# Yêu cầu với BTL NHẬP MÔN KHAI PHÁ DỮ LIỆU VÀ MÁY HỌC

1. Tên đề tài có dạng  
   “**Ứng dụng khai phá dữ liệu và máy học tích hợp vào hệ thống <tên hệ thống> [giải quyết bài toán <tên bài toán>]**”
2. Sử dụng dữ liệu có nguồn gốc rõ ràng, tin cậy, đủ lớn và có liên quan đến bài toán định xử lý; được cung cấp sẵn trên các cộng đồng DMML hoặc SV tự thu thập.
3. Áp dụng Khai phá luật kết hợp (1) hoặc Phân lớp (2) hoặc Phân cụm (3) để xây dựng giải pháp ứng dụng cho bài toán cụ thể
4. Mô hình được lựa chọn sau huấn luyện (/khai phá) phải căn cứ trên kết quả đánh giá tin cậy, chính xác và có căn cứ khoa học.
5. Mô hình cần được tích hợp vào hệ thống cụ thể, ở chức năng liên quan tiêu biểu để chứng minh tính ứng dụng - giải quyết bài toán đã xác định (các chức năng khác không phải là mục tiêu của học phần này)
6. Đáp ứng đúng mục tiêu theo đề cương của học phần NHẬP MÔN KHAI PHÁ DỮ LIỆU VÀ MÁY HỌC
7. Từng SV cần thực hiện nhiệm vụ để chứng minh đạt được Chuẩn đầu ra của học phần.

## 

***Báo cáo BTL học phần này cần trình bày theo cấu trúc dưới đây (SV soạn thảo***

***trực tiếp vào file này), tuân thủ Quy định trình bày của khoa***

# 

# Tên đề tài: Ứng dụng khai phá dữ liệu và máy học tích hợp vào “hệ thống dự đoán nhu cầu sản phẩm của khách hàng tại các cửa hàng thời trang bán lẻ” giải quyết bài toán “dự đoán nhu cầu sản phẩm trong tương lai nhằm tối ưu hóa tồn kho và điều chỉnh chiến lược kinh doanh hiệu quả trong năm tiếp theo”.

## Chương 1. Tổng quan về đề tài

### Giới thiệu bài toán (nêu vấn đề)

Bài toán dự đoán nhu cầu sản phẩm trong tương lai trong một cửa hàng thời trang bán lẻ trực tuyến là một yếu tố quan trọng đối với quản lý tồn kho, kế hoạch sản xuất và vận chuyển, cũng như chiến lược tiếp thị và quảng cáo của doanh nghiệp.

Vấn đề chính mà bài toán này đặt ra là làm thế nào để dự đoán một cách chính xác và hiệu quả nhu cầu của khách hàng đối với các sản phẩm trong tương lai, từ đó giúp cửa hàng tối ưu hóa tồn kho, cải thiện trải nghiệm mua sắm và tăng doanh số bán hàng. Điều này đòi hỏi sự kết hợp giữa dữ liệu lịch sử bán hàng, thông tin về sản phẩm và các yếu tố môi trường để đưa ra dự đoán chính xác về nhu cầu sản phẩm.

### Đánh giá đề tài liên quan (vấn đề hiện đã được giải quyết như thế nào, hiệu quả đến đâu)

**Dự đoán xu hướng mua hàng:** Bài toán này tập trung vào việc dự đoán xu hướng mua hàng của khách hàng dựa trên dữ liệu lịch sử bán hàng và các yếu tố bên ngoài như mùa vụ, sự kiện đặc biệt, thời tiết. Các phương pháp máy học như học sâu và học tăng cường đã được áp dụng để dự đoán xu hướng mua hàng với độ chính xác cao.

**Nguồn:**[SuNT's Blog | AI in Practical (tiensu.github.io)](https://tiensu.github.io/blog/109_ml_in_demand_prediction/)

**Tối ưu hóa tồn kho:** Bài toán này tập trung vào việc quản lý tồn kho sao cho phù hợp với nhu cầu thực tế của khách hàng và giảm thiểu chi phí lưu kho. Các mô hình dự đoán nhu cầu sản phẩm có thể giúp cải thiện quản lý tồn kho bằng cách dự đoán mức độ cần thiết của từng sản phẩm trong tương lai.

