

EAR207 : ANALYSE DE DONNÉES ET DÉCISIONS DANS L'ENTREPRISE

MÉMOIRE D'ANALYSE EXPLORATOIRE DE DONNÉES

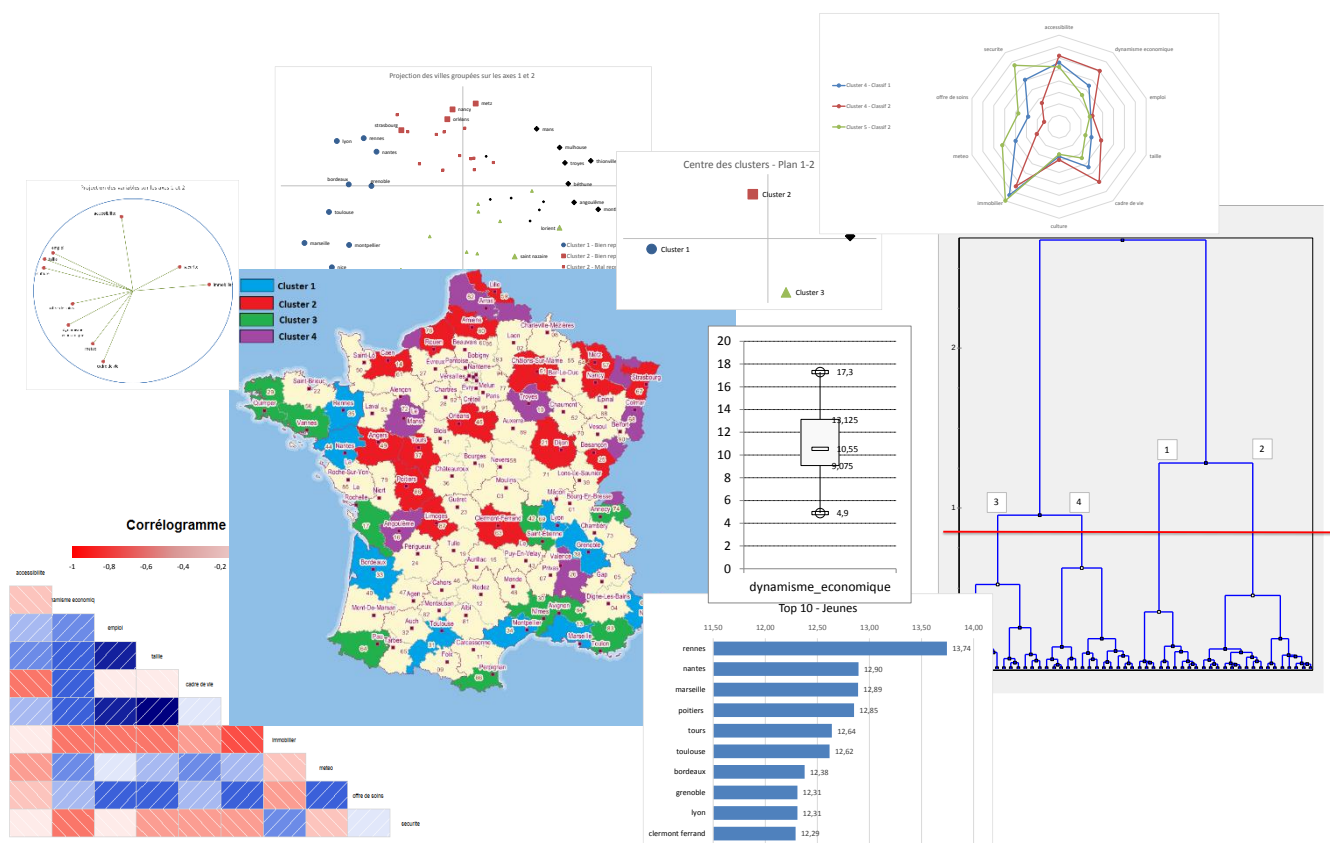
APPLICATION AUX 50 AGGLOMÉRATIONS DE PROVINCE LES PLUS PEUPLÉES EN FRANCE

Réalisé par : Mohamed Amine EL AFRIT

Encadré par : Mathilde Guergoat-Larivière

<http://www.mohamedelafrit.com/ear207>

Juin 2015



1 REFERENCES ET DOCUMENTS CONNEXES

1.1 Objet du document

L'objectif de ce travail est d'analyser une base de données des 50 agglomérations de province les plus peuplées en France. Ces villes sont décrites avec quelques indices sociaux et économiques permettant de réaliser des classements.

Ce document a pour objet de synthétiser le travail d'analyse réalisé dans le classeur Excel joint.

1.2 Références et documents connexes

Les documents suivants ont été utilisés pour ce projet et sont joint à ce rapport. Ces documents et ce rapport sont disponibles à l'adresse suivante : <http://www.mohamedelafrir.com/ear207>

Intitulé	Document	Version	Nom du fichier
Cours EAR206 - 2015 – Gilbert LAFFOND Cours EAR207 - 2015 – Mathilde Guergoat-Larivière		2015	
Analyse de la base quantitative avec Tanagra	 EAR207-ACP-Classification-EL_AFRIT.bdm		EAR207-ACP-Classification-EL_AFRIT.bdm
Analyse de la base qualitative (variables à 2 modalités) avec Tanagra	 EAR207-ACM-2-modalités-EL_AFRIT.bdm		EAR207-ACM-2-modalités-EL_AFRIT.bdm
Analyse de la base qualitative (variables à 3 modalités) avec Tanagra	 EAR207-ACM-3-modalités-EL_AFRIT.bdm		EAR207-ACM-3-modalités-EL_AFRIT.bdm
Analyse et mise en forme avec Excel	 EAR207-Mémoire-EL_AFRIT.xlsx	Excel 2013	EAR207-Mémoire-EL_AFRIT.xlsx
Réalisation du corrélogramme avec R	 EAR207-Mémoire-EL_AFRIT.R		EAR207-Mémoire-EL_AFRIT.R
Cartes de France réalisées avec Géoconcept	 France-de-base00.gcm	6.6	France-de-base00.gcm
Outils et macros d'analyse de données	 DataAnalysisTools.xlsm	Excel 2013	DataAnalysisTools.xlsm

SOMMAIRE

1	Références et documents connexes	2
1.1	Objet du document	2
1.2	Références et documents connexes	2
	Sommaire	3
2	Données étudiées et démarche d'analyse	5
2.1	Démarche d'analyse	5
2.2	Base de données initiale (Quantitative)	6
3	Exploration préliminaire de la base	7
3.1	Description des données de base	7
3.2	Description des individus	7
3.3	Description des variables	8
3.4	Classement des villes par critères	11
3.5	Classement des villes par catégorie de personne	11
3.5.1	Palmarès des villes où il fait bon vivre pour les jeunes	12
3.5.2	Palmarès des villes où il fait bon vivre pour les familles	12
3.5.3	Palmarès des villes où il fait bon vivre pour les seniors	13
4	Analyse en composante principale	14
4.1	Valeurs propres	14
4.2	Analyse des variables	14
4.2.1	Corrélations entre les variables	14
4.2.2	Contributions et qualité de projection des variables	15
4.2.3	Contributions et qualités de projection des variables	16
4.2.4	Cercles de corrélation et interprétation des axes	17
4.3	Analyse des individus	19
4.4	Projection sur les plans factoriels	22
4.4.1	Projection sur le plan 1-2	22
4.4.2	Projection sur le plan 1-3	23
4.4.3	Projection sur le plan 1-5	24
4.4.4	Projection sur le plan 2-4	25
5	Classification des villes	27
5.1	Dendrogrammes	27
5.2	Classification en 4 clusters	27
5.3	Classification en 5 clusters	29
5.4	Classification en 8 clusters	30
6	ACP et classification	31

6.1	ACP et classification en 4 clusters	31
6.2	ACP et classification en 5 clusters	35
6.3	Comparaison des classifications	35
7	Analyse des correspondances multiples	37
7.1	Bases de données transformées (Qualitatives)	37
7.1.1	Variables qualitatives à deux modalités	37
7.1.2	Variables qualitatives à trois modalités	38
7.2	Dendrogrammes	39
7.3	ACM et classification avec variables à 2 modalités	39
7.4	ACM et classification avec variables à 3 modalités	39
	Conclusion et perspectives	40

2 DONNEES ETUDIEES ET DEMARCHE D'ANALYSE

2.1 Démarche d'analyse

Les grandes étapes de la démarche suivie dans ce travail sont :

- Exploration préliminaire de la base
- Analyse en composantes principales
- Classification CAH
- Analyse en composantes principales avec classification
- Analyse des correspondances multiples

L'exploration préliminaire de la base consiste d'abord à comprendre le contexte et à décrire les données à l'aide d'outils statistiques descriptives de base (moyennes, écart-type, médianes, quartiles, étendus, histogrammes, boîtes à moustaches etc.).

L'analyse en composantes principales consiste d'abord à observer les corrélations entre les variables et les valeurs propres pour identifier le nombre d'axes factoriels pertinents à retenir. Ensuite, on analysera les projections des variables sur les axes et plans factoriels pour identifier les significations des axes. Après l'interprétation des axes, on analysera les projections des individus (ici les villes) en identifiant les individus bien représentés et les individus aberrants. On pourra donc interpréter et mieux comprendre les caractéristiques des villes.

L'analyse de la classification CAH, consiste d'abord à réaliser le dendrogramme et à choisir un nombre pertinent de clusters à retenir, ensuite à identifier, pour chaque cluster, les variables dont le test-value est plus grand que 2 ou plus petit que -2 pour caractériser ce groupe. On utilisera des outils de visualisation pour observer les profils des clusters.

L'analyse simultanée de l'ACP et de la classification consiste à faire une synthèse et de vérifier la cohérence des deux analyses précédentes en représentant les individus sur les plans factoriels retenus et en mettant en évidence les clusters identifiés pendant l'étape de classification.

L'analyse en correspondances multiples consiste d'abord à transformer les variables quantitatives en variables qualitatives. On fera deux transformations : une transformation avec des variables qualitatives à 2 modalités et une transformation avec des variables à 3 modalités. Faute de temps, l'analyse en correspondances multiple n'est pas terminée encore mais le début a été fait avec Tanagra et dans le fichier Excel ci-joint.

2.2 Base de donnée initiale (Quantitative)

ville	accessibilite	dynamisme economique	emploi	taille	cadre de vie	culture	immobilier	meteo	offre de soins	securite
amiens	14,3	9,3	7,2	12,4	10,5	8,4	7,6	4,5	12,6	8,3
angers	12,4	11,5	12,2	11,6	8,8	10,6	9,7	9,3	14,9	19,5
angoulême	9,7	5,9	3,2	4,8	9,4	3,8	18,2	12,9	11	16,6
annecy	5,5	8,3	10,8	6,2	8,6	7	1,7	10,7	14,2	12,7
annemasse	8,5	8,5	7,4	3,2	7,4	1,5	5,6	10,3	5,8	12,7
avignon	9,6	14,8	7,9	10,2	12	13,7	8,8	18,7	8,4	5,6
bayonne	5,6	9,5	8,5	4,6	18,4	9,7	6	9,3	17	16,3
besançon	8,9	10,9	9,4	12,6	5,9	8,6	12,9	7,6	19	14,7
béthune	13,7	10,5	3,3	5,6	11	2,4	17,2	4,7	2,3	8,5
bordeaux	9,5	14,5	15,9	19,2	11	16,6	3,9	11,3	18,1	7,4
brest	6,8	9,5	11,5	7	14,4	7,4	18,9	0,4	10,7	17,9
caen	11,7	11,3	12,3	13,2	11,3	13,8	5,2	5,3	17,8	12,7
clermont ferrand	9,5	7,9	15,4	15,6	9,4	10,2	13,2	14,9	16,8	15,6
dijon	12,1	9,5	14,6	14,2	4,8	11,3	7,8	10,5	13,9	15,7
douai-lens	14,5	13,2	8,2	12,4	9,9	10,1	14,6	4,7	2	5,4
dunkerque	11,9	9,5	4,7	5,2	14,6	3,9	12,9	4,7	3	3
grenoble	9,9	12,5	17,5	14,8	10,3	15,9	4,8	15,9	15,8	9,2
havre	10,3	9,8	7,3	6,8	14,9	8,6	10,6	1,7	3,9	5,7
Lille	18,3	10,4	12,5	15,8	10,3	18,9	2,4	4,7	6,6	3
limoges	9,2	10,5	10,4	12	7,5	11,9	17,6	9,1	11,4	18,5
lorient	5,6	9	6,2	3,6	14,5	6,1	15,3	6,8	6,5	19
lyon	14,4	12,9	18,5	20	8,3	19,8	3,2	14,7	10,1	2,6
mans	15,6	5,8	10,7	9,2	7,5	5,2	17,6	9,7	9,7	17
marseille	10,7	15,1	13,8	19,8	18,9	19,7	2	20	17,1	0,4
metz	15,1	10,2	12,3	15,8	5,9	11,9	14,6	3,9	7,5	14,1
montbéliard	5,5	10,1	2,4	3	5,1	3,5	19,5	8,7	3,3	14,7
montpellier	8,7	14,4	10,7	16	14,8	16,5	2,8	18,7	20	3,7
mulhouse	10,3	4,9	6,9	6	3,1	10,2	18,4	10,5	4,9	12,4
nancy	14,3	6,8	13,7	12,6	5,3	15,6	10,8	6,5	13	10,1
nantes	13,9	16,5	17,9	15	11,2	16,6	5,9	7,6	9,8	9,8
nice	6,8	14,2	15,5	14,8	19,5	16,7	0,4	18,7	15,2	0,8
nîmes	8,1	14,2	3,6	8,6	14	3,4	14,1	18	19,7	3,9
orléans	15,1	8,6	14,7	14,6	6,3	12,1	10	9,5	8,2	8,8
pau	4,9	10,3	10,2	8,2	12,7	8,2	11,3	9,3	9	16,3
perpignan	5,7	12,9	4,9	8,6	16,2	7	14,3	18,4	16,4	5
poitiers	13,9	11,1	11,1	10,6	8,6	11,5	15,2	13,6	14,6	18,3
reims	15,9	5,3	11,1	11,6	7,3	12,1	9,6	9,2	12,3	5,5
rennes	13,2	14,9	17,9	18,2	10,1	17,1	5,4	6,8	17,8	14,9
rochelle	8,7	10,6	4,7	5,8	15,9	9,3	1,2	14,5	13,8	12,9
rouen	14	10,8	12,7	17,6	11	16,6	5,2	1,7	9,4	5,7
saint etienne	9,1	12,7	10,3	9,6	9,4	12,2	19,2	15,9	8,5	10,9
saint nazaire	7,9	14,2	5,5	3,2	15,1	6,5	11,5	12,1	3,6	9,8
strasbourg	12,3	11,3	14,5	15,6	5,4	17,3	5,6	8,5	15,5	12,1
thionville	11,3	6,6	3,6	2,2	5	2,6	13	3,9	7,2	14,1
toulon	6,3	14,5	11,8	12,8	18,5	13,5	8,6	19,3	9,7	6,3
toulouse	6,5	16,4	16,2	19,4	9,8	18,6	5,3	16	18,4	4,7
tours	16,4	11,4	14,4	11,2	7,8	12,8	9,8	11,9	14,2	14,6
troyes	13,2	6,5	4,6	5,2	5,7	5,1	17,6	10,3	6,8	9,9
valence	9,5	6,4	6,3	5	10,9	6,7	17,8	17,2	11,3	8,7
valenciennes	11,6	17,3	7,1	8,6	9,1	3,9	9,1	4,7	2,6	3

Figure 1 : Base de donnée initiale

Il est nécessaire de s'assurer que le séparateur décimal utilisé est la virgule et non pas le point avant d'importer cette base dans Tanagra pour que les variables quantitatives soient considérées comme variables continues.

3 EXPLORATION PRELIMINAIRE DE LA BASE

3.1 Description des données de base

Il s'agit des résultats d'une enquête portant sur les 50 agglomérations de province les plus peuplées en France. Toutes ces structures intercommunales (communauté urbaine ou d'agglomération) comptent au moins 100 000 habitants rassemblés autour d'une ville centre d'au moins 50 000 habitants. La signification des variables et les valeurs possibles sont expliquées dans le Tableau 1.

