

# Retail Transaction Dataset<sup>1</sup>

## A. Giới thiệu về tập dữ liệu

Bộ CSDL này gồm 1.000.000 records, ghi nhận các giao dịch mua hàng tại 10 thành phố diễn ra từ ngày 1/1/2020 đến 18/5/2024.

## B. Các tính năng chính của tập dữ liệu

- (i).- **Transaction\_ID**: (int) mã số của giao dịch. Được đánh số từ 1.000.000.001.
- (ii).- **Date**: (datetime) ngày giao dịch diễn ra, ghi nhận từ ngày 1/1/2020 đến 18/5/2024.
- (iii).- **Customer\_Name**: (string) thông tin về khách hàng.
- (iv).- **Product**: (string) Danh sách tên các sản phẩm được mua trong mỗi giao dịch.
- (v).- **Total Items**: (int) Thông tin về số lượng mặt hàng của mỗi giao dịch. Giá trị nằm trong khoảng từ 1 đến 10.
- (vi).- **Total Cost**: (float) Thông tin về tổng chi phí của mỗi giao dịch. Giá trị nằm trong khoảng từ 5 đến 100.
- (vii).- **Payment Information**: (float) Phương thức thanh toán được sử dụng cho mỗi lần mua hàng. Gồm 4 hình thức: *Cash, Credit Card, Debit Card, Mobile Payment*.
- (viii).- **City**: (string) Thành phố nơi giao dịch diễn ra. Gồm 10 thành phố: *Atlanta, Boston, Chicago, Dallas, Houston, Los Angeles, Miami, New York, San Francisco, Seattle*.
- (ix).- **Store Type**: (string) loại cửa hàng nơi giao dịch diễn ra. Gồm 6 dạng cửa hàng: *Convenience Store, Department Store, Pharmacy, Specialty Store, Supermarket, Warehouse Club*.
- (x).- **Discounts**: (boolean) Cho biết giao dịch có được hưởng việc giảm giá hay không?.
- (xi).- **Customer\_Categories**: (string) thể hiện danh mục khách hàng dựa trên nền tảng hoặc nhóm tuổi. Gồm 8 giá trị: *Homemaker, Middle-Aged, Professional, Retiree, Senior Citizen, Student, Teenager, Young Adult*.
- (xii).- **Season**: (string) giao dịch diễn ra trong mùa nào trong năm. Gồm 4 mùa: *Fall, Spring, Summer, Winter*.
- (xiii).- **Promotions**: (string) Các dạng khuyến mãi áp dụng cho các giao dịch. Gồm 3 giá trị: *BOGO (Buy One Get One), Discount on Selected Items, None*.

## C. Yêu cầu thực hiện

- I.- Field Product hiện bao gồm tên nhiều sản phẩm. Yêu cầu SV thực hiện lần lượt như sau:
  - Tách mỗi sản phẩm thành 1 field với kiểu dữ liệu của các cột mới tạo là boolean.
  - Gộp tất cả các field vừa có này trở lại thành 1 field kiểu số (làm khóa ngoại). Trong quá trình thực hiện, có thể cần tạo mới 1 table để lưu lại danh mục các hàng hóa được mua kèm nhau và field kiểu số làm khóa chính.

---

<sup>1</sup> [Retail Transactions](#)