

```
In [1]: #Python Imports
import os
import sys
import csv
import json
import time
import itertools
import numpy as np
import pandas as pd
from fuzzywuzzy import fuzz
from fuzzywuzzy import process
from selenium import webdriver
from IPython.display import Image
from selenium.webdriver.common.by import By
from selenium.webdriver.common.keys import Keys
from selenium.webdriver.chrome.options import Options
from selenium.webdriver.chrome.service import Service
from selenium.webdriver.support.ui import WebDriverWait
from selenium.webdriver.support import expected_conditions as EC
```

```
In [2]: from fuzzywuzzy import process

# Define the list to search through
my_list = ['apple', 'banana', 'orange', 'grape', 'watermelon']

# Define the search string
search_string = 'Mark Cross, Informaton Specialist'

# Use FuzzyWuzzy to find the most similar string in the list
best_match = process.extractOne(search_string, my_list)

# Print the best match
print(f"The best match for '{search_string}' is '{best_match[0]}' with a score of {best_match[1]}")

The best match for 'aple' is 'apple' with a score of 89
```

```
In [28]: true_string = "Mark Cross Women's Handbags"
string1 = 'Mark Cross, Information Specialist'
string2 = 'Designer Handbags for Women – Womens Designer Fashion'

similarity_score = fuzz.token_set_ratio(true_string, string1)
similarity_score2 = fuzz.token_set_ratio(true_string, string2)

print(f"The token set score between '{true_string}' and '{string1}' \nis {similarity_score}")
print(f"The token set score between '{true_string}' and '{string2}' \nis {similarity_score2}")

The token set score between 'Mark Cross Women's Handbags' and 'Mark Cross, Information Specialist'
is 54
The token set score between 'Mark Cross Women's Handbags' and 'Designer Handbags for Women – Womens Designer Fashion'
is 68
```

```
In [3]: str(best_match[0])
```

Out [3]: 'apple'

```
In [4]: roc_skincare = pd.read_csv('roc_skincare_mega.csv')
ulla_johnson = pd.read_csv('ulla_johnson_mega.csv')
hurley = pd.read_csv('hurley_mega.csv')
le_specs = pd.read_csv('le_specs_mega.csv')
jl_audio = pd.read_csv('jl_audio_mega.csv')
obey_clothing = pd.read_csv('obey_clothing_mega.csv')
mark_cross = pd.read_csv('mark_cross_mega.csv')
```

```
In [5]: result1 = pd.concat([roc_skincare, ulla_johnson, hurley,
                             le_specs, jl_audio, obey_clothing, mark_cross], axis=0)
```

```
In [48]: filtered = result1[
            #(result1['search_term_raw'] == 'ulla johnson') &
            #(result1['zipcode'] == 19104) #&
            #(result1['search_engine'] == 'store_locator')
            #]

filtered['search_engine'].unique()
```

Out [48]: array(['bing', 'yahoo', 'mojeek', 'startpage', 'duckduckgo',  
'yellow\_pages', 'google\_search', 'google\_shopping',  
'store\_locator'], dtype=object)

```
In [49]: filtered_ddd = result1[
            (result1['search_term_raw'] == 'ulla johnson') &
            (result1['zipcode'] == 19104) &
            (result1['search_engine'] == 'bing')
        ]
```

```
In [58]: filtered_ddd[['title', 'page_rank', 'extractTop_name', 'is_stockist_store']]
```

Out [58]:

	title	page_rank	extractTop_name	is_stockist_store
<b>1234</b>	Ulla Johnson Store Locator   Ulla Johnson	1	ulla johnson Stockist Store Locator	1.0
<b>1235</b>	Ulla Johnson - Luxury Women's Fashion	2	ulla johnson Stockist Store Locator	0.0
<b>1236</b>	Ulla Johnson   NET-A-PORTER	3	ulla johnson Stockist Store Locator	0.0
<b>1237</b>	Ulla Johnson - Luxury Women's Fashion	4	ulla johnson Stockist Store Locator	0.0
<b>1238</b>	Inside Fashion Designer Ulla Johnson's Bohemia...	5	ulla johnson Stockist Store Locator	1.0
<b>1239</b>	Shop Ulla Johnson Online   Nordstrom	6	Warm Shop	1.0
<b>1240</b>	Women's Ulla Johnson Clothing   Nordstrom	7	Nordstrom Short Hills	1.0

```
In [32]: filtered_ddd
```

```
Out[32]: Unnamed: 0 title url page_rank zipcode search_engine search_term address search_t
```

