

Web Scraping Perfume Data: A Comprehensive Project Proposal

****Introduction****

In the world of e-commerce, data is king. Understanding consumer behavior, product trends, and competitor strategies is crucial for businesses to stay ahead of the curve. Web scraping, a technique for extracting data from websites, offers a powerful tool for gathering valuable insights from the vast ocean of online information.

This project proposes the utilization of web scraping to gather comprehensive data on perfumes, a popular and diverse product category. The data collected will be analyzed to gain valuable insights into perfume trends, consumer preferences, and market dynamics.



****Objectives****

The primary objectives of this project are as follows:

1. **Gather extensive perfume data:** Utilize web scraping techniques to extract data on various perfume attributes, including brand, name, description, price, ingredients, reviews, and ratings.
2. **Analyze fragrance profiles:** Employ data analysis techniques to identify common fragrance notes, popular fragrance families, and trending scent profiles.
3. **Understand consumer preferences:** Analyze consumer reviews and ratings to understand consumer preferences for different perfume attributes, such as fragrance intensity, longevity, and sillage.
4. **Assess market trends:** Analyze perfume sales data and industry reports to identify emerging trends in the perfume market, including popular brands, price segments, and consumer preferences.

****Data Sources****

The project will target reputable e-commerce websites and online perfume retailers, such as:

1. **Fragrantica**: comprehensive fragrance database with user reviews and ratings.
2. **FragranceNet**: A leading online perfume retailer with a wide selection of brands and products.
3. **Sephora**: A global cosmetics and fragrance retailer with a strong online presence.
4. **Ultra Beauty**: A leading American beauty retailer with a wide variety of perfumes.
5. **Amazon**: The world's largest online marketplace with a vast selection of perfumes.
6. **Snapdeal**: online market place

****Web Scraping Methodology****

The project will employ web scraping libraries such as BeautifulSoup (Python) or Scrappy (Python) to extract data from the target websites. These libraries provide efficient tools for parsing HTML content, identifying relevant data elements, and extracting structured information.

****Data Analysis****

The extracted perfume data will be analyzed using various techniques, including:

1. **Descriptive statistics**: To summarize the characteristics of the perfume data, such as average price, most popular brands, and common fragrance notes.
2. **Text analysis**: To analyze consumer reviews and identify common themes, sentiment, and keywords.
3. **Visualization**: To create charts and graphs to visualize trends, patterns, and relationships within the data.

****Deliverables****

The project will deliver the following deliverables:

1. **A comprehensive perfume dataset**: A structured dataset containing detailed information on various perfumes, including brand, name, description, price, ingredients, reviews, and ratings.
2. **An analysis of fragrance profiles**: A report summarizing the most common fragrance notes, popular fragrance families, and trending scent profiles.
3. **Insights into consumer preferences**: A report highlighting consumer preferences for different perfume attributes, such as fragrance intensity, longevity, and sillage.
4. **A market trends assessment**: A report identifying emerging trends in the perfume market, including popular brands, price segments, and consumer preferences.



Important Libraries for Project :-

Python has several web scrapping libraries. We will use the following libraries:

- **Requests :-** for making HTTP requests to website
- **BeautifulSoup :-** for parsing the HTML code
- **Pandas :-** for storing the scraped data in data frame

Project Outline :-

From following website we will grab following information :-

- Name of the product.
- Price of the product.
- Different colours options for the product.
- Customer names.
- Customer Ratings (☆)
- Customer reviews.

Benefits of Data :-

- Access to valuable data for analysis or research.
- Automation of data collection, saving time and effort.
- Stay up to date with change changes on the target websites.

Exporting of extracted data :- We'll export the data as a CSV file. We will use the pandas library. We'll use the pandas library.

