Proarbeit

Olatunji Ojuko

February 1, 2021

1 Dataset 1 - Campaign Performance Analysis

Marketing campaigns have been launched on Facebook Ads and Google Ads to help a business attract potential customers. Here we will analyse the performance of these campaigns within different age groups of the target audience between end of the previous year and start of the new year.

Goal - Plan next marketing budget around response rates of different age groups. Eliminate or keep current advertisement campaigns depending on group performance.

1.1 Features (16834 records & 16 Columns)

Date - Date of event
product - Product been advertised
phase - What this dataset is meant to evaluate
campaign_platform - Google Ads or Facebook Ads
campaign_type - Search for Google Ads and Conversion for Facebook Ads
communication_medium - Keywords for Google Ads and CTA items for Facebook Ads
subchannel - Relevant for Google Ads only. Own Brand, Competitor or Generic related keyword
audience_type - Relevant for Facebook Ads. Classes of audience
creative_type - Relevant for Facebook Ads. Type of Image/Video/Carousel used
creative_name - Classes of *creative_type*
age - Age group of user
spends - Monetary cost of advertisement in that session
impressions - Count of views of impression item
clicks - Count of Clicks on impression item
link_clicks - Count of Clicks on call-to-action links within impression item

1.2 Dataset Justification

The dataset contains campaign from two of the most relevant digital platforms, Facebook Ads and Google Ads. Insights into campaign success within each and between both platforms can be derived, contrasted and compared. Below is an overview of the relevance of the features to each campaign. .

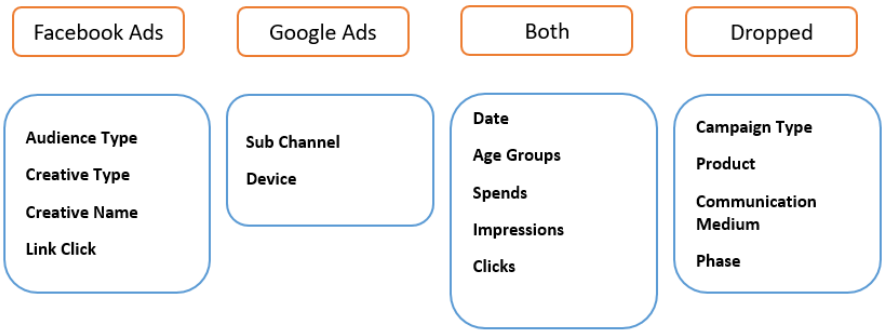


Figure 1: Feature Overview

1.3 Transfer and ETL

1.3.1 Transfer into SQL Server with 'Import'

The csv file was loaded into SQL Server with the Import module and the table structure is shown in the provided image *dataset1_structure_sqlserver.PNG*.

1.3.2 Transformation

Here the data will be processed. A primary is be added, missing values dealt with, columns renamed and so on. The well annotated query script of this transformation (*Dataset_1 Transformation*) is supplied. Categorical missing values are replaced with 'Undetermined' while those of numeric values are replaced with '0'. Specifically *AudienceType*, *CreativeType* and *CreativeName* were affected.

1.4 Preparing for Visualization

A SQL Server connection to Tableau Prep was established and the *Dataset_1* table taken. Supplied image *sql_server_tableau_prep_connection.PNG* shows the connection.

1.4.1 Dropped Columns

The following columns are dropped as they either have single values throughout the dataset or two values, one for each campaign, and will therefore not help in establishing patterns: *CampaignType*, *Product*, *CommunicationMedium* and *Phase*.

1.4.2 Data Type Changes

Data Type change - *Spends*, *Impressions*, *Clicks* and *LinkClicks* were changed from string to Number data types.

1.4.3 Feature Engineering

Facebook Ads
CTA-CTR - Can be taken as a sort of conversion rate, since user makes en-

quiry into offer: (Link Clicks/Impressions)*100
CTA - CPC - Spends/Link Clicks
Return on Ad Spend (ROAS) - Link Clicks/Spends. This is taken as so since the highest scope of performance in this dataset is that users click on the offer as there is no *Revenue* data.

