Team Project

Section and Team Number D16

Team Members:

1. Akash Deep Kumar
2. Che-Chih Lee
3. Kuan-Chih Lee
4. Shitij Gusain

Guidelines:

- Please submit your project report along with your presentation file (please see next point for details) following the Project submission link by the due date. Please name your file, **Project1_[Section][Team]** (e.g., Project1_B1.doc and Project1_B1.ppt). This will serve as evidence that your homework is done by the deadline.

Objectives: Through this project, I would like you to learn some skills on

- how to ask interesting and practical question(s)
- how to conduct data requirements analyses (what data is needed, where to get the data, and how to assemble the data together)
- how to address the question(s) and demonstrate the value of answering the questions

Grading: Grading will be based on:

- How interesting your question(s) is(are) and the potential value?
- Whether your design and approach support what you aim to do. Here is the caveat: the bigger or more "questions" you try to address, (likely) the higher the potential value, but (also likely) the more complex your design will be, and the more likely an error occurs.
- Effectiveness of your report: your report needs to be clear enough so people can understand and evaluate them.

Introduction

You have been consulted by a major footwear manufacturer, say, adidas or Nike, to provide advanced analytics to help advance their business. As with most businesses, the ultimate goal is to increase sales (profits) and decrease costs, and there are of course several ways to achieve the goal. After certain sessions with your client, you understand that as a manufacturer, they do not know a lot about their consumers, but they might be missing important opportunities for not knowing enough about their customers. Therefore, they are hoping, through data analytics, to learn more about their customer profiles, to understand where they stand with other competitors or potential competitors, to tap on

opportunities, to be able to predict future trends. As learned in class, successful business analytics start with questions that are relevant and could have impact. Consider the following the data that is currently available from the client company.

- Any transactions from your client's ecommerce website.
- Through partnership, you also have data from one specialty retail chain (e.g., Dicks' sporting goods).
- Through subscription, you also have access to aggregate weekly sales (in dollars and in units) at product subcategory - brand level (e.g. running shoes - Nike) at different types of retailers, including major department stores, mid-size department stores, shoe chains and specialty stores. You can also break the data into different geographical regions (northeast, west, central, south).

Of course, there is a lot of data out there that has not been collected by your client company but is available to grab, for example, social media data, consumer reviews data, as well as data from public sources (e.g. census bureau) etc.

Your client is seeking help from you to help them advance/protect their business and to help them design an appropriate data infrastructure that will help address important questions that have impact. Given this background, please:

1. Identifying the problem and relevant data:
   a. List one specific question (or a few related questions) you would like to focus on.
   b. Discuss how addressing this question may benefit and/or provide useful insights to your client.
   c. List the high-level data source (e.g. consumer's data, review data, etc.) you would need in order to answer the question you identify. Be realistic on what you could get and what you cannot have, and how you will get it.

Answer 1 a.

Since the goal is to increase profit and decrease cost, we should focus on these questions:

Questions:
1. What are the total sales per product?
2. How much revenue is generated per customer?
3. How does Weekly Sales Report look like?
4. What is the Return on Ad Spend?
5. How many impressions do we spend on per Advertisement?
6. What is the Lifetime Value of each customer?

Assumptions:
i. The shoe manufacturing company sells more than one product.
ii. This company uses social media platform for product promotion.
iii. We have customer level details.

Answer 1 b.

1. What are the total sales per product?

   - This will help us learn about our popular and unpopular products.

Picture reference: https://www.sbnation.com/2017/9/6/16255002/nba-signature-shoe-lebron-james-steph-curry-kyrie-irving-kevin-durant-paul-george-james-harden

- If we plot a trendline over time we can learn more about sales trends which will help the company strategize manufacturing, marketing, and inventory.

Picture reference: https://fizfy.com/list-of-marketing-strategies-to-increase-sales-in-2021/

- This will give the company an opportunity to relaunch a product based on the performance.

Picture reference: https://nicklashellb.org/diesel-yuk-the-pre-internet-shoe

- The company can also strategize to optimize the production cost of the popular products.

2. What are the total sales per customer?

- Total sales per customer would give us an idea of customer demands on an average.

- This would help the company estimate of production quantity (manage supply).

