

Les Evénements de l'année
en économie médicale 2015
Paris, les 22&23 janvier 2015

Thème

*La fouille de données médico-administratives comme moyen
d'aide à l'intelligence territoriale.*

- *R. BENAMIROUCHE*

- *Email: rbena2002@hotmail.com*



De la Statistique ...

- Quelques centaines d'individus,
- Quelques variables,
- Fortes hypothèses sur les lois statistiques suivies,
- Importance accordée au calcul,
- Échantillon aléatoire.

... au Data mining

- Des millions d'individus,
- Des centaines de variables,
- Données recueillies sans étude préalable,
- Nécessité de calculs rapides,
- Pas un échantillon aléatoire.

Techniques utilisées selon leur « origine » :

Statistiques

Théorie de l'estimation, tests,
Econométrie, Maximum de
vraisemblance et moindres carrés,
régression logistique.

Analyse de données (Statistique exploratoire)

Descriptive factorielle, Discrimination,
clustering
Méthodes géométriques, probabilités,
ACP, AFCM,...

	Var1	Var2	VarJ
individu 1				
individu 2		Valeurs		
.....				
individu n				

Informatique (Intelligence artificielle)

Apprentissage symbolique
Reconnaissance de formes
Une étape de l'intelligence artificielle
Réseaux neurones, algorithmes génétiques,...

Informatique (Base de données)

Exploration des bases de données
Volumétrie
Règles d'association, motifs fréquents,...

Phase 01 : préparation des données (pré-traitement)

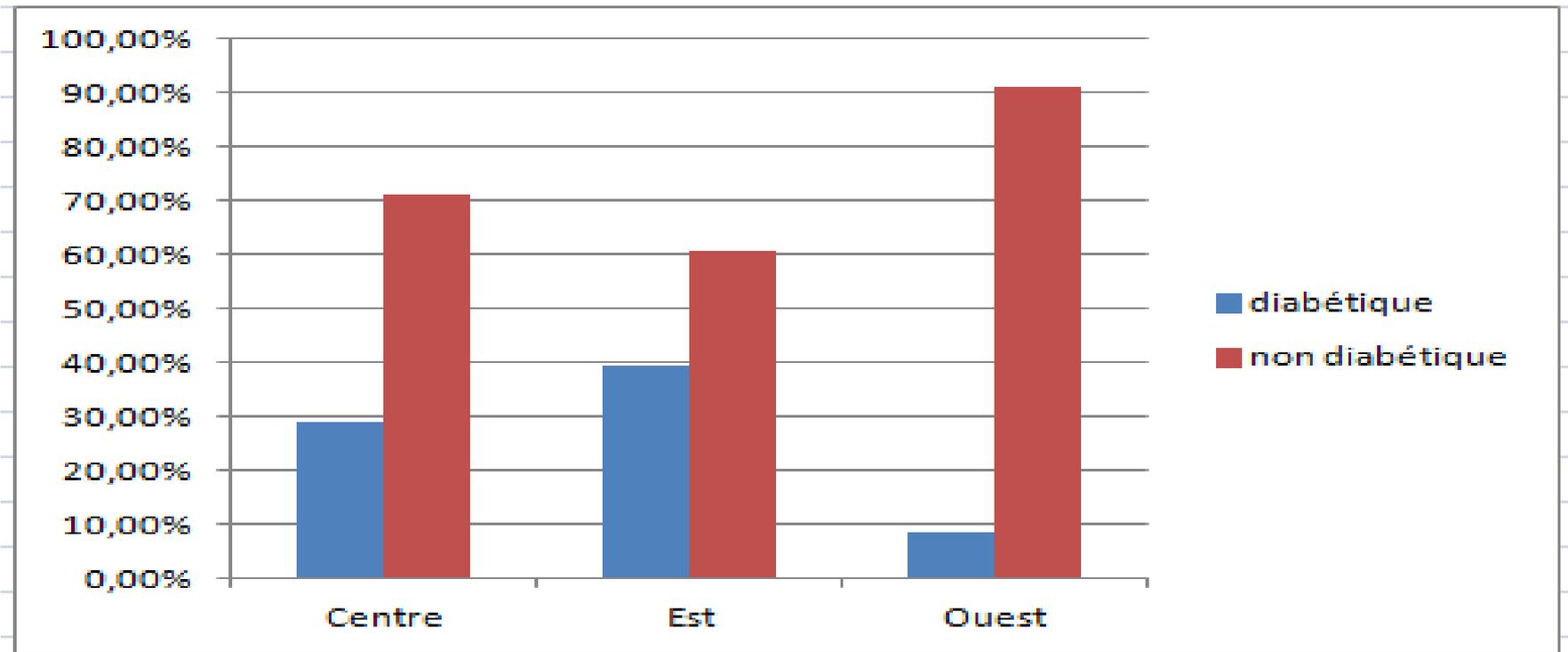
- Dans cette phase on a procédé :
 - Au regroupement des données des wilayas en régions (Est, Ouest, centre).
 - À la segmentation des âges par tranches d'âge

Phase 02 : exploration des données

- Dans cette phase on a procédé à l'exploration des données afin :
 - De tirer les informations et les règles cachés;
 - D'avoir une vision sur le phénomène à étudier.

