



Certification

Bloc de compétences BC06

RNCP niv.7 n°32123

Yen Phi Do | Hugo Alpiste | Sébastien Martel | Morgane Geoffroy

04/09/2023

Introduction

Identifiant	Liste de compétences
A1	Recevoir la demande et rédiger un cahier de charges techniques pour la conception et la mise en place d'une solution d'analyse des données volumineuse
A2	Installer et configurer l'écosystème Hadoop
A3	Concevoir et déployer un système d'entrepôt de données structurées et non-structurées
A4	Définir l'architecture des données
A5	Ecrire des algorithmes d'analyse de données
A6	Maîtriser la recherche étendue (ElasticSearch)
A7	Concevoir un système d'intelligence artificielle et d'apprentissage automatique (Machine Learning)
A8	Maîtriser l'analyse et la science de données
A9	Développer des requêtes SQL et NO SQL pour traiter des données volumineuses
A10	Sécuriser les bases de données et mettre en place des procédures de sauvegarde et de restauration

Plan



Projet 1



Projet 2



Projet 3

Plan

Projet 1

Cahier
des charges

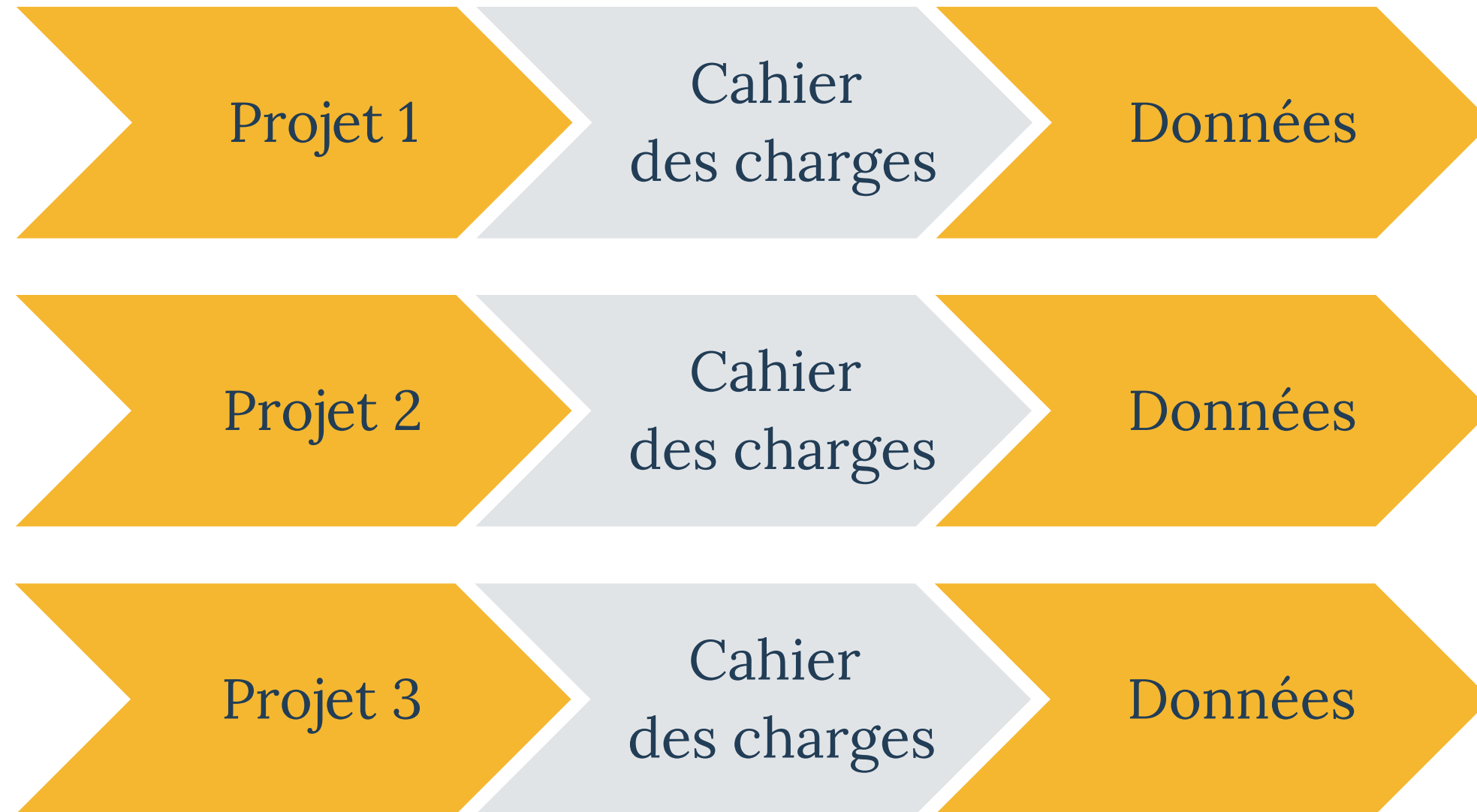
Projet 2

Cahier
des charges

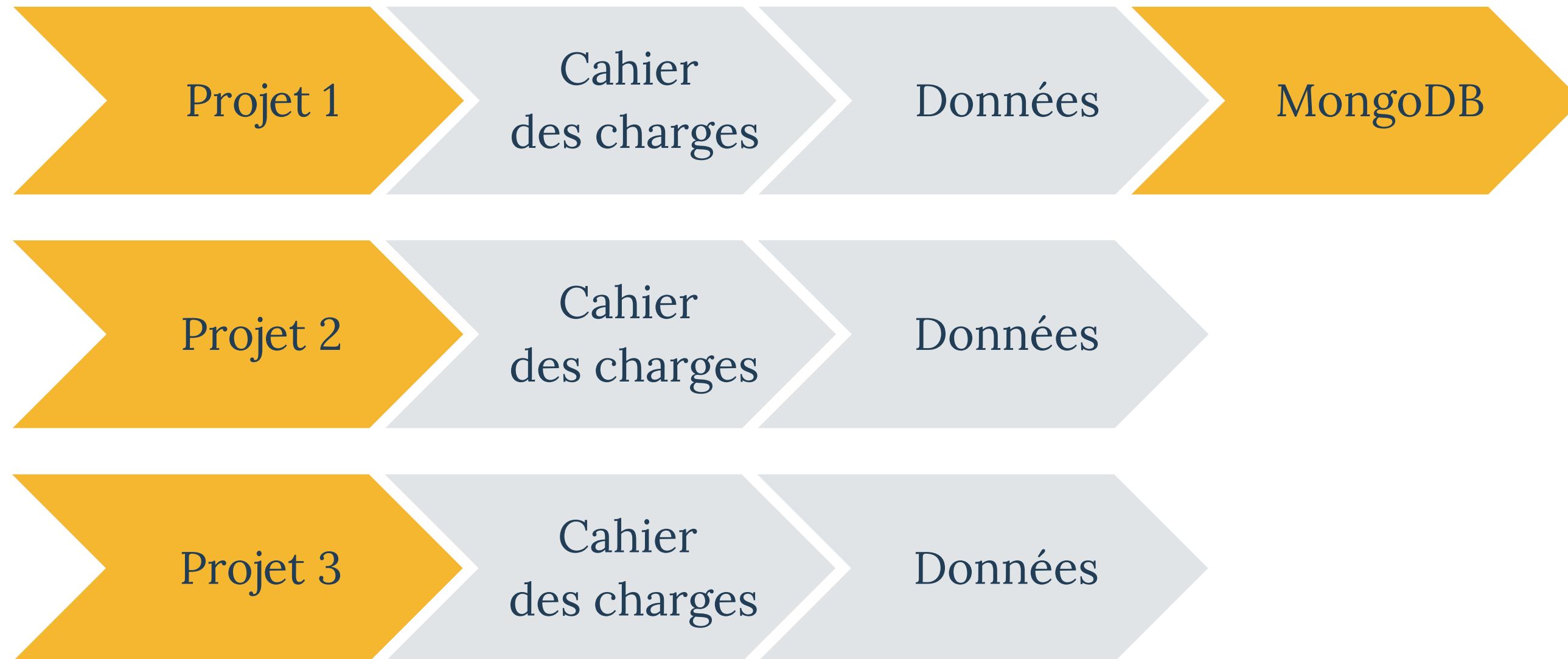
Projet 3

Cahier
des charges

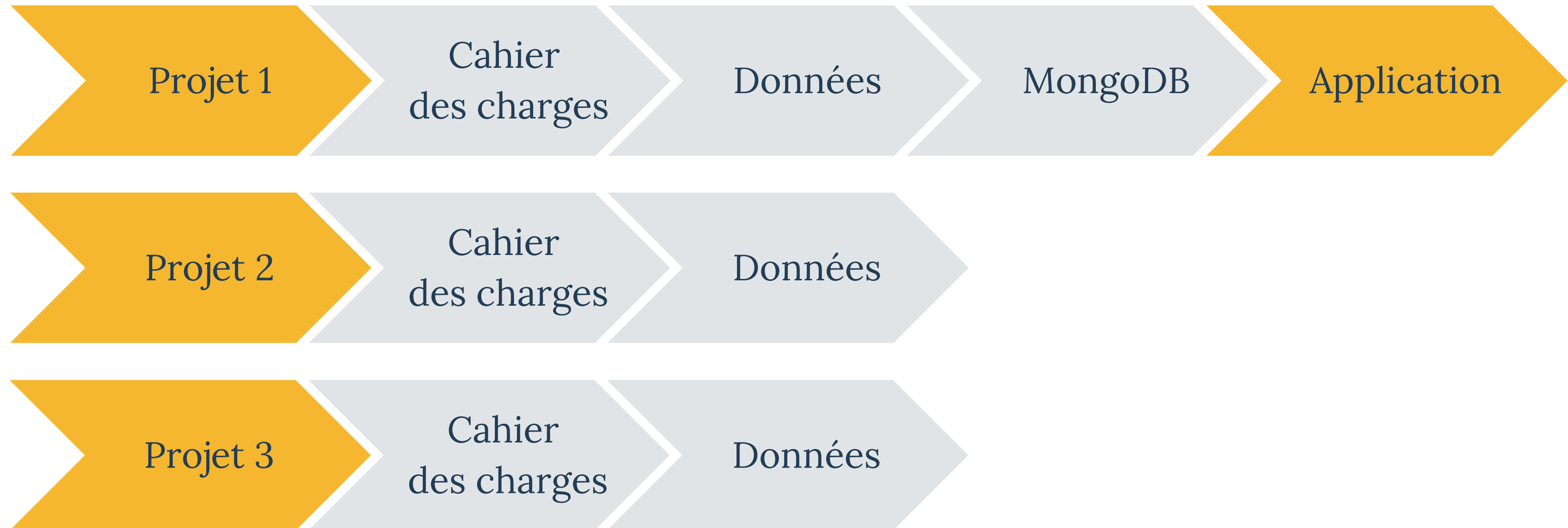
Plan



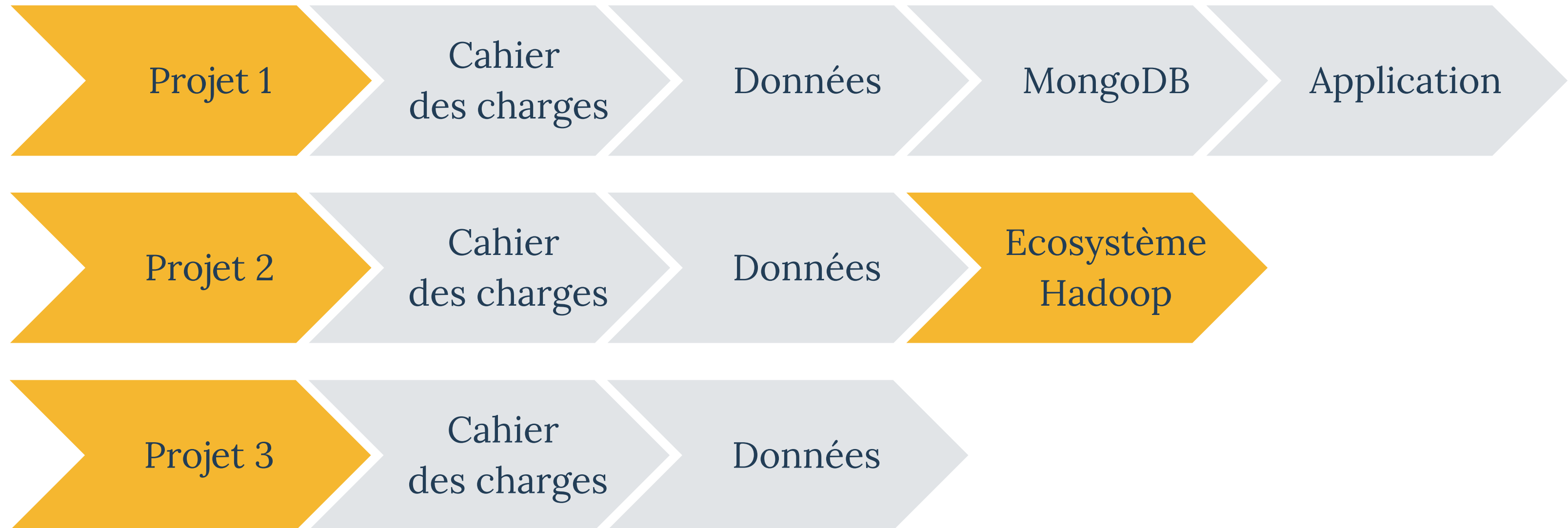
Plan



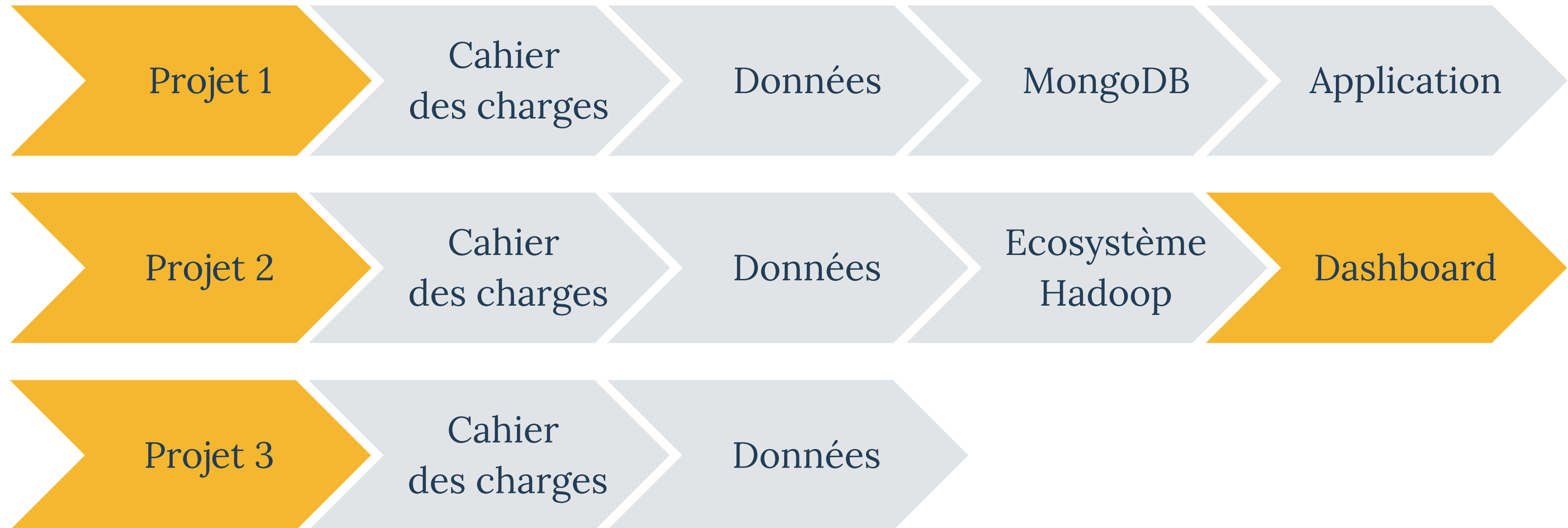
Plan



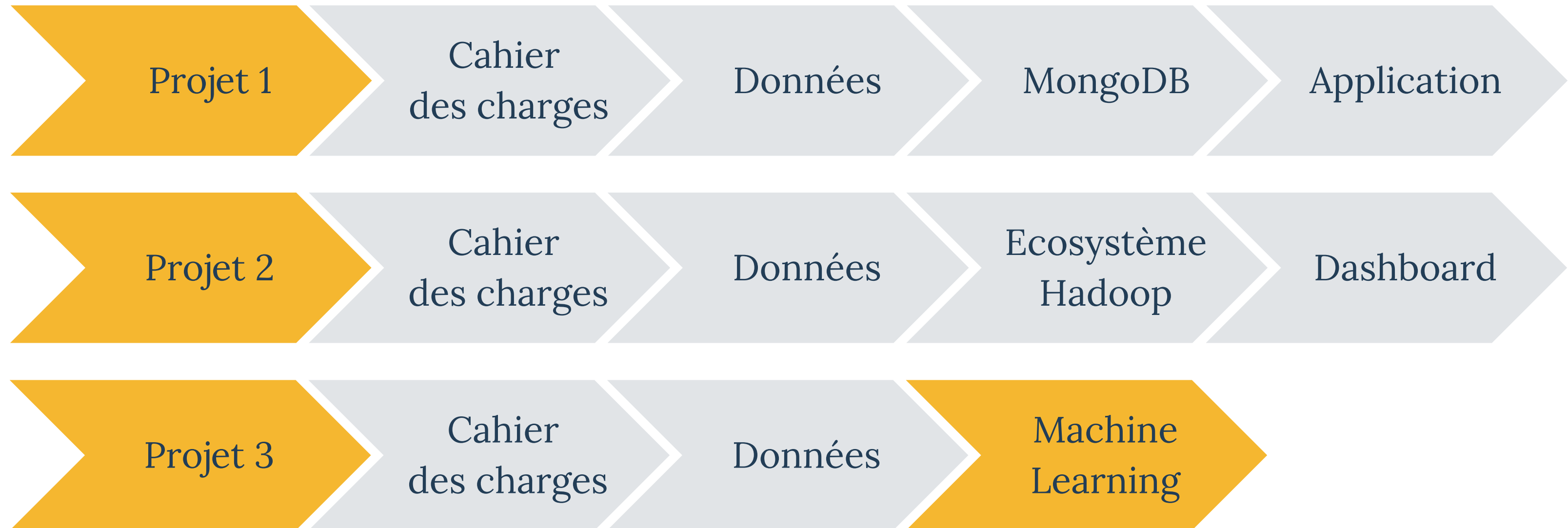
Plan



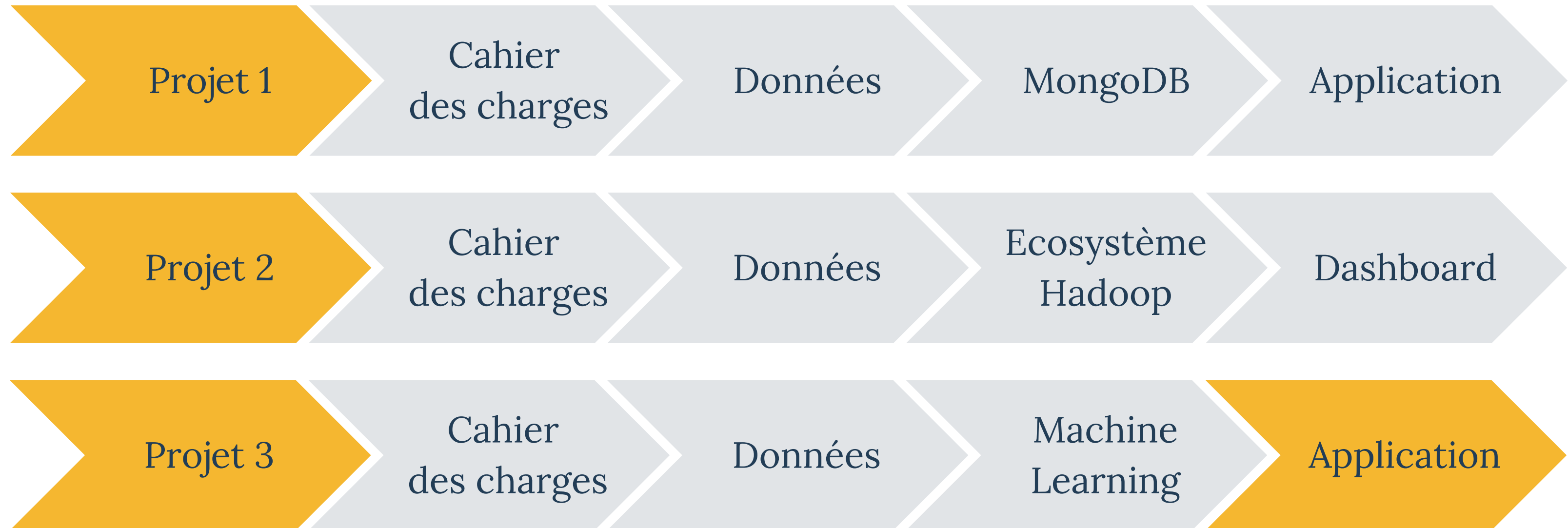
Plan



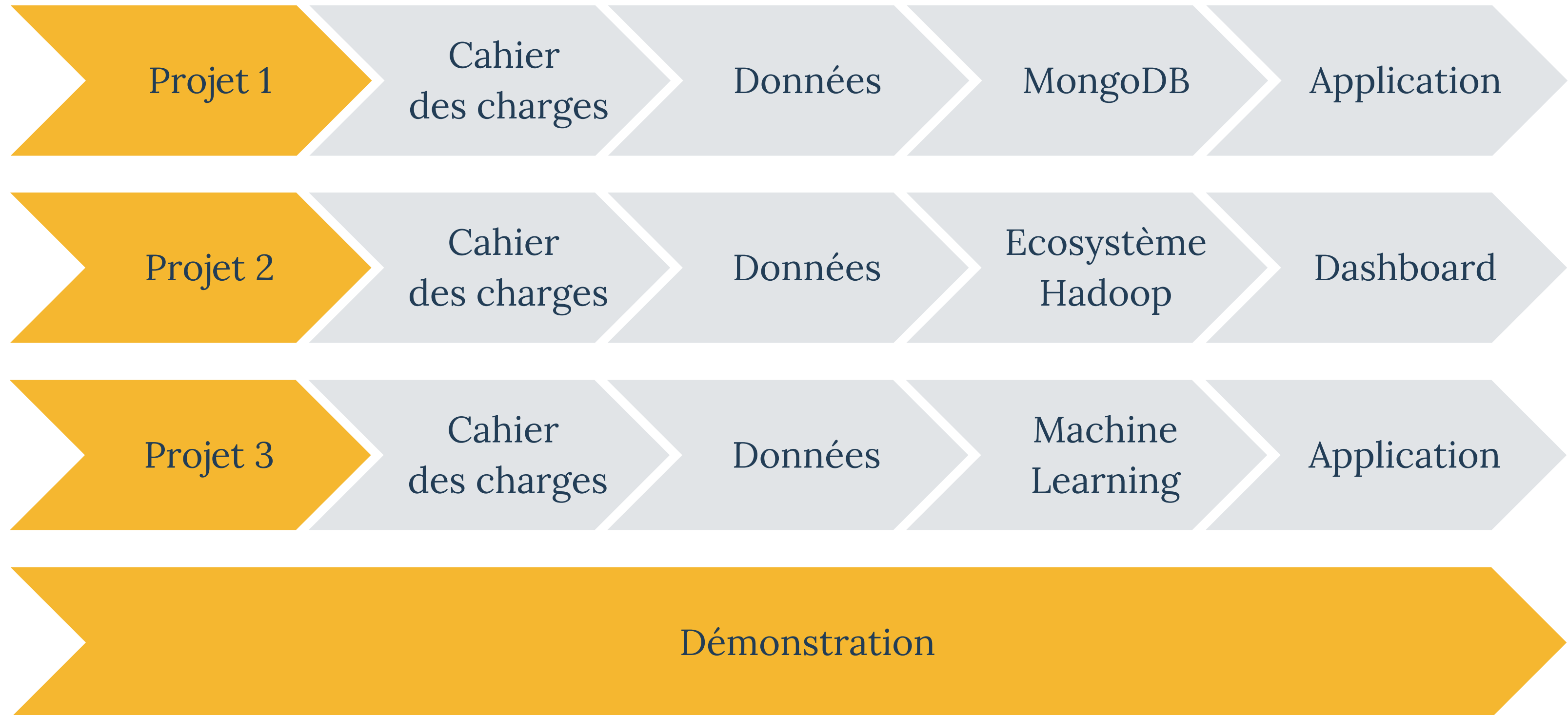
Plan



Plan



Plan





Projet n°1

“Développement d'une application de traitement de données NoSQL avec MongoDB et Python”

Cahier des charges

“Développement d'une application de traitement de données NoSQL avec MongoDB et Python.”

Etapes de réalisation du projet

1 - Contraintes générales

Utilisation de python OO dans un environnement virtuel.
venv, 1 classe

Cahier des charges

“Développement d'une application de traitement de données NoSQL avec MongoDB et Python.”

Etapes de réalisation du projet

2 - Analyse et nettoyage des données

Exploration des données et traitement des valeurs le nécessitant.

(Dans notre cas, jeu parfait)

Cahier des charges

“Développement d'une application de traitement de données NoSQL avec MongoDB et Python.”

Etapes de réalisation du projet

3 - Utilisation d'une base NoSQL sous MongoDB

setup.py : création et remplissage de la base

Connexions : singleton

Requêtes d'agrégation

Vues

Cahier des charges

“Développement d'une application de traitement de données NoSQL avec MongoDB et Python.”

Etapes de réalisation du projet

4 - Interface graphique

TKinter ou CLI -> TKinter

Deux onglets et graphes en pop-up

Cahier des charges

“Développement d'une application de traitement de données NoSQL avec MongoDB et Python.”

Etapes de réalisation du projet