Access perfume website by using import requests URL and BeautifulSoup library.

```
In [15]: import requests
         from bs4 import BeautifulSoup
         import pandas as pd
```

```
In [16]: URL = "https://www.snapdeal.com/search?keyword=perfume%20men&sort=r1vncy"
```

```
In [17]: r = requests.get(URL)
htmlcontent=r.content
htmlcontent
```

```
Out[17]: b'<!DOCTYPE html>\n<!--[if IE 8]><html lang="en" class="ie ie8 lt-ie9 lt-ie10"> <![endif-->\n<!--[if IE 9]><html lang="en" class="ie ie9 lt-ie10"> <![endif-->\n<!--[if IE]><html lang="en" class="ie"><![endif-->\n<!--[if gt IE 9]>!--><html lang="en"><!--<![endif-->\n<t<head prefix="og: https://ogp.me/ns# fb: https://ogp.me/ns/fb# snapdeallog: https://ogp.me/ns/fb/snapdeallog#>\n<t<link rel="dns-prefetch" href="https://i1.sdcdn.com" />\n<t<link rel="dns-prefetch" href="https://i2.sdcdn.com" />\n<t<link rel="dns-prefetch" href="https://i3.sdcdn.com" />\n<t<link rel="dns-prefetch" href="https://i4.sdcdn.com" />\n<t<link rel="dns-prefetch" href="https://n1.sdcdn.com" />\n<t<link rel="dns-prefetch" href="https://n2.sdcdn.com" />\n<t<link rel="dns-prefetch" href="https://n3.sdcdn.com" />\n<t<link rel="dns-prefetch" href="https://n4.sdcdn.com" />\n<t<link rel="dns-prefetch" href="https://sa.snapdeal.com" />\n<t<link rel="dns-prefetch" href="https://search-suggester.snapdeal.com" />\n<t<link rel="dns-prefetch" href="https://mobileapi.snapdeal.com" />\n<t<link rel="manifest" href="/manifest.json">\n<t<link rel="preload" href="https://i4.sdcdn.com/js/1639746993560/snap/base.jquery111.min.js" crossorigin="anonym
```

Parse the HTML link.

```
In [18]: soup = BeautifulSoup(htmlcontent, "html.parser")
         print(soup)
         soup.prettify
```

```
<!DOCTYPE html>

<!--[if IE 8]><html lang="en" class="ie ie8 lt-ie9 lt-ie10"> <![endif]-->
<!--[if IE 9]><html lang="en" class="ie ie9 lt-ie10"> <![endif]-->
<!--[if IE 10]><html lang="en" class="ie"> <![endif]-->
<!--[if gt IE 9]><!--><html lang="en"><!--<![endif]-->

<head prefix="og: https://ogp.me/ns# fb: https://ogp.me/ns/fb# snapdeallog: https://ogp.me/ns/fb/snapdeallog#">
<link href="https://i1.sdcdn.com" rel="dns-prefetch"/>
<link href="https://i2.sdcdn.com" rel="dns-prefetch"/>
<link href="https://i3.sdcdn.com" rel="dns-prefetch"/>
<link href="https://i4.sdcdn.com" rel="dns-prefetch"/>
<link href="https://n1.sdcdn.com" rel="dns-prefetch"/>
<link href="https://n2.sdcdn.com" rel="dns-prefetch"/>
<link href="https://n3.sdcdn.com" rel="dns-prefetch"/>
<link href="https://n4.sdcdn.com" rel="dns-prefetch"/>
<link href="https://sa.snapdeal.com" rel="dns-prefetch"/>
<link href="https://search-suggester.snapdeal.com" rel="dns-prefetch"/>
<link href="https://mobileapi.snapdeal.com" rel="dns-prefetch"/>
<link href="/manifest.json" rel="manifest"/>
```

