

# 4. OUTLIER

Mục tiêu OUTLIER: giúp xác định các điểm dữ liệu không phù hợp hoặc không bình thường so với phần còn lại của tập dữ liệu.

- [Thực hiện trên hệ thống theo các bước hướng dẫn sau:](#)

# Thực hiện trên hệ thống theo các bước hướng dẫn sau:

Bước 1: Vào hệ thống chọn module Explore, chọn tập dữ liệu cần tổng hợp và tạo mới một tập dữ liệu tổng hợp từ OUTLIER.

The screenshot shows the SQL Lab interface. On the left, the 'EXPLORE' button is highlighted in the sidebar. The main area displays a dataset named 'storagehub.hanh\_chinh.quanly\_t...'. The dataset is a table with columns: 'abc', 'cccc', 'dia\_diem\_tai\_nau', 'quan\_huyen', 'so\_nguoi\_bi\_thu', 'so\_nguoi\_chet', 'ten', and 'thoi\_gian\_xay'. The table contains 15 rows of data. On the right, the 'OUTLIER' button is highlighted in the 'MORE TOOLS' dropdown menu.

abc	cccc	dia_diem_tai_nau	quan_huyen	so_nguoi_bi_thu	so_nguoi_chet	ten	thoi_gian_xay
693156748023		Ngã tư Kim Liên - ...	Hai Bà Trưng	0	1	Trần Thị Bình	
918304582715		Giao lộ Lê Duẩn - ...	Hoàn Kiếm	0	1	Lê Văn Cường	Hà Nội
572639104820		Cầu Thanh Trì.	Hai Bà Trưng	0	1	Phạm Thị Dung	Hà Nội
239714865301		Đường Trần Phú - ...	Đống Đa	1	0	Hoàng Văn Đức	Hà Nội
820173495624		Cầu Long Biên.	Long Biên	1	0	Vũ Thị Hà	Hà Nội
461028573916		Đường Lê Trọng T...	Hoàn Kiếm	1	0	Đặng Văn Hiếu	Hà Nội
187492056318		Đoạn đường qua k...	Hoàn Kiếm	1	0	Bùi Thị Hoa	Hà Nội
574820196347		Giao lộ Nguyễn V...	Hoàn Kiếm	1	0	Đỗ Văn Hùng	Hà Nội
329165487201		Đường Nguyễn Tr...	Hai Bà Trưng	1	0	Ngô Thị Lan	Hà Nội
841693705286		Đường Phạm Hùn...	Hai Bà Trưng	1	0	Hồ Văn Long	Hà Nội
509728436162		Giao lộ Nguyễn Tr...	Hai Bà Trưng	1	0	Dương Thị Mai	Hà Nội
267490158374		Đường Phùng Hùn...	Hai Bà Trưng	1	0	Đào Văn Nam	Hà Nội
361985247092		Đoạn đường qua k...	Hoàn Kiếm	1	0	Lý Thị Ngọc	Hà Nội

Bước 2: Chọn cột dữ liệu cần xác định các điểm không phù hợp hoặc không bình thường so với phần còn lại của tập dữ liệu.

The screenshot shows the SQL Lab interface with the 'OUTLIER' tool selected. The 'SOURCE' dropdown is set to 'quanly\_tainangiaothong'. The 'PHÂN TÍCH OUTLIER' section is active, showing a list of columns to analyze: 'so\_nguoi\_chet' and 'so\_nguoi\_bi\_thuong'. The 'NHÓM GIÁ TRỊ' (Value Groups) section shows three groups: 'Thấp' (Low) with a value of 1,75, 'Bình Thường' (Normal) with a value of 2,05, and 'Cao' (High) with a value of 2,7. The 'KẾT QUẢ' (Results) section shows a table with columns: 'abc', 'cccc', 'dia\_diem\_tai\_nau', 'quan\_huyen', 'so\_nguoi\_bi\_thu', 'so\_nguoi\_chet', 'ten', 'thoi\_gian\_xay', and 'tinh'. The table contains 15 rows of data. The 'tinh' column shows the results of the outlier analysis for each row.

