



Projet Big Data

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

18/08/2023

Plan de présentation

Introduction

Jeu(x) de données

Présentation des jeux de données initiaux et des étapes de traitement préliminaires

Mapper

Présentation du fichier mapper.py, permettant le filtrage des lignes

Reducer

Présentation du fichier reducer.py, permettant l'envoi des données en base de données Hbase

Récupération

Description de la méthodologie de récupération des données en base de données Hbase

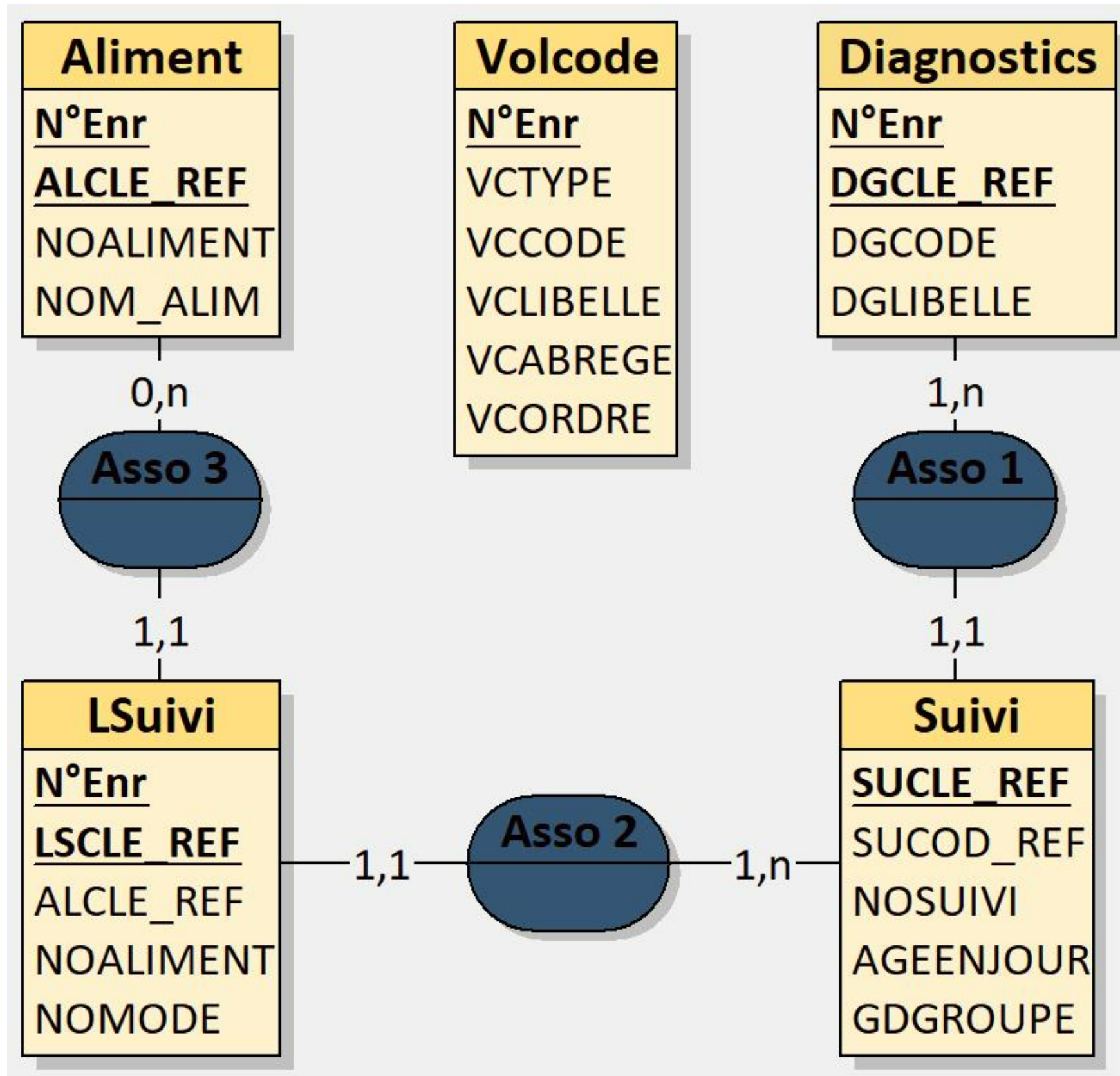
Visualisation

Présentation du dashboard généré à l'aide de Microsoft PowerBI

Problématiques rencontrées

Présentation des problématiques rencontrées lors de la réalisation du projet

Jeux de 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.

