PREDICTING FOOTBALL WINS FOR 2003 TO 2019 FOR TENNESSEE STATE UNIVERSITY USING PYTHON AND DATA FROM THE SCHOOL WEBSITES

by Erika Harrell

Disclaimer: All opinions and conclusions in this project are those of the author and not of the U.S. Department of Justice.

Executive Summary

This data science project was conducted to predict whether Tennesse State University (TSU) would win a football game and what factors are involved in winning a football game. Data on scores and other football statistics were scraped from multiple pages from the TSU athletics website to create a dataframe with information from 191 football games that TSU played from 2003 to 2019. The data scraped included the final scores from each game for TSU and their opponents; game attendance; year and location of each game; and statistics about rushing, receiving, tackles, and punts that TSU made in each game. After web scraping, all data was placed into a dataframe with each row representing a single game. After data wrangling and cleaning were completed, exploratory data anlaysis was conducted. Continuous variables had their outliers replaced by the variable mean. Bivariate correlations showed several pairs of features were at least moderately correlated (r>=0.5). Machine learning was conducted as the dataframe was split into a test data set (30% of cases) and a training dataset (70% of cases). As a result of evidence of potential multicollinearity among the features, instead of logistc regression models, two decision tree models predicting whether TSU won or loss a game were created using the training data , one created using a Gini index and another with an Entropy index. After comparing accuracy scores of the two trees using the test dataset, it was determined that the decision tree with the Entropy index was better. In that tree, it was shown that the opponent score, TSU score, TSU kick return yards, TSU receiving yards, and TSU rushing yards all played a role in determining if TSU won or loss a game.

In [1]:

```
#importing libraries for project
#python's lxml library parses xml and html files
from lxml import html
#python requests library gets data from web pages
import requests
#libraries for data wrangling & cleaning
import pandas as pd
import numpy as np
import datetime as dt
#libraries for data visualization
import matplotlib.pyplot as plt
import seaborn as sns
#libraries for machine learning
from sklearn.tree import DecisionTreeClassifier # Decision tree algorithm
from sklearn.model_selection import train_test_split # data split
from sklearn.metrics import accuracy_score # evaluation metric
from sklearn import tree
```

Web Scraping-getting data on location of games and scores from TSU websites

Data Cleaning-correcting names and deleting extra text and spaces from lists

Data Wrangling-putting data from TSU websites into lists (one list per year)

In [2]:

```
#2019 data
page = requests.get('https://tennstate_ftp.sidearmsports.com/custompages/tsutigers/E018AECE-A1B8-46E7-B1B7-46FE
mytree = html.fromstring(page.content)
schedule2019 = mytree.xpath('body//tr/td/font[@color="#000000"]/text()')
print(schedule2019)
```

```
['Nov 23, 2019\xa0\xa0', 'Cookeville, Tenn.   \xa0\xa0', 'Tennessee State 37, Tennessee Tech 27\xa0\xa0', '\xa
0', 'Nov 16, 2019\xa0\xa0', 'Martin, Tenn.       \xa0\xa0', 'UT Martin 28, Tennessee State 17\xa0\xa0', '\xa0',
'Nov 09, 2019\xa0\xa0', 'Nashville, Tenn.    \xa0\xa0', 'EIU 49, Tennessee State 38\xa0\xa0', '\xa0', 'Nov 02,
2019\xa0\xa0', 'Nashville, Tenn.    \xa0\xa0', 'SEMO 32, Tennessee State 13\xa0\xa0', '\xa0', 'Oct 19, 2019\xa0
\xa0', 'Nashville, Tenn.    \xa0\xa0', 'Tennessee State 26, Austin Peay 24\xa0\xa0', '\xa0', 'Oct 12, 2019\xa0
\xa0', 'Nashville, Tenn.    \xa0\xa0', 'Murray State 31, Tennessee State 17\xa0\xa0', '\xa0', 'Oct 05, 2019\xa0
\xa0', 'Jacksonville, Ala.  \xa0\xa0', 'Jacksonville State 31, Tennessee State 23\xa0\xa0', '\xa0', 'Sep 28, 20
19\xa0\xa0', 'Richmond, Ky.       \xa0\xa0', 'Eastern Kentucky 42, Tennessee State 16\xa0\xa0', '\xa0', 'Sep 2
1, 2019\xa0\xa0', 'Nashville, Tenn.    \xa0\xa0', 'UAPB 37, Tennessee State 31\xa0\xa0', '\xa0', 'Sep 14, 2019
\xa0\xa0', 'Memphis, Tenn.      \xa0\xa0', 'Jackson State 49, Tennessee State 44\xa0\xa0', '\xa0', 'Sep 7, 2019
\xa0\xa0', 'Murfreesboro, TN    \xa0\xa0', 'Middle Tennessee 45, Tennessee State 26\xa0\xa0', '\xa0', 'Aug 31,
2019\xa0\xa0', 'Nashville, Tenn.    \xa0\xa0', 'Tennessee State 26, Mississippi Valley 20\xa0\xa0', '\xa0']
```

In [3]:

```
#cleaning 2019 list
```

```
schedule2019=[s.replace("UAPB","Pine Bluff") for s in schedule2019]
schedule2019=[t.replace("SEMO","Southeast Missouri") for t in schedule2019]
schedule2019=[u.replace("EIU","Eastern Illinois") for u in schedule2019]
schedule2019=[v.replace("UT","Tennessee") for v in schedule2019]
schedule2019=[w.replace(".", "") for w in schedule2019]
schedule2019=[x.replace("\xa0", "") for x in schedule2019]
schedule2019=[y.replace(",","") for y in schedule2019]
schedule2019=[z.strip() for z in schedule2019]
print(schedule2019)
```

```
['Nov 23 2019', 'Cookeville Tenn', 'Tennessee State 37 Tennessee Tech 27', '', 'Nov 16 2019', 'Martin Tenn', 'T
ennessee Martin 28 Tennessee State 17', '', 'Nov 09 2019', 'Nashville Tenn', 'Eastern Illinois 49 Tennessee Sta
te 38', '', 'Nov 02 2019', 'Nashville Tenn', 'Southeast Missouri 32 Tennessee State 13', '', 'Oct 19 2019', 'Na
shville Tenn', 'Tennessee State 26 Austin Peay 24', '', 'Oct 12 2019', 'Nashville Tenn', 'Murray State 31 Tenne
ssee State 17', '', 'Oct 05 2019', 'Jacksonville Ala', 'Jacksonville State 31 Tennessee State 23', '', 'Sep 28
2019', 'Richmond Ky', 'Eastern Kentucky 42 Tennessee State 16', '', 'Sep 21 2019', 'Nashville Tenn', 'Pine Bluf
f 37 Tennessee State 31', '', 'Sep 14 2019', 'Memphis Tenn', 'Jackson State 49 Tennessee State 44', '', 'Sep 7
2019', 'Murfreesboro TN', 'Middle Tennessee 45 Tennessee State 26', '', 'Aug 31 2019', 'Nashville Tenn', 'Tenne
ssee State 26 Mississippi Valley 20', '']
```

In [4]:

```
#2018 data
page = requests.get('https://tennstate_ftp.sidearmsports.com/custompages/tsutigers/125803E8-C113-4C2F-890C-645D
mytree = html.fromstring(page.content)
schedule2018 = mytree.xpath('body//tr/td/font[@color="#000000"]/text()')
print(schedule2018)
```

```
['Nov 17, 2018\xa0\xa0', 'Nashville, Tenn.    \xa0\xa0', 'Tennessee State 31, UTM 28\xa0\xa0', '\xa0', 'Nov 10,
2018\xa0\xa0', 'Nashville, Tenn.    \xa0\xa0', 'Jacksonville State 41, Tennessee State 14\xa0\xa0', '\xa0', '11
-03-18    \xa0\xa0', 'Cape Girardeau, Mo. \xa0\xa0', 'Southeast Missouri 38, Tennessee State 21\xa0\xa0', '\xa
0', 'Oct 20, 2018\xa0\xa0', 'Nashville, TN       \xa0\xa0', 'Tennessee State 41, Tennessee Tech 14\xa0\xa0',
'\xa0', 'Oct 13, 2018\xa0\xa0', 'Murray, Ky.         \xa0\xa0', 'Murray St. 45, Tennessee State 21\xa0\xa0',
'\xa0', 'Oct 06, 2018\xa0\xa0', 'Clarksville, Tenn.  \xa0\xa0', 'Austin Peay 49, Tennessee State 34\xa0\xa0',
'\xa0', 'Sep 29, 2018\xa0\xa0', 'Nashville, Tenn.    \xa0\xa0', 'Vanderbilt 31, Tennessee State 27\xa0\xa0',
'\xa0', 'Sep 22, 2018\xa0\xa0', 'Charleston, Ill.    \xa0\xa0', 'Tennessee State 41, Eastern Illinois 40\xa0\xa
0', '\xa0', 'Sep 01, 2018\xa0\xa0', 'Nashville, Tenn.    \xa0\xa0', 'Tennessee State 34, Bethune-Cookman 3\xa0
\xa0', '\xa0']
```

In [5]:

```
#cleaning 2018 list
schedule2018=[q.replace('Vanderbilt','Vanderbilt University') for q in schedule2018]
schedule2018=[p.replace('Cape Girardeau','CapeGirardeau') for p in schedule2018]
schedule2018=[r.replace('11-03-18','Nov 03 2018') for r in schedule2018]
schedule2018=[s.replace("UTM","Tennessee Martin") for s in schedule2018]
schedule2018=[t.replace("SEMO","Southeast Missouri") for t in schedule2018]
schedule2018=[u.replace("Bethune-Cookman","Bethune Cookman") for u in schedule2018]
schedule2018=[w.replace(".", "") for w in schedule2018]
schedule2018=[x.replace("\xa0", "") for x in schedule2018]
schedule2018=[y.replace(",","") for y in schedule2018]
schedule2018=[z.strip() for z in schedule2018]
print(schedule2018)
```

```
['Nov 17 2018', 'Nashville Tenn', 'Tennessee State 31 Tennessee Martin 28', '', 'Nov 10 2018', 'Nashville Ten
n', 'Jacksonville State 41 Tennessee State 14', '', 'Nov 03 2018', 'CapeGirardeau Mo', 'Southeast Missouri 38 T
ennessee State 21', '', 'Oct 20 2018', 'Nashville TN', 'Tennessee State 41 Tennessee Tech 14', '', 'Oct 13 201
8', 'Murray Ky', 'Murray St 45 Tennessee State 21', '', 'Oct 06 2018', 'Clarksville Tenn', 'Austin Peay 49 Tenn
essee State 34', '', 'Sep 29 2018', 'Nashville Tenn', 'Vanderbilt University 31 Tennessee State 27', '', 'Sep 2
2 2018', 'Charleston Ill', 'Tennessee State 41 Eastern Illinois 40', '', 'Sep 01 2018', 'Nashville Tenn', 'Tenn
essee State 34 Bethune Cookman 3', '']
```

In [6]:

```
#2017 data
page = requests.get('https://tennstate_ftp.sidearmsports.com/custompages/tsutigers/1EEBECA9-B01E-43AB-B6EB-AFDE
mytree = html.fromstring(page.content)
schedule2017 = mytree.xpath('body//tr/td/font[@color="#000000"]/text()')
print(schedule2017)
```

```
['Nov 16, 2017\xa0\xa0', 'Jacksonville, Ala.  \xa0\xa0', 'Jacksonville State 36, Tennessee State 6\xa0\xa0',
'\xa0', 'Nov 11, 2017\xa0\xa0', 'Nashville, Tenn.    \xa0\xa0', 'Tennessee State 23, SEMO 20\xa0\xa0', '\xa0',
'Nov 04, 2017\xa0\xa0', 'Nashville, Tenn.    \xa0\xa0', 'Tennessee State 60, VUL 0\xa0\xa0', '\xa0', 'Oct 28, 2
017\xa0\xa0', 'Cookeville, Tenn.   \xa0\xa0', 'Tennessee Tech 30, Tennessee State 26\xa0\xa0', '\xa0', 'Oct 14,
2017\xa0\xa0', 'Nashville, Tenn.    \xa0\xa0', 'Austin Peay 21, Tennessee State 17\xa0\xa0', '\xa0', 'Oct 07, 2
017\xa0\xa0', 'Richmond, Ky.       \xa0\xa0', 'Tennessee State 45, Eastern Kentucky 21\xa0\xa0', '\xa0', 'Sep 3
0, 2017\xa0\xa0', 'Nashville, Tenn.    \xa0\xa0', 'Eastern Illinois 19, Tennessee State 16\xa0\xa0', '\xa0', 'S
ep 23, 2017\xa0\xa0', 'Martin, Tenn.       \xa0\xa0', 'UT Martin 31, Tennessee State 16\xa0\xa0', '\xa0', 'Sep
17, 2017\xa0\xa0', 'Tampa, Fla          \xa0\xa0', 'Tennessee State 24, Florida A&M 13\xa0\xa0', '\xa0', 'Sep 0
9, 2017\xa0\xa0', 'Memphis, Tenn.      \xa0\xa0', 'Tennessee State 17, Jackson State 15\xa0\xa0', '\xa0', 'Aug
31, 2017\xa0\xa0', 'Atlanta, Ga.        \xa0\xa0', 'Tennessee State 17, Georgia State 10\xa0\xa0', '\xa0']
```

In [7]:

```
#cleaning 2017 list
schedule2017=[t.replace("VUL", "VirginiaU Lynchburg") for t in schedule2017]
schedule2017=[u.replace("SEMO","Southeast Missouri") for u in schedule2017]
schedule2017=[v.replace("UT","Tennessee")for v in schedule2017]
```

```
schedule2017=[w.replace(".", "") for w in schedule2017]
schedule2017=[x.replace("\xa0", "") for x in schedule2017]
schedule2017=[y.replace(",","") for y in schedule2017]
schedule2017=[z.strip() for z in schedule2017]
print(schedule2017)
```

```
['Nov 16 2017', 'Jacksonville Ala', 'Jacksonville State 36 Tennessee State 6', '', 'Nov 11 2017', 'Nashville Te
nn', 'Tennessee State 23 Southeast Missouri 20', '', 'Nov 04 2017', 'Nashville Tenn', 'Tennessee State 60 Virgi
niaU Lynchburg 0', '', 'Oct 28 2017', 'Cookeville Tenn', 'Tennessee Tech 30 Tennessee State 26', '', 'Oct 14 20
17', 'Nashville Tenn', 'Austin Peay 21 Tennessee State 17', '', 'Oct 07 2017', 'Richmond Ky', 'Tennessee State
45 Eastern Kentucky 21', '', 'Sep 30 2017', 'Nashville Tenn', 'Eastern Illinois 19 Tennessee State 16', '', 'Se
p 23 2017', 'Martin Tenn', 'Tennessee Martin 31 Tennessee State 16', '', 'Sep 17 2017', 'Tampa Fla', 'Tennessee
State 24 Florida A&M 13', '', 'Sep 09 2017', 'Memphis Tenn', 'Tennessee State 17 Jackson State 15', '', 'Aug 31
2017', 'Atlanta Ga', 'Tennessee State 17 Georgia State 10', '']
```

In [8]:

```
#2016 data
page = requests.get('https://tennstate_ftp.sidearmsports.com/custompages/tsutigers/F8A9B17C-4555-45FB-AAF3-D889
mytree = html.fromstring(page.content)
schedule2016 = mytree.xpath('body//tr/td/font[@color="#000000"]/text()')
print(schedule2016)
```

```
['11-19-16    \xa0\xa0', 'Cape Girardeau, Mo. \xa0\xa0', 'Tennessee State 32, Southeast Missouri 31\xa0\xa0',
'\xa0', 'Nov 12, 2016\xa0\xa0', 'Nashville, TN       \xa0\xa0', 'Tennessee Tech 44, Tennessee State 16\xa0\xa
0', '\xa0', 'Nov 05, 2016\xa0\xa0', 'Clarksville, Tenn.  \xa0\xa0', 'Tennessee State 41, Austin Peay 40\xa0\xa
0', '\xa0', 'Oct 29, 2016\xa0\xa0', 'Murray, Ky.         \xa0\xa0', 'Murray St. 38, Tennessee State 31\xa0\xa
0', '\xa0', 'Oct 22, 2016\xa0\xa0', 'Nashville, Tenn.    \xa0\xa0', 'Vanderbilt 35, Tennessee State 17\xa0\xa
0', '\xa0', 'Oct 15, 2016\xa0\xa0', 'Nashville, Tenn.    \xa0\xa0', 'Tennessee State 35, Eastern Kentucky 28\xa
0\xa0', '\xa0', 'Oct 08, 2016\xa0\xa0', 'Charleston, Ill.    \xa0\xa0', 'Eastern Illinois 35, Tennessee State 3
4\xa0\xa0', '\xa0', 'Oct 01, 2016\xa0\xa0', 'Nashville, Tenn.    \xa0\xa0', 'Tennessee State 34, UT Martin 30\x
a0\xa0', '\xa0', 'Sep 17, 2016\xa0\xa0', 'Daytona Beach, Fla. \xa0\xa0', 'Tennessee State 31, Bethune-Cookman 2
4\xa0\xa0', '\xa0', 'Sep 10, 2016\xa0\xa0', 'Memphis, Tenn.      \xa0\xa0', 'Tennessee State 40, Jackson State
26\xa0\xa0', '\xa0', 'Sep 03, 2016\xa0\xa0', 'Nashville, Tenn.    \xa0\xa0', 'Tennessee State 44, Arkansas-Pine
Bluff 0\xa0\xa0', '\xa0']
```

In [9]:

```
#cleaning 2016 data
schedule2016=[o.replace('Pine Bluff',"PineBluff") for o in schedule2016]
schedule2016=[p.replace('11-19-16','Nov 19 2016') for p in schedule2016]
schedule2016=[r.replace('Cape Girardeau','CapeGirardeau') for r in schedule2016]
schedule2016=[s.replace("Vanderbilt","Vanderbilt University") for s in schedule2016]
schedule2016=[t.replace("UT","Tennessee") for t in schedule2016]
schedule2016=[u.replace("Daytona Beach", "DaytonaBeach") for u in schedule2016]
schedule2016=[v.replace("-"," ") for v in schedule2016]
schedule2016=[w.replace(".", "") for w in schedule2016]
schedule2016=[x.replace("\xa0", "") for x in schedule2016]
schedule2016=[y.replace(",","") for y in schedule2016]
schedule2016=[z.strip() for z in schedule2016]
print(schedule2016)
```

```
['Nov 19 2016', 'CapeGirardeau Mo', 'Tennessee State 32 Southeast Missouri 31', '', 'Nov 12 2016', 'Nashville T
N', 'Tennessee Tech 44 Tennessee State 16', '', 'Nov 05 2016', 'Clarksville Tenn', 'Tennessee State 41 Austin P
eay 40', '', 'Oct 29 2016', 'Murray Ky', 'Murray St 38 Tennessee State 31', '', 'Oct 22 2016', 'Nashville Ten
n', 'Vanderbilt University 35 Tennessee State 17', '', 'Oct 15 2016', 'Nashville Tenn', 'Tennessee State 35 Eas
tern Kentucky 28', '', 'Oct 08 2016', 'Charleston Ill', 'Eastern Illinois 35 Tennessee State 34', '', 'Oct 01 2
016', 'Nashville Tenn', 'Tennessee State 34 Tennessee Martin 30', '', 'Sep 17 2016', 'DaytonaBeach Fla', 'Tenne
ssee State 31 Bethune Cookman 24', '', 'Sep 10 2016', 'Memphis Tenn', 'Tennessee State 40 Jackson State 26',
'', 'Sep 03 2016', 'Nashville Tenn', 'Tennessee State 44 Arkansas PineBluff 0', '']
```

In [10]:

```
#2015 data
page = requests.get('https://tennstate_ftp.sidearmsports.com/custompages/tsutigers/B8713968-6888-4BB4-B12E-9BA4
mytree = html.fromstring(page.content)
schedule2015 = mytree.xpath('body//tr/td/font[@color="#000000"]/text()')
print(schedule2015)
```

```
['Nov 21, 2015\xa0\xa0', 'Cookeville, Tenn.   \xa0\xa0', 'Tennessee Tech 30, Tennessee State 24\xa0\xa0', '\xa
0', 'Nov 07, 2015\xa0\xa0', 'Nashville, Tenn.    \xa0\xa0', 'Murray State 46, Tennessee State 43\xa0\xa0', '\xa
0', 'Oct 31, 2015\xa0\xa0', 'Nashville, Tenn.    \xa0\xa0', 'Tennessee State 20, Austin Peay 6\xa0\xa0', '\xa
0', 'Oct 24, 2015\xa0\xa0', 'Richmond, Ky.       \xa0\xa0', 'Eastern Kentucky 45, Tennessee State 21\xa0\xa0',
'\xa0', 'Oct 17, 2015\xa0\xa0', 'Nashville, Tenn.    \xa0\xa0', 'Eastern Illinois 25, Tennessee State 22\xa0\xa
0', '\xa0', 'Oct 10, 2015\xa0\xa0', 'Martin, Tenn.       \xa0\xa0', 'UT Martin 28, Tennessee State 14\xa0\xa0',
'\xa0', 'Sep 26, 2015\xa0\xa0', 'Tallahassee, Fla.   \xa0\xa0', 'Tennessee State 30, FAMU 14\xa0\xa0', '\xa0',
'Sep 19, 2015\xa0\xa0', 'Jacksonville, Ala.  \xa0\xa0', 'Jacksonville State 48, Tennessee State 13\xa0\xa0',
'\xa0', 'Sep 12, 2015\xa0\xa0', 'Memphis, TN         \xa0\xa0', 'Tennessee State 35, Jackson State Tigers 25\xa
0\xa0', '\xa0', 'Sep 06, 2015\xa0\xa0', 'Nashville, Tenn.    \xa0\xa0', 'Tennessee State 24, Alabama State 14\x
a0\xa0', '\xa0']
```

In [11]:

```
#cleaning 2015 list
schedule2015=[s.replace("Tigers","") for s in schedule2015]
schedule2015=[t.replace("UT","Tennessee") for t in schedule2015]
schedule2015=[u.replace("FAMU","Florida A&M") for u in schedule2015]
schedule2015=[v.replace("-"," ") for v in schedule2015]
schedule2015=[w.replace(".", "") for w in schedule2015]
schedule2015=[x.replace("\xa0", "") for x in schedule2015]
```

```
schedule2015=[y.replace(",","") for y in schedule2015]
schedule2015=[z.strip() for z in schedule2015]
print(schedule2015)
```

```
['Nov 21 2015', 'Cookeville Tenn', 'Tennessee Tech 30 Tennessee State 24', '', 'Nov 07 2015', 'Nashville Tenn',
'Murray State 46 Tennessee State 43', '', 'Oct 31 2015', 'Nashville Tenn', 'Tennessee State 20 Austin Peay 6',
'', 'Oct 24 2015', 'Richmond Ky', 'Eastern Kentucky 45 Tennessee State 21', '', 'Oct 17 2015', 'Nashville Ten
n', 'Eastern Illinois 25 Tennessee State 22', '', 'Oct 10 2015', 'Martin Tenn', 'Tennessee Martin 28 Tennessee
State 14', '', 'Sep 26 2015', 'Tallahassee Fla', 'Tennessee State 30 Florida A&M 14', '', 'Sep 19 2015', 'Jacks
onville Ala', 'Jacksonville State 48 Tennessee State 13', '', 'Sep 12 2015', 'Memphis TN', 'Tennessee State 35
Jackson State  25', '', 'Sep 06 2015', 'Nashville Tenn', 'Tennessee State 24 Alabama State 14', '']
```

In [12]:

```
#2014 data
page = requests.get('https://tennstate_ftp.sidearmsports.com/custompages/tsutigers/26C45AE2-D036-475D-A4EF-38BE
mytree = html.fromstring(page.content)
schedule2014 = mytree.xpath('body//tr/td/font[@color="#000000"]/text()')
print(schedule2014)
```

```
['Nov 22, 2014\xa0\xa0', 'Murray, Ky.         \xa0\xa0', 'Tennessee State 48, Murray St. 33\xa0\xa0', '\xa0',
'Nov 08, 2014\xa0\xa0', 'Clarksville, Tenn.  \xa0\xa0', 'Tennessee State 31, Austin Peay 27\xa0\xa0', '\xa0',
'Nov 01, 2014\xa0\xa0', 'Nashville, Tenn.    \xa0\xa0', 'Eastern Kentucky 56, Tennessee State 42\xa0\xa0', '\xa
0', 'Oct 25, 2014\xa0\xa0', 'Charleston, Ill.    \xa0\xa0', 'Eastern Illinois 28, Tennessee State 3\xa0\xa0',
'\xa0', 'Oct 18, 2014\xa0\xa0', 'Hale Stadium        \xa0\xa0', 'UT Martin 21, Tennessee State 16\xa0\xa0', '\x
a0', 'Oct 11, 2014\xa0\xa0', 'Nashville, Tenn.    \xa0\xa0', 'Jacksonville State 27, Tennessee State 20\xa0\xa
0', '\xa0', '10-04-14    \xa0\xa0', 'Cape Girardeau, Mo. \xa0\xa0', 'Southeast Missouri 28, Tennessee State 21
\xa0\xa0', '\xa0', 'Sep 27, 2014\xa0\xa0', 'Nashville, Tenn.    \xa0\xa0', 'Tennessee State 27, Florida A&M 7\x
a0\xa0', '\xa0', 'Sep 20, 2014\xa0\xa0', 'Nashville, TN       \xa0\xa0', 'Tennessee State 10, Tennessee Tech 7
\xa0\xa0', '\xa0', 'Sep 13, 2014\xa0\xa0', 'Memphis, Tenn.      \xa0\xa0', 'Tennessee State 35, Jackson State 7
\xa0\xa0', '\xa0', 'Sep 06, 2014\xa0\xa0', 'Montgomery, Ala.    \xa0\xa0', 'Alabama State 27, Tennessee State 2
1\xa0\xa0', '\xa0', 'Aug 30, 2014\xa0\xa0', 'Nashville, Tenn.    \xa0\xa0', 'Tennessee State 58, Edward Waters
6\xa0\xa0', '\xa0']
```

In [13]:

```
#cleaning 2014 list
schedule2014=[s.replace("Hale Stadium","Nashville, Tenn.") for s in schedule2014]
schedule2014=[t.replace("10-04-14","Oct 04, 2014") for t in schedule2014]
schedule2014=[u.replace("Cape Girardeau","CapeGirardeau") for u in schedule2014]
schedule2014=[v.replace("-"," ") for v in schedule2014]
schedule2014=[w.replace(".", "") for w in schedule2014]
schedule2014=[x.replace("\xa0", "") for x in schedule2014]
schedule2014=[y.replace(",","") for y in schedule2014]
schedule2014=[z.strip() for z in schedule2014]
print(schedule2014)
```

```
['Nov 22 2014', 'Murray Ky', 'Tennessee State 48 Murray St 33', '', 'Nov 08 2014', 'Clarksville Tenn', 'Tenness
ee State 31 Austin Peay 27', '', 'Nov 01 2014', 'Nashville Tenn', 'Eastern Kentucky 56 Tennessee State 42', '',
'Oct 25 2014', 'Charleston Ill', 'Eastern Illinois 28 Tennessee State 3', '', 'Oct 18 2014', 'Nashville Tenn',
'UT Martin 21 Tennessee State 16', '', 'Oct 11 2014', 'Nashville Tenn', 'Jacksonville State 27 Tennessee State
20', '', 'Oct 04 2014', 'CapeGirardeau Mo', 'Southeast Missouri 28 Tennessee State 21', '', 'Sep 27 2014', 'Nas
hville Tenn', 'Tennessee State 27 Florida A&M 7', '', 'Sep 20 2014', 'Nashville TN', 'Tennessee State 10 Tennes
see Tech 7', '', 'Sep 13 2014', 'Memphis Tenn', 'Tennessee State 35 Jackson State 7', '', 'Sep 06 2014', 'Montg
omery Ala', 'Alabama State 27 Tennessee State 21', '', 'Aug 30 2014', 'Nashville Tenn', 'Tennessee State 58 Edw
ard Waters 6', '']
```

In [14]:

```
#2013 data
page = requests.get('https://tennstate_ftp.sidearmsports.com/custompages/tsutigers/E394BBB1-387B-4C6A-9B21-E33D
mytree = html.fromstring(page.content)
schedule2013 = mytree.xpath('body//tr/td/font[@color="#000000"]/text()')
print(schedule2013)
```

```
['Dec 07, 2013\xa0\xa0', 'Charleston, Ill.    \xa0\xa0', 'Eastern Illinois 51, Tennessee State 10\xa0\xa0', '\x
a0', 'Nov 30, 2013\xa0\xa0', 'Indianapolis, Ind.  \xa0\xa0', 'Tennessee State 31, Butler 0\xa0\xa0', '\xa0', 'N
ov 09, 2013\xa0\xa0', 'Nashville, Tenn.    \xa0\xa0', 'Tennessee State 31, Austin Peay 6\xa0\xa0', '\xa0', 'Nov
16, 2013\xa0\xa0', 'Nashville, Tenn.    \xa0\xa0', 'Tennessee State 17, Murray State 10\xa0\xa0', '\xa0', 'Nov
02, 2013\xa0\xa0', 'Richmond, Ky.       \xa0\xa0', 'Eastern Kentucky 44, Tennessee State 0\xa0\xa0', '\xa0', 'O
ct 26, 2013\xa0\xa0', 'Nashville, Tenn.    \xa0\xa0', 'Eastern Illinois 34, Tennessee State 16\xa0\xa0', '\xa
0', 'Oct 19, 2013\xa0\xa0', 'Martin, Tenn.       \xa0\xa0', 'Tennessee State 29, UT Martin 15\xa0\xa0', '\xa0',
'Oct 12, 2013\xa0\xa0', 'Jacksonville, Ala.  \xa0\xa0', 'Tennessee State 31, Jacksonville State 15\xa0\xa0',
'\xa0', 'Oct 05, 2013\xa0\xa0', 'Nashville, Tenn.    \xa0\xa0', 'Tennessee State 40, Southeast Missouri 16\xa0
\xa0', '\xa0', 'Sep 28, 2013\xa0\xa0', 'St. Louis, Missouri \xa0\xa0', 'Tennessee State 73, Central State 6\xa0
\xa0', '\xa0', 'Sep 21, 2013\xa0\xa0', 'Cookeville, Tenn.   \xa0\xa0', 'Tennessee State 41, Tennessee Tech 21\x
a0\xa0', '\xa0', 'Sep 14, 2013\xa0\xa0', 'Memphis, Tenn.      \xa0\xa0', 'Tennessee State 26, Jackson State Tig
ers 16\xa0\xa0', '\xa0', 'Sep 07, 2013\xa0\xa0', 'Tallahassee, Fla.   \xa0\xa0', 'Tennessee State 27, Florida A
&M 7\xa0\xa0', '\xa0', 'Sep 01, 2013\xa0\xa0', 'Nashville, Tenn.    \xa0\xa0', 'Bethune-Cookman 12, Tennessee S
tate 9\xa0\xa0', '\xa0']
```

In [15]:

```
#cleaning 2013 list
schedule2013=[r.replace("St. Louis","StLouis") for r in schedule2013]
schedule2013=[s.replace("Butler","Butler University") for s in schedule2013]
schedule2013=[t.replace("UT","Tennessee") for t in schedule2013]
schedule2013=[u.replace("Tigers","") for u in schedule2013]
schedule2013=[v.replace("-"," ") for v in schedule2013]
```

```
schedule2013=[w.replace(".", "") for w in schedule2013]
schedule2013=[x.replace("\xa0", "") for x in schedule2013]
schedule2013=[y.replace(",","") for y in schedule2013]
schedule2013=[z.strip() for z in schedule2013]
print(schedule2013)
```

```
['Dec 07 2013', 'Charleston Ill', 'Eastern Illinois 51 Tennessee State 10', '', 'Nov 30 2013', 'Indianapolis In
d', 'Tennessee State 31 Butler University 0', '', 'Nov 09 2013', 'Nashville Tenn', 'Tennessee State 31 Austin P
eay 6', '', 'Nov 16 2013', 'Nashville Tenn', 'Tennessee State 17 Murray State 10', '', 'Nov 02 2013', 'Richmond
Ky', 'Eastern Kentucky 44 Tennessee State 0', '', 'Oct 26 2013', 'Nashville Tenn', 'Eastern Illinois 34 Tenness
ee State 16', '', 'Oct 19 2013', 'Martin Tenn', 'Tennessee State 29 Tennessee Martin 15', '', 'Oct 12 2013', 'J
acksonville Ala', 'Tennessee State 31 Jacksonville State 15', '', 'Oct 05 2013', 'Nashville Tenn', 'Tennessee S
tate 40 Southeast Missouri 16', '', 'Sep 28 2013', 'StLouis Missouri', 'Tennessee State 73 Central State 6',
'', 'Sep 21 2013', 'Cookeville Tenn', 'Tennessee State 41 Tennessee Tech 21', '', 'Sep 14 2013', 'Memphis Ten
n', 'Tennessee State 26 Jackson State  16', '', 'Sep 07 2013', 'Tallahassee Fla', 'Tennessee State 27 Florida A
&M 7', '', 'Sep 01 2013', 'Nashville Tenn', 'Bethune Cookman 12 Tennessee State 9', '']
```