abc	cccc	dia_diem_tai_nau	quan_huyen	so_nguoi_bi_thu	so_nguoi_chet	ten	thoi_gian_xay	tinh
693156748023		Ngã tư Kim Liên - ...	Hai Bà Trưng	0	1	Trần Thị Bình		Hà Nội
918304582715		Giao lộ Lê Duẩn - ...	Hoàn Kiếm	0	1	Lê Văn Cường		Hà Nội
572639104820		Cầu Thanh Trì.	Hai Bà Trưng	0	1	Phạm Thị Dung		Hà Nội
239714865301		Đường Trần Phú - ...	Đống Đa	1	0	Hoàng Văn Đức		Hà Nội
820173495624		Cầu Long Biên.	Long Biên	1	0	Vũ Thị Hà		Hà Nội
461028573916		Đường Lê Trọng T...	Hoàn Kiếm	1	0	Đặng Văn Hiếu		Hà Nội
187492056318		Đoạn đường qua k...	Hoàn Kiếm	1	0	Bùi Thị Hoa		Hà Nội
574820196347		Giao lộ Nguyễn V...	Hoàn Kiếm	1	0	Đỗ Văn Hùng		Hà Nội
329165487201		Đường Nguyễn Tr...	Hai Bà Trưng	1	0	Ngô Thị Lan		Hà Nội
841693705286		Đường Phạm Hùn...	Hai Bà Trưng	1	0	Hồ Văn Long		Hà Nội

Bước 3: Chọn thuật toán để xử lý

SQL Lab > Tạo mới

Share

Nhân Hạnh Nguyễn

demo1@incloud.vn

Nhập tên ...

LƯU

SOURCE

quany\_tainangiaothong

PHÂN TÍCH OUTLIER

so\_nguoi\_bi\_thuong

THUẬT TOÁN

Z-Scores

1,75 2,05 2,7

Thấp là một số từ 1.0 đến 2.0  
Bình thường là một số từ 1.0 đến 3.0  
Cao là một số từ 2.5 đến 6.0

TEST

KẾT QUẢ

Tổng cộng có 100 dữ liệu được trả về. Thời gian truy vấn 00:00:00.02

Chọn thuộc tính

Nhập từ khóa tìm kiếm

1 - 20 / 100

1 / 5

abc	cccc	abc	dia_diem_tai_na	abc	quanhuyen	#	so_nguoi_bi_thu	#	so_nguoi_chet	abc	ten	thoi_gian_xay_r	abc	tinhtp
693156748023			Ngã tư Kim Liên - ...		Hai Bà Trưng	0		1			Trần Thị Bình			Hà Nội
918304582715			Giao lộ Lê Duẩn - ...		Hoàn Kiếm	0		1			Lê Văn Cường			Hà Nội
572639104820			Cầu Thanh Trì.		Hai Bà Trưng	0		1			Phạm Thị Dung			Hà Nội
239714865301			Đường Trần Phú - ...		Đống Đa	1		0			Hoàng Văn Đức			Hà Nội
820173495624			Cầu Long Biên.		Long Biên	1		0			Vũ Thị Hà			Hà Nội
461028573916			Đường Lê Trọng T...		Hoàn Kiếm	1		0			Đặng Văn Hiếu			Hà Nội
187492056318			Đoạn đường qua ...		Hoàn Kiếm	1		0			Bùi Thị Hoa			Hà Nội
574820196347			Giao lộ Nguyễn V...		Hoàn Kiếm	1		0			Đỗ Văn Hùng			Hà Nội
329165487201			Đường Nguyễn Tr...		Hai Bà Trưng	1		0			Ngô Thị Lan			Hà Nội
841693705286			Đường Phạm Hùn...		Hai Bà Trưng	1		0			Hồ Văn Long			Hà Nội
509728436162			Giao lộ Nguyễn Tr...		Hai Bà Trưng	1		0			Dương Thị Mai			Hà Nội

Trong đó:

- o Thuật toán Z-Scores được sử dụng để xác định các giá trị ngoại lệ trong tập dữ liệu.
- o Thuật toán IQR được sử dụng chủ yếu để phát hiện và xử lý các giá trị ngoại lệ trong tập dữ liệu

Bước 4: Điều chỉnh giá trị xác định, so sánh

SQL Lab > Tạo mới

Share

Nhân Hạnh Nguyễn

demo1@incloud.vn

Nhập tên ...

