DATA ANALYTICS REPORT

EXPLORATORY DATA ANALYSIS OF GOOGLE MERCHANDISE STORE USER BEHAVIOR & LIFETIME VALUE

User Behaviour, Purchase Patterns & Lifetime Value

TOOLS USED

Python, Pandas, Matplotlib, Seaborn & Plotly

PRESENTED BY

Tejas Jadhav

GITHUB

TABLE OF CONTENTS

PROBLEM STATEMENT & OBJECTIVE

Problem Statement:

The Google Merchandise Store serves a diverse user base with varying browsing, purchase, and spending behaviors, making it challenging to identify the key drivers of conversion efficiency and long-term customer value.

Objective:

To explore the Google Merchandise Store e-commerce dataset to uncover patterns, insights, and trends that can help businesses make data-driven decisions around user behaviour, conversion performance and revenue generation.

EDA WORKFLOW

For this analysis, a structured workflow was followed, involving data collection, understanding, cleaning, exploration, and summarization of insights to allow a clear understanding of the dataset and its trends.

01 Data Collection

- Gathered the Google Merchandise Dataset from Kaggle

02 Data Understanding & Anomaly Detection

- Looked at data distributions
- Found missing values, outliers and unusual patterns

03 Data Cleaning & Treatment

- Fixed missing or incorrect values
- Standardized formats for consistency

04 Exploratory Analysis

- Univariate Analysis: e.g Device Distribution, Top 10 Categories, Brand Distribution etc
- **Bivariate:** Price vs LTV, Event Type vs Device etc
- Multivariate Analysis: Device vs Category vs Average LTV

05 Insights & Reporting

- Summarized patterns and trends
- Highlighted key findings for developers and businesses

KEY QUESTIONS EXPLORED

- How do user location and device type influence browsing behavior, purchase activity, and lifetime value?
- Which stages of the shopping funnel (add to cart, checkout, purchase) contribute most to conversion drop-offs?
- Do one-time purchasers represent a larger growth opportunity compared to repeat buyers in terms of lifetime value expansion?
- How strongly does repeat session behavior influence purchase frequency and overall customer lifetime value?
- What role do product pricing and category mix play in conversion rates and revenue concentration?
- How does checkout friction impact final purchase completion and cart abandonment?
- Which product categories and brands are most effective at driving sustained revenue and repeat purchases?
- To what extent does purchase frequency versus order size influence long-term customer value?

DATA OVERVIEW

The dataset provides insights into SaaS customer behavior, covering demographics, subscription characteristics, product usage, billing activity & support interactions. It also captures customer engagement signals, satisfaction metrics, and feedback indicators that are relevant for understanding retention and churn dynamics.

Data Source: Kaggle

Dataset Size

7,58,884

Records

14

Features

Purchase Diversity

21

Product Categories

109

Countries

5

Brands

DATA OVERVIEW

Below is a detailed description of the feature set:

Dataset Features	Type	Feature Description
user_id	Numerical (Discrete)	Unique identifier assigned to each user in the store
ga_session_id	Numerical (Discrete)	Unique identifier representing a single user session (visit)
country	Categorical	Country from which the user accessed the Google Merchandise Store
device	Categorical	Device type used during the session (e.g., desktop, mobile, tablet)
type	Categorical	Event type performed by the user (e.g., add_to_cart, begin_checkout, purchase)
item_id	Numerical (Discrete)	Unique identifier of the product involved in the event
date_x	Date / Time	Timestamp of the event interaction (e.g., cart addition or purchase)
name	Categorical	Type of subscription contract (Monthly, Quarterly, Annual)
brand	Categorical	Name of the product involved in the interaction
variant	Categorical	Specific product variation such as size or color (may contain missing values)
category	Categorical	Product category (e.g Apparel, Bags, Fun, New)
price_in_usd	Numerical (Discrete)	Price of the product in U.S. dollars
ltv	Numerical (Discrete)	Total lifetime revenue generated by the user
date_y	Date / Time	Timestamp associated with the user lifecycle (e.g., first interaction date)

DATA QUALITY CHALLENGES & ANOMALIES

Few inconsistencies were found in the dataset, which could have affected the analysis if left unaddressed.

DATA ANOMALIES

- Several fields (country, device, type, name, brand, category) are stored as strings and should be converted to categorical types, while date_x and date_y should be converted to datetime for time-based analysis.
- The variant column has a very high proportion of missing values (~83.84%), making reliable imputation impractical and limiting its analytical usefulness.
- The country column has minimal missing data (~0.6%).

DATA CLEANING & TREATMENT

Inconsistent and missing values were addressed, and key features were cleaned and standardized for analysis.

DATA CLEANING SUMMARY

- Categorical fields such as country, device, type, name, brand, and category were converted to appropriate categorical data types, while date_x and date_y were standardized to datetime format to support efficient and accurate analysis.
- Due to the extremely high proportion of missing values in the variant column (~83.84%), the feature was excluded from imputation and used only where non-null values were analytically meaningful.
- The country column contains only a small fraction of missing values (~0.6%), which were safely filled with "Unknown" to preserve data completeness without introducing analytical bias.

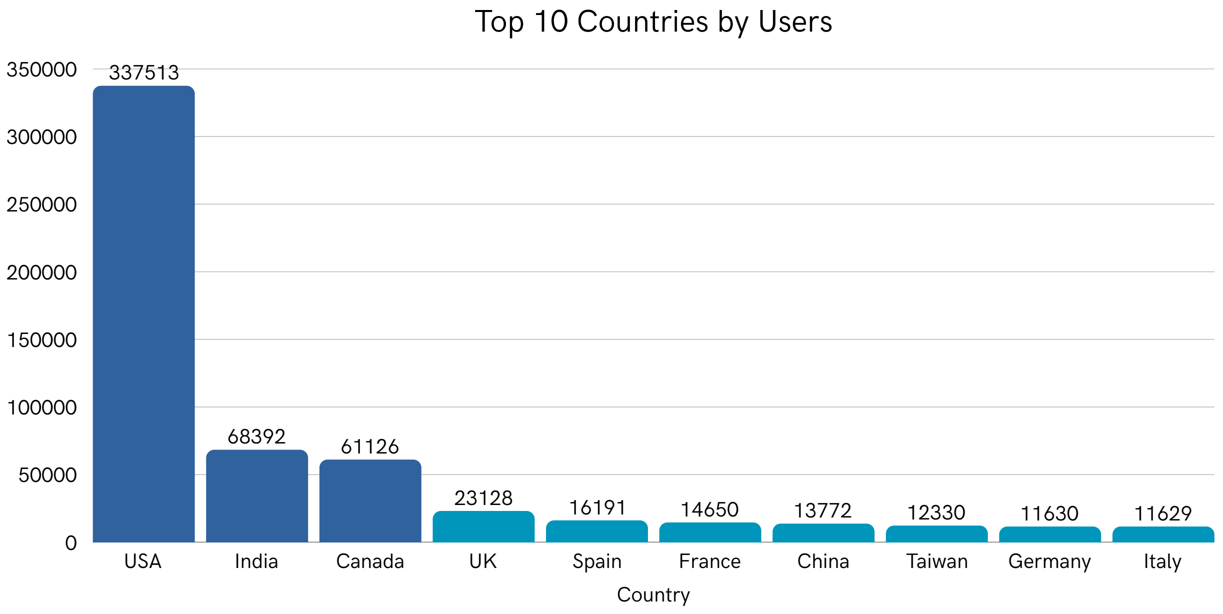
INSIGHTS

Customer Profile & Engagement

User traffic is overwhelmingly concentrated in the US, with international markets contributing a long tail of engagement

44.47%

User Traffic is from the United States



Key observations

- The United States accounts for the majority of users by a large margin, far exceeding all other countries in the dataset.
- After the top few countries (US, India, Canada), user counts drop sharply, indicating a highly skewed country-level distribution.

