



Application de l'Analyse en Composantes Principales (ACP) dans le domaine d'impôt sur les véhicules : Evaluation de la solvabilité de contribuables à la Direction de Recettes du Haut Katanga (DRHKAT)

Ruphin NYAMI¹, Philippes NKAYA², Jonathan KABASU³

- 1. Université de Lubumbashi / Facultés de Sciences et Technologies / Département de Mathématiques et Informatique, RDC*
- 2. Université Protestante de Lubumbashi / Facultés de Sciences Informatiques, RDC*
- 3. Institut Supérieur de Commerce d'Ilebo, RDC*

This is an open access article under the [CC BY-NC-ND](https://creativecommons.org/licenses/by-nc-nd/4.0/) license.



Résumé

Cet article offre une analyse méthodique et approfondie de l'évaluation de la solvabilité des contribuables en ce qui concerne l'impôt sur les véhicules dans la Direction des Recettes du Haut-Katanga. L'utilisation de l'Analyse en Composantes Principales (ACP) pour détecter les similitudes entre les contribuables et identifier les tendances cachées est une approche novatrice et pertinente. Le papier propose une approche innovante pour évaluer la solvabilité des contribuables en matière d'impôt sur les véhicules dans la Direction des Recettes du Haut-Katanga. Les résultats de l'étude, qui mettent en lumière la corrélation entre la solvabilité et des variables telles que le statut du contribuable et la localisation géographique, offrent des perspectives précieuses pour les politiques fiscales et la gestion des recettes dans la région. En outre, nous démontrons une maîtrise remarquable des techniques d'exploration de données et de réduction de dimensions, ce qui facilite grandement l'interprétation des résultats et

permet une sensibilisation ciblée. En conclusion, cet article constitue une contribution significative à la littérature sur la fiscalité locale et mérite d'être publié.

Mots-clés : Data Mining d'impôts ; ACP ; cercle de corrélation ; paiement de pénalités

Abstract

This article provides a methodical and in-depth analysis of the assessment of taxpayer solvency with regard to vehicle tax in the Haut-Katanga Revenue Directorate. Using Principal Component Analysis (PCA) to detect similarities between taxpayers and identify hidden trends is an innovative and relevant approach. The paper proposes an innovative approach to assess the solvency of taxpayers in terms of vehicle tax in the Haut-Katanga Revenue Directorate. The study results, which highlight the correlation between creditworthiness and variables such as taxpayer status and geographic location, offer valuable insights for tax policies and revenue management in the region. Additionally, we demonstrate remarkable mastery of data mining and dimension reduction techniques, which greatly facilitates the interpretation of results and enables targeted outreach. In conclusion, this article constitutes a significant contribution to the literature on local taxation and deserves to be published.

Keywords: Tax data mining; PCA; correlation circle; payment of penalties

Digital Object Identifier (DOI): <https://doi.org/10.5281/zenodo.11454841>

1 Introduction

Les enjeux de développement de la RDC sont en constante évolution et propulsent les entités décentralisées à mobiliser davantage les recettes pour la construction d'une société stable. Chaque entité décentralisée se voit dotée d'une Direction de Recettes exerçant en matière de recettes fiscales et non fiscales, toutes les missions et prérogatives sur les droits revenants à l'entité décentralisée. La mobilisation de recettes donne à l'Etat les moyens pour jouer le rôle primordial de garantir les conditions de vies de la population en termes d'emploi, de la création de services, de la construction des infrastructures à intérêt, garantir l'enseignement de base etc. L'un des axes financiers important considéré comme source de recettes provinciales de premier rang est l'impôt sur le véhicule dit « vignette automobile » [1]. Payer l'impôt est une obligation légale et un devoir civique pour chaque citoyen institué dans la constitution selon l'esprit de

l'Art. 174 de la Constitution de la RDC du 18 février 2006 ; complété par l'Art 65 qui surenchère : « Tout Congolais est tenu de remplir loyalement ses obligations vis-à-vis l'Etat » et constitue donc un devoir civique de booster le développement du pays [2]. Néanmoins, une entité décentralisée ne peut jouer pleinement son rôle du moteur de développement qu'à partir de solvabilité de contribuables aux impôts et taxes. En outre, la compétitivité de développement des entités décentralisées réside sur la solvabilité des contribuables. La mobilisation recettes de l'impôt sur les véhicules nécessite la mise à disposition des agents de recouvrement et de décideurs une bonne information, une bonne analyse sur les comportements de contribuables afin de faire de prévisions fiables.

En général, lors de la perception de l'impôt, le système collecte quotidiennement de masses importantes de données tant sur les caractéristiques administratives de contribuables, de détails sur les paiements d'impôts et taxes automobiles, des informations sur les pénalités dues aux défauts de paiement à l'échéance. In fine, tant de décisions sont prises dans un bureau qui se veut être proactive et souple sur la budgétisation, le plan de mobilisation et le recouvrement forcé de recettes. Toutefois, prendre une décision critique nécessite de disposer d'une information précise, ponctuelle et exhaustive provenant du système d'information existant. Le système d'information doit analyser, traiter les données recueillies dans les exercices fiscaux passés dans différents points de perception d'impôts pour prédire une information[3]. Les données sur les paiements de vignettes automobiles regorgent un énorme potentiel pour les bureaux de recettes d'user dans leur processus de prise de décisions, l'exploration de données dans les processus tels que le recouvrement, la gestion de contentieux, le planning de recettes à temps. Dans ces conditions, la DRHKAT peut faire de prévisions fiables, axer le recouvrement de l'impôt avec précision en tenant compte de défauts de paiements de certains contribuables défaillants. Habituellement, les contribuables sont regroupés dans les différents bureaux de recettes où ils sont sensé payer l'impôt ; ces bureaux fournissent chacun un ensemble d'information sur la marque du véhicule, le nombre de chevaux, la date de paiement, les pénalités à payés. Toutes ces informations constituent de variables utiles pour l'analyse en composantes principales pour déterminer les corrélations existantes entre ces variables et classifier les individus (assujettis) selon leur habitude de paiement de la vignette. Comme le souligne *Stéphane* [4], face aux progrès techniques et l'envie de bien faire les choses en matière d'exploration de données, cette masse de données sur les vignettes constitue une mine d'or pour l'organisation. Des connaissances précieuses et des modèles intéressants sont cachés dans ces données. L'absence de l'information sur la caractérisation des contribuables en fonction de leur

habitude ou sur les variables les caractérisant peut occasionner l'illusion sur la mobilisation de recettes et handicaper ainsi la mise en œuvre de la politique de développement définie.

