

THE CHINESE UNIVERSITY OF HONG KONG, SHENZHEN

CSC3170

DATABASE SYSTEMS

---

# A Database System for Sephora's Online Shop

---

*Group 6*

Guiming Chen 119010012

Jiarui Liang 119010170

Wanting Tang 119010278

Jialin Pan 119020040

Yunhan Zhang 119020076

Yuwei Chen 120090256

Kexuan Ma 120090651

Mingrui Liang 120090723

2023.05.08

# Contents

<b>1 Introduction</b>	<b>1</b>
<b>2 Database Design</b>	<b>1</b>
2.1 Entities, Attributes & Relationships	1
2.2 ER Diagram	3
2.3 Normalization & Relational Schema	3
2.4 Index & Hashing	4
<b>3 Data Collection &amp; Generation</b>	<b>5</b>
3.1 Data Collection	5
3.2 Data Generation	5
<b>4 Implementation, Results &amp; Analysis</b>	<b>6</b>
4.1 Creating Tables	6
4.2 Load Data	6
4.3 Basic Queries	7
4.3.1 Adding, Deleting and Updating	7
4.3.2 Retrieving	7
4.4 Analytical Queries	9
4.4.1 Customers' age groups	9
4.4.2 Relationship between ratings and customers' age groups	10
4.4.3 Relationship between ratings and product revenues	10
4.4.4 Seasonal consumption	10
4.4.5 Popular brands for each skin type	11
4.4.6 Complaints addressing efficiency	11
<b>5 Conclusion &amp; Self-evaluation</b>	<b>12</b>
5.1 Limitation & Conclusion	12
5.2 Self-evaluation	12
<b>Reference</b>	<b>13</b>
<b>Appendices</b>	<b>14</b>
A1. Comments generated by <i>phoenix-inst-chat-7b</i>	14

# 1 Introduction

Sephora is a worldwide retailer selling makeup and personal care products with hundreds of brands. It has online and offline shops, providing customers with readily accessible ways to purchase makeup and self-care products. The focus of this project will be on the online shop website of Sephora.

Like any online retailer, the Sephora website provides users with a platform to quickly check for the available products, their names, prices, the category of the products, and whether the product is in stock or unavailable. With one click on the link of each product, users will find a more detailed description of the products, including the applicable skin types and reviews of the products. The review section includes an overall rating of the product and specific user comments on certain dates. Moreover, after making an order, a record will be created that shows the total cost, shipping address, and other information related to that particular purchase. After receiving the items, the website also provides customers with a way to make complaints about the products if customers feel unsatisfied.

The basic functionalities of the Sephora online shop website motivate our group to create a database system that stores attribute information of different entities on the websites while connecting them in a dynamic manner.

## 2 Database Design

### 2.1 Entities, Attributes & Relationships

To simplify the database design, we proposed the following assumptions:

1. Each product only belongs to one brand; each brand has multiple products.
2. Each user can issue multiple reviews on multiple products; each product can be reviewed by multiple users.
3. Each user can place multiple orders; each order can be placed by only one user.
4. Each order can contain multiple items of only one product; each product can appear in multiple orders.
5. Each user can file multiple complaints; each complaint can only be filed by one user.
6. Each complaint directs to only one brand; each brand can be complained about multiple times.
7. Products of type "Hair & Body", "Makeup", or "Skincare" are applicable to multiple skin types. Each skin type can have multiple applicable products. Skin type contains "dry", "oily", and "normal".
8. Product price may be changed by the seller, the "Price" attribute in the "Product" table records the latest price, and the "Item price" attribute in the "Order" table is the product price at the time the order was placed.
9. Each user can issue at most one review on one product.
10. There are no other candidate keys except primary keys for each entity.

Based on the assumptions above, we designed the database, which comprises six distinct entities, each with unique attributes. This arrangement of entities and attributes serves to elucidate the comprehensive logical framework of the database lucidly and coherently.

The following part provides a more comprehensive explanation of the concepts of entities and their corresponding attributes with their constraints and properties:

1. **Entity:** brand

**Attribute:** BRAND\_NAME

As each brand on the platform contains a unique name, BRAND\_NAME can be considered the primary key. The entity is responsible for organizing the records of the various brands featured on the platform.

2. **Entity:** complaint

**Attribute:** COMPLAINT\_ID, STATUS, COMPLAINT\_DATE, RESOLUTION\_DATE

The unique COMPLAINT\_ID assigned to each complaint designates it as the primary key. STATUS provides information on the resolution status of each complaint, while COMPLAINT\_DATE and RESOLUTION\_DATE, respectively, record the dates on which the complaint was raised and resolved by the platform (Constraints: RESOLUTION\_DATE should be later than COMPLAINT\_DATE, and if a complaint hasn't been resolved, it's RESOLUTION\_DATE should be null). Together, these attributes and the entity enable efficient and organized tracking and management of user complaints within the database.

