

Webscraping With Python

Michael Babington Christopher Clapp James Freeland

Department of Economics
Florida State University

Strozier Library, 2015



Uses Of Webscraping

Data Gathering



Source: www.phdcomics.com

Webscraping is used to extract information from websites
It has been used to collect data on everything from [airline seat price and availability](#) to [journal article citations](#)



Uses of Webscraping

The basic method is to find patterns in the HTML code, then use a Python program to methodically extract the data you want



Source: www.python.org

Our example will collect data on [FSU Football statistics](#)



Webscrape FSU Football Statistics (seminoles.com)

Date	Opponent	Rushing				Receiving				Passing				Kick Ret				Punt Ret				tot	
		no.	yds	td	lg	no.	yds	td	lg	c-a-i	yds	td	lg	no.	yds	td	lg	no.	yds	td	lg	off	def
9/2/13	at PITT	34	156	1	20	27	377	4	42	27-29-0	377	4	42	3	56	0	26	1	26	0	26	533	
Sep 14, 2013	NEVADA	41	377	6	65	17	240	2	29	17-22-1	240	2	29	2	54	0	28	7	82	0	17	617	
Sep 21, 2013	BETHUNE-COOKMAN	36	266	4	35	13	226	2	42	13-25-0	226	2	42	3	86	0	41	0	0	0	0	492	
Sep 28, 2013	at Boston College	36	159	1	20	17	330	4	56	17-27-1	330	4	56	1	17	0	17	0	0	0	0	489	
Oct 05, 2013	MARYLAND	43	183	4	24	26	431	5	48	26-39-0	431	5	48	1	12	0	12	5	50	0	28	614	
Oct 19, 2013	at Clemson	38	121	2	18	22	444	3	94	22-35-1	444	3	94	2	73	0	43	1	24	0	24	565	
Oct 26, 2013	NC STATE	34	224	4	31	20	342	3	42	20-34-2	342	3	42	3	79	0	34	5	42	0	23	566	
Nov 02, 2013	MIAMI	44	192	4	21	21	325	1	48	21-29-2	325	1	48	2	36	0	23	0	0	0	0	517	
Nov 09, 2013	at Wake Forest	34	89	3	22	22	207	2	18	22-39-1	207	2	18	2	101	1	97	4	20	0	10	296	
Nov 16, 2013	SYRACUSE	19	225	4	74	22	298	3	34	22-26-1	298	3	34	2	7	0	4	3	45	0	25	523	
Nov 23, 2013	IDAHO	40	336	4	60	23	309	5	46	23-39-1	309	5	46	2	96	0	58	4	43	0	20	645	
Nov 30, 2013	at Florida	32	129	1	23	19	327	3	56	19-31-1	327	3	56	2	24	0	19	1	-1	0	0	456	
Dec 07, 2013	vs Duke	43	239	3	22	19	330	3	54	19-32-2	330	3	54	2	60	0	44	3	9	0	6	569	
Jan 06, 2014	vs #2 AUBURN	31	148	1	21	20	237	2	49	20-35-0	237	2	49	4	172	1	100	0	0	0	0	385	
Florida State		505	2844	42	74	288	4423	42	94	288-442-13	4423	42	94	31	873	2	100	34	340	0	28	7267	
Opponent		536	1747	7	72	222	2193	14	68	222-427-26	2193	14	68	70	1301	0	71	14	237	0	45	3940	



Webscrape FSU Football Statistics (seminoles.com)

Date	Opponent	Rushing				Receiving				Passing				Kick Ret				Punt Ret				tot	
		no.	yds	td	lg	no.	yds	td	lg	c-a-i	yds	td	lg	no.	yds	td	lg	no.	yds	td	lg	off	def
9/2/13	at DUTCHMAN	24	156	1	20	27	377	4	42	27-29-0	377	4	42	3	56	0	26	1	26	0	26	533	
Sep 1																							
Sep 2																							
Sep 2																							
Oct 05	AdBlock																						
Oct 19																							
Oct 26, 2013	NC STATE	34	224	4	31	20	342	3	42	20-34-2	342	3	42	3	79	0	34	5	42	0	23	566	
Nov 02, 2013	MIAMI	44	192	4	21	21	325	1	48	21-29-2	325	1	48	2	36	0	23	0	0	0	0	517	
Nov 09, 2013	at Wake Forest	34	89	3	22	22	207	2	18	22-39-1	207	2	18	2	101	1	97	4	20	0	10	296	
Nov 16, 2013	SYRACUSE	19	225	4	74	22	298	3	34	22-26-1	298	3	34	2	7	0	4	3	45	0	25	523	
Nov 23, 2013	IDAHO	40	336	4	60	23	309	5	46	23-39-1	309	5	46	2	96	0	58	4	43	0	20	645	
Nov 30, 2013	at Florida	32	129	1	23	19	327	3	56	19-31-1	327	3	56	2	24	0	19	1	-1	0	0	456	
Dec 07, 2013	vs Duke	43	239	3	22	19	330	3	54	19-32-2	330	3	54	2	60	0	44	3	9	0	6	569	
Jan 06, 2014	vs #2 AUBURN	31	148	1	21	20	237	2	49	20-35-0	237	2	49	4	172	1	100	0	0	0	0	385	
Florida State		505	2844	42	74	288	4423	42	94	288-442-13	4423	42	94	31	873	2	100	34	340	0	28	7267	
Opponent		536	1747	7	72	222	2193	14	68	222-427-26	2193	14	68	70	1301	0	71	14	237	0	45	3940	



