

# DA-webscraper

April 26, 2019

## Data Scraping

```
In [1]: #import basic libraries
import pandas as pd
import re
import logging
import scrapy
from scrapy.crawler import CrawlerProcess

In [2]: from scrapy.pipelines.images import ImagesPipeline
from scrapy.exceptions import DropItem

class ImageItem(scrapy.Item):

    #direct link to image file for downloading via ImagePipeline
    image_urls = scrapy.Field()

    #link to specific image page for scraping more stats
    image_links = scrapy.Field()

    #image attributes
    titles = scrapy.Field() #image title
    date_posted = scrapy.Field() #date posted
    hashtags = scrapy.Field() #hashtags

    #image stats
    views = scrapy.Field() #number of views of the current image
    faves = scrapy.Field() #number of faves of the current image
    comments = scrapy.Field() #number of comments of the current image
    downloads = scrapy.Field() #number of downloads of the current image

    #artist details
    artists = scrapy.Field() #artist's name
    artist_urls = scrapy.Field() #link to the artist's account
    artist_deviations = scrapy.Field() #number of deviations (images) posted
    artist_comments = scrapy.Field() #number of total comments received
    artist_page_views = scrapy.Field() #number of total page views received
    artist_scraps = scrapy.Field() #number of scraps (WIPs or archived art)
    artist_watchers = scrapy.Field() #number of watchers (followers)
    artist_critiques = scrapy.Field() #number of critiques given
    artist_forum_posts = scrapy.Field() #number of forum posts made
    artist_faves = scrapy.Field() #number of total faves received
    artist_asl = scrapy.Field() #age, sex, location
    artist_dob = scrapy.Field() #date of birth
    account_age = scrapy.Field() #how old the account is

    # to be filled in by ImagePipeline
    image_paths = scrapy.Field() #location of image in local storage

class MyImagesPipeline(ImagesPipeline):
    """
```

```

Image pipeline for downloading images.
'''
def get_media_requests(self, item, info):
    for image_url in item['image_urls']:
        yield scrapy.Request(image_url)

def item_completed(self, results, item, info):
    image_paths = [x['path'] for ok, x in results if ok]
    if not image_paths:
        raise DropItem("Item contains no images")
    item['image_paths'] = image_paths
    return item

In [3]: class ImageSpider(scrapy.Spider):

    name = 'images'

    start_urls = ['https://www.deviantart.com/popular-all-time/?q=sherlock&offset=0']

    #initialize offset at 0
    offset = 0
    #set offset limit to control the amount of images downloaded
    offset_limit = 16000

    custom_settings = {
        'LOG_LEVEL': logging.INFO,
        'ITEM_PIPELINES': {'__main__.MyImagesPipeline': 1}, #enable image download
        'IMAGES_STORE': 'DA-images-2', #store images in DA-images-2 folder
        'FEED_FORMAT': 'json',
        'FEED_URI': 'image-data-2.json', #store image data in image-data-2.json
        'DOWNLOAD_FAIL_ON_DATALOSS': False, #if image download fails (due to various
issues), don't send error message, just flag it.
        'DOWNLOAD_DELAY': 0.25 #250 ms download delay, with inbuilt scrapy randomization
    }

    def parse(self, response):

        #get page body
        page = response.css('div.page-results span.thumb')

        for img in page:

            #thumbnail link
            thumbnail = img.css('::attr(data-super-img)').get()

            #full link that leads to the individual image post
            img_link = img.css('::attr(href)').get()

            #if there is a thumbnail, aka the post is an image, follow url to scrape
image details
            if thumbnail:
                yield scrapy.Request(img_link, callback = self.parse_image)

            #go to next page
            while self.offset < self.offset_limit:
                self.offset += 24 #DA's natural offset scroll is set at increments of 24
                next_page = f'https://www.deviantart.com/popular-all-
time/?q=sherlock&offset={self.offset}'
                yield scrapy.Request(next_page, callback=self.parse)

    def parse_image(self, response):

        #initialize image item
        image = ImageItem()

        #get image url (for downloading via ImagePipeline)

```

```

        image["image_urls"] = [response.css('div.dev-view-deviation img
::attr(src)').get()]
        #get other image info
        image["image_links"] = response.url
        image['titles'] = response.xpath("//a[@class='title']/text()").extract()[0]
        image['date_posted'] = response.xpath("//div[@class='dev-right-bar-content dev-
metainfo-content dev-metainfo-details']/dl/dd/span/text()").extract()[0]

        #check whether image has hashtags (some don't)
        hashtag = response.xpath("//div[@class='dev-about-tags-cc dev-about-
breadcrumb']/a/text()").extract()
        if hashtag: image['hashtags'] = hashtag

        #get image stats
        stats = response.xpath("//div[@class='dev-right-bar-content dev-metainfo-
content dev-metainfo-stats']/dl/dd/text()").extract()

        #check that stats list only contains numbers (sometimes irregular data falls in)
        stats = [re.sub("\D", "", s) for s in stats]

        #remove any None types from list
        stats = list(filter(None, stats))

        #the responses are ordered in: views, faves, comments, downloads
        #sometimes comments are disabled, sometimes downloads are disabled
        headers = ['views', 'faves', 'comments', 'downloads']

        #if comments/downloads are disabled, they will not be looped over for a given
image
        for i in range(len(stats)):
            image[headers[i]] = stats[i]

        #get artist info
        artist_name = response.xpath("//small[@class='author']/span/a/text()").extract()

        if artist_name:
            image['artists'] =
response.xpath("//small[@class='author']/span/a/text()").extract()[-1]
            image['artist_urls'] =
response.xpath("//small[@class='author']/span/a/@href").extract()[-1]
            request = scrapy.Request(image['artist_urls'], callback=self.parse_artist,
meta={'image': image})
            yield request
        else: #if no artist name (sometimes artists are banned), just yield the image
            image['artists'] = 'Banned'
            yield image

    def parse_artist(self, response):

        #get image item for the higher-level parser
        image = response.meta['image']

        #get artist account stats
        artist_stats = response.xpath("//div[@id='super-secret-
stats']/div/div/div/strong/text()").extract()

        headers = ['artist_deviations', 'artist_comments', 'artist_page_views', 'artist_scr
aps', 'artist_watchers', 'artist_critiques', 'artist_forum_posts', 'artist_faves']

        for i in range(len(artist_stats)):
            image[headers[i]] = artist_stats[i]

        #get account age and membership details
        age_membership = response.xpath("//a[@href='#super-secret-
activity']/div/text()").extract()

        #sometimes the age is wrapped up in a span
        if len(age_membership) == 0:

```

```

        age = response.xpath("//a[@href='#super-secret-activity']/span/div/text()").extract()[0]
        age_membership.append(age)
        image['account_age'] = age_membership[0]

    #get artist personal details
    artist_details = response.xpath("//div[@id='super-secret-why']/div/div/div/dl/dd/text()").extract()
    details = ['artist_asl', 'artist_dob'] #some artists do not share their dobs
    for i in range(len(artist_details)):
        image[details[i]] = artist_details[i]

    return image

process = CrawlerProcess()
process.crawl(ImageSpider)
process.start()

