



Optimizing Marketing Strategies with Customer Segmentation in Thames E-Commerce

By Riza Arfiqih



Project Overview



01

**Business & Data
Understanding**

02

Data Wrangling

03

**Exploratory Data
Analysis**

04

Model Building

05

Model Interpretation

03

Conclusion

Strategic planning project plan



Business Context

Thames E-Commerce faces the critical challenge of understanding customer behavior and optimizing marketing strategies

Vision

This project can empower the company to make more informed decisions and elevate overall business performance.

Objectives

Conduct an in-depth analysis of the Thames E-Commerce dataset using RFM Analysis to segment customers based on their recency, frequency, and monetary value, enabling targeted marketing campaigns.

Outcomes

Provide actionable recommendations for Thames E-Commerce to optimize its marketing strategies, operational efficiency, and customer experience.

Data Understanding

This project leverages the dataset "*E-Commerce Data | Actual transactions from UK retailer*" available on [Kaggle](#). This dataset offers a comprehensive record of transactions from a UK-based online retailer, Thames E-Commerce, spanning from 01/12/2010 to 09/12/2011.



The dataset contains 541909 rows and 8 columns which is:

| # | Column | Non-Null | Count | Dtype |
|---|-------------|----------|----------|---------|
| 0 | InvoiceNo | 541909 | non-null | object |
| 1 | StockCode | 541909 | non-null | object |
| 2 | Description | 540455 | non-null | object |
| 3 | Quantity | 541909 | non-null | int64 |
| 4 | InvoiceDate | 541909 | non-null | object |
| 5 | UnitPrice | 541909 | non-null | float64 |
| 6 | CustomerID | 406829 | non-null | float64 |
| 7 | Country | 541909 | non-null | object |

- **InvoiceNo** : Categorical, Invoice of transactions in unique code.
- **StockCode** : Categorical, Specific product with unique code.
- **Description** : Categorical, Description of detailed product information.
- **Quantity** : Numerical, Amount of units had been sold.
- **InvoiceDate** : Categorical, transactions time.
- **UnitPrice** : Numerical, Price of each unit.
- **CustomerID** : Numerical, Specific customer with unique code.
- **Country** : Categorical, Customer location by country.



Data Transform

- Change data type into true format
- Create new feature '**Sales**' (Quantity*Unitprice).

Duplicate

Drop 5.192 duplicated rows.

Missing Values

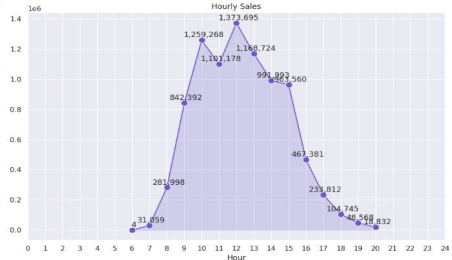
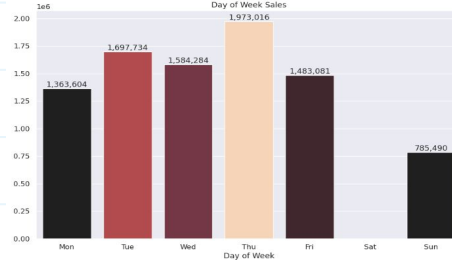
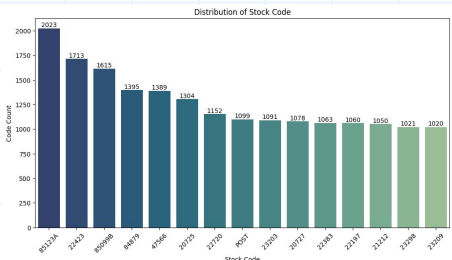
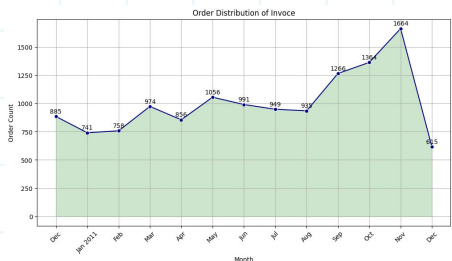
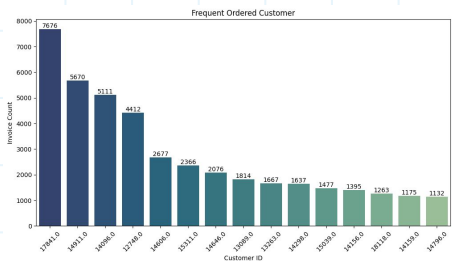
- 25% of the CustomerID were missing and 0.2% of the products didn't have any description.
- Drop feature CustomerID and Description, which both consist of 'NaN' when UnitPrice consists of 0.00. Also, feature Quantity seems abnormal, consisting of minus/negative quantity.