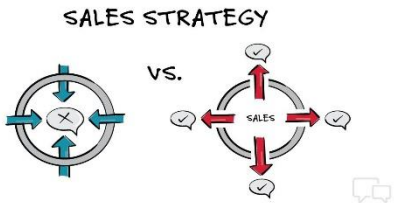
Picture Reference: https://smartdelivery.trimble.com/new-customer-demands-for-advanced-returns/

- Once of the most important and insightful analysis would be to understand target audience. Based on the various demographics (mentioned below) this company can target their audience by segmenting its customers.
  i. Age
  ii. Gender
  iii. Location

Picture Reference: https://www.adjust.com/blog/audience-segmentation-with-audience-builder/

iv. Income
v. Marital Status
vi. Number of children
Note: While segmentation we may want to segment the data into Household Level as it is not necessary that the consumer will be same as the customer.

3. How does Weekly Sales Report look like?
   - Would get the highest and lowest selling week.
   - The weekly report will also help the shoe company to re-strategize sales based on real-time response. It will allow them to give store/region wise discounts if needed.



Picture Reference: https://corporatevisions.com/sales-strategy/

   - Weekly Sales would help the shoe company to perform employee management, like hiring new employees or plan a hiring freeze or layoff.

4. What is the Return on Ad Spend?

Since we have target audience (using question 3), the company can use marketing to increase visibility of their products to eventually increase sales and profit.
Return on Ad Spend or ROAS would help the shoe company to understand the revenue generated per every dollar that was spent on advertising.

Picture Reference: https://streamingwords.com/why-roas-matters-in-digital-marketing/

As we have the transaction data, we will link that to the customer data to get revenue generated per customer. Each customer would have their social media account and that can help us determine if the impression matched that customer or not. Joining both these information we can get ROAS and can extrapolate the numbers to learn if the campaign was effective or not.

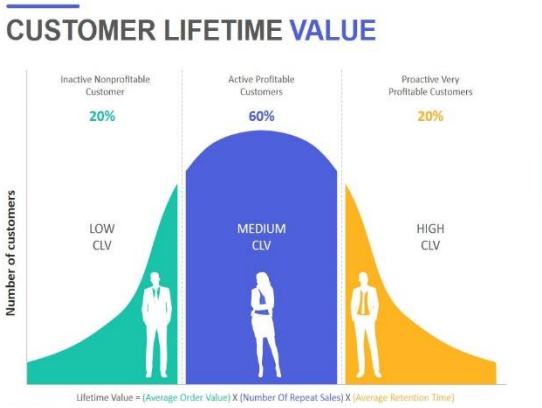
5. How many impressions do we spend on per Advertisement?

- Impressions are basically clicks/views and to understand how much money the shoe company spent of each click we count this for them. Impressions can be shared only with target audience and not the test.
- This can further help the shoe company to analyze how the different segments (frequent buyers, occasional buyers and new buyers) of customer perform.

- The shoe company can compare this with the performance of the test audience and plan the next campaigns and the frequency of campaigns.

6. What is the Lifetime Value of each customer?

- The shoe company can do more niche segmenting using the CLV technique. This data will help them learn about the behavior and patterns of spend based on how much was spent by a customer during their lifetime.
- The shoe company can learn more about their loyal customers.
- If we look at the data, more carefully this can help the shoe company identify the gap between the purchases of expensive items which will help the design teams to plan new product launch.



Picture reference: https://powerslides.com/product-tag/customer-lifetime-value/

2. Please provide the database schema(s) (based on either relational schema or star schema) that will help you address the questions you identify above. This should cover:
   a. Specific data in terms of attributes/columns you need to track and collect and how they help address the question(s) you identify?

      Answer: Please consider the table and column descriptions below:

      Table 1. Customer

      The Customer Table provides a list of the customer and their details. As we know that the shoe company is not aware of its customer profile, it is important for them to have this table in order to know their customer base.

      Attributes:
      a. Customer ID: Unique ID given to each customer
      b. Customer_FName: Customer First Name
      c. Customer_LName: Customer Last Name
      d. Age: Age of Customer
      e. Gender: Gender of Customer
      f. Married: Marital Status
      g. Children: # of Children the Customer has
      h. Income: Annual Income of the Customer
      i. Customer_Address: Customer Address
      j. Customer_City: City of Residence
      h. Customer_Country: Country of Residence

      Most of the above attributes help us to identify customer based on demographics and will eventually help the shoe company to segment and target it's audience.

