BIA | BOSTON INSTITUTE OF ANALYTICS ®

E-Commerce Customer Segmentation

Agenda

1. Introduction and Objectives
    - Problem Statement
    - Business Objectives
    - Expected Outcomes
2. Data Methodology
    - Data Collection and Sources
    - Key Data Insights
    - Segmentation Methodology
    - Churn Prediction Methodology
    - Feature Engineering

Agenda

3. Customer Segmentation Results
    - Customer Segmentation Overview
    - Segment Profiles
    - Actionable Segment Strategies
4. Churn Prediction Results
    - Churn Prediction Model Overview
    - Key Drivers of Churn
    - Customer Churn Risk Levels.
5. Visualization and Deployment
    - Dashboard Overview
    - Deployment of Churn Model

Agenda

6. Future Enhancement
7. Conclusion

PROBLEM STATEMENT

- In an increasingly competitive e-commerce landscape, understanding and predicting customer behavior is essential to drive targeted marketing, improve retention, and manage resources efficiently.
- Currently, the company lacks a precise mechanism to categorize customers based on purchasing patterns, leading to generalized marketing strategies that may not resonate with different customer segments.
- This gap results in missed opportunities for customer engagement, suboptimal resource allocation, and potential revenue loss from unsatisfied or inactive customers.

OBJECTIVE OF THIS PROJECT

- The objective of this project is to develop a machine learning-based solution that segments customers according to their purchasing behaviors and predicts future purchasing tendencies
- By accurately segmenting customers, predicting churn likelihood, and pinpointing those who may need incentives for further engagement, the company can implement targeted retention and personalized marketing strategies.
- This approach will enhance customer satisfaction, foster loyalty, and drive repeat purchases while ensuring efficient resource allocation and optimized marketing strategies.

EXPECTED OUTCOMES

- **Customer Segmentation Insights:**

  Clear identification and categorization of customer segments based on purchasing behavior, allowing the marketing team to understand distinct characteristics and needs of each group.

- **Churn Prediction Model:**

  A robust model capable of identifying customers at risk of churning, enabling the company to implement timely, targeted retention strategies to reduce churn rates and enhance customer lifetime value.

EXPECTED OUTCOMES (cont..)

- **Enhanced Marketing Efficiency:**
- Targeted marketing strategies for each customer segment, focusing on high-value and high-risk groups, optimizing budget allocation and maximizing return on marketing investment.
- **Increased Revenue Potential:**
- Boosted revenue through reduced churn rates and increased repeat purchases, driving overall profitability by nurturing long-term customer relationships.


DATA AND METHOLOGY

DATA SUMMARY

The dataset consists of 8 attributes that provide crucial insights into customer transactions in a retail setting.
•Each attribute plays a vital role in analyzing purchasing patterns, customer behavior, and inventory management.

Sno	Attribute	Description
1	**InvoiceNo**	Unique identifier for each invoice, representing a specific transaction.
2	**StockCode**	Unique code for each product in stock, useful for tracking individual items in inventory.
3	**Description**	Text description of the product, providing the product name or additional details about the item.
4	**Quantity**	Number of units purchased per item, helpful for understanding sales volume and stock depletion.

DATA SUMMARY (cont…)

5	**InvoiceDate**	Date and time when the invoice was generated, crucial for tracking sales trends and seasonality.
6	**UnitPrice**	Price per unit of the product, essential for calculating total transaction amounts and revenue.
7	**CustomerID**	Unique identifier for each customer, allowing for individual customer tracking and behavior analysis.
8	**Country**	Country of the customer or purchase location, aiding in geographic sales analysis and segmentation.

DATA CLEANING AND PREPROCESSING TECHNIQUES

- **Missing Values:**
  - Identified and imputed or removed missing data in fields like CustomerID and Description to ensure data
- **Removing Duplicates:**
  - Eliminated duplicate records, especially in InvoiceNo and CustomerID, to maintain data accuracy.
- **Data Type Conversion:**
  - Converted fields such as InvoiceDate to datetime and Quantity to numerical types to enable proper computations.

DATA CLEANING AND PREPROCESSING TECHNIQUES

- **Feature Scaling:**
  - Applied scaling to numerical attributes like Quantity and UnitPrice for consistency in analysis and modeling.
- **Creating New Features:**
  - Derived features such as TotalPrice (Quantity * UnitPrice) to enhance transaction value insights.
- **Feature Engineering:**
  - Extracted components like month and day of the week from InvoiceDate to analyze seasonal trends and patterns.

