

Mise en place d'un score géomarketing de potentiel bancaire dans les villes françaises de plus de 15 000 habitants

Tuteur : Nicolas Telmon

Contexte et objectifs.

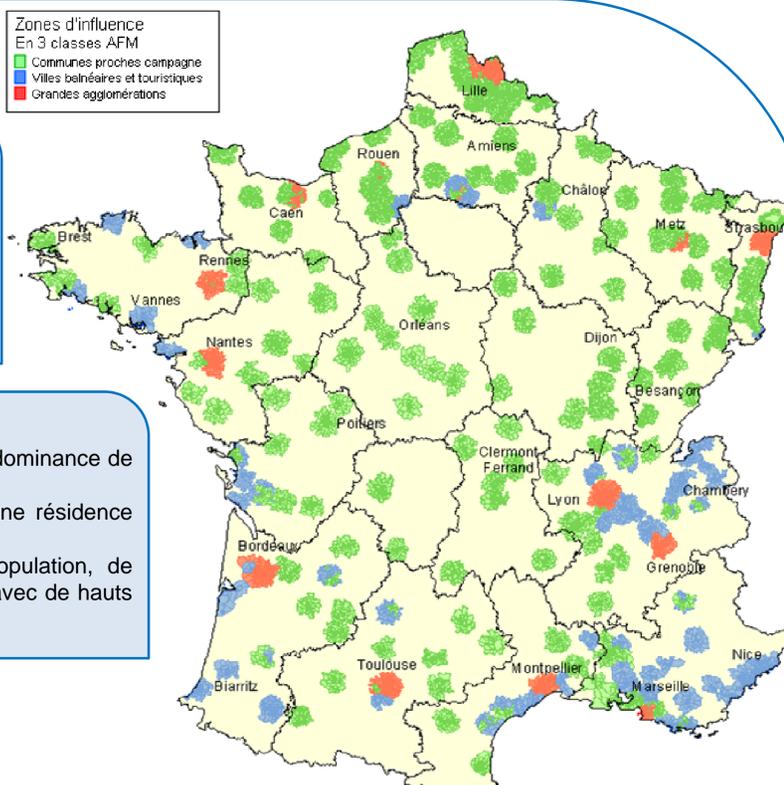
Renforcer sa présence sur le territoire national permet de faire face à la concurrence accrue du système bancaire. C'est dans cette optique que se situe ce travail. Il s'agit donc de mener une étude de marché portant sur la population française et visant à donner des perspectives de croissance du bénéfice selon l'implantation géographique des agences de la Banque Postale. Il faudra d'abord établir les profils des agences actuellement implantées à partir de leur caractéristiques, de leur environnement géoéconomique et sociodémographique. Ensuite, nous tenterons de définir les facteurs clés de la rentabilité des agences.



Typographie des agences

Une approche via l'Analyse Factorielle Multiple (AFM) :

Principe de l'Analyse Factorielle Multiple : permet de réaliser une analyse factorielle dans laquelle l'influence des groupes de variables est équilibrée (voir le tableau ci-dessous). Ainsi, même si l'on a deux fois plus de variables sur la richesse de la population que sur les caractéristiques du logement, ces deux groupes auront le même poids dans la constitution des classes. Cet équilibre est nécessaire, car les groupes diffèrent toujours selon le nombre ou la structure des variables.



L'AFM nous a permis de distinguer 3 classes :

- Villes proches de la **campagne** : familles à revenus moyens, prédominance de l'emploi primaire et secondaire.
- Villes **touristiques** : habitant assez riche, retraité, possédant une résidence secondaire. Proche de littoraux ou de stations de skis.
- **Centres-villes** de grandes agglomérations : beaucoup de population, de commerces de tous types et d'agences postales, individus qualifiés avec de hauts revenus et un fort patrimoine.

La classification obtenue est sensiblement la même que celle issue de l'étude préliminaire où une ACP fut mise en œuvre. La différence remarquable : la classe des villes touristiques comprend à présent des agglomérations des Alpes (Isère, Haute-Savoie), où les stations de ski sont, à l'instar des villes balnéaires, de hauts lieux du tourisme de masse. Ci-dessous, un tableau détaille par groupe les caractéristiques des classes obtenues suite à l'AFM.

Carte de la partition en trois classes issue de la Classification Ascendante Hiérarchique fondée sur les résultats de l'AFM.

Groupes de variables (effectif)	Communes proches de la campagne (827)	Villes balnéaires et touristiques (256)	Grandes agglomérations (437)
Groupe 1 : Population	plus de 40 ans, secteur secondaire (ouvriers, employés)	retraité, plus de 65 ans, secteur primaire et tertiaire	population jeune, cadres et professions intermédiaires
Groupe 2 : Habitat	peu peuplé, logements vacants et résidences principales	plus de résidences secondaires, moins de résidences principales	très peuplé, peu de logements vacants, beaucoup de résidences principales
Groupe 4 : Indicateurs de richesse	population relativement moins riche	population assez riche	population assez riche
Groupe 5 : Densité de la concurrence	peu d'agences concurrentes, hormis Crédit Agricole et Caisse d'Epargne	Banques citadines (Société Générale, BNP, banques virtuelles...)	beaucoup d'agences concurrentes, et fort taux de pénétration de la Poste
Groupe 6 : Densité des commerces	faible	plutôt faible	très forte
Groupe 7 : Caractéristiques des agences	peu de conseiller mais de nombreuses ventes, les agences sont éloignées	beaucoup de conseillers, des agences éloignées et peu de distributeurs	beaucoup de conseillers mais peu de ventes, de retraités, d'automates
Groupe 8 : Caractéristiques des distributeurs et guichets à billets	beaucoup de distributeurs présents en agences ou à proximité de commerce	peu de distributeurs en agences ou à proximité de gare, aéroport ou aire d'autoroute	un grand nombre de distributeurs concurrents
Groupe 9 : bénéfice de l'agence (illustratif)	plutôt faible	moyennement élevé	plutôt élevé

Modélisation

L'objectif de cette étude est également de modéliser le potentiel bancaire qu'une agence peut espérer dans une certaine zone géographique. Pour cela, nous avons testé différents modèles afin de voir lequel prédirait au mieux le bénéfice d'une agence.