Phase 02 : exploration des données

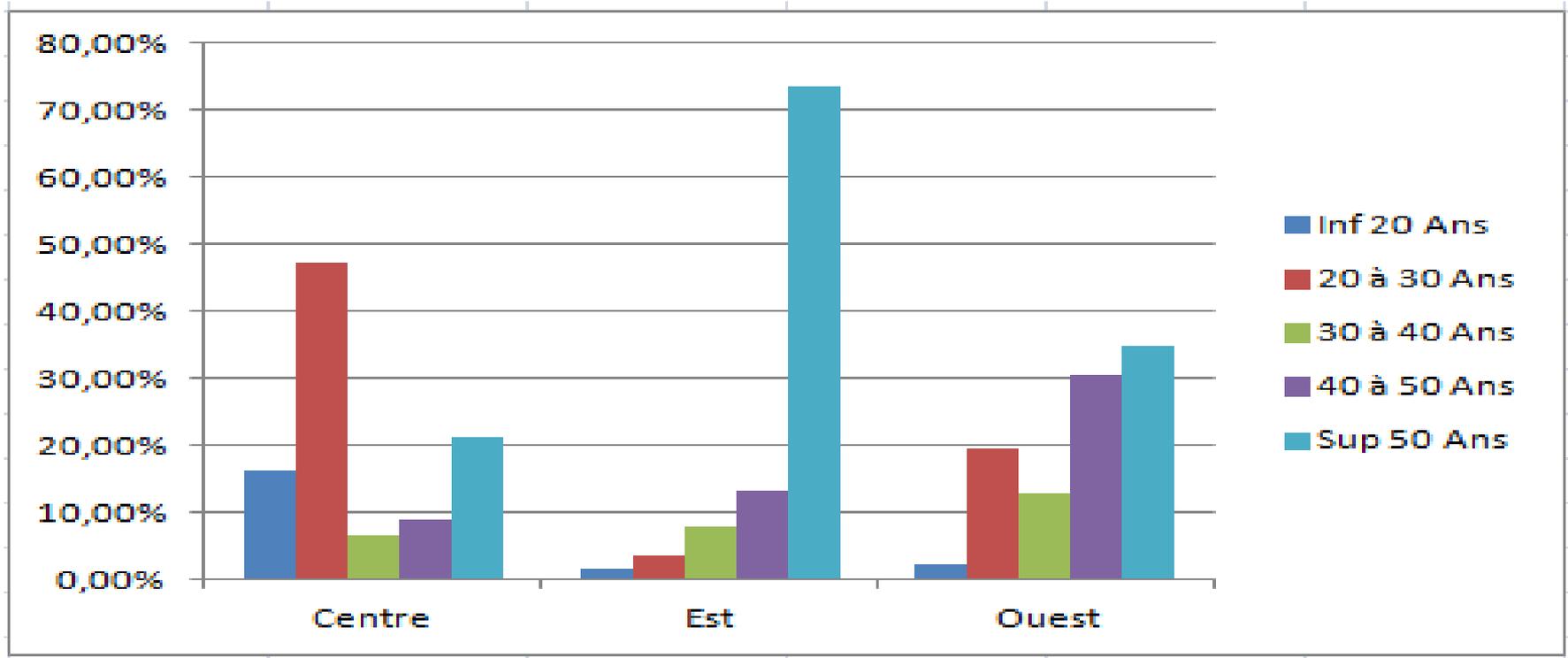
Vue globale



- On remarque que le plus grand nombre des diabétiques se concentre sur la région de l'est

Phase 02 : exploration des données

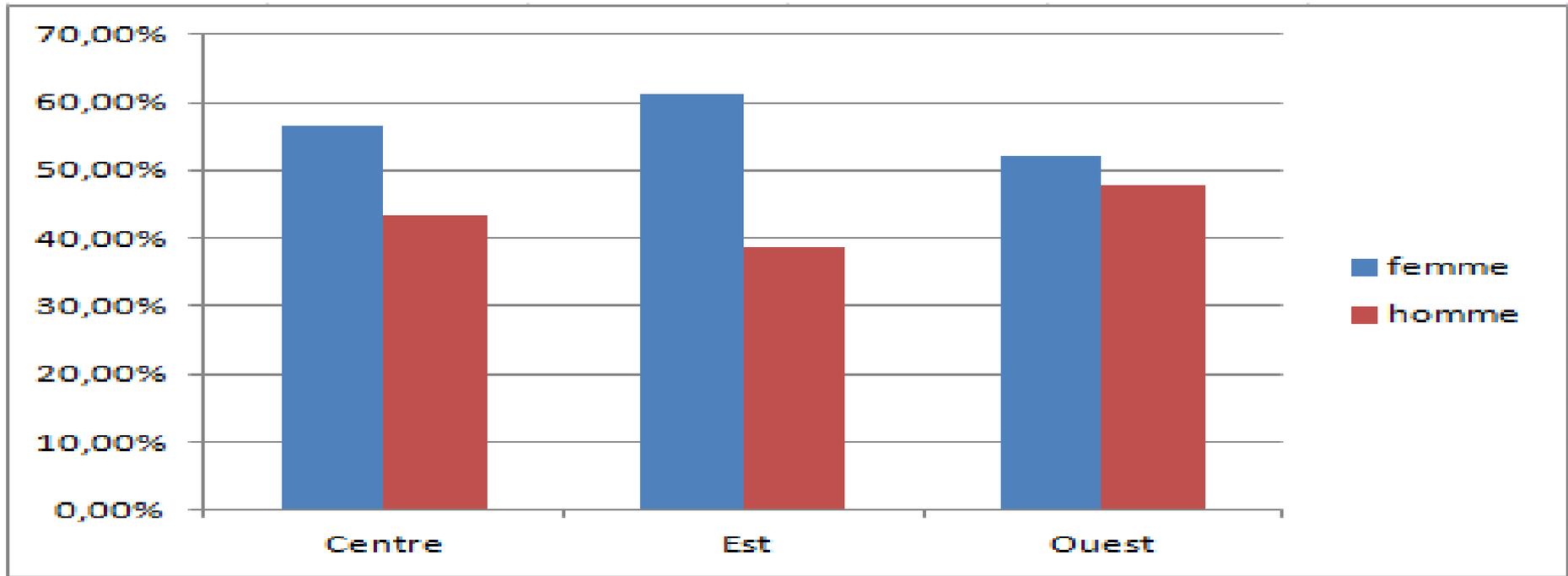
Répartition par Age



- On remarque que la population jeune se concentre dans le centre tandis que la population vieille se concentre dans l'est, par contre, l'ouest regroupe la population d'âge moyen.

