Market Insights for E-commerce Company in 2019

Turing College
Linas Sutkaitis
2024-02-01

Executive summary

The e-commerce data provided relates to a fictional company based in the USA, specializing in the online retail of home and office appliances catering to both men and women. The data source for this project is a publicly available dataset obtained from Kaggle.com, which can be accessed here .

The primary objective of the company is to enhance customer experience, elevate satisfaction levels, and ultimately boost revenue. To achieve these goals, the project employed various data analysis techniques, including RFM segmentation, retention analysis using cohorts, assessment of customer lifetime value (CLV), and product analysis.

The project's intended audience is the management team of the e-commerce company, keen on comprehending the company's performance and pinpointing areas for improvement. It aims to deliver detailed insights into various facets of the company's operations, such as sales trends, customer behavior, and products analysis.

Ultimately, the goal is to provide actionable insights to the management team, facilitating improved company operations and fostering growth.

The project aimed to answer several questions, including:

- Who are the company's most valuable customers, and how can the company segment them to improve the effectiveness of targeted marketing campaigns?

- How can cohort analysis assist business in identifying trends and patterns in customer behavior over time?

- What are the Customer Lifetime Value (CLV) and what are the differences between genders, and how can the company enhance its value?

- How can the company identify the products that contribute the most to revenue, and what strategies can be used to improve their performance using product analysis?

- How can the company understand customer purchasing behavior and then optimize product offerings and promotions accordingly using market basket analysis?

After analyzing the dataset, several key insights and recommendations could be derived:

- The analysis revealed that the company's most valuable customers are the ‘Top Customers’ and ‘Loyal Customers’, generating approximately 58% (3.1 million) in revenue. By targeting these segments with personalized marketing strategies, businesses can enhance their efforts to nurture and grow these relationships, ultimately leading to better customer satisfaction and increased overall company revenue

- The analysis revealed a notable ‘At-Risk’, ‘Cant Loose Them’ and ‘Customers Needing Attention’ customer segments, accounting for roughly 31% (1.68 million) of revenue. Prior to refining our targeting for this segment, it's essential to engage with these customers through surveys, feedback forms, or one-on-one conversations to grasp their concerns. Tailoring our communication based on their feedback and preferences is crucial to demonstrate value. Once their situation is understood, we can offer solutions or incentives such as discounts or personalized offers to encourage sustained engagement, highlighting the benefits of loyalty to our brand.

- The cohort analysis revealed varying percentages of customers making their first purchase in the same month across all cohorts. Initially, after 2-4 months from their initial first purchase, we observed a decline in the first cohorts, while over a longer period, the number of retained customers increased, indicating the presence of top customers or other contributing factors. However, the most recent cohorts exhibited a significant decrease in the number of retained customers in later months.

- Through the calculation of Customer Lifetime Value (CLV) and cohort analysis, we employed predictive modeling to account for changing revenue patterns across different cohorts. This approach yielded more accurate results, with the overall estimate standing at 5166 USD. Further segmentation by gender indicates that female customers have an estimated CLV of 5102 USD, while male customers have an estimated CLV of 5166 USD.

- The insights reveal that certain product categories, namely Nest-USA, Apparel, Office, Drinkware, and Lifestyle, generate the highest revenue, indicating strong customer demand in these areas. This suggests potential for increased overall revenue and customer satisfaction by focusing on these high-revenue categories. Additionally, while the Office category may not generate the most revenue, it is among the top 5 categories and holds significant popularity among customers. This popularity underscores its importance as a driver of customer acquisition and retention.

- Taking a product-centric view, the company offers 404 distinct products purchased by customers. Applying the Pareto principle, we observe that 80% of revenue is generated from just around 15% of these products, translating to approximately 58 products contributing to around 80% (4.3 million) of total revenue. Prioritizing these key products,

organizations can allocate resources more efficiently, aiding stakeholders in pinpointing where to focus their efforts for maximum impact and which marketing strategies to emphasize.

- Through market basket analysis with two-product combinations, it's evident that customers who purchase the 'Nest Cam Indoor Security Camera - USA' tend to also buy the 'Nest Cam Outdoor Security Camera - USA,' occurring 459 times, which represents 9% of all occurrences involving the purchase of two products. However, when examining three-product combinations, it's notable that 'Google Sunglasses' added to the basket occurs 112 times, nearly 2% of occurrences in three-product combinations. Analyzing these three-product combinations reveals the importance of having additional information about the products (e.g., size, color) to recommend the right product and enhance upselling efforts, ultimately increasing revenue and customer satisfaction.

Introduction to the dataset

The e-commerce data provided relates to a fictional company based in the USA, specializing in the online retail of home and office appliances catering to both men and women.

The dataset comprises approximately 53,000 rows spread across 5 tables, containing diverse information on customers, online sales, product details, taxes, and discounts for the year 2019. Throughout this period, the company generated a revenue of approximately $5.4 million, engaging 1,468 unique customers in roughly 26,631 distinct transactions, with an average transaction value of approximately $203.

Designed for analysts and data scientists, the dataset serves as a valuable resource for gaining deeper insights into the operations and performance of e-commerce companies from various perspectives.

Data preparation and preprocessing

Before commencing any data analysis, data preparation and preprocessing steps were executed to guarantee the accuracy and consistency of the dataset. This process entailed identifying and rectifying duplicate records, as well as ensuring the appropriate data types were utilized and additional fields were created where necessary.

Accurate calculation of the invoice value is vital for sound financial management. To ensure precision, it is crucial to adhere to the correct formula. The primary formula for calculating the invoice value was provided, which accounts for quantity, product price (avg_price), discount percentages, goods and services taxes (GST) and delivery charges, the data preprocessing approach was implemented for users who utilized any discount codes.

Invoice Value = ((Quantity * Avg_price) * (1 - Discount_pct) * (1 + GST)) + Delivery_Charges

For customers who did not use any discount codes or for whom there is no information available regarding discounts:

Invoice Value = ((Quantity * Avg_price) * (1 + GST)) + Delivery_Charges

In my analysis, I identified customers who purchased products with and without discounts to guarantee precise calculations of the customer invoice value. This meticulous approach is essential for maintaining financial transparency, making informed decisions, and fostering trust with customers.

Exploratory Analysis

Once the data was prepared and preprocessed using SQL, exploratory analysis was conducted to provide an overview of the company's performance and gain initial insights. The visualizations were prepared using the business intelligence tool Tableau, where interactive dashboards were created. You can access the interactive dashboard through this link.

Firstly, the main Key Performance Indicator (KPI) for the team was established to assess the overall performance of the business.

Customers	Revenue | (AOV)	Customer Lifetime Value	Marketing Costs	Marketing ROI %
1 468	$5 397 367 | $202,67	$5 123	$1 733 546	211,35%

Table 1.1 Main KPI's

From these data points, we observe that we have 1,468 unique customers, generating approximately $5.4 million in revenue in 2019, with an average order value of around $203. Utilizing the predictive model, we determined our predictive customer lifetime value to be $5,134. Additionally, we allocated approximately $1.7 million to marketing campaigns across online and offline channels, representing 32% of the revenue generated in 2019. Looking at generated revenue and marketing costs we have 211 % of return of investments.