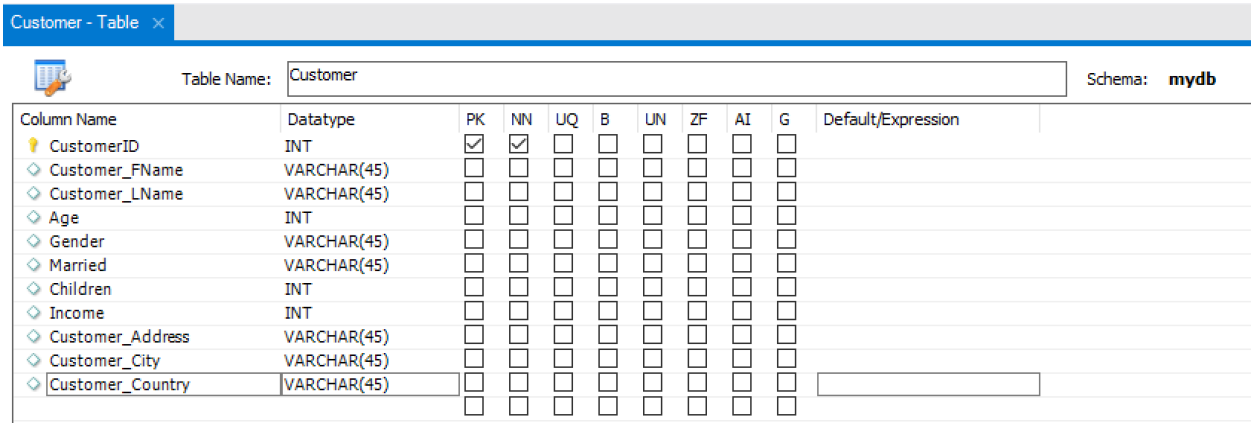


Table 2: Orders

This table keeps tracks of all the orders made so that Shoe company can track the trends of orders and the count of purchases made over a due course of time.

Attributes
a. Order_ID – Order ID (Primary key and unique to each order).
b. Date – Date when the order was made to help us track the number of orders overtime.
c. Customer_ID – Foreign Key from Customer table to link to each Customer
d. Store_ID – Foreign Key to link the store from where this order was made
e. E-Commerce_PlatformID - Foreign Key to link the store from where this order was placed

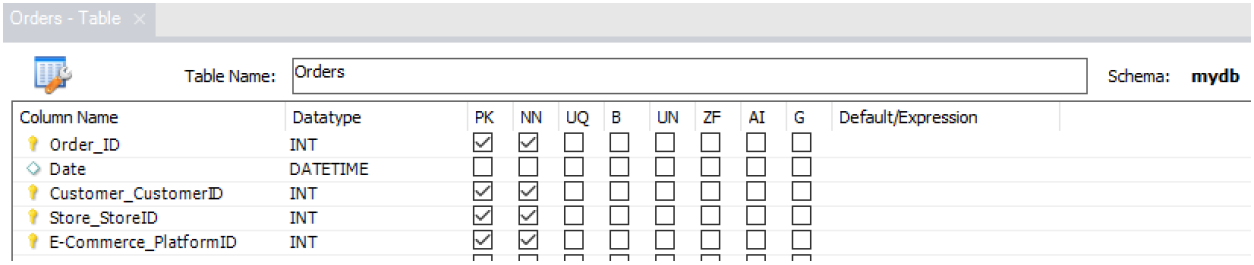


Table 3: Product

It is important for the shoe company to have a list of the products that they are selling so that they can get granular in tracking the product consumption rate and make smart business decisions of manufacturing and warehousing.

Attributes
a. ProductID: ID of the Product
b. Product_Name: Name of the Product
c. Category: Category in which the product falls into
d. Product_Price: Price of the Product
e. Country_of_Origin: Country in which the Product was made

