

Analyse des Projets - Rapport RMarkdown

Ermilsonn DERYS

2025-02-05

Introduction

Ce rapport analyse les données du fichier **DATA.xlsx**, qui contient des informations sur divers projets. L'objectif est de fournir une analyse descriptive et visuelle pour comprendre la répartition des projets par état, type, année, et d'autres dimensions clés.

Pour traiter ce (grand) jeu de données, nous chargerons la bibliothèque tidyverse. La bibliothèque regroupe plusieurs packages, dont ggplot2 et dplyr, indispensables pour l'analyse et la visualisation des données.

Description des Données

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
## v dplyr      1.1.4      v readr      2.1.5
## v forcats    1.0.0      v stringr   1.5.1
## v ggplot2    3.5.1      v tibble    3.2.1
## v lubridate  1.9.4      v tidyr     1.3.1
## v purrr      1.0.2
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
df = readxl::read_xls("C:/Users/ermil/Desktop/Data (1).xls")
```

Le fichier de données contient les colonnes suivantes :

- **Clé** : Identifie les différents projets.
- **Résumé** : Description détaillée du projet ou ticket.
- **État** : État actuel du projet (ex. "Fini", "Fermée", "En cours", etc.).
- **WSJF** : Une métrique prioritaire pour les projets (Weighted Shortest Job First).
- **Origine** : Source du projet (CAPA, EPIC, etc.).
- **Création** : La date de création du projet.
- **Responsable** : Responsable associé au projet.
- **Type de ticket** : Nature du projet (Feature, Enabler, etc.).
- **Étiquettes** : Tags associés au projet pour mieux le classer.

Nettoyage des données

On va maintenant séparer les tickets en fonction de deux natures possibles : Fonctionnelle et Technique. Et on va regrouper les tickets en fonction de leur projet. Nous savons que les projets commencent par "TFRA".

```
# Nettoyage des données et ajout de colonnes utiles
df1 = df |>
  filter(!is.na(Clé)) |> # Supprimer les lignes sans clé
  mutate(
    Nature = case_when(
      grepl("Technique", Résumé, ignore.case = TRUE) ~ "Technique",
      TRUE ~ "Fonctionnelle"
    ),
    Projets = ifelse(str_starts(Clé, "TFRA"), Clé, NA),
    Création = as.Date(Création) # Convertir la colonne Création en date
  ) |>
  fill(c(Nature, Projets), .direction = "down")
```

Analyse descriptive

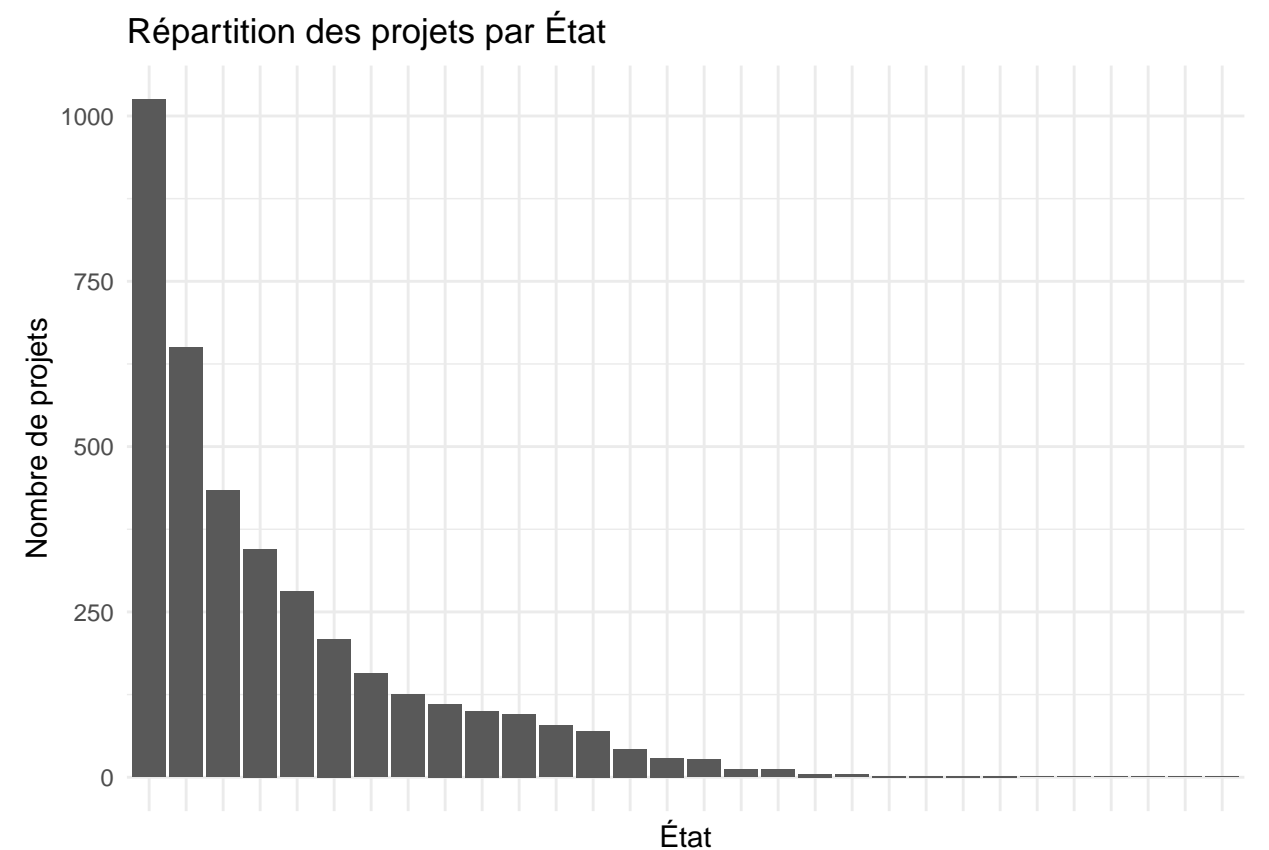
Répartition des projets par État

