Kimberly Bastien

Yash Jain

Rahul Sivaram

Linxi Sun

7 October 2021

CIS 505: Footwear Manufacturer Team Project

Identify the Problem and Relevant Data

Our client, a major footwear manufacturer, is requesting our team to provide advanced analytics in order to improve its business brand through customer relationships. Lacking information about their customers, our client is seeking our help to design an appropriate data infrastructure that helps to address our main question—***"How can our client identify its consumer base to meet customer requirements and expectations in order to improve customer acquisition and retention?"***

Identify the Benefits and Insights

Addressing the question above would benefit our client's ability to identify, understand, and meet brand and product expectations of the customer base. In turn, this would improve our client's ability to expand within the market, generate revenue, and increase return on investment. There are many factors that will determine how to best identify a customer base, meet customer's brand and product expectations, and improve customer acquisition and retention; therefore, our team has developed sub-questions to provide further insight.

1) ***Which locations promote highest sales for each shoe type?***

This question will provide insight into how our client can better gauge demand and inventory for each shoe type. Additionally, evaluating location decisions will

increase our client's company's long-term performance; for example, in terms of targeted opening of stores in the future.

2) ***How to expand the customer portfolio, by curating shoes that focus on consumer's intended wear, price points and feedback, in order to optimize our client's manufacturing and design decisions to meet market expectations and demand.***

This question would be further narrowed down by design (graphics, shapes, colors) and comfort (materials, sizing options) based on location and product category. This question would allow for our client to understand its consumer base's feedback in order to increase our client's chances of reaching a wider audience and purchasing market, as well as optimize its manufacturing process to meet market expectations and demand; effectively making footwear in designs that its consumers truly want to own.

3) ***How to optimize the customer rewards program.***

This would include discounts based on purchase location, purchase quantity, age, or season. This question provides insights on how our client can understand and improve its customer rewards program effectiveness. This will benefit our client since a memorable user experience increases sales, number of customers, and customer loyalty, and also helps refine the levels of rewards it offers to various demographics of customers in an attempt to maximize their lifetime value.

4) ***How to gain an understanding of current market trends.***

The goal of this question would be to measure competitor data in order to compare popularity and pricing of products by location and time (seasonality). This would provide our client with valuable insight on marketplace trends and bring the

advantage of discovering opportunities where our client can differentiate its products and services.

5) ***How to understand the role of social media.***

Our team will focus on social media metrics, such as the number of likes and shares on product posts, along with its impact on revenue and units sold. Understanding our client's social media performance can offer insight into how to best reach and inform its audience about our client's brand and products.

Identify the High Level Data Sources

Our team will be utilizing multiple data sources for this project, including:

1) ***Transaction Data***

   This will be acquired from all currently available transactional databases, such as e-commerce and partner transaction data

2) ***Feedback & Review Data***

   This will be gathered from survey and feedback forms.

3) ***Social Media Data***

   This will be collected from posts made on the client's social media platform.

4) ***Market & Competitor Data***

   This will be gathered via subscription through various retail channels.

