## Webscraping Yelp

In this tutorial, we will go through the various steps needed to gather user review data from websites. Although this example is based on the yelp.com website, the general techniques used here can be applied to other websites. As a word of CAUTION scraping sometimes breaks a company terms of use, this could get you in trouble with the law, thus one should be careful about which sites you scrape.

To Achieve our goals, we will make use of the Python. Python is a versatile programming language and I *highly* recommend it for text mining. Because Python is a real programming language, it contains a lot of packages to perform the various text mining steps and great data structures to arrange the data neatly. The packages we will use for this process are requests, BeautifullSoup, and numpy

```
In [1]: import pandas as pd
import numpy as np
import collections, csv, re
import requests
import time
import os.path
from bs4 import BeautifulSoup
```

```
In [2]: url = 'https://www.yelp.com/biz/dominos-pizza-brooklyn-6'
```

```
In [3]: webpage = requests.get(url)
```

/usr/local/lib/python2.7/dist-packages/requests/packages/urllib3/util/ssl\_.py:100: InsecurePlatfor mWarning: A true SSLContext object is not available. This prevents urllib3 from configuring SSL ap propriately and may cause certain SSL connections to fail. For more information, see https://urllib3.readthedocs.org/en/latest/security.html#insecureplatformwarning.

InsecurePlatformWarning

Before proceeding to the analysis we need to check the status code of our request. Here are a few Status Code Definition:

```
* 200 : OK, Our request was accepted by the website
* 400 : Bad Request, Our request was bad, i.e. not properly formatted
* 401 : Unauthorized, our request is Unauthorized, thus authentification is required
* 403 : Forbidden, our request was received but the server is *NOT* processing it! probably due to rate limits
```

For more information on Http Status code: <a href="https://www.w3.org/Protocols/rfc2616/rfc2616-sec10.html">https://www.w3.org/Protocols/rfc2616-sec10.html</a> (<a href="https://www.w3.org/Protocols/rfc2616-sec10.html">https://www.w3.org/Protocols/rfc2616-sec10.html</a>)

```
In [4]: webpage
Out[4]: <Response [200]>
```

Our request was accepted by the server. When automating this process one should always check if our request was process. But what did we get exactly? Let us take a peek at the data.

```
In [5]: webpage.text[:1000]
Out[5]: u'<!DOCTYPE HTML>\n\n<!--[if lt IE 7 ]> <html xmlns:fb="http://www.facebook.com/2008/fbml" class
       ="ie6 ie ltie9 ltie8 no-js" lang="en"> <![endif]-->\n<!--[if IE 7 ]>
                                                                          <html xmlns:fb="http://ww
       w.facebook.com/2008/fbml" class="ie7 ie ltie9 ltie8 no-js" lang="en"> <![endif]-->\n<!--[if IE 8 ]
            <html xmlns:fb="http://www.facebook.com/2008/fbml" class="ie8 ie ltie9 no-js" lang="en"> <![e</pre>
       ndif]-->\n<!--[if IE 9 ]>
                                  <html xmlns:fb="http://www.facebook.com/2008/fbml" class="ie9 ie no-j</pre>
       m/2008/fbml" class="no-js" lang="en"> <!--<![endif]-->\n
                                                              <head>\n\n<script type="text/javascr</pre>
                         window.yPageStart = new Date().getTime();\n\n</script>\n\n\n<script type="tex</pre>
       ipt">\n
       t/javascript">\n
                                 var initialVisibilityState = document.webkitVisibilityState;\n\n
                yPerfTimings = [];\n\n
                                                       ySitRepParams = {"clientIP": "128.59.172.127",
        "datacenter": "iad1", "action": "default", "uni'
```

As we can see this is just the raw html output.

An Html is just a markup language used in creating websites. Html consists of several keys and tags (h1, div(multi column layout), class, p etc.). Organizing and finding these keys and tags makes it possible for your program to access the content you are trying to scrape.

To organize it in a structured way, we will use BeautifulSoup.

```
In [6]: soup = BeautifulSoup(webpage.text)
```

Let us start by extracting the first review from the website. To achieve this, we will make used of the different tags in the raw html. Historically, people use to look at the entire raw files to find the tags that they needed. This process used to be long and tedious. However, nowadays this process is greatly simplified with modern internet browser. Thanks in large part to features like *inspect* in google chrome.

Using this feature, we see that the first review is the . Using this knowledge, we can look for this tag in the "soup" object created above.

```
In [7]: review = soup.find("p", {"itemprop" : "description"})
```

In [8]: review

Unfortunately, we see that this output contains a lot of html such as < br > and the . To obtain the content from the data,we make use of the get\_text method

```
In [9]: review.get_text()
```

Out[9]: u'From the "fake pizza" spots Dominos by far is my favorite! When I refer to fake pizza spots I me an Papa Johns, Ceaser\'s & Etc. The one that is not too far from my house is located on neptune av e and west 12th street. pretty close. I believe they have a delivery guy who drives the big body r ange rover!!!! ballin! Must be a busy location and he must be doing a lot of deliveries. Good for him. Always happy when good honest people are making money and enjoying it. This past Saturday my wife sent me on some errands of which one of those was to buy pizza for our daughter. I was contem plating where to go. and said f it lets do dominos so I can pig out my self and get myself a whole pepperoni medium pie. I called in for a pick up. The guy was helpful on the phone with placing my order quick and offering me the appropriate coupon.I ran quickly to the tailors and on my ride bac k the 2 pies were ready for pick up. the kid at the window was a bit weird but I was happy with the fact that the pies were ready and I could keep things moving.Once again! the best "fake pizza" a round.P.S. yes it is mind boggling for me as well. That with so many freaking delicious Pizza spot s around our neighborhood and what Brooklyn is known for that we still sometimes turn to Dominos.'

It should be said that soup.find("p",{"lang": "en"}) would return the same results(In this case!). The next question, one could ask is what if I want all the reviews from this page? Thankfully, BeautifulSoup has a findAll method that makes it very easy to obtain.

```
In [10]: all_review = soup.findAll("p", {"itemprop" : "description"})
```

In [11]: all\_review

Out[11]: [From the "fake pizza" spots Dominos by far is my favorite! When I refer to fake pizza spots I mean Papa Johns, Ceaser's & Etc. <br/>
busy location and he must be doing a lot of deliveries. Good for him. Always happy when good hon est people are making money and enjoying it. <br/>
busy location and he must be doing a lot of deliveries. Good for him. Always happy when good hon est people are making money and enjoying it. <br/>
busy location and he must be doing a lot of deliveries. Good for him. Always happy when good hon est people are making money and enjoying it. <br/>
busy location and he must be doing a lot of deliveries. Good for him. Always happy when good hon est people are making money and enjoying it. <br/>
busy location and he must be doing a lot of deliveries. Good for him. Always happy when good hon est people are making money and enjoying it. <br/>
busy location and he must be doing a lot of deliveries. Good for him. Always happy when good hon est people are making wnoney and enjoying it. <br/>
busy location halways happy when good hon est people are making where to go. and said f it lets do dominos so I can pig out my self and get myself a whole pepperoni medi um pie. I called in for a pick up. The guy was helpful on the phone with placing my order quick and offering me the appropriate coupon.<br/>
busy location halways happy when good hon est people are making where to go. and said f it lets do dominos so I can pig out my self and get myself a whole pepperoni medi um pie. I called in for a pick up. The guy was helpful on the phone with placing my order quick and offering me the appropriate coupon.<br/>
busy location halways happy when good hon est people are making where to go. and said f it lets do dominos so I can pig out my self and get myself a whole pepperoni medi um pie. I called in for a pick up. The guy was helpful on the phone with placing my wife sent me or some peroni location with placing my wife sent me or some peroni location halways