Mapper

Filtrage des valeurs aberrantes et/ou inexploitable

SUCLE_REF	SUCOD_REF	PACLE_REF	SVPOIDS	SVTAILLE	AGEENJOUR	DGLIBELLE	VCLIBELLE_G	ALCLE_REF	QUANTITE	NOM ALIM	FAMILLE_AL
1	NECKER001	1225	8,28	69	265	ACIDEMIE ISOVALERIQUE	ENFANT 6 MOIS A 1 AN				
14	NECKER001	94	0	0	4960	LEUCINOSE	HOMME 11 A 15 ANS				
19	NECKER001	40	55	161	9733	TYROSINOSE OCULO CUTANEE	FEMME 25 A 50 ANS	1231	60	SUCRE	13
20	NECKER001	479	46	0	13769	CARNITINE PALMITYL TRANSFERASE	FEMME 25 A 50 ANS	1231	150	SUCRE	13
453	NECKER001	471	3,3	0	47	ACIDEMIE METHYLMALONIQUE	ENFANT MOINS DE 6 MOIS	183	0	CARNITINE	9
5315	NECKER001	937	8	0	0	TRIPLE H	ENFANT MOINS DE 6 MOIS	1537	150	CHOCOLAT NESTLE NOIR FRUITE	13
7929	NECKER001	1174	60	170	-22047	TRIPLE H	ENFANT MOINS DE 6 MOIS	569	10	HUILE TOURNESOL 1ERE PRESSION	3
13660	NECKER001	1511	5,96	62	174	ATTENTE DE DIAG	ENFANT MOINS DE 6 MOIS	382	650	EAU D'EVIAN	1
19327	NECKER001	1563	-13,7	1	1465	ORNITHINE CARBAMYL TRANSFERASE	ENFANT 4 A 7 ANS	646	60	LAIT ENTIER CANDIA	7
19436	NECKER001	1735	2,11	0	17	DEF. EN CPS	ENFANT MOINS DE 6 MOIS				
21722	DS-MS 001	1873	70	170	13603	ORNITHINE CARBAMYL TRANSFERASE	HOMMES ADULTES	1452	220	RIZ PRECUIT LONG GRAIN	5
21918	DS-MS 001	1848	36	157	9112	PHENYLCETONURIE HYPERPHE.	HOMMES ADULTES	642	200	LAIT DEMI ECREME UHT	7
24988	NECKER001	1258	33,5	1,325	3780	GLYCOGENOSE TYPEIa et Ib	ENFANT 7 A 11 ANS	869	42	PAIN DE BLE	5

- SUCOD_REF = "NECKER001"
- SVPOIDS > 0
- SVTAILLE >= 0
- AGEENJOUR >= 0
- QUANTITE > 0 ou vide

Reducer

SUCLE_REF	SVPOIDS	SVTAILLE	DGLIBELLE	VCLIBELLE_G	LSCLE_REF	QUANTITE	NOM_ALIM	SANSSEL
15	55	161	TYROSINOSE OCULO CUTANEE	FEMME 25 A 50 ANS	-	-	-	-
16	16,7	95	ACIDEMIE METHYLMALONIQUE	ENFANT 4 A 7 ANS	-	-	-	-
17	55	161	TYROSINOSE OCULO CUTANEE	FEMME 25 A 50 ANS	-	-	-	-
18	55	161	TYROSINOSE OCULO CUTANEE	FEMME 25 A 50 ANS	-	-	-	-
19	55	161	TYROSINOSE OCULO CUTANEE	FEMME 25 A 50 ANS	8058	60	SUCRE	FALSE
19	55	161	TYROSINOSE OCULO CUTANEE	FEMME 25 A 50 ANS	9581	140	POMMES	FALSE
19	55	161	TYROSINOSE OCULO CUTANEE	FEMME 25 A 50 ANS	11105	9	BISCOTTE APROTEN	FALSE
19	55	161	TYROSINOSE OCULO CUTANEE	FEMME 25 A 50 ANS	12629	225	RATATOUILLE	FALSE
20	46	0	CARNITINE PALMITYL TRANSFERASE	FEMME 25 A 50 ANS	14153	150	SUCRE	FALSE
20	46	0	CARNITINE PALMITYL TRANSFERASE	FEMME 25 A 50 ANS	15677	8,8	BISCOTTE 6CEREALES	FALSE
20	46	0	CARNITINE PALMITYL TRANSFERASE	FEMME 25 A 50 ANS	17201	10	NUTELLA	FALSE
20	46	0	CARNITINE PALMITYL TRANSFERASE	FEMME 25 A 50 ANS	18725	10	HUILE MAIS	TRUE

