**SCALA CASE STUDY**  
Công ty: Voizfm

​​​​

**Voizfm** là nền tảng nghe Sách nói và Podcast chất lượng cao, 100% Bản quyền lớn nhất Việt Nam.

Voizfm lưu trữ các nguồn dữ liệu sau:

1. **Firebase Event Data**: dữ liệu hành vi của users dưới dạng events, thông qua [Firebase](https://firebase.google.com/) ([Firebase Event Analytics](https://towardsdatascience.com/firebase-event-analytics-with-google-bigquery-ec756230f768)). Các events này bao gồm: install, register, trial, purchase/renew, hành vi nghe, các nội dung và người dùng tương tác
2. **Dữ liệu Demography:** Có những những người dùng opt-in khai báo (khoảng 30%)
3. **Dữ liệu về kinh doanh:** Server nội bộ, App Stores & Ads Manager. Hiện nay quá trình tổng hợp dữ liệu vẫn đang thực hiện thủ công, thông qua spreadsheet
4. **Metadata về các contents**: đang trong quá trình chuẩn hoá và hệ thống lại

Các dashboard chính:

* **Business KPI Dashboard**: để cập nhật tình hình kinh doanh, doanh số
* **Internal Dashboard:** thống kê một số chỉ số như số người dùng đăng ký mới, VIP, người dùng cancel
* **Product Dashboard** (từ Firebase event): các chỉ số liên quan đến in-app events, như lượt nghe, thời lượng nghe, crash, app remove
* **User Clusters Dashboard:** chia người dùng thành các nhóm dựa trên quá khứ giao dịch, nhiều dữ liệu về hành vi (in-app events) chưa được tận dụng như hành vi nghe, hệ diều hành

**Mục Đích**

Voizfm mong muốn phát triển thêm User Clusters Dashboard hiện tại để có các hiểu biết sâu sắc hơn về người dùng, phục vụ cho các chiến lược kinh doanh và phát triển sản phẩm:

* Khai thác và tận dụng thêm các thông tin về hành vi trong in-app events trong việc phân loại người dùng
* Tìm ra mối liên kết và tương quan giữa nhiều yếu tố khác nhau trong thông tin và hành vi của người dùng
* Từ mô tả, khám phá, tạo nền tảng cho các mô hình tiên toán: LTV, Churn, content recommend

**Yêu Cầu**

Tiếp cận và giải quyết bài toán theo hướng [hypothesis-based](https://www.consultantsmind.com/2017/07/10/hypothesis-based-consulting-2/)

1. Tìm hiểu và thảo luận để hình thành các giả thuyết về hành vi người dùng
2. Khai thác và phân tích mẫu in-app events data
3. Báo cáo các phát hiện và kết luận theo giả thuyết của (1)
4. Xây dựng các mô hình giải thích, tiên đoán, clustering
5. Dựa trên kết quả bước (3) + (4), thiết kế, nâng cấp User Clusters Dashboard hiện có
6. Bàn giao và đề xuất các bước tiếp theo

**Research**

**Yêu Cầu**

**Tiếp cận và giải quyết bài toán theo hướng** [**hypothesis-based**](https://www.consultantsmind.com/2017/07/10/hypothesis-based-consulting-2/)

1. Tìm hiểu và thảo luận để hình thành các giả thuyết về hành vi người dùng
2. Khai thác và phân tích mẫu in-app events data
3. Báo cáo các phát hiện và kết luận theo giả thuyết của (1)
4. Xây dựng các mô hình giải thích, tiên đoán, clustering
5. Dựa trên kết quả bước (3) + (4), thiết kế, nâng cấp User Clusters Dashboard hiện có
6. Bàn giao và đề xuất các bước tiếp theo

**1. Tìm hiểu và thảo luận để hình thành các giả thuyết về hành vi người dùng**:

Cần:

* Đọc 1 số bài toán về hành vi người dùng
* Hỏi doanh nghiệp về 1 số giả thuyết mà doanh nghiệp đã có

**2. Khai thác và phân tích mẫu in-app events data**

Cần:

* Xin data mẫu của doanh nghiệp (lưu ý: chỉ lấy mẫu in-app events và nếu có thể thì xin 1 file template mẫu của data)

**3. Báo cáo các phát hiện và kết luận theo giả thuyết của (1)**

Cần:

* Phân tích dữ liệu, đưa ra những minh chứng cụ thể cho những giả thuyết ở trên

1. **Xây dựng mô hình (clusteing)**

* RFM (Recency, Frequency, Monetary)
* RFD (Recency, Frequency, Duration) – Duration here is time spent. Particularly useful while analyzing consumer behaviour of viewership/readership/surfing oriented products.
* RFE (Recency, Frequency, Engagement) – Engagement can be a composite value based on time spent on page, pages per visit, bounce rate, social media engagement etc. Particularly useful for online businesses.

1. **Nâng cấp user clusters Dashboard hiện có**

Cần:

* Xem qua user cluster dashboard hiện có của doanh nghiệp
* Thu thập những thiếu sót của user cluster dashboard hiện có
* Khoanh vùng phạm vi giải quyết, đưa ra phương án cải tiến
* Chốt phương án và thực hiện

1. **Bàn giao và đề xuất phương án tiếp theo**

**Bàn giao:** User clustering Dashboard đã cải tiến có tích hợp mô hình clustering đã xây dựng ở bước 4

**Phương án đề xuất:**

* Collect data (xây dựng data warehouse)
* Nâng cấp các dashboard hiện có:

Tích hợp lại tất cả các dashboards hiện được xây dựng rời rạc (user clustering, Product dashboard, Product dashboard, Business KPI) vào trong 1 hệ thống

* Xây dựng các hệ recommend systems (đề xuất sản phẩm phù hợp với người dùng dựa trên các thông tin nhân khẩu học như độ tuổi, giới tính, sở thích, thời gian nghe, thể loại hay theo dõi trong thời gian gần đây, trend…)
* Design các gói khuyến mãi phù hợp đối với từng đối tượng
* Xây dựng các mô hình tiên đoán về LTV, Churn để nâng cao chỉ số kinh doanh của doanh nghiệp

Note:

1. **Calculating RFM**

A more formal way of assigning points is to assign a number from 1 to 5 for each category, where 5 is the highest. In order to calculate RFM you’ll need some data on your customers:

Their most recent purchase date.

Number of purchases within a set time period (i.e. one year).

Total sales from that customer (you could also use average sales or average margin).

Next, decide on categories and assign numbers for those categories. Your exact categories will depend on your type of business, how long your business has been open and other factors, but as an example:

Recency: within a week (5), a month (4), a quarter (3) six months (2) a year (1)

Frequency: 10 purchases in the last year (5), 8 purchases (4), 6 purchases (3), 4 purchases (2), 3 purchases or under (1)

Total sales: Over $10,000 (5), $8,000-$9,999 (4), $6,000-$7,999 (3), $4,000-$5,999 (2) $3,999 or less (1).