AI-01 · Bài 3

Dữ liệu trông có vẻ hữu ích mà thật ra không nói lên điều gì

Lớp Dữ liệu nào dùng được trong phát triển sản phẩm

m.school by MES LAB

Học xong bài này bạn sẽ

- gọi tên được bốn dạng dữ liệu trông giống bằng chứng mà không phải.
- Soi được một báo cáo bạn đang nhận hằng tháng và chỉ ra dạng nào đang có trong đó.
- Nói được vì sao một con số hoàn toàn đúng vẫn có thể không giúp gì cho quyết định đang chờ.

1 · Một số đo cứ tăng đều mà không đổi được việc gì thì nó chỉ còn để trưng ra

- Dạng thứ nhất là một con số luôn đi lên, nhìn rất dễ chịu, mà không làm ai phải quyết khác đi.
- Dấu hiệu dễ thấy nhất là con số ấy được cộng dồn từ ngày đầu: tổng số sản phẩm đã bán từ khi mở xưởng, tổng số lượt người mở trang giới thiệu, tổng số hồ sơ khách đã thu về.
- Phép thử không nằm ở bản thân con số, mà nằm ở quyết định đứng sau nó.
- Còn một dấu hiệu nữa: con số ấy có chia nhỏ ra được không.
- Vì sao loại con số này lại nhiều tới vậy?
- Chúng tôi không nói những con số ấy phải bỏ đi.

2 · Dữ liệu chỉ còn lại của những người và những vật đã trụ lại được tới lúc nó được ghi

- Dạng thứ hai không nằm ở con số, mà nằm ở chỗ ai đã không có mặt trong bảng số.
- Thiên lệch do chỉ nhìn thấy phần còn sống sót
- Chuyện này có mặt khắp nơi quanh một sản phẩm.
- Chỗ nguy nằm ở chỗ bảng số vẫn đầy và vẫn đúng.
- Câu hỏi để nhận ra dạng này chỉ gồm một vế: trước khi bộ dữ liệu này được ghi, ai đã rơi ra khỏi nó.
- Cách xử lý thì không nằm trong bảng số.

3 · Một số đo đứng thay cho thứ khác, dùng lâu thì người ta quên mất nó chỉ đứng thay

- Dạng thứ ba bắt đầu từ một việc hoàn toàn hợp lý.
- Số đo thay thế không phải chuyện xấu.
- Số cuộc gọi khiếu nại đứng thay cho mức hài lòng của khách.
- Chỗ thứ nhất: thứ đứng thay bao giờ cũng bỏ sót một phần của thứ thật.
- Khi một số đo thay thế được dùng để chấm điểm người, thì người ta sẽ tìm cách làm cho con số ấy đẹp lên, và cách rẻ nhất để làm nó đẹp lên thường không động gì tới thứ thật.
- Cách giữ cho mình tỉnh táo là viết một câu và đặt nó ngay cạnh con số: con số này đang đứng thay cho điều gì, và nó lệch khỏi điều ấy ở chỗ nào.

4 · Số bình quân gộp hai nhóm khác hẳn nhau thì nó không mô tả đúng nhóm nào

- Dạng thứ tư là số bình quân.
- Lấy một ví dụ dựng ra cho dễ thấy.
- Chuyện còn đi xa thêm một bước.
- Vậy trước khi đọc một số bình quân, hãy hỏi nó đang gộp những nhóm nào vào với nhau.
- Có một cách làm nhẹ mà đủ dùng ở mức này.

Bảng tổng hợp

Bảng tổng hợp bốn dạng dữ liệu trông giống bằng chứng mà không phải

Dạng	Nó hiện ra thế nào trong báo cáo	Vì sao nó không giúp quyết định	Câu hỏi để nhận ra nó
Số đo chỉ để trưng ra	một con số cộng dồn, tháng nào cũng cao hơn tháng trước, không chia nhỏ được	nó lên trong cả tháng tốt lẫn tháng xấu, nên nó không phân biệt được hai tháng ấy	nếu con số này tăng mạnh, và nếu nó đứng yên, tôi có làm gì khác đi không
Chỉ còn phần đã trụ lại	một bảng đầy đủ, dựng từ hồ sơ của những người hoặc những vật còn ở lại trong hệ thống	phần đã rời đi không để lại dòng nào, mà đó thường là phần đáng xem	trước khi bảng này được ghi, ai đã rơi ra khỏi nó

