

**UNIVERSITE NATIONALE DU VIETNAM (UNV)**  
**/ Institut Francophone pour l'Innovation (IFI)**

**Option : Système Intelligent et Multimédia**

**Niveau : Master I**

## **Cours : Fouille de Données**

Professeur : NGUYEN Thà Minh Huyen

Titre du devoir

**Rapport du TP2 & TP3 du cours de Fouille de Données**

- **Analyse Factorielle**
- **Clustering**

**Préparé par le Binôme 13 :**

**Mongetro GOINT & Myderson SEMEURAND**

**Date: 19 Mars 2018**

## Table des matières

Tableaux et figures.....	2
Analyses Factorielles .....	4
1.- Application des méthodes factorielles à notre jeu de données “Credit Scoring” .....	4
1.1.- ACP – Analyses en Composante Principale :.....	4
2.- Analyse Factorielle des Correspondances .....	8
2.1.- Les résultats numérique de l’Analyse Factorielle des Correspondances (AFC) .....	8
2.2.- Représentation graphique de l’Analyse Factorielle des Correspondances (AFC) .....	9
Clustering.....	11
3.- Classifications.....	11
3.1.- Répartition des individus .....	11
3.2.- Part d’explication de la partition dans l’inertie totale.....	11
3.4.- Groupe de caractérisation .....	12
3.5.- La projection des individus suivant leur cluster dans le premier plan factoriel .....	12
3.6.- HAC.....	13
Conclusion .....	<b>Error! Bookmark not defined.</b>

## Tableaux et figures

Figure 1: Importance des composantes principales .....	5
Figure 2 : Corrélation entres les variables et les axes principaux.....	5
Figure 3 : Plan factoriel sur le premier axe .....	6
Figure 4 : Plan factoriel du premier axe modifié .....	7
Figure 5 : Cercle des corrélations entre les variables dans le premier plan factoriel.....	7
Figure 6 : Test CHI 2 d’indépendance .....	8
Figure 7 : Valeurs propres calculées .....	8
Figure 8 : AFC dans le premier plan factoriel .....	9
Figure 9 : Coordonnées factorielles des profils lignes.....	10
Figure 10 : Coordonnées factorielles des profils colonnes. ....	10
Figure 11 : Répartition des individus dans les clusters.....	11
Figure 12 : Part d’explication de la partition dans l’inertie totale. ....	11
Figure 13 : Coordonnées des centres de chacun des cinq clusters.....	12
Figure 14 : Caractérisation des clusters .....	12
Figure 15 : Projection des individus par cluster.....	13
Figure 16 : Sélection du nombre optimal de clusters.....	13

Figure 17 : Résultat du Clustering .....	14
Figure 18 : Coordonnées des centres des clusters.....	14

## Analyses Factorielles

### 1.- Application des méthodes factorielles à notre jeu de données “Credit Scoring”

#### 1.1.- ACP – Analyses en Composante Principale :

Notre objectif en réalisant cette analyse en composante principale est de considérer l'ensemble des données contenant 24 observations et 5 variables continues afin de résumer l'information disponible à l'aide de variables synthétiques appelées composantes principales. Au moyen de la fonction « Principal Analysis » située sous l'onglet « Factoriel Analysis », nous réalisons l'analyse en composante principale. Ci-dessous, les résultats obtenus après cette opération :

#### Eigen values

Matrix trace	5.000000
Average	1.000000






Axis	Eigen value	Difference	Proportion (%)	Histogram	Cumulative (%)
1	1.699882	0.614191	34.00 %		34.00 %
2	1.085691	0.094146	21.71 %		55.71 %
3	0.991545	0.163913	19.83 %		75.54 %
4	0.827632	0.432383	16.55 %		92.10 %
5	0.395249	-	7.90 %		100.00 %
Tot.	5.000000	-	-	-	-

Tableau 1 :

Valeurs propres

##### 1.1.1.- Valeurs propres

D'après le tableau ci-dessous ([voir le tableau 1](#)), l'inertie décrite par le premier axe principal est  $\lambda_1=1.70$  et la part de celle décrite par ce même axe principal est  $\lambda_1/p=0.34$ . Autrement dire, la première composante principale décrite à elle seule 34% (soit un peu moins de la moitié) de la variance totale.

En considérant le critère de Kaiser-Guttman proposé et utilisé par notre outil de travail [Tanagra](#) pour la définition des composantes principales les plus significatives, nous observons que deux des cinq valeurs propres sont importantes et pour la suite, ce sont les seules qui seront retenues. ([voir la figure 1](#))

## Significance of Principal Components

Global critical values	
Kaiser-Guttman	1
Kartis-Saporta-Spinaki	1.83406

### Eigenvalue table - Test for significance

Eigenvalues - Significance		
Axis	Eigenvalue	Broken-stick critical values
1	1.699882	2.283333
2	1.085691	1.283333
3	0.991545	0.783333
4	0.827632	0.450000
5	0.395249	0.200000

Figure 1: Importance des composantes principales

### 1.1.2.- Corrélation entre les variables et les axes principaux

Cette partie des résultats ([voir la figure 2](#)) indique la corrélation entre les variables et les axes principaux.

