Project B : Retail

Objective:

The project's goal is to replicate a real-world situation in which you study consumer data from e-commerce to create predictive models and extract useful insights. Finding trends in consumer behaviour, maximising marketing initiatives, and guiding corporate plans to improve client retention, boost revenue, and customise marketing strategies are the objectives. You can make use of any open source libraries in python.

Dataset:

The UCI Machine Learning Repository's Online Retail II dataset is a real-world dataset that includes transactional data from a UK-based online retailer. It records different facets of retail transactions, including item purchases, customer information, and sales dates. This dataset offers a comprehensive understanding of consumer buying patterns and can be used for a variety of analytics, such as product recommendation systems, sales trends, and customer segmentation. The dataset can be found here:

https://archive.ics.uci.edu/dataset/502/online+retail+ii

Key columns include **InvoiceNo**, **StockCode**, **Description**, **Quantity**, **InvoiceDate**, **UnitPrice**, **CustomerID**, and **Country**.

InvoiceNo is a unique identifier for each transaction, while **StockCode** and **Description** provide product-specific information.

Quantity and **UnitPrice** offer insights into the volume and value of each transaction, allowing for revenue calculations.

CustomerID enables segmentation and customer-based analyses, while **Country** supports geographical segmentation.

Project Components & Key Insights:

1. **Data Cleaning and Exploration:**

    - Analyse historical data, such as customer demographics, browsing sessions, product views, cart additions, and completed sales, as part of the initial dataset analysis.
    - Address missing values, eliminate duplicates, and fix irregularities. To determine the best times to shop, add new options like "Time of Day" and "Day of Week."
    - To uncover significant aspects, use visualisation tools to investigate customer trends, distributions, and correlations.

2. **Customer Segmentation**

    Use clustering techniques (such K-Means or DBSCAN) to divide up your clientele based on their browsing habits, preferences, and purchasing habits.

    - Personalised Marketing: Make use of segments to advise tailored promotions (e.g., offering personalised discounts or suggesting products).
    - Optimised Communication Times: By examining when each consumer category is most engaged, you can ascertain the ideal times to interact with them.

3. **Customer Purchase Prediction:**

    Using information such as time spent on the website, products viewed, and previous purchases, create a classification model to forecast whether a customer's browsing session will end in a purchase.

    - Conversion-Boosting Strategies: To increase conversions, use forecasts to instantly present special offers or free shipping incentives to prospective customers.

    - When to Offer a Discount to Cart Abandoners: Determine the best times and kinds of discounts to entice cart abandoners to finish their purchases.

4. **Sales Forecasting:**

    Use attributes such as goods in a cart, average transaction size, and frequency of purchases to create regression models that predict future sales.

- Seasonal Discount Timing: To ensure steady revenue, pinpoint periods of low sales and recommend well-timed discounts or exclusive deals during off-peak hours.

5. **Product Bundling:**

Using measures like support, confidence, and lift, use association rule mining (such as the Apriori algorithm) to identify products that are frequently purchased together.

- Product Bundling: Provide discounts for combined items to raise the average order size and recommend product bundles based on consumer purchasing patterns.

- Cross-Promotional Opportunities: Find related items and provide real-time, targeted marketing for customers (e.g., "Customers who bought X also bought Y").

Final Report:

A detailed report less than ten pages with the following outline:

1- Introduction: Explain the problem and dataset, and briefly describe your methodology, findings and insights.

2 - Data preprocessing: explaining all steps taken for data cleaning, preprocessing, and feature engineering.

3 - Methodology: Explain model architecture, optimisation policy, and training process.

4 - Experiments: Explain the logic behind using every algorithm that you have used during this project.

5 - Discussion: Cover all requested tasks on visualisations and analyses, providing insights. Based on these findings, you may offer actionable recommendations for strategies, and promotional activities (discounts).

6 - Conclusion: Summarize your findings and understanding of the problem.

Deliverables:

1. **Models for Clustering, Classification, and Regression:** Including code and step-by- step instructions. (30 points)
2. **Association Guidelines and Suggestions:** Code, analysis, and recommendations for cross-promotions and product bundling. (40 points)
3. **Final Report:** An extensive document that includes all methods, conclusions, and practical business suggestions. (30 points)