Pour le reste, le volume de données à extraire dans plusieurs points de perceptions constitue le premier handicap de la prise de décisions et complexifie l'extraction de connaissances cachées. C'est l'apanage de l'Analyse en Composantes Principales (ACP) qui, est un élément constitutif des méthodes descriptives multidimensionnelles dites méthodes factorielles apparue dans les années 60 [5]. Introduite par K. Pearson en 1901 et ensuite développée par H. Heotelling 1933[6], l'ACP est l'une des puissantes méthodes utilisées pour explorer la structure d'une masse importante de données quantitatives. L'ACP est l'une de techniques de l'exploration de données (Data Mining) la plus connue [7]. Elle est souvent utilisée dans le cadre d'analyse de données multi-variées difficiles à représenter graphiquement. Elle permet d'étudier des ensembles de données multidimensionnelles avec des variables quantitatives. Comme toute autre méthode descriptive, l'ACP se base sur un modèle géométrique issu d'une structure de données rectangulaire [8]. Cet outil d'analyse descriptive produit de représentations géométriques à partir de variables quantitatives afin de représenter les unités comparables, connues à priori afin d'étudier l'organisation de données linéaires. L'idée derrière l'ACP est de se noyer dans une masse de données sans avoir à priori une idée quelconque sur les données multi-facteurs. L'ACP vise à concentrer le maximum d'informations sur les premières dimensions et ensuite de segmenter les individus selon les valeurs propres dans une dimension [8]. *Dominique Desbois*[9] a démontré l'apport de l'ACP dans la classification par intervalles aux estimations quantiles de coûts d'engrais, qui redistribue l'information contenue dans les premiers facteurs. En 2016 Yamina et son équipe ont démontré l'application de l'ACP dans une base de données de 500 liquides ioniques différents afin de desceller le facteur de la conductivité électrique[10]. Le résultat remarquable obtenu grâce à l'ACP, montre que la forte conductivité électrique est corrélée positivement avec la faible température de la transition vitreuse. En outre, comme le souligne *Nassima* en 2021 [11], l'usage de l'ACP dans le secteur du PME permet de répartir les contribuables en trois groupes dont le premier est caractérisé par une survie passive, dont le défaut de paiement est inévitable ; le second groupe de contribuables est en situation financière stable et enfin, le troisième groupe est celui ayant une situation financière satisfaisante.

Dans le cadre de cet article, l'usage d'ACP permettra de résumer un jeu de données sur les pénalités payées sur les vignettes automobiles, ayant plusieurs colonnes sur quelques graphiques afin d'étudier les liens entre les variables (pénalités sur les remorques, les véhicules

de différents tonnages et de chevaux variés; évaluer la proximité des contribuables attachés aux différentes communes entre eux ; et enfin la caractérisation des contribuables par des variables selon le type de véhicules en termes de tonnages et chevaux. Concrètement, nous partons d'un jeu de données de départ où la quasi-totalité de l'information est répartie harmonieusement et contenue dans plusieurs colonnes pour fabriquer artificiellement un tableau de données avec la densité de l'information dans les premiers facteurs (dimensions)[12].

Comme effleuré ci-haut, la prise de décision dans une Direction des Recettes d'une entité donnée requiert de rapports issus de différents bureaux de recettes (points de perceptions). Chaque rapport périodique généré par le système d'information d'impôts est exploité manuellement dans le processus de prise de décision. Toutefois, il est possible d'extraire les données du SI existant faire présenter de rapports sous-forme de nuages de points, tableau de contingences. Néanmoins il est difficile d'intégrer plusieurs variables et ressortir l'information cachée. De surcroît, les possibilités qu'offre l'analyse manuelle restent limitées car le volume de données à explorer est énorme donc les décisions qui en découlent peuvent être hasardeuses et risquées [3]. Par exemple, il est possible qu'une catégorie de contribuables soit solvable qu'après un recouvrement forcé ou qu'une autre catégorie reste insolvable. Cette tendance négative alarmante dans la caisse de l'Etat peut avoir de conséquences néfastes sur les objectifs de développement de la ville. Il sied de constater que les contribuables disposants de Motos communément appelés « *Manseba ou Motards* » sont réputés à la délinquance fiscale et qu'un recouvrement sur cette couche de la population va jusqu'à troubler la paix sociale[13].

Ces pépites utiles à la prise de décision ne sont pas détectables facilement par des processus manuels. C'est sans doute que cette masse de données opérationnelles et archivées cache des connaissances précieuses et qu'elles peuvent servir de soubassement à la prise de décisions critiques si elles sont découvertes et exploitées en bon escient par des outils appropriés [14]. Par exemple, l'utilisation de l'ACP pour l'exploration de données peut aider à lancer la sensibilisation fiscale ciblée pour un groupe d'assujettis. La figure 2 montre comment l'exploration de données peut améliorer le processus de prise de décision. La mobilisation de recettes au-delà des objectifs fixés procure non seulement de la motivation exceptionnelle mais aussi une récompense en termes de promotions, primes et autres avantages ; Les responsables des bureaux de recettes doivent être proactives dans l'analyse des préférences et habitudes de contribuables afin d'adapter leur politique de recouvrement en conséquence pour conserver leur leadership. En segmentant les contribuables en solvable prompt, solvable tardivement et non solvable, la Direction des Recettes peut réadapter sa politique de

mobilisation de recettes si tôt et réduire les pertes avant qu'il ne soit trop tard [15] [16]. Ces fonctionnalités hautement souhaitées sont réalisées grâce à l'exploration de données. L'impôt sur les véhicules automobiles « vignette » est encadré par la loi [1]. La nomenclature de tous les impôts, droits, taxes et redevances des entités territoriales décentralisées, la modalité de paiement. Tout véhicule en circulation sur la voie publique est d'office frappé par l'impôt. Néanmoins il existe une catégorie de véhicules exonérés à cet impôt [17] [1]. Outre les véhicules exonérés, le montant de la vignette dépend de la catégorie, du poids et de la puissance fiscale du véhicule. Par ailleurs, la DRHKAT regroupe les assujettis en deux classes à savoir les personnes physiques et morales. Une division est réservée aux personnes morales et une autre division pour les personnes physiques. Certains bureaux sont attachés aux entités administratives ou communes dont les potentiels contribuables sont les habitants de l'entité. La souscription à la vignette est lancée par une sensibilisation de la population et, elle est délimitée dans une période de temps à laquelle tout assujetti doit se procurer la vignette automobile sous peine de pénalités. Dépassé cette période, tout véhicule non estampillé par une vignette est obligé de payer cet impôt avec pénalités.

