



# Analyzing Customer Behavior and Churn Prediction in Retail

## Project Title: Analyzing Customer Behavior and Churn Prediction in Retail

### Business Context

A leading retail company has been experiencing a gradual decline in customer retention rates over the past few quarters. This trend has raised concerns regarding long-term revenue stability and market competitiveness. The company offers a wide range of products and frequently runs promotional campaigns to attract and retain customers. Despite these efforts, the churn rate has been increasing, indicating that customers are not sufficiently engaged or satisfied to remain loyal.

The retail company has collected extensive data on customer demographics, purchasing patterns, product interactions, and promotional campaign responses. Recognizing the need to leverage this data for strategic insights, the company has enlisted the expertise of a data science consultancy to identify underlying issues and opportunities for improving customer loyalty.

### Problem Statement

As a data science consultancy assigned to this project, our objective is to analyze the provided datasets to uncover factors that influence customer churn and identify patterns that could predict future churn. The specific goals include:

1. **Understanding Customer Behavior**: Analyze customer demographics, purchasing behaviour, and interactions with promotions to identify distinctive characteristics of customers who churn versus those who remain loyal.
2. **Evaluating Promotional Campaigns**: Assess the effectiveness of promotional campaigns in influencing customer purchasing behaviour and identify which aspects of promotions are most positively associated with customer retention.
3. **Predicting Customer Churn**: Develop a predictive model that can accurately identify customers at high risk of churning based on their behaviour, demographics, and interactions with the company's products and promotions.
4. **Strategic Recommendations**: Provide actionable insights and strategic recommendations to the retail company on how to enhance customer engagement, optimize promotional campaigns, and ultimately reduce the churn rate.

The outcome of this project is expected to empower the retail company with data-driven strategies to improve customer retention, optimize marketing efforts, and enhance overall customer satisfaction, thereby contributing to sustained business growth and competitiveness in the market.

#### Data Description:

The project utilises five main datasets:

1. **Churn Data**: Customer demographics, churn status, and RFM (Recency, Frequency, Monetary) scores.
2. **Customer**: Basic information about customers including demographics and profession.
3. **Products**: Details on products including name, category, brand, price, and release date.
4. **Orders**: Records of customer orders, including product details, quantities, and prices.
5. **Promotions**: Information on promotional campaigns, including discount percentages and campaign durations.

Note on Data Relationships:

* The datasets on customers, products, orders, and promotions are interrelated and should be compatible with each other, forming a consistent pattern reflecting sales and marketing operations.
* Churn Data, however, is separate and distinct from the other four datasets. It may include summaries like the last and first purchase dates or total spent amounts, but these figures might not match precisely with the detailed transactional data captured in the Orders dataset. This is because the transaction data in the Orders dataset could be a subsample of a larger database, potentially used for more granular analyses, while the Churn Data is likely used for higher-level customer retention and behaviour analyses.

The CRISP-DM (Cross-Industry Standard Process for Data Mining) methodology provides a structured approach to planning and executing data mining projects. It consists of six phases: Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment. Here's how you can apply CRISP-DM to solve the capstone project focused on analyzing the impact of the customer retention:

### 1. Business Understanding

* **Objective**: Identify and understand the business objectives and requirements. The goal is to understand factors influencing customer churn and loyalty, and to predict customer churn based on purchasing patterns, demographics, and interactions with products and promotions.

### 2. Data Understanding

* **Data Collection**: Describe the data collection process for the five datasets: churn data, customer, products, orders, and promotions.
* **Initial Data Analysis**: Conduct preliminary analysis to familiarize yourself with the data, understand the data format, quality, and basic properties.
* **Generating Plots:**

1. **Brand and Category Promotions Count**:
   * How many promotions are run by each brand across different product categories?
   * Which brands and categories have the highest number of promotions?
   * Generate a plot that visualizes the count of promotions by brand and category.
2. **Brand and Category Sales Change Percentage**:
   * What is the percentage change in sales for each brand and category before and after promotions?
   * Identify the brands and categories with the most significant positive and negative changes in sales percentage.
   * Generate a plot that compares the percentage change in sales across brands and categories.
3. **Brand and Category Promotion Effectiveness**:
   * How effective are promotions for each brand within various categories in terms of sales volume, customer acquisition, and revenue?
   * Which brands and categories see the most substantial impact from promotions?
   * Generate a plot that measures the effectiveness of promotions by brand and category.