In [16]:

```
#2012 data
page = requests.get('https://tennstate_ftp.sidearmsports.com/custompages/tsutigers/BDF8C4F3-D7B6-4D50-847F-9227
mytree = html.fromstring(page.content)
schedule2012 = mytree.xpath('body//tr/td/font[@color="#000000"]/text()')
print(schedule2012)
```

```
['Nov 17, 2012\xa0\xa0', 'Martin, Tenn.       \xa0\xa0', 'UT Martin 35, Tennessee State 26\xa0\xa0', '\xa0', 'N
ov 03, 2012\xa0\xa0', 'Murray, Ky.         \xa0\xa0', 'Murray State 49, Tennessee State 28\xa0\xa0', '\xa0', 'O
ct 27, 2012\xa0\xa0', 'Nashville, Tennessee\xa0\xa0', 'Tennessee State 22, Tennessee Tech 21\xa0\xa0', '\xa0',
'Oct 20, 2012\xa0\xa0', 'Jacksonville, Ala.  \xa0\xa0', 'Jacksonville State 31, Tennessee State 28\xa0\xa0',
'\xa0', 'Oct 13, 2012\xa0\xa0', 'Cape Girardeau, Mo. \xa0\xa0', 'Tennessee State 40, Southeast Missouri 28\xa0
\xa0', '\xa0', 'Oct 05, 2012\xa0\xa0', 'Nashville, Tennessee\xa0\xa0', 'Tennessee State 23, Eastern Kentucky 20
\xa0\xa0', '\xa0', 'Sep 29, 2012\xa0\xa0', 'Nashville, Tennessee\xa0\xa0', 'Tennessee State 40, Arkansas Pine B
luff 13\xa0\xa0', '\xa0', 'Sep 22, 2012\xa0\xa0', 'Daytona Beach, Fla. \xa0\xa0', 'Tennessee State 21, Bethune-
Cookman 14\xa0\xa0', '\xa0', 'Sep 15, 2012\xa0\xa0', 'Nashville, Tennessee\xa0\xa0', 'Tennessee State 34, Austi
n Peay 14\xa0\xa0', '\xa0', 'Sep 08, 2012\xa0\xa0', 'Memphis. Tennessee  \xa0\xa0', 'Tennessee State 38, Jackso
n State 12\xa0\xa0', '\xa0', 'Sep 01, 2012\xa0\xa0', 'Nashville, Tennessee\xa0\xa0', 'Tennessee State 17, Flori
da A&M 14\xa0\xa0', '\xa0']
```

In [17]:

```
#cleaning 2012 list
schedule2012=[r.replace("Cape Girardeau","CapeGirardeau") for r in schedule2012]
schedule2012=[s.replace("Daytona Beach","DaytonaBeach") for s in schedule2012]
schedule2012=[t.replace("UT","Tennessee") for t in schedule2012]
schedule2012=[u.replace("Pine Bluff","PineBluff") for u in schedule2012]
schedule2012=[v.replace("-"," ") for v in schedule2012]
schedule2012=[w.replace(".", "") for w in schedule2012]
schedule2012=[x.replace("\xa0", "") for x in schedule2012]
schedule2012=[y.replace(",","") for y in schedule2012]
schedule2012=[z.strip() for z in schedule2012]
print(schedule2012)
```

```
['Nov 17 2012', 'Martin Tenn', 'Tennessee Martin 35 Tennessee State 26', '', 'Nov 03 2012', 'Murray Ky', 'Murra
y State 49 Tennessee State 28', '', 'Oct 27 2012', 'Nashville Tennessee', 'Tennessee State 22 Tennessee Tech 2
1', '', 'Oct 20 2012', 'Jacksonville Ala', 'Jacksonville State 31 Tennessee State 28', '', 'Oct 13 2012', 'Cape
Girardeau Mo', 'Tennessee State 40 Southeast Missouri 28', '', 'Oct 05 2012', 'Nashville Tennessee', 'Tennessee
State 23 Eastern Kentucky 20', '', 'Sep 29 2012', 'Nashville Tennessee', 'Tennessee State 40 Arkansas PineBluff
13', '', 'Sep 22 2012', 'DaytonaBeach Fla', 'Tennessee State 21 Bethune Cookman 14', '', 'Sep 15 2012', 'Nashvi
lle Tennessee', 'Tennessee State 34 Austin Peay 14', '', 'Sep 08 2012', 'Memphis Tennessee', 'Tennessee State 3
8 Jackson State 12', '', 'Sep 01 2012', 'Nashville Tennessee', 'Tennessee State 17 Florida A&M 14', '']
```

In [18]:

```
#2011 data
page = requests.get('https://tennstate_ftp.sidearmsports.com/custompages/tsutigers/CA835442-6528-4937-854B-D964
mytree = html.fromstring(page.content)
schedule2011 = mytree.xpath('body//tr/td/font[@color="#000000"]/text()')
print(schedule2011)
```

```
['Nov 19, 2011\xa0\xa0', 'Nashville, TN       \xa0\xa0', 'Jacksonville State 38, Tennessee State 16\xa0\xa0',
'\xa0', 'Nov 12, 2011\xa0\xa0', 'Nashville, Tenn.    \xa0\xa0', 'Tennessee State 35, UT Martin 30\xa0\xa0', '\x
a0', 'Nov 05, 2011\xa0\xa0', 'Charleston, Ill.    \xa0\xa0', 'Tennessee State 18, Eastern Illinois 17\xa0\xa0',
'\xa0', 'Oct 22, 2011\xa0\xa0', 'Richmond, Ky.       \xa0\xa0', 'Eastern Kentucky 33, Tennessee State 22\xa0\xa
0', '\xa0', 'Oct 15, 2011\xa0\xa0', 'Cookeville, Tenn.   \xa0\xa0', 'Tennessee State 42, Tennessee Tech 40\xa0
\xa0', '\xa0', 'Oct 08, 2011\xa0\xa0', 'Nashville, Tenn.    \xa0\xa0', 'Tennessee State 55, Southeast Missouri
3\xa0\xa0', '\xa0', 'Oct 01, 2011\xa0\xa0', 'Clarksville, Tenn.  \xa0\xa0', 'Austin Peay 37, Tennessee State 34
\xa0\xa0', '\xa0', 'Sep 24, 2011\xa0\xa0', 'USAFA, Colo.        \xa0\xa0', 'Air Force 63, Tennessee State 24\xa
0\xa0', '\xa0', 'Sep 17, 2011\xa0\xa0', 'Murray, Ky.         \xa0\xa0', 'Murray State 58, Tennessee State 27\xa
0\xa0', '\xa0', 'Sep 10, 2011\xa0\xa0', 'Memphis, TN         \xa0\xa0', 'Jackson State 35, Tennessee State 29\x
a0\xa0', '\xa0', 'Sep 03, 2011\xa0\xa0', 'Nashville, TN       \xa0\xa0', 'Tennessee State 33, Southern U. 7\xa0
\xa0', '\xa0']
```

In [19]:

```
#cleaning 2011 list
schedule2011=[r.replace("USAFA","AirForceAcademy") for r in schedule2011]
schedule2011=[s.replace("UT","Tennessee") for s in schedule2011]
schedule2011=[t.replace("Southern U.","Southern University") for t in schedule2011]
schedule2011=[v.replace("-"," ") for v in schedule2011]
```

```
schedule2011=[w.replace(".", "") for w in schedule2011]
schedule2011=[x.replace("\xa0", "") for x in schedule2011]
schedule2011=[y.replace(",","") for y in schedule2011]
schedule2011=[z.strip() for z in schedule2011]
print(schedule2011)
```

```
['Nov 19 2011', 'Nashville TN', 'Jacksonville State 38 Tennessee State 16', '', 'Nov 12 2011', 'Nashville Ten
n', 'Tennessee State 35 Tennessee Martin 30', '', 'Nov 05 2011', 'Charleston Ill', 'Tennessee State 18 Eastern
Illinois 17', '', 'Oct 22 2011', 'Richmond Ky', 'Eastern Kentucky 33 Tennessee State 22', '', 'Oct 15 2011', 'C
ookeville Tenn', 'Tennessee State 42 Tennessee Tech 40', '', 'Oct 08 2011', 'Nashville Tenn', 'Tennessee State
55 Southeast Missouri 3', '', 'Oct 01 2011', 'Clarksville Tenn', 'Austin Peay 37 Tennessee State 34', '', 'Sep
24 2011', 'AirForceAcademy Colo', 'Air Force 63 Tennessee State 24', '', 'Sep 17 2011', 'Murray Ky', 'Murray St
ate 58 Tennessee State 27', '', 'Sep 10 2011', 'Memphis TN', 'Jackson State 35 Tennessee State 29', '', 'Sep 03
2011', 'Nashville TN', 'Tennessee State 33 Southern University 7', '']
```

In [20]:

```
#2010 data
page = requests.get('https://tennstate_ftp.sidearmsports.com/custompages/tsutigers/108733F9-E3AE-4C3D-95EE-76C0
mytree = html.fromstring(page.content)
schedule2010 = mytree.xpath('body//tr/td/font[@color="#000000"]/text()')
print(schedule2010)
```

```
['Nov 20, 2010\xa0\xa0', 'Murray, Ky.         \xa0\xa0', 'Murray State 28, Tennessee State 23\xa0\xa0', '\xa0',
'Nov 13, 2010\xa0\xa0', 'Martin, Tenn.       \xa0\xa0', 'UT Martin 37, Tennessee State 0\xa0\xa0', '\xa0', 'Nov
06, 2010\xa0\xa0', 'Nashville, TN       \xa0\xa0', 'Eastern Illinois 31, Tennessee State 28\xa0\xa0', '\xa0',
'Oct 23, 2010\xa0\xa0', 'Nashville, TN       \xa0\xa0', 'Tennessee Tech 21, Tennessee State 10\xa0\xa0', '\xa
0', 'Oct 16, 2010\xa0\xa0', 'Jacksonville, Ala.  \xa0\xa0', 'Jacksonville State 24, Tennessee State 0\xa0\xa0',
'\xa0', 'Oct. 9, 2010\xa0\xa0', 'Cape Girardeau, Mo. \xa0\xa0', 'Southeast Missouri 19, Tennessee State 17\xa0
\xa0', '\xa0', 'Oct. 2, 2010\xa0\xa0', 'Indianapolis, Ind.  \xa0\xa0', 'Tennessee State 37, North Carolina A&T
7\xa0\xa0', '\xa0', 'Sep 25, 2010\xa0\xa0', 'Atlanta, GA         \xa0\xa0', 'Tennessee State 29, Florida A&M 18
\xa0\xa0', '\xa0', 'Sep 18, 2010\xa0\xa0', 'Nashville, TN       \xa0\xa0', 'Austin Peay 26, Tennessee State 23
\xa0\xa0', '\xa0', 'Sep 11, 2010\xa0\xa0', 'Memphis, TN         \xa0\xa0', 'Jackson State 33, Tennessee State 2
6\xa0\xa0', '\xa0', 'Sep 04, 2010\xa0\xa0', 'Nashville, TN       \xa0\xa0', 'Tennessee State 27, Alabama A&M 14
\xa0\xa0', '\xa0']
```

In [21]:

```
#cleaning 2010 list
schedule2010=[r.replace("Oct. 9, 2010","Oct. 09, 2010") for r in schedule2010]
schedule2010=[r.replace("Oct. 2, 2010","Oct. 02, 2010") for r in schedule2010]
schedule2010=[r.replace("Cape Girardeau","CapeGirardeau") for r in schedule2010]
schedule2010=[s.replace("UT","Tennessee") for s in schedule2010]
schedule2010=[t.replace("North Carolina ","NorthCarolina ") for t in schedule2010]
schedule2010=[v.replace("-"," ") for v in schedule2010]
schedule2010=[w.replace(".", "") for w in schedule2010]
schedule2010=[x.replace("\xa0", "") for x in schedule2010]
schedule2010=[y.replace(",","") for y in schedule2010]
schedule2010=[z.strip() for z in schedule2010]
print(schedule2010)
```

```
['Nov 20 2010', 'Murray Ky', 'Murray State 28 Tennessee State 23', '', 'Nov 13 2010', 'Martin Tenn', 'Tennessee
Martin 37 Tennessee State 0', '', 'Nov 06 2010', 'Nashville TN', 'Eastern Illinois 31 Tennessee State 28', '',
'Oct 23 2010', 'Nashville TN', 'Tennessee Tech 21 Tennessee State 10', '', 'Oct 16 2010', 'Jacksonville Ala',
'Jacksonville State 24 Tennessee State 0', '', 'Oct 09 2010', 'CapeGirardeau Mo', 'Southeast Missouri 19 Tennes
see State 17', '', 'Oct 02 2010', 'Indianapolis Ind', 'Tennessee State 37 NorthCarolina A&T 7', '', 'Sep 25 201
0', 'Atlanta GA', 'Tennessee State 29 Florida A&M 18', '', 'Sep 18 2010', 'Nashville TN', 'Austin Peay 26 Tenne
ssee State 23', '', 'Sep 11 2010', 'Memphis TN', 'Jackson State 33 Tennessee State 26', '', 'Sep 04 2010', 'Nas
hville TN', 'Tennessee State 27 Alabama A&M 14', '']
```

In [22]:

```
#2009 data
page = requests.get('https://tennstate_ftp.sidearmsports.com/custompages/tsutigers/DEDBD687-051A-4E37-98F7-ACE4
mytree = html.fromstring(page.content)
schedule2009 = mytree.xpath('body//tr/td/font[@color="#000000"]/text()')
print(schedule2009)
```

```
['Nov 19, 2009\xa0\xa0', 'Charleston, Ill.    \xa0\xa0', 'Tennessee State 21, Eastern Illinois 10\xa0\xa0', '\x
a0', 'Nov 14, 2009\xa0\xa0', 'Clarksville, Tenn.  \xa0\xa0', 'Austin Peay 24, Tennessee State 21\xa0\xa0', '\xa
0', 'Nov 07, 2009\xa0\xa0', 'Nashville, TN       \xa0\xa0', 'Tennessee Martin 28, Tennessee State 7\xa0\xa0',
'\xa0', 'Oct 31, 2009\xa0\xa0', 'Cookeville, Tenn.   \xa0\xa0', 'Tennessee Tech 20, Tennessee State 13\xa0\xa
0', '\xa0', 'Oct 17, 2009\xa0\xa0', 'Nashville, TN       \xa0\xa0', 'Murray State 9, Tennessee State 6\xa0\xa
0', '\xa0', 'Oct 10, 2009\xa0\xa0', 'Richmond, Ky.       \xa0\xa0', 'Tennessee State 20, Eastern Kentucky 17\xa
0\xa0', '\xa0', 'Oct 03, 2009\xa0\xa0', 'Nashville, TN       \xa0\xa0', 'Tennessee State 23, Southeast Missouri
17\xa0\xa0', '\xa0', 'Sep 26, 2009\xa0\xa0', 'Atlanta, Ga.        \xa0\xa0', '#25 Florida A&M 31, Tennessee Sta
te 12\xa0\xa0', '\xa0', 'Sep 19, 2009\xa0\xa0', 'Baton Rouge, LA     \xa0\xa0', 'Southern University 21, Tennes
see State 17\xa0\xa0', '\xa0', 'Sep 12, 2009\xa0\xa0', 'Memphis, TN         \xa0\xa0', 'Tennessee State 14, Jac
kson State Tigers 7\xa0\xa0', '\xa0', 'Sep 05, 2009\xa0\xa0', 'Nashville, TN       \xa0\xa0', 'Alabama A&M 24,
Tennessee State 7\xa0\xa0', '\xa0']
```

In [23]:

```
#cleaning 2009 list
schedule2009=[r.replace("#25","") for r in schedule2009]
schedule2009=[s.replace("Tigers","") for s in schedule2009]
schedule2009=[t.replace("Baton Rouge","BatonRouge") for t in schedule2009]
schedule2009=[v.replace("-"," ") for v in schedule2009]
schedule2009=[w.replace(".", "") for w in schedule2009]
schedule2009=[x.replace("\xa0", "") for x in schedule2009]
```

```
schedule2009=[y.replace(",","") for y in schedule2009]
schedule2009=[z.strip() for z in schedule2009]
print(schedule2009)
```

```
['Nov 19 2009', 'Charleston Ill', 'Tennessee State 21 Eastern Illinois 10', '', 'Nov 14 2009', 'Clarksville Ten
n', 'Austin Peay 24 Tennessee State 21', '', 'Nov 07 2009', 'Nashville TN', 'Tennessee Martin 28 Tennessee Stat
e 7', '', 'Oct 31 2009', 'Cookeville Tenn', 'Tennessee Tech 20 Tennessee State 13', '', 'Oct 17 2009', 'Nashvil
le TN', 'Murray State 9 Tennessee State 6', '', 'Oct 10 2009', 'Richmond Ky', 'Tennessee State 20 Eastern Kentu
cky 17', '', 'Oct 03 2009', 'Nashville TN', 'Tennessee State 23 Southeast Missouri 17', '', 'Sep 26 2009', 'Atl
anta Ga', 'Florida A&M 31 Tennessee State 12', '', 'Sep 19 2009', 'BatonRouge LA', 'Southern University 21 Tenn
essee State 17', '', 'Sep 12 2009', 'Memphis TN', 'Tennessee State 14 Jackson State  7', '', 'Sep 05 2009', 'Na
shville TN', 'Alabama A&M 24 Tennessee State 7', '']
```

In [24]:

```
#2008 data
page = requests.get('https://tennstate_ftp.sidearmsports.com/custompages/tsutigers/E0564A83-37A0-4B85-8811-B7EE
mytree = html.fromstring(page.content)
schedule2008 = mytree.xpath('body//tr/td/font[@color="#000000"]/text()')
print(schedule2008)
```

```
['Nov 22, 2008\xa0\xa0', 'Murray, Ky.         \xa0\xa0', 'Murray State 24, Tennessee State 17\xa0\xa0', '\xa0',
'Nov 15, 2008\xa0\xa0', 'Jacksonville, Ala.  \xa0\xa0', 'Jacksonville State 26, Tennessee State 21\xa0\xa0',
'\xa0', 'Nov 08, 2008\xa0\xa0', 'Nashville, TN       \xa0\xa0', 'Tennessee State 45, Eastern Illinois 24\xa0\xa
0', '\xa0', 'Nov 01, 2008\xa0\xa0', 'Nashville, TN       \xa0\xa0', 'Tennessee State 41, Tennessee Tech 14\xa0
\xa0', '\xa0', 'Oct 25, 2008\xa0\xa0', 'Cape Girardeau, Mo. \xa0\xa0', 'Southeast Missouri 27, Tennessee State
20\xa0\xa0', '\xa0', 'Oct 18, 2008\xa0\xa0', 'Nashville, TN       \xa0\xa0', 'Tennessee State 37, Austin Peay 3
4\xa0\xa0', '\xa0', 'Oct 04, 2008\xa0\xa0', 'Martin, TN          \xa0\xa0', 'Tennessee State 30, UT Martin 27\x
a0\xa0', '\xa0', 'Sep 27, 2008\xa0\xa0', 'Atlanta, GA         \xa0\xa0', 'Florida A&M 28, Tennessee State 21\xa
0\xa0', '\xa0', 'Sep 20, 2008\xa0\xa0', 'Nashville, TN       \xa0\xa0', 'Tennessee State 34, Eastern Kentucky 2
0\xa0\xa0', '\xa0', 'Sep 13, 2008\xa0\xa0', 'Memphis, TN         \xa0\xa0', 'Tennessee State 41, Jackson State
18\xa0\xa0', '\xa0', 'Sep 06, 2008\xa0\xa0', 'Nashville, TN       \xa0\xa0', 'Tennessee State 34, Southern 32\x
a0\xa0', '\xa0', 'Aug 30, 2008\xa0\xa0', 'Huntsville, AL      \xa0\xa0', 'Tennessee State 34, Alabama A&M 13\xa
0\xa0', '\xa0']
```

In [25]:

```
#cleaning 2008 list
schedule2008=[r.replace("Southern","Southern University") for r in schedule2008]
schedule2008=[s.replace("UT","Tennessee") for s in schedule2008]
schedule2008=[t.replace("Cape Girardeau","CapeGirardeau") for t in schedule2008]
schedule2008=[v.replace("-"," ") for v in schedule2008]
schedule2008=[w.replace(".", "") for w in schedule2008]
schedule2008=[x.replace("\xa0", "") for x in schedule2008]
schedule2008=[y.replace(",","") for y in schedule2008]
schedule2008=[z.strip() for z in schedule2008]
print(schedule2008)
```

```
['Nov 22 2008', 'Murray Ky', 'Murray State 24 Tennessee State 17', '', 'Nov 15 2008', 'Jacksonville Ala', 'Jack
sonville State 26 Tennessee State 21', '', 'Nov 08 2008', 'Nashville TN', 'Tennessee State 45 Eastern Illinois
24', '', 'Nov 01 2008', 'Nashville TN', 'Tennessee State 41 Tennessee Tech 14', '', 'Oct 25 2008', 'CapeGirarde
au Mo', 'Southeast Missouri 27 Tennessee State 20', '', 'Oct 18 2008', 'Nashville TN', 'Tennessee State 37 Aust
in Peay 34', '', 'Oct 04 2008', 'Martin TN', 'Tennessee State 30 Tennessee Martin 27', '', 'Sep 27 2008', 'Atla
nta GA', 'Florida A&M 28 Tennessee State 21', '', 'Sep 20 2008', 'Nashville TN', 'Tennessee State 34 Eastern Ke
ntucky 20', '', 'Sep 13 2008', 'Memphis TN', 'Tennessee State 41 Jackson State 18', '', 'Sep 06 2008', 'Nashvil
le TN', 'Tennessee State 34 Southern University 32', '', 'Aug 30 2008', 'Huntsville AL', 'Tennessee State 34 Al
abama A&M 13', '']
```

In [26]:

```
#2007 data
page = requests.get('https://tennstate_ftp.sidearmsports.com/custompages/tsutigers/4C48DA5B-76BE-4378-8E95-9BB2
mytree = html.fromstring(page.content)
schedule2007 = mytree.xpath('body//tr/td/font[@color="#000000"]/text()')
print(schedule2007)
```

```
['Nov 17, 2007\xa0\xa0', 'Nashville, TN       \xa0\xa0', 'Tennessee-Martin 43, Tennessee State 38\xa0\xa0', '\x
a0', 'Nov 08, 2007\xa0\xa0', 'Birmingham, AL      \xa0\xa0', 'Tennessee State 38, Samford 28\xa0\xa0', '\xa0',
'Nov 03, 2007\xa0\xa0', 'Nashville, TN       \xa0\xa0', 'Tennessee State 42, Murray State 28\xa0\xa0', '\xa0',
'Oct 27, 2007\xa0\xa0', 'Nashville, TN       \xa0\xa0', 'Eastern Illinois 38, Tennessee State 35\xa0\xa0', '\xa
0', 'Oct 20, 2007\xa0\xa0', 'Richmond, Ky.       \xa0\xa0', 'Eastern Kentucky 49, Tennessee State 7\xa0\xa0',
'\xa0', 'Oct 11, 2007\xa0\xa0', 'Cookeville, TN      \xa0\xa0', 'Tennessee State 45, Tennessee Tech 28\xa0\xa
0', '\xa0', 'Sep 29, 2007\xa0\xa0', 'Atlanta, Ga         \xa0\xa0', 'Florida A&M 18, Tennessee State 17\xa0\xa
0', '\xa0', 'Sep 22, 2007\xa0\xa0', 'Baton Rouge, LA     \xa0\xa0', 'Southern 41, Tennessee State 34\xa0\xa0',
'\xa0', 'Sep 15, 2007\xa0\xa0', 'Clarksville, Tenn.  \xa0\xa0', 'Tennessee State 33, Austin Peay 32\xa0\xa0',
'\xa0', 'Sep 08, 2007\xa0\xa0', 'Memphis, TN         \xa0\xa0', 'Tennessee State 16, Jackson State 13\xa0\xa0',
'\xa0', 'Sep 01, 2007\xa0\xa0', 'Nashville, TN       \xa0\xa0', 'Alabama A&M 49, Tennessee State 23\xa0\xa0',
'\xa0']
```

In [27]:

```
#cleaning 2007 list
schedule2007=[r.replace("Southern","Southern University") for r in schedule2007]
schedule2007=[s.replace("Samford","Samford University") for s in schedule2007]
schedule2007=[t.replace("Baton Rouge","BatonRouge") for t in schedule2007]
schedule2007=[v.replace("-"," ") for v in schedule2007]
schedule2007=[w.replace(".", "") for w in schedule2007]
schedule2007=[x.replace("\xa0", "") for x in schedule2007]
```

```
schedule2007=[y.replace(",","") for y in schedule2007]
schedule2007=[z.strip() for z in schedule2007]
print(schedule2007)
```

```
['Nov 17 2007', 'Nashville TN', 'Tennessee Martin 43 Tennessee State 38', '', 'Nov 08 2007', 'Birmingham AL',
'Tennessee State 38 Samford University 28', '', 'Nov 03 2007', 'Nashville TN', 'Tennessee State 42 Murray State
28', '', 'Oct 27 2007', 'Nashville TN', 'Eastern Illinois 38 Tennessee State 35', '', 'Oct 20 2007', 'Richmond
Ky', 'Eastern Kentucky 49 Tennessee State 7', '', 'Oct 11 2007', 'Cookeville TN', 'Tennessee State 45 Tennessee
Tech 28', '', 'Sep 29 2007', 'Atlanta Ga', 'Florida A&M 18 Tennessee State 17', '', 'Sep 22 2007', 'BatonRouge
LA', 'Southern University 41 Tennessee State 34', '', 'Sep 15 2007', 'Clarksville Tenn', 'Tennessee State 33 Au
stin Peay 32', '', 'Sep 08 2007', 'Memphis TN', 'Tennessee State 16 Jackson State 13', '', 'Sep 01 2007', 'Nash
ville TN', 'Alabama A&M 49 Tennessee State 23', '']
```

In [28]:

```
#2006 data
page = requests.get('https://tennstate_ftp.sidearmsports.com/custompages/tsutigers/902C2D72-E881-4838-87F2-28BA
mytree = html.fromstring(page.content)
schedule2006 = mytree.xpath('body//tr/td/font[@color="#000000"]/text()')
print(schedule2006)
```

```
['Nov 18, 2006\xa0\xa0', 'Richmond, Ky.       \xa0\xa0', 'Eastern Kentucky 20, Tennessee State 3\xa0\xa0', '\xa
0', 'Nov 11, 2006\xa0\xa0', 'Nashville, TN       \xa0\xa0', 'Tennessee State 31, Southeast Missouri 0\xa0\xa0',
'\xa0', 'Nov 04, 2006\xa0\xa0', 'Charleston, IL      \xa0\xa0', 'EASTERN ILLINOIS 29, Tennessee State 3\xa0\xa
0', '\xa0', 'Oct 28, 2006\xa0\xa0', 'Nashville, TN       \xa0\xa0', 'Tennessee State 29, Samford 7\xa0\xa0',
'\xa0', 'Oct 21, 2006\xa0\xa0', 'Jacksonville, Ala.  \xa0\xa0', 'Tennessee State 38, Jacksonville State 31\xa0
\xa0', '\xa0', 'Oct 14, 2006\xa0\xa0', 'Nashville, TN       \xa0\xa0', 'Tennessee State 30, Tennessee Tech 20\x
a0\xa0', '\xa0', 'Sep 30, 2006\xa0\xa0', 'Atlanta, GA         \xa0\xa0', 'Florida A&M 25, Tennessee State 22\xa
0\xa0', '\xa0', 'Sep 23, 2006\xa0\xa0', 'Nashville, TN       \xa0\xa0', 'Vanderbilt 38, Tennessee State 9\xa0\x
a0', '\xa0', 'Sep 16, 2006\xa0\xa0', 'Memphis, TN         \xa0\xa0', 'Tennessee State 31, Jackson State 30\xa0
\xa0', '\xa0', 'Sep 09, 2006\xa0\xa0', 'Nashville, TN       \xa0\xa0', 'Tennessee State 25, Murray State Univ 1
5\xa0\xa0', '\xa0', 'Sep 02, 2006\xa0\xa0', 'Nashville,TN        \xa0\xa0', 'Alabama A&M 27, Tennessee State 20
\xa0\xa0', '\xa0']
```

In [29]:

```
#cleaning 2006 list
schedule2006=[p.replace("Nashville,TN","Nashville, TN") for p in schedule2006]
schedule2006=[q.replace("Murray State Univ","Murray State") for q in schedule2006]
schedule2006=[r.replace("Vanderbilt","Vanderbilt University") for r in schedule2006]
schedule2006=[s.replace("Samford","Samford University") for s in schedule2006]
schedule2006=[t.replace("EASTERN ILLINOIS","Eastern Illinois") for t in schedule2006]
schedule2006=[v.replace("-"," ") for v in schedule2006]
schedule2006=[w.replace(".", "") for w in schedule2006]
schedule2006=[x.replace("\xa0", "") for x in schedule2006]
schedule2006=[y.replace(",","") for y in schedule2006]
schedule2006=[z.strip() for z in schedule2006]
print(schedule2006)
```

```
['Nov 18 2006', 'Richmond Ky', 'Eastern Kentucky 20 Tennessee State 3', '', 'Nov 11 2006', 'Nashville TN', 'Ten
nessee State 31 Southeast Missouri 0', '', 'Nov 04 2006', 'Charleston IL', 'Eastern Illinois 29 Tennessee State
3', '', 'Oct 28 2006', 'Nashville TN', 'Tennessee State 29 Samford University 7', '', 'Oct 21 2006', 'Jacksonvi
lle Ala', 'Tennessee State 38 Jacksonville State 31', '', 'Oct 14 2006', 'Nashville TN', 'Tennessee State 30 Te
nnessee Tech 20', '', 'Sep 30 2006', 'Atlanta GA', 'Florida A&M 25 Tennessee State 22', '', 'Sep 23 2006', 'Nas
hville TN', 'Vanderbilt University 38 Tennessee State 9', '', 'Sep 16 2006', 'Memphis TN', 'Tennessee State 31
Jackson State 30', '', 'Sep 09 2006', 'Nashville TN', 'Tennessee State 25 Murray State 15', '', 'Sep 02 2006',
'Nashville TN', 'Alabama A&M 27 Tennessee State 20', '']
```

In [30]:

```
#2005 data
page = requests.get('https://tennstate_ftp.sidearmsports.com/custompages/tsutigers/99B728E2-12D4-4887-A990-2498
mytree = html.fromstring(page.content)
schedule2005 = mytree.xpath('body//tr/td/font[@color="#000000"]/text()')
print(schedule2005)
```

```
['Nov 19, 2005\xa0\xa0', 'Nashville, TN       \xa0\xa0', 'EKU 49, Tennessee State 0\xa0\xa0', '\xa0', 'Nov 12,
2005\xa0\xa0', 'Cape Girardeau, Mo. \xa0\xa0', 'SEMO 32, Tennessee State 24\xa0\xa0', '\xa0', 'Nov 05, 2005\xa0
\xa0', 'Nashville, TN       \xa0\xa0', 'Eastern Illinois Uni 27, Tennessee State 3\xa0\xa0', '\xa0', 'Oct 29, 2
005\xa0\xa0', 'Birmingham, Ala.    \xa0\xa0', 'Samford 31, Tennessee State 11\xa0\xa0', '\xa0', 'Oct 22, 2005\x
a0\xa0', 'Nashville,TN        \xa0\xa0', 'Jacksonville State 33, Tennessee State 3\xa0\xa0', '\xa0', 'Oct 13, 2
005\xa0\xa0', 'Cookeville, Tenn.   \xa0\xa0', 'Tennessee State 31, Tennessee Tech 20\xa0\xa0', '\xa0', 'Oct 01,
2005\xa0\xa0', 'Indianapolis, Ind.  \xa0\xa0', 'No. Carolina A&T St. 16, Tennessee State 3\xa0\xa0', '\xa0', 'S
ep 24, 2005\xa0\xa0', 'Atlanta, GA         \xa0\xa0', 'Florida A&M 12, Tennessee State 7\xa0\xa0', '\xa0', 'Sep
17, 2005\xa0\xa0', 'Martin, Tenn        \xa0\xa0', 'Tennessee-Martin 42, Tennessee State 20\xa0\xa0', '\xa0',
'Sep 10, 2005\xa0\xa0', 'Memphis, TN         \xa0\xa0', 'Tennessee State 20, Jackson State 14\xa0\xa0', '\xa0',
'Sep 03, 2005\xa0\xa0', 'Nashville,TN        \xa0\xa0', 'Alabama A&M Univ 27, Tennessee State 14\xa0\xa0', '\xa
0']
```

