

# Application du data mining à la segmentation du marché des meubles aux États-Unis

THI THU HOA LE<sup>1</sup>, BRUNO AGARD<sup>1</sup>, STÉPHANE DEVEAULT<sup>2</sup>

<sup>1</sup> ÉCOLE POLYTECHNIQUE DE MONTRÉAL  
C.P. 6079, succ. Centre-ville, Montréal (Québec), H3C 3A7, Canada  
[hoa.le@polymtl.ca](mailto:hoa.le@polymtl.ca); [bruno.agard@polymtl.ca](mailto:bruno.agard@polymtl.ca)

<sup>2</sup> CANADEL FURNITURE  
700, av. Canadel, P.O. Box 5, Louiseville (Québec), J5V 2L6, Canada  
[StephaneDeveault@canadel.ca](mailto:StephaneDeveault@canadel.ca)

---

## Résumé

La découverte des marchés potentiels est la clé du succès pour le développement d'une entreprise dans le contexte actuel de globalisation de l'économie. Différentes techniques de marketing et d'analyse des données appliquent à la segmentation du marché en se basant sur les caractéristiques socioéconomiques des clients dans différentes zones géographiques. Ces données sont ensuite croisées avec des historiques de ventes.

Dans cet article, nous appliquons les techniques de segmentation pour identifier des groupes de consommateurs potentiels sur le marché des meubles haut de gamme aux États-Unis par l'analyse de données secondaires. Le résultat obtenu, sur un cas d'étude réel, sera intégré par la suite aux données historiques de ventes afin de donner une vue concrète et graphique des ventes réelles dans chaque segment. Cela permettra de prendre les décisions pertinentes pour augmenter les ventes dans les marchés identifiés.

## Abstract

Determining the potential markets is a successful key for the development of any company in changing and globalization economy contexts. Several data analysis techniques are applied for segmenting a market, especially based on the typical social-economic factors in different geographical areas, and then crossed with historical sales data.

In this article, we will apply clustering techniques to identify groups of prospective buyers of the high-end furniture market in the United State by analysing the secondary data. These practical results, going along with the reported sales data, will give the company a concrete and visual demonstration of sales in each segment. This will contribute to assist decision makers to expanding sales to their targets market.

**Mots clés – segmentation du marché, clustering, système d'aide à la décision, marketing, data mining.**

**Keywords – market segmentation, clustering, decision support system, marketing, data mining.**

---

## 1 INTRODUCTION

La segmentation du marché est définie la première fois en 1956 par Smith (Smith, 1956). La définition de la segmentation est simple mais elle demeure un élément essentiel du marketing. Elle permet aux entreprises de bien comprendre leurs clients et de se concentrer sur leurs besoins. Selon Croft «la segmentation du marché est le processus d'identifier différents groupes d'utilisateurs dans un marché qui pourrait probablement être visé avec les produits séparés ou les programmes de marketing» (Croft, 1994).

De nos jours, avec les changements du marché, le marketing évolue du marketing de masse vers le marketing direct. Cela accentue l'importance de la segmentation dans la stratégie marketing d'une entreprise. L'entreprise concentre maintenant son attention vers les petits segments de marché avec une approche de marketing direct. La segmentation permet d'identifier certaines caractéristiques du marché pour le développement de nouveaux produits.

L'approche traditionnelle de la segmentation du marché est basée sur les sondages. La collecte des données dans la plupart des cas est très coûteuse et demande beaucoup de temps.

D'autre part, le résultat de la segmentation est parfois difficile à analyser pour la prise de décision. Dans cet article, le premier objectif est d'aider une entreprise à identifier les marchés potentiels en utilisant les données secondaires sociodémographiques du territoire où s'effectue l'essentiel de ses ventes. Le second objectif consiste à proposer un modèle intégrant le résultat de la segmentation aux données de ventes réelles afin de traduire le résultat trouvé en stratégie.

Dans une première partie, le contexte du marché des meubles aux États-Unis et les principales approches de segmentation seront présentés. Le potentiel des données secondaires pour la segmentation sera expliqué par la suite. Le modèle intégrant le résultat de la segmentation aux données de ventes est présenté dans la section de méthodologie. Enfin, la dernière partie abordera l'explication et l'évaluation du résultat à travers la relation entre les groupes obtenus et les ventes réelles de l'entreprise.

