```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline
# beauty_soup.py
import re
from bs4 import BeautifulSoup
from urllib.request import urlopen

url = "http://amazon.in"
page = urlopen(url)
html = page.read().decode("utf-8")
soup = BeautifulSoup(html, "html.parser")
print(soup.get_text())
soup.find_all("img")
image1= soup.find_all("img")
#image1.name
#image1['src']
#image2['src']
soup.title
soup.title.string
soup.find_all("img", src="/static/
dionysus.jpg")
soup.find_all('a')
rows = soup.find_all('tr')
print(rows[:10])
for row in rows:
    row_td = row.find_all('td')
print(row_td)
type(row_td)
str_cells = str(row_td)
cleantext = BeautifulSoup(str_cells,
"lxml").get_text()
print(cleantext)
list_rows = []
for row in rows:
    cells = row.find_all('td')
    str_cells = str(cells)
    clean = re.compile('<.*?>')
    clean2 = (re.sub(clean, '',str_cells))
    list_rows.append(clean2)
print(clean2)
type(clean2)
df = pd.DataFrame(list_rows)
df.head(10)
df1 = df[0].str.split(',', expand=True)
df1.head(5)
col_labels = soup.find_all('th')
all_header = []
col_str = str(col_labels)
#cleantext2 = BeautifulSoup(col_str,
"lxml").get_text()
all_header.append(cleantext2)
print(all_header)
df2 = pd.DataFrame(all_header)
df2.head()
df3 = df2[0].str.split(',', expand=True)
df3.head()
frames = [df3, df1]
df4 = pd.concat(frames)
df4.head(10)
df5 = df4.rename(columns=df4.iloc[0])
df5.head()
df5.info()
df5.shape
df6 = df5.dropna(axis=0, how='any')
df7 = df6.drop(df6.index[0])
df7.head()
```