# E-commerce Data Description

The database comprises 7 interconnected datasets for an e-commerce platform. It includes information on distribution centers, user events, inventory items, order details, products, and user profiles.

The data provides a holistic view of business analytics and decision-making in the e-commerce domain.

1. Distribution Centers: Here is a data description for a distribution\_center dataset which has 10 rows and 4 columns:

* id (INTEGER): A unique identifier for each distribution center.
* name (STRING): The name or label associated with the distribution center.
* latitude (FLOAT): The geographical latitude coordinate of the distribution center's location.
* longitude (FLOAT): The geographical longitude coordinate of the distribution center's location.

1. Events: Here is a data description for events dataset, which has 24,27,931 rows and 13 columns:

* id (INTEGER): A unique identifier for each event, serving as a primary key.
* user\_id (INTEGER): The identifier associated with the user who generated the event. This links the event to a specific user.
* sequence\_number (INTEGER): An integer indicating the order or sequence of the event within a session. It helps to understand the chronological flow of events for a given user.
* session\_id (STRING): A unique identifier for the session during which the event occurred. Sessions group together a series of related user interactions.
* created\_at (TIMESTAMP): The timestamp indicating when the event occurred, capturing both date and time information.
* ip\_address (STRING): The IP address associated with the user when the event was generated. IP addresses can be used to identify the location or network from which the event originated.
* city (STRING): The city associated with the user's location when the event occurred.
* state (STRING): The state or region associated with the user's location when the event occurred.
* postal\_code (STRING): The postal code associated with the user's location when the event occurred.
* browser (STRING): The web browser used by the user when generating the event.
* traffic\_source (STRING): The source from which the user's traffic originated, indicating how the user arrived at the platform or website.
* uri (STRING): The Uniform Resource Identifier associated with the event, providing information about the specific resource or page the user interacted with.
* event\_type (STRING): The type or category of the event, providing insight into the nature of the user's interaction

1. Inventory: Here's a data description for the "inventory\_items" dataset which has 4,91,098 rows and 12 columns

* id (INTEGER): A unique identifier for each inventory item, serving as a primary key.
* product\_id (INTEGER): The identifier associated with the specific product to which the inventory item corresponds.
* created\_at (TIMESTAMP): The timestamp indicating when the inventory item was created or added to the inventory.
* sold\_at (TIMESTAMP): The timestamp indicating when the inventory item was sold. This provides information on the date and time of the item's sale.
* cost (FLOAT): The cost associated with the inventory item, representing the expense incurred to acquire or produce the item.
* product\_category (STRING): The category to which the product belongs, providing a high-level classification of the item.
* product\_name (STRING): The name or label of the product associated with the inventory item.
* product\_brand (STRING): The brand associated with the product.
* product\_retail\_price (FLOAT): The retail price of the product, indicating the price at which the product is sold to customers.
* product\_department (STRING): The department to which the product belongs
* product\_sku (STRING): Stock Keeping Unit (SKU) associated with the product, providing a unique identifier for tracking the item in inventory.
* product\_distribution\_center\_id (INTEGER): The identifier of the distribution center associated with the product.

1. Order items: Here's a data description for the "order\_items" dataset which has 1,81,891 rows and 11 columns

* id (INTEGER): A unique identifier for each order item, serving as a primary key.
* order\_id (INTEGER): The identifier associated with the specific order to which the item belongs.
* user\_id (INTEGER): The identifier associated with the user who placed the order.
* product\_id (INTEGER): The identifier associated with the specific product included in the order.
* inventory\_item\_id (INTEGER): The identifier of the inventory item associated with the ordered product. This links the order item to a specific inventory item.
* status (STRING): The status of the ordered item
* created\_at (TIMESTAMP): The timestamp indicates when the order item was created
* shipped\_at (TIMESTAMP): The timestamp indicating when the order item was shipped
* delivered\_at (TIMESTAMP): The timestamp indicating when the order item was delivered to the customer.
* returned\_at (TIMESTAMP): The timestamp indicating when the order item was returned by the customer, if applicable.
* sale\_price (FLOAT): The price at which the product was sold as part of this order item.

1. Orders: Here's a data description for the "orders" dataset which has 1,25,530 rows and 9 columns

* order\_id (INTEGER): A unique identifier for each order, serving as a primary key.
* user\_id (INTEGER): The identifier associated with the user who placed the order.
* status (STRING): The status of the order, indicating its current state in the order fulfillment process
* gender (STRING): The gender associated with the user who placed the order.
* created\_at (TIMESTAMP): The timestamp indicating when the order was created.
* returned\_at (TIMESTAMP): The timestamp indicating when the order was returned by the customer, if applicable.
* shipped\_at (TIMESTAMP): The timestamp indicating when the order was shipped from the distribution center or warehouse.
* delivered\_at (TIMESTAMP): The timestamp indicating when the order was delivered to the customer.
* num\_of\_item (INTEGER): The number of items included in the order. This provides information about the size or quantity of the order.

1. Products: Here's a data description for the "products" dataset which has 29,120 rows and 9 columns

* id (INTEGER): A unique identifier for each product, serving as a primary key.
* cost (FLOAT): The cost associated with producing or acquiring the product.
* category (STRING): The category to which the product belongs, providing a high-level classification.
* name (STRING): The name or label of the product.
* brand (STRING): The brand associated with the product.
* retail\_price (FLOAT): The retail price at which the product is sold to customers.
* department (STRING): The department to which the product belongs, providing a more granular classification within the product categories.
* sku (STRING): Stock Keeping Unit (SKU) associated with the product, providing a unique identifier for tracking the product.
* distribution\_center\_id (INTEGER): The identifier of the distribution center associated with the product.

1. Users: Here's a data description for the "users" dataset, which has 1,00,000 rows and 15 columns

* id (INTEGER): A unique identifier for each user, serving as a primary key.
* first\_name (STRING): The first name of the user.
* last\_name (STRING): The last name of the user.
* email (STRING): The email address associated with the user.
* age (INTEGER): The age of the user.
* gender (STRING): The gender of the user.
* state (STRING): The state or region where the user is located.
* street\_address (STRING): The street address of the user.
* postal\_code (STRING): The postal code associated with the user's location.
* city (STRING): The city where the user is located.
* country (STRING): The country where the user is located.
* latitude (FLOAT): The geographical latitude coordinate of the user's location.
* longitude (FLOAT): The geographical longitude coordinate of the user's location.
* traffic\_source (STRING): The source from which the user's traffic originated, indicating how the user discovered or accessed the platform.
* created\_at (TIMESTAMP): The timestamp indicates when the user account was created.