BUSINESS MARKETING AND CUSTOMER SENTIMENT ANALYSIS REPORT

Aditya Kumar

TABLE OF CONTENT

BUSINESS PROBLEM

an online retail business, is facing reduced customer engagement and conversion rates despite launching several new online marketing campaigns. They are reaching out to you to help conduct a detailed analysis and identify areas for improvement in their marketing strategies.

Key Point:

Reduced Customer Engagement: The number of customer interactions and engagement with the site and marketing content has declined.

Decreased Conversion Rates: Fewer site visitors are converting into paying customers.

High Marketing Expenses: Significant investments in marketing campaigns are not yielding expected returns.

Need for Customer Feedback Analysis: Understanding customer opinions about products and services is crucial for improving engagement and conversions.

TOOLS USED

- **SQL Server :** To store and manage raw datasets such as customer reviews, product info.

- **Python :** To analyze review text and generate sentiment scores (Positive, Negative, Neutral, Mixed). Exported analyzed data to CSV for Power BI import

- **Microsoft Power BI :** For building interactive dashboards and visualizing review data, sentiment trends, and performance metrics.

DATASET DESCITPTION

This section includes a detailed description of data from the number of instances to the missing values, outliers, types of data attributes, and many more. This description helps us to understand our data.

The raw dataset consists of **6 relational tables** Those are **stored in a SQL server database** :

- **Customer_journey :** detailed interaction logs between customers and products on a website or digital platform.
  - Columns : JourneyID, CustomerID, ProductID, VisitDate,Stage(checkout/homepage/productpage), Action, Duration
  - Count of records : 4,011

- **Cutomer_reviews:** captures valuable customer feedback about products
  - Columns: ReviewID, CustomerID, ProductID, ReviewDate, Rating(1-5), ReviewText
  - Count of records: 1,363

- **Customers :** Cotain Customers Information
  - Columns : CustomerID, CustomerName, Email, Gender, Age, GeographyID
  - Count of records: 100

- **Engagement_data :** tracks customer interaction with marketing content
  - Columns: EngagementID, ContentID, ContentType, Likes, EngagementDate, CampaignID, ProductID, ViewsClicksCombined
  - Count of records: 4,623

- **Geography:** providing location-based segmentation for customers or campaigns.
  - Columns: GeographyID, Country, City
  - Count of records: 10

- **Products:** Stores product details
  - Columns: ProductID, ProductName, Category, Price
  - Count of records: 20

Customer_reviews_with_sentiment.csv is extracted from the table customers_reviews which consist of sentiment analysis of customers

- Columns : ReviewID, CustomerID, ProductID, ReviewDate, Rating, ReviewText, SentimentScore, SentimentCategory, SentimentBucket
- Count of records : 1363

DATA PREPROCESSING

Using SQL

1. **Joined the table customers and georaphy on the GeographyID** field to match customers with their geographic information and dropped GeographyId as after join there is no use of it.

```
SELECT
    c.CustomerID,
    c.CustomerName,
    c.Email,
    c.Gender,
    c.Age,
    g.Country,
    g.City
FROM
    dbo.customers as c
LEFT JOIN
    dbo.geography g
ON
    c.GeographyID = g.GeographyID;
```

2. Feature Extraction in the table Products and **categorized product into low, medium and high price category** in product table.

```
SELECT
    ProductID,
    ProductName,
    Price,

    CASE -- Categorizes the products into price categories: Low, Medium, or High
        WHEN Price < 50 THEN 'Low'
        WHEN Price BETWEEN 50 AND 200 THEN 'Medium'
        ELSE 'High'
    END AS PriceCategory  -- Names the new column as PriceCategory

FROM
    dbo.products;  -- Specifies the source table from which to select the data
```

3. **Removed The unnecessary spaces** in the ReviewText of customer_reviews table

```
SELECT
      ReviewID,
      CustomerID,
      ProductID,
      ReviewDate,
      Rating,
      REPLACE(ReviewText,'  ',' ') AS ReviewText FROM dbo.customer_reviews;
```

