**Analyzing NCAA Baseball dataset**

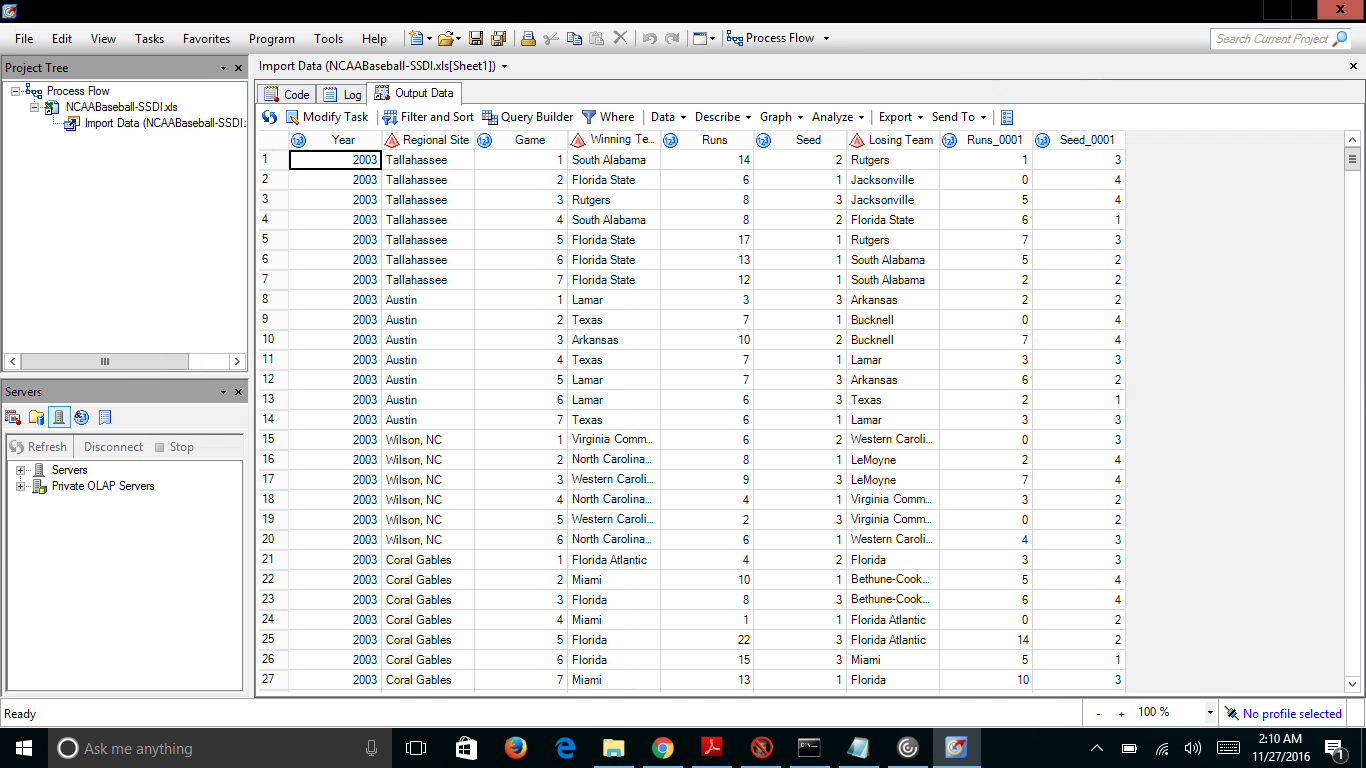
**USER DOCUMENTATION:**

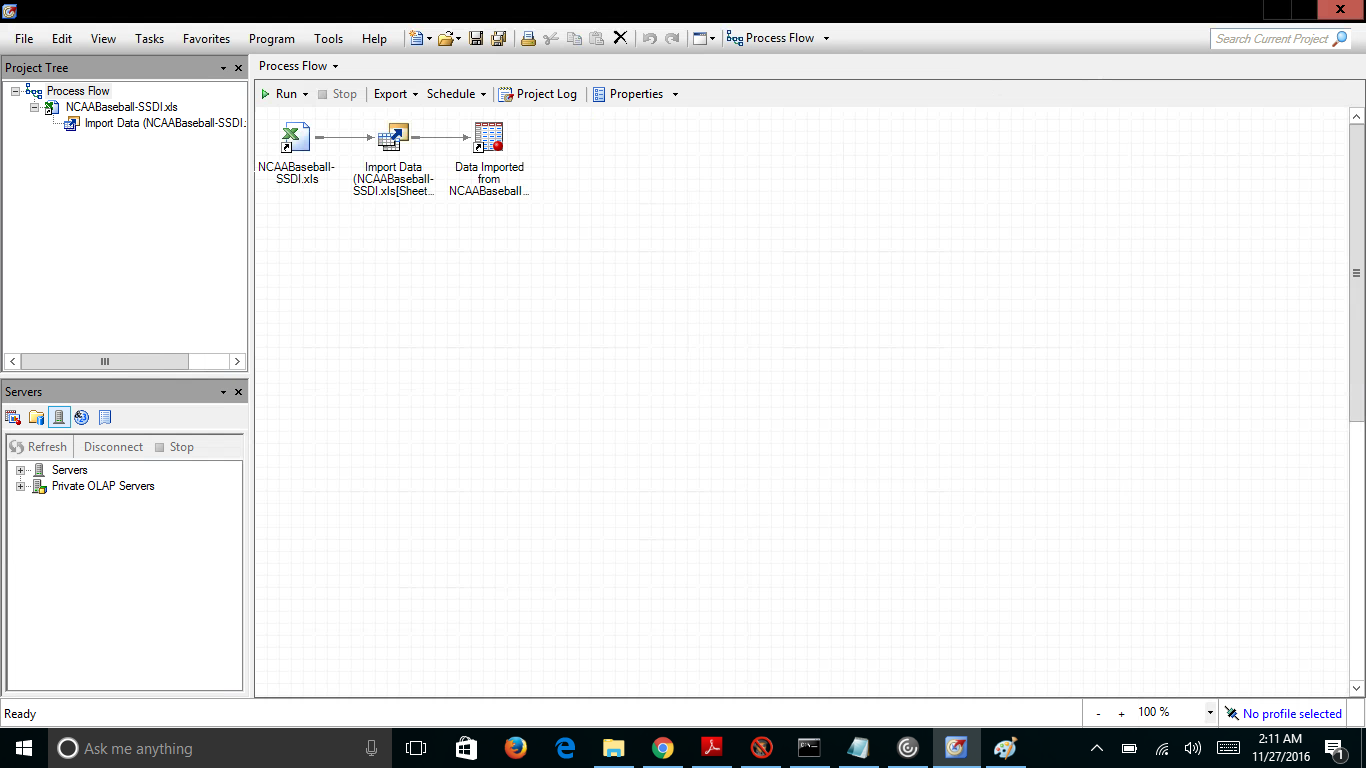
1. SAS Enterprise guide

Analysis of dataset is done in SAS Enterprise guide, where first the dataset is imported into the SAS Enterprise guide and