3. **Entity:** product

**Attribute:** PRODUCT\_ID, PRODUCT\_NAME, INVENTORY, PRICE, CATEGORY

Each product within the database is allocated a unique PRODUCT\_ID, making it the primary key. The entity is responsible for storing and organizing information about a wide range of products, each of which also consists of PRODUCT\_NAME. INVENTORY is maintained to track the quantity of each remaining product in the warehouse, while the PRICE records the latest pricing information. Additionally, CATEGORY is utilized to differentiate between different levels of perfumes, thereby enabling efficient organization and retrieval of data within the database.

4. **Entity:** user

**Attribute:** USER\_ID, USER\_NAME, AGE

A unique id is assigned to each user on the platform so USER\_ID is regarded as the primary key. The entity is responsible for organizing the detailed information of each user, including their USER\_NAME and AGE.

5. **Entity:** order

**Attribute:** ORDRE\_ID, ITEM\_PRICE, ITEM\_QUANTITY, SHIPPING\_ADDRESS, TOTAL\_SHIPPING\_COST, TOTAL\_COST, TIMESTAMP

Since every order has a distinct id number, ORDRE\_ID is the primary key. The entity is utilized to keep track of records of all user-generated orders. Within each order record, ITEM\_PRICE and ITEM\_QUANTITY, respectively, store information on the price and quantity of products. TOTAL\_COST attribute is computed by multiplying these two values, providing an overview of the total cost of the order. Finally, TIMESTAMP records the time when the order was placed.

6. **Entity:** skin

**Attribute:** SKIN\_TYPE

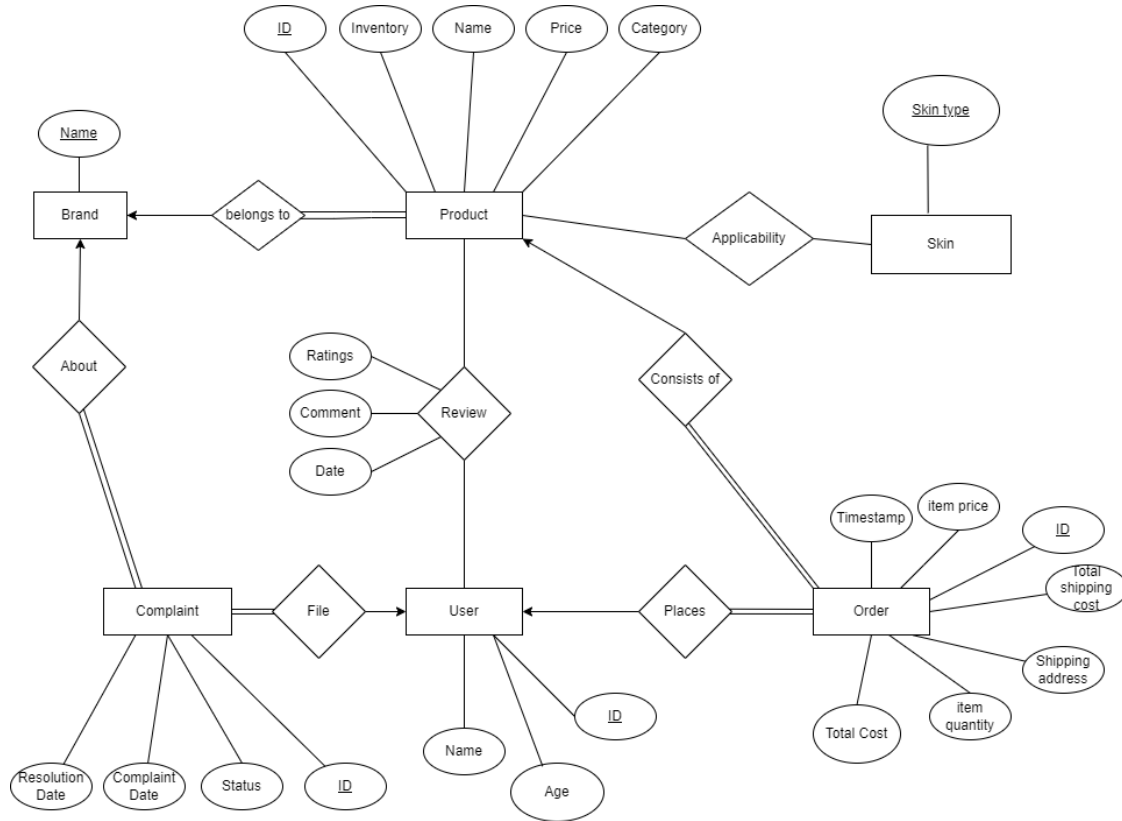
Since SKIN\_TYPE is the only attribute of this entity, it is regarded as the primary key. The entity is used to indicate different skin types for which various products are suitable.

The relationships are listed below:

1. **Belongs to:** Many to one relationship between Product and Brand
2. **Applicable for:** Many to many relationship between Product and Skin type
3. **About:** Many to one relationship between Complaint and Brand
4. **Review:** Many to many relationships between User and Product, each review has corresponding attributes Review\_DATE, COMMENT, and RATINGS. (Constraints: Review\_DATE should be later than order.TIMESTAMP)
5. **File:** Many to one relationship between User and Complaint
6. **Places:** Many to one relationship between User and Order
7. **Consists of:** Many to one relationship between Order and Product

## 2.2 ER Diagram

According to the database design above, we generated an ER diagram:



## 2.3 Normalization & Relational Schema

To guarantee that the schema is in normalized form, we examine all entities one by one. For the *brand* entity, it is obvious to see because there is only one attribute. For the *complaint* entity, each complaint is relatively independent. The status depends on each event, the resolution date is also random, and the user, as well as the brand, is a one-time issue as well. So there is no other key that is functionally dependent on any keys. As for the *product* entity, since Sephora is a collection of brands, one product name can appear in multiple brands as categories. However, in the initial schema, skin type as a multi-valued attribute exists in the *product* entity, which causes redundancy. Thus, we set up a *skin* entity relating to the *product* entity. For *user*, user name and age are also uncorrelated. For the *order* entity, it keeps records at an operational level, which is also a once-only issue. And each user can have multiple shipping addresses, so again, there are no functional dependencies. For associative entity applicability, there are no other keys except the composite primary key, so it is normalized. Lastly, as for the Review entity, the ratings, dates, and comments are not correlated, so any one of them cannot determine the others.

Based on the ER diagram and the normalization process, we obtained eight schemas in total:

1. **brand** (BRAND\_NAME)
2. **complaint** (COMPLAINT\_ID, STATUS, COMPLAINT\_DATE, RESOLUTION\_DATE, USER\_ID, BRAND\_NAME)

Foreign keys:

- USER\_ID references user.USER\_ID

- BRAND\_NAME references brand.BRAND\_NAME
- Functional dependencies:
- COMPLAINT\_ID → STATUS, COMPLAINT\_DATE, RESOLUTION\_DATE, USER\_ID, BRAND\_NAME
3. **product** (PRODUCT\_ID, PRODUCT\_NAME, INVENTORY, PRICE, CATEGORY, BRAND\_NAME)
- Foreign keys:
- BRAND\_NAME references brand.BRAND\_NAME
- Functional dependencies:
- PRODUCT\_ID → PRODUCT\_NAME, INVENTORY, PRICE, CATEGORY, BRAND\_NAME
4. **user** (USER\_ID, USER\_NAME, AGE)
- Functional dependencies:
- USER\_ID → USER\_NAME, AGE
5. **order** (ORDER\_ID, ITEM\_PRICE, ITEM\_QUANTITY, SHIPPING\_ADDRESS, TOTAL\_SHIPPING\_COST, TOTAL\_COST, TIMESTAMP, PRODUCT\_ID, USER\_ID)
- Foreign keys:
- PRODUCT\_ID references product.PRODUCT\_ID
  - USER\_ID references user.USER\_ID
- Functional dependencies:
- ORDER\_ID → ITEM\_PRICE, ITEM\_QUANTITY, SHIPPING\_ADDRESS, TOTAL\_SHIPPING\_COST, TOTAL\_COST, TIMESTAMP, PRODUCT\_ID, USER\_ID
6. **skin** (SKIN\_TYPE)
7. **applicability** (SKIN\_TYPE, PRODUCT\_ID)
- Foreign keys:
- SKIN\_TYPE references skin.SKIN\_TYPE
  - PRODUCT\_ID references product.PRODUCT\_ID
8. **review** (PRODUCT\_ID, USER\_ID, RATINGS, COMMENT, REVIEW\_DATE)
- Foreign keys:
- PRODUCT\_ID references product.PRODUCT\_ID
  - USER\_ID references user.USER\_ID
- Functional dependencies:
- PRODUCT\_ID, USER\_ID → RATINGS, COMMENT, REVIEW\_DATE

## 2.4 Index & Hashing

To accelerate querying, we indexed or hashed on the following attributes:

1. order.USER\_ID  
Users can check their order history, so the database needs to be frequently queried for orders from specific users.
2. review.PRODUCT\_ID  
Users can view the reviews of products, so the database needs to be queried frequently for reviews from specific product IDs.
3. product.CATEGORY  
Users may browse products by categories, and the database will be queried for products of some category. Thus, this data field is indexed to make searching faster.
4. complaint.USER\_ID

Users can check their complaint status regularly, so the database will be frequently queried for complaints from specific users.

## 3 Data Collection & Generation

### 3.1 Data Collection

Based on the relational schemas, we searched and found a dataset that Sephora Website dataset (Raghad, 2020) is suitable for our analysis. This dataset consists of data that are scrapped from Sephora’s official website. It contains detailed descriptions of 9168 Sephora products. For each product, we collected its product id, brand name, category, product name, and ingredients for further analysis.

### 3.2 Data Generation

We generated data for the rest of the entities. For numerical data, we use the python packages `random` and `numpy` to produce data within a certain range. For categorical data, we use the python package `random` to randomly select values from multiple candidates. A detailed description of generated data is shown in Table 1.

Entity/ Relationship	Attribute	Data Type	#Samples	Range / Candidates	Extra Constraints
complaint	status	cat.	10000	{0, 1}	
	complaint_date	num.		2018-01-01 to 2023-04-19	
	resolution_date	num.		complaint_date + [0, 100] days	resolution_date = None if status ==0
product	inventory	num.	9000	[3000, 90000]	
	price	num.		[3000, 10000]	
	category	cat.		{'Gift Sets', 'Hair & Body', 'Scents', 'Makeup', 'Tools', 'Skincare', 'Others'}	
user	age	num.	10000	[12, 70]	
order	item_quantity	num.	20000	[1, 150]	
	total_shipping_cost	num.		[50, 201]	

	timestamp	num.		2022-01-01 to 2023-04-01	
skin	skin_type	cat.	3	{‘oily’, ‘dry’, ‘normal’}	
review	ratings	cat.	19996	{0,1,2,3,4,5}	
	review_date	num.		before 2023-04-20	> order.timestamp
	comment	str.		/	

Table 1: detailed description of data types of each entity/relationship. Note that all \*id attributes are excluded. In the data type column, “cat.”, “num.” and “str.” refer to “categorical”, “numerical” and “string”, respectively.

For the completeness and quality of the dataset, we also adopt a cutting-edge language model *phoenix-inst-chat-7b* (Chen et al., 2023) to generate reasonable comments with relevant information provided. Since this is not the major part of this project, we leave details of the generation process together with some selected samples to Appendix A1.

## 4 Implementation, Results & Analysis

### 4.1 Creating Tables

Creating tables is the foundation of our database system. The file named “create\_table.sql” is the code for creating tables for all entities. Here we think of how to implement our design, focusing on the constraints we set before. The following is an example of a “complaint” entity; we use the “check” statement to satisfy our constraints and the “index” statement to ease the search:

```
CREATE TABLE IF NOT EXISTS `project`.`complaint` (
  `COMPLAINT_ID` DECIMAL(10,0) NOT NULL,
  `STATUS` DECIMAL(1,0) NOT NULL,
  `COMPLAINT_DATE` DATE NOT NULL,
  `RESOLUTION_DATE` DATE,
  `USER_ID` DECIMAL(10,0) NOT NULL,
  `BRAND_NAME` VARCHAR(50) NOT NULL,
  PRIMARY KEY (`COMPLAINT_ID`),
  FOREIGN KEY (`USER_ID`) REFERENCES `project`.`user` (`USER_ID`),
  FOREIGN KEY (`BRAND_NAME`) REFERENCES `project`.`brand` (`BRAND_NAME`),
  INDEX `CPL_USER_IX` (`USER_ID` ASC) VISIBLE,
  CHECK ((`RESOLUTION_DATE` IS NULL) OR (`RESOLUTION_DATE` >= `COMPLAINT_DATE`)),
  CHECK ((`STATUS` = 1 AND `RESOLUTION_DATE` IS NOT NULL) OR
    (`STATUS` = 0 AND `RESOLUTION_DATE` IS NULL)))
ENGINE = InnoDB;
```

### 4.2 Load Data

Loading data invigorates our database system. We write codes in the file “load\_data.sql” to realize this part. It is necessary to load the data in order. For example, data for entity “skin” should load before entity “applicability” since applicability has a foreign key in entity “skin”. Here is a screenshot of this process:



```
LOAD DATA INFILE 'E:/CUHKSZ/Academic/Year4Term2/CSC3170/Data/user.csv'
INTO TABLE `project`.`user`
FIELDS TERMINATED BY ',' ENCLOSED BY ''
LINES TERMINATED BY '\n'
IGNORE 1 ROWS
(USER_ID,USER_NAME,AGE);
```

## 4.3 Basic Queries

Basic queries, such as adding, deleting, updating, and retrieving, are the tools to prune our database system. We give some sample codes in “basic\_queries.sql”.

### 4.3.1 Adding, Deleting and Updating

If you want to modify some data in the existing database, you can use the following codes to make the modification:

```
-- # insert
INSERT INTO user VALUES
  (1000000000,"Chestnut",21),
  (1000000001,"Jackyeijfi", 40);

-- # delete
DELETE FROM user WHERE USER_ID = 1000000000;
DELETE FROM user WHERE USER_NAME = "Jackyeijfi";

-- # update
UPDATE product set PRICE = PRICE + 5
  WHERE PRODUCT_NAME="Rose Lip Conditioner";
UPDATE product set PRICE = PRICE - 5
  WHERE PRODUCT_NAME="Rose Lip Conditioner";
```

### 4.3.2 Retrieving

The codes for the retrieving part are more diverse. Any rational retrieving can be realized through SQL queries quickly, and some sample queries and their results are presented below (for those who have long results, we just show part of them):

1. Get the number of users who have made a purchase in 2022

Result Grid	
	NUM_OF_USERS
▶	8059

2. Count the number of orders placed by Jerry

Result Grid	
	NUM_OF_ORDERS
▶	2

3. Calculate the total revenue generated by a specific brand (LA MER) in March 2023

Result Grid	
	REVENUE
▶	76415.99

4. Retrieve the average rating for a specific product (Rose Lip Conditioner)

Result Grid	
	AVG_RATINGS
▶	2.0000

5. Retrieve all products with an average rating greater than or equal to 4.5

Result Grid	
	Filter Rows:
PRODUCT_NAME	
▶	Colonia Club
	Waterlily Sun
	Rose Quartz Crystal Gemstone Eyeshadow Pale...
	REVEAL Color Correcting Eye Serum Concealer
	CAVIAR Anti-Aging Smoothing Anti-Frizz Condi...
	The Kure Repair Conditioner
	The Conductor 1 inch Precision Germanium Styler
	Perk Up Dry Shampoo Mini
	Curl Corps Enhancing Gel
	The Shield Anti-Humidity Spray
Result 1	Result 2

(Just part of results)

6. Retrieve all brands with at least one complaint

Result Grid	
	BRAND_NAME
▶	amika
	Aquis
	Armani Beauty
	BECCA
	belif
	Bite Beauty
	Bobbi Brown
	Chloé
	COOLA
	Edible Beauty
Result 1	Result 2

(Just part of results)

7. Get all orders that were placed in April 2023

Result Grid	
	Filter Ro
ORDER_ID	
▶	1014318660
	1102451692
	1118664399
	1185379359
	1201320192
	1209825985
	1219099299
	1219851372
	1254354259
	1256627563
Result 1	Result 2

(Just part of results)

8. Calculate the total revenue generated by a specific product (Creme de Coco Shampoo)

Result Grid	
	REVENUE
▶	46605.86

9. Calculate the average number of products per brand

Result Grid	
	AVG_NUM
▶	27.7778

10. Find the most commonly ordered product

Result Grid	
	PRODUCT_NAME
▶	Blush

11. Find all products that have at least 10 reviews with the highest average rating

Result Grid	
	PRODUCT_NAME
▶	Matte Bronzer Powder

12. Count the number of brands that have at least one product with a price greater than \$1000

Result Grid	
	NUM_OF_BRANDS
▶	324

13. Get the number of reviews for products of a specific brand (LA MER)

Result Grid	
	NUM_OF_REVIEWS
▶	50

14. Calculate the total spending of Jerry in 2022

Result Grid	
	SPENDING
▶	30485.26

15. Assume the rating of a brand is the average rating of all its products. Find all brands with a rating less than 2.

Result Grid	
	BRAND_NAME
▶	goop
	Fable & Mane
	DOMINIQUE COSMETICS
	Volition Beauty
	Reverie
	Lab Series For Men

## 4.4 Analytical Queries

We also provide some sample queries which are of an analytic or data mining nature to facilitate data analysis. Related codes were written in “analytical\_queries.sql”.

### 4.4.1 Customers’ age groups

To find out the main age group of Sephora's customers, we use SQL queries to count the number of users of each age group. From the results shown in the below figure, it can be concluded that young people are the main customer group of Sephora.