4. **clean and normalize** the engagement_data table

```
SELECT
    EngagementID,
    ContentID,
       CampaignID,
    ProductID,
```

```
    UPPER(REPLACE(ContentType, 'Socialmedia', 'Social Media')) AS ContentType,  -
- Replacing "Socialmedia" with "Social Media" and then converting all ContentType
values to uppercase
    Likes,
    LEFT(ViewsClicksCombined, CHARINDEX('-', ViewsClicksCombined) - 1) AS Views,
-- Extracts the Views part from the ViewsClicksCombined column by taking the
substring before the '-' character
    RIGHT(ViewsClicksCombined, LEN(ViewsClicksCombined) - CHARINDEX('-',
ViewsClicksCombined)) AS Clicks,  -- Extracts the Clicks part from the
ViewsClicksCombined column by taking the substring after the '-' character
    FORMAT(CONVERT(DATE, EngagementDate), 'dd.MM.yyyy') AS EngagementDate  --
formating the date as dd.mm.yyyy
FROM dbo.engagement_data
WHERE
       ContentType != 'Newsletter';  -- Filters out rows where ContentType is
   'Newsletter' as these are not relevant for our analysis
```

5. Removal of **Duplicate and Null Values** in Customers_journey table

```
SELECT
    JourneyID,
    CustomerID,
    ProductID,
    VisitDate,
    Stage,
    Action,
    COALESCE(Duration, avg_duration) AS Duration
FROM
    (
        -- Subquery to process and clean the data
        SELECT
            JourneyID,
            CustomerID,
            ProductID,
            VisitDate,
            UPPER(Stage) AS Stage,
            Action,
            Duration,
            AVG(Duration) OVER (PARTITION BY VisitDate) AS avg_duration,  --
Calculates the average duration for each date, using only numeric values
            ROW_NUMBER() OVER (
                PARTITION BY CustomerID, ProductID, VisitDate, UPPER(Stage),
Action  -- Groups by these columns to identify duplicate records
                ORDER BY JourneyID  -- Orders by JourneyID to keep the first
occurrence of each duplicate
            ) AS row_num  -- Assigns a row number to each row within the
partition to identify duplicates
        FROM
            dbo.customer_journey  -- Specifies the source table from which to
select the data
    ) AS subquery  -- Names the subquery for reference in the outer query
WHERE
       row_num = 1;  -- Keeps only the first occurrence of each duplicate group
   identified in the subquery
```

Using Python

1. **Feature Extraction -SentimentScore**: calculate sentiment scores using VADER

```
def calculate_sentiment(review):
    # Get the sentiment scores for the review text
    sentiment = sia.polarity_scores(review)
    # Return the compound score, which is a normalized score between -1 (most
negative) and 1 (most positive)
    return sentiment['compound']
```

2. **Feature Extraction – SentimentCategory:** categorize sentiment using both the sentiment score and the review rating

