Table of Contents

Executive Summary2

1. Introduction3
2. Methodology3

*2.1 Data Preparation and Cleaning3*

*2.2 Clustering Analysis4*

* 1. *Segmentation Analysis6*
  2. *Predictive Modelling8*

3. Findings and Analysis10

*3.1 Customer Engagement clusters10*

* 1. *Demographic Insights11*

*3.3 Engagement Segments14*

*3.4 Promotional Spend Analysis14*

3.5 Insights from Feature Importance14

4. Recommendations15

5. Conclusion15

Strategic Impact and Recommendations16

# Executive Summary

This report provides a comprehensive analysis of customer engagement and marketing effectiveness using a simulated dataset modelled on retail consumer behaviour. The primary objectives are to identify patterns in customer behaviour, evaluate the impact of promotional spending, and suggest actionable recommendations to enhance customer acquisition and retention. This analysis leverages clustering, segmentation, and predictive modelling techniques to derive insights from the dataset.

Key findings from the analysis include:

* **Customer Engagement Clusters**: Three distinct clusters were identified, representing varying levels of engagement. High-value customers contributed significantly to subsequent order profits and were most responsive to promotions.
* **Demographic Insights**: Male customers aged 30-50 formed the most engaged demographic, predominantly sourced through direct and organic search channels.
* **Promotional Effectiveness**: Promotions were a critical driver for engagement, with moderate-spend promotions achieving the best balance between cost and returns.

Recommendations for improving customer engagement and marketing effectiveness include:

1. Prioritizing high-engagement customers for targeted marketing campaigns.
2. Optimizing paid marketing channels (e.g., paid search) to attract more engaged customers.
3. Refining promotional strategies to focus on moderate-spend campaigns for maximum ROI.

This report concludes with insights and recommendations designed to align with the retail companies’ strategic goals, offering data-driven pathways to enhance marketing outcomes and customer experiences.

# **Analysis Report**

## 1. Introduction

This report focuses on analysing customer engagement data to derive actionable insights. This analysis aims to identify patterns in customer behaviour, evaluate the impact of promotional spending, and recommend strategies to enhance customer acquisition and retention.

The dataset used in this analysis includes information on marketing costs, customer demographics, and engagement metrics. By employing techniques such as clustering, segmentation, and predictive modelling, this report uncovers key drivers of customer behaviour and offers data-driven recommendations to optimize marketing strategies.

This analysis is designed to showcase the application of data analytics in addressing business challenges and informing strategic decision-making for customer engagement.

## 2. Methodology

#### 2.1 Data Preparation and cleaning

The dataset provided for this analysis included:

1. **Cost Data**: Monthly marketing expenditures across various channels.
2. **Value Info Data**: Information on initial and subsequent orders, profits, and total promotional values.
3. **Demographic Info Data**: Customer demographics and their acquisition sources.

**Key Steps in Data Preparation:**

* **Data Cleaning**:
  + Missing values were imputed using the mean for numerical data and the mode for categorical data. This ensured that the analysis proceeded without data gaps affecting the results.
  + Created new data frames by merging relevant sheets and isolating customers from Dublin
* **Feature Engineering**:
  + Derived new metrics such as customer Recency, and Engagement Segments based on subsequent order counts.
  + Aggregated promotional spend values to analyse their effect on engagement.
* **Encoding Categorical Data**:
  + Categorical variables like Gender and Customer Source were label-encoded to ensure compatibility with machine learning algorithms.
* **Scaling Numeric Features**:
  + Applied standard scaling to numeric columns, including Subsequent Orders Count and Total Promotions, to normalize data ranges for clustering and predictive modelling.

This rigorous data preparation step ensured that the datasets were clean, relevant, and ready for advanced analysis, forming the foundation for clustering and predictive modelling.

#### 2.2 Clustering Analysis

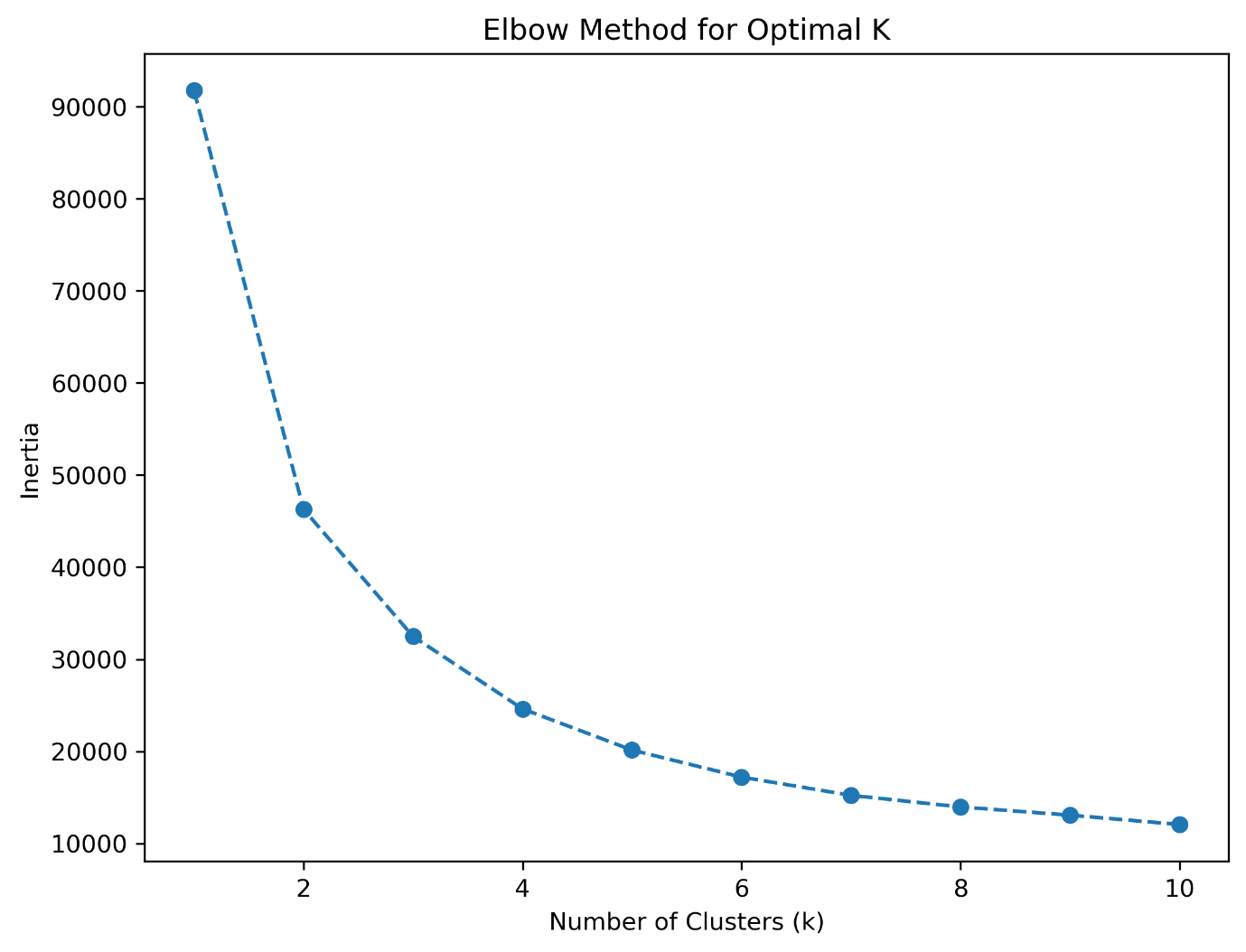
The clustering analysis aimed to segment customers based on their engagement patterns and key behavioural metrics. The following steps and findings highlight the approach and outcomes:

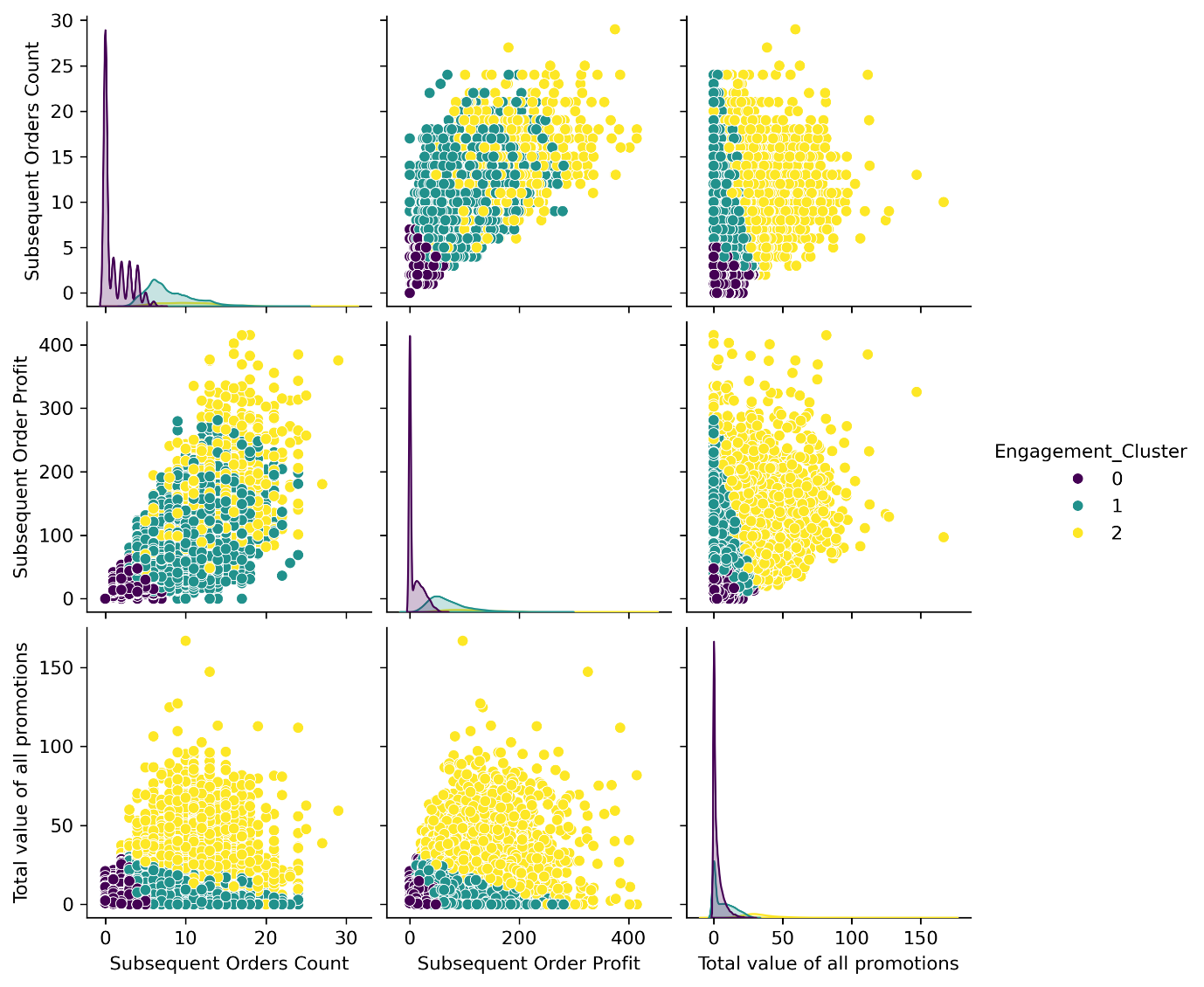
**Methodology**

1. **Feature Selection**:
   * Key features such as Subsequent Orders Count, Subsequent Order Profit, and Total Promotions were used to define customer engagement.
   * Features were standardized using a scaler to ensure equal weight during clustering.
2. **K-Means Clustering**:
   * The K-Means algorithm was chosen for its simplicity and effectiveness in creating distinct groups.
   * The optimal number of clusters (k=3) was determined using the elbow method, where inertia was plotted against the number of clusters to find the point of diminishing returns.
3. **Cluster Labelling**:
   * Each customer was assigned to one of three clusters based on their engagement patterns:
     + **Cluster 0**: Low engagement with minimal order counts and profits.
     + **Cluster 1**: Moderate engagement, characterized by consistent order activity and average profitability.
     + **Cluster 2**: High engagement, defined by frequent orders, high profits, and significant responsiveness to promotions.

**Findings:**

1. **Cluster Characteristics**:
   * **Cluster 0** (Low Engagement):
     + Smallest group with limited interaction.
     + Contributed minimally to overall profits.
   * **Cluster 1** (Moderate Engagement):
     + Largest group with steady but average profitability.
   * **Cluster 2** (High Engagement):
     + Highest contribution to profits and strong responsiveness to promotional campaigns.
2. **Insights from Clustering**:
   * High-engagement customers (Cluster 2) are the most valuable segment and should be prioritized in targeted marketing efforts.
   * Low-engagement customers (Cluster 0) present an opportunity for reactivation strategies.

* **Elbow Method Plot**: Showed a clear inflection point at k=3, supporting the decision to use three clusters. 
* **Cluster Scatterplot**: Highlighted the separation of clusters, with Cluster 2 standing out for its high values in subsequent orders and profits.



The clustering analysis provided a foundation for understanding customer segments, enabling targeted recommendations for improving engagement and profitability.

#### 2.3 Segmentation Analysis

The segmentation analysis builds on the clustering results to group customers into four distinct engagement segments based on their subsequent order counts. This process allows for more targeted recommendations by identifying customers with varying levels of interaction and profitability.

