

11 JUIN 2025

SEWA, ZAKARIA, MARIE

LES CRISES CARDIAQUES



ETAPE 1(a) Sélection et Exploration du Dataset

Les maladies cardiovasculaires, et en particulier les crises cardiaques, représentent aujourd'hui l'une des principales causes de mortalité dans le monde. Comprendre les facteurs qui influencent la survenue d'un arrêt cardiaque est un enjeu majeur en santé publique, tant pour améliorer la prévention que pour optimiser la prise en charge des patients à risque.

Ce projet a pour objectif d'analyser un jeu de données médicales comprenant diverses informations cliniques (âge, sexe, rythme cardiaque, tension artérielle, taux de sucre, troponine, etc.) afin de mieux comprendre les caractéristiques associées à la survenue d'une crise cardiaque.

L'objectif est donc d'identifier le facteur permettant de différencier un patient sain d'un patient ayant déjà été victime d'une crise cardiaque, ou plus largement d'un événement cardiaque ayant causé des lésions détectables, afin de mieux prévenir d'éventuelles rechutes.

À travers des visualisations interactives et une modélisation prédictive, l'ambition est d'identifier les variables les plus déterminantes dans l'apparition de ce type d'événement critique. L'analyse repose sur l'utilisation d'outils de data science, notamment le langage R via RStudio, pour explorer, nettoyer, visualiser et modéliser les données.

age	age du patient
Sexe	homme/femme
Fréquence cardiaque	nombre de battement cardiaque par minutes
Pression artérielle sytolique	la pression dans les artères lorsque le coeur se contracte
Pression artérielle diastolique	la pression artérielle entre les battements cardiaque
glycémie	taux de glucose dans le sang
CK-mb	une enzyme cardiaque libérée lors des lésions du muscle cardiaque
Troponine	Un biomarqueur de protéines très spécifique pour les lésions du muscle cardiaque
Résultat	L'étiquette de résultat indiquant si le patient a eu ou non une crise cardiaque

	A	B	C	D	E	F	G	H	I
1	Age	Gender	Heart rate	Systolic blood press	Diastolic blood press	Blood sugar	CK-MB	Troponin	Result
2	64	1	66	160	83	160	1.8	0.012	negative
3	21	1	94	98	46	296	6.75	1.06	positive
4	55	1	64	160	77	270	1.99	0.003	negative
5	64	1	70	120	55	270	13.87	0.122	positive
6	55	1	64	112	65	300	1.08	0.003	negative
7	58	0	61	112	58	87	1.83	0.004	negative
8	32	0	40	179	68	102	0.71	0.003	negative
9	63	1	60	214	82	87	300	2.37	positive
10	44	0	60	154	81	135	2.35	0.004	negative
11	67	1	61	160	95	100	2.84	0.011	negative
12	44	0	60	166	90	102	2.39	0.006	negative
13	63	0	60	150	83	198	2.39	0.013	negative
14	64	1	60	199	99	92	3.43	5.37	positive
15	54	0	94	122	67	97	1.42	0.012	negative

ETAPE 1(b) IDENTIFICATION DES PROBLEME DU DATASET :

→ Quels sont les principaux problèmes du dataset ?

Supprimer les doublons, nettoyer les types de données,
supprimer ou recoder les valeurs aberrantes,
et nettoyer les colonnes de textes.

→ Quelles variables nécessitent un nettoyage en priorité ?

Vérification des valeurs abérante de toutes les variables

Passage de Gender en binaire

Renommer les résultats positif ou négatif en Sain – Malade pour plus de clarté

ETAPE 2 NETOYAGE DES DONNEES

→ Comment avez-vous corrigé les problèmes de qualité des données ?