```

2019-04-21 01:03:58 [scrapy.utils.log] INFO: Scrapy 1.5.2 started (bot: scrapybot)  
 2019-04-21 01:03:58 [scrapy.utils.log] INFO: Versions: lxml 4.2.1.0, libxml2 2.9.8, cssselect 1.0.3, parsel 1.5.1, w3lib 1.20.0, Twisted 17.5.0, Python 3.6.8 |Anaconda, Inc.| (default, Dec 29 2018, 19:04:46) - [GCC 4.2.1 Compatible Clang 4.0.1 (tags/RELEASE\_401/final)], pyOpenSSL 18.0.0 (OpenSSL 1.1.1b 26 Feb 2019), cryptography 2.6.1, Platform Darwin-18.2.0-x86\_64-i386-64bit  
 2019-04-21 01:03:58 [scrapy.crawler] INFO: Overridden settings: {'DOWNLOAD\_DELAY': 0.25, 'DOWNLOAD\_FAIL\_ON\_DATALOSS': False, 'FEED\_FORMAT': 'json', 'FEED\_URI': 'image-data-2.json', 'LOG\_LEVEL': 20}  
 2019-04-21 01:03:58 [scrapy.extensions.telnet] INFO: Telnet Password: Ode246c7e41b8d30  
 2019-04-21 01:03:58 [scrapy.middleware] INFO: Enabled extensions:  
 ['scrapy.extensions.corestats.CoreStats',  
 'scrapy.extensions.telnet.TelnetConsole',  
 'scrapy.extensions.memusage.MemoryUsage',  
 'scrapy.extensions.feedexport.FeedExporter',  
 'scrapy.extensions.logstats.LogStats']  
 2019-04-21 01:03:58 [scrapy.middleware] INFO: Enabled downloader middlewares:  
 ['scrapy.downloadermiddlewares.httppauth.HttpAuthMiddleware',  
 'scrapy.downloadermiddlewares.downloadtimeout.DownloadTimeoutMiddleware',  
 'scrapy.downloadermiddlewares.defaultheaders.DefaultHeadersMiddleware',  
 'scrapy.downloadermiddlewares.useragent.UserAgentMiddleware',  
 'scrapy.downloadermiddlewares.retry.RetryMiddleware',  
 'scrapy.downloadermiddlewares.redirect.MetaRefreshMiddleware',  
 'scrapy.downloadermiddlewares.httpcompression.HttpCompressionMiddleware',  
 'scrapy.downloadermiddlewares.redirect.RedirectMiddleware',  
 'scrapy.downloadermiddlewares.cookies.CookiesMiddleware',  
 'scrapy.downloadermiddlewares.httpproxy.HttpProxyMiddleware',  
 'scrapy.downloadermiddlewares.stats.DownloaderStats']  
 2019-04-21 01:03:58 [scrapy.middleware] INFO: Enabled spider middlewares:  
 ['scrapy.spidermiddlewares.httperror.HttpErrorMiddleware',  
 'scrapy.spidermiddlewares.offsite.OffsiteMiddleware',  
 'scrapy.spidermiddlewares.referer.RefererMiddleware',  
 'scrapy.spidermiddlewares.urllength.UrlLengthMiddleware',  
 'scrapy.spidermiddlewares.depth.DepthMiddleware']  
 2019-04-21 01:03:58 [scrapy.middleware] INFO: Enabled item pipelines:  
 ['\_\_main\_\_.MyImagesPipeline']  
 2019-04-21 01:03:58 [scrapy.core.engine] INFO: Spider opened  
 2019-04-21 01:03:58 [scrapy.extensions.logstats] INFO: Crawled 0 pages (at 0 pages/min), scraped 0 items (at 0 items/min)  
 2019-04-21 01:04:58 [scrapy.extensions.logstats] INFO: Crawled 269 pages (at 269 pages/min), scraped 77 items (at 77 items/min)  
 2019-04-21 01:05:58 [scrapy.extensions.logstats] INFO: Crawled 552 pages (at 283

```

pages/min), scraped 167 items (at 90 items/min)
2019-04-21 01:06:58 [scrapy.extensions.logstats] INFO: Crawled 832 pages (at 280
pages/min), scraped 253 items (at 86 items/min)
2019-04-21 01:07:43 [scrapy.pipelines.files] WARNING: File (code: 404): Error
downloading file from <GET https://img00.deviantart.net/1ef1/i/2015/115/1/2/dumbasses_
by_princessblackrabbit-d4q8ggn.jpg> referred in <None>
2019-04-21 01:07:43 [scrapy.core.scrapers] WARNING: Dropped: Item contains no images
{'account_age': 'Deviant for 11 Years',
 'artist_asl': '31/Female/Germany',
 'artist_comments': '58,004 ',
 'artist_critiques': '0 ',
 'artist_deviations': '1,978 ',
 'artist_dob': 'December 12, 1987',
 'artist_faves': '5,555 ',
 'artist_forum_posts': '2 ',
 'artist_page_views': '227,344 ',
 'artist_scraps': '7 ',
 'artist_urls': 'https://www.deviantart.com/princessblackrabbit',
 'artist_watchers': '1,349 ',
 'artists': 'PrincessBlackRabbit',
 'comments': '38',
 'date_posted': 'February 19, 2012',
 'faves': '77',
 'image_links':
'https://www.deviantart.com/princessblackrabbit/art/dumbasses-285929555',
 'image_urls': ['https://img00.deviantart.net/1ef1/i/2015/115/1/2/dumbasses_by_princes
sblackrabbit-d4q8ggn.jpg'],
 'titles': 'dumbasses',
 'views': '1438'}
2019-04-21 01:07:58 [scrapy.extensions.logstats] INFO: Crawled 1125 pages (at 293
pages/min), scraped 348 items (at 95 items/min)
2019-04-21 01:08:58 [scrapy.extensions.logstats] INFO: Crawled 1399 pages (at 274
pages/min), scraped 427 items (at 79 items/min)
2019-04-21 01:09:58 [scrapy.extensions.logstats] INFO: Crawled 1675 pages (at 276
pages/min), scraped 507 items (at 80 items/min)
2019-04-21 01:10:58 [scrapy.extensions.logstats] INFO: Crawled 1946 pages (at 271
pages/min), scraped 583 items (at 76 items/min)
2019-04-21 01:11:58 [scrapy.extensions.logstats] INFO: Crawled 2218 pages (at 272
pages/min), scraped 662 items (at 79 items/min)
2019-04-21 01:12:58 [scrapy.extensions.logstats] INFO: Crawled 2486 pages (at 268
pages/min), scraped 737 items (at 75 items/min)
2019-04-21 01:13:58 [scrapy.extensions.logstats] INFO: Crawled 2757 pages (at 271
pages/min), scraped 810 items (at 73 items/min)
2019-04-21 01:14:58 [scrapy.extensions.logstats] INFO: Crawled 3024 pages (at 267
pages/min), scraped 880 items (at 70 items/min)
2019-04-21 01:15:58 [scrapy.extensions.logstats] INFO: Crawled 3287 pages (at 263
pages/min), scraped 946 items (at 66 items/min)
2019-04-21 01:16:58 [scrapy.extensions.logstats] INFO: Crawled 3553 pages (at 266
pages/min), scraped 1018 items (at 72 items/min)
2019-04-21 01:17:58 [scrapy.extensions.logstats] INFO: Crawled 3810 pages (at 257
pages/min), scraped 1084 items (at 66 items/min)
2019-04-21 01:18:58 [scrapy.extensions.logstats] INFO: Crawled 4064 pages (at 254
pages/min), scraped 1146 items (at 62 items/min)
2019-04-21 01:19:58 [scrapy.extensions.logstats] INFO: Crawled 4317 pages (at 253
pages/min), scraped 1203 items (at 57 items/min)
2019-04-21 01:20:58 [scrapy.extensions.logstats] INFO: Crawled 4577 pages (at 260
pages/min), scraped 1265 items (at 62 items/min)
2019-04-21 01:21:58 [scrapy.extensions.logstats] INFO: Crawled 4830 pages (at 253
pages/min), scraped 1319 items (at 54 items/min)

