Digital Marketing Analytics
Vivian Nguyen
October 2024

Project Overview

Mission Statement

The goal of this project is to build a practical database that marketers can use to evaluate how consumers behave and respond to digital marketing campaigns. By analyzing this data, marketers can make more informed decisions about how to allocate ad spend, boost conversion rates, and improve click-through rates at key stages of the customer journey.

Mission Objectives

1. Campaign Effectiveness: Analyze the different marketing channels and campaign types to determine which leads to the highest conversion rates.
2. Customer Engagement Analysis: Explore how engagement metrics, like social shares and email clicks, impact conversions and overall customer behavior.
3. Customer Segmentation: Group consumers by demographic factors, like age and income, to analyze their behavior and gain valuable insights for better marketing strategies.

Purpose

This project focuses on using SQL to analyze marketing data and extract actionable insights that can help businesses understand their customers to make informed decisions. By simulating the role of a marketing manager, I aim to deepen my understanding of how data-driven strategies influence marketing outcomes.

Dataset Overview

The data used in this project was sourced from Kaggle and provides a detailed view of customer interactions with digital marketing campaigns. It includes key information such as customer demographic data, engagement metrics, and marketing-specific performance indicators. This dataset is valuable for analyzing how customers respond to different campaigns, helping to uncover patterns and insights that businesses can use to improve their marketing strategies.

Database Management Section

Trigger	
Business Purpose	This trigger ensures that each time a customer completes a purchase (conversion), their loyalty points are updated based on the current point system of 2 points per $1 spent. The loyalty points are calculated directly within the trigger rather than pre-calculated before the insert statement. This ensures that even if the points system changes later, marketers would not have to modify the insert logic elsewhere. This way, marketers can handle the point system dynamically and avoid hard-coding it in the insert logic. Although placeholder values were used in the insert statement, these values would be dynamically calculated or retrieved from other tables or business processes in a production environment. For the scope of this project, I settled with placeholder values. Overall, while simplified, this trigger demonstrates the process for automating loyalty updates in real-world marketing systems.

SQL Query Text	<pre>CREATE TABLE orders (
	    id NUMBER GENERATED BY DEFAULT AS IDENTITY PRIMARY KEY,
	    customerID NUMBER,
	    date TIMESTAMP NOT NULL,
	    channel VARCHAR2(25) NOT NULL, -- Indicates whether an order was made online or in-store
	    productID NUMBER NOT NULL,
	    quantity NUMBER NOT NULL,
	    total NUMBER NOT NULL,
	    loyaltyPoints NUMBER NOT NULL, -- Based on a point system of 2 points per 1 dollar spent
	    referralSource VARCHAR2(25) NOT NULL, -- Captures how a customer heard about the company or its products
	    FOREIGN KEY(customerID) REFERENCES customers(id),
	    CONSTRAINT check_channel_input CHECK(channel IN ('online', 'in-store'))
	);
	
	CREATE TRIGGER update_loyalty_points
	AFTER INSERT ON orders
	FOR EACH ROW
	DECLARE
	    customer_exists NUMBER; -- Declare as a number instead of BOOLEAN to store the COUNT
	BEGIN
	    -- Check if the customer exists in the conversions table
	    SELECT COUNT(*) INTO customer_exists -- Store the result of COUNT into the variable
	    FROM conversions
	    WHERE customerID = :NEW.customerID;
	
	    IF customer_exists > 0 THEN -- If the customer exists, proceed with update
	        -- Update loyalty points if customer exists
	        UPDATE conversions
	        SET loyaltyPoints = loyaltyPoints + (:NEW.total * 2)
	        WHERE customerID = :NEW.customerID;
	    ELSE
	        -- Insert a new customer record into conversions
	        INSERT INTO conversions(customerID, campaignID, previousPurchases, loyaltyPoints, conversion)
	        VALUES (:NEW.customerID, 1, 0, :NEW.total * 2, 0); -- Include other necessary columns
	    END;
	END;</pre>

Insights	This trigger ensures that loyalty points are automatically updated whenever a customer makes a new purchase, reinforcing the brand's customer

	retention strategy. By automating this process, marketers can maintain an accurate record of loyalty points, which is critical for delivering personalized offers and increasing long-term customer engagement. The dynamic calculation also ensures scalability as marketing strategies evolve.