Zero/Negative Values

- Drop transactions with a big amount of quantity but the price is 0.
- Drop the 'Canceling' invoices which consisted of minus/negative quantity.

Total Customer : 4.338
Total Product : 3.665
Total Order : 392.688
Total Invoice : 18.532
Highest Monthly Order : Nov 2011, 1.664 order
Total Sales : 8.8 M
Highest Monthly Sales : Nov 2011, 1.1 M
Overall Sales trend : Slightly upward , increasing orders over the year.

Exploratory Data Analysis

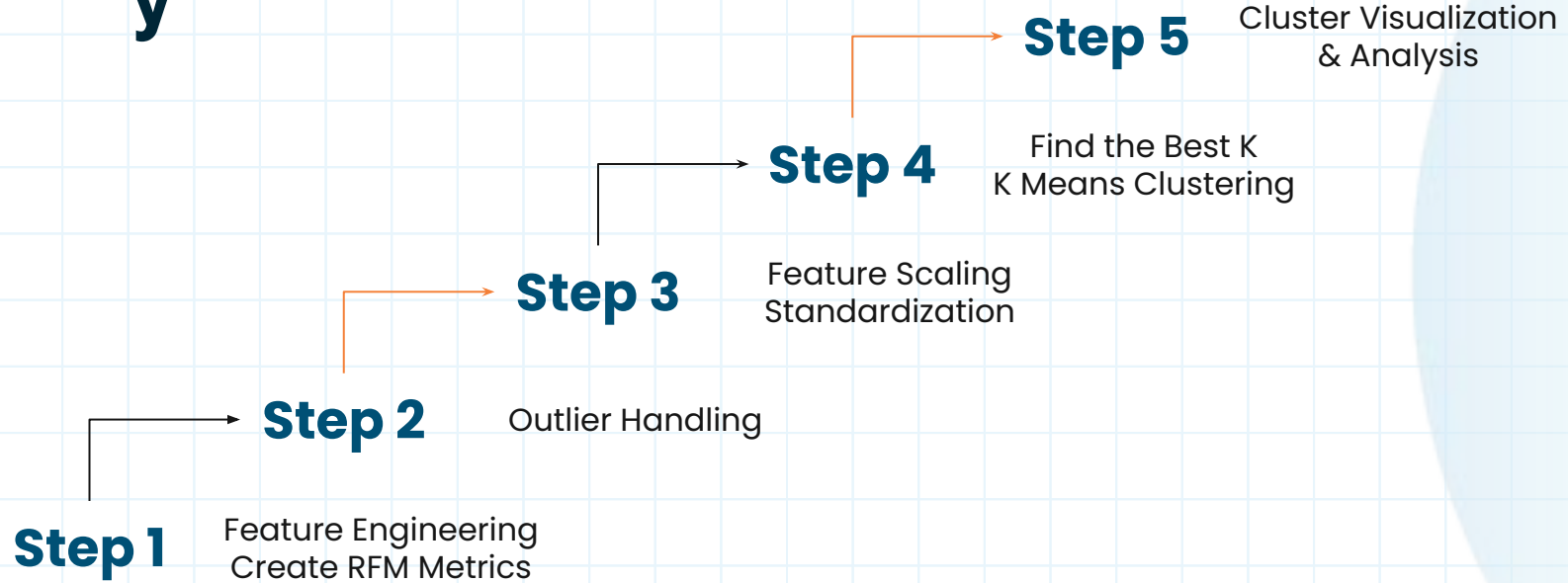


The sales trend is **upward at the beginning of the week** and **downward to the weekend** due to the day off.
The hourly sales **peak around 10 am to around 3 pm**. This aligns with the typical shopping hours of vendors and might be influenced by working hours because some of the customers are resellers.

* Please visit [my notebook](#) for more explorative analysis



Methodology



Build the K Means clustering model with 3 features according to RFM.