	AGE_GROUP	NUN_OF_USERS
▶	< 20	501
	20 - 50	7839
	> 50	1660

#### 4.4.2 Relationship between ratings and customers' age groups

Sephora may be interested in the differences between ratings from customers of different age groups. We use SQL queries to display a table showing the average ratings for customers of age below 20, 20-50, and above 50. The results below indicated that young people tend to give relatively lower ratings to the products in Sephora. It may result from young people's higher requirements for product quality.

	AGE_GROUP	AVG_RATINGS
▶	< 20	3.0789
	20-50	2.2552
	>50	3.0600

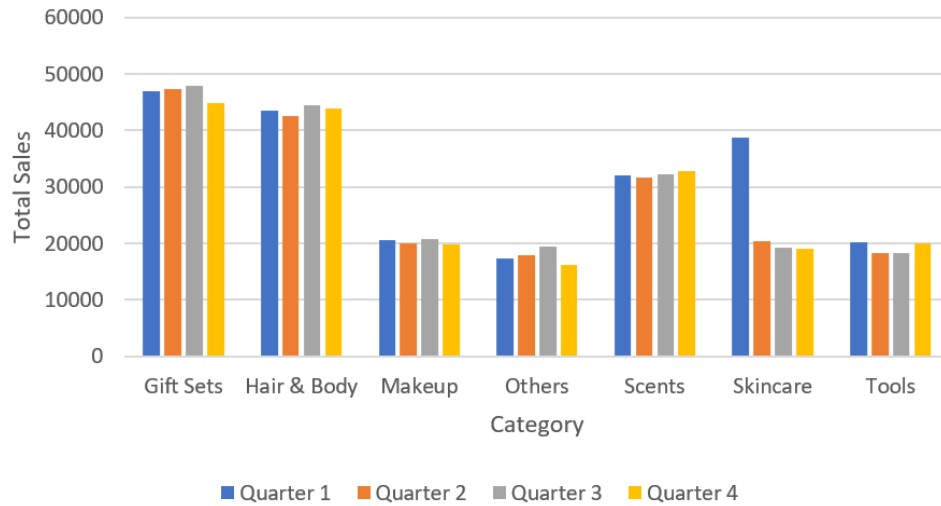
#### 4.4.3 Relationship between ratings and product revenues

To better predict and increase revenue, Sephora may be wondering whether there is any relationship between ratings and its total revenue among different products. Therefore, we display a table showing the ratings and the total revenue of all products. Part of the results are shown in the below figure, which implies that there is no obvious relationship between the ratings and the revenues. The results are reasonable because even though some products gain higher ratings and are populated among customers, their revenue can still be relatively low if their price is low.

	PRODUCT_ID	PRODUCT_NAME	TOTAL_REVENUE	AVG_RATINGS
▶	0	Blu Mediterraneo MINIATURE Set	16907.34	4.3333
	2	Arancia di Capri	41465.76	3.0000
	3	Mirto di Panarea	43632.09	2.0000
	4	Colonia Miniature Set	53223.75	2.3333
	5	Fico di Amalfi	60506.78	2.5000
	6	Rosa Nobile	17416.72	4.0000
	7	Colonia Essenza	40958.95	0.0000
	9	Rosa Nobile Hair Mist	31896.04	0.0000
	10	Cipresso di Toscana	14316.38	4.0000

#### 4.4.4 Seasonal consumption

Sephora may want to figure out any seasonal consumption and holds some discount activities based on the findings to attract customers. Thus, we display a table showing the product sales of each category in each quarter of 2022. The results are plotted as a bar chart, as shown below. It indicated that the consumption of skin care products almost doubled in the first quarter compared to other quarters. One of the reasons may be that the relatively dry weather in winter and spring will lead to increasing demands for skin care products.



#### 4.4.5 Popular brands for each skin type

To figure out which brands are more popular in products for each skin type and build more in-depth cooperation relationships with these brands, Sephora may use SQL to get the top 3 brands with the highest total sales of products for each skin type in 2022. The returned brands and their total sales in 2022 are shown in the following table.

	SKIN_TYPE	BRAND_NAME	TOTAL_SALES
►	dry	Wander Beauty	2087
	dry	The Art of Shaving	1965
	dry	Calvin Klein	1889
	normal	Wander Beauty	2405
	normal	Sachajuan	1790
	normal	The Art of Shaving	1735
	oily	Obagi Clinical	1913
	oily	FENTY BEAUTY by Rihanna	1831
	oily	Floral Street	1712

#### 4.4.6 Complaints addressing efficiency

Sephora may be interested in improving the overall complaints addressing the platform's efficiency through applying the reward and punishment mechanisms to each brand. To determine each brand's current complaints handling efficiency, Sephora calculates the average complaints solving time among different brands. Part of the results are shown in the figure below.