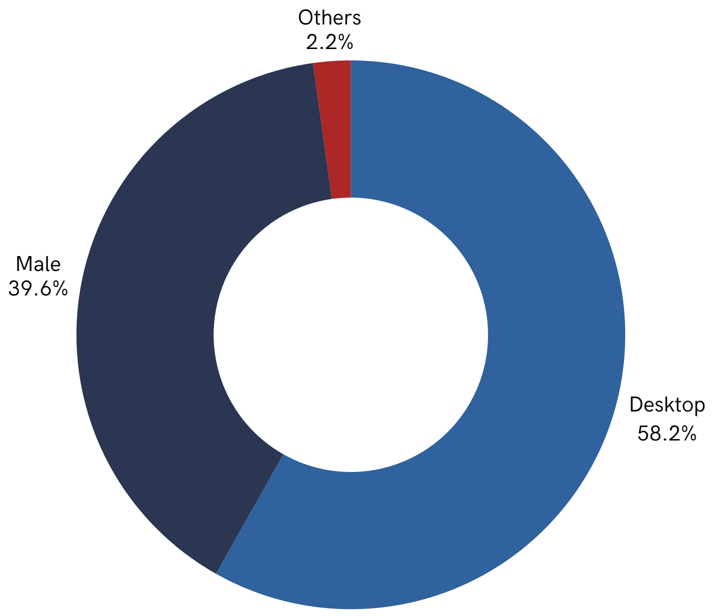
Business Insights

- Heavy reliance on the US market makes overall traffic and revenue vulnerable to shifts in a single geography, increasing concentration risk.
- Secondary markets like India and Canada present the strongest opportunities for international growth through targeted localization and expansion efforts.

Desktop dominates user sessions, while mobile represents a substantial secondary access channel

58.20%

of User Sessions come from Desktops



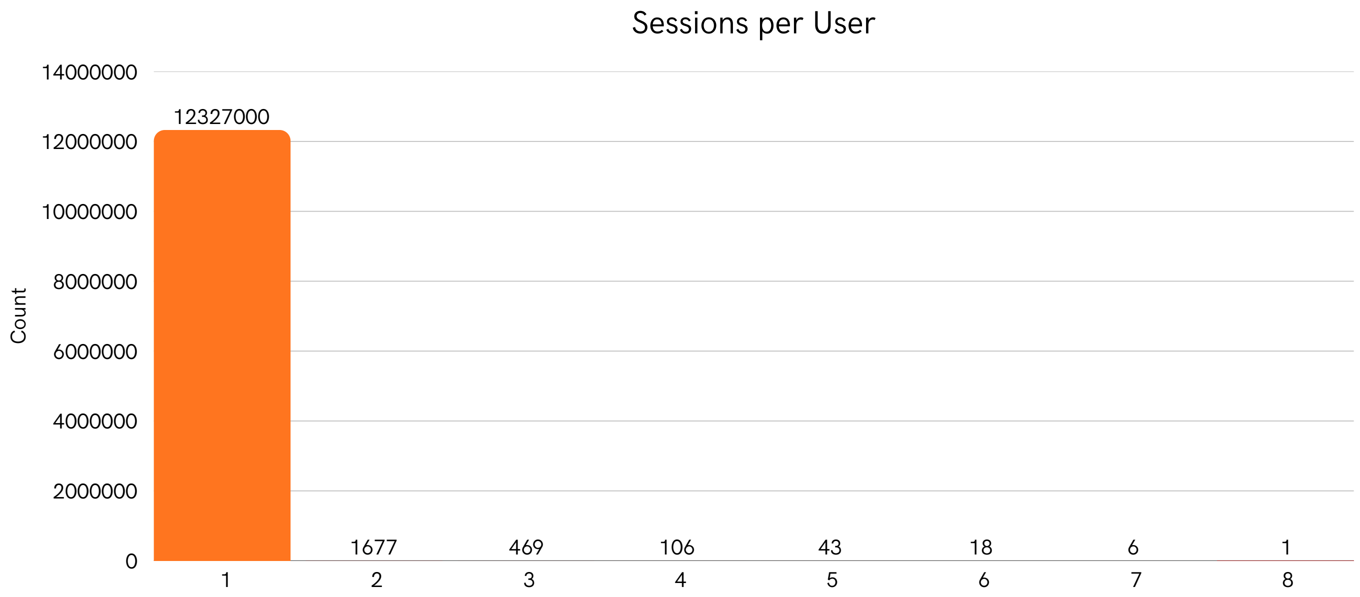
Key observations

- Desktop accounts for the majority of user sessions, contributing over half of total traffic in the dataset.
- Mobile usage is significant but clearly trails desktop, while tablet usage remains negligible.

Business Insights

- The store’s experience and conversion performance are likely most sensitive to desktop UX, making it a critical optimization priority.
- The sizable mobile user base presents an opportunity to improve mobile-specific usability and conversion flows to unlock incremental revenue.

Most users interact with the store only once, with repeat engagement limited to a small cohort



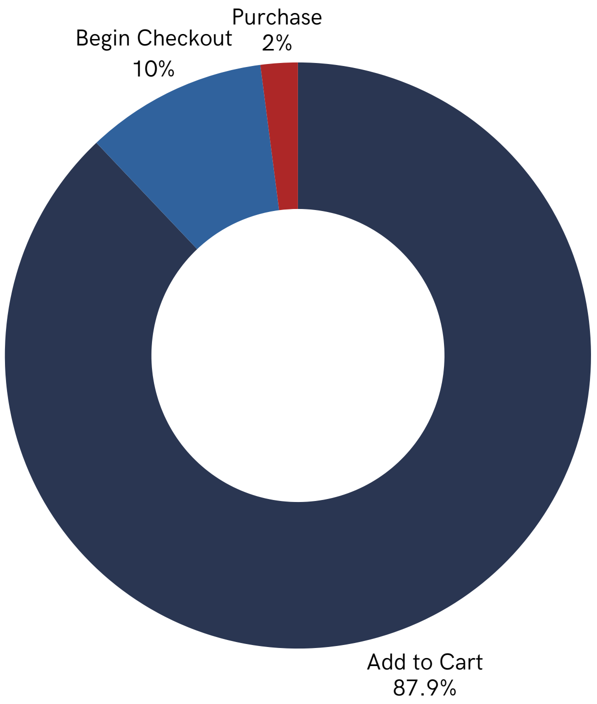
Key observations

- The majority of users have a single session, while user counts decline sharply as the number of sessions increases.
- Only a very small fraction of users return for multiple visits, indicating a highly right-skewed engagement distribution.