```

2019-04-21 01:22:58 [scrapy.extensions.logstats] INFO: Crawled 5096 pages (at 266 pages/min), scraped 1390 items (at 71 items/min)  
 2019-04-21 01:23:58 [scrapy.extensions.logstats] INFO: Crawled 5349 pages (at 253 pages/min), scraped 1446 items (at 56 items/min)  
 2019-04-21 01:24:58 [scrapy.extensions.logstats] INFO: Crawled 5604 pages (at 255 pages/min), scraped 1503 items (at 57 items/min)  
 2019-04-21 01:25:58 [scrapy.extensions.logstats] INFO: Crawled 5861 pages (at 257 pages/min), scraped 1564 items (at 61 items/min)  
 2019-04-21 01:26:58 [scrapy.extensions.logstats] INFO: Crawled 6118 pages (at 257 pages/min), scraped 1623 items (at 59 items/min)  
 2019-04-21 01:27:58 [scrapy.extensions.logstats] INFO: Crawled 6370 pages (at 252 pages/min), scraped 1680 items (at 57 items/min)  
 2019-04-21 01:28:58 [scrapy.extensions.logstats] INFO: Crawled 6624 pages (at 254 pages/min), scraped 1738 items (at 58 items/min)  
 2019-04-21 01:29:58 [scrapy.extensions.logstats] INFO: Crawled 6868 pages (at 244 pages/min), scraped 1789 items (at 51 items/min)  
 2019-04-21 01:30:58 [scrapy.extensions.logstats] INFO: Crawled 7115 pages (at 247 pages/min), scraped 1842 items (at 53 items/min)  
 2019-04-21 01:31:58 [scrapy.extensions.logstats] INFO: Crawled 7361 pages (at 246 pages/min), scraped 1894 items (at 52 items/min)  
 2019-04-21 01:32:58 [scrapy.extensions.logstats] INFO: Crawled 7611 pages (at 250 pages/min), scraped 1949 items (at 55 items/min)  
 2019-04-21 01:33:58 [scrapy.extensions.logstats] INFO: Crawled 7870 pages (at 259 pages/min), scraped 2010 items (at 61 items/min)  
 2019-04-21 01:34:58 [scrapy.extensions.logstats] INFO: Crawled 8124 pages (at 254 pages/min), scraped 2068 items (at 58 items/min)  
 2019-04-21 01:35:58 [scrapy.extensions.logstats] INFO: Crawled 8370 pages (at 246 pages/min), scraped 2121 items (at 53 items/min)  
 2019-04-21 01:36:58 [scrapy.extensions.logstats] INFO: Crawled 8615 pages (at 245 pages/min), scraped 2175 items (at 54 items/min)  
 2019-04-21 01:37:58 [scrapy.extensions.logstats] INFO: Crawled 8865 pages (at 250 pages/min), scraped 2230 items (at 55 items/min)  
 2019-04-21 01:38:58 [scrapy.extensions.logstats] INFO: Crawled 9111 pages (at 246 pages/min), scraped 2285 items (at 55 items/min)  
 2019-04-21 01:39:58 [scrapy.extensions.logstats] INFO: Crawled 9370 pages (at 259 pages/min), scraped 2347 items (at 62 items/min)  
 2019-04-21 01:40:58 [scrapy.extensions.logstats] INFO: Crawled 9622 pages (at 252 pages/min), scraped 2406 items (at 59 items/min)  
 2019-04-21 01:41:58 [scrapy.extensions.logstats] INFO: Crawled 9867 pages (at 245 pages/min), scraped 2453 items (at 47 items/min)  
 2019-04-21 01:42:58 [scrapy.extensions.logstats] INFO: Crawled 10117 pages (at 250 pages/min), scraped 2514 items (at 61 items/min)  
 2019-04-21 01:43:58 [scrapy.extensions.logstats] INFO: Crawled 10362 pages (at 245 pages/min), scraped 2567 items (at 53 items/min)  
 2019-04-21 01:44:58 [scrapy.extensions.logstats] INFO: Crawled 10626 pages (at 264 pages/min), scraped 2635 items (at 68 items/min)  
 2019-04-21 01:45:58 [scrapy.extensions.logstats] INFO: Crawled 10883 pages (at 257 pages/min), scraped 2696 items (at 61 items/min)  
 2019-04-21 01:46:58 [scrapy.extensions.logstats] INFO: Crawled 11130 pages (at 247 pages/min), scraped 2750 items (at 54 items/min)  
 2019-04-21 01:47:58 [scrapy.extensions.logstats] INFO: Crawled 11382 pages (at 252 pages/min), scraped 2808 items (at 58 items/min)  
 2019-04-21 01:48:58 [scrapy.extensions.logstats] INFO: Crawled 11634 pages (at 252 pages/min), scraped 2866 items (at 58 items/min)  
 2019-04-21 01:49:58 [scrapy.extensions.logstats] INFO: Crawled 11895 pages (at 261 pages/min), scraped 2934 items (at 68 items/min)  
 2019-04-21 01:50:58 [scrapy.extensions.logstats] INFO: Crawled 12145 pages (at 250 pages/min), scraped 2988 items (at 54 items/min)  
 2019-04-21 01:51:58 [scrapy.extensions.logstats] INFO: Crawled 12401 pages (at 256

```

pages/min), scraped 3049 items (at 61 items/min)
2019-04-21 01:52:58 [scrapy.extensions.logstats] INFO: Crawled 12668 pages (at 267
pages/min), scraped 3121 items (at 72 items/min)
2019-04-21 01:53:58 [scrapy.extensions.logstats] INFO: Crawled 12916 pages (at 248
pages/min), scraped 3174 items (at 53 items/min)
2019-04-21 01:54:58 [scrapy.extensions.logstats] INFO: Crawled 13173 pages (at 257
pages/min), scraped 3237 items (at 63 items/min)
2019-04-21 01:55:58 [scrapy.extensions.logstats] INFO: Crawled 13420 pages (at 247
pages/min), scraped 3292 items (at 55 items/min)
2019-04-21 01:56:58 [scrapy.extensions.logstats] INFO: Crawled 13682 pages (at 262
pages/min), scraped 3356 items (at 64 items/min)
2019-04-21 01:57:58 [scrapy.extensions.logstats] INFO: Crawled 13942 pages (at 260
pages/min), scraped 3423 items (at 67 items/min)
2019-04-21 01:58:58 [scrapy.extensions.logstats] INFO: Crawled 14191 pages (at 249
pages/min), scraped 3480 items (at 57 items/min)
2019-04-21 01:59:58 [scrapy.extensions.logstats] INFO: Crawled 14453 pages (at 262
pages/min), scraped 3547 items (at 67 items/min)
2019-04-21 02:00:58 [scrapy.extensions.logstats] INFO: Crawled 14707 pages (at 254
pages/min), scraped 3605 items (at 58 items/min)
2019-04-21 02:01:58 [scrapy.extensions.logstats] INFO: Crawled 14963 pages (at 256
pages/min), scraped 3663 items (at 58 items/min)
2019-04-21 02:02:58 [scrapy.extensions.logstats] INFO: Crawled 15211 pages (at 248
pages/min), scraped 3720 items (at 57 items/min)
2019-04-21 02:03:58 [scrapy.extensions.logstats] INFO: Crawled 15461 pages (at 250
pages/min), scraped 3771 items (at 51 items/min)
2019-04-21 02:04:58 [scrapy.extensions.logstats] INFO: Crawled 15710 pages (at 249
pages/min), scraped 3828 items (at 57 items/min)
2019-04-21 02:05:58 [scrapy.extensions.logstats] INFO: Crawled 15966 pages (at 256
pages/min), scraped 3889 items (at 61 items/min)
2019-04-21 02:06:58 [scrapy.extensions.logstats] INFO: Crawled 16218 pages (at 252
pages/min), scraped 3948 items (at 59 items/min)
2019-04-21 02:07:58 [scrapy.extensions.logstats] INFO: Crawled 16472 pages (at 254
pages/min), scraped 4007 items (at 59 items/min)
2019-04-21 02:08:58 [scrapy.extensions.logstats] INFO: Crawled 16721 pages (at 249
pages/min), scraped 4061 items (at 54 items/min)
2019-04-21 02:09:58 [scrapy.extensions.logstats] INFO: Crawled 16968 pages (at 247
pages/min), scraped 4113 items (at 52 items/min)
2019-04-21 02:10:58 [scrapy.extensions.logstats] INFO: Crawled 17217 pages (at 249
pages/min), scraped 4167 items (at 54 items/min)
2019-04-21 02:11:58 [scrapy.extensions.logstats] INFO: Crawled 17473 pages (at 256
pages/min), scraped 4228 items (at 61 items/min)
2019-04-21 02:12:58 [scrapy.extensions.logstats] INFO: Crawled 17721 pages (at 248
pages/min), scraped 4280 items (at 52 items/min)
2019-04-21 02:13:58 [scrapy.extensions.logstats] INFO: Crawled 17964 pages (at 243
pages/min), scraped 4331 items (at 51 items/min)
2019-04-21 02:14:58 [scrapy.extensions.logstats] INFO: Crawled 18203 pages (at 239
pages/min), scraped 4375 items (at 44 items/min)
2019-04-21 02:15:58 [scrapy.extensions.logstats] INFO: Crawled 18445 pages (at 242
pages/min), scraped 4422 items (at 47 items/min)
2019-04-21 02:16:58 [scrapy.extensions.logstats] INFO: Crawled 18696 pages (at 251
pages/min), scraped 4473 items (at 51 items/min)
2019-04-21 02:17:07 [scrapy.core.scrapers] ERROR: Spider error processing <GET
https://www.deviantart.com/sherlock-holmes> (referer:
https://www.deviantart.com/sherlock-holmes/art/Sherlock-Holmes-Club-ID-8867259)
Traceback (most recent call last):
  File "/anaconda3/lib/python3.6/site-packages/twisted/internet/defer.py", line 653,
in _runCallbacks
    current.result = callback(current.result, *args, **kw)
  File "<ipython-input-3-b0621a444cf4>", line 110, in parse_artist

