**Customer Personality Analysis**

**1. Introduction**

The purpose of this analysis was clustering our customers through some segments. The data set comprised information about customers, their purchasing behaviour, and engagement with promotional campaigns. The goal was to derive actionable insights that could help a business and understand it’s customers and make it easier for them to modify products according to the specific needs, behaviours and concerns of different types of customers.

**About the dataset**:

The dataset from a groceries firm's database on the customer's records consists of **2240** data points and **29** attributes. It can be categorised into the following subsets:-

**People**

**ID**: Customer's unique identifier

**Year\_Birth**: Customer's birth year

**Education**: Customer's education level

**Marital\_Status**: Customer's marital status

**Income**: Customer's yearly household income

**Kidhome**: Number of children in customer's household

**Teenhome**: Number of teenagers in customer's household

**Dt\_Customer**: Date of customer's enrollment with the company

**Recency**: Number of days since customer's last purchase

**Complain**: 1 if the customer complained in the last 2 years, 0 otherwise

**Products**

**MntJuiceProducts**: Amount spent on Juice in last 2 years

**MntFruits**: Amount spent on fruits in last 2 years

**MntMeatProducts**: Amount spent on meat in last 2 years

**MntFishProducts**: Amount spent on fish in last 2 years

**MntSweetProducts**: Amount spent on sweets in last 2 years

**MntGoldProds**: Amount spent on gold in last 2 years

**Promotion**

**NumDealsPurchases**: Number of purchases made with a discount

**AcceptedCmp1**: 1 if customer accepted the offer in the 1st campaign, 0 otherwise

**AcceptedCmp2**: 1 if customer accepted the offer in the 2nd campaign, 0 otherwise

**AcceptedCmp3**: 1 if customer accepted the offer in the 3rd campaign, 0 otherwise

**AcceptedCmp4**: 1 if customer accepted the offer in the 4th campaign, 0 otherwise

**AcceptedCmp5**: 1 if customer accepted the offer in the 5th campaign, 0 otherwise

**Response**: 1 if customer accepted the offer in the last campaign, 0 otherwise

**Place**

**NumWebPurchases**: Number of purchases made through the company’s website

**NumCatalogPurchases**: Number of purchases made using a catalogue

**NumStorePurchases**: Number of purchases made directly in stores

**NumWebVisitsMonth**: Number of visits to company’s website in the last month

**2. Data Preparation**

The initial step involved cleaning and transforming the data to ensure its usability for analysis. The following tasks were performed:

**Data Cleaning:**   
1. Missing values handled by Excel Duplicates were identified and removed.

2. Used SQL queries to remove the outliers in Income and Year\_Birth

DELETE FROM marketing\_campaign WHERE Income = 666666   
DELETE FROM marketing\_campaign WHERE Year\_Birth IN ('1893','1899','1900')

**Data Transformation:**

The dataset was normalised to ensure consistency in formats. Dates were converted to standard formats, and categorical data was encoded where necessary. We also used SQL to convert the datatypes of columns to be consistent.  
We converted **Kidhome**, **Teenhome**, **Recency**, **NumDealsPurchases**, **NumWebPurchases**, **NumCatalogPurchases**, **NumStorePurchases**, **NumWebVisitsMonth**, **AcceptedCmp1**, **AcceptedCmp2**, **AcceptedCmp3**, **AcceptedCmp4**, **AcceptedCmp5**, **Complain** and **Response** from **varchar(50)** to **int**.

also, **Income**, **MntJuiceProducts**, **MntFruits**, **MntMeatProducts**, **MntFishProducts**, **MntSweetProducts**, **MntGoldProds** from **varchar(50)** to **float**.

And finally **Dt\_Customer** from **varchar(50)** to **date**. Using this query  
ALTER TABLE marketing\_campaign

ALTER COLUMN [column name] [converted datatype]