Business Insights

- Low repeat visitation suggests that most users do not form sustained engagement, limiting long-term lifetime value growth.
- Improving post-visit retention through remarketing, personalized recommendations, or follow-up communication could convert one-time visitors into repeat users.

A steep funnel drop-off occurs between cart addition and final purchase



Key observations

- Add-to-cart events dominate user activity, while only a small fraction of sessions progress to checkout and an even smaller share convert to purchases.
- The sharp decline from begin_checkout to purchase highlights a significant loss of users late in the shopping journey.

Business Insights

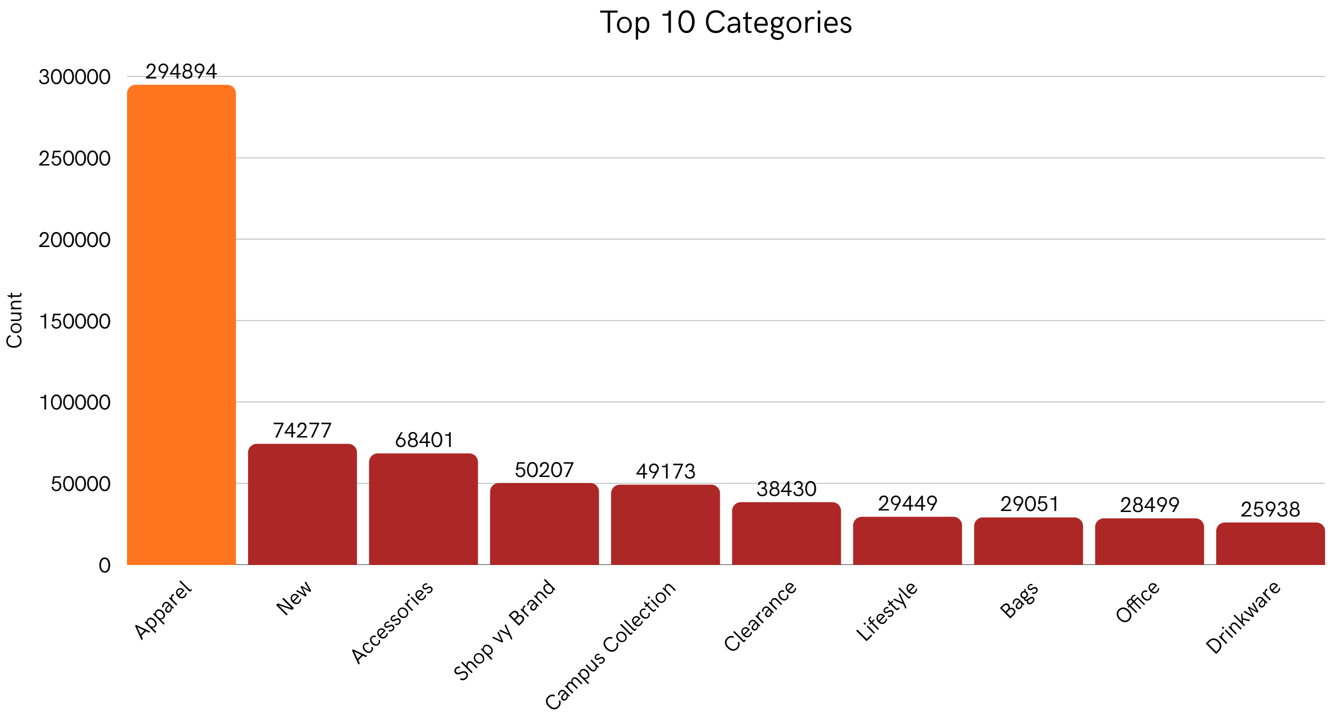
- Friction in the checkout or payment stages is likely suppressing conversions, making checkout optimization a high-impact opportunity.
- Reducing late-stage drop-offs through clearer pricing, simplified checkout flows, or trust signals could materially improve purchase conversion rates.

Product Performance & Preferences

Apparel overwhelmingly drives user interactions, with other categories contributing a fragmented long tail

38.85%

Of the Interactions is driven by Apparel



Key observations

- Apparel dominates category-level interactions by a wide margin, far surpassing all other product categories.
- The remaining categories show a steep drop in engagement, with relatively similar and much lower interaction volumes.

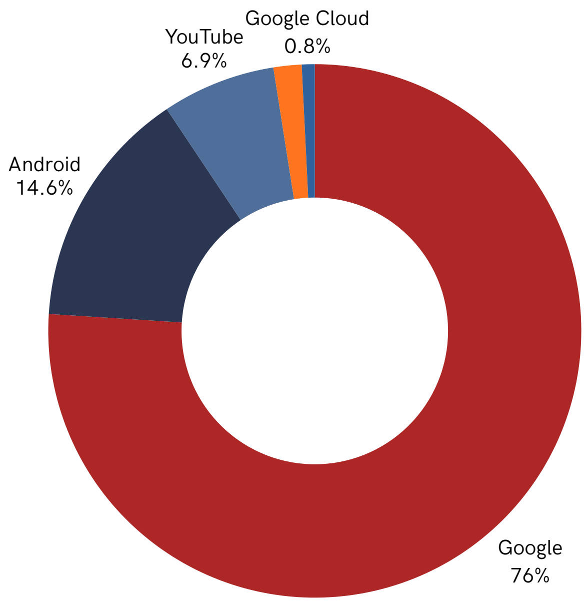
Business Insights

- The store’s performance is heavily reliant on Apparel, making merchandising, pricing, and availability in this category critical to overall success.
- Diversifying engagement through cross-selling and bundling from Apparel into lower-performing categories could help broaden revenue contribution and reduce category dependence.

User interactions are heavily concentrated on Google branded products, with limited traction for other brands

