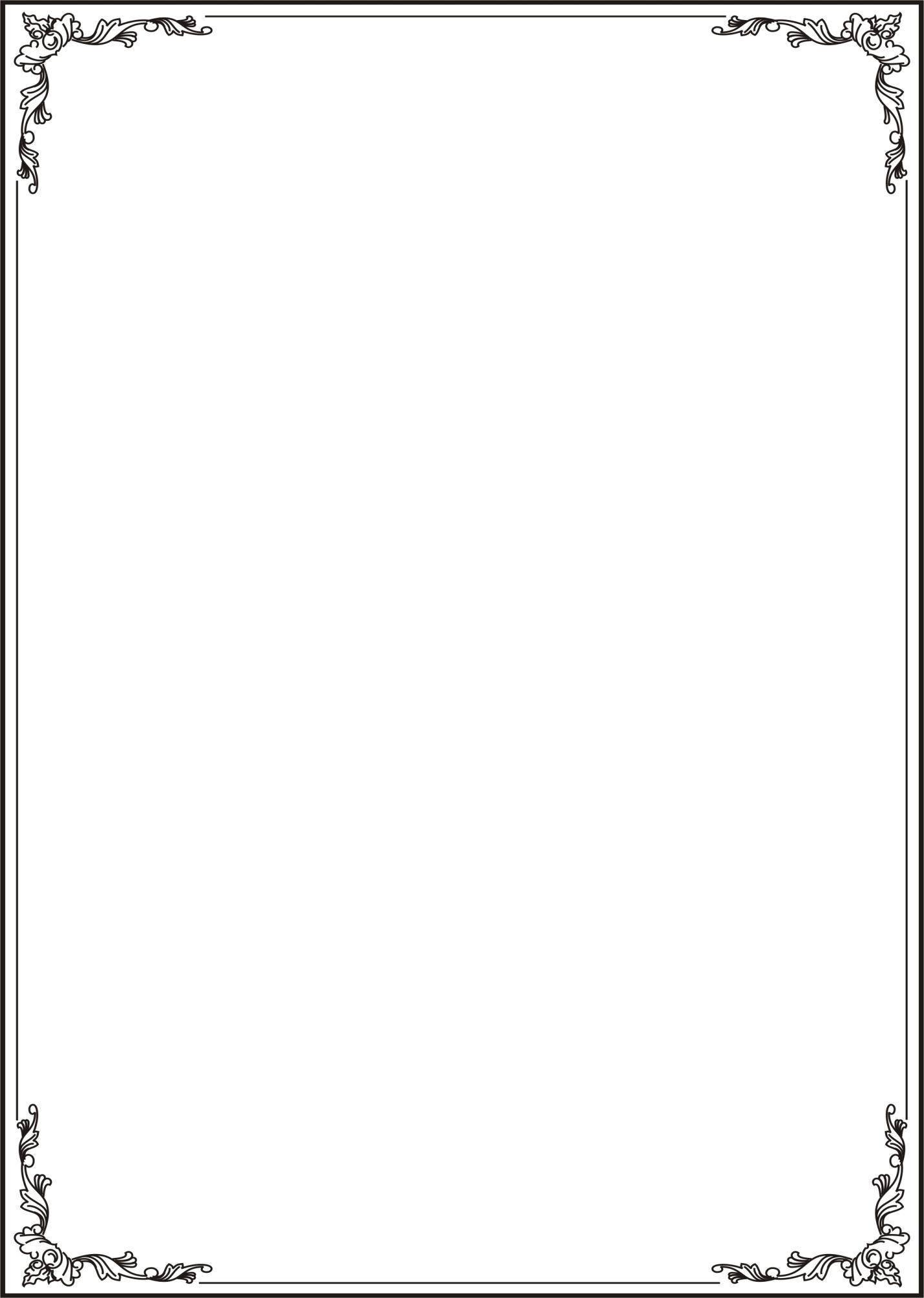
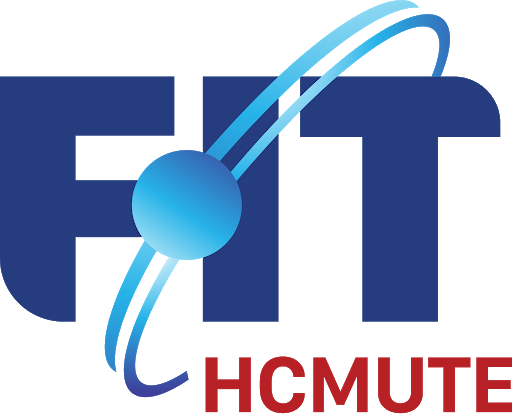
**.**

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

**KHOA CÔNG NGHỆ THÔNG TIN**

**BỘ MÔN TƯƠNG TÁC DỮ LIỆU TRỰC QUAN**

-----🙞🙜🕮🙞🙜-----

****

**Đề Tài:**

**SỬ DỤNG APACHE SUPERSET ĐỂ PHÂN TÍCH TRỰC QUAN CÁC ỨNG DỤNG TỪ GOOGLE PLAY**

**GIẢNG VIÊN HƯỚNG DẪN:** TS. HUỲNH XUÂN PHỤNG

**Sinh Viên Thực Hiện:**

Hồng Tiến Hào – 19133022

Dương Thanh Tuấn – 19133004

Nguyễn Thanh Tùng – 19133065

PHIẾU NHẬN XÉT CỦA GIÁO VIÊN HƯỚNG DẪN

Họ và tên Sinh viên 1: **Dương Thanh Tuấn** MSSV 1: **19133004**

Họ và tên Sinh viên 2: **Nguyễn Thanh Tùng** MSSV 2: **19133065**

Họ và tên Sinh viên 3: **Hồng Tiến Hào** MSSV 3: **19133022**

Ngành: **Kỹ Thuật Dữ Liệu**

Tên đề tài: **SỬ DỤNG APACHE SUPERSET ĐỂ PHÂN TÍCH ỨNG DỤNG TỪ GOOGLE PLAY**

Họ và tên Giáo viên hướng dẫn: **TS. Huỳnh Xuân Phụng**

**NHẬN XÉT**

1. Về nội dung đề tài & khối lượng thực hiện:

2. Ưu điểm:

3. Khuyết điểm:

4. Đề nghị cho bảo vệ hay không?

5. Đánh giá loại:

6. Điểm:

**LỜI CẢM ƠN**

Lời đầu tiên nhóm em xin phép được gửi lời cảm ơn chân thành đến với Khoa Công Nghệ Thông Tin – Trường Đại Học Sư Phạm Kỹ Thuật Thành Phố Hồ Chí Minh đã tạo điều kiện cho nhóm em được học tập và phát triển nền tảng kiến thức để thực hiện đề tài này.

Và bên cạnh đó nhóm chúng em xin gửi đến thầy Huỳnh Xuân Phụng lời cảm ơn tha thiết nhất. Trải qua một quá trình học tập và thực hiện đề tài trong thời gian qua, Thầy đã tận tâm chỉ bảo nhóm em trong suốt quá trình từ lúc bắt đầu đến lúc kết thúc đề tài này.

Dưới sự giảng dạy nhiệt tình và tận tâm của thầy Huỳnh Xuân Phụng, chúng emi đã học và hiểu những kiến ​​thức cơ bản về Apache Superset - một ứng dụng phần mềm mã nguồn mở để khai phá và trực quan hóa dữ liệu, có thể xử lý hàng petabyte dữ liệu từ biểu đồ hình tròn đơn giản đến biểu đồ không gian địa lý deck.gl rất chi tiết. Như vậy, chúng em đã biết cách cài đặt và sử dụng Apache Superset.

Tuy nhiên, lượng kiến ​​thức là vô hạn và khả năng có hạn, chúng em đã rất nỗ lực để thực hiện nó một cách tốt nhất có thể. Vì vậy, những sai sót là không thể tránh khỏi. Chúng em rất mong nhận được những ý kiến ​​đóng góp nhiệt tình của quý thầy(cô) để có thể rút kinh nghiệm và trau dồi, hoàn thiện và nâng cấp sản phẩm của mình một cách tốt nhất có thể.

Chúng em xin chân thành cảm ơn!

**Nhóm thực hiện**

Hồng Tiến Hào 19133022

Dương Thanh Tuấn 19133004

Nguyễn Thanh Tùng 19133065

| **BẢNG PHÂN CÔNG CÔNG VIỆC** | | | | |
| --- | --- | --- | --- | --- |
| STT | THÀNH VIÊN | MSSV | CÔNG VIỆC | MỨC ĐỘ  HOÀN THÀNH |
| 1 | HỒNG TIẾN HÀO | 19133022 | * tạo database kết nối superset * tiền xử lý dữ liệu bằng R * Vẽ biểu đồ chart * thiết kế cấu trúc dashboard | 100% |
| 2 | DƯƠNG THANH TUẤN | 19133004 | * Cập nhật dữ liệu * jinja template * phân quyền | 100% |
| 3 | NGUYỄN THANH TÙNG | 19133065 | * Viết báo cáo, slide * đánh giá dashboard, trả lời câu hỏi các nhóm bằng biểu đồ * Chọn biểu đồ * Vẽ biểu đồ chart, sql | 100% |

**MỤC LỤC**

[**PHẦN MỞ ĐẦU**](#_heading=h.2a4cqfnlcjxv) **6**

[1. Tính cấp thiết của đề tài](#_heading=h.tyjcwt) 6

[2. Đối tượng nghiên cứu](#_heading=h.3dy6vkm) 6

[3. Phạm vi nghiên cứu](#_heading=h.1t3h5sf) 6

[4. Kết quả dự kiến đạt được](#_heading=h.4d34og8) 6

[**CHƯƠNG 1: CƠ SỞ LÝ THUYẾT**](#_heading=h.pzenvewyuo0b) **8**

[Tổng quan về APACHE SUPERSET](#_heading=h.9xgjcvcxxza7) 8

[Các tính năng của APACHE SUPERSET ?](#_heading=h.6319lm4vuop5) 8

[Nền tảng APACHE SUPERSET](#_heading=h.p34midg1nswh) 8

[Các kiểu dữ liệu (data type) được hỗ trợ bởi superset](#_heading=h.gexx4o5syfo9) 9