#### Factor Loadings [Communality Estimates]

Attribute	Axis_1		Axis_2		Axis_3		Axis_4		Axis_5	
	Corr.	% (Tot. %)	Corr.	% (Tot. %)	Corr.	% (Tot. %)	Corr.	% (Tot. %)	Corr.	% (Tot. %)
Durée(mois)	0.83120	69 % (69 %)	0.21891	5 % (74 %)	-0.07946	1 % (75 %)	-0.27751	8 % (82 %)	-0.42174	18 % (100 %)
Montant	-0.77978	61 % (61 %)	0.34444	12 % (73 %)	0.17134	3 % (76 %)	0.29339	9 % (84 %)	-0.39732	16 % (100 %)
Age	0.56774	32 % (32 %)	-0.18003	3 % (35 %)	0.02836	0 % (36 %)	0.80229	64 % (100 %)	-0.02810	0 % (100 %)
Solde	-0.26699	7 % (7 %)	-0.84223	71 % (78 %)	-0.40205	16 % (94 %)	-0.05733	0 % (95 %)	-0.23329	5 % (100 %)
Nombres de prêts antérieurs dans cette banque	-0.08555	1 % (1 %)	0.42115	18 % (18 %)	-0.89074	79 % (98 %)	0.13261	2 % (100 %)	0.06560	0 % (100 %)
Var. Expl.	1.69988	34 % (34 %)	1.08569	22 % (56 %)	0.99155	20 % (76 %)	0.82763	17 % (92 %)	0.39525	8 % (100 %)

Figure 2 : Corrélation entre les variables et les axes principaux

A remarquer qu'ici, le premier axe est fortement corrélé positivement avec les variables : Durée(mois) et Age, et, négativement avec la variable : Montant. Ce qui indique qu'il prend en compte une bonne partie des données quantitatives de notre « Credit Scoring ». Mais cependant, l'évolution de la variable associée à cet axe est opposée à celle de la variable : Montant.

Le premier axe ainsi observé est déterminant dans l'identification d'un dossier « bon » ou « mauvais » concernant la fidélité des clients chez l'établissement de crédit.

Les trois autres axes suivants par contre, sont peu corrélé et le cinquième est quant à lui faiblement corrélé.

### 1.1.3.- Plans factoriels

La popularité de l'ACP est en majeure partie liée à sa capacité de fournir des représentations graphiques, permettant d'apprécier par visualisation, les proximités entre les observations.

Dans notre cas, nous projetons les observations dans le premier plan factoriel où nous visons d'associer les identifiants aux points. Nous utilisons pour cela le composant SCATTERPLOT

WITH LABEL (onglet DATA VISUALIZATION) que nous plaçons en dessous de l'ACP. Nous le paramétrons de manière à avoir en abscisse le premier facteur, en ordonnée le second facteur.

Notons qu'il est très aisé de passer d'un plan factoriel à un autre.