In [31]:

```
#cleaning 2005 list
schedule2005=[m.replace("Eastern Illinois Uni","Eastern Illinois") for m in schedule2005]
schedule2005=[n.replace("No. Carolina A&T St.","NorthCarolina A&T") for n in schedule2005]
schedule2005=[o.replace("Alabama A&M Univ","Alabama A&M") for o in schedule2005]
schedule2005=[p.replace("Nashville,TN","Nashville, TN") for p in schedule2005]
schedule2005=[q.replace("Samford","Samford University") for q in schedule2005]
schedule2005=[r.replace("SEMO","Southeast Missouri") for r in schedule2005]
schedule2005=[s.replace("EKU","Eastern Kentucky") for s in schedule2005]
schedule2005=[t.replace("Cape Girardeau","CapeGirardeau") for t in schedule2005]
```

```
schedule2005=[v.replace("-"," ") for v in schedule2005]
schedule2005=[w.replace(".", "") for w in schedule2005]
schedule2005=[x.replace("\xa0", "") for x in schedule2005]
schedule2005=[y.replace(",","") for y in schedule2005]
schedule2005=[z.strip() for z in schedule2005]
print(schedule2005)
```

```
['Nov 19 2005', 'Nashville TN', 'Eastern Kentucky 49 Tennessee State 0', '', 'Nov 12 2005', 'CapeGirardeau Mo',
'Southeast Missouri 32 Tennessee State 24', '', 'Nov 05 2005', 'Nashville TN', 'Eastern Illinois 27 Tennessee S
tate 3', '', 'Oct 29 2005', 'Birmingham Ala', 'Samford University 31 Tennessee State 11', '', 'Oct 22 2005', 'N
ashville TN', 'Jacksonville State 33 Tennessee State 3', '', 'Oct 13 2005', 'Cookeville Tenn', 'Tennessee State
31 Tennessee Tech 20', '', 'Oct 01 2005', 'Indianapolis Ind', 'NorthCarolina A&T 16 Tennessee State 3', '', 'Se
p 24 2005', 'Atlanta GA', 'Florida A&M 12 Tennessee State 7', '', 'Sep 17 2005', 'Martin Tenn', 'Tennessee Mart
in 42 Tennessee State 20', '', 'Sep 10 2005', 'Memphis TN', 'Tennessee State 20 Jackson State 14', '', 'Sep 03
2005', 'Nashville TN', 'Alabama A&M 27 Tennessee State 14', '']
```

In [32]:

```
#2004 data
page = requests.get('https://tennstate_ftp.sidearmsports.com/custompages/tsutigers/57C54C44-DE63-4F94-99EB-2BB3
mytree = html.fromstring(page.content)
schedule2004 = mytree.xpath('body//tr/td/font[@color="#000000"]/text()')
print(schedule2004)
```

```
['Nov 20, 2004\xa0\xa0', 'Nashville, TN       \xa0\xa0', 'Murray State 30, Tennessee State 13\xa0\xa0', 'Nov 1
3, 2004\xa0\xa0', 'Richmond KY         \xa0\xa0', 'Eastern Kentucky 29, Tennessee State 14\xa0\xa0', 'Nov 06, 2
004\xa0\xa0', 'Nashville,TN        \xa0\xa0', 'Tennessee State 38, Southeast Missouri 36\xa0\xa0', 'Oct 30, 200
4\xa0\xa0', 'Charleston, Illinois\xa0\xa0', 'Eastern Illinois 34, Tennessee State 24\xa0\xa0', 'Oct 23, 2004\xa
0\xa0', 'Nashville, TN       \xa0\xa0', 'Samford University 42, Tennessee State 36\xa0\xa0', 'Oct 16, 2004\xa0
\xa0', 'Jacksonville, Ala.  \xa0\xa0', 'Jacksonville State 49, Tennessee State 35\xa0\xa0', 'Oct 02, 2004\xa0\x
a0', 'RCA Dome            \xa0\xa0', 'South Carolina State 30, Tennessee State 13\xa0\xa0', 'Sep 25, 2004\xa0\x
a0', 'Atlanta, Georgia    \xa0\xa0', 'Florida A&M 21, Tennessee State 15\xa0\xa0', 'Sep 18, 2004\xa0\xa0', 'Mem
phis, Tennessee  \xa0\xa0', 'Tennessee State 21, Jackson State 20\xa0\xa0', 'Sep 09, 2004\xa0\xa0', 'Martin, Te
nnessee   \xa0\xa0', 'Tennessee State 27, Tennessee-Martin 13\xa0\xa0', 'Sep. 4, 2003\xa0\xa0', 'Nashville, TN
      \xa0\xa0', 'Tennessee State 42, Alabama A&M 7\xa0\xa0']
```

In [33]:

```
#cleaning 2004 list
schedule2004=[p.replace('Sep. 4, 2003','Sep 04 2004') for p in schedule2004]
schedule2004=[r.replace("Nashville,TN","Nashville, TN") for r in schedule2004]
schedule2004=[s.replace("South Carolina State","SouthCarolina State") for s in schedule2004]
schedule2004=[t.replace("RCA Dome","Indianapolis, IN") for t in schedule2004]
schedule2004=[v.replace("-"," ") for v in schedule2004]
schedule2004=[w.replace(".", "") for w in schedule2004]
schedule2004=[x.replace("\xa0", "") for x in schedule2004]
schedule2004=[y.replace(",","") for y in schedule2004]
schedule2004=[z.strip() for z in schedule2004]
schedule2004.insert(3, "")
schedule2004.insert(7, "")
schedule2004.insert(11, "")
schedule2004.insert(15, "")
schedule2004.insert(19, "")
schedule2004.insert(23, "")
schedule2004.insert(27, "")
schedule2004.insert(31, "")
schedule2004.insert(35, "")
schedule2004.insert(39, "")
schedule2004.insert(43, "")
print(schedule2004)
```

```
['Nov 20 2004', 'Nashville TN', 'Murray State 30 Tennessee State 13', '', 'Nov 13 2004', 'Richmond KY', 'Easter
n Kentucky 29 Tennessee State 14', '', 'Nov 06 2004', 'Nashville TN', 'Tennessee State 38 Southeast Missouri 3
6', '', 'Oct 30 2004', 'Charleston Illinois', 'Eastern Illinois 34 Tennessee State 24', '', 'Oct 23 2004', 'Nas
hville TN', 'Samford University 42 Tennessee State 36', '', 'Oct 16 2004', 'Jacksonville Ala', 'Jacksonville St
ate 49 Tennessee State 35', '', 'Oct 02 2004', 'Indianapolis IN', 'SouthCarolina State 30 Tennessee State 13',
'', 'Sep 25 2004', 'Atlanta Georgia', 'Florida A&M 21 Tennessee State 15', '', 'Sep 18 2004', 'Memphis Tennesse
e', 'Tennessee State 21 Jackson State 20', '', 'Sep 09 2004', 'Martin Tennessee', 'Tennessee State 27 Tennessee
Martin 13', '', 'Sep 04 2004', 'Nashville TN', 'Tennessee State 42 Alabama A&M 7', '']
```

In [34]:

```
#2003 data
page = requests.get('https://tennstate_ftp.sidearmsports.com/custompages/tsutigers/A5FB3AA1-8151-496A-A022-754E
mytree = html.fromstring(page.content)
schedule2003 = mytree.xpath('body//tr/td/font[@color="#000000"]/text()')
print(schedule2003)
```

```
['Nov 22, 2003\xa0\xa0', 'Murray, Ky.         \xa0\xa0', 'Tennessee State 35, Murray State 10\xa0\xa0', 'Nov 1
5, 2003\xa0\xa0', 'Nashville, Tenn.    \xa0\xa0', 'Eastern Kentucky 43, Tennessee State 38\xa0\xa0', 'Nov 08, 2
003\xa0\xa0', 'Cape Girardeau, Mo. \xa0\xa0', 'Southeast Missouri 52, Tennessee State 35\xa0\xa0', 'Nov 01, 200
3\xa0\xa0', 'Nashville, Tenn.    \xa0\xa0', 'Tennessee State 24, Eastern Illinois 14\xa0\xa0', 'Oct 25, 2003\xa
0\xa0', 'Birmingham, Ala.    \xa0\xa0', 'Tennessee State 29, Samford 24\xa0\xa0', 'Oct 18, 2003\xa0\xa0', 'Nash
ville, Tennessee\xa0\xa0', 'Jacksonville State 34, Tennessee State 7\xa0\xa0', 'Oct 11, 2003\xa0\xa0', 'Cookevi
```

```
lle, Tenn.   \xa0\xa0', 'Tennessee State 27, Tennessee Tech 23\xa0\xa0', 'Sep 27, 2003\xa0\xa0', 'Nashville, Te
nn.    \xa0\xa0', 'Tennessee State 41, Tennessee-Martin 10\xa0\xa0', 'Sep 20, 2003\xa0\xa0', 'Atlanta, Ga.
   \xa0\xa0', 'Florida A&M Univ. 10, Tennessee State 7\xa0\xa0', 'Sep 13, 2003\xa0\xa0', 'Memphis, Tennessee
\xa0\xa0', 'Tennessee State 44, Jackson State Tigers 14\xa0\xa0', 'Sep. 6, 2003\xa0\xa0', 'Huntsville, Alabama
\xa0\xa0', 'Alabama A&M 31, Tennessee State 24\xa0\xa0', 'Aug 30, 2003\xa0\xa0', 'Nashville, Tenn.    \xa0\xa
0', 'Tennessee State 37, South Carolina State 20\xa0\xa0']
```

In [35]:

```
#cleaning 2003 list
schedule2003=[q.replace("Univ.","") for q in schedule2003]
schedule2003=[r.replace("South Carolina State","SouthCarolina State") for r in schedule2003]
schedule2003=[s.replace("Samford","Samford University") for s in schedule2003]
schedule2003=[s.replace("Tigers","") for s in schedule2003]
schedule2003=[t.replace("Cape Girardeau","CapeGirardeau") for t in schedule2003]
schedule2003=[v.replace("-"," ") for v in schedule2003]
schedule2003=[w.replace(".", "") for w in schedule2003]
schedule2003=[x.replace("\xa0", "") for x in schedule2003]
schedule2003=[y.replace(",","") for y in schedule2003]
schedule2003=[z.strip() for z in schedule2003]
schedule2003
schedule2003.insert(3, "")
schedule2003.insert(7, "")
schedule2003.insert(11, "")
schedule2003.insert(15, "")
schedule2003.insert(19, "")
schedule2003.insert(23, "")
schedule2003.insert(27, "")
schedule2003.insert(31, "")
schedule2003.insert(35, "")
schedule2003.insert(39, "")
schedule2003.insert(43, "")
schedule2003.insert(47, "")
print(schedule2003)
```

```
['Nov 22 2003', 'Murray Ky', 'Tennessee State 35 Murray State 10', '', 'Nov 15 2003', 'Nashville Tenn', 'Easter
n Kentucky 43 Tennessee State 38', '', 'Nov 08 2003', 'CapeGirardeau Mo', 'Southeast Missouri 52 Tennessee Stat
e 35', '', 'Nov 01 2003', 'Nashville Tenn', 'Tennessee State 24 Eastern Illinois 14', '', 'Oct 25 2003', 'Birmi
ngham Ala', 'Tennessee State 29 Samford University 24', '', 'Oct 18 2003', 'Nashville Tennessee', 'Jacksonville
State 34 Tennessee State 7', '', 'Oct 11 2003', 'Cookeville Tenn', 'Tennessee State 27 Tennessee Tech 23', '',
'Sep 27 2003', 'Nashville Tenn', 'Tennessee State 41 Tennessee Martin 10', '', 'Sep 20 2003', 'Atlanta Ga', 'Fl
orida A&M  10 Tennessee State 7', '', 'Sep 13 2003', 'Memphis Tennessee', 'Tennessee State 44 Jackson State  1
4', '', 'Sep 6 2003', 'Huntsville Alabama', 'Alabama A&M 31 Tennessee State 24', '', 'Aug 30 2003', 'Nashville
Tenn', 'Tennessee State 37 SouthCarolina State 20', '']
```

In [36]:

```
#get number of elements for 2003 list
len(schedule2003)
```

Out[36]:

```
48
```

In [37]:

```
#get number of elements for 2004 list
len(schedule2004)
```

Out[37]:

```
44
```

In [38]:

```
#get number of elements for 2005 list
len(schedule2005)
```

Out[38]:

```
44
```

In [39]:

```
#get number of elements for 2006 list
len(schedule2006)
```

Out[39]:

```
44
```

In [40]:

```
#get number of elements for 2007 list
len(schedule2007)
```

Out[40]:

```
44
```

In [41]:

```
#get number of elements for 2008 list
len(schedule2008)
```

Out[41]:

```
48
```

In [42]:

```
#get number of elements for 2009 list
len(schedule2009)
```

Out[42]: 44

In [43]:
```
#get number of elements for 2010 list
len(schedule2010)
```

Out[43]: 44

In [44]:
```
#get number of elements for 2011 list
len(schedule2011)
```

Out[44]: 44

In [45]:
```
#get number of elements for 2012 list
len(schedule2012)
```

Out[45]: 44

In [46]:
```
#get number of elements for 2013 list
len(schedule2013)
```

Out[46]: 56

In [47]:
```
# get number of elements for 2014 list
len(schedule2014)
```

Out[47]: 48

In [48]:
```
# get number of elements for 2015 list
len(schedule2015)
```

Out[48]: 40

In [49]:
```
# get number of elements for 2016 list
len(schedule2016)
```

Out[49]: 44

In [50]:
```
# get number of elements for 2017 list
len(schedule2017)
```

Out[50]: 44

In [51]:
```
# get number of elements for 2018 list
len(schedule2018)
```

Out[51]: 36

In [52]:
```
#get number of elements for 2019 list
len(schedule2019)
```
Out[52]: 48

In [53]:
```
#create a single list for all years
fulldata=schedule2003+schedule2004+schedule2005+schedule2006+schedule2007+schedule2008+schedule2009+schedule201
#get number of elements for full data
len(fulldata)
```
Out[53]: 764

In [54]:
```
#create list of sublists from fulldata list
list_of_sublists=[fulldata[i:i+4] for i in range(0,len(fulldata),4)]
print(list_of_sublists[:5])
```
[['Nov 22 2003', 'Murray Ky', 'Tennessee State 35 Murray State 10', ''], ['Nov 15 2003', 'Nashville Tenn', 'Eas tern Kentucky 43 Tennessee State 38', ''], ['Nov 08 2003', 'CapeGirardeau Mo', 'Southeast Missouri 52 Tennessee State 35', ''], ['Nov 01 2003', 'Nashville Tenn', 'Tennessee State 24 Eastern Illinois 14', ''], ['Oct 25 200 3', 'Birmingham Ala', 'Tennessee State 29 Samford University 24', '']]

In [55]:
```
#create dataframe from list of sublists
df=pd.DataFrame(list_of_sublists, columns=['date','location','scores','reserve'])
df.head()
```
Out[55]:

	date	location	scores	reserve
0	Nov 22 2003	Murray Ky	Tennessee State 35 Murray State 10	
1	Nov 15 2003	Nashville Tenn	Eastern Kentucky 43 Tennessee State 38	
2	Nov 08 2003	CapeGirardeau Mo	Southeast Missouri 52 Tennessee State 35	
3	Nov 01 2003	Nashville Tenn	Tennessee State 24 Eastern Illinois 14	
4	Oct 25 2003	Birmingham Ala	Tennessee State 29 Samford University 24	

In [56]:
```
#check reserve column
df['reserve'].value_counts()
```
Out[56]:
```
    191
Name: reserve, dtype: int64
```

Data Cleaning- dropping unneeded variables and creating needed variables

In [57]:
```
#drop reserve column
df=df.drop('reserve',axis=1)
df.head()
```
Out[57]:

	date	location	scores
0	Nov 22 2003	Murray Ky	Tennessee State 35 Murray State 10
1	Nov 15 2003	Nashville Tenn	Eastern Kentucky 43 Tennessee State 38
2	Nov 08 2003	CapeGirardeau Mo	Southeast Missouri 52 Tennessee State 35
3	Nov 01 2003	Nashville Tenn	Tennessee State 24 Eastern Illinois 14
4	Oct 25 2003	Birmingham Ala	Tennessee State 29 Samford University 24

In [58]:
```
#get number of rows and columns in dataframe
df.shape
```
Out[58]: (191, 3)

In [59]:
```
#creating relevant columns & dropping unnecessary columns
df['date']=pd.to_datetime(df['date'],format='%b %d %Y')
df[['city','state']]=df.location.str.split(expand=True)
df[['win1','win2','winscore','los1','los2','loscore']]=df.scores.str.split(expand=True)
df['loser']=df['los1']+' '+df['los2']
df['winner']=df['win1']+' '+df['win2']
df['locale'] = np.where(df['city']=='Nashville', 'Home', 'Away')
df['TSU score']=pd.to_numeric(np.where(df['winner']=='Tennessee State',df['winscore'],df['loscore']))
df['opponent score']=pd.to_numeric(np.where(df['winner']=='Tennessee State',df['loscore'],df['winscore']))
df['winscore']=pd.to_numeric(df['winscore'])
df['loscore']=pd.to_numeric(df['loscore'])
df['scorediff']=df['TSU score']-df['opponent score']
df['scorediff_abs']=abs(df['scorediff'])
df['winloss']=np.where(df['winner']=='Tennessee State','Win','Loss')
df['year']=df.date.dt.year
df.head()
df=df.drop(['location','scores','win1','win2','los1','los2'], axis=1)
df.columns
```
Out[59]:
```
Index(['date', 'city', 'state', 'winscore', 'loscore', 'loser', 'winner',
       'locale', 'TSU score', 'opponent score', 'scorediff', 'scorediff_abs',
       'winloss', 'year'],
      dtype='object')
```

In [60]:
```
#clean up city column
df['city'].unique()
```
Out[60]:
```
array(['Murray', 'Nashville', 'CapeGirardeau', 'Birmingham', 'Cookeville',
       'Atlanta', 'Memphis', 'Huntsville', 'Richmond', 'Charleston',
       'Clarksville', 'AirForceAcademy', 'DaytonaBeach', 'BatonRouge',
       'Tallahassee', 'Montgomery', 'Tampa', 'Murfreesboro', 'StLouis'], dtype=object)
```

In [61]:
```
df['city']=df['city'].replace("CapeGirardeau","Cape Girardeau").replace("AirForceAcademy","Air Force Academy")
df['city']=df['city'].replace("DaytonaBeach","Daytona Beach").replace("StLouis", "St. Louis")
df['city']=df['city'].replace("BatonRouge","Baton Rouge")
df['city'].unique()
```
Out[61]:
```
array(['Murray', 'Nashville', 'Cape Girardeau', 'Birmingham',
       'Cookeville', 'Atlanta', 'Memphis', 'Huntsville', 'Richmond',
       'Charleston', 'Jacksonville', 'Indianapolis', 'Martin',
       'Baton Rouge', 'Clarksville', 'Air Force Academy', 'Daytona Beach',
       'St. Louis', 'Tallahassee', 'Montgomery', 'Tampa', 'Murfreesboro'],
      dtype=object)
```

In [62]:
```
#clean up state column
df['state'].unique()
```
Out[62]:
```
array(['Ky', 'Tenn', 'Mo', 'Ala', 'Tennessee', 'Ga', 'Alabama', 'TN',
       'KY', 'Illinois', 'IN', 'Georgia', 'Ind', 'GA', 'LA',
       'Ill', 'Colo', 'Fla', 'Florida', 'CO'], dtype=object)
```

In [63]:
```
df['state']=df['state'].replace(["Ga","Georgia"],"GA").replace(["Tenn","Tennessee"],"TN").replace(["Ind","IN"]
df['state']=df['state'].replace(["Mo","Missouri"],"MO").replace(["Ala","Alabama"],"AL").replace(["Ky","KY"]
df['state']=df['state'].replace(["Ill","Illinois"],"IL").replace(["Fla","Florida"],"FL").replace(["Colo","CO"]).replace("Ind"
df['state'].unique()
```
Out[63]:
```
array(['KY', 'TN', 'MO', 'AL', 'GA', 'IL', 'IN', 'LA', 'CO', 'FL'],
```

In [64]:
```
#clean up loser column values
df['loser'].unique()
```
Out[64]:
```
array(['Murray State', 'Tennessee State', 'Eastern Illinois',
       'Samford University', 'Tennessee Tech', 'Tennessee Martin',
       'Jackson State', 'SouthCarolina State', 'Southeast Missouri',
       'Eastern Kentucky', 'Jacksonville State', 'Austin Peay',
       'Alabama A&M', 'Southern University', 'NorthCarolina A&T',
       'Florida A&M', 'Arkansas PineBluff', 'Bethune Cookman',
       'Butler University', 'Central State', 'Edward Waters',
       'Alabama State', 'VirginiaU Lynchburg', 'Georgia State',
       'Mississippi Valley'], dtype=object)
```

In [65]:
```
df['loser']=df['loser'].replace("Murray State","Murray State University")
df['loser']=df['loser'].replace("Tennessee State","Tennessee State University")
df['loser']=df['loser'].replace("Eastern Illinois","Eastern Illinois University")
df['loser']=df['loser'].replace("Tennessee Tech","Tennessee Tech University")
df['loser']=df['loser'].replace("Tennessee Martin","University of Tennessee Martin")
df['loser']=df['loser'].replace("Jackson State","Jackson State University")
df['loser']=df['loser'].replace("SouthCarolina State","South Carolina State University")
df['loser']=df['loser'].replace("Southeast Missouri","Southeast Missouri State University")
df['loser']=df['loser'].replace("Alabama A&M","Alabama A&M University")
df['loser']=df['loser'].replace("Jacksonville State","Jacksonville State University")
df['loser']=df['loser'].replace("Austin Peay","Austin Peay State University")
df['loser']=df['loser'].replace("Eastern Kentucky","Eastern Kentucky University")
df['loser']=df['loser'].replace("Southern University","Southern University & A&M College")
df['loser']=df['loser'].replace("NorthCarolina A&T","North Carolina A&T State University")
df['loser']=df['loser'].replace("Florida A&M","Florida A&M University")
df['loser']=df['loser'].replace("Arkansas PineBluff","University of Arkansas Pine Bluff")
df['loser']=df['loser'].replace("Bethune Cookman","Bethune Cookman University")
df['loser']=df['loser'].replace("Central State","Central State University")
df['loser']=df['loser'].replace("Edward Waters","Edward Waters College")
df['loser']=df['loser'].replace("Alabama State","Alabama State University")
df['loser']=df['loser'].replace("VirginiaU Lynchburg","Virginia University Lynchburg")
df['loser']=df['loser'].replace("Georgia State","Georgia State University")
df['loser']=df['loser'].replace("Mississippi Valley","Mississippi Valley State University")
df['loser'].unique()
```
Out[65]:
```
array(['Murray State University', 'Tennessee State University',
       'Eastern Illinois University', 'Samford University',
       'Tennessee Tech University', 'University of Tennessee Martin',
       'Jackson State University', 'South Carolina State University',
       'Southeast Missouri State University', 'Alabama A&M University',
       'Jacksonville State University', 'Austin Peay State University',
       'Eastern Kentucky University', 'Southern University & A&M College',
       'North Carolina A&T State University', 'Florida A&M University',
       'University of Arkansas Pine Bluff', 'Bethune Cookman University',
       'Butler University', 'Central State University',
       'Edward Waters College', 'Alabama State University',
       'Virginia University Lynchburg', 'Georgia State University',
       'Mississippi Valley State University'], dtype=object)
```

In [66]:
```
#clean up winner column values
df['winner'].unique()
```
Out[66]:
```
array(['Tennessee State', 'Eastern Kentucky', 'Southeast Missouri',
       'Alabama State', 'Florida A&M', 'Alabama A&M', 'Murray State',
       'Eastern Illinois', 'Jacksonville State', 'SouthCarolina State',
       'NorthCarolina A&T', 'Tennessee Martin', 'Vanderbilt University',
       'Southern University', 'Austin Peay', 'Tennessee Tech',
       'Jackson State', 'Air Force', 'Bethune Cookman',
       'Alabama State', 'Murray St', 'Pine Bluff', 'Middle Tennessee'],
      dtype=object)
```

In [67]:
```
df['winner']=df['winner'].replace("Tennessee State","Tennessee State University")
df['winner']=df['winner'].replace("Eastern Kentucky","Eastern Kentucky University")
df['winner']=df['winner'].replace("Southeast Missouri","Southeast Missouri State University")
df['winner']=df['winner'].replace("Jacksonville State","Jacksonville State University")
df['winner']=df['winner'].replace("Florida A&M", "Florida A&M University")
df['winner']=df['winner'].replace("Alabama A&M","Alabama A&M University")
df['winner']=df['winner'].replace("Murray St","Murray State University")
df['winner']=df['winner'].replace("Eastern Illinois","Eastern Illinois University")
df['winner']=df['winner'].replace("SouthCarolina State","South Carolina State University")
df['winner']=df['winner'].replace("NorthCarolina A&T","North Carolina A&T State University")
df['winner']=df['winner'].replace("UT Martin","Tennessee Martin","University of Tennessee Martin")
df['winner']=df['winner'].replace("Southern University","Southern University & A&M College")
df['winner']=df['winner'].replace("Austin Peay","Austin Peay State University")
df['winner']=df['winner'].replace("Tennessee Tech","Tennessee Tech University")
df['winner']=df['winner'].replace("Jackson State","Jackson State University")
df['winner']=df['winner'].replace("Air Force","Air Force Academy")
df['winner']=df['winner'].replace("Bethune Cookman","Bethune Cookman University")
df['winner']=df['winner'].replace("Alabama State","Alabama State University")
df['winner']=df['winner'].replace("Pine Bluff","University of Arkansas Pine Bluff")
df['winner']=df['winner'].replace("Middle Tennessee","Middle Tennessee State University")
df['winner'].unique()
```
Out[67]:
```
array(['Tennessee State University', 'Eastern Kentucky University',
       'Southeast Missouri State University',
       'Alabama A&M University', 'Murray State University',
       'Eastern Illinois University', 'Samford University',
       'South Carolina State University',
       'North Carolina A&T State University',
       'University of Tennessee Martin', 'Vanderbilt University',
       'Southern University & A&M College',
       'Austin Peay State University', 'Tennessee Tech University',
       'Jackson State University', 'Air Force Academy',
       'Bethune Cookman University', 'Alabama State University',
       'University of Arkansas Pine Bluff',
       'Middle Tennessee State University'], dtype=object)
```

In [68]:
```
#sort rows by date
df=df.sort_values('date',ignore_index=True)
df.head()
```
Out[68]:

	date	city	state	winscore	loscore	loser	winner	locale	TSU score	opponent score	scorediff	scorediff_abs	winloss	year
0	2003-08-30	Nashville	TN	37	20	South Carolina State University	Tennessee State University	Home	37	20	17	17	Win	2003
1	2003-09-06	Huntsville	AL	31	24	Tennessee State University	Alabama A&M University	Away	24	31	-7	7	Loss	2003
2	2003-09-13	Memphis	TN	44	14	Jackson State University	Tennessee State University	Away	44	14	30	30	Win	2003
3	2003-09-20	Atlanta	GA	10	7	Tennessee State University	Florida A&M University	Away	7	10	-3	3	Loss	2003
4	2003-09-27	Nashville	TN	41	10	University of Tennessee Martin	Tennessee State University	Home	41	10	31	31	Win	2003

In [69]:
```
#get basic information on dataframe
df.info
```
Out[69]:
```
<bound method DataFrame.info of          date       city state  winscore  loscore  \
0   2003-08-30  Nashville    TN        37       20
1   2003-09-06 Huntsville    AL        31       24
2   2003-09-13    Memphis    TN        44       14
3   2003-09-20    Atlanta    GA        10        7
4   2003-09-27  Nashville    TN        41       10
..         ...        ...   ...       ...      ...
186 2019-10-19  Nashville    TN        26       24
187 2019-11-02  Nashville    TN        32       13
188 2019-11-09  Nashville    TN        49       38
189 2019-11-16     Martin    TN        28       17
190 2019-11-23 Cookeville    TN        37       27

                              loser                         winner  \
0    South Carolina State University     Tennessee State University
1         Tennessee State University         Alabama A&M University
2           Jackson State University     Tennessee State University
3         Tennessee State University         Florida A&M University
4    University of Tennessee Martin      Tennessee State University
..                              ...                            ...
186     Austin Peay State University     Tennessee State University
187       Tennessee State University Southeast Missouri State University
188       Tennessee State University     Eastern Illinois University
189       Tennessee State University University of Tennessee Martin
190       Tennessee Tech University      Tennessee State University

    locale  TSU score  opponent score  scorediff  scorediff_abs winloss  year
0     Home         37              20         17             17     Win  2003
1     Away         24              31         -7              7    Loss  2003
2     Away         44              14         30             30     Win  2003
3     Away          7              10         -3              3    Loss  2003
4     Home         41              10         31             31     Win  2003
..     ...        ...             ...        ...            ...     ...   ...
186   Home         26              24          2              2     Win  2019
187   Home         13              32        -19             19    Loss  2019
188   Home         38              49        -11             11    Loss  2019
189   Away         17              28        -11             11    Loss  2019
190   Away         37              27         10             10     Win  2019

[191 rows x 14 columns]>
```

In [70]:
```
#save data frame with scores to folder
df.to_csv("scores.csv",encoding="utf-8")
```

Web Scraping-getting individual years of stats data from TSU stats websites

Data Cleaning-correcting text and deleting extra text and spaces

Data Wrangling-putting data into lists and data frame

In [71]:
```
#2003 data
page = requests.get('https://tennstate_ftp.sidearmsports.com/custompages/tsutigers/A5FB3AA1-8151-496A-A022-754
mytree = html.fromstring(page.content)
```

In [72]:
```
#get date data
date = mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[1:120:10]
date=[a.replace("\xa0","") for a in date]
date=[a.replace("Sep. 6, 2003","Sep 06 2003") for a in date]
date=[a.replace(",","") for a in date]
print(date)
```
['Aug 30 2003', 'Sep 06 2003', 'Sep 13 2003', 'Sep 20 2003', 'Sep 27 2003', 'Oct 11 2003', 'Oct 18 2003', 'Oct 25 2003', 'Nov 01 2003', 'Nov 08 2003', 'Nov 15 2003', 'Nov 22 2003']

In [73]:
```
#get attendance data
attendance=mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[9:120:10]
attendance=[a.replace("\xa0","") for a in attendance]
attendance=[a.strip() for a in attendance]
attendance=pd.to_numeric(attendance)
attendance
```
Out[73]:
```
array([18124, 18085, 52603, 70185,  8434,  6127,  8023, 10360, 25037,
        5375,  3875, 28161], dtype=int64)
```

In [74]:
```
#get TSU rushing yards
TSUrushyards=mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[124:380:23]
TSUrushyards=[a.replace("\xa0","") for a in TSUrushyards]
TSUrushyards=pd.to_numeric(TSUrushyards)
TSUrushyards
```
Out[74]:
```
array([113, 141, 209, 153, 142, 182,  47, 132, 223, 149, 144, 191],
      dtype=int64)
```

In [75]:
```
#get TSU receiving yards
TSUreceiveyards=mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[128:400:23]
TSUreceiveyards=[a.replace("\xa0","") for a in TSUreceiveyards]
TSUreceiveyards=pd.to_numeric(TSUreceiveyards)
TSUreceiveyards
```
Out[75]:
```
array([365, 146, 132, 138, 285, 297, 175, 271, 184, 196, 371, 195],
      dtype=int64)
```

In [76]:
```
#get TSU kick return yards
TSUkreturnyards=mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[136:400:23]
TSUkreturnyards=[a.replace("\xa0","") for a in TSUkreturnyards]
TSUkreturnyards=pd.to_numeric(TSUkreturnyards)
TSUkreturnyards
```
Out[76]:
```
array([ 69, 147, 138,  86,  11,  33,  90,  50,   8, 139, 112,  63],
```

In [77]:
```
#get TSU punt return yards
TSUpreturnyards=mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[140:400:23]
TSUpreturnyards=[a.replace("\xa0","") for a in TSUpreturnyards]
TSUpreturnyards=pd.to_numeric(TSUpreturnyards)
TSUpreturnyards
```
Out[77]:
```
array([ 88,   8, 107,   2,  29,  52, -17,  45,   9,  18,   0,  15],
```

In [78]:
```
#get TSU total tackles
TSUtackles=mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[447:700:22]
TSUtackles=[a.replace("\xa0","") for a in TSUtackles]
TSUtackles=pd.to_numeric(TSUtackles)
TSUtackles
```
Out[78]:
```
array([69, 68, 84, 50, 76, 74, 79, 83, 65, 65, 75, 67], dtype=int64)
```

In [79]:
```
#get TSU tackle yards
TSUtackleyd=mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[449:700:22]
TSUtackleyd=[a.replace("\xa0","") for a in TSUtackleyd]
TSUtackleyd=pd.to_numeric(TSUtackleyd)
TSUtackleyd
```
Out[79]:
```
array([33, 22, 26, 22, 46, 46,  9, 36, 23, 33, 32, 37], dtype=int64)
```

In [80]:
```
#get TSU sacks
TSUsacks=mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[450:700:22]
TSUsacks=[a.replace("\xa0","") for a in TSUsacks]
TSUsacks=pd.to_numeric(TSUsacks)
TSUsacks
```
Out[80]:
```
array([3., 0., 1., 0., 4., 1., 2., 2., 4., 5., 0., 1.])
```

In [81]:
```
#get TSU sack yards
TSUsackyd=mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[451:700:22]
TSUsackyd=[a.replace("\xa0","") for a in TSUsackyd]
TSUsackyd=pd.to_numeric(TSUsackyd)
TSUsackyd
```
Out[81]:
```
array([21,  0,  4,  0, 22, 30,  9, 24, 12, 14, 20, 28], dtype=int64)
```

In [82]:
```
#TSU punts
TSUpunt=mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[753:969:19]
TSUpunt=[a.replace("\xa0","") for a in TSUpunt]
TSUpunt=pd.to_numeric(TSUpunt)
TSUpunt
```
Out[82]:
```
array([ 5,  7,  5,  6,  3,  2,  6,  8,  5,  5, 10,  7], dtype=int64)
```

