

HEC MONTRÉAL

**Identification de critères influençant la défection de clientèle dans un contexte
d'arrivée de nouveaux compétiteurs**

Par

Sébastien Rodriguez

Sciences de la gestion

(Intelligence d'affaires)

Mémoire présenté en vue de l'obtention du grade de maîtrise ès sciences

(M.Sc.)

Janvier 2013

© Sébastien Rodriguez, 2013

**AVIS DE CONFORMITÉ À LA POLITIQUE EN MATIÈRE D'ÉTHIQUE
DE LA RECHERCHE AVEC DES ÊTRES HUMAINS DE HEC MONTRÉAL**

La présente atteste que le projet de recherche décrit ci-dessous a fait l'objet d'une évaluation en matière d'éthique de la recherche avec des êtres humains et qu'il satisfait les exigences de notre politique en cette matière.

Titre du projet de recherche:

Évaluation de la propension d'un client à changer de magasin d'alimentation principal suite à l'ouverture d'un concurrent à proximité

Chercheur principal:

Chercheur : Sébastien Rodriguez

Titre : Étudiant(e) Maîtrise

Service/Option : Intelligence d'affaires

Directeur : Marc Fredette

Titre : Professeur(e) agrégé(e)

Service/Option : Méthodes quantitatives de gestion

Date de déclaration du projet au Comité d'éthique de la recherche:

22 mars 2011

Date d'approbation du projet:

05 mai 2011

Date de publication de l'avis:

05 mai 2011



Silvana Pozzebon, Présidente
Comité d'éthique de la recherche

Sommaire

La prédiction de la défection d'un individu aussi appelée *churn*, a largement été couverte à travers la littérature. Différentes approches ont été utilisées afin de modéliser ce phénomène variant en fonction des variables disponibles et du but recherché. Bien que plusieurs se soient penchés sur la question, la majorité des travaux disponibles à ce jour émanent du monde des télécommunications et des services financiers.

Ce mémoire a comme objectif de modéliser une problématique de défection de consommateurs dans le monde du commerce de détail alimentaires. De plus, ce travail tentera de comprendre les caractéristiques qui rendent un consommateur à risque de quitter une entreprise dans un contexte où un nouveau compétiteur fait son apparition sur le marché.

Pour y parvenir, un modèle de régression linéaire pour données corrélées considérant l'effet du temps a été proposé dans ce travail.

L'application du modèle sur les données transactionnelles d'un détaillant alimentaire Nord-Américain a permis de déterminer un ensemble de variables permettant d'identifier des clients à risque de quitter l'entreprise.

Mots clés : Défection, Churn, Analyse longitudinale, Modèle explicatif

Liste des tableaux

<u>Tableau 1</u> : Modèles prédictifs du churn vs. type de défection.....	18
<u>Tableau 2</u> : Variables indépendantes incluses dans la base de données finale.....	28
<u>Tableau 3</u> : Moyennes des dépenses hebdomadaires des clients membres du programme de fidélité avant, pendant et après l'arrivée du compétiteur A	34
<u>Tableau 4</u> : Moyennes des dépenses hebdomadaires des clients membres du programme de fidélité avant, pendant et après l'arrivée du compétiteur B.....	35
<u>Tableau 5</u> : Moyennes des dépenses départementales des clients du magasin 1.....	36
<u>Tableau 6</u> : Moyennes des dépenses départementales des clients du magasin 2.....	37
<u>Tableau 7</u> : Moyennes des dépenses départementales des clients du magasin 3.....	38
<u>Tableau 8</u> : Moyennes des dépenses départementales des clients du magasin 4.....	38
<u>Tableau 9</u> : Moyennes des dépenses départementales des clients du magasin 5.....	39
<u>Tableau 10</u> : Moyennes des dépenses départementales des clients du magasin 6.....	40
<u>Tableau 11</u> : Moyennes des dépenses départementales des clients du magasin 7.....	40
<u>Tableau 12</u> : Moyennes des dépenses départementales des clients du magasin 8.....	41
<u>Tableau 13</u> : Moyennes des variations départementales avant/après l'arrivée des compétiteurs A et B.....	41
<u>Tableau 14</u> : Moyennes des dépenses hebdomadaires par segment de clientèle avant/après l'arrivée du compétiteur A	43
<u>Tableau 15</u> : Moyennes des dépenses hebdomadaires par segment de clientèle avant/après l'arrivée du compétiteur B	44
<u>Tableau 16</u> : Moyennes des dépenses hebdomadaires en fonction du type de milieu avant/après l'arrivée du compétiteur	45
<u>Tableau 17</u> : Moyennes des dépenses hebdomadaires en fonction du type de magasin étant le plus près des clients étudiés	46
<u>Tableau 18</u> : Ensemble des effets simples et des interactions faisant partie du modèle.....	47
<u>Tableau 19</u> : Estimation du paramètre <i>RFV_Code</i>	49
<u>Tableau 20</u> : Estimation du paramètre <i>Semaine</i>	50
<u>Tableau 21</u> : Estimation du paramètre <i>Semaine*RFV_Code</i>	52
<u>Tableau 22</u> : Estimation du paramètre <i>Compétiteur</i>	53
<u>Tableau 23</u> : Estimation du paramètre <i>RFV_Code*Compétiteur</i>	54
<u>Tableau 24</u> : Estimation du paramètre <i>Semaine*Compétiteur</i>	55
<u>Tableau 25</u> : Estimation du paramètre <i>Noël</i>	55
<u>Tableau 26</u> : Estimation du paramètre <i>Compétiteur_Plus_Proche</i>	56

Liste des tableaux

<u>Tableau27</u> : Estimation du paramètre <i>Urbain_Rural</i>	56
<u>Tableau28</u> : Estimation du paramètre <i>Compétiteur_Plus_Proche*Urbain_Rural</i>	58
<u>Tableau29</u> : Estimation du paramètre <i>Compétiteur*Compétiteur_Plus_Proche</i>	59
<u>Tableau30</u> : Estimation du paramètre <i>Ratio_Département</i>	60
<u>Tableau31</u> : Estimation du paramètre <i>Ratio_Déli*Compétiteur</i>	61
<u>Tableau32</u> : Estimation du paramètre <i>Ratio_Fruits&Légumes*Compétiteur</i>	62
<u>Tableau33</u> : Estimation du paramètre <i>Ratio_Non_alimentaire*Compétiteur</i>	62

Liste des figures

<u>Figure 1</u> : Évolution des dépenses hebdomadaires des clients membres du programme de fidélité ayant vécu l'arrivée du compétiteur A.....	33
<u>Figure 2</u> : Évolution des dépenses hebdomadaires des clients membres du programme de fidélité ayant vécu l'arrivée du compétiteur B.....	34
<u>Figure 3</u> : Évolution des dépenses départementales chez les clients analysés du magasin 1.....	36
<u>Figure 4</u> : Évolution des dépenses départementales chez les clients analysés du magasin 2.....	37
<u>Figure 5</u> : Évolution des dépenses départementales chez les clients analysés du magasin 3.....	37
<u>Figure 6</u> : Évolution des dépenses départementales chez les clients analysés du magasin 4.....	38
<u>Figure 7</u> : Évolution des dépenses départementales chez les clients analysés du magasin 5.....	39
<u>Figure 8</u> : Évolution des dépenses départementales chez les clients analysés du magasin 6.....	39
<u>Figure 9</u> : Évolution des dépenses départementales chez les clients analysés du magasin 7.....	40
<u>Figure 10</u> : Évolution des dépenses départementales chez les clients analysés du magasin 8...	41
<u>Figure 11</u> : Évolution des dépenses hebdomadaires par segment de clientèle impacté par le compétiteur A	43
<u>Figure 12</u> : Évolution des dépenses hebdomadaires par segment de clientèle impacté par le compétiteur B	43
<u>Figure 13</u> : Évolution des dépenses hebdomadaires pour les magasins se trouvant en milieu urbain	44
<u>Figure 14</u> : Évolution des dépenses hebdomadaires pour les magasins se trouvant en milieu rural.....	45
<u>Figure 15</u> : Évolution des dépenses hebdomadaires en fonction du magasin le plus près des clients étudiés	46
<u>Figure 16</u> : Effet du temps sur les différents segments de clients étudiés	50
<u>Figure 17</u> : Effet de la présence du compétiteur A ou B sur le type de segment de clientèle	53
<u>Figure 18</u> : Effet de la proximité du magasin sur le type de ville où il se trouve	57
<u>Figure 19</u> : Effet de la proximité du magasin sur le type de compétiteur présent dans le marché	58

Remerciement

Je tiens à dédier ce travail à l'ensemble des membres de ma famille qui chacun à leur façon, ont su jouer un rôle indispensable dans l'accomplissement de ce travail. Leurs encouragements m'ont permis d'aller jusqu'au bout et je leur en serai éternellement reconnaissant.

Je souhaite aussi remercier M. Marc Fredette pour avoir accepté de diriger mon mémoire. Sa disponibilité, sa patience et ses judicieux conseils ont aussi joué un rôle déterminant dans la réalisation de ce travail.

CHAPITRE 1 : INTRODUCTION ET REVUE DE LITTÉRATURE

1.1-Introduction

Depuis déjà quelques années, il est de plus en plus commun chez les entreprises œuvrant dans le monde du commerce de détail, d'offrir un programme de fidélité à leur clientèle. Par l'entremise de ces programmes, les entreprises sont en mesure de recueillir l'information rattachée à chacune des transactions effectuées par leurs clients. Bien que ces programmes aient un coût non négligeable – Aeroplan ayant remis 2.3 millions de primes en 2011 au Canada (Aimia 2012) – ceux qui les utilisent comprennent qu'il s'agit d'un outil permettant de mieux connaître les habitudes des clients et ainsi développer des stratégies marketing plus efficaces et mieux ciblées.

À ce chapitre, une des stratégies marketing ayant suscité beaucoup d'intérêt auprès des chercheurs est celle de la modélisation permettant de prédire la défection ou le « churn » d'un client face à la compagnie avec laquelle il fait affaire. Plusieurs travaux de recherche ont traité de la question et bon nombre d'entre eux se sont penchés plus particulièrement sur des problématiques reliées au monde de la téléphonie cellulaire, des assurances ainsi que des services bancaires.

Pour cet essai, le thème abordé sera lui aussi celui de la défection de clientèle, mais il se distinguera des études déjà publiées quant à 2 aspects principaux: 1- Il portera sur une problématique vécue par une entreprise Nord-Américaine œuvrant dans le commerce de détail alimentaire, un secteur qui à notre connaissance, n'a pratiquement pas été étudié dans la littérature pour ce genre d'analyse. 2- Il combinera à la fois des données générées par des clients au sein de leur entreprise avec l'occurrence d'un évènement extérieur : l'arrivée d'un nouveau concurrent à proximité d'un point de vente. À notre connaissance, un tel évènement n'a jamais été analysé dans un contexte de défection de clientèle du monde alimentaire. L'objectif que se fixe ce mémoire est de valider l'existence de facteurs clés permettant d'identifier les clients ayant une forte propension à quitter ou délaisser l'entreprise étudiée suite à l'arrivée d'un concurrent, en plus de comprendre et de quantifier l'effet que ces facteurs ont sur les magasins.

