Northeastern University
College of Professional Studies

Module 4 | Assignment (Team Report)
"Consumer Database/Dataset Overview"

Team Members:
DIVYA CHENTHAMARAKSHAN
NISHTHA JATINBHAI PATEL
SANIKA MAHESH POTPHPDE
YASH SANJAY SINGH

Northeastern University
ITC 6000: Database Management Systems
Prof. Venkata Duvvuri

INTRODUCTION

The dataset contains 3,900 entries with details about customer shopping behavior. It includes information like customer age, gender, and location, along with what they bought, how much they spent, and their preferred product size, color, and season. It also tracks feedback ratings, subscription status, use of discounts, shipping methods, and payment preferences. Additionally, it shows how often customers shop and their purchase history. This data helps analyze customer habits, satisfaction, and preferences, providing insights to improve sales, marketing, and customer experience.

PROJECT OVERVIEW

We chose 7 KPIs to better understand customer behavior and improve business performance. The Customer Satisfaction (CSAT) Score measures customer happiness to enhance their experience, while Sales Performance tracks discounts, promo codes, and shipping choices to boost sales. Product Performance / Profit Margin evaluates product quality and profitability, and Customer Lifetime Value (CLTV) measures the total revenue generated by a customer over time. Average Purchase Value (APV) calculates the average amount spent per transaction, and Seasonal Purchase Patternsidentifies how purchases change with seasons for better planning. Finally, Review Rating Distribution analyzes feedback to improve products and customer satisfaction. Together, these KPIs help to make data-driven decisions that enhance customer experience and drive business growth.

EXPLAINING KPIs

- **Customer Satisfaction (CSAT) Score:** Measures how happy customers are with a product or service, often based on reviews. It helps to identify what customers like or dislike, guiding improvements.
- **Sales Performance:**Tracks sales trends using metrics like discounts applied, promo codes used, and shipping types chosen. It also helps to understand which offers or services boost sales.
- **Product Performance / Product Profit Margin:**Analyzes product quality using review ratings and measures profitability. It identifies high-performing and profitable products for better focus.
- **Customer Lifetime Value (CLTV):**Measures the total revenue a customer generates over their relationship with the business. This helps to prioritize retaining high-value customers for long-term growth.
- **Average Purchase Value (APV):**Calculates the average amount spent per transaction. Helps understand customer spending habits and improve marketing strategies.
- **Seasonal Purchase Patterns:**Tracks how seasons affect customer buying behavior. Guides inventory and marketing efforts during peak seasons.
- **Review Rating Distribution:**Shows the spread of customer feedback ratings across products or services.Helps pinpoint areas of excellence or products needing improvement.

QUERIES AND ANALYSIS

Creating the New Shopping Behaviour : A new table,new shopping behaviour was created to store customer shopping data in an organized manner. It included key details like Customer ID Age,Purchase Amount USD,Shipping Type,Discount Applied, and other attributes relevant for analysis. After creating the table, data was migrated from the old tables, and the unnecessary ones were dropped to maintain a clean structure. The table was then renamed to Shopping behaviour 1 for consistency.

1) **Customer Satisfaction (CSAT) Score :**
The CSAT score was calculated using review ratings. The query summed up all the review ratings to get the Total Review Rating and it also averaged the ratings to determine the Average CSAT Score.

```
33        -------Customer Satisfaction (CSAT) Score
34
35        SELECT
36          SUM("Review_Rating") AS "Total Review Rating",
37          AVG("Review_Rating") AS "Average CSAT Score"
38        FROM
39          shopping_behavior1;
40
```

	Total Review Rating	Average CSAT Score
1	14624.8	3.74898743911817

The overall review rating sum is 14,624.8, and the average CSAT score is 3.75, showing moderate customer satisfaction levels. This suggests room for improvement in customer experience.