```
# Compter les projets par état
nbEtat = df1 |>
  group_by(État) |>
  summarise(Count = n(), .groups = "drop") |>
  arrange(desc(Count))
print(head(nbEtat))
```

```
## # A tibble: 6 x 2
##   État      Count
##   <chr>    <int>
## 1 Fini      1025
## 2 En Prod    650
## 3 Canceled  434
## 4 Mis en service 345
## 5 Ready     281
## 6 Funnel    209
```

```
# Graphique des projets par état
df1 |>
  group_by(État) |>
  summarise(Count = n(), .groups = "drop") |>
  ggplot(aes(x = reorder(État, -Count), y = Count)) +
  geom_bar(stat = "identity") +
  labs(
    title = "Répartition des projets par État",
    x = "État",
    y = "Nombre de projets"
  ) +
```

```
theme_minimal()+
theme(
  axis.text.x = element_blank())
```



On remarque que 3 statuts se démarquent : En cours, En Prod et Canceled.

Répartition des projets par Type de ticket

```
nbProjets = df1 |>
  group_by(Projets) |>
  summarise(Count = n(), .groups = "drop") |>
  arrange(desc(Count))

# Compter les projets par type de ticket
nbTicket = df1 |>
  group_by(`Type de ticket`) |>
  summarise(Count = n(), .groups = "drop") |>
  arrange(desc(Count))

print(nbTicket)
```

```
## # A tibble: 11 x 2
##   'Type de ticket' Count
```

```
##      <chr>          <int>
##  1 Feature          2271
##  2 Enabler Feature   994
##  3 Capability        312
##  4 Enabler Capability 100
##  5 Story             93
##  6 Dependency        35
##  7 Bogue             11
##  8 Support           5
##  9 Enabler           2
## 10 Evolution         2
## 11 Tâche             2
```

2 types de tickets se démarquent des autres : Feature et Enable Feature.

Nombre de projets par Année et État

```
# Analyse par année et état
Etat_an = df1 |>
  mutate(Année = as.integer(format(Création, "%Y"))) |>
  filter(!is.na(Année)) |>
  group_by(Année, État) |>
  summarise(Count = n(), .groups = "drop") |>
  arrange(desc(Count)) |>
  slice_head(n = 5)

