Approach with GenAI Integration in Customer Segmentation and Personalized Marketing

Incorporating Generative AI (GenAI) into the customer segmentation and personalized marketing approach allows for dynamic content generation, such as email copywriting, taglines, and images, based on customer data and segmentation. We will use large language models (LLMs) like GPT for text generation (e.g., personalized marketing emails and taglines) and image generation models like Stable Diffusion for category-specific visuals. These models are key to creating more engaging and personalized marketing experiences.

1. Data Ingestion & Preprocessing

- **Step 1.1**: **Ingestion of Data**
    - The data from users_data.csv and transactions_data.csv is ingested into the system for preprocessing.
- **Step 1.2**: **Preprocessing of Transaction Data**
    - Transaction data is cleaned (e.g., removing dollar symbols, handling missing values), and key features like monthly_spend, avg_transaction_amount, and chip_transaction_pct are calculated.
- **Step 1.3**: **Feature Engineering**
    - Calculate and create additional features, such as monthly_spend, debt_to_income_ratio, discretionary_income, and credit_score_category, that describe a user's financial status.

2. Clustering & Segmentation

- **Step 2.1**: **Clustering for Customer Segmentation**
    - Use K-means clustering to categorize customers based on features such as transaction habits, monthly spend, and financial status (e.g., debt-to-income ratio).
- **Step 2.2**: **Optimal Cluster Determination**
    - Evaluate clustering performance using metrics like silhouette scores to select the optimal number of clusters. This step defines which customer segment the user belongs to (e.g., "Budget-Conscious Shoppers", "Luxury Spenders", "Medium Range Spender").

3. GenAI Integration for Content Generation

- **Step 3.1**: **Email Content Generation (LLM-Based)**
    - **Model**: Use a Generative AI model (e.g., GPT-based model like GPT-3, GPT-Neo) to generate personalized marketing email content based on the customer’s profile and cluster.
    - **Process**:
        - Once a customer’s segment is identified, a prompt is created based on their financial profile and cluster category.
        - The prompt is passed to the LLM to generate personalized email content, which includes a subject line, greeting, and body content specific to their category (e.g., emphasizing savings for Budget-Conscious customers).
- **Step 3.2**: **Tagline Generation (LLM-Based)**
    - **Model**: A generative AI model (e.g., GPT) can also be used to create a short, catchy tagline for the marketing campaign.
    - **Process**:
        - A prompt is generated based on the customer’s segment (e.g., “Generate a tagline for Budget-Conscious Shoppers”) and passed to the model.
        - The generated tagline will be concise and aligned with the marketing goal (e.g., "Maximize savings with every spend").
- **Step 3.3**: **Image Generation (GenAI for Visuals)**

- **Model**: Use an image generation model like Stable Diffusion to create category-based visual assets (e.g., "Budget-Conscious Shopper" visual, "Luxury Spender" visual).
- **Process**:
  - A prompt based on the segment is passed to the image generation model (e.g., “Generate a visual for Luxury Spenders”).
  - The model generates a visual representation that can be used in the marketing campaign (e.g., a luxury-themed image for high-spending customers).

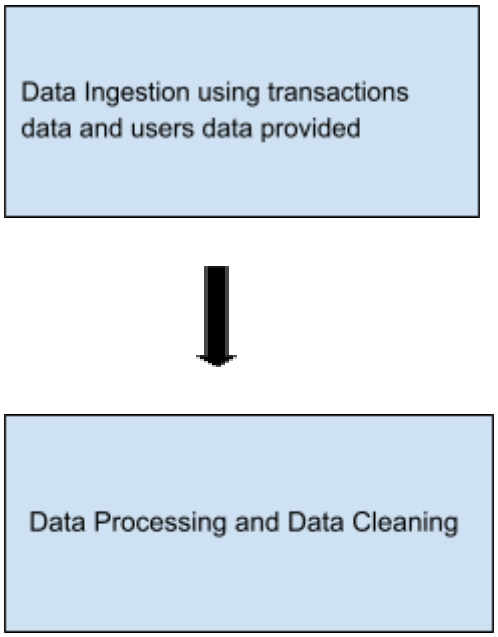
4. Result Presentation & Output

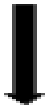
- **Step 4.1**: **Presentation of Marketing Content**
  - The system generates the following for each customer:
    - **Personalized Marketing Email**: Generated using the LLM based on the customer's segment and profile.
    - **Tagline**: A brief, attractive tagline generated using the LLM.
    - **Category-Based Image**: Visuals created using Stable Diffusion that correspond to the customer’s behavior and segment.
- **Step 4.2**: **Integration & Automation**
  - The generated content (email, tagline, and image) can be sent directly to the customers using python CLI.
  - The final result can be generated in the python command prompt window.

5. GenAI Models Used

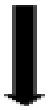
- **LLM Models (Text Generation)**:
  - **GPT-Neo** or **GPT-3** (for generating email content and taglines based on customer profile and segmentation).
- **Image Generation Models**:
  - **Stable Diffusion** (for generating category-specific images based on customer segmentation).

System Architecture

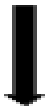




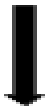
Feature Engineering(On transaction data, chip data, monthly spend, average transaction etc.



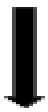
Merging Transactions data and Users data into the final_data dataframe



Clustering the final_data into three different clusters with the help of KMeans Algorithm



Text Generation for email content generation and tagline generation (Have used GPT3 NEO open source model for this purpose)



Based on the 3 cluster category Budget Conscious Shoppers, Luxury Spenders, Medium Range Spenders generate text to image with the help of Stable Diffusion Model

Prompt Engineering Strategy

1. **Text Generation:**

   For each customer profile, carefully crafted prompts are passed to the text-generation model to generate:

   - Offers aligned with cluster behavior.
   - Personalized marketing emails, ensuring the tone and content match the customer's profile.

2. **Taglines:**

   Short prompts are used to generate concise, impactful taglines reflecting the core value of the profile.

3. **Visuals:**

   Prompts for Stable Diffusion specify category-related visual descriptions to generate relevant marketing images.

Assumptions:

1. Data is clean and complete after preprocessing (e.g., no outliers or major inconsistencies).
2. Customers behave consistently over time; recent data reflects current behavior.
3. The text-generation model's output is free from inaccuracies and irrelevant suggestions.

Limitations:

1. **Data Limitations**: Missing or incomplete data might lead to biased clusters or suboptimal feature engineering.
2. **Cluster Interpretability**: K-Means assumes spherical clusters, which may not capture complex patterns in customer behavior.
3. **Model Constraints**: The GPT-Neo model and Stable Diffusion might occasionally generate less relevant content or visuals.
4. **Static Profiles**: Profiles do not account for dynamic changes in customer behavior over time.

Potential Improvements:

1. **Dynamic Clustering**: Implement time-series clustering or use a hierarchical approach for more nuanced segmentation.
2. **Model Tuning**: Fine-tune the text-generation and diffusion models for industry-specific applications.
3. **Feature Expansion**: Introduce more behavioral and contextual features, such as geographic data or time-based transaction patterns.
4. **User Feedback Loop**: Incorporate user engagement metrics to continuously refine clusters and personalize offers.
5. **System Scalability**: Optimize the pipeline for faster processing and real-time marketing campaign generation.