In [83]:
```
#create data frame
list_of_dicts={'date':date,'attendance':attendance,'TSUrushyards':TSUrushyards, 'TSUreceiveyards':TSUreceiveyar
    'TSUkreturnyards':TSUkreturnyards,'TSUpreturnyards':TSUpreturnyards,'TSUtackles':TSUtackles,'TSUtackleyd':
    'TSUsacks':TSUsacks,'TSUsackyd':TSUsackyd,'TSUpunt':TSUpunt}
df2003=pd.DataFrame(list_of_dicts)
df2003.head()
```
Out[83]:

	date	attendance	TSUrushyards	TSUreceiveyards	TSUkreturnyards	TSUpreturnyards	TSUtackles	TSUtackleyd	TSUsacks	TSUsackyd	TSUp
0	Aug 30 2003	18124	113	365	69	88	69	33	3.0	21	
1	Sep 06 2003	18085	141	146	147	8	68	22	0.0	0	
2	Sep 13 2003	52603	209	132	138	107	84	26	1.0	4	
3	Sep 20 2003	70185	153	138	86	2	50	22	0.0	0	
4	Sep 27 2003	8434	142	285	11	29	76	46	4.0	22	

In [84]:
```
#creating date column as a datetime column
df2003['date']=pd.to_datetime(df2003['date'],format='%b %d %Y')
#creating year column
df2003['year']=df2003.date.dt.year
df2003.columns
```
Out[84]:
```
Index(['date', 'attendance', 'TSUrushyards', 'TSUreceiveyards',
       'TSUkreturnyards', 'TSUpreturnyards', 'TSUtackles', 'TSUtackleyd',
       'TSUsacks', 'TSUsackyd', 'TSUpunt', 'year'],
      dtype='object')
```

In [85]:
```
#get info on dataframe
df2003.info()
```
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 12 entries, 0 to 11
Data columns (total 12 columns):
 #   Column           Non-Null Count  Dtype
---  ------           --------------  -----
 0   date             12 non-null     datetime64[ns]
 1   attendance       12 non-null     int64
 2   TSUrushyards     12 non-null     int64
 3   TSUreceiveyards  12 non-null     int64
 4   TSUkreturnyards  12 non-null     int64
 5   TSUpreturnyards  12 non-null     int64
 6   TSUtackles       12 non-null     int64
 7   TSUtackleyd      12 non-null     int64
 8   TSUsacks         12 non-null     float64
 9   TSUsackyd        12 non-null     int64
 10  TSUpunt          12 non-null     int64
 11  year             12 non-null     int64
dtypes: datetime64[ns](1), float64(1), int64(10)
memory usage: 1.2 KB
```

In [86]:
```
#sort rows by date
df2003=df2003.sort_values('date',ignore_index=True)
df2003['date'].head()
```
Out[86]:
```
0   2003-08-30
1   2003-09-06
2   2003-09-13
3   2003-09-20
4   2003-09-27
Name: date, dtype: datetime64[ns]
```

In [87]:
```
#save data frame to folder
df2003.to_csv("2003.csv",encoding="utf-8")
```

In [88]:
```
#2004 data
page = requests.get('https://tennstate_ftp.sidearmsports.com/custompages/tsutigers/57C54C44-DE63-4F94-99EB-2BB
mytree = html.fromstring(page.content)
```

In [89]:
```
#get date data
date = mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[1:110:10]
date=[a.replace("\xa0","") for a in date]
date=[a.replace("2003","2004") for a in date]
date=[a.replace(",","") for a in date]
date=[a.replace("4" ","4" 04") for a in date]
print(date)
```
['Sep 04 2004', 'Sep 09 2004', 'Sep 18 2004', 'Sep 25 2004', 'Oct 02 2004', 'Oct 16 2004', 'Oct 23 2004', 'Oct 30 2004', 'Nov 06 2004', 'Nov 13 2004', 'Nov 20 2004']

In [90]:
```
#get attendance data
attendance=mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[9:110:10]
attendance=[a.replace("\xa0","") for a in attendance]
attendance=[a.strip() for a in attendance]
attendance=pd.to_numeric(attendance)
attendance
```
Out[90]:
```
array([25117,  7019, 55015, 67712, 51082, 14722,  5041,  5422, 23421,
        7200, 48848], dtype=int64)
```

In [91]:
```
#get TSU rushing yards
TSUrushyards=mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[114:357:23]
TSUrushyards=[a.replace("\xa0","") for a in TSUrushyards]
TSUrushyards=pd.to_numeric(TSUrushyards)
TSUrushyards
```
Out[91]:
```
array([152, 156, 164, 161,  22, 106, 290,  22, 243, 194,  83], dtype=int64)
```

In [92]:
```
#get TSU receiving yards
TSUreceiveyards=mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[118:360:23]
TSUreceiveyards=[a.replace("\xa0","") for a in TSUreceiveyards]
TSUreceiveyards=pd.to_numeric(TSUreceiveyards)
TSUreceiveyards
```
Out[92]:
```
array([203, 158, 149, 163, 173, 381, 211, 192, 215, 162, 142], dtype=int64)
```

In [93]:
```
#get TSU kick return yards
TSUkreturnyards=mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[126:360:23]
TSUkreturnyards=[a.replace("\xa0","") for a in TSUkreturnyards]
TSUkreturnyards=pd.to_numeric(TSUkreturnyards)
TSUkreturnyards
```
Out[93]:
```
array([ 24,  34,  39,  49,  57, 120, 127, 100,  76,  95, 130], dtype=int64)
```

In [94]:
```
#get TSU punt return yards
TSUpreturnyards=mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[130:380:23]
TSUpreturnyards=[a.replace("\xa0","") for a in TSUpreturnyards]
TSUpreturnyards=pd.to_numeric(TSUpreturnyards)
TSUpreturnyards
```
Out[94]:
```
array([26, 32, 10, 32,  0, 36, 77, 22, 63,  0, 0], dtype=int64)
```

In [95]:
```
#get TSU total tackles
TSUtackles=mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[414:650:22]
TSUtackles=[a.replace("\xa0","") for a in TSUtackles]
TSUtackles=pd.to_numeric(TSUtackles)
TSUtackles
```
Out[95]:
```
array([74, 65, 63, 78, 91, 77, 88, 66, 63, 67, 66], dtype=int64)
```

In [96]:
```
#get TSU tackle yards
TSUtackleyd=mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[416:650:22]
TSUtackleyd=[a.replace("\xa0","") for a in TSUtackleyd]
TSUtackleyd=pd.to_numeric(TSUtackleyd)
TSUtackleyd
```
Out[96]:
```
array([35, 60, 19, 70, 17,  0, 23,  6, 31, 25, 16], dtype=int64)
```

In [97]:
```
#get TSU sacks
TSUsacks=mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[417:650:22]
TSUsacks=[a.replace("\xa0","") for a in TSUsacks]
TSUsacks=pd.to_numeric(TSUsacks)
TSUsacks
```
Out[97]:
```
array([2., 4., 2., 5., 1., 0., 1., 0., 2., 1., 0.])
```

In [98]:
```
#get TSU sack yards
TSUsackyd=mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[418:650:22]
TSUsackyd=[a.replace("\xa0","") for a in TSUsackyd]
TSUsackyd=pd.to_numeric(TSUsackyd)
TSUsackyd
```
Out[98]:
```
array([12, 30,  7, 49,  4,  0,  5,  0, 12,  5,  0], dtype=int64)
```

In [99]:
```
#TSU punts
TSUpunt=mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[698:900:19]
TSUpunt=[a.replace("\xa0","") for a in TSUpunt]
TSUpunt=pd.to_numeric(TSUpunt)
TSUpunt
```
Out[99]:
```
array([4, 7, 7, 9, 9, 4, 6, 4, 7, 6, 8], dtype=int64)
```

In [100]:
```
#create data frame
list_of_dicts={'date':date,'attendance':attendance,'TSUrushyards':TSUrushyards, 'TSUreceiveyards':TSUreceiveyar
    'TSUkreturnyards':TSUkreturnyards,'TSUpreturnyards':TSUpreturnyards,'TSUtackles':TSUtackles,'TSUtackleyd':
    'TSUsacks':TSUsacks,'TSUsackyd':TSUsackyd,'TSUpunt':TSUpunt}
df2004=pd.DataFrame(list_of_dicts)
df2004.head()
```
Out[100]:

	date	attendance	TSUrushyards	TSUreceiveyards	TSUkreturnyards	TSUpreturnyards	TSUtackles	TSUtackleyd	TSUsacks	TSUsackyd	TSUp
0	Sep 04 2004	25117	152	203	24	26	74	35	2.0	12	
1	Sep 09 2004	7019	156	158	34	32	65	60	4.0	30	
2	Sep 18 2004	55015	164	149	39	10	63	19	2.0	7	
3	Sep 25 2004	67712	161	163	49	32	78	70	5.0	49	
4	Oct 02 2004	51082	22	173	57	0	91	17	1.0	4	

In [101]:
```
#creating date column as a datetime column
df2004['date']=pd.to_datetime(df2004['date'],format='%b %d %Y')
#creating year column
df2004['year']=df2004.date.dt.year
df2004.columns
```
Out[101]:
```
Index(['date', 'attendance', 'TSUrushyards', 'TSUreceiveyards',
       'TSUkreturnyards', 'TSUpreturnyards', 'TSUtackles', 'TSUtackleyd',
       'TSUsacks', 'TSUsackyd', 'TSUpunt', 'year'],
      dtype='object')
```

In [102]:
```
#get info on dataframe
df2004.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   TSUrushyards     11 non-null     int64
 3   TSUreceiveyards  11 non-null     int64
 4   TSUkreturnyards  11 non-null     int64
 5   TSUpreturnyards  11 non-null     int64
 6   TSUtackles       11 non-null     int64
 7   TSUtackleyd      11 non-null     int64
 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
```

In [103]:
```
#sort data
df2004=df2004.sort_values('date',ignore_index=True)
df2004['date'].head()
```
Out[103]:
```
0   2004-09-04
1   2004-09-09
2   2004-09-18
3   2004-09-25
4   2004-10-02
Name: date, dtype: datetime64[ns]
```

In [104]:
```
#save data frame to folder
df2004.to_csv("2004.csv",encoding="utf-8")
```

In [105]:
```
#2005 data
page = requests.get('https://tennstate_ftp.sidearmsports.com/custompages/tsutigers/99B728E2-12D4-4887-A990-249
mytree = html.fromstring(page.content)
```

In [106]:
```
#get date data
date = mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[1:110:10]
date=[a.replace("\xa0","") for a in date]
date=[a.replace(",","") for a in date]
date=[a.replace(".","") for a in date]
print(date)
```
['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']

In [107]:
```
#get attendance data
attendance=mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[9:110:10]
attendance=[a.replace("\xa0","") for a in attendance]
attendance=[a.strip() for a in attendance]
attendance=pd.to_numeric(attendance)
attendance
```
Out[107]:
```
array([25342, 48300,  5263, 56297, 42310, 10226,  6490,  8278, 23481,
        2512,  6779], dtype=int64)
```

In [108]:
```
#get TSU rushing yards
TSUrushyards=mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[114:357:23]
TSUrushyards=[a.replace("\xa0","") for a in TSUrushyards]
TSUrushyards=pd.to_numeric(TSUrushyards)
TSUrushyards
```
Out[108]:
```
array([ 72, 286,  77,  26, 117, 233, 157,  87,  50, 141, -18], dtype=int64)
```

In [109]:
```
#get TSU receiving yards
TSUreceiveyards=mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[118:360:23]
TSUreceiveyards=[a.replace("\xa0","") for a in TSUreceiveyards]
TSUreceiveyards=pd.to_numeric(TSUreceiveyards)
TSUreceiveyards
```
Out[109]:
```
array([154, 103, 170, 238, 115,  78, 185, 160, 196, 286, 150], dtype=int64)
```

In [110]:
```
#get TSU kick return yards
TSUkreturnyards=mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[126:360:23]
TSUkreturnyards=[a.replace("\xa0","") for a in TSUkreturnyards]
TSUkreturnyards=pd.to_numeric(TSUkreturnyards)
TSUkreturnyards
```
Out[110]:
```
array([ 62,  11, 110,  10,  99,  71, 114,  78,  90, 128, 138], dtype=int64)
```

In [111]:
```
#get TSU punt return yards
TSUpreturnyards=mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[130:380:23]
TSUpreturnyards=[a.replace("\xa0","") for a in TSUpreturnyards]
TSUpreturnyards=pd.to_numeric(TSUpreturnyards)
TSUpreturnyards
```
Out[111]:
```
array([32, 47, -2, 67, 49, 53, 14,  3,  0,  1,  0], dtype=int64)
```

In [112]:
```
#get TSU total tackles
TSUtackles=mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[414:650:22]
TSUtackles=[a.replace("\xa0","") for a in TSUtackles]
TSUtackles=pd.to_numeric(TSUtackles)
TSUtackles
```
Out[112]:
```
array([47, 59, 95, 70, 75, 89, 62, 72, 61, 53, 64], dtype=int64)
```

In [113]:
```
#get TSU tackle yards
TSUtackleyd=mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[416:650:22]
TSUtackleyd=[a.replace("\xa0","") for a in TSUtackleyd]
TSUtackleyd=pd.to_numeric(TSUtackleyd)
TSUtackleyd
```
Out[113]:
```
array([20, 29, 17, 56, 56, 41,  9, 20, 26, 20,  6], dtype=int64)
```

In [114]:
```
#get TSU sacks
TSUsacks=mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[417:650:22]
TSUsacks=[a.replace("\xa0","") for a in TSUsacks]
TSUsacks=pd.to_numeric(TSUsacks)
TSUsacks
```
Out[114]:
```
array([2., 3., 1., 7., 6., 4., 1., 2., 1., 2., 0.])
```

In [115]:
```
#get TSU sack yards
TSUsackyd=mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[418:650:22]
TSUsackyd=[a.replace("\xa0","") for a in TSUsackyd]
TSUsackyd=pd.to_numeric(TSUsackyd)
TSUsackyd
```
Out[115]:
```
array([13, 16,  6, 50, 37, 11,  0,  8, 14,  7,  0], dtype=int64)
```

In [116]:
```
#TSU punts
TSUpunt=mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[698:900:19]
TSUpunt=[a.replace("\xa0","") for a in TSUpunt]
TSUpunt=pd.to_numeric(TSUpunt)
TSUpunt
```
Out[116]:
```
array([4, 3, 6, 6, 6, 8, 5, 7, 5, 5, 8], dtype=int64)
```

In [117]:
```
#create data frame
list_of_dicts={'date':date,'attendance':attendance,'TSUrushyards':TSUrushyards, 'TSUreceiveyards':TSUreceiveyar
    'TSUkreturnyards':TSUkreturnyards,'TSUpreturnyards':TSUpreturnyards,'TSUtackles':TSUtackles,'TSUtackleyd':
    'TSUsacks':TSUsacks,'TSUsackyd':TSUsackyd,'TSUpunt':TSUpunt}
df2005=pd.DataFrame(list_of_dicts)
df2005.head()
```
Out[117]:

	date	attendance	TSUrushyards	TSUreceiveyards	TSUkreturnyards	TSUpreturnyards	TSUtackles	TSUtackleyd	TSUsacks	TSUsackyd	TSUp
0	Sep 03 2005	25342	72	154	62	32	47	20	2.0	13	
1	Sep 10 2005	48300	286	103	11	47	59	29	3.0	16	
2	Sep 17 2005	5263	77	170	110	-2	95	17	1.0	6	
3	Sep 24 2005	56297	26	238	10	67	70	56	7.0	50	
4	Oct 01 2005	42310	117	115	99	49	75	56	6.0	37	

In [118]:
```
#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.columns
```
Out[118]:
```
Index(['date', 'attendance', 'TSUrushyards', 'TSUreceiveyards',
       'TSUkreturnyards', 'TSUpreturnyards', 'TSUtackles', 'TSUtackleyd',
       'TSUsacks', 'TSUsackyd', 'TSUpunt', 'year'],
      dtype='object')
```

In [119]:
```
#get info about dataframe
df2005.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   TSUrushyards     11 non-null     int64
 3   TSUreceiveyards  11 non-null     int64
 4   TSUkreturnyards  11 non-null     int64
 5   TSUpreturnyards  11 non-null     int64
 6   TSUtackles       11 non-null     int64
 7   TSUtackleyd      11 non-null     int64
 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
```

In [120]:
```
#sort data
df2005=df2005.sort_values('date',ignore_index=True)
df2005['date'].head()
```
Out[120]:
```
0   2005-09-03
1   2005-09-10
2   2005-09-17
3   2005-09-24
4   2005-10-01
Name: date, dtype: datetime64[ns]
```
In [121]:
```
#save data
df2005.to_csv('2005.csv',encoding="utf-8")
```
In [122]:
```
#2006 data
page = requests.get('https://tennstate_ftp.sidearmsports.com/custompages/tsutigers/902C2D72-E881-4838-87F2-28BA
mytree = html.fromstring(page.content)
```
In [123]:
```
#get date data
date = mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[1:110:10]
date=[a.replace("\xa0","") for a in date]
date=[a.replace(",","") for a in date]
date=[a.replace(".","") for a in date]
print(date)
```
```
['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']
```
In [124]:
```
#get attendance data
attendance=mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[9:110:10]
attendance=[a.replace("\xa0","") for a in attendance]
attendance=[a.strip() for a in attendance]
attendance=pd.to_numeric(attendance)
attendance
```
Out[124]:
```
array([19487, 10613, 53441, 27460, 57885,  9720, 11800, 18758,  5912,
        4271,  5500], dtype=int64)
```
In [125]:
```
#get TSU rushing yards
TSUrushyards=mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[114:357:23]
TSUrushyards=[a.replace("\xa0","") for a in TSUrushyards]
TSUrushyards=pd.to_numeric(TSUrushyards)
TSUrushyards
```
Out[125]:
```
array([156, 111, 146,  71, 227, 310, 103, 247,  49, 133,  95], dtype=int64)
```
In [126]:
```
#get TSU receiving yards
TSUreceiveyards=mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[118:360:23]
TSUreceiveyards=[a.replace("\xa0","") for a in TSUreceiveyards]
TSUreceiveyards=pd.to_numeric(TSUreceiveyards)
TSUreceiveyards
```
Out[126]:
```
array([206, 230, 167, 181, 210, 169, 260, 105, 162, 152, 162], dtype=int64)
```
In [127]:
```
#get TSU kick return yards
TSUkreturnyards=mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[126:360:23]
TSUkreturnyards=[a.replace("\xa0","") for a in TSUkreturnyards]
TSUkreturnyards=pd.to_numeric(TSUkreturnyards)
TSUkreturnyards
```
Out[127]:
```
array([105, 102,  77,  70,  98,  45, 119,  23, 115,   0, 126], dtype=int64)
```
In [128]:
```
#get TSU punt return yards
TSUpreturnyards=mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[130:380:23]
TSUpreturnyards=[a.replace("\xa0","") for a in TSUpreturnyards]
TSUpreturnyards=pd.to_numeric(TSUpreturnyards)
TSUpreturnyards
```
Out[128]:
```
array([ 0,  0, 24,  0, 16,  4, 20, 13,  0, 17,  7], dtype=int64)
```
In [129]:
```
#get TSU total tackles
TSUtackles=mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[414:650:22]
TSUtackles=[a.replace("\xa0","") for a in TSUtackles]
TSUtackles=pd.to_numeric(TSUtackles)
TSUtackles
```
Out[129]:
```
array([63, 76, 63, 59, 54, 64, 77, 45, 88, 44, 70], dtype=int64)
```
In [130]:
```
#get TSU tackle yards
TSUtackleyd=mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[416:650:22]
TSUtackleyd=[a.replace("\xa0","") for a in TSUtackleyd]
TSUtackleyd=pd.to_numeric(TSUtackleyd)
TSUtackleyd
```
Out[130]:
```
array([ 3, 38, 18, 25, 14, 15, 26, 28, 22, 27, 30], dtype=int64)
```
In [131]:
```
#get TSU sacks
TSUsacks=mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[417:650:22]
TSUsacks=[a.replace("\xa0","") for a in TSUsacks]
TSUsacks=pd.to_numeric(TSUsacks)
TSUsacks
```
Out[131]:
```
array([0., 3., 1., 1., 1., 1., 3., 3., 1., 4., 2.])
```
In [132]:
```
#get TSU sack yards
TSUsackyd=mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[418:650:22]
TSUsackyd=[a.replace("\xa0","") for a in TSUsackyd]
TSUsackyd=pd.to_numeric(TSUsackyd)
TSUsackyd
```
Out[132]:
```
array([ 0, 23,  6,  7,  4,  8, 20, 16, 10, 23, 14], dtype=int64)
```
In [133]:
```
#TSU punts
TSUpunt=mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[698:900:19]
TSUpunt=[a.replace("\xa0","") for a in TSUpunt]
TSUpunt=pd.to_numeric(TSUpunt)
TSUpunt
```
Out[133]:
```
array([2, 5, 2, 3, 4, 4, 6, 3, 6, 4, 1], dtype=int64)
```
In [134]:
```
#create data frame
list_of_dicts={'date':date,'attendance':attendance,'TSUrushyards':TSUrushyards, 'TSUreceiveyards':TSUreceiveya
     'TSUkreturnyards':TSUkreturnyards,'TSUpreturnyards':TSUpreturnyards,'TSUtackles':TSUtackles,'TSUtackleyd':
     'TSUsacks':TSUsacks,'TSUsackyd':TSUsackyd,'TSUpunt':TSUpunt}
df2006=pd.DataFrame(list_of_dicts)
df2006.head()
```
Out[134]:

	date	attendance	TSUrushyards	TSUreceiveyards	TSUkreturnyards	TSUpreturnyards	TSUtackles	TSUtackleyd	TSUsacks	TSUsackyd
0	Sep 02 2006	19487	156	206	105	0	63	3	0.0	0
1	Sep 09 2006	10613	111	230	102	0	76	38	3.0	23
2	Sep 16 2006	53441	146	167	77	24	63	18	1.0	6
3	Sep 23 2006	27460	71	181	70	0	59	25	1.0	7
4	Sep 30 2006	57885	227	210	98	16	54	14	1.0	4

In [135]:
```
#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.columns
```
Out[135]:
```
Index(['date', 'attendance', 'TSUrushyards', 'TSUreceiveyards',
       'TSUkreturnyards', 'TSUpreturnyards', 'TSUtackles', 'TSUtackleyd',
       'TSUsacks', 'TSUsackyd', 'TSUpunt', 'year'],
      dtype='object')
```
In [136]:
```
#get info on data frame
df2006.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   TSUrushyards     11 non-null     int64
 3   TSUreceiveyards  11 non-null     int64
 4   TSUkreturnyards  11 non-null     int64
 5   TSUpreturnyards  11 non-null     int64
 6   TSUtackles       11 non-null     int64
 7   TSUtackleyd      11 non-null     int64
 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
```
In [137]:
```
#sort data
df2006=df2006.sort_values('date',ignore_index=True)
df2006['date'].head()
```
Out[137]:
```
0   2006-09-02
1   2006-09-09
2   2006-09-16
3   2006-09-23
4   2006-09-30
Name: date, dtype: datetime64[ns]
```
In [138]:
```
#save data
df2006.to_csv('2006.csv',encoding="utf-8")
```
In [139]:
```
#2007 data
page = requests.get('https://tennstate_ftp.sidearmsports.com/custompages/tsutigers/4C48DA5B-76BE-4378-8E95-9BB2
mytree = html.fromstring(page.content)
```
In [140]:
```
#get date data
date = mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[1:110:10]
date=[a.replace("\xa0","") for a in date]
date=[a.replace(",","") for a in date]
date=[a.replace(".","") for a in date]
print(date)
```
```
['Sep 01 2007', 'Sep 08 2007', 'Sep 15 2007', 'Sep 22 2007', 'Sep 29 2007', 'Oct 11 2007', 'Oct 20 2007', 'Oct
27 2007', 'Nov 03 2007', 'Nov 08 2007', 'Nov 17 2007']
```
In [141]:
```
#get attendance data
attendance=mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[9:110:10]
attendance=[a.replace("\xa0","") for a in attendance]
attendance=[a.strip() for a in attendance]
attendance=pd.to_numeric(attendance)
attendance
```
Out[141]:
```
array([23440, 50879,  8359, 15371, 56990,  9369, 11500,  8935, 24878,
        4193,  7859], dtype=int64)
```
In [142]:
```
#get TSU rushing yards
TSUrushyards=mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[114:357:23]
TSUrushyards=[a.replace("\xa0","") for a in TSUrushyards]
TSUrushyards=pd.to_numeric(TSUrushyards)
TSUrushyards
```
Out[142]:
```
array([206, 153, 238, 133, 177, 320, 169, 166, 280, 187, 290], dtype=int64)
```
In [143]:
```
#get TSU receiving yards
TSUreceiveyards=mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[118:360:23]
TSUreceiveyards=[a.replace("\xa0","") for a in TSUreceiveyards]
TSUreceiveyards=pd.to_numeric(TSUreceiveyards)
TSUreceiveyards
```
Out[143]:
```
array([176, 232, 233, 316, 309, 138, 127, 189,  69, 125, 173], dtype=int64)
```
In [144]:
```
#get TSU kick return yards
TSUkreturnyards=mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[126:360:23]
TSUkreturnyards=[a.replace("\xa0","") for a in TSUkreturnyards]
TSUkreturnyards=pd.to_numeric(TSUkreturnyards)
TSUkreturnyards
```
Out[144]:
```
array([154,  84,  83, 135, 111,  64, 122,  84,   6,  94, 222], dtype=int64)
```
In [145]:
```
#get TSU punt return yards
TSUpreturnyards=mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[130:380:23]
TSUpreturnyards=[a.replace("\xa0","") for a in TSUpreturnyards]
TSUpreturnyards=pd.to_numeric(TSUpreturnyards)
TSUpreturnyards
```
Out[145]:
```
array([35, 24, 14, 15, 33,  8,  0, 20, 77, -1, 31], dtype=int64)
```
In [146]:
```
#get TSU total tackles
TSUtackles=mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[414:650:22]
TSUtackles=[a.replace("\xa0","") for a in TSUtackles]
TSUtackles=pd.to_numeric(TSUtackles)
TSUtackles
```
Out[146]:
```
array([56, 63, 64, 80, 78, 70, 77, 88, 93, 75, 92], dtype=int64)
```
In [147]:
```
#get TSU tackle yards
TSUtackleyd=mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[416:650:22]
TSUtackleyd=[a.replace("\xa0","") for a in TSUtackleyd]
TSUtackleyd=pd.to_numeric(TSUtackleyd)
TSUtackleyd
```
Out[147]:
```
array([17, 30, 44, 56, 28, 23, 19, 12, 25, 35, 14], dtype=int64)
```
In [148]:
```
#get TSU sacks
TSUsacks=mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[417:650:22]
TSUsacks=[a.replace("\xa0","") for a in TSUsacks]
TSUsacks=pd.to_numeric(TSUsacks)
TSUsacks
```
Out[148]:
```
array([2, 2, 3, 3, 3, 0, 1, 0, 3, 1, 1], dtype=int64)
```
In [149]:
```
#get TSU sack yards
TSUsackyd=mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[418:650:22]
TSUsackyd=[a.replace("\xa0","") for a in TSUsackyd]
TSUsackyd=pd.to_numeric(TSUsackyd)
TSUsackyd
```
Out[149]:
```
array([15, 16, 26, 36, 23,  0, 10,  0, 15, 11,  8], dtype=int64)
```
In [150]:
```
#TSU punts
TSUpunt=mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[698:900:19]
TSUpunt=[a.replace("\xa0","") for a in TSUpunt]
TSUpunt=pd.to_numeric(TSUpunt)
TSUpunt
```
Out[150]:
```
array([ 6,  5,  3,  7,  4,  4, 10,  5,  3,  5,  3], dtype=int64)
```
In [151]:
```
#create data frame
list_of_dicts={'date':date,'attendance':attendance,'TSUrushyards':TSUrushyards, 'TSUreceiveyards':TSUreceiveya
     'TSUkreturnyards':TSUkreturnyards,'TSUpreturnyards':TSUpreturnyards,'TSUtackles':TSUtackles,'TSUtackleyd':
     'TSUsacks':TSUsacks,'TSUsackyd':TSUsackyd,'TSUpunt':TSUpunt}
df2007=pd.DataFrame(list_of_dicts)
df2007.head()
```
Out[151]:

	date	attendance	TSUrushyards	TSUreceiveyards	TSUkreturnyards	TSUpreturnyards	TSUtackles	TSUtackleyd	TSUsacks	TSUsackyd
0	Sep 01 2007	23440	206	176	154	35	56	17	2	15
1	Sep 08 2007	50879	153	232	84	24	63	30	2	16
2	Sep 15 2007	8359	238	233	83	14	64	44	3	26
3	Sep 22 2007	15371	133	316	135	15	80	56	3	36
4	Sep 29 2007	56990	177	309	111	33	78	28	3	23

In [152]:
```
#creating date column as a datetime column
df2007['date']=pd.to_datetime(df2007['date'],format='%b %d %Y')
#creating year column
df2007['year']=df2007.date.dt.year
df2007.columns
```
Out[152]:
```
Index(['date', 'attendance', 'TSUrushyards', 'TSUreceiveyards',
       'TSUkreturnyards', 'TSUpreturnyards', 'TSUtackles', 'TSUtackleyd',
       'TSUsacks', 'TSUsackyd', 'TSUpunt', 'year'],
      dtype='object')
```
In [153]:
```
#get info for data frame
df2007.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   TSUrushyards     11 non-null     int64
 3   TSUreceiveyards  11 non-null     int64
 4   TSUkreturnyards  11 non-null     int64
 5   TSUpreturnyards  11 non-null     int64
 6   TSUtackles       11 non-null     int64
 7   TSUtackleyd      11 non-null     int64
 8   TSUsacks         11 non-null     int64
 9   TSUsackyd        11 non-null     int64
 10  TSUpunt          11 non-null     int64
 11  year             11 non-null     int64
dtypes: datetime64[ns](1), int64(11)
memory usage: 1.2 KB
```
In [154]:
```
#sort values
df2007=df2007.sort_values('date',ignore_index=True)
df2007['date'].head()
```
Out[154]:
```
0   2007-09-01
1   2007-09-08
2   2007-09-15
3   2007-09-22
4   2007-09-29
Name: date, dtype: datetime64[ns]
```
In [155]:
```
#save data
df2007.to_csv('2007.csv',encoding='utf-8')
```
In [156]:
```
#2008 data
page = requests.get('https://tennstate_ftp.sidearmsports.com/custompages/tsutigers/E0564A83-37A0-4B85-8811-B7EE
mytree = html.fromstring(page.content)
```
In [157]:
```
#get date data
date = mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[1:120:10]
date=[a.replace("\xa0","") for a in date]
date=[a.replace(",","") for a in date]
date=[a.replace(".","") for a in date]
print(date)
```
```
['Aug 30 2008', 'Sep 06 2008', 'Sep 13 2008', 'Sep 20 2008', 'Sep 27 2008', 'Oct 04 2008', 'Oct 18 2008', 'Oct
25 2008', 'Nov 01 2008', 'Nov 08 2008', 'Nov 15 2008', 'Nov 22 2008']
```
In [158]:
```
#get attendance data
attendance=mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[9:120:10]
attendance=[a.replace("\xa0","") for a in attendance]
attendance=[a.strip() for a in attendance]
attendance=pd.to_numeric(attendance)
attendance
```
Out[158]:
```
array([10072, 28830, 50794,  8276, 50428,  7549,  9358,  9750, 24361,
        6393,  7956,  2292], dtype=int64)
```
In [159]:
```
#get TSU rushing yards
TSUrushyards=mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[124:380:23]
TSUrushyards=[a.replace("\xa0","") for a in TSUrushyards]
TSUrushyards=pd.to_numeric(TSUrushyards)
TSUrushyards
```
Out[159]:
```
array([107, 148, 148, 229, 136, 177, 377, 217, 212, 309, 177, 113],
      dtype=int64)
```
In [160]:
```
#get TSU receiving yards
TSUreceiveyards=mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[128:383:23]
TSUreceiveyards=[a.replace("\xa0","") for a in TSUreceiveyards]
TSUreceiveyards=pd.to_numeric(TSUreceiveyards)
TSUreceiveyards
```
Out[160]:
```
array([292, 315, 137, 314, 241, 155, 108, 233, 138, 162, 170, 210],
      dtype=int64)
```
In [161]:
```
#get TSU kick return yards
TSUkreturnyards=mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[136:406:23]
TSUkreturnyards=[a.replace("\xa0","") for a in TSUkreturnyards]
TSUkreturnyards=pd.to_numeric(TSUkreturnyards)
TSUkreturnyards
```
Out[161]:
```
array([160,  66,  34,  84,  95,  93, 101,  72,  32,  59,  57,  57],
      dtype=int64)
```
In [162]:
```
#get TSU punt return yards
TSUpreturnyards=mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[140:403:23]
TSUpreturnyards=[a.replace("\xa0","") for a in TSUpreturnyards]
TSUpreturnyards=pd.to_numeric(TSUpreturnyards)
TSUpreturnyards
```
Out[162]:
```
array([ 2, 37, 38,  6,  2, 21,  0, 18, 37,  4, 42, -2], dtype=int64)
```
In [163]:
```
#get TSU total tackles
TSUtackles=mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[447:694:22]
TSUtackles=[a.replace("\xa0","") for a in TSUtackles]
TSUtackles=pd.to_numeric(TSUtackles)
TSUtackles
```
Out[163]:
```
array([46, 76, 58, 51, 61, 74, 62, 54, 67, 68, 69, 60], dtype=int64)
```
In [164]:
```
#get TSU tackle yards
TSUtackleyd=mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[449:694:22]
TSUtackleyd=[a.replace("\xa0","") for a in TSUtackleyd]
TSUtackleyd=pd.to_numeric(TSUtackleyd)
TSUtackleyd
```
Out[164]:
```
array([20, 29,  9, 48, 37, 13, 25, 27, 30, 19, 10, 26], dtype=int64)
```
In [165]:
```
#get TSU sacks
TSUsacks=mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[450:694:22]
TSUsacks=[a.replace("\xa0","") for a in TSUsacks]
TSUsacks=pd.to_numeric(TSUsacks)
TSUsacks
```
Out[165]:
```
array([1, 1, 0, 3, 3, 0, 2, 2, 2, 2, 1, 2], dtype=int64)
```
In [166]:
```
#get TSU sack yards
TSUsackyd=mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[451:694:22]
TSUsackyd=[a.replace("\xa0","") for a in TSUsackyd]
TSUsackyd=pd.to_numeric(TSUsackyd)
TSUsackyd
```
Out[166]:
```
array([10,  4,  0, 27, 23,  0, 22, 15, 16, 13,  8, 11], dtype=int64)
```
In [167]:
```
#get TSU punts
TSUpunt=[]
TSUpunt.append(mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[753][0])
TSUpunt.append(mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[772][0])
TSUpunt.append(mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[791][0])
TSUpunt.append(mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[810][0])
TSUpunt.append(mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[829][0])
TSUpunt.append(mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[848][0])
TSUpunt.append(mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[867][0])
TSUpunt.append(mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[886][0])
TSUpunt.append(mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[905][0])
TSUpunt.append(mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[924][0])
TSUpunt.append(mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[943][0])
TSUpunt.append(mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[962][0])
TSUpunt=pd.to_numeric(TSUpunt)
TSUpunt
```
Out[167]:
```
array([3, 3, 3, 3, 5, 6, 1, 1, 4, 4, 7, 5], dtype=int64)
```
In [168]:
```
#create data frame
list_of_dicts={'date':date,'attendance':attendance,'TSUrushyards':TSUrushyards, 'TSUreceiveyards':TSUreceiveya
     'TSUkreturnyards':TSUkreturnyards,'TSUpreturnyards':TSUpreturnyards,'TSUtackles':TSUtackles,'TSUtackleyd':
     'TSUsacks':TSUsacks,'TSUsackyd':TSUsackyd,'TSUpunt':TSUpunt}
df2008=pd.DataFrame(list_of_dicts)
df2008.head()
```
Out[168]:

	date	attendance	TSUrushyards	TSUreceiveyards	TSUkreturnyards	TSUpreturnyards	TSUtackles	TSUtackleyd	TSUsacks	TSUsackyd
0	Aug 30 2008	10072	107	292	160	2	46	20	1	10
1	Sep 06 2008	28830	148	315	66	37	76	29	1	4
2	Sep 13 2008	50794	148	137	34	38	58	9	0	0
3	Sep 20 2008	8276	229	314	84	6	51	48	3	27
4	Sep 27 2008	50428	136	241	95	2	61	37	3	23

In [169]:
```
#creating date column as a datetime column
df2008['date']=pd.to_datetime(df2008['date'],format='%b %d %Y')
#creating year column
df2008['year']=df2008.date.dt.year
df2008.columns
```
Out[169]:
```
Index(['date', 'attendance', 'TSUrushyards', 'TSUreceiveyards',
       'TSUkreturnyards', 'TSUpreturnyards', 'TSUtackles', 'TSUtackleyd',
       'TSUsacks', 'TSUsackyd', 'TSUpunt', 'year'],
      dtype='object')
```
In [170]:
```
#get info on data frame
df2008.info()
```
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 12 entries, 0 to 11
Data columns (total 12 columns):
 #   Column           Non-Null Count  Dtype
---  ------           --------------  -----
 0   date             12 non-null     datetime64[ns]
 1   attendance       12 non-null     int64
 2   TSUrushyards     12 non-null     int64
 3   TSUreceiveyards  12 non-null     int64
 4   TSUkreturnyards  12 non-null     int64
 5   TSUpreturnyards  12 non-null     int64
 6   TSUtackles       12 non-null     int64
 7   TSUtackleyd      12 non-null     int64
 8   TSUsacks         12 non-null     int64
 9   TSUsackyd        12 non-null     int64
 10  TSUpunt          12 non-null     int64
 11  year             12 non-null     int64
dtypes: datetime64[ns](1), int64(11)
memory usage: 1.2 KB
```
In [171]:
```
#sort data
df2008=df2008.sort_values('date',ignore_index=True)
df2008['date'].head()
```
Out[171]:
```
0   2008-08-30
1   2008-09-06
2   2008-09-13
3   2008-09-20
4   2008-09-27
Name: date, dtype: datetime64[ns]
```
In [172]:
```
#save data
df2008.to_csv('2008.csv',encoding='utf-8')
```
In [173]:
```
#2009 data
page = requests.get('https://tennstate_ftp.sidearmsports.com/custompages/tsutigers/DEDBD687-051A-4E37-98F7-ACE
mytree = html.fromstring(page.content)
```
In [174]:
```
#get date data
date = mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[1:110:10]
date=[a.replace("\xa0","") for a in date]
date=[a.replace(",","") for a in date]
date=[a.replace(".","") for a in date]
print(date)
```
```
['Sep 05 2009', 'Sep 12 2009', 'Sep 19 2009', 'Sep 26 2009', 'Oct 03 2009', 'Oct 10 2009', 'Oct 17 2009', 'Oct
31 2009', 'Nov 07 2009', 'Nov 14 2009', 'Nov 19 2009']
```
In [175]:
```
#get attendance data
attendance=mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[9:110:10]
attendance=[a.replace("\xa0","") for a in attendance]
attendance=[a.strip() for a in attendance]
attendance=pd.to_numeric(attendance)
attendance
```
Out[175]:
```
array([23871, 43306, 12247, 51950,  6314,  7100,  5572,  7999, 22092,
        6968,  3509], dtype=int64)
```
In [176]:
```
#get TSU rushing yards
TSUrushyards=mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[114:357:23]
TSUrushyards=[a.replace("\xa0","") for a in TSUrushyards]
TSUrushyards=pd.to_numeric(TSUrushyards)
TSUrushyards
```
Out[176]:
```
array([ 13, 217, 169, 167, 259, 159, 169, 152,  62, 104, 176], dtype=int64)
```
In [177]:
```
#get TSU receiving yards
TSUreceiveyards=mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[118:360:23]
TSUreceiveyards=[a.replace("\xa0","") for a in TSUreceiveyards]
TSUreceiveyards=pd.to_numeric(TSUreceiveyards)
TSUreceiveyards
```
Out[177]:
```
array([178,  43,  86, 107,  95, 124,  68, 128, 207, 189, 135], dtype=int64)
```
In [178]:
```
#get TSU kick return yards
TSUkreturnyards=mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[126:360:23]
TSUkreturnyards=[a.replace("\xa0","") for a in TSUkreturnyards]
TSUkreturnyards=pd.to_numeric(TSUkreturnyards)
TSUkreturnyards
```
Out[178]:
```
array([46, 19, 75, 95, 72, 81, 64, 78, 97, 60, 55], dtype=int64)
```
In [179]:
```
#get TSU punt return yards
TSUpreturnyards=mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[130:383:23]
TSUpreturnyards=[a.replace("\xa0","") for a in TSUpreturnyards]
TSUpreturnyards=pd.to_numeric(TSUpreturnyards)
TSUpreturnyards
```
Out[179]:
```
array([ 0, 20, 13,  2, 30, 24,  5,  7,  7,  2, -4], dtype=int64)
```
In [180]:
```
#get TSU total tackles
TSUtackles=mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[414:644:22]
TSUtackles=[a.replace("\xa0","") for a in TSUtackles]
TSUtackles=pd.to_numeric(TSUtackles)
TSUtackles
```
Out[180]:
```
array([71, 60, 81, 63, 58, 52, 69, 56, 83, 73, 47], dtype=int64)
```
In [181]:
```
#get TSU tackle yards
TSUtackleyd=mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[416:646:22]
TSUtackleyd=[a.replace("\xa0","") for a in TSUtackleyd]
TSUtackleyd=pd.to_numeric(TSUtackleyd)
TSUtackleyd
```
Out[181]:
```
array([18, 42, 28, 14, 20, 24, 21, 35,  9,  6, 19], dtype=int64)
```
In [182]:
```
#get TSU sacks
TSUsacks=mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[417:646:22]
TSUsacks=[a.replace("\xa0","") for a in TSUsacks]
TSUsacks=pd.to_numeric(TSUsacks)
TSUsacks
```
Out[182]:
```
array([1, 3, 3, 1, 3, 3, 2, 2, 0, 0, 2], dtype=int64)
```
In [183]:
```
#get TSU sack yards
TSUsackyd=mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[418:646:22]
TSUsackyd=[a.replace("\xa0","") for a in TSUsackyd]
TSUsackyd=pd.to_numeric(TSUsackyd)
TSUsackyd
```
Out[183]:
```
array([ 7, 26, 15,  5, 15, 14, 12, 18,  0,  0, 11], dtype=int64)
```
In [184]:
```
#TSU punts
TSUpunt=mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[698:900:19]
TSUpunt=[a.replace("\xa0","") for a in TSUpunt]
TSUpunt=pd.to_numeric(TSUpunt)
TSUpunt
```
Out[184]:
```
array([6, 9, 8, 6, 5, 8, 6, 7, 7, 8, 8], dtype=int64)
```
In [185]:
```
#create data frame
#change dictionary of lists to data frame
list_of_dicts={'date':date,'attendance':attendance,'TSUrushyards':TSUrushyards, 'TSUreceiveyards':TSUreceiveya
     'TSUkreturnyards':TSUkreturnyards,'TSUpreturnyards':TSUpreturnyards,'TSUtackles':TSUtackles,'TSUtackleyd':
     'TSUsacks':TSUsacks,'TSUsackyd':TSUsackyd,'TSUpunt':TSUpunt}
df2009=pd.DataFrame(list_of_dicts)
df2009.head()
```
Out[185]:

	date	attendance	TSUrushyards	TSUreceiveyards	TSUkreturnyards	TSUpreturnyards	TSUtackles	TSUtackleyd	TSUsacks	TSUsackyd
0	Sep 05 2009	23871	13	178	46	0	71	18	1	7
1	Sep 12 2009	43306	217	43	19	20	60	42	3	26
2	Sep 19 2009	12247	169	86	75	13	81	28	3	15
3	Sep 26 2009	51950	167	107	95	2	63	14	1	5
4	Oct 03 2009	6314	259	95	72	30	58	20	3	15

In [186]:
```
#creating date column as a datetime column
df2009['date']=pd.to_datetime(df2009['date'],format='%b %d %Y')
#creating year column
df2009['year']=df2009.date.dt.year
df2009.columns
```
Out[186]:
```
Index(['date', 'attendance', 'TSUrushyards', 'TSUreceiveyards',
       'TSUkreturnyards', 'TSUpreturnyards', 'TSUtackles', 'TSUtackleyd',
       'TSUsacks', 'TSUsackyd', 'TSUpunt', 'year'],
      dtype='object')
```
In [187]:
```
#get info on data frame
df2009.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   TSUrushyards     11 non-null     int64
 3   TSUreceiveyards  11 non-null     int64
 4   TSUkreturnyards  11 non-null     int64
 5   TSUpreturnyards  11 non-null     int64
 6   TSUtackles       11 non-null     int64
 7   TSUtackleyd      11 non-null     int64
 8   TSUsacks         11 non-null     int64
 9   TSUsackyd        11 non-null     int64
 10  TSUpunt          11 non-null     int64
 11  year             11 non-null     int64
dtypes: datetime64[ns](1), int64(11)
memory usage: 1.2 KB
```
In [188]:
```
#sort data
df2009=df2009.sort_values('date',ignore_index=True)
df2009['date'].head()
```
Out[188]:
```
0   2009-09-05
1   2009-09-12
2   2009-09-19
3   2009-09-26
4   2009-10-03
Name: date, dtype: datetime64[ns]
```
In [189]:
```
#save data
df2009.to_csv('2009.csv',encoding='utf-8')
```
In [190]:
```
#2010 data
page = requests.get('https://tennstate_ftp.sidearmsports.com/custompages/tsutigers/108733F9-E3AE-4C3D-95EE-76C
mytree = html.fromstring(page.content)
```
In [191]:
```
#get date data
date = mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[1:110:10]
date=[a.replace("\xa0","") for a in date]
date=[a.replace(",","") for a in date]
date=[a.replace(".","") for a in date]
date=[a.replace("2 ","02 ") for a in date]
date=[a.replace("9 ","09 ") for a in date]
print(date)
```
```
['Sep 04 2010', 'Sep 11 2010', 'Sep 18 2010', 'Sep 25 2010', 'Oct 02 2010', 'Oct 09 2010', 'Oct 16 2010', 'Oct
23 2010', 'Nov 06 2010', 'Nov 13 2010', 'Nov 20 2010']
```
In [192]:
```
#get attendance data
attendance=mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[9:110:10]
attendance=[a.replace("\xa0","") for a in attendance]
attendance=[a.strip() for a in attendance]
attendance=pd.to_numeric(attendance)
attendance
```
Out[192]:
```
array([22607, 44688,  8502, 54202, 35217, 10316, 15218,  6739, 21596,
        3141,  2904], dtype=int64)
```
In [193]:
```
#get TSU rushing yards
TSUrushyards=mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[114:357:23]
TSUrushyards=[a.replace("\xa0","") for a in TSUrushyards]
TSUrushyards=pd.to_numeric(TSUrushyards)
TSUrushyards
```
Out[193]:
```
array([107, 224, 162, 304, 379, 152, 154, 158, 180,  72, 198], dtype=int64)
```
In [194]:
```
#get TSU receiving yards
TSUreceiveyards=mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[118:360:23]
TSUreceiveyards=[a.replace("\xa0","") for a in TSUreceiveyards]
TSUreceiveyards=pd.to_numeric(TSUreceiveyards)
TSUreceiveyards
```
Out[194]:
```
array([127, 185, 225, 109, 142, 147,  65, 203, 277,  51, 289], dtype=int64)
```
In [195]:
```
#get TSU kick return yards
TSUkreturnyards=mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[126:360:23]
TSUkreturnyards=[a.replace("\xa0","") for a in TSUkreturnyards]
TSUkreturnyards=pd.to_numeric(TSUkreturnyards)
TSUkreturnyards
```
Out[195]:
```
array([127, 133, 134,  33,  20,  90,  61,  56,  29, 116, 134], dtype=int64)
```
In [196]:
```
#get TSU punt return yards
TSUpreturnyards=mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[130:383:23]
TSUpreturnyards=[a.replace("\xa0","") for a in TSUpreturnyards]
TSUpreturnyards=pd.to_numeric(TSUpreturnyards)
TSUpreturnyards
```
Out[196]:
```
array([ 36,  28,  25, 142,  13,   0, -19,   1,  21,   2,   1], dtype=int64)
```
In [197]:
```
#get TSU total tackles
TSUtackles=mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[414:644:22]
TSUtackles=[a.replace("\xa0","") for a in TSUtackles]
TSUtackles=pd.to_numeric(TSUtackles)
TSUtackles
```
Out[197]:
```
array([64, 60, 57, 58, 66, 86, 66, 67, 66, 55, 65], dtype=int64)
```
In [198]:
```
#get TSU tackle yards
TSUtackleyd=mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[416:646:22]
TSUtackleyd=[a.replace("\xa0","") for a in TSUtackleyd]
TSUtackleyd=pd.to_numeric(TSUtackleyd)
TSUtackleyd
```
Out[198]:
```
array([42, 26, 24, 60, 34, 17, 11, 24, 19, 12, 30], dtype=int64)
```
In [199]:
```
#get TSU sacks
TSUsacks=mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[417:646:22]
TSUsacks=[a.replace("\xa0","") for a in TSUsacks]
TSUsacks=pd.to_numeric(TSUsacks)
TSUsacks
```
Out[199]:
```
array([7., 0., 2., 8., 5., 1., 0., 1., 1., 1., 1.])
```

In [200]:
```
TSUsackyd=mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[418:646:22]
TSUsackyd=[a.replace("\xa0","") for a in TSUsackyd]
TSUsackyd=pd.to_numeric(TSUsackyd)
TSUsackyd
```
Out[200]:
```
array([40,  0, 13, 48, 31,  9,  0,  4,  7,  7,  4], dtype=int64)
```
In [201]:
```
#TSU punts
TSUpunt=mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[698:900:19]
TSUpunt=[a.replace("\xa0","") for a in TSUpunt]
TSUpunt=pd.to_numeric(TSUpunt)
TSUpunt
```
Out[201]:
```
array([6, 6, 3, 5, 4, 5, 8, 3, 7, 9, 5], dtype=int64)
```
In [202]:
```
#create data frame
list_of_dicts={'date':date,'attendance':attendance,'TSUrushyards':TSUrushyards, 'TSUreceiveyards':TSUreceiveya
    'TSUkreturnyards':TSUkreturnyards,'TSUpreturnyards':TSUpreturnyards,'TSUtackles':TSUtackles,'TSUtackleyd':
    'TSUsacks':TSUsacks,'TSUsackyd':TSUsackyd,'TSUpunt':TSUpunt}
df2010=pd.DataFrame(list_of_dicts)
df2010.head()
```
Out[202]:

	date	attendance	TSUrushyards	TSUreceiveyards	TSUkreturnyards	TSUpreturnyards	TSUtackles	TSUtackleyd	TSUsacks	TSUsackyd	TSUp
0	Sep 04 2010	22607	107	127	127	36	64	42	7.0	40	
1	Sep 11 2010	44688	224	185	133	28	60	26	0.0	0	
2	Sep 18 2010	8502	162	225	134	25	57	24	2.0	13	
3	Sep 25 2010	54202	304	109	33	142	58	60	8.0	48	
4	Oct 02 2010	35217	379	142	20	13	66	34	5.0	31	

In [203]:
```
#creating date column as a datetime column
df2010['date']=pd.to_datetime(df2010['date'],format='%b %d %Y')
#creating year column
df2010['year']=df2010.date.dt.year
df2010.columns
```
Out[203]:
```
Index(['date', 'attendance', 'TSUrushyards', 'TSUreceiveyards',
       'TSUkreturnyards', 'TSUpreturnyards', 'TSUtackles', 'TSUtackleyd',
       'TSUsacks', 'TSUsackyd', 'TSUpunt', 'year'],
      dtype='object')
```
In [204]:
```
#get info for data frame
df2010.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   TSUrushyards     11 non-null     int64
 3   TSUreceiveyards  11 non-null     int64
 4   TSUkreturnyards  11 non-null     int64
 5   TSUpreturnyards  11 non-null     int64
 6   TSUtackles       11 non-null     int64
 7   TSUtackleyd      11 non-null     int64
 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
```
In [205]:
```
#sort data
df2010=df2010.sort_values('date',ignore_index=True)
df2010['date'].head()
```
Out[205]:
```
0   2010-09-04
1   2010-09-11
2   2010-09-18
3   2010-09-25
4   2010-10-02
Name: date, dtype: datetime64[ns]
```
In [206]:
```
#save data
df2010.to_csv('2010.csv',encoding='utf-8')
```
In [207]:
```
#2011 data
page = requests.get('https://tennstate_ftp.sidearmsports.com/custompages/tsutigers/CA835442-6528-4937-854B-D96
mytree = html.fromstring(page.content)
```
In [208]:
```
#get date data
date = mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[1:110:10]
date=[a.replace("\xa0","") for a in date]
date=[a.replace(",","") for a in date]
date=[a.replace(".","") for a in date]
print(date)
```
```
['Sep 03 2011', 'Sep 10 2011', 'Sep 17 2011', 'Sep 24 2011', 'Oct 01 2011', 'Oct 08 2011', 'Oct 15 2011', 'Oct
22 2011', 'Nov 05 2011', 'Nov 12 2011', 'Nov 19 2011']
```
In [209]:
```
#get attendance data
attendance=mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[9:110:10]
attendance=[a.replace("\xa0","") for a in attendance]
attendance=[a.strip() for a in attendance]
attendance=pd.to_numeric(attendance)
attendance
```
Out[209]:
```
array([25209, 43532, 10031, 33487,  8614,  6234,  8676, 10800,  6774,
       19537,  6137], dtype=int64)
```
In [210]:
```
#get TSU rushing yards
TSUrushyards=mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[114:357:23]
TSUrushyards=[a.replace("\xa0","") for a in TSUrushyards]
TSUrushyards=pd.to_numeric(TSUrushyards)
TSUrushyards
```
Out[210]:
```
array([342,  78, 168, 188, 162, 258, 166,  65, 200, 253,  87], dtype=int64)
```
In [211]:
```
#get TSU receiving yards
TSUreceiveyards=mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[118:360:23]
TSUreceiveyards=[a.replace("\xa0","") for a in TSUreceiveyards]
TSUreceiveyards=pd.to_numeric(TSUreceiveyards)
TSUreceiveyards
```
Out[211]:
```
array([170, 226, 207, 206, 304, 278, 232, 343, 118, 263, 262], dtype=int64)
```
In [212]:
```
#get TSU kick return yards
TSUkreturnyards=mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[126:360:23]
TSUkreturnyards=[a.replace("\xa0","") for a in TSUkreturnyards]
TSUkreturnyards=pd.to_numeric(TSUkreturnyards)
TSUkreturnyards
```
Out[212]:
```
array([ 48, 146, 197,  97, 158,  15, 118, 112,  64,  46, 129], dtype=int64)
```
In [213]:
```
#get TSU punt return yards
TSUpreturnyards=mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[130:383:23]
TSUpreturnyards=[a.replace("\xa0","") for a in TSUpreturnyards]
TSUpreturnyards=pd.to_numeric(TSUpreturnyards)
TSUpreturnyards
```
Out[213]:
```
array([  0,   5,  17,   0,   0, 115,   5,  17,   0,   5,  16], dtype=int64)
```
In [214]:
```
#get TSU total tackles
TSUtackles=mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[414:644:22]
TSUtackles=[a.replace("\xa0","") for a in TSUtackles]
TSUtackles=pd.to_numeric(TSUtackles)
TSUtackles
```
Out[214]:
```
array([45, 64, 94, 88, 61, 54, 88, 80, 59, 53, 50], dtype=int64)
```
In [215]:
```
#get TSU tackle yards
TSUtackleyd=mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[416:646:22]
TSUtackleyd=[a.replace("\xa0","") for a in TSUtackleyd]
TSUtackleyd=pd.to_numeric(TSUtackleyd)
TSUtackleyd
```
Out[215]:
```
array([41, 45, 32,  0, 20, 17, 38, 20, 23, 14, 18], dtype=int64)
```
In [216]:
```
#get TSU sacks
TSUsacks=mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[417:646:22]
TSUsacks=[a.replace("\xa0","") for a in TSUsacks]
TSUsacks=pd.to_numeric(TSUsacks)
TSUsacks
```
Out[216]:
```
array([5., 3., 3., 0., 2., 1., 5., 3., 3., 1., 2.])
```
In [217]:
```
#get TSU sack yards
TSUsackyd=mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[418:646:22]
TSUsackyd=[a.replace("\xa0","") for a in TSUsackyd]
TSUsackyd=pd.to_numeric(TSUsackyd)
TSUsackyd
```
Out[217]:
```
array([30, 31, 21,  0, 10,  9, 32, 14, 17,  1, 10], dtype=int64)
```
In [218]:
```
#TSU punts
TSUpunt=mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[698:900:19]
TSUpunt=[a.replace("\xa0","") for a in TSUpunt]
TSUpunt=pd.to_numeric(TSUpunt)
TSUpunt
```
Out[218]:
```
array([3, 6, 9, 8, 2, 2, 5, 4, 7, 5, 5], dtype=int64)
```
In [219]:
```
#create data frame
list_of_dicts={'date':date,'attendance':attendance,'TSUrushyards':TSUrushyards, 'TSUreceiveyards':TSUreceiveya
    'TSUkreturnyards':TSUkreturnyards,'TSUpreturnyards':TSUpreturnyards,'TSUtackles':TSUtackles,'TSUtackleyd':
    'TSUsacks':TSUsacks,'TSUsackyd':TSUsackyd,'TSUpunt':TSUpunt}
df2011=pd.DataFrame(list_of_dicts)
df2011.head()
```
Out[219]:

	date	attendance	TSUrushyards	TSUreceiveyards	TSUkreturnyards	TSUpreturnyards	TSUtackles	TSUtackleyd	TSUsacks	TSUsackyd	TSUp
0	Sep 03 2011	25209	342	170	48	0	45	41	5.0	30	
1	Sep 10 2011	43532	78	226	146	5	64	45	3.0	31	
2	Sep 17 2011	10031	168	207	197	17	94	32	3.0	21	
3	Sep 24 2011	33487	188	206	97	0	88	0	0.0	0	
4	Oct 01 2011	8614	162	304	158	0	61	20	2.0	10	

In [220]:
```
#creating date column as a datetime column
df2011['date']=pd.to_datetime(df2011['date'],format='%b %d %Y')
#creating year column
df2011['year']=df2011.date.dt.year
df2011.columns
```
Out[220]:
```
Index(['date', 'attendance', 'TSUrushyards', 'TSUreceiveyards',
       'TSUkreturnyards', 'TSUpreturnyards', 'TSUtackles', 'TSUtackleyd',
       'TSUsacks', 'TSUsackyd', 'TSUpunt', 'year'],
      dtype='object')
```
In [221]:
```
#get info from data frame
df2011.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   TSUrushyards     11 non-null     int64
 3   TSUreceiveyards  11 non-null     int64
 4   TSUkreturnyards  11 non-null     int64
 5   TSUpreturnyards  11 non-null     int64
 6   TSUtackles       11 non-null     int64
 7   TSUtackleyd      11 non-null     int64
 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
```
In [222]:
```
#sort data
df2011=df2011.sort_values('date',ignore_index=True)
df2011['date'].head()
```
Out[222]:
```
0   2011-09-03
1   2011-09-10
2   2011-09-17
3   2011-09-24
4   2011-10-01
Name: date, dtype: datetime64[ns]
```
In [223]:
```
#save data
df2011.to_csv('2011.csv',encoding='utf-8')
```
In [224]:
```
#2012 data
page = requests.get('https://tennstate_ftp.sidearmsports.com/custompages/tsutigers/BDF8C4F3-D7B6-4D50-847F-9227
mytree = html.fromstring(page.content)
```
In [225]:
```
#get date data
date = mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[1:110:10]
date=[a.replace("\xa0","") for a in date]
date=[a.replace(",","") for a in date]
date=[a.replace(".","") for a in date]
print(date)
```
```
['Sep 01 2012', 'Sep 08 2012', 'Sep 15 2012', 'Sep 22 2012', 'Sep 29 2012', 'Oct 05 2012', 'Oct 13 2012', 'Oct
20 2012', 'Oct 27 2012', 'Nov 03 2012', 'Nov 17 2012']
```
In [226]:
```
#get attendance data
attendance=mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[9:110:10]
attendance=[a.replace("\xa0","") for a in attendance]
attendance=[a.strip() for a in attendance]
attendance=pd.to_numeric(attendance)
attendance
```
Out[226]:
```
array([15652, 42257, 14264,  9461, 31765,  9878,  4800, 14867, 11373,
        3112,  6322], dtype=int64)
```
In [227]:
```
#get TSU rushing yards
TSUrushyards=mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[114:357:23]
TSUrushyards=[a.replace("\xa0","") for a in TSUrushyards]
TSUrushyards=pd.to_numeric(TSUrushyards)
TSUrushyards
```
Out[227]:
```
array([138, 235, 112, 200, 201, 139, 306, 184, 158, 138, 122], dtype=int64)
```
In [228]:
```
#get TSU receiving yards
TSUreceiveyards=mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[118:360:23]
TSUreceiveyards=[a.replace("\xa0","") for a in TSUreceiveyards]
TSUreceiveyards=pd.to_numeric(TSUreceiveyards)
TSUreceiveyards
```
Out[228]:
```
array([263, 137, 322, 157, 262, 227, 154, 311, 294, 273, 355], dtype=int64)
```
In [229]:
```
#get TSU kick return yards
TSUkreturnyards=mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[126:360:23]
TSUkreturnyards=[a.replace("\xa0","") for a in TSUkreturnyards]
TSUkreturnyards=pd.to_numeric(TSUkreturnyards)
TSUkreturnyards
```
Out[229]:
```
array([ 52,  69,  24,  41,  47,  63,  56,  62,  54, 195, 101], dtype=int64)
```
In [230]:
```
#get TSU punt return yards
TSUpreturnyards=mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[130:383:23]
TSUpreturnyards=[a.replace("\xa0","") for a in TSUpreturnyards]
TSUpreturnyards=pd.to_numeric(TSUpreturnyards)
TSUpreturnyards
```
Out[230]:
```
array([ 8, 42,  7, 37,  4,  0,  6,  0, 44,  0,  0], dtype=int64)
```
In [231]:
```
#get TSU total tackles
TSUtackles=mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[414:644:22]
TSUtackles=[a.replace("\xa0","") for a in TSUtackles]
TSUtackles=pd.to_numeric(TSUtackles)
TSUtackles
```
Out[231]:
```
array([49, 66, 55, 84, 62, 67, 74, 62, 48, 72, 79], dtype=int64)
```
In [232]:
```
#get TSU tackle yards
TSUtackleyd=mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[416:646:22]
TSUtackleyd=[a.replace("\xa0","") for a in TSUtackleyd]
TSUtackleyd=pd.to_numeric(TSUtackleyd)
TSUtackleyd
```
Out[232]:
```
array([28, 21, 13, 52, 40, 11, 21, 15, 13, 11, 24], dtype=int64)
```
In [233]:
```
#get TSU sacks
TSUsacks=mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[417:646:22]
TSUsacks=[a.replace("\xa0","") for a in TSUsacks]
TSUsacks=pd.to_numeric(TSUsacks)
TSUsacks
```
Out[233]:
```
array([1, 3, 1, 4, 3, 3, 2, 1, 1, 1, 1], dtype=int64)
```
In [234]:
```
#get TSU sack yards
TSUsackyd=mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[418:646:22]
TSUsackyd=[a.replace("\xa0","") for a in TSUsackyd]
TSUsackyd=pd.to_numeric(TSUsackyd)
TSUsackyd
```
Out[234]:
```
array([ 5, 20,  9, 30, 29,  6, 13,  9,  8,  7,  6], dtype=int64)
```
In [235]:
```
#TSU punts
TSUpunt=mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[698:900:19]
TSUpunt=[a.replace("\xa0","") for a in TSUpunt]
TSUpunt=pd.to_numeric(TSUpunt)
TSUpunt
```
Out[235]:
```
array([5, 8, 3, 7, 4, 5, 1, 5, 4, 6, 5], dtype=int64)
```
In [236]:
```
#create data frame
list_of_dicts={'date':date,'attendance':attendance,'TSUrushyards':TSUrushyards, 'TSUreceiveyards':TSUreceiveya
    'TSUkreturnyards':TSUkreturnyards,'TSUpreturnyards':TSUpreturnyards,'TSUtackles':TSUtackles,'TSUtackleyd':
    'TSUsacks':TSUsacks,'TSUsackyd':TSUsackyd,'TSUpunt':TSUpunt}
df2012=pd.DataFrame(list_of_dicts)
df2012.head()
```
Out[236]:

	date	attendance	TSUrushyards	TSUreceiveyards	TSUkreturnyards	TSUpreturnyards	TSUtackles	TSUtackleyd	TSUsacks	TSUsackyd	TSUp
0	Sep 01 2012	15652	138	263	52	8	49	28	1	5	
1	Sep 08 2012	42257	235	137	69	42	66	21	3	20	
2	Sep 15 2012	14264	112	322	24	7	55	13	1	9	
3	Sep 22 2012	9461	200	157	41	37	84	52	4	30	
4	Sep 29 2012	31765	201	262	47	4	62	40	3	29	