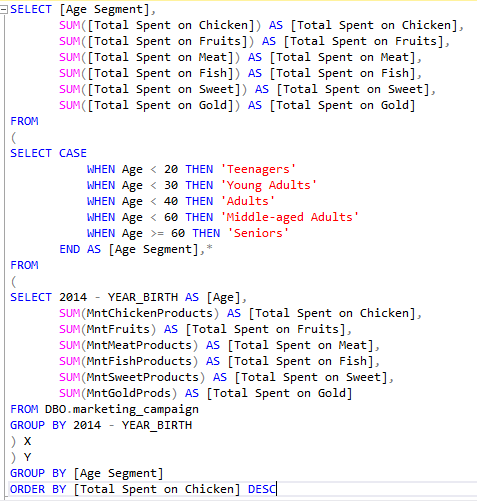
**Calculations:**

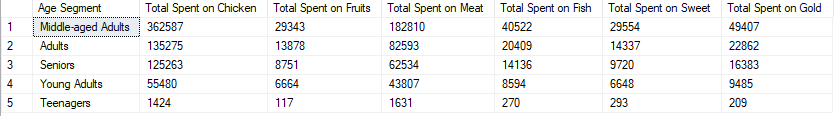
Once the data was cleaned, We had some questions to answer using SQL server queries.

1. What are the main purchasing habits across different product categories?

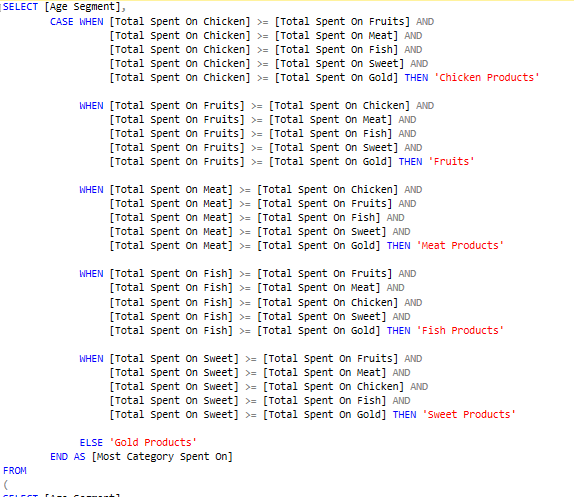
We used this query to calculate the sum of each category and grouped customers by age

segment to highlight the total spent of each category of each age segment.





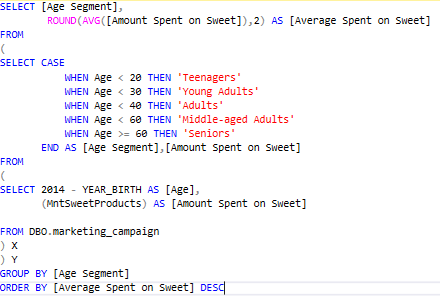
We wanted to see what the most product category that customer has spent by age segment, so we added subquery to the previous query.

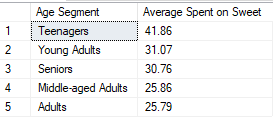


****

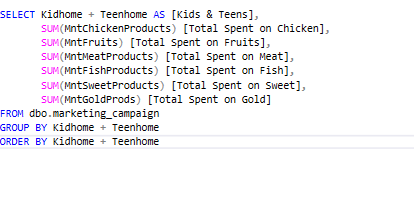
2. What is the average spent on Sweet products by customer Age segment?

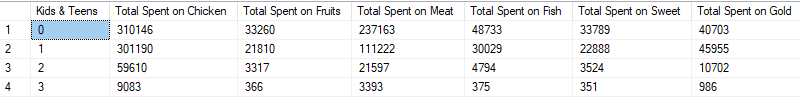
We created a subquery calculating age of the customer then grouping customers by age segments ending with calculating average of each segment.

****

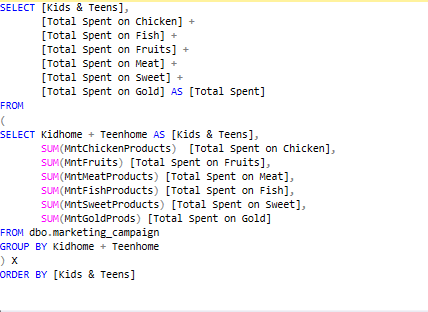
****

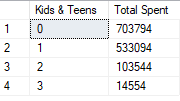
3. Do customers with children and teenagers have distinct purchasing behaviours?

Actually we added **Kidhome** and **Teenhome** into one column to calculate the total number of household’s kids and teens then calculating the total spent of each product category ending with grouping by the new column we have made.  
  