later in it described using characterize data where the columns of the dataset are divided into categorical and numerical variables. Date variable is not shown as there are no date format variables in the dataset.

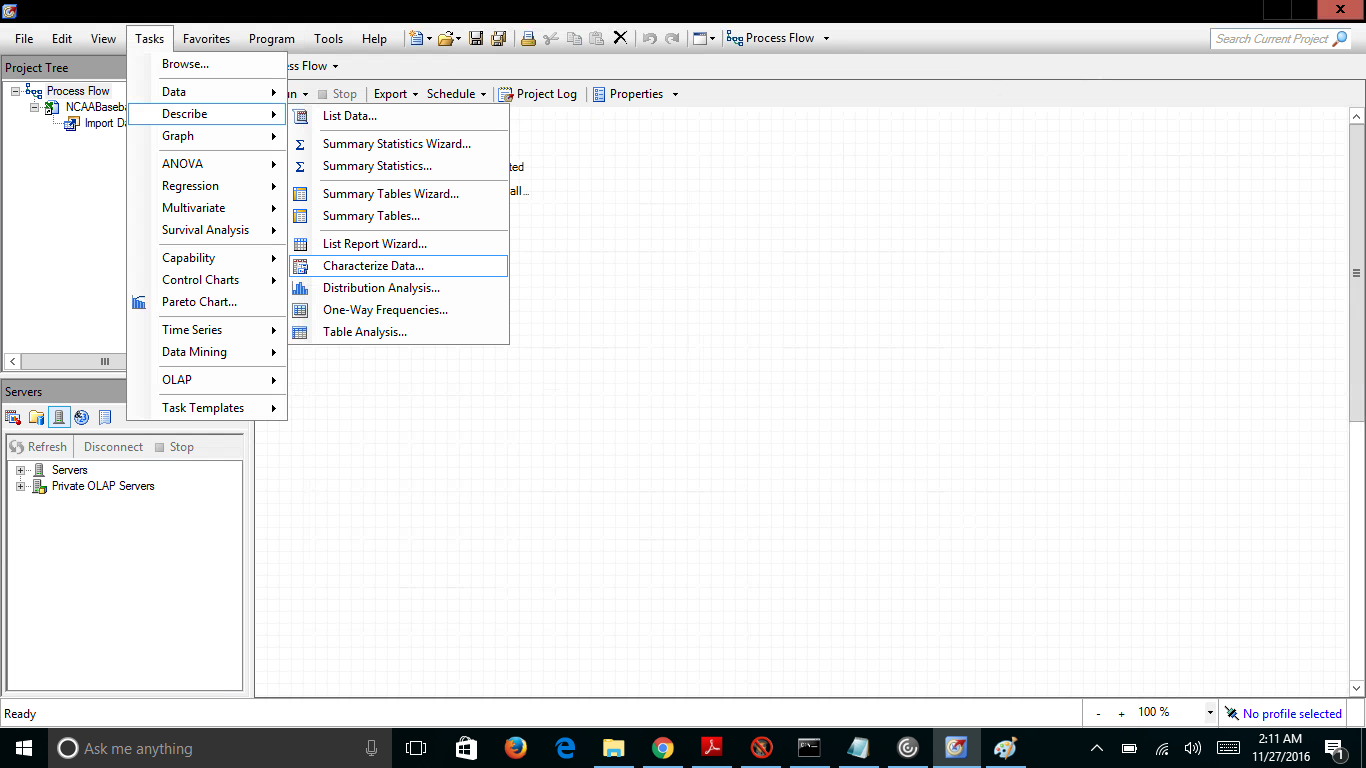
The process is followed as shown below:

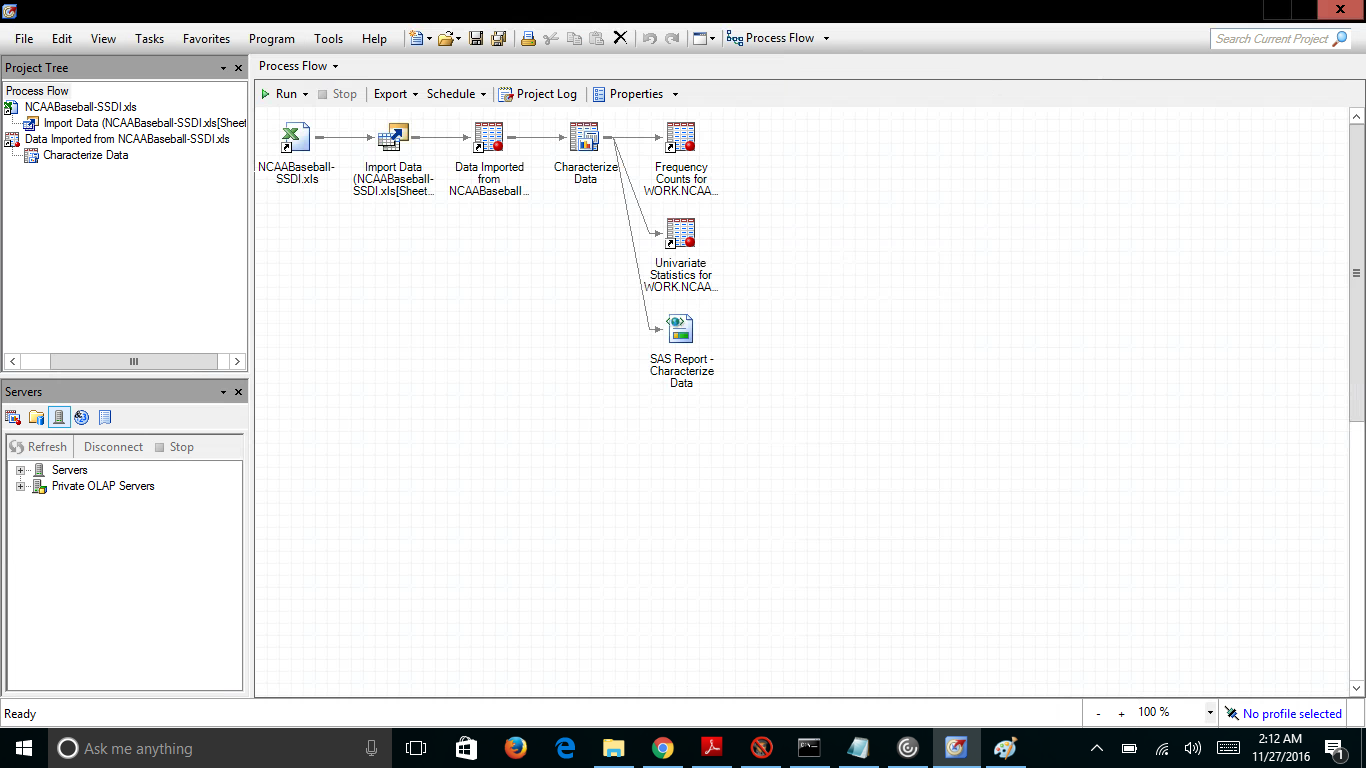
When the file is imported, the SAS Enterprise guide displays it as follows





