**Abstract:** This paper presents a DIVE (Discover, Investigate, Validate, Extend) analysis of retail customer transaction data, with a focus on enabling digitally-driven customer strategies. Using clustering techniques and statistical validation, we uncover actionable insights into customer segments, purchasing behavior, and market dynamics. These findings contribute to a broader digital transformation initiative, helping retailers transition from intuition-based decisions to real-time, data-driven personalization, segmentation, and retention strategies.

**1. Discover: Understanding Segments, Purchasing Patterns, and Market Trends**

**1.1 Dataset Overview:** The dataset, sourced from Google BigQuery, includes one thousand retail transactions. It captures attributes such as date, category, price, discount, units sold and ordered, inventory, demand forecast, competitor pricing, and store identifiers.

**1.2 Data Preparation:** All records were complete with no missing values. Date columns were transformed into time-based features like month and day of the week. Numeric features were standardized to prepare the data for clustering algorithms.

**1.3 Key Observations:** There was a wide variance in price and discount activity, reflecting active promotional strategies. Demand forecasts were closely aligned with actual sales, suggesting a well-tuned planning process.

**2. Investigate: Who Are Our Most Valuable Customers and What Drives Them**

**2.1 Segmenting Customers:** We applied K-Means clustering on features related to purchasing behavior. The Elbow Method revealed that three clusters offered the best balance between simplicity and differentiation. Cluster zero represented high-value generalists who made large, consistent purchases across a wide range of categories. Cluster one included mid-value toy shoppers who spent moderately and displayed some price sensitivity. Cluster two comprised budget-conscious furniture buyers with a lower overall spend and a seasonal purchasing pattern.

**2.2 Value by Cluster:** Cluster zero had high average units sold and high average price, resulting in an average transaction value of $108.01. Cluster one had moderate values across the board, with an average transaction value of $21.42. Cluster two showed lower units sold and price points, with an average transaction value of $21.03.

**2.3 Behavioral Insights:** Promotions had a strong effect on Cluster zero, with increased activity in response to discounts. Cluster one showed a clear preference for toy products, while Cluster two leaned heavily toward furniture. Seasonality trends indicated that Cluster zero maintained consistent purchasing throughout the year, while Clusters one and two showed more cyclical behavior.

**3. Validate: Do These Patterns Hold Up**

**3.1 Hypothesis Testing:** We used ANOVA to test whether the clusters differed meaningfully in their purchase behavior. The resulting F-statistic was very high, and the p-value was 2.82e-198, confirming that the clusters represented statistically distinct groups.

**3.2 Supporting Evidence:** There was a very strong correlation between demand forecasts and units sold, with a coefficient of 0.997. Prices were closely aligned with competitor pricing, showing a correlation of 0.994. A few outliers were identified in inventory and units sold, suggesting operational inconsistencies. The product category mix within each cluster reinforced the behavioral segmentation, with toys dominating Cluster one and furniture dominating Cluster two.

**4. Extend: How Can We Use This to Grow and Retain Customers**

**4.1 Tailored Strategies for Each Cluster with Digital Enablement:** The segmentation provides a strong foundation for delivering digitally personalized customer experiences. By integrating this segmentation into CRM and marketing automation systems, businesses can move from generalized messaging to tailored engagement.

For Cluster zero, high-value generalists, acquisition efforts can focus on premium ad placements and digital referral campaigns. Retention strategies can include personalized rewards, early access to product launches, and dedicated VIP support through email or app-based channels.

For Cluster one, mid-value toy shoppers, outreach can include campaigns in partnership with toy brands and targeted ads on family-focused digital platforms. Retention can be improved through gamified loyalty programs, personalized toy bundles, and recommendations powered by machine learning.

For Cluster two, budget-conscious furniture buyers, acquisition should highlight value messaging and product durability through social and search ads. Retention tactics might include time-sensitive discounts and helpful digital content such as virtual room styling tools and furniture care guides.

**4.2 Recommendations Going Forward: Scaling With Digital Infrastructure:** To operationalize these insights at scale, businesses should automate segmentation updates using scheduled BigQuery ML models or dbt workflows. These segment definitions can be pushed directly into CRM platforms for activation.

A B testing tools like Google Optimize or Braze should be used to continuously test and refine strategies within each segment. Real-time behavior data can be integrated into dynamic segmentation models, allowing businesses to respond immediately to shifts in customer intent.

Internal dashboards should be built and shared across marketing, merchandising, and product teams to ensure consistent access to actionable insights. Embedding these workflows into business-as-usual operations represents a significant step forward in digital maturity.

**5. Conclusion:** This analysis uncovered meaningful and measurable customer segments that differ in value, behavior, and preference. These findings support not only more targeted acquisition and retention strategies, but also the broader digital transformation of customer engagement. By embedding segmentation and behavior insights into automated systems, businesses can move from one-size-fits-all marketing to dynamic, personalized experiences. Future exploration into real-time promotions, journey orchestration, and time-based behavior could further elevate the level of digital intelligence and customer relevance.