Predicting Commercial Success of Fragrances: A Machine Learning Approach Using Olfactory and Market Features

Domenico Copia
Matr. Num. 0001113950
University of Bologna
Digital Transformation Management
Machine Learning
Email: domenico.copia@studio.unibo.it

Abstract—**This project investigates whether the commercial success of fragrances can be predicted through machine learning techniques applied to both olfactory and market-related features. Using a dataset of over 24,000 perfumes from Fragrantica.com, the data were preprocessed and enriched with attributes capturing temporal trends, brand reputation, perfumer expertise, geographical context, and olfactory complexity. Success was operationalized through a composite Success Score that balances consumer-perceived quality (ratings) with popularity (number of reviews). Several predictive models were implemented and evaluated, with particular attention to ensemble methods such as Random Forest. Results indicate that predictive performance is moderate (R^2 0.39), but consistently highlight brand reputation, launch year, perfumer portfolio, and specific accords (e.g., woody, citrus) as the strongest predictors of success. These findings suggest that fragrance outcomes are shaped not only by intrinsic scent composition but also by extrinsic signals of credibility and visibility. Overall, the study demonstrates the potential of machine learning to uncover systematic drivers of success in the fragrance industry, while recognizing the limits imposed by external factors such as marketing and cultural trends. Beyond its immediate scope, the project illustrates how data-driven approaches can be applied to creative and experiential industries, bridging sensory analytics with predictive modeling.**

1. Introduction

Predicting the commercial success of consumer products has long been a central challenge in business, marketing, and data science. In industries such as music, film, and gaming, researchers have sought to anticipate which creative works will resonate with audiences. The fragrance industry is no exception: perfumes are launched every year by both established fashion houses and niche brands, yet only a fraction achieve lasting market success. Understanding which factors drive popularity and consumer approval is thus of great interest to both practitioners and scholars.

This project investigates whether the commercial success of a fragrance can be predicted based on its intrinsic olfactory properties and contextual market characteristics. Specifically, the analysis examines the extent to which features such as brand reputation, perfumer expertise, note complexity, and main accords can explain or forecast the reception of a fragrance among consumers. To address this research question, a machine learning system is built and evaluated, integrating data from Fragrantica, one of the largest publicly available fragrance databases, containing over 24,000 entries.

The methodology proceeds through several stages. First, the dataset is preprocessed and enriched by handling missing values, standardizing formats, and engineering additional features that capture temporal, brand-level, geographical, perfumer-related, and olfactory aspects of each perfume. A composite metric of Success Score is introduced to serve as the prediction target, balancing quality (user ratings) with popularity (number of reviews). This metric ensures that the model does not merely identify critically acclaimed fragrances but also accounts for their broader visibility and acceptance in the market.

Two complementary modeling perspectives are considered. In the regression setting, the task is framed as the prediction of the continuous Success Score. Several algorithms are tested, with ensemble methods such as Random Forest and Gradient Boosting providing the best results, although predictive power remains moderate (R^2 0.35–0.38). In parallel, the task is reformulated as a binary classification problem, where perfumes are categorized as high- or low-rated according to the median Rating Value (3.97). This alternative approach yields an accuracy of approximately 71% and a ROC-AUC of 0.78, confirming the ability of the models to distinguish between more and less appreciated fragrances. Importantly, both regression and classification highlight a consistent set of key predictors, including brand reputation, launch year, perfumer portfolio size, and certain accords such as woody and citrus.

The report is structured as follows. Section 2 reviews related work, highlighting previous attempts to model product success in creative industries and techniques applicable to fragrance analytics. Section 3 details the proposed method, covering data preparation, feature engineering, and model selection. Section 4 presents the results, evaluating performance metrics and extracting insights from feature importance rankings. Section 5 concludes by summarizing contributions, limitations, and directions for future research.

2. Related Work

The problem of predicting commercial success has been investigated across multiple domains, ranging from cultural products such as movies, music, and books to consumer goods and online services. In each case, researchers have sought to combine intrinsic attributes of the product with contextual or market-level indicators to anticipate future performance. For instance, in the film industry, studies have used features such as cast, genre, and production budget to forecast box office revenues, while in music analytics, audio features and streaming patterns have been linked to chart performance. These studies consistently highlight that success is a multifaceted construct shaped by both quality signals and market visibility. In the fragrance domain, academic contributions are relatively limited compared to other creative industries. However, previous work in sensory analytics and computational olfaction has laid the groundwork for structured analysis. Efforts to classify perfumes based on their notes and accords often draw on natural language processing and clustering methods, showing that consumer perception can be mapped into a structured olfactory space. Beyond classification, researchers have also applied recommender systems to perfume databases, suggesting fragrances to users based on similarity in scent profiles or user reviews. These approaches, while not explicitly predictive of success, demonstrate the feasibility of leveraging structured fragrance data for machine learning tasks. More broadly, techniques from predictive modeling and machine learning are directly applicable to this problem. Ensemble learning methods such as Random Forests and Gradient Boosting Machines have been shown to capture complex, non-linear interactions in tabular data, making them well-suited for integrating heterogeneous fragrance features (e.g., brand reputation, perfumer expertise, geographical origin, and olfactory composition). Similarly, feature engineering strategies, such as constructing differential or aggregated attributes, have been employed in consumer behavior studies to reduce dimensionality and emphasize relational patterns. Another relevant stream of work concerns popularity prediction in online platforms. Studies analyzing user-generated ratings, reviews, and social media engagement emphasize the importance of social proof and visibility as determinants of perceived success. This insight underpins the decision to define a composite Success Score, balancing intrinsic quality (average user rating) with market reach (number of ratings). Such approaches have parallels in recommendation systems where hybrid metrics are used to avoid bias toward either critically acclaimed but obscure products or popular but low-quality items. Taken together, the literature suggests that combining intrinsic product features (olfactory complexity, accords, perfumer expertise) with extrinsic contextual signals (brand portfolio, country averages, temporal trends) can produce reliable predictive insights. While no prior work has comprehensively modeled fragrance success using a large-scale dataset such as Fragrantica's, this project extends existing methodologies from related fields and demonstrates their application to the perfume industry.

3. Proposed Method

The predictive system was developed through a structured pipeline, beginning with dataset preparation, followed by feature engineering, and concluding with model development and evaluation. The primary objective was to design a supervised learning framework capable of estimating the commercial success of a fragrance, represented by a composite Success Score metric. The dataset was first preprocessed and enriched through several steps, including the imputation of missing values, standardization of numerical variables, and the creation of additional features capturing temporal dynamics, brand reputation, perfumer expertise, geographical context, and olfactory complexity. This process ensured a clean and consistent foundation for the subsequent modeling phase.

In the regression setting, the task was framed as the prediction of the continuous Success Score. Linear Regression was employed as a baseline due to its interpretability, while ensemble methods such as Random Forest and Gradient Boosting were prioritized for their ability to capture nonlinear interactions across heterogeneous features. Evaluation metrics for this task included the coefficient of determination (R^2) and error measures such as RMSE, which provide a comprehensive assessment of predictive accuracy.

In addition to regression, the problem was reformulated as a classification task in order to provide an alternative perspective on fragrance success. Instead of predicting a continuous score, perfumes were divided into two groups according to their Rating Value, with the median threshold of 3.97 used to separate high-rated from low-rated perfumes. This binary framing allowed the models to focus on distinguishing between fragrances that consumers perceived as more or less appreciated. Classification was primarily addressed using a Random Forest Classifier, with performance evaluated in terms of accuracy, precision, recall, and ROC-AUC.

Taken together, the regression and classification formulations offered complementary perspectives: the former provided a continuous estimation of commercial success, while the latter emphasized the model's discriminative ability in identifying perceptual differences in fragrance quality. Both approaches relied on feature importance analyses to identify the factors most strongly associated with success, enabling interpretability alongside predictive modeling.

3.1. Feature Engineering

To capture both intrinsic and extrinsic determinants of fragrance success, a range of additional features was engineered. Temporal indicators included perfume age, a binary variable for recent releases (5 years), and a vintage status indicator (20 years). Brand-level attributes summarized portfolio size, average ratings, popularity, and years active, reflecting reputation and longevity.

Perfumer-related features included portfolio size, average ratings, and historical success. Olfactory characteristics were quantified through the number of notes in the top, middle, and base layers, as well as the overall note count. In addition, the forty most frequent accords (such as woody, citrus, aromatic, and sweet) were one-hot encoded to capture recurring scent families. Finally, gender categories (men, women, unisex) were transformed into binary indicators.

Following feature engineering, the dataset expanded from 18 to more than 80 structured predictors, enhancing the model's ability to represent the multifaceted drivers of success.

3.2. Data Preparation

The dataset was obtained from Kaggle, where it had been previously scraped from *Fragrantica.com*, one of the largest online fragrance repositories. It contains more than 24,000 perfumes and their associated attributes, including perfume name, brand, country of origin, gender, release year, perfumer(s), fragrance notes (top, middle, base), main accords, as well as consumer ratings and review counts.

Several preprocessing steps were necessary to ensure consistency and prepare the data for modeling. A key challenge involved missing values in the release year, with around 2,000 entries incomplete. Median imputation at 2015 resulted in an unrealistic concentration of records, distorting temporal trends. To avoid this, missing years were imputed through a random assignment weighted by the empirical distribution of existing years. Perfumer names were left as "unknown" where not provided, thereby distinguishing anonymous creations from those signed by established professionals.

Other transformations addressed formatting and scale issues. Numerical variables such as *Rating Value*, originally recorded in European decimal notation with commas, were converted into floating-point values, while *Rating Count* was standardized as an integer. To account for scale differences, user ratings were normalized using min–max scaling, while review counts were log-transformed to reduce skewness from highly popular fragrances.

Categorical attributes also required transformation. Gender was encoded using one-hot encoding (men, women, unisex), while the 40 most frequent fragrance accords were similarly expanded into binary indicators, allowing the models to capture recurring scent families such as woody, citrus, sweet, and aromatic.

In addition, correlation analysis was performed to detect redundant predictors. Highly collinear features were removed in order to reduce multicollinearity and improve model stability. Continuous variables were then standardized to ensure comparability across features with different ranges.

After preprocessing, the dataset was reduced from 62 to 57 features, producing a cleaner and more robust representation of both intrinsic fragrance characteristics and contextual market factors, ready to be used in supervised learning tasks.

3.3. Defining Commercial Success

Since success cannot be reduced to either ratings or popularity alone, a composite index was developed. The *Success Score* was defined as:

$$\text{Success\_Score} = 0.8 \times NormalizedRating + 0.2 \times NormalizedLogRatingCount$$

This formulation assigned greater weight to quality, as expressed by user ratings, while still incorporating visibility through review counts. The resulting measure ranged from 0.05 to 0.90, with a mean of approximately 0.63, providing a continuous and interpretable target for regression modeling.

3.4. Modeling Approach

The task was framed as a regression problem, where the model aimed to predict the continuous *Success Score*. Several approaches were considered. Linear Regression served as a baseline due to its interpretability, though it was limited in capturing non-linear interactions. Random Forest, an ensemble of decision trees, was prioritized for its robustness with heterogeneous data and its capacity to estimate feature importance. In addition, lightweight AutoML frameworks such as FLAML were considered to automate hyperparameter tuning and provide benchmarking against manually designed models.

Beyond regression, the problem was also reformulated as a classification task. In this setting, the target variable was the *Rating Value*, with perfumes divided into two balanced groups using the median threshold of 3.97. Fragrances with a rating above this value were classified as high-rated, while those at or below it were categorized as low-rated. This binary framing enabled the evaluation of the model's discriminative ability. The classification task was primarily addressed through a *Random Forest Classifier*, with performance assessed using standard metrics such as accuracy, precision, recall, and ROC-AUC.

4. Results

The predictive experiments were carried out on the held-out test set, with performance assessed using the coefficient of determination (R^2) and Mean Squared Error (MSE). Both regression and classification settings were explored to evaluate the feasibility of predicting fragrance success from intrinsic and contextual features.

4.1. Model Performance (Regression)

The linear regression baseline achieved an R^2 of approximately 0.34 with an RMSE of 0.060. This result confirmed that linear models can capture some variance, but struggle with the complex, non-linear relationships characterizing fragrance success. Ridge regression produced nearly identical results, while Lasso regression performed worse, indicating that aggressive feature shrinkage removed useful predictors.

The **Random Forest Regressor** improved predictive capacity slightly, achieving an R^2 of 0.378 with an RMSE of 0.059. The ensemble's ability to capture non-linear interactions was evident in its tighter clustering of predicted vs. actual scores. Further hyperparameter tuning through randomized search did not significantly improve performance, confirming a natural ceiling.

The best-performing model was obtained via **AutoML** (FLAML), which selected a LightGBM regressor. This model achieved an R^2 of 0.384 and an MSE of 0.00340, representing the lowest error among all tested approaches. While the gain compared to Random Forest was incremental, it highlighted the benefit of gradient boosting methods for structured, tabular datasets.

Overall, the regression results suggest moderate predictive power, consistent with the literature on cultural product success prediction, where subjective consumer dynamics introduce significant irreducible noise.

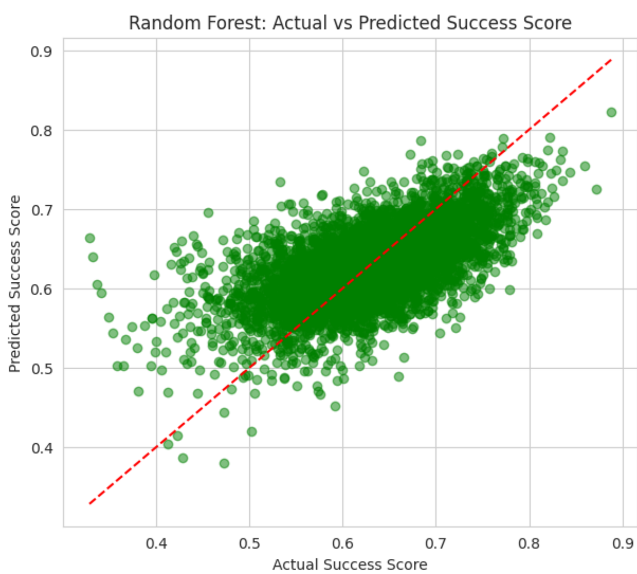

Figure 1. Random Forest Regressor: Actual vs. Predicted Success Score.

4.2. Feature Importance

Analysis of feature importance across Random Forest and LightGBM revealed a consistent set of dominant predictors.

- **Brand reputation** (`Brand_Avg_Rating`) was by far the strongest driver, with importance scores between 0.18 and 0.25. This underscores the centrality of brand equity in consumer evaluations.
- **Temporal features** such as release year and fragrance age ranked next, reflecting both novelty effects and the enduring appeal of vintage products.
- **Brand visibility** through portfolio size and popularity also contributed significantly, indicating that established houses with broader offerings gain reputational and distribution advantages.
- **Perfumer expertise** (portfolio size and track record) provided incremental explanatory power, highlighting the reputational spillover of master perfumers.
- **Olfactory structure** measured through note counts (top, middle, base) contributed modestly, suggesting that compositional richness influences success but remains secondary to extrinsic factors.
- **Accords** such as woody, citrus, aromatic, and sweet consistently appeared among the top-ranked categorical features, confirming that certain olfactory families resonate more strongly with consumers.