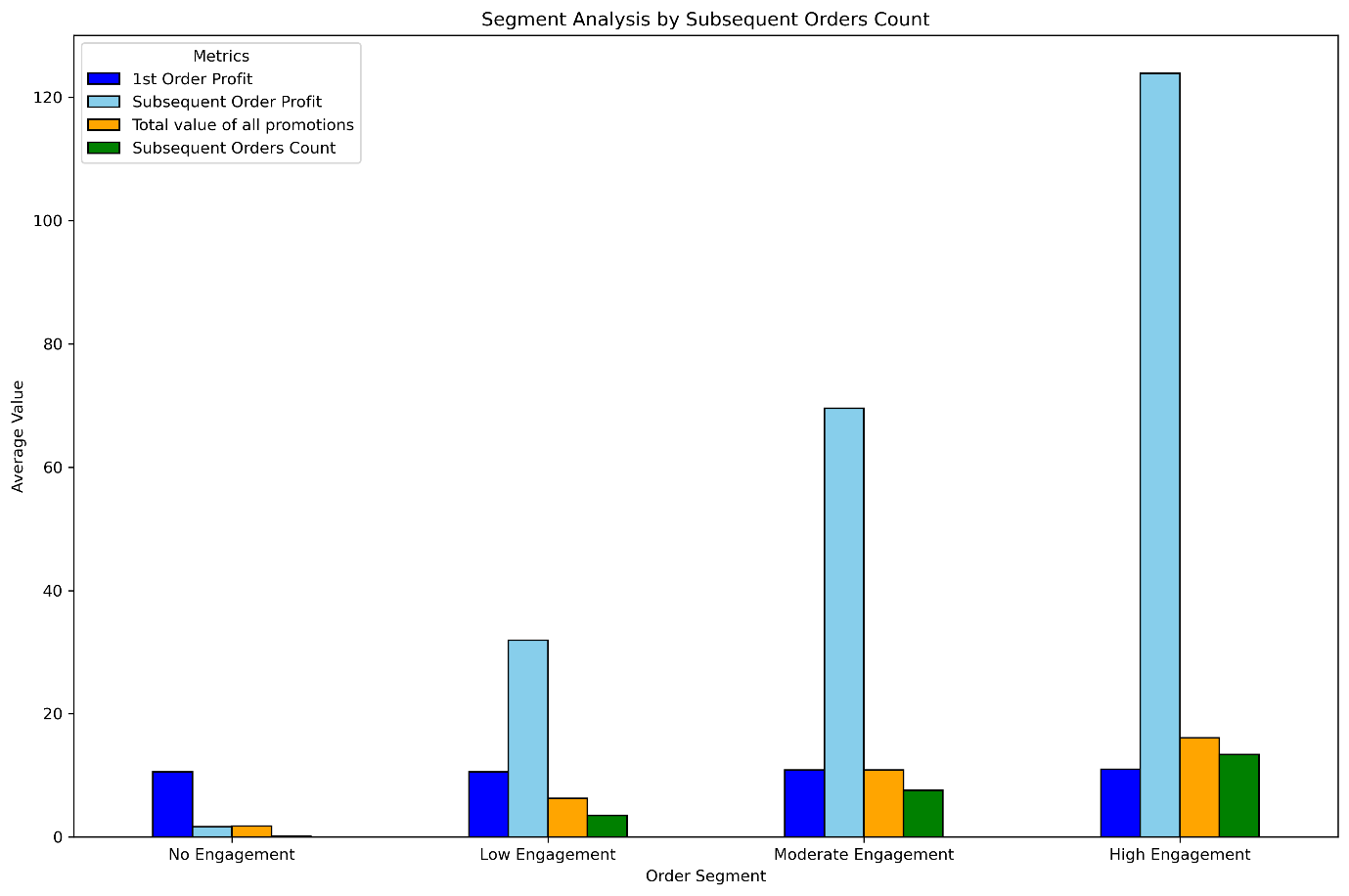
**Methodology**

1. **Engagement Segments**:
   * Customers were divided into the following segments based on their subsequent orders:
     + **No Engagement**: No subsequent orders after their initial interaction.
     + **Low Engagement**: Between 1–3 subsequent orders.
     + **Moderate Engagement**: Between 4–10 subsequent orders.
     + **High Engagement**: More than 10 subsequent orders.
2. **Features for Segmentation**:
   * **Subsequent Orders Count**: Used as the primary metric to categorize engagement.
   * **Subsequent Order Profit**: Analysed to evaluate the profitability of each segment.
   * **Promotional Spend**: Included to understand how promotions drive engagement.

**Findings**

1. **Segment Distribution**:
   * Most customers fell into the **Low Engagement** segment, reflecting a common behaviour of limited repeat activity.
   * A smaller proportion of customers were classified as **High Engagement**, indicating a highly valuable group that contributes disproportionately to profits.
2. **Profitability**:
   * **High Engagement** customers generated the highest average profits per order and were most responsive to promotional campaigns.
   * **No Engagement** customers had negligible profitability, highlighting a potential opportunity for re-engagement strategies.
3. **Promotional Effectiveness**:
   * Promotions had a significant impact on **Moderate** and **High Engagement** segments, with average promotional spending correlating positively with subsequent profits.
   * For **Low Engagement** customers, promotions were less effective, suggesting diminishing returns on investment for this group.

* **Segment Analysis chart:** Shows higher subsequent order profit with only a moderate increase in promotion spending



#### 2.4 Predictive Modelling

Predictive modelling was used to forecast customer engagement levels and profitability. While the models did not achieve high predictive accuracy, the feature importance derived from the models provides valuable insights into the factors most strongly affecting subsequent order count and profit. By employing machine learning techniques, I aimed to identify key drivers of customer behaviour and provide actionable insights for improving marketing strategies.

**Methodology**

1. **Feature Selection**:
   * Key predictors included Subsequent Orders Count, Promotional Spend, and Recency.
   * Features were standardized to ensure uniform scale and reduce bias in model performance.
2. **Models Used**:
   * **Linear Regression**: For predicting subsequent order profits.
   * **Random Forest Classifier**: For classifying customers into engagement levels (low, medium, high).
   * **XGBoost**: For improving classification accuracy and addressing imbalanced data.
3. **Evaluation Metrics**:
   * **Regression**: Evaluated using R² score, mean squared error (MSE), and mean absolute error (MAE).
   * **Classification**: Assessed using accuracy, precision, recall, and F1-score.

