

Task 2: Lookalike Model

1. Load and Prepare Data:

- Read customer, product, and transaction datasets.
- Extract relevant customer attributes (e.g., signup year) and aggregate transaction data (e.g., total spending, product purchase history).

2. Feature Engineering:

- Create a combined profile for each customer using their region, purchase history (concatenated product IDs), and signup year.

3. Vectorize Profiles:

- Use **TF-IDF Vectorizer** to convert the text-based customer profiles into numerical feature vectors for similarity analysis.

4. Calculate Similarities:

- Compute **cosine similarity** scores between all customers, measuring how closely each customer resembles the others.

5. Generate Recommendations:

- For the first 20 customers, find the top 3 most similar customers (excluding themselves) based on similarity scores.
- Store these lookalikes along with their similarity scores.

6. Save Results:

- Format the output as a DataFrame and save it to a CSV file (Vineet_Pundpal_Lookalike.csv) for future use or analysis.