FINAL PROJECT AND CERTIFICATION OF GROWIA

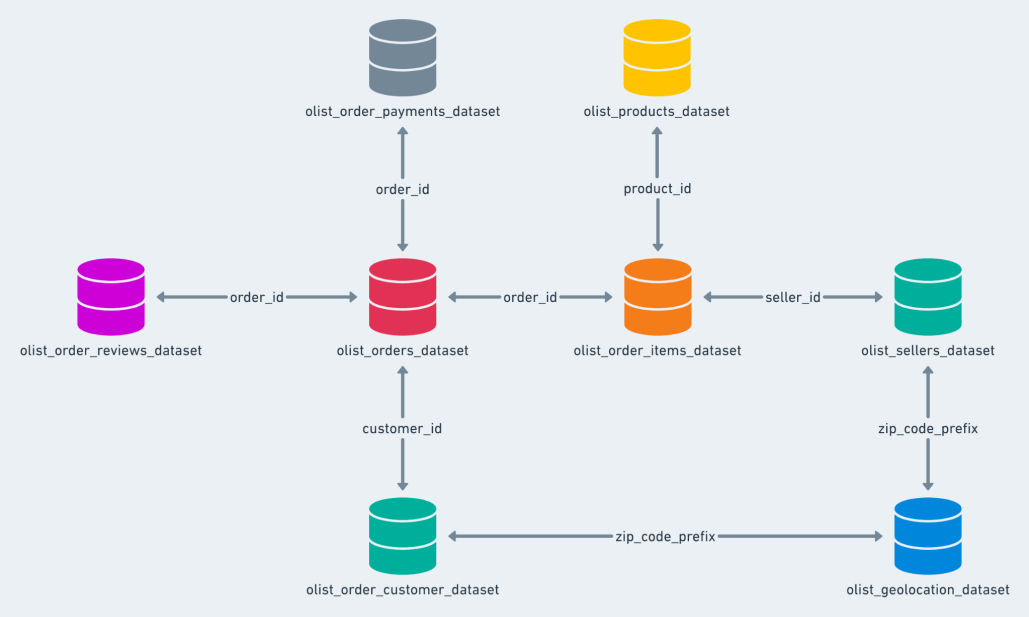
Study Case:

Halo Students, welcome to Growia final project, It will be fun and make you more confident in the data field.

OLIST is one of Brazil's largest department stores. OLIST's sales business concept is to connect small businesses from all over Brazil to channels so that these merchants can sell their products through OLIST Stores and deliver them to customers through OLIST's logistics partners.

As part of the data team, you are asked to do analysis with components that come from the business team. At this stage of the analysis, you are asked to use Python as an analysis and visualization tool. The information that appears is Key Metrics Business that has been defined by stakeholders.

The dataset framework used is as follows



***Source: Data Framework* OLIST Dataset**

The following are the steps used before working on this project. Some of the following business questions can be answered by exploring data and performing visualizations, so that you can dig into insights and make recommendations.

STEP 1: Import dataset by sql query file (*download here*) to your local computer and create connection between your sql and python

STEP 2: Import library needed (pandas, matplotlib, and seaborn)

STEP 3: Merge the dataset using their key connection based on dataset framework above

STEP 4: Deal Null Values

1. Null Value Identification
2. Would you please describe to us, why null values happen?
3. How to deal with our null value?

STEP 5: Add columns

1. Add columns based on the time period from the order purchase timestamp column
2. Add columns by year
3. Add columns by month
4. Add columns by name of the day
5. Add columns by hour
6. Add columns based on the actual price the customer purchased

   a. Total price value

   b. Total freight value

   c. Total order value (total price + total freight)

   E.g. Total Order Value

   a. Total price value = $95 (harga)

   b. Total freight value = $100 (kargo/pengantaran)

   c. Total order value = $95 + $100 = $195

STEP 6: Customer Demographic

1. Change the name of the state and city columns to address static and address city then do a grouping to calculate the total unique customer id based on the state address and it's visualization

   state → address_static

   city → address_city

2. Create a grouping to calculate the total unique customer id based on the address city and its visualization

3. Make a table regarding the total orders for each customer, then create a column that describes the frequency of orders for each customer (assuming = if below equals 5 it is called a little, otherwise it is a lot)
4. Create a table that lists the first and last purchase dates for each customer, then combine the first purchase, last purchase, and total order tables to generate the difference total order between the last purchase date and the first purchase date
5. What are conclusions from your exploratory?

STEP 7: Order Behaviour

1. Create a grouping table that calculates the average number of items ordered each based on total customers (customer_unique_id), then create a column that describes the frequency of orders for each customer (assuming = if below equals 5 it is called a little, otherwise it is a lot)
2. Create a grouping table that calculates average item weight and standard deviation of item weight
3. Create a grouping table that calculates the most frequent order times by month and day, and its time (morning, evening, or night)
4. Create a grouping table that calculates the average total order value per order and classifies the average order price for each customer with the assumption that prices above 2500 are called expensive, prices between 1000 - 2500 are called normal, prices below 1000 are called cheap
5. Create a grouping table that calculates the most popular payment methods and the preference of each customer's product category
6. What are the most ordered categories each year and how are the categories developing each year?

STEP 8: Review Customer

1. What is the average review score for each customer and create a column for the satisfaction level of each customer assuming a score of 4 & 5 is satisfied, otherwise not satisfied
2. Select a customer who fills in the title and message review, then create a table to calculate the total reviews for each customer and combine it with the total order table to find information on the review ratio per transaction for each customer

STEP 9: Time Series Analysis

1. Time series analysis is an analysis to see the trend of a phenomenon based on time. Try totaling the sales of each item per day. What are the sales trends? Is there an increase, decrease, or stagnant. What should business do?
2. Perform a seasonality analysis by looking at the total orders per day from Monday to Sunday.

STEP 10: Create your own visualization and business presentation!

Include :

Background

Objective

Data Explanation

Analysis

Suggestions/Conclusions/Recommendations

Goodluck