**Nguồn:**[4 Mô hình quản lý hàng tồn kho EOQ, POQ, PDM, ABC (vti-solutions.vn)](https://vti-solutions.vn/mo-hinh-quan-ly-hang-ton-kho/)

**Quản lý chuỗi cung ứng:** Bài toán này liên quan đến việc dự đoán nhu cầu của cả chuỗi cung ứng từ nguồn cung đến điểm bán hàng. Dự đoán chính xác nhu cầu sản phẩm giúp cải thiện quá trình lập kế hoạch sản xuất và vận chuyển trong chuỗi cung ứng.

**Nguồn:**[Mô hình Just In Time: Vai trò, Ứng dụng và Quy trình triển khai (atalink.com)](https://vietnam.atalink.com/blog/tim-hieu-ve-mo-hinh-just-in-time/)

### Mục đích đề tài

(đã sửa lần 1)

**Dự đoán nhu cầu về sản phẩm trong tương lai** của khách hàng tại các cửa hàng thời trang bán lẻ nhằm tối ưu hóa tồn kho và điều chỉnh chiến lược kinh doanh hiệu quả trong năm tiếp theo, giúp cửa hàng thời trang bán lẻ hoạt động một cách hiệu quả, linh hoạt và có thể vững vàng hơn trong môi trường kinh doanh ngày càng cạnh tranh.

### Mục tiêu đề tài

(làm lại, bổ sung mục tiêu)

Phát triển và triển khai một hệ thống dự đoán nhu cầu sản phẩm trong tương lai cho một cửa hàng thời trang bán lẻ, bằng cách tích hợp các phương pháp và công nghệ khai phá dữ liệu và máy học. Mục tiêu bao gồm:

* **Áp dụng các kỹ thuật và mô hình máy học** để phân tích và dự đoán nhu cầu sản phẩm trong tương lai (của cái gì?)
* **Tích hợp hệ thống dự đoán vào quy trình quản lý tồn kho** kế hoạch sản xuất và vận chuyển của cửa hàng thời trang bán lẻ, nhằm tối ưu hóa tồn kho, cải thiện trải nghiệm mua sắm và tăng doanh số bán hàng. (chỉnh sửa thêm cho phù hợp với mục đích của bài toán)

### Phân công và Kế hoạch thực hiện

| Nội dung (Khoảng thời gian thực hiện) | Công việc chi tiết | Thời gian hoàn thành | Chu Quỳnh Anh | Lê Văn Đạt | Nguyễn Thị Thu Hằng |
| --- | --- | --- | --- | --- | --- |
| Chương 1  (2/4 - 4/4) | - Giới thiệu bài toán  - Đánh giá đề tài liên quan | 3/4/2024 |  | **x** |  |
|  | - Mục đích đề tài | 4/4/2024 | **x** |  |  |
|  | - Mục tiêu đề tài | 3/4/2024 |  |  | **x** |
|  | - Phân công và lên kế hoạch | 3/4/2024 | **x** | **x** | **x** |
| Chương 2  (10/4 - 11/4) | - Tìm Dataset | 10/4/2024 | **x** |  |  |
|  | - Thực hiện phân tích dữ liệu | 11/4/2024 |  |  | **x** |
|  | - Tiến hành các bước tiền xử lý dữ liệu | 11/4/2024 |  | **x** |  |
|  | - Biểu diễn trực quan hóa dữ liệu | 11/4/2024 | **x** |  |  |
|  | - Làm slide báo cáo | 11/4/2024 | **x** |  |  |
|  | - Thuyết trình | 11/4/2024 | **x** | **x** | **x** |

#### 

## Tiền xử lý dữ liệu

Thực hiện phân tích về dữ liệu, tiến hành các bước tiền xử lý dữ liệu, chuẩn bị dữ liệu cho giai đoạn sau.

