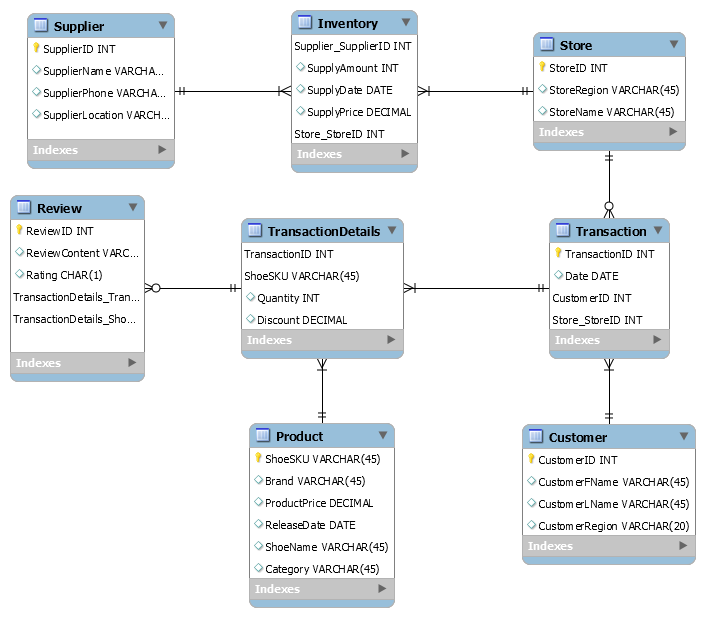
**Database Schema**



*Note: We put all the stores in the transaction table: Store\_StoreID, including all the online stores and retailing stores, so that no extra attributes are needed. If a purchase is made online, the store ID is coded as ‘1’.*

Tracking transaction information is vital to recognizing customer behaviors and identifying trends within the data. We can see how many shoes were purchased by brand, category (shoe style), and shoe SKU. Knowing what date and within what region shoes were purchased is necessary in order to group transactions and determine patterns in shoe popularity. Knowing who our customers are and where they are making their purchases is important when our goal is to track consumer buying habits. We would also like to have data on our retailers and suppliers, so that we may see which shoes are purchased and restocked often. All of this data will allow us to better understand our customers and make more informed business decisions in the future.

**Sample Queries**

1. What are the internet sales behaviors of Reebok in the northeast region during one month when ‘Nike’ has released its latest product one month (30 days) before?

select sum(p.ProductPrice\*td.Quantity\*td.Discount) as “Revenue”

from TransactionDetails td, Product p, Store s,Transaction t

where p.shoeSKU=td.shoeSKU and t.TransactionID=td.TransactionID

and s.StoreID=t.Store\_StoreID and p.Brand= ’Reebok’ and c.CustomerRegion=”northeast” and s.StoreID=1 and

(t.date between (date(select max(date(ReleaseDate) ) from Product where Brand=’Nike’)) and (date(select max(date(ReleaseDate) ) from Product where Brand=’Nike’)+30));

This query allows us to see what the sales of Reebok in the Northeast region looked like after Nike released a new product. We can determine if customers are more likely to purchase the new product during the month after the release, thus impacting our sales. If we recognize that our competitor’s shoes sold more than our brand’s shoes, our company might increase marketing and/or have special promotions or discounts on our shoes to entice consumers.

1. How many customers have bought multiple pairs of shoes last year?

select customerId,sum(td.Quantity) as ‘Purchasing Count’

from Customer c, Transaction t, TransactionDetails td

where t.TransactionID=td.TransactionID and t.CustomerId=c.customerID

group by CustomerID

having sum(td.Quantity)>1;

This query shows us which customers purchase many shoes, regardless of brand. These customers have the potential to be top customers for our brand, and we can market our shoes accordingly to them.

1. Which shoe is the most popular (highest purchase quantity) in each region? Which brand?

select max(TD.Quantity), P.ShoeSKU, P.ShoeName, P.Brand, S.Region

from TransactionDetails TD, Product P, Store S, Transaction T

where P.ShoeSKU=TD.ShoeSKU and TD.transactionID=T.transactionID and T.StoreID= S.StoreID

group by S.Region, P.Brand

order by Quantity DESC limit 1;

This query will help guide our supply and inventory decisions. If we know that a certain shoe is purchased most often in a certain region, then we need to make sure to maintain enough inventory of that shoe within that region. We can also see which brands are preferred by region, which allows us to focus marketing efforts on the regions where our shoes are not as popular.

