

**Businesss Intelligence and DATABASE Management System**

**Business Intelligence Project on Sephora**

*Prepared by:*

Doua Etteyeb

Fadwa Chibani

*--Junior ACCT /IT--*

Academic year: 2024-2025

***Contents***

1. Introduction .................................................................3
   * Objectives
   * Goals
   * Deliverables
2. Implementation..............................................................3

* 2.1 Data Gathering
* 2.2 Data Preparation (ETL : Extract, Transform, Load)

1. Data Storage / Modeling and DWH creation...................5

* 3.1 Storage
* 3.2 Data Warehouse Schema
* 3.3 Olap Process

1. Data Visualization………………………………………………………10
2. Conclusion.....................................................................11
   * 5.1 Analysis Outcomes
   * 5.2 Strategic Recommendations
   * 5.3General Conclusion
3. **Introduction**

In the competitive beauty market, understanding customer preferences and brand performance is crucial. This project analyzes Sephora's product and brand data to uncover insights on product popularity, customer satisfaction, and brand positioning. These insights will enable Sephora to optimize product assortments, refine marketing strategies, enhance customer engagement, and drive business growth. By leveraging these findings, Sephora can navigate the evolving market and strengthen its industry leadership.

**1.1. Objectives**

* Identify the most popular and profitable products.
* Analyze product performance across categories
* Analyze brand product portfolio strategies
* Identify Top-performing brand based on customer engagements

**1.2. Goals**

* Optimize product assortment (based on popularity, ratings, and product distribution).
* Enhance brand performance insights (based on popularity and ratings).
* Improve customer engagement (by understanding preferences and feedback).
* Drive Revenue Growth (By identifying top-selling products, categories).

**1.3. Deliverables**

This project delivers :

* Optimized Database : MySQL database schema designed for efficient data storage and analysis of Sephora product, brand, and category data.
* Interactive Power BI Dashboards : Visualizations of key performance indicators (KPIs) related to product popularity, customer satisfaction, and brand performance, enabling data exploration and actionable insights
* Comprehensive Project Report : A document detailing the project lifecycle, from data gathering and preparation to data modeling, visualization, and recommendations.

**2.Implementations**

**2.1 Data Gathering**

Data was gathered from **Kaggle** This data was organized into three Pandas DataFrames :

* **df1** (Brands with Number of Products): Contains information about 100 cosmetic brands, including **brand** (brand name), **COUNT** (number of products), and brand\_id (unique brand identifier).
* **df2** (Most Expensive Products)**:** Contains information about a set of products, including **brand**, **name**, and **price**.
* **df3** (Product Information)**:** Contains detailed information about individual products, including **product\_id, product\_name, brand\_id, brand\_name, loves\_count, rating, reviews, size, price\_usd, value\_price\_usd, sale\_price\_usd, online\_only, sephora\_exclusive, primary\_category, secondary\_category,** and **tertiary\_category**.

**2.2 Data Preparation (ETL: Extract, Transform, Load)**

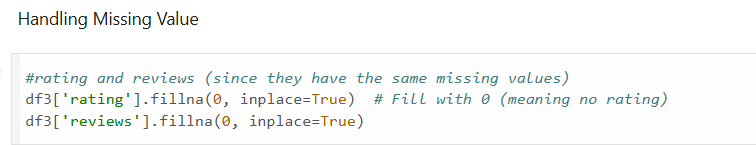
**Data Cleaning**

For the Sephora product and brand performance analysis, the data provided was pre-processed and organized in a structured format, allowing for efficient analysis and visualization. The key steps involved in the data preparation process are as follows :

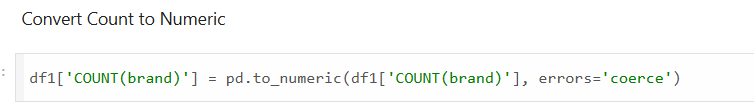
* **Handling missing values**
* **Data Type Conversion**
* **Performing text cleaning**
* **Creating unique identifiers**
* **Merging related tables**

This ensured the dataset was clean, standardized, and ready for further transformations. You can find some figures representing examples of these processes below :

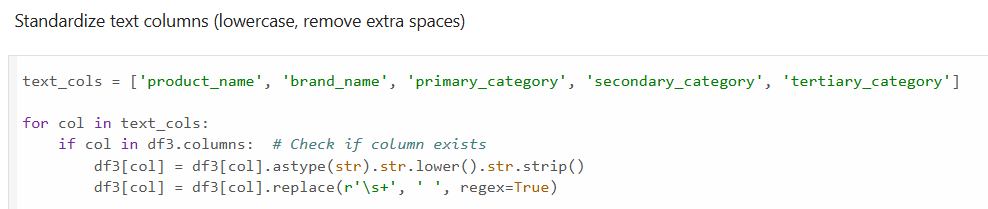
*Figure 1 : Handling missing values*



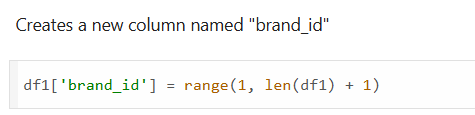
*Figure 2 : Data Type Conversion*



*Figure 3 : Performing Text Cleaning*



*Figure 4 : Creating unique identifiers*



*Figure 5 : Merging related tables*



**Data Loading :**

After cleaning the data, I used **TableConvert** to restructure and format it for compatibility with the MySQL data warehouse. This step ensured that the data adhered to a well-organized structure, making it suitable for creating the fact and dimension tables. By preparing the data in this way, I streamlined the process of loading it into MySQL, enabling seamless integration with Power BI for advanced analysis and visualization.

**3.Data Storage \ Modeling and DWH creation**

**3.1 Storage**

Sephora's product and customer data is consolidated into a structured **MySQL database**, laying the foundation for the analysis. This data model captures key facts, such as product sales, and measures, including total sales, customer engagement, and average ratings. These metrics are organized across product and brand dimensions, enabling efficient querying and exploration of performance, satisfaction and preference insights.

**A-Facts and measures :**

1. **Facts :**

* **factproductmetrics**

1. **Measures :**

* Sales
* Total loves
* Average rating
* Total reviews
* Number of products per band

**B-Dimensions :**

1. **Product :**

***Attributes:* product\_id (PK), product\_name, online\_only, sephora\_exclusive.**

**A close-up of a computer screen

Description automatically generated**

1. **Brand :**

***Attributes:* brand\_id (PK), brand\_name.**

**A close up of text

Description automatically generated**

1. **Category:**

***Attributes:* category\_id (PK), catproduct\_id (FK), primary\_category, secondary\_category, tertiary\_category.**

**A screenshot of a computer code

Description automatically generated**

1. **Price:**

***Attributes:* price\_id (PK), priceproduct\_id (FK), price\_usd, price\_value\_usd, sale\_price\_usd.**

**A screen shot of a computer code

Description automatically generated**

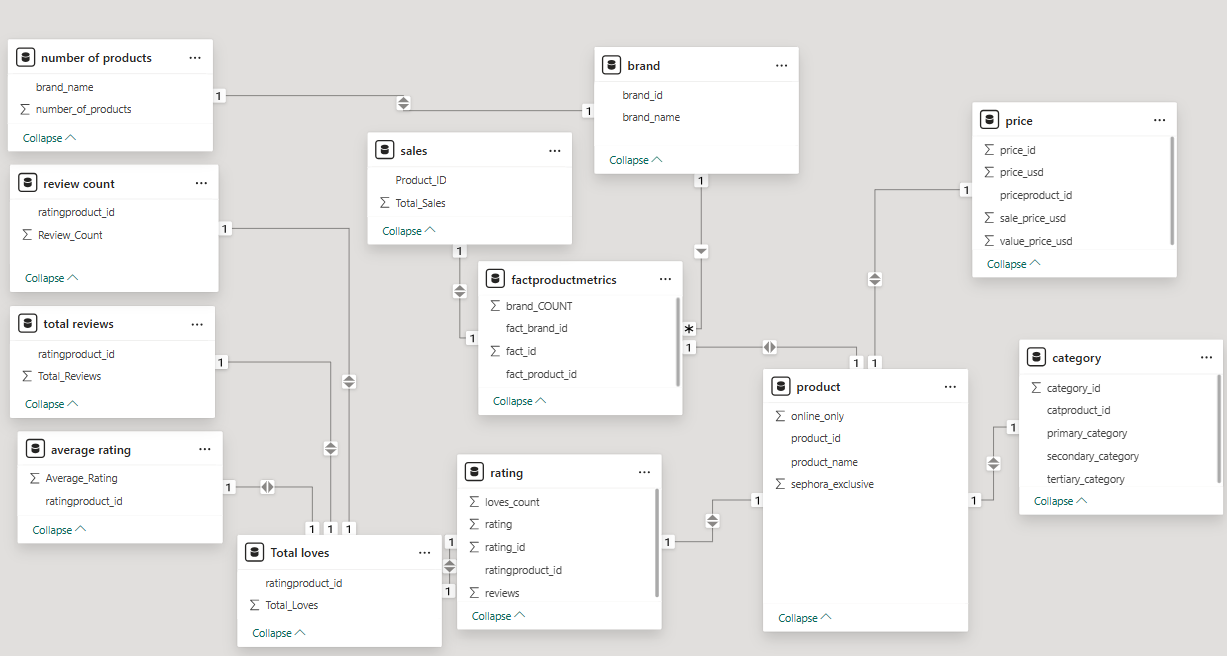
1. **Rating:**

***Attributes:* rating\_id (PK), ratingproduct\_id (FK), lobes\_count, rating, reviews.**

**A screenshot of a computer code

Description automatically generated**

**3.2 Data Warehouse Schema**

*Snowflake Schema*

**3.3 OLAP Process**

We chose to implement **ROLAP** (Relational OLAP) for this project. This decision was driven by the need to perform flexible and detailed analysis directly on the relational data stored in our MySQL database.

