Amazon

October 18, 2024

```python
[2]: # Import necessary libraries
     from bs4 import BeautifulSoup
     import pandas as pd

     # Step 1: Read the Amazon.html file
     with open('Amazon.html', 'r', encoding='utf-8') as file:
         html_content = file.read()

     # Step 2: Parse the HTML content with BeautifulSoup
     soup = BeautifulSoup(html_content, 'html.parser')

     # Step 3: Find all <div> with the specified class
     divs = soup.find_all('div', class_="puis-card-container s-card-container
      ↪s-overflow-hidden aok-relative puis-include-content-margin puis
      ↪puis-v3az2sezl0cvhm234391inayc4r s-latency-cf-section puis-card-border")

     # Initialize lists to store the extracted data
     product_names = []
     product_prices = []

     # Step 4: Loop through each <div> and find the <span> for product name and price
     for div in divs:
         # Find product name
         product_name_span = div.find('span', class_='a-size-medium a-color-base
      ↪a-text-normal')
         product_name = product_name_span.get_text(strip=True) if product_name_span
      ↪else 'N/A'

         # Find product price
         product_price_span = div.find('span', class_='a-price-whole')
         product_price = product_price_span.get_text(strip=True) if
      ↪product_price_span else 'N/A'

         # Append the data to the lists
         product_names.append(product_name)
         product_prices.append(product_price)
```

1

```python
# Step 5: Create a DataFrame and write to an Excel file
data = {
    'Product Name': product_names,
    'Product Price': product_prices
}

# Convert the data to a DataFrame
df = pd.DataFrame(data)

# Write the DataFrame to an Excel file
df.to_excel('Amazon_Products.xlsx', index=False)

print("Data has been successfully written to Amazon_Products.xlsx")
```

```
Data has been successfully written to Amazon_Products.xlsx
```

```python
[1]: # Example using a full path
with open('Amazon.html', 'r', encoding='utf-8') as file:
    html_content = file.read()
```

```python
[4]: # Import necessary libraries
from bs4 import BeautifulSoup
import pandas as pd

# Step 1: Read the Amazon.html file
with open('Amazon2.html', 'r', encoding='utf-8') as file:
    html_content = file.read()

# Step 2: Parse the HTML content with BeautifulSoup
soup = BeautifulSoup(html_content, 'html.parser')

# Step 3: Find all <div> with the specified class
divs = soup.find_all('div', class_="puis-card-container s-card-container
  ↪s-overflow-hidden aok-relative puis-include-content-margin puis
  ↪puis-v3az2sezl0cvhm234391inayc4r s-latency-cf-section puis-card-border")

# Initialize lists to store the extracted data
product_names = []
product_prices = []

# Step 4: Loop through each <div> and find the <span> for product name and price
for div in divs:
    # Find product name
    product_name_span = div.find('span', class_='a-size-medium a-color-base
  ↪a-text-normal')
    product_name = product_name_span.get_text(strip=True) if product_name_span
  ↪else 'N/A'
```

```python
    # Find product price
    product_price_span = div.find('span', class_='a-price-whole')
    product_price = product_price_span.get_text(strip=True) if 
 ↪product_price_span else 'N/A'

    # Append the data to the lists
    product_names.append(product_name)
    product_prices.append(product_price)

# Step 5: Create a DataFrame and write to an Excel file
data = {
    'Product Name': product_names,
    'Product Price': product_prices
}

# Convert the data to a DataFrame
df = pd.DataFrame(data)

# Write the DataFrame to an Excel file
df.to_excel('Amazon_Products.xlsx', index=False)

print("Data has been successfully written to Amazon_Products.xlsx")
```

```
Data has been successfully written to Amazon_Products.xlsx
```

[ ]: