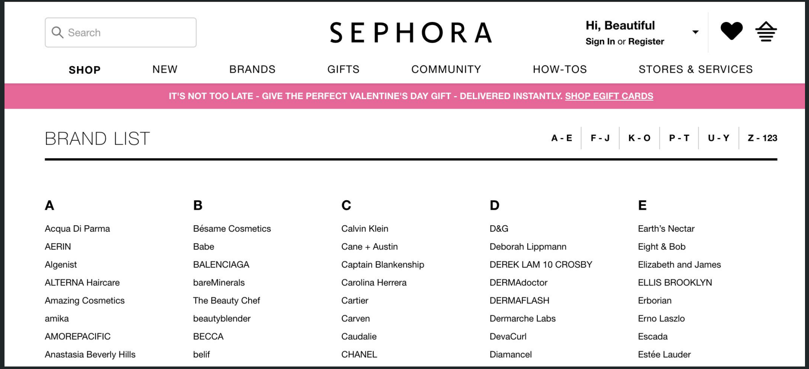
What is my goal for this project?

1. Scrape Sephora
2. Learning how to use Flask and creating an Application which can do simple queries of Sephora reviews data using Flask and sqlite3

How to Scrape Sephora’s website

1. Getting the brand links
2. Getting the product links
3. Getting product details
4. Constructing the API link using information from (#2)
5. Getting data from the API - Reviews and Reviewer details

Getting the brand links



BRAND LIST

A

Acqua Di Parma
AERIN
Algenist
ALTERNA Haircare
Amazing Cosmetics
amika
AMOREPACIFIC
Anastasia Beverly Hills

B

Bésame Cosmetics
Babe
BALENCIAGA
bareMinerals
The Beauty Chef
beautyblender
BECCA
belif

C

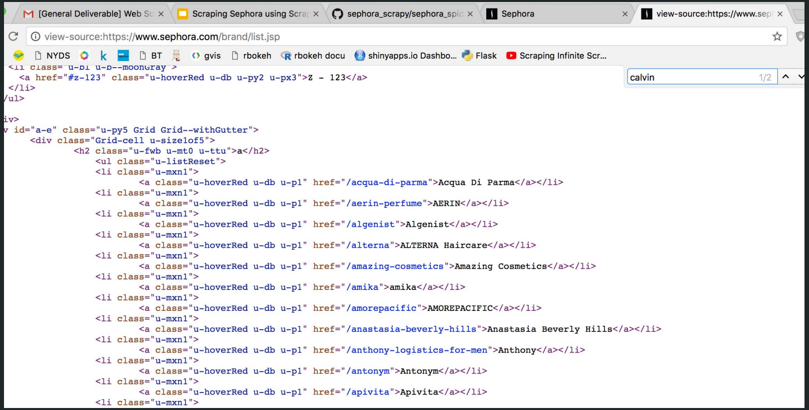
Calvin Klein
Cane + Austin
Captain Blankenship
Carolina Herrera
Cartier
Carven
Caudalie
CHANEL

D

D&G
Deborah Lippmann
DEREK LAM 10 CROSBY
DERMAdoctor
DERMAFLASH
Dermarche Labs
DevaCurl
Diamancel

E

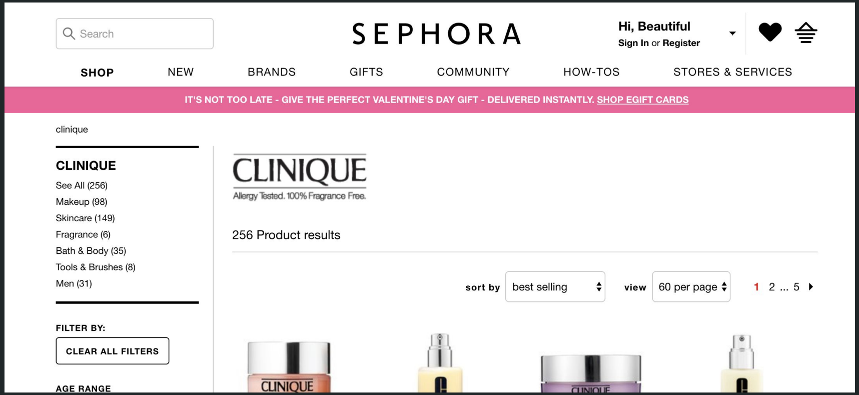
Earth's Nectar
Eight & Bob
Elizabeth and James
ELLIS BROOKLYN
Erborian
Erno Laszlo
Escada
Estée Lauder

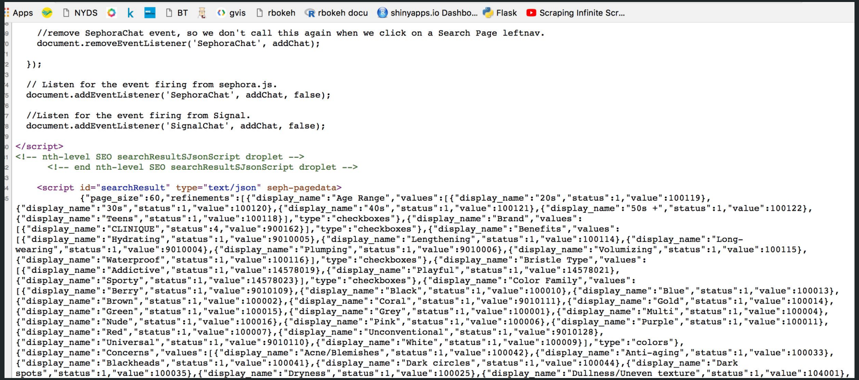

[General Deliverable] Web Sc
Scraping Sephora using Scra
sephora_scrapy/sephora_spid
Sephora
view-source:https://www.sep
view-source:https://www.sephora.com/brand/list.jsp
NYDS
BT
gvis
rbokeh
rbokeh docu
shinyapps.io Dashbo...
Flask
Scraping Infinite Scr...
calvin
1/2

```
<li class="u-pl u-b--moonGray">
  <a href="#z-123" class="u-hoverRed u-db u-py2 u-px3">Z - 123</a>
</li>
/ul>

iv>
v id="a-e" class="u-py5 Grid Grid--withGutter">
     <div class="Grid-cell u-size1of5">
             <h2 class="u-fwb u-mt0 u-ttu">a</h2>
                 <ul class="u-listReset">
                 <li class="u-mxn1">
                         <a class="u-hoverRed u-db u-p1" href="/acqua-di-parma">Acqua Di Parma</a></li>
                 <li class="u-mxn1">
                         <a class="u-hoverRed u-db u-p1" href="/aerin-perfume">AERIN</a></li>
                 <li class="u-mxn1">
                         <a class="u-hoverRed u-db u-p1" href="/algenist">Algenist</a></li>
                 <li class="u-mxn1">
                         <a class="u-hoverRed u-db u-p1" href="/alterna">ALTERNA Haircare</a></li>
                 <li class="u-mxn1">
                         <a class="u-hoverRed u-db u-p1" href="/amazing-cosmetics">Amazing Cosmetics</a></li>
                 <li class="u-mxn1">
                         <a class="u-hoverRed u-db u-p1" href="/amika">amika</a></li>
                 <li class="u-mxn1">
                         <a class="u-hoverRed u-db u-p1" href="/amorepacific">AMOREPACIFIC</a></li>
                 <li class="u-mxn1">
                         <a class="u-hoverRed u-db u-p1" href="/anastasia-beverly-hills">Anastasia Beverly Hills</a></li>
                 <li class="u-mxn1">
                         <a class="u-hoverRed u-db u-p1" href="/anthony-logistics-for-men">Anthony</a></li>
                 <li class="u-mxn1">
                         <a class="u-hoverRed u-db u-p1" href="/antonym">Antonym</a></li>
                 <li class="u-mxn1">
                         <a class="u-hoverRed u-db u-p1" href="/apivita">Apivita</a></li>
                 <li class="u-mxn1">
```

