

ĐẠI HỌC QUỐC GIA TP HCM
TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN
KHOA CÔNG NGHỆ THÔNG TIN



Báo cáo Đồ án Cuối kỳ

ĐỀ TÀI: PHÂN TÍCH ĐIỂM THI, ĐIỂM CHUẨN CÁC PHƯƠNG
THỨC XÉT TUYỂN ĐẠI HỌC TẠI VIỆT NAM QUA CÁC NĂM

Môn học: Trực quan hóa dữ liệu

Nhóm thực hiện:

Nhóm 10

Giảng viên hướng dẫn:

TS. Bùi Tiến Lên

Ngày 13 tháng 7 năm 2024

Mục lục

| | |
|---|-----------|
| Tổng quan đồ án | 4 |
| Thông tin thành viên | 4 |
| Yêu cầu đồ án | 4 |
| Yêu cầu về dữ liệu | 4 |
| Chủ đề của nhóm | 4 |
| Lời cảm ơn | 4 |
| 1 Giới thiệu về bộ dữ liệu | 6 |
| 1.1 Dữ liệu điểm thi | 6 |
| 1.1.1 Ngữ cảnh | 6 |
| 1.1.2 Nguồn dữ liệu | 6 |
| 1.1.3 Cách thức thu thập dữ liệu | 6 |
| 1.1.4 Mô tả dữ liệu | 7 |
| 1.1.5 Xử lý dữ liệu | 8 |
| 1.2 Dữ liệu điểm chuẩn | 11 |
| 1.2.1 Ngữ cảnh | 11 |
| 1.2.2 Nguồn dữ liệu | 11 |
| 1.2.3 Cách thức thu thập dữ liệu | 11 |
| 1.2.3.1 Cách thức truy cập và tải dữ liệu | 11 |
| 1.2.3.2 Các bước thực hiện cụ thể | 12 |
| 1.2.4 Mô tả dữ liệu | 13 |
| 1.2.5 Xử lý dữ liệu | 13 |
| 2 Phân tích Dashboard | 15 |
| 2.1 Tổng quan | 15 |
| 2.1.1 Đối tượng | 15 |
| 2.1.2 Mục tiêu | 15 |
| 2.1.3 Các dashboard thành phần | 16 |
| 2.2 Phân tích Dashboard "Điểm thi theo môn" | 16 |
| 2.2.1 Mục tiêu dashboard | 16 |

| | | |
|----------|--|-----------|
| 2.2.2 | Các thành phần trực quan của dashboard | 17 |
| 2.2.3 | Phân tích dashboard | 20 |
| 2.2.4 | Liên hệ | 31 |
| 2.3 | Phân tích Dashboard "Điểm thi theo tổ hợp" | 33 |
| 2.3.1 | Mục tiêu dashboard | 33 |
| 2.3.2 | Các thành phần trực quan của dashboard | 36 |
| 2.3.3 | Phân tích dashboard | 38 |
| 2.4 | Phân tích Dashboard "Điểm chuẩn" | 43 |
| 2.4.1 | Mục tiêu dashboard | 43 |
| 2.4.2 | Các thành phần trực quan của dashboard | 45 |
| 2.4.3 | Phân tích dashboard | 47 |
| 2.4.3.1 | Năm 2018 | 47 |
| 2.4.3.2 | Năm 2019 | 48 |
| 2.4.3.3 | Năm 2020 | 48 |
| 2.4.3.4 | Năm 2021 | 49 |
| 2.4.3.5 | Năm 2022 | 50 |
| 2.4.3.6 | Năm 2023 | 51 |
| 2.5 | Phân tích Dashboard "Gợi ý chọn trường" | 53 |
| 2.5.1 | Mục tiêu dashboard | 53 |
| 2.5.2 | Các thành phần trực quan của dashboard | 54 |
| 3 | Đánh giá Dashboard | 54 |
| 3.1 | Nguồn dữ liệu đáng tin cậy | 54 |
| 3.2 | Phù hợp với mục đích | 55 |
| 3.3 | Rõ ràng và dễ hiểu | 55 |
| 3.3.1 | Biểu đồ dễ hiểu | 55 |
| 3.3.2 | Đối tượng trực quan rõ ràng | 56 |
| 3.4 | Tích hợp và liên kết | 56 |
| 3.4.1 | Liên kết về màu sắc | 56 |
| 3.4.2 | Liên kết về chức năng | 56 |
| 3.4.3 | Đồng bộ | 56 |

| | | |
|----------|---------------------------------------|-----------|
| 3.5 | Tương tác và điều hướng | 57 |
| 3.5.1 | Tương tác | 57 |
| 3.6 | Thiết kế hấp dẫn | 58 |
| 3.6.1 | Cân đối | 58 |
| 3.6.2 | Rõ ràng | 58 |
| 3.6.3 | Đơn giản | 58 |
| 3.6.4 | Hài hòa | 58 |
| 3.6.5 | Thân thiện với người mù màu | 58 |
| 3.7 | Phân tích dữ liệu | 59 |
| 3.8 | Tích hợp AI | 61 |
| 3.9 | Các tiêu chí đánh giá khác | 66 |
| 4 | Tổng hợp đánh giá | 68 |
| | Tài liệu tham khảo | 69 |
| | Phụ lục | 69 |

TỔNG QUAN ĐỒ ÁN

THÔNG TIN THÀNH VIÊN:

| MSSV | Họ và tên |
|----------|--------------------|
| 20120454 | Lê Công Đất |
| 21120279 | Lê Trần Minh Khuê |
| 21120290 | Hoàng Trung Nam |
| 21120296 | Lê Trần Như Ngọc |
| 21120533 | Lê Thị Minh Phương |

Bảng 1: Danh sách thành viên nhóm 10

YÊU CẦU ĐỒ ÁN:

- Trình bày một dự án ứng dụng trực quan sử dụng dashboard để phân tích dữ liệu
- Ngữ cảnh của dữ liệu là Việt Nam
- Trình bày trong buổi vấn đáp

YÊU CẦU VỀ DỮ LIỆU:

- Dữ liệu là dữ liệu thật có liên quan đất nước Việt Nam.
- Dữ liệu phải có tối thiểu 6 biến độc lập và tối thiểu 1000 dòng.
- Tổng số dữ liệu liên quan đến Việt Nam phải trên 50% dữ liệu.

CHỦ ĐỀ NHÓM CHỌN:

Phân tích điểm thi, điểm chuẩn các phương thức xét tuyển đại học tại Việt Nam qua các năm.

LỜI CẢM ƠN

Chúng em xin chân thành cảm ơn thầy - TS. Bùi Tiến Lên đã tận tình hướng dẫn và nhận xét, góp ý phù hợp để nhóm chúng em có thể hoàn thành đồ án này.

Trân trọng.

TP. Hồ Chí Minh, ngày 22 tháng 04 năm 2024

NHÓM SINH VIÊN THỰC HIỆN

Đại diện

Lê Thị Minh Phương

1 Giới thiệu về bộ dữ liệu

1.1 Dữ liệu điểm thi

1.1.1 Ngữ cảnh

Dữ liệu điểm thi THPT Quốc gia là một tập hợp thông tin quan trọng phản ánh kết quả học tập của học sinh sau 12 năm học phổ thông. Dữ liệu này thường được sử dụng cho nhiều mục đích khác nhau, bao gồm:

- Xét tuyển vào các trường đại học, cao đẳng.
- Đánh giá chất lượng giáo dục của các trường, địa phương.
- Nghiên cứu và phân tích xu hướng giáo dục.

1.1.2 Nguồn dữ liệu

Dữ liệu điểm thi THPT Quốc gia năm 2022 và 2023 được thu thập từ trang tra cứu kết quả thi THPT Quốc gia của các trang báo điện tử. Đây là các trang báo uy tín cập nhật dữ liệu tra cứu điểm thi tốt nghiệp THPT từ Bộ GD&ĐT, giúp phụ huynh, học sinh biết điểm thi tốt nghiệp THPT một cách nhanh chóng và hoàn toàn miễn phí.

- **Năm 2022:** Trang web báo điện tử Tiền Phong (<https://tienphong.vn/>).
- **Năm 2023:** Trang web VietTimes (<https://viettimes.vn/>).

1.1.3 Cách thức thu thập dữ liệu

Mã nguồn crawl dữ liệu vui lòng xem trong thư mục THPTQG/source/2022 và THPTQG/source/2023.

Được tham khảo tại https://github.com/anhdung98/diem_thi_2022 và https://github.com/anhdung98/diem_thi_2023

Công cụ Crawl dữ liệu được viết bằng Python Framework Scrapy

Mô tả các bước chính:

1. **Import các thư viện cần thiết:** Các thư viện như `scrapy`, `json`, `requests`, `BeautifulSoup` và `sqlite` được import để hỗ trợ việc crawl dữ liệu và xử lý dữ liệu.

2. **Định nghĩa lớp** `DiemThiSpider`: Đây là lớp chính kế thừa từ `scrapy.Spider`, chịu trách nhiệm thực hiện việc crawl dữ liệu.

3. **Khởi tạo các biến và kết nối CSDL SQLite**:

- Các mẫu URL, API endpoints để truy xuất dữ liệu từ API của các trang báo điện tử.
- Kết nối tới cơ sở dữ liệu SQLite và tạo con trỏ để thao tác với cơ sở dữ liệu.

4. **Hàm** `start_requests`: Hàm này bắt đầu quá trình gửi yêu cầu để lấy dữ liệu.

- Lặp qua các tỉnh thành (mã tỉnh thành từ 01 đến 64).
- Gửi yêu cầu HTTP để lấy thông tin điểm thi của mỗi tỉnh.
- Kiểm tra xem tỉnh đó có dữ liệu không bằng cách đọc `error_code` trong phản hồi JSON.
- Phân tích HTML để tìm `end_id` (mã số cuối cùng của học sinh trong tỉnh đó).
- Dựa vào `end_id`, tính toán `last_group` (nhóm cuối cùng của học sinh).
- Gửi yêu cầu lấy dữ liệu cho nhóm học sinh tương ứng nếu dữ liệu trong CSDL chưa đủ.

5. **Hàm** `parse`: Hàm này xử lý phản hồi từ các yêu cầu gửi đi trong `start_requests`.

- Đọc và giải mã JSON từ phản hồi.
- Nếu `error_code` là 0, nghĩa là có dữ liệu.
- Phân tích HTML để lấy danh sách các học sinh (`trs`).
- Nếu số lượng học sinh trong nhóm bằng với dữ liệu đã có, bỏ qua nhóm đó.
- Lặp qua từng học sinh và trích xuất dữ liệu từ các cột (`td`), sau đó yield dữ liệu.

Đoạn mã này hoạt động bằng cách lặp lại quá trình crawl thông tin điểm thi từ từng tỉnh và nhóm trong đó, và lưu trữ kết quả vào cơ sở dữ liệu SQLite để phân tích và sử dụng sau này.

1.1.4 Mô tả dữ liệu

Dữ liệu điểm thi bao gồm các thông tin sau cho mỗi thí sinh:

- **Số báo danh**: Mã định danh duy nhất của thí sinh.

- **Điểm các môn thi:** Điểm của từng môn thi (Toán, Ngữ văn, Ngoại ngữ, Vật lý, Hóa học, Sinh học, Lịch sử, Địa lý, Giáo dục công dân).

| Tên thuộc tính | Mô tả | Loại dữ liệu |
|---------------------------|----------------------|--------------|
| <code>sbd</code> | Số báo danh | Định danh |
| <code>toan</code> | Điểm môn Toán | Liên tục |
| <code>ngu_van</code> | Điểm môn Ngữ văn | Liên tục |
| <code>ngoai_ngu</code> | Điểm môn Ngoại ngữ | Liên tục |
| <code>vat_li</code> | Điểm môn Vật lý | Liên tục |
| <code>hoa_hoc</code> | Điểm môn Hóa học | Liên tục |
| <code>sinh_hoc</code> | Điểm môn Sinh học | Liên tục |
| <code>lich_su</code> | Điểm môn Lịch sử | Liên tục |
| <code>dia_li</code> | Điểm môn Địa lý | Liên tục |
| <code>gdcd</code> | Điểm môn GD&CD | Liên tục |
| <code>ma_ngoai_ngu</code> | Mã của môn Ngoại ngữ | Định danh |

Bảng 2: Mô tả bộ dữ liệu thô điểm thi THPT quốc gia

Dữ liệu được lưu trữ trong các tệp CSV riêng biệt cho từng năm (2022 và 2023).

1.1.5 Xử lý dữ liệu

Quá trình xử lý dữ liệu bao gồm các bước sau:

- **Import thư viện:** Nhóm các thư viện như `numpy`, `pandas`, `matplotlib`, `seaborn`, `plotly` được import để sử dụng cho việc xử lý và trực quan hóa dữ liệu.
- **Đường dẫn và đọc dữ liệu:** Dữ liệu từ năm 2022 và 2023 được đọc từ các file CSV và lưu vào DataFrame `df_2022` và `df_2023` tương ứng.
- **Xóa cột không cần thiết:** Cột `ma_ngoai_ngu` được loại bỏ khỏi DataFrame.
- **Tính điểm cho môn tổ hợp:** Tính toán điểm trung bình cho tổ hợp môn Khoa học tự nhiên (khtn) và Khoa học xã hội (kxhx) cho cả `df_2022` và `df_2023`.
- **Xử lý dữ liệu điểm môn Ngữ văn:** Loại bỏ các điểm không tròn 0.25 và thay thế bằng giá trị tròn gần nhất cho cả `df_2022` và `df_2023`.
- **Thêm cột Tỉnh/Thành phố vào DataFrame:** Thêm cột `Province` và `Region` dựa trên mã tỉnh/thành phố từ cột `sbd`.

- **Tính tần số xuất hiện điểm thi:** Tính toán tần số xuất hiện của từng điểm số theo môn học, tỉnh thành và năm. Kết quả được lưu vào file Excel `score_frequencies.xlsx`.

| Tên thuộc tính | Mô tả | Loại dữ liệu |
|------------------|----------------|--------------|
| Year | Năm | Định danh |
| Region | Vùng miền | Định danh |
| Province | Tỉnh/Thành phố | Định danh |
| Subject | Môn học | Định danh |
| Score | Điểm | Liên tục |
| Frequency | Tần số | Rời rạc |
| Longitude | Kinh độ | Liên tục |
| Latitude | Vĩ độ | Liên tục |

Bảng 3: Mô tả dữ liệu file `score_frequencies.xlsx`

- **Tính tổng điểm cho các tổ hợp môn:** Tính tổng điểm cho các tổ hợp môn theo định nghĩa A00, A01, A16, B00, C00, C01, D01, D07, D90, D96 và lưu vào `df_2022` và `df_2023`.
- **Tính tần số xuất hiện điểm theo từng tổ hợp thi:** Tính toán tần số xuất hiện của từng điểm số theo tổ hợp, tỉnh thành và năm. Kết quả được lưu vào file Excel `combination_score_frequencies.xlsx`.

| Tên thuộc tính | Mô tả | Loại dữ liệu |
|--------------------|----------------|--------------|
| Year | Năm | Định danh |
| Region | Vùng miền | Định danh |
| Province | Tỉnh/Thành phố | Định danh |
| Combination | Tổ hợp môn | Định danh |
| Score | Điểm | Liên tục |
| Frequency | Tần số | Rời rạc |
| Longitude | Kinh độ | Liên tục |
| Latitude | Vĩ độ | Liên tục |

Bảng 4: Mô tả dữ liệu file `combination_score_frequencies.xlsx`

- **Tính thống kê cho từng tỉnh thành:** Tính toán các số liệu thống kê cho từng môn thi và cho từng tổ hợp thi. Kết quả được lưu vào file Excel `subject_statistics.xlsx` và `combination_statistics.xlsx`.

| Tên thuộc tính | Mô tả | Loại dữ liệu |
|---|------------------------|--------------|
| Year | Năm | Định danh |
| Region | Vùng miền | Định danh |
| Province | Tỉnh/Thành phố | Định danh |
| Subject | Môn học | Định danh |
| Number of Students | Số thí sinh | Rời rạc |
| Average Score | Điểm trung bình | Liên tục |
| Number of Scores ≤ 1 | Số lượng điểm ≤ 1 | Rời rạc |
| Number of Scores < 5 | Số lượng điểm < 5 | Rời rạc |
| Number of Scores ≥ 9 | Số lượng điểm ≥ 9 | Rời rạc |
| Max Score | Điểm cao nhất | Liên tục |
| Min Score | Điểm thấp nhất | Liên tục |
| Longitude | Kinh độ | Liên tục |
| Latitude | Vĩ độ | Liên tục |

Bảng 5: Mô tả dữ liệu file `subject_statistics.xlsx`

| Tên thuộc tính | Mô tả | Loại dữ liệu |
|--|-------------------------|--------------|
| Year | Năm | Định danh |
| Region | Vùng miền | Định danh |
| Province | Tỉnh/Thành phố | Định danh |
| Combination | Tổ hợp thi | Định danh |
| Number of Students | Số thí sinh | Rời rạc |
| Average Score | Điểm trung bình | Liên tục |
| Number of Scores < 15 | Số lượng điểm < 15 | Rời rạc |
| Number of Scores < 20 | Số lượng điểm < 20 | Rời rạc |
| Number of Scores ≥ 27 | Số lượng điểm ≥ 27 | Rời rạc |
| Max Score | Điểm cao nhất | Liên tục |
| Min Score | Điểm thấp nhất | Liên tục |
| Longitude | Kinh độ | Liên tục |
| Latitude | Vĩ độ | Liên tục |

Bảng 6: Mô tả dữ liệu file `combination_statistics.xlsx`

Các bước trên giúp chuẩn bị dữ liệu để tiến hành phân tích và trực quan hóa kết quả của kỳ thi THPT Quốc gia 2022 và 2023 một cách chi tiết và có hệ thống.

1.2 Dữ liệu điểm chuẩn

1.2.1 Ngữ cảnh

Tại Việt Nam, mỗi năm, các trường đại học công bố điểm chuẩn tuyển sinh dựa trên nhiều phương thức như kết quả kỳ thi tốt nghiệp Trung học phổ thông quốc gia, kết quả các kỳ thi đánh giá năng lực riêng biệt của từng cơ sở đào tạo, hoặc kết quả học tập trung học phổ thông (điểm học bạ). Dữ liệu này phản ánh sự biến động của chất lượng đầu vào sinh viên, xu hướng chọn ngành nghề của thí sinh, cũng như sự thay đổi trong cơ cấu và chất lượng giáo dục. Việc nghiên cứu và phân tích dữ liệu điểm chuẩn không chỉ giúp học sinh và phụ huynh có cái nhìn tổng quan và chính xác hơn về xu hướng điểm chuẩn và mức độ cạnh tranh vào các trường đại học, mà còn cung cấp cho các nhà nghiên cứu khả năng theo dõi xu hướng điểm chuẩn của các nhóm ngành qua các năm.

1.2.2 Nguồn dữ liệu

Dữ liệu điểm chuẩn được thu thập từ trang thông tin hướng nghiệp của Hệ thống giáo dục HOCMAI (sau đây gọi tắt là Học mãi). Đây là trang thông tin uy tín tổng hợp và cung cấp các thông tin tuyển sinh của các trường Đại học, Cao đẳng,... trên cả nước. Giúp thí sinh và phụ huynh có thể dễ dàng truy cập và tra cứu một cách miễn phí.

Tất cả dữ liệu điểm chuẩn của các trường đại học (từ năm 2018 - 2023) được Học mãi tổng hợp và cho phép tra cứu tại: huongnghiep.hocmai.vn/diem-chuan

Ngoài ra, trong quá trình xử lý dữ liệu, nhóm có tiếp cận và sử dụng thêm các dữ liệu về: Trường và Mã trường đại học theo khu vực của Thư viện pháp luật ([xem tại đây](#))[1], và dữ liệu về mã ngành, phân ngành theo Thông tư số: 09/2022/TT-BGDĐT ([xem tại đây](#))[2]..

1.2.3 Cách thức thu thập dữ liệu

Nhóm thiết kế bộ thu thập dữ liệu để lấy thông tin điểm chuẩn của các trường đại học tại Việt Nam qua nhiều năm.

1.2.3.1 Cách thức truy cập và tải dữ liệu

- *Truy cập trang web*

Sử dụng `requests` để gửi yêu cầu GET và lấy nội dung HTML của trang web.

- *Điều khiển trình duyệt web*

- + Sử dụng Selenium để mở trang web và điều khiển các thao tác như chọn năm học cụ thể từ dropdown menu.
- + Chờ trang tải hoàn chỉnh rồi trích xuất dữ liệu từ bảng HTML.
- ***Đa luồng để tăng tốc quá trình***
Sử dụng `concurrent.futures.ThreadPoolExecutor` để thực hiện các tác vụ đa luồng, giúp tăng tốc quá trình thu thập dữ liệu từ nhiều trang web cùng một lúc.
- ***Lưu trữ dữ liệu***
Dữ liệu được lưu dưới dạng bảng trong các file CSV, sử dụng thư viện Pandas để chuyển đổi từ JSON sang DataFrame và lưu vào file.

1.2.3.2 Các bước thực hiện cụ thể

Bước 1: Lấy danh sách các trường đại học

- Gửi yêu cầu HTTP GET đến URL <https://huongnghiep.hocmai.vn/diem-chuan/> để lấy nội dung HTML của trang web.
- Sử dụng BeautifulSoup để phân tích cú pháp HTML và trích xuất dữ liệu các trường đại học.
- Tạo danh sách chứa các thông tin về URL, mã trường và tên trường đại học.

Bước 2: Lấy dữ liệu điểm chuẩn

- Khởi tạo trình duyệt web bằng Selenium để truy cập vào từng trang web cụ thể của các trường.
- Chọn năm cụ thể từ danh sách các năm (2018-2023) để lấy dữ liệu điểm chuẩn.
- Trích xuất dữ liệu điểm chuẩn từ bảng HTML trên trang web.
- Lưu dữ liệu điểm chuẩn vào một danh sách dưới dạng JSON.

