

07

## Faça o que eu fiz na aula

Vamos criar uma variável chamada “quant” com as colunas quantitativas que iremos utilizar.

```
quant = ['bounces', 'hits', 'newVisits',
         'pageviews', 'transactionRevenue']
```

Agora vamos garantir que todas essas colunas estejam identificadas como tipos numéricos.

```
for coluna in quant:
    df[coluna] = pd.to_numeric(df[coluna])
```

Ao verificar os tipos das colunas, a saída deve ser igual a:

```
df.dtypes
```

channelGrouping	object
date	object
fullvisitorId	object
sessionId	object
visitId	int64
visitNumber	int64
visitStartTime	int64
browser	object
deviceCategory	object
isMobile	bool
operatingSystem	object
city	object
continent	object
country	object
metro	object
networkDomain	object
region	object
subContinent	object
adContent	object
campaign	object
campaignCode	object
isTrueDirect	object
keyword	object
medium	object
referralPath	object
source	object
bounces	float64
hits	int64
newVisits	float64
pageviews	int64
transactionRevenue	float64

Para o caso dos dados que tiveram um Revenue vazio, vamos preencher com 0.

```
df.transactionRevenue.fillna(0, inplace=True)
```

Segundo a documentação do Google Analytics, o valor do Revenue é multiplicado por 1.000.000. Para trabalhar com o número real, vamos dividir essa coluna por 1.000.000.

```
df.transactionRevenue.fillna(0, inplace=True)
```

Aqui chegamos na parte principal da criação de modelos de machine learning com esse tipo de dados. O Dataset possui mais de uma linha por usuário, e o nosso desafio é prever o valor total gasto por usuário.

Para conseguir treinar um algoritmo de machine learning nesse caso, primeiro precisamos agregar as informações por usuário e criar as variáveis preditoras baseando no conhecimento que temos sobre a base de dados.

A nossa primeira estratégia será treinar um algoritmo baseado nas informações quantitativas dos usuários.

```
df_quant = df.groupby('fullVisitorId', as_index=False)[quant].sum()
```

A final da criação do df\_quant, o head deverá ser igual a:

```
df_quant.head()
```

	fullVisitorId	bounces	hits	pageviews	visits	transactionRevenue
0	0002365800130207040	1.0	1	1	1	0.0
1	0010286039787739137	0.0	4	3	1	0.0
2	0011056874471185769	0.0	2	2	1	0.0
3	0014443856125569702	1.0	1	1	1	0.0
4	0017260116665815114	0.0	2	2	1	0.0