Fulbright
SCHOOL OF PUBLIC POLICY AND MANAGEMENT

Disruptive Technologies
for Public Asset Governance

THE WORLD BANK
IBRD • IDA | WORLD BANK GROUP

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Embassy of Switzerland
**Swiss Cooperation Office in Vietnam**

# Năng lực dữ liệu - Một số khái niệm và tình huống thực tế

**Huỳnh Nhật Nam**

*Trường Chính sách công và Quản lý Fulbright*
*TPHCM, 11/2025*

- Năng lực dữ liệu
- Đọc hiểu dữ liệu
- Khai thác dữ liệu
- Quản trị dữ liệu
- Thảo luận

# Năng lực dữ liệu

# Một vài định nghĩa

> … là năng lực **đọc, hiểu, tạo ra dữ liệu, và giao tiếp nội dung thông tin của dữ liệu.**

https://en.wikipedia.org/wiki/Data_literacy

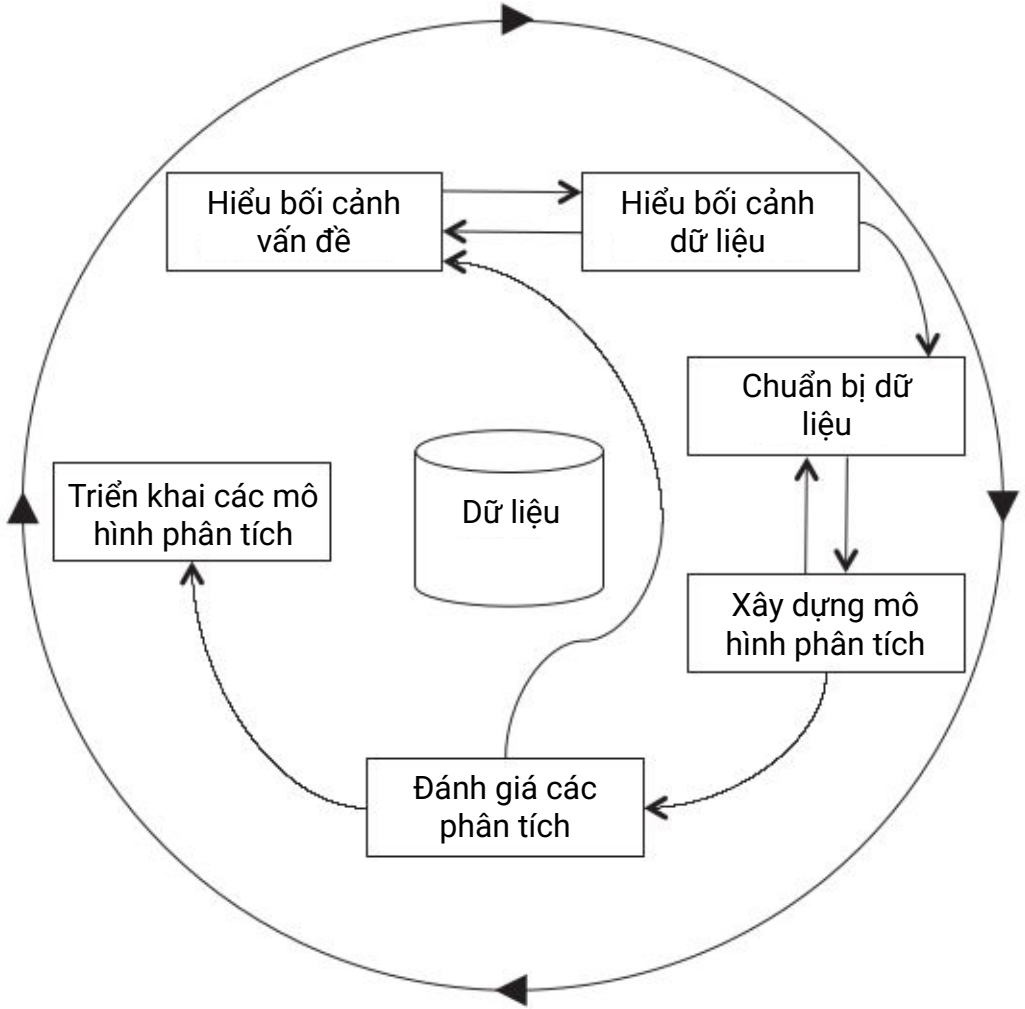
> … mô tả mức độ một cá nhân hoặc tổ chức **hiểu, làm việc, phân tích, hiển thị, và ứng dụng dữ liệu để đạt được các mục tiêu công việc**.

https://www.dataversity.net/what-is-data-literacy/

> … là năng lực **hiểu dữ liệu và các cách thực hành về dữ liệu** một cách đầy đủ để **diễn giải dữ liệu một cách có ý nghĩa và truyền đạt ý nghĩa đó một cách hiệu quả**.

https://www.nnlm.gov/guides/data-glossary/data-literacy#:~:text=Definition,and%20effectively%20communicate%20that%20meaning.

# Vòng đời phân tích khai phá dữ liệu CRISP-DM




Để **"đạt được mục tiêu công việc"** cần đặt dữ liệu trong bối cảnh của vấn đề cần giải quyết

- Một vấn đề được xác định rõ ràng đã được giải quyết một nửa.
- Hiểu vấn đề (và các giải pháp cũng như các tác động dự kiến) trước khi đi sâu vào các phân tích.
  - Vì sao vấn đề này quan trọng?
  - Vấn đề này tác động những ai?
  - Nếu chúng ta không có đủ dữ liệu (phù hợp) thì sẽ như thế nào?
  - Khi nào dự án kết thúc?
  - Nếu chúng ta không thích các kết quả phân tích thì sẽ như thế nào?

Provost & Fawcett, 2013, *Data Science for Business*, O'Reilly

Gutman & Goldmeier, 2021, Becoming a *Data Head*, Wiley



# Đọc hiểu dữ liệu

# Các bước chính

- Các đặc điểm tổng quan của tập dữ liệu

- Phân tích mô tả
  - Mô tả và hiển thị các giá trị đơn biến
  - Mô tả và hiển thị các giá trị của cặp biến

- Các giá trị bất thường
  - Phát hiện và xử lý các giá trị ngoại lệ và các giá trị *inliers*
  - Phát hiện và xử lý các giá trị bị mất

# Các đặc điểm tổng quan của tập dữ liệu

- **Dữ liệu phi cấu trúc**, ví dụ hình ảnh, chữ viết, mạng lưới, âm thanh.

- **Dữ liệu có cấu trúc**, ví dụ các bộ dữ liệu có dạng bảng như điều tra khảo sát và chuỗi thời gian.
  - số hàng dữ liệu và mỗi hàng dữ liệu đại diện thông tin gì, ví dụ một cá nhân, một hộ gia đình, một sự kiện ngập đường.
  - Số lượng và định dạng của các cột giá trị
    - *Giá trị định danh (nominal values).* Các giá trị loại trừ lẫn nhau, không có ý nghĩa về số lượng và thứ tự, ví dụ Nam và Nữ.
    - *Giá trị thứ tự (ordinal values).* Các giá trị có thứ tự lớn nhỏ, nhưng sự khác nhau giữa các giá trị không được định nghĩa, ví dụ các giá trị của thang đo Likert.
    - *Giá trị khoảng (interval values).* Các giá trị có thứ tự, sự khác nhau giữa các giá trị được định nghĩa, nhưng không có khái niệm 'giá trị không', ví dụ nhiệt độ, giờ (theo đồng hồ)
    - *Giá trị tỷ lệ (ratio values).* Tương tự giá trị khoảng nhưng có khái niệm "giá trị không", ví dụ thu nhập, độ sâu ngập.

# Phân tích mô tả

| Loại giá trị | Các phép toán | Đại lượng đo lường sự định tâm | Các đại lượng đo lường sự biến thiên |
|---|---|---|---|
| Định danh (nominal) | ● So sánh giống nhau (=, ≠) | ● Giá trị yếu vị | ● Không có |
| Thứ tự (ordinal) | ● So sánh giống nhau (=, ≠)<br>● So sánh lớn nhỏ (>, <) | ● Giá trị yếu vị<br>● Giá trị trung vị | ● Khoảng giá trị<br>● Khoảng tứ phân vị |
| Khoảng (interval) | ● So sánh giống nhau (=, ≠)<br>● So sánh lớn nhỏ (>, <)<br>● Cộng, trừ (+, -) | ● Giá trị yếu vị<br>● Giá trị trung vị<br>● Trung bình số học | ● Khoảng giá trị<br>● Khoảng tứ phân vị<br>● Độ lệch chuẩn<br>● Phương sai |
| Tỷ lệ (ratio) | ● So sánh giống nhau (=, ≠)<br>● So sánh lớn nhỏ (>, <)<br>● Cộng, trừ (+, -)<br>● Nhân, chia (x, ÷) | ● Giá trị yếu vị<br>● Giá trị trung vị<br>● Trung bình số học<br>● Trung bình hình học | ● Khoảng giá trị<br>● Khoảng tứ phân vị<br>● Độ lệch chuẩn<br>● Phương sai<br>● Độ lệch chuẩn tương đối |

Source:
https://www.scribbr.com/statistics/levels-of-measurement/

# Phân tích mô tả

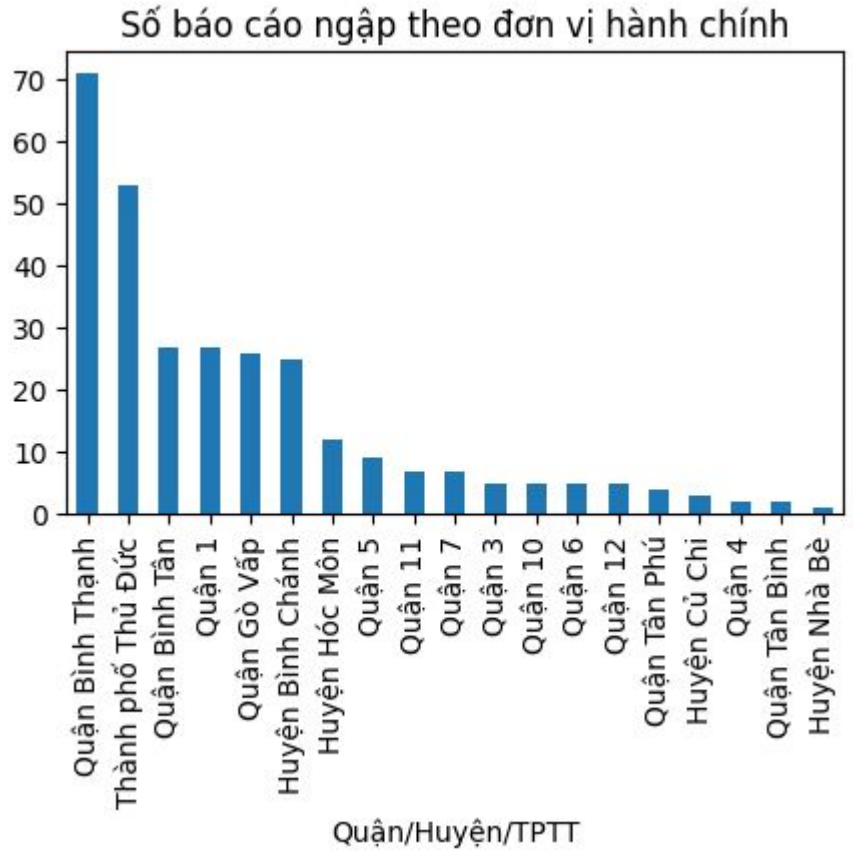
Dữ liệu mẫu từ nền tảng ‘Báo cáo ngập và hỗ trợ ra quyết định’ của TPHCM (FEDS)

| Số hiệu báo cáo | Loại hình học | Quận/Huyện/TP TT | Phường/Xã | Đường/Phố | Vị trí ngập đầu | Vị trí ngập cuối | Mức ngập | Nguyên nhân ngập | Thời điểm tạo báo cáo | Thời điểm bắt đầu ngập | Thời điểm kết thúc ngập | Thời lượng ngập (phút) | Thời gian tiêu thoát (phút) | Độ sâu ngập | Chiều dài ngập (m) | Chiều rộng ngập (m) | Diện tích ngập (m²) | Vũ lượng (mm) | Thời điểm bắt đầu mưa | Thời điểm kết thúc mưa |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2875 | Đường | Quận Gò | Phường 1 | Đường Ph | SN 520 | Lê Đức T | Chưa xác | Do mưa | 45511.80 | 45511.78 | 45511.81 | 45 | | 0.1 | 435.58 | 9 | 3920.22 | | | |
| 2866 | Đường | Huyện Bì | Thị Trấn | Đường Bi | 22 | 26 | Chưa xác | Do mưa | 45504.66 | 45504.60 | 45504.66 | 77 | | 0.2 | 117.77 | 10 | 1177.7 | | | |
| 1260 | Điểm | Huyện Bì | Thị Trấn | Đường Bi | 33 | 55 | Chưa xác | Do mưa | 45504.65 | 45503.64 | 45503.67 | 56 | | 1 | 100 | 20 | 2000 | | | |
| 2865 | Đường | Huyện Bì | Thị Trấn | Đường K | 22 | 56 | Chưa xác | Do mưa | 45503.70 | 45502.70 | 45502.71 | 19 | | 0.51 | 199.15 | 10 | 1991.5 | | | |
| 1258 | Điểm | Quận Bìn | Phường A | Đường Số | 48 | 128 | Chưa xác | Do mưa | 45503.62 | 45502.62 | 45502.70 | 120 | | 0.1 | 160 | 16 | 2560 | | | |
| 1257 | Điểm | Quận 4 | Phường 3 | Hẻm 256 | 1 | 5 | Chưa xác | Do mưa | 45503.62 | 45503.62 | 45503.62 | 3 | | 1 | 66 | 22 | 1452 | | | |
| 2864 | Đường | Quận Bìn | Phường A | Hẻm 532/ | 2 | 48 | Chưa xác | Do mưa | 45503.61 | 45491.37 | 45491.44 | 101 | | 0.1 | 121.55 | 8 | 972.4 | | | |
| 1256 | Điểm | Huyện Bì | Thị Trấn | Quốc Lộ | 12 | 123 | Chưa xác | Do mưa | 45503.60 | 45502.60 | 45502.64 | 62 | | 0.3 | 106.66 | 10 | 1066.6 | | | |
| 2863 | Đường | Thành phố | Phường A | Đường Tr | Sn300 | Sn500 | Chưa xác | Do mưa | 45503.59 | 45503.55 | 45503.58 | 43 | | 0.2 | 119.43 | 12 | 1433.16 | | | |
| 2862 | Đường | Thành phố | Phường A | Đường Hi | Kha Vạn | CA phườn | Chưa xác | Do mưa | 45503.59 | 45502.59 | 45502.63 | 62 | | 0.1 | 258.5 | 8 | 2068 | | | |
| 2861 | Đường | Thành phố | Phường A | Đường Cà | 100 | 50 | Chưa xác | Do mưa | 45503.59 | 45503.59 | 45503.60 | 12 | | 0.1 | 128.57 | 150 | 19285.5 | | | |
| 1255 | Điểm | Quận Bìn | Phường 1 | Đường Lé | 12 | 123 | Chưa xác | Do mưa | 45502.43 | 45495.43 | 45495.47 | 60 | | 0.5 | 106.63 | 10 | 1066.3 | | | |
| 1254 | Điểm | Quận 1 | Phường B | Cầu Thị N | 12a | 112b | Chưa xác | Do mưa | 45499.65 | 45498.59 | 45498.63 | 60 | | 0.5 | 100 | 10 | 1000 | | | |
| 1253 | Điểm | Huyện Bì | Thị Trấn | Đường Ng | 1 | 88 | Chưa xác | Do mưa | 45498.66 | 45496.54 | 45496.55 | 20 | | 0.15 | 106.66 | 20 | 2133.2 | | | |
| 1252 | Điểm | Huyện Bì | Thị Trấn | Đường Ng | 12 | 123 | Chưa xác | Do mưa | 45498.66 | 45496.54 | 45496.55 | 20 | | 0.1 | 100 | 20 | 2000 | | | |
| 1251 | Điểm | Quận 6 | Phường 1 | Đường Ly | 3 | 88 | Ngập tron | Do mưa | 45497.73 | 45496.52 | 45496.55 | 45 | 25 | 0.1 | 100 | 20 | 2000 | 1 | 45496.25 | 45496.54 |
| 1250 | Điểm | Quận 1 | Phường B | Đường Lé | 3 | 123 | Ngập tron | Do mưa | 45497.73 | 45496.52 | 45496.55 | 55 | 25 | 0.1 | 22 | 10 | 220 | 1 | 45496.25 | 45496.54 |
| 2860 | Đường | Quận Gò | Phường 1 | Đường Qu | 1 | 90 | Ngập vừa | Do mưa | 45497.73 | 45496.52 | 45496.56 | 70 | 40 | 0.14 | 1331.38 | 10 | 13313.8 | 1 | 45496.25 | 45496.54 |
| 2859 | Đường | Quận Bìn | Phường 1 | Đường Dị | 12 | 123 | Ngập vừa | Do mưa | 45497.72 | 45496.51 | 45496.56 | 75 | 40 | 0.15 | 1652.39 | 10 | 16523.9 | 1 | 45496.25 | 45496.54 |
| 1249 | Điểm | Thành phố | Phường A | Đường M | 2 | 123 | Chưa xác | Do mưa | 45497.72 | 45496.51 | 45496.56 | 69 | | 0.15 | 10 | 10 | 100 | | | |

