Assignment: Customer Segmentation using K-Means Clustering

Objective:

Segment wholesale customers based on their annual spending habits to identify distinct customer groups and propose targeted business strategies for each segment.

📂 Dataset Overview:

Dataset: Wholesale customers data
Attributes:

- **Channel** (1=Horeca, 2=Retail)
- **Region** (1=Lisbon, 2=Oporto, 3=Other)
- **Fresh** – Annual spending on fresh products
- **Milk**
- **Grocery**
- **Frozen**
- **Detergents_Paper**
- **Delicatessen**

Assignment Tasks:

- **Part 1: Data Preprocessing and EDA(10 visuals)**

  1. Load the dataset using `pandas`.
  2. Drop or retain the `Channel` and `Region` columns depending on your approach.
  3. Explore the distribution and relationships of key features (Fresh, Milk, Grocery, etc.).
  4. Analyze spending patterns by Channel and Region.
  5. Scale the features using `StandardScaler`.

- **Part 2: Determine the Optimal Number of Clusters**

  1. Apply **KMeans clustering** for values of K from 2 to 10.
  2. For each K:
     - Calculate **inertia**
     - Calculate **silhouette score**
  3. Plot:
     - Elbow curve (K vs Inertia)
     - Silhouette score plot (K vs Silhouette Score)
  4. Use both metrics to **choose the optimal number of clusters**.

- **Part 3: Apply KMeans with Optimal K**

  1. Fit a KMeans model using the chosen K.
  2. Assign each customer to a cluster.
  3. Add the cluster label to the original dataset.

🔹 **Part 4: Cluster Profiling**

1. Calculate the **mean annual spending** per cluster across product categories.
2. Interpret what kind of customers belong to each cluster.

🔹 **Part 5: Business Application – Targeted Strategy**

1. Based on each cluster's behavior, propose a **targeted business strategy**.
    - Examples: Offer discounts on frozen goods, increase grocery upselling, or reduce churn for high-value clients.
2. Add a new column `"Suggested Strategy"` to the DataFrame based on the cluster.

Deliverables:

- Python script or Jupyter Notebook with:
    - Clean and well-commented code
    - Plots and metrics for Elbow Method and Silhouette Score
    - Cluster summaries and proposed strategies
    -
- PowerPoint report (.pptx) summarizing EDA (min. 10 slides)

Submit here: https://forms.gle/MRPvsRZPjdgotnWh8