---

```
In [6]: result1['search_term_raw'] = result1.apply(lambda row: row['search_term'] if
```

```
In [7]: result2 = result1[
        (result1['search_term_raw'] == str('ulla johnson')) &
        (result1['zipcode'] == int(94123))
        #(result1['search_engine'] == 'store_locator')
    ]
```

```
In [47]: result2
```

```
Out[47]: Unnamed: 0 title url page_rank zipcode search_engine search_term address search_t
```

---

```
In [9]: result1
```

Out [9] :

	Unnamed: 0	title	url	page_rank	zipcc
0	0	Store Locator – RoC® Skincare	https://www.rocskincare.com/pages/store-locator	1	94'
1	1	RoC® Retinol Skincare   Clinically Proven – Ro...	https://www.bing.com/rebates/welcome?url=https...	2	94'
2	2	Roc Skincare Review - Must Read This Before Bu...	https://www.honestbrandreviews.com/reviews/roc...	3	94'
3	3	Roc Skincare Review - Must Read This Before Bu...	https://www.honestbrandreviews.com/reviews/roc...	4	94'
4	4	Roc Skincare Review - Must Read This Before Bu...	https://www.honestbrandreviews.com/reviews/roc...	5	94'
...	...	...	...	...	...
2252	39	Five Story	https://www.google.com/maps/dir/?api=1&destina...	2	191
2253	40	Jonathan Cohen	https://www.google.com/maps/dir/?api=1&destina...	3	191
2254	41	Julianne	https://www.google.com/maps/dir/?api=1&destina...	4	191

	Unnamed: 0	title	url	page_rank	zipco
2255	42	Capitol	https://www.google.com/maps/dir/?api=1&destina...	1	303
2256	43	Amazon	https://www.google.com/maps/dir/?api=1&destina...	1	981

16182 rows × 12 columns

```
In [10]: result1.to_csv("megabase.csv")
```

```
In [11]: result1['search_term_raw'].unique()
```

```
Out[11]: array(['roc skincare', nan, 'ulla johnson', 'hurley clothing', 'le specs',
                'jl audio', 'obey clothing', 'mark cross'], dtype=object)
```

```
In [12]: best_match = 'ulla johnson'
result1['search_term_raw'] = result1['search_term_raw'].fillna('N/A')
filtered = result1[result1['search_term_raw'].str.contains(best_match)]
print(filtered)
```

	Unnamed: 0	title \
0	0	Ulla Johnson Store Locator   Ulla Johnson
1	1	Ulla Johnson - Luxury Women's Fashion
2	2	Shop Ulla Johnson Online   Nordstrom
3	3	Ulla Johnson   Designer   NET-A-PORTER
4	4	Ulla Johnson Store Locator   Ulla Johnson
...	...	...
1454	1454	JCPenney
1455	1455	JCPenney
1456	1456	Gap
1457	1457	Gap
1458	1458	Gap

	url	page_rank	zipcode
0	https://ullajohnson.com/pages/store-locator	1	94110
1	https://ullajohnson.com/	2	94110
2	https://www.nordstrom.com/brands/ulla-johnson-...	3	94110
3	https://www.bing.com/rebates/welcome?url=https...	4	94110
4	https://ullajohnson.com/pages/store-locator	1	94110
...	...	...	...
1454	https://maps.google.com/maps?daddr=18601+33rd+...	10	98109
1455	https://maps.google.com/maps?daddr=1249+Southc...	11	98109
1456	https://maps.google.com/maps?daddr=183+Bellevu...	12	98109
1457	https://maps.google.com/maps?daddr=7430+164th+...	13	98109
1458	https://maps.google.com/maps?daddr=4625+27th+A...	14	98109