Key Data Insights:

High Revenue Generating Products:
Identified the top products Contributing to overall revenue through analysis of Total Price and Quantity sold, This Indicate the key items to focus on for Further Sales

Sales Seasonality:
Observed that certain months and days show increased sales, highlighting seasonal demand patterns that can guide inventory and promotional strategies.

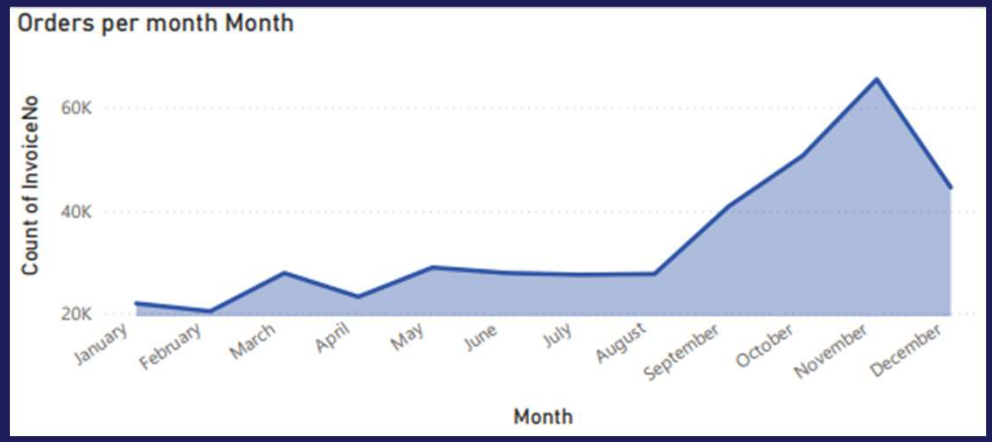
Description	TotalRevenue
REGENCY CAKESTAND 3 TIER	132870.40
WHITE HANGING HEART T-LIGHT HOLDER	93823.85
JUMBO BAG RED RETROSPOT	83236.76
PARTY BUNTING	67687.53
POSTAGE	66710.24
ASSORTED COLOUR BIRD ORNAMENT	56499.22
RABBIT NIGHT LIGHT	51137.80
CHILLI LIGHTS	45936.81
PAPER CHAIN KIT 50'S CHRISTMAS	41500.48
PICNIC BASKET WICKER 60 PIECES	39619.50

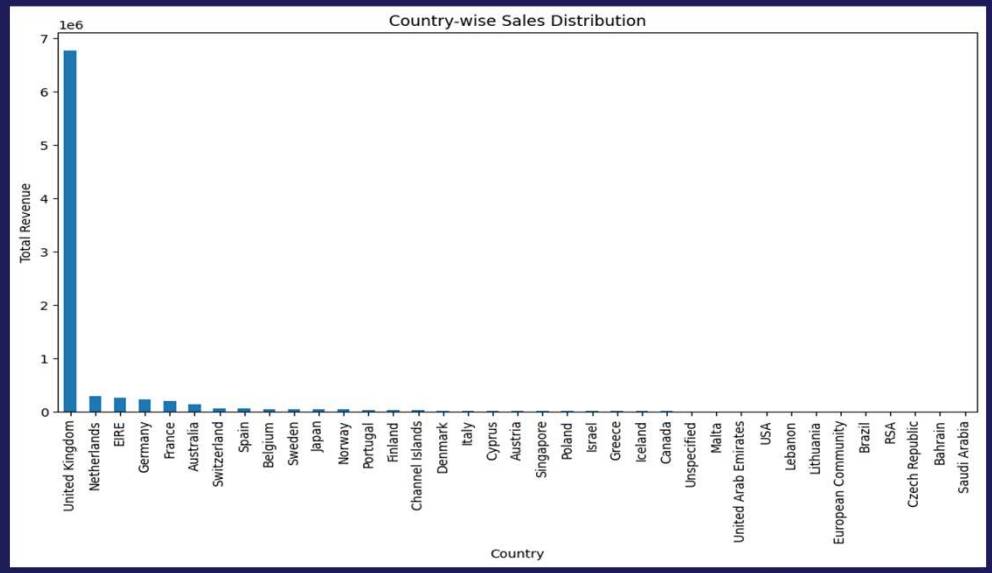
BIA BOSTON INSTITUTE OF ANALYTICS

Key Insights (cont…)

Customer Purchase:
Patterns Noticed distinct purchasing trends, such as peak shopping periods and frequency of repeat purchases by specific Customer IDs, useful for targeted marketing.