**Lưu ý**. Các giá trị chỉ mang tính minh họa hỗ trợ bài giảng, có thể không là giá trị thực.

# Phân tích mô tả

Mô tả và hiển thị cột giá trị định danh (dữ liệu minh họa từ nền tảng FEDS)



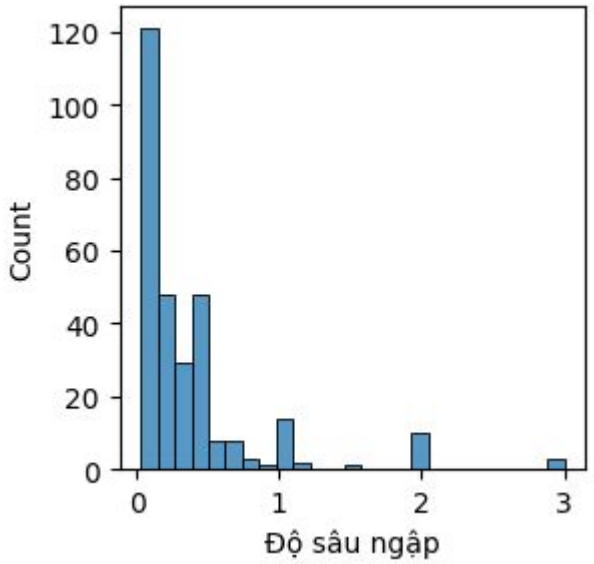

| Quận/Huyện/TPTT | |
|---|---|
| Quận Bình Thạnh | 71 |
| Thành phố Thủ Đức | 53 |
| Quận Bình Tân | 27 |
| Quận 1 | 27 |
| Quận Gò Vấp | 26 |
| Huyện Bình Chánh | 25 |
| Huyện Hóc Môn | 12 |
| Quận 5 | 9 |
| Quận 11 | 7 |
| Quận 7 | 7 |
| Quận 3 | 5 |
| Quận 10 | 5 |
| Quận 6 | 5 |
| Quận 12 | 5 |
| Quận Tân Phú | 4 |
| Huyện Củ Chi | 3 |
| Quận 4 | 2 |
| Quận Tân Bình | 2 |
| Huyện Nhà Bè | 1 |

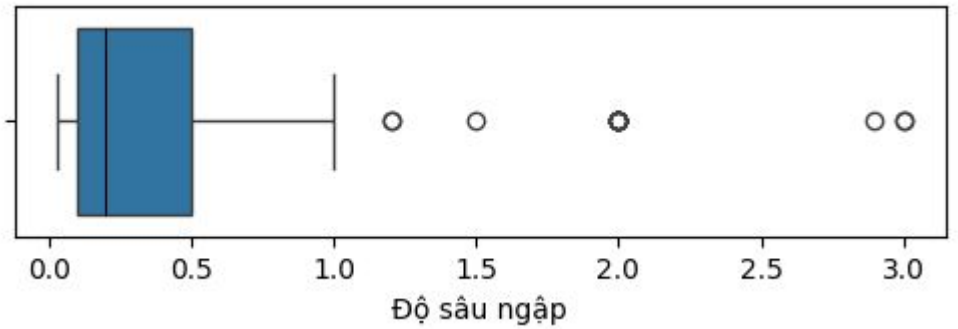
# Phân tích mô tả

Mô tả và hiển thị cột giá trị số (dữ liệu minh họa từ nền tảng FEDS)



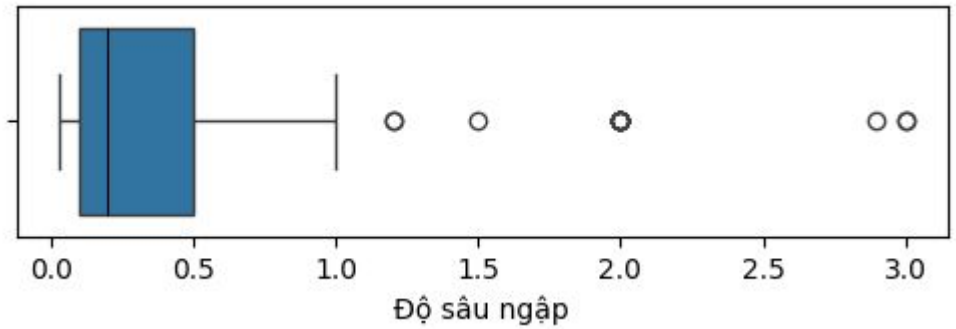

| | Độ sâu ngập |
|---|---|
| count | 296.000000 |
| mean | 0.378649 |
| std | 0.482299 |
| min | 0.030000 |
| 25% | 0.100000 |
| 50% | 0.200000 |
| 75% | 0.500000 |
| max | 3.000000 |




*Loại đồ thị nào thể hiện đầy đủ thông tin hơn về (các) cột dữ liệu đang được tìm hiểu?*

# Phân tích mô tả







Vui lòng giải thích boxplot trên cho một nhà quản lý. Trục ngang mô tả giá trị độ sâu ngập (đơn vị mét).

Phân tích

Boxplot này mô tả sự phân bố của dữ liệu về **Độ sâu ngập** và cung cấp một cái nhìn nhanh chóng, toàn diện về các chỉ số thống kê quan trọng.

Dưới đây là giải thích dành cho một nhà quản lý, tập trung vào các điểm chính:

**Tóm tắt Phân bố Độ sâu Ngập**

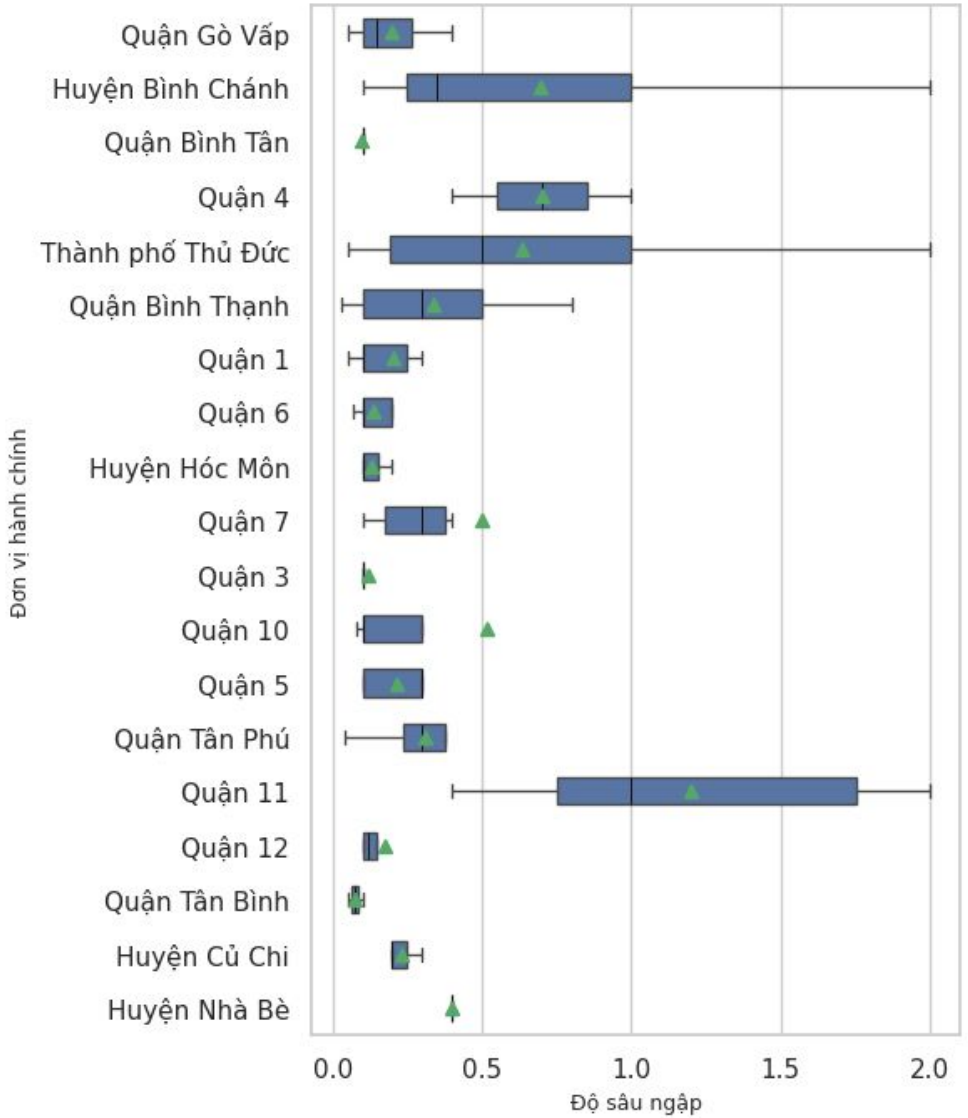
Boxplot (biểu đồ hộp) này cho thấy phần lớn dữ liệu về độ sâu ngập của chúng ta tập trung ở mức rất thấp, nhưng có một số ít trường hợp ngập sâu bất thường.

- **Giá trị Trung tâm (Median):** Đường kẻ dọc màu đen nằm bên trong hộp là **Trung vị (Median)**, cho thấy **50%** các trường hợp ngập có độ sâu **khoảng 0.2 mét**.
- **Phần lớn dữ liệu (Hộp):** Chiều dài của hộp (từ 0.0 mét đến khoảng 0.5 mét) biểu thị **Phạm vi Liên Tứ phân vị (IQR)**, nơi chứa **50%** các điểm dữ liệu ở giữa. Điều này cho thấy đa số các trường hợp ngập không vượt quá 0.5 mét.
  - **Tứ phân vị thứ nhất (Q1):** Khoảng 0.1 mét.
  - **Tứ phân vị thứ ba (Q3):** Khoảng 0.5 mét.
- **Phạm vi không có giá trị ngoại lai (Râu):** "Râu" (whiskers) của biểu đồ kéo dài đến khoảng **1.0 mét**. Đây là mức độ ngập sâu tối đa được coi là **"bình thường"** trong mô hình phân bố dữ liệu này.

# Phân tích mô tả

Hiển thị cặp giá trị định tính - định lượng (dữ liệu minh họa từ nền tảng FEDS)

