Schrödinger's Cat or Lululemon's Discounts?

Xiaofei Xue

Agenda

- What's the problem?
- Background
- Scraped data and its process
- Checking the data
- Add-on features
- Future work
- Q&A
- Reference and acknowledgement

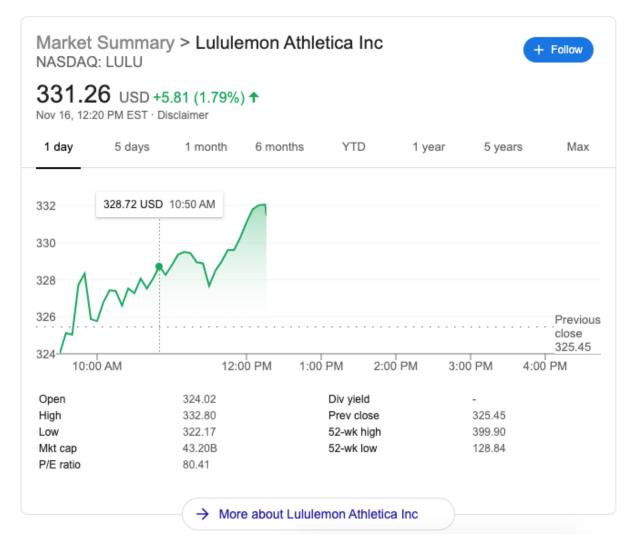
Prompt 3

• Lululemon's management is telling us that most of their clothes are sold at full price and they rarely discount. We want to know if this is true.

Background on Lululemon

- Founded in 1998 as a retailer of yoga pants and other yoga wear, Lululemon has since expanded to sell its products internationally in 460 stores as well as online.
- The company has expanded to sell a variety of athletic wear, including performance shirts, shorts, and pants, as well as lifestyle apparel and yoga accessories.
- The company's initial public offering was in July 2007, raising \$327.6 million by selling 18.2 million shares.
- In 2019, lululemon announced an investment in Mirror, a fitness startup that sells an interactive mirror with a camera and speakers for at-home workouts. The companies planned to create new content for the device, starting with meditation classes.
- In June 2020, Lululemon announced a \$500 million deal to purchase Mirror, capitalizing on a growing trend of people conducting virtual workouts at home instead of going to a gym due to the COVID-19 pandemic.

Background in Lululemon



Scrapped Data

In [408]: acc sale

Out[408]:

Unique ID	Category	Item Category	Display Name	Color	URL	Sale Or Not	Sale Price	Original Price	Sizes In Stock	Sizes OOS
us_104428440	Accessories	women	Find Your Calm Wrap	Serene Blue	https://shop.lululemon.com/p/women/Find-Your- C	1	\$99.00	\$148.00	'4', '6', '8'	'2', '10', '12'
us_122825361	Accessories	Sweaters	Opened Up Poncho	Iron Blue/True Navy	https://shop.lululemon.com/p/women-sweaters- an		\$89.00	\$148.00	'XS/S', 'M/L'	
us_122825361	Accessories	Sweaters	Opened Up Poncho	Heathered Core Ultra Light Grey/Heathered Asph	https://shop.lululemon.com/p/women-sweaters- an		\$89.00	\$148.00	'XS/S', 'M/L'	
us_122825361	Accessories	Sweaters	Opened Up Poncho	Black	https://shop.lululemon.com/p/women-sweaters- an		\$89.00	\$148.00	'XS/S', 'M/L'	
us_3798493	Accessories	Hats	Lightspeed Run Hat	Black	https://shop.lululemon.com/p/hats/Lightspeed-R	1	\$19.00	\$38.00	'ONE SIZE'	
us_114750380	Accessories	Hats	Days Shade Ball Cap	White	https://shop.lululemon.com/p/hats/Days-Shade-B	1	\$19.00	\$38.00	'ONE SIZE'	