76%

Of Users search for Google related products



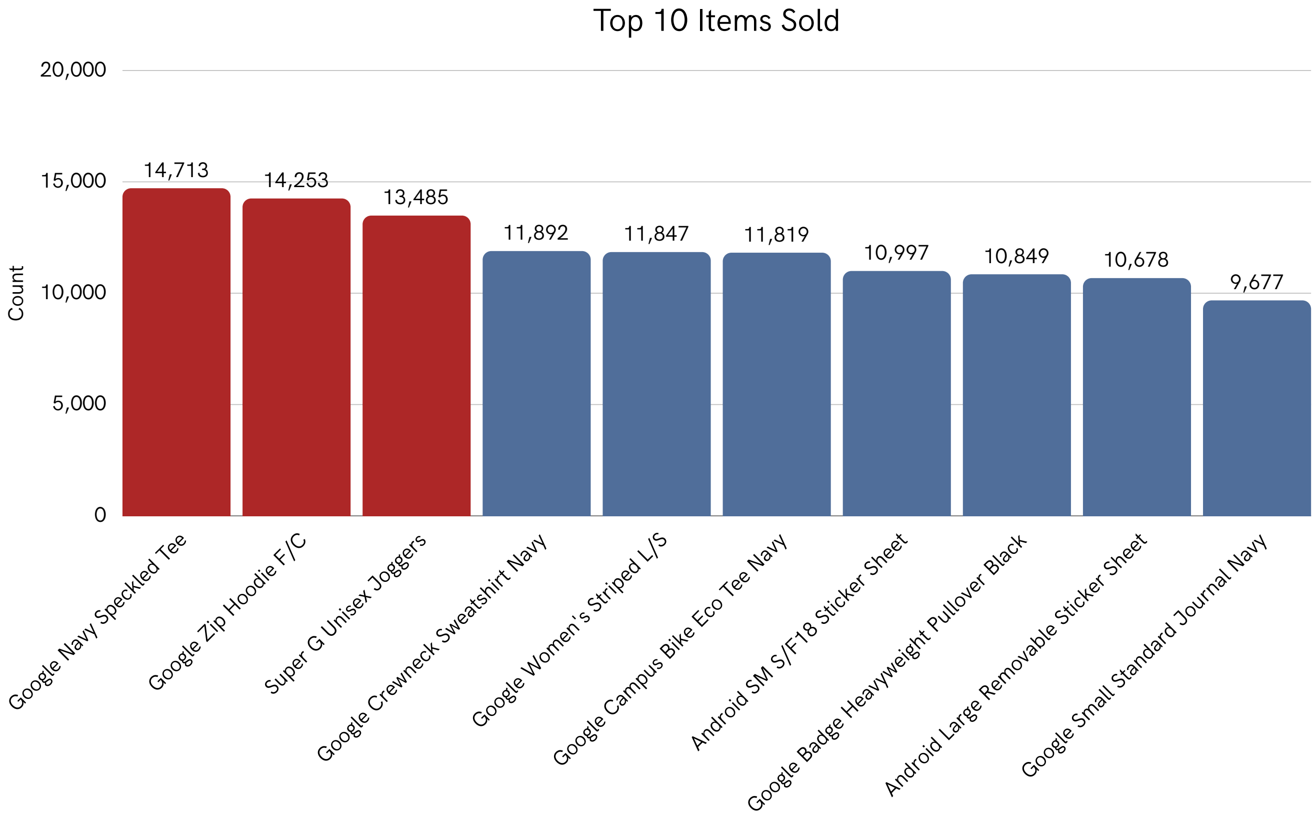
Key observations

- Google-branded products account for the vast majority of interactions, dwarfing engagement with Android, YouTube, and other brands.
- Non-Google brands form a small long tail, each contributing only a minor share of overall activity.

Business Insights

- Performance of the store is strongly tied to the Google brand, making brand-led demand a primary driver of engagement and sales.
- Strengthening visibility, positioning, or bundling of non-Google brands could help diversify demand and reduce reliance on a single dominant brand.

A small set of core merchandise items drives a significant share of total sales



Key observations

- The top-selling items significantly outperform the rest, with relatively similar high sales volumes across the top 10 products.
- Bestsellers are dominated by apparel and low-priced branded merchandise, indicating consistent demand for staple products.

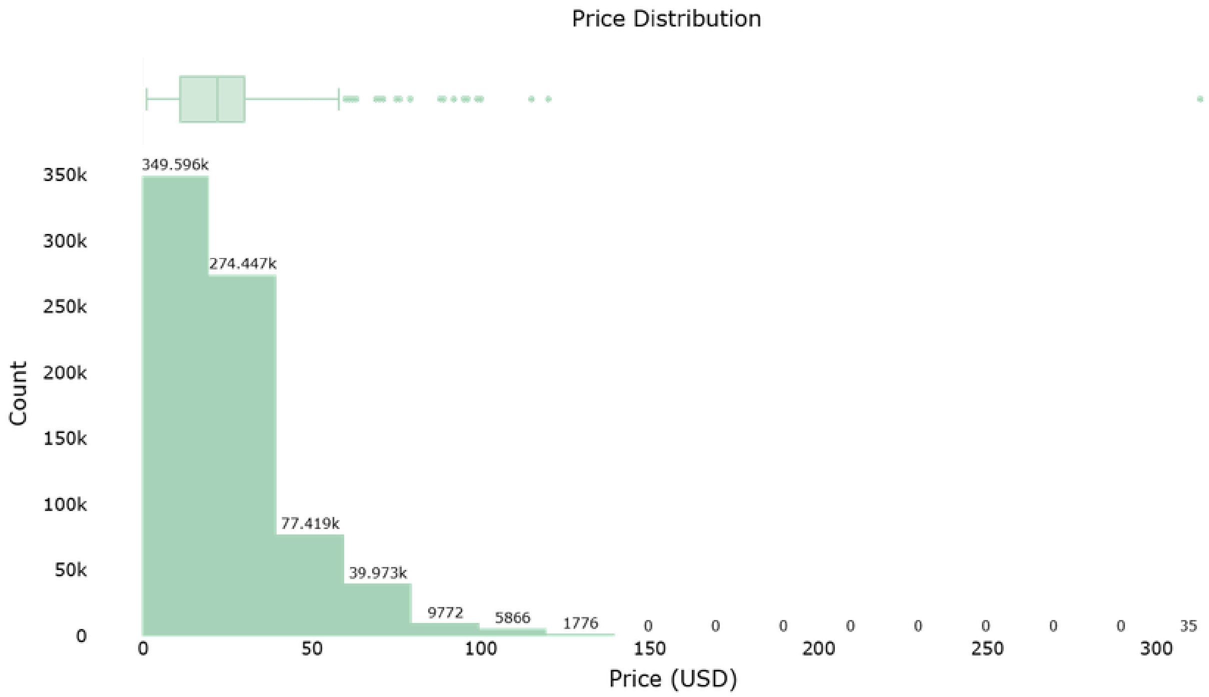
Business Insights

- Ensuring high availability, optimal pricing, and prominent placement of these core items is critical to sustaining overall sales performance.
- These top sellers present strong opportunities for bundling, cross-selling, and upsell strategies to increase average order value.

Product interactions are concentrated at lower price points, with premium items forming a small high-value tail

$22

Typical Price (Median)