Bước 3: Xử lý và lưu trữ dữ liệu

- Chuyển đổi dữ liệu thô (JSON) sang dạng bảng (CSV, Pandas DataFrame).
- Lưu dữ liệu vào các file CSV để dễ dàng xử lý và phân tích sau này.

Quy trình này đảm bảo rằng dữ liệu được thu thập một cách tự động, hiệu quả và tiết kiệm rất nhiều thời gian so với thực hiện thủ công.

1.2.4 Mô tả dữ liệu

Mỗi dòng dữ liệu thô về điểm chuẩn bao gồm:

- **Thông tin về cơ sở đào tạo:** bao gồm mã trường, tên trường.
- **Thông tin về ngành tuyển sinh:** gồm mã ngành xét tuyển và tên ngành.
- **Tổ hợp:** Các tổ hợp được dùng để tuyển sinh cho ngành này.
- **Điểm chuẩn và ghi chú:** thông tin về điểm chuẩn và ghi chú về phương thức, chương trình đào tạo, hoặc các điều kiện xét tuyển khác.

| Tên thuộc tính | Mô tả | Loại dữ liệu |
|----------------|--------------------------------|--------------|
| Mã trường | Mã viết tắt của trường Đại học | Định danh |
| Tên trường | Tên trường Đại học | Định danh |
| Mã ngành | Mã tuyển sinh của ngành | Định danh |
| Tên ngành | Tên ngành | Định danh |
| Tổ hợp | Các tổ hợp xét tuyển | Chuỗi ký tự |
| Điểm chuẩn | Điểm trúng tuyển vào ngành | Liên tục |
| Ghi chú | Thông tin thêm | Chuỗi ký tự |
| Năm | Năm xét tuyển | Thứ tự |

Bảng 7: Mô tả bộ dữ liệu thô điểm chuẩn các trường Đại học

Các file dữ liệu thô từ năm 2018 đến năm 2023 được lưu trữ trong thư mục `data/full` với tên file theo cấu trúc `diemchuan<nam>_full.csv`.

1.2.5 Xử lý dữ liệu

- **Import các thư viện và chỉ định đường dẫn thư mục:** các thư viện như `Pandas`, `NumPy`, `os`, ... được import để xử lý dữ liệu. Đường dẫn thư mục chứa các file CSV được chỉ định là `./data/full`.
- **Đọc và lưu trữ dữ liệu từ các file CSV:** Dữ liệu từ mỗi file CSV là các thông tin điểm chuẩn của các trường đại học cho từng năm, được đọc vào `DataFrame` và lưu trữ trong một dictionary. Sau đó gộp lại và lưu vào một dataframe duy nhất. (tại bước này, nhóm cũng mở rộng thêm 4 trường dữ liệu là "Tỉnh/ Thành phố", "Khu vực", "Mã trường xét tuyển", và "Loại đơn vị")

- **Kiểm tra tính đầy đủ của dữ liệu và xử lý dữ liệu thiếu:** Dữ liệu được kiểm tra tính đầy đủ, xử lý các dòng có dữ liệu thiếu, và loại bỏ các dòng trùng lặp.
- **Khám phá dữ liệu:** Dữ liệu được khám phá qua việc xem mẫu dữ liệu và kiểm tra kiểu dữ liệu của từng cột. Ngoài ra, nhóm tiến hành xem phân phối/ phân bố của dữ liệu. Kiểm tra những điểm bất thường.
- **Tiền xử lý dữ liệu:**
 - + Xử lý các điểm bất thường phát hiện được.
 - + Thêm thông tin về ngành học (bao gồm "Mã nhóm ngành", "Nhóm ngành", "Mã phân ngành", "Phân ngành" từ file `major_info.csv`,
 - + Tách cột "Tổ hợp" để mỗi dòng chỉ chứa một tổ hợp môn.
 - + Phân lớp loại điểm theo phương thức xét tuyển.
- **Sắp xếp lại cấu trúc dữ liệu:** cấu trúc dữ liệu được sắp xếp lại để chuẩn hóa các thông tin, xuất ra file Excel `diemchuan_full.xlsx`, chuẩn bị cho các phân tích và báo cáo sau này.

| Tên thuộc tính | Mô tả | Loại dữ liệu |
|---------------------|---|--------------|
| Khu vực | Vùng miền của trường Đại học | Định danh |
| Tỉnh/ Thành phố | Vị trí của trường Đại học | Định danh |
| Mã trường xét tuyển | Mã của trường trên hệ thống của Bộ Giáo dục | Định danh |
| Mã trường | Mã viết tắt của trường Đại học | Định danh |
| Tên trường | Tên trường Đại học | Định danh |
| Loại đơn vị | Cơ sở đào tạo công lập/ dân lập | Định danh |
| Mã nhóm ngành | Mã nhóm ngành đào tạo | Định danh |
| Nhóm ngành | Nhóm ngành đào tạo | Định danh |
| Mã phân ngành | Mã phân ngành đào tạo | Định danh |
| Phân ngành | Phân ngành đào tạo | Định danh |
| Mã xét tuyển | Mã tuyển sinh của ngành | Định danh |
| Tên ngành | Tên ngành | Định danh |
| Loại điểm | Loại điểm/ Phương thức xét tuyển | Định danh |
| Tổ hợp | Tổ hợp xét tuyển | Định danh |
| Điểm chuẩn | Điểm trúng tuyển vào ngành | Liên tục |
| Ghi chú | Thông tin thêm | Chuỗi ký tự |
| Năm | Năm xét tuyển | Thứ tự |

Bảng 8: Mô tả bộ dữ liệu điểm chuẩn các trường Đại học sau khi xử lý

2 Phân tích Dashboard

2.1 Tổng quan

2.1.1 Đối tượng

Đối tượng: Học sinh, phụ huynh, giáo viên và những người quan tâm đến kỳ xét tuyển Đại học ở Việt Nam.

Trong đó, đối tượng chính và quan trọng của dashboard là các học sinh hoặc phụ huynh của các học sinh sắp hoặc đang trong thời gian xét tuyển đại học có nhu cầu theo dõi tình hình điểm thi THPTQG và điểm chuẩn để đưa ra các lựa chọn ngành học, trường học hoặc các kế hoạch phù hợp.

2.1.2 Mục tiêu

Mục tiêu chung của dashboard:

- Cung cấp các thông tin tổng quan về điểm thi Trung học Phổ thông Quốc gia (kỳ thi chính để xét tuyển đại học ở Việt Nam) theo môn, theo tổ hợp thi từ năm 2022-2023.
- Cung cấp các thông tin tổng quan về điểm chuẩn của các trường đại học tại Việt Nam theo các phương thức xét tuyển từ năm 2018-2023.

Mục tiêu cụ thể của dashboard:

Về điểm thi Trung học Phổ thông Quốc gia, dashboard cung cấp cái nhìn tổng quan về:

- Số lượng thí sinh dự thi trên cả nước và tương quan về số lượng thí sinh dự thi giữa các tỉnh thành.
- Điểm thi từng môn: Điểm trung bình của môn, điểm cao nhất và thấp nhất, phổ điểm chung.
- Điểm thi đối với các tổ hợp tiêu biểu: Điểm trung bình của của tổ hợp, điểm cao nhất và thấp nhất, phổ điểm chung.
- Điểm thi của các tỉnh thành trên cả nước: Thứ hạng tương đối về điểm trung bình ở các môn thi hoặc ở các tổ hợp.
- Số thí sinh dưới điểm liệt, dưới điểm trung bình, số thí sinh đạt điểm cao ở các môn thi (tương tự với tổ hợp).

Về điểm chuẩn các trường đại học tại Việt Nam, dashboard cung cấp:

- Cái nhìn tổng quan về điểm chuẩn xét tuyển trung bình, cao nhất và thấp nhất theo từng nhóm ngành/phân ngành/ngành học; Các trường đại học có điểm chuẩn xét tuyển cao nhất và thấp nhất theo từng ngành.
- Thông tin cơ bản về trường đại học: Vị trí địa lý của trường, loại cơ sở đào tạo (công lập/dân lập).
- Gợi ý chi tiết về các ngành học và trường đại học phù hợp với điểm xét tuyển của người dùng.

2.1.3 Các dashboard thành phần

Dashboard bao gồm 4 dashboard thành phần:

- Dashboard "Điểm thi theo môn": Cung cấp thông tin tổng quan về số lượng thí sinh và tình hình điểm thi Trung học Phổ thông Quốc gia (2022-2023) theo từng môn.
- Dashboard "Điểm thi theo tổ hợp": Cung cấp thông tin tổng quan về số lượng thí sinh và tình hình điểm thi Trung học Phổ thông Quốc gia (2022-2023) theo một số tổ hợp tiêu biểu.
- Dashboard "Điểm chuẩn": Cung cấp thông tin tổng quan về điểm chuẩn xét tuyển của các trường đại học theo nhóm ngành/phân ngành/ngành học.
- Dashboard "Gợi ý chọn trường phù hợp": Đưa ra gợi ý các ngành học và trường đại học phù hợp với điểm xét tuyển do người dùng cung cấp.

2.2 Phân tích Dashboard "Điểm thi theo môn"

2.2.1 Mục tiêu dashboard

Mục tiêu chung của dashboard:

- Cung cấp các thông tin tổng quan về điểm thi theo Trung học Phổ thông Quốc gia tại Việt Nam theo **môn thi**
- Thời gian: năm 2022 và 2023
- Phạm vi: 63 tỉnh thành trên cả nước

- Đối tượng: Tất cả các môn thi của kỳ thi

Mục tiêu cụ thể của dashboard:

- Cung cấp thông tin tổng quan về số lượng thí sinh dự thi trên cả nước và tại mỗi tỉnh thành theo từng môn thi.
- Cung cấp thông tin tổng quan về tình hình điểm thi: Điểm trung bình, cao nhất và thấp nhất; phổ điểm.
- Cung cấp thông tin tổng quan về các mức điểm đặc biệt: số lượng bài thi dưới điểm liệt, dưới điểm trung bình, trên điểm 9 trên cả nước và của mỗi tỉnh thành.
- So sánh tình hình điểm của các tỉnh thành: tỉnh nào có nhiều/ít thí sinh, tỉnh nào có điểm trung bình cao nhất/thấp nhất, tỉnh nào có nhiều bài điểm liệt, điểm cao (trên 9),...

2.2.2 Các thành phần trực quan của dashboard

Dashboard "Điểm thi theo môn" (hình 1) bao gồm các thành phần chính sau đây:

Tiêu đề

Tiêu đề của dashboard cung cấp thông tin về nội dung được thể hiện trong dashboard (Phân tích điểm thi Trung học Phổ thông Quốc gia) và thời gian trực quan (năm 2022 - 2023).

Tiêu đề được đặt đơn giản, ngắn gọn, bao hàm vừa đủ lượng thông tin. Kích cỡ chữ của thông tin và thời gian khác nhau giúp người dùng nhanh chóng phân biệt được thông tin chính và phụ.

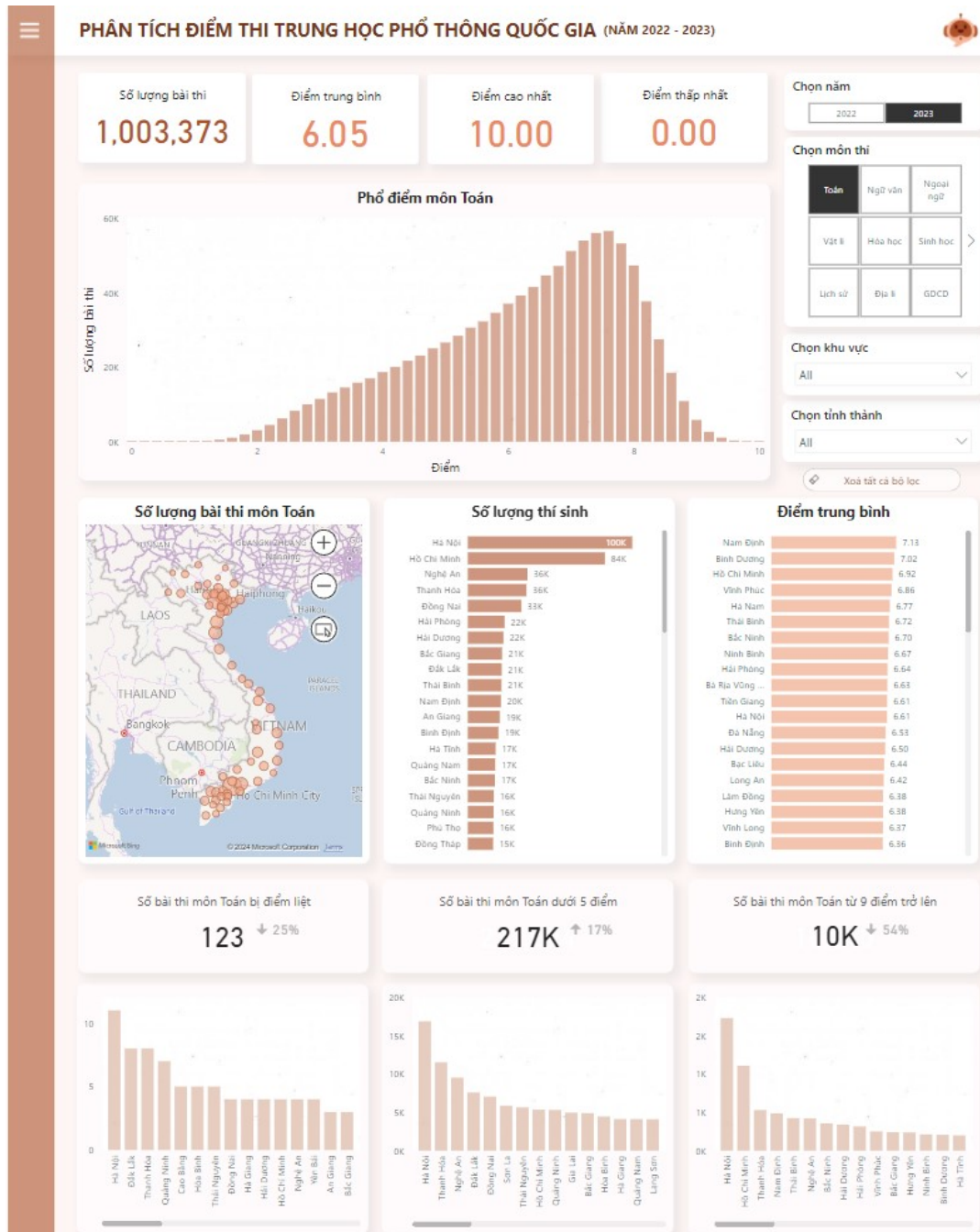
Bộ lọc

Khu vực bộ lọc cho phép người dùng chọn các đối tượng muốn theo dõi trong dashboard, bao gồm: Năm, Môn thi, Khu vực, Tỉnh thành.

Các bộ lọc Năm, Môn thi là bộ lọc chỉ chọn một giá trị (việc theo dõi số lượng thí sinh và điểm thi tổng của 2 năm hoặc 2 môn học không mang nhiều ý nghĩa).

Các bộ lọc Khu vực, Tỉnh thành có thể chọn một hoặc nhiều giá trị, cho phép người dùng theo dõi tình hình điểm thi theo các phạm vi khác nhau: trên cả nước, theo từng khu vực, theo từng tỉnh thành.

Các bộ lọc được bố trí gần nhau tạo thành một nhóm, giúp người dùng dễ nhận diện và thao tác.



Hình 1: Tổng quan dashboard "Điểm thi theo môn"

Số liệu thống kê tổng quan

Loại trực quan: Số liệu

Số liệu thống kê tổng quan cung cấp số liệu chính xác và cụ thể về:

- Số lượng thí sinh dự thi môn được chọn trong phạm vi được chọn
- Điểm trung bình, điểm cao nhất và thấp nhất của môn được chọn trong phạm vi được chọn

Các số liệu được bố trí trong các ô lớn, kích cỡ chữ lớn, màu sắc nổi bật và nằm ở một trong những vị trí trọng tâm nhất của biểu đồ giúp thu hút sự chú ý của người dùng. Màu sắc của số liệu cũng đồng bộ với cách encode màu sắc trong toàn dashboard: màu xanh dùng cho biến số lượng thí sinh, màu cam dùng cho các biến liên quan đến điểm.

Biểu đồ phổ điểm

Loại trực quan: Biểu đồ cột đứng

Phổ điểm là thông tin về điểm thi được quan tâm nhiều nhất. Chính vì vậy, biểu đồ này được đặt ở vị trí trung tâm của dashboard, là vị trí dễ thu hút sự chú ý của người dùng nhất và dễ quan sát nhất.

Các biểu đồ về tình hình dự thi chung trên cả nước

Loại trực quan: Bản đồ bong bóng, biểu đồ cột ngang

Các biểu đồ này cung cấp cái nhìn tổng quan về tình hình dự thi chung trên phạm vi cả nước, bao gồm: số lượng thí sinh và điểm trung bình theo từng tỉnh thành, bao gồm số liệu và mối tương quan giữa những số liệu đó.

- Bản đồ bong bóng: Thể hiện trên bản đồ Địa lý sự tương quan về số thí sinh dự thi giữa các tỉnh thành; sự tập trung thí sinh dự thi ở một số khu vực nhất định trên bản đồ.
- Biểu đồ cột ngang 1: Hỗ trợ cho bản đồ bong bóng trong việc thể hiện riêng biệt và có thứ tự tương quan về số lượng thí sinh (vì trên bản đồ, các đối tượng có thể bị chồng lấp lên nhau, và kích thước của bong bóng khó so sánh hơn).
- Biểu đồ cột ngang 2: Thể hiện thứ tự xếp hạng cùng số liệu cụ thể về điểm trung bình giữa

các tỉnh thành, phần nào đó có thể cung cấp đánh giá nhanh về tình hình điểm thi của các tỉnh thành.

Các số liệu và biểu đồ về các mức điểm đặc biệt

Loại trực quan: Số liệu, Biểu đồ cột đứng

Khu vực cuối cùng của dashboard cung cấp thông tin về tình hình điểm thi tại các tỉnh thành khi xem xét các mức điểm đặc biệt và thường được quan tâm:

- Số bài thi dưới điểm liệt trên cả nước và thứ tự giữa các tỉnh thành (tỉnh nào có ít, nhiều thí sinh dưới điểm liệt)
- Số bài thi dưới điểm trung bình trên cả nước và thứ tự giữa các tỉnh thành (tỉnh nào có ít, nhiều thí sinh dưới điểm trung bình)
- Số bài thi điểm cao (từ 9 điểm) trên cả nước và thứ tự giữa các tỉnh thành (tỉnh nào có ít, nhiều thí sinh đạt điểm cao)

Trong đó, tổng số bài thi trên cả nước được biểu diễn bằng số liệu, còn thứ tự giữa các tỉnh thành được biểu diễn bằng biểu đồ cột. Số liệu và biểu đồ được nhóm lại với nhau từng cặp theo chiều dọc giúp người xem dễ quan sát hơn, đồng thời giúp hai thành phần này có thể dùng chung tiêu đề, làm cho dashboard đơn giản, tinh gọn.

Các thành phần khác

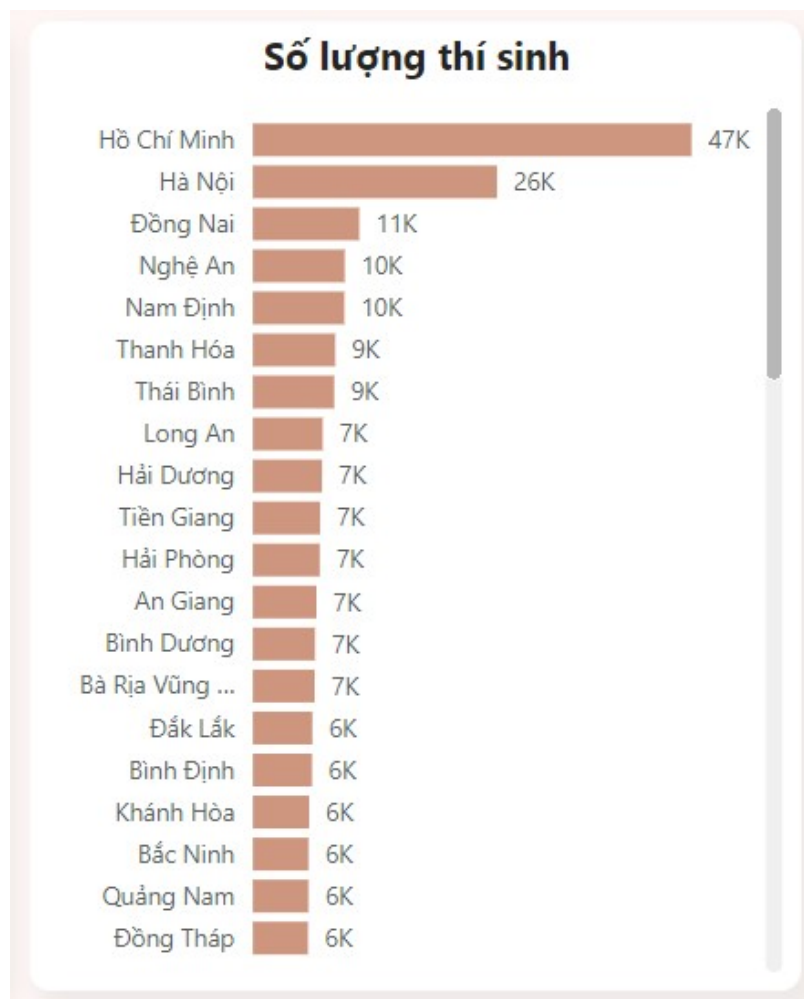
- Nút xóa bộ lọc: được đặt ở vị trí góc phải trên của dashboard, gần khu vực bộ lọc. Nút này cho phép người dùng xóa nhanh tất cả các bộ lọc đang áp dụng trên dashboard và đưa về lựa chọn mặc định ban đầu.
- Nút thông tin: được đặt ở vị trí góc phải trên của dashboard. Nút này cung cấp thông tin về tác giả và quyền sở hữu đối với dashboard.

