ML Assignment-2: Analytics File Creation – Customer 360 Creation

Business Context:
The client is one of the leading online marketplaces in India and would like to partner with Analytixlabs. The client seeks assistance in measuring, managing, and analyzing business performance.

The client wants to analyze their transaction data and understand the drivers for churners and high spenders. Along with that, the company wants to understand overall customer behavior so that it can help define the strategy to improve the revenues/margins.

What do you require to work?

To achieve the solution,

1. The team would like you to work on the "Customer360 data set" using instructions (description) provided in "Customer360 Instruction.xlsx". As part of this file, every record will be at the customer level (aggregating the information at the customer level).
2. Once you created the Customer 360 data set from step 1, you are required to formulate business problems related to regression, classification, and segmentation. Also, implement the solutions using Python & traditional algorithms like linear regression, logistic regression, and Heuristic Segmentation.

As part of this project, you are expected to clean the data (if required) before analyzing it.

Available Data:
Data has been provided for the period from September 2016 to October 2018, and the following is the data model.

Note: you are required to use two years of data (recent data – Nov-2016 to Oct-2018) and use period-1 as Nov-2016 to Oct-2017 and period-2 as Nov-2017 to October -2018

Tables:

Customers: Customers information

- Customer_id: Customer Unique ID
- customer_zip_code_prefix: Customer location zip code
- customer_city: Customer City
- customer_state: Customer state

Sellers: Sellers information

- seller_id: Seller Unique ID
- seller_zip_code_prefix: Seller location zip code
- seller_city: Seller City
- seller_state: Seller State

Products: Product information

Products: Product information

product_id: Product unique id
product_category_name: Product Category Name
product_name_length: Length of Product Name
product_description_length: Length of Product description
product_photos_qty: Number of photos
product_weight_gm: Product weight
product_length_cm: Product Length
product_height_cm: Product Height
product_width_cm: Product width

Orders: Orders info like ordered, product id, status, order dates etc…

order_id: Tranaction Unique ID
customer_id: Customer Unique ID
order_status: Status of Order
order_purchase_timestamp: Transaction Date & Time
order_approved_at: Order approved time
order_delivered_carrier_date: Order Shipping date & time
order_delivered_customer_date: Order delivered date & time
order_estimated_delivery_date: Estimated delivery date & time informed when order placed

Order_Items: Order level information

order_id - Order Unique ID
order_item_id - Quantity
product_id - Product unique id

seller_id - Seller ID
shipping_limit_date - Shipping date (Shipping limit data)
Price - Price per item
freight_value - Shipping cost

Order_Payments: Order payment information

order_id - Transaction ID
payment_sequential - Amortization schedule
payment_type - Payment Type
payment_installments - Number of instalments
payment_value - Order value

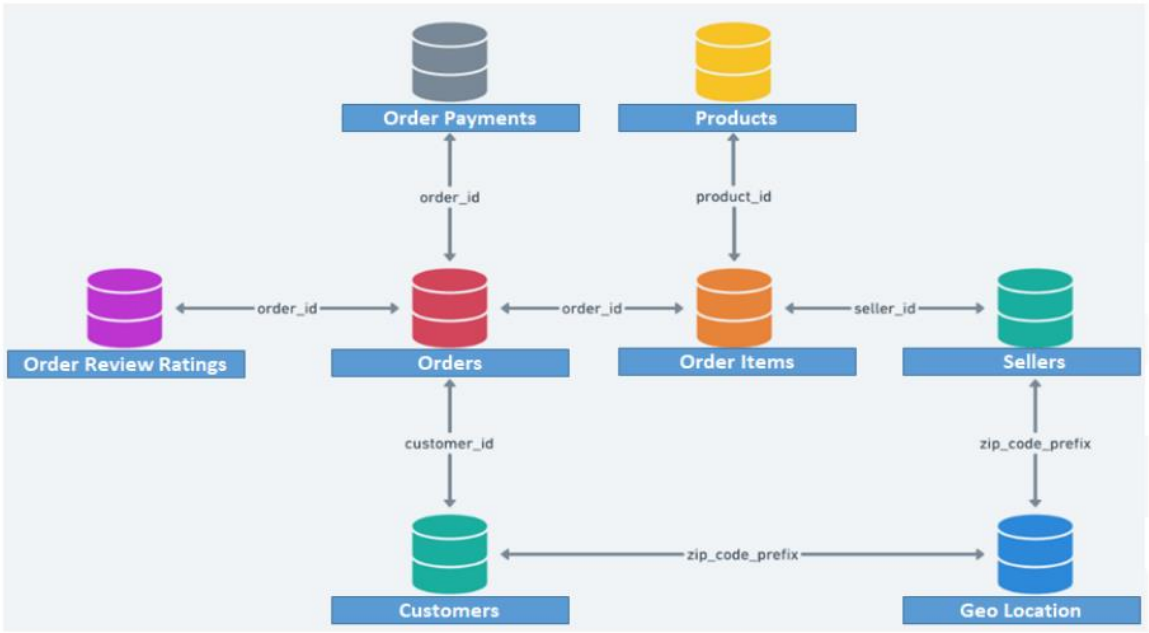
Order_Review_Ratings: Customer ratings at order level

review_id: Unique id for review
order_id: Order ID (Transaction ID)
review_score: Review rating score
review_creation_date: Review date
review_answer_timestamp: Review time

Geo-Location: Location details

geolocation_zip_code_prefix - Location zip code
geolocation_lat - Location Latitude
geolocation_lng- Location Longitude
geolocation_city - City
geolocation_state – State

Data Model:


Order Payments
Products
order_id
product_id
Order Review Ratings
order_id
Orders
order_id
Order Items
seller_id
Sellers
customer_id
zip_code_prefix
Customers
zip_code_prefix
Geo Location