ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH

TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN

KHOA CÔNG NGHỆ THÔNG TIN

🙧🙧🙧✯🙥🙥🙥

**BÁO CÁO**

**TRỰC QUAN HOÁ DỮ LIỆU**

**LAB 02: TRỰC QUAN HOÁ DỮ LIỆU VỚI TABLEAU**

Nhóm 5:

1. 1712117 – Nguyễn Huỳnh Thảo Nhi

2. 1712710 – Lê Quang Quí

3. 1712713 – Lê Bá Quyền

4. 1712775 – Nguyễn Lê Trường Thành

**GVHD: Th.S Lê Ngọc Thành**

**Lớp: Trực quan hoá dữ liệu CQ2017/1**

*TP.HCM, 2019-2020*

Mục Lục

[I. Giới thiệu đồ án: 3](#_Toc43764822)

[II. Phân công công việc: 3](#_Toc43764823)

[III. Môi trường cài đặt: 4](#_Toc43764824)

[IV. Tìm hiểu công cụ Tableau: 4](#_Toc43764825)

[**1. Giới thiệu về Tableau:** 4](#_Toc43764826)

[**2. Các tính năng hỗ trợ của Tableau:** 4](#_Toc43764827)

[V. Trực quan dữ liệu với Tableau: 4](#_Toc43764828)

[**1. Trực quan một số loại biểu đồ với Tableau:** 4](#_Toc43764829)

[**a.** 4](#_Toc43764830)

[**b.** 4](#_Toc43764831)

[**c. Heat map:** 4](#_Toc43764832)

[VI. Áp dụng một số thuật toán máy học: 5](#_Toc43764833)

[**1. PCA:** 5](#_Toc43764834)

[**2. K-means Clustering:** 6](#_Toc43764835)

[**3. Linear Regression:** 6](#_Toc43764836)

[VII. Tham khảo: 6](#_Toc43764837)

# **I. Giới thiệu đồ án:**

* 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:
    - 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.
* 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 đối 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.

# **II. Phân công công việc:**

|  |  |  |  |
| --- | --- | --- | --- |
| **STT** | **Công việc** | **Thành viên thực hiện** | **Ghi chú** |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

# **III. Môi trường cài đặt:**

# **IV. Tìm hiểu công cụ Tableau:**

## **1. Giới thiệu về Tableau:**

* **Tableau** là phần mềm hỗ trợ phân tích và trực quan hóa dữ liệu (***Data Visualization),*** phát triển mạnh mẽ và được dùng nhiều trong ngành BI (***Business Intelligence).*** Nó giúp tổng hợp các dữ liệu dạng thô và chuyển đổi những liệu này từ các dãy số thành những hình ảnh, biểu đồ trực quan.
* Phân tích dữ liệu với **Tableau** rất nhanh chóng và hiệu quả, các hình ảnh được tạo ra ở dạng ***Dashboard*** (Bảng điều khiển) và ***WorkSheet*** (Bảng tính). Dữ liệu được tạo bằng **Tableau** có thể được hiểu bởi chuyên gia ở mọi cấp độ trong một tổ chức. Nó thậm chí còn cho phép người dùng không có kỹ thuật tạo bảng điều khiển (***Dashboard***) tùy chỉnh.
* Mục tiêu của **Tableau** rất đơn giản: "Help people see and understand their data".