Dans notre code R , nous avons utilisée df\$gender et df\$result

```
# 🔄 Transformation des données
df$gender <- factor(df$gender, levels = c(0, 1), labels = c("Femme", "Homme"))
df$result <- factor(df$result, levels = c("negative", "positive"), labels = c("Sain", "Pathologique"))
df <- df %>% distinct()
```

Ici nous avons le nétoyage des noms de colonnes

```
# 💡 Nettoyage des noms de colonnes
df <- clean_names(df)
```

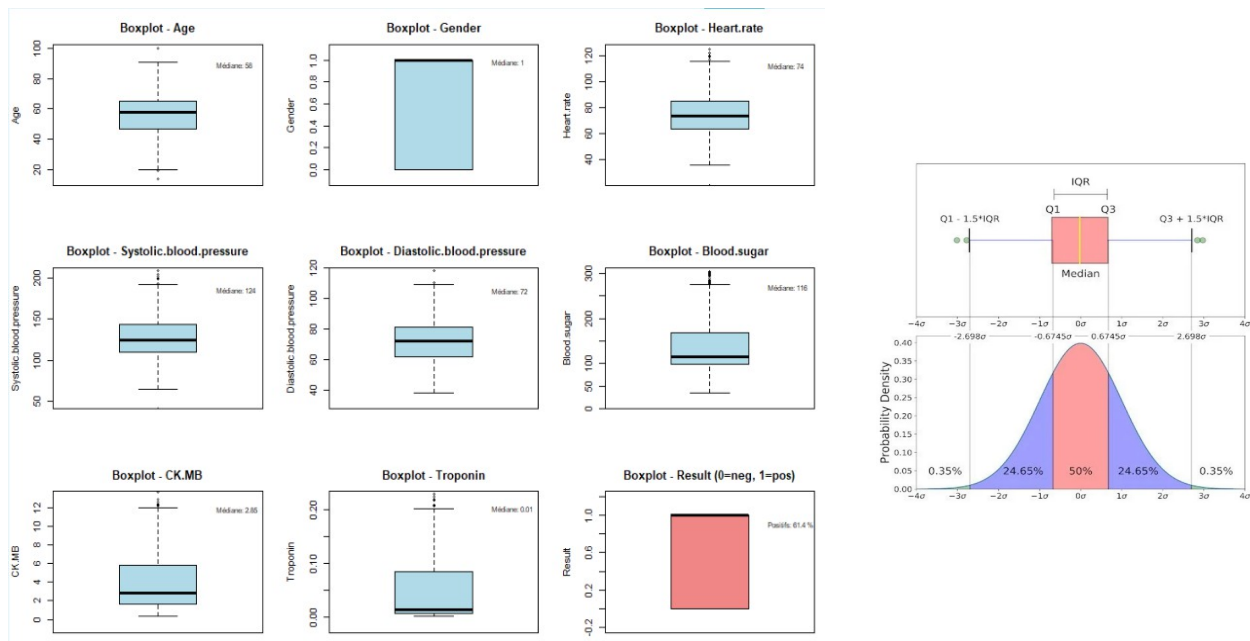
Ici nous avons la détection des doublons

```
# 🗑️ Détection des doublons
df <- df %>% distinct()
```

Détection avec Boxplot des valeurs abérantes

```
# 📊 Détection des valeurs aberrantes pour CK-MB
boxplot(df$ck_mb, main = "Boxplot CK-MB", col = "lightblue")

# 📄 Filtrage des valeurs CK-MB extrêmes (facultatif, à ajuster)
df <- df %>% filter(ck_mb < 50)
```



Face à l'impossibilité de mettre de côté toutes les données aberrantes, car dans notre cas, ce sont justement les valeurs qui se distinguent qui nous permettent de différencier les patients sains des patients pathologiques — nous avons mis en place un filtre qui nous permet d'ajuster le maximum ou le minimum que l'on souhaite conserver.

```

21
22 # ===== Filtrage des valeurs aberrantes =====
23 df <- df %>%
24   filter(!(result == "Pathologique" & ck_mb > 20)) %>% # CK-MB > 20 pour pathologiques
25   filter(troponin <= 2.5) # Troponine > 2.5
26
27 # 🚨 Filtrage des valeurs selon critères médicaux spécifiques
28 print(paste("Nombre de lignes avant filtrage:", nrow(df)))
29
30 df <- df %>%
31   filter(
32     troponin <= 0.35, # Troponine ≤ 1.0 ng/mL
33     ck_mb <= 10, # CK-MB ≤ 10
34     diastolic_blood_pressure <= 100, # Pression diastolique ≤ 100 mmHg
35     blood_sugar <= 300, # Glycémie ≤ 300 mg/dL
36     systolic_blood_pressure <= 170, # Pression systolique ≤ 170 mmHg
37     heart_rate <= 110 # Fréquence cardiaque ≤ 110 bpm
38   )
39
40 print(paste("Nombre de lignes après filtrage:", nrow(df)))
41
42 # =====
43

```

➔Quels traitements ont eu le plus d'impact sur la structure des données ?