Country-wise Sales Distribution:
Sales primarily concentrated in a few countries, with the highest volume observed in the UK, presenting opportunities for geographical expansion.





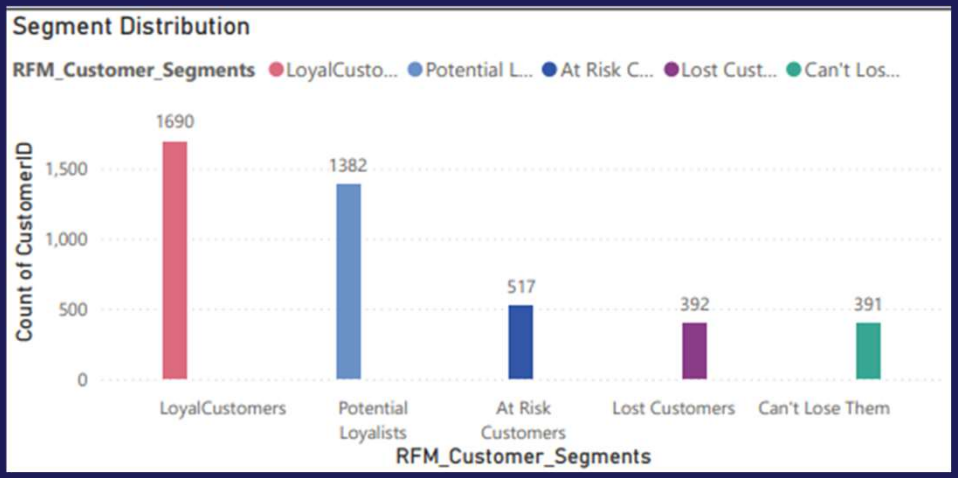
BIA BOSTON INSTITUTE OF ANALYTICS

Key Insights (cont…)

High-Value Customer Segments

•Identified customers with
•High values,
•Mid values,
•Low values base on Recency, Frequency, Monetary Values.

They are further segments as
•Loyal Customer
•Potential Customer
•At Risk Customer
•Lost Customer
•Can’t Loose Them
•useful for loyalty programs and customer retention strategies.



BIA BOSTON INSTITUTE OF ANALYTICS

METHODOLOGY OVERVIEW:

- **Clustering Techniques for Segmentation:**
  To identify distinct customer segments, clustering techniques were employed. Specifically:
  - **K-means Clustering** was used to partition customers into groups based on their purchasing patterns, allowing for tailored marketing strategies.
  - **Hierarchical Clustering** offered insights into customer hierarchy and relationships within segments, providing a more granular view of segment similarities.
  - **DBSCAN (Density-Based Spatial Clustering of Applications with Noise)**: This technique identified core and noise points in the dataset, making it effective for detecting natural clusters with irregular shapes, especially useful in separating dense areas from outliers.

METHODOLOGY OVERVIEW:

- **Machine Learning Models for Churn Prediction:**
  Several machine learning models were implemented to predict customer churn and enhance retention efforts:
  - **Random Forest** provided robust performance with feature importance analysis, improving model accuracy and aiding in identifying key drivers of churn.

FEATURE ENGINEERING:

- **Key Behavioral Features:**
  Important features were engineered to capture customer behavior patterns, improving model performance and interpretability:
  - **Recency, Frequency, Monetary (RFM) Analysis** features were derived to assess the likelihood of future purchases:
    - **Recency** measures time since the last purchase, indicating engagement levels.
    - **Frequency** counts the total purchases, capturing repeat customer value.
    - **Monetary** calculates total spending, which reflects customer value.