1. Which type of shoe are we currently supplying the most of?

select p.Category, sum(i.SupplyAmount) as Supply\_Amount

from Product p, TransactionDetails td, Transaction t, Store s, Inventory i

where i.StoreID=s.StoreID and s.StoreID=t.StoreID and t.TransactionID=td.TransactionID and

td.ShoeSKU=p.ShoeSKU

group by p.Category

order by Supply\_Amount DESC limit 1;

We would like to know which type of shoes (sneakers, sandals, etc) we are supplying the most so that we can compare to the type of shoe that is most popular. These values should be consistent with each other if Reebok is making smart supply and inventory decisions. If we find that our most-supplied shoe is not as popular among customers, our company needs to make adjustments in order to decrease costs and increase revenues.

1. What retailer performs the best (having the best selling record) in our four regions last year and what is their location?

select s. StoreRegion,s.StoreID, sum(p.Productprice\*td.Quantity\*td.Discount) as Revenue

from TransactionDetails td, Product p, Store s,Transaction t

where p.ShoeSKU=td.ShoeSKU and t.TransactionID=td.TransactionID

and s.StoreID=t.Store\_StoreID

group by s.storeregion

order by Revenue DESC limit 1;

Knowing which retailer performs the best by region gives us insight into our customer’s buying habits. If we know which stores are the most frequented by our customers, we can infer the areas within each region where our top customers are located. Additionally, we can see whether our shoes sell better within high-end stores, specialty stores, or another type of store (using general field knowledge about the different retailers). We can also investigate how our shoes are displayed and marketed in the most popular retailer, and make changes within other stores to boost customer purchases.

1. Which region produced the highest revenue for our company in the third quarter of last year?

select s.StoreRegion,sum(p.ProductPrice\*td.Quantity\*td.Discount) as Revenue

from TransactionDetails td, Product p, Store s,Transaction t

where p.shoeSKU=td.ShoeSKU and t.TransactionID=td.TransactionID

and s.StoreID=t.Store\_StoreID and

YEAR(t.date)=2018 and (month(t.date) between 7 and 9)

group by s.StoreRegion

order by “Revenue” DESC limit 1;

This query shows us the region in which our shoes are most popular by revenue. We can also investigate the revenue by region in the other quarters to determine if there are any seasonality trends when it comes to our customer’s buying habits. This information gives us insight into when and where we can seek improvement in our sales.

1. Who are the top 50 customers and what % of sales they contribute to?

select c.CustomerID, c.CustomerFName, c.CustomerLName, sum(p.ProductPrice\*td.Quantity\*td.Discount) as Revenue

from Customer c, Transaction t, TransactionDetails td

where c.CustomerID= t.CustomerID and t. TransactionID=td.TransactionID and p.ShoeSKU=td.ShoeSKU

group by CustomerID

order by Revenue DESC

limit 50;

select SUM(p.ProductPrice\*td.Quantity\*td.Discount) as Revenue

from Product p

join TransactionDetails td on td.shoesku=p.ShoeSKU;

Knowing who our top 50 customers are and recognizing how much they contribute to our overall sales can allow us to investigate trends among those customers. If these customers come from a similar region or have a similar background, we can figure out which subset of the population prefers our brand’s shoes. This can show us who is likely to purchase our shoes in the future, and from this information, we can identify target customers to market our shoes to.

1. How many products have 5 star ratings and happen to be the top selling product?

select p.ShoeSKU, p.ShoeName, td.Quantity

from Review r, Product p, TransactionDetails td

where r.TransactionDetails\_TransactionID=td.TransactionID and td.ShoeSKU=p.ShoeSKU

and r.Rating=’5’

group by p.ShoeSKU

order by td.Quantity DESC;

9. Is the rating and the product sales related?

select avg(rating) as ‘Rating Average‘, sum(p.ProductPrice\*td.Quantity\*td.Discount) as Revenue, p.ShoeName

from Product p, TransactionDetails td, Review r

where td.ShoeSKU=p.ShoeSKU and r.TransactionDetails\_TransactionID=td.TransactionID

group by ShoeSKU

order by sum(p.ProductPrice\*td.Quantity\*td.Discount) DESC;

The query tells us the products with the best ratings and enables us to compare the ratings to the revenue. If, for example, Reebok finds that the highest-rated shoes do not contribute to the greatest amount of revenue, we can then further investigate the reasons behind this difference.