Scrapped Data

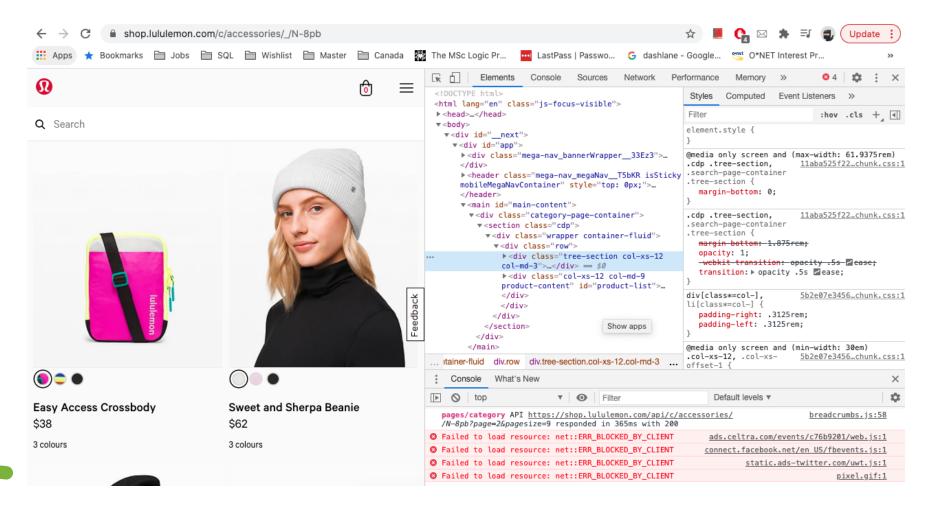
In [331]: acc_origin

	Unique ID	Category	Item Category	Display Name	Color	URL	Sale Or Not	Sale Price	Original Price	Sizes In Stock	Sizes
0	us_122976244	Accessories	Bags	Easy Access Crossbody	Vapor/Highlight Purple/Black	https://shop.lululemon.com/p/bags/Easy- Access	0	NA	[38.00]	'ONE SIZE'	
1	us_122976244	Accessories	Bags	Easy Access Crossbody	Saffron/Cerulean Blue/Cassis	https://shop.lululemon.com/p/bags/Easy- Access	0	NA	[38.00]	'ONE SIZE'	
2	us_122976244	Accessories	Bags	Easy Access Crossbody	Black	https://shop.lululemon.com/p/bags/Easy- Access	0	NA	[38.00]	'ONE SIZE'	
3	us_113000633	Accessories	Hats	Sweet and Sherpa Beanie	Heathered Core Ultra Light Grey	https://shop.lululemon.com/p/hats/Sweet-and-Sh	0	NA	[62.00]	'ONE SIZE'	
4	us_113000633	Accessories	Hats	Sweet and Sherpa Beanie	Heathered Pink Taupe	https://shop.lululemon.com/p/hats/Sweet-and-Sh	0	NA	[62.00]	'ONE SIZE'	
5	us_113000633	Accessories	Hats	Sweet and Sherpa Beanie	Black	https://shop.lululemon.com/p/hats/Sweet-and-Sh	0	NA	[62.00]	'ONE SIZE'	

Scraping - Finding API

- Official Website > women/men/accessories > right click for "Inspect"
- "Console" is on the lower left part of the page
- "API" is listed under "Console"