## 2 REVUE DE LA LITTÉRATURE

### 2.1 Contexte du marché des meubles aux États-Unis

Les États-Unis sont le plus grand marché des meubles au monde en plus d'être le plus concurrentiel avec plus de 120 pays exportateurs (Florence, 1990).

Selon les données du Bureau du recensement des États-Unis (Census Bureau), en 2007, la valeur des importations des meubles aux États-Unis s'élevait à 27,24 milliards de dollars. Il s'agit d'une hausse de 154% par rapport à 1998.

En 1998, le Canada était le premier fournisseur, suivi par la Chine et le Mexique. Depuis 2000, la part canadienne du marché des meubles américain a chuté, tandis que la part asiatique (principalement la Chine) a augmenté rapidement. En 2006, la Chine a surpassé le Canada en tant que premier exportateur de meubles vers les États-Unis (50%); le Canada prend alors la deuxième place avec une part de 16%, Figure 1 (Murillo, 2007).

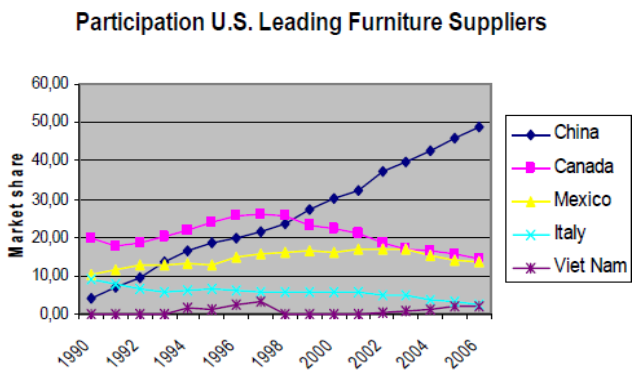


Figure 1. Participation des grands fournisseurs de fourniture aux États-Unis

Cela met les entreprises de meuble du Canada dans un contexte difficile qui nécessite de trouver des stratégies pertinentes pour qu'elles reprennent plus de place dans le marché américain.

## 2.2 Revue des méthodes de la segmentation du marché

Il existe différents outils pour la segmentation appliquée au marketing. Ces méthodes peuvent utiliser différents type de données (données primaires ou secondaire) et une multitude de format (discret, continue, numérique, textuel...).

(Wedel & Kamakura, 1998) ont classé les méthodes de segmentation en quatre groupes, voir aussi (Green, 1977) et (Wind, 1978), comme représenté dans le tableau 1.

	A priori	Post hoc
<b>Descriptive</b>	Tableau de contingence. Modèle log-linéaire.	Méthode de clustering.
<b>Prédictive</b>	Tableau croisés. Régression.	AID, CART, ANN, Modèle de mélange.

Tableau 1. Classification des méthodes de la segmentation

Les sections suivantes vont détailler les éléments essentiels de cette classification.

### 2.2.1 L'approche a-priori

Dans l'approche « a priori », le marché est segmenté selon des critères préexistants tels que l'âge, le sexe ou le statut économique et social. Les méthodes ont alors pour objectif de découvrir ou de décrire les caractéristiques des clients dans les segments déjà connus. Les méthodes : modèle log-linéaire, tableau de contingence (Pearson, 1904), les tableaux croisés et

la régression sont de type « a-priori ». Cette approche est très efficace dans le cas où les variables d'analyse sont bien définies. Cependant, il est parfois difficile d'identifier les variables à utiliser. Les méthodes « a-priori » sont donc utilisées soit dans des cas particuliers, quand on dispose a-priori d'une certaine connaissance du domaine, soit en combinaison avec des méthodes « post-hoc » (hybrid segmentation).

Par exemple, (Umesh, 1987) a appliqué le modèle de régression pour expliquer les préférences de transports des consommateurs pour aller au centre d'achats.

### 2.2.2 L'approche post-hoc :

Les méthodes de segmentation sont les plus utilisées dans l'approche post-hoc car elles sont faciles à mettre en œuvre et peuvent s'appliquer à divers types de données. La plupart des techniques de segmentation se divisent dans une des deux catégories principales suivantes : les méthodes hiérarchiques et les méthodes par partitionnement.

L'idée principale de ces méthodes est basée sur une mesure de similarité (ou distance) pour grouper les données. Le but de la segmentation est de maximiser la similarité intra-classes et de minimiser la similarité inter-classes. La distance entre les objets peut être calculée par une des méthodes ci-dessous, dépendamment de types des données :

- Distance euclidienne de deux points en n-dimension. C'est la distance la plus populaire dans les méthodes de segmentation :

$$d_2(x_i, x_j) = \left( \sum_{k=1}^n (x_{ik} - x_{jk})^2 \right)^{1/2}$$

- Distance de Minkowsky : (l'extension de la distance euclidienne)

$$d_p(x_i, x_j) = \left( \sum_{k=1}^n |x_{ik} - x_{jk}|^p \right)^{1/p}$$

- Distance Manhattan :

$$d_M(x_i, x_j) = (x_i - x_j) \sum^{-1} (x_i - x_j)^T$$

La distance utilisée dépendra du type de données disponible, de son format et de sa précision de la base de données.

- **Les méthodes hiérarchiques :**

Les méthodes de segmentation hiérarchiques sont récursives et peuvent être soit de type par agglomération soit de type par division.

Les méthodes par agglomération partent d'une partition où chaque donnée représente un segment et à chaque itération les deux segments les plus proches sont fusionnés jusqu'à ce que tous les points se trouvent dans un seul grand segment.

Les méthodes par division créent une hiérarchie descendante en procédant par divisions successives. La distance entre deux segments peut être donnée par la distance entre les deux membres les plus proches, les plus éloignés ou entre leur centroïdes.

Les méthodes hiérarchiques retournent le résultat sous la forme d'un dendrogramme, qui non seulement indique les objets et les segments à chaque itération, mais aussi la valeur du critère choisi pour chaque partition rencontrée. Cela permet de déterminer le nombre de segment à utiliser dépendamment de

l'objectif poursuivit. Cependant, ces méthodes sont difficilement applicables avec de grandes bases de données. Dans la segmentation du marché, (Greeno, Summers, & Kernan, 1973) ont utilisé la méthode hiérarchique pour identifier les segments en se basant sur les caractéristiques personnelles et comportementales des consommateurs.

- **Les méthodes par partitionnement**

Dans ces méthodes, le nombre de segments voulu (k) doit être connu, il peut être déterminé par des experts ou par itérations successives (en croisant différentes méthodes de segmentation).

Ces méthodes produisent les segments autour de noyaux choisis initialement, puis elles améliorent itérativement ces segments initiaux en se basant sur une fonction de coût à minimiser. Elles restent efficaces avec de grandes bases de données et des types de données variés : données énumératives, numériques et textuelles.

Néanmoins, ces méthodes rencontrent deux problèmes majeurs :

- Il est souvent difficile de déterminer le nombre de segments optimal (k) et le résultat dépend de l'initialisation (les noyaux de départ).
- Ces méthodes peuvent converger vers un optimum local.

Dans leur revue de littérature, Punj et Stewart (1983) ont conclu que les méthodes par partitionnement sont plus efficaces que les méthodes hiérarchiques dans la segmentation du marché.

La méthode des K-moyennes (Mc Queen, 1967) est la plus connue et la plus couramment utilisée dans la pratique. D'autres méthodes peuvent aussi être utilisées tels que : K-medoids, CLARANS, EM-clustering,...

Des extensions pour les méthodes par partitionnement sont proposées. (DeSarbo, Carrol, & Clark, 1984) ont proposé une méthode en considérant la pondération des variables. Les points indiquent l'importance de chaque variable. Cependant, cette méthode est difficile à implémenter. (Dolnicar & Leisch, 2004) ont intégré la méthode hiérarchique à la méthode k-moyenne (bagged clustering) pour améliorer la stabilité et l'interopérabilité des segments concernant les données binaires. D'autres améliorations concernent le choix du point de départ et/ou la détermination du nombre de cluster. (Maulika & Bandyopadhyay, 2000) ont appliqué l'algorithme génétique pour choisir les noyaux au départ.

Il existe beaucoup de cas d'applications de ces méthodes en segmentation du marché. Nous pouvons citer ici quelques exemples. En utilisant la méthode bagged clustering, (Dolnicar & Leisch, 2004) a identifié cinq classes de touristes différents dans le marché du tourisme en Australie. (H.Sung, 2004) a utilisé la méthode de k-moyenne pour identifier six groupes de voyageurs d'aventure aux États-Unis afin de déterminer les marchés potentiels. (B. Agard, C. Morency & M. Trépanier, 2006) ont appliqué la méthode k-moyenne couplée à un algorithme hiérarchique pour déterminer les comportements des clients en transport publique.

### 2.3 L'importance des données secondaires en segmentation

Les données utilisées pour la segmentation du marché peuvent être collectées à partir de soit source primaire, soit à partir de source secondaire.