Phase 02 : exploration des données

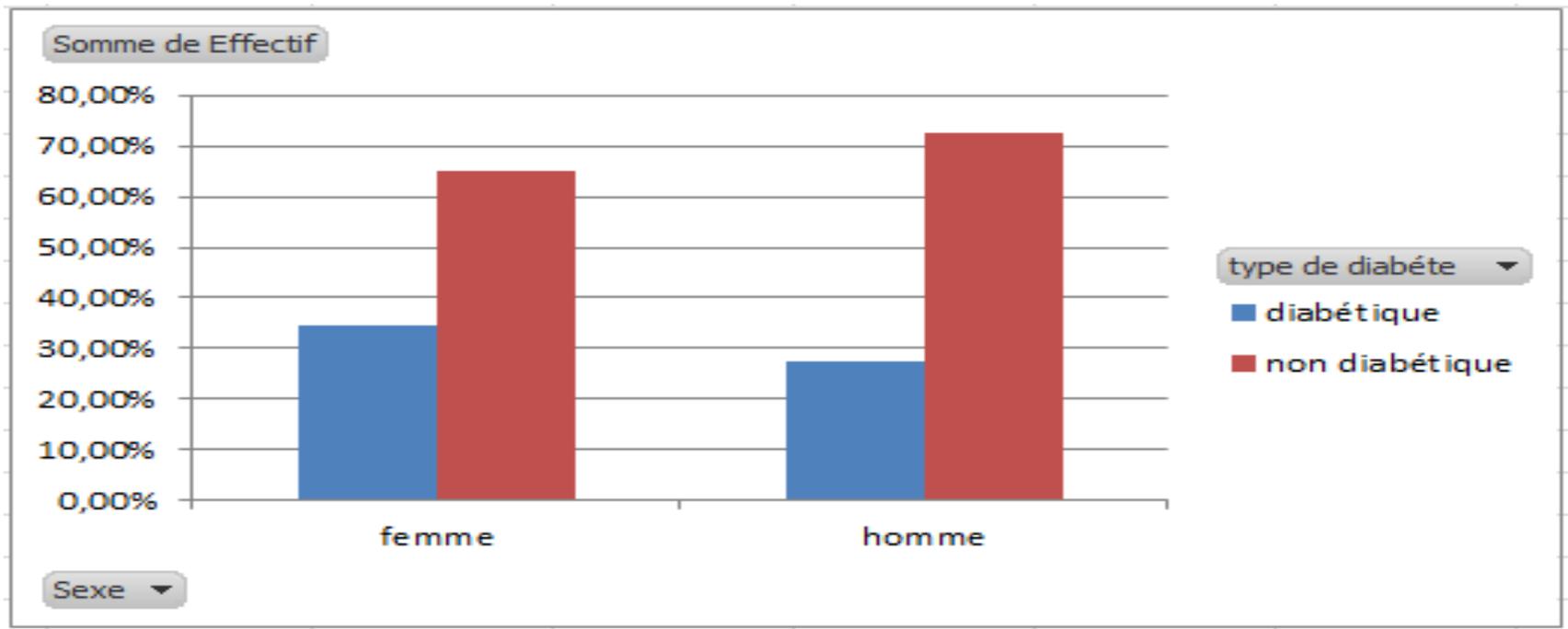
Répartition par Sexe



- On remarque que la distribution par sexe est homogène dans les différentes régions