Customer Engagement Analysis

Query A	Weighted Engagement Score
Business Purpose	This query exemplifies how a weighted engagement score is calculated. The weights are assigned accordingly to the engagement metrics depending on the campaign type before its correlation to conversion is analyzed. This is more insightful as opposed to a consistent engagement score calculated across all marketing campaigns. Engagement metrics hold weight differently in each stage of the marketing funnel. For example, website visits and time on site might be more critical in the earlier stages where the focus is to gauge the interest of the customers. Given later stages prioritize action and building long-term relationships with customers, metrics like email clicks and social shares will naturally be more important. Moreover, the metrics were scaled since each is measured on a different scale. This ensures the larger numbers would not disproportionately affect the final engagement score.
Data Output Expected	This view calculates an engagement score specifically for channels within the Retention stage of the marketing funnel. The select statement then queries each channel along with its respective engagement score and average conversion rate.

| **SQL Query Text** | ```sql
CREATE VIEW retentionEngagement AS
SELECT
    c.id,
    c.channel,
    -- Scaled and weighted engagement metrics
    ((e.emailOpens / 19 * .3) +
     (e.emailClicks / 9 * .3) +
     (e.socialShares / 99 * .2) +
     (e.websiteVisits / 49 * .1) +
     (e.timeOnSite / 15 * .1)) AS engagementScore,
    c.conversionRate,
    c.adSpend
FROM campaigns c
JOIN engagement e ON c.id = e.campaignID
WHERE c.type = 'Retention'
ORDER BY 1, 2;

SELECT
    channel,
    ROUND(AVG(engagementScore), 2) AS avgEngagementScore,
    ROUND(AVG(conversionRate) * 100, 2) AS avgConversionRate
FROM retentionEngagement
GROUP BY channel
ORDER BY 2 DESC;
``` |
| **SQL Query Screenshot** | |

	ID	CHANNEL	ENGAGEMENTSCORE	CONVERSIONRATE	ADSPEND
1	1198	Email	0.5131256648113790970933828076685219542363	0.166	3809.74
2	1199	Email	0.6939604205318491032776747062461348175634	0.0185	9856.94
3	1200	Email	0.4948464646464646464646464646464646464646	0.027	8325.37
4	1201	Email	0.7923290455576169861884147598433312719028	0.1468	5062.99
5	1202	Email	0.6323885753344399960941314324772971389513	0.0585	2818.19

	CHANNEL	AVGENGAGEMENTSCORE	AVGCONVERSIONRATE
1	Social Media	0.5	10.16
2	PPC	0.5	10.41
3	SEO	0.5	10.45
4	Referral	0.5	10.39
5	Email	0.5	10.17

Insights	Interestingly enough, the engagement score is consistent across all campaign channels. However, there are slight differences in each channel's average conversion rate. This either tells marketers that the engagement score requires recalculating and/or that customers interact consistently with this company's ad regardless of the channel. However, marketers must identify other factors that are influencing conversion through running further analysis.

Query B	Time-on-Site Thresholds to Conversion
Business Purpose	This query breaks down time-on-site into bins. This adds more value as opposed to calculating the average time on site across the marketing campaign channels. Given there are 8000 records, data binning makes it easier for marketers to identify trends in the data and is also less sensitive to outliers. For instance, if it's noticeably clear that one or two of the intervals were shown to have significantly higher average conversion rates, marketers can use this to optimize website layout or content to better match how long a customer preferably wanders through the site to better improve customer retention.
Data Output Expected	This query outputs each time-on-site interval with its respective total count and average conversion rate.
SQL Query Text	<pre>WITH timeInterval AS (
SELECT
campaignID,
CASE
WHEN timeOnSite BETWEEN 0 AND 3 THEN '0-3'
WHEN timeOnSite BETWEEN 4 AND 7 THEN '4-7'
WHEN timeOnSite BETWEEN 8 AND 11 THEN '8-11'
ELSE '11-15'
END AS interval
FROM engagement
)
SELECT
t.interval,
COUNT(*),
ROUND(AVG(c.conversionRate) * 100, 2) AS avgConversionRate
FROM timeInterval t
JOIN campaigns c ON t.campaignID = c.id
GROUP BY t.interval
ORDER BY 1;</pre>
SQL Query Screenshot	

	INTERVAL	COUNT(*)	AVGCONVERSIONRATE
1	0-3	1411	10.48
2	11-15	3380	10.43
3	4-7	1648	10.3
4	8-11	1561	10.56

Insights	The result set indicates to marketers that there is essentially one of two groups that browse a company's ad. The first group is users who

	accidentally came across the ad and clicked out of pure curiosity despite not caring much at all for the product or service advertised. The second group is potential customers who genuinely are interested in the product or service and want to learn more about it. This is based on the fact that the two highest average conversion rates are for the intervals with the quickest time-on-site (0-3, 10.48%) and the longest time-on-site (8-11, 10.56%).

Customer Segmentation Analysis

Query F	Generation and Income Class Grouping
Business Purpose	For marketing purposes, the grouping of the age and income columns in the customers table allows for easier marketing analysis and strategic decision-making. Categorizing customers this way is useful because it allows marketers to tailor campaigns more effectively, as each group tends to have shared behaviors, preferences, and values. Creating a view based on these groupings is effective especially since it'd likely be frequently used to track customer engagement or conversion rates by demographics.
Expected Data Output	This view selects all the data from the customers table but includes two extra columns that group the data from the age and income columns into generations and income classes, respectively.
SQL Query Text	(see code below)