Tableau 1 : Données étudiées

Variables	Type de variable	Unité	Valeurs possibles	Signification de la variable
accessibilite	Continue	Sans unité	[0,4 - 20]	Infrastructure et réseau de transport ferroviaire ou routière et liaison avec les autres villes
dynamisme_economique	Continue	Sans unité	[0,4 - 20]	Décrit le dynamisme économique dans la ville
emploi	Continue	Sans unité	[0,4 - 20]	Les offres et perspectives d'emploi
taille	Continue	Sans unité	[0,4 - 20]	Taille de la ville
cadre_de_vie	Continue	Sans unité	[0,4 - 20]	Les espaces verts
culture	Continue	Sans unité	[0,4 - 20]	L'offre culturelle de la ville
immobilier	Continue	Sans unité	[0,4 - 20]	Les prix de l'immobilier (prix élevé ==> note faible)
meteo	Continue	Sans unité	[0,4 - 20]	Caractéristiques du climat
offre_de_soins	Continue	Sans unité	[0,4 - 20]	L'offre des soins
securite	Continue	Sans unité	[0,4 - 20]	Niveau de sécurité dans la ville (délinquance élevée ==> note faible)

3.2 Description des individus

Les provinces étudiées sont mis en évidence en bleu sur la carte de la figure Figure 2.

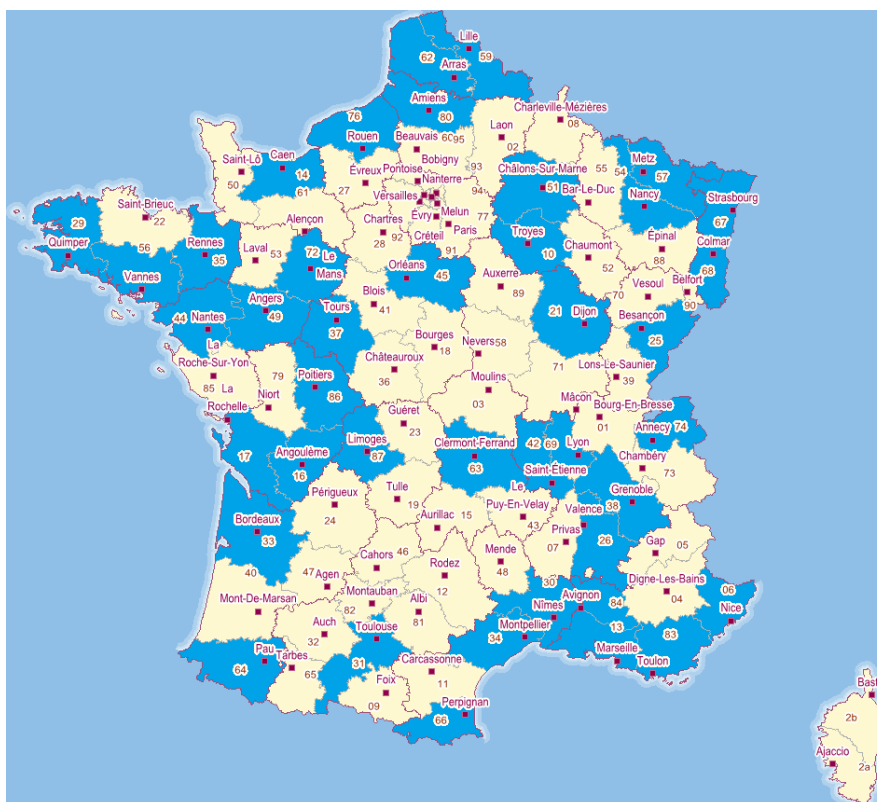


Figure 2 : Carte des provinces étudiées

On complète la base de données initiale avec la région géographique de chaque ville étudiée de la manière suivante :

- NO pour les villes situées dans le Nord-Ouest
- NE pour les villes situées dans le Nord-Est
- SO pour les villes situées dans le Sud-Ouest
- SE pour les villes situées dans le Sud-Est



Table 1 : Nombre de villes étudiées par région

Régions	Nbre de villes étudiée
NE	16
NO	11
SE	15
SO	8

On constate que la majorité des villes de l'échantillon étudiées sont dans l'est de la France : 16 dans le Nord-Est et 15 dans le Sud-Est contre 19 dans l'Ouest.

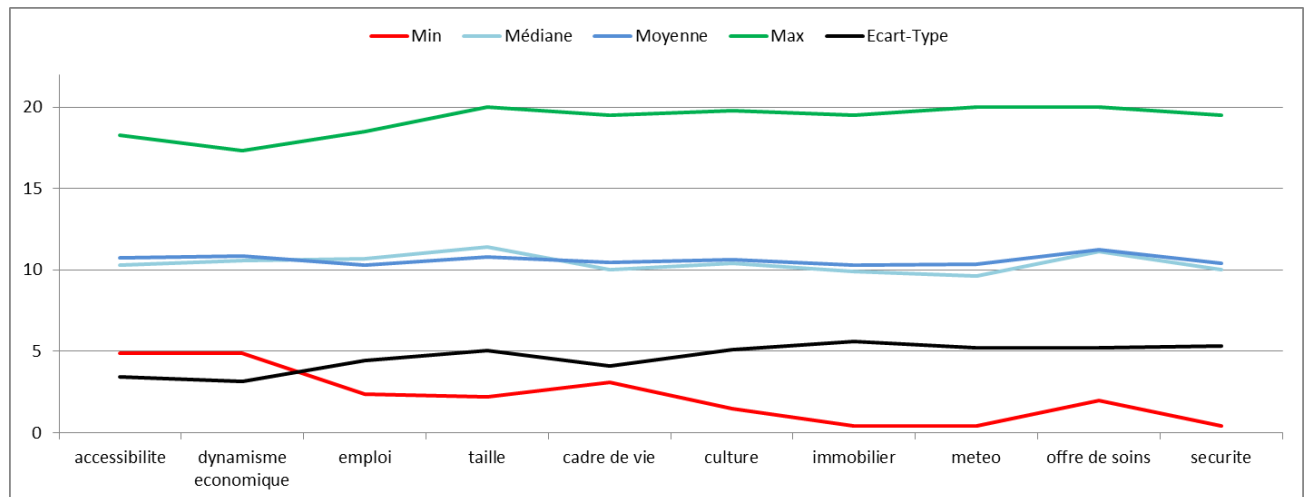
3.3 Description des variables

Pour décrire les variables, on a utilisé d'abord les fonctions « Univariate continius stat » et « More univariate cont stat » de Tanagra. On obtient les résultats du Tableau 2.

Tableau 2 : Statistiques de base

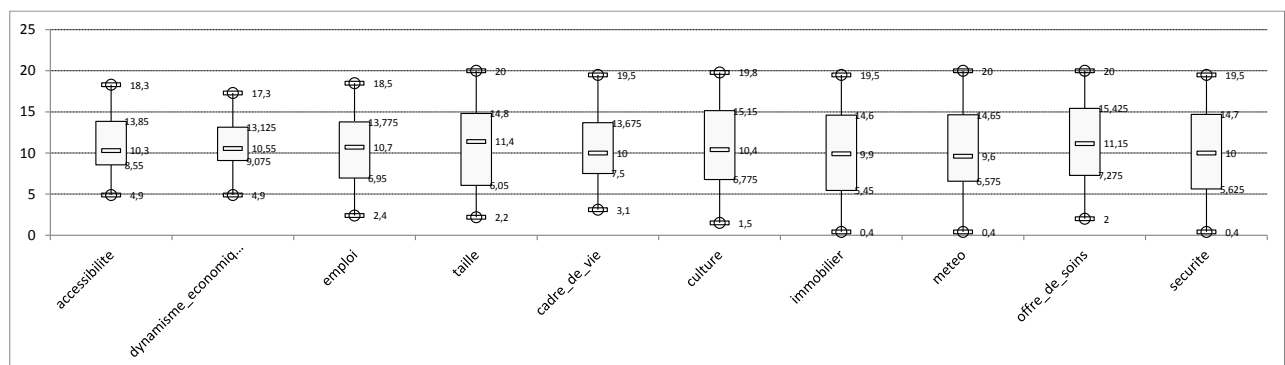
Variabes	Min	Max	Moyenne	Médiane	Ecart-type	Ecart-type / Moyenne
accessibilite	4,9	18,3	10,728	10,3	3,4198	0,3188
dynamisme_economique	4,9	17,3	10,874	10,55	3,129	0,2878
emploi	2,4	18,5	10,28	10,7	4,4551	0,4334
taille	2,2	20	10,796	11,4	5,0699	0,4696
cadre_de_vie	3,1	19,5	10,466	10	4,0844	0,3903
culture	1,5	19,8	10,652	10,4	5,1252	0,4812
immobilier	0,4	19,5	10,278	9,9	5,6175	0,5466
meteo	0,4	20	10,356	9,6	5,242	0,5062
offre_de_soins	2	20	11,226	11,15	5,1965	0,4629
securite	0,4	19,5	10,38	10	5,3516	0,5156

Comme montre le Tableau 2 et le Graphique 1, les variables étudiées ont sensiblement les mêmes moyennes et médianes (à peu près égales à 10). Les écarts types sont aussi de même ordre de grandeur (de l'ordre de 5). Les valeurs prises par les variables sont généralement situées entre 0,4 et 20. On remarque juste que les valeurs minimales pour les variables « accessibilités » et « dynamisme économique » sont plus élevées (proche de 5) ce qui est normal pour un pays développé.

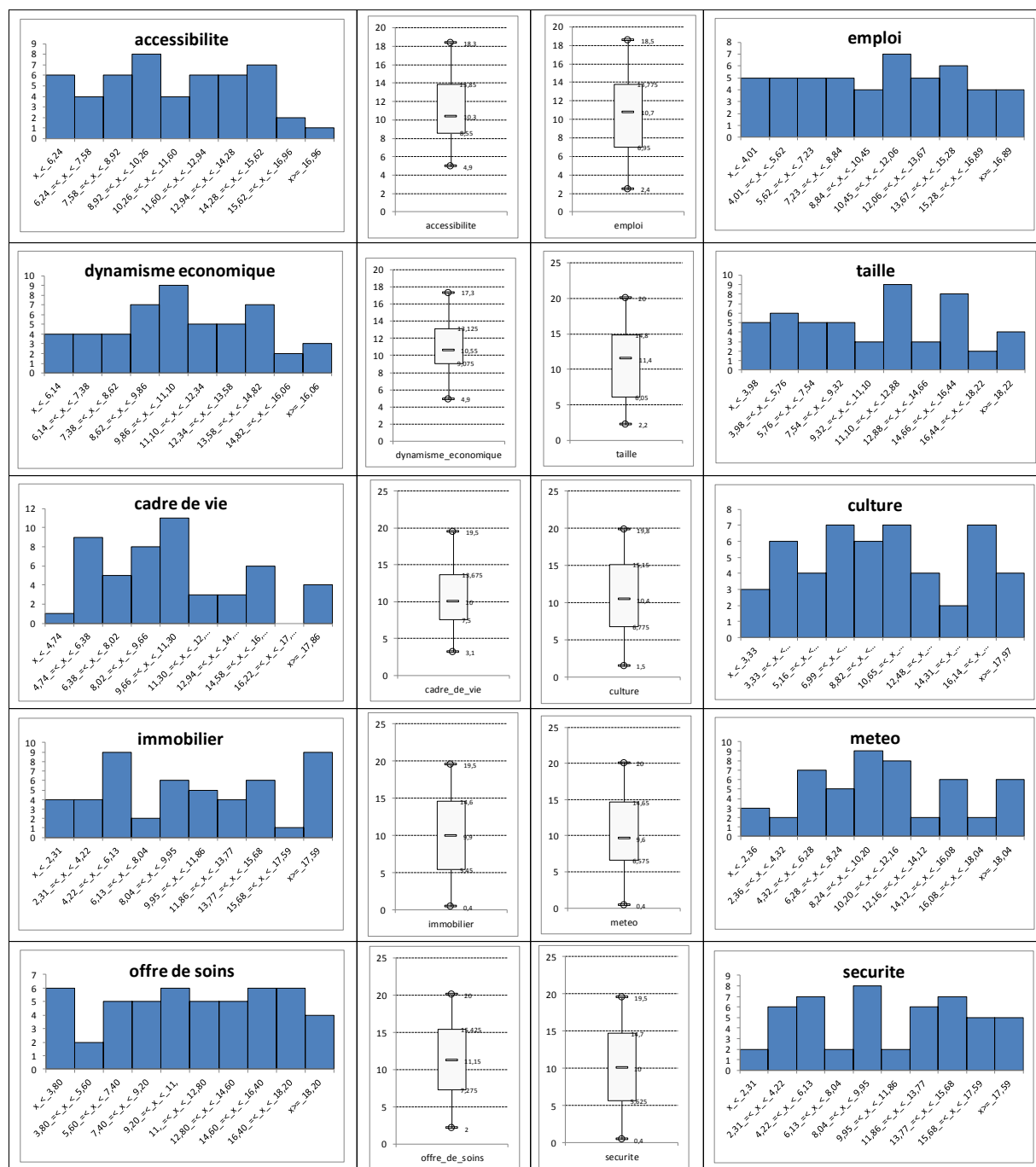


Graphique 1 : Comparaison des statistiques de base des variables

On représente la dispersion des variables en utilisant la macro « boîte à moustache » disponible dans les outils joint à cette étude. On obtient le Graphique 2. On constate que la dispersion est plus ou moins régulière dans l'intervalle [0,4 – 20] pour l'ensemble des variables et il n'y a pas d'aberration. Ceci est normal car par construction les variables étudiées sont construites artificiellement à partir des indicateurs de l'enquête en attribuant la meilleure note de 20 à la ville qui termine en tête, ensuite 19,6 à la ville suivante jusqu'à 0,4 à la dernière ville du classement.



Graphique 2 : Dispersion des variables



Graphique 3 : Description des variables

La description des variables avec des statistiques classiques comme dans le Graphique 3, ne nous apporte pas grand-chose car par construction, ces variables sont construites pour classer les villes.

3.4 Classement des villes par critères

On classe les villes en prenant une seule variable à chaque fois. Lille est la meilleure ville accessible de province. Valenciennes est la meilleure ville d'un point de vue dynamisme économique. Lyon est la première ville où il y a des opportunités d'emploi c'est aussi la plus grande ville et la meilleure ville culturelle. Le cadre de vie est plus agréable à Nice. Le climat est mieux apprécié à Marseille. L'immobilier est moins cher à Montbéliard. L'offre de soin est plus performante à Montpellier. Enfin, Angers est la ville où il y a moins de délinquance.



3.5 Classement des villes par catégorie de personne

Maintenant on va expliquer la méthode de mise en place d'un palmarès des villes. Comme on ne dispose pas d'informations sur les personnes interrogées pour réaliser cette enquête, on va faire des hypothèses. On suppose qu'il y a 3 classements selon 3 types de préférences : préférences des jeunes, préférence des familles et préférences des seniors (personnes âgés ou retraitées).

On traduit les préférences en accordant des poids plus ou moins grands pour traduire l'importance de ces variables pour une certaine catégorie des personnes. Pour simplifier, un critère sera considéré comme « Important » (poids 3) ou bien « Moyen » (poids 2) ou bien « Négligé » (poids 1).

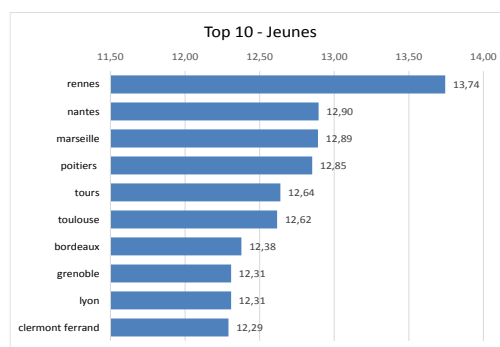
Les calculs sont réalisés dans la feuille « Data base » et présentées dans la feuille « Descrip-Palmarès » du document Excel joint à ce rapport.

3.5.1 Palmarès des villes où il fait bon vivre pour les jeunes

Tableau 3 : Ordre de priorité des critères pour les jeunes

Graphique 4 : Top 10 des meilleurs choix pour les jeunes

Critères	Signification	Importance des critères	Coeff
Emploi	Bonne perspectives d'emploi	Important	3
dynamisme économique			3
Immobilier	Prix de l'immobilier attractifs		3
Culture	Activités culturelles		3
Accessibilité	La possibilité de rejoindre facilement les autres villes		3
cadre de vie		Moyen	2
offre de soins			2
sécurité			2
Taille	La taille des villes n'est pas un critère	Négligé	1
météo			1



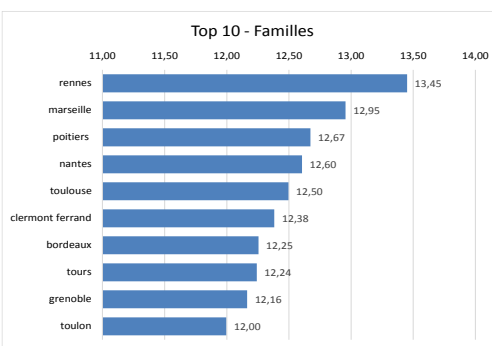
Pour les jeunes, on suppose que les perspectives d'emploi, les prix de l'immobilier, l'offre culturelle et la possibilité de rejoindre facilement les autres villes sont des critères importants. Le cadre de vie, l'offre de soins et la sécurité sont moyennement importants et la taille de la ville et la météo ne sont pas importants. Avec ces hypothèses, Rennes vient en tête du classement des villes.