Ce travail sera divisé de la façon suivante : Le premier chapitre consiste en une revue de littérature dans laquelle on expose divers contextes d'affaires dans lesquels la modélisation du « churn » a été réalisée tout en portant un regard attentif sur les méthodologies utilisées dans chacun de ces différents contextes. Le second chapitre détaille la problématique à laquelle ce mémoire s'attarde, passe en revue les diverses étapes ayant mené à la création de la base de données finale et expose la méthodologie adoptée pour modéliser la défection d'un client. Le troisième chapitre donne un aperçu général de l'impact des compétiteurs sur les magasins étudiés et explique les résultats obtenus à l'aide du modèle créé. Nous terminerons avec un dernier chapitre proposant une réflexion sur les résultats obtenus ainsi que des recommandations managériales en fonction de ceux-ci.

1.2-Revue de littérature

Bien que nous sachions que la modélisation du « churn » soit une technique utilisée chez les entreprises traditionnelles de commerce de détail, notre analyse de la littérature aura permis de constater que très peu d'articles scientifiques se sont penchés sur ce domaine d'activité. Pour cette raison, ce chapitre exposera d'autres contextes d'affaires ayant souvent fait l'objet de modélisation du « churn » et contrastera ces derniers avec la littérature existante sur le commerce de détail en ce qui a trait aux méthodologies utilisées et aux variables étudiées. Nous introduirons également les concepts de défection totale et de défection partielle.

1.2.1 Contextes d'affaires les plus courants dans la modélisation du churn

Un des sujets les plus souvent abordé dans la modélisation du « churn » est celui de la défection de clientèle face aux abonnements de téléphonie cellulaire. Au début des années 2000 aux États-Unis, le taux d'abonnement annuel de service téléphonique mobile était passé de 50% à 15% alors que celui de la défection de clientèle avait grimpé de 28 à 46% (Duke Teradata 2002). Cette situation força les compagnies à trouver une façon de cibler les clients à risque de quitter la compagnie avant que ceux-ci ne passent à l'action. Nath et Behara (2002) figurent parmi les chercheurs

s'étant intéressés aux techniques de prospection de données (Data Mining) permettant de prédire les clients délaissant leur compagnie de cellulaire. Pour étudier la question, le Centre Teradata de l'Université de Duke leur a fourni des données provenant d'une compagnie de téléphonie cellulaire américaine. Cette base de données contenait de l'information sur environ 170 variables et sur plus de 100 000 clients. Afin de créer la variable binaire dépendante indiquant l'état de défection, ils décidèrent qu'un client n'ayant pas utilisé son service de téléphonie mobile pendant plus de 30 jours serait considéré comme un client ayant quitté la compagnie (1 : Ayant quitté la compagnie et 0 : N'ayant pas quitté la compagnie). Pour ce qui est des variables prédictives étudiées, elles se regroupent en 4 catégories différentes :

- A- Les variables sociodémographiques*
- B- Les variables comportementales (Utilisation du cellulaire, date, durée, minutes additionnelles, etc.)*
- C- La qualité du service reçu*
- D- Caractéristiques marketing (Type de forfaits, options, etc.)*

Afin de déterminer quelles variables aidaient à prédire la défection d'un client, Nath et Behara (2002) ont utilisé une version simplifiée du théorème de Bayes sur les probabilités conditionnelles. Cette approche consiste d'abord à calculer la probabilité que le « churn » survienne pour chacune des variables prédictives du modèle et ensuite multiplier chacune de ces probabilités entre-elles afin d'obtenir la probabilité « totale » que la défection survienne. Cette approche a permis à Nath et Behara (2002) de prédire correctement le comportement de 68% des clients étudiés. Les variables avec le meilleur pouvoir prédictif furent les suivantes, en ordre d'importance :

- 1) DUALBAND : Type de téléphone
- 2) CARTYPE : Moyen de transport privilégié
- 3) EDUC1 : Niveau d'éducation
- 4) ETHNIC : Groupe ethnique
- 5) TOT_ACPT : Nombre d'offres acceptées par l'équipe de rétention
- 6) OCCU1 : Occupation du 1^{er} membre du ménage
- 7) AREA : Situation géographique
- 8) INCOME : Revenu estimé du ménage
- 9) DWLLSIZE : Taille du logement
- 10) PROPTYPE : Type de propriété

L'autre contexte d'affaires ayant souvent été couvert dans la littérature de la défection est celui des services financiers. Contrairement à la téléphonie cellulaire, les services financiers sont considérés comme étant un contexte d'affaires non-contractuel, puisqu'aucune obligation ne lie normalement les individus à leur banque en ce qui a trait aux services financiers de base. En 2008, Baesens et al. se sont penchés sur le cas d'une banque belge qui souhaitait cibler des clients potentiellement à risque de quitter l'institution. Pour y parvenir, Baesens et al. (2008) ont utilisé une approche qui avait auparavant uniquement été utilisée dans un contexte contractuel. Il s'agit de la modélisation du « churn » avec comme variable dépendante une variable de type CLV (Customer Lifetime Value ou Valeur à Vie du Consommateur - Shaw 1988). Ce type de variable consiste à déterminer la valeur actualisée des profits que génèrera un client dans le futur et ce, en ne considérant pas les coûts déjà engendrés par la compagnie. Pour créer leur modèle, ils ont utilisé un jeu de données ayant été fourni par une institution bancaire belge et qui comptait neuf mois de transactions effectuées par plus de dix milles clients. Ce jeu de données contenait les catégories d'informations suivantes sur les transactions :

- A- Données temporelles (Date des transactions)*
- B- Informations sur le type de transactions (Retraits, dépôts, transferts, etc.)*
- C- Informations sur la facturation (Nombre de factures envoyées, montants facturés, type de frais, etc.)*

Ce qui frappe dans l'approche de Baesens et al. (2008), est qu'ils ont utilisé un raisonnement moins traditionnel dans l'identification d'un client en état de défection. En effet, les cas étudiés dans la littérature ont l'habitude de considérer un client en défection lorsque ce dernier a complètement cessé ses activités à l'égard de la compagnie avec laquelle il fait affaire. À titre d'exemple, il suffit de penser au cas étudié par Nath et Behara (2002) qui considéraient la défection d'un client à partir du moment où celui-ci avait complètement cessé ses interactions avec l'entreprise pendant au moins trente jours. Dans les travaux de Baesens et al. (2008), ils ont plutôt considéré que l'état de défection d'un client débutait à partir du moment où ce dernier commençait à réduire son interaction avec la compagnie. Ce type de défection est ce que l'on appelle : défection partielle. Par rapport à cette définition, leur modélisation

visait à prédire le moment où les clients commençaient à montrer une décroissance au niveau de la valeur dépendante CLV ou, en d'autres mots, à partir du moment où la profitabilité de ces derniers commençait à décroître. D'un point de vue managérial, ils trouvaient pertinent pour une institution financière de considérer un client en défection dès que ce dernier commençait à décroître en profitabilité et non pas uniquement à partir du moment où il avait quitté définitivement l'institution.

Pour ce qui est des variables prédictives, Baesens et al. (2008) ont uniquement utilisé une variable de type RFV (Recense, Fréquence, Valeur monétaire - Fader et al. 2005) en se basant sur l'étude de Buckinx et Van den Poel (2004) qui avait démontré que les variables de type RFV étaient de bons prédicteurs pour une variable de type CLV. La modélisation de la CLV en utilisant une variable RFV aura donné raison à Baesens et al. (2008) puisqu'ils obtinrent eux aussi des résultats empiriques probants. Différentes techniques de modélisation furent utilisées afin de valider quelle approche prédisait le mieux la défection partielle des clients. Les principales furent : La régression logistique, les réseaux de neurones et les arbres de décision avec fonction de coût pénalisant la mauvaise classification des « churners ». L'approche la plus performante fut celle de l'arbre de décision avec fonction de coût.

Tout comme Nath et Bahara en 2002, Lemmens et Croux (2006) se sont aussi intéressés au domaine de la téléphonie cellulaire, mais ont plutôt voulu examiner si les techniques de « bagging » (Breiman 1996) et de « boosting » (Freund & Schapire 1999) arrivaient à mieux prédire le churn que l'approche par régression logistique qui selon eux, était la plus utilisée dans le domaine du marketing. Afin de réaliser cette comparaison, Lemmens et Croux (2006) ont travaillé sur le même jeu de données fourni par le centre Teradata de l'Université Duke que Nath et Behara avaient utilisé en 2002. Contrairement à ces derniers qui avaient utilisé l'ensemble des variables indépendantes disponibles, Lemmens et Croux (2006) ont plutôt décidé de garder uniquement les variables ayant moins de 30% de valeurs manquantes. Ce critère de sélection aura permis de réduire à 46 le nombre de variables à utiliser pour la modélisation. Ces 46 variables se répartissent dans les trois catégories suivantes :

- A- *Les variables sociodémographiques*
- B- *Les variables comportementales (Utilisation du cellulaire, date, durée des appels, minutes additionnelles, etc.)*
- C- *Les variables démontrant l'interaction entre la compagnie et le client (Minutes passées avec le service à la clientèle et réponses aux offres promotionnelles envoyées par courriel)*

Les techniques de bagging et de boosting sont considérées comme étant de type apprentissage automatique. La première consiste à faire la moyenne de plusieurs échantillons issus de tirage avec remise (ou de réchantillonnage) alors que la deuxième consiste à combiner plusieurs algorithmes et classifieurs (faibles et forts) afin de permettre au modèle d'optimiser au maximum les performances des algorithmes. En plus de ces deux techniques, Lemmens et Croux (2006) ont testé l'approche par régression logistique afin de voir laquelle de ces trois techniques performait le mieux. Le coefficient de Gini et le « lift » du meilleur décile furent choisis comme indicateurs déterminant la performance des différents modèles.