Examining sales trends, we observe revenue fluctuations, possibly reflecting purchases of products with longer expiration periods for inventory replenishment. Peak sales predominantly occur at year-end, driven by heightened gift-buying and November's Black Friday campaigns. Conversely, the lowest sales period typically falls at the start of summer in June. Transactionally, there is steady growth in customer transactions.

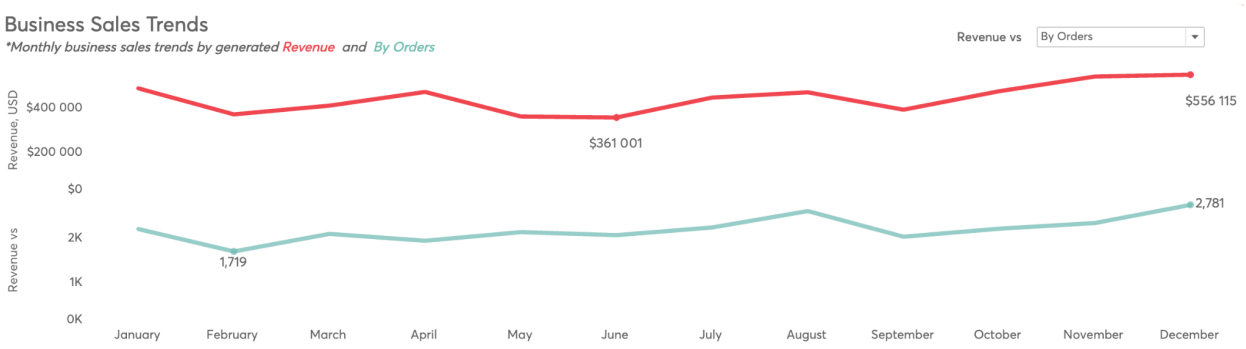


Table 1.2 Business sales trends revenue versus number of orders

From a gender perspective, we observe that the majority of revenue, 62% (approximately $3.34 million), is generated by female customers, while males contribute only 38% (around $2 million) to the revenue. When analyzing monthly revenue distribution, females consistently lead in revenue percentage across all months, although there is a slight increase in the percentage contributed by males in January and September.

With this information, we can tailor our marketing strategies to target female customers more effectively, capitalizing on their higher contribution to revenue. Additionally, we can explore ways to engage male customers further during months where their percentage of revenue increases, such as offering targeted promotions or product bundles.

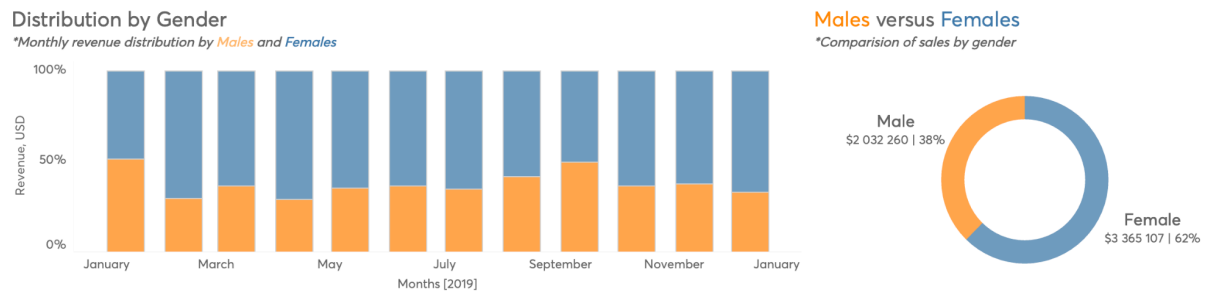


Table 1.3 Revenue distribution by gender

Looking through customer segments we can see that the majority of revenue is generated by ‘Top Customers’ and ‘Loyal Customers’ generating approximately 58% (3.1 million) in revenue. By targeting these segments with personalized marketing strategies, businesses can enhance their efforts to nurture and grow these relationships, ultimately leading to better customer satisfaction and increased overall company revenue

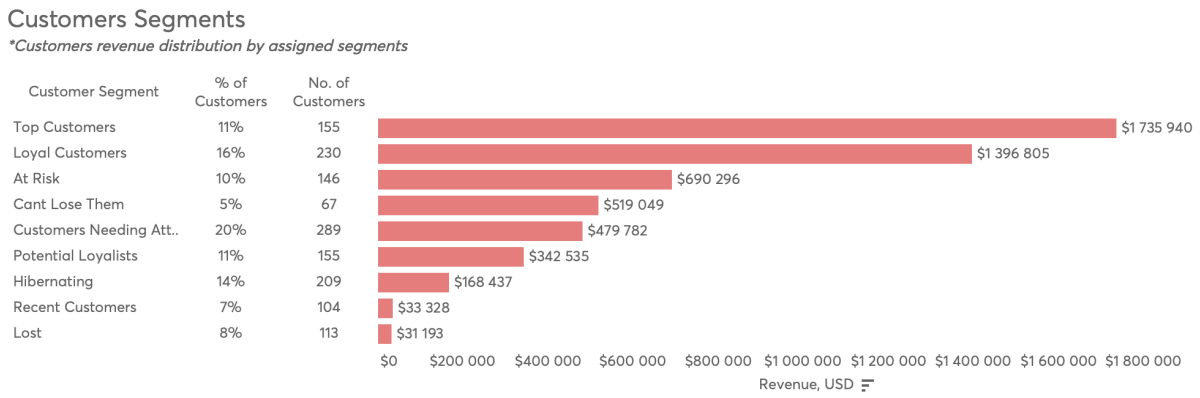


Table 1.4 Revenue distribution by customers segments

From a Top 5 product categories point of view, we observe that the majority of revenue, around $2.7 million, is generated from products categorized under Nest-USA, which are designed for region-specific compatibility and possibly regulatory certifications. Conversely, revenue from apparel, nest, and other categories is comparatively lower.

With this product category insight, we can focus marketing efforts on promoting the high-performing Nest-USA category while strategizing ways to improve sales in lower-performing categories like apparel and nest and other categories.