https://www.sephora.com/brand/list.jsp

Getting the product links

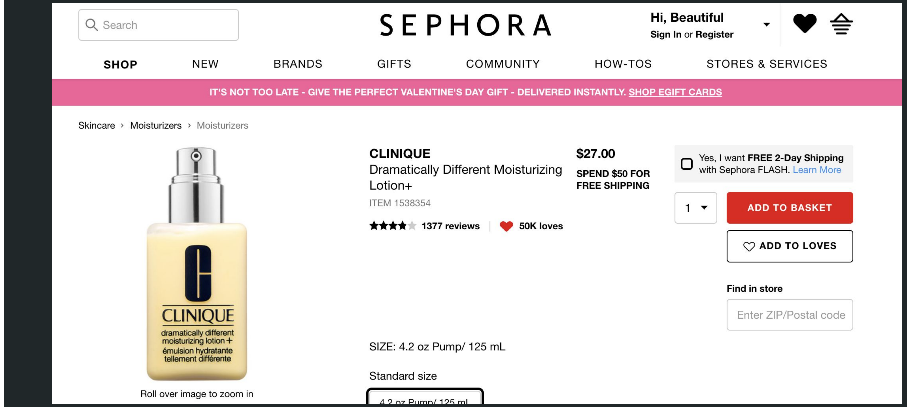




```
    //remove SephoraChat event, so we don't call this again when we click on a Search Page leftnav.
    document.removeEventListener('SephoraChat', addChat);

  });

  // Listen for the event firing from sephora.js.
  document.addEventListener('SephoraChat', addChat, false);

  //Listen for the event firing from Signal.
  document.addEventListener('SignalChat', addChat, false);

</script>
<!-- nth-level SEO searchResultSJsonScript droplet -->
      <!-- end nth-level SEO searchResultSJsonScript droplet -->

    <script id="searchResult" type="text/json" seph-pagedata>
            {"page_size":60,"refinements":[{"display_name":"Age Range","values":[{"display_name":"20s","status":1,"value":100119},
{"display_name":"30s","status":1,"value":100120},{"display_name":"40s","status":1,"value":100121},{"display_name":"50s +","status":1,"value":100122},
{"display_name":"Teens","status":1,"value":100118}],"type":"checkboxes"},{"display_name":"Brand","values":
[{"display_name":"CLINIQUE","status":4,"value":900162}],"type":"checkboxes"},{"display_name":"Benefits","values":
[{"display_name":"Hydrating","status":1,"value":9010005},{"display_name":"Lengthening","status":1,"value":100114},{"display_name":"Long-
wearing","status":1,"value":9010004},{"display_name":"Plumping","status":1,"value":9010006},{"display_name":"Volumizing","status":1,"value":100115},
{"display_name":"Waterproof","status":1,"value":100116}],"type":"checkboxes"},{"display_name":"Bristle Type","values":
[{"display_name":"Addictive","status":1,"value":14578019},{"display_name":"Playful","status":1,"value":14578021},
{"display_name":"Sporty","status":1,"value":14578023}],"type":"checkboxes"},{"display_name":"Color Family","values":
[{"display_name":"Berry","status":1,"value":9010109},{"display_name":"Black","status":1,"value":100010},{"display_name":"Blue","status":1,"value":100013},
{"display_name":"Brown","status":1,"value":100002},{"display_name":"Coral","status":1,"value":9010111},{"display_name":"Gold","status":1,"value":100014},
{"display_name":"Green","status":1,"value":100015},{"display_name":"Grey","status":1,"value":100001},{"display_name":"Multi","status":1,"value":100004},
{"display_name":"Nude","status":1,"value":100016},{"display_name":"Pink","status":1,"value":100006},{"display_name":"Purple","status":1,"value":100011},
{"display_name":"Red","status":1,"value":100007},{"display_name":"Unconventional","status":1,"value":9010128},
{"display_name":"Universal","status":1,"value":9010110},{"display_name":"White","status":1,"value":100009}],"type":"colors"},
{"display_name":"Concerns","values":[{"display_name":"Acne/Blemishes","status":1,"value":100042},{"display_name":"Anti-aging","status":1,"value":100033},
{"display_name":"Blackheads","status":1,"value":100041},{"display_name":"Dark circles","status":1,"value":100044},{"display_name":"Dark
spots","status":1,"value":100035},{"display_name":"Dryness","status":1,"value":100025},{"display_name":"Dullness/Uneven texture","status":1,"value":104001},
```

