# Analytics Plan for Predicting Repeat Customer Behavior in E-Commerce

### Executive Summary: Project Goal & Approach

This plan outlines how we will predict which e-commerce customers are likely to buy again. Our main goal is to help businesses turn one-time buyers into loyal customers. This is important because repeat customers spend more and cost less to keep than finding new ones.

We will use data analysis on the ecommerce\_customer\_data.csv dataset to identify the key factors that predict repeat purchases. The final result will be clear insights and recommendations for marketing teams to improve customer retention.

**The Challenge:** Our journey begins with a critical business challenge faced by e-commerce companies: converting one-time buyers into loyal, repeat customers. We know that repeat customers are the lifeblood of a sustainable online business, contributing significantly more revenue with lower marketing costs than constantly acquiring new ones. The core problem is identifying *who* is likely to become a repeat buyer and *why*.

**The Data:** To tackle this, we have a dataset (ecommerce\_customer\_data.csv) capturing a snapshot of customer behavior and demographics. It contains valuable information like purchasing history (Total Purchases, Average Order Value, Customer Lifetime Value), engagement metrics (Email Engagement Rate, Social Media Engagement Rate, Mobile App Usage), customer service interactions, satisfaction scores, conversion rates from different channels, membership status, return history, and demographic details (Age, Gender, Income Level, Country, City). Crucially, it includes our target: whether a customer is a Repeat Customer.

**Understanding the Data Landscape:** Our initial exploration revealed the dataset's structure and the types of data we're working with. We immediately noticed a **class imbalance** in our target variable (Repeat Customer), with a large majority of customers being repeat buyers. This is a key finding that we'll need to address during modeling to ensure we can accurately identify the minority class (non-repeat customers), which is vital for churn prediction and retention strategies.

We also checked for data quality issues. We found and handled **missing values** through imputation (using median for numerical and mode/None for categorical). Visualizing missingness with a heatmap helped us see the pattern of these gaps. We also identified and prepared to handle **duplicate records** and potential **constant value columns**, although none were found in our initial check after cleaning. We addressed obvious **outliers** like negative age or lifetime value by converting them to missing values for imputation.

We created a new feature, **Customer Tenure**, from the registration date, which provides valuable context about the customer's relationship duration.

**Exploring Relationships:** Through visualizations, we started to uncover patterns:

* **Univariate Plots** showed the distributions of individual features. We saw the range of purchase values, engagement rates, and the spread of demographics. Skewness in some numerical features was noted, which might be considered for transformations later.
* **Bivariate Analysis** gave us insights into how different features relate to repeat customer status. We could visually see if certain income levels, countries, favorite categories, or engagement levels tend to have a higher proportion of repeat customers. Similarly, boxplots showed us how the distributions of numerical features like Total Purchases or Customer Lifetime Value differ between repeat and non-repeat customers – likely showing higher values for repeat buyers, as expected.
* The **Correlation Matrix** highlighted linear relationships between numerical features. This helps us understand potential multicollinearity issues and see which numerical features have stronger correlations with each other and, importantly, with our numerical target variable (Repeat Customer\_Num).

**The Path Forward:** Based on these insights, our plan is to build a **Logistic Regression model**. This model is chosen for its interpretability, which will allow us to understand *which* specific factors identified in our EDA are the strongest predictors of repeat customer behavior. We will address the class imbalance using a technique like SMOTE to ensure our model is not biased towards the majority class and can effectively identify customers who are *not* likely to repeat, allowing for targeted retention efforts.

**The Goal:** The ultimate goal is to translate the model's findings – the key predictive factors and their influence – into **actionable recommendations** for e-commerce marketers. This will empower them to move beyond generic campaigns and implement data-backed strategies to nurture one-time buyers, increase customer loyalty, and ultimately drive profitability.

In essence, we are using data to understand the drivers of customer loyalty, build a predictive tool, and provide strategic guidance to turn insights into business action.

