

# CS364 Documentation Final

## Differences from ER Diagram

The original ER diagram planned for tables like `Products`, `Customers`, `Orders`, `Reviews`, and `Categories`. In practice, due to the unavailability of the customer dataset, I had to adapt and work with the available "eCommerce Behavior Data". This led to a focus on `Products`, `Users`, and `Events` tables primarily, reflecting the dataset's structure of user interactions and product details.

## Schema Changes

The actual database schema is significantly different than the initial ER diagram. I removed multiple interconnected entities like `Customers`, `Orders`, and `Reviews`, and simplified it. The `Products` table was retained but adjusted to the dataset specifics, and `Users` and `Events` tables were introduced to capture the user activity and product interactions.

## Data Cleaning and Preparation

For data insertion, the raw dataset required cleaning and formatting to fit the SQL schema. This included parsing and converting timestamps to a MySQL-friendly format. I also had to download the 2019 csv file from original Kaggle since the 2020 was too large. Moreover, I removed all but 50 rows of the data to have an easier time opening it.

## Challenges in Data Cleaning

One challenge was the date-time format conversion, where the original 'UTC' timestamp needed to be reformatted to match MySQL's datetime standards. Additionally, handling missing data, especially for categories and brands in the `Products` table, required careful consideration to maintain data integrity.

## View and Trigger Implementation

The `ProductSummary` view was created to provide a quick summary of product interactions, offering insights into views and purchases per product. This aligns with the database's purpose of tracking user-product interactions and aids in decision-making for inventory and marketing strategies.

The `BeforePriceUpdate` trigger was implemented to log changes in product prices, enhancing the database's ability to audit and track significant data changes. This addition serves to maintain a

history of price adjustments, crucial for financial analysis and strategic planning.

## **Conclusion and Future Directions**

The shift from the proposed comprehensive e-commerce platform to a more focused user-product interaction database was a significant pivot in the project's direction.

The challenges faced, particularly in data preparation and schema adjustment, underscored the need for robust data cleaning and validation processes. Going forward, improving data validation, expanding the database to include more purchasing data from customers. Also, if I had to choose a new database, I would try to understand a dataset by actually finishing the ERD diagram.