2) **Sales Performance:**
This analyzed sales performance by shipping type. It provided insights into:
- Total order count for each shipping type.
- Total sales revenue based on purchase amounts.
- The number of orders where discounts and promo codes were used.

```
41        ------- Sales Performance
42        SELECT
43          "Shipping_Type",
44          COUNT(*) AS "Orders Count",
45          SUM("Purchase_Amount_USD") AS "Total Sales",
46          COUNT(CASE WHEN "Discount_Applied" = 'Yes' THEN 1 END) AS "Discounted Sales",
47          COUNT(CASE WHEN "Promo_Code_Used" = 'Yes' THEN 1 END) AS "Promo Code Usage"
48        FROM shopping_behavior1
49        GROUP BY "Shipping_Type";
50
```

	Shipping_Type	Orders Count	Total Sales	Discounted Sales	Promo Code Usage
1	2-Day Shipping	627	38080.0	255	255
2	Express	646	39067.0	279	279
3	Free Shipping	675	40777.0	278	278
4	Next Day Air	648	37993.0	295	295
5	Shipping Type	1	0.0	0	0
6	Standard	654	38233.0	285	285
7	Store Pickup	650	38931.0	285	285

Execution finished without errors.
Result: 7 rows returned in 23ms
At line 41:

Free Shipping saw the highest sales and orders, showing its popularity. Discounts and promo codes were frequently used, reflecting customer inclination towards promotions.

3) **Product Profit Margin**
The Product Profit Margin analysis focuses on understanding the revenue and profit for each product type. The total revenue for each product was calculated by summing up the purchase amounts and the profit was estimated by assuming a 30% profit margin.

This analysis shows the revenue and estimated profit for each product type, assuming 30% profit margin. For example, Blouse had the highest revenue of $10,410, with an estimated profit of $3,123.

```
52    ------- Product Profit Margin
53
54    SELECT
55       "Item_Purchased",
56       ROUND(AVG("Review_Rating"), 2) AS "Average Review Rating",
57       SUM("Purchase_Amount_USD") AS "Total Revenue",
58       ROUND(SUM("Purchase_Amount_USD") * 0.3, 2) AS "Estimated Profit" -- Assuming a 30% margin
59    FROM shopping_behavior1
60    GROUP BY "Item_Purchased"
61    ORDER BY "Total Revenue" DESC;
62
```

	Item_Purchased	Average Review Rating	Total Revenue	Estimated Profit
1	Blouse	3.68	10410.0	3123.0
2	Shirt	3.63	10332.0	3099.6
3	Dress	3.75	10320.0	3096.0
4	Pants	3.72	10090.0	3027.0
5	Jewelry	3.76	10010.0	3003.0
6	Sunglasses	3.74	9649.0	2894.7
7	Belt	3.76	9635.0	2890.5

```
Execution finished without errors.
Result: 26 rows returned in 22ms
At line 52:
------- Product Profit Margin
```

4) Customer Lifetime Value (CLTV)

The Customer Lifetime Value (CLTV) analysis calculates the total spending of each customer over their lifetime. This was achieved by summing up all purchases made by each customer and grouping them by their Customer ID.The data was sorted in descending order to highlight the highest-spending customers.

```
64    ------------------Customer Lifetime Value (CLTV)
65
66    SELECT
67       "Customer_ID",
68       SUM("Purchase_Amount_USD") AS "Lifetime Value"
69    FROM shopping_behavior1
70    GROUP BY "Customer_ID"
71    ORDER BY "Lifetime Value" DESC;
72
```

	Customer_ID	Lifetime Value
1	96	100.0
2	862	100.0
3	770	100.0
4	616	100.0
5	582	100.0
6	519	100.0
7	456	100.0

```
Execution finished without errors.
Result: 3901 rows returned in 24ms
At line 64:
------------------Customer Lifetime Value (CLTV)
```

CLTV measures how much a customer has spent in total. Customers 96, 862, and 770 each spent $100, the highest in this dataset.

5) Average Purchase Value (APV)

The Average Purchase Value (APV) measures the average amount spent per transaction. This was calculated by taking the average of all purchase amounts across all transactions.

```
74    ------------------Average Purchase Value (APV)
75
76    SELECT
77       ROUND(AVG("Purchase_Amount_USD"), 2) AS "Average Purchase Value"
78    FROM shopping_behavior1;
79
80
```

	Average Purchase Value
1	59.75

The average amount spent per transaction is $59.75, means customers typically spent around $60 per purchase.