Lemmens et Croux (2006) déterminèrent finalement que l'approche par régression logistique était celle qui performait le moins bien. Après quelques itérations, les résultats ont démontré que les techniques de boosting et de bagging performaient mieux que l'approche logistique. En effet, le bagging s'est révélé comme la meilleure technique de modélisation, puisqu'elle performait tout aussi bien que le boosting, mais en ayant l'avantage d'être conceptuellement plus facile à interpréter (Lemmens et Croux 2006). Les variables prédictives s'étant avérées les plus efficaces selon Lemmens et Croux (2006) furent les suivantes en ordre d'importance :

1. *Âge de l'équipement de téléphonie*
2. *Changement dans les minutes utilisées par mois*
3. *Coût de base du plan forfaitaire*
4. *Coût du téléphone*
5. *Moyenne d'appels en période de fort achalandage du réseau*

1.2.2 La modélisation du churn dans le commerce de détail

Buckinx et Van den Poel (2004) ont été parmi les premiers à étudier la défection de clients dans le monde du commerce de détail de type B2C (Business to Consumer). Comme dans ce mémoire, Buckinx et Van den Poel (2004) se sont penchés sur une problématique liée à la défection de clientèle chez un détaillant alimentaire. Tel qu'abordé précédemment par Baesens et al. (2008), l'objectif de Buckinx et Van den Poel (2004) était également de cibler la défection des clients à partir du moment où ceux-ci commençaient à diminuer en profitabilité (défection partielle). Comme dans le domaine des services bancaires, le commerce de détail alimentaire est un contexte non-contractuel cherchant à identifier les clients dès qu'ils diminuent en profitabilité et non pas lorsqu'ils ont quitté complètement le détaillant (défection totale). C'est donc à partir d'une base de données contenant de l'information transactionnelle sur 158 884 clients de l'un des plus importants détaillant alimentaire au monde, que Buckinx et Van den Poel (2004) ont effectué leur recherche. Ces données furent récoltées via un système de carte fidélité entre avril 2000 et janvier 2001 et ce, uniquement auprès des clients les plus loyaux. Ils ont utilisé 61 variables prédictives pouvant être regroupées dans les catégories suivantes :

- A- *Temps écoulé entre les achats*
- B- *Fréquence d'achat*
- C- *Valeur d'achat*
- D- *Département d'épicerie*
- E- *Type de marques achetées (privées, nationales, économiques)*
- F- *Durée de la relation avec le client*
- G- *Heure d'achat*
- H- *Méthode de paiement*
- I- *Promotions*
- J- *Variables sociodémographiques*

Comme variable dépendante, une variable binaire fut créée afin de déterminer si le client étudié se trouvait ou non dans un état de défection partielle. Pour être considéré en état de défection partielle, Buckinx et Van den Poel (2004) ont déterminé que l'une des deux conditions suivantes devaient être remplies : 1- Avoir une fréquence d'achat inférieure à la moyenne 2- Avoir un taux de rachat sous la moyenne (indice que le temps écoulé entre les visites est irrégulier). Comme technique de modélisation, Buckinx et Van den Poel (2004) ont utilisé la même approche que Baesens et al. (2006), c'est-à-dire comparer les résultats de la régression logistique et des réseaux de neurones. De plus, ils ont ajouté la modélisation par forêts aléatoires dans leur comparaison entre différents modèles. Buckinx et Van den Poel (2004) conclurent finalement que ces trois types de modélisation performaient relativement bien et que les variables contribuant le plus à la prédiction des clients ayant partiellement fait défection étaient les suivantes en ordre d'importance :

1. *Variables en lien avec la fréquence d'achat du client*
2. *Variables en lien avec la récurrence des achats du client*
3. *Durée de la relation avec le client*
4. *Valeur monétaire des achats du client*
5. *% des visites avec dépenses supérieures à la moyenne*
6. *Variable signifiant qu'un client ne veut pas donner ses informations démographiques*
7. *Montant dépensé par le client dans les différents départements*
8. *Montant retourné aux clients pour les consignes de bouteilles vides*

Le tableau 1 synthétise ce qui a été couvert dans la revue de littérature en lien avec la modélisation du « churn ».

Tableau 1 : Modèles prédictifs du churn vs type de défection

Auteur	Année	Contexte d'affaires / Existence de contrat	Variables prédictives avec effet significatif sur la prédiction du churn	Variable dépendante utilisée pour la modélisation du churn	Type de défection	Technique de modélisation préconisée
Nath & Behara	2002	Téléphonie cellulaire/ Avec contrat	Type de téléphone, Moyen de transport, Niveau d'éducation, Groupe ethnique, Nombre d'offres promotionnelles acceptées, Occupation du premier ménage, Situation géographique, Revenu du ménage, Taille du logement, Type de propriété	Churn lorsqu'il y a inactivité du client pendant au moins 30 jours	Totale	Approche par probabilité avec le théorème de Bayes
Baesens & al.	2008	Banques / Sans contrat	Variable RFV (Récence, Fréquence, Valeur monétaire)	CLV (Valeur à vie du consommateur)	Partielle	Régression logistique, réseaux de neurones, arbres de décision
Lemmens & Croux	2006	Téléphonie cellulaire/ Avec contrat	Âge de l'équipement, Changement des minutes mensuelles utilisées, Coût de base du forfait, Coût du téléphone, Moyenne d'appels en période d'achalandage du réseau	Churn lorsqu'il y a inactivité du client pendant au moins 30 jours	Totale	Arbres avec Bagging, Boosting
Buckinx & Van den Poel	2004	Détaillant en alimentation/ Sans contrat	Récence, Fréquence et valeur des achats, Durée de la relation, % des transactions plus élevées que la moyenne, Variable indiquant qu'un client ne donne pas ses renseignements personnels, \$ dépensés par département, \$ liés au retour de bouteilles	Variable déterminant si le client est en état de défection partielle ou non	Partielle	Régression logistique, réseaux de neurones, forêts aléatoires

Le tableau 1 permet d'apprécier les points communs et divergents qu'ont ces travaux quant aux contextes d'affaires dans lesquels la modélisation du « churn » est utilisée, les variables prédictives utilisées, la création de la variable dépendante, les techniques de modélisation ainsi que sur le type de défection (totale ou partielle) que l'on tente de prédire. Les différences et les similitudes observées permettent de voir qu'il n'existe pas un seul type de variable ou une seule technique de modélisation donnant à tout coup la meilleure solution. Chaque contexte est unique et nécessite une analyse approfondie de la situation afin de trouver l'approche répondant le mieux à la problématique étudiée.

Cette revue de littérature a permis d'observer que lors d'une modélisation de défection d'un client, cette dernière peut soit être totale ou partielle. De plus, il a été possible de constater que dans une perspective de rétention de clientèle, la modélisation d'une défection partielle a comme avantage de pouvoir identifier un client dès qu'il commence à diminuer en profitabilité ou en interaction avec l'entreprise. Cet avantage permet à une entreprise de mettre en marche une stratégie de rétention de clientèle dès qu'elle commence à présenter des risques de quitter l'entreprise.

Pour ce qui est des techniques de modélisation du churn utilisées dans la littérature, elles sont principalement de type logit (régression logistique/ théorème de Bayes) ou de type apprentissage automatique (forêts aléatoires, bagging, boosting). À la lumière de ce qui a été couvert dans cette revue de littérature, les techniques de type apprentissage automatique proposent les modèles avec les meilleurs résultats prédictifs. Cependant, ces dernières ont le désavantage de fournir des résultats plus difficiles à interpréter au niveau des effets qu'ont les variables prédictives sur la variable d'influence. À ce registre, c'est l'approche par régression qui propose le meilleur compromis entre de bons résultats prédictifs et une interprétation plus facile des variables prédictives.

Puisque l'objectif que s'est fixé ce mémoire est d'arriver à identifier et à expliquer les facteurs permettant de cibler les clients à risque de défection, la modélisation de type apprentissage automatique sera délaissée au profit d'une modélisation de type régression linéaire ayant comme avantage sa simplicité d'analyse. Cette approche permettra d'identifier et d'expliquer plus facilement l'effet qu'ont les variables prédictives et leur combinaison (variables d'interaction) sur la défection des clients.

L'ensemble des études citées dans notre revue de littérature ont basé leurs analyses sur des données ayant été récoltées sur une certaine période de temps pour ensuite être agrégées par individu. Cette approche considère chaque individu comme une observation unique, ce qui rend donc impossible l'étude de l'effet direct du temps sur la défection de ces individus. Ce travail se démarquera donc de ceux ayant été couverts dans cette revue de littérature de par son analyse de données non-agrégées permettant de considérer l'effet du temps sur la défection des clients.

CHAPITRE 2 : MÉTHODOLOGIE

2.1-Préparation des données

La collecte et la préparation de données représentent une grande partie du travail de recherche ayant été accompli dans ce projet. En plus des diverses extractions effectuées afin de récolter les données nécessaires à la recherche, de multiples tâches de nettoyage, de recodage et de création de variables ont aussi été accomplies.

2.1.1. – Entrepôt de données transactionnelles du détaillant étudié

Les données utilisées pour ce mémoire ont été extraites directement de l'entrepôt de données transactionnelles de la bannière alimentaire étudiée. Cet entrepôt conserve l'information sur la totalité des transactions effectuées dans l'ensemble des magasins de la province pendant vingt-sept mois. Chacune des transactions possède un identifiant unique et peut être associée à un individu s'il a présenté sa carte de fidélité au moment de régler son achat. Pour chacune des transactions, il est également possible d'obtenir le détail de tous les produits associés au numéro de transaction unique.

2.1.2. – Construction de la base de données de recherche

L'extraction des données fut effectuée à l'aide du logiciel SAS et du langage de requête SQL.

Étape 1 : Extraction de l'ensemble des transactions liées à la problématique d'affaires

L'objectif de recherche étant de déterminer les facteurs influençant la défection d'un client suite à l'arrivée d'un compétiteur, la première étape fut de trouver les magasins ayant vécu ce scénario dans les vingt-sept mois précédant le moment de l'extraction (données disponibles). Huit magasins issus de différentes régions et ayant été touchés soit par le compétiteur A ou le compétiteur B furent alors identifiés. Pour chacun de ces magasins, l'ensemble des transactions effectuées huit semaines avant l'arrivée du compétiteur ainsi que celles effectuées huit semaines après l'arrivée du compétiteur, excluant les deux premières semaines d'ouverture, furent extraites. Les ventes ayant été effectuées pendant les deux premières semaines d'ouverture des compétiteurs ne

furent pas incluses dans l'analyse, car selon l'entreprise étudiée, ces dernières ont l'habitude d'être considérablement plus élevées que ce qu'elles deviennent par la suite. En effet, les épiceries ont l'habitude d'effectuer des événements spéciaux lors de leur première semaine d'ouverture en plus d'offrir des prix très agressifs, ce qui a pour effet d'attirer de nouveaux clients curieux d'essayer le nouvel établissement.

Étape 2 : Élimination des transactions non-pertinentes à la problématique de recherche

Afin de pouvoir identifier un client unique dans le temps (avant et après l'arrivée du compétiteur), seules les transactions ayant été effectuées à l'aide d'une carte de fidélité ont été conservées. Le numéro du membre participant au programme de fidélité constitue le seul moyen d'associer un individu à une transaction. La variable permettant d'identifier le membre est *Household_id*. Il est raisonnable de croire que les membres du programme de fidélité soient des clients plus fidèles que ceux n'en faisant pas partie. Bien que la totalité des clients analysés dans cette étude fassent partie de ce programme de fidélité et que cela puisse être vu comme un biais, il est important de se rappeler que ces clients étaient les seuls que l'on pouvait étudier dans un contexte longitudinal. Les autres clients non-membres ne peuvent en effet pas être associés à aucune transaction.

À noter également que les transactions des membres ayant transigé uniquement pendant la période post-ouverture du compétiteur ont toutes été éliminées de la base de données. Puisque ce travail s'intéresse à l'effet causé par l'arrivée d'un nouveau compétiteur sur les clients de la bannière, il était essentiel de sélectionner uniquement des membres ayant effectué au moins une transaction dans la période pré-ouverture du compétiteur.

Étape 3 : Liaison des variables d'intérêt aux transactions sélectionnées

Depense_Hebdo est la variable dépendante du modèle et représente le montant hebdomadaire total dépensé par les clients de la bannière étudiée. Afin de détailler le contenu de ces dépenses hebdomadaires, une table indiquant le département du magasin dans lequel les produits ont été achetés fut liée à la table des transactions créée à l'étape 2. Le prix de vente de chacun des items fut aussi ajouté aux items.

Au chapitre des variables au niveau du consommateur, une variable sociodémographique donnant le code postal de l'individu ainsi qu'une variable comportementale de type FV (Fréquence Valeur monétaire) ont été ajoutées à l'analyse. Grâce au logiciel de géo-localisation ARC-GIS, la distance à vol d'oiseau entre le code postal de l'individu et les magasins étudiés (magasin de la bannière étudiée & magasin de la compétition) fut calculée pour chacun des individus. La variable de type FV regroupe les individus étudiés en six différents segments de fidélité en fonction de leur fréquence de visites en magasin et de leurs dépenses durant une période de quinze semaines. Cette variable est rafraîchie à tous les quatre mois en utilisant les quinze dernières semaines de transactions. De plus, elle se distingue d'une variable RFV traditionnelle par le fait qu'elle n'utilise pas la récence des transactions dans sa composition, puisque dans ce cas-ci, elle est considérée comme intrinsèque à la fréquence des transactions. Dans un contexte d'achats alimentaires, la récence est jugée comme intrinsèque à la fréquence puisque le temps écoulé entre les transactions est assez rapproché. La récence au sens propre est donc étroitement liée à la fréquence des transactions.

La variable FV permet de préciser le niveau de fidélité du client. Un client dépensant un montant élevé par visite et visitant régulièrement le magasin est considéré comme faisant partie du segment le plus fidèle (*Premium*). Les segments *Lapsed* et *Gone Away* représentent quant à eux les segments les moins fidèles puisqu'ils n'ont respectivement pas magasiné chez la bannière étudiée depuis plusieurs semaines. Entre ces deux niveaux de segments, on retrouve les trois segments suivants: *Standard*, *Low* et *Uncommitted*. À noter que les variables FV qui ont été liées aux

individus sont issues de la période de rafraîchissement se rapprochant le plus de la période pré-ouverture de chacun des magasins étudiés.

Étape 4 : Création et recodage de variables

À partir des données récoltées à l'étape précédente, plusieurs nouvelles variables furent créées. La variable ordinale *Semaine* a été codée à partir de la date à laquelle la transaction a été réalisée. Elle représente la période de temps durant laquelle les transactions ont été faites par rapport à l'ouverture du compétiteur. La valeur 0 est associée à toutes les transactions faites durant les huit semaines précédant l'ouverture du compétiteur. La valeur 1 représente les transactions effectuées durant les 2 premières semaines suivant l'ouverture du compétiteur. Rappelons que les deux semaines immédiatement après l'ouverture du compétiteur sont exclues de l'analyse. La valeur 2 représente les 3 semaines suivant immédiatement les semaines de la période 1 et la valeur 3 les 3 semaines suivant immédiatement les semaines de la période 2. Les huit semaines post-ouverture du compétiteur furent séparées en trois catégories afin de faciliter la création et l'interprétation de termes d'interaction durant l'analyse des effets du modèle. Cette codification a également été réalisée afin d'équilibrer le nombre de clients par groupe.

La variable *RFV_Code* est une recodification de la variable de type FV ajoutée à l'étape 3. Cette recodification avait le même objectif que pour la codification de la variable *Semaine*, c'est-à-dire pour faire des groupes homogènes et répartis équitablement afin de faciliter la création et l'interprétation de termes d'interaction lors de l'analyse des effets du modèle. La valeur 1 de la variable *RFV_Code* représente le segment le plus fidèle (Premium), la valeur 2 le segment intermédiaire (Standard) et la valeur 3 les segments les moins fidèles (Low, Uncommitted, Lapsed et Gone Away).

Les variables *Ratio_Département_X* ont été créées à partir des dépenses faites par départements durant les huit semaines précédant l'arrivée de la compétition. La totalité des achats faits dans les huit départements les plus importants du magasin

furent comptabilisés par individu afin d'en déduire la proportion attribuable à chacun des départements. Ces variables permettent donc de comprendre le genre de produits qu'achetaient les clients avant la venue du compétiteur. Les neuf variables créées sont les suivantes : *Ratio_Département_Déli*, *Ratio_Département_Viande*, *Ratio_Département_Épicerie*, *Ratio_Département_Non-Alimentaire*, *Ratio_Département_Fruits_Légumes*, *Ratio_Département_Prêt_a_Manger*, *Ratio_Département_Boulanges*, *Ratio_Département_Cigarettes_Vin_Bière* et *Ratio_Département_Dairy*.

Voici le détail des différents départements : *Déli* regroupe les charcuteries et fromages, *Viande* regroupe les viandes, volailles et fruits de mer, *Épicerie* regroupe les produits emballés non-périssables, *Prêt à Manger* regroupe tous les mets déjà préparés ou cuisinés, *Boulanges* regroupe les produits de boulangerie fraîche et les pâtisseries, *Non-Alimentaire* regroupe tout ce qui n'est pas alimentaire, *Dairy* regroupe tous les produits laitiers et finalement *Cigarettes Vin Bière* regroupe les vins, les bières et les produits de tabac.

La variable nominale *Compétiteur* indique quant à elle la présence ou non d'un compétiteur durant la période d'analyse. La valeur de cette variable est de 0 pour les huit semaines précédant l'arrivée du compétiteur. Elle prend la valeur 0 pour les huit premières semaines de l'analyse, car aucun compétiteur n'a encore ouvert ses portes à ce moment. Pour ce qui est des huit semaines post-ouverture, la variable compétiteur prend la valeur de 1 lorsque c'est le compétiteur B qui fait son apparition dans le marché et la valeur de 2 lorsqu'il s'agit du compétiteur A. La variable *Urbain_Rural* est une variable binaire créée en fonction de la grandeur du marché où se trouvent les magasins étudiés. Si la taille de la population du marché étudié est supérieure à 50 000 habitants, la variable *Urbain_Rural* prend la valeur de 1 alors que la valeur 2 est attribuée aux marchés ayant moins de 50 000 habitants.

La variable *Compétiteur_Plus_Proche* est une variable nominale créée à partir de la distance entre l'individu et l'épicerie étudiée ainsi qu'entre l'individu et le magasin

de la compétition. Si le compétiteur est plus près de l'individu que le magasin de la bannière étudiée et ce, à une distance vol d'oiseau, la variable prend la valeur 1 alors que si c'est plutôt le magasin de la bannière étudiée, la variable prend la valeur 2. À noter que cette variable prend la valeur 0 pour les 8 semaines précédant l'arrivée du compétiteur puisqu'il n'y a aucun compétiteur à ce moment.

La variable binaire *Noël* indique les transactions effectuées durant les deux semaines précédant la fête de Noël. Cette variable a été ajoutée afin de faire ressortir les ventes effectuées durant *l'avant-Noël*, puisque ces semaines sont reconnues comme ayant des ventes plus élevées que durant les autres semaines de l'année.

2.1.3. – Base de données finale

Au final, la base de données servant à l'analyse avait un total de 689 136 observations provenant de 43 071 clients uniques qui ont effectué des transactions sur une période de 16 semaines. La variable dépendante est la dépense hebdomadaire totale du client. Pour ce qui est des variables indépendantes, voici un tableau récapitulatif énumérant l'ensemble de ces dernières.

Tableau 2: Variables indépendantes incluses dans la base de données finale

Nom de la variable indépendante	Type	Explication
Semaine	Nominale	Semaine = 0: 8 Semaines Pré-Ouverture Semaine = 1: 2 premières semaines 2 semaines après l'ouverture Semaine = 2: 3 semaines suivant la période 1 Semaine = 3: 3 semaines suivant la période 2
RFV	Nominale	RFV = 1: Clients du segment Premium RFV = 2: Clients du segment Standard RFV = 3: Clients des segments Low Lapsed et Gone Away
Compétiteur	Nominale	Compétiteur = 0: Pas de présence de compétiteur (Pré-ouverture) Compétiteur = 1: Présence du compétiteur B Compétiteur = 2: Présence du compétiteur A
Ratio_Departement_X	Ordinale	Ratio du total des dépenses pré-ouverture attribuable à chacun de 9 départements du magasin
Urbain_Rural	Nominale	Urbain_Rural = 1: Ville de plus de 50 000 habitants Urbain_Rural = 2: Ville de moins de 50 000 habitants
Compétiteur_Plus_Proche	Nominale	Comp_Plus_Proche = 0: Pas de compétiteur (Pré-ouverture) Comp_Plus_Proche = 1: Compétiteur est le magasin le plus près de l'individu Comp_Plus_Proche = 2: Magasin de la bannière étudiée est le plus près de l'individu
Noël	Nominale	Noël = 0: Pas durant la semaine précédant Noël Noël = 1: Durant la semaine précédent Noël

À ces variables, s'ajoute une variable de contrôle, la variable *household_id* permettant d'identifier l'individu.

2.2-Modélisation

L'objectif que se fixe ce mémoire est de déterminer les facteurs influençant la défection d'un client suite à l'arrivée d'un compétiteur en plus de quantifier leur effet sur les magasins étudiés. Dans la littérature, on aborde la défection de clientèle de deux façons. Soit on tente de créer un modèle prédictif à partir d'observations existantes pour prédire les clients qui feront défection dans le futur ou soit on tente plutôt d'expliquer les facteurs ayant un impact sur la défection via un modèle explicatif. Dans le cadre de ce travail, une modélisation de type explicative sera préconisée, car elle permettra de mieux répondre à l'objectif que s'est fixé ce mémoire.

De plus, il fut décidé d'étudier la défection partielle des clients, car cette dernière fut utilisée dans la littérature pour une problématique similaire touchant au commerce de détail alimentaire. D'un point de vue managérial, ce type de défection permettra à l'entreprise étudiée d'identifier les clients faisant défection dès qu'ils commenceront à diminuer leurs achats et non pas lorsqu'ils auront cessé complètement leur relation avec la bannière.

2.2.1. – Termes d'interaction

Afin de comprendre si l'effet combiné de variables indépendantes expliquait davantage la diminution des dépenses hebdomadaires que l'effet isolé de ces prédicateurs, des termes d'interaction double furent intégrés à la modélisation. Les termes d'interaction créés furent les suivants : 1- L'interaction de la variable de segmentation *RFV_Code* avec la variable *Compétiteur* afin de voir si des segments de clientèle sont plus ou moins impactés en fonction du type de compétiteur présent (A ou B). 2- L'interaction de la variable *RFV_Code* et de la variable *Urbain_Rural* afin de voir si les différents segments de clients sont touchés de la même façon en milieu urbain et rural. 3- L'interaction de la variable *Compétiteur_Plus_Proche* et la variable *Urbain_Rural* afin de voir si la proximité d'un compétiteur face au client a le même impact en milieu urbain et rural. 4- L'interaction de la variable *RFV_Code* avec la variable *Semaine* afin de voir comment les segments de clientèle sont impactés en fonction du temps. 5- L'interaction de la variable *Compétiteur* avec la variable *Semaine* afin de voir comment les clients sont touchés par l'arrivée de deux différents compétiteurs dans le temps. 6-L'interaction entre la variable *Compétiteur* et la variable *Compétiteur_Plus_Proche* afin de voir si l'impact de la proximité d'un compétiteur est le même selon le type de compétiteur en place. 7-Huit termes d'interaction entre tous les ratios d'achat des départements et le type de compétiteur (*Compétiteur*) présent afin d'évaluer si le genre de produits achetés avant l'arrivée du compétiteur (*Ratio_Département_X*) permet d'identifier les clients à risque de diminuer leurs achats en fonction du type de concurrent sur place.

2.1.2. – Modèle linéaire pour données corrélées

La modélisation réalisée consiste en une régression linéaire pour données corrélées et longitudinales. Tel que mentionné précédemment, les données étudiées n'ont pas été agrégées par client, mais fournissent plutôt de l'information sur chacune des seize semaines étudiées chez ce dernier. Cette particularité nous force à considérer qu'il existe une corrélation entre les transactions d'un même client effectuées durant les seize semaines d'analyse. Une régression linéaire ordinaire n'aurait pas pu être envisagée dans un contexte semblable, car cette dernière suppose que toutes les observations d'un échantillon sont indépendantes, ce qui n'est pas le cas dans ce travail. Laroque (2008) prévient du risque inhérent à ne pas considérer une corrélation intra-individu dans un échantillon. En effet, ce genre d'oubli augmente les chances de faussement rejeter une hypothèse nulle lorsqu'elle est vraie (erreur de type 1) et du même coup, empêche d'effectuer une inférence statistique efficace à l'aide des paramètres fournis par le modèle. Laroque (2008) démontre que ce risque a tendance à augmenter en fonction du nombre d'observations par groupe intra-individu, qui est au nombre de seize dans notre situation. Afin d'éviter ce genre de risque sur notre modèle, il fut décidé de modéliser directement la structure de covariance des clients étudiés à l'aide d'une structure de covariance appelée équi-corrélation. Cette dernière suppose qu'il existe une corrélation égale entre chacun des groupes intra-clients étudiés.

Du côté plus technique, c'est la procédure GENMOD du logiciel SAS qui fut utilisée pour réaliser la modélisation. L'option `subject= household_id` a permis d'identifier les différents groupes intra-clients et l'option `type=exch corrw` a permis de spécifier la structure de corrélation équivalente (équi-corrélation) entre les groupes.

Il est vrai que pour une modélisation de type régression linéaire mixte, différentes procédures telles que MIXED, GLIMMIX et HPMIXED auraient aussi pu être utilisées tout en considérant la corrélation intra-individu. Cependant, c'est la procédure GENMOD qui a été retenue pour ses qualités de rapidité de calcul sur des jeux de données volumineux (base de données finale ayant près de 700 000

observations). L'aspect longitudinal a pu être pris en considération à l'aide de la variable *Semaine* qui permettait de voir à quel moment la transaction du client avait été réalisée.

Ce genre d'approche de type « régression linéaire » possède aussi l'avantage de donner des résultats plus facilement interprétables dans un contexte longitudinal. Il suffit d'observer si les termes simples et les termes d'interaction sont significatifs afin de conclure s'ils ont un effet ou non sur la variable dépendante. De plus, les coefficients d'un modèle de type « régression linéaire » permettent de quantifier directement l'effet que ces variables significatives ont sur cette même variable dépendante.

CHAPITRE 3 : LES RÉSULTATS

Dans ce chapitre, nous présenterons les résultats obtenus suite aux différentes analyses effectuées sur la base de données finale. Tout d'abord, une analyse descriptive des variables prédictives les plus importantes donnera un aperçu global sur la façon dont ces dernières réagissent dans le temps. En deuxième lieu, une analyse des résultats proposés par le modèle permettra de répondre aux objectifs de recherche que se fixe ce mémoire.

3.1- Statistiques descriptives

3.1.1. – Effet sur les ventes des magasins

Tel que mentionné précédemment, la base de données finale comporte de l'information sur huit magasins ayant tous subi l'arrivée d'un nouveau compétiteur dans leur environnement. Puisque la moitié des magasins ont subi l'arrivée du compétiteur A et que l'autre moitié ont subi l'arrivée du compétiteur B, il était légitime de s'interroger sur la possibilité qu'un des compétiteurs impacte plus sévèrement les magasins de la bannière étudiée que l'autre compétiteur. Voici donc un aperçu de l'évolution des dépenses hebdomadaires avant et après l'arrivée des deux différents compétiteurs.

Figure 1 : Évolution des dépenses hebdomadaires des clients membres du programme de fidélité ayant vécu l'arrivée du compétiteur A.

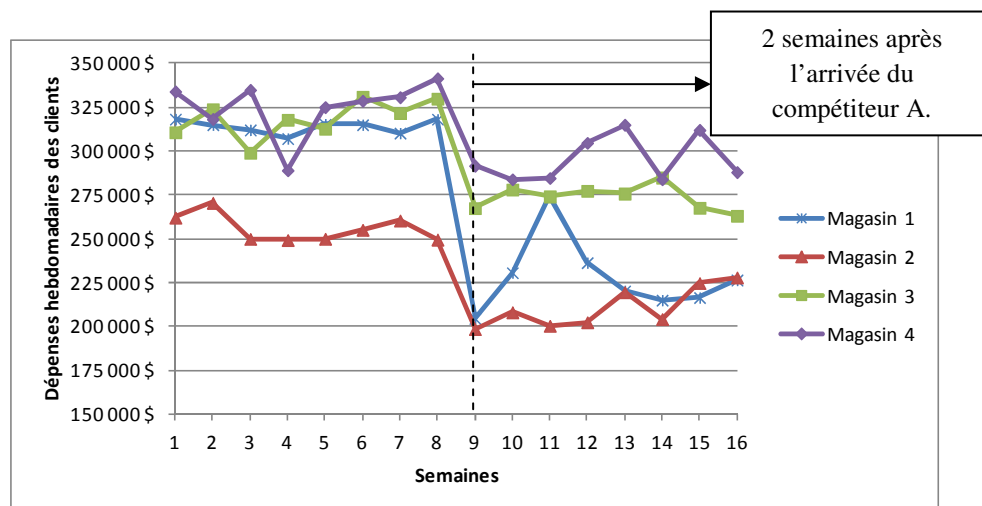


Tableau 3 : Moyennes des dépenses hebdomadaires des clients membres du programme de fidélité avant et après l'arrivée du compétiteur A.

Statistique sur les dépenses hebdomadaires	Magasin 1	Magasin 2	Magasin 3	Magasin 4
Moyenne des 8 semaines pré-ouverture	313 908,02 \$	255 843,69 \$	318 354,62 \$	325 277,52 \$
Moyenne des 8 semaines suivant les 2 premières semaines d'ouverture	227 998,29 \$	210 739,51 \$	273 580,94 \$	295 502,34 \$
Variation Avant/Après	-27%	-18%	-14%	-9%

Pour tous les magasins ayant subi l'arrivée du compétiteur A, il est possible de constater qu'il y a eu une diminution significative des dépenses hebdomadaires chez les clients étudiés. On constate cependant que même si le compétiteur est le même pour ces quatre magasins, l'impact n'est pas le même pour tous. Rappelons aussi que les deux premières semaines d'ouverture du compétiteur n'ont pas été incluses dans l'analyse, puisqu'elles sont reconnues par l'entreprise étudiée comme n'étant pas représentatives du climat compétitif habituel.

Pour ce qui est des magasins ayant subi l'arrivée du compétiteur B, on observe une tendance similaire. Voici un aperçu de l'évolution des dépenses des clients confrontés au second compétiteur :

Figure 2 : Évolution des dépenses hebdomadaires des clients membres du programme de fidélité ayant vécu l'arrivée du compétiteur B.

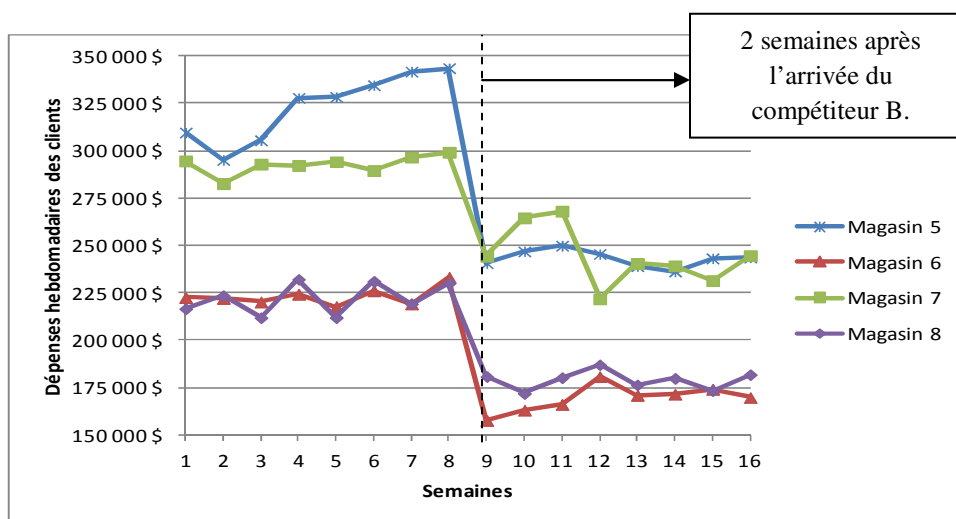


Tableau 4 : Moyennes des dépenses hebdomadaires des clients membres du programme de fidélité avant, pendant et après l'arrivée du compétiteur B.

Statistique sur les dépenses hebdomadaires	Magasin 5	Magasin 6	Magasin 7	Magasin 8
Moyenne des 8 semaines pré-ouverture	323 371,74 \$	223 191,10 \$	292 826,91 \$	222 130,00 \$
Moyenne des 8 semaines suivant les 2 premières semaines d'ouverture	243 215,84 \$	169 269,64 \$	244 386,67 \$	179 028,33 \$
Variation Avant/Après	-25%	-24%	-17%	-19%

À première vue, on constate que comme pour les magasins ayant été affectés par le compétiteur A, ceux ayant été affectés par le compétiteur B ont aussi subi une diminution significative des dépenses hebdomadaires suite à l'ouverture. En comparant l'impact du compétiteur A avec celui du compétiteur B, on observe une diminution plus forte sur les dépenses hebdomadaires des magasins impactés par le compétiteur B. L'arrivée du compétiteur B semble donc avoir un impact négatif plus grand sur les clients de ces magasins que pour ceux qui ont été affectés par le compétiteur A.

Après avoir observé l'effet global des deux différents compétiteurs sur les huit magasins étudiés, nous allons maintenant analyser l'impact lié aux différents départements.

3.1.2. – Effet sur les départements des magasins

Les compétiteurs A et B sont tous deux des magasins d'alimentation, mais ces derniers proposent deux concepts différents. Le compétiteur A est une entreprise qui met l'emphase sur les prix les plus bas à l'aide de produits de moins grande qualité en plus d'offrir moins de service en magasin. Le compétiteur B possède quant à lui un concept de magasin similaire à celui de la bannière étudiée, c'est-à-dire des magasins offrant du service et des produits de plus haute qualité. Pour cette raison, il y a lieu de s'interroger sur l'impact que ces compétiteurs ont sur les différents départements des

huit magasins étudiés. À noter que les deux premières semaines d'ouverture ne figurent pas dans les tableaux et figures suivantes.

3.1.2.1 – Effet sur les départements des magasins faisant face au compétiteur A

Les figures et les tableaux suivants permettront de comparer les dépenses effectuées dans chacun des départements durant les huit semaines pré-ouverture et les huit semaines post-ouverture.

Figure 3: Évolution des dépenses départementales chez les clients du magasin 1

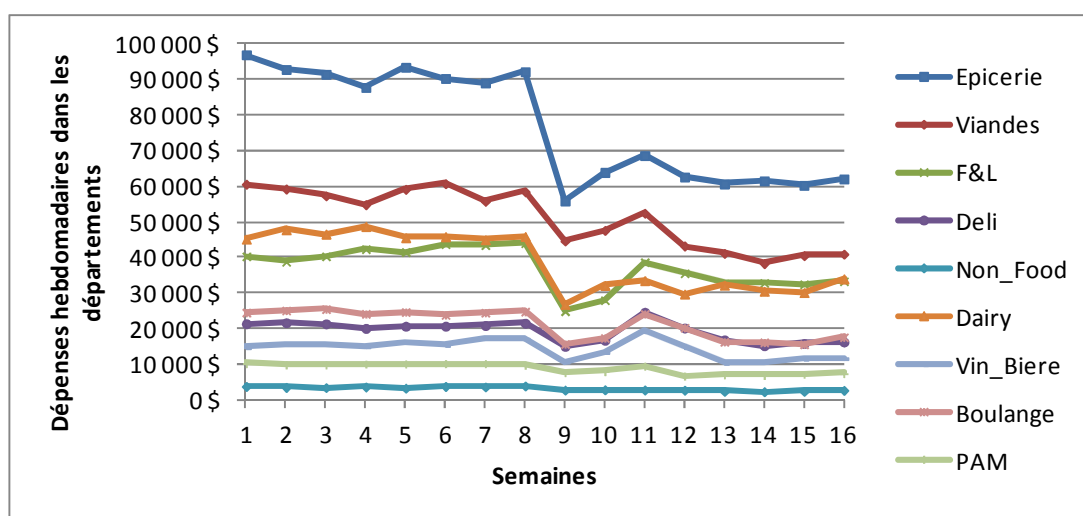


Tableau 5: Moyennes des dépenses départementales des clients du magasin 1

	Epicerie	Viandes	F&L	Deli	Non_Food	Dairy	Vin_Biere	Boulange	PAM
Moyenne des 8 semaines pré-ouverture	91 782 \$	58 311 \$	41 879 \$	21 114 \$	3 689 \$	46 363 \$	15 965 \$	24 707 \$	10 098 \$
Moyenne des 8 semaines post-ouverture	65 353 \$	45 310 \$	33 670 \$	18 107 \$	2 825 \$	32 837 \$	13 328 \$	18 731 \$	7 859 \$
Variation Avant/Après	-29%	-22%	-20%	-14%	-23%	-29%	-17%	-24%	-22%

Figure 4: Évolution des dépenses départementales chez les clients du magasin 2

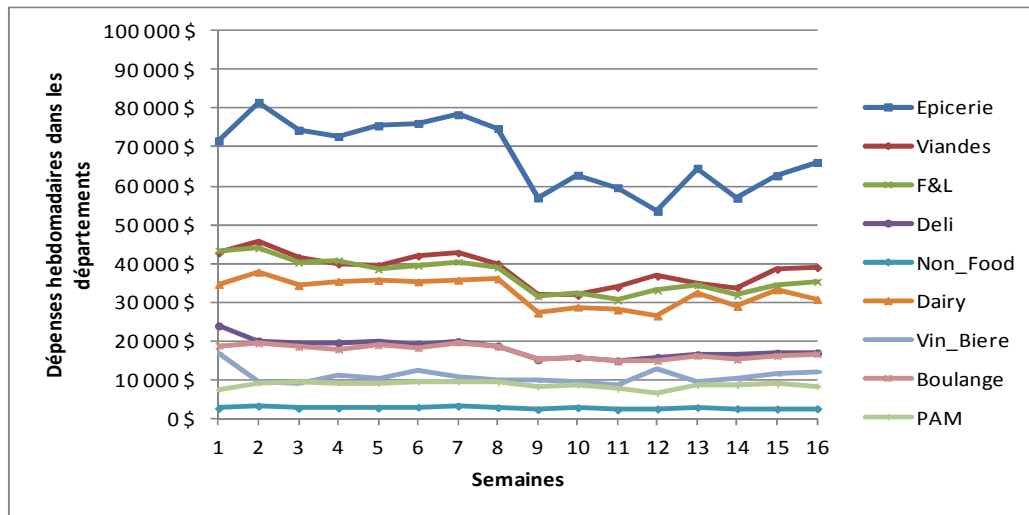


Tableau 6: Moyennes des dépenses départementales des clients du magasin 2

	Epicerie	Viandes	F&L	Deli	Non_Food	Dairy	Vin_Biere	Boulange	PAM
Moyenne des 8 semaines pré-ouverture	75 648 \$	41 744 \$	40 703 \$	20 061 \$	2 896 \$	35 613 \$	11 281 \$	18 739 \$	9 158 \$
Moyenne des 8 semaines post-ouverture	60 341 \$	35 033 \$	32 898 \$	16 054 \$	2 579 \$	29 496 \$	10 517 \$	15 629 \$	8 193 \$
Variation Avant/Après	-20%	-16%	-19%	-20%	-11%	-17%	-7%	-17%	-11%

Figure 5: Évolution des dépenses départementales chez les clients du magasin 3

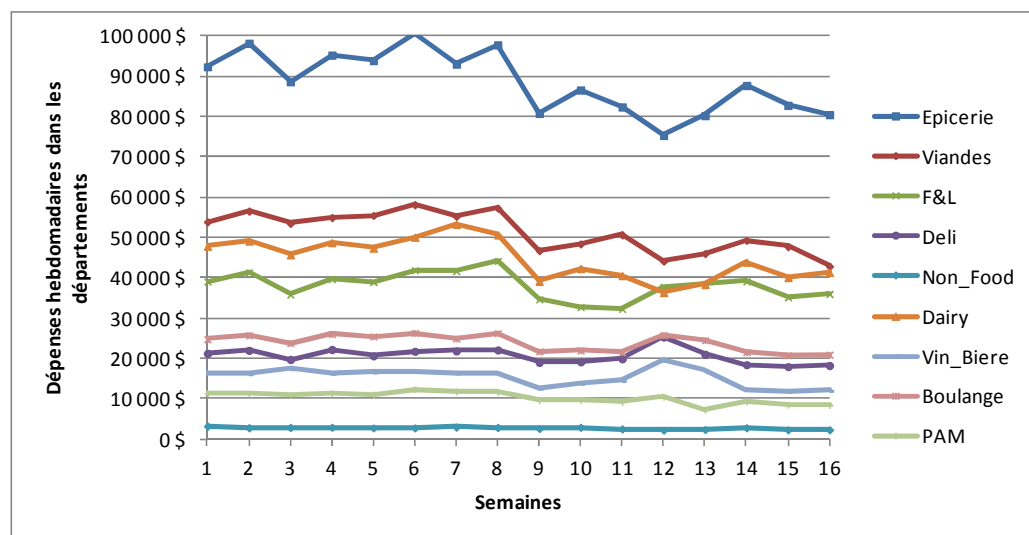


Tableau 7: Moyennes des dépenses départementales des clients du magasin 3

	Epicerie	Viandes	F&L	Deli	Non_Food	Dairy	Vin_Biere	Boulanges	PAM
Moyenne des 8 semaines pré-ouverture	94 946 \$	55 678 \$	40 391 \$	21 526 \$	2 989 \$	49 142 \$	16 624 \$	25 490 \$	11 568 \$
Moyenne des 8 semaines post-ouverture	82 061 \$	47 036 \$	35 836 \$	19 965 \$	2 578 \$	40 296 \$	14 236 \$	22 428 \$	9 144 \$
Variation Avant/Après	-14%	-16%	-11%	-7%	-14%	-18%	-14%	-12%	-21%