Les données primaires sont des données recueillies dans un but de recherche précis. Les données secondaires sont des données collectées non pas pour les objectifs de la recherche en cours, mais pour d'autres objectifs.

Normalement, les données primaires ne sont pas disponibles telles quel au moment du besoin et nécessitent un effort particulier de collecte par des sondages, des entrevues, etc. En général, les données secondaires sont déjà disponibles parce qu'elles sont produites pour d'autres objectifs (Castleberry, 2001). Ces données peuvent être fournies par des moyens externes (source gouvernementales) ou par des moyens internes comme les données historiques de l'entreprise (les données de facturations, les données de ventes,...) (Patze, 1995) et (Aaker, V.Kumar, & S.Day, 2007) ont souligné l'importance des sources de données secondaires dans la recherche en marketing. Ils citent les avantages de l'utilisation des sources secondaires par rapport aux données primaires :

- La collecte des données secondaires est moins coûteuse que celle des données primaires.
- Le temps nécessaire à la collecte des données secondaires est de loin inférieur à celui requis pour les données primaires.
- Dans certains cas, les sources secondaires peuvent fournir des données plus précises que celles obtenues par la recherche primaire. Les données statistiques sur la population fournies par les gouvernements sont plus fiables que celles d'autres sources similaires.
- Les données secondaires sont importantes pour définir le problème et établir des hypothèses avant de collecter des données primaires. Elles peuvent aussi aider à caractériser la population d'un marché.

### 2.4 Conclusion :

Le choix de la source des données est basé sur la disponibilité des données et l'objectif de l'entreprise et doit se faire attentivement avec les spécialistes du domaine concerné. La méthode appropriée est ensuite déterminée en se basant sur les caractéristiques des données et l'objectif de la segmentation.

## 3 MÉTHODOLOGIE

### 3.1 Contexte

Sous la pression de la globalisation, la concurrence est de plus en plus féroce. La demande évolue et devient de plus en plus segmentée (Kaynak & Hassan, 1994). Cela présente de nouveaux défis à relever et de nouvelles voies d'amélioration à explorer pour la conception, la production et la commercialisation de produits adaptés au marché. La segmentation du marché est alors un élément essentiel qui va permettre de définir les éléments clés nécessaires pour définir les besoins à satisfaire. Avec plus de 300 millions d'habitants, un territoire vaste et des cultures diverses, le marché des États-Unis est un défi pour l'analyse commerciale. À partir d'un cas d'étude développé en collaboration avec une entreprise qui conçoit et fabrique des meubles haut de gamme, commercialisés en grande majorité sur le territoire des États-Unis, nous montrons comment la segmentation, basée sur des données secondaires peut enrichir les connaissances nécessaires au développement de l'entreprise.

Dans le cas du marché des meubles aux États-Unis, la plupart des entreprises de fabrication de meubles n'ont pas leurs propres magasins. Elles doivent, par conséquent, recourir un canal de distribution passant par diverses chaînes de magasins. Ceci induit une complexité élevée et des coûts prohibitifs pour obtenir des informations sur les clients finaux. En faisant l'hypothèse que les caractéristiques démographiques, sociales et économiques sont des facteurs influençant sur le marché des meubles haut de gamme, nous analysons ces données pour obtenir une segmentation adaptée au marché du meuble haut de gamme. Les données statistiques fournies par le Bureau du recensement des États-Unis (Census Bureau) seront exploitées.

Le Census Bureau procède au recensement de la population américaine. Le recensement inclut des informations démographiques (l'âge, l'éducation, la composition du foyer, l'occupation,...) et économiques (le revenu, l'endettement, ...). Les données fournies par le Census Bureau constituent une source potentielle et fiable pour les chercheurs du marché Américains.

Pour la segmentation du marché des meubles aux États-Unis, nous avons tout d'abord examiné les caractéristiques des données pour choisir la méthode appropriée de segmentation. Le résultat obtenu a été intégré par la suite aux données historiques de ventes afin de fournir à l'entreprise une vue concrète et graphique de ses ventes dans chaque groupe.

### 3.2 Les méthodes et outils du Data Mining

Dans notre étude de cas, l'objectif est d'identifier des groupes d'États homogènes d'un point de vue socio-économique. Nous ne connaissons pas encore les critères, ni le nombre de segments, donc par la méthode de segmentation la plus appropriée.