Bảng tổng hợp

Bảng tổng hợp bốn dạng dữ liệu trông giống bằng chứng mà không phải tiếp

Dạng	Nó hiện ra thế nào trong báo cáo	Vì sao nó không giúp quyết định	Câu hỏi để nhận ra nó
Số đo đứng thay cho thứ khác	một con số đo được, đứng thay cho thứ muốn biết mà không đo thẳng được	nó bỏ sót một phần của thứ thật, và nó làm đẹp lên được mà thứ thật không đổi	con số này đang đứng thay cho điều gì, và lệch khỏi điều ấy ở chỗ nào
Số bình quân gộp hai nhóm khác nhau	một con số mô tả cả nhóm, thường đi một mình, không kèm khoảng rộng hẹp	nó rơi vào giữa hai nhóm nên không mô tả đúng nhóm nào, và nó đổi khi tỉ lệ hai nhóm đổi	con số này gộp những nhóm nào, tách ra thì mỗi nhóm bao nhiêu

Phần giá trị nhất của bài

Chỗ người ta hay làm sai

- **Chọn số đo theo mức dễ đo, chứ không theo quyết định đang chờ**

  Cách tránh là bắt đầu từ quyết định chứ không từ hệ thống: viết ra quyết định đang chờ, rồi mới hỏi số đo nào đổi được quyết định ấy.

- **Đọc dữ liệu của những người còn quay lại rồi kết luận cho cả những người đã bỏ đi**

  Cách tránh là mỗi lần đọc một nguồn hậu mãi thì viết thêm một dòng nói rõ nguồn này bỏ sót ai, và tìm ít nhất một nguồn chạm được nhóm bị bỏ sót.

- **Khen một con số vì nó tăng, mà không hỏi nó tăng thì mình làm gì khác đi**

  Cách tránh là đặt một câu hỏi cố định cho mọi con số trong báo cáo định kỳ, hỏi cả khi nó đẹp: con số này đổi thì tuần sau chúng ta làm gì khác đi.

Ví dụ để khái niệm sáng ra

Ví dụ minh hoạ

- Năm 1943, nhà thống kê Abraham Wald làm việc trong nhóm nghiên cứu thống kê của trường đại học Columbia và viết một loạt bản ghi nhớ về cách ước lượng mức dễ tổn thương của...
- Điểm then chốt của cách làm nằm ở phần dữ liệu không có: những chiếc không bay về thì không có mặt trong bảng số.
- Loạt bản ghi nhớ ấy không được công bố ra ngoài khi đó.

Điều rút ra: chỗ bảng số im lặng có thể là chỗ đáng đọc hơn cả phần nó nói.

Ví dụ dưới đây là một công trình có thật, bạn tra lại được tên tài liệu.

Bài tập tự chấm

Thử nhìn lại việc của mình

- [ ] Lấy một báo cáo định kỳ mà bạn đang nhận hằng tháng hoặc hằng quý, loại có liên quan tới sản phẩm bạn đang làm.
- [ ] In ra hoặc mở ra, rồi đi từng con số một.
- [ ] Với mỗi con số, đánh dấu nó thuộc dạng nào trong bốn dạng của bài này, hoặc ghi là không thuộc dạng nào.
- [ ] Với những con số bị đánh dấu, viết thêm một dòng: con số này đang thiếu điều gì để dùng được.
- [ ] Cuối cùng, chọn ra một con số và viết lại nó theo cách bạn thấy dùng được hơn, ví dụ tách nhóm ra, hay đổi từ cộng dồn sang theo từng kỳ.

Cách tự chấm và lời giải mẫu nằm trong bài trên trang web.

m.school

Đọc trọn bài và làm bài tập tại

mschool.meslab.org/lop/AI-01/bai-03.html

m.school by MES LAB · học không mất phí, không cần đăng nhập