### 1. Project Overview and Business Context

This foundational section explains the project's purpose, why it's important for the business, and what we expect to achieve.

| **Field** | **Description** |
| --- | --- |
| **Project** | Predicting Repeat Customer Behavior in E-Commerce |
| **Preparer** | Chidiebere F. Odurukwe, 301309004 |
| **Requestor** | Bilal Hasanzadah |
| **Date of Request** | July 2nd 2025 |
| **Target Quarter for Delivery** | 2nd Quarter |
| **Epic Link(s)** | <https://www.kaggle.com/datasets/noir1112/e-commerce-customer-engagement?resource=download> |
| **Business Impact** | Optimizing customer retention and increasing overall profitability |

#### 1.0 Problem Statement: Who, What, When, Where, Why, and How

This project addresses a critical business challenge within the e-commerce sector, aiming to transform one-time buyers into loyal, recurring customers.

* Who:
  + Focus: Individual e-commerce customers.
  + Requester: Chidiebere F. Odurukwe.
  + Audience for Insights: E-commerce marketers.
* What (Problem & Goal):
  + Problem: E-commerce businesses struggle to convert one-time buyers into loyal, repeat customers, leading to inefficient marketing spend and missed revenue opportunities.
  + Goal: To identify and quantify the key factors that influence a customer's likelihood of making subsequent purchases, thereby predicting repeat customer behavior. The ultimate objective is to provide actionable insights and recommendations to improve customer retention.
* When:
  + Data Observation Window: The ecommerce\_customer\_data.csv dataset covers customer behavior from 2018-01-11 to 2023-12-31.
  + Project Timeline: The project is scheduled from June 16, 2025, to August 13, 2025, with delivery targeted for the 2nd Quarter.
* Where:
  + Context: The e-commerce sector.
  + Primary Data Source: The ecommerce\_customer\_data.csv dataset, publicly available on Kaggle.
* Why:
  + Business Rationale: Repeat customers are crucial for sustainable e-commerce success because they drive significantly more revenue and require fewer marketing resources compared to acquiring new customers.
  + Impact of Success: This project will optimize customer retention, increase Customer Lifetime Value (CLTV), enable smarter marketing spending, and enhance overall customer loyalty and satisfaction.
  + Implications of Inaction: Without this analysis, the business risks continued inefficient spending on customer acquisition, stagnant repeat purchase rates, and potential loss of competitive advantage.
* How:
  + Methodology: Data analysis, primarily utilizing Logistic Regression for its interpretability and suitability for binary classification.
  + Key Steps: This includes comprehensive data acquisition, preprocessing (cleaning, handling missing values and outliers, feature engineering, categorical encoding), Exploratory Data Analysis (EDA), addressing class imbalance using SMOTE, predictive model building, rigorous evaluation (using Recall, F1-Score, ROC AUC), interpretation of results, and clear visualization/reporting.

#### 1.1 Business Opportunity Brief

The core business challenge addressed by this project is the imperative for e-commerce businesses to convert one-time buyers into loyal, repeat customers. Repeat customers are crucial for e-commerce success because they drive revenue and are cheaper to keep. This project aims to use data to figure out what makes customers buy again. By understanding these factors, businesses can improve their marketing to keep more customers and make more profit.

We will use the provided ecommerce\_customer\_data.csv dataset to predict if a customer will become a repeat buyer. This involves finding out which customer traits and behaviors are most linked to repeat purchases.

#### 1.2 Supporting Insights

Keeping customers is a key business focus in e-commerce. Our project uses data to support this by predicting repeat purchases. We'll use insights from marketing research to guide our analysis.

(Note: In a real business project, we would also look at what competitors are doing for customer retention. For this project, we focus on analyzing our own data.)

#### 1.3 Project Gains

This project will help marketing teams make better decisions based on data. By knowing which factors predict repeat buying, they can create more effective campaigns to keep customers.