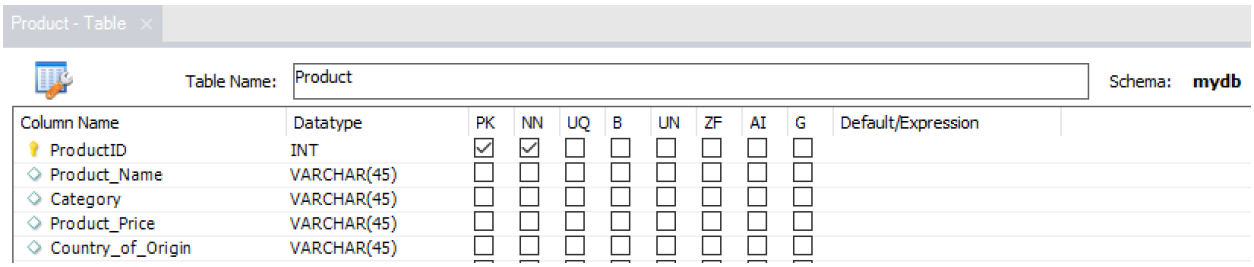


Table 4: Supplier

We assume that not all the products are made at one location and the company would have a supplier responsible for assembling and spare parts. This table give information of that which would help the shoe company make logistics decisions

a. SupplierID: ID of the Supplier
b. Supplier_Name: Name of the Supplier

c. Supplier_Address: Address of the Supplier
d. Supplier_City: City where the Supplier is located
e. Supplier_Country: Country where the Supplier is located
f. Supplier_Phone: Contact Number of the Supplier

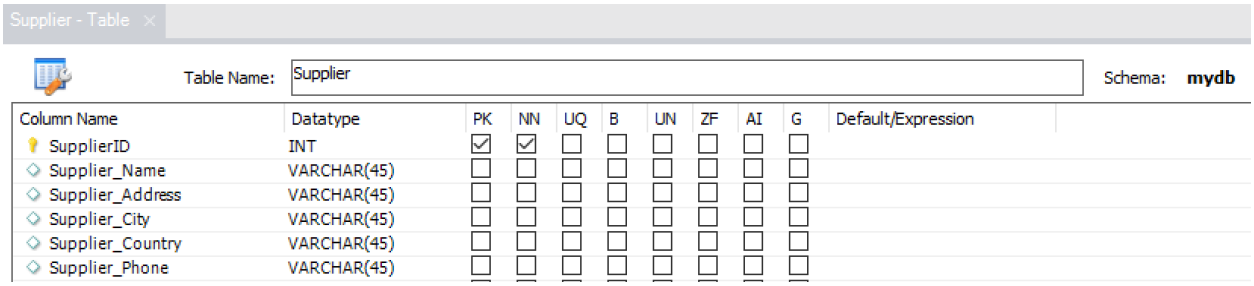


Table 5: Store

As mentioned above, the shoe company has stores with which it has partnered and may incur cost for the same, this table will help us locate the store that sells products of the shoe company and that will help us understand the extra operational cost for which the shoe company is liable. This table will give us important information to make decisions on cost-cutting or expanding our business.

Attributes
a. StoreID – ID of the Store/Outlet (eg. Dicks' sporting goods)
b. Store_Name – Store Name
c. EmployeeNum – Employees working in the Store
d. RunningCost – Cost of running the store
e. Store_Address – Address of the Store
f. Store_City – City where the store is located
g. Store_Country – Country where the store is located

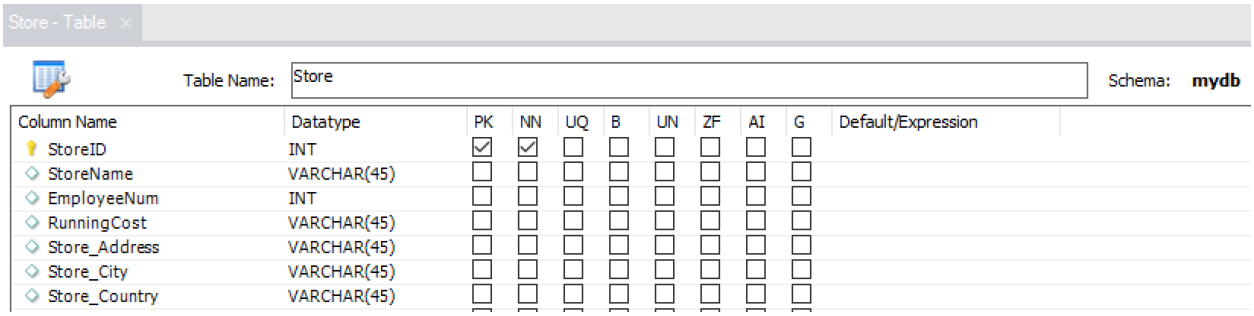


Table 6: E-Commerce

The company also sells its products online and this table gives them a list of portals or E-Commerce websites where it sells the products.