****

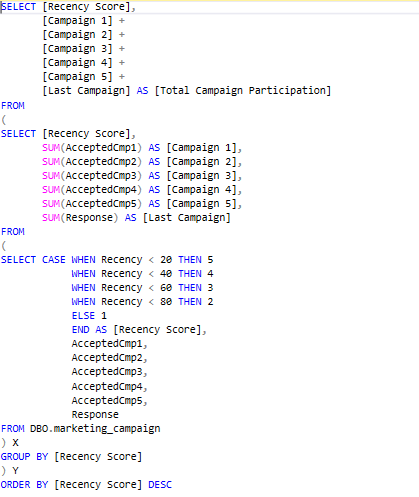
We used a previous query to calculate the overall total spent.

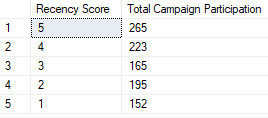
****

****

4. How do recent purchases correlate with campaign acceptance?

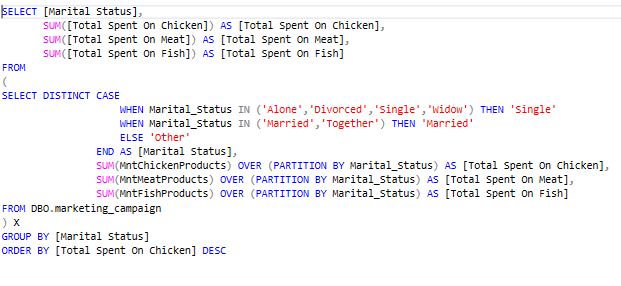
We noticed that the minimum value of recency is **0** and the maximum is **99**, so we calculated a recency score in which from 1 to 5 (Higher score will be better recency) then we summed up each campaign according to recency score.

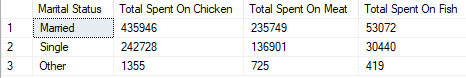
****

****

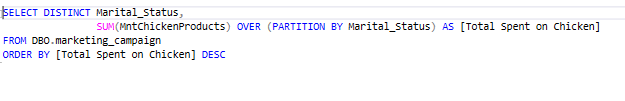
5. How does marital status and education level influence purchasing preferences?

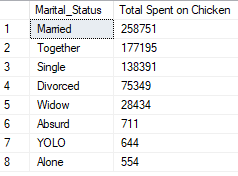
We focused on Juice, Meat and Fish to calculate total spent by marital status, also grouped Alone, Divorced, Single and Widow into Single status and Married, Together into Married status and YOLO, Absurd into Other.



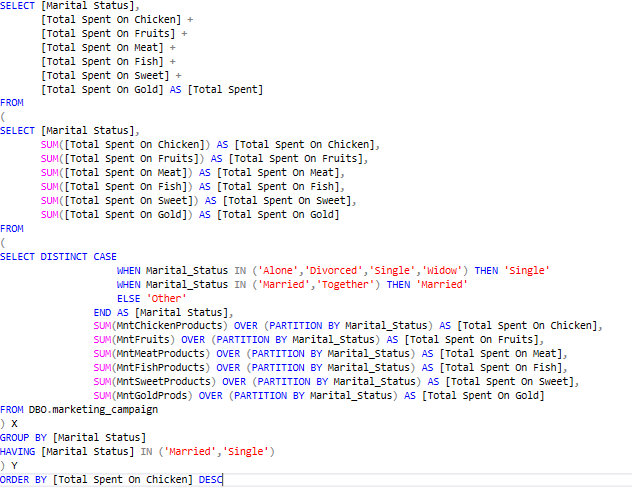


Another way to see the data of marital status is by calculating total spent on Juice for all marital status. We chose this approach because the Juice is the most product category customers spent on.



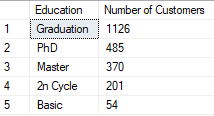
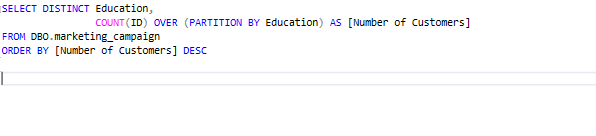


We calculated the total spent of all product categories by marital status, also grouped Alone, Divorced, Single and Widow into Single status and Married, Together into Married status and YOLO, Absurd into Other.