2.2.3 Phân tích dashboard

Số lượng thí sinh

Hà Nội và TP. Hồ Chí Minh là hai tỉnh thành có số lượng thí sinh dự thi đông nhất.

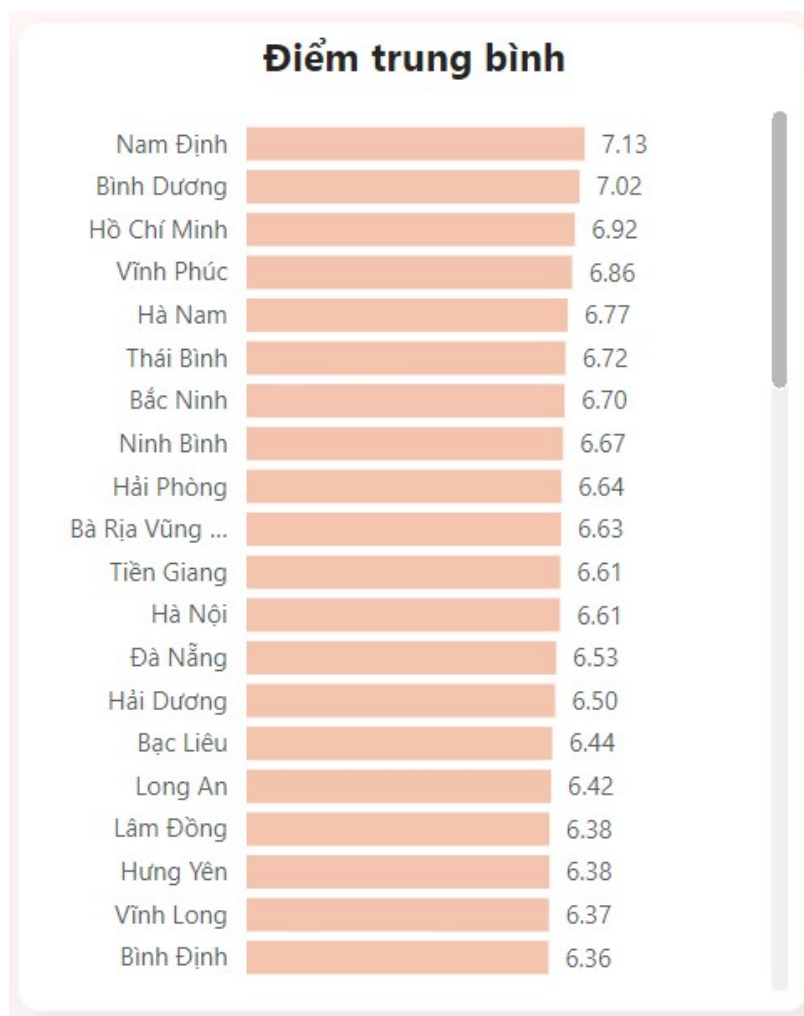
- Đối với các môn thi tốt nghiệp chung (Ngữ văn, Toán, Ngoại ngữ): Hà Nội là tỉnh thành có nhiều thí sinh dự thi nhất trên cả nước.
- Đối với các môn thi khối Khoa học Xã hội (Lịch sử, Địa lí, Giáo dục công dân): Hà Nội là tỉnh thành có nhiều thí sinh dự thi nhất trên cả nước.
- Đối với các môn thi khối Khoa học Tự nhiên (Vật lí, Hóa học, Sinh học): TP. Hồ Chí Minh là tỉnh thành có nhiều thí sinh dự thi nhất trên cả nước.



Hình 2: Các tỉnh có số lượng thí sinh thi môn Vật lí cao nhất năm 2023

Điểm trung bình cao nhất các môn thi

- Môn Toán: Nam Định, Bình Dương (2022, 2023)
- Môn Ngữ văn: Hải Phòng, Vĩnh Phúc (2022), Hà Nam, Vĩnh Phúc (2023)
- Môn Ngoại ngữ: TP. Hồ Chí Minh, Bình Dương (2022, 2023)
- Môn Vật lí: Bắc Ninh, Ninh Bình (2022), Vĩnh Phúc, Bắc ninh (2023)
- Môn Hóa học: Tuyên Quang, Ninh Bình (2022), Nam Định, Tuyên Quang (2023)
- Môn Sinh học: Tuyên Quang, Vĩnh Long (2022), Vĩnh Long, Bình Dương (2023)
- Môn Lịch sử: Vĩnh Phúc, Nam Định (2022), Vĩnh Phúc, Ninh Bình (2023)
- Môn Địa lí: Nam Định, Ninh Bình (2022), Vĩnh Phúc, Nam Định (2023)
- Môn Giáo dục công dân: Vĩnh Phúc, Bình Dương (2022, 2023)



Hình 3: Các tỉnh có điểm trung bình cao nhất môn Toán năm 2023

Nhìn chung, các tỉnh đứng đầu trong các môn thường tập trung ở vùng đồng bằng Sông Hồng. Tỉnh Tuyên Quang là tỉnh trung du và miền núi phía Bắc duy nhất nằm trong top 2. TP. Hồ Chí Minh chỉ đứng đầu duy nhất cả nước ở môn Ngoại ngữ (điểm khá gần với Bình Dương, và điểm của 2 tỉnh này cách biệt khá nhiều so với các tỉnh còn lại). Vĩnh Long là tỉnh duy nhất ở khu vực đồng bằng Sông Cửu Long nằm trong top 2.

Điểm trung bình thấp nhất các môn thi

- Môn Toán: Cao Bằng, Hà Giang
- Môn Ngữ văn: Cao Bằng, Hà Giang (2022), Hà Giang, Lai Châu (2023).
- Môn Ngoại ngữ: Sơn La, Hà Giang (2022), Điện Biên, Hà Giang (2023)

- Môn Vật lí: Cà Mau, Trà Vinh
- Môn Hóa học: Cà Mau, Bình Phước (2023), Cà Mau, Hà Nội (2023)
- Môn Sinh học: Quảng Ngãi, Hưng Yên
- Môn Lịch sử: Đắk Lắk, Hà Giang
- Môn Địa lí: Đắk Lắk, Hà Giang (2022), Quảng Ngãi, Hà Giang (2023)
- Môn Giáo dục công dân: Quảng Ngãi, Hà Giang

Hầu hết các tỉnh nằm trong danh sách điểm trung bình thấp nhất thuộc về khu vực Trung du và miền núi phía Bắc, đặc biệt là tỉnh Hà Giang (thấp nhất ở nhiều môn).



Hình 4: Các tỉnh có điểm trung bình thấp nhất môn Toán năm 2023

Hà Nội là thành phố trực thuộc Trung ương duy nhất nằm trong danh sách, có điểm trung bình thấp nhất ở môn Hóa học. Ở miền Trung ta có các tỉnh Đắk Lắk, Quảng Ngãi. Ở miền Nam có tỉnh Cà Mau, Trà Vinh.

Khu vực miền Bắc

- Hà Nội có số lượng thí sinh rất lớn, gấp 2.5 lần so với tỉnh đứng thứ 2. Các tỉnh cũng có số lượng thí sinh dự thi nhiều là: Nghệ An, Nam Định, Thanh Hóa.
- Các tỉnh xếp hạng cao nhất về điểm trung bình: Bắc Ninh (Vật lí), Ninh Bình, Vĩnh Phúc (GD&ĐT, Lịch sử, Địa lí, Vật lí), Tuyên Quang (Hóa học, Sinh học), Hà Nội (Ngoại ngữ), Hải Phòng (Ngữ văn), Nam Định (Địa lí, Toán, Hóa học), Hà Nam (Ngữ văn).
- Các tỉnh xếp hạng thấp nhất về điểm trung bình: Điện Biên (Vật lí), Hà Giang (Toán, Ngoại ngữ, Lịch sử, GD&ĐT, Địa lí), Hưng Yên (Sinh học), Lai Châu (Ngữ văn), Hà Nội (Hóa học).

Khu vực miền Trung

- Các tỉnh có nhiều thí sinh dự thi nhất: Đắk Lắk, Bình Định.
- Các tỉnh xếp hạng cao nhất về điểm trung bình: Lâm Đồng (Địa lí, GD&ĐT, Ngữ văn, Sinh học), Thừa Thiên - Huế (Hóa học, Vật lí), Bình Định (Vật lí, Lịch sử), Đà Nẵng (Toán, Ngoại ngữ), Quảng Bình (Lịch sử)
- Các tỉnh xếp hạng thấp nhất về điểm trung bình: Đắk Lắk,

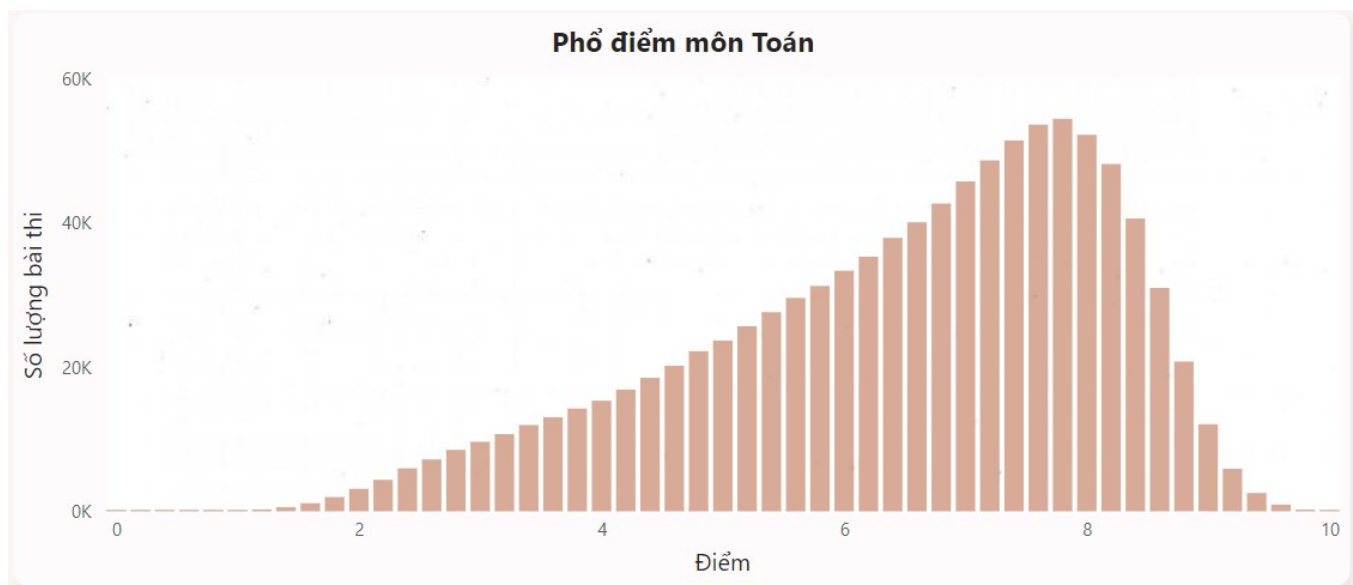
Khu vực miền Nam

- TP. Hồ Chí Minh có số lượng thí sinh rất lớn, gấp 2.5 lần so với tỉnh đứng thứ 2. Tỉnh cũng có số lượng thí sinh dự thi nhiều là: Đồng Nai, An Giang.
- Các tỉnh xếp hạng cao nhất về điểm trung bình: Bình Dương (Địa lí, GD&ĐT, Lịch sử, Hóa học, Vật lí, Toán, Ngữ văn), TP. Hồ Chí Minh (Ngoại ngữ), Tiền Giang (Ngữ văn), Vĩnh Long (Sinh học). Ta thấy tỉnh Bình Dương xếp hạng nhất ở hầu hết tất cả các môn.

- Các tỉnh xếp hạng thấp nhất về điểm trung bình: Cà Mau (Vật lí, Hóa học), Trà Vinh (Toán, Ngữ văn, GDCD, Địa lí, Lịch sử, Ngoại ngữ, Vật lí), Bà Rịa - Vũng Tàu (Sinh học), Sóc Trăng (Ngoại ngữ), Ninh Thuận (Lịch sử, Địa lí, GDCD), Bình Phước (Hóa học), Long An (Ngữ văn). Ta thấy, tỉnh Trà Vinh xếp hạng cuối ở hầu hết tất cả các môn.

Phổ điểm

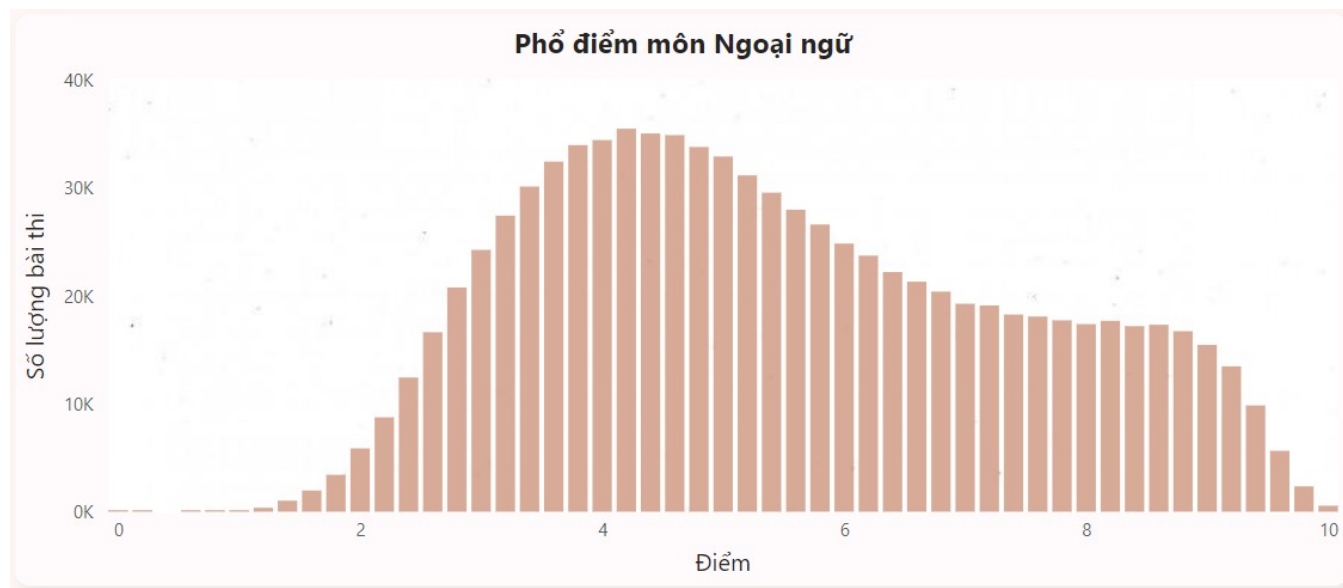
Môn Toán: Phổ điểm tương đối giống phân phối tam giác và lệch về phía bên phải ở cả 2 năm 2022 và 2023. Mode năm 2023 rơi vào khoảng 7.6 điểm.



Hình 5: Phổ điểm thi THPTQG môn Toán năm 2023

Môn Ngữ văn: Phổ điểm hình chuông khá cân đối ở cả 2 năm 2022 và 2023. Phổ điểm tương đối chuẩn. Mode năm 2023 rơi vào khoảng 7 điểm.

Môn Ngoại ngữ: Phổ điểm khác với các môn còn lại, lệch về bên trái và có hình yên ngựa ở cả năm 2022 và 2023. Mode năm 2023 là khoảng 4.2 điểm, rất thấp so với các môn khác. Năm 2023 thì phân phối chuyển dịch về phía bên phải nhiều hơn, cho thấy có sự cải thiện về tình hình điểm thi của môn này.

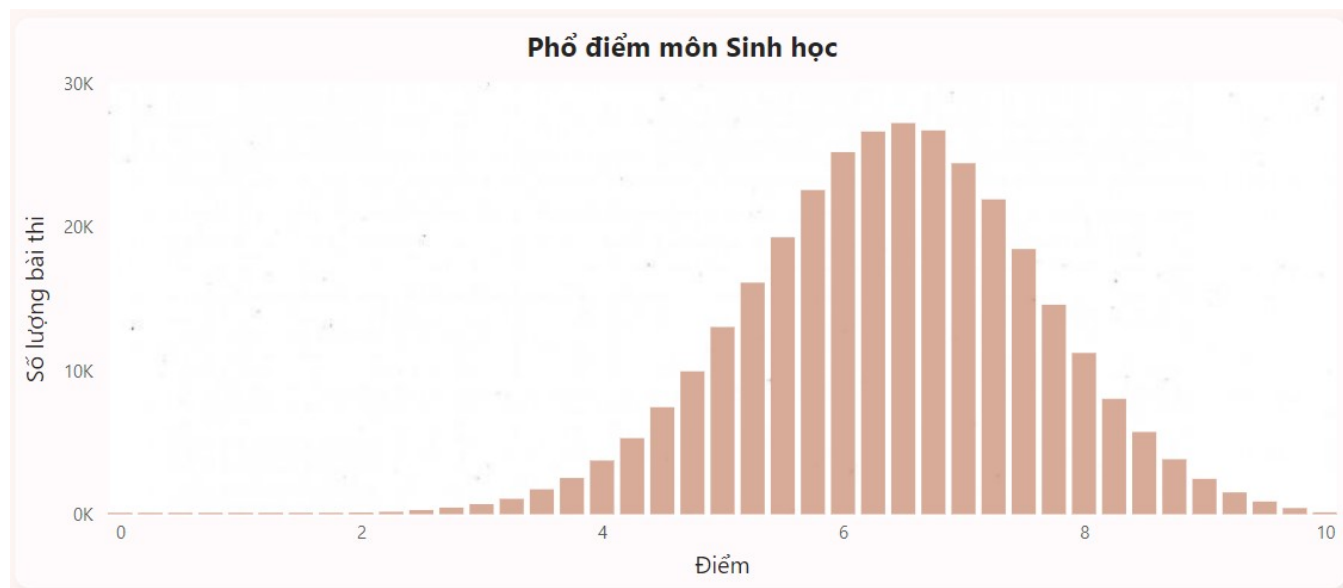


Hình 6: Phổ điểm thi THPTQG môn Ngoại ngữ năm 2023

Môn Vật lí: Phổ điểm khá giống môn Toán, lệch về phía bên phải với giá trị mode cao 7.5 điểm (2023).

Môn Hóa học: Phổ điểm khá giống môn Toán, lệch về phía bên phải với giá trị mode cao 7.5 điểm (2023). So với năm 2022, phổ điểm của năm 2023 ít lệch phải hơn.

Môn Sinh học: Phổ điểm hình chuông cân đối ở cả 2 năm 2022 và 2023. Giá trị mode có sự tăng rõ rệt từ 4.5 vào năm 2022 lên 6.5 vào năm 2023.

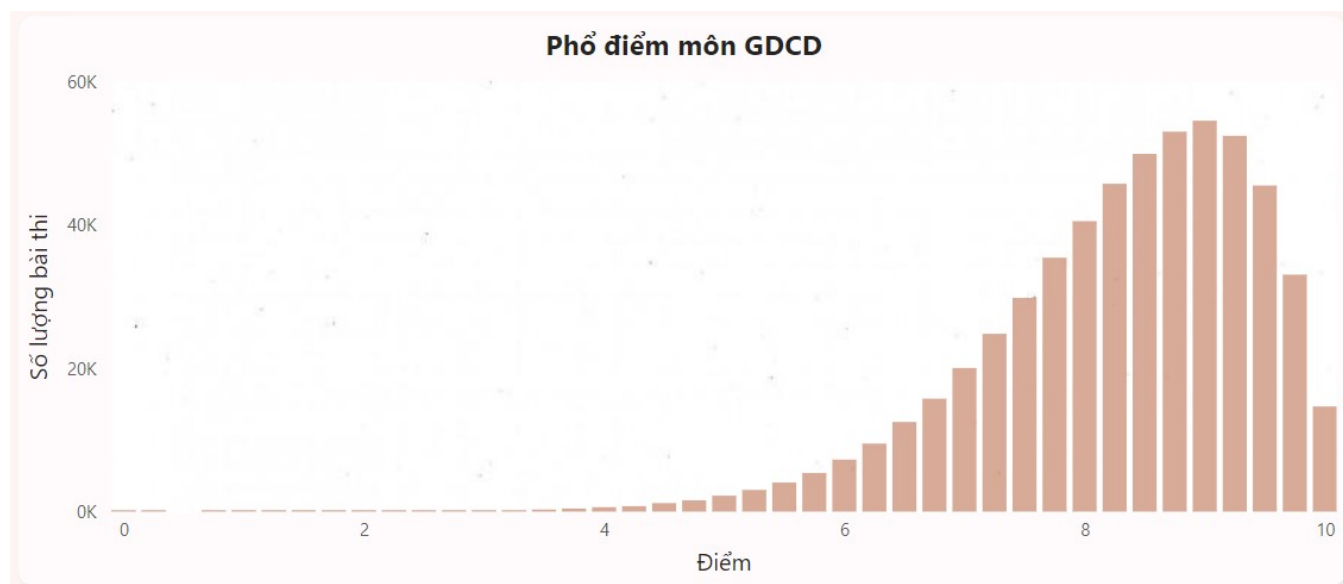


Hình 7: Phổ điểm thi THPTQG môn Sinh học năm 2023

Môn Lịch sử: Phổ điểm năm 2022 lệch nhẹ về bên phải và có giá trị mode khoảng 7 điểm. Tuy nhiên, đến năm 2023 thì phổ điểm lại lệch nhẹ về bên trái và có giá trị trung vị giảm đáng kể chỉ còn 5.75.

Môn Địa lí: Phổ điểm hình chuông cân đối ở cả 2 năm 2022 và 2023. Giá trị mode năm 2023 tăng so với năm 2022.

Môn GDCD: Phổ điểm lệch mạnh về phía bên phải, đặc biệt là ở năm 2023 có nhiều thí sinh đạt điểm tuyệt đối.