	search_engine	search_term \
0	bing	ulla johnson in 94110
1	bing	ulla johnson in 94110
2	bing	ulla johnson in 94110
3	bing	ulla johnson in 94110
4	yahoo	ulla johnson in 94110
...	...	...
1454	google_shopping	ulla+johnson+in+98109
1455	google_shopping	ulla+johnson+in+98109
1456	google_shopping	ulla+johnson+in+98109
1457	google_shopping	ulla+johnson+in+98109
1458	google_shopping	ulla+johnson+in+98109

	address	search_term_raw \
0	NaN	ulla johnson
1	NaN	ulla johnson
2	NaN	ulla johnson
3	NaN	ulla johnson
4	NaN	ulla johnson
...	...	...
1454	18601 33rd Ave W, Lynnwood·13.9 miles	ulla johnson
1455	1249 Southcenter Mall, Tukwila·12.8 miles	ulla johnson
1456	183 Bellevue Square, 183, Bellevue·6.6 miles	ulla johnson
1457	7430 164th Ave NE, B115, Redmond·10.6 miles	ulla johnson
1458	4625 27th Ave NE, Ste #23, Seattle·2.9 miles	ulla johnson

	extractTop_name	extractTop_value	is_stockist_st
0	ulla johnson	Stockist Store Locator	97.5

1.0		
1	ulla johnson Stockist Store Locator	58.5
0.0		
2	ulla johnson Stockist Store Locator	62.5
0.0		
3	ulla johnson Stockist Store Locator	62.5
0.0		
4	ulla johnson Stockist Store Locator	97.5
1.0		
...	...	...
...		
1454	ulla johnson Stockist Store Locator	26.5
0.0		
1455	ulla johnson Stockist Store Locator	26.5
0.0		
1456	Alhambra	24.0
0.0		
1457	Alhambra	24.0
0.0		
1458	Alhambra	24.0
0.0		

[1459 rows x 12 columns]

```
In [13]: result1
```

Out[13]:

	Unnamed: 0	title	url	page_rank	zipcc
0	0	Store Locator – RoC® Skincare	<a href="https://www.rocskincare.com/pages/store-locator">https://www.rocskincare.com/pages/store-locator</a>	1	94'
1	1	RoC® Retinol Skincare   Clinically Proven – Ro...	<a href="https://www.bing.com/rebates/welcome?url=https...">https://www.bing.com/rebates/welcome?url=https...</a>	2	94'
2	2	Roc Skincare Review - Must Read This Before Bu...	<a href="https://www.honestbrandreviews.com/reviews/roc...">https://www.honestbrandreviews.com/reviews/roc...</a>	3	94'
3	3	Roc Skincare Review - Must Read This Before Bu...	<a href="https://www.honestbrandreviews.com/reviews/roc...">https://www.honestbrandreviews.com/reviews/roc...</a>	4	94'
4	4	Roc Skincare Review - Must Read This Before Bu...	<a href="https://www.honestbrandreviews.com/reviews/roc...">https://www.honestbrandreviews.com/reviews/roc...</a>	5	94'
...	...	...	...	...	...
2252	39	Five Story	<a href="https://www.google.com/maps/dir/?api=1&amp;destina...">https://www.google.com/maps/dir/?api=1&amp;destina...</a>	2	191
2253	40	Jonathan Cohen	<a href="https://www.google.com/maps/dir/?api=1&amp;destina...">https://www.google.com/maps/dir/?api=1&amp;destina...</a>	3	191
2254	41	Julianne	<a href="https://www.google.com/maps/dir/?api=1&amp;destina...">https://www.google.com/maps/dir/?api=1&amp;destina...</a>	4	191



	Unnamed: 0	title	url	page_rank	zipco
2255	42	Capitol	https://www.google.com/maps/dir/?api=1&destina...	1	303
2256	43	Amazon	https://www.google.com/maps/dir/?api=1&destina...	1	981

16182 rows x 12 columns

```
In [14]: result1['search_term_raw'].unique()
```

```
Out[14]: array(['roc skincare', 'N/A', 'ulla johnson', 'hurley clothing',  
                'le specs', 'j1 audio', 'obey clothing', 'mark cross'],  
              dtype=object)
```