Business Data Management (BDM)

Project Application

Target User: Zalora

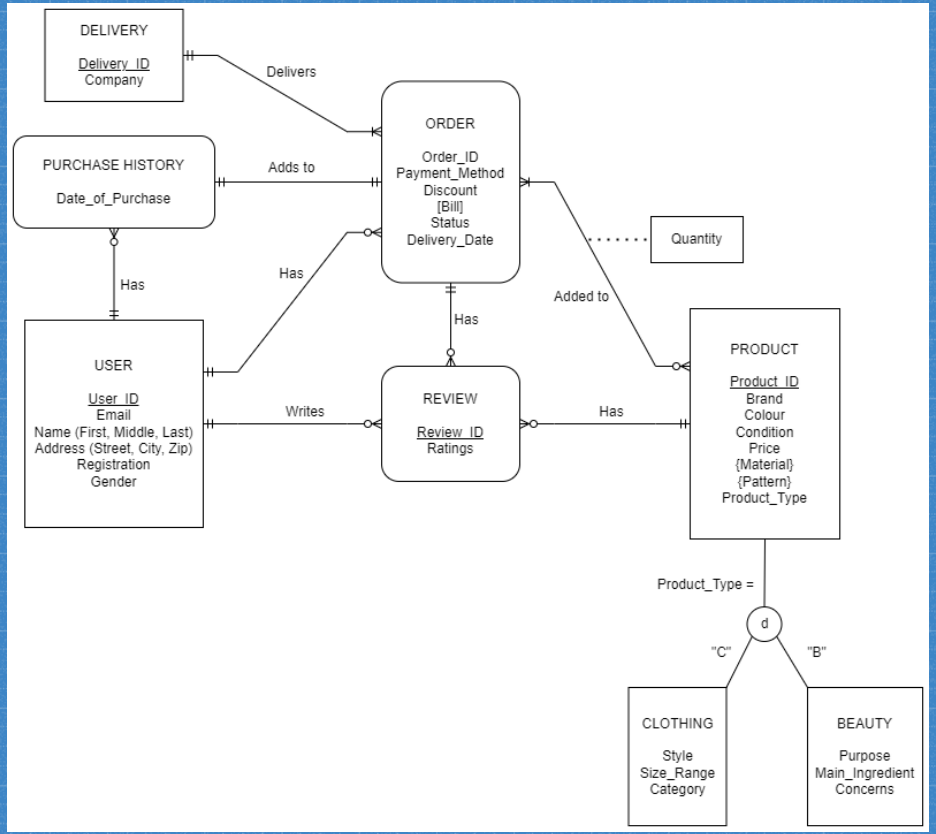
Aim: To successfully compile customer and product data into a database and use queries to identify trends or necessary data