Database Context & Assumptions

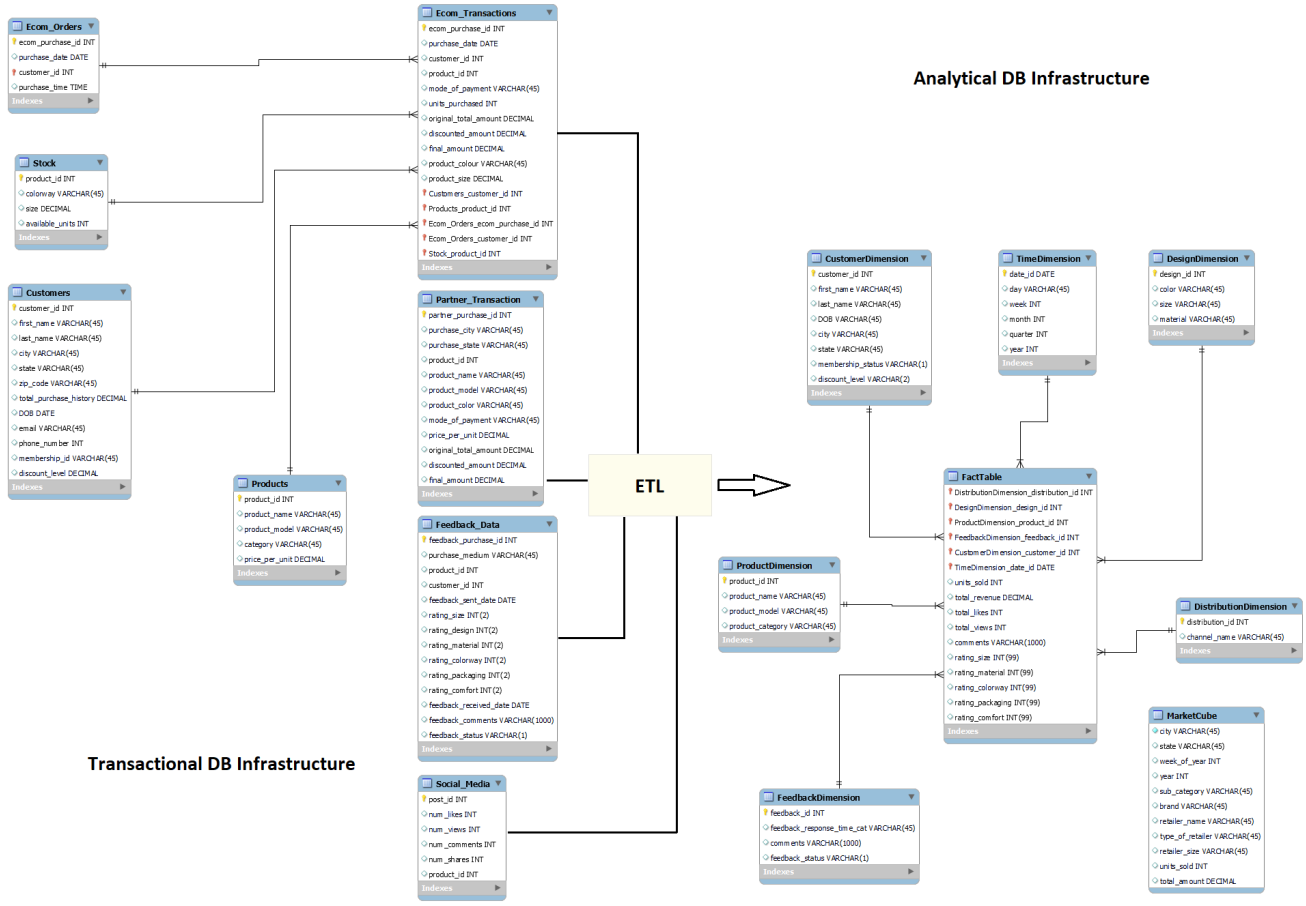
Through the creation of the database, the following context and assumptions can be made:

- **Overall Database Architecture:** The database schema is split into the transactional, ETL, and analytical infrastructures. The transactional infrastructure (e-commerce,

partner, feedback, and social media) goes into the ETL process, where the attributes and dimensions are arranged logically together into the analytical warehouse. For example, all of the customer data from the e-commerce and partner table gets consolidated into the ETL process and flows into the customer dimension of the data warehouse. An additional example is that the sales information from the e-commerce and the partner table, which gets consolidated into the ETL process and is put into the fact table. Moreover, the market data is aggregated and then put into the analytical warehouse as a cube.

- **Discount Program**: Our client's company offers a membership and customers must pay a one-time fee in order to become a member. Discounts are only provided to customers who are members. There are discount levels for members, which depends on the total purchase history; therefore, it can be concluded that each time a member makes a purchase, the data system is updated to track the member's purchases.
- **Social Media:** Our client's only social media platform is Facebook. Our team mines the client's social media platform in order to track the activity of posts at an end consumer level, which is consolidated in the ETL process associated with a customer in the analytical infrastructure.
- **Feedback:** After every purchase, a customer is sent a feedback/survey form, and a record is created for it in the database. If the customer fills out the feedback/survey form, the record is updated accordingly.

Database Schema



Queries

1) Which locations promote highest sales for each shoe type?