SEMINOLES.com
 THE OFFICIAL ATHLETIC WEBSITE OF FLORIDA STATE UNIVERSITY

 SUPPORT YOUR NOLES

 START SHOPPING

 BUY TICKETS

[SPORTS](#) | [TICKETS](#) | [MULTIMEDIA](#) | [DEPARTMENT](#) | [FACILITIES](#) | [FANS](#) | [SHOP](#)

		Rushing			Receiving			Passing			Kick Ret			Punt Ret			tot					
Date	Opponent	no.	yds	td	lg	no.	yds	td	lg	c-a	yds	td	lg	no.	yds	td	lg	no.	yds	td	lg	off
9/2/13	at PITT	34	156	1	20	27	377	4	42	27-29-0	377	4	42	3	56	0	26	1	26	0	26	533
Sep 14, 2013	NEVADA	41	377	6	65	17	240	2	29	17-22-1	240	2	29	2	54	0	28	7	82	0	17	617
Sep 21, 2013	BETHUNE-COOKMAN	36	266	4	35	13	226	2	42	13-25-0	226	2	42	3	86	0	41	0	0	0	0	492
Sep 28, 2013	at Boston College	36	159	1	20	17	330	4	56	17-27-1	330	4	56	1	17	0	17	0	0	0	0	489
Oct 05, 2013	MARYLAND	33	183	4	24	26	431	5	48	26-39-0	431	5	48	1	12	0	12	5	50	0	28	614
Oct 19, 2013	at Clemson	38	121	2	18	22	444	3	94	22-35-1	444	3	94	2	73	0	43	1	24	0	24	565
Oct 26, 2013	NC STATE	34	224	4	31	20	342	3	42	20-34-2	342	3	42	3	79	0	34	5	42	0	23	566
Nov 02, 2013	MIAMI	44	192	4	21	21	325	1	48	21-29-2	325	1	48	2	36	0	23	0	0	0	0	516
Nov 09, 2013	at Wake Forest	34	89	3	22	22	207	2	18	22-39-1	207	2	18	2	101	1	97	4	20	0	10	297
Nov 16, 2013	SYRACUSE	19	225	4	74	22	298	3	34	22-26-1	298	3	34	2	7	0	4	3	45	0	25	523
Nov 23, 2013	IDAHO	40	336	4	60	23	309	5	46	23-39-1	309	5	46	2	96	0	58	4	43	0	20	645
Nov 30, 2013	at Florida	32	129	1	23	19	327	3	56	19-31-1	327	3	56	2	24	0	19	1	-1	0	0	456
Dec 07, 2013	vs Duke	43	239	3	22	19	330	3	54	19-32-2	330	3	54	2	60	0	44	3	9	0	6	569
Jan 06, 2014	vs #2 AUBURN	31	148	1	21	20	237	2	49	20-35-0	237	2	49	4	172	1	100	0	0	0	0	385
Florida State		505	2844	42	74	288	4293	42	94	288-442-13	4423	42	94	31	873	2	100	34	340	0	28	7267
Opponent		536	1747	7	72	222	2193	14	68	222-427-26	2193	14	68	70	1301	0	71	14	237	0	45	3949

[illegible]

SEMINOLES.com
THE OFFICIAL RECREATION WEBSITE OF FLORIDA STATE UNIVERSITY

[SPORTS](#) | [TICKETS](#) | [MULTIMEDIA](#) | [DEPARTMENT](#) | [FACILITIES](#) | [FANS](#) | [SHOP](#)