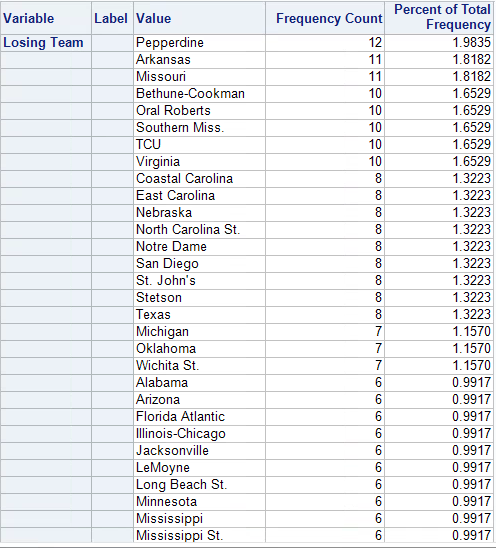
Then it is described using characterize data

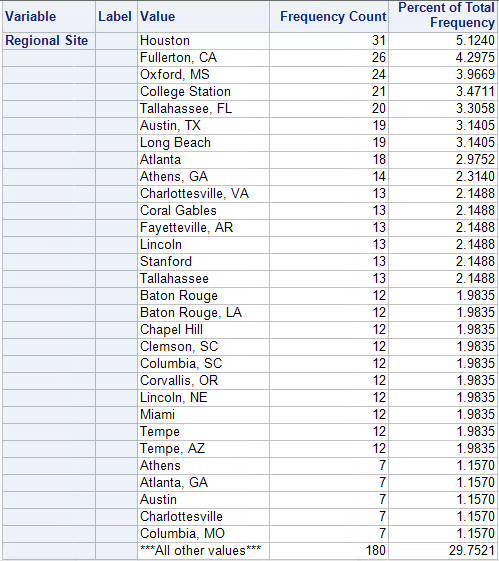
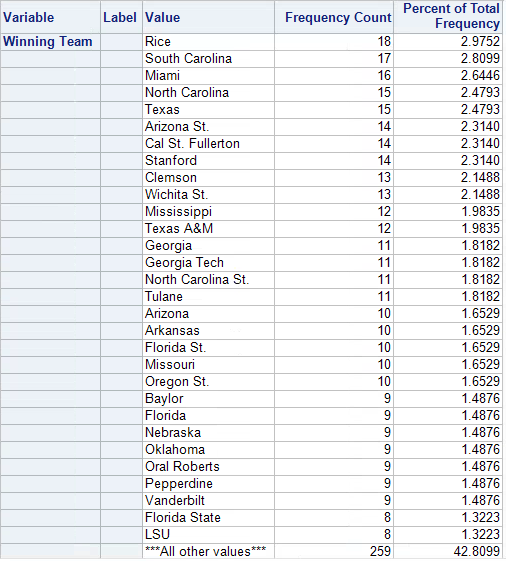
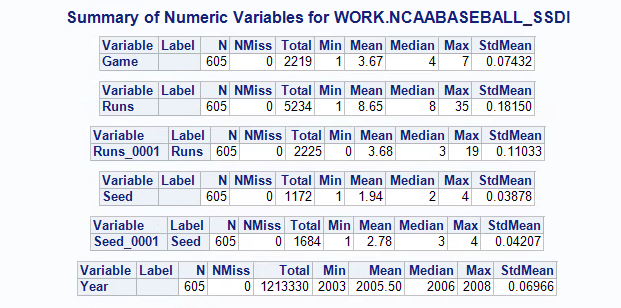


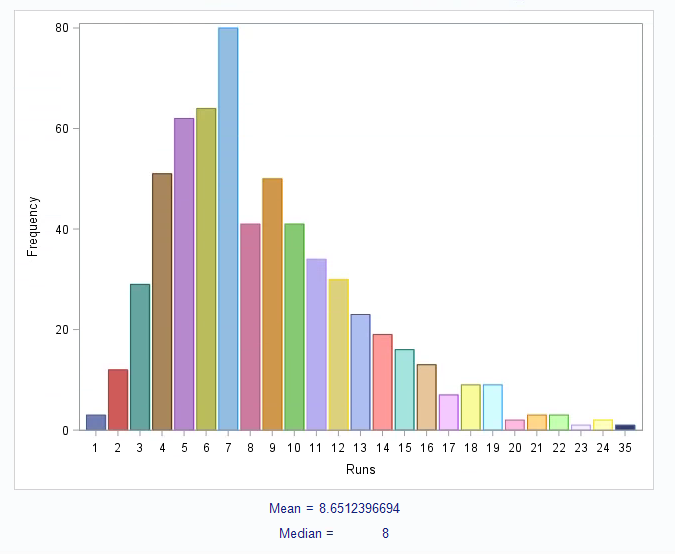


Categorical and numerical variables are







Program:

The python program is run using python interpreter. The program baseball.py is run from the command prompt where the user is asked to select an year. As the dataset contains the data from years 2003-2008, it is better to select year from these years. Once the year is selected, it displays a HTML page where a link is provided for the anal.txt file. The analysis based on that year is stored in that anal.txt file.

