

Báo Cáo Đề Tài Thu Thập Dữ Liệu

Nguồn Nước Tp. Hà Nội

Môn Kho Dữ Liệu - Nhóm 2

Thành Viên Nhóm

Vũ Tùng Dương - 16020921 (Trưởng Nhóm)

Nguyễn Thế Chí Dũng - 16020906

Nguyễn Văn Dương - 16020919

Trương Văn Dương - 16020081

Lê Quang Đạo - 16020875

Mục lục

Chương 1: Giới thiệu	2
Chương 2: Quá trình thu thập dữ liệu	2
Tìm dữ liệu khu vực đồng bằng sông Hồng	2
Tìm hiểu dữ liệu theo 5 khu vực thuộc Thành Phố Hà Nội	2
Làm mịn và cắt lọc dữ liệu quan trọng	5
Chương 3: Hàm số tính toán chất lượng nước	5
1. Quy đổi dữ liệu	5
2. Hàm tính toán chất lượng nước	6
Chương 4: Công nghệ được sử dụng	7
Chương 5: Tổng kết	10
Tài liệu tham khảo	10

Chương 1: Giới thiệu

Trong quá trình học tập, lớp quyết định thực hiện một dự án về đánh giá chất lượng sống của thành phố Hà Nội, bao gồm nhiều tiêu chí như điện, nước, xử lý nước thải, rác thải, về các dịch vụ như cứu hỏa, cảnh sát, bệnh viện, trường học, ... Theo yêu cầu của giảng viên, nhóm 2 thực hiện phần việc tìm hiểu về chất lượng nước của thành phố Hà Nội, sau đó đưa dữ liệu lên một bản đồ tổng hợp.

Chương 2: Quá trình thu thập dữ liệu

Tìm dữ liệu khu vực đồng bằng sông Hồng

Dựa trên sự tương đồng về việc sử dụng nước giữa khu vực Đồng Bằng Sông Hồng và Thành Phố Hà Nội nên nhóm quyết định bắt đầu tìm kiếm dữ liệu ở khu vực này trước.

Doanh Nghiệp	Tỷ lệ dịch vụ(%)	Mức nước sinh hoạt tiêu thụ bình quân đầu người (l/người/ngày)	Số lần vỡ ống trên toàn hệ thống (lần/km)	Tính liên tục trong cấp nước (giờ/ngày)	Tỉ lệ mẫu nước thử đạt tiêu chuẩn (%)	Tỉ số giữa tổng số lần khiếu nại trên tổng số đầu nối (%)	Giá nước bình quân thực tế (VND / m3)	Tỷ lệ thu tiền nước (%)
Công ty nước sạch Hà Nội	100	82	0.11	24	99	null	7100	87
VIWACO	null	null	null	24	null	null	null	null
Công ty TNHH MTV Cấp nước Hà Đông	27	140	0.5	24	92	null	15000	87.78
Công ty CP Cấp nước Sơn Tây	50	100.8	27	23.5	100	null	6	100
Công ty CP Cấp thoát nước số 1 Vĩnh Phúc	62	null	null	24	null	null	null	null
Công ty CP nước sạch Vĩnh Phúc	75.3	94.97	3	22	100	8.36	9	95.24
Công ty TNHH MTV Cấp thoát nước Bắc Ninh	91.01	91.16	1	24	91.85	0.02	8.37	90.91
Nhà máy nước Từ Sơn	10.61	94.06	7	23.5	100	6.05	7	93.21
Công ty CP Nước Sạch Hà Nam	87.63	78.83	8	21	75	0.22	6	100

Dữ liệu nguồn nước theo các doanh nghiệp nước sạch năm 2015.

Dữ liệu được thu thập chủ yếu từ trang web <http://www.vnwd.vn>

Tìm hiểu dữ liệu theo 5 khu vực thuộc Thành Phố Hà Nội

Sau khi có một cái nhìn tổng quan về dữ liệu nguồn nước của khu vực, nhóm bắt tay vào tìm kiếm dữ liệu có giá trị hơn. Mỗi thành viên trong nhóm được phân chia tìm kiếm ở một khu vực cụ thể.

