

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import requests
import re
from bs4 import BeautifulSoup as bs
import warnings
warnings.filterwarnings("ignore")

#The goal of my project is to track the improvement of various high school debaters (including myself)
#from an earlier debate tournament to a later one. While the results of this project are not that difficult
#to obtain, this project is important to me because it has both allowed me to work on my current Data Science skills
#while learning new techniques and has personal significance as someone who had a passion for debate in high school.
#In this project, I learned how to web scrape an HTML Table into a Pandas Dataframe, improved my advanced data
#cleaning skills, and used my data wrangling skills to merge datasets.
```

```
res1 = requests.get('https://www.tabroom.com/index/tourn/results/event_results.mhtml?
tourn_id=10887&result_id=73142')
```

```
soup1 = bs(res1.content, 'lxml')
```

```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN"
"http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd">\n<html>\n <head>\n <meta
content="text/html; charset=utf-8" http-equiv="Content-Type"/>\n <link href="/lib/css/reset.css"
rel="stylesheet" type="text/css"/>\n <link href="/lib/css/jqueryui.css" rel="stylesheet"
type="text/css"/>\n <link href="/lib/css/uniform.css" rel="stylesheet" type="text/css"/>\n <link
href="/lib/css/fonts/fonts.css" rel="stylesheet" type="text/css"/>\n <link
href="/lib/css/alertify.min.css" rel="stylesheet" type="text/css"/>\n <link
href="/lib/css/alertify-bootstrap.min.css" rel="stylesheet" type="text/css"/>\n <link
href="/lib/css/tabroom.css" rel="stylesheet" type="text/css"/>\n <title>\n Tabroom.com\n
</title>\n <script src="/lib/javascript/tabroom.min.js" type="text/javascript">\n </script>\n
<script src="/lib/javascript/tinymce/tinymce.min.js" type="text/javascript">\n </script>\n
<script>\n $(".alert").on("click", function(e){\n\t\t\t\t\tte.preventDefault();\n\t\t\t\t\ttvar
href = this.href;\n\t\t\t\t\ttalertify.confirm("Are you sure?", function (e) {\n\t\t\t\t\t\t\ttif (e)
{\n\t\t\t\t\t\t\t\t\twindow.location.href =
href;\n\t\t\t\t\t\t\t\t\t});\n\t\t\t\t\t\t});\n\n\t\t\t\t\t$(document).ready(function()
{\n\n\t\t\t\t\t\t\t// adds zebra striping to elements. in\n\t\t\t\t\t\t\t//
web/lib/javascript/tabroom.js\n\t\t\t\t\t\t\t\tzebraRows(); \n\n\t\t\t\t\t\t\t// put the focus on the first
editable element.\n\n\t\t\t\t\t\t\t$("*:input.enabled:visible:not(.notfirst):first").focus();\n\t\t\t\t\t\t\t$("*.starthere")
ocus());\n\n\t\t\t\t\t\t\t// Adds the searchable select menus to eligible
elements\n\t\t\t\t\t\t\t$("select:not(.plain)").chosen({\n\t\t\t\t\t\t\t\ttno_results_text: "Oops, nothing
found!", \n\t\t\t\t\t\t\t\ttdisable_search_threshold: 5 \n\t\t\t\t\t\t\t\t});\n\n\t\t\t\t\t\t\t// Fixes the bug w
here selects in hidden divs would have\n\t\t\t\t\t\t\t// zero width.\n\n\t\t\t\t\t\t\t$(".chosen-
container").each(function(i, chosenOne) {\n\t\t\t\t\t\t\t\t\tif ($(chosenOne).width() == 0)
{\n\t\t\t\t\t\t\t\t\t\t(chosenOne).width("100%");\n\t\t\t\t\t\t\t\t\t}\n\t\t\t\t\t\t\t\t});\n\n\t\t\t\t\t\t\t//
Control-s gives you the search box\n\t\t\t\t\t\t\t\tMousetrap.bind("'mod+s'", function(e)
{\n\t\t\t\t\t\t\t\t\t$("#searchtext").focus();\n\t\t\t\t\t\t\t\t\treturn
false;\n\t\t\t\t\t\t\t\t});\n\n\t\t\t\t\t\t\t// Sets up the little arrows to close and open the
sidebars\n\n\t\t\t\t\t\t\t$(".menu").prepend("<div class='shade openshade fa fa-forward'>
</div>");\n\n\t\t\t\t\t\t\t$("#content").prepend("<div class='hidden shade closedshade fa fa-
backward'></div>");\n\n\t\t\t\t\t\t\t$(".shade").click( function() {
\n\t\t\t\t\t\t\t\t\t$(".menu").toggleClass("hidden");\n\t\t\t\t\t\t\t\t\t$(".closedshade").toggleClass("hidden"
)\n\t\t\t\t\t\t\t\t\t$(".main").toggleClass("mainfull");\n\t\t\t\t\t\t\t\t\t};\n\n\t\t\t\t\t\t\t!-- This
```

[illegible]

[illegible]

[illegible]

