Aromatic Analytics

Decoding the DNA of best seller perfumes

Aditya Sindhavad | Adithya Murali | Advaith Shankar
Manasa Maganti | Shashank Rao | Varsha Manju Jayakumar
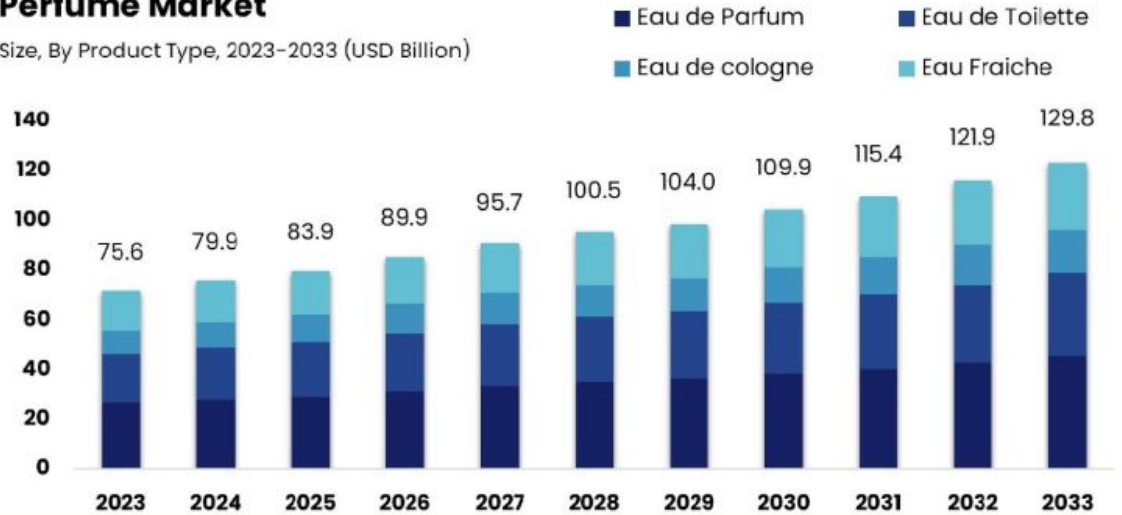
est. 1987

Predict best seller – but why is it important to solve?

- Believe it or not, a staggering 773 million people worldwide now rely on online platforms to find their perfect scent.
- Websites can position their product rightly that matches with user's notes and preferences

Perfume Market
Size, By Product Type, 2023–2033 (USD Billion)

■ Eau de Parfum ■ Eau de Toilette
■ Eau de cologne ■ Eau Fraiche

2023: 75.6
2024: 79.9
2025: 83.9
2026: 89.9
2027: 95.7
2028: 100.5
2029: 104.0
2030: 109.9
2031: 115.4
2032: 121.9
2033: 129.8