Thành Viên Nhóm	Phân chia công việc
Vũ Tùng Dương - 16020921 (Trưởng Nhóm)	Trung Tâm
Nguyễn Thế Chí Dũng - 16020906	Phía Bắc
Nguyễn Văn Dương - 16020919	Phía Nam
Trương Văn Dương - 16020081	Phía Tây
Lê Quang Đạo - 16020875	Phía Đông

Bảng phân chia khu vực cho từng cá nhân

Hà Nội có 30 quận, huyện. Vì phải tìm kiếm dữ liệu đến mức phường xã nên công việc đòi hỏi rất nhiều thời gian tìm kiếm dữ liệu.

Kết thúc giai đoạn này, nhóm đã tìm kiếm được thông tin về tất cả các nhà máy nước trên khu vực địa bàn Hà Nội và thông tin về nguồn nước như *nồng độ Asen nước ngầm trung bình, màu sắc, mùi vị, độ đục, độ Ph, nồng độ các chất khác...*

Nguồn nước (sông hay ngầm)	Nồng độ Asen nước ngầm trung bình	Màu sắc	Mùi vị	Độ đục	pH	Nitrit	Nitrat	Độ cứng	Clorua	Clo dư	Sunfat	Chỉ số Pecmanganat	Cd	Cr
ngầm	20.7	Không màu	Không mùi	null	null									
ngầm	20.7	Không màu	Không mùi	null	null									
ngầm	20.7	Không màu	Không mùi	null	null									
ngầm	20.7	Không màu	Không mùi	null	null									
ngầm	20.7	Không màu	Không mùi	null	null									
ngầm	20.7	Không màu	Không mùi	null	null									
ngầm	20.7	Không màu	Không mùi	null	null									
ngầm	20.7	Không màu	Không mùi	null	null									
ngầm	20.7	Không màu	Không mùi	null	null									
ngầm	20.7	Không màu	Không mùi	null	null									
ngầm	20.7	Không màu	Không mùi	null	null									
ngầm	20.7	Không màu	Không mùi	null	null									
ngầm	20.7	Không màu	Không mùi	null	null									
Ngầm	1.1	Không màu	Không mùi	<0.5	7.5	<100	3000	162000	20900	700	42800	1300	0.2	15.6
	1.1	Không màu	Không mùi											
	1.1	Không màu	Không mùi											
	1.1	Không màu	Không mùi											
	1.1	Không màu	Không mùi											
	1.1	Không màu	Không mùi											
	1.1	Không màu	Không mùi											
	1.1	Không màu	Không mùi											
Ngầm	1.1	Không màu	Không mùi	<0.5	7.3	<100	<1000	88000	31900	500	20900	1600	0.2	15.6
	1.1	Không màu	Không mùi											

Dữ liệu về nguồn nước khu vực phía Đông thuộc Thành Phố Hà Nội.

Quận	Năm 2012	Năm 2013	Năm 2014	Năm 2015	Năm 2017
Tây Hồ	3	0	1	8	8
Ba Đình	1	0	0	18	6
Hoàn Kiếm	1	0	0	6	7
Hai Bà Trưng	8	2	2	12	2
Thanh Xuân	null	null	null	null	null

Số lần dừng cấp nước khu vực trung tâm Thành Phố Hà Nội.

Giá bán nước		
Giá bán nước sạch cho đối tượng là hộ gia đình sử dụng vào mục đích sinh hoạt		
TT	Mức sử dụng nước sinh hoạt của hộ dân cư (m3/tháng/hộ)	Giá bán nước từ 01/10/2015 (đồng/m3)
1	Mức 10 m3 đầu tiên	5.973
2	Từ trên 10 m3 đến 20 m3	7.052
3	Từ trên 20 m3 đến 30m3	8.669
4	Trên 30 m3	15.929
Giá bán nước sạch cho các đối tượng là các cơ quan		
TT	Đối tượng sử dụng	Giá bán nước từ 01/10/2015 (đồng/m3)
1	Cơ quan hành chính sự nghiệp	9.955
2	Đơn vị sự nghiệp, dịch vụ công cộng	9.955
3	Sản xuất vật chất	11.615
4	Kinh doanh dịch vụ	22.068
Số liệu lấy từ công ty Nước Sạch Hà Nội		http://hawacom.vn

Giá bán nước khu vực trung tâm Thành Phố Hà Nội.