### Dataset đã chọn

(đã thêm nguồn gốc data)

Tập dữ liệu được lựa chọn gồm 3000 bản ghi lịch sử mua hàng của khách hàng tại một cửa hàng thời trang bán lẻ trong năm 2022.

Nguồn thu thập dataset: <https://www.kaggle.com/datasets/nishchay331/retail-store/data>

Nguồn lưu trữ dataset được sử dụng trong đề tài:

<https://drive.google.com/file/d/1tvu5qsEP4cHSQW_s6uuT-Sc4Z3k5-uNH/view?usp=drive_link>

### Thực hiện phân tích về dữ liệu:

(phân tích và chọn ra những đặc trưng sẽ sử dụng trong bài toán của mình, kết quả phân tích, tính khả tín, -> mô tả bước làm) + trực quan hóa dữ liệu

#### Chuẩn bị

##### Khai báo thư viện



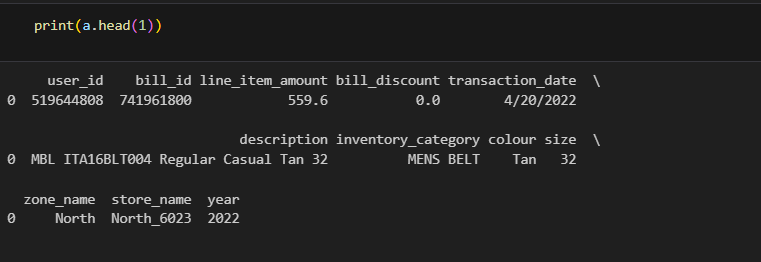
##### Lấy dữ liệu

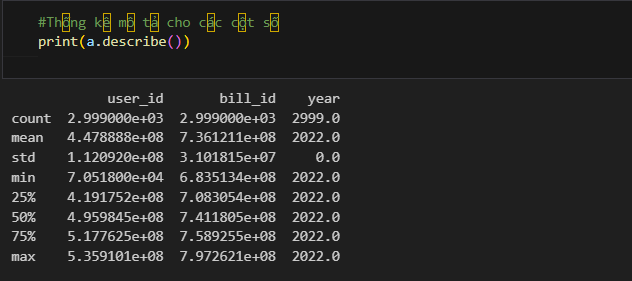


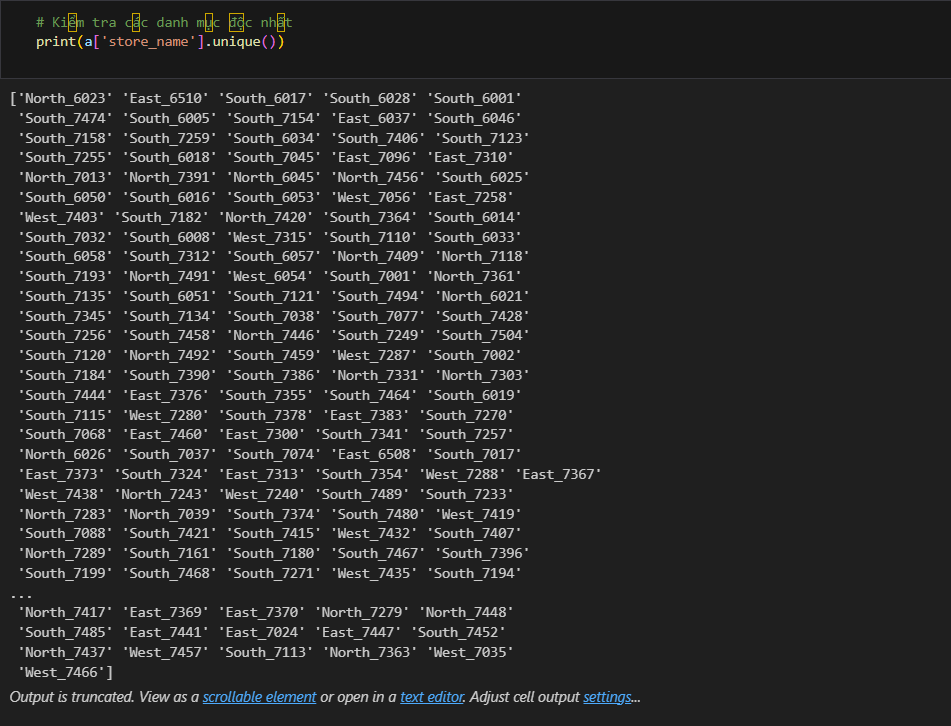
##### Đọc dữ liệu

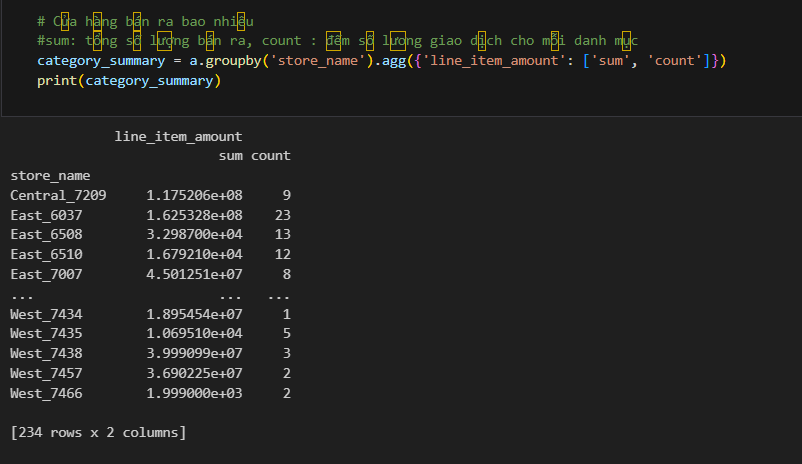


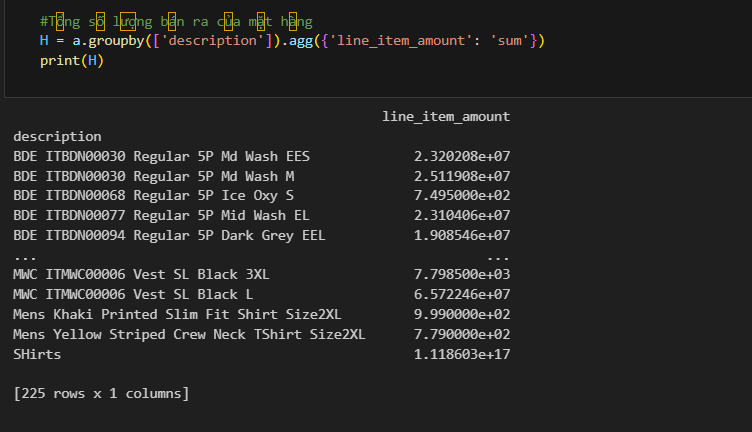
##### Phân tích dữ liệu

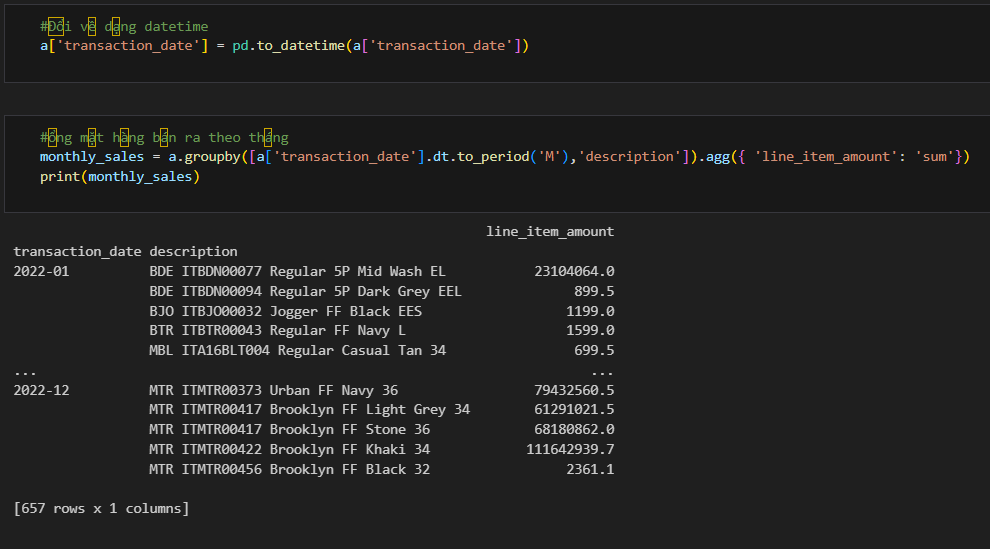












### Tiến hành các bước tiền xử lý dữ liệu:

#### Phân tích dữ liệu

● Dữ liệu thiếu

○ Có 51 Sản phẩm thiếu danh mục(inventory\_category)

–chỉ ra minh chứng nhận định dữ liệu lỗi, chỉ ra dữ liệu bị lỗi, rồi mới thực hiện xử lý dữ liệu lỗi

