**Customer Segmentation Proposal**

**Juan Segovia, Ernesto Garcia, Julia McKinnon, Daphine Nyangena, Grant Ritzwoller**

**Project Proposal: Customer Segmentation for E-commerce Companies (Unsupervised Approach)**

**Project Overview:**

This project aims to perform customer segmentation for e-commerce companies by analyzing transaction data and customer behaviors. By leveraging unsupervised machine learning (ML) techniques, we will cluster customers based on purchasing patterns, demographic information, and behavioral metrics. This segmentation will provide insights into customer groups, enabling personalized marketing, enhanced user experiences, and improved customer retention strategies. The project will include interactive visualizations and a user-friendly interface to explore customer segments.

**Problem Statement:**

E-commerce platforms cater to a diverse customer base with varying purchasing behaviors and preferences. Understanding these customer segments is crucial for targeted marketing, product recommendations, and loyalty programs. This project addresses the challenge of identifying distinct customer segments by analyzing transactional, behavioral, and demographic data through clustering techniques.

**Objectives:**

* Segment customers based on purchasing behavior, transaction patterns, and demographics.
* Analyze customer lifetime value (CLV) and purchasing frequency.
* Visualize customer segments and key behavioral trends interactively.

**Project Workflow:**

1. **Data Collection**

* **Transactional Data:** Gather data on customer purchases, including transaction amount, frequency, and items purchased.
* **Demographic Data:** Collect customer age, location, and account age details.
* **Behavioral Data:** Extract information on payment methods, device usage, and declined transactions.
* Ensure the dataset contains at least 100 records for robust clustering.

1. **Data Preprocessing**

* Use Python Pandas to clean and merge datasets from multiple sources.
* Handle missing or inconsistent data through imputation or removal.
* Perform standardization or normalization to scale features for clustering.

1. **Feature Engineering**

* Aggregate transactional features (total spending, average order value, transaction count).
* Extract time-based features (recency, preferred transaction hour).
* Engineer behavioral features (fraud rate, decline rate, payment method diversity).
* Derive geographic and demographic insights (average customer location, age clustering).

1. **Machine Learning (Unsupervised)**

* Train clustering models to segment customers based on engineered features.
* Apply Principal Component Analysis (PCA) to reduce dimensionality for better cluster visualization.
* Evaluate cluster quality using silhouette scores and within-cluster sum of squares (WCSS).

1. **Visualization**

* Use Python Matplotlib to create static visualizations of customer segments.
* Develop interactive dashboards with Plotly or Tableau to explore segment characteristics.

**Technologies and Tools:**

* **Programming Languages:** Python.
* **Libraries and Frameworks:**
  + **Machine Learning:** Scikit-learn, TensorFlow.
  + **Data Manipulation:** Python Pandas, NumPy.
  + **Visualization:** Matplotlib, Plotly, Tableau.

**Team Roles:**

* **Data Engineer:**
  + Collect and preprocess transactional, demographic, and behavioral data.
* **Machine Learning Engineer:**
  + Develop and optimize clustering models.
  + Evaluate model performance and refine features.
* **Visualization Specialists:**
  + Visualize distributions and customer behavior patterns.
* **Deliverables:**
* An unsupervised machine learning model for segmenting e-commerce customers.
* Interactive visualizations of customer segments, purchasing patterns, and demographics.
* Comprehensive project documentation, including methodology, results, and limitations.

**Feature Engineering Additions:**

* Aggregated transaction features (Total Amount, Average Amount, Transaction Count).
* Time-based features (Recency, Preferred Hour of transactions).
* Behavioral features (Fraud Rate, Decline Rate, Unique Payment Methods).
* Location-based features (Latitude and Longitude clustering).
* Merchant category and transaction status features (Completed, Failed, Pending rates).