ĐẠI HỌC QUỐC GIA TP HỒ CHÍ MINH

TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN

KHOA CÔNG NGHỆ THÔNG TIN

BỘ MÔN HỆ THỐNG THÔNG TIN

**BÁO CÁO ĐỒ ÁN**

**TRỰC QUAN HÓA DỮ LIỆU**

**Đề tài :**

**CÔNG CỤ TRỰC QUAN HÓA**

**TÌNH TRẠNG TAI NẠN GIAO THÔNG TẠI BANG CALIFORNIA**

**GIẢNG VIÊN HƯỚNG DẪN :** TS. Lê Thị Nhàn

**HỌC VIÊN THỰC HIỆN :**

20C12007 – Trần Đình Lâm

20C12030 – Huỳnh Lâm Phú Sĩ

20C12031 – Huỳnh Ngọc Đan Thanh

**KHÓA :** K30

TP.HCM, 16 tháng 10 năm 2021

**Mục Lục**

[**Mục Lục** 1](#_Toc85304376)

[**Danh sách các bảng** 2](#_Toc85304377)

[**Danh sách các hình** 2](#_Toc85304378)

[**Thông tin chi tiết nhóm** 3](#_Toc85304379)

[**1.** **Giới thiệu nghiệp vụ** 4](#_Toc85304380)

[**1.1.** **Chức năng và đối tượng của công cụ** 4](#_Toc85304381)

[**1.2.** **Minh họa công cụ** 4](#_Toc85304382)

[**2.** **Các công trình liên quan** 5](#_Toc85304383)

[**2.1.** **Công cụ TIMS** [1] 5](#_Toc85304384)

[**2.2.** **Công cụ SAVE-T** [3] 6](#_Toc85304385)

[**3.** **Tổng quan và phân tích dữ liệu** 7](#_Toc85304386)

[**3.1.** **Tổng quan dữ liệu** 7](#_Toc85304387)

[**3.3.** **Phân tích các yếu tố con người** 8](#_Toc85304388)

[**3.4.** **Phân tích các yêu tố phương tiện** 10](#_Toc85304389)

[**3.5.** **Phân tích các yếu tố ngoại cảnh** 12](#_Toc85304390)

[**4.** **Kế hoạch thực hiện** 14](#_Toc85304391)

[References 15](#_Toc85304392)

# **Danh sách các bảng**

[Bảng 1: Thông tin chi tiết nhóm 3](#_Toc85304231)

[Bảng 2: So sánh các công trình liên quan 7](#_Toc85304232)

[Bảng 3: Kế hoạch thực hiện 14](#_Toc85304233)

# **Danh sách các hình**

[Hình 1: Minh họa giao diện của công cụ 5](file:///C:\Users\LAP11777-local\Documents\GitHub\Nhom1DataVisualization\ReportDV_GiuaKy.docx#_Toc85304223)

[Hình 2: Tương quan nhóm tuổi 8](#_Toc85304224)

[Hình 3: Mô phỏng tương quan về độ tỉnh táo 9](file:///C:\Users\LAP11777-local\Documents\GitHub\Nhom1DataVisualization\ReportDV_GiuaKy.docx#_Toc85304225)

[Hình 4: Một trực quan hóa các loại phương tiện khác nhau 10](file:///C:\Users\LAP11777-local\Documents\GitHub\Nhom1DataVisualization\ReportDV_GiuaKy.docx#_Toc85304226)

[Hình 5: Biểu đồ số người bị thương 11](#_Toc85304227)

[Hình 6: Một biểu đồ số người tử vong 12](file:///C:\Users\LAP11777-local\Documents\GitHub\Nhom1DataVisualization\ReportDV_GiuaKy.docx#_Toc85304228)

[Hình 7: Phân bố số vụ theo giờ 13](file:///C:\Users\LAP11777-local\Documents\GitHub\Nhom1DataVisualization\ReportDV_GiuaKy.docx#_Toc85304229)

[Hình 8: Phân bố số vụ tai nạn theo ngày 13](file:///C:\Users\LAP11777-local\Documents\GitHub\Nhom1DataVisualization\ReportDV_GiuaKy.docx#_Toc85304230)