view-source:https://www.sephora.com/clinique?products=all

Getting the product details



Search

SEPHORA

Hi, Beautiful
Sign In or Register

SHOP NEW BRANDS GIFTS COMMUNITY HOW-TOS STORES & SERVICES

IT'S NOT TOO LATE - GIVE THE PERFECT VALENTINE'S DAY GIFT - DELIVERED INSTANTLY. SHOP EGIFT CARDS

Skincare › Moisturizers › Moisturizers

CLINIQUE
Dramatically Different Moisturizing Lotion+
ITEM 1538354

★★★★☆ 1377 reviews | ♥ 50K loves

$27.00
SPEND $50 FOR FREE SHIPPING

□ Yes, I want FREE 2-Day Shipping with Sephora FLASH. Learn More

1 ▾ ADD TO BASKET

♡ ADD TO LOVES

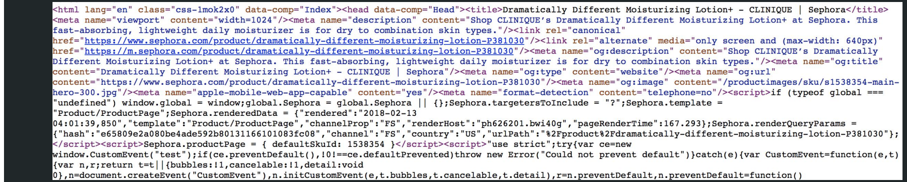
Find in store
Enter ZIP/Postal code

SIZE: 4.2 oz Pump/ 125 mL

Standard size

4.2 oz Pump/ 125 mL

Roll over image to zoom in



```
<html lang="en" class="css-1mok2x0" data-comp="Index"><head data-comp="Head"><title>Dramatically Different Moisturizing Lotion+ - CLINIQUE | Sephora</title>
<meta name="viewport" content="width=1024"/><meta name="description" content="Shop CLINIQUE’s Dramatically Different Moisturizing Lotion+ at Sephora. This
fast-absorbing, lightweight daily moisturizer is for dry to combination skin types."/><link rel="canonical"
href="https://www.sephora.com/product/dramatically-different-moisturizing-lotion-P381030"/><link rel="alternate" media="only screen and (max-width: 640px)"
href="https://m.sephora.com/product/dramatically-different-moisturizing-lotion-P381030"/><meta name="og:description" content="Shop CLINIQUE’s Dramatically
Different Moisturizing Lotion+ at Sephora. This fast-absorbing, lightweight daily moisturizer is for dry to combination skin types."/><meta name="og:title"
content="Dramatically Different Moisturizing Lotion+ - CLINIQUE | Sephora"/><meta name="og:type" content="website"/><meta name="og:url"
content="https://www.sephora.com/product/dramatically-different-moisturizing-lotion-P381030"/><meta name="og:image" content="/productimages/sku/s1538354-main-
hero-300.jpg"/><meta name="apple-mobile-web-app-capable" content="yes"/><meta name="format-detection" content="telephone=no"/><script>if (typeof global ===
"undefined") window.global = window;global.Sephora = global.Sephora || {};Sephora.targetersToInclude = "?";Sephora.template =
"Product/ProductPage";Sephora.renderedData = {"rendered":"2018-02-13
04:01:39,850","template":"Product/ProductPage","channelProp":"FS","renderHost":"ph626201.bwi40g","pageRenderTime":167.293};Sephora.renderQueryParams =
{"hash":"e65809e2a080be4ade592b80131166101083fc08","channel":"FS","country":"US","urlPath":"%2Fproduct%2Fdramatically-different-moisturizing-lotion-P381030"};
</script><script>Sephora.productPage = { defaultSkuId: 1538354 }</script><script>"use strict";try{var ce=new
window.CustomEvent("test");if(ce.preventDefault(),!0!==ce.defaultPrevented)throw new Error("Could not prevent default")}catch(e){var CustomEvent=function(e,t)
{var n,r;return t=t||{bubbles:!1,cancelable:!1,detail:void
0},n=document.createEvent("CustomEvent"),n.initCustomEvent(e,t.bubbles,t.cancelable,t.detail),r=n.preventDefault,n.preventDefault=function()
```

https://www.sephora.com/product/dramatically-different-moisturizing-lotion-P381030?skuId=1538354&icid2=products%20grid:p381030

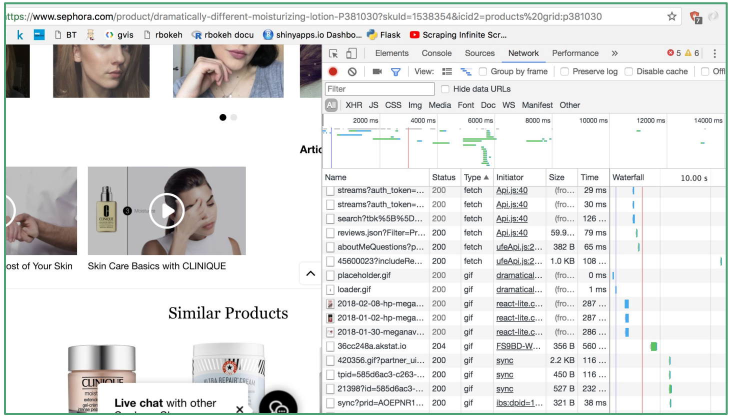
Getting the Reviews - Infinite Scroll Problem

How to solve this?