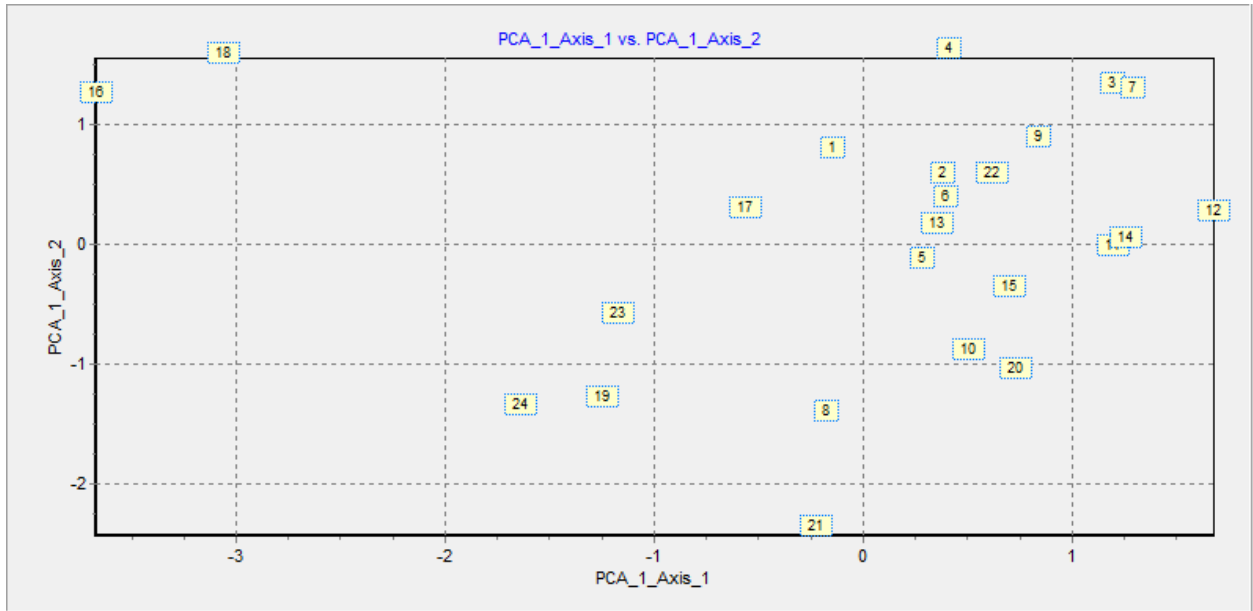


Figure 3 : Plan factoriel sur le premier axe

En activant l'option LEGEND / ATTRIBUTE VALUE, et en choisissant la référence « Dossier Prêt », nous obtenons la carte des points étiquetés par leurs identifiants. À noter que cette option est pratique que si les observations sont peu nombreuses. Mais, à mesure que les observations augmentent, le risque d'avoir un nombre de point illisible devient évidente.

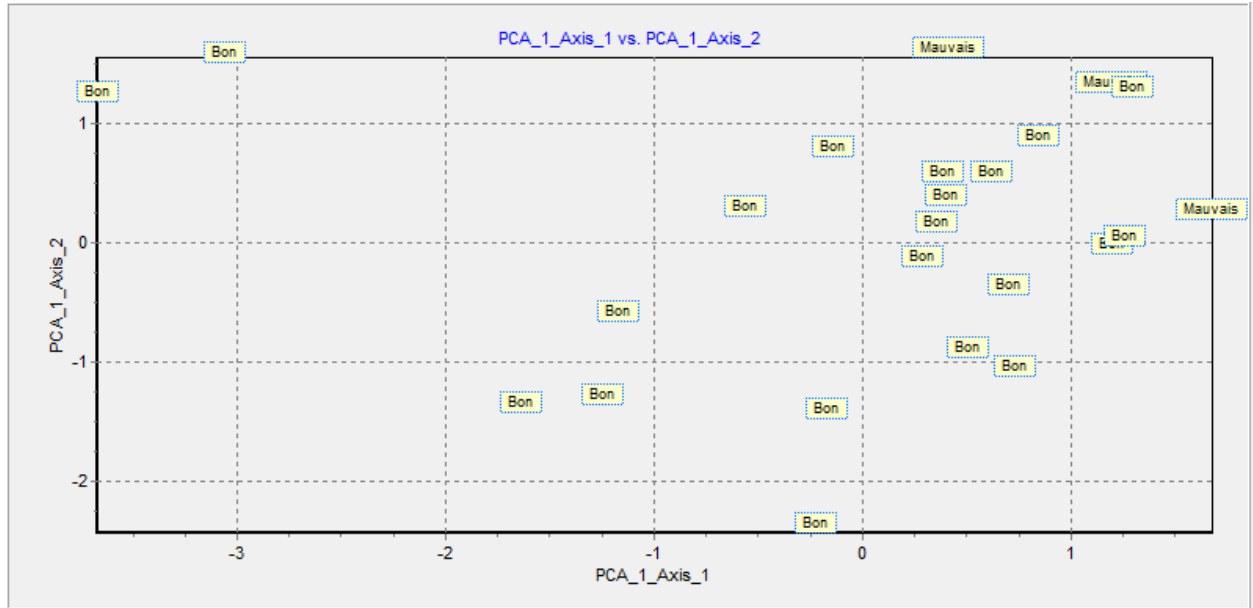


Figure 4 : Plan factoriel du premier axe modifié

#### 1.1.4.- Cercle des corrélations

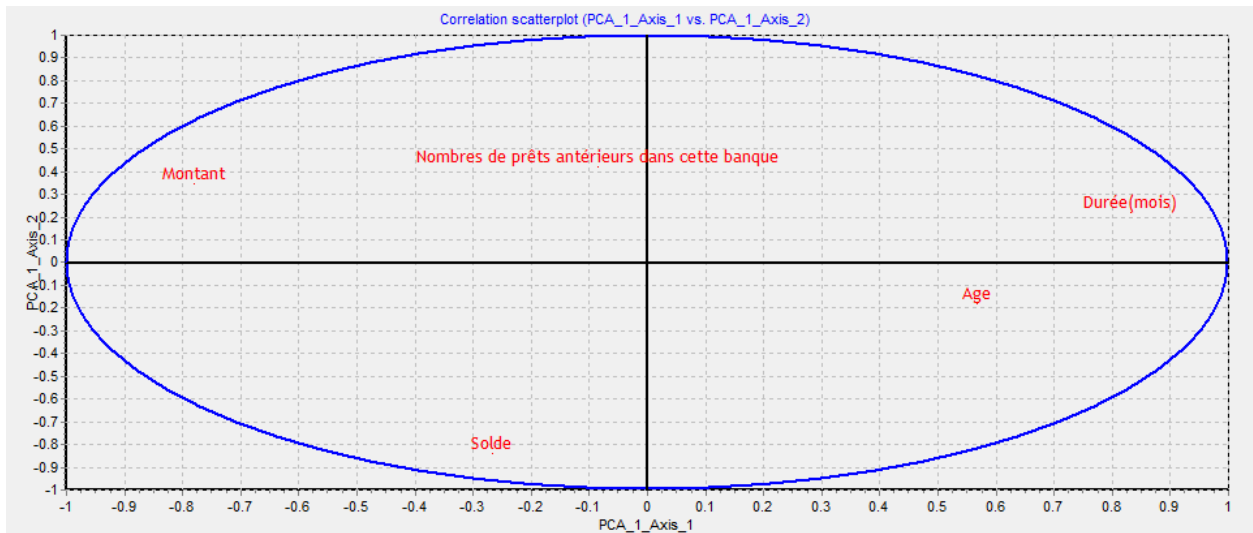


Figure 5 : Cercle des corrélations entre les variables dans le premier plan factoriel