Le traitement qui a eu le plus d'impact est la détection et gestion des valeurs aberrantes, cela nous

a permis de pouvoir comparer sur une meme echelle les patients sain et patient pathologique de manière visible .

ÉTAPE 3 : Développement du Dashboard Interactif sous R Shiny

METHODE

PARTIE FILTRE

Nous avons choisis de séparer les facteurs en fonction de leur nature , la fréquence cardiaque et l'âge se soumettent bien au filtre de plage de mesure (tranches d'âge ou tranche de fréquence cardiaques) . Le sexe et le caractère sain ou pathologique étaient binaires donc très propices à mettre dans la partie filtre

PARTIE DIAGNOSTIQUE

Le diagnostic est la partie où l'on peut voir chaque facteur (glycémie , pression ..) indépendamment donc voir lequel est présent chez le pathologique et pas chez le sain

PARTIE CORRELATIONS

Cette partie existe afin de confronter les facteurs entre eux et détecter une corrélation ou une spécificité . Ce qui est intéressant c'est que l'on est face à des mesures quantitatives alors on peut les placer sur une abscisse et une ordonnée afin d'avoir un graphique

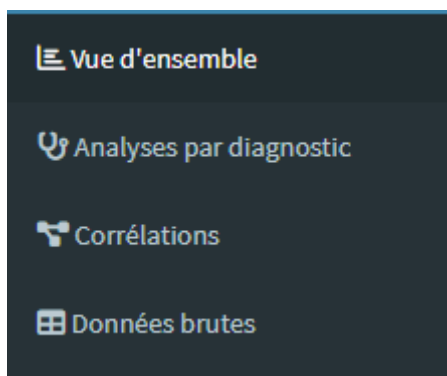
➔Quels types d'interactions avez-vous intégrés ? Nous avons 4 pages différentes

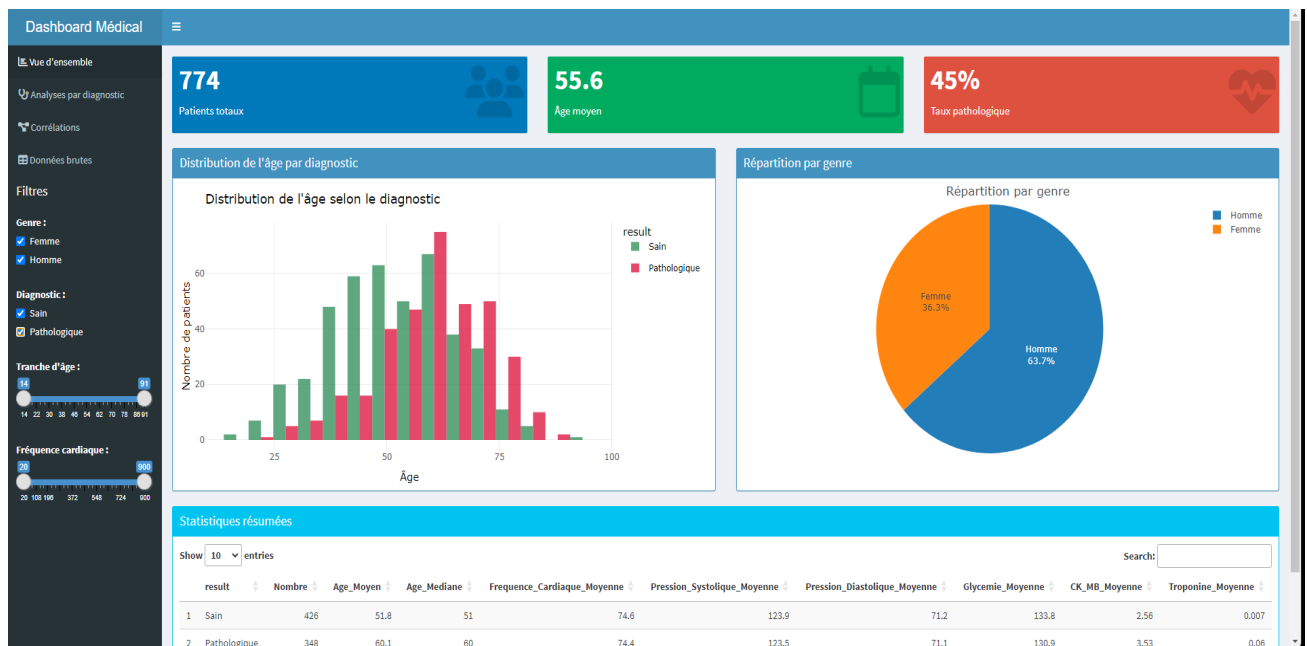
La première page est "Vue d'ensemble" .

