



02/05/2018

PROJET STA101

Analyse de données : méthodes descriptives

A l'attention de Monsieur Pierre-Louis GONZALEZ,
Professeur et Maître de Conférences au CNAM, Paris.

Guillaume CHEVRON

CONSERVATOIRE NATIONAL DES ARTS ET METIERS

le **cnam**

Table des matières

Table des matières	1
Table des illustrations	2
I. Introduction.....	3
A. Objectif de l'analyse	3
B. Sources des données.....	4
C. Préparation des données	5
1. Individus actifs et illustratifs.....	5
2. Variables actives et illustratives	6
D. Choix des données	7
E. Statistiques préliminaires	7
II. Analyse en composantes principales.....	8
A. Statistiques sommaires	8
B. Choix du nombre d'axes factoriels	10
C. Interprétation des principaux axes choisis.....	11
1. Premier facteur principal.....	11
2. Second facteur principal.....	12
3. Troisième facteur principal.....	13
4. Quatrième facteur principal	14
5. Cinquième facteur principal	15
D. Interprétation des plans factoriels.....	16
1. Interprétation des cercles des corrélations.....	16
2. Projection des individus sur les plans factoriels.....	19
E. Conclusion de l'analyse en composante principale.....	20
III. Classification avec la méthode hiérarchique de Ward	21
A. Objectif de la classification.....	21
B. Analyse des résultats de la classification	21
C. Composition des classes.....	23
Annexes	24

Table des illustrations

Figure 1 – Individus actifs (communes avec une population supérieure à 75 000 habitants)	5
Figure 2 – Variables liées au secteur de l'autocar en France	6
Figure 3 – Variables de nature démographiques et socioprofessionnelles issues de l'INSEE	6
Figure 4 – statistiques préliminaires sur les variables actives.....	7
Figure 5 – Matrice des valeurs-tests des variables actives	8
Figure 6 – Matrice des corrélations des variables actives	9
Figure 7 – Tableau des valeurs propres	10
Figure 8 – Individus actifs à forte contribution sur le premier facteur principal.....	11
Figure 9 – Variables actives à forte contribution sur le premier facteur principal.....	11
Figure 10 – Individus actifs à forte contribution sur le second facteur principal.....	12
Figure 11 – Variables actives à forte contribution sur le second facteur principal	12
Figure 12 – Individus actifs à forte contribution sur le troisième facteur principal.....	13
Figure 13 – Variables actives à forte contribution sur le troisième facteur principal.....	13
Figure 14 – Variables actives à forte contribution sur le quatrième facteur principal.....	14
Figure 15 – Individus actifs à forte contribution sur le quatrième facteur principal	14
Figure 16 – Variables actives à forte contribution sur le cinquième facteur principal.....	15
Figure 17 – Individus actifs à forte contribution sur le cinquième facteur principal	15
Figure 18 – Cercle des corrélations des variables actives dans le plan factoriel 1-2.....	16
Figure 19 – Cercle des corrélations des variables actives dans le plan factoriel 1-3.....	17
Figure 20 – Cercle des corrélations des variables actives dans le plan factoriel 2-3.....	18
Figure 21 – Cercle des corrélations des variables actives dans le plan factoriel 2-4.....	19
Figure 22 – Indices de niveau & perte d'inertie	21
Figure 23 – Inertie interclasses et intraclasses.....	21
Figure 24 – Dendrogramme des individus actifs	22
Figure 25 – Informations sur les classes	23
Figure 26 – Cercle des corrélations des variables actives dans le plan factoriel 1-4.....	24
Figure 27 – Cercle des corrélations des variables actives dans le plan factoriel 1-5.....	24
Figure 28 – Cercle des corrélations des variables actives dans le plan factoriel 2-5.....	25
Figure 29 – Cercle des corrélations des variables actives dans le plan factoriel 3-4.....	25
Figure 30 – Cercle des corrélations des variables actives dans le plan factoriel 3-5.....	26
Figure 31 – Cercle des corrélations des variables actives dans le plan factoriel 4-5.....	26
Figure 32 – Projection des individus actifs sur le plan factoriel 1-2	27
Figure 33 – Projection des individus actifs sur le plan factoriel 1-3	28
Figure 34 – Projection des individus actifs sur le plan factoriel 1-4	29
Figure 35 – Projection des individus actifs sur le plan factoriel 1-5	30
Figure 36 – Projection des individus actifs sur le plan factoriel 2-3	31
Figure 37 – Projection des individus actifs sur le plan factoriel 2-4	32
Figure 39 – Projection des individus actifs sur le plan factoriel 2-5	33
Figure 40 – Projection des individus actifs sur le plan factoriel 3-5	34
Figure 41 – Projection des individus actifs sur le plan factoriel 4-5	35

I. Introduction

A. Objectif de l'analyse

La récente loi pour la croissance, l'activité et l'égalité des chances économiques du 6 août 2015 a libéralisé le transport régulier interurbain de voyageurs par autocar. Cette libéralisation a permis à plusieurs fournisseurs de transport spécialisé dans les autocars longue distance de se lancer à la conquête de ce marché et d'exploiter un nombre grandissant de lignes et de dessertes, reliant les différentes communes et régions de France.

Alors que de nombreuses sociétés privées et publiques ont tenté ce pari audacieux de concurrencer d'autres moyens de transport comme le covoiturage, le train ou encore l'avion, seules trois ont réussi à tirer leur épingle du jeu et à perdurer jusqu'à aujourd'hui. Ces trois fournisseurs de transport autocariste relient aujourd'hui un grand nombre de communes et espèrent ainsi devenir rentables à moyen terme,

Afin d'améliorer le ciblage de sa clientèle, l'une de ces trois entreprises souhaite faire le lien entre les différents canaux d'achat dont elle dispose et le profil de ses potentiels clients par le biais d'une analyse en composantes principales. Cette analyse a pour objet de décrire un jeu de données à partir de l'étude des variables et des individus, de le résumer et d'en réduire la dimensionnalité. L'étude des liaisons linéaires entre les variables – qu'elles soient quantitatives ou qualitatives – permet ainsi de chercher des variables synthétiques pouvant potentiellement résumer les différences entre les trois canaux de revenus de cette entreprise autocariste. Ces trois canaux sont les suivants :

- ➔ Les points de vente revendeurs : ce sont l'ensemble des sociétés partenaires (agences de voyages, offices de tourisme, boutiques spécialisées, restaurants, bureaux de tabac, gares routières, etc.) qui peuvent revendre des billets d'autocar de l'entreprise en question.
- ➔ Le site internet de la société : il s'agit de la plateforme web qui commercialise directement en ligne ses billets d'autocar
- ➔ L'application mobile : il s'agit de l'application mobile de la marque qui permet également l'achat de billets d'autocar par le biais d'un accès à internet

L'étude des individus du jeu de données permet quant à lui de s'intéresser à la variabilité des individus observés afin d'établir des profils distincts. Ici, les individus prennent la forme d'un échantillon de 51 communes françaises.

En d'autres termes, on cherche donc à identifier au sein d'un échantillon de 51 communes françaises (individus) de potentielles corrélations linéaires, liées au contexte démographique et socioprofessionnel (variables exogènes), permettant de résumer les différences entre les trois canaux de revenus de la société d'autocars (variables endogènes).

L'étude de ces corrélations et la projection des individus actifs dans les différents plans factoriels pourra dans un second temps permettre d'opérer une classification ascendante hiérarchique afin de segmenter cet échantillon en classes similaires.

B. Sources des données

Le jeu de données étudié dans cette analyse provient du recoupement de plusieurs sources de données :

La première partie du jeu de données est issue de la base de données de l'un des fournisseurs de transport spécialisé dans les autocars longue distance en France sur le dernier trimestre de l'année 2017 (Q4 2017). Pour des raisons de confidentialité, le nom de la société ne sera pas mentionné, et les données utilisées dans le cadre de cette étude pourraient avoir été modifiées volontairement au préalable.

La seconde partie du jeu de données provient directement de données INSEE en libre accès contenant notamment la liste des communes géolocalisées contenant l'intégralité des codes et informations [démographiques](#), [socioprofessionnelles](#) et [géographiques](#) sur les 36000 communes françaises (hors Mayotte).

C. Préparation des données

La préparation des données consiste à identifier la nature des individus et des variables que l'on souhaite prendre en compte au sein de l'étude en définissant :

- ➔ Les individus actifs (voire illustratifs)
- ➔ Les variables quantitatives et qualitatives

Une fois l'ensemble des individus et variables identifiées, il faut procéder à un travail de collecte des données requises afin de les agréger au sein d'un fichier texte (.txt) importable sous SPAD, et qui prend la forme d'une matrice de n individus et de p variables.

1. Individus actifs et illustratifs

Par souci de lisibilité, et afin de se concentrer sur les principaux points d'intérêt, les individus actifs ont été sélectionnés en fonction de leur taille. Dans le cas présent, ils sont représentés par les communes avec une population supérieure à soixante-quinze mille habitants. Etant donnée la nature du marché de l'autocar qui nécessite un maillage routier important et connecté, il n'est pas utile d'inclure à l'étude les communes hors France métropolitaine ainsi que les îles proches de l'océan Atlantique, de la Manche et de la mer Méditerranée, comme la Corse.

Chacune de ces communes possède un poids équivalent égal à 1/51^{ème} afin de ne pas donner plus d'importance à l'un des individus actifs plutôt qu'un autre. Il n'est pas nécessaire d'inclure des individus illustratifs à l'étude.

On obtient alors la liste des communes suivantes, numérotées aléatoirement de 1 à 51, incluant ainsi le code INSEE :

ID	Code INSEE	Communes	ID	Code INSEE	Communes
1	28267	Montreuil	26	34172	Montpellier
2	13001	Aix-En-Provence	27	68224	Mulhouse
3	94028	Créteil	28	54395	Nancy
4	92004	Asnières-Sur-Seine	29	44109	Nantes
5	93005	Aulnay-Sous-Bois	30	6088	Nice
6	92063	Rueil-Malmaison	31	30189	Nîmes
7	80021	Amiens	32	45234	Orleans
8	49007	Angers	33	75056	Paris
9	6004	Antibes	34	64445	Pau
10	84007	Avignon	35	66136	Perpignan
11	25056	Besançon	36	86194	Poitiers
12	33063	Bordeaux	37	51454	Reims
13	29019	Brest	38	35238	Rennes
14	14118	Caen	39	59512	Roubaix
15	63113	Clermont-Ferrand	40	76540	Rouen
16	21231	Dijon	41	11339	Saint-Denis
17	59183	Dunkerque	42	42218	Saint-Etienne
18	38185	Grenoble	43	57618	Saint-Louis
19	76351	Le Havre	44	67482	Strasbourg
20	72181	Le Mans	45	83137	Toulon
21	59350	Lille	46	31555	Toulouse
22	87085	Limoges	47	59599	Tourcoing
23	69123	Lyon	48	37261	Tours
24	13055	Marseille	49	78646	Versailles
25	57463	Metz	50	69266	Villeurbanne
			51	94081	Vitry-Sur-Seine

Figure 1 – Individus actifs (communes avec une population supérieure à 75 000 habitants)

2. Variables actives et illustratives

Comme indiqué précédemment, une partie des variables actives et illustratives provient de la base de données de l'un des fournisseurs de transport spécialisé dans les autocars longue distance en France sur le dernier trimestre de l'année 2017 :

Variables quantitative/qualitative	Statut	Définitions et commentaires
Pourcentage Revenus Part Revendeurs	Active	Pourcentage des revenus totaux générés à partir les revendeurs
Pourcentage Revenus Part Mobile	Active	Pourcentage des revenus totaux générés à partir l'application mobile
Pourcentage Revenus Part Web	Active	Pourcentage des revenus totaux générés à partir le site internet
Pourcentage de points de vente actifs	Active	Pourcentage des points de vente revendeurs ayant vendus au moins un billet
Distance moyenne de l'arrêt le plus proche	Active	Distance moyenne entre les points de vente revendeurs et l'arrêt de car le plus proche
Nombre de connexions	Active	Nombre de destinations possibles à partir d'une ville donnée
Nombre de points de vente revendeurs	Active	Nombre de points de vente qui peuvent revendre des billets
Nombre de « Ticket Shops »	Active	Nombre de revendeurs situés à moins de deux cents mètres de l'arrêt le plus proche
Nombre de « Ticket Shops Like »	Active	Nombre de revendeurs situés entre deux cents et huit cents mètres de l'arrêt le plus proche
Gares routières	Illustrative	La ville possède-t-elle une gare routière (Oui/Non) ?
Nombre de points de vente en compétition	Active	Nombre moyen de points de vente revendeurs pour un même arrêt

Figure 2 – Variables liées au secteur de l'autocar en France

Les termes de « Ticket Shops » et « Ticket Shop Like » sont des abréviations utilisées en interne par la société d'autocar pour désigner les points de vente revendeurs situés respectivement à moins de deux cents mètres et entre deux cent mètres des arrêts d'autocar les plus proches. Ces distances ont été choisies de manière empirique pour construire ces deux catégories.

L'autre partie du jeu de données provient d'un regroupement de données brutes INSEE (datant de 2013, 2014 et 2015) de natures démographiques, géographiques et socioprofessionnelles.