Code:

import csv

import cgi, cgitb

import os

import webbrowser

f=open('baseball.csv')

csv\_f=csv.reader(f)

d=0

yr=[]

rs=[]

game=[]

wt=[]

runsw=[]

seedw=[]

lt=[]

runsl=[]

seedl=[]

for row in csv\_f:

yr.append(row[0])

rs.append(row[1])

game.append(row[2])

wt.append(row[3])

runsw.append(row[4])

seedw.append(row[5])

lt.append(row[6])

runsl.append(row[7])

seedl.append(row[8])

d+=1

w=[]

dw=[]

l=[]

dl=[]

r=[]

dr=[]

a=0

aa=1

ab=1

b=0

c=0

m=0

n=0

var = input("Year for which you want analysis: ")

for i in range(d):

if(yr[i]==var):

w.append(wt[i])

l.append(lt[i])

r.append(rs[i])

dw=set (w)

a=len(dw)

dw=list(dw)

dl=set(l)

b=len(dl)

dl=list(dl)

dr=set(r)

c=len(dr)

dr=list(dr)

file=open('anal.txt','w')

if a>0:

file.write ("\n Number of teams which won in year " +var+ " are ")

file.write (str(a))

file.write ("\n Teams which won in year " +var+ " are ")

file.write ("\n"+str(dw))

file.write ("\n")

file.write ("\n")

file.write ("\n")

file.write ("\n------------------------------------------------------------------------------- ")

file.write ("\n")

file.write ("\n")

file.write ("\n Number of teams which lost in year " +var+ " are ")

file.write (str(b))

file.write ("\n Teams which lost in year " +var+ " are ")

file.write ("\n"+str(dl))

file.write ("\n")

file.write ("\n")

file.write ("\n")

file.write ("\n------------------------------------------------------------------------------- ")

file.write ("\n")

file.write ("\n")

file.write ("\n Number of regions in which match was played in year " +var+ " are ")

file.write (str(c))

file.write ("\n All the Regions in which match was played in year " +var+ " are ")

file.write ("\n"+str(dr))

file.write ("\n")

file.write ("\n")

file.write ("\n")

file.write ("\n------------------------------------------------------------------------------- ")

file.write ("\n")

file.write ("\n")

for i in range(d):

if yr[i]==var:

if int(runsw[aa])<int(runsw[i]):

aa=i

#file.write(aa)

file.write("\n Team which did max runs in year " +var+ " are ")

file.write(wt[aa])

file.write("\n runs scored is ")

file.write(runsw[aa])

file.write ("\n")

file.write ("\n")

file.write ("\n")

file.write ("\n------------------------------------------------------------------------------- ")

file.write ("\n")

file.write ("\n")

for i in range(d):

if yr[i]==var:

if int(runsl[aa])>int(runsl[i]):

ab=i

#file.write(aa)

file.write("\n Team which did least runs in year " +var+ " are ")

file.write(lt[ab])

file.write("\n runs scored is ")

file.write(runsl[ab])

else:

file.write("\n Dataset contains only years from 2003 to 2008")

file.close()

f = open('ssd.html','w')

message = """<html>

<head></head>

<body><p>"NCAA BASEBALL DATA ANALYSIS "</p></body>

<p></p>

THE ANALYSIS IS STORED IN THE TEXT FILE

<p></p>

<a href="anal.txt" target="\_explorer.exe">anal</a>

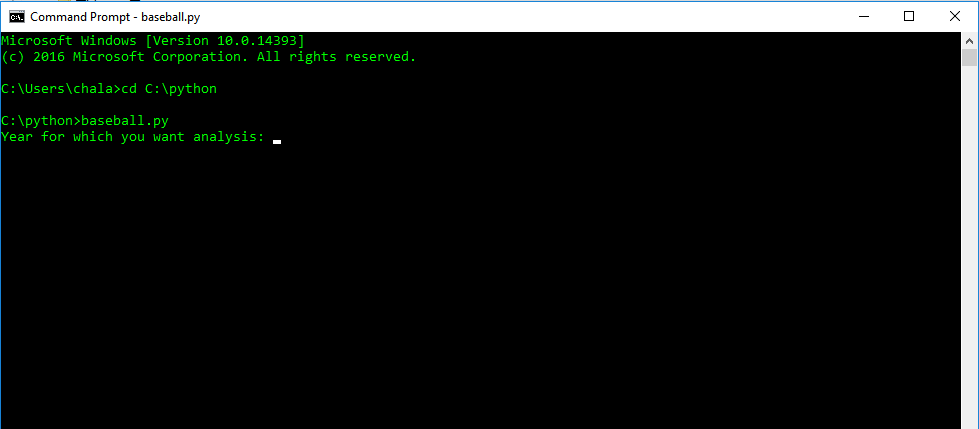
</html>"""