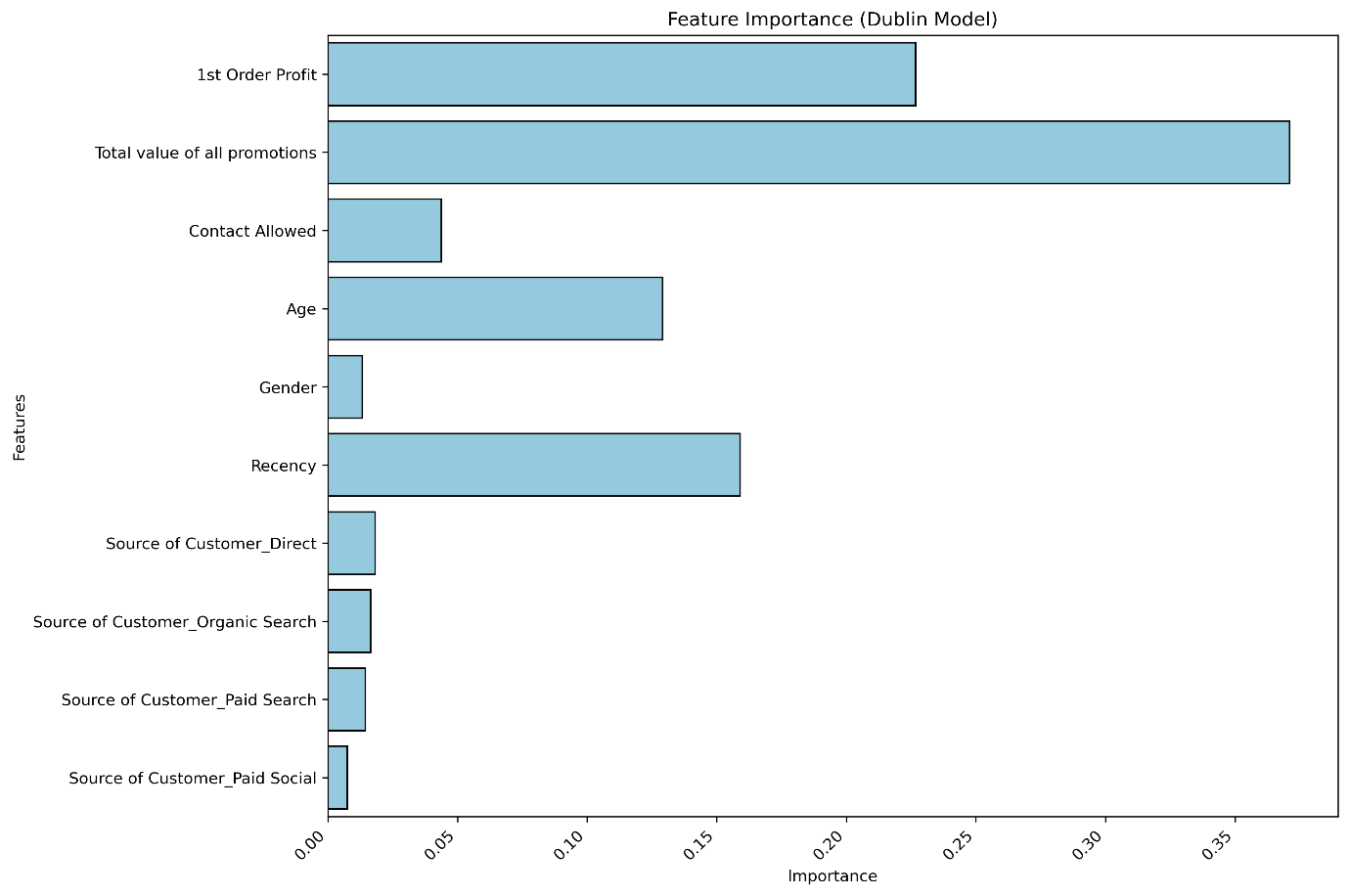
**Findings**

* **Key Drivers of Engagement**:
  + Promotional spending, recency and first order profit were the most significant predictors of engagement.
  + High-spend promotions had a clear impact on transitioning customers to higher engagement levels.
* **Model Performance**:
  + Several models were tested, however most indicated insufficient data to build an accurate predictive model for subsequent order count or profit
  + However, feature importance analysis showed that features like total value of promotions, initial purchase profit, and subscription recency consistently had more weight in predicting subsequent order count and profit
* **Challenges**:
  + Imbalanced data across engagement levels reduced classification performance for low-frequency segments.
  + Several manufactured features warped the results of models, which required features to be dropped to prevent cyclical reasoning

**Insights**

Predictive modelling highlights the potential of using targeted promotions and recency-based strategies to increase customer engagement and profitability. While the overall model performance was poor, the feature importance analysis identified key drivers such as promotional spending and recency, which are critical for influencing subsequent orders and profits. These insights can inform strategic decisions even if the predictive accuracy was limited. Improving data balance and incorporating additional features, such as customer feedback, could enhance model accuracy.

* **Feature importance chart**: model testing showed promotion value, 1st order profit, recency and age were key indicators in predicting subsequent order count and profit, however poor model performance may suggest this should be considered with caution.



## 3. Findings and Analysis

### 3.1 Customer Engagement Clusters

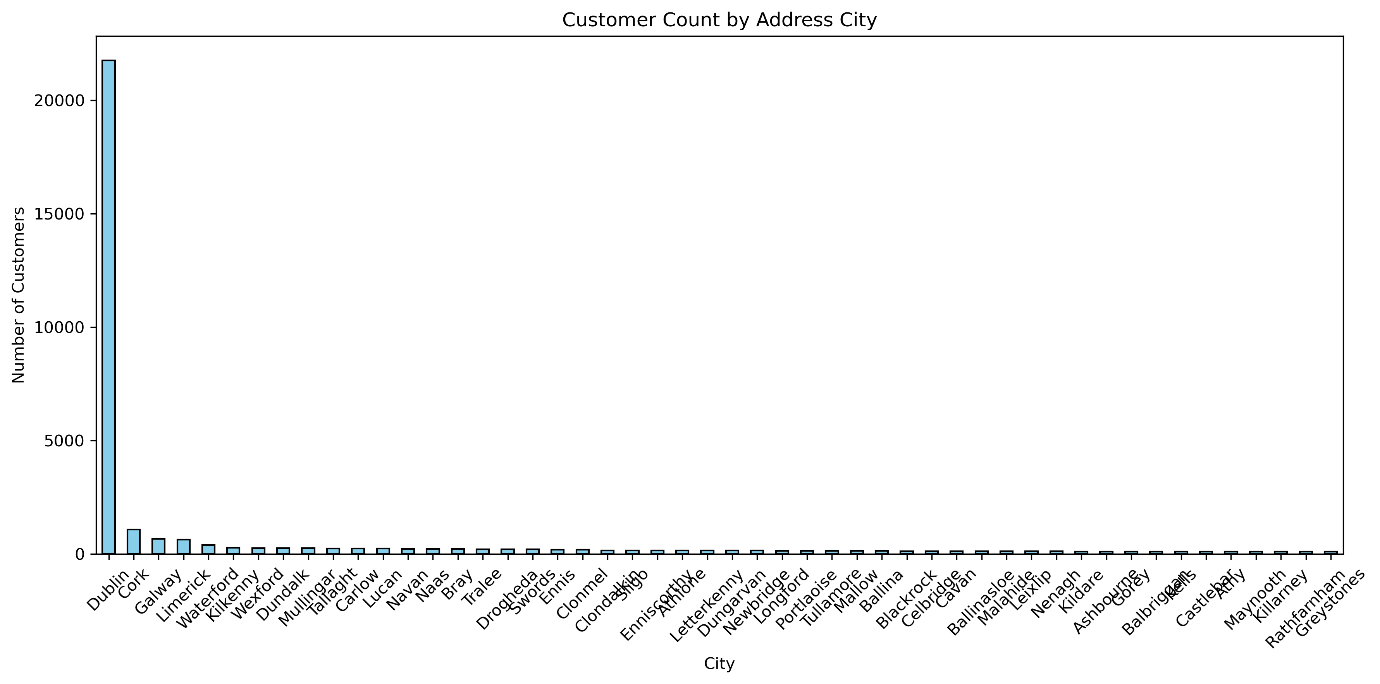
The clustering analysis provided a clear segmentation of customers based on their engagement levels:

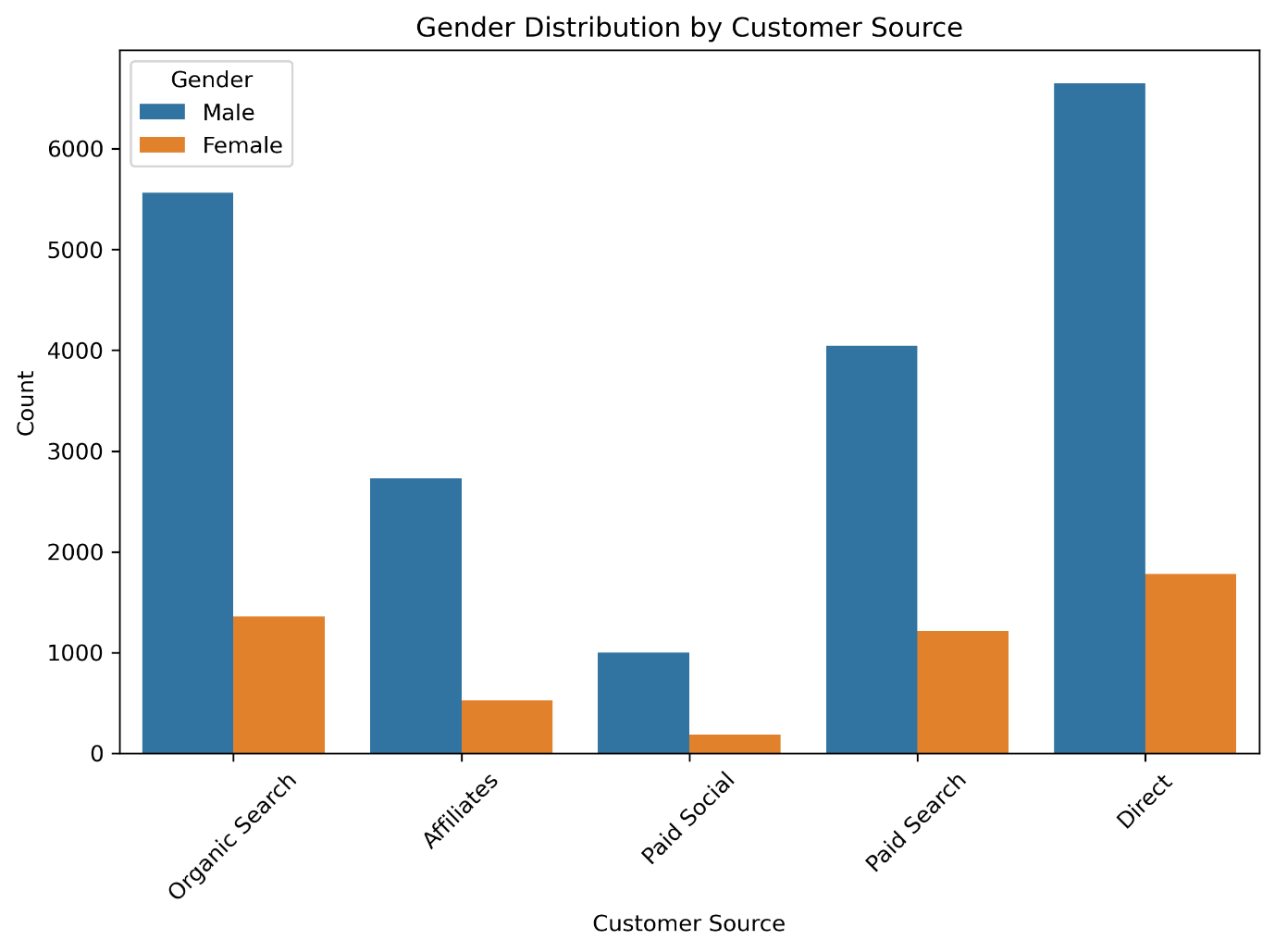
* Cluster 0 (Low engagement):
  + Represented customers with minimal subsequent orders and negligible profits.
  + This group contributed the least to overall revenue, highlighting a cluster with low ROI for promotional efforts.
  + Strategies targeting reactivation may help improve their engagement.
* Cluster 1 (Moderate engagement):
  + Comprising the largest cluster, these customers exhibited steady engagement and average profitability.
  + Their consistent order activity suggests potential for upselling and cross-selling strategies to move them into the high engagement cluster.
* Cluster 2 (High engagement):
  + The smallest but most profitable cluster, defined by frequent orders, high profits, and significant responsiveness to promotions.
  + Prioritizing retention for this group is critical to sustaining long-term revenue growth.