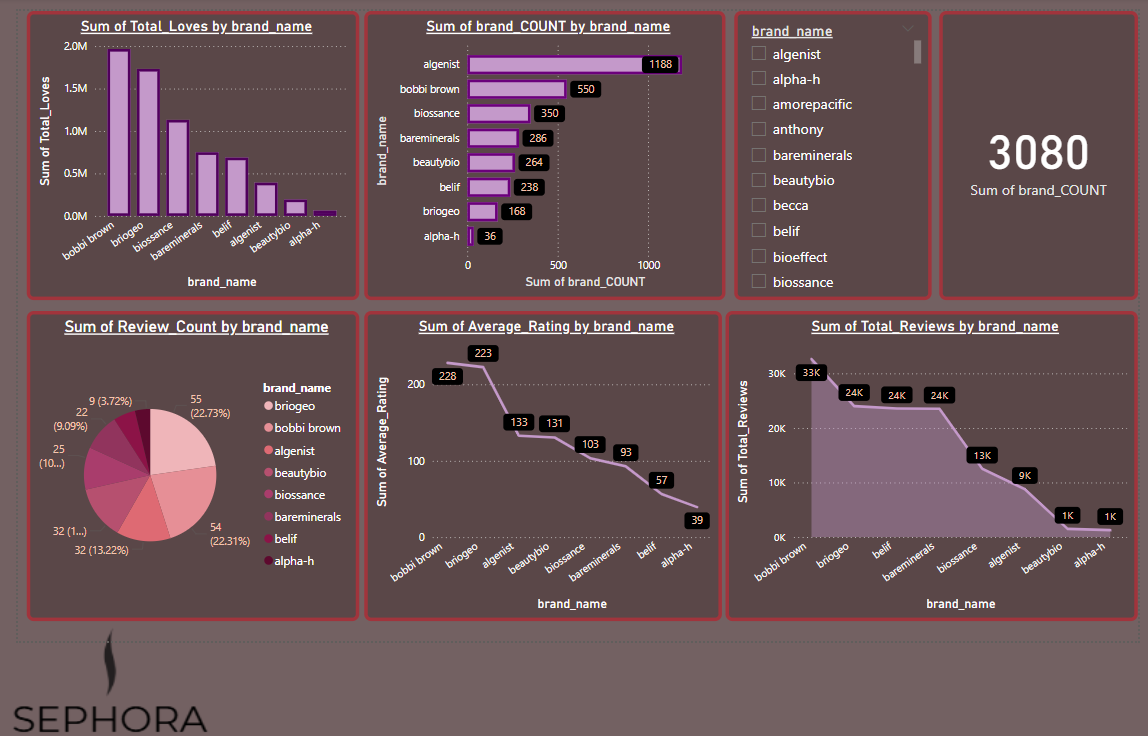
The following SQL queries demonstrate the ROLAP process :