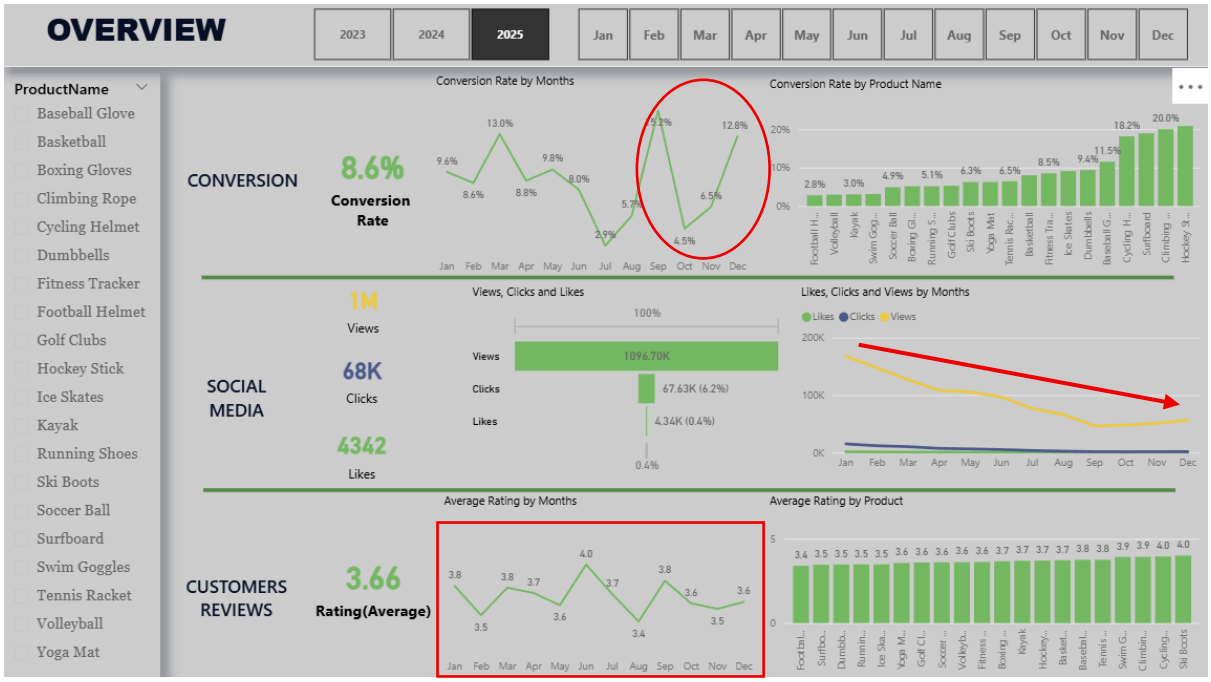
```
def categorize_sentiment(score, rating):
    # Use both the text sentiment score and the numerical rating to determine
sentiment category
    if score > 0.05:  # Positive sentiment score
        if rating >= 4:
            return 'Positive'  # High rating and positive sentiment
        elif rating == 3:
            return 'Mixed Positive'  # Neutral rating but positive sentiment
        else:
            return 'Mixed Negative'  # Low rating but positive sentiment
    elif score < -0.05:  # Negative sentiment score
        if rating <= 2:
            return 'Negative'  # Low rating and negative sentiment
        elif rating == 3:
            return 'Mixed Negative'  # Neutral rating but negative sentiment
        else:
            return 'Mixed Positive'  # High rating but negative sentiment
    else:  # Neutral sentiment score
        if rating >= 4:
            return 'Positive'  # High rating with neutral sentiment
        elif rating <= 2:
            return 'Negative'  # Low rating with neutral sentiment
        else:
            return 'Neutral'  # Neutral rating and neutral sentiment
```

3. **Feature Extraction – SentimentBucket :** Extracting bucket sentiment scores

```
ef sentiment_bucket(score):
    if score >= 0.5:
        return '0.5 to 1.0'  # Strongly positive sentiment
    elif 0.0 <= score < 0.5:
        return '0.0 to 0.49'  # Mildly positive sentiment
    elif -0.5 <= score < 0.0:
        return '-0.49 to 0.0'  # Mildly negative sentiment
    else:
        return '-1.0 to -0.5'  # Strongly negative sentiment
```

KEY FINDING AND INSIGHTS

Overview



Decreased Conversion Rates: The conversion rate demonstrated a strong rebound in December, reaching 10.3%, despite a notable dip to 5.1% in October,2025.