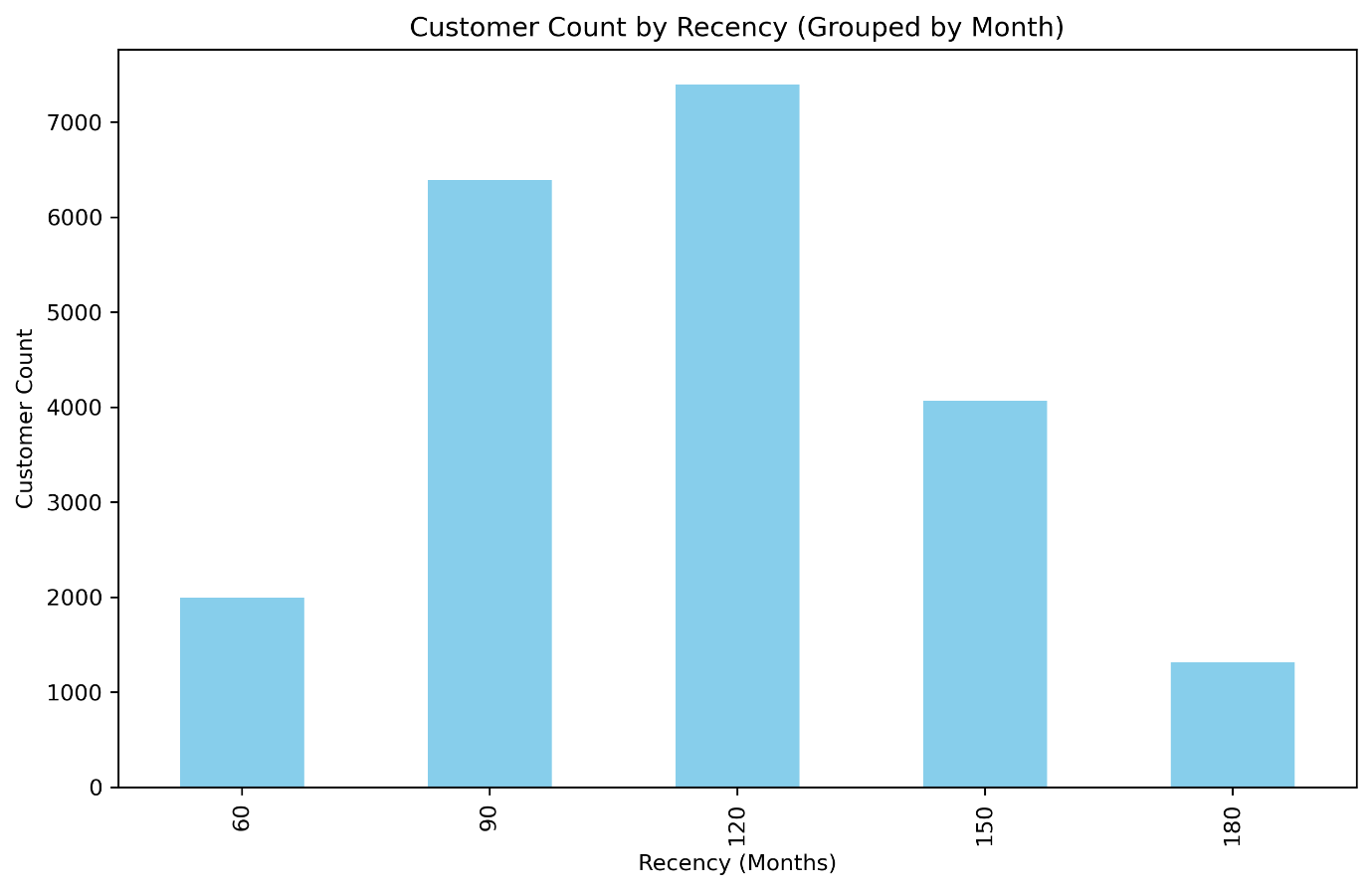
### 3.2 Demographic Insights

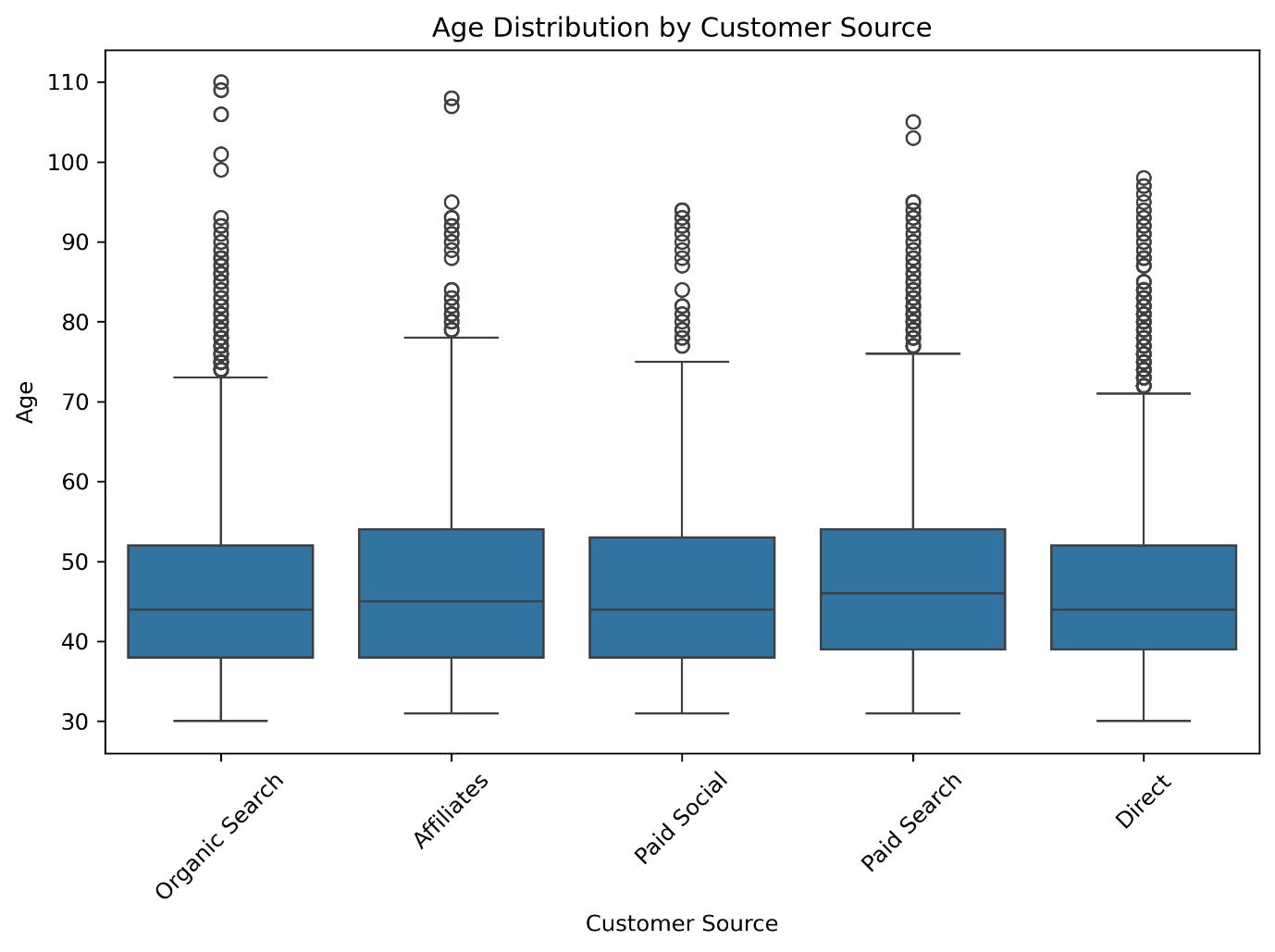
Demographic analysis revealed key trends:

* **Gender Distribution**:
  + Male customers aged 30-50 formed most engaged users, particularly in high engagement clusters.
  + Marketing campaigns focusing on this demographic may yield better results.
* **Customer Acquisition Channels**:
  + Direct and organic search channels performed the best, contributing to the highest engagement and profitability levels.
  + Paid channels, such as paid search and affiliates, showed lower engagement rates, suggesting a need for optimization in these areas.
* **Customer Location**
  + Data shows that most customers are located in Dublin, which suggests that the business should prioritize its marketing efforts in this location.
* **Further Visualizations:**









### 3.3 Engagement Segments

The segmentation analysis categorized customers into four distinct engagement levels:

* **No Engagement**:
  + Customers with no subsequent orders contributed little to revenue.
  + Reactivation campaigns, such as first-order discounts or re-engagement emails, could be effective.
* **Low Engagement**:
  + Representing a significant proportion, these customers had 1-3 subsequent orders and limited profitability.
  + Cost-effective promotional strategies may help improve their activity.
* **Moderate Engagement**:
  + Customers with 4-10 subsequent orders contributed average profits.
  + These users are strong candidates for upselling and loyalty programs to transition them into the high engagement segment.
* **High Engagement**:
  + The most profitable group, with over 10 subsequent orders, demonstrating the highest responsiveness to promotional efforts.
  + Retention campaigns and personalized offers can help maintain their activity and profitability.

### 3.4 Promotional Spend Analysis

Promotions had varying levels of effectiveness across segments:

* High engagement customers showed a strong correlation between promotional spend and subsequent profits, emphasizing the value of targeted campaigns.
* Moderate engagement customers benefited from moderate promotional spend, achieving a good balance between investment and returns.
* Low engagement and no engagement customers demonstrated diminishing returns, suggesting a need for more tailored and cost-effective strategies.

### 3.5 Insights from Feature Importance

Feature importance analysis from predictive models highlighted key factors influencing customer behaviour:

* **Promotional Spend**:
  + A major driver of engagement and profitability, underscoring the importance of optimizing promotional budgets.
* **Recency**:
  + Recent customer interactions significantly impacted subsequent orders, indicating the need for timely follow-ups.
* **Customer Source**:
  + Acquisition channels played a critical role in determining engagement, with direct and organic channels showing the highest potential for profitability.
* **First Order Profit:**
  + Customers who make larger initial purchases are more likely to make higher value subsequent orders, which suggests that targeting these customers may lead to increased ROI.

These findings collectively provide a robust foundation for actionable strategies to enhance customer engagement, optimize marketing efforts, and drive long-term profitability.

## 4. Recommendations

1. **Prioritize High Engagement Customers**:
   * Retention strategies, personalized offers, and loyalty programs.
2. **Optimize Paid Channels**:
   * Focus on channels with high ROI and reassess underperforming campaigns.
3. **Promotional Spend Optimization**:
   * Allocate resources to moderate-spend campaigns for maximum profitability.
4. **Improve Predictive Models**:
   * Address data imbalance and include additional features for enhanced accuracy.

## 5. Conclusion

The analysis provided actionable insights into customer engagement, identifying high-value segments and strategies to improve marketing effectiveness. By targeting high-engagement customers and refining promotional strategies, the business can enhance customer retention and profitability.

# Strategic Impact and Recommendations

This analysis highlights the value of data-driven strategies in optimizing marketing efforts and enhancing customer engagement for the retail company. By leveraging insights into customer behaviour, the outlined recommendations can improve marketing efficiency and ROI.

To build on these findings, the following steps are recommended:

1. **Target High-Value Clusters**:
   * Focus on high-engagement customers through personalized campaigns.
2. **Continuous Optimization**:
   * Regularly assess promotional effectiveness and adjust based on engagement metrics.
   * Enhance clustering models as new data becomes available.
3. **Future-Proofing**:
   * Expand analytics capabilities to identify further growth opportunities.

These actions align with the companies’ goals of leveraging analytics for strategic growth, ensuring long-term success and innovation.