5 - Filtres et export

Filtres de sélection des attributs (formulaire)

Bouton d'export csv / xlsx

Données

Proviennent du CDC et constituent une partie importante du Système de Surveillance des Facteurs de Risque Environnementaux (BRFSS), qui mène des enquêtes téléphoniques annuelles pour recueillir des données sur l'état de santé des résidents américains.

Données

Proviennent du CDC et constituent une partie importante du Système de Surveillance des Facteurs de Risque Environnementaux (BRFSS), qui mène des enquêtes téléphoniques annuelles pour recueillir des données sur l'état de santé des résidents américains.

HeartDisease	AgeCategory
BMI	Race
Smoking	Diabetic
AlcoholDrinking	PhysicalActivity
Stroke	GenHealth
PhysicalHealth	SleepTime
MentalHealth	Asthma
DiffWalking	KidneyDisease
Sex	SkinCancer

Données

Proviennent du CDC et constituent une partie importante du Système de Surveillance des Facteurs de Risque Environnementaux (BRFSS), qui mène des enquêtes téléphoniques annuelles pour recueillir des données sur l'état de santé des résidents américains.



Données

Proviennent du CDC et constituent une partie importante du Système de Surveillance des Facteurs de Risque Environnementaux (BRFSS), qui mène des enquêtes téléphoniques annuelles pour recueillir des données sur l'état de santé des résidents américains.



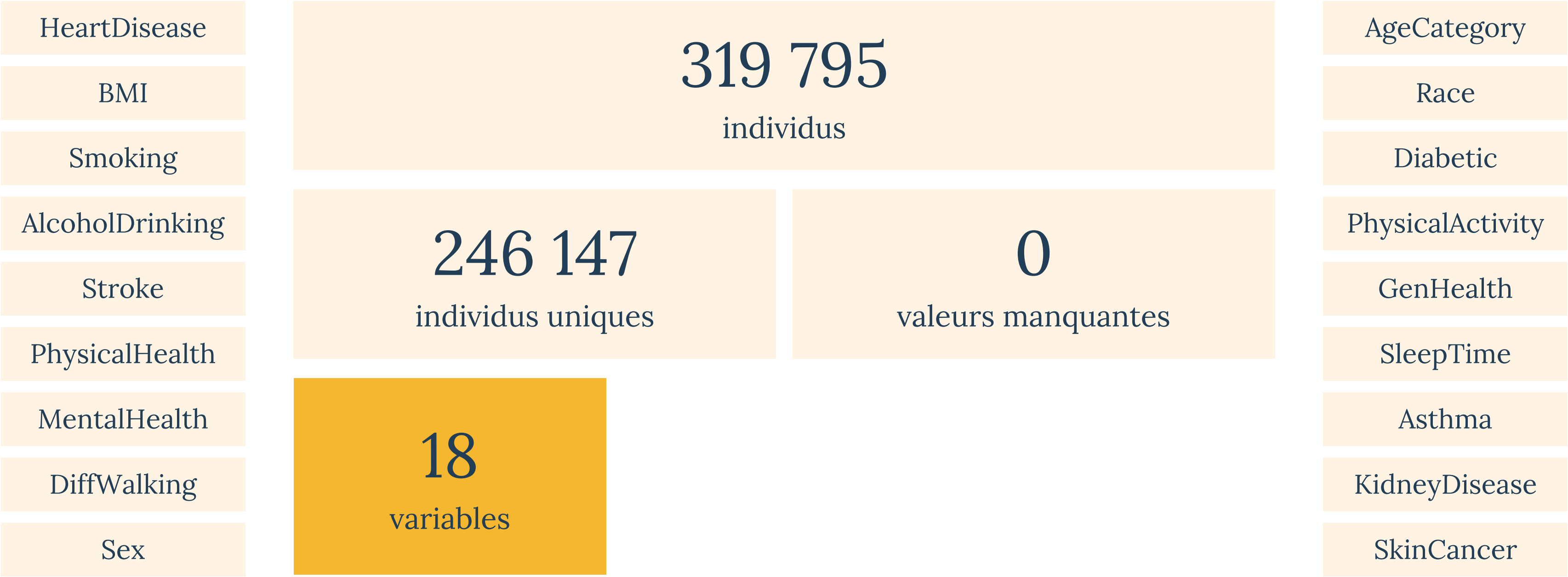
Données

Proviennent du CDC et constituent une partie importante du Système de Surveillance des Facteurs de Risque Environnementaux (BRFSS), qui mène des enquêtes téléphoniques annuelles pour recueillir des données sur l'état de santé des résidents américains.

HeartDisease	319 795		AgeCategory
BMI	individus		Race
Smoking			Diabetic
AlcoholDrinking	246 147	0	PhysicalActivity
Stroke	individus uniques	valeurs manquantes	GenHealth
PhysicalHealth			SleepTime
MentalHealth			Asthma
DiffWalking			KidneyDisease
Sex			SkinCancer

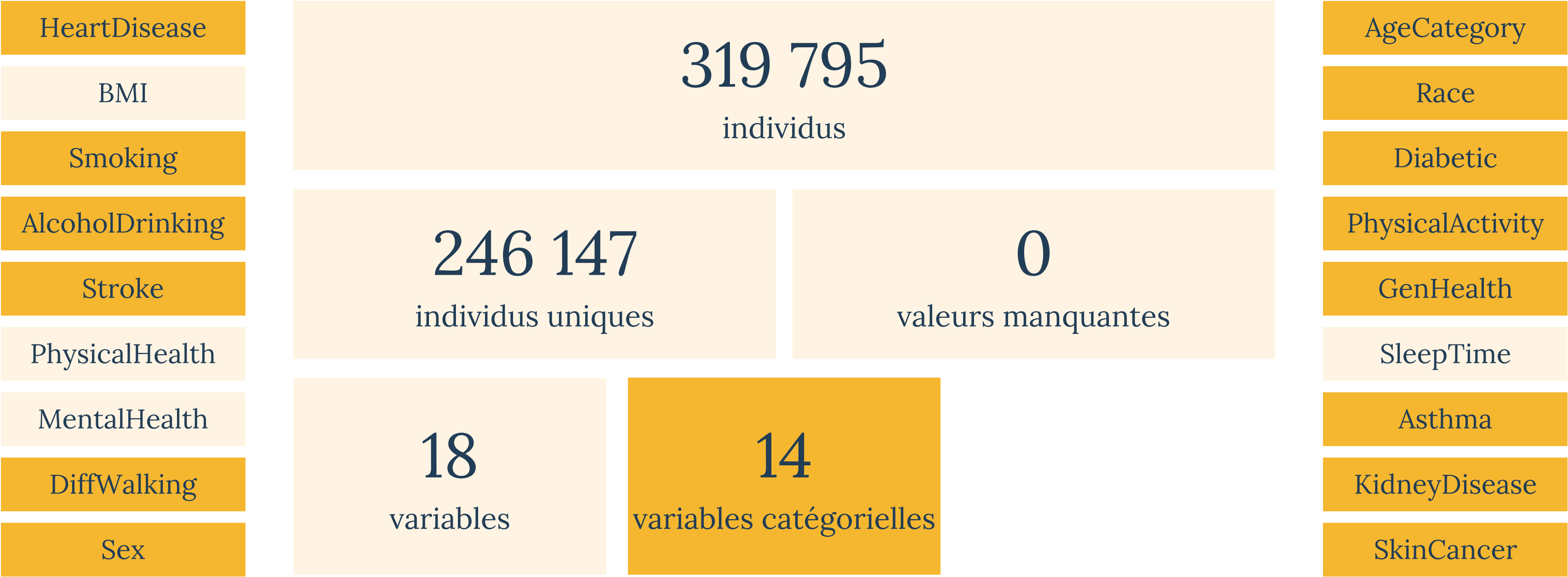
Données

Proviennent du CDC et constituent une partie importante du Système de Surveillance des Facteurs de Risque Environnementaux (BRFSS), qui mène des enquêtes téléphoniques annuelles pour recueillir des données sur l'état de santé des résidents américains.



Données

Proviennent du CDC et constituent une partie importante du Système de Surveillance des Facteurs de Risque Environnementaux (BRFSS), qui mène des enquêtes téléphoniques annuelles pour recueillir des données sur l'état de santé des résidents américains.



Données

Proviennent du CDC et constituent une partie importante du Système de Surveillance des Facteurs de Risque Environnementaux (BRFSS), qui mène des enquêtes téléphoniques annuelles pour recueillir des données sur l'état de santé des résidents américains.



MongoDB



Prétraitement

MongoDB

Prétraitement

Typage des attributs numériques en “int” afin d’homogénéiser au sein de la base de données, les types des attributs numériques

MongoDB

Prétraitement

Typage des attributs numériques en “int” afin d’homogénéiser au sein de la base de données, les types des attributs numériques

Base
de données

MongoDB

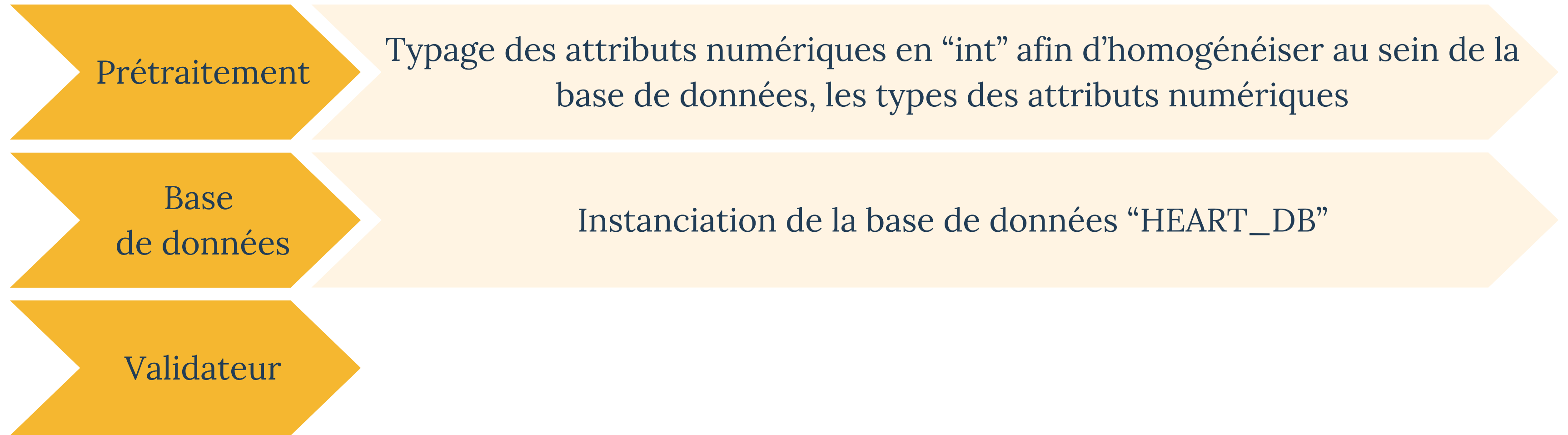
Prétraitement

Typage des attributs numériques en “int” afin d’homogénéiser au sein de la base de données, les types des attributs numériques

Base
de données

Instanciation de la base de données “HEART_DB”

MongoDB



MongoDB

Prétraitement

Typage des attributs numériques en “int” afin d’homogénéiser au sein de la base de données, les types des attributs numériques

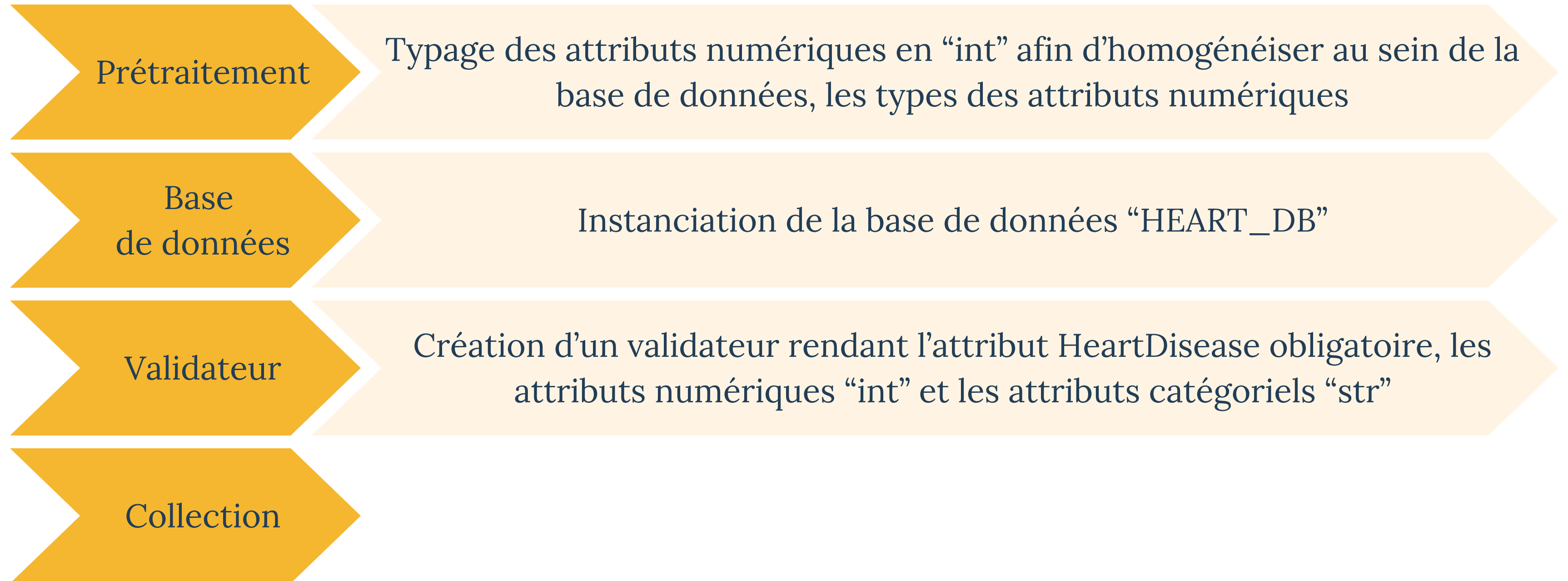
Base de données

Instanciation de la base de données “HEART_DB”

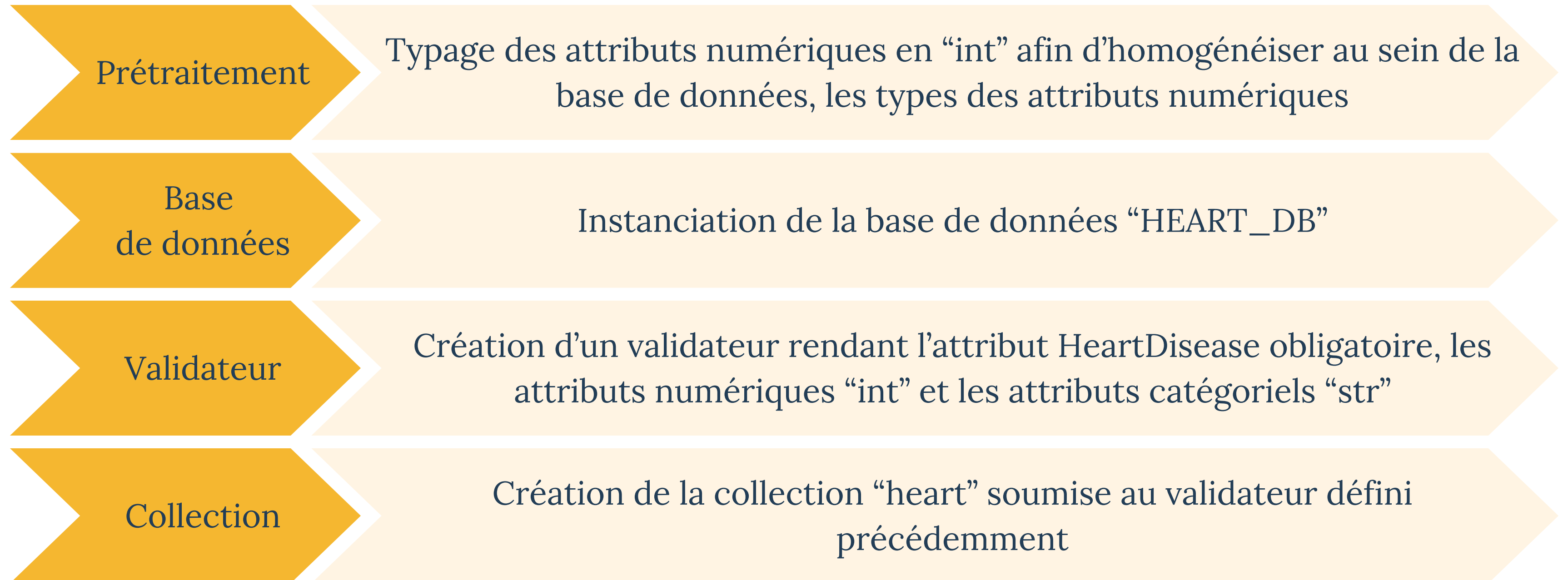
Valdateur

Création d’un validateur rendant l’attribut HeartDisease obligatoire, les attributs numériques “int” et les attributs catégoriels “str”

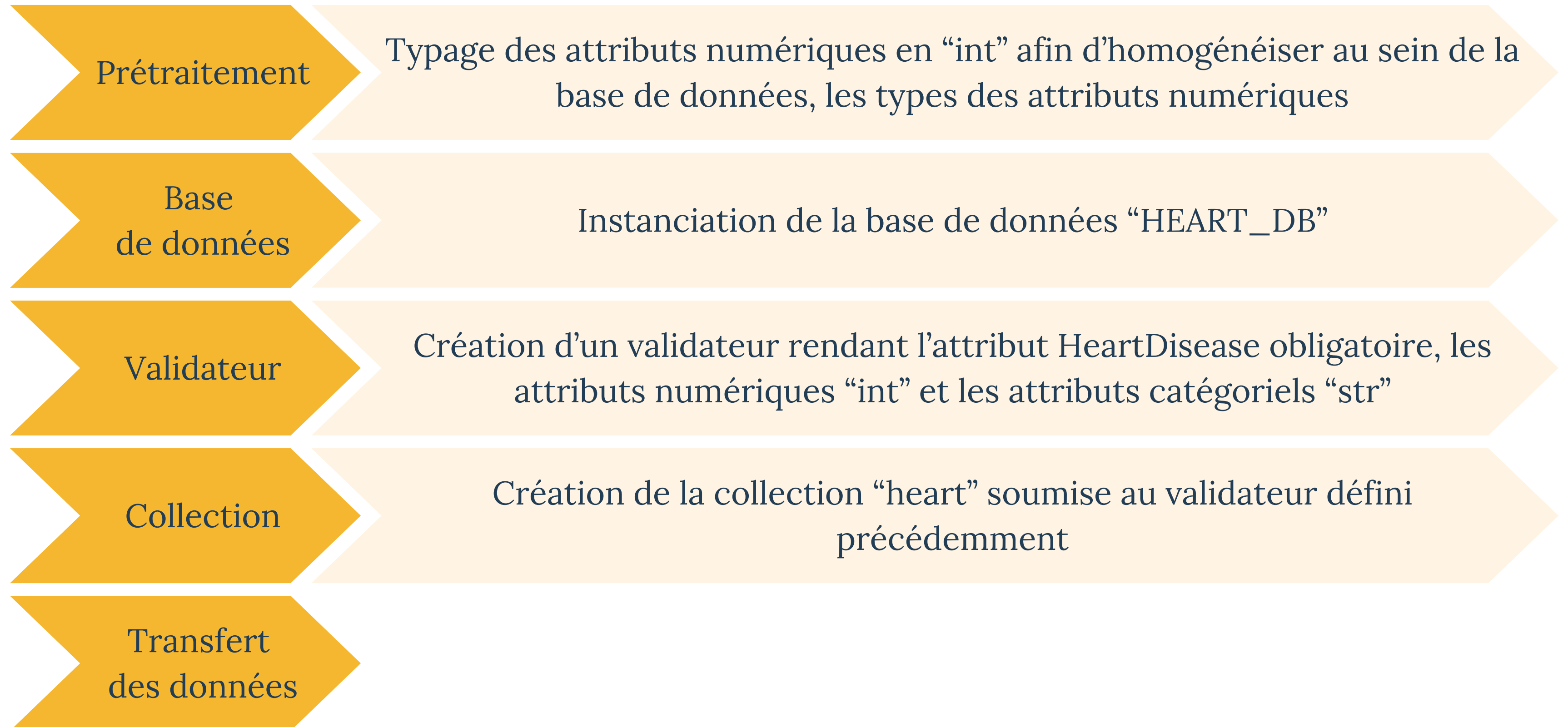
MongoDB



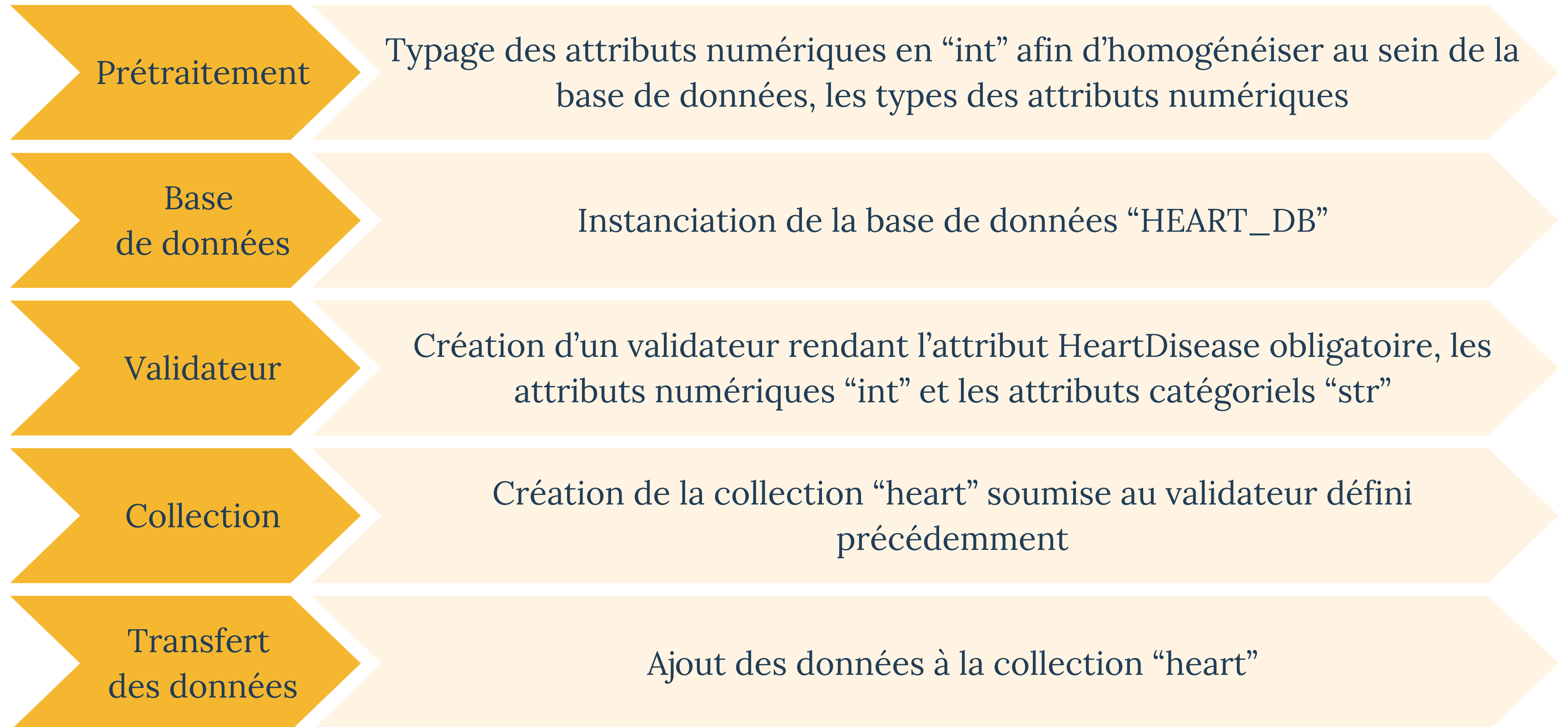
MongoDB



MongoDB



MongoDB



Application

Affichage de la liste des individus présents en base de données correspondant aux critères renseignés dans le formulaire, de leur répartition au sein de chaque attributs et de mesures pertinentes.

Remplissage du formulaire

Interrogation de la base de données

Affichage des résultats

Heart disease :	<input type="text"/>
Smoking :	<input type="text"/>
Alcohol drinking :	<input type="text"/>
Stroke :	<input type="text"/>
Walking difficulty :	<input type="text"/>
Sex :	<input type="text"/>
Age category :	<input type="text"/>
Ethnicity :	<input type="text"/>
Diabetic :	<input type="text"/>
Physical activity :	<input type="text"/>
General health :	<input type="text"/>
Asthma :	<input type="text"/>
Kidney disease :	<input type="text"/>
Skin cancer :	<input type="text"/>
BMI :	<input type="text"/>
Physical health :	<input type="text"/>
Mental health :	<input type="text"/>
Sleep time :	<input type="text"/>

Récupération des individus
`get_patients_data()`

Génération de métriques
`disease_estimate()`

Graphiques
`singleGraphsGeneration()`



Projet n°2

“Conception et développement d'une solution de collecte, de stockage et traitement de données volumineuses et hétérogènes avec Hadoop”

Cahier des charges

“Mettre en œuvre des solutions de traitement et d'analyse de données à grande échelle en utilisant la plateforme Hadoop, de l'exploration initiale des données jusqu'à la création d'une application finale.”

Cahier des charges

“Mettre en œuvre des solutions de traitement et d'analyse de données à grande échelle en utilisant la plateforme Hadoop, de l'exploration initiale des données jusqu'à la création d'une application finale.”

Etapes de réalisation du projet

1 - Analyse exploratoire des données

Analyse exploratoire des données et résumé des
caractéristiques clés

Cahier des charges

“Mettre en œuvre des solutions de traitement et d'analyse de données à grande échelle en utilisant la plateforme Hadoop, de l'exploration initiale des données jusqu'à la création d'une application finale.”

Etapes de réalisation du projet

2 - Pré-traitement des données

Le pré-traitement prépare les données pour Hadoop en les nettoyant et les ajustant en fusionnant les tables

Cahier des charges

“Mettre en œuvre des solutions de traitement et d'analyse de données à grande échelle en utilisant la plateforme Hadoop, de l'exploration initiale des données jusqu'à la création d'une application finale.”

Etapes de réalisation du projet

3 - Implémentation MapReduce

Conception et implémentation des tâches Map et Reduce
pour le traitement distribué avec Hadoop

Cahier des charges

“Mettre en œuvre des solutions de traitement et d'analyse de données à grande échelle en utilisant la plateforme Hadoop, de l'exploration initiale des données jusqu'à la création d'une application finale.”