Both Ad Platforms
Click Through Rate (CTR) - (Clicks/Impressions)*100
Cost Per Click (CPC) - Spends/Clicks
Cost Per Mile (CPM) - Cost of every thousand views: (Spends/Impressions)*1000

P.S - *Null Values* caused by division-by-zero in the calculated fields were replaced by 0.

1.5 Analysis

1.5.1 Facebook Ads

Determine underperforming group(s)

We check the average response rate of the different groups using the average *CTA-CTR*. We use average to prevent population sizes from cause skews data. As can be seen in figure 2 the 55 - 64 group has seen the greatest decline. That group will be further analysed to find actionable information. .

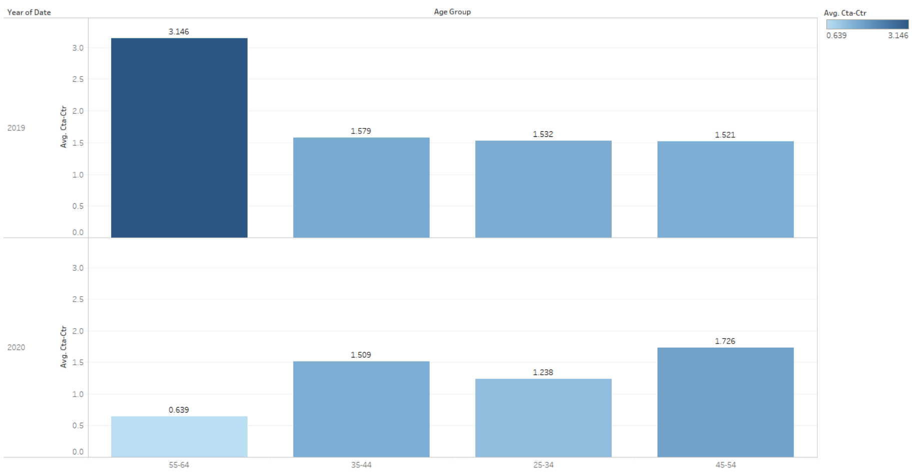


Figure 2: Facebook Ads Response

Age Group 55 - 64: Findings

Facebook Ads categorical variables are *AudienceType* and *CreativeName*.
Audience Type - From figure 3 below we can see that *Audience 2* group are responding much better than *Audience 1*. Next we will look at other variables before drawing conclusions. .

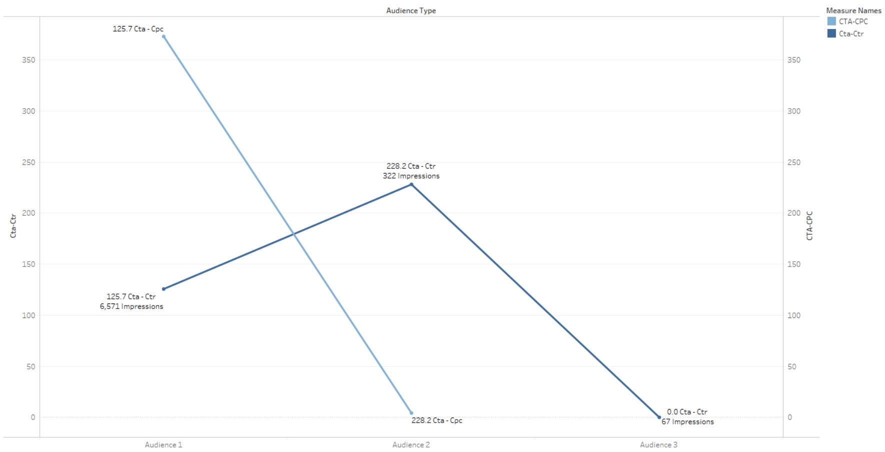


Figure 3: Audience Types Analysis

Creative Names - A quick glance at figure 4 shows a much better response to *Carousal* than to the other options. Also, losses have been incured on *Clicks*. We will cross both *AudienceType* and *CreativeName* to get better details. .