#### Các bước xử lý dữ liệu (xử lý trên Google Sheet)

TẠO BỘ LỌC CHO TOÀN BỘ DỮ LIỆU

##### Bước 1: Xử lý dữ liệu trùng lặp

1.1. Chọn toàn bộ bảng dữ liệu

1.2. Trong tab "Data" (Dữ liệu), chọn "Remove Duplicates" (Loại bỏ các dòng trùng lặp).

=> từ 2999 dòng còn 2207 dòng loại bỏ 792 bản ghi trùng nhau

##### Bước 2: Xử lý dữ liệu thiếu

2.1. Cột inventory\_category

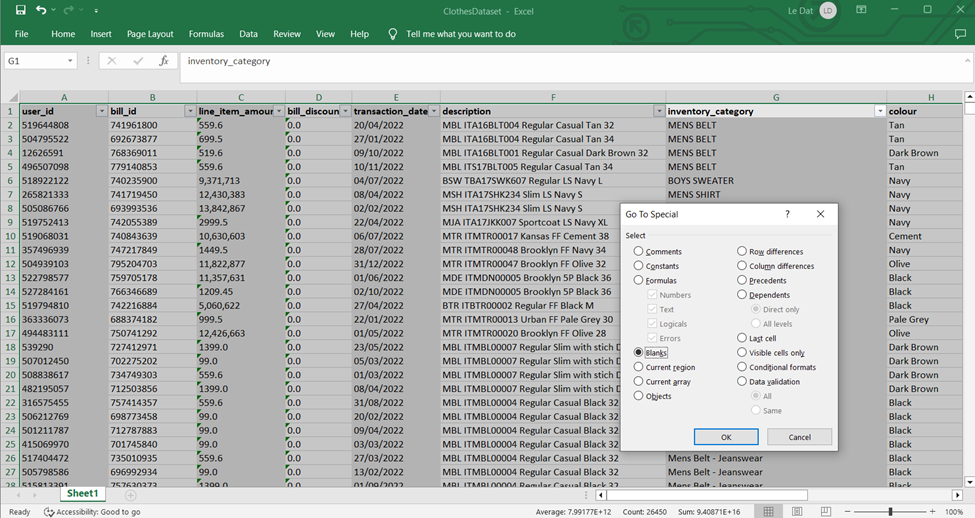
Bước 1: Chọn toàn bộ dữ liệu

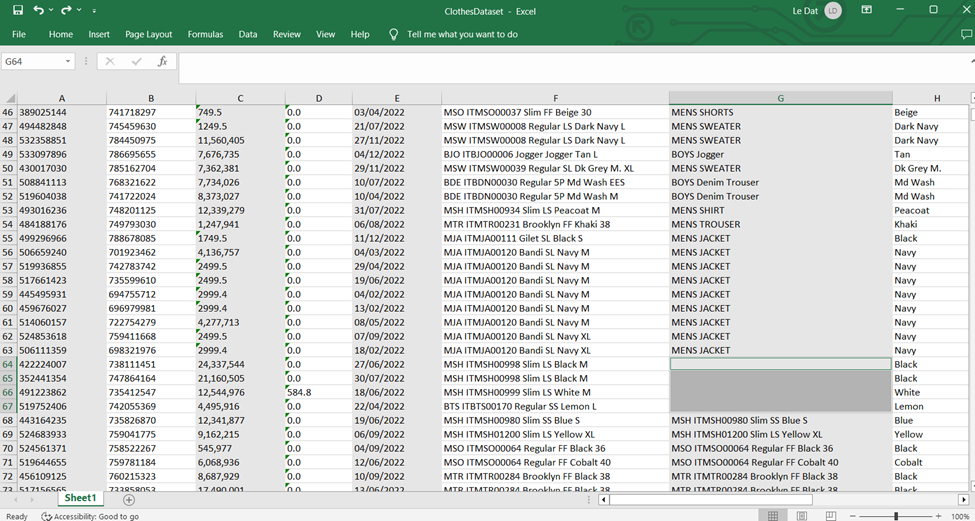
Bước 2: Nhấn F5, hộp thoại Go To xuất hiện.