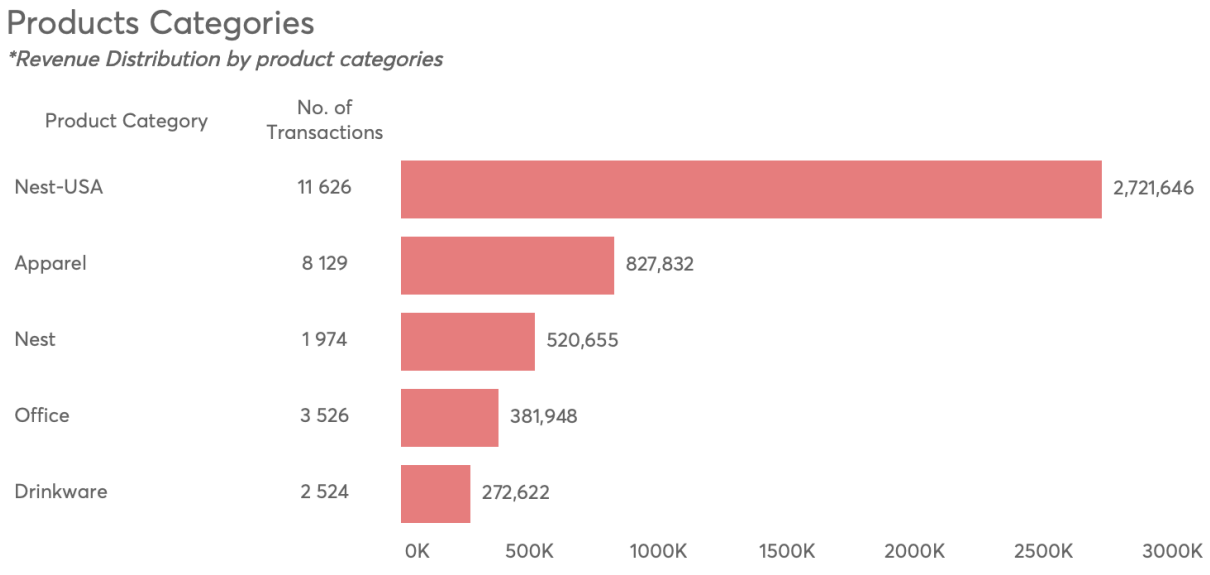


Table 1.4 Revenue distribution by product categories

RFM (Customers Segmentation) Analysis

RFM analysis segments customers based on their transactional behavior, where RFM refers to Recency (time elapsed since a customer's last interaction), Frequency (number of interactions within a specific period), and Monetary (total spending within the same period). This segmentation technique aids in identifying valuable customers for retention and upselling opportunities while also enabling strategies to retain other customers.

The process of assigning scores involved calculating RFM scores for each customer using SQL code. This was done by determining percentiles for the customers' Recency, Frequency, and Monetary values. Customers were then categorized into one of nine segments based on these RFM scores. The code calculated percentiles for Monetary, Frequency, and Recency values by dividing them into five equal parts (25th, 50th, 75th, and 100th percentiles). Subsequently, RFM scores were computed for each customer by comparing their Monetary, Frequency, and Recency values to these percentiles. Finally, customers were assigned to one of the nine segments based on their combined R (Recency), F (Frequency), and M (Monetary) scores.

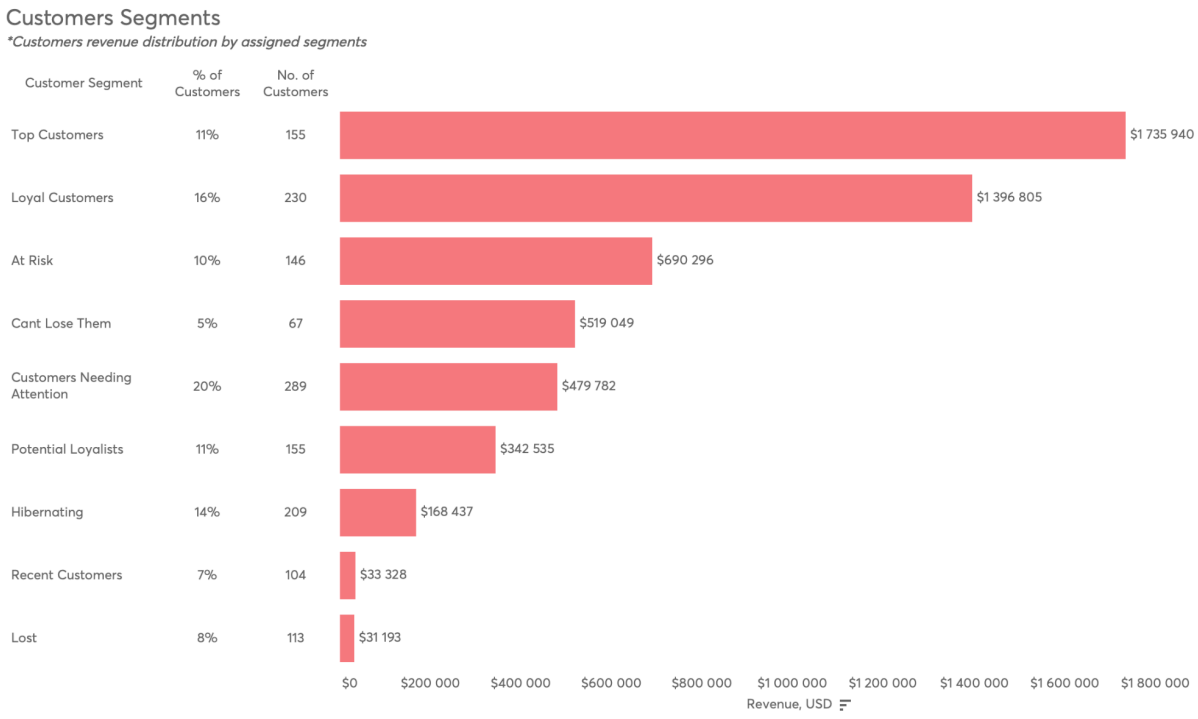


Table 2.1 Customers revenue distribution by assigned segments

The analysis of revenue distribution across RFM segments provides valuable insights into the diverse spending behaviors exhibited by different customer groups. This understanding enables businesses to effectively allocate resources and tailor marketing efforts to maximize returns.

Focusing on high-potential customer segments, such as 'Top Customers' and 'Loyal Customers,' is particularly strategic. These segments often contribute significantly to overall revenue and represent loyal clients who are more likely to make repeat purchases and engage with the brand over time. By implementing targeted marketing strategies, businesses can nurture these relationships, deepen customer engagement, and foster long-term loyalty. Investing in personalized marketing initiatives, exclusive offers, loyalty programs, and exceptional customer service for these segments can enhance their overall experience

Analyzing revenue distribution for customer segments labeled as 'At Risk' and 'Customers Needing Attention' provides critical insights into potential challenges in customer retention and engagement. Before fine-tuning our focus on these groups, it's imperative to engage with them through surveys, feedback forms, or direct discussions to grasp their concerns. Adapting our communication based on their feedback and preferences is crucial to demonstrating value. Implementing targeted re-engagement strategies, such as personalized communications and tailored incentives, can reignite their interest in the brand and encourage renewed purchasing activity. By proactively re-engaging with these segments and addressing their needs, businesses can prevent potential revenue loss and strengthen customer relationships and loyalty over time.

Looking at revenue distribution among segments like 'Can't Lose Them' and 'Potential Loyalists' reveals their importance for long-term success. Re-engaging with these segments is crucial to solidify their loyalty and drive growth. Tailored communication and offerings, such as personalized promotions and rewards, can reignite their interest. Gathering feedback directly from these segments helps understand their needs better. By proactively re-engaging with them, businesses can secure their revenue and foster a loyal customer base essential for sustained success.

Furthermore, alternative customer segments including Hibernating, Recent Customers, and Lost exist. Tailored marketing strategies can be developed to serve these customer groups effectively. For instance, Recent Customers could be offered exclusive incentives to facilitate their transition into loyal, long-term clients. Reactivation campaigns aimed at Hibernating or Lost customers have the potential to reintegrate them into the business ecosystem.

In summary, it is essential for the business to uphold strong relationships with its core customer base while simultaneously addressing the needs of other customer segments in order to optimize customer retention and revenue generation.

So the question before conducting RFM analysis was: Who are the company's most valuable customers, and how can the company segment them to improve the effectiveness of targeted marketing campaigns? The analysis showed that the companies most valuable are those whose fall under the ‘Top Customers’ and ‘Loyal Customers’ segments based on the RFM analysis. These customers have highest recency, frequency and monetary values indicating consistent

engagement and high-value purchases from the company. For better targeting these customers the following approach can be done:

Top Customers:

Goal: retain, engage and upsell other products
Strategy: Company can provide exclusive deals, personalized suggestions, and loyalty incentives to stimulate ongoing engagement and sales. Investigate options for upselling or cross-selling complementary products or services to optimize revenue potential.