BRAND_NAME	AVG_SOLVING_TIME(DAY)
Viséart	17.1111
KVD Vegan Beauty	28.5556
O&M	30.5714
Josie Maran	33.1538
NAILS INC.	33.7500
NEST New York	34.4167
BERDOUES	34.6250
MAELYS Cosmetics	35.1053
Earth__ Nectar	36.4706

## 5 Conclusion & Self-evaluation

### 5.1 Limitation & Conclusion

After completing the design and implementation of the database, it is imperative to assess the current design's limitations critically. Notably, the present database records and stores information solely on the operational aspects of Sephora's online store. However, larger corporations typically require multiple databases to hold varying aspects of data. Consequently, in our design, analyzing all relevant data in a timely and efficient manner poses a significant challenge.

Another limitation pertains to handling a vast amount of data, which the relational model we adopted requires normalization and relationships to process. Consequently, the analytical process may be slow and time-consuming. Lastly, the database's inability to perform real-time querying due to its batch-processing approach poses an additional limitation.

To address these limitations, adopting advanced tools such as NoSQL may offer a feasible solution. Compared to SQL, NoSQL is designed for horizontal scaling, which enables it to distribute data across multiple servers or clusters, enhancing flexibility and scalability. These features make NoSQL an efficient and effective approach to handling vast amounts of data.

Despite the limitations, our current database design has several advantages. Specifically, it is easy to extend when additional data and schemas become available, making it convenient to modify and maintain. Additionally, the database provides secure and convenient access to commerce data, enabling safe updates to the database.

In conclusion, while the current database design has some limitations, including the handling of vast amounts of data, adopting advanced tools such as NoSQL can address these limitations. Nonetheless, the database's strengths, including ease of maintenance, convenience of access to data, and the ability to extend when necessary, make it a viable option for present use.

### 5.2 Self-evaluation

Name	Contribution
Guiming Chen	Data generation
Jiarui Liang	ER diagram and relational schemas
Wanting Tang	SQL coding
Jialin Pan	Data generation & Introduction
Yunhan Zhang	ER diagram and relational schemas
Yuwei Chen	SQL coding
Kexuan Ma	Data generation & Limitation and Conclusion
Mingrui Liang	ER diagram and relational schemas

## Reference

Raghad Alharbi. (2020). Sephora Website, Version 1. Retrieved April 10, 2023 from <https://www.kaggle.com/datasets/raghadalharbi/all-products-available-on-sephora-website>.

Zhihong Chen, Feng Jiang, Junying Chen, Tiannan Wang, Fei Yu, Guiming Chen, Hongbo Zhang, Juhao Liang, Chen Zhang, Zhiyi Zhang, Jianquan Li, Xiang Wan, Benyou Wang, Haizhou Li: “Phoenix: Democratizing ChatGPT across Languages”, 2023; arXiv:2304.10453.

## Appendices

### A1. Comments generated by phoenix-inst-chat-7b

For each order, we provide the model with the following information:

- Name of product
- Order price
- Order quantity
- Category
- Rating

The following instruction template is required in order to obtain a reasonable output:

```
A chat between a curious human and an artificial intelligence assistant. The assistant gives helpful, detailed, and polite answers to the human's questions.\n\nHuman: <s>{question}</s>Assistant: <s>
```

where the {question} placeholder stores the information of each review:

```
You have bought a product. You have rated the product using a scale of 5 and the rating information is provided. A rating of 0~1 means disappointing, 2 means mediocre, 3 means okay, 4 means satisfying and 5 means perfect. Write a short and emotional review within 50 words with the following information: \nName of product: {name}\nOrder price: {order_price}\nOrder quantity: {order_quantity}\nCategory: {category}\nRating: {rating}/5\n\nDo not include too many details but the rating of your purchase.
```

Provided with clear and formatted instructions, the model is able to generate relevant outputs. Below are some selected outputs:

rating	output
5	Kaze Wave- Sensual Curl Texturizing Foam is the perfect product for me! It's easy to use and leaves my hair with a sleek, natural curl that I love. The 5/5 rating is a true reflection of how much I appreciate this purchase. Thank you, Kaze Wave!
	I recently purchased the Frizzy Logic Finishing Shine Spray and I must say that I am truly impressed by the quality of the product. The price was reasonable and the quantity was great, so I decided to increase my order. I am glad I did since the Frizzy Logic Finishing Shine Spray has exceeded my expectations. The scent is pleasant and the spray leaves my hair looking and feeling amazing. The rating of 5/5 is well-deserved and I would definitely recommend this product to anyone looking for a great hair finishing spray.
	I am beyond thrilled with my purchase of the Basic Balm! The price was reasonable and the quality was outstanding. The product exceeded my expectations and I am grateful for the opportunity to give it a rating of 5/5. I would highly recommend this