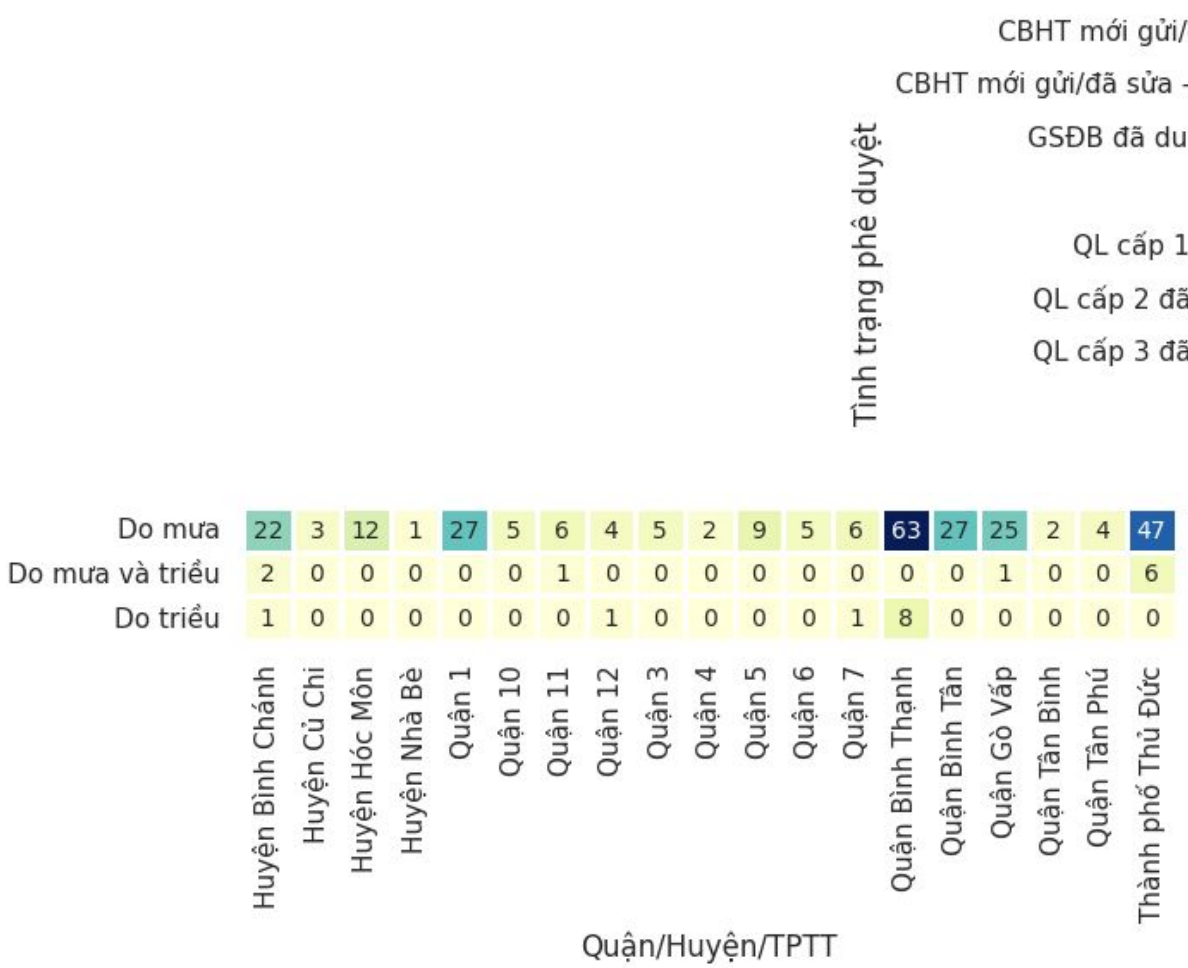


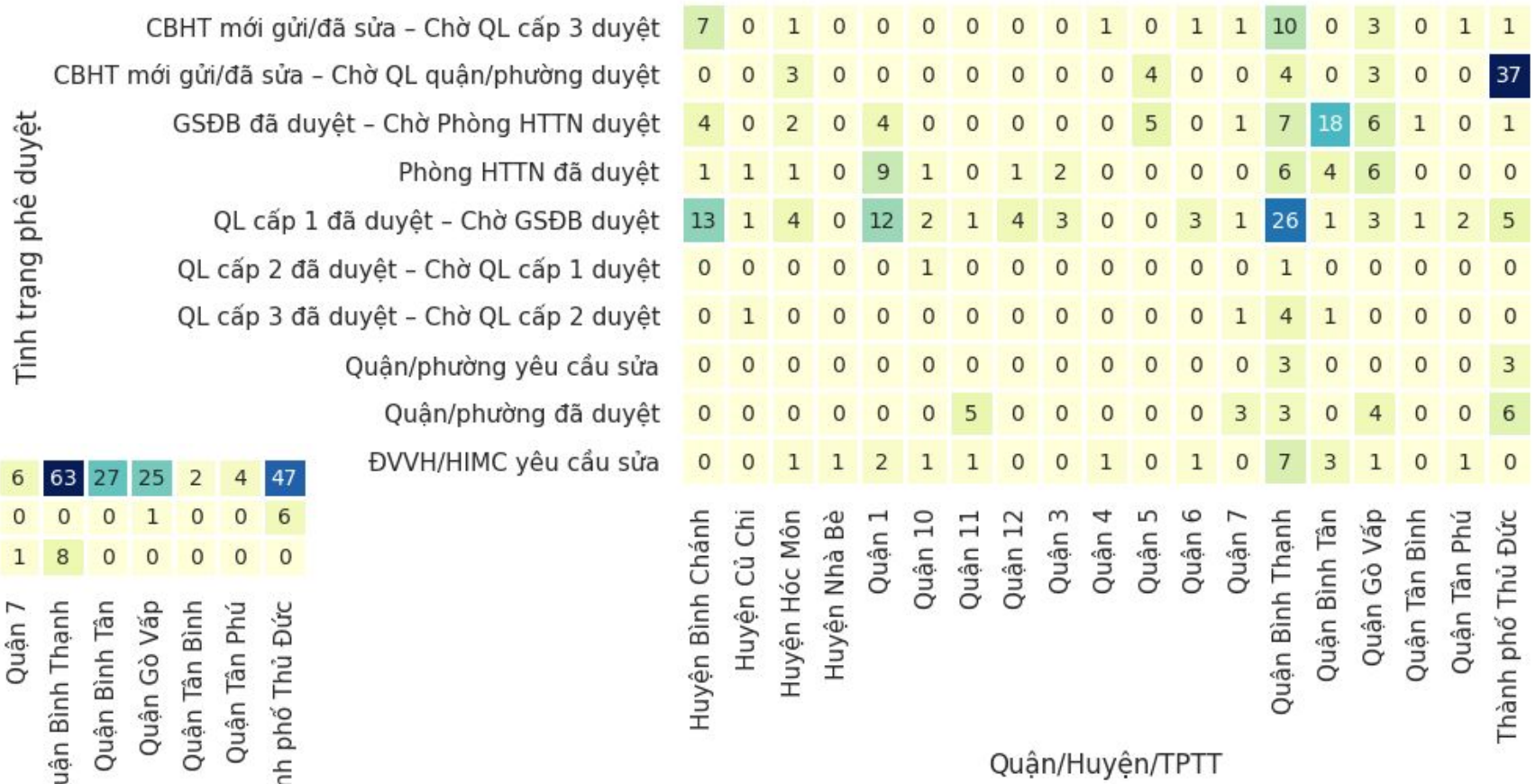




*Loại đồ thị nào thể hiện đầy đủ thông tin hơn về (các) cột dữ liệu đang được tìm hiểu?*

# Phân tích mô tả

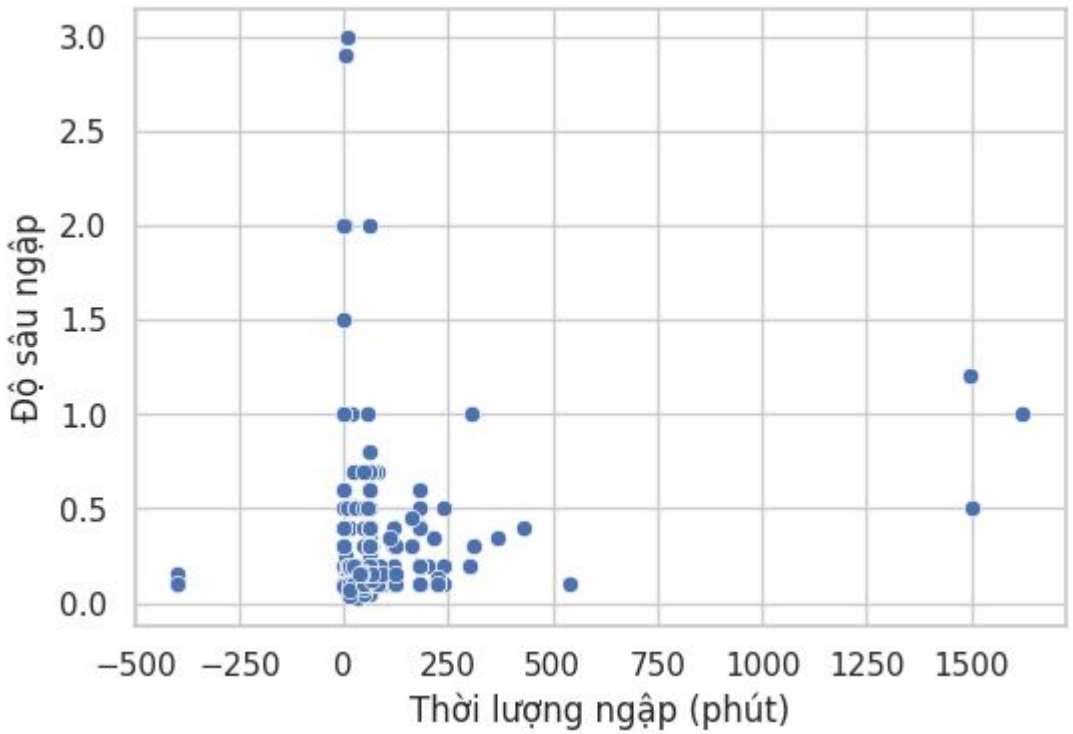
Hiển thị cặp giá trị định tính - định tính (dữ liệu minh họa từ nền tảng FEDS)

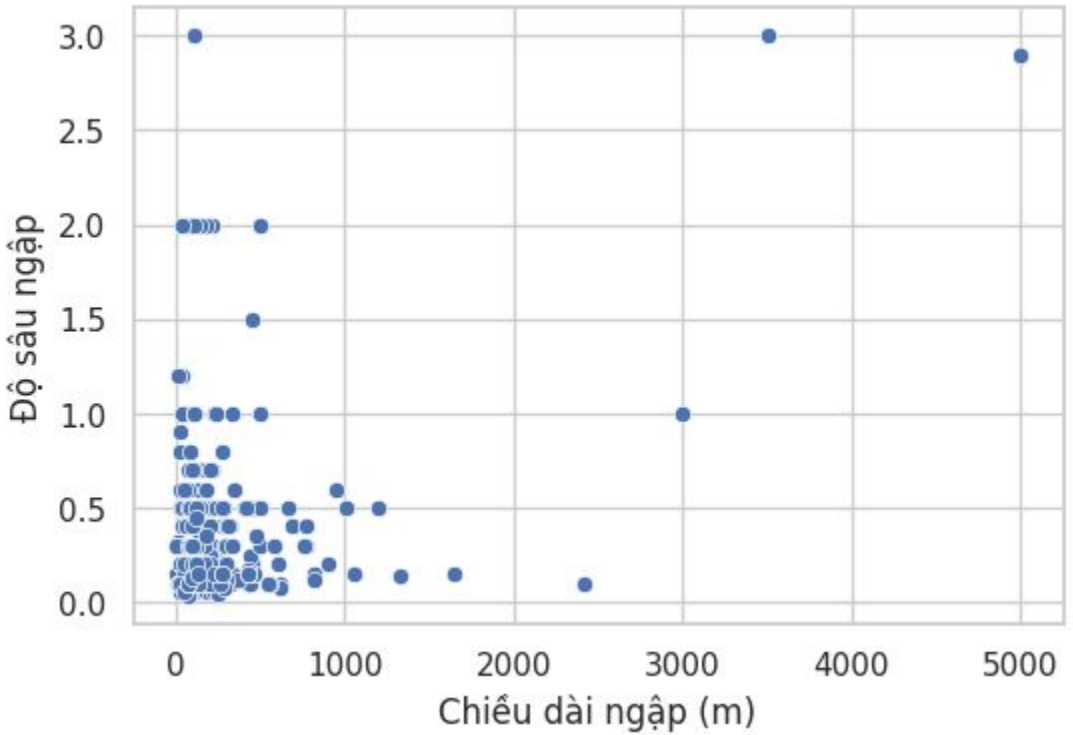


# Phân tích mô tả

Hiển thị cặp giá trị định lượng - định lượng (dữ liệu minh họa từ nền tảng FEDS)



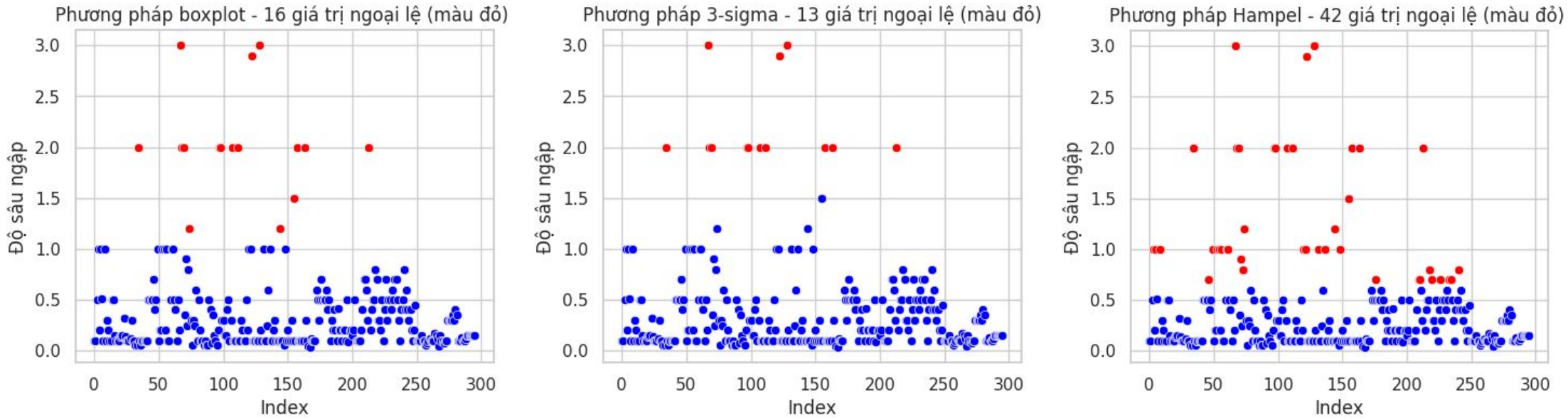

*Thông tin được thể hiện trong các đồ thị trên có tương ứng với các hiểu biết trước đây về tình hình ngập đường ở TPHCM không?*

# Các giá trị ngoại lệ

> ... là các quan sát có giá trị không nhất quán với các quan sát còn lại của tập dữ liệu đó.

V. Barnett and T. Lewis. *Outliers in Statistical Data*. Wiley, 2nd edition, 1984

- Các giá trị lớn hoặc nhỏ bất thường làm biến dạng các đặc tính của một tập dữ liệu.
- Các giá trị ngoại lệ có thể do *tổng hợp các lỗi dữ liệu* hoặc là *các trường hợp đặc biệt*.
- Các phương pháp toán *có thể giúp xác định* nhưng *không thể giúp diễn giải* các giá trị ngoại lệ.




*Nguyên nhân của các điểm giá trị "Độ sâu ngập" bất thường (màu đỏ trong hình trên) trong dữ liệu mẫu từ nền tảng FEDS là gì?*

# Các giá trị inliers

> ... là các giá trị lỗi nhưng nằm trong phạm vi ý nghĩa thống kê của một phân phối xác suất.