By contrast, **gender labels** (men, women, unisex) exhibited low importance, suggesting that marketing segmentation by gender is less decisive in shaping consumer ratings compared to perceived quality and olfactory identity.

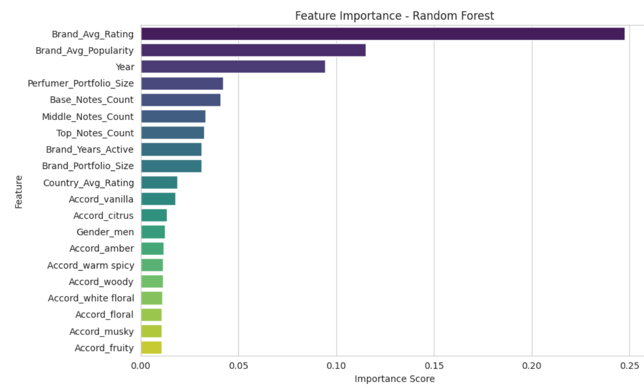

Figure 2. Feature importance from the Random Forest Regressor (Success Score).

4.3. Model Performance (Classification)

To complement regression, a binary classification task was conducted by splitting perfumes based on whether their rating value exceeded the median (3.97). This produced two balanced classes of "low-rated" and "high-rated" perfumes.

A **Random Forest Classifier** achieved an overall accuracy of 71%, with precision and recall balanced across both classes (precision = 0.73 for class 0, 0.69 for class 1). The ROC-AUC score of 0.78 indicated a good ability to discriminate between higher- and lower-rated fragrances. The confusion matrix confirmed that the classifier maintained balanced performance, avoiding major class bias.

Feature importance in classification largely mirrored the regression results. The most influential predictors were again `Brand_Avg_Rating`, release year, brand popularity, and brand portfolio size. Among accords, *citrus* and *woody* emerged prominently, further confirming the generalizability of these olfactory themes across predictive tasks.

4.4. Insights

Several insights emerge from the combined regression and classification analyses:

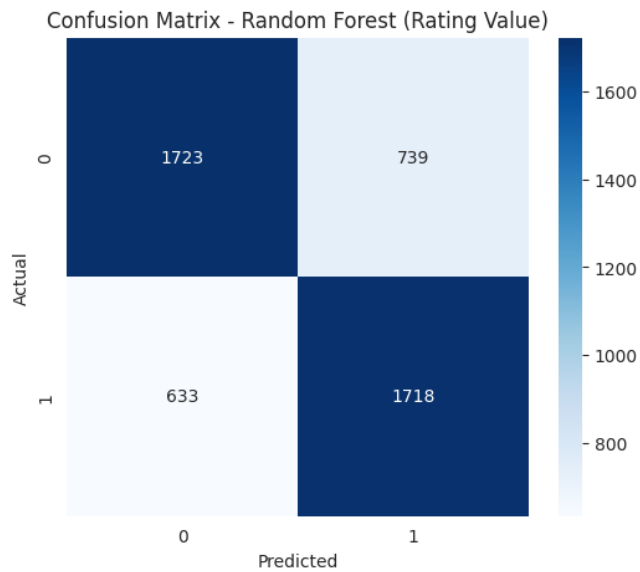

Figure 3. Confusion Matrix of the Random Forest Classifier (Rating Value).

- **Reputation dominates**: both brand equity and perfumer expertise are central determinants of success, underscoring the path-dependent nature of consumer preferences.
- **Temporal effects matter**: novelty and heritage both contribute, but in different ways—recent releases benefit from trend dynamics, while vintage fragrances leverage prestige.
- **Olfactory complexity and accords play a role**: while not as decisive as reputation, structural richness and certain scent families (woody, citrus) consistently enhance success probability.
- **Predictive ceiling**: the moderate R^2 values and only fair classification performance highlight the strong influence of unobserved external drivers such as advertising campaigns, cultural positioning, and celebrity endorsements.