### 3. Data Preparation

* **Data Cleaning**: Address missing values, especially in the 'Profession' column, and convert date columns to datetime format.
* **Merging Datasets**: Combine datasets on relevant keys to create a unified view for analysis.
* **Data Transformation**: Prepare the data for modeling, including normalization, encoding categorical variables, and splitting the data into training and test sets.
* **Feature Engineering**:
  + Deriving new features that may influence churn, such as customer lifetime value, average discount availed, and purchase frequency.
  + Segmenting customers based on RFM analysis for targeted marketing strategies.

### 4. Modeling

* **Model Development**:
  + Implementing machine learning models (e.g., Logistic Regression, Random Forest, Gradient Boosting) to predict customer churn.
  + Evaluating model performance using appropriate metrics (Accuracy, Precision, Recall, F1 Score, ROC-AUC).

### 5. Evaluation

* **Evaluate Results**: Assess the models using metrics such as Accuracy, Precision, Recall, F1 Score, and ROC-AUC to evaluate their performance.
* **Review Process**: Review the modeling and evaluation process to identify any steps that could be improved.
* **Determine Next Steps**: Decide on strategies for model improvement or additional analyses based on the evaluation results.

### 6. Deployment

* **Code Optimization**: Refactor your code for efficiency and readability. Ensure that it adheres to coding standards and best practices.
* **Model Serialization**: Serialize your final model using libraries like pickle or joblib in Python, making it easy to save and load for future predictions.
* **Deployment Script**: Create a deployment script that can load the serialized model and make predictions on new data. This script should handle data preprocessing in a way that mirrors the preprocessing done during the model's training phase.

This CRISP-DM-based structure provides a systematic approach to your project, ensuring that each phase is aligned with both the project's objectives and the standards of the data mining process.

### Comprehensive Deliverables for Submission

1. **Detailed Report/Jupyter Notebook**
   * **Data Analysis**: Include all exploratory data analysis (EDA), feature engineering, and preprocessing steps.
   * **Modeling**: Document the model-building process, including baseline models, advanced models, and neural network implementations. Include code, explanations, and evaluation metrics.
   * **Insights and Recommendations**: Provide a detailed analysis of findings, with data-driven insights and actionable recommendations for the business.
   * **Code and Model Files**: Submit all code files, serialized models, and any scripts for deployment.
2. **PowerPoint Presentation**
   * **Project Overview**: Introduce the project, requirements, and constraints.
   * **Business Understanding**: Discuss the business model, project goals, and any constraints.
   * **Data Exploration and Feature Engineering**: Summarize key insights from the EDA, and describe the feature engineering process.
   * **Modeling Approach and Evaluation**: Present the models tested, evaluation metrics, and model selection rationale.
   * **Insights and Business Impact**: Highlight key findings, potential impacts on customer retention and conversion, and strategic recommendations.
   * **Conclusion and Next Steps**: Recap major findings and outline future work or areas for further investigation.

### What to Include

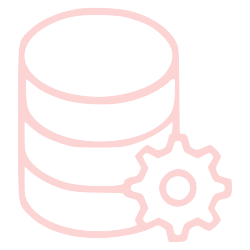
* **Data Analysis and Modeling**: Focus on the methodologies, including any challenges faced and how they were overcome.
* **Clear Visuals**: Use charts, graphs, and models' performance metrics to support your findings.
* **Business Insights**: Emphasize actionable insights and how they can drive business strategy regarding Retail Customer Retention.

### What Not to Include

* **Excessive Technical Details**: While technical depth is necessary, avoid overwhelming non-technical audiences with complex jargon or in-depth coding specifics.
* **Irrelevant Information**: Do not include data or findings that do not contribute directly to understanding the Retail Customer Retention feature's impact or improving business strategies.
* **Unverified Assumptions**: Ensure all claims are supported by data or established research.

### Tips for Effective Presentation

* **Storytelling**: Present your findings as a cohesive narrative that leads the audience through your analytical journey to your conclusions.
* **Engagement**: Design your presentation to engage the audience, using questions, interactive elements, or scenarios to make your points relatable.
* **Professionalism**: Maintain a professional tone and appearance in both your documentation and presentation. Use a consistent theme, legible fonts, and appropriate visuals.

****