Date	Opponent	Rushing				Receiving				Passing				Kick Ret				Punt Ret				tot
		no.	yds	td	lg	no.	yds	td	lg	c-a-i	yds	td	lg	no.	yds	td	lg	no.	yds	td	lg	
9/2/13	@ PITT	34	156	1	20	27	377	4	42	27-29-0	377	4	42	3	56	0	26	1	26	0	26	533
tr 7:56a x 23:20a	AIDA	41	377	6	65	17	240	2	29	17-22-1	240	2	29	2	54	0	28	7	82	0	17	617
Sep 21, 2013	BETHUNE-COOKMAN	36	266	4	35	13	226	2	42	13-25-0	226	2	42	3	86	0	41	0	0	0	0	492
Sep 28, 2013	at Boston College	36	159	1	20	17	330	4	56	17-27-1	330	4	56	1	17	0	17	0	0	0	0	489
Oct 05, 2013	MARYLAND	43	183	4	24	26	431	5	48	26-39-0	431	5	48	1	12	0	12	5	50	0	28	614
Oct 19, 2013	at Clemson	38	121	2	18	22	444	3	94	22-35-1	444	3	94	2	73	0	43	1	24	0	24	565
Oct 26, 2013	NC STATE	34	224	4	31	20	342	3	42	20-34-2	342	3	42	3	79	0	34	5	42	0	23	566
Nov 02, 2013	MIAMI	44	192	4	21	21	325	1	48	21-29-2	325	1	48	2	36	0	23	0	0	0	0	516
Nov 09, 2013	at Wake Forest	34	89	3	22	22	207	2	18	22-29-1	207	2	18	2	101	1	97	4	20	0	10	297
Nov 16, 2013	SYRACUSE	19	225	4	74	22	298	3	34	22-26-1	298	3	34	2	7	0	4	3	45	0	25	523
Nov 23, 2013	IDAHO	40	336	4	60	23	309	5	46	23-39-1	309	5	46	2	96	0	58	4	43	0	20	645
Nov 30, 2013	at Florida	32	129	1	23	19	327	3	56	19-31-1	327	3	56	2	24	0	19	1	-1	0	0	456
Dec 07, 2013	vs Duke	43	239	3	22	19	320	3	54	19-32-2	330	3	54	2	60	0	44	3	9	0	6	569
Jan 06, 2014	vs #2 AUBURN	31	148	1	21	20	237	2	49	20-35-0	237	2	49	4	172	1	100	0	0	0	0	385
Florida State		505	2844	47	74	288	4423	42	94	288-442-13	4423	42	94	31	873	2	100	34	340	0	28	7267
Opponent		536	1747	7	72	222	2193	14	68	222-427-26	2193	14	68	70	1301	0	71	14	237	0	45	3941

[illegible]

Webscrape FSU Football Statistics (seminoles.com)


SEMINOLES.com
THE OFFICIAL BULLETIN WEBSITE OF FLORIDA STATE UNIVERSITY


SUPPORT YOUR NOLES

START SHOPPING

BUY TICKETS

[SPORTS](#) | [TICKETS](#) | [MULTIMEDIA](#) | [DEPARTMENT](#) | [FACILITIES](#) | [FANS](#) | [SHOP](#)

TEAM STATISTICS

		Rushing				Receiving				Passing				Kick Ret				Punt Ret				tot
Date	Opponent	no.	yds	td	lg	no.	yds	td	lg	at	yds	td	lg	no.	yds	td	lg	no.	yds	td	lg	off
9/2/13	at PITT	34	156	1	20	27	377	4	42	27-29-0	377	4	42	3	56	0	26	1	26	0	26	533
Sep 14, 2013	NEVADA	41	377	6	65	17	240	2	29	17-22-1	240	2	29	2	54	0	28	7	82	0	17	617
by 756px × 23.2px	HUNE-COOKMAN	36	266	4	35	13	226	2	42	13-25-0	226	2	42	3	86	0	41	0	0	0	0	492
Sep 28, 2013	at Boston College	35	159	1	20	17	330	4	56	17-27-1	330	4	56	1	17	0	17	0	0	0	0	489
Oct 05, 2013	MARYLAND	43	183	4	24	26	431	5	48	26-39-0	431	5	48	1	12	0	12	5	50	0	28	614
Oct 19, 2013	at Clemson	38	121	2	18	22	444	3	94	22-35-1	444	3	94	2	73	0	43	1	24	0	24	565
Oct 26, 2013	NC STATE	34	224	4	31	20	342	3	42	20-34-2	342	3	42	3	79	0	34	5	42	0	23	566
Nov 09, 2013	MIAMI	44	192	4	21	21	325	1	48	21-29-2	325	1	48	2	36	0	23	0	0	0	0	516
Nov 02, 2013	at Wake Forest	34	89	3	22	22	207	2	18	22-29-1	207	2	18	2	101	1	97	4	20	0	10	297
Nov 16, 2013	SYRACUSE	19	225	4	74	22	298	3	34	22-26-1	298	3	34	2	7	0	4	3	45	0	25	523
Nov 23, 2013	IDAHO	40	336	4	60	23	309	5	46	23-39-1	309	5	46	2	96	0	58	4	43	0	20	645
Nov 30, 2013	at Florida	32	129	1	23	19	327	3	56	19-31-1	327	3	56	2	24	0	19	1	-1	0	0	456
Dec 07, 2013	vs Duke	43	239	3	22	19	330	3	54	19-32-2	330	3	54	2	60	0	44	3	9	0	6	569
Jan 06, 2014	vs #2 AUBURN	31	148	1	21	20	237	2	49	20-35-0	237	2	49	4	172	1	100	0	0	0	0	385
Florida State		505	2844	42	74	288	4423	42	94	288-442-13	4423	42	94	31	873	2	100	34	340	0	28	7267
Opponent		536	1747	7	72	222	2193	14	68	222-442-26	2193	14	68	70	1301	0	71	14	237	0	45	3969