D. DesJardins. Paper 169: Outliers, inliers and just plain liars – new eda+ techniques for understanding data. In *Proceedings SAS User's Group International Conference*, SUG126. Cary, NC, USA, 2001

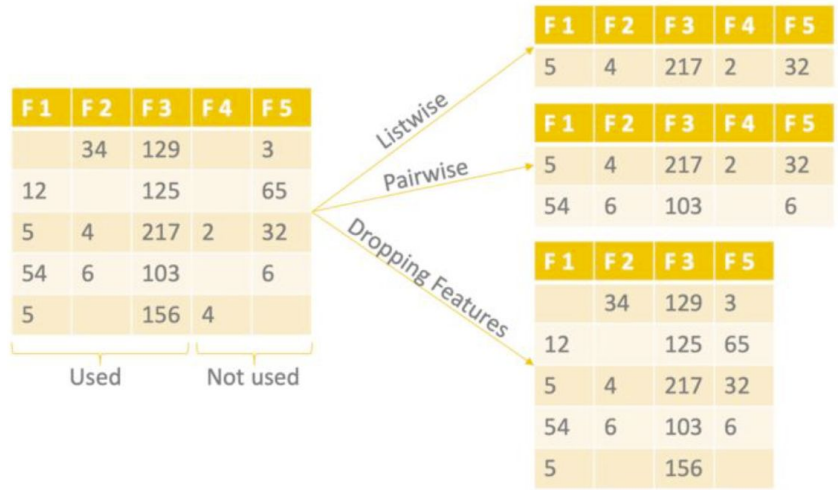
- Các giá trị inliers thường ở dưới dạng các giá trị có tần suất xuất hiện cao bất thường.
- Tương tự các giá trị ngoại lệ, các giá trị inliers không nhất thiết là các điểm dữ liệu xấu, nhưng cần được tìm hiểu thêm.

| | Chiều dài ngập (m) | count |
|---|---|---|
| 0 | 150.00 | 10 |
| 1 | 100.00 | 9 |
| 2 | 200.00 | 6 |
| 3 | 435.58 | 6 |
| 4 | 232.68 | 6 |

*Nguyên nhân của các giá trị 'Chiều dài ngập (m)' có tần suất xuất hiện bất thường được xác định như trên là gì?*

# Các giá trị bị mất

- Mất giá trị có thể do lỗi con người, lỗi thiết bị, hoặc do tích hợp các cơ sở dữ liệu.
- Các giá trị ngoại lệ và inliers, nếu được xác định là giá trị xấu, cũng có thể xem bị mất.
- Các giá trị bị mất có thể được phân thành 3 loại
  - Mất ngẫu nhiên hoàn toàn (Missing completely at random)
  - Mất ngẫu nhiên (Missing at random)
  - Mất không ngẫu nhiên (Missing not at random)
- Các cách xử lý dữ liệu bị mất
  - xóa các hàng dữ liệu hoặc cột dữ liệu bị mất giá trị, hoặc
  - thay thế các dữ liệu bị mất
    - phụ thuộc vào loại hình giá trị bị mất.
    - không dùng 1 giá trị duy nhất để thay thế các giá trị bị mất.




https://www.kdnuggets.com/2020/09/missing-value-imputation-review.html

https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.578.5816&rep=rep1&type=pdf
http://www.stat.columbia.edu/~gelman/arm/missing.pdf

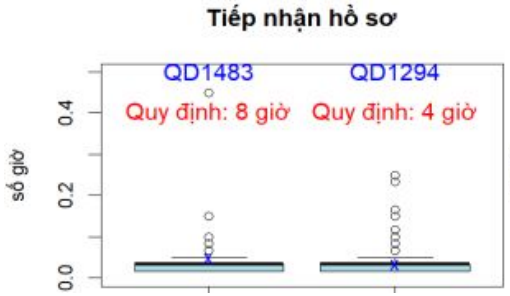


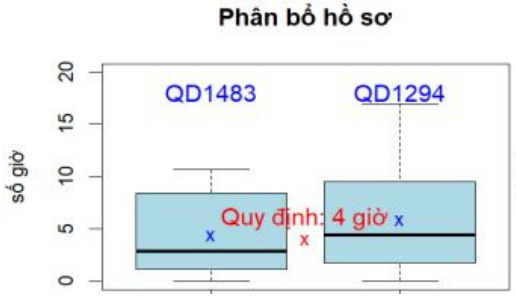
# Khai thác dữ liệu

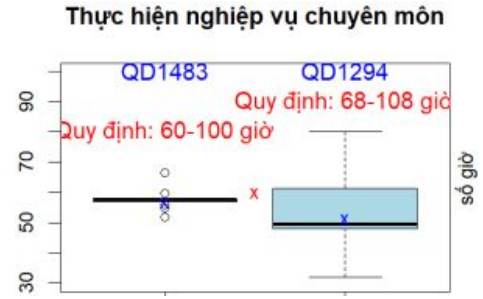
# Đánh giá hiệu quả quy trình cấp LLTP An Giang

Quy trình thực hiện thủ tục cấp LLTP của Sở Tư pháp An Giang (Quyết định 1483/QĐ-UBND)

| Bước 1 | Bước 2 | Bước 3 | Bước 4 | Bước 5 | Bước 6 |
|---|---|---|---|---|---|
| •Tiếp nhận<br>•08 giờ | •Phân công cán bộ xử lý hồ sơ<br>•04 giờ | •Thực hiện nghiệp vụ chuyên môn<br>•60 hoặc 100 giờ | •Ký duyệt hồ sơ<br>•04 giờ | •Lưu sổ văn thư<br>•04 giờ | •Trả kết quả<br>•Không tính thời gian |

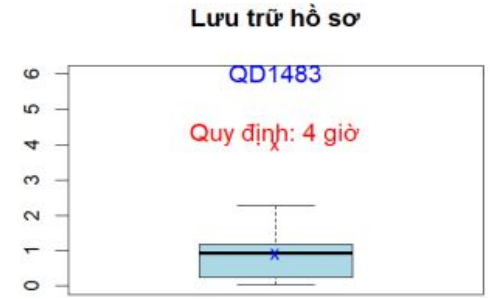


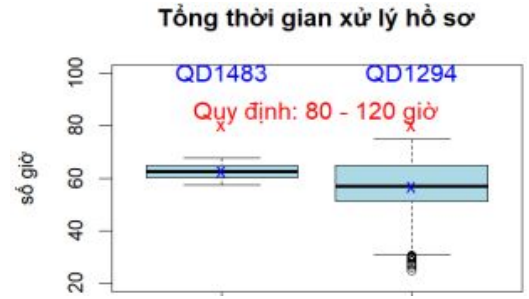





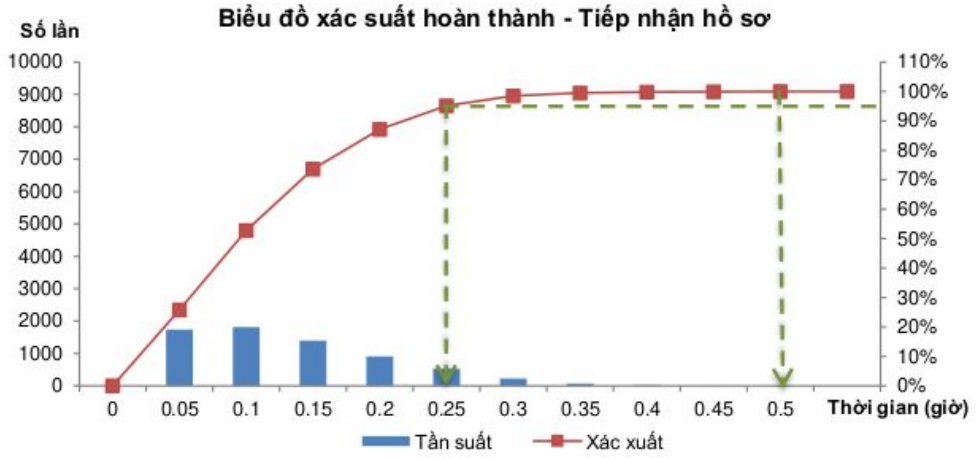


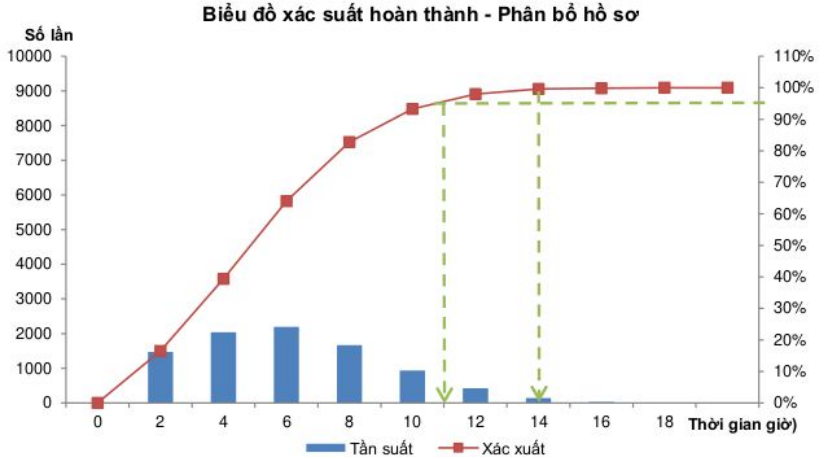


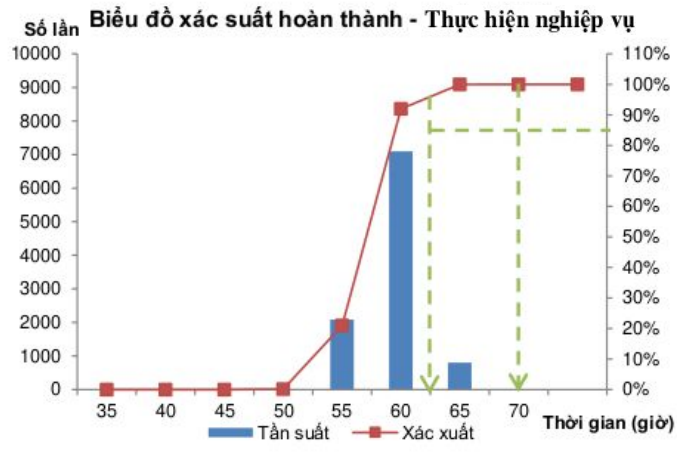



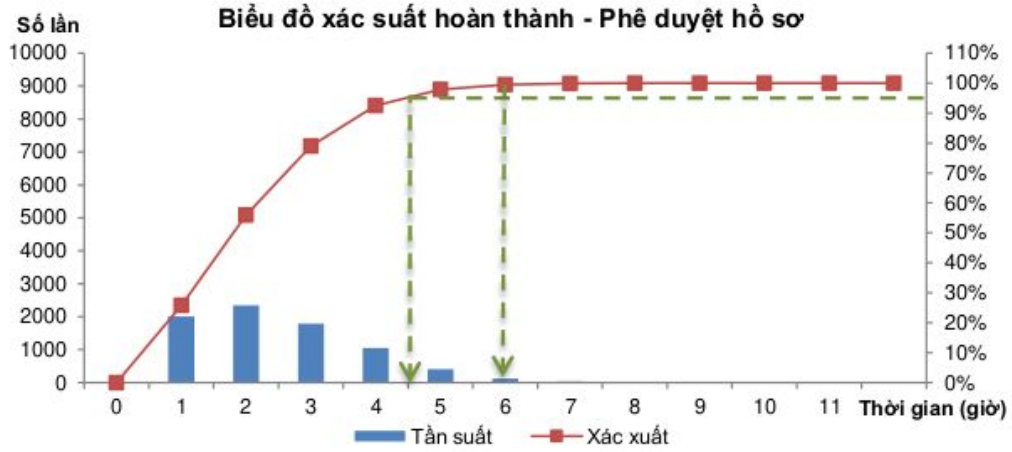
# Đánh giá hiệu quả quy trình cấp LLTP An Giang

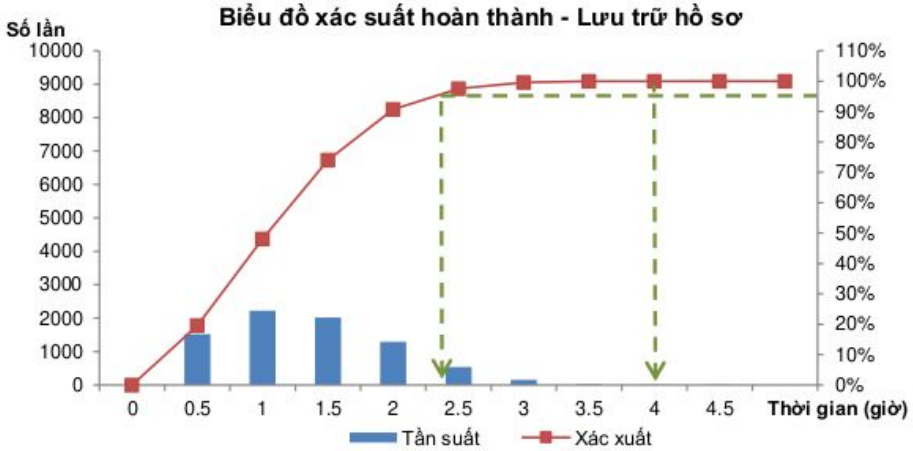
Mô phỏng thời gian xử lý của từng bước (10000 lần)



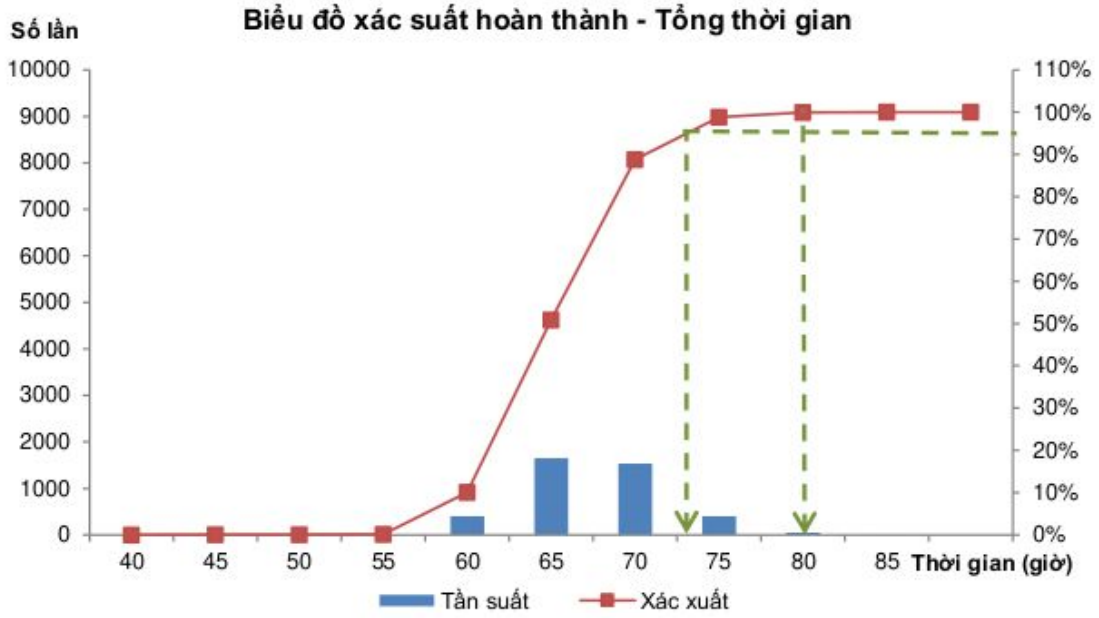



# Đánh giá hiệu quả quy trình cấp LLTP An Giang

Mô phỏng thời gian xử lý của từng bước (10000 lần)




