

Du défaut à ce qui l'a produit

Cinq catégories de cause, une règle d'arrêt pour les cinq pourquoi, et pourquoi « il faut former le personnel » est la réponse que l'on écrit quand personne n'a cherché. Testez la cause sur les données avant de l'inscrire au rapport.

Pierre Robentz Cassion

Leçon 7 sur 8

Évaluation de la qualité des données — Transformer les constats en changement

Ce que couvre cette leçon

- Un constat n'est pas une cause, et seule une cause se corrige
- Cinq catégories, et elles n'exigent pas le même argent
- Laissez les données réduire le champ avant d'interroger quiconque
- Puis interrogez le formulaire
- Les cinq pourquoi, avec une règle d'arrêt
- Testez la cause sur les données
- Placez la cause au niveau où vit le remède
- La suite

Un constat n'est pas une cause, et seule une cause se corrige

Recommandation. Former le personnel à la qualité des données.

Cinq catégories, et elles n'exigent pas le même argent

Catégorie	Ressemble à	Se corrige par
Définition	Deux sites comptant des choses différentes sous un même nom d'indicateur	Une fiche de référence d'indicateur écrite
Outil	Un formulaire incapable d'exprimer ce qui s'est passé, ou un système qui perd des valeurs	Une modification du formulaire ou du paramétrage
Capacité	Le bon outil mal utilisé, systématiquement, par quelqu'un à qui on ne l'a jamais montré	Supervision et aide-mémoire, occasionnellement formation
Incitation	Un rapportage qui s'améliore quand on le regarde, ou des chiffres qui atteignent exactement une cible	Changer l'usage qui est fait du nombre
Système	Tout le monde échoue en même temps — ni formulaires, ni carburant, ni réseau, ni superviseur	Logistique, budget, effectifs

Cinq catégories, et elles n'exigent pas le même argent

- La définition est la plus grande catégorie et la moins soupçonnée — Un facteur de vérification éloigné de 1 traduit. . .
- La capacité est la plus petite catégorie et la plus fréquemment accusée — car c'est la seule dont le remède est un. . .

Laissez les données réduire le champ avant d'interroger quiconque — En Python

```
by_month = vax.groupby(vax["period"].dt.strftime("%Y-%m"))["reported"].mean()
by_facility = vax.groupby("facility_id")["reported"].mean()

print("spread across months: ", round(by_month.max() - by_month.min(), 2))
print("spread across facilities:", round(by_facility.max() - by_facility.min(), 2))
```

Laissez les données réduire le champ avant d'interroger quiconque — En R

```
by_month <- vax |> mutate(m = format(period, "%Y-%m")) |>  
  summarise(r = mean(report_submitted), .by = m)  
by_facility <- vax |> summarise(r = mean(report_submitted), .by = facility_id)  
  
c(months = diff(range(by_month$r)), facilities = diff(range(by_facility$r)))
```

Laissez les données réduire le champ avant d'interroger quiconque

- **Concentrée dans le temps, étalée sur les unités.** Août, partout. C'est une cause système — quelque chose a lâché. . .
- **Concentrée sur une unité, étalée dans le temps.** Le codage Y/N de SCH09, tous les mois. C'est une cause d'outil. . .
- **Étalée sur les deux.** L'attraction des âges, toutes équipes, toute l'année. C'est une cause de méthode, inhérente à. . .
- Écrivez la forme avant de proposer une cause — Cela élimine gratuitement trois des cinq catégories, et c'est l'étape. . .

Puis interrogez le formulaire

- Quelles valeurs le formulaire peut-il réellement accepter ? Les registres de SCH09 offraient des cases Y/N ; la...
- Le champ est-il obligatoire ? Un champ facultatif avec 40 % de non-réponse est une décision de conception de...
- Que dit l'intitulé du champ ? Une colonne intitulée « cas » recueille ce que la personne qui la remplit croit être un...

Les cinq pourquoi, avec une règle d'arrêt

- Arrêtez-vous quand vous atteignez quelque chose qu'une personne nommée peut changer avec un budget que vous pouvez. . .

Les cinq pourquoi, avec une règle d'arrêt

La présence a été sous-déclarée de 42 % à SCH09. — *Pourquoi ?* L'extraction a lu Y comme une valeur manquante. — *Pourquoi ?* Le registre employait Y/N et le formulaire attendait true/false. — *Pourquoi ?* Le registre papier imprimé en 2019 a des cases oui/non ; le formulaire numérique a été conçu plus tard, séparément. — **Arrêt.** Le bureau de l'éducation peut réimprimer le registre ou l'équipe SNIS peut accepter les deux codages. Les deux se chiffrent en une après-midi.

Testez la cause sur les données — En Python

```
# Cause: one school's register uses Y/N. Prediction: Y/N appears in that  
# school only, and in most of its months.  
yn = attendance[attendance["present"].isin(["Y", "N"])].merge(  
    roster[["student_id", "school_id"]], on="student_id", how="left"  
)  
print(yn["school_id"].value_counts())  
print(yn.groupby("school_id")["attendance_date"].nunique())
```

Testez la cause sur les données — En R

```
attendance |>
  filter(present %in% c("Y", "N")) |>
  left_join(select(roster, student_id, school_id), by = "student_id") |>
  count(school_id)
```

Placez la cause au niveau où vit le remède

- Une cause de **niveau formation** appelle une action de niveau formation. « La fiche de pointage de FAC020 manque » se. . .
- Une cause de **niveau district** appelle une action de niveau district. « Aucune formation n'a reçu de formulaires en. . .
- Une cause de **niveau système** appelle une action de niveau système, et la nommer est souvent ce que le rapport. . .

Placez la cause au niveau où vit le remède

Nommer les causes sert à rendre le tour suivant meilleur. Un rapport qui répartit les responsabilités avec exactitude et ne change rien a échoué à la seule chose pour laquelle il existait.

- Vous avez des constats, et chacun porte désormais une cause avec son niveau et son test.

Lire la leçon complète, avec le code exécutable

[https://data-analysis.cassion.dev/fr/cours/data-quality-assessment/
dqa-07-root-cause](https://data-analysis.cassion.dev/fr/cours/data-quality-assessment/dqa-07-root-cause)