Loyal Customers:
Goal: Increase loyalty, prevent them from churning
Strategy: Nurture these customers with tailored communication, special promotions, and exclusive access to new products or services. Encourage referrals by offering incentives for referring friends or family members.

By concentrating on these two valuable segments, you can efficiently allocate resources to sustain and enhance relationships with these customers, resulting in increased revenue and customer satisfaction.

Cohort Analysis

Cohort analysis provides invaluable insights into our customers' behavior and lifecycle, enabling businesses to make data-driven decisions and further refine their marketing strategies. By segmenting customers into cohorts based on shared characteristics at the time of their initial purchase, we can assess performance across different groups and periods to uncover important trends.

This analytical approach helps identify customer engagement patterns, evaluate promotional campaigns, and assess the impact of product or service changes on customer churn rates. Additionally, cohort analysis offers insights into customer retention, churn rates, and customer lifetime value, all of which are potential revenue streams for the business.

In summary, cohort analysis emerges as a potent instrument for comprehending customer behavior across time, informing strategic decision-making, and ultimately fostering business growth and success. By scrutinizing cohorts from the present year, the business endeavors to unveil valuable insights that will steer efforts toward refining and elevating the customer experience in the coming future.

The cohort analysis process begins with cohort identification, where customers are grouped into monthly cohorts based on their initial first purchase. Each cohort represents customers who made their initial purchase during the same month. Cohort Metrics Calculation involves determining the cohort size and the number of customers still active in subsequent months for each cohort and using conditional aggregation to find these metrics, counting customers based on certain conditions like whether they bought something in a month using different intervals or if their last purchase was within a certain time period.

The retention rate for each cohort was calculated by dividing the number of customers still active in the subsequent months by the initial cohort size. This gives us a percentage that represents the proportion of customers retained over time, allowing us to compare the performance of different cohorts.

To present the cohorts more clearly, the results were organized in a heatmap format using Tableau. This table displays the start monthly cohorts, the number of customers in each cohort, and the number of customers in subsequent months. This layout makes it easier to comprehend the various patterns in customer behavior across different periods within each cohort.

Customers Retention

**Retained customers by theirs initial first transaction*

Monthly Cohorts	# Customers	0	1	2	3	4	5	6	7	8	9	10	11
2019 January	215	215	13	24	34	23	44	35	47	23	28	20	34
2019 February	96	96	7	9	16	17	22	19	15	12	11	16	
2019 March	177	177	18	35	25	32	33	22	22	15	19		
2019 April	163	163	14	24	24	18	15	10	16	12			
2019 May	112	112	12	9	13	10	13	14	8				
2019 June	137	137	20	22	12	11	14	11					
2019 July	94	94	13	4	6	11	9						
2019 August	135	135	14	15	10	8							
2019 September	78	78	6	3	2								
2019 October	87	87	6	4									
2019 November	68	68	7										
2019 December	106	106											

Table 3.1 Retained customers by monthly cohorts

Customers Retention Rates

*Percentages of retained customers by theirs initial first transaction

Monthly Cohorts	# Customers	0	1	2	3	4	5	6	7	8	9	10	11
2019 January	215	100%	6%	11%	16%	11%	20%	16%	22%	11%	13%	9%	16%
2019 February	96	100%	7%	9%	17%	18%	23%	20%	16%	13%	11%	17%	
2019 March	177	100%	10%	20%	14%	18%	19%	12%	12%	8%	11%		
2019 April	163	100%	9%	15%	15%	11%	9%	6%	10%	7%			
2019 May	112	100%	11%	8%	12%	9%	12%	13%	7%				
2019 June	137	100%	15%	16%	9%	8%	10%	8%					
2019 July	94	100%	14%	4%	6%	12%	10%						
2019 August	135	100%	10%	11%	7%	6%							
2019 September	78	100%	8%	4%	3%								
2019 October	87	100%	7%	5%									
2019 November	68	100%	10%										
2019 December	106	100%											

Table 3.2 Percentages of retained customers by monthly cohorts

Based on the 3.1 and 3.2 results here are recommendations and insights:

The "initial size" column indicates the total customer count in each cohort (100%). The "Month 0" column displays the percentage of customers who made a purchase in the same month as their initial purchase. Looking through the cohorts, we notice a significant decrease in the percentage of retained customers after the first month, ranging from 8% to 12% on average. However, for the oldest cohorts (from January 2019 to March 2019), despite a substantial drop in customers during the longer period of 3 to 6 months after their initial purchase, we observe an increase in retention rates from 16% to 22%. This indicates that certain strategies may have been implemented during that time period, prompting customers to return and make additional purchases. It's crucial to investigate these strategies further and adapt them to create even more effective retention strategies for retaining more customers in future periods

These insights can guide marketing strategies and highlight areas needing improvement. For example, focusing on enhancing customer engagement and retention, especially for cohorts with lower retention rates, could be beneficial and it could be an eye opener about the customers behavior. Continuously monitoring the data over subsequent months can offer deeper insights into evolving customer behavior patterns. We need to explore strategies to re-engage customers who haven't made repeat purchases, such as personalized email campaigns or promotional

offers, to encourage their return to the platform as we saw that only a small portion of customers are buying the product in the subsequent months.

Recommendations:

Improving the customer experience is crucial, particularly during the first month of engagement. This can be achieved through personalized recommendations, targeted promotions, exceptional customer support, and gathering feedback through surveys or one-on-one interactions to better understand our customers and represent our brand.

Implement continuous engagement strategies, including loyalty programs, targeted marketing campaigns, and personalized offers, to sustain customer interest and foster repeat purchases over time.

Continuously monitor and update the data to ensure accurate insights and recommendations for implementing marketing strategies. Assess whether the marketing approach is effective and allow additional time for incomplete cohorts before making any definitive decisions.

So the question was raised at the beginning of the project: How can cohort analysis assist business in identifying trends and patterns in customer behavior over time? Analyzing cohorts enables the identification of trends in customer retention rates and behavior over various timeframes. This understanding allows businesses to customize their marketing strategies, product offerings, and enhancements to the customer experience, targeting high-value segments and address underperforming cohorts, thereby fostering long-term growth.

Customer Lifetime Value (CLV) Analysis

Customer Lifetime Value (CLV) is a metric used to estimate the total value a customer will generate for the company throughout their lifetime. It helps identify the most valuable customers for retention and upsell opportunities. CLV can be calculated using various methods, including standard formula, historical averages or predictive modeling. For this analysis, a more in-depth approach such as predictive modeling using historical averages on cohort analysis was utilized, allowing for continuous updates based on new data.

The Customer Lifetime Value (CLV) analysis starts by pinpointing the customer's initial purchase time, marking their introduction to the company for the current year as our starting cohort. This is followed by reviewing all transactions made in the following months and calculating the revenue generated during those periods. This revenue is then divided by the cohort size to determine the average order value and assess how it varies over time.