Key observations

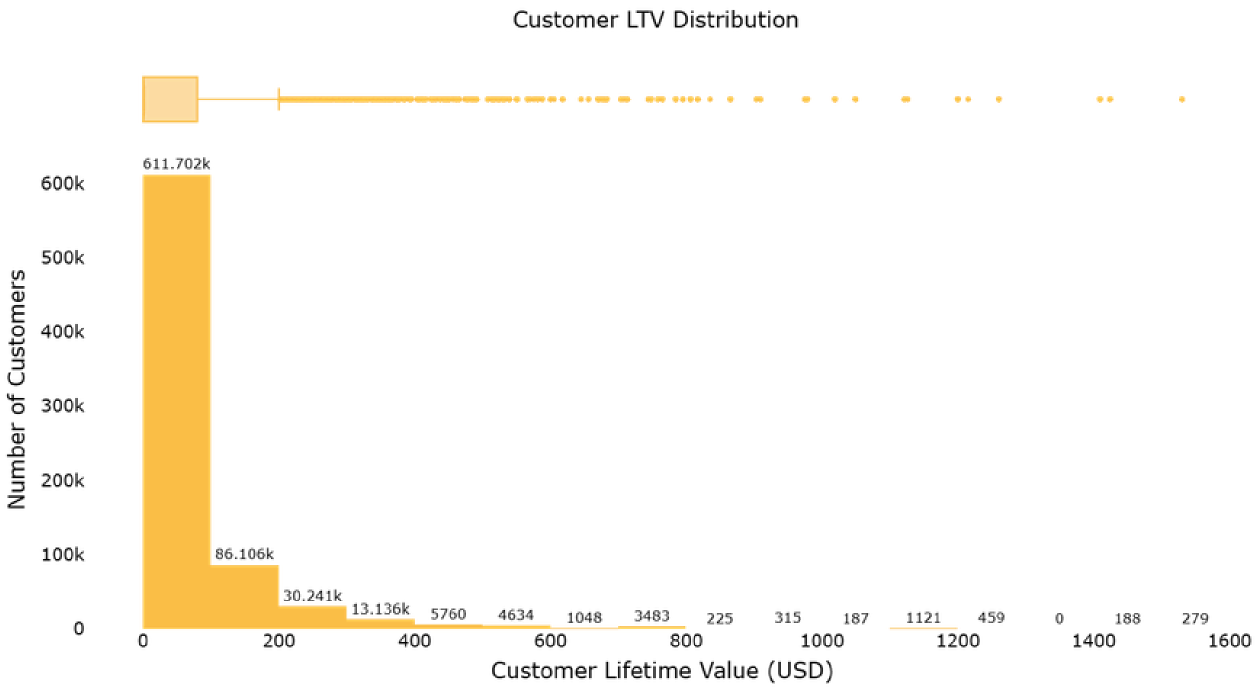
- The majority of product interactions occur at lower price ranges, with counts declining rapidly as prices increase.
- A small number of high-priced items appear as outliers, indicating limited but present demand for premium products.

Business Insights

- Revenue volume is likely driven by affordable, high-frequency products, making pricing and availability in lower tiers critical for scale.
- Premium products, while niche, may offer opportunities for margin expansion through targeted positioning rather than mass promotion.

Revenue & Customer Value

Customer lifetime value is highly concentrated, with a small cohort driving the majority of revenue



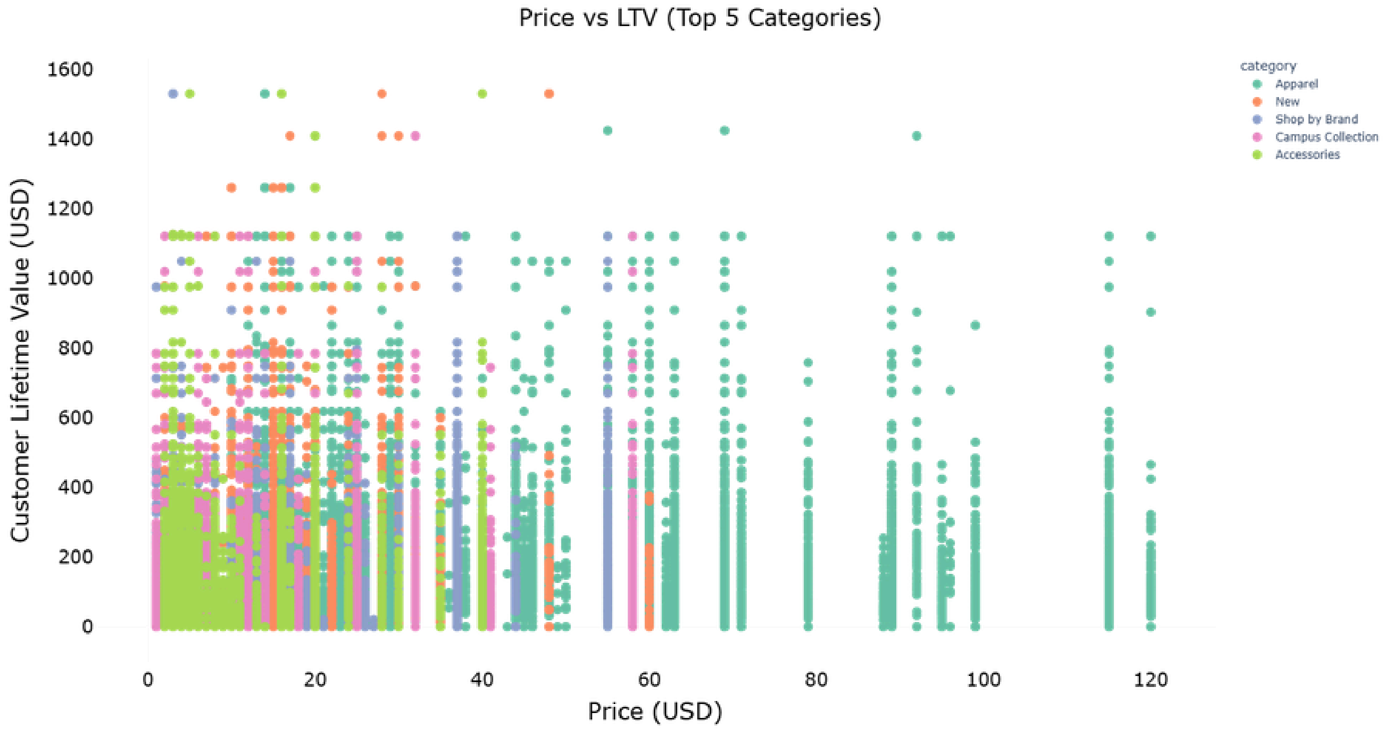
Key observations

- Most customers generate low lifetime value, with a sharp drop-off as LTV increases.
- A long right tail indicates a small group of customers with exceptionally high lifetime value.

Business Insights

- Overall revenue is disproportionately driven by a minority of high-value customers, making their retention critical to business performance.
- Understanding and replicating the behaviors of these high-LTV customers can inform targeted retention, loyalty, and upsell strategies to maximize long-term value.

Higher-priced items do not guarantee higher lifetime value, with repeat engagement driving LTV across categories