5. Conclusion

This project set out to investigate whether the commercial success of fragrances can be predicted using machine learning, by combining intrinsic olfactory features (notes, accords, and overall complexity) with extrinsic market characteristics (brand reputation, perfumer expertise, and temporal as well as geographical context). The construction of a composite *Success Score*—defined as a weighted function of normalized ratings and popularity—enabled the establishment of a practical and interpretable target variable that captures both consumer satisfaction and visibility within the marketplace.

The methodological pipeline developed for this work integrated multiple stages: rigorous preprocessing and cleaning of the raw dataset, imputation of missing values, scaling and normalization of numerical variables, reduction of collinearity, feature engineering, and ultimately model training and evaluation. Baseline results obtained with Linear Regression highlighted the presence of predictive signals in the data ($R^2 \approx 0.34$, RMSE ≈ 0.060). Ensemble models substantially improved performance: Random Forest achieved $R^2 \approx 0.38$ (RMSE ≈ 0.059), while an AutoML procedure identified LightGBM as the best performing estimator ($R^2 \approx 0.384$, MSE ≈ 0.0034). Although the magnitude of these values indicates only moderate predictive accuracy, the consistency across models is noteworthy.

Feature importance analysis provided valuable insights into the drivers of fragrance success. Across models, brand-level reputation, in particular *Brand Average Rating* and *Brand Popularity*, emerged as the most influential predictors, far outweighing most olfactory descriptors. Temporal variables such as release year and brand years of activity also played a key role, suggesting that longevity and timing of release influence consumer perception. Perfumer portfolio size, together with note complexity (counts of base, middle, and top notes), contributed additional explanatory power. These findings underline the interplay between creativity, craftsmanship, and market positioning in shaping consumer evaluations.

Beyond regression tasks, a complementary classification experiment was conducted using a binary target defined by the median of *Rating Value*. The Random Forest Classifier achieved approximately 71% accuracy and a ROC AUC of 0.78. The stability of these results reinforces the central role of brand reputation while also demonstrating that accords such as citrus, woody, amber, vanilla, and floral act as meaningful discriminators of high- and low-rated perfumes.

Taken together, these results suggest that success in the fragrance industry is far from arbitrary. Instead, it emerges from the interaction between intrinsic product features and extrinsic market signals. Nevertheless, a considerable portion of unexplained variance remains. Factors such as marketing investments, advertising strategies, distribution networks, celebrity endorsements, and evolving cultural trends were excluded from this analysis but undoubtedly exert a strong influence on consumer choice and overall market outcomes.

The study also has methodological limitations. The reliance on crowd-sourced data from *Fragrantica.com* introduces potential biases, as ratings may reflect niche communities rather than the global consumer base. The absence of direct sales figures, retail performance indicators, and detailed marketing expenditure data constrains the capacity of the models to approximate real-world commercial success. Moreover, the focus on structured features meant that unstructured textual data such as consumer reviews, which might offer rich sentiment information, were not leveraged in this work.

Future research could extend the present framework in several directions. Integrating retail sales data and market share indicators would substantially increase external validity. Incorporating social media analytics and sentiment analysis of online reviews could provide real-time signals of consumer engagement and cultural momentum. Deep learning models, particularly those capable of handling multimodal data (e.g., textual reviews, olfactory note embeddings, and visual advertising content), may capture more complex patterns and improve predictive performance. From a managerial perspective, linking predictive outputs with strategic levers such as pricing, positioning, and promotional intensity could translate data-driven insights into actionable recommendations.

In conclusion, this project demonstrates the feasibility and value of applying machine learning to the domain of fragrance analytics. While no predictive model can fully capture the intricacies of consumer taste and cultural influence, the analysis highlights key structural drivers of success and uncovers systematic patterns in both brand-related attributes and olfactory families. By combining predictive modeling with interpretable clustering, the study not only quantifies the determinants of fragrance success but also offers a framework for exploring creativity and market performance through the lens of data science. Such an approach opens promising avenues for bridging artistic creation with empirical evidence, thereby contributing to a more transparent and strategic understanding of consumer preferences in creative and experiential industries.