Ainsi, il y a de s'interroger au sens de savoir : *existe-il un lien entre le type de véhicule et la solvabilité de la vignette ? Existe-il une corrélation entre la région d'habitation et la solvabilité ?*

L'objectif de cette étude est d'appliquer d'exploration de données avec ACP sur les données d'impôts sur les véhicules automobiles « vignette » en vue de d'identifier les comportements de contribuables de différentes communes, redynamiser la politique de maximisation de recettes, de projeter et analyser les corrélations entre les différentes variables notamment, d'identifier les principaux facteurs à la base de paiement de pénalités. Pour y arriver, cette exploration est structurée autour de deux points à savoir : la méthodologie et les résultats.

1. Matériels et Méthodologie

Cette étude se base sur les données recueillies à la DRHKAT pour la période allant de l'exercice fiscal 2021 au deuxième trimestre de l'exercice fiscal 2023. Ces données sont issues de rapports périodiques produits à partir du système d'information existant.

Les données concernent 13 bureaux de recettes regroupées en deux catégories à savoir les personnes morales « PM » et celle des personnes physiques « PP » comme suit :

- GRANAT (Bureau Personnes Morale de la ville de LUBUMBASHI & KIPUSHI),
- KPBA (Bureau secondaire de la commune KAMPEMBA),

- KLSA_PP (Bureau Personnes Physique de la ville de KASUMBALESA),
- KLSA_PM (Bureau Personnes Morale de la ville de KASUMBALESA),
- KENYA (Bureau secondaire de la commune KENYA & KAMALONDO),
- RSHI (Bureau secondaire de la commune RUASHI),
- ANX (Bureau secondaire de la commune ANNEXE)
- LIKSI_PM (Bureau personnes Morales de la ville de LIKASI)
- LIKSI_PP (Bureau personnes Physique de la ville de LIKASI),
- DRL1 (Bureau de la Division Urbaine de Lubumbashi 1)
- DRL2 (Bureau de la Division Urbaine de Lubumbashi 2)
- APHK (Bureau de la Division Urbaine de Lubumbashi de l'Assemblée Provinciale du Haut KATANGA)
- KPSHI_PP (Bureau Personnes Physique de la ville de KIPUSHI)

Sur les 13 bureaux, nous avons analysé les paiements de pénalités de la vignette de $P = 9$ variables (types de véhicules) suivantes :

- P_11_15_CV : (Pénalités de véhicules ayant une puissance fiscale allant de 11 à 15 chevaux);
- P_PLUS_15CV : (Pénalités de véhicules de plus de 15 chevaux);
- P_PLUS_10T : (Pénalités de véhicules de plus de 10 tonnes) ;
- P_2_5_10T : (Pénalités de véhicules de 2.5 à 10 tonnes)
- P_REMO_10T : (Pénalités de camions Remorque de 10 tonnes) ;
- P_1_10CV : (Pénalités de de véhicules de 1 à 10 chevaux) ;
- P_2_5_10_REMO : (Pénalités de véhicules de 2.5 à 10 tonnes remorques)
- P_MOINS_2_5T : (pénalités de véhicules de moins de 2.5 tonnes)
- P_MOTO : (pénalités de motos)

Nous voulons décrire les comportements de chaque entité observée par rapport à une catégorie de véhicules d'une part et catégoriser les comportements de contribuables en fonction de leur solvabilité comme illustré dans le *Tableau 1* suivante. Ainsi, le traitement en vue de la consolidation de rapports de toutes les directions nous a permis d'avoir un jeu de données facilement explorable avec ACP. L'exploration de données à l'aide de l'ACP utilisée dans ce document suit la démarche suivante :

- *La sélection de données* où l'extraction de données pertinentes est faite ;

- *Le prétraitement de données* : dans cette étape où le nettoyage de données est réalisé, l'objectif est d'éliminer les données non pertinentes, les bruits ainsi que les données incohérentes ;
- *L'intégration de données* : Dans un environnement de production, il peut y avoir plusieurs bases de données stockant les mêmes informations. Ces sources de données hétérogènes sont combinées dans une source commune.
- *La Transformation de données* : à ce niveau, nous appliquons les opérations de résumé ou d'agrégation de données de plusieurs directions afin qu'elles soient facile à manipuler.
- *Extraction de données* : nous appliquons ici la technique d'ACP pour extraire les données d'un fichier CSV (Comma Separated Values) en DataFrame facilement manipulable par des fonctions d'ACP ;
- *Présentation de connaissances* : à ce niveau il est question d'interpréter les résultats obtenus de différents graphiques. Signalons que les connaissances déduites doivent être neuves toutefois elles ne peuvent pas être évidentes.

Le cheminement logique adopté dans cet article basé sur l'exploration de données d'impôts «vignette » et la découverte de connaissances est illustré à la figure 1 suivante.