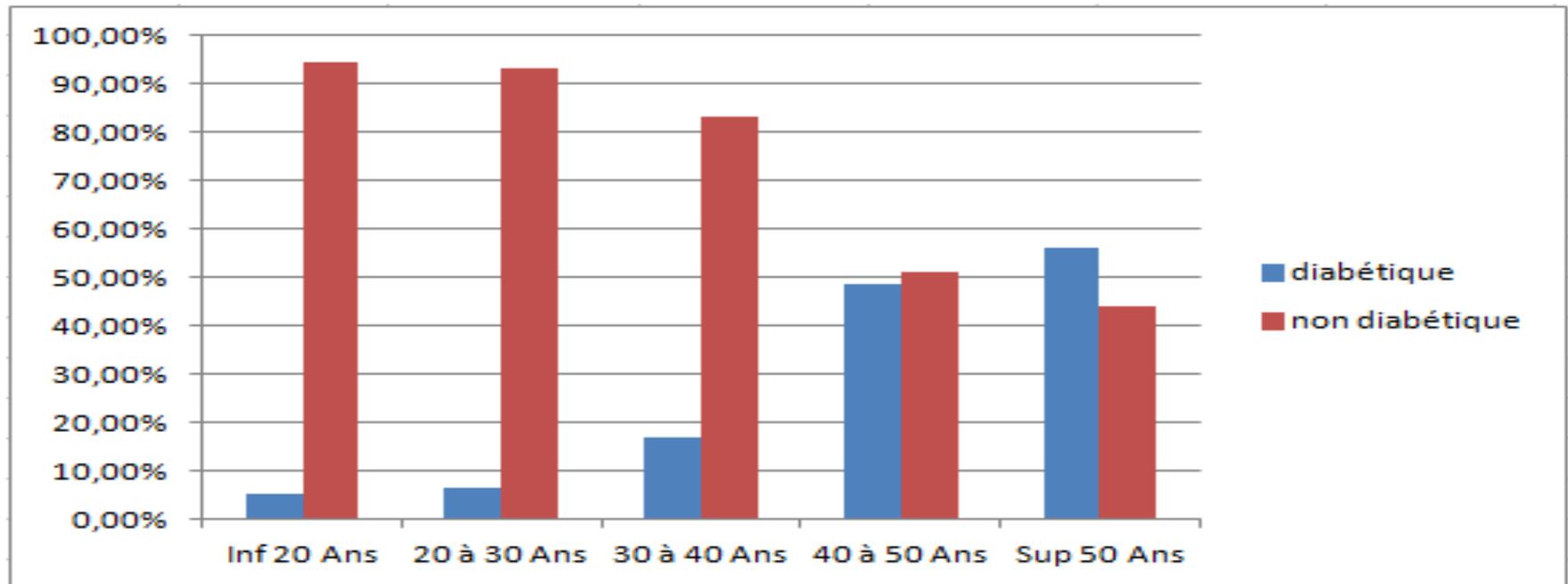
Phase 02 : exploration des données

Répartition du diabète par Sexe



Phase 02 : exploration des données

Répartition du diabète par Age



- On remarque le diabète est plus présent chez les personnes âgées

Phase 03 : modélisation data mining

- Durant cette phase on a procédé à la modélisation du diabète par région. Et cela en utilisant l'algorithme d'arbre de décision qui nous a permis d'établir les règles d'associations propres à chaque région.

Phase 03 : modélisation data mining

règles d'association région est

tour de taille <= 85,500 => non diabétique

age <= 32,500 => diabétique

pression artérielle <= 105 => diabétique

Pression artérielle > 105 => non diabétique

age > 32,500 => non diabétique

tour de taille > 85,500 => non diabétique

IMC in ["24,2215" "25,3906" "28,3737" "29,3962" "29,4118" "35,4979" "36,8889"] => diabétique

age <= 46 => non diabétique

age > 46 => diabétique

age <= 68,500 => diabétique

age > 68,500 => non diabétique

IMC in ["20,5714" "20,8307" "23,4586" "24,3911" "44,4444"] => non diabétique

antécédent familial in ["absence"] => non diabétique

Sexe in ["femme"] => non diabétique

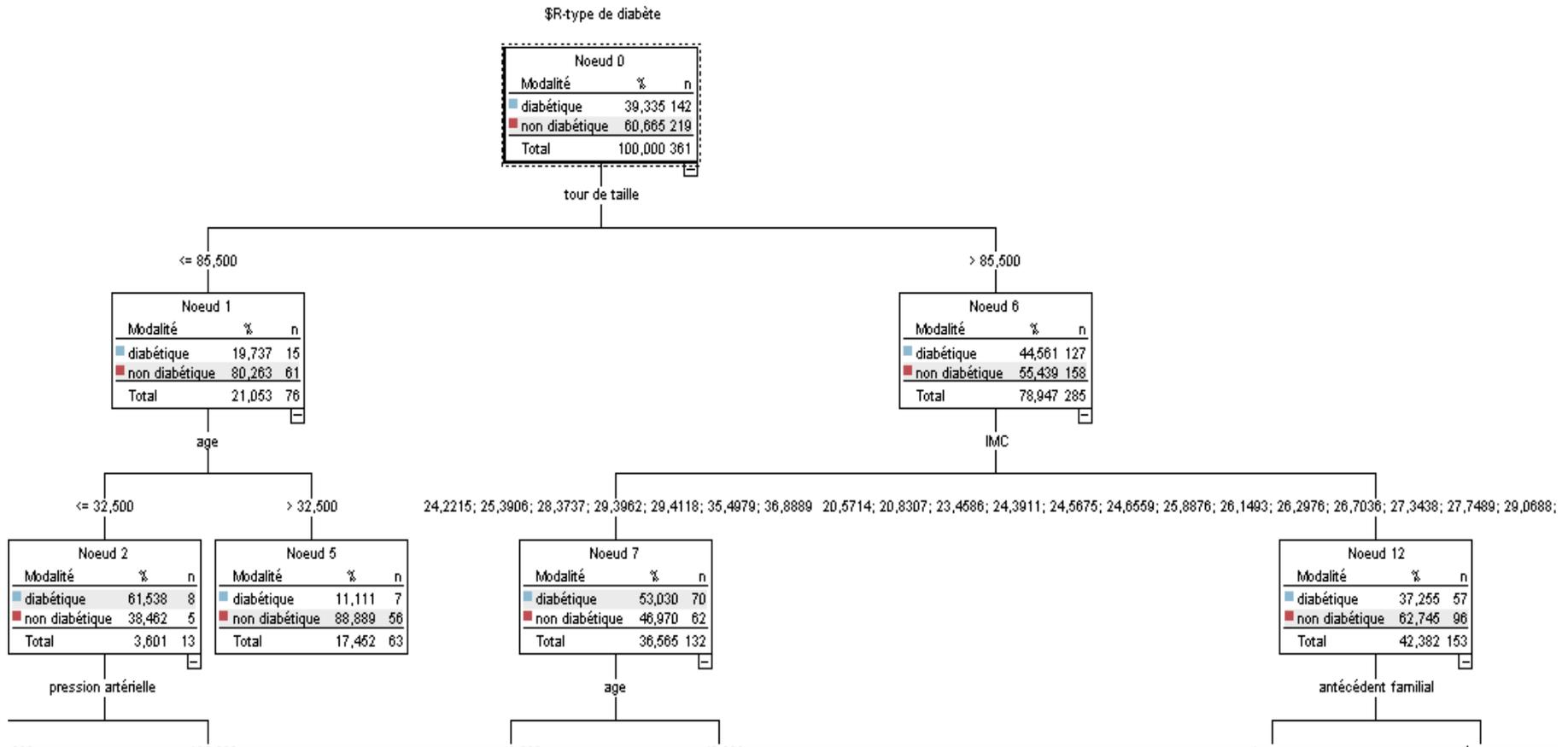
pression artérielle <= 165 => non diabétique

