CHƯƠNG 3: TIỀN XỬ LÝ DỮ LIỆU VÀ CHỌN LỌC FIELD

3.1 Các kỹ thuật tiền xử lí dữ liệu

3.2 Chọn lọc và chuẩn hóa field

Trong quá trình bạn tải lên từng dataset và tôi hỗ trợ bạn tổng hợp các dữ liệu lại, chúng ta đã áp dụng một số phép biến đổi tiền xử lý dữ liệu. Dưới đây là tổng hợp các bước và phép biến đổi mà chúng ta đã thực hiện:

Chuẩn hóa Tên Cột:

Đổi tên các cột để đồng nhất giữa các tập dữ liệu, giúp việc hợp nhất dữ liệu sau này trở nên dễ dàng hơn. Ví dụ, đổi tên cột "title" thành "car\_name", "Price" thành "price" v.v.

Chuyển Đổi Đơn Vị Tiền Tệ:

Chuyển đổi giá từ GBP và Lakh sang USD để thống nhất đơn vị tiền tệ trong tất cả các dataset. Tỷ giá đã được sử dụng là 1 GBP = 1.24 USD và 1 Lakh = 1199 USD.

Xử Lý Dữ Liệu Động Cơ (Engine):

Chuyển đổi thông tin động cơ từ "L" sang "cc" (xilanh) và đảm bảo tất cả các giá trị đều thể hiện bằng đơn vị "cc". Phép chuyển đổi này cũng bao gồm việc xử lý các chuỗi và chuyển đổi giá trị số.

Xóa Cột Không Cần Thiết:

Xóa các cột không cần thiết như "Unnamed: 0" để làm sạch và giảm kích thước của dataset.

Hợp Nhất Dữ Liệu:

Sau khi chuẩn hóa và làm sạch, chúng ta đã hợp nhất ba tập dữ liệu vào một DataFrame duy nhất dựa trên các trường thông tin chung.

Làm Sạch Dữ Liệu Người Sở Hữu:

Chuẩn hóa các giá trị trong cột liên quan đến thông tin sở hữu xe từ chuỗi sang số, như chuyển "First", "Second", v.v. thành 1, 2, và sử dụng các giá trị ngẫu nhiên cho các dữ liệu bị thiếu hoặc không phù hợp.

3.3 Áp dụng nội suy để xử lí dữ liệu thiếu

CHƯƠNG 4: PHÂN TÍCH THỐNG KÊ VÀ XÁC SUẤT

4.1 Cơ bản về xác suất và thống kê

4.2 Áp dụng Central Limit Theorem trong lấy mẫu và ước lượng