Variables quantitative & qualitative	Statut	Définitions et commentaires
Nombre d'habitants	Illustrative	Nombre total d'habitants
Nombre d'habitants 15-29 ans	Active	Nombre d'habitants dont l'âge est compris entre 15 et 29 ans
Nombre d'étrangers ¹	Active	Nombre d'habitants étrangers
Nombre d'immigrés	Active	Nombre d'habitants immigrés
Nombre d'étudiants	Active	Nombre d'habitants étudiants
Latitude	Active	Latitude
Longitude	Active	Longitude
Salaire net horaire moyen	Active	Valeur du salaire net horaire moyen (2015)
Salaire net horaire cadres sup.	Active	Valeur du salaire net horaire moyen des cadres supérieurs (2015)
Salaire net horaire prof. inter.	Active	Salaire net horaire moyen des professions intermédiaires (2015)
Salaire net horaire employés	Active	Salaire net horaire moyen des employés (2015)
Salaire net horaire ouvriers	Active	Salaire net horaire moyen des ouvriers (2015)

Figure 3 – Variables de nature démographiques et socioprofessionnelles issues de l'INSEE

¹ Selon la définition adoptée par le Haut Conseil à l'Intégration, un immigré est une personne née étrangère à l'étranger et résidant en France. Les personnes nées françaises à l'étranger et vivant en France ne sont donc pas comptabilisées.

Les variables illustratives « Gares routières » et « Nombre d'habitants » ont pour objectif de détecter de potentielles liaisons linéaires d'une part en fonction de la présence ou non d'une gare routière au sein des villes observées ; et d'autre part en fonction de la taille des villes par le biais de leur population.

D. Choix des données

Les données ont été sélectionnées de manière empirique et/ou intuitive, principalement à partir des résultats d'études passées, de la connaissance du marché et de l'approche terrain. Certaines variables annexes ont été retenues afin de vérifier ou de réfuter des observations effectuées de façon pratique.

E. Statistiques préliminaires

Les statistiques préliminaires peuvent être obtenues par le biais de la fonctionnalité « Statistiques sommaires des variables continues pour les individus actifs » de SPAD. Cet outil complémentaire calcule la moyenne, l'écart-type, le minimum et le maximum de chaque variable active prise en compte par l'étude, et donne les résultats suivants :

Libellé de la variable	Moyenne	Écart-type	Minimum	Maximum
Pourcentage Revenus Part Revendeurs	0,058	0,082	0,000	0,378
Pourcentage Revenus Part Mobile	0,159	0,059	0,000	0,325
Pourcentage Revenus Part Web	0,794	0,084	0,493	1,000
Pourcentage de points de vente actifs	0,423	0,190	0,048	1,000
Distance moyenne de l'arrêt le plus proche	2,189	1,896	0,400	8,710
Nombre de connexions	92,255	162,155	0,000	1112,000
Nombre de points de vente revendeurs	8,039	11,051	1,000	71,000
Nombre de « Ticket Shops »	0,333	0,548	0,000	2,000
Nombre de « Ticket Shops Like »	1,843	2,703	0,000	15,000
Nombre de points de vente en compétition	27,804	16,629	7,000	88,366
Nombre d'habitants entre 15-29 ans	53814,605	74227,960	12342,268	522244,885
Nombre total d'étrangers	22706,572	45370,867	3167,892	330167,096
Nombre total d'immigrés	31458,303	62971,431	4662,500	455485,854
Nombre total d'étudiants	23119,350	28069,044	2521,448	193636,000
Salaire net horaire moyen	14,256	2,842	11,820	25,152
Salaire net horaire cadres supérieurs	24,009	2,702	20,982	34,391
Salaire net horaire professions intermédiaires	14,172	1,022	12,750	17,612
Salaire net horaire employés	10,500	0,728	9,701	13,399
Salaire net horaire ouvriers	10,974	1,190	9,805	16,677
Latitude	47,004	2,418	42,683	51,050
Longitude	3,015	2,682	-4,483	7,750

Figure 4 – statistiques préliminaires sur les variables actives

Etant donné la très importante dispersion des écart-types entre les différentes variables actives (74227,960 pour le nombre d'habitant entre 15-29 ans et 0,059 pour le pourcentage de revenus générés par l'application mobile), il est très important d'effectuer une analyse en composantes principales en normalisant/standardisant les données, c'était-à-dire en pondérant chacune des variables par son écart-type et sa moyenne, afin de leur donner un point similaire. De ce fait, les variables sont dites « centrées réduites ».

II. Analyse en composantes principales

A. Statistiques sommaires

L'hypothèse de liaison linéaire entre deux variables est validée au seuil de 5% si et seulement si la valeur absolue du coefficient correspondant à ces variables dans la matrice des valeurs-tests est supérieure au seuil $R (= \frac{2}{\sqrt{96+2}} \approx 0.20)$, qui est spécifiée en rouge dans la matrice des valeurs-tests calculée sous SPAD ci-dessous :

	%REVEN	%MOBILE	%WEB	%ACTIF	DISTANC	NBCONEX	NBREVEN	NBTS	NBTSL	NBCOMP	NB15-29
%REVEN											
%MOBILE	-3,834										
%WEB	-2,666	-7,276									
%ACTIF	-1,112	-2,267	-0,083								
DISTANC	-2,976	3,143	-3,105	2,177							
NBCONEX	1,605	1,830	-1,584	-2,225	-1,785						
NBREVEN	1,986	1,242	-1,063	-1,141	-1,891	>10					
NBTS	1,999	-2,515	1,025	-2,570	-4,207	6,232	6,156				
NBTSL	2,811	-1,126	0,385	-0,088	-4,420	4,994	6,733	4,703			
NBCOMP	1,625	1,266	-1,074	-3,981	0,405	7,941	>10	4,633	3,134		
NB15-29	1,541	2,016	-1,914	-2,623	-1,303	>10	>10	6,610	5,395	>10	
NBETRAN	1,116	3,420	-2,952	-2,238	-0,446	>10	>10	5,589	4,629	>10	>10
NBIMMIG	1,163	3,414	-3,032	-2,200	-0,405	>10	>10	5,604	4,913	>10	>10
NBETUDI	1,670	1,245	-1,273	-2,650	-1,612	>10	>10	7,036	5,840	>10	>10
SALMOY	-1,673	-0,088	-0,725	1,348	3,430	4,490	3,629	1,201	0,262	2,352	4,513
SMOYCAD	-1,715	1,557	-1,283	0,982	3,092	4,378	3,449	1,073	-0,063	1,183	4,028
SMOYPRO	-1,500	2,069	-3,727	0,913	5,209	3,745	2,899	0,383	-0,560	2,036	3,670
SMOYEMP	-1,929	0,326	-1,456	1,914	5,126	3,278	2,562	0,031	-0,351	2,314	3,375
SMOYUUV	-2,001	0,877	-0,604	1,947	4,409	2,100	1,634	-0,718	-0,603	0,528	2,176
LATI	-3,664	1,808	0,517	1,375	1,615	-0,866	-1,193	-1,567	-2,238	-1,933	-0,494
LONG	3,558	0,925	-2,248	-0,989	1,821	0,762	0,867	-2,891	0,553	0,654	-0,359

VARIABLES	NBETRAN	NBIMMIG	NBETUDI	SALMOY	SMOYCAD	SMOYPRO	SMOYEMP	SMOYUUV	LATT	LONG
%REVEN										
%MOBILE										
%WEB										
%ACTIF										
DISTANC										
NBCONEX										
NBREVEN										
NBTS										
NBTSL										
NBCOMP										
NB15-29										
NBETRAN										
NBIMMIG	>10									
NBETUDI	>10	>10								
SALMOY	4,724	4,740	4,264							
SMOYCAD	4,803	4,821	3,651	>10						
SMOYPRO	4,604	4,662	3,154	>10	>10					
SMOYEMP	3,915	3,950	3,028	>10	>10	>10				
SMOYUUV	2,781	2,819	1,818	>10	>10	>10	>10			
LATI	-0,110	-0,226	-0,443	1,182	1,741	0,672	1,054	2,356		
LONG	0,496	0,612	-0,625	-0,538	-0,342	1,165	-0,276	-0,075	-2,639	

Figure 5 – Matrice des valeurs-tests des variables actives

L'étude des valeurs-tests permet dès lors d'identifier la présence ou l'absence de liaison linéaire entre variables actives.

	%REVEN	%MOBILE	%WEB	%ACTIF	DISTANC	NBCONEX	NBREVEN	NBTS	NBTSL	NBCOMP	NB15-29
%REVEN	1,000										
%MOBILE	-0,313	1,000									
%WEB	-0,221	-0,557	1,000								
%ACTIF	-0,093	-0,188	-0,007	1,000							
DISTANC	-0,245	0,259	-0,256	0,181	1,000						
NBCONEX	0,134	0,152	-0,132	-0,185	-0,149	1,000					
NBREVEN	0,165	0,104	-0,089	-0,095	-0,157	0,931	1,000				
NBTS	0,166	-0,208	0,086	-0,213	-0,342	0,488	0,483	1,000			
NBTSL	0,232	-0,094	0,032	-0,007	-0,358	0,401	0,522	0,379	1,000		
NBCOMP	0,136	0,106	-0,090	-0,324	0,034	0,598	0,688	0,374	0,258	1,000	
NB15-29	0,129	0,168	-0,159	-0,217	-0,109	0,940	0,939	0,514	0,430	0,711	1,000
NBETRAN	0,093	0,281	-0,244	-0,186	-0,037	0,927	0,901	0,444	0,374	0,664	0,962
NBIMMIG	0,097	0,280	-0,250	-0,183	-0,034	0,927	0,906	0,445	0,395	0,666	0,964
NBETUDI	0,139	0,104	-0,106	-0,219	-0,134	0,929	0,934	0,542	0,461	0,688	0,991
SALMOY	-0,139	-0,007	-0,061	0,113	0,281	0,363	0,297	0,100	0,022	0,195	0,365
SMOYCAD	-0,143	0,130	-0,107	0,082	0,255	0,355	0,283	0,090	-0,005	0,099	0,328
SMOYPRO	-0,125	0,172	-0,305	0,076	0,416	0,306	0,239	0,032	-0,047	0,169	0,300
SMOYEMP	-0,161	0,027	-0,122	0,159	0,410	0,270	0,212	0,003	-0,029	0,192	0,277
SMOYUUV	-0,166	0,073	-0,051	0,162	0,357	0,175	0,136	-0,060	-0,050	0,044	0,181
LATI	-0,300	0,151	0,043	0,115	0,135	-0,072	-0,100	-0,131	-0,186	-0,161	-0,041
LONG	0,292	0,077	-0,187	-0,083	0,152	0,064	0,072	-0,239	0,046	0,055	-0,030

	NBETRAN	NBIMMIG	NBETUDI	SALMOY	SMOYCAD	SMOYPRO	SMOYEMP	SMOYUUV	LATT	LONG
%REVEN										
%MOBILE										
%WEB										
%ACTIF										
DISTANC										
NBCONEX										
NBREVEN										
NBTS										
NBTSL										
NBCOMP										
NB15-29										
NBETRAN	1,000									
NBIMMIG	0,999	1,000								
NBETUDI	0,933	0,936	1,000							
SALMOY	0,381	0,382	0,346	1,000						
SMOYCAD	0,387	0,388	0,299	0,926	1,000					
SMOYPRO	0,372	0,376	0,260	0,890	0,851	1,000				
SMOYEMP	0,319	0,322	0,250	0,962	0,874	0,918	1,000			
SMOYUUV	0,230	0,233	0,151	0,898	0,881	0,873	0,903	1,000		
LATT	-0,009	-0,019	-0,037	0,099	0,145	0,056	0,088	0,195	1,000	
LONG	0,041	0,051	-0,052	-0,045	-0,029	0,097	-0,023	-0,006	-0,218	1,000

Figure 6 – Matrice des corrélations des variables actives

On en déduit donc que l'hypothèse de liaison linéaire au seuil de 5% est rejetée pour les couples de variables suivants :

- ➔ Le salaire net horaire moyen avec la part des revenus générés via l'application mobile
- ➔ Le pourcentage de revendeurs actifs (ayant vendus) avec la part des revenus générés via le site internet
- ➔ Le nombre de « Ticket Shops Like » avec le pourcentage de revendeurs actifs (ayant vendus)
- ➔ Le salaire horaire net moyen des employés avec le nombre de « Ticket Shops »
- ➔ Le salaire horaire net moyen des cadres supérieurs avec le nombre de « Ticket Shops Like »
- ➔ La latitude avec le nombre avec le nombre d'étrangers
- ➔ La longitude avec le salaire horaire net moyen des ouvriers

Le reste des couples de variables semble avoir été validée par l'hypothèse de liaison linéaire au seuil de 5%.

