The US Soccer commercial team invites innovative minds to participate in our hackathon challenge, where you’ll design a cutting-edge product recommendation engine for our dynamic e-commerce store. We operate an agile digital storefront where new collections and distinctive merchandise are routinely rolled out to cater to passionate fans. With new product launches happening regularly and a high volume of single-time purchasers, our goal is to convert casual shoppers into loyal, repeat customers. Your task is to build a model that not only anticipates customer needs but also adapts to our ever-evolving catalog.

Challenge Details

**Objective**

Develop a product recommendation model we can use to select products placed in customized emails to fans. The email template has three product pictures, so your model should generate three product suggestions for each customer. The challenge is evaluated on the test set that has been withheld from the training set provided: if any one of the three recommendations matches the customer’s actual next purchase, it counts as a “correct” prediction.

**Datasets**

* Products: Products from our US Soccer store. This dataset includes metadata on all historical and active products ever sold on our store.
* Product\_tags: Tags that are manually assigned to each product as they are set up on our store. The tags act as descriptors to help describe the product in a standardized way.
* Customers: Metadata on customers who have made previous purchases
* Line\_items: The order and item information for all historical sales from the US Soccer Store

**Challenges**

* *Dynamic Inventory:* With continuous launches of new merchandise, your model must adapt to frequent changes in product availability and relevance.
* *Customer Conversion:*A significant portion of our customer base are one-time buyers. The model should help identify patterns that encourage repeat purchasing by understanding customer preferences and predicting future buying behavior despite limited historical repeat data.
* *Data Quality:* This is a sample of a *real* dataset! This means there is a possibility of dirty data within the dataset that will need to be dealt with in whatever way you find most effective. Fix the data issues, when necessary, in the most logical way possible, just as we do in the real world!

**How to Submit your Model for Testing**

Create a CSV titled YOUR\_NAME\_commercial\_submission.csv formatted with the following columns:

* customer\_id
* product\_id\_1
* product\_id\_2
* product\_id\_3

You may include other columns as well if you like, however these four columns are required for assessment. Please find an example CSV in the S3 bucket titled “gtu\_hack\_commercial\_sample\_output.csv”

**Evaluation Criteria**

* Accuracy: The percentage of the test set where the last purchased product is within your model’s 3 predictions.
* Innovation: Creative approaches and novel use of data to improve recommendations in a dynamic retail environment.
* Clarity & Documentation: Code readability, well-documented processes, and clear presentation/understanding of your approach.

**Additional Notes**

When building product recommendation models, three primary approaches come into play:

*Collaborative Filtering*: This method leverages the purchasing patterns and behaviors of users. By analyzing the collective interactions of customers, it identifies similar users and suggests products based on what peers have bought. This modeling type, however, can struggle with “cold start” issues for new users or products.

*Content-Based Filtering*: Instead of relying on user interactions, this approach focuses on the intrinsic features of products—such as category, style, and price. It matches these attributes with customer preferences to generate recommendations. While effective when rich product metadata is available, relying solely on this method can sometimes restrict the diversity of recommendations.

*Hybrid Models*: These models combine collaborative and content-based techniques to balance their strengths and mitigate weaknesses. Hybrid approaches often yield more robust and adaptive recommendations, making them particularly valuable in dynamic environments with frequent new product launches.

This challenge invites you to experiment with these methodologies to create a recommendation engine that excels in a fast-paced, evolving e-commerce setting. Good luck!