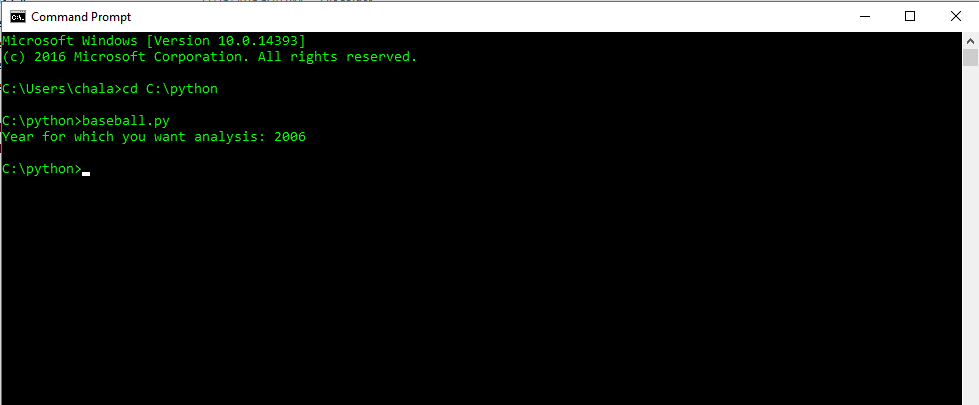
f.write(message)

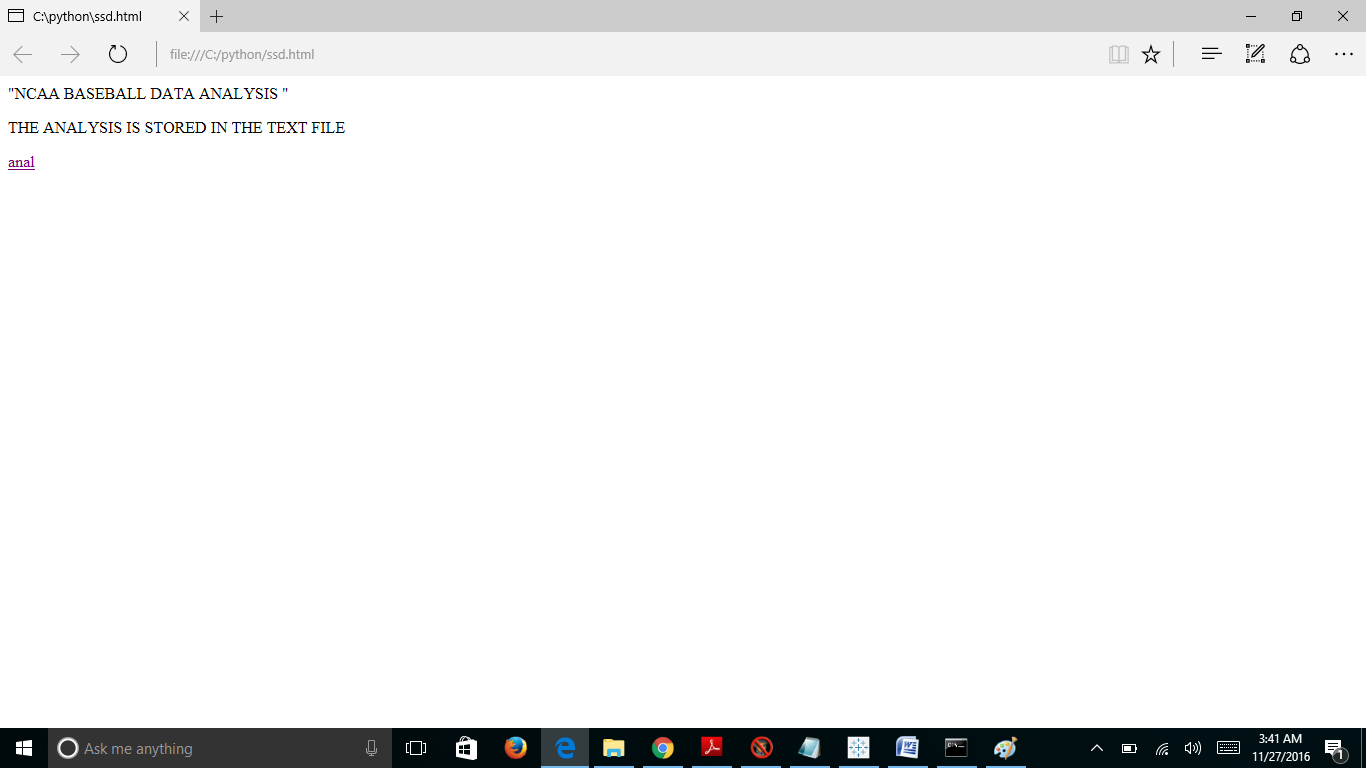
f.close()

