









	<pre>df2004.info()  &lt;class 'pandas.core.frame.DataFrame'&gt; RangeIndex: 11 entries, 0 to 10 Data columns (total 12 columns):  #   Column      Non-Null Count  Dtype ---  --  0   date        11 non-null      datetime64[ns]  1   attendance  11 non-null      int64  2   TSUreceivingyards  11 non-null      int64  3   TSUreceiveyards  11 non-null      int64  4   TSUkreturnyards  11 non-null      int64  5   TSUpreturmyards  11 non-null      int64  6   TSUackles     11 non-null      int64  7   TSUackleyd    11 non-null      float64  8   TSUsacks     11 non-null      float64  9   TSUsackyd    11 non-null      int64 10   TSUpunt      11 non-null      int64 11   year         11 non-null      int64 dtypes: datetime64[ns](1), float64(1), int64(10) memory usage: 1.2 KB</pre>																																																																																			
In [104]	<pre>#sort data df2004=df2004.sort_values('date',ignore_index=True) df2004.head()</pre>																																																																																			
Out[104]	<table><tr><th></th><th>date</th><th>attendance</th><th>TSUurshyards</th><th>TSUreceiveyards</th><th>TSUkreturnyards</th><th>TSUpreturmyards</th><th>TSUackles</th><th>TSUackleyd</th><th>TSUsacks</th><th>TSUsackyd</th><th>TSUpunt</th></tr><tr><td>0</td><td>2004-09-04</td><td>25117</td><td>152</td><td>203</td><td>24</td><td>26</td><td>74</td><td>35</td><td>2.0</td><td>12</td><td></td></tr><tr><td>1</td><td>2004-09-09</td><td>7019</td><td>156</td><td>158</td><td>34</td><td>32</td><td>65</td><td>60</td><td>4.0</td><td>30</td><td></td></tr><tr><td>2</td><td>2004-09-18</td><td>55015</td><td>164</td><td>149</td><td>39</td><td>10</td><td>63</td><td>19</td><td>2.0</td><td>7</td><td></td></tr><tr><td>3</td><td>2004-09-25</td><td>67712</td><td>161</td><td>163</td><td>49</td><td>32</td><td>78</td><td>70</td><td>5.0</td><td>49</td><td></td></tr><tr><td>4</td><td>2004-10-02</td><td>51082</td><td>22</td><td>173</td><td>57</td><td>0</td><td>91</td><td>17</td><td>1.0</td><td>4</td><td></td></tr></table>													date	attendance	TSUurshyards	TSUreceiveyards	TSUkreturnyards	TSUpreturmyards	TSUackles	TSUackleyd	TSUsacks	TSUsackyd	TSUpunt	0	2004-09-04	25117	152	203	24	26	74	35	2.0	12		1	2004-09-09	7019	156	158	34	32	65	60	4.0	30		2	2004-09-18	55015	164	149	39	10	63	19	2.0	7		3	2004-09-25	67712	161	163	49	32	78	70	5.0	49		4	2004-10-02	51082	22	173	57	0	91	17	1.0	4	
	date	attendance	TSUurshyards	TSUreceiveyards	TSUkreturnyards	TSUpreturmyards	TSUackles	TSUackleyd	TSUsacks	TSUsackyd	TSUpunt																																																																									
0	2004-09-04	25117	152	203	24	26	74	35	2.0	12																																																																										
1	2004-09-09	7019	156	158	34	32	65	60	4.0	30																																																																										
2	2004-09-18	55015	164	149	39	10	63	19	2.0	7																																																																										
3	2004-09-25	67712	161	163	49	32	78	70	5.0	49																																																																										
4	2004-10-02	51082	22	173	57	0	91	17	1.0	4																																																																										
In [105]	<pre>#save data frame to folder df2004.to_csv('2004.csv',encoding='utf-8')</pre>																																																																																			
In [106]	<pre>#2005 data #use requests.get() to get web page with data page = requests.get('https://tennstate.fsp.sidearmsports.com/custompages/tcutigers/99872882-1284-4887-8990-249f0041info') #parse data on web page using html module.fromstring mytree = html.fromstring(page.content)  #go to web address above , right click on page and select inspect to get HTML code for data from right side of #create XPath query and use xpath function to get data #date data date = mytree.xpath('body/center//tr/td/font[&amp;color="#000000"]/text()')[11:10:10] date=[a.replace("&lt;br&gt;","") for a in date] date=[a.replace(",","") for a in date] date=[a.replace(".", "") for a in date] print(date)</pre>																																																																																			
In [107]	<pre>['Sep 03 2005', 'Sep 10 2005', 'Sep 17 2005', 'Sep 24 2005', 'Oct 01 2005', 'Oct 13 2005', 'Oct 22 2005', 'Oct 29 2005', 'Nov 05 2005', 'Nov 12 2005', 'Nov 19 2005']</pre>																																																																																			
In [107]	<pre>#get attendance data attendance=mytree.xpath('body/center//tr/td/font[&amp;color="#000000"]/text()')[9:110:10] attendance=[a.replace("&lt;br&gt;","") for a in attendance] attendance=[a.strip() for a in attendance] attendance=pd.to_numeric(attendance) attendance</pre>																																																																																			
Out[107]	<pre>array([25342, 48300, 5263, 56297, 42310, 10226, 6490, 8278, 23481,        2512, 4779], dtype=int64)</pre>																																																																																			
In [108]	<pre>#get TSU rushing yards TSUurshyards=mytree.xpath('body/center//tr/td/font[&amp;color="#000000"]/text()')[114:357:23] TSUreceiveyards=[a.replace("&lt;br&gt;","") for a in TSUurshyards] TSUurshyards=pd.to_numeric(TSUurshyards) TSUurshyards</pre>																																																																																			
Out[108]	<pre>array([72, 286, 77, 26, 117, 233, 157, 87, 50, 141, -18], dtype=int64)</pre>																																																																																			
In [109]	<pre>#get TSU receiving yards TSUreceiv yards=mytree.xpath('body/center//tr/td/font[&amp;color="#000000"]/text()')[118:360:23] TSUreceiveyards=[a.replace("&lt;br&gt;","") for a in TSUreceiv yards] TSUreceiveyards=pd.to_numeric(TSUreceiveyards) TSUreceiveyards</pre>																																																																																			
Out[109]	<pre>array([154, 103, 170, 238, 115, 78, 185, 160, 196, 286, 150], dtype=int64)</pre>																																																																																			
In [110]	<pre>#get TSU kick return yards TSUkreturnyards=mytree.xpath('body/center//tr/td/font[&amp;color="#000000"]/text()')[126:360:23] TSUkreturnyards=[a.replace("&lt;br&gt;","") for a in TSUkreturnyards] TSUkreturnyards=pd.to_numeric(TSUkreturnyards) TSUkreturnyards</pre>																																																																																			
Out[110]	<pre>array([62, 11, 110, 10, 99, 71, 114, 78, 90, 128, 138], dtype=int64)</pre>																																																																																			
In [111]	<pre>#get TSU punt return yards TSUpreturmyards=mytree.xpath('body/center//tr/td/font[&amp;color="#000000"]/text()')[130:380:23] TSUpreturmyards=[a.replace("&lt;br&gt;","") for a in TSUpreturmyards] TSUpreturmyards=pd.to_numeric(TSUpreturmyards) TSUpreturmyards</pre>																																																																																			
Out[111]	<pre>array([32, 47, -2, 67, 49, 53, 14, 3, 0, 1, 0], dtype=int64)</pre>																																																																																			
In [112]	<pre>#get TSU total tackles TSUackles=mytree.xpath('body/center//tr/td/font[&amp;color="#000000"]/text()')[414:650:22] TSUackles=[a.replace("&lt;br&gt;","") for a in TSUackles] TSUackles=pd.to_numeric(TSUackles) TSUackles</pre>																																																																																			
Out[112]	<pre>array([47, 59, 95, 70, 75, 89, 62, 72, 61, 53, 64], dtype=int64)</pre>																																																																																			
In [113]	<pre>#get TSU tackle yards TSUackleyd=mytree.xpath('body/center//tr/td/font[&amp;color="#000000"]/text()')[416:650:22] TSUackleyd=[a.replace("&lt;br&gt;","") for a in TSUackleyd] TSUackleyd=pd.to_numeric(TSUackleyd) TSUackleyd</pre>																																																																																			
Out[113]	<pre>array([20, 29, 17, 56, 56, 43, 9, 20, 26, 20, 6], dtype=int64)</pre>																																																																																			
In [114]	<pre>#get TSU sacks TSUsacks=mytree.xpath('body/center//tr/td/font[&amp;color="#000000"]/text()')[417:650:22] TSUsacks=[a.replace("&lt;br&gt;","") for a in TSUsacks] TSUsacks=pd.to_numeric(TSUsacks) TSUsacks</pre>																																																																																			
Out[114]	<pre>array([2, 3, 1, 7, 6, 2, 6, 2, 1, 1, 0, 1])</pre>																																																																																			
In [115]	<pre>#get TSU sack yards TSUsackyd=mytree.xpath('body/center//tr/td/font[&amp;color="#000000"]/text()')[418:650:22] TSUsackyd=[a.replace("&lt;br&gt;","") for a in TSUsackyd] TSUsackyd=pd.to_numeric(TSUsackyd) TSUsackyd</pre>																																																																																			
Out[115]	<pre>array([13, 16, 6, 50, 37, 15, 0, 8, 14, 9, 0], dtype=int64)</pre>																																																																																			
In [116]	<pre>#TSU punts TSUpunt=mytree.xpath('body/center//tr/td/font[&amp;color="#000000"]/text()')[698:900:19] TSUpunt=[a.replace("&lt;br&gt;","") for a in TSUpunt] TSUpunt=pd.to_numeric(TSUpunt) TSUpunt</pre>																																																																																			
Out[116]	<pre>array([4, 3, 6, 6, 6, 8, 5, 7, 5, 5, 8], dtype=int64)</pre>																																																																																			
In [117]	<pre>#create data frame #change dictionary of lists to data frame list_of_dicts='date',date,                'attendance',attendance,                'TSUurshyards',TSUurshyards,                'TSUreceiveyards',TSUreceiveyards,                'TSUkreturnyards',TSUkreturnyards,                'TSUpreturmyards',TSUpreturmyards,                'TSUackles',TSUackles,                'TSUackleyd',TSUackleyd,                'TSUsacks',TSUsacks,                'TSUsackyd',TSUsackyd,                'TSUpunt',TSUpunt) df2005=pd.DataFrame(list_of_dicts) df2005.head()</pre>																																																																																			
Out[117]	<table><tr><th></th><th>date</th><th>attendance</th><th>TSUurshyards</th><th>TSUreceiveyards</th><th>TSUkreturnyards</th><th>TSUpreturmyards</th><th>TSUackles</th><th>TSUackleyd</th><th>TSUsacks</th><th>TSUsackyd</th><th>TSUpunt</th></tr><tr><td>0</td><td>2005-09-03</td><td>25342</td><td>72</td><td>154</td><td>62</td><td>32</td><td>47</td><td>20</td><td>2.0</td><td>13</td><td></td></tr><tr><td>1</td><td>2005-09-10</td><td>48300</td><td>286</td><td>103</td><td>11</td><td>47</td><td>59</td><td>29</td><td>3.0</td><td>16</td><td></td></tr><tr><td>2</td><td>2005-09-17</td><td>5263</td><td>77</td><td>170</td><td>110</td><td>-2</td><td>95</td><td>17</td><td>1.0</td><td>6</td><td></td></tr><tr><td>3</td><td>2005-09-24</td><td>56297</td><td>26</td><td>238</td><td>10</td><td>67</td><td>70</td><td>56</td><td>7.0</td><td>50</td><td></td></tr><tr><td>4</td><td>2005-10-01</td><td>42310</td><td>117</td><td>115</td><td>99</td><td>49</td><td>75</td><td>56</td><td>6.0</td><td>37</td><td></td></tr></table>													date	attendance	TSUurshyards	TSUreceiveyards	TSUkreturnyards	TSUpreturmyards	TSUackles	TSUackleyd	TSUsacks	TSUsackyd	TSUpunt	0	2005-09-03	25342	72	154	62	32	47	20	2.0	13		1	2005-09-10	48300	286	103	11	47	59	29	3.0	16		2	2005-09-17	5263	77	170	110	-2	95	17	1.0	6		3	2005-09-24	56297	26	238	10	67	70	56	7.0	50		4	2005-10-01	42310	117	115	99	49	75	56	6.0	37	
	date	attendance	TSUurshyards	TSUreceiveyards	TSUkreturnyards	TSUpreturmyards	TSUackles	TSUackleyd	TSUsacks	TSUsackyd	TSUpunt																																																																									
0	2005-09-03	25342	72	154	62	32	47	20	2.0	13																																																																										
1	2005-09-10	48300	286	103	11	47	59	29	3.0	16																																																																										
2	2005-09-17	5263	77	170	110	-2	95	17	1.0	6																																																																										
3	2005-09-24	56297	26	238	10	67	70	56	7.0	50																																																																										
4	2005-10-01	42310	117	115	99	49	75	56	6.0	37																																																																										
In [118]	<pre>#creating date column as a datetime column df2005['date']=pd.to_datetime(df2005['date'],format='%b %d %Y') #creating year column df2005['year']=df2005.date.dt.year df2005.head()</pre>																																																																																			
Out[118]	<table><tr><th></th><th>date</th><th>attendance</th><th>TSUurshyards</th><th>TSUreceiveyards</th><th>TSUkreturnyards</th><th>TSUpreturmyards</th><th>TSUackles</th><th>TSUackleyd</th><th>TSUsacks</th><th>TSUsackyd</th><th>TSUpunt</th></tr><tr><td>0</td><td>2005-09-03</td><td>25342</td><td>72</td><td>154</td><td>62</td><td>32</td><td>47</td><td>20</td><td>2.0</td><td>13</td><td></td></tr><tr><td>1</td><td>2005-09-10</td><td>48300</td><td>286</td><td>103</td><td>11</td><td>47</td><td>59</td><td>29</td><td>3.0</td><td>16</td><td></td></tr><tr><td>2</td><td>2005-09-17</td><td>5263</td><td>77</td><td>170</td><td>110</td><td>-2</td><td>95</td><td>17</td><td>1.0</td><td>6</td><td></td></tr><tr><td>3</td><td>2005-09-24</td><td>56297</td><td>26</td><td>238</td><td>10</td><td>67</td><td>70</td><td>56</td><td>7.0</td><td>50</td><td></td></tr><tr><td>4</td><td>2005-10-01</td><td>42310</td><td>117</td><td>115</td><td>99</td><td>49</td><td>75</td><td>56</td><td>6.0</td><td>37</td><td></td></tr></table>													date	attendance	TSUurshyards	TSUreceiveyards	TSUkreturnyards	TSUpreturmyards	TSUackles	TSUackleyd	TSUsacks	TSUsackyd	TSUpunt	0	2005-09-03	25342	72	154	62	32	47	20	2.0	13		1	2005-09-10	48300	286	103	11	47	59	29	3.0	16		2	2005-09-17	5263	77	170	110	-2	95	17	1.0	6		3	2005-09-24	56297	26	238	10	67	70	56	7.0	50		4	2005-10-01	42310	117	115	99	49	75	56	6.0	37	
	date	attendance	TSUurshyards	TSUreceiveyards	TSUkreturnyards	TSUpreturmyards	TSUackles	TSUackleyd	TSUsacks	TSUsackyd	TSUpunt																																																																									
0	2005-09-03	25342	72	154	62	32	47	20	2.0	13																																																																										
1	2005-09-10	48300	286	103	11	47	59	29	3.0	16																																																																										
2	2005-09-17	5263	77	170	110	-2	95	17	1.0	6																																																																										
3	2005-09-24	56297	26	238	10	67	70	56	7.0	50																																																																										
4	2005-10-01	42310	117	115	99	49	75	56	6.0	37																																																																										
In [119]	<pre>#get info about dataframe df2005.info()</pre>																																																																																			
Out[119]	<pre>&lt;class 'pandas.core.frame.DataFrame'&gt; RangeIndex: 11 entries, 0 to 10 Data columns (total 12 columns):  #   Column      Non-Null Count  Dtype ---  --  0   date        11 non-null      datetime64[ns]  1   attendance  11 non-null      int64  2   TSUurshyards  11 non-null      int64  3   TSUreceiveyards  11 non-null      int64  4   TSUkreturnyards  11 non-null      int64  5   TSUpreturmyards  11 non-null      int64  6   TSUackles     11 non-null      int64  7   TSUackleyd    11 non-null      float64  8   TSUsacks     11 non-null      float64  9   TSUsackyd    11 non-null      int64 10   TSUpunt      11 non-null      int64 11   year         11 non-null      int64 dtypes: datetime64[ns](1), float64(1), int64(10) memory usage: 1.2 KB</pre>																																																																																			
In [120]	<pre>#sort data df2005=df2005.sort_values('date',ignore_index=True) df2005.head()</pre>																																																																																			
Out[120]	<table><tr><th></th><th>date</th><th>attendance</th><th>TSUurshyards</th><th>TSUreceiveyards</th><th>TSUkreturnyards</th><th>TSUpreturmyards</th><th>TSUackles</th><th>TSUackleyd</th><th>TSUsacks</th><th>TSUsackyd</th><th>TSUpunt</th></tr><tr><td>0</td><td>2005-09-03</td><td>25342</td><td>72</td><td>154</td><td>62</td><td>32</td><td>47</td><td>20</td><td>2.0</td><td>13</td><td></td></tr><tr><td>1</td><td>2005-09-10</td><td>48300</td><td>286</td><td>103</td><td>11</td><td>47</td><td>59</td><td>29</td><td>3.0</td><td>16</td><td></td></tr><tr><td>2</td><td>2005-09-17</td><td>5263</td><td>77</td><td>170</td><td>110</td><td>-2</td><td>95</td><td>17</td><td>1.0</td><td>6</td><td></td></tr><tr><td>3</td><td>2005-09-24</td><td>56297</td><td>26</td><td>238</td><td>10</td><td>67</td><td>70</td><td>56</td><td>7.0</td><td>50</td><td></td></tr><tr><td>4</td><td>2005-10-01</td><td>42310</td><td>117</td><td>115</td><td>99</td><td>49</td><td>75</td><td>56</td><td>6.0</td><td>37</td><td></td></tr></table>													date	attendance	TSUurshyards	TSUreceiveyards	TSUkreturnyards	TSUpreturmyards	TSUackles	TSUackleyd	TSUsacks	TSUsackyd	TSUpunt	0	2005-09-03	25342	72	154	62	32	47	20	2.0	13		1	2005-09-10	48300	286	103	11	47	59	29	3.0	16		2	2005-09-17	5263	77	170	110	-2	95	17	1.0	6		3	2005-09-24	56297	26	238	10	67	70	56	7.0	50		4	2005-10-01	42310	117	115	99	49	75	56	6.0	37	
	date	attendance	TSUurshyards	TSUreceiveyards	TSUkreturnyards	TSUpreturmyards	TSUackles	TSUackleyd	TSUsacks	TSUsackyd	TSUpunt																																																																									
0	2005-09-03	25342	72	154	62	32	47	20	2.0	13																																																																										
1	2005-09-10	48300	286	103	11	47	59	29	3.0	16																																																																										
2	2005-09-17	5263	77	170	110	-2	95	17	1.0	6																																																																										
3	2005-09-24	56297	26	238	10	67	70	56	7.0	50																																																																										
4	2005-10-01	42310	117	115	99	49	75	56	6.0	37																																																																										
In [121]	<pre>#save data df2005.to_csv('2005.csv',encoding='utf-8')</pre>																																																																																			
In [122]	<pre>#2006 data #use requests.get() to get web page with data page = requests.get('https://tennstate.fsp.sidearmsports.com/custompages/tcutigers/902C2D72-B881-4838-87F2-288f0041info') #parse data on web page using html module.fromstring mytree = html.fromstring(page.content)  #go to web address above , right click on page and select inspect to get HTML code for data from right side of #create XPath query and use xpath function to get data #date data date = mytree.xpath('body/center//tr/td/font[&amp;color="#000000"]/text()')[11:10:10] date=[a.replace("&lt;br&gt;","") for a in date] date=[a.replace(",","") for a in date] date=[a.replace(".", "") for a in date] print(date)</pre>																																																																																			
In [123]	<pre>['Sep 02 2006', 'Sep 09 2006', 'Sep 16 2006', 'Sep 23 2006', 'Sep 30 2006', 'Oct 14 2006', 'Oct 21 2006', 'Oct 28 2006', 'Nov 04 2006', 'Nov 11 2006', 'Nov 18 2006']</pre>																																																																																			
In [123]	<pre>#get attendance data attendance=mytree.xpath('body/center//tr/td/font[&amp;color="#000000"]/text()')[9:110:10] attendance=[a.replace("&lt;br&gt;","") for a in attendance] attendance=[a.strip() for a in attendance] attendance=pd.to_numeric(attendance) attendance</pre>																																																																																			
Out[123]	<pre>array([19487, 10613, 53441, 27460, 57885, 9720, 11800, 18758, 5912,        4271, 5500], dtype=int64)</pre>																																																																																			
In [124]	<pre>#get TSU rushing yards TSUurshyards=mytree.xpath('body/center//tr/td/font[&amp;color="#000000"]/text()')[114:357:23] TSUurshyards=[a.replace("&lt;br&gt;","") for a in TSUurshyards] TSUurshyards=pd.to_numeric(TSUurshyards) TSUurshyards</pre>																																																																																			
Out[124]	<pre>array([156, 111, 146, 71, 227, 310, 103, 247, 49, 133, 95], dtype=int64)</pre>																																																																																			
In [125]	<pre>#get TSU receiving yards TSUreceiv yards=mytree.xpath('body/center//tr/td/font[&amp;color="#000000"]/text()')[118:360:23] TSUreceiveyards=[a.replace("&lt;br&gt;","") for a in TSUreceiv yards] TSUreceiveyards=pd.to_numeric(TSUreceiveyards) TSUreceiveyards</pre>																																																																																			
Out[125]	<pre>array([206, 230, 167, 181, 210, 169, 260, 105, 162, 152, 162], dtype=int64)</pre>																																																																																			
In [126]	<pre>#get TSU kick return yards TSUkreturnyards=mytree.xpath('body/center//tr/td/font[&amp;color="#000000"]/text()')[126:360:23] TSUkreturnyards=[a.replace("&lt;br&gt;","") for a in TSUkreturnyards] TSUkreturnyards=pd.to_numeric(TSUkreturnyards) TSUkreturnyards</pre>																																																																																			
Out[126]	<pre>array([105, 102, 97, 70, 98, 45, 119, 23, 115, 0, 126], dtype=int64)</pre>																																																																																			
In [127]	<pre>#get TSU punt return yards TSUpreturmyards=mytree.xpath('body/center//tr/td/font[&amp;color="#000000"]/text()')[130:380:23] TSUpreturmyards=[a.replace("&lt;br&gt;","") for a in TSUpreturmyards] TSUpreturmyards=pd.to_numeric(TSUpreturmyards) TSUpreturmyards</pre>																																																																																			
Out[127]	<pre>array([ 0, 0, 24, 0, 16, 4, 20, 13, 0, 17, 7], dtype=int64)</pre>																																																																																			
In [128]	<pre>#get TSU total tackles TSUackles=mytree.xpath('body/center//tr/td/font[&amp;color="#000000"]/text()')[414:650:22] TSUackles=[a.replace("&lt;br&gt;","") for a in TSUackles] TSUackles=pd.to_numeric(TSUackles) TSUackles</pre>																																																																																			
Out[128]	<pre>array([63, 76, 63, 59, 54, 64, 77, 45, 88, 44, 70], dtype=int64)</pre>																																																																																			
In [129]	<pre>#get TSU tackle yards TSUackleyd=mytree.xpath('body/center//tr/td/font[&amp;color="#000000"]/text()')[416:650:22] TSUackleyd=[a.replace("&lt;br&gt;","") for a in TSUackleyd] TSUackleyd=pd.to_numeric(TSUackleyd) TSUackleyd</pre>																																																																																			
Out[129]	<pre>array([3, 38, 18, 25, 14, 15, 26, 28, 22, 27, 30], dtype=int64)</pre>																																																																																			
In [130]	<pre>#get TSU sacks TSUsacks=mytree.xpath('body/center//tr/td/font[&amp;color="#000000"]/text()')[417:650:22] TSUsacks=[a.replace("&lt;br&gt;","") for a in TSUsacks] TSUsacks=pd.to_numeric(TSUsacks) TSUsacks</pre>																																																																																			
Out[130]	<pre>array([0, 3, 1, 1, 1, 1, 1, 3, 3, 1, 4, 2, 1])</pre>																																																																																			
In [131]	<pre>#get TSU sack yards TSUsackyd=mytree.xpath('body/center//tr/td/font[&amp;color="#000000"]/text()')[418:650:22] TSUsackyd=[a.replace("&lt;br&gt;","") for a in TSUsackyd] TSUsackyd=pd.to_numeric(TSUsackyd) TSUsackyd</pre>																																																																																			
Out[131]	<pre>array([0, 23, 6, 7, 4, 8, 20, 16, 10, 23, 14], dtype=int64)</pre>																																																																																			
In [132]	<pre>#TSU punts TSUpunt=mytree.xpath('body/center//tr/td/font[&amp;color="#000000']/text()')[698:900:19] TSUpunt=[a.replace("&lt;br&gt;","") for a in TSUpunt] TSUpunt=pd.to_numeric(TSUpunt) TSUpunt</pre>																																																																																			
Out[132]	<pre>array([2, 5, 2, 3, 4, 4, 6, 3, 6, 4, 1], dtype=int64)</pre>																																																																																			
In [133]	<pre>#create data frame #change dictionary of lists to data frame list_of_dicts='date',date,                'attendance',attendance,                'TSUurshyards',TSUurshyards,                'TSUreceiveyards',TSUreceiveyards,                'TSUkreturnyards',TSUkreturnyards,                'TSUpreturmyards',TSUpreturmyards,                'TSUackles',TSUackles,                'TSUackleyd',TSUackleyd,                'TSUsacks',TSUsacks,                'TSUsackyd',TSUsackyd,                'TSUpunt',TSUpunt) df2006=pd.DataFrame(list_of_dicts) df2006.head()</pre>																																																																																			
Out[133]	<table><tr><th></th><th>date</th><th>attendance</th><th>TSUurshyards</th><th>TSUreceiveyards</th><th>TSUkreturnyards</th><th>TSUpreturmyards</th><th>TSUackles</th><th>TSUackleyd</th><th>TSUsacks</th><th>TSUsackyd</th><th>TSUpunt</th></tr><tr><td>0</td><td>2006-09-02</td><td>19487</td><td>156</td><td>206</td><td>105</td><td>0</td><td>63</td><td>3</td><td>0.0</td><td>0</td><td></td></tr><tr><td>1</td><td>2006-09-09</td><td>10613</td><td>111</td><td>230</td><td>102</td><td>0</td><td>76</td><td>38</td><td>3.0</td><td>23</td><td></td></tr><tr><td>2</td><td>2006-09-16</td><td>53441</td><td>146</td><td>167</td><td>77</td><td>24</td><td>63</td><td>18</td><td>1.0</td><td>6</td><td></td></tr><tr><td>3</td><td>2006-09-23</td><td>27460</td><td>71</td><td>181</td><td>70</td><td>0</td><td>59</td><td>25</td><td>1.0</td><td>7</td><td></td></tr><tr><td>4</td><td>2006-09-30</td><td>57885</td><td>227</td><td>210</td><td>98</td><td>16</td><td>54</td><td>14</td><td>1.0</td><td>4</td><td></td></tr></table>													date	attendance	TSUurshyards	TSUreceiveyards	TSUkreturnyards	TSUpreturmyards	TSUackles	TSUackleyd	TSUsacks	TSUsackyd	TSUpunt	0	2006-09-02	19487	156	206	105	0	63	3	0.0	0		1	2006-09-09	10613	111	230	102	0	76	38	3.0	23		2	2006-09-16	53441	146	167	77	24	63	18	1.0	6		3	2006-09-23	27460	71	181	70	0	59	25	1.0	7		4	2006-09-30	57885	227	210	98	16	54	14	1.0	4	
	date	attendance	TSUurshyards	TSUreceiveyards	TSUkreturnyards	TSUpreturmyards	TSUackles	TSUackleyd	TSUsacks	TSUsackyd	TSUpunt																																																																									
0	2006-09-02	19487	156	206	105	0	63	3	0.0	0																																																																										
1	2006-09-09	10613	111	230	102	0	76	38	3.0	23																																																																										
2	2006-09-16	53441	146	167	77	24	63	18	1.0	6																																																																										
3	2006-09-23	27460	71	181	70	0	59	25	1.0	7																																																																										
4	2006-09-30	57885	227	210	98	16	54	14	1.0	4																																																																										
In [134]	<pre>#creating date column as a datetime column df2006['date']=pd.to_datetime(df2006['date'],format='%b %d %Y') #creating year column df2006['year']=df2006.date.dt.year df2006.head()</pre>																																																																																			
Out[134]	<table><tr><th></th><th>date</th><th>attendance</th><th>TSUurshyards</th><th>TSUreceiveyards</th><th>TSUkreturnyards</th><th>TSUpreturmyards</th><th>TSUackles</th><th>TSUackleyd</th><th>TSUsacks</th><th>TSUsackyd</th><th>TSUpunt</th></tr><tr><td>0</td><td>2006-09-02</td><td>19487</td><td>156</td><td>206</td><td>105</td><td>0</td><td>63</td><td>3</td><td>0.0</td><td>0</td><td></td></tr><tr><td>1</td><td>2006-09-09</td><td>10613</td><td>111</td><td>230</td><td>102</td><td>0</td><td>76</td><td>38</td><td>3.0</td><td>23</td><td></td></tr><tr><td>2</td><td>2006-09-16</td><td>53441</td><td>146</td><td>167</td><td>77</td><td>24</td><td>63</td><td>18</td><td>1.0</td><td>6</td><td></td></tr><tr><td>3</td><td>2006-09-23</td><td>27460</td><td>71</td><td>181</td><td>70</td><td>0</td><td>59</td><td>25</td><td>1.0</td><td>7</td><td></td></tr><tr><td>4</td><td>2006-09-30</td><td>57885</td><td>227</td><td>210</td><td>98</td><td>16</td><td>54</td><td>14</td><td>1.0</td><td>4</td><td></td></tr></table>													date	attendance	TSUurshyards	TSUreceiveyards	TSUkreturnyards	TSUpreturmyards	TSUackles	TSUackleyd	TSUsacks	TSUsackyd	TSUpunt	0	2006-09-02	19487	156	206	105	0	63	3	0.0	0		1	2006-09-09	10613	111	230	102	0	76	38	3.0	23		2	2006-09-16	53441	146	167	77	24	63	18	1.0	6		3	2006-09-23	27460	71	181	70	0	59	25	1.0	7		4	2006-09-30	57885	227	210	98	16	54	14	1.0	4	
	date	attendance	TSUurshyards	TSUreceiveyards	TSUkreturnyards	TSUpreturmyards	TSUackles	TSUackleyd	TSUsacks	TSUsackyd	TSUpunt																																																																									
0	2006-09-02	19487	156	206	105	0	63	3	0.0	0																																																																										
1	2006-09-09	10613	111	230	102	0	76	38	3.0	23																																																																										
2	2006-09-16	53441	146	167	77	24	63	18	1.0	6																																																																										
3	2006-09-23	27460	71	181	70	0	59	25	1.0	7																																																																										
4	2006-09-30	57885	227	210	98	16	54	14	1.0	4																																																																										
In [135]	<pre>#get info on data frame df2006.info()</pre>																																																																																			
Out[135]	<pre>&lt;class 'pandas.core.frame.DataFrame'&gt; RangeIndex: 11 entries, 0 to 10 Data columns (total 12 columns):  #   Column      Non-Null Count  Dtype ---  --  0   date        11 non-null      datetime64[ns]  1   attendance  11 non-null      int64  2   TSUurshyards  11 non-null      int64  3   TSUreceiveyards  11 non-null      int64  4   TSUkreturnyards  11 non-null      int64  5   TSUpreturmyards  11 non-null      int64  6   TSUackles     11 non-null      int64  7   TSUackleyd    11 non-null      float64  8   TSUsacks     11 non-null      float64  9   TSUsackyd    11 non-null      int64 10   year         11 non-null      int64 11   year         11 non-null      int64 dtypes: datetime64[ns](1), float64(1), int64(10) memory usage: 1.2 KB</pre>																																																																																			
In [136]	<pre>#sort data df2006=df2006.sort_values('date',ignore_index=True) df2006.head()</pre>																																																																																			
Out[136]	<table><tr><th></th><th>date</th><th>attendance</th><th>TSUurshyards</th><th>TSUreceiveyards</th><th>TSUkreturnyards</th><th>TSUpreturmyards</th><th>TSUackles</th><th>TSUackleyd</th><th>TSUsacks</th><th>TSUsackyd</th><th>TSUpunt</th></tr><tr><td>0</td><td>2006-09-02</td><td>19487</td><td>156</td><td>206</td><td>105</td><td>0</td><td>63</td><td>3</td><td>0.0</td><td>0</td><td></td></tr><tr><td>1</td><td>2006-09-09</td><td>10613</td><td>111</td><td>230</td><td>102</td><td>0</td><td>76</td><td>38</td><td>3.0</td><td>23</td><td></td></tr><tr><td>2</td><td>2006-09-16</td><td>53441</td><td>146</td><td>167</td><td>77</td><td>24</td><td>63</td><td>18</td><td>1.0</td><td>6</td><td></td></tr><tr><td>3</td><td>2006-09-23</td><td>27460</td><td>71</td><td>181</td><td>70</td><td>0</td><td>59</td><td>25</td><td>1.0</td><td>7</td><td></td></tr><tr><td>4</td><td>2006-09-30</td><td>57885</td><td>227</td><td>210</td><td>98</td><td>16</td><td>54</td><td>14</td><td>1.0</td><td>4</td><td></td></tr></table>													date	attendance	TSUurshyards	TSUreceiveyards	TSUkreturnyards	TSUpreturmyards	TSUackles	TSUackleyd	TSUsacks	TSUsackyd	TSUpunt	0	2006-09-02	19487	156	206	105	0	63	3	0.0	0		1	2006-09-09	10613	111	230	102	0	76	38	3.0	23		2	2006-09-16	53441	146	167	77	24	63	18	1.0	6		3	2006-09-23	27460	71	181	70	0	59	25	1.0	7		4	2006-09-30	57885	227	210	98	16	54	14	1.0	4	
	date	attendance	TSUurshyards	TSUreceiveyards	TSUkreturnyards	TSUpreturmyards	TSUackles	TSUackleyd	TSUsacks	TSUsackyd	TSUpunt																																																																									
0	2006-09-02	19487	156	206	105	0	63	3	0.0	0																																																																										
1	2006-09-09	10613	111	230	102	0	76	38	3.0	23																																																																										
2	2006-09-16	53441	146	167	77	24	63	18	1.0	6																																																																										
3	2006-09-23	27460	71	181	70	0	59	25	1.0	7																																																																										
4	2006-09-30	57885	227	210	98	16	54	14	1.0	4																																																																										
In [137]	<pre>#save data df2006.to_csv('2006.csv',encoding='utf-8')</pre>																																																																																			
In [138]	<pre>#2007 data #use requests.get() to get web page with data page = requests.get('https://tennstate.fsp.sidearmsports.com/custompages/tcutigers/4C064A83-37A0-4B85-8B11-B78f0041info') #parse data on web page using html module.fromstring mytree = html.fromstring(page.content)  #go to web address above , right click on page and select inspect to get HTML code for data from right side of #create XPath query and use xpath function to get data #date data date = mytree.xpath('body/center//tr/td/font[&amp;color="#000000']/text()')[11:1')[</pre>																																																																																			



<



	array([15, 6, 3, 7, 4, 5, 1, 5, 4, 6, 5], dtype=int64)										
	<pre>#create data frame #change dictionary of lists to data frame list_of_dicts={'date':date,                 'attendance':attendance,                 'TSUrushyards':TSUrushyards,                 'TSUreceiveways':TSUreceiveways,                 'TSUkretunyards':TSUkretunyards,                 'TSUpretunyards':TSUpretunyards,                 'TSUacks':TSUacks,                 'TSUackleyd':TSUackleyd,                 'TSUsacks':TSUsacks,                 'TSUsackyd':TSUsackyd,                 'TSUpunt':TSUpunt) df2012=mpd.DataFrame(list_of_dicts) df2012.head()</pre>										
Out[229]	<b>date</b>	<b>attendance</b>	<b>TSUrushyards</b>	<b>TSUreceiveways</b>	<b>TSUkretunyards</b>	<b>TSUpretunyards</b>	<b>TSUacks</b>	<b>TSUackleyd</b>	<b>TSUsacks</b>	<b>TSUsackyd</b>	<b>TSUpunt</b>
	Sep 01	15652	138	263	52	8	49	28	1	5	
	Sep 02	42257	235	137	69	42	66	21	3	20	
	Sep 03	14264	112	322	24	7	55	13	1	9	
	Sep 04	9461	200	157	41	37	84	52	4	30	
	Sep 05	31765	201	262	47	4	62	40	3	29	
In [230]	<pre>#creating date column as a datetime column df2012['date']=mpd.to_datetime(df2012['date'],format='%b %d %Y') #creating year column df2012['year']=df2012.date.dt.year df2012.head()</pre>										
Out[230]	<b>date</b>	<b>attendance</b>	<b>TSUrushyards</b>	<b>TSUreceiveways</b>	<b>TSUkretunyards</b>	<b>TSUpretunyards</b>	<b>TSUacks</b>	<b>TSUackleyd</b>	<b>TSUsacks</b>	<b>TSUsackyd</b>	<b>TSUpunt</b>
	Sep 01-01	15652	138	263	52	8	49	28	1	5	
	Sep 02-02	42257	235	137	69	42	66	21	3	20	
	Sep 03-03	14264	112	322	24	7	55	13	1	9	
	Sep 04-04	9461	200	157	41	37	84	52	4	30	
	Sep 05-05	31765	201	262	47	4	62	40	3	29	
In [231]	<pre>#get info for data frame df2012.info() &lt;class 'pandas.core.frame.DataFrame'&gt; RangeIndex: 11 entries, 0 to 10 Data columns (total 12 columns):  # Column Non-Null Count Dtype ---  --  0 date 11 non-null datetime64[ns]  1 attendance 11 non-null int64  2 TSUrushyards 11 non-null int64  3 TSUreceiveways 11 non-null int64  4 TSUkretunyards 11 non-null int64  5 TSUpretunyards 11 non-null int64  6 TSUacks 11 non-null int64  7 TSUackleyd 11 non-null int64  8 TSUsacks 11 non-null int64  9 TSUsackyd 11 non-null int64 10 TSUpunt 11 non-null int64 dtypes: datetime64[ns](1), int64(11) memory usage: 1.2 KB</pre>										
In [232]	<pre>#sort data df2012=df2012.sort_values('date',ignore_index=True) df2012.head()</pre>										
Out[232]	<b>date</b>	<b>attendance</b>	<b>TSUrushyards</b>	<b>TSUreceiveways</b>	<b>TSUkretunyards</b>	<b>TSUpretunyards</b>	<b>TSUacks</b>	<b>TSUackleyd</b>	<b>TSUsacks</b>	<b>TSUsackyd</b>	<b>TSUpunt</b>
	Sep 01-01	15652	138	263	52	8	49	28	1	5	
	Sep 02-02	42257	235	137	69	42	66	21	3	20	
	Sep 03-03	14264	112	322	24	7	55	13	1	9	
	Sep 04-04	9461	200	157	41	37	84	52	4	30	
	Sep 05-05	31765	201	262	47	4	62	40	3	29	
In [233]	<pre>#get attendance data attendance=mytree.xpath('//tbody/center//tr/td/font[@color="#000000']/text()')[9:140:10] attendance=[a.replace("xao","") for a in attendance] attendance=[a.strip() for a in attendance] attendance=pd.to_numeric(attendance) attendance</pre>										
Out[233]	array([16108, 14237, 42400, 10044, 22000, 7374, 19092, 4166, 22157, 5700, 6412, 5286, 1928, 4825], dtype=int64)										
In [236]	<pre>#get TSU rushing yards TSUrushyards=mytree.xpath('//tbody/center//tr/td/font[@color="#000000']/text()')[144:446:23] TSUrushyards=[a.replace("xao","") for a in TSUrushyards] TSUrushyards=pd.to_numeric(TSUrushyards) TSUrushyards</pre>										
Out[236]	array([116, 268, 174, 95, 311, 264, 215, 69, 126, 71, 241, 146, 152, 440], dtype=int64)										
In [237]	<pre>#get TSU receiving yards TSUreceiveways=mytree.xpath('//tbody/center//tr/td/font[@color="#000000']/text()')[148:466:23] TSUreceiveways=[a.replace("xao","") for a in TSUreceiveways] TSUreceiveways=pd.to_numeric(TSUreceiveways) TSUreceiveways</pre>										
Out[237]	array([132, 131, 111, 343, 228, 280, 133, 170, 212, 101, 60, 173, 263, 242], dtype=int64)										
In [238]	<pre>#get TSU kick return yards TSUkretunyards=mytree.xpath('//tbody/center//tr/td/font[@color="#000000']/text()')[156:460:23] TSUkretunyards=[a.replace("xao","") for a in TSUkretunyards] TSUkretunyards=pd.to_numeric(TSUkretunyards) TSUkretunyards</pre>										
Out[238]	array([84, 111, 64, 63, 46, 24, 48, 64, 19, 96, 45, 8, -9], dtype=int64)										
In [239]	<pre>#get TSU punt return yards TSUpretunyards=mytree.xpath('//tbody/center//tr/td/font[@color="#000000']/text()')[160:446:23] TSUpretunyards=[a.replace("xao","") for a in TSUpretunyards] TSUpretunyards=pd.to_numeric(TSUpretunyards) TSUpretunyards</pre>										
Out[239]	array([51, 17, 11, 23, 75, 7, 46, 0, 56, -1, 19, 13, -3, 21], dtype=int64)										
In [240]	<pre>#get TSU total tackles TSUacks=mytree.xpath('//tbody/center//tr/td/font[@color="#000000']/text()')[513:800:22] TSUacks=[a.replace("xao","") for a in TSUacks] TSUacks=pd.to_numeric(TSUacks) TSUacks</pre>										
Out[240]	array([62, 41, 69, 76, 70, 37, 70, 78, 70, 63, 54, 60, 47, 73], dtype=int64)										
In [241]	<pre>#get TSU tackle yards TSUackleyd=mytree.xpath('//tbody/center//tr/td/font[@color="#000000']/text()')[515:822:22] TSUackleyd=[a.replace("xao","") for a in TSUackleyd] TSUackleyd=pd.to_numeric(TSUackleyd) TSUackleyd</pre>										
Out[241]	array([7, 35, 26, 51, 44, 12, 61, 56, 31, 3, 31, 59, 21, 13], dtype=int64)										
In [242]	<pre>#get TSU sacks TSUsacks=mytree.xpath('//tbody/center//tr/td/font[@color="#000000']/text()')[516:822:22] TSUsacks=[a.replace("xao","") for a in TSUsacks] TSUsacks=pd.to_numeric(TSUsacks) TSUsacks</pre>										
Out[242]	array([1, 3, 2, 3, 3, 3, 1, 5, 5, 3, 0, 3, 0, 5, 2, 1, 1])										
In [243]	<pre>#get TSU sack yards TSUsackyd=mytree.xpath('//tbody/center//tr/td/font[@color="#000000']/text()')[517:822:22] TSUsackyd=[a.replace("xao","") for a in TSUsackyd] TSUsackyd=pd.to_numeric(TSUsackyd) TSUsackyd</pre>										
Out[243]	array([4, 18, 13, 28, 17, 8, 18, 35, 27, 0, 20, 37, 7, 5], dtype=int64)										
In [244]	<pre>#TSU punts TSUpunt=mytree.xpath('//tbody/center//tr/td/font[@color="#000000']/text()')[863:1119:19] TSUpunt=[a.replace("xao","") for a in TSUpunt] TSUpunt=pd.to_numeric(TSUpunt) TSUpunt</pre>										
Out[244]	array([5, 4, 3, 5, 1, 3, 6, 6, 8, 9, 8, 5, 4, 8], dtype=int64)										
In [245]	<pre>#create data frame #change dictionary of lists to data frame list_of_dicts={'date':date,                 'attendance':attendance,                 'TSUrushyards':TSUrushyards,                 'TSUreceiveways':TSUreceiveways,                 'TSUkretunyards':TSUkretunyards,                 'TSUpretunyards':TSUpretunyards,                 'TSUacks':TSUacks,                 'TSUackleyd':TSUackleyd,                 'TSUsacks':TSUsacks,                 'TSUsackyd':TSUsackyd,                 'TSUpunt':TSUpunt) df2013=mpd.DataFrame(list_of_dicts) df2013.head()</pre>										
Out[245]	<b>date</b>	<b>attendance</b>	<b>TSUrushyards</b>	<b>TSUreceiveways</b>	<b>TSUkretunyards</b>	<b>TSUpretunyards</b>	<b>TSUacks</b>	<b>TSUackleyd</b>	<b>TSUsacks</b>	<b>TSUsackyd</b>	<b>TSUpunt</b>
	Sep 01	16108	116	132	84	51	62	7	1.0	4	
	Sep 07	14237	268	131	111	17	41	35	3.0	18	
	Sep 08	42400	174	111	64	11	69	26	2.0	13	
	Sep 21	10044	95	343	63	23	76	51	3.0	28	
	Sep 23	22000	311	228	46	75	70	44	3.0	17	
In [246]	<pre>#creating date column as a datetime column df2013['date']=mpd.to_datetime(df2013['date'],format='%b %d %Y') #creating year column df2013['year']=df2013.date.dt.year df2013.head()</pre>										
Out[246]	<b>date</b>	<b>attendance</b>	<b>TSUrushyards</b>	<b>TSUreceiveways</b>	<b>TSUkretunyards</b>	<b>TSUpretunyards</b>	<b>TSUacks</b>	<b>TSUackleyd</b>	<b>TSUsacks</b>	<b>TSUsackyd</b>	<b>TSUpunt</b>
	Sep 01-01	16108	116	132	84	51	62	7	1.0	4	
	Sep 07-07	14237	268	131	111	17	41	35	3.0	18	
	Sep 08-08	42400	174	111	64	11	69	26	2.0	13	
	Sep 21-21	10044	95	343	63	23	76	51	3.0	28	
	Sep 23-23	22000	311	228	46	75	70	44	3.0	17	
In [247]	<pre>#get info on data frame df2013.info() &lt;class 'pandas.core.frame.DataFrame'&gt; RangeIndex: 14 entries, 0 to 13 Data columns (total 12 columns):  # Column Non-Null Count Dtype ---  --  0 date 14 non-null datetime64[ns]  1 attendance 14 non-null int64  2 TSUrushyards 14 non-null int64  3 TSUreceiveways 14 non-null int64  4 TSUkretunyards 14 non-null int64  5 TSUpretunyards 14 non-null int64  6 TSUacks 14 non-null int64  7 TSUackleyd 14 non-null int64  8 TSUsacks 14 non-null float64  9 TSUsackyd 14 non-null int64 10 TSUpunt 14 non-null int64 11 year 14 non-null int64 dtypes: datetime64[ns](1), float64(1), int64(10) memory usage: 1.4 KB</pre>										
In [248]	<pre>#sort data df2013=df2013.sort_values('date',ignore_index=True) df2013.head()</pre>										
Out[248]	<b>date</b>	<b>attendance</b>	<b>TSUrushyards</b>	<b>TSUreceiveways</b>	<b>TSUkretunyards</b>	<b>TSUpretunyards</b>	<b>TSUacks</b>	<b>TSUackleyd</b>	<b>TSUsacks</b>	<b>TSUsackyd</b>	<b>TSUpunt</b>
	Sep 01-01	16108	116	132	84	51	62	7	1.0	4	
	Sep 07-07	14237	268	131	111	17	41	35	3.0	18	
	Sep 08-08	42400	174	111	64	11	69	26	2.0	13	
	Sep 21-21	10044	95	343	63	23	76	51	3.0	28	
	Sep 23-23	22000	311	228	46	75	70	44	3.0	17	
In [249]	<pre>#save data df2013.to_csv('2013.csv',encoding='utf-8')</pre>										
In [250]	<pre>#2014 data #use requests.get() to get web page with data page = requests.get('https://tennstate.fcp.sidearmsports.com/custompages/tautigers/26045A82-D036-475D-A8EF-3881') #parse data on web page using html module.fronstring mytree = html.fronstring(page.content)  #go to web address above , right click on page and select inspect to get HTML code for data from right side of #create XPath query and use xpath function to get data #date data date = mytree.xpath('//tbody/center//tr/td/font[@color="#000000']/text()')[1:120:10] date=[a.replace("xao","") for a in date] date=[a.strip() for a in date] date=[a.replace("10-04-14", "Oct 04 2014") for a in date] print(date)  ['Aug 10 2014', 'Sep 06 2014', 'Sep 13 2014', 'Sep 20 2014', 'Sep 27 2014', 'Oct 04 2014', 'Oct 11 2014', 'Oct 18 2014', 'Oct 25 2014', 'Nov 01 2014', 'Nov 08 2014', 'Nov 22 2014']</pre>										
In [251]	<pre>#get attendance data attendance=mytree.xpath('//tbody/center//tr/td/font[@color="#000000']/text()')[9:120:10] attendance=[a.replace("xao","") for a in attendance] attendance=[a.strip() for a in attendance] attendance=pd.to_numeric(attendance) attendance</pre>										
Out[251]	array([170543, 15725, 46914, 9217, 29225, 8089, 5845, 6738, 8289, 5052, 6143, 1962], dtype=int64)										
In [252]	<pre>#get TSU rushing yards TSUrushyards=mytree.xpath('//tbody/center//tr/td/font[@color="#000000']/text()')[124:400:23] TSUrushyards=[a.replace("xao","") for a in TSUrushyards] TSUrushyards=pd.to_numeric(TSUrushyards) TSUrushyards</pre>										
Out[252]	array([439, 92, 137, 92, 125, 143, 43, 53, 107, 105, 111, 114], dtype=int64)										
In [253]	<pre>#get TSU receiving yards TSUreceiveways=mytree.xpath('//tbody/center//tr/td/font[@color="#000000']/text()')[128:400:23] TSUreceiveways=[a.replace("xao","") for a in TSUreceiveways] TSUreceiveways=pd.to_numeric(TSUreceiveways) TSUreceiveways</pre>										
Out[253]	array([71, 340, 187, 113, 153, 360, 364, 314, 180, 326, 395, 463], dtype=int64)										
In [254]	<pre>#get TSU kick return yards TSUkretunyards=mytree.xpath('//tbody/center//tr/td/font[@color="#000000']/text()')[136:400:23] TSUkretunyards=[a.replace("xao","") for a in TSUkretunyards] TSUkretunyards=pd.to_numeric(TSUkretunyards) TSUkretunyards</pre>										
Out[254]	array([77, 111, 29, 0, 24, 53, 63, 32, 136, 198, 58, 48], dtype=int64)										
In [255]	<pre>#get TSU punt return yards TSUpretunyards=mytree.xpath('//tbody/center//tr/td/font[@color="#000000']/text()')[140:400:23] TSUpretunyards=[a.replace("xao","") for a in TSUpretunyards] TSUpretunyards=pd.to_numeric(TSUpretunyards) TSUpretunyards</pre>										
Out[255]	array([17, 22, 40, 15, 56, 24, 12, -1, 1, 0, -2, 0], dtype=int64)										
In [256]	<pre>#get TSU total tackles TSUacks=mytree.xpath('//tbody/center//tr/td/font[@color="#000000']/text()')[147:723:23] TSUacks=[a.replace("xao","") for a in TSUacks] TSUacks=pd.to_numeric(TSUacks) TSUacks</pre>										
Out[256]	array([53, 89, 54, 48, 64, 81, 15, 73, 54, 84, 85, 81], dtype=int64)										
In [257]	<pre>#get TSU tackle yards TSUackleyd=mytree.xpath('//tbody/center//tr/td/font[@color="#000000']/text()')[149:711:23] TSUackleyd=[a.replace("xao","") for a in TSUackleyd] TSUackleyd=pd.to_numeric(TSUackleyd) TSUackleyd</pre>										
Out[257]	array([38, 28, 40, 31, 51, 9, 21, 48, 17, 19, 29, 63], dtype=int64)										
In [257]	<pre>#get TSU sacks TSUsacks=mytree.xpath('//tbody/center//tr/td/font[@color="#000000']/text()')[1</pre>										



```
#for TSU sacks
TSUsacks=mytree.xpath('//body/center//tr/cd/font[&color="#000000"]//text()') [417:660:23]
TSUsacks=[a.replace("xao","") for a in TSUsacks]
TSUsacks=pd.to_numeric(TSUsacks)
TSUsacks

Out[298]:
array([5., 3., 1., 2., 1., 0., 1., 2., 2., 1., 3.])

In [291]:
#get TSU sack yards
TSUsackyd=mytree.xpath('//body/center//tr/cd/font[&color="#000000"]//text()') [418:660:23]
TSUsackyd=[a.replace("xao","") for a in TSUsackyd]
TSUsackyd=pd.to_numeric(TSUsackyd)
TSUsackyd

Out[291]:
array([31, 21, 1, 12, 3, 0, 5, 14, 13, 0, 26], dtype=int64)

In [292]:
#TSU punts
TSUpunt=mytree.xpath('//body/center//tr/cd/font[&color="#000000"]//text()') [711:919:19]
TSUpunt=[a.replace("xao","") for a in TSUpunt]
TSUpunt=pd.to_numeric(TSUpunt)
TSUpunt

Out[292]:
array([3, 6, 5, 1, 5, 3, 3, 2, 5, 3, 3], dtype=int64)

In [293]:
#create data frame
#change dictionary of lists to data frame
list_of_dicts={'date':date,
               'attendance':attendance,
               'TSUruhsyards':TSUruhsyards,
               'TSUreceiveyards':TSUreceiveyards,
               'TSUkreturnyards':TSUkreturnyards,
               'TSUpreturnyards':TSUpreturnyards,
               'TSUackles':TSUackles,
               'TSUacks':TSUacks,
               'TSUackleyd':TSUackleyd,
               'TSUpunt':TSUpunt}
df2016=pd.DataFrame(list_of_dicts)
df2016.head()
```

	date	attendance	TSUruhsyards	TSUreceiveyards	TSUkreturnyards	TSUpreturnyards	TSUackles	TSUackleyd	TSUacks	TSUackyd	TSUp
0	Sep 2016	15078	202	259	29	64	52	39	5.0	31	
1	Sep 2016	46263	121	273	96	11	61	40	3.0	21	
2	Sep 2016	9385	210	184	122	23	52	17	1.0	1	
3	Oct 2016	10001	141	223	80	0	63	28	2.0	12	
4	Oct 2016	4319	76	303	113	0	75	6	1.0	3	

```
df2016['date'].mpd.to_datetime(df2016['date'],format='%b %d %Y')
#creating year column
df2016['year']=df2016.date.dt.year
df2016.head()
```

	date	attendance	TSUruhsyards	TSUreceiveyards	TSUkreturnyards	TSUpreturnyards	TSUackles	TSUackleyd	TSUacks	TSUackyd	TSUp
0	2016-09-03	15078	202	259	29	64	52	39	5.0	31	
1	2016-09-10	46263	121	273	96	11	61	40	3.0	21	
2	2016-09-17	9385	210	184	122	23	52	17	1.0	1	
3	2016-10-01	10001	141	223	80	0	63	28	2.0	12	
4	2016-10-08	4319	76	303	113	0	75	6	1.0	3	

```
#get info on data frame
df2016.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 11 entries, 0 to 10
Data columns (total 12 columns):
# Column Non-Null Count Dtype
---
0 date 11 non-null datetime64[ns]
1 attendance 11 non-null int64
2 TSUruhsyards 11 non-null int64
3 TSUreceiveyards 11 non-null int64
4 TSUkreturnyards 11 non-null int64
5 TSUpreturnyards 11 non-null int64
6 TSUackles 11 non-null int64
7 TSUackleyd 11 non-null float64
8 TSUacks 11 non-null float64
9 TSUackyd 11 non-null int64
10 TSUpunt 11 non-null int64
11 year 11 non-null int64
dtypes: datetime64[ns](1), float64(1), int64(10)
memory usage: 1.2 KB

In [296]:
#sort data
df2016=df2016.sort_values('date',ignore_index=True)
df2016.head()
```

	date	attendance	TSUruhsyards	TSUreceiveyards	TSUkreturnyards	TSUpreturnyards	TSUackles	TSUackleyd	TSUacks	TSUackyd	TSUp
0	2016-09-03	15078	202	259	29	64	52	39	5.0	31	
1	2016-09-10	46263	121	273	96	11	61	40	3.0	21	
2	2016-09-17	9385	210	184	122	23	52	17	1.0	1	
3	2016-10-01	10001	141	223	80	0	63	28	2.0	12	
4	2016-10-08	4319	76	303	113	0	75	6	1.0	3	

```
#save data
df2016.to_csv('2016.csv',encoding='utf-8')
```

```
#2017 data
#use requests.get() to get web page with data
page = requests.get('https://tennstate.fip.sidaexsports.com/customreports/tautgers/125803E8-418B-422C-980C-6451')
#parse data on web page using html module.fromstring
mytree = html.fromstring(page.content)

#go to web address above , right click on page and select inspect to get HTML code for data from right side of
#create XPath query and use xpath function to get data
#date data
date = mytree.xpath('//body/center//tr/cd/font[&color="#000000"]//text()') [1:110:10]
date=[a.replace("xao","") for a in date]
date=[a.replace(" ",",") for a in date]
date=[a.strip() for a in date]
print(date)

['Aug 31 2017', 'Sep 09 2017', 'Sep 17 2017', 'Sep 23 2017', 'Sep 30 2017', 'Oct 07 2017', 'Oct 14 2017', 'Oct
28 2017', 'Nov 04 2017', 'Nov 11 2017', 'Nov 16 2017']

In [299]:
#get attendance data
attendance=mytree.xpath('//body/center//tr/cd/font[&color="#000000"]//text()') [9:110:10]
attendance=[a.replace("xao","") for a in attendance]
attendance=[a.strip() for a in attendance]
attendance=pd.to_numeric(attendance)
attendance

Out[299]:
array([24333, 47407, 17102, 6484, 11013, 8410, 21127, 5235, 7487,
       8693, 18782], dtype=int64)
```

```
#get TSU rushing yards
TSUruhsyards=mytree.xpath('//body/center//tr/cd/font[&color="#000000"]//text()') [114:346:23]
TSUruhsyards=[a.replace("xao","") for a in TSUruhsyards]
TSUruhsyards=pd.to_numeric(TSUruhsyards)
TSUruhsyards

Out[300]:
array([238, 160, 241, 200, 83, 193, 106, 83, 174, 147, 15], dtype=int64)
```

```
#get TSU receiving yards
TSUreceiveyards=mytree.xpath('//body/center//tr/cd/font[&color="#000000"]//text()') [118:369:23]
TSUreceiveyards=[a.replace("xao","") for a in TSUreceiveyards]
TSUreceiveyards=pd.to_numeric(TSUreceiveyards)
TSUreceiveyards

Out[301]:
array([145, 78, 273, 195, 106, 208, 331, 274, 196, 82], dtype=int64)
```

```
#get TSU kick return yards
TSUkreturnyards=mytree.xpath('//body/center//tr/cd/font[&color="#000000"]//text()') [126:363:23]
TSUkreturnyards=[a.replace("xao","") for a in TSUkreturnyards]
TSUkreturnyards=pd.to_numeric(TSUkreturnyards)
TSUkreturnyards

Out[302]:
array([ 44, 49, 63, 160, 48, 25, 64, 133, 18, 103, 140], dtype=int64)
```

```
#get TSU punt return yards
TSUpreturnyards=mytree.xpath('//body/center//tr/cd/font[&color="#000000"]//text()') [130:363:23]
TSUpreturnyards=[a.replace("xao","") for a in TSUpreturnyards]
TSUpreturnyards=pd.to_numeric(TSUpreturnyards)
TSUpreturnyards

Out[303]:
array([ 0, 41, 7, 3, 0, -3, 13, -2, 94, 23, 1], dtype=int64)
```

```
#get TSU total tackles
TSUackles=mytree.xpath('//body/center//tr/cd/font[&color="#000000"]//text()') [418:646:23]
TSUackles=[a.replace("xao","") for a in TSUackles]
TSUackles=pd.to_numeric(TSUackles)
TSUackles

Out[304]:
array([59, 76, 38, 77, 73, 60, 61, 62, 41, 57, 71], dtype=int64)
```

```
#get TSU tackle yards
TSUackleyd=mytree.xpath('//body/center//tr/cd/font[&color="#000000"]//text()') [416:660:23]
TSUackleyd=[a.replace("xao","") for a in TSUackleyd]
TSUackleyd=pd.to_numeric(TSUackleyd)
TSUackleyd

Out[305]:
array([26, 21, 19, 21, 14, 8, 11, 28, 45, 14, 18], dtype=int64)
```

```
#get TSU sacks
TSUsacks=mytree.xpath('//body/center//tr/cd/font[&color="#000000"]//text()') [417:660:23]
TSUsacks=[a.replace("xao","") for a in TSUsacks]
TSUsacks=pd.to_numeric(TSUsacks)
TSUsacks

Out[306]:
array([3., 4., 0., 1., 0., 1., 1., 1., 6., 1., 1])

In [307]:
#get TSU sack yards
TSUsackyd=mytree.xpath('//body/center//tr/cd/font[&color="#000000"]//text()') [418:660:23]
TSUsackyd=[a.replace("xao","") for a in TSUsackyd]
TSUsackyd=pd.to_numeric(TSUsackyd)
TSUsackyd

Out[307]:
array([17, 18, 0, 5, 0, 1, 0, 9, 34, 5, 9], dtype=int64)
```

```
#TSU punts
TSUpunt=mytree.xpath('//body/center//tr/cd/font[&color="#000000"]//text()') [711:919:19]
TSUpunt=[a.replace("xao","") for a in TSUpunt]
TSUpunt=pd.to_numeric(TSUpunt)
TSUpunt

Out[308]:
array([ 5, 6, 2, 6, 6, 5, 4, 4, 3, 4, 11], dtype=int64)
```

```
#create data frame
#change dictionary of lists to data frame
list_of_dicts={'date':date,
               'attendance':attendance,
               'TSUruhsyards':TSUruhsyards,
               'TSUreceiveyards':TSUreceiveyards,
               'TSUkreturnyards':TSUkreturnyards,
               'TSUpreturnyards':TSUpreturnyards,
               'TSUackles':TSUackles,
               'TSUacks':TSUacks,
               'TSUackleyd':TSUackleyd,
               'TSUpunt':TSUpunt}
df2017=pd.DataFrame(list_of_dicts)
df2017.head()
```

	date	attendance	TSUruhsyards	TSUreceiveyards	TSUkreturnyards	TSUpreturnyards	TSUackles	TSUackleyd	TSUacks	TSUackyd	TSUp
0	2017-08-31	24333	238	145	44	0	59	26	3.0	17	
1	2017-09-09	47407	160	78	49	41	76	21	4.0	18	
2	2017-09-17	17102	241	273	63	7	38	19	0.0	0	
3	2017-09-23	6484	100	159	160	3	77	21	1.0	5	
4	2017-09-30	11013	83	195	48	0	73	14	0.0	0	

```
df2017['date'].mpd.to_datetime(df2017['date'],format='%b %d %Y')
#creating year column
df2017['year']=df2017.date.dt.year
df2017.head()
```

	date	attendance	TSUruhsyards	TSUreceiveyards	TSUkreturnyards	TSUpreturnyards	TSUackles	TSUackleyd	TSUacks	TSUackyd	TSUp
0	2017-08-31	24333	238	145	44	0	59	26	3.0	17	
1	2017-09-09	47407	160	78	49	41	76	21	4.0	18	
2	2017-09-17	17102	241	273	63	7	38	19	0.0	0	
3	2017-09-23	6484	100	159	160	3	77	21	1.0	5	
4	2017-09-30	11013	83	195	48	0	73	14	0.0	0	

```
#get info on data frame
df2017.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 11 entries, 0 to 10
Data columns (total 12 columns):
# Column Non-Null Count Dtype
---
0 date 11 non-null datetime64[ns]
1 attendance 11 non-null int64
2 TSUruhsyards 11 non-null int64
3 TSUreceiveyards 11 non-null int64
4 TSUkreturnyards 11 non-null int64
5 TSUpreturnyards 11 non-null int64
6 TSUackles 11 non-null int64
7 TSUackleyd 11 non-null float64
8 TSUacks 11 non-null float64
9 TSUackyd 11 non-null int64
10 TSUpunt 11 non-null int64
11 year 11 non-null int64
dtypes: datetime64[ns](1), float64(1), int64(10)
memory usage: 1.2 KB

In [312]:
#sort data
df2017=df2017.sort_values('date',ignore_index=True)
df2017.head()
```

	date	attendance	TSUruhsyards	TSUreceiveyards	TSUkreturnyards	TSUpreturnyards	TSUackles	TSUackleyd	TSUacks	TSUackyd	TSUp
0	2017-08-31	24333	238	145	44	0	59	26	3.0	17	
1	2017-09-09	47407	160	78	49	41	76	21	4.0	18	
2	2017-09-17	17102	241	273	63	7	38	19	0.0	0	
3	2017-09-23	6484	100	159	160	3	77	21	1.0	5	
4	2017-09-30	11013	83	195	48	0	73	14	0.0	0	

```
#save data
df2017.to_csv('2017.csv',encoding='utf-8')
```

```
#2018 data
#use requests.get() to get web page with data
page = requests.get('https://tennstate.fip.sidaexsports.com/customreports/tautgers/125803E8-418B-422C-980C-6451')
#parse data on web page using html module.fromstring
mytree = html.fromstring(page.content)

#go to web address above , right click on page and select inspect to get HTML code for data from right side of
#create XPath query and use xpath function to get data
#date data
date = mytree.xpath('//body/center//tr/cd/font[&color="#000000"]//text()') [1:120:10]
date=[a.replace("xao","") for a in date]
date=[a.replace(" ",",") for a in date]
date=[a.strip() for a in date]
print(date)

['Sep 01 2018', 'Sep 22 2018', 'Sep 29 2018', 'Oct 06 2018', 'Oct 13 2018', 'Oct 20 2018', 'Nov 03 2018', 'Nov
10 2018', 'Nov 17 2018']

In [315]:
#get attendance data
attendance=mytree.xpath('//body/center//tr/cd/font[&color="#000000"]//text()') [9:120:10]
attendance=[a.replace("xao","") for a in attendance]
attendance=[a.strip() for a in attendance]
attendance=pd.to_numeric(attendance)
attendance

Out[315]:
array([14069, 7670, 27340, 12201, 3318, 17283, 3481, 6718, 3618],
      dtype=int64)
```

```
#get TSU rushing yards
TSUruhsyards=mytree.xpath('//body/center//tr/cd/font[&color="#000000"]//text()') [94:300:23]
TSUruhsyards=[a.replace("xao","") for a in TSUruhsyards]
TSUruhsyards=pd.to_numeric(TSUruhsyards)
TSUruhsyards

Out[316]:
array([201, 195, 104, 164, 63, 149, 253, 178, 89], dtype=int64)
```

```
#get TSU receiving yards
TSUreceiveyards=mytree.xpath('//body/center//tr/cd/font[&color="#000000"]//text()') [98:300:23]
TSUreceiveyards=[a.replace("xao","") for a in TSUreceiveyards]
TSUreceiveyards=pd.to_numeric(TSUreceiveyards)
TSUreceiveyards

Out[317]:
array([324, 349, 269, 325, 307, 525, 255, 164, 170], dtype=int64)
```

```
#get TSU kick return yards
TSUkreturnyards=mytree.xpath('//body/center//tr/cd/font[&color="#000000"]//text()') [106:300:23]
TSUkreturnyards=[a.replace("xao","") for a in TSUkreturnyards]
TSUkreturnyards=pd.to_numeric(TSUkreturnyards)
TSUkreturnyards

Out[318]:
array([ 49, 150, 105, 134, 140, 48, 78, 36, 63], dtype=int64)
```

```
#get TSU punt return yards
TSUpreturnyards=mytree.xpath('//body/center//tr/cd/font[&color="#000000"]//text()') [110:300:23]
TSUpreturnyards=[a.replace("xao","") for a in TSUpreturnyards]
TSUpreturnyards=pd.to_numeric(TSUpreturnyards)
TSUpreturnyards

Out[319]:
array([64, 51, 3, 34, -1, 14, 0, 0, 29], dtype=int64)
```

```
#get TSU total tackles
TSUackles=mytree.xpath('//body/center//tr/cd/font[&color="#000000"]//text()') [348:546:23]
TSUackles=[a.replace("xao","") for a in TSUackles]
TSUackles=pd.to_numeric(TSUackles)
TSUackles

Out[320]:
array([54, 63, 69, 77, 74, 77, 65, 60, 64], dtype=int64)
```

```
#get TSU tackle yards
TSUackleyd=mytree.xpath('//body/center//tr/cd/font[&color="#000000"]//text()') [350:546:23]
TSUackleyd=[a.replace("xao","") for a in TSUackleyd]
TSUackleyd=pd.to_numeric(TSUackleyd)
TSUackleyd

Out[321]:
array([43, 30, 5, 19, 19, 28, 28, 27, 38], dtype=int64)
```

```
#get TSU sacks
TSUsacks=mytree.xpath('//body/center//tr/cd/font[&color="#000000"]//text()') [351:540:23]
TSUsacks=[a.replace("xao","") for a in TSUsacks]
TSUsacks=pd.to_numeric(TSUsacks)
TSUsacks

Out[322]:
array([5., 2., 0., 0., 1., 4., 2., 2., 2.])

In [323]:
#get TSU sack yards
TSUsackyd=mytree.xpath('//body/center//tr/cd/font[&color="#000000"]//text()') [352:540:23]
TSUsackyd=[a.replace("xao","") for a in TSUsackyd]
TSUsackyd=pd.to_numeric(TSUsackyd)
TSUsackyd

Out[323]:
array([35, 15, 0, 0, 7, 21, 14, 19, 18], dtype=int64)
```

```
#TSU punts
TSUpunt=mytree.xpath('//body/center//tr/cd/font[&color="#000000"]//text()') [599:768:19]
TSUpunt=[a.replace("xao","") for a in TSUpunt]
TSUpunt=pd.to_numeric(TSUpunt)
TSUpunt

Out[324]:
array([ 6, 4, 4, 4, 5, 4, 3, 5, 7], dtype=int64)
```

```
#create data frame
#change dictionary of lists to data frame
list_of_dicts={'date':date,
               'attendance':attendance,
               'TSUruhsyards':TSUruhsyards,
               'TSUreceiveyards':TSUreceiveyards,
               'TSUkreturnyards':TSUkreturnyards,
               'TSUpreturnyards':TSUpreturnyards,
               'TSUackles':TSUackles,
               'TSUacks':TSUacks,
               'TSUackleyd':TSUackleyd,
               'TSUpunt':TSUpunt}
df2018=pd.DataFrame(list_of_dicts)
df2018.head()
```

	date	attendance	TSUruhsyards	TSUreceiveyards	TSUkreturnyards	TSUpreturnyards	TSUackles	TSUackleyd	TSUacks	TSUackyd	TSUp
0	2018-01-01	14069	201	324	49	64	54	43	5.0	35	
1	2018-09-22	7670	195	349	150	51	63	30	2.0	15	
2	2018-09-29	27340	104	269	105	3	69	5	0.0	0	
3	2018-10-06	12201	164	325	134	34	77	19	0.0	0	
4	2018-10-13	3318	63	307	140	-1	74	19	1.0	7	

```
df2018['date'].mpd.to_datetime(df2018['date'],format='%b %d %Y')
#creating year column
df2018['year']=df2018.date.dt.year
df2018.head()
```

	date	attendance	TSUruhsyards	TSUreceiveyards	TSUkreturnyards	TSUpreturnyards	TSUackles	TSUackleyd	TSUacks	TSUackyd	TSUp
0	2018-01-01	14069	201	324	49	64	54	43	5.0	35	
1	2018-09-22	7670	195	349	150	51	63	30	2.0	15	
2	2018-09-29	27340	104	269	105	3	69	5	0.0	0	
3	2018-10-06	12201	164	325	134	34	77	19	0.0	0	
4	2018-10-13	3318	63	307	140	-1	74	19	1.0	7	

```
#get info from data frame
df2018.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 9 entries, 0 to 8
Data columns (total 12 columns):
# Column Non-Null Count Dtype
---
0 date 9 non-null datetime64[ns]
1 attendance 9 non-null int64
2 TSUruhsyards 9 non-null int64
3 TSUreceiveyards 9 non-null int64
4 TSUkreturnyards 9 non-null int64
5 TSUpreturnyards 9 non-null int64
6 TSUackles 9 non-null int64
7 TSUackleyd 9 non-null float64
8 TSUacks 9 non-null float64
9 TSUackyd 9 non-null int64
10 TSUpunt 9 non-null int64
11 year 9 non-null int64
dtypes: datetime64[ns](1), float64(1), int64(10)
memory usage: 992.0 bytes

In [328]:
#sort values
df2018=df2018.sort_values('date',ignore_index=True)
df2018.head()
```

	date	attendance	TSUruhsyards	TSUreceiveyards	TSUkreturnyards	TSUpreturnyards	TSUackles	TSUackleyd	TSUacks	TSUackyd	TSUp
0	2018-01-01	14069	201	324	49	64	54	43	5.0	35	
1	2018-09-22	7670	195	349	150	51	63	30	2.0	15	
2	2018-09-29	27340	104	269	105	3	69	5	0.0	0	
3	2018-10-06	12201	164	325	134	34	77	19	0.0	0	
4	2018-10-13	3318	63	307	140	-1	74	19	1.0	7	

```
#save data
df2018.to_csv('2018.csv',encoding='utf-8')
```

```
#2019 data
#use requests.get() to get web page with data
page = requests.get('https://tennstate.fip.sidaexsports.com/customreports/tautgers/0018AECB-A18B-4682-B187-46F1')
#parse data on web page using html module.fromstring
mytree = html.fromstring(page.content)

#go to web address above , right click on page and select inspect to get HTML code for data from right side of
#create XPath query and use xpath function to get data
#date data
date = mytree.xpath('//body/center//tr/cd/font[&color="#000000"]//text()') [1:120:10]
date=[a.replace("xao","") for a in date]
date=[a.replace(" ",",") for a in date]
date=[a.strip() for a in date]
print(date)

['Aug 31 2019', 'Sep 07 2019', 'Sep 14 2019', 'Sep 21 2019', 'Sep 28 2019', 'Oct 05 2019', 'Oct 12 2019', 'Oct
19 2019', 'Oct 26 2019', 'Nov 03 2019', 'Nov 10 2019', 'Nov 17 2019']

In [331]:
#get attendance data
attendance=mytree.xpath('//body/center//tr/cd/font[&color="#000000"]//text()') [9:120:10]
attendance=[a.replace("xao","") for a in attendance]
attendance=[a.strip() for a in attendance]
attendance=pd.to_numeric(attendance)
attendance

Out[331]:
array([13458, 20912, 48347, 8683, 8861, 16389, 5324, 16389, 4738,
       4132, 1776, 2728], dtype=int64)
```

```
#get TSU rushing yards
TSUruhsyards=mytree.xpath('//body/center//tr/cd/font[&color="#000000"]//text()') [124:383:23]
TSUruhsyards=[a.replace("xao","") for a in TSUruhsyards]
TSUruhsyards=pd.to_numeric(TSUruhsyards)
TSUruhsyards

Out[332]:
array([117, 207, 141, 71, 112, 230, 108, 116, 83, 97, 193, 198],
      dtype=int64)
```

```
#get TSU receiving yards
TSUreceiveyards=mytree.xpath('//body/center//tr/cd/font[&color="#000000"]//text()') [128:383:23]
TSUreceiveyards=[a.replace("xao","") for a in TSUreceiveyards]
TSUreceiveyards=pd.to_numeric(TSUreceiveyards)
TSUreceiveyards

Out[333]:
array([130, 259, 389, 334, 270, 240, 332, 229, 209, 123, 225],
      dtype=int64)
```

```
#get TSU kick return yards
TSUkreturnyards=mytree.xpath('//body/center//tr/cd/font[&color="#000000"]//text()') [136:400:23]
TSUkreturnyards=[a.replace("xao","") for a in TSUkreturnyards]
TSUkreturnyards=pd.to_numeric(TSUkreturnyards)
TSUkreturnyards

Out[334]:
array([ 13, 108, 155, 74, 38, 56, 55, 66, 113, 162, 105, 51],
      dtype=int64)
```

```
#get TSU punt return yards
TSUpreturnyards=mytree.xpath('//body/center//tr/cd/font[&color="#000000"]//text()') [140:400:23]
TSUpreturnyards=[a.replace("xao","") for a in TSUpreturnyards]
TSUpreturnyards=pd.to_numeric(TSUpreturnyards)
TSUpreturnyards

Out[335]:
array([61, 0, 8, 15, 7, 0, 8, 0, 13, 62, 6, 25], dtype=int64)
```

```
#get TSU total tackles
TSUackles=mytree.xpath('//body/center//tr/cd/font[&color="#000000"]//text()')
```



	date	attendance	TSUrushyards	TSUreceiveyards	TSUketurnyards	TSUpurturnyards	TSUstackles	TSUstackleyd	TSUacks	TSUackeyd	TSUsfz
0	2003-09-03	25342	172	186	154	62	32	47	20	2.0	13
1	2003-09-06	48300	120	103	103	61	47	59	29	3.0	16
2	2003-09-13	22631	111	107	105	61	47	59	29	3.0	16
3	2003-09-20	27460	71	181	70	0	59	25	1.0	7	
4	2003-09-27	57885	227	210	98	16	54	14	1.0	4	

```
#read in pd.read_csv('2008.csv')
#get rid of unnamed list variable that's just the index put on when file was saved to a csv file
df2008=df2008.iloc[:,1:]
df2008.head()
```

	date	attendance	TSUrushyards	TSUreceiveyards	TSUketurnyards	TSUpurturnyards	TSUstackles	TSUstackleyd	TSUacks	TSUackeyd	TSUsfz
0	2008-09-02	15487	156	206	105	0	63	3	0.0	0	
1	2008-09-06	19613	111	230	102	0	76	38	3.0	23	
2	2008-09-13	53441	146	167	77	24	63	18	1.0	6	
3	2008-09-20	27460	71	181	70	0	59	25	1.0	7	
4	2008-09-27	57885	227	210	98	16	54	14	1.0	4	

```
#read in pd.read_csv('2007.csv')
#get rid of unnamed list variable that's just the index put on when file was saved to a csv file
df2007=df2007.iloc[:,1:]
df2007.head()
```

	date	attendance	TSUrushyards	TSUreceiveyards	TSUketurnyards	TSUpurturnyards	TSUstackles	TSUstackleyd	TSUacks	TSUackeyd	TSUsfz
0	2007-09-01	23440	206	176	154	35	56	17	2	15	
1	2007-09-08	50879	153	232	84	24	63	30	2	16	
2	2007-09-15	10861	238	233	83	14	64	44	3	26	
3	2007-09-22	15371	133	316	135	15	80	56	3	36	
4	2007-09-29	56990	177	309	111	33	78	28	3	23	

```
#read in pd.read_csv('2008.csv')
#get rid of unnamed list variable that's just the index put on when file was saved to a csv file
df2008=df2008.iloc[:,1:]
df2008.head()
```

	date	attendance	TSUrushyards	TSUreceiveyards	TSUketurnyards	TSUpurturnyards	TSUstackles	TSUstackleyd	TSUacks	TSUackeyd	TSUsfz
0	2008-08-30	10072	107	292	160	2	46	20	1	10	
1	2008-09-06	28830	148	315	66	37	76	29	1	4	
2	2008-09-13	50794	148	137	34	38	58	9	0	0	
3	2008-09-20	8276	229	314	84	6	51	48	3	27	
4	2008-09-27	50428	136	241	95	2	61	37	3	23	

```
#read in pd.read_csv('2009.csv')
#get rid of unnamed list variable that's just the index put on when file was saved to a csv file
df2009=df2009.iloc[:,1:]
df2009.head()
```

	date	attendance	TSUrushyards	TSUreceiveyards	TSUketurnyards	TSUpurturnyards	TSUstackles	TSUstackleyd	TSUacks	TSUackeyd	TSUsfz
0	2009-09-05	23871	13	173	46	0	71	18	1	7	
1	2009-09-12	43306	217	43	19	20	60	42	3	26	
2	2009-09-19	12247	169	86	75	13	81	28	3	15	
3	2009-09-26	51950	167	107	95	2	63	14	1	5	
4	2009-09-29	6314	259	95	72	30	58	20	3	15	

```
#read in pd.read_csv('2010.csv')
#get rid of unnamed list variable that's just the index put on when file was saved to a csv file
df2010=df2010.iloc[:,1:]
df2010.head()
```

	date	attendance	TSUrushyards	TSUreceiveyards	TSUketurnyards	TSUpurturnyards	TSUstackles	TSUstackleyd	TSUacks	TSUackeyd	TSUsfz
0	2010-09-04	22607	107	127	127	36	64	42	7.0	40	
1	2010-09-11	44688	224	185	133	28	60	26	0.0	0	
2	2010-09-18	8502	162	225	134	25	57	24	2.0	13	
3	2010-09-25	54202	304	109	33	142	58	60	8.0	48	
4	2010-09-30	35217	379	142	20	13	66	34	5.0	31	

```
#read in pd.read_csv('2011.csv')
#get rid of unnamed list variable that's just the index put on when file was saved to a csv file
df2011=df2011.iloc[:,1:]
df2011.head()
```

	date	attendance	TSUrushyards	TSUreceiveyards	TSUketurnyards	TSUpurturnyards	TSUstackles	TSUstackleyd	TSUacks	TSUackeyd	TSUsfz
0	2011-09-03	25209	342	170	48	0	45	41	5.0	30	
1	2011-09-10	43532	78	226	146	5	64	45	3.0	31	
2	2011-09-17	10031	168	207	197	17	94	32	3.0	21	
3	2011-09-24	33487	188	206	97	0	88	0	0.0	0	
4	2011-09-27	8614	162	304	158	0	61	20	2.0	10	

```
#read in pd.read_csv('2012.csv')
#get rid of unnamed list variable that's just the index put on when file was saved to a csv file
df2012=df2012.iloc[:,1:]
df2012.head()
```

	date	attendance	TSUrushyards	TSUreceiveyards	TSUketurnyards	TSUpurturnyards	TSUstackles	TSUstackleyd	TSUacks	TSUackeyd	TSUsfz
0	2012-09-01	15652	138	263	52	8	49	28	1	5	
1	2012-09-08	42257	225	337	69	42	66	21	3	20	
2	2012-09-15	14264	112	322	24	7	55	13	1	9	
3	2012-09-22	9461	200	157	41	37	84	52	4	30	
4	2012-09-29	31765	201	262	47	4	62	40	3	29	

```
#read in pd.read_csv('2013.csv')
#get rid of unnamed list variable that's just the index put on when file was saved to a csv file
df2013=df2013.iloc[:,1:]
df2013.head()
```

	date	attendance	TSUrushyards	TSUreceiveyards	TSUketurnyards	TSUpurturnyards	TSUstackles	TSUstackleyd	TSUacks	TSUackeyd	TSUsfz
0	2013-09-01	16108	116	132	84	51	62	7	1.0	4	
1	2013-09-07	14237	268	131	111	17	41	35	3.0	18	
2	2013-09-14	42400	174	111	64	11	69	26	2.0	13	
3	2013-09-21	10044	95	343	63	23	76	51	3.0	28	
4	2013-09-28	22000	311	228	46	75	70	44	3.0	17	

```
#read in pd.read_csv('2014.csv')
#get rid of unnamed list variable that's just the index put on when file was saved to a csv file
df2014=df2014.iloc[:,1:]
df2014.head()
```

	date	attendance	TSUrushyards	TSUreceiveyards	TSUketurnyards	TSUpurturnyards	TSUstackles	TSUstackleyd	TSUacks	TSUackeyd	TSUsfz
0	2014-08-30	10541	439	71	77	57	53	38	2.0	13	
1	2014-09-06	15725	92	340	111	22	89	28	3.0	13	
2	2014-09-13	46914	137	187	29	40	54	40	7.0	33	
3	2014-09-20	9217	92	113	0	15	48	31	6.0	28	
4	2014-09-27	29225	125	153	24	56	64	51	7.0	44	

```
#read in pd.read_csv('2015.csv')
#get rid of unnamed list variable that's just the index put on when file was saved to a csv file
df2015=df2015.iloc[:,1:]
df2015.head()
```

	date	attendance	TSUrushyards	TSUreceiveyards	TSUketurnyards	TSUpurturnyards	TSUstackles	TSUstackleyd	TSUacks	TSUackeyd	TSUsfz
0	2015-09-06	22455	142	188	58	38	47	21	1.0	12	
1	2015-09-12	48385	110	304	152	48	83	38	5.0	31	
2	2015-09-19	23413	24	184	102	0	83	4	0.0	0	
3	2015-09-26	18020	169	238	46	15	68	44	4.0	22	
4	2015-09-29	7123	85	205	40	4	71	7	0.0	0	

```
#read in pd.read_csv('2016.csv')
#get rid of unnamed list variable that's just the index put on when file was saved to a csv file
df2016=df2016.iloc[:,1:]
df2016.head()
```

	date	attendance	TSUrushyards	TSUreceiveyards	TSUketurnyards	TSUpurturnyards	TSUstackles	TSUstackleyd	TSUacks	TSUackeyd	TSUsfz
0	2016-09-03	15078	202	259	29	64	52	39	5.0	31	
1	2016-09-10	46263	121	273	96	11	61	40	3.0	21	
2	2016-09-17	9385	210	184	122	23	52	17	1.0	1	
3	2016-09-24	10001	141	223	80	0	63	28	2.0	12	
4	2016-09-27	4319	76	303	113	0	75	6	1.0	3	

```
#read in pd.read_csv('2017.csv')
#get rid of unnamed list variable that's just the index put on when file was saved to a csv file
df2017=df2017.iloc[:,1:]
df2017.head()
```

	date	attendance	TSUrushyards	TSUreceiveyards	TSUketurnyards	TSUpurturnyards	TSUstackles	TSUstackleyd	TSUacks	TSUackeyd	TSUsfz
0	2017-08-31	24333	238	145	44	0	59	26	3.0	17	
1	2017-09-07	47407	160	78	49	41	76	21	4.0	18	
2	2017-09-14	17102	241	273	63	7	38	19	0.0	0	
3	2017-09-21	6484	100	159	160	3	77	21	1.0	5	
4	2017-09-28	11013	83	195	48	0	73	14	0.0	0	

```
#read in pd.read_csv('2018.csv')
#get rid of unnamed list variable that's just the index put on when file was saved to a csv file
df2018=df2018.iloc[:,1:]
df2018.head()
```

	date	attendance	TSUrushyards	TSUreceiveyards	TSUketurnyards	TSUpurturnyards	TSUstackles	TSUstackleyd	TSUacks	TSUackeyd	TSUsfz
0	2018-09-01	14069	201	324	49	64	54	43	5.0	35	
1	2018-09-22	7670	195	349	150	51	63	30	2.0	15	
2	2018-09-29	27340	104	269	105	3	69	5	0.0	0	
3	2018-10-06	12201	164	325	134	34	77	19	0.0	0	
4	2018-10-13	3318	63	307	140	-1	74	19	1.0	7	

```
#read in pd.read_csv('2019.csv')
#get rid of unnamed list variable that's just the index put on when file was saved to a csv file
df2019=df2019.iloc[:,1:]
df2019.head()
```

	date	attendance	TSUrushyards	TSUreceiveyards	TSUketurnyards	TSUpurturnyards	TSUstackles	TSUstackleyd	TSUacks	TSUackeyd	TSUsfz
0	2019-08-31	13458	117	330	13	61	58	3	0.0	0	
1	2019-09-07	20912	207	259	108	0	49	13	1.0	7	
2	2019-09-14	48347	141	389	155	0	55	19	1.0	8	
3	2019-09-21	8683	71	334	74	15	65	13	1.0	3	
4	2019-09-28	8861	112	270	38	7	74	11	1.0	5	

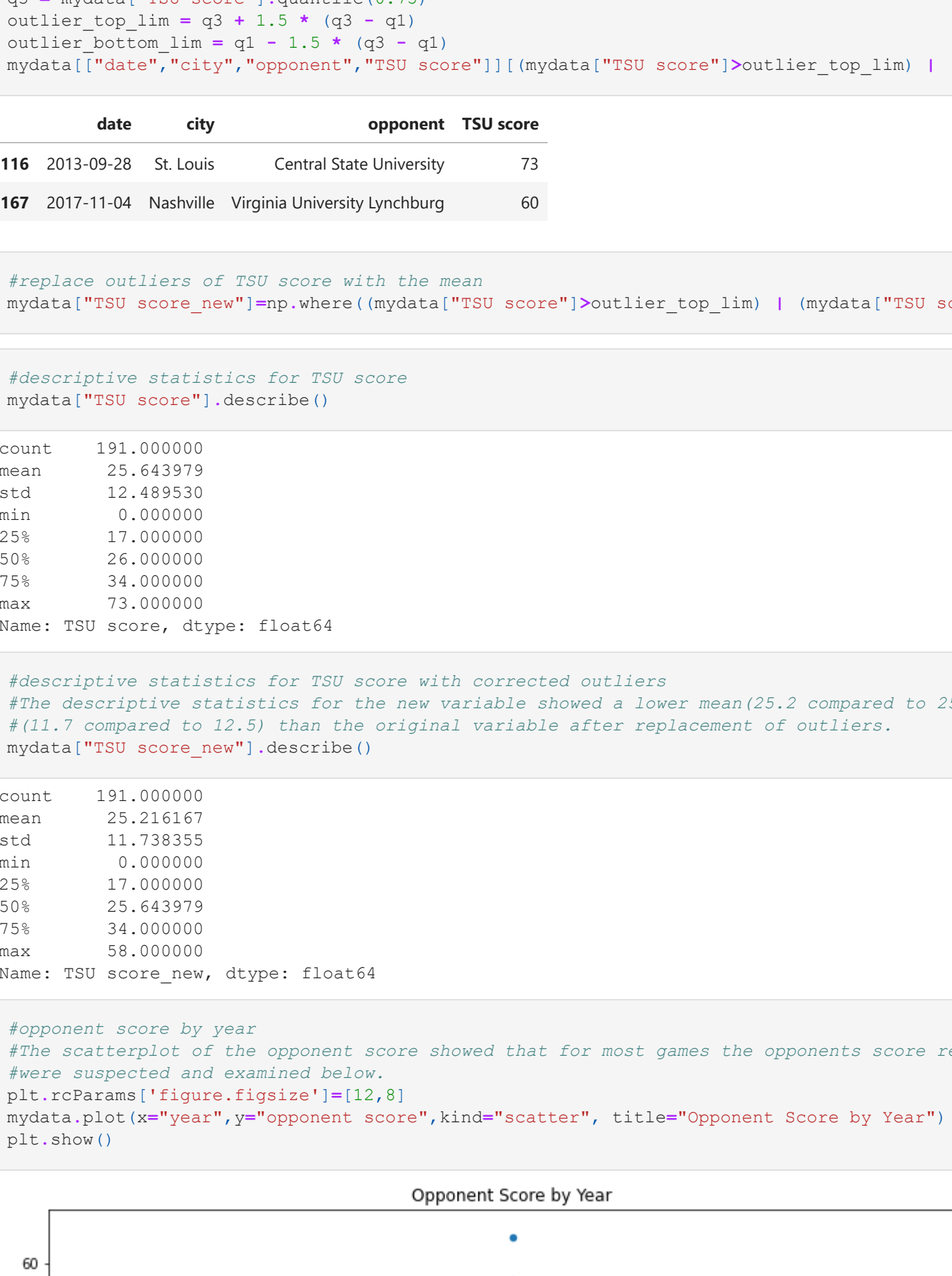
```
#append all annual state dataframes into a single data frame (stacking rows)
#the annual number of away games from 2003 to 
```



<AxesSubplot:title='Number of Home and Away Games by Year', xlabel='year, locale')>



#TSU score by year  
#For most games during the year, the TSU score remained between about 10 and 45 points. Based on the scatterplot, there appeared to be some outliers that warranted further examination.  
plt.rcParams['figure.figsize']=(12,8)  
mydata.plot(x='year',y='TSU score',kind='scatter', title='TSU Score by Year')  
plt.show()



#get outliers of TSU score  
#There were 10 outliers on the TSU score, a 73 points against Central State University and 60 points against Virginia University Lynchburg.  
#Another TSU score variable was made with the values of these outliers being replaced the mean.  
q1 = mydata["TSU score"].quantile(0.25)  
q3 = mydata["TSU score"].quantile(0.75)  
outlier\_top\_lim = q3 + 1.5 \* (q3 - q1)  
outlier\_bottom\_lim = q1 - 1.5 \* (q3 - q1)  
mydata["date","city","opponent","TSU score"][[mydata["TSU score"]>outlier\_top\_lim | (mydata["TSU score"]<outlier\_bottom\_lim)]]

```
#get outliers of opponent score
#There was one outlier on the opponent score variable, 63 points scored by the Air Force
#This was replaced with the mean below.
q1 = mydata["opponent score"].quantile(0.25)
q3 = mydata["opponent score"].quantile(0.75)
outlier_top_lim = q3 + 1.5 * (q3 - q1)
outlier_bottom_lim = q1 - 1.5 * (q3 - q1)
mydata["date"] = c("11/6", "opponent", "opponent score")[[mydata["opponent score"] > outlier_top_lim | mydata["opponent score"] < outlier_bottom_lim]]
```

#replace outliers of TSU score with the mean  
mydata["TSU score\_new"] = mydata.where(mydata["TSU score"]>outlier\_top\_lim | (mydata["TSU score"]<outlier\_bottom\_lim), mydata["TSU score\_new"], mydata["TSU score"])

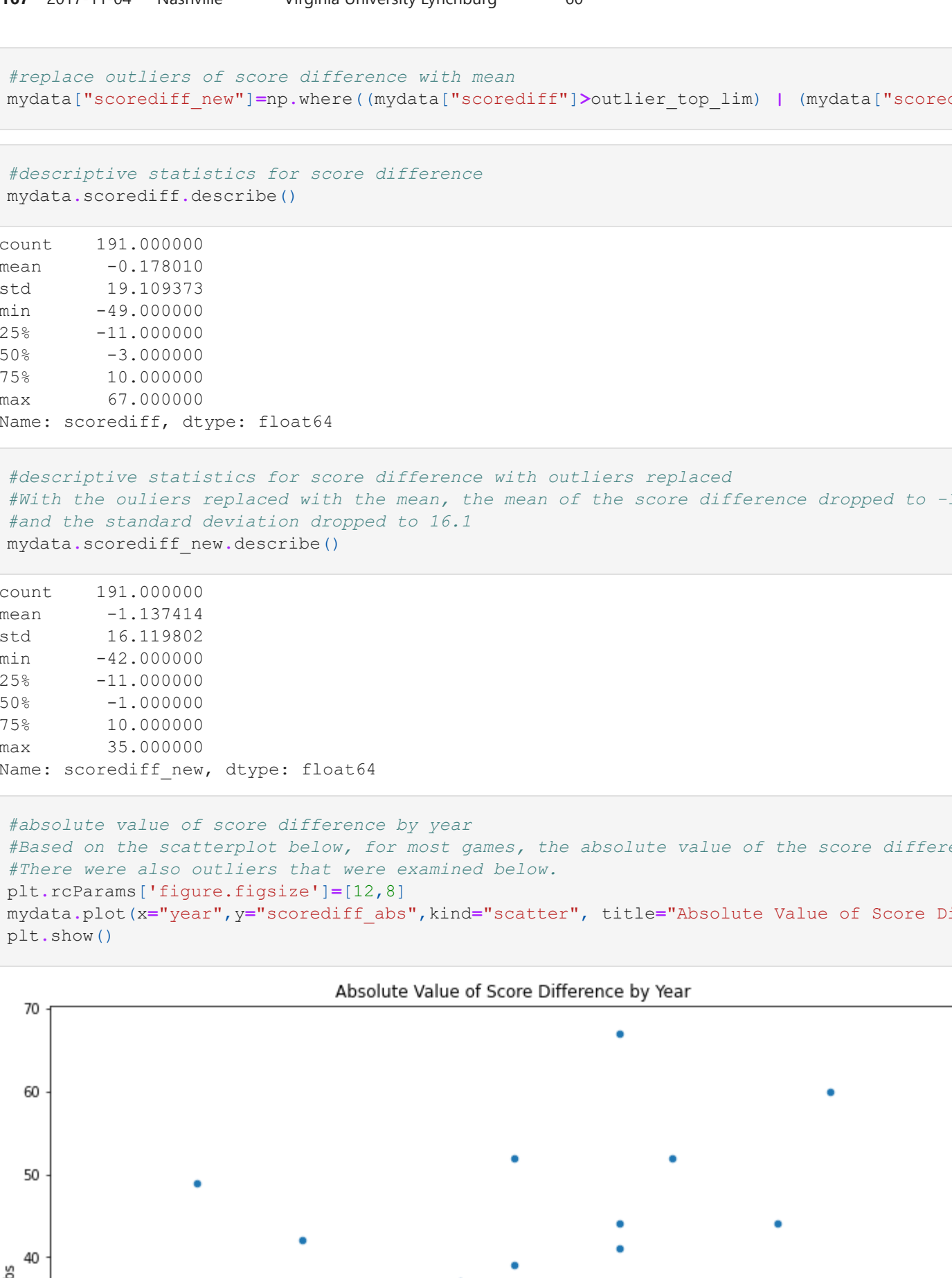
#descriptive statistics for TSU score  
mydata["TSU score"].describe()

count	191.000000
mean	25.463979
std	12.489530
min	0.000000
25%	17.000000
50%	26.000000
75%	34.000000
max	73.000000
Name:	TSU score, dtype: float64

#descriptive statistics for TSU score with corrected outliers  
#The descriptive statistics for the new variable showed a lower mean (25.2 compared to 25.6) and standard deviation (11.7 compared to 12.5) than the original variable after replacement of outliers.  
mydata["TSU score\_new"].describe()

count	191.000000
mean	25.216167
std	11.738355
min	0.000000
25%	17.000000
50%	25.000000
75%	34.000000
max	58.000000
Name:	TSU score_new, dtype: float64

#opponent score by year  
#Based on the scatterplot of the opponent score showed that for most games the opponents score remained between 0 and 50 points. There were also outliers that were examined below.  
plt.rcParams['figure.figsize']=(12,8)  
mydata.plot(x='year',y='opponent score',kind='scatter', title='Opponent Score by Year')  
plt.show()



#get outliers of opponent score  
#There were 63 outliers on the opponent score variable, 63 points scored by the Air Force Academy.  
#TSU was replaced with the mean below.  
q1 = mydata["opponent score"].quantile(0.25)  
q3 = mydata["opponent score"].quantile(0.75)  
outlier\_top\_lim = q3 + 1.5 \* (q3 - q1)  
outlier\_bottom\_lim = q1 - 1.5 \* (q3 - q1)  
mydata["date","city","opponent","opponent score"][[mydata["opponent score"]>outlier\_top\_lim | (mydata["opponent score"]<outlier\_bottom\_lim)]]

	date	city	opponent	scorediff_abs
33	2005-11-19	Nashville	Eastern Kentucky University	49
95	2011-10-08	Nashville	Southeast Missouri State University	52
116	2013-09-28	St. Louis	Central State University	67
126	2014-08-30	Nashville	Edward Waters College	52

#replace outliers of opponent score with the mean  
mydata["opponent score\_new"] = mydata.where(mydata["opponent score"]>outlier\_top\_lim | (mydata["opponent score"]<outlier\_bottom\_lim), mydata["opponent score\_new"], mydata["opponent score"])

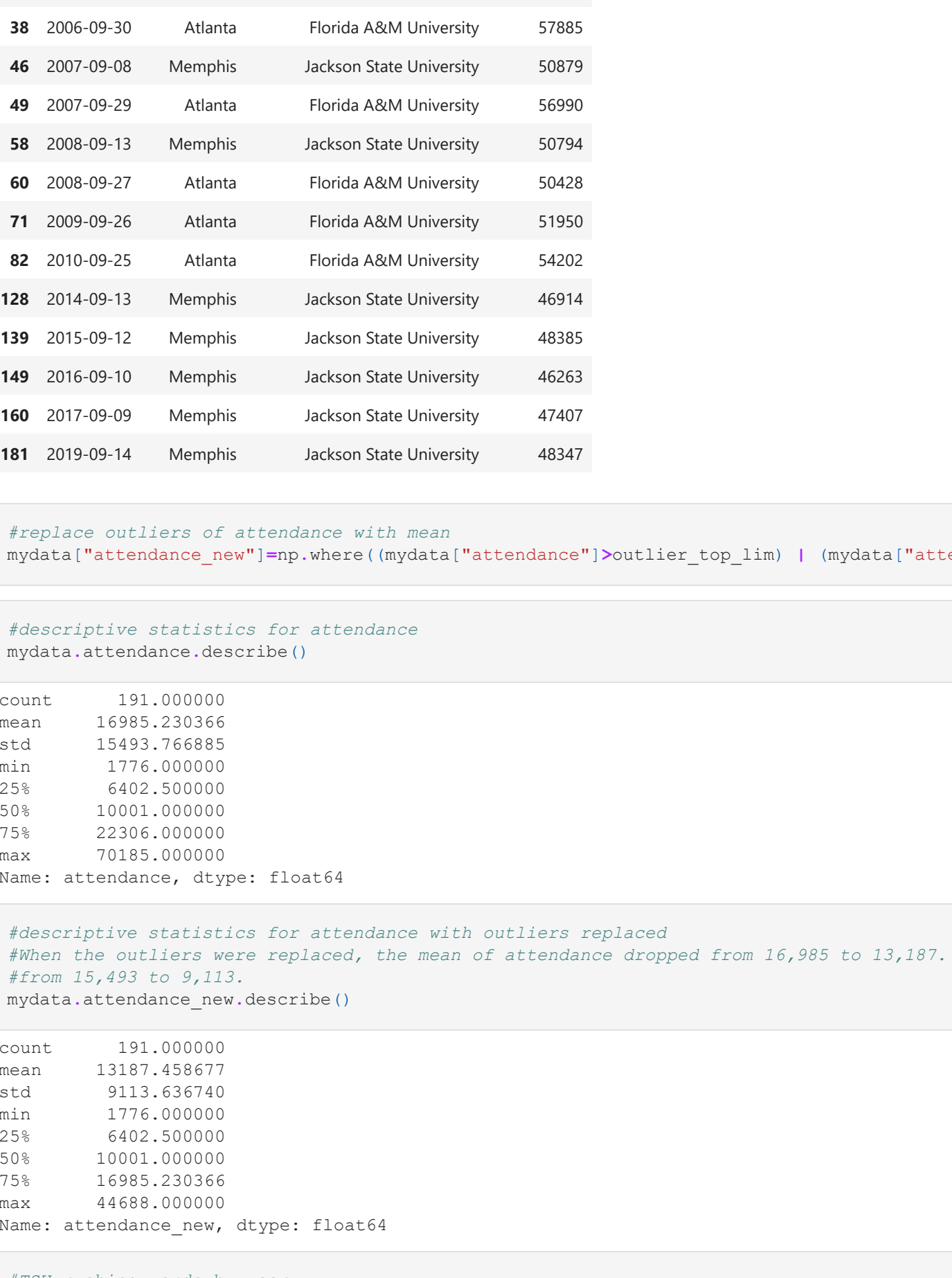
#descriptive statistics for opponent score  
mydata["opponent score"].describe()

count	191.000000
mean	12.627340
std	12.94969
min	0.000000
25%	15.000000
50%	27.000000
75%	33.500000
max	63.000000
Name:	opponent score, dtype: float64

#descriptive statistics for opponent score with outliers replaced  
#The descriptive statistics for the new variable showed a lower mean (12.9 to 12.7) for opponent score and standard deviation (12.9 to 12.7) for opponent score.  
mydata["opponent score\_new"].describe()

count	191.000000
mean	12.664186
std	12.664186
min	0.000000
25%	15.000000
50%	27.000000
75%	33.500000
max	58.000000
Name:	opponent score_new, dtype: float64

#score difference by year  
#Based on the scatterplot of the score difference showed that it remained between -20 and 40 for most games. It also showed that there were outliers that were examined below.  
plt.rcParams['figure.figsize']=(12,8)  
mydata.plot(x='year',y='scorediff',kind='scatter', title='Score Difference by Year')  
plt.show()



#get outliers of score difference  
#There were 7 outliers on the score difference variable. 2 were games in which TSU lost to Eastern Kentucky University and 5 were games in which TSU won by at least 44 points. A new variable was made with the values of these outliers being replaced with the mean.  
q1 = mydata["scorediff"].quantile(0.25)  
q3 = mydata["scorediff"].quantile(0.75)  
outlier\_top\_lim = q3 + 1.5 \* (q3 - q1)  
outlier\_bottom\_lim = q1 - 1.5 \* (q3 - q1)  
mydata["date","city","opponent","scorediff"][[mydata["scorediff"]>outlier\_top\_lim | (mydata["scorediff"]<outlier\_bottom\_lim)]]

date	city	opponent	scorediff	
33	2005-11-19	Nashville	Eastern Kentucky University	-49
95	2011-10-08	Nashville	Southeast Missouri State University	52
116	2013-09-28	St. Louis	Central State University	67
121	2013-11-02	Richmond	Eastern Kentucky University	44
126	2014-08-30	Nashville	Edward Waters College	52
48	2016-09-03	Nashville	University of Arkansas Pine Bluff	54
167	2017-11-04	Nashville	Virginia University Lynchburg	60

#replace outliers of score difference with mean  
mydata["scorediff\_new"] = mydata.where(mydata["scorediff"]>outlier\_top\_lim | (mydata["scorediff"]<outlier\_bottom\_lim), mydata["scorediff\_new"], mydata["scorediff"])

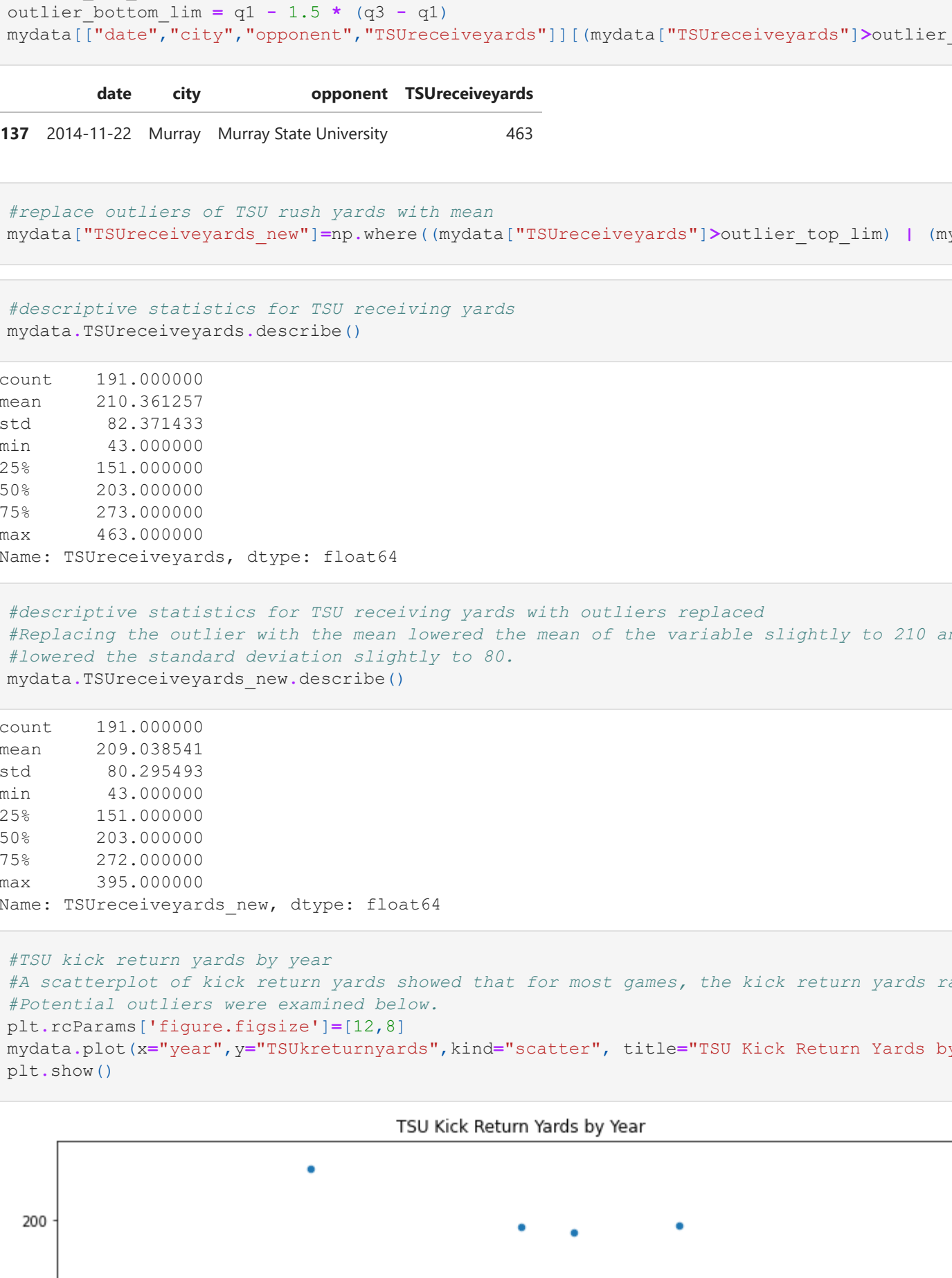
#descriptive statistics for score difference  
mydata["scorediff"].describe()

count	191.000000
mean	-0.178010
std	10.089373
min	-49.000000
25%	-11.000000
50%	-3.000000
75%	10.000000
max	67.000000
Name:	scorediff, dtype: float64

#descriptive statistics for score difference with outliers replaced  
#With the outliers replaced with the mean, the mean of the score difference dropped to -1.1 and the standard deviation dropped to 10.1.  
mydata["scorediff\_new"].describe()

count	191.000000
mean	-1.137414
std	10.108702
min	-42.000000
25%	-11.000000
50%	-1.000000
75%	10.000000
max	35.000000
Name:	scorediff_new, dtype: float64

#absolute value of score difference by year  
#Based on the scatterplot below, for most games, the absolute value of the score difference remained between 0 and 40 points. There were also outliers that were examined below.  
plt.rcParams['figure.figsize']=(12,8)  
mydata.plot(x='year',y='scorediff\_abs',kind='scatter', title='Absolute Value of Score Difference by Year')  
plt.show()