```sql
CREATE VIEW customerSegView AS
SELECT
    id,
    gender,
    age,
    -- Partition age in generations
    CASE
        WHEN age BETWEEN 60 AND 78 THEN 'boomers'
        WHEN age BETWEEN 44 and 59 THEN 'genX'
        WHEN age BETWEEN 28 AND 43 THEN 'millenials'
        WHEN age BETWEEN 12 AND 27 THEN 'genZ'
        ELSE 'other' -- Adds a fallback in case the age doesn't match
    END AS generation,
    income,
    -- Partition income in separate income classes
    CASE
        WHEN income <= 30000 THEN 'low'
        WHEN income > 30000 AND income <= 60000 THEN 'lowerMiddle'
        WHEN income > 60000 AND income <= 100000 THEN 'upperMiddle'
        WHEN income > 100000 THEN 'high'
        ELSE 'other' -- Adds a fallback in case the income doesn't match
    END AS incomeClass
FROM customers;
```

SQL Query Screenshot						
		⬍ ID	⬍ GENDER	⬍ AGE	⬍ GENERATION	⬍ INCOME	⬍ INCOMECLASS

	ID	GENDER	AGE	GENERATION	INCOME	INCOMECLASS
1	8000	Female	56	genX	136912	high
2	8001	Male	69	boomers	41760	lowerMiddle
3	8002	Female	46	genX	88456	upperMiddle
4	8003	Female	32	millenials	44085	lowerMiddle
5	8004	Female	60	boomers	83964	upperMiddle
6	8005	Female	25	genZ	42925	lowerMiddle
7	8006	Female	38	millenials	25615	low
8	8007	Female	56	genX	57083	lowerMiddle
9	8008	Female	36	millenials	140788	high
10	8009	Male	40	millenials	130764	high

Query G	Percentage Breakdown by Generation
Business Purpose	This query calculates the percentage of each generation that engages with different marketing channels as well as their average conversion rate. This query is particularly useful because it shows marketers which generation(s) comprise the majority of a marketing campaign channel's audience and whether the campaign does an effective job of converting them. Knowing this will provide valuable insights into how marketing campaigns should be planned and executed to effectively reach different target demographics. NOTE: Using percent as opposed to COUNT() counteracts the uneven distribution of generations within the dataset. The query effectively executes this statement, "For each marketing channel, show me the percentage of each generation based on the total number of customers in that generation."
Expected Data Output	This query outputs the campaign channel combined with each generation along with their respective distribution percentage and average conversion rate, but only if they match the Conversion stage.

SQL Query Text	```WITH genCTE AS (
SELECT
generation,
COUNT(generation) AS genCount
FROM customerSegView
GROUP BY generation
)
SELECT
c.channel,
v.generation,
ROUND(COUNT(v.generation)/g.genCount * 100, 2) AS percentage,
ROUND(AVG(c.conversionRate) * 100, 2) AS avgConversionRate
FROM campaigns c
JOIN engagement e ON c.id = e.campaignID
JOIN customerSegView v ON e.customerID = v.id
JOIN genCTE g ON v.generation = g.generation
WHERE c.type = 'Conversion'
GROUP BY c.channel, v.generation, g.genCount
ORDER BY 1, 3 DESC;```
SQL Query Screenshot	<table><thead><tr><th></th><th>CHANNEL</th><th>GENERATION</th><th>PERCENTAGE</th><th>AVGCONVERSIONRATE</th></tr></thead><tbody><tr><td>2</td><td>Email</td><td>millenials</td><td>5.5</td><td>10.36</td></tr><tr><td>3</td><td>Email</td><td>boomers</td><td>5.09</td><td>10.04</td></tr><tr><td>4</td><td>Email</td><td>genX</td><td>4.61</td><td>10.92</td></tr><tr><td>5</td><td>PPC</td><td>genZ</td><td>5.77</td><td>9.34</td></tr><tr><td>6</td><td>PPC</td><td>genX</td><td>5.72</td><td>11.14</td></tr><tr><td>7</td><td>PPC</td><td>millenials</td><td>5.46</td><td>9.75</td></tr><tr><td>8</td><td>PPC</td><td>boomers</td><td>5.41</td><td>9.72</td></tr><tr><td>9</td><td>Referral</td><td>boomers</td><td>6.06</td><td>10.8</td></tr><tr><td>10</td><td>Referral</td><td>genX</td><td>6.05</td><td>10.52</td></tr><tr><td>11</td><td>Referral</td><td>genZ</td><td>5.37</td><td>9.02</td></tr><tr><td>12</td><td>Referral</td><td>millenials</td><td>4.91</td><td>10.46</td></tr><tr><td>13</td><td>SEO</td><td>boomers</td><td>6.12</td><td>9.76</td></tr><tr><td>14</td><td>SEO</td><td>millenials</td><td>5.15</td><td>9.92</td></tr><tr><td>15</td><td>SEO</td><td>genZ</td><td>4.76</td><td>10.93</td></tr><tr><td>16</td><td>SEO</td><td>genX</td><td>4.36</td><td>10.39</td></tr><tr><td>17</td><td>Social Media</td><td>genX</td><td>5.02</td><td>11.58</td></tr><tr><td>18</td><td>Social Media</td><td>millenials</td><td>4.68</td><td>11.34</td></tr><tr><td>19</td><td>Social Media</td><td>genZ</td><td>4.21</td><td>11.49</td></tr><tr><td>20</td><td>Social Media</td><td>boomers</td><td>4.12</td><td>10.68</td></tr></tbody></table>