Etapes de réalisation du projet

4 - Implémentation HBase

Création et mise en place des fonctionnalités liées à
HBase

Cahier des charges

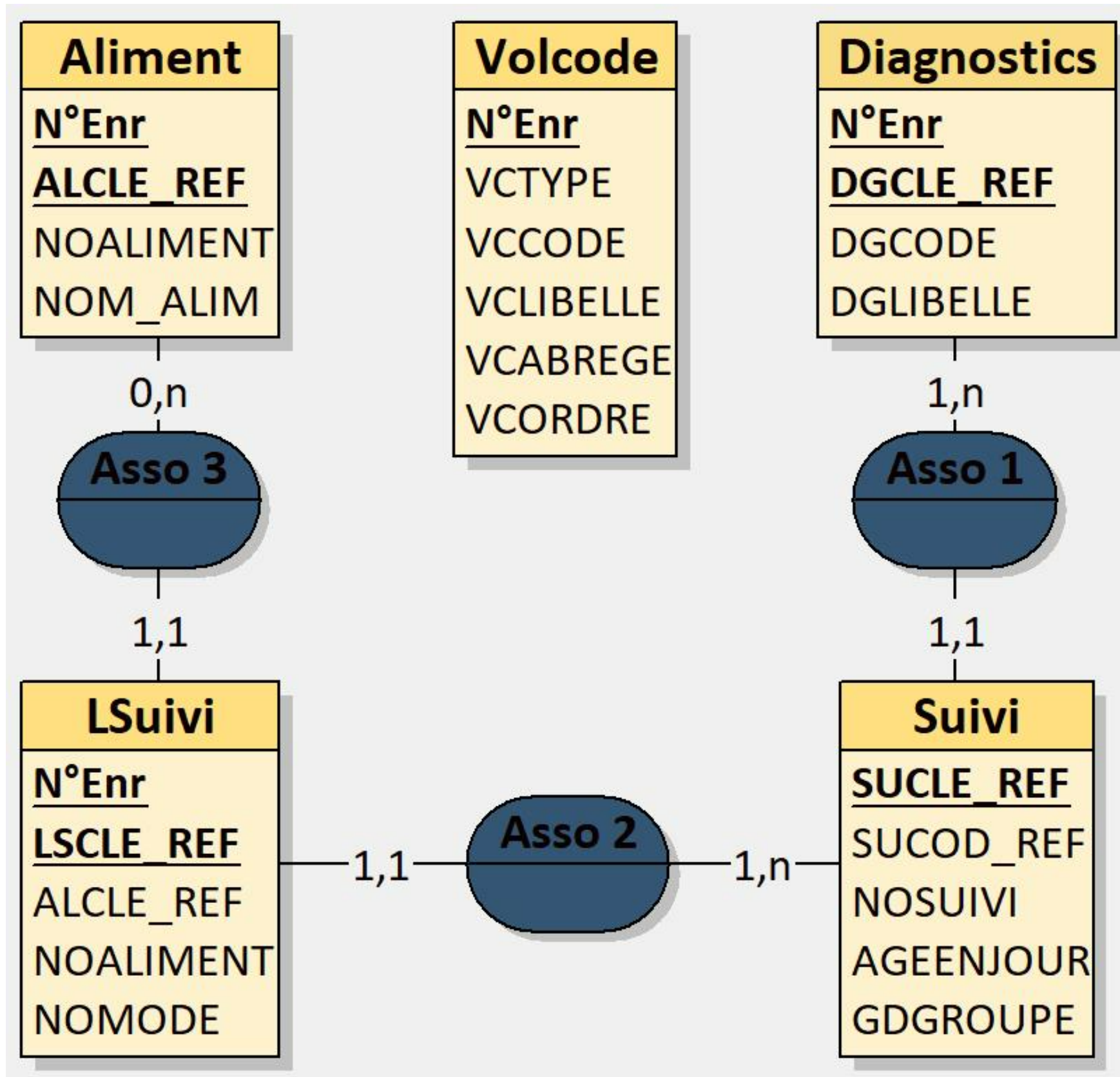
“Mettre en œuvre des solutions de traitement et d'analyse de données à grande échelle en utilisant la plateforme Hadoop, de l'exploration initiale des données jusqu'à la création d'une application finale.”

Etapes de réalisation du projet

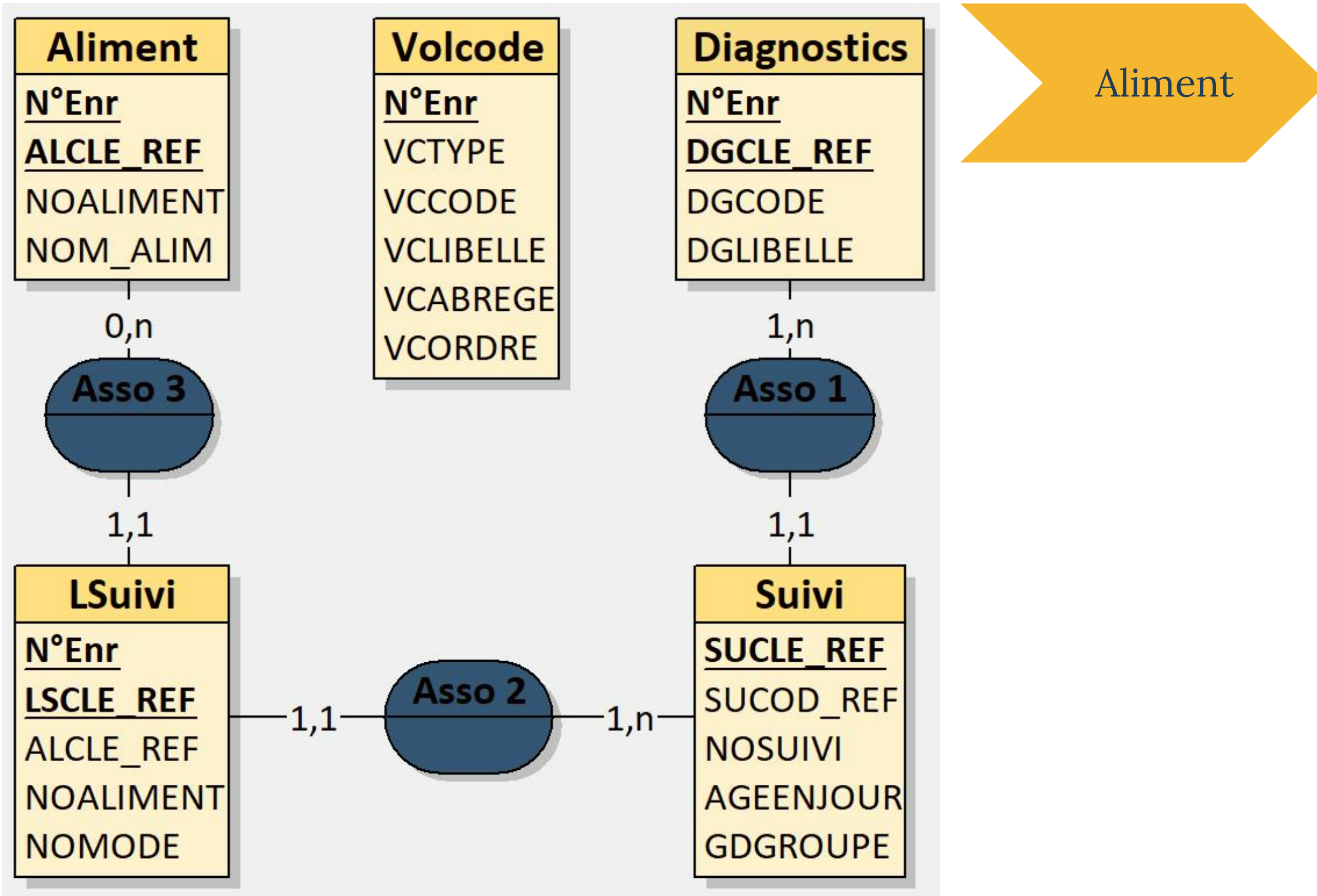
5 - Visualisation des résultats

Création de tableaux de bord à l'aide de Power BI pour présenter les résultats du traitement des données.

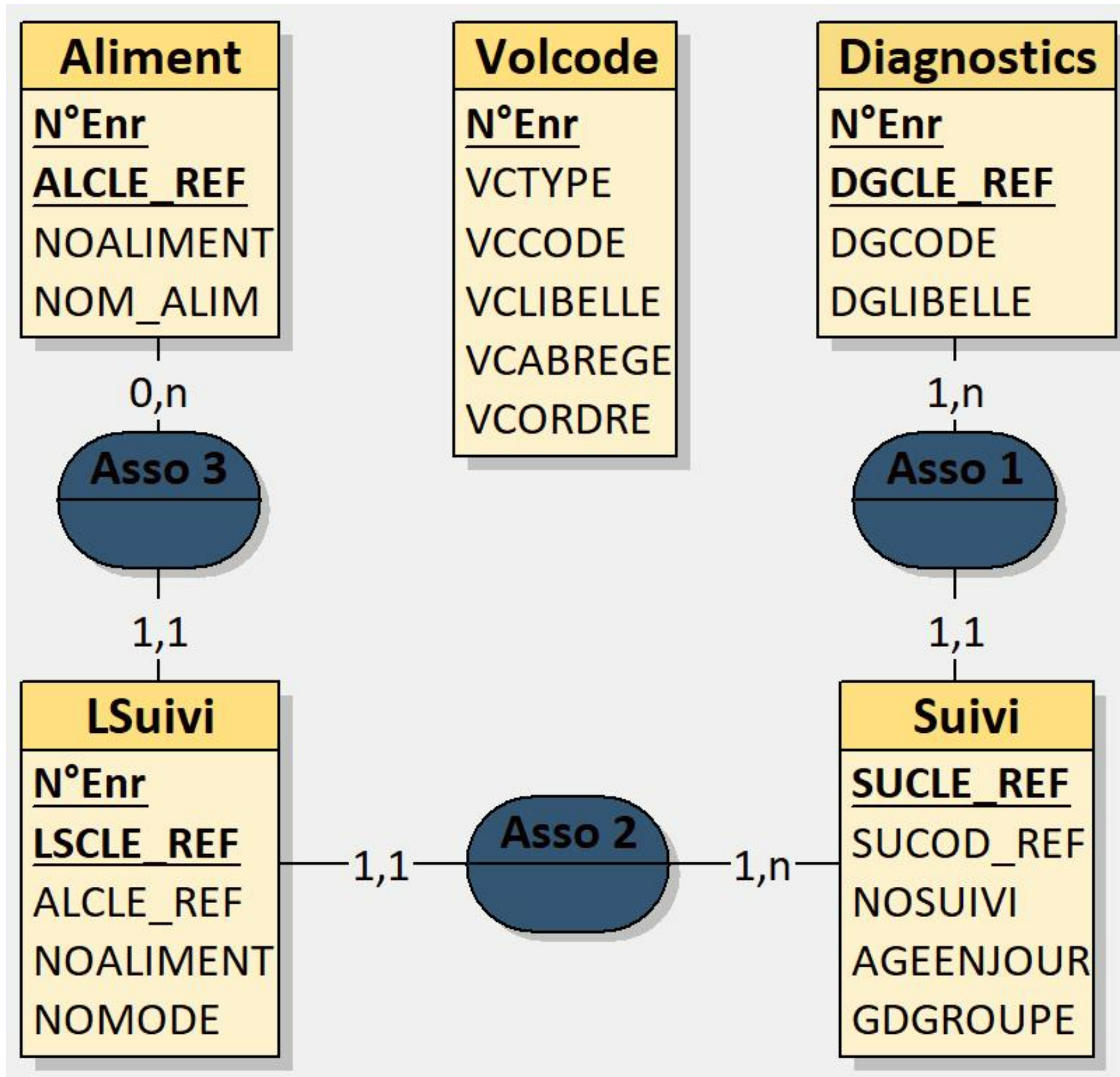
Données



Données



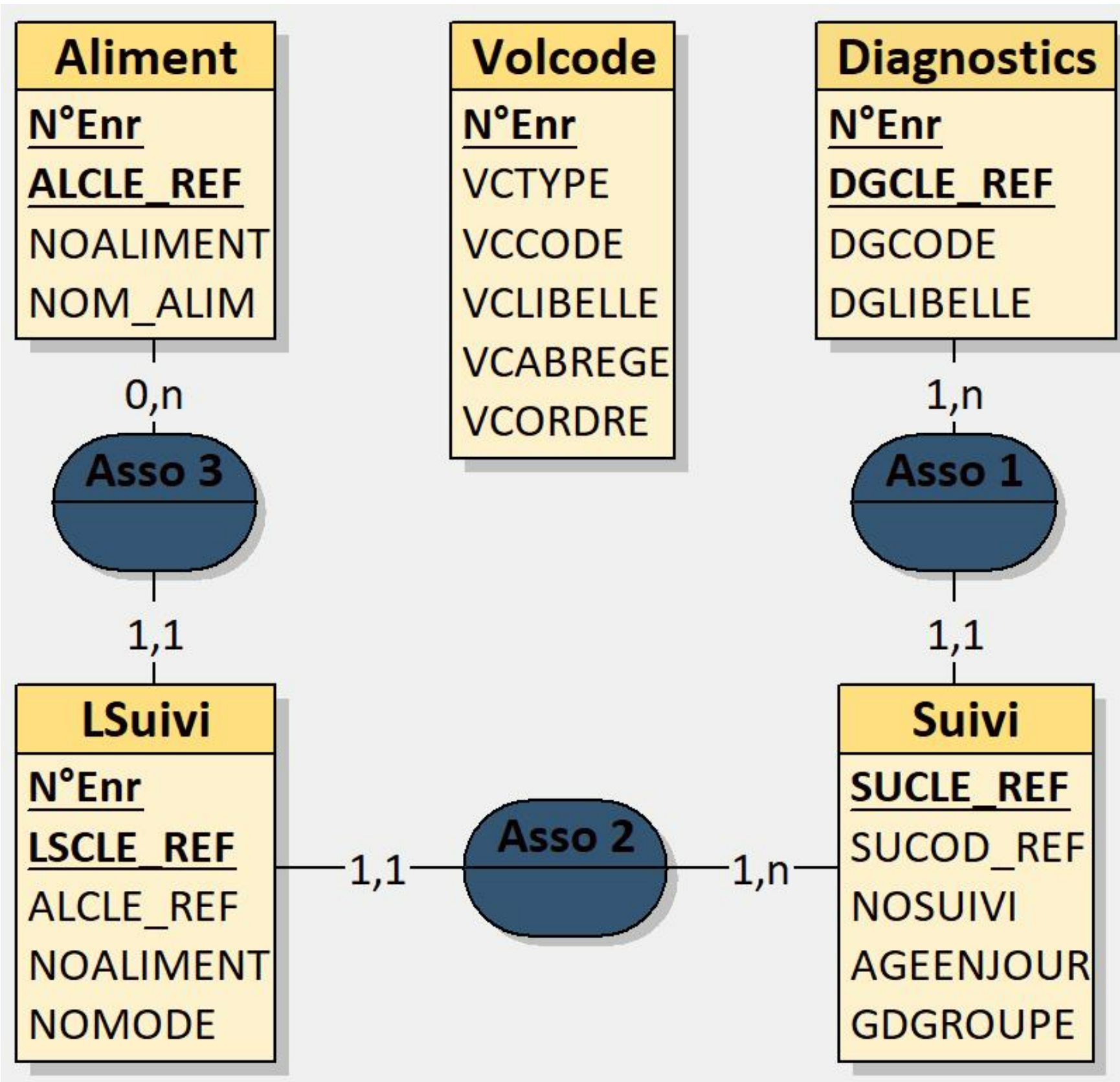
Données



Aliment

Table répertoriant les différents aliments qui peuvent être prescrits.

Données

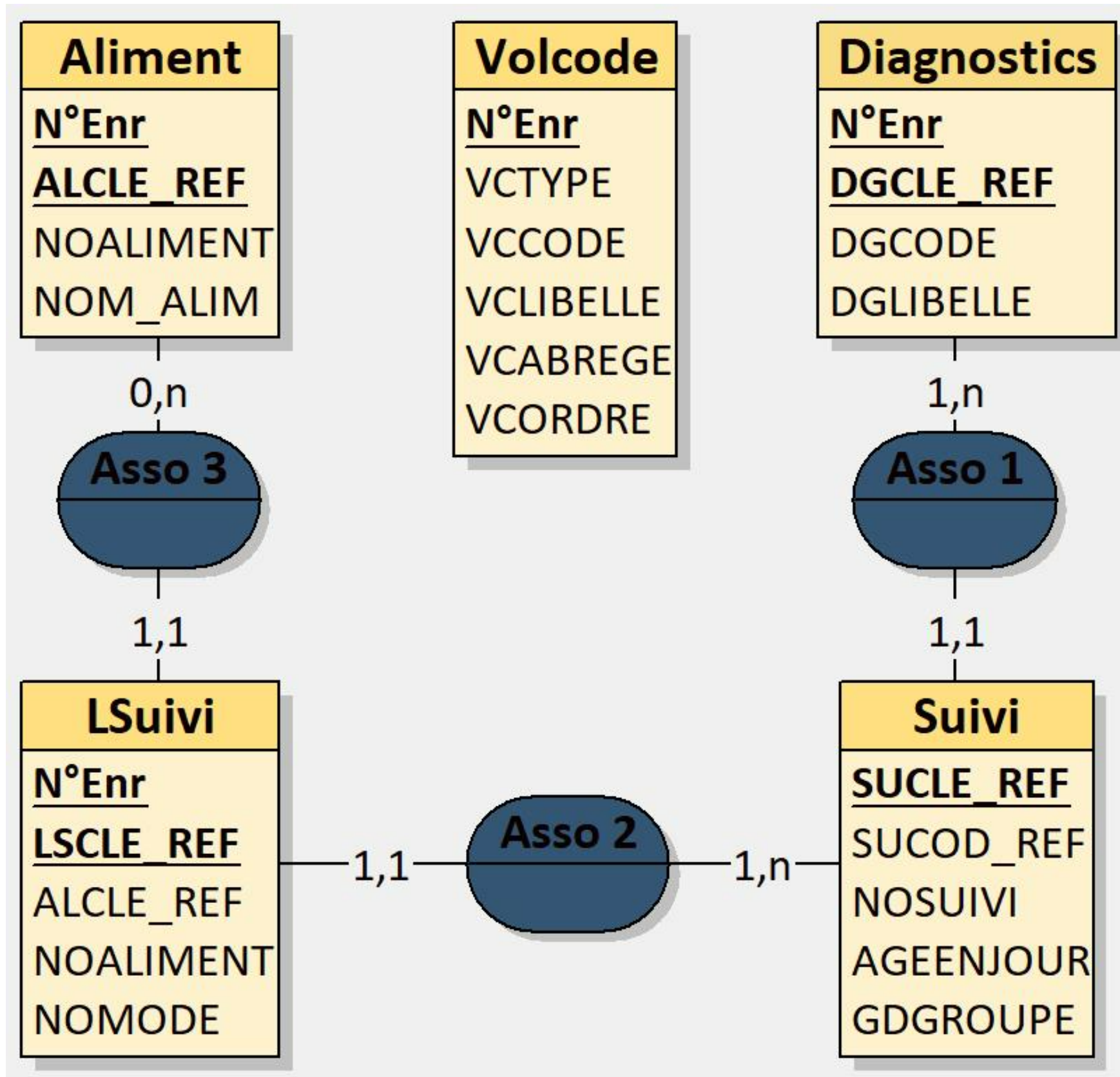


Aliment

Diagnostics

Table répertoriant les différents aliments qui peuvent être prescrits.