3.5.2 Palmarès des villes où il fait bon vivre pour les familles

Tableau 4 : Ordre de priorité des critères pour les familles

Graphique 5 : Top 10 des meilleurs choix pour les familles

Critères	Signification	Importance des critères	Coeff
cadre de vie	Cadre de vie agréable	Important	3
emploi	Offres d'emploi		3
Immobilier	Prix de l'immobilier attractifs		3
dynamisme économique			3
Culture		Moyen	2
offre de soins			2
accessibilité			2
sécurité			2
météo		Négligé	1
taille			1



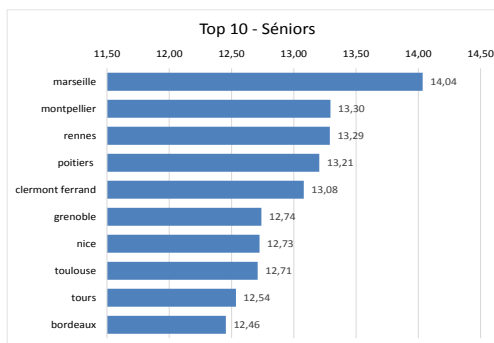
Pour les familles, le cadre de vie, l'emploi, le dynamisme économique et le prix de l'immobilier sont des critères importants. La culture, la santé, l'accessibilité et la sécurité sont d'une importance moyenne et la météo et la taille de la ville sont négligeables pour cette catégorie de personnes. Avec ces hypothèses, Rennes est aussi en tête du classement.

3.5.3 Palmarès des villes où il fait bon vivre pour les seniors

Tableau 5 : Ordre de priorité des critères pour les seniors

Graphique 6 : Top 10 des meilleurs choix pour les seniors

Critères	Signification	Importance des critères	Coeff
cadre de vie	cadre de vie agréable	Important	3
securite	délinquance faible		3
offre de soins	offre de soins performante		3
Météo	Conditions climatiques agréables		3
accessibilite	accessibilité ferroviaire ou routière	Moyen	2
culture	offre culturelle		2
dynamisme economique		Négligé	1
taille			1
Immobilier			1
Emploi			1



Pour les seniors, on suppose qu'un cadre de vie agréable, un faible niveau de délinquance, et une offre de soins performante sont des critères importants. L'accessibilité et l'offre de culture sont moyennement importantes. Enfin, pour les personnes âgées, on suppose que le dynamisme économique, la taille de la ville, le prix de l'immobilier et les offres d'emploi ne sont pas importants. Avec ces hypothèses, Marseille est considérée comme la meilleure ville où il fait bon vivre pour les seniors.

4 ANALYSE EN COMPOSANTE PRINCIPALE

4.1 Valeurs propres

Tanagra a permis de calculer les valeurs propres de la matrice de corrélation du Tableau 7.

Tableau 6 : Valeurs propres

Axis	Eigen value	Difference	Proportion (%)	Cumulative (%)
1	4,100377	1,954109	41,00%	41,00%
2	2,146269	0,827923	21,46%	62,47%
3	1,318346	0,600376	13,18%	75,65%
4	0,71797	0,147369	7,18%	82,83%
5	0,5706	0,203354	5,71%	88,54%
6	0,367246	0,02222	3,67%	92,21%
7	0,345026	0,096473	3,45%	95,66%
8	0,248552	0,123675	2,49%	98,14%
9	0,124877	0,064141	1,25%	99,39%
10	0,060737	-	0,61%	100,00%
Tot.	10	-	-	-

D'après la règle de Kaiser, on garde les axes dont la valeur propre est supérieure à 1. Donc avec cette règle on garde au moins les 3 premiers axes factoriels pour interpréter les résultats.

D'après la règle du coude, on garde les valeurs propres avant l'apparition du coude en se basant sur le graphique associé aux valeurs propres (ce qui est matérialisé ici par l'histogramme dans la colonne « Proportion »). Il est difficile de voir apparaître un coude dans notre cas, donc plus on prend en compte des axes mieux l'interprétation sera proche de la réalité.

Nous allons interpréter les résultats obtenus jusqu'à l'axe 5 dans cette étude car comme nous allons le voir dans l'analyse des variables, certaines variables seront mieux représentées par les axes 4 et 5 (Cf. Tableau 8).

Avec les 3 premiers axes, il est possible d'expliquer au moins 75% de l'inertie totale et avec les 5 premiers axes, il est possible d'en expliquer 88%.

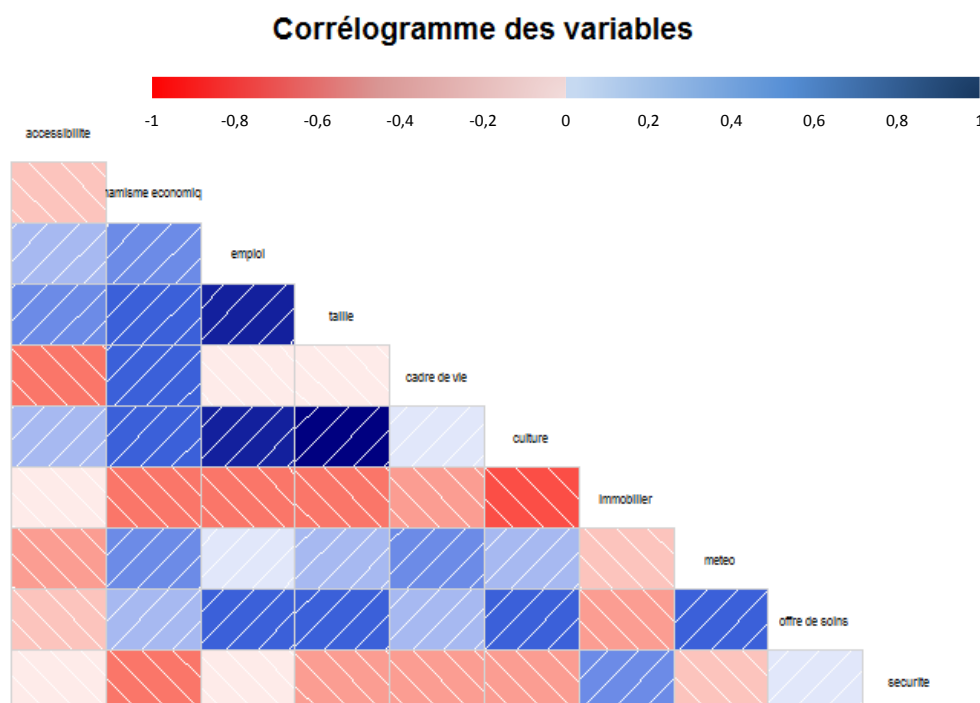
4.2 Analyse des variables

4.2.1 Corrélations entre les variables

L'analyse en composantes principales avec Tanagra a permis de générer la matrice de corrélation suivante :

Tableau 7 : Matrice de corrélation

	accessibilité	dynamisme économique	emploi	taille	cadre de vie	culture	immobilier	meteo	offre de soins	securite
accessibilité	1	-0,16707	0,27883	0,35979	-0,45753	0,25948	-0,0441	-0,38029	-0,17936	-0,12211
dynamisme économique	-0,16707	1	0,3351	0,47655	0,45223	0,44086	-0,44214	0,30793	0,22142	-0,44708
emploi	0,27883	0,3351	1	0,84364	-0,07052	0,83085	-0,5407	0,10996	0,46183	-0,06687
taille	0,35979	0,47655	0,84364	1	-0,01035	0,87343	-0,53762	0,16959	0,50065	-0,34288
cadre de vie	-0,45753	0,45223	-0,07052	-0,01035	1	0,10241	-0,31839	0,33193	0,175	-0,37126
culture	0,25948	0,44086	0,83085	0,87343	0,10241	1	-0,61273	0,23689	0,45299	-0,3218
immobilier	-0,0441	-0,44214	-0,5407	-0,53762	-0,31839	-0,61273	1	-0,18229	-0,42374	0,41837
meteo	-0,38029	0,30793	0,10996	0,16959	0,33193	0,23689	-0,18229	1	0,45693	-0,27308
offre de soins	-0,17936	0,22142	0,46183	0,50065	0,175	0,45299	-0,42374	0,45693	1	0,06095
securite	-0,12211	-0,44708	-0,06687	-0,34288	-0,37126	-0,3218	0,41837	-0,27308	0,06095	1



Graphique 7 : Visualisation de la matrice de corrélation à l'aide du corrélogramme

Le corrélogramme du Graphique 7 permet de visualiser les variables corrélées positivement et négativement et l'intensité de cette corrélation avec des codes couleurs. Les cellules coloriées en bleu montrent les variables corrélées positivement, celles qui sont en rouge montrent les corrélations négatives. Plus la couleur est foncée, plus la corrélation est grande en valeur absolue. Ce graphique a été obtenu avec le logiciel R. Les détails sont expliqués dans la feuille « Outils d'analyse » du document Excel. Il est possible d'exécuter directement le code en ligne sur le site <http://www.r-fiddle.org>.

Comme le montre la matrice de corrélation et le corrélogramme, il y a beaucoup de corrélations positives et négatives entre les variables. En particulier, les variables « Emploi », « Taille » et « Culture » sont très corrélées positivement (plus que 0.8) et sont corrélées négativement avec la variable « Immobilier ». L'offre de soins est corrélée positivement avec la météo, la culture la taille de la ville et l'offre d'emploi. La sécurité est corrélée positivement avec le prix d'immobilier mais négativement avec le dynamisme économique, la taille de la ville, le cadre de vie et la culture. L'accessibilité est corrélée négativement avec le cadre de vie ce qui laisse penser à que plus il y a des espaces verts dans la ville, moins cette ville sera accessible ce qui semble logique pour les campagnes par exemple.

4.2.2 Contributions et qualité de projection des variables

Tableau 8 : Identification des meilleurs plans pour interpréter les différentes dimensions (variables)

Variable	Meilleurs plans de projections			Cos²										CTR			
	1er plan	2ème plan	3ème plan	1-2	1-3	1-4	1-5	2-3	2-4	2-5	3-4	3-5	4-5	1-2	1-3	1-5	2-4
accessibilité	2-3	2-4	1-2	0,664	0,142	0,056	0,029	0,779	0,692	0,666	0,171	0,144	0,058	31%	10%	3%	36%
dynamisme économique	1-5	1-2	1-3	0,556	0,495	0,449	0,667	0,210	0,164	0,382	0,102	0,321	0,275	17%	16%	54%	10%
emploi	1-2	1-3	1-4	0,817	0,691	0,662	0,660	0,215	0,186	0,184	0,060	0,058	0,029	24%	19%	18%	10%
taille	1-2	1-5	1-4	0,909	0,789	0,793	0,807	0,120	0,124	0,137	0,004	0,017	0,021	25%	19%	22%	6%
cadre de vie	1-2	2-4	2-3	0,674	0,130	0,170	0,105	0,626	0,667	0,601	0,123	0,057	0,097	29%	5%	5%	39%
culture	1-2	1-5	1-3	0,868	0,805	0,805	0,807	0,064	0,063	0,065	0,001	0,003	0,002	23%	20%	20%	3%
immobilier	1-5	1-4	1-3	0,592	0,602	0,662	0,776	0,020	0,080	0,194	0,090	0,204	0,265	15%	15%	47%	11%
météo	2-4	1-4	1-2	0,492	0,249	0,522	0,164	0,416	0,689	0,330	0,446	0,087	0,361	19%	11%	4%	65%
offre de soins	1-3	2-3	3-4	0,386	0,786	0,368	0,411	0,438	0,020	0,063	0,420	0,464	0,046	10%	41%	17%	1%
sécurité	1-3	3-4	2-3	0,289	0,729	0,331	0,245	0,578	0,180	0,093	0,620	0,533	0,135	9%	44%	10%	19%
Inertie projetée				Valeurs propres cumulées				62%	54%	48%	47%	35%	29%	27%	20%	19%	13%

4.2.3 Contributions et qualités de projection des variables

Le Tableau 8 **Error! Reference source not found.** permet de trouver les plans factoriels permettant de mieux interpréter les villes selon un critère donné. Par exemple, pour comparer la taille des villes, il est intéressant de projeter sur le plan 1-2 car le Cos^2 de la variable « Taille » sur ce plan est proche de 1. Pour comparer les villes en tenant en compte du climat (critère « météo »), il est intéressant de projeter sur le plan 2-4. Pour mieux comprendre les différences de dynamisme économique entre les villes, on projettera sur le plan 1-5.

Les plans de projections les plus intéressants à voir sont donc 1-2, 1-3, 1-5 et 2-4.

Pour interpréter les axes factoriels, nous avons identifié d'abord les variables bien projetées ($\text{Cos}^2 > 0,6$) et celles qui contribuent plus que la contribution théorique moyenne à la formation de chaque axe :

$$\text{Contribution théorique moyenne} = \frac{100}{\text{nombre de variables}} = \frac{100}{10} = 10\%$$

Les variables qui contribuent le plus à la constitution des axes sont résumées dans le Tableau 9.

Tableau 9 : Variables qui contribuent le plus à la constitution des axes

Axe 1			
Variable	Coord 1	CTR 1	$\text{Cos}^2 1$
culture	-0,897	20%	0,80
taille	-0,888	19%	0,79
emploi	-0,804	16%	0,65
immobilier	0,766	14%	0,59

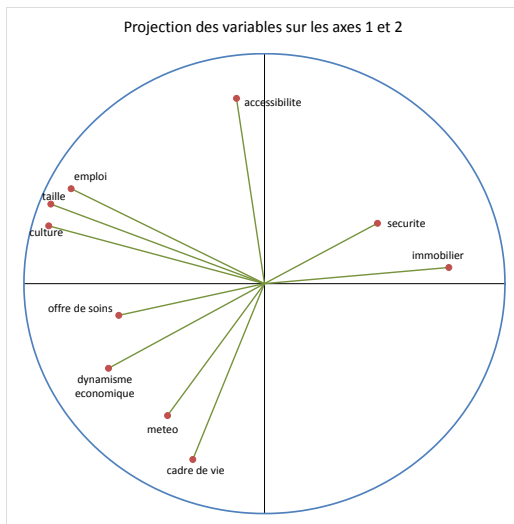
Axe 2			
Variable	Coord 2	CTR 2	$\text{Cos}^2 2$
accessibilite	0,806	30%	0,65
cadre de vie	-0,765	27%	0,59
meteo	-0,574	15%	0,33

Axe 3			
Variable	Coord 3	CTR 3	$\text{Cos}^2 3$
securite	-0,713	39%	0,51
offre de soins	-0,647	32%	0,42

Axe 4			
Variable	Coord 4	CTR 4	$\text{Cos}^2 4$
meteo	-0,600	50%	0,36
securite	0,333	15%	0,11
cadre de vie	0,285	11%	0,08

Axe 5			
Variable	Coord 5	CTR 5	$\text{Cos}^2 5$
dynamisme economique	0,497	43%	0,25
immobilier	0,435	33%	0,19

4.2.4 Cercles de corrélation et interprétation des axes



Graphique 8 : Cercle de corrélation sur le plan 1-2

Tableau 10 : Interprétation des cadrans du plan 1-2

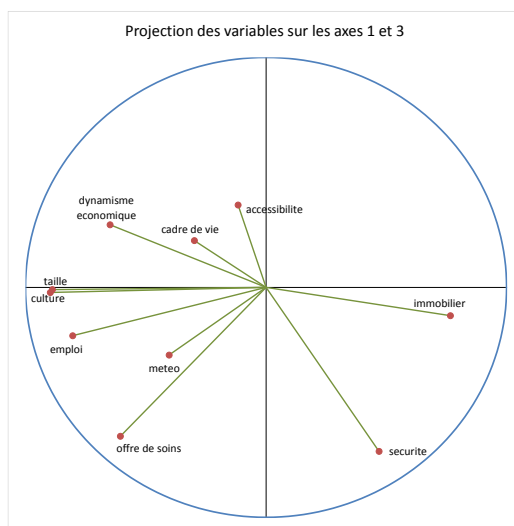
Axe 2		Axe 1
Grandes villes culturelles et dynamiques accessibles avec un cadre de vie moins apprécié	Petites villes accessibles avec immobilier pas cher mais pas dynamiques économiquement	
Grandes villes culturelles et dynamiques moins accessibles avec un cadre de vie apprécié	Petites villes difficilement accessibles où l'immobilier n'est pas cher	

Dans les cercles de corrélation, on ne prend en compte que les variables bien présentées, c'est-à-dire proche du cercle.