print(Etat_an)
```

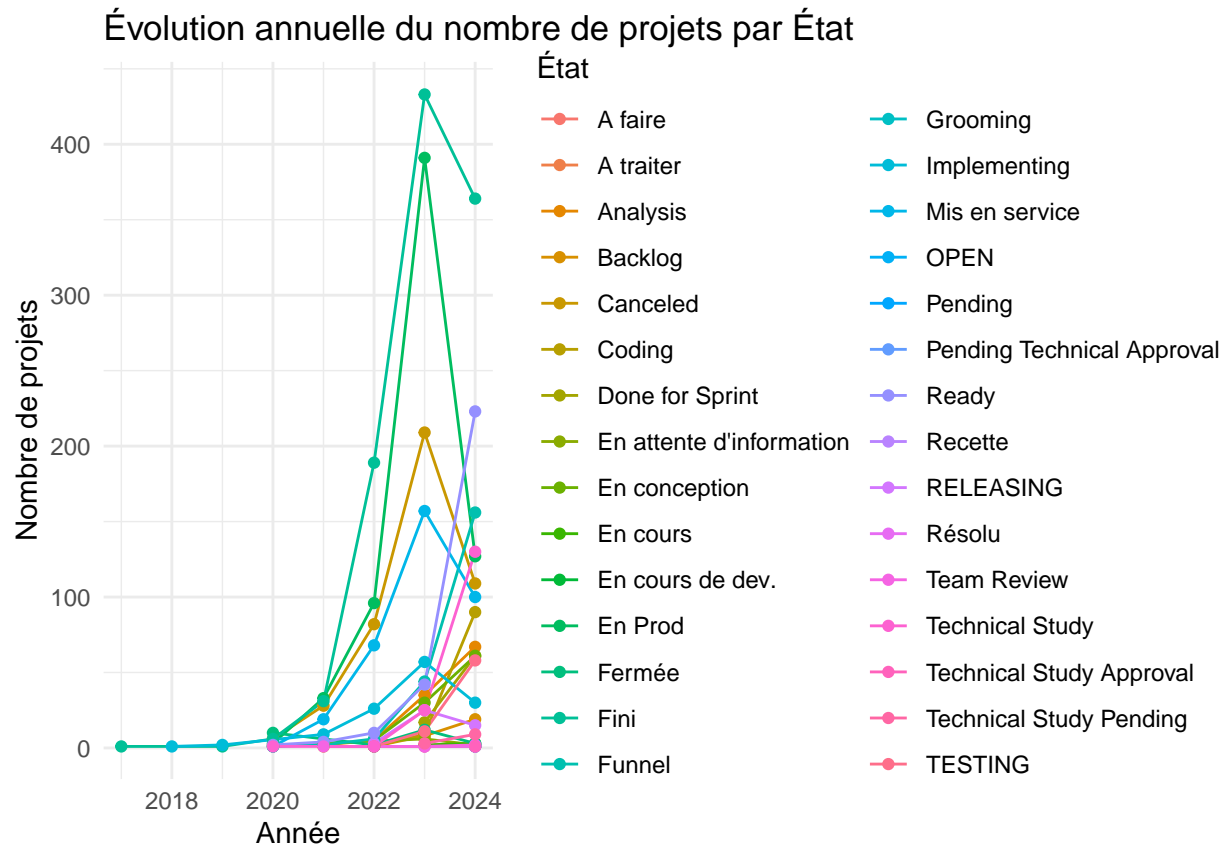
```
## # A tibble: 5 x 3
##   Année État      Count
##   <int> <chr>    <int>
## 1  2023 Fini      433
## 2  2023 En Prod   391
## 3  2024 Fini      364
## 4  2024 Ready     223
## 5  2023 Canceled  209
```

Les années 2023 et 2024 sont bien représentées : Une occurrence normale, étant donné que plus on avance dans le temps, plus l'entreprise a de projets, et donc plus elle a de tickets.

Visualisation de l'évolution annuelle par État

```
# Graphique de l'évolution du nombre de projets par état
df1 |>
  mutate(Année = as.integer(format(Création, "%Y"))) |>
  filter(!is.na(Année)) |>
  group_by(Année, État) |>
  summarise(Count = n(), .groups = "drop") |>
  ggplot(aes(x = Année, y = Count, color = État)) +
```

```
geom_line() +
geom_point() +
labs(
  title = "Évolution annuelle du nombre de projets par État",
  x = "Année",
  y = "Nombre de projets",
  color = "État"
) +
theme_minimal()
```



On remarque qu'en 2023, il y a un pic de changement d'état pour la très grande majorité des projets. Intéressons nous maintenant au tickets associé au projets

```
# Calculer le nombre de tickets par projet et leurs statistiques
stat_projet = df1 |>
  group_by(Projets) |>
  summarise(Nombre_Ticket = n(), .groups = "drop") |>
  arrange(desc(Nombre_Ticket))

# Afficher les 5 premiers projets
print(head(stat_projet, 5))
```

```
## # A tibble: 5 x 2
##   Projets   Nombre_Ticket
##   <chr>         <int>
```

```
## 1 TFRA-2144      152
## 2 TFRA-2007      98
## 3 TFRA-1995      88
## 4 TFRA-2008      87
## 5 TFRA-2001      72
```

```
# Calculer la moyenne et l'écart-type
stat_ticket = stat_projet |>
  summarise(
    Moyenne_Ticket = mean(Nombre_Ticket, na.rm = TRUE),
    sd_Ticket = sd(Nombre_Ticket, na.rm = TRUE)
  )

# Afficher les statistiques globales
print(stat_ticket)
```

```
## # A tibble: 1 x 2
##   Moyenne_Ticket sd_Ticket
##           <dbl>     <dbl>
## 1           9.31      14.9
```

Dans un premier temps on voit qu'il y a 3799 tickets pour les projets de nature fonctionnelle, et 28 pour les projets de nature technique. Le projet TFRA-2144 cumule 152 tickets. En moyenne, il y a 9.3 tickets par projet et un écart-type de 14.92. Un écart-type assez conséquent.