Scraping - Finding API



Scraping - Finding API

```
shop.lululemon.com/api/c/accessories/
                                                                                                                                                         Update
 🚃 Apps 🛊 Bookmarks 🗎 Jobs 🗎 SQL 🗎 Wishlist 🗎 Master 🗎 Canada 🐯 The MSc Logic Pr... 🚾 LastPass | Passwo... 💪 dashlane - Google... 🥰 O*NET Interest Pr..
{"links": {"self": "/c/accessories/?page=1", "first": "/c/accessories/?page=1", "prev": "/c/accessories/?page=1", "next": "/c/accessories/?page=2", "last":
"/c/accessories/?page=12"}, "meta": {"topratedenabled": "False"}, "data": {"id": "cdp", "type": "ProductCategoryDetailsPage", "name": "Product Category Details
Pages", "h1-title-tag": "Accessories", "isFusionQuery": true, "isBopisEnabled": true, "fusion-query-id": "BQBrcCzjGE", "fusionExperimentVariant": "false",
"fusionExperimentId": "", "attributes": {"main-content": [{"type": "CDPResultsList", "paging-action-template": {"content-path": "/c", "navigation-state":
"/accessories/_(N-8pb?No=%7Boffset%7D", "site-root-path": "/pages", "label": null}, "name": "CDP Results List", "total-num-recs": 227, "first-rec-num": 1, "last-
rec-num": 20, "recs-per-page": 20, "records": [{"default-parent-category": "Bags", "product-silver": "false", "common-id": "125050038 prod9960678", "list-price":
["38.00"], "product-sale-price": [], "display-name-with-br": "", "default-sku": "us 122976244", "product-whats-new": "1", "product-category-hierarchy": "['Shop
Outfits^cat2200003', 'women^uswomen', 'Gifts for Him^cat1460011', 'Accessories accessories Bags^cat1350014', 'men^usmen', 'Gifts for Her^cat1460013']", "linked-
products": [""], "all-available-sizes": ["ONE SIZE"], "sku-sku-images": ["https://images.lululemon.com/is/image/lululemon/LU9AKRS 047202 1",
"https://images.lululemon.com/is/image/lululemon/LU9AKRS 047202 2", "https://images.lululemon.com/is/image/lululemon/LU9AJLS 047120 1",
"https://images.lululemon.com/is/image/lululemon/LU9AJLS 047120 2", "https://images.lululemon.com/is/image/lululemon/LU9AJJS 0001 1",
"https://images.lululemon.com/is/image/lululemon/LU9AJJS 0001 2"], "pdp-url": "/p/bags/Easy-Access-Crossbody/ /prod9960678", "bazaar-voice-id":
"Easy_Access_Crossbody", "f5-bckimg-url": "", "title": "Easy Access Crossbody | Bags", "parent-category-unified-id": "bags", "sku-style-number": ["LU9AKRS-
047202", "LU9AJLS-047120", "LU9AJJS-0001"], "unified-id": "Easy-Access-Crossbody", "parent-category-display-name": "Bags", "repository-id": "prod9960678", "sku-
style-order": [{"color-group": "multiNeonpink", "sku-style-order-id": "5", "style-id": "LU9AKRS-047202", "style-id-01": "LU9AKRS", "color-name": "Vapor/Highlight
Purple/Black", "color-id": "47202", "sku": "us 125050038", "size": "ONE SIZE", "style-id-02": "047202"}, {"color-group": "multiblue", "sku-style-order-id": "2",
"style-id": "LU9AJLS-047120", "style-id-01": "LU9AJLS", "color-name": "Saffron/Cerulean Blue/Cassis", "color-id": "47120", "sku": "us 122976266", "size": "ONE
SIZE", "style-id-02": "047120"}, {"color-group": "black", "sku-style-order-id": "1", "style-id": "LU9AJJS-0001", "style-id-01": "LU9AJJS", "color-name": "Black",
"color-id": "0001", "sku": "us 122976244", "size": "ONE SIZE", "style-id-02": "0001"}], "color-group": [{"name": "multi", "url":
"https://images.lululemon.com/is/image/lululemon/43899?$swatch$"}, {"name": "Neon", "url": "https://images.lululemon.com/is/image/lululemon/46171?$swatch$"},
{"name": "pink", "url": "https://images.lululemon.com/is/image/lululemon/pink swatch?\swatch?\swatch?\, {"name": "blue", "url":
"https://images.lululemon.com/is/image/lululemon/blue swatch?$swatch$"}, {"name": "black", "url": "https://images.lululemon.com/is/image/lululemon/black swatch?
$swatch$"}], "has-linked-products": "false", "product-applicable-systems": "All", "currency-code": "USD", "image-scheme": "Old", "display-name": "Easy Access
Crossbody", "product-on-sale": "0"}, {"default-parent-category": "Hats", "product-silver": "false", "common-id": "113000651_prod9600678", "list-price": ["62.00"],
"product-sale-price": [], "display-name-with-br": "", "default-sku": "us 113000633", "product-whats-new": "0", "product-category-hierarchy": '
['Loungewear^cat1460084', 'Cold Weather^cat1880004', 'Shop Outfits^cat2200003', 'women'uswomen', 'Accessories^usaccessories|Hats^ushats', 'Gifts for
Her^cat1460013']", "linked-products": [""], "all-available-sizes": ["ONE SIZE"], "sku-sku-images":
["https://images.lululemon.com/is/image/lululemon/LW9CMWS 032493 1", "https://images.lululemon.com/is/image/lululemon/LW9CMWS 032493 2",
"https://images.lululemon.com/is/image/lululemon/LW9CMWS_041975_1", "https://images.lululemon.com/is/image/lululemon/LW9CMWS_041975_2",
"https://images.lululemon.com/is/image/lululemon/LW9CHUS 0001 1", "https://images.lululemon.com/is/image/lululemon/LW9CHUS 0001 2"], "pdp-url": "/p/hats/Sweet-
and-Sherpa-Beanie/ /prod9600678", "bazaar-voice-id": "Sweet and Sherpa Beanie", "f5-bckimg-url": "https://images.lululemon.com/is/image/lululemon/LW1BMIS 0001",
"title": "Sweet and Sherpa Beanie | Women's Hats", "parent-category-unified-id": "hats", "sku-style-number": ["LW9CMWS-032493", "LW9CMWS-041975", "LW9CHUS-0001"],
"unified-id": "Sweet-and-Sherpa-Beanie", "parent-category-display-name": "Hats", "repository-id": "prod9600678", "sku-style-order": [{"color-group": "grey", "sku-
style-order-id": "3", "style-id": "LW9CMWS-032493", "style-id-01": "LW9CMWS", "color-name": "Heathered Core Ultra Light Grey", "color-id": "32493", "sku":
"us 113000651", "size": "ONE SIZE", "style-id-02": "032493"}, {"color-group": "pink", "sku-style-order-id": "2", "style-id": "LW9CMWS-041975", "style-id-01":
"LW9CMWS", "color-name": "Heathered Pink Taupe", "color-id": "41975", "sku": "us 113000652", "size": "ONE SIZE", "style-id-02": "041975"}, {"color-group":
"black", "sku-style-order-id": "1", "style-id": "LW9CHUS-0001", "style-id-01": "LW9CHUS", "color-name": "Black", "color-id": "0001", "sku": "us 113000633",
"size": "ONE SIZE", "style-id-02": "0001"}], "color-group": [{"name": "grey", "url": "https://images.lululemon.com/is/image/lululemon/grey swatch?$swatch$"},
{"name": "pink", "url": "https://images.lululemon.com/is/image/lululemon/pink swatch?$swatch$"}, {"name": "black", "url":
"https://images.lululemon.com/is/image/lululemon/black_swatch?$swatch$"}], "has-linked-products": "false", "product-applicable-systems": "All", "currency-code":
"USD", "image-scheme": "Old", "display-name": "Sweet and Sherpa Beanie", "product-on-sale": "0"}, {"default-parent-category": "Hats", "product-silver": "false",
"common-id": "122750497 prod9960888", "list-price": ["48.00"], "product-sale-price": [], "display-name-with-br": "", "default-sku": "us 122750497", "product-
```