Why using RFM?

- RFM (Recency, Frequency, Monetary) Analysis is a segmentation technique that uses customer purchase behavior to group them into segments. And very effective as the purchase behavior can be summarized by using a small number of variables.
- It is also very useful for targeted marketing campaigns and other customer relationship management initiatives.

Why using K-means?

- K-means is a effective clustering algorithm for customer segmentation and RFM analysis due to its simplicity, efficiency, and interpretability.
- Computationally efficient algorithm, suitable for clustering large datasets.
- Can be relied on to determine appropriate value for 'k' (the number of clusters) using techniques like the elbow method or silhouette analysis.

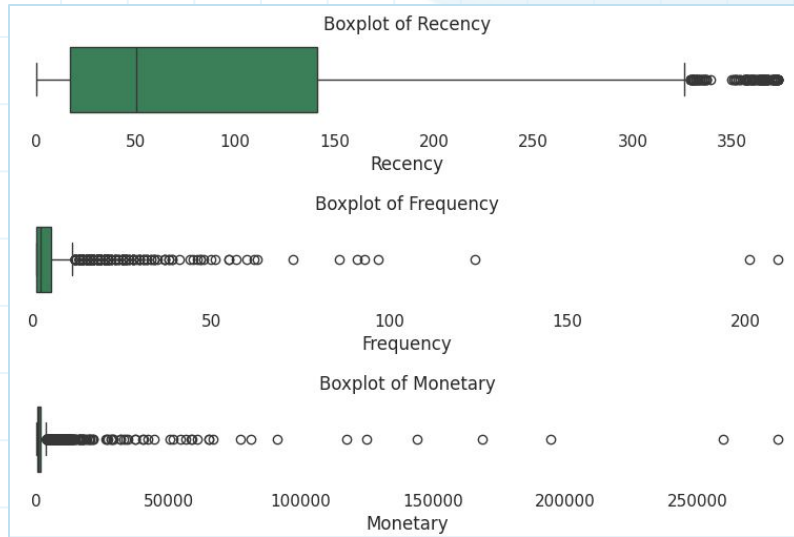
Define the RFM features as follows:

- **Recency (R):** InvoiceDate, calculate the number of days since the customer's last purchase.
- **Frequency (F):** InvoiceNo, calculate the number of unique invoices for each customer. This represents how often the customer has made purchases.
- **Monetary (M):** Sales, calculate the total sales for each customer. This represents the total value of the customer's purchases.

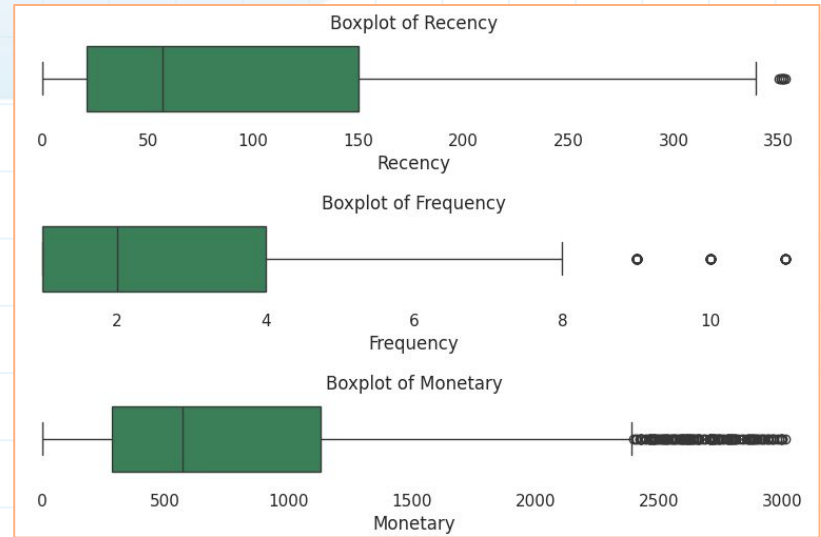
| | Recency | Frequency | Monetary |
|------------|---------|-----------|----------|
| CustomerID | | | |
| 12346.00 | 325 | 1 | 77183.60 |
| 12347.00 | 1 | 7 | 4310.00 |
| 12348.00 | 74 | 4 | 1797.24 |
| 12349.00 | 18 | 1 | 1757.55 |
| 12350.00 | 309 | 1 | 334.40 |

Outlier Handling

Handling the outliers of the **rfm_df** into more clean it's data for modeling!



