

Tari n'est pas en panne

Les puits protégés tournent à 74,7 % en saison humide et à 36,4 % en saison sèche. Les forages à pompe manuelle ne bougent presque pas. Le champ d'état ne distingue pas ces deux défaillances, la séquence des visites le peut.

Pierre Robentz Cassion

Leçon 6 sur 8

Analyse EAH — Le service est une année, non une journée

Ce que couvre cette leçon

- Un seul état, plusieurs défaillances différentes
- Le signal saisonnier, et là où il n'est pas
- Classer chaque point à partir de sa propre séquence
- La valeur d'état qui ressemble à la réponse sans l'être
- La durée d'arrêt est une statistique de gestion
- Rapportez les formes, non le seul taux
- La suite

Un seul état, plusieurs défaillances différentes

Forme sur l'année	Ce que c'est	Que faire
À l'arrêt les mêmes mois chaque année, en service entre-temps	Saisonnier — la source se tarit	Approfondir, ou prévoir une alternative de saison sèche
À l'arrêt une fois, puis en service après un intervalle	Panne — réparée	Regarder la longueur de l'intervalle
À l'arrêt et le reste jusqu'à la fin	Abandonné ou en attente de pièces	Déterminer lequel ; ce n'est pas la même chose
En service et à l'arrêt à répétition	Chronique — sous-entretenu ou surchargé	Problème de gestion, non d'ouvrage

Le signal saisonnier, et là où il n'est pas — En Python

```
import pandas as pd

points = pd.read_csv("water-point-monitoring-2024.v1.csv", parse_dates=["visit_date"])
WORKING = {"functional", "partially-functional"}
DRY_MONTHS = {1, 2, 3, 11, 12}

points["dry_season"] = points["visit_date"].dt.month.isin(DRY_MONTHS)
points["ok"] = points["functional_status"].isin(WORKING)

print(points.groupby("dry_season")["ok"].agg(["mean", "size"]).round(3))
```

Le signal saisonnier, et là où il n'est pas — En R

```
library(dplyr)

points |>
  mutate(dry = lubridate::month(visit_date) %in% c(1, 2, 3, 11, 12),
         ok = functional_status %in% c("functional", "partially-functional")) |>
  summarise(functionality = mean(ok), n = n(), .by = dry)
```

Le signal saisonnier, et là où il n'est pas

- 67,8 % en saison sèche contre 79,4 % en saison humide — Onze points d'écart, qu'un rapport comparant une évaluation de...

Le signal saisonnier, et là où il n'est pas — En Python

```
seasonal = (  
    points.groupby(["source_type", "dry_season"])["ok"].mean().unstack()  
)  
seasonal.columns = ["wet", "dry"]  
seasonal["swing"] = seasonal["wet"] - seasonal["dry"]  
print((seasonal * 100).round(1).sort_values("swing", ascending=False))
```

Le signal saisonnier, et là où il n'est pas — En R

```
points |>  
  summarise(functionality = mean(ok), .by = c(source_type, dry)) |>  
  tidyr::pivot_wider(names_from = dry, values_from = functionality)
```

Le signal saisonnier, et là où il n'est pas

Type de source	Humide	Sèche	Écart
Puits protégé	74,7 %	36,4 %	38,3
Source protégée	84,8 %	41,0 %	43,8
Forage à pompe manuelle	75,0 %	76,1 %	−1,1
Borne de réseau	95,9 %	98,8 %	−2,9

Le signal saisonnier, et là où il n'est pas

- Deux types de source portent tout l'effet saisonnier et deux ne bougent pas du tout — Les puits peu profonds et les. . .

Classer chaque point à partir de sa propre séquence — En Python

```
def classify(group):
    group = group.sort_values("visit_date")
    failures = group.loc[~group["ok"]]
    if failures.empty:
        return "never failed"
    if not group["ok"].tail(3).any() and len(group) >= 3:
        return "down at year end"
    if failures["dry_season"].all():
        return "seasonal"
    return "intermittent"

shape = points.groupby("water_point_id").apply(classify, include_groups=False)
print(shape.value_counts())
```

Classer chaque point à partir de sa propre séquence — En R

```
points |>
  arrange(visit_date) |>
  summarise(
    shape = case_when(
      all(ok) ~ "never failed",
      !any(tail(ok, 3)) & n() >= 3 ~ "down at year end",
      all(dry[!ok]) ~ "seasonal",
      TRUE ~ "intermittent"
    ), .by = water_point_id) |>
  count(shape)
```

Classer chaque point à partir de sa propre séquence

Forme	Points
Jamais défailli	82
Intermittent	80
Saisonnier	46
À l'arrêt en fin d'année	34

Classer chaque point à partir de sa propre séquence

- L'ordre des tests compte et c'est un jugement, non un détail — Un point qui ne défaille qu'en mois secs et qui est. . .
- Notez l'ordre employé — Deux analystes avec le même fichier et des ordres différents produiront des effectifs. . .

La valeur d'état qui ressemble à la réponse sans l'être — En Python

```
ever_abandoned = points.loc[points["functional_status"].eq("abandoned"),  
                             "water_point_id"].nunique()  
print(f"points ever recorded as abandoned: {ever_abandoned}")  
print(f"points down at the end of the year: 34")
```

La valeur d'état qui ressemble à la réponse sans l'être — En R

```
points |> filter(functional_status == "abandoned") |>  
  summarise(points = n_distinct(water_point_id))
```

La valeur d'état qui ressemble à la réponse sans l'être

- Dérivez l'état de l'historique là où vous le pouvez, et employez l'état enregistré comme recoupement — Un champ qu'un...

La durée d'arrêt est une statistique de gestion — En Python

```
down = points.loc[~points["ok"] & points["days_since_breakdown"].notna()]
print(down.groupby("management")["days_since_breakdown"].agg(
    median="median", visits="size"
).sort_values("median"))
```

La durée d'arrêt est une statistique de gestion — En R

```
points |>  
  filter(!ok, !is.na(days_since_breakdown)) |>  
  summarise(median_days = median(days_since_breakdown), n = n(), .by = management)
```

La durée d'arrêt est une statistique de gestion

Gestion	Jours d'arrêt médians à la visite
Opérateur privé	29
Service public	32
Géré par une ONG	48
Comité communautaire	68

Rapportez les formes, non le seul taux — Exemple

Water point failure, 2024, 242 points

Never failed	82	33.9%
Intermittent	80	33.1%
Seasonal (dry months only)	46	19.0%
Down at year end	34	14.0%

Seasonality is concentrated in protected wells (74.7% wet, 36.4% dry) and protected springs (84.8% wet, 41.0% dry). Handpump boreholes and piped schemes show no seasonal effect.

Points classified as down at year end if not working at any of the last three visits; seasonal if every recorded failure fell in a dry month. Permanence tested before seasonality.

- Chaque nombre de cette leçon est calculé sur les visites qui ont été faites.

Lire la leçon complète, avec le code exécutable

```
https://data-analysis.cassion.dev/fr/cours/wash-analysis/  
wash-06-seasonal-or-abandoned
```