Business requirements

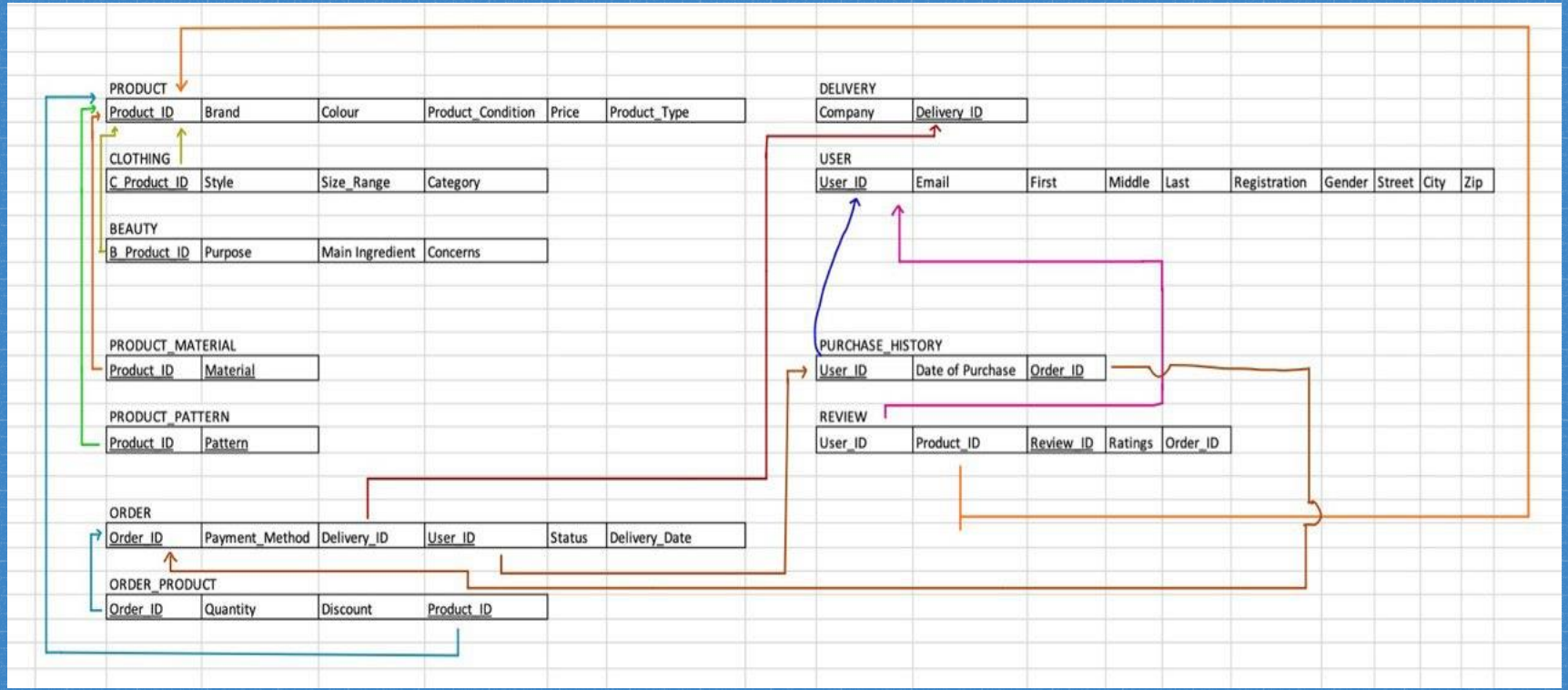
1. A user can have **one address** associated with themselves. The date that the user signs up on is also logged.
2. Whenever an order is made by the user, the data would be stored in the purchase history, which contains the number of purchases and the date purchased.
3. The order would also be attached to the user’s **associated address.**
4. A review contains a unique review ID and ratings.

ER Model

1. Wishlist was removed
2. Removed clothing subtypes (Adult+Children)
3. Removed supplier due to the lack of meaningful attributes (changed supplier/brand to Brand attribute under Product)
4. Created Purchase History associative entity (removed purchase history as an attribute under user_data)
5. Changes to user address format
6. Quantity is a relationship attribute between ORDER and PRODUCT



LD Model (updated)



QUERY 1

Which product type & brand is the most popular?

```
select p.brand, p.product_type, AVG(rd.ratings) as average_ratings
from product p
inner join review rd
on p.product_id = rd.product_id
group by p.brand
order by average_ratings desc;
```

	brand	product_type	average_ratings
▶	BOURJOIS	Beauty	5.0000
	Sleep No More	Clothing	5.0000
	Versace	Beauty	5.0000
	ADIDAS	Clothing	4.5000
	Nike	Clothing	4.0000
	Abercrombie & Fitch	Clothing	4.0000
	Twenty Eight Shoes	Clothing	3.5000
	H&M	Clothing	3.5000
	Moschnino	Clothing	3.0000
	Under Armour	Clothing	3.0000
	Dsquared2	Clothing	1.5000
	Clinique	Beauty	1.5000

Top Brands include BOURJOIS, Sleep No More, Twenty Eight Shoes, Versace that have a 5/5 rating

- Continue to offer these brands on Zalora

Brands with low ratings include Dsquared2 and Clinique

- Discontinue once contract ends or haggle for lower price

OUTPUT

QUERY 2

Most common material used in clothing / Number of products that use cotton and how much

```
select pm.material, count(distinct p.product_id) as num from product_material pm, product p
where p.product_id = pm.product_id and pm.material like "%cotton%"
group by material;
```

	material	num
▶	100% Cotton	4
	100% organic cotton	3
	95% Cotton	1
	99% Cotton	1
	Cotton Blend	1

We found that a majority of users purchased clothes made out of 100% Cotton and 100% Organic cotton.

- The 100% keyword seems to resonate with users.
- We propose that Zalora looks into increasing supply for such products.

