Scrape data from following webpage, create a data frame and find:

- the top 5 countries with most internet users for 2009 and 2010.
- the top 5 countries with least internet users for 2009 and 2010

https://www.nationmaster.com/country-info/stats/Media/Internet-users

Note: For the data frame:

- Please just keep COUNTRY, AMOUNT and DATE columns in your dataframe.
- Please change the name of DATE column to "YEAR"

```
import pandas as pd
from bs4 import BeautifulSoup
import requests
```

1. Request the content from the website

```
# Send a GET request to the webpage
res = requests.get('https://www.nationmaster.com/country-info/stats/Media/Internet-users
```

```
/usr/local/lib/python3.10/dist-packages/urllib3/connectionpool.py:1045: InsecureRe
  warnings.warn(
```

```
res
```

```
<Response [200]>
```

2. Parse the html document and identify the elements we need (Use the BeautifulSoup library)

```
# Create a BeautifulSoup object to parse the HTML content
soup = BeautifulSoup(res.content, 'lxml')
```

```
# Find the table containing the data
table = soup.find_all('table')[0]
```

3. Create a data frame using the Pandas library

```
df = pd.read_html(str(table))[0]
```

```
df.head()
```

✓ 0s completed at 12:34 PM

0	1	China	389 million	2009	NaN	NaN
1	2	United States	245 million	2009	NaN	NaN
2	3	Japan	99.18 million	2009	NaN	NaN
3	NaN	Group of 7 countries (G7) average (profile)	80.32 million	2009	NaN	NaN
4	4	Brazil	75.98 million	2009	NaN	NaN

4. Data Wrangling

```
df1 = df[['COUNTRY', 'AMOUNT', 'DATE']]
```

```
df1.head()
```

	COUNTRY	AMOUNT	DATE
0	China	389 million	2009
1	United States	245 million	2009
2	Japan	99.18 million	2009
3	Group of 7 countries (G7) average (profile)	80.32 million	2009
4	Brazil	75.98 million	2009

Rename the columns

```
df1.rename(columns = {'COUNTRY': 'country', 'AMOUNT': 'amount', 'DATE': 'year'}, inplace
```

```
<ipython-input-12-807a77c7c381>:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable
  df1.rename(columns = {'COUNTRY': 'country', 'AMOUNT': 'amount', 'DATE': 'year'},
```

```
df1.head()
```

	country	amount	year
0	China	389 million	2009
1	United States	245 million	2009
2	Japan	99.18 million	2009
3	Group of 7 countries (G7) average (profile)	80.32 million	2009
4	Brazil	75.98 million	2009

Data types

```
df1.dtypes
```

```
country    object
amount     object
year        int64
dtype: object
```

Convert amount column to numerical

```
df1['amount'] = df1['amount'].replace({"million": "*1e6"}, regex=True).map(pd.eval).asty
df1.head()
```

```
<ipython-input-15-28215a6671fe>:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable
  df1['amount'] = df1['amount'].replace({"million": "*1e6"}, regex=True).map(pd.ev
```

	country	amount	year
0	China	389000000.0	2009
1	United States	245000000.0	2009
2	Japan	99180000.0	2009
3	Group of 7 countries (G7) average (profile)	80320000.0	2009
4	Brazil	75980000.0	2009

Filter the data for 2009 and 2010

```
filtered_year = [2009, 2010]
filtered_df1 = df1[df1['year'].isin(filtered_year)]
filtered_df1
```

	country	amount	year
0	China	389000000.0	2009
1	United States	245000000.0	2009
2	Japan	99180000.0	2009
3	Group of 7 countries (G7) average (profile)	80320000.0	2009
4	Brazil	75980000.0	2009
...	...	...	...
242	Wallis and Futuna	1300.0	2009

242	Wallis and Futuna	1300.0	2009
243	Montserrat	1200.0	2009
244	Niue	1100.0	2009
245	Saint Helena, Ascension, and Tristan da Cunha	900.0	2009
246	Saint Helena	900.0	2009

237 rows × 3 columns

Sort the data by the "amount" column in descending order

```
sorted_df1 = filtered_df1.sort_values('amount', ascending = False)
sorted_df1
```

	country	amount	year
0	China	389000000.0	2009
1	United States	245000000.0	2009
2	Japan	99180000.0	2009
3	Group of 7 countries (G7) average (profile)	80320000.0	2009
4	Brazil	75980000.0	2009
...	...	...	...
242	Wallis and Futuna	1300.0	2009
243	Montserrat	1200.0	2009
244	Niue	1100.0	2009
245	Saint Helena, Ascension, and Tristan da Cunha	900.0	2009
246	Saint Helena	900.0	2009

237 rows × 3 columns

Top 5 countries with the most internet users for 2009 and 2010

```
top_five_countries_most_users = sorted_df1.head()
print("Top 5 countries with most internet users for 2009 and 2010:")
print(top_five_countries_most_users)
```

```
Top 5 countries with most internet users for 2009 and 2010:
                                       country       amount  year
0                                        China  389000000.0  2009
1                                United States  245000000.0  2009
2                                        Japan   99180000.0  2009
3  Group of 7 countries (G7) average (profile)   80320000.0  2009
4                                       Brazil   75980000.0  2009
```

Top 5 countries with the least internet users for 2009 and 2010

```
bottom_five_countries_least_users = sorted_df1.tail()
print("Top 5 countries with least internet users for 2009 and 2010:")
print(bottom_five_countries_least_users)
```

```
Top 5 countries with least internet users for 2009 and 2010:
                                          country  amount  year
242                            Wallis and Futuna  1300.0  2009
243                                   Montserrat  1200.0  2009
244                                         Niue  1100.0  2009
245  Saint Helena, Ascension, and Tristan da Cunha   900.0  2009
246                                  Saint Helena   900.0  2009
```