Hình 8: Phổ điểm thi THPTQG môn Giáo dục công dân năm 2023

Môn Khoa học Tự nhiên: Phổ điểm lệch tương đối nhẹ về phía bên phải, giá trị mode khoảng 7 điểm (2023).

Môn Khoa học Xã hội: Phổ điểm cân đối, giá trị mode khoảng 5.25 điểm (2023).

Số bài thi dưới điểm liệt

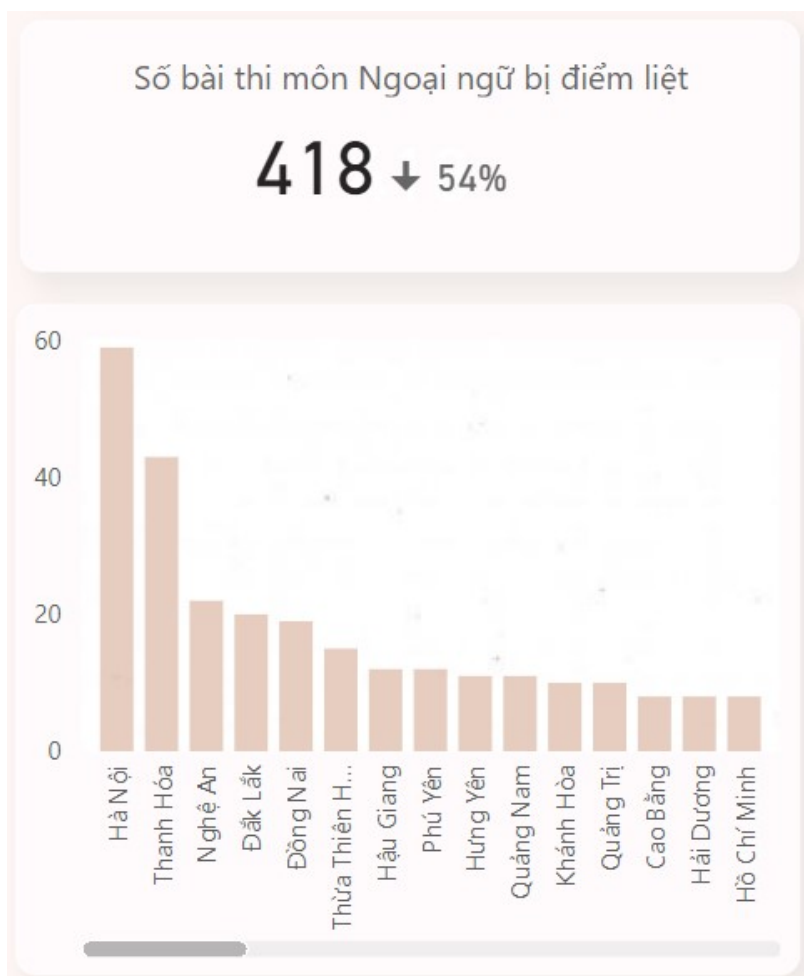
Điểm liệt là 1 điểm.

Các tỉnh có nhiều bài thi dưới điểm liệt:

- Môn Toán: Hà Nội (số lượng thí sinh nhiều nhất cả nước), Đắk Lắk (số lượng thí sinh nhiều nhất miền Trung - Tây Nguyên)
- Môn Ngữ văn: Hồ Chí Minh, Long An

- Môn Ngoại ngữ: Hà Nội

Mặc dù số lượng thí sinh năm 2023 tăng so với năm 2022, tình hình điểm liệt trong năm 2023 lại giảm mạnh và giảm ở hầu hết môn thi (trừ Vật lí và Địa lí).



Hình 9: Số lượng bài thi bị điểm liệt môn Ngoại ngữ năm 2023

Số bài thi dưới điểm trung bình

Điểm trung bình là 5 điểm.

Sự thay đổi về số bài thi dưới điểm trung bình của năm 2023 so với năm 2022:

- Tăng ở các môn: Toán, Vật lí, Lịch sử, Địa lí
- Giảm ở các môn: Ngữ văn, Ngoại ngữ, Hóa học, Sinh học, GDCD

Số bài thi điểm cao

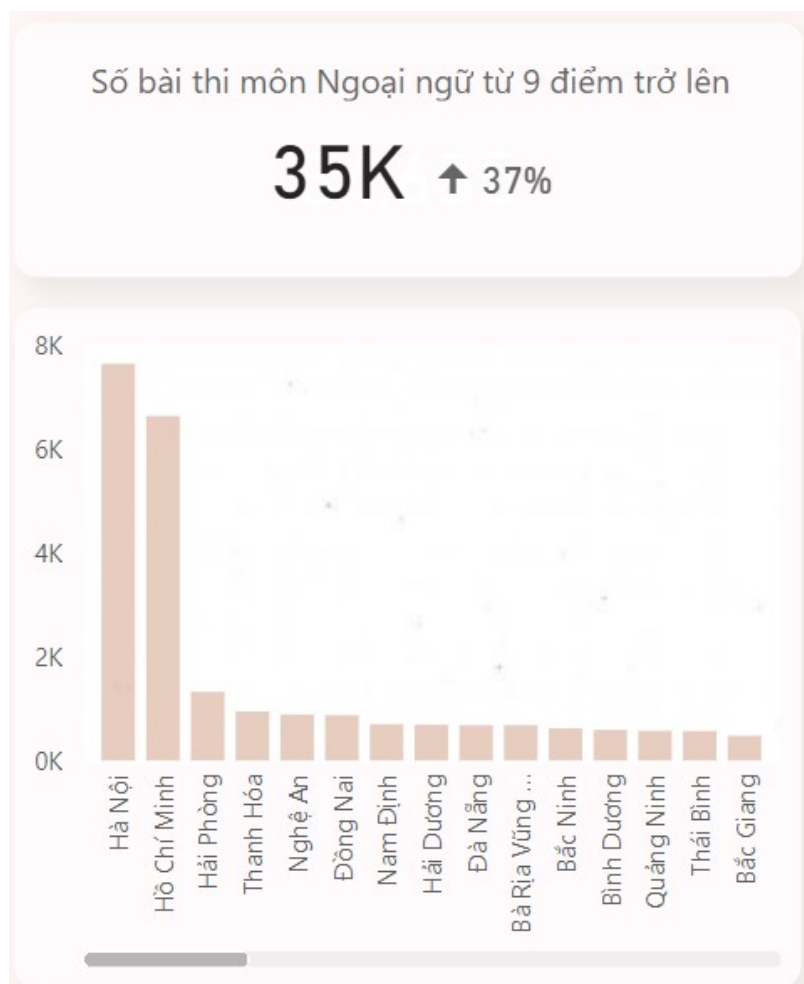
Điểm cao là điểm từ 9 trở lên.

Các tỉnh có nhiều bài thi điểm cao:

- Môn Toán: Hà Nội, TP. Hồ Chí Minh
- Môn Ngữ văn: Hải Phòng, Hà Nội
- Môn Ngoại ngữ: Hà Nội, TP. Hồ Chí Minh

Sự thay đổi về số bài thi điểm cao của năm 2023 so với năm 2022:

- Tăng ở các môn: Ngữ văn, Ngoại ngữ, GDCD
- Giảm ở các môn: Toán, Vật lí, Hóa học, Lịch sử, Địa lí



Hình 10: Số lượng bài thi đạt từ 9 điểm trở lên môn Ngoại ngữ năm 2023

2.2.4 Liên hệ

Thực trạng giáo dục ở các tỉnh thuộc khu vực Trung du và miền núi phía Bắc

Các phân tích phía trên cho thấy thực trạng giáo dục đáng báo động ở các tỉnh thành thuộc khu vực Trung du và miền núi phía Bắc, đặc biệt trong đó là tỉnh Hà Giang.

Hội nghị "Nâng cao chất lượng giáo dục giai đoạn 2023-2030" [3] được tổ chức tại tỉnh Hà Giang vào tháng 8/2023 đã chỉ ra thực trạng giáo dục của tỉnh Hà Giang. Tỷ lệ thi đỗ tốt nghiệp THPT và kết quả điểm thi trung bình các môn thi tốt nghiệp THPT xếp ở vị trí thấp nhất so với các tỉnh/TP trong cả nước (63/63 địa phương). Tỷ lệ trường được công nhận đạt chuẩn quốc gia chưa cao. Tình trạng xã trắng (thiếu trường học) đã được giải quyết nhưng cơ sở vật chất vẫn còn nhiều thiếu thốn. Đội ngũ giáo viên thiếu và yếu, trong đó năm học tới toàn tỉnh thiếu 2.947 giáo viên, nhân viên. Đội ngũ cán bộ quản lý trường học một số nơi hạn chế nghiệp vụ quản lý và năng lực đội ngũ nhà giáo cũng chưa đáp ứng yêu cầu. Đây cũng là vấn đề chung của nhiều tỉnh trong khu vực Trung du và miền núi phía Bắc.

Nguyên nhân chủ yếu là do đời sống nhân dân địa phương còn nghèo. Ở tỉnh Hà Giang có tới 7 huyện nghèo, tỷ lệ hộ nghèo chiếm 37,08%; đồng bào dân tộc thiểu số chiếm tới 87%; hơn 1.900 điểm trường nhỏ lẻ; thu nhập bình quân đầu người thấp so với các tỉnh trong khu vực. Kinh phí UBND tỉnh chi cho giáo dục thấp hơn so với các tỉnh. Chính quyền địa phương vẫn đang nỗ lực quan tâm, sửa đổi, bổ sung các chính sách cho phù hợp, nhất là các chính sách cho người dạy, người học, các điều kiện đảm bảo chất lượng giáo dục.

Ngoài ra, các tỉnh Hà Giang, Hòa Bình, Sơn La cũng là tâm điểm nhức tối của vụ gian lận thi cử trong Kỳ thi quốc gia năm 2018. Cụ thể, tại Hà Giang có 330 bài thi của 114 thí sinh được sửa điểm thi từ 1-8 điểm. Hàng trăm bài thi bị chỉnh sửa trực tiếp, không đúng với bài thi gốc gây nhiều khó khăn cho cơ quan điều tra.

Thực trạng học Ngoại ngữ chung ở Việt Nam

Ngoại ngữ là môn thi có nhiều điểm đặc biệt. Phổ điểm của môn thi này cho thấy số lượng thí sinh đạt điểm thấp nhiều. Ngoại ngữ cũng đồng thời là môn thi có nhiều điểm liệt. Điều này thể hiện

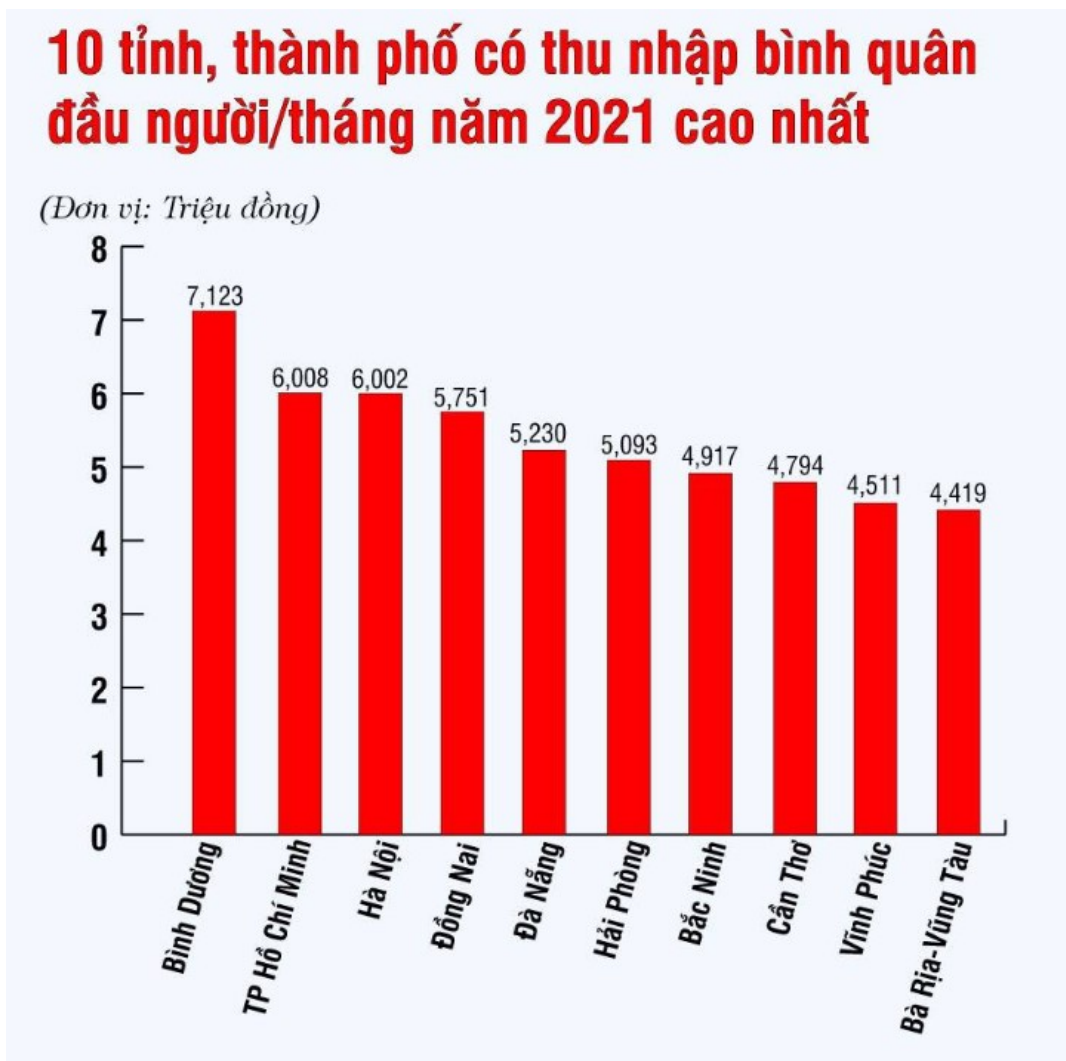
thực trạng đáng quan ngại về dạy và học Ngoại ngữ ở Việt Nam.

Một trong những nguyên nhân chính dẫn đến hạn chế, yếu kém trong việc dạy và học ngoại ngữ là do sự phát triển kinh tế - xã hội không đồng đều giữa khu vực đồng bằng, thành thị với vùng nông thôn, miền núi. Đặc biệt ở những địa phương vùng núi cao, xa xôi, hẻo lánh, nhiều học sinh nói chưa thạo tiếng kinh, đội ngũ giáo viên thừa, thiếu cục bộ ở các cấp học. Nhiều địa phương không có đủ giáo viên Ngoại ngữ. [4]

Năng lực Ngoại ngữ của học sinh TP. Hồ Chí Minh và Bình Dương

Qua phân tích, có thể thấy mặc dù TP. Hồ Chí Minh chỉ dẫn đầu cả nước duy nhất ở điểm thi Ngoại ngữ, tuy nhiên sự cách biệt này rất đáng kể và duy trì trong nhiều năm liên tiếp. Điều này phản ánh phần nào năng lực và điều kiện học Ngoại ngữ ở địa phương. Cách đây 20 năm, TP. Hồ Chí Minh đi tiên phong trong việc dạy chương trình tiếng Anh tăng cường ở bậc tiểu học. Để nâng cao chất lượng dạy học tiếng Anh trong nhà trường phổ thông, từ năm học 2015-2016, TP.HCM thực hiện đề án “Dạy và học các môn toán, khoa học và tiếng Anh tích hợp chương trình Anh và Việt Nam” (gọi tắt tiếng Anh tích hợp) tại các trường công lập trên địa bàn TP.HCM. Có thể nói TP. Hồ Chí Minh tập trung chú trọng triển khai nhiều mô hình mới, cách làm hiệu quả và mở rộng chương trình giáo dục tiên tiến theo xu hướng hội nhập quốc tế.

Ngoài TP. Hồ Chí Minh thì còn có Bình Dương là tỉnh đứng thứ hai với điểm trung bình môn Ngoại ngữ và số bài thi đạt điểm cao dẫn đầu cả nước (rất sát với TP. Hồ Chí Minh). Một điểm chung đó là Bình Dương và TP. Hồ Chí Minh là hai tỉnh thành dẫn đầu cả nước về thu nhập bình quân đầu người (2021) [5]. Mặc dù vậy, bên cạnh điều kiện kinh tế vẫn còn nhiều tác động của chính sách giáo dục hiệu quả tác động tích cực đến kết quả dạy và học Ngoại ngữ của hai tỉnh thành này.



Hình 11: 10 tỉnh thành có thu nhập bình quân đầu người cao nhất cả nước năm 2021
(Báo Quảng Ninh)

2.3 Phân tích Dashboard "Điểm thi theo tổ hợp"

2.3.1 Mục tiêu dashboard

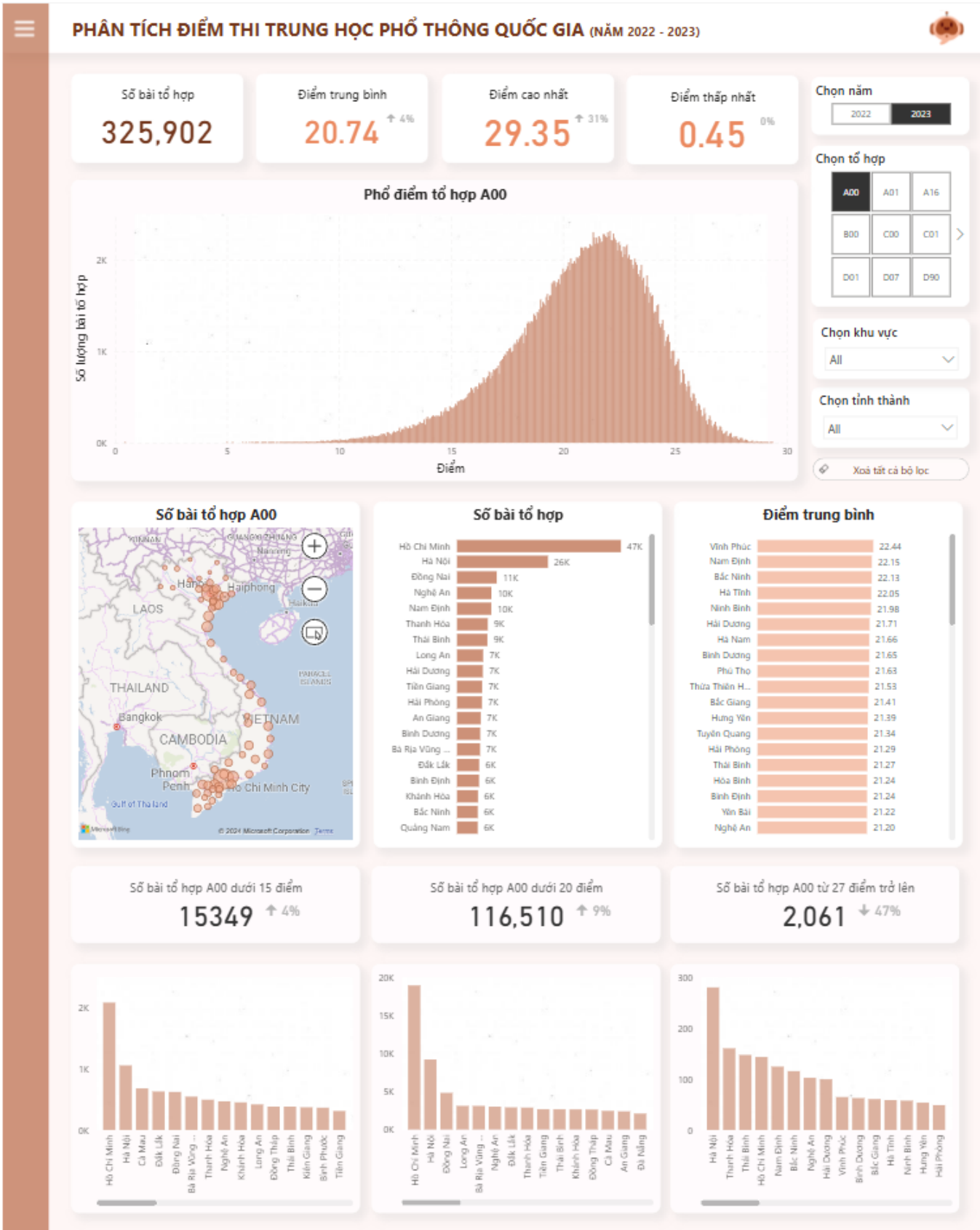
Mục tiêu chung của dashboard:

- Cung cấp các thông tin tổng quan về điểm thi Trung học Phổ thông Quốc gia tại Việt Nam theo **tổ hợp môn thi**
- Thời gian: năm 2022 và 2023
- Phạm vi: 63 tỉnh thành trên cả nước

- Đối tượng: Một số tổ hợp tiêu biểu:
 - A00: Toán, Vật lí, Hóa học
 - A01: Toán, Vật lí, Ngoại ngữ
 - A16: Toán, KHTN, Ngữ văn
 - B00: Toán, Hóa học, Sinh học
 - C00: Ngữ văn, Lịch sử, Địa lí
 - C01: Ngữ văn, Toán, Vật lí
 - D01: Toán, Ngữ văn, Ngoại ngữ
 - D07: Toán, Hóa học, Ngoại ngữ
 - D90: Toán, KHTN, Ngoại ngữ
 - D96: Toán, KHXH, Ngoại ngữ

Mục tiêu cụ thể của dashboard:

- Cung cấp cái nhìn tổng quan về kết quả thi THPTQG từ năm 2022 đến 2023, bao gồm số lượng bài thi, điểm trung bình, điểm cao nhất và điểm thấp nhất.
- Phân tích phổ điểm cho các tổ hợp môn thi, giúp người xem dễ dàng nhận biết xu hướng điểm số của từng tổ hợp môn.
- Hiện thị số lượng bài thi theo từng khu vực và tỉnh thành, giúp đánh giá sự phân bố và mức độ tham gia thi của các địa phương.
- Cung cấp thông tin về điểm trung bình theo từng tỉnh thành, giúp so sánh hiệu quả học tập giữa các khu vực.
- Phân tích số lượng bài thi có điểm dưới các mức điểm quan trọng như 15, 20 và từ 27 điểm trở lên, nhằm đánh giá chất lượng giáo dục và hiệu quả học tập.
- Tạo điều kiện cho người dùng lọc và xem dữ liệu theo các tiêu chí như năm, môn thi, khu vực và tỉnh thành, giúp tùy chỉnh thông tin hiển thị theo nhu cầu cá nhân.