ROWKEY	SUIVI				LSUIVI		
	SVPOIDS	SVTAILLE	DGLIBELLE	VCLIBELLE_G	QUANTITE	NOM_ALIM	SANSSEL
15-	55	161	TYROSINOSE OCULO CUTANEE	FEMME 25 A 50 ANS	-	-	-
16-	16,7	95	ACIDEMIE METHYLMALONIQUE	ENFANT 4 A 7 ANS	-	-	-
17-	55	161	TYROSINOSE OCULO CUTANEE	FEMME 25 A 50 ANS	-	-	-
18-	55	161	TYROSINOSE OCULO CUTANEE	FEMME 25 A 50 ANS	-	-	-
19-8058	55	161	TYROSINOSE OCULO CUTANEE	FEMME 25 A 50 ANS	60	SUCRE	FALSE
19-9581	55	161	TYROSINOSE OCULO CUTANEE	FEMME 25 A 50 ANS	140	POMMES	FALSE
19-11105	55	161	TYROSINOSE OCULO CUTANEE	FEMME 25 A 50 ANS	9	BISCOTTE APROTEN	FALSE
19-12629	55	161	TYROSINOSE OCULO CUTANEE	FEMME 25 A 50 ANS	225	RATATOUILLE	FALSE
20-14153	46	0	CARNITINE PALMITYL TRANSFERASE	FEMME 25 A 50 ANS	150	SUCRE	FALSE
20-15677	46	0	CARNITINE PALMITYL TRANSFERASE	FEMME 25 A 50 ANS	8,8	BISCOTTE 6CEREALES	FALSE
20-17201	46	0	CARNITINE PALMITYL TRANSFERASE	FEMME 25 A 50 ANS	10	NUTELLA	FALSE
20-18725	46	0	CARNITINE PALMITYL TRANSFERASE	FEMME 25 A 50 ANS	10	HUILE MAIS	TRUE

Data



Hbase

Création des rowkeys (uniques) en fusionnant les colonnes SUCLE_REF et LSCLE_REF.

Répartition des variables dans deux familles de colonnes SUIVI et LSUIVI.