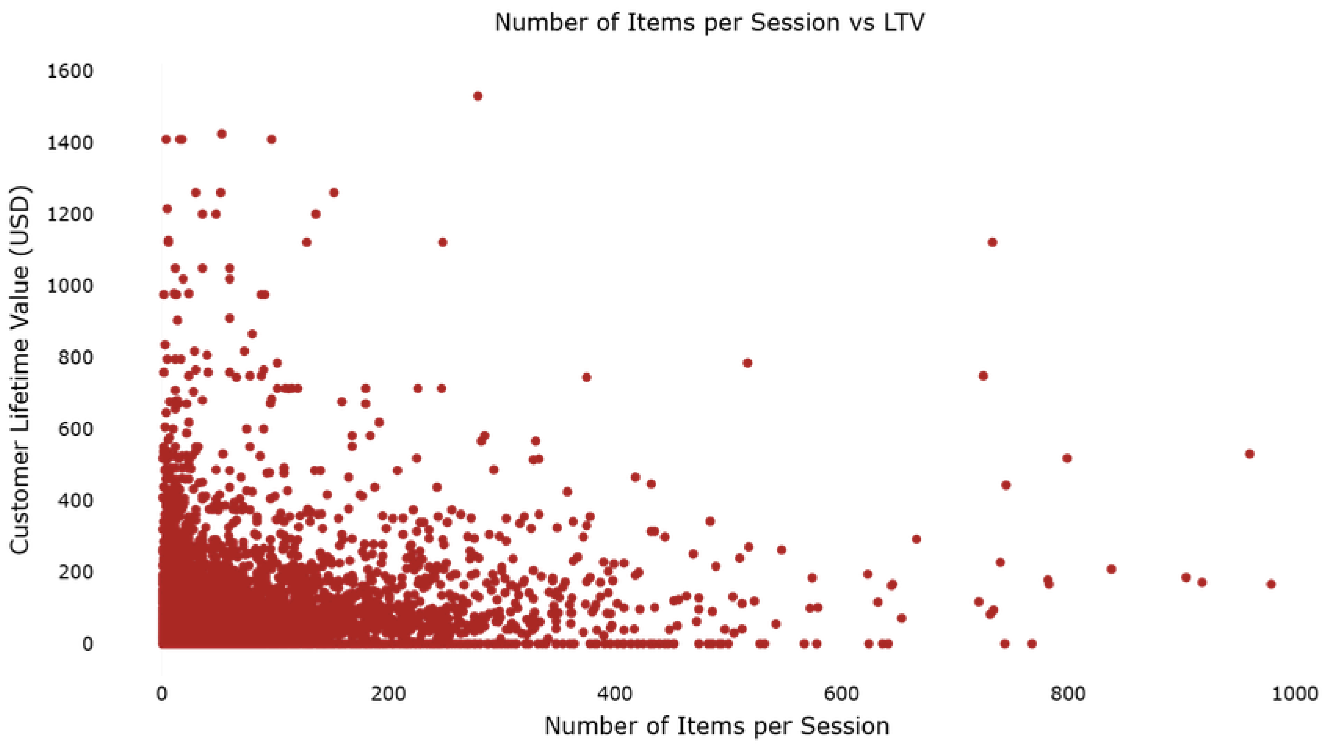
Key observations

- There is no clear linear relationship between product price and customer lifetime value across the top categories.
- High-LTV customers appear across a wide range of price points, including low- and mid-priced items.

Business Insights

- Customer lifetime value is driven more by repeat purchasing behavior than by one-time purchases of expensive products.
- Focusing on retention and repeatability of popular lower-priced items may yield greater LTV gains than emphasizing premium pricing alone.

Higher Number of Items per Session does not guarantee increased Lifetime Value



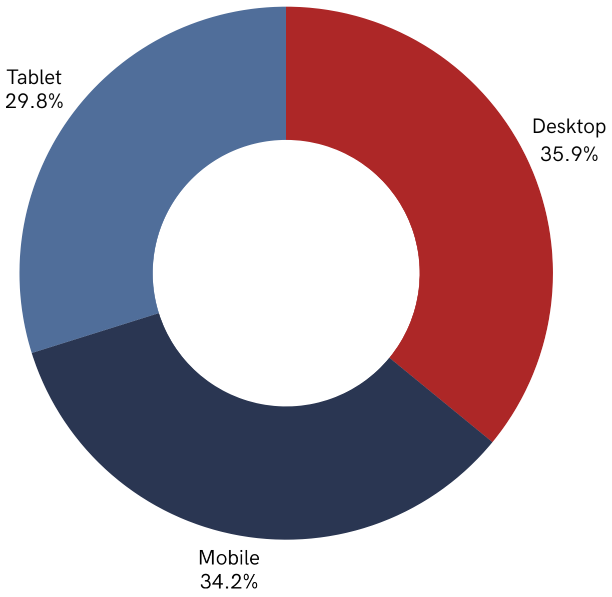
Key observations

- Most sessions with a high number of items show low or zero LTV, indicating bulk activity doesn't always translate to long-term revenue.
- The majority of users with higher LTV tend to have moderate item counts per session rather than extreme quantities.

Business Insights

- Focus should be on the quality and repeat engagement of customers rather than just boosting the quantity of items per session.
- Strategies targeting user retention and repeat purchases may yield higher LTV than encouraging one-time large cart sizes.

Desktop Users deliver the Highest Lifetime Value compared to Mobile and Tablet



Key observations

- Desktop users have the highest average LTV, followed closely by mobile users, with tablet users trailing behind.
- The gap between desktop and tablet LTV suggests device type influences customer spending behavior and engagement.

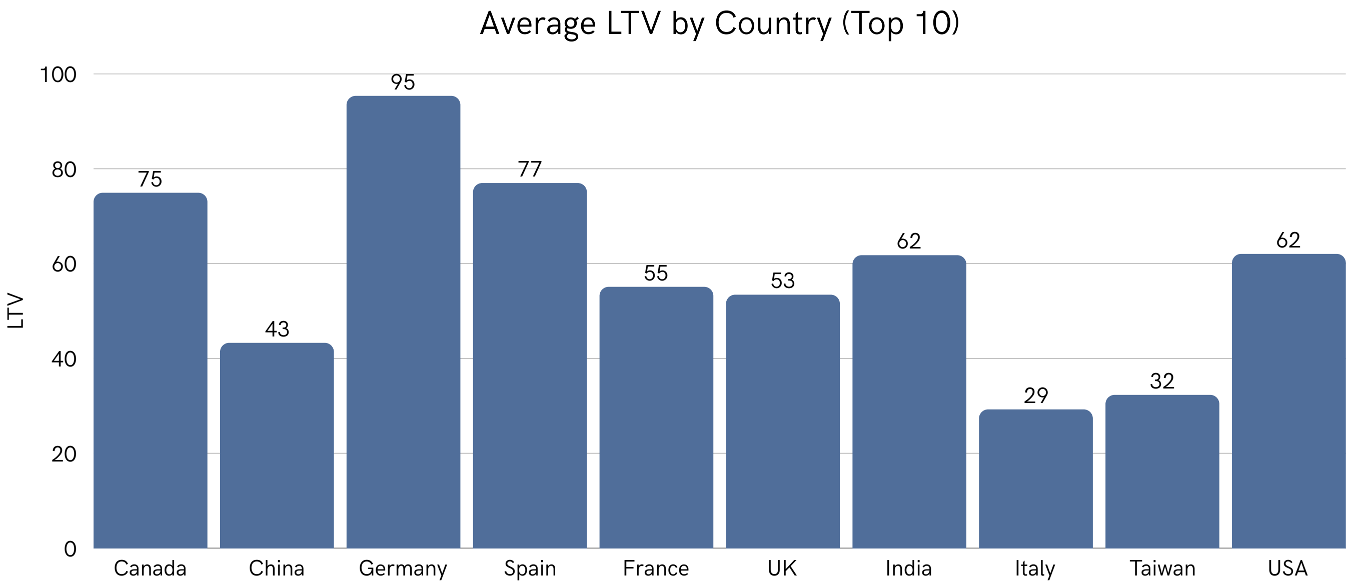
Business Insights

- Prioritize optimizing the desktop shopping experience to capitalize on the higher-value customer segment.
- Consider tailored marketing and UX improvements for tablet users to boost their LTV and reduce the performance gap.