Customers Average Order Value

**Customers average order value by number of customers cohorts size*

Monthly Cohorts	# Customers	Months 0	1	2	3	4	5	6	7	8	9	10	11
2019 January	215	$2298	$216	$275	$616	$167	$270	$465	$283	$211	$248	$202	$457
2019 February	96	$3423	$101	$159	$276	$276	$531	$294	$509	$205	$321	$529	
2019 March	177	$1956	$273	$239	$187	$319	$446	$296	$259	$152	$219		
2019 April	163	$1727	$169	$113	$220	$176	$202	$112	$372	$120			
2019 May	112	$2083	$67	$133	$159	$160	$379	$363	$107				
2019 June	137	$1586	$97	$112	$128	$88	$295	$121					
2019 July	94	$1919	$181	$84	$177	$275	$320						
2019 August	135	$1695	$84	$95	$230	$134							
2019 September	78	$2083	$28	$35	$9								
2019 October	87	$2959	$122	$33									
2019 November	68	$3450	$72										
2019 December	106	$2485											

Table 4.1 Customers average order value by number of cohort size

Table 4.1 illustrates the average revenue per customer for various cohorts from January 2019 to December 2019. Each row corresponds to a cohort's initial purchases, and each column indicates the average monthly revenue for that cohort relative to the starting month. For instance, in January 2019 (Month 0), the average revenue per customer was $2298, while in February 2019 (Month 1), it decreased to $216. The table demonstrates the changes in average revenue over time.

The observation from the CLV cohort analysis: The average revenue per user in the starting month (Month 0) varies across cohorts, with the highest revenue in November 2019 ($3450) and the lowest in June 2019 ($1586). The average revenue in Month 1 generally greatly decreases compared to Month 0 for all cohorts. This is anticipated since customers generally spend more during their initial interaction with the platform. However, revenue trends vary among cohorts over time, with some cohorts exhibiting higher revenues in specific months while others show lower revenues. This variance could be attributed to factors like seasonality, promotional activities, or shifts in user behavior.

The most recent cohorts, spanning from September 2019 to December 2019, exhibit the highest starting revenue per customer compared to all other cohorts. However, since it is the most recent

cohort, data for subsequent months is unavailable. Nevertheless, this cohort analysis offers valuable insights into the revenue patterns of various cohorts over time, informing potential adjustments to marketing and product strategies.

Cumulative Customers Average Order Value

**Customers are by the month of they initial first transaction*

		Months											
Monthly Cohorts	# Customers	0	1	2	3	4	5	6	7	8	9	10	11
2019 January	215	$2298	$2514	$2790	$3406	$3573	$3843	$4309	$4591	$4802	$5050	$5252	$5709
2019 February	96	$3423	$3524	$3683	$3960	$4235	$4767	$5061	$5570	$5775	$6096	$6626	
2019 March	177	$1956	$2229	$2468	$2656	$2975	$3421	$3717	$3976	$4127	$4346		
2019 April	163	$1727	$1896	$2009	$2229	$2405	$2606	$2719	$3091	$3211			
2019 May	112	$2083	$2149	$2282	$2441	$2601	$2979	$3342	$3449				
2019 June	137	$1586	$1683	$1796	$1924	$2012	$2307	$2428					
2019 July	94	$1919	$2100	$2184	$2360	$2636	$2955						
2019 August	135	$1695	$1780	$1875	$2105	$2238							
2019 September	78	$2083	$2111	$2146	$2155								
2019 October	87	$2959	$3081	$3114									
2019 November	68	$3450	$3523										
2019 December	106	$2485											
Average		$2305	$2434	$2562	$2784	$2983	$3332	$3608	$3914	$4086	$4349	$4714	$5171
Growth %			5,56%	5,26%	8,68%	7,16%	11,70%	8,26%	8,48%	4,39%	6,44%	8,41%	9,69%

Table 4.2 Cumulative customers average order value

Table 4.2 presents the same data as Table 4.1, but the revenue for each cohort in a particular month is expressed as a cumulative sum. This means that the revenue from the previous month is added to the revenue from the current month. Subsequently, the averages for all month numbers (representing months since the first purchase) were calculated, followed by the calculation of percentage growth based on these average numbers.

Essentially, the table below displays the revenue growth by customers' first purchase cohort over X months after registration. The figures below summarize these values in terms of cumulative averages (Monetary) and percentages (Growth %). This offers a comprehensive perspective on the anticipated revenue growth for the business based on historical data.

Next, predictive modeling will be utilized to forecast the missing data, which in this case refers to the revenue expected from later-acquired customer cohorts. For instance, for users whose first purchase occurred in December 2019, we currently have data only for their first month's purchase revenue, which amounts to $2485 per customer. However, the revenue for the subsequent months remains unclear.

The previously calculated Cumulative Growth Percentage (Growth%) will be utilized, and predictions for all 11 future months' values will be made. For example, for this cohort, the expected revenue for month 1 can be calculated as $2485 x (1 + 5.56%) = $2623, and for month 2 as $2623 x (1 + 5.26%) = $2761. By using the average cumulative growth for each month, it can be projected that based on the initial value of $2485, the revenue for month 11 is expected to be $5574.

Future Prediction of posible customers purchases

**Future prediction of customers sales trends using cumulative growth percentages and last cohorts months values*

	Months										
Monthly Cohorts	1	2	3	4	5	6	7	8	9	10	11
2019 February											$7268
2019 March										$4712	$5169
2019 April									$3418	$3706	$4065
2019 May								$3601	$3832	$4155	$4557
2019 June							$2634	$2750	$2927	$3173	$3481
2019 July						$3200	$3471	$3623	$3857	$4181	$4586
2019 August					$2500	$2707	$2936	$3065	$3263	$3537	$3880
2019 September				$2309	$2580	$2793	$3030	$3163	$3366	$3649	$4003
2019 October			$3385	$3627	$4051	$4386	$4758	$4967	$5287	$5731	$6287
2019 November		$3708	$4030	$4318	$4824	$5222	$5665	$5914	$6295	$6824	$7485
2019 December	$2623	$2761	$3001	$3216	$3592	$3889	$4219	$4404	$4687	$5081	$5574

CLV: $5 123

Table 4.3 Future prediction of possible customers purchases

Utilizing predictive modeling to compute a CLV of $5123 showcases a data-driven and precise method for determining customer lifetime value. This approach leverages historical data, patterns, and trends identified through cohort analysis to offer a more accurate estimation of a customer's expected spending over their lifetime with the company.

While the CLV method itself doesn't inherently differentiate between genders, it can be used to understand customer behavior across different genders by analyzing various factors contributing to CLV. By examining the purchasing patterns of different genders, businesses can customize their marketing strategies and product offerings to better align with the needs and preferences of each gender segment.

Table 4.4 Future prediction of females customers purchases

Table 4.5 Future prediction of Male customers purchases

For this case, we can see that male customers demonstrate a higher CLV for specific product categories, the business may allocate more resources towards marketing those products to male

customers. Additionally, businesses can devise gender-specific retention strategies informed by CLV insights. By considering the diverse revenue patterns among different cohorts and genders over time, this method yields a more data-driven CLV estimate.

In conclusion, based on CLV estimates, the overall CLV for all customers is $5123. When examining CLV by gender, females have a CLV of $5102 and males have a CLV of $5166. These insights enable us to more effectively target our customers through comprehensive and data-driven analysis. It is recommended to employ a more data-driven approach, such as cohort analysis and predictive modeling, to enhance the accuracy and actionability of CLV estimation and allocate resources more effectively towards marketing products to different genders.