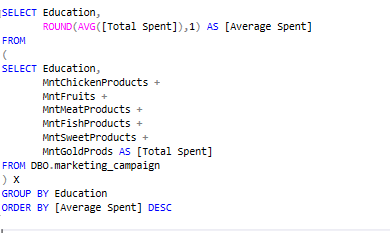


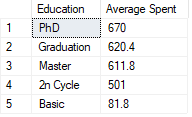


Let’s take Education insights, we calculated the total number of customers of each Education segment.



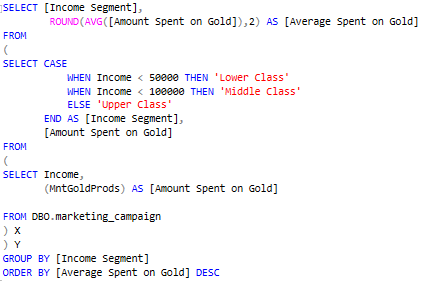
Here calculated the average spent of all categories by Education segment.

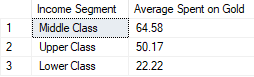




6. What is the average spent on Gold products by customer Income segment?

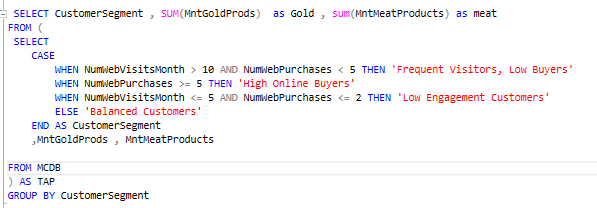
We created a subquery calculating Income of the customer then grouping customers by Income segments ending with calculating average of each segment.



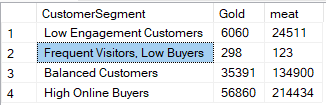


**7.Are there distinct customer segments based on complaint behavior?**

Purpose: Determine whether higher-income customers have different spending patterns or are more likely to purchase premium products like Meat and gold

**Query**  :   
   


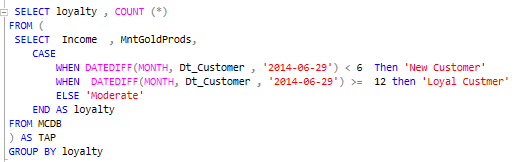
**Result :**



8.How do enrollment dates correlate with customer loyalty and spending?

* Purpose: Analyse if long-term customers (based on enrollment date) spend more, complain less, or respond differently to campaigns compared to newer customers.

**Query :**

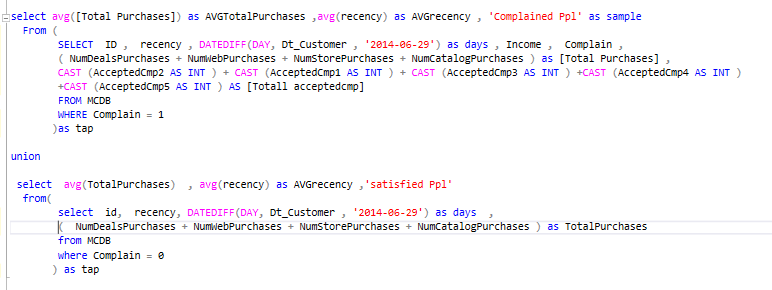


**Result :**



9.Are there distinct customer segments based on complaint behavior?

* **Purpose**: Find out if customers who have filed complaints in the last two years behave differently in terms of loyalty, spending, or campaign responsiveness.  
    
    
    
    
    
  **Query :**

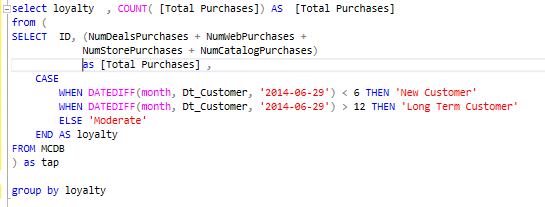


**Result :**

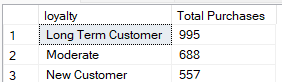


10. What segments of customers show the highest potential for upselling and cross-selling?

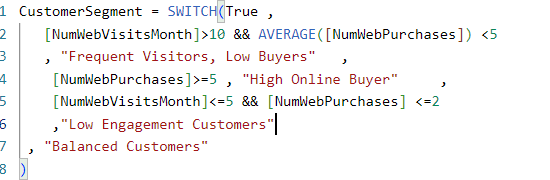
* **Purpose**: Identify which customer groups have high purchasing power but haven't yet maximized their potential in certain product categories, like gold or Meat.  
    