Sur le cercle des corrélations ci-dessus, nous pouvons constater des variables comme (Durée et Age) sont très proches du cercle et du seconde axe (position des valeurs négatives), ce qu'en quelque sorte confirme leur forte corrélation positive avec la seconde composante principale. En opposition évidemment à la variable (Montant).

Quant à elle la variable (Nombre de prêts antérieurs dans cette banque), elle est plus proche du centre cercle du cercle et ne donne aucune information sur sa corrélation avec les deux premières composantes

principales. Le cas de notre variable (Solde) situé proche du centre de notre cercle mais ne présentant aucune information sur sa corrélation avec les deux premières composantes principales.

## 2.- Analyse Factorielle des Correspondances

D'abord, rappelons qu'au niveau de la première partie de ce travail, nous avons réalisé une analyse factorielle des correspondances entre les attributs qualitatifs de notre « CreditScoring ». Ces relations entre les paires de variables qualitatives sont représentées dans un tableau de contingence répartis en plusieurs figures (soit Figure 15 à 27).

### 2.1.- Les résultats numérique de l'Analyse Factorielle des Correspondances (AFC)

Dans cette partie nous présentons les résultats numériques de l'ACP à l'aide du Test de CHI 2 et des variables propres, en nous basant sur les variables discrètes surtout.

#### 2.1.1.- Le Test de CHI 2

Avec les 14 variables qualitatives de notre « CreditScoring » les Test de CHI 2 sont nombreux, par contre ici nous ne présentons que l'un d'entre-deux.

Stat	Value
d.f.	2
Tschuprow's t	0.342125
Cramer's v	0.406857
Phi <sup>2</sup>	0.165533
Chi <sup>2</sup> (p-value)	3.97 (0.1372)

Figure 6 : Test CHI 2 d'indépendance

#### 2.1.2.- Les valeurs propres

##### Eigen values

Matrix trace = 0.2634

SQRT(Matrix trace) = 0.5132




Axis	Eigen value	% explained	Histogram	% cumulated
1	0.236805	89.90%		89.90%
2	0.018498	7.02%		96.92%
3	0.005824	2.21%		99.14%
4	0.002278	0.86%		100.00%
Tot.	0.263405	-	-	-

Figure 7 : Valeurs propres calculées



## 2.2.- Représentation graphique de l'Analyse Factorielle des Correspondances (AFC)

Nous ne devons pas oublier que notre outil Data Mining de travail « Tanagra » offre une fenêtre de visualisation où nous pouvons observer les graphiques suivant les différents plans factoriels à partir de l'onglet CHART ([voir la figure 8](#)).