Data Sources & Scraping info
The John Wick moment
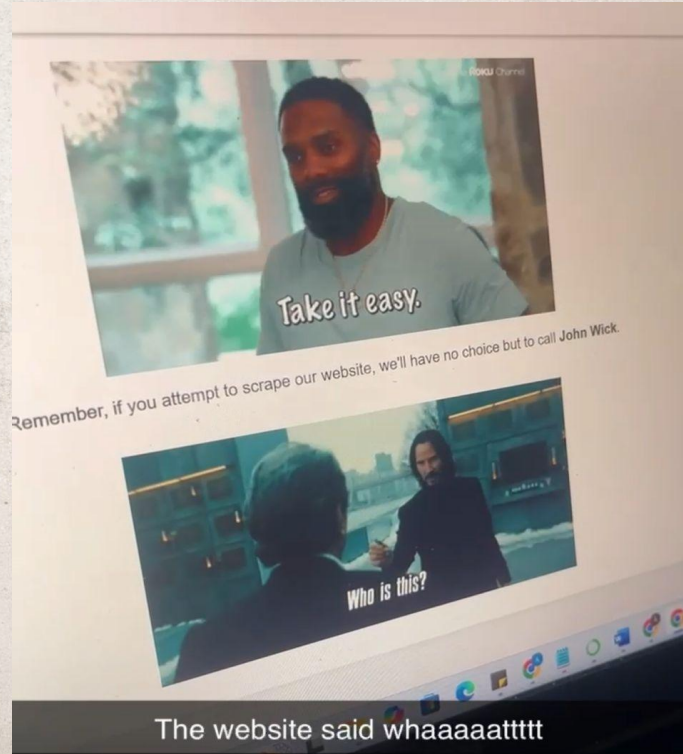
Scraped user reviews for -
- Train set: top 60 best seller and bottom 60 perfumes (identifying the key attributes using their user reviews)
- Test set: 30 newly released perfumes (predicting which of these would be best sellers)
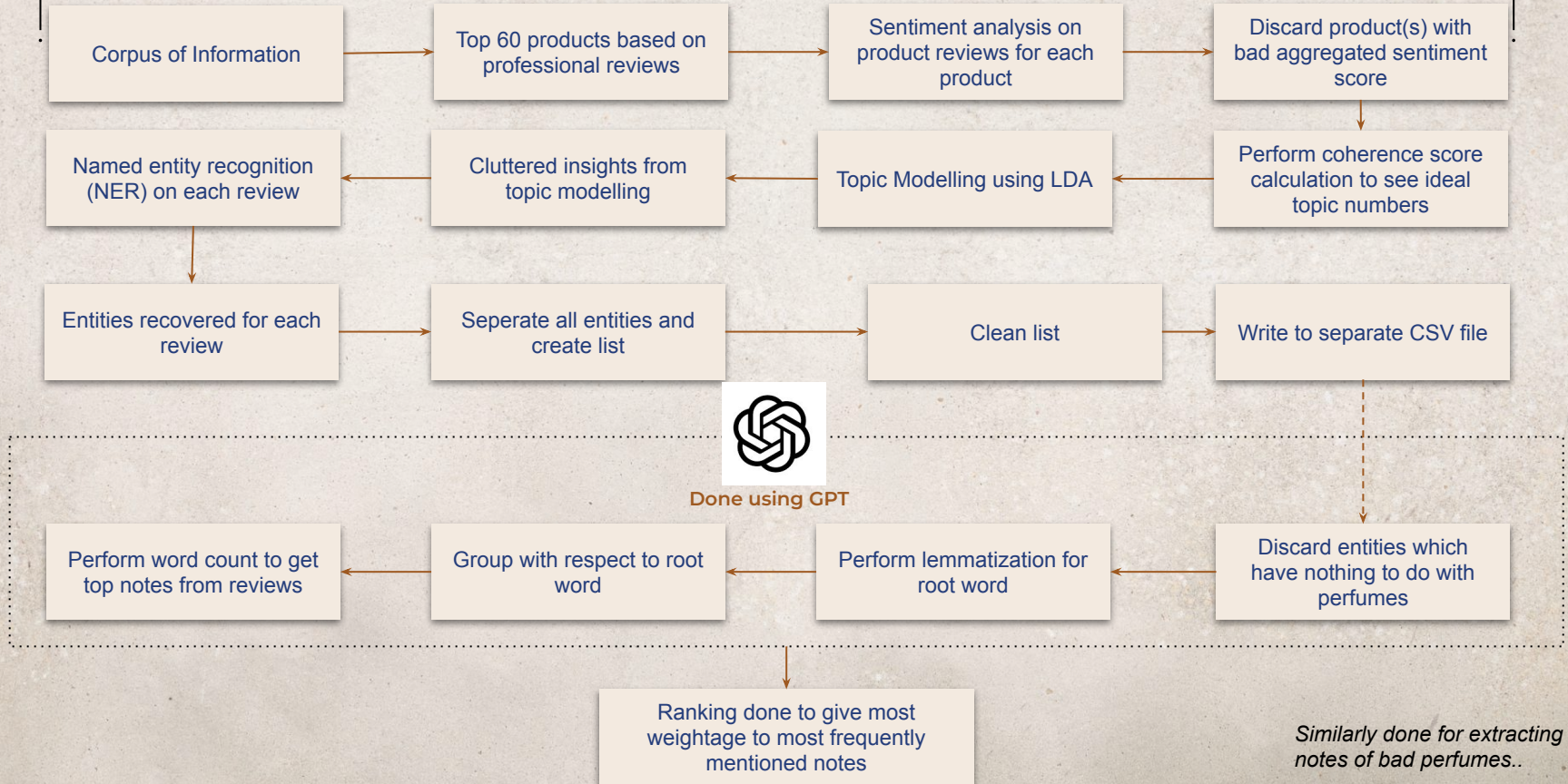

429 Too Many Requests

Wow, you're quite the enthusiast! Unfortunately, our server can't handle the high volume of requests. Moreover, we need to protect our website from malicious activities.

