1.0 Introduction

1.1 Background

The continuous growth of electronic commerce, also known as e-commerce, has generated great interest in the studies of online consumer behavior as highlighted by Nalchigar et al. (2016). Therefore, a proper method to study and visualize the complex consumer trend in an ever-increasing digital marketplace is needed to tackle some business problem statements.

1.2 Research scope

Our research scope is confined to an e-commerce customer behavior dataset found on Kaggle. It comprises five main datasets, which are labeled as “Customer data”, “Discount Coupon”, “Marketing Spend”, “Online Sales”, and “Tax Amount”. Our data analytics, visualizations, and insights will only revolve around these datasets.

1.3 Problem statement

- How can we optimize customer segmentation strategy to improve customer growth?

In understanding the customer domain problems, we reference a study by Kim et al. (2009) that emphasizes the role of e-loyalty, e-satisfaction, and e-trust in the development of online loyalty in highlighting the importance of customer-centric factors in e-commerce. Furthermore, the work of Qiu et al. (2015) focuses on the need to predict customer purchase behavior in the e-commerce context, giving basic knowledge on the significance of understanding customer preferences and buying habits. In addition to that, there are some studies highlighting the significance of RFM analysis in assessing customer lifetime value and segmenting customer behavior Fader et al. (2005) Liu & Shih, 2005; Gustriansyah et al., 2020). These studies emphasize the importance of understanding customer purchasing patterns and the value of customers to the business.

- How can we optimize marketing strategy to improve revenue?

For the second problem statement, we attempt to optimize marketing strategy to improve revenue. The study by (Salamai et al., 2022) attempts to link the impact of short-term product marketing goals like promotion, and seasonal factors on e-commerce sales performance and also highlights the importance of forecasting and understanding the lifecycle of e-commerce platforms during seasonal sales. In addition to that, Stephen & Galak (2012) found that both traditional and social marketing can affect sales, with social earned media's sales elasticity being significantly greater than traditional earned media. This could suggest that social media marketing can have a better impact on sales compared to traditional media.

- How can we improve product performance to optimize supply and demand?

The third problem statement focuses on improving product performance and optimizing supply and demand of e-commerce items. Research by Yustiana et al. (2021) highlights the usage of market basket analysis products by determining what products customers will buy simultaneously by analyzing customer transactions. From the insights, we can identify items cross-selling opportunities and understand customer purchasing patterns.

1.4 Project Aim

In this project, the aim is to use PowerBI as a visualization tool for the chosen e-commerce Customer Behavior Dataset to obtain insights into addressing business domain problems focusing on customer, business, and product aspects. PowerBI is a business analytics service provided by Microsoft. PowerBI is considered a powerful visualization tool (Obaid, 2023) because it enables users to analyze data through visualization to draw insights.

1.5 Objectives

The objective of this project is to provide detailed research from the insights gained from the visualization of the e-commerce customer behavior dataset.

Objective 1	In relation to Problem Statement 1, this project aims to provide insights to optimize customer growth via customer segmentation.

Objective 2	In relation to Problem Statement 2, this project aims to provide insights to optimize sales and revenue by improving the company's marketing strategy.
Objective 3	In relation to Problem Statement 3, this project aims to provide insights on product performance to optimize the company's supply and demand.

2.0 Methodology

This section outlines the meta-data of the dataset along with the data visualization created by PowerBI. The visualizations will assist in understanding the dataset in an uncomplicated manner, allowing readers to instantly extract insight with ease.

2.1 Data Source

2.1.1 Data Source

The dataset was retrieved from the website Kaggle. Rishi Kumar was the author of the dataset and 28 October 2023 was the dataset release date. Please click here to view the link of the website.

2.1.2 Data Structure

Before creating the visualizations, it is important to understand the variables (data structures) within each table. It would allow the analyst to efficiently create the necessary relationship within PowerBI while performing data cleaning if necessary.

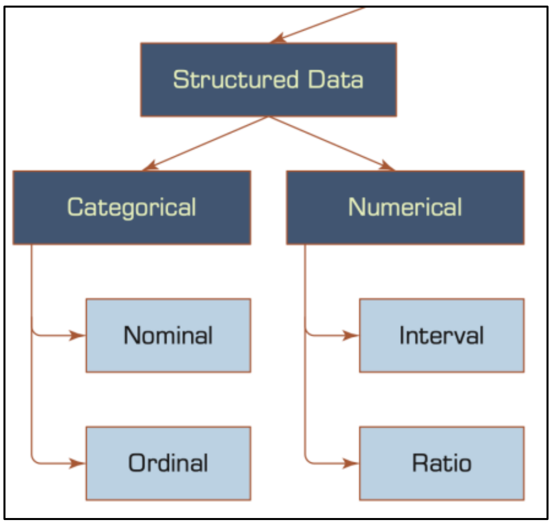

Figure 1 - Data Structure (Sharda, Delen, Turban & King, 2017, pg 87)

Upon reviewing, there are two categories for structure data.

Categorical Structure:

- Nominal Data
  - This data obeys the simplest scale of measurement which does not follow a hierarchy order within the scale. Moreover, each value within the database are a distinct value and does not overlap. (Jafari, Chen, Mirzaie & Tang, 2022)
- Ordinal Data
  - In comparison, ordinal data strictly follows a ranking system which allows the data to be categorized. An example of categorization would be low, medium, and high categories (Sharda et al, 2017, pg 87).

Numeric Structure:

- Interval Data
  - Interval data represent measurement that does not have an absolute zero point. With the absence of the measurement attribute, information could be inferred from zero point. An example would be Celsius where zero does not reflect the absence of temperature (Sharda et al, 2017, pg 87).
- Ratio Data

- In comparison, ratio data is the absence of a measurement attribute. Possessing an absolute absence of measure resulted in an absolute zero point (Sharda et al, 2017, pg 87).

The following table describes the data structure of the dataset.

Name	***Data Structure***
CustomerID	Nominal Data
Gender	Ordinal Data
Location	Ordinal Data - Density of CustomerID
Tenure_Months	Ratio Data

Table 1: Customer table

Name	***Data Structure***
Month	Ordinal Data - Month ranking
Product_Category	Nominal Data
Coupon_Code	Nominal Data
Discount_pct	Ratio Data

Table 2: Discount_Coupon

Name	***Data Structure***
Data	Ordinal Data
Offline_Spend	Ratio Data

Online_Spend	Ratio Data

Table 3: Marketing Spend

Name	***Data Structure***
CustomerID	Nominal Data
Transaction_ID	Nominal Data
transaction_Date	Ordinal Data
Product_SKU	Ordinal Data
Product_Description	Ordinal Data
Product_Category	Ordinal Data
Quantity	Ratio Data
Avg_Price	Ratio Data
Delivery_Charges	Ratio Data
Coupon_Status	Nominal Data

Table 4: Online Sales

Name	***Data Structure***
Product_Category	Ordinal Data
GST	Ratio Data

Table 5: Tax Amount

Data Structure	***Number***

Nominal Data	6
Ordinal Data	9
Ratio Data	8
Interval Data	0

Table 6: Summary of all data structure

2.1.3 Data Cleaning & Transformation

Upon inspecting the dataset, there are no missing values or NaN values found, thus requiring no data-cleaning process to be done. To solve the problem statements, the following datasets needed to be transformed. In return, it would benefit in terms of extracting insights in a digestible manner to assist in achieving the vision and mission of the project. The data transformation (measure & calculated column) would be categorized into three parts:

1. Customer Segmentation
    a. RFM Table (Calculated Column)
        i. R, F & M Score
        ii. R, F & M Value
    b. Online Sales (2)
        i. Last transaction date - created a measure by using Max Function
        ii. R, F & M value, created a measure by using Distinct Count
        iii. Total Quantity - created a measure by using the SUMX function to calculate individual sum of quantity.
        iv. Total_sale_per_customer - created a measure to calculate SUMX of the Avg_price multiplied by Quantity.
2. Sales & Marketing
    a. Return on Investment (ROI) - created a measure that divides revenue by marketing spend.
3. Customer Lifetime Value (LTV)
    a. Average Purchase Value - created a measure that divides the total sales per customer by the total number of orders.

b. Average Purchase Frequency - created a measure that divides the number of orders by the total number of customers.
c. Customer Lifetime Value - created a measure that multiplies average purchase value with average purchase frequency.

4. Market Basket Analysis
    a. Product Table - created using Excel by compiling all columns related to the product as a preparation for the Basket Analysis Table
    b. Basket Analysis Table
        i. Product Item & Product Item 2 - using the DAX formula to create one row for each couple of products items.
        ii. Basket - created a calculated column with the concatenation of the two items.
        iii. Support of the basket - using the DAX formula to calculate the percentage of the transactions containing the two specific product items.
        iv. Confidence items 1 & 2 - using the DAX formula to calculate the Support of the basket divided by the percentage of transactions with only one product.
        v. Lift - using the DAX formula to calculate the Support of the basket divided by the percentage of times product 1 is bought, multiplied by the percentage of times product 2 is bought.

2.2 Data Visualization

2.2.1 Customer Segmentation

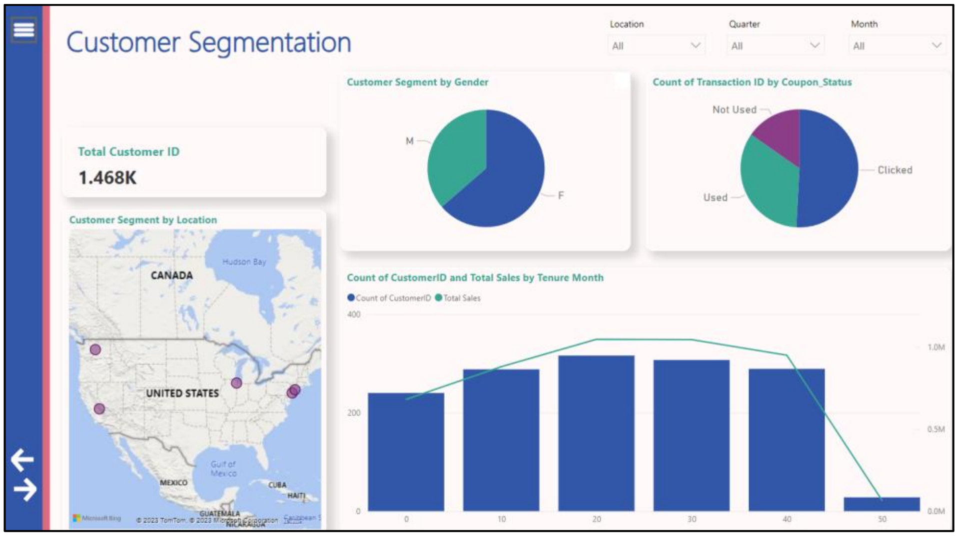

Figure 2 - PowerBI Dashboard for Customer Segmentation

Figure 2 shows the PowerBI Dashboard for Basic Customer Segmentation. It includes metrics such as:

1. Pie Chart of gender
2. Total Count of Customer ID
3. Pie Chart of coupon status
4. Location of the client
5. Histogram with relation to total sales by tenure month (Bin)

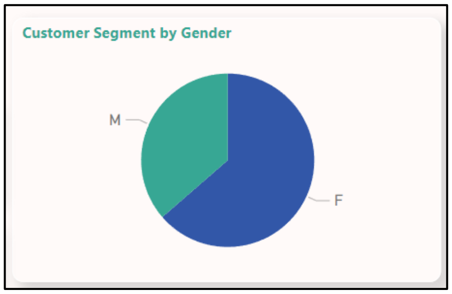

Figure 3 - Gender Segmentation

Descriptive Statistics:

- 64% of the users are female.
- 36% of the users are male.