[Các bước để trực quan hóa dữ liệu:](#_heading=h.jet5stpd1k9u) 9

[Các bước để chuyển đổi từ raw data sang visualization?](#_heading=h.9zlw70yivbw0) 9

[Cách xác định màu (color palette) trong các chart](#_heading=h.675olxh66kpk) 9

[Các giải pháp cho trường dữ liệu bị missing data](#_heading=h.cgjc5jid7ck1) 10

[Khái niệm về Data modeling](#_heading=h.41nzbdb5bwa9) **10**

[**CHƯƠNG 2: CÀI ĐẶT APACHE SUPERSET**](#_heading=h.otdf672g284t) **11**

[Cài Đặt Apache Superset](#_heading=h.1ksv4uv) 11

[Cài Đặt Một Database Kết Nối](#_heading=h.f03jm9yb5znz) 11

[Trang Chủ Apache Superset](#_heading=h.1cjbhhxc74un) 13

[Kết Nối Database và Upload Dataset](#_heading=h.ll082vw6cbxj) 13

[Biểu Đồ](#_heading=h.xc8ge8qw1l4v) 15

[Xử lý dữ liệu bằng R](#_heading=h.1ci93xb) 15

[Jinja Template](#_heading=h.3whwml4) 15

[Phân quyền trên Superset.](#_heading=h.wvroz6sdflfe) 16

[**CHƯƠNG 3: XÂY DỰNG DASHBOARD**](#_heading=h.jyjlo5vii82) **18**

[Cấu Trúc Dashboard](#_heading=h.rehm22i9g639) 18

[Tổng Quan Sự Phát Triển](#_heading=h.xsut91f8wc97) 18

[So Sánh Các Hạng Mục](#_heading=h.xfmy8zg4u6z3) 19

[Ứng Dụng Cho Một Số Người Dùng Nhất Định](#_heading=h.nfl68s96wk4k) 21

[Xu Hướng Ứng Dụng và Sự Hỗ Trợ Từ Nhà Phát Hành](#_heading=h.9r48wh8nk6ac) 21

[Khác](#_heading=h.9wgrhh40dllh) 22

[Dashboard](#_heading=h.8p08a0u2q53s) **23**

[**KẾT LUẬN**](#_heading=h.2vq23axt8tt4) **24**

[Kết quả đạt được](#_heading=h.qsh70q) 24

[Ưu điểm](#_heading=h.2p2csry) 24

[Nhược điểm](#_heading=h.147n2zr) 24

[Hướng phát triển](#_heading=h.3o7alnk) 24

[**TÀI LIỆU THAM KHẢO**](#_heading=h.23ckvvd) **25**

# 

# **PHẦN MỞ ĐẦU**

## **1. Tính cấp thiết của đề tài**

Trong thời đại công nghệ thông tin phát triển như vũ bão ngày nay, các bộ dữ liệu rất lớn và phức tạp với nhiều kiểu dữ liệu ngày càng trở nên phổ biến, tính thời gian thực ngày càng cao. Nếu giải quyết được các bộ dữ liệu lớn này, các nhà nghiên cứu có thể giải quyết được nhiều vấn đề trong xã hội và giáo dục, doanh nghiệp có thể nâng cao lợi nhuận của doanh nghiệp thông qua phân tích. Nhưng để có thể xử lý, phân tích và thực hiện trực quan hóa các tập dữ liệu lớn, chúng ta cần một ứng dụng để khám phá và hình dung quy mô terabyte của nó. Do đó, Apache Superset ra đời và trở thành một trong những ứng cử viên sáng giá nhất, vì đây là một ứng dụng web thông minh và sẵn sàng đáp ứng nhu cầu của doanh nghiệp. Có nhiều biểu đồ nhanh, nhẹ, linh hoạt và hỗ trợ tất cả người dùng thuộc mọi kỹ năng dễ dàng khám phá và hình dung dữ liệu.

## **2. Đối tượng nghiên cứu**

* Apache Superset: đối tượng nghiên cứu chính, phục vụ trực quan tập dữ liệu lớn.
* PostgreSQL: một hệ quản trị cơ sở dữ liệu quan hệ mã nguồn mở (RDBMS), phục vụ cho việc lưu trữ dữ liệu.
* RStudio: là một môi trường phát triển tích hợp (IDE) mã nguồn mở phục vụ mục đích phân tích và xử lý dữ liệu.
* Jinja Template: một công cụ tạo khuôn mẫu nhanh thuận tiện, phục vụ việc lọc dữ liệu và hỗ trợ truy vấn trên tập dữ liệu. .

## **3. Phạm vi nghiên cứu**

Sử dụng Superset trên EC2 của AWS Cloud để trực quan hoá tập dữ liệu Google Play Store

## **4. Kết quả dự kiến đạt được**

* Cài đặt Apache Superset
* Tải tập dữ liệu đã qua xử lý lên nền superset
* Trực quan hóa và truy vấn trên tập dữ liệu.
* Tạo ra ứng dụng tự động cập nhật tập dữ liệu theo thời gian thực.

# 

# **CHƯƠNG 1: CƠ SỞ LÝ THUYẾT**

## **Tổng quan về APACHE SUPERSET**

Apache Superset đã ra đời và trở thành 1 trong những ứng cử viên sáng giá vì nó là ứng dụng web thông minh và sẵn sàng để phục vụ cho các nhu cầu của doanh nghiệp. Với lợi ích nhanh, nhẹ, đa dạng với nhiều biểu đồ và hỗ trợ cho tất cả người dùng thuộc tất cả các bộ kỹ năng dễ dàng khám phá và hình dung dữ liệu của họ, từ biểu đồ realtime series, đến bar chart rất chi tiết.

Một số cách khác nhau để có thể bắt đầu cài đặt và sử dụng với Superset :

* Tải xuống [từ trang web của Apache Foundation](https://dist.apache.org/repos/dist/release/superset/1.4.1/)
* Tải xuống phiên bản Superset mới nhất từ [​​Pypi tại đây](https://pypi.org/project/apache-superset/)
* Thiết lập Superset cục bộ bằng một lệnh sử dụng [Docker Compose](https://superset.apache.org/docs/installation/installing-superset-using-docker-compose)
* Cài đặt phiên bản mới nhất của Superset [từ GitHub](https://github.com/apache/superset/tree/latest)

## **Các tính năng của APACHE SUPERSET ?**

* Tạo, xử lí và quản lý dashboard
* Xác thực với các hệ thống xác thực của doanh nghiệp (OpenID, OAuth,LDAP...)
* Hỗ trợ nhiều loại cơ sở dữ liệu như: SQL,MySQL, Postgres,...
* Hỗ trợ truy vấn trực tiếp dữ liệu trên dashboard

## **Nền tảng APACHE SUPERSET**

Apache Superset là nền tảng đám mây, được thiết kế để mở rộng ra các môi trường lớn. Mặc dù có thể dễ dàng kiểm tra ổ đĩa Superset trên máy tính xách tay của mình, nhưng hầu như không có giới hạn về việc mở rộng nền tảng.

Superset linh hoạt và cho phép chọn:

* Máy chủ web (Nginx, Apache, Gunicorn),
* Công cụ cơ sở dữ liệu siêu dữ liệu (Postgres, MariaDB,MySQL, v.v.),
* Hàng chờ tin nhắn (RabbitMQ, SQS, Redis, v.v.),
* Phụ trợ kết quả (Redis, Memcached, S3, v.v.),
* Lớp bộ nhớ đệm (Redis, Memcached, v.v.),

Superset vẫn hoạt động tốt với các dịch vụ như StatsD,NewRelic, và DataDog, đồng thời có khả năng chạy khối lượng lớn công việc phân tích và trực quan dựa trên các công nghệ cơ sở dữ liệu phổ biến và hiện đang được sử dụng trên quy mô lớn tại nhiều công ty.

## **Các kiểu dữ liệu (data type) được hỗ trợ bởi superset**

* TEXT
* BIGINT
* TIMESTAMP WITHOUT TIME ZONE)
* TIMESTAMP (TIME ZONE)
* DOUBLE PRECISION
* VARCHAR

## **Các bước để trực quan hóa dữ liệu:**

* Bước 1. Xác định câu hỏi
* Bước 2. Thu thập dữ liệu
* Bước 3. làm sạch dữ liệu
* Bước 4. Tìm biểu đồ thích hợp
* Bước 5. chọn công cụ
* Bước 6. Xác định các hàm thống kê đưa vào biểu đồ
* Bước 7. Tạo biểu đồ

## **Các bước để chuyển đổi từ raw data sang visualization?**

* Bước 1: Thu thập dữ liệu (Data Source)
* Bước 2: Tổng hợp dữ liệu thành dữ liệu thô (dữ liệu chưa được xử lý) (Raw Data)
* Bước 3: Xử lý dữ liệu thô thành dữ liệu đầu vào cho chương trình (Data) Bước 4: Phân tích dữ liệu (View)
* Bước 5: Mô tả dữ liệu thành dạng bảng, đồ thị… (Graph)
* Bước 6: Thiết kế bảng điều khiển (Dashboard)

## **Cách xác định màu (color palette) trong các chart**

Ba loại bảng màu chính để trực quan hóa dữ liệu:

* Bảng màu định tính: Một bảng màu định tính được sử dụng khi biến có thể phân loại trong tự nhiên. Các biến phân loại là các biến có nhãn riêng biệt mà không cần đặt hàng vốn có. Các ví dụ bao gồm quốc gia hoặc nhà nước, chủng tộc và giới tính. Mỗi giá trị có thể của biến được gán một màu từ bảng định tính.
* Bảng màu tuần tự: Khi biến được gán để được tô màu là số hoặc đã sắp xếp, thì nó có thể được mô tả bằng bảng màu tuần tự. Màu sắc được gán cho các giá trị dữ liệu theo liên tục, thường dựa trên độ sáng, màu sắc hoặc cả hai.
* Bảng màu phân kỳ: Nếu biến số của ta có giá trị trung tâm có ý nghĩa, như 0, thì chúng tôi có thể áp dụng một bảng màu phân kỳ. Một bảng màu chuyển hướng về cơ bản là sự kết hợp của hai bảng màu tuần tự với điểm cuối được chia sẻ ngồi ở giá trị trung tâm. Các giá trị lớn hơn trung tâm được gán cho màu ở một bên của trung tâm, trong khi các giá trị nhỏ hơn được gán cho màu ở phía đối diện

→ Loại bảng màu mà bạn sử dụng trong một hình ảnh phụ thuộc vào bản chất của dữ liệu được ánh xạ thành màu.