- I Ordered two pizzas, using the app. The pizza was taking s
  o long like 3 hours so I decided to call the place but they hang up on me. Long story short I ne
  ver got my pizzas. Needles to say I will not order from this place again, and I would't recommen
  d it to anyone.,
- Will never order from again, refuse to come inside and actu
  ally deliver your pizza. Worst delivery ever. Tried more than once. Once it was "company" polic
  y, this time it's because he double parked. Nonsense. Pizza came uncooked, waste of money.,
- Worst dominos ever!took almost 2hrs to arrive, pizza was bu
  rnt. When I called to talk to a manager the girl left me on hold for 15 mins. I had to call usin
  g another phone while on hold. Then she hung up on me. I called corporate office, the manager c
  alled me hours later offering me a new pizza. No THANKS! Never again
- Is there a zero star I ordered a med pizza Pepperoni and ba
  con and it was crap the dough was under cooked but get this my Bbq Bacon Chicken was over cooked
  like a hot pile of shit just one stiff mess for food. It is like who the the hell is watching th
  ese people. The good news I did not get sick. Why write the review well the last time at this sa
  me store my food was over an hour late and I told the drive I wanted my food hot, he said he wou
  ld bring me another. Never shoes up and calling the store the phone was off the hook, I called t
  he complaint line and nothing happened and no one called me. My advice order local pizza not fro
  m here. Order number 217110
- They had a sanitation grade of B when I came here before th
  en had nothing up the last time I came here. Also, it's all through glass so you have to wonder
  how violent the area around here is for them to do the entire transaction through glass windows.
  Any other one I've been to was NOT set up that way; some even had seating for people. Ghetto m
  uch?
- So I see that this domino's has some bad reviews but I can
  honestly say that anytime I've dealt with this location they were awesome on the phone or in per
  son very professional and polite makes it feel like they really care about the customers.... no
  w the 13th Avenue location needs to be completely shut down they need to rehire ppl for that loc
  ation the worse yet !!!!!,
- If I could give zero stars I would. The food is good, if yo
  u ever get it. They refuse to come in buildings. They say they try to call (but don't). It was
  n't just one time. We've tried three times now. Each time, they take over twice the quoted time
  and we have to meet them. They call refusing to get out of their car! It's absolutely rediculou
  s..
- This was by far the most horrid experience I have ever went
  through ordering food. The first time I ordered food online I got a call 20 min later letting me
  know a driver can t deliver... That was great but I thought let me try again... After I got hung
  up on 3 times and called back yet again to place my order it took around and hr and 57 min for t
  hem to deliver me my food, not only did I try to call yet again and no one would even answer me
  and when they did they would hang up... I never received such poor customer service. When the dr
  iver got here he called and told me to come out even though direction specified for him to come
  up. As I came down he said he didn't want to look for parking then handed me someone else's orde
  r... I asked for my receipt and he said he had forgotten it... I asked where my garlic sauce was
  and he said they ran out. Are you kidding ?! Never again.
- I called this Dominoes and a girl named Kiana picked up. Te
  rrible service, oh my she was extremely RATCHET it's was horrible.. Coney Island has to do bette
  r,
- Ordered pizza two times. Pizza was under cooked!! Don't ord
  er from this location. Waste of time and money,

Worst experience ever. I made a yelp account only to leave
this review it was that bad.,

I was about to order pizza from this place but something to
ld me to check yelp reviews first lol thank u all,

Definitely the worst experience I have had in New York Cit y. I love Dominos, but this place is horrendous. Fire your staff, Fire your drivers, and start o ver in Noah Ark fashion. If you want dominos, your best bet is to choose somewhere else and just order take out. I ordered and was giving an email with a 21-31 wait time. At 35 minutes, I calle d and was told the driver was on the way. Ok, no problem. At 45 minutes, no show, so I called ag ain...and again I was told the same thing, the driver is on the way and he will call me once he gets there, but this time there was attitude on the other side of the phone. Look woman, I am hu ngry and the order is late, and you are the one getting attitude?!?!?!?! I decided to wait more, and at 55 minutes I called back...the woman who answered (the one with attitude) told me that sh e was sure that the wait time wasn't 21-31 minutes, even though they had already told me the dri ver left 20 minutes ago. I told her, please call the driver and get an update, but she refused t o do so or do anything that was helpful. She proceeded to hang up on me. At 1 hour, I decided to call back and this time a guy picked up the phone. I told him the order was to be canceled and t he money refunded. I was out of time and had to leave my home so couldn't wait any longer. He to ld me I could keep the pizza, but I said no thank you...If the driver really did leave 30 minute s ago (which I am sure he didn't) then that pizza is ice cold, or at least, I don't want to eat something that has been sitting in a car for 30 minutes. <br/><br/>I am not the only 1 star rati ng of this place, so this is definitely not an isolated incident. Blame the workers all you wan t, but essentially, this is a management issue. Poor managers lead to poor hiring, and poor hiri ng leads to poor service. <br/>
<br/>Never ordering from here again and you should save your time and money and do the same.,

So drunchies hit us at midnight!! Placed an order on their
app on my phone and an hour ago the tracking feature said my food has arrived and to dig in.. St
ill no food! I called and they legit don't know their ass from their elbow! They said it's alrea
dy been sent out...still nothing. I called back again and they said the guy is just leaving now
and it'll take 15 more mins. I was furious but said okay.. Half an hour later still nothing. I c
alled back and this went on 3 more times. Worst of all they call me to come down and no one was
there! In the end I eventually got my food an hour later! Can't believe how terribly run it is,
how are they still running? !! Should of listened to the other yelpers.. Not worth the headache!

I called in for the 5,5,5 deal and the customer service re
p, Crystal, was the sweetest. Incredibly patient, repeated my order, and took time to be pleasan
t. Will definitely order again. Thank you, Crystal!,

Horrible, ordered at 1:30am and was told an hour. Called at
2:15am to ensure they're on their way and I'm told another 20min. Called again at 3am and told a
nother 15min. Its 3:25am and no call back or arrival. If they couldn't fulfill the order they sh
ouldn't have lead me on. Worst Dominoes service.]