**Bảng 4.1. Xác suất thời gian hoàn thành các bước xử lý hồ sơ**

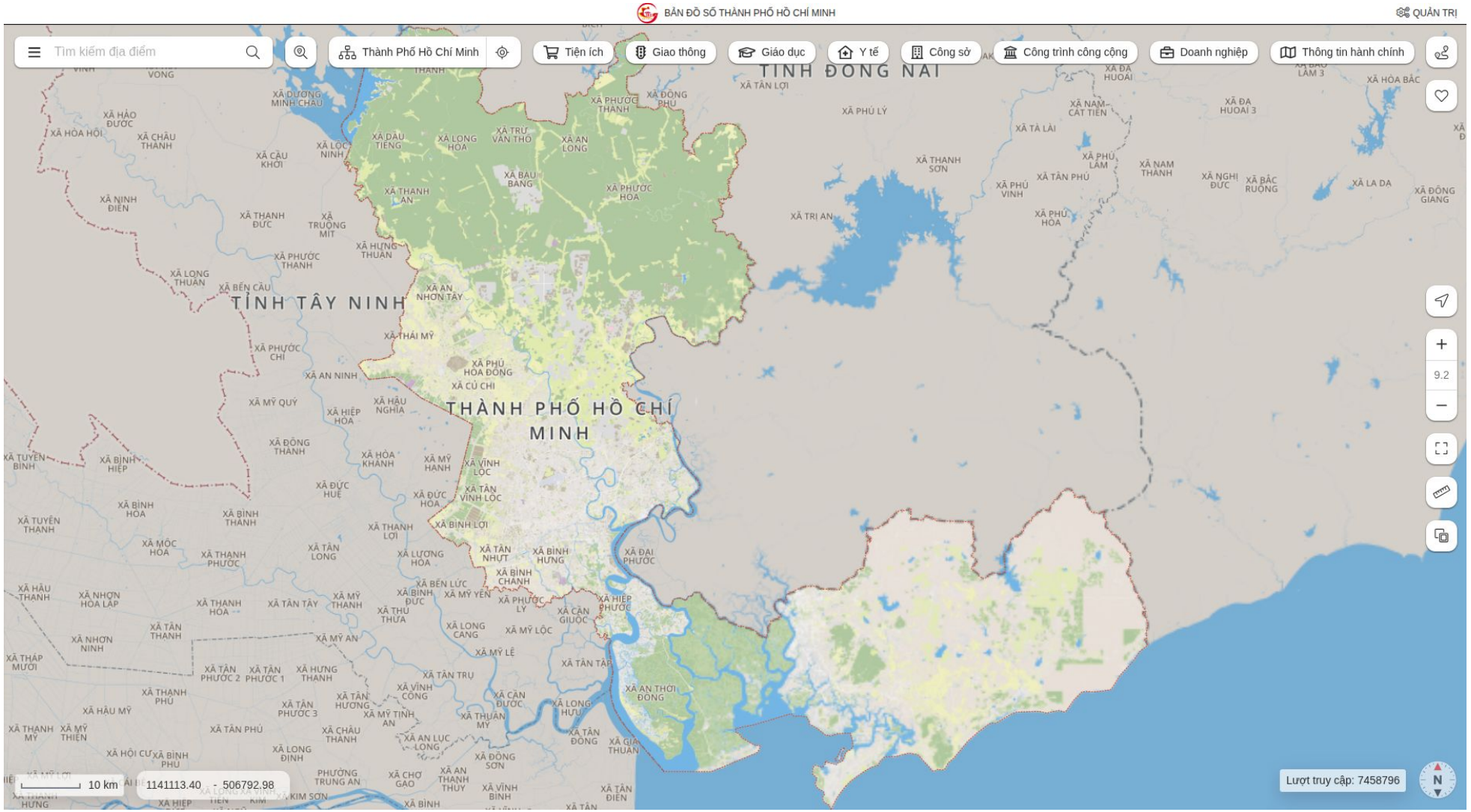
*Đơn vị tính: giờ*

| Xác suất hoàn thành đúng hạn theo thời gian | Tổng thời gian | Tiếp nhận | Phân bổ hồ sơ | Nghiệp vụ | Phê duyệt | Lưu trữ |
|---|---|---|---|---|---|---|
| 95% | 80,85 | 0,35 | 11 | 62,5 | 4,5 | 2,5 |
| 100% | 94,5 | 0,5 | 14 | 70 | 6 | 4 |

**Bảng 4.2. Thời gian hoàn thành theo quy định hiện nay**

| | Tổng thời gian | Tiếp nhận | Phân bổ hồ sơ | Nghiệp vụ | Phê duyệt | Lưu trữ |
|---|---|---|---|---|---|---|
| Quy định hiện tại (theo giờ) | 80 hoặc 120 | 8 | 4 | 60 hoặc 100 | 4 | 4 |

# Bản đồ số TPHCM




https://bando.tphcm.gov.vn/

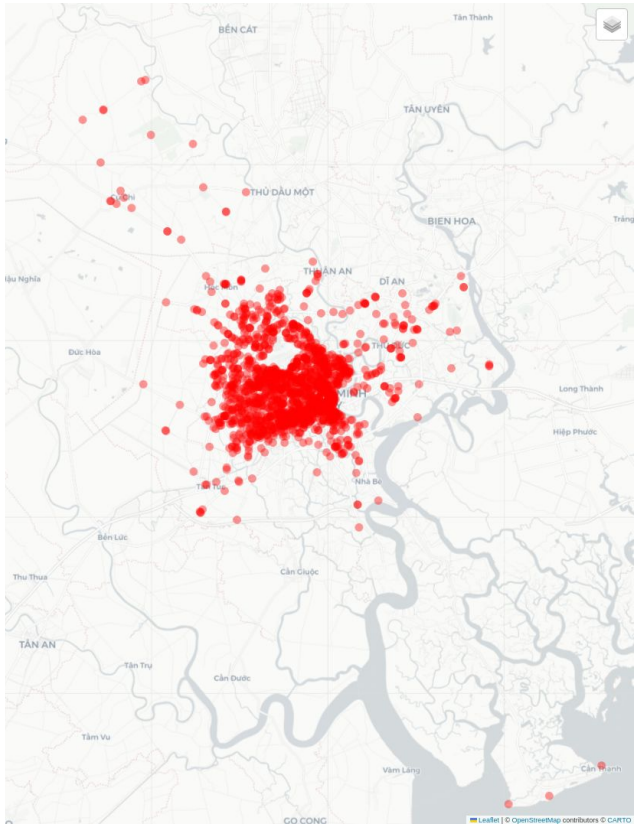
# Bản đồ số TPHCM - Các lớp dữ liệu hạ tầng xã hội

| | | |
|---|---|---|
| **Tiện ích** | Chợ<br>Cửa hàng tiện lợi<br>Siêu thị<br>Trung tâm thương mại | **Mua sắm (1891)** |
| | Ngân hàng | **Tài chính (2318)** |
| **Giao thông** | Bến xe, trạm dừng<br>Bến buýt đường sông<br>Nhà ga MRT | **Giao thông công cộng (5905)** |
| | Bãi xe<br>Cầu<br>Đèn tín hiệu<br>Đường | |
| **Giáo dục** | Mầm non<br>Tiểu học<br>Trung học cơ sở<br>Trung học phổ thông<br>Cao đẳng<br>Đại học | **Giáo dục (3520)** |
| | Phòng giáo dục | |

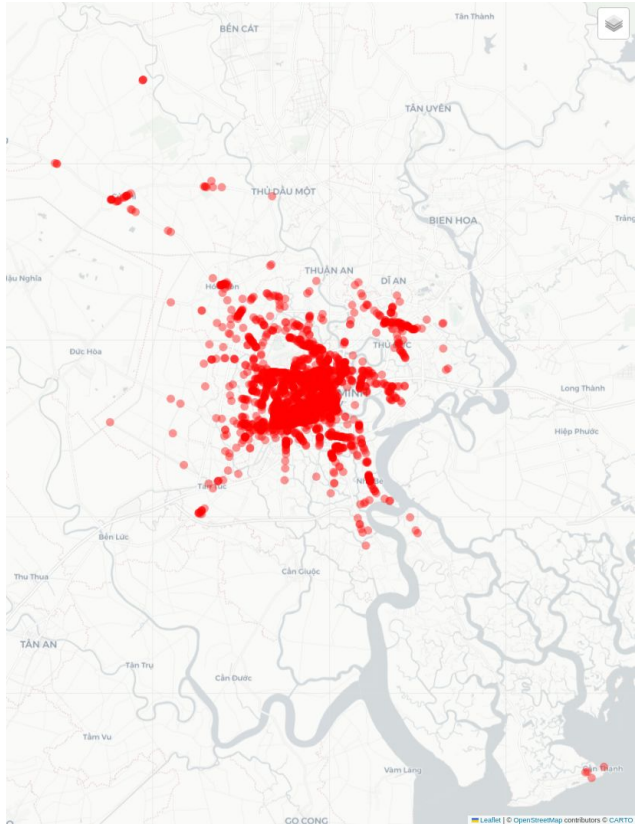
| | | |
|---|---|---|
| **Y tế** | Bệnh viện<br>Phòng khám<br>Trung tâm y tế | **Y tế (267)** |
| **Công sở** | Công an HCM<br>Công sở | **Hành chính (982)** |
| **Công trình công cộng** | Bảo tàng, khu di tích<br>Công viên<br>Khu du lịch | **Giải trí (723)** |
| | Khách sạn<br>Nhà hàng, quán ăn<br>Chùa<br>Nhà thờ | |
| **Doanh nghiệp** | Công ty<br>Doanh nghiệp tư nhân | |

# Dữ liệu sử dụng trong tình huống nghiên cứu

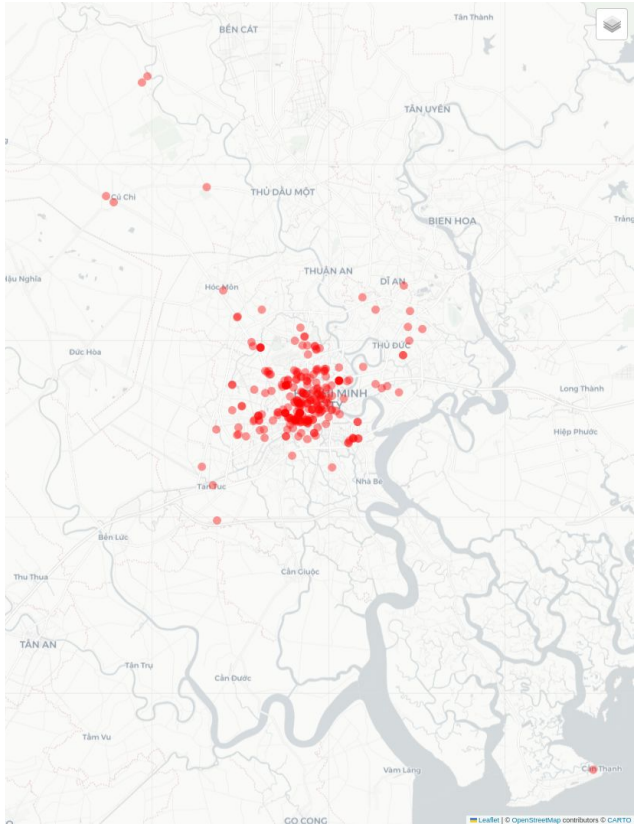
POI Mua sắm (1891)



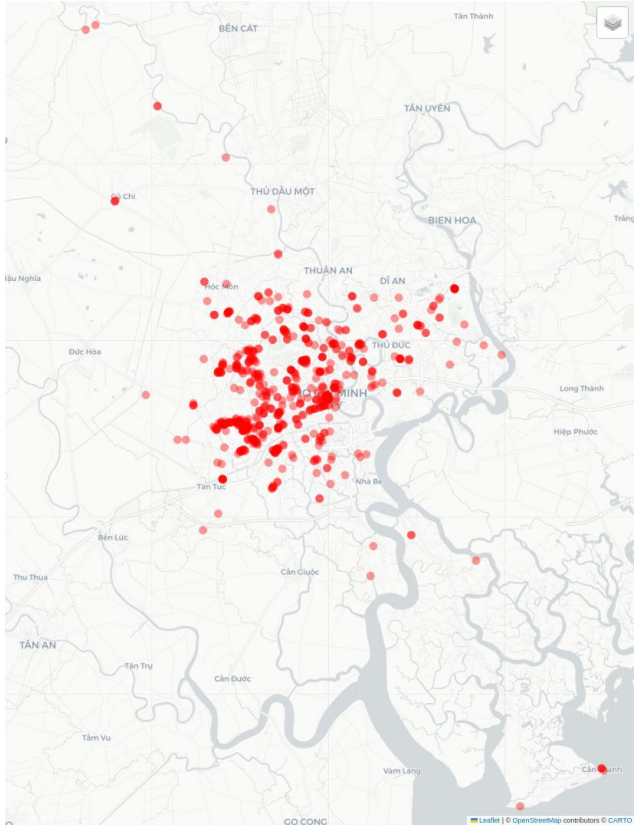

POI Tài chính (2318)




POI Y tế (267)

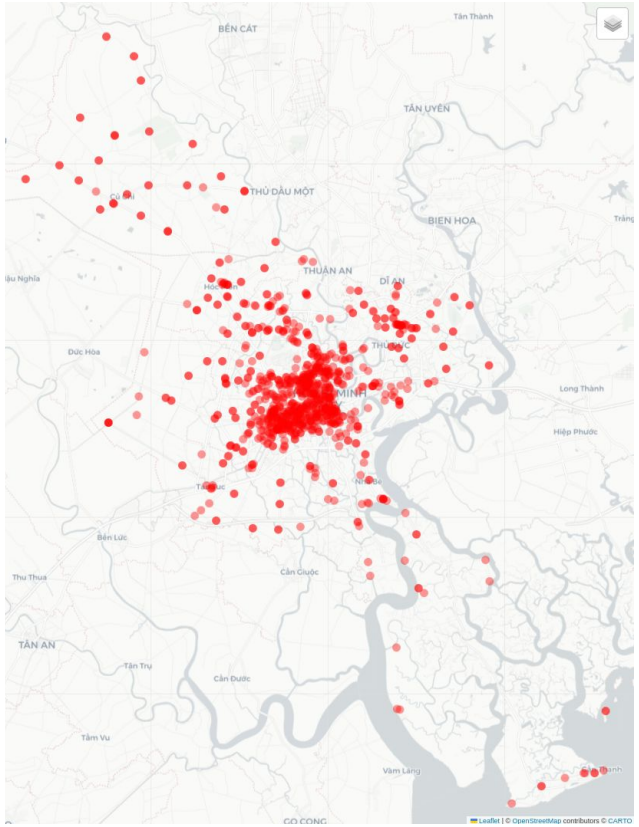



POI Giải trí (723)

