

E-Commerce Consumer Behavior Analysis: From Raw Data to Actionable Insights

A Data-Driven Project for Optimizing Marketing Strategy

BY MOHAMED AMINE EL JARROUDI



PROJECT OVERVIEW & OBJECTIVES

Project Goal: To analyze a comprehensive e-commerce dataset to understand consumer behavior, identify key customer segments, and build actionable BI dashboards.

Key Objectives:

- Perform end-to-end data cleaning and transformation.
- Engineer new, insightful features (e.g., High-Value Customer, Season).
- Analyze demographic, behavioral, and sales trends.
- Visualize findings in two deliverables: an Excel dashboard and a multi-page Looker Studio report.

Tools Used: Excel, Looker Studio



DATASET OVERVIEW

Source: [Kaggle – E-commerce Consumer Behavior Analysis Dataset](#)

Records: 1,000 customers

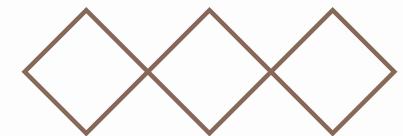
Key Data Points:

- **Customer Demographics:** Age, Gender, Income Level, Education.
- **Purchase Behavior:** Purchase Amount, Frequency, Product Category, Channel.
- **Behavioral Metrics:** Product Rating, Customer Satisfaction, Device Used, Purchase Intent.

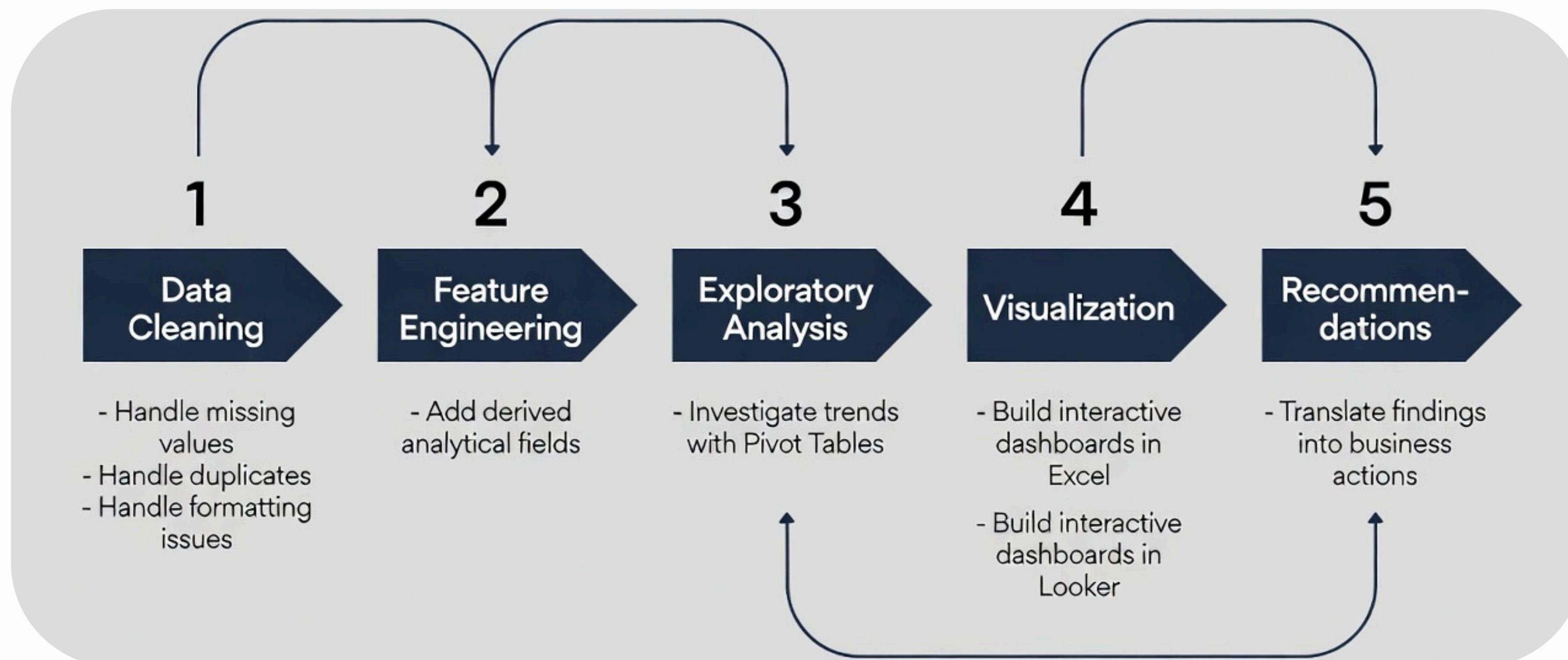


METHODOLOGY

The 4-Phase Process

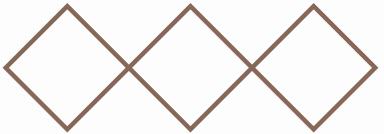


My approach was a 4-phase process, starting in Excel for deep cleaning and feature engineering, and finishing in Looker Studio for interactive visualization



METHODOLOGY

Data Cleaning in Excel



Data Integrity & Quality Assurance:

- **Encoding:** Resolved special character errors by setting file encoding to UTF-8 with BOM.
- **Missing Values:** Identified and analyzed blank cells using COUNTBLANK and pivot tables to ensure data completeness.
- **Duplicate Removal:** Ensured data integrity by removing duplicate records based on customer-id.
- **Data Type Correction:** Formatted columns appropriately (e.g., IDs as 'Text', currency as 'Number') to prepare for analysis.