Scraping – Understanding the Format

- https://jsonformatter.org/
- https://jsonformatter.curiousconcept.com/#
- https://www.json.cn/
- Transform the page into a more readable format

Scraping – Understanding the Format

```
{"links": {"self": "/c/accessories/?page=1", "first":
"/c/accessories/?page=1", "prev": "/c/accessories/?page=1",
"next": "/c/accessories/?page=2", "last": "/c/accessories/?
page=12"}, "meta": {"topratedenabled": "False"}, "data": {"id":
"cdp", "type": "ProductCategoryDetailsPage", "name": "Product
Category Details Pages", "h1-title-tag": "Accessories",
"isFusionQuery": true, "isBopisEnabled": true, "fusion-query-id":
"BQBrcCzjGE", "fusionExperimentVariant": "false",
"fusionExperimentId": "", "attributes": {"main-content":
[{"type": "CDPResultsList", "paging-action-template": {"content-
path": "/c", "navigation-state": "/accessories/_/N-8pb?
No=%7Boffset%7D", "site-root-path": "/pages", "label": null},
"name": "CDP Results List", "total-num-recs": 227, "first-rec-
num": 1, "last-rec-num": 20, "recs-per-page": 20, "records":
[{"default-parent-category": "Bags", "product-silver": "false",
"common-id": "125050038_prod9960678", "list-price": ["38.00"],
"product-sale-price": [], "display-name-with-br": "", "default-
sku": "us_122976244", "product-whats-new": "1", "product-
category-hierarchy": "['Shop Outfits^cat2200003',
'women^uswomen', 'Gifts for Him^cat1460011',
'Accessories\usaccessories|Bags\cat1350014', 'men\usmen', 'Gifts
for Her^cat1460013']", "linked-products": [""], "all-available-
sizes": ["ONE SIZE"], "sku-sku-images":
["https://images.lululemon.com/is/image/lululemon/LU9AKRS_047202
"https://images.lululemon.com/is/image/lululemon/LU9AKRS 047202 2
"https://images.lululemon.com/is/image/lululemon/LU9AJLS 047120 1
```

```
"links":⊟{
    "self":"/c/accessories/?page=1",
   "first": "/c/accessories/?page=1",
    "prev": "/c/accessories/?page=1",
   "next": "/c/accessories/?page=2",
   "last": "/c/accessories/?page=12"
"meta":⊟{
    "topratedenabled": "False"
"data":⊟{
    "id":"cdp",
    "type": "ProductCategoryDetailsPage",
    "name": "Product Category Details Pages",
    "h1-title-tag": "Accessories",
    "isFusionQuery":true,
    "isBopisEnabled":true,
    "fusion-query-id": "BQBrcCzjGE",
    "fusionExperimentVariant": "false",
    "fusionExperimentId":"",
    "attributes":⊟{
        "main-content":⊟[
            ⊟{
                "type": "CDPResultsList",
                "paging-action-template":⊟{
                    "content-path": "/c",
```