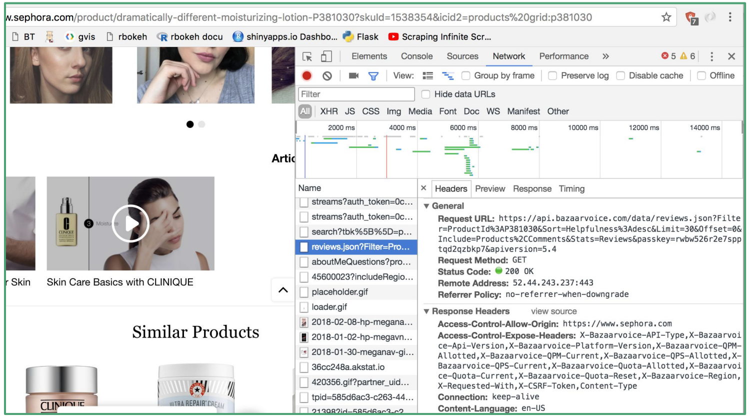
A. Inspect Element

B. Go to the Network, reload the page

C. Notice that when you scroll to make the reviews appear, a json fetch item appears



Name	Status	Type ▲	Initiator	Size	Time	Waterfall 10.00 s
streams?auth_token=…	200	fetch	Api.js:40	(fro…	29 ms	
streams?auth_token=…	200	fetch	Api.js:40	(fro…	30 ms	
search?tbk%5B%5D…	200	fetch	Api.js:40	(fro…	126 …	
reviews.json?Filter=Pr…	200	fetch	Api.js:40	59.9…	79 ms	
aboutMeQuestions?p…	200	fetch	ufeApi.js:2…	382 B	65 ms	
45600023?includeRe…	200	fetch	ufeApi.js:2…	1.0 KB	108 …	
placeholder.gif	200	gif	dramatical…	(fro…	0 ms	
loader.gif	200	gif	dramatical…	(fro…	1 ms	
2018-02-08-hp-mega…	200	gif	react-lite.c…	(fro…	287 …	
2018-01-02-hp-mega…	200	gif	react-lite.c…	(fro…	287 …	
2018-01-30-meganav…	200	gif	react-lite.c…	(fro…	286 …	
36cc248a.akstat.io	204	gif	FS9BD-W…	356 B	560 …	
420356.gif?partner_ui…	200	gif	sync	2.2 KB	116 …	
tpid=585d6ac3-c263-…	200	gif	sync	450 B	116 …	
21398?id=585d6ac3-…	200	gif	sync	527 B	232 …	
sync?prid=AOEPNR1…	200	gif	ibs:dpid=1…	321 B	38 ms	



Name
streams?auth_token=0c…
streams?auth_token=0c…
search?tbk%5B%5D=p…
reviews.json?Filter=Pro…
aboutMeQuestions?pro…
45600023?includeRegio…
placeholder.gif
loader.gif
2018-02-08-hp-megana…
2018-01-02-hp-megavn…
2018-01-30-meganav-gi…
36cc248a.akstat.io
420356.gif?partner_uid…
tpid=585d6ac3-c263-44…

▼ General

Request URL: `https://api.bazaarvoice.com/data/reviews.json?Filter=ProductId%3AP381030&Sort=Helpfulness%3Adesc&Limit=30&Offset=0&Include=Products%2CComments&Stats=Reviews&passkey=rwbw526r2e7spptqd2qzbkp7&apiversion=5.4`

