

PROJET : R

Analyse de données relatives aux personnes vaccinées contre la Covid-19

|  |  |  |
| --- | --- | --- |
| Doray Balassoupramanien Lahoucine Iberdi Hicham ECHARIF |  |  |

f

Table des matières

1. [Identité du document 2](#_Toc122680097)

[Projet](#_Toc122680098)

1. [Présentation des intervenants 3](#_Toc122680099)

[Qui sommons nous ?](#_Toc122680100)

[Constitution du groupe](#_Toc122680101)

1. [Présentation du projet 3](#_Toc122680102)

[Contexte](#_Toc122680103)

1. [Présentation du cahier des charges 3](#_Toc122680104)

[Objectif](#_Toc122680105)

1. [Démarche gestion de projet 4](#_Toc122680106)

[Méthodologie Agile](#_Toc122680107)

[Trello](#_Toc122680108)

1. [Description détaillée 5](#_Toc122680109)

[Analyse 1 (Doray)](#_Toc122680110)

[Analyse 2 (Doray)](#_Toc122680111)

1. [Conclusion 11](#_Toc122680112)
2. [La partie technique 11](#_Toc122680113)
3. [Annexes 12](#_Toc122680114)

[Données importantes](#_Toc122680115)

# Identité du document

## Projet

|  |  |
| --- | --- |
| Titre | rapport\_projetR |
| Project concerné | Analyse de données relatives aux personnes vaccinés |
| Date | 22/12/2022 |
| Version | V1 |

# Présentation des intervenants

## Qui sommons nous ?

Etudiant en Mastère spécialité Dev, Big Data & IA.

## Constitution du groupe

L’équipe est constituée de Doray , Lahoucine et Hicham.

# Présentation du projet

## Contexte

Le but est de ce projet est de réaliser une solution avec le langage R pour faire des traitements de données relatives aux personnes vaccinées contre le COVID-19.

L'application des techniques statistiques dans le milieu médical est courante. Cela fait partie du travail d'investigation où l'on doit recueillir, analyser et interpréter des données pour en tirer et dégager des recommandations.

Nous allons utiliser comme source de données ce fichier « vacsi-s-a-reg-2022-12-19-19h00.csv » disponible sur le site <https://www.data.gouv.fr/fr/datasets/donnees-relatives-aux-personnes-vaccinees-contre-la-covid-19-1/#resources>

# Présentation du cahier des charges

## Objectif

L’objectif est de réaliser 5 analyse de données grâce aux langage de programmation R. Proposer des graphiques pour la présentation de la solution finale.