Insights	Older generations are less tech-savvy compared to their younger counterparts. This explains why boomers make up the smallest percent of social media marketing, but are the highest in referral marketing. This indicates to marketers that word-of-mouth marketing is most effective on boomers.

Query C	Ad Spend Efficiency
Business Purpose	This query examines cost per conversion by dividing total ad spend by the total number of conversions across each marketing campaign channel and type. By combining this with the average conversion rate, marketers can see how successful a campaign is and the cost it takes to achieve that success. If there's a mismatch such as high ad spend but poor conversion rate, marketers will know what queries to run to figure out the issue.
Data Output Expected	This query outputs the channel, its type, the total count of marketing campaigns within those two categories, the cost per conversion, and the average conversion rate.

SQL Query Text	<pre>SELECT
	    cl.channel,
	    cl.type,
	    COUNT(*) AS totalCampaigns,
	    -- Cost per conversion = total ad spend / total conversion
	    ROUND(SUM(cl.adSpend)/SUM(c2.conversion), 2) AS costPerConversion,
	    ROUND(AVG(cl.conversionRate) * 100, 2) AS avgConversionRate
	FROM campaigns cl
	JOIN conversions c2 ON cl.id = c2.campaignID
	GROUP BY cl.channel, cl.type
	ORDER BY 1, 2;</pre>

SQL Query Screenshot	(see table below)

	CHANNEL	TYPE	TOTALCAMPAIGNS	COSTPERCONVERSION	AVGCONVERSIONRATE
1	Email	Awareness	378	5802.15	10.42
2	Email	Consideration	403	5803.34	10.67
3	Email	Conversion	416	5543.01	10.66
4	Email	Retention	360	6163.33	10.17
5	PPC	Awareness	412	5810.15	10.87
6	PPC	Consideration	404	5758.74	10.3
7	PPC	Conversion	447	5137.65	10.1
8	PPC	Retention	392	5843.19	10.41
9	Referral	Awareness	431	5827.46	10.52
10	Referral	Consideration	407	5852.58	9.99
11	Referral	Conversion	445	5377.27	10.29
12	Referral	Retention	436	5787.65	10.39

Insights	Across conversion campaigns, social media leads with the highest average conversion rate of 11.33 percent despite being the third highest for cost per conversion of $5,264 and lowest in total campaigns. These statistics indicate to marketers that social media conversion campaigns have the best balance of ad spend and conversions. The campaign count could either mean fewer marketing efforts are put towards social media conversion campaigns than there should be or the complete opposite — because there

	are fewer of these campaigns more marketing efforts are directed towards each one.

Query D	Conversion Rate
Business Purpose	This query uses a correlated subquery to compare each campaign's conversion rate to the average within its channel and type. This helps marketers focus on specific campaigns to identify the factors behind their success. By comparing campaigns within the same channel, marketers can make more targeted decisions without relying on a broad, less informative overall average.
Expected Data Output	This query outputs only the marketing channels that have an above-average conversion rate within its category, specifically specifying the channel, its type, and its conversion rate.
| **SQL Query Text** | ```sql
SELECT
    channel,
    type,
    (conversionRate * 100) AS conversionRate
FROM campaigns c1
WHERE c1.conversionRate > (
    -- Calculates the avg conversion rate for each combination of channel and type
    SELECT AVG(c2.conversionRate)
    FROM campaigns c2
    WHERE c1.channel = c2.channel -- Correlate by the channel
    AND c1.type = c2.type -- Correlate by the type
)
ORDER BY 1, 2;
``` |
| **SQL Query Screenshot** | |