B. Choix du nombre d'axes factoriels

Le calcul des valeurs propres à partir de la diagonalisation de la matrice des variances/covariances permet d'obtenir le tableau suivant :

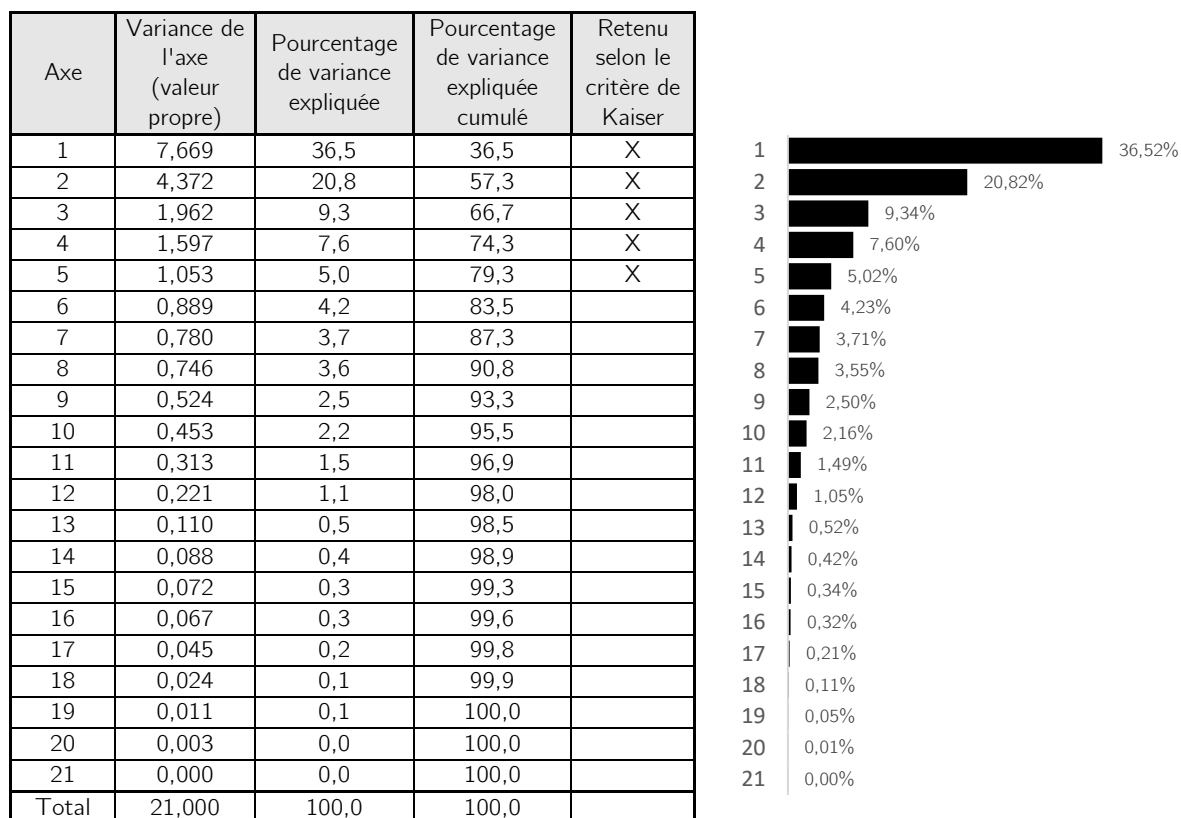


Figure 7 – Tableau des valeurs propres

La somme des valeurs propres est bien égale au nombre de variables actives centrées réduites, ce qui vérifie la fiabilité du calcul.

On observe une décroissance (ou un éboulis) des valeurs propres. Il faut alors définir le seuil auquel on s'arrête afin de d'obtenir une inertie suffisante à la bonne fiabilité de l'étude. En suivant le critère du coude, on peut tenir compte d'au moins au moins quatre axes (voire cinq) parmi les vingt-et-un proposés. Par souci de rigueur et de simplicité, et en suivant le critère de Kaiser, on choisit donc de retenir cinq axes, ce qui représente une part d'inertie de 79,3%.

C. Interprétation des principaux axes choisis

1. Premier facteur principal

Le tableau ci-dessous indique la contribution des individus actifs à l'inertie du premier axe principal et la qualité de leur représentation par cet axe :

Identifiant	Poids	Coordonnée	Contribution
Brest	1,000	-3,421	1,052
Roubaix	1,000	-3,109	0,869
Tourcoing	1,000	-3,083	0,855
Saint-Louis	1,000	-2,918	0,765
ZONE CENTRALE			
Toulouse	1,000	3,052	0,838
Rueil-Malmaison	1,000	3,525	1,118
Versailles	1,000	4,240	1,616
Lyon	1,000	5,318	2,543
Marseille	1,000	7,157	4,606
Paris	1,000	29,266	77,023
TOTAL			91,285

Figure 8 – Individus actifs à forte contribution sur le premier facteur principal

On constate que l'individu « Paris » est le plus éloigné de l'origine des 51 individus actifs et contribue très fortement (à près de 84%) à ce premier axe. On aurait pu intuitivement deviner cette observation compte tenu de la nature particulière de la capitale française :

- ➔ Capitale économique française
- ➔ Ville la plus visitée au monde
- ➔ Maillage routier centralisé
- ➔ Importants pôles universitaires

Cette intuition se vérifie à l'observation des variables actives ayant le plus contribué à ce premier axe principal :

Libellé de la variable	Poids	Moyenne	Écart-type	Coordonnée	Contribution
ZONE CENTRALE					
Salaire net horaire moyen	51	14,256	2,842	0,625	5,1
Nombre de points de vente en compétition	51	27,804	16,629	0,674	5,9
Nombre de points de vente revendeurs	51	8,039	11,051	0,888	10,3
Nombre de connexions	51	92,255	162,155	0,906	10,7
Nombre total d'étudiants	51	23119,350	28069,044	0,915	10,9
Nombre d'habitants entre 15-29 ans	51	53814,605	74227,960	0,937	11,4
Nombre total d'étrangers	51	22706,572	45370,867	0,939	11,5
Nombre total d'immigrés	51	31458,303	62971,431	0,942	11,6
TOTAL					77,5

Figure 9 – Variables actives à forte contribution sur le premier facteur principal

En effet, la majorité des variables ayant les plus grandes contributions sont fortement liées à la taille de la ville observée.

2. Second facteur principal

Le tableau ci-dessous indique la contribution des individus actifs à l'inertie du second axe principal et la qualité de leur représentation par cet axe :

Identifiant	Poids	Coordonnée	Contribution
Marseille	1,000	-5,451	4,686
Toulon	1,000	-3,447	1,874
Toulouse	1,000	-3,443	1,870
Angers	1,000	-2,972	1,393
Paris	1,000	-2,595	1,062
Lille	1,000	-2,532	1,011
Avignon	1,000	-2,529	1,009
Montpellier	1,000	-2,446	0,943
Saint-Etienne	1,000	-2,411	0,917
ZONE CENTRALE			
Le Havre	1,000	2,302	0,836
Antibes	1,000	3,163	1,578
Aulnay-Sous-Bois	1,000	4,048	2,585
Créteil	1,000	4,436	3,104
Asnières-Sur-Seine	1,000	7,573	9,045
Versailles	1,000	13,048	26,852
Rueil-Malmaison	1,000	14,045	31,114
TOTAL			89,878

Figure 10 – Individus actifs à forte contribution sur le second facteur principal

Deux individus en particulier, Versailles et Rueil-Malmaison, se démarquent du reste de l'échantillon. D'autres villes comme Asnières-Sur-Seine, Créteil ou encore Aulnay-Sous-Bois ressortent également.

En observant les variables actives ayant le plus contribuées à ce second axe, on constate que cet axe est presque exclusivement composé de variables ayant attiré au salaire moyen, ce qui laisse supposer que les individus contributeurs se démarquent fortement de par leur particularité en termes de salaire horaire net moyen : certains sont probablement élevés (Rueil-Malmaison, Versailles et Asnières-Sur-Seine) quand d'autres sont probablement plus faibles (Aulnay-Sous-Bois, Créteil).

Libellé de la variable	Poids	Moyenne	Écart-type	Coordonnée	Contribution
ZONE CENTRALE					
Distance moyenne de l'arrêt le plus proche	51	2,189	1,896	0,592	8,0
Salaire net horaire cadres supérieurs	51	24,009	2,702	0,699	11,2
Salaire net horaire moyen	51	14,256	2,842	0,705	11,4
Salaire net horaire professions intermédiaires	51	14,172	1,022	0,738	12,5
Salaire net horaire employés	51	10,500	0,728	0,773	13,7
Salaire net horaire ouvriers	51	10,974	1,190	0,808	14,9
TOTAL					71,7

Figure 11 – Variables actives à forte contribution sur le second facteur principal

3. Troisième facteur principal

Difficile de comprendre intuitivement le positionnement des individus ayant les plus importantes contributions à partir de la valeur de ces mêmes contributions et des coordonnées calculées. En effet, le troisième axe semble scindé en deux groupes, l'un ayant des coordonnées positives et l'autre ayant des coordonnées négatives.

Identifiant	Poids	Coordonnée	Contribution
Aulnay-Sous-Bois	1,000	-6,144	13,266
Saint-Louis	1,000	-4,395	6,788
Vitry-Sur-Seine	1,000	-4,273	6,416
Montreuil	1,000	-3,532	4,383
Créteil	1,000	-3,143	3,471
Saint-Denis	1,000	-3,131	3,446
Asnières-Sur-Seine	1,000	-2,671	2,508
Strasbourg	1,000	-2,572	2,325
Nîmes	1,000	-2,542	2,271
Villeurbanne	1,000	-1,788	1,124
Antibes	1,000	-1,756	1,084
Paris	1,000	-1,715	1,033
ZONE CENTRALE			
Grenoble	1,000	1,406	0,695
Lille	1,000	1,496	0,787
Caen	1,000	1,788	1,123
Poitiers	1,000	1,797	1,135
Pau	1,000	2,488	2,175
Bordeaux	1,000	2,631	2,432
Rennes	1,000	2,779	2,713
Brest	1,000	2,898	2,952
Nantes	1,000	3,092	3,359
Angers	1,000	3,388	4,035
Versailles	1,000	4,595	7,421
Rueil-Malmaison	1,000	4,701	7,767
Saint-Etienne	1,000	4,748	7,924
TOTAL			92,634

Figure 12 – Individus actifs à forte contribution sur le troisième facteur principal

En observant les variables ayant le plus contribué à ce troisième axe, et compte tenu du poids des deux principales variables actives (32,1 pour la part de revenus mobile et 25,9 pour la part de revenus web), on comprend qu'il s'agit probablement d'une représentation qui fait ressortir les individus ayant un comportement d'achat très marqué : les uns préfèrent acheter via le site internet quand les autres optent pour l'application mobile.

Libellé de la variable	Poids	Moyenne	Écart-type	Coordonnée	Contribution
Pourcentage Revenus Part Mobile	51	0,159	0,059	-0,794	32,1
Distance moyenne de l'arrêt le plus proche	51	2,189	1,896	-0,425	9,2
Longitude	51	3,015	2,682	-0,402	8,2
ZONE CENTRALE					
Nombre de « Ticket Shops »	51	0,333	0,548	0,400	8,1
Pourcentage Revenus Part Web	51	0,794	0,084	0,713	25,9
TOTAL					83,6

Figure 13 – Variables actives à forte contribution sur le troisième facteur principal

4. Quatrième facteur principal

En observant les variables les plus contributrices au quatrième axe principal, on observe qu'il s'agit avant tout d'un axe géographique qui oppose longitude (26.1) et latitude (23.7).

Toutefois, la part de revenus générée par les revendeurs semble également être liée linéairement à la position géographique de l'individu avec une contribution très élevée (31.6).

Libellé de la variable	Poids	Moyenne	Écart-type	Coordonnée	Contribution
Pourcentage Revenus Part Revendeurs	51	0,058	0,082	-0,711	31,6
Longitude	51	3,015	2,682	-0,646	26,1
ZONE CENTRALE					
Pourcentage Revenus Part Mobile	51	0,159	0,059	0,327	6,7
Latitude	51	47,004	2,418	0,615	23,7
TOTAL					88,2

Figure 14 – Variables actives à forte contribution sur le quatrième facteur principal

On retrouve cette dichotomie dans le découpage des individus ayant le plus contribué au quatrième axe :

- ➔ Sud et Est : Toulon, Nice, Marseille, Mulhouse, Toulouse
- ➔ Nord et Ouest : Brest, Rennes, Caen, Le Havre, Paris, Montreuil

Identifiant	Poids	Coordonnée	Contribution
Toulon	1,000	-6,857	20,302
Nice	1,000	-6,090	16,015
Marseille	1,000	-3,415	5,036
Mulhouse	1,000	-2,917	3,674
Toulouse	1,000	-2,602	2,925
Asnières-Sur-Seine	1,000	-2,309	2,303
Antibes	1,000	-2,242	2,170
Clermont-Ferrand	1,000	-1,965	1,668
Grenoble	1,000	-1,712	1,266
Besançon	1,000	-1,659	1,189
Rueil-Malmaison	1,000	-1,590	1,092
Aix-En-Provence	1,000	-1,442	0,898
ZONE CENTRALE			
Le Mans	1,000	1,524	1,003
Angers	1,000	1,729	1,291
Dunkerque	1,000	1,822	1,434
Aulnay-Sous-Bois	1,000	1,823	1,435
Lille	1,000	1,894	1,549
Vitry-Sur-Seine	1,000	1,925	1,601
Amiens	1,000	2,208	2,105
Le Havre	1,000	2,452	2,596
Caen	1,000	2,778	3,333
Rennes	1,000	2,860	3,532
Montreuil	1,000	2,908	3,652
Paris	1,000	3,191	4,397
Brest	1,000	3,627	5,680
TOTAL			92,149

Figure 15 – Individus actifs à forte contribution sur le quatrième facteur principal

5. Cinquième facteur principal

Le cinquième et dernier axe semble consacré aux variables liées aux points de vente des revendeurs, en particulier au nombre de points de vente ayant vendus au moins un billet d'autocars sur la période observée. Cette variable semble s'opposer au nombre de points de vente pour un même arrêt.

Libellé de la variable	Poids	Moyenne	Écart-type	Coordonnée	Contribution
Nombre de points de vente en compétition	51	27,804	16,629	-0,285	7,7
Pourcentage Revenus Part Web	51	0,794	0,084	-0,240	5,5
ZONE CENTRALE					
Latitude	51	47,004	2,418	0,244	5,7
Nombre de « Ticket Shops Like »	51	1,843	2,703	0,336	10,7
Pourcentage de points de vente actifs	51	0,423	0,190	0,813	62,8
TOTAL					92,4

Figure 16 – Variables actives à forte contribution sur le cinquième facteur principal

Deux individus, Créteil (17,03) et Saint-Denis (12,64), semblent contribuer davantage au cinquième axe.