Récupération



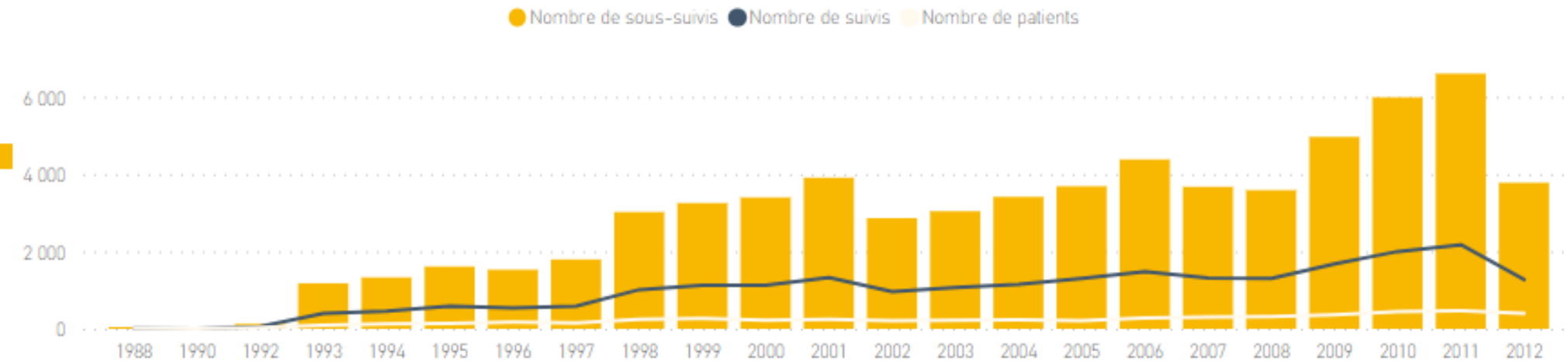
Visualisation



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