Làm mịn và chất lọc dữ liệu quan trọng

Dữ liệu thu được từ bước thứ 2 khá dồi dào tuy nhiên lại tạp nham và thiếu sự thống nhất. Nhóm đã thu gọn kho dữ liệu dựa trên những tính chất quan trọng quyết định đến chất

lượng nước sinh hoạt của người dân để làm đầu vào cho việc xây dựng Bản Đồ Chất Lượng Nước Thành Phố Hà Nội.

Dữ liệu mà nhóm thu thập được: [Kết quả thu thập dữ liệu](#)

Chương 3: Hàm số tính toán chất lượng nước

Dữ liệu về nguồn nước thu thập được ảnh hưởng rất lớn đến việc lựa chọn nhà cấp nước, địa bàn sinh sống và thói quen sinh hoạt.

1. Quy đổi dữ liệu

- Đơn vị : Trước tiên có sự khác nhau về đơn vị thu thập được, ví dụ như có các đơn vị mg/l, ppb, ug/l ta sẽ quy đổi hết về đơn vị ppb.
- Tính chất : Do tính chất các dữ liệu thu thập được một phần không cụ thể bằng số liệu mà chỉ phân ra các mức độ cao thấp nên sẽ phải quy đổi hết tất cả các dữ liệu đã có ra 3 mức là cao, trung bình và thấp.
- Phân loại : Mức độ cao thấp của các kim loại và gốc amoni sẽ phân theo bảng quy chuẩn nước ngầm sinh hoạt theo tiêu chuẩn xây dựng 233 năm 1999 :

TCXD 233 : 1999

Bảng 1. Phân loại chất lượng nguồn nước mặt – Giá trị giới hạn các thông số và nồng độ của các chất thành phần trong từng loại nước mặt

Số thứ tự	Các thông số	Đơn vị	Các loại nước		
			Loại A	Loại B	Loại C
1	2	3	4	5	6
1	Độ pH		6,5 đến 8,5	6,0 đến 9,0	pH > 9 và pH < 6
2	Độ đục	NTU	< 20	< 500	< 1.000
3	Độ màu	mg/l Pt	< 10	< 100	< 200
4	Độ oxy hóa KMnO ₄	mg/l O ₂	< 2,0	2 – 5	< 10
5	Độ cứng toàn phần	°dH	4 đến 8	< 4 hoặc 8 đến 13	< 28
6	Sulfua H ₂ S	mg/l	0	0	< 0,5
7	Clorua Cl ⁻	mg/l	< 25	< 200	< 400
8	Sunfat SO ₄ ²⁻	mg/l	< 25	< 250	< 400
9	Nitrit NO ₂ ⁻	mg/l	< 0,1	< 1	< 2
10	Nitrat NO ₃ ⁻	mg/l N	0	< 6	< 10
11	Photphat PO ₄ ³⁻	mg/l	0	< 1,5	< 2
12	Sắt tổng Fe	mg/l	< 0,3	< 1	< 2
13	Mangan tổng Mn	mg/l	< 0,2	< 0,5	< 1
14	Amonium NH ₄ ⁺	mg/l	< 0,2	< 0,5	< 1
15	Florua F ⁻	mg/l	0,5 đến 1,0	< 1,5	< 2
16	Xianua CN ⁻	µg/l	0	< 50	< 100
17	Phenol	µg/l	0	0,5	< 100
18	Asen As	µg/l	0	50	< 100
19	Cadmi Cd	µg/l	0	< 1	< 5
20	Crom tổng Cr	µg/l	0	< 10	< 50
21	Selen Se	µg/l	0	< 5	< 10
22	Thủy ngân Hg	µg/l	0	0	< 1
23	Đồng Cu	µg/l	< 50	< 1.000	< 3.000
24	Chì Pb	µg/l	0	< 10	< 50
25	Kẽm Zn	µg/l	< 50	< 1.000	< 5.000
26	E. Coli	MPN/100ml	< 20	< 100	< 200
27	Tổng hóa chất bảo vệ thực vật (trừ DDT)	mg/l	0	< 0,15	< 0,15
28	DDT	mg/l	0	< 0,01	< 0,01
29	Tổng hoạt độ phóng xạ α	Bq/l	< 0,1	< 0,1	< 0,1
30	Tổng hoạt độ phóng xạ β	Bq/l	< 1	< 1	< 1

