E-Commerce: Live Shopping Analytics

Kaylyn Nguyen
Rimsa Shrestha
Ji Yeon (Jenna) Woo
Mahnoor Shahid
Xiangwen (Sarah) Xiong

Business Problem

Domain: E-Commerce

Use Case: Live-stream shopping analytics for viewer engagement and conversion performance

Problem Statement: Brands lack real-time analytics to monitor live viewer engagement and detect drop-offs, limiting their ability to optimize sales dynamically.

Why It Matters: Live shopping is projected to drive 20%+ of e-commerce sales. Real-time insights, like RFM to identify high-value customers, enable brands to boost retention, increase conversions, and personalize offers on the fly.

Roadmap to Success

Week 1 (Jul 7–13)

- Event simulation
- Pub/Sub setup
- BigQuery ingestion

Week 2 (Jul 14–20)

- dbt transformations
- ML model training and deployment

Week 3 (Jul 21–26)

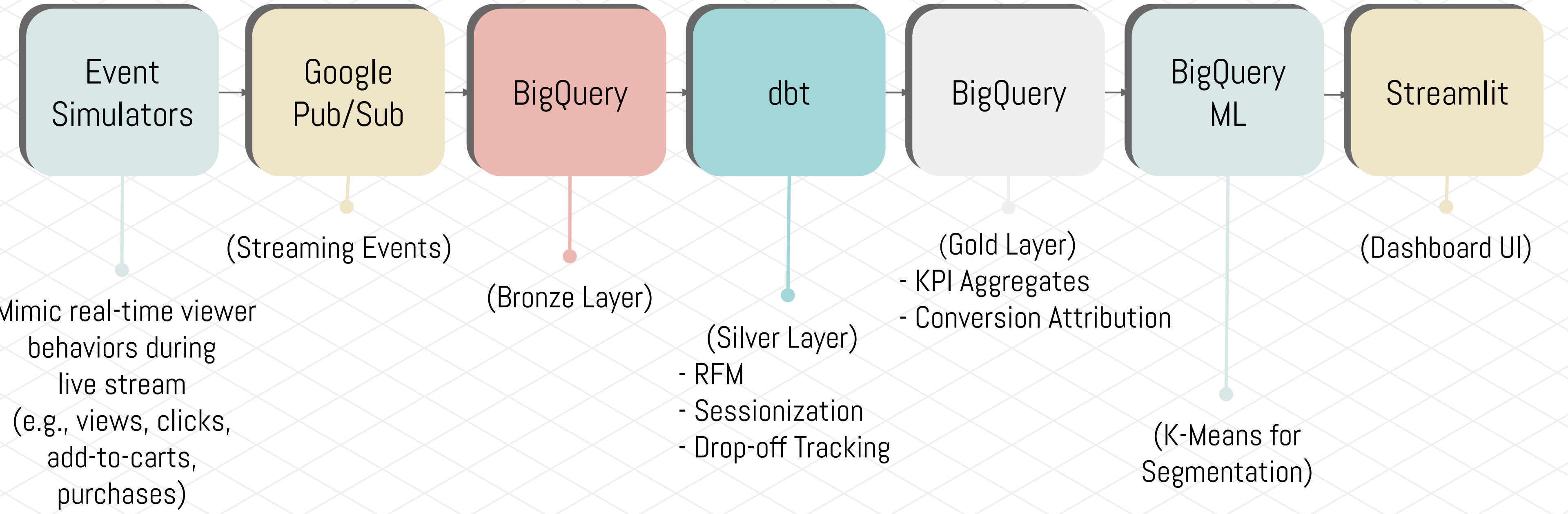
- Dashboard development

1

2

3

High-Level Architecture Diagram



Data Dictionary

Raw Tables (BigQuery)

- **fct_sessions:** Viewer session logs (duration, events, cart adds, orders, drop-off)
- **predicted_conversions:** ML scores per session (conversion probability, device, viewer state)
- **rfm_analysis:** RFM scores per viewer (recency, frequency, monetary)
- **train_customer_segments:** KMeans clustering model trained on RFM scores
- **customer_segments_scored:** Viewers labeled with predicted customer segment
- **customer_segments_named:** Final labeled segments with human-readable names

Transformed Tables (dbt + BigQuery ML)