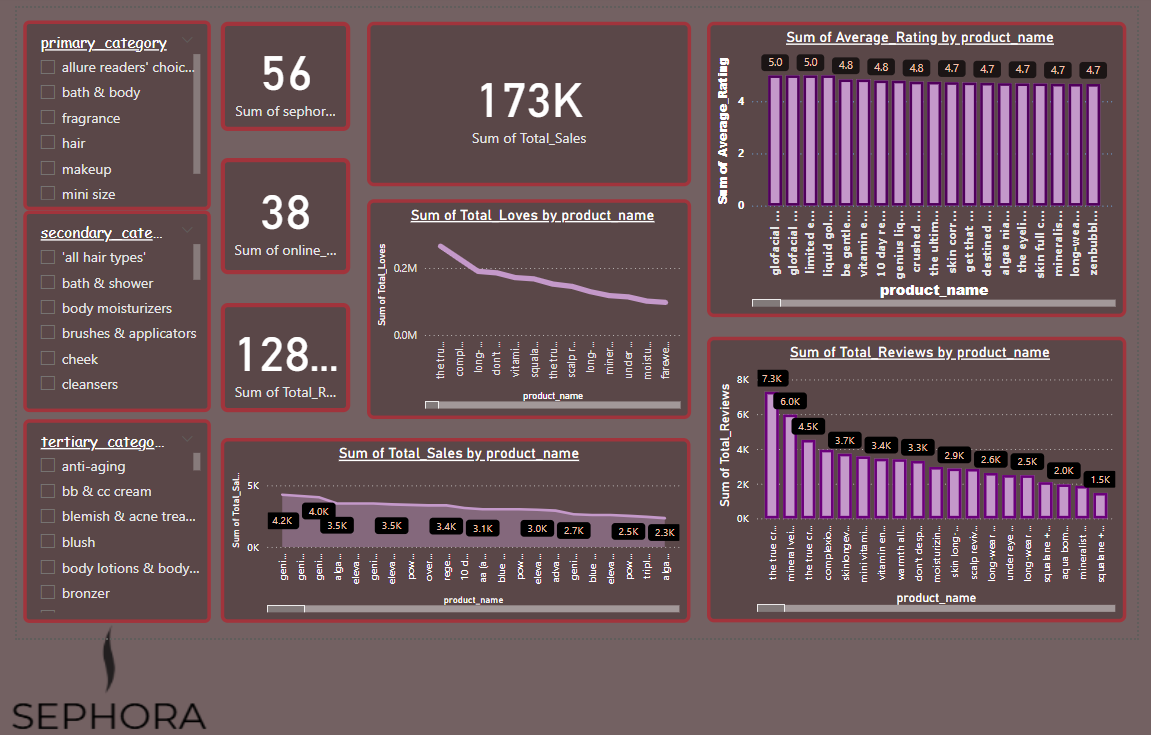
|  |  |
| --- | --- |
| **SQL queries** | **Output** |
| ***Total loves count by primary category***    ***Top 5 brand avg rating***  ***Average rating and total review for each brand*** |  |

**4. Data Visualization:**

We leveraged Power BI to conduct a comprehensive analysis of customer behavior and satisfaction with Sephora's brand offerings. By diving deep into Sephora's robust data, we were able to uncover valuable insights that can guide the company's strategies to better attract and serve its customers.

*Figure 1: Brand performance*



*Figure 2: Product/Category performance*

**5.Conclusions:**

**5.1 Analysis Outcomes**

1. **Identify the most popular and profitable products :**

* Most Profitable Product : Based on Aggregated Sales Data : "genius ultimate anti-aging vitamin c +serum" ($4.2K total sales)
* Most Popular Product : Based on Customer Engagement (Loves, Reviews) : "true cream aqua bomb"
* Highest Rated Product : Based on Highest Average Rating : "glofacial antimicrobial treatment tips & cleaning brush accessories".

1. **Analyze product performance across categories.**

Top Performing Categories (Based on Sales) **:**

* Primary Category : Skincare ($138k)
* Secondary Category (within Skincare) : Moisturizers ($52k)
* Tertiary Category (within Moisturizers) : Face Serums ($22k)

1. **Analyze brand product portfolio strategies :**

* **"algenist"** employs a strategy of offering a wide product range with approximately 1188 products.
* **"bobbi brown"** demonstrates strong performance with a more focused product portfolio of approximately 550 products.

1. **Identify Top-performing brand based on customer engagements**

* **Top Brand (Based on Engagement and Satisfaction) :** "bobbi brown" performs strongly in both customer engagement ("Loves" and reviews) and satisfaction (Average Rating).

**5.2 Strategic Recommendations**

**Maximize Bobbi Brown's Potential** : Feature prominently in marketing and promotions. \* Analyze and replicate their success factors (e.g., product quality, marketing).

**Optimize Algenist's Portfolio :**

\* Refine assortment by identifying and potentially discontinuing underperforming products. \* Implement targeted marketing campaigns based on product type.

**Boost Customer Engagement Across Brands :**

\* Incentivize reviews and "Loves." \* Implement/improve personalized product recommendations.

**Prioritize Skincare Category :**

\* Continue investment in Skincare (especially Moisturizers and Face Serums) due to high aggregated sales.

**5.3 General Conclusion**

This analysis provided key insights into Sephora's brand performance, focusing on customer engagement, satisfaction, and product portfolio strategies. "bobbi brown" emerged as a leader in customer engagement and satisfaction, while "algenist" demonstrated a contrasting strategy with a much larger product portfolio. Skincare, particularly moisturizers and face serums, proved to be a high-performing category. These findings offer actionable insights for Sephora and its brands to refine marketing, optimize product assortments, and enhance customer experiences.

**A full repository containing all project materials can be found at [https://github.com/fadwa-chb/Sephora\_Analysis\_Project]**