```

```

age = response.xpath("//a[@href='#super-secret-activity']/span/div/text()").extract()[0]
IndexError: list index out of range
2019-04-21 02:17:58 [scrapy.extensions.logstats] INFO: Crawled 18948 pages (at 252
pages/min), scraped 4524 items (at 51 items/min)
2019-04-21 02:18:58 [scrapy.extensions.logstats] INFO: Crawled 19202 pages (at 254
pages/min), scraped 4578 items (at 54 items/min)
2019-04-21 02:19:58 [scrapy.extensions.logstats] INFO: Crawled 19436 pages (at 234
pages/min), scraped 4621 items (at 43 items/min)
2019-04-21 02:20:58 [scrapy.extensions.logstats] INFO: Crawled 19693 pages (at 257
pages/min), scraped 4679 items (at 58 items/min)
2019-04-21 02:21:58 [scrapy.extensions.logstats] INFO: Crawled 19938 pages (at 245
pages/min), scraped 4728 items (at 49 items/min)
2019-04-21 02:22:58 [scrapy.extensions.logstats] INFO: Crawled 20180 pages (at 242
pages/min), scraped 4773 items (at 45 items/min)
2019-04-21 02:23:58 [scrapy.extensions.logstats] INFO: Crawled 20423 pages (at 243
pages/min), scraped 4816 items (at 43 items/min)
2019-04-21 02:24:58 [scrapy.extensions.logstats] INFO: Crawled 20658 pages (at 235
pages/min), scraped 4858 items (at 42 items/min)
2019-04-21 02:25:58 [scrapy.extensions.logstats] INFO: Crawled 20893 pages (at 235
pages/min), scraped 4899 items (at 41 items/min)
2019-04-21 02:26:58 [scrapy.extensions.logstats] INFO: Crawled 21136 pages (at 243
pages/min), scraped 4942 items (at 43 items/min)
2019-04-21 02:27:58 [scrapy.extensions.logstats] INFO: Crawled 21382 pages (at 246
pages/min), scraped 4993 items (at 51 items/min)
2019-04-21 02:28:58 [scrapy.extensions.logstats] INFO: Crawled 21618 pages (at 236
pages/min), scraped 5033 items (at 40 items/min)
2019-04-21 02:29:58 [scrapy.extensions.logstats] INFO: Crawled 21867 pages (at 249
pages/min), scraped 5082 items (at 49 items/min)
2019-04-21 02:30:58 [scrapy.extensions.logstats] INFO: Crawled 22106 pages (at 239
pages/min), scraped 5126 items (at 44 items/min)
2019-04-21 02:31:58 [scrapy.extensions.logstats] INFO: Crawled 22342 pages (at 236
pages/min), scraped 5166 items (at 40 items/min)
2019-04-21 02:32:58 [scrapy.extensions.logstats] INFO: Crawled 22577 pages (at 235
pages/min), scraped 5210 items (at 44 items/min)
2019-04-21 02:33:58 [scrapy.extensions.logstats] INFO: Crawled 22807 pages (at 230
pages/min), scraped 5252 items (at 42 items/min)
2019-04-21 02:34:58 [scrapy.extensions.logstats] INFO: Crawled 23037 pages (at 230
pages/min), scraped 5289 items (at 37 items/min)
2019-04-21 02:35:58 [scrapy.extensions.logstats] INFO: Crawled 23268 pages (at 231
pages/min), scraped 5330 items (at 41 items/min)
2019-04-21 02:36:58 [scrapy.extensions.logstats] INFO: Crawled 23508 pages (at 240
pages/min), scraped 5372 items (at 42 items/min)
2019-04-21 02:37:58 [scrapy.extensions.logstats] INFO: Crawled 23736 pages (at 228
pages/min), scraped 5408 items (at 36 items/min)
2019-04-21 02:38:58 [scrapy.extensions.logstats] INFO: Crawled 23959 pages (at 223
pages/min), scraped 5440 items (at 32 items/min)
2019-04-21 02:39:51 [scrapy.pipelines.files] ERROR: File (unknown-error): Error
processing file from <GET https://images-wixmp-
ed30a86b8c4ca887773594c2.wixmp.com/f/0dc70551-b190-44e8-9b59-5e9378a30049/d2y7tid-cc06
bf13-115d-41d9-b9ec-00c33c8cab98.jpg?token=eyJ0eXAiOiJKV1QiLCJhbGciOiJIUzI1NiJ9.eyJzdW
IiOiJlcm46YXBwOjdlMGQxODg5ODIyNjQzNzNhNWYwZDQxNWVhMGQyNmUwIiwiaXNzIjojdXJuOmFwcDo3ZTBk
MTg4OTgyMjYOMzcZYTVMGQOMTVlYTBkMjZlMCIsIm9iaaiI6W1t7InBhdGgiOiJcL2ZcLzBkYzZcNTUxLWlXOT
AtNDRlOC05YjU5LTVlOTM3OGEzMDA0OVVwZDZlJ5N3R3ZC1jYzA2YmYxMy0xMTVvLTQxZDktYjllYy0wMGZmM2M4
Y2FiOTguanBnInIdXSwiYXVkJpbInVybjpzZXJ2aWNlOmZpbGUuZG93bmVvYXQiXXO.3RRjMd6_q6cBQZh-
YJCavU7n72FfQbHmg1KuoEJpdk8> referred in <None>
Traceback (most recent call last):
  File "/anaconda3/lib/python3.6/site-packages/twisted/internet/defer.py", line 1386,
in _inlineCallbacks

```



```

        result = g.send(result)
    File "/anaconda3/lib/python3.6/site-packages/scrapy/core/downloader/middleware.py",
line 43, in process_request
        defer.returnValue((yield download_func(request=request,spider=spider)))
    File "/anaconda3/lib/python3.6/site-packages/twisted/internet/defer.py", line 1363,
in returnValue
        raise _DefGen_Return(val)
twisted.internet.defer._DefGen_Return: <200 https://images-wixmp-
ed30a86b8c4ca887773594c2.wixmp.com/f/0dc70551-b190-44e8-9b59-5e9378a30049/d2y7tid-cc06
bf13-115d-41d9-b9ec-00c33c8cab98.jpg?token=eyJ0eXAiOiJKV1QiLCJhbGciOiJIUzI1NiJ9.eyJzdW
IiOiJlcm46YXBwOjdlMGQxODg5ODIyNjQzNzNhNWYwZDQxNWVhMGQyNmUwIiwiaXNzIjoIdXJuOmFwcDo3ZTBk
MTg4OTgyMjYOMzczYTVMGQOMTVlYTlkMjZlMCIscIm9iaaiI6W1t7InBhdGgiOiJcL2ZcLzBkYzZcwNTUxLWlXOT
AtNDRlOC05YjU5LTVlOTM3OGEzMDA0OVwvZDJSN3RpZC1jYzA2YmYxMy0xMTVhLTQxZDktYjllYy0wMGZmM2M4
Y2FiOTguanBnIn1dXSwiYXVkJpbInVybjpzZXJ2aWNlOmZpbGUuZG93bmxxvYWQixXO.3RRjMd6_q6cBQZh-
YJCavU7n72FfQbHmg1KuoeJpdK8>

```

During handling of the above exception, another exception occurred:

Traceback (most recent call last):

```

    File "/anaconda3/lib/python3.6/site-packages/scrapy/pipelines/files.py", line 401,
in media_downloaded
        checksum = self.file_downloaded(response, request, info)
    File "/anaconda3/lib/python3.6/site-packages/scrapy/pipelines/images.py", line 101,
in file_downloaded
        return self.image_downloaded(response, request, info)
    File "/anaconda3/lib/python3.6/site-packages/scrapy/pipelines/images.py", line 105,
in image_downloaded
        for path, image, buf in self.get_images(response, request, info):
    File "/anaconda3/lib/python3.6/site-packages/scrapy/pipelines/images.py", line 125,
in get_images
        image, buf = self.convert_image(orig_image)
    File "/anaconda3/lib/python3.6/site-packages/scrapy/pipelines/images.py", line 151,
in convert_image
        image.save(buf, 'JPEG')
    File "/anaconda3/lib/python3.6/site-packages/PIL/Image.py", line 1899, in save
        self.load()
    File "/anaconda3/lib/python3.6/site-packages/PIL/ImageFile.py", line 228, in load
        "(%d bytes not processed)" % len(b))
OSError: image file is truncated (50 bytes not processed)
2019-04-21 02:39:51 [scrapy.core.scrapers] WARNING: Dropped: Item contains no images
{'account_age': 'Deviant for 15 Years',
 'artist_asl': 'Philippines',
 'artist_comments': '731 ',
 'artist_critiques': '0 ',
 'artist_deviations': '75 ',
 'artist_faves': '100 ',
 'artist_forum_posts': '1 ',
 'artist_page_views': '38,875 ',
 'artist_scraps': '0 ',
 'artist_urls': 'https://www.deviantart.com/melrosetormhaven',
 'artist_watchers': '454 ',
 'artists': 'melrosetormhaven',
 'comments': '678',
 'date_posted': 'September 7, 2010',
 'downloads': '818',
 'faves': '4856',
 'image_links': 'https://www.deviantart.com/melrosetormhaven/art/BBC-Sherlock-and-
John-178404133',
 'image_urls': ['https://images-wixmp-

