

1. About the Data:

The dataset under analysis comprises comprehensive product reviews from an e-commerce platform (nykaa-India) specializing in cosmetic products. It contains several key fields, including:

- **Product Information:** Unique identifiers (**product_id**), product titles (**product_title**), and URLs (**product_url**). It also includes price (**price**), overall product ratings (**product_rating**), the number of ratings (**product_rating_count**), and associated tags (**product_tags**).
- **Review Details:** Each product review is uniquely identified (**review_id**) and includes a title (**review_title**), the full text of the review (**review_text**), the date it was posted (**review_date**), and the rating given (**review_rating**).
- **Author Details:** Information about the reviewer, such as their username (**author**), whether they have purchased the product (**is_a_buyer**), and if they are considered a professional user (**pro_user**).
- **Brand Information:** The name of the brand (**brand_name**) that the product belongs to, providing insights into brand-specific consumer feedback.
- **Review Label:** A categorical label (**review_label**) that could potentially indicate the sentiment of the review or other classifications derived from the review content.

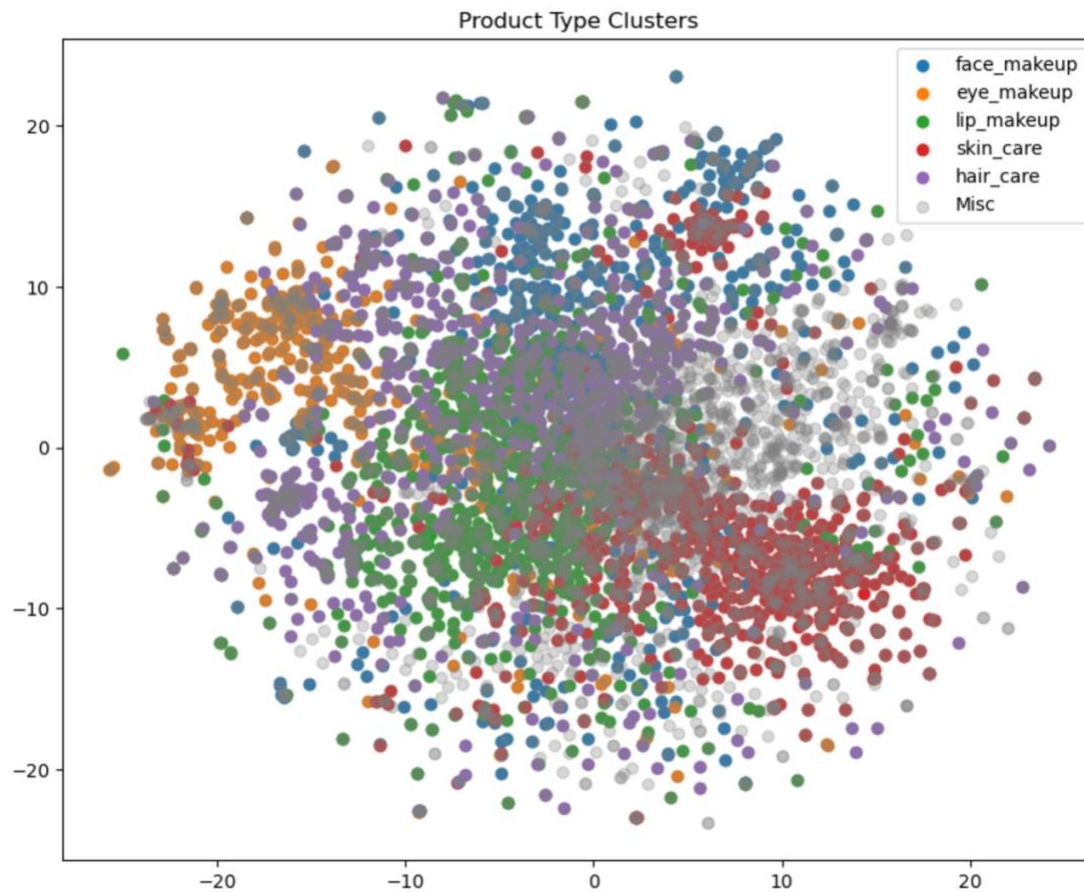
This dataset enables a multifaceted analysis of consumer preferences, behaviors, and trends across various brands and product types in the cosmetic industry. The detailed review data, combined with product metadata, offers a rich resource for understanding customer satisfaction and product performance on a granular level. The dataset is publicly available on Kaggle and can be accessed here:

<https://www.kaggle.com/datasets/jithinanievarghese/cosmetics-and-beauty-products-reviews-top-brands/code>