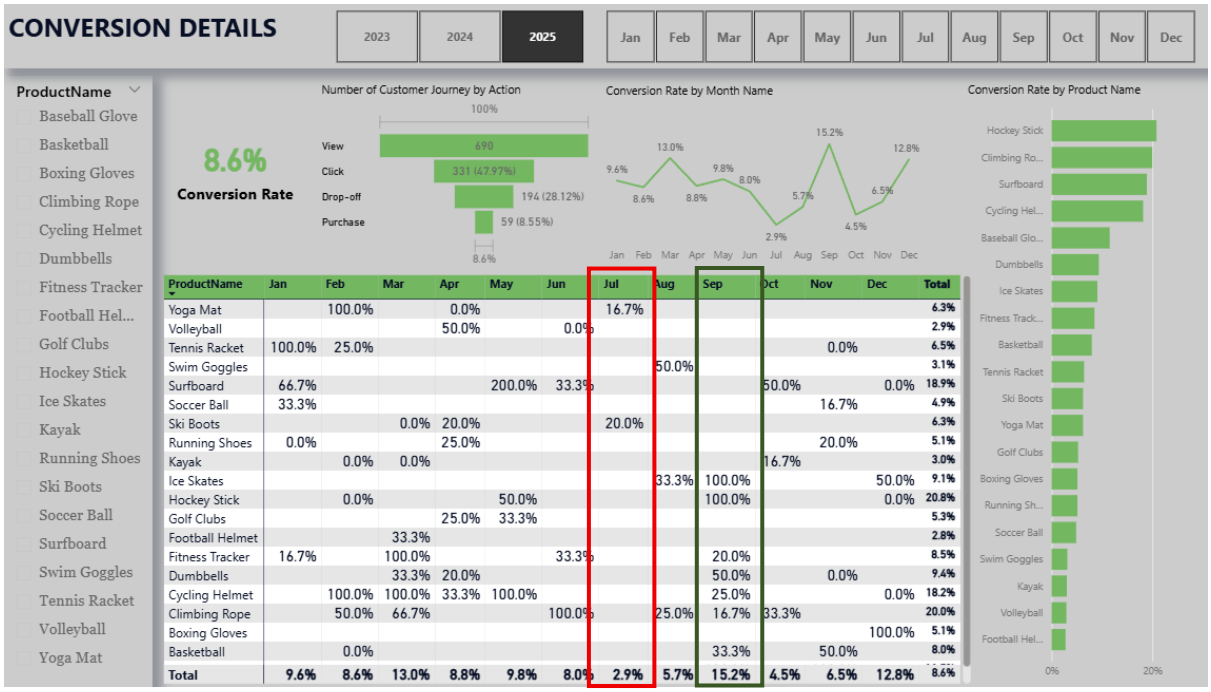
Reduced Customer Engagement:

- There is a decline in overall social media engagement, with views dropping throughout the year.
- While clicks and likes are also extremely low compared to views, the click-through rate stands at 6.2%, and likes is also 0.4% which is also a not good indicator

Customer Feedback Analysis:

- Customer ratings have remained consistent, averaging around 3.66 throughout the year.
- Although stable, the average rating is below the target of 4.0, suggesting a need for focused improvements in customer satisfaction, for products below 3.5

Decreased Conversion Rates



ProductName	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Total
Yoga Mat		100.0%		0.0%			16.7%						6.3%
Volleyball				50.0%		0.0%							2.9%
Tennis Racket	100.0%	25.0%									0.0%		6.5%
Swim Goggles								50.0%					3.1%
Surfboard	66.7%				200.0%	33.3%				50.0%		0.0%	18.9%
Soccer Ball	33.3%										16.7%		4.9%
Ski Boots			0.0%	20.0%			20.0%						6.3%
Running Shoes	0.0%			25.0%							20.0%		5.1%
Kayak		0.0%	0.0%							16.7%			3.0%
Ice Skates								33.3%	100.0%			50.0%	9.1%
Hockey Stick		0.0%			50.0%				100.0%			0.0%	20.8%
Golf Clubs				25.0%	33.3%								5.3%
Football Helmet			33.3%										2.8%
Fitness Tracker	16.7%		100.0%			33.3%			20.0%				8.5%
Dumbbells			33.3%	20.0%					50.0%		0.0%		9.4%
Cycling Helmet		100.0%	100.0%	33.3%	100.0%				25.0%			0.0%	18.2%
Climbing Rope		50.0%	66.7%			100.0%		25.0%	16.7%	33.3%			20.0%
Boxing Gloves												100.0%	5.1%
Basketball		0.0%							33.3%		50.0%		8.0%
Total	9.6%	8.6%	13.0%	8.8%	9.8%	8.0%	2.9%	5.7%	15.2%	4.5%	6.5%	12.8%	8.6%

General Conversion Trend:

- Throughout the year, conversion rates varied, with higher numbers of products converting successfully in months like September(15.2%) and March(13%). This suggests that while some products had strong seasonal peaks, there is potential to improve conversions in lower-performing months through targeted interventions.

Lowest Conversion Month:

- July experienced the lowest overall conversion rate at 2.9%, with no products standing out significantly in terms of conversion. This trend continues in August also This indicates a potential need to revisit marketing strategies or promotions during this period to boost performance.

Highest Conversion Rates:

- September recorded the highest overall conversion rate at 15.2%, driven significantly by the Ski Boots and Running Shoes and Basketball gloves with 100% conversion rate. March and December also shows good conversion rate which can be due to the seasonal offer or demand.

