**Project Title: Sephora: Beauty by Numbers**

**Project Description:**

This project delves into the Sephora dataset, a renowned name in cosmetics and beauty retailing. The main goal is to extract valuable insights from the dataset, exploring product variety, customer preferences and pricing.

**The project steps included:**

1. **Data Acquisition**: The initial dataset for this analysis was downloaded from Kaggle.com for use in a Sephora Python project. The dataset comprises information on over 8,000 beauty products from Sephora's online store and was collected in March 2023 using a Python web scraper. It includes details such as product and brand names, prices, ingredients, ratings and more. The data was prepared in Python and saved to a CSV file, which is used in this analysis.
2. **Data Connection**: A connection was established to an csv file within Microsoft Power BI, enabling an access and integration of the dataset.
3. **Data Preparation**: No data cleaning was required as the data had already been

pre-processed in Python. Some columns were renamed for convenience and additional columns were created as needed for the Power BI report.

1. **Comprehensive Analysis**: Data Analysis Expressions (DAX) measures were utilized to create a comprehensive report in Power BI, enabling in-depth analysis and visualization of the dataset.
2. **Detailed Report**: The project was concluded by generating a 5-tab report in Power BI, comprising "Overview", "Categories", "Brands", "Products" and "Prices". This multi-tab report was used to provide valuable insights and present the analysis findings in a structured manner.

**Key findings from the data analysis are as follows:**

1. **Dataset Overview:** The dataset comprises 8,494 products from 304 unique brands. The average rating across all products is 4.1. Most products fall within the rating range of 4 to 4.75, with the highest count of ratings occurring between 4.25 and 4.5.

The average price of products is $52 USD, with a median price of $35 USD.

1. **Data Analysis Approach:** To minimize the impact of outliers and variations due to brand sizes, the analysis primarily relies on the average number of Likes Count, Reviews Count and Prices.
2. **Category and Subcategory Insights:** The dataset covers a total of 9 main categories, along with 41 subcategories**.** In terms of average Likes Count, Makeup is the highest, followed by Skincare and Mini size categories. For average prices, Fragrance tops the list, followed by Skincare and Gifts.

Mini size, Makeup and Skincare have the highest average Reviews Count. Interestingly, the Gifts category contains only 4 products, all priced at $50 USD each.

1. **Brand Analysis:** Among brands, OLAPLEX leads with the highest average Likes Count, followed by RARE BEAUTY BY SELENA GOMES and THE ORDINARY.

BUXOM stands out with the highest average number of reviews, with ROSEBUD PERFUME CO. as a distant second. Illuminage has the highest average price, followed by NUFACE and SHANI DARDEN SKIN CARE.

1. **Product Insights:** The most expensive product in the dataset is "Shani Darden by Déesse PRO LED Light Musk" by SHANI DARDEN SKIN CARE, priced at $1,900 USD. "Soft Pinch Liquid Blush" by RARE BEAUTY BY SELENA GOMES has the highest Likes Count, with a staggering 1,401,068 Likes Count. "Tattoo Liner Vegan Waterproof Liquid Eyeliner" by KVD BEAUTY has the highest number of reviews, totalling 21,281.

Separate analysis is required to delve deeper into these anomalies.

1. **Insights from Product Category Analysis:** In the analysis, a close examination was conducted on various product categories, including SEPHORA EXCLUSIVE, LIMITED EDITION, ONLINE ONLY, NEW and OUT OF STOCK. Here's what emerged:

SEPHORA EXCLUSIVE products make up 28% of the entire product range and are generally priced more affordably than other products. LIMITED EDITION and NEW Products: These products had a very low presence in the dataset and came with higher price tags. ONLINE ONLY products make up 28% of the total, and they are 20% more expensive than non-online items. OUT OF STOCK products have limited presence in the dataset and are typically priced lower than similar options.

1. **Ingredient Analysis:** A word cloud was generated to visualize the most common ingredients, but further analysis is necessary to understand their impact on product prices and other metrics.

**Conclusion:**

In conclusion, the "Sephora: Beauty by Numbers" project has delved into Sephora's dataset, providing valuable insights into product variety, customer preferences and pricing. Through data organization and thorough analysis, the findings have been presented in a structured report. This project offers useful information for both professionals and consumers interested in the cosmetics and beauty retail sector, enabling them to make more informed choices.