Request Method: GET

Status Code: ● 200 OK

Remote Address: 52.44.243.237:443

Referrer Policy: no-referrer-when-downgrade

▼ Response Headers view source

Access-Control-Allow-Origin: https://www.sephora.com

Access-Control-Expose-Headers: X-Bazaarvoice-API-Type,X-Bazaarvoice-Api-Version,X-Bazaarvoice-Platform-Version,X-Bazaarvoice-QPM-Allotted,X-Bazaarvoice-QPM-Current,X-Bazaarvoice-QPS-Allotted,X-Bazaarvoice-QPS-Current,X-Bazaarvoice-Quota-Allotted,X-Bazaarvoice-Quota-Current,X-Bazaarvoice-Quota-Reset,X-Bazaarvoice-Region,X-Requested-With,X-CSRF-Token,Content-Type

Connection: keep-alive

Content-Language: en-US

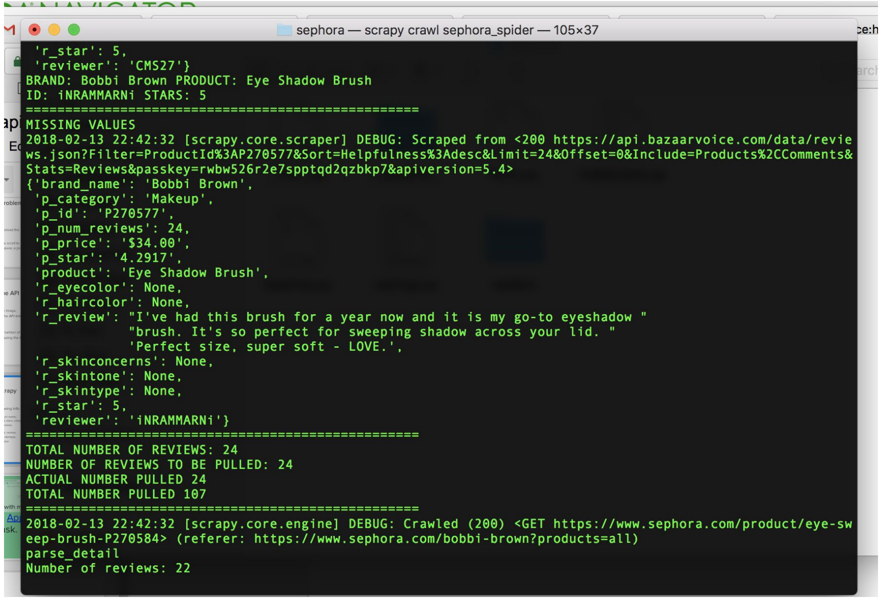
Constructing the API link

- As we can see in the image, there’s a pattern to the API link
- Product ID
- And we can set the number of reviews we can call using the API

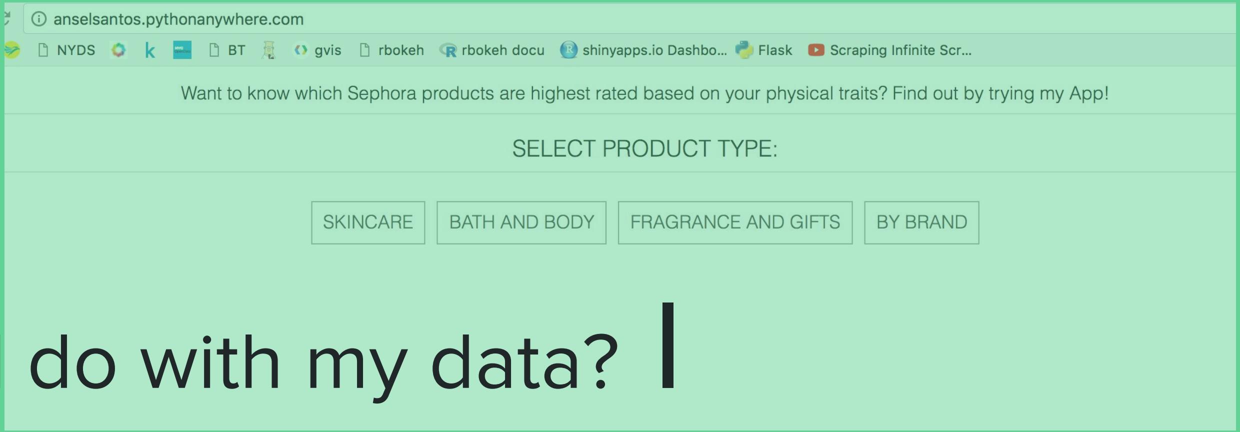
https://api.bazaarvoice.com/data/reviews.json?Filter=ProductId%3A**P381030**&Sort=Helpfulness%3Adesc&**Limit=30**&Offset=0&Include=Products%2CComments&Stats=Reviews&passkey=rwbw526r2e7spptqd2qzbkp7&apiversion=5.4