Insights:

- As our users are a majority female, it could be inferred that females are more likely to conduct online purchases in comparison to the male gender.
- With limited customer data available, it could be inferred that the male user within the e-commerce platform would tend to display stronger shopping behavior than the average male.

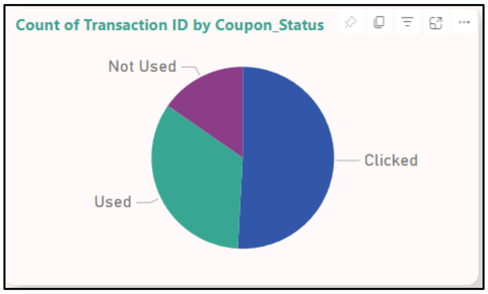

Figure 4 - Segmentation by coupon status

Descriptive Statistics:

- 51% of the users clicked the coupon.
- 34% of users used the coupon.
- The remaining 15% of users purchase without utilizing the coupon.

Insights:

- As seen on the pie chart, it could be inferred that the users are attracted to the platform by the concept of coupons.
- The following assumption is that the users within the database are regular shoppers therefore relying on coupons to chase a better price.

Figure 5 - Customer location by State

Descriptive Statistics:

- The users that use e-commerce platforms are located in four states within the United States.
  - New York
  - New Jersey
  - Chicago
  - California
  - Washington

Insights:

- As our users mainly are located in these four states, it could be assumed that the demographic within the respective states would be interested in e-commerce platforms.
- Therefore, it draws an educated conclusion that demographics could be an important element for an e-commerce business.
- Moreover, the cities mentioned are metropolitan with a significant commercial presence within the environment. Therefore, it could lead users to have a larger trust towards a reputable e-commerce platform.

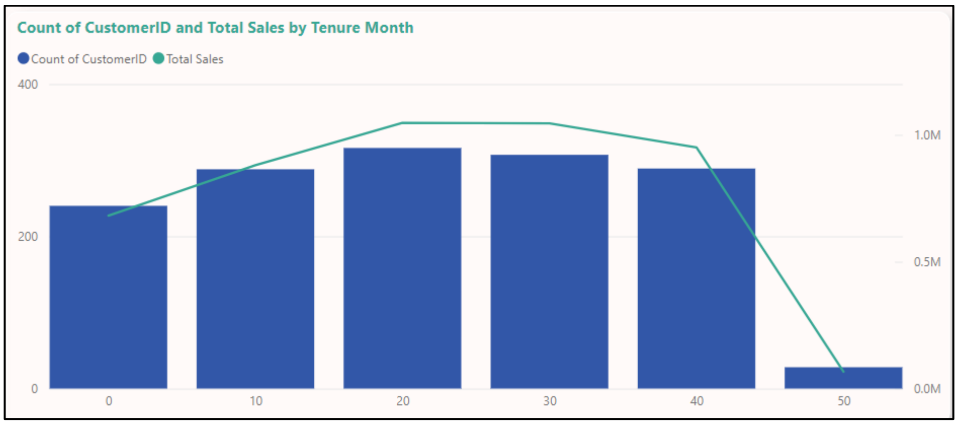

Figure 6 - Histogram in relation to Tenure Month and Count of customer ID

Descriptive Statistics:

- The line stack column chart displays that users with tenure below 50 months are consistently above the 200 counts of Customer ID.
- Users who exceed the 50-tenure month threshold hold a significantly low proportion of the overall platform users.
- The line chart represents the total sales based on tenure month. It displayed that the Sum of Customer ID has a positive correlation with Total Sales.
- Especially on the tenure month bins of 20,30 & 40. It is seen that the total sales exceed the height of the column.

Insights:

- Based on the column and the line chart of the total sales the analyst assumes that users with less than 9 tenure months tend to purchase a conservative amount. It could reflect that the users were in the initial phase of testing the platform.
- From the 10-tenure month bin onwards, it could be seen that users are starting to be familiar with the platform therefore reflecting a slightly higher average spend per customer.
- The tenure month bin of 20 displays the peak in terms of CustomerID and total sales. The analyst drew a conclusion that it would take an estimated 20 months for a customer to be familiar with the platform before engaging in more purchases or larger purchases in terms of monetary value.
- The bins of 30 and 40 tenure months reflect the number of users starting to reduce. It seems that the e-commerce platform could be facing intensified competition within the industry. Therefore, some users might be poached by competitors.
- On the bin of the tenure month of 50, the amount of “customerID” dropped significantly. It would be difficult to pinpoint the cause, however it could be speculated that the e-commerce platform may lack innovation to continuously engage the user.

2.2.2 RFM analysis

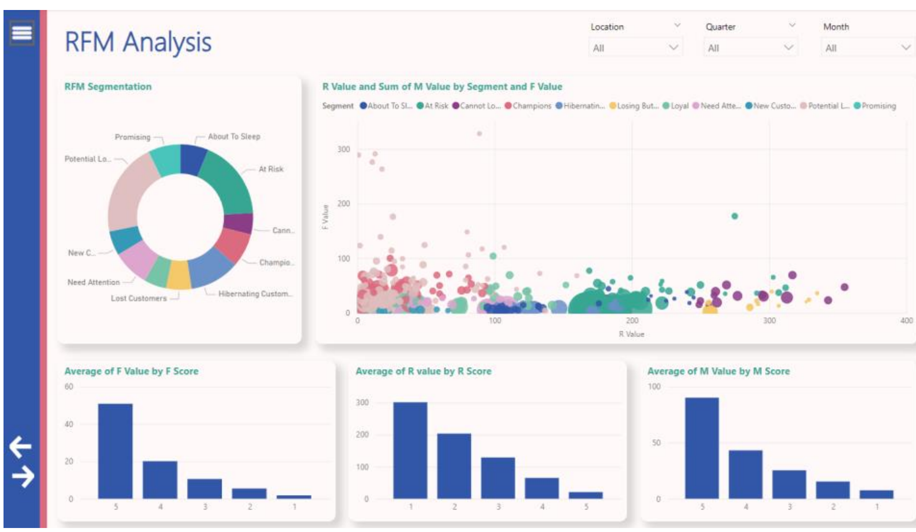

Figure 7 - PowerBI Dashboard for RFM analysis

Figure 7 shows the PowerBI Dashboard for Basic Customer Segmentation. It includes metrics such as:

1. Pie Chart of RFM Segmentation
2. Scatter plot for R, F & M value
3. Stacked column chart for R, F & M score

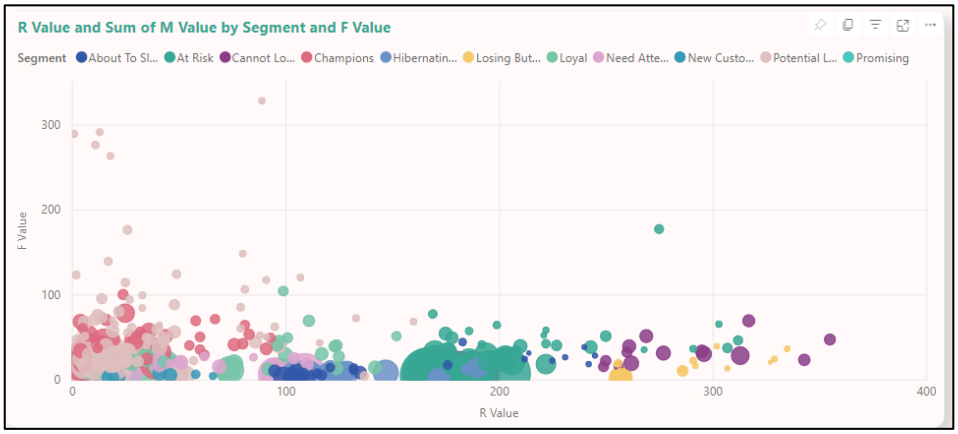

Figure 8 - RFM Analysis (Scatter Plot)

Descriptive Statistics:

- The RFM (Recency, Frequency, and Monetary) analysis displays the relationship between three variables within the scatter plot.
- The RFM scatter plot exhibits high density for 'Potential Loyalist' users which exhibit F higher F value and R value less than 50 days.
- Within the R-value range of 150 to 210 days, it is seen that a group of customers is labeled as 'At Risk'.

Insights:

- As seen on the left side of the scatter plot, our top users typically exhibit behavior such as F value (purchase frequency) hovering around the 20 - 80 range along with less than 100 of R-value (purchase recency) which indicates repeated purchase along with a short interval between purchases.

- Users exceeding the 100 value (R-value) highlights that the user tends to exhibit a low F value. Therefore, highlighting that these users could be one-time users or low-frequency purchase users.

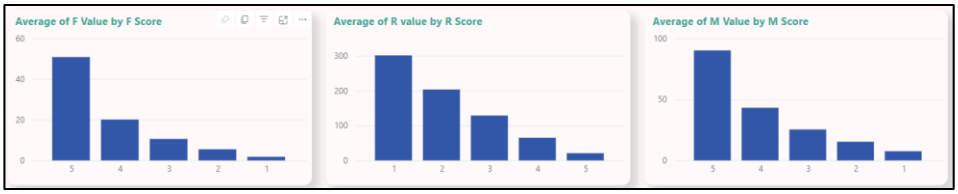

Figure 9 - Stacked Column chart in relation to F, R&M Score

Descriptive Statistics:

- The method of calculating would be based on the percentile of the transaction ID. The DAX formula will create a bracket of 25% of the transaction within the data.
- The F Score shows the top 25% of purchases were repeated purchases. The average repeated purchase for each user would be 50.81 repeat purchase.
- On the R Score, it displays that most of the users tend to exhibit low scores. Informing the analyst that users have probability that users will repeat a purchase within 300 days.
- On the contrary, the M Score displays that the average user spends an estimate of $90.10. In addition, the column is the highest therefore suggesting that most of the users tend to spend around the $90.10 range.

Insights:

- It could be inferred that the platform has a strong reason to believe that a large proportion of users are on the lower spectrum in terms of duration of purchase interval.
- Furthermore, it suggests that each user would have a higher probability of repeating a purchase.
- As seen on the coupon pie chart mentioned earlier, there could be a probability that the coupon release by e-commerce could be a factor affecting the R-value of the users.

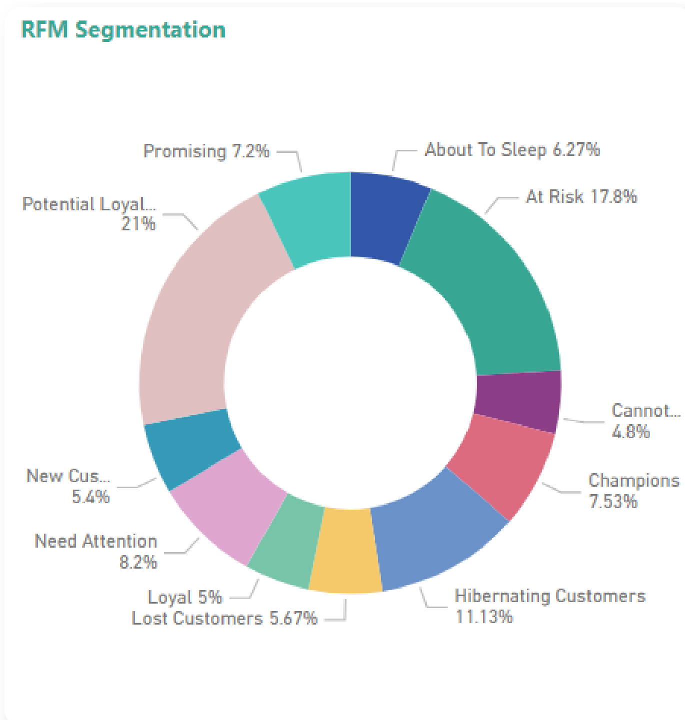

Figure 10 - Pie Chart by RFM segmentation

Descriptive Statistics:

- 21% of the users within the platform qualify as 'Potential Loyalists' while only 7.53% of the users are 'Champions'. Moreover, only 7.2% of the users qualify as 'Promising' users, and 5% of the users are labeled as 'Loyal'.
- On the contrary, 26% of the users are 'At Risk', and 11.13% of users are segmented as 'Hibernating Customers'.Furthermore, within the platform, 7.53% of users are labeled as 'Lost Customers'.