# 

# **Thông tin chi tiết nhóm**

|  |  |  |  |
| --- | --- | --- | --- |
| **BẢNG PHÂN CÔNG & ĐÁNH GIÁ HOÀN THÀNH CÔNG VIỆC** | | | |
| **Người**  **thực hiện** | **Công việc thực hiện** | **Mức độ**  **hoàn thành** | **Đánh giá của nhóm** |
| 20C12007  Trần Đình Lâm | Lên kế hoạch và phân công nhiệm vụ |  |  |
| Tìm kiếm & So sánh các công trình liên quan |  |
| Tiền xử lý dữ liệu |  |
| Phân tích & trực quan hóa theo yếu tố ngoại cảnh |  |
| Soạn thảo slide & báo cáo |  |
| 20C12030  Huỳnh Lâm  Phú Sĩ | Tìm dữ liệu SWITRS cho bang California |  |  |
| Phân tích & trực quan hóa theo yếu tố phương tiện |  |
| Thao tác truy vấn trên bản đồ |  |
| Soạn thảo slide & báo cáo |  |
| 20C12031  Huỳnh Ngọc  Đan Thanh | Tiền xử lý dữ liệu |  |  |
| Phân tích & trực quan hóa theo yếu tố con người |  |
| Thao tác truy vấn trên bản đồ |  |
| Soạn thảo slide & báo cáo |  |

Bảng 1: Thông tin chi tiết nhóm

1. **Giới thiệu nghiệp vụ**

Ngày nay, một trong những nguyên nhân phổ biến gây nguy hại đến sự an toàn của con người là tai nạn khi tham gia giao thông. Điều này đặc biệt phổ biến ở các thành phố, đô thị lớn và là vấn đề nan giải tại đối với các nhà quản lý để đảm bảo an toàn cho người dân. Vì vậy ở bang California và một số bang khác của Mỹ đã tiến hành thu thập dữ liệu các vụ tai nạn, vụ va chạm xe ở nhiều mức độ khác nhau, nhằm mục đích xây dựng các công cụ hỗ trợ truy vấn, phân tích các vấn đề liên quan nhằm đưa ra các quyết định phù hợp trong quản lý giao thông đô thị.

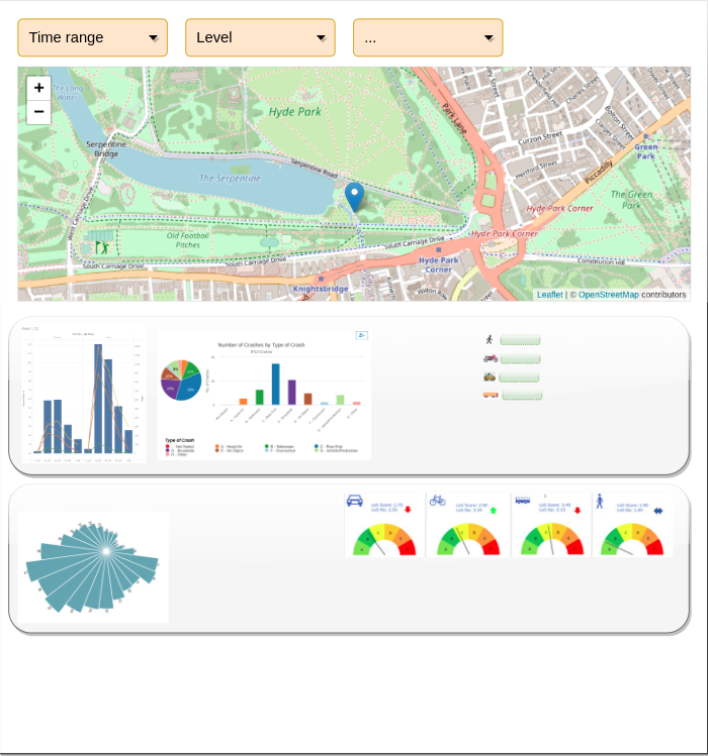
1. **Chức năng và đối tượng của công cụ**

Bằng cách sử dụng công cụ phân tích dữ liệu Tableu, nhóm đang tiến hành xây dựng một công cụ tổng hợp và trực quan các thông tin liên quan nhất, nhằm cố gắng đánh giá được mức độ an toàn cục bộ ở một khu vực hoặc một đoạn đường cụ thể ở bang California. Đối tượng sử dụng công cụ có thể là người dùng phổ thông,

1. **Minh họa công cụ**

Phác thảo các thành phần chính của công cụ được thể hiện ở giao diện minh họa **Error! Not a valid bookmark self-reference.**. Cụ thể ý tưởng vận hành cơ bản như sau:

1. User tìm chọn một phân vùng nhỏ được định sẵn trên bản đồ. Ở mức độ nâng cao, user tùy chọn phân vùng hoặc đoạn đường cụ thế trên bản đồ.
2. User chọn các tiêu chí về thời gian, mức độ, loại phương tiện,…
3. Hệ thống dashboard hiển thị giao diện các biểu đồ liên quan.
4. Hiển thị đánh giá **mức độ an toàn** dựa trên thông tin cục bộ về phân đoạn khu vực đã chọn, so sánh với thống kê toàn cục của toàn khu.