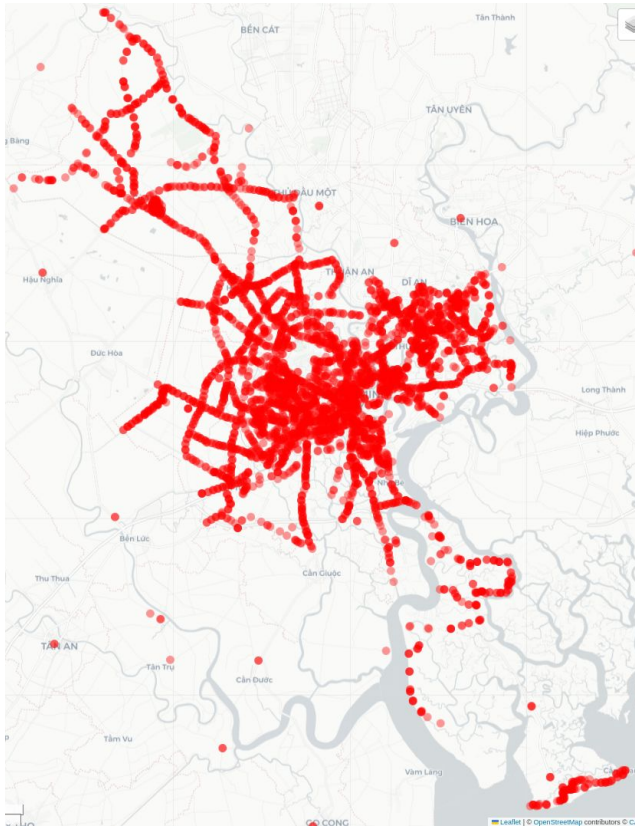


# Dữ liệu sử dụng trong tình huống nghiên cứu

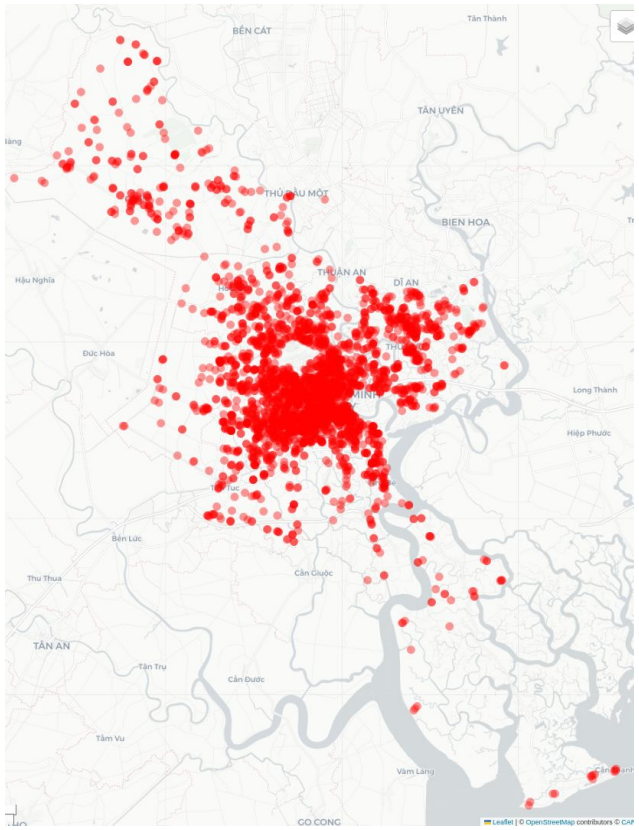
POI Hành chính (982)




POI Giao thông công cộng (5905)



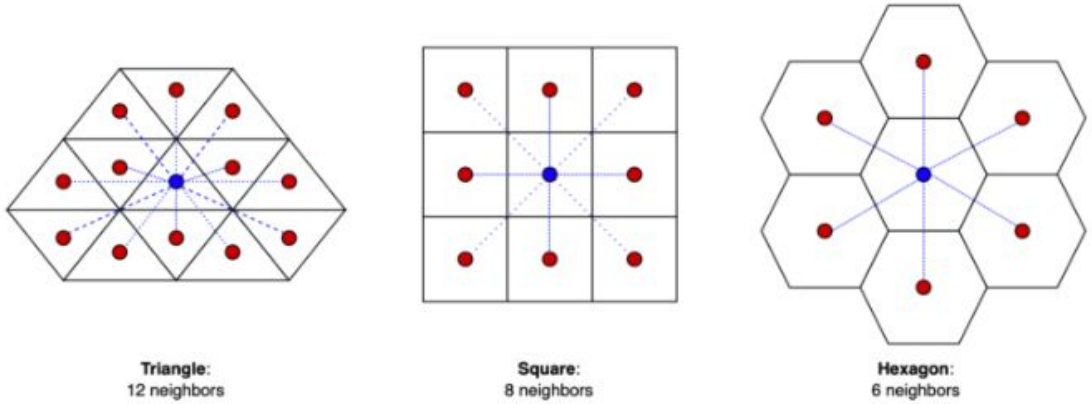

POI Giáo dục (3520)



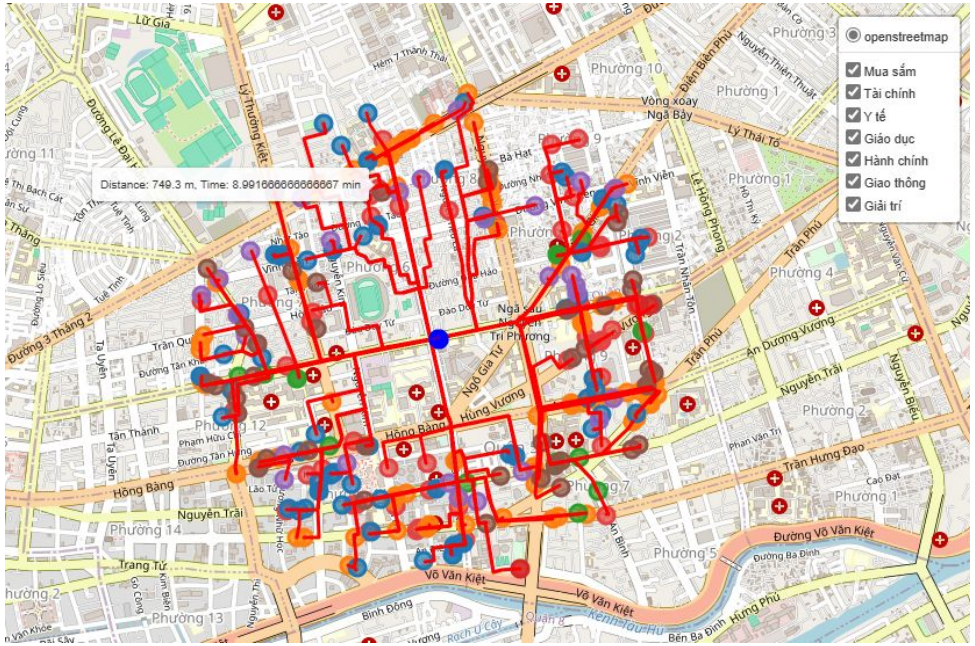
# Thời gian đến các POIs từ các vị trí địa lý khác nhau

- Diện tích TPHCM được chia thành ~17,000 ô lục giác.
- Chiều dài mỗi cạnh ô lục giác khoảng 200 m.
- Thời gian đi bộ từ trọng tâm mỗi ô lục giác đến từng POI được tính bằng dịch vụ Routing của Maps API của Vietmap.

Vì sao dùng “thời gian đi bộ”?



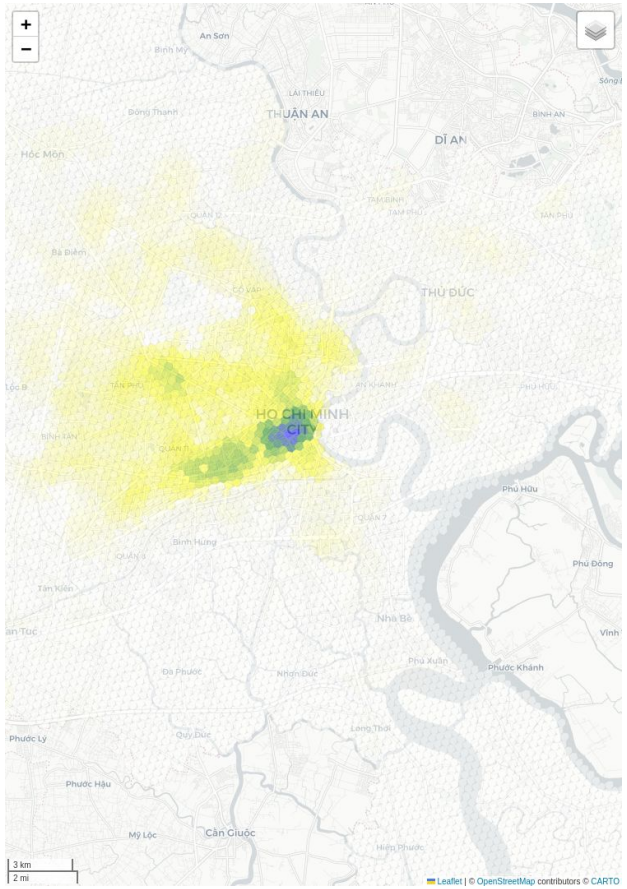

https://towardsdatascience.com/exploring-location-data-using-a-hexagon-grid-3509b68b04a2/