Bước 3: Nhấn nút Special… trong hộp thoại.

Bước 4: Trong hộp thoại Go to special, chọn nút Blanks và nhấn OK.

Bước 5: Nhấp phải vào bất kỳ ô được chọn vào và chọn Delete.

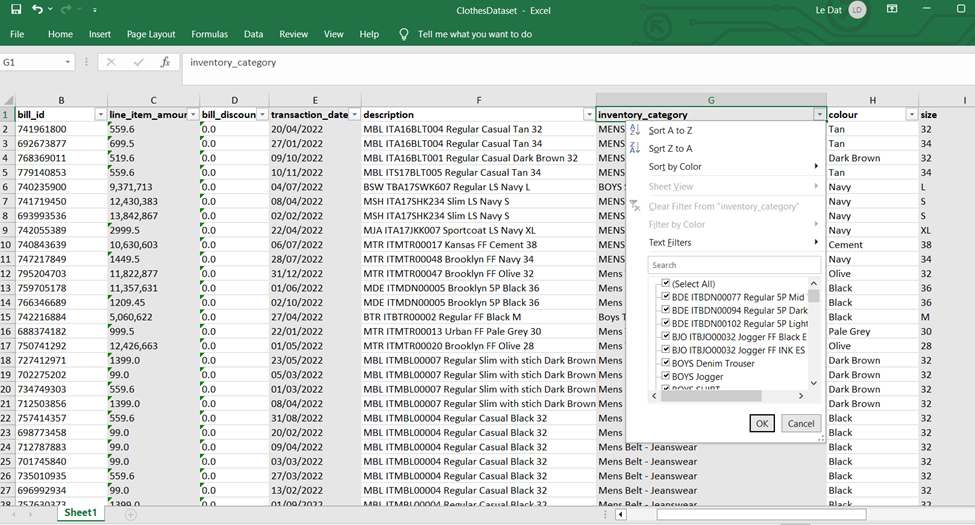




=> Còn lại 2161 bản ghi

##### Bước 3: Xử lý dữ liệu nhiễu

Trong cột inventory\_category bấm vào bộ lọc tìm thấy các bản ghi có dữ liệu nhiễu (tên trùng với cột description).



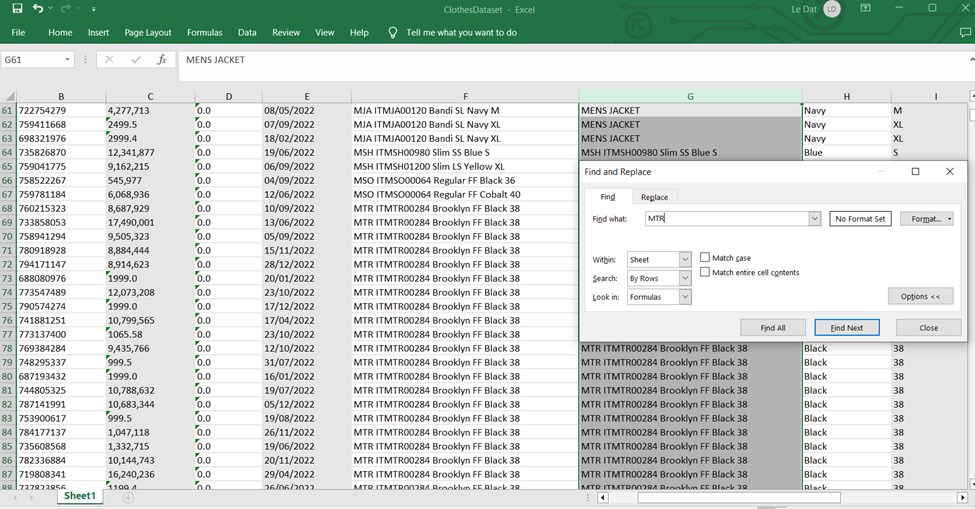
Tiến hành dùng công cụ tìm kiếm và thay thế các bản ghi bị nhiễu dựa trên dữ liệu đúng từ các bản ghi khác ví dụ cột description là BTR ITBTR00002 Regular FF Black M thì cột inventory\_category là Boys Trouser – Sportswear.

