

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 shown in a table view 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, a dropdown menu is open, showing options: 'IMPUTER', 'OUTLIER', 'XTABLE', and 'FORECAST'. The 'OUTLIER' option is highlighted.

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' analysis configuration. The 'SOURCE' is set to 'quanly_tainangiaothong'. The 'PHÂN TÍCH OUTLIER' section is highlighted, 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 range of 1,75 to 2,05, 'Bình Thường' (Normal) with a range of 2,05 to 2,7, and 'Cao' (High) with a range of 2,7 to 6,0. The 'KẾT QUẢ' (Results) section shows a table with 100 rows of data, including columns for 'abc', 'cccc', 'dia_diem_tai_nau', 'quan_huyen', 'so_nguoi_bi_thu', 'so_nguoi_chet', 'ten', and 'thoi_gian_xay'.

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

IQR

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

Table with 7 columns: id, dia_diem_tai_na, quan_huyen, so_nguoi_bi_thu, so_nguoi_chet, ten, thoi_gian_xay_ra, tinh_hp

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

Table with 7 columns: id, dia_diem_tai_na, quan_huyen, so_nguoi_bi_thu, so_nguoi_chet, ten, thoi_gian_xay_ra, tinh_hp

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_nanabc	quan_huyen	# so_nguoi_bi_thuabc	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_nanabc	quan_huyen	# so_nguoi_bi_thuabc	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