



KHOA CÔNG NGHỆ THÔNG TIN
TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN
227 Nguyễn Văn Cừ, Phường 4, Quận 5, TP.HCM
Điện Thoại: (08) 38.354.266 - Fax: (08) 38.350.096

Trực Quan Hóa Dữ Liệu

LAB TRỰC QUAN HÓA DỮ LIỆU VỚI TABLEAU

Mã: Lab2

Biên soạn:
Lê Ngọc Thành

1. Nội dung

Sử dụng Tableau để trực quan hóa dữ liệu.

2. Yêu cầu

Project được thực hiện theo nhóm. Thời gian và cách thức nộp, xem trên Moodle.

Nội dung cần nộp:

- Báo cáo trình bày trong file .doc/.docx/pdf chứa:
 - o Thông tin nhóm: tên nhóm, mssv...
 - o Mức độ hoàn thành tổng thể của mỗi yêu cầu.
 - o Mức độ hoàn thành của từng thành viên.
 - o Chi tiết các bước thực hiện (kèm hình ảnh), thuật toán, chạy ví dụ, nhận xét.
- Khuyến khích trình bày đơn giản, có hình minh họa.
- Source code được tích hợp (nếu có).
- Dataset được lấy gốc theo từng ngày, nếu có modify thì tạo file riêng.
- Ngôn ngữ lập trình cho code: Python
 - o Cho phép sử dụng các thư viện đã được giới thiệu trong lý thuyết.

3. Yêu cầu chi tiết

[Report coronavirus cases](#)

Now Yesterday		Search: <input type="text"/>											
All	Europe	North America	Asia	South America	Africa	Oceania							
#	Country, Other	Total Cases	New Cases	Total Deaths	New Deaths	Total Recovered	Active Cases	Serious, Critical	Tot Cases/ 1M pop	Deaths/ 1M pop	Total Tests	Tests/ 1M pop	Population
	World	4,885,970	+86,704	319,878	+3,358	1,903,201	2,662,891	44,754	627	41.0			
1	USA	1,549,359	+21,695	91,955	+977	354,340	1,103,064	16,857	4,684	278	12,255,697	37,052	330,769,370
2	Russia	290,678	+8,926	2,722	+91	70,209	217,747	2,300	1,992	19	7,147,014	48,977	145,927,122
3	Spain	278,188	+469	27,709	+59	196,958	53,521	1,152	5,950	593	3,037,840	64,977	46,752,654
4	Brazil	254,220	+13,140	16,792	+674	100,459	136,969	8,318	1,197	79	735,224	3,462	212,376,810
5	UK	246,406	+2,711	34,796	+160	N/A	N/A	1,559	3,632	513	2,682,716	39,543	67,843,268
6	Italy	225,886	+451	32,007	+99	127,326	66,553	749	3,735	529	3,041,366	50,294	60,472,166
7	France	179,927	+358	28,239	+131	61,728	89,960	1,998	2,757	433	1,384,633	21,218	65,256,433
8	Germany	177,289	+638	8,123	+74	154,600	14,566	1,133	2,117	97	3,147,771	37,584	83,752,125
9	Turkey	150,593	+1,158	4,171	+31	111,577	34,845	903	1,788	50	1,650,135	19,591	84,227,597
10	Iran	122,492	+2,294	7,057	+69	95,661	19,774	2,294	1,461	84	701,640	8,367	83,859,705
11	India	100,340	+4,642	3,156	+131	39,233	57,951		73	2	2,302,792	1,671	1,378,344,732
12	Peru	94,933	+2,660	2,789	+141	30,306	61,838	866	2,884	85	661,132	20,086	32,914,644

Hình 1. Dữ liệu thống kê từng ngày ca nhiễm virus Covid-19 từ tổ chức Worldometer

Đây là bài tiếp nối bài lab01 mà phần trước chúng ta đã thực hiện việc lấy dữ liệu, tìm hiểu một số quan hệ trên đó. Trong lab này, chúng ta vận dụng Tableau để việc trực quan hóa dữ liệu trở nên tốt hơn.