	CHANNEL	TYPE	CONVERSIONRATE
1	Email	Awareness	16.62
2	Email	Awareness	14.85
3	Email	Awareness	18.02
4	Email	Awareness	11.12
5	Email	Awareness	18
6	Email	Awareness	19.68
7	Email	Awareness	14.2
8	Email	Awareness	10.48
9	Email	Awareness	19.21
10	Email	Awareness	15.59
11	Email	Awareness	13.8
12	Email	Awareness	15.94

Insights	This analysis highlights which campaigns outperform within their channels. By identifying the top campaigns, marketers can explore which creative, targeting, or other factors contributed to higher conversion rates and use these insights to improve other campaigns in the same channel. This

	focused comparison ensures better allocation of resources by scaling the campaigns that are already performing well.

Query E	Bottom 10% Campaigns Based on Click-Through Rate
Business Purpose	This query identifies the bottom 10 percent of campaigns with each channel and type based on its CTR, specifically in the Awareness and Consideration stages. CTR is a critical metric especially since CTR helps gauge initial engagement and interest, both of which the Awareness and Consideration stages primarily focus on. The conversion rate, on the other hand, measures the success of turning that interest into desired actions. This makes it more vital in the Conversion and Retention stages. Because this query provides a snapshot of poorly performing campaigns, there are many highly valuable insights marketers can gain to better inform business decisions. For example, marketers can do A/B testing to correctly identify factors that prevent campaigns from being successful, such as bad website design. This, in turn, allows marketers to reallocate budget and direct marketing efforts where it matters.
Expected Data Output	This query outputs all the campaign IDs, their channel and type, and CTR within the Awareness and Consideration stages that fall within the bottom 10 percent of their categories based on CTR alone.
| **SQL Query Text** | ```sql
WITH rankedCampaigns AS (
    SELECT
        id,
        channel,
        type,
        clickThroughRate AS CTR,
        ROUND(PERCENT_RANK() OVER (
            PARTITION BY channel, type ORDER BY clickThroughRate ASC
        ), 2) AS ctrRank
    FROM campaigns
    WHERE type IN ('Awareness', 'Consideration')
)
SELECT *
FROM rankedCampaigns
WHERE ctrRank >= 0.90; -- Bottom 10%
``` |

		SQL Query Screenshot			
SQL Query Screenshot					

	ID	CHANNEL	TYPE	CTR	CTRRANK
1	353	Email	Awareness	0.2674	0.9
2	207	Email	Awareness	0.269	0.9
3	363	Email	Awareness	0.2693	0.9
4	282	Email	Awareness	0.2693	0.9
5	371	Email	Awareness	0.2713	0.91
6	88	Email	Awareness	0.2715	0.91
7	178	Email	Awareness	0.2716	0.91
8	16	Email	Awareness	0.2746	0.92
9	10	Email	Awareness	0.2749	0.92
10	53	Email	Awareness	0.2754	0.92
11	308	Email	Awareness	0.2755	0.92
12	299	Email	Awareness	0.2757	0.93

Insights

The result set encompasses all marketing campaigns within the awareness and consideration stages of the marketing funnel with a CTR between 26 and 30 percent only. This indicates to marketers that they can use this as a benchmark to measure future campaigns to determine whether they're successful in attracting potential customers. Anything that falls below 30 percent means they're low-performing campaigns.