Hình 12: Tổng quan dashboard Điểm thi theo tổ hợp

2.3.2 Các thành phần trực quan của dashboard

Dashboard "Điểm thi theo tổ hợp" bao gồm các thành phần chính sau đây:

Tiêu đề

Tiêu đề của dashboard cung cấp thông tin về nội dung được thể hiện trong dashboard (Phân tích điểm thi Trung học Phổ thông Quốc gia) và thời gian trực quan (năm 2022 - 2023).

Tiêu đề được đặt đơn giản, ngắn gọn, bao hàm vừa đủ lượng thông tin. Kích cỡ chữ của thông tin và thời gian khác nhau giúp người dùng nhanh chóng phân biệt được thông tin chính và phụ.

Bộ lọc

Khu vực bộ lọc cho phép người dùng lựa chọn các tiêu chí để theo dõi thông tin trên dashboard, bao gồm: Năm, Tổ hợp, Khu vực và Tỉnh thành.

Các bộ lọc này giúp người dùng tùy chỉnh dữ liệu hiển thị theo nhu cầu, tập trung vào những tổ hợp môn cụ thể hoặc khoảng thời gian nhất định. Bộ lọc Khu vực và Tỉnh thành có thể chọn nhiều giá trị, cho phép theo dõi thông tin theo các phạm vi khác nhau: trên toàn quốc, theo từng khu vực hoặc từng tỉnh thành.

Các bộ lọc được bố trí ở vị trí dễ thao tác, gần với tiêu đề của dashboard, tạo sự thuận tiện cho người dùng.

Số liệu thống kê tổng quan

Loại trực quan: Số liệu

Phần số liệu thống kê tổng quan cung cấp các số liệu chính xác và cụ thể về:

- Số lượng thí sinh dự thi theo từng tổ hợp môn trong phạm vi được chọn.
- Điểm trung bình, điểm cao nhất và điểm thấp nhất của các tổ hợp môn trong phạm vi được chọn.

Các số liệu này được trình bày rõ ràng trong các ô lớn, với kích cỡ chữ lớn và màu sắc nổi bật, giúp thu hút sự chú ý của người dùng. Màu sắc của các số liệu cũng đồng bộ với cách mã hóa màu sắc trong toàn bộ dashboard, giúp người dùng dễ dàng nhận biết và phân biệt thông tin.

Biểu đồ phân bố điểm

Loại trực quan: Biểu đồ cột đứng

Biểu đồ phân bố điểm là thành phần chính, cung cấp thông tin chi tiết về phổ điểm của các tổ hợp môn. Biểu đồ này được đặt ở vị trí trung tâm của dashboard, là vị trí dễ thu hút sự chú ý và quan sát nhất của người dùng.

Biểu đồ cột đứng giúp hiển thị trực quan sự phân bố điểm thi của thí sinh, cho phép người dùng dễ dàng so sánh và phân tích dữ liệu.

Các biểu đồ về tình hình dự thi chung trên cả nước

Loại trực quan: Bản đồ bong bóng, biểu đồ cột ngang

Các biểu đồ này cung cấp cái nhìn tổng quan về tình hình dự thi chung trên phạm vi cả nước, bao gồm: số lượng thí sinh và điểm trung bình theo từng tỉnh thành, bao gồm số liệu và mối tương quan giữa những số liệu đó.

- **Bản đồ phân bố thí sinh trên cả nước**

Loại trực quan: Bản đồ bong bóng

Biểu đồ này thể hiện trên bản đồ địa lý sự tương quan về số thí sinh dự thi giữa các tỉnh thành; sự tập trung thí sinh dự thi ở một số khu vực nhất định trên bản đồ.

- **Biểu đồ số lượng thí sinh dự thi theo từng tỉnh thành**

Loại trực quan: Biểu đồ cột ngang

Biểu đồ này cung cấp thông tin về số lượng thí sinh dự thi theo từng tỉnh, giúp người dùng nắm bắt tình hình tham gia thi của các tỉnh thành trên cả nước.

Biểu đồ cột ngang giúp hiển thị sự so sánh rõ ràng giữa các tỉnh thành, cho phép người dùng dễ dàng nhận biết sự chênh lệch về số lượng thí sinh dự thi.

- **Biểu đồ điểm trung bình theo từng tỉnh thành**

Loại trực quan: Biểu đồ cột ngang

Biểu đồ này cung cấp thông tin về điểm trung bình theo từng tỉnh, giúp người dùng nắm bắt được chất lượng thi cử của các tỉnh thành trên cả nước.

Biểu đồ cột ngang hiển thị sự so sánh điểm trung bình giữa các tỉnh, giúp người dùng dễ dàng đánh giá và phân tích sự chênh lệch về điểm thi giữa các tỉnh thành.

Các số liệu và biểu đồ về các mức điểm đặc biệt

Loại trực quan: Số liệu, Biểu đồ cột đứng

Khu vực cuối cùng của dashboard cung cấp thông tin về tình hình điểm thi tại các tỉnh thành khi xem xét các mức điểm đặc biệt và thường được quan tâm:

- Số bài thi tổ hợp dưới điểm 15 trên cả nước và thứ tự giữa các tỉnh thành.
- Số bài thi tổ hợp dưới điểm 20 trên cả nước và thứ tự giữa các tỉnh thành.
- Số bài thi điểm cao (từ 27 điểm trở lên) trên cả nước và thứ tự giữa các tỉnh thành (tỉnh nào có ít, nhiều thí sinh đạt điểm cao)

Trong đó, tổng số bài thi trên cả nước được biểu diễn bằng số liệu, còn thứ tự giữa các tỉnh thành được biểu diễn bằng biểu đồ cột. Số liệu và biểu đồ được nhóm lại với nhau từng cặp theo chiều dọc giúp người xem dễ quan sát hơn, đồng thời giúp hai thành phần này có thể dùng chung tiêu đề, làm cho dashboard đơn giản, tinh gọn.

Các thành phần khác

- Nút xóa bộ lọc: Được đặt ở vị trí dưới cùng của khu vực bộ lọc. Nút này cho phép người dùng xóa nhanh tất cả các bộ lọc đang áp dụng trên dashboard và đưa về lựa chọn mặc định ban đầu.
- Nút thông tin: Được đặt ở vị trí góc phải trên của dashboard. Nút này cung cấp thông tin về tác giả và quyền sở hữu đối với dashboard.

2.3.3 Phân tích dashboard

Số lượng thí sinh

Hà Nội và TP. Hồ Chí Minh là hai tỉnh thành có số lượng thí sinh dự thi đông nhất.

- Đối với các tổ hợp môn thi có các môn Khoa học Xã hội (C00, D96) và tổ hợp môn chung Toán, Ngữ văn, Ngoại ngữ (D01): Hà Nội là tỉnh thành có nhiều thí sinh dự thi nhất trên cả nước.

- Đối với các tổ hợp môn thi có các môn Khoa học Tự nhiên (A00, A01, A16, B00, C01, D07, D90): TP. Hồ Chí Minh là tỉnh thành có nhiều thí sinh dự thi nhất trên cả nước.

Điểm trung bình cao nhất các tổ hợp môn thi

- A00: Ninh Bình, Bắc Ninh (2022), Vĩnh Phúc, Nam Định (2023)
- A01: Bình Dương, Hải Phòng(2022), Vĩnh Phúc, Hà Nội (2023)
- A16: Ninh Bình, Vĩnh Phúc (2022), Vĩnh Phúc, Hà Nam (2023)
- B00: Tuyên Quang, Ninh Bình (2022), Vĩnh Phúc, Bình Dương (2023)
- C00: Vĩnh Phúc, Nam Định (2022), Vĩnh Phúc, Ninh Bình (2023)
- C01: Ninh Bình, Vĩnh Phúc (2022), Vĩnh Phúc, Bắc Ninh (2023)
- D01: Vĩnh Phúc, Bình Dương (2022, 2023)
- D07: Bình Dương, Yên Bái (2022), Bình Dương, Vĩnh Phúc (2023)
- D90: Yên Bái, Hải Phòng (2022), Vĩnh Phúc, Hải Phòng (2023)
- D96: Vĩnh Phúc, Bình Dương (2022), Bình Dương, Vĩnh Phúc (2023)

Nhìn chung, các tỉnh đứng đầu trong các tổ hợp môn thi tập trung ở vùng đồng bằng Sông Hồng. Tuyên Quang, Yên Bái là 2 tỉnh trung du và miền núi phía Bắc nằm trong top 2. Trong khi đó, Bình Dương (Đông Nam Bộ) là tỉnh duy nhất không nằm ở miền Bắc góp mặt trong danh sách.

Điểm trung bình thấp nhất các tổ hợp môn thi

- A00: Cà Mau, Kiên Giang (2022), Cà Mau, Trà Vinh (2023)
- A01: Cà Mau, Hậu Giang (2022, 2023)
- A16: Đắc Nông, Lai Châu (2022), Cà Mau, Hà Giang (2023)
- B00: Bình Phước, Bà Rịa Vũng Tàu (2022), Cà Mau, Lai Châu (2023)

- C00: Hà Giang, Phú Yên (2022), Hà Giang, Đắk Lắk (2023)
- C01: Đắk Nông, Lai Châu (2022), Cà Mau, Đắk Nông (2023)
- D01: Hà Giang, Cao Bằng (2022, 2023)
- D07: Cà Mau, Hậu Giang (2022), Lai Châu, Cà Mau (2023)
- D90: Cà Mau, Hậu Giang (2022), Lai Châu, Cà Mau (2023)
- D96: Hà Giang, Đắk Lắk (2022), Hà Giang, Cao Bằng (2023)

Nhìn chung, các tỉnh có điểm trung bình thấp nhất theo các tổ hợp môn thi tập trung chủ yếu ở khu vực Trung du và miền núi phía Bắc và Đồng bằng sông Cửu Long. Cà Mau là tỉnh có số lượng tổ hợp môn nằm trong danh sách điểm trung bình thấp nhất cao nhất, xuất hiện trong nhiều tổ hợp và cả hai năm. Điều này cho thấy cần có các biện pháp cải thiện chất lượng giáo dục tại tỉnh này. Tỉnh Hà Giang cũng xuất hiện nhiều trong danh sách, đặc biệt là trong các tổ hợp C00, D01, và D96. Điều này phản ánh một phần chất lượng giáo dục chưa đồng đều giữa các tỉnh miền núi và các tỉnh đồng bằng.

Các tỉnh khác như Hậu Giang, Lai Châu, Đắk Nông, và Cao Bằng cũng xuất hiện nhiều lần trong danh sách, cho thấy sự cần thiết của việc tăng cường đầu tư vào giáo dục ở các khu vực này. Tóm lại, phân tích này cho thấy sự cần thiết của việc cải thiện chất lượng giáo dục tại một số tỉnh thành, đặc biệt là các tỉnh thường xuyên có điểm trung bình thấp nhất. Các biện pháp có thể bao gồm tăng cường đào tạo giáo viên, cải thiện cơ sở vật chất, và cung cấp thêm các tài liệu học tập cho học sinh.

Khu vực miền Bắc

- Hà Nội có số lượng thí sinh rất lớn, gấp 2.5 lần so với tỉnh đứng thứ 2. Các tỉnh cũng có số lượng thí sinh dự thi nhiều là: Hải Phòng, Nam Định.
- Các tỉnh xếp hạng cao nhất về điểm trung bình: Ninh Bình (A00, A16, C01), Vĩnh Phúc (A00, A01, A16, B00, C00, C01, D01, D07, D90, D96), Hải Phòng (A01), Tuyên Quang (B00), Yên Bái (D07, D90)

- Các tỉnh xếp hạng thấp nhất về điểm trung bình: Lai Châu (A00, A01, A16, B00, C01, D07, D90), Hà Giang (A16, B00, C00, C01, D01, D96).

Khu vực miền Trung

- Các tỉnh có nhiều thí sinh dự thi nhất: Nghệ An, Thanh Hóa.
- Các tỉnh xếp hạng cao nhất về điểm trung bình: Hà Tĩnh (A00, A01, A16, B00, C00, C01, D01, D96), Bình Định (A01, D07, D90)
- Các tỉnh xếp hạng thấp nhất về điểm trung bình: Đắk Lắk (A00, B00, C00, D96), Đắk Nông (A00, A01, A16, C01, D01, D07, D90), Đà Nẵng (A16), Phú Yên (C00)

Khu vực miền Nam

- TP. Hồ Chí Minh có số lượng thí sinh rất lớn, gấp 2.5 lần so với tỉnh đứng thứ 2. Tỉnh cũng có số lượng thí sinh dự thi nhiều là: Đồng Nai, An Giang.
- Các tỉnh xếp hạng cao nhất về điểm trung bình: Bình Dương (A00, A01, A16, B00, C00, C01, D01, D07, D90, D96)
- Các tỉnh xếp hạng thấp nhất về điểm trung bình: Cà Mau (A00, A01, A16, B00, C01, D07, D90), Bình Phước (A16, B00, C01), Trà Vinh (C00, D01, D96)

Phổ điểm

A00: Phổ điểm cả 2 năm đều lệch nhẹ về phía bên phải. Số điểm có nhiều thí sinh đạt được nhất rơi vào khoảng 21-23 điểm

A01: Phổ điểm cả 2 năm đều lệch nhẹ về phía bên phải. Số điểm có nhiều thí sinh đạt được nhất rơi vào khoảng 20-21 điểm

A16: Phổ điểm cả 2 năm đều lệch nhẹ về phía bên phải. Số điểm có nhiều thí sinh đạt được nhất rơi vào khoảng 21-22 điểm

B00: Phổ điểm cả 2 năm đều lệch nhẹ về phía bên phải. Phổ điểm năm 2023 lệch nhiều về bên phải hơn 2022. Số điểm có nhiều thí sinh đạt được nhất rơi vào khoảng 19-21 điểm (2022), 21-22 điểm (2023)

C00: Phổ điểm cả 2 năm đều lệch nhẹ về phía bên phải. Phổ điểm năm 2023 lệch nhiều về bên phải hơn 2022. Số điểm có nhiều thí sinh đạt được nhất rơi vào khoảng 19-21 điểm (2022), 18-20 điểm (2023)

C01: Phổ điểm cả 2 năm đều lệch nhẹ về phía bên phải. Số điểm có nhiều thí sinh đạt được nhất rơi vào khoảng 21-23 điểm

D01: Phổ điểm cả 2 năm đều lệch nhẹ về phía bên phải. Phổ điểm năm 2023 lệch nhiều về bên phải hơn 2022. Số điểm có nhiều thí sinh đạt được nhất rơi vào khoảng 18-20 điểm

D07: Phổ điểm cả 2 năm đều lệch nhẹ về phía bên phải. Số điểm có nhiều thí sinh đạt được nhất rơi vào khoảng 20-22 điểm

D90: Phổ điểm cả 2 năm đều lệch nhẹ về phía bên phải. Phổ điểm năm 2023 lệch nhiều về bên phải hơn 2022. Số điểm có nhiều thí sinh đạt được nhất rơi vào khoảng 18,5-21 điểm (2022), 19,5-22 điểm (2023)

D96: Phổ điểm cả 2 năm đều lệch nhẹ về phía bên phải. Số điểm có nhiều thí sinh đạt được nhất rơi vào khoảng 16-19 điểm. Tổ hợp D96 cũng là tổ hợp có điểm trung bình thấp nhất trong các tổ hợp.

Số bài thi điểm thấp (dưới 15 điểm)

Các tỉnh thành có nhiều bài thi dưới 15 điểm:

- Các tỉnh thành có điểm trung bình thấp cũng là các tỉnh có nhiều bài thi dưới 15 điểm như: Đắk Lắk, Cà Mau, Hà Giang
- Các tỉnh thành có nhiều bài thi 15 điểm do có số lượng thí sinh cao hơn hẳn các tỉnh thành còn lại: Hà Nội, Hồ Chí Minh, Thanh Hóa, Nghệ An, Đồng Nai

Số lượng bài thi dưới 15 điểm ở 2 năm 2022, 2023 khá tương đương nhau.

Số bài thi điểm khá thấp (dưới 20 điểm)

Sự thay đổi về số bài thi dưới 20 điểm của năm 2023 so với năm 2022:

- Tăng ở các tổ hợp: A00, C00, D96
- Giảm ở các tổ hợp: A16, B00, D01, D07, D90
- Không thay đổi nhiều đối với tổ hợp: A01, C01

Số bài thi điểm cao (từ 27 trở lên)

Các tỉnh có nhiều bài thi điểm cao:

- A00: Hà Nội
- A01: Hà Nội, Hồ Chí Minh
- A16: Bắc Ninh, Nam Định, Hải Phòng, Thanh Hóa
- B00: Hồ Chí Minh
- C00: Thanh Hóa, Vĩnh Phúc, Hải Phòng, Nghệ An
- C01: Hải Phòng, Bắc Ninh, Thanh Hóa
- D01: Hà Nội
- D07: Hồ Chí Minh, Hà Nội
- D90: Hồ Chí Minh, Hà Nội
- D96: Hà Nội

Sự thay đổi về số bài thi điểm cao của năm 2023 so với năm 2022:

- Tăng ở các tổ hợp: A16, B00, D01, D90
- Giảm ở các tổ hợp: A00, A01, C00, C01, D07, D96

2.4 Phân tích Dashboard "Điểm chuẩn"

2.4.1 Mục tiêu dashboard

Mục tiêu chung của dashboard:

- Cung cấp một công cụ mạnh mẽ và trực quan để hỗ trợ người dùng trong việc tìm kiếm và lựa chọn trường đại học phù hợp với mức điểm và các tiêu chí cá nhân của họ
- Thời gian: năm 2023

- Phạm vi: 63 tỉnh thành trên cả nước
- Đối tượng: Một số phương thức xét tuyển tiêu biểu (tùy theo năm):
 - Học bạ
 - THPTQG (Trung học Phổ thông Quốc gia)
 - THPTQG (Trung học Phổ thông Quốc gia) - Thang 40
 - DGNLHCM (Đánh giá năng lực Hồ Chí Minh)
 - DGNLHCM (Đánh giá năng lực Hồ Chí Minh) - Quy đổi
 - DGNLQGHN (Đánh giá năng lực Quốc gia Hà Nội)
 - DGNLQGHN (Đánh giá năng lực Quốc gia Hà Nội) - Quy đổi
 - DGTD (Đánh giá tư duy)
 - DGTD (Đánh giá tư duy) - Quy đổi
 - BKHCM (Bách Khoa Hồ Chí Minh)

Mục tiêu cụ thể của dashboard:

- Cung cấp cái nhìn tổng quan về điểm chuẩn của một số phương thức xét tuyển tiêu biểu (tùy theo năm) từ năm 2018 đến 2023, bao gồm điểm chuẩn trung bình của năm đang chọn và phần trăm tăng/ giảm so với 1 năm trước, điểm chuẩn trung bình của năm 2023, điểm chuẩn cao nhất và điểm chuẩn thấp nhất của năm đang chọn và phần trăm tăng/ giảm so với 1 năm trước
- Giúp người dùng nhận diện các trường đại học hàng đầu có điểm chuẩn cao nhất để hỗ trợ người dùng hiểu rõ hơn về những trường có yêu cầu đầu vào cao và có thể được coi là có chất lượng đào tạo tốt hơn
- Giúp người dùng nhận diện các trường có yêu cầu đầu vào thấp hơn. Điều này có thể hữu ích cho học sinh có kết quả thi không cao, hoặc những người tìm kiếm các lựa chọn học tập khác dễ tiếp cận hơn
- Cung cấp cái nhìn tổng quan về số lượng các trường đại học công lập và dân lập tại Việt Nam, giúp người dùng hiểu rõ hơn về cấu trúc và sự phân bố của hệ thống giáo dục đại học. Ngoài ra còn cung cấp thêm tỉ lệ tăng/ giảm của tổng số trường so với 1 năm trước

- Cung cấp bộ lọc linh hoạt để người dùng xem dữ liệu theo nhiều tiêu chí khác nhau như năm, khu vực, ngành học, phương thức xét tuyển... Điều này giúp tùy chỉnh thông tin hiển thị theo nhu cầu cá nhân

2.4.2 Các thành phần trực quan của dashboard

Dashboard "Điểm chuẩn" bao gồm các thành phần chính sau đây:

Tiêu đề

Tiêu đề của dashboard cung cấp thông tin về nội dung được thể hiện trong dashboard (Phân tích điểm chuẩn Đại học tại Việt Nam) và thời gian trực quan (năm 2018 đến 2023).

Tiêu đề được đặt đơn giản, ngắn gọn, bao hàm vừa đủ lượng thông tin. Kích cỡ chữ của thông tin và thời gian khác nhau giúp người dùng nhanh chóng phân biệt được thông tin chính và phụ.

Bộ lọc

Khu vực bộ lọc cho phép người dùng lựa chọn các tiêu chí để theo dõi thông tin trên dashboard, bao gồm: Năm, Khu vực, Ngành, Phân ngành, Tên ngành và Phương thức xét tuyển.

Các bộ lọc này giúp người dùng tùy chỉnh dữ liệu hiển thị theo nhu cầu, tập trung vào 1 phương thức xét tuyển cụ thể hoặc khoảng thời gian nhất định. Bộ lọc Khu vực, Ngành, Phân ngành, Tên ngành có thể chọn nhiều giá trị, cho phép theo dõi thông tin theo các phạm vi khác nhau: trên toàn quốc, theo những ngành, phân ngành hoặc tên ngành.