```
ish">\n                2.00\n            </td>\n            <td class="rightalign smallish">\n                170.00\n        </td>\n        <td class="rightalign smallish">\n            282.00\n        </td>\n        <td class="rightalign smallish">\n            0.00\n        </td>\n        <td class="rightalign smallish">\n            6.93\n        </td>\n        <td class="hiddencsv">\n            R1:W2929R2:I925R3:W2929R4:L2828R5:L2828\n        </td>\n    </tr>\n    <tr>\n        <td class="centeralign">\n            19\n        </td>\n        <td>\n            Eagan HS\n        </td>\n        <td>\n            Hagape & Shah\n        </td>\n        <td class="rightalign smallish">\n            169.50\n        </td>\n        <td class="rightalign smallish">\n            283.00\n        </td>\n        <td class="rightalign smallish">\n            0.00\n        </td>\n        <td class="rightalign smallish">\n            5.9\n        </td>\n        <td class="hiddencsv">\n            R1:W3028.5R2:L2827R3:L2928R4:L27.529R5:W2927\n        </td>\n    </tr>\n    <tr>\n        <td class="centeralign">\n            20\n        </td>\n        <td>\n            Eagan LS\n        </td>\n        <td>\n            Lee & San Jose\n        </td>\n        <td class="rightalign smallish">\n            169.50\n        </td>\n        <td class="rightalign smallish">\n            281.50\n        </td>\n        <td class="rightalign smallish">\n            0.00\n        </td>\n        <td class="rightalign smallish">\n            0.08\n        </td>\n        <td class="hiddencsv">\n            R1:L2928.5R2:L2727R3:W2929R4:W2928R5:L27.527.5\n        </td>\n    </tr>\n    <tr>\n        <td class="centeralign">\n            21\n        </td>\n        <td>\n            SPA MS\n        </td>\n        <td>\n            McCauley & Singh\n        </td>\n        <td class="rightalign smallish">\n            169.00\n        </td>\n        <td class="rightalign smallish">\n            280.00\n        </td>\n        <td class="rightalign smallish">\n            0.00\n        </td>\n        <td class="hiddencsv">\n            R1:L2727R2:W2829R3:L28.527R4:W2828.5R5:L2829\n        </td>\n    </tr>\n    <tr>\n        <td class="centeralign">\n            22\n        </td>\n        <td>\n            Edina WJ\n        </td>\n        <td>\n            Wu & Jiang\n        </td>\n        <td class="rightalign smallish">\n            169.00\n        </td>\n        <td class="rightalign smallish">\n            279.50\n        </td>\n        <td class="rightalign smallish">\n            0.00\n        </td>\n        <td class="rightalign smallish">\n            0.68\n        </td>\n        <td class="hiddencsv">\n            R1:L2728R2:I729R3:W29.529R4:W2626R5:L2929\n        </td>\n    </tr>\n    <tr>\n        <td class="centeralign">\n            23\n        </td>\n        <td>\n            SPA HM\n        </td>\n        <td>\n            Hoeglund & Moran\n        </td>\n        <td class="rightalign smallish">\n            168.00\n        </td>\n        <td class="rightalign smallish">\n            279.00\n        </td>\n        <td class="rightalign smallish">\n            0.00\n        </td>\n        <td class="rightalign smallish">\n            8.7\n        </td>\n        <td class="hiddencsv">\n            R1:W2928R2:L2629R3:L2729R4:L2727R5:W2928\n        </td>\n    </tr>\n    <tr>\n        <td class="centeralign">\n            24\n        </td>\n        <td>\n            Eagan HL\n        </td>\n        <td>\n            Hedeman & Li\n        </td>\n        <td class="rightalign smallish">\n            167.50\n        </td>\n        <td class="rightalign smallish">\n            279.00\n        </td>\n        <td class="rightalign smallish">\n            0.00\n        </td>\n        <td class="rightalign smallish">\n            4.14\n        </td>\n        <td class="hiddencsv">\n            R1:L2828R2:W829.5R3:L29.529R4:L2627R5:W2727\n        </td>\n    </tr>\n    <tr>\n        <td class="centeralign">\n            25\n        </td>\n        <td>\n            Maple Grove GA\n        </td>\n        <td>\n            Golakoti & Ali\n        </td>\n        <td class="rightalign smallish">\n            167.00\n        </td>\n        <td class="rightalign smallish">\n            279.50\n        </td>\n        <td class="rightalign smallish">\n            0.00\n        </td>\n        <td class="rightalign smallish">\n            3.3\n        </td>\n        <td class="hiddencsv">\n            R1:L2828R2:W2828R3:W2827R4:L28.529R5:L2728\n        </td>\n    </tr>\n    <tr>\n        <td class="centeralign">\n            26\n        </td>\n        <td>\n            SPA CK\n        </td>\n        <td>\n            Chang & Keogh\n        </td>\n        <td class="rightalign smallish">\n            166.00\n        </td>\n        <td class="rightalign smallish">\n            277.00\n        </td>\n        <td class="rightalign smallish">\n            0.00\n        </td>\n        <td class="rightalign smallish">\n            8.45\n        </td>\n        <td class="hiddencsv">\n            R1:W2727R2:W829R3:L2828R4:L2727R5:L2828\n        </td>\n    </tr>\n    <tr>\n        <td class="centeralign">\n            27\n        </td>\n        <td>\n            SPA SV\n        </td>\n        <td>\n            Shaker-Check & Vliestra\n        </td>\n        <td class="rightalign smallish">\n            167.00\n        </td>\n        <td class="rightalign smallish">\n            273.00\n        </td>\n        <td class="rightalign smallish">\n            0.00\n        </td>\n        <td class="rightalign smallish">\n            4.71\n        </td>\n        <td class="hiddencsv">\n            R1:L2827R2:W626R3:W2828R4:L2828R5:L26.527.5\n        </td>\n    </tr>\n    <tr>\n        <td class="centeralign">\n            28\n        </td>\n        <td>\n            Eastview GV\n        </td>\n        <td>\n            Guttipatti & Vij\n        </td>\n        <td class="rightalign smallish">\n            169.00\n        </td>\n        <td class="rightalign smallish">\n            281.00\n        </td>\n        <td class="rightalign smallish">\n            0.00\n        </td>\n        <td class="rightalign smallish">\n            8.1\n        </td>\n        <td class="hiddencsv">\n            R1:L2728R2:L2828R3:W2928R4:L2828.5R5:L28.528\n        </td>\n    </tr>\n    <tr>\n        <td class="centeralign">\n            29\n        </td>\n        <td>\n            Benilde-St. Margarita's JK\n        </td>\n        <td>\n            Jensen & Krueger\n        </td>
```

```

class="rightalign smallish">\n          1.00\n          </td>\n          <td class="rightalign  

smallish">\n          166.50\n          </td>\n          <td class="rightalign smallish">\n          0.00\n          </td>\n
76.00\n          </td>\n          <td class="rightalign smallish">\n          0.00\n          </td>\n
<td class="rightalign smallish">\n          9.52\n          </td>\n          <td class="hiddencsv">\n
R1:L2828R2:W2528R3:L28.528R4:L2727R5:L27.529\n          </td>\n          </tr>\n          <tr>\n
<td class="centeralign">\n          30\n          </td>\n          <td>\n          Bloomington Jeffer  

on MM\n          </td>\n          <td>\n          Maki & Matias\n          </td>\n          <td cla  

s="rightalign smallish">\n          1.00\n          </td>\n          <td class="rightalign  

smallish">\n          163.50\n          </td>\n          <td class="rightalign smallish">\n          0.00\n          </td>\n
75.50\n          </td>\n          <td class="rightalign smallish">\n          0.00\n          </td>\n
<td class="rightalign smallish">\n          5.09\n          </td>\n          <td class="hiddencsv">\n
R1:L2628R2:L2728R3:L2627R4:W2930R5:L28.526\n          </td>\n          </tr>\n          <tr>\n          <
d class="centeralign">\n          31\n          </td>\n          <td>\n          Edina ST\n          <
td>\n          <td>\n          Smalley & Tessmer-Tuck\n          </td>\n          <td class="right  

align smallish">\n          1.00\n          </td>\n          <td class="rightalign smallish">\n          270.50\n          </td>
n          <td class="rightalign smallish">\n          0.00\n          </td>\n          <td class="rig  

talign smallish">\n          2.88\n          </td>\n          <td class="hiddencsv">\n          R1:L2  

