IIT MADRAS
ZANZIBAR

Data Engineering & Management

ZZ104

Shopping Cart Analysis

From Wishlist to Cart... and Sometimes, Goodbye

Problem Statement

Using e-commerce transaction data, identify customer purchasing patterns, seasonal trends, and abandoned cart characteristics to generate insights for improving conversion rates.

Tech Stack

- Backend: Python, WebSocket, psycopg2
- Database: PostgreSQL
- Dashboard: Streamlit, Plotly
- Data Generation: Faker, FakeStore API
- Real-time: asyncio, websockets
- Containerization: Docker, Docker Compose

Project Methodology

- We got product information from Fakestore API
- Then we got 100 customers using the Faker Library
- We had to find a way to get realtime data flowing - so we created a simulator that randomly simulates customer actions (adding/removing from cart, purchasing products)
- Websocket server listens in, stores it into PostgreSQL and broadcasts it to the dashboard

Key Assumptions

3 actions - purchasing products, adding to cart or removing it.

Add - remove - purchased = abandonment rate

The data is randomly generated - so we do not know the true accuracy of the data.

That being said, we can connect our dashboard to realtime data

Database Design

Customers Table

Column Name	Type	Description
customer_id	SERIAL (PK)	Unique customer identifier
name	VARCHAR(100)	Customer's full name
age	INT	Customer's age (0–120, validated)
email	VARCHAR(255)	Unique email address
created_at	TIMESTAMP	When the customer was created

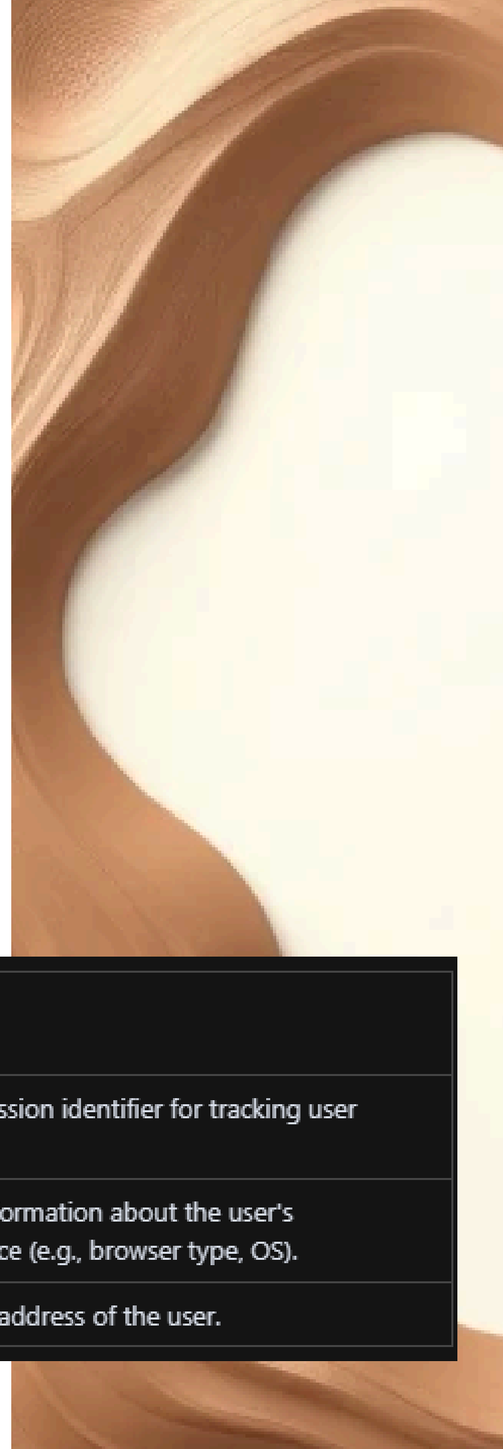
Products Table

Column Name	Type	Description
product_id	SERIAL (PK)	Unique product identifier
title	VARCHAR(255)	Product name/title
price	DECIMAL(10,2)	Product price
description	TEXT	Product description
image_url	VARCHAR(500)	URL to product image
category	VARCHAR(100)	Product category
created_at	TIMESTAMP	When the product was added

Database Design

Events Table

Column Name	Type	Description
event_id	VARCHAR(255)	Unique identifier for each event (UUID).
customer_id	INTEGER	References the customer who performed the action.
product_id	INTEGER	References the product involved in the event.
product_title	VARCHAR(255)	The name/title of the product (from FakeStore API).
product_price	DECIMAL(10,2)	The price of the product at the time of the event.
product_image	VARCHAR(500)	URL to the product image.
action	VARCHAR(50)	The type of event: 'add_to_cart', 'remove_from_cart', or 'purchase_cart'.
description	TEXT	Human-readable description of the event (e.g., "Added X to cart").
timestamp	TIMESTAMP	When the event occurred.