Identifiant	Poids	Coordonnée	Contribution
Saint-Denis	1,000	-4,394	12,641
Aix-En-Provence	1,000	-2,921	5,588
Rennes	1,000	-2,447	3,922
Antibes	1,000	-2,412	3,809
Versailles	1,000	-2,388	3,734
Vitry-Sur-Seine	1,000	-2,061	2,780
Toulouse	1,000	-1,922	2,418
Caen	1,000	-1,837	2,209
Saint-Etienne	1,000	-1,790	2,099
Reims	1,000	-1,658	1,801
Orléans	1,000	-1,482	1,437
Metz	1,000	-1,105	0,799
ZONE CENTRALE			
Rouen	1,000	1,116	0,816
Besançon	1,000	1,200	0,943
Roubaix	1,000	1,366	1,222
Villeurbanne	1,000	1,529	1,530
Grenoble	1,000	1,590	1,654
Marseille	1,000	1,614	1,706
Strasbourg	1,000	1,647	1,776
Mulhouse	1,000	1,924	2,423
Poitiers	1,000	1,995	2,606
Rueil-Malmaison	1,000	2,074	2,816
Tours	1,000	2,303	3,474
Dunkerque	1,000	2,501	4,095
Brest	1,000	3,640	8,673
Créteil	1,000	5,100	17,026
TOTAL			93,993

Figure 17 – Individus actifs à forte contribution sur le cinquième facteur principal

D. Interprétation des plans factoriels

1. Interprétation des cercles des corrélations

Le plan factoriel 1-2 représente 57,3% de l'inertie totale, dont près de 36.5% et 20.8% proviennent respectivement des axes 1 et 2.

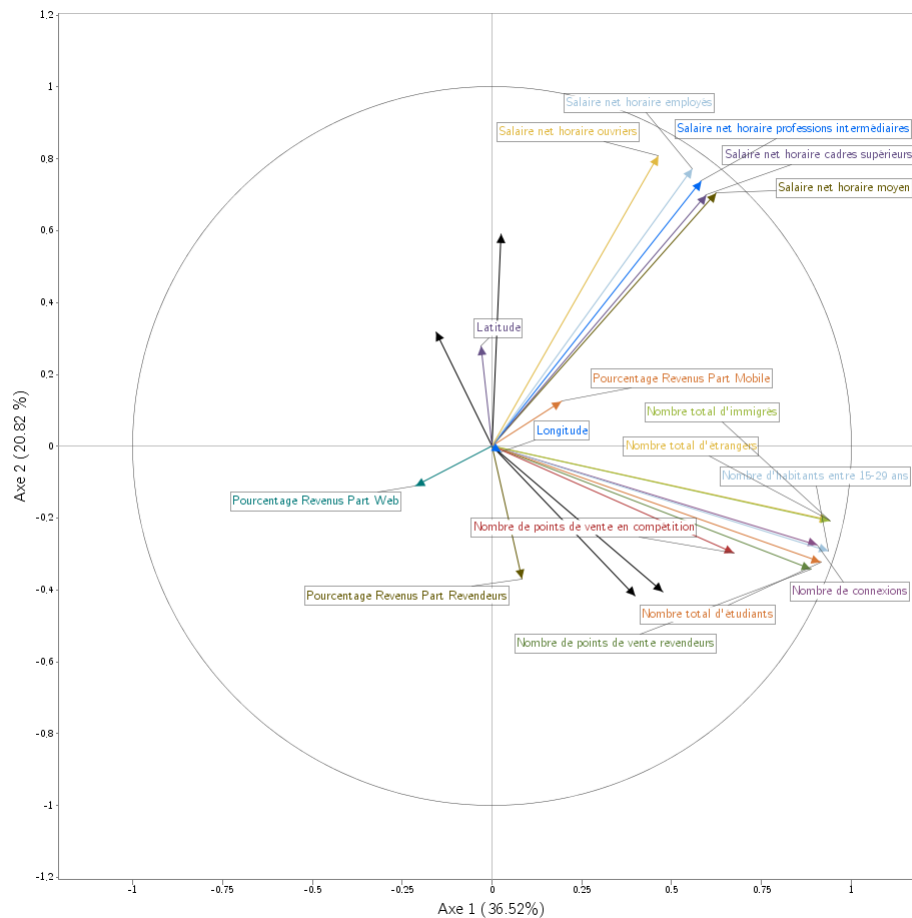


Figure 18 – Cercle des corrélations des variables actives dans le plan factoriel 1-2

Les variables les mieux représentées sont celles qui ont les plus importantes contributions au sein des deux axes observés. Globalement, elles se répartissent en trois groupes :

- ➔ Des données socioprofessionnelles au travers du salaire horaire net moyen par catégories
- ➔ Des données démographiques par le biais de variables liées à la population
- ➔ Des données business propres à la société d'autocars

On constate notamment que ce plan factoriel semble ne pas indiquer de véritables corrélations entre salaire moyen par catégorie et nombre d'habitants. En effet, l'angle observé semble se rapprocher fortement de 90°, ce qui implique une corrélation mineure, voire quasi-inexistante entre les deux groupes de variables. De même, les trois variables liées aux canaux de revenus sont mal représentées sur ce plan factoriel, ce qui ne permet pas d'en déduire de potentielles liaisons linéaires pour le moment.

Toutefois, on observe une liaison linéaire entre le nombre de point de vente revendeurs de billets d'autocar et le nombre d'étudiants. Il est donc possible que la société d'autocar ait davantage ciblée les villes fortement peuplées ou possédant une population étudiante importante dans sa stratégie d'acquisition de points de vente revendeurs.

De même, le nombre de connexions, c'est-à-dire le nombre de destinations en partance d'une ville observée, est également fortement corrélé à la taille de la ville, ce qui s'explique assez intuitivement par l'importance du maillage routier français.

Ces deux liaisons linéaires sont également très visibles sur le plan factoriel 1-3 avec des angles très faibles entre le nombre de points de vente revendeurs et le nombre d'étudiants ainsi qu'entre le nombre de connexions et le nombre d'habitants entre 15 et 29 ans.

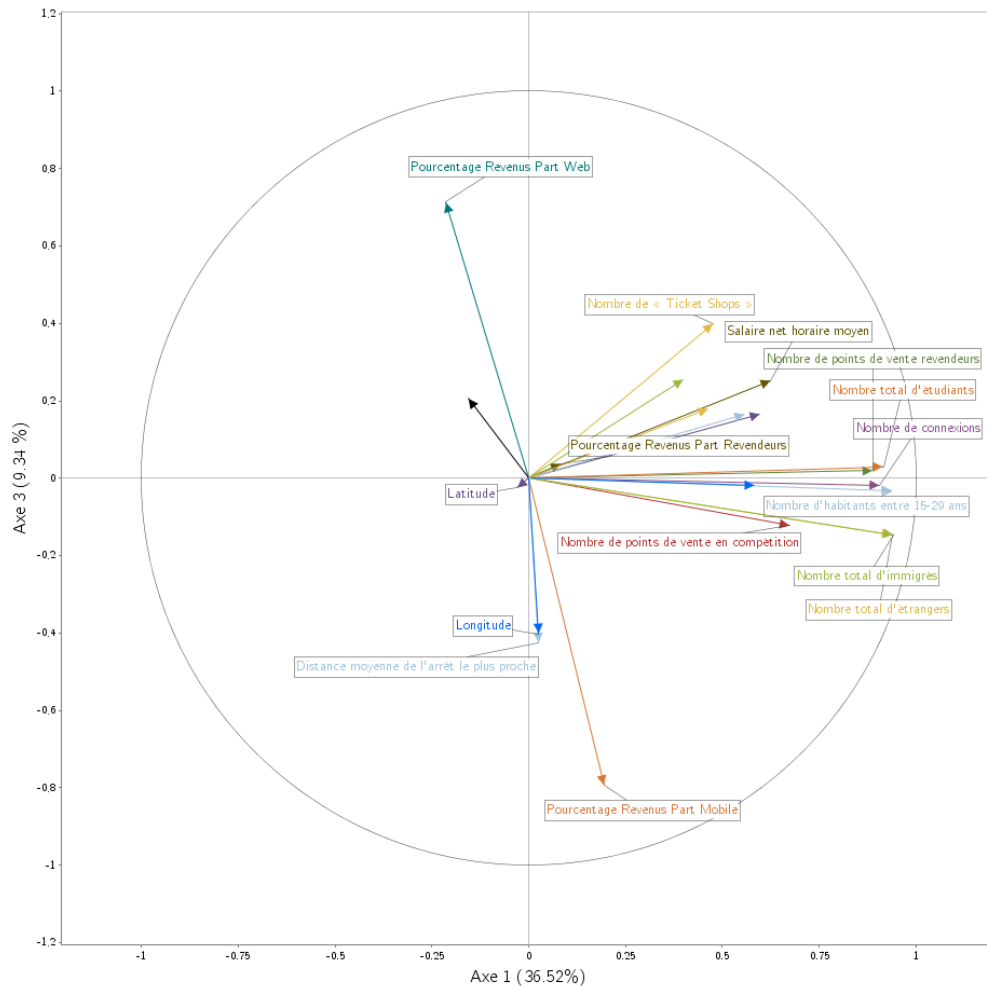


Figure 19 – Cercle des corrélations des variables actives dans le plan factoriel 1-3

A noter que deux des trois variables liées aux canaux de revenus sont également bien représentées au travers de ce second plan. En effet, on constate que les pourcentages de revenus générés par le biais de l'application mobile ou par le site internet semblent tous deux être peu corrélés avec les variables démographiques sélectionnées, en particulier avec le nombre d'étudiants ou de personnes entre 15 et 29 ans. De même, cette observation se confirme sur le plan factoriel 2-3 qui semble indiquer que ces deux variables sont également très peu corrélées aux variables socioprofessionnelles avec des angles entre variables proches de 90°. En somme, cela indique une quasi-absence de liaison linéaire entre le salaire et le choix du canal d'achat, et entre la taille de la population et ce même choix du canal d'achat.

Toutefois, il existe une forte corrélation entre ces deux variables liées aux canaux de revenus. En effet, elles se trouvent à près de 180° l'une de l'autre sur le cercle des corrélations du plan factoriel 1-3. Ce phénomène est également observable sur les plans factoriels 2-3 et 3-5 (figure 30 en annexes) avec un angle également proche de 180° entre les deux variables.

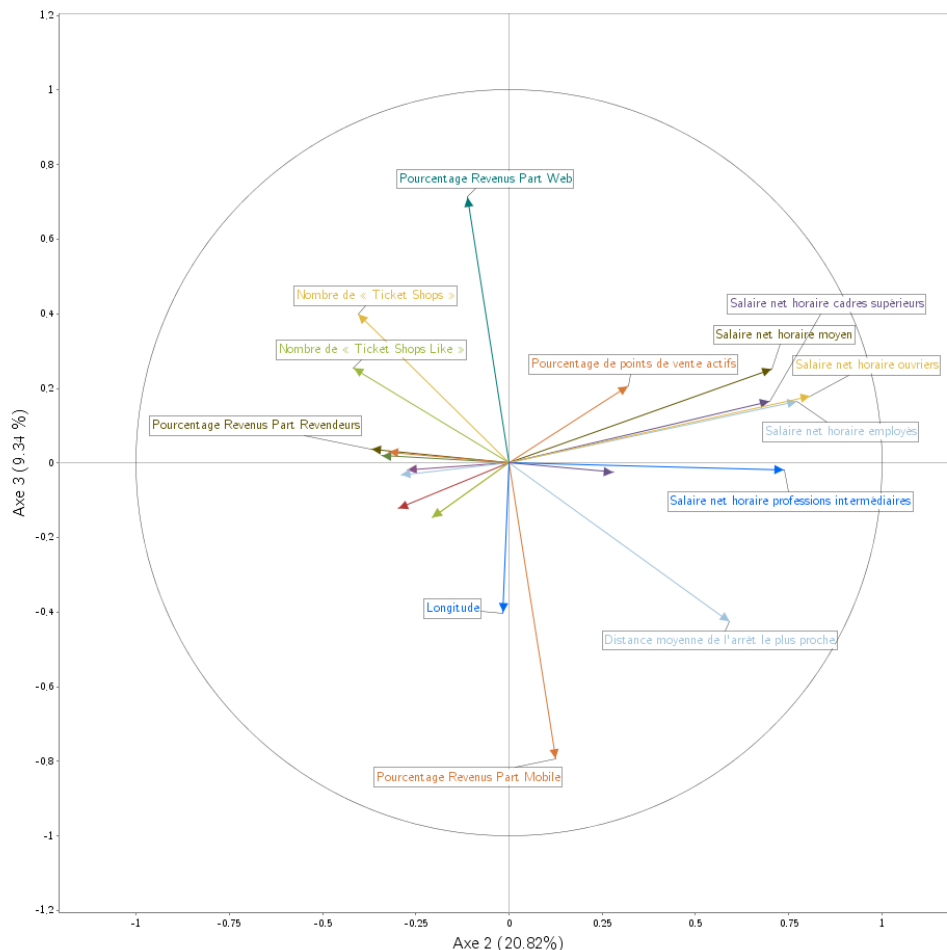


Figure 20 – Cercle des corrélations des variables actives dans le plan factoriel 2-3

Cela peut intuitivement s'expliquer par le fait que l'augmentation ou la diminution en pourcentage des revenus générés par le site internet entraîne mécaniquement et respectivement une diminution ou une augmentation des revenus générés par l'application mobile.