Figure 6: Évolution des dépenses départementales chez les clients du magasin 4

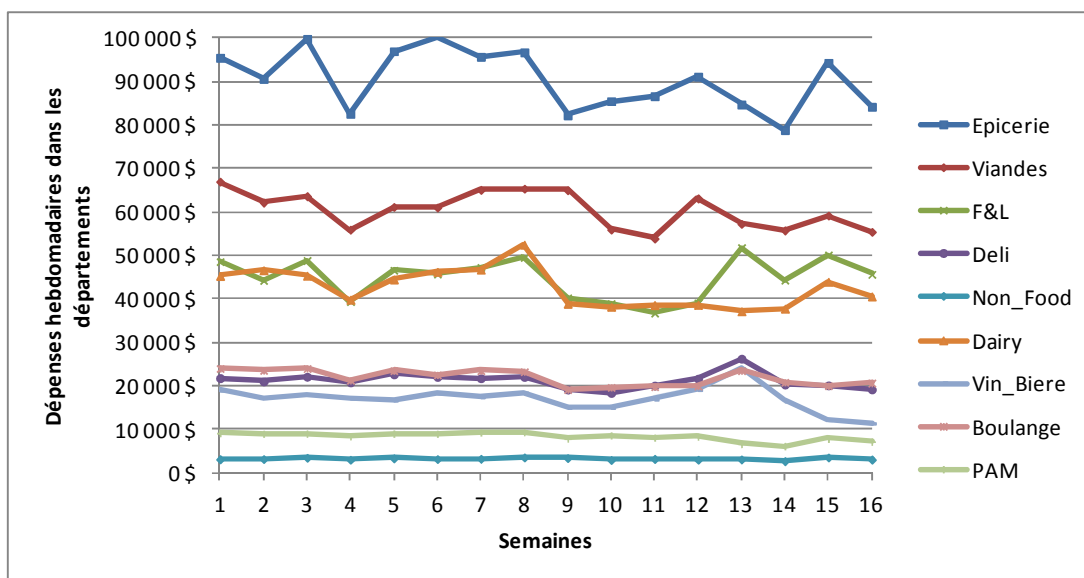


Tableau 8: Moyennes des dépenses départementales des clients du magasin 4

	Epicerie	Viandes	F&L	Deli	Non_Food	Dairy	Vin_Biere	Boulanges	PAM
Moyenne des 8 semaines pré-ouverture	94 762 \$	62 717 \$	46 333 \$	21 874 \$	3 364 \$	45 931 \$	17 903 \$	23 277 \$	9 118 \$
Moyenne des 8 semaines post-ouverture	85 977 \$	58 253 \$	43 437 \$	20 733 \$	3 215 \$	39 285 \$	16 380 \$	20 526 \$	7 697 \$
Variation Avant/Après	-9%	-7%	-6%	-5%	-4%	-14%	-9%	-12%	-16%

3.1.2.2 – Effet sur les départements des magasins faisant face au compétiteur B

Figure 7 : Évolution des dépenses départementales chez les clients du magasin 5

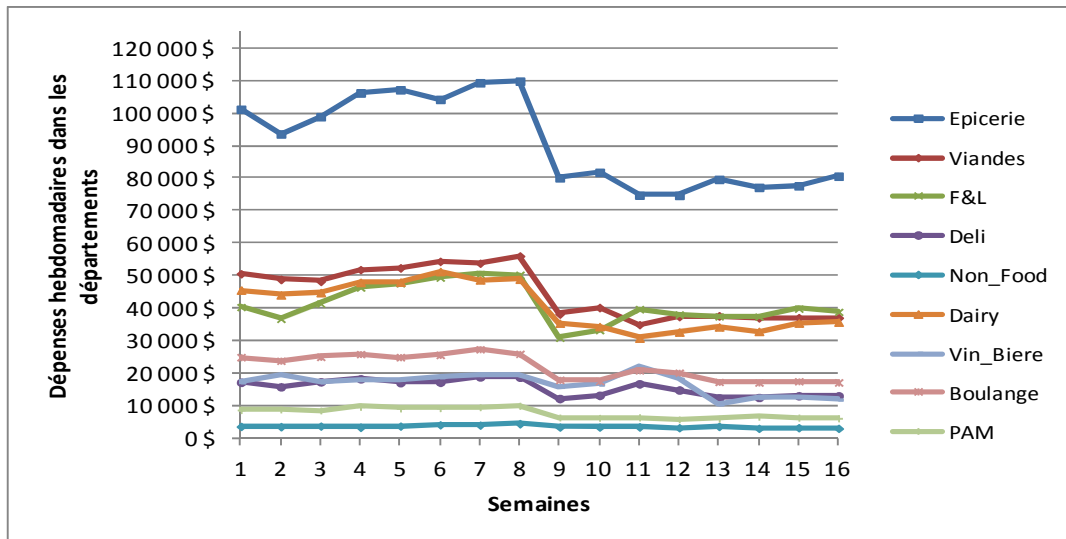


Tableau 9 : Moyennes des dépenses départementales des clients du magasin 5

	Epicerie	Viandes	F&L	Deli	Non_Food	Dairy	Vin_Biere	Boulanges	PAM
Moyenne des 8 semaines pré-ouverture	103 797 \$	52 056 \$	45 357 \$	17 602 \$	3 873 \$	47 428 \$	18 522 \$	25 478 \$	9 259 \$
Moyenne des 8 semaines post-ouverture	78 307 \$	37 476 \$	36 964 \$	13 541 \$	3 349 \$	33 947 \$	15 106 \$	18 247 \$	6 280 \$
Variation Avant/Après	-25%	-28%	-19%	-23%	-14%	-28%	-18%	-28%	-32%

Figure 8 : Évolution des dépenses départementales chez les clients du magasin 6

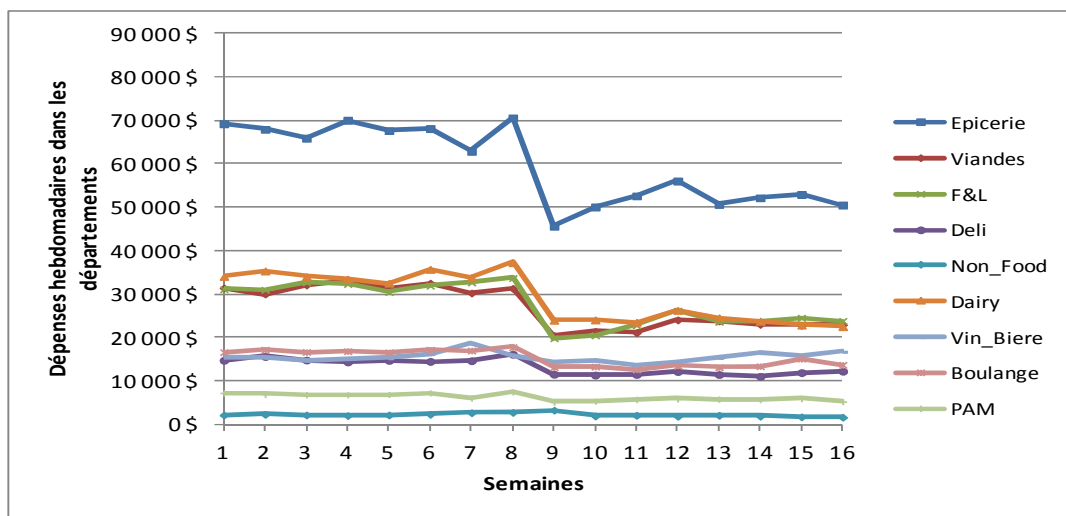


Tableau 10 : Moyennes des dépenses départementales des clients du magasin 6

	Epicerie	Viandes	F&L	Deli	Non_Food	Dairy	Vin_Biere	Boulanges	PAM
Moyenne des 8 semaines pré-ouverture	67 817 \$	31 507 \$	32 021 \$	14 995 \$	2 360 \$	34 548 \$	15 920 \$	17 028 \$	6 995 \$
Moyenne des 8 semaines post-ouverture	51 353 \$	22 498 \$	23 167 \$	11 698 \$	2 114 \$	23 973 \$	15 265 \$	13 548 \$	5 655 \$
Variation Avant/Après	-24%	-29%	-28%	-22%	-10%	-31%	-4%	-20%	-19%

Figure 9: Évolution des dépenses départementales chez les clients du magasin 7

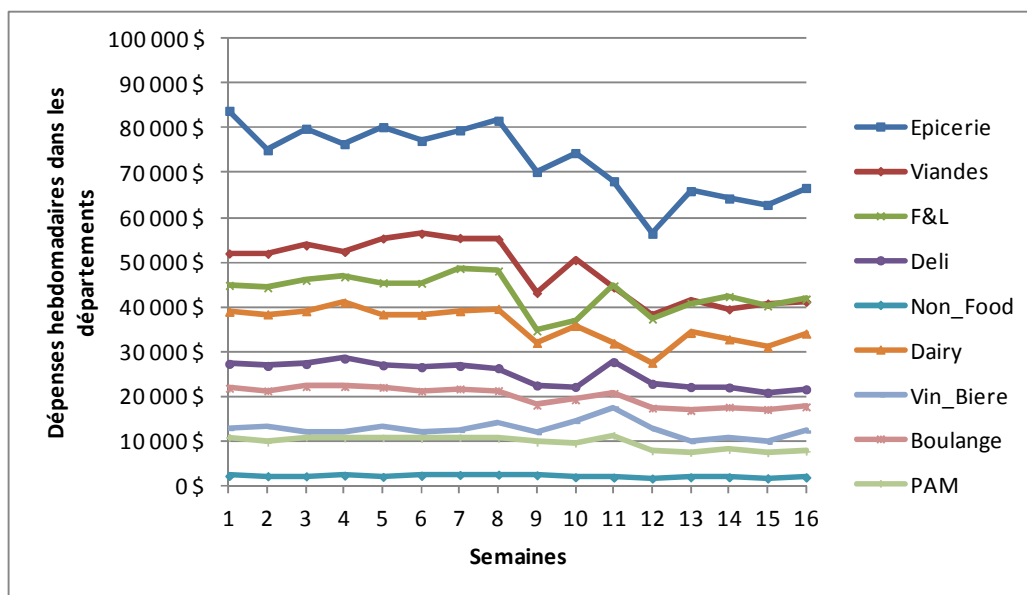


Tableau 11 : Moyennes des dépenses départementales des clients du magasin 7

	Epicerie	Viandes	F&L	Deli	Non_Food	Dairy	Vin_Biere	Boulanges	PAM
Moyenne des 8 semaines pré-ouverture	79 229 \$	54 033 \$	46 137 \$	27 056 \$	2 246 \$	39 081 \$	12 748 \$	21 702 \$	10 595 \$
Moyenne des 8 semaines post-ouverture	66 101 \$	42 373 \$	39 825 \$	22 697 \$	1 890 \$	32 393 \$	12 402 \$	18 070 \$	8 635 \$
Variation Avant/Après	-17%	-22%	-14%	-16%	-16%	-17%	-3%	-17%	-19%