Bước 1: chọn toàn bộ cột inventory\_category.

Bước 2: Ctrl H.

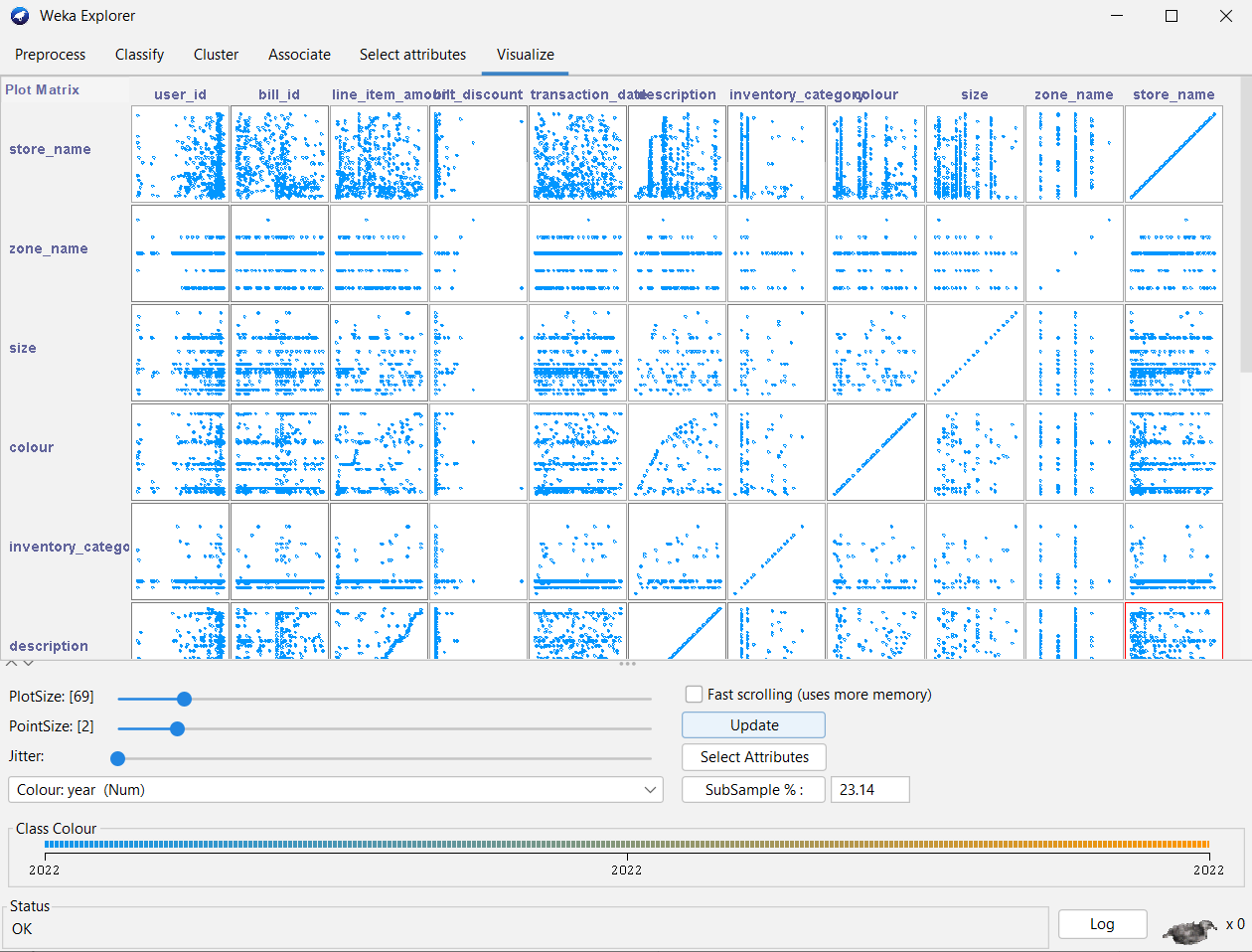
Bước 3: tìm kiếm với ten cột description Chọn Find All.