Cela ne semble néanmoins pas affecter de la même façon les revenus générés par la revente de billets car cette variable se trouve mal représentée sur le plan factoriel 1-3. De plus, il est difficile d'apporter une interprétation complémentaire à la lecture du plan factoriel 1-4 (figure 26 en annexes), sur lequel la variable liée aux revenus générés par la revente est cette fois-ci bien représentée, mais les deux autres variables se trouvent à leurs tours mal représentées.

Néanmoins, les plans factoriels 2-4, 3-4 (figure 29 en annexes) et 4-5 (figure 31 en annexes) permettent tout de même de déduire que la variable liée au pourcentage de revenus générés par la revente de billets semble être plutôt corrélée à la géolocalisation au travers des variables de latitude et de longitude.

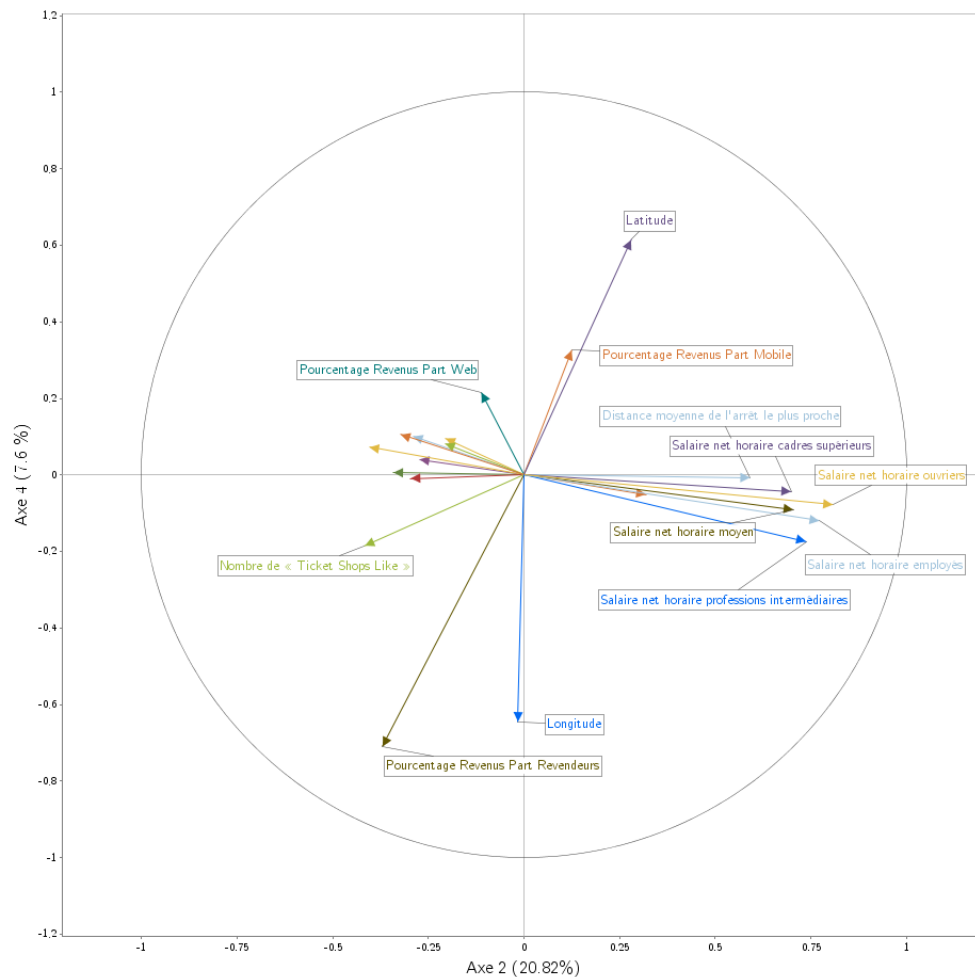


Figure 21 – Cercle des corrélations des variables actives dans le plan factoriel 2-4

Aussi, plus la longitude est grande (angle aux alentours de 30°) et la latitude faible (angle à 180°), plus le pourcentage de revenus générés par la revente de billets est important.

Les plans factoriels 1-5 et 2-5 (figure 28 en annexes) permet quant à lui de conclure qu'il n'y a pas ou peu de corrélation entre la variable mesurant le pourcentage de points de vente revendeurs ayant vendu au moins un billet sur la période observée et les autres variables actives.

2. Projection des individus sur les plans factoriels

Une fois l'analyse des variables effectuée, on peut procéder à l'analyse de la projection des individus actifs sur les différents plans factoriels retenus. Afin de simplifier la lecture, seule l'interprétation sera détaillée dans cette partie de l'étude, l'intégralité des figures se trouvant en annexes.

Les projections sur les plans factoriels 1-2, 1-3, 1-4 et 1-5 sont fortement marquées par l'opposition entre l'individu actif « Paris » et le reste des individus retenus. En effet, comme précisé précédemment dans la partie consacrée à l'interprétation des plans factoriels et des contributions des différentes variables, ce premier groupe de plans factoriels tient fortement compte de la taille démographique de la ville, notamment au travers du nombre d'habitants.

Pour les autres projections, l'analyse des contributions des individus dans la partie dédiée à l'interprétation des plans factoriels permet de comprendre le positionnement et les coordonnées des individus actifs projetés.

E. Conclusion de l'analyse en composante principale

A partir de l'étude des corrélations, de l'interprétation des différents plans factoriels et de la projection des individus actifs, on procède à plusieurs conclusions.

Tandis que le salaire moyen par catégorie ne semble pas être corrélé à la taille des individus actifs, c'est-à-dire au nombre d'habitants, le nombre de points de vente revendeurs et le nombre de connexions quant à eux sont directement liés linéairement à ces variables démographiques. De ce fait, la société d'autocar a probablement davantage ciblé les villes fortement peuplées ou possédant une population étudiante importante dans sa stratégie d'acquisition de points de vente revendeurs.

Toutefois, on note également une quasi-absence de liaison linéaire entre d'une part le choix du canal d'achat, et d'autre part la taille de la population et le salaire moyen. Aussi, la société d'autocars ne doit probablement pas tenir compte à l'avenir des données socioprofessionnelles et démographiques dans sa stratégie de croissance de l'un de ces trois canaux.

De même, le pourcentage de point de ventes revendeurs ayant réalisés au moins une vente sur la période observée semble indépendante des autres variables. Aussi, cet indicateur, empiriquement très utilisé en interne par les différentes équipes de la compagnie d'autocars, semble ne pas être pertinent dans l'amélioration du ciblage de la clientèle et doit probablement être abandonné au profit d'autres indicateurs.

En revanche, l'évolution positive ou négative des revenus générés en pourcentage par le site internet semble entraîner mécaniquement une évolution inverse des revenus générés par l'application mobile, ce qui semble ne pas être le cas pour les revenus générés par la revente de billets. Néanmoins ces revenus apparaissent comme sensiblement corrélés à la géolocalisation au travers des variables de latitude et de longitude. Cette tendance laisse penser que le comportement d'achat des français dans le Sud et l'Est de la France métropolitaine est davantage tourné vers l'achat en boutique auprès des revendeurs que dans le reste du pays (au Nord et à l'Ouest).

III. Classification avec la méthode hiérarchique de Ward

A. Objectif de la classification

Afin de partitionner les 51 individus actifs retenus dans cette étude, on souhaite procéder à une classification ascendante hiérarchique sous SPAD. Cette méthode semble être la plus appropriée compte tenu du fait que l'échantillon de communes retenues ne comporte que peu d'individus. Toutefois, dans le cas où l'on souhaiterait partitionner l'intégralité des 36 000 communes françaises, il aurait sans doute été bon de choisir une autre méthode telle que les nuées dynamiques ou les centres mobiles. Afin d'avoir une meilleure idée du nombre de classes à sélectionner, on paramètre le nombre de classes par partition de la classification CAH entre cinq et sept classes.

B. Analyse des résultats de la classification

L'analyse des résultats consiste à étudier le diagramme des indices de niveaux, qui représente graphiquement la perte de l'énergie interclasse à chaque étape de regroupement de nœuds, ainsi que le dendrogramme.

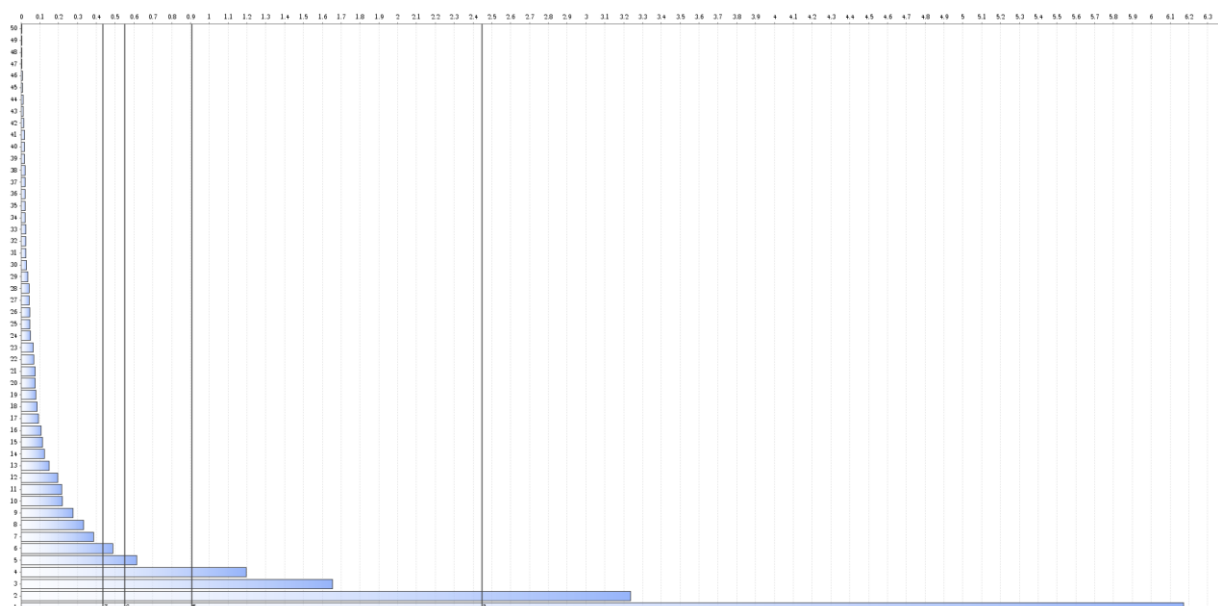


Figure 22 – Indices de niveau & perte d'inertie

A partir de la méthode empirique du coude précédemment citée, et les simulations SPAD du nombre de classes par partition² (5, 6, 7), on constate que le meilleur choix du nombre de classes semble être de cinq compte tenu de la forte perte d'inertie observée. Du fait de la méthode de classification employée, l'inertie interclasse est maximisée puis consolidée.

Nom	Avant consolidation	Après consolidation
Inertie intra-classes	4,396	4,164
Inertie inter-classes	12,258	12,491
Inertie expliquée (%)	73,604	74,998
Critère de Calinski-Harabasz (pseudo F)	32,068	34,496
Indice de Davies-Bouldin	0,947	0,865

Figure 23 – Inertie interclasses et intraclasses

² Le nombre de classes est indiqué par une droite verticale en cinq, six et sept classes.

L'étude de la classification permet ainsi d'obtenir un dendrogramme listant l'ensemble des individus actifs regroupés au sein des cinq classes obtenues.

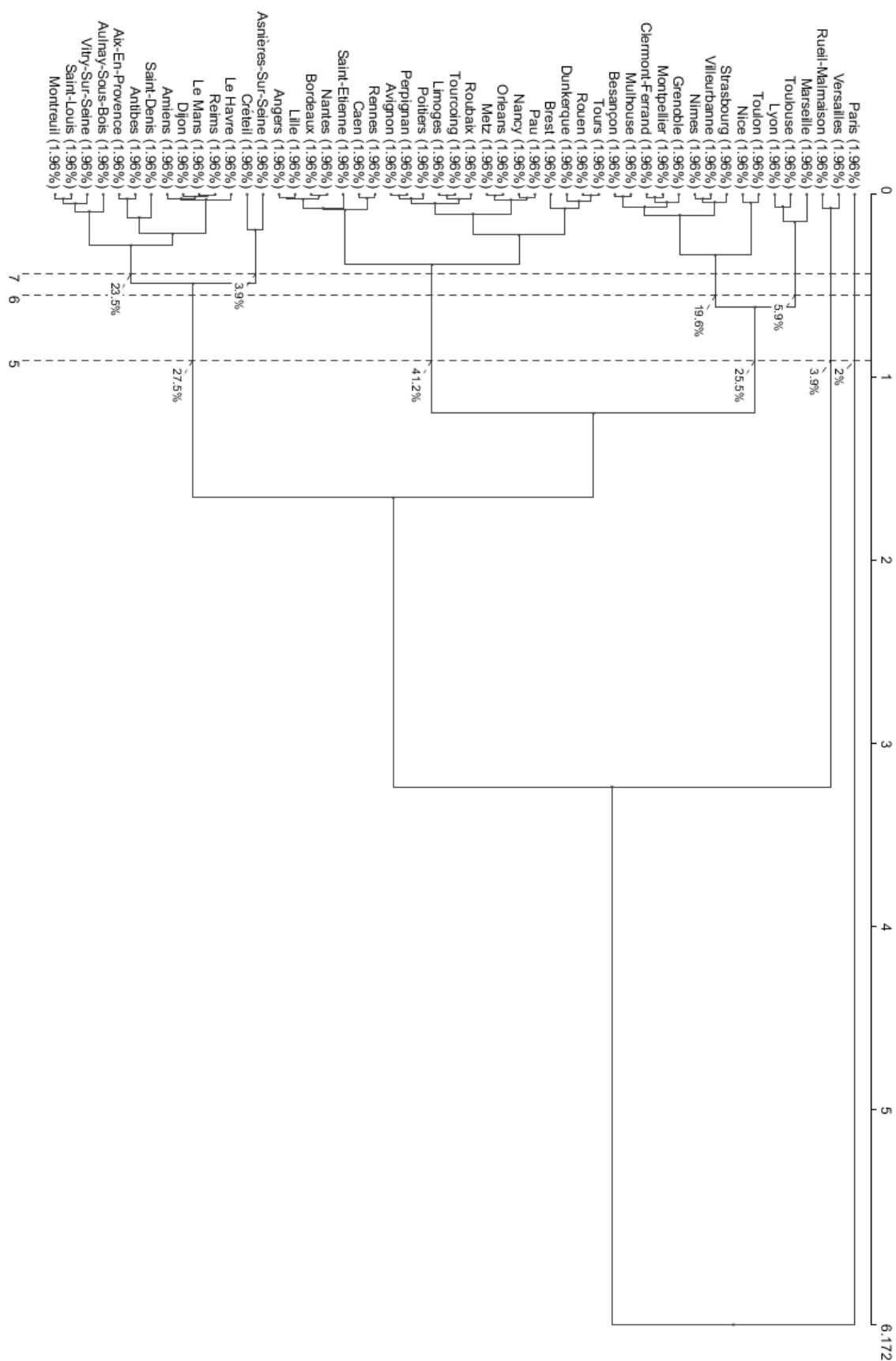


Figure 24 – Dendrogramme des individus actifs

C. Composition des classes

Une fois le nombre de classes déterminé, on peut étudier la composition des cinq classes obtenues.