```

```

ed30a86b8c4ca887773594c2.wixmp.com/f/0dc70551-b190-44e8-9b59-5e9378a30049/d2y7tid-cc06
bf13-115d-41d9-b9ec-00c33c8cab98.jpg?token=eyJ0eXAiOiJKV1QiLCJhbGciOiJIUzI1NiJ9.eyJzdW
IiOiJ1cm46YXBwOjdlMGQxODg5ODIyNjQzNzNhNWYwZDQxNWVhMGQyNmUwIiwiaXNzIjoibXJ0eFwDo3ZTBk
MTg4OTgyMjYOMzcZYTVMGQOMTVlYTBkMjZlMCIsIm9iaaiI6W1t7InBhdGgiOiJcL2ZcLzBkYzZcwNTUxLWIxOT
AtNDRlOC05YjU5LTVlOTM3OGEzMDA0OVVwZDZJ5N3RpZC1jYzA2YmYxMy0xMTVhLTQxZDktYjllYy0wMGMzM2M4
Y2FiOTguanBnIn1dXSwiYXVkiJpbInVybipzZXJ2aWN1OmZpbGUuZG93bmxxvYwQiXX0.3RRjMd6_q6cBQZh-
YJCavU7n72FfQbHmg1KuoeJpdk8'],
'titles': 'BBC Sherlock and John',
'views': '58451'}
2019-04-21 02:39:58 [scrapy.extensions.logstats] INFO: Crawled 24201 pages (at 242
pages/min), scraped 5483 items (at 43 items/min)
2019-04-21 02:40:58 [scrapy.extensions.logstats] INFO: Crawled 24440 pages (at 239
pages/min), scraped 5529 items (at 46 items/min)
2019-04-21 02:41:58 [scrapy.extensions.logstats] INFO: Crawled 24668 pages (at 228
pages/min), scraped 5562 items (at 33 items/min)
2019-04-21 02:42:58 [scrapy.extensions.logstats] INFO: Crawled 24903 pages (at 235
pages/min), scraped 5602 items (at 40 items/min)
2019-04-21 02:43:58 [scrapy.extensions.logstats] INFO: Crawled 25127 pages (at 224
pages/min), scraped 5634 items (at 32 items/min)
2019-04-21 02:44:01 [scrapy.core.engine] INFO: Closing spider (finished)
2019-04-21 02:44:01 [scrapy.extensions.feedexport] INFO: Stored json feed (5638 items)
in: image-data-2.json
2019-04-21 02:44:01 [scrapy.statscollectors] INFO: Dumping Scrapy stats:
{'downloader/request_bytes': 17258826,
 'downloader/request_count': 25136,
 'downloader/request_method_count/GET': 25136,
 'downloader/response_bytes': 2029325194,
 'downloader/response_count': 25136,
 'downloader/response_status_count/200': 25134,
 'downloader/response_status_count/404': 1,
 'downloader/response_status_count/502': 1,
 'dupefilter/filtered': 7563,
 'file_count': 5639,
 'file_status_count/downloaded': 5639,
 'finish_reason': 'finished',
 'finish_time': datetime.datetime(2019, 4, 21, 5, 44, 1, 103890),
 'item_dropped_count': 2,
 'item_dropped_reasons_count/DropItem': 2,
 'item_scraped_count': 5638,
 'log_count/ERROR': 2,
 'log_count/INFO': 109,
 'log_count/WARNING': 3,
 'memusage/max': 234598400,
 'memusage/startup': 104931328,
 'request_depth_max': 3,
 'response_received_count': 25135,
 'retry/count': 1,
 'retry/reason_count/502 Bad Gateway': 1,
 'scheduler/dequeued': 19495,
 'scheduler/dequeued/memory': 19495,
 'scheduler/enqueued': 19495,
 'scheduler/enqueued/memory': 19495,
 'spider_exceptions/IndexError': 1,
 'start_time': datetime.datetime(2019, 4, 21, 4, 3, 58, 630997)}
2019-04-21 02:44:01 [scrapy.core.engine] INFO: Spider closed (finished)

```

```

In [4]: df = pd.read_json('image-data-2.json')
df

```

```

Out[4]:
      account_age      artist_asl artist_comments \
0      Deviant for 5 Years      Female/France      468
1      Deviant for 8 Years      Female/Russia      169
2      Deviant for 6 Years      Female/Japan      23
3      Deviant for 13 Years      Female/France      322
4      Deviant for 7 Years      Female/South Korea      436
5      Deviant for 10 Years      United Kingdom      17
6      Deviant for 7 Years      28/Female/Russia      8,264
7      Deviant for 12 Years      France      376
8      Deviant for 6 Years      Female/United Kingdom      128
9      Deviant for 7 Years      Female/United States      32
10     Deviant for 9 Years      Female/Belgium      780
11     Deviant for 15 Years      Female/Australia      5,301
12     Deviant for 9 Years      29/Female/South Africa      52
13     Deviant for 15 Years      Female/Canada      4,021
14     Deviant for 12 Years      Latvia      3,400
15     Deviant for 6 Years      Australia      15,748
16     Deviant for 7 Years      Female/United Kingdom      164
17     Deviant for 8 Years      Female/Unknown      950
18     Deviant for 9 Years      Female/Poland      1,551
19     Deviant for 8 Years      Female/Germany      143
20     Deviant for 7 Years      24/Female/United States      5,300
21     Deviant for 8 Years      Spain      28,352
22     Deviant for 14 Years      Female/United States      3,784
23     Deviant for 11 Years      Venezuela      6,558
24     Deviant for 11 Years      Female/Canada      134
25     Deviant for 9 Years      30/Female/Russia      173
26     Deviant for 8 Years      Taiwan      541
27     Deviant for 13 Years      Female/United Kingdom      784
28     Deviant for 7 Years      Female/Moldova      376
29     Deviant for 11 Years      28/Female/Ukraine      685
...     ...     ...     ...
5608    Deviant for 7 Years      Female/Spain      329
5609    Deviant for 6 Years      Female/Canada      1,983
5610    Deviant for 12 Years      United States      4,634
5611    Deviant for 6 Years      21/Male/India      8,613
5612    Deviant for 8 Years      Male/United States      1,477
5613    Deviant for 11 Years      Female/Russia      775
5614    Deviant for 7 Years      Female/United States      12,004
5615    Deviant for 6 Years      Male/India      676
5616    Deviant for 11 Years      Male/Malaysia      117
5617    Deviant for 11 Years      32/Male/Russia      1,251
5618    Deviant for 6 Years      Male/Philippines      149
5619    Deviant for 12 Years      France      473
5620    Deviant for 9 Years      36/Male/Spain      2,667
5621    Deviant for 5 Years      United Kingdom      107
5622    Deviant for 9 Years      Female/United States      2,253
5623    Deviant for 8 Years      Female/United States      1,688

```

5624	Deviant for 10 Years	Female/France	2,126
5625	Deviant for 11 Years	Male/United Kingdom	7,471
5626	Deviant for 5 Years	20/Female/United Kingdom	519
5627	Deviant for 14 Years	Female/Germany	3,608
5628	Deviant for 10 Years	33/Female/United Kingdom	196
5629	Deviant for 11 Years	Female/Finland	12,610
5630	Deviant for 8 Years	Female/United Kingdom	2,180
5631	Deviant for 9 Years	25/Female/Canada	5,881
5632	Deviant for 9 Years	Female/Japan	10,445
5633	Deviant for 12 Years	Female/United States	1,586
5634	Deviant for 9 Years	Female/Russia	623
5635	Deviant for 13 Years	Female/Spain	5,973
5636	Deviant for 14 Years	Male/United Kingdom	89,573
5637	Deviant for 6 Years	Female/Russia	3,925

	artist_critiques	artist_deviations	artist_dob	artist_faves \
0	0.0	183	March 5	1,830
1	0.0	36	March 5	216
2	0.0	28	September 9	53
3	0.0	55	March 29	396
4	0.0	73	September 17	37
5	0.0	77	NaN	6
6	0.0	150	April 21, 1990	3,760
7	0.0	13	NaN	114
8	0.0	97	NaN	0
9	0.0	8	March 27	30
10	0.0	43	NaN	45
11	2.0	1,010	December 6	505
12	0.0	50	January 25, 1990	356
13	0.0	184	NaN	3,066
14	0.0	415	NaN	237
15	0.0	210	September 15	1,256
16	0.0	18	NaN	92
17	0.0	200	February 14	357
18	0.0	136	NaN	456
19	0.0	68	December 22	215
20	0.0	206	February 8, 1995	1,607
21	0.0	398	January 27	3,821
22	0.0	239	February 11	788
23	0.0	758	February 7	8,778
24	0.0	27	October 15	4,098
25	0.0	73	February 26, 1989	183
26	0.0	173	September 17	1,182
27	0.0	237	September 19	652
28	0.0	140	June 21	417
29	0.0	30	September 2, 1990	1,599
...	...	...	...	...
5608	0.0	41	April 3	372