6) Seasonal Purchase Patterns

This query analyzes how purchases and sales vary across seasons. It counts total purchases and calculates total sales for each season, sorting them by sales in descending order. Businesses can use this to identify their peak sales seasons and optimize inventory accordingly. Additionally, it can guide marketing strategies to focus on high-revenue seasons.

```
81    ----------------Seasonal Purchase Patterns
82    SELECT
83      "Season",
84      COUNT(*) AS "Total Purchases",
85      SUM("Purchase_Amount_USD") AS "Total Sales"
86    FROM shopping_behavior
87    GROUP BY "Season"
88    ORDER BY "Total Sales" DESC;
89
```

	Season	Total Purchases	Total Sales
1	Fall	975	60018.0
2	Spring	999	58679.0
3	Winter	971	58607.0
4	Summer	955	55777.0

Fall had the highest sales ($60,018) despite fewer purchases than Spring, while Summer had the lowest sales and purchases. This indicates that Fall transactions have a higher average value compared to other seasons.

7) Review Rating Distribution

This query examines how customer review ratings are distributed. It counts the number of reviews for each rating, showing the frequency of ratings received. This helps businesses understand customer satisfaction trends. Additionally, it highlights areas where improvements might boost customer ratings.

```
91    ---------Review Rating Distribution
92
93    SELECT
94      "Review_Rating",
95      COUNT(*) AS "Number of Reviews"
96    FROM shopping_behavior
97    GROUP BY "Review_Rating"
98    ORDER BY "Review_Rating";
99
```

	Review_Rating	Number of Reviews
1	2.5	66
2	2.6	159
3	2.7	154
4	2.8	136
5	2.9	170
6	3	162
7	3.1	157

```
Execution finished without errors.
Result: 26 rows returned in 22ms
At line 91:
```

Most ratings range is for 3.4, indicating mid-level satisfaction among customers. This distribution suggests room for improvement in product or service quality to achieve higher ratings.

Data Visualization

We chose R Studio to create data visualizations for our four KPIs:
1. Seasonal Purchase Patterns
2. Review Rating Distribution
3. Product Performance
4. Sales Performance by Shipping Type.

This tool allows us to represent the data visually, making it easier to identify trends, patterns, and actionable insights. It provides clear, professional graphs that simplify complex data, helping us make informed decisions based on visual trends.

1. **Seasonal Purchase Patterns** : The attached graph shows Seasonal Purchase Pattern**s**, with total sales (in USD) compared across seasons. Fall has the highest sales ($60,018), while Summer has the lowest ($55,777). Sales drop in Summer but pick up again in Winter. This means sales are better during Fall and Winte**r**, likely because of holidays or seasonal demand, while Summer sees fewer purchases. Businesses can use this information to plan better for the seasons with higher sales