# Démarche gestion de projet

## Méthodologie Agile

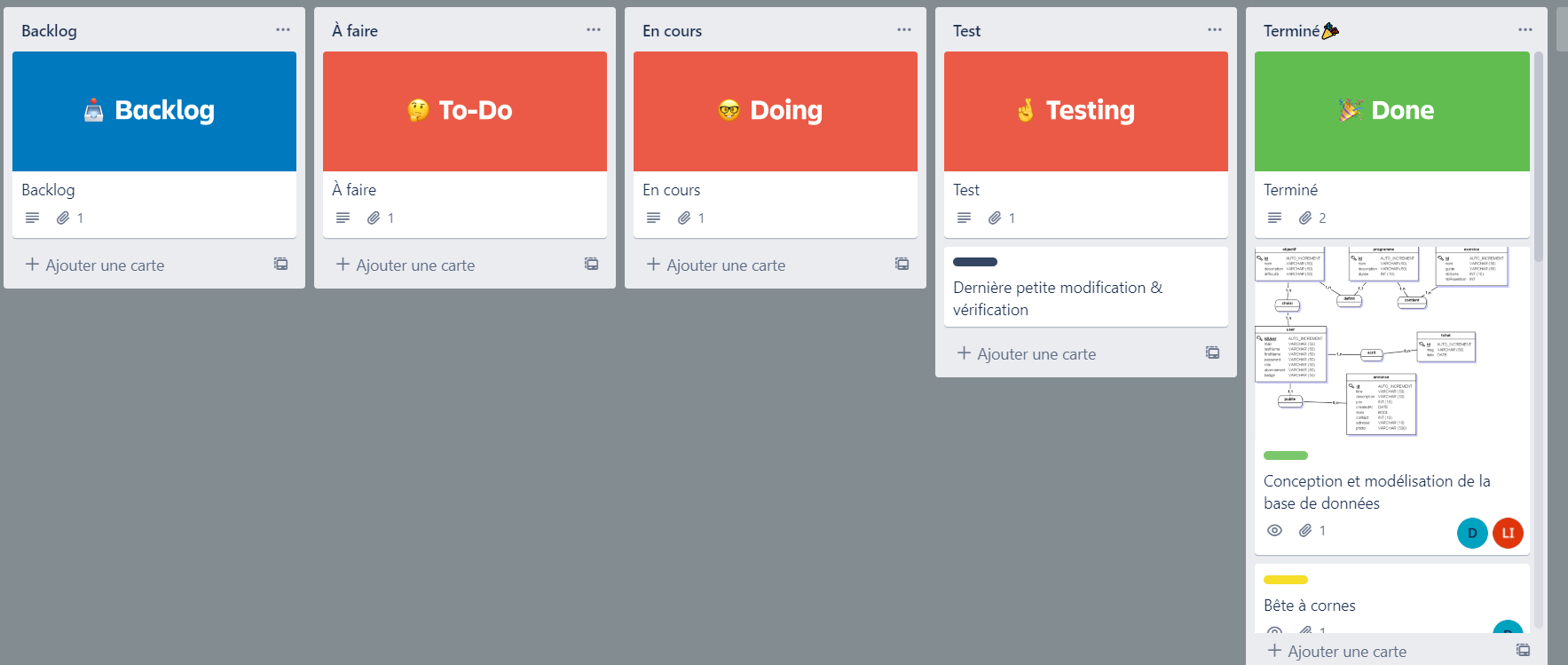
La mise en place de la méthodologie agile permet de se fixer des objectifs à courts terme, le projet est converti en plusieurs mini-projet. Cette approche est plus flexible. Puisqu'il est impossible de tout prévoir et de tout anticiper, elle laisse la place aux imprévus et aux changements.

## Trello

Dans le but d’organiser le projet nous avons décidé d’utiliser un outil pour pouvoir tout planifier. Trello fonctionne en utilisant la méthode kanban permettant d’améliorer une application ou un processus en production.

L’objectif est de pouvoir répondre aux besoins des utilisateurs et cette méthode est qualifiée comme méthode agile.

Le système kanban fonctionne avec un workflow par l’utilisation d’un tableau et vous permet d’avoir un dashboard du suivi de l’avancement des tâches confier. Le fonctionnement est assez ludique, il fonctionne avec un système de ticket.



# Description détaillée

## Analyse 1 (Doray)

L’étude de cette analyse comporte une prévision du nombre de dose total c’est-à-dire la somme des doses. Dans cette étude nous allons démontrer les différentes prévisions pour 2023.

Dans un premier temps nous récupérons les données via le site du gouvernement sous format Excel puis nous appliquons des traitement afin de nettoyer les données récupérer. Une conversion en data frame puis un filtre sur la date.