5609	0.0	404	January 28	1,971
5610	0.0	293	June 25	378
5611	2.0	156	May 16, 1997	2,367
5612	0.0	122	September 16	283
5613	0.0	40	October 6	429
5614	1.0	197	September 10	8,137
5615	0.0	128	September 18	114
5616	0.0	65	October 23	47
5617	0.0	173	May 23, 1986	914
5618	0.0	135	February 12	121
5619	0.0	61	NaN	217
5620	3.0	215	April 15, 1983	214
5621	0.0	28	NaN	2
5622	0.0	65	August 21	564
5623	4.0	123	NaN	1,847
5624	0.0	135	December 16	229
5625	0.0	639	December 8	6,183
5626	0.0	82	September 12, 1998	471
5627	0.0	170	NaN	272
5628	0.0	259	June 3, 1985	28
5629	0.0	570	July 18	2,718
5630	0.0	173	January 29	1,424
5631	0.0	116	May 17, 1993	3,196
5632	2.0	292	May 24	3,435
5633	0.0	417	NaN	145
5634	0.0	134	NaN	561
5635	0.0	287	February 8	1,316
5636	0.0	80	December 6	70,833
5637	0.0	218	July 1	59

	artist_forum_posts	artist_page_views	artist_scraps	...	comments \
0	0	10,634	0	...	22.0
1	0	11,377	2	...	45.0
2	0	9,907	0	...	102.0
3	0	36,431	3	...	35.0
4	0	45,119	0	...	8.0
5	0	18,474	0	...	21.0
6	0	68,208	0	...	46.0
7	0	6,566	0	...	5.0
8	0	29,442	0	...	22.0
9	0	2,958	2	...	38.0
10	0	29,133	0	...	23.0
11	45	4,256,227	184	...	66.0
12	0	4,660	0	...	34.0
13	4	86,094	24	...	73.0
14	0	84,301	0	...	81.0
15	0	236,955	0	...	18.0
16	0	7,223	9	...	129.0

17	0	30,365	0	...	252.0
18	0	37,202	0	...	61.0
19	1	5,863	50	...	26.0
20	0	357,800	2	...	93.0
21	1	307,901	11	...	131.0
22	2	151,427	8	...	45.0
23	7	179,005	479	...	8.0
24	0	5,447	0	...	25.0
25	0	86,717	0	...	20.0
26	0	31,372	0	...	16.0
27	10	42,535	0	...	31.0
28	0	23,300	0	...	15.0
29	0	75,854	0	...	30.0
...	...	...	...	...	...
5608	0	6,738	0	...	24.0
5609	0	278,550	9	...	30.0
5610	0	31,471	3	...	18.0
5611	0	24,715	1	...	64.0
5612	0	13,515	9	...	77.0
5613	0	32,967	0	...	103.0
5614	364	42,202	17	...	39.0
5615	0	9,881	0	...	16.0
5616	0	14,384	0	...	15.0
5617	0	98,343	9	...	20.0
5618	0	18,515	0	...	10.0
5619	0	393,967	0	...	107.0
5620	1	67,880	0	...	32.0
5621	0	9,081	0	...	4.0
5622	0	9,858	0	...	245.0
5623	24	48,473	6	...	10.0
5624	0	24,844	14	...	6.0
5625	1	716,043	4	...	22.0
5626	0	8,012	0	...	14.0
5627	0	52,423	9	...	27.0
5628	0	17,532	0	...	14.0
5629	0	391,745	2	...	38.0
5630	0	32,462	10	...	22.0
5631	0	155,664	5	...	83.0
5632	58	181,201	5	...	116.0
5633	182	27,771	80	...	9.0
5634	0	20,083	0	...	21.0
5635	0	214,965	126	...	66.0
5636	37	440,468	0	...	72.0
5637	0	514,945	0	...	149.0

	date_posted	downloads	faves \
0	January 17, 2014	92.0	142.0
1	December 30, 2011	186.0	1135.0

2	August 11, 2012	225.0	1959.0
3	January 21, 2012	48.0	551.0
4	August 19, 2011	34.0	91.0
5	August 1, 2010	59.0	696.0
6	February 20, 2012	46.0	218.0
7	January 27, 2012	14.0	96.0
8	July 13, 2013	8.0	66.0
9	December 26, 2011	39.0	877.0
10	April 14, 2016	16.0	157.0
11	April 9, 2012	146.0	1262.0
12	September 13, 2011	41.0	246.0
13	August 21, 2012	129.0	383.0
14	July 29, 2010	91.0	202.0
15	July 11, 2012	22.0	100.0
16	June 26, 2011	189.0	948.0
17	August 7, 2010	393.0	2515.0
18	September 16, 2013	NaN	375.0
19	May 16, 2013	34.0	258.0
20	January 6, 2013	NaN	1462.0
21	April 23, 2012	211.0	801.0
22	February 27, 2012	184.0	1101.0
23	November 20, 2012	NaN	242.0
24	January 15, 2012	20.0	350.0
25	January 17, 2017	NaN	449.0
26	August 23, 2010	42.0	280.0
27	January 24, 2012	281.0	1167.0
28	September 19, 2013	105.0	377.0
29	November 10, 2012	89.0	508.0
...	...	...	...
5608	September 24, 2012	26.0	72.0
5609	December 9, 2015	NaN	183.0
5610	May 1, 2013	8.0	156.0
5611	March 8, 2014	8.0	152.0
5612	January 14, 2012	96.0	527.0
5613	June 14, 2011	93.0	1186.0
5614	September 1, 2013	14.0	103.0
5615	May 7, 2016	3.0	85.0
5616	November 10, 2009	335.0	275.0
5617	July 14, 2012	138.0	290.0
5618	January 1, 2015	9.0	81.0
5619	December 9, 2008	542.0	1051.0
5620	March 8, 2011	92.0	161.0
5621	March 3, 2014	16.0	73.0
5622	November 15, 2010	203.0	1328.0
5623	June 8, 2012	NaN	125.0
5624	January 9, 2011	14.0	39.0
5625	January 16, 2012	NaN	86.0
5626	March 13, 2014	NaN	102.0

5627	January 29, 2012	24.0	63.0
5628	April 30, 2010	NaN	152.0
5629	August 20, 2012	125.0	451.0
5630	February 8, 2012	83.0	221.0
5631	September 7, 2011	958.0	2249.0
5632	June 17, 2014	NaN	956.0
5633	August 17, 2011	NaN	80.0
5634	October 10, 2010	20.0	66.0
5635	September 15, 2013	121.0	884.0
5636	June 1, 2016	119.0	950.0
5637	March 5, 2014	NaN	2285.0

	hashtags \
0	NaN
1	NaN
2	NaN
3	NaN
4	NaN
5	NaN
6	NaN
7	NaN
8	NaN
9	NaN
10	[sherlock, sherlockholmes, benedictcumberbatch...
11	NaN
12	NaN
13	NaN
14	NaN
15	NaN
16	NaN
17	NaN
18	[sherlockdrawing]
19	NaN
20	NaN
21	NaN
22	NaN
23	NaN
24	NaN
25	[tv, thelastproblem, art, amazing, artdigital,...
26	NaN
27	NaN
28	NaN
29	NaN
...	...
5608	[fanmade, toys, tuy, cumberbatch, sherlockfan,...
5609	[bird, dark, fan, funny, ipad, owl, sherlock, ...
5610	NaN
5611	NaN



5612	NaN
5613	NaN
5614	NaN
5615	[sherlock, tej, art, teart, art_of_tej, tej_a...
5616	NaN
5617	NaN
5618	[adler, bbc, blend, holmes, hooper, irene, joh...
5619	NaN
5620	NaN
5621	NaN
5622	NaN
5623	NaN
5624	NaN
5625	NaN
5626	NaN
5627	NaN
5628	NaN
5629	NaN
5630	NaN
5631	NaN
5632	NaN
5633	NaN
5634	NaN
5635	NaN
5636	[benedict, bride, brideoffrankenstein, canvas,...
5637	NaN