## **2. Các tính năng hỗ trợ của Tableau:**

# **V. Trực quan dữ liệu với Tableau:**

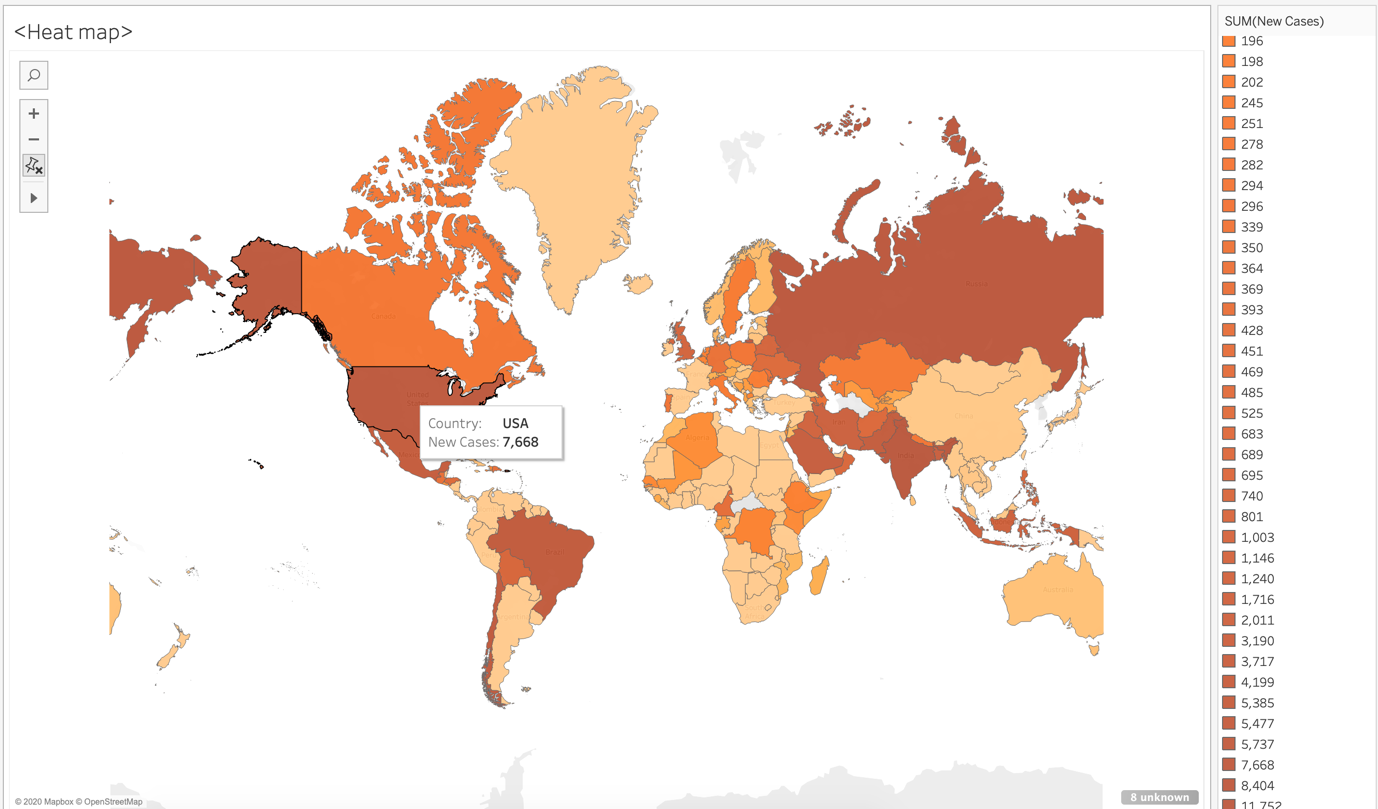
## **1. Trực quan một số loại biểu đồ với Tableau:**

### **a.**

### **b.**

### **c. Global Heatmap:**

* Global Heatmap là …
* Để trực quan biểu đồ trên, nhóm lựa chọn hai trường dữ liệu là Country và New Cases để thể hiện thông tin số lượng ca nhiễm bệnh mới của tất cả các quốc gia trên thế giới.
* Ngoài ra, với lựa chọn màu sắc tông cam từ nhạt đến đậm thể hiện cho mức độ số lượng ca nhiễm. Mức độ màu cam càng đậm chứng tỏ cho quốc gia có số ca nhiễm mới càng lớn, ngược lại, khi sắc cam càng nhạt, chứng tỏ quốc gia có số ca nhiễm càng thấp.



* + - * Với biểu đồ trên, ta có thể quan sát được số lượng ca nhiễm ứng với từng quốc gia bằng cách di chuyển chuột vào vùng quốc gia tương ứng.
      * Như hình trên, ta có thể thấy Mỹ đang có 7668 ca nhiễm bệnh mới.
      * Qua biểu đồ trên, ta có thể nhìn được khái quát toàn cảnh tình trạng nhiễm bệnh mới của các quốc gia trên toàn thế giới. Biết được quốc gia nào vẫn đang trong tầm nguy hiểm, quốc gia nào không còn ca nhiễm mới. Từ đó, có thể tổng kết được tình trạng dịch bệnh đang diễn tiến thế nào.

d.

e.

f.

**2. Thể hiện trực quan một số dữ liệu biến đối qua từng ngày. Rút ra ý nghĩa.**

**3. Sử dụng các kỹ thuật được giới thiệu trong bài Manipulate View, Facet, Reduce,**

**Embed để trình diễn trên Tableau với dữ liệu Woldometer.**

# **VI. Áp dụng một số thuật toán máy học:**

## **1. PCA:**

* Đây là thuật toán sinh ra để giải quyết vấn đề dữ liệu có quá nhiều chiều dữ liệu, cần giảm bớt chiều dữ liệu nhằm tăng tốc độ xử lí, nhưng vẫn giữ lại thông tin nhiều nhất có thể (high variance).

## **2. K-means Clustering:**

## **3. Linear Regression:**

# **VII. Tham khảo:**

[1]. https://viblo.asia/p/gioi-thieu-principal-component-analysis-07LKXpq2KV4