Nguồn : [tcxd233-1999](#)

2. Hàm tính toán chất lượng nước

Đối với các mức độ phù hợp của mẫu nước với tiêu chuẩn sinh hoạt, nhóm sẽ tìm ra mức độ KHÔNG an toàn của nước đối với nhu cầu sinh hoạt của cư dân

Dựa vào bảng thống kê, ta có hàm tính toán chất lượng nước :

$$F = \frac{\sum_1^n X}{n}$$

Giải thích số liệu:

- F : giá trị chất lượng nước
- X : giá trị của tiêu chí về chất lượng nước
- n : số tiêu chí

Quy đổi giá trị với các mức tương ứng

- Thấp = 0
- Trung bình = 1
- Cao = 2

Quy đổi giá trị chất lượng nước

- $0 \leq F \leq 0.6$: Thấp
- $0.6 < F \leq 1.3$: Trung bình
- $1.3 < F \leq 2$: Cao

Chương 4: Công nghệ được sử dụng

Quá trình thu thập dữ liệu được thực hiện bằng việc thu thập thông tin trên mạng Internet. Sau đó sử dụng HTML + CSS + javascript, các thư viện leaflet và bản đồ việt nam (vnmap) để tạo nên bản đồ chất lượng nước Hà Nội. Leaflet là một thư viện hỗ trợ việc vẽ bản đồ, Vnmap là nguồn cung cấp các hình ảnh để tạo nên bản đồ, ở đây chúng em sử dụng http request thông qua api à vnmap cung cấp.

Bản đồ gồm 2 phần: bên trái là tên bản đồ, các thông tin chú thích, cùng với các tùy chọn tương tác cơ bản; bên phải là phần hiển thị bản đồ.

Bản đồ chất lượng nước Hà Nội

Thống kê trên bản đồ theo loại chất:

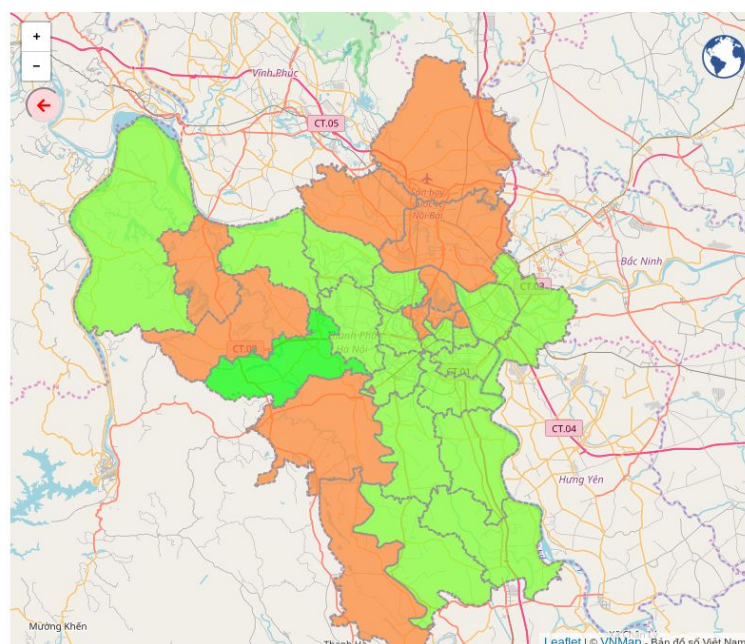
Tổng hợp

Hiện thị các trạm/công ty cấp nước

Ẩn các trạm cấp nước

Chú thích:

- Cao (trên 7 điểm)
- Trung bình (từ 5 đến 7 điểm)
- Thấp (dưới 5 điểm)



Tùy chọn
Bản đồ chất lượng diện
Bản đồ chất lượng nước
Quản lý nước thải
Quản lý rác thải
Trạm cứu hỏa
Đồn cảnh sát
Bệnh viện
Siêu thị
Công viên
Chợ
Trường mẫu giáo
Trường tiểu học
Trường THCS
Trường THPT
Đại học
Trạm xe bus
Quán karaoke
Rạp chiếu phim
Quán cà phê

Giao diện của bản đồ

Thống kê về mức độ theo từng loại chất trong nguồn nước, bao gồm 4 chất: Asen (As), Chì (Pb), Amoni (NH₄⁺), Magie (Mg).

Bản đồ chất lượng nước Hà Nội

Thống kê trên bản đồ theo loại chất:

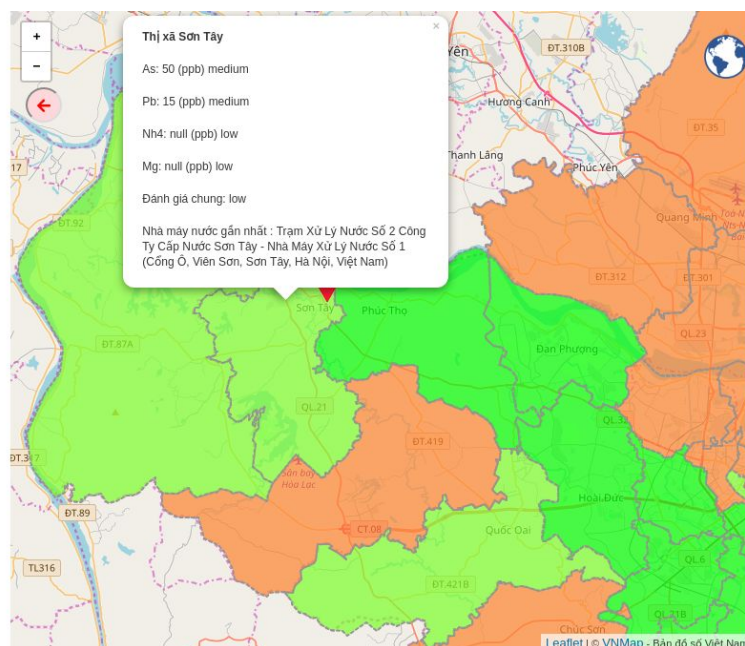
Asen

Hiện thị các trạm/công ty cấp nước

Ẩn các trạm cấp nước

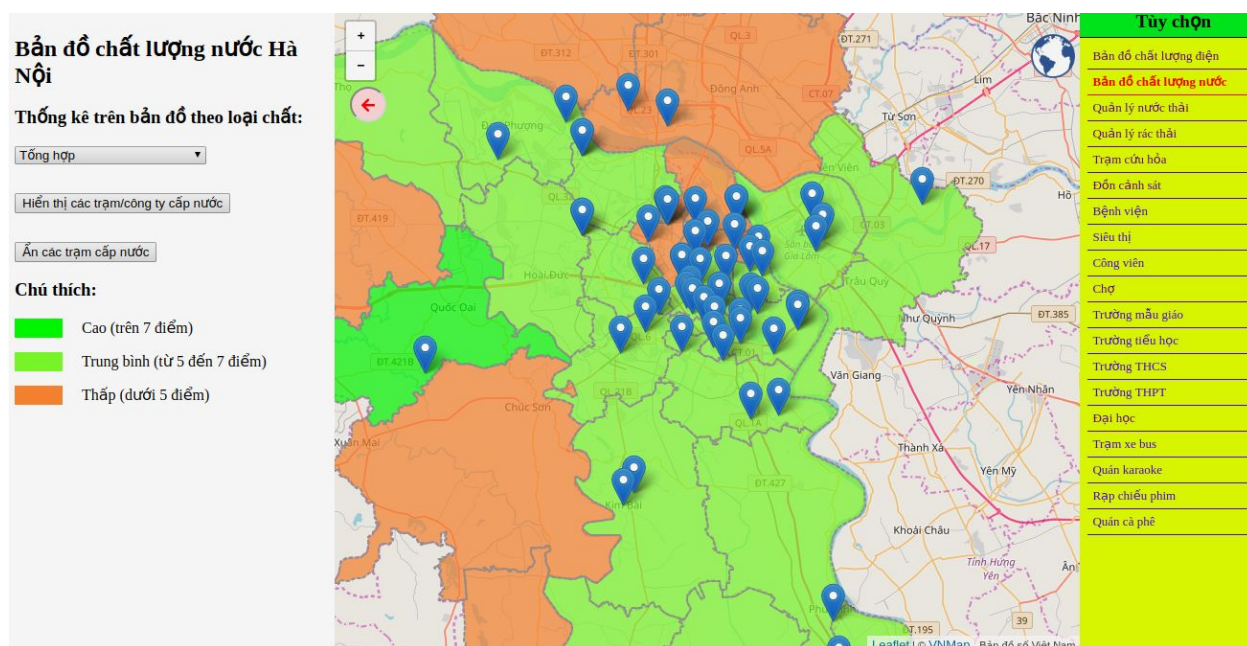
Chú thích:

- Cao (trên 7 điểm)
- Trung bình (từ 5 đến 7 điểm)
- Thấp (dưới 5 điểm)



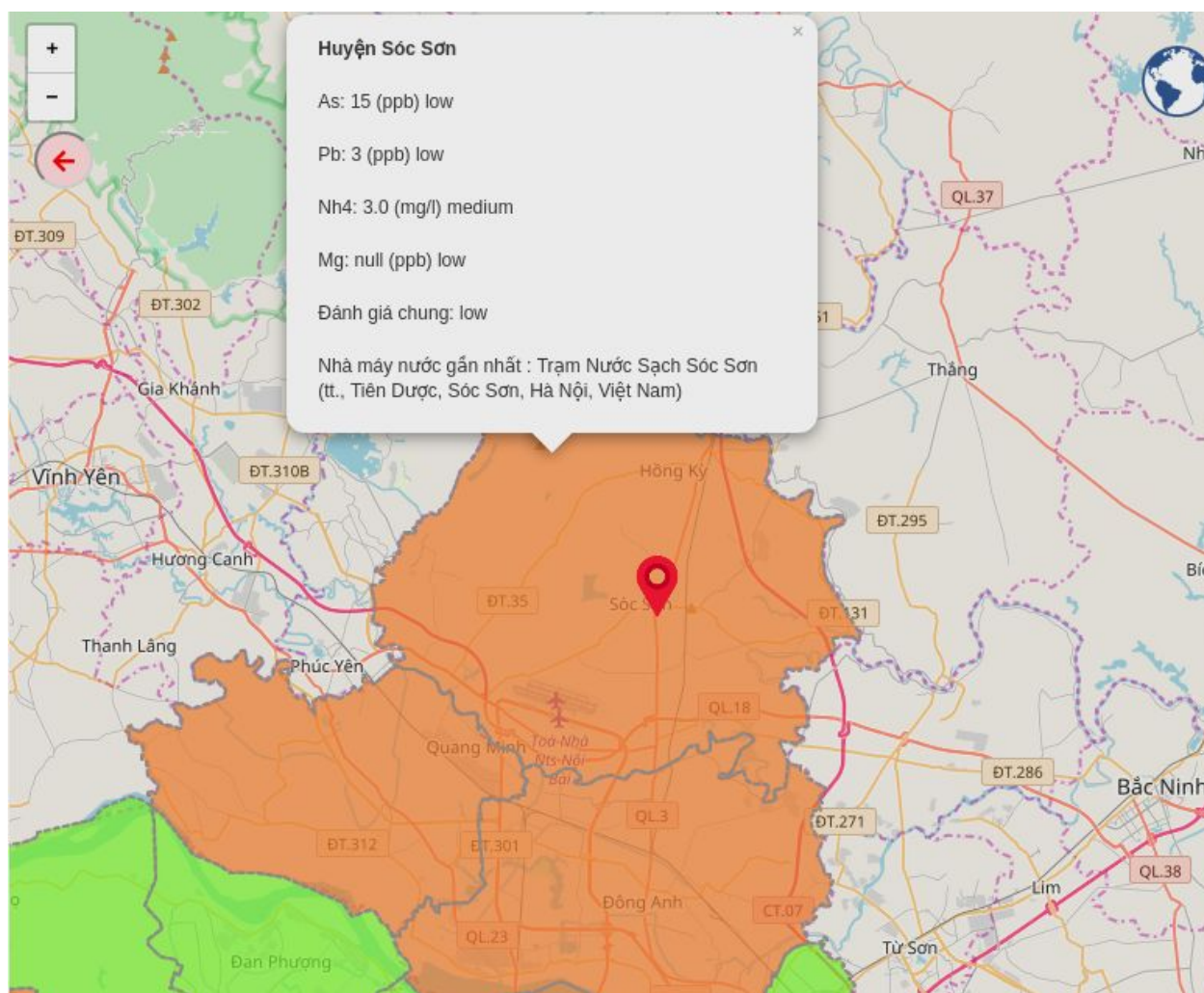
Tùy chọn
Bản đồ chất lượng diện
Bản đồ chất lượng nước
Quản lý nước thải
Quản lý rác thải
Trạm cứu hỏa
Đồn cảnh sát
Bệnh viện
Siêu thị
Công viên
Chợ
Trường mẫu giáo
Trường tiểu học
Trường THCS
Trường THPT
Đại học
Trạm xe bus
Quán karaoke
Rạp chiếu phim
Quán cà phê