Let's again clean the data and input a line break to separate the different review

In [12]: for review in all\_review:
 print review.get\_text()
 print "\n"

From the "fake pizza" spots Dominos by far is my favorite! When I refer to fake pizza spots I mean Papa Johns, Ceaser's & Etc. The one that is not too far from my house is located on neptune ave an d west 12th street. pretty close. I believe they have a delivery guy who drives the big body range rover!!!! ballin! Must be a busy location and he must be doing a lot of deliveries. Good for him. Always happy when good honest people are making money and enjoying it. This past Saturday my wife sent me on some errands of which one of those was to buy pizza for our daughter. I was contemplating where to go. and said f it lets do dominos so I can pig out my self and get myself a whole pepp eroni medium pie. I called in for a pick up. The guy was helpful on the phone with placing my order quick and offering me the appropriate coupon.I ran quickly to the tailors and on my ride back the 2 pies were ready for pick up. the kid at the window was a bit weird but I was happy with the fact that the pies were ready and I could keep things moving.Once again! the best "fake pizza" around d.P.S. yes it is mind boggling for me as well. That with so many freaking delicious Pizza spots around our neighborhood and what Brooklyn is known for that we still sometimes turn to Dominos.

I Ordered two pizzas, using the app. The pizza was taking so long like 3 hours so I decided to cal 1 the place but they hang up on me. Long story short I never got my pizzas. Needles to say I will not order from this place again, and I would't recommend it to anyone.

Will never order from again, refuse to come inside and actually deliver your pizza. Worst delivery ever. Tried more than once. Once it was "company" policy, this time it's because he double parked. Nonsense. Pizza came uncooked, waste of money.

Worst dominos ever!took almost 2hrs to arrive, pizza was burnt. When I called to talk to a manager the girl left me on hold for 15 mins. I had to call using another phone while on hold. Then she h ung up on me. I called corporate office, the manager called me hours later offering me a new pizz a. No THANKS! Never again

Is there a zero star I ordered a med pizza Pepperoni and bacon and it was crap the dough was under cooked but get this my Bbq Bacon Chicken was over cooked like a hot pile of shit just one stiff me ss for food. It is like who the the hell is watching these people. The good news I did not get sic k. Why write the review well the last time at this same store my food was over an hour late and I told the drive I wanted my food hot, he said he would bring me another. Never shoes up and calling the store the phone was off the hook, I called the complaint line and nothing happened and no one called me. My advice order local pizza not from here. Order number 217110

Horrible Service. They take forever to create your pizza and quality check for a long time. When on e tries to call the store and contact them and or have concerns the store hangs up the phone on yo u and the manager is extremely rude. They also never seem to have staff so your delivery will take 10 hours. HORRIBLE PLACE!!! RUN BY RETARDS THAT DON'T APPRECIATE THEIR CUSTOMERS. HORRIBLE!!!! THI S SHOULD RECEIVE -5 stars.

They had a sanitation grade of B when I came here before then had nothing up the last time I came here. Also, it's all through glass so you have to wonder how violent the area around here is for them to do the entire transaction through glass windows. Any other one I've been to was NOT set up that way; some even had seating for people. Ghetto much?

HORRIBLE!Just like everyone else on Yelp, my order came after 2 hours. Of course, by the time it c ame it was cold. Worst experience ever. And when I call them to ask when is it going to come, I'm on hold for 30 minutes. What pizza store puts their customer on hold for 30 minutes?Will never ord er again. Oh, and the delivery guy just sat in his car and rolled down his window to give me the p izza. Why am I paying a delivery charge again? If i have to go out to the street while he sits his lazy ass in his car?Papa John's only from now on, it is! STAY FAR AWAY!!!!

So I see that this domino's has some bad reviews but I can honestly say that anytime I've dealt wi th this location they were awesome on the phone or in person very professional and polite makes it feel like they really care about the customers.... now the 13th Avenue location needs to be compl etely shut down they need to rehire ppl for that location the worse yet !!!!!

If I could give zero stars I would. The food is good, if you ever get it. They refuse to come in b uildings. They say they try to call (but don't). It wasn't just one time. We've tried three times now. Each time, they take over twice the quoted time and we have to meet them. They call refusing to get out of their car! It's absolutely rediculous.