Attributes
a. E-CommerceID: ID of the Website
b. E-Commerce_Name: Name of the Website

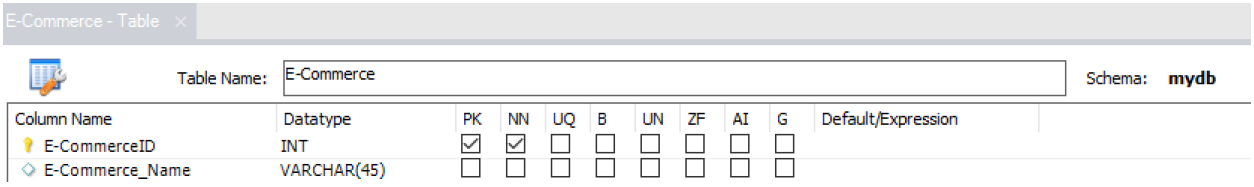


Table 7: Retailers

Some other retailers also sell the products of a shoe company, and this will help them keep a track of those, so that the shoe company knows the sources where its products are placed.

Attributes
a. RetailerID – Unique ID of the Retailer
b. Retailer_Name – Name of the Retailer
c. Retailer_Category – Category of the retailer (Based on the products the retailer sells)

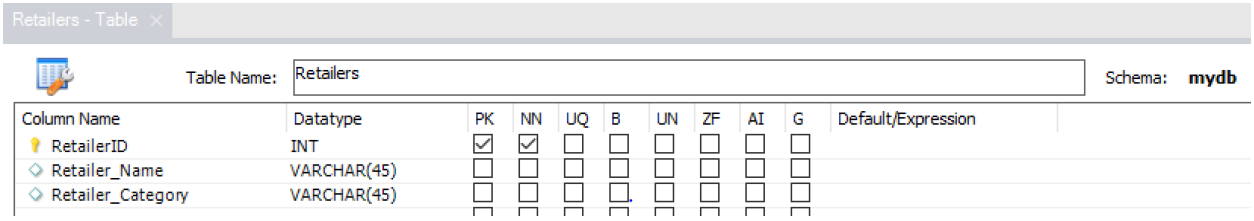


Table 8: Platform

Marketing is very important key for any business to flourish and we assume that in today's generation it's important for shoe company to use an online social media platform to market their products so this table saves all the platforms which can help the shoe company to advertise their product.

Attributes
a. PlatformID – Unique ID of the Social Media Platform
b. Retailer_Name – Name of the Platfrom

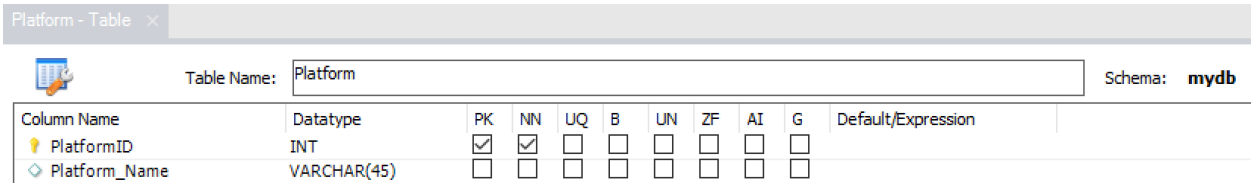


Table 8: Advertisement

Advertisement is a major part of Marketing, and this table stores all the advertising data for the shoe company.

Attributes
a. AdvertisementID – Unique ID of the Social Media Platform
b. Campaign_StartDate – Start Date of the AD Campaign
c. Campiagn_EndDate – End date of the AD Campaign
d. Campaign_Cost – Amount spent to run the entire Campaign
e. Impression_Count – The total number of impression that the campaign had.
f. Advertisement_Name – Name of the AD Campaign