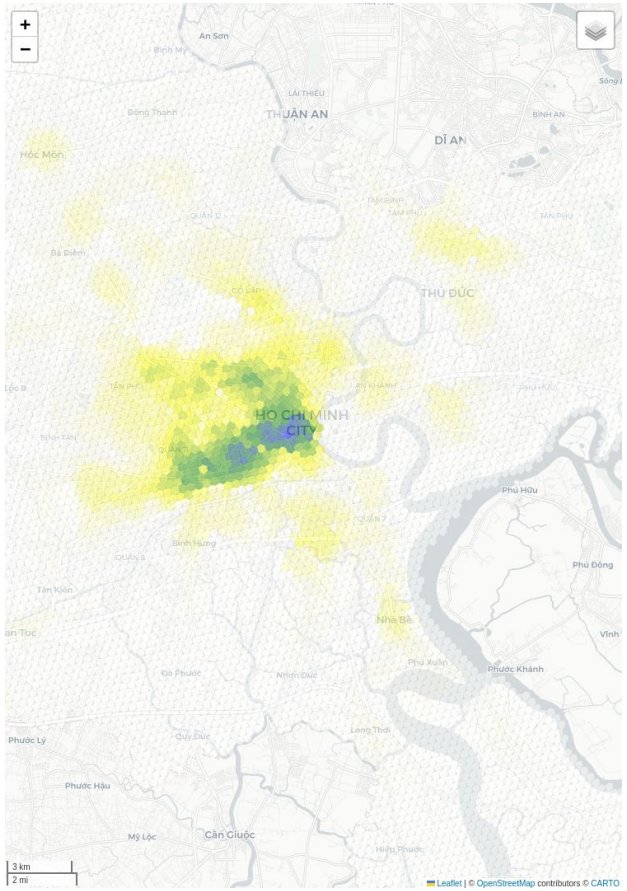
# Khả năng tiếp cận tiện ích xã hội

Số POIs trong bán kính 15 phút đi bộ

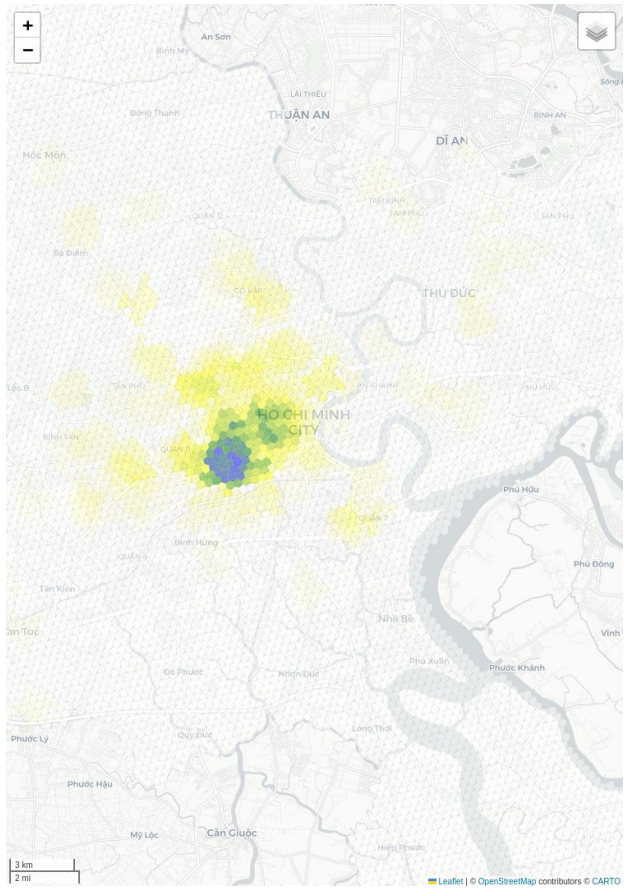
POI Mua sắm



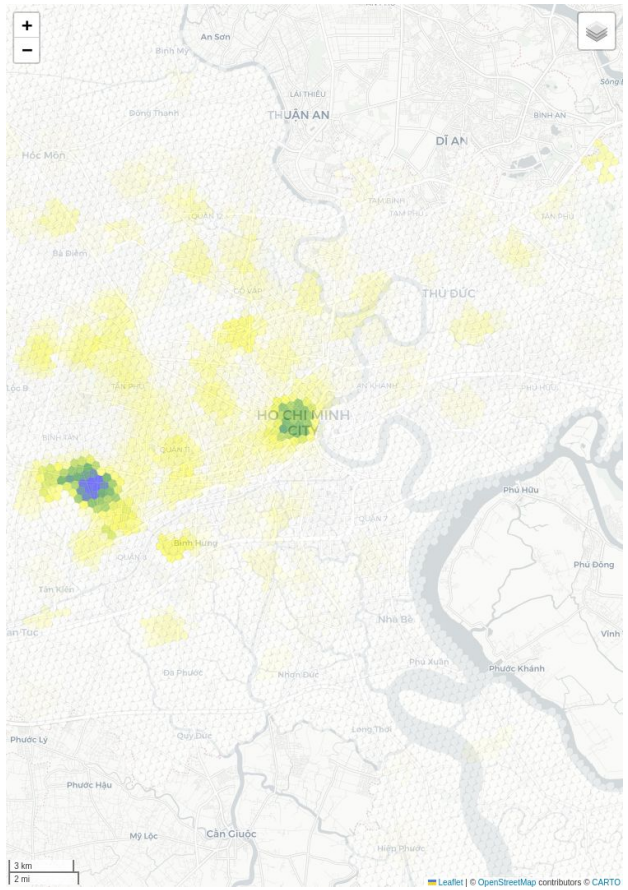

POI Tài chính




POI Y tế



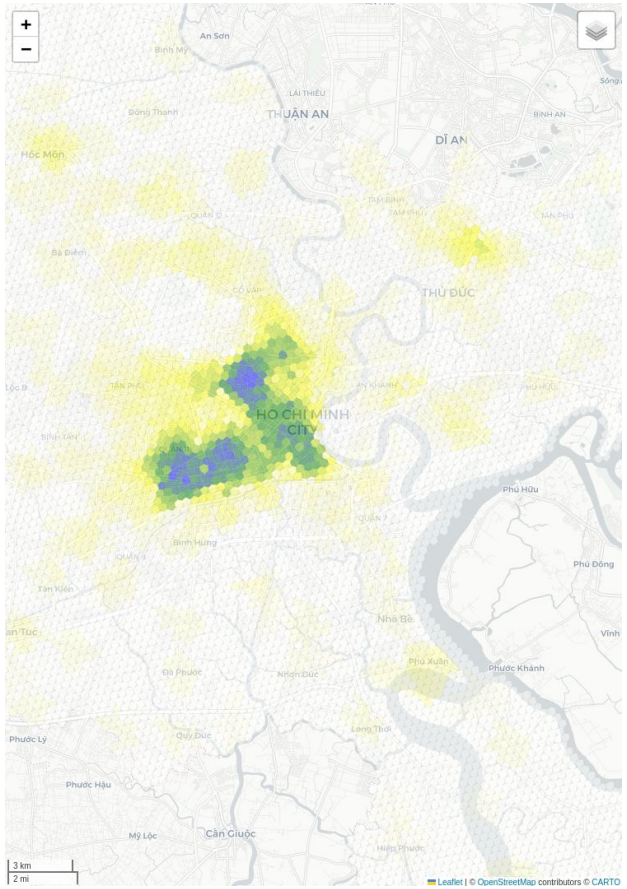

POI Giải trí



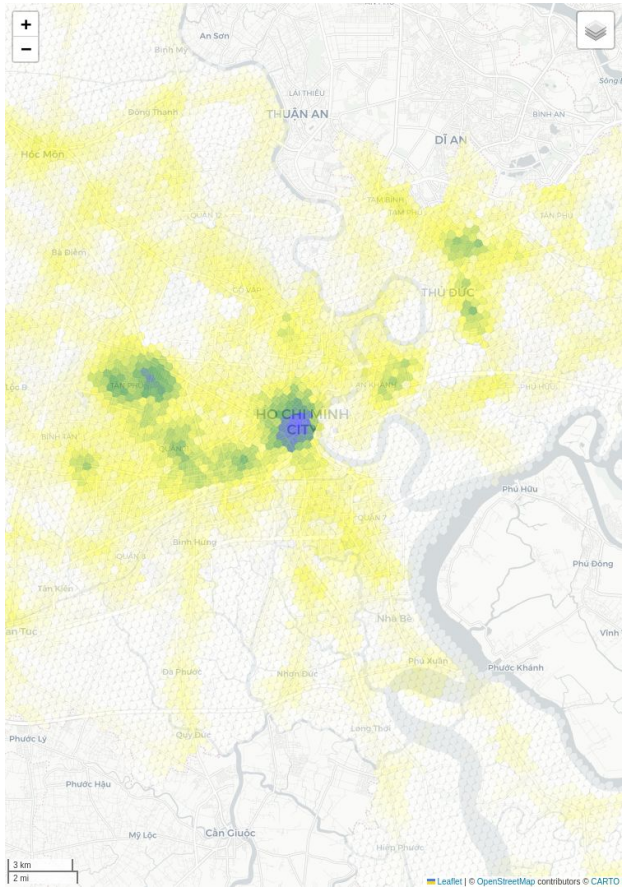
# Khả năng tiếp cận tiện ích xã hội

Số POIs trong bán kính 15 phút đi bộ

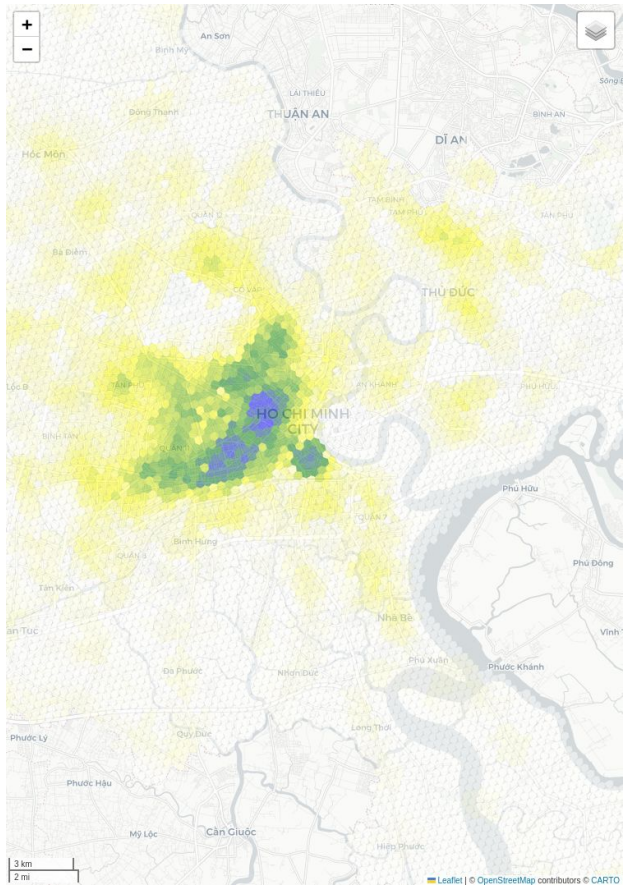
POI Hành chính



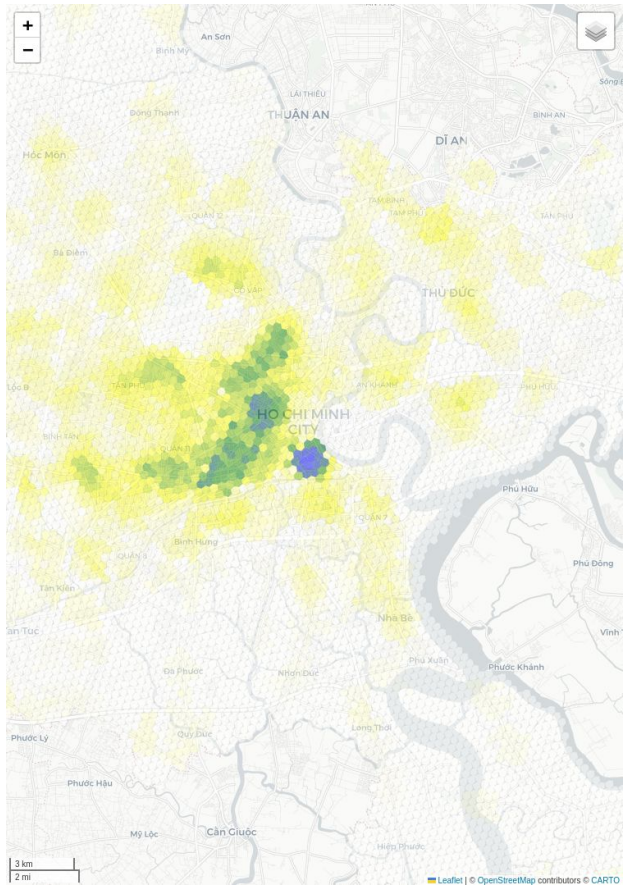

POI Giao thông công cộng



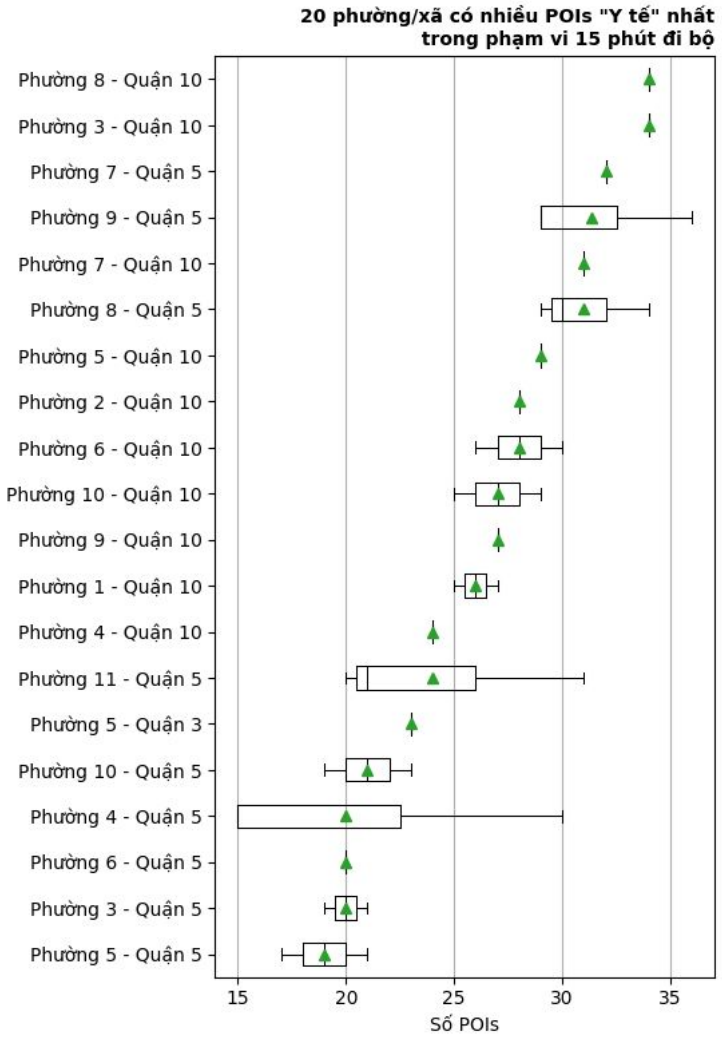

POI Giáo dục

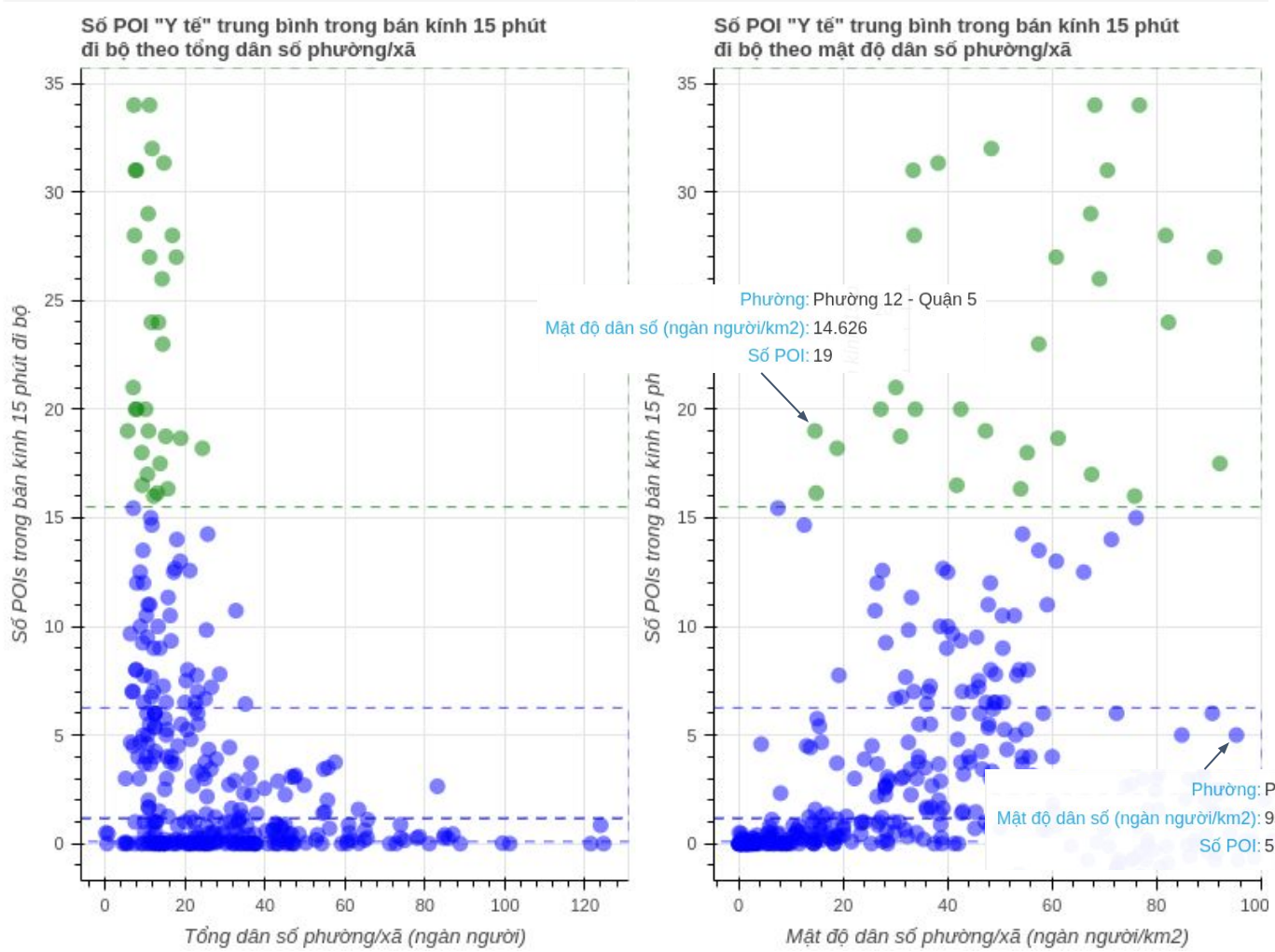



POI Mầm non



# Khả năng tiếp cận tiện ích xã hội theo phường/xã



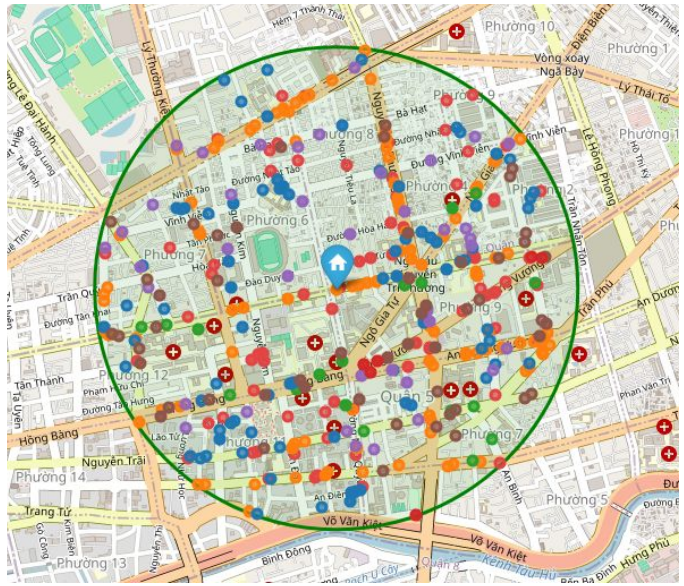



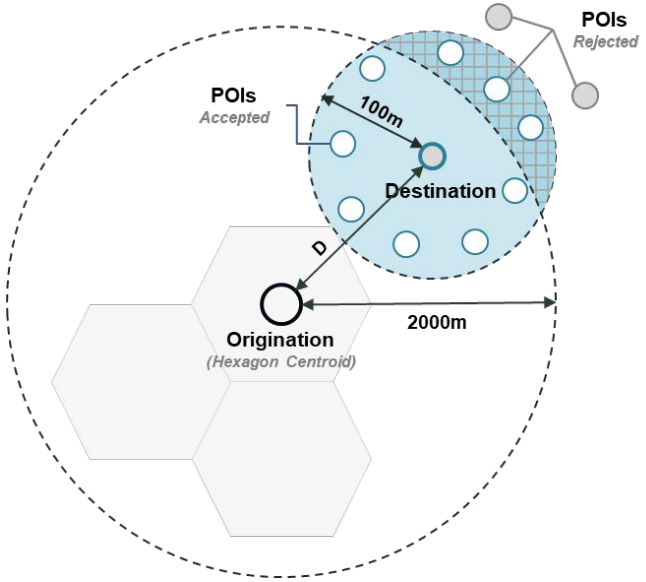
**Tất cả mô hình đều sai, nhưng một vài mô hình hữu ích.**