Données



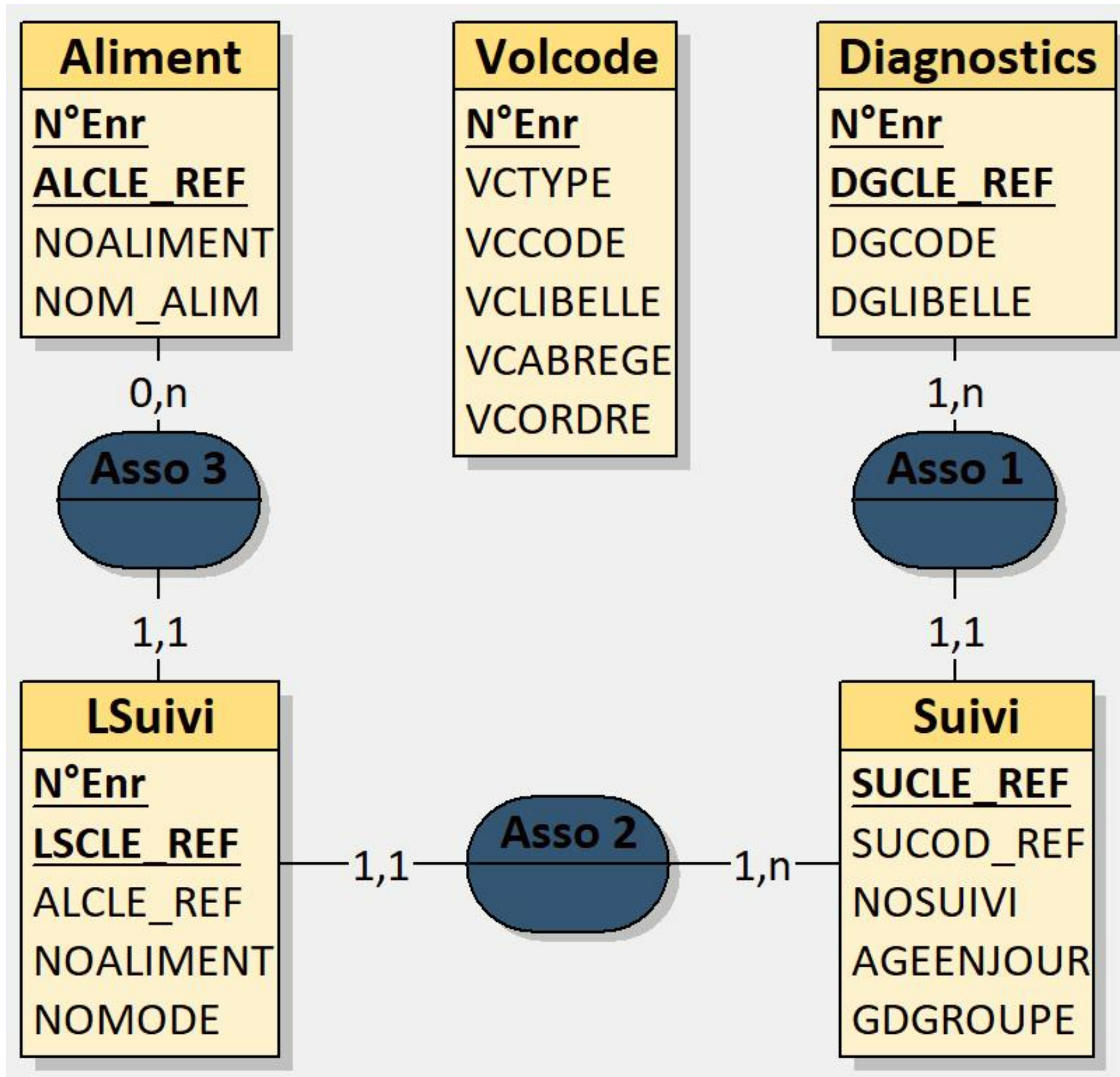
Aliment

Table répertoriant les différents aliments qui peuvent être prescrits.

Diagnostics

Table répertoriant les différents diagnostics pouvant être établis.

Données



Aliment

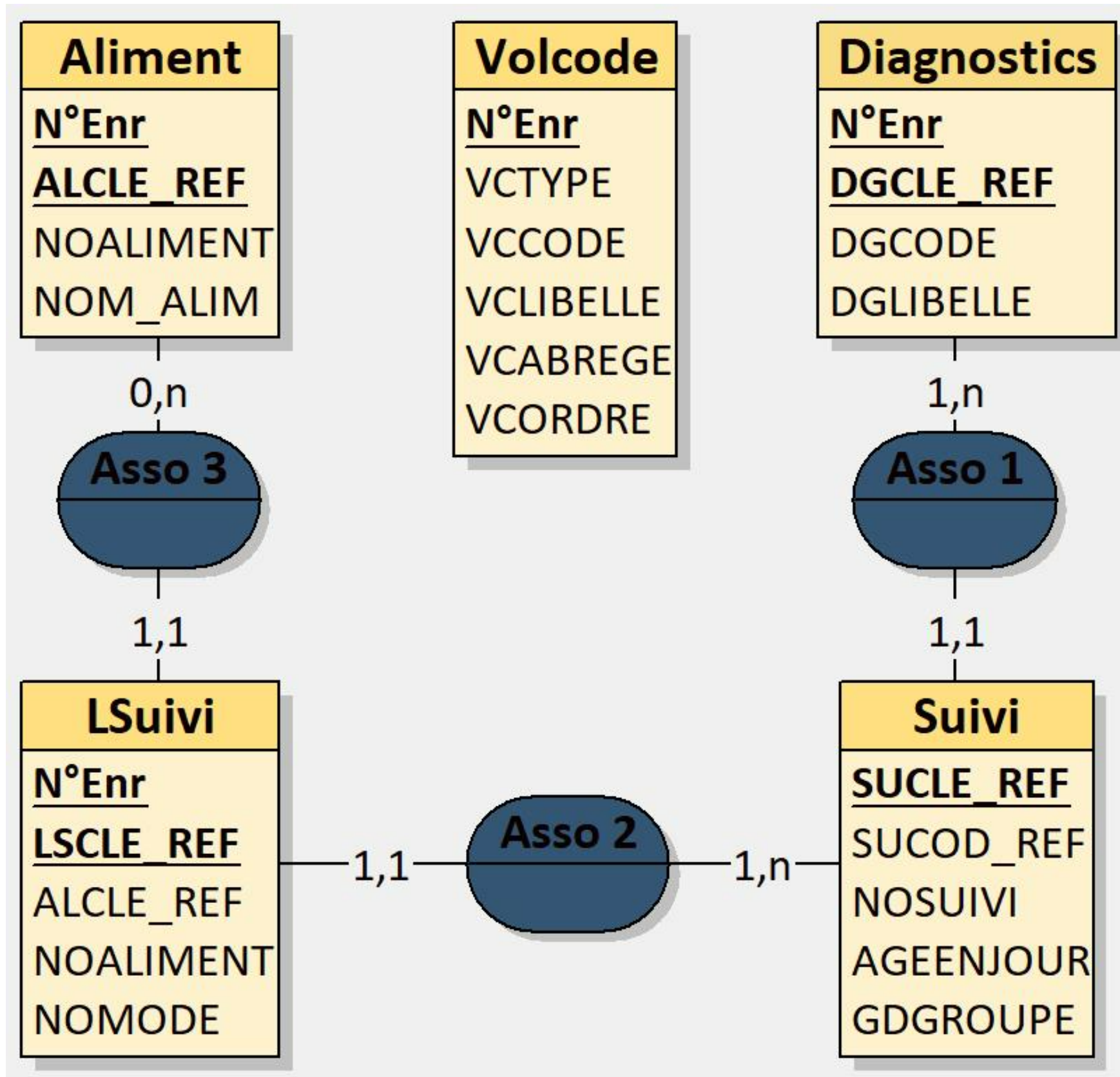
Table répertoriant les différents aliments qui peuvent être prescrits.

Diagnostics

Table répertoriant les différents diagnostics pouvant être établis.

Suivi

Données



Aliment

Table répertoriant les différents aliments qui peuvent être prescrits.

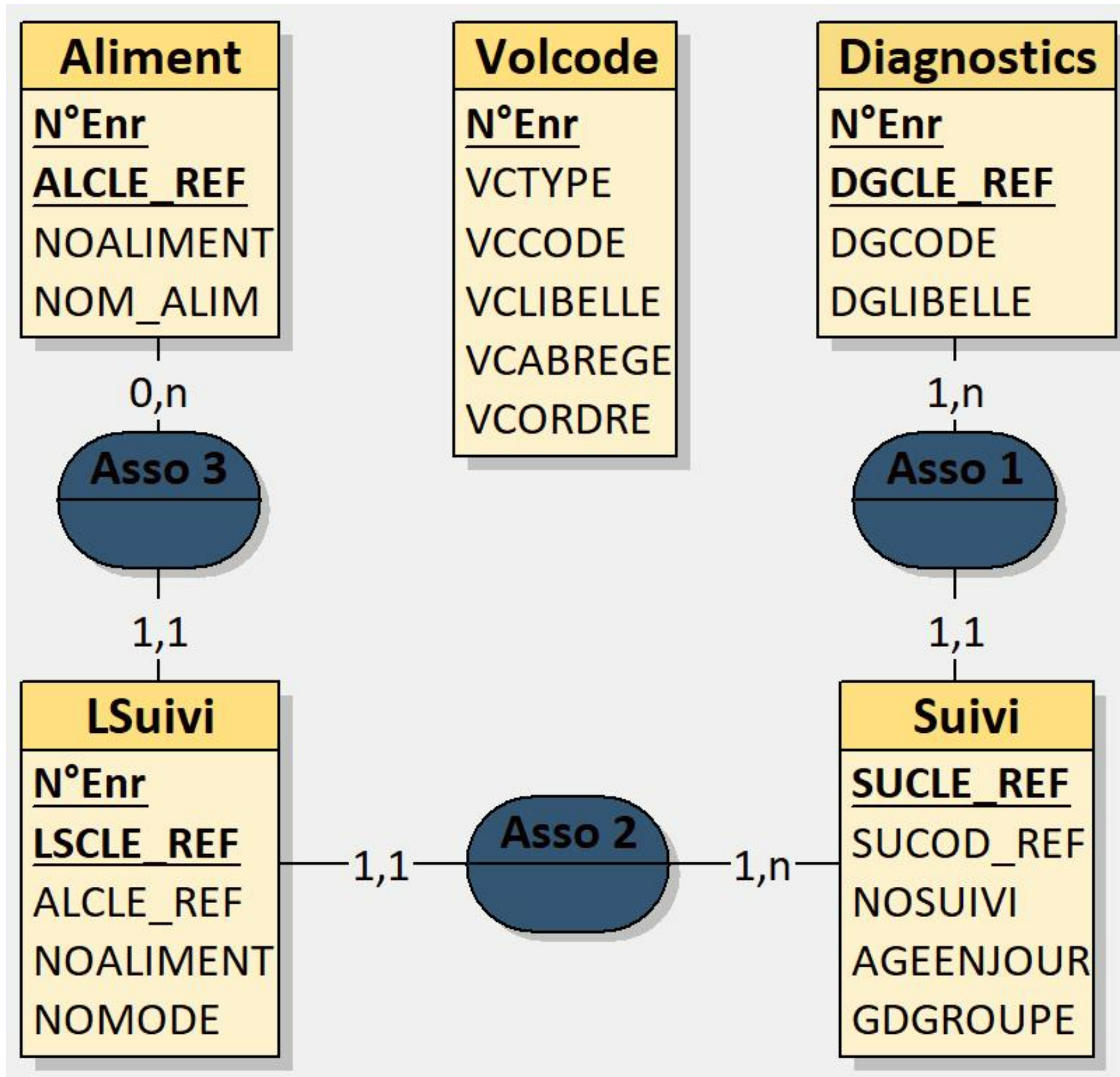
Diagnostics

Table répertoriant les différents diagnostics pouvant être établis.

Suivi

Table répertoriant les différents suivis effectués sur une période de temps donnée.

Données



Aliment

Table répertoriant les différents aliments qui peuvent être prescrits.

Diagnostics

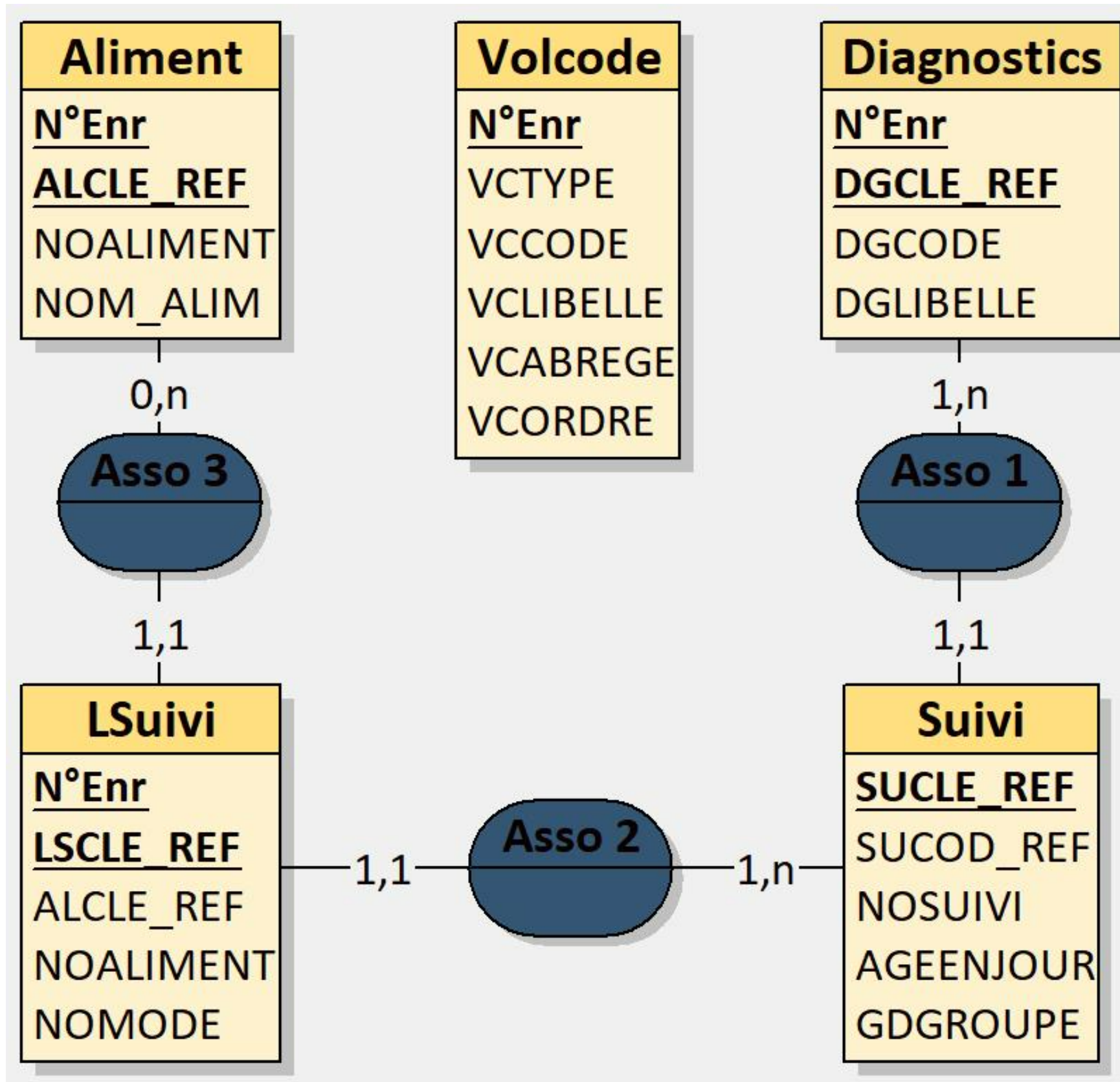
Table répertoriant les différents diagnostics pouvant être établis.

Suivi

Table répertoriant les différents suivis effectués sur une période de temps donnée.

Lsuivi

Données



Aliment

Table répertoriant les différents aliments qui peuvent être prescrits.

Diagnostics

Table répertoriant les différents diagnostics pouvant être établis.