- **fct_sessions:** raw sessions/events table
- **rfm_analysis:** calculates Recency, Frequency, Monetary scores
- **train_customer_segments:** applies k-means clustering via BigQuery ML
- **customer_segments_scored:** assigns each viewer to a cluster
- **customer_segments_named:** labels each cluster with human-readable segments (e.g. "High-Value Loyal")

Key Metrics

Viewer Count per Minute	# of active users every 60 seconds
Average Session Duration	Mean time spent per stream
Drop-off Rate	% of users who leave before mid-point
Click-to-Purchase Lag	Time between click & purchase
Engagement Score	ML-based metric that combines views, likes, comments, and clicks
RFM Scores	Recency, Frequency, Monetary value used for segmentation

Streamlit Dashboard

Displays live metrics:

- Top Sessions to Retarget based on engagement and churn risk
- Viewer Count Per Minute for tracking live traffic trends
- Customer Segmentation using behavioral and RFM insights

Interactive Features:

- Filters by Device Type and Viewer State
- Sorting Options for dynamic targeting and prioritization

Link to Dashboard:

https://live-shopping-app-wsrkvcs8kneneycowsbmvp.streamlit.app/

Streamlit Dashboard

🛍️ Session Conversion Predictions

Device Type

streamlitApp

All

Viewer State

All

Total Sessions
500

Unique Viewers
100

Avg. Conversion
82.86%

High-Intent Sessions
323

Drop-off Rate
29.0%

Avg. Session Duration
36.4 min

Click-to-Purchase Lag
18.9 min

Engagement Score
15.2

Average Conversion Probability
82.86%

	session_id	viewer_id	stream_id	device_type	viewer_state	conversion_probability
0	CUST097-1	CUST097	STRM1	tablet	FL	0.9888
1	CUST019-1	CUST019	STRM1	tablet	FL	0.9737
2	CUST027-1	CUST027	STRM2	desktop	IL	0.9731
3	CUST032-1	CUST032	STRM1	tablet	NY	0.9673
4	CUST037-1	CUST037	STRM1	desktop	FL	0.9652
5	CUST017-1	CUST017	STRM1	tablet	FL	0.9632
6	CUST033-1	CUST033	STRM1	tablet	GA	0.9624
7	CUST095-1	CUST095	STRM1	desktop	IL	0.95
8	CUST021-1	CUST021	STRM1	tablet	GA	0.9473
9	CUST100-1	CUST100	STRM1	tablet	GA	0.936

Business Action

Brands can:

- Identify at-risk viewers via churn prediction and send targeted offers
- Optimize live stream content based on drop-off and engagement data
- Segment customers for personalized marketing campaigns using RFM scores

Total Sessions 500	Unique Viewers 100	Avg. Conversion 82.86%	High-Intent Sessions 323
Drop-off Rate 29.0%	Avg. Session Duration 36.4 min	Click-to-Purchase Lag 18.9 min	Engagement Score 15.2
Average Conversion Probability 82.86%			

Business Outcome

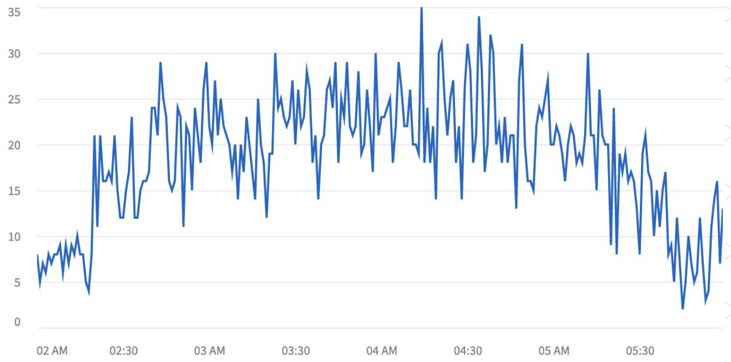
Real-time insights led to:

- Improved viewer retention during live streams
- Increased conversion rates via targeted interventions
- More efficient marketing spend through data-driven customer segmentation

🧠 Customer Segments

	viewer_id	recency	frequency	monetary	segment_name
13	CUST093	6	1	200	New & Promising
14	CUST015	6	1	0	Churned or Inactive
15	CUST045	6	1	200	New & Promising
16	CUST090	6	1	260	At Risk
17	CUST074	6	1	80	Churned or Inactive
18	CUST053	6	2	100	Low-Value Infrequent
19	CUST005	6	1	200	New & Promising

📊 Viewer Count Per Minute



Thank You