Running the Scrapy program

- I collected the following info:
  - Brand Name, product name, product ID, average stars, category, price, number of reviews
  - Reviewer nickname, review, reviewer skintone, skintype, haircolor and eyecolor
  - I collected around **80k+** rows of data



```
 'r_star': 5,
 'reviewer': 'CMS27'}
BRAND: Bobbi Brown PRODUCT: Eye Shadow Brush
ID: iNRAMMARNi STARS: 5
==================================================
MISSING VALUES
2018-02-13 22:42:32 [scrapy.core.scraper] DEBUG: Scraped from <200 https://api.bazaarvoice.com/data/revie
ws.json?Filter=ProductId%3AP270577&Sort=Helpfulness%3Adesc&Limit=24&Offset=0&Include=Products%2CComments&
Stats=Reviews&passkey=rwbw526r2e7spptqd2qzbkp7&apiversion=5.4>
{'brand_name': 'Bobbi Brown',
 'p_category': 'Makeup',
 'p_id': 'P270577',
 'p_num_reviews': 24,
 'p_price': '$34.00',
 'p_star': '4.2917',
 'product': 'Eye Shadow Brush',
 'r_eyecolor': None,
 'r_haircolor': None,
 'r_review': "I've had this brush for a year now and it is my go-to eyeshadow "
             "brush. It's so perfect for sweeping shadow across your lid. "
             'Perfect size, super soft - LOVE.',
 'r_skinconcerns': None,
 'r_skintone': None,
 'r_skintype': None,
 'r_star': 5,
 'reviewer': 'iNRAMMARNi'}
==================================================
TOTAL NUMBER OF REVIEWS: 24
NUMBER OF REVIEWS TO BE PULLED: 24
ACTUAL NUMBER PULLED 24
TOTAL NUMBER PULLED 107
==================================================
2018-02-13 22:42:32 [scrapy.core.engine] DEBUG: Crawled (200) <GET https://www.sephora.com/product/eye-sw
eep-brush-P270584> (referer: https://www.sephora.com/bobbi-brown?products=all)
parse_detail
Number of reviews: 22
```

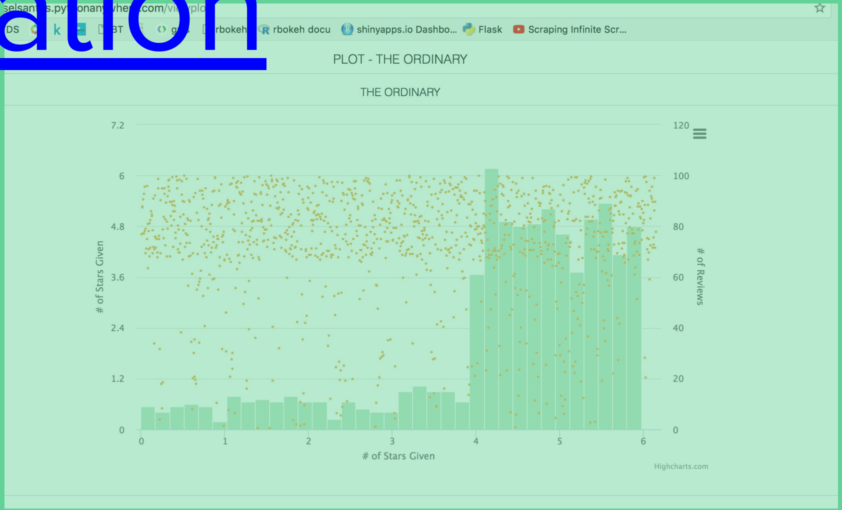


Want to know which Sephora products are highest rated based on your physical traits? Find out by trying my App!

SELECT PRODUCT TYPE:

SKINCARE | BATH AND BODY | FRAGRANCE AND GIFTS | BY BRAND

What did I do with my data? I made an Application using Flask.



PLOT - THE ORDINARY

THE ORDINARY

Conclusion

To sum things up, I was able to show that Sephora’s website can be scraped relatively easily and I learned how to create an Application using Flask and sqlite3, achieving my goals for the project.