

Case Study: E-Commerce Payment and Delivery Analytics

Background:

You are working as a data analyst for a large e-commerce company that operates across various product categories, including electronics, fashion, beauty, pet supplies, and more. The company receives millions of orders each year and relies on various forms of payments such as credit cards, debit cards, bank transfers, and digital wallets. Customers provide feedback through reviews, which range from 1 to 5 stars, and these reviews are linked to specific orders. Additionally, products are delivered by different carriers, with delivery times varying across regions.

Your task is to perform an in-depth analysis of payment statistics, delivery performance, and customer satisfaction to help improve operational efficiency, customer experience, and marketing strategies.

Objectives:

The case study consists of several key performance indicators (KPIs) that will guide your analysis. These KPIs will help you understand customer behavior, evaluate delivery efficiency, and explore the impact of different variables on customer satisfaction.

KPI 1: Weekday vs Weekend Payment Statistics

Task: Analyze the payment statistics by separating orders made on weekdays from those made on weekends. Determine if there are any notable differences in payment behavior between these two groups.

- **Steps:**
 1. Classify orders as weekday or weekend based on the `order_purchase_timestamp`.
 2. Calculate payment statistics such as total payments, average payment value, and the breakdown of payment types (credit card, debit card, etc.) for both weekdays and weekends.
 3. Identify any significant differences in payment behavior (e.g., is there a higher preference for credit card payments on weekends?).

Expected Outcome: Insights into how payment methods vary between weekdays and weekends, which could influence promotional campaigns or payment system optimizations.

KPI 2: Count of Orders with Review Score 5 and Payment Type as Credit Card

Task: Determine how many orders received a review score of 5 (the highest rating) and were paid using a credit card. This helps assess customer satisfaction and payment method preferences.

- **Steps:**
 1. Filter the dataset to find orders with a review score of 5 and a payment type of "credit card."

2. Count the total number of such orders.
3. Investigate the correlation between payment method and satisfaction levels.

Expected Outcome: Understanding of how many satisfied customers (rating 5) prefer to use credit cards, which could help tailor marketing and loyalty programs.

KPI 3: Average Delivery Time for Pet Shop Products

Task: Calculate the average number of days it takes to deliver orders from the "pet_shop" category to customers. This analysis helps assess delivery efficiency for this category.

- **Steps:**

1. Filter orders that belong to the "pet_shop" category.
2. Calculate the difference between the order_delivered_customer_date and order_purchase_timestamp for each order.
3. Calculate the average delivery time across all "pet_shop" orders.

Expected Outcome: Insights into the delivery efficiency of the "pet_shop" category, which can be used to improve customer satisfaction and identify potential logistical bottlenecks.

KPI 4: Average Order Price and Payment Amount for Customers in São Paulo

Task: Calculate the average order price and the average payment amount for customers located in São Paulo. This analysis will help understand the purchasing behavior of customers in one of the company's key markets.

- **Steps:**

1. Filter the dataset for customers located in São Paulo.
2. Calculate the average order price (product_price) and average payment value (payment_value) for these customers.
3. Compare the results to customers from other regions (optional).

Expected Outcome: A deeper understanding of the spending behavior of São Paulo customers, which could inform regional marketing strategies, pricing decisions, and stock allocation.

KPI 5: Relationship Between Shipping Days and Review Scores

Task: Analyze the relationship between the number of shipping days (time taken for delivery) and the review scores given by customers. Does faster delivery correlate with higher customer satisfaction?

- **Steps:**

1. Calculate the number of shipping days for each order by subtracting the order_purchase_timestamp from the order_delivered_customer_date.

- 2. Group the data by review score and calculate the average number of shipping days for each score (1 to 5).
- 3. Analyze whether shorter delivery times lead to higher review scores, and identify any trends.

Expected Outcome: Insights into whether delivery speed affects customer satisfaction, and how companies can optimize their delivery strategies to improve reviews.