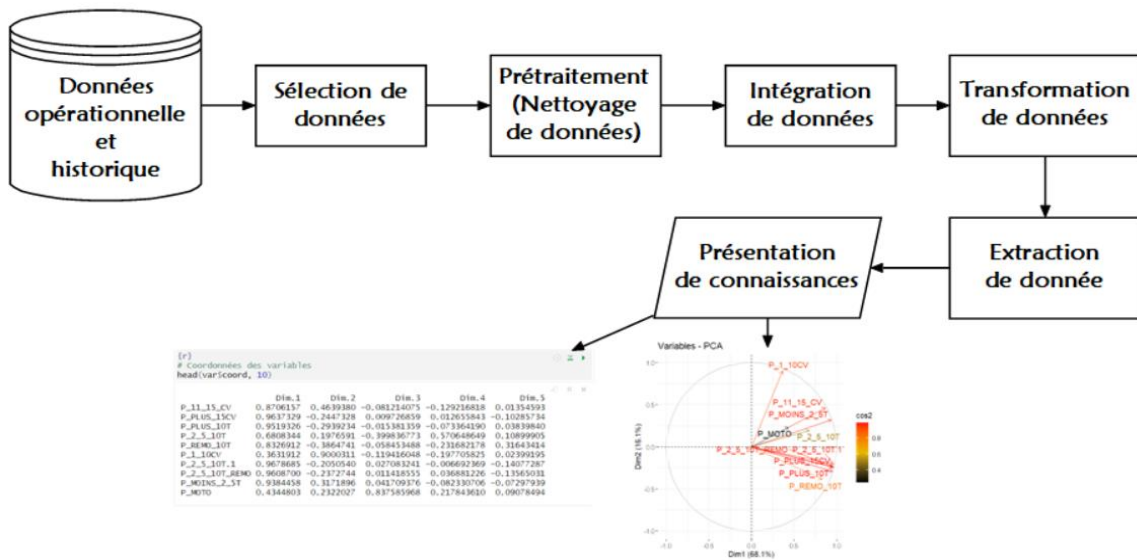


Figure 1—Processus d'exploration de données adoptée

Source : Auteurs

L'activité principale de la DRHKAT est accomplir les prérogatives légales dans la mobilisation de Recettes fiscales, non fiscales ainsi que de droits de la Province du Haut-Katanga sur toute son étendue. De manière spécifique, la DRHKAT s'occupe de la perception de recettes, de la

taxation, l'ordonnancement ainsi que le contrôle de pièces ou le recouvrement et le contentieux des impôts provinciaux et locaux.

Les données sur le paiement des amendes pendant la période d'étude sont présentées dans le tableau 1.

Tableau 1—Données de paiements de pénalités de la vignette dans la période d'étude

	DIRECTION	P_11_15_CV	P_PLUS_15CV	P_PLUS_10T	P_REMO_10T	P_1_10CV	P_2_5_10T	P_2_5_10T_REMO	P_MOINS_2_5T	P_MOTO
1	ANX	15.18	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2	APKHK	6462.46	357.64	975.00	38.16	3311.68	1620.14	0.00	624.91	104.00
3	DRL_UN	3731.92	49.90	323.28	70.56	2222.06	746.00	0.00	346.86	0.00
4	DRL_DEUX	3765.60	223.30	797.16	0.00	2207.22	730.45	149.18	282.12	5.00
5	GRANAT	9018.00	38661.53	70237.56	25260.46	1449.52	29044.86	46974.44	1125.30	116.32
6	KENYA	685.74	24.24	25.25	0.00	306.48	148.58	0.00	134.16	31.00
7	KPBA	976.46	47.16	0.00	0.00	641.10	210.35	14.65	168.34	206.40
8	KPSHI_PP	813.84	44.40	6.48	0.00	628.98	438.12	166.80	85.86	26.20
9	LIKSI	2915.80	1518.92	11687.76	11392.92	1229.76	1083.96	550.44	216.24	0.00
10	LIKSI_PM	301.63	1267.32	11428.56	11392.92	0.00	0.00	0.00	0.00	0.00
11	LIKSI_PP	2614.17	251.60	259.20	0.00	1229.76	1083.96	550.44	216.24	8.80
12	RSHI	631.92	19.24	10.14	0.00	452.52	741.18	145.56	38.88	0.00
13	KLSA_PM	55.60	0.00	0.00	0.00	38.61	0.00	0.00	0.00	0.00

Source : Auteurs

Dans le tableau ci-haut, sont reprises les communes ayant payé hors délai l'impôt sur les véhicules suivant la puissance fiscale du véhicule. Le résultat brut montre que les personnes morales paient plus les pénalités par rapport aux personnes physiques, pour la simple raison que les pénalités sont fonction de la puissance fiscale ; et celle cette catégorie de contribuable dispose la quasi-totalité de véhicules poids lourds.

Les pénalités sont calculées sur base critères de puissances fiscales ainsi le taux d'imposition même de la vignette[1]. Ainsi, peut être considéré comme assujetti, tout détenteur d'un véhicule admis à circuler sur l'étendue de la province du Haut Katanga. Néanmoins, le tableau 1 montre un déséquilibre de la répartition de types de véhicules dans différentes régions en termes de paiement de pénalités.

Pour le souci de résumer ce jeu de données obtenu dans le paiement hors délai d'impôts sur les véhicules, une ACP a été réalisée sur les puissances fiscales comme variables et sur les communes comme individu les 14 directions. La méthode ACP est utilisée afin de desceller les tendances, les liens et les phénomènes susceptibles d'influencer le paiement de pénalités dans certaines communes.

1.1. La méthode coude Elbow

L'un des inconnus dans l'ACP c'est comment choisir le nombre optimal des axes factoriels dans un jeu de données ? C'est-à-dire à quel point l'analyse doit s'arrêter sur un jeu de données

visant la maximisation de la variance expliquée. En général, le premier axe factoriel en ACP explique la plus grande variance expliquée, les restes des axes s'occupent de l'explication des restes de variances[8]. Quel que soit le nombre d'axes factoriels, le principe de non-redondance de l'information impose une orthogonalité des axes factoriels. Une des méthodes utilisées pour déterminer le nombre des axes factoriels optimal de la variance expliquée est la méthode du **coude Elbow** [18], [19], stipulent que pour avoir un nombre maximal de la variance expliquée, il faut que prendre le coude qui maximise le plus d'inertie d'information.

2. Résultats et discussions

Suit à l'ACP, il est aisé de constater que la masse importante informations est concentrée et expliquée par les deux premières dimensions (facteurs ou axes factoriels). Considérant le plan factoriel Dim1 x Dim2, les valeurs non nulles (propres) des deux dimensions Dim1 et Dim2 ainsi que leur inertie totale sont repris dans le Figure 3 et Figure 4. Les deux premiers axes factoriels (dimensions) sont à considérer pour notre analyse afin décrire les variables liées en termes de corrélations aux et détiennent à eux seuls à **88.58%** de l'inertie totale avec écart de de variance respective de 71.0 et pour la 1^{ère} dimension et **17.57 %** pour la deuxième dimension. L'axe 1, est exprimé vers son pôle positif par les véhicules allant de 2.5 à 10 tonnes, de 11 à 15 chevaux, de plus de 15 chevaux, Remorque de plus de 10 et de moins de 2.5 tonnes. Par contre la variable moto a un cosinus carré faible d'où la mauvaise qualité de la projection. (Figure 4). Alors que le deuxième facteur (Dim 2) est défini par les véhicules dont la puissance fiscale varie entre 1 et 10 chevaux vers son pôle positif. Le Tableau 2 suivant montre que plus la corrélation est élevée c'est-à-dire proche de 1, plus la variable n'est liée à la composante et la qualité de projection est meilleure. A l'inverse, plus le cosinus carré est proche de 0 moins la qualité de la projection est meilleure, alors la variable est liée à cette composante, il faut essayer dans une composante.