OUTPUT

QUERY 3

Which payment method is most commonly used for orders > $50?

```
select payment_method, count(order_id) as number
from `order`
where order_id in
(
select op.order_id
from product p
join order_product op
on p.product_id = op.product_id
where p.price*(1-ifnull(op.discount,0))*op.quantity > 50
)
group by payment_method
order by number DESC;
```

payment_method	number
Visa	5
Cash	3
Mastercard	2
GrabPay	2

Most common payment method is Visa

- Provide discounts or incentives targeted at Visa users may drive up sales as users are likely to spend more

OUTPUT

QUERY 4

Sales volume of clothing vs beauty products by month

```
select month(date_of_purchase) as date_month, count(case p.product_type when "clothing" then 1 else null end) as no_of_clothing,
count(case p.product_type when "beauty" then 1 else null end) as no_of_beauty, count(p.product_id) as total
from purchase_history ph
inner join order_product op on ph.order_id = op.order_id
inner join product p on op.product_id = p.product_id
group by date_month
order by date_month;
```

	date_month	no_of_clothing	no_of_beauty	total
▶	1	1	2	3
	2	2	0	2
	3	3	0	3
	4	1	0	1
	5	1	0	1
	6	1	0	1
	9	2	0	2
	10	1	0	1
	11	1	0	1
	12	5	2	7

Sales tend to peak around December, which is expected due to Christmas

Clothing product sales are significantly higher than beauty product sales, especially in December

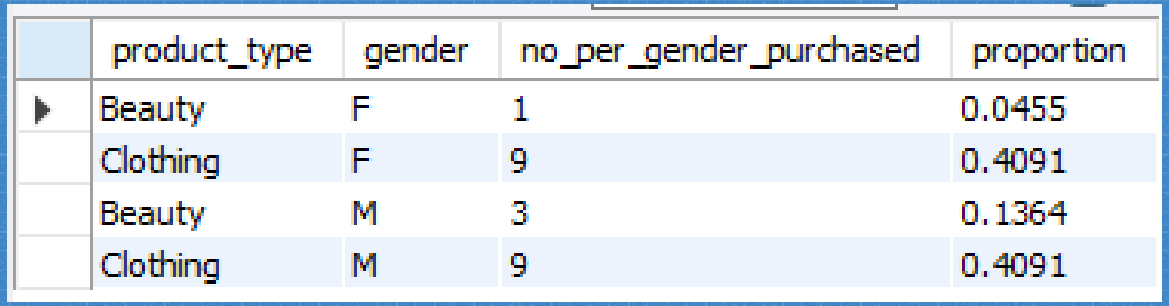
- Provide discounts for the middle of the year to increase sales in that time frame
- Increase marketing effort for beauty products

OUTPUT

QUERY 5

Gender proportion of user purchases for each product

```
select d.product_type, u.gender, count(u.gender) as no_per_gender_purchased, count(u.gender) / sum(count(u.gender)) over() as proportion
from `user` u
inner join `order` o on o.user_id  = u.user_id
inner join order_product p on p.order_id = o.order_id
inner join product d on d.product_id = p.product_id
group by u.gender, d.product_type
order by u.gender
;
```



	product_type	gender	no_per_gender_purchased	proportion
▶	Beauty	F	1	0.0455
	Clothing	F	9	0.4091
	Beauty	M	3	0.1364
	Clothing	M	9	0.4091

OUTPUT

- By segregating each product type by gender, as well as their respective counts, we obtain this result.
- We observe how interestingly, the male customers purchased more clothing/ beauty products.
- We propose that Zalora increase their marketing efforts towards their male audiences

QUERY 6

Which delivery service is the most efficient?

```
select d.Company, AVG(DATEDIFF(o.Delivery_Date,p.Date_of_Purchase)) as difference_days
from Delivery d,`Order` o,Purchase_History p
where o.Order_ID=p.Order_ID
and o.Delivery_ID=d.Delivery_ID
GROUP BY d.Company
ORDER BY difference_days desc;
```

Company	difference_days
S.F. Express Singapore	19.5000
NinjaVan	17.0000
SingPost	16.7500
Yamato Transport Singapore	15.3333
Janio Asia	15.2000

OUTPUT

- Janio Asia is the most efficient delivery service as it takes the least average number of days (15.2 days) for the delivery to be completed
- S.F Express Singapore is the least inefficient delivery service as it has the highest average number of days (19.5 days) for delivery to be completed
- Zalora should focus on Janio Asia’s delivery service to provide better customer experience

Lessons Learnt

- Data needs to be formatted properly to ease the data analysis process
  - Need to think carefully about the requirements of the database before data collection
- Importance of order of statements in MySQL
  - Eg. create parent tables first
  - What we did was to find the entities that had the least links in the ER diagram and create those tables first such as USER_DATA, PRODUCT, DELIVERY
- Additional ways in which we can write SQL queries to provide insight using our data
- The reasoning behind logical diagrams to ensure the integrity of our tables.