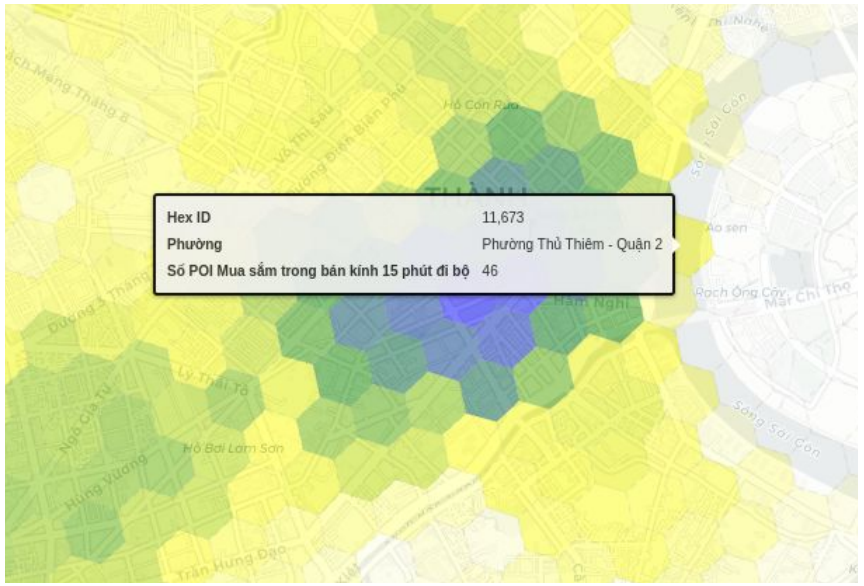
*[All models are wrong, some are useful.]*

# Những yếu tố ảnh hưởng đến kết quả phân tích - Phương pháp

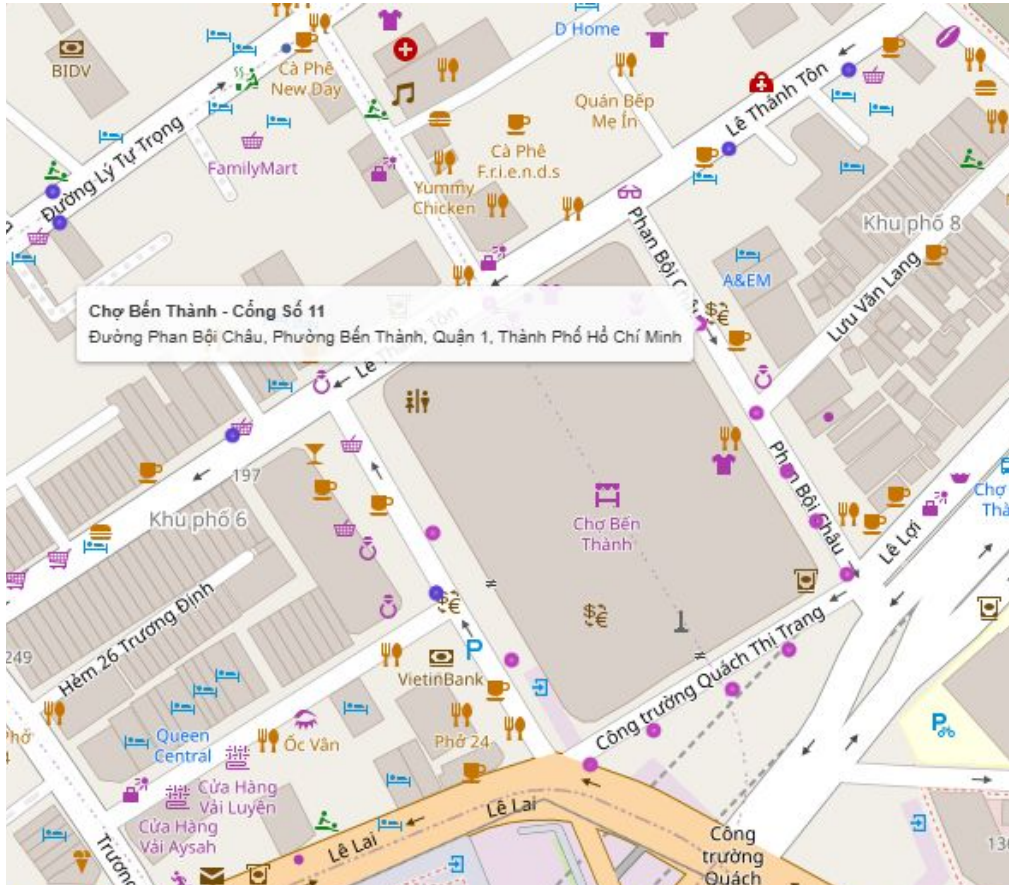
> *Chất lượng của một mô hình phụ thuộc chất lượng các giả định của mô hình đó.*
>
> *[A model is only as good as its assumptions.]*



# Những yếu tố ảnh hưởng đến kết quả phân tích - Dữ liệu




Các POIs của chợ Bến Thành (cho từng cổng)




Các POIs của một công viên nội khu chung cư



# Quản trị dữ liệu

# “... nhưng một vài mô hình hữu ích”

- Dữ liệu từ Bản đồ số TPHCM mở ra nhiều tiềm năng khai thác ứng dụng cho nghiên cứu, quản lý nhà nước, và thương mại.
- Một số gợi ý để hiện thực hóa các tiềm năng này, góp phần tạo các phân tích và sản phẩm dữ liệu tốt (“*mô hình hữu ích*”)
  - Đảm bảo chất lượng dữ liệu
  - Mô tả chất lượng dữ liệu
  - Đáp ứng đúng nhu cầu của người dùng

# Đảm bảo chất lượng dữ liệu

- "Dữ liệu mở phải **toàn vẹn và phản ánh đầy đủ nội dung thông tin cơ quan nhà nước cung cấp**" và "Dữ liệu mở được cung cấp là **dữ liệu được cập nhật mới nhất**" (Điều 17, Nghị định 47/2020/NĐ-CP)
- "Danh mục dữ liệu mở được **rà soát, sửa đổi (nếu có) theo định kỳ, tối thiểu 6 tháng một lần**; dữ liệu mở đã được công bố theo danh mục phải được cập nhật hoặc bổ sung (nếu có) trong thời hạn không quá 03 tháng kể từ thời điểm được công bố." (Điều 20, Nghị định 47/2020/NĐ-CP)

*Đây là cơ sở để "Bản đồ số TPHCM là nền tảng bản đồ số dùng chung cho toàn Thành phố, thay thế các ứng dụng sử dụng bản đồ như Openstreetmap, Google Maps."*

https://chuyendoiso.hochiminhcity.gov.vn/-/tp-hcm-se-ua-vao-su-dung-ban-o-so-tich-hop-du-lieu-a-nganh?redirect=/tin-tuc-tong-hop

# Mô tả chất lượng dữ liệu

- Nội dung kiểm tra, đánh giá dữ liệu bao gồm: (Điều 16, Nghị định 47/2020/NĐ-CP)
  - (c) Đánh giá chất lượng dữ liệu bao gồm: **độ chính xác dữ liệu, độ toàn vẹn của dữ liệu, dữ liệu bất thường**;
  - (d) Đánh giá về **duy trì, vận hành, cập nhật dữ liệu, nhật ký cập nhật, khai thác của dữ liệu**;
  - (đ) Đánh giá về chia sẻ dữ liệu, đối tượng sử dụng, mục đích chia sẻ phù hợp với các quy định, quy chế khai thác, sử dụng dữ liệu chia sẻ;
- Kết quả đánh giá phải **lập thành văn bản**, đồng thời đề xuất các hoạt động cần thiết để duy trì dữ liệu (nếu có) (Điều 16, Nghị định 47/2020/NĐ-CP).

*Mô tả chất lượng các tập dữ liệu **"đúng đủ sạch sống" như thế nào tại thời điểm nào** là mấu chốt để người dùng **đánh giá tính phù hợp** và tăng khả năng ứng dụng của các tập dữ liệu này.*

Introduction to data.gov.sg

What's new

USER GUIDE

Content quality guidelines

Data types

Data principles

For data consumers

How to search for real-time APIs

For data owners

How to onboard as an admin

How to login

How to publish data

How to unpublish and/or delete datasets

How to edit datasets

How to generate API keys

Write APIs

DEVELOPER GUIDE

API overview

Collection APIs

Dataset APIs

USER GUIDE

# Content quality guidelines

As an open data collective, we're focused on providing quality data that users can not only explore, analyse or use for development, but derive value and meaning from. If you are uploading data, we've created a guide to help you get it ready for publishing.

## Title of Dataset

Your title is what first attracts users to your data. To help users find the data they need, it is important that your title is not only accurate but easy to understand.

We recommend titles:

- **Stick only to what is essential. As each title is accompanied by a description, your title should only convey what a user needs to know at a glance.**
  - For collections, titles can give a broad picture of the datasets contained. You might consider adding slightly more details to the dataset files themselves.
  - You might title a collection, "Annual pet registration", with the datasets, "Pet registration by type", "Pet registration by location", and "Pet registration by month".
  - For individual datasets, while titles should be broad thematically, they should still be specific to the data being shared.
  - You might title an individual dataset, "Monthly pet registration by location".
- **Avoid having information that already appears as separate data points**
  - Coverage years: Your chosen coverage period will be shown below your title. Unless differentiating between datasets in a collection, try to avoid adding these in your titles.
  - Agency name: While your agency will be shown as the data provider, you might consider adding this in only if you believe it is beneficial for users to know. This could be to avoid confusion if you are publishing data similar to another agency, for example.

    While it is possible to use the abbreviated form of your agency name, this should ideally be consistent across the datasets and collections you upload.
- **Be consistent. For example, if you add coverage years to dataset titles within a collection, do try to add years for all the datasets.**

## Description of your dataset

Title of Dataset

Description of your dataset

Column descriptions

Data principles

Changelog

Get in touch

Was this helpful?

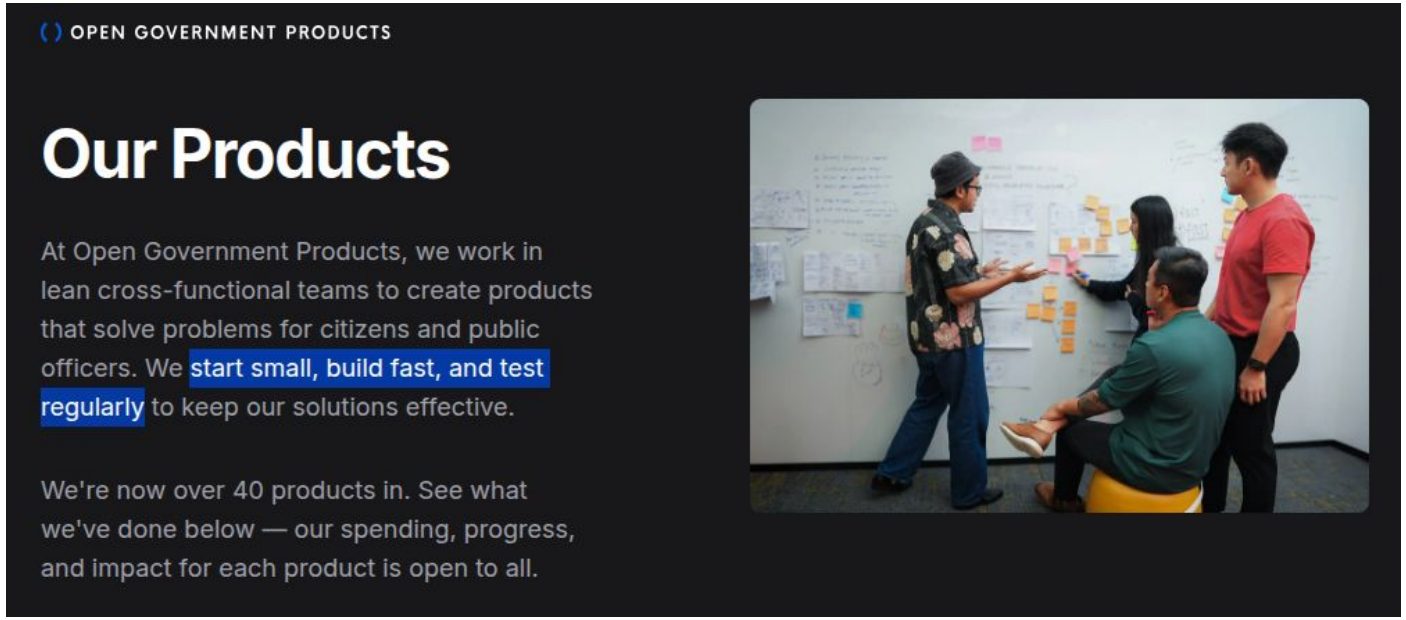
# Đáp ứng nhu cầu người dùng

- Thực hiện các giải pháp khuyến khích tổ chức, cá nhân sử dụng dữ liệu mở (Điều 19, Nghị định 47/2020/NĐ-CP)
  - Tiếp nhận các ý kiến phản hồi của tổ chức, cá nhân để xác định mức độ ưu tiên và thực hiện công bố dữ liệu mở **phù hợp với nhu cầu**;
  - Hỗ trợ các **tổ chức, cá nhân trong xã hội**, cộng đồng thực hiện các hoạt động xây dựng, đóng góp để **mở rộng dữ liệu mở**.

*Các tổ chức và cá nhân, đặc biệt các viện/trường, tham gia **xây dựng các sản phẩm mẫu (prototypes)** từ dữ liệu mở sẽ đẩy nhanh việc **xác định nhu cầu, cải thiện chất lượng** và **hiệu quả sử dụng** của dữ liệu.*

# Đáp ứng nhu cầu người dùng

## Open Government Products (Singapore)




https://products.open.gov.sg/




https://www.open.gov.sg/how-we-build/

# Thảo luận

Fulbright
SCHOOL OF PUBLIC POLICY AND MANAGEMENT

Disruptive Technologies
for Public Asset Governance

THE WORLD BANK
IBRD • IDA | WORLD BANK GROUP

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Embassy of Switzerland
**Swiss Cooperation Office in Vietnam**

# Cảm ơn

***Contacts:*** *Huỳnh Nhật Nam – Giảng viên, Trường Chính sách công và Quản lý Fulbright, nam.huynh@fulbright.edu.vn*
*Vũ Thúy Vinh – Quản lý dự án, Trường Chính sách công và Quản lý Fulbright, vinh.vu@fulbright.edu.vn*