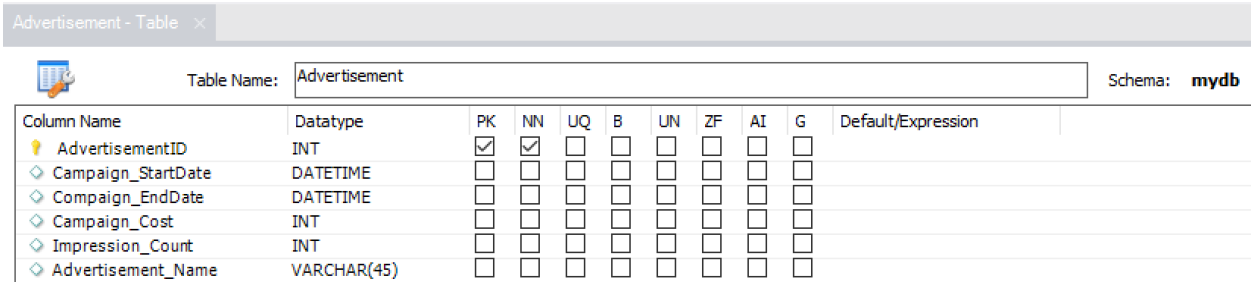
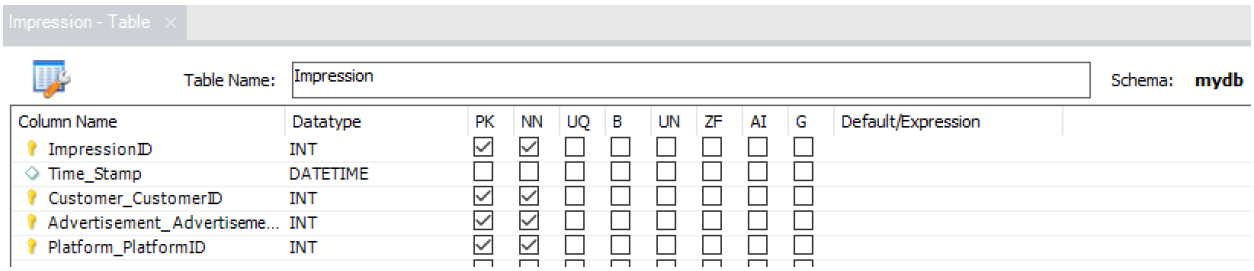


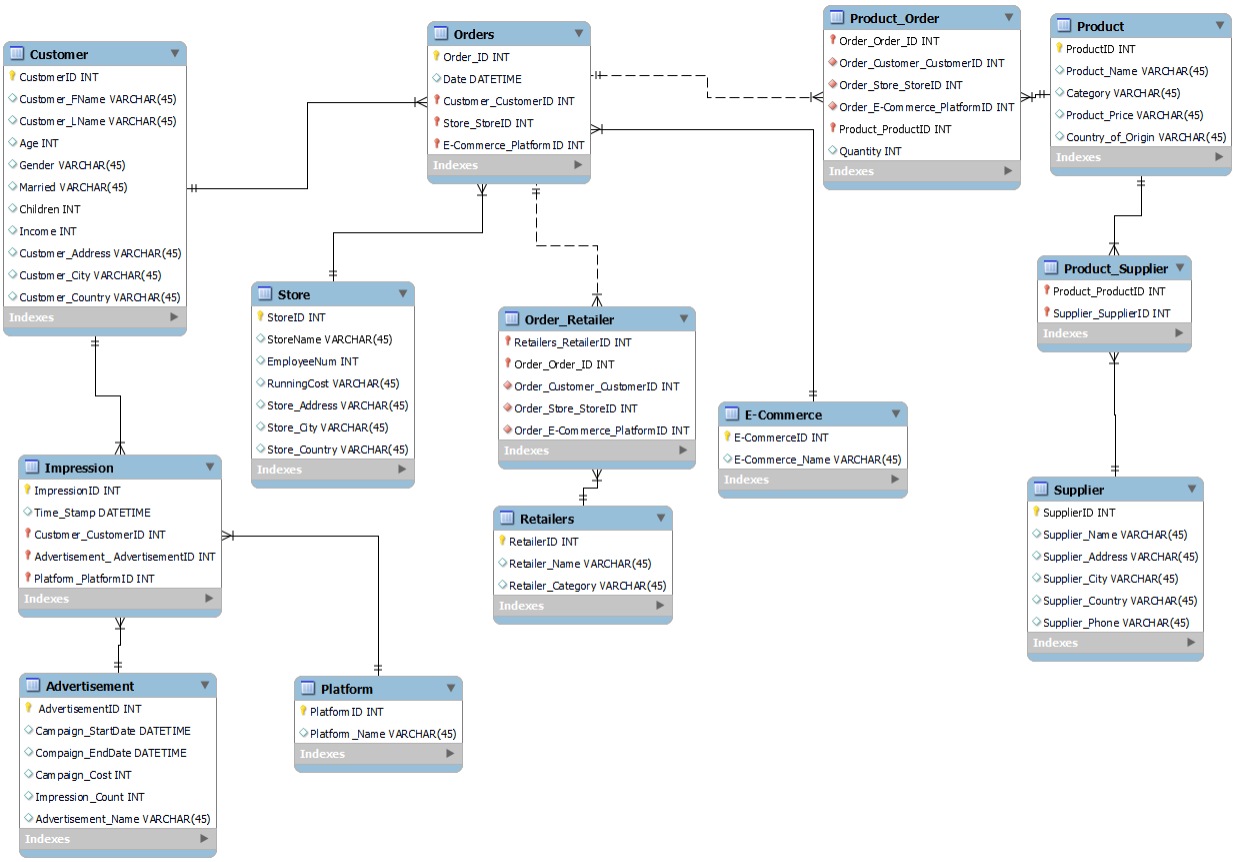
Table 9: Impression

The number of impressions is the number of views that the AD Campaign has. This table gives a detailed picture of each impression

Attributes
a. ImpressionID – Unique ID of every Impression
b. Time_Stamp – Time when the AD Campaign was viewed
c. Customer_CustomerID – Customer ID to link who viewed the AD.
d. Advertisement_AdevrtisementID – Advertisement ID associated to this campaign.
e. Platform_PlatformID – Gives us the information of which platform was used to display this AD.


Impression - Table
Table Name:
Impression
Schema:
mydb
Column Name
Datatype
PK
NN
UQ
B
UN
ZF
AI
G
Default/Expression
ImpressionID
INT
Time_Stamp
DATETIME
Customer_CustomerID
INT
Advertisement_Advertiseme...
INT
Platform_PlatformID
INT

b. Provide database schema(s). Note in the case that you have many attributes and could not be shown in a page, please just show the most important attributes, and the rest could be provided as additional notes accompanying your schema(s).



3. Based on your schema(s), provide sample queries, and discuss how you will use the query results to address your question(s).

Answer: We have scripted some queries in order to answer the questions that would help the shoe company increase profit and decrease cost.

Sum of order per customer per product:

```
select
c.Customer_Fname,
c.Customer_Lname,
c.CustomerID,
p.ProductID,
p.Product_Name,
sum(po.Quantity)*sum(p.Product_Price) as Revenue
from
Orders o
inner join customer c on o.CustomerID = c.CustomerID
inner join product_order po on o.OrderID = po.OrderID
inner join product p on po.OrderID = p. OrderID
group by
c.CustomerID,
p.ProductID;
```

Sum of order per product: (This helps us answer the total sales per product)

```
Select
EachRevenue.ProductID,
EachRevenue.Product_Name,
Sum(EachRevenue. Revenue) as RevenuePerProduct
From(
select
c.Customer_Fname,
```

```
c.Customer_Lname,
c.CustomerID,
p.ProductID,
p.Product_Name,
sum(po.Quantity)*sum(p.Product_Price) as Revenue
from
Orders o
inner join customer c on o.CustomerID = c.CustomerID
inner join product_order po on o.OrderID = po.OrderID
inner join product p on po.OrderID = p. OrderID
group by
c.CustomerID,
p.ProductID) EachRevenue
group by EachRevenue. ProductID;
```

Sum of order per customer: (This helps us answer revenue generated per customer)

```
Select
EachRevenue.Customer_Fname,
EachRevenue.Customer_Lname,
EachRevenue.CustomerID,
Sum(EachRevenue. Revenue) as RevenuePerCustomer
From(
select
c.Customer_Fname,
c.Customer_Lname,
c.CustomerID,
p.ProductID,
p.Product_Name,
sum(po.Quantity)*sum(p.Product_Price) as Revenue
```

```
from
Orders o
inner join customer c on o.CustomerID = c.CustomerID
inner join product_order po on o.OrderID = po.OrderID
inner join product p on po.OrderID = p. OrderID
group by
c.CustomerID,
p.ProductID) EachRevenue
group by EachRevenue.CustomerID;
```