Classe	Avant consolidation			Après consolidation		
	Effectif	Pourcentage	Inertie	Effectif	Pourcentage	Inertie
1	14	27,451	1,563	15	29,412	1,579
2	21	41,176	1,232	23	45,098	1,389
3	13	25,490	1,527	10	19,608	1,121
4	2	3,922	0,074	2	3,922	0,074
5	1	1,961	0,000	1	1,961	0,000
Ensemble	51	100,000	4,396	51	100,000	4,164

Figure 25 – Informations sur les classes

La classe 1 regroupe les villes suivantes : Asnières-Sur-Seine, Créteil, Le Havre, Reims, Le Mans, Dijon, Amiens, Saint-Denis, Antibes, Aix-En-Provence, Aulnay-Sous-Bois, Vitry-Sur-Seine, Saint-Louis et Montreuil. Ces villes possèdent un salaire horaire net moyen plutôt élevé et sont davantage situées dans la moitié nord de la France métropolitaine et possèdent un réseau de revendeurs de billets très bien localisé, avec une partie importante de ces points de vente se trouvant à proximité des arrêts d'autocar (moins de deux cents mètres). Les clients de la société d'autocar inclus au sein de ces individus actifs ont plutôt tendance à se tourner l'achat de billet d'autocar par le biais de l'application mobile, au détriment du site internet.

La classe 2 regroupe les individus suivants : Angers, Lille, Bordeaux, Nantes, Saint-Etienne, Caen, Rennes, Avignon, Perpignan, Poitiers, Limoges, Tourcoing, Roubaix, Metz, Orléans, Nancy, Pau, Brest, Dunkerque, Rouen et Tours. Les clients de la société d'autocar habitant au sein de ces villes ont plutôt tendance à procéder à leurs achats par le biais du site internet, délaissant ainsi davantage l'application mobile et l'achat auprès d'un revendeur. Ces villes, situées essentiellement au Nord et à l'Ouest de la France métropolitaine, possèdent pour les plus peuplées une gare routière.

La classe 3 regroupe les individus suivants : Besançon, Mulhouse, Clermont-Ferrand, Montpellier, Grenoble, Nîmes, Villeurbanne, Strasbourg, Nice, Toulon, Lyon, Toulouse et Marseille. Ces villes sont localisées principalement au Sud et à l'Est de la France métropolitaine et possèdent bien souvent une gare routière compte tenu de leur taille importante. Le comportement d'achat au sein de ces villes est quant à lui davantage orienté vers l'achat auprès d'un revendeur.

La classe 4 regroupe les villes de Versailles et de Rueil-Malmaison. Ces deux villes de région parisienne possèdent les salaires moyens les plus élevés de l'échantillon, et ont tendance à préférer l'achat de billets d'autocar par le biais du site internet malgré une distance moyenne plutôt faible entre arrêts et revendeurs.

La classe 5 ne comporte qu'un seul individu, la ville de Paris, du fait de ses caractéristiques démographiques et socioprofessionnelles atypiques, bénéficiant d'une place prépondérante au sein du maillage routier français du fait de son positionnement très centralisé, propre à la France métropolitaine.

Annexes

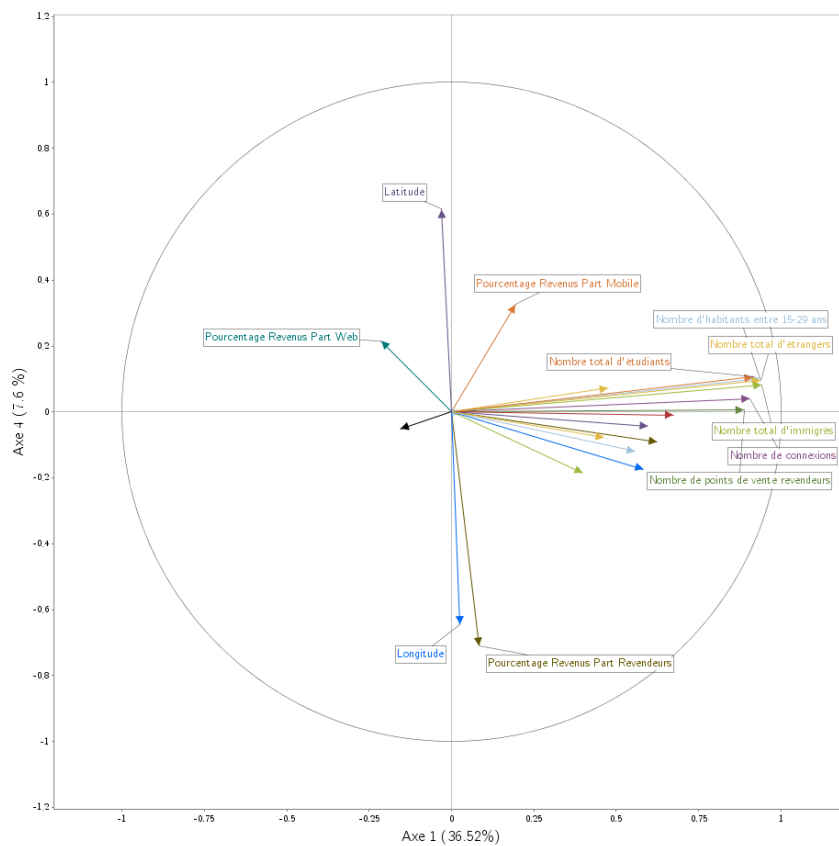


Figure 26 – Cercle des corrélations des variables actives dans le plan factoriel 1-4

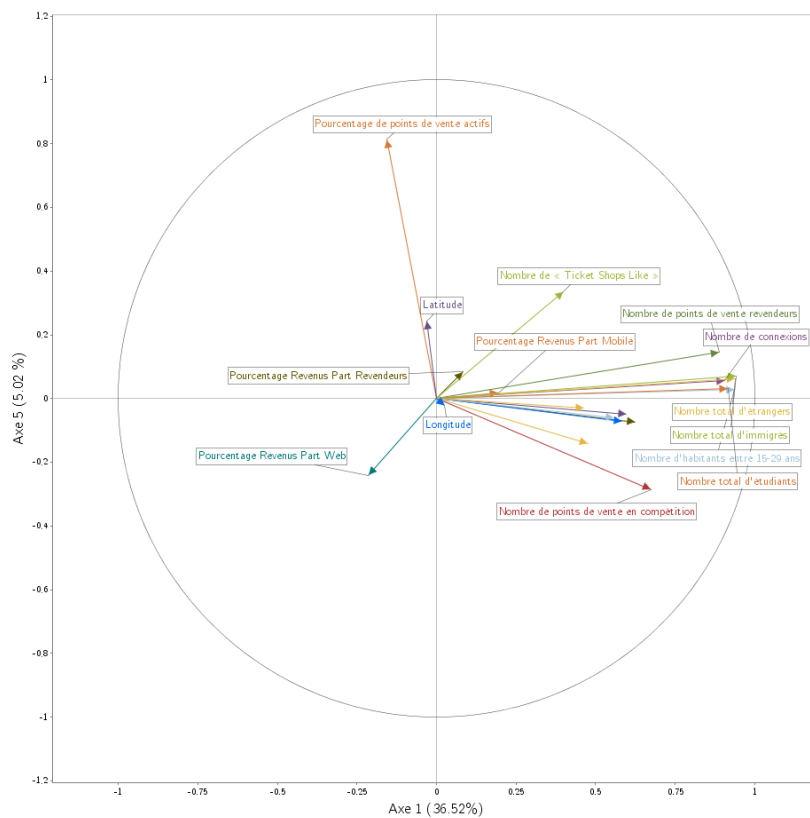


Figure 27 – Cercle des corrélations des variables actives dans le plan factoriel 1-5

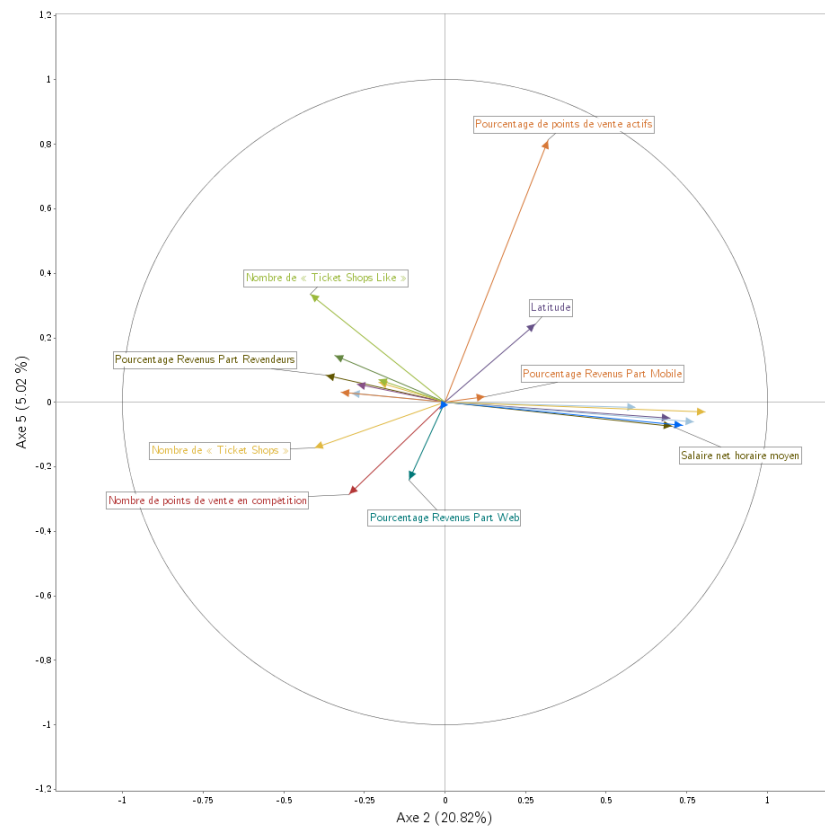


Figure 28 – Cercle des corrélations des variables actives dans le plan factoriel 2-5

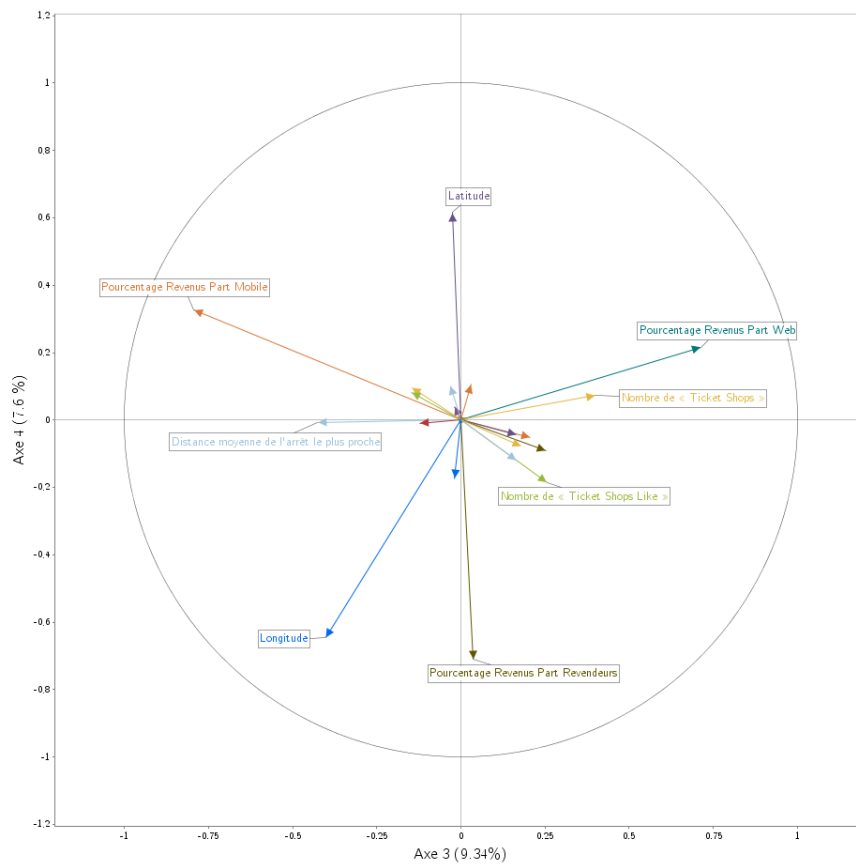


Figure 29 – Cercle des corrélations des variables actives dans le plan factoriel 3-4