LƯU

SOURCE

quany\_tainangiaothong

PHÂN TÍCH OUTLIER

so\_nguoi\_bi\_thuong

THUẬT TOÁN

Z-Scores

NHÓM GIÁ TRỊ

Thấp Bình Thường Cao

1 2,05 2,7

Thấp là một số từ 1.0 đến 2.0  
Bình thường là một số từ 1.0 đến 3.0  
Cao là một số từ 2.5 đến 6.0

TEST

KẾT QUẢ

Tổng cộng có 100 dữ liệu được trả về. Thời gian truy vấn 00:00:00.02

Chọn thuộc tính

Nhập từ khóa tìm kiếm

1 - 20 / 100

1 / 5

abc	cccc	abc	dia_diem_tai_na	abc	quanhuyen	#	so_nguoi_bi_thu	#	so_nguoi_chet	abc	ten	thoi_gian_xay_r	abc	tinhtp
693156748023			Ngã tư Kim Liên - ...		Hai Bà Trưng	0		1			Trần Thị Bình			Hà Nội
918304582715			Giao lộ Lê Duẩn - ...		Hoàn Kiếm	0		1			Lê Văn Cường			Hà Nội
572639104820			Cầu Thanh Trì.		Hai Bà Trưng	0		1			Phạm Thị Dung			Hà Nội
239714865301			Đường Trần Phú - ...		Đống Đa	1		0			Hoàng Văn Đức			Hà Nội
820173495624			Cầu Long Biên.		Long Biên	1		0			Vũ Thị Hà			Hà Nội
461028573916			Đường Lê Trọng T...		Hoàn Kiếm	1		0			Đặng Văn Hiếu			Hà Nội
187492056318			Đoạn đường qua ...		Hoàn Kiếm	1		0			Bùi Thị Hoa			Hà Nội
574820196347			Giao lộ Nguyễn V...		Hoàn Kiếm	1		0			Đỗ Văn Hùng			Hà Nội
329165487201			Đường Nguyễn Tr...		Hai Bà Trưng	1		0			Ngô Thị Lan			Hà Nội
841693705286			Đường Phạm Hùn...		Hai Bà Trưng	1		0			Hồ Văn Long			Hà Nội
509728436162			Giao lộ Nguyễn Tr...		Hai Bà Trưng	1		0			Dương Thị Mai			Hà Nội

Bước 5: “RUN” chạy test dữ liệu được tổng hợp

SQL lab

Share

Nhân Hạnh Nguyễn

demo1@inetacloud.vn

SQL Lab > Tạo mới

Nhập tên ...