Une image contenant texte

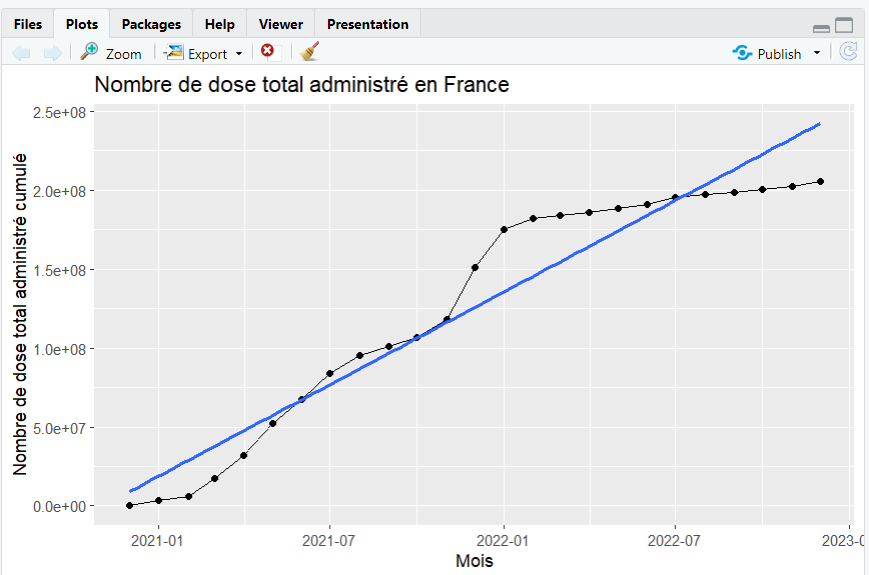
Description générée automatiquement

Ensuite, pour récupérer le nombre de dose effectué sur la journée nous allons appliquer une additions des différents colonnes (voir L-24 – Annexe-Analyse-1) et voici le résultat :

**Une image contenant table

Description générée automatiquement**

Puis, une fois ces données récupérer nous allons les interpréter dans un graphique, pour cela nous allons utiliser plot (voir L-32 – Annexe-Analyse-1).

****

Et pour conclure cette analyse nous allons prédire pour janvier 2023.

## Analyse 2 (Doray)

L’étude de cette analyse comporte une prévision du nombre de vaccine (3ème dose ou 2ème dose de rappel) pour homme/femme et dans cette prédiction nous allons prédire pour les deux premiers mois de l’année 2023.

Dans un premier temps nous récupérons le fichier et nous appliquons les différents traitement (voir Analyse 1 pour plus de détail).

Ensuite, pour récupérer le nombre de dose homme/femme nous allons appliquer des additions des différents lignes correspondant à leurs région (voir L-23/24 – Annexe-Analyse-2) et voici le résultat :

**Une image contenant table

Description générée automatiquement**

Nous obtenons une liste contenant plusieurs listes chacune correspondent à une région.

Prenons la région numéro 20 comme exemple.

Une image contenant table

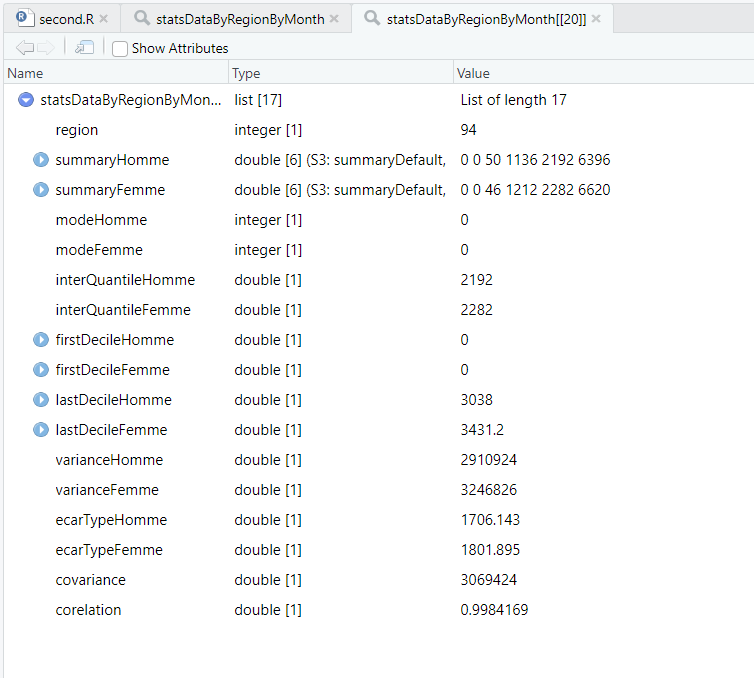
Description générée automatiquement

Puis grâce à ses informations nous allons établir des données statistiques, pour cela nous allons créer une liste puis dedans on affecte les différents calculs (moyenne, variance, écart-type etc. …) pour cela nous allons itérer sur l’ensemble de toutes les régions et affecté ceci dans le nouvelle liste (voir L-35/54 – Annexe-Analyse-2). Ce qui va nous créer une liste contenant les statistiques de toutes les régions :

Une image contenant table

Description générée automatiquement

Comme vous pouvez le voir ci-dessous nous avons dans la région (94) les différents informations statistiques.