Expected benefits:

* **More Repeat Purchases:** Directly increases customer loyalty.
* **Higher Customer Value:** Customers stay longer and spend more over time.
* **Smarter Marketing Spending:** Focus efforts on customers likely to become repeat buyers.
* **Happier Customers:** More relevant offers and better experiences.

If we don't do this, the business might keep spending a lot to find new customers without improving loyalty, missing out on potential profit and falling behind competitors.

### 

### 2. Defining Analytical Objectives and Success Measures

This section states our specific goals, what we assume about the data, and how we will know if our project is successful.

#### 2.0 Analytical Objective

Our main question is: "What factors best predict if a customer will buy again, and how can businesses use this to improve marketing?"

We believe certain factors like total purchases, order value, engagement rates (email, social media), and membership status will be strong predictors. Our analysis will test this.

The final result will be clear insights, simple rules, and visual reports that marketing teams can use to make decisions. Asking the right questions helps us choose the best data and methods.

#### 2.1 Other Related Questions and Underlying Assumptions

Every project has assumptions. Ours include:

* The dataset represents typical e-commerce customer behavior.
* The 'Repeat Customer' data accurately shows repeat buying.
* The features in the dataset are good predictors.
* We can fix data issues like missing values and outliers.

Knowing these assumptions helps us be aware of possible limits and focus on cleaning the data well.

#### 2.2 Success Measures and Metrics

Success means building a model that accurately predicts repeat buyers and provides useful recommendations for marketers.

We'll use specific measures to check our model's performance:

**Technical Success Metrics:**

* **Recall:** How well we find *all* the actual repeat customers (important to avoid missing potential loyal buyers).
* **F1-Score:** A balance between finding repeat customers (Recall) and being correct in our predictions (Precision).
* **ROC AUC:** Measures how well the model separates repeat from non-repeat customers overall.
* **Precision:** How often our prediction of a repeat customer is correct.
* **Accuracy:** Overall correct predictions (less reliable for imbalanced data).
* **Model Validation:** The model's performance will be rigorously validated to ensure its reliability and generalizability. A standard train-test split will be used to evaluate the model on unseen data. Given the class imbalance, metrics beyond accuracy (specifically Recall, F1-Score, and ROC AUC) will be prioritized for evaluating the model's effectiveness in identifying the minority class ('No' - potential churners).

**Business Success Indicators:**

* **Improved Repeat Purchase Rate:** This is the ultimate business outcome, directly reflecting an increase in customer retention and loyalty.
* **Increased Customer Lifetime Value (CLTV):** By effectively retaining customers, their long-term financial contribution to the business is expected to grow.
* **Optimized Marketing Spend:** The project's findings will enable more efficient allocation of marketing resources by allowing businesses to target high-potential repeat customers more effectively.
* **Enhanced Customer Loyalty:** This qualitative measure reflects an improvement in customer relationships and brand affinity, driven by more personalized and relevant interactions.

Showing how this model results connect to business gains is key.

### 

### 3. Proposed Analytical Approach and Techniques

This section describes the steps and methods we will use to analyze the data and build the model.

#### 2.3 Methodology and Approach

We will follow a clear process from getting the data to sharing our findings.

Our main method is **Logistic Regression**. This is a good choice because:

* It's designed for predicting a Yes/No outcome like 'Repeat Customer'.
* It's **interpretable**: We can see how each factor influences the likelihood of repeat buying, which is great for making business recommendations.
* It's efficient.
* It serves as a strong, interpretable baseline model.