Product Analysis

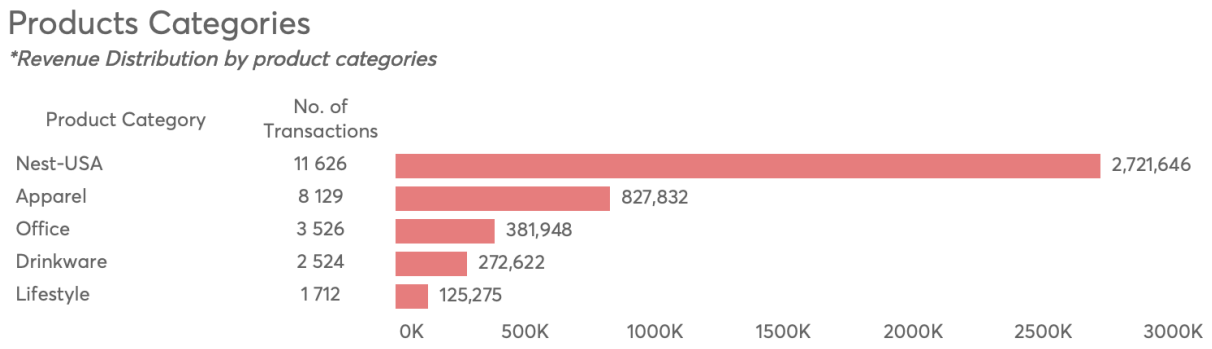
Product analysis is a crucial aspect of any business strategy as it provides valuable insights into the performance, and customer demand for various products within a company's portfolio. By conducting product analysis, businesses will gain a deeper understanding of their product offerings, identify strengths and weaknesses, and make informed decisions to optimize their product mix and marketing strategies. The analysis allows businesses to comprehend which products are driving revenue, which products may need improvement or need to reconsider them, and which products resonate most with their targeted audiences.

From here we will look more at product categories, products, Pareto Analysis (80/20 rule) and market basket analysis. As purchases per product categories, individual product performance, Pareto analysis (80/20 rule), and market basket analysis are integral components of product analysis. Purchases per product categories help identify the most revenue generated product categories, while individual product performance analysis sheds light on the success of specific products. Pareto analysis highlights the vital few products that generate the majority of revenue, enabling businesses to focus their resources on the key products. Market basket analysis reveals patterns of product co-purchases, uncovering cross-selling opportunities and informing merchandising strategies. Overall, product analysis plays a fundamental role in driving strategic decision-making, optimizing product offerings, and maximizing overall business performance.

1. Product Purchases

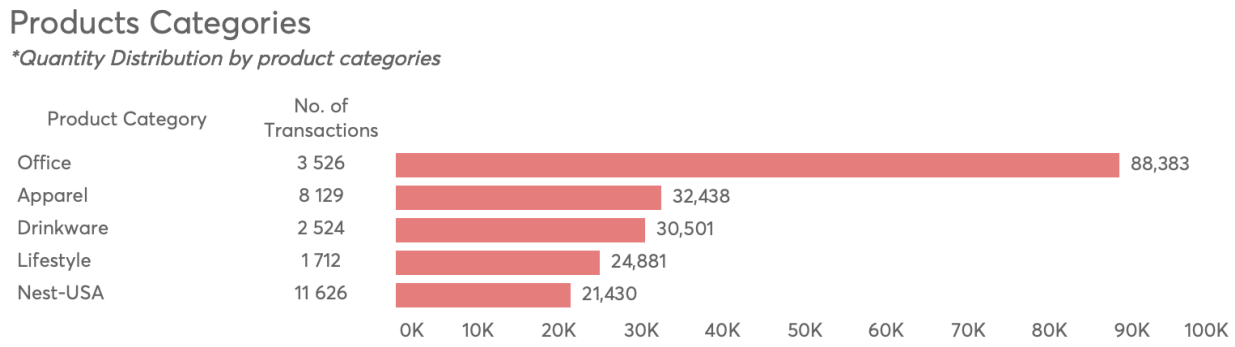
Product sales analysis involves examining and evaluating the sales performance of various products within a business. This analysis assists businesses in understanding which products and product categories are performing well, identifying areas for improvement, and making data-driven decisions for future product development, marketing, and pricing strategies. The objective is to maximize sales, and customer satisfaction. Sales by product category is a metric that indicates the total sales for each product category, aiding businesses in identifying their most

popular categories. Understanding the top-performing categories can guide marketing efforts and product development strategies effectively.



5.1.1 Top 5 Product Categories by Revenue Distribution

The product categories generating the most revenue include Nest-USA, Apparel, Office, Drinkware, and Lifestyle. These categories appear to generate the highest revenue, suggesting that customers are willing to spend more on these types of product categories. This insight can guide marketing and inventory decisions, as focusing on these high-revenue categories may lead to increased overall revenue and customer satisfaction.