Figure 10: Évolution des dépenses départementales chez les clients du magasin 8

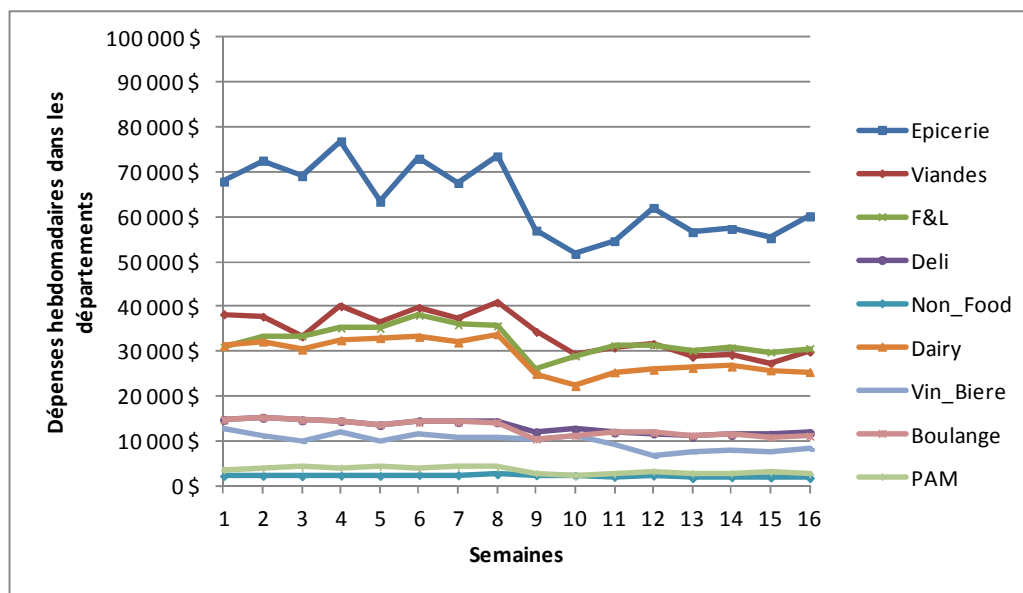


Tableau 12 : Moyennes des dépenses départementales des clients du magasin 8

	Epicerie	Viandes	F&L	Deli	Non_Food	Dairy	Vin_Biere	Boulange	PAM
Moyenne des 8 semaines pré-ouverture	70 477 \$	38 010 \$	34 728 \$	14 487 \$	2 377 \$	32 336 \$	11 183 \$	14 466 \$	4 067 \$
Moyenne des 8 semaines post-ouverture	56 841 \$	30 217 \$	29 951 \$	11 769 \$	2 102 \$	25 378 \$	8 603 \$	11 313 \$	2 855 \$
Variation Avant/Après	-19%	-21%	-14%	-19%	-12%	-22%	-23%	-22%	-30%

3.1.2.3 – Comparaison de l'effet sur les départements entre les magasins affectés par le compétiteur A et le compétiteur B

Afin de mieux observer les différences entre les variations départementales des magasins affectés par les deux compétiteurs, voici un tableau qui synthétise les moyennes de ces variations afin de voir si les départements semblent touchés de la même façon selon le compétiteur.

Tableau 13 : Moyennes des variations départementales avant/après l'arrivée des compétiteurs A et B

	Epicerie	Viandes	F&L	Deli	Non_Food	Dairy	Vin_Biere	Boulange	PAM
Moyenne des variations Avant/Après avec compétiteur A	-18%	-15%	-14%	-12%	-13%	-20%	-12%	-16%	-17%
Moyenne des variations Avant/Après avec compétiteur B	-21%	-25%	-18%	-20%	-13%	-24%	-12%	-22%	-25%

À la section 3.1.1, il a été possible d'observer que l'impact global sur les ventes des magasins affectés par le compétiteur B semblait plus fort que pour ceux affectés par le compétiteur A. Malgré cet impact global plus fort pour le compétiteur B, le tableau 13 permet de constater que ce portrait n'est pas identique pour tous les départements. En effet, on observe que le compétiteur B affecte de façon plus importante le département des viandes, du déli, de la boulange et du prêt à manger que le compétiteur A. Ce phénomène semble logique, car pour ces départements, le compétiteur A offrent des produits de qualité inférieure à ceux de la bannière étudiée. Il est donc possible de supposer que malgré un prix plus bas de ces produits chez le compétiteur A, la différence de qualité puisse être trop importante pour affecter l'entreprise de la même façon que l'a fait le compétiteur B. Pour les départements d'épicerie, de fruits et légumes, de non-alimentaire, de produits laitiers et de vins et bière, on constate que l'impact entre les deux compétiteurs est presque le même. Il aurait été attendu de voir un impact négatif plus grand par le compétiteur A, car ce dernier offre des produits de qualité similaire à ceux de la bannière étudiée et du compétiteur B pour ces départements, mais à un prix plus bas.

3.1.3 – Effet sur les segments de clientèle

Les figures et le tableau suivants permettent d'observer l'effet du temps sur les dépenses totales des différents segments de clients. Les résultats ont été séparés en fonction des deux types de compétiteur afin de voir si les segments se comportaient différemment dans les deux environnements différents. Rappelons que RFV1 représente le segment de clients le plus fidèles (*Premium*), que RFV2 représente le segment de clients moyennement fidèles (*Standard*) et que RFV3 représente le segment de clients les moins fidèles (*Low, Uncommitted, Lapsed et Gone Away*).

Figure 11: Évolution des dépenses hebdomadaires par segment de clientèle impacté par le compétiteur A

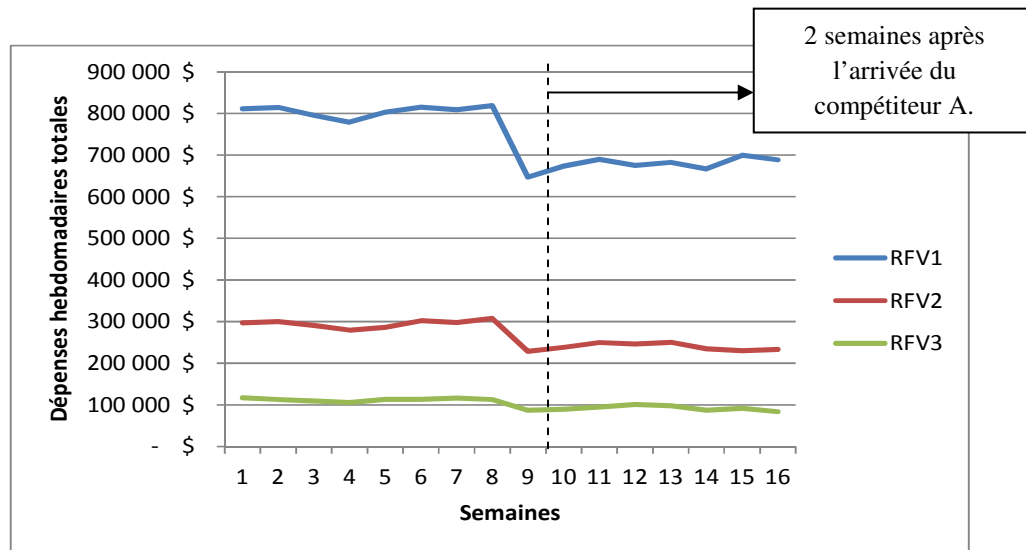


Tableau 14: Moyennes des dépenses hebdomadaires par segment de clientèle avant/après l'arrivée du compétiteur A

Statistique sur les dépenses hebdomadaires	Segment de clientèle		
	RFV1	RFV2	RFV3
Moyenne des 8 semaines pré-ouverture	805 832 \$	295 162 \$	112 389 \$
Moyenne des 8 semaines suivant les 2 premières semaines d'ouverture	677 886 \$	238 709 \$	91 226 \$
Variation Avant/Après	-16%	-19%	-19%

Figure 12: Évolution des dépenses hebdomadaires par segment de clientèle impacté par le compétiteur B

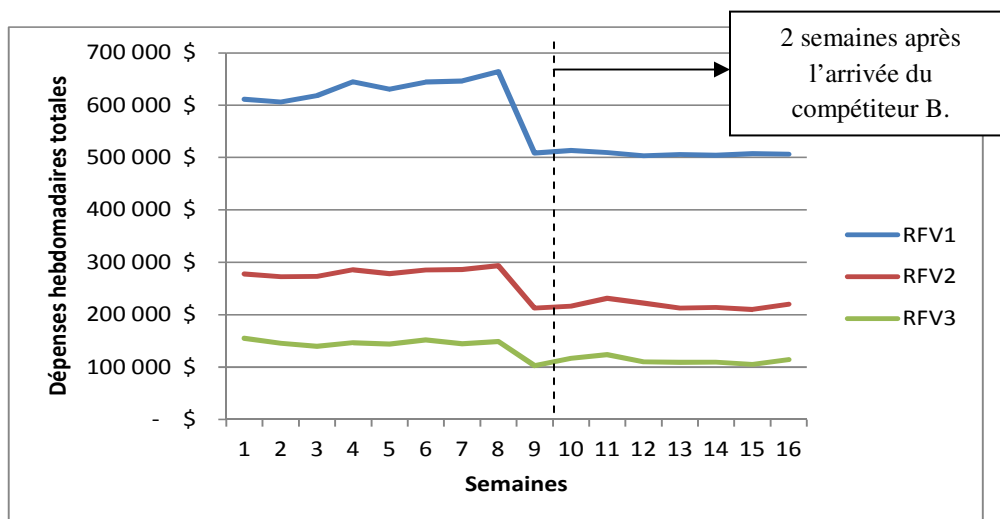


Tableau 15: Moyennes des dépenses hebdomadaires par segment de clientèle avant/après l'arrivée du compétiteur B

Statistique sur les dépenses hebdomadaires	Segment de clientèle		
	RFV1	RFV2	RFV3
Moyenne des 8 semaines pré-ouverture	633 125 \$	281 571 \$	146 824 \$
Moyenne des 8 semaines suivant les 2 premières semaines d'ouverture	507 388 \$	217 287 \$	111 226 \$
Variation Avant/Après	-20%	-23%	-24%

En regardant les figures et tableaux précédents, il est possible de dégager quelques tendances. Peu importe le type de compétiteur ayant fait son arrivée, on observe que le segment RFV1 représentant les clients les plus fidèles, semble légèrement moins impacté que les deux autres segments. On remarque aussi que le segment RFV3 représentant les clients les moins fidèles et le segment RFV2 représentant les clients moyennement fidèles semblent se comporter de façon semblable, mais avec une tendance un peu plus à la baisse lorsqu'ils sont en présence du compétiteur B.

3.1.4 – Effet sur les dépenses hebdomadaires selon le type de région (Urbain/Rural)

Les figures et le tableau suivants mettent en évidence l'évolution des dépenses hebdomadaires totales des clients par type d'environnement où se trouve leur magasin.

Figure 13: Évolution des dépenses hebdomadaires pour les magasins se trouvant en milieu urbain