Scraping – Get the Data

```
def get data(url):
   r = requests.get(url)
    soup = BeautifulSoup(r.text, 'html.parser')
    # use BS to parse the HTML format
   data = soup.findAll(text=True)
    # 'data' is a ResultSet but needs to be a string to parse as json
   full string = ""
    for each in data:
        str(each)
        full string += each
    parsed = json.loads(full string)
    text = json.dumps(parsed, sort_keys=True, indent=4)
    return parsed
```

```
get_data('https://shop.lululemon.com/api/c/accessories')
```

Scraping – Last Page

Return how many pages to iterate through

```
parsed['links']
{'self': '/c/accessories?page=1',
 'first': '/c/accessories?page=1',
 'prev': '/c/accessories?page=1',
 'next': '/c/accessories?page=2',
 'last': '/c/accessories?page=12'}
def get last page(parsed):
   last page json = parsed['links']['last']
   last_page = last_page_json.split('=')[1]
    return last page
get last page(parsed)
'12'
```

Scraping – Build the right links

```
category 1 = 'Accessories'
cat = category to scrape(category 1)
print(cat)
https://shop.lululemon.com/api/c/accessories?page=
lulu = 'https://shop.lululemon.com'
scrape url = page url builder(cat, 1)
print(scrape_url)
https://shop.lululemon.com/api/c/accessories?page=1&pagesize=45
parsed = get data(scrape url)
print(parsed)
```

```
parsed = get_data('https://shop.lululemon.com/api/c/accessories')

for key in parsed:
    print(key)

links
meta
data
```

```
parsed['data']
{'id': 'cdp',
 'type': 'ProductCategoryDetailsPage',
 'name': 'Product Category Details Pages',
 'h1-title-tag': 'Accessories',
 'isFusionQuery': True,
 'isBopisEnabled': True,
 'fusion-query-id': '05PsaDXzTk',
 'fusionExperimentVariant': '',
 'fusionExperimentId': '',
 'attributes': {'main-content': [{'last-rec-num': 20,
    'recs-per-page': 20,
    'sort-options': [{'navigation-state': '/accessories/_/N-8pb',
      'label': 'Featured',
      'selected': False,
      'content-path': '/c'},
     {'navigation-state': '/accessories/ /N-8pb?Ns=product.last SKU addition dateTime%7C1',
      'label': 'New Arrivals',
      'selected': False,
      'content-path': '/c'},
len(parsed['data'])
```

```
[print(key) for key in parsed['data']]
id
type
name
h1-title-tag
isFusionQuery
isBopisEnabled
fusion-query-id
fusionExperimentVariant
fusionExperimentId
attributes
```

```
parsed['data']['attributes']
{'main-content': [{'last-rec-num': 20,
   'recs-per-page': 20,
   'sort-options': [{'navigation-state': '/accessories/_/N-8pb',
     'label': 'Featured',
     'selected': False,
     'content-path': '/c'},
    {'navigation-state': '/accessories/ /N-8pb?Ns=product.last SKU addition dateTime%7C1',
     'label': 'New Arrivals',
     'selected': False,
     'content-path': '/c'},
    {'navigation-state': '/accessories/ /N-8pb?Ns=price%7C1',
     'label': 'Price: High to Low',
     'selected': False,
     'content-path': '/c'},
    {'navigation-state': '/accessories/ /N-8pb?Ns=price%7C0',
     'label': 'Price: Low to High',
     'selected': False,
     'content-path': '/c'}],
   'paging-action-template': {'site-root-path': '/pages',
len(parsed['data']['attributes'])
2
[print(key) for key in parsed['data']['attributes']]
main-content
secondary-content
```

```
len(parsed['data']['attributes']['main-content'])
2
parsed['data']['attributes']['main-content'][0]
  'navigation-state': '/accessories/ /N-8pb?No=%7Boffset%7D',
  'label': None,
  'content-path': '/c'},
 'first-rec-num': 1,
 'records': [{'repository-id': 'prod9960678',
   'common-id': '125050038 prod9960678',
   'sku-sku-images': ['https://images.lululemon.com/is/image/lululemon/LU9AKRS 047202 1',
    'https://images.lululemon.com/is/image/lululemon/LU9AKRS 047202 2',
    'https://images.lululemon.com/is/image/lululemon/LU9AJLS 047120 1',
    'https://images.lululemon.com/is/image/lululemon/LU9AJLS 047120 2',
    'https://images.lululemon.com/is/image/lululemon/LU9AJJS 0001 1',
    'https://images.lululemon.com/is/image/lululemon/LU9AJJS 0001 2'],
   'image-scheme': 'Old',
   'unified-id': 'Easy-Access-Crossbody',
   'product-category-hierarchy': "['Shop Outfits^cat2200003', 'women^uswomen', 'Gifts for Him^cat1460011', 'Accessori
es^usaccessories Bags^cat1350014', 'men^usmen', 'Gifts for Her^cat1460013']",
   'sku-style-order': [{'color-group': 'multiNeonpink',
     'sku': 'us 125050038',
     'style-id': 'LU9AKRS-047202',
     'style-id-02': '047202'
parsed['data']['attributes']['main-content'][0]['records']
```

```
20
There are 20 products on this parsed page.
1st item, after pdp-url, Easy Access Crossbody
20th item, after pdp-url, Sit In Lotus Wrap II
 unique id = parsed 2['data']['attributes']['main-content'][0]['records'][17]['default-sku']
 print(unique_id)
 us 117226621
 type_clothing = parsed_2['data']['attributes']['main-content'][0]['records'][17]['default-parent-category']
 print(type_clothing)
 women
 display_name = parsed 2['data']['attributes']['main-content'][0]['records'][17]['display-name']
 print(display_name)
 Sundown Wrap
 list_price = parsed_2['data']['attributes']['main-content'][0]['records'][17]['list-price']
 print(list_price)
 ['148.00']
```