Insights:

- From a quick observation of the donut chart, it could be viewed that a significant number of users would be in the process of transitioning into 'At Risk' and 'Hibernating

Customers'. Thus, if the e-commerce platform continues with its current strategy. It would be a possibility that it would lose out on a significant number of users.

__2.2.3 Marketing Performance__

Figure 11 - PowerBI Dashboard for Sales & Marketing Analysis

Figure 10 shows the PowerBI Dashboard for Sales & Marketing Analysis. It focuses on the relationship between marketing spend and total sales in 2019. The dashboard includes key metrics such as:

1. Total offline spend: Channels such as TV, radio, newspapers, and others.
2. Total online spend: Channels such as Facebook, Instagram, Google Ads, and others.
3. Total marketing spends: The sum of total offline and online spend.
4. Total sales amount: The sum of revenue gained in 2019.
5. ROI meter: Displaying the return on investment from the selected period.

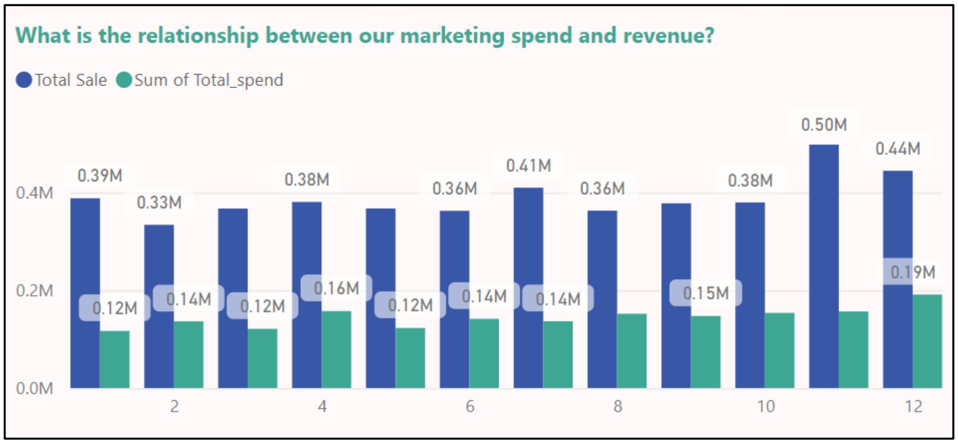

Figure 12 - Total Sales with Total Marketing Spend by Month (2019)

Descriptive Statistics:

- It could be seen that the lowest total marketing spend was $120K and the highest would be $160K.
- It could be seen that in November, it resulted in the highest sales (500K) despite the total marketing spend being the second highest.

Insights:

- In November, it could be inferred that the total sales were the highest due to the influences of the festive season and the positive impact on consumer behavior (Getter & Behe, 2013; Waturandang et al., 2020).
- Based on figure 11, it could be seen that the total marketing spend ranges from $120k and $190k, gradually increasing from May to December.

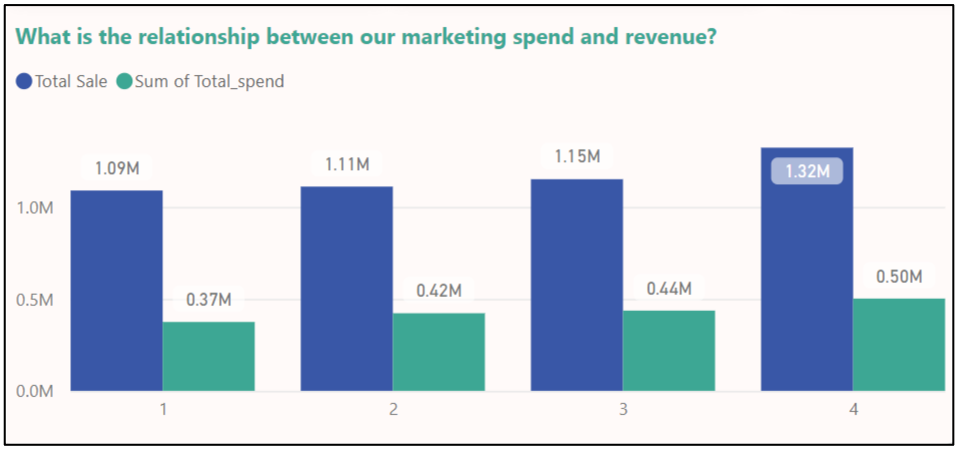

Figure 13 - Total Sales with Total Marketing Spend by Quarter (Drill-up to quarterly)

Descriptive Statistics:

- It can be seen that Q4 has the highest marketing spend of $500K.
- The marketing budget for the remaining quarter remained consistent at $400K.

Insights:

- It can be seen from a quarterly perspective that Q4 had the highest total sales. With the marketing budget increased by 20%. The e-commerce company yielded the best sales within Q4 of $1.3M.

Based on the information provided by the e-commerce platform, the analyst introduced a metric that measures the ROI value based on the revenue divided by total marketing spend.

Period	**Revenue**	**ROI value**
Q1	$1.1m	3.56
Q2	$1.1m	2.93
Q3	$1.2m	2.75

Q4	$1.3m	2.74

Table 7

Descriptive Statistics:

- Based on the ROI value metric, the value decreases gradually over the course of 2019 despite an increase in sales.

Insights:

- Even though the revenue was the highest in Q4, the ROI value was at the lowest point at 2.74. This could be due to the increase of competitors penetrating the market to compete for multiple huge festive seasons such as Black Friday, Thanksgiving, and Christmas all happening in Q4.
- Thus, the e-commerce platform must deploy an insightful strategy to optimize the ROI value while seeking an increase in revenue.

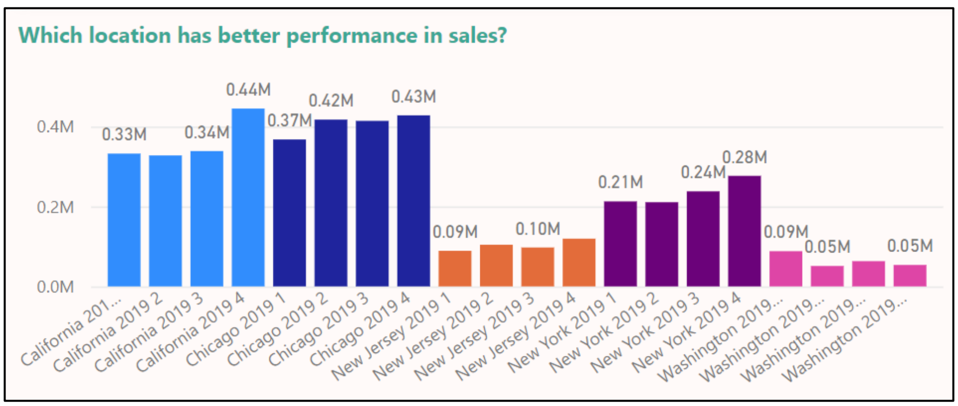

Figure 14 - Total Quarterly Sales based on Location.

Descriptive Statistics:

- As seen in Figure 14, California had the highest sales in Q4, 2019. While Washington Q2, 2019 performed the worst.

Insights:

- It could be seen that California and Chicago contributed the majority of the total sales throughout every quarter.
- Interestingly, most of the locations consistently display that Q4 had the highest total sales throughout 2019. Washington is proven to be an exception where it had the highest sales in Q1 instead.

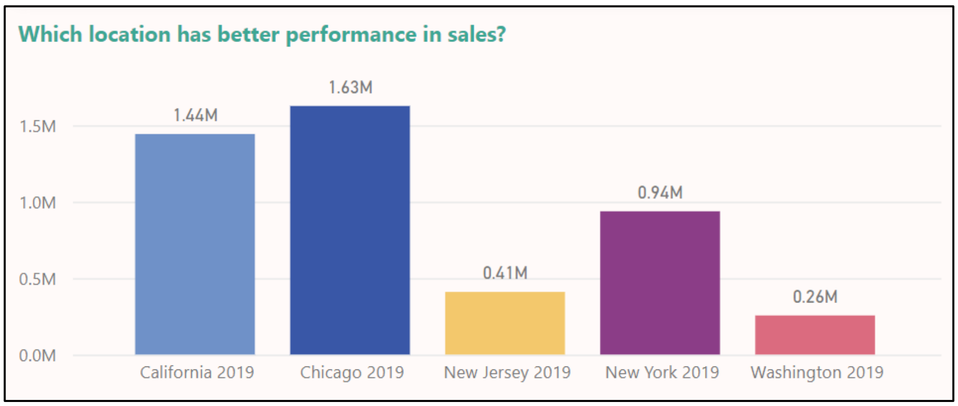

Figure 15 - Total Sales by Location by Year (Drill up)

Descriptive Statistics:

- From an annual sales perspective, Chicago achieved a total sale of $1.63M, followed by California ($1.44M), New York ($0.94M), New Jersey ($0.41M), and Washington ($0.26M) at the lowest.

Insights:

- With Chicago and California exceeding $1 Million in sales during 2019, it is recommended to allocate a larger proportion budget towards these two locations. Anchoring strong brand presence within the prime location to solidify the e-commerce platform brand positioning.

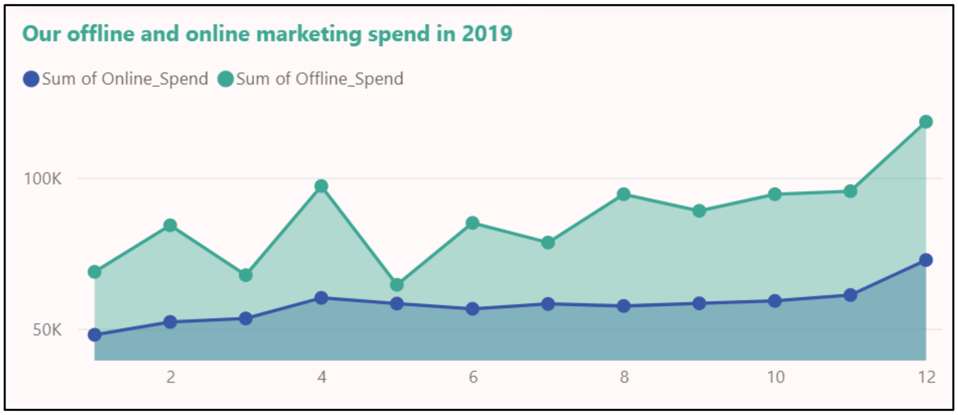

Figure 16 - Spend on online & offline marketing spend by month

Descriptive Statistics:

- Based on the area graph, it would be seen that the e-commerce platform emphasizes offline marketing over online marketing.
- During December, the e-commerce platform had the largest expenditure on marketing. Resulting in $118K being spent on offline marketing while online marketing only spent a total of $72K.
- The second largest expenditure on marketing was in April, with offline and marketing spending at $97k and $60k respectively.

Insights:

- It could be seen that the e-commerce platform strategy implemented a marketing strategy that heavily relies on offline marketing, such as TV, radio channels, and newspapers.
- However, it could be seen based on Table 7 that the ROI value decreases by the passing quarter. Therefore, it could be viewed that offline marketing was ineffective towards e-commerce platform sales.