[illegible]

Webscrape FSU Football Statistics (seminoles.com)



TEAM STATISTICS

		Rushing				Receiving				Passing				Kick Ret		Punt Ret		tot					
Date	Opponent	no.	yds	tg	no.	yds	tg	no.	yds	tg	no.	yds	tg	no.	yds	tg	tot						
9/2/13	at PITT	34	156	1	20	27	377	4	42	27-29-0	377	4	42	3	56	0	26	1	26	0	26	533	
td 75.2px	td 23.2px	ADA	41	377	6	65	17	240	2	49	17-22-1	240	2	29	2	54	0	28	7	82	0	17	617
Sep 21, 2013	BETHUNE-COOKMAN	36	266	4	35	13	226	2	42	13-25-0	226	2	42	3	86	0	41	0	0	0	0	0	492
Sep 28, 2013	at Boston College	36	159	1	20	17	330	4	56	17-27-1	330	4	56	1	17	0	17	0	0	0	0	0	489
Oct 05, 2013	MARYLAND	43	183	4	24	26	431	5	48	26-39-0	431	5	48	1	12	0	12	5	50	0	28	614	
Oct 19, 2013	at Clemson	38	121	2	18	22	444	3	94	22-35-1	444	3	94	2	73	0	43	1	24	0	24	565	
Oct 26, 2013	NC STATE	34	224	4	31	20	342	3	42	20-34-2	342	3	42	3	79	0	34	5	42	0	23	566	
Nov 02, 2013	MIAMI	44	192	4	21	21	325	1	48	21-29-1	325	1	48	2	36	0	23	0	0	0	0	0	516
Nov 09, 2013	at Wake Forest	34	89	3	22	22	207	2	18	22-39-2	207	2	18	2	101	1	97	4	20	0	10	297	
Nov 16, 2013	SYRACUSE	19	225	4	74	22	298	3	34	22-26-1	298	3	34	2	7	0	4	3	45	0	25	523	
Nov 23, 2013	IDAHO	40	336	4	60	23	309	5	46	23-39-1	309	5	46	2	96	0	58	4	43	0	20	645	
Nov 30, 2013	at Florida	32	129	1	23	19	327	3	56	19-31-1	327	3	56	2	24	0	19	1	-1	0	0	456	
Dec 07, 2013	vs Duke	43	239	3	22	19	330	3	54	19-32-2	330	3	54	2	60	0	44	3	9	0	6	569	
Jan 06, 2014	vs #2 AUBURN	31	148	1	21	20	237	2	49	20-35-0	237	2	49	4	172	1	100	0	0	0	0	0	385
Florida State		505	2844	42	74	288	4423	42	94	288-442-13	4423	42	94	31	873	2	100	34	340	0	28	7267	
Opponent		536	1747	7	72	222	2193	14	68	222-427-26	2193	14	68	70	1301	0	71	14	237	0	45	3949	



Webscrape FSU Football Statistics (seminoles.com)