## **Các giải pháp cho trường dữ liệu bị missing data**

* Xóa hết tất cả dòng chứa missing data
* Thay thể missing data bằng 1 giá trị (ví dụ như mean hay median...)
* Thu thập dữ liệu missing data lại

## **Khái niệm về Data modeling**

Là quá trình tạo ra một mô hình dữ liệu để dữ liệu được lưu trữ trong cơ sở dữ liệu. Mô hình dữ liệu này là một biểu diễn khái niệm của các đối tượng dữ liệu, sự liên kết giữa các đối tượng dữ liệu khác nhau và các quy tắc. Mô hình hóa dữ liệu giúp thể hiện trực quan dữ liệu và thực thi các quy tắc kinh doanh, tuân thủ quy định và chính sách của chính phủ về dữ liệu.

Mô hình dữ liệu đảm bảo tính nhất quán trong quy ước đặt tên, giá trị mặc định, ngữ nghĩa, bảo mật đồng thời đảm bảo chất lượng của dữ liệu. Mục đích của data modeling là tạo ra phương pháp hiệu quả nhất để lưu trữ thông tin, đông thời vẫn vẫn cung cấp các quy trình truy cập và báo cáo hoàn chỉnh

# 

# **CHƯƠNG 2: CÀI ĐẶT APACHE SUPERSET**

## **Cài Đặt Apache Superset**

* Cài Đặt Apache Superset theo hưởng dẫn sau: <https://youtu.be/X6aG396HTnc>
* Đọc thêm tại: <https://superset.apache.org/docs/installation/installing-superset-using-docker-compose/>

## **Cài Đặt Một Database Kết Nối**

1. Cài đặt PostgreSQL:

Chúng ta truy cập vào <https://www.postgresql.org/download/>. Tải phiên bản mới nhất

Timeline

Description automatically generated

1. Cài đặt Pgadmin 4:

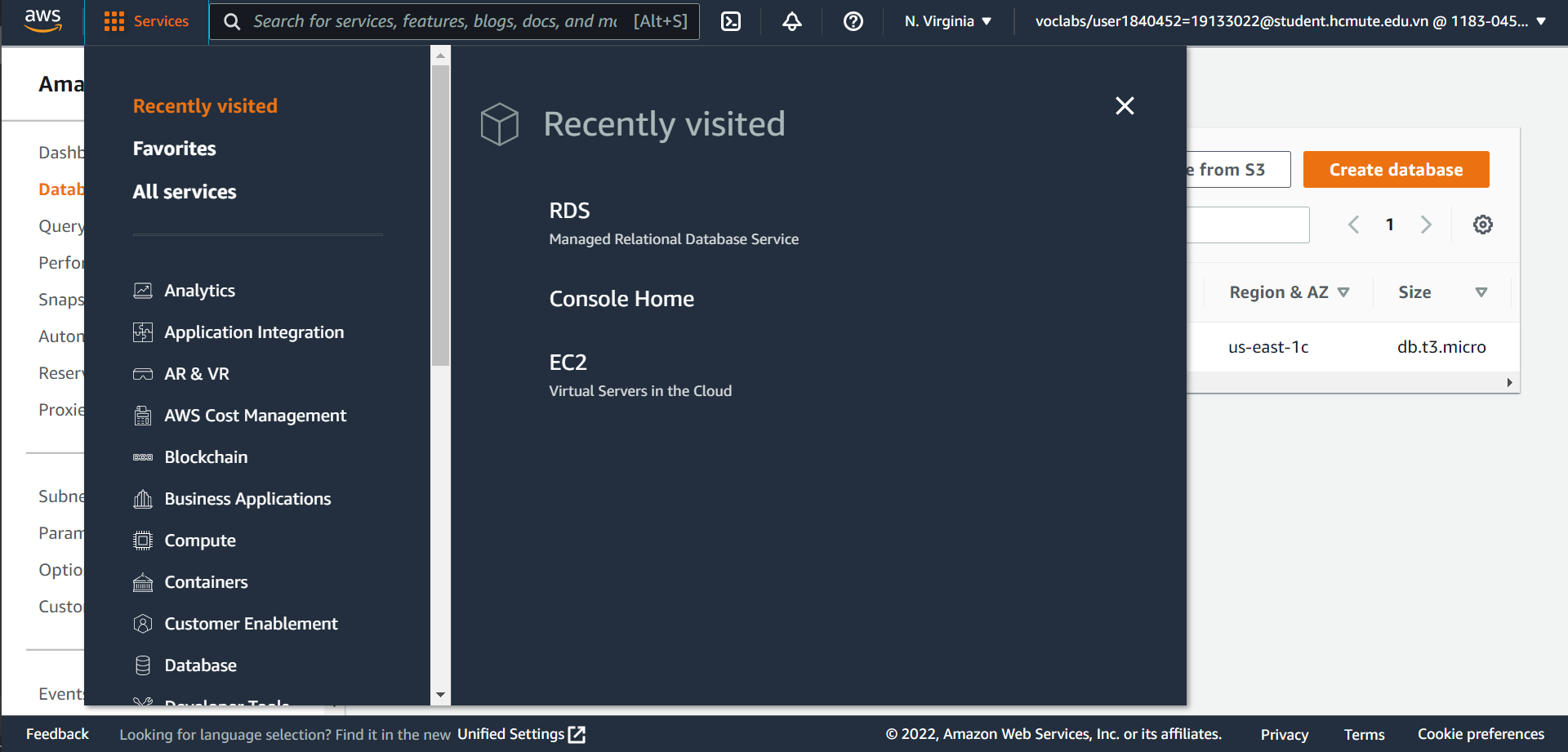
Chúng ta truy cập vào: <https://www.pgadmin.org/download/>

Graphical user interface, text, application, email

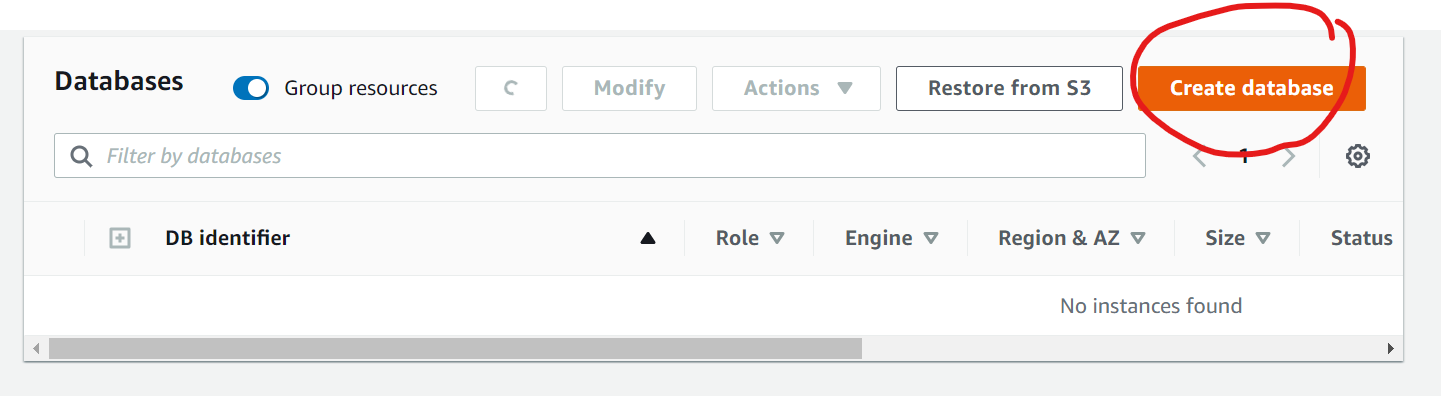
Description automatically generated

1. Tạo cơ sở dữ liệu PostgreSQL trên AWS RDS:

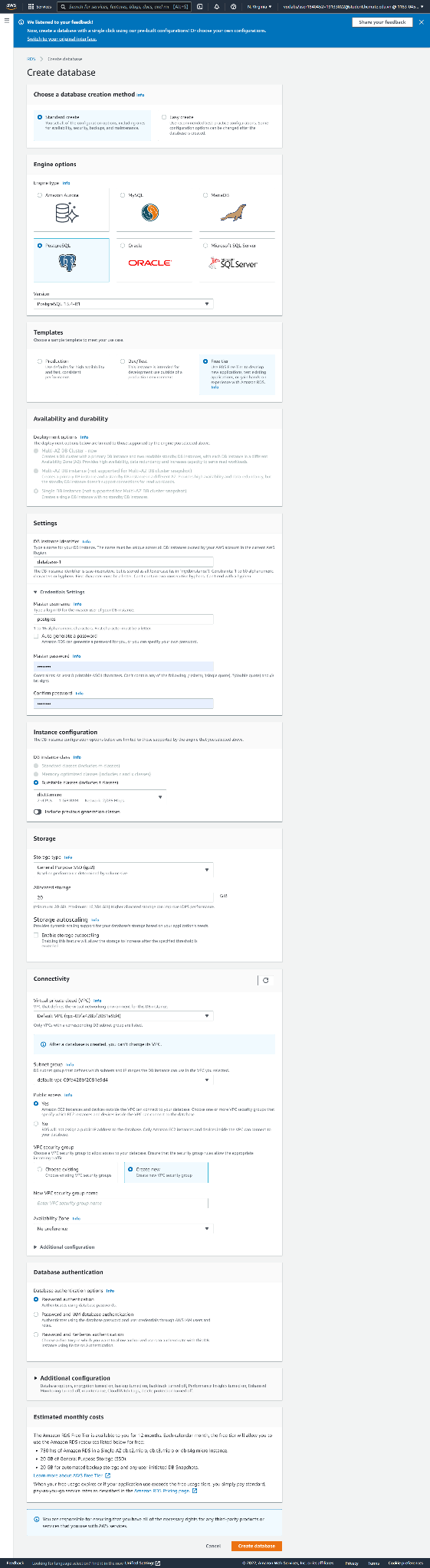
* Chúng ta vào service và chọn RDS.