CUSTOMER SEGMENATION RESULTS

- **Overview of Segments**: The customer base was categorized into distinct segments, each with unique characteristics to tailor marketing strategies effectively:
  - At Risk Customers
  - Can't Lose Them
  - Lost Customers
  - Loyal Customers
  - Potential Loyalists

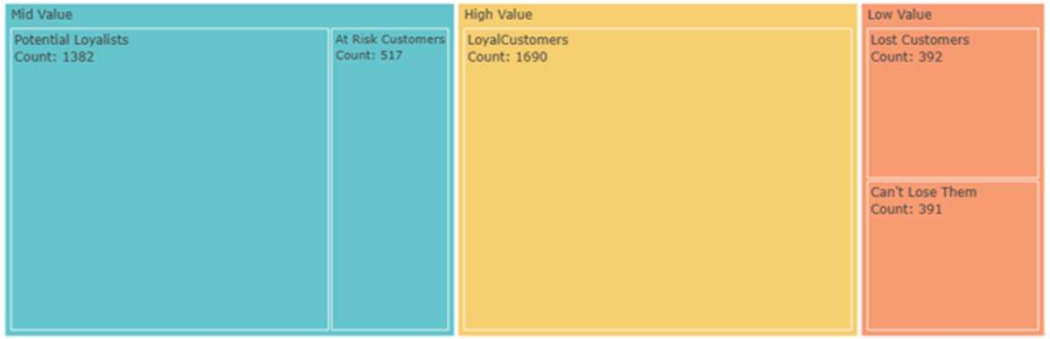
BIA BOSTON INSTITUTE OF ANALYTICS ®

OVERVIEW OF SEGMENTS

- **At Risk Customers:**
  - Customers with a declining purchase frequency who may potentially churn if not re-engaged. They show lower recent activity and could benefit from re-engagement campaigns or reminders.
- **Can’t Lose Them:**
  - High-value, highly engaged customers critical to revenue, but with signs of reduced engagement. Strategic retention efforts, such as personalized outreach or exclusive perks, are crucial for this group.
- **Lost Customers:**
  - Customers who haven’t made a purchase in a long time and show low engagement likelihood. Win-back campaigns or reactivation offers could help in rekindling their interest.

OVERVIEW OF SEGMENTS

- **Loyal Customers:**
  - Regular purchasers with high engagement and brand loyalty. They respond well to loyalty rewards, exclusive offers, and brand communication, making them key to retention efforts.
- **Potential Loyalists:**
  - New or occasional customers showing increasing engagement and potential for higher loyalty. Conversion efforts, such as onboarding programs or incentives, can help nurture these customers into becoming loyal patrons.



BIA | BOSTON INSTITUTE OF ANALYTICS ®

Actionable Insights

- **At Risk Customers**: Implement re-engagement campaigns like targeted discounts or personalized product recommendations.
- **Can’t Lose Them**: Offer special incentives or exclusive perks to maintain their interest and prevent churn.
- **Lost Customers**: Run reactivation campaigns with attractive offers to regain their interest and bring them back into active engagement.
- **Loyal Customers**: Maintain strong engagement through loyalty programs and premium customer service, ensuring continued retention.
- **Potential Loyalists**: Foster a relationship with incentives and personalized outreach to build loyalty and encourage repeat purchases.

Each segment benefits from tailored strategies that address their unique needs and potential to maximize engagement and lifetime value.

CHURN PREDICTION RESULTS ANALYSIS

- **Model Overview**:
  - **Model**: The Random Forest algorithm was implemented to predict customer churn based on Recency, Frequency, and Monetary (RFM) values, identifying customers at higher risk of leaving.
- **Performance Metrics**: **Accuracy**: The model achieved an accuracy of 99%, indicating reliable predictions based on RFM values.
- **ROC-AUC**: With a ROC-AUC score of 99%, the model effectively distinguishes churners from loyal customers.
- **Precision-Recall**: Optimized precision-recall balances were reached, minimizing false positives and ensuring high-risk customers were correctly identified.

BIA BOSTON INSTITUTE OF ANALYTICS ®

KEY DRIVERS OF CHURN

- **Recency**: Customers who hadn’t engaged recently had a higher risk of churning, emphasizing the importance of maintaining regular contact.
- **Frequency**: Low purchase frequency was a strong indicator of churn, with infrequent buyers showing less attachment.
- **Monetary Value**: Customers with lower overall spending levels were likelier to churn, potentially highlighting pricing sensitivity or lack of perceived value

CHURN RISK LEVELS

- **High Churn Risk:** Customers with low frequency, high recency, and low monetary value, indicating a need for retention strategies like special offers or re-engagement campaigns.
- **Medium Churn Risk:** Customers with moderate RFM scores, where targeted loyalty programs can strengthen engagement.
- **Low Churn Risk:** Customers with high purchase frequency, recent activity, and significant spending, suggesting low churn probability. Retention strategies should focus on satisfaction maintenance.