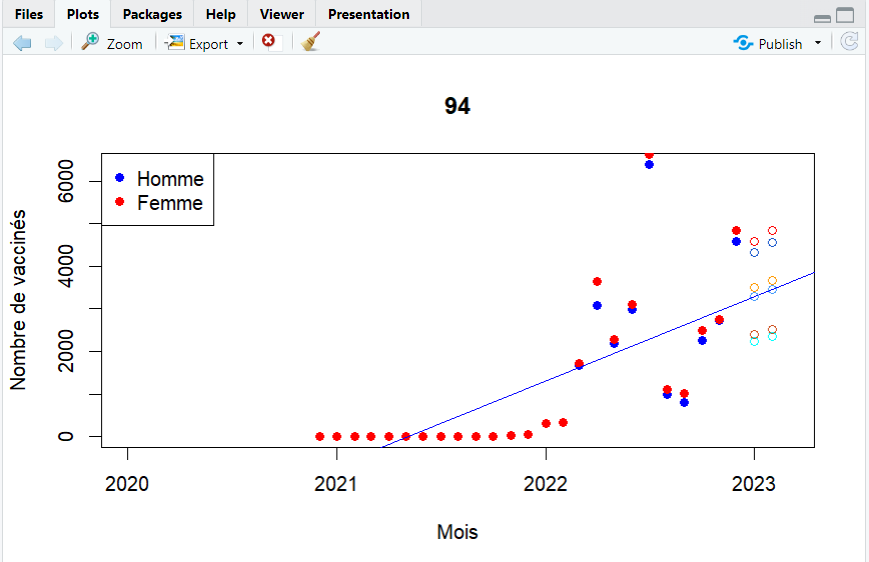
Ensuite ce data frame « statsDataByRegionByMonth » on va l’utiliser pour faire un jolie graphique (voir L-59 – Annexe-Analyse-2).

Pour conclure cette analyse nous allons procéder à la prédiction de cette analyse portant à savoir le nombre d’homme et de femme qui aurait pris un 2 -ème dose de rappel.

Pour cela nous utilisons un modèle pour l’utilisation d’un graphique linéaire (voir L-79 ou L-84 – Annexe-Analyse-2).

Ayant ce modèle nous allons utiliser pour créer une prédiction sur le mois de janvier et février 2023 pour les hommes et femmes dont chacun dispose de 3 catégorie de prédiction, prédiction dit « fit » qui correspond à la prédiction moyenne, prédiction dite « lower » l’estimation pessimiste et prédiction dit « upper » la prédiction optimiste.

Grâce à ceci nous pouvons ajuster notre graphe est ajouté les différents points de prédilections voir image ci-dessous, on peut voir qu’ici pour la région 94 les femmes sont plus préoccupés à faire le 2ème dose de rappel que les hommes qui sont légèrement moins préoccupés.



# Conclusion

**Doray**

Grâce à ce mini projet j’ai pu consolide mon niveau d’apprentissage concernant les statistiques ayant un niveau néophyte dans ce domaine cela ma permis d’apprendre un nouveau langage de programmation que j’espère réutiliser lors des manœuvres de projet big data très utile pour faire des prédictions sur une analyse qu’on souhaite faire.

Concernant le projet au début il est assez dur de comprendre le résultat à obtenir , on peut s’y perds assez facilement et des remise en questions permet de revenir sur l’analyse afin d’avancés sur celle-ci. C’est assez dommage de ne pas avoir fait d’IHM pour interpréter nos résultats avec un peu plus de temps et de meilleur compréhension dès le départ cette idée aurait pu être envisageable

# La partie technique

**Informations concernant le projet**

**Lien GITHUB code source :** https://github.com/dorayy/projet-r

# Annexes

**Analyse-1**

**Une image contenant texte

Description générée automatiquement**

**Analyse-2**

**Une image contenant texte

Description générée automatiquement**

## Données importantes

**Une image contenant texte

Description générée automatiquement**

**Une image contenant table

Description générée automatiquement**

**Une image contenant texte

Description générée automatiquement**