* Tạo database.



* Chọn PostgreSQL
* Chọn Templates là Free tier.
* Đặt tên Database
* Chọn DB instance class và Storage.
* Bỏ tích Storage autoscaling.
* Chọn Create database để tạo cơ sở dữ liệu.
* Sau khi tạo ta có database test.
* Giờ ta vào database test vừa tạo, sau đó vào vpc security groups của test.
* Chọn phần inbound rules.
* Chọn edit inbound rules.
* Thêm rule và thiết lập cổng port là 5432. Sau đó save rules

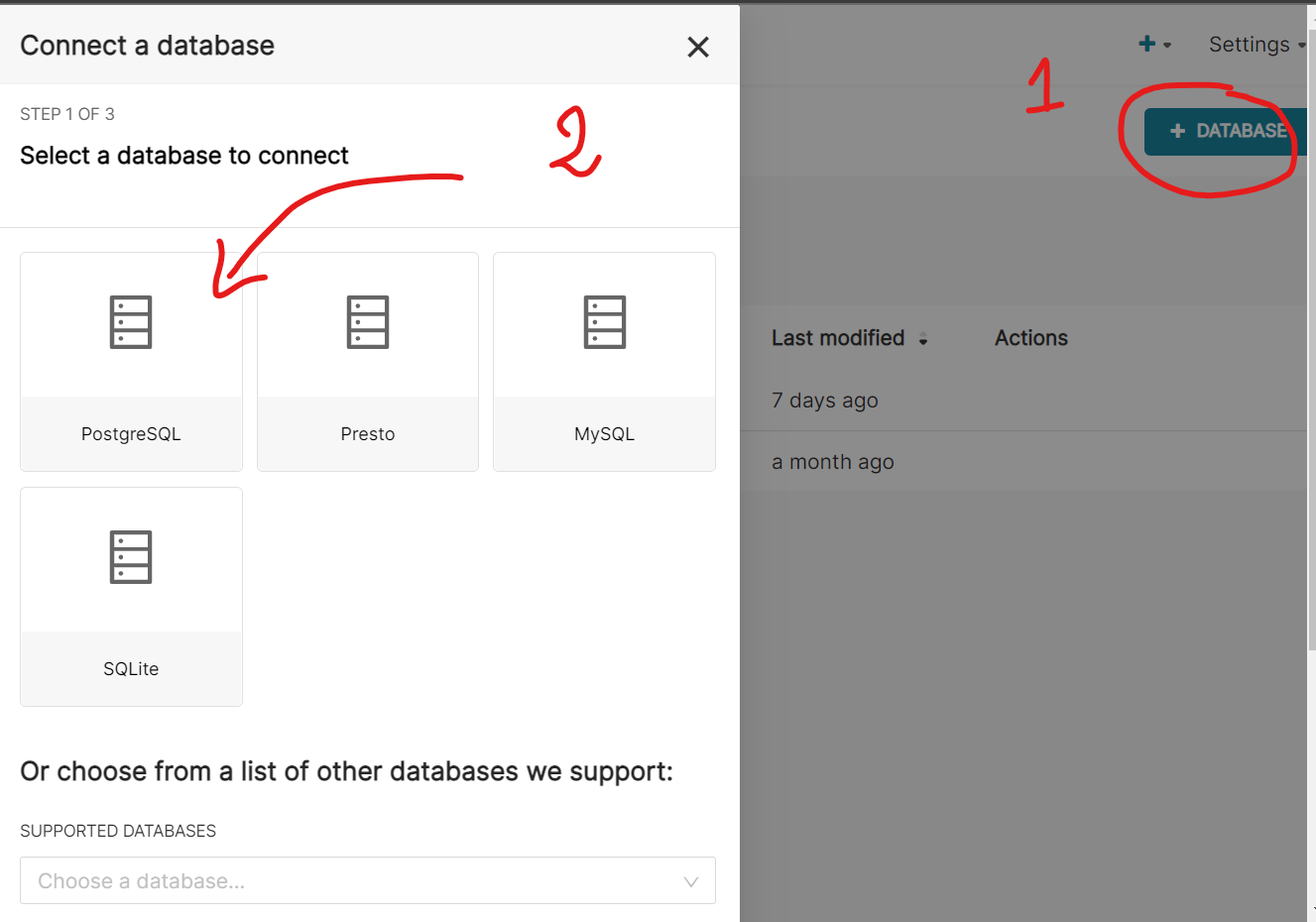


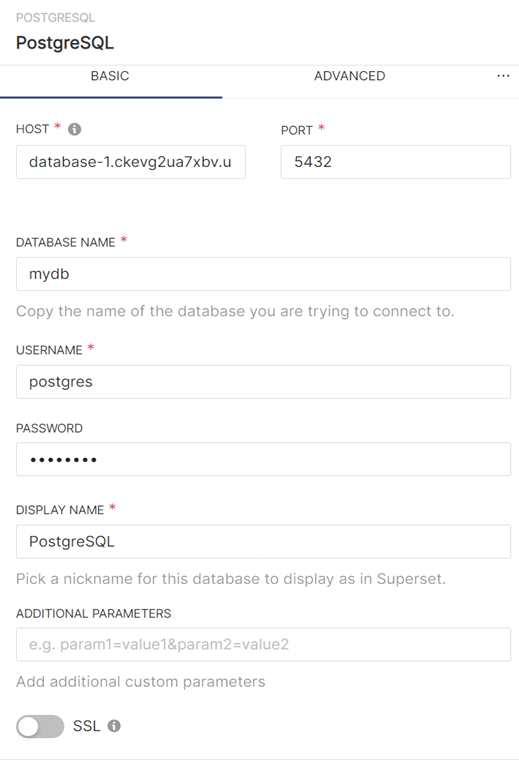
## **Trang Chủ Apache Superset**



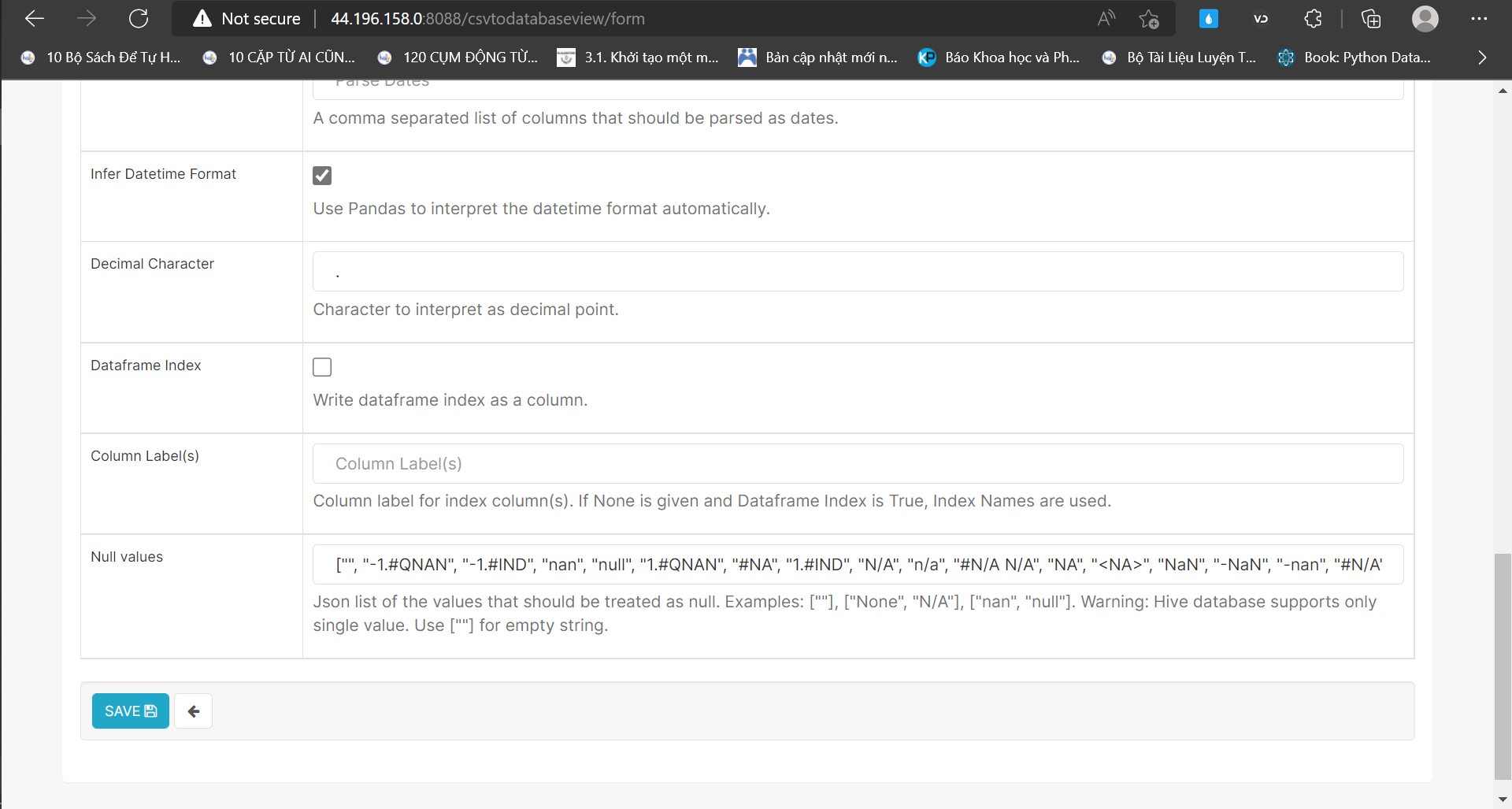
## **Kết Nối Database và Upload Dataset**

Apache Superset không lưu trữ dữ liệu của chúng ta cần phân tích nên cần phải kết nối đến một database.





Sau khi kết nối database, điền đầy đủ thông tin để upload csv.



## **Biểu Đồ**

Một vài biểu đồ trên Apache Superset hỗ trợ



## **Xử lý dữ liệu bằng R**

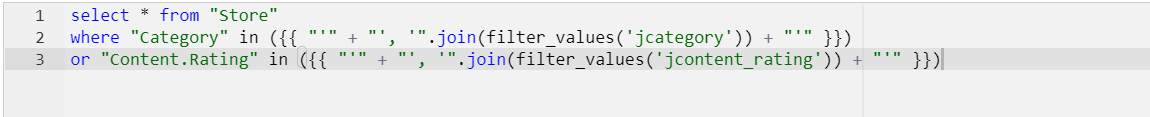
- Chúng ta sẽ sử dụng ngôn ngữ R trên phần mềm R Studio để tiến hành xử lý dữ liệu.

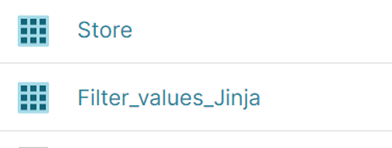
Graphical user interface, text

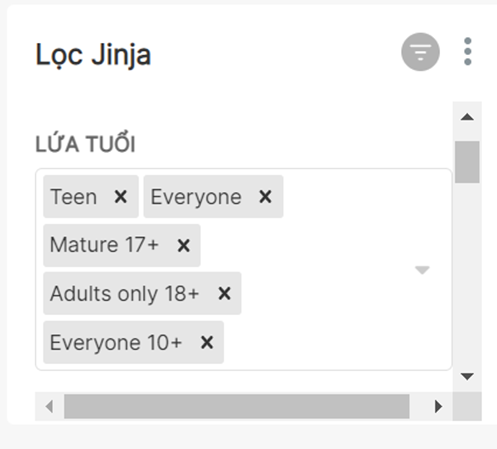
Description automatically generated

## **Jinja Template**

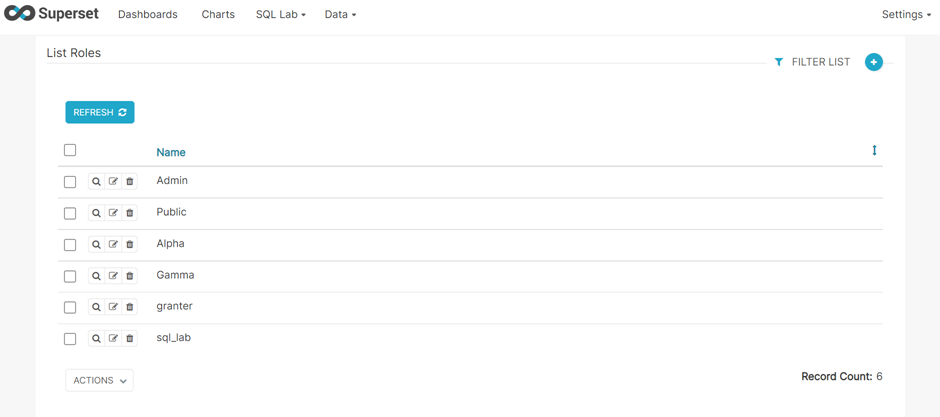
* Sử dụng Jinja để tạo Filter Box cho Dashboard.



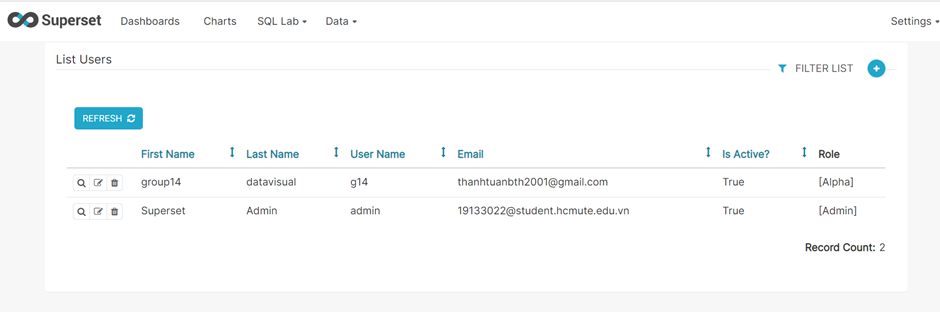




## **Phân quyền trên Superset.**



Gồm các quyền như admin, public, alpha, gamma, granter và cuối cùng là sql\_lab. Ở đây sẽ sử dụng quyền alpha để cho người khác có thể tham khảo dashboard và đưa ra đánh giá nhận xét



Tạo thêm 1 account với username là g14 và password là g14 để cho mọi người có thể truy cập vào dashboard của nhóm

# 

# **CHƯƠNG 3: XÂY DỰNG DASHBOARD**

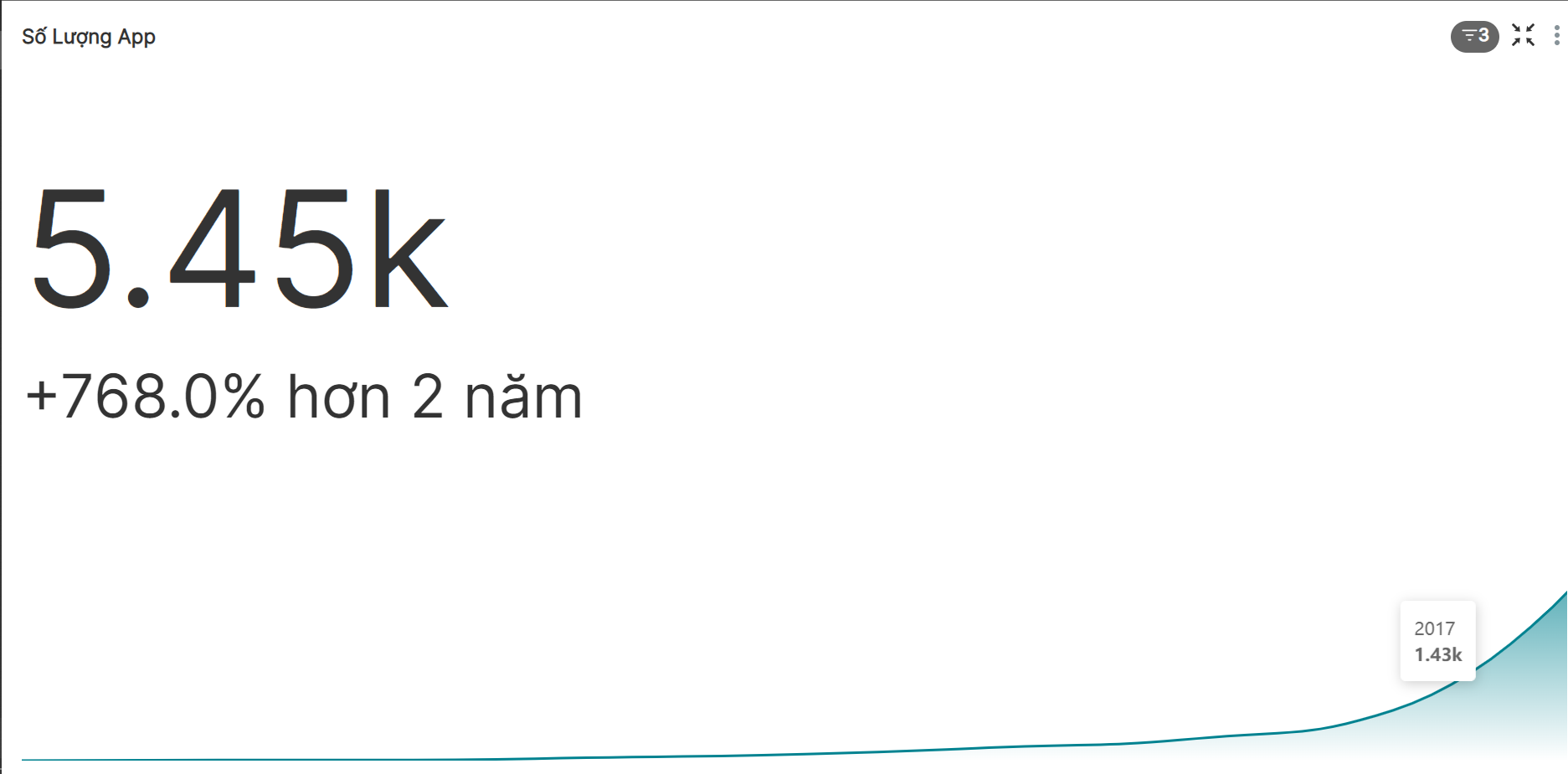
## **Cấu Trúc Dashboard**

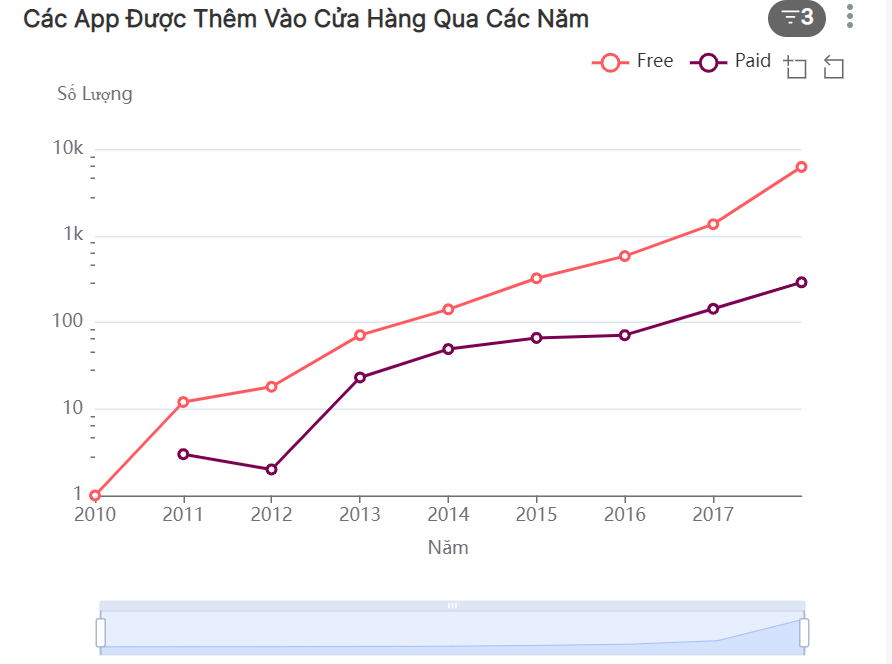
Cấu trúc Dashboard được chia thành 4 phần sau để trực quan:

* Tổng quan sự phát triển của ứng dụng theo thời gian (miễn phí và trả phí)
* Các hạng mục ứng dụng (so sánh số lượng, đánh giá, lượt tải…)
* Các ứng dụng tập trung vào những người dùng nhất định
* Xu hướng ứng dụng và vai trò của nhà phát hành

## **Tổng Quan Sự Phát Triển**

Cửa hàng trực tuyến Google Play Store ra mắt với hình thức miễn phí vào năm 2008 (2009 với ứng dụng trả phí), số lượng ứng dụng tăng dần qua các năm và bùng nổ kể từ 2016. Những ứng dụng trả phí có lẽ không mang lại nhiều lợi ích cho nhà phát hành, ta thấy số lượng các ứng dụng trả phí phát triển rất là khiêm tốn.

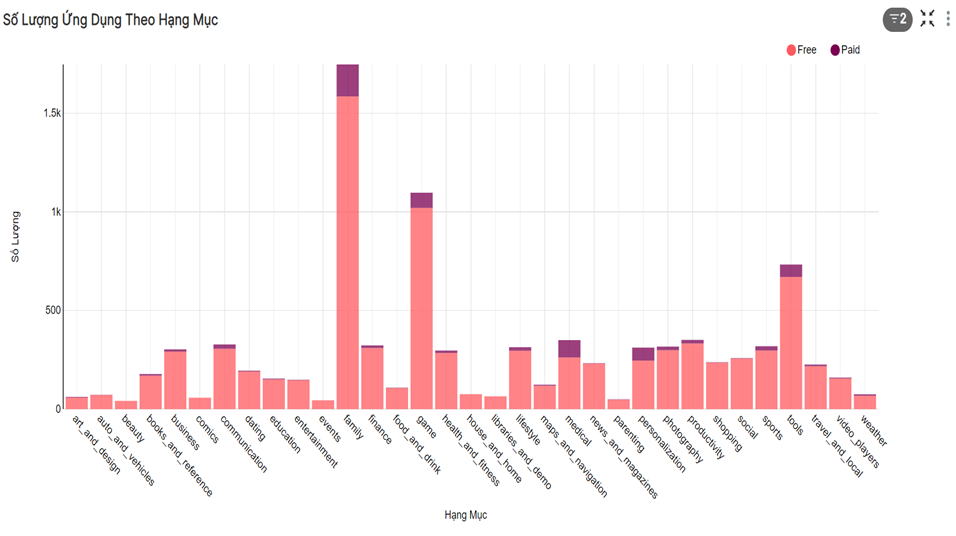




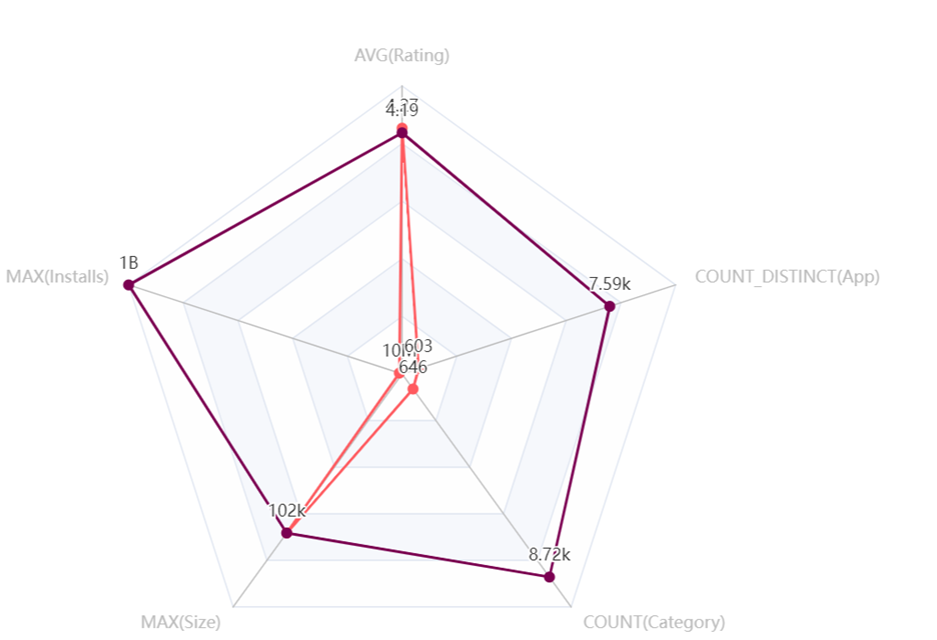
## **So Sánh Các Hạng Mục**

Tuy nhiên khi số lượng Apps tăng đáng kể, thì để trở nên được phổ biến đòi hỏi phải có sự khác biệt. Hàng loạt các danh mục ứng dụng đa dạng nổi lên với mục đích làm hài lòng người dùng. Trong 3 ứng dụng nhiều nhất được phát hành trên cửa hàng ta thấy nó thuộc về family, game và tools.

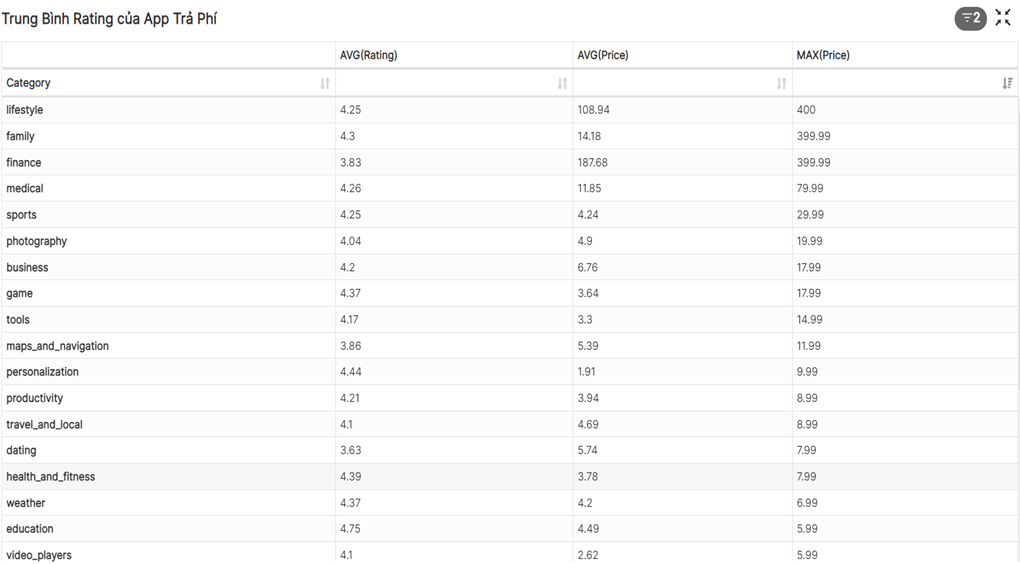




Các ứng dụng trả phí cũng là một định hướng cho các nhà phát triển. Nhưng cần cân nhắc, vì có vẻ như các Apps với mức giá **càng cao** thì mong muốn từ người dùng cũng sẽ tăng và **Rating sẽ thấp nếu không đáp ứng được yêu cầu của họ**.