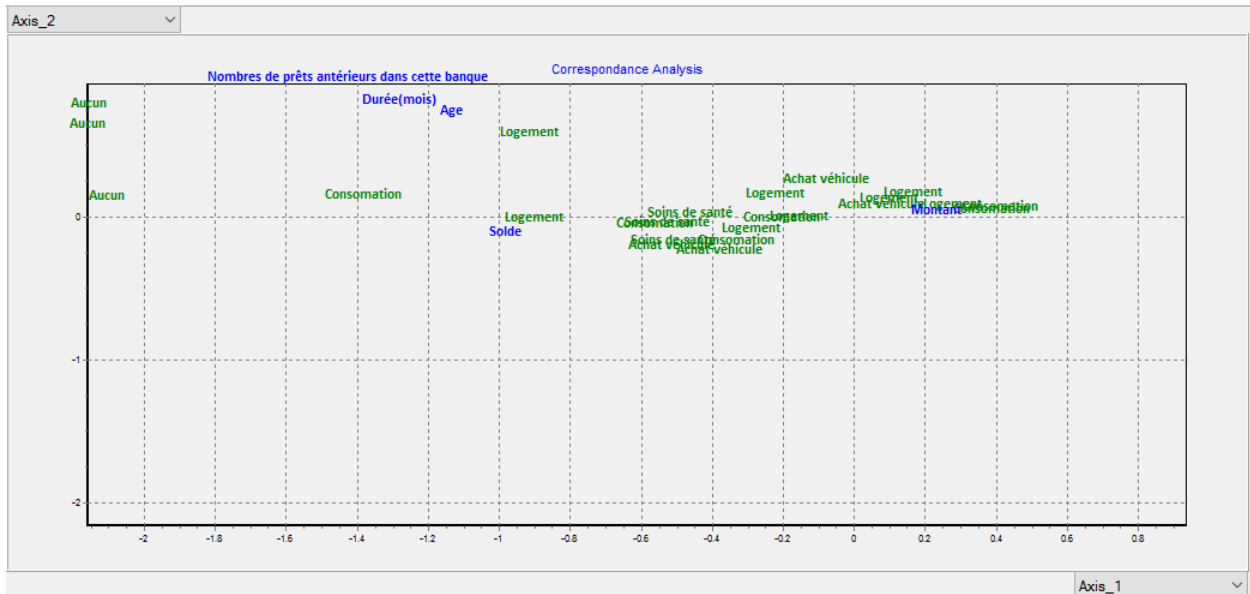


Figure 8 : AFC dans le premier plan factoriel

### 2.2.3.- Coordonnées factorielles et interprétations des axes par les points individus :

Les points individus dans notre cas sont les différents Objectif de prêt (Aucun, Consommation, Logement, Achat de Véhicules et Soins de santé).

#### ✓ Coordonnées factorielles

Le tableau de la figure suivantes est une démonstration du résultat des coordonnées factorielles obtenues pour chaque point individu ([voir la figure 9](#)).

#### **Interprétation :**

L'axe 1 présente les crédits à la consommation comme étant dominante tandis que l'axe 2 met accent sur les crédits pour achat de véhicules. Au niveau des axes 3 et 4 la remarque à faire n'est autre que quasiment le même degré d'importance de tous les points individus.

## Rapport TP2 et TP3 – Analyses Factorielles et Clustering

### Rows analysis

Characterization				Coord.				Contributions (%)				COS²			
Values	Weight	Sq. Dist.	Inertia	coord 1	coord 2	coord 3	coord 4	ctr 1	ctr 2	ctr 3	ctr 4	cos² 1	cos² 2	cos² 3	cos² 4
Aucun	0.00820	4.44741	0.03648	-2.10528	0.10052	-0.01958	-0.06868	15.35	0.45	0.05	1.70	1.00 (1.00)	0.00 (1.00)	0.00 (1.00)	0.00 (1.00)
Consomation	0.19510	0.16926	0.03302	0.41053	0.02509	-0.00619	0.00783	13.89	0.66	0.13	0.52	1.00 (1.00)	0.00 (1.00)	0.00 (1.00)	0.00 (1.00)
Consomation	0.19696	0.15083	0.02971	0.38795	0.00903	-0.00905	0.01275	12.52	0.09	0.28	1.41	1.00 (1.00)	0.00 (1.00)	0.00 (1.00)	0.00 (1.00)
Aucun	0.00558	5.28347	0.02950	-2.15480	0.75116	-0.02493	0.27466	10.95	17.03	0.06	18.49	0.88 (0.88)	0.11 (0.99)	0.00 (0.99)	0.01 (1.00)
Aucun	0.00396	5.47601	0.02167	-2.15844	0.60404	-0.66234	0.11655	7.79	7.81	29.81	2.36	0.85 (0.85)	0.07 (0.92)	0.08 (1.00)	0.00 (1.00)
Consomation	0.00900	2.06860	0.01862	-1.38061	0.11329	0.38689	0.00236	7.25	0.62	23.14	0.00	0.92 (0.92)	0.01 (0.93)	0.07 (1.00)	0.00 (1.00)
Consomation	0.03314	0.32757	0.01086	-0.56115	-0.09175	0.03785	0.05319	4.41	1.51	0.82	4.12	0.96 (0.96)	0.03 (0.99)	0.00 (0.99)	0.01 (1.00)
Logement	0.00923	1.24328	0.01147	-0.91244	0.54945	0.29681	-0.14402	3.24	15.06	13.96	8.40	0.67 (0.67)	0.24 (0.91)	0.07 (0.98)	0.02 (1.00)
Achat véhicule	0.02785	0.22132	0.00616	-0.37778	-0.27740	0.03959	0.00927	1.68	11.58	0.75	0.10	0.64 (0.64)	0.35 (0.99)	0.01 (1.00)	0.00 (1.00)
Achat véhicule	0.03230	0.33336	0.01077	-0.51263	-0.24316	-0.06796	0.08264	3.58	10.32	2.56	9.68	0.79 (0.79)	0.18 (0.97)	0.01 (0.98)	0.02 (1.00)
Soins de santé	0.03304	0.30917	0.01021	-0.50812	-0.20729	-0.07131	-0.05414	3.60	7.67	2.88	4.25	0.84 (0.84)	0.14 (0.97)	0.02 (0.99)	0.01 (1.00)
Consomation	0.02804	0.16305	0.00457	-0.33338	-0.21271	0.08026	0.01489	1.32	6.86	3.10	0.27	0.68 (0.68)	0.28 (0.96)	0.04 (1.00)	0.00 (1.00)
Achat véhicule	0.02491	0.05952	0.00148	-0.08019	-0.22260	0.02127	0.05557	0.07	6.67	0.19	3.38	0.11 (0.11)	0.83 (0.94)	0.01 (0.95)	0.05 (1.00)
Logement	0.04343	0.11156	0.00485	-0.28888	-0.12846	0.10728	0.01003	1.53	3.87	8.58	0.19	0.75 (0.75)	0.15 (0.90)	0.10 (1.00)	0.00 (1.00)
Logement	0.01164	0.86327	0.01005	-0.89838	-0.04756	-0.14691	-0.17984	3.97	0.14	4.31	16.52	0.93 (0.93)	0.00 (0.94)	0.03 (0.96)	0.04 (1.00)
Consomation	0.02679	0.05208	0.00140	-0.20204	-0.04484	0.03780	0.08844	0.46	0.29	0.66	9.20	0.78 (0.78)	0.04 (0.82)	0.03 (0.85)	0.15 (1.00)
Logement	0.01879	0.06970	0.00131	-0.21993	0.11788	-0.03534	-0.07866	0.38	1.41	0.40	5.11	0.69 (0.69)	0.20 (0.89)	0.02 (0.91)	0.09 (1.00)
Soins de santé	0.02154	0.29498	0.00636	-0.52536	-0.08501	-0.08730	-0.06433	2.51	0.84	2.82	3.91	0.94 (0.94)	0.02 (0.96)	0.03 (0.99)	0.01 (1.00)
Logement	0.04178	0.02720	0.00114	-0.15425	-0.03971	-0.00113	-0.04282	0.42	0.36	0.00	3.36	0.87 (0.87)	0.06 (0.93)	0.00 (0.93)	0.07 (1.00)
Logement	0.03859	0.04576	0.00177	0.16831	0.12564	-0.00642	-0.04011	0.46	3.29	0.03	2.73	0.62 (0.62)	0.34 (0.96)	0.00 (0.96)	0.04 (1.00)
Logement	0.04851	0.02057	0.00100	0.10095	0.08850	0.04038	-0.03034	0.21	2.05	1.36	1.96	0.50 (0.50)	0.38 (0.88)	0.08 (0.96)	0.04 (1.00)
Logement	0.07446	0.08117	0.00604	0.28001	0.04472	-0.01960	-0.01955	2.47	0.80	0.49	1.25	0.97 (0.97)	0.02 (0.99)	0.00 (1.00)	0.00 (1.00)
Soins de santé	0.02061	0.22309	0.00460	-0.46007	-0.01678	-0.10009	-0.03360	1.84	0.03	3.55	1.02	0.95 (0.95)	0.00 (0.95)	0.04 (0.99)	0.01 (1.00)
Achat véhicule	0.04653	0.00814	0.00038	0.07606	0.04725	-0.00977	0.00521	0.11	0.56	0.08	0.06	0.71 (0.71)	0.27 (0.98)	0.01 (1.00)	0.00 (1.00)