Cụ thể trong lab này, nhóm được yêu cầu thực hiện các nhiệm vụ sau:

- Lý thuyết: tìm hiểu công cụ Tableau
 - o Giới thiệu về Tableau

- Các tính năng hỗ trợ của Tableau kèm các ví dụ minh họa. Các ví dụ được thực hiện trên các mẫu dataset có sẵn, không liên quan đến dataset chính của bài.
- Thực hành: Vận dụng Tableau để trực quan hóa dữ liệu Woldometer
 - Nhóm sinh viên (NSV) dựa trên bài đã thực hiện ở phần trước để thực hiện tiếp.
 - NSV cũng quan sát dữ liệu từ đơn giản đến phức tạp, từ thuộc tính đơn đến kết hợp thuộc tính, từ quan hệ độc lập đến quan hệ phụ thuộc, ...
 - Chọn lựa nhiều dạng biểu đồ khác nhau, đánh giá sự phù hợp, có thể dùng lại các lý luận ở phần trước. Thể hiện các biểu đồ này trong Tableau.
 - Sử dụng màu sắc để thể hiện dữ liệu, giải thích ý nghĩa các màu và tại sao mình sử dụng màu như vậy. Sau khi bổ sung màu, NSV có rút ra thêm ý nghĩa gì không?
 - Thể hiện trực quan một số dữ liệu biến đổi qua từng ngày. Rút ra ý nghĩa.
 - Sử dụng các kỹ thuật được giới thiệu trong bài Manipulate View, Facet, Reduce, Embed để trình diễn trên Tableau với dữ liệu Woldometer. Giải thích việc chọn lựa và ý nghĩa mang lại.
 - Chạy một số thuật toán học máy đơn giản để hiểu thêm về dữ liệu, thuật toán học máy được quyền sử dụng code có sẵn, ghi rõ nguồn gốc.

4. Những giới hạn

- Bài lab này được giới hạn trong môi trường Tableau có tích hợp thêm thư viện/code để minh họa.
- Một số thư viện như numpy, pandas, seaborn, matplotlib có thể được sử dụng. Các thư viện khác trừ thư viện học máy muốn sử dụng cần phải hỏi ý kiến của giáo viên thực hành.

5. Đánh giá

- Các tiêu chí đánh giá:
 1. Tìm hiểu về Tableau (20%)
 2. Áp dụng Tableau để trực quan dữ liệu (50%)
 3. Rút ra ý nghĩa hợp lý sau mỗi dữ liệu được trực quan (10%)
 4. Báo cáo trình bày bố cục và định dạng hợp lý, rõ ràng (20%)

Lưu ý: nếu số biểu diễn trực quan sử dụng Tableau quá ít thì sẽ xem xét giảm tỉ lệ ở mức 2.

6. Quy định

- Bài không có báo cáo sẽ không chấm.
- Thành viên không tham gia sẽ không có điểm.
- Các nguồn tài liệu tham khảo (nếu có) cần ghi đầy đủ trong báo cáo ở mục *Tài liệu tham khảo*. Lưu ý cần phân biệt giữa tham khảo và đạo văn.
- Đặt tên thứ mục bài làm là MSSV1_MSSV2_MSSV3_..._Lab01, với MSSV là mã số sinh viên, nén toàn bộ bài nộp thành 1 tập tin trước khi nộp. Nếu kích thước >10MB thì upload lên server ngoài như Google Drive, ..., nộp link và giữ link public ít nhất trong 2 năm.
- **Bài giống nhau sẽ 0 điểm môn học.**

7. Liên hệ

Mọi thắc mắc trong quá trình thực hiện vui lòng gửi mail về lnthanh@fit.hcmus.edu.vn