	image_links \
0	<a href="https://www.deviantart.com/get-sherlock/art/Sh...">https://www.deviantart.com/get-sherlock/art/Sh...</a>
1	<a href="https://www.deviantart.com/masterhalfling/art/...">https://www.deviantart.com/masterhalfling/art/...</a>
2	<a href="https://www.deviantart.com/donperico/art/SHERL...">https://www.deviantart.com/donperico/art/SHERL...</a>
3	<a href="https://www.deviantart.com/cheeky-bee/art/Sher...">https://www.deviantart.com/cheeky-bee/art/Sher...</a>
4	<a href="https://www.deviantart.com/hahaaaaaaaaaaaaa/art...">https://www.deviantart.com/hahaaaaaaaaaaaaa/art...</a>
5	<a href="https://www.deviantart.com/1stclasstamps/art/...">https://www.deviantart.com/1stclasstamps/art/...</a>
6	<a href="https://www.deviantart.com/feyjane/art/Sherloc...">https://www.deviantart.com/feyjane/art/Sherloc...</a>
7	<a href="https://www.deviantart.com/elsias/art/Sherlock...">https://www.deviantart.com/elsias/art/Sherlock...</a>
8	<a href="https://www.deviantart.com/sherlockthegame/art...">https://www.deviantart.com/sherlockthegame/art...</a>
9	<a href="https://www.deviantart.com/fractionofadot/art/...">https://www.deviantart.com/fractionofadot/art/...</a>
10	<a href="https://www.deviantart.com/nathalief87/art/She...">https://www.deviantart.com/nathalief87/art/She...</a>
11	<a href="https://www.deviantart.com/risachantag/art/She...">https://www.deviantart.com/risachantag/art/She...</a>
12	<a href="https://www.deviantart.com/coraline125/art/She...">https://www.deviantart.com/coraline125/art/She...</a>
13	<a href="https://www.deviantart.com/intervain/art/Sherl...">https://www.deviantart.com/intervain/art/Sherl...</a>
14	<a href="https://www.deviantart.com/luluha/art/Sherlock...">https://www.deviantart.com/luluha/art/Sherlock...</a>
15	<a href="https://www.deviantart.com/teralilac/art/Sherl...">https://www.deviantart.com/teralilac/art/Sherl...</a>
16	<a href="https://www.deviantart.com/sherlockrocksmyssock...">https://www.deviantart.com/sherlockrocksmyssock...</a>
17	<a href="https://www.deviantart.com/carrienloveyou/art/...">https://www.deviantart.com/carrienloveyou/art/...</a>
18	<a href="https://www.deviantart.com/kuliszu/art/Sherloc...">https://www.deviantart.com/kuliszu/art/Sherloc...</a>
19	<a href="https://www.deviantart.com/jessixo/art/Sherloc...">https://www.deviantart.com/jessixo/art/Sherloc...</a>

20 <https://www.deviantart.com/jasric/art/Sherlock...>  
 21 <https://www.deviantart.com/mibu-no-ookami/art/...>  
 22 <https://www.deviantart.com/aerettberg/art/Sher...>  
 23 <https://www.deviantart.com/eilyn-chan/art/SHER...>  
 24 <https://www.deviantart.com/jadeapprentice/art/...>  
 25 <https://www.deviantart.com/olga-tereshenko/art...>  
 26 <https://www.deviantart.com/kd666/art/SHERLOCK-...>  
 27 <https://www.deviantart.com/sueworld/art/Sherlo...>  
 28 <https://www.deviantart.com/nadia-ch/art/Sherlo...>  
 29 <https://www.deviantart.com/loga90/art/Sherlock...>  
 ...  
 5608 <https://www.deviantart.com/coralfg/art/Sherloc...>  
 5609 <https://www.deviantart.com/tapwing/art/Sherloc...>  
 5610 <https://www.deviantart.com/jadethefirefox/art/...>  
 5611 <https://www.deviantart.com/vedabitsaha/art/She...>  
 5612 <https://www.deviantart.com/naydshiko/art/Sherl...>  
 5613 <https://www.deviantart.com/alice-insain/art/Sh...>  
 5614 <https://www.deviantart.com/yiamme/art/Sherlock...>  
 5615 <https://www.deviantart.com/artoftej/art/Sherlo...>  
 5616 <https://www.deviantart.com/bjak/art/Sherlock-h...>  
 5617 <https://www.deviantart.com/tovmauzer/art/Sherl...>  
 5618 <https://www.deviantart.com/ashiharanakatsu/art...>  
 5619 <https://www.deviantart.com/artofgrelin/art/She...>  
 5620 <https://www.deviantart.com/jaumecullell/art/SH...>  
 5621 <https://www.deviantart.com/crumblincookie/art/...>  
 5622 <https://www.deviantart.com/floating-hobbit/art...>  
 5623 <https://www.deviantart.com/kierstenolsen/art/S...>  
 5624 <https://www.deviantart.com/miss-azalis/art/She...>  
 5625 <https://www.deviantart.com/cerberuslives/art/S...>  
 5626 <https://www.deviantart.com/chunsmunkey/art/She...>  
 5627 <https://www.deviantart.com/enednoviel/art/Sher...>  
 5628 <https://www.deviantart.com/xredderz/art/Sherlo...>  
 5629 <https://www.deviantart.com/bloodlinev/art/Sulk...>  
 5630 <https://www.deviantart.com/dracilla112/art/She...>  
 5631 <https://www.deviantart.com/hyuman/art/SHERLOCK...>  
 5632 <https://www.deviantart.com/greyradian/art/Sher...>  
 5633 <https://www.deviantart.com/iuiuiulia/art/Sherl...>  
 5634 <https://www.deviantart.com/jit-seven/art/Sherl...>  
 5635 <https://www.deviantart.com/gigai/art/Sherlock-...>  
 5636 <https://www.deviantart.com/joe-roberts/art/She...>  
 5637 <https://www.deviantart.com/kimberly80/art/Sher...>

image\_paths \

0 [full/c11741fda71c47368b3ecc42de439de62c48cf30...  
 1 [full/09fbbe05751a1b08cf03445e0821ca44dd5b547b...  
 2 [full/0f13dd8004eb7f0ce2b45c826f4e76df5978ac13...  
 3 [full/2d3083ae52b6ea7b69fdd83c0f74c5fd8525a489...  
 4 [full/be06d405f58195ef9d4c5c42c1c6cb2edb673754...

5 [full/11dc8450503f97976fba168972410295b76c72ed...  
 6 [full/587559fae93b85cbfa8802d39e7a2e1d3cbe415c...  
 7 [full/9507a844991ab37afec9bc0a0f619bfa0d033b04...  
 8 [full/45901b4a9092329595f2e6739d213b2d169e45f9...  
 9 [full/c752fa217f32f5510a49c4f17f3980db9a442c68...  
 10 [full/ecc8785c5591d6de7a3107bea75a24fae0b9150c...  
 11 [full/3bd907ff9ca9af2fafaed00019fe9753be055728...  
 12 [full/03914aae095df72509b4e3018d8b23dcd549e25c...  
 13 [full/74786a06d872dbb80c572834de2f382948591b55...  
 14 [full/dc017adc4ea20ed2d26a004649d295c722676649...  
 15 [full/b0bb868d4f7b569aa5b6f0fa6d5c32387520419e...  
 16 [full/4140b964daef77f89d71d428e561d36e6d2e9fab...  
 17 [full/13445d30465ae6843965d0509f9d8fbd33eccdc1...  
 18 [full/a65c1d68d67daaae2c24253af270129d38af7fda...  
 19 [full/913074bf1d4f1123244320815abec91e53cb701a...  
 20 [full/91bc5581858b6b2e779ad4f0dcd1468a3cf54fe0...  
 21 [full/ea1d1a057bb6f62a41efcb3d21fc0923e1d630a8...  
 22 [full/c2ae2f8172ed019019fbee72038641499c5adf11...  
 23 [full/8d6dbf4ac59c4af6c98115470dc7ff572c8e74fa...  
 24 [full/1b68f0154f53ebfb54c3730395dc8990a90c77d0...  
 25 [full/326ecc620d8c1a816111369f01551666bbd59a1b...  
 26 [full/37875736ae424ee0945998459c4b5b590ad7db31...  
 27 [full/267d5b8f85bff7b75d79351b53fdbfd8402d5f58...  
 28 [full/b492082b1b7fb9282f67268d76492156bc582742...  
 29 [full/e330cf67b33a22c9b640564f7b6fe9e7fe61ff6d...  
 ...  
 5608 [full/b93da70b2fb0a7567390ff1a44dadfa370fa8e8c...  
 5609 [full/4c0dc58a7cac6fb08d76679e4328e77dd46ed1cf...  
 5610 [full/67531b928465ed15073e6bc49699f0a6a334a20f...  
 5611 [full/4d9ccda45b4256337ab926453f58373d6f4dccc6...  
 5612 [full/6588fe2c2832c66fa3cba12b0a8c2cea3622371d...  
 5613 [full/cea652cb3765d5b74d13d30340f1da8873b4d534...  
 5614 [full/0dcc71157ca6c6f73029a321bcc1e27b5e09fb57...  
 5615 [full/130308117de6a4c3a19d2656aa52190e28248759...  
 5616 [full/4d1213476048008fe6a3d79dd60751421000114f...  
 5617 [full/68987fd59d38dc8f8f4cdccdfa1ea2b83616b28f...  
 5618 [full/905b773566b436e1dc043533d6eaa4e6f0261750...  
 5619 [full/a4fc24a8689b3931aa72adb33778e059227ddfe0...  
 5620 [full/638fb0ac58fe5f5849f1a021c5ad119242b12ec3...  
 5621 [full/c5a31ff34a43fb82c9eb1b3622d0434b887bf5de...  
 5622 [full/3c7af4c7e736baeee3e681d5a23790718b64b4d1...  
 5623 [full/afb62280f31544ca90ebeae352a46481349f36c6...  
 5624 [full/8b65ccb3df679c0386f43526b66b460fb134820b...  
 5625 [full/8c8ce91927c013023e9e80dfb5c4a7ea75e61f9b...  
 5626 [full/9d749f3a5369212f08bb2f6d6fef550f21be99a0...  
 5627 [full/fb759e5ac781a852c7e505744176a1d5da386468...  
 5628 [full/6850b8352f8407eb919a262743218939ea3d24e4...  
 5629 [full/4549c11d29cae6b6aeb782cdd5ba5089359fedd5...