Các bộ lọc được bố trí ở vị trí dễ thao tác, gần với tiêu đề của dashboard, tạo sự thuận tiện cho người dùng.

Số liệu thống kê tổng quan

Loại trực quan: Số liệu

Phần số liệu thống kê tổng quan cung cấp các số liệu chính xác và cụ thể về:

- Điểm chuẩn trung bình của năm đang chọn và phần trăm tăng/ giảm so với 1 năm trước, điểm chuẩn trung bình của năm 2023, điểm chuẩn cao nhất và điểm chuẩn thấp nhất của năm đang chọn và phần trăm tăng/ giảm so với 1 năm trước theo phương thức xét tuyển đang được chọn

Các số liệu này được trình bày rõ ràng trong các ô lớn, với kích cỡ chữ lớn và màu sắc nổi bật, giúp thu hút sự chú ý của người dùng. Màu sắc của các số liệu cũng đồng bộ với cách mã hóa màu sắc trong toàn bộ dashboard, giúp người dùng dễ dàng nhận biết và phân biệt thông tin.

Biểu đồ Phân bố trường đại học theo tỉnh/ thành phố

Loại trực quan: bản đồ

Biểu đồ này thể hiện thông tin phân bố số lượng trường công bố điểm chuẩn theo các bộ lọc được chọn trên đất nước Việt Nam.

Màu trên biểu đồ được phân bố từ màu nhạt đến màu đậm. Màu càng nhạt thì số lượng trường công bố điểm chuẩn ở tỉnh thành đó càng ít và ngược lại.

Note: Vì với hầu như mỗi năm và phương thức được phân tích thì số lượng trường công bố điểm chuẩn đều tập trung nhiều nhất ở Hà Nội sau đó đến TPHCM nên ở mục kế tiếp sẽ không cần thiết phải phân tích kỹ lại ở biểu đồ này.

Biểu đồ Top 5 trường có điểm chuẩn cao nhất và thấp nhất

Loại trực quan: Biểu đồ cột đứng

Biểu đồ này thể hiện thông tin top 5 trường có điểm chuẩn cao nhất và thấp nhất theo phương thức xét tuyển và năm đang được chọn, giúp người dùng có cái nhìn về yêu cầu đầu vào để có thể lựa chọn ngôi trường phù hợp.

Biểu đồ cột đứng giúp hiển thị trực quan và so sánh giữa các trường có điểm thi cao nhất và thấp nhất theo phương thức xét tuyển và năm đang được chọn.

Biểu đồ Số đơn vị đào tạo

Loại trực quan: Biểu đồ donut

Biểu đồ này cung cấp cái nhìn tổng quan về số lượng các trường đại học công lập và dân lập tại Việt Nam, giúp người dùng hiểu rõ hơn về cấu trúc và sự phân bố của hệ thống giáo dục đại học.

Biểu đồ Donut thể hiện rõ ràng phân phối giữa số lượng các trường công lập và dân lập, giúp người dùng dễ dàng nhận biết tỉ lệ tương đối của hai loại trường này trong tổng số đơn vị đào tạo.

Ngoài ra nhóm cũng cung cấp thêm tỉ lệ tăng/ giảm của tổng số trường so với 1 năm trước trên biểu đồ này

Các thành phần khác

- Nút thông tin: được đặt ở vị trí góc phải trên của dashboard. Nút này cung cấp thông tin về tác giả và quyền sở hữu đối với dashboard.

2.4.3 Phân tích dashboard

2.4.3.1 Năm 2018

Là khi mà các phương thức xét tuyển vẫn còn đang khá hạn chế, xét tuyển bằng điểm chuẩn THPTQG ngày đó được xem là thịnh hành nhất:

- **Số liệu thống kê tổng quan:**
 - Điểm chuẩn trung bình: 16.76 vì năm 2018 là năm có độ khó của đề thi rất khắc nghiệt dẫn đến điểm thi của học sinh không cao và điểm chuẩn trung bình cũng thấp
 - Điểm chuẩn cao nhất: 27.75 là một điểm chuẩn rất cao so với mặt bằng điểm chuẩn trung bình
 - Điểm chuẩn thấp nhất: 13.00 là một điểm chuẩn không quá bất ngờ vì độ khó vốn có của đề thi năm 2018
- **Biểu đồ top 5 trường có điểm chuẩn cao nhất và thấp nhất:**
 - Điểm chuẩn cao nhất thuộc về Học Viện Biên Phòng với 27.75 điểm, tiếp đến là các trường Đại Học Quy Nhơn, Đại Học Sư Phạm Kỹ Thuật TPHCM, Học Viện Cảnh Sát Nhân Dân, UEB - Đại Học Kinh Tế - Đại Học Quốc Gia Hà Nội
 - Có hơn 5 trường có mức điểm thấp nhất là Đại Học Công Nghệ Vạn Xuân, Đại Học Công Nghiệp Quảng Ninh, Đại Học Công Nghiệp Việt Trì, Đại Học Cửu Long, Đại Học Quang Trung...
- **Biểu đồ Số đơn vị đào tạo**
 - Có tổng 225 đơn vị đào tạo với số lượng đại học công lập chiếm phần lớn
- Ngoài ra còn có nhiều bộ lọc để giúp người dùng tiến hành cá nhân hóa thông tin cần thiết

2.4.3.2 Năm 2019

Lúc này các phương thức xét tuyển vẫn còn đang khá hạn chế, xét tuyển bằng điểm chuẩn THPTQG ngày đó được xem là thịnh hành nhất:

- **Số liệu thống kê tổng quan:**

- Điểm chuẩn trung bình: 17.36 là một mức điểm đã được tăng nhẹ thêm 4% lên so với năm 2018 chứng tỏ đề thi đã được giảm nhẹ độ khó đi
- Điểm chuẩn cao nhất: 29.5 là một điểm chuẩn rất cao so với mặt bằng điểm chuẩn trung bình
- Điểm chuẩn thấp nhất: 10.5

- **Biểu đồ top 5 trường có điểm chuẩn cao nhất và thấp nhất:**

- Điểm chuẩn cao nhất thuộc về Đại Học Sư Phạm Thành Phố Hồ Chí Minh với 29.5 điểm, tiếp đến là các trường Đại Học Văn Hóa Hà Nội, Đại Học Ngoại Ngữ - Đại Học Quốc Gia Hà Nội, USSH - Đại Học Khoa Học Xã Hội Và Nhân Văn - Đại Học Quốc Gia Hà Nội, Đại Học Kiểm Sát Hà Nội
- Điểm chuẩn thấp nhất (10.5) nằm ở Viện Đại Học Mở Hà Nội, tiếp đến là Đại học Kinh Tế Nghệ An, Đại Học Kỹ Thuật Công Nghệ Cần Thơ...

- **Biểu đồ Số đơn vị đào tạo**

- Có tổng 244 đơn vị đào tạo với số lượng đại học công lập chiếm phần lớn

- Ngoài ra còn có nhiều bộ lọc để giúp người dùng tiến hành cá nhân hóa thông tin cần thiết

2.4.3.3 Năm 2020

Là khi mà các phương thức xét tuyển vẫn còn đang khá hạn chế, xét tuyển bằng điểm chuẩn THPTQG ngày đó được xem là thịnh hành nhất:

- **Số liệu thống kê tổng quan:**

- Điểm chuẩn trung bình: 18.59 là một mức điểm đã được tăng lên khá đáng kể so với năm 2019 (tăng 7%) chứng tỏ đề thi năm nay mặt bằng thí sinh đạt điểm cao nhiều hơn khiến điểm chuẩn phải tăng lên

- Điểm chuẩn cao nhất: 30.00 là một điểm chuẩn tuyệt đối và rất cao so với mặt bằng điểm chuẩn trung bình
- Điểm chuẩn thấp nhất: 10.00

• **Biểu đồ top 5 trường có điểm chuẩn cao nhất và thấp nhất:**

- Điểm chuẩn cao nhất thuộc về Đại Học Hàng Hải với 30.00 điểm, tiếp đến là các trường USSH - Đại Học Khoa Học Xã Hội Và Nhân Văn - Đại Học Quốc Gia Hà Nội, Đại Học Kiểm Sát Hà Nội, Học Viện Báo Chí - Tuyên Truyền, Đại Học Bách Khoa Hà Nội
- Điểm chuẩn thấp nhất là 10.00 nằm ở Đại Học Nguyễn Tất Thành, tiếp đến là Đại Học Phòng Cháy Chữa Cháy (Hệ Dân sự Phía Nam), Đại Học Phòng Cháy Chữa Cháy (Hệ Dân sự Phía Bắc)...

• **Biểu đồ Số đơn vị đào tạo**

- Có tổng 220 đơn vị đào tạo với số lượng đại học công lập chiếm phần lớn
- Ngoài ra còn có nhiều bộ lọc để giúp người dùng tiến hành cá nhân hóa thông tin cần thiết

2.4.3.4 Năm 2021

Năm đại dịch Covid-19 hoành hành trên khắp đất nước Việt Nam đã khiến cho việc học và thi cử trở nên hạn chế hơn. Mặc dù vậy, số lượng trường xét điểm chuẩn THPTQG vẫn còn nhiều và chiếm đa số trên cả nước. Ngoài ra, một phương thức tuyển sinh khác đã nổi lên, với số lượng trường xét tuyển bằng Học bạ tăng từ 36 lên 102 trường (tăng 158%). Tuy nhiên, do điểm học bạ ở nhiều trường đại học có mức scale điểm khác nhau nên khi phân tích sẽ không được công bằng, vì vậy chúng ta sẽ tiếp tục phân tích phương thức xét tuyển THPTQG.

• **Số liệu thống kê tổng quan:**

- Điểm chuẩn trung bình: 19.99 là một mức điểm đã được tăng lên khá đáng kể so với năm 2020 (tăng 5%) chứng tỏ đề thi năm nay mặt bằng thí sinh đạt điểm cao nhiều hơn khiến điểm chuẩn phải tăng lên
- Điểm chuẩn cao nhất: 30.5 là một điểm chuẩn rất cao so với mặt bằng điểm chuẩn trung bình

– Điểm chuẩn thấp nhất: 5.00

• **Biểu đồ top 5 trường có điểm chuẩn cao nhất và thấp nhất:**

- Điểm chuẩn cao nhất thuộc về Đại Học Hồng Đức với 30.5 điểm, tiếp đến là các trường USSH - Đại Học Khoa Học Xã Hội Và Nhân Văn - Đại Học Quốc Gia Hà Nội, Học Viện An Ninh Nhân Dân, Học Viện Cảnh Sát Nhân Dân, Học Viện Chính Trị Công An Nhân Dân
- Điểm chuẩn thấp nhất là 5.5 điểm nằm ở Đại Học Gia Định, tiếp đến là Đại Học Nha Trang, Đại Học Cửu Long...

• **Biểu đồ Số đơn vị đào tạo**

– Có tổng 211 đơn vị đào tạo với số lượng đại học công lập chiếm phần lớn

• Ngoài ra còn có nhiều bộ lọc để giúp người dùng tiến hành cá nhân hóa thông tin cần thiết

2.4.3.5 Năm 2022

Tác động của dịch Covid-19 trước đó đã gây ra nhiều khó khăn cho việc tổ chức các kỳ thi và quá trình học tập của học sinh. Vì vậy ở năm 2022 các trường đại học đã đa dạng hóa phương thức xét tuyển nhằm giảm áp lực thi cử, đảm bảo công bằng cho thí sinh và đảm bảo quá trình tuyển sinh không bị gián đoạn. Tới năm 2022 thì đã có đến 9 phương thức xét tuyển trên toàn đất nước, trong đó THPTQG và DGNLHCM là 2 phương thức tiêu biểu nhất.

• **THPTQG**

– **Số liệu thống kê tổng quan:**

- * Điểm chuẩn trung bình: 19.67 là mức điểm đã được giảm đi 2% so với năm 2021 chứng tỏ đề thi đã mang mức độ khó cao hơn so với năm trước dẫn đến việc điểm chuẩn của các trường cũng giảm đi
- * Điểm chuẩn cao nhất: 29.95 là một điểm chuẩn rất cao so với mặt bằng điểm chuẩn trung bình
- * Điểm chuẩn thấp nhất: 13,00

– **Biểu đồ top 5 trường có điểm chuẩn cao nhất và thấp nhất:**

- * Điểm chuẩn cao nhất thuộc về trường USSH - Đại Học Khoa Học Xã Hội Và Nhân Văn - Đại Học Quốc Gia Hà Nội với 29.95 điểm, tiếp đến là các trường Học Viện Khoa Học Quân Sự - Hệ Quân Sự, Đại Học Hồng Đức, Đại Học Hải Phòng, Đại Học Luật Hà Nội
- * Điểm chuẩn thấp nhất 13.00 thuộc về Đại Học Kiên Giang, tiếp sau đó là Đại Học Kiến Trúc Đà Nẵng, Đại Học Xây Dựng Miền Tây,...

– **Biểu đồ Số đơn vị đào tạo**

- * Có tổng 196 đơn vị đào tạo với số lượng đại học công lập chiếm phần lớn

• **ĐGNLHCM**

– **Số liệu thống kê tổng quan:**

- * Điểm chuẩn trung bình: 667.37
- * Điểm chuẩn cao nhất: 1001
- * Điểm chuẩn thấp nhất: 500

– **Biểu đồ top 5 trường có điểm chuẩn cao nhất và thấp nhất:**

- * Điểm chuẩn cao nhất thuộc về trường Đại Học Khoa Học Tự Nhiên - Đại Học Quốc Gia TPHCM với 1001.00 điểm, tiếp đến là các trường Đại Học Bách Khoa - Đại Học Đà Nẵng, Đại Học Kinh Tế TPHCM, Khoa Y - Đại Học Quốc Gia TPHCM, UIT - Đại Học Công Nghệ Thông Tin - Đại Học Quốc Gia TPHCM
- * .Đại học Bình Dương, Đại Học Hùng Vương - TPHCM có điểm chuẩn thấp nhất với 500.00 điểm

– **Biểu đồ Số đơn vị đào tạo**

- * Có tổng 47 đơn vị đào tạo (tăng 53% so với 2021) với số lượng đại học công lập chiếm phần lớn

- Ngoài ra còn có nhiều bộ lọc để giúp người dùng tiến hành cá nhân hóa thông tin cần thiết

2.4.3.6 Năm 2023

Các trường đại học tại Việt Nam tiếp tục đa dạng hóa các phương thức xét tuyển để tạo điều kiện

thuận lợi cho thí sinh và đảm bảo tính công bằng. Ở năm 2023, Đại Học Bách Khoa Hà Nội đã tổ chức thêm phương thức Đánh giá tư duy nên số lượng phương thức xét tuyển trên cả nước lên 10. Tuy vậy, số lượng trường sử dụng phương thức xét tuyển bằng điểm THPTQG và ĐGNLHCM.

• THPTQG

– Số liệu thống kê tổng quan:

- * Điểm chuẩn trung bình: 20.52 là một mức điểm đã được tăng lên khá đáng kể so với năm 2022 (tăng 4%) chứng tỏ đề thi năm nay mặt bằng thí sinh đạt điểm cao nhiều hơn khiến điểm chuẩn phải tăng lên
- * Điểm chuẩn cao nhất: 29.42 là một điểm chuẩn rất cao so với mặt bằng điểm chuẩn trung bình
- * Điểm chuẩn thấp nhất: 14,00

– Biểu đồ top 5 trường có điểm chuẩn cao nhất và thấp nhất:

- * Điểm chuẩn cao nhất thuộc về trường Đại Học Bách Khoa Hà Nội với 29.42 điểm, tiếp đến là các trường Học Viện Báo Chí - Tuyên Truyền, Đại Học Ngoại Thương (Cơ sở phía Bắc), Đại Học Sư Phạm Hà Nội, Học Viện Ngoại Giao
- * Điểm chuẩn thấp nhất nằm ở Đại Học Dân Lập Duy Tân, Đại Học Kiên Giang và Đại Học Quảng Nam với 14 điểm.

– Biểu đồ Số đơn vị đào tạo

- * Có tổng 150 đơn vị đào tạo với số lượng đại học công lập chiếm phần lớn

• ĐGNLHCM

– Số liệu thống kê tổng quan:

- * Điểm chuẩn trung bình: 666.27
- * Điểm chuẩn cao nhất: 1035
- * Điểm chuẩn thấp nhất: 500

– Biểu đồ top 5 trường có điểm chuẩn cao nhất và thấp nhất:

- * Điểm chuẩn cao nhất thuộc về trường Đại Học Khoa Học Tự Nhiên - Đại Học Quốc Gia TPHCM với 29.42 điểm, tiếp đến là các trường Đại Học Bách Khoa - Đại Học

Đà Nẵng, UIT - Đại Học Công Nghệ Thông Tin - Đại Học Quốc Gia TPHCM, Đại Học Ngoại Ngữ - Đại Học Quốc Gia Hà Nội, Khoa Y - Đại Học Quốc Gia TPHCM

* .Đại học Bình Dương, Đại Học Hùng Vương - TPHCM, Đại Học Quang Trung có điểm chuẩn thấp nhất với 500.00 điểm

– Biểu đồ Số đơn vị đào tạo

* Có tổng 43 đơn vị đào tạo với số lượng đại học công lập chiếm phần lớn

- Ngoài ra còn có nhiều bộ lọc để giúp người dùng tiến hành cá nhân hóa thông tin cần thiết

2.5 Phân tích Dashboard "Gợi ý chọn trường"

2.5.1 Mục tiêu dashboard

Mục tiêu chung của dashboard:

- Cung cấp thông tin tổng quan về điểm chuẩn của các trường đại học tại Việt Nam
- Thời gian: năm 2023
- Phạm vi: 63 tỉnh thành trên cả nước
- Đối tượng: Các người dùng đặc biệt là học sinh và phụ huynh muốn tìm kiếm và lựa chọn ngành học phù hợp với khoảng điểm mà học sinh ước tính
- Cung cấp khả năng tìm kiếm các ngành học và trường đại học phù hợp với khoảng điểm xét tuyển của người dùng đã .
- Cho phép người dùng lọc dữ liệu theo các tiêu chí như trường, khu vực, phương thức xét tuyển, và tổ hợp môn thi.
- Hiển thị các thông tin thống kê tổng quan về số phân ngành, số ngành phù hợp, điểm chuẩn cao nhất và điểm chuẩn trung bình.
- Trực quan hóa dữ liệu bằng biểu đồ cây (tree map), giúp người dùng dễ dàng nhận diện và so sánh các ngành học dựa trên phân ngành và tên ngành.
- Cung cấp chi tiết về các ngành học cụ thể khi người dùng tương tác với biểu đồ cột.

2.5.2 Các thành phần trực quan của dashboard

Dashboard "Gợi ý chọn trường phù hợp" bao gồm các thành phần chính sau đây:

Số liệu thống kê tổng quan

Loại trực quan: Số liệu

Phần số liệu thống kê tổng quan cung cấp các số liệu (theo khoảng điểm và filter của người dùng) chính xác và cụ thể về:

- Số phân ngành
- Số ngành phù hợp
- Điểm chuẩn cao nhất
- Điểm chuẩn trung bình

Biểu đồ cây phân rã

Loại trực quan: Biểu đồ cây phân rã

Hiển thị số ngành, các phân ngành và tên ngành cùng số lượng ngành con tương ứng.

Người dùng có thể nhấp vào các thanh ngang trong tree map để xem chi tiết các ngành học cụ thể phù hợp với khoảng điểm xét tuyển đã nhập.

Các thành phần khác

- Nút xóa tất cả các bộ lọc: Cho phép người dùng làm mới kết quả tìm kiếm bằng cách xóa tất cả bộ lọc đã chọn trước đó
- Thông tin chú thích: Có chú thích ở phía dưới biểu đồ cây để người dùng có thể dễ dàng hiểu và sử dụng biểu đồ này

3 Đánh giá Dashboard

3.1 Nguồn dữ liệu đáng tin cậy

Dữ liệu được sử dụng trong Dashboard được nhóm crawl từ các trang báo điện tử và các trang web uy tín: Tiền Phong, VietTimes, Học mãi, Thư viện pháp luật, Bộ Giáo dục và Đào tạo. Sau đó, dữ

liệu được tiền xử lý (như đã nêu ở phần xử lý dữ liệu) để phù hợp mục đích sử dụng.

3.2 Phù hợp với mục đích

Dashboard sử dụng các thành phần trực quan phù hợp với mục đích đã đặt ra.

Đối với dashboard "Điểm thi theo môn" và "Điểm thi theo tổ hợp": Ô số liệu giúp cung cấp nhanh những thông tin quan trọng cho người dùng. Biểu đồ histogram là loại biểu đồ phù hợp để thể hiện phổ điểm thi. Biểu đồ cột ngang phù hợp cho mục đích so sánh, được sử dụng để so sánh và thể hiện thứ hạng về số lượng thí sinh và điểm trung bình giữa các tỉnh thành. Bản đồ bong bóng phù hợp để so sánh tương quan và mật độ phân bố của các điểm, dùng để thể hiện sự tập trung số lượng thí sinh của các tỉnh thành.

Đối với dashboard "Điểm chuẩn": Ô số liệu giúp cung cấp nhanh những thông tin quan trọng cho người dùng. Biểu đồ cột đứng dùng để so sánh và thể hiện thứ hạng về điểm chuẩn của các trường đại học. Bản đồ thể hiện vị trí địa lý của các trường đại học, giúp cung cấp nhanh thông tin về khu vực địa lý của trường đó. Biểu đồ Donut thể hiện tỉ lệ giữa số trường công lập và dân lập kèm số liệu, giúp người dùng nắm bắt được tỉ lệ và số lượng cơ sở đào tạo mà họ quan tâm.

Đối với dashboard "Gợi ý chọn trường phù hợp": Ô số liệu cung cấp nhanh những thông tin quan trọng cho người dùng. Biểu đồ cây phân rã (decomposition tree) phù hợp để biểu diễn dữ liệu phân cấp và giúp ích trong việc phân tích đưa ra quyết định (người dùng có thể thay đổi thứ tự của các thuộc tính quyết định chính là thứ tự của các node cha/con). Các thông tin không được quan tâm có thể được đóng gói lại trong một node và có thể được chọn để theo dõi.

