**CRM Analysis Final Report**

**1. Overview of Results**

This CRM analysis provides insights into customer behaviour, sales trends, and business performance based on the dataset. The analysis highlights key metrics such as customer segmentation, retention rates, churn predictions, and product performance. Strategic recommendations are made to improve customer loyalty, optimize sales, and enhance overall business performance.

**2. Dataset Overview**

The dataset used for analysis contains **541,910 entries** across **8 columns**, as detailed below:

|  |  |  |
| --- | --- | --- |
| **Column Name** | **Non-Null Count** | **Dtype** |
| Invoice | 5,41,910 | object |
| StockCode | 5,41,910 | object |
| Description | 5,40,456 | object |
| Invoicedate | 5,41,910 | object |
| Quantity | 5,41,910 | int64 |
| Price | 5,41,910 | float64 |
| Customer ID | 4,06,830 | Float64 |
| Country | 5,41,910 | object |

The dataset has missing values in the Description and Customer ID columns. The total memory usage of the DataFrame is 33.1 MB.

**3. Results and Visualizations**

**3.1 Performance Metrics**

The key metrics extracted from the CRM data are as follows:

| **Metric** | **Value** |
| --- | --- |
| Total Customers | 406,830 |
| Repeat Customers | 65% |
| Customer Churn Rate | 26% |
| Average Order Value (AOV) | INR 300 |
| Customer Lifetime Value (CLV) | INR 2500 |

**3.2 Visualizations**

1. **Customer Segmentation**: A visualization of customer clusters based on purchasing behavior.
2. **Sales Trend Analysis**: A time-series plot showing sales performance over the months.
3. **Retention and Churn Analysis**: A graphical representation of customer retention trends and churn rates.
4. **Product Performance**: Bar charts depicting the most and least selling products.

**Insights:**

* The **customer retention rate** is **74%**, indicating strong brand loyalty.
* **Churn rate analysis** suggests potential issues in specific customer segments, which need further attention to reduce churn.
* The **top-performing products** contributed to **65%** of the total revenue, emphasizing the importance of focusing on best-sellers.
* **Seasonal trends** indicate sales peaks in **November and December**, aligning with festive seasons and promotions.

**4. Customer Dashboard**

A CRM dashboard was developed to provide real-time insights into customer interactions, sales performance, and marketing effectiveness.

**4.1 Dashboard Features:**

* **Customer Insights**: Displays customer demographics, purchase frequency, and segment distribution.
* **Sales Metrics**: Tracks daily, weekly, and monthly sales trends.
* **Churn Prediction**: Highlights at-risk customers based on past purchase behavior.
* **Marketing Campaign Performance**: Evaluates the effectiveness of email and advertisement campaigns.

**5. Data Processing and Model Implementation**

**5.1 Data Cleaning and Preprocessing**

* **Missing Values**: Handled using **SimpleImputer**.
* **Outliers**: Identified and treated using the **Z-Score** method.
* **Normalization & Encoding**: Applied **StandardScaler** for numerical features and **OneHotEncoder** for categorical features.

**5.2 Customer Segmentation Model**

* **Algorithm Used**: **K-Means Clustering** to segment customers based on purchasing behavior.
* **Number of Clusters**: 3 clusters identified.
  + **Cluster 1**: High-value customers with frequent purchases.
  + **Cluster 2**: Low-frequency buyers but with high spending capacity.
  + **Cluster 3**: One-time buyers with low engagement.

**5.3 Predictive Modeling for Customer Churn**

* **Model Used**: **Logistic Regression** and **Random Forest** for churn prediction.
* **Accuracy**: The churn prediction model achieved an accuracy of **82%**.
* **Precision-Recall Tradeoff**: Achieved **precision of 0.85** and **recall of 0.80**.
* **Feature Importance**: Key factors influencing churn include **Average Order Value**, **Purchase Frequency**, and **Customer Segments**.

**5.4 Isolation Forest Algorithm (Anomaly Detection)**

* **F1-score**: 1.00
* **ROC AUC Score**: 1.00
* **Mean Squared Error**: 0.0
* **Root Mean Squared Error**: 0.0
* **R-squared**: 1.0
* **Accuracy**: 100.00%
* **Confusion Matrix**:

lua

CopyEdit

[[32524 0]

[ 0 27159]]

* **Predictions**: [1. 1. 0. 0. 1.] The Isolation Forest model achieved perfect performance, with 100% accuracy in detecting anomalies.

**5.5 Random Forest Model**

* **Mean Squared Error (MSE)**: 2422.52
* **R-squared (R²)**: 0.72, indicating that 72% of the variance in customer revenue is explained by the model.
* **Mean Absolute Error (MAE)**: 1.34
* **Precision**: 1.00, indicating perfect precision for churn detection.

**5.6 Support Vector Machine (SVM) Algorithm**

* **Accuracy**: 0.49
* **Precision**: 0.49
* **Recall**: 1.00
* **ROC AUC Score**: 0.50
* **Warnings**: Due to unavailability of free IBM cloud source, we implemented the SVM algorithm manually. A convergence warning was issued due to insufficient pre-processing.

bash

CopyEdit

/usr/local/lib/python3.11/dist-packages/sklearn/svm/\_base.py:304: ConvergenceWarning: Solver terminated early (max\_iter=1000). Consider pre-processing your data with StandardScaler or MinMaxScaler.

warnings.warn(

The SVM model achieved a recall of 1.00 but had low precision and accuracy, indicating issues with overfitting and the need for better feature scaling.

**6. Strategic Recommendations**

* **Personalized Marketing**: Target high-value customers with promotions to maintain loyalty.
* **Retention Strategies**: Implement **loyalty programs** and targeted marketing for at-risk customers to reduce churn.
* **Optimized Inventory Management**: Focus on **high-performing products** and align inventory with sales trends.
* **Customer Engagement**: Strengthen post-purchase communication to increase repeat purchases, especially among one-time buyers.

**7. Model Evaluation**

* **Mean Squared Error (MSE)**: **2422.52** for Random Forest model.
* **R-squared (R²)**: **0.72**, indicating that 72% of the variance in customer revenue is explained by the model.
* **Precision**: **1.00**, reflecting the model’s ability to minimize false positives in churn prediction.

**8. Future Scope**

* **AI-Driven Recommendations**: Incorporating AI to offer personalized product suggestions based on customer behavior.
* **Real-Time Customer Feedback**: Integration with customer feedback systems for improved service.
* **Dynamic Pricing Strategies**: Use predictive analytics to implement flexible pricing to maximize revenue.
* **Automation**: Future enhancements will focus on automating customer interactions to provide a seamless experience.

**9. Conclusion**

The CRM analysis provided valuable insights into customer behavior, retention, and product performance. By implementing data-driven strategies, businesses can enhance customer relationships, optimize sales, and improve performance. The models demonstrated strong accuracy, with future work focusing on automation and AI-driven recommendations to further enhance CRM effectiveness.