Query:

```
select
t.`year`,t.`quarter`,c.state,c.city,p.product_category,p.product_model,sum(f.units_sold),su
m(f.total_revenue)
from FactTable f inner join CustomerDimension c on
f.CustomerDimension_customer_id=c.customer_id
inner join ProductDimension p on p.product_id=f.ProductDimension_Product_id
inner join TimeDimension t on t.date_id=f.TimeDimension_date_id
```

```
group by t.`year`,t.`quarter`,c.state,c.city,p.product_category,p.product_model;
```

The results of this query will allow the client to view the total units sold and the revenue generated by each year and each quarter by location, category, and model.

2) How to expand the consumer portfolio by curating shoes that focus on consumer's intended wear and price points in order to optimize our client's manufacturing to meet market expectations and demand.

Query:

```
select
p.product_category,p.product_model,d.color,d.size,d.material,c.state,c.city,avg(f.rating_si
ze),avg(f.rating_material),avg(f.rating_colorway),avg(f.rating_packaging),avg(f.rating_c
omfort),sum(f.units_sold)
from FactTable f inner join ProductDimension p on
p.product_id=f.ProductDimension_Product_id
inner join DesignDimension d on d.design_id=f.DesignDimension_design_id
inner join CustomerDimension c on f.CustomerDimension_customer_id=c.customer_id
group by p.product_category,p.product_model,d.color,d.size,d.material,c.state,c.city;
```

The results of this query will allow the client to view the popularity of products by various dimensions (e.g.: size, color, material, packaging, and comfort) in terms of ratings from feedback and units sold.

3) How to optimize the customer rewards program?

Query:

```
select c.discount_level,t.`year`,t.`quarter`,c.state,c.city,
case
        when timestampdiff(year,c.dob,curdate()) < 18 then 'Below 18'
       when timestampdiff(year,c.dob,curdate()) < 26 then '18-25'
```

```
        when timestampdiff(year,c.dob,curdate()) < 36 then '26-35'
        when timestampdiff(year,c.dob,curdate()) < 46 then '36-45'
        when timestampdiff(year,c.dob,curdate()) < 56 then '46-55'
        when timestampdiff(year,c.dob,curdate()) < 66 then '56-65'
        else '65+'
    end as AgeCategory,sum(f.units_sold),sum(f.total_revenue)
from FactTable f inner join CustomerDimension c on
f.CustomerDimension_customer_id=c.customer_id
inner join TimeDimension t on t.date_id=f.TimeDimension_date_id
group by c.discount_level,t.`year`,t.`quarter`,c.state,c.city,AgeCategory;
```

The results of this query will allow the client to view the effectiveness of the discount program from the viewpoint of the customer's age, location, and purchasing period.

4) How to gain understanding of current market trends?

Query:

```
With MarketAnalysis as (
select
`year`,week_of_year,state,city,sub_category,sum(units_sold)'Market_units_sold',sum(tota
l_revenue)'Market_total_revenue'
from MarketCube
group by `year`,week_of_year,state,city,sub_category
),
ClientAnalysis as
(
select
t.`year`,t.`week`,c.state,c.city,p.product_category,sum(f.units_sold)'Client_units_sold',su
m(f.total_revenue)'Client_total_revenue'
from FactTable f inner join CustomerDimension c on
f.CustomerDimension_customer_id=c.customer_id
```

```
inner join TimeDimension t on t.date_id=f.TimeDimension_date_id
inner join ProductDimension p on p.product_id=f.ProductDimension_Product_id
group by t.`year`,t.`week`,c.state,c.city,p.product_category
)
select
c.`year`,c.`week`,c.state,c.city,c.product_category,c.Client_units_sold,c.Client_total_reve
nue,m.Market_units_sold,m.Market_total_revenue
from ClientAnalysis c inner join MarketAnalysis m on c.`year`=m.`year`
and c.`week`= m.week_of_year and c.state=m.state
and c.city=m.city and c.product_category=m.sub_category;
```