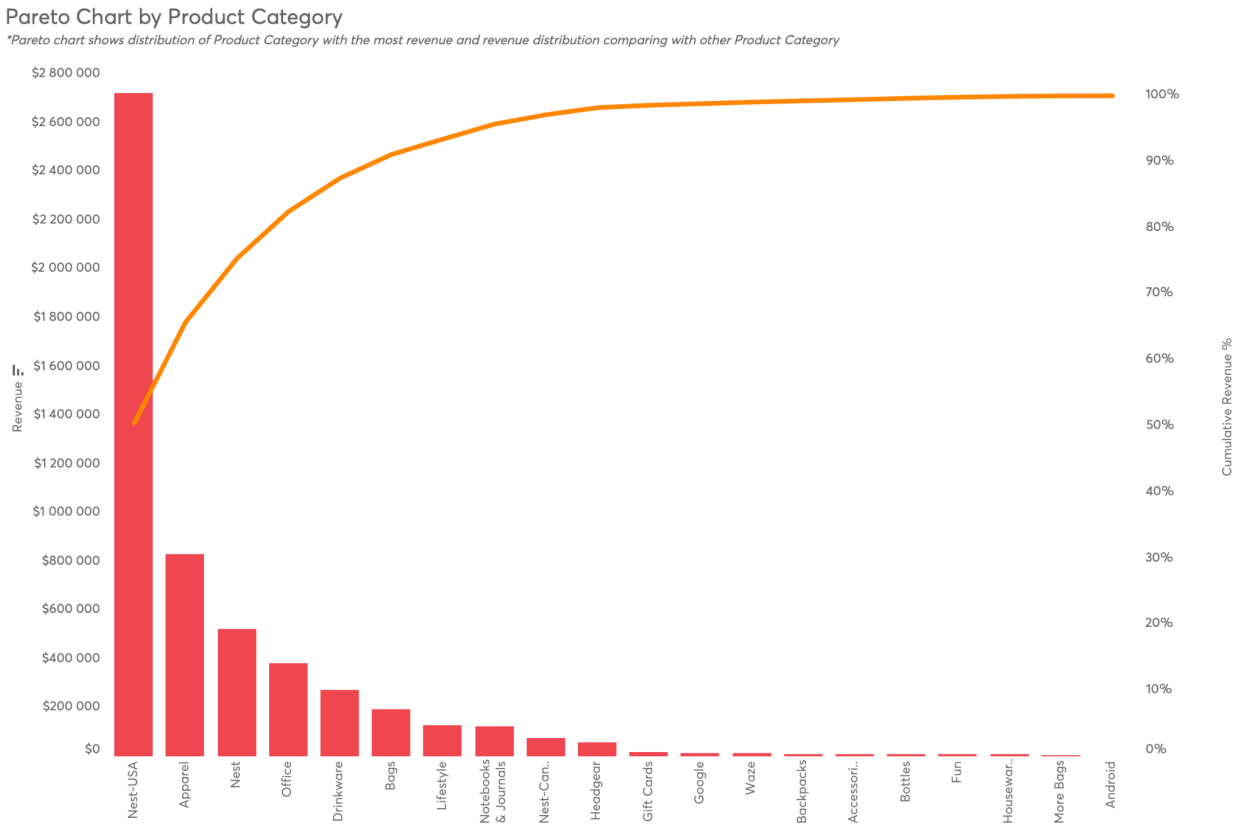


5.1.2 Top 5 Product Categories by Quantity Distribution

Most popular categories: The office category appears to be the most popular among customers, likely indicating high demand for these products. While this category may not generate the most revenue, it still ranks among our top 5 categories. Its popularity suggests that it could be a key driver of customer acquisition and retention. E-commerce should consider optimizing its product assortment, pricing, and promotions within this category to capitalize on its popularity and potentially increase revenue.

2. Pareto Analysis

Based on the Pareto Principle or the 80/20 rule, is an essential tool in product analysis that helps businesses identify the vital few products that contribute the majority of revenue or sales. This analysis is crucial because it allows businesses to focus their resources and efforts on the most impactful products, thereby maximizing revenue and efficiency. By conducting Pareto analysis from a product perspective, the e-commerce can understand which products are the main drivers of revenue and which products may require further attention or optimization. This insight enables strategic decision-making regarding product development, marketing strategies, inventory management, and resource allocation, ultimately leading to improved profitability and overall business performance.



5.2.1 Pareto chart by product categories

Delving into the Pareto analysis reveals that four main product categories - Nest-USA, Apparel, Nest, and Office collectively generate approximately 83% of total revenue. This insight underscores the significance of these categories in driving revenue and guides strategies for further improvement. By focusing efforts on optimizing these categories, businesses can enhance revenue, streamline inventory management, and tailor marketing strategies to maximize customer engagement and retention.

But let's delve deeper from the product perspective: how can Pareto analysis provide insights when viewed through the lens of individual products

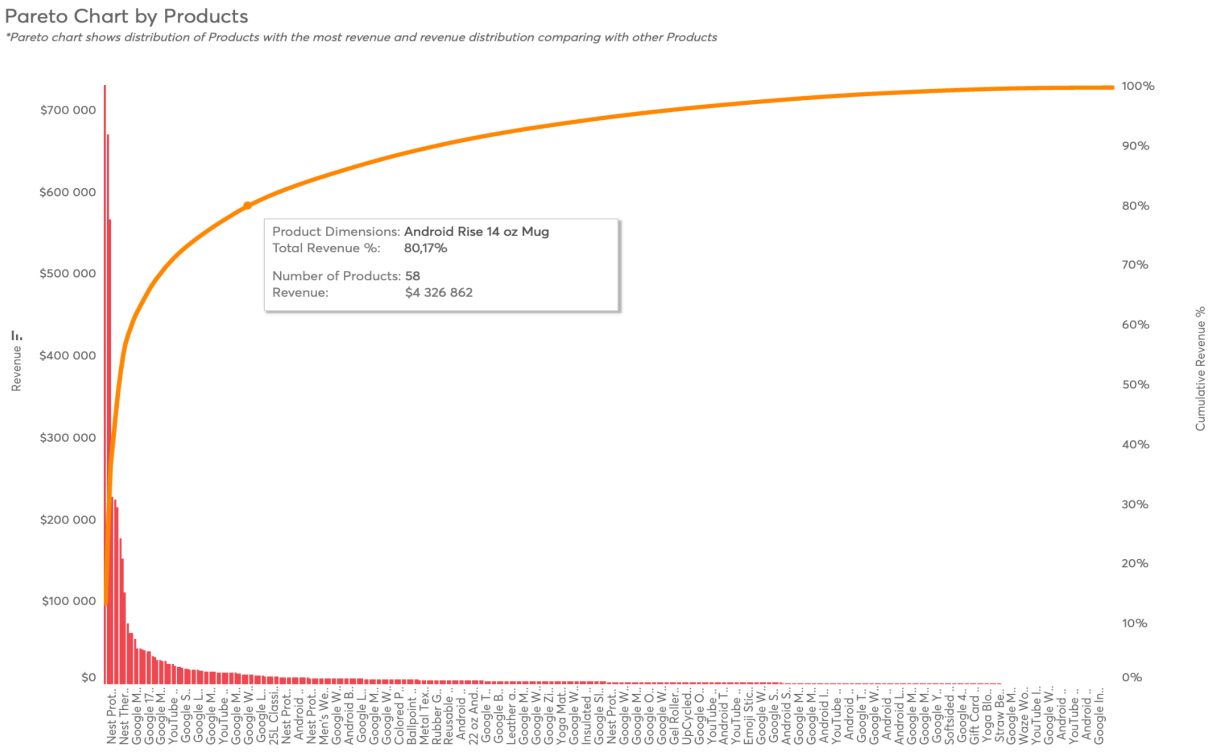


Table 5.2.2 Pareto chart by products

Looking at the insights from a product perspective, it's evident that out of 404 different products initially, Pareto analysis highlights that only 58 products, roughly 15% of the total, generate over 80% of all revenue, amounting to approximately $4.3 million. This emphasizes the significant contribution of a select few products to the business's overall revenue. Understanding this distribution helps better focus on inventory management, prioritize marketing efforts, and allocate resources effectively. It also guides product development strategies, enabling the business to enhance existing products, introduce new ones aligned with revenue goals and customer preferences, and make decisions regarding underperforming products.

3. Market Basket Analysis

Market basket analysis is an important tool for e-commerce businesses as it allows them to understand customer purchasing patterns and relationships between products. By analyzing the contents of customers' shopping baskets, e-commerce can uncover which products are commonly purchased together. This insight is invaluable for optimizing product recommendations, cross-selling strategies, and promotional campaigns. Having a deeper knowledge of these associations enables businesses to personalize the shopping experience, improve product bundling strategies, and increase average order value. Moreover, market basket

analysis provides insights into customer preferences, helping businesses anticipate demand and adjust inventory accordingly. Overall, leveraging market basket analysis empowers e-commerce businesses to enhance customer satisfaction, drive sales, and maximize revenue by delivering targeted and relevant product offerings to their customers.

