CIS 505 Team Project

Group A2

Updated: 09/28/2020

Team Members

- **Chaithanya Muninagappa Sridhara Murthy**
- **Jinghuan Li**
- **Nathaniel Englert**
- **Suparva Paruthy**

Table of Contents

Executive Summary

This report contains the analysis for a shoe manufacturer who wants to maximize sales and optimize costs along with understanding the customer's profiles. We focus on deriving insights by answering a few relevant questions via querying the available databases obtained after conducting data requirement analysis to gather relevant information. We need to perform advanced analytics which includes modifying, aggregating and optimizing the (attributes) data to be included in our analytical database and then execute our queries and analyze the outcome to provide useful insights to our client company for making informed business decisions.

One of the problems we faced in this modelling project was the inability to gather all relevant information about the customer and the purchases from our client. However with all the available data we draw ER diagrams and analyze relationships and cardinalities before investigating the database schema the client already uses to store the transactional data. We consider our client to be Nike in this project.

Introduction

Nike was founded in 1964 and currently has an evaluation of $34.8B[1] with offices in more than 45 countries and occupies more than 31%[2] of the global athletic footwear market. Nike management makes all business decisions based on real data based on sale numbers, revenue etc. Our goal is to maximize the revenue earned by Nike for its stores and minimize the cost of production using the analytical insights we derive from the provided data supplementing with the public data.

Background of the data: Customer entity has all the personal information of the customer which includes customer id, first name, last name, emailId among others. Product entity has product id, name, gender as attributes which belongs to category entity whose attributes are category id and name and also has a relationship with Reviews entity where we collect aggregated data of positive and negative reviews along with average rating of the products if any. The Customer places an order, this order information is gathered in an order entity whose attributes are order id, shipping details, productid, quantity, color, price among others. For simplicity we accommodate all the data about products sold in brand-stores, partner retail stores, online platforms, and subscribed stores in a single entity named store whose attributes are store id, name, type, location among others.

