📌Project Title:

E-commerce Customers Segmentation

📌Dataset Overview:

You will be working with a dataset consisting of five interrelated tables, each containing crucial information about customers, transactions, branches, and merchants. The tables are described as follows:

1. **Customers Table:**
   - *customer_id*: Unique identifier for each customer.
   - *join_date*: The date the customer joined.
   - *city_id*: The ID representing the customer's city.
   - *gender_id*: The ID representing the customer's gender.
2. **Genders Table:**
   - *gender_id*: Unique identifier for each gender.
   - *gender_name*: Name of the gender (e.g., male, female).
3. **Cities Table:**
   - *city_id*: Unique identifier for each city.
   - *city_name*: Name of the city.
4. **Transactions Table:**
   - *transaction_id*: Unique identifier for each coupon transaction.
   - *customer_id*: ID of the customer who performed the transaction.
   - *transaction_date*: The date the coupon was claimed.
   - *transaction_status*: Status of the coupon (e.g., claimed, burnt).

- *coupon_name*: The name of the coupon.
- *burn_date*: The date the coupon was burnt.
- *branch_id*: ID of the branch where the coupon was burnt.

5. **Branches Table:**
   - *branch_id*: Unique identifier for each branch.
   - *merchant_id*: ID of the merchant who owns the branch.
6. **Merchants Table:**
   - *merchant_id*: Unique identifier for each merchant.
   - *merchant_name*: Name of the merchant.

📌Project Requirements:

#1: Build a Dashboard for Stakeholders

You are required to design and develop a **well-structured and informative dashboard** that presents key insights from the dataset. The dashboard should be designed for stakeholders, such as business managers and marketing teams, and should help them make data-driven decisions.

The dashboard should include, but is not limited to, the following visualizations:

- **Customer Demographics:** A breakdown of customers by gender and city.
- **Coupon Usage:** Insights into coupon transaction status (e.g., claimed vs. burnt coupons) over time.
- **Top-performing Cities/Branches:** Cities or branches with the highest number of successful coupon burns.
- **Customer Retention and Loyalty:** Patterns in customer coupon usage frequency and trends over time.

Recommended visualizations include:

- *Pie charts*: for customer gender and city distributions.
- *Time-series line charts*: for coupon transactions over time.
- *Heatmaps or bar charts*: for city-wise or branch-wise coupon burn rates.
- *Histograms*: for customer retention and burn frequency patterns.

Submission Requirements:

☐ Submit the dashboard as a **PDF file** with all key insights clearly displayed.

#2: Customer Segmentation Using Unsupervised Learning

The second part of the task requires you to develop and train an unsupervised machine learning model to **segment customers** based on their transactional behavior and other relevant features from the dataset. The goal is to identify customer groups that share similar behaviors and use these segments to discuss strategies for offering coupons to increase loyalty and satisfaction.

Key instructions:

1. **Feature Selection:** Utilize customer demographic features (e.g., gender, city) and transactional features (e.g., coupon usage frequency, transaction status) for segmentation.
2. **Model Development:**
   - Train and evaluate an **unsupervised machine learning model** (such as K-Means, DBSCAN, or hierarchical clustering) for customer segmentation.
   - Ensure to explore different numbers of segments (clusters) and optimize your model for meaningful customer groups.
3. **Model Evaluation:** Use appropriate evaluation metrics for unsupervised learning, such as:
   - Silhouette score.
   - Inertia (for K-Means).
   - … etc.
4. **Segment Analysis:** Analyze and describe each segment based on customer behavior, and make recommendations on which segments should be given coupons to maximize their loyalty and satisfaction with the store.

Submission Requirements:

- ☐ Submit your code in a **GitHub repository**.
- ☐ The repository should include:
  - ☐ A well-documented **README file** explaining the project, model approach, evaluation metrics, and key findings.
  - ☐ A **Jupyter Notebook** with clean and clear code, properly structured with comments and explanations.
- ☐ Include visualizations of the customer segments and a discussion of the results.

📌Submission Deadline:

The deadline for submission is *Thursday, 19th September at 11:59 PM.* Please submit your PDF File & GitHub repository link through the classroom platform by this time.