pression artérielle > 165 => diabétique

Sexe in ["homme"] => non diabétique

antécédent familial in ["présence"] => diabétique

L'arbre de la région est



Phase 03 : modélisation data mining

règles d'association région ouest

pression diastolique $\leq 91,500 \Rightarrow$ non diabétique

age $\leq 76,500 \Rightarrow$ non diabétique

age $> 76,500 \Rightarrow$ diabétique

pression diastolique $> 91,500 \Rightarrow$ non diabétique

age $\leq 54 \Rightarrow$ diabétique

age $\leq 49,500 \Rightarrow$ non diabétique

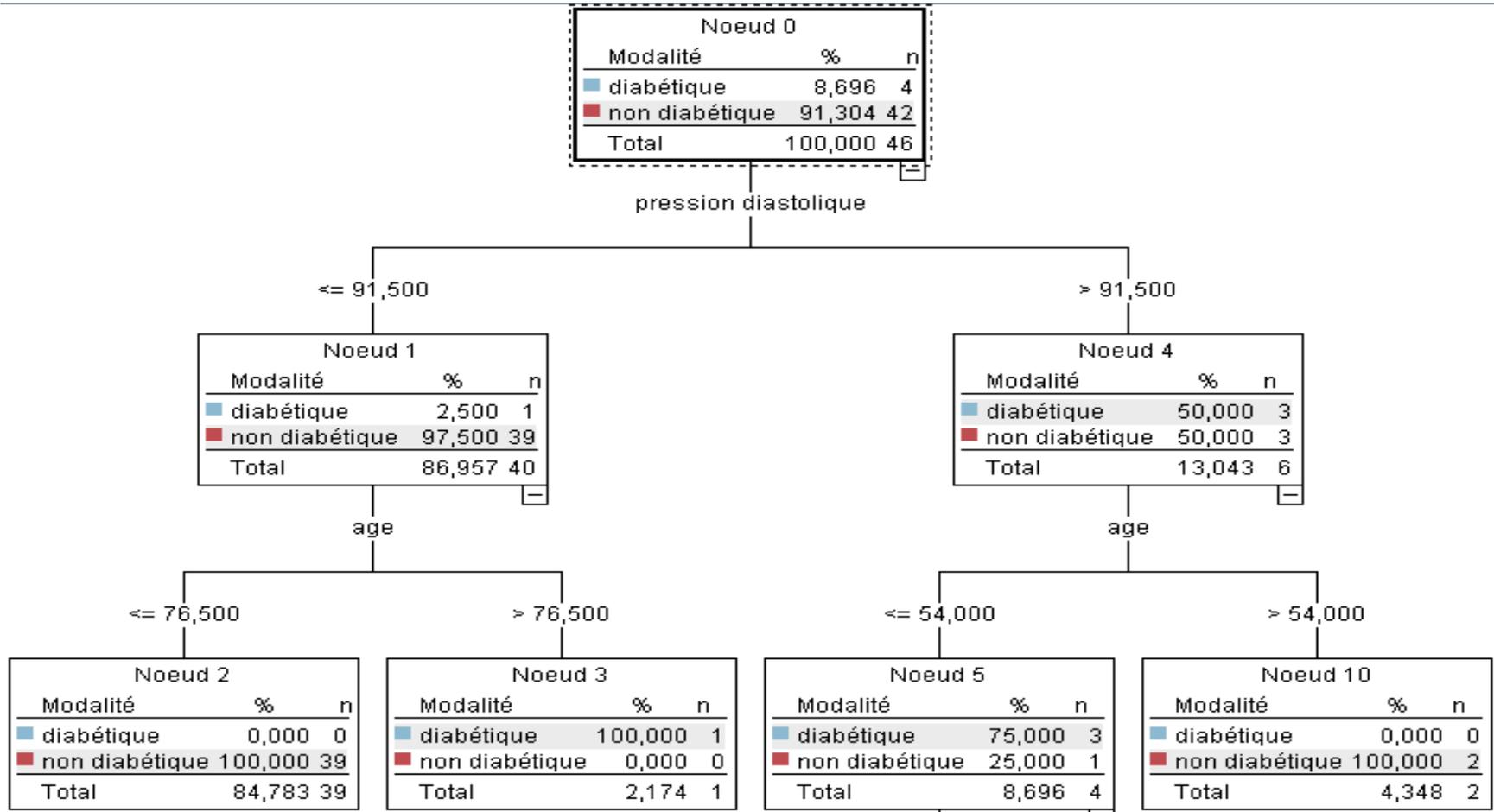
age $\leq 45 \Rightarrow$ diabétique

age $> 45 \Rightarrow$ non diabétique

age $> 49,500 \Rightarrow$ diabétique