2. **Review Rating Distribution** : This bar chart illustrates the distribution of product review ratings, highlighting the number of reviews for ratings ranging from 1 to 5. The majority of reviews fall within the 3 to 5 range, indicating that most customers are moderately to highly satisfied. There are fewer reviews with ratings at the extreme ends, suggesting that strongly negative or overly enthusiastic feedback is less common.
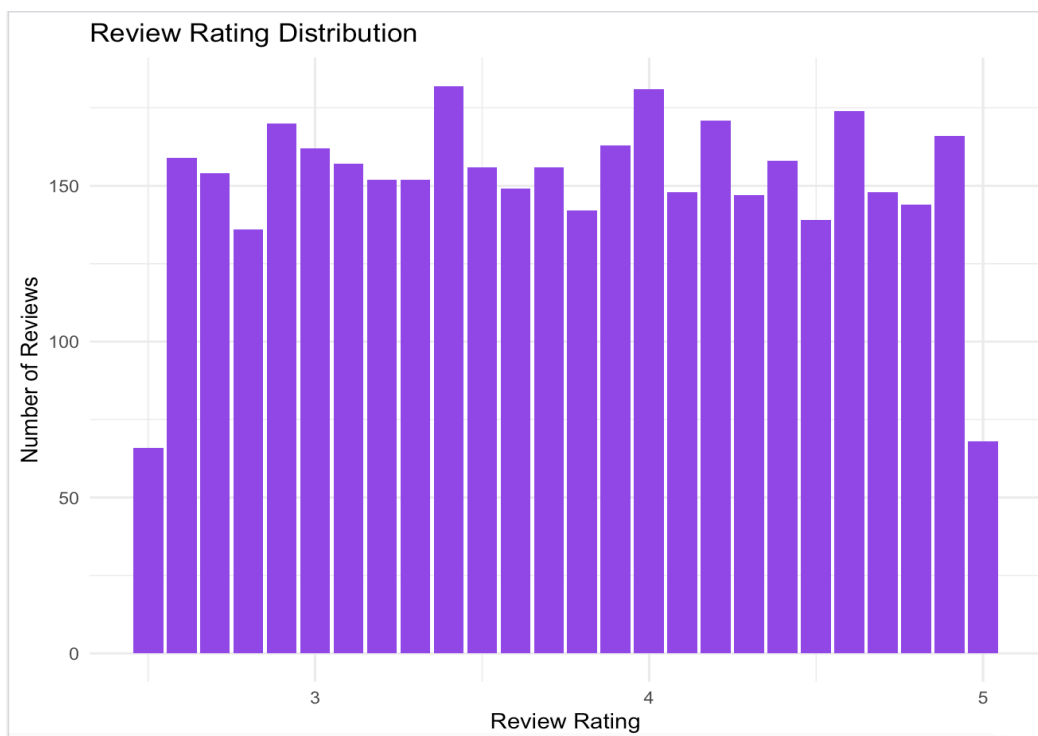
3. Product Performance : This bar chart displays the total revenue generated by various product categories, such as backpacks, belts, and blouses. High-performing items like backpacks and belts generate the most revenue, while products like jeans and jewelry contribute less. This analysis helps identify the most and least profitable items for the business.
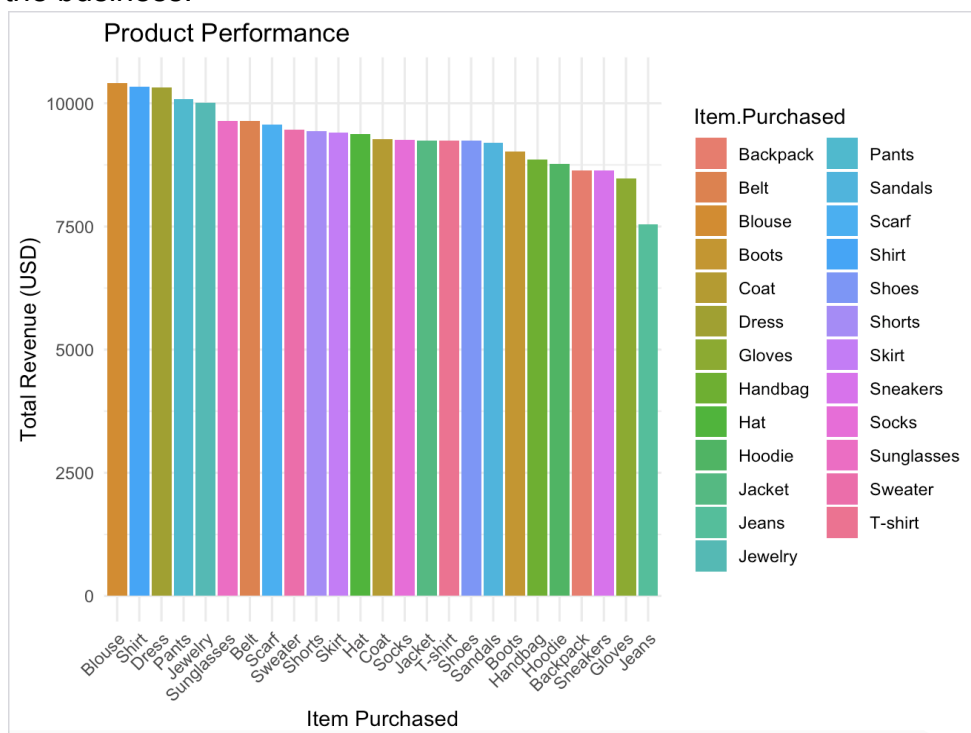
4. Sales Performance by Shipping Type

This bar chart depicts total sales (in USD) categorized by different shipping options, including 2-Day Shipping, Express, Free Shipping, Next Day Air, Standard, and Store Pickup. Free Shipping accounts for the highest sales, indicating its popularity among customers. The other shipping types generate similar sales figures, but none surpass the revenue brought in by Free Shipping.