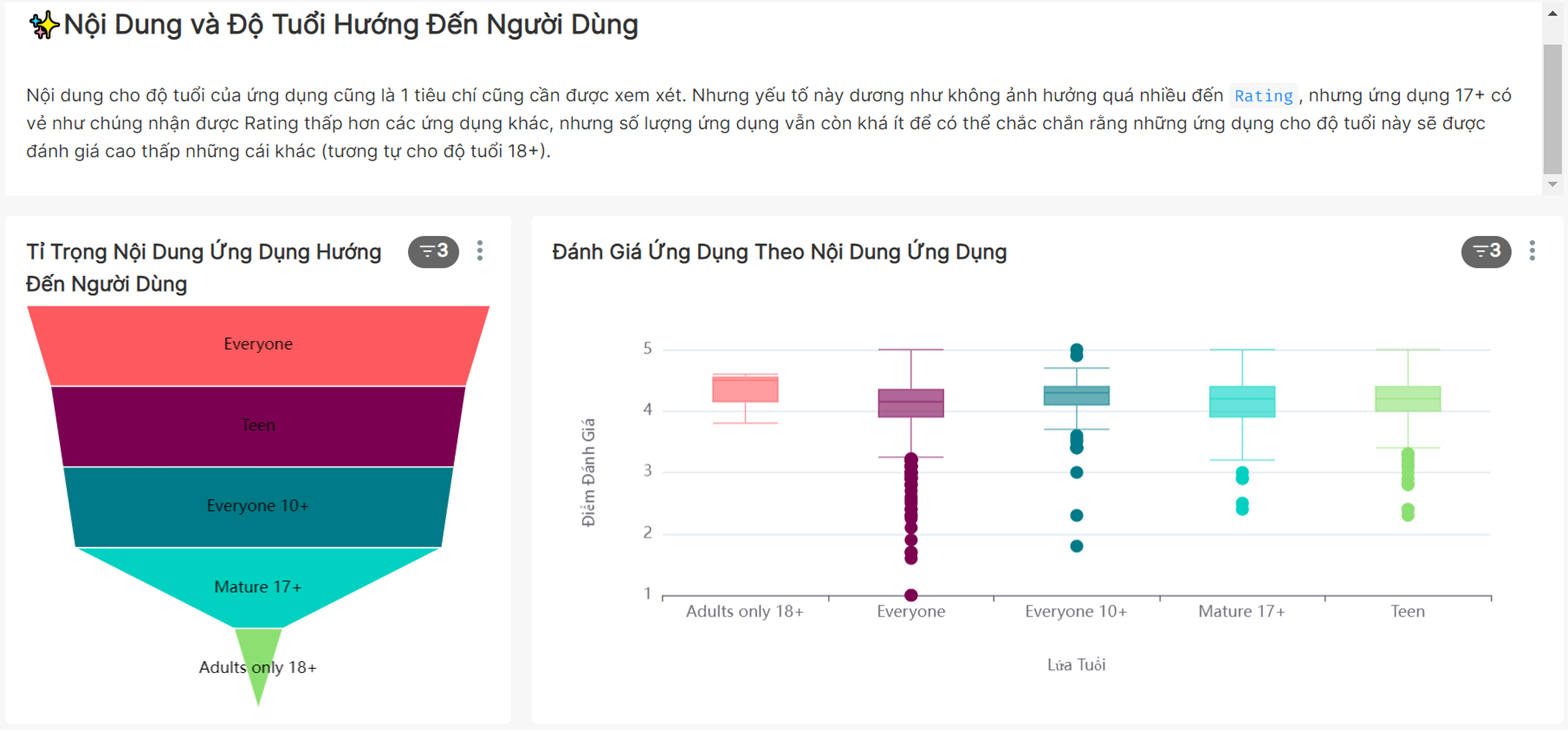
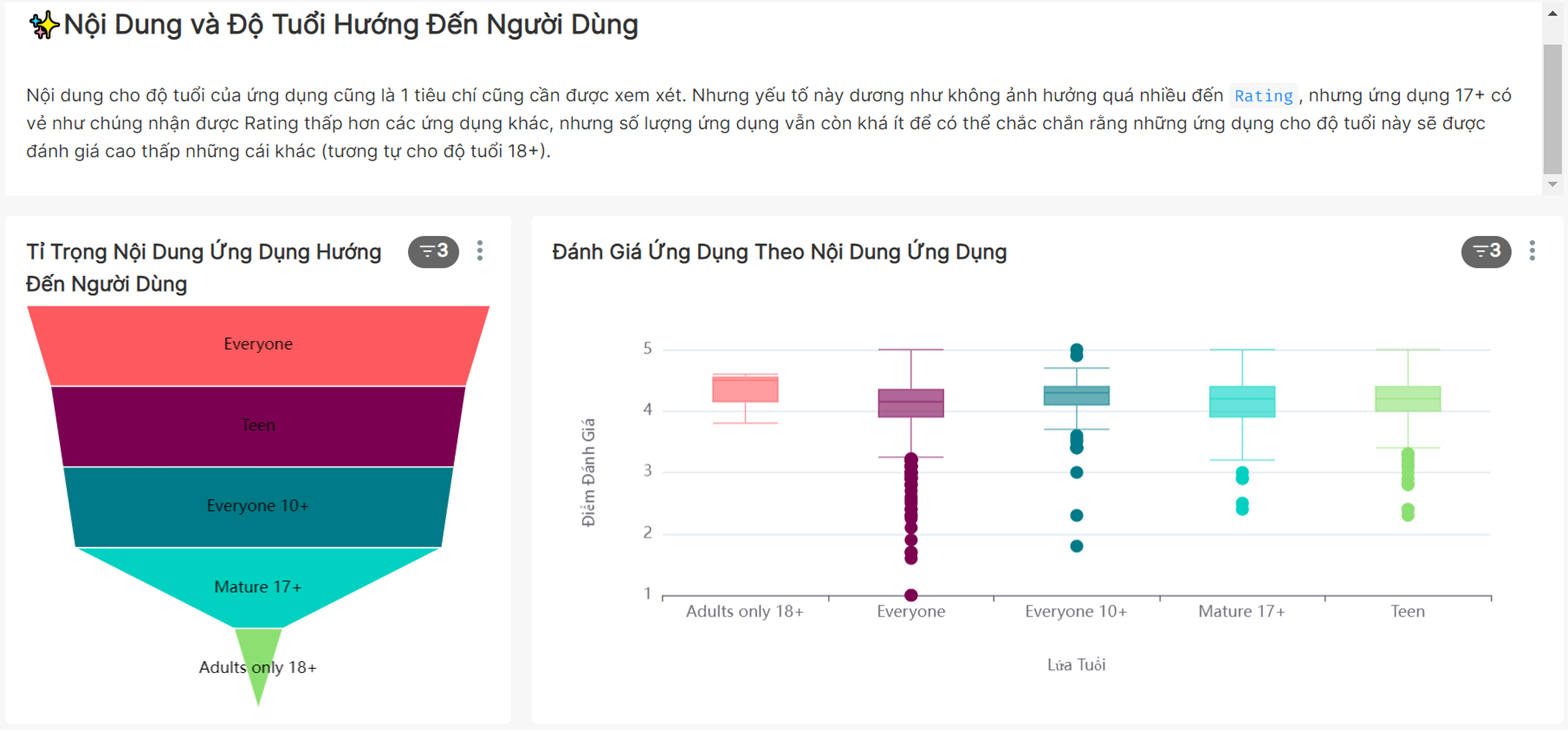


Nhưng **Rating trung bình không khác biệt nhiều**. Có vẻ như điểm đánh giá không đến từ các danh mục ứng dụng.



## **Ứng Dụng Cho Một Số Người Dùng Nhất Định**

Nội dung cho độ tuổi của ứng dụng cũng là 1 tiêu chí cũng cần được xem xét. Nhưng yếu tố này dương như không ảnh hưởng quá nhiều đến Rating, nhưng ứng dụng 17+ có vẻ như chúng nhận được Rating thấp hơn các ứng dụng khác, nhưng số lượng ứng dụng vẫn còn khá ít để có thể chắc chắn rằng những ứng dụng cho độ tuổi này sẽ được đánh giá cao thấp những cái khác (tương tự cho độ tuổi 18+).