Le nombre de cluster est un paramètre difficile à déterminer dans les méthodes par partitionnement. Pour trouver un nombre raisonnable de cluster, plusieurs méthodes sont proposées telles que : tableaux croisés, tests de permutation et rééchantillonnage. Par contre, ces méthodes ne fonctionnent pas très bien en pratique avec les données volumineuses ou multidimensionnelles (Salvador & Chan, 2004). La combinaison de méthodes hiérarchique et méthodes par partitionnement est une approche efficace pour ce type de données (Singh, 1990).

Dans ce cas, nous avons utilisé la méthode hiérarchique pour déterminer le nombre raisonnable de cluster, puis la méthode k- moyenne pour la détermination des groupes.

### 3.3 Modèle intégrant le résultat de la segmentation avec les chiffres de vente réels

(Berry & Linoff, 1997) ont souligné que le résultat de la détection automatique de cluster doit être intégrée à d'autres méthodes pour en dégager la signification. La plupart des applications de la segmentation du marché s'arrêtent à la fin de la méthode de segmentation, c'est-à-dire à la constitution des groupes homogènes. Cependant, il est difficile de prendre quelque décision à partir de ce résultat partiel.

Nous proposons ici un modèle intégrant le résultat de la segmentation du marché aux données historiques de ventes en vue d'aider une entreprise à prendre des décisions de développement à partir des groupes formés (Figure 2). En

analysant les relations entre les données historiques de ventes et les groupes formés, le modèle aide à identifier les magasins en difficultés et les endroits stratégiques.

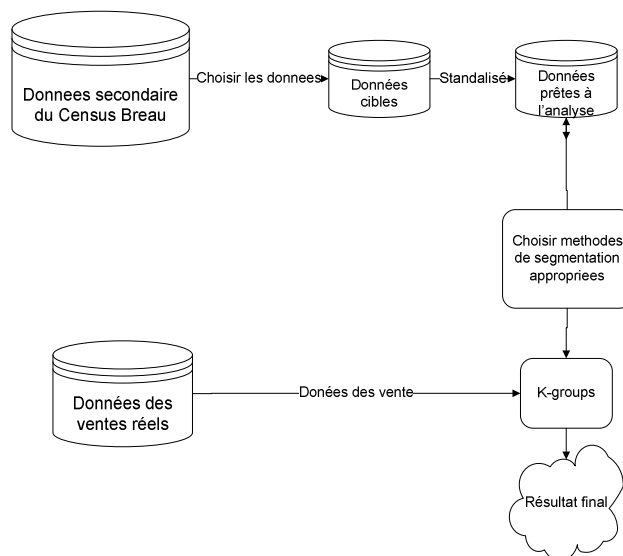


Figure 2. Modèle intégrant le résultat de la segmentation avec les données de ventes réelles.

## 4 CAS D'ÉTUDE

### 4.1 Les données :

Dans la segmentation du marché, le choix des variables doit être guidé par des experts du domaine de l'application.

Dans le cas du marché des meubles aux États-Unis, avec l'aide des experts de l'entreprise concernée, nous avons choisi 60 variables parmi l'ensemble des données disponibles dans les informations collectées par le Census Bureau pour les 50 états à analyser. Ces données ont été prétraitées pour donner une information palpable pour l'utilisateur. Le Tableau 2 ci-dessous montre un aperçu des données considérées.

Etat	Population	% de 35 a 60 ans	% famille en foyer sur nb foyer ....	% population dans la meme maison	Couple marié de 35 à 65 ans	% taux Employée
Alabama	4,442,558	35.29%	68.41% ....	82.70%	32.0%	7.83
Alaska	641,724	37.87%	67.39% ....	80.24%	34.7%	0.53
Arizona	5,829,839	32.02%	66.22% ....	76.76%	28.6%	5.75
Arkansas	2,701,431	34.26%	68.27% ....	80.31%	30.5%	4.94
....	....	....	....	....	....	....
West Virg	1,771,750	37.26%	67.52% ....	87.26%	33.9%	3.35
Wisconsin	5,375,751	36.87%	64.91% ....	84.12%	33.3%	6.92
Wyoming	495,226	37.41%	65.34% ....	80.68%	35.4%	0.66

Tableau 2. Exemple des données.

Les données utilisées dans le tableau 2 concernent la taille de la population, son évolution démographique, le pourcentage de personnes dans certaines tranches d'âges, quelques caractéristiques de logement, de situation familiale, de revenus, et d'emplois.