Tableau 2—Tableau de Corrélation groupée par direction

	P_11_15_CV	P_PLUS_15CV	P_PLUS_10T	P_REMO_10T	P_1_10CV	P_2_5_10T	P_2_5_10T_REMO	P_MOINS_2_5T	P_MOTO
P_11_15_CV	1.00	0.72	0.70	0.58	0.77	0.75	0.72	0.97	0.39
P_PLUS_15CV	0.72	1.00	0.98	0.86	0.12	1.00	1.00	0.83	0.36
P_PLUS_10T	0.70	0.98	1.00	0.94	0.10	0.97	0.98	0.80	0.32
P_REMO_10T	0.58	0.86	0.94	1.00	0.02	0.84	0.84	0.65	0.20
P_1_10CV	0.77	0.12	0.10	0.02	1.00	0.18	0.12	0.63	0.23
P_2_5_10T	0.75	1.00	0.97	0.84	0.18	1.00	1.00	0.86	0.38
P_2_5_10T_REMO	0.72	1.00	0.98	0.84	0.12	1.00	1.00	0.83	0.37
P_MOINS_2_5T	0.97	0.83	0.80	0.65	0.63	0.86	0.83	1.00	0.49
P_MOTO	0.39	0.36	0.32	0.20	0.23	0.38	0.37	0.49	1.00

Source : Auteurs

Nous remarquons le résultat évident dans le Tableau 2 et la Figure 2 suivante :

- Une corrélation positive entre les véhicules ayant une puissance fiscale allant de 11 à 15 CV et ceux de Moins de 2.5 Tonnes, de plus de 15 CV, de plus de 10 tonnes, de remorques, de véhicules légers : Généralement les propriétaires de véhicules de moins de 2.5 tonnes possèdent également les véhicules ayant une puissance fiscale allant de 11 à 15 CV et paient bien l'impôt sur les véhicules.
- Une corrélation faible entre les motos et les autres véhicules : cela veut tout simplement dire que les propriétaires ont rarement de motos et vice versa (plus on a de motos, moins on aimerait avoir un véhicule à partir de 4 roues), par conséquent difficile de payer les pénalités de la moto si on en dispose pas. Il semblerait que pour les détenteurs de motos habitant des zones autres que le centre-ville, sont insolvable. La raison la plus frappante est que cette catégorie de véhicules œuvre dans les périphéries de la ville où la campagne fiscale n'est pas accentuée. Exception faite aux particuliers détenteurs de motos qui aiment entrer au centre-ville.

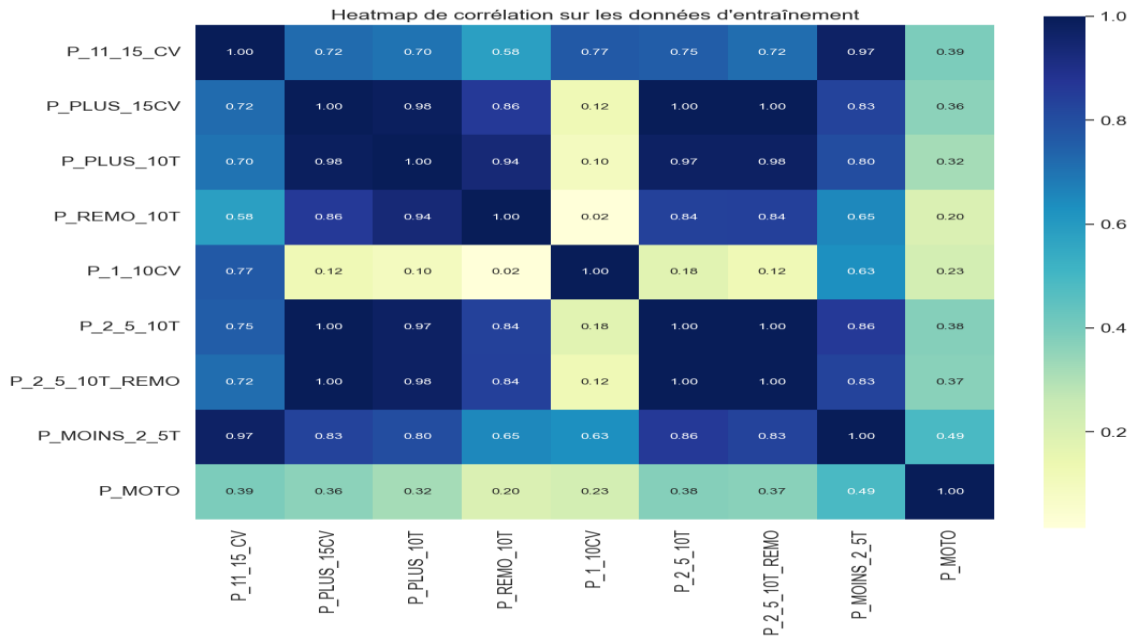


Figure 2—Matrice de corrélation

Pour de raisons de qualité de notre ACP, la Figure 3 suivante présente des informations importantes «eigenvalues» indiquant ainsi l’inertie totale du jeu de données de départ regroupée par les composantes premières composantes de l’analyse.

 	eigenvalue	variance.percent	cumulative.variance.percent
Dim. 1	6.391	71.01	71.01
Dim. 2	1.581	17.57	88.58
Dim. 3	0.8063	8.959	97.54
Dim. 4	0.2056	2.285	99.83
Dim. 5	0.0114	0.1267	99.95
Dim. 6	0.003776	0.04196	99.99
Dim. 7	0.0004383	0.00487	100
Dim. 8	5.303e-05	0.0005893	100
Dim. 9	1.227e-06	1.363e-05	100

Figure 3—Inertie du jeu de données de départ
Source : Auteurs

Nous remarquons que le maximum d’informations est regroupé dans les 2 premières dimensions avec l’inertie totale de 88.58%. Nous allons nous concentrer dans ce premier bloque (Dimensions). Comme illustré dans la Figure 4 suivant, où la représentation graphique de la variance expliquée cumulée en fonction de de nombre de composantes principales montre clairement qu’à deux coudes, l’inertie totale atteint 88.5%. Pour de raisons de simplifications et suivant les objectifs d’analyse, nous nous limitons volontiers à deux première dimensions, pourtant à 3 dimensions nous atteignons Identification 98 % de la variance expliquée.

