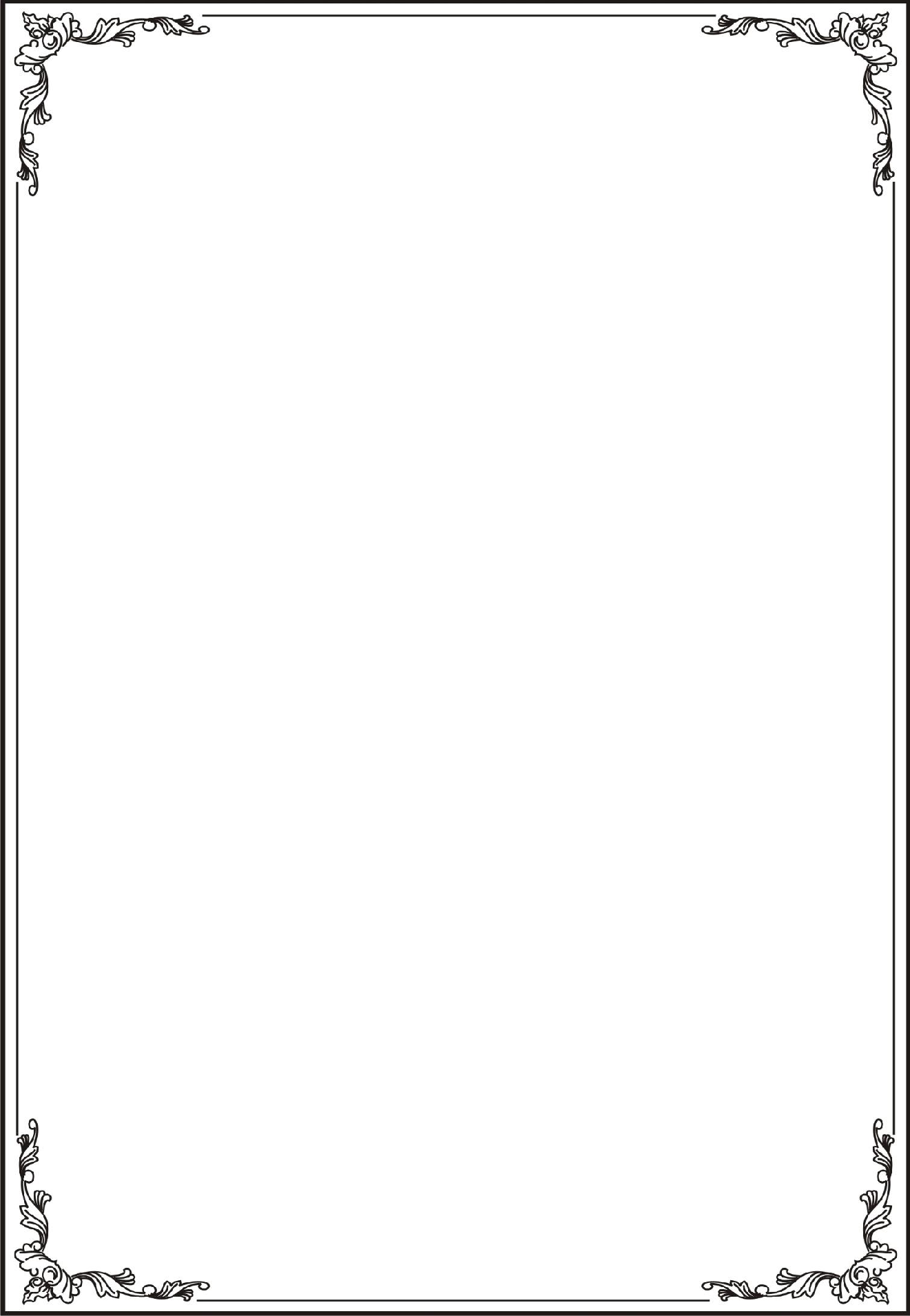
**TRƯỜNG ĐẠI HỌC SƯ PHẠM KỸ THUẬT TP. HỒ CHÍ MINH**

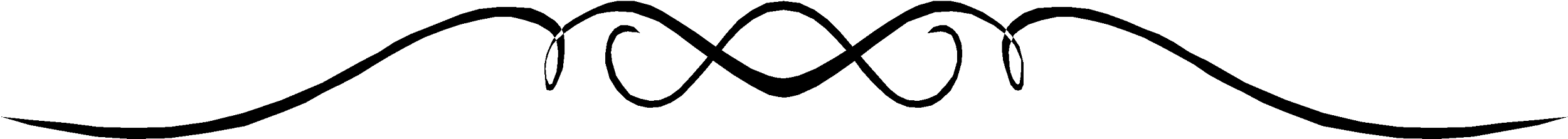


**KHOA CÔNG NGHỆ THÔNG TIN**

**BÁO CÁO CUỐI KỲ MÔN HỌC TRỰC QUAN HOÁ DỮ LIỆU**

**ĐỀ TÀI**

**Ứng dụng Apache Superset để xây dựng dashboard nghiên cứu về các vụ tai nạn máy bay**

**

**GVHD:** Ths. Huỳnh Xuân Phụng

**Lớp:** Thứ 5 (tiết 12-15)

**Sinh viên thực hiện:** Nhóm 25

Lý Quốc Dũng – 19133015

Đoàn Trần Đăng Khoa – 19133028

Lê Thị Kim Ngân – 19133038

|  |  |
| --- | --- |
| **ĐH SƯ PHẠM KỸ THUẬT TP.HCM KHOA CNTT**  \*\*\*\*\*\*\* | **XÃ HỘI CHỦ NGHĨA VIỆT NAM**  **Độc lập – Tự do – Hạnh Phúc**  \*\*\*\*\*\*\* |

**PHIẾU NHẬN XÉT CỦA GIÁO VIÊN HƯỚNG DẪN**

Họ và tên Sinh viên 1: Lý Quốc Dũng MSSV: 19133015

Họ và tên Sinh viên 2: Đoàn Trần Đăng Khoa MSSV: 19133028

Họ và tên Sinh viên 3: Lê Thị Kim Ngân MSSV: 19133038

Ngành: Kỹ thuật dữ liệu

Tên đề tài: Ứng dụng Apache Superset để xây dựng dashboard nghiên cứu về các vụ tai nạn máy bay

Họ và tên Giáo viên hướng dẫn: Huỳnh Xuân Phụng

**NHẬN XÉT**

Về nội dung đề tài khối lượng thực hiện:

..................................................................................................................................................

..................................................................................................................................................

..................................................................................................................................................

..................................................................................................................................................

1. Ưu điểm:

..................................................................................................................................................

.................................................................................................................................................

..................................................................................................................................................

..................................................................................................................................................

1. Khuyết điểm

..................................................................................................................................................

..................................................................................................................................................

..................................................................................................................................................

..................................................................................................................................................

1. Đề nghị cho bảo vệ hay không?
2. Đánh giá loại:
3. Điểm:

Tp*. Hồ Chí Minh, ngày tháng năm 2022*

Giáo viên hướng dẫn

*(Ký & ghi rõ họ tên)*

|  |  |
| --- | --- |
| **ĐH SƯ PHẠM KỸ THUẬT TP.HCM KHOA CNTT**  \*\*\*\*\*\*\* | **XÃ HỘI CHỦ NGHĨA VIỆT NAM**  **Độc lập – Tự do – Hạnh Phúc**  \*\*\*\*\*\*\* |

**PHIẾU NHẬN XÉT CỦA GIÁO VIÊN PHẢN BIỆN**

Họ và tên Sinh viên 1: Lý Quốc Dũng MSSV: 19133015

Họ và tên Sinh viên 2: Đoàn Trần Đăng Khoa MSSV: 19133028

Họ và tên Sinh viên 3: Lê Thị Kim Ngân MSSV: 19133038

Ngành: Kỹ thuật dữ liệu

Tên đề tài: Ứng dụng Apache Superset để xây dựng dashboard nghiên cứu về các vụ tai nạn máy bay

Họ và tên Giáo viên phản biện:

**NHẬN XÉT**

Về nội dung đề tài khối lượng thực hiện:

..................................................................................................................................................

..................................................................................................................................................

..................................................................................................................................................

..................................................................................................................................................

1. Ưu điểm:

..................................................................................................................................................

.................................................................................................................................................

..................................................................................................................................................

..................................................................................................................................................

1. Khuyết điểm

..................................................................................................................................................

..................................................................................................................................................

..................................................................................................................................................

..................................................................................................................................................

1. Đề nghị cho bảo vệ hay không?
2. Đánh giá loại:
3. Điểm:

Tp*. Hồ Chí Minh, ngày tháng năm 2022*

Giáo viên phản biện

*(Ký & ghi rõ họ tên)*

# **LỜI CẢM ƠN**

Lời nói đầu tiên, nhóm thực hiện xin được gửi đến thầy Huỳnh Xuân Phụng – giảng viên bộ môn lời cảm ơn chân thành và sâu sắc nhất.

Thứ hai, nhóm cũng xin gửi lời cảm ơn chân thành đến các thành viên và các nhóm trong lớp học đã gửi những câu hỏi và lời góp ý đến nhóm. Thông qua những góp ý này, nhóm chúng tôi đã hoàn thiện và khắc phục những điểm thiếu sót.

Nhóm thực hiện xin cảm ơn sự quan tâm và giúp đỡ tận tình của thầy trong suốt quá trình giảng dạy. Cảm ơn thầy đã luôn giải đáp những thắc mắc cũng như đưa ra những nhận xét, góp ý giúp nhóm thực hiện cải thiện chất lượng công việc của nhóm.

Vì khả năng còn hạn chế nên trong quá trình thực hiện báo cáo không tránh khỏi sai sót, kính mong nhận được những ý kiến đóng góp từ thầy để nhóm có thể cải thiện hơn sau này.

Nhóm thực hiện xin chân thành cảm ơn thầy.

**MỤC LỤC**

[**LỜI CẢM ƠN** 4](#_Toc104465924)

[**DANH MỤC HÌNH** 8](#_Toc104465925)

[1. Chọn List Roles 8](#_Toc104465926)

[2. Edit List Roles 8](#_Toc104465927)

[3. Details Edit Role 8](#_Toc104465928)

[4. Add User 8](#_Toc104465929)

[5. Kết quả Add User 8](#_Toc104465930)

[6. Add Chart 8](#_Toc104465931)

[7. Create New Chart 8](#_Toc104465932)

[8. Tổng tai nạn 8](#_Toc104465933)

[9. Tổng hành khách 8](#_Toc104465934)

[10. Tổng số người chết 8](#_Toc104465935)

[11. Tỷ lệ tai nạn các hãng bay 8](#_Toc104465936)

[12. Tỷ lệ loại máy bay bị tai nạn 8](#_Toc104465937)

[13. Ranking các hãng bị tai nạn 8](#_Toc104465938)

[14. Giai đoạn các hãng xảy ra tai nạn 8](#_Toc104465939)

[15. Tổng số người sống và chết trên tổng hành khách do tai nạn máy bay 8](#_Toc104465940)

[16. Thông tin top 5 hãng tai nạn máy bay 8](#_Toc104465941)

[17. Giai đoạn top 5 hãng tai nạn máy bay 8](#_Toc104465942)

[18. Truy cập vào SQL Editor 8](#_Toc104465943)

[19. Filter và Explorer 8](#_Toc104465944)

[20. Filter box 8](#_Toc104465945)

[21. Cấu hình tham số Filter 8](#_Toc104465946)

[22. Kết quả Filter 8](#_Toc104465947)

[PHẦN 1: MỞ ĐẦU 10](#_Toc104465948)

[1.1 Tính cấp thiết của đề tài 10](#_Toc104465949)

[1.2 Mục đích của đề tài 10](#_Toc104465950)

[1.3 Đối tượng và phạm vi nghiên cứu 10](#_Toc104465951)

[1.4 Phân tích, đánh giá các cuộc tai nạn máy bay 10](#_Toc104465952)

[1.5 Kết quả dự kiến đạt được 10](#_Toc104465953)

[PHẦN 2: NỘI DUNG 11](#_Toc104465954)

[CHƯƠNG 1: CƠ SỞ LÝ THUYẾT 11](#_Toc104465955)

[1.1 Sơ lược về AWS 11](#_Toc104465956)

[1.1.1 Giới thiệu AWS 11](#_Toc104465957)

[1.1.2 Các dịch vụ đã sử dụng 11](#_Toc104465958)

[1.2 Sơ lược về Docker và các dịch vụ chính 11](#_Toc104465959)

[1.2.1 Khái niệm 11](#_Toc104465960)

[1.2.2 Các thành phần cơ bản Docker 11](#_Toc104465961)

[1.2.3 Kiến trúc docker 13](#_Toc104465962)

[1.3 Giới thiệu về Apache Superset 13](#_Toc104465963)

[1.4 Một số khái niệm khác 14](#_Toc104465964)

[CHƯƠNG 2: THIẾT KẾ VÀ CÀI ĐẶT 15](#_Toc104465965)

[2.1. Thiết kế cơ sở dữ liệu 15](#_Toc104465966)

[2.1.1. Mô tả dữ liệu 15](#_Toc104465967)

[2.1.3. Kết nối cơ sở dữ liệu 15](#_Toc104465968)

[2.1.4. Điều chỉnh cơ sở dữ liệu 15](#_Toc104465969)

[2.3. Xây dựng dashboard 16](#_Toc104465970)

[2.3.1. Bảo mật dashboard 16](#_Toc104465971)

[Phân quyền cho user người xem 16](#_Toc104465972)

[2.3.1.1. Chọn List Roles 16](#_Toc104465973)

[2.3.1.2. Edit List Roles 16](#_Toc104465974)

[2.3.1.3.Details Edit Role 17](#_Toc104465975)

[2.3.1.4.Add User 17](#_Toc104465976)

[2.3.1.5.Kết quả Add User 18](#_Toc104465977)

[2.3.2. Xây dựng các chart 18](#_Toc104465978)

[2.3.2.1. Add Chart 18](#_Toc104465979)

[2.3.2.2. Create New Chart 18](#_Toc104465980)

[2.3.2.3.Tổng tai nạn 19](#_Toc104465981)

[2.3.2.4.Tổng hành khách 19](#_Toc104465982)

[2.3.2.5.Tổng số người chết 20](#_Toc104465983)

[2.3.2.6.Tỷ lệ tai nạn các hãng bay 20](#_Toc104465984)

[2.3.2.7.Tỷ lệ loại máy bay bị tai nạn 21](#_Toc104465985)

[2.3.2.8.Ranking các hãng bị tai nạn 21](#_Toc104465986)

[2.3.2.9.Giai đoạn các hãng xảy ra tai nạn 22](#_Toc104465987)

[2.3.2.10. Tổng số người sống và chết trên tổng hành khách do tai nạn máy bay 22](#_Toc104465988)

[2.3.2.11.Thông tin top 5 hãng tai nạn máy bay 23](#_Toc104465989)

[2.3.2.12.Giai đoạn top 5 hãng tai nạn máy bay 23](#_Toc104465990)

[2.3.4. Xây dựng filter 23](#_Toc104465991)

[2.3.4.1.Truy cập vào SQL Editor 24](#_Toc104465992)

[2.3.4.2.Filter và Explorer 24](#_Toc104465993)

[2.3.4.3.Filter box 25](#_Toc104465994)

[2.3.4.4.Cấu hình tham số Filter 25](#_Toc104465995)

[2.3.4.5.Kết quả Filter 26](#_Toc104465996)

[2.4. Các tiện ích mở rộng 26](#_Toc104465997)

[ *Cấu hình file superset\_config.py* 27](#_Toc104465998)

[PHẦN 3: KẾT LUẬN 27](#_Toc104465999)

[3.1 Kết quả đạt được 27](#_Toc104466000)

[3.2 Ưu nhược điểm của đề tài 27](#_Toc104466001)

[3.2.1 Ưu điểm 27](#_Toc104466002)

[3.2.2 Nhược điểm 27](#_Toc104466003)

[3.3 Hướng phát triển 27](#_Toc104466004)

# **DANH MỤC HÌNH**

## [Chọn List Roles](#_2.3.1.1._Chọn_List)

## [Edit List Roles](#_2.3.1.2._Edit_List)

## [Details Edit Role](#_2.3.1.3.Details_Edit_Role)

## [Add User](#_2.3.1.4.Add_User)

## [Kết quả Add User](#_2.3.1.5.Kết_quả_Add)

## [Add Chart](#_2.3.2.1._Add_Chart)

## [Create New Chart](#_2.3.2.2._Create_New)

## [Tổng tai nạn](#_2.3.2.3.Tổng_tai_nạn)

## [Tổng hành khách](#_2.3.2.4.Tổng_hành_khách)

## [Tổng số người chết](#_2.3.2.5.Tổng_số_người)

## [Tỷ lệ tai nạn các hãng bay](#_2.3.2.6.Tỷ_lệ_tai)

## [Tỷ lệ loại máy bay bị tai nạn](#_2.3.2.7.Tỷ_lệ_loại)

## [Ranking các hãng bị tai nạn](#_2.3.2.8.Ranking_các_hãng)

## [Giai đoạn các hãng xảy ra tai nạn](#_2.3.2.9.Giai_đoạn_các)

## [Tổng số người sống và chết trên tổng hành khách do tai nạn máy bay](#_2.3.2.10._Tổng_số)

## [Thông tin top 5 hãng tai nạn máy bay](#_2.3.2.11.Thông_tin_top)

## [Giai đoạn top 5 hãng tai nạn máy bay](#_2.3.2.12.Giai_đoạn_top)

## [Truy cập vào SQL Editor](#_2.3.4.1.Truy_cập_vào)

## [Filter và Explorer](#_2.3.4.2.Filter_và_Explorer)

## [Filter box](#_2.3.4.3.Filter_box)

## [Cấu hình tham số Filter](#_2.3.4.4.Cấu_hình_tham)

## [Kết quả Filter](#_2.3.4.5.Kết_quả_Filter)

**DANH MỤC TỪ VIẾT TẮT**

|  |  |
| --- | --- |
| Date | Date of accident, in the format - January 01, 2001 |
| Time | Local time, in 24 hr. format unless otherwise specified |
| Location | Location of the accident |
| Operator | Airline or operator of the aircraft |
| Flight | Flight number assigned by the aircraft operator |
| Router | Complete or partial route flown prior to the accident |
| Type | Aircraft type |
| Registration | ICAO registration of the aircraft |
| cn/In | Construction or serial number / Line or fuselage number |
| Aboard | Total aboard (passengers / crew) |
| AWS | Amazon Web Services |

# **PHẦN 1: MỞ ĐẦU**

## **Tính cấp thiết của đề tài**

Trong thời đại ngày nay đời sống con người ngày càng hiện đại, việc di chuyển giữa các quốc gia khu vực trên toàn thế giới ngày càng thuận tiện, dễ dàng hơn. Tuy thế đi kèm với sự tiện lợi đó là những nguy cơ tìm ẩn, và đã có các cuộc tai nạn máy bay thương tâm xảy ra. Vì vậy, việc phân tích bộ dữ liệu chứa dữ liệu về các vụ tai nạn máy bay liên quan đến vận tải dân dụng, thương mại và quân sự trên toàn thế giới từ năm 1908-09-17 đến 2009-06-08 là cần thiết.

## **Mục đích của đề tài**

Do tính xã hội của dữ liệu, mục đích của nhóm khi phân tích dữ liệu để đưa ra các phương án phù hợp để hạn chế tối đa các vụ tai nạn tương tự xảy ra. Ta tìm ra các yếu tố ảnh hưởng có thể gây ra các vụ tai nạn trong quá khứ. Từ các phân tích từ Dashboard đưa ra các đề xuất để phân tích , ứng dụng giảm các vụ tai nạn máy bay trong tương lai.

## **Đối tượng và phạm vi nghiên cứu**

* Đối tượng nghiên cứu: Các nhà nghiên cứu , sinh viên học tập nghiên cứu các lĩnh vực liên quan đến Big Data, Business, các nhà đầu tư,…
* Phạm vi nghiên cứu: Sinh viên học tập phân tích dữ liệu trên Superset.

## **Phân tích, đánh giá các cuộc tai nạn máy bay**

Dựa vào nguồn dữ liệu, ta có thể thấy được các cuộc tai nạn máy bay, rải khắp nhiều nơi, hầu hết các thời gian trong năm, nhiều hãng bay khác nhau, thời tiết khác nhau,… Do đó, nguồn dữ liệu cần được cập nhật liên tục để đạt được sự chính xác cao, và khách quan hơn.

## **Kết quả dự kiến đạt được**

Trong phạm vi đề tài nghiên cứu, nhóm sẽ nghiên cứu và tìm hiểu các cuộc tai nạn máy bay trong khoảng thời gian nhất định. Phân tích : Tổng chuyến bay, tổng số hành khách, tổng số người chết, tỷ lệ tai nạn các hãng, tỷ lệ tai nạn các loại máy bay, xếp hạng các hãng tai nạn, giai đoạn các hãng xảy ra tai nạn, thông tin top 5 hãng bay bị tai nạn, giai đoạn top 5 hãng tai nạn, số lượng máy bay của hãng bị tai nạn, thống kê giai đoạn số lượng tổng sống chết hành khách của hãng, filter chi tiết của một hãng bất kỳ,…

# **PHẦN 2: NỘI DUNG**

## **CHƯƠNG 1: CƠ SỞ LÝ THUYẾT**

## **Sơ lược về AWS**

### **Giới thiệu AWS**

Amazon Web Services (AWS) là nền tảng đám mây toàn diện và được sử dụng rộng rãi nhất, cung cấp trên 200 dịch vụ đầy đủ tính năng từ các trung tâm dữ liệu trên toàn thế giới. Hàng triệu khách hàng, bao gồm các công ty khởi nghiệp tăng trưởng nhanh nhất, các tập đoàn lớn nhất cũng như các cơ quan hàng đầu của chính phủ đều tin tưởng vào AWS để giảm chi phí, trở nên linh hoạt hơn và đổi mới nhanh hơn.

### **Các dịch vụ đã sử dụng**

AWS có nhiều dịch vụ và nhiều tính năng, cung cấp nhiều loại cơ sở dữ liệu nhất được xây dựng có mục đích cho các loại ứng dụng khác nhau để có thể chọn công cụ phù hợp cho công việc để có chi phí và hiệu suất tốt nhất. Nó có nhiều dịch vụ nổi bật như: Amazon EC2, Amazon Simple Storage Service (S3), Amazon Aurora, Amazon DynamoDB, Amazon RDS, AWS Lambda, Amazon VPC, Amazon Lightsail,…

Trong đó có hai dịch vụ nhóm đã sử dụng :

* Amazon EC2 (Elastic Compute Cloud), một trong những dịch vụ nổi tiếng nhất của Amazon Web Services, cung cấp cho các doanh nghiệp khả năng chạy các ứng dụng trên đám mây công cộng.
* Amazon S3 là dịch vụ lưu trữ đối tượng được xây dựng để lưu trữ và truy xuất bất kỳ lượng dữ liệu nào từ bất cứ nơi nào. Đây là dịch vụ lưu trữ đơn giản có độ bền, độ sẵn có, hiệu suất, tính bảo mật dẫn đầu ngành và khả năng thay đổi quy mô gần như không giới hạn với chi phí cực kỳ thấp.

## **Sơ lược về Docker và các dịch vụ chính**

### **Khái niệm**

Docker là một nền tảng để cung cấp cách để building, deploying và running ứng dụng dễ dàng hơn bằng cách sử dụng các containers (trên nền tảng ảo hóa). Ban đầu viết bằng Python, hiện tại đã chuyển sang Golang.

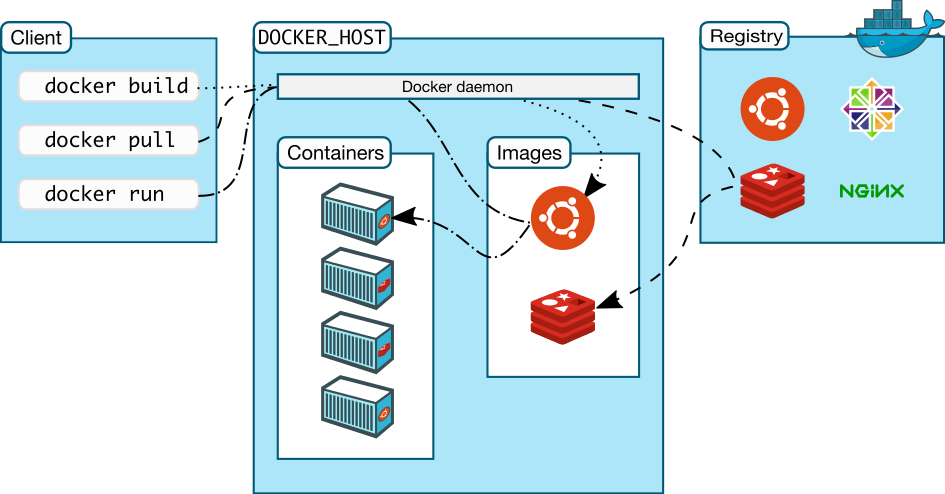
### **Các thành phần cơ bản Docker**

* ***Docker Engine***
* Docker Engine là công cụ Client - Server hỗ trợ công nghệ container để xử lý các nhiệm vụ và quy trình công việc liên quan đến việc xây dựng các ứng dụng dựa trên vùng chứa (container).
* Engine tạo ra một quy trình daemon phía máy chủ lưu trữ images, containers, networks và storage volumes.
* Images: là thành phần để đóng gói ứng dụng và các thành phần mà ứng dụng phụ thuộc để chạy. Và image được lưu trữ ở trên local hoặc trên một Registry (là nơi lưu trữ và cung cấp kho chứa các image)
* Containers: là một instance của image, và nó hoạt động như một thư mục, chứa tất cả những thứ cần thiết để chạy một ứng dụng
* Network: cung cấp một private network chỉ tồn tại giữa container và host
* volume: Volume trong Docker được dùng để chia sẻ dữ liệu cho container
* Daemon cũng cung cấp giao diện dòng lệnh phía máy khách (CLI) cho phép người dùng tương tác với daemon thông qua giao diện lập trình ứng dụng Docker.
* ***Distribution tools***
* Là các công cụ phân tán giúp chúng ta lưu trữ và quản lý các Docker Images như: Docker Registry, Docker Trusted Registry, Docker Hub
* ***Orchestration tools***
* Docker Machine: Machine tạo Docker Engine trên laptop của bạn hoặc trên bất cứ dịch vụ cloud phổ biến nào như AWS, Azure, Google Cloud, Softlayer hoặc trên hệ thống data center như VMware, OpenStack. Docker Machine sẽ tạo các máy ảo và cài Docker Engine lên chúng và cuối cùng nó sẽ cấu hình Docker Client để giao tiếp với Docker Engine một cách bảo mật
* Docker Compose: là công cụ giúp định nghĩa và khởi chạy multi-container Docker applications
* Docker Swarm: là một công cụ giúp chúng ta tạo ra một clustering Docker. Nó giúp chúng ta gom nhiều Docker Engine lại với nhau và ta có thể "nhìn" nó như duy nhất một virtual Docker Engine

Ngoài các thành phần trên ta còn có một số thành phần khác như:

* Dockerfile: như một script dùng để build các image trong container. Dockerfile bao gồm các câu lệnh liên tiếp nhau được thực hiện tự động trên một image gốc để tạo ra một image mới. Dockerfile giúp đơn giản hóa tiến trình từ lúc bắt đầu đến khi kết thúc
* Docker Toolbox: Bởi vì Docker Engine dùng một số feature của kernel Linux nên ta sẽ không thể chạy Docker Engine natively trên Windows hoặc BSD được. Ở các phiên bản trước đây thì ta sẽ cần một máy ảo cài một phiên bản Linux nào đó và sau đó cài Docker Engine lên máy ảo đó

### **Kiến trúc docker**



* Docker sử dụng kiến trúc client-server. Docker client sẽ liên lạc với các Docker daemon, các Docker daemon sẽ thực hiện các tác vụ build, run và distribuing các Docker container. Cả Docker client và Docker daemon có thể chạy trên cùng 1 máy, hoặc có thể kết nối theo kiểu Docker client điều khiển các docker daemon như hình trên. Docker client và daemon giao tiếp với nhau thông qua socket hoặc RESTful API
* Docker daemon chạy trên các máy host. Người dùng sẽ không tương tác trực tiếp với các daemon, mà thông qua Docker Client.

## **Giới thiệu về Apache Superset**

***1.3.1. Khái niệm***

Apache Superset là một ứng dụng web thông minh kinh doanh hiện đại, sẵn sàng cho doanh nghiệp. Nó nhanh, nhẹ, trực quan và được tải với các tùy chọn giúp người dùng thuộc tất cả các bộ kỹ năng dễ dàng khám phá và hình dung dữ liệu của họ, từ biểu đồ hình tròn đơn giản đến biểu đồ không gian địa lý deck.gl rất chi tiết.

Superset cung cấp:

* Giao diện trực quan để trực quan hóa tập dữ liệu và tạo bảng điều khiển tương tác
* Một loạt các hình ảnh trực quan đẹp mắt để giới thiệu dữ liệu của bạn
* Trình tạo trực quan hóa không cần mã để trích xuất và trình bày bộ dữ liệu
* Một SQL IDE đẳng cấp thế giới để chuẩn bị dữ liệu để hiển thị, bao gồm một trình duyệt siêu dữ liệu phong phú
* Một lớp ngữ nghĩa nhẹ cho phép các nhà phân tích dữ liệu nhanh chóng xác định các thứ nguyên và chỉ số tùy chỉnh
* Hỗ trợ out-of-the-box cho hầu hết các cơ sở dữ liệu nói tiếng SQL
* Bộ nhớ đệm và truy vấn không đồng bộ liền mạch, trong bộ nhớ
* Một mô hình bảo mật có thể mở rộng cho phép cấu hình các quy tắc rất phức tạp về việc ai có thể truy cập các tính năng và bộ dữ liệu của sản phẩm.
* Tích hợp với các phụ trợ xác thực chính (cơ sở dữ liệu, OpenID, LDAP, OAuth, REMOTE\_USER, v.v.)
* Khả năng thêm các plugin trực quan hóa tùy chỉnh
* Một API để tùy chỉnh có lập trình
* Một kiến ​​trúc gốc đám mây được thiết kế từ đầu cho quy mô

***1.3.2. Chart***

Là biểu đồ dùng để trực quan hóa các nguồn dữ liệu, giúp người dùng dễ dàng phân tích,, Superset có các kiểu biểu đồ rất đa dạng: pie chart, Time-series Line Chart, Time-series Area chart, Time-series Scatter Plot, Nightingale Rose Chart,…

***1.3.3. Dashboard***

Nơi hiển thi danh sách các Chart được tạo, đồng thời hiển thị các Modified, Status, Created by, Owners , Actions của Chart đó.

***1.3.4. Filter***

Dùng để lọc các dữ liệu theo ý muốn, nhầm mục đích phân tích các dữ liệu tạo ra các Chart phục vụ cho mục đích phân tích cụ thể

***1.3.5. Superset API***

Superset luôn có API, nhưng không phải lúc nào nó cũng có API công khai, RESTful, nghiêm ngặt, an toàn về loại. Ban đầu, API là một tập hợp các điểm cuối CRUD ( Create Read Update Delete) được tạo tự động được cung cấp bởi khung công tác web Flask App Builder , cũng như một tập hợp các điểm cuối ngày càng tăng đáp ứng nhu cầu của React frontend ngày càng tăng các ứng dụng.

## **Một số khái niệm khác**

* Điều chỉnh file superset\_config.py

Vào thư mục “superset/docker/pythonpath\_dev/” sau đó chỉnh sửa ở các row ở file superset\_config.py như hình:

Text

Description automatically generated

## **CHƯƠNG 2: THIẾT KẾ VÀ CÀI ĐẶT**

## **2.1. Thiết kế cơ sở dữ liệu**

### **2.1.1. Mô tả dữ liệu**

Bộ dữ liệu này bao gồm thông tin:

* Tất cả các vụ tai nạn hàng không dân dụng và thương mại của các hãng hàng không chở khách theo lịch trình và không theo lịch trình trên toàn thế giới, dẫn đến tử vong (bao gồm tất cả các vụ tai nạn chết người theo Phần 121 và Phần 135 của Hoa Kỳ)
* Tất cả các tai nạn chết người về hàng hóa, định vị, phà và chuyến bay thử nghiệm.
* Tất cả các vụ tai nạn vận tải quân sự có 10 người chết trở lên.
* Tất cả các vụ tai nạn máy bay trực thăng thương mại và quân sự với hơn 10 người thiệt mạng.
* Tất cả các vụ tai nạn tàu bay dân dụng và quân sự liên quan đến tử vong.
* Tai nạn hàng không liên quan đến cái chết của những người nổi tiếng.
* Tai nạn hàng không hoặc sự cố đáng quan tâm.

Bộ dữ liệu gồm 10 cột và 5269 dòng :

* Date: Ngày xảy ra tai nạn, theo định dạng - ngày 1 tháng 1 năm 2001
* Time: Giờ địa phương, 24 giờ.
* Location: Vị trí xảy ra tai nạn
* Operator: Hãng hàng không hoặc người điều hành máy bay
* Flight: Số hiệu chuyến bay
* Route: Toàn bộ hoặc một phần tuyến đường bay trước khi xảy ra tai nạn
* Type: Loại tàu bay
* Registration: ICAO đăng ký máy bay
* cn/In: Cấu tạo hoặc số sê-ri / Dòng hoặc số thân máy bay
* Aboard: Tổng số người trên tàu (hành khách / phi hành đoàn)
* Fatalities: Tổng số người chết trên tàu (hành khách / phi hành đoàn)
* Ground: Tổng số bị chết trên mặt đất
* Summary: Mô tả ngắn gọn về vụ tai nạn và nguyên nhân nếu biết

**2.1.2. Cơ sở dữ liệu**

- Tập dữ liệu Airplane\_Crashes\_and\_Fatalities\_Since\_1908.csv lấy từ kaggle

-Link: https://www.kaggle.com/datasets/saurograndi/airplane-crashes-since-1908?fbclid=IwAR3VuIGSk7h01b8M9UxFrWfI0reqrBlOXxvFY1nfadotdaj-wvI6lnyMy\_0

### **2.1.3. Kết nối cơ sở dữ liệu**

Sử dụng cơ sở dữ liệu Mysql

### **2.1.4. Điều chỉnh cơ sở dữ liệu**

* Chuyển Date từ dạng M-D-Y sang Y-M-D

select str\_to\_date(Date, '%m/%d/%Y') from Airplane3;

update Airplane3 set Date = str\_to\_date(Date,'%m/%d/%Y');

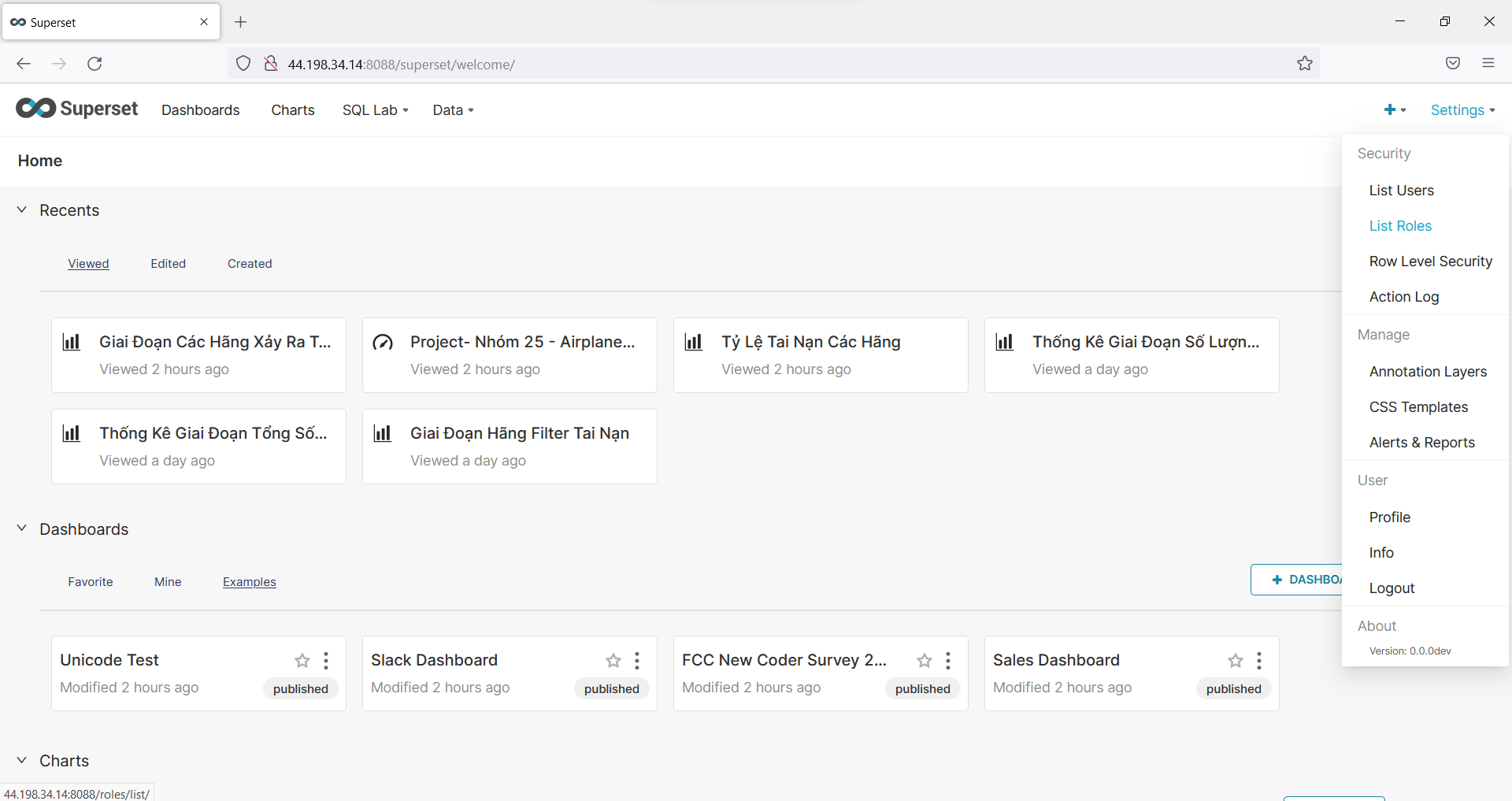
ALTER TABLE Airplane3 MODIFY Date date;

## **2.3. Xây dựng dashboard**

### **2.3.1. Bảo mật dashboard**

### **Phân quyền cho user người xem**

Vào Settings chọn List Roles



#### 2.3.1.1. Chọn List Roles

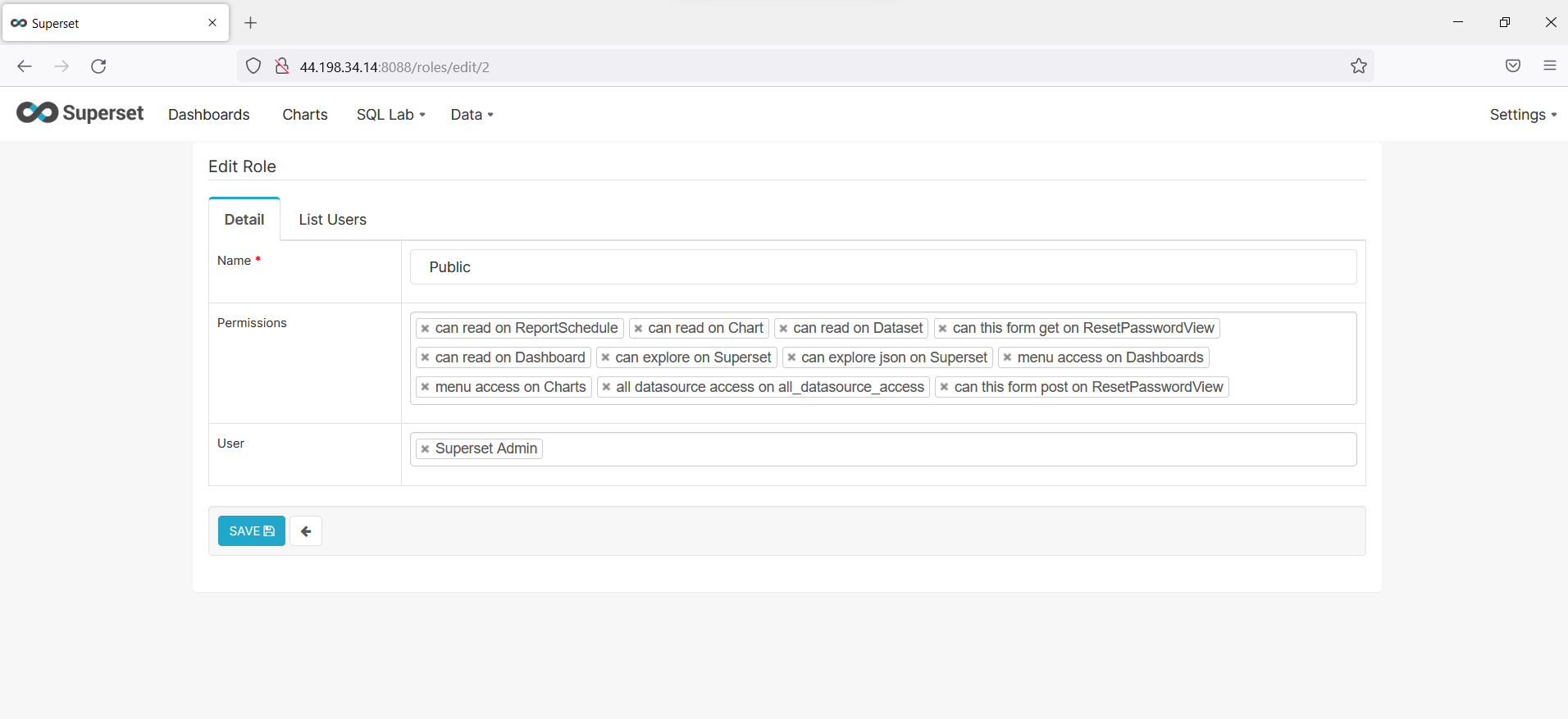
Thực hiện edit ở public

Ảnh có chứa văn bản

Mô tả được tạo tự động

#### 2.3.1.2. Edit List Roles

Chọn các mục trong Edit Role tương tự như hình, sau đó Save để lưu lại



#### 2.3.1.3.Details Edit Role

Nhập những thông tin để Add User, sau đó Save

#### 2.3.1.4.Add User

Kết quả sau khi Save, đã thành công thêm User

Ảnh có chứa văn bản

Mô tả được tạo tự động

#### 2.3.1.5.Kết quả Add User

### **2.3.2. Xây dựng các chart**

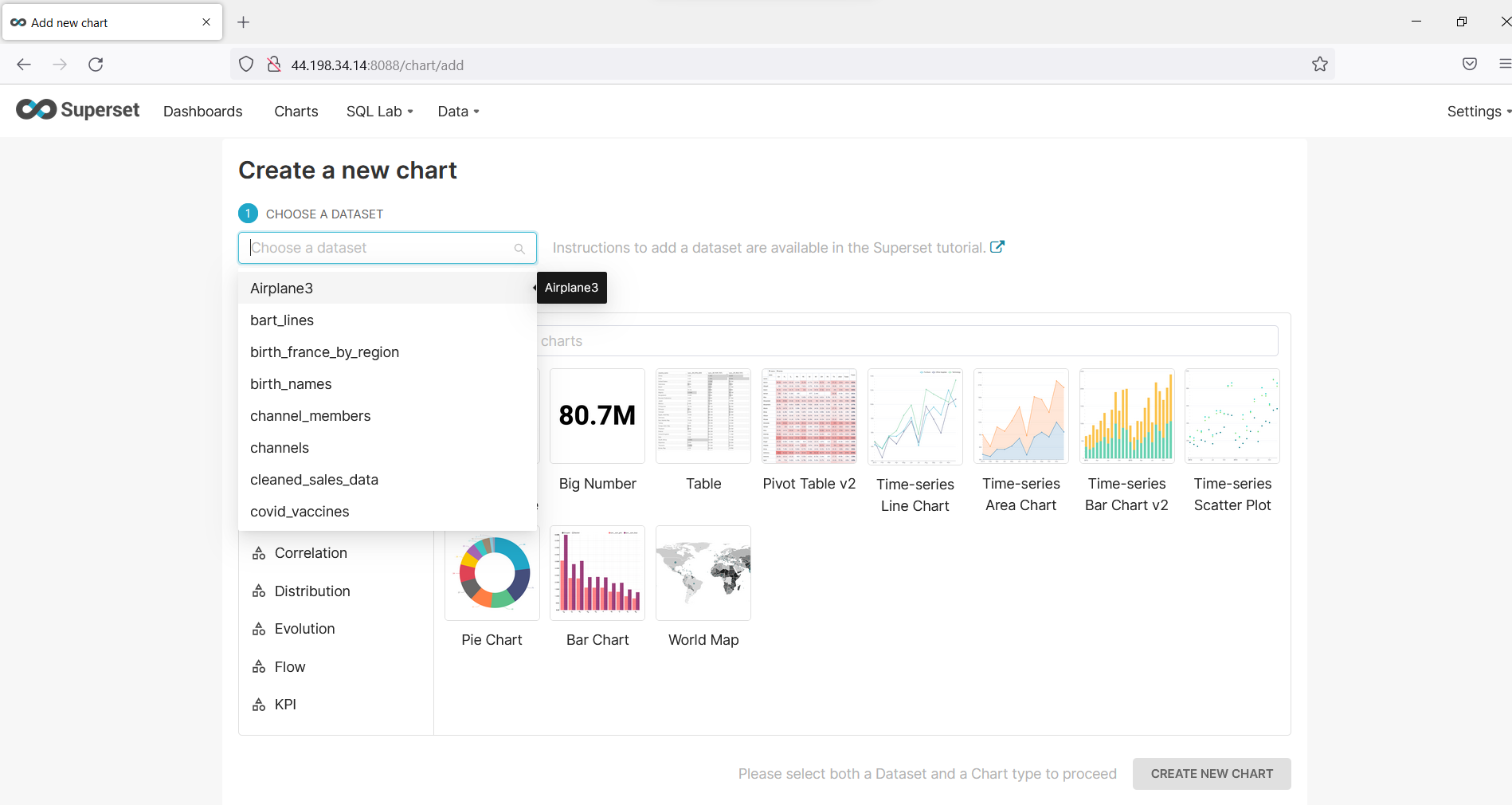
Trên thanh , chọn new chart

Ảnh có chứa văn bản

Mô tả được tạo tự động

#### 2.3.2.1. Add Chart

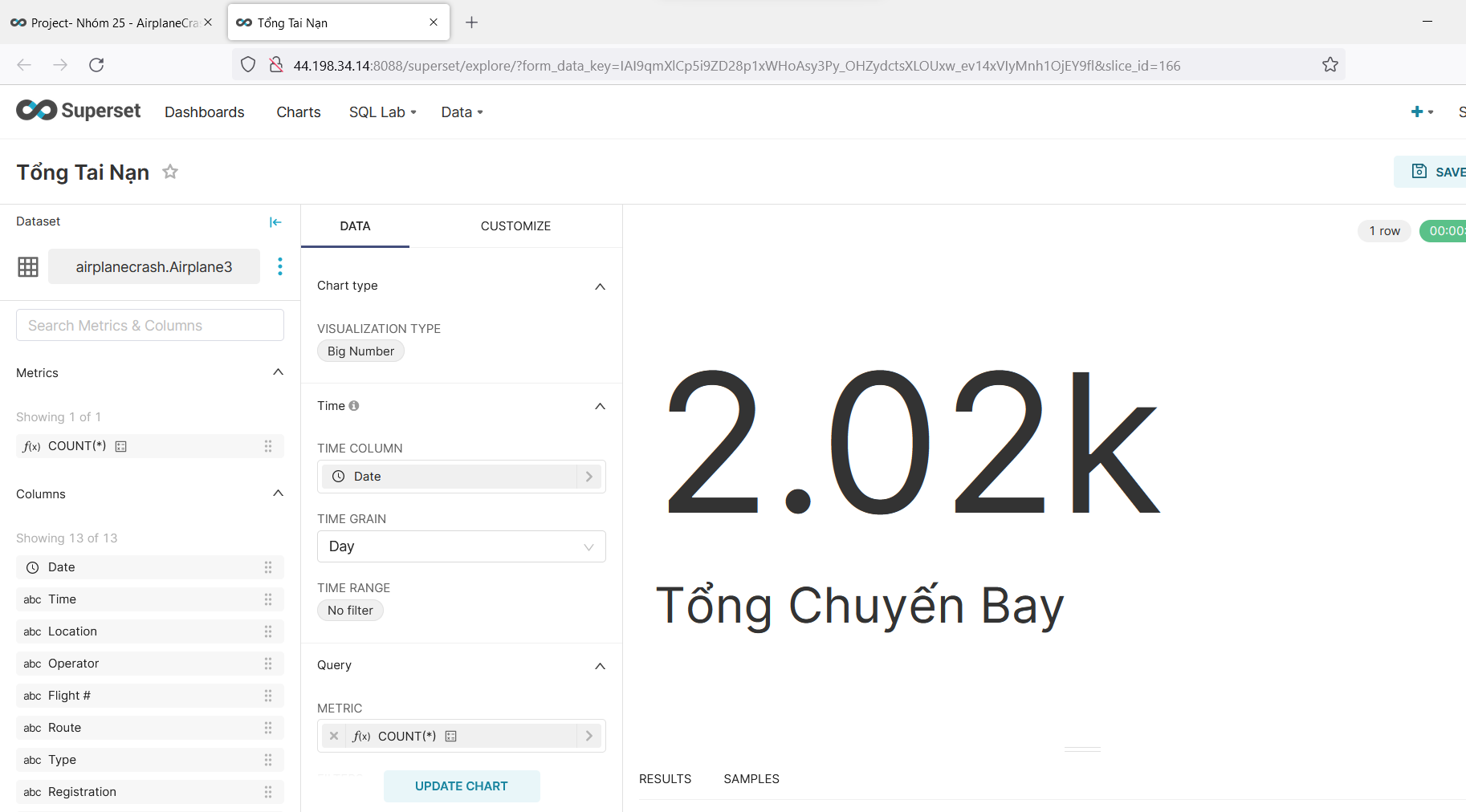
Chọn biểu đồ muốn vẽ:



#### 2.3.2.2. Create New Chart

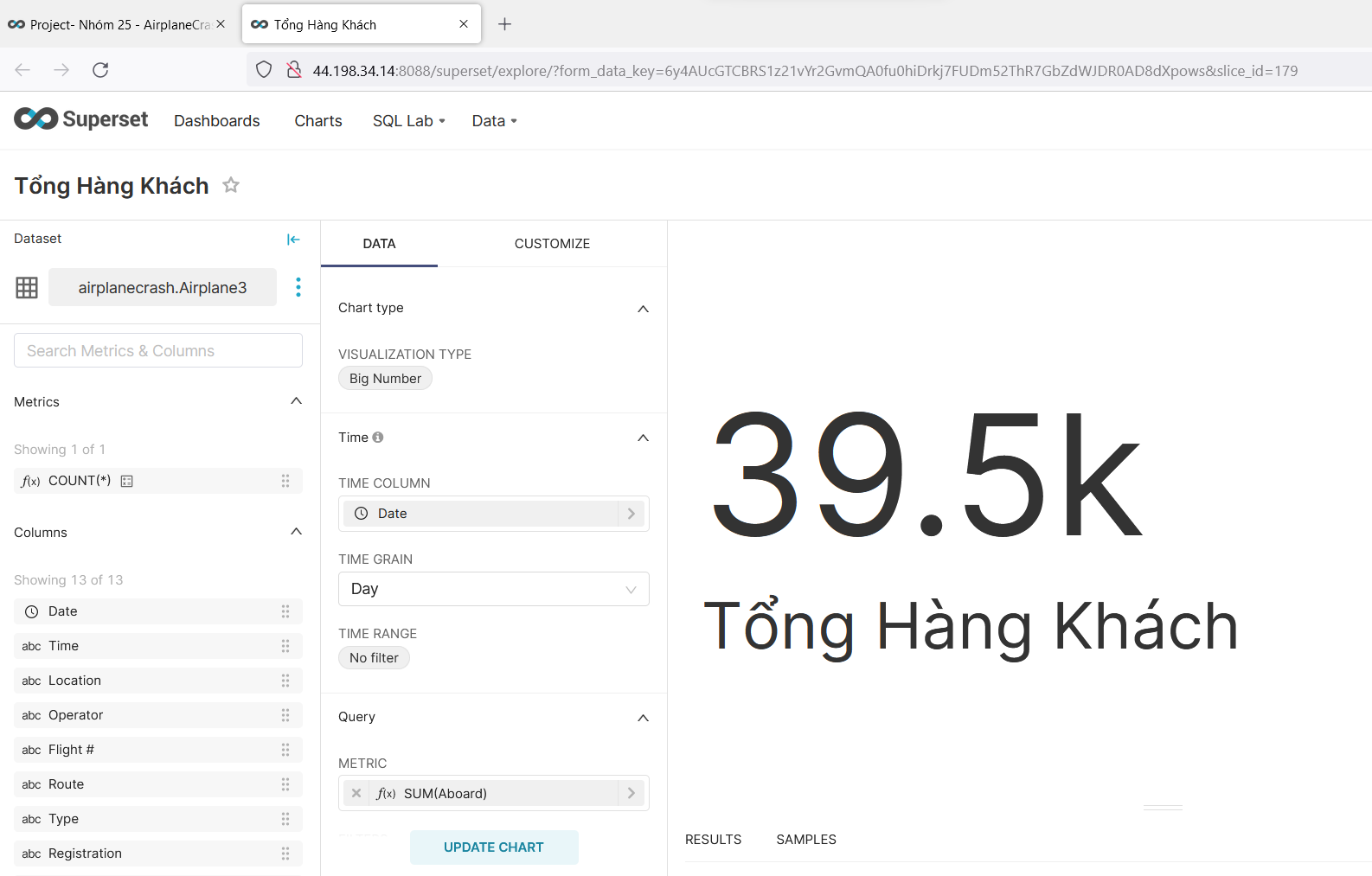
**Tổng quan dữ liệu**

* Tổng tai nạn



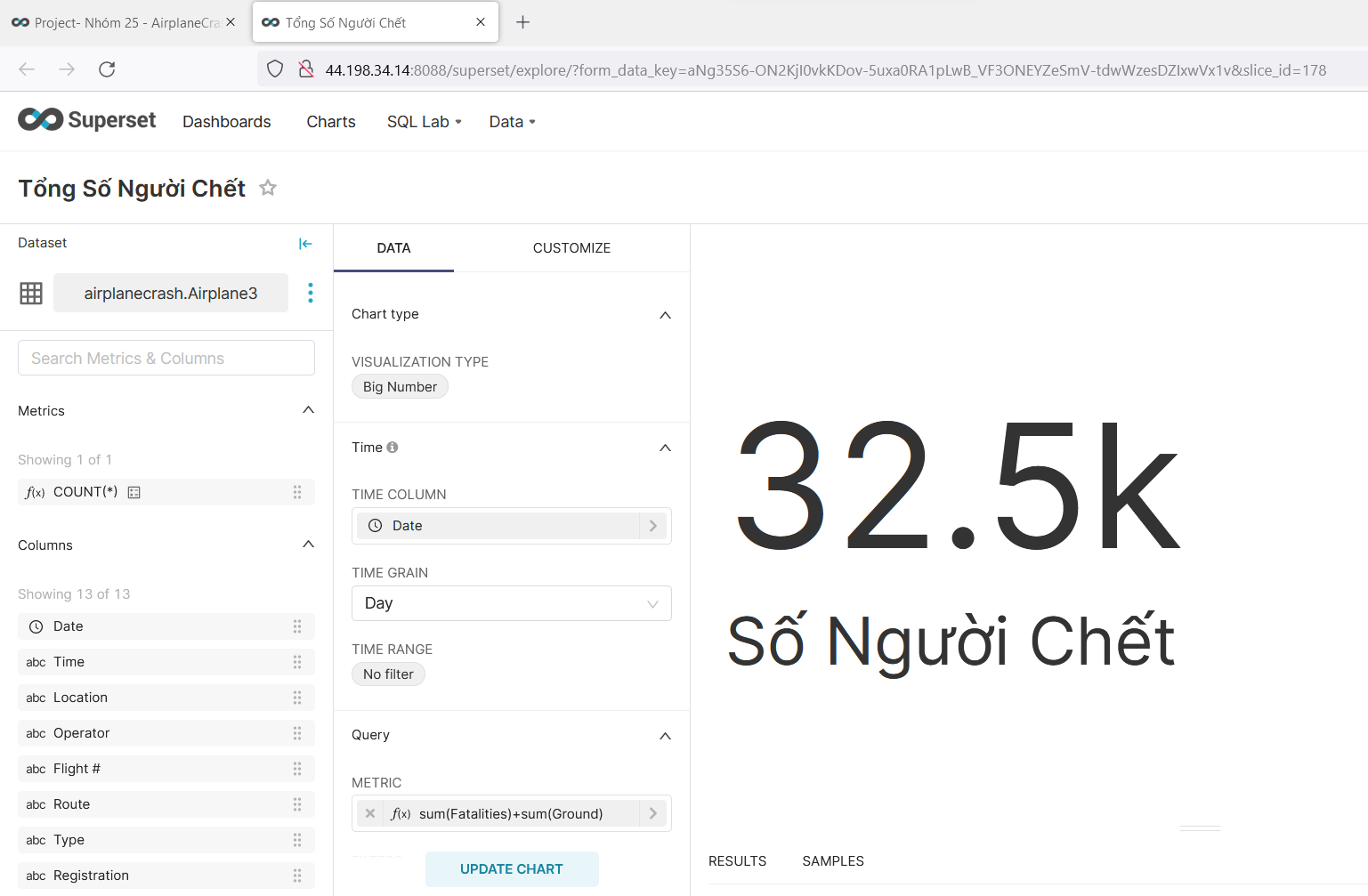
#### 2.3.2.3.Tổng tai nạn

* Tổng hành khách



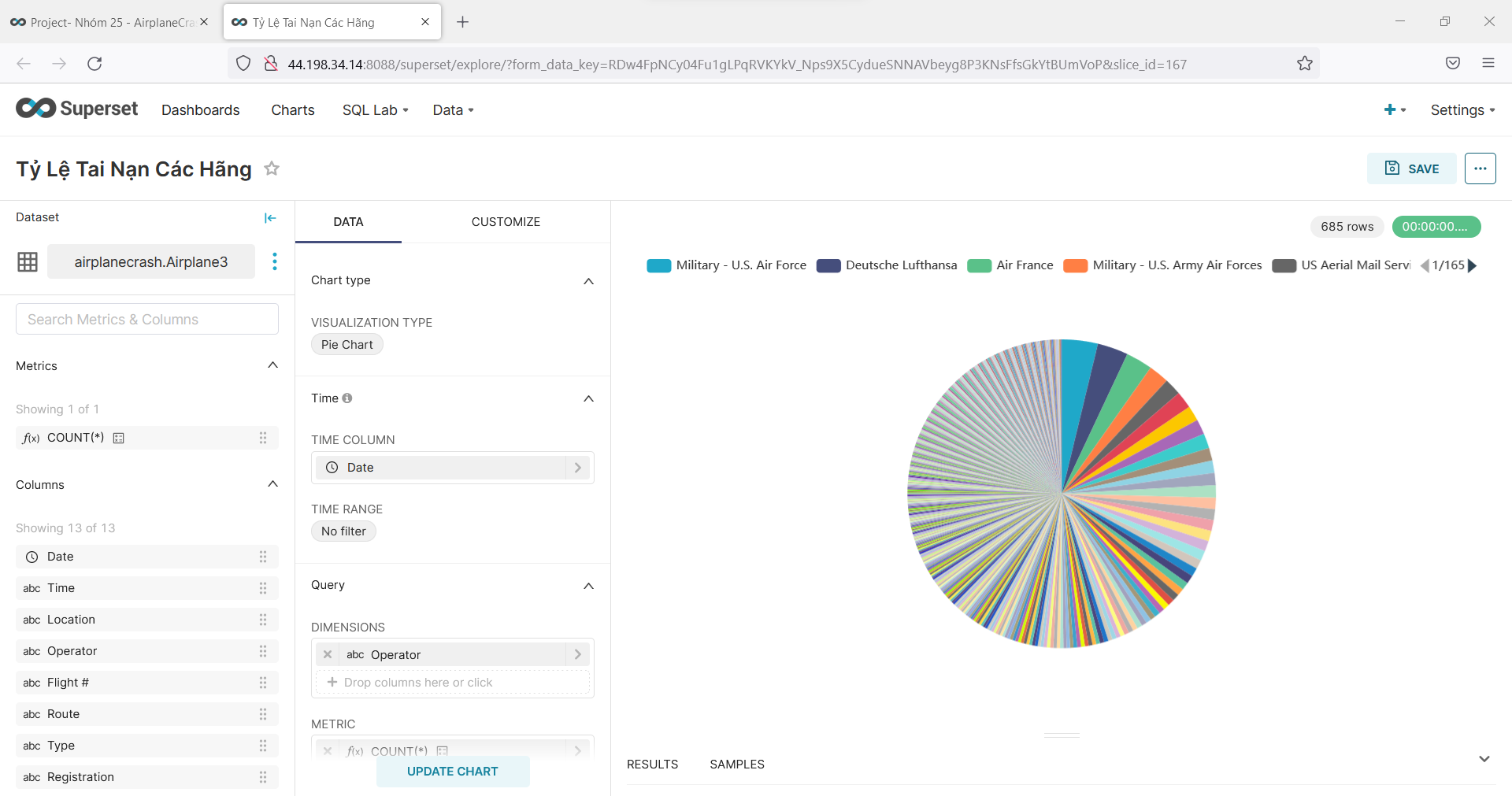
#### 2.3.2.4.Tổng hành khách

* Tổng số người chết



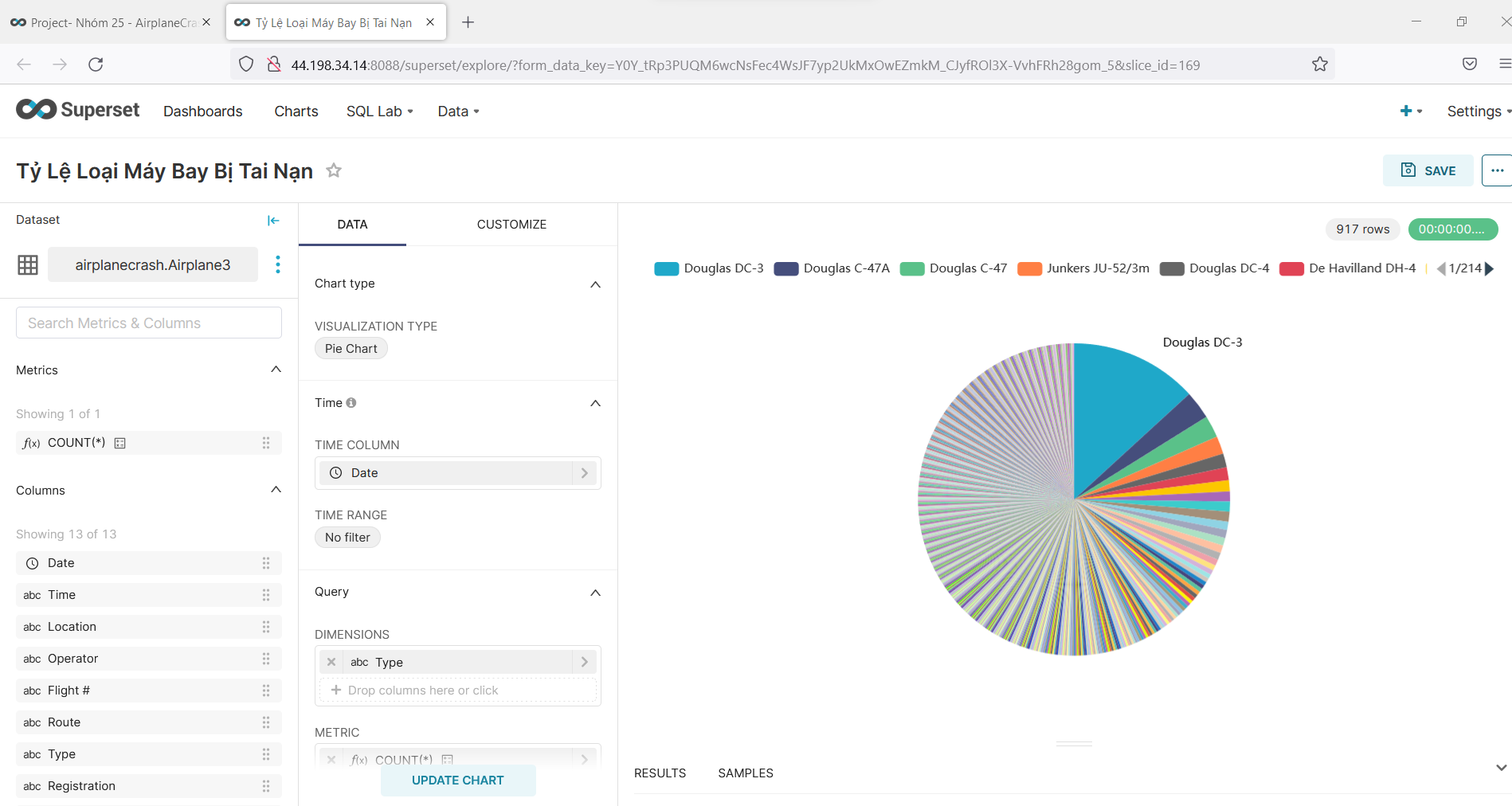
#### 2.3.2.5.Tổng số người chết

* Tỷ lệ tai nạn các hãng bay



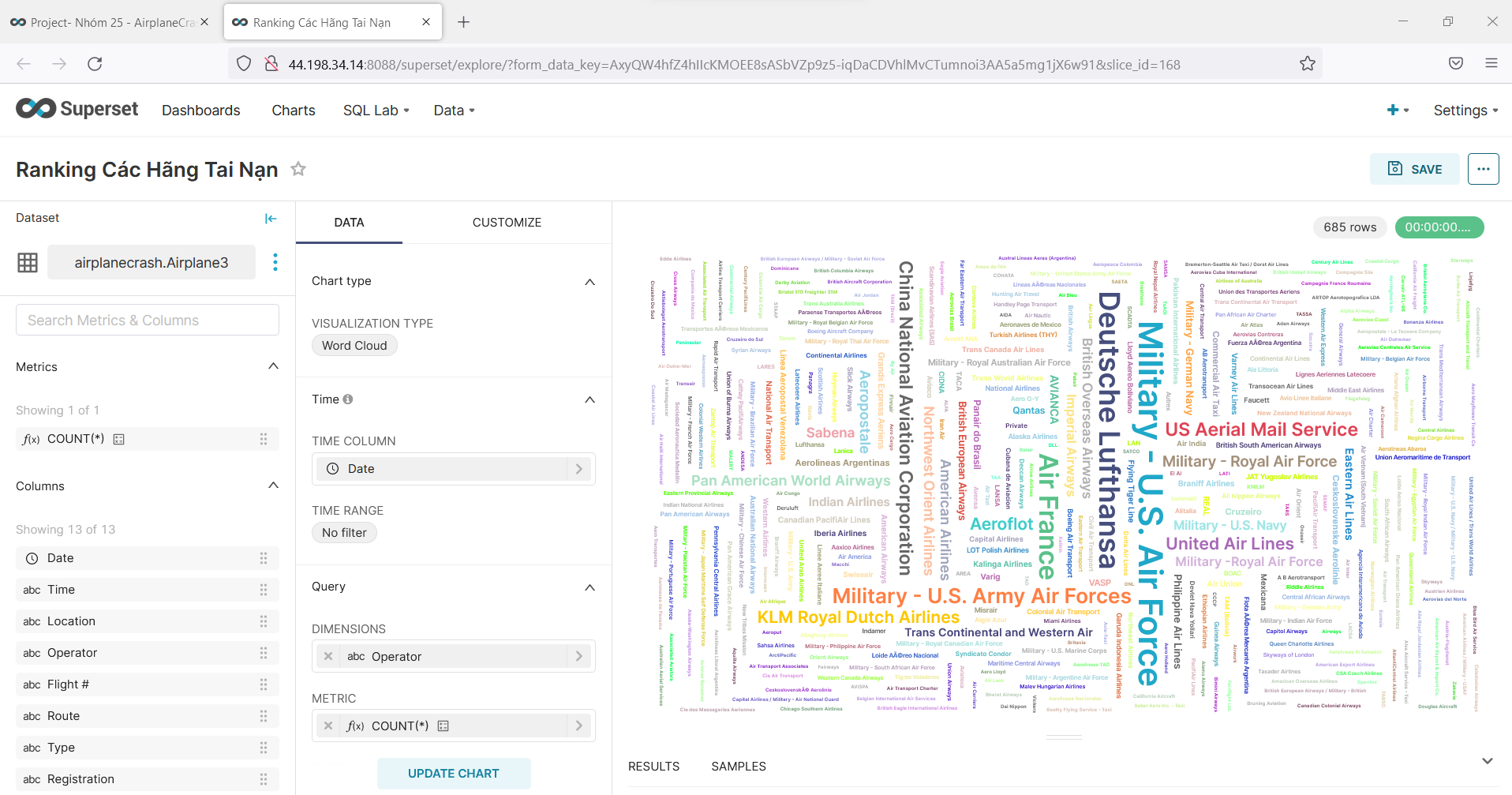
#### 2.3.2.6.Tỷ lệ tai nạn các hãng bay

* Tỷ lệ loại máy bay bị tai nạn



#### 2.3.2.7.Tỷ lệ loại máy bay bị tai nạn

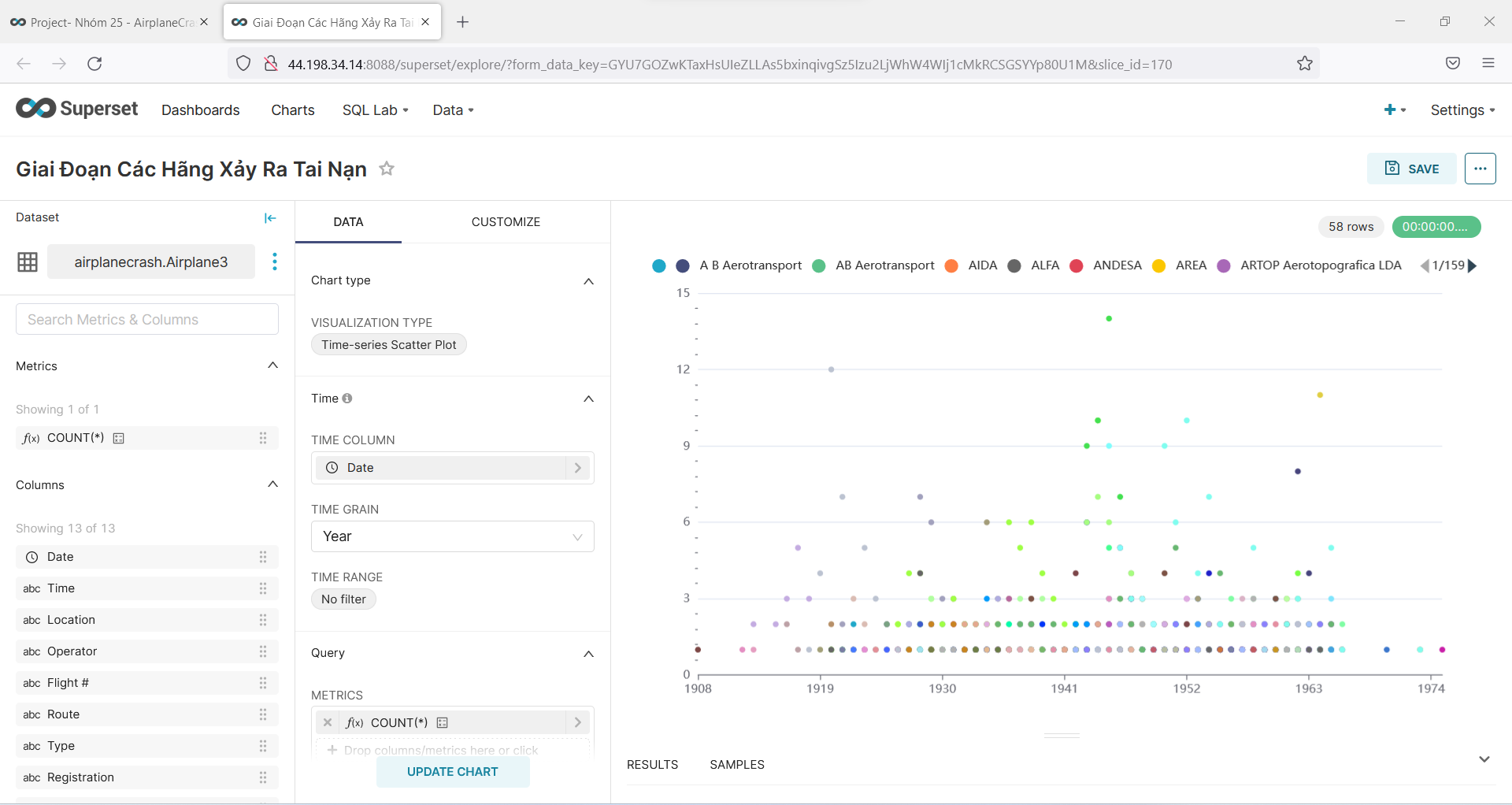
* Ranking các hãng bị tai nạn



#### 2.3.2.8.Ranking các hãng bị tai nạn

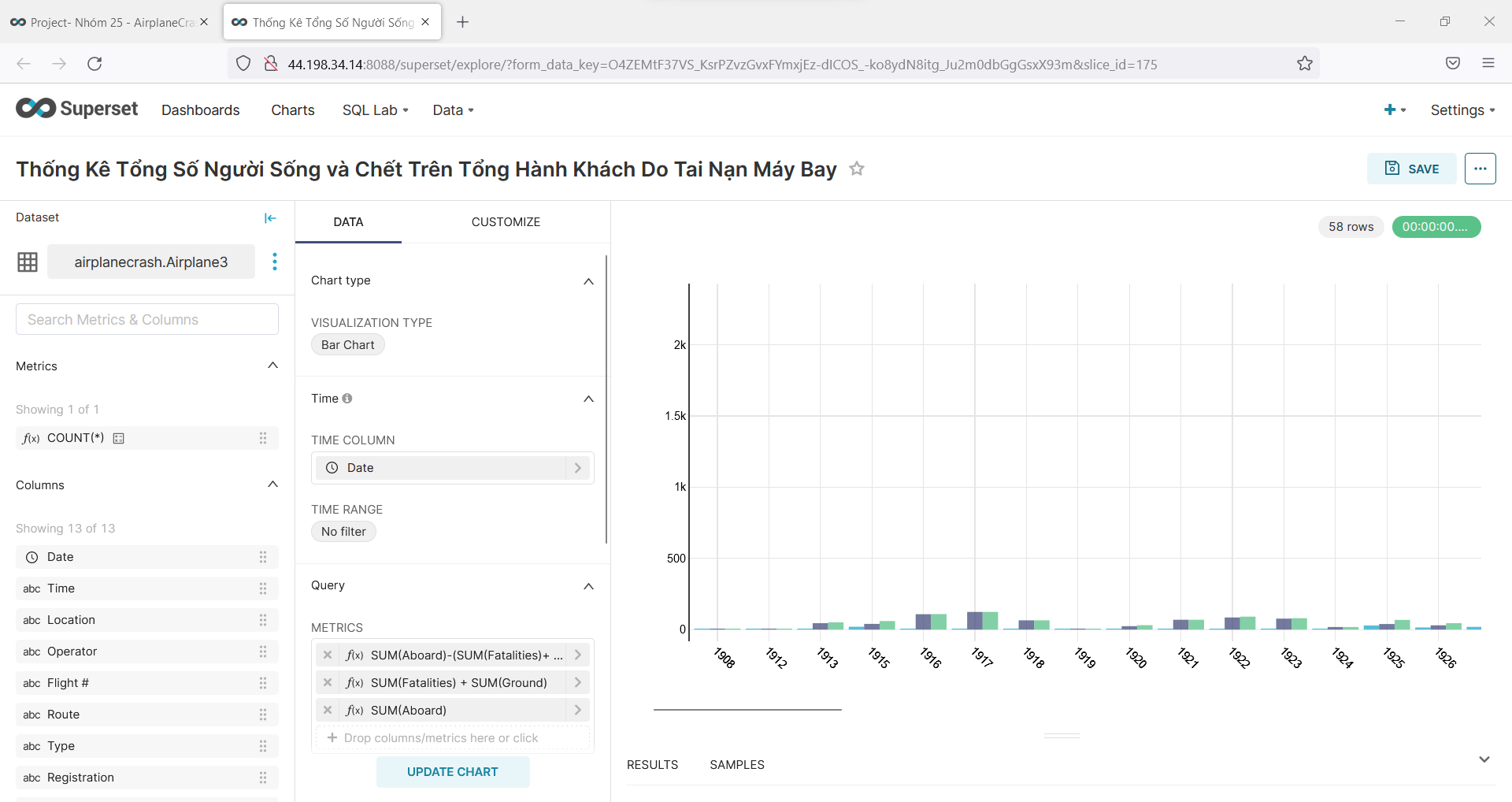
**Phân tích chi tiết**

* Giai đoạn các hãng xảy ra tai nạn



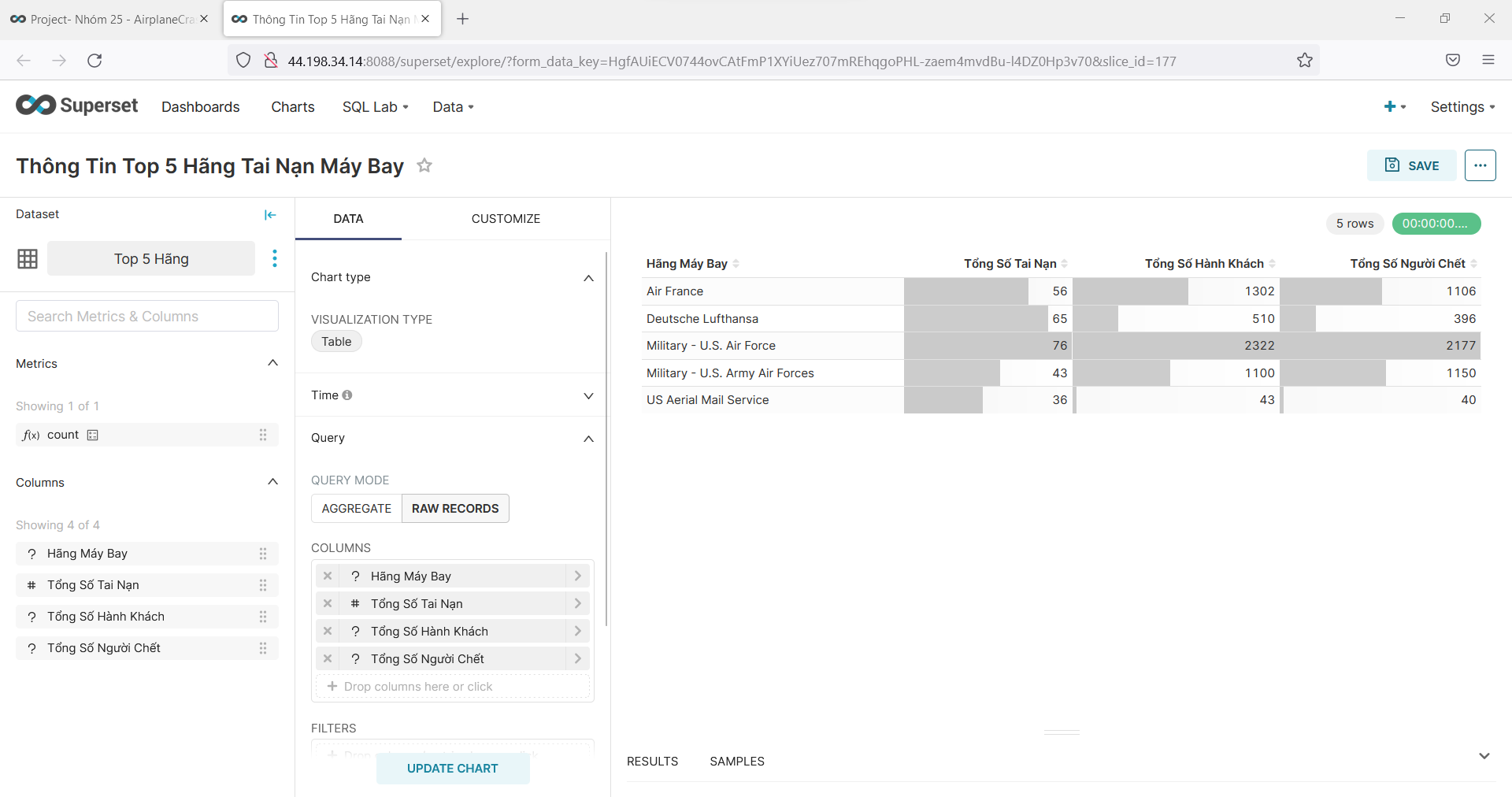
#### 2.3.2.9.Giai đoạn các hãng xảy ra tai nạn

* Thống kê tổng số người sống và chết trên tổng hành khách do tai nạn máy bay



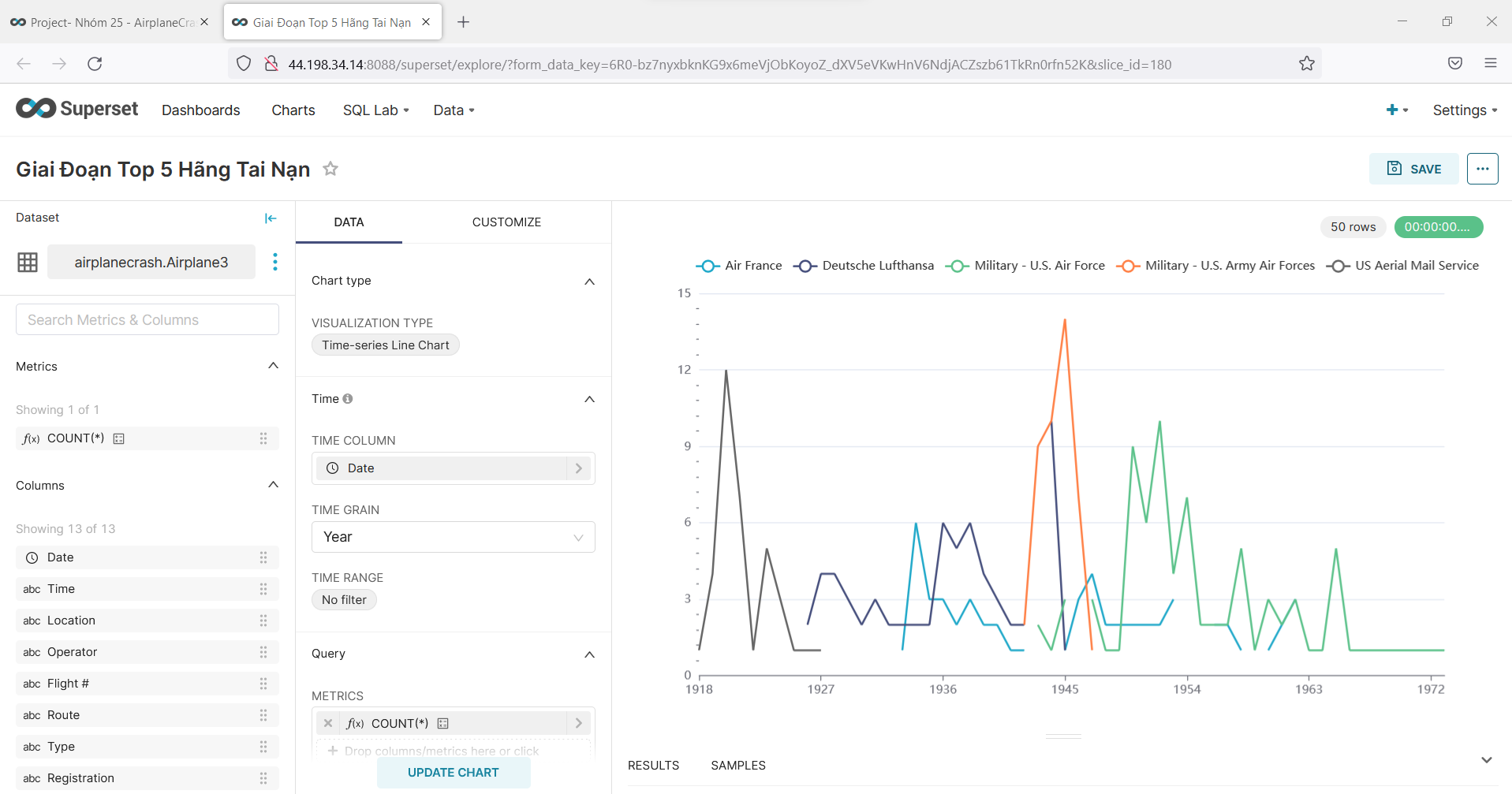
#### 2.3.2.10. Tổng số người sống và chết trên tổng hành khách do tai nạn máy bay

* Thông tin top 5 hãng tai nạn máy bay



#### 2.3.2.11.Thông tin top 5 hãng tai nạn máy bay

* Giai đoạn top 5 hãng tai nạn máy bay



#### 2.3.2.12.Giai đoạn top 5 hãng tai nạn máy bay

### **2.3.4. Xây dựng filter**

Ta truy cập vào SQL Editor

A screenshot of a computer

Description automatically generated

#### 2.3.4.1.Truy cập vào SQL Editor

Tạo câu truy vấn những thuộc tính mà ta cần filter và Explorer

Graphical user interface, text, application

Description automatically generated

#### 2.3.4.2.Filter và Explorer

Tạo chart filter box để filter

Graphical user interface, application

Description automatically generated

#### 2.3.4.3.Filter box

Cấu hình các tham số cần filter

Graphical user interface, application

Description automatically generated

#### 2.3.4.4.Cấu hình tham số Filter

Kết quả filter

Graphical user interface, text, application

Description automatically generated

#### 2.3.4.5.Kết quả Filter

**2.3.5. Sử dụng SQL template (Jinja)**

* *Vào đường dẫn sau ta có thể thấy file superset\_config.py*



* *Truy cập vào file superset\_config.py để sửa*



* *Thực hiện thêm những câu lệnh sau và file superset\_config.py*

Text

Description automatically generated

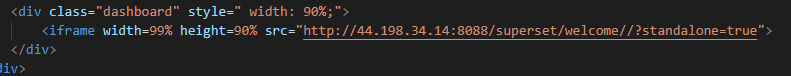
Câu lệnh đầu tiên dùng để kích hoạt thư viện Jinja template

Ba câu lệnh tiếp theo dùng để kích hoạt để có thể sử dụng trên trang web

## **2.4. Các tiện ích mở rộng**

**2.4.1. Tích hợp vào website**

* *Tạo file HTML và them iframe chứa đường dẫn:*



## *Cấu hình file superset\_config.py*

Text

Description automatically generated

* *Đưa project lên S3*

Graphical user interface, table

Description automatically generated

# **PHẦN 3: KẾT LUẬN**

### **3.1 Kết quả đạt được**

Nhóm sau khi tìm hiểu và phát triển đề tài , chúng em đã hoàn thành cơ bản một website thông tin phân tích các vụ tai nạn máy bay.

### **3.2 Ưu nhược điểm của đề tài**

### **3.2.1 Ưu điểm**

* Đề tài phân tích được trực quan nguồn dữ liệu, có tính thực tiễn.
* Áp dụng được các kiến thức đã học.
* Các biểu đồ trực quan, đa dạng, thể hiện rõ được mục đích phân tích.

### **3.2.2 Nhược điểm**

Nguồn dữ liệu là tương đối và khá cũ nên việc phân tích chưa đạt hiệu quả cao, tính chính xác của dữ liệu đến thời gian hiện tại là quá cũ chưa sát thực tế. Các đề xuất còn hạn chế.

### **3.3 Hướng phát triển**

Tìm hiểu và tạo thêm hàm cập nhật dữ liệu định kỹ để dữ liệu được chính xác và trực quan hơn. Đồng thời thêm cái filter, chart để phân tích, trực quan rõ hơn các dữ liệu để có thể đưa ra các nhận xét chính xác hơn.

**TÀI LIỆU THAM KHẢO**

[Introduction | Superset (apache.org)](https://superset.apache.org/docs/intro/)

<https://superset.apache.org/docs/installation/installing-superset-using-docker-compose/>

<https://superset.apache.org/docs/intro>

<https://viblo.asia/p/gioi-thieu-ve-rds-OeVKBJB0KkW>