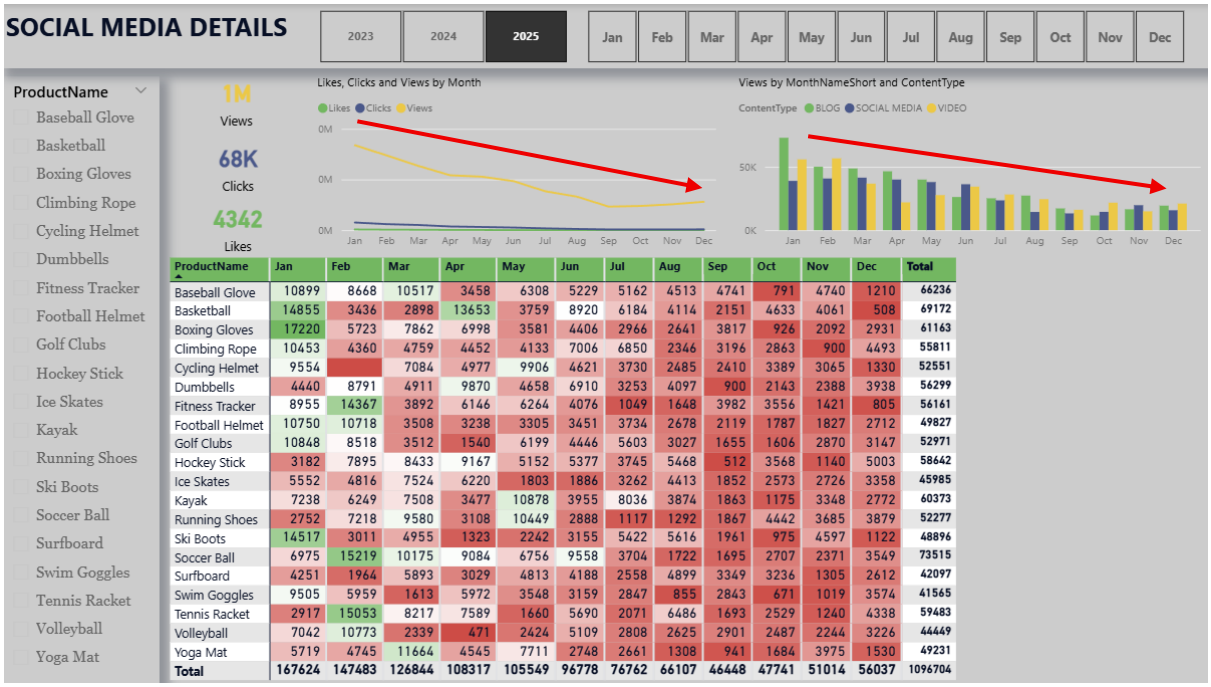
Low Performing Products

- Football helmet, volleyball, kayak, swim goggles these the poorly performing products need to work on these products.

High Performing Products

- Hockey Stick, Climbing rope, Surfboard, Cycling helmet, basketball gloves are the high performing products.

Reduced Customer Engagement



ProductName	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Total
Baseball Glove	10899	8668	10517	3458	6308	5229	5162	4513	4741	791	4740	1210	66236
Basketball	14855	3436	2898	13653	3759	8920	6184	4114	2151	4633	4061	508	69172
Boxing Gloves	17220	5723	7862	6998	3581	4406	2966	2641	3817	926	2092	2931	61163
Climbing Rope	10453	4360	4759	4452	4133	7006	6850	2346	3196	2863	900	4493	55811
Cycling Helmet	9554		7084	4977	9906	4621	3730	2485	2410	3389	3065	1330	52551
Dumbbells	4440	8791	4911	9870	4658	6910	3253	4097	900	2143	2388	3938	56299
Fitness Tracker	8955	14367	3892	6146	6264	4076	1049	1648	3982	3556	1421	805	56161
Football Helmet	10750	10718	3508	3238	3305	3451	3734	2678	2119	1787	1827	2712	49827
Golf Clubs	10848	8518	3512	1540	6199	4446	5603	3027	1655	1606	2870	3147	52971
Hockey Stick	3182	7895	8433	9167	5152	5377	3745	5468	512	3568	1140	5003	58642
Ice Skates	5552	4816	7524	6220	1803	1886	3262	4413	1852	2573	2726	3358	45985
Kayak	7238	6249	7508	3477	10878	3955	8036	3874	1863	1175	3348	2772	60373
Running Shoes	2752	7218	9580	3108	10449	2888	1117	1292	1867	4442	3685	3879	52277
Ski Boots	14517	3011	4955	1323	2242	3155	5422	5616	1961	975	4597	1122	48896
Soccer Ball	6975	15219	10175	9084	6756	9558	3704	1722	1695	2707	2371	3549	73515
Surfboard	4251	1964	5893	3029	4813	4188	2558	4899	3349	3236	1305	2612	42097
Swim Goggles	9505	5959	1613	5972	3548	3159	2847	855	2843	671	1019	3574	41565
Tennis Racket	2917	15053	8217	7589	1660	5690	2071	6486	1693	2529	1240	4338	59483
Volleyball	7042	10773	2339	471	2424	5109	2808	2625	2901	2487	2244	3226	44449
Yoga Mat	5719	4745	11664	4545	7711	2748	2661	1308	941	1684	3975	1530	49231
Total	**167624**	**147483**	**126844**	**108317**	**105549**	**96778**	**76762**	**66107**	**46448**	**47741**	**51014**	**56037**	**1096704**