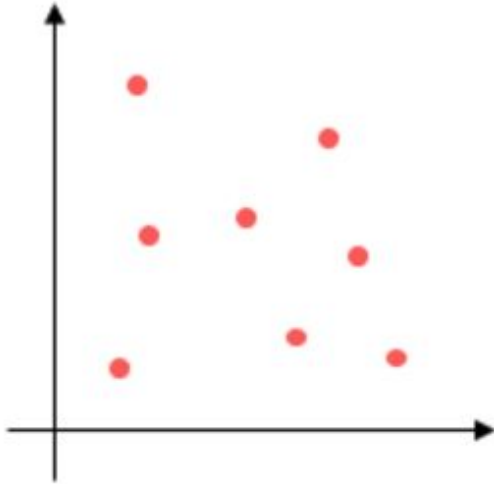
Before



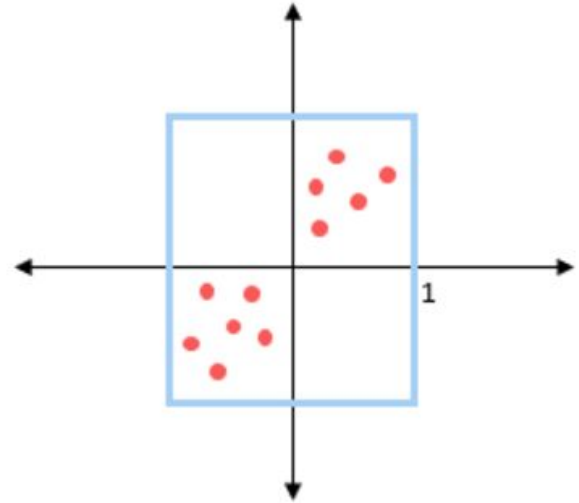
After

Feature Scaling

Ensuring that Recency, Frequency, and Monetary (RFM) features are on a similar scale before applying the K Means algorithm. This will help KMeans create more meaningful and accurate customer segments based on their RFM behavior.

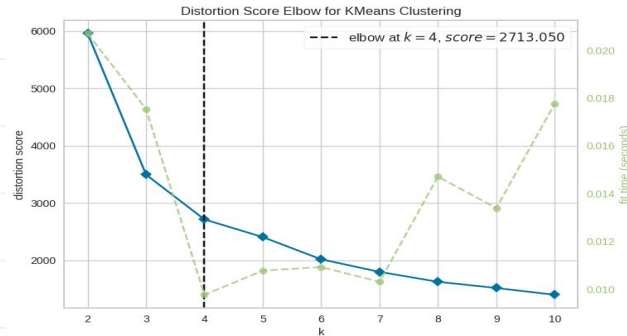
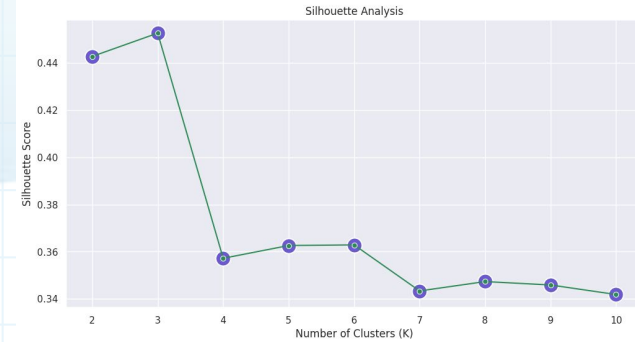
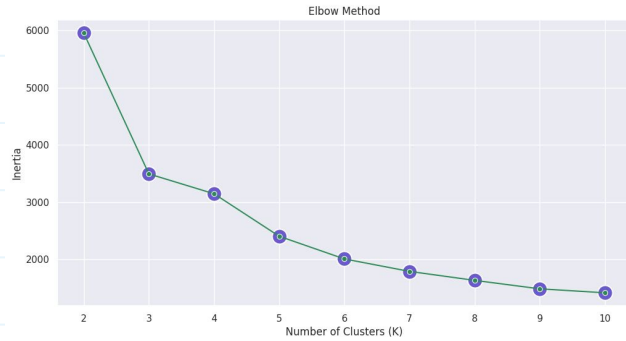