METHODOLOGY

Data Transformation in Excel

Standardization & Normalization:

Text Normalization:

- Corrected inconsistent text cases (e.g., from UPPER to Proper) using PROPER() and LOWER() functions.
- Removed all extra/leading/trailing spaces using TRIM() and Find & Replace.

Data Consistency:

- Merged related values to create single, unified categories.
 - 1.Example: Identified "Health Supplements" and "Hotels" via pivot tables and merged them into "Health Care" and "Travel & Leisure" to ensure consistency.

METHODOLOGY

Feature Engineering in Excel

Creating New Analytical Features in Excel:

Demographic: Created an '*Age Group*' column (e.g., "18-25", "26-35") using the IFS function.

Behavioral:

- *High-Value Customer* column: Built a True/False column using IF(AND(...)) to isolate customers with high purchase amounts AND high frequency.
- *Season*: Mapped months to "Winter," "Spring," "Summer," "Autumn."
- Text Labels: Converted numerical 1-10 scores for *Customer_Satisfaction* into intuitive text (e.g., "Satisfied," "Neutral") using IFS.

Time-Series: Extracted *Purchase Month* and *Weekday* from the date using the TEXT function.

KEY INSIGHT 1

The "Who" - Customer Demographics

A high-level view of the 1,000-customer base, revealing a diverse and highly-educated audience.

Key Business KPIs:

- Total Customers: 1,000
- Total Sales: \$275,064
- Avg. Purchase Amount: \$275

Key Demographic Insights:

- Income: The customer base has a slight skew towards **High-Income earners (51.5%)**.
- Education: High prevalence of advanced degrees (**Bachelor's, High School, and Master's are all >30%**).
- Age: The customer base is relatively young; **26-35 and 36-44** are the largest segments.



KEY INSIGHT 2

The "VIPs" - Profiling High-Value Customers

A "High-Value Customer" (HVC) segment was engineered to identify the most impactful customers. This group shows a distinct and valuable profile.

Key Insights:

- **Channel:** High-Value Customers prefer Online (36.9%), followed by In-Store (32.8%) and Mixed (30.3%).
- **Education:** Education: HVCs are highly educated, with 62.5% holding degrees—29.2% with Master's degrees, and 33.3% with Bachelor's degrees. While the remaining 37.5% are high-school graduates.
- **Demographics:** The 26-35 age group forms the largest cohort of these key customers, with females accounting for 51.79%. Married customers represent the highest segment at 28.72%.
- **Top Categories:** Their top purchase categories by a large margin are Luxury Goods & Accessories, and Health Care.

KEY INSIGHT 3

The "When" - Timing & Seasonality

Understanding when customers buy reveals clear opportunities for optimizing marketing spend and timing.

Key Timing Insights:

- **Seasonal Peak:** Summer is the highest-grossing season (\$77.99K in sales), with a specific peak in August (\$27.4K). Spring is not far behind, with \$74.69K in sales.
- **Weekday Trend:** While sales are steady, **Monday** shows the highest average purchase amount (\$289.57), indicating a day of high-intent purchases.

Data-Driven Recommendations:

1. Shift marketing budget to launch major campaigns in late March, building anticipation for the proven April sales peak, followed by the summer peak.
2. Launch targeted email and social media campaigns on Mondays to capture these high-intent, high-value customers.

KEY INSIGHT 4

The "How" - Behavior & Channels

Analyzing how customers shop identifies the most effective channels and the mindsets driving their purchases.

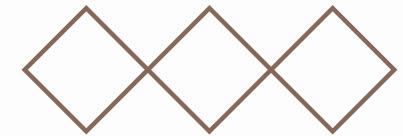
Key Behavioral Insights:

- **Channel Performance:** The In-Store channel has the highest average purchase amount (\$429.17), suggesting strong upsell/cross-sell opportunities.
- **Purchase Intent:** The primary purchase drivers are split between "Need-based" (71K in sales) and "Wants-based" (72K in sales).
- **Payment method:** PayPal accounts for 21.9% of all sales, which makes it number one.
- **Device used:** 35% of orders are placed using a desktop.

Data-Driven Recommendations:

1. Invest in in-store sales training to capitalize on the already-higher average order value.
2. Improve our UX on the desktop version of the website and investigate promoting PayPal across all locations, as it is a key payment method.

CONCLUSION & NEXT STEPS

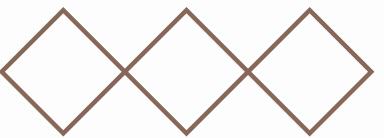


Conclusion:

- Successfully transformed a raw dataset into an actionable, multi-platform BI report.
- Identified a key customer profile: A 26-35 year old, highly-educated, "High-Value" customer who shops online and is sensitive to discounts.
- Isolated clear sales peaks (April, Summer, Mondays) and top products (Luxury, Health Care) for targeted marketing and inventory planning.

Future Recommendations:

- **Marketing:** Launch targeted campaigns for the key "26-35 HVC" profile.
- **Promotions:** Test a "Monday" promotion to capitalize on high-value weekly purchases.
- **Investigation:** Deep-dive into why Goteborg is such a high-performing location and attempt to replicate its success.
- **Seasonal Optimization:** Allocate higher ad budgets in Spring and Summer.
- **Customer Satisfaction:** Investigate causes of low ratings and high returns (954 returned items).
- **Digital Marketing:** Use influencers to reach discount-sensitive, younger demographics.
- **Channel Improvement:** Encourage cross-channel engagement (online + in-store).



THANK YOU

LINKS:

THE LOOKER DASHBOARD

GITHUB PORTFOLIO

LINKEDIN ACCOUNT

