

Dataset Source Selection

The dataset was obtained from Kaggle, a platform known for hosting datasets for various purposes. The specific dataset used in this assignment is titled "E-commerce Customer Churn Analysis and Prediction" and can be accessed [here](#).

Key Attributes

Identified and highlighted essential attributes in the Kaggle dataset that are crucial for the analysis. These include "CustomerID," a unique identifier for each customer, "Churn," indicating customer retention status, and "Tenure," reflecting the duration of customer association with the e-commerce platform.

Attribute Suitability and Replacement

Replaced the "Location" variable with "WarehousetoHome" to quantify the geographical distance between a customer's home and the warehouse. This enhancement provides valuable insights into delivery times and service coverage, aligning with the project's goals. Additionally, variable replacements were implemented, such as substituting "Total Purchases" with "OrderCount" for a more relevant metric aligning with the assignment's objectives. These modifications contribute to a nuanced understanding of customer behavior and enhance the dataset for the intended analysis. Substituted the "Total Purchases" variable with "OrderCount" to better align with the assignment's objectives. "OrderCount" represents the total number of orders placed in the last month, providing a more relevant metric for customer purchasing behavior.

Variable Generation:

A new variable, "Membership Level," has been created, categorizing customers into different membership tiers (Not Member, Bronze, Silver, Gold, Platinum) based on specific conditions involving city tier, tenure, and app engagement.

The "Membership Level" variable was created with specific conditions for different tiers:

Not Member: Default level assigned to all customers.

Bronze: Customers in city tiers 1 and 2, with tenure greater than 2, and spending more than 2 hours on the app.

Silver: Customers in city tiers 2 and 3, with tenure greater than 4, and spending more than 2 hours on the app.

Gold: Customers in city tiers 2 and 3, with tenure greater than 8, and spending more than 3 hours on the app.

Platinum: Customers in city tier 3, with the highest tenure, and spending more than 3 hours on the app.

Identified and included various additional attributes present in the Kaggle dataset that were not explicitly mentioned in the assignment guidelines. These attributes, such as "PreferredLoginDevice," "PreferredPaymentMode," "HourSpendOnApp," and others, contribute to a more comprehensive analysis of customer behavior and preferences.

Jupyter Notebook for Data Generation:

Executed the necessary code in a Jupyter Notebook to implement the proposed data modifications. This includes creating the "Membership Level" variable and performing other adjustments to enhance the dataset's relevance to the project objectives.

Dataset Saving:

Saved the modified dataset as a CSV file using Jupyter Notebook, ensuring that the changes made during the data preprocessing and modeling are done using at least three no-code softwares.

Objective Alignment:

Ensured that the dataset structure closely aligns with the outlined objectives of the assignment, emphasizing the exploration of customer behavior, predictive analysis, and the evaluation of different models.

Dataset Overview:

The resulting dataset provides a comprehensive view of customer behavior, satisfaction, and feedback. It serves as a valuable resource for e-commerce analysis, offering insights that can inform strategic decision-making processes within the business.