On constate que le prix bas de l'immobilier est corrélé positivement avec l'axe 1, les offres d'emploi, la taille de la ville, l'offre culturelle et le dynamisme économique sont corrélées négativement avec cet axe. L'axe 1 oppose donc les petites villes avec immobilier pas cher (à droite) et les grandes villes culturelles avec plus de dynamisme économiques et d'opportunités d'emploi (à gauche).

L'axe 2 est corrélé positivement avec la variable « accessibilité » et corrélé négativement avec le cadre de vie. Cet axe oppose donc les villes accessibles avec moins d'espaces verts (en haut) et les villes moins accessibles avec plus d'espaces verts et un cadre de vie plus agréable (en bas).

On en déduit donc les interprétations des cadrans du plan 1-2 comme détaillé dans le Tableau 10.

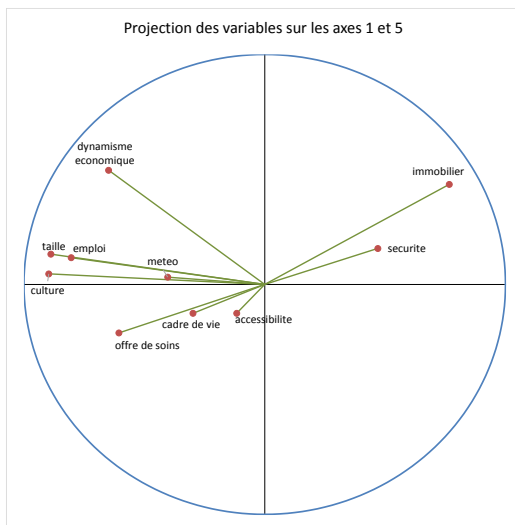


Graphique 9 : Cercle de corrélation sur le plan 1-3

Tableau 11 : Interprétation des cadrans du plan 1-3

Axe 3		Axe 1
Grandes villes mais moins de sécurité et immobiliers chers	Faible qualité de l'offre de soin	
Bonne offre de soin	Petites villes bien sécurisées avec immobilier pas cher	

L'interprétation de l'axe 1 est déjà réalisée. On constate d'après le cercle de corrélation que l'axe 3 oppose les villes avec une offre de soin performante et avec un taux de délinquance faible (en bas) aux villes avec moins de sécurité et une offre de soin plus faible que la moyenne (en haut).

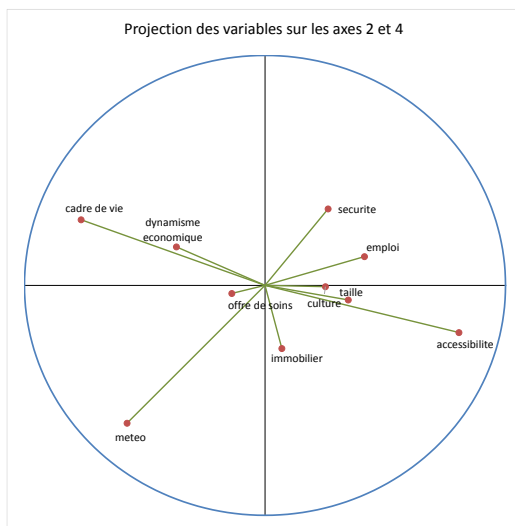


Graphique 10 : Cercle de corrélation sur le plan 1-5

Comme vu au paragraphe 4.2.2, on projette sur le plan 1-5 pour mieux visualiser les villes d'un point de vue dynamisme économiques. En effet, la variable « Dynamisme économique » est corrélée positivement avec l'axe 5 et négativement avec l'axe 1. Les villes les plus dynamiques sont donc proches de la deuxième bissectrice et situées dans le quadrant gauche-haut du plan 1-5.

Tableau 12 : Interprétation des cadrans du plan 1-5

Axe 5		Axe 1
Grandes villes culturelles et dynamiques économiquement	Petites villes avec immobiliers pas chers	
Grandes villes culturelles avec immobilier cher et dynamique économique moyen	Petites villes non dynamiques économiquement	



Graphique 11 : Cercle de corrélation sur le plan 2-4

Comme vu au paragraphe 4.2.2, le plan 2-4 est le meilleur plan qui permet de visualiser les villes en mettant en évidence l'accessibilité, le cadre de vie et les conditions climatiques. Les interprétations des cadrants de ce plan sont résumées dans la Tableau 13.

Tableau 13 : Interprétation des cadrans du plan 2-4

Axe 4		Axe 2
Plus d'espaces verts mais moins accessibles dans un climat océanique ou méditerranéen		
Bon climat méditerranéen et montagnard du sud-est avec des espaces verts appréciés	Villes accessibles mais où le climat continental est moins apprécié et avec des espaces verts limités	

4.3 Analyse des individus

Le Tableau 14 montre les villes qui contribuent le plus à la constitution des plans de projection. Ici on a 50 individus (villes) donc la contribution théorique moyenne est de $\frac{100}{50} = 2\%$. Une ville contribue significativement à la constitution du plan si son CTR sur ce plan est plus grand que 2%. Le CTR d'un individu dans un plan factoriel est la somme des CTR de cet individu sur les 2 axes factoriels formant ce plan.

Tableau 14 : Villes avec meilleures contribution à la constitution des axes

Plan 1-2		Plan 1-3		Plan 1-5		Plan 2-4	
Ville	Total CTR 1-2	Ville	Total CTR 1-3	Ville	Total CTR 1-5	Ville	Total CTR 2-4
marseille	12,00	valenciennes	12,06	montbéliard	11,99	bayonne	12,53
nice	11,95	dunkerque	11,24	marseille	9,77	nîmes	12,26
perpignan	9,20	marseille	9,98	toulouse	9,60	perpignan	11,70
montpellier	7,73	béthune	9,82	saint etienne	9,35	brest	11,61
nîmes	7,57	Lille	9,80	reims	8,55	valence	9,10
lyon	7,53	douai-lens	8,78	rochelle	8,05	lorient	7,18
toulon	7,53	lyon	7,25	thionville	7,98	reims	6,49
montbéliard	7,22	toulouse	7,12	nice	7,68	toulon	6,21
toulouse	7,17	clermont ferrand	6,90	nantes	6,76	rennes	6,18
thionville	6,56	havre	6,83	annecy	6,53	mulhouse	5,93
Lille	5,86	montbéliard	6,78	lyon	5,88	metz	5,83
metz	5,79	nice	6,70	limoges	5,69	troyes	5,68
rennes	5,53	angoulême	6,65	montpellier	5,67	nice	5,65
saint nazaire	5,23	thionville	6,04	Lille	5,19	nancy	5,40
nancy	4,98	bordeaux	5,04	angoulême	5,13	orléans	5,03
lorient	4,94	montpellier	4,83	bayonne	5,03	avignon	4,96
angoulême	4,79	rennes	4,62	rennes	5,03	saint nazaire	4,85
bordeaux	4,79	rouen	4,34	bordeaux	4,97	rouen	4,84
mans	4,76	lorient	4,10	annemasse	4,89	pau	4,73
mulhouse	4,69	grenoble	4,08	valenciennes	4,68	rochelle	4,73
troyes	4,28	troyes	4,03	béthune	4,43	Lille	4,32
béthune	4,10	besançon	3,93	troyes	4,23	caen	4,15
strasbourg	4,00	nantes	3,85	amiens	4,06	mans	4,15
bayonne	3,98	mulhouse	3,66	douai-lens	3,82	lyon	4,13
orléans	3,84	saint nazaire	3,24	lorient	3,82	marseille	4,04
rochelle	3,73	limoges	3,16	brest	3,71	montpellier	4,00
nantes	3,72	mans	2,93	mulhouse	3,45	havre	3,93
rouen	3,59	poitiers	2,84	dunkerque	3,45	nantes	3,64
grenoble	3,07	angers	2,81	grenoble	3,07	saint etienne	3,38
reims	2,82	brest	2,80	toulon	2,82	strasbourg	2,70
valence	2,72	bayonne	2,78	saint nazaire	2,73	tours	2,59
dunkerque	2,65	dijon	2,66	valence	2,66	dijon	2,47
annemasse	2,62	annemasse	2,41	metz	2,50	annecy	2,12
dijon	2,58	annecy	2,31	mans	2,16	angoulême	2,09
avignon	2,58	strasbourg	2,22	caen	2,09		
tours	2,49	valence	2,11				

Le Tableau 15 représente les villes mal projetées sur les plans étudiées. On considère qu'un individu est bien représenté dans un plan si son Cos^2 est supérieure à 0,6 et que c'est moins bien représenté si le Cos^2 est entre 0,4 et 0,6 et que c'est très mal représenté si son Cos^2 est inférieure à 0,4. Comme le CTR, le Cos^2 d'un individu sur un plan est la somme des Cos^2 de cet individu sur les axes factoriels formant ce plan. Par exemple, sur le plan 1-2, le Cos^2 de Valenciennes et Besançon sont de l'ordre de 0,06 donc inférieure à 0,4 et très proche de 0. Ces deux villes sont donc très mal représentées sur ce plan. On n'interprètera que les villes bien représentées sur le plan.

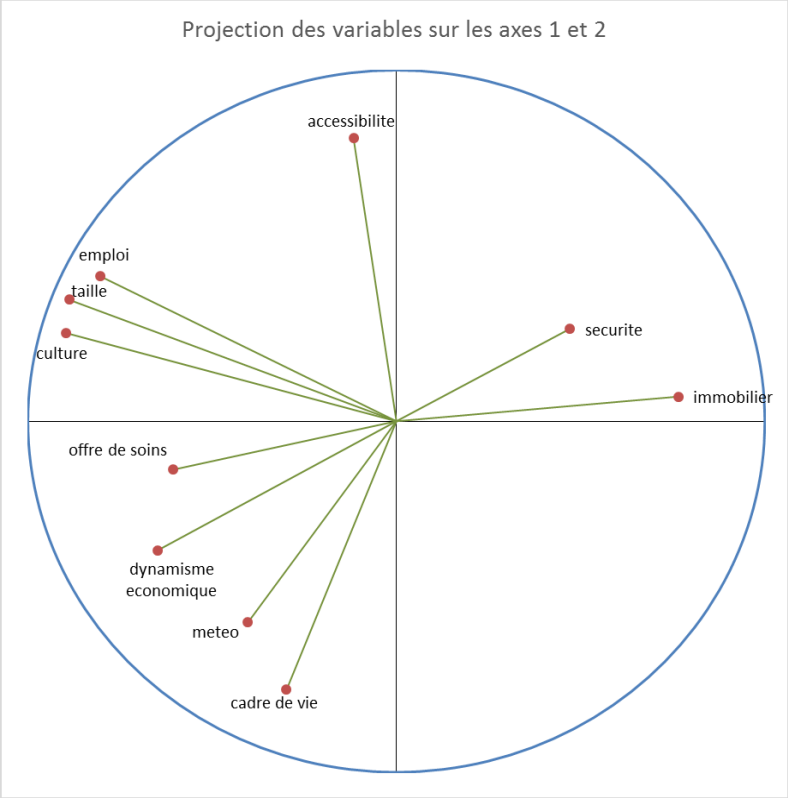
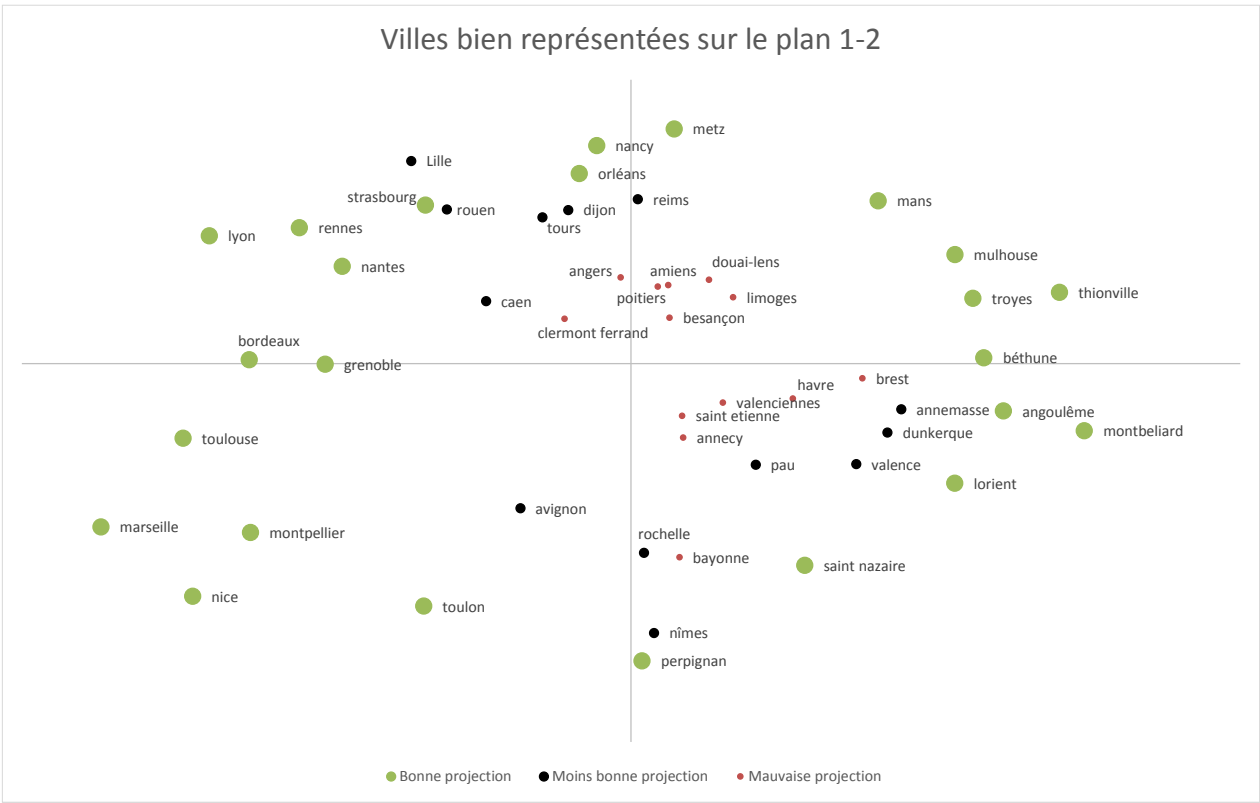
Tableau 15 : Villes mal projetées