Data Dictionaries

Table Name: Customer_Details

Description: This table contains customer demographic and location details from the Olist platform.

Column Name	Description	Data Type	Explanation
customer_id	Unique customer identifier	String	A unique key representing each customer in the Olist system (used internally for operations and mapping).
customer_unique_id	Persistent customer identifier	String	A unique value representing a customer across all their orders, enabling tracking and analysis of customer behavior over time.
customer_zip_code_prefix	Customer's ZIP code prefix	Integer	The first five digits of the ZIP code of the customer, giving information about their locality or area.
customer_city	Customer's city	String	The name of the city where the customer resides, providing geographical segmentation for analysis.
customer_state	Customer's state	String	The two-letter abbreviation of the state where the customer resides (e.g., SP for São Paulo, MG for Minas Gerais). Useful for regional analysis.

Table Name: Geolocation_Details

Description: This table contains geographical information for locations associated with customers or orders, including coordinates and city/state details.

Column Name	Description	Data Type	Explanation
geolocation_zip_code_prefix	ZIP code prefix of the geographical location	Integer	The first five digits of the ZIP code for the given geographical location, used to categorize areas for mapping and analysis.
geolocation_lat	Latitude coordinate of the location	Float	The geographic latitude (north-south position) of the location, useful for mapping and geospatial analysis.
geolocation_lng	Longitude coordinate of the location	Float	The geographic longitude (east-west position) of the location, allowing for precise geospatial mapping and distance calculations.
geolocation_city	The city associated with the geographical location	String	The name of the city linked to the geographical coordinates, useful for regional grouping and analysis.
geolocation_state	State abbreviation of the geographical location	String	The two-letter abbreviation for the state of the given location (e.g., SP for São Paulo), aiding in regional segmentation for analysis.

Table Name: Order_Details

Description: This table contains detailed information about customer orders, including order IDs, product IDs, pricing, shipping details, and associated seller information.

Column Name	Description	Data Type	Explanation
order_id	Unique identifier for the order	String	A unique ID for each order placed by customers, allowing tracking of individual purchases.
order_item_id	Unique identifier for the order item	Integer	A unique ID for each item in the order, used to track individual products within an order.

Column Name	Description	Data Type	Explanation
product_id	Unique identifier for the product	String	The unique identifier for the product being purchased, used to link to the product catalog.
seller_id	Unique identifier for the seller	String	The unique ID of the seller fulfilling the order, linking to the seller's store in the system.
shipping_limit_date	Date and time by which the order should be shipped	DateTime	The deadline by which the seller must ship the order. Provides important logistics information.
price	Price of the product in the order	Float	The price of the individual product being sold, excluding any additional fees like freight or taxes.
freight_value	Shipping cost for the order	Float	The cost of shipping the product, representing the additional charges for delivery.

Table Name: Payment_Details

Description: This table contains payment information for each order, including payment method, installments, and payment values.

Column Name	Description	Data Type	Explanation
order_id	Unique identifier for the order	String	A unique ID for each order placed, used to track payments related to that specific order.
payment_sequential	Sequential number for the payment in the order	Integer	The sequence number of the payment in the order. Multiple payments can exist for a single order, especially if paid in installments.
payment_type	Type of payment method used	String	The method used for payment, such as credit_card, debit_card, boleto, etc.
payment_installments	Number of installments for the payment	Integer	The number of installments chosen for the payment, indicating whether the payment is split into multiple parts.

Column Name	Description	Data Type	Explanation
payment_value	Total value of the payment	Float	The amount paid in the specific installment or full payment, representing the monetary value processed for the order.

Table Name: Product_Reviews

Description: This table contains product reviews provided by customers, including the review score, comments, and timestamps for when the review was created and answered.