len(parsed['data']['attributes']['main-content'][0]['records'])

```
def get_target_feature(target_key, records):
    return [record[target_key] for record in records]

records = parsed['data']['attributes']['main-content'][0]['records']
target_keys = ['default-sku','product-on-sale', 'product-sale-price', 'list-price']

extracted_prices = {}
for key in target_keys:
    extracted_prices[key] = get_target_feature(key, records)

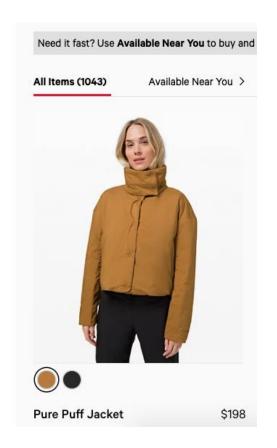
extracted_prices = pd.DataFrame(extracted_prices)
extracted_prices
```

Scraping - Extracting

	default-sku	product-on-sale	product-sale-price	list-price
0	us_122976244	0	0	[38.00]
1	us_113000633	0	0	[62.00]
2	us_122750497	0	0	[48.00]
3	us_119750057	0	0	[42.00]
4	us_118800039	0	0	[28.00]
5	us_123000045	0	0	[48.00]
6	us_122800039	0	0	[48.00]
7	us_3788498	0	0	[38.00]
8	us_3797320	0	0	[68.00, 78.00]
9	us_123052806	0	0	[38.00, 42.00]
10	us_114826689	0	0	[68.00]
11	us_113150067	0	0	[24.00]
12	us_124877210	0	0	[48.00]
13	us_3816655	0	0	[12.00]
14	us_122800278	0	0	[68.00]
15	us_3721293	0	0	[14.00]

Checking the data

If scraped by women, men and accessories, the total number should be 1733, which is 1043 + 486 + 222, which includes the repeated items that falls into both women and men. (Timestamp 2020/11/15 10:21:37 EST)







Checking the data

```
num 1 = len(pd.unique(acc sale['Unique ID']))
num 1
22
num_2 = len(pd.unique(acc_origin['Unique ID']))
num 3 = len(pd.unique(women sale['Unique ID']))
num_4 = len(pd.unique(women_origin['Unique ID']))
num 5 = len(pd.unique(men sale['Unique ID']))
num 6 = len(pd.unique(men origin['Unique ID']))
total = num 1 + num 2 + num 3 + num 4 + num 5 + num 6
print(num 1 + num 2)
print(num 3 + num 4)
print(num 5 + num 6)
220
945
454
total
1619
```

```
num_1/num_2 * 100
```

11.11111111111111

52.41935483870967

37,99392097264438

Checking the data - Basic Statistics

women_origin_copy.describe()