age $> 54 \Rightarrow$ non diabétique

L'arbre de la région ouest



Phase 03 : modélisation data mining

règles d'association région Centre

IMC in ["25,3906" "26,2976" "27,6817"] => diabétique

age <= 40,500 => non diabétique

age > 40,500 => diabétique

tour de taille <= 78,500 => non diabétique

tour de taille > 78,500 => diabétique

age <= 56,500 => diabétique

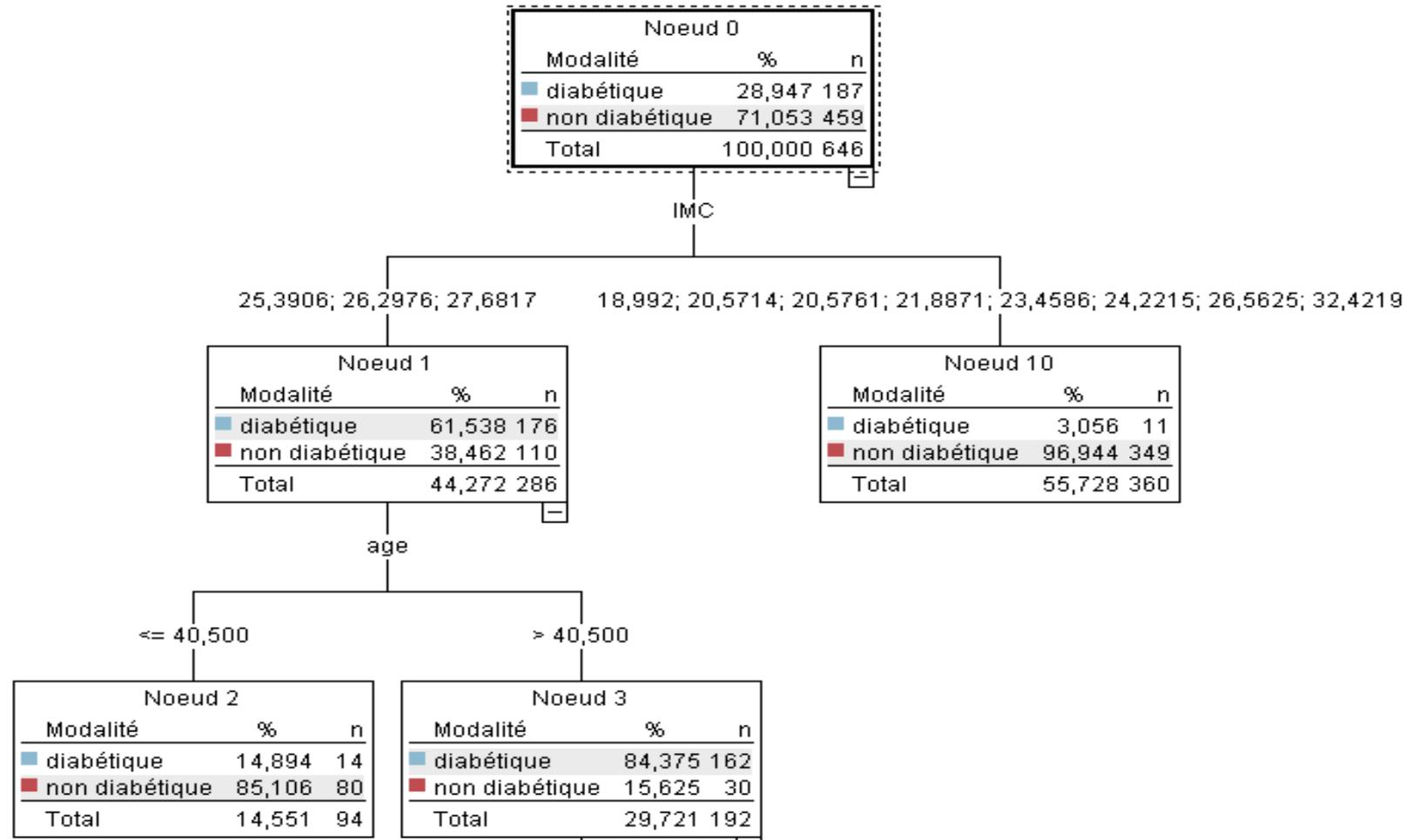
pression diastolique <= 82,500 => diabétique

pression diastolique > 82,500 => non diabétique

age > 56,500 => diabétique

IMC in ["18,992" "20,5714" "20,5761" "21,8871" "23,4586" "24,2215" "26,5625" "32,4219"] => non diabétique

