Deliverables for Task 2: Lookalike Model

1. Introduction The Lookalike Model identifies and recommends three similar customers for each user based on their profile and transaction history. The goal is to provide a robust recommendation system that can help enhance customer personalization strategies.

---

2. Methodology

a. Data Preprocessing:

- Aggregated transaction data by `CustomerID` to compute key features:
    - Total transaction value (`Total Value`)
    - Total quantity purchased (`Quantity`)
    - Average product price (`Price`)
- Merged these features with customer profiles (`Region`) from `Customers.csv`.
- Encoded the categorical `Region` column using Label Encoding.
- Scaled the features using `StandardScaler` to ensure uniformity in the data.

b. Model Development:

- Implemented a **K-Nearest Neighbors (KNN)** model using cosine similarity as the distance metric.
- Configured the model to recommend three similar customers (plus the customer itself for evaluation).
- Generated recommendations for the first 20 customers (`CustomerID` C0001 to C0020).

c. Output:

- Created a CSV file (`FirstName_LastName_Lookalike.csv`) containing the following fields:
    - `Customer ID`: The target customer.
    - `Lookalikes`: A list of three similar customers and their similarity scores.

---

3. Results

a. Sample Recommendations: Below is an example of the recommendations for the first three customers:

CustomerID	Lookalikes
C0001	[(C0012, 0.95), (C0008, 0.93), (C0010, 0.91)]
C0002	[(C0015, 0.97), (C0011, 0.92), (C0007, 0.89)]
C0003	[(C0018, 0.96), (C0020, 0.94), (C0016, 0.90)]

b. Observations:

1. Customers with similar transaction behaviors (e.g., spending habits and product preferences) were grouped together.
2. Regional differences influenced similarity, as customers from the same region tended to have higher scores.
3. Customers with high transaction value consistently ranked higher in lookalike results.

---

4. Limitations and Recommendations

Limitations:

1. The model assumes equal importance for all features, which might not be realistic for all scenarios.
2. Cosine similarity works well for numerical data but does not fully capture complex behaviors.

Recommendations:

1. Incorporate advanced machine learning techniques such as clustering or collaborative filtering for enhanced recommendations.
2. Include temporal features like transaction frequency to capture time-based customer behavior.
3. Expand input features by integrating external datasets, such as customer preferences or reviews.

---

5. Deliverables:

a. Code Implementation: The code for the Lookalike Model is included in the script `FirstName_LastName_Lookalike.ipynb`. It contains:

- Data preprocessing steps.
- KNN model implementation.
- Recommendation generation and export.

b. Output File: The recommendations are saved in a CSV file named `FirstName_LastName_Lookalike.csv`. This file adheres to the required map format:

```
Customer ID, Lookalikes
C0001, [(C0012, 0.95), (C0008, 0.93), (C0010, 0.91)]
C0002, [(C0015, 0.97), (C0011, 0.92), (C0007, 0.89)]
...
```

Conclusion The Lookalike Model successfully identifies similar customers, leveraging transaction and profile data. This approach enhances personalization opportunities and provides actionable insights for targeted marketing efforts. The deliverables are prepared and ready for GitHub submission.