

Une grève qui ressemble à une urgence

SCH07 est à 88,2 % de présence et SCH18 à 85,2 %. Construisez le registre comme un calendrier complet et comptez chaque jour manquant comme une absence, et ils affichent 66,1 % et 63,9 % — les pires écoles du district, sur quinze jours où personne n'était censé être là.

Pierre Robentz Cassion

Leçon 4 sur 8

Analyse de programmes d'éducation — Inscrit n'est pas présent

Ce que couvre cette leçon

- Les lignes qui n'y sont pas
- Ce qui arrive si on ne regarde pas
- La règle
- Quand une fermeture est le constat
- La forme générale
- La suite

Les lignes qui n'y sont pas — En Python

```
import pandas as pd

attendance = pd.read_csv("school-attendance-2024.v1.csv")
roster = pd.read_csv("school-roster-2024.v1.csv")

joined = attendance.merge(roster[["student_id", "school_id"]], on="student_id")
calendar = sorted(joined["attendance_date"].unique())

days_per_school = joined.groupby("school_id")["attendance_date"].nunique()
print(f"school days in the calendar: {len(calendar)}")
print(days_per_school.sort_values().head())
```

Les lignes qui n'y sont pas — En R

```
library(dplyr)

attendance |>
  left_join(roster, by = "student_id") |>
  summarise(days = n_distinct(attendance_date), .by = school_id) |>
  arrange(days)
```

Les lignes qui n'y sont pas

- Vingt-deux écoles ont les soixante jours. SCH07 et SCH18 en ont quarante-cinq

Les lignes qui n'y sont pas — En Python

```
missing = {  
    school: sorted(set(calendar) - set(group["attendance_date"]))  
    for school, group in joined.groupby("school_id")  
}  
for school, dates in missing.items():  
    if dates:  
        print(f"{school}: {len(dates)} days, {dates[0]} to {dates[-1]}")
```

Les lignes qui n'y sont pas — En R

Which dates, not just how many. The pattern is the diagnosis.

Les lignes qui n'y sont pas

- Aux deux il manque exactement du 11 au 29 mars — et les mêmes quinze jours
- Des jours manquants consécutifs dans plusieurs écoles sont une fermeture jusqu'à preuve du contraire — Des jours. . .

Ce qui arrive si on ne regarde pas — En Python

```
def attendance_rate(school, missing_counts_as_absent):
    students = roster.loc[roster["school_id"] == school, "student_id"]
    marks = attendance[attendance["student_id"].isin(students)]
    present = marks["present"].isin(["true", "Y"]).sum()
    if missing_counts_as_absent:
        denominator = len(students) * len(calendar)
    else:
        denominator = marks["present"].isin(["true", "Y", "false", "N"]).sum()
    return present / denominator

for school in ("SCH07", "SCH18", "SCH01"):
    observed = attendance_rate(school, False)
    filled = attendance_rate(school, True)
    print(f"{school}: observed {observed:.1%}, grid-filled {filled:.1%}")
```

Ce qui arrive si on ne regarde pas — En R

Two denominators: marks made, and student-days in the calendar.

Ce qui arrive si on ne regarde pas

École	Sur les pointages faits	Sur une grille complète
SCH07	88,2 %	66,1 %
SCH18	85,2 %	63,9 %
SCH01	91,3 %	91,3 %

Ce qui arrive si on ne regarde pas

- Vingt-deux points, inventés — Sur les chiffres de la grille complète, SCH07 et SCH18 sont de loin les deux pires écoles. . .
- L'école non touchée est inchangée — ce qui rend l'erreur si difficile à attraper : le tableau a l'air correct, . . .

- Construisez le dénominateur à partir des jours d'ouverture de l'école, non du calendrier

```
open_days = joined.groupby("school_id")["attendance_date"].nunique()
enrolled = roster.groupby("school_id")["student_id"].nunique()
expected = (open_days * enrolled).rename("student_days_expected")

actual = joined.groupby("school_id").size().rename("marks_made")
coverage = (actual / expected).rename("mark_coverage")
print(pd.concat([expected, actual, coverage.round(3)], axis=1).head())
```

Expected student-days uses each school's own open days.

Quand une fermeture est le constat — Exemple

Attendance, February to April 2024

Average daily attendance, all schools	88.4%	on marks made
SCH07	88.2%	
SCH18	85.2%	

SCH07 and SCH18 were closed for the fifteen school days from 11 to 29 March, a quarter of the term. Their attendance figures are computed on the 45 days they were open and are comparable with other schools; their instructional time is not.

Counting the closure days as absences would report these two schools at 66.1% and 63.9% and rank them worst in the district.

Quand une fermeture est le constat

- Rapportez la fermeture en jours d'enseignement perdus, séparément de la présence
 - Les deux répondent à des questions. . .

La forme générale

Cours	La chose absente	Ce à quoi cela ressemblait
Données de routine et DHIS2	Une formation qui n'a pas rapporté	Un district à couverture en baisse
Analyse EAH	Une tournée que personne n'a conduite	Un district à fonctionnalité en hausse
Éducation	Un jour où l'école était fermée	Deux écoles en urgence d'abandon

- Dans les trois cas, l'absence n'est pas une valeur et rien dans le fichier ne l'annonce — L'habitude qui attrape les. . .

- La présence décrit un trimestre.

Lire la leçon complète, avec le code exécutable

[https://data-analysis.cassion.dev/fr/cours/education-programme-analysis/
edu-04-the-days-that-are-not-there](https://data-analysis.cassion.dev/fr/cours/education-programme-analysis/edu-04-the-days-that-are-not-there)