TEAM STATISTICS

		Rushing				Receiving				Passing				Kick Ret				Punt Ret				Tot	
Date	Opponent	no.	yds	tg	no.	yds	tg	no.	yds	tg	no.	yds	tg	no.	yds	tg	no.	yds	tg	no.	yds	off	
9/2/13	at PITT	34	156	1	20	27	377	4	42	27-29-0	377	4	42	3	56	0	26	1	26	0	26	533	
Sep 14, 2013	td 113.6px × 23.2px	41	377	6	65	17	440	2	29	17-22-1	240	2	29	2	54	0	28	7	82	0	17	617	
Sep 21, 2013	BETHUNE-COOKMAN	36	266	4	35	13	226	2	42	13-25-0	226	2	42	3	86	0	41	0	0	0	0	492	
Sep 28, 2013	at Boston College	36	159	1	20	17	330	4	56	17-27-1	330	4	56	1	17	0	17	0	0	0	0	489	
Oct 05, 2013	MARYLAND	43	183	4	24	26	431	4	48	26-39-0	431	5	48	1	12	0	12	5	50	0	28	614	
Oct 19, 2013	at Clemson	38	121	2	18	22	444	3	94	22-35-1	444	3	94	2	73	0	43	1	24	0	24	565	
Oct 26, 2013	NC STATE	34	224	4	31	20	342	3	42	20-34-2	342	3	42	3	79	0	34	5	42	0	23	566	
Nov 02, 2013	MIAMI	44	192	4	21	21	325	1	48	21-29-2	325	1	48	2	36	0	23	0	0	0	0	517	
Nov 09, 2013	at Wake Forest	34	89	3	22	22	207	2	18	22-39-1	207	2	18	2	101	1	97	4	20	0	10	298	
Nov 16, 2013	SYRACUSE	19	225	4	74	22	298	3	34	22-26-1	298	3	34	2	7	0	4	3	45	0	25	523	
Nov 23, 2013	IDAHO	40	336	4	60	23	309	5	46	23-39-1	309	5	46	2	96	0	58	4	43	0	20	645	
Nov 30, 2013	at Florida	32	129	1	23	19	327	3	56	19-31-1	327	3	56	2	24	0	19	1	-1	0	0	450	
Dec 07, 2013	vs Duke	43	239	3	22	19	330	3	54	19-32-2	330	3	54	2	60	0	44	3	9	0	6	569	
Jan 06, 2014	vs #2 AUBURN	31	148	1	21	20	237	2	49	20-35-0	237	2	49	4	172	1	100	0	0	0	0	385	
Florida State		505	2844	42	74	288	4423	42	94	288-442-13	4423	42	94	31	873	2	100	34	340	0	28	7267	
Opponent		536	1747	7	72	222	2193	14	68	222-427-26	2193	14	68	70	1301	0	71	14	237	0	45	3905	

[illegible]

Coding Outline

Program Outline

Use `requests` to import HTML code into Python

Use `Beautiful Soup` to make HTML code “readable”
Python

Point Python to the location of the data you want

Loop over the data to get it into a usable form

Full code will be available online



Results

	A	B	C	D	E	F	G	H	I	J
1	Date	Opponent	Rushing Att	Rushing Ya	Rushing Tc	Longest Ru	Number Ru	Receiving	Receiving	Longest Re
2	9/2/13	at PITT	34	156	1	20	27	377	4	42
3	Sep 14, 20	NEVADA	41	377	6	65	17	240	2	29
4	Sep 21, 20	BETHUNE	36	266	4	35	13	226	2	42
5	Sep 28, 20	at Boston	36	159	1	20	17	330	4	56
6	Oct 05, 20	MARYLAN	43	183	4	24	26	431	5	48
7	Oct 19, 20	at Clemso	38	121	2	18	22	444	3	94
8	Oct 26, 20	NC STATE	34	224	4	31	20	342	3	42
9	Nov 02, 20	MIAMI	44	192	4	21	21	325	1	48
10	Nov 09, 20	at Wake F	34	89	3	22	22	207	2	18
11	Nov 16, 20	SYRACUSE	19	225	4	74	22	298	3	34
12	Nov 23, 20	IDAHO	40	336	4	60	23	309	5	46
13	Nov 30, 20	at Florida	32	129	1	23	19	327	3	56
14	Dec 07, 20	vs Duke	43	239	3	22	19	330	3	54
15	Jan 06, 20	vs #2 AUB	31	148	1	21	20	237	2	49
16	Florida Sta		505	2844	42	74	288	4423	42	94
17	Opponent		536	1747	7	72	222	2193	14	68



Contact Information And Resources

Contact Information

Michael Babington: mb13m@my.fsu.edu

James Freeland: jef08e@my.fsu.edu

Resources

[Youtube tutorial using Yellow Pages](#)

[Beautiful Soup Documentation](#)

[Download Python](#)



Conclusion

Webscraping is a useful research tool

It gives you access to new and exciting data

Presentation and Code will be available at
<http://chrisclapp.org/teaching/>