26R2:L2828R3:L2728R4:W25.527R5:L2727\n          </td>\n          </tr>\n          <tr>\n          <td cla  

s="centeralign">\n          32\n          </td>\n          <td>\n          Eastview GP\n          </tc  

\n          <td>\n          Gouette & Pruidze\n          </td>\n          <td class="rightalign sm  

llish">\n          0.00\n          </td>\n          <td class="rightalign smallish">\n          158.C  

\n          </td>\n          <td class="rightalign smallish">\n          264.00\n          </td>\n
<td class="rightalign smallish">\n          0.00\n          </td>\n          <td class="rightalign  

smallish">\n          0.10\n          </td>\n          <td class="hiddencsv">\n          R1:L2525R2:I  

6.529.5R3:L2726R4:L2526R5:L2727\n          </td>\n          </tr>\n          </tbody>\n          </table>\n
</div>\n          </div>\n          </div>\n          <div id="footer">\n          <span class="leftlinks twothirds">\n          <
a href="https://www.speechanddebate.org/join" tabindex="-1">\n          Join the National Speech &
Debate Association\n          </a>\n          <a href="https://www.speechanddebate.org/mission" tabindex="-1"
">\n          About\n          </a>\n          <a href="mailto:help@tabroom.com" tabindex="-1">\n          Help\n
</a>\n          <a href="https://www.tabroom.com/index/about.mhtml" tabindex="-1">\n          Contact\n
</a>\n          </span>\n          </div>\n          </div>\n          </body>\n          </html>\n          '

```

In [5]:

```
table1 = soup1.select('table')[0]
```

In [6]:

```
columns1 = table1.find('thead').find_all('th')
```

In [7]:

```
column_names1 = [c.string for c in columns1]
```

In [8]:

```
table_rows1 = table1.find('tbody').find_all('tr')
```

In [9]:

```
list1 = []
```

In [10]:

```

#After extracting and soupifying the HTML table, I put the contents into a list for my Pandas Data
Frame.

for tr in table_rows1:
    td = tr.find_all('td')
    row = [str(tr.get_text()).strip() for tr in td]
    list1.append(row)

```

In [11]:

```
df1 = pd.DataFrame(list1, columns = column_names1)
```

In [12]:

Out[12]:

In [13]:

In [14]:

Out [14]:

In [15]:

In [16]:

Out[16]:

In [17]:

In [18]:

[illegible]