### 4.2 Analyse sociodémographique

La méthode hiérarchique a été utilisée et la structure hiérarchique montre qu'un partage de la population en trois ou quatre groupes (Figure 3) est approprié.

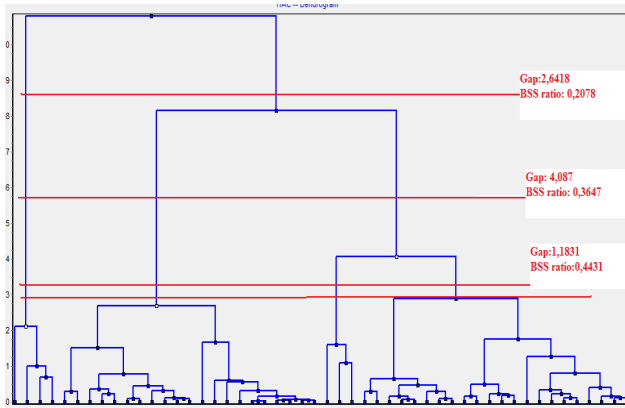


Figure 3 : Résultat de la méthode hiérarchique

Après quelques itérations et confrontations avec les experts de l'entreprise concernés nous sommes arrivés à un partitionnement en trois groupes, qui couvre la totalité du territoire des États-Unis (voir la Figure 4) (l'Alaska est volontairement retiré).

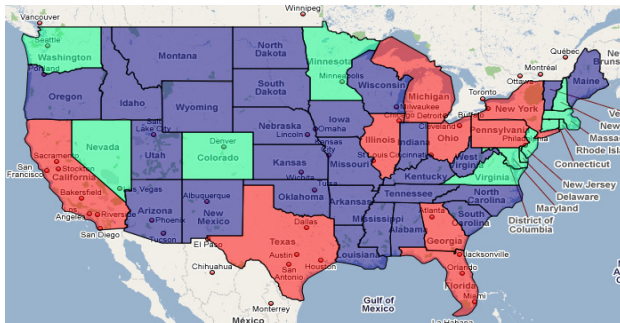


Figure 4. Distribution des segments dans la carte des É-U À partir des données sociodémographiques du Census Bureau, nous arrivons à l'analyse suivante :

➤ **Groupe 1 : « Majorité de revenus faibles »**

Le groupe 1 contient la plupart des états (29 parmi 50 états). Les états dans le groupe 1 sont très étalés sur le territoire, principalement, les terres. De plus, ce sont les états ayant le salaire moyen par personne le plus faible. Le pourcentage de familles avec des salaires élevés est plus faible par rapport la totalité des États-Unis. La majorité des personnes dans ces états gagnent peu d'argent : moins que 29% du salaire moyen max (sur l'ensemble des états) (Figure 5).

**Caractéristiques du Groupe 1**

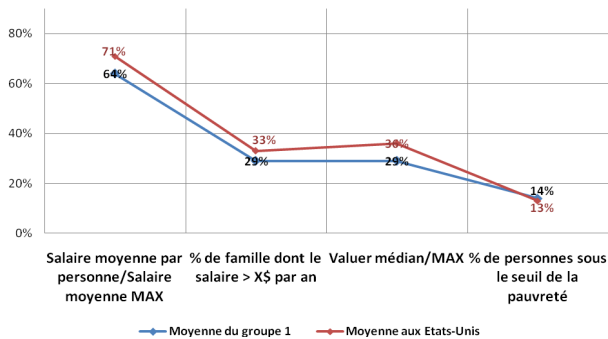


Figure 5. Les moyennes des caractéristiques du groupe 1 par rapport à celles aux États-Unis.

➤ **Groupe 2 : « Majorité de classe moyenne supérieure »**

Le groupe 2 contient 12 états situés principalement dans l'Est. Le pourcentage de familles avec des bons salaires est plus élevé (43% au lieu de 33 %) que sur l'ensemble des États-Unis.

La majorité des personnes dans ces états (>50% - médiane) gagnent plus que dans les autres états et le salaire moyen par personne est plus élevé (Figure 6).