Entity-Relationship Diagram

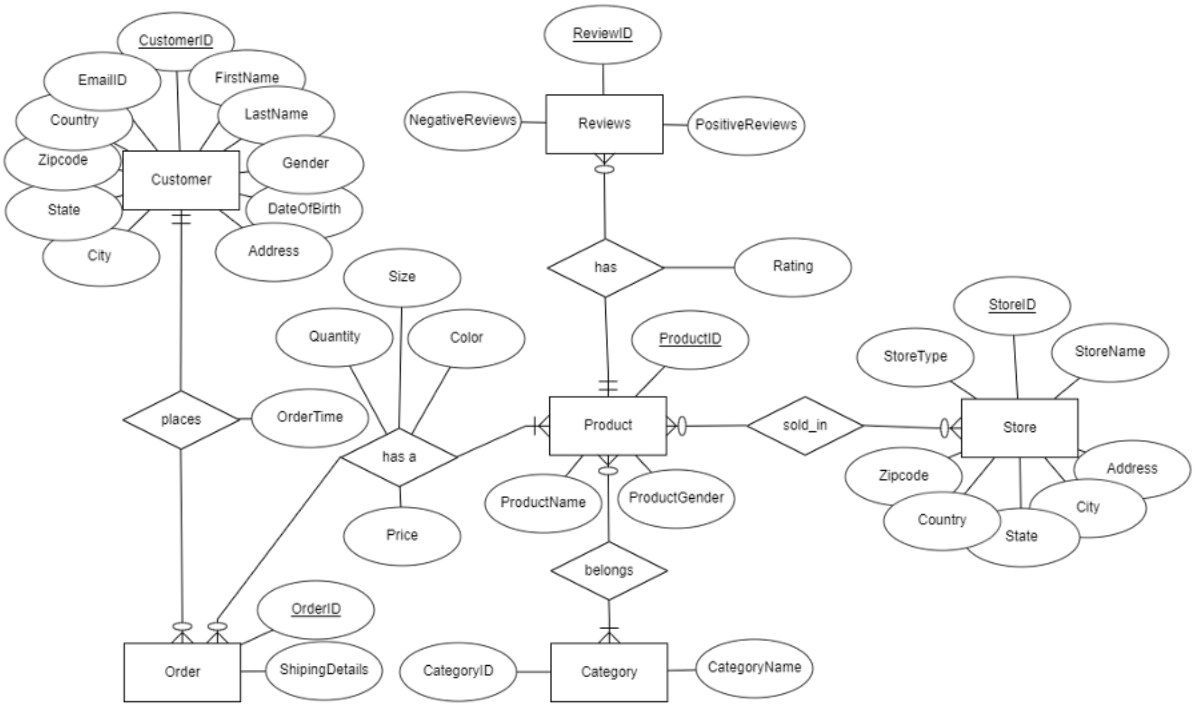

figure 1: Entity Relationship Diagram

The Entity Relationship Diagram of our analytical database is shown in figure 1. We have 6 entities, Customer, Order, Reviews, Product, Store and Category. It also showcases the attributes of corresponding entities and relationships. Apart from the ones shown in ERD, we have also taken into account the public data as tables for our queries like weather_info, census_data etc. The cardinality is explained as a customer can place one or more orders, a particular order must be associated with 1 and only one customer. An order can have one or more products whereas a product can be in zero or more orders. Similarly we define cardinalities for product belongs to category, product sold in stores and product has reviews relationships. A normalized database schema considered in this project work is as shown in figure 2.

Normalized Relational Schema

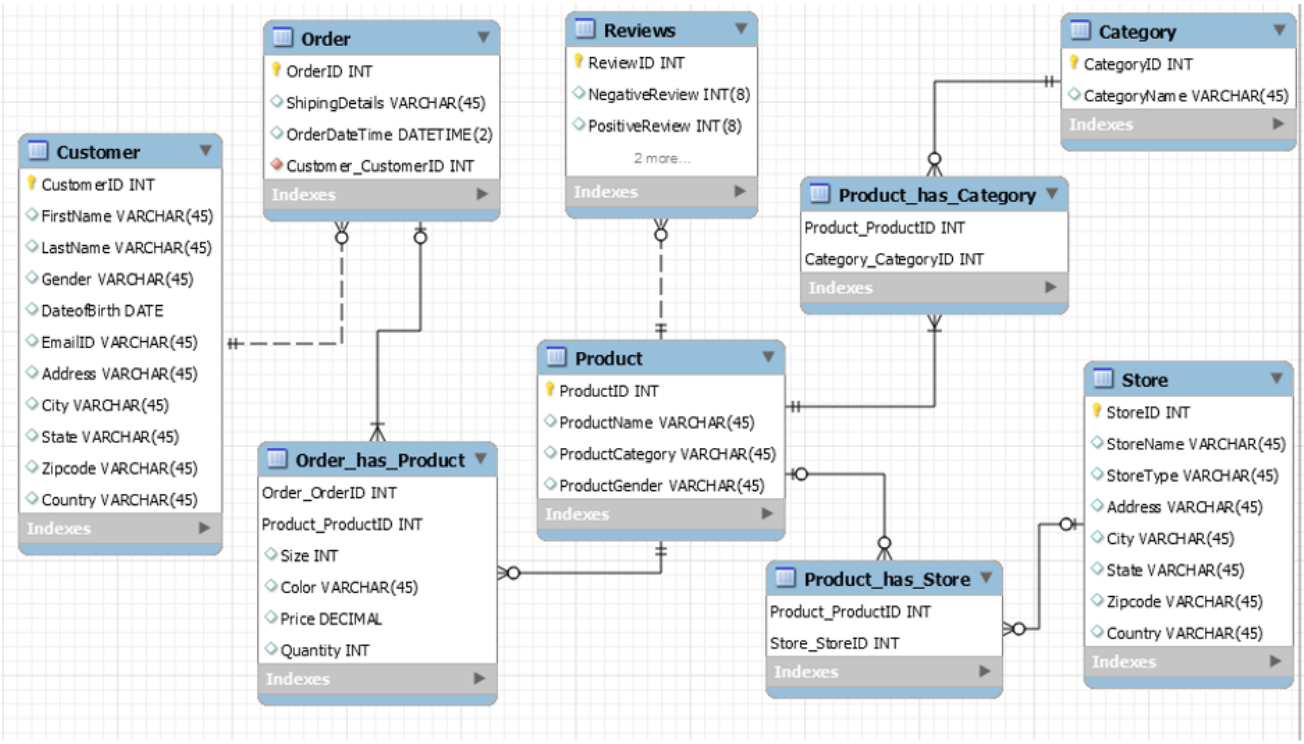

figure 2: Normalized Relational Schema

Question

How does Supply chain management affect sales and production? What are the recommended changes that can be made in supply chain management to obtain positive increase in sales/ profit?