Germany Leads in Customer Lifetime Value among Top countries, while Italy and Taiwan lag behind

61.42

Typical LTV (Average)



Key observations

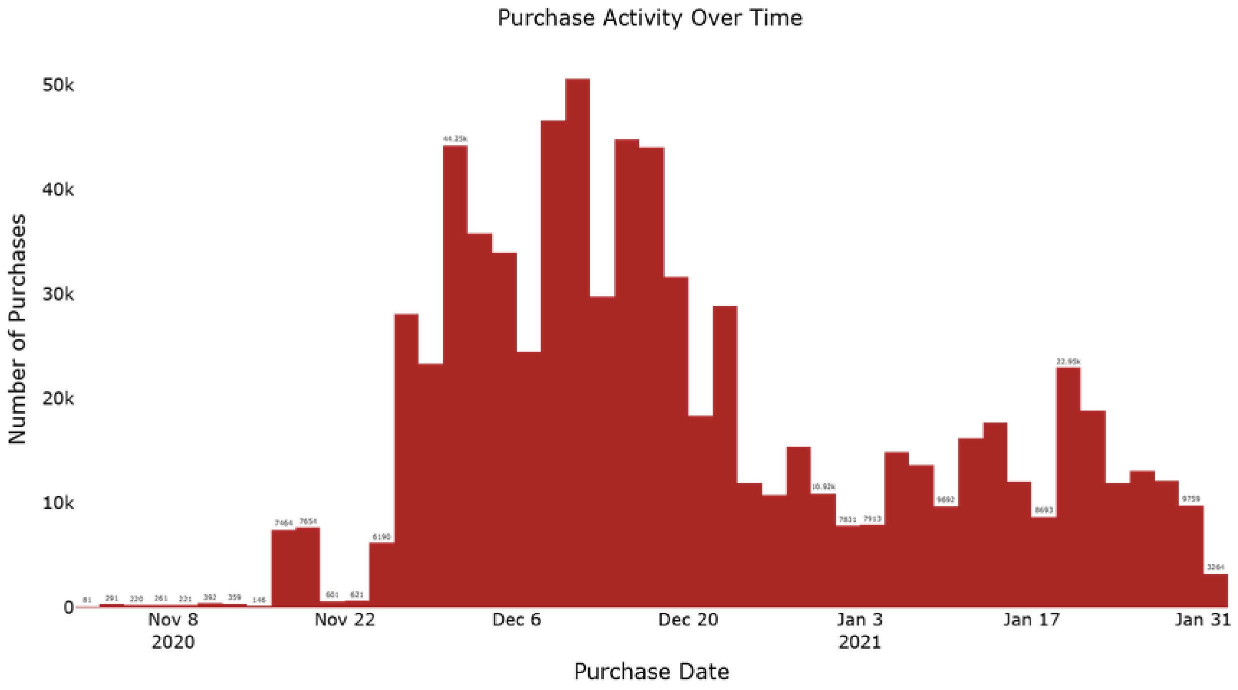
- Germany (DE) shows the highest average LTV, significantly outperforming other major markets like Canada (CA) and the US (US).
- Italy (IT) and Taiwan (TW) record the lowest LTV among the top 10 countries, indicating potential regional challenges or lower customer engagement.

Business Insights

- Invest in marketing and localized strategies in high-LTV countries like Germany to maximize revenue potential.
- Explore tailored initiatives to understand and improve customer retention and spending in lower-performing markets such as Italy and Taiwan.

Temporal Trends & Conversion Funnel

Holiday Season Drives Peak Purchases, Followed by Post-Season Slowdown



Key observations

- Purchase activity rises sharply from late November and peaks in early-to-mid December, indicating strong holiday-driven demand.
- After mid-December, purchase volume declines steadily into January, stabilizing at significantly lower levels compared to the December peak.

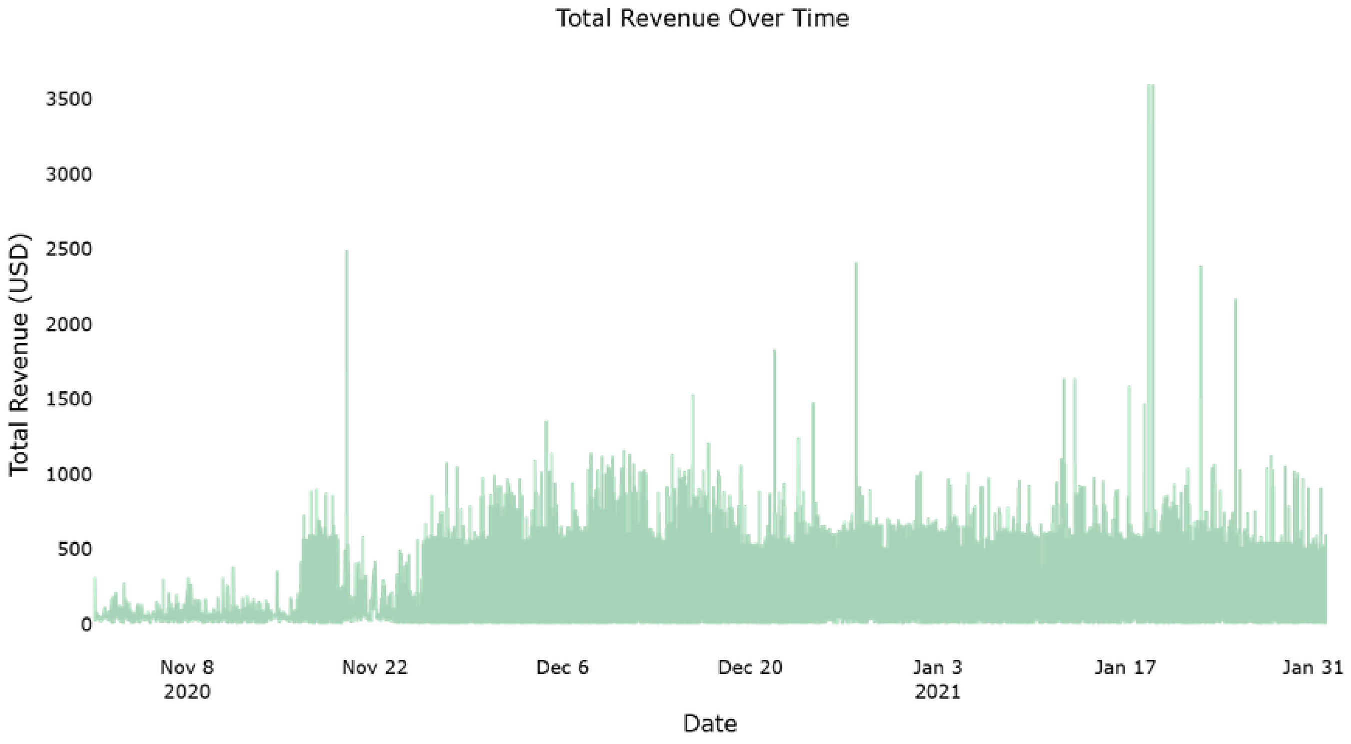
Business Insights

- Revenue is highly seasonal, with December acting as a critical sales window for marketing, promotions and inventory planning should be heavily optimized for this period.
- The January drop suggests post-holiday demand fatigue, highlighting an opportunity to introduce retention campaigns, New Year promotions, or loyalty incentives to sustain momentum.