Plan 1-2		Plan 1-3		Plan 1-5		Plan 2-4	
Ville	Total Cos^2 1-2	Ville	Total Cos^2 1-3	Ville	Total Cos^2 1-5	Ville	Total Cos^2 2-4
valenciennes	0,06	nîmes	0,00	perpignan	0,00	béthune	0,00
besançon	0,06	perpignan	0,01	nîmes	0,00	grenoble	0,01
clermont ferrand	0,08	rochelle	0,01	angers	0,02	bordeaux	0,01
saint etienne	0,10	reims	0,02	besançon	0,03	montbeliard	0,03
annecy	0,11	metz	0,03	orléans	0,04	annemasse	0,03
poitiers	0,14	nancy	0,03	clermont ferrand	0,05	valenciennes	0,03
douai-lens	0,15	saint etienne	0,05	dijon	0,05	thionville	0,04
angers	0,20	orléans	0,07	nancy	0,08	toulouse	0,04
amiens	0,20	tours	0,17	poitiers	0,10	dunkerque	0,05
havre	0,23	bayonne	0,17	tours	0,10	besançon	0,05
limoges	0,24	amiens	0,18	metz	0,11	angoulême	0,08
brest	0,31	annecy	0,22	bayonne	0,14	clermont ferrand	0,08
bayonne	0,39	avignon	0,24	valenciennes	0,14	douai-lens	0,10
		toulon	0,27	douai-lens	0,18	limoges	0,10
		pau	0,34	avignon	0,20	marseille	0,14
		saint nazaire	0,36	pau	0,23	annecy	0,15
		brest	0,37	havre	0,24	lyon	0,16
		dijon	0,37	saint nazaire	0,26	havre	0,18
		poitiers	0,38	annecy	0,27	poitiers	0,18
		caen	0,38	rochelle	0,27	nantes	0,20
				rouen	0,28	amiens	0,20
				Lille	0,29	troyes	0,22
				toulon	0,29	mulhouse	0,24
				amiens	0,32	montpellier	0,25
				reims	0,34	rennes	0,28
				brest	0,36	nice	0,29
				caen	0,40	saint etienne	0,30
						lorient	0,30
						angers	0,30
						Lille	0,31
						mans	0,35
						brest	0,36
						caen	0,39
						rouen	0,39

le **cnam**

4.4 Projection sur les plans factoriels

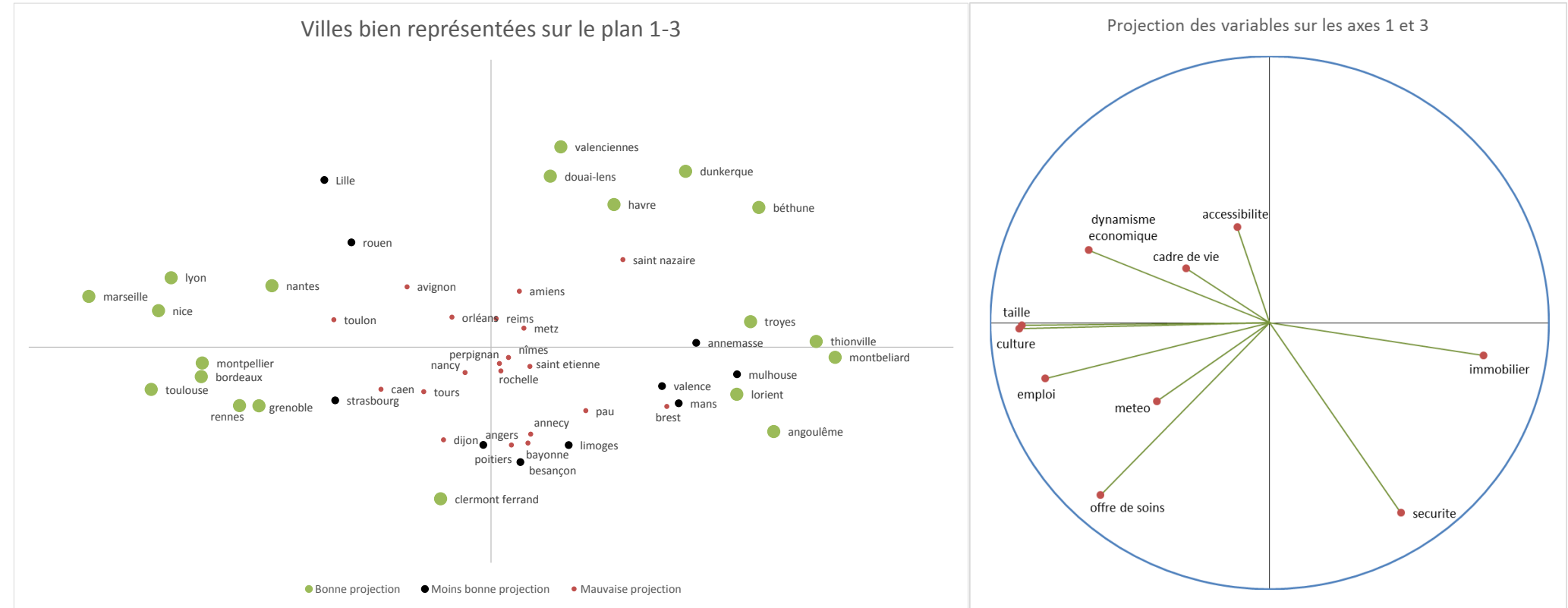
4.4.1 Projection sur le plan 1-2



Les villes bien représentées sur le plan 1-2 sont en verts. On constate que l'infrastructure de transport est bien appréciée à Metz, Nancy et Orléans. Le prix de l'immobilier est moins cher à Thionville, Troyes, Béthune, Mulhouse, Angoulême et Montbéliard. Le cadre de vie est agréable à Nice, Toulon, Montpellier, et Perpignan. Les plus grandes agglomérations où l'offre culturelle est appréciée et où il y a de bonnes opportunités d'emploi sont Lyon, Rennes, Bordeaux, Grenoble, Nantes et Strasbourg. Saint Nazaire est une petite ville (par rapport aux autres) où le prix de l'immobilier n'est pas très cher et où le cadre de vie est agréable mais il n'y a pas beaucoup d'opportunités d'emploi.

Axe 2		Axe 1
Grandes villes culturelles et dynamiques accessibles avec un cadre de vie moins apprécié	Petites villes accessibles avec immobilier pas cher mais pas dynamiques économiquement	
Grandes villes culturelles et dynamiques moins accessibles avec un cadre de vie apprécié	Petites villes difficilement accessibles où l'immobilier n'est pas cher	

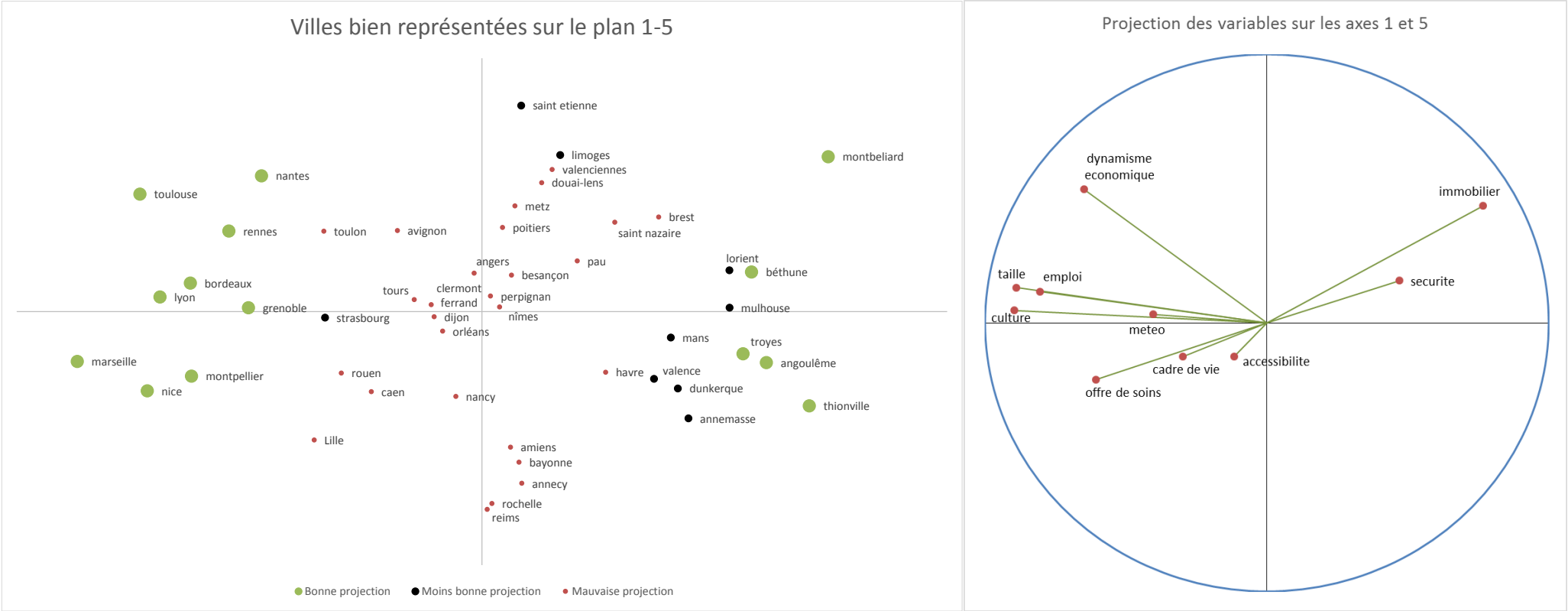
4.4.2 Projection sur le plan 1-3



Il semble que l'offre de soin est performante à Montpellier, Bordeaux, Toulouse, Rennes, Grenoble et Clermont-Ferrand. Dans cette dernière ville, il y a un faible taux de délinquance. Les villes à gauche du graphique et proches de l'axe horizontal sont des grandes villes culturelles où il y a de bonnes opportunités d'emploi. Il s'agit par exemple de Lyon, Marseille, Nice, Montpellier, Bordeaux et Toulouse.

Axe 3		Axe 1
Grandes villes mais moins de sécurité et immobiliers chers	Faible qualité de l'offre de soin	
Bonne offre de soin	Petites villes bien sécurisées avec immobilier pas cher	

4.4.3 Projection sur le plan 1-5



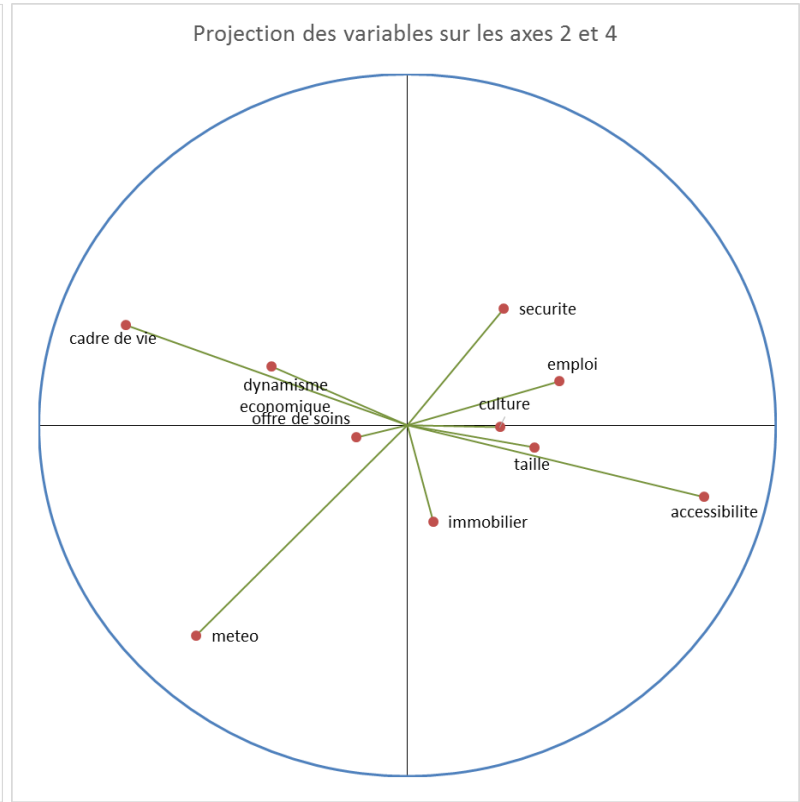
Les provinces les plus dynamiques économiquement sont Toulouse, Nantes, Rennes, Bordeaux, Lyon et Grenoble. Le prix de l'immobilier est élevé dans ces villes. Par contre, Troyes, Angoulême et Thionville ne sont pas dynamique mais le prix de l'immobilier est plus faible.

Axe 5		Axe 1
Grandes villes culturelles et dynamiques économiquement	Petites villes avec immobiliers pas chers	
Grandes villes culturelles avec immobilier cher et dynamique économique moyen	Petites villes non dynamiques économiquement	

Villes bien représentées sur le plan 2-4

The scatter plot displays the following cities and their approximate coordinates (x, y) on a scale from 0 to 100:

City	Color	X (approx)	Y (approx)
brest	Red	55	85
bayonne	Green	30	75
lorient	Red	40	70
pau	Black	45	65
havre	Red	55	65
caen	Red	65	65
rennes	Red	75	65
nantes	Red	70	60
rouen	Red	80	60
angers	Red	70	55
amiens	Red	70	50
limoges	Red	65	50
thionville	Red	65	48
besançon	Red	60	48
valenciennes	Red	55	48
bordeaux	Red	55	45
annemasse	Red	50	45
dunkerque	Red	45	45
nice	Red	25	55
rochelle	Black	30	45
saint nazaire	Black	30	48
strasbourg	Black	75	50
Lille	Red	85	50
metz	Green	90	50
dijon	Black	75	55
douai-lens	Red	70	58
béthune	Red	55	58
grenoble	Red	55	62
montbéliard	Red	45	62
toulouse	Red	45	65
montpellier	Red	35	70
marseille	Red	35	72
perpignan	Green	15	75
nîmes	Green	20	85
avignon	Black	40	78
angoulême	Red	50	75
saint etienne	Red	50	80
clermont ferrand	Red	60	70
poitiers	Red	65	68
tours	Black	75	65
nancy	Green	85	65
orléans	Green	80	70
lyon	Red	75	75
reims	Black	80	80
mulhouse	Red	70	85
troyes	Red	65	85
valence	Black	45	95



Metz, Nancy et Orléans sont bien accessibles comme villes.

Axe 4		Axe 2
Plus d'espaces vers mais moins accessibles dans un climat océanique ou méditerranéen		
Bon climat méditerranéen et montagnard du sud-est avec des espaces verts appréciés	Villes accessibles mais où le climat continental est moins apprécié et avec des espaces verts limités	

5 CLASSIFICATION DES VILLES

5.1 Dendrogrammes

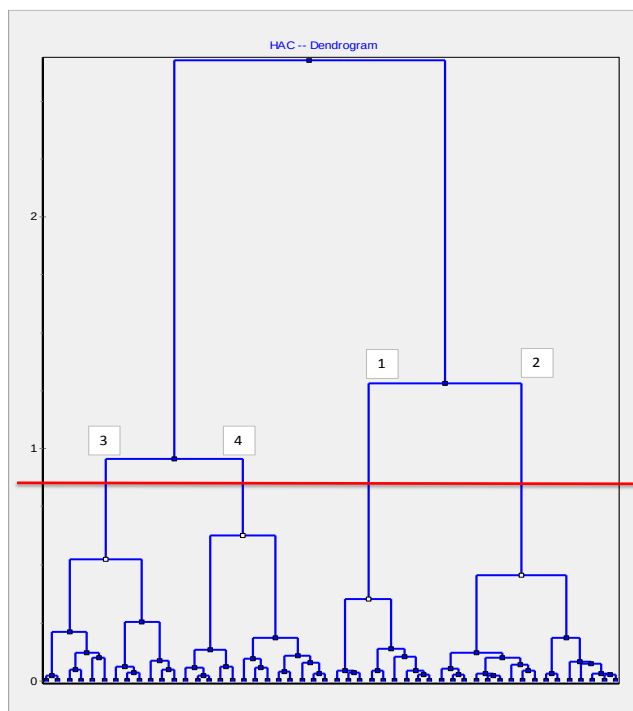


Figure 3 : Dendrogramme pour la 1ère classification à 4 clusters

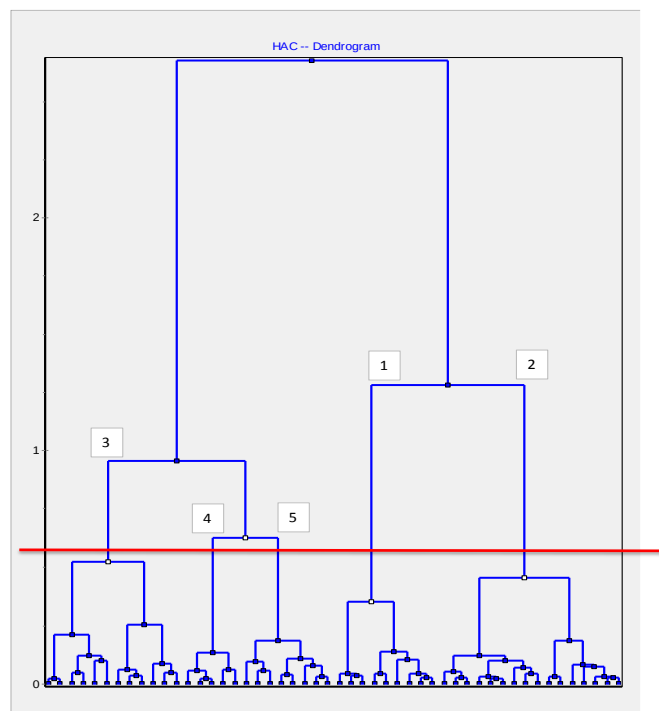


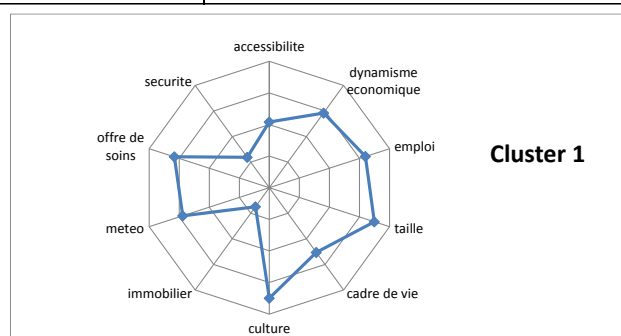
Figure 4 : Dendrogramme pour la 2ème classification à 5 clusters

Tanagra propose 4 clusters pour ces données. Nous avons fait aussi une classification à 5 et 8 clusters, détaillées dans le document Excel ci-joint.