	Sale Price	Original Price	#_Sizes_In_Stock	#_Sizes_OOS	Discount%
count	2058.0	2058.000000	2058.000000	2058.000000	2058.0
mean	0.0	85.431487	5.370748	1.295918	0.0
std	0.0	67.397928	3.175786	1.869436	0.0
min	0.0	8.000000	1.000000	0.000000	0.0
25%	0.0	54.000000	3.000000	0.000000	0.0
50%	0.0	68.000000	6.000000	1.000000	0.0
75%	0.0	108.000000	7.000000	2.000000	0.0
max	0.0	1200.000000	29.000000	13.000000	0.0

Checking the data - Basic Statistics

women_sale_copy.describe()

	Sale Price	Original Price	#_Sizes_In_Stock	#_Sizes_OOS	Discount%
count	568.000000	568.000000	568.000000	568.000000	568.000000
mean	64.466549	103.139085	3.491197	3.387324	36.396391
std	26.719569	42.637474	2.131462	3.330366	13.578058
min	14.000000	20.000000	1.000000	0.000000	11.360000
25%	39.000000	68.000000	2.000000	1.000000	26.850000
50%	69.000000	98.000000	3.000000	3.000000	32.760000
75%	81.500000	128.000000	5.000000	5.000000	46.090000
max	179.000000	298.000000	8.000000	24.000000	85.350000

Checking the data - Random Sample

	<pre>ample= women_origin.sample(n = 50) ample</pre>																
	Unique ID	Category	Item Category	Display Name	Color	URL	Sale Or Not	Sale Price	Original Price	Sizes In Stock	Sizes OOS						
87	us_112975499	Womens	Capris	lululemon Align™ Crop *21"	Heritage 365 Camo Green Twill Multi	https://shop.lululemon.com/p/women- crops/Align	0	NA	[88.00]	'0', '2', '4', '6', '8', '10', '12', '21"',	'14'						
550	us_110025838	Womens	Skirts	Pace Rival Skirt (Regular) *4- way Stretch 13"	White	https://shop.lululemon.com/p/skirts-and-dresse	0	NA	[68.00]	'8', '10', '12', '14'	'0', '2', '4', '6'						
813	us_108125741	Womens	Coats & Jackets	Non-Stop Bomber *Reversible	Frontier	https://shop.lululemon.com/p/jackets-and- hoodi	0	NA	[148.00]	'0', '2', '4', '6', '8', '10', '12'	'14'						
samp	le.to_csv('	lulu_samp	ole', inde	ex = False, h	neader = True)		sample.to_csv('lulu_sample', index = False, header = True)										

Checking the data – Remove Duplicates

```
def find_dup_rows(df):
    num_row_df = df.shape[0]
    uni_series = df.nunique()
    uni_df = pd.DataFrame({'Column_Name': uni_series.index, '#_Unique': uni_series.values})

num_uni_url = uni_df.loc[5, '#_Unique']
    if num_row_df > num_uni_url:
        print('There are duplicated rows in df.')
        dup_in_df = df[df.duplicated()]
        print(dup_in_df['URL'].to_list())
        print(len(dup_in_df['URL'].to_list()))
    return dup_in_df
```

```
find_dup_rows(women_sale_copy)

There are duplicated rows in df.
['https://shop.lululemon.com/p/womens-outerwear/All-Yours-Zip-Hoodie-MD/_/prod9370338?color=43948', 'https://shop.lululemon.com/p/womens-outerwear/All-Yours-Zip-Hoodie-MD/_/prod9370338?color=0001', 'https://shop.lululemon.com/p/womens-outerwear/All-Yours-Zip-Hoodie-MD/_/prod9370338?color=43608', 'https://shop.lululemon.com/p/women/Open-Mind-Wrap-MD/_/prod10080177?color=36756', 'https://shop.lululemon.com/p/women/Open-Mind-Wrap-MD/_/prod10080177?color=30437']
```

Checking the data – Remove Duplicates

```
def find_dup_url(df):
    num_row_df = df.shape[0]
    uni_series = df.nunique()
    uni_df = pd.DataFrame({'Column_Name': uni_series.index, '#_Unique': uni_series.values})

num_uni_url = uni_df.loc[5, '#_Unique']
    if num_row_df > num_uni_url:
        print('There are duplicated rows in df.')
        url_series = df['URL'].value_counts()
        url_pd = pd.DataFrame({'URL': url_series.index, '#_URL': url_series.values})
        dup_url_pd = url_pd[url_pd['#_URL'] > 1]
        print(dup_url_pd['URL'].to_list())
        print(len(dup_url_pd['URL'].to_list()))
    else:
        print('No duplicated rows in df.')
```

```
find_dup_url(women_sale_copy)
```