Figure 9 : Coordonnées factorielles des profils lignes

### Columns analysis

Characterization				Coord.				Contributions (%)				COS²			
Values	Weight	Sq. Dist.	Inertia	coord 1	coord 2	coord 3	coord 4	ctr 1	ctr 2	ctr 3	ctr 4	cos² 1	cos² 2	cos² 3	cos² 4
Solde	0.15942	0.98394	0.15686	-0.98057	-0.14973	-0.00088	0.00090	64.73	19.32	0.00	0.01	0.98 (0.98)	0.02 (1.00)	0.00 (1.00)	0.00 (1.00)
Durée(mois)	0.01283	2.41507	0.03098	-1.28170	0.77740	-0.36902	-0.17830	8.90	41.91	30.00	17.91	0.68 (0.68)	0.25 (0.93)	0.06 (0.99)	0.01 (1.00)
Age	0.01287	2.02733	0.02609	-1.13344	0.69676	0.50387	0.05737	6.98	33.77	56.10	1.86	0.63 (0.63)	0.24 (0.87)	0.13 (1.00)	0.00 (1.00)
Nombres de prêts antérieurs dans cette banque	0.00104	5.45153	0.00564	-1.42536	0.93444	-0.88403	1.32860	0.89	4.89	13.89	80.23	0.37 (0.37)	0.16 (0.53)	0.14 (0.68)	0.32 (1.00)
Montant	0.81385	0.05386	0.04383	0.23202	0.00487	-0.00085	0.00004	18.50	0.10	0.01	0.00	1.00 (1.00)	0.00 (1.00)	0.00 (1.00)	0.00 (1.00)

Figure 10 : Coordonnées factorielles des profils colonnes.

### Interprétation :

Sur l'axe 1, le solde est le point individu à remarquer tandis que l'axe 2, et 4 sont respectivement dominés par la durée du prêt, le nombre de prêts antérieurs d'un de cette banque.

# Clustering

## 3.- Classifications

Pour illustrer la méthode **K-Means**, nous choisissons de classer les objectifs suivants la similarité de leurs individus (valeurs des variables quantitatives). Pour cela, nous utilisons le composant **K-Means** de l'onglet **Clustering** de Tanagra avec la variable Objectif en Target et les cinq variables continues en Input. Les paramètres utilisés sont :

- 5 Clusters
- 10 Itérations au maximum
- 5 Essais

### 3.1.- Répartition des individus

Nous utilisons tous les variables Continues de notre « CreditScoring ».