Weekly sales per week per product:

```
set datefirst = 7;
select datepart(week, o,Date) as week,
p.ProductID,
p.Product_Name,
sum(po.Quantity)*sum(p.Product_Price) as Revenue
from
Orders o
inner join product_order po on o.OrderID = po.OrderID
inner join product p on po.OrderID = p. OrderID
group by
week,
p.ProductID;
```

Weekly sales per week: (We can determine how does the Weekly Sales Report looks like)

```
set DATEFIRST = 7;
select
WeekProductSale.week,
Sum(WeekProductSale. Revenue) as WeekRevenue
```

```
from
(select datepart(week, o,Date) as week,
p.ProductID,
p.Product_Name,
sum(po.Quantity)*sum(p.Product_Price) as Revenue
from
Orders o
inner join product_order po on o.OrderID = po.OrderID
inner join product p on po.OrderID = p. OrderID
group by
week,
p.ProductID) WeekProductSale
group by WeekProductSale.week
order by WeekRevenue;
```

Advertisement:

Return on advertising spend (ROAS): (This covers how much is the return on Ad Spend)

```
Select
Cost. AdevertisementID,
Revenue.Revenue as Revenue,
Cost. Campaign_Cost as Cost,
Revenue.Revenue/ Cost. Campaign_Cost as ROAS
from
(
Select
A.Campaign_Cost,
A.AdevertisementID
From
Advertisement A ) Cost
Inner join (
```

```
Select
ad.AdvertisementID,
sum(po.Quantity)*sum(p.Product_Price) as Revenue
from
Orders o
inner join customer c on o.CustomerID = c.CustomerID
inner join product_order po on o.OrderID = po.OrderID
inner join product p on po.OrderID = p. OrderID
inner join Impression i on i.Customer = c.Customer
inner join advertisement ad on ad.AdvertisementID = I.AdvertisementID
where o.Date between ad.Campaign_StartDate and ad.Campaign_EndDate
and I.Time_Stamp between ad.Campaign_StartDate and ad.Campaign_EndDate
group by
ad.AdvertisementID) Revenue on Cost.AdevertisementID = Revenue.AdevertisementID;
```

Impression Spend: (We can get a count of impressions per Advertisement)

```
Select
Ad.AdvertisementID,
Ad.Advertisement_Name,
P.Platform_Name,
AD. Campaign_Cost/Ad.Impression_Count as ImpressionSpend
From
Impression I
Inner join Platform P on P.PlatformID = I.PlatformID
Inner join Advertisement Ad on Ad.AdvertisementID = I.AdvertisementID;
```

Life time Value(LTV):

LTV per customer: (As the name suggests we get the LifeTime Value per customer using this query)

```
Select
Final.Customer_Fname,
Final.Customer_Lname,
Final.CustomerID,
Final.CRevPerMon * Final.LifeCycle as LTV
From(
Select
CTotalRevenue.Customer_Fname,
CTotalRevenue.Customer_Lname,
CTotalRevenue.CustomerID,
CTotalRevenue. Revenue/ CLifeMonth. LifeMonth as CRevPerMon,
TLifeCycle.LifeCycle
From(
select
c.Customer_Fname,
c.Customer_Lname,
c.CustomerID,
sum(po.Quantity)*sum(p.Product_Price) as Revenue
from
Orders o
inner join customer c on o.CustomerID = c.CustomerID
inner join product_order po on o.OrderID = po.OrderID
inner join product p on po.OrderID = p. OrderID
group by
c.CustomerID) CTotalRevenue
```

```
inner join
(select
c.Customer_Fname,
c.Customer_Lname,
c.CustomerID,
Ceil(timestampdiff (day,date(min(o.Date)),now())/30) as LifeMonth
from
Orders o
inner join customer c on o.CustomerID = c.CustomerID
group by c.CustomerID
) CLifeMonth on CLifeMonth.CustomerID = CTotalRevenue.CustomerID
left join (
select
"TEST" as TEST,
Sum(LifeCycle.LifeCycPerC) as LifeCycle
from(
select
c.Customer_Fname,
c.Customer_Lname,
c.CustomerID,
Ceil(timestampdiff (day,date(min(o.Date))), (day,date(max(o.Date)))/30) as LifeCycPerC
from
Orders o
inner join customer c on o.CustomerID = c.CustomerID
group by c.CustomerID) LifeCycle
) TLifeCycle on CTotalRevenue.CustomerID <> TLifeCycle.TEST) Final
```