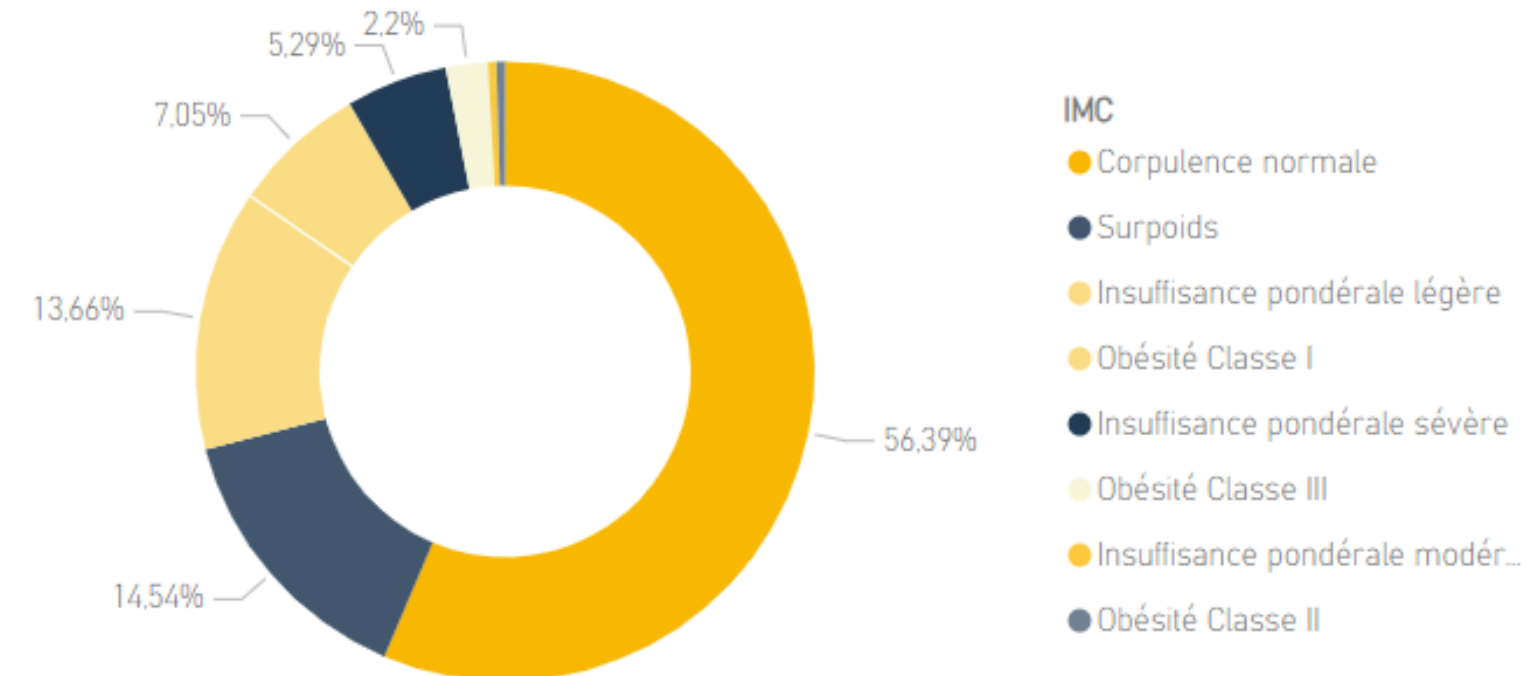
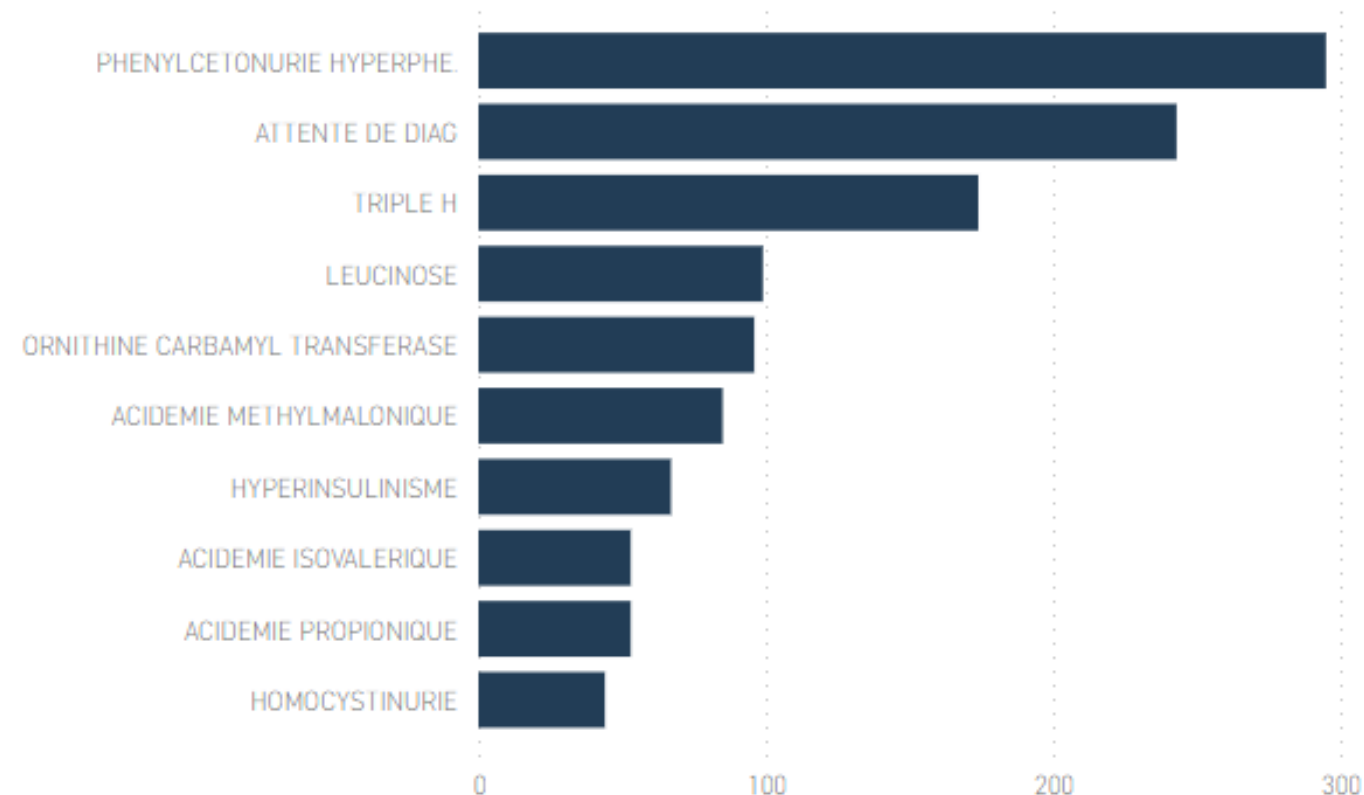
1673