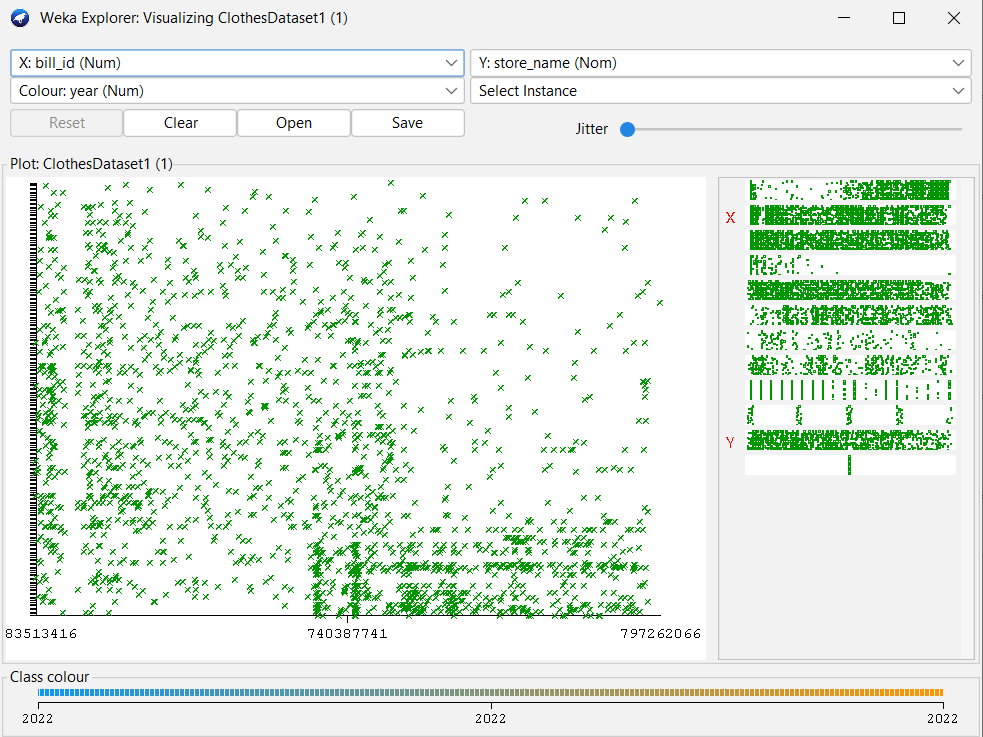
Bước 4: chọn toàn bộ và thay thế.



### Trực quan hóa dữ liệu:

### (thêm nhận định khi visualize lần 2 sau khi thực hiện tiền xử lý dữ liệu )





## Chương 3. Xây dựng mô hình

Lựa chọn hướng giải quyết, lựa chọn thuật toán, công nghệ sử dụng cần căn cứ theo tính chất của dữ liệu và mục tiêu của đề tài, kết hợp với việc tham khảo hướng giải quyết từ các đề tài liên quan.

## Chương 4. Tích hợp mô hình vào hệ thống

Mô hình cần được tích hợp vào chức năng cụ thể, thỏa mãn yêu cầu của đề bài

## Kết luận

Kết luận theo từng mục tiêu của đề tài

## Tài liệu tham khảo

**Phụ lục**