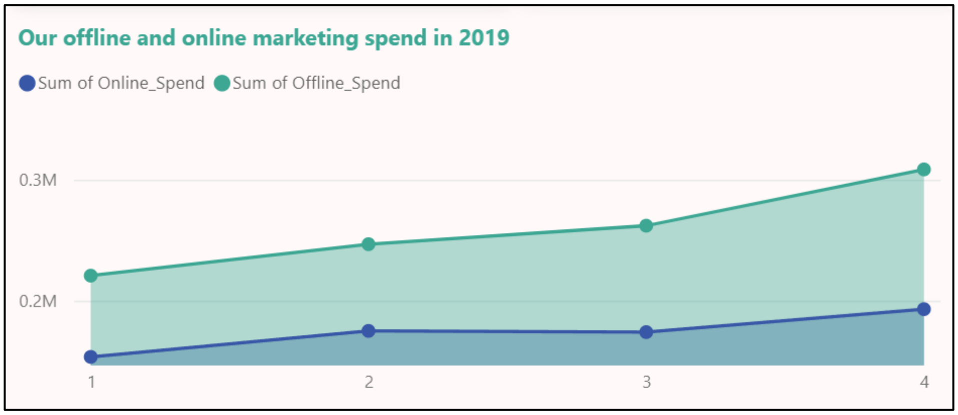

Figure 17 - Online & offline marketing spend by quarter (Drill-up)

Descriptive Statistics:

- It could be seen that offline marketing increased from $220K (Q1) to $308K (Q4). A significant increase of 40%.
- While online marketing increased by 26% from $153k (Q1) to $193k (Q4).

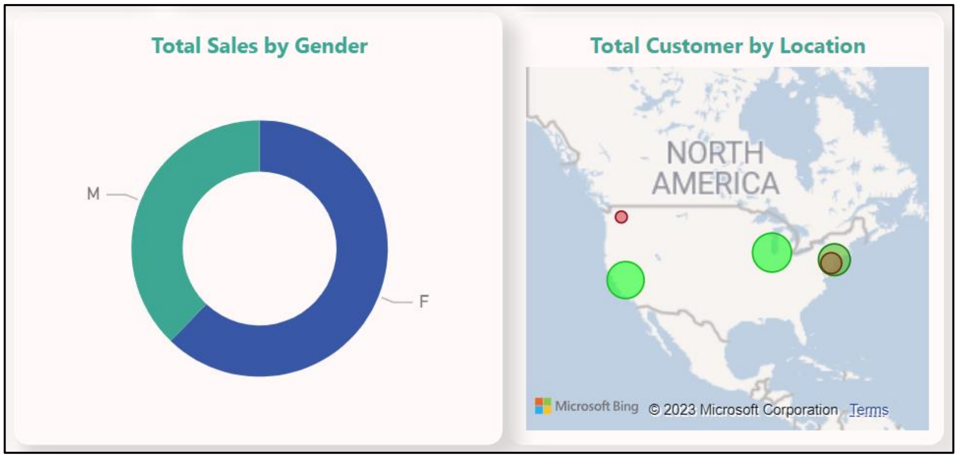

Figure 18 - Interactive chart between gender segmentation (Donut Chart) and location

Descriptive Statistics:

- In 2019, it is observed that $2.7M sales were generated by female customers (62% of the users). While male users only contributed to the remaining 37% of the sales at $1.77M.
- It could be seen on the map that the two brightest bubbles in green belong to Chicago and California. The dark green shade bubble belongs to New York, while the remaining greenish brown and red bubbles belong to New Jersey and Washington.

Period	**Revenue by Female**	**Revenue by Male**
Q1	$619k (57%)	$469k (43%)
Q2	$748k (68%)	$361k (32%)
Q3	$676k (59%)	$474k (41%)
Q4	$860k (65%)	$461k (35%)

Table 8 - Displayed the sales generated by gender by quarter.

Descriptive Statistics:

- Within Q2, female users had the largest contribution towards sales. Achieving 68% while male users only make up 32% of the sales contribution.
- In terms of sales volume, the female users almost achieved $860K during Q4.

Insights:

- With female users contributing a large volume of sales, the e-commerce platform could feature more female-centric products to capitalize on higher revenue.
- Moreover, personalizing the market basket analysis towards female-centric products could prove to be an effective method to drive higher sales.

2.2.4 Customer Lifetime Value

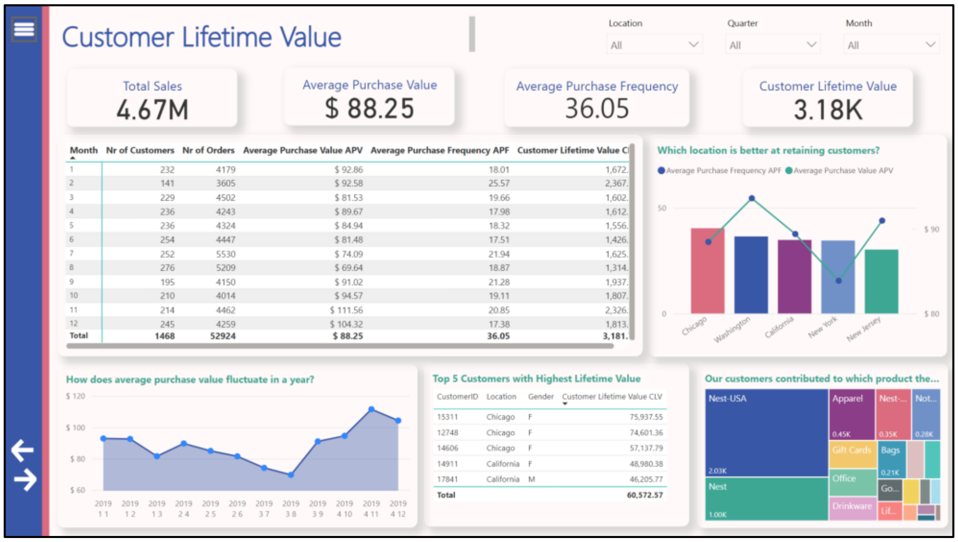

Figure 19 - PowerBI Dashboard for Customer Lifetime Value

The above PowerBI dashboard for Customer Lifetime Value (CLV). It focuses on attributes that impact LTV. The dashboard includes key metrics such as:

1. Average Purchase Value (APV): The average amount that customers have spent on the e-commerce product.
2. Average Purchase Frequency (APF): The average number of times that customers have bought an item from the e-commerce platform.
3. Customer Lifetime Value (CLV): How much is a customer worth throughout the entire relationship with the e-commerce platform.

Month	Nr of Customers	Nr of Orders	Average Purchase Value APV	Average Purchase Frequency APF	Customer Lifetime Value CLV
1	232	4179	92.86	18.01	1,672.61
2	141	3605	92.58	25.57	2,367.00
3	229	4502	81.53	19.66	1,602.79
4	236	4243	89.67	17.98	1,612.23
5	236	4324	84.94	18.32	1,556.22
6	254	4447	81.48	17.51	1,426.58
7	252	5530	74.09	21.94	1,625.93
8	276	5209	69.64	18.87	1,314.27
9	195	4150	91.02	21.28	1,937.09
10	210	4014	94.57	19.11	1,807.56
11	214	4462	111.56	20.85	2,326.04
12	245	4259	104.32	17.38	1,813.42
Total	**1468**	**52924**	**88.25**	**36.05**	**3,181.74**

Figure 20 - CLV on a granular aspect

Descriptive Statistics:

- The matrix chart allows the viewer to analyze data on a granular level. It displays metrics such as average purchase value (APV), average purchase frequency (APF), and customer lifetime value (CLV) on a monthly period.
- In February, CLV and APF achieved the highest at $2367 and 25.57 respectively.
- In December it had the highest number of customers at 245.
- July has the highest number of orders at 5530.
- November had the higher APV of $111.56 per user.

Insights:

- The range for APV typically remained within the $69 to $95 area. However, the APV increased by a minimum of 10% throughout November and December.
- With February and November securing the top spot in terms of CLV of $2,367 and $2,326 respectively. The remaining month did not exceed the $2,000 threshold, implying a difference of $300. It is led to believe that customers within these two months should be the e-commerce platform priority as it suggests users spend more per transaction.

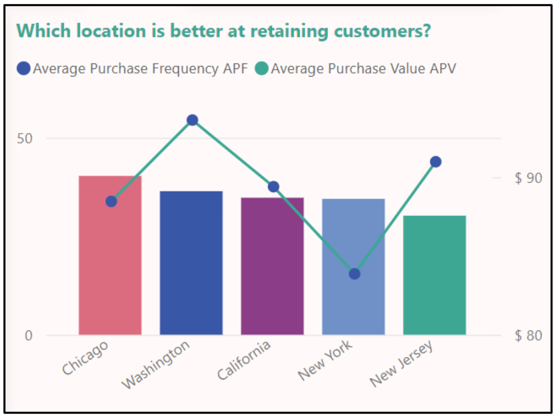

Figure 21 - The AVP and APV by location

Descriptive Statistics:

- As seen in Figure 21, it is seen that Chicago has the highest APF at 40, followed by Washington at 36, both California and New York at 34, with New Jersey at last, which is at 30.
- APV wise, Washington has the highest value at $93, followed by New Jersey at $90, California at $89, Chicago at $88, and New York at the lowest, which is $83.

Insights:

- Combining both standpoints, although Chicago has the highest APF, the APV aspect performed lower than expected. Thus, translating into an area where customers have higher shopping frequency but lower purchase value. It is led to believe that users purchase items with low-profit margins.
- On the other hand, despite the total sales by location. Washington APV achieved the highest APV while APF was the second highest among the other locations. This information allows the analyst to infer that Washington tends to have a higher tolerance towards higher spending. It leads to a higher average CLV.

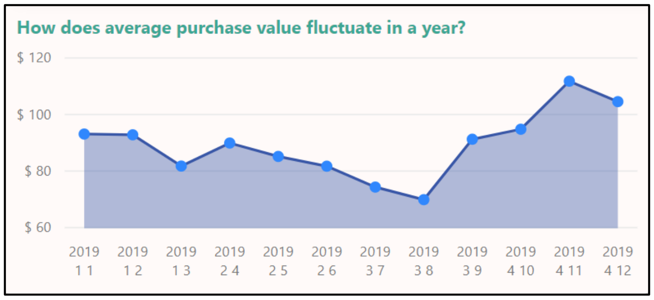

Figure 22 - Fluctuation of average purchase price by month

Insights:

- Even though the data is identical to the APV value presented in the matrix chart above, this visualization allows the user to drill up or drill down for a broader or granular inspection of the APV metric across a specific timeline. Meanwhile, the above matrix chart only provides data points on every month.

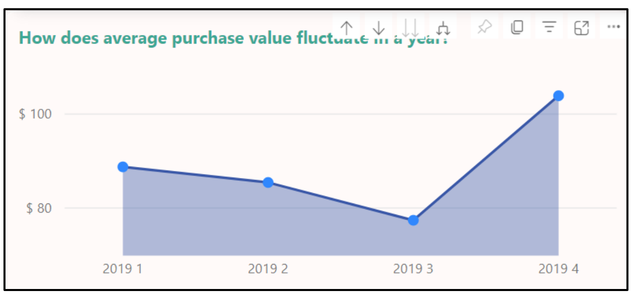

Figure 23 - Fluctuation of the average purchase price by quarter (Drill up)

Descriptive Statistics:

- From a drill-up perspective, it is observed that the average purchase value decreased from $88 to $77 during the first three quarters. However, Q4 APV had a tremendous increase up to $107, a 40% increase from Q3.

Insights:

- It could be inferred that the users had higher tolerance towards spending on the last quarter due to the festive season.

Top 5 Customers with Highest Lifetime Value

CustomerID	Location	Gender	Customer Lifetime Value CLV
15311	Chicago	F	75,937.55
12748	Chicago	F	74,601.36
14606	Chicago	F	57,137.79
14911	California	F	48,980.38
17841	California	M	46,205.77
Total			**60,572.57**

Figure 24 - Top 5 users with the highest CLV

Descriptive Statistics:

- Three of the five customers were from Chicago and the remaining were from California.
- It could be seen that the top four users with female and the highest CLV was $75K. On the other hand, the maximum spend for the male user reaches only $46K.