Access Perfume Name using this source code

```
per_name = soup.find_all(class_="product-title")
print(per_name)
```

```
[<p class="product-title" title="Denver - Deodorant Deodorant Spray & Perfume for Men 50 ml ( Pack of 4 )">Denver - Deodora
nt Deodorant Spray & Perfume for Men 50 ml ( Pack of 4 )</p>, <p class="product-title" title="Wild Stone Edge and Forest Sp
ice Perfume Combo for Men Eau de Parfum - 60 ml (For Men)">Wild Stone Edge and Forest Spice Perfume Combo for Men Eau de Parfum
- 60 ml (For Men)</p>, <p class="product-title" title="NORD - Gunther Perfume Body Spray for Men 120 ml ( Pack of 1 )">NORD - G
unther Perfume Body Spray for Men 120 ml ( Pack of 1 )</p>, <p class="product-title" title="INDRA SUGANDH BHANDAR Attar For Men
|Women|Pujan Shahi Kesar Chandan Pure and Original Perfume 24 Hours Long Lasting Fragrance 15ml Rollon Cubic Fancy Pack">INDRA
SUGANDH BHANDAR Attar For Men|Women|Pujan Shahi Kesar Chandan Pure and Original Perfume 24 Hours Long Lasting Fragrance 15ml Ro
llon Cubic Fancy Pack</p>, <p class="product-title" title="St. John - Cobra & Black Current Perfume For Men 10ml Each Eau
De Parfum (EDP) For Men 10ML ( Pack of 2 )">St. John - Cobra & Black Current Perfume For Men 10ml Each Eau De Parfum (EDP)
For Men 10ML ( Pack of 2 )</p>, <p class="product-title" title="LA FRENCH -Oudh Perfume Gift Set (4x20ml) Eau De Parfum (EDP) F
or Men,Women 80 ml ( Pack of 1 )">LA FRENCH -Oudh Perfume Gift Set (4x20ml) Eau De Parfum (EDP) For Men,Women 80 ml ( Pack of 1
)</p>, <p class="product-title" title="NORD Luxury Perfume Gift Set (Pride, Lust, Wrath, Greed) for Men, Pack of 4 (20 ml eac
h)">NORD Luxury Perfume Gift Set (Pride, Lust, Wrath, Greed) for Men, Pack of 4 (20 ml each)</p>, <p class="product-title" titl
e="NORD - Beast Perfume Body Spray for Men 120 ml ( Pack of 1 )">NORD - Beast Perfume Body Spray for Men 120 ml ( Pack of 1 )</
p>, <p class="product-title" title="NORD - Rage Perfume Body Spray for Men 120 ml ( Pack of 1 )">NORD - Rage Perfume Body Spray
for Men 120 ml ( Pack of 1 )</p>, <p class="product-title" title="NORD - Rage, Gunther, Bolt, Beast Perfume Body Spray for Men
120 ml Each ( Pack of 4 )">NORD - Rage, Gunther, Bolt, Beast Perfume Body Spray for Men 120 ml Each ( Pack of 4 )</p>, <p class
="product-title" title="St. John | Black Berry Perfume | 100 ml | Pack Of 1 | Eau de Parfum - 100 ml (For Men)">St. John | Blac
k Berry Perfume | 100 ml | Pack Of 1 | Eau de Parfum - 100 ml (For Men)</p>, <p class="product-title" title="Ustraa After Dark
Cologne - 100 ml - Perfume for Men">Ustraa After Dark Cologne - 100 ml - Perfume for Men</p>]
```

ss/snapdeal_perfume_project_final/invn

```
per_name_list = []
for i in range(0,len(per_name)):
    per_name_list.append(per_name[i].get_text())
print(per_name_list)
```

```
['Denver - Deodorant Deodorant Spray & Perfume for Men 50 ml ( Pack of 4 )', 'Wild Stone Edge and Forest Spice Perfume Combo fo
r Men Eau de Parfum - 60 ml (For Men)', 'NORD - Gunther Perfume Body Spray for Men 120 ml ( Pack of 1 )', 'INDRA SUGANDH BHANDA
R Attar For Men|Women|Pujan Shahi Kesar Chandan Pure and Original Perfume 24 Hours Long Lasting Fragrance 15ml Rollon Cubic Fan
cy Pack', 'St. John - Cobra & Black Current Perfume For Men 10ml Each Eau De Parfum (EDP) For Men 10ML ( Pack of 2 )', 'LA FRE
NCH -Oudh Perfume Gift Set (4x20ml) Eau De Parfum (EDP) For Men,Women 80 ml ( Pack of 1 )', 'NORD Luxury Perfume Gift Set (Prid
e, Lust, Wrath, Greed) for Men, Pack of 4 (20 ml each)', 'NORD - Beast Perfume Body Spray for Men 120 ml ( Pack of 1 )', 'NORD
- Rage Perfume Body Spray for Men 120 ml ( Pack of 1 )', 'NORD - Rage, Gunther, Bolt, Beast Perfume Body Spray for Men 120 ml E
ach ( Pack of 4 )', 'St. John | Black Berry Perfume | 100 ml | Pack Of 1 | Eau de Parfum - 100 ml (For Men)', 'Ustraa After Dar
k Cologne - 100 ml - Perfume for Men', 'Wild Stone Intense Black No Gas Body Perfume, Pack of 2 Perfume Body Spray - For Men (2
40 ml, Pack of 2)', 'Bella Vita Organic Unisex Luxury Perfume Gift Set 4x20 ML (SKAI, FRESH, WHITEOUD, PATCHOULI) for Men & Wom
en (Pack of 4)', 'Ajmal Al-Wafi Concentrated Perfume Oil 10ml Attar for Men & Women', 'Wild Stone Iron Pack of 3 Perfume Body S
pray - For Men (360 ml, Pack of 3)', 'INDRA SUGANDH BHANDAR Attar For Men|Women|Pujan Shahi White Musk Perfume Mild Kasturi 24
Hours Long Lasting Fragrance 15ml Rollon Cubic Fancy Pack', 'Ajmal Wisal Dhabab Perfume Deodorant 200ml Body Spray Gift For me
n', 'SAFIRF DAMASK ROSE & JASMTNF SAMBRAC ATTAR (COMBO PACK 6ML*2) ROLL-ON PERFUME OIL FOR MEN AND WOMEN', 'Archies - Black is B
```