The business value of answering the above question can be deduced by answering the following relevant questions and making some recommendations based on the analysis of the outcomes.

Implementation

a. Which is the least bought product and product category across different regions among store types?

Query:

select a.PName, a.Pcategory, a.sales, s.State from (select p.ProductName as Pname, p.ProductCategory as pcategory, sum(o.Quantity) as salesfrom Order_has_Product o Join Product p on o.Product_ProductID = p.ProductID)a Join Product_has_store ps on ps.Product_ProductID = p.ProductID Join Store s on Store_StoreID = StoreID group by s.State;

Scope: we could eliminate manufacturing of such products thereby reducing the costs incurred and also supply of those products to the respective region could be reduced and sometimes stopped.

b. List the product and product types with respect to demand, gender across regions and seasons.

Query:

select p.ProductName, p.ProductCategory, p.ProductGender, sum(op.Quantity) as sales, month(o.OrderDateTime) from Order_has__Product op join `Order` o on op.Order_OrderID = o.OrderID join Product p on p.ProductID = op.Order_OrderID group by op.Product_ProductID order by sales desc;

Scope: Enables clients to understand the demand of the products they produce along with the preference of their customers in varied conditions.

c. Which is the most happening product among product types and brands i.e., which is the product, product type and brand that makes the most sales across regions in the subscription table?

Query:

Select Product, Product Type, Brand, sum(Quantity), Region from subscription group by ProductType, Brand, Region order by Quantity desc;

Scope: It is important to understand the current and future trend in the market and change our strategies in production; hence we must know where we stand amongst the competitors. We can use the data available with our subscription stores for aggregation on weekly sales in comparison with brand and product categories.

d. Which store type has the most and least revenue generation?

Query:

select s.StoreType, sum(o.price*o.quantity) as revenue from Order_has_product O, Store S, Order R
where r.OrderID = o.Order_OrderID AND timestampdiff(day, curdate(), OrderDateTime) < 366
group by s.StoreType order by revenue desc limit 1 #can also do asc

Scope: Clients can make informed decisions about where they must increase or adapt newer marketing strategies to improve their turnover.

e. What is the recommended future increase in production with respect to size and gender of the product?

Query:

```
select gender, count(gender) as total from census_info
where age < 17  AND age > 12
group by gender
order by total desc #Shows total gender for the total young population

select distinct(city, state) as area, count(gender) as total from census_info
where age < 20 and gender = 'M' #Would rerun with 'F' for female
group by area
order by total desc
#Shows total males/females by city, can also put a state constraint in the where to show certain regions or other specific features (coastal states, higher population states)
```

Scope: Learning what to expect next and focusing production on those terms increases the sales and profitably, hence analysis on the publicly available census data in combination with the available client data will help predict the age groups and gender spread across different regions and accordingly the sizes and material type could be decided for production.

f. Which products have the highest and lowest accumulated rating, highest positive reviews, and highest negative reviews?

Query:

```
Select ProductID, sum(PositiveReview)/(sum(PositiveReview) + sum(NegativeReview)) as Rating, PositiveReview, NegativeReview from Reviews Order by Rating desc #or asc. Can also order by PositiveReview and NegativeReview;
```

Scope: Clients can easily understand their customer's feedback, what do they like, which products require improvement. After all, enhancing the products based on customer feedback not only evolves the features of the product but also eases the reach to the target audience.

Queries utilizing Census/Weather Information w.r.t to reviews

Finding Reviews for Expensive Products Amongst Higher Salary Zip Codes

Query:

```
Select p.ProductID, o.price r.CustomerID, r.product_rating, c.zip from Order_has_Product o, Reviews r, Customer c, Census_Info i, Product p Join Reviews On p.ProductID = r.ProductID
Where r.CustomerID = c.CustomerID AND c.zipcode = i.zipcode AND p.price > 150
Group by r.ProductID HAVING i.median_salary > 100000
```

Scope: With shoe companies, they will create shoes with different price ranges intended to attract a different demographic of customers. If Nike or Adidas wants to know how they can create a more appealing $250 shoe, as a business analyst we are going to receive the most insight from demographics with higher salaries as these customers will likely have a larger budget and stricter criteria on these types of shoes.

Finding Reviews for Clothing/Shoes Amongst Intended Gender

Query:

```
Select r.CustomerID, r.PositiveReview, r.NegativeReview, p.ProductID, o.gender from Reviews R, Product P, Order_has_product O, Customer C
JOIN Reviews ON p.ProductID = r.ProductID
Where r.CustomerID = c.CustomerID AND o.Gender = c.Gender
GROUP BY p.ProductID
Order By r.PositiveReview Asc #Can also order by NegativeReview
```

Additional Queries that could be used if we were able to collect product reviews as is (wordings)

Finding Common Themes Amongst Bad Reviews

This will require an initial aspect of concern for a particular product, for this query I will be examining how often comfort and value are brought up in negative reviews for a particular product

Query:

```
Select count(r.NegativeReview), p.ProductID from Reviews R, Product P
JOIN Reviews ON p.ProductID = r.ProductID
Where p.ProductID = _____  AND r.Review LIKE '%comfort%'
```

```
Select count(r.NegativeReview), p.ProductID from Reviews R, Product P
JOIN Reviews ON p.ProductID = r.ProductID Where p.ProductID = ____ AND (r.Review LIKE '%price%' OR r.Review LIKE '%expensive%' OR r.Review LIKE '%value%')
```

Finding Common Themes of Bad Reviews of Expensive Products Amongst Higher Salary Zip Codes

Finding count of reviews that mention style/look for all of the company's expensive products from higher salary zip codes

```
SELECT r.ProductID, o.price, count(r.NegativeReview) AS Total FROM Reviews R, Order_has_product O, Product P, Customer C, census_info I JOIN Reviews ON p.ProductID = r.ProductID
WHERE r.CustomerID = c.CustomerID AND c.zip = i.zip
AND i.median_salary > 100000 AND o.price > 150
AND (r.Review LIKE '%style%' OR r.Review LIKE '%look%' OR r.Review LIKE '%ugly%')
ORDER BY Total DESC
```

Summary/ Recommendations

After conducting the data requirement analysis, we arrived at the conclusion that certain attributes are more significant than the others in determining our Revenue Optimization. Factors such as Name of the customer are not significant contributors in achieving our objective but where a customer lives might affect his spending capacity and the category of shoe purchased which can effectively contribute to our objective.

Similarly, Product ID and sizes will not influence our revenue but ratings of particular products will play a significant role in influencing our business decisions. We can effectively conclude from our queries which products perform well and which products are not so popular. We can use this information and sort our demand by seasons and regions to stock particular warehouses and stores with more popular products for the season in order to fully satisfy customer demand.

Reviews can also provide insight to the company rather than simply impact overall purchases. For example through filtering data to a certain demographic we can pinpoint how a specific audience rates a certain product, which will allow the company to assess how their high price products appeal or not appeal to higher salary zip codes, how a hoodie or thermal clothing appeals to those that live in a colder climate, how kids clothing appeals to households with children, etc.. If the company has data set up to where a rating/user id can load up what was said in the review, we can also filter for keywords from the target audience to see what improvements a company can make in their product, such as comfort, durability, styles/colors, etc..

We can also use our queries to determine highest and lowest ratings of particular products and alter them from our product line if evidence is suggestive that it is not generating enough demand and also use the feedback for future product development. There is a lot of useful information that we can deduce from our dataset that can help us make analytical decisions.

References

1. https://www.statista.com/statistics/632210/nike-brand-value/

2.https://profitworks.ca/blog/marketing-strategy/545-nike-strategy-how-nike-became-successful-and-the-leader-in-the-sports-product-market.html