This was by far the most horrid experience I have ever went through ordering food. The first time I ordered food online I got a call 20 min later letting me know a driver can t deliver... That was great but I thought let me try again... After I got hung up on 3 times and called back yet again t

o place my order it took around and hr and 57 min for them to deliver me my food, not only did I t ry to call yet again and no one would even answer me and when they did they would hang up... I nev er received such poor customer service. When the driver got here he called and told me to come out even though direction specified for him to come up. As I came down he said he didn't want to look for parking then handed me someone else's order... I asked for my receipt and he said he had forgo tten it... I asked where my garlic sauce was and he said they ran out. Are you kidding ?! Never a gain.

I called this Dominoes and a girl named Kiana picked up. Terrible service, oh my she was extremely RATCHET it's was horrible.. Coney Island has to do better

Ordered pizza two times. Pizza was under cooked!! Don't order from this location. Waste of time an d money

Worst experience ever. I made a yelp account only to leave this review it was that bad.

I was about to order pizza from this place but something told me to check yelp reviews first lol t

Definitely the worst experience I have had in New York City. I love Dominos, but this place is hor rendous. Fire your staff, Fire your drivers, and start over in Noah Ark fashion. If you want domin os, your best bet is to choose somewhere else and just order take out. I ordered and was giving an email with a 21-31 wait time. At 35 minutes, I called and was told the driver was on the way. Ok, no problem. At 45 minutes, no show, so I called again...and again I was told the same thing, the d river is on the way and he will call me once he gets there, but this time there was attitude on th e other side of the phone. Look woman, I am hungry and the order is late, and you are the one gett ing attitude?!?!?!?! I decided to wait more, and at 55 minutes I called back...the woman who answe red (the one with attitude) told me that she was sure that the wait time wasn't 21-31 minutes, eve n though they had already told me the driver left 20 minutes ago. I told her, please call the driv er and get an update, but she refused to do so or do anything that was helpful. She proceeded to h ang up on me. At 1 hour, I decided to call back and this time a guy picked up the phone. I told hi m the order was to be canceled and the money refunded. I was out of time and had to leave my home so couldn't wait any longer. He told me I could keep the pizza, but I said no thank you...If the d river really did leave 30 minutes ago (which I am sure he didn't) then that pizza is ice cold, or at least, I don't want to eat something that has been sitting in a car for 30 minutes. I am not th e only 1 star rating of this place, so this is definitely not an isolated incident. Blame the work ers all you want, but essentially, this is a management issue. Poor managers lead to poor hiring, and poor hiring leads to poor service. Never ordering from here again and you should save your tim e and money and do the same.

So drunchies hit us at midnight!! Placed an order on their app on my phone and an hour ago the tracking feature said my food has arrived and to dig in.. Still no food! I called and they legit do n't know their ass from their elbow! They said it's already been sent out...still nothing. I called back again and they said the guy is just leaving now and it'll take 15 more mins. I was furious but said okay.. Half an hour later still nothing. I called back and this went on 3 more times. Wor st of all they call me to come down and no one was there! In the end I eventually got my food an h our later! Can't believe how terribly run it is, how are they still running? !! Should of listened to the other yelpers.. Not worth the headache!

I called in for the 5,5,5 deal and the customer service rep, Crystal, was the sweetest. Incredibly patient, repeated my order, and took time to be pleasant. Will definitely order again. Thank you, Crystal!

I absolutely love Domino's. It's one of my favorite guilty pleasures. Unfortunately I can no lon ger order delivery to my house. For some reason some drivers will deliver to your door and some w on't. The last driver spent 20 MINUTES calling me and ringing my bell from the lobby but would no t come to my door. I had injured my foot so I was unable to leave and pick up the food (which is why I ordered delivery in the first place). I eventually told the driver to either come upstairs or get the hell out, and he choose to leave. In 20 minutes he could've taken a dozen trips from my apartment to his car. WTF?To charge a delivery fee and expect a tip when you won't even come to someone's door is absolutely ridiculous. This is Brooklyn. We have some of the BEST pizza in the world. L&B's delivers straight to your door, no delivery fee, no hassle, amazing food. So does Lenny's, Pino's, and if you want the good crappy stuff, so does Papa John's. Why bother with the hassle from ordering from this place? It's not worth it.Get it together Domino's!!

Horrible, ordered at 1:30am and was told an hour. Called at 2:15am to ensure they're on their way and I'm told another 20min. Called again at 3am and told another 15min. Its 3:25am and no call bac

 $\boldsymbol{k}$  or arrival. If they couldn't fulfill the order they shouldn't have lead me on. Worst Dominoes se rvice.

## Conclusion

In this brief tutorial, I gave you an idea of how to go about the scraping a websites. Since we don't have a lot of time, I am purposely omitting some important steps. Since most of theses steps are more programming related then text mining.

- \* try except
- \* regular expressions
- \* sleep timer

	_	
Tm [	٦.	
In I		
1		