Hình 1: Minh họa giao diện của công cụ

1. **Các công trình liên quan**
2. **Công cụ TIMS** [1]

Công cụ TIMS là một website tổ hợp của nhiều tính năng và truy vấn cực kỳ đa dạng, phát triển bởi Trung tâm nghiên cứu SafeTREC, Đại học California Berkeley.

#### **Chức năng:**

Truy vấn và trực quan hóa hầu hết các thông tin có được trong bộ dữ liệu SWITRS, trong đó nhiều chức năng tập trung vào mô tả dữ liệu trên bản đồ.

#### **Các thư viện được sử dụng:**

Bootstrap, JQuery, Google Maps APIs, Dojo Toolkit, Google Charts

#### **Đối tượng sử dụng:**

Tập trung phát triển cho đội ngũ nghiên cứ và các cơ quan quản lý.

#### **Dữ liệu sử dụng:**

California Statewide Integrated Traffic Records System (SWITRS) [2]

1. **Công cụ SAVE-T** [3]

Công cụ SAVE-T (Safety Analysis Visualization and Evaluation Tool) là một công cụ web hỗ trợ truy vấn và tạo báo cáo tự động, trong đó có chèn các yếu tố hình ảnh, bảng biểu để trực quan hóa thông tin giao thông trên một số đường cao tốc ở New Jersey.

#### **Chức năng:**

Truy vấn dữ liệu và tạo báo cáo tự động về tình hình giao thông. Ngoài ra web còn có chức năng hiển thị truy vấn theo phân đoạn đường cao tốc cho sẵn.

#### **Các thư viện được sử dụng:**

D3.js để vẽ báo cáo và biểu đồ

Leaflet.js để trực quan và tương tác với bản đồ

#### **Đối tượng sử dụng:**

Các cơ quan quản lý giao thông đường cao tốc

#### **Dữ liệu sử dụng:**

New Jersey Department of Transportation (NJDOT)​

Tổng hợp lại, Bảng 2 so sánh giữa các công trình liên quan và sản phẩm mà nhóm đề xuất:

|  |  |  |
| --- | --- | --- |
| Công cụ TIMS [1] | Công cụ SAVE-T [3] | Công cụ nhóm đề xuất |
| Trực quan hóa phong phú, đa dạng;  Có tương tác​;  Dùng bản đồ để mô tả​ | Chủ yếu là báo cáo, ít hình ảnh​;  Có truy vấn trên bản đồ;  Tùy chỉnh truy vấn​ | Tạo các trực quan hóa cục bộ;  Có truy vấn trên bản đồ;  Tương tác được trên hình​ |
| Google Map,​ Google Chart, ​ ArcGIS, jQuery, Bootstrap, Dojo | D3.js​  Leaflet.js ​ | Tableu  ​Python​ |
| Dùng trong nghiên cứu​ | ​ Hỗ trợ chuyên gia,  quản lý giao thông | Hỗ trợ người dùng  không chuyên​ |
| California Traffic Collision Data from SWITRS​ | ​ New Jersey Department of Transportation (NJDOT) | California Traffic Collision Data from SWITRS​ |

Bảng 2: So sánh các công trình liên quan

1. **Tổng quan và phân tích dữ liệu**
2. **Tổng quan dữ liệu**

Cơ sở dữ liệu về các vụ va chạm giao thông ở bang California, lấy từ Hệ thống tích hợp giao thông trên toàn tiểu bang. ​Dữ liệu từ 01-2000 đến 06-2021.Cụ thể dữ liệu gồm 3 bảng:​

* Collision: Chứa thông tin về vụ va chạm như ngày giờ, nơi nó xảy ra (kinh độ, vĩ độ), thời tiết tại thời điểm xảy ra tại nạn, tình trạng mặt đường, những phương tiện có liên quan đến vụ tai nạn.​
* Party: Chứa thông tin về những người có liên quan đến vụ va chạm bao gồm tuổi tác, giới tính và mức độ tỉnh táo (rượu bia/ thuốc …).​
* Victim: Chứa thông tin về thương tích của nạn nhân trong vụ va chạm, vị trí ngồi của nạn nhâ, tình trạng thiết bị an toàn trên xe.