BIA BOSTON INSTITUTE OF ANALYTICS ®


BUSINESS IMPACT AND RECOMENDATIONS
Here is breakdown for each customer segment, focusing on strategies and impacts

Segmentation Recommendations:

1. **At-Risk Customers**:

- **Characteristics**: These customers show a decrease in engagement or purchasing frequency.
- **Strategy**: Offer personalized incentives, such as targeted discounts or reminders of loyalty benefits, to encourage renewed engagement.
- **Impact**: Re-engagement can prevent these customers from moving to the "Lost Customers" category, reducing churn rate.

Segmentation Recommendations:

2. Can't Lose Them:

- **Characteristics**: High-value customers essential to revenue, showing signs of wavering loyalty.
- **Strategy**: Implement VIP benefits like dedicated support, exclusive offers, or early product access to reinforce their importance.
- **Impact**: Retaining these customers preserves critical revenue streams and improves brand loyalty.

Segmentation Recommendations:

3. Lost Customers:

- **Characteristics**: These customers have disengaged and are no longer active.
- **Strategy**: Attempt reactivation with special offers or limited-time discounts; consider exit surveys to understand the reasons for disengagement.
- **Impact**: Reactivating even a small percentage can generate additional revenue and reveal insights into common churn causes.

Segmentation Recommendations:

4. Loyal Customers:

- **Characteristics**: Regular, repeat purchasers with high engagement and brand commitment.
- **Strategy**: Strengthen loyalty through rewards programs, referral incentives, and exclusive previews, fostering long-term relationships.
- **Impact**: Increasing loyalty among these customers maximizes lifetime value and boosts brand advocacy.

Segmentation Recommendations:

5. Potential Loyalists:

- **Characteristics**: Customers with occasional purchases who could become more frequent buyers.
- **Strategy**: Nurture this group by providing relevant product recommendations and tailored offers that encourage repeat purchases.
- **Impact**: Converting these customers into loyal customers increases engagement and future revenue potential.

CHURN INTERVENTION STRATEGIES

- **High-Risk Customers**: Deploy personalized offers to high-risk segments based on RFM profiles, enhancing the likelihood of re-engagement. Provide these customers with high-touch support, including priority customer service and exclusive deals.
- **Medium-Risk Customers**: Encourage engagement through re-targeting ads and email marketing that highlight relevant products and deals.
- **Low-Risk Customers**: Focus on maintaining satisfaction through loyalty programs, VIP memberships, or access to premium features to keep retention high.

POTENTIAL IMPACT

Increased Revenue

By improving retention of high-value customers, we could achieve an 25% increase in revenue.

Example: For “At-Risk Customers,” estimate the revenue saved by re-engaging a percentage of them. If re-engagement drives $100,000 in retained revenue and the cost of retention strategies (discounts, marketing) is $20,000, the ROI would be 400%.

Cost Efficiency

Targeted strategies reduce marketing costs by focusing on customer segments that offer the highest return