Access Perfume Price using this source code.

```
price = soup.find_all(class_="lfloat product-price")
print(price)
```

```
[<span class="lfloat product-price" data-price="213" display-price="213" id="display-price-654657119016">Rs. 213</span>, <span
class="lfloat product-price" data-price="374" display-price="374" id="display-price-669740050412">Rs. 374</span>, <span class
="lfloat product-price" data-price="179" display-price="179" id="display-price-640691703101">Rs. 179</span>, <span class="lflo
at product-price" data-price="289" display-price="289" id="display-price-625824621124">Rs. 289</span>, <span class="lfloat pro
duct-price" data-price="110" display-price="110" id="display-price-638247118654">Rs. 110</span>, <span class="lfloat product-p
rice" data-price="419" display-price="419" id="display-price-633683980418">Rs. 419</span>, <span class="lfloat product-price"
data-price="359" display-price="359" id="display-price-662760353501">Rs. 359</span>, <span class="lfloat product-price" data-p
rices="179" display-price="179" id="display-price-625854280343">Rs. 179</span>, <span class="lfloat product-price" data-price
="179" display-price="179" id="display-price-635015178843">Rs. 179</span>, <span class="lfloat product-price" data-price="359"
display-price="359" id="display-price-623310609552">Rs. 359</span>, <span class="lfloat product-price" data-price="259" displa
y-price="259" id="display-price-621942038808">Rs. 259</span>, <span class="lfloat product-price" data-price="521" display-pric
e="521" id="display-price-635641710056">Rs. 521</span>, <span class="lfloat product-price" data-price="273" display-price="27
3" id="display-price-683390464309">Rs. 273</span>, <span class="lfloat product-price" data-price="565" display-price="565" id
="display-price-653209608295">Rs. 565</span>, <span class="lfloat product-price" data-price="279" display-price="279" id="disp
lay-price-670580735018">Rs. 279</span>, <span class="lfloat product-price" data-price="338" display-price="338" id="display-pr
ice-686039599550">Rs. 338</span>, <span class="lfloat product-price" data-price="419" display-price="419" id="display-price-65
4774551787">Rs. 419</span>, <span class="lfloat product-price" data-price="149" display-price="149" id="display-price-64463458
```

```

price_list = []
price_int=[]
for i in range(0,len(price)):
    price_list.append(price[i].get_text())
    p1=price_list[i][5:]
    price_int.append(p1)
print(price_int)

```

```

['213', '374', '179', '289', '110', '419', '359', '179', '179', '359', '259', '521', '273', '565', '279', '338', '419', '149', '140', '309']

```

Access Perfume Ratings.

```

reviews = soup.find_all(class_="product-rating-count")
print(reviews)