```

orm action="/index/search.shtml">\n        <input autocapitalize="off" autocomplete="off" autocorrec
t="off" class="notfirst" id="searchtext" maxlength="128" name="search" placeholder="SEARCH
TOURNAMENTS" size="15" spellcheck="false" tabindex="-1" type="text"/>\n            <input name="caller"
type="hidden" value="/index/tourn/results/event_results.mhtml?
tourn_id=10671&amp;result_id=70544"/>\n                <button class="search notfirst" type="submit">\n
\n                    </button>\n                </form>\n            </span>\n            <span id="
helpbutton" title="Tabroom Help">\n                <a class="fa fa-question-circle"
href="http://docs.tabroom.com" tabindex="-1" target="_blank">\n                    </a>\n                </span>\n            </spar
>\n            <span id="menus">\n                <ul id="nav">\n                    <li class="top">\n                        <a class="centeralign
top_link" href="/">\n                            <span class="down">\n                                Home\n                            </span>\n                        </a>\n
</li>\n                    <li class="top">\n                        <a class="centeralign top_link"
href="/index/circuits.mhtml">\n                            <span class="down">\n                                Circuits\n                            </span>\n
</a>\n                    </li>\n                    <li class="top">\n                        <a class="centeralign top_link"
href="/index/results">\n                            <span class="down">\n                                Results\n                            </span>\n
</a>\n                    </li>\n                    <li class="top">\n                        <a class="centeralign top_link"
href="/index/paradigm.mhtml">\n                            <span class="down">\n                                Paradigms\n                            </span>\n
</a>\n                    </li>\n                    <li class="top">\n                        <a class="centeralign top_link"
href="/index/help.mhtml">\n                            <span class="down">\n                                Help\n                            </span>\n
</a>\n                    </li>\n                    <li class="top">\n                        <a class="centeralign top_link"
href="/index/about.mhtml">\n                            <span class="down">\n                                About\n                            </span>\n
</a>\n                    </li>\n                </ul>\n            <span>\n            </span>\n            </div>\n            <div
id="wrapper">\n                <div id="content">\n                    <div class="menu">\n                        <div class="sidenote">\n
<h4>\n                            Event Results\n                        </h4>\n                        <form action="index.mhtml" method="post">\n
<input name="tourn_id" type="hidden" value="10671"/>\n                            <div class="row centeralign">\n
<span class="threequarters">\n                                <select class="fixedmost" name="event_id"
onchange="this.form.submit();">\n                                    <option value="88088">\n                                        Congress\n
</option>\n                                    <option value="88089">\n                                        Junior Varsity Lincoln-Douglas\n
</option>\n                                    <option value="88090">\n                                        Junior Varsity Policy\n
</option>\n                                    <option value="88091">\n                                        Junior Varsity Public Forum\n
</option>\n                                    <option value="88092">\n                                        Novice Lincoln-Douglas\n
</option>\n                                    <option value="88093">\n                                        Novice Policy\n
</option>\n                                    <option
value="88094">\n                                        Novice Public Forum\n
</option>\n                                    <option value="88
95">\n                                        Varsity Lincoln-Douglas\n
</option>\n                                    <option value="88096">\n
Varsity Policy\n
</option>\n                                    <option selected="" value="88097">\n
Varsity Public Forum\n
</option>\n                                </select>\n                            </span>\n                            <span c
ass="quarter">\n                                <input class="thin" type="submit" value="Go"/>\n                            </span>\n
</div>\n                        </form>\n                    </div>\n                    <div class="sidenote">\n                        <h4>\n                            VPF
Results\n                        </h4>\n                        <a class="dkyellow full nowrap"
href="/index/tourn/results/event_results.mhtml?tourn_id=10671&amp;result_id=70544">\n                            Final
Places\n                        </a>\n                        <a class="yellow full nowrap"
href="/index/tourn/results/event_results.mhtml?tourn_id=10671&amp;result_id=70547">\n                            Speake
r Awards\n                        </a>\n                        <a class="yellow full nowrap" href="debate_cumesheet.mhtml?event_id=8
8097&amp;tourn_id=10671">\n                            Cume Sheet\n                        </a>\n                        <a class="blue full nowrap" href
="/index/tourn/results/round_results.mhtml?tourn_id=10671&amp;round_id=342075">\n                            <span clas
s="half padno marno semibold">\n                                Round 5\n                            </span>\n                            <span class="half
rightalign nowrap padno marno">\n                                Round results\n                            </span>\n                        </a>\n                        <a
class="blue full nowrap" href="/index/tourn/results/round_results.mhtml?
tourn_id=10671&amp;round_id=342074">\n                            <span class="half padno marno semibold">\n                                Rc
und 4\n                            </span>\n                            <span class="half rightalign nowrap padno marno">\n                                Round r
esults\n                            </span>\n                        </a>\n                        <a class="blue full nowrap"