L'arbre de la région centre



	Wilaya	Wilaya-7	Wilaya-16	Wilaya-19	Wilaya-20	Wilaya-21	Wilaya-23	Wilaya-24	Wilaya-25	Wilaya-26	Wilaya-31	Wilaya-46
Type de diabète	Diabétique	+	+	-		+		+	+	-		
	non diabétique	-	-	+		-		-	-	+		
Sexe	femme	+	+	-		+		+	+	-		
	homme	-	-	+		-		-	-	+		
Antécédent familial	absence											
	présence											
AGE	<= 18	-	-	+		-		-	-	+		
	19 - 28	-	-	+		-		-	-	+		
	29 - 38	-	-	+		-		-	-	+		
	39 - 48	+	+	-		+		+	+	-		
	49 - 58	+	+	-		+		+	+	-		
	59 - 68	+	+	-		+		+	+	-		
	69 - 78	+	+	-		+		+	+	-		
	79 - 88	+	+	-		+		+	+	-		
	89+											
Tour de Taille	<= 47	-	-	+		-		-	-	+		
	48 - 62	-	-	+		-		-	-	+		
	63 - 73	-	-	+		-		-	-	+		
	74 - 83	-	-	+		-		-	-	+		
	84 - 98	+	+	-		+		+	+	-		
	99 - 114	+	+	-		+		+	+	-		
	115+	+	+	-		+		+	+	-		
PAS	<= 83											
	84 - 102	-	-	+		-		-	-	+		
	103 - 116	-	-	+		-		-	-	+		
	117 - 134	-	-	+		-		-	-	+		
	135 - 149	+	+	-		+		+	+	-		
	150 - 164	+	+	-		+		+	+	-		
	165 - 179	+	+	-		+		+	+	-		
180+	+	+	-		+		+	+	-			
IMC	<= 16.4382	-	-	+		-		-	-	+		
	16.4383 - 21.1403	-	-	+		-		-	-	+		
	21.1404 - 26.0103	-	-	+		-		-	-	+		
	26.0104 - 30.8803	+	+	-		+		+	+	-		
	30.8804 - 35.5824	+	+	-		+		+	+	-		
	35.5825 - 40.4524	+	+	-		+		+	+	-		
	40.4525 - 45.0000	+	+	-		+		+	+	-		
	45.0001 - 50.0000	+	+	-		+		+	+	-		
	50.0001+											
PAD	<= 50											
	51 - 60											
	61 - 70	-	-	+		-		-	-	+		
	71 - 80											
	81 - 90	+	+	-		+		+	+	-		
	91+	+	+	-		+		+	+	-		

	Wilaya	Wilaya-7	Wilaya-16	Wilaya-19	Wilaya-20	Wilaya-21	Wilaya-23	Wilaya-24	Wilaya-25	Wilaya-26	Wilaya-31	Wilaya-46
Type de diabète	Diabétique	+	+		-	-	-	+		-		+
	non diabétique	-	-		+	+	+	-		+		-
Sexe	femme	+	+		-	-	-	+		-		+
	homme	-	-		+	+	+	-		+		-
Antécédent familial	absence	+	+		-	-	-	+		-		+
	présence	-	-		+	+	+	-		+		-
AGE	<= 18	+	+		-	-	-	+		-		+
	19 - 28											
	29 - 38	+	+		-	-	-	+		-		+
	39 - 48	+	+		-	-	-	+		-		+
	49 - 58											
	59 - 68	-	-		+	+	+	-		+		-
	69 - 78	-	-		+	+	+	-		+		-
	79 - 88	-	-		+	+	+	-		+		-
89+	-	-		+	+	+	-		+		-	
Tour de Taille	<= 47											
	48 - 62	+	+		-	-	-	+		-		+
	63 - 73	+	+		-	-	-	+		-		+
	74 - 83											
	84 - 98	-	-		+	+	+	-		+		-
	99 - 114											
	115+	+	+		-	-	-	+		-		+
PAS	<= 83	+	+		-	-	-	+		-		+
	84 - 102	+	+		-	-	-	+		-		+
	103 - 116	+	+		-	-	-	+		-		+
	117 - 134	-	-		+	+	+	-		+		-
	135 - 149	-	-		+	+	+	-		+		-
	150 - 164	-	-		+	+	+	-		+		-
	165 - 179	-	-		+	+	+	-		+		-
	180+	-	-		+	+	+	-		+		-
IMC	<= 16,4382	+	+		-	-	-	+		-		+
	16,4383 - 21,1403	+	+		-	-	-	+		-		+
	21,1404 - 26,0103	-	-		+	+	+	-		+		-
	26,0104 - 30,8803	-	-		+	+	+	-		+		-
	30,8804 - 35,5824	+	+		-	-	-	+		-		+
	35,5825 - 40,4524	+	+		-	-	-	+		-		+
	40,4525 - 45,0000	+	+		-	-	-	+		-		+
	45,0001 - 50,0000	+	+		-	-	-	+		-		+
50,0001+												
PAD	<= 50	+	+		-	-	-	+		-		+
	51 - 60	+	+		-	-	-	+		-		+
	61 - 70											
	71 - 80	-	-		+	+	+	-		+		-
	81 - 90	-	-		+	+	+	-		+		-
	91+	-	-		+	+	+	-		+		-