Declining Views:

- Views peaked in starting of the year , it can be due to the new year resolution or seasonal offer and marketing but it's showing a steep declining after the month of January but in the month of aprill and may
- There can be slight increase in views after the September month

- **Low Interaction Rates:**
  - Clicks and likes remained consistently low compared to views, suggesting the need for more engaging content or stronger calls to action.
- **Content Type Performance:**
  - Blog content drove the most views, especially in January , while social media and video content maintained steady but slightly lower engagement.

Customer Feedback Analysis

ReviewDate	CustomerID	ReviewText	SentimentCategory	Rating
01 January 2025	85	Amazing value for the price.	Positive	5
03 January 2025	33	Five stars for the quick delivery.	Positive	5
03 January 2025	46	The quality is top-notch.	Positive	4
04 January 2025	91	Great purchase, very satisfied.	Positive	4
05 January 2025	48	The product is okay, but the instructions were unclear.	Negative	2
05 January 2025	85	Exceeded my expectations!	Positive	5
05 January 2025	93	Five stars for the quick delivery.	Positive	4
06 January 2025	29	Five stars for the quick delivery.	Positive	4
08 January 2025	63	Exceeded my expectations!	Positive	5
08 January 2025	98	Excellent product, highly recommend!	Positive	5
09 January 2025	1	Exceeded my expectations!	Positive	5
09 January 2025	5	I love this product, will buy again!	Positive	4
09 January 2025	30	Good quality, but could be cheaper.	Mixed Positive	3
09 January 2025	58	Shipping was fast and the item was well-packaged.	Positive	5
10 January 2025	27	Good quality, but could be cheaper.	Mixed Negative	2
11 January 2025	38	Exceeded my expectations!	Positive	5
11 January 2025	65	Not worth the money.	Negative	2
12 January 2025	92	Product did not meet my expectations.	Negative	2
13 January 2025	77	Product did not meet my expectations.	Negative	2
14 January 2025	46	Customer support was very helpful.	Positive	5
14 January 2025	66	Amazing value for the price.	Positive	5
15 January 2025	11	Shipping was fast and the item was well-packaged.	Neutral	3

Customer Ratings Distribution:

- The majority of customer reviews are in the higher ratings, with 154 reviews at 4 stars and 138 reviews at 5 stars, indicating overall positive feedback. Lower ratings (1-2 stars) account for a smaller proportion, with 27 reviews at 1 star and 60 reviews at 2 stars.

Sentiment Analysis:

- Positive sentiment dominates with 292 reviews, reflecting a generally satisfied customer base. Negative sentiment is present in 81 reviews, with a smaller number of mixed and neutral sentiments, suggesting some areas for improvement but overall strong customer approval.
- Sudden spike in negative sentiment in the month of May, august and October need to determine the root cause and resolve it

Opportunity for Improvement:

- The presence of mixed positive and mixed negative sentiments suggests that there are opportunities to convert those mixed experiences into more clearly positive ones, potentially boosting overall ratings. Addressing the specific concerns in mixed reviews could elevate customer satisfaction.