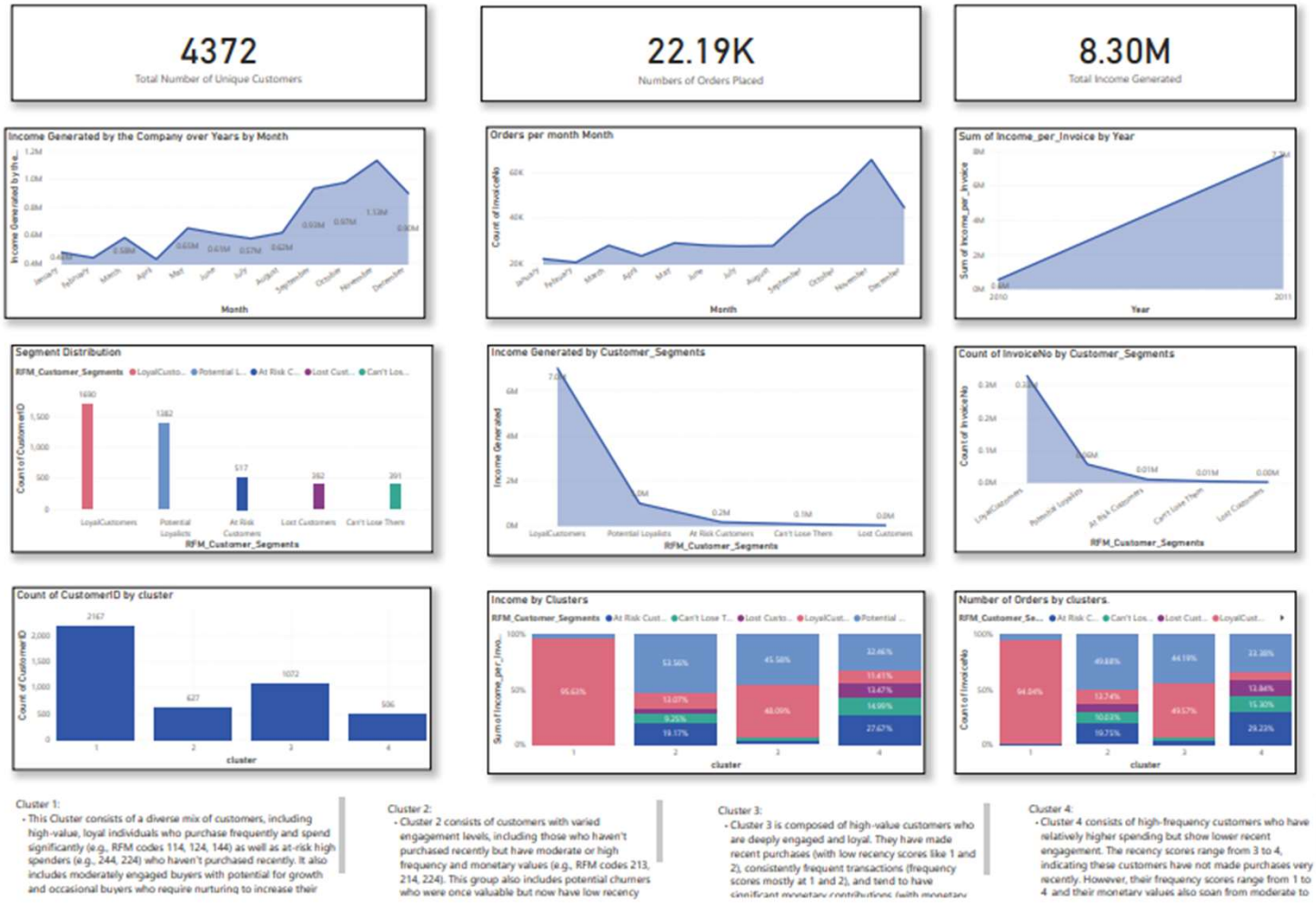
Example: If churn prediction prevents 5% of customers from leaving, and each retained customer represents $500 in lifetime value, the savings on 1000 customers would be $25,000. If the cost of retention initiatives is $5,000, the ROI is 400%.

Projected ROI

Based on our analysis, a **15% reduction** in churn could yield an estimated **ROI of 100%** over the next year, due to retained customer spending and minimized acquisition costs.