Additional Analytics

1. Customer Segmentation Analysis
This analysis groups customers based on their age, gender, how often they shop, and how much they spend. By doing this, we can identify different types of customers, such as those who spend a lot (high-value customers), those who shop regularly (frequent shoppers), or those who mainly buy during sales (discount seekers). Understanding these groups helps businesses create targeted marketing campaigns, suggest the right products, and build loyalty programs that suit each type of customer.

2. Sales Trends Analysis
This analysis looks at sales patterns over time by focusing on product categories, seasonal changes, or regions. It helps identify trends, such as which products sell best at certain times of the year or in specific locations. Knowing this allows businesses to plan better, stock up on popular items during busy periods, and run promotions that match customer demand.

Demonstration Of The Analysis

Customer Segmentation Analysis (Demonstration)

To demonstrate this, we can group customers by their total spending and purchase frequency. For example, customers who spend more than $1,000 and shop at least 10 times a year can be categorized as "High-Value Customers." Those who shop frequently but spend less per visit could be classified as "Frequent Shoppers," while customers who primarily buy discounted items may fall into the "Discount Seekers" category. This

segmentation helps businesses focus their efforts on retaining high-value customers, encouraging frequent shoppers, and crafting targeted promotions for discount seekers.

Sales Trend Analysis (Demonstration)

For this demonstration, we can track monthly sales data for a particular product category, such as electronics. By plotting the sales numbers over time, we might observe a spike during holiday seasons, indicating a seasonal trend. Similarly, analyzing sales by region might reveal that certain areas have higher demand for specific products. This information can be used to prepare inventory and launch targeted marketing campaigns during peak seasons or in high-demand regions.

Conclusion

This report provides a detailed analysis of the business's performance and customer behaviour based on seven key metrics. These KPIs CSAT Score, Sales Performance, Product Profit Margin, Customer Lifetime Value (CLTV), Average Purchase Value (APV), Seasonal Purchase Patterns, and Review Rating Distribution offer a comprehensive understanding of how the business is performing across critical areas.

The analysis highlights important findings. The CSAT Score reflects customer satisfaction, giving insight into service quality. Sales Performance identifies trends in revenue and helps track growth. Product Profit Margins highlight which products drive profitability, while CLTV and APV shed light on customer spending behavior and overall value to the business. Seasonal Purchase Patterns reveal when demand peaks, helping the business plan inventory and marketing effectively. Lastly, the Review Rating Distribution provides feedback on product quality and customer experience.

To make the data clearer, visualizations were created using R for Seasonal Purchase Patterns, Review Rating Distribution, Product Performance, and Sales Performance by Shipping Type. These visualizations help identify trends, customer preferences, and areas needing improvement.

By combining these insights, the business can focus on improving customer satisfaction, optimizing product offerings, preparing for seasonal demand, and refining sales strategies. This data-driven approach provides clear steps to improve performance and create a better customer experience.

Reference:

1. Spiryn, K., & Spiryn, K. (2024, September 3). *6 Most important customer service KPIs | Brand24*. Brand24 Blog. https://brand24.com/blog/customer-service-kpis/
2. *W3Schools.com*. (n.d.). https://www.w3schools.com/sql/
3. GeeksforGeeks. (2024, July 5). *Getting started with Data Visualization in R*. GeeksforGeeks. https://www.geeksforgeeks.org/data-visualization-in-r/
4. *Consumer Behavior and Shopping Habits Dataset:* (2023, October 19). Kaggle. https://www.kaggle.com/datasets/zeesolver/consumer-behavior-and-shopping-habits-dataset/?select=shopping_behavior_updated.csv