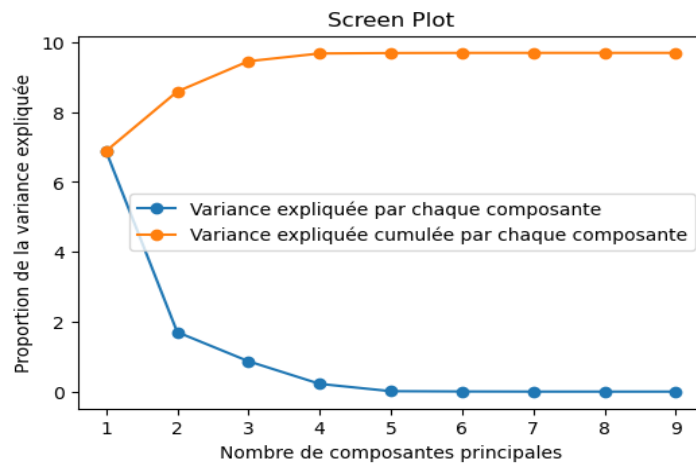


Figure 4—Calcul de la variance expliquée cumulée de différentes composantes
Source : Auteurs

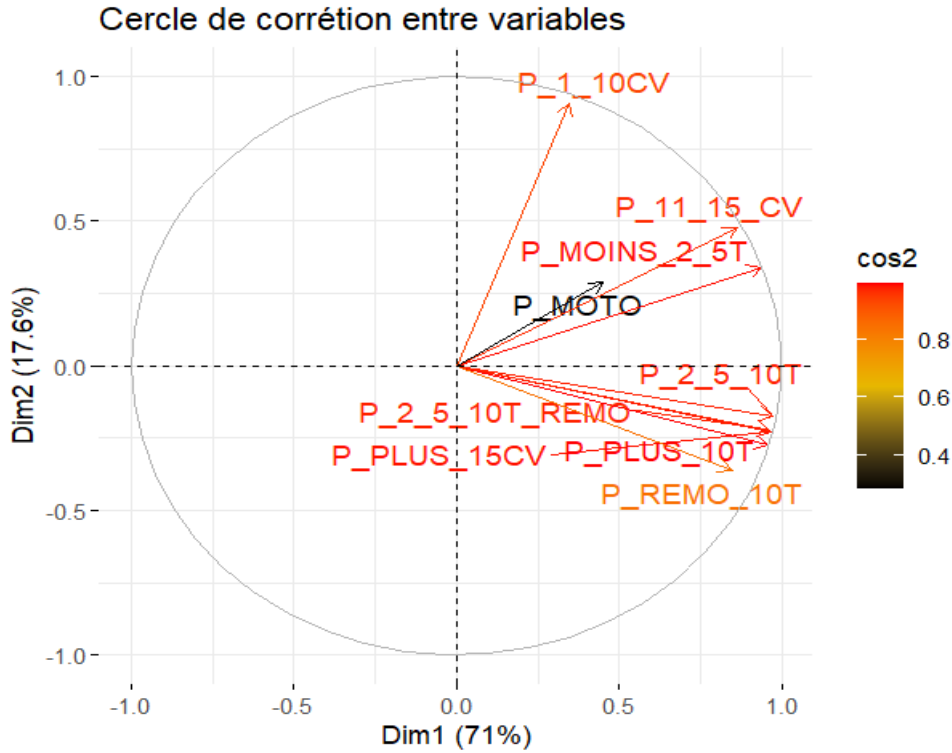


Figure 5—cercle de corrélation de variables
Source : Auteurs

En plus de la corrélation entre les variables de jeu de données (puissance fiscale du véhicule), ACP nous offre la possibilité d'analyser la corrélation entre les individus statistiques ici, les contribuables regroupés de de commune (bureau). Dans la Figure 6 il ressort les informations importantes de paiement de pénalités de l'impôt sur les véhicules. Nous remarquons que seule GRANAT (Figure 6 A) est l'un de points le plus parlant et intéressant qui est assez proche d'un des axes, loin de l'origine. C'est pour dire qu'en matière du civisme fiscal, les personnes morales sont solvables par rapport aux particuliers.

On remarque également qu'APHK mène une politique fiscale sur ses assujettis, car ces derniers excellent dans le paiement de l'impôt de véhicule de 1 à 10 chevaux. Par contre, nous avons une corrélation faible entre les assujettis d'APHK et ceux du GRANAT. Ce qui veut dire que plus l'on dispose de véhicules poids lourds, moins on aura les petits véhicules. Cela peut s'expliquer aussi du fait que les personnes morales sont en majorité dans le secteur minier et sont les seules à disposer de véhicules de puissances fiscales élevées. Les autres bureaux (DRL1 et 2, LKSI_PP, LKSI_PM, KPSHI_PP, KLSA_PP, ANX, RSHI) sont à gauche et proche de l'origine et cela dénote une insuffisance du civisme fiscal, dans ces communes. La Figure 6 (B) regroupe les assujettis et les types de véhicules préférés et nous montre clairement que les contribuables de la Commune KAMPEMBA (KPBA), LIKASI, de deux bureaux du

centre-ville de Lubumbashi disposent en majorité de véhicules légers (de 1 à 15 CV). Cela se justifie du fait que bon nombre de ces véhicules sont dans le transport et préfèrent payer hors délai en cas de recouvrement forcé.