5.2 Classification en 4 clusters

Cluster 1 :			
Caractéristique	Test value	Interprétation	Description du cluster
culture	4,38	Villes culturelles	Grandes villes culturelles et dynamiques où il fait beau et il y a une bonne offre de soin mais l'immobilier est cher et le niveau de sécurité est faible.
taille	4,32	Grandes villes	
emploi	4,2	Il y a de l'emploi	
dynamisme économique	3,91	un bon dynamisme économique	
offre de soins	2,89	Bonne offre de soin	
meteo	2,54	Bon climat	
securite	-2,72	Manque de sécurité	
immobilier	-3,81	L'immobilier est cher	

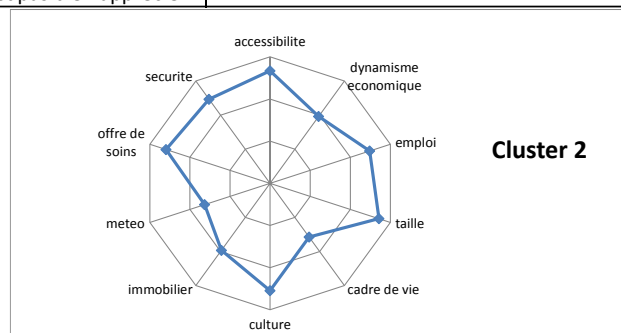
Tableau 16 : Caractéristiques du cluster 1



Graphique 12 : Profil du Cluster 1

Cluster 2 :			
Caractéristique	Test value	Interprétation	Description du cluster
accessibilite	3,66	Villes accessibles	Grandes villes culturelles et accessibles où il y a beaucoup d'offre d'emploi mais le climat et les espaces verts ne sont pas bien appréciés.
taille	2,62	Grandes villes	
emploi	2,29	Il y a des offres d'emploi	
culture	1,94	Villes culturelles	
meteo	-2	Climat moins apprécié	
cadre de vie	-3,04	Le cadre de vie n'est pas bien apprécié	

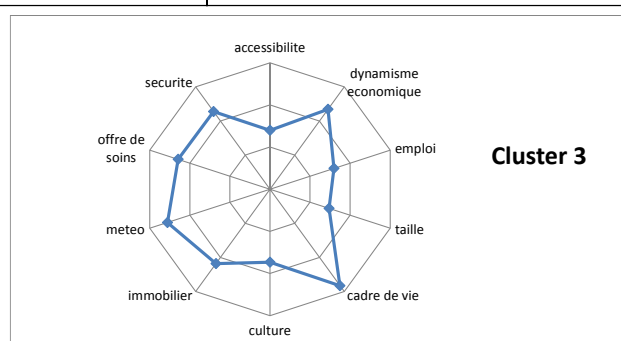
Tableau 17 : Caractéristiques du cluster 2



Graphique 13 : Profil du Cluster 2

Cluster 3 :			
Caractéristique	Test value	Interprétation	Description du cluster
cadre de vie	3,54	Villes avec beaucoup d'espaces verts	Petites villes avec beaucoup d'espaces verts mais avec une faible infrastructures adaptée aux personnes handicapées (moins accessibles)
emploi	-2,02	Villes avec moins d'offres d'emploi	
taille	-2,66	Petites villes	
accessibilite	-4,31	Villes difficilement accessibles	

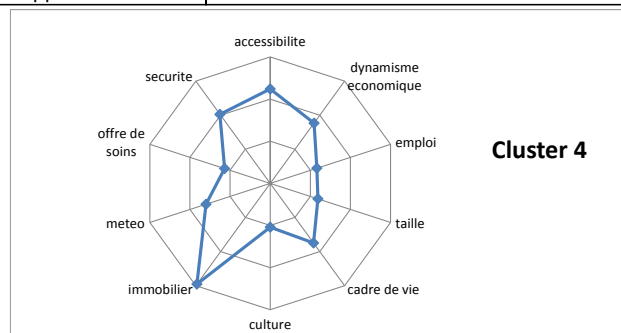
Tableau 18 : Caractéristiques du cluster 3



Graphique 14 : Profil du Cluster 3

Cluster 4 :			
Caractéristique	Test value	Interprétation	Description du cluster
immobilier	3,32	Immobilier pas cher	Villes plus petites, moins dynamiques où l'immobilier n'est pas cher.
dynamisme économique	-2,69	Villes moins dynamiques	
taille	-3,98	Petites villes	
emploi	-4,15	Pas beaucoup d'offres d'emploi	
culture	-4,42	Villes avec faibles activités culturelles	
offre de soins	-4,43	Offre de soins moins appréciée	

Tableau 19 : Caractéristiques du cluster 4



Graphique 15 : Profil du Cluster 4

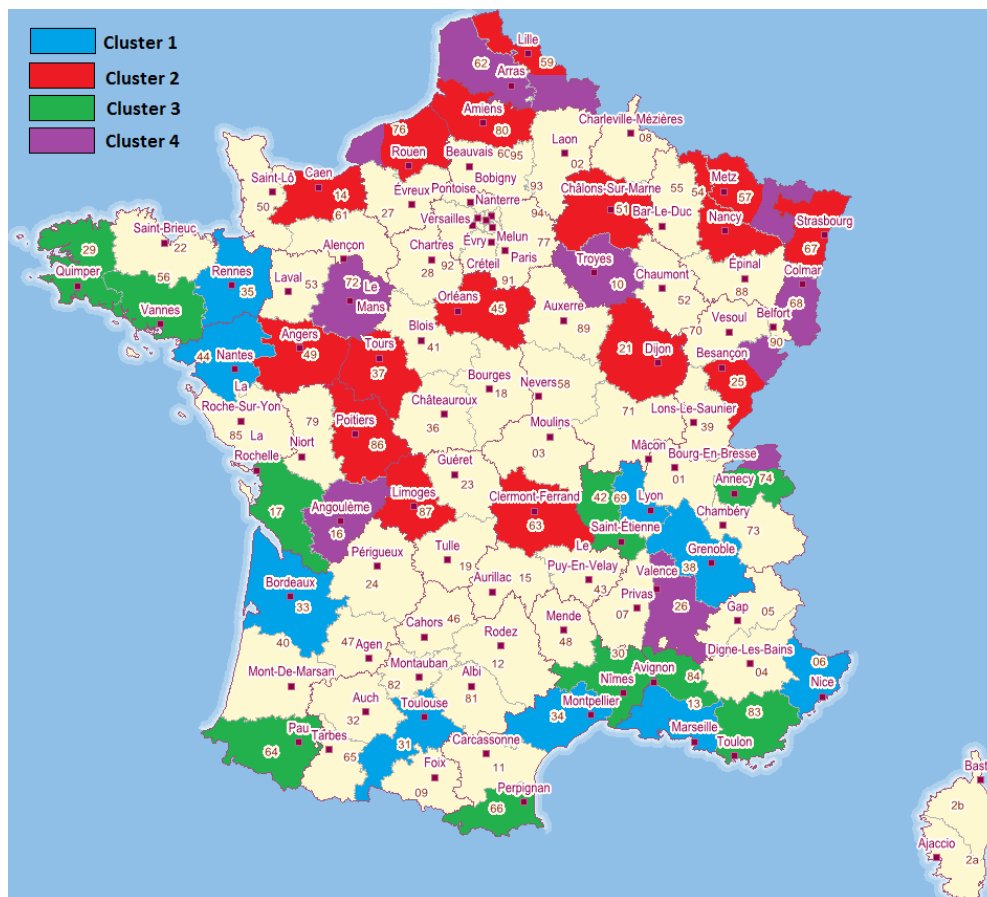


Figure 5 : Visualisation géographique des clusters

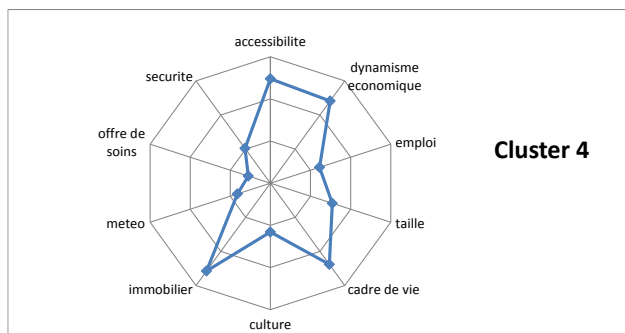
Les villes des clusters 1 et 3 sont dans le sud ou bien dans l'ouest et les villes des clusters 2 et 4 sont plutôt vers le nord et l'est.

5.3 Classification en 5 clusters

Comme le montre les dendrogrammes, la différence entre ces deux classifications est au niveau des clusters 4 et 5.

Cluster 4 :			
Caractéristique	Test value	Interprétation	Description du cluster
emploi	-2,18		Villes avec faible niveau de sécurité, de cultures et d'offre de soin. Le climat dans ces villes n'est pas bien apprécié.
culture	-2,22		
securite	-2,29	Faibles sécurités par rapport au top 50 plus grande	
meteo	-2,78	Climat apprécié	
offre de soins	-3,8	Faible offre de soin	

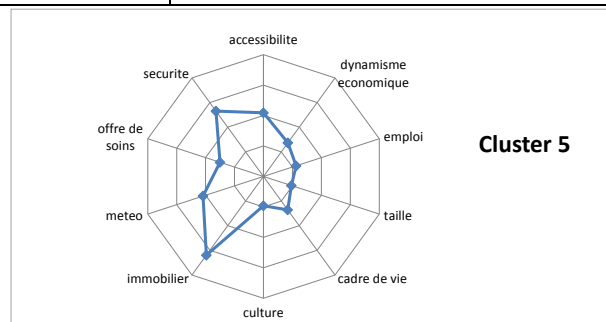
Tableau 20 : Caractéristiques du cluster 4 de la 2ème classification



Graphique 16 : Profil du Cluster 4 de la 2ème classification

Cluster 5 :			
Caractéristique	Test value	Interprétation	Description du cluster
immobilier	3,09	Immobilier pas cher	Petites villes calmes non dynamiques avec des immobilier pas chers mais des espaces verts limités et des offres de soin moins appréciées.
offre de soins	-2,19	Offre de soin moins appréciée	
cadre de vie	-2,77	Espaces verts limités	
emploi	-3,18	Pas assez d'offres d'emplois	
culture	-3,47	Pas d'activités culturelles	
taille	-3,6	Petites villes	
dynamisme économique	-3,94	Pas de dynamismes économiques	

Tableau 21 : Caractéristiques du cluster 5 de la 2ème classification



Graphique 17 : Profil du Cluster 5 de la 2ème classification

Ces deux clusters étaient fusionnés dans la première classification du paragraphe 5.2. On constate qu'il y a une séparation en deux groupes en se basant sur la sécurité, le dynamisme économique, le cadre de vie et l'immobilier.

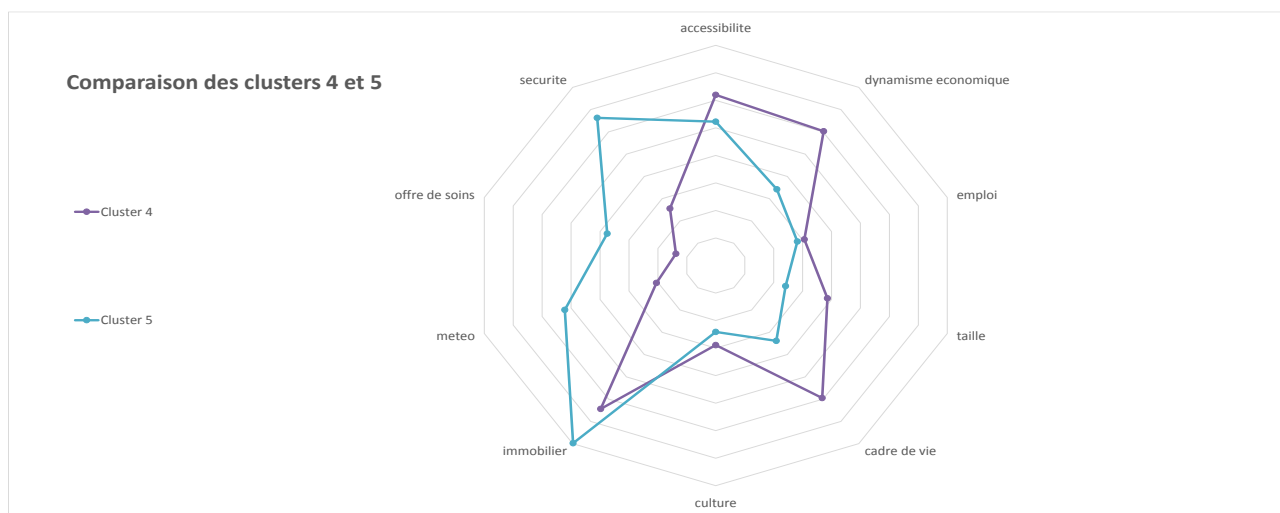


Tableau 22 : Calcul du profil du cluster 1

Graphique 18 : Comparaison des clusters 4 et 5

ville	accessibilité	dynamisme économique	emploi	taille	cadre de vie	culture	immobilier	météo	offre de soins	sécurité
lyon	14,40	12,90	18,50	20,00	8,30	19,80	3,20	14,70	10,10	2,60
nantes	13,90	16,50	17,90	15,00	11,20	16,60	5,90	7,60	9,80	9,80
rennes	13,20	14,90	17,90	18,20	10,10	17,10	5,40	6,80	17,80	14,90
marseille	10,70	15,10	13,80	19,80	18,90	19,70	2,00	20,00	17,10	0,40
grenoble	9,90	12,50	17,50	14,80	10,30	15,90	4,80	15,90	15,80	9,20
bordeaux	9,50	14,50	15,90	19,20	11,00	16,60	3,90	11,30	18,10	7,40
montpellier	8,70	14,40	10,70	16,00	14,80	16,50	2,80	18,70	20,00	3,70
nice	6,80	14,20	15,50	14,80	19,50	16,70	0,40	18,70	15,20	0,80
toulouse	6,50	16,40	16,20	19,40	9,80	18,60	5,30	16,00	18,40	4,70
Cluster 1	10,40	14,60	15,99	17,47	12,66	17,50	3,74	14,41	15,81	5,94

Comme montre le Tableau 22, le profil du cluster 1, par exemple, est la moyenne des profils des individus formant ce cluster.

5.4 Classification en 8 clusters

Cf. détails dans le document Excel.

6 ACP ET CLASSIFICATION

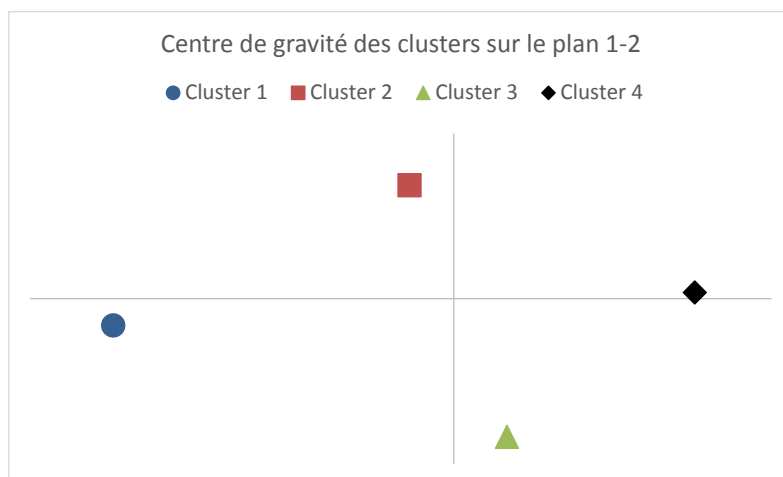
6.1 ACP et classification en 4 clusters

Ici on s'intéressera uniquement au plan 1-2. Le principe est le même sur les autres plans.