Here are the main steps:

1. **Data Acquisition:** Obtain the ecommerce\_customer\_data.csv dataset.
2. **Data Preprocessing and Cleaning:** This critical phase involves preparing the raw data for modeling, ensuring data quality and suitability.
   * **Variable Selection and Exclusion:** Review variables for relevance.
   * **Handling Missing Values:** Address missing values (e.g., nan or blank entries) using appropriate imputation strategies (median for numerical, mode/placeholder for categorical).
   * **Handling Outliers:** Identify and manage outliers (e.g., negative values in Age/CLTV) by replacing them with NaN and then imputing.
   * **Data Type Conversion and Consistency:** Ensure data types are appropriate for analysis (e.g., converting date strings to datetime objects).
   * **Handling Duplicates:** Check for and remove any completely duplicate records.
   * **Handling Skewness:** For numerical predictors, assess skewness and consider transformations if necessary.
   * **Feature Engineering:** Create new features like CustomerTenureDays from RegistrationDate.
   * **Categorical Encoding:** Convert categorical features into a numerical format suitable for logistic regression using techniques like One-Hot Encoding.
   * **Dimensionality Reduction:** Consider if features become excessively large, though interpretability will be prioritized.
3. **Exploratory Data Analysis (EDA):** Conduct in-depth EDA to understand data patterns, relationships between variables, and distributions using appropriate visualizations.
   * **Univariate Analysis:** Examine individual feature distributions.
   * **Bivariate Analysis:** Analyze relationships between predictors and the target variable.
   * **Correlation Analysis:** Assess correlations between numerical features.
   * **Class Imbalance Check:** Explicitly check the distribution of the target variable.
4. **Addressing Class Imbalance:** Given the identified significant class imbalance in the target variable (RepeatCustomer), resampling techniques will be applied to the training data to mitigate its impact.
   * **Strategy:** Oversampling of the minority class using **SMOTE (Synthetic Minority Over-sampling Technique)** will be the primary strategy. SMOTE generates synthetic samples for the minority class, helping to balance the class distribution in the training data without simply duplicating existing instances.
   * **Justification for SMOTE:** SMOTE is chosen because it helps the model learn the characteristics of the minority class more effectively. Compared to undersampling, it retains all information from the majority class, which can be beneficial given the dataset size. This is particularly important for improving recall for the minority class, allowing the model to identify more potential non-repeat customers.
   * **Implementation:** SMOTE will be applied *only* to the training data *after* the train-test split to prevent data leakage.
5. **Predictive Modeling with Logistic Regression:**
   * **Target Variable Definition:** RepeatCustomer (Yes/NO) will be explicitly defined as the binary outcome variable.
   * **Objective:** Identify and quantify the influence of various predictor variables on the likelihood of a repeat purchase.
   * **Model Building:** Split data into training and testing sets. Apply SMOTE to the training set. Train the logistic regression model on the oversampled training data.
6. **Evaluation:** Rigorously measure model performance, prioritizing metrics suitable for imbalanced datasets.
   * **Key Metrics:** Recall and F1-Score will be primary.
   * **Supplementary Metrics:** ROC AUC, Precision, and Accuracy will also be used.
   * **Confusion Matrix:** Generated for detailed breakdown of predictions.
7. **Understand Results:** Analyze the trained model's coefficients to understand the impact and significance of different features on repeat customer likelihood.
8. **Share Findings:** Create clear reports and visuals for the marketing team.

The final output will be actionable insights and recommendations to help businesses improve retention.

### 

### 4. Data Landscape and Variable Considerations

This section provides information about the data we are using, its size, and any important considerations.

#### 3.0 Population, Variable Selection, and Considerations

We are mainly using the ecommerce\_customer\_data.csv dataset from Kaggle. It contains 10,000 rows and 23 columns, covering customer details, purchase history, and engagement. The data covers a period from 2018-01-11 to 2023-12-31.

We will also be aware of other potential datasets (like the UCI Online Shoppers dataset or Google Analytics data) for context, though our main analysis uses the primary CSV.

The data is about individual customers.

**Data Structure and Types:**

* **Shape:** The dataset contains 10000 rows and 23 columns.
* **Data Types:** Includes numerical (float64) and categorical (object) data types. Careful conversion and consistency checks will be performed.