Market basket

Product Combinations: 2 Product Combination

**Market basket by most valueable product by their unit price in 2 Product Combinations*

Rank	Product Basket	Occurrence	% of Occurence
1	Nest Cam Indoor Security Camera - USA -> Nest Cam Outdoor Security Camera - USA	459.0	9,01%
2	Nest Cam Outdoor Security Camera - USA -> Nest Learning Thermostat 3rd Gen-USA - Stainless Steel	155.0	3,04%
3	Nest Protect Smoke + CO White Battery Alarm-USA -> Nest Learning Thermostat 3rd Gen-USA - Stainless Steel	127.0	2,49%
4	Nest Protect Smoke + CO White Wired Alarm-USA -> Nest Learning Thermostat 3rd Gen-USA - Stainless Steel	113.0	2,22%
5	Nest Cam Indoor Security Camera - USA -> Nest Learning Thermostat 3rd Gen-USA - Stainless Steel	109.0	2,14%
6	Nest Protect Smoke + CO White Battery Alarm-USA -> Nest Cam Outdoor Security Camera - USA	68.0	1,33%
7	Nest Protect Smoke + CO White Battery Alarm-USA -> Nest Protect Smoke + CO White Wired Alarm-USA	63.0	1,24%
8	Nest Protect Smoke + CO White Battery Alarm-USA -> Nest Learning Thermostat 3rd Gen-USA - White	52.0	1,02%
9	Nest Cam Outdoor Security Camera - USA -> Nest Learning Thermostat 3rd Gen-USA - White	50.0	0,98%
10	Nest Protect Smoke + CO White Wired Alarm-USA -> Nest Cam Outdoor Security Camera - USA	49.0	0,96%

Table 5.3.1 Market Basket analysis with two product combinations

Through market basket analysis, analyzing combinations of two main products by unit price per transaction in order to see most valuable items reveals valuable insights. For example, we observe that customers who purchase the 'Nest Cam Indoor Security Camera - USA' also tend to buy the 'Nest Cam Outdoor Security Camera - USA,' with 459 occurrences, representing 9% of all occurrences involving the purchase of two products.

Upon deeper examination of three-product combinations, it becomes apparent that one of the frequently chosen products is "Google Sunglasses" -> "Goggle Sunglasses" -> "Google Sunglasses," occurring 112 times, representing nearly 2% of occurrences in three-product combinations. From this data, it's evident that obtaining additional information about the products (such as sizes, colors, etc.) is essential for gaining a better understanding of product variations at deeper levels. This information is crucial for accurately managing inventory, offering the right products to our customers, and adjusting our marketing campaigns accordingly.

Market basket

**Market basket by most valueable product by their unit price in 3 Product Combinations*

Product Combinations: 3 Product Combination

Rank	Product Basket	Occurrence	% of Occurence
1	Google Sunglasses -> Google Sunglasses -> Google Sunglasses	112.0	1,99%
2	Nest Cam Indoor Security Camera - USA -> Nest Cam Outdoor Security Camera - USA -> Nest Learning Thermostat 3rd Gen-USA - Stainless Steel	54.0	0,96%
3	Nest Protect Smoke + CO White Battery Alarm-USA -> Nest Cam Indoor Security Camera - USA -> Nest Cam Outdoor Security Camera - USA	41.0	0,73%
4	Nest Protect Smoke + CO White Wired Alarm-USA -> Nest Cam Indoor Security Camera - USA -> Nest Cam Outdoor Security Camera - USA	21.0	0,37%
5	Nest Protect Smoke + CO White Battery Alarm-USA -> Nest Cam Outdoor Security Camera - USA -> Nest Learning Thermostat 3rd Gen-USA - Stainless Steel	20.0	0,36%
6	Nest Protect Smoke + CO White Wired Alarm-USA -> Nest Cam Outdoor Security Camera - USA -> Nest Learning Thermostat 3rd Gen-USA - Stainless Steel	19.0	0,34%
7	BLM Sweatshirt -> BLM Sweatshirt -> BLM Sweatshirt	16.0	0,28%
8	Nest Protect Smoke + CO White Battery Alarm-USA -> Nest Cam Indoor Security Camera - USA -> Nest Learning Thermostat 3rd Gen-USA - Stainless Steel	15.0	0,27%
9	Nest Cam Indoor Security Camera - USA -> Nest Cam Outdoor Security Camera - USA -> Nest Learning Thermostat 3rd Gen-USA - White	15.0	0,27%
10	22 oz Android Bottle -> Google 22 oz Water Bottle -> Google 22 oz Water Bottle	13.0	0,23%

Table 5.3.2 Market Basket analysis with three product combinations

Market basket analysis is crucial for e-commerce businesses as it provides valuable insights into customer purchasing patterns and relationships between products. By understanding the correlation between products, businesses can optimize product recommendations and implement effective cross-selling and up-selling strategies, ultimately enhancing revenue and customer satisfaction. By suggesting complementary products when a customer adds one item to their cart, e-commerce businesses not only increase sales and revenue but also improve the overall customer experience. These insights are essential for businesses to enhance the customer shopping journey, boost average order value, and drive sales. Leveraging market basket analysis allows businesses to tailor their product offerings, inventory management, marketing campaigns and maximize revenue by delivering targeted and relevant product combinations to their customers. Overall, market basket analysis is indispensable for e-commerce businesses looking to optimize their operations and provide a seamless and personalized shopping experience for their customers.

Conclusions

In conclusion, this comprehensive analysis underscores the importance of understanding customer behavior and preferences to drive growth and enhance customer satisfaction. With 27% of revenue attributed to 'Top Customers' and 'Loyal Customers', it's crucial to not overlook other customer segments, especially the 20% categorized as "At-Risk" and other specific segments

needing attention. The company should focus on retention and re-engagement strategies to address these segments. Analyzing Customer Lifetime Value, sales trends, and marketing spendings provides valuable insights into the business's overall performance. Additionally, product analysis has provided a better understanding of our main product offerings and customer demands, highlighting which products and categories require more attention and which ones are underperforming. Market basket analysis has shed light on customers' buying behavior, offering insights into cross-selling and up-selling opportunities. By leveraging these insights and understanding the factors contributing to these metrics, along with customer segmentation and product buying behaviors, the company can optimize its marketing and inventory management efforts to maximize profitability and customer satisfaction.

Recommendations

Based on the analysis results the following recommendation can be made:

Focus on high-value segments: Prioritize marketing and retention efforts on Loyal Customers, Potential Loyalists, as they contribute significantly to the company's revenue 58% ($3.1 million). Tailor marketing messages and offers specifically to these high-value segments to maximize their engagement and lifetime value.

Address the "At-Risk", "Can't Loose Them", and "Customers Needing Attention" segments: These segments account for roughly 31% ($1.68 million) of revenue. Prior to refining our targeting for this segment, it's important to engage with these customers through surveys, feedback forms, or one-on-one conversations to understand their concerns. Once their situation is understood, we can offer better solutions to encourage sustained engagement, highlighting the benefits of loyalty to our brand.

Target Female Customers: Given that female customers contribute more to revenue than male customers, concentrate marketing efforts on engaging female customers. Develop targeted marketing campaigns and promotions that align with their preferences and needs.

Product Portfolio Optimization: Use insights from the Pareto analysis to optimize the product portfolio. Consider phasing out or reevaluating low-performing products to focus resources on those that drive the most significant revenue and align with customer preferences and market demand.

Leverage Most Popular Products and Revenue-Generating Product Categories: Capitalize on the popularity and revenue generated by Nest, Apparel, Office, and Drinkware, as these products are highly demanded by customers. Develop marketing campaigns and promotional offers around these product categories to boost sales and customer engagement.

Enhanced Product Recommendations: Utilize the insights from two-product and three-product combinations to enhance product recommendations for customers. Implement algorithms and personalized recommendation systems that take into account the observed patterns of product associations to suggest related items during the purchasing process, thereby improving upselling opportunities and increasing revenue.