Revenue Peaks Extend Beyond Holiday Season, Driven by High-Value Purchase Spikes

$18.7M

Total revenue over selected period



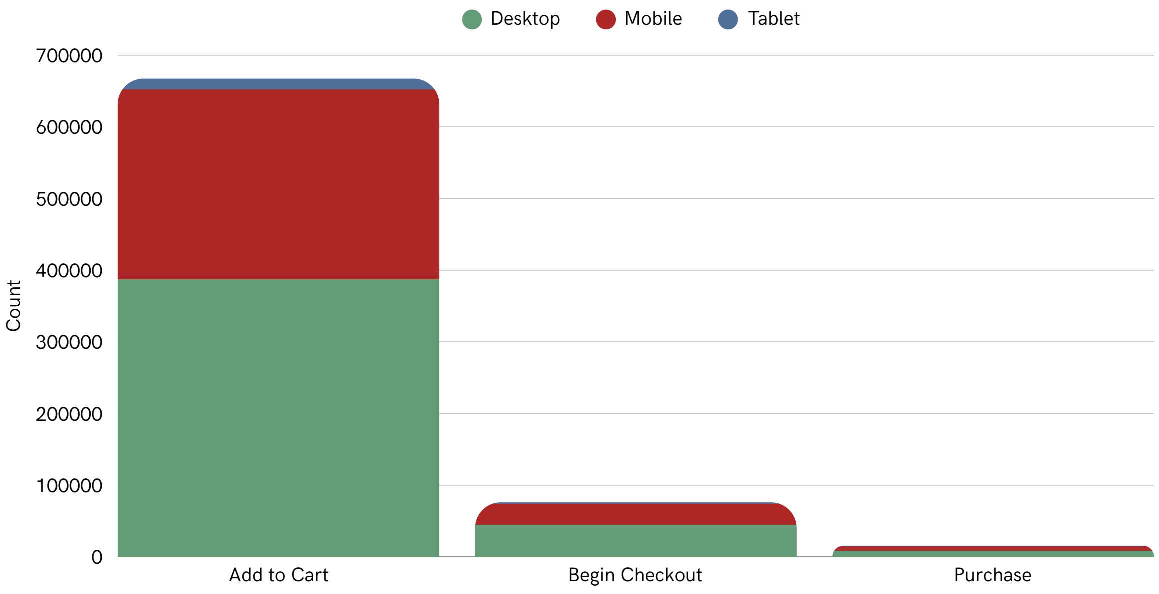
Key observations

- Revenue increases significantly from late November through December, aligning with the surge in purchase activity during the holiday season.
- Unlike purchase volume, revenue continues to show sharp spikes into January, indicating intermittent high-value transactions despite lower overall activity.

Business Insights

- A portion of total revenue appears to be driven by high-ticket purchases or bulk orders, suggesting the presence of premium or high-intent buyers.
- Post-holiday revenue resilience indicates an opportunity to promote higher-priced bundles, premium variants, or targeted upsell campaigns to sustain profitability even when overall traffic declines.

Desktop dominates revenue driving actions, while mobile shows higher drop-off across the purchase funnel



Key observations

- Desktop records the highest volume across all funnel stages (add_to_cart, begin_checkout, purchase), with a notably stronger progression to checkout and purchase compared to mobile.
- Mobile generates substantial add-to-cart activity but converts proportionally fewer users into completed purchases, indicating a steeper funnel drop-off.

Business Insights

- Desktop is the primary revenue-converting device, suggesting higher purchase intent or a smoother checkout experience on larger screens.
- Improving mobile checkout UX and reducing friction (payment flow, speed, form optimization) could materially increase overall conversion and unlock incremental revenue.

BUSINESS / DEVELOPER TAKEAWAYS

- **Geographic Concentration Drives Risk & Growth Potential:** Heavy reliance on the US market heightens revenue concentration risk, while India and Canada present the strongest opportunities for international expansion.
- **Desktop Experience Is the Primary Conversion Engine:** Desktop users generate the majority of conversions and revenue, making desktop UX optimization the most impactful short-term lever.
- **Mobile Scale Exists, but Conversion Friction Limits Value:** Strong mobile traffic is undermined by weaker conversion performance, indicating usability and checkout friction on smaller screens.
- **One-Time Visitors Dominate, Constraining Lifetime Value:** Most users purchase only once, making repeat buying the primary driver of long-term value.
- **Checkout Friction Suppresses Final Conversions:** Significant drop-offs during checkout suggest pricing clarity, trust signals, and payment flow optimization as high-impact priorities.

BUSINESS / DEVELOPER TAKEAWAYS

- **Apparel Anchors Store Performance**: Revenue and engagement are heavily concentrated in Apparel, making merchandising and pricing decisions in this category disproportionately influential.

- **Brand-Led Demand Limits Portfolio Diversification**: Google-branded products dominate interactions, while non-Google brands remain under-leveraged growth opportunities.

- **Bestsellers Sustain Revenue, Bundles Drive Expansion:** A small group of core products generates most sales, creating strong opportunities for bundling, cross-selling, and upsell strategies.

- **Repeat Purchasing Drives High Lifetime Value:** High-LTV customers are defined by purchase frequency rather than high ticket size, reinforcing the importance of repeat engagement strategies.

- **Seasonality Shapes Revenue Planning:** December is a critical sales peak, while post-holiday slowdowns highlight the need for targeted promotions to sustain momentum.

CHALLENGES & OPPORTUNITIES

Limitations

- The dataset is based on tracked user sessions and events from the Google Merchandise Store and may not fully capture offline purchases or interactions outside the platform.
- User behavior reflects only observed interactions within the dataset timeframe, potentially underrepresenting long-term or infrequent purchasers.
- Timestamped events provide a snapshot of activity during the observed period, limiting the ability to infer long-term behavioral shifts beyond the available data window.

Future Work / Opportunities:

- Incorporate customer reviews or feedback data (if available) to analyze sentiment and better understand product perception and satisfaction.
- Extend time-series analysis to longer observation windows to study how browsing, purchasing, and revenue patterns evolve over time.
- Perform deeper comparisons between repeat purchasers and one-time buyers to identify behaviors associated with higher lifetime value.
- Enrich the dataset with recommendation exposure or product review metrics to assess their impact on purchase decisions and conversion behavior.

CONCLUSION

This analysis of the Google Merchandise Store dataset provides insight into user browsing behavior, purchase patterns, and key revenue drivers across products, devices, and geographies. The findings demonstrate how funnel dynamics, product mix, pricing, and repeat purchasing behavior influence overall performance and customer lifetime value, while future work could extend this through longer-term time-series analysis and deeper comparisons between one-time and repeat purchasers.

This analysis was conducted using Python, Pandas, Matplotlib, Seaborn and Plotly.

Author:

Tejas Jadhav

GitHub: **@tejas-jadhav**

LinkedIn: **Tejas Jadhav**