6. Extra: Clustering of Fragrance Accords

In addition to the predictive models developed for regression and classification tasks, an exploratory analysis was carried out through clustering techniques applied to fragrance accords. The motivation for this step was not only to understand whether perfumes could be grouped into coherent olfactory families, but also to evaluate the potential of such groupings for applications beyond pure prediction.

The methodological approach relied on a TF–IDF representation of fragrance accords, where the hierarchical importance of each accord within a perfume was preserved through a weighting scheme. By repeating the first accord five times, the second four times, and so on down to the fifth, the transformation ensured that the most defining notes of a perfume had greater influence in the numerical representation. The use of TF–IDF further allowed us to highlight distinctive descriptors, such as leather or oud, while attenuating the impact of overly common labels such as floral or woody. This representation was subsequently clustered with k–means, producing ten distinct groups of perfumes which, when examined, corresponded closely to well-known olfactory archetypes.

The results of clustering were highly interpretable. Some groups aggregated playful, fruity, and sweet compositions often associated with younger consumers and gourmand preferences, while others were dominated by fresh and aromatic accords typical of fougère and energetic masculine scents. At the opposite end of the spectrum, clusters emerged around leather, animalic, and amber profiles, which are usually perceived as niche or bold creations. Between these extremes, floral and vanilla-oriented categories reflected classic feminine signatures, while powdery and musky accords were characteristic of more elegant and timeless creations.

The importance of these findings goes beyond descriptive categorization. Clustering confirms that consumer preferences and creative practices in perfumery gravitate toward recurring structures rather than being randomly distributed. This implies that unsupervised learning can be employed not only as an analytical tool for understanding the fragrance market, but also as the foundation for future recommender systems. For instance, knowing that a consumer expresses a preference for perfumes dominated by woody and amber accords would allow the system to suggest other fragrances belonging to the same cluster, or to recommend adjacent olfactory families with a higher probability of acceptance.

In future work, this approach could be extended to personalized recommendation platforms, where a user's historical preferences or ratings are mapped to one or more clusters, and new perfumes are suggested accordingly. Such systems would bridge the gap between quantitative data science and the experiential dimension of perfumery, offering consumers more meaningful guidance in navigating an increasingly saturated market. Moreover, the capacity to identify niche-oriented clusters may assist brands in targeting specific audiences and positioning their creations strategically.

In conclusion, the clustering of accords not only validated the existence of coherent olfactory families, but also demonstrated the potential of unsupervised methods to power recommendation and personalization in the fragrance industry. By aligning consumer tastes with data-driven insights, this type of analysis paves the way for innovative digital tools that could enhance both consumer experience and managerial decision-making in creative industries.

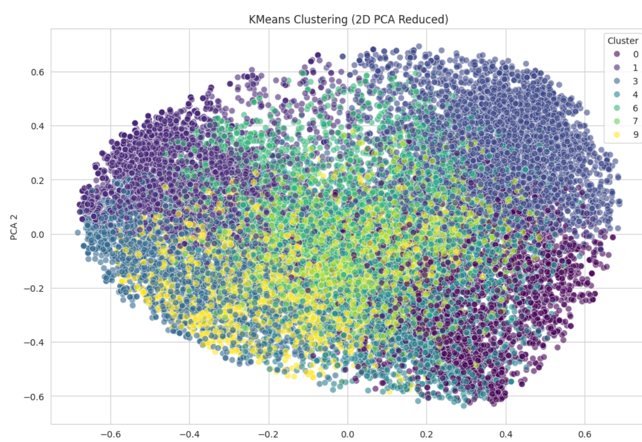

Figure 4. Visualization of fragrance clusters.