It looks like you've opened more pages in a short time than one can possibly read. If you're a regular user, please take a break, explore our other interesting content, and come back later.

If you're crawling our website, please note that it's against our terms of service. Scraping and stealing proprietary database information is illegal, so kindly cease such actions. Fragrantica content is only allowed for private browsing purposes.

Take it easy.

Remember, if you attempt to scrape our website, we'll have no choice but to call John Wick.

Who is this?

The website said whaaaaattttt

Analytics on the data – Pulling out notes of a perfume

Corpus of Information → Top 60 products based on professional reviews → Sentiment analysis on product reviews for each product → Discard product(s) with bad aggregated sentiment score

Named entity recognition (NER) on each review ← Cluttered insights from topic modelling ← Topic Modelling using LDA ← Perform coherence score calculation to see ideal topic numbers

Entities recovered for each review → Seperate all entities and create list → Clean list → Write to separate CSV file

Done using GPT

Perform word count to get top notes from reviews ← Group with respect to root word ← Perform lemmatization for root word ← Discard entities which have nothing to do with perfumes

Ranking done to give most weightage to most frequently mentioned notes

Similarly done for extracting notes of bad perfumes..

Vanilla is hated as much it is loved – here's why it makes "scents"

Vanilla in combination with feminine notes show a positive sentiment, and in combination with masculine notes shows negative sentiment

Bottom Notes

Notes	Count
amber	2681
musk	2520
orange	2488
bergamot	2411
jasmine	2387
rose	2371
vanilla	2252
sandalwood	2197
patchouli	2056
cedar	1921

Top Notes

Notes	Count
sweet	2282
vanilla	1746
intense	1161
wood	1064
leather	990
floral	945
tobacco	692
warm	677
fresh	654
citrus	650

Review of 9pm Afnan cologne - Men's Fragrance

"Bubblegum candy vanilla scent. Not suitable for 25+ guy. I really like it but I'm too old for it and don't want to smell like a kid. I wouldn't be able to take seriously a person wearing it."

Review of Coco Eau de Parfum Chanel - Women's Fragrance

"I have been wearing Coco since I was 23, it is timeless and doesn't represent an age, it exudes confidence, agency, and soft power. Unapologetic, two sprays last all day, and it envelops into a plush, warm, spicy vanilla. So feminine, sensual, intelligent, it is a perfume that represents success to me."

Ranking Fragrances: A Data-Driven Approach to Evaluation

Evaluation score = rank of positive accord - rank of negative accord + 0 for unknown accord

Final score = normalized_sentiment * 0.2 + normalized_rating * 0.3 + normalized_evaluation * 0.5

- Evaluation score is calculated based on the balance of positive and negative accords
- Final score is determined through a weighted combination of sentiment, rating, and evaluation score to ensure robust analysis

Fragrances are categorized from "Masterstroke" to "Not Fitting" based on their Eval score:

- Masterstroke
- Likely hit

- Fair set
- Low attribute match

- Not fitting

Title	Accords	Rating	Evaluation_Score	Evaluation_Score_Category	Sentiment_Score_x	Final_Score2
Pear Gelato Theodoros Kalotinis perfume - a new fragrance for women and men 2024	sweet,vanilla, fruity, lactonic, aquatic, fresh	4.02	18	likely hit	0.695663	0.846613
Asad Zanzibar Lattafa Perfumes cologne - a new fragrance for men 2024	Lavender, Black Pepper,Coconut Water, Iris, Sa...	3.76	8	fair set	0.459361	0.465825
Narcotic Delight Initio Parfums Prives perfume - a new fragrance for women and men 2024	Cherry, Pink Pepper, Black Pepper, Cognac, Hed...	4.20	6	fair set	0.615551	0.671871

Future work and scope

With additional data

Utilize more geographic, demographic, psychographic data to understand the sentiment associated with various attributes

Further project enhancements

Trend Forecasting: Incorporate global and regional market trends, seasonality, and user reviews to forecast which fragrance notes and compositions will dominate future best-seller lists

For example, a perfume brand noticed an uptick in positive sentiment towards floral notes on social media before launching a successful jasmine-based fragrance line

Sustainability and Ingredient Preferences: Expand the analysis to include consumer preferences for eco-friendly and sustainable ingredients, as there is a growing demand for transparency in the perfume industry. This can help brands align with ethical consumer trends.

Questions?