1. **Phân tích các yếu tố con người**

Phác họa sơ bộ thông tin về các yếu tố sinh học của người tham gia vào tai nạn và mức độ thương tật.Trực quan hóa hiện tại được xây dựng để tìm hiểu về bộ dữ liệu, chưa phải đứng trên khái cạnh người dùng của sản phẩm.

* **Chart, bar chart, histogram

  Description automatically generated**Tương quan về nhóm tuổi, giới tính đối với mức độ tỉnh táo của người tham gia vụ tai nạn. Cụ thể ở đây tập trung vào đối tượng là tài xế lái xe.

Hình 2: Tương quan nhóm tuổi

* Chart, histogram

  Description automatically generatedTương quan về độ tỉnh táo của người tham gia trong vụ tai nạn với mức độ thương tật của họ

Hình 3: Mô phỏng tương quan về độ tỉnh táo

***Nhận xét:***

* + Thông tin thời gian (ngày, tháng, năm) không ảnh hưởng đến yếu tố sinh học.
  + Nam giới lái xe nhiều hơn nữ và tỉ lệ gây tai nạn cũng cao hơn nhiều lần so với nữ giới.
  + Trong các vụ tai nạn, tỉ lệ người có sử dụng rượu bia/ thuốc bị chết nhiều hơn so với các mức độ thương tật còn lại.

**Đánh giá:**

* Thông tin về giờ có thể ảnh hưởng đến yếu tố rượu bia/ thuốc. Mọi người thường dùng rượu bia trước khi tham gia giao thông vào buổi tối trong ngày.
* Các trực quan hóa đang thể hiện trên đối tượng là 1 người trong vụ tai nạn nên chưa thể hiện rõ mức độ nghiêm trọng của vụ tai nạn. Thay đổi đối tượng cụ thể là 1 vụ tai nạn để đánh giá.

**Mục tiêu:**

* + Trả lời được các yếu tố sinh học của con người như tuổi, giới tính, mức độ tỉnh táo ảnh hưởng như thế nào (lượng nạn nhân, mức độ tổn thương của nạn nhân) đến mức độ nghiêm trọng của vụ tai nạn.

1. **Phân tích các yêu tố phương tiện**

Chart, bar chart

Description automatically generated**3.3.1. Các loại phương tiện khác nhau**

Hình 4: Một trực quan hóa các loại phương tiện khác nhau

Nhận xét về biểu đồ: Xe tải lớn gây tai nạn nhiều nhất trong các loại phương tiện giao thông khác nhau. Các phương tiện còn lại có mức độ tai nạn ít hơn nhiều so với xe tải lớn. Trong đó xe đạp có mức độ tai nạn tương đối thấp trong cả bốn loại phương tiện

**3.3.2. Số người bị thương trong tai nạn**

Nhận xét: Cả ba loại phương tiện đều có mức độ bị thương khi gặp tai nạn tương đối giống nhau.

Chart, bar chart, histogram

Description automatically generated

Hình 5: Biểu đồ số người bị thương

**3.3.3. Số người chết trong tai nạn**

Nhận xét: số ngừoi tử vong khi đi bộ chiếm rất cao, chứng tỏ đi bộ mặ dù không có tai nạn nhiều nhưng số lượng ngừoi đi bộ tử vong khi xảy ra tai nạn cao. Trong đó di chuyển bằng xe đạp có ít ca tử vong được Chart, bar chart

Description automatically generatedghi nhận nhất.

