes e o número de vítimas. (Utilize groupby)

OUTPUT

	media_idade	vivos	mortos
Sex			
female	27.915709	233.0	81.0
male	30.726645	109.0	468.0

A partir do dataset 'train.csv' mostre na tela qual era a proporção de passageiros homens e mulheres no Titanic, apresente os dados em porcentagem.

OUTPUT

```
Homens no titanic:  64.75869809203144 %  
Mulheres no titanic:  35.24130190796858 %
```

O arquivo “cat\_breeds\_clean.csv” é composto pelos seguintes features:

- “raca” – A raça do gato, pode ser “Angora”, “Maine coon” ou “Ragdoll”;
- “meses\_idade” – Quantos meses o gato tem de idade;
- “sexo” – “macho” ou “femea”;
- “peso” – O peso do gato em Kg;
- “comprimento” – O comprimento do gato em cm;
- “cor” – A cor do gato, pode ser “branco”, “preto”, “cinza”, “marrom” ou “laranja”;
- “horas\_dormindo” – Quantas horas por dia o gato dorme;
- “pais” – O país onde o gato mora, pode ser “EUA”, “Inglaterra”, “França”, “Canadá” ou “Alemanha”.

Desses dados extraia:

- A lista de todos os gatos da raça Angora.
- A lista de todos os gatos que moram em países da Europa.
- Os países dos gatos laranjas.
- A razão entre o peso (Kg) e comprimento (m) dos gatos com mais de 6 meses de idade.
- A raça, idade e peso de todos os gatos machos.
- A porcentagem do dia que cada gato Ragdoll passa dormindo (como string).
- A soma do comprimento de todos os gatos.
- A média de horas dormidas.
- O peso total de todos os gatos entre 3 e 12 meses de idade.
- A média da razão entre comprimento (cm) e idade dos gatos brancos e pretos da América do Norte

Um cliente deseja comprar uma casa no Rio de Janeiro. Suas exigências são: 3 quartos, no bairro Tijuca, área maior que 130m<sup>2</sup>. Mostre a opção mais barata presente na tabela “dados.csv”.

#### OUTPUT

	condominio	quartos	suites	vagas	area	bairro	preco	pm2
1971	480	3	1.0	3.0	142	Tijuca	600000	4225.35

Faça um tratamento no arquivo “master.csv”:

Separe a coluna “country-year” em duas colunas chamadas “country” e outra chamada year usando duas funções para criar cada coluna.

```
      sex      age suicides_no ... generation year country
0     male 15-24 years      21 ... Generation X 1987  Albania
1     male 35-54 years      16 ...      Silent 1987  Albania
2    female 15-24 years      14 ... Generation X 1987  Albania
3     male  75+ years       1 ... G.I. Generation 1987  Albania
4     male 25-34 years       9 ...      Boomers 1987  Albania
...    ...    ...      ... ...    ...    ...    ...
27815 female 35-54 years     107 ... Generation X 2014  Uzbekistan
27816 female  75+ years       9 ...      Silent 2014  Uzbekistan
27817 male   5-14 years      60 ... Generation Z 2014  Uzbekistan
27818 female  5-14 years      44 ... Generation Z 2014  Uzbekistan
27819 female 55-74 years      21 ...      Boomers 2014  Uzbekistan
```

Excluir a coluna “HDI for year”.

E fazer um filtro que apresente somente as informações da linha com o país = Brazil, ano = 1987, sex=female, e a geração = Boomers.

```
      sex      age suicides_no ... generation year country
4204 female 25-34 years      266 ...      Boomers 1987  Brazil
```

Com o arquivo “tips.csv”:

- Crie uma coluna tip\_pct com a porcentagem de gorjeta em relação ao total da conta.
- Encontre a mesa com maior gorjeta percentual.
- Calcule a média de gorjeta por gênero
- Calcule a quantidade de atendimentos
- Crie uma nova coluna e defina como “pagador ruim” gorjetas abaixo da média
- Em qual dia há a maior proporção de maus pagadores?
- Fumantes ou não fumantes pagam pior?
- Qual é a média de gorjeta apenas para mesas com mais de 3 pessoas, jantar, não fumantes, no sábado?