

## Introduction to the Look E-commerce Dataset

The Look E-commerce dataset provides a comprehensive view of online retail operations, covering sales transactions, customer information, product attributes, website engagement metrics, logistics, and metadata. It serves as a valuable resource for understanding and analyzing e-commerce business activities.

Data sources:

- [Big query](#)
  - [Kaggle](#)
- 

## Data Dictionary

### 1. distribution\_centers

- **id:** Unique identifier for each distribution center.
- **name:** Name of the distribution center.
- **latitude:** Latitude coordinate of the distribution center.
- **longitude:** Longitude coordinate of the distribution center.
- **distribution\_center\_geom:** Geometric data representing the location of the distribution center.

### 2. events

- **id:** Unique identifier for the event.
- **user\_id:** Identifier linking the event to a specific user.
- **sequence\_number:** Order of the events in a session.
- **session\_id:** Identifier for the user's session.
- **created\_at:** Timestamp when the event occurred.
- **ip\_address:** IP address from which the event was generated.
- **city:** City associated with the event.
- **state:** State where the event occurred.
- **postal\_code:** Postal code associated with the event.
- **browser:** Browser used by the user during the event.
- **traffic\_source:** Source through which the user arrived. Common values include:
  - **Adwords:** Traffic from paid Google search advertisements.
  - **Email:** Visitors arriving via links in marketing or transactional emails.
  - **YouTube:** Traffic from YouTube ads or video descriptions.
  - **Organic:** Traffic from unpaid search results.
  - **Facebook:** Traffic originating from Facebook ads or posts.
- **uri:** Uniform Resource Identifier (URI) visited during the event.
- **event\_type:** Type of action performed by the user. Common values include:
  - **cancel:** Action to cancel an order or session.

- **cart:** Interaction with the shopping cart (e.g., adding or removing items).
- **department:** Browsing or interacting with a product department.
- **home:** Navigation on the home page of the website.
- **product:** Viewing or interacting with a product page.
- **purchase:** Completing a purchase transaction.

### Example of Events Table:

id	user_id	sequence_number	session_id	country	ip	city	state	postal_code	brand	traffic_source	url	event_type
1040586	79964	1	00002769-7a3b-43ee-8fa3-6c6c8f0a6d96							Email	/department	department
1040587	79964	2	00002769-7a3b-43ee-8fa3-6c6c8f0a6d96							Email		product
1040588	79964	3	00002769-7a3b-43ee-8fa3-6c6c8f0a6d96							Email		cart
1040589	79964	4	00002769-7a3b-43ee-8fa3-6c6c8f0a6d96							Email	/department	department
1040590	79964	5	00002769-7a3b-43ee-8fa3-6c6c8f0a6d96							Email		product
1040591	79964	6	00002769-7a3b-43ee-8fa3-6c6c8f0a6d96							Email		cart
1040592	79964	7	00002769-7a3b-43ee-8fa3-6c6c8f0a6d96							Email		purchase
746440	57137	1	00005a06-8a06-459e-95ae-74db0859a1...							Adwords		home

### 3. inventory\_items

- **id:** Unique identifier for the inventory item.
- **product\_id:** Identifier linking the item to a product.
- **created\_at:** Timestamp when the item was added to inventory.
- **sold\_at:** Timestamp when the item was sold.
- **cost:** Cost price of the inventory item. This reflects the actual cost of the item in inventory, which may vary due to specific purchase agreements or shipment costs.
- **product\_category:** Category of the product.
- **product\_name:** Name of the product.
- **product\_brand:** Brand of the product.
- **product\_retail\_price:** Retail price of the product.
- **product\_department:** Department where the product belongs.
- **product\_sku:** Stock keeping unit for the product.

### 4. order\_items

- **id:** Unique identifier for the order item.
- **order\_id:** Identifier linking the item to an order.

- **user\_id**: Identifier for the user who placed the order.
- **product\_id**: Identifier linking the item to a product.
- **inventory\_item\_id**: Identifier linking the item to inventory.
- **status**: Status of the order item (e.g., shipped, delivered).
- **created\_at**: Timestamp when the item was ordered.
- **shipped\_at**: Timestamp when the item was shipped.
- **delivered\_at**: Timestamp when the item was delivered.
- **returned\_at**: Timestamp when the item was returned.
- **sale\_price**: Price at which the item was sold. This may differ from the product's retail price due to discounts, promotions, or custom pricing.

Example of orders item Table:

Row	id	order_id	user_id	product_id	inventory_item_id	st	cr	sh	de	re	sale_price
1	158470	109106	87559	14322	427754						82.0
2	158471	109106	87559	15013	427758						130.0
3	158472	109106	87559	15013	427761						130.0
4	158473	109106	87559	8556	427765						285.0

## 5. orders

- **order\_id**: Unique identifier for the order.
- **user\_id**: Identifier for the user who placed the order.
- **status**: Current status of the order.
- **gender**: Gender of the user who placed the order.
- **created\_at**: Timestamp when the order was created.
- **returned\_at**: Timestamp if the order was returned.
- **shipped\_at**: Timestamp when the order was shipped.
- **delivered\_at**: Timestamp when the order was delivered.
- **num\_of\_items**: Number of items in the order.

## 6. products

- **id**: Unique identifier for the product.
- **cost**: Cost price of the product. This represents the standard cost of the product as listed in the product catalog, excluding any real-time inventory-specific adjustments.
- **category**: Category of the product.
- **name**: Name of the product.
- **brand**: Brand of the product.
- **retail\_price**: Retail price of the product. This represents the standard listed price, excluding discounts or promotions applied during sales.
- **department**: Department of the product.
- **sku**: Stock keeping unit of the product.
- **distribution\_center**: Identifier linking the product to a distribution center.

## 7. users

- **id:** Unique identifier for the user.
- **first\_name:** First name of the user.
- **last\_name:** Last name of the user.
- **email:** Email address of the user.
- **age:** Age of the user.
- **gender:** Gender of the user.
- **state:** State of residence.
- **street\_address:** Street address of the user.
- **postal\_code:** Postal code of the user.
- **city:** City of residence.
- **country:** Country of residence.
- **traffic\_source:** Source of the user's website traffic (e.g., organic, paid, referral).
- **created\_at:** Timestamp when the user account was created.
- **user\_geom:** Geometric data representing the user's location.
- **latitude:** Latitude coordinate of the user's location.
- **longitude:** Longitude coordinate of the user's location.

---

## Project Brief

### 1. Select an Area of Focus:

Analyze a specific business aspect, such as:

- Sales performance
- Customer behavior
- Inventory efficiency
- Website performance
- Marketing effectiveness

### 2. Define a Problem Statement Aligned with Business Goals

Think about strategic goals such as:

- Increasing sales
- Improving customer retention
- Optimizing inventory
- Optimizing marketing budget allocation

*Before defining your problem statement, take time to research the eCommerce industry and understand common business challenges. Explore customer behavior, sales trends, inventory management, and marketing strategies to gain domain knowledge. This will help you ask the right questions and align your analysis with real business needs.*

### **3. Identify Key Business Questions:**

Formulate specific, measurable, and data-driven business questions to address the issue. Below are some examples:

- Sales Performance:
  - What percentage of orders are returned, and what are the most common reasons for returns?
- Customer Behavior:
  - What percentage of customers make repeat purchases, and how does this vary by demographics?

**Good Luck and Have Fun!**