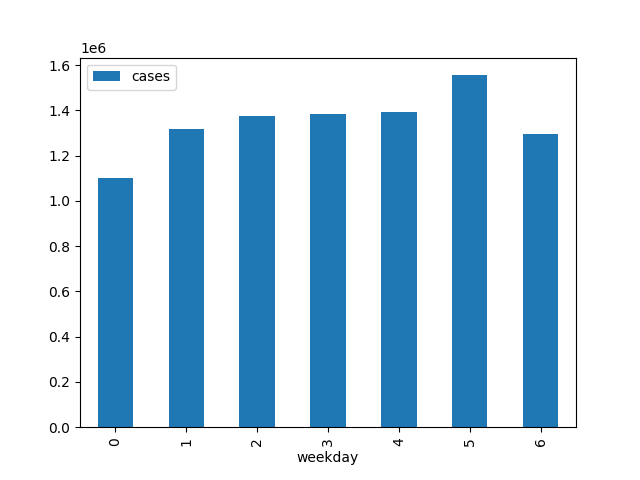
Hình 6: Một biểu đồ số người tử vong

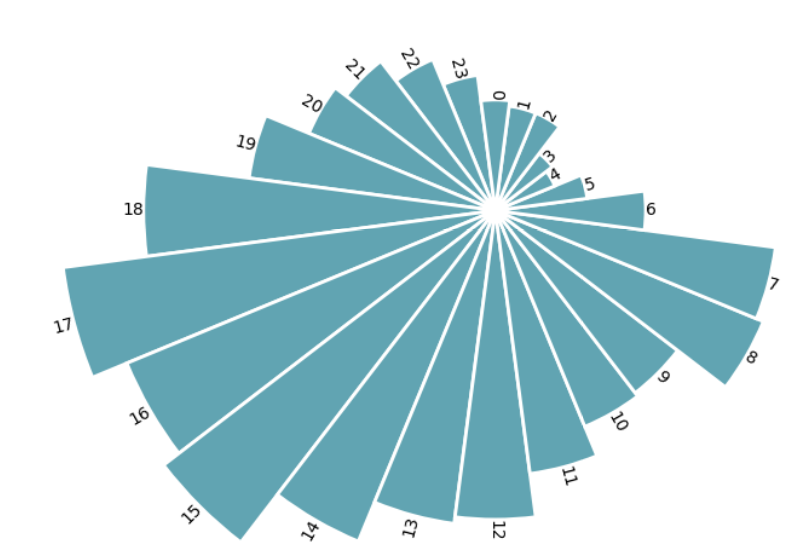
**3.3.4. Kết luận chung:**

Kết luận chung: Tỉ lệ người tử vong trong các vụ tai nạn giao thông ở ban California khá thấp, tuy nhiên số người tử vong khi đi bộ lại cao, do đó cần phải có thêm các biện pháp bảo vệ hoặc cảnh báo cho người đi bộ khi tham gia giao thông. Đối với xe tải cần có các biện pháp an toàn cao hơn, hệ thống làm ra cần phải cảnh báo được nguy cơ trên các đoạn đường có xảy ra tai nạn do xe tải lớn gây ra.

1. **Phân tích các yếu tố ngoại cảnh**

Dữ liệu về các yêu tố ngoại cảnh có thể được chia làm nhiều nhóm nhỏ tương đối riêng biệt nhau bao gồm: Thời gian xảy ra sự kiện, địa điểm, thời tiết, môi trường và chất lượng đường sá. Dựa vào một số yếu tố này, nhóm có thể tiến hành phân tích khám phá các mối tương quan để trả lời một số câu hỏi cụ thể như:

* Số lượng tai nạn tăng như thế nào theo năm? Theo tháng quý?
* Phân bố số lượng tai nạn theo các quận và tình hình dân cư ra sao?
* Điểu kiện nắng, mây, mưa có tác động đến khả năng gây tai nạn?

****

Hình 7: Phân bố số vụ theo giờ

Hình 8: Phân bố số vụ tai nạn theo ngày

1. **Kế hoạch thực hiện**

|  |  |  |
| --- | --- | --- |
| **Mô tả​** | **Thành viên​** | **Thời gian​** |
| Tiền xử lý dữ liệu​ | Lâm​  Thanh​ | 1 tuần (18/10-24/10)​ |
| Trích xuất truy vấn từ bản đồ​  Thao tác trên bản đồ​ | Thanh​  Sĩ​ | 1 tuần (25/10-31/10)​ |
| Trực quan hóa các biểu đồ theo thứ tự ưu tiên ​ | Lâm​  Sĩ​ | 1 tuần (01/11-07/11)​  ​ |
| Tinh chỉnh & cải tiến hiển thị​ | Lâm​  Sĩ​ | 1 tuần (08/11-14/11)​  ​ |
| Soạn thảo báo cáo ​  Slide trình bày​ | Thanh​  Sĩ​ | 1 tuần (15/11-20/11)​ |

Bảng 3: Kế hoạch thực hiện

# References

|  |  |
| --- | --- |
| [1] | Yuan Zhu, Sami Demiroluk, Kaan Ozbay, Kun Xie, Hong Yang, Di Sha, "SAVE-T: Safety Analysis Visualization and Evaluation Tool," *Journal of Advanced Transportation,* vol. 2021, 2021. |
| [2] | U. o. C. B. 2. afe Transportation Research and Education Center, "Transportation Injury Mapping System (TIMS)," University of California, Berkeley, 2021. [Online]. |
| [3] | "California Traffic Collision Data from SWITRS," 2021. [Online]. Available: http://iswitrs.chp.ca.gov/ . |