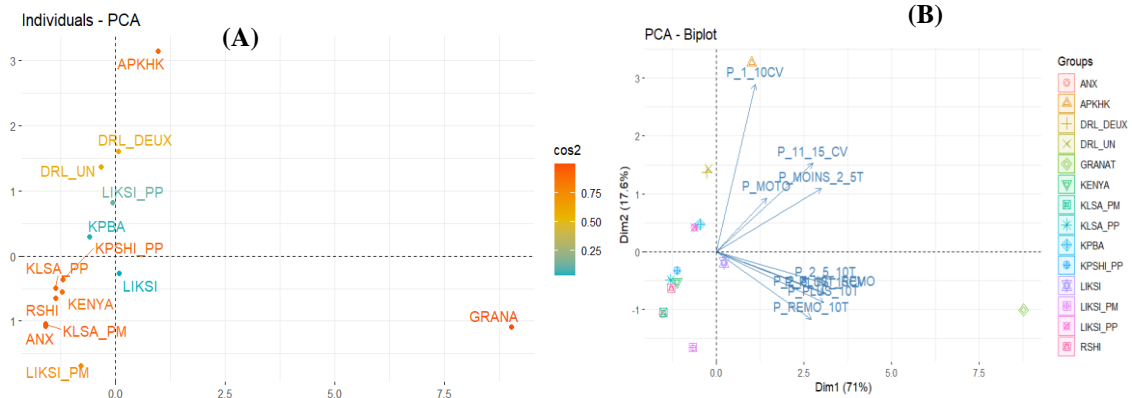


Figure 6—Cercle de corrélation des individus et variables
Source : Auteurs

En ce qui concerne le paiement de pénalités, il ressort de ce graphique que les contribuables les plus exemplaires sont ceux du bâtiment GRANA dans la première dimension et détiennent les véhicules ayant une puissance fiscale de plus de 10 CV au remorques, par contre la ville de APHK, la DRL_UN et DEUX sont bien corrélés dans la 2^{ème} dimension ; la majorité dispose de petit véhicule (1 à 10 CV). On peut extrapoler que la commune KENYA, RUASHI, KIPUSHI, KASUMBALESA et la commune annexe se trouvent au centre de gravité c'est-à-dire paient moins de pénalités. En général, ces communes sont peuplées par la classe moyenne nécessitant une campagne de marketing et un recouvrement musclés.

Tableau 3—Tableau de contribution à l'inertie de deux premières composantes

	Dim.1	Dim.2
P_11_15_CV	0.8615617	0.4765874
P_PLUS_15CV	0.9672817	-0.2278568
P_PLUS_10T	0.9590163	-0.2744236
P_REMO_10T	0.8465479	-0.3635417
P_1_10CV	0.3478320	0.9042942
P_2_5_10T	0.9746039	-0.1718572
P_2_5_10T_REMO	0.9632941	-0.2217146
P_MOINS_2_5T	0.9357955	0.3393915
P_MOTO	0.4499147	0.2881656

Source : Auteurs

Dans le tableau de l'inertie (Tableau 3), la 1^{ère} composante (Dim.1), contribuant avec 71 % voire le cercle de corrélation Figure 5 (B) d'inertie, est défini par les véhicules à puissance fiscale élevé à savoir la P_2_5_10T (0,974), les P_2_5_10t_REMO (0,963), les P_MOINS_2_5_T (0,935), la MOTO (0,441), les P_1_10_CV (0,347) et les P_11_15_CV (0,861). Avec une inertie de 17,6 % la 2^{ème} dimension (Dim.2) est définie par les véhicules ayant la puissance fiscale : P_1_10_CV (0,904), les P_11_15_CV (476), faiblement par les véhicules de moins de 2.5 tonnes (339), les motos (228), du côté négatif et les véhicules de plus de 15_CV (-0,227), les véhicules poids lourds de plus de 2.5 tonnes (-0.363), et les autres de (-0.227). L'examen de la matrice de corrélation entre catégories de véhicules (Figure 2) révèle la présence d'une intimité de la culture fiscale, constitué de variables bien corrélées entre elles. Il s'agit de :

- Véhicules légers : de 1 à 10, de 11 à 15 chevaux, de moins de 2.5 tonnes et de motos dont la majorité est détenue par de particuliers
- Les véhicules poids lourds de 2.5 à 10 tonnes, de plus tonnes, ainsi que de remorques dont la majorité appartient aux personnes morales.

Ainsi, nous pouvons dire que les assujettis ayant de petits véhicules (particuliers) préfèrent payer l'impôt en cas de forçing (recouvrement forcé) d'où la nécessité de faire une sensibilisation ciblée au civisme fiscal au particuliers pour la maximisation de recettes. Du côté personne morale, la DRHKAT arrive à récupérer l'impôt bien que tardif.

Conclusion

Cet article visé l'évaluation de Cet article visé l'évaluation de la solvabilité de contribuable à l'impôt sur les véhicules. Au des résultats à l'issu de l'examen de groupes de contribuables et suivant la puissance fiscale de véhicules recueillies dans les 13 bureaux de la DRHKAT, on constate un déficit du civisme fiscal de la part des contribuables (personnes physiques) de la commune Annexe, Ruashi, Kenya, Kampemba, Kipushi, Kasumbalesa. Néanmoins incivisme fiscal à tendance à diminuer dans les bureaux du centre-ville notamment DRL1, DRL2 et à l'APHK où les particuliers sont obligés de conduire en toute période. Mieux encore, les résultats révèlent la solvabilité du deuxième groupe de contribuables (personnes morales) qui arrivent à s'acquitter de l'impôt tardivement. Ayant fait l'objet d'un examen approfondi en Composantes Principales, ces résultats ont fait l'objet d'une interprétation sur les corrélations existantes entre les différents contribuables et entre les véhicules de différentes puissances

fiscales issus de différentes entités administratives. L'Analyse en Composantes Principales sur la solvabilité de contribuables nous a permis de différencier une solvabilité et civisme fiscal intrinsèque et relatif selon la puissance fiscale du véhicule et la commune (bureau) d'habitation. La solvabilité de contribuables varie d'un bureau et du type de véhicule à l'autre en fonction de la puissance fiscale et de la sensibilisation. En outre, l'ACP a rendu possible la découverte de connaissances cachées par la description de l'accroissement de pénalité suivant en gradient à savoir un gradient de la puissance fiscale croissante vers le tonnage faisons croire la solvabilité de l'impôt sur les véhicules ; Un gradient d'insolvabilité fiscale décroissant au prorata de la puissance fiscale et de la commune de résidence.