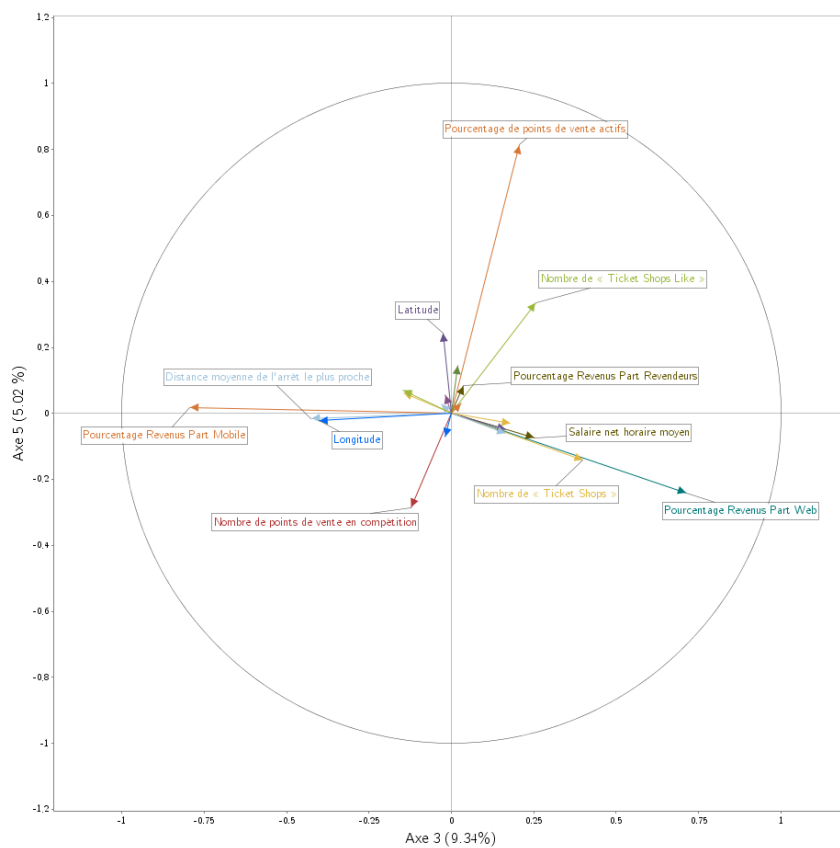


Figure 30 – Cercle des corrélations des variables actives dans le plan factoriel 3-5

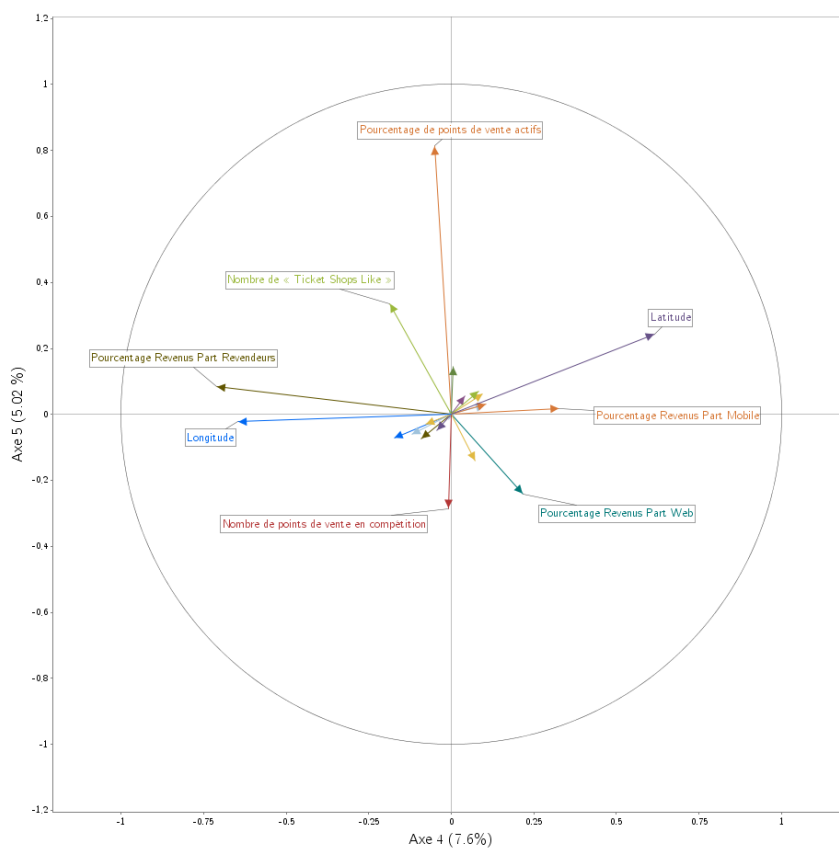


Figure 31 – Cercle des corrélations des variables actives dans le plan factoriel 4-5

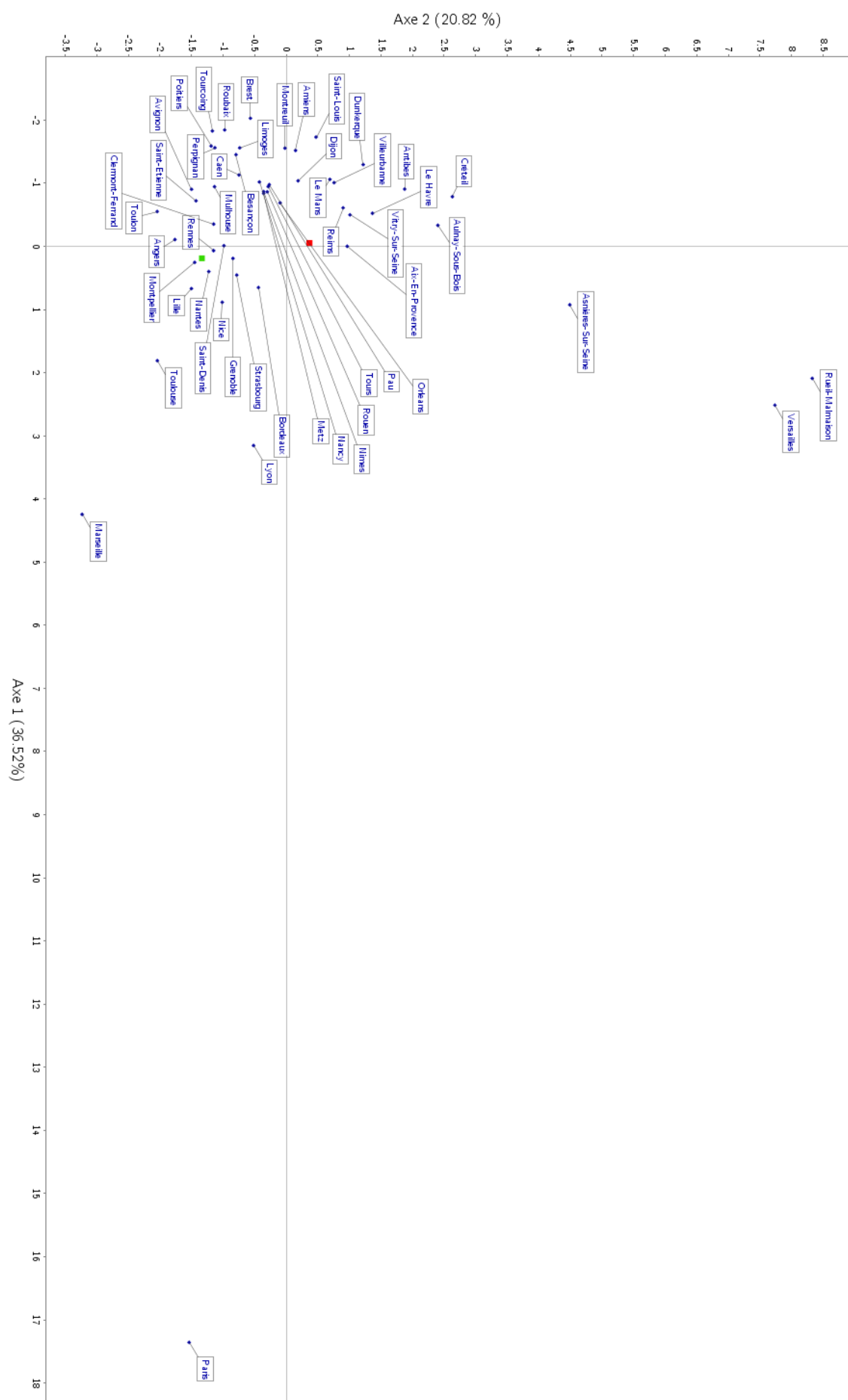
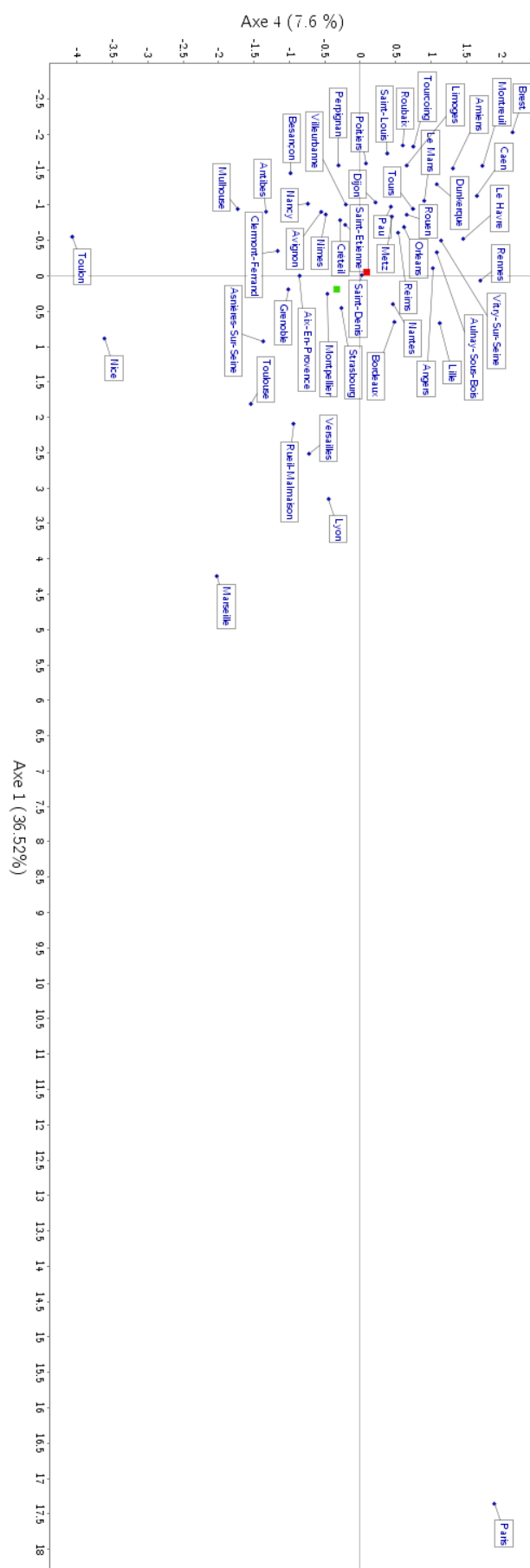


