

## ✓ **Radio Data Extraction from Amazon**

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

def get_title(soup):

    try:

        title = soup.find('span', attrs = {"id" : 'productTitle'})
        title_value = title.text
        title_string = title_value.strip()

    except AttributeError:
        title_string = ""

    return title_string

def get_price(soup):

    try:
        price = soup.find('span', attrs = {"class" : 'a-price-whole'}).text.strip()
    except AttributeError:
        try:
            price = soup.find('span', attrs = {"class" : 'a-price-whole'}).text.strip()
        except:
            price = ""

    return price

def get_brand(soup):

    try:
        brand = soup.find('span', attrs = {"class" : "a-size-base po-break-word"})
        brand_value = brand.text
        brand_string = brand_value.strip()

    except AttributeError:
        brand_string = ""

    return brand_string

def get_ratings(soup):

    try:
        rating = soup.find('span', attrs = {"id" : "acrCustomerReviewText"})
        rating_value = rating.text
        rating_string = rating_value.strip()
```

```
except AttributeError:
    rating_string = ""

return rating_string


def get_colour(soup):

    try:
        colour = soup.find('tr', attrs = {"class" : 'a-spacing-small po-color'})
        colour_value = colour.text
        colour_string = colour_value.strip()

    except AttributeError:
        colour_string = ""

    return colour_string


def get_feature(soup):

    try:
        feature = soup.find('tr', attrs = {"class" : 'a-spacing-small po-special_feature'})
        feature_value = feature.text
        feature_string = feature_value.strip()

    except AttributeError:
        feature_string = ""

    return feature_string


def get_Dimensions(soup):

    try:
        Dimensions = soup.find('tr', attrs = {"class" : 'a-spacing-small po-item_length_width_height'})
        Dimensions_value = Dimensions.text
        Dimensions_string = Dimensions_value.strip()

    except AttributeError:
        Dimensions_string = ""

    return Dimensions_string


def get_tuner(soup):

    try:
        tuner = soup.find('tr', attrs = {"class" : 'a-spacing-small po-tuner_technology'}).text.strip()
        tuner_value = tuner.text
        tuner_string = tuner_value.strip()

    except AttributeError:
        tuner_string = ""
```

```

    return tuner_string

import requests
from bs4 import BeautifulSoup
import pandas as pd

def scrape_page(url):
    HEADERS = {
        'User-Agent': 'Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/120.0.0.0 Safari/537.36'
    }
    webpage = requests.get(url, headers=HEADERS)
    soup = BeautifulSoup(webpage.content, "html.parser")
    links = soup.find_all("a",
                           attrs={'class': "a-link-normal s-underline-text s-underline-link-text s-link-style a-text-normal"})

    data = {'title': [], 'price': [], 'brand': [], 'rating': [], 'feature': [], 'Dimensions': [], 'tuner': [], 'colour': []}

    for link in links:
        if 'https://www.amazon.in' not in link['href']:
            full_url = f"https://www.amazon.in{link['href']}"
            new_webpage = requests.get(full_url, headers=HEADERS)
            new_soup = BeautifulSoup(new_webpage.content, "html.parser")
            data['title'].append(get_title(new_soup))
            data['price'].append(get_price(new_soup))
            data['brand'].append(get_brand(new_soup))
            data['rating'].append(get_ratings(new_soup)) # KeyError here
            data['feature'].append(get_feature(new_soup))
            data['Dimensions'].append(get_Dimensions(new_soup))
            data['tuner'].append(get_tuner(new_soup))
            data['colour'].append(get_colour(new_soup))

    return data

if __name__ == '__main__':
    base_url = "https://www.amazon.in/s?k=radio&crid=X02ZA33B7ZKS&prefix=radio%2Caps%2C323&ref=nb_sb_noss_1&page="

    all_data = {'title': [], 'price': [], 'brand': [], 'rating': [], 'feature': [], 'Dimensions': [], 'tuner': [], 'colour': []}

    for page_number in range(1, 6):
        url = f"{base_url}{page_number}"
        page_data = scrape_page(url)

        all_data['title'].extend(page_data['title'])
        all_data['price'].extend(page_data['price'])
        all_data['brand'].extend(page_data['brand'])
        all_data['rating'].extend(page_data['rating'])
        all_data['feature'].extend(page_data['feature'])
        all_data['Dimensions'].extend(page_data['Dimensions'])
        all_data['tuner'].extend(page_data['tuner'])
        all_data['colour'].extend(page_data['colour'])