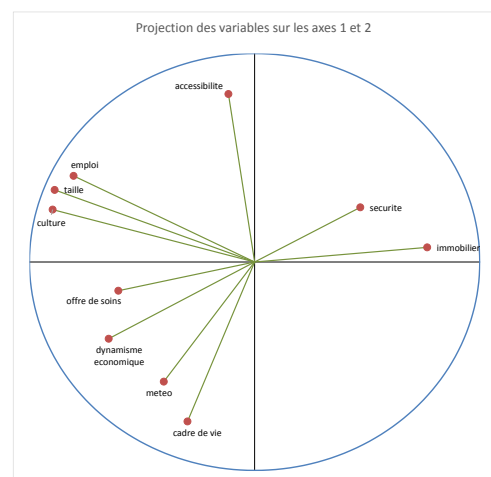
Tableau 23 : Calcul du centre de gravité du cluster 1 sur le plan 1-2

		Axe horizontal	Axe vertical	Centres de gravité	
		1	2	Coord cluster axe x	Coord cluster axe y
Cluster	ville	Coord 1	Coord 2		
1	lyon	-3,459746838	1,34876072		
1	nantes	-2,36976099	1,02718973		
1	rennes	-2,721017122	1,4347167		
1	marseille	-4,34914732	-1,72622776		
1	grenoble	-2,50943017	-0,00697178		
1	bordeaux	-3,133709192	0,04013611		
1	montpellier	-3,122710705	-1,78678429		
1	nice	-3,596777201	-2,45986795		
1	toulouse	-3,675391436	-0,79006886	-3,215299	-0,32434638

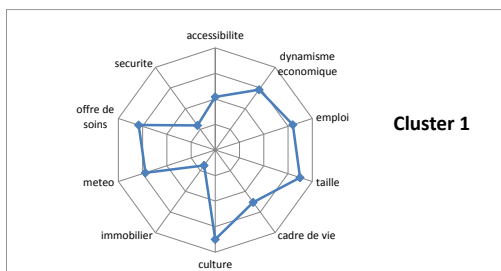
Comme montre le Tableau 23, les coordonnées du centre du cluster 1 sur le plan 1-2 sont les moyennes des coordonnées des individus formant ce cluster sur ce même plan.



Graphique 20 : Projection des centres des clusters sur la plan 1-2



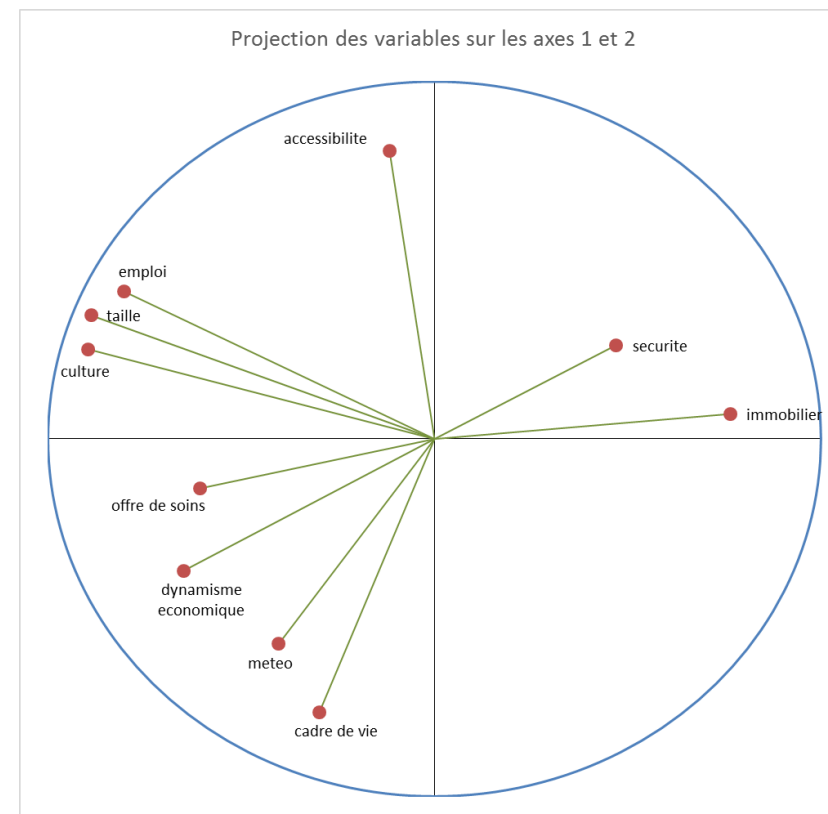
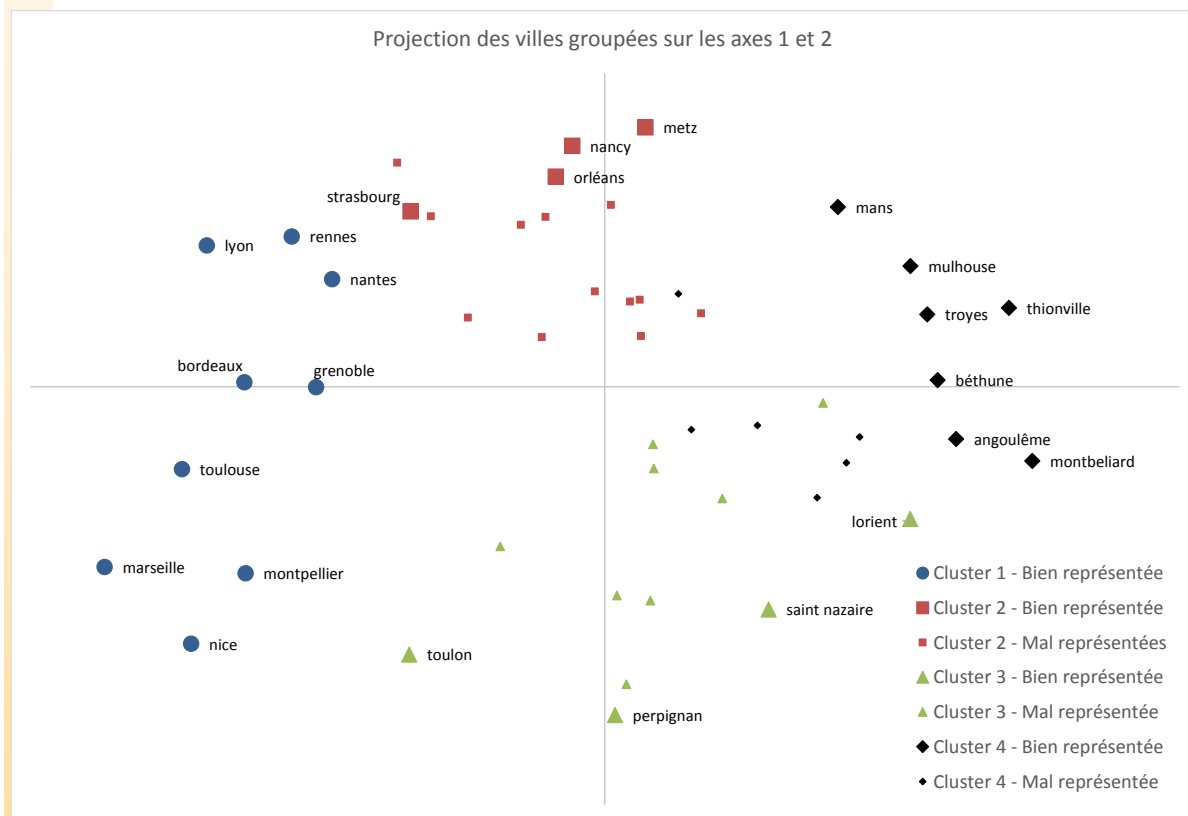
Graphique 19 : Cercle de corrélation sur le plan 1-2



Graphique 21 : Profil du cluster 1

Comme vu dans le paragraphe 5.2, le profil du cluster 1 est caractérisé par des grandes villes culturelles et dynamiques où le climat est apprécié et il y a une bonne offre de soin mais l'immobilier est cher et le niveau de sécurité est faible. Ces caractéristiques sont bien cohérentes avec la position du centre du cluster dans le plan 1-2.

le **cnam**



Graphique 22 : Projection des villes groupées par cluster sur le plan 1-2

Graphique 23 : Cercle de corrélation sur le plan 1-2

Les villes bien projetée sur le plan 1-2 (Cos^2 supérieure à 0,6) sont mises en évidence et regroupées par cluster dans le Graphique 22. Cela est en cohérence avec l'analyse des clusters obtenue dans le paragraphe 5.2. En effet, on retrouve les caractéristiques du cluster 1 constitué par les grandes villes culturelles et dynamiques où le climat est apprécié et il y a une offre de soin performante mais l'immobilier est cher et le taux de délinquance est à priori plus élevé que la moyenne.

Voici dans le Tableau 24 les villes qui constituent chaque cluster obtenus avec la première classification.

Tableau 24 : Villes bien représentées dans le plan 1-2 par cluster

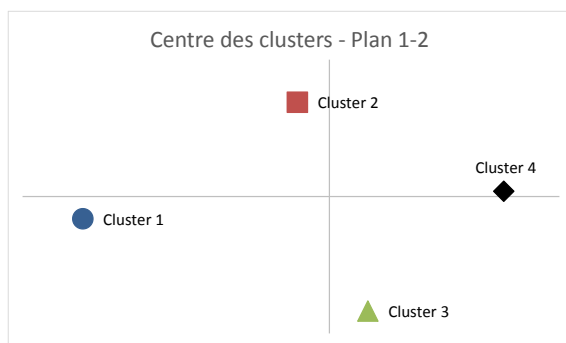
Cluster 1	Cluster 2	Cluster 3	Cluster 4
lyon	Lille	toulon	douai-lens
nantes	rouen	avignon	mans
rennes	strasbourg	saint etienne	béthune
marseille	tours	brest	dunkerque
grenoble	reims	perpignan	valenciennes
bordeaux	orléans	rochelle	mulhouse
montpellier	metz	nîmes	havre
nice	amiens	saint nazaire	annemasse
toulouse	nancy	bayonne	troyes
	angers	lorient	thionville
	dijon	annecy	angoulême
	caen	pau	valence
	clermont ferrand		montbeliard
	limoges		
	poitiers		
	besançon		

6.2 ACP et classification en 5 clusters

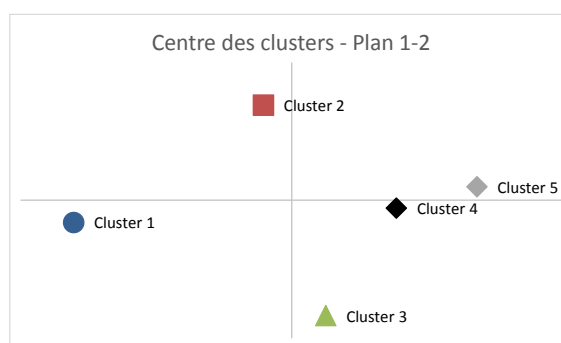
Cf. détails dans le document Excel.

6.3 Comparaison des classifications

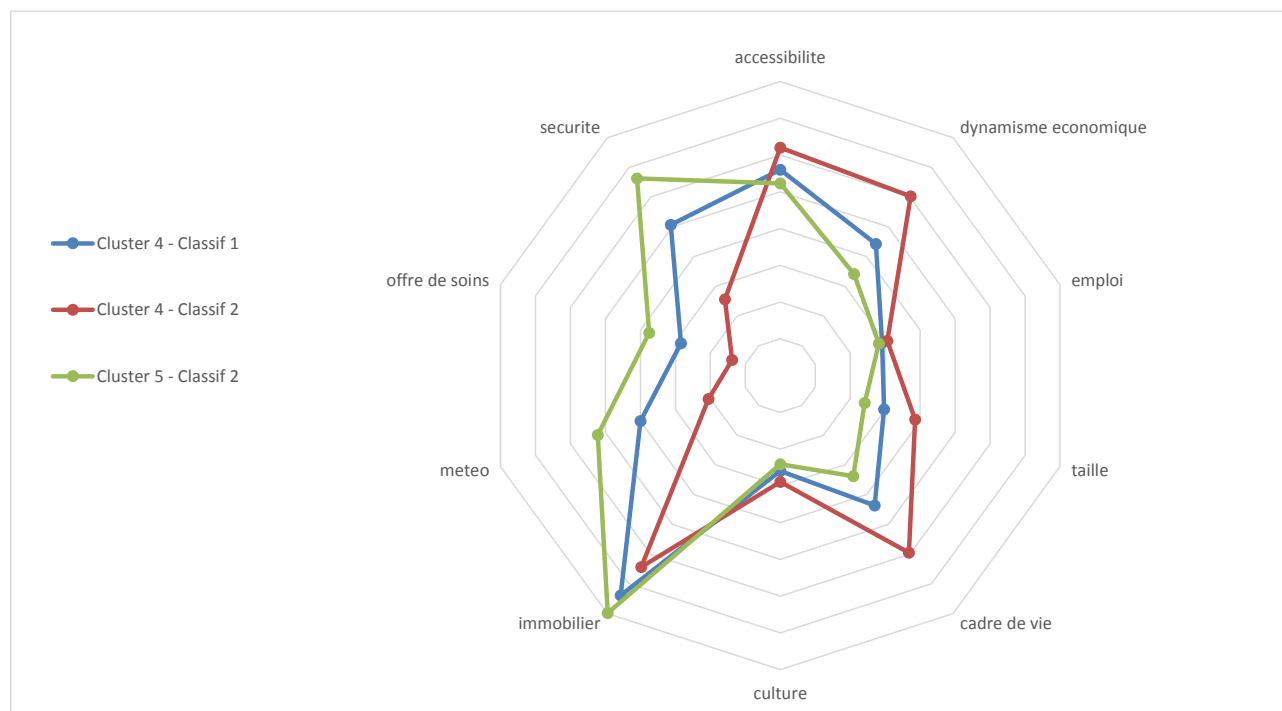
On compare ici le cluster 4 de la 1^{ère} classification (cf. dendrogramme de la Figure 3 et Graphique 25) avec les clusters 4 et 5 de la 2^{ème} classification (cf. dendrogramme de la figure Figure 4 et Graphique 24). Les autres comparaisons sont dans le fichier Excel.



Graphique 25 : Projection des centres des clusters de la 1^{ère} classification sur le plan 1-2



Graphique 24 : Projection des centres des clusters de la 2^{ème} classification sur le plan 1-2



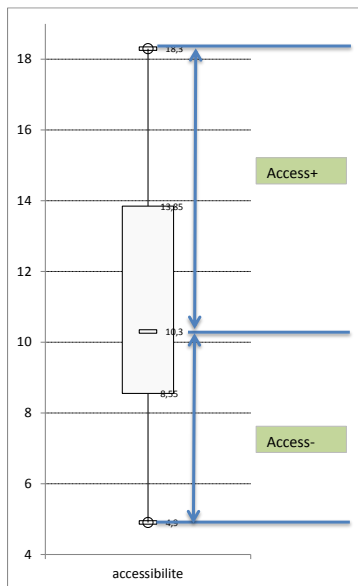
Graphique 26 : Comparaison du profil du cluster 4 de la 1ère classification avec les profils des clusters 4 et 5 de la 2ème classification

Le Graphique 26 montre d'une part, qu'il y a plus de dynamisme économique mais moins de sécurité dans le cluster 4 de la 2^{ème} classification (en rouge) et c'est exactement l'inverse pour le cluster 5 (en vert). D'autre part, dans le cluster 4 de la 1^{ère} classification (en bleu), le dynamisme économique et la sécurité sont entre les deux valeurs des autres clusters. Plus généralement, le profil du cluster consolidé (ici en bleu) de deux clusters est situé au barycentre des deux nouveaux clusters (en rouge et vert) sur toutes les dimensions (variables).

7 ANALYSE DES CORRESPONDANCES MULTIPLES

7.1 Bases de données transformées (Qualitatives)

7.1.1 Variables qualitatives à deux modalités

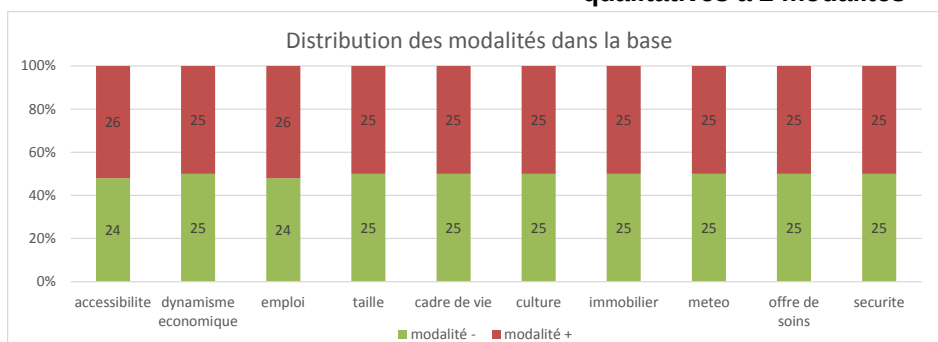


Graphique 27 : Transformation d'une variable continue en variable discrète à 2 modalités

Pour chaque variable, on considère que les valeurs supérieures à la médiane sont appréciées et notées (+) et les valeurs inférieures à la médiane sont défavorables et notées (-). Par exemple, pour la variable « Accessibilité », comme le montre le Graphique 27, si la valeur de la variable « Accessibilité » est inférieure à la médiane (10,3) alors la nouvelle variable discrète prend la modalité « Access- », et si cette valeur est plus grande que la médiane alors la nouvelle variable prend la modalité « Access+ ».