Thống kê theo chỉ số Asen.



Hiện thị các trạm cấp nước. công ty cấp nước trên Hà Nội

Nếu chọn điểm bất kì trên bản đồ, sẽ hiển thị các thông tin về chất lượng nước tại vị trí đó và hiển thị trạm cấp nước gần vị trí đó nhất (các chỉ số chưa có số liệu cụ thể được để dạng null).



Chọn điểm bất kì trên bản đồ.

Chương 5: Tổng kết

Thông qua bài tập lớn, nhóm đã thành công làm các công việc được giao: thu thập dữ liệu nguồn nước ở Hà Nội, thiết kế một hàm số lượng hóa chất lượng nước, và đưa kết quả thu thập và tính toán hiển thị lên một bản đồ. Tất cả các quá trình trên giúp nhóm hiểu rõ hơn về quá trình ETL trong kho dữ liệu. Quá trình thu thập dữ liệu tương ứng với công đoạn Extract; thiết kế hàm số lượng hóa là nhằm mục đích chuyển đổi dữ liệu thu thập được thành một loại dữ liệu phân tích khác và làm sạch nó, tương ứng với quá trình Transform; đưa dữ liệu trên hiển thị lên bản đồ chính là quá trình Load, cả ba việc này cấu thành quá trình ETL cơ bản.

Tài liệu tham khảo

- [1] Dữ liệu về lịch dừng cấp nước của CÔNG TY TNHH MỘT THÀNH VIÊN NƯỚC SẠCH HÀ NỘI.
- [2] Dữ liệu về các nhà máy nước thuộc CÔNG TY TNHH MỘT THÀNH VIÊN NƯỚC SẠCH HÀ NỘI.
- [3] Dữ liệu về các nhà máy nước thuộc CÔNG TY CỔ PHẦN VIWACO.
- [4] Bản đồ mô phỏng mức độ ô nhiễm ASEN tại Hà Nội.