Bellevue University

Milestone 4 Screenshots

Timothy Jelinek

DSC350-T301

7/28/2024

JelinekMilestone4.py

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

# scrapes the website
url = 'https://nflcombineresults.com/nflcombinedata.php'
response = requests.get(url)
soup = BeautifulSoup(response.content, 'html.parser')

# extracts the table body
table_body = soup.find('tbody')

# extracts data from the table body
rows = []
for tr in table_body.find_all('tr'):
    cells = [td.get_text().strip() for td in tr.find_all('td')]
    if cells:  # avoid empty rows
        rows.append(cells)

# converts rows to dataframe for processing
headers = ['Year', 'Name', 'College', 'Pos', 'Height (in)', 'Weight (lbs)', 'Wonderlic', '40 Yard', 'Bench Press', 'Vert Leap (in)', 'Broad Jump (in)', 'Shuttle', '3Cone']
df = pd.DataFrame(rows, columns=headers)

# identify duplicate rows in all columns
duplicates = df[df.duplicated(keep='first')]
```

```
26
27  # create new row
28  new_row1 = {
29      'Year': '2020',
30      'Name': 'Bob Stewart',
31      'College': 'Bellevue University',
32      'Pos': 'RB',
33      'Height (in)': '80',
34      'Weight (lbs)': '190',
35      'Wonderlic': '',
36      '40 Yard': '2.0',
37      'Bench Press': '18',
38      'Vert Leap (in)': '25',
39      'Broad Jump (in)': '110',
40      'Shuttle': '4.7',
41      '3Cone': '6.7'
42  }
43
44  # convert row to dataframe
45  new_row_df1 = pd.DataFrame([new_row1], columns=df.columns)
46
47  # add element using concat
48  df = pd.concat([df, new_row_df1], ignore_index=True)
49
```

```python
# create new row
new_row2 = {
    'Year': '1998',
    'Name': 'Matthew Cowards',
    'College': 'Cowards College',
    'Pos': 'CB',
    'Height (in)': '70',
    'Weight (lbs)': '170',
    'Wonderlic': '20',
    '40 Yard': '4.5',
    'Bench Press': '15',
    'Vert Leap (in)': '25',
    'Broad Jump (in)': '100',
    'Shuttle': '4.8',
    '3Cone': '7.5'
}

# convert new row to data frame
new_row_df2 = pd.DataFrame([new_row2], columns=df.columns)

# add element using concat
df = pd.concat([df, new_row_df2], ignore_index=True)

```

```python
# create new row
new_row3 = {
    'Year': '2024',
    'Name': 'John Doe',
    'College': 'Sample College',
    'Pos': 'QB',
    'Height (in)': '75',
    'Weight (lbs)': '230',
    'Wonderlic': '25',
    '40 Yard': '4.8',
    'Bench Press': '20',
    'Vert Leap (in)': '32',
    'Broad Jump (in)': '120',
    'Shuttle': '4.2',
    '3Cone': '7.0'
}

# convert the new row to dataframe
new_row_df3 = pd.DataFrame([new_row3], columns=df.columns)

# add element using concat
df = pd.concat([df, new_row_df3], ignore_index=True)

# clean the dataframe
df = df.applymap(lambda x: x.strip().lower())  # converts to lowercase and strips whitespace

# makes sure there is no NaN values in the dataframe
df.dropna(inplace=True)
```

```
102   # clear existing tbody
103   table_body.clear()
104
105   # add modified rows to the tbody
106   for index, row in df.iterrows():
107       tr = soup.new_tag('tr')
108       for value in row:
109           td = soup.new_tag('td')
110           td.string = value
111           tr.append(td)
112       table_body.append(tr)
113
114   # save updated HTML to HTML file
115   with open('JelinekMilestone4.html', 'w', encoding='utf-8') as
      file:
116       file.write(str(soup))
```

2024	zak zinter	michigan	og	77.88	309						
2020	bob stewart	bellevue university	rb	80	190		2.0	18	25	110	4.
1998	matthew cowards	cowards college	cb	70	170	20	4.5	15	25	100	4.
2024	john doe	sample college	qb	75	230	25	4.8	20	32	120	4.