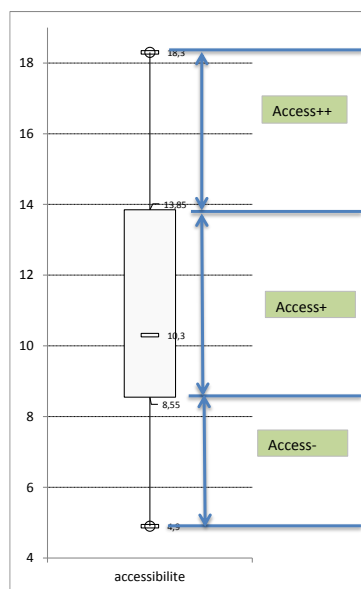
ville	accessibilité	dynamisme économique	emploi	taille	cadre de vie	culture	immobilier	meteo	offre de soins	securite
lyon	access+	dynamis-	emploi-	taille+	cadre+	culture-	immo-	meteo-	soins+	securite-
marseille	access+	dynamis+	emploi+	taille+	cadre-	culture+	immo-	meteo-	soins+	securite+
toulouse	access-	dynamis-	emploi-	taille-	cadre-	culture-	immo+	meteo+	soins-	securite+
bordeaux	access-	dynamis-	emploi+	taille-	cadre-	culture-	immo-	meteo+	soins+	securite+
rennes	access-	dynamis-	emploi-	taille-	cadre-	culture-	immo-	meteo+	soins-	securite+
rouen	access-	dynamis+	emploi-	taille-	cadre+	culture+	immo-	meteo+	soins-	securite-
montpellier	access-	dynamis-	emploi-	taille-	cadre+	culture-	immo-	meteo-	soins+	securite+
Lille	access-	dynamis+	emploi-	taille+	cadre-	culture-	immo+	meteo-	soins+	securite+
metz	access+	dynamis-	emploi-	taille-	cadre+	culture-	immo-	meteo-	soins-	securite-
strasbourg	access-	dynamis+	emploi+	taille+	cadre+	culture+	immo-	meteo+	soins+	securite-
clermont ferrand	access-	dynamis-	emploi+	taille-	cadre+	culture-	immo+	meteo-	soins-	securite+
nantes	access+	dynamis+	emploi+	taille+	cadre+	culture+	immo-	meteo-	soins+	securite+
nice	access-	dynamis-	emploi+	taille+	cadre-	culture-	immo+	meteo+	soins+	securite+
grenoble	access+	dynamis-	emploi+	taille+	cadre-	culture-	immo-	meteo+	soins+	securite+
orléans	access+	dynamis+	emploi-	taille+	cadre-	culture-	immo+	meteo-	soins-	securite-
dijon	access+	dynamis-	emploi-	taille-	cadre+	culture-	immo+	meteo-	soins-	securite-
caen	access-	dynamis+	emploi+	taille+	cadre+	culture+	immo-	meteo+	soins+	securite-
toulon	access+	dynamis-	emploi-	taille-	cadre+	culture-	immo+	meteo-	soins-	securite-
nancy	access+	dynamis-	emploi+	taille+	cadre+	culture+	immo-	meteo-	soins-	securite-
besançon	access-	dynamis-	emploi-	taille+	cadre-	culture+	immo+	meteo-	soins+	securite+
douai-lens	access-	dynamis-	emploi-	taille-	cadre+	culture-	immo+	meteo+	soins-	securite+
amiens	access+	dynamis+	emploi+	taille+	cadre-	culture+	immo-	meteo+	soins-	securite-
limoges	access+	dynamis-	emploi+	taille-	cadre-	culture-	immo+	meteo+	soins-	securite+
reims	access+	dynamis+	emploi+	taille+	cadre+	culture+	immo-	meteo+	soins+	securite-
angers	access+	dynamis-	emploi+	taille+	cadre-	culture+	immo+	meteo-	soins-	securite+
tours	access-	dynamis-	emploi-	taille-	cadre-	culture-	immo+	meteo-	soins-	securite+
poitiers	access-	dynamis+	emploi+	taille+	cadre+	culture+	immo-	meteo+	soins+	securite-
avignon	access+	dynamis-	emploi-	taille-	cadre-	culture-	immo+	meteo+	soins-	securite+
saint etienne	access+	dynamis-	emploi+	taille+	cadre-	culture+	immo+	meteo-	soins+	securite+
mans	access+	dynamis+	emploi+	taille+	cadre+	culture+	immo-	meteo-	soins-	securite-
valenciennes	access-	dynamis+	emploi+	taille+	cadre+	culture+	immo-	meteo+	soins+	securite-
nîmes	access-	dynamis+	emploi-	taille-	cadre+	culture-	immo+	meteo+	soins+	securite-
perpignan	access+	dynamis-	emploi+	taille+	cadre-	culture+	immo+	meteo-	soins-	securite-
pau	access-	dynamis-	emploi-	taille-	cadre+	culture-	immo+	meteo-	soins-	securite+
brest	access-	dynamis-	emploi-	taille-	cadre-	culture-	immo+	meteo+	soins+	securite-
havre	access+	dynamis+	emploi+	taille-	cadre-	culture+	immo+	meteo+	soins+	securite+
annecy	access+	dynamis-	emploi+	taille+	cadre-	culture+	immo-	meteo-	soins+	securite-
mulhouse	access+	dynamis+	emploi+	taille+	cadre+	culture+	immo-	meteo-	soins+	securite+
rochelle	access-	dynamis+	emploi-	taille-	cadre+	culture-	immo-	meteo+	soins+	securite+
béthune	access+	dynamis+	emploi+	taille+	cadre+	culture+	immo-	meteo-	soins-	securite-
dunkerque	access-	dynamis+	emploi-	taille-	cadre-	culture+	immo+	meteo+	soins-	securite+
troyes	access-	dynamis-	emploi-	taille-	cadre+	culture-	immo+	meteo+	soins-	securite-
valence	access+	dynamis+	emploi+	taille+	cadre-	culture+	immo-	meteo-	soins+	securite+
angoulême	access+	dynamis-	emploi-	taille-	cadre-	culture-	immo+	meteo-	soins-	securite+
bayonne	access-	dynamis+	emploi+	taille+	cadre+	culture+	immo-	meteo+	soins-	securite-
lorient	access-	dynamis+	emploi+	taille+	cadre-	culture+	immo-	meteo+	soins+	securite-
saint nazaire	access+	dynamis+	emploi+	taille-	cadre-	culture+	immo-	meteo+	soins+	securite+
annemasse	access+	dynamis-	emploi-	taille-	cadre-	culture-	immo+	meteo+	soins-	securite-
montbéliard	access-	dynamis-	emploi-	taille-	cadre+	culture-	immo+	meteo+	soins+	securite-

Figure 6 : Base de données transformée - variables qualitatives à 2 modalités



Graphique 28 : Répartition des modalités + et - dans la base transformée à 2 modalités.

7.1.2 Variables qualitatives à trois modalités

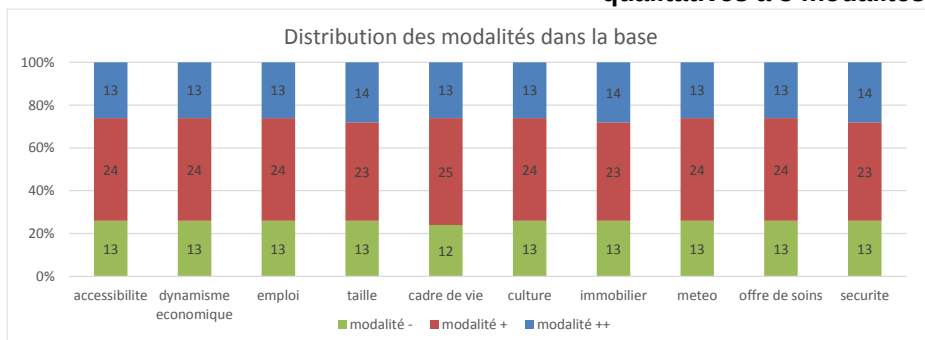


Graphique 29 : Transformation d'une variable continue en variable discrète à 3 modalités

Pour chaque variable, on considère que les valeurs supérieures au 3^{ème} quartile sont très appréciées et notées (++), les valeurs dans l'intervalle interquartile sont appréciées et notées (+) et les valeurs inférieures au 1^{er} quartile sont défavorables et notées (-). Par exemple, pour la variable « Accessibilité », comme le montre le Graphique 29, si la valeur de la variable « Accessibilité » est inférieure au 1^{er} quartile (8,55) alors la nouvelle variable discrète prend la modalité « Access- », si cette valeur est dans l'intervalle interquartile (entre 8,55 et 13,85) alors la nouvelle variable prend la modalité « Access+ » et si la valeur est supérieure au 3^{ème} quartile (13,85) alors la nouvelle variable prend la modalité « Access ++ ».

ville	accessibilité	dynamisme économique	emploi	taille	cadre de vie	culture	immobilier	meteo	offre de soins	securite
lyon	access++	dynamis+	emploi+	taille+	cadre+	culture+	immo+	meteo+	soins+	securite+
marseille	access+	dynamis+	emploi+	taille+	cadre+	culture+	immo+	meteo+	soins+	securite++
toulouse	access+	dynamis-	emploi-	taille-	cadre+	culture-	immo++	meteo+	soins++	securite++
bordeaux	access-	dynamis-	emploi+	taille+	cadre+	culture+	immo-	meteo+	soins+	securite+
rennes	access-	dynamis-	emploi+	taille-	cadre-	culture-	immo+	meteo+	soins-	securite+
rouen	access+	dynamis++	emploi+	taille+	cadre+	culture+	immo+	meteo++	soins+	securite-
montpellier	access-	dynamis+	emploi+	taille-	cadre++	culture+	immo+	meteo+	soins++	securite++
Lille	access+	dynamis+	emploi+	taille+	cadre-	culture+	immo+	meteo+	soins++	securite++
metz	access+	dynamis+	emploi-	taille-	cadre+	culture-	immo++	meteo-	soins-	securite+
strasbourg	access+	dynamis++	emploi++	taille++	cadre+	culture++	immo-	meteo+	soins++	securite+
dermont ferrand	access-	dynamis+	emploi+	taille+	cadre++	culture+	immo++	meteo-	soins+	securite++
nantes	access+	dynamis+	emploi+	taille+	cadre+	culture+	immo-	meteo-	soins++	securite+
nice	access+	dynamis-	emploi++	taille++	cadre+	culture+	immo+	meteo++	soins++	securite++
grenoble	access+	dynamis+	emploi++	taille+	cadre-	culture+	immo+	meteo+	soins+	securite++
orléans	access++	dynamis++	emploi+	taille+	cadre+	culture+	immo++	meteo+	soins-	securite-
dijon	access+	dynamis+	emploi-	taille-	cadre++	culture-	immo+	meteo-	soins-	securite-
caen	access+	dynamis+	emploi++	taille++	cadre+	culture++	immo-	meteo++	soins++	securite+
toulon	access+	dynamis+	emploi+	taille+	cadre++	culture+	immo+	meteo-	soins-	securite+
nancy	access++	dynamis+	emploi+	taille++	cadre+	culture++	immo-	meteo-	soins-	securite-
besancon	access+	dynamis+	emploi+	taille+	cadre+	culture+	immo++	meteo+	soins+	securite++
douai-lens	access-	dynamis-	emploi-	taille-	cadre++	culture-	immo++	meteo+	soins-	securite++
amiens	access++	dynamis+	emploi++	taille++	cadre+	culture++	immo-	meteo++	soins-	securite-
limoges	access++	dynamis-	emploi+	taille+	cadre+	culture-	immo++	meteo+	soins+	securite++
reims	access+	dynamis++	emploi++	taille++	cadre++	culture++	immo-	meteo++	soins++	securite-
angers	access++	dynamis+	emploi+	taille++	cadre-	culture+	immo++	meteo-	soins+	securite+
tours	access-	dynamis+	emploi-	taille-	cadre-	culture-	immo++	meteo+	soins-	securite++
poitiers	access+	dynamis++	emploi+	taille++	cadre++	culture++	immo-	meteo++	soins++	securite-
avignon	access+	dynamis-	emploi-	taille-	cadre-	culture+	immo+	meteo+	soins-	securite+
saint etienne	access++	dynamis-	emploi+	taille+	cadre-	culture++	immo+	meteo-	soins+	securite+
mans	access++	dynamis++	emploi++	taille++	cadre+	culture++	immo+	meteo+	soins+	securite+
valenciennes	access-	dynamis++	emploi++	taille++	cadre++	culture++	immo-	meteo++	soins+	securite-
nîmes	access-	dynamis++	emploi-	taille+	cadre++	culture-	immo+	meteo++	soins++	securite-
perpignan	access++	dynamis-	emploi++	taille+	cadre-	culture+	immo+	meteo+	soins+	securite+
pau	access-	dynamis+	emploi+	taille+	cadre+	culture+	immo+	meteo+	soins+	securite++
brest	access-	dynamis+	emploi-	taille+	cadre++	culture+	immo+	meteo++	soins++	securite-
havre	access++	dynamis+	emploi+	taille+	cadre+	culture+	immo++	meteo+	soins+	securite++
annecy	access++	dynamis-	emploi+	taille+	cadre-	culture-	immo+	meteo+	soins+	securite-
mulhouse	access+	dynamis++	emploi++	taille++	cadre+	culture++	immo-	meteo+	soins++	securite++
rochelle	access+	dynamis+	emploi-	taille-	cadre++	culture+	immo-	meteo+	soins+	securite+
béthune	access++	dynamis+	emploi+	taille++	cadre+	culture++	immo-	meteo-	soins+	securite+
dunkerque	access+	dynamis+	emploi+	taille+	cadre+	culture+	immo++	meteo++	soins+	securite+
troyes	access-	dynamis++	emploi-	taille-	cadre++	culture-	immo+	meteo+	soins-	securite+
valence	access+	dynamis+	emploi++	taille++	cadre-	culture++	immo+	meteo+	soins++	securite+
angoulême	access+	dynamis-	emploi-	taille-	cadre-	culture-	immo+	meteo-	soins-	securite+
bayonne	access-	dynamis++	emploi+	taille+	cadre++	culture+	immo+	meteo++	soins++	securite+
lorient	access-	dynamis++	emploi++	taille++	cadre+	culture++	immo-	meteo++	soins+	securite-
saint nazaire	access++	dynamis+	emploi++	taille+	cadre+	culture+	immo+	meteo+	soins+	securite+
annemasse	access+	dynamis-	emploi-	taille-	cadre-	culture-	immo++	meteo+	soins-	securite+
montbéliard	access+	dynamis-	emploi-	taille-	cadre+	culture-	immo++	meteo++	soins+	securite+
thionville	access+	dynamis++	emploi+	taille+	cadre+	culture-	immo+	meteo-	soins-	securite-

Figure 7 : Base de données transformée - variables qualitatives à 3 modalités



Graphique 30 : Répartition des modalités + et - dans la base transformée à 3 modalités.

7.2 Dendrogrammes

Cf. détails dans le document Excel ci-joint.

7.3 ACM et classification avec variables à 2 modalités

Cf. détails dans le document Excel ci-joint.

7.4 ACM et classification avec variables à 3 modalités

Cf. détails dans le document Excel ci-joint.

CONCLUSION ET PERSPECTIVES

L'essentiel de la démarche d'analyse a été présenté dans ce document et il reste encore d'autres détails intéressants à voir dans le document Excel joint à cette analyse.

Nous avons utilisé des outils et des méthodes d'analyse de données mais il reste d'autres outils que nous n'avons pas pu détailler encore par manque de temps. Par exemple l'analyse des correspondances multiples et les autres méthodes de classification peuvent être exploitées dans cette analyse. Les autres outils et méthodes peuvent donner des résultats un peu différents. C'est pour cela qu'on reste prudent en donnant des conclusions ou des caractéristiques des individus ou groupes d'individus.