**Caractéristiques du Groupe 2**

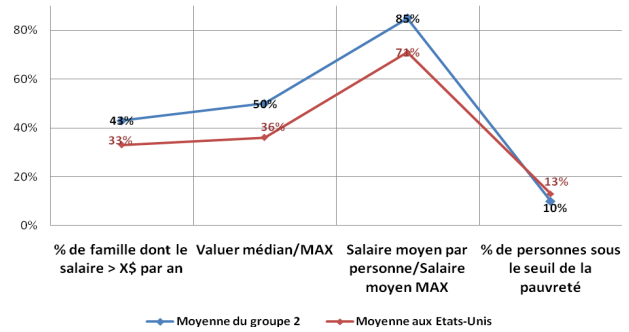


Figure 6. Les moyennes des caractéristiques du groupe 2 par rapport à celles aux États-Unis.

➤ **Groupe 3 : « Majorité de Très riches »**

Le groupe 3 est le groupe le plus riche. Avec 9 états qui se situent principalement autour des Grands lacs avec la Floride et la Californie, il comporte beaucoup de foyers sans hypothèque et énormément de très grandes maisons. De plus, le nombre de foyer avec un bon revenu est supérieur à la moyenne aux États-Unis (Figure 7).

**Caractéristiques du Groupe 3**

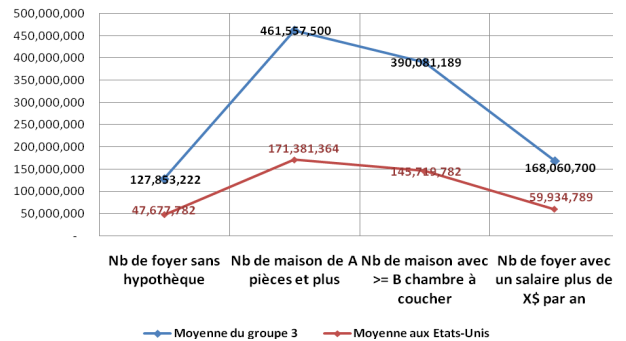
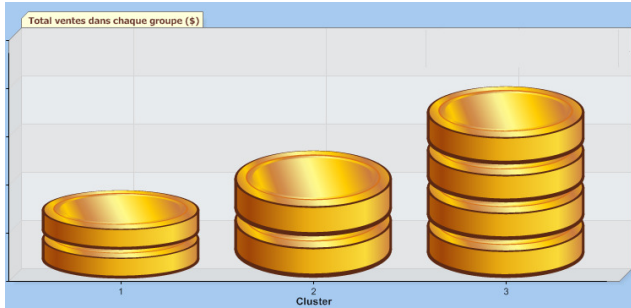


Figure 7. Les moyennes des caractéristiques du groupe 3 par rapport à celles aux États-Unis.

4.2 Analyse croisée avec les ventes

L'analyse des groupes sociodémographiques est croisée avec les données réelles des ventes, pour chaque type de produit, dans chaque groupe et chaque état afin d'approfondir la caractérisation du marché, vis-à-vis de la pénétration effective des ventes réalisées. Le modèle permet aussi de comparer, au sein de chaque groupe, les ventes des principaux types de produits, afin de mesurer la marge de progrès à réaliser pour atteindre la meilleure vente dans chaque groupe. En regardant les données historiques des ventes dans chaque État, le modèle permet à l'entreprise de détecter automatiquement les états et ainsi les magasins en difficulté (baisse des ventes).



**Figure 8 : Ventes totales dans chaque groupe.**

La Figure 8, montre que les ventes du groupe 1 sont plus faibles même s'il contient 29 états. Par contre, le groupe 3 compte seulement 9 états mais les ventes sont les meilleures (près de 3 fois le groupe 1). Les 9 états dans le groupe 3 semblent les marchés potentiels.

Avec ces résultats, l'entreprise peut formuler des stratégies pour répondre à cibler efficacement les segments et avoir des stratégies pertinentes et directes pour améliorer ses ventes dans les états potentiels :

- Augmenter le budget de marketing dans les états potentiels.
- Augmenter la chaîne de magasin dans les états potentiels.
- Centraliser sur les produits préférés dans chaque région, chaque magasin.

## 5 CONCLUSIONS ET PERSPECTIVES

Les résultats confirment que l'utilisation des données secondaires avec techniques du data mining s'applique bien pour la segmentation du marché des meubles. Les trois segments formés se distinguent par des caractéristiques différentes.

De plus, le modèle intégré les données historiques au résultat de la segmentation ce qui aide à visualiser les relations entre les segments et les ventes réelles afin de fournir un outil d'aide à la décision.

A partir du résultat final, l'entreprise pourra établir différentes stratégies ou décider d'adapter la chaîne de distribution dans chaque segment.