Cette étude constitue une interpellation de chaque entité pour la sensibilisation à la culture fiscale et fourni un bon exemple l'analyse de paramètres de maximisations de recettes et de la planification budgétaires dans les entités décentralisées en RDC. Bon nombre de contribuables arrivent à s'acquitter de l'impôt selon qu'on est personne morale ou physique d'une part et la puissance fiscale ainsi que l'usage qu'on fait du véhicule. Par ailleurs, en dehors du bureau du centre-ville (GRANAT) où est concentré la quasi-totalité des entreprises, la solvabilité la diminue progressivement plus on s'éloigne de la ville de Lubumbashi. On constate que les contribuables ayant de motos mérite une campagne ciblée ou de lois fortes pour l'impulsion fiscale.

REFERENCES

- [1] E. M. Kazadi, « Apport et stratégies de recouvrement de l'impôt sur les véhicules dans le budget de l'Etat Congolais. », *KAS*, vol. 8, n° 4, p. 419-438, 2021, doi: 10.5771/2363-6262-2021-4-419.
- [2] M. Kamba-Kibatshi, « L'influence des petites et moyennes entreprises au développement économique de la République Démocratique du Congo », *Nierówności spo\leczne a wzrost gospodarczy*, n° 46, p. 478-505, 2016.
- [3] P. B. S. Diouf, « La business intelligence comme moteur de réussite des projet en période de crise », PhD Thesis, Université du Québec à Trois-Rivières, 2023. Consulté le: 12 novembre 2023. [En ligne]. Disponible à: <https://depot-e.uqtr.ca/id/eprint/10781/1/eprint10781.pdf>
- [4] T. Stéphane, *Data mining et statistique décisionnelle: l'intelligence des données*. Editions Technip, 2012. Consulté le: 12 novembre 2023. [En ligne]. Disponible à: https://books.google.com/books?hl=fr&lr=&id=Cs2fCgAAQBAJ&oi=fnd&pg=PR5&dq=Data+mining+dans+le+domaine+bancaires&ots=MLtJ7tuY0F&sig=-9X1pF3HiYO8J_vWbUTiWOZ2dMY

- [5] C. Duby et S. Robin, « Analyse en Composantes Principales ».
- [6] R. He, B.-G. Hu, W.-S. Zheng, et X.-W. Kong, « Robust principal component analysis based on maximum correntropy criterion », *IEEE Transactions on Image Processing*, vol. 20, n° 6, p. 1485-1494, 2011.
- [7] J.-L. Berger, « Analyse factorielle exploratoire et analyse en composantes principales: guide pratique », 2021, Consulté le: 12 novembre 2023. [En ligne]. Disponible à: <https://hal.science/hal-03436771/document>
- [8] F. Husson et S. Lê, « Analyse en composantes principales (ACP) », *Anal. données avec R*, p. 27-28, 2016.
- [9] D. Desbois, « Application de l'ACP et de la classification par intervalles aux estimations de quantiles des distributions empiriques du coût des engrais pour les cultures annuelles dans les pays de l'UE », dans *14ièmes Journées de Recherches en Sciences sociales*, 2021. Consulté le: 14 janvier 2024. [En ligne]. Disponible à: <https://hal.science/hal-03193404/>
- [10] M. C. Yamina, B. E. Habib, D. Mansour, A. Khaldia, et A. Rachida, « APPLICATION DE L'ACP À UNE BASE DE DONNÉES DE LIQUIDES IONIQUES », 2016.
- [11] B. Nassima, « Performance et secteur d'activité des PME de la wilaya d'Oran: une analyse statistique par l'ACP Performance and activity's sector of SMEs in the province of Oran: a statistical analysis by the ACP », Consulté le: 16 janvier 2024. [En ligne]. Disponible à: <https://www.asjp.cerist.dz/en/downArticle/299/9/1/162372>
- [12] I. Chaouki, L. Mouhir, M. Fekhaoui, A. El Abidi, A. Bellaouchou, et M. El Morhit, « Application de l'analyse en composantes principales (ACP) pour l'évaluation de la qualité des eaux usées industrielles de Salam Gaz-Skhirat », *Journal of Materials and Environmental Science*, vol. 6, n° 2, p. 455-464, 2015.
- [13] S. Ayimpam, B. K. OKITO, et M. B. KIBUL, « La gouvernance au quotidien des taximotos à Kinshasa. Une régulation difficile », *Géotransports*, vol. 2022, n° 17-18, p. 135-149, 2023.
- [14] G. Bothorel, M. Serrurier, et C. Hurter, « Utilisation d'outils de visual data mining pour l'exploration d'un ensemble de règles d'association », dans *Proceedings of the 23rd Conference on l'Interaction Homme-Machine*, Sophia Antipolis France: ACM, oct. 2011, p. 1-4. doi: 10.1145/2044354.2044369.
- [15] J. Silva, N. Varela, L. A. B. López, et R. H. R. Millán, « Association rules extraction for customer segmentation in the SMEs sector using the apriori algorithm », *Procedia Computer Science*, vol. 151, p. 1207-1212, 2019.
- [16] M. Massoud, « La réduction de la biométrie cérébelleuse fœtale: modèle précoce de l'hypoplasie cérébelleuse de diagnostic postnatal? », PhD Thesis, Université Claude Bernard-Lyon I, 2022. Consulté le: 28 septembre 2023. [En ligne]. Disponible à: <https://theses.hal.science/tel-04209088/>
- [17] M. Yaosuwa et I. Gaya, « La Fiscalite Congolaise Face Aux Enjeux De La Decentralisation: Cas De La Province De L'ituri De 2016 A 2018 », 2019, Consulté le: 13 novembre 2023. [En ligne]. Disponible à: <http://scholarsmepub.com/wp-content/uploads/2019/11/SB-511-624-634.pdf>

- [18] J. C. RATSIMAVO, A. A. RATIARISON, O. RAMIARINJANAHARY, J. E. RANDRIANANTENAINA, N. A. RAKOTOVAO, et M. HAJARISOA, « Impact des indices climatiques (AAO, IOD et ENSO) sur la prévision de la sécheresse dans toute la partie sud de Madagascar », *Afrique SCIENCE*, vol. 13, n° 6, p. 58-75, 2017.
- [19] I. BOUSRI, A. BOUCETTA, et S. SAHABI-ABED, « Régionalisation des normales annuelles des températures en Algérie par la méthode K-means. », *JAMA*, vol. 6, p. 40-46, 2022.