Insights:

- It could be deduced that users within Chicago could have higher CLV.

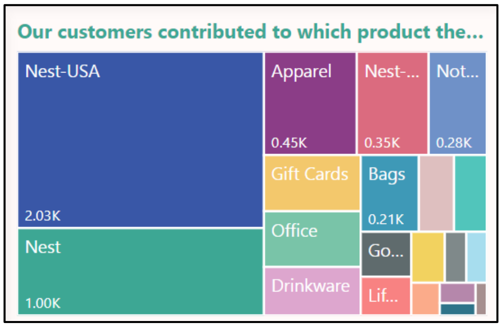

Figure 25 - Treemap of Customer's Lifetime Value by Product Category

Descriptive Statistics:

- Based on figure 25, it could be seen that the top 5 users typically spend more on the Nest-USA category followed by Nest.

2.2.5 Product Performance

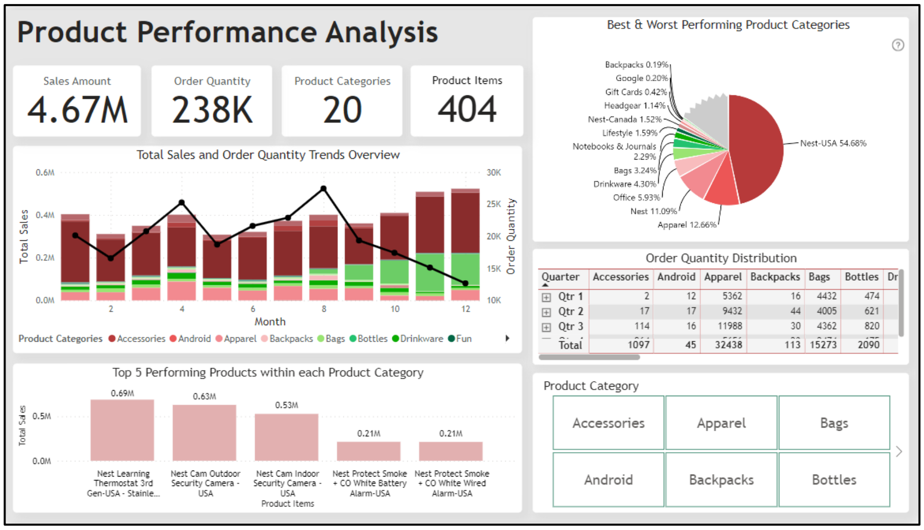

Figure 26 - PowerBI Dashboard for Product Performance Analysis

The Product Performance Analysis Dashboard provides a comprehensive view of the overall performance of the product categories along with the inclusion of product items. The key metrics used to evaluate the performance are:

1. Total Sales and Order Quantity Trends Overview
2. Best & Worst Performing Product Categories (Product Categories Sales Distribution)
3. Order Quantity Distribution
4. Top 5 Performing Product Items within each category.

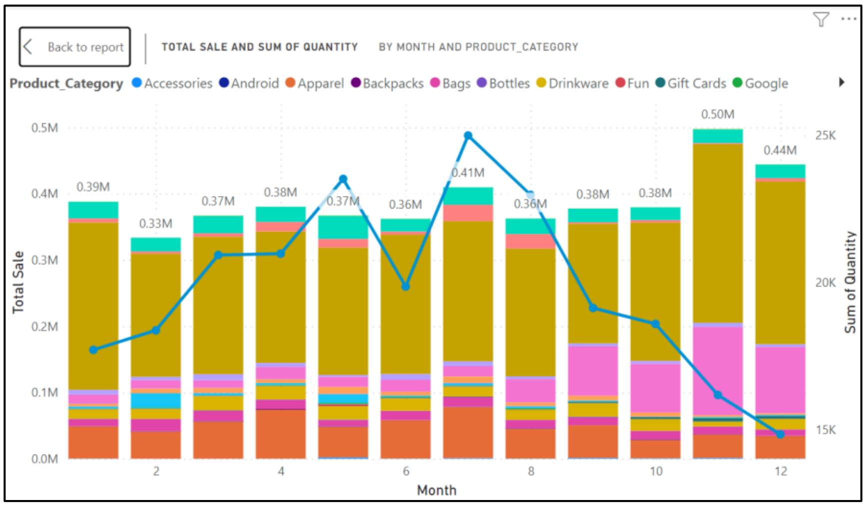

Figure 27 - Overview of Total Sales & Order Quantity Trends

Descriptive Statistic

- From Figure 27, the Nest-USA category (dark gold) was the best performing category as it held a significant amount of sales and inventory on a monthly basis.
- November had the best sales within the year, achieving the highest of $500K in that month.
- The line chart display reflects the movement of inventory towards the users. It could be seen that the line chart peaked in July at 25K inventory shipped towards users while December had the lowest at 14.83K.

Insight

- With the ongoing fluctuation of the sales amount over the 12 months throughout 2019, the overall trend of the total sales has proven to be an uptrend.
- Interestingly, it could be seen that although the column chart gradually increased, the line graph fell drastically. It is translated to the analyst that the number of quantities sold fell but the sales increased. Indicating that each product category price point is significantly different.
- In July, Apparel products sold the most at 410,000 pieces while only achieving a total sale of $77K. In November, Nest-USA sold 500,000 pieces and achieved a total sale of $270K. This information allows analysts to understand that the Apparel product mainly consists of items which have a lower price.

Commented [1]: To gain a holistic understanding of the relationship between sales performance and order quantity for the product categories throughout the year, we present a line and stacked column chart that consist of dual-axis, one depicting the sales amount and another indicating the order quantity trend.

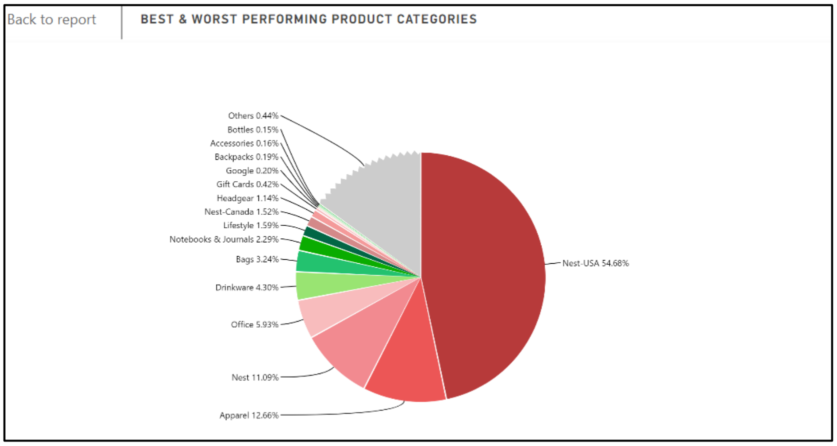

Figure 28 - Product Category (Best Product Category)

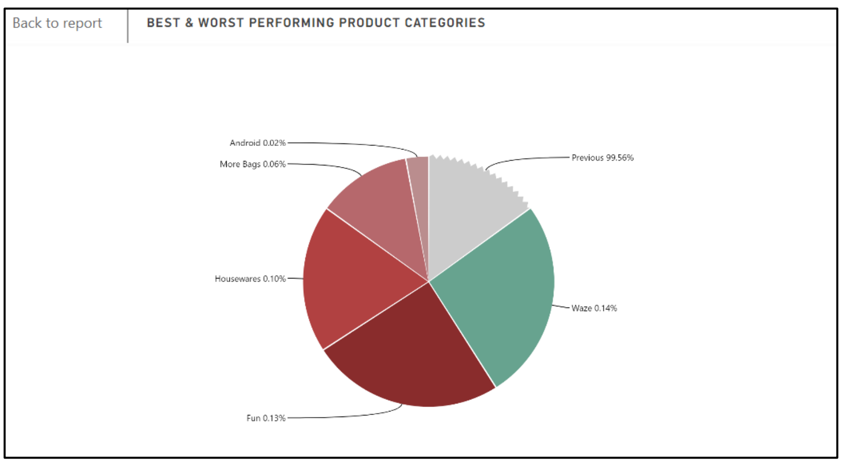

Figure 29 - Product Category (Worst Product Category) - Drill down

Descriptive Statistic

- From the figure 28, the top 5 performing category are:
    - Nest-USA (55%)
    - Apparel (13%)
    - Nest (11%)
    - Office (6%)
    - Drinkware (4.5%)

- From Figure 29, the worst 5 performing categories are:
    - Android (0.02%)
    - More Bags (0.06%)
    - Houseware(0.10%)
    - Fun (0.13%)
    - Waze (0.14%)

Quarter	Accessories	Android	Apparel	Backpacks	Bags	Bottles	Drinkware	Fun	Gift Cards	Google	Headgear	Housewares
⊞ 1	22	10	6574	24	4269	540	7471	159	21	173	1866	536
⊞ 2	116	13	9096	39	4291	523	7865	406	55	178	982	532
⊞ 3	280	20	10957	22	3971	804	8851	202	56	166	502	1333
⊞ 4	679	2	5811	28	2742	223	6314	67	74	58	183	83
Total	**1097**	**45**	**32438**	**113**	**15273**	**2090**	**30501**	**834**	**206**	**575**	**3533**	**2484**

Figure 30 - Order Quantity Distribution (Part 1)

Lifestyle	More Bags	Nest	Nest-Canada	Nest-USA	Notebooks & Journals	Office	Waze	**Total**
5746	100	176	142	5327	1317	22274	253	**57000**
7425	37	246	120	5090	2557	24354	426	**64351**
7782	4	634	100	4966	4728	21554	165	**67097**
3928	3	1781	107	6047	954	20201	300	**49585**
24881	**144**	**2837**	**469**	**21430**	**9556**	**88383**	**1144**	**238033**

Figure 31 - Order Quantity Distribution (Part 2)

With the matrix shown on Figure 30, the purpose of it would be to allow operation level staff to evaluate the inventory of each product category on a granular level. The matrix is interactive and allows operation staff to slice and dice the matrix accordingly to predict trends for inventory movement on a monthly or seasonal basis.

Descriptive Statistics:

- It could be seen that the Apparel category has the highest inventory movement throughout the year. A total of 32,438 units were sold. In contrast, products within the Android category only had 45 units sold throughout the year.
- Within 2019, Q3 has the highest order quantity at 67,097 units while Q4 has the lowest at 49,585 units.

Insights:

- Compared with Figure 27, it could be seen in Figure 30 that certain categories (Apparel, bags, bottles, drinkware, google, headgear, housewares, lifestyle) had a significant drop in terms of quantity volume while Nest-USA and Nest category had an increase of quantity sold when transitioning from Q3 to Q4.
- With a range of category drop in terms of quantity volume yet sales remain higher could reflect that Nest-USA and Nest could be an ideal selection to focus on promoting these categories to increase sales.