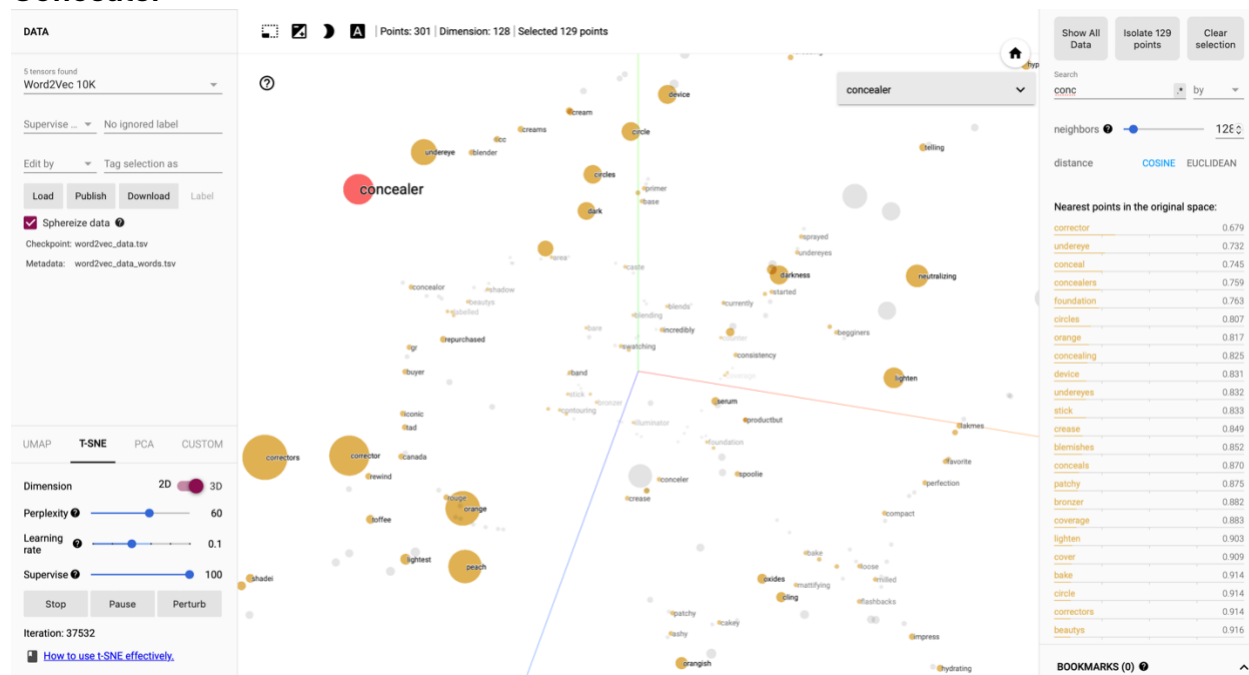
2. **Code:** The code to create word vectors is uploaded as a Jupyter notebook file.

3. Screenshot of output visualization

Below is the visualization of the product clusters to determine what kind of reviews were popular. The clusters seem to have enough representation of all the categories I defined.



Concealer



4. Knowledge Discovery

Given the visualizations above from the t-SNE analysis of word vectors related to beauty_reviews, I picked compact and concealer as they were most related to foundation. I assumed that a new business specializing in foundations wants to expand their market with other products and these were the most related.

- a. **Integration with Foundation:** The high similarity between "concealer" and "foundation" indicates that consumers may be using these products together.
- b. **Correction and Coverage:** The terms "corrector" and "coverage" being closely associated with "concealer" suggests that consumers prioritize concealers that offer substantial coverage and have corrective properties.
- c. **Under-Eye Focus:** The association of "undereye" and "circles" with "concealer" underscores the importance of concealers in targeting under-eye issues. Consumers may be looking for products that specifically address dark circles and under-eye puffiness.
- d. **Consistency and Finish:** The relationship between "concealer" and descriptors like "matte," "dewy," and "neutralizing" provides insight into the desired finish and consistency of concealers. The emphasis may be on a natural finish that doesn't crease, particularly under the eyes, and a texture that allows for easy blending.
- e. **Skin Concerns:** Terms such as "blemishes," "neutralizing," and "cancel" reflect a concern for skin imperfections. This suggests a market preference for concealers that not only cover but also treat or neutralize various skin discolorations and imperfections.
- f. **Formulation:** The composite vector of "concealer," "undereye," and "neutralizing" with a high similarity score back to "concealer" could indicate that the ideal concealer formula should specifically cater to neutralizing and correcting under-eye concerns.

5. Results from Word2Vec that support findings.

- a. **Integration with Foundation:** The visual clustering of 'concealer' near 'foundation' supports the notion that consumers often discuss and perhaps use these products in tandem. The visualization reflects a consumer trend where the discussion of foundation naturally segues into concealers, suggesting that the products are complementary in the consumer's makeup routine.
- b. **Correction and Coverage:** In the visualization, 'corrector' and 'coverage' are located near 'concealer', which supports the idea that consumers are looking for multifunctionality in their concealers. This insight can guide product development to focus on a formula that

emphasizes high coverage and corrective features, responding to this clear consumer demand.

- c. **Under-Eye Focus:** The t-SNE visualization shows 'undereye' and 'circles' in the same neighborhood as 'concealer,' which underlines the consumer's focus on concealers as a solution for under-eye issues. It is likely that when consumers are considering concealers, their primary concern is the ability of the product to address under-eye problems, which should be a focal point in concealer product development.
- d. **Consistency and Finish:** The proximity of terms related to finish, such as 'matte', 'dewy', and 'neutralizing', to 'concealer' in the visualization, supports the diversity in consumer preferences for concealer finishes. This pattern in the data indicates a market that values a range of options, suggesting that a line of concealers offering various finishes could be well-received.
- e. **Skin Concerns:** Words like 'blemishes,' 'neutralizing,' and 'cancel' appear near 'concealer' in the visualization, which confirms a consumer interest in concealers that go beyond mere coverage. The visualization suggests a preference for products that address specific skin concerns, pointing towards a market for concealers that offer corrective benefits.
- f. **Formulation:** The high similarity score of the composite vector combining 'concealer,' 'undereye,' and 'neutralizing' suggests that these attributes are core to the consumer's expectations of a concealer. The visualization thus implies that a successful concealer formula in the market would be one that excels in correcting and neutralizing, especially for under-eye concerns.