**Kafka Streaming Learning Project**

**Purpose: The purpose of this project is to build a real working Kafka streaming application to gain better understanding of Kafka Streams.**

**Project Name:** E-commerce Sales Analysis

**Problem Statement:** An e-commerce platform wants to analyze its sales data in real-time to understand sales trends and improve customer engagement.

**Data Source:** The e-commerce platform's sales data stream, which contains orders and customers with the following fields:

* **Order ID**: Unique identifier for each order
* **Timestamp**: Timestamp for when the order was placed
* **Product ID**: Unique identifier for the product ordered
* **Customer ID**: Unique identifier for the customer who placed the order
* **Quantity**: Number of units ordered
* **Price**: Price per unit
* **Customer Gender**: Gender of the customer
* **Customer Age**: Age of the customer

**Project Requirements:**

1. **Real-time processing:** The system should be able to process sales data in real-time and perform analysis on the data as it arrives.
2. **Joining and aggregation:** The system should join sales data with customer data based on the **Customer ID** field and perform real-time aggregation to calculate sales metrics (e.g., total sales, average sales price, sales by product category, sales by customer segment).
3. **Output to different topics:** The system should output the aggregated data to different Kafka topics based on the type of analysis performed (e.g., total sales by product category to one topic, sales by customer segment to another topic).
4. **Microservices architecture:** The system should be designed using a microservices architecture, with each microservice responsible for a specific task (e.g., data ingestion, data enrichment, real-time aggregation, output to different topics).

**Project Outline:**

1. **Data ingestion:** Set up a Kafka producer to ingest sales data from the e-commerce platform's sales data stream and publish it to a Kafka topic.
2. **Data enrichment:** Set up a Kafka streaming application to join sales data with customer data in real-time based on the **Customer ID** field. Use a lookup table to enrich the sales data with customer data.
3. **Real-time aggregation:** Set up a Kafka streaming application to perform real-time aggregation on the enriched data to calculate sales metrics (e.g., total sales, average sales price, sales by product category, sales by customer segment). Use Kafka Streams' aggregation feature to perform real-time aggregation.
4. **Output to different topics:** Output the aggregated data to different Kafka topics based on the type of analysis performed (e.g., total sales by product category to one topic, sales by customer segment to another topic).
5. **Microservices architecture:** Implement the above components as separate microservices using Docker and Kubernetes, with each microservice responsible for a specific task.