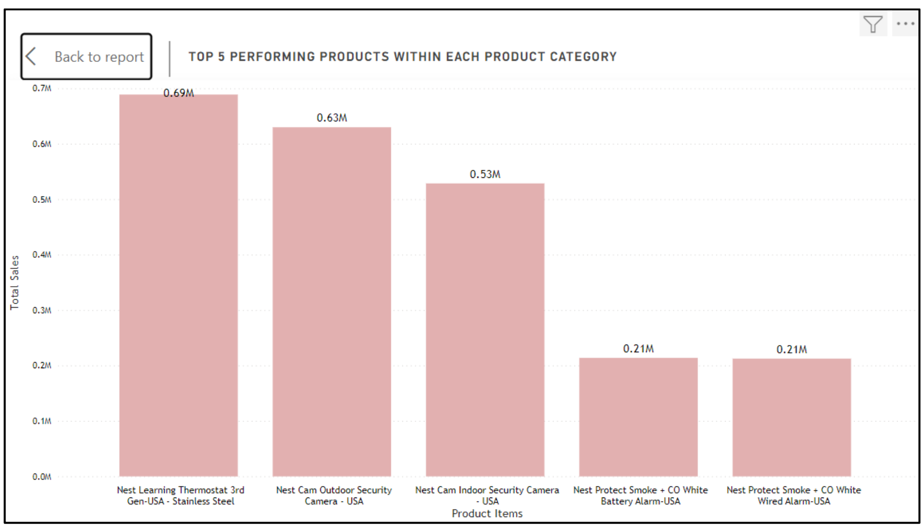

Figure 32- Top 5 Performing Product Items within each category

Descriptive Statistic

- Within the best performing category (Nest-USA), the 5 products that contributed the most for this category are:
  - Nest Learning Thermostat 3rd Gen-USA - Stainless Steel (30.3%)
  - Nest Cam Outdoor Security Camera - USA (27.71%)
  - Nest Cam Indoor Security Camera - USA (23.25%)
  - Nest Protect Smoke + CO White Battery Alarm-USA (9.40%)
  - Nest Protect Smoke + CO White Wired Alarm-USA (9.35%)

2.2.6 Market Basket Analysis

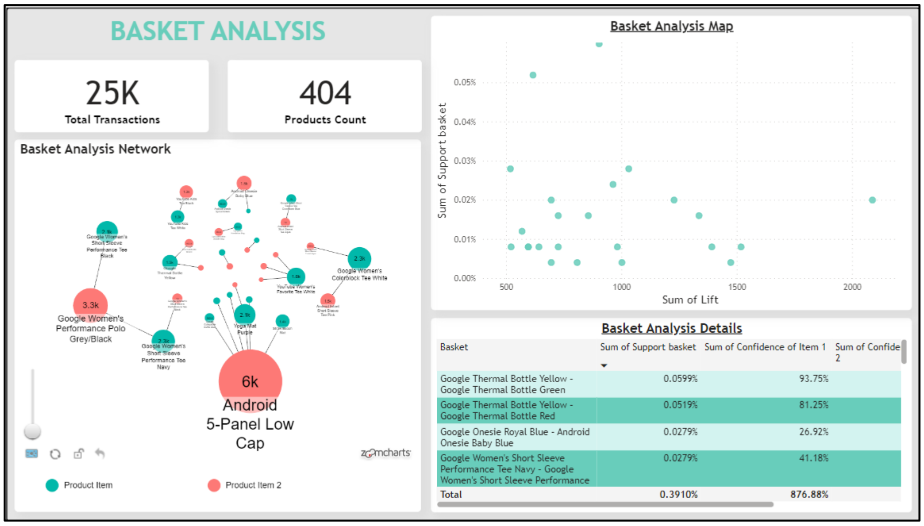

Figure 33 - PowerBI Dashboard for Market Basket Analysis

The Market Basket Analysis Dashboard provides insight regarding consumer purchasing behavior. The interactive dashboard identifies relationships between products to create an itemset. The key metrics used to measure support, confidence, and lift values between products are:

1. Basket Analysis Network
2. Basket Analysis Map
3. Basket Analysis Details

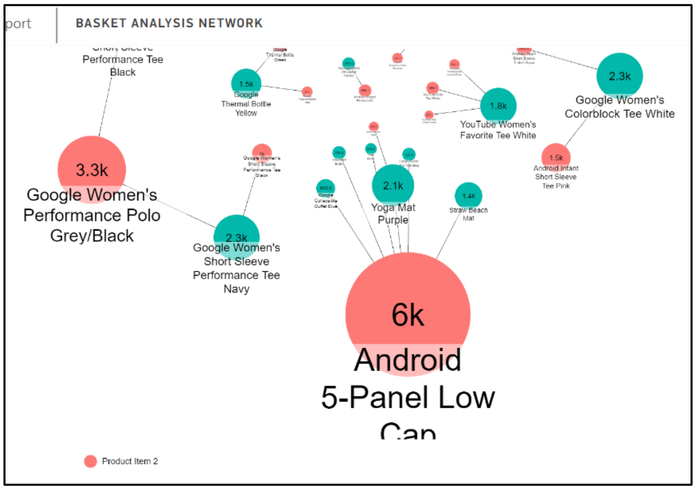

Figure 34 - Overview of Basket Analysis Network

Descriptive Statistics:

- A network visual is used to present the basket analysis network that visually depicts the relationship between items in the basket. As shown in Figure 34, each node represents an individual product (antecedent) item where the connected line reflects the co-occurrence of items (consequent) within a basket.

Insights:

- As seen in Figure 34, there are several red nodes which are the antecedent. These nodes are the following:
  - Android 5-Panel Low Cap
  - Google Women’s Performance Polo Grey/ Black
  - Google Women’s Colorblock Tee White
  - Youtube Women’s Favorite Tee White
- These nodes (key product) are the core of the cluster as it is connected to multiple nodes which reflect the probability of the users purchasing the consequence item.

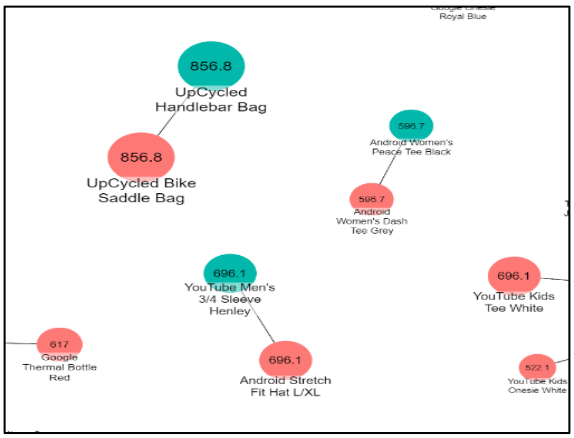

Figure 35 - Detail of Basket Analysis Network

Descriptive Statistics:

- Based on Figure 35, the network chart displays several clusters of interconnected nodes. These clusters are product groups that are often purchased together.
- As seen in Figure 35, several isolated nodes can be observed within.
    - UpCycled Handlebar Bag → UpCycled Bike Saddle Bag
    - Android Women's Peace Tee Black → Android Women's Dash Tee Grey
    - YouTube Men's ¾ Sleeve Henley → Android Stretch Fit Hat L/XL

Insights:

- With the suggested isolated node, the antecedent and consequent item process high complementary relationship therefore it suggests the creation of these combinations as itemset.

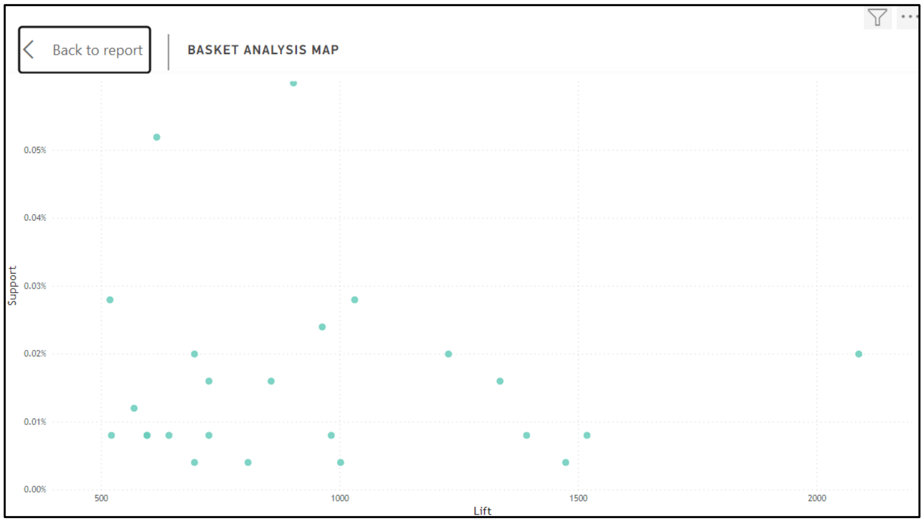

Figure 36 - Basket Analysis Map

A scatter chart is employed to serve as a basket analysis map, mapping product combinations based on their support and lift values. Each point on the chart corresponds to a specific combination, with the x-axis representing Support and the y-axis representing Lift.

Descriptive Statistics:

- Based on the chart, it can be observed that the majority of the points are mapped on the bottom left.
- Based on Figure 36, a minimum threshold of 500 for lift is applied to the map due to an excessive combination of products.
- Within Figure 36, the data point could be seen clustering into several groups. The goal would be to create a generalized insight regarding the map. Therefore, the analyst only assumes a large cluster with the lift range of 500 to 1500 and the support of 0.00% to 0.03%.

Insights:

- Support reflects the frequency of itemset purchases.
- Lift indicates the probability of items being purchased together rather than individually.
- With the cluster, it could be assumed that the average user's purchasing habit would exhibit low support and high lift. Although it is not a guarantee, the analyst creates a situation where these itemsets are more likely to be purchased frequently. The higher the life value, it reinforces the higher the probability of the itemset being sold together.
- With the outliers, the data point located on the top left reflects a strong relationship between these itemset. Typically, these products complement each other. The e-commerce platform should market these products together to increase total sales.

Back to report | BASKET ANALYSIS DETAILS

Basket	Sum of Support basket	Sum of Confidence of Item 1	Sum of Confidence of Item 2	Sum of Lift
Nest Cam Outdoor Security Camera - USA - Nest Cam Indoor Security Camera - USA	2.7653%	20.82%	21.46%	1.62
Nest Learning Thermostat 3rd Gen-USA - Stainless Steel - Nest Cam Outdoor Security Camera - USA	1.2011%	8.57%	9.04%	0.65
Google Sunglasses - Google 22 oz Water Bottle	1.0175%	16.74%	28.27%	4.65
Nest Learning Thermostat 3rd Gen-USA - Stainless Steel - Nest Cam Indoor Security Camera - USA	0.9098%	6.49%	7.06%	0.50
Nest Protect Smoke + CO White Battery Alarm-USA - Nest Learning Thermostat 3rd Gen-USA - Stainless Steel	0.9018%	16.61%	6.44%	1.19
Nest Protect Smoke + CO White Battery Alarm-USA - Nest Cam Outdoor Security Camera - USA	0.7661%	14.11%	5.77%	1.06
Google Sunglasses - Google Laptop and Cell Phone Stickers	0.7302%	12.02%	22.70%	3.74
Nest Protect Smoke + CO White Wired Alarm-USA - Nest Learning Thermostat 3rd Gen-USA - Stainless Steel	0.7222%	17.00%	5.16%	1.21
Google Sunglasses - Google Kick Ball	0.6265%	10.31%	55.87%	9.19
Nest Protect Smoke + CO White Battery Alarm-USA - Nest Cam Indoor Security Camera - USA	0.5985%	11.02%	4.64%	0.86
YouTube Custom Decals - Google Laptop and Cell Phone Stickers	0.5347%	31.98%	16.63%	9.94
Total	305.7779%	75203.66%	76634.19%	248850....

Figure 37 - Basket Analysis Details

To facilitate information to be easily digested, the analyst presented a matrix table which provides a comprehensive view of basket item combinations, support values, confidence value of both items in a combination, and the lift values.

Descriptive Statistics:

- Each row represents a unique combination, and columns offer specific insights into the strength, direction, and significance of these combinations.

Insights:

- The combinations include:
  - low-support and high-lift
  - high-support and high-lift
  - low-support and high-lift
  - high-support and low-lift
- By segregation the itemset into the proposed combination, it allows the analyst to extract insight regarding the itemset performance. These insights would be relayed to the e-

commerce platform to re-strategize their product recommendation section to improve user experiences.