Nombre ajusté de patients

243

En attente de diagnostic

Nombre de patients par diagnostic



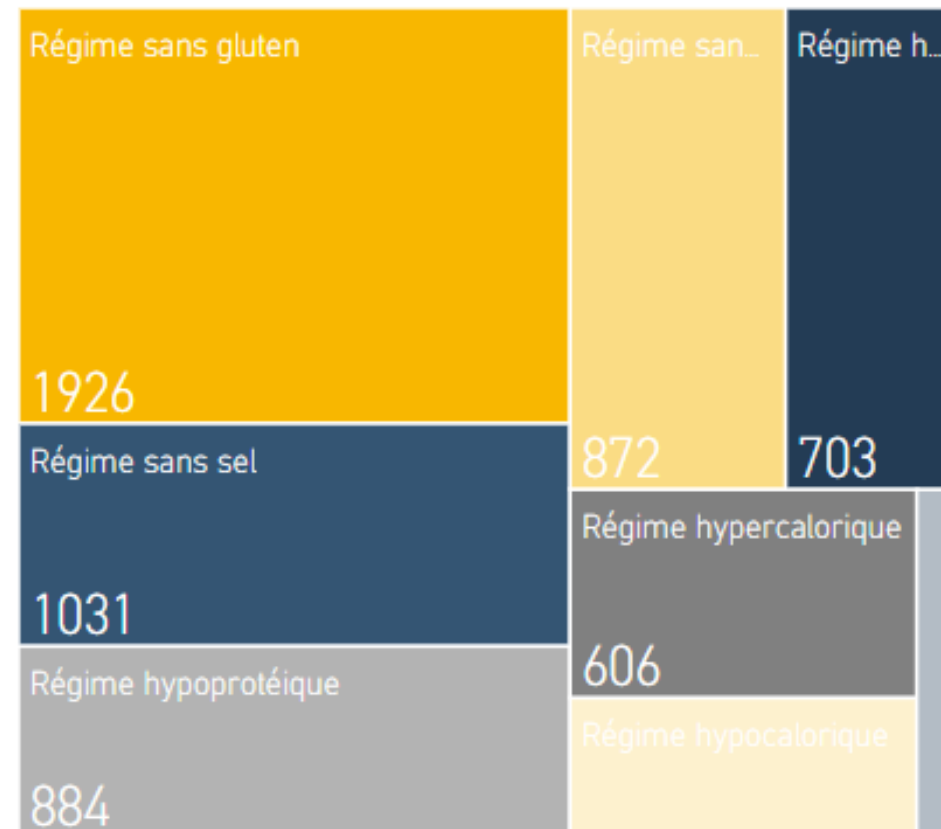
Visualisation



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

Projet Big Data

Répartition des régimes prescrits

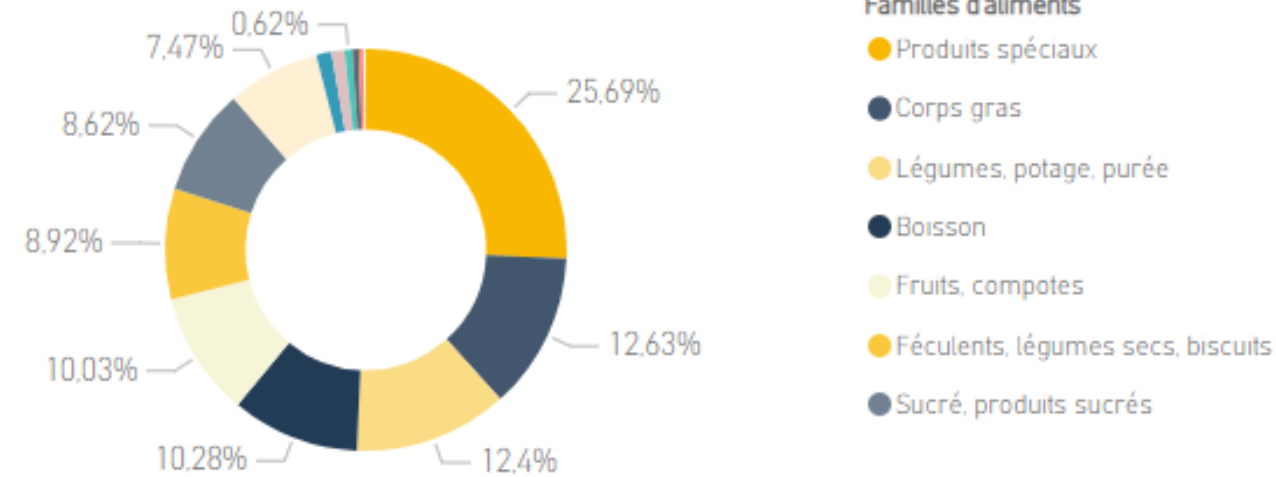


Nombre d'aliments par régime alimentaire



Focus sur le diagnostic de la Phénylcétonurie Hyperphe

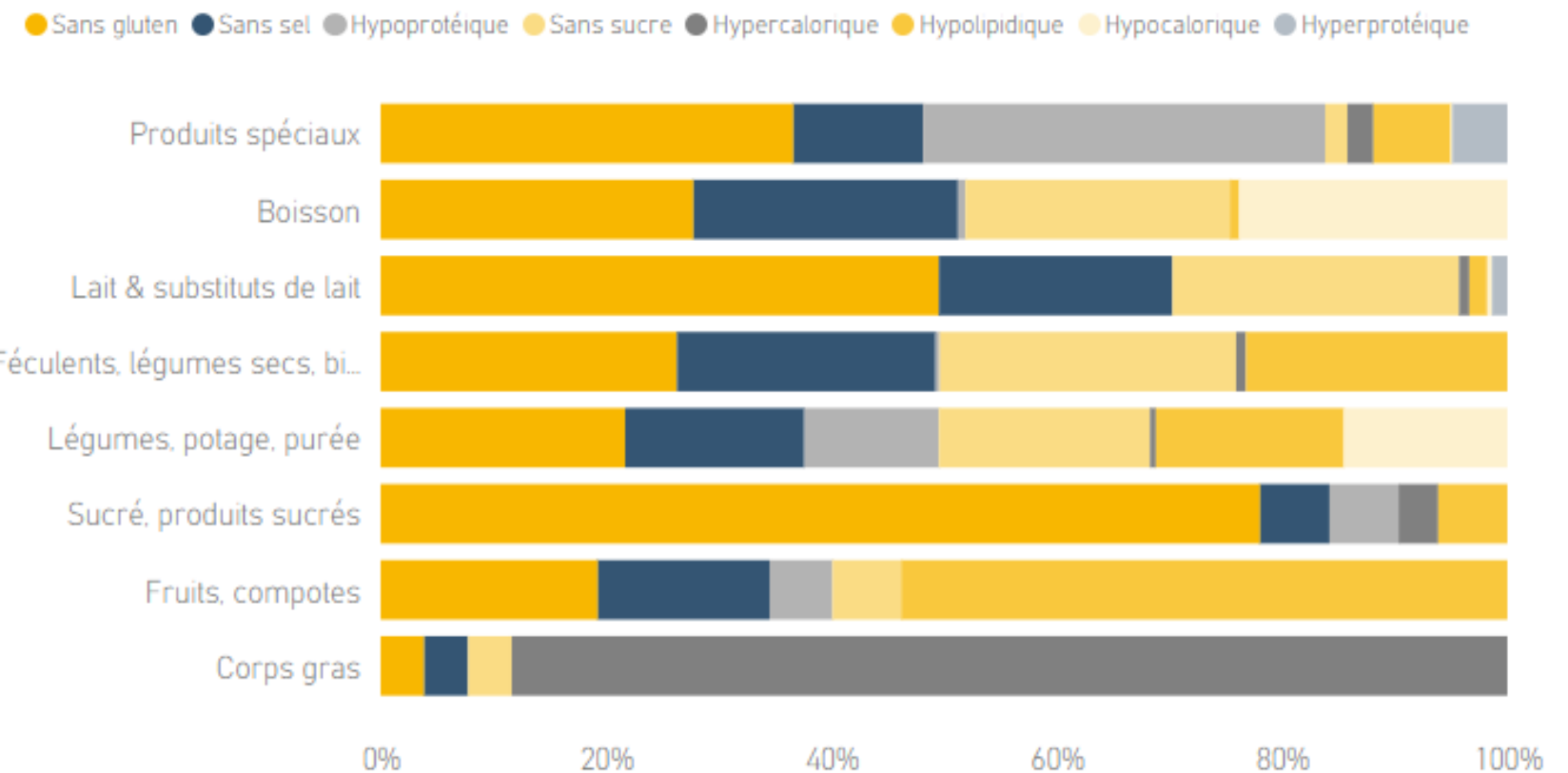
Répartition des familles d'aliments prescrites



6317
Nombre total de sous-suivis

6317
Nombre ajusté de sous-suivis

Répartition des régimes alimentaires par famille d'aliments



Problématiques rencontrées

- Appropriation du jeu de données
- Définition des objectifs pour le mapper et le reducer
- Déconnexions intempestives à la base durant les développements. Dès que celle-ci doit être interrogée. (+ jar)
- Problème de configuration des conteneurs et manque de compétences linux; f-strings, CRLF, '/udcc3'
- Volonté d'envoyer les données issues de la base à la suite ELK pour la génération d'un rapport. Pas faisable (explorable) dans le temps imparti. Utilisation de PowerBI.



Merci pour votre attention !

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