LƯU

SOURCE

quany\_tainangiaothong

PHÂN TÍCH OUTLIER

so\_nguoi\_bi\_thuong

THUẬT TOÁN

Z-Scores

NHÓM GIÁ TRỊ

Thấp Bình Thường Cao

1 2,05 2,7

Thấp là một số từ 1.0 đến 2.0

Bình thường là một số từ 1.0 đến 3.0

Cao là một số từ 2.5 đến 6.0

TEST

KẾT QUẢ

Tổng cộng có 17 dữ liệu được trả về. Thời gian truy vấn 00:00:00.09

abc	dia_diem_tai_nan	abc	quan_huyen	#	so_nguoi_bi_thu	abc	so_nguoi_bi_thuong_outlier_rank	#	so_nguoi_chet	abc	ten
1748023	Ngã tư Kim Liên - ...		Hai Bà Trưng	0			L2		1		Trần Thị Bình
1582715	Giao lộ Lê Duẩn - Đ...		Hoàn Kiếm	0			L2		1		Lê Văn Cường
1104820	Cầu Thanh Trì.		Hai Bà Trưng	0			L2		1		Phạm Thị Dung
1274801	Đường Nguyễn Xiế...		Hoàn Kiếm	0			L2		1		Trịnh Thị Quỳnh
1385021	Giao lộ Nguyễn Kh...		Hoàn Kiếm	0			L2		1		Lương Văn Sơn
1847562	Đường Hồ Tùng Mậ...		Hoàn Kiếm	0			L2		1		Mai Thị Thu
1946281	Đoạn đường qua k...		Đống Đa	0			L2		1		Lưu Văn Thanh
17631420	Đường Nguyễn Tuấ...		Đống Đa	0			L2		1		Đinh Thị Trang
1084573	Đường Lê Văn Lươ...		Đống Đa	0			L2		1		Đoàn Văn Trường
12301974	Đường Lê Văn Lươ...		Long Biên	0			L2		1		Dương Thị Hà
1093658	Đường Nguyễn Chí ...		Ba Đình	0			L2		1		Đào Văn Hải
1265710	Đoạn đường qua k...		Ba Đình	0			L2		1		Lý Thị Hạnh
1386297	Đường Hàm Nghi - ...		Đống Đa	0			L2		1		Châu Văn Hiệp

Bước 6: Nhập tên cho tập dữ liệu mới được tổng hợp, chọn Lưu thông tin tập dữ liệu được tổng hợp bằng công cụ OUTLIER

SQL lab

Share

Nhân Hạnh Nguyễn

demo1@inetacloud.vn

SQL Lab > Tạo mới

outlier\_tainangiaothong

LƯU

SOURCE

quany\_tainangiaothong

PHÂN TÍCH OUTLIER

so\_nguoi\_bi\_thuong

THUẬT TOÁN

Z-Scores

NHÓM GIÁ TRỊ

Thấp Bình Thường Cao

1 2,05 2,7

Thấp là một số từ 1.0 đến 2.0

Bình thường là một số từ 1.0 đến 3.0

Cao là một số từ 2.5 đến 6.0

TEST

KẾT QUẢ

Tổng cộng có 17 dữ liệu được trả về. Thời gian truy vấn 00:00:00.09

abc	dia_diem_tai_nan	abc	quan_huyen	#	so_nguoi_bi_thu	abc	so_nguoi_bi_thuong_outlier_rank	#	so_nguoi_chet	abc	ten
1748023	Ngã tư Kim Liên - ...		Hai Bà Trưng	0			L2		1		Trần Thị Bình
1582715	Giao lộ Lê Duẩn - Đ...		Hoàn Kiếm	0			L2		1		Lê Văn Cường
1104820	Cầu Thanh Trì.		Hai Bà Trưng	0			L2		1		Phạm Thị Dung
1274801	Đường Nguyễn Xiế...		Hoàn Kiếm	0			L2		1		Trịnh Thị Quỳnh
1385021	Giao lộ Nguyễn Kh...		Hoàn Kiếm	0			L2		1		Lương Văn Sơn
1847562	Đường Hồ Tùng Mậ...		Hoàn Kiếm	0			L2		1		Mai Thị Thu
1946281	Đoạn đường qua k...		Đống Đa	0			L2		1		Lưu Văn Thanh
17631420	Đường Nguyễn Tuấ...		Đống Đa	0			L2		1		Đinh Thị Trang
1084573	Đường Lê Văn Lươ...		Đống Đa	0			L2		1		Đoàn Văn Trường
12301974	Đường Lê Văn Lươ...		Long Biên	0			L2		1		Dương Thị Hà
1093658	Đường Nguyễn Chí ...		Ba Đình	0			L2		1		Đào Văn Hải
1265710	Đoạn đường qua k...		Ba Đình	0			L2		1		Lý Thị Hạnh
1386297	Đường Hàm Nghi - ...		Đống Đa	0			L2		1		Châu Văn Hiệp