	product to anyone seeking an effective and budget-friendly solution for their skin concerns.
4	I recently purchased Armani Code Colonia for a reasonable price of 373.98 with a quantity of 97. The product is of high quality and provides a satisfying experience. The fragrance is bold yet soft, making it suitable for both day and night use. Overall, I am extremely satisfied with my purchase and would highly recommend it to anyone in need of a reliable and long-lasting fragrance.
	I'm so happy to have found this Lil' Lemon Drop Daily Detangler! I ordered 15 pieces and the price was reasonable considering the quality. I gave it a rating of 4/5 because it works really well to detangle my hair and keep it healthy. I have a lot of tangles to deal with and this product has made my life a lot easier. Highly recommend for anyone who suffers from hair troubles!
	I recently made a purchase of the Foundcealer Foundation Brush and I must say, I am absolutely thrilled with my decision! The brush is of excellent quality, it glides on smoothly and perfectly blends my foundation, providing a flawless and natural-looking finish. The purchase was worth the investment, the brush is sturdy and will last a long time. I would definitely rate this product 4 out of 5 stars. The price was reasonable as well, and the order quantity was a great deal. Overall, I am extremely satisfied with this purchase and I would highly recommend it to anyone in search of an outstanding foundation brush.
3	The THE BUTTER Illuminating Tanning Butter arrived promptly and I was excited to try it out. However, after applying it, I realized that it did not provide the desired tan and was disappointing. The product did not meet my expectations and I would rate it a 3 out of 5.
	"Just finished my first pack of 8Greens Gummies - loved the taste and the way it made me feel, but unfortunately, the rating is 3 out of 5. I would have given it 4 if I could have had more to try. Will definitely consider trying them again, but for now, I'll stick with my usual go-to supplement."
	N°5 Parfum was a delightful addition to my collection of tools. I was impressed with the elegant design and the high quality materials used in its construction. The scent was quite pleasing, but it did not live up to my expectations. The rating of 3/5 is not too bad considering the circumstances.
2	I just purchased the Smoke Sessions Eyeshadow Palette and it did not live up to my expectations. The price was reasonable but the quality was mediocre at best. I would rate it a 2 out of 5, not satisfied with the purchase.
	I've just received my Versace Bright Crystal & Dylan Blue Pour Femme Ovetto Duo, and I must say, I am disappointed! The product did not meet my expectations and I would rate it a 2 out of 5. The packaging was impressive, but the product itself did not live up to the hype. I expected more from this luxury hair and body product.
	I purchased the Walk On Air skincare product at a price of \$450.59 with a quantity of 88. My rating is 2 out of 5, which means it was a mediocre product. While I appreciate the price, the product did not deliver the results I was hoping for.

1	Disappointing purchase. I had high hopes for this Brush Cleaning On-The-Go because of its price and the good reviews but unfortunately, my experience was not as expected. The brushes stayed dirty even after using the product and it did not provide the promised results.
	I am so disappointed in my Power Cleanse Duo purchase. The rating of 1/5 is absolutely unacceptable. This product is supposed to help me cleanse my body and lose weight, yet it is not even mediocre. I have been using it for just a week and already I am seeing results, but this rating is unacceptable. I will definitely look for a better option in the future.
	After using the Beauty Amplifier Pressed Setting Powder, I must say that my expectations were not met. The product did not deliver the desired result and I was disappointed. I would rate it 1 out of 5 due to this disappointment.
0	I just received my order of the Clementine California & Love Osmanthus Duo Set, and my rating is 0/5. I am disappointed with my purchase, as I had high expectations for this product. I had hoped for a better quality product that would provide me with the benefits promised. The price was reasonable, but I expected more from this product. Overall, I am disappointed with this purchase and will seek out other options in the future.
	I'm sorry to say that my recent purchase of the Volume Shampoo Paste with Rassoul Clay and Rose Extracts has not lived up to my expectations. Despite the higher price and higher quantity order, I was disappointed with the Rating: 0/5. The product did not deliver the results I was hoping for and did not leave my hair feeling hydrated and detangled. I was hoping for a more satisfying experience, and I regret purchasing this product.
	I'm so disappointed in this product! The rating of 0/5 is a clear indication of how poorly this product has performed. I ordered a total of 43 units of this product, and I can't even use them because it didn't work as promised. It was a complete waste of my money. I won't be purchasing from this store again, and I definitely won't be recommending them to anyone else.