3.3 Rõ ràng và dễ hiểu

Dashboard hướng đến đối tượng người sử dụng là học sinh, phụ huynh, giáo viên và những người quan tâm đến kỳ thi tuyển sinh Đại học, do đó tiêu chí rõ ràng và dễ hiểu là một trong những tiêu chí quan trọng. Dashboard được thiết kế sao cho:

3.3.1 Biểu đồ dễ hiểu

Thiết kế của Dashboard hướng đến việc lựa chọn những công cụ trực quan dễ hiểu cùng với cách thể hiện thông tin rõ ràng: Chủ yếu sử dụng biểu đồ cột quen thuộc, bản đồ giúp biểu thị thông

tin tỉnh thành và trường đại học dễ nắm bắt hơn,... Biểu đồ cây phân rã khá mới lạ nhưng cũng không khó sử dụng và có hướng dẫn kèm theo.

3.3.2 Đối tượng trực quan rõ ràng

Để giúp người sử dụng có thể hiểu được các biểu đồ thành phần trong Dashboard một cách hiệu quả thì thông tin về đối tượng được trực quan trong mỗi biểu đồ cần được thể hiện chính xác và dễ nắm bắt. Trong thiết kế Dashboard này, tiêu đề của mỗi biểu đồ thành phần được đặt ngắn gọn, rõ ràng và mang thông tin chính, vừa đủ về đối tượng được trực quan.

3.4 Tích hợp và liên kết

3.4.1 Liên kết về màu sắc

Các biểu đồ thành phần trong Dashboard được liên kết và nhất quán xuyên suốt thông qua màu sắc của đối tượng trực quan chính mà các biểu đồ đang thể hiện (điểm chuẩn, điểm thi): Tông màu chủ đạo của các biểu đồ là màu cam ấm.

3.4.2 Liên kết về chức năng

Các biểu đồ thành phần trong Dashboard được liên kết theo từng nhóm về mặt chức năng. Các biểu đồ cùng chức năng được đặt cạnh nhau: nhóm biểu đồ so sánh tình hình điểm chung của các tỉnh thành, nhóm biểu đồ về thứ hạng các mức điểm đặc biệt (điểm liệt, điểm trung bình, điểm cao), nhóm biểu đồ top 5 trường có điểm chuẩn cao/thấp nhất. Các số liệu liên quan cũng được đặt cạnh nhau và đặt gần các biểu đồ tương ứng với nó. Các filter được đặt gần nhau thành một khu vực riêng biệt.

3.4.3 Đồng bộ

Thao tác trên các biểu đồ thành phần hoạt động đồng bộ và ảnh hưởng đến các biểu đồ còn lại. Các biểu đồ có thể được sử dụng như một bộ lọc (môn thi, khối thi, khu vực, tỉnh) để lọc trong toàn bộ Dashboard. Dashboard cho phép highlight liên biểu đồ đối với nhóm biểu đồ so sánh để người sử dụng có thể quan sát số liệu trong cùng loại giá trị (điểm chuẩn, điểm thi) dễ dàng hơn.

3.5 Tương tác và điều hướng

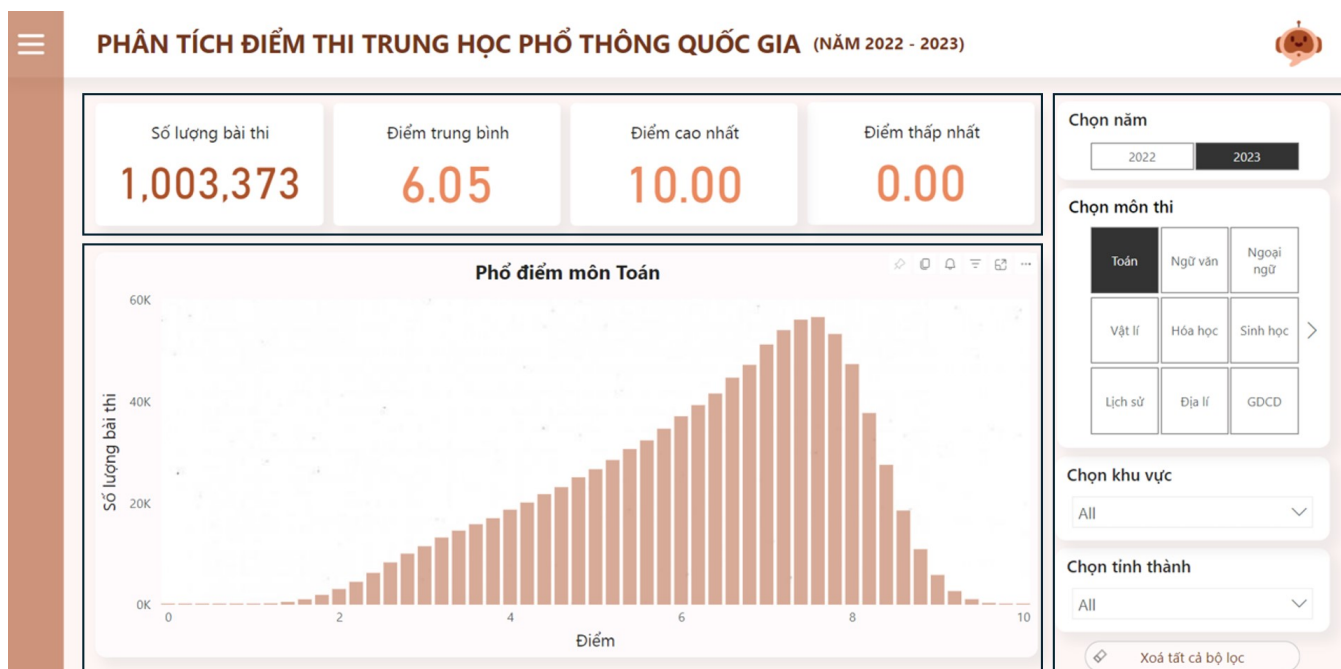
3.5.1 Tương tác

Các công cụ tương tác trong Dashboard được lựa chọn có chủ đích, phù hợp mức độ quan trọng của đối tượng. Số lượng các loại công cụ tương tác được tối giản hóa cần thiết, tránh sự quá tải cho người sử dụng, nhưng vẫn đảm bảo sự đa dạng.

- Bộ lọc dạng nút bấm dễ nhận diện và thao tác, dùng cho những biến thường xuyên được thay đổi (năm, môn thi/tổ hợp), phương thức xét tuyển, nhóm ngành (trong dashboard "Gợi ý chọn trường").
- Bộ lọc dạng thả xuống giúp tiết kiệm không gian, dùng cho những thông tin lọc ít được thay đổi hơn (khu vực, tỉnh thành,...).
- Bộ lọc nhanh (mức độ nhận diện không cao nhưng tính tiện lợi cao) hỗ trợ lọc nhanh và trực tiếp khi nháy chuột (click) trên biểu đồ.
- Tooltip cho phép xem chú thích khi di chuột (hover) trên biểu đồ.

Điều hướng bố cục chung

Bố cục của Dashboard được định hình bằng việc sử dụng nguyên lý proximity.



Hình 13: Sử dụng nguyên lý proximity để tạo bố cục cho dashboard

Điều hướng yếu tố quan trọng Khi trực quan, có những đối tượng có vai trò quan trọng trong việc giúp người dùng nắm bắt thông tin quan trọng hoặc nhận diện được những công cụ tương tác chính. Vì vậy Dashboard cũng chú trọng thiết kế các yếu tố điều hướng phù hợp.

3.6 Thiết kế hấp dẫn

3.6.1 Cân đối

Dashboard sử dụng bố cục đối xứng tạo cảm giác cân đối, dễ chịu.

3.6.2 Rõ ràng

Dashboard sử dụng nguyên lý enclosure để đóng khung từng khu vực, giúp bố cục rõ ràng, có sự phân biệt giữa các biểu đồ thành phần.

3.6.3 Đơn giản

- Số lượng các bộ lọc phù hợp, các bộ lọc không cần thiết được giản lược.
- Số lượng màu sắc vừa đủ, tránh sự quá tải cho người dùng.

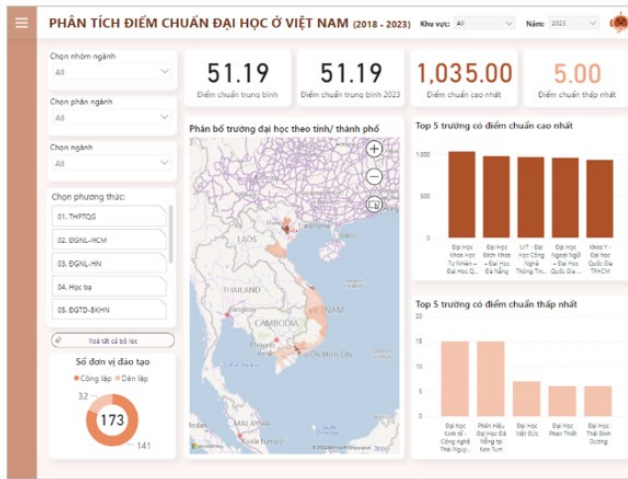
3.6.4 hài hòa

- Màu sắc được sử dụng thống nhất trong toàn bộ Dashboard với tông màu chủ đạo là cam.
- Màu nền của các thành phần trực quan là màu trắng trên màu nền chung cam nhạt. Màu nền được chọn cùng sắc với tông màu chủ đạo đảm bảo sự hài hòa mà vẫn làm nổi bật được các thành phần trực quan.

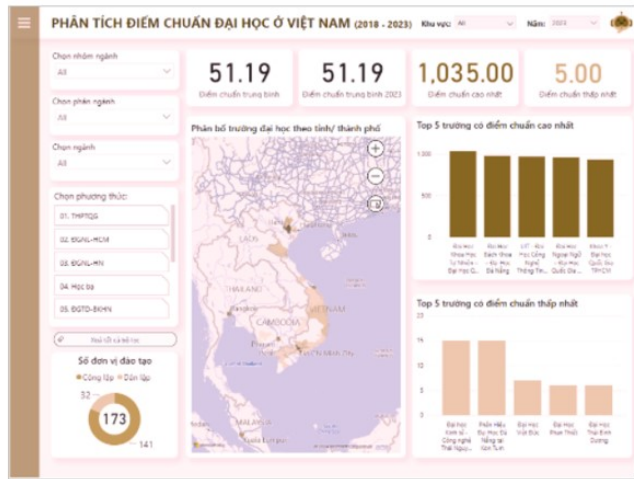
3.6.5 Thân thiện với người mù màu

Màu sắc được sử dụng trong dashboard có thể phân biệt được với các nhóm người mù màu khác nhau mà không cần chuyển giữa các chế độ:

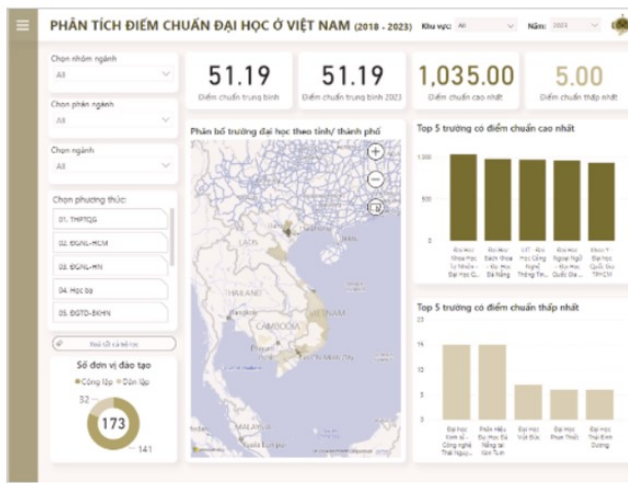
Màu gốc



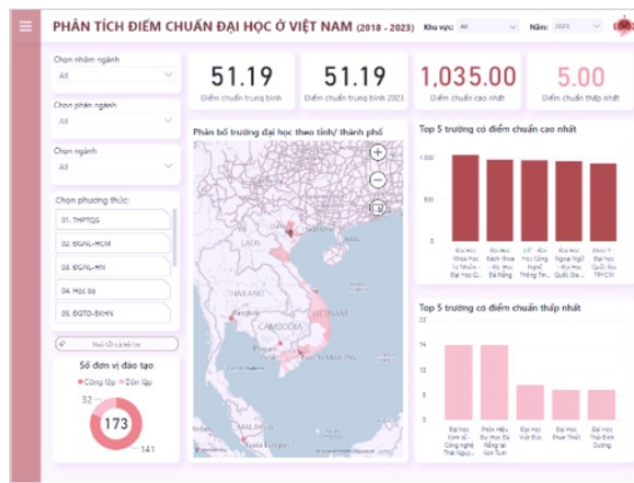
Màu màu xanh lá



Màu màu đỏ



Màu màu xanh dương



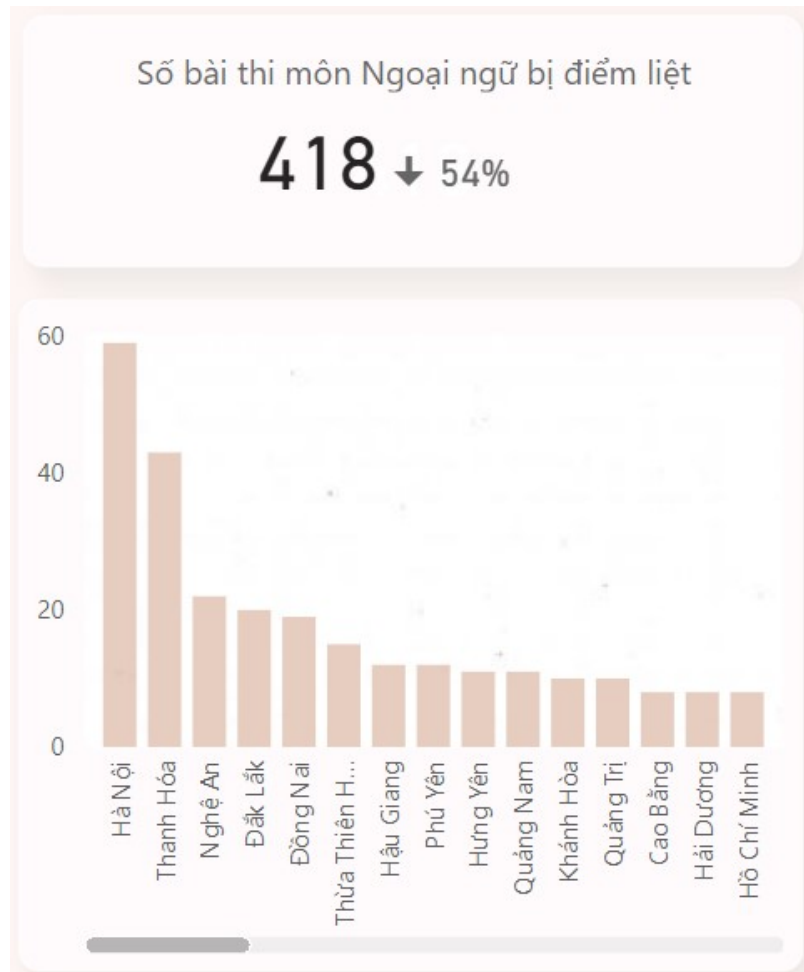
Hình 14: Màu sắc trong dashboard có tính phân biệt cao với các nhóm người mù màu

3.7 Phân tích dữ liệu

Xu hướng

Mục tiêu và đối tượng hướng đến của dashboard không chú trọng việc thể hiện xu hướng thay đổi theo thời gian mà tập trung vào nhiệm vụ cung cấp cái nhìn tổng quan về tình hình điểm thi của các môn thi/tổ hợp và các tỉnh thành; tổng quan về điểm chuẩn của các trường đại học; gợi ý các ngành và trường hợp theo điểm. Các nhiệm vụ này tập trung làm rõ thông tin trong một năm ở nhiều khía cạnh khác nhau hơn là thể hiện sự thay đổi theo thời gian (phù hợp hơn với mục tiêu trợ giúp đưa ra quyết định hơn là nghiên cứu về tình hình dài hạn).

Tuy nhiên, dashboard cũng có những thành phần trực quan giúp thể hiện xu hướng. Các ô số liệu trong dashboard không chỉ cung cấp thông tin về số lượng mà còn có tỉ lệ phần trăm tăng, giảm so với năm trước đó, từ đó giúp người dùng nắm bắt được sự thay đổi của dữ liệu qua thời gian:



Hình 15: Số lượng thí sinh bị điểm liệt môn Ngoại ngữ năm 2023

Những kết luận và câu chuyện liên quan

Các dashboard cung cấp được nhiều thông tin có ý nghĩa và gắn liền với những vấn đề thực tế và phản ánh thực trạng giáo dục tại Việt Nam. Các liên hệ này đã được trình bày trong mục "Liên hệ" của phần 2:

- Thực trạng giáo dục ở các tỉnh thuộc khu vực Trung du và miền núi phía Bắc [2.2.4](#)
- Thực trạng học Ngoại ngữ chung ở Việt Nam [2.2.4](#)
- Năng lực Ngoại ngữ của học sinh TP. Hồ Chí Minh và Bình Dương [2.2.4](#)

Dashboard "Điểm chuẩn" đã cung cấp cho chúng ta thêm nhiều thông tin về các phương thức xét tuyển được thể hiện rõ nhất ở năm 2021 và 2022 khi dịch bệnh Covid-19 ảnh hưởng nhiều đến việc tuyển sinh của các trường đại học:

- Số lượng trường xét học bạ tăng đột biến năm 2021 [2.4.3.4](#)
- Số phương thức xét tuyển tăng và phương thức xét tuyển bằng điểm DGNLHCM trở nên phổ biến hơn [2.4.3.5](#)

3.8 Tích hợp AI

Dashboard tích hợp công cụ AI để hỗ trợ tìm kiếm và phân tích nâng cao; cá nhân hóa trải nghiệm cho người dùng dashboard.

Trong quá trình xây dựng dashboard, nhóm đã thử nghiệm 3 loại công cụ AI chính:

API của ChatGPT được hỗ trợ bởi bên thứ 3

Power BI Apps cho phép sử dụng dịch vụ của các bên thứ ba để tích hợp trực tiếp các mô hình ngôn ngữ lớn sẵn có bằng API, ví dụ như [AI Lens](#), [chatPowerBI](#). Ưu điểm của các công cụ này là khả năng của các mô hình ngôn ngữ lớn phổ biến hiện nay: hiểu ngôn ngữ tự nhiên rất tốt kể cả tiếng Việt, trả lời được những câu hỏi phức tạp. Việc sử dụng dịch vụ của bên thứ ba giúp giảm nhiều công sức trong việc xây dựng giao diện cho chatbot trong Power BI (chỉ cần cung cấp API key của OpenAI hoặc License key của bên thứ ba).

Tuy nhiên, loại công cụ này không phù hợp với mục tiêu nhóm hướng tới là giúp người dùng hiểu dữ liệu vì nó không hiểu tốt dữ liệu của dashboard để trả lời các thông tin liên quan một cách chính xác, thường trả lời sai và đôi khi không hiểu câu hỏi.

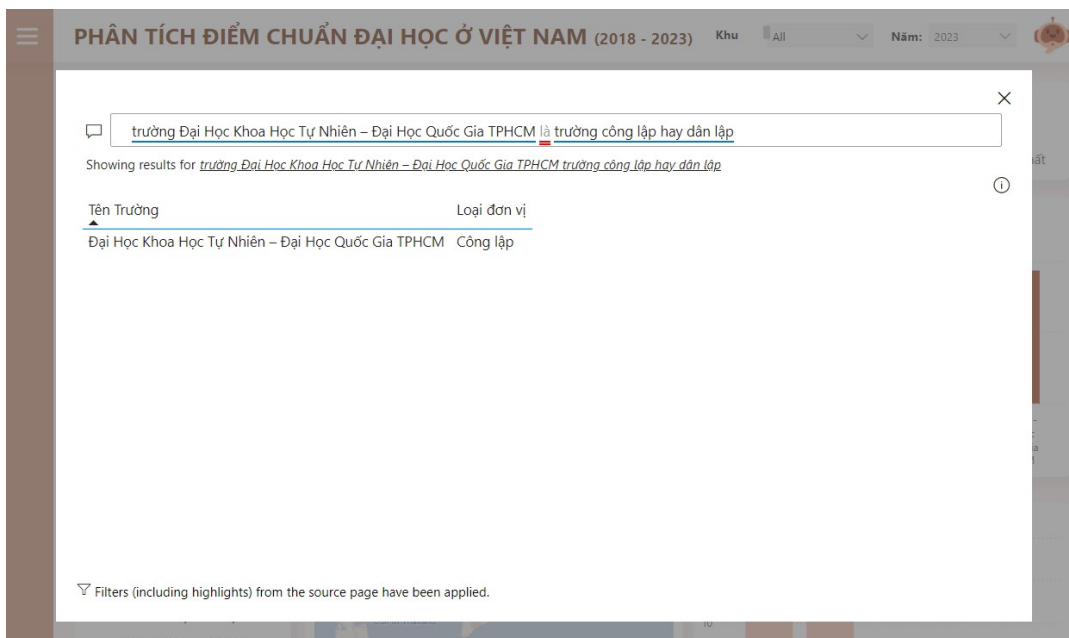


Hình 16: Tích hợp ChatGPT vào Power BI với công cụ AI Lens do Len Visuals cung cấp

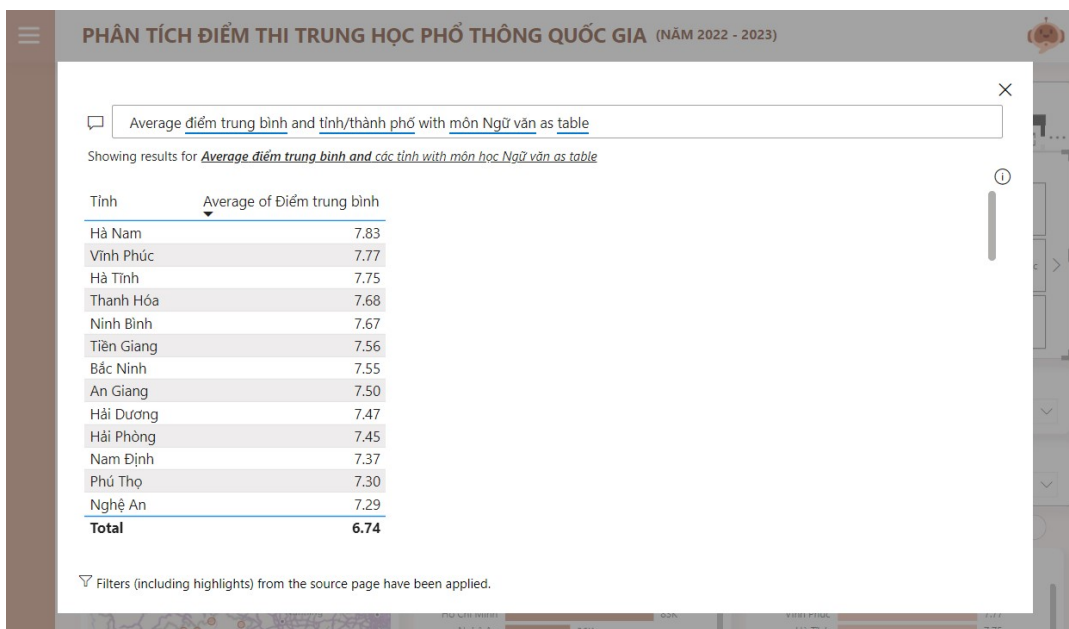
Công cụ Q&A của Power BI

Power BI cung cấp công cụ Q&A hỗ trợ người dùng sử dụng ngôn ngữ tự nhiên để truy vấn dữ liệu của dashboard. Người dùng có thể hỏi về số liệu hoặc yêu cầu vẽ các biểu đồ bằng ngôn ngữ tự nhiên (các hình bên dưới). Tuy nhiên, Q&A có nhược điểm là không hỗ trợ đa ngôn ngữ (hiện tại chỉ hỗ trợ tiếng Anh và bản preview hỗ trợ tiếng Tây Ban Nha); đồng thời nó cũng không thể trả lời được những câu hỏi quá phức tạp và cần nhiều suy luận hoặc các kiến thức bên ngoài như những mô hình ngôn ngữ lớn.