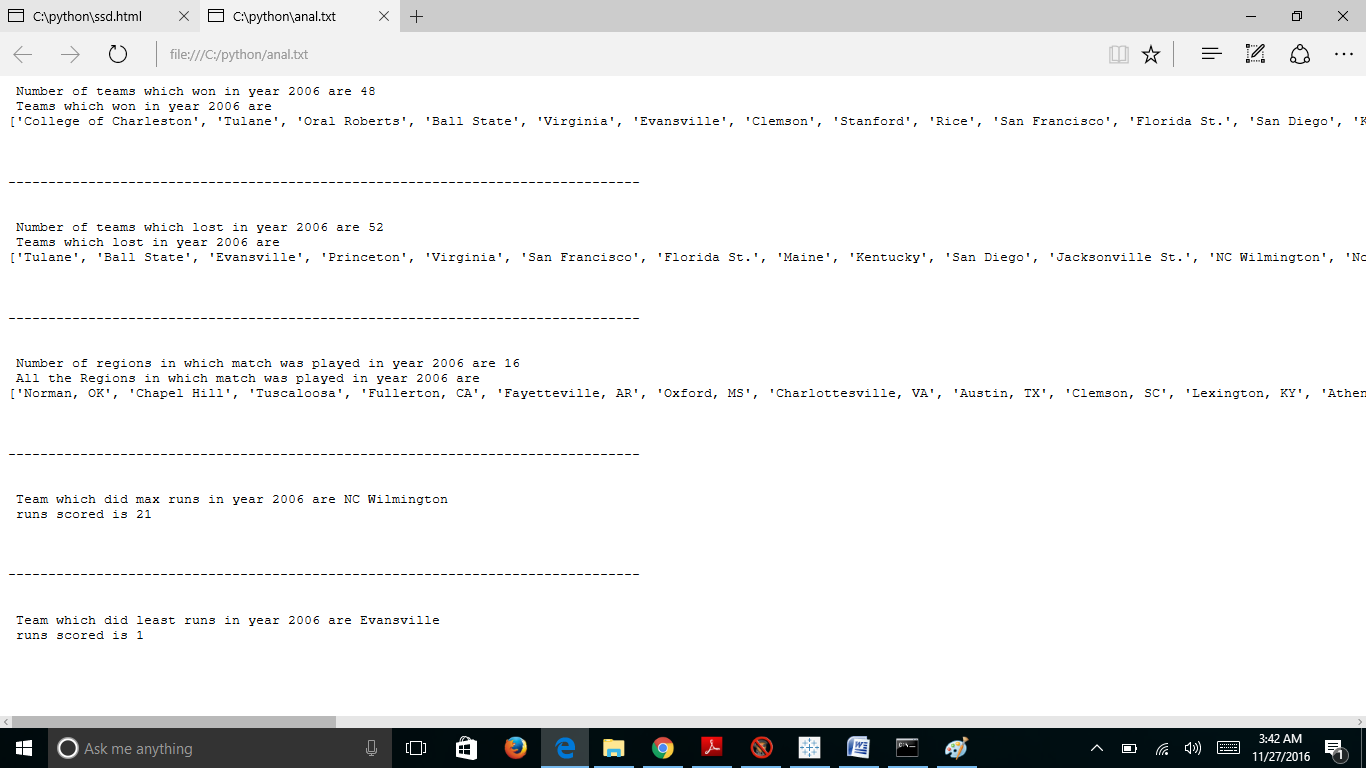
webbrowser.open("ssd.html")

Output screenshots:









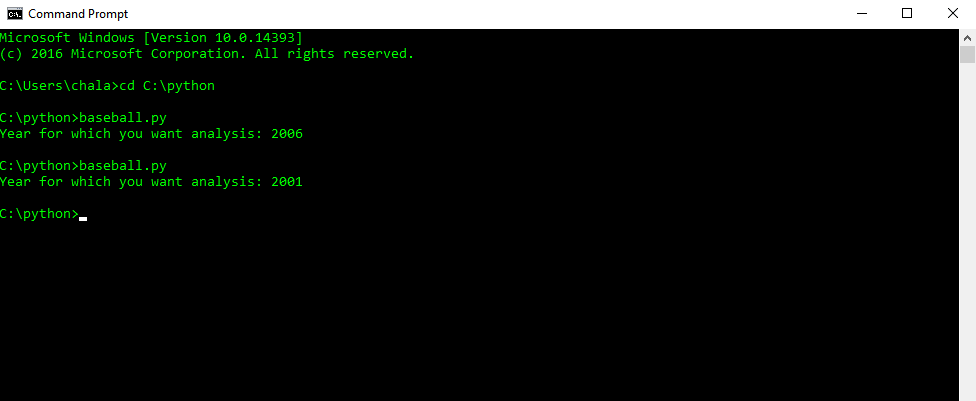
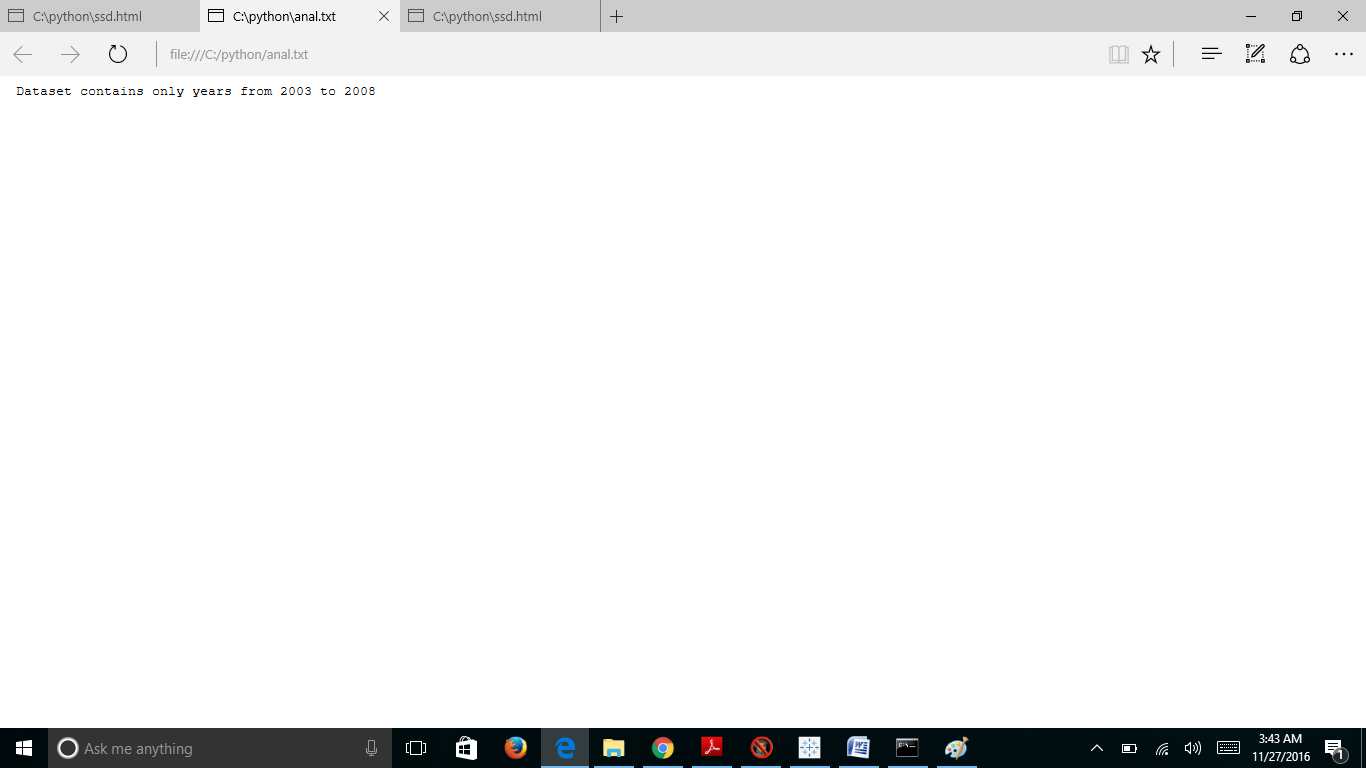
 

Tableau:

Tableau Desktop can be installed from the link below and the zipped folder contains the file “tab.twb” which can be opened after Tableau Desktop is installed. When tab.twb is opened in the Tableau it asks you to locate the NCAABaseball.xls file, it is also uploaded in the zip file. All the worksheets get displayed once the file is located.

<http://www.tableau.com/products/desktop>

Ta