#get outliers of absolute value of score difference  
#There were 5 outliers on the absolute value of the score difference variable, all which had values of at least 44 points. A new variable was made with the values of these outliers being replaced with the mean.  
q1 = mydata["scorediff\_abs"].quantile(0.25)  
q3 = mydata["scorediff\_abs"].quantile(0.75)  
outlier\_top\_lim = q3 + 1.5 \* (q3 - q1)  
outlier\_bottom\_lim = q1 - 1.5 \* (q3 - q1)  
mydata["date","city","opponent","scorediff\_abs"][[mydata["scorediff\_abs"]>outlier\_top\_lim | (mydata["scorediff\_abs"]<outlier\_bottom\_lim)]]

```
#with 222 yards. This was replaced with the variable mean.
q1 = mydata["TSUkretturnyards"].quantile(0.25)
q3 = mydata["TSUkretturnyards"].quantile(0.75)
outlier_top_lim = q3 + 1.5 * (q3 - q1)
outlier_bottom_lim = q1 - 1.5 * (q3 - q1)
mydata[["date", "city", "opponent", "TSUkretturnyards"]][mydata["TSUkretturnyards"]>outlier_top_lim | mydata["TSUkretturnyards"]<outlier_bottom_lim] = NA
```

#replace outliers of absolute value of score difference with mean  
mydata["scorediff\_abs\_new"] = mydata.where(mydata["scorediff\_abs"]>outlier\_top\_lim | (mydata["scorediff\_abs"]<outlier\_bottom\_lim), mydata["scorediff\_abs\_new"], mydata["scorediff\_abs"])

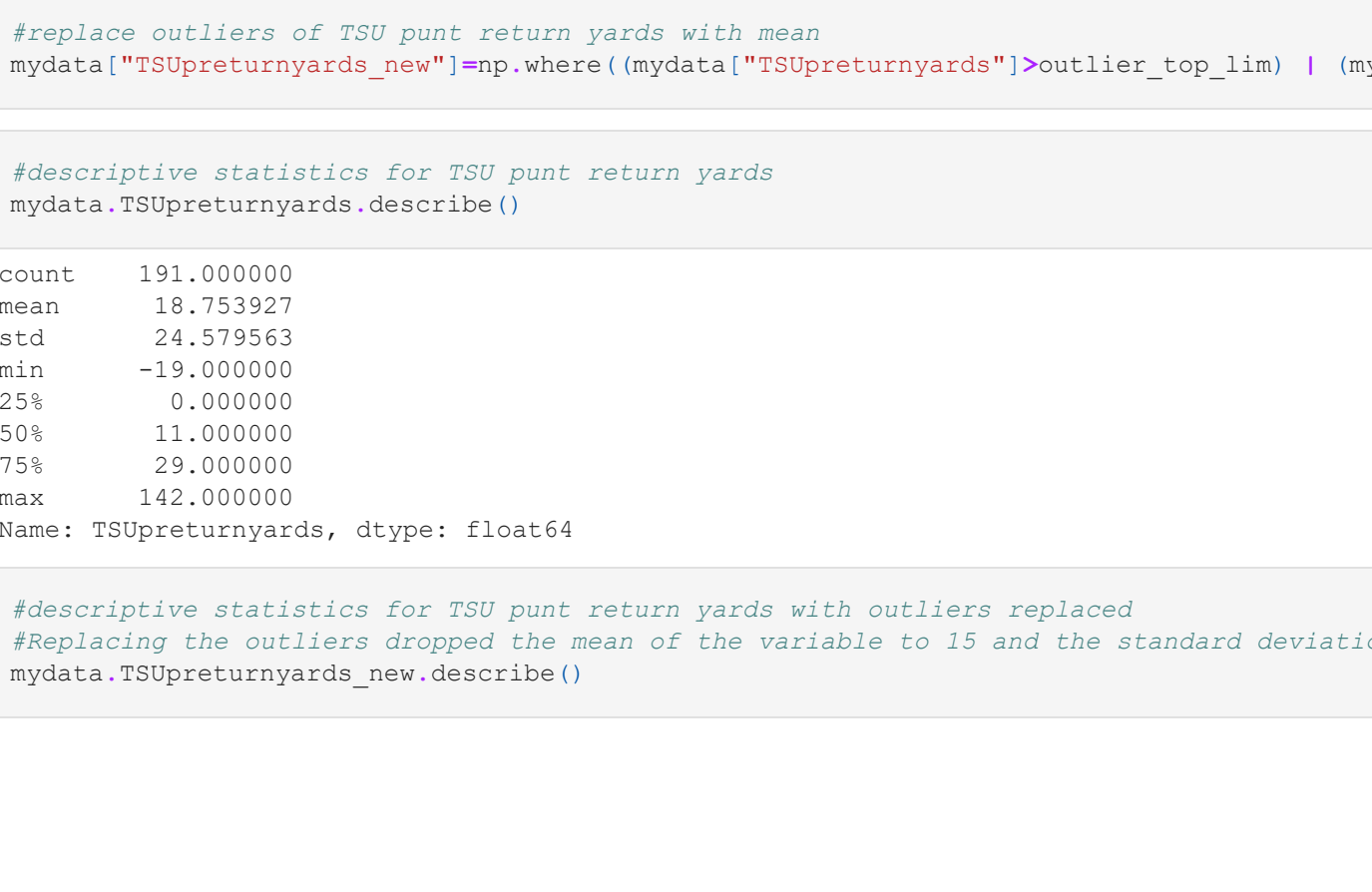
#descriptive statistics for absolute difference of score difference  
mydata["scorediff\_abs"].describe()

count	191.000000
mean	14.502618
std	12.400286
min	1.000000
25%	5.000000
50%	11.000000
75%	21.000000
max	67.000000
Name:	scorediff_abs, dtype: float64

#descriptive statistics for absolute difference of score difference with outliers replaced  
#As with the previous variables that replaced the outlier with the mean, the mean and standard deviation of the absolute value of the score difference dropped once the outliers were replaced.  
mydata["scorediff\_abs\_new"].describe()

count	191.000000
mean	10.313300
std	10.313300
min	1.000000
25%	5.000000
50%	11.000000
75%	20.000000
max	44.000000
Name:	scorediff_abs_new, dtype: float64

#attendance by year  
#The scatter plot for attendance showed that for most games, the attendance was less than 35,000. However, there were 4 outliers on the attendance variable, all of which had values of at least 45,000. The outliers were assessed below.  
plt.rcParams['figure.figsize']=(12,8)  
mydata.plot(x='year',y='attendance',kind='scatter', title='Attendance by Year')  
plt.show()



#get outliers of attendance  
#There were 20 outliers for attendance. These games had attendance figures of at least about 47,000. These games were against Jacksonville State University, Florida A&M University, and South Carolina State University. The attendance for these games was replaced with the mean below.  
q1 = mydata["attendance"].quantile(0.25)  
q3 = mydata["attendance"].quantile(0.75)  
outlier\_top\_lim = q3 + 1.5 \* (q3 - q1)  
outlier\_bottom\_lim = q1 - 1.5 \* (q3 - q1)  
mydata["date","city","opponent","attendance"][[mydata["attendance"]>outlier\_top\_lim | (mydata["attendance"]<outlier\_bottom\_lim)]]

date	city	opponent	attendance	
2	2003-09-13	Memphis	Jackson State University	52603
3	2003-09-20	Atlanta	Florida A&M University	70185
14	2004-09-18	Memphis	Jackson State University	55015
15	2004-09-25	Atlanta	Florida A&M University	67712
16	2004-10-02	Indianapolis	South Carolina State University	51802
24	2005-09-10	Memphis	Jackson State University	48300
26	2005-09-24	Atlanta	Florida A&M University	56297
36	2006-09-16	Memphis	Jackson State University	53441
38	2006-09-30	Atlanta	Florida A&M University	57885
46	2007-09-08	Memphis	Jackson State University	50879
49	2007-09-29	Atlanta	Florida A&M University	56990
58	2008-09-13	Memphis	Jackson State University	50794
60	2008-09-27	Atlanta	Florida A&M University	50428
71	2009-09-26	Atlanta	Florida A&M University	51950
82	2010-09-25	Atlanta	Florida A&M University	54202
128	2014-09-13	Memphis	Jackson State University	46914
139	2015-09-12	Memphis	Jackson State University	48385
149	2016-09-10	Memphis	Jackson State University	46263
160	2017-09-09	Memphis	Jackson State University	47407
181	2019-09-14	Memphis	Jackson State University	48347

#replace outliers of attendance with mean  
mydata["attendance\_new"] = mydata.where(mydata["attendance"]>outlier\_top\_lim | (mydata["attendance"]<outlier\_bottom\_lim), mydata["attendance\_new"], mydata["attendance"])

#descriptive statistics for attendance  
mydata["attendance"].describe()

count	191.000000
mean	16985.20366
std	15493.76885
min	1776.000000
25%	6402.500000
50%	10001.000000
75%	22306.000000
max	70185.000000
Name:	attendance, dtype: float64

#descriptive statistics for attendance with outliers replaced  
#When the outliers were replaced, the mean of attendance dropped from 16,985 to 13,187. The standard deviation dropped from 15,493 to 9,113.  
mydata["attendance\_new"].describe()

count	191.000000
mean	13187.458677
std	9113.636740
min	1776.000000
25%	6402.500000
50%	10001.000000
75%	16985.20366
max	44688.000000
Name:	attendance_new, dtype: float64

#TSU rushing yards by year  
#The scatter plot for TSU rushing yards showed that the rushing yards ranged from 0 to 300 per game annually. There were 4 outliers on the TSU rushing yards variable, all of which had values of at least 342 rushing yards. These outliers were replaced with the mean below.  
plt.rcParams['figure.figsize']=(12,8)  
mydata.plot(x='year',y='TSU rushing yards',kind='scatter', title='TSU Rushing Yards by Year')  
plt.show()



#get outliers of TSU rushing yards  
#There were 4 outliers on the TSU rushing yards variable. In these 4 games, TSU had at least 342 rushing yards. These outliers were replaced with the mean below.  
q1 = mydata["TSU rushing yards"].quantile(0.25)  
q3 = mydata["TSU rushing yards"].quantile(0.75)  
outlier\_top\_lim = q3 + 1.5 \* (q3 - q1)  
outlier\_bottom\_lim = q1 - 1.5 \* (q3 - q1)  
mydata["date","city","opponent","TSU rushing yards"][[mydata["TSU rushing yards"]>outlier\_top\_lim | (mydata["TSU rushing yards"]<outlier\_bottom\_lim)]]

date	city	opponent	TSU rushing yards	
62	2008-10-18	Nashville	Austin Peay State University	377
83	2010-10-02	Indianapolis	North Carolina A&T State University	379
90	2011-09-03	Nashville	Southern University & A&M College	342
126	2014-08-30	Nashville	Edward Waters College	439

#replace outliers of TSU rush yards with mean  
mydata["TSU rushing yards\_new"] = mydata.where(mydata["TSU rushing yards"]>outlier\_top\_lim | (mydata["TSU rushing yards"]<outlier\_bottom\_lim), mydata["TSU rushing yards\_new"], mydata["TSU rushing yards"])

#descriptive statistics for TSU rushing yards  
mydata["TSU rushing yards"].describe()

count	191.000000
mean	137.769634
std	74.855646
min	-18.000000
25%	107.000000
50%	152.000000
75%	200.000000
max	439.000000
Name:	TSU rushing yards, dtype: float64

#descriptive statistics for TSU rushing yards with outliers replaced  
#Replacing the outliers with the mean lowered the mean of the variable slightly to 153 and about 67, respectively when the outliers were replaced.  
mydata["TSU rushing yards\_new"].describe()

count	191.000000
mean	153.026589
std	67.919739
min	-18.000000
25%	107.000000
50%	152.000000
75%	195.500000
max	320.000000
Name:	TSU rushing yards_new, dtype: float64

#TSU receiving yards by year  
#The scatter plot for TSU receiving yards showed that for most games the receiving yard ranged between 100 and 300 yards. There were 8 outliers on the TSU receiving yards variable, all of which had values of at least 342 receiving yards. These outliers were replaced with the mean below.  
plt.rcParams['figure.figsize']=(12,8)  
mydata.plot(x='year',y='TSU receiving yards',kind='scatter', title='TSU Receiving Yards by Year')  
plt.show()



#get outliers of TSU receiving yards  
#There were 8 outliers on the TSU receiving yards variable, a game against Murray State University where TSU had 463 receiving yards. These outliers were replaced with the mean below.  
q1 = mydata["TSU receiving yards"].quantile(0.25)  
q3 = mydata["TSU receiving yards"].quantile(0.75)  
outlier\_top\_lim = q3 + 1.5 \* (q3 - q1)  
outlier\_bottom\_lim = q1 - 1.5 \* (q3 - q1)  
mydata["date","city","opponent","TSU receiving yards"][[mydata["TSU receiving yards"]>outlier\_top\_lim | (mydata["TSU receiving yards"]<outlier\_bottom\_lim)]]

date	city	opponent	TSU receiving yards	
137	2014-11-22	Murray	Murray State University	463

#replace outliers of TSU rush yards with mean  
mydata["TSU receiving yards\_new"] = mydata.where(mydata["TSU receiving yards"]>outlier\_top\_lim | (mydata["TSU receiving yards"]<outlier\_bottom\_lim), mydata["TSU receiving yards\_new"], mydata["TSU receiving yards"])

#descriptive statistics for TSU receiving yards  
mydata["TSU receiving yards"].describe()

count	191.000000
mean	187.769634
std	74.855646
min	-18.000000
25%	107.000000
50%	152.000000
75%	200.000000
max	463.000000
Name:	TSU receiving yards, dtype: float64

#descriptive statistics for TSU receiving yards with outliers replaced  
#Replacing the outliers with the mean lowered the mean of the variable slightly to 210 and lowered the standard deviation slightly to 80.  
mydata["TSU receiving yards\_new"].describe()

count	191.000000
mean	209.038541
std	80.295493
min	43.000000
25%	151.000000
50%	203.000000
75%	273.000000
max	395.000000
Name:	TSU receiving yards_new, dtype: float64

#TSU punt return yards by year  
#A scatter plot of TSU punt return yards showed that for most games it ranged from 0 to 55. It also showed that there were outliers that were examined below.  
plt.rcParams['figure.figsize']=(12,8)  
mydata.plot(x='year',y='TSU punt return yards',kind='scatter', title='TSU Punt Return Yards by Year')  
plt.show()



#get outliers of TSU punt return yards  
#There were 8 outliers on the TSU punt return yards variable. Each of these games had a value of at least 75 punt return yards for TSU. These were replaced with the mean below.  
q1 = mydata["TSU punt return yards"].quantile(0.25)  
q3 = mydata["TSU punt return yards"].quantile(0.75)  
outlier\_top\_lim = q3 + 1.5 \* (q3 - q1)  
outlier\_bottom\_lim = q1 - 1.5 \* (q3 - q1)  
mydata["date","city","opponent","TSU punt return yards"][[mydata["TSU punt return yards"]>outlier\_top\_lim | (mydata["TSU punt return yards"]<outlier\_bottom\_lim)]]

date	city	opponent	TSU punt return yards	
0	2003-08-30	Nashville	South Carolina State University	88
2	2003-09-13	Memphis	Jackson State University	77
19	2004-10-30	Charleston	Eastern Illinois University	107
53	2007-11-03	Nashville	Murray State University	77
82	2010-09-25	Atlanta	Florida A&M University	142
95	2011-10-08	Nashville	Southeast Missouri State University	115
116	2013-09-28	St. Louis	Central State University	75
167	2017-11-04	Nashville	Virginia University Lynchburg	94

#replace outliers of TSU punt return yards with mean  
mydata["TSU punt return yards\_new"] = mydata.where(mydata["TSU punt return yards"]>outlier\_top\_lim | (mydata["TSU punt return yards"]<outlier\_bottom\_lim), mydata["TSU punt return yards\_new"], mydata["TSU punt return yards"])