Notre recherche envisage différentes perspectives de poursuite :

- Une mise en direct avec les données de l'entreprise pour analyser les changements du marché en temps réel.
- Analyse de la concurrence du marché dans chaque groupe.
- Analyse de la corrélation entre le contenu d'un magasin et les ventes locales.
- Mesure de l'impact de l'introduction d'un nouveau produit sur les planchers de vente.

## 6 REMERCIEMENT

Nous tenons à remercier le soutien de l'entreprise Canadel qui collabore étroitement à cette étude. Cette recherche est également soutenue par le programme de bourse canadien pour la francophonie (PCBF).

## 7 RÉFÉRENCES

Aaker, V.Kumar, & S.Day, G. (2007). *Marketing Research*. NY: John Wiley and Sons.  
 Berry, M. J., & Linoff, G. (1997). *Data Mining: For Marketing, Sales, and Customer Support*. John Wiley & sons, Inc.

Bruno, A., Catherine, M., & Martin, T. (2006). Mining Public Transport User Behavior From Smart Card Data. *12 th IFAC Symposium on Information Control Problems in Manufacturing*, 14, pp. 193-203.  
 Castleberry, S. B. (2001). Using Secondary Data in Marketing Research : A Project That Melds Web and Off-Web Sources. *Journal of Marketing Education* , 23, 195-203.  
 Croft, M. J. (1994). *Market Segmentation: A Step-by-step Guide to Profitable New Business*. Routledge.  
 DeSarbo, W., Carrol, J., & Clark, L. (1984). Synthesized Clustering: A method for Amalgamating Alternative Clustering Bases with Differential Weighting of Variables. *Psychometrika* , 49, 57-78.  
 Dolnicar, S., & Leisch, F. (2004). Segmenting Markets by Baged Clustering. *Australasian Marketing Journal* , 12, 51-65.  
 Florence, M. (1990). International Trade and the Furniture Industry. *Alabama International Trade Center* , 18-32.  
 Green, P. (1977). A new approach to market segmentation. *Business Horizons* , 20, 61-73.  
 Greeno, D., Summers, N., & Kernan, J. (1973). Personality and Implicit Behavior Patterns. *Journal of Marketing Research* , 10, 63-69.  
 H.Sung, H. (2004). Classification of Adventure Travelers:Behavior, Decision Making,and Target Markets. *Journal of Travel Research* , 42, 343-356.  
 Kaynak, E., & Hassan, S. (1994). *Globalization of Consumer Markets: Structures and Strategies*. Haworth Press.  
 Maulika, U., & Bandyopadhyay, S. (2000). Genetic algorithm-based clustering technique. *Pattern Recognition* , 33, 1455-1465.  
 Mc Queen, J. (1967). Some methods for classification and analysis of multivariate observations. *Proceedings of the 5th Berkeley symposium on mathematical statistics and probability*, (pp. 281-297).  
 Murillo, L. M. (2007). Manufacturers-Retailers: The New Actor in the U.S.Furniture Industry. Characteristics and Implications for the Chinese Industry. *Proceedings of world cademy of science, Engineering and technology*.  
 Patze, G. L. (1995). *Using Secondary Data in Marketing Research*. Quorum Books.  
 Pearson, K. (1904). On the Theory of Contingency and its Relation to Association and Normal Correlation. *Draper's Co. Res. Mem. Biometric Ser* , 1, 1-35.  
 Punj, G., & Stewart, D. W. (1983). Cluster Analysis in Marketing Research: Review and Suggestions for Application. *Journal of Marketing Research* , 20 (2), 134-148.  
 Salvador, S., & Chan, P. (2004). Determining the Number of Clusters/Segments in Hierarchical Clustering/Segmentation Algorithms. *Proceedings of the 16th IEEE International Conference on Tools with Artificial Intelligence*, (pp. 576-584).  
 Singh, J. (1990). A typology of consumer response styles. *Journal of Retailing* , 66 (1), 57-99.  
 Smith, W. (1956). Product differentiation and Market Segmentation as Alternative Marketing Strategies. *Journal of marketing* , 3-8.  
 Umesh, U. (1987). Transferability of preference models across segments and geographic areas. *Journal of Marketing* , 51, 59-70.  
 Wedel, M., & Kamakura, W. A. (1998). *Market Segmentation : Conceptual and Methodological Foundation*. Malsachusetts: Kluwer Academic Publishers.  
 Wind, Y. (1978). Issues and Advances in Segmentation Research. *Journal of Marketing Research* , 15 (3), 317-337.