Customer Segmentation

You are given a dataset containing data about customer behavior in an e-commerce platform. The dataset contains information about customer interactions, purchases, and browsing patterns. Your task is to identify distinct customer segments (clusters) based on their behavior.

The dataset contains 6 features:

- **customer_id** : Unique id for the customer.
- **total_purchases** : Total number of purchases made by the customer.
- **avg_cart_value** : Average value of items in the customer's cart.
- **total_time_spent** : Total time spent on the platform (in minutes).
- **product_click** : Number of products viewed by the customer.
- **discount_count** : Number of times the customer used a discount code.

The dataset has 3 hidden clusters, each representing a distinct customer segment:

1. Bargain Hunters
2. High Spenders
3. Window Shoppers

1. Bargain Hunters

- **total_purchases**: High (frequent purchases).
- **avg_cart_value** : Low (they buy cheaper items).
- **total_time_spent** : Moderate (they spend some time browsing but focus on purchasing).
- **product_click** : Moderate (they view a reasonable number of products).
- **discount_count** : High (they frequently use discount codes).

Behavior These customers are deal-seekers who make frequent purchases of low-value items and heavily rely on discounts.

2. High Spenders

- **total_purchases**: Moderate (they make fewer but high-value purchases).
- **avg_cart_value**: High (they buy expensive items).
- **time_spent**: Moderate (they spend time browsing but focus on high-value items).
- **product_click** : Moderate (they view a reasonable number of products).
- **discount_usage**: Low (they rarely use discount codes).

Behavior: These customers are premium buyers who focus on high-value purchases and are less influenced by discounts.

3. Window Shoppers

- **total_purchases**: Low (they make very few purchases).
- **avg_cart_value**: Moderate (they view items of varying prices).
- **time_spent**: High (they spend a lot of time browsing).
- **product_click** : High (they view a large number of products).
- **discount_usage**: Low (they rarely use discount codes).

Your goal is to identify and clearly visualize these cluster separations.

Click here to download dataset here

Marking:

EDA - 40%
Model Selection - 20%
Model Evaluation - 10 %
Identifying Clusters - 10%
Code - 20%

Participants Should Deliver

1. Source JupyterNotebook Script
2. Report (do not exceed 15 pages except cover)