    df_all = pd.DataFrame(all_data)

```

```
print("Combined Data from All Pages:")
print(df_all)
```

Combined Data from All Pages:

	title	price	brand \
0	Philips Audio RL191/94 with MW/FM Bands 180mW ...	999.	Philips Audio
1	Philips Radio RL118/94 with MW/SW/FM Bands, 20...	1,529.	Philips Audio
2	Philips Radio DL167/94 with MW/SW/FM Bands, 2x...	1,499.	Philips Audio
3	iBELL FM700BT Portable FM Radio with Bluetooth...	1,590.	IBELL
4	Zebronics Zeb-County Bluetooth Speaker with Bu...	549.	ZEBRONICS
..	...	...	...
80	Nokia 130 Music   Built-in Powerful Loud Speak...	1,949.	Nokia
81	Saregama Carvaan Hindi - Portable Music Player...	6,790.	Saregama
82	Xech Bluetooth Speaker with Phone Holder 5W Mi...	899.	Xech
83	Tech-lobby 5 Band Rechargeable with Remote Con...	1,349.	Tech-Lobby
84	Portronics SoundDrum 1 10W TWS Portable Blueto...	999.	Portronics

	rating	feature \
0	16,795 ratings	Special Feature Built-In AM Antenna
1	16,795 ratings	Special Feature Rechargeable
2	16,795 ratings	Special Feature Portable
3	1,367 ratings	Special Feature Bluetooth, portable
4	60,793 ratings	Special Feature SD Card, Radio
..	...	...
80	4,602 ratings	
81	9,719 ratings	Special Feature Portable
82	3,040 ratings	Special Feature Retro Radio Design, USB Pe...
83	182 ratings	Special Feature MANUAL CONTROL
84	7,083 ratings	Special Feature Handsfree Calling, Portable,...

	Dimensions tuner \
0	Product Dimensions 12.2L x 6.6W x 18.7H Cent...
1	Product Dimensions 40.2L x 21.7W x 17.9H Cen...
2	Product Dimensions 21.4L x 16.6W x 10H Centi...
3	Product Dimensions 18.5L x 6.5W x 12H Centim...
4	
..	...
80	
81	
82	
83	
84	

	colour
0	Colour Black
1	Colour White
2	Colour Silver
3	Colour Brown 700BT
4	
..	...
80	
81	
82	
83	Colour Black
84	

[85 rows x 8 columns]

```
df = df_all
```



```
# Replace empty strings with NaN
#df.replace('', np.nan, inplace=True)

# Extract only the feature information
#df['feature'] = df['feature'].str.replace('Special Feature ', '')

# Extract only the numerical dimensions
#df['Dimensions'] = df['Dimensions'].str.extract(r'(\d+\.\d+L x \d+\.\d+W x \d+\.\d+H)')

# Convert 'price' and 'rating' columns to numeric types
#df['price'] = pd.to_numeric(df['price'], errors='coerce')
#df['rating'] = df['rating'].str.replace(' ratings', '').str.replace(', ', '')
#df['rating'] = pd.to_numeric(df['rating'], errors='coerce')

df.head(25)
```

	title	price	brand	rating	feature	Dimensions	tuner	colour	
0	Philips Audio RL191/94 with MW/FM Bands 180mW ...	999.0	Philips Audio	16795.0	Built-In AM Antenna	12.2L x 6.6W x 18.7H	NaN	Colour Black	
1	Philips Radio RL118/94 with MW/SW/FM Bands, 20... Philips Radio RL 167/04 with	NaN	Philips Audio	16795.0	Rechargeable	40.2L x 21.7W x 17.9H	NaN	Colour White	
3	iBELL FM700BT Portable FM Radio with Bluetooth...	NaN	IBELL	1367.0	Bluetooth, portable	NaN	NaN	Colour Brown 700BT	
4	Zebronics Zeb-County Bluetooth Speaker with Bu...	549.0	ZEBRONICS	60793.0	SD Card, Radio	NaN	NaN	NaN	
5	Saregama Carvaan Mini Hindi 2.0- Music Player ...	NaN	Carvaan	21274.0	Auto Speaker Setup	NaN	NaN	Colour Moonlight Black	
6	Saregama Carvaan Premium Hindi - Portable Musi...	NaN	Carvaan	9719.0	Portable	NaN	NaN	Colour Emerald Green	
7	Amkette Pocket Mate FM Radio with Bluetooth Sp...	NaN	AMKETTE	709.0	High Quality FM Speaker with Antenna, Up t...	11.8L x 6.4W x 2.8H	NaN	Colour Grey	
8	Amkette Pocket Blast FM Radio with Bluetooth S...	NaN	AMKETTE	3759.0	Inbuilt Mic for Voice & FM Recording & Cal...	NaN	NaN	Colour Brown	
9	Amkette Pocket Mate FM Radio with Bluetooth Sp...	NaN	AMKETTE	709.0	High Quality FM Speaker with Antenna, Up t...	11.8L x 6.4W x 2.8H	NaN	Colour Blue	
10	Vemax Base 3-Band (FM/AM/MW) Portable Radio (B...	799.0	VEmax	704.0	Portable, Lightweight	NaN	NaN	Colour Black	