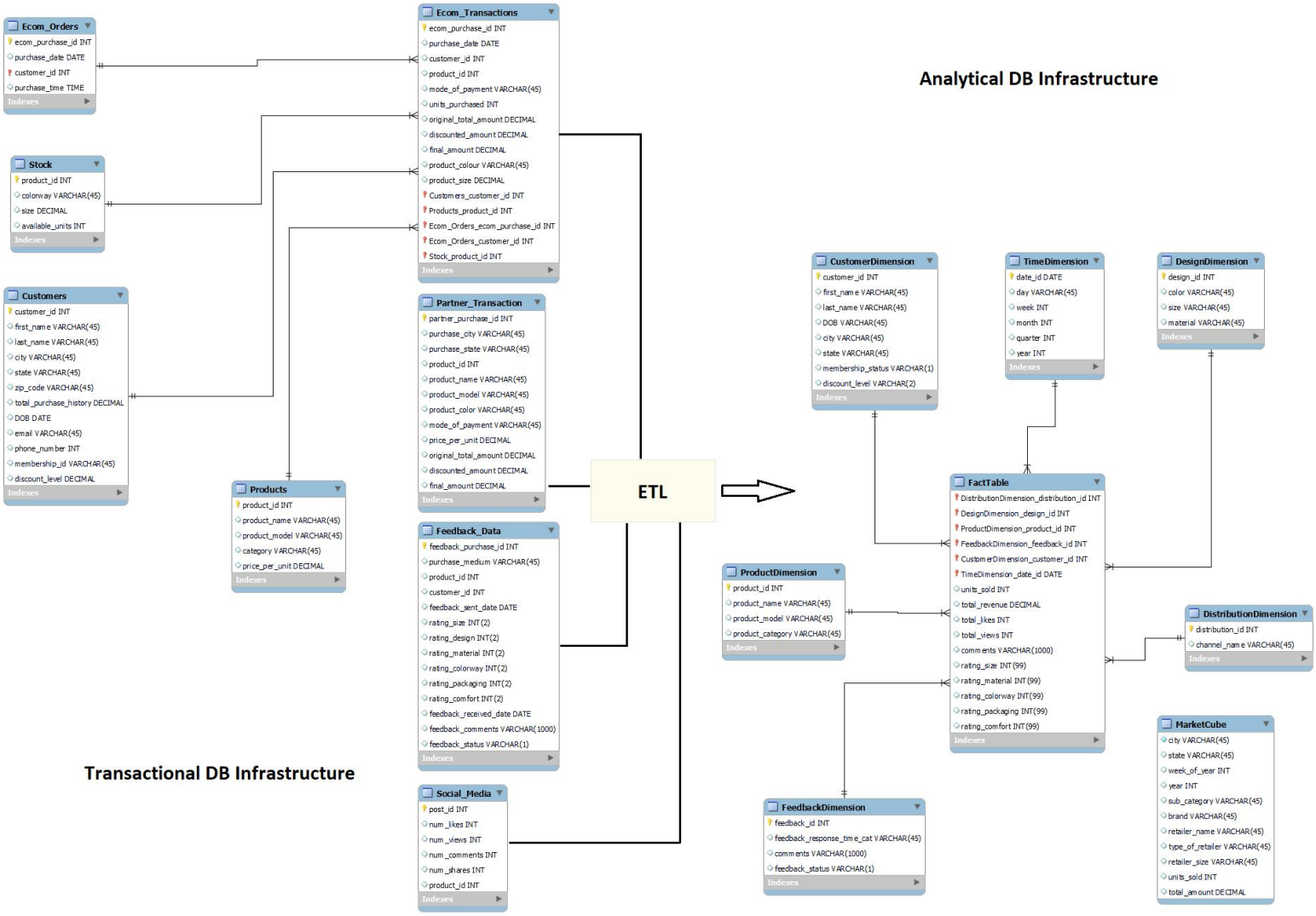
The results of this query will allow the client to compare its product sales performance to the competitor's & market's product sales performance at the product category, and time level.

5) How to understand the role of social media?

Query:

```
select
p.product_category,p.product_model,sum(f.total_views)'Total_views',sum(f.total_likes)'T
otal_likes',
sum(f.units_sold) 'Total_units_sold',sum(total_revenue)'Total_revenue' from FactTable f
inner join ProductDimension p on p.product_id=f.ProductDimension_Product_id
group by p.product_category,p.product_model;
```

The results of this query will allow the client to analyze the number of units sold and revenue versus the engagement on its social media platform in order to determine the effectiveness of the client's social media advertising.


Ecom_Orders
ecom_purchase_id INT
purchase_date DATE
customer_id INT
purchase_time TIME
Indexes
Ecom_Transactions
ecom_purchase_id INT
purchase_date DATE
customer_id INT
product_id INT
mode_of_payment VARCHAR(45)
units_purchased INT
original_total_amount DECIMAL
discounted_amount DECIMAL
final_amount DECIMAL
product_colour VARCHAR(45)
product_size DECIMAL
Customers_customer_id INT
Products_product_id INT
Ecom_Orders_ecom_purchase_id INT
Ecom_Orders_customer_id INT
Stock_product_id INT
Indexes
Stock
product_id INT
colorway VARCHAR(45)
size DECIMAL
available_units INT
Indexes
Customers
customer_id INT
first_name VARCHAR(45)
last_name VARCHAR(45)
city VARCHAR(45)
state VARCHAR(45)
zip_code VARCHAR(45)
total_purchase_history DECIMAL
DOB DATE
email VARCHAR(45)
phone_number INT
membership_id VARCHAR(45)
discount_level DECIMAL
Indexes
Products
product_id INT
product_name VARCHAR(45)
product_model VARCHAR(45)
category VARCHAR(45)
price_per_unit DECIMAL
Indexes
Partner_Transaction
partner_purchase_id INT
purchase_city VARCHAR(45)
purchase_state VARCHAR(45)
product_id INT
product_name VARCHAR(45)
product_model VARCHAR(45)
product_color VARCHAR(45)
mode_of_payment VARCHAR(45)
price_per_unit DECIMAL
original_total_amount DECIMAL
discounted_amount DECIMAL
final_amount DECIMAL
Indexes
Feedback_Data
feedback_purchase_id INT
purchase_medium VARCHAR(45)
product_id INT
customer_id INT
feedback_sent_date DATE
rating_size INT(2)
rating_design INT(2)
rating_material INT(2)
rating_colorway INT(2)
rating_packaging INT(2)
rating_comfort INT(2)
feedback_received_date DATE
feedback_comments VARCHAR(1000)
feedback_status VARCHAR(1)
Indexes
Social_Media
post_id INT
num_likes INT
num_views INT
num_comments INT
num_shares INT
product_id INT
Indexes
ETL
Transactional DB Infrastructure
Analytical DB Infrastructure
CustomerDimension
customer_id INT
first_name VARCHAR(45)
last_name VARCHAR(45)
DOB VARCHAR(45)
city VARCHAR(45)
state VARCHAR(45)
membership_status VARCHAR(1)
discount_level VARCHAR(2)
Indexes
TimeDimension
date_id DATE
day VARCHAR(45)
week INT
month INT
quarter INT
year INT
Indexes
DesignDimension
design_id INT
color VARCHAR(45)
size VARCHAR(45)
material VARCHAR(45)
Indexes
FactTable
DistributionDimension_distribution_id INT
DesignDimension_design_id INT
ProductDimension_product_id INT
FeedbackDimension_feedback_id INT
CustomerDimension_customer_id INT
TimeDimension_date_id DATE
units_sold INT
total_revenue DECIMAL
total_likes INT
total_views INT
comments VARCHAR(1000)
rating_size INT(99)
rating_material INT(99)
rating_colorway INT(99)
rating_packaging INT(99)
rating_comfort INT(99)
Indexes
ProductDimension
product_id INT
product_name VARCHAR(45)
product_model VARCHAR(45)
product_category VARCHAR(45)
Indexes
DistributionDimension
distribution_id INT
channel_name VARCHAR(45)
Indexes
FeedbackDimension
feedback_id INT
feedback_response_time_cat VARCHAR(45)
comments VARCHAR(1000)
feedback_status VARCHAR(1)
Indexes
MarketCube
city VARCHAR(45)
state VARCHAR(45)
week_of_year INT
year INT
sub_category VARCHAR(45)
brand VARCHAR(45)
retailer_name VARCHAR(45)
type_of_retailer VARCHAR(45)
retailer_size VARCHAR(45)
units_sold INT
total_amount DECIMAL