There are duplicated rows in df. ['https://shop.lululemon.com/p/womens-outerwear/All-Yours-Zip-Hoodie-MD/_/prod9370338?color=0001', 'https://shop.lululemon.com/p/women/Open-Mind-Wrap-MD/_/prod10080177?color=30437', 'https://shop.lululemon.com/p/womens-outerwear/All-Yours-Zip-Hoodie-MD/_/prod9370338?color=43608', 'https://shop.lululemon.com/p/womens-outerwear/All-Yours-Zip-Hoodie-MD/_/prod9370338?color=43948', 'https://shop.lululemon.com/p/women/Open-Mind-Wrap-MD/_/prod10080177?color=36756']

Checking the data – Drop Duplicates

```
def get_uni(df):
    copy_df = df.copy()
    new_df = copy_df.drop_duplicates()

num_row_df = new_df.shape[0]
    uni_series = new_df.nunique()
    uni_df = pd.DataFrame({'Column_Name': uni_series.index, '#_Unique': uni_series.values})

num_uni_url = uni_df.loc[5, '#_Unique']
if num_row_df > num_uni_url:
    print('Still duplicates')
else:
    print('no duplicates')
return new_df
```

```
get uni(women sale copy)
no duplicates
                                             Display
         Unique ID Category
                                                                                                                                               #_Sizes_In_St
                                                             Color
                                                                                                                                         oos
                              Category
                                              Name
                                                                                                            Not
                                                                                                                                  Stock
                                          On the Fly
                                                                                                                                 '0', '2',
                                           Wide-Leg
                                                                        https://shop.lululemon.com/p/womens-
                                                                                                                                 '4', '6',
                                                                                                                                         '12',
                     Womens Trousers
   0 us_107385321
                                                          True Navy
                                                                                                              1 79.0
                                                                                                                          118.0
                                            7/8 Pant
                                                                                               trousers/O...
                                             *Woven
```

Add-On Features

- Number of sizes that are in stock
- Number of sizes that are out of stock
- Percentage of the discount: (original price sale price)/original price * 100

Add-On Features

```
In [459]:

def add_feature_sale(dataframe_sale):
    copy_df = dataframe_sale.copy()
    copy_df['#_Sizes_In_Stock'] = copy_df['Sizes In Stock'].apply(lambda x: len(x.split(',')))
    copy_df['#_Sizes_00S'] = copy_df['Sizes 00S'].apply(lambda x: 0 if len(x) == 0 else len(x.split(',')))
    copy_df['Sale Price'] = copy_df['Sale Price'].str.replace('$', '').astype(float)
    copy_df['Original Price'] = copy_df['Original Price'].str.replace('$', '').astype(float)
    copy_df['Discount%'] = round(((copy_df['Original Price'] - copy_df['Sale Price'])/copy_df['Original Price'])*100,2)
    return copy_df
```

```
def add_feature_origin(dataframe_origin):
    copy_df = dataframe_origin.copy()
    copy_df['#_Sizes_In_Stock'] = copy_df['Sizes In Stock'].apply(lambda x: len(x.split(',')))
    copy_df['#_Sizes_OOS'] = copy_df['Sizes OOS'].apply(lambda x: 0 if len(x) == 0 else len(x.split(',')))

    copy_df['Sale Price'] = copy_df['Sale Price'].replace('NA', 0)
    copy_df['Original Price'] = copy_df['Original Price'].apply(lambda x: float(x[0]))
    copy_df['Discount%'] = 0
    return copy_df
```

Future Work

- Find a better identification ID
- What would be a better way to store color and sizes?
- In-depth analysis



Reference and Acknowledgement

- https://shop.lululemon.com/
- https://en.wikipedia.org/wiki/Lululemon_Athletica
- https://www.json.cn/
- https://www.reddit.com/r/lululemon/comments/iz057l/i_created_a_lululemon_scraper_that_outputs_a/
- https://docs.google.com/spreadsheets/d/15EJLz_LOzGfGS69PZvCfvTSuUXex0AAS3Q8pEh996r M/edit#gid=1243014411
- https://github.com/mmdo7/lululemon-wmtm/blob/master/scraper.py
- https://github.com/stellahkim93/webscraping_lululemon/blob/master/product_info_selenium.p
- https://github.com/aromeeo/web-scraping-lululemon/blob/master/lulu_scraper.py
- Rui An: https://www.linkedin.com/in/an-rui-402ba8191/
 - Tianqiao Liu: https://www.linkedin.com/in/liutianqiao18/