In [237]:
```
#creating date column as a datetime column
df2012['date']=pd.to_datetime(df2012['date'],format='%b %d %Y')
#creating year column
df2012['year']=df2012.date.dt.year
df2012.columns
```
Out[237]:
```
Index(['date', 'attendance', 'TSUrushyards', 'TSUreceiveyards',
       'TSUkreturnyards', 'TSUpreturnyards', 'TSUtackles', 'TSUtackleyd',
       'TSUsacks', 'TSUsackyd', 'TSUpunt', 'year'],
      dtype='object')
```
In [238]:
```
#get info for data frame
df2012.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   TSUrushyards     11 non-null     int64
 3   TSUreceiveyards  11 non-null     int64
 4   TSUkreturnyards  11 non-null     int64
 5   TSUpreturnyards  11 non-null     int64
 6   TSUtackles       11 non-null     int64
 7   TSUtackleyd      11 non-null     int64
 8   TSUsacks         11 non-null     int64
 9   TSUsackyd        11 non-null     int64
 10  TSUpunt          11 non-null     int64
 11  year             11 non-null     int64
dtypes: datetime64[ns](1), int64(11)
memory usage: 1.2 KB
```
In [239]:
```
#sort data
df2012=df2012.sort_values('date',ignore_index=True)
df2012['date'].head()
```
Out[239]:
```
0   2012-09-01
1   2012-09-08
2   2012-09-15
3   2012-09-22
4   2012-09-29
Name: date, dtype: datetime64[ns]
```
In [240]:
```
#save data
df2012.to_csv('2012.csv',encoding='utf-8')
```
In [241]:
```
#2013 data
page = requests.get('https://tennstate_ftp.sidearmsports.com/custompages/tsutigers/E394BBB1-387B-4C6A-9B21-E33
mytree = html.fromstring(page.content)
```
In [242]:
```
#get date data
date = mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[1:140:10]
date=[a.replace("\xa0","") for a in date]
date=[a.replace(",","") for a in date]
date=[a.replace(".","") for a in date]
print(date)
```
```
['Sep 01 2013', 'Sep 07 2013', 'Sep 14 2013', 'Sep 21 2013', 'Sep 28 2013', 'Oct 05 2013', 'Oct 12 2013', 'Oct
19 2013', 'Oct 26 2013', 'Nov 02 2013', 'Nov 16 2013', 'Nov 09 2013', 'Nov 30 2013', 'Dec 07 2013']
```
In [243]:
```
#get attendance data
attendance=mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[9:140:10]
attendance=[a.replace("\xa0","") for a in attendance]
attendance=[a.strip() for a in attendance]
attendance=pd.to_numeric(attendance)
attendance
```
Out[243]:
```
array([16108, 14237, 42400, 10044, 22000,  7374, 19092,  4166, 22157,
        5700,  6412,  5258,  1928,  4825], dtype=int64)
```
In [244]:
```
#get TSU rushing yards
TSUrushyards=mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[144:446:23]
TSUrushyards=[a.replace("\xa0","") for a in TSUrushyards]
TSUrushyards=pd.to_numeric(TSUrushyards)
TSUrushyards
```
Out[244]:
```
array([116, 268, 174,  95, 311, 264, 215,  69, 126,  71, 241, 146, 152,
        45], dtype=int64)
```
In [245]:
```
#get TSU receiving yards
TSUreceiveyards=mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[148:446:23]
TSUreceiveyards=[a.replace("\xa0","") for a in TSUreceiveyards]
TSUreceiveyards=pd.to_numeric(TSUreceiveyards)
TSUreceiveyards
```
Out[245]:
```
array([132, 131, 111, 343, 228, 280, 133, 170, 212, 101,  60, 173, 263,
       242], dtype=int64)
```
In [246]:
```
#get TSU kick return yards
TSUkreturnyards=mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[156:460:23]
TSUkreturnyards=[a.replace("\xa0","") for a in TSUkreturnyards]
TSUkreturnyards=pd.to_numeric(TSUkreturnyards)
TSUkreturnyards
```
Out[246]:
```
array([ 84, 111,  64,  63,  46,  24,  48,  64,  19,  96,  45,   8,  -9,
        52], dtype=int64)
```
In [247]:
```
#get TSU punt return yards
TSUpreturnyards=mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[160:460:23]
TSUpreturnyards=[a.replace("\xa0","") for a in TSUpreturnyards]
TSUpreturnyards=pd.to_numeric(TSUpreturnyards)
TSUpreturnyards
```
Out[247]:
```
array([51, 17, 11, 23, 75,  7, 46,  0, 56, -1, 19, 13, -3, 21],
      dtype=int64)
```
In [248]:
```
#get TSU total tackles
TSUtackles=mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[513:800:22]
TSUtackles=[a.replace("\xa0","") for a in TSUtackles]
TSUtackles=pd.to_numeric(TSUtackles)
TSUtackles
```
Out[248]:
```
array([62, 41, 69, 76, 70, 37, 70, 78, 70, 63, 54, 60, 47, 73],
      dtype=int64)
```
In [249]:
```
#get TSU tackle yards
TSUtackleyd=mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[515:822:22]
TSUtackleyd=[a.replace("\xa0","") for a in TSUtackleyd]
TSUtackleyd=pd.to_numeric(TSUtackleyd)
TSUtackleyd
```
Out[249]:
```
array([ 7, 35, 26, 51, 44, 12, 61, 56, 31,  3, 31, 59, 21, 13],
      dtype=int64)
```
In [250]:
```
#get TSU sacks
TSUsacks=mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[516:822:22]
TSUsacks=[a.replace("\xa0","") for a in TSUsacks]
TSUsacks=pd.to_numeric(TSUsacks)
TSUsacks
```
Out[250]:
```
array([1., 3., 2., 3., 3., 1., 5., 5., 3., 0., 3., 5., 2., 1.])
```
In [251]:
```
#get TSU sack yards
TSUsackyd=mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[517:822:22]
TSUsackyd=[a.replace("\xa0","") for a in TSUsackyd]
TSUsackyd=pd.to_numeric(TSUsackyd)
TSUsackyd
```
Out[251]:
```
array([ 4, 18, 13, 28, 17,  8, 18, 35, 27,  0, 20, 37,  7,  5],
      dtype=int64)
```
In [252]:
```
#TSU punts
TSUpunt=mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[863:1119:19]
TSUpunt=[a.replace("\xa0","") for a in TSUpunt]
TSUpunt=pd.to_numeric(TSUpunt)
TSUpunt
```
Out[252]:
```
array([5, 4, 3, 5, 1, 3, 6, 6, 8, 9, 8, 5, 4, 8], dtype=int64)
```
In [253]:
```
#create data frame
list_of_dicts={'date':date,'attendance':attendance,'TSUrushyards':TSUrushyards, 'TSUreceiveyards':TSUreceiveya
    'TSUkreturnyards':TSUkreturnyards,'TSUpreturnyards':TSUpreturnyards,'TSUtackles':TSUtackles,'TSUtackleyd':
    'TSUsacks':TSUsacks,'TSUsackyd':TSUsackyd,'TSUpunt':TSUpunt}
df2013=pd.DataFrame(list_of_dicts)
df2013.head()
```
Out[253]:

	date	attendance	TSUrushyards	TSUreceiveyards	TSUkreturnyards	TSUpreturnyards	TSUtackles	TSUtackleyd	TSUsacks	TSUsackyd	TSUp
0	Sep 01 2013	16108	116	132	84	51	62	7	1.0	4	
1	Sep 07 2013	14237	268	131	111	17	41	35	3.0	18	
2	Sep 14 2013	42400	174	111	64	11	69	26	2.0	13	
3	Sep 21 2013	10044	95	343	63	23	76	51	3.0	28	
4	Sep 28 2013	22000	311	228	46	75	70	44	3.0	17	

In [254]:
```
#creating date column as a datetime column
df2013['date']=pd.to_datetime(df2013['date'],format='%b %d %Y')
#creating year column
df2013['year']=df2013.date.dt.year
df2013.columns
```
Out[254]:
```
Index(['date', 'attendance', 'TSUrushyards', 'TSUreceiveyards',
       'TSUkreturnyards', 'TSUpreturnyards', 'TSUtackles', 'TSUtackleyd',
       'TSUsacks', 'TSUsackyd', 'TSUpunt', 'year'],
      dtype='object')
```
In [255]:
```
#get info on data frame
df2013.info()
```
```
<class 'pandas.core.frame.DataFrame'>
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   TSUreceiveyards  14 non-null     int64
 4   TSUkreturnyards  14 non-null     int64
 5   TSUpreturnyards  14 non-null     int64
 6   TSUtackles       14 non-null     int64
 7   TSUtackleyd      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
```
In [256]:
```
#sort data
df2013=df2013.sort_values('date',ignore_index=True)
df2013['date'].head()
```
Out[256]:
```
0   2013-09-01
1   2013-09-07
2   2013-09-14
3   2013-09-21
4   2013-09-28
Name: date, dtype: datetime64[ns]
```
In [257]:
```
#save data
df2013.to_csv('2013.csv',encoding='utf-8')
```
In [258]:
```
#2014 data
page = requests.get('https://tennstate_ftp.sidearmsports.com/custompages/tsutigers/26C45AE2-D036-475D-A4EF-38B
mytree = html.fromstring(page.content)
```
In [259]:
```
#get date data
date = mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[1:120:10]
date=[a.replace("\xa0","") for a in date]
date=[a.replace(",","") for a in date]
date=[a.replace(".","") for a in date]
date=[a.replace('10-04-14    ','Oct 04 2014') for a in date]
print(date)
```
```
['Aug 30 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']
```
In [260]:
```
#get attendance data
attendance=mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[9:120:10]
attendance=[a.replace("\xa0","") for a in attendance]
attendance=[a.strip() for a in attendance]
attendance=pd.to_numeric(attendance)
attendance
```
Out[260]:
```
array([10541, 15725, 46914,  9217, 29225,  8089,  5849,  6738,  8289,
        5052,  6143,  1962], dtype=int64)
```
In [261]:
```
#get TSU rushing yards
TSUrushyards=mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[124:400:23]
TSUrushyards=[a.replace("\xa0","") for a in TSUrushyards]
TSUrushyards=pd.to_numeric(TSUrushyards)
TSUrushyards
```
Out[261]:
```
array([439,  92, 137,  92, 125, 143,  43,  53, 107, 105, 111, 114],
      dtype=int64)
```
In [262]:
```
#get TSU receiving yards
TSUreceiveyards=mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[128:400:23]
TSUreceiveyards=[a.replace("\xa0","") for a in TSUreceiveyards]
TSUreceiveyards=pd.to_numeric(TSUreceiveyards)
TSUreceiveyards
```
Out[262]:
```
array([ 71, 340, 187, 113, 153, 362, 364, 314, 180, 326, 395, 463],
      dtype=int64)
```
In [263]:
```
#get TSU kick return yards
TSUkreturnyards=mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[136:400:23]
TSUkreturnyards=[a.replace("\xa0","") for a in TSUkreturnyards]
TSUkreturnyards=pd.to_numeric(TSUkreturnyards)
TSUkreturnyards
```
Out[263]:
```
array([ 77, 111,  29,   0,  24,  53,  63,  32, 136, 198,  58,  48],
      dtype=int64)
```
In [264]:
```
#get TSU punt return yards
TSUpreturnyards=mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[140:400:23]
TSUpreturnyards=[a.replace("\xa0","") for a in TSUpreturnyards]
TSUpreturnyards=pd.to_numeric(TSUpreturnyards)
TSUpreturnyards
```
Out[264]:
```
array([57, 22, 40, 15, 56, 24, 12, -1,  1,  0, -2,  0], dtype=int64)
```
In [265]:
```
#get TSU total tackles
TSUtackles=mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[447:723:23]
TSUtackles=[a.replace("\xa0","") for a in TSUtackles]
TSUtackles=pd.to_numeric(TSUtackles)
TSUtackles
```
Out[265]:
```
array([53, 89, 54, 48, 64, 81, 75, 73, 54, 84, 85, 81], dtype=int64)
```
In [266]:
```
#get TSU tackle yards
TSUtackleyd=mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[449:711:23]
TSUtackleyd=[a.replace("\xa0","") for a in TSUtackleyd]
TSUtackleyd=pd.to_numeric(TSUtackleyd)
TSUtackleyd
```
Out[266]:
```
array([38, 28, 40, 31, 51,  9, 21, 48, 17, 19, 29, 63], dtype=int64)
```
In [267]:
```
#get TSU sacks
TSUsacks=mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[450:711:23]
TSUsacks=[a.replace("\xa0","") for a in TSUsacks]
TSUsacks=pd.to_numeric(TSUsacks)
TSUsacks
```
Out[267]:
```
array([2., 3., 7., 6., 7., 0., 2., 3., 1., 2., 2., 8.])
```
In [268]:
```
#get TSU sack yards
TSUsackyd=mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[451:711:23]
TSUsackyd=[a.replace("\xa0","") for a in TSUsackyd]
TSUsackyd=pd.to_numeric(TSUsackyd)
TSUsackyd
```
Out[268]:
```
array([13, 13, 33, 28, 44,  0, 13, 22,  9,  4, 15, 51], dtype=int64)
```
In [269]:
```
#TSU punts
TSUpunt=mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[767:981:19]
TSUpunt=[a.replace("\xa0","") for a in TSUpunt]
TSUpunt=pd.to_numeric(TSUpunt)
TSUpunt
```
Out[269]:
```
array([1, 8, 8, 9, 7, 5, 3, 8, 6, 5, 0, 6], dtype=int64)
```
In [270]:
```
#create data frame
list_of_dicts={'date':date,'attendance':attendance,'TSUrushyards':TSUrushyards, 'TSUreceiveyards':TSUreceiveya
    'TSUkreturnyards':TSUkreturnyards,'TSUpreturnyards':TSUpreturnyards,'TSUtackles':TSUtackles,'TSUtackleyd':
    'TSUsacks':TSUsacks,'TSUsackyd':TSUsackyd,'TSUpunt':TSUpunt}
df2014=pd.DataFrame(list_of_dicts)
df2014.head()
```
Out[270]:

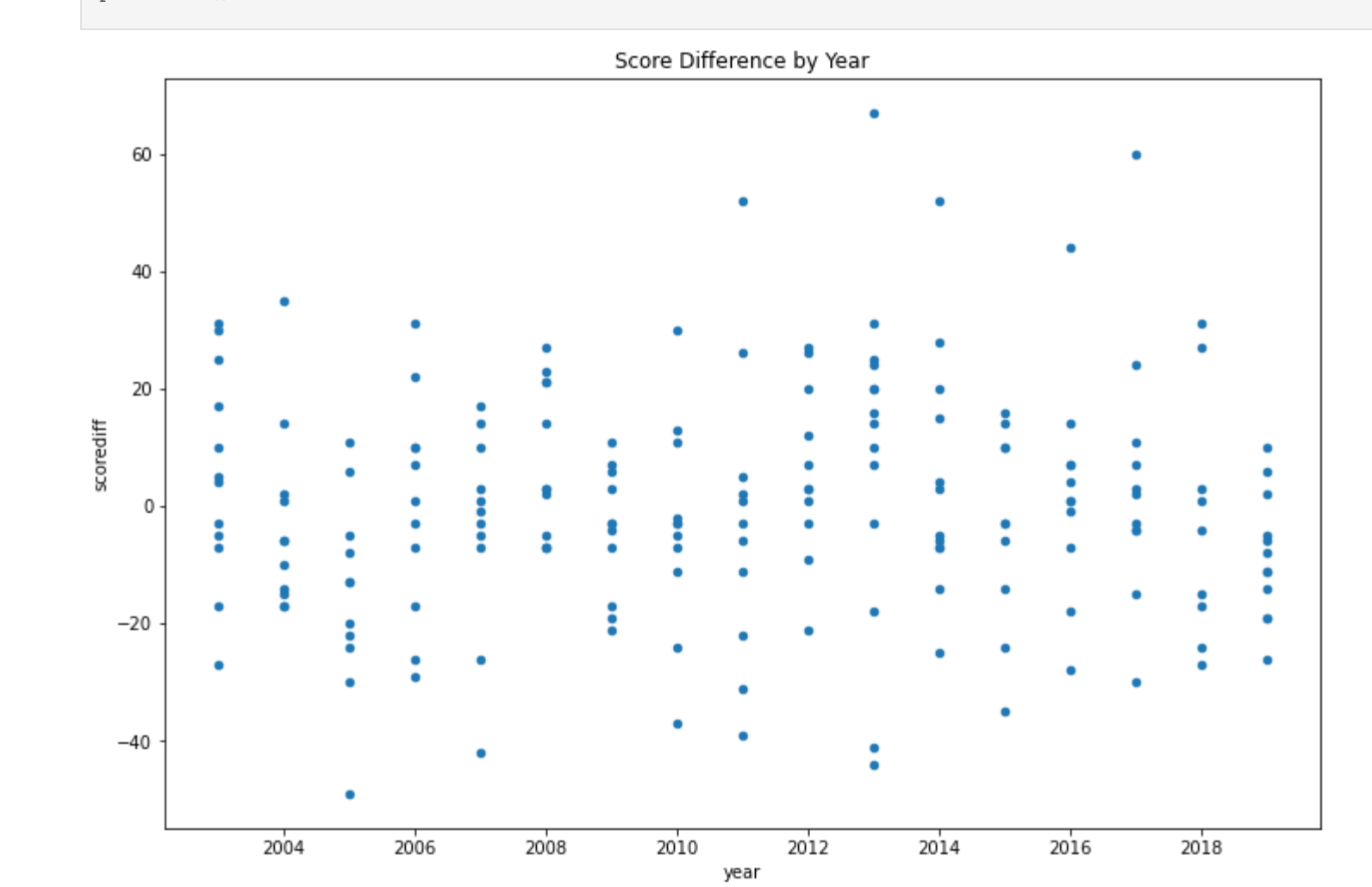
	date	attendance	TSUrushyards	TSUreceiveyards	TSUkreturnyards	TSUpreturnyards	TSUtackles	TSUtackleyd	TSUsacks	TSUsackyd	TSUp
0	Aug 30 2014	10541	439	71	77	57	53	38	2.0	13	
1	Sep 06 2014	15725	92	340	111	22	89	28	3.0	13	
2	Sep 13 2014	46914	137	187	29	40	54	40	7.0	33	
3	Sep 20 2014	9217	92	113	0	15	48	31	6.0	28	
4	Sep 27 2014	29225	125	153	24	56	64	51	7.0	44	

In [271]:
```
#creating date column as a datetime column
df2014['date']=pd.to_datetime(df2014['date'],format='%b %d %Y')
#creating year column
df2014['year']=df2014.date.dt.year
df2014.columns
```
Out[271]:
```
Index(['date', 'attendance', 'TSUrushyards', 'TSUreceiveyards',
       'TSUkreturnyards', 'TSUpreturnyards', 'TSUtackles', 'TSUtackleyd',
       'TSUsacks', 'TSUsackyd', 'TSUpunt', 'year'],
      dtype='object')
```
In [272]:
```
#get info on data frame
df2014.info()
```
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 12 entries, 0 to 11
Data columns (total 12 columns):
 #   Column           Non-Null Count  Dtype
---  ------           --------------  -----
 0   date             12 non-null     datetime64[ns]
 1   attendance       12 non-null     int64
 2   TSUrushyards     12 non-null     int64
 3   TSUreceiveyards  12 non-null     int64
 4   TSUkreturnyards  12 non-null     int64
 5   TSUpreturnyards  12 non-null     int64
 6   TSUtackles       12 non-null     int64
 7   TSUtackleyd      12 non-null     int64
 8   TSUsacks         12 non-null     float64
 9   TSUsackyd        12 non-null     int64
 10  TSUpunt          12 non-null     int64
 11  year             12 non-null     int64
dtypes: datetime64[ns](1), float64(1), int64(10)
memory usage: 1.2 KB
```
In [273]:
```
#sort data
df2014=df2014.sort_values('date',ignore_index=True)
df2014['date'].head()
```
Out[273]:
```
0   2014-08-30
1   2014-09-06
2   2014-09-13
3   2014-09-20
4   2014-09-27
Name: date, dtype: datetime64[ns]
```
In [274]:
```
#save data
df2014.to_csv('2014.csv',encoding='utf-8')
```
In [275]:
```
#2015 data
page = requests.get('https://tennstate_ftp.sidearmsports.com/custompages/tsutigers/88713968-6888-4884-B12E-9B6
mytree = html.fromstring(page.content)
```
In [276]:
```
#get date data
date = mytree.xpath('body/center//tr/td/font[@color="#000000"]/text()')[1:100:10]
date=[a.replace("\xa0","") for a in date]
date=[a.replace(",","") for a in date]
date=[a.replace(".","") for a in date]
print(date)
```
```
['Sep 06 2015', 'Sep 12 2015', 'Sep 19 2015', 'Sep 26 2015', 'Oct 10 2015', 'Oct 17 2015', 'Oct 24 2015', 'Oct
31 2015', 'Nov 07 2015', 'Nov 21 2015']
```

```
count    191.000000
mean      25.427340
std       12.664186
min        0.000000
25%       15.000000
50%       27.000000
75%       33.000000
max       58.000000
Name: opponent score_new, dtype: float64
```

In [403]:

```
#score difference by year
#The scatterplot of the score difference showed that it remained between -20 and 40 for most games. It also sho...
plt.rcParams["figure.figsize"]=[12,8]
mydata.plot(x="year", y="scorediff",kind="scatter", title="Score Difference by Year")
plt.show()
```



In [404]:

```
#get outliers of score difference
#There were 7 outliers on the score difference variable. 2 were games in which TSU lost to Easter Kentucky Univ...
#at least 44 points. The other 5 games where those in which TSU won by at least 44 points. A new variable was c...
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"]<out...
```

Out[404]:

	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
148	2016-09-03	Nashville	University of Arkansas Pine Bluff	44
167	2017-11-04	Nashville	Virginia University Lynchburg	60

In [405]:

```
#replace outliers of score difference with mean
mydata["scorediff_new"]=np.where((mydata["scorediff"]>outlier_top_lim) | (mydata["scorediff"]<outlier_bottom_li...
```

In [406]:

```
#descriptive statistics for score difference
mydata.scorediff.describe()
```

Out[406]:

```
count    191.000000
mean      -0.178010
std       19.109373
min      -49.000000
25%      -11.000000
50%       -3.000000
75%       10.000000
max       67.000000
Name: scorediff, dtype: float64
```

In [407]:

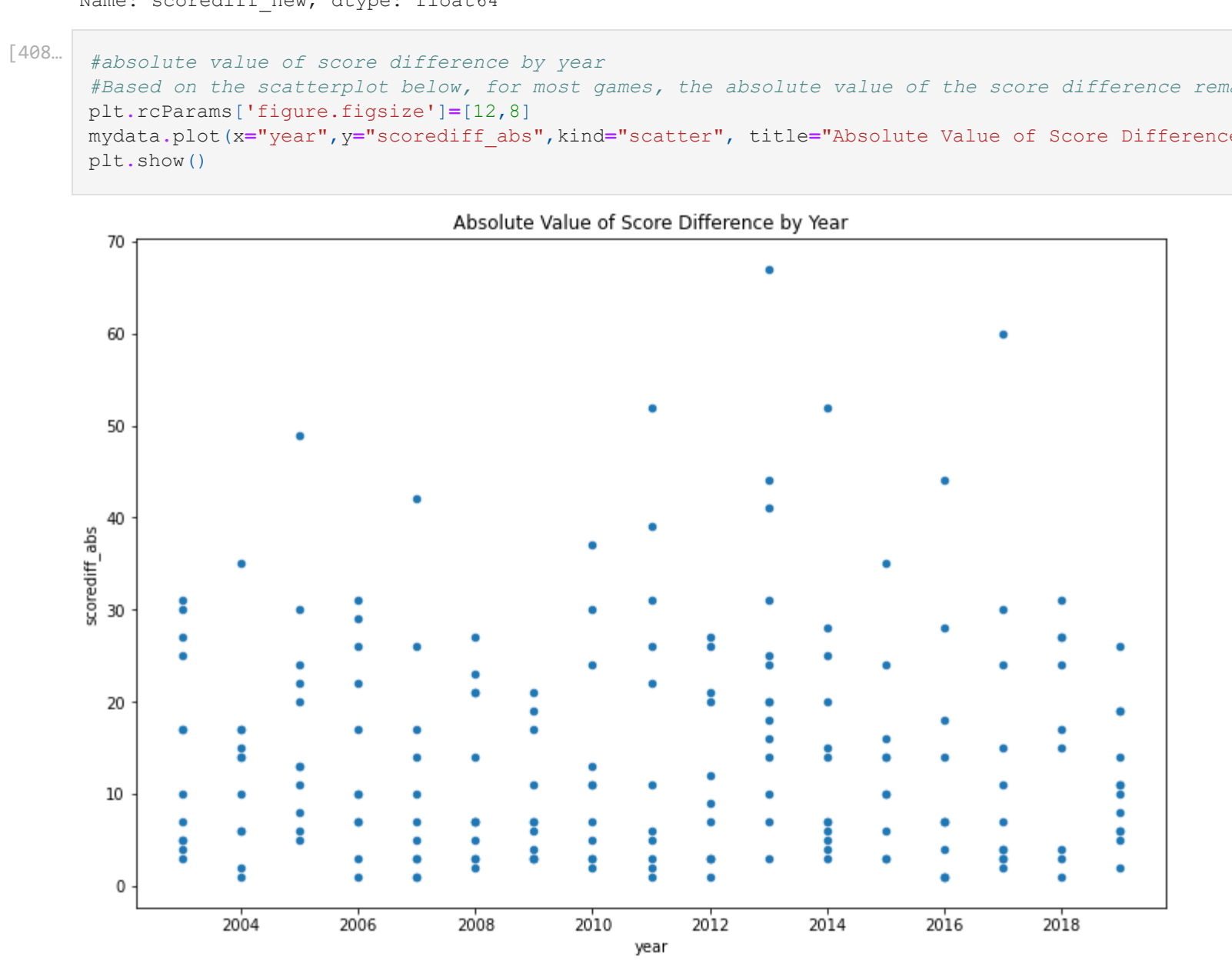
```
#descriptive statistics for score difference with outliers replaced
#With the outliers replaced with the mean, the mean of the score difference dropped to -1.1 and the standard dev...
mydata.scorediff_new.describe()
```

Out[407]:

```
count    191.000000
mean      -1.137414
std       16.119802
min      -42.000000
25%      -11.000000
50%       -3.000000
75%       10.000000
max       35.000000
Name: scorediff_new, dtype: float64
```

In [408]:

```
#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 ...
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()
```



In [409]:

```
#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...
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["scoredi...
```

Out[409]:

	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
167	2017-11-04	Nashville	Virginia University Lynchburg	60

In [410]:

```
#replace outliers of absolute value of score difference with mean
mydata["scorediff_abs_new"]=np.where((mydata["scorediff_abs"]>outlier_top_lim) | (mydata["scorediff_abs"]<outli...
```

In [411]:

```
#descriptive statistics for absolute value of score difference
mydata.scorediff_abs.describe()
```

Out[411]:

```
count    191.000000
mean      14.502618
std       12.400286
min        1.000000
25%        5.000000
50%       11.000000
75%       21.500000
max       67.000000
Name: scorediff_abs, dtype: float64
```

In [412]:

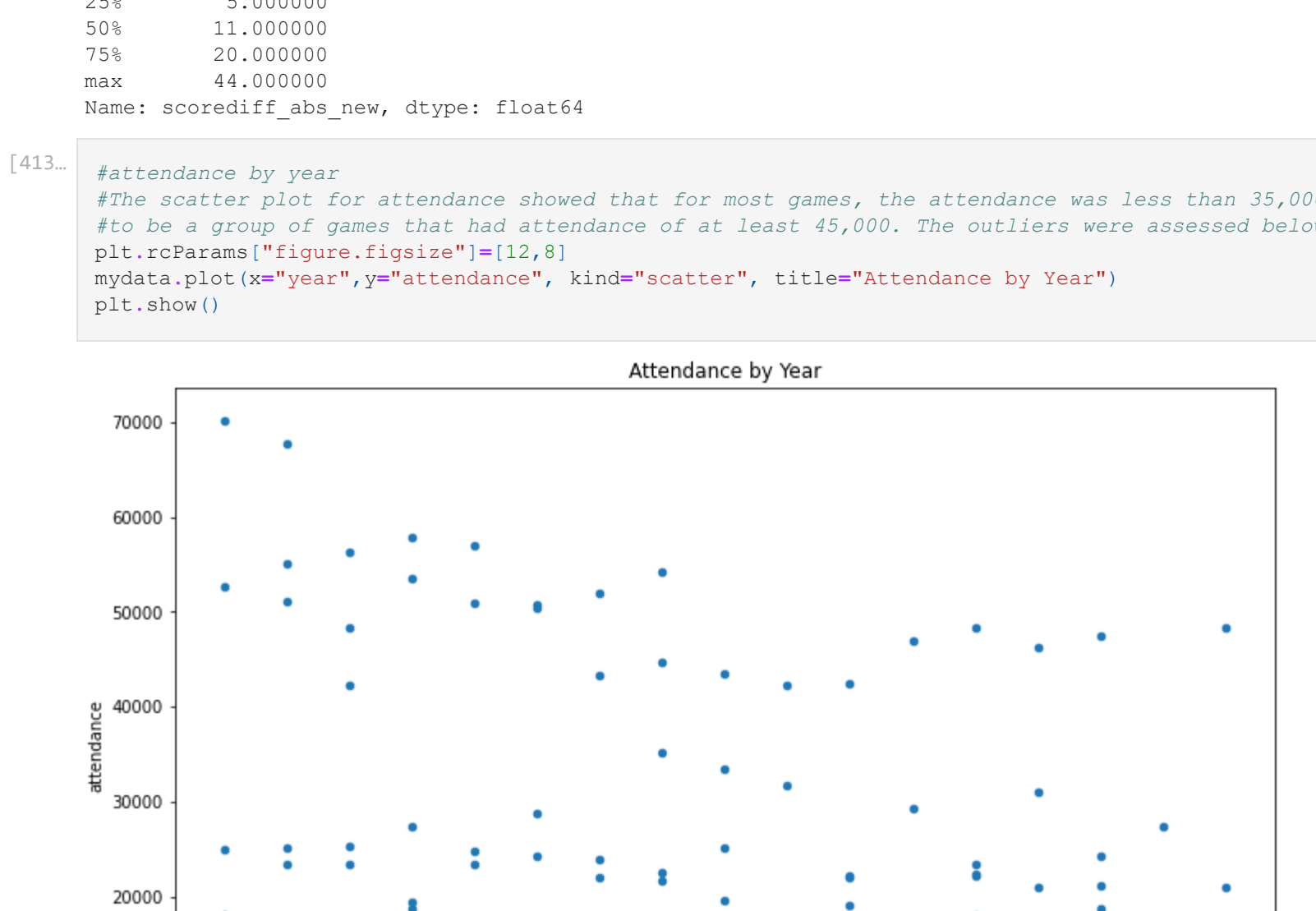
```
#descriptive statistics for absolute value of score difference with outliers replaced
#As with the previous variables that replaced the outliers with the mean, the mean and standard deviation of th...
mydata.scorediff_abs_new.describe()
```

Out[412]:

```
count    191.000000
mean      13.416299
std       10.301300
min        1.000000
25%        5.000000
50%       11.000000
75%       20.000000
max       44.000000
Name: scorediff_abs_new, dtype: float64
```

In [413]:

```
#attendance by year
#The scatter plot for attendance showed that for most games, the attendance was less than 35,000. However, ther...
#to be a group of games that had attendance 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()
```



In [414]:

```
#get outliers of attendance
#There were 20 outliers for attendance. These games had attendance figures of at least about 47,000. These game...
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"]<...
```

Out[414]:

	date	city	opponent	attendance
2	2003-09-13	Memphis	Jackson State University	52603
13	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	51082
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

In [415]:

```
#replace outliers of attendance with mean
mydata["attendance_new"]=np.where((mydata["attendance"]>outlier_top_lim) | (mydata["attendance"]<outlier_botto...
```

In [416]:

```
#descriptive statistics for attendance
mydata.attendance.describe()
```

Out[416]:

```
count      191.000000
mean     16985.230366
std      15493.766885
min       1776.000000
25%       6402.500000
50%      10001.000000
75%      22306.000000
max      70185.000000
Name: attendance, dtype: float64
```

In [417]:

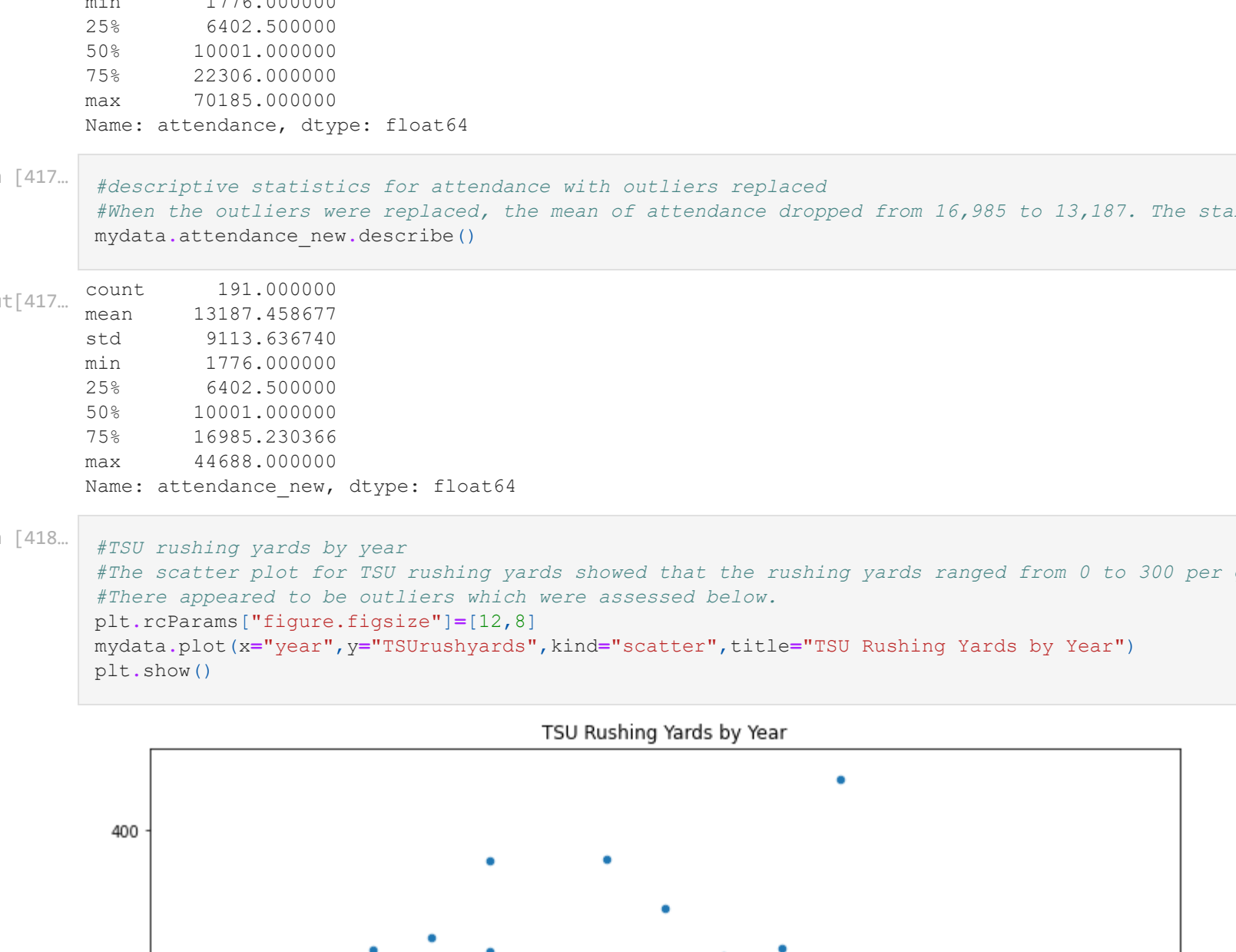
```
#descriptive statistics for attendance with outliers replaced
#When the outliers were replaced, the mean of attendance dropped from 16,985 to 13,187. The standard deviation ...
mydata.attendance_new.describe()
```

Out[417]:

```
count      191.000000
mean     13187.458677
std       9113.636740
min       1776.000000
25%       6402.500000
50%      10001.000000
75%      16985.230366
max      44688.000000
Name: attendance_new, dtype: float64
```

In [418]:

```
#TSU rushing yards by year
#The scatter plot for TSU rushing yards showed that the rushing yards ranged from 0 to 300 per game.
#There appeared to be outliers which were assessed below.
plt.rcParams["figure.figsize"]=[12,8]
mydata.plot(x="year",y="TSUrushyards",kind="scatter",title="TSU Rushing Yards by Year")
plt.show()
```



In [419]:

```
#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["TSUrushyards"].quantile(0.25)
q3 = mydata["TSUrushyards"].quantile(0.75)
outlier_top_lim = q3 + 1.5 * (q3 - q1)
outlier_bottom_lim = q1 - 1.5 * (q3 - q1)
mydata[["date","city","opponent","TSUrushyards"]][(mydata["TSUrushyards"]>outlier_top_lim) | (mydata["TSUrushy...
```

Out[419]:

	date	city	opponent	TSUrushyards
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

In [420]:

```
#replace outliers of TSU rush yards with mean
mydata["TSUrushyards_new"]=np.where((mydata["TSUrushyards"]>outlier_top_lim) | (mydata["TSUrushyards"]<outlier_...
```

In [421]:

```
#descriptive statistics for TSU rushing yards
mydata.TSUrushyards.describe()
```

Out[421]:

```
count    191.000000
mean     157.769634
std       74.855646
min      -18.000000
25%      107.000000
50%      152.000000
75%      200.000000
max      439.000000
Name: TSUrushyards, dtype: float64
```

In [422]:

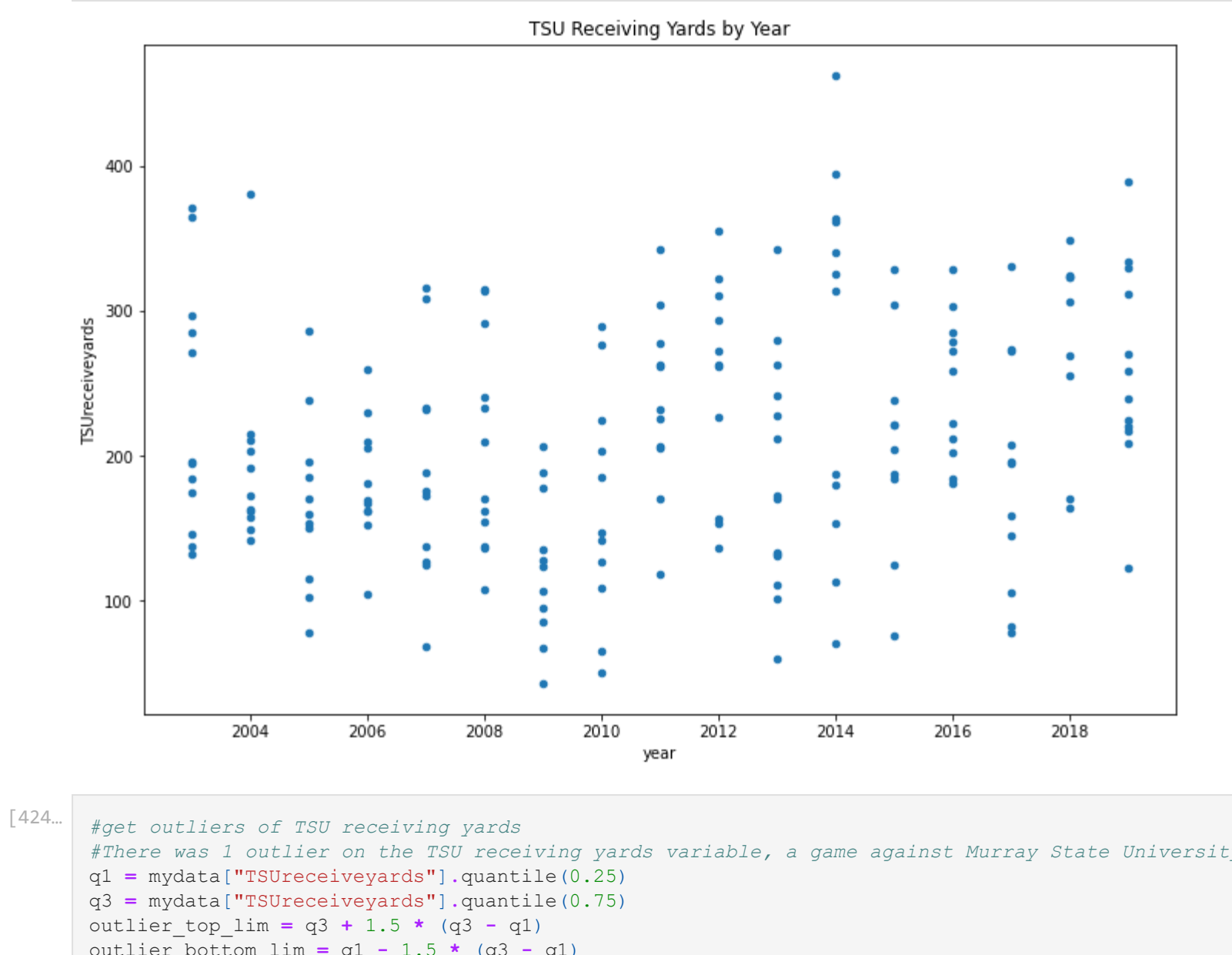
```
#descriptive statistics for TSU rushing yards with outliers replaced
#The mean and standard deviation of TSU rushing yards dropped to 153 and about 67, respectively when the outli...
mydata.TSUrushyards_new.describe()
```

Out[422]:

```
count    191.000000
mean     153.528589
std       66.897678
min      -18.000000
25%      107.000000
50%      152.000000
75%      195.500000
max      320.000000
Name: TSUrushyards_new, dtype: float64
```

In [423]:

```
#TSU receiving yards by year
#A scatterplot of TSU receiving yards showed that for most games the receiving  yard ranged between 100 and 300...
plt.rcParams["figure.figsize"]=[12,8]
mydata.plot(x="year",y="TSUreceiveyards",kind="scatter",title="TSU Receiving Yards by Year")
plt.show()
```



In [424]:

```
#get outliers of TSU receiving yards
#There was 1 outlier on the TSU receiving yards variable, a game against Murray State University where TSU had ...
q1 = mydata["TSUreceiveyards"].quantile(0.25)
q3 = mydata["TSUreceiveyards"].quantile(0.75)
outlier_top_lim = q3 + 1.5 * (q3 - q1)
outlier_bottom_lim = q1 - 1.5 * (q3 - q1)
mydata[["date","city","opponent","TSUreceiveyards"]][(mydata["TSUreceiveyards"]>outlier_top_lim) | (mydata["TSU...
```

Out[424]:

	date	city	opponent	TSUreceiveyards
137	2014-11-22	Murray	Murray State University	463

In [425]:

```
#replace outliers of TSU rush yards with mean
mydata["TSUreceiveyards_new"]=np.where((mydata["TSUreceiveyards"]>outlier_top_lim) | (mydata["TSUreceiveyards"]...
```

In [426]:

```
#descriptive statistics for TSU receiving yards
mydata.TSUreceiveyards.describe()
```

Out[426]:

```
count    191.000000
mean     210.361257
std       82.371433
min       43.000000
25%      151.000000
50%      203.000000
75%      273.000000
max      463.000000
Name: TSUreceiveyards, dtype: float64
```

In [427]:

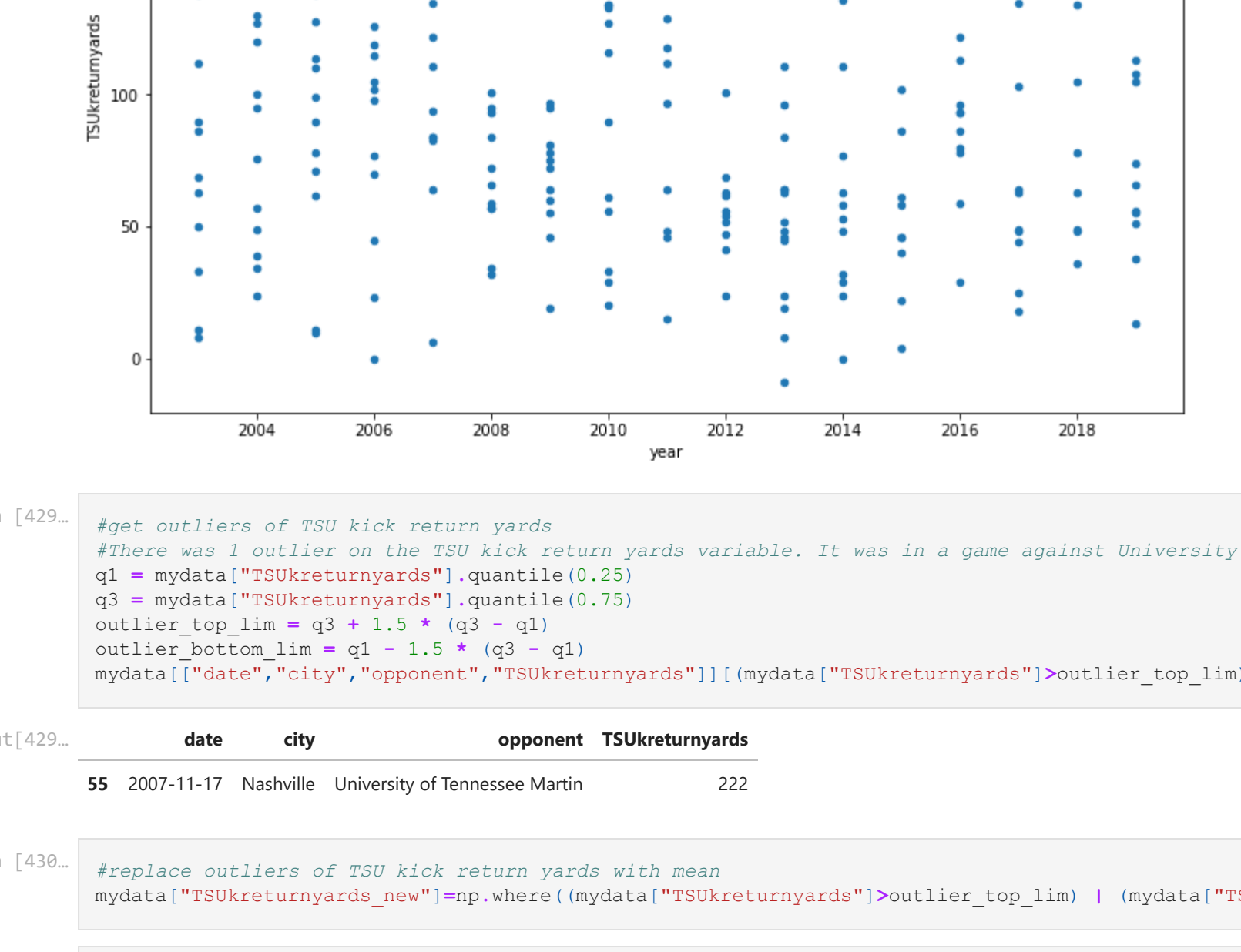
```
#descriptive statistics for TSU receiving yards with outliers replaced
#Replacing the outlier with the mean increased the mean of the variable slightly to 210 and lowered the standard ...
mydata.TSUreceiveyards_new.describe()
```

Out[427]:

```
count    191.000000
mean     209.338541
std       80.295493
min       43.000000
25%      151.000000
50%      205.000000
75%      272.000000
max      395.000000
Name: TSUreceiveyards_new, dtype: float64
```

In [428]:

```
#TSU kick return yards by year
#A scatterplot of kick return yards showed that for most games, the kick return yards ranged from 0 to about 1...
plt.rcParams["figure.figsize"]=[12,8]
mydata.plot(x="year", y="TSUkreturnyards",kind="scatter", title="TSU Kick Return Yards by Year")
plt.show()
```



In [429]:

```
#get outliers of TSU kick return yards
#There was 1 outlier on the TSU kick return yards variable. It was in a game against University of Tennessee M...
q1 = mydata["TSUkreturnyards"].quantile(0.25)
q3 = mydata["TSUkreturnyards"].quantile(0.75)
outlier_top_lim = q3 + 1.5 * (q3 - q1)
outlier_bottom_lim = q1 - 1.5 * (q3 - q1)
mydata[["date","city","opponent","TSUkreturnyards"]][(mydata["TSUkreturnyards"]>outlier_top_lim) | (mydata["TSU...
```

Out[429]:

	date	city	opponent	TSUkreturnyards
55	2007-11-17	Nashville	University of Tennessee Martin	222

In [430]:

```
#replace outliers of TSU kick return yards with mean
mydata["TSUkreturnyards_new"]=np.where((mydata["TSUkreturnyards"]>outlier_top_lim) | (mydata["TSUkreturnyards"...
```

In [431]:

```
#descriptive statistics for TSU kick return yards
mydata.TSUkreturnyards.describe()
```

Out[431]:

```
count    191.000000
mean      78.670157
std       44.059642
min       -9.000000
25%       48.000000
50%       72.000000
75%      110.500000
max      222.000000
Name: TSUkreturnyards, dtype: float64
```

In [432]:

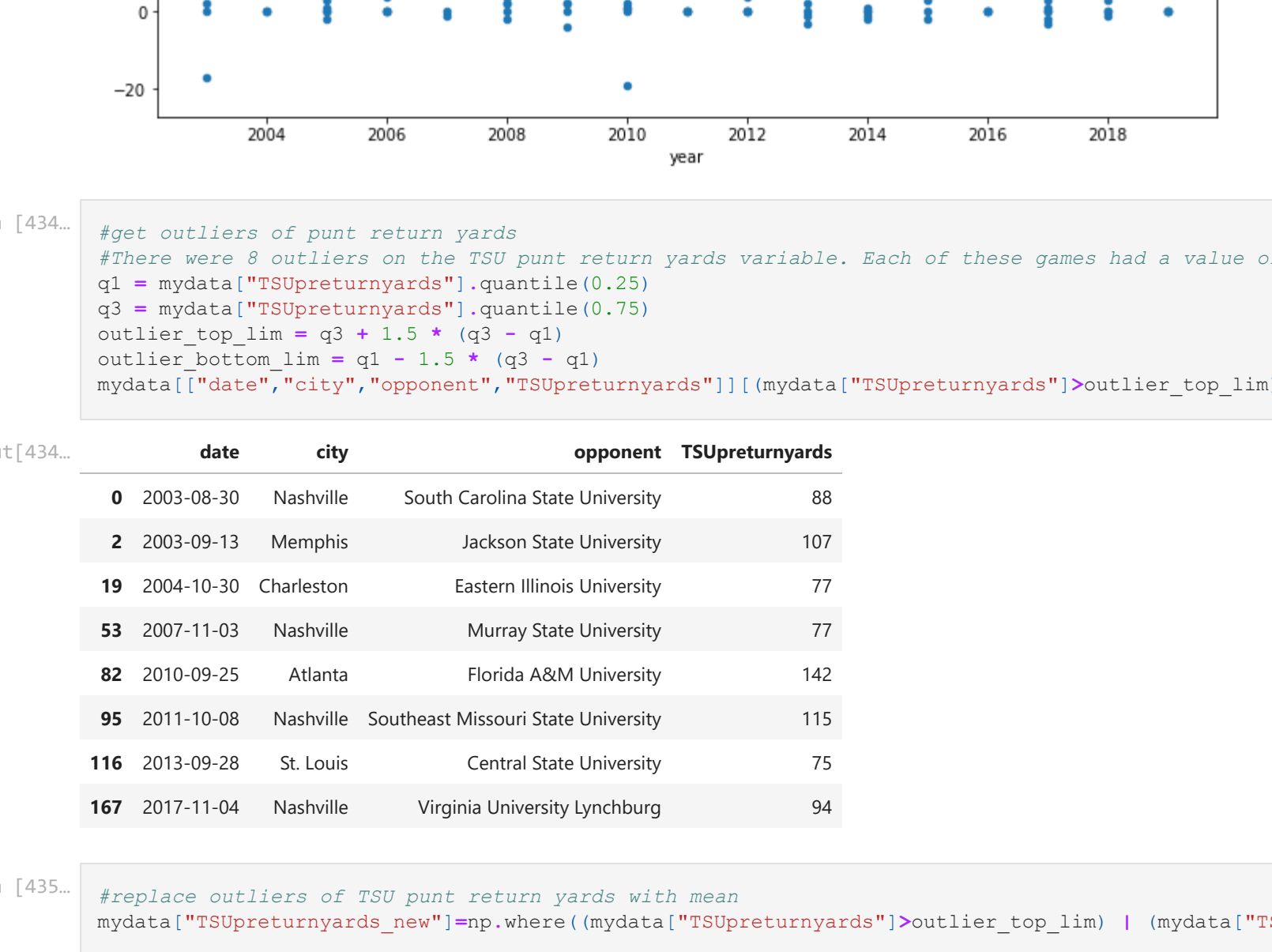
```
#descriptive statistics for TSU kick return yards with outliers replaced
#Replacing the value on the outlier caused the mean and standard deviation of TSU kick return yards to drop sl...
mydata.TSUkreturnyards_new.describe()
```

Out[432]:

```
count    191.000000
mean      77.919739
std       42.808461
min       -9.000000
25%       48.000000
50%       72.000000
75%      109.000000
max      198.000000
Name: TSUkreturnyards_new, dtype: float64
```

In [433]:

```
#TSU punt return yards by year
#A scatter plot of the TSU punt return yad showed that for most games it ranged from 0 to 55. It also showed t...
plt.rcParams["figure.figsize"]=[12,8]
mydata.plot(x="year", y="TSUpreturnyards",kind="scatter", title="TSU Punt Return Yards by Year")
plt.show()
```



In [434]:

```
#get outliers of 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 p...
q1 = mydata["TSUpreturnyards"].quantile(0.25)
q3 = mydata["TSUpreturnyards"].quantile(0.75)
outlier_top_lim = q3 + 1.5 * (q3 - q1)
outlier_bottom_lim = q1 - 1.5 * (q3 - q1)
mydata[["date","city","opponent","TSUpreturnyards"]][(mydata["TSUpreturnyards"]>outlier_top_lim) | (mydata["TSU...
```

Out[434]:

	date	city	opponent	TSUpreturnyards
0	2003-08-30	Nashville	South Carolina State University	88
2	2003-09-13	Memphis	Jackson State University	107
19	2004-10-30	Charleston	Eastern Illinois University	77
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

In [435]:

```
#replace outliers of punt return yards with mean
mydata["TSUpreturnyards_new"]=np.where((mydata["TSUpreturnyards"]>outlier_top_lim) | (mydata["TSUpreturnyards"...
```

In [436]:

```
#descriptive statistics for TSU punt return yards
mydata.TSUpreturnyards.describe()
```

Out[436]:

```
count    191.000000
mean      18.753927
std       24.579563
min      -19.000000
25%        0.000000
50%       11.000000
75%       29.000000
max      142.000000
Name: TSUpreturnyards, dtype: float64
```

In [437]:

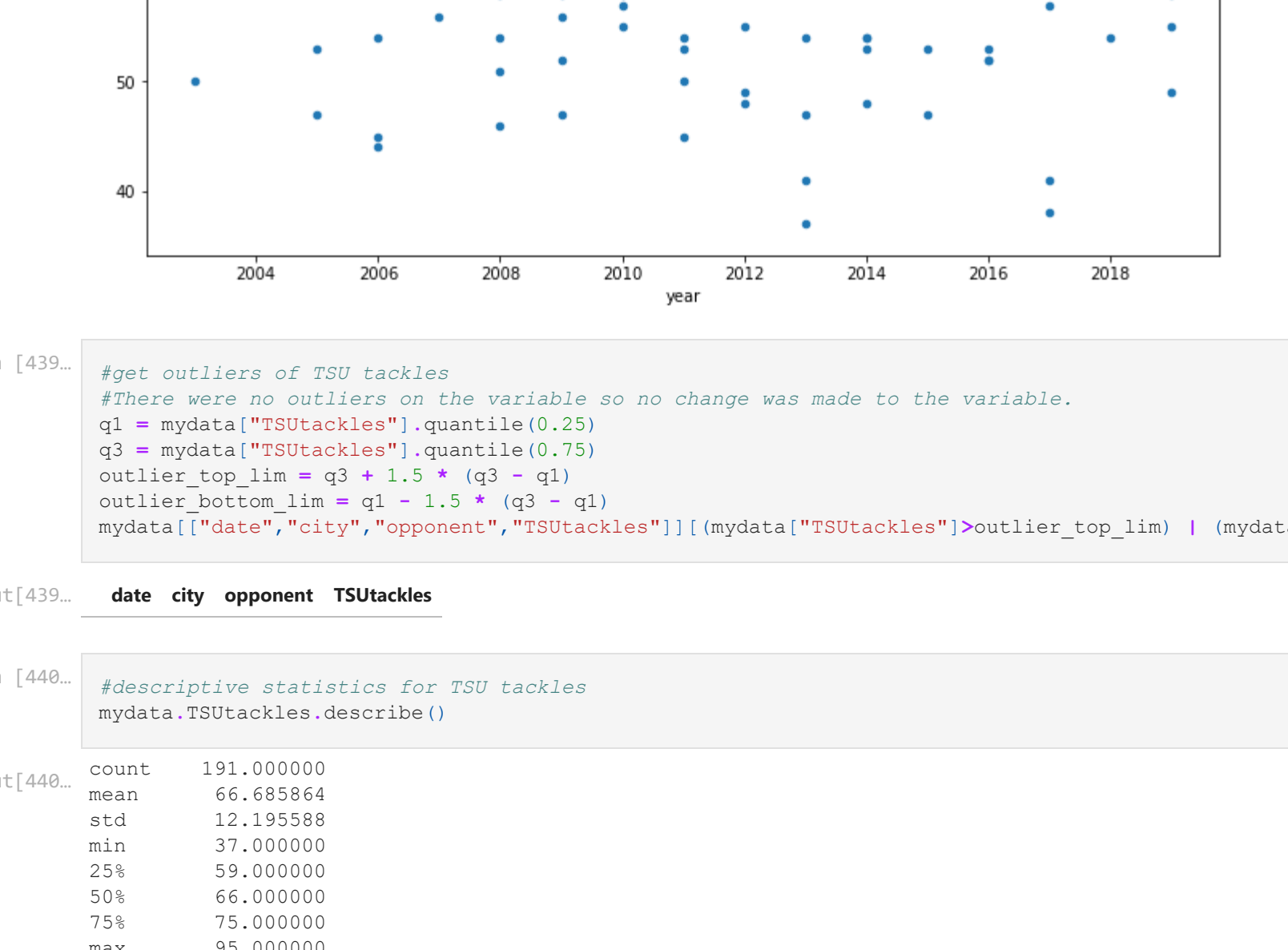
```
#descriptive statistics for TSU punt return yards with outliers replaced
#Replacing the outliers dropped the mean of the variable to 15 and the standard deviation to about 18.
mydata.TSUpreturnyards_new.describe()
```

Out[437]:

```
count    191.000000
mean      15.481840
std       17.785031
min      -19.000000
25%        0.000000
50%       11.000000
75%       24.000000
max       67.000000
Name: TSUpreturnyards_new, dtype: float64
```

In [438]:

```
#TSU tackles by year
#A scatter plot of the TSU tackles per game showed that for most games they ranged from 45 to 85. Analysis below lo...
plt.rcParams["figure.figsize"]=[12,8]
mydata.plot(x="year", y="TSUtackles",kind="scatter", title="TSU tackles by Year")
plt.show()
```



In [439]:

```
#get outliers of TSU tackles
#There were no outliers on the variable so no change was made to the variable.
q1 = mydata["TSUtackles"].quantile(0.25)
q3 = mydata["TSUtackles"].quantile(0.75)
outlier_top_lim = q3 + 1.5 * (q3 - q1)
outlier_bottom_lim = q1 - 1.5 * (q3 - q1)
mydata[["date","city","opponent","TSUtackles"]][(mydata["TSUtackles"]>outlier_top_lim) | (mydata["TSUtackles"]<...
```

Out[439]:

date	city	opponent	TSUtackles

In [440]:

```
#descriptive statistics for TSU tackles
mydata.TSUtackles.describe()
```

Out[440]:

```
count    191.000000
mean      66.685864
std       12.195588
min       37.000000
25%       59.000000
50%       66.000000
75%       75.000000
max       95.000000
Name: TSUtackles, dtype: float64
```

In [441]:

```
#TSU tackle yards by year
#TSU tackle yards for most games from 2003 to 2019 ranged from 0 to about 45 per game, based on a scatterplot o...
plt.rcParams["figure.figsize"]=[12,8]
mydata.plot(x="year", y="TSUtackleyd",kind="scatter", title="TSU Tackle Yards by Year")
plt.show()
```

In [442]:

```
#get outliers of TSU tackle yards
#There were 6 outliers for TSU tackle yards. They came in 2 games against Florida A&M University, 2 games agai...
q1 = mydata["TSUtackleyd"].quantile(0.25)
q3 = mydata["TSUtackleyd"].quantile(0.75)
outlier_top_lim = q3 + 1.5 * (q3 - q1)
outlier_bottom_lim = q1 - 1.5 * (q3 - q1)
mydata[["date","city","opponent","TSUtackleyd"]][(mydata["TSUtackleyd"]>outlier_top_lim) | (mydata["TSUtackleyd...
```

Out[442]:

	date	city	opponent	TSUtackleyd
13	2004-09-09	Martin	University of Tennessee Martin	60
15	2004-09-25	Atlanta	Florida A&M University	70
82	2010-09-25	Atlanta	Florida A&M University	60
118	2013-10-12	Jacksonville	Jacksonville State University	61
137	2014-11-22	Murray	Murray State University	63
146	2015-11-07	Nashville	Murray State University	69

In [443]:

```
#replace outliers of TSU tackle yards with mean
mydata["TSUtackleyd_new"]=np.where((mydata["TSUtackleyd"]>outlier_top_lim) | (mydata["TSUtackleyd"]<outlier_bot...
```

In [444]:

```
#descriptive statistics for TSU tackle yards
mydata.TSUtackleyd.describe()
```

Out[444]:

```
count    191.000000
mean      25.350785
std       14.392157
min        0.000000
25%       14.500000
50%       23.000000
75%       32.500000
max       70.000000
Name: TSUtackleyd, dtype: float64
```

In [445]:

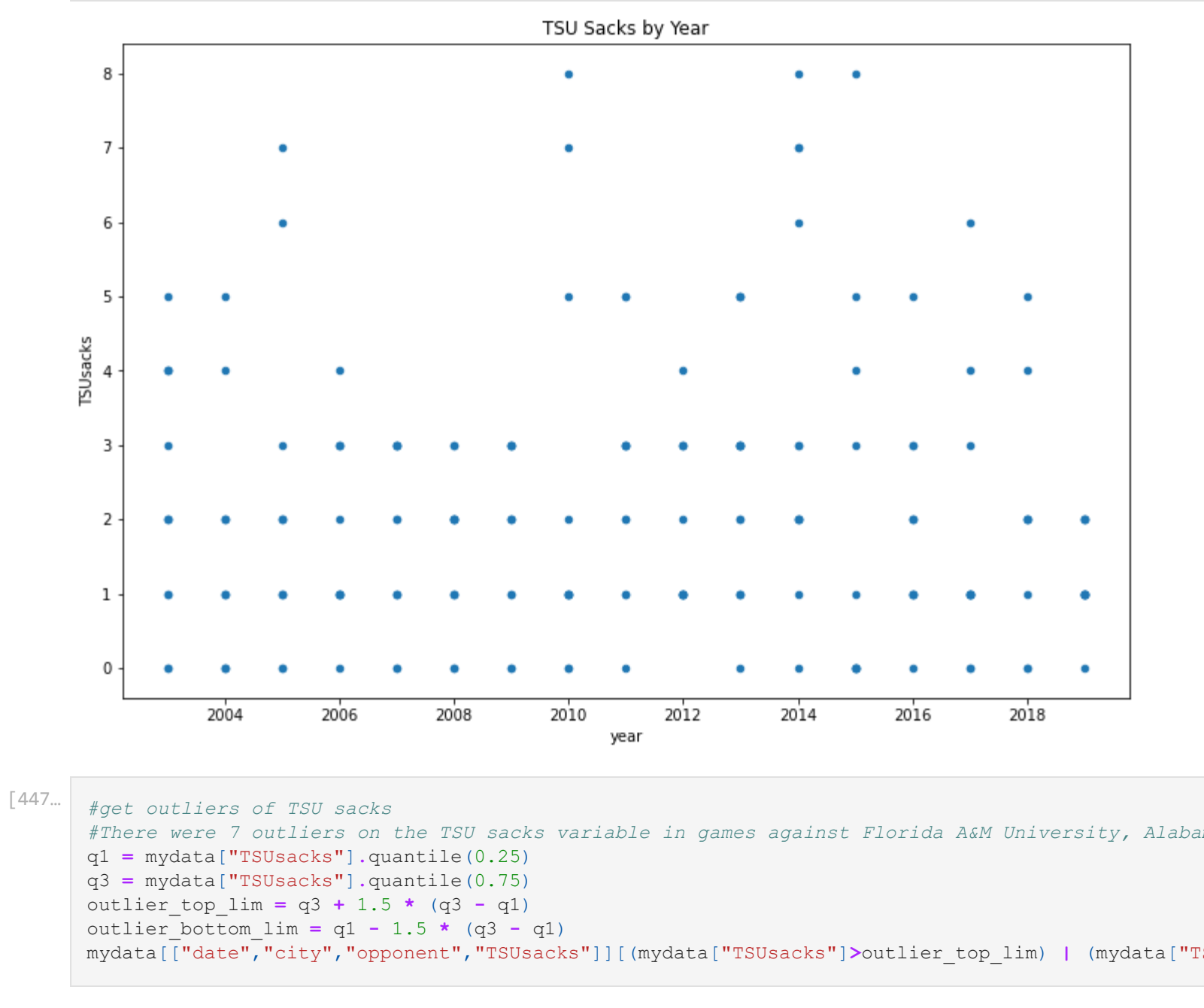
```
#descriptive statistics for TSU tackle yards with outliers replaced
#Replacing the outliers on TSU tackle yards lowered the mean slightly to 24 and
#lowered its standard deviation slightly to about 13.
mydata.TSUtackleyd_new.describe()
```

Out[445]:

```
count    191.000000
mean      24.141909
std       12.584055
min        0.000000
25%       14.500000
50%       23.000000
75%       30.500000
max       59.000000
Name: TSUtackleyd_new, dtype: float64
```

In [446]:

```
#TSU sacks by year
#For most games from 2003 to 2019, TSU had 5 or fewer sacks. The scatterplot of the TSU sacks variable showed o...
plt.rcParams["figure.figsize"]=[12,8]
mydata.plot(x="year", y="TSUsacks",kind="scatter", title="TSU Sacks by Year")
plt.show()
```



In [447]:

```
#get outliers of TSU sacks
#There were 7 outliers on the TSU sacks variable in games against Florida A&M University, Alabama A&M Universit...
q1 = mydata["TSUsacks"].quantile(0.25)
q3 = mydata["TSUsacks"].quantile(0.75)
outlier_top_lim = q3 + 1.5 * (q3 - q1)
outlier_bottom_lim = q1 - 1.5 * (q3 - q1)
mydata[["date","city","opponent","TSUsacks"]][(mydata["TSUsacks"]>outlier_top_lim) | (mydata["TSUsacks"]<outli...
```

	date	city	opponent	TSUsacks
26	2005-09-24	Atlanta	Florida A&M University	7.0
79	2010-09-04	Nashville	Alabama A&M University	7.0
82	2010-09-25	Atlanta	Florida A&M University	8.0
128	2014-09-13	Memphis	Jackson State University	7.0
130	2014-09-27	Nashville	Florida A&M University	7.0
137	2014-11-22	Murray	Murray State University	8.0
146	2015-11-07	Murray	Murray State University	7.0

In [448]:
```
#replace outliers of TSU sacks with mean
mydata["TSUsacks_new"]=np.where((mydata["TSUsacks"]>outlier_top_lim) | (mydata["TSUsacks"]<outlier_bottom_lim),...
```

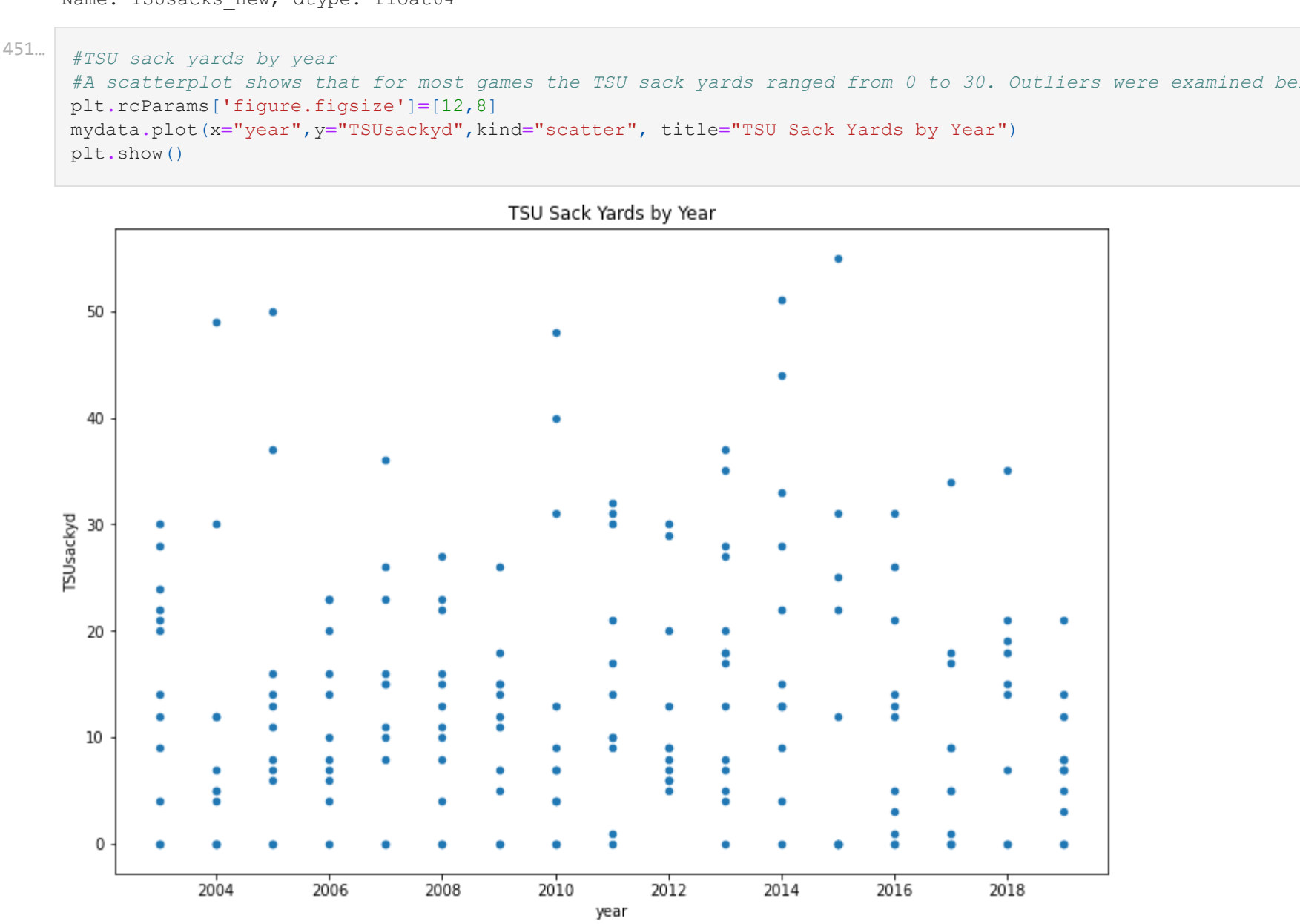
In [449]:
```
#descriptive statistics for TSU sacks
mydata.TSUsacks.describe()
```

Out[449]:
```
count    191.000000
mean       2.089005
std        1.776287
min        0.000000
25%        1.000000
50%        2.000000
75%        3.000000
max        8.000000
Name: TSUsacks, dtype: float64
```

In [450]:
```
#descriptive statistics for TSU sacks with outliers replaced
#With the outliers replaced, the mean TSU sacks were 1.9 with a standard deviation of 1.4.
mydata.TSUsacks_new.describe()
```

Out[450]:
```
count    191.000000
mean       1.893314
std        1.434319
min        0.000000
25%        1.000000
50%        2.000000
75%        3.000000
max        6.000000
Name: TSUsacks_new, dtype: float64
```

In [451]:
```
#TSU sack yards by year
#A scatterplot shows that for most games the TSU sack yards ranged from 0 to 30. Outliers were examined below.
plt.rcParams['figure.figsize']=[12,8]
mydata.plot(x="year",y="TSUsackyd",kind="scatter", title="TSU Sack Yards by Year")
plt.show()
```



In [452]:
```
#get outliers of TSU sack yards
#There were 6 outliers on the TSU sack yards variable. These were in games against Florida A&M University and ...
q1 = mydata["TSUsackyd"].quantile(0.25)
q3 = mydata["TSUsackyd"].quantile(0.75)
outlier_top_lim = q3 + 1.5 * (q3 - q1)
outlier_bottom_lim = q1 - 1.5 * (q3 - q1)
mydata[["date","city","opponent","TSUsackyd"]][(mydata["TSUsackyd"]>outlier_top_lim) | (mydata["TSUsackyd"]<out...
```

Out[452]:

	date	city	opponent	TSUsackyd
15	2004-09-25	Atlanta	Florida A&M University	49
26	2005-09-24	Atlanta	Florida A&M University	50
82	2010-09-25	Atlanta	Florida A&M University	48
130	2014-09-27	Nashville	Florida A&M University	44
137	2014-11-22	Murray	Murray State University	51
146	2015-11-07	Murray	Murray State University	55

In [453]:
```
#replace outliers of TSU sack yards with mean
mydata["TSUsackyd_new"]=np.where((mydata["TSUsackyd"]>outlier_top_lim) | (mydata["TSUsackyd"]<outlier_bottom_li...
```

In [454]:
```
#descriptive statistics for TSU sack yards
mydata.TSUsackyd.describe()
```

Out[454]:
```
count    191.000000
mean      13.534031
std       11.944946
min        0.000000
25%        5.000000
50%       11.000000
75%       20.000000
max       55.000000
Name: TSUsackyd, dtype: float64
```

In [455]:
```
#descriptive statistics for TSU sack yards with outliers replaced
#Replacing these outliers, caused the mean of TSU sack yards to drop to 12.4 while the standard deviation dropp...
mydata.TSUsackyd_new.describe()
```

Out[455]:
```
count    191.000000
mean      12.404210
std       10.010235
min        0.000000
25%        5.000000
50%       11.000000
75%       18.000000
max       40.000000
Name: TSUsackyd_new, dtype: float64
```

In [456]:
```
#TSU punts by year
#A scatter plot of TSU punts shows that for most games, TSU had between 2 and 8 punts. Outliers were assessed ...
plt.rcParams['figure.figsize']=[12,8]
mydata.plot(x="year",y="TSUpunt",kind="scatter", title="TSU Punts by Year")
plt.show()
```

In [457]:
```
#get outliers of TSU punt
#There were 4 outliers on the TSU punt variable. They occurred in a game against Austin Peay State University ...
q1 = mydata["TSUpunt"].quantile(0.25)
q3 = mydata["TSUpunt"].quantile(0.75)
outlier_top_lim = q3 + 1.5 * (q3 - q1)
outlier_bottom_lim = q1 - 1.5 * (q3 - q1)
mydata[["date","city","opponent","TSUpunt"]][(mydata["TSUpunt"]>outlier_top_lim) | (mydata["TSUpunt"]<outlier_b...
```

Out[457]:

	date	city	opponent	TSUpunt
10	2003-11-15	Nashville	Eastern Kentucky University	10
51	2007-10-20	Richmond	Eastern Kentucky University	10
136	2014-11-08	Clarksville	Austin Peay State University	0
169	2017-11-04	Jacksonville	Jacksonville State University	11

In [458]:
```
#replace outliers of TSU punts with mean
mydata["TSUpunt_new"]=np.where((mydata["TSUpunt"]>outlier_top_lim) | (mydata["TSUpunt"]<outlier_bottom_lim),myd...
```

In [459]:
```
#descriptive statistics for TSU punts
mydata.TSUpunt.describe()
```

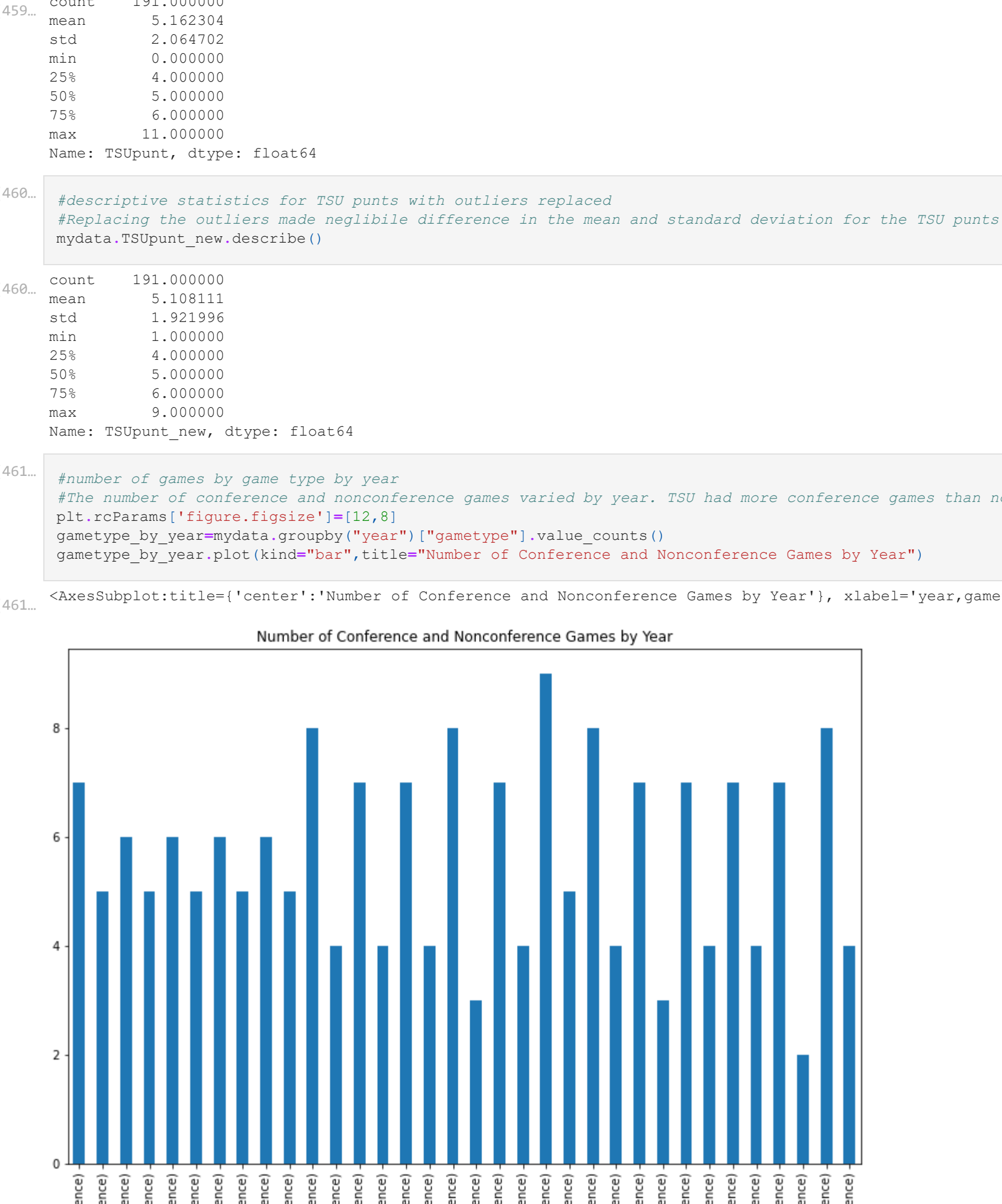
Out[459]:
```
count    191.000000
mean       5.162304
std        2.064702
min        0.000000
25%        4.000000
50%        5.000000
75%        6.000000
max       11.000000
Name: TSUpunt, dtype: float64
```

In [460]:
```
#descriptive statistics for TSU punts with outliers replaced
#Replacing the outliers made negligible difference in the mean and standard deviation for the TSU punts variab...
mydata.TSUpunt_new.describe()
```

Out[460]:
```
count    191.000000
mean       5.108111
std        1.921996
min        1.000000
25%        4.000000
50%        5.000000
75%        6.000000
max        9.000000
Name: TSUpunt_new, dtype: float64
```

In [461]:
```
#number of games by game type by year
#The number of conference and nonconference games varied by year. TSU had more conference games than nonconfere...
plt.rcParams['figure.figsize']=[12,8]
gametype_by_year=mydata.groupby("year")["gametype"].value_counts()
gametype_by_year.plot(kind="bar",title="Number of Conference and Nonconference Games by Year")
```

Out[461]: <AxesSubplot:title={'center':'Number of Conference and Nonconference Games by Year'}, xlabel='year,gametype'>



More data wrangling to prep for machine learning

In [462]:
```
#creating and saving new file with variables with outliers replaced
mydata_1=mydata
mydata_1.to_csv("mydata_1.csv",encoding="utf-8")
```

In [463]:
```
mydata_1=pd.read_csv("mydata_1.csv")
mydata_1=mydata_1.iloc[:,1:]
mydata_1.columns
```

Out[463]:
```
Index(['date', 'city', 'state', 'winscore', 'losscore', 'loser', 'winner',
       'locale', 'TSU score', 'opponent score', 'scorediff', 'scorediff_abs',
       'winloss', 'attendance', 'TSUrushyards', 'TSUreceiveyards',
       'TSUkreturnyards', 'TSUpreturnyards', 'TSUtackles', 'TSUtackleyd',
       'TSUsacks', 'TSUsackyd', 'TSUpunt', 'opponent', 'gametype', 'year',
       'TSU score_new', 'opponent score_new', 'scorediff_new',
       'scorediff_abs_new', 'attendance_new', 'TSUrushyards_new',
       'TSUreceiveyards_new', 'TSUkreturnyards_new', 'TSUpreturnyards_new',
       'TSUtackles_new', 'TSUtackleyd_new', 'TSUsacks_new', 'TSUsackyd_new',
       'TSUpunt_new'],
      dtype='object')
```

In [464]:
```
mydata_1["locale_01"]=np.where(mydata_1["locale"]=="home",1,0)
mydata_1["gametype_01"]=np.where(mydata_1["gametype"]=="conference",1,0)
mydata_1.columns
```

Out[464]:
```
Index(['date', 'city', 'state', 'winscore', 'losscore', 'loser', 'winner',
       'locale', 'TSU score', 'opponent score', 'scorediff', 'scorediff_abs',
       'winloss', 'attendance', 'TSUrushyards', 'TSUreceiveyards',
       'TSUkreturnyards', 'TSUpreturnyards', 'TSUtackles', 'TSUtackleyd',
       'TSUsacks', 'TSUsackyd', 'TSUpunt', 'opponent', 'gametype', 'year',
       'TSU score_new', 'opponent score_new', 'scorediff_new',
       'scorediff_abs_new', 'attendance_new', 'TSUrushyards_new',
       'TSUreceiveyards_new', 'TSUkreturnyards_new', 'TSUpreturnyards_new',
       'TSUtackles_new', 'TSUtackleyd_new', 'TSUsacks_new', 'TSUsackyd_new',
       'TSUpunt_new', 'locale_01', 'gametype_01'],
      dtype='object')
```

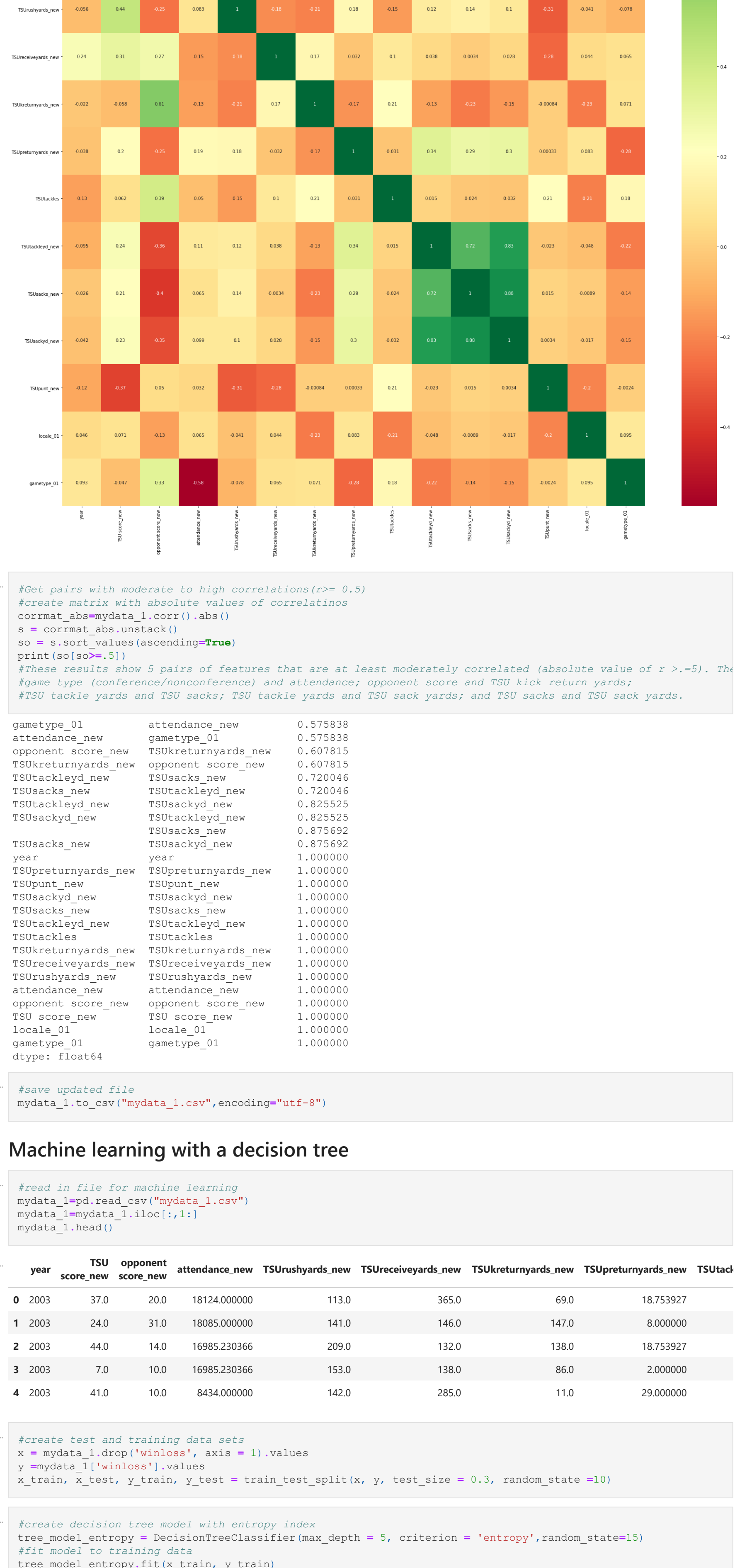
In [465]:
```
mydata_1=mydata_1[['year','TSU score_new','opponent score_new',
    'attendance_new','TSUreceiveyards_new','TSUrushyards_new','TSUkreturnyards_new',
    'TSUpreturnyards_new','TSUtackles_new','TSUtackleyd_new','TSUsacks_new',
    'TSUsackyd_new','TSUpunt_new','locale_01','winloss','gametype_01']]
mydata_1.columns
```

Out[465]:
```
Index(['year', 'TSU score_new', 'opponent score_new', 'attendance_new',
       'TSUreceiveyards_new', 'TSUrushyards_new', 'TSUkreturnyards_new',
       'TSUpreturnyards_new', 'TSUtackles_new', 'TSUtackleyd_new', 'TSUsacks_new',
       'TSUsackyd_new', 'TSUpunt_new', 'locale_01', 'winloss', 'gametype_01'],
      dtype='object')
```

In [466]:
```
#get info on dataset
mydata_1.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 191 entries, 0 to 190
Data columns (total 16 columns):
 #   Column               Non-Null Count  Dtype
---  ------               --------------  -----
 0   year                 191 non-null    int64
 1   TSU score_new        191 non-null    float64
 2   opponent score_new   191 non-null    float64
 3   attendance_new       191 non-null    float64
 4   TSUreceiveyards_new  191 non-null    float64
 5   TSUrushyards_new     191 non-null    float64
 6   TSUkreturnyards_new  191 non-null    float64
 7   TSUpreturnyards_new  191 non-null    float64
 8   TSUtackles_new       191 non-null    int64
 9   TSUtackleyd_new      191 non-null    float64
 10  TSUsacks_new         191 non-null    float64
 11  TSUsackyd_new        191 non-null    float64
 12  TSUpunt_new          191 non-null    float64
 13  locale_01            191 non-null    int32
 14  winloss              191 non-null    object
 15  gametype_01          191 non-null    int32
dtypes: float64(11), int32(2), int64(2), object(1)
memory usage: 22.5+ KB
```

In [467]:
```
#correlation heatmap
cormat=mydata_1.corr()
feature_ind=cormat.index
plt.figure(figsize=(30,30))
sns.heatmap(mydata_1[feature_ind].corr(), annot=True,cmap="RdYlGn")
plt.show()
```



In [468]:
```
#get pairs with moderate to high correlations(r>= 0.5)
#create matrix with absolute values of correlations
cormat_abs=mydata_1.corr().abs()
s = cormat_abs.unstack()
so = s.sort_values(ascending=True)
print(so[so>=.5])
#These results show 9 pairs of features that are at least moderately correlated (absolute value of r >.5). The...
#(TSU gametype/nonconference/conference) and attendance; opponent score and TSU kick return yards;
#TSU tackle yards and TSU sacks; TSU tackle yards and TSU sacks and TSU sack yards.
```

```
gametype_01          attendance_new         0.575838
attendance_new       gametype_01            0.575838
opponent score_new   TSUkreturnyards_new    0.607815
TSUkreturnyards_new  opponent score_new     0.607815
TSUtackleyd_new      TSUsacks_new           0.720046
TSUsacks_new         TSUtackleyd_new        0.720046
TSUtackleyd_new      TSUsackyd_new          0.825525
TSUsackyd_new        TSUtackleyd_new        0.825525
TSUsacks_new         TSUsackyd_new          0.875692
TSUsackyd_new        TSUsacks_new           0.875692
year                 year                   1.000000
TSUkreturnyards_new  TSUkreturnyards_new    1.000000
TSUpunt_new          TSUpunt_new            1.000000
TSUsackyd_new        TSUsackyd_new          1.000000
TSUsacks_new         TSUsacks_new           1.000000
TSUtackleyd_new      TSUtackleyd_new        1.000000
TSUtackles_new       TSUtackles_new         1.000000
TSUpreturnyards_new  TSUpreturnyards_new    1.000000
TSUrushyards_new     TSUrushyards_new       1.000000
TSUreceiveyards_new  TSUreceiveyards_new    1.000000
attendance_new       attendance_new         1.000000
opponent score_new   opponent score_new     1.000000
TSU score_new        TSU score_new          1.000000
locale_01            locale_01              1.000000
gametype_01          gametype_01            1.000000
dtype: float64
```

In [469]:
```
#save updated file
mydata_1.to_csv("mydata_1.csv",encoding="utf-8")
```

Machine learning with a decision tree

In [470]:
```
#read file for machine learning
mydata_1=pd.read_csv("mydata_1.csv")
mydata_1=mydata_1.iloc[:,1:]
mydata_1.head()
```

Out[470]:

	year	TSU score_new	opponent score_new	attendance_new	TSUrushyards_new	TSUreceiveyards_new	TSUkreturnyards_new	TSUpreturnyards_new
0	2003	37.0	20.0	18124.000000	113.0	365.0	69.0	18.753927
1	2003	24.0	31.0	18085.000000	141.0	146.0	147.0	8.000000
2	2003	44.0	14.0	16985.230366	209.0	132.0	158.0	18.753927
3	2003	14.0	17.0	16985.230366	153.0	138.0	86.0	2.000000
4	2003	41.0	10.0	8434.000000	142.0	285.0	11.0	29.000000

In [471]:
```
#create test and training data sets
x =mydata_1.drop(['winloss'],axis=1).values
y=mydata_1['winloss'].values
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size = 0.3, random_state=10)
```

In [472]:
```
#create decision tree model with entropy index
tree_model_entropy = DecisionTreeClassifier(max_depth = 5, criterion = 'entropy',random_state=10)
#fit model to training data
tree_model_entropy.fit(x_train, y_train)
#use model to predict on test data
tree_yhat = tree_model_entropy.predict(x_test)
```

In [473]:
```
#accuracy score
print('Accuracy score of the decision tree model with the Entropy index is {}'.format(accuracy_score(y_test, tr...
#The accuracy score of the decision tree model with entropy index is about .88 or 88% which means that the mode...
```

Accuracy score of the decision tree model with the Entropy index is 0.8793103448275862

In [474]:
```
#create decision tree model with gini index
tree_model_gini = DecisionTreeClassifier(max_depth = 5, criterion = 'gini',random_state=10)
#fit model to training data
tree_model_gini.fit(x_train, y_train)
#use model to predict on test data
tree_yhat = tree_model_gini.predict(x_test)
```

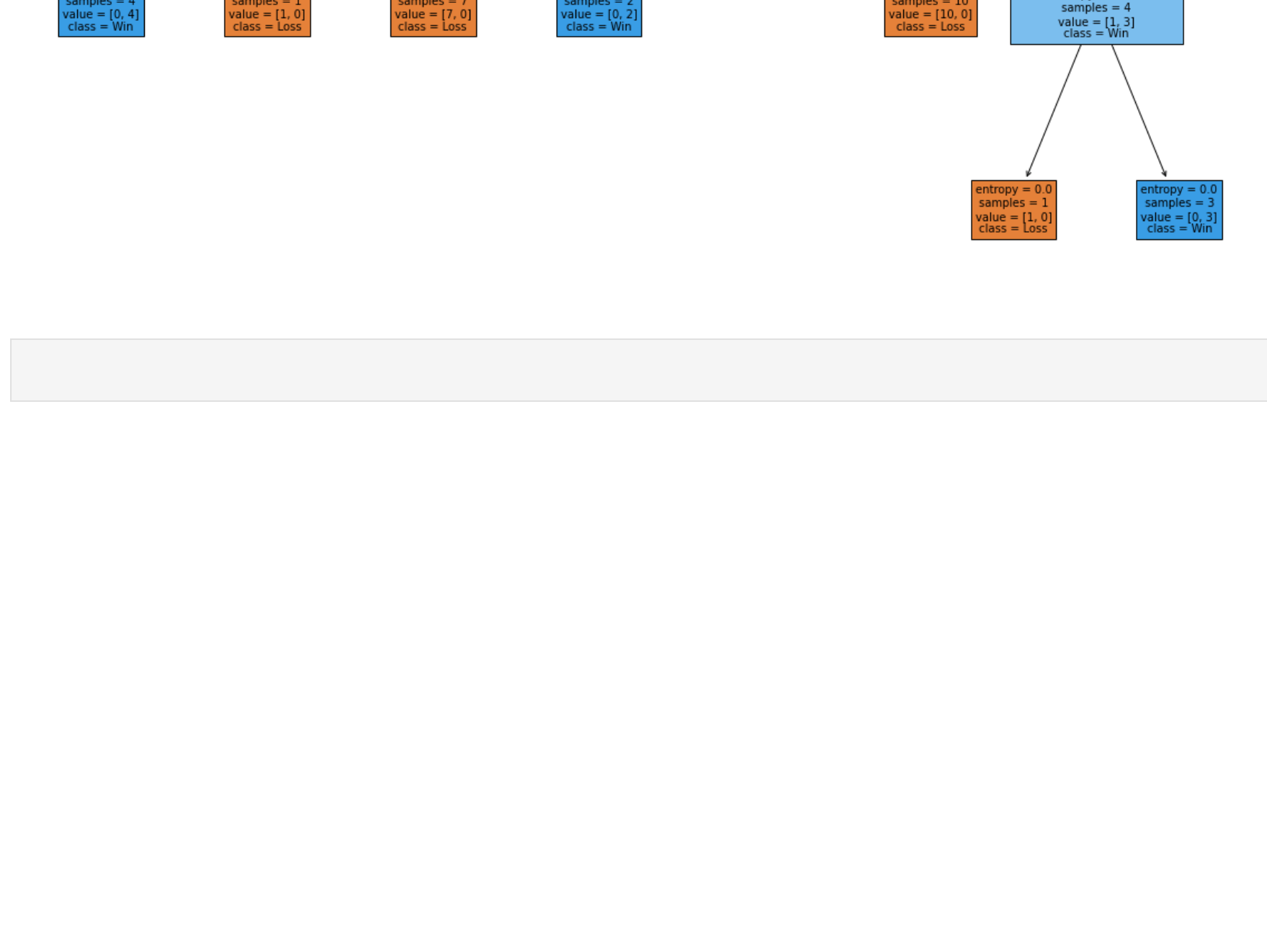
In [475]:
```
#accuracy score
print('Accuracy score of the decision tree model with the Gini index is {}'.format(accuracy_score(y_test, tree_...
#The accuracy score of the decision tree model with gini index is about .78 or 78% which means that the model ...
```

Accuracy score of the decision tree model with the Gini index is 0.7758620689655172

Since the decision tree with the Entropy index has a higher accuracy score than the decision tree with the Entropy index (88% vs. 78%), I decided to go with the decision tree with the Gini index.

In [476]:
```
#plot decision tree with entropy index
fn=['year','TSU score_new','opponent score_new','attendance_new',
    'TSUreceiveyards_new','TSUrushyards_new','TSUkreturnyards_new',
    'TSUpreturnyards_new','TSUtackles_new','TSUtackleyd_new','TSUsacks_new',
    'TSUsackyd_new','TSUpunt_new','locale_01','gametype_01']
cn=['Loss','Win']
plt.rcParams['figure.figsize']=[20,20]
plt.plot_tree(tree_model_entropy,class_names=cn, feature_names=fn,filled=True,fontsize=10)
```

Out[476]:
```
[Text(483.6, 996.6, 'opponent score_new <= 20.5\nentropy = 0.997\nsamples = 133\nvalue = [71, 62]\nclass = Los
s'),
 Text(297.6, 815.4000000000001, 'TSU score_new <= 18.5\nentropy = 0.607\nsamples = 47\nvalue = [7, 40]\nclass = Win'),
 Text(223.20000000000002, 634.2, 'TSUkreturnyards_new <= 56.5\nentropy = 0.946\nsamples = 11\nvalue = [7, 4]\nclass = Loss'),
 Text(148.8, 453.0, 'TSUreceiveyards_new <= 170.0\nentropy = 0.722\nsamples = 5\nvalue = [1, 4]\nclass = Win'),
 Text(74.4, 271.79999999999995, 'entropy = 0.0\nsamples = 4\nvalue = [0, 4]\nclass = Win'),
 Text(223.20000000000002, 271.79999999999995, 'entropy = 0.0\nsamples = 1\nvalue = [1, 0]\nclass = Loss'),
 Text(297.6, 453.0, 'entropy = 0.0\nsamples = 6\nvalue = [6, 0]\nclass = Loss'),
 Text(372.0, 634.2, 'entropy = 0.0\nsamples = 36\nvalue = [0, 36]\nclass = Win'),
 Text(669.6, 815.4000000000001, 'TSU score_new <= 30.0\nentropy = 0.82\nsamples = 86\nvalue = [64, 22]\nclass = Loss'),
 Text(520.8000000000001, 634.2, 'opponent score_new <= 24.5\nentropy = 0.225\nsamples = 55\nvalue = [53, 2]\ncl
ass = Loss'),
 Text(446.40000000000003, 453.0, 'TSUreceiveyards_new <= 215.0\nentropy = 0.764\nsamples = 9\nvalue = [7, 2]\nc
lass = Loss'),
 Text(372.0, 271.79999999999995, 'entropy = 0.0\nsamples = 7\nvalue = [7, 0]\nclass = Loss'),
 Text(520.8000000000001, 271.79999999999995, 'entropy = 0.0\nsamples = 2\nvalue = [0, 2]\nclass = Win'),
 Text(595.2, 453.0, 'entropy = 0.0\nsamples = 46\nvalue = [46, 0]\nclass = Loss'),
 Text(818.4000000000001, 634.2, 'opponent score_new <= 34.5\nentropy = 0.938\nsamples = 31\nvalue = [11, 20]\nc
lass = Win'),
 Text(744.0, 453.0, 'entropy = 0.0\nsamples = 17\nvalue = [0, 17]\nclass = Win'),
 Text(892.8000000000001, 453.0, 'TSU score_new <= 39.5\nentropy = 0.75\nsamples = 14\nvalue = [11, 3]\nclass = Loss'),
 Text(818.4000000000001, 271.79999999999995, 'entropy = 0.0\nsamples = 10\nvalue = [10, 0]\nclass = Loss'),
 Text(967.2, 271.79999999999995, 'TSUrushyards_new <= 135.5\nentropy = 0.811\nsamples = 4\nvalue = [1, 3]\nclass = Win'),
 Text(892.8000000000001, 90.59999999999991, 'entropy = 0.0\nsamples = 1\nvalue = [1, 0]\nclass = Loss'),
 Text(1041.6000000000001, 90.59999999999991, 'entropy = 0.0\nsamples = 3\nvalue = [0, 3]\nclass = Win')]
```



In [ ]: