A PROJECT PROPOSAL ON

CUSTOMER LIFETIME VALUE PREDICTION

Group Number – 11

Presented By-

1] Mrunal Madhukar Gavit
2] Srushti Anil Kamble
3] Pradnya Sham Jagtap
4] Akshay Kamlakar Parate

Under the guidance of-
Prof. Hong Do

Submitted to

STEVENS INSTITUTE OF TECHNOLOGY

IN PARTIAL FULFILLMENT OF THE REQUIREMENT FOR THE COURSE

STATISTICAL METHODS
(MA 541-A)

1) Source of dataset

The dataset originates from the UC Irvine Machine Learning Repository.

Here is the link to the dataset source: https://archive.ics.uci.edu/dataset/502/online+retail+ii

2) Summary of dataset

Dataset Overview:
This dataset is well-suited for the analysis of online retail transactions, customer behavior, and sales patterns. Preceding analysis, the dataset has undergone thorough cleaning, eliminating the need for further processing. It comprises eight key attributes:

i. Invoice: Uniquely identifies each transaction.
ii. Stock Code: Represents a unique code for each product.
iii. Description: Provides details about the product associated with the stock code.
iv. Quantity: Indicates the quantity of each product in a given transaction.
v. Price: Represents the unit price of each product.
vi. Invoice Date: Specifies the date and time of each transaction.
vii. Customer ID: Identifies the customer linked to the transaction.
viii. Country: Indicates the location of the customer, with a specific focus on transactions in France.

The dataset encompasses qualitative elements such as invoice numbers, stock codes, product descriptions, customer IDs, and country names. Additionally, quantitative variables include quantities and prices, while temporal information is captured by the invoice and date attributes. It is noteworthy that although not explicitly mentioned, the 'Description' attribute likely contains textual data describing the products.

3) Propose strategic plan

The forthcoming analysis focuses on Customer Lifetime Value (CLV), a critical metric for enterprises aspiring to gain comprehensive insights into the enduring value each customer brings to the organization. The analysis aims to provide a nuanced understanding by delving into various dimensions, enabling the computation of key metrics such as:

i. Purchase Frequency: Unveiling how frequently, on average, customers engage in transactions within the observed timeframe.
ii. Average Order Value: Rigorously assessing the average monetary value of each customer's transaction.
iii. Fluctuations in Average Order Value: Investigating potential variations in average order value over time or among distinct customer segments.
iv. Customer Retention Rate: Evaluating the overall customer retention rate and discerning potential fluctuations across different periods or customer cohorts.

The multifaceted exploration of these dimensions promises to offer valuable insights into customer loyalty dynamics and identification of high-value customer segments. This, in turn, empowers businesses to formulate nuanced and targeted strategies for enhancing customer retention, optimizing acquisition efforts, and ultimately driving revenue growth. By leveraging the rich dataset, this comprehensive CLV analysis seeks to be a strategic compass for businesses navigating the intricate landscape of customer relationship management.

4) Suggestion section

i. Kindly provide guidance on the specific tasks and objectives expected in the statistical inference segment.
ii. Please advise on the optimal number of variables to be selected for inclusion in the statistical inference portion of the analysis.