**Data Sources:**

1. **Primary Dataset:** ecommerce\_customer\_data.csv
   * **Description:** Contains e-commerce metrics such as RepeatCustomer status, TotalPurchases, AverageOrderValue, CustomerLifetimeValue, EmailEngagementRate, SocialMediaEngagementRate, MobileAppUsage, CustomerServiceInteractions, AverageSatisfactionScore, EmailConversionRate, SocialMediaConversionRate, SearchEngineConversionRate, PremiumMember, and HasReturnedItems, along with demographic information like Age, Gender, IncomeLevel, Country, City, and FavoriteCategory/SecondFavoriteCategory.
   * **Role:** This dataset will serve as the primary source for building and rigorously testing the classification model designed to predict repeat purchase behavior.

The audience level for this analysis is at the individual customer level.

Variable Selection:

All relevant variables from the ecommerce\_customer\_data.csv dataset have been identified for this project, comprising one target variable and twenty-one potential predictor variables. A clear overview of these variables is provided below:

| **Variable Name** | **Variable Type** | **Description** | **Relevance to Repeat Customer Behavior** |
| --- | --- | --- | --- |
| RepeatCustomer | Target (Binary) | Indicates whether the customer has made more than one purchase (Yes/No). | The primary outcome variable the model aims to predict. |
| CustomerID | Identifier | Unique identifier for each customer. | Used for tracking individual customers; not a predictor. |
| RegistrationDate | Date | Date of customer registration. | Can be used to derive customer tenure or recency. |
| Age | Predictor (Numeric) | Age of the customer. | Demographic factors that may influence purchasing habits and loyalty. |
| Gender | Predictor (Categorical) | Gender of the customer. | Demographic factors that may influence product preferences. |
| IncomeLevel | Predictor (Categorical) | Customer's income level (e.g., Low, Medium, High, Very High). | Economic factors influencing purchasing power and frequency. |
| Country | Predictor (Categorical) | Customer's country of residence. | Geographic factors that may influence market trends and product availability. |
| City | Predictor (Categorical) | Customer's city of residence. | Geographic factors that may influence local market trends. |
| TotalPurchases | Predictor (Numeric) | Total number of purchases made by the customer. | Direct indicator of past purchasing activity, highly relevant for repeat behavior. |
| AverageOrderValue | Predictor (Numeric) | Average monetary value of a customer's orders. | Indicates spending habits; higher values may correlate with loyalty. |
| CustomerLifetimeValue | Predictor (Numeric) | Predicted total revenue a customer will generate over their relationship with the business. | A key metric for customer value, often correlated with repeat behavior. |
| FavoriteCategory | Predictor (Categorical) | Customer's most frequently purchased product category. | Indicates product preference, which can drive repeat purchases within that category. |
| SecondFavoriteCategory | Predictor (Categorical) | Customer's second most frequently purchased product category. | Provides additional insight into product preferences. |
| EmailEngagementRate | Predictor (Numeric) | Rate at which customers interact with marketing emails. | Higher engagement may indicate stronger brand interest and loyalty. |
| SocialMediaEngagementRate | Predictor (Numeric) | Rate at which customers interact with social media content. | Reflects brand affinity and active participation. |
| MobileAppUsage | Predictor (Categorical) | Level of mobile app usage (High, Low, Never). | Indicates preferred shopping channel, potentially influencing convenience and repeat visits. |
| CustomerServiceInteractions | Predictor (Numeric) | Number of times a customer has interacted with customer service. | Can indicate issues or strong engagement; may correlate with satisfaction. |
| AverageSatisfactionScore | Predictor (Numeric) | Average customer satisfaction score. | Higher satisfaction is generally linked to increased loyalty and repeat purchases. |
| EmailConversionRate | Predictor (Numeric) | Rate at which email interactions lead to purchases. | Direct measure of marketing effectiveness and customer responsiveness. |
| SocialMediaConversionRate | Predictor (Numeric) | Rate at which social media interactions lead to purchases. | Direct measure of social media marketing effectiveness. |
| SearchEngineConversionRate | Predictor (Numeric) | Rate at which search engine interactions lead to purchases. | Indicates effectiveness of organic/paid search channels. |
| PremiumMember | Predictor (Binary) | Indicates if the customer is a premium member (Yes/No). | Membership status often correlates with higher loyalty and repeat purchases. |
| HasReturnedItems | Predictor (Binary) | Indicates if the customer has returned items (Yes/No). | Can indicate dissatisfaction or a complex purchasing pattern. |