Actual Data

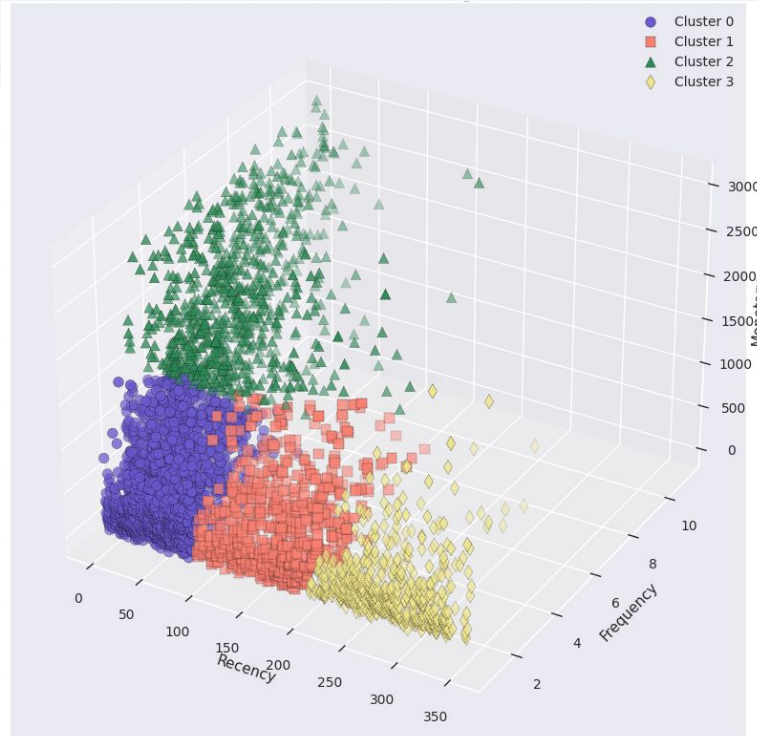


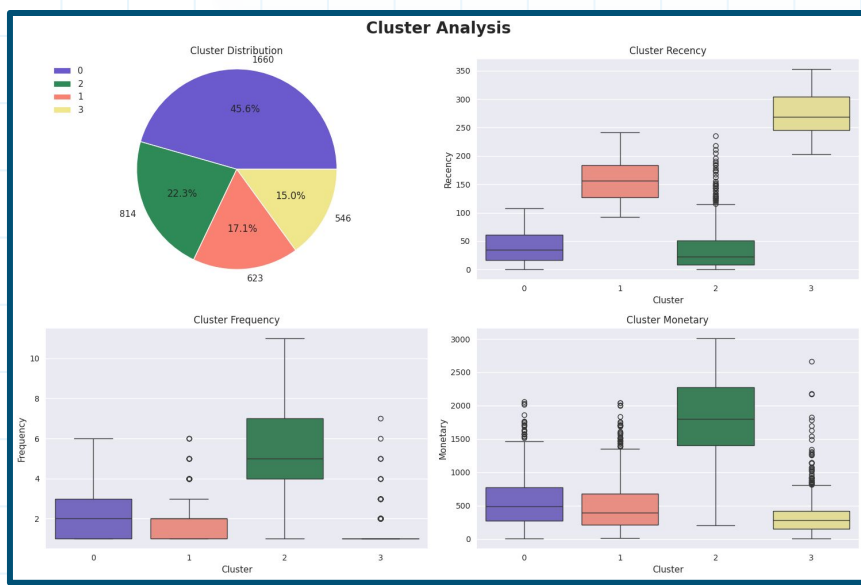
After Standardization

Based on the **Elbow Method** and **Silhouette Analysis** suggest 4 clusters for better interpretation and more effective customer segmentation.