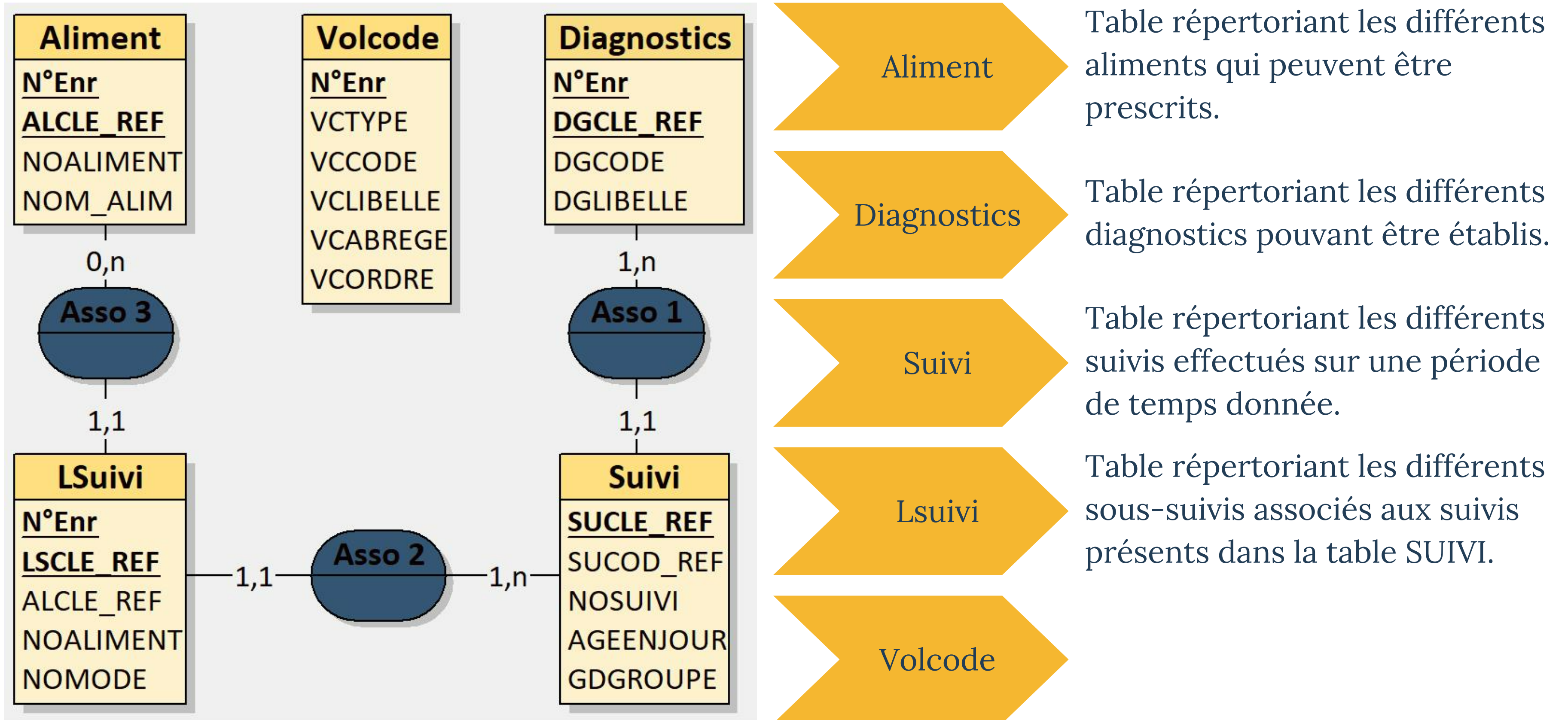
Suivi

Table répertoriant les différents suivis effectués sur une période de temps donnée.

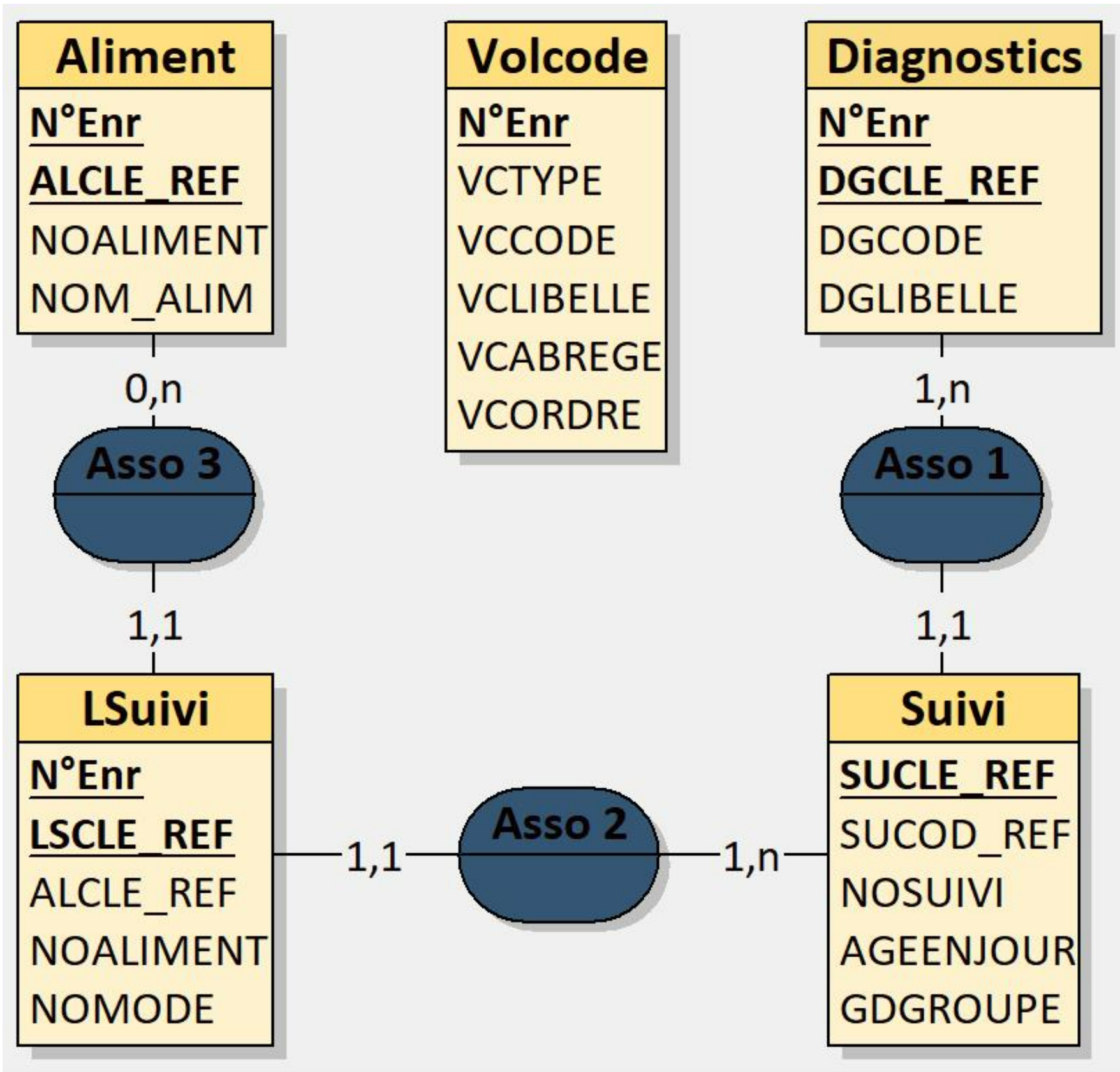
Lsuivi

Table répertoriant les différents sous-suivis associés aux suivis présents dans la table SUIVI.

Données



Données



- Aliment**: Table répertoriant les différents aliments qui peuvent être prescrits.
- Diagnostics**: Table répertoriant les différents diagnostics pouvant être établis.
- Suivi**: Table répertoriant les différents suivis effectués sur une période de temps donnée.
- Lsuivi**: Table répertoriant les différents sous-suivis associés aux suivis présents dans la table SUIVI.
- Volcode**: Table de contingence.

Ecosystème Hadoop

Calcul et stockage distribués

hadoop **HDFS**, **MapReduce** : traitement et insertion, base NoSQL **HBase**

Trois conteneurs **Docker** / un réseau : NameNode et deux Slaves

setup.ps1 :

- Envoi sur le NameNode : données, mapper et reducer
 - Envoi et lancement du script de configuration

1 ligne de code pour lancer le traitement

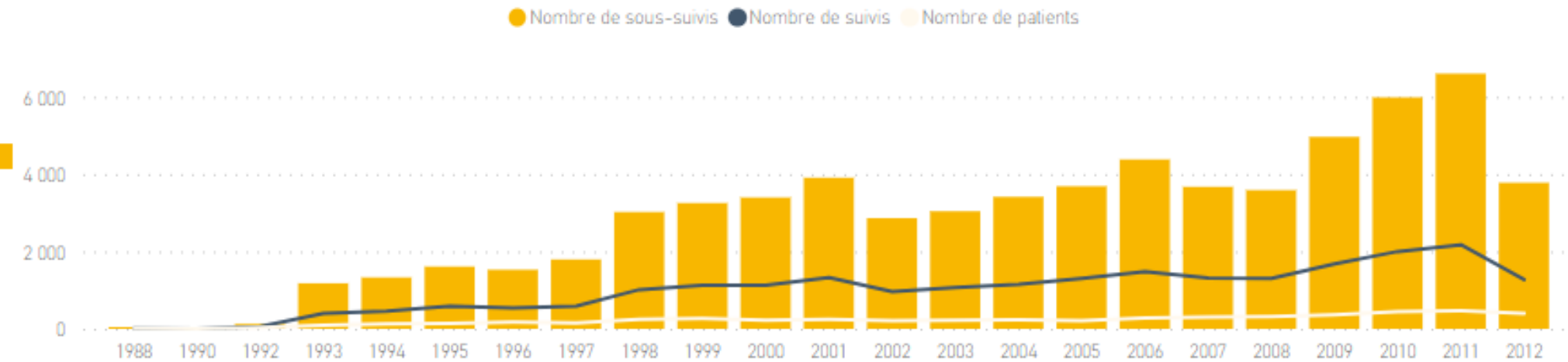
Dashboard



Yen Phi Do | Hugo Alpiste | Sébastien Martel | Morgane Geoffroy

Projet Big Data

Nombre de patients, de suivis et de sous-suivis par année



1673

Nombre total de patients

Population

Adulte

Enfant

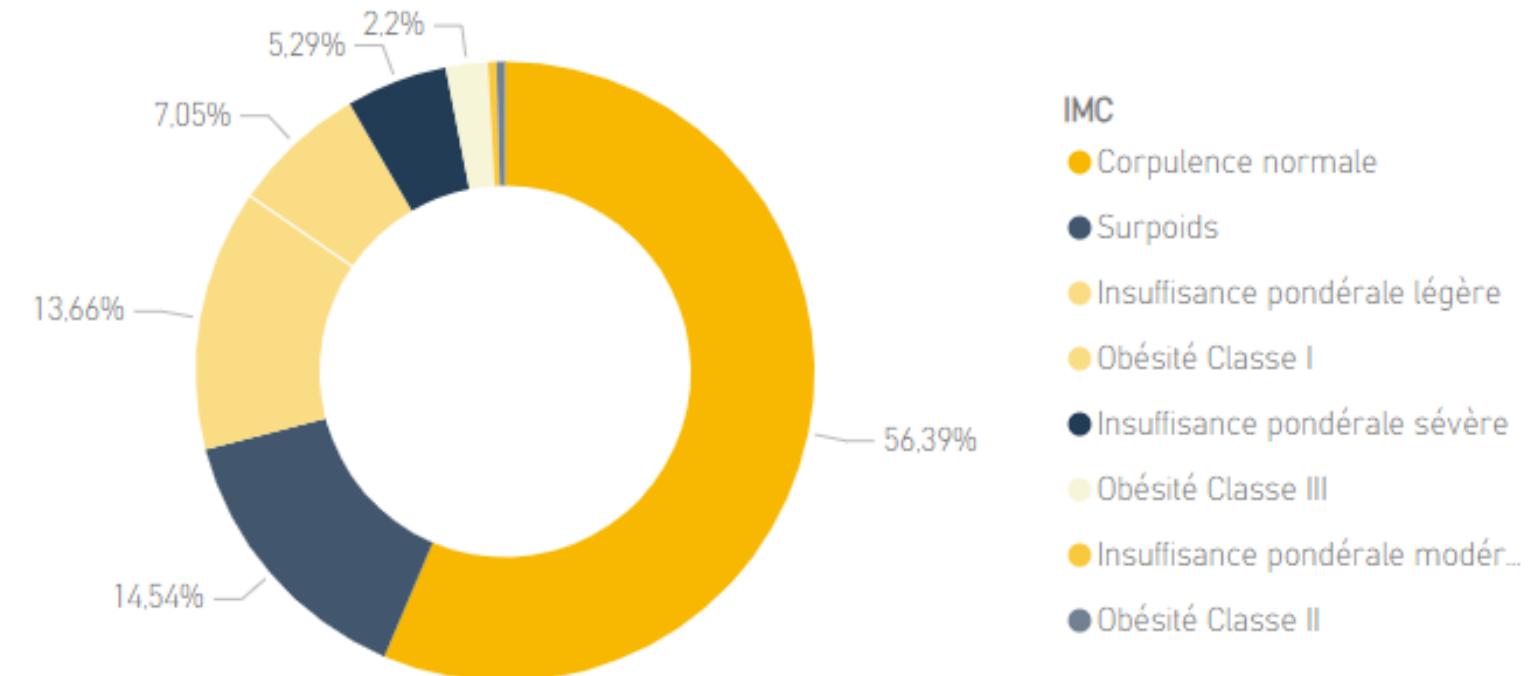
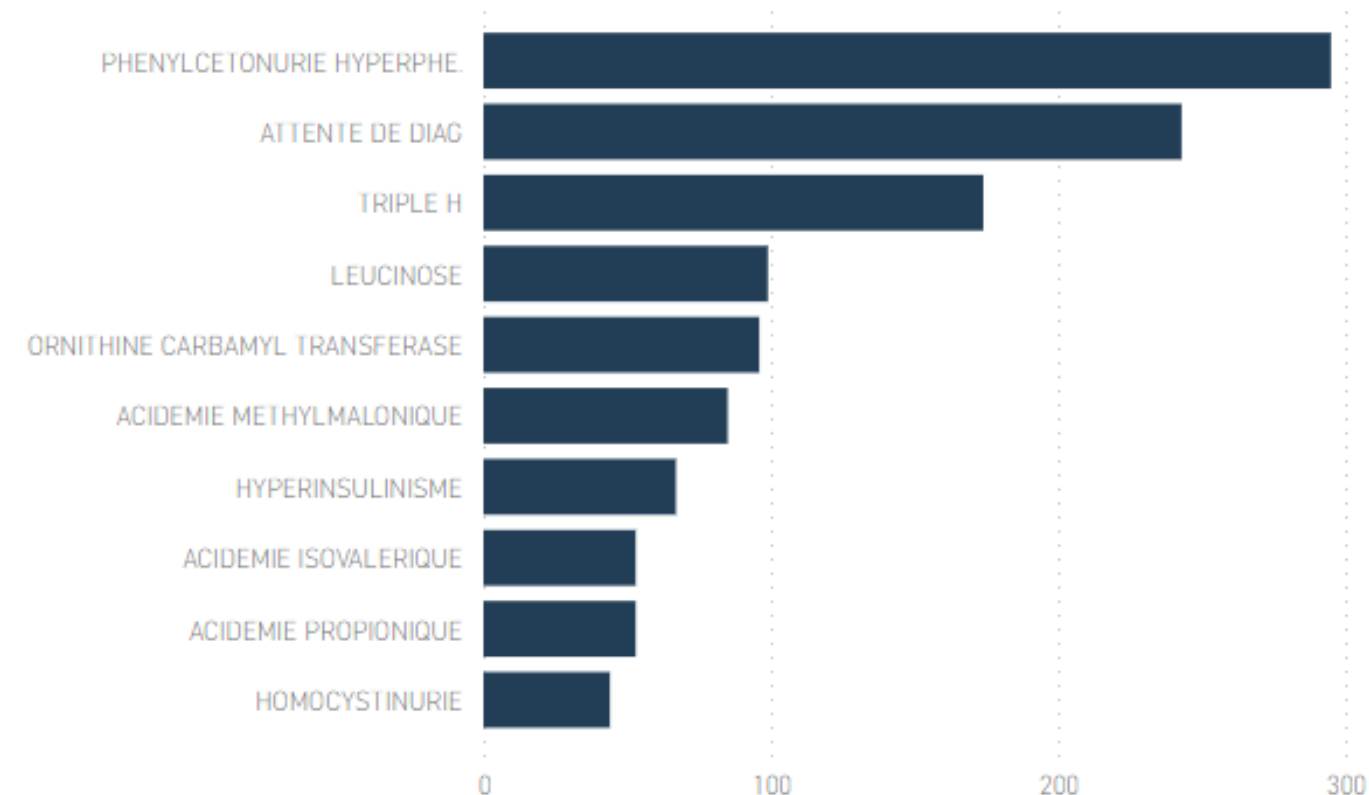
1673