    
    
   **Query :**

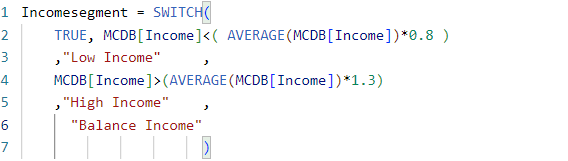


**Result :**

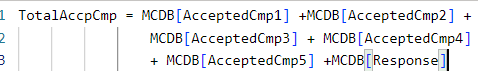


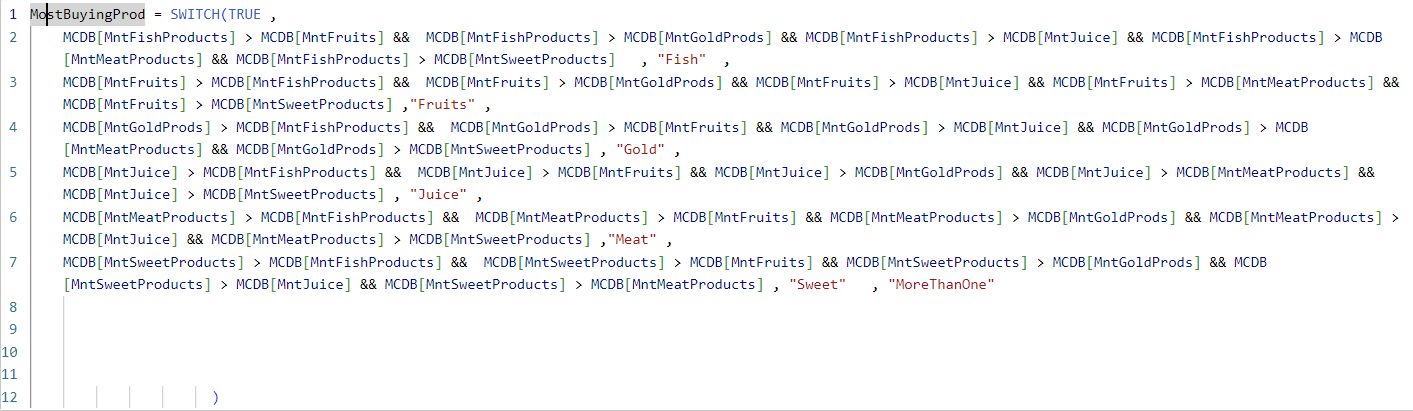
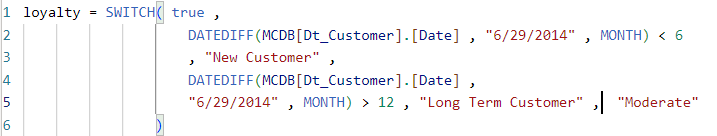
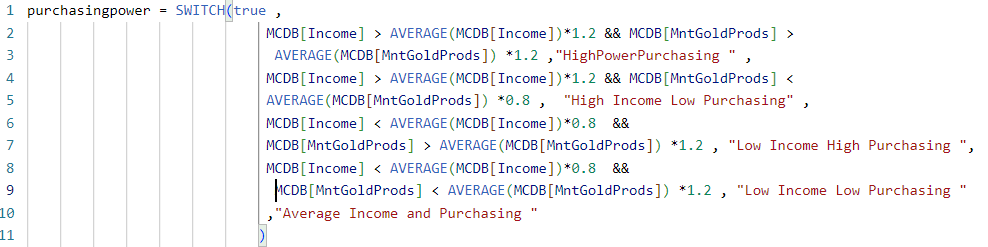
After SQL analysis, we moved to Power BI to do some Data transformation and Data visualisation.