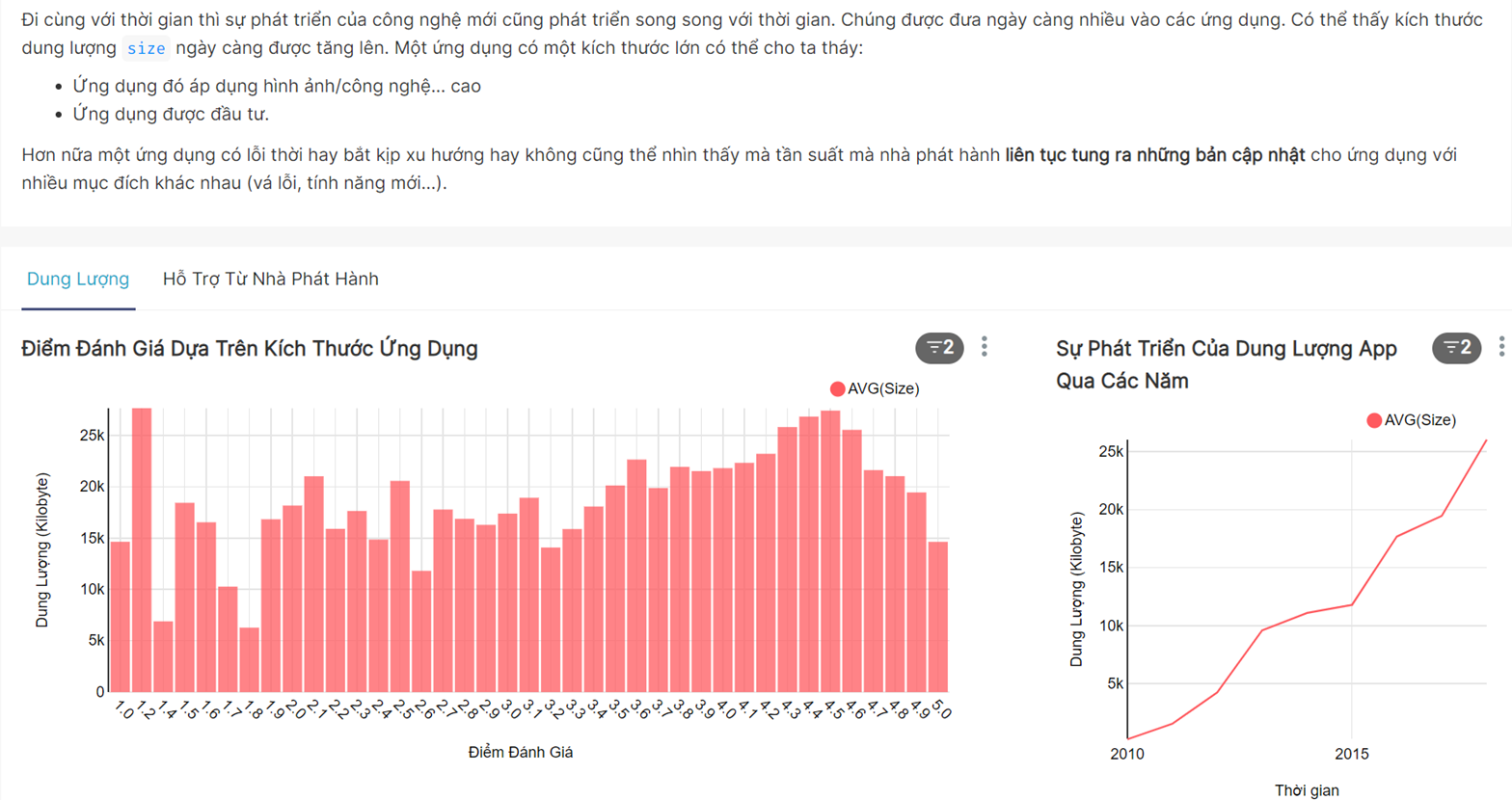


## **Xu Hướng Ứng Dụng và Sự Hỗ Trợ Từ Nhà Phát Hành**

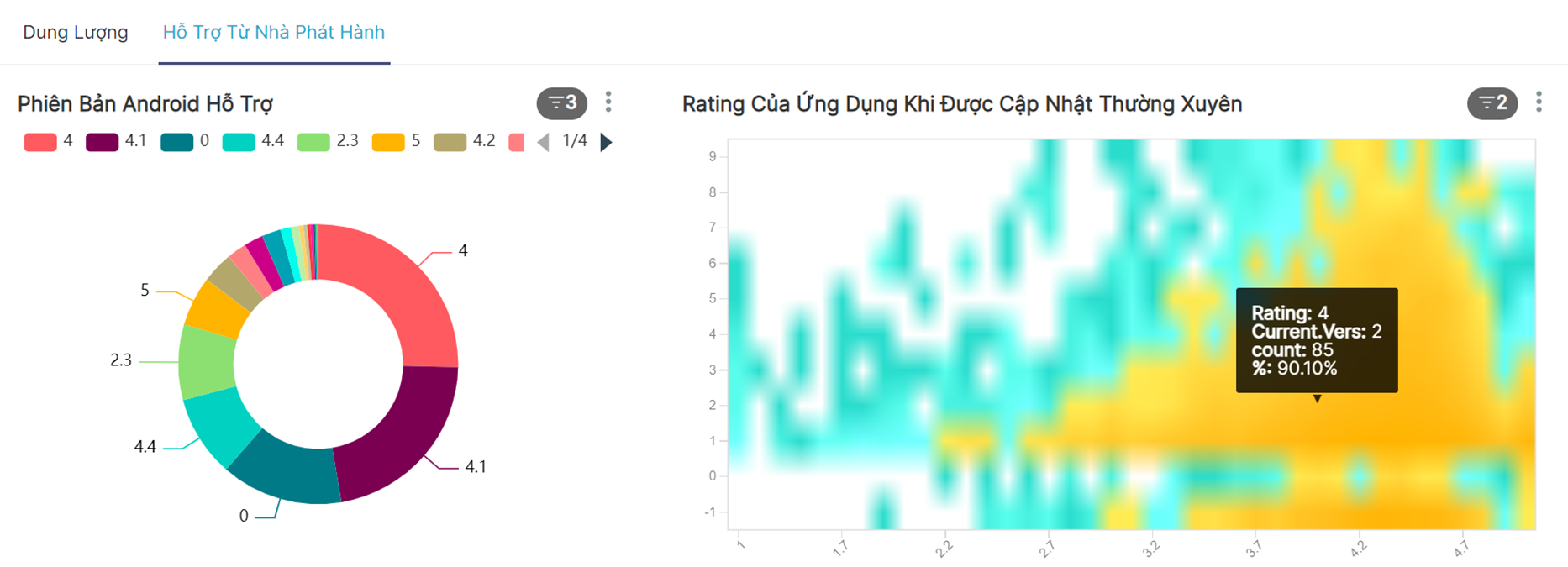
Đi cùng với thời gian thì sự phát triển của công nghệ mới cũng phát triển song song với thời gian. Chúng được đưa ngày càng nhiều vào các ứng dụng. Có thể thấy kích thước dung lượng size ngày càng được tăng lên. Một ứng dụng có một kích thước lớn có thể cho ta tháy:

* Ứng dụng đó áp dụng hình ảnh/công nghệ... cao
* Ứng dụng được đầu tư.

Hơn nữa một ứng dụng có lỗi thời hay bắt kịp xu hướng hay không cũng thể nhìn thấy mà tần suất mà nhà phát hành liên tục tung ra những bản cập nhật cho ứng dụng với nhiều mục đích khác nhau (vá lỗi, tính năng mới...).

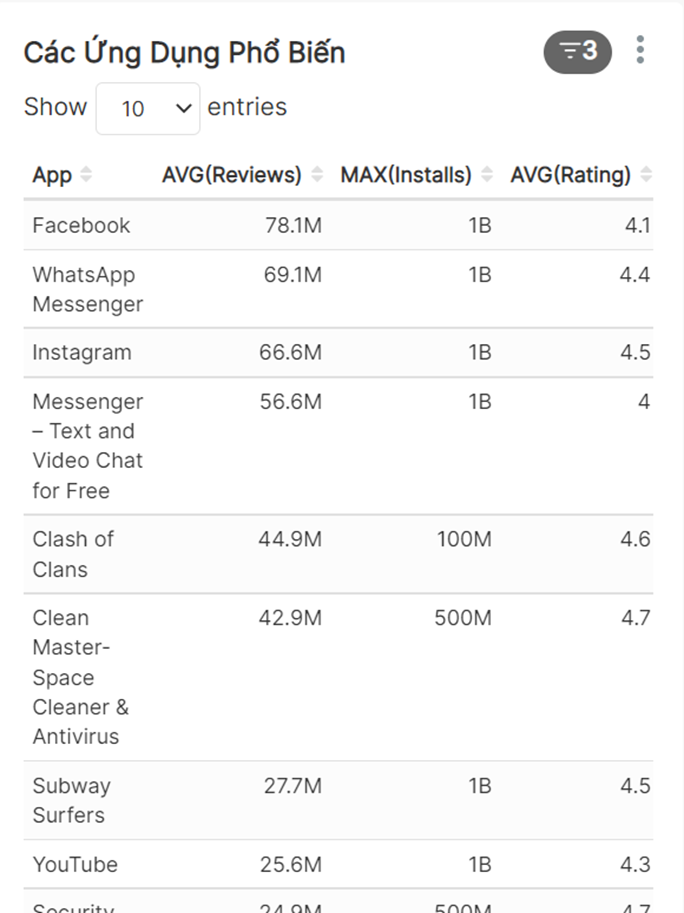


Ứng dụng ngày càng có xu hướng gia tăng dung lượng, và lượng ứng dụng có kích thước lớn cho điểm đánh giá trung bình cao hơn mộ số ứng dụng khác



* Ảnh trái cho ta thấy số lượng phiên bản android mà ứng dụng được hỗ trợ
  + 0 tương ứng với ứng dụng hỗ trợ toàn bộ phiên bản android
* Ảnh phải cho biết sự tương quan giữa việc cập nhật phiên bản thường xuyên sẽ ảnh hưởng đến điểm đánh giá như thế nào

## **Khác**



## **Dashboard**

Chart

Description automatically generated

# **KẾT LUẬN**

## **Kết quả đạt được**

Sau một thời gian nghiên cứu và thực hiện đề tài “***Tìm hiểu và sử dụng Apache Superset***”, nhóm chúng em đã đạt được những kết quả như sau:

1. *Kiến thức tìm hiểu được*

Nắm bắt được các kiến thức cũng như những vấn đề liên quan trọng về Apache Superset. Biết được cách tạo các biểu đồ trực quan một cách hệ thống, biết tải dữ liệu lên Superset, hiểu được công dụng và nền tảng của nó. Thực hiện truy vấn với sql và áp dụng kiến thức để thiết kế và trực quan hóa các biểu đồ trên Superset sử dụng SQL.

1. *Chương trình đã làm được*

* Cài đặt và sử dụng Superset.
* Sử dụng PosgreSQL thông qua RDS của amazon web service
* Tạo biểu đồ trực quan hóa, cũng như tạo các filter cho việc trực quan được rõ ràng, cụ thể hơn.
* Tiền xử lý bằng R.
* Sử dụng sql để thực hiện truy vấn và tạo biểu đồ.
* Sử dụng jinja code để lọc các dữ liệu.

## **Ưu điểm**

* Dashboard được tổ chức hợp lý theo từng mục dễ quan sát
* Biểu đồ đa dạng
* Có một vài biểu đồ được tạo từ code SQL.
* Tập dữ liệu đã được xử lý trước bằng R
* Áp dụng jinja cho filter

## **Nhược điểm**

* Tập dữ liệu còn khá ít
* Superset chưa hỗ trợ tốt một số biểu đồ

## **Hướng phát triển**

* Tiếp tục hoàn thiện các chức năng còn thiếu.
* Cập nhật thêm tập dữ liệu

# **TÀI LIỆU THAM KHẢO**

https://www.youtube.com/watch?v=dE1QlhunWMs

GooglePlay Store App https://www.kaggle.com/datasets/lava18/google-play-store-apps?fbclid=IwAR2LOWLtuWVJQ3-cAsCXP\_OffN\_lMAzshCL-Wu6yT1HhAPtNcXTovMIb3yE