Funnel Analysis on Google Analytics Sample Data

🧠 Project Overview

This project uses Google Analytics session data from BigQuery's public dataset `google_analytics_sample,` to perform a **funnel analysis** across key ecommerce steps:

Landing Page → Product Page → Cart → Checkout → Thank You

With over 10,000 sessions, this dataset offers a rich opportunity to identify **conversion drop-offs**, assess **user behavior by segment**, and suggest **UX and technical improvements** based on actual funnel performance.

Defining the Funnel

The raw GA data doesn't come with ready-made funnel steps, so I had to build them manually using the `pagePath` column. I tried two approaches:

1. Manual Keyword Mapping

I initially built a dictionary mapping funnel steps to specific `pagePath` patterns (e.g., `Cart` → `/basket.html`, `Thank You` → `/ordercompleted.html`).
However, this approach lost **over half of the sessions** due to overly narrow matching.

2. Smarter Path Classification

To fix this, I implemented a classifier function using regex and keyword heuristics.
It labeled each `pagePath` into funnel steps, then aggregated at the session level.
This method captured **significantly more sessions** and improved funnel fidelity.

Two Types of Funnels: Standard vs Fast-Track

While analyzing, I noticed **114 sessions went straight to Checkout**, skipping Cart entirely — likely due to quick-buy buttons or direct checkout links. So I split the funnel into two paths:

Step	Fast-Track Sessions	Standard Sessions
Landing Page	49	1574
Product Page	30	1086
Cart	0	252
Checkout	114	154
Thank You	0	51

🔍 Insights:

Fast-tracked sessions had **0 completions**, suggesting:

- Users dropped off mid-checkout
- Something broke in the quick checkout flow
  Or possibly bot/misclassified sessions

👉 **Recommendation**: Investigate fast-track checkout UX and session origins. Consider bot filtering.

📉 Standard Funnel Results

From 1574 standard sessions:

- **1086** reached a Product Page
- **252** reached the Cart (76% drop-off)
- **51** reached the Thank You page

That's a **3.2% overall completion rate**.

👉 **Suggestions**:

- Improve Cart transition UX (e.g., clearer CTAs, trust badges)
- Simplify checkout forms & reduce steps
- Run A/B tests on Product Page layout and CTA visibility

📱 Device-Level Funnel Drop-Off

Segmenting by `deviceCategory` revealed huge behavioral differences:

Step	Desktop	Mobile	Tablet
Landing	1176	397	50
Product	804	280	32
Cart	207	41	4
Checkout	225	39	4
Thank You	49	2	0

🔍 Insights:

- Mobile conversion = 0.5%, vs Desktop = 4.2%
- Tablet = 0% conversion

👉 **Recommendations**:

- Prioritize mobile UX and error tracking
  Consider deprioritizing tablets unless traffic increases

🌍 Funnel Drop-Offs by Country

Using country-level segmentation, I found that:

Country	Step	Drop-Off Count	Drop-Off Rate
🇺🇸 USA	Product → Cart	530	74.9%
🇩🇪 Germany	Product → Cart	17	89.5%
🇮🇳 India	Product → Cart	50	84.7%
🇬🇧 UK	Product → Cart	27	84.4%

🔍 Insights:

- While the US has the **highest drop-off volume**, international users like Germany, India, and the UK show **higher drop-off rates**, pointing to potential **trust or localization issues**.

👉 **Recommendations**:

- Run localized A/B tests on product + cart flow
  Add country-specific shipping info and currency displays

🌐 Drop-Offs by Browser

I categorized 15+ browsers into High and Low volume groups.
Among High-volume browsers:

- **Chrome:**
  - Largest drop-off volume across the funnel
  - 654 users dropped from Product → Cart
  - 191 dropped at Checkout → Thank You
- **Safari & Firefox:**
  - Drop-off rates > 80% at both Cart and Checkout stages
- **Internet Explorer:**
  - 100% abandonment at Checkout
  - Likely poor compatibility or broken UX

Common Drop-Off Issues:

1. Tracking gaps (negative drop-offs → session stitching issues)
2. Checkout flow inconsistencies
3. Cart not tracked properly

👉 **Recommendations**:

- QA checkout pages in Firefox, Safari, and IE
- Improve session stitching and tag firing reliability

Final Takeaways

✅ Mapped a 5-step ecommerce funnel across 10K+ GA sessions
✅ Identified drop-offs by segment: **device, country, browser**
✅ Created actionable insights for product, UX, and data teams
✅ Built two funnel models: Standard and Fast-Track
✅ Used BigQuery + Colab for data processing and Plotly for visualization

---

Appendix: SQL Data Extraction

```
SELECT
  fullVisitorId,
  visitId,
  CONCAT(CAST(fullVisitorId AS STRING), '_', CAST(visitId AS STRING)) AS
session_id,
  hits.page.pagePath AS pagePath,
  device.browser,
  device.deviceCategory,
  geoNetwork.country
FROM `bigquery-public-data.google_analytics_sample.ga_sessions_20170801`,
UNNEST(hits) AS hits
WHERE hits.type = 'PAGE'
```