* **DAX Calculations:** Using DAX (Data Analysis Expressions), the following key metrics were calculated:
  + **[Metric 1]** (TotalNumProducts): [addition of this columns :( MntJuice , MntFruits, MntMeatProducts, MntFishProducts, MntSweetProducts, MntGoldProds)]
  + **[Metric 2]** (TotalNumPurchases): [addition of this columns :( NumDealsPurchases + NumWebPurchases + NumCatalogPurchases+ NumStorePurchases + NumWebVisitsMonth )]
  + **[Metric 3]** (CustomerSegment):
  + 
  + **[Metric 4]** (Incomesegment):



* + **[Metric 5]** (TotalAccpCmp):



* + **[Metric 6]** (MostBuyingProd):   
      
      
      
    
  + **[Metric 7]** (loyalty):  
      
      
      
    
  + **[Metric 8]** (purchasingpower):  
      
    

**3. Challenges Encountered**

Throughout the process, I encountered several challenges:

* **Data-Related Challenges:** Some data points were on wrong Format, which required changing data format like : Dates were converted to standard formats and some of data the year were two digit only Additionally, the dataset had inconsistencies in format that required significant preprocessing Do Through Excel and SQL Server.
* **Visualization Challenges:** Visualizing the data was tricky due to Cluster first Customer on some certain Behaviors To resolve this, I used Power BI (DAX)
* **Storytelling Challenges:** Presenting the story in a clear, concise way was challenging due to the complexity of the findings. However, I used Power BI to address this.

**4. Findings and Insights**

After conducting the analysis, I identified the following key insights:

**Customer Distribution by Income Segment**:

* The majority of customers fall under the "Balance Income" group (859), followed by "Low Income" (812), and "High Income" (569).

**Product Sales by Category**:

* Juice is the most sold product, with 1,507 units sold, followed by meat (433 units). The least sold products are fruits (11 units) and sweets (14 units).

**Purchases by Income Segment**:

* Customers in the "Balance Income" segment account for 44.41% of purchases, followed by "High Income" (35.53%) and "Low Income" (20.06%).

**Gold Product Sales by Income Segment**:

* "Balance Income" customers bought the highest amount of gold products (42K units), followed closely by "High Income" (41K), while "Low Income" customers purchased only 16K units.

**Customer Loyalty and Web Visits**:

* Long-term customers visit the website the most (995 visits on average), while new customers make the fewest visits (557 visits). Long-term customers also have a higher purchasing rate.

**Campaign Success by Loyalty**:

* Long-term customers accepted the last campaign the most (219 customers), while new customers showed the lowest campaign acceptance (48 customers).

**Purchases by Purchasing Power**:

* Customers with "Average Income and High Purchasing Power" dominate the purchases (38.82%), while "Low Income, High Purchasing Power" represents only 2.36% of the total.

**Campaign Acceptance by Purchasing Power**:

* "High Power, High Income" customers accepted the campaign the most (332), followed by "Average Income and Purchasing Power" (302). "Low Income, Low Purchasing Power" accepted the least (21).

**Total spent by Customer’s household (Kids & Teens)**:

* With **0** kids**/**teens customers tends to spend more **702K**,customers with **1** kid**/**teen spend **533K**, customers with **2** kids**/**teens spend **103K**, and with **3** kids**/**teens customers spend **15K**.

**Campaign Participation by Recency Score**:

* With **5** Recency Score, the Total Campaign participation is **265**. With **4** Recency Score, the Total Campaign participation is **223**. With **3** Recency Score, the Total Campaign participation is **165**. With **2** Recency Score, the Total Campaign participation is **195**. With **1** Recency Score, the Total Campaign participation is **152**.

**5. Conclusion**

The analysis reveals key insights into customer behaviors and purchasing patterns for a grocery firm. Customers are segmented by income, age, household composition, education, and marital status, with "Balance Income" and "High Income" groups driving most of the sales, especially on premium products like meat and gold.

Key findings include:

* **Juice** is the most popular product across segments, while **fruits** and **sweets** lag behind.
* Long-term, loyal customers are more responsive to campaigns and engage more with the company (e.g., website visits).
* **Families with children** tend to spend more, while married customers and those with higher education also show higher purchasing power.
* **Upselling opportunities** exist for customers with high purchasing power but low current spending on premium items.
* As number of kids & teens increases, customers tend to spend less.
* Campaign Participation increases with higher recency score (Customers with low recency)

### **Recommendations:**

* Focus marketing efforts on **targeted campaigns** for high-value and long-term customers.
* Promote **low-selling products** to lower-income and newer customers.
* Use **upsell strategies** for premium products to high-income segments.

This approach will enhance customer retention, optimize marketing, and increase sales across all customer segments.