	Wilaya	Wilaya-7	Wilaya-16	Wilaya-19	Wilaya-20	Wilaya-21	Wilaya-23	Wilaya-24	Wilaya-25	Wilaya-26	Wilaya-31	Wilaya-46
Type de diabète	Diabétique	+		+	+	-			+	-		-
	non diabétique	-		-	-	+			-	+		+
Sexe	femme	-		-	-	+			-	+		+
	homme	+		+	+	-			+	-		-
Antécédent familial	absence	-		-	-	+			-	+		+
	présence	+		+	+	-			+	-		-
AGE	<= 18	-		-	-	+			-	+		+
	19 - 28											
	29 - 38	+		+	+	-			+	-		-
	39 - 48	+		+	+	-			+	-		-
	49 - 58	+		+	+	-			+	-		-
	59 - 68	-		-	-	+			-	+		+
	69 - 78	-		-	-	+			-	+		+
	79 - 88											
	89+	-		-	-	+			-	+		+
Tour de Taille	<= 47	-		-	-	+			-	+		+
	48 - 62	-		-	-	+			-	+		+
	63 - 73	-		-	-	+			-	+		+
	74 - 83											
	84 - 98	+		+	+	-			+	-		-
	99 - 114	-		-	-	+			-	+		+
	115+	-		-	-	+			-	+		+
PAS	<= 83	+		+	+	-			+	-		-
	84 - 102	+		+	+	-			+	-		-
	103 - 116											
	117 - 134											
	135 - 149	-		-	-	+			-	+		+
	150 - 164	-		-	-	+			-	+		+
	165 - 179	-		-	-	+			-	+		+
	180+	-		-	-	+			-	+		+
IMC	<= 16.4382											
	16.4383 - 21.1403											
	21.1404 - 26.0103	+		+	+	-			+	-		-
	26.0104 - 30.8803	+		+	+	-			+	-		-
	30.8804 - 35.5824	-		-	-	+			-	+		+
	35.5825 - 40.4524	-		-	-	+			-	+		+
	40.4525 - 45.0000	-		-	-	+			-	+		+
	45.0001 - 50.0000	-		-	-	+			-	+		+
	50.0001+	-		-	-	+			-	+		+
PAD	<= 50	+		+	+	-			+	-		-
	51 - 60											
	61 - 70	+		+	+	-			+	-		-
	71 - 80	+		+	+	-			+	-		-
	81 - 90											
	91+	-		-	-	+			-	+		+

	Wilaya	Wilaya-7	Wilaya-16	Wilaya-19	Wilaya-20	Wilaya-21	Wilaya-23	Wilaya-24	Wilaya-25	Wilaya-26	Wilaya-31	Wilaya-46
Type de diabète	Diabétique	+	-	-				+	+	+		-
	non diabétique	-	+	+				-	-	-		+
Sexe	femme	-	+	+				-	-	-		+
	homme	+	-	-				+	+	+		-
Antécédent familial	absence	+	-	-				+	+	+		-
	présence	-	+	+				-	-	-		+
AGE	<= 18	+	-	-				+	+	+		-
	19 - 28	-	+	+				-	-	-		+
	29 - 38	-	+	+				-	-	-		+
	39 - 48	+	-	-				+	+	+		-
	49 - 58											
	59 - 68											
	69 - 78	+	-	-				+	+	+		-
	79 - 88											
89+												
Tour de Taille	<= 47	+	-	-				+	+	+		-
	48 - 62	+	-	-				+	+	+		-
	63 - 73											
	74 - 83	-	+	+				-	-	-		+
	84 - 98	+	-	-				+	+	+		-
	99 - 114	-	+	+				-	-	-		+
	115+	-	+	+				-	-	-		+
PAS	<= 83	+	-	-				+	+	+		-
	84 - 102	+	-	-				+	+	+		-
	103 - 116	-	+	+				-	-	-		+
	117 - 134											
	135 - 149	-	+	+				-	-	-		+
	150 - 164	+	-	-				+	+	+		-
	165 - 179	-	+	+				-	-	-		+
	180+											
IMC	<= 16.4382											
	16.4383 - 21.1403	+	-	-				+	+	+		-
	21.1404 - 26.0103											
	26.0104 - 30.8803	+	-	-				+	+	+		-
	30.8804 - 35.5824	-	+	+				-	-	-		+
	35.5825 - 40.4524	-	+	+				-	-	-		+
	40.4525 - 45.0000	-	+	+				-	-	-		+
	45.0001 - 50.0000	-	+	+				-	-	-		+
50.0001+	-	+	+				-	-	-		+	
PAD	<= 50	+	-	-				+	+	+		-
	51 - 60	+	-	-				+	+	+		-
	61 - 70	+	-	-				+	+	+		-
	71 - 80	-	+	+				-	-	-		+
	81 - 90											
	91+	-	+	+				-	-	-		+

	AXE 1	AXE 2	AXE 3	AXE 4
Wilaya-7	+	+	+	+
Wilaya-16	+	+		-
Wilaya-19	-		+	-
Wilaya-20		-	+	
Wilaya-21	+	-	-	
Wilaya-23		-		
Wilaya-24	+	+		+
Wilaya-25	+		+	+
Wilaya-26	-	-	-	+
Wilaya-31				
Wilaya-46		+	-	-

Objectifs

- *Considérer les données médico-administrative comme « patrimoine »;*
- -Mettre à la disposition des acteurs de la santé d'un territoire les informations dans un souci d'une meilleure prise de décision en matière :
 - de prévention,
 - d'affectation des ressources matérielles et humaines;
- Exploiter les données médico administratives numérisées en vue de tirer les meilleures connaissances de la santé de la population au sein du territoire concerné;
- Concevoir des systèmes d'information géographique innovants dans une optique de réseaux de soins;
- Optimiser l'usage des ressources par la disponibilité de l'information.

Caractéristiques d'un territoire

- Le territoire comme système complexe
(*Problèmes de frontières*);
- Les collectivités doivent gérer la santé de la population vivant dans l'espace qui correspond à leurs prérogatives;
- Pour mieux gérer, ces données sont tributaires de l'information disponible;
- Le territoire doit être appréhendé par approche systémique.

Conclusion:

- Les règles d'associations diffèrent d'une région à une autre, et c'est dans cette différence que réside l'importance de la fouille de données qui nous a permis d'extraire des règles importantes ;
- Il est important d'adjoindre d'autres variables pour mieux cerner le détail des informations :
Climat, mode de vie, ruralité ; le sol, catégorie socio-professionnelle.....;