href="/index/tourn/results/round_results.mhtml?tourn_id=10671&amp;round_id=342073">\n                            <span
class="half padno marno semibold">\n                                Round 3\n                            </span>\n                            <span class="half
rightalign nowrap padno marno">\n                                Round results\n                            </span>\n                        </a>\n                        <a
class="blue full nowrap" href="/index/tourn/results/round_results.mhtml?
tourn_id=10671&amp;round_id=342072">\n                            <span class="half padno marno semibold">\n                                Rc
und 2\n                            </span>\n                            <span class="half rightalign nowrap padno marno">\n                                Round r
esults\n                            </span>\n                        </a>\n                        <a class="blue full nowrap"
href="/index/tourn/results/round_results.mhtml?tourn_id=10671&amp;round_id=342071">\n                            <span
class="half padno marno semibold">\n                                Round 1\n                            </span>\n                            <span class="half
rightalign nowrap padno marno">\n                                Round results\n                            </span>\n                        </a>\n                        <a
class="blue full" href="ranked_list.mhtml?event_id=88097&amp;tourn_id=10671">\n                            Prelim Recor
ds\n                        </a>\n                    </div>\n                    <div class="sidenote">\n                        <h4>\n                            Tournament-Wide\rs
</h4>\n                        <a class="yellow full" href="https://s3.amazonaws.com/tabroom-
files/tourns/10671/results/13045/CongressScores-RosemountIrishInvitational-1538412919.pdf">\n
CongressScores.Prelims.Super\n                        </a>\n                    </div>\n                    <div class="sidenote">\n                        <h4>\n
                            Tiebreakers\n                        </h4>\n                        <p>\n                            You can hover your mouse over the title of
tiebreaker column\n\t\t\t\tfor a fuller explanation of that tiebreaker.\n                        </p>\n                    </div>\n
                    <div class="main">\n                        <h2 class="centeralign marno">\n                            Rosemount Irish
Invitational\n                        </h2>\n                        <h5 class="full centeralign marno">\n                            2018 \n\t\t\t-\n\t\t\tRose
mount,\n\t\t\tMN/US\n                        </h5>\n                        <ul id="tabnav">\n                            <li class="">\n                                <a
href="/index/tourn/index.mhtml?tourn_id=10671">\n                                    Invite\n                                </a>\n                            </li>\n
                            <li class="">\n                                <a href="/index/tourn/fields.mhtml?tourn_id=10671">\n                                    Entries\n
                                </a>\n                            </li>\n                            <li class="">\n                                <a href="/index/tourn/judges.mhtml?
tourn_id=10671">\n                                    Judges\n                                </a>\n                            </li>\n                            <li class="">\n                                <a hre
="/index/tourn/postings/index.mhtml?tourn_id=10671">\n                                    Pairings\n                                </a>\n
                            </li>\n                            <liclass="selected">\n                                <a href="/index/tourn/results/index.mhtml?tourn_id=106
71">\n                                    Results\n                                </a>\n                            </li>\n                        </ul>\n                    <div class="full nospace pad

```

[illegible]

R1:W27.529R2:W2829R3:W2830R4:W28.528.7R5:L2826				
lish">\n	4\n	</td>\n	<td class="rightalign smallish">\n	170.7\n
</td>\n	<td class="rightalign smallish">\n	57\n	</td>\n	<td class="rightalign smallish">\n
rightalign smallish">\n	284.7\n	</td>\n	<td class="rightalign smallish">\n	r
2.8\n	</td>\n	<td class="rightalign smallish">\n	3.95\n	</td>\n
</tr>\n	<tr>\n	<td class="centeralign">\n	Prelim\n	</td>\n
td>\n	Eastview AA\n	</td>\n	<td>\n	Alapati & Addagudi\n
</td>\n	<td class="hiddecsv">\n	R1:W28.929.1R2:W2928R3:L3029R4:L2828R5:W2929\n		
</td>\n	<td class="rightalign smallish">\n	3\n	</td>\n	<td class="rightalign smallish">\n
ightalign smallish">\n	173\n	</td>\n	<td class="rightalign smallish">\n	
58\n	</td>\n	<td class="rightalign smallish">\n	288\n	</td>\n
<td class="rightalign smallish">\n	2.6\n	</td>\n	<td class="rightalign smallish">\n	
smallish">\n	6.8\n	</td>\n	</tr>\n	<tr>\n
class="centeralign">\n	Prelim\n	</td>\n	<td>\n	Eastview GV\n
</td>\n	<td>\n	Guttipatti & Vij\n	</td>\n	<td>\n
class="hiddecsv">\n	R1:W29.529R2:W28.528.8R3:L26.526.5R4:L2928.7R5:W2829\n			</td>
n	<td class="rightalign smallish">\n	3\n	</td>\n	<td class="rightalign smallish">\n
ign smallish">\n	172\n	</td>\n	<td class="rightalign smallish">\n	
57.3\n	</td>\n	<td class="rightalign smallish">\n	283.5\n	</td>\n
<td class="rightalign smallish">\n	2.8\n	</td>\n	<td class="rightalign smallish">\n	
smallish">\n	4.67\n	</td>\n	</tr>\n	<tr>\n
class="centeralign">\n	Prelim\n	</td>\n	<td>\n	Eastview SC\n
</td>\n	<td>\n	Saravanapandi & Chatterjee\n	</td>\n	<td class="rightalign smallish">\n
hiddecsv">\n	R1:W2727R2:W2928R3:L2828R4:L2828.3R5:W3029\n		</td>\n	<td>\n
class="rightalign smallish">\n	3\n	</td>\n	<td class="rightalign smallish">\n	
smallish">\n	169.3\n	</td>\n	<td class="rightalign smallish">\n	5
.3\n	</td>\n	<td class="rightalign smallish">\n	282.3\n	</td>\n
<td class="rightalign smallish">\n	2.8\n	</td>\n	<td class="rightalign smallish">\n	
smallish">\n	0.46\n	</td>\n	</tr>\n	<tr>\n
class="centeralign">\n	Prelim\n	</td>\n	<td>\n	Eagan VH\n
</td>\n	<td>\n	Victor & Hagape\n	</td>\n	<td>\n
class="hiddecsv">\n	R1:L28.528R2:L28.128.2R3:W2828R4:W2830R5:W2728\n			</td>
<td class="rightalign smallish">\n	3\n	</td>\n	<td class="rightalign smallish">\n	
smallish">\n	168.8\n	</td>\n	<td class="rightalign smallish">\n	5
.3\n	</td>\n	<td class="rightalign smallish">\n	281.8\n	</td>\n
<td class="rightalign smallish">\n	2.6\n	</td>\n	<td class="rightalign smallish">\n	
smallish">\n	2.73\n	</td>\n	</tr>\n	<tr>\n
class="centeralign">\n	Prelim\n	</td>\n	<td>\n	Eastview DZ\n
</td>\n	<td>\n	Dulla & Zhu\n	</td>\n	<td class="hiddecsv">\n
R1:W2827R2:W29.528.5R3:L2727R4:W27.527.5R5:L2829\n		</td>\n	<td class="rightalign smallish">\n	sm
allish">\n	3\n	</td>\n	<td class="rightalign smallish">\n	167\n
</td>\n	<td class="rightalign smallish">\n	55\n	</td>\n	<td class="rightalign smallish">\n
rightalign smallish">\n	279\n	</td>\n	<td class="rightalign smallish">\n	
3.2\n	</td>\n	<td class="rightalign smallish">\n	9.48\n	</td>\n
</tr>\n	<tr>\n	<td class="centeralign">\n	Prelim\n	</td>\n
td>\n	Edina SI\n	</td>\n	<td>\n	Shilts & Ishani\n
>\n	<td class="hiddecsv">\n	R1:W2827R2:W2829R3:W2728R4:L2526R5:L2827\n		
</td>\n	<td class="rightalign smallish">\n	3\n	</td>\n	<td class="rightalign smallish">\n
ightalign smallish">\n	165\n	</td>\n	<td class="rightalign smallish">\n	
55\n	</td>\n	<td class="rightalign smallish">\n	273\n	</td>\n
<td class="rightalign smallish">\n	2.8\n	</td>\n	<td class="rightalign smallish">\n	
smallish">\n	4.27\n	</td>\n	</tr>\n	<tr>\n
class="centeralign">\n	Prelim\n	</td>\n	<td>\n	Eagan SG\n
</td>\n	<td>\n	Shah & Garishakurti\n	</td>\n	<td class="hiddecsv">\n
sv">\n	R1:L2526R2:W2827R3:W2829R4:L27.627.4R5:W27.527\n		</td>\n	<td>\n
class="rightalign smallish">\n	3\n	</td>\n	<td class="rightalign smallish">\n	
smallish">\n	164.5\n	</td>\n	<td class="rightalign smallish">\n	5
\n	</td>\n	<td class="rightalign smallish">\n	272.5\n	</td>\n
<td class="rightalign smallish">\n	2.8\n	</td>\n	<td class="rightalign smallish">\n	
smallish">\n	3.7\n	</td>\n	</tr>\n	<tr>\n
class="centeralign">\n	Prelim\n	</td>\n	<td>\n	Eagan CU\n
</td>\n	<td>\n	Chino & Ugboajah\n	</td>\n	<td>\n
class="hiddecsv">\n	R1:L29.229R2:L2829R3:W2629R4:W2828.5R5:L2929\n			</td>
<td class="rightalign smallish">\n				

ey & Tessmer-Tuck\n

R1:L2827R2:W2829R3:L2827R4:W2828.5R5:L28.529\n			
sh">\n	2\n		168.5\n
</td>\n	<td class="rightalign smallish">\n	56.5\n	</td>\n
<td class="rightalign smallish">\n	281\n	</td>\n	<td class="rightalign smallish">\n
2.4\n	</td>\n	<td class="rightalign smallish">\n	1.59\n
</tr>\n	<tr>\n	<td class="centralalign">\n	Prelim\n
td>\n	Eagan SP\n	</td>\n	Shailja Panchal\n
<td class="hiddencsv">\n	R1:L26R2:L28R3:W27.7R4:W28.5R5:L28\n	</td>\n	<td
class="rightalign smallish">\n	2\n	</td>\n	<td class="rightalign
smallish">\n	166.62\n	</td>\n	<td class="rightalign smallish">\n
5.64\n	</td>\n	<td class="rightalign smallish">\n	276.4\n
<td class="rightalign smallish">\n	2.6\n	</td>\n	<td class="rightalign
smallish">\n	7.49\n	</td>\n	</tr>\n
class="centralalign">\n	Prelim\n	</td>\n	<td>\n
n	</td>\n	<td>\n	Lakeville South NF
s="hiddencsv">\n	R1:BR2:W29.529R3:L2727R4:L2626R5:L2727\n	</td>\n	<td cla
class="rightalign smallish">\n	2\n	</td>\n	<td class="rightalign
smallish">\n	162.62\n	</td>\n	<td class="rightalign smallish">\n
4.62\n	</td>\n	<td class="rightalign smallish">\n	273.12\n
<td class="rightalign smallish">\n	2.75\n	</td>\n	<td class="rightalign
smallish">\n	7.42\n	</td>\n	</tr>\n
class="centralalign">\n	Prelim\n	</td>\n	<td>\n
</td>\n	<td>\n	Gouette & Pruidze\n	<td
class="hiddencsv">\n	R1:L2928R2:W2825R3:W2727R4:L27.528R5:L2626\n	</td>\n	<td>\n
<td class="rightalign smallish">\n	2\n	</td>\n	<td class="rightalign
smallish">\n	162.5\n	</td>\n	<td class="rightalign smallish">\n
\n	</td>\n	<td class="rightalign smallish">\n	271.5\n
<td class="rightalign smallish">\n	3\n	</td>\n	<td class="rightalign
smallish">\n	4.12\n	</td>\n	</tr>\n
class="centralalign">\n	Prelim\n	</td>\n	<td>\n
</td>\n	<td>\n	Sathe & Qian\n	<td class="hiddencsv">\n
R1:L2626R2:L2728.5R3:BR4:L2827.5R5:W26.526.5\n	</td>\n	<td class="rightalign smallish">\n	162.5\n
sh">\n	2\n	</td>\n	<td class="rightalign smallish">\n
</td>\n	<td class="rightalign smallish">\n	54\n	</td>\n
rightalign smallish">\n	270\n	</td>\n	<td class="rightalign smallish">\n
2.25\n	</td>\n	<td class="rightalign smallish">\n	6.96\n
</tr>\n	<tr>\n	<td class="centralalign">\n	Prelim\n
td>\n	Blake CZ\n	</td>\n	<td>\n
</td>\n	<td class="hiddencsv">\n	R1:L2626R2:L27.528R3:L2727R4:BR5:W27.626.8\n	</td>\n
</td>\n	<td class="rightalign smallish">\n	2\n	<td class="
ightalign smallish">\n	162.37\n	</td>\n	<td class="rightalign smallish">\n
53.97\n	</td>\n	<td class="rightalign smallish">\n	269.87\n
<td class="rightalign smallish">\n	3\n	</td>\n	<td class="rightalign
smallish">\n	7.28\n	</td>\n	</tr>\n
class="centralalign">\n	Prelim\n	</td>\n	<td>\n
</td>\n	<td>\n	Lu & Patel\n	<td class="hiddencsv">\n
R1:L2727R2:L2625R3:W2827R4:L2525R5:W28.228.1\n	</td>\n	<td class="rightalign smallish">\n	160\n
sh">\n	2\n	</td>\n	<td class="rightalign smallish">\n
</td>\n	<td class="rightalign smallish">\n	54\n	</td>\n
rightalign smallish">\n	266.3\n	</td>\n	<td class="rightalign smallish">\n
2.2\n	</td>\n	<td class="rightalign smallish">\n	9.89\n
</tr>\n	<tr>\n	<td class="centralalign">\n	Prelim\n
td>\n	Eastview SG\n	</td>\n	<td>\n
</td>\n	<td class="hiddencsv">\n	R1:W2827R2:L2828R3:L27.528.5R4:L2727R5:L2828\n	</td>\n
</td>\n	<td class="rightalign smallish">\n	1\n	<td class="
ightalign smallish">\n	167\n	</td>\n	<td class="rightalign smallish">\n
56\n	</td>\n	<td class="rightalign smallish">\n	277\n
<td class="rightalign smallish">\n	2.4\n	</td>\n	<td class="rightalign
smallish">\n	5.82\n	</td>\n	</tr>\n
class="centralalign">\n	Prelim\n	</td>\n	<td>\n
</td>\n	<td>\n	Eliza Beyers\n	<td class="hiddencsv">\n
R1:L27.5R2:L28R3:W28R4:L28R5:L26\n	</td>\n	<td class="rightalign smallish">\n	</td>\n
1\n	</td>\n	<td class="rightalign smallish">\n	166\n
<td class="rightalign smallish">\n	55.5\n	</td>\n	<td class="rightalign
smallish">\n	275\n	</td>\n	<td class="rightalign smallish">\n
n	</td>\n	<td class="rightalign smallish">\n	3.68\n
</tr>\n	<tr>\n	<td class="centralalign">\n	Prelim\n
td>\n	Bloomington Jefferson MR\n	</td>\n	<td>\n
n	</td>\n	<td class="hiddencsv">\n	</td>\n
R1:L2827R2:L2927.5R3:L27.626.8R4:L2829R5:W2726\n	</td>\n	<td class="rightalign smallish">\n	165.9\n
lish">\n	1\n	</td>\n	<td class="rightalign smallish">\n
</td>\n	<td class="rightalign smallish">\n	55\n	</td>\n
rightalign smallish">\n	275.9\n	</td>\n	<td class="rightalign smallish">\n
2\n	</td>\n	<td class="rightalign smallish">\n	1.28\n
</tr>\n	<tr>\n	<td class="centralalign">\n	Prelim\n
td>\n	Blake HP\n	</td>\n	<td>\n
			Homan & Pradhan\n

```

</td>\n          <td class="hiddencsv">\n          R1:L2828R2:L2728R3:L2727.5R4:L2727.5R5:B\n
</td>\n          <td class="rightalign smallish">\n          1\n          </td>\n          <td class="rightalign smallish">\n          164.5\n          </td>\n          <td class="rightalign smallish">\n          55\n          </td>\n          <td class="rightalign smallish">\n          275\n          </td>\n
<td class="rightalign smallish">\n          2.25\n          </td>\n          <td class="rightalign smallish">\n          4.85\n          </td>\n          </tr>\n          <tr>\n          <td class="centeralign">\n          Prelim\n          </td>\n          <td>\n          Sciamanda & Hentges\n          </td>\n          <td class="hiddencsv">\n          R1:L2727R2:BR3:L2727R4:L2828.5R5:L26.526.5\n          </td>\n          <td class="rightalign smallish">\n          1\n          </td>\n          <td class="rightalign smallish">\n          162.37\n          </td>\n          <td class="rightalign smallish">\n          4.37\n          </td>\n          <td class="rightalign smallish">\n          271.87\n          </td>\n
<td class="rightalign smallish">\n          2.75\n          </td>\n          <td class="rightalign smallish">\n          4.66\n          </td>\n          </tr>\n          <tr>\n          <td class="centeralign">\n          Prelim\n          </td>\n          <td>\n          Nova debaters FG\n          </td>\n          <td>\n          Flesher & Gidwani\n          </td>\n          <td class="hiddencsv">\n          R1:L2727R2:L2627R3:L2526R4:W3028R5:L26.226\n          </td>\n          <td class="rightalign smallish">\n          1\n          </td>\n          <td class="rightalign smallish">\n          159.2\n          </td>\n          <td class="rightalign smallish">\n          268.2\n          </td>\n
<td class="rightalign smallish">\n          2.4\n          </td>\n          <td class="rightalign smallish">\n          6.91\n          </td>\n          </tr>\n          </tbody>\n          </table>\n          </div>\n          </div>\n          </div>\n          <div id="footer">\n          <span class="leftlinks twothirds">\n          <a href="https://www.speechanddebate.org/join" tabindex="-1">\n          Join the National Speech & Debate Association\n          </a>\n          <a href="https://www.speechanddebate.org/mission" tabindex="-1">\n          >\n          About\n          </a>\n          <a href="mailto:help@tabroom.com" tabindex="-1">\n          Help\n          </a>\n          <a href="https://www.tabroom.com/index/about.mhtml" tabindex="-1">\n          Contact\n          </a>\n          </span>\n          </div>\n          </div>\n          </body>\n          </html>\n          '

```

In [20]:

```
table2 = soup2.select('table')[0]
```

In [21]:

```
columns2 = table2.find('thead').find_all('th')
```

In [22]:

```
column_names2 = [c.string for c in columns2]
```

In [23]:

```
table_rows2 = table2.find('tbody').find_all('tr')
```

In [24]:

```
list2 = []
```

In [25]:

```

for tr in table_rows2:
    td = tr.find_all('td')
    row = [str(tr.get_text()).strip() for tr in td]
    list2.append(row)

```

In [26]:

```
df2 = pd.DataFrame(list2, columns = column_names2)
```

In [27]:

```

#Here is our second Pandas DataFrame scraped from an HTML Webpage.
df2.head()

```

Out[27]:

	Place Place	Code Code	Name Name	Ballots	PWin	PPts	HL
0	Prelim	Eagan PS	Patel & Scheller	R1:W2829R2:W29.429.9R3:W28.529R4:W2828.5R5:W3030	5	173.8	
1	Prelim	Edina RF	Rewalliwar & Fu	R1:W2726R2:W2728R3:W27.127.5R4:W2929R5:W2929	5	167.6	
2	Prelim	Lakeville South MG	McCoy & Gerlach	R1:W29.529R2:W3030R3:L27.125.1R4:W3030R5:W2929	4	176.5	
3	Prelim	Lakeville South CN	Cook & Norcia	R1:W2929R2:L2929R3:W2928.5R4:W2929R5:W27.527	4	173.5	
4	Prelim	Eastview TK	Thota & Kapur	R1:W3029R2:L29.229.1R3:W2928.5R4:W2626R5:W2829	4	172.8	

In [28]:

```
#More Data Cleaning Processes Here
df2 = df2.iloc[:, :6]
```

In [29]:

```
df2.columns
```

Out[29]:

```
Index([' Place ', ' Code ', ' Name ', '\n\t\t\t\t\t\t\tBallots\n\t\t\t\t\t\t\t',
      '\n\t\t\t\t\t\t\tPWin\n\t\t\t\t\t\t\t',
      '\n\t\t\t\t\t\t\tPPts -1HL\n\t\t\t\t\t\t\t'],
      dtype='object')
```

In [30]:

```
del df2['\n\t\t\t\t\t\t\tBallots\n\t\t\t\t\t\t\t']
```

In [31]:

```
df2 = df2.rename({'\n\t\t\t\t\t\t\tPWin\n\t\t\t\t\t\t\t': 'Wins', '\n\t\t\t\t\t\t\tPPts -1HL\n\t\t\t\t\t\t\t': 'Speaker Points'}, axis=1)
```

In [32]:

```
for index, row in df2.iterrows():
    row[' Place '] = index + 1
```

In [33]:

```
df2.head()
```

Out[33]:

	Place	Code	Name	Wins	Speaker Points
0	1	Eagan PS	Patel & Scheller	5	173.8
1	2	Edina RF	Rewalliwar & Fu	5	167.6
2	3	Lakeville South MG	McCoy & Gerlach	4	176.5
3	4	Lakeville South CN	Cook & Norcia	4	173.5
4	5	Eastview TK	Thota & Kapur	4	172.8

In [34]:

```
df1['Wins'] = pd.to_numeric(df1["Wins"])
df1[' Place '] = pd.to_numeric(df1[' Place '])
df1['Speaker Points'] = pd.to_numeric(df1['Speaker Points'])
```

In [35]:

```
df1['Wins'] = df1['Wins'].astype(int)
```

In [36]:

```
df2['Wins'] = pd.to_numeric(df2["Wins"])
df2[' Place '] = pd.to_numeric(df2[' Place '])
df2['Speaker Points'] = pd.to_numeric(df2['Speaker Points'])
```

In [37]:

```
#This was a very advanced data cleaning process. In our data, sometimes the 'Code' column has different values
#but represent the same thing. For example, Eastview TK is the same as Eastview KT, but based on the tournament,
#some use the first code, others use the second code. This would be very troublesome when it comes to merging our
#data based on the 'Code' column, so we need to standardize each DataFrame by changing the codes so that all the
#letter codes (e.g. TK) would be formatted in alphabetical order (e.g. Eastview TK -> Eastview KT)
for index, row in df1.iterrows():
    temp1 = df1[' Code '][index].split()
    temp4 = temp1[0]
    if len(temp1) == 3:
        temp5 = temp1[1]
        temp2 = temp1[-1]
        temp3 = temp2[::-1]
        if temp2 > temp3:
            temp2 = temp3
        if len(temp1) == 3:
            df1[' Code '][index] = temp4 + ' ' + temp5 + ' ' + temp2
        else:
            df1[' Code '][index] = temp4 + ' ' + temp2
```

In [38]:

```
#I apply the same process for the second DataFrame.
for index, row in df2.iterrows():
    temp1 = df2[' Code '][index].split()
    temp4 = temp1[0]
    if len(temp1) == 3:
        temp5 = temp1[1]
        temp2 = temp1[-1]
        temp3 = temp2[::-1]
        if temp2 > temp3:
            temp2 = temp3
        if len(temp1) == 3:
            df2[' Code '][index] = temp4 + ' ' + temp5 + ' ' + temp2
        else:
            df2[' Code '][index] = temp4 + ' ' + temp2
```

In [39]:

```
df1.head()
```

Out[39]:

	Place	Code	Name	Wins	Speaker Points
0	1	Eagan PS	Patel & Scheller	5	176.5
1	2	Eastview DZ	Dulla & Zhu	4	173.5
2	3	Denfeld AZ	Autumn Zierman	4	172.1
3	4	Eastview KT	Kapur & Thota	4	172.0
4	5	SPA BG	Bijwadia & Guinan	4	172.0

In [40]:

```
df2.head()
```

Out[40]:

	Place	Code	Name	Wins	Speaker Points
0	1	Eagan PS	Patel & Scheller	5	173.8
1	2	Edina FR	Rewalliwar & Fu	5	167.6
2	3	Lakeville South GM	McCoy & Gerlach	4	176.5
3	4	Lakeville South CN	Cook & Norcia	4	173.5
4	5	Eastview KT	Thota & Kapur	4	172.8

In [41]:

```
df2 = df2.drop_duplicates(' Code ')
```

In [42]:

```
#Now I begin the Data Wrangling process. Here, I merged the two dataframes based on the 'Code' column.  
df3 = pd.merge(df2, df1, on=' Code ', suffixes=('_T1', '_T2'))
```

In [43]:

```
#You may notice there is only 11 teams here instead of the 32 that were present in the first dataframe or the 29  
#in the second dataframe. This is because we performed an inner join on the merge, meaning that we are only going  
#to merge on the 'Code' column key for teams that participated in both tournaments. As we can see, only 11 teams  
#participated in both tournaments, so only 11 teams are present in the new, merged dataframe.  
df3
```

Out[43]:

	Place_T1	Code	Name_T1	Wins_T1	Speaker Points_T1	Place_T2	Name_T2	Wins_T2	Speaker Points_T2
0	1	Eagan PS	Patel & Scheller	5	173.8	1	Patel & Scheller	5	176.5
1	2	Edina FR	Rewalliwar & Fu	5	167.6	7	Fu & Rewalliwar	3	172.5
2	5	Eastview KT	Thota & Kapur	4	172.8	4	Kapur & Thota	4	172.0
3	6	Edina JM	Jones & Maynor	4	172.5	13	Jones & Maynor	3	169.0
4	7	Edina NT	Nissen & Teske	4	170.7	8	Teske & Nissen	3	171.5
5	9	Eastview GV	Guttiatti & Vij	3	172.0	28	Guttiatti & Vij	1	169.0
6	10	Eastview CS	Saravanapandi & Chatterjee	3	169.3	18	Chatterjee & Saravanapandi	2	170.0
7	12	Eastview DZ	Dulla & Zhu	3	167.0	2	Dulla & Zhu	4	173.5
8	17	Edina ST	Smalley & Tessmer-Tuck	2	168.5	31	Smalley & Tessmer-Tuck	1	162.0
9	20	Eastview GP	Gouette & Pruidze	2	162.5	32	Gouette & Pruidze	0	158.0
10	24	Eastview GS	Schlegel & Gupta	1	167.0	14	Gupta & Schlegel	3	168.0

In [44]:

```
#Creating some new data while still Data Cleaning.  
df3['Win Change'] = df3['Wins_T2'] - df3['Wins_T1']  
df3['Points Change'] = df3['Speaker Points_T2'] - df3['Speaker Points_T1']  
df3['Rank Change'] = df3[' Place_T1'] - df3[' Place_T2']
```


In [45]:

```
del df3[' Name _T2']
```

In [46]:

```
#Here I apply a function to my merged dataframe to see improvement among the debate teams.
def check_improvement(x,y):
    if (x > 0 and y >= 0) or (x >= 0 and y > 0):
        return('Yes')
    else:
        return('No')
```

In [47]:

```
df3['Improvement'] = df3.apply(lambda row: check_improvement(row['Win Change'], row['Points Change']
)),axis=1)
```

In [48]:

```
#Final Data Cleaning
df3 = df3.rename({' Name _T1': 'Name'}, axis=1)
```

In [49]:

```
df3 = df3.reindex(columns=[' Code ', 'Name', ' Place _T1', 'Wins_T1', 'Speaker Points_T1',
    ' Place _T2', 'Wins_T2', 'Speaker Points_T2', 'Win Change',
    'Points Change', 'Rank Change', 'Improvement'])
```

In [50]:

```
#Here's the final result. Not too shabby. This was a really cool project for me to work on my data
science skills
#while having fun with a dataset that has had a lot of significance in my life.
df3
```

Out[50]:

	Code	Name	Place _T1	Wins_T1	Speaker Points_T1	Place _T2	Wins_T2	Speaker Points_T2	Win Change	Points Change	Rank Change	Improvement
0	Eagan PS	Patel & Scheller	1	5	173.8	1	5	176.5	0	2.7	0	Yes
1	Edina FR	Rewalliwar & Fu	2	5	167.6	7	3	172.5	-2	4.9	-5	No
2	Eastview KT	Thota & Kapur	5	4	172.8	4	4	172.0	0	-0.8	1	No
3	Edina JM	Jones & Maynor	6	4	172.5	13	3	169.0	-1	-3.5	-7	No
4	Edina NT	Nissen & Teske	7	4	170.7	8	3	171.5	-1	0.8	-1	No
5	Eastview GV	Guttipatti & Vij	9	3	172.0	28	1	169.0	-2	-3.0	-19	No
6	Eastview CS	Saravanapandi & Chatterjee	10	3	169.3	18	2	170.0	-1	0.7	-8	No
7	Eastview DZ	Dulla & Zhu	12	3	167.0	2	4	173.5	1	6.5	10	Yes
8	Edina ST	Smalley & Tessmer-Tuck	17	2	168.5	31	1	162.0	-1	-6.5	-14	No
9	Eastview GP	Gouette & Pruidze	20	2	162.5	32	0	158.0	-2	-4.5	-12	No
10	Eastview GS	Schlegel & Gupta	24	1	167.0	14	3	168.0	2	1.0	10	Yes