Elle contient le nombre de nos patients , l'âge moyen total , et le taux de pathologies (de personnes qui ont fait une crise cardiaque)

En dessous 2 Visuels , le premier un graphique qui présente la Distribution de l'âge selon le diagnostic , le deuxième présente la répartition par genre des patients .

En fin nous avons la moyenne de tous nos relevés .

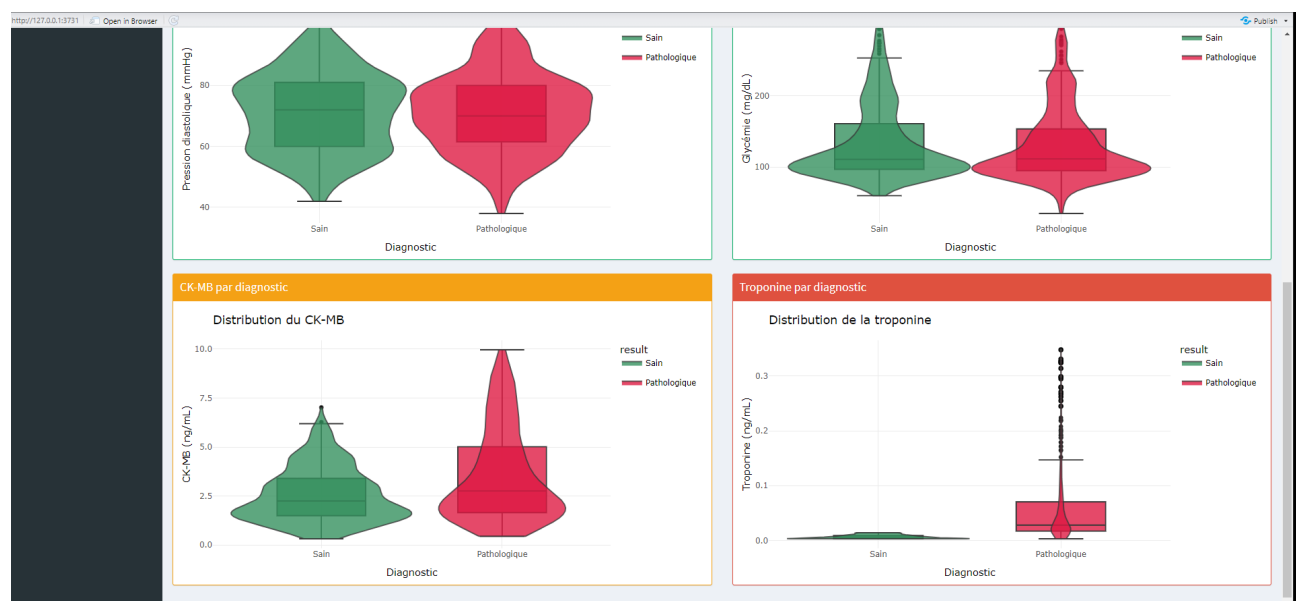




La seconde page est "Analyse par Diagnostique"

Cette page présente tous les diagnostique (saint pathologique) face aux différents relevés mesurés , on peut alors voir quels sont les facteurs clés qui expliquent les crises cardiaques .