5630 [full/3e24b33c326afe011c29aadafe6563974876b25fc...  
 5631 [full/669f2975001098e2c334bf5e3e8f49bc8b6adf5f...  
 5632 [full/f860c98585500dbecd4150c12e180e68467ac92c...  
 5633 [full/61c1ecb85714048574a9696e570bc131447fe677...  
 5634 [full/f14d8bfedcf579019d5719bb75c6222eafba214a...  
 5635 [full/a212ffab3456626f176885fe1d5dcb5fc9fa0f0c...  
 5636 [full/071c38ac67037aca590719c02d623a41b100c17e...  
 5637 [full/47c23bc68e46067af1f60f20ab9951d0bdc88c47...

image\_urls \

0 [https://images-wixmp-ed30a86b8c4ca887773594c2...  
 1 [https://images-wixmp-ed30a86b8c4ca887773594c2...  
 2 [https://images-wixmp-ed30a86b8c4ca887773594c2...  
 3 [https://images-wixmp-ed30a86b8c4ca887773594c2...  
 4 [https://images-wixmp-ed30a86b8c4ca887773594c2...  
 5 [https://images-wixmp-ed30a86b8c4ca887773594c2...  
 6 [https://images-wixmp-ed30a86b8c4ca887773594c2...  
 7 [https://images-wixmp-ed30a86b8c4ca887773594c2...  
 8 [https://images-wixmp-ed30a86b8c4ca887773594c2...  
 9 [https://images-wixmp-ed30a86b8c4ca887773594c2...  
 10 [https://images-wixmp-ed30a86b8c4ca887773594c2...  
 11 [https://images-wixmp-ed30a86b8c4ca887773594c2...  
 12 [https://images-wixmp-ed30a86b8c4ca887773594c2...  
 13 [https://images-wixmp-ed30a86b8c4ca887773594c2...  
 14 [https://images-wixmp-ed30a86b8c4ca887773594c2...  
 15 [https://images-wixmp-ed30a86b8c4ca887773594c2...  
 16 [https://images-wixmp-ed30a86b8c4ca887773594c2...  
 17 [https://images-wixmp-ed30a86b8c4ca887773594c2...  
 18 [https://images-wixmp-ed30a86b8c4ca887773594c2...  
 19 [https://images-wixmp-ed30a86b8c4ca887773594c2...  
 20 [https://images-wixmp-ed30a86b8c4ca887773594c2...  
 21 [https://images-wixmp-ed30a86b8c4ca887773594c2...  
 22 [https://images-wixmp-ed30a86b8c4ca887773594c2...  
 23 [https://img00.deviantart.net/f19f/i/2015/113/...  
 24 [https://images-wixmp-ed30a86b8c4ca887773594c2...  
 25 [https://images-wixmp-ed30a86b8c4ca887773594c2...  
 26 [https://images-wixmp-ed30a86b8c4ca887773594c2...  
 27 [https://images-wixmp-ed30a86b8c4ca887773594c2...  
 28 [https://images-wixmp-ed30a86b8c4ca887773594c2...  
 29 [https://images-wixmp-ed30a86b8c4ca887773594c2...  
 ...  
 5608 [https://images-wixmp-ed30a86b8c4ca887773594c2...  
 5609 [https://images-wixmp-ed30a86b8c4ca887773594c2...  
 5610 [https://images-wixmp-ed30a86b8c4ca887773594c2...  
 5611 [https://images-wixmp-ed30a86b8c4ca887773594c2...  
 5612 [https://images-wixmp-ed30a86b8c4ca887773594c2...  
 5613 [https://images-wixmp-ed30a86b8c4ca887773594c2...  
 5614 [https://images-wixmp-ed30a86b8c4ca887773594c2...

5615 [https://images-wixmp-ed30a86b8c4ca887773594c2...  
 5616 [https://images-wixmp-ed30a86b8c4ca887773594c2...  
 5617 [https://images-wixmp-ed30a86b8c4ca887773594c2...  
 5618 [https://images-wixmp-ed30a86b8c4ca887773594c2...  
 5619 [https://images-wixmp-ed30a86b8c4ca887773594c2...  
 5620 [https://images-wixmp-ed30a86b8c4ca887773594c2...  
 5621 [https://images-wixmp-ed30a86b8c4ca887773594c2...  
 5622 [https://images-wixmp-ed30a86b8c4ca887773594c2...  
 5623 [https://img00.deviantart.net/b49c/i/2015/114/...  
 5624 [https://images-wixmp-ed30a86b8c4ca887773594c2...  
 5625 [https://images-wixmp-ed30a86b8c4ca887773594c2...  
 5626 [https://images-wixmp-ed30a86b8c4ca887773594c2...  
 5627 [https://images-wixmp-ed30a86b8c4ca887773594c2...  
 5628 [https://images-wixmp-ed30a86b8c4ca887773594c2...  
 5629 [https://images-wixmp-ed30a86b8c4ca887773594c2...  
 5630 [https://images-wixmp-ed30a86b8c4ca887773594c2...  
 5631 [https://images-wixmp-ed30a86b8c4ca887773594c2...  
 5632 [https://images-wixmp-ed30a86b8c4ca887773594c2...  
 5633 [https://images-wixmp-ed30a86b8c4ca887773594c2...  
 5634 [https://images-wixmp-ed30a86b8c4ca887773594c2...  
 5635 [https://images-wixmp-ed30a86b8c4ca887773594c2...  
 5636 [https://images-wixmp-ed30a86b8c4ca887773594c2...  
 5637 [https://pre00.deviantart.net/f400/th/pre/i/20...

	titles	views
0	Sherlock	1606.0
1	Sherlock	14153.0
2	SHERLOCK	21163.0
3	Sherlock Of the Dead	6988.0
4	Sherlock	2742.0
5	Sherlock Holmes	14566.0
6	Sherlock	3139.0
7	Sherlock	2208.0
8	Sherlock: The Game Is On (Character Contest)	2983.0
9	Sherlock	10189.0
10	Sherlock	1430.0
11	Sherlock	16771.0
12	Sherlock	4751.0
13	Sherlock	8399.0
14	Sherlock	12179.0
15	Sherlock	1923.0
16	Abbey Road - Sherlock Style	16569.0
17	sherlock	43269.0
18	Sherlock	5226.0
19	Sherlock	3215.0
20	Sherlock	14131.0
21	Sherlock	14674.0
22	Sherlock	19440.0

23	SHERLOCK	2694.0
24	Sherlock	6218.0
25	Sherlock	2475.0
26	SHERLOCK	10417.0
27	'Sherlock'	16220.0
28	Sherlock	4734.0
29	Sherlock	7000.0
...	...	...
5608	Sherlock	2695.0
5609	Sherlock	1400.0
5610	British Gothic -Sherlock	2442.0
5611	Sherlock Holmes	1422.0
5612	Sherlock on the Tube	15199.0
5613	Sherlock	19117.0
5614	Sherlock!	2218.0
5615	Sherlock	533.0
5616	Sherlock holmes	11341.0
5617	Sherlock Holmes	9582.0
5618	BLEND - Sherlock	1195.0
5619	Sherlock holmes	25954.0
5620	SHERLOCK	4364.0
5621	Sherlock	1093.0
5622	Sherlock: Hey John	39280.0
5623	Sherlock	2400.0
5624	Sherlock	1604.0
5625	Sherlock	3381.0
5626	Sherlock	794.0
5627	Sherlock	1302.0
5628	Sherlock	3848.0
5629	Sulking Sherlock	4751.0
5630	Sherlock	3280.0
5631	SHERLOCK: BORED	41651.0
5632	Sherlock	8377.0
5633	Sherlock	1947.0
5634	Sherlock	2720.0
5635	Sherlock	11722.0
5636	Sherlock	8390.0
5637	Sherlock	20086.0

[5638 rows x 23 columns]