Figure 32 – Projection des individus actifs sur le plan factoriel 1-2



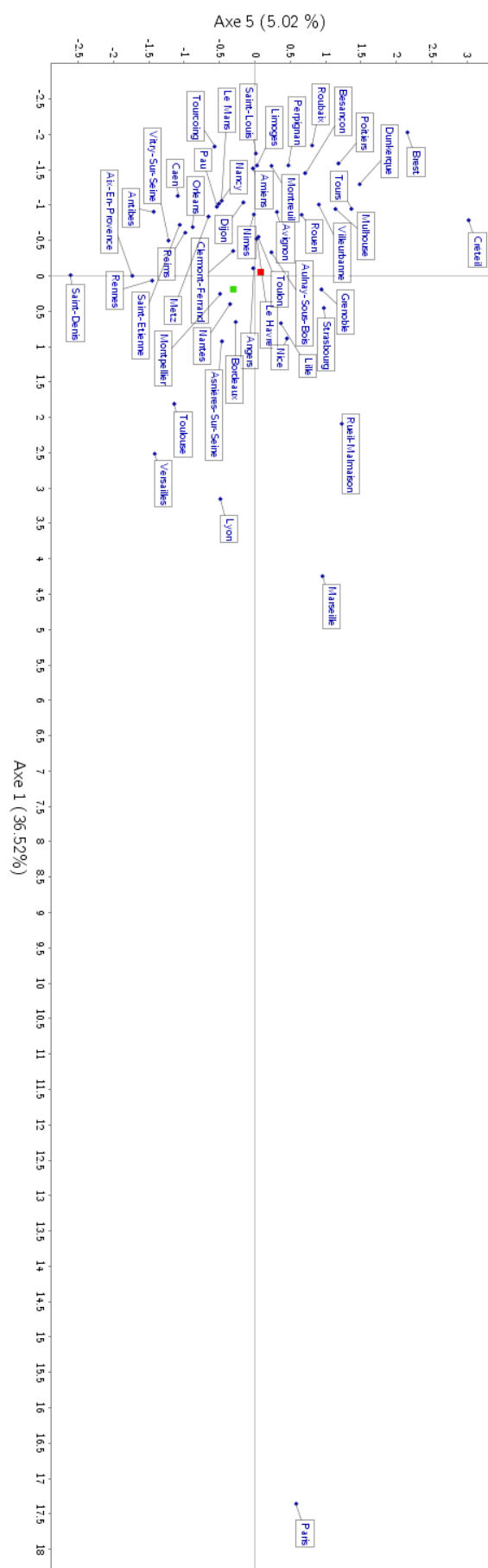


Figure 35 – Projection des individus actifs sur le plan factoriel 1-5

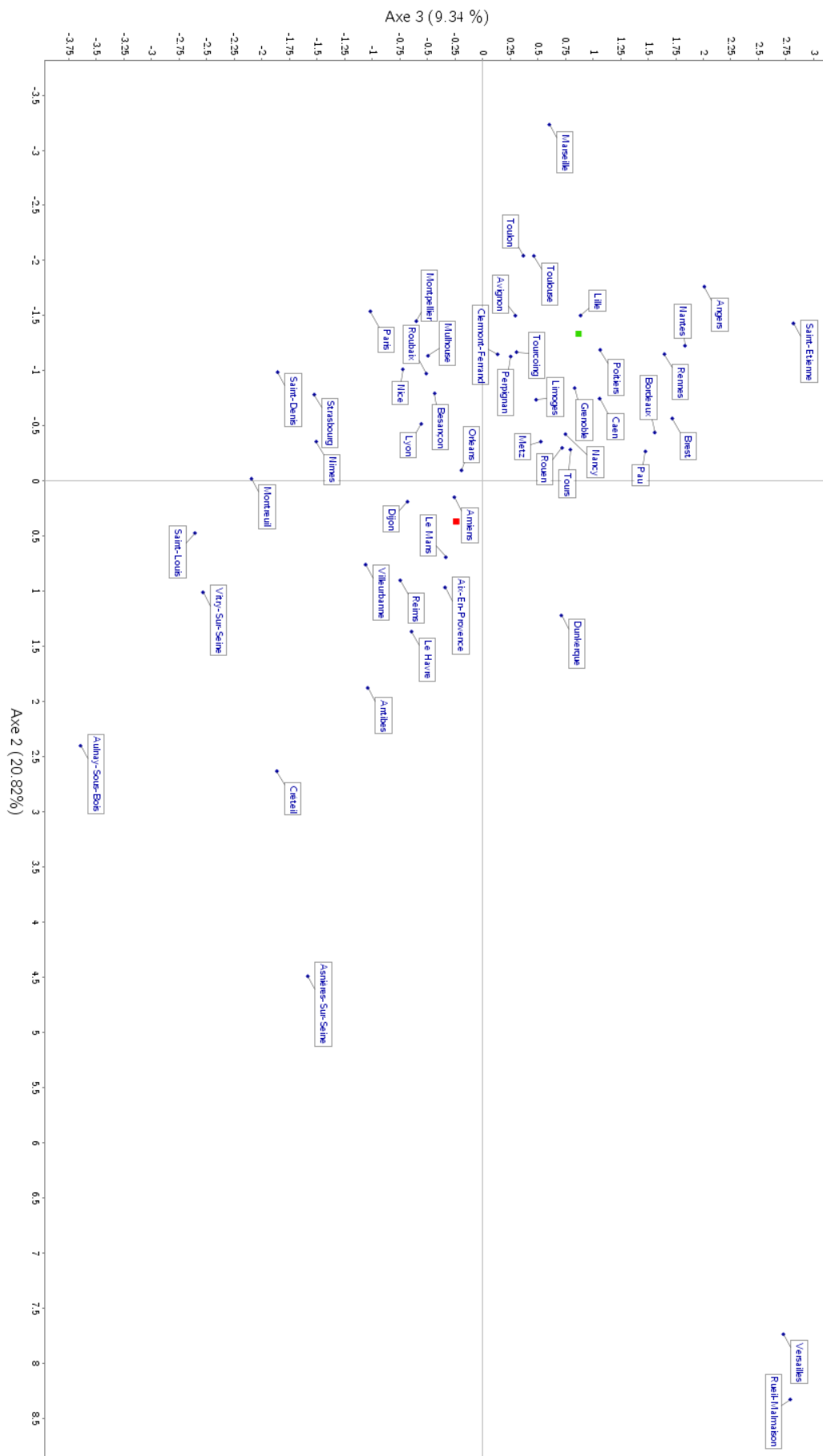


Figure 36 – Projection des individus actifs sur le plan factoriel 2-3

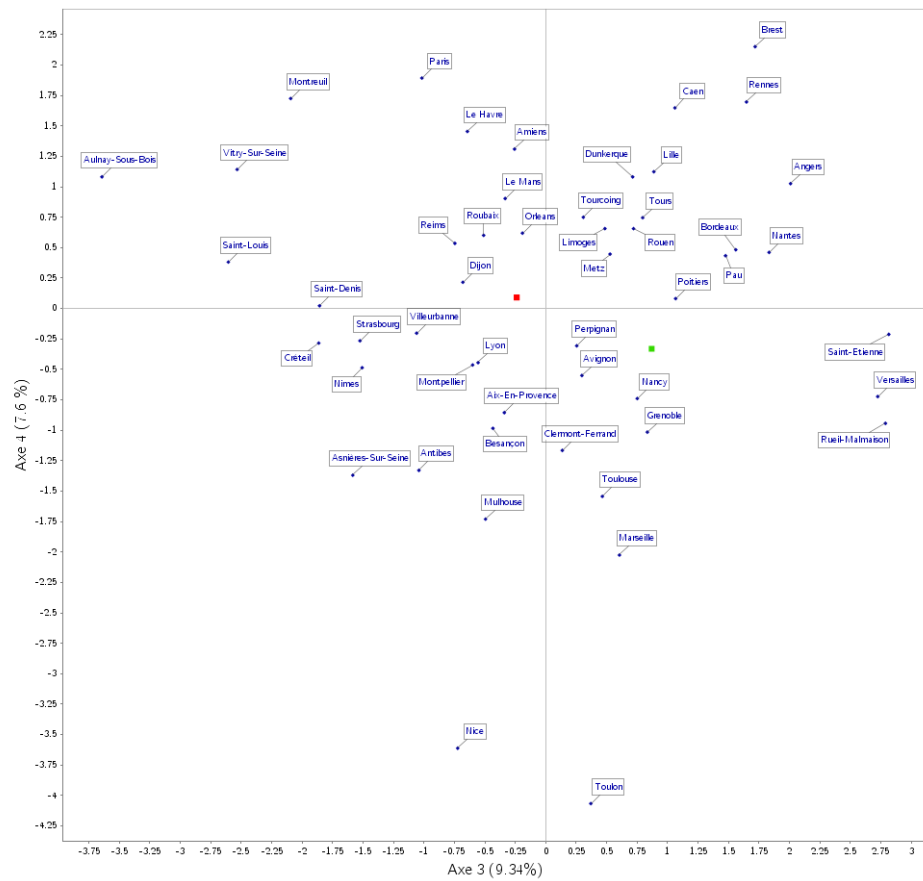


Figure 38 – Projection des individus actifs sur le plan factoriel 3-4

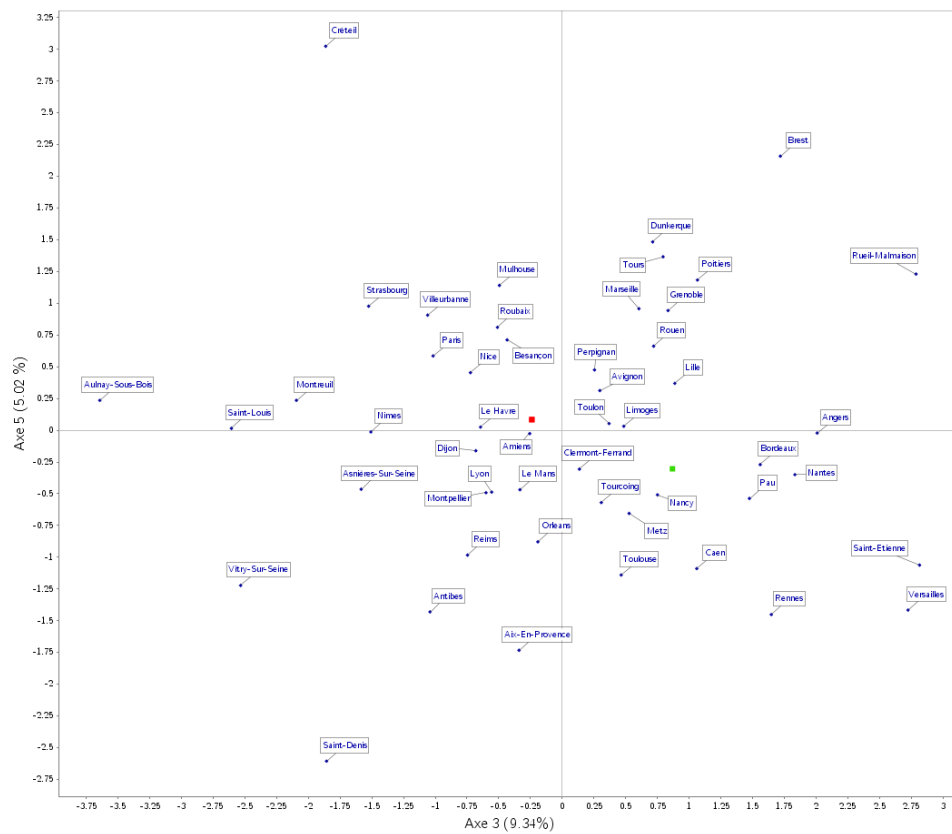


Figure 40 – Projection des individus actifs sur le plan factoriel 3-5

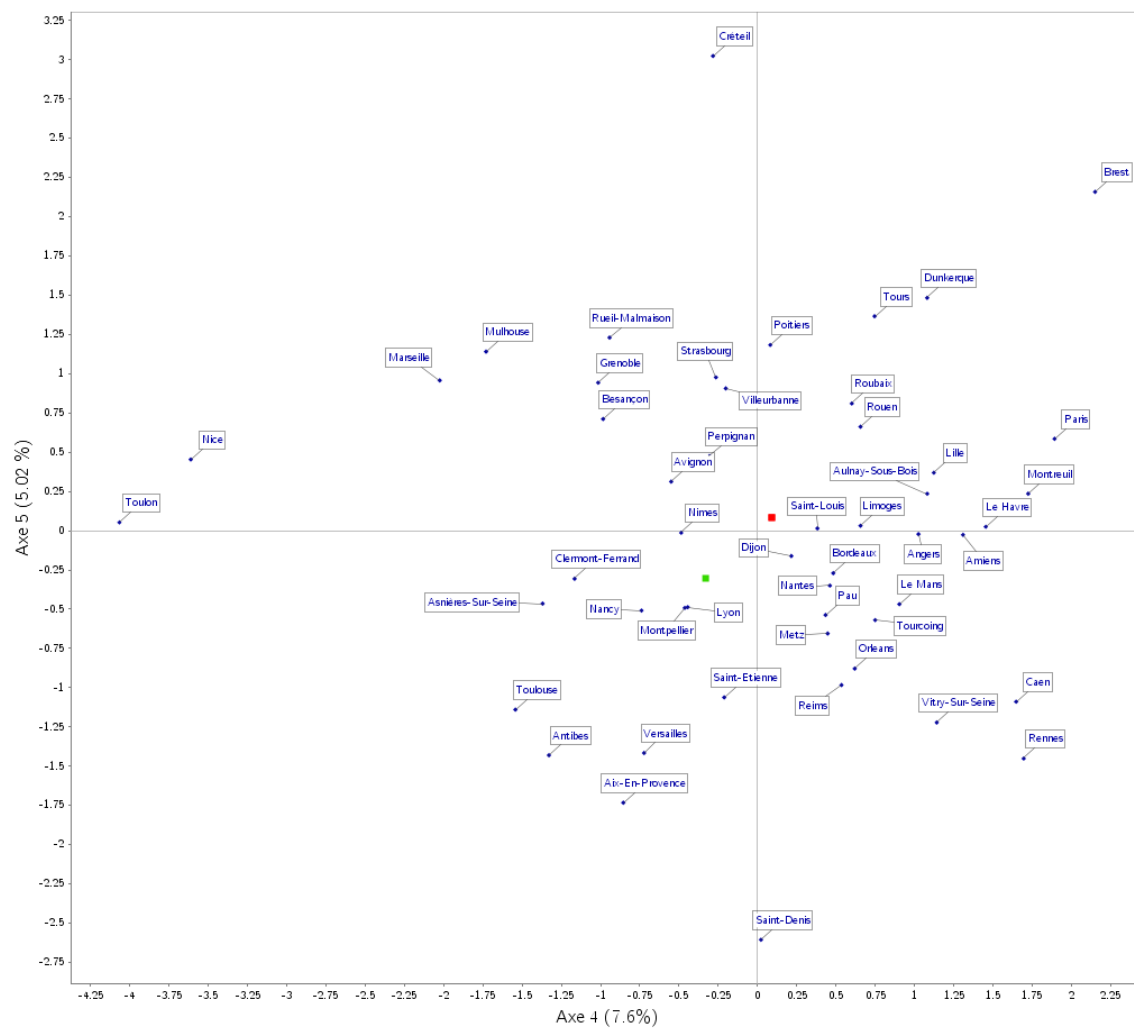


Figure 41 – Projection des individus actifs sur le plan factoriel 4-5