Nombre ajusté de patients

243

En attente de diagnostic

Nombre de patients par diagnostic



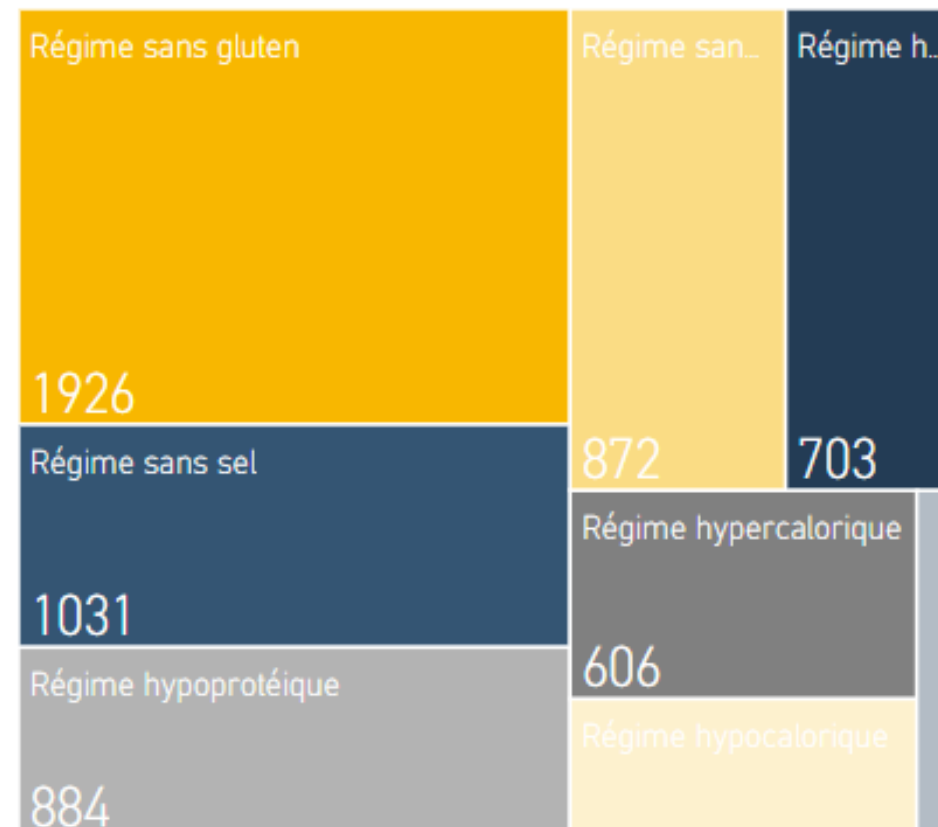
Dashboard



Yen Phi Do | Hugo Alpiste | Sébastien Martel | Morgane Geoffroy

Projet Big Data

Répartition des régimes prescrits

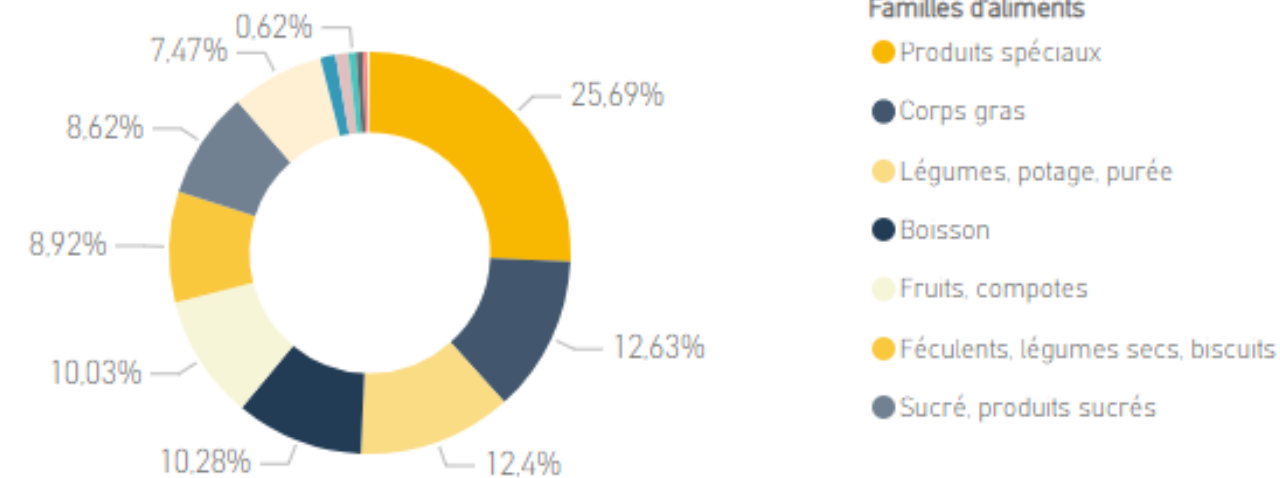


Nombre d'aliments par régime alimentaire



Focus sur le diagnostic de la Phénylcétonurie Hyperphe

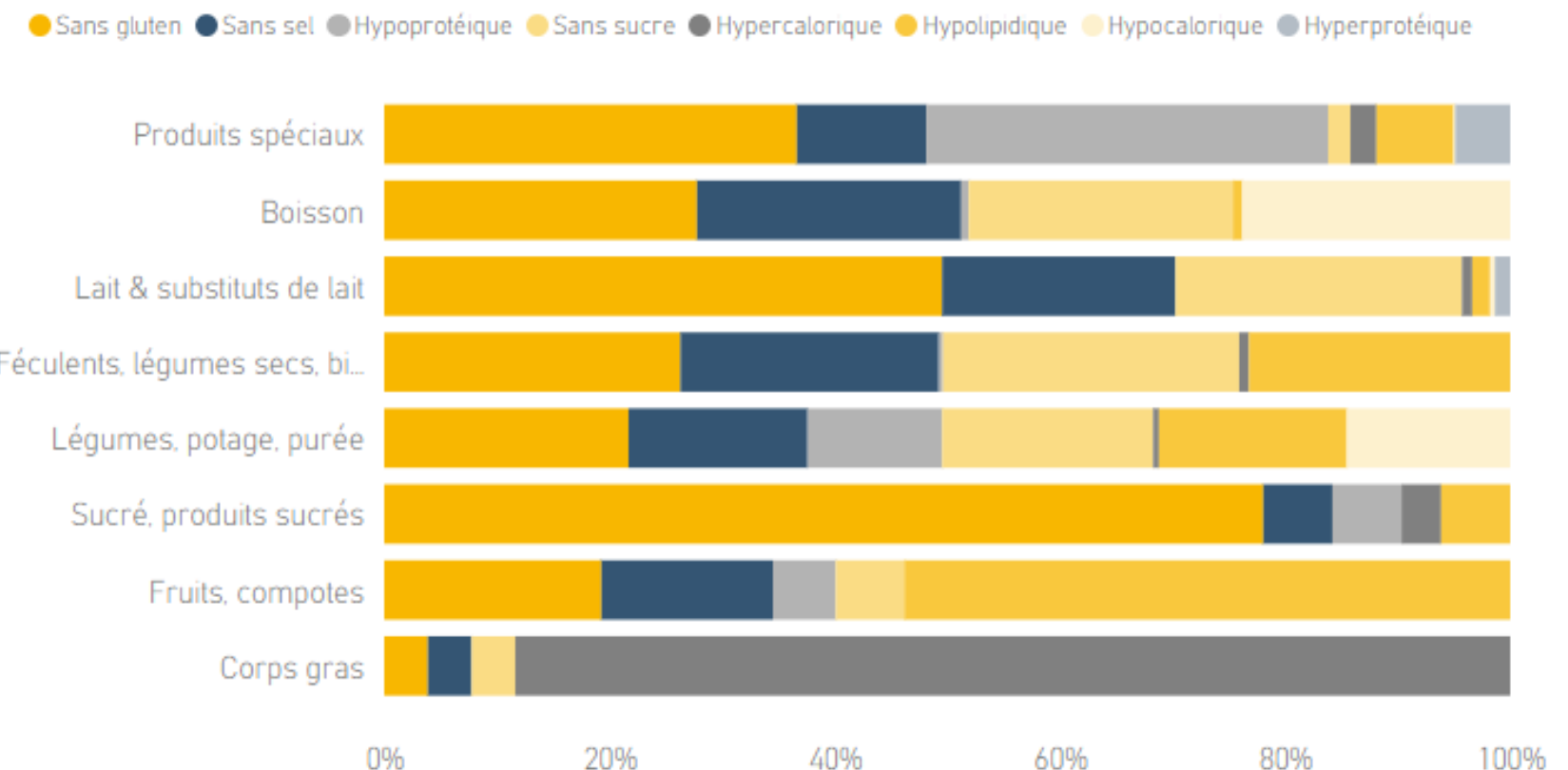
Répartition des familles d'aliments prescrites



6317
Nombre total de sous-suivis

6317
Nombre ajusté de sous-suivis

Répartition des régimes alimentaires par famille d'aliments





Projet n°3

“Concevoir un système d’intelligence artificielle et d’apprentissage automatique”

Cahier des charges

Prédire la survenue probable d'un incident cardiaque sur la base de réponses renseignées par le patient dans un questionnaire.

Etapes de réalisation du projet

1 - Analyse exploratoire

Exploration des données afin de pouvoir en évaluer la pertinence et résumer leurs principales caractéristiques.

Cahier des charges

Prédire la survenue probable d'un incident cardiaque sur la base de réponses renseignées par le patient dans un questionnaire.

Etapes de réalisation du projet

2 - Pré-traitement

Formatage du jeu de données initial, résultant de l'analyse exploratoire, dans le but de l'adapter à une utilisation dans des algorithmes de Machine Learning

Cahier des charges

Prédire la survenue probable d'un incident cardiaque sur la base de réponses renseignées par le patient dans un questionnaire.

Etapes de réalisation du projet

3 - Entraînement, optimisation et évaluation des performances de plusieurs modèles

Comparaison de trois modèles adaptés à la classification dont les hyperparamètres auront été optimisés et dont les performances auront été évaluées.

Cahier des charges

Prédire la survenue probable d'un incident cardiaque sur la base de réponses renseignées par le patient dans un questionnaire.

Etapes de réalisation du projet

4 - Choix d'un modèle

Choix du modèle qui permettra de prédire la possibilité d'une atteinte cardiaque sur la base d'un certain nombre de paramètres renseignés en interface.

Cahier des charges

Prédire la survenue probable d'un incident cardiaque sur la base de réponses renseignées par le patient dans un questionnaire.

Etapes de réalisation du projet

5 - Application finale

Mise en place d'une interface web fonctionnelle et facile d'utilisation afin de fournir une prédiction à partir des données d'entrée.

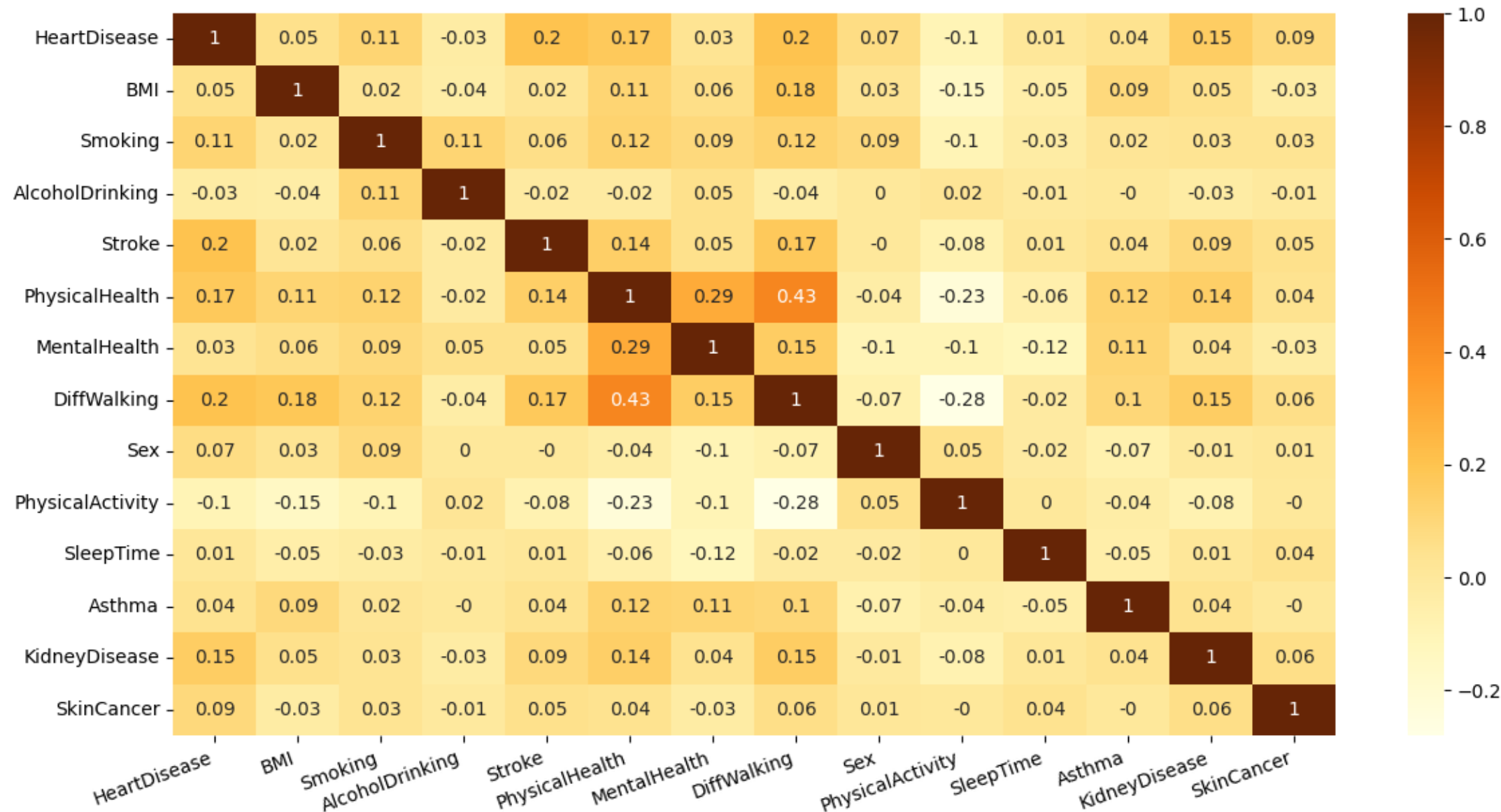
Données

Existe-t-il des variables fortement corrélées ?

Données

Existe-t-il des variables fortement corrélées ?

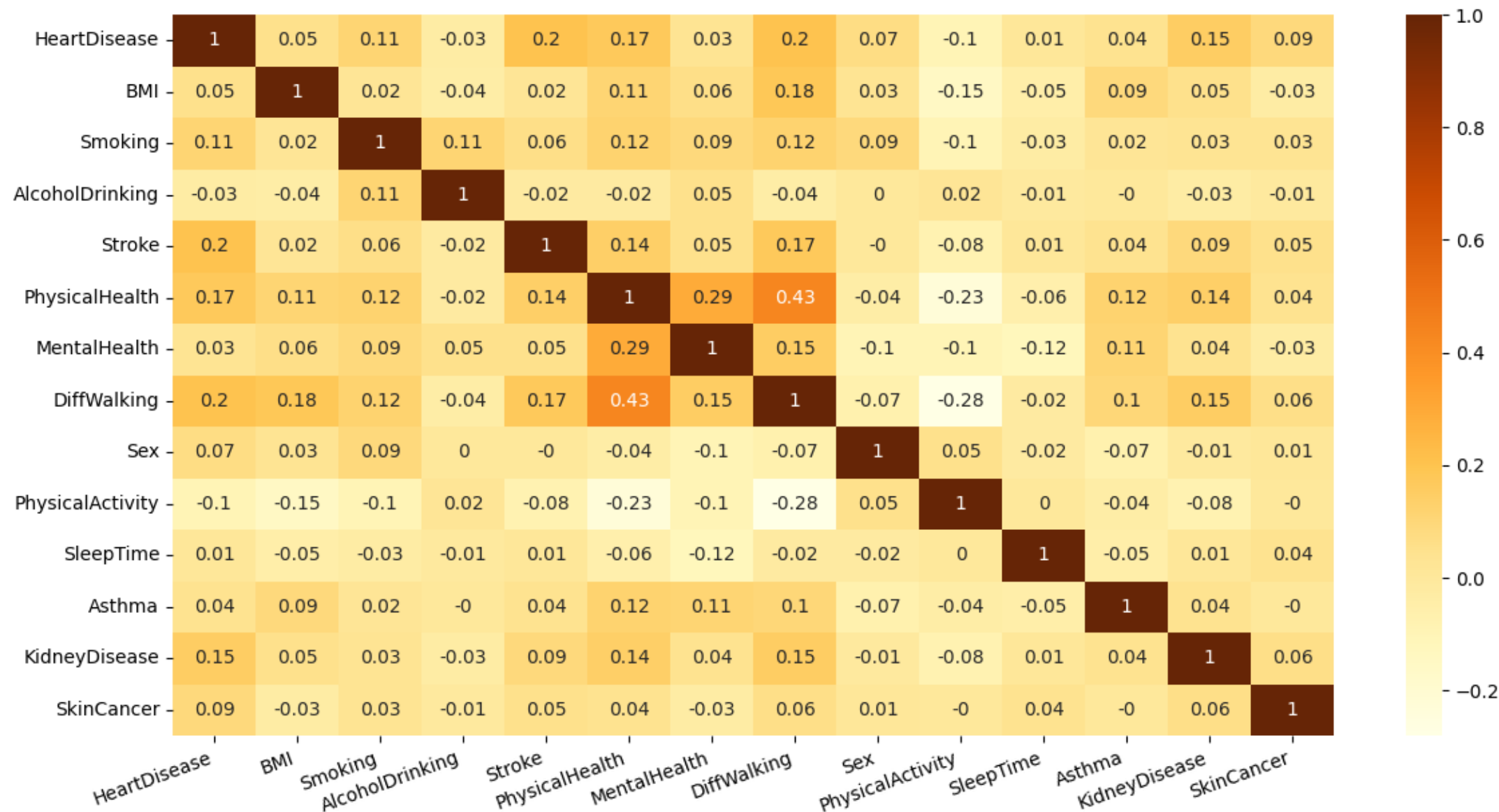
Matrice de corrélations



Données

Existe-t-il des variables fortement corrélées ?