Nous pouvons voir que tous les facteurs sont relativement proches mais seul le facteur Troponine (biomarqueur qui révèle la présence de lésions cardiaques) est ici significativement plus élevée chez les patients pathologiques



=====

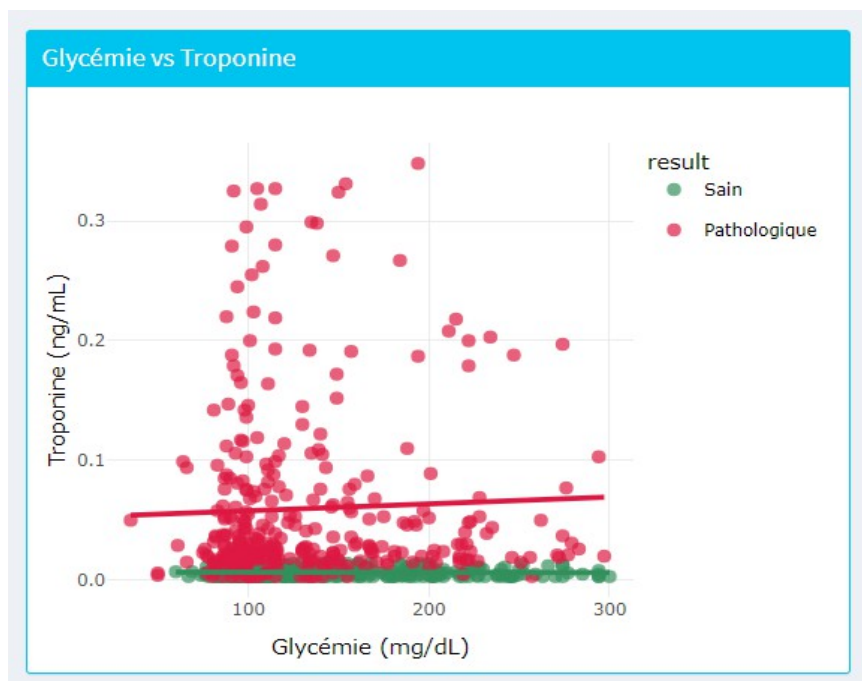
La troisième page est "Correlations"

Ici nous avons la comparaison entre :

Fraquence cardiaques - Pression systolique - Pression diastolique- Glycémie - CK-MB - Troponine



On peut observer que aucune autre corrélation ne soit aussi bien révélatrice que la Tromponine avec le reste des facteurs relevé ,nous pouvons conclure que le taux de Tromponine biomarqueur des lésions cardiaques est le premier élément a vérifié lorsque le patient à deja était victime d'une crise cardiaque



➔Comment les utilisateurs peuvent-ils manipuler les données en temps réel ?

Dans notre Dashboard l'utilisateur peut sélectionner un profil de patient type attraver un filtre sur le coté de la fenetre . Il peut choisir le résultat (sain pathologique) ,la plage d'age ,le sexe et la plage de frèquence cardiaque , afin de voir les different graphiques de comparaison ne présenté que le profil de patient qu'il souhaite .



Conclusion

L'analyse menée à travers notre dashboard interactif sous R Shiny permet de visualiser et de comparer différents facteurs biologiques et vitaux entre patients sains et pathologiques. Grâce à la segmentation des données et aux visualisations proposées, il ressort de manière évidente que la troponine est le facteur le plus discriminant entre les deux groupes.

En effet, parmi l'ensemble des variables mesurées (fréquence cardiaque, pression artérielle, glycémie, CK-MB...), seul le taux de troponine montre une élévation significative chez les patients ayant subi une crise cardiaque. Ce biomarqueur, bien connu pour révéler des lésions cardiaques, se distingue par sa forte spécificité dans ce contexte.

Nous pouvons donc conclure que **la troponine est l'indicateur le plus fiable pour confirmer qu'un patient a été victime d'un infarctus du myocarde**, et qu'elle doit être systématiquement vérifiée en cas de suspicion clinique.

Axes d'amélioration

Identifier un jeu de données plus riche, contenant davantage de relevés, et intégrant si possible des variables encore plus étroitement liées aux facteurs de risque connus des crises cardiaques — par exemple le taux de cholestérol, le tabagisme ou les antécédents familiaux.