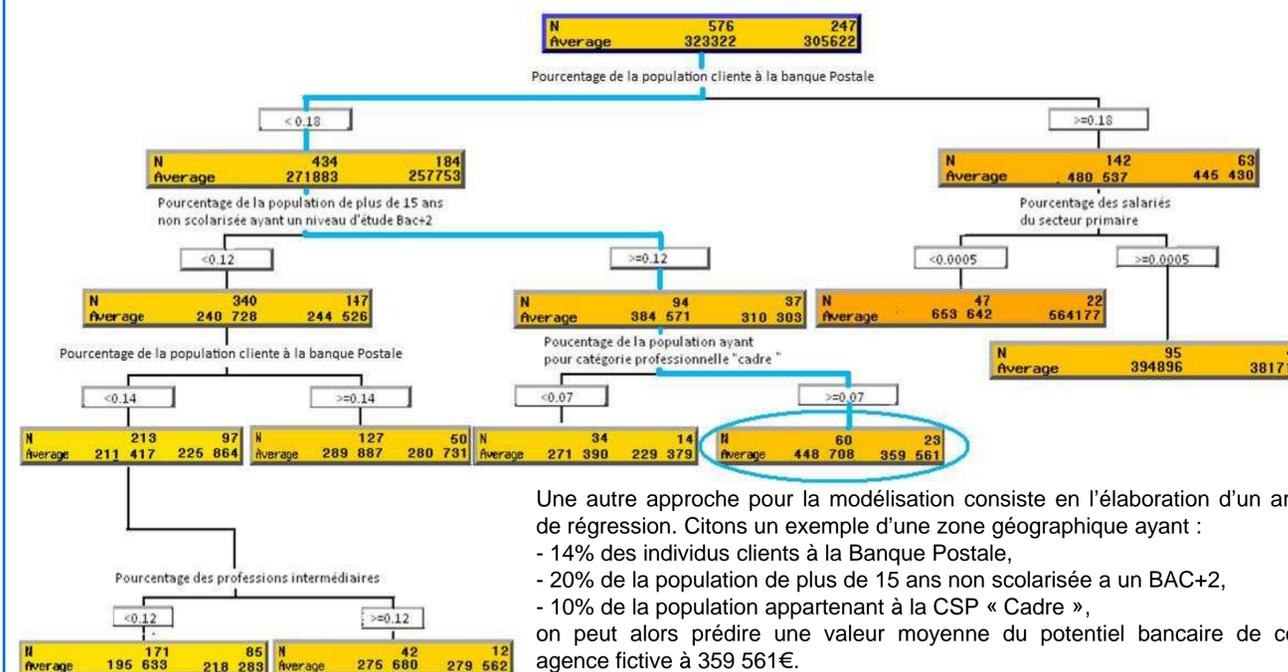
Au terme de ces tests, le critère du plus grand R² ajusté s'est révélé être le meilleur critère de sélection de modèle afin de minimiser l'erreur de prédiction (RMSE – Root Mean Square Error).

Néanmoins, les résultats obtenus ne peuvent être considérés comme satisfaisants. En effet, la variable à expliquer (profit d'une agence) ne suit pas une loi normale, et les liens entre cette variable et ses différents régresseurs ne sont modélisables par aucune fonction connue. Mais cela n'est pas en contradiction avec le fait que l'action simultanée de ces variables soit modélisable par une régression linéaire multiple. Cependant, l'analyse des résidus de nos modèles pose problème, ces derniers étant de variances différentes et non-gaussiens.

Le modèle sélectionné au critère du plus grand R² ajusté s'est donc révélé être le meilleur modèle en terme de prédiction du potentiel bancaire d'une agence.

Modèles Critères	Sélection Forward	Sélection Stepwise	Sélection Backward	Plus grand R ² ajusté	Bootstrap
R ² ajusté	0.233	0.233	0.2373	0.2501	0.2442
RMSE	202021	202021	201462	199758	200538

Bootstrap : le modèle « Bootstrap » consiste en une sélection de variable après régressions sur 1000 échantillons Bootstrap (individus tirés aléatoirement avec remise). Les variables sélectionnées sont celles étant apparues le plus de fois de façon significative dans les différents modèles.



Une autre approche pour la modélisation consiste en l'élaboration d'un arbre de régression. Citons un exemple d'une zone géographique ayant :

- 14% des individus clients à la Banque Postale,
- 20% de la population de plus de 15 ans non scolarisée à un BAC+2,
- 10% de la population appartenant à la CSP « Cadre »,

on peut alors prédire une valeur moyenne du potentiel bancaire de cette agence fictive à 359 561€.

Pour aller plus loin ...

Les modèles obtenus n'étant pas entièrement satisfaisants, d'autres méthodes sont envisageables :

- utiliser la régression non paramétrique
- faire une régression à l'intérieur de chaque classe issue de la classification ascendante hiérarchique, afin d'avoir un système de notation propre à chaque type de villes.