

Marketing Attribution Analysis Case Study

Executive Summary

I built a first- vs last-click attribution pipeline in BigQuery and a Looker Studio dashboard to show how model choice changes perceived campaign and channel performance. Using a 7-day lookback (pre-purchase) and excluding unattributed orders, MTD revenue appears equal under both models (~\$94.3K each), but credit shifts across campaigns/channels. Paid Search dominates overall; Display/Paid Social gain more under last-click. AOV remains stable (~\$158), indicating attribution affects *allocation of credit*, not customer spend.

Phase 1 - Understand & Prepare the Data (EDA + Cleaning Plan)

Source files

- **ad_events.csv** — user-level events
Fields: user_id, event_type (click|purchase), event_ts, campaign_id, campaign_name, channel, order_id, order_value
- **ad_spend_daily.csv** — daily media metrics
Fields: date, campaign_id, campaign_name, channel, spend, impressions, clicks

What to clean (principles applied)

ad_events (raw → clicks & purchases):

- Keep only click and purchase
- Trim/standardize text fields (normalize channel casing)
- Remove purchases with NULL/non-positive order_value or NULL order_id
- De-duplicate purchases on order_id (keep most recent / highest value)
- De-duplicate exact duplicate clicks (user_id + event_ts + campaign_id + channel)
- Keep timestamps as TIMESTAMP; flag (not drop) extreme order_value outliers

ad_spend_daily (raw → clean):

- Trim/standardize campaign and channel fields
- Replace NULL metrics with 0; clamp negatives to 0
- Roll up to one row per date × campaign_id × channel (sum spend/impressions/clicks)

Phase 2 - Data Preparation (BigQuery)

Datasets

- **Raw:** suvimarketingdataanalysis.marketing_data
- **Staging/Clean:** suvimarketingdataanalysis.marketing_data_stg

Staging tables created

- ad_events_purchases_clean — valid, deduped purchases with standardized dimensions
- ad_events_clicks_clean — valid, deduped clicks with standardized dimensions
- ad_spend_daily_clean — rolled-up media metrics by date × campaign × channel

Attribution table

- revenue_by_model_campaign_daily
Logic:
 - Join each purchase to eligible clicks within 7 days before the purchase timestamp; ignore post-purchase clicks
 - Exclude purchases with no eligible click (unattributed by design)
 - Assign revenue to:
 - First Click (earliest eligible click)
 - Last Click (latest eligible click)
 - Aggregate to date × campaign × channel × model with:
 - attributed_revenue
 - attributed_orders

Rollups (dashboard-ready)

- attr_rollup_by_campaign — totals by campaign × channel × model
- attr_rollup_by_channel — totals by channel × model

Phase 3 - Dashboard Implementation (Looker Studio)

Structure (3 pages)

1. Landing Page — Objective & methodology

- 7-day lookback (pre-purchase)
- Unattributed purchases excluded
- First vs Last click definitions

2. Revenue Trends & Campaign Performance

- MTD Revenue (scorecards): First vs Last (~\$94.3K each)
- Daily Attributed Revenue (Last 30 Days): two series by model
- Top Campaigns by Model: side-by-side bar charts

3. Channel Mix & Model Disagreement

- Channel Mix: revenue share by channel for each model
- Orders & AOV by Model: count and AOV (\approx \$158)
- Model Disagreement: table showing Revenue Delta (Last – First) with conditional coloring (green = closer, red = introducer)

Key Findings (Marketing Interpretation)

- **Model choice changes the story.**
MTD totals are similar across models (as expected), but *who* gets credit shifts.
- **Campaign roles diverge.**
 - Brand Search gains under first-click → strong entry/awareness driver.
 - Awareness Blast and parts of Paid Social gain under last-click → strong closers/remarketing.
- **Channel mix is stable but reweighted.**
Paid Search remains the largest contributor overall; Paid Social/Display pick up more credit under last-click.
- **Customer spend is steady.**
AOV \approx \$158 under both models → attribution affects allocation, not basket size.
- **Attribution coverage.**
~47% of purchases attributed within the 7-day window; ~53% unattributed (organic/other/missing touchpoints). This is typical and should be communicated to stakeholders.

Business Implications

- **Budgeting:** Avoid over-relying on last-click; you risk under-funding awareness drivers (e.g., Brand Search) that fill the funnel.
- **Planning:** Treat Display/Paid Social as closers in retargeting flighting; keep Brand Search funded as a consistent entry lever.
- **Measurement:** Consider blended (e.g., 50/50) or multi-touch models for a balanced view; retain first vs last comparison in monthly reviews.

Quality Checks (performed)

- **Spec compliance:** 7-day pre-purchase lookback; post-purchase clicks ignored; unattributed purchases excluded.
- **Conservation:** Total attributed revenue and order counts match across models (only allocation differs).
- **Deduping:** No duplicate order_id in clean purchases; no exact-duplicate clicks after staging.
- **Media sanity:** No negative spend; clicks \leq impressions; one row per date \times campaign \times channel.

Deliverables Summary

- **BigQuery (SQL implementation):**
 - Clean staging tables in marketing_data_stg
 - Attribution table revenue_by_model_campaign_daily (first vs last)
 - Rollups by campaign/channel for dashboarding
- **Looker Studio dashboard:**
 - MTD model comparison, 30-day trends, top campaigns (side-by-side), model disagreement (delta with conditional colors), channel mix, orders & AOV
- **Insights:**
 - First- vs last-click materially reweights campaign/channel credit; AOV stable; clear guidance for budget allocation and flighting.