#### Cluster size and WSS

Clusters	5		
Cluster	Description	Size	WSS
cluster n°1	c_kmeans_1	7	7.9151
cluster n°2	c_kmeans_2	2	0.5889
cluster n°3	c_kmeans_3	3	8.5441
cluster n°4	c_kmeans_4	9	23.1201
cluster n°5	c_kmeans_5	3	5.5733

Figure 11 : Répartition des individus dans les clusters

Remarque pertinente : Le cluster No 2 est celui contenant le moins d'individu soit 2 au total, tandis que le No 4 est celui du plus grand nombre, soit 9 au total.

### 3.2.- Part d'explication de la partition dans l'inertie totale

D'après la figure suivante, la partition explique 61.88% de l'inertie totale.

#### Global evaluation

Within Sum of Squares	45.7415
Total Sum of Squares	120.0000
R-Square	0.6188

Figure 12 : Part d'explication de la partition dans l'inertie totale.

### 3.3.- Coordonnées des centres de chacun des cinq clusters sont présentés ci-dessous : Cluster centroids

Attribute	Cluster n°1	Cluster n°2	Cluster n°3	Cluster n°4	Cluster n°5
Solde	368.285714	422.500000	832.666667	442.666667	753.333333
Durée(mois)	52.142857	8.500000	47.000000	47.222222	10.333333
Montant	1616.142857	14500.000000	1813.649984	1243.322222	1720.000000
Age	28.000000	23.000000	27.666667	58.777778	42.666667
Nombres de prêts antérieurs dans cette banque	2.285714	5.000000	3.666667	4.666667	0.000000

Figure 13 : Coordonnées des centres de chacun des cinq clusters.

### 3.4.- Groupe de caractérisation

Pour être explicite, nous présentons la répartition des individus au sein des différents groupes (clusters) à partir d'une caractérisation de ces derniers en utilisant l'outil **Group Characterization** de **Tanagra**.

Results																			
Description of "Objectif"																			
Objectif=Logement				Objectif=Achat véhicule				Objectif=Aucun				Objectif=Soins de santé				Objectif=Consommation			
Exemples [ 33.3 %] 8				Exemples [ 16.7 %] 4				Exemples [ 12.5 %] 3				Exemples [ 12.5 %] 3				Exemples [ 25.0 %] 6			
Att - Desc	Test value	Group	Overall	Att - Desc	Test value	Group	Overall	Att - Desc	Test value	Group	Overall	Att - Desc	Test value	Group	Overall	Att - Desc	Test value	Group	Overall
Continuous attributes : Mean (StdDev)				Continuous attributes : Mean (StdDev)				Continuous attributes : Mean (StdDev)				Continuous attributes : Mean (StdDev)				Continuous attributes : Mean (StdDev)			
Durée(mois)	1.36	48.38 (11.69)	40.79 (18.92)	Nombres de prêts antérieurs dans cette banque	1.55	5.75 (4.35)	3.29 (3.39)	Nombres de prêts antérieurs dans cette banque	1.80	6.67 (2.52)	3.29 (3.39)	Durée(mois)	1.33	54.67 (0.58)	40.79 (18.92)	Montant	2.16	5606.82 (6906.82)	2587.66 (3862.28)
Age	1.26	48.13 (20.64)	40.92 (19.43)									Solde	1.17	652.33 (222.09)	506.88 (224.85)	Nombres de prêts antérieurs dans cette banque	0.85	4.33 (3.27)	3.29 (3.39)
Montant	-0.35	2190.25 (1623.42)	2587.66 (3862.28)	Solde	0.73	583.25 (275.73)	506.88 (224.85)	Durée(mois)	1.36	55.00 (5.00)	40.79 (18.92)	Nombres de prêts antérieurs dans cette banque	-0.51	2.33 (1.53)	3.29 (3.39)	Solde	0.47	545.00 (202.48)	506.88 (224.85)
Solde	-0.93	445.00 (234.84)	506.88 (224.85)	Montant	-0.41	1850.00 (768.11)	2587.66 (3862.28)	Age	0.04	41.33 (19.50)	40.92 (19.43)	Montant	-0.66	1180.32 (312.32)	2587.66 (3862.28)	Age	0.27	42.83 (20.62)	40.92 (19.43)
Nombres de prêts antérieurs dans cette banque	-2.91	0.38 (0.52)	3.29 (3.39)	Age	-0.46	36.75 (20.97)	40.92 (19.43)	Montant	-1.21	0.00 (0.00)	2587.66 (3862.28)	Age	-1.67	23.00 (3.61)	40.92 (19.43)	Durée(mois)	-2.95	20.67 (17.57)	40.79 (18.92)
				Durée(mois)	-0.68	34.75 (22.10)	40.79 (18.92)	Solde	-1.28	348.33 (147.41)	506.88 (224.85)								
Discrete attributes : [Recall] Accuracy				Discrete attributes : [Recall] Accuracy				Discrete attributes : [Recall] Accuracy				Discrete attributes : [Recall] Accuracy				Discrete attributes : [Recall] Accuracy			