Column Name	Description	Data Type	Explanation
review_id	Unique identifier for the review	String	A unique ID assigned to each review, used to track individual reviews and link them to specific orders.
order_id	Unique identifier for the order associated with the review	String	The unique ID of the order related to the review, used to correlate customer feedback with order data.
review_score	Rating score given by the customer	Integer	The numerical score (usually from 1 to 5) that the customer assigns to the product based on their experience.
review_comment_title	Title of the customer's review	String	A short title provided by the customer summarizing their review or feedback. It is often a brief expression of their opinion.
review_comment_message	Detailed message provided by the customer in the review	String	A detailed message from the customer providing feedback or comments about the product, order experience, or shipping.
review_creation_date	Date and time when the review was created	DateTime	The timestamp indicating when the review was created by the customer. It helps track review submission timings.
review_answer_timestamp	Date and time when the review was	DateTime	The timestamp indicating when the seller or customer service responded to the review.

Column Name	Description	Data Type	Explanation
	answered (if applicable)		

Table Name: Orders

Description: This table contains order information including order status, purchase timestamp, and delivery details.

Column Name	Description	Data Type	Explanation
order_id	Unique identifier for the order	String	A unique ID assigned to each order, used to track individual orders.
customer_id	Unique identifier for the customer	String	A unique ID assigned to each customer, used to associate the order with the customer who made the purchase.
order_status	Current status of the order	String	The status of the order, such as 'delivered', 'invoiced', etc.
order_purchase_timestamp	Date and time when the order was purchased	DateTime	The timestamp when the order was created or purchased by the customer.
order_approved_at	Date and time when the order was approved	DateTime	The timestamp when the order was approved, indicating when the payment was processed.
order_delivered_carrier_date	Date and time when the order was delivered to the carrier	DateTime	The timestamp when the order was handed over to the delivery carrier.
order_delivered_customer_date	Date and time when the order was delivered to the customer	DateTime	The timestamp when the order was received by the customer.
order_estimated_delivery_date	Estimated date for the order delivery	DateTime	The estimated date when the order is expected to be delivered to the customer.

Table Name: Products

Description: This table contains product information such as product category, dimensions, weight, and description lengths.

Column Name	Description	Data Type	Explanation
product_id	Unique identifier for the product	String	A unique ID assigned to each product.
product_category_name	Name of the product category	String	The category under which the product falls, e.g., 'perfumaria', 'artes', etc.
product_name_length	Length of the product name	Integer	The number of characters in the product's name.
product_description_length	Length of the product description	Integer	The number of characters in the product's description.
product_photos_qty	Quantity of product photos	Integer	The number of photos associated with the product.
product_weight_g	Product weight in grams	Integer	The weight of the product in grams.
product_length_cm	Product length in centimeters	Integer	The length of the product in centimeters.
product_height_cm	Product height in centimeters	Integer	The height of the product in centimeters.
product_width_cm	Product width in centimeters	Integer	The width of the product in centimeters.

Table Name: Sellers

Description: This table contains details about sellers, including their location and associated ZIP code.

Column Name	Description	Data Type	Explanation
seller_id	Unique identifier for the seller	String	A unique ID assigned to each seller.

Column Name	Description	Data Type	Explanation
seller_zip_code_prefix	ZIP code prefix of the seller's location	Integer	The prefix of the ZIP code where the seller is located.
seller_city	City where the seller is located	String	The name of the city where the seller's business is based.
seller_state	State where the seller is located	String	The state where the seller's business is based (e.g., 'SP', 'RJ', 'PE').

Table Name: Product Categories

Description: This table includes the mapping of product categories to their English equivalents.

Column Name	Description	Data Type	Explanation
product_category_name	Original product category name	String	The product category name in the native language (e.g., 'beleza_saude' for Health & Beauty).
product_category_name_english	English translation of the product category name	String	The English translation for the product category (e.g., 'health_beauty' for Health & Beauty).

Conclusion:

At the end of this case study, you will have gained a deeper understanding of customer purchasing behavior, delivery efficiency, and satisfaction levels. You will be able to provide actionable insights to help optimize payment methods, improve delivery logistics, and enhance the customer experience, leading to increased sales and customer loyalty.