Fields for future use

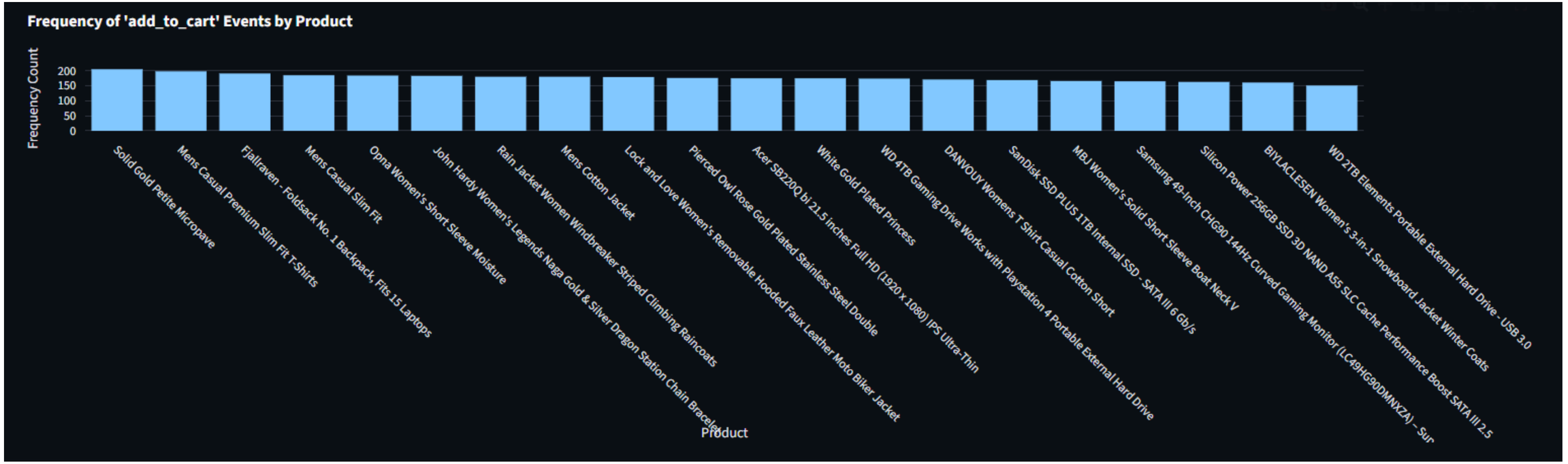
Column Name	Type	Description
session_id	VARCHAR(255)	(Optional) Session identifier for tracking user sessions.
user_agent	TEXT	(Optional) Information about the user's browser/device (e.g., browser type, OS).
ip_address	INET	(Optional) IP address of the user.