VISUALISATION

E - Commerce Customer Segmentation and Sales Analysis



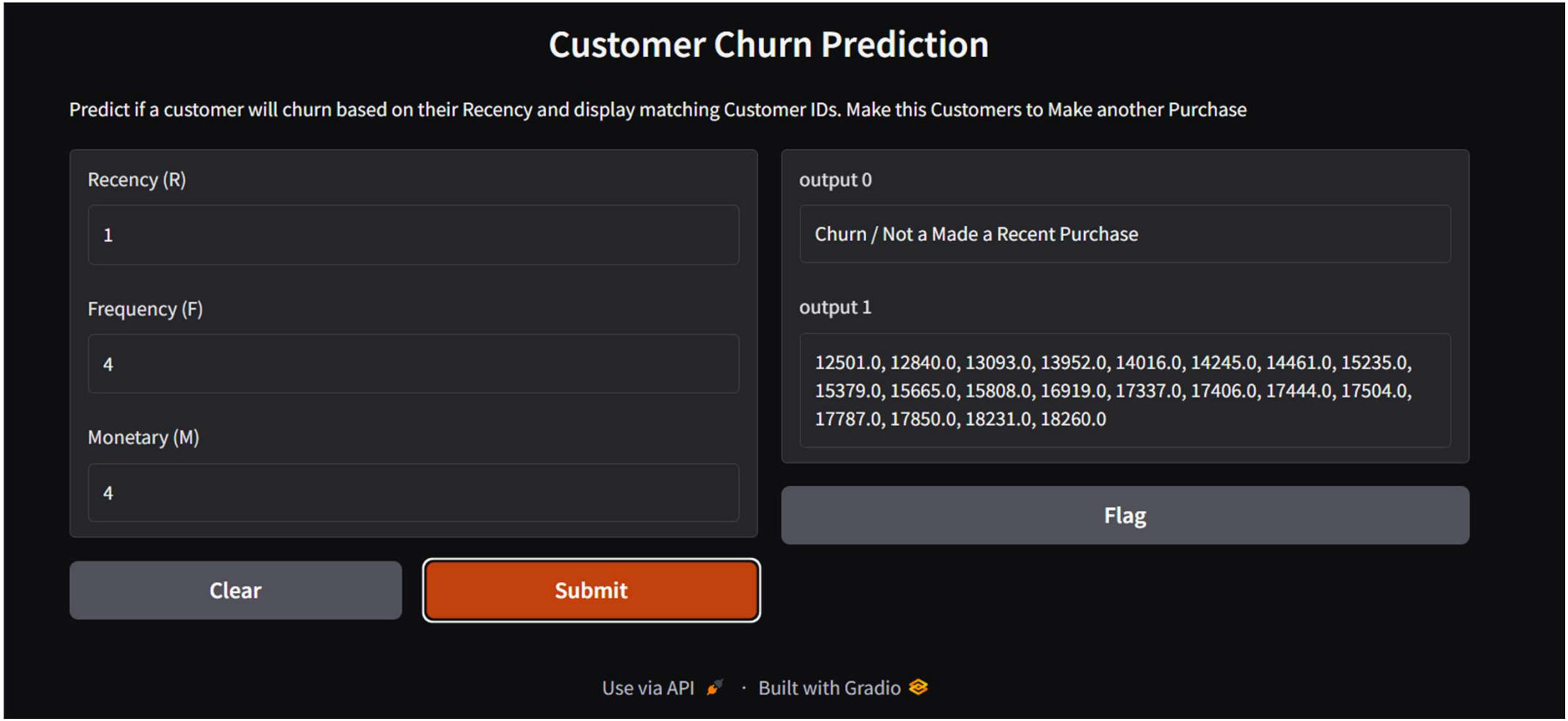
Cluster 1:
- This Cluster consists of a diverse mix of customers, including high-value, loyal individuals who purchase frequently and spend significantly (e.g., RFM codes 114, 124, 144) as well as at-risk high spenders (e.g., 244, 224) who haven't purchased recently. It also includes moderately engaged buyers with potential for growth and occasional buyers who require nurturing to increase their

Cluster 2:
- Cluster 2 consists of customers with varied engagement levels, including those who haven't purchased recently but have moderate or high frequency and monetary values (e.g., RFM codes 213, 214, 224). This group also includes potential churners who were once valuable but now have low recency

Cluster 3:
- Cluster 3 is composed of high-value customers who are deeply engaged and loyal. They have made recent purchases (with low recency scores like 1 and 2), consistently frequent transactions (frequency scores mostly at 1 and 2), and tend to have significant monetary contributions (with monetary

Cluster 4:
- Cluster 4 consists of high-frequency customers who have relatively higher spending but show lower recent engagement. The recency scores range from 3 to 4, indicating these customers have not made purchases very recently. However, their frequency scores range from 1 to 4 and their monetary values also span from moderate to

BIA BOSTON INSTITUTE OF ANALYTICS ®



Future Enhancements

- **Advanced Analytics:** Explore machine learning models beyond Random Forest to improve churn prediction accuracy, such as Gradient Boosting Machines (GBM) or Neural Networks. Utilize predictive analytics to identify emerging trends and adapt strategies proactively.
- **Real-time Customer Engagement**: Develop real-time analytics dashboards to monitor customer interactions and identify churn signals immediately. Implement automated customer communication systems for timely outreach to high-risk customers.
- **Customer Review Analysis**: Implement advanced Natural Language Processing (NLP) techniques to analyze customer feedback and sentiment.Integrate insights from reviews to refine product offerings and customer service strategies.

Conclusion

- In summary, our analysis has provided valuable insights into customer behavior, segmentation, and churn prediction. By leveraging clustering techniques and machine learning models, specifically Random Forest, we have effectively identified distinct customer segments and key drivers of churn.
- The actionable insights derived from this analysis enable tailored marketing strategies, enhancing customer engagement and retention. Implementing targeted interventions for high-risk segments can significantly reduce churn rates, yielding substantial ROI.


Thank You!