Figure 14 : Caractérisation des clusters

### 3.5.- La projection des individus suivant leur cluster dans le premier plan factoriel

Elle est décrite dans la figure suivante :

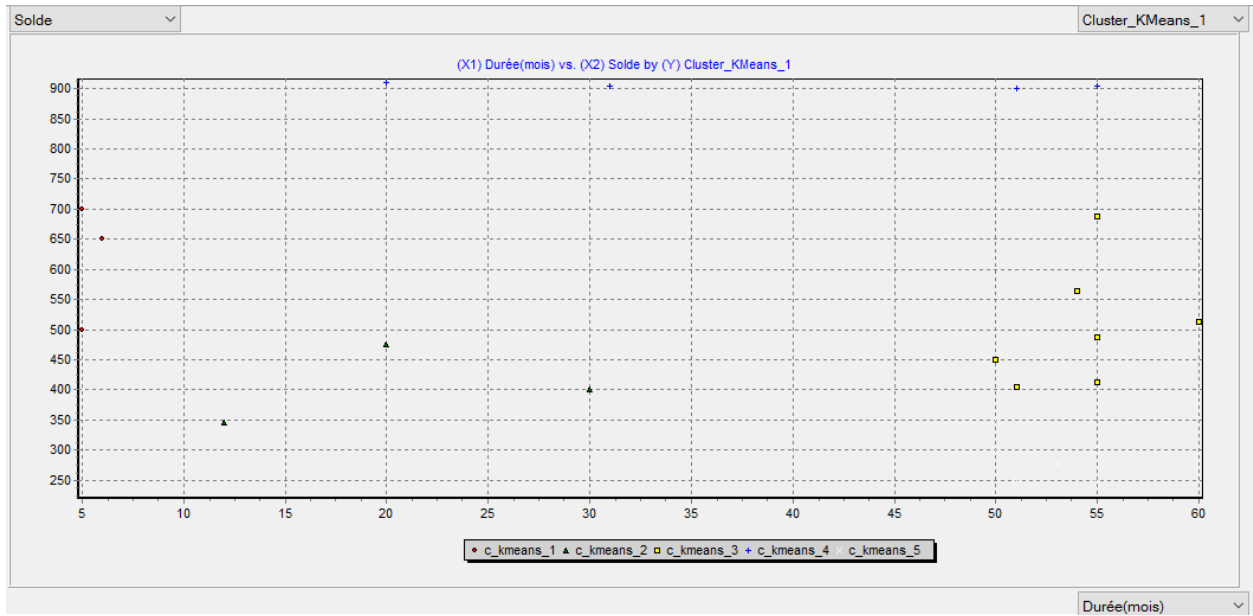


Figure 15 : Projection des individus par cluster

### 3.6.- HAC

Pour illustrer la classification hiérarchique, nous utilisons toutes nos variables Continues et une variable discrète (Objet). Nous appliquons l'outil HAC de l'onglet Clustering de Tanagra.

#### Best cluster selection

Clusters	BSS ratio	Gap
1	0.0000	0.0000
2	0.1524	0.3875
3	0.2273	0.0868
4	0.2849	0.0999
5	0.3225	0.1880

Figure 16 : Sélection du nombre optimal de clusters

- ✓ Les quatre clusters obtenus après réorganisation des données sont présentés à la figure 17 et les coordonnées de leurs centres à la figure 18.

## Clustering results

Clusters	From the dendrogram	After one-pass relocation
cluster n°1	10	10
cluster n°2	3	3
cluster n°3	8	8
cluster n°4	3	3

Figure 17 : Résultat du Clustering

### Interprétation :

Dans la figure 17 où nous apparaissent les quatre clusters, nous pouvons remarquer que les valeurs restent même au départ et après relocation. En plus, le cluster n°1 est celui ayant la plus grande valeur.

## Cluster centroids

Attribute	Cluster n°1	Cluster n°2	Cluster n°3	Cluster n°4
Solde	560.300000	348.333333	445.000000	652.333333
Durée(mois)	26.300000	55.000000	48.375000	54.666667
Montant	4104.089999	0.000000	2190.250000	1180.316650
Age	40.400000	41.333333	48.125000	23.000000
Nombres de prêts antérieurs dans cette banque	4.900000	6.666667	0.375000	2.333333

Les COORDONNÉES CENTRALES des clusters obtenus

Figure 18 : Coordonnées des centres des clusters

### Interprétation :

Après avoir représenté les clusters, il est à remarquer que dans la figure 18 où représentent les coordonnées de leurs centres, l'attribut **Montant** est celui ayant la plus grande valeur dans tous les clusters, sauf dans le n° 2. Alors que l'attribut **Nombre de prêts antérieurs dans cette banque** est celui ayant la plus petite valeur dans tous les clusters, sauf dans le n° 2.