For research purposes I initially tried to include some variables that were hypothesized, such as TimeOnWebsite, Likes, Shares, Comments, PageViews, Bounce Rate, DeviceType, and ReferralSource, are not directly available in the ecommerce\_customer\_data.csv dataset.[1] However, related metrics like EmailEngagementRate, SocialMediaEngagementRate, MobileAppUsage, and various conversion rates (EmailConversionRate, SocialMediaConversionRate, SearchEngineConversionRate) will serve as valuable proxies for customer engagement and acquisition channels.

Assumptions and Data Limitations:

Several considerations regarding data structure and limitations are acknowledged.

* The dataset is cross-sectional, limiting analysis of dynamic temporal trends.
* RepeatCustomer is binary, simplifying loyalty nuances.
* Anticipate missing values (nan or blank entries) and negative values (e.g., in Age, CustomerLifetimeValue), requiring robust preprocessing.
* The target variable (RepeatCustomer) has a significant class imbalance (80% Yes, 20% No), which needs to be addressed during modeling to avoid biased predictions.

### 

### 5. Potential Challenges and Mitigation Strategies

This section lists potential issues that could come up during the project and our plans to manage them.

#### 4.0 Dependencies and Risks

Identifying potential risks before they materialize is a critical component of effective project management in analytics.

| **Risk** | **Description** | **Likelihood** | **Impact** | **Mitigation** |
| --- | --- | --- | --- | --- |
| **Data Quality Issues** | Missing values (nan, blank entries), errors (negative values in Age, CustomerLifetimeValue), or inconsistencies could compromise analysis. | Medium | Potential for project delays, reduced model performance, inaccurate findings. | Implement thorough EDA, robust cleaning (imputation, outlier handling), and clear documentation. |
| **Limited Record Count** | If usable data size decreases significantly after cleaning, it might limit model statistical power or generalizability. | Low | Increased risk of model overfitting, less reliable predictions. | Utilize cross-validation, prioritize robust feature selection, clearly articulate limitations. |
| **Multicollinearity among Predictors** | Highly correlated variables (e.g., EmailEngagementRate and EmailConversionRate) can affect model interpretability and stability. | Medium | Difficulty interpreting individual predictor contributions, less reliable coefficients. | Conduct correlation analysis, consider feature engineering (composite scores), or regularization. |
| **Definition Ambiguity of "RepeatCustomer"** | The binary Yes/No definition might not capture all nuances of customer loyalty (e.g., purchase frequency, recency, monetary value). | Low | Model predictions might align with definition but miss broader business interpretations. | Adhere strictly to defined target; acknowledge limitation in report, suggest future nuanced definitions. |
| **Class Imbalance** | The target variable (RepeatCustomer) is significantly imbalanced (80% Yes, 20% No), potentially biasing the model against the minority class (non-repeat customers). | High | Reduced model performance on minority class, limiting actionable insights for churn risk. | Employ resampling techniques (e.g., SMOTE) on training data; use appropriate metrics (Precision, Recall, F1-score, ROC AUC). |

Thinking about these problems ahead of time helps us plan better and increases our chances of success.

### 6. Project Milestones and Deliverables

This section shows the main steps of the project and when we plan to complete them, presented as a project roadmap.

#### 5.0 Deliverables

The project will adhere to a structured, phased approach, aligning with typical analytics project lifecycles and academic deadlines. This work-back schedule sets clear expectations for stakeholders and provides a framework for tracking progress.

| **Item** | **Major Events / Milestones** | **Description** | **Scope** | **Days** | **Date** |
| --- | --- | --- | --- | --- | --- |
| 1. | **Kick-off / Formal Request** | Official start of the project and confirmation of scope (Capstone Declaration Submission). | Define project scope, goals, and initial requirements. | 1 day | June 16, 2025 |
| 2. | **Assessment / Triage** | Initial review of project scope, resources, time, data availability (ecommerce\_customer\_data.csv). | Confirm project feasibility and refine initial plan. | 2 days | June 17-18, 2025 |
| 3. | **Prioritization** | Align project goals with business objectives and confirm priorities. Finalize analytical approach. | Confirm key objectives and analytical methods. | 2 days | June 19-20, 2025 |
| 4. | **Data Exploration & Analysis** | Comprehensive data loading, cleaning, preprocessing, feature engineering, and in-depth EDA. Includes addressing issues like duplicates, missing values, outliers, and checking data distributions and relationships. | Prepare data for modeling, understand data patterns and issues. | 15 days | June 23 - July 11, 2025 |
| 5. | **Story Board 1 (Initial Model)** | Build the initial Logistic Regression model, handle class imbalance (SMOTE), train, validate, and interpret results. Develop draft findings and preliminary visualizations. | Develop initial model and insights. | 5 days | July 14-18, 2025 |
| 6. | **QA Output** | Quality assurance of model outputs, data accuracy, and consistency of conclusions. | Verify model results and findings. | 2 days | July 21-22, 2025 |
| 7. | **Internal Team Presentation** | Present initial findings and model to the internal project team for feedback. | Gather feedback and ensure alignment. | 1 day | July 23, 2025 |
| 8. | **Go/No Go** | Formal review point to decide whether to proceed to the next phase based on internal feedback and model performance. | Decision to finalize report or require further work. | 1 day | July 24, 2025 |
| 9. | **Story Board 2 (Refined Model)** | Refine the model based on feedback, optimize performance, develop detailed actionable recommendations. Enhance reports and visualizations. | Finalize model, insights, and recommendations. | 6 days | July 25 - August 1, 2025 |
| 10. | **Pilot (Conceptual)** | Outline how the project's recommendations *would* be implemented and tested in a real-world scenario (e.g., A/B testing). | Plan for real-world application of insights. | 5 days | August 4-8, 2025 |
| 11. | **Delivery & Sign-off** | Finalize all project deliverables (report, presentation) and submit. Obtain formal sign-off. | Submit final project deliverables. | 3 days | August 11-13, 2025 |

**Justification for Milestone Inclusion:** The inclusion of milestones such as "Go/No Go" and "Pilot," while standard in industry, serves a vital purpose in this academic project. These stages demonstrate a sophisticated understanding of comprehensive project management principles. The "Go/No Go" milestone functions as a critical internal review gate, ensuring that the project progresses only after rigorous evaluation and alignment with objectives, thereby minimizing risks and maximizing the likelihood of a successful outcome. Similarly, the "Pilot" stage, even when conceptualized for a capstone, showcases an appreciation for the full project lifecycle by requiring a detailed proposal for real-world implementation. This reflects a practical, business-oriented mindset and enhances the academic rigor by connecting theoretical analysis to practical application.

### 

### 7. Expected Outcomes and Actionable Business Impact

This final section summarizes the main results we expect and how they will help the business.

The project is expected to clearly show which marketing efforts lead to repeat customers. The main result will be **actionable recommendations** that marketing teams can use to improve their strategies, boost customer loyalty, and increase repeat purchases.

These insights should lead to:

* More revenue from loyal customers.
* Saving money on marketing by focusing on retention.
* Better customer relationships.

Not doing this project would mean missing out on these benefits, continuing to spend inefficiently, and potentially losing ground to competitors.

Sharing the findings with clear explanations and visuals will show the value of the analysis and meet the project's goals.