```

```

[<p class="product-rating-count">(3391)</p>, <p class="product-rating-count">(49)</p>, <p class="product-rating-count">(62)</p>, <p class="product-rating-count">(81)</p>, <p class="product-rating-count">(8)</p>, <p class="product-rating-count">(37)</p>, <p class="product-rating-count">(459)</p>, <p class="product-rating-count">(176)</p>, <p class="product-rating-count">(218)</p>, <p class="product-rating-count">(411)</p>, <p class="product-rating-count">(251)</p>, <p class="product-rating-count">(72)</p>, <p class="product-rating-count">(34)</p>, <p class="product-rating-count">(413)</p>, <p class="product-rating-count">(124)</p>, <p class="product-rating-count">(55)</p>, <p class="product-rating-count">(17)</p>, <p class="product-rating-count">(1466)</p>, <p class="product-rating-count">(3)</p>, <p class="product-rating-count">(11)</p>]

```

```
reviews_list = []
for i in range(0,len(reviews)):
    reviews_list.append(reviews[i].get_text())
print(reviews_list)
```

```
['(3391)', '(49)', '(62)', '(81)', '(8)', '(37)', '(459)', '(176)', '(218)', '(411)', '(251)', '(72)', '(34)', '(413)', '(124)', '(55)', '(17)', '(1466)', '(3)', '(11)']
```

```
final_review_list = []
reviews_int=[]
for i in range(0,len(reviews_list)):
    final_review_list.append(reviews_list[i])
    p1=final_review_list[i][1:-1]
    reviews_int.append(p1)
print(reviews_int)
```

```
['3391', '49', '62', '81', '8', '37', '459', '176', '218', '411', '251', '72', '34', '413', '124', '55', '17', '1466', '3', '11']
```

Access data frame.

```
import pandas as pd
df = pd.DataFrame({"Perfume_name":per_name_list,"Perfume_price":price_int,"Reviews":reviews_int})
print(df)
```

	Perfume_name	Perfume_price	Reviews
0	Denver - Deodorant Deodorant Spray & Perfume f...	213	3391
1	Wild Stone Edge and Forest Spice Perfume Combo...	374	49
2	NORD - Gunther Perfume Body Spray for Men 120 ...	179	62
3	INDRA SUGANDH BHANDAR Attar For Men Women Puja...	289	81
4	St. John - Cobra & Black Current Perfume For ...	110	8
5	LA FRENCH -Oudh Perfume Gift Set (4x20ml) Eau ...	419	37
6	NORD Luxury Perfume Gift Set (Pride, Lust, Wra...	359	459
7	NORD - Beast Perfume Body Spray for Men 120 ml...	179	176
8	NORD - Rage Perfume Body Spray for Men 120 ml ...	179	218
9	NORD - Rage, Gunther, Bolt, Beast Perfume Body...	359	411
10	St. John Black Berry Perfume 100 ml Pack...	259	251
11	Ustraa After Dark Cologne - 100 ml - Perfume f...	521	72
12	Wild Stone Intense Black No Gas Body Perfume, ...	273	34
13	Bella Vita Organic Unisex Luxury Perfume Gift ...	565	413

Read csv file.

```
import os
os.getcwd()
df.to_csv("snapdeal_perfume.csv",index=False)
df1 = pd.read_csv("snapdeal_perfume.csv")
df1
```

	Perfume_name	Perfume_price	Reviews
0	Denver - Deodorant Deodorant Spray & Perfume f...	213	3391
1	Wild Stone Edge and Forest Spice Perfume Combo...	374	49
2	NORD - Gunther Perfume Body Spray for Men 120 ...	179	62
3	INDRA SUGANDH BHANDAR Attar For Men Women Puja...	289	81
4	St. John - Cobra & Black Current Perfume For ...	110	8
5	LA FRENCH -Oudh Perfume Gift Set (4x20ml) Eau ...	419	37
6	NORD Luxury Perfume Gift Set (Pride, Lust, Wra...	359	459
7	NORD - Beast Perfume Body Spray for Men 120 ml...	179	176
8	NORD - Rage Perfume Body Spray for Men 120 ml ...	179	218

Conclusion

scraping offers a valuable tool for gathering insights into the perfume industry. By extracting and analyzing data from various online sources, this project will provide comprehensive insights into fragrance trends, consumer preferences, and market dynamics. This Web information can be utilized by businesses to make informed decisions regarding product development, marketing strategies, and customer engagement.

**THANK
YOU**