Matrice de corrélations

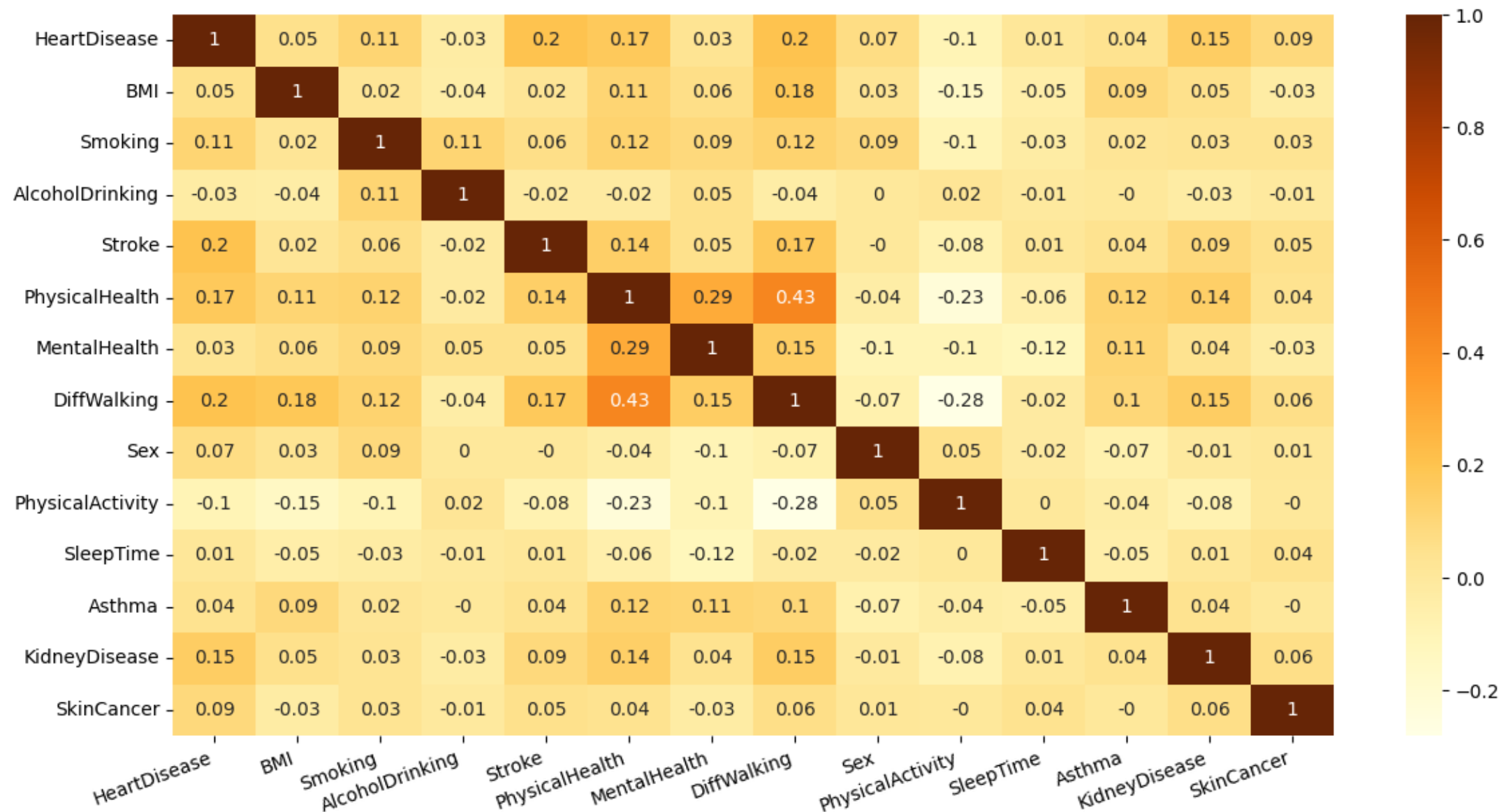


Des corrélations relativement faibles nous indiquent l'**importance de chacune des variables** pour la mesure de différentes caractéristiques

Données

Existe-t-il des variables fortement corrélées ?

Matrice de corrélations



Des corrélations relativement faibles nous indiquent l'**importance de chacune des variables** pour la mesure de différentes caractéristiques

Deux corrélations plus importantes

DiffWalking et PhysicalHealth
PhysicalHealth et MentalHealth

Données

Existe-t-il, pour chaque variable, une différence significative entre le groupe de patients sains et celui de patients atteints d'une maladie cardiaque ?

Données

Existe-t-il, pour chaque variable, une différence significative entre le groupe de patients sains et celui de patients atteints d'une maladie cardiaque ?

Analyses univariées

Données

Existe-t-il, pour chaque variable, une différence significative entre le groupe de patients sains et celui de patients atteints d'une maladie cardiaque ?

Analyses univariées

En analyse préliminaire, on observe une différence significative ($p\text{value} < 0.001$) entre le groupe de patients sains et celui de patients atteints d'une maladie cardiaque pour chaque variable.

Données

Existe-t-il, pour chaque variable, une différence significative entre le groupe de patients sains et celui de patients atteints d'une maladie cardiaque ?

Analyses univariées

En analyse préliminaire, on observe une différence significative ($pvalue < 0.001$) entre le groupe de patients sains et celui de patients atteints d'une maladie cardiaque pour chaque variable.



Variables
catégorielles

Données

Existe-t-il, pour chaque variable, une différence significative entre le groupe de patients sains et celui de patients atteints d'une maladie cardiaque ?

Analyses univariées

En analyse préliminaire, on observe une différence significative ($p\text{value} < 0.001$) entre le groupe de patients sains et celui de patients atteints d'une maladie cardiaque pour chaque variable.



Données

Existe-t-il, pour chaque variable, une différence significative entre le groupe de patients sains et celui de patients atteints d'une maladie cardiaque ?

Analyses univariées

En analyse préliminaire, on observe une différence significative ($pvalue < 0.001$) entre le groupe de patients sains et celui de patients atteints d'une maladie cardiaque pour chaque variable.



Données

Existe-t-il, pour chaque variable, une différence significative entre le groupe de patients sains et celui de patients atteints d'une maladie cardiaque ?

Analyses univariées

En analyse préliminaire, on observe une différence significative ($pvalue < 0.001$) entre le groupe de patients sains et celui de patients atteints d'une maladie cardiaque pour chaque variable.

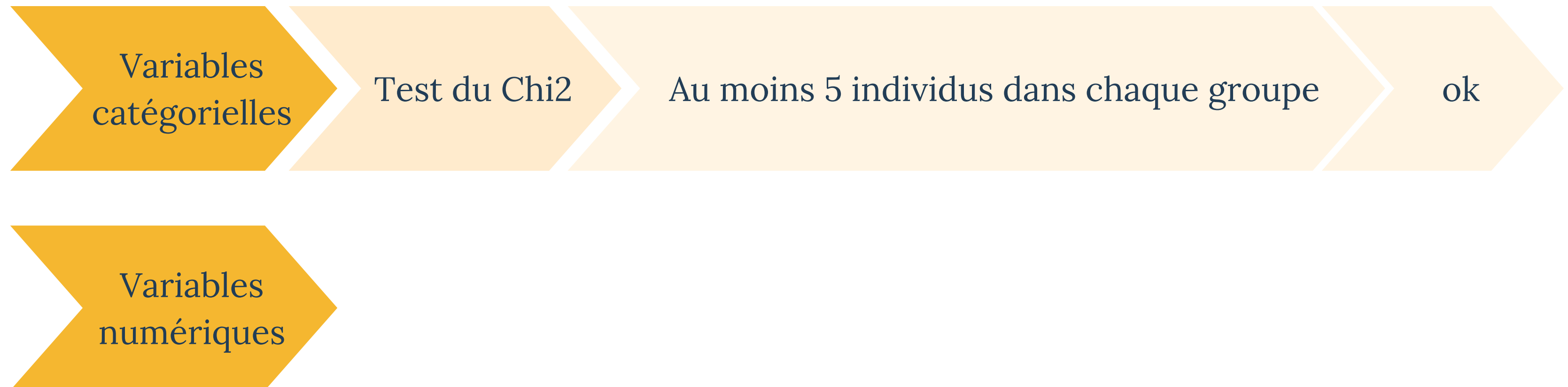


Données

Existe-t-il, pour chaque variable, une différence significative entre le groupe de patients sains et celui de patients atteints d'une maladie cardiaque ?

Analyses univariées

En analyse préliminaire, on observe une différence significative ($pvalue < 0.001$) entre le groupe de patients sains et celui de patients atteints d'une maladie cardiaque pour chaque variable.

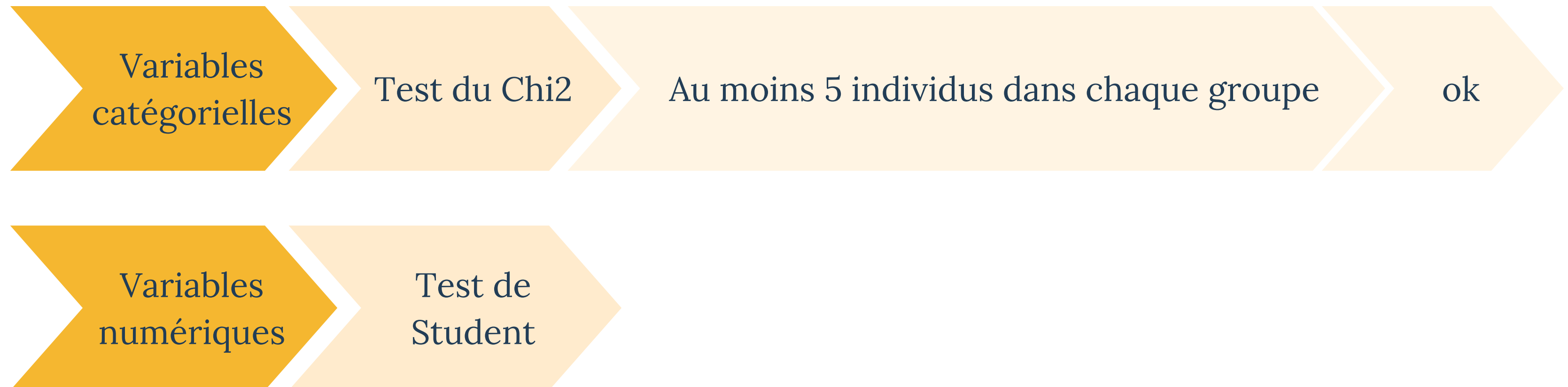


Données

Existe-t-il, pour chaque variable, une différence significative entre le groupe de patients sains et celui de patients atteints d'une maladie cardiaque ?

Analyses univariées

En analyse préliminaire, on observe une différence significative ($pvalue < 0.001$) entre le groupe de patients sains et celui de patients atteints d'une maladie cardiaque pour chaque variable.



Données

Existe-t-il, pour chaque variable, une différence significative entre le groupe de patients sains et celui de patients atteints d'une maladie cardiaque ?

Analyses univariées

En analyse préliminaire, on observe une différence significative ($p\text{-value} < 0.001$) entre le groupe de patients sains et celui de patients atteints d'une maladie cardiaque pour chaque variable.



Données

Existe-t-il, pour chaque variable, une différence significative entre le groupe de patients sains et celui de patients atteints d'une maladie cardiaque ?

Analyses univariées

En analyse préliminaire, on observe une différence significative ($pvalue < 0.001$) entre le groupe de patients sains et celui de patients atteints d'une maladie cardiaque pour chaque variable.



Données

Existe-t-il, pour chaque variable, une différence significative entre le groupe de patients sains et celui de patients atteints d'une maladie cardiaque ?

Variables catégorielles

Observations

- Condition d'application respectée : $n > 5$
- Très (trop ?) petits effectifs dans certaines catégories de AgeCategory
- Formatage des variables non optimisé

Décisions

- Regroupement des classes d'âge deux à deux afin d'en étoffer les effectifs.
- Recodage des variables dichotomiques en binaire
- One-hot encodeage des variables à plus de deux catégories

Données

Existe-t-il, pour chaque variable, une différence significative entre le groupe de patients sains et celui de patients atteints d'une maladie cardiaque ?

Variables numériques

Indépendance des échantillons

Les valeurs observées dans les différents groupes sont indépendantes les unes des autres.

Normalité

Vérification visuelle car supposée (TCL) pour de grands effectifs.

Homoscédasticité

Vérification de l'égalité des variances au sein des deux groupes (sains et malades) pour les variables BMI et SleepTime à l'aide du test de Bartlett. Sans surprise, elle n'est pas vérifiée, on décide donc de s'en affranchir.

Machine Learning

Quel algorithme utiliser ?

Machine Learning

Quel algorithme utiliser ?

K-Nearest
Neighbors

Modèle qui repose sur le principe que des points similaires peuvent être trouvés à proximité les uns des autres.

Logistic
Regression

Modèle statistique de régression qui permet de prédire la probabilité qu'un événement arrive ou non.

Random
Forest

Effectue un apprentissage sur de multiples arbres de décision entraînés sur des sous-ensembles de données légèrement différents.

Machine Learning

Quels sont les différentes étapes d'implémentation ?
(bibliothèque scikit_learn)

1 - Séparation des données

`train_test_split()`

Génération d'un jeu d'entraînement et d'un jeu test.

“random_state” : assure la reproductibilité des résultats

“stratify” : assure le respect de la répartition de la cible
dans les nouveaux jeux de données.

Machine Learning

Quels sont les différentes étapes d'implémentation ?
(librairie scikit_learn)

2 - Optimisation des paramètres GridSearchCV ()

Création d'un dictionnaire de paramètres à tester que l'on passe à la fonction GridSearchCV() qui teste les différentes combinaisons de ces paramètres.

Machine Learning

Quels sont les différentes étapes d'implémentation ?
(librairie scikit_learn)

3 - Récupération des informations du modèle

`gscv.best_estimator_` et `gscv.best_score_`

Enregistrement du modèle dans une variable afin de pouvoir le réutiliser et récupération du meilleur score (accuracy) d'apprentissage afin de pouvoir évaluer la performance du modèle.

Machine Learning

Quels sont les différentes étapes d'implémentation ?
(librairie scikit_learn)

4 - Prédictions sur le jeu test

`predict()`

On teste le modèle sur le jeu de données afin de récupérer les prédictions associées à chacun des individus testés.

Machine Learning

Quels sont les différentes étapes d'implémentation ?
(librairie scikit_learn)

5 - Mesure de l'adéquation des prédictions

`accuracy_score()`

Vérification de l'adéquation des prédictions aux
étiquettes connues.

Machine Learning

K-Nearest Neighbors

Accuracy
91.6 % (91.6)

Logistic Regression

Accuracy
91.6 % (91.6)

Random Forest

Accuracy
90.6 % (90.5)

Machine Learning

K-Nearest Neighbors

Accuracy
91.6 % (91.6)

Logistic Regression

Accuracy
91.6 % (91.6)

Random Forest

Accuracy
90.6 % (90.5)

Overfitting : Non ! Underfitting : Non !

Modèles cohérents et valides

Machine Learning

K-Nearest Neighbors

Accuracy
91.6 % (91.6)

Logistic Regression

Accuracy
91.6 % (91.6)

Random Forest

Accuracy
90.6 % (90.5)

Overfitting : Non ! Underfitting : Non !

Modèles cohérents et valides

Quel modèle choisir ?

Machine Learning

K-Nearest Neighbors

Accuracy
91.6 % (91.6)

Logistic Regression

Accuracy
91.6 % (91.6)

Random Forest

Accuracy
90.6 % (90.5)

Overfitting : Non ! Underfitting : Non !

Modèles cohérents et valides

Quel modèle choisir ?

K-Nearest Neighbors

Application



Streamlit

Application



Streamlit

Submit

Prise en main de l'outil

Mise en place d'un formulaire adapté

Asthma

☒ Yes

☐ No

Kidney disease

☒ Yes

☐ No

Skin cancer

☒ Yes

☐ No

Physical health

0 15 30

Mental health

0 15 30

BMI

1 50 100

Sleeptime

1 8 24

Submit

Done

D'apres notre étude ce patient n'a pas de prédispositions à une maladie cardiaque

Application



Streamlit

Submit

dataPreprocessing

Prise en main de l'outil

Mise en place d'un formulaire adapté

Utilisation de la fonction dataPreprocessing

Asthma

☒ Yes

☐ No

Kidney disease

☒ Yes

☐ No

Skin cancer

☒ Yes

☐ No

Physical health

0 15 30

Mental health

0 15 30

BMI

1 50 100

Sleeptime

1 8 24

Submit

Done

D'apres notre étude ce patient n'a pas de prédispositions à une maladie cardiaque

Application



Streamlit

Submit

dataPreprocessing

Machine
Learning

Prise en main de l'outil

Mise en place d'un formulaire adapté

Utilisation de la fonction dataPreprocessing

Utilisation du modèle pour prédire la donnée nouvellement entrée

Asthma

☒ Yes

☐ No

Kidney disease

☒ Yes

☐ No

Skin cancer

☒ Yes

☐ No

Physical health

0 15 30

Mental health

0 15 30

BMI

1 50 100

Sleeptime

1 8 24

Submit

Done

D'après notre étude ce patient n'a pas de prédispositions à une maladie cardiaque

Application



Streamlit

Submit

dataPreprocessing

Machine
Learning

Résultat

Prise en main de l'outil

Mise en place d'un formulaire adapté

Utilisation de la fonction dataPreprocessing

Utilisation du modèle pour prédire la donnée nouvellement entrée

Asthma

☒ Yes

☐ No

Kidney disease

☒ Yes

☐ No

Skin cancer

☒ Yes

☐ No

Physical health

0 15 30

Mental health

0 15 30

BMI

1 50 100

Sleeptime

1 8 24

Submit

Done

D'après notre étude ce patient n'a pas de prédispositions à une maladie cardiaque

Démonstration

Conclusion

Projets multiples

Améliorations significatives

Défis technique

Engagement envers l'Innovation



Merci pour votre attention !

Yen Phi Do | Hugo Alpiste | Sébastien Martel | Morgane Geoffroy