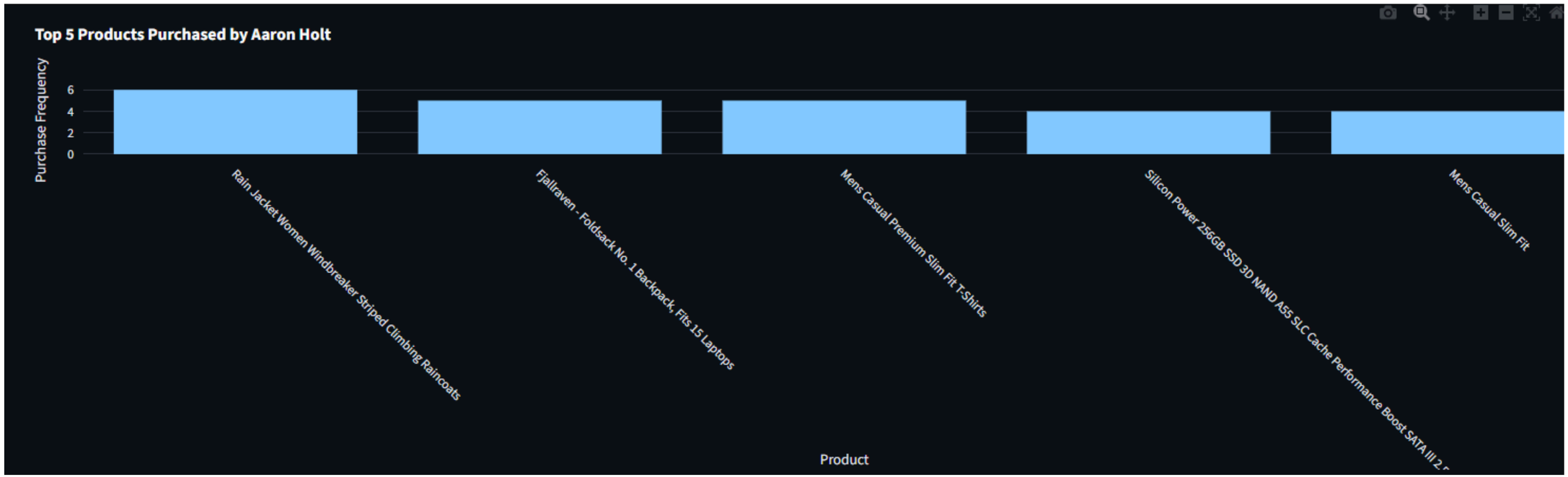
Design Choices

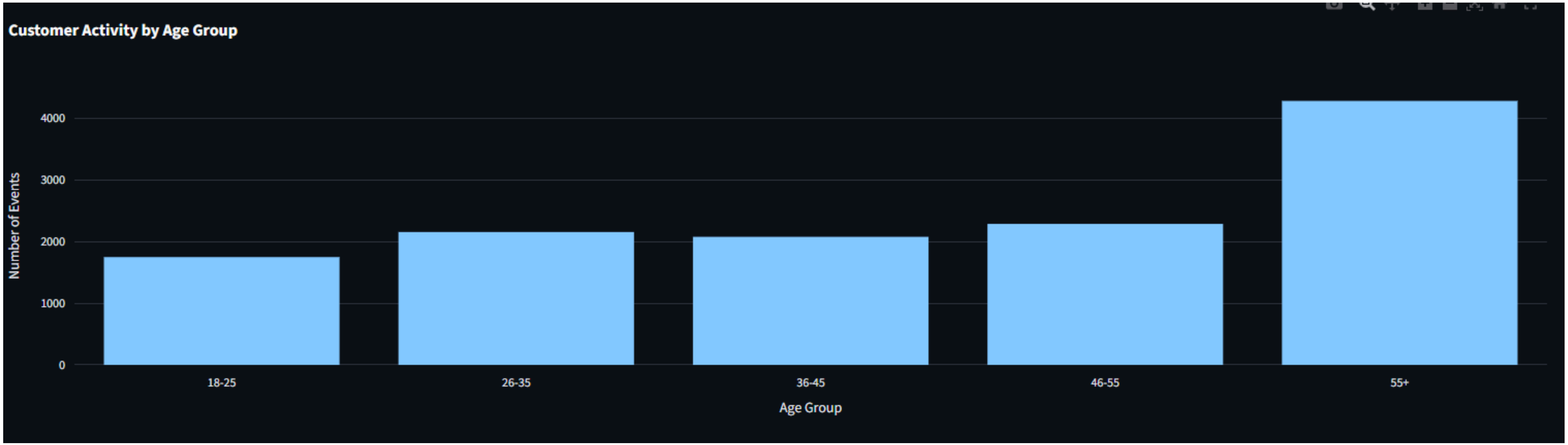
- SQL has strong ACID properties – standardized and simple.
- Livestreaming data to see observe time-wise changes
- Optimizes performance and usability with indexes and analytic SQL views.
- Makes the system flexible for future features and reliable for both real-time and historical analysis.

Key Findings & Visualizations

1. Customer Patterns:
2. High-value customers and age group analysis
- Cart-to-purchase rates and abandonment rates
3. Seasonal Trends:
- Peak activity hours/days/months
4. Product Performance:
- Most viewed/purchased products and revenue leaders
5. Abandoned Carts:
- Most abandoned products and total value


Frequency of 'add_to_cart' Events by Product
Frequency Count
200
150
100
50
0
Solid Gold Petite Micropave
Mens Casual Premium Slim Fit T-Shirts
Fjallraven - Foldsack No. 1 Backpack, Fits 15 Laptops
Mens Casual Slim Fit
Opna Women's Short Sleeve Moisture
John Hardy Women's Legends Naga Gold & Silver Dragon Station Chain Bracelet
Rain Jacket Women Windbreaker Striped Climbing Raincoats
Mens Cotton Jacket
Lock and Love Women's Removable Hooded Faux Leather Moto Biker Jacket
Pierced Owl Rose Gold Plated Stainless Steel Double
Acer SB220Q bi 21.5 inches Full HD (1920 x 1080) IPS Ultra-Thin
White Gold Plated Princess
WD 4TB Gaming Drive Works with Playstation 4 Portable External Hard Drive
DANVOUY Womens T Shirt Casual Cotton Short
SanDisk SSD PLUS 1TB Internal SSD - SATA III 6 Gb/s
MBJ Women's Solid Short Sleeve Boat Neck V
Samsung 49-Inch CHG90 144Hz Curved Gaming Monitor (LC49HG90DMNXZA) - Sur
Silicon Power 256GB SSD 3D NAND A55 SLC Cache Performance Boost SATA III 2.5
BIYLACLESEN Women's 3-in-1 Snowboard Jacket Winter Coats
WD 2TB Elements Portable External Hard Drive - USB 3.0
Product


Top 5 Products Purchased by Aaron Holt
Purchase Frequency
6
4
2
0
Rain Jacket Women Windbreaker Striped Climbing Raincoats
Fjallraven - Foldsack No. 1 Backpack, Fits 15 Laptops
Mens Casual Premium Slim Fit T-Shirts
Silicon Power 256GB SSD 3D NAND A55 SLC Cache Performance Boost SATA III 2.
Mens Casual Slim Fit
Product


Customer Activity by Age Group
4000
3000
2000
1000
0
Number of Events
18-25
26-35
36-45
46-55
55+
Age Group

thank
you