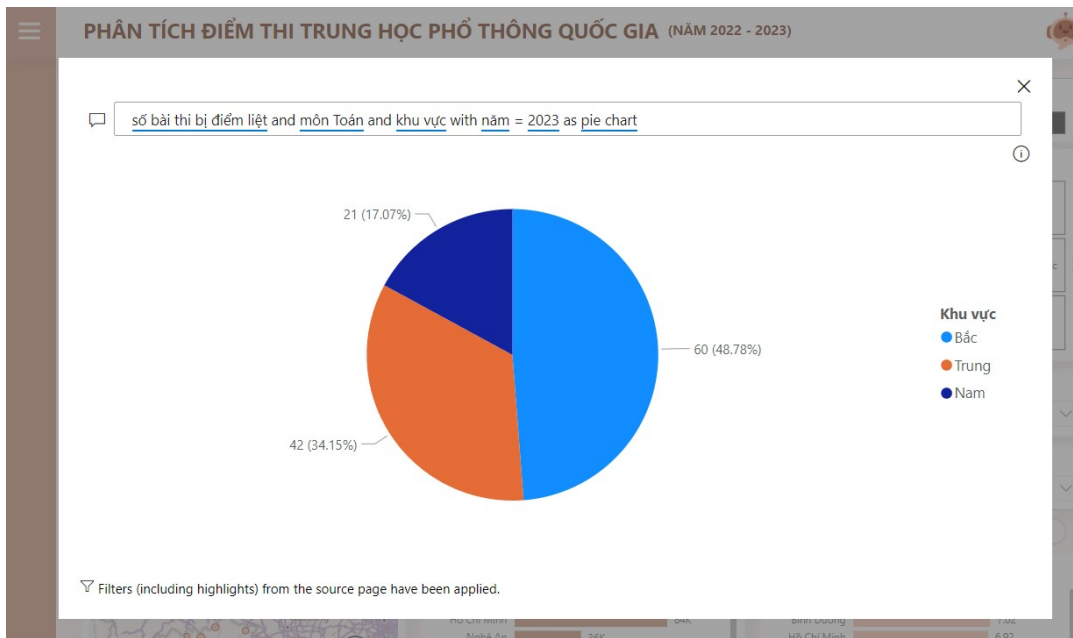
Để có thể sử dụng được cho tiếng Việt thì quá trình tinh chỉnh gặp nhiều khó khăn vì phải thiết lập các luật một cách thủ công và không hiệu quả (từ đồng nghĩa, mối quan hệ ngữ nghĩa giữa các bảng,...). Công cụ tương đối khó sử dụng để hỏi và cần có kinh nghiệm về cách prompting để mô hình có thể hiểu và trả lời.



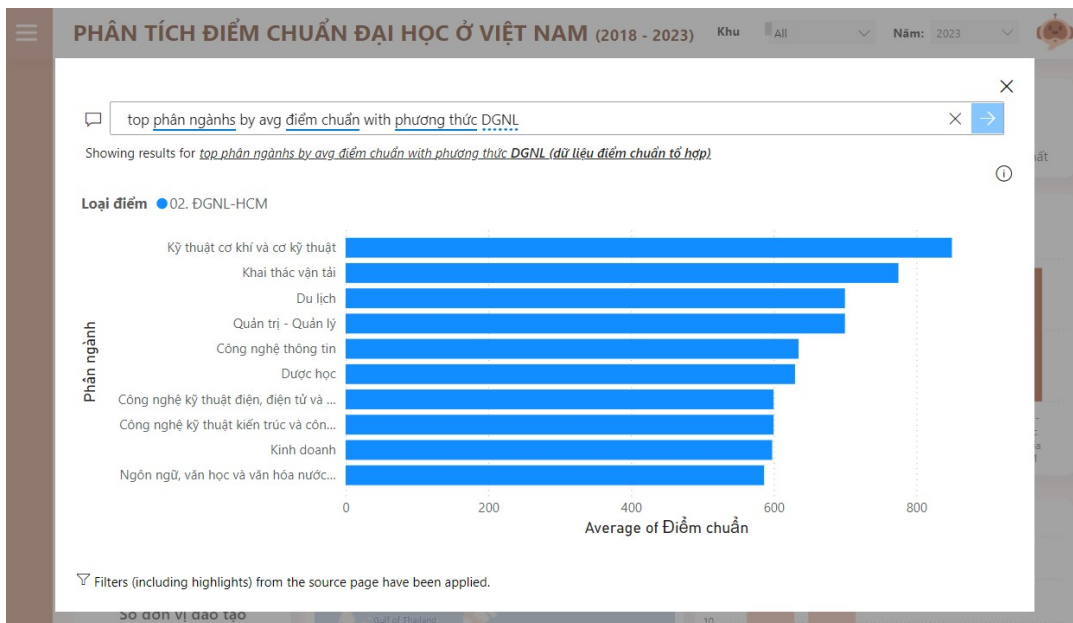
Hình 17: Sử dụng Q&A của Power BI để hỏi về thông tin đã có trong cơ sở dữ liệu



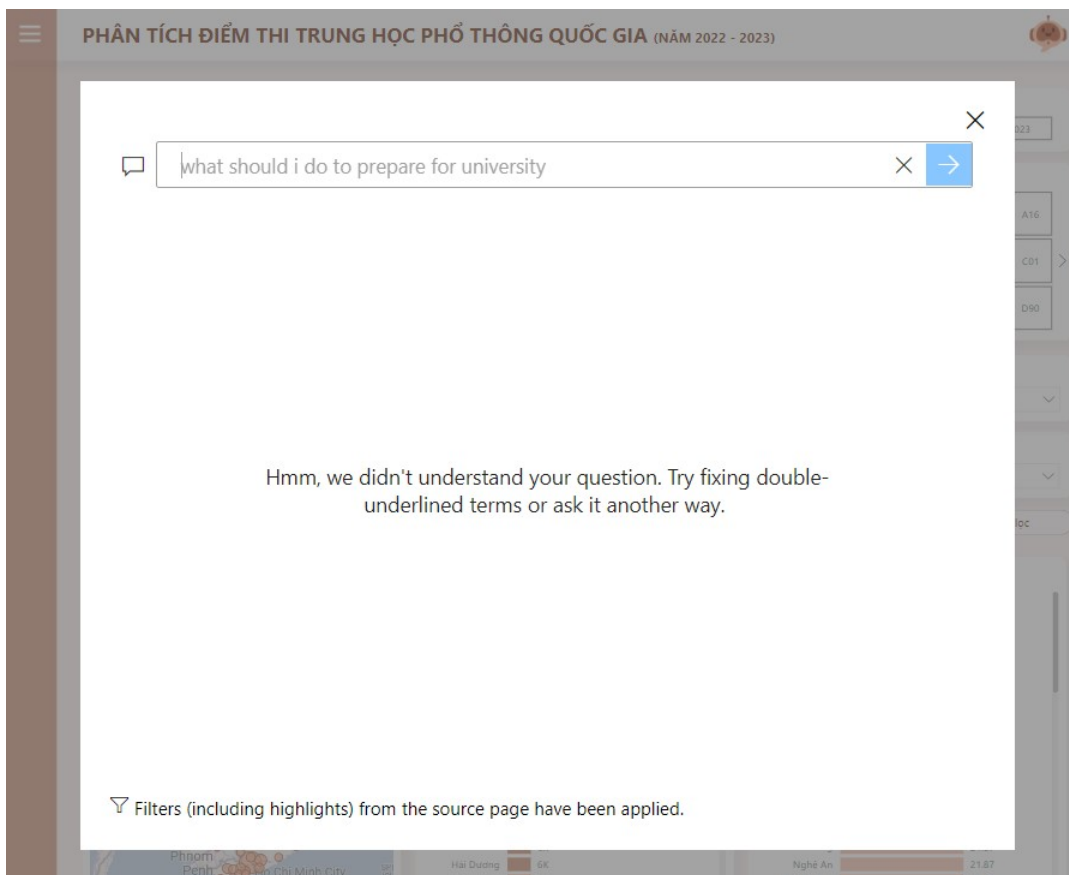
Hình 18: Sử dụng Q&A của Power BI để truy vấn dữ liệu



Hình 19: Sử dụng Q&A của Power BI để vẽ biểu đồ về dữ liệu điểm thi



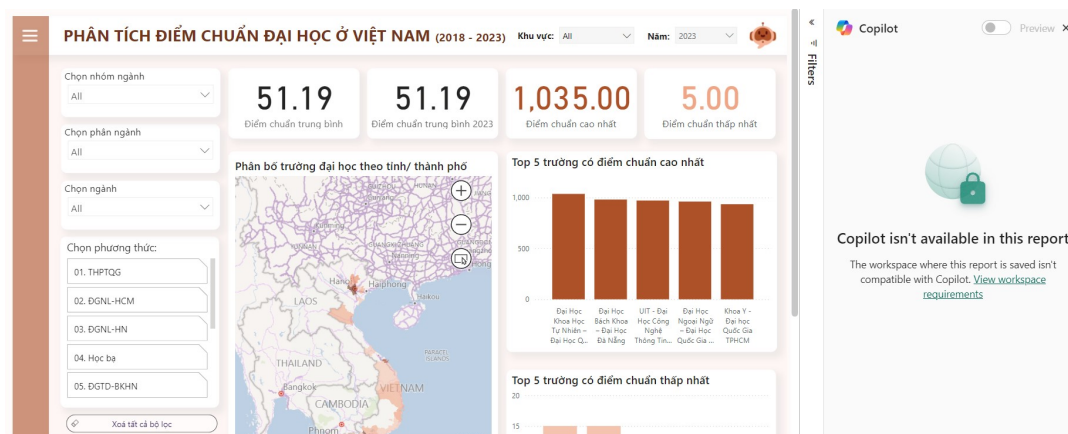
Hình 20: Sử dụng Q&A của Power BI để vẽ biểu đồ về dữ liệu điểm chuẩn



Hình 21: Sử dụng Q&A của Power BI để hỏi về thông tin không có trong cơ sở dữ liệu

Công cụ Q&A và Copilot tích hợp sẵn trong Power BI

Đây là lựa chọn có phần hiệu quả và phù hợp nhất vì nó giải quyết được các nhược điểm của hai công cụ bên trên: hiểu được dữ liệu riêng tư của dashboard (Len Visuals chưa hỗ trợ); hỗ trợ đa ngôn ngữ, có thể trả lời được những câu hỏi phức tạp và liên quan đến kiến thức bên ngoài (Q&A chưa hỗ trợ); cập nhật thông tin mới trên mạng Internet (cả Q&A và Len Visuals đều chưa hỗ trợ). Tuy nhiên, tính năng này là tính năng preview của Power BI và đồng thời yêu cầu có tài khoản Premium để có thể sử dụng. Vì điều kiện tài chính giới hạn nên trong dashboard này nhóm chỉ sử dụng công cụ Q&A cơ bản của Power BI như đã trình bày ở trên.



Hình 22: Power BI hỗ trợ tích hợp Copilot nhưng cần có tài khoản Premium

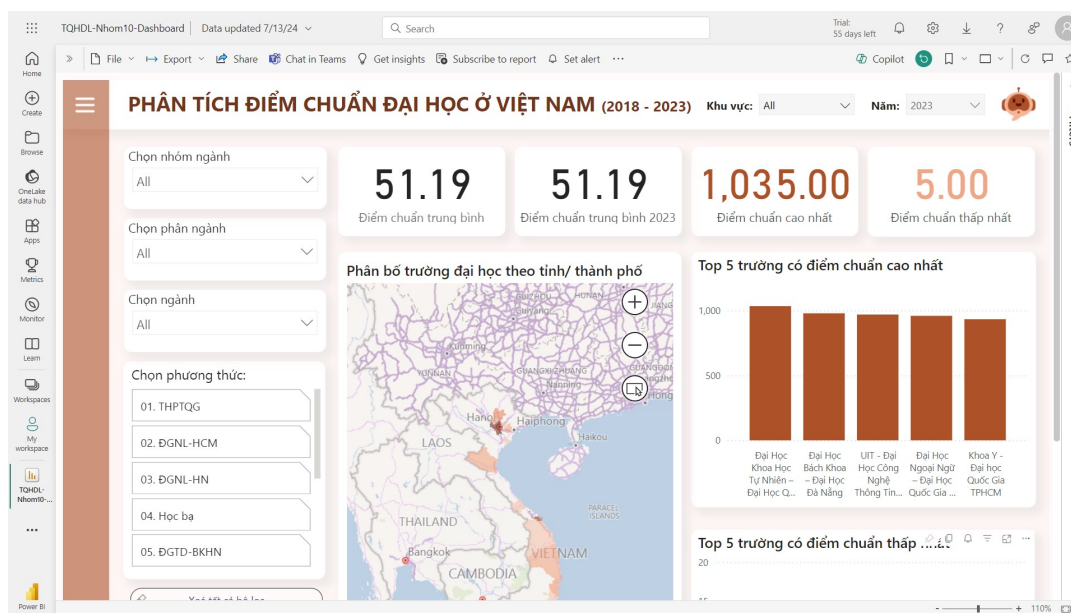
3.9 Các tiêu chí đánh giá khác

Trong phần này, nhóm sẽ đánh giá nhanh về các tiêu chí mở rộng dù không được yêu cầu trong đồ án cuối kỳ nhưng vẫn là những yếu tố quan trọng trong quá trình xây dựng dashboard.

Tích hợp và chia sẻ

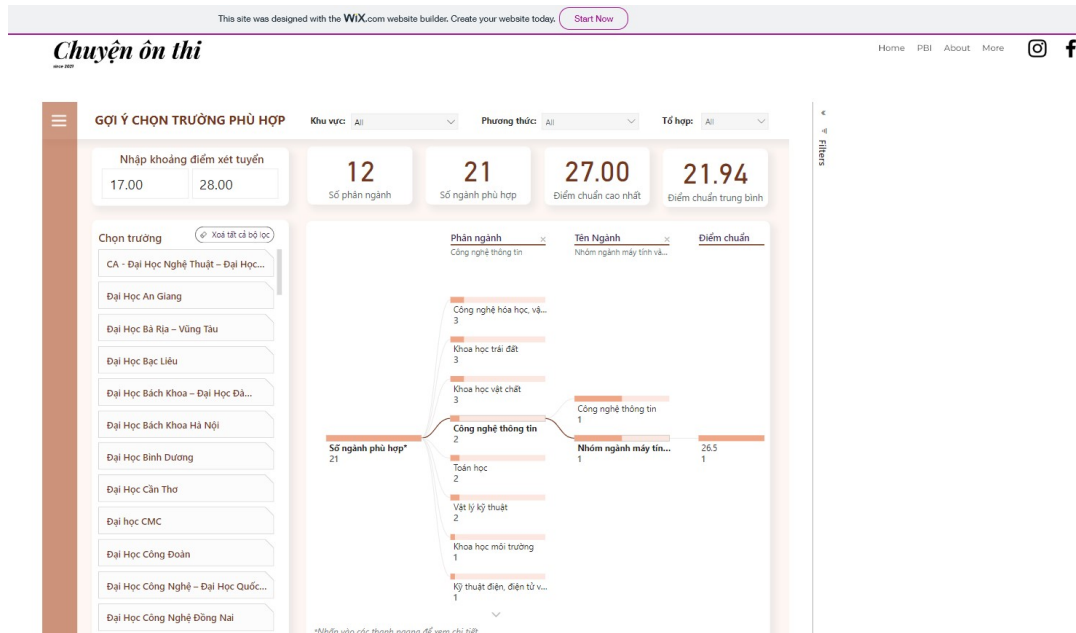
Nhóm đã thực hiện tích hợp/chia sẻ dashboard trên các nền tảng khác nhau:

- Publish dashboard trên Power BI Service: Chia sẻ với người trong tổ chức trên nền tảng Web (cần đăng nhập nên rất hạn chế về khả năng chia sẻ)



Hình 23: Publish Dashboard trên Power BI service (xem tại đây)

- Nhúng vào bản thuyết trình Powerpoint
- Nhúng vào website cá nhân: Có yêu cầu đăng nhập tương tự như trong Power BI Service



Hình 24: Nhúng và chia sẻ dashboard trên website ([xem tại đây](#))

Có thể thấy so với Tableau thì Power BI có phần hạn chế chia sẻ hơn.

Hiệu suất

Kích thước dữ liệu được sử dụng trong dashboard này lớn hơn nhiều so với dashboard "Phân tích tình hình dịch bệnh Covid-19" trong đồ án giữa kỳ. Hiệu suất của Power BI tốt hơn Tableau trên tập dữ liệu có kích thước lớn, cụ thể là dữ liệu và biểu đồ được tải gần như ngay lập tức khi thay đổi giá trị lọc. Ngoài ra, số lượng dashboard thành phần khá nhiều nên có một chút độ trễ khi thay đổi giữa các dashboard nhưng không đáng kể.

4 Tổng hợp đánh giá

Bảng 9: Đánh giá mức độ hoàn thành đồ án

| Nội dung | Đánh giá |
|-------------------------|-------------|
| Nguồn dữ liệu tin cậy | Đạt yêu cầu |
| Phù hợp với mục đích | Đạt yêu cầu |
| Rõ ràng và dễ hiểu | Đạt yêu cầu |
| Đồng bộ và liên kết | Đạt yêu cầu |
| Thể hiện được xu hướng | Đạt yêu cầu |
| Tương tác và điều hướng | Đạt yêu cầu |
| Thiết kế hấp dẫn | Đạt yêu cầu |
| Phản hồi ý kiến | Đạt yêu cầu |
| Tích hợp và chia sẻ | Đạt yêu cầu |
| Hiệu suất | Đạt yêu cầu |

Tài liệu tham khảo

- [1] *Danh sách các trường đại học kèm mã trường được phân chia theo từng khu vực chi tiết nhất 2024.* **Thư viện pháp luật**. Truy cập lần cuối: 12/07/2024.
- [2] *Thông tư số: 09/2022/TT-BGDDT về việc quy định mã ngành, phân ngành đào tạo giáo dục đại học, cao đẳng.* **Bộ Giáo dục và Đào tạo**. Truy cập lần cuối: 12/07/2024.
- [3] *Hà Giang bàn kế sách 'thoát đáy' giáo dục.* **Báo Tiền Phong**. Truy cập lần cuối: 12/07/2024.
- [4] *Nâng cao chất lượng dạy và học ngoại ngữ tại địa phương.* **Tạp chí điện tử Người làm báo**. Truy cập lần cuối: 12/07/2024.
- [5] *10 tỉnh, thành phố có thu nhập bình quân đầu người cao nhất và thấp nhất cả nước.* **Báo Quảng Ninh**. Truy cập lần cuối: 12/07/2024.
- [6] *Tra cứu điểm chuẩn các trường đại học.* **Hệ thống giáo dục HOCMAI**. Truy cập lần cuối: 12/07/2024.
- [7] *Vì sao TP.HCM luôn có kết quả thi tiếng Anh cao nhất?* **Báo Phụ Nữ**. Truy cập lần cuối: 12/07/2024.
- [8] *Vùng Trung du và miền núi Bắc Bộ đang đối mặt nhiều thứ 'thiếu' trong giáo dục.* **Báo Giáo dục**. Truy cập lần cuối: 12/07/2024.

Phụ lục

- Dashboard trên Power BI: **Dashboard Điểm thi, điểm chuẩn - Power BI Service**
- Website tích hợp Power BI dashboard: **Dashboard Điểm thi, điểm chuẩn - Website Nhóm 10**
- Slide thuyết trình: **Dashboard Điểm thi, điểm chuẩn - Slide đồ án cuối kì**