This diagram shows the distribution of customers which is divided into clusters according to the K-Means Clustering algorithm.





| Percentile | Recency | Frequency | Monetary |
|------------|---------|-----------|----------|
| min | 0.00 | 1.00 | 33.00 |
| 10% | 23.00 | 1.00 | 70.00 |
| 20% | 48.00 | 1.00 | 109.00 |
| 30% | 72.00 | 1.00 | 122.00 |
| 40% | 99.00 | 1.00 | 159.00 |
| 50% | 133.00 | 1.00 | 207.00 |
| 60% | 163.00 | 1.00 | 218.00 |
| 70% | 198.00 | 1.00 | 228.00 |
| 80% | 236.00 | 2.00 | 250.00 |
| 90% | 287.00 | 2.00 | 354.10 |
| max | 363.00 | 6.00 | 994.00 |

High <= 40% → Low <= 40%

Moderate <= 50% → Moderate >= 50%

Low >= 80% → High >= 80%

Clusters breakdown and their potential customer segments based on percentiles to rank the customers from the RFM metrics to identify behavior patterns:

| Cluster | Total Customer | High | Moderate | Low |
|---------|----------------|-------|----------|-------|
| 0 | 1.660 | R | F,M | X |
| 1 | 623 | X | F,M | R |
| 2 | 814 | R,F,M | X | X |
| 3 | 546 | X | X | R,F,M |



Cluster 0

High recency, moderate frequency, moderate monetary

Loyal Customers

Distribution (45.57%): Dominant cluster with 1,660 customers.

Recency (around 30%): Customers who had made purchases very recently.

Frequency (above 50%): Their purchase frequency is moderate.

Monetary (around 50%): They spend a good amount of money.



Cluster 2

High recency, high frequency, high monetary

Champions Customers

Distribution (22.34%): 814 customers.

Recency (around 20%): These customers have purchased very recently.

Frequency (above 80%): They purchase very frequently.

Monetary (above 80%): They spend a significant high amount of money.



Cluster 1

Low recency, moderate frequency, moderate monetary

Promising Customers

Distribution (17.10%): 623 customers.

Recency (around 70%): They had not made purchased slightly in a long time.

Frequency (above 50%): Their purchase frequency is moderate too.

Monetary (around 50%): They spend a moderate amount of money.



Cluster 3

Low recency, low frequency, low monetary

At-Risk/Lost Customers

Distribution (14.99%): The minority of cluster, with 546 customers.

Recency (above 90%): Customers who hadn't made purchased in a long time.

Frequency (below 30%): Their purchase frequency is low, only 1 order in average.

Monetary (around 30%): They spend a small amount of money.

Conclusion



- There are 4 clusters of customers based on RFM/purchased behavior:
 - **At-Risk/Lost Customers** 14.99%
 - **Promising Customers** 17.10%
 - **Loyal Customers** 45.57%
 - **Champions Customers** 22.34%
- It's a good sign that the company has only a minority problem with customer churn and dominates with loyal customers throughout the year.
- Almost a quarter of its customers are high-value customers who drive significant revenue.
- Treat the promising customers well then I promise they'll transform into loyal.

Recommendations for Optimizing Marketing Strategies by Customer Segments:



Loyal Customers

- **Personalized Marketing:** Utilize data to tailor promotions and offers.
- **Loyalty Programs:** Implement tiered loyalty programs with exclusive rewards.
- **Customer Advocacy:** Encourage customer reviews and referrals.
- **Early Access:** Provide early access to new products or sales.



Promising Customers

- **Targeted Marketing:** Use email marketing and WhatsApp to promote relevant products.
- **Cross-Selling and Upselling:** Suggest complementary products or higher-tier options.
- **Personalized Offers:** Offer discounts or special deals to incentivize repeat purchases.
- **Customer Surveys:** Gather feedback to understand their needs and preferences.



Recommendations for Optimizing Marketing Strategies by Customer Segments:



Champion Customers

- **VIP Treatment:** Provide exceptional customer service and exclusive benefits.
- **Co-Creation:** Involve them in product development or marketing campaigns.
- **Personalized Experiences:** Tailor experiences to their individual preferences.
- **Exclusive Events:** Invite them to special events or product launches.



At-Risk/Lost Customers

- **Win-Back Campaigns:** Offer limited-time discounts or special promotions.
- **Personalized Communication:** Reach out with a sincere apology and offer to resolve any issues.
- **Re-Engagement:** Send targeted emails or WhatsApp highlighting new products or promotions.
- **Social Media Engagement:** Interact with them on social media to rekindle interest.

Thanks!

Any questions?

Email

WhatsApp

LinkedIn

GitHub

Link to Code



CREDITS: This presentation template was created by Slidesgo, and includes icons by Flaticon, and infographics & images by Freepik