- Products with high support and low lift suggested that product items in these pairs are individually popular but do not impact the purchase of the other.
- Moreover, the confidence value of items can also be analyzed to determine the direction of the purchase within the product combination.

3.0 Business Strategy

3.1 Creation of Business Intelligence Competency Centre (BICC)

In the scenario that the e-commerce company is reaching out to the analyst as a consultant. It would be a one-way communication and the analyst would follow the instructions of the e-commerce company. It displayed the scenario in which there would be a limited link between strategy and deployment of the Business Intelligence function.

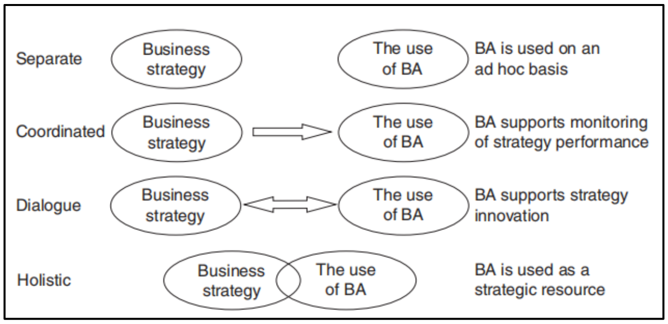

Figure 38 (Laursen & Thorlund, 2018, pg 21)

As seen in the image above, it would indicate that the link between the consultant and the e-commerce company displayed as scenario 2. With no backflow of information from the e-commerce company, the analyst would be only able to provide support in terms of reporting on a functional level. Therefore, the role of the consultant in the scenario would be reactive rather than proactive. The downfall of the current scenario would be that the analyst would be only able to provide support in terms of monitoring the department's Key Performance Index (KPI). Thus, it would be difficult to effectively review a strategy proposed by the business intelligence.

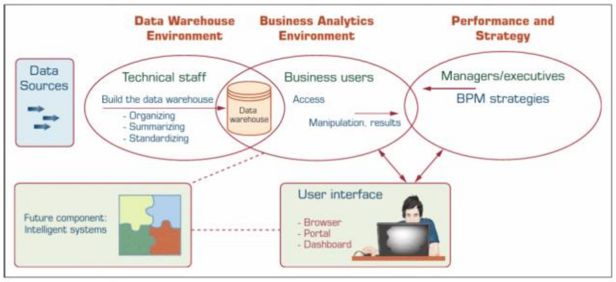

Figure 39 (Sharda et al, 2017, pg 43)

The e-commerce company requires business intelligence and analytics to support decision-making. It is recommended that the analyst should include information as a vision and mission. With the alignment of the vision and mission, information would be viewed as a strategic resource for the e-commerce company. By creating a holistic scenario where information would be a strategic resource, the BICC would be able to calibrate or create a data warehouse which would enable a systematic approach to data collection (Online Transactional Processing), consolidating of data and enabling data visualization with ease to allow efficient reporting (Online Analytical Processing). As PowerBI does have a DirectQuery function, it allows Business Intelligence Analysts to focus on monitoring the performance of the strategy with live data. With the implementation of data warehousing, visualization would be automatically refreshed therefore allowing Business Intelligence Analyst to identify any changes within its business (Laursen & Thorlund, 2018, pg 21).

3.2 Rockart Model

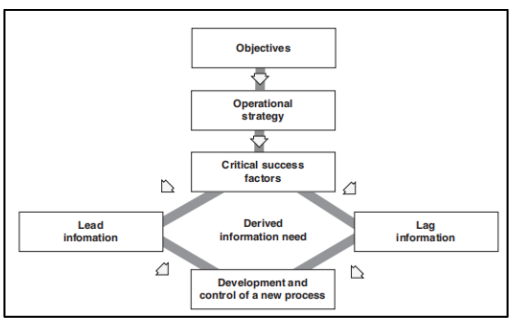

Figure 40 (Laursen & Thorlund, 2018, pg 60)

With the proposal of the creation of BICC, the analyst would be proposing a Rockart Model to create a flow of information that circulates within the organization level. Allowing on-demand business intelligence applications to view live data. As suggested by Anderson and Mittal (2000), the researcher introduces the dimension of Temporal Variation within the paper with a focus on sustaining or enhancing customer satisfaction. It emphasizes the timing of engaging in an effective activity at the right moment.

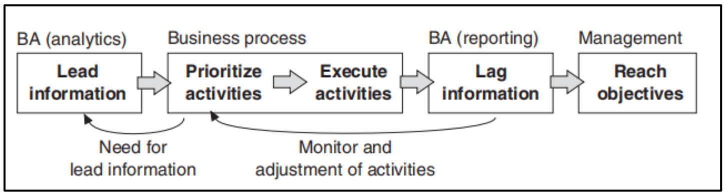

Figure 41 (Laursen & Thorlund, 2018, pg 60)

By identifying the lead and lag information within the strategy, it creates a closed-loop system that would enable feedback to circulate back to the management (Laursen & Thorlund, 2018, pg 60).

3.3 Strategy Implementation

1. E-service strategy

Level 1 - Objective	Maintain customer retention and enhancing customer loyalty
Level 2 - Identifying an operation strategy	Strategically re-evaluating seven dimensions of e-SERVQUAL
Level 3 - Identifying Critical Success Factors	The success factors for the current strategy would be targeted towards users who are active within the platform (Loyal, Potential Loyalist and etc).
Level 4 - Identifying Lead & Lag information (KPI)	● KPI 1: Reduce the R Value of the users by 15%. ● KPI 2: Increase the F Value of the users by 10%.

Table 7

Since it is an e-commerce company, the main strategy is to implement e-service, which is defined as ‘the extent to which a website facilitates efficient and effective shopping, purchasing, and delivery of products and services’ (Lin, Luo, Cai, Ma & Rong, 2016, pg 391). It means that the creation of better experiences does reflect the retention of users.

- Strategy 1 - Improve the website’s User Interface (UI) design.
    - Fu, Jiang, Zhang & Zhang (2019) identify the perceptual differences in UI between the user and designer. It created a cycle where the design consistently dynamically changes (Chen, Feng, Liu, Xing & Zhao, 2020). Thus, redesigning the UI based on the current trend along with the group segmentation would create an optimized level of modernity while engaging the users within the platform.
    - Moreover, with the data provided by the e-commerce platform. The UI team could take this into consideration upon optimizing the UI design with a more feminine outlook:
        - 63% of users are female.
    - The Key Performance Indicators (KPIs) for the UI Design would be:

        - Increase the Conversion Rate Optimization (CRO) by 15%
        - Collect User Feedback Surveys and evaluate the feedback.

- Strategy 2 - Improve website's User Experience (UX)
    - Typically, the users would emphasize the visualization of the website. Yet, users expect a certain level of usability (speed and loading time) of the website (Hasan, 2022). Therefore, the UX team would focus on developing and monitoring the KPI mentioned below to maintain a certain level of usability for the users.
    - The Key Performance Indicators (KPIs) for the UX Design would be:
        - Improve Task Success Rate by 10%
        - Increase Retention Rate by 10%
        - Reduce Time on Task rate by 10%

- Strategy 3 - Invest in Research and Development (R&D)
    - The goal of R&D was to conduct market research to either innovate new products or services within the industry. With ongoing development on Fintech, new innovations such as Buy Now Pay Later (BNPL) could be introduced to e-commerce to remain competitive (Laursen & Thorlund, 2018).
    - The Key Performance Indicators (KPIs) for R&D would be:
        - Increase the number of Features.
        - Competitive Benchmarking with competitor

2. Sales & Marketing Strategy

Level 1 - Objective	Increase the e-commerce platform revenue growth
Level 2 - Identifying an operation strategy	Optimizing marketing spend in terms of offline and online channels
Level 3 - Identifying Critical Success Factors	● Improve budget distribution for marketing channels. ● Targeting female audience ● Increase marketing activity during festive seasons
Level 4 - Identifying Lead & Lag information (KPI)	● KPI 1: Increase annual revenue growth by 15% ● KPI 2: Improve monthly ROI above 3.0. ● KPI 3: Increase average customer lifetime value to $5k

Table 8

- Strategy 1 - Re-evaluating budget for online and offline marketing
  - Although implementing the mixture of both offline and online marketing channels would offer a multidimensional shopping experience, it is undeniable that offline channels such as TV, radio, and newspapers must be altered to maintain their relevance in the Internet era (Liu et al., 2019). In other words, offline channels fail to capture an accurate customer journey across the entire marketing funnel, unlike online channels with access to APIs and website cookies to track customers' footprints (Liu & Zhang, 2022).
  - Thus, offline channels should focus on brand awareness as the main success factor, rather than leads and conversions. With that said, it makes sense to decrease the overall offline marketing budget to save costs.
  - The Key Performance Indicators (KPIs) for offline marketing would be:
    - Downsize monthly offline marketing budget to 25% of the total budget.
    - Increase Q4 online marketing budget by 20% for festive seasons.

- Strategy 2 - Optimize online marketing budget.
    - With online marketing channels continuously proven to lower costs in leads generation, expand new audiences, and spread valuable information (Syafrizal, 2021; Han et al., 2021; Etcheverry, 2022). The analyst recommends increasing the budget allocated toward online marketing.
    - The e-commerce platform currently has a stronger presence within Chicago and California. It is recommended to allocate a larger budget towards these locations to build a solid foundation for the e-commerce platform.
    - Meanwhile, it is recommended to allocate a smaller budget to new locations within the United States to experiment with a new user base.
    - Moreover, with the previous statistics showing that Q4 had the highest sales due to the festive seasons, it is worth increasing the budget in Q4 to promote extra rebates for festivals such as Black Friday and Christmas.
    - With 60% of users being female, the marketing and advertising strategy should be female-centric to maximize the Return on Investment (ROI).
    - The main success factor depends on the quality of leads and the amount of conversion gained from each online channel.
    - The Key Performance Indicators (KPIs) for online marketing would be:
        - Increase quarterly website visits by 20% via paid traffic.
        - Increase quarterly website visits by 10% via organic traffic.
        - Increase quarterly website conversions by 20%
        - Increase Q4 online marketing budget by 25% for festive seasons.
- Strategy 3 - Improve customer's lifetime value (LTV)
    - According to Weinstein (2002), the cost of acquiring a new customer would cost 5 times more than retaining an existing customer. Thus, companies are advised to invest more of the marketing budget towards retargeting old customers rather than spending the majority of the budget to acquire new customers. It would result in a healthier ROI value since old customers are most likely to purchase from the same company, given that they are satisfied with the products.
    - The main strategy is to allocate our online marketing budget toward areas with higher average purchase value and average purchase frequency, both of which

directly influence the customer lifetime value. By referring to Figure 10, it proves to be crucial to spend more to maintain and grow customers in Washington. Meanwhile, it is recommended to provide exclusive promotions via online coupons to increase the frequency of recurrent purchases.

- The Key Performance Indicators (KPIs) to evaluate customers purchasing behavior would be:
  - Increase the average purchase frequency for all locations by 20%
  - Increase the average purchase value for all locations by 25%
  - Increase the average customer lifetime value to $5,000.

3. Product recommendation Strategy

Level 1 - Objective	Increase the e-commerce platform revenue growth
Level 2 - Identifying an operation strategy	Strategically performing market basket analysis to increase cross-selling.
Level 3 - Identifying Critical Success Factors	Identifying itemset with strong lift and support to create itemset bundle or product recommendation.
Level 4 - Identifying Lead & Lag information (KPI)	• KPI 1: 10% increase of the average value of customer transaction

Table 9

In the evolving landscape of e-commerce, the imperative for businesses to rapidly access and analyze customer data to explore needs, preferences, and purchasing behaviors remains vital (Tsiptsis & Chorianopoulos, 2009). The large volume of customer and transaction data available presents a compelling opportunity to employ Market Basket Analysis that aims to introduce cross-selling and up-selling strategies.

