**Commerce Data Generator**

**Overview**

The Commerce Data Generator is a Python script designed to create synthetic transaction data for a commerce dataset. It leverages the Faker library to generate realistic customer names and dates, producing a CSV file containing transaction information. This README provides an overview of the script's functionality and guidance on how to customize the generated data.

**Usage**

Prerequisites

Ensure you have Python installed on your system.

Install the required libraries using the following command:

bash pip install faker pandas



**Running the Script**

To run the script and generate synthetic transaction data:

1. Open a terminal or command prompt.

2. Navigate to the directory containing the script.

3. Run the script:

Bash python commerce\_data\_generator.py



4. The script will generate a CSV file named "CommerceTable.csv" containing the synthetic transaction data.

**Customization Options**

The script offers several variables that can be adjusted to customize the generated data:

**1. Number of Rows (`num\_rows`)**

The `num\_rows` variable controls the total number of rows or transactions to be generated. Update this variable to specify the desired size of your dataset.

**2. Number of Transactions per Customer (`min\_transactions` and `max\_transactions`)**

Adjust the `min\_transactions` and `max\_transactions` variables to define the range for the number of transactions per customer. This introduces variability in the number of transactions each customer can have.

**3. Gender Distribution (`female\_weight`, `male\_weight`)**

The `female\_weight` variable determines the percentage of female customers in the generated dataset. By default, it is set to 62%. Modify this value to achieve the desired gender distribution.

**4. Countries (`country`)**

Currently, the script is configured to generate transactions for the "United States." If you want transactions in different countries, you can modify the `country` variable within the script.

**5. Additional Fields**

The script includes various other fields such as `Product\_ID`, `Category\_ID`, `Brand\_ID`, etc. You can explore the script and update these fields based on your specific requirements.

Feel free to experiment with these variables to generate datasets tailored to your specific needs.

**Example:**

Update the number of rows to generate 5000 transactions

num\_rows = 5000



Update the range for the number of transactions per customer

min\_transactions = 1



max\_transactions = 5



Adjust the female weight to 50%

female\_weight = 0.5



Change the target country to "Canada"

country = "Canada"



By customizing these variables, you can generate diverse datasets suitable for testing, analysis, or any other purposes  
  
This Script was was put together by Steve Puma. For any additional questions you can reach out directly to him at [stephen.poff@sisense.com](mailto:stephen.poff@sisense.com) or slack Steve Puma.