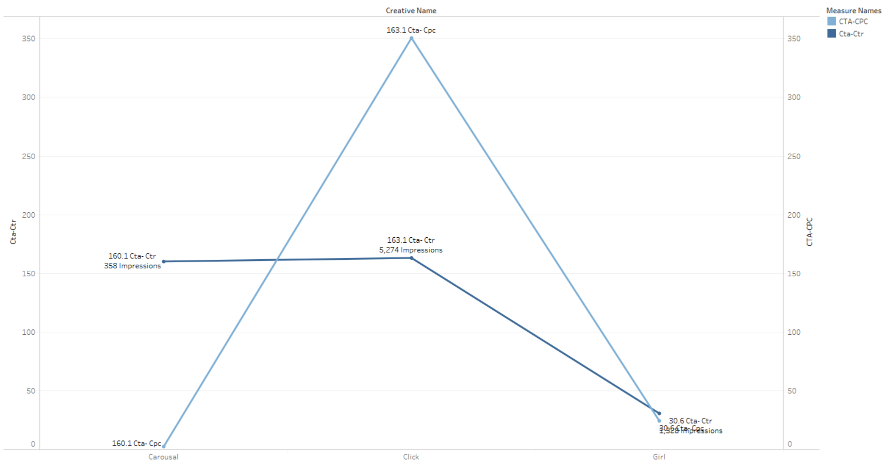


Figure 4: Creative Names Analysis

Figure 5 shows that huge losses have been taken on *Click* campaigns on *Audience 1*. . Furthermore, figure 6 shows very good responses by *Audience 2* to both *Click* and *Carousal* campaigns.

. *ROAS* has high correlation with *CTA-CTR* (0.7910) so further analysis of campaign cost with *CTA-CPC* yields the same results.

Facebook Ads Conclusions

- Suspend *Click* campaigns targeted at *Audience 1*.
- Divert the budget to pushing more *Carousal* and *Click* campaigns to *Audience 2*.

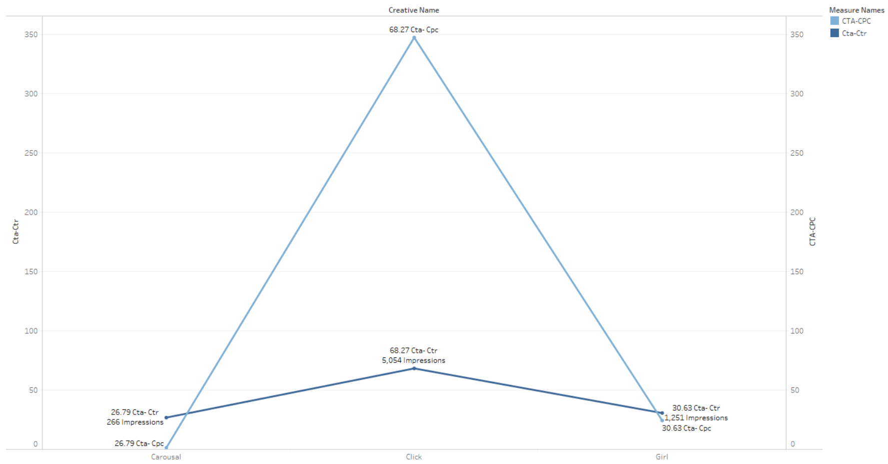


Figure 5: Audience 1: Losses on Click Campaigns

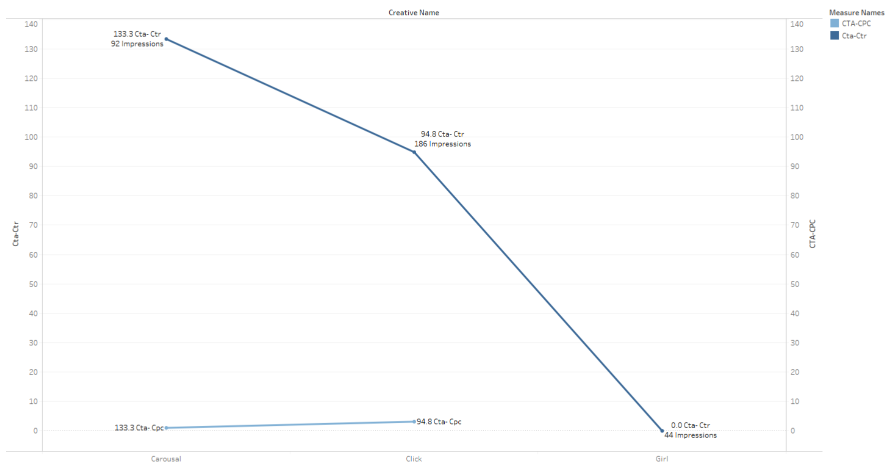


Figure 6: Audience 2 Positive Responses

1.5.2 Google Ads

The *CTR* is used to assess response here as there are no *CTA-CTR* measures. Figure 7 gives a great overview of the Google Ads performance.
Observation 1 - Far more users have come from other brand options with comparable price and quality.
Observation 2 - In 2019, cost (*CPC* in Line) outstrips response (*CTR in bars*) in all age groups and in all channels.
Observation 3 - In 2020, response among all ages cost when coming from other brand options. Age groups 55 - 64 and 65 and above also outstrip cost in their responses showing they are open minded to alternatives of different prices and qualities.
Observation 4 - Costs of users coming from generic products of different but related functionalities remained too high among among all groups in both years.

.

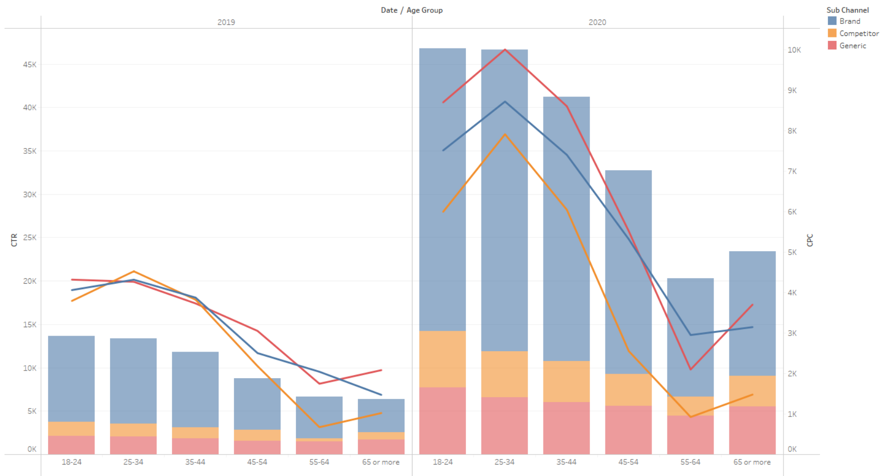


Figure 7: Google Ads Performance

Google Ads Conclusions

- Observation 1 shows majority of people will only checkout product of similar prices and quality.
- Observation 4 shows a contrast of above to ads placed on unrelated products. Hence emphasis should be placed on advertising on related products pages.
- There was significant improvements in responses from 2019 to 2020. Except for generic channels.
- users that are 55 and above are more open minded to products of differing prices and qualities.

1.5.3 Dashboard

. **Total Spend** - The total spend on both campaigns so far.
Platform Response Split - A percentage split of spends on the two advertisement platforms. Showing almost 8 times as much is spent on Google Ads.
Platform Response Split - A split of all responses in the forms of clicks on the adverts on both platforms. This shows Facebook Ads returns a larger share

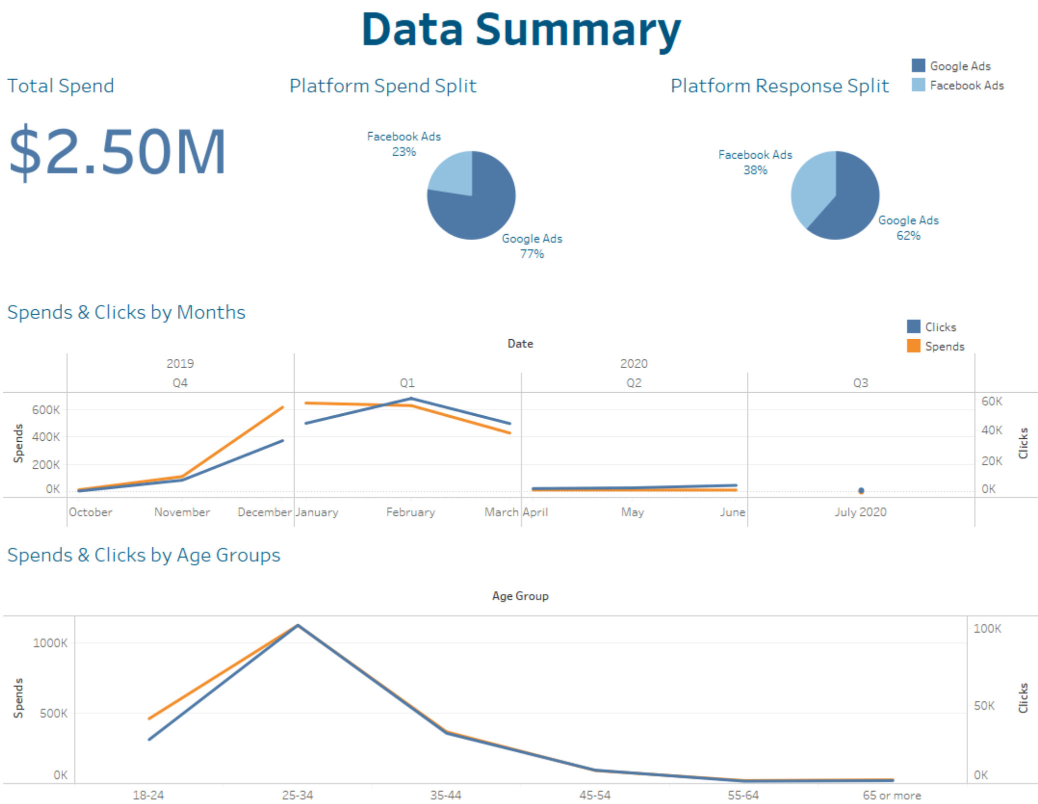


Figure 8: Dashboard

of the responses compared to how much is spent on it. The opposite applies to Google Ads.
Spends and Clicks by Months - This chart shows a highly correlated relationship between spends and responses. **Performance was worst between November and December 2019**, but the new year saw a gradual improvement in performance. Eventually in February 2020, spending dipped below response. **February to March of 2020 saw the saw best performance**.
Spends and Clicks by Age Groups - Only the 18 - 24 age group have a relative lower response than the other groups even as the correlation of spends and responses is high among the age groups. It is important to note that while the 18 - 24 age group data is exclusive to Google Ads, the exact same pattern seen here in relation to the other age groups is maintained on that platform too.

Conclusions and Recommendations - Overall, to engage more users more money will be spent. However, efficiency of cost - response ratio can be improved with platform specific analysis like those seen in previous sections of this report. Also, the higher ratio of response on Facebook Ads compared to spend suggests there should be more focus on it and it should get a bigger portion of the spending budget than the current 23%.

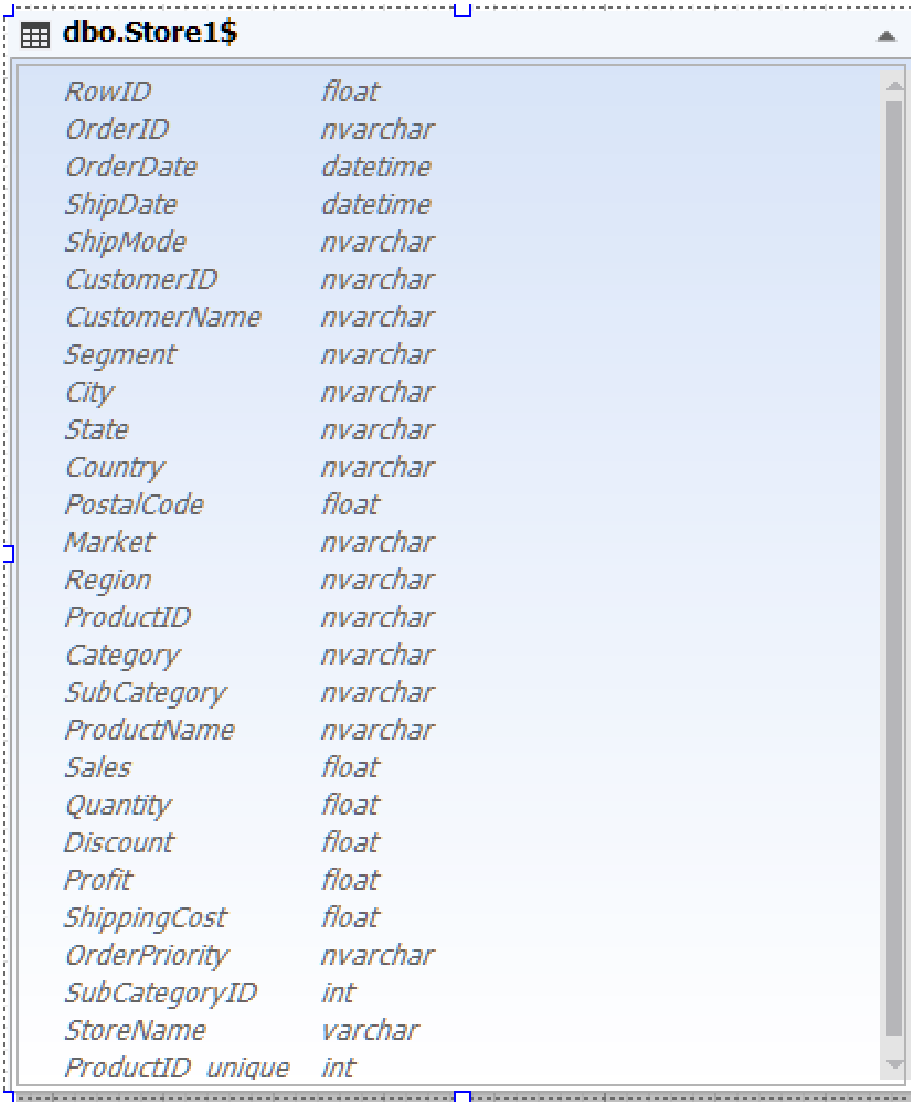
2 Dataset 2 - Orders ETL

The first dataset, while suitable for Digital Analysis tasks, doesn't present ETL task opportunities. Hence, I present a brief ETL process of two datasets, each representing online retail stores. Below are features of both datasets, *Copany1* and *Company2*: .

. Goal is to separate relevant entities in the datasets into single relational tables and then put them into analysis relevant tables.

2.1 Transform

Step 1 Flatten *Company 2*'s datasets into a single flat file with an inner join on the *OrderID* columns.
Step 2 Create and select relevant data from both compannies' datasets into *CustomerTable*, *ProductTable*, *CategoryTable*,*SubCategoryTable* and *Ordertable*. Primary keys to all and *StoreName* (Company 1/Company 2) columns added where necessary.
Step 3 *ProductID* in *Company 1* had duplicates and *Company 2* had no such column so a new primary key (*ProductID_unique*) was generated for the merged *ProductsTable*.
Step 4 Categories and subcategories in both tables had identical values so no need for the separating column *StoreName*.
Step 5 Relationships were created and our relational database is setup as shown in figure 11. Data from both company data now merged into relevant tables. .
P.S - I didn't create a separate set of tables for location features because of time pressure due to work on my abschlussarbeit. I thought the goal and thinking is already well established.



dbo.Store1$

RowID	float
OrderID	nvarchar
OrderDate	datetime
ShipDate	datetime
ShipMode	nvarchar
CustomerID	nvarchar
CustomerName	nvarchar
Segment	nvarchar
City	nvarchar
State	nvarchar
Country	nvarchar
PostalCode	float
Market	nvarchar
Region	nvarchar
ProductID	nvarchar
Category	nvarchar
SubCategory	nvarchar
ProductName	nvarchar
Sales	float
Quantity	float
Discount	float
Profit	float
ShippingCost	float
OrderPriority	nvarchar
SubCategoryID	int
StoreName	varchar
ProductID_unique	int

Figure 9: Company 1 Features

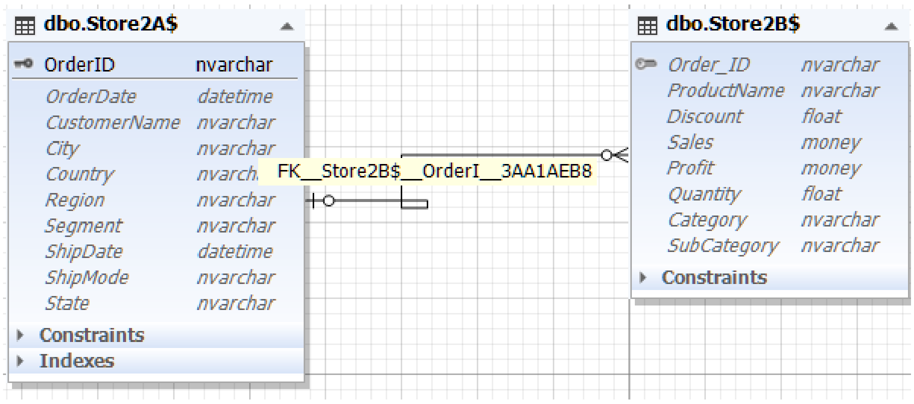


Figure 10: Company 2 Features

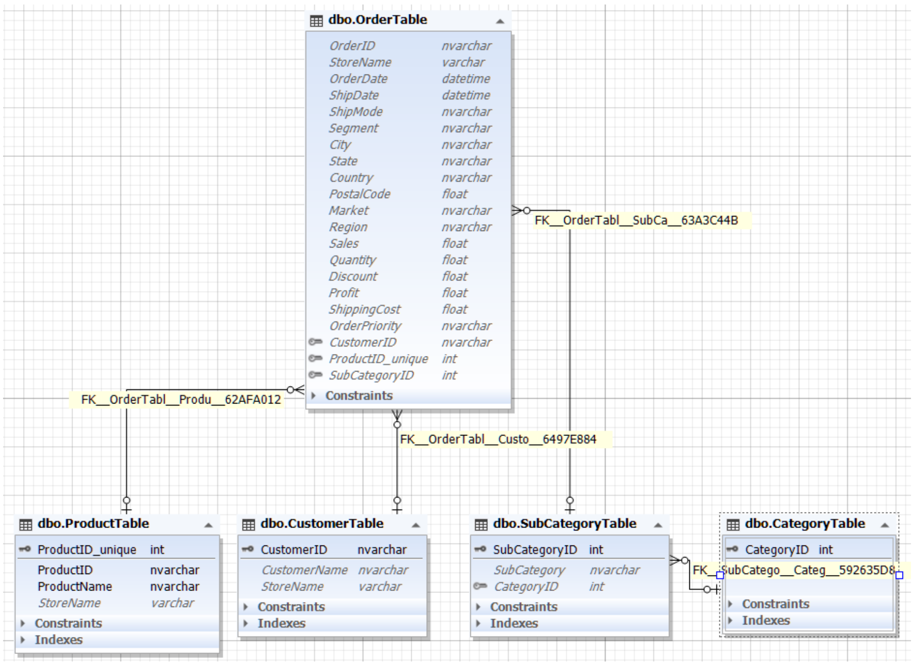


Figure 11: Entity Relationship Diagram (ERD)

2.2 Load

The final leg here will be separating the relational database into analysis ready tables. This depends very much on the day to day analysis objectives. For example, if daily analysis will be done separately on each company's data, then we can create splits of each company's tables. The aim is to put as many data that will be probed together often in one place as much as we can. Speed is king in analysis so we wouldn't want to be joining data from diferent tables during analysis. It could be splitting departmental data (Sales, HR, Marketing, by location, etc) into separate data marts.

3 Data Sources

3.1 Dataset 1

https://www.kaggle.com/avinashlalith/merkle-sokrati-advertising-campaign

3.2 Dataset 2

https://www.kaggle.com/apoorvaappz/global-super-store-dataset?select=Global_Superstore2.xlsx