- Strategy 1: Creation of bundle deal for itemset

Combinations of Product Items	Support	Lift
Google Thermal Bottle Yellow → Google Thermal Bottle Green	0.0599%	903.64
Google Thermal Bottle Yellow → Google Thermal Bottle Red	0.0519%	617.03
Google Women's Short Sleeve Performance Tee Navy → Google Women's Short Sleeve Performance Tee Black	0.0279	1031.92
Google Onesie Royal Blue → Android Onesie Baby Blue	0.0279%	519.01
Insulated Bottle → Android 5-Panel Low Cap	0.016%	716.41

Table 10

- With the past transaction, the data suggest that the item shown on table 10 had high support and high lift values. It reflects that these items were the most sought itemset among the various combinations.
- The Key Performance Indicators (KPIs) to evaluate the bundle deal are:
    - Increase the sales of the following itemset by 10%
    - Re-evaluate the top five itemset monthly.

- Strategy 2: Implementation of basket market analytics recommendation
    - With the market basket analysis, the insight display of a certain product would be the core node which forms multiple connections with several products. Thus, it allows the e-commerce platform to create better recommendations which would prompt users to increase the purchase.
    - The Key Performance Indicators (KPIs) to evaluate the market basket analysis are:
        - Increase by 10% in user clicks on the recommendations.
        - Increase the conversion by 10% from users clicking the recommendation

4.0 Conclusion

In conclusion, the report focuses on allowing the e-commerce company to review the data visualization created from the data provided. These meaningful insights essentially assist the company in understanding their current situation in terms of customer segmentation, customer RFM, sales, marketing, customer lifetime value, and product performance. Meanwhile, "3.0 Business Strategy" provides further insights with actionable business decisions to achieve the intended problem statements. The usage of Rockart Model has provided a detailed business strategy directed to each problem statement.

4.1 Project achievements

Problem statement	**Insights & Outcomes**
How can we optimize customer segmentation strategy to improve customer growth?	This project has provided two main dashboards showing the main customer insights on PowerBI, which are: ● Figure 2 - PowerBI Dashboard for Customer Segmentation ● Figure 7 - PowerBI Dashboard for RFM analysis Based on these insights, the strategy was created based on the insight. As seen on the e-service strategy, the analyst recommend the re-evaluate the User Interface (UI) to personalize the design. For example, by inferring that 60% of the users are female, the UI team could focus on the design of the UI design which appeals to the female gender. Moreover, from the RFM analysis, it allows the analyst to segregate the users within the e-commerce platform into a category based on their last purchase, spending and purchase interval. By improving usability and design to provide better experiences, the analyst would like to re-evaluate the RFM Analysis in the future to determine the effectiveness.
How can we optimize marketing strategy to improve revenue?	This project has provided two main dashboards showing the overall marketing insights on PowerBi, which are: ● Figure 10 - PowerBI Dashboard for Sales & Marketing Analysis ● Figure 18 - PowerBI Dashboard for Customer Lifetime Value

	Based on these insights, in order to optimize the marketing strategy to improve revenue, the company must increase their online budget and decrease their offline budget. Meanwhile, the company must target the correct demographic, which are females living in places like Chicago, which has contributed the largest sales amount in terms of location. However, the company should focus on Washington since it has the highest customer lifetime value (CLV) even though the location is the lowest in sales contribution. This could mean that customers in this location are extremely underrated and the company should increase their marketing budget targeting Washington as well.
How can we improve product performance to optimize supply and demand?	This project has provided two main dashboards showing the overall product insights on PowerBi, which are: ● Figure 26 - Power BI Dashboard for Product Performance ● Figure 33 - Power BI Dashboard for Market Basket Analysis By analyzing the dashboard, the company should allocate more resources on their best performing categories by using cross-selling strategies. Also, companies should consider removing the worst performing product categories whose order quantity is very low and reallocate the investment for the popular product categories such as Nest-USA. On another perspective, the company should employ basket analysis to identify the item sets with high support and high lift among the product categories. In a result

Table 11

4.2 Limitation

4.2.1 Lack of more detailed data

Various limitations occurred during data analysis. The main issue is the lack of sufficient data. Within several instances, the analyst has to create an assumption regarding the data provided. For example, for sales and marketing, the dataset only contains two main marketing streams, which are offline and online. A better analysis could have been implemented if the dataset contained data from each marketing channel such as Facebook, Google, and Instagram from online channels. Thus, it resulted in overgeneralizing the data therefore the analyst was unable to create an analyst accurately.

4.2.2 Lack of financial dataset

Aside from total sale amounts, the dataset does not include the financial aspect of the e-commerce company. Hence, the analyst was unable to determine if the e-commerce platform was operating in a profit or loss during 2019.

4.2.3 Data inconsistency

In addition, the given dataset contains five different sheets which are interrelated. For example, a relationship can be formed between “CustomersData.xlsx” and “Online_Sales.csv” via the “CustomerID” attribute. However, certain sheets could not be connected due to the inconsistency of the data format. For example, the date columns in “Discount_Coupon.csv” and “Coupon_Status.csv” are of different formats. The former date column is stored as a string for month values, while the latter is stored as a date format for exact date values.

4.3 Recommendation

For future recommendations, the analyst would like to recommend incorporating a BICC department to achieve a holistic approach within the organization. Currently, the analyst plays a supporting role that only provides document support. In addition, it would be recommended to introduce a data warehouse to allow live queries to be utilized to measure live data and allow management to keep track of the KPI recommended by the analyst. In addition, the analyst recommended the e-commerce platform to include the recommended metric suggested by the analyst to allow the e-commerce platform to track the progress. Moreover, by introducing a data warehouse, it creates a single source of truth dataset within the depositary to create data consistency and reduce confusion as to which dataset would be the latest version. In addition, the analyst would like to recommend a company to improve data collection to allow analysts to improve the analysis.

5.0 References

1. Anderson, E. W. and Mittal, V. (2000). Strengthening the satisfaction-profit chain. Journal of Service Research, 3(2), 107-120. https://doi.org/10.1177/109467050032001
2. Barbu, C. M., Florea, D. L., Dabija, D., & Barbu, M. C. R. (2021). Customer experience in fintech. Journal of Theoretical and Applied Electronic Commerce Research, 16(5), 1415-1433. https://doi.org/10.3390/jtaer16050080
3. Chen, C., Feng, S., Liu, Z., Xing, Z., & Zhao, S. (2020). From lost to found. Proceedings of the ACM on Human-Computer Interaction, 4(CSCW2), 1-22. https://doi.org/10.1145/3415194
4. Elfirdoussi, S., Lachgar, M., Kabaili, H., Rochdi, A., Goujdami, D., & Firdoussi, L. E. (2020). Assessing distance learning in higher education during the covid-19 pandemic. Education Research International, 2020, 1-13. https://doi.org/10.1155/2020/8890633
5. Etcheverry, C. (2022). The effect of e-commerce on grocery store choice and retail competition. Review of Network Economics, 21(1), 1-28. https://doi.org/10.1515/rne-2023-0017
6. Fader, P. S., Hardie, B. G. S., & Lee, K. L. (2005). Rfm and clv: using iso-value curves for customer base analysis. Journal of Marketing Research, 42(4), 415-430. https://doi.org/10.1509/jmkr.2005.42.4.415
7. Fu, Y., Jiang, H., Zhang, D., & Zhang, X. (2019). Comparison of perceptual differences between users and designers in mobile shopping app interface design: implications for evaluation practice. IEEE Access, 7, 23459-23470. https://doi.org/10.1109/access.2019.2899671
8. Gustriansyah, R., Suhandi, N., & Antony, F. (2020). Clustering optimization in rfm analysis based on k-means. Indonesian Journal of Electrical Engineering and Computer Science, 18(1), 470. https://doi.org/10.11591/ijeecs.v18.i1.pp470-477
9. Han, Q., Yang, Z., Zhang, Z., & Shen, L. (2021). Offline and online channel selection of low-carbon supply chain under carbon trading market. Mathematical Problems in Engineering, 2021, 1-17. https://doi.org/10.1155/2021/6627937
10. Hasan, L. (2022). Investigating user experience (ux) on the top most frequently visited e-commerce websites in malaysia. https://doi.org/10.21203/rs.3.rs-2137280/v1

11. Jafari, M., Chen, C., Mirzaie, M., & Tang, J. (2022). Nimaa: anr/cran package to accomplish nominal data mining analysis.. https://doi.org/10.1101/2022.01.13.475835
12. Kim, J., Jin, B., & Swinney, J. (2009). The role of etail quality, e-satisfaction and e-trust in online loyalty development process. Journal of Retailing and Consumer Services, 16(4), 239-247. https://doi.org/10.1016/j.jretconser.2008.11.019
13. Laursen, G. H. N., & Thorlund, J. (2018). Business Analytics for Managers : Taking Business Intelligence beyond Reporting (Second Edition). Hoboken, New Jersey: Wiley.
14. Liu, D. and Shih, Y. (2005). Integrating ahp and data mining for product recommendation based on customer lifetime value. Information &Amp; Management, 42(3), 387-400. https://doi.org/10.1016/j.im.2004.01.008
15. Liu, Y. and Zhang, H. (2022). Reassessing research on marketing channel strategies in the internet era: opportunities, challenges, and responses. Journal of Contemporary Marketing Science, 5(1), 81-91. https://doi.org/10.1108/jcmars-01-2022-0001
16. Liu, Y., Lan, Y., & Gao, W. (2019). An empirical study of customers' willingness to use offline channels under the background of new retail. Proceedings of the 1st International Conference on Business, Economics, Management Science (BEMS 2019). https://doi.org/10.2991/bems-19.2019.64
17. Nalchigar, S., Weber, I., Lak, P., & Bener, A. (2016). A large-scale study of online shopping behavior. Proceedings of the 20th International Database Engineering &Amp; Applications Symposium on - IDEAS '16. https://doi.org/10.1145/2938503.2938534
18. Qiu, J., Lin, Z., & Li, Y. (2015). Predicting customer purchase behavior in the e-commerce context. Electronic Commerce Research, 15(4), 427-452. https://doi.org/10.1007/s10660-015-9191-6
19. Salamai, A. A., Ageeli, A. A., & El-kenawy, E. M. (2022). Forecasting e-commerce adoption based on bidirectional recurrent neural networks. Computers, Materials &Amp; Continua, 70(3), 5091-5106. https://doi.org/10.32604/cmc.2022.021268
20. Sharda, R., Delen, D., Turban, E. & King, D. (2017). Business Intelligence : a Managerial Approach, Global Edition. Harlow, United Kingdom: Pearson.
21. Stephen, A. T. and Galak, J. (2012). The effects of traditional and social earned media on sales: a study of a microlending marketplace. Journal of Marketing Research, 49(5), 624-639. https://doi.org/10.1509/jmr.09.0401

22. Syafrizal, M. (2021). Web-based sme online marketing system (e-commerce). International Journal Software Engineering and Computer Science (IJSECS), 1(2), 75-79. https://doi.org/10.35870/ijsecs.v1i2.599
23. Tsiptsis, K., Chorianopoulos, A. (2009). Data Mining Techniques in CRM: Inside Customer Segmentation, John Wiley & Sons Ltd., Chichester, United Kingdom.
24. Weinstein, A. (2002). Customer retention: a usage segmentation and customer value approach. Journal of Targeting, Measurement and Analysis for Marketing, 10(3), 259-268. https://doi.org/10.1057/palgrave.jt.5740051
25. Yustiana, I., Sutarha, I., Maulawati, N., & Yusuf, I. M. (2021). Basket market analysis using r-based apriori algorithm to find information from sales data. International Journal Engineering and Applied Technology (Ijeat), 4(2), 79-92. https://doi.org/10.52005/ijeat.v4i2.52