Projects/Probelm Statement - Target Retail Market Analysis – SQL Project

Business Case: Target SQL
Context:

Target is a globally renowned brand and a prominent retailer in the United States. Target makes itself a preferred shopping destination by offering outstanding value, inspiration, innovation and an exceptional guest experience that no other retailer can deliver.

This particular business case focuses on the operations of Target in Brazil and provides insightful information about 100,000 orders placed between 2016 and 2018. The dataset offers a comprehensive view of various dimensions including the order status, price, payment and freight performance, customer location, product attributes, and customer reviews.

By analyzing this extensive dataset, it becomes possible to gain valuable insights into Target's operations in Brazil. The information can shed light on various aspects of the business, such as order processing, pricing strategies, payment and shipping efficiency, customer demographics, product characteristics, and customer satisfaction levels.

The data is available in 8 csv files:

customers.csv
sellers.csv
order_items.csv
geolocation.csv
payments.csv
reviews.csv
orders.csv
products.csv

The column description for these csv files is given below.

The customers.csv contain following features:

Features

Description

customer_id

ID of the consumer who made the purchase

customer_unique_id

Unique ID of the consumer

customer_zip_code_prefix

Zip Code of consumer’s location

customer_city

Name of the City from where order is made

customer_state

State Code from where order is made (Eg. são paulo - SP)

The sellers.csv contains following features:

Features

Description

seller_id

Unique ID of the seller registered

seller_zip_code_prefix

Zip Code of the seller’s location

seller_city

Name of the City of the seller

seller_state

State Code (Eg. são paulo - SP)

The order_items.csv contain following features:

Features

Description

order_id

A Unique ID of order made by the consumers

order_item_id

A Unique ID given to each item ordered in the order

product_id

A Unique ID given to each product available on the site

seller_id

Unique ID of the seller registered in Target

shipping_limit_date

The date before which the ordered product must be shipped

price

Actual price of the products ordered

freight_value

Price rate at which a product is delivered from one point to another

The geolocations.csv contain following features:

Features

Description

geolocation_zip_code_prefix

First 5 digits of Zip Code

geolocation_lat

Latitude

geolocation_lng

Longitude

geolocation_city

City

geolocation_state

State

The payments.csv contain following features:

Features

Description

order_id

A Unique ID of order made by the consumers

payment_sequential

Sequences of the payments made in case of EMI

payment_type

Mode of payment used (Eg. Credit Card)

payment_installments

Number of installments in case of EMI purchase

payment_value

Total amount paid for the purchase order

The orders.csv contain following features:

Features

Description

order_id

A Unique ID of order made by the consumers

customer_id

ID of the consumer who made the purchase

order_status

Status of the order made i.e. delivered, shipped, etc.

order_purchase_timestamp

Timestamp of the purchase

order_delivered_carrier_date

Delivery date at which carrier made the delivery

order_delivered_customer_date

Date at which customer got the product

order_estimated_delivery_date

Estimated delivery date of the products

The reviews.csv contain following features:

Features

Description

review_id

ID of the review given on the product ordered by the order id

order_id

A Unique ID of order made by the consumers

review_score

Review score given by the customer for each order on a scale of 1-5

review_comment_title

Title of the review

review_comment_message

Review comments posted by the consumer for each order

review_creation_date

Timestamp of the review when it is created

review_answer_timestamp

Timestamp of the review answered

The products.csv contain following features:

Features

Description

product_id

A Unique identifier for the proposed project.

product_category_name

Name of the product category

product_name_lenght

Length of the string which specifies the name given to the products ordered

product_description_lenght

Length of the description written for each product ordered on the site

product_photos_qty

Number of photos of each product ordered available on the shopping portal

product_weight_g

Weight of the products ordered in grams

product_length_cm

Length of the products ordered in centimeters

product_height_cm

Height of the products ordered in centimeters

product_width_cm

Width of the product ordered in centimeters

Problem Statement:

Assuming you are a data analyst/ scientist at Target, you have been assigned the task of analyzing the given dataset to extract valuable insights and provide actionable recommendations.

What does 'good' look like?

Import the dataset and do usual exploratory analysis steps like checking the structure & characteristics of the dataset:
Data type of all columns in the "customers" table.
Get the time range between which the orders were placed.
Count the Cities & States of customers who ordered during the given period.
In-depth Exploration:
Is there a growing trend in the no. of orders placed over the past years?
Can we see some kind of monthly seasonality in terms of the no. of orders being placed?
During what time of the day, do the Brazilian customers mostly place their orders? (Dawn, Morning, Afternoon or Night)
0-6 hrs : Dawn
7-12 hrs : Mornings
13-18 hrs : Afternoon
19-23 hrs : Night
Evolution of E-commerce orders in the Brazil region:
Get the month on month no. of orders placed in each state.
How are the customers distributed across all the states?
Impact on Economy: Analyze the money movement by e-commerce by looking at order prices,

freight and others.
Get the % increase in the cost of orders from year 2017 to 2018 (include months between Jan to Aug only).
You can use the "payment_value" column in the payments table to get the cost of orders.
Calculate the Total & Average value of order price for each state.
Calculate the Total & Average value of order freight for each state.
Analysis based on sales, freight and delivery time.
Find the no. of days taken to deliver each order from the order's purchase date as delivery time.
Also, calculate the difference (in days) between the estimated & actual delivery date of an order.
Do this in a single query.

You can calculate the delivery time and the difference between the estimated & actual delivery date using the given formula:
time_to_deliver = order_delivered_customer_date - order_purchase_timestamp
diff_estimated_delivery = order_estimated_delivery_date - order_delivered_customer_date
Find out the top 5 states with the highest & lowest average freight value.
Find out the top 5 states with the highest & lowest average delivery time.
Find out the top 5 states where the order delivery is really fast as compared to the estimated date of delivery.
You can use the difference between the averages of actual & estimated delivery date to figure out how fast the delivery was for each state.
Analysis based on the payments:
Find the month on month no. of orders placed using different payment types.
Find the no. of orders placed on the basis of the payment installments that have been paid.
_____