Olist Brazil Data Analysis

Introduction

This data analysis project focuses on Olist, Brazil's prominent online marketplace platform that bridges small business/independent sellers and customers across the country. By leveraging Olist's comprehensive dataset, which encompasses a variety of metrics including order details, customer information, seller information, payment information, and review scores, this study aims to extract actionable insights to enhance business operations, improve customer satisfaction, and drive strategic growth. The analysis will delve into patterns of consumer behavior, sales performance, and product popularity, aiming to identify key factors contributing to the platform's success and areas for potential improvement.

Dataset

The datasets, kindly provided by Olist, encompasses 99,441 order records from Brazil, covering the period from 2016 to 2018. It offers a comprehensive view of each order through multiple dimensions, including order status, pricing, payment methods, geolocations, and freight performance, as well as customer location, product attributes, and customer reviews. The complete datasets can be found here: [Brazilian E-Commerce Public Dataset by Olist](https://www.kaggle.com/datasets/olistbr/brazilian-ecommerce).

Data Preprocessing

After importing the datasets into MySQL, I noticed some columns are missing values and of wrong datatypes, we need to fix them before our analysis.

```sql
UPDATE orders
SET order_id = NULLIF(order_id, ''),
    customer_id = NULLIF(customer_id, ''),
    order_status = NULLIF(order_status, ''),
    order_purchase_timestamp = NULLIF(order_purchase_timestamp, ''),
    order_approved_at = NULLIF(order_approved_at, ''),
    order_delivered_carrier_date = NULLIF(order_delivered_carrier_date, ''),
    order_delivered_customer_date = NULLIF(order_delivered_customer_date, ''),
    order_estimated_delivery_date = NULLIF(order_estimated_delivery_date, '');

UPDATE order_items
SET order_id = NULLIF(order_id, ''),
    order_item_id = NULLIF(order_item_id, ''),
    product_id = NULLIF(product_id, ''),
    seller_id = NULLIF(seller_id, ''),
    shipping_limit_date = NULLIF(shipping_limit_date, ''),
    price = NULLIF(price, ''),
    freight_value = NULLIF(freight_value, '');

UPDATE products
SET product_id = NULLIF(product_id, ''),
    product_category_name = NULLIF(product_category_name, ''),
    product_name_lenght = NULLIF(product_name_lenght, ''),
    product_description_lenght = NULLIF(product_description_lenght, ''),
    product_photos_qty = NULLIF(product_photos_qty, ''),
    product_weight_g = NULLIF(product_weight_g, ''),
    product_length_cm = NULLIF(product_length_cm, ''),
    product_height_cm = NULLIF(product_height_cm, ''),
    product_width_cm = NULLIF(product_width_cm, '');

UPDATE payments
SET order_id = NULLIF(order_id, ''),
    payment_sequential = NULLIF(payment_sequential, ''),
    payment_type = NULLIF(payment_type, ''),
    payment_installments = NULLIF(payment_installments, ''),
    payment_value = NULLIF(payment_value, '');
```

```sql
ALTER TABLE orders
    MODIFY COLUMN order_id VARCHAR(255),
    MODIFY COLUMN customer_id VARCHAR(255),
    MODIFY COLUMN order_status VARCHAR(25),
    MODIFY COLUMN order_purchase_timestamp DATETIME,
    MODIFY COLUMN order_approved_at DATETIME,
    MODIFY COLUMN order_delivered_carrier_date DATETIME,
    MODIFY COLUMN order_delivered_customer_date DATETIME,
    MODIFY COLUMN order_estimated_delivery_date DATETIME;

ALTER TABLE order_items
    MODIFY COLUMN order_id VARCHAR(255),
    MODIFY COLUMN order_item_id INT,
    MODIFY COLUMN seller_id VARCHAR(255),
    MODIFY COLUMN shipping_limit_date DATETIME;

ALTER TABLE products
    MODIFY COLUMN product_id VARCHAR(255),
    MODIFY COLUMN product_category_name VARCHAR(255);

ALTER TABLE payments
    MODIFY COLUMN order_id VARCHAR(255),
    MODIFY COLUMN payment_type VARCHAR(25);

ALTER TABLE customers
    MODIFY COLUMN customer_id VARCHAR(255),
    MODIFY COLUMN customer_unique_id VARCHAR(255),
    MODIFY COLUMN customer_zip_code_prefix VARCHAR(5),
    MODIFY COLUMN customer_city VARCHAR(50),
    MODIFY COLUMN customer_state VARCHAR(2);

ALTER TABLE sellers
    MODIFY COLUMN seller_id VARCHAR(255),
    MODIFY COLUMN seller_zip_code_prefix VARCHAR(5),
    MODIFY COLUMN seller_city VARCHAR(50),
    MODIFY COLUMN seller_state VARCHAR(2);

ALTER TABLE reviews
    MODIFY COLUMN review_id VARCHAR(255),
    MODIFY COLUMN order_id VARCHAR(255),
    MODIFY COLUMN review_creation_date DATETIME,
    MODIFY COLUMN review_answer_timestamp DATETIME;
```