

# Análise dos dados da Copa do Mundo

Com base na página sobre a copa do mundo na Wikipedia

Tabela: Público pagante na final de cada edição do Mundial

link => [https://pt.wikipedia.org/wiki/Copa\\_do\\_Mundo\\_FIFA](https://pt.wikipedia.org/wiki/Copa_do_Mundo_FIFA)  
([https://pt.wikipedia.org/wiki/Copa\\_do\\_Mundo\\_FIFA](https://pt.wikipedia.org/wiki/Copa_do_Mundo_FIFA))

Willer Gomes Junior 09/08/2021

```
In [1]: # Importar bibliotecas
# por convencao vamos chamar o pandas de pd
import pandas as pd
import requests

# Configurando o matplotlib
%matplotlib inline

# Endereco no Wikipedia
url='https://pt.wikipedia.org/wiki/Copa_do_Mundo_FIFA'

# Obterndo os dados
resposta = requests.get(url)

# Obtendo todo o conteúdo do endereço
#resposta.text

# Obtendo todo o conteúdo do endereço
table = pd.read_html(resposta.text)
#table

# Queremos apenas a tabela com pagantes por final
# Essa é a 12o tabela com dados retornados
df = table[11]
df.head()
```

Out[1]:

	Edição	Ano	Estádio	Local	Público pagante
0	I	1930	Estádio Centenário	Montevideu	68.346
1	II	1934	Estádio do Partido Nacional Fascista	Roma	55.000
2	III	1938	Estádio Olímpico Yves-du-Manoir	Colombes	45.000
3	IV	1950	Estádio do Maracanã	Rio de Janeiro	173.850
4	V	1954	Estádio Wankdorf	Berna	62.500

```
In [2]: # Muda o nome das colunas usando a primeira linha / Quando a linha de titulos
        #df.columns = df.iloc[0]
        #df.head
```

```
In [3]: # Apagando a primeira linha
df.drop(df.index[0], inplace=True)
df.head
```

```
Out[3]: <bound method NDFrame.head of      Edição    Ano      E
stádio      Local \
1      II  1934  Estádio do Partido Nacional Fascista      Roma
2      III 1938      Estádio Olímpico Yves-du-Manoir      Colombes
3      IV 1950      Estádio do Maracanã      Rio de Janeiro
4      V  1954      Estádio Wankdorf      Berna
5      VI 1958      Estádio Råsunda      Solna
6      VII 1962      Estádio Nacional do Chile      Santiago
7      VIII 1966      Estádio de Wembley      Londres
8      IX  1970      Estádio Azteca      Cidade do México
9      X  1974      Estádio Olímpico de Munique      Munique
10     XI  1978      Estádio Monumental de Núñez      Buenos Aires
11     XII 1982      Estádio Santiago Bernabéu      Madrid
12     XIII 1986      Estádio Azteca      Cidade do México
13     XIV 1990      Estádio Olímpico de Roma      Roma
14     XV  1994      Rose Bowl      Pasadena
15     XVI 1998      Stade de France      Saint-Denis
16     XVII 2002      Estádio Internacional de Yokohama      Yokohama
17     XVIII 2006      Estádio Olímpico de Berlim      Berlim
18     XIX 2010      Soccer City      Joanesburgo
19     XX  2014      Estádio do Maracanã      Rio de Janeiro
20     XXI 2018      Estádio Lujniki      Moscou

      Público pagante
1      55.000
2      45.000
3     173.850
4      62.500
5      49.737
6      68.679
7      96.924
8     107.412
9      78.200
10     71.483
11     90.000
12    114.600
13     73.603
14     94.194
15     80.000
16     69.029
17     69.000
18     84.490
19     74.738
20     78.011 >
```

```
In [4]: # Obtendo todo o conteúdo do endereço
table1 = pd.read_html(resposta.text)
#table

# Queremos apenas a tabela com pagantes por final
# Essa é a 12 tabela dos dados retornados
df1 = table1[11]
df1.head()
```

Out[4]:

	Edição	Ano	Estádio	Local	Público pagante
0	I	1930	Estádio Centenário	Montevideu	68.346
1	II	1934	Estádio do Partido Nacional Fascista	Roma	55.000
2	III	1938	Estádio Olímpico Yves-du-Manoir	Colombes	45.000
3	IV	1950	Estádio do Maracanã	Rio de Janeiro	173.850
4	V	1954	Estádio Wankdorf	Berna	62.500

```
In [5]: df1.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 21 entries, 0 to 20
Data columns (total 5 columns):
Edição          21 non-null object
Ano             21 non-null int64
Estádio         21 non-null object
Local           21 non-null object
Público pagante 21 non-null float64
dtypes: float64(1), int64(1), object(3)
memory usage: 652.0+ bytes
```

```
In [6]: # Quando o Público Pagante aparece como object pois existe um ponto no milhar
# Retirando o ponto do milhar e convertendo para int
#df1['Público pagante'] = df1['Público pagante'].str.replace('.', '').astype
(int)

# Público Pagante aparece como float pois existe um ponto no milhar
df1['Público pagante'] = df1['Público pagante'] * 1000
```

```
In [7]: df1['Público pagante'].astype(int)
df1.head()
```

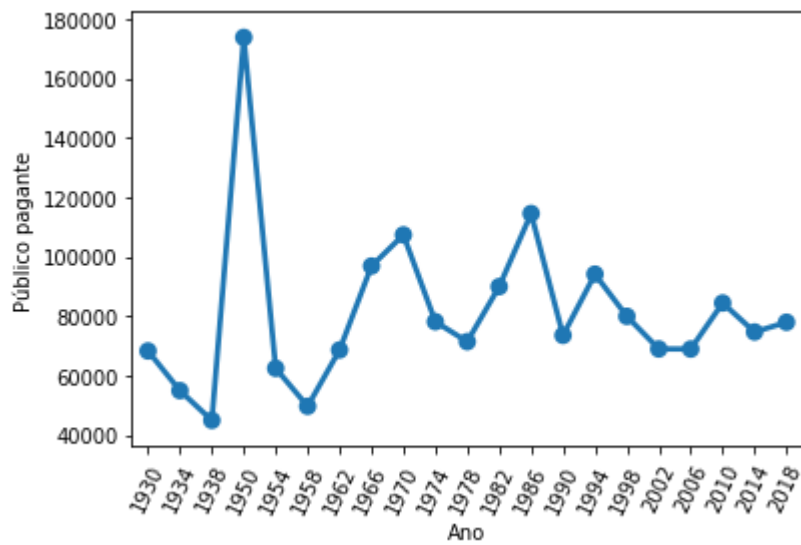
Out[7]:

	Edição	Ano	Estádio	Local	Público pagante
0	I	1930	Estádio Centenário	Montevideu	68346.0
1	II	1934	Estádio do Partido Nacional Fascista	Roma	55000.0
2	III	1938	Estádio Olímpico Yves-du-Manoir	Colombes	45000.0
3	IV	1950	Estádio do Maracanã	Rio de Janeiro	173850.0
4	V	1954	Estádio Wankdorf	Berna	62500.0

```
In [8]: # Construindo um gráfico com a evolução do publico pagante
# importando uma biblioteca para gráficos
import seaborn as sns
import matplotlib.pyplot as plt
```

```
In [9]: # Exibindo o gráfico de linhas com marcação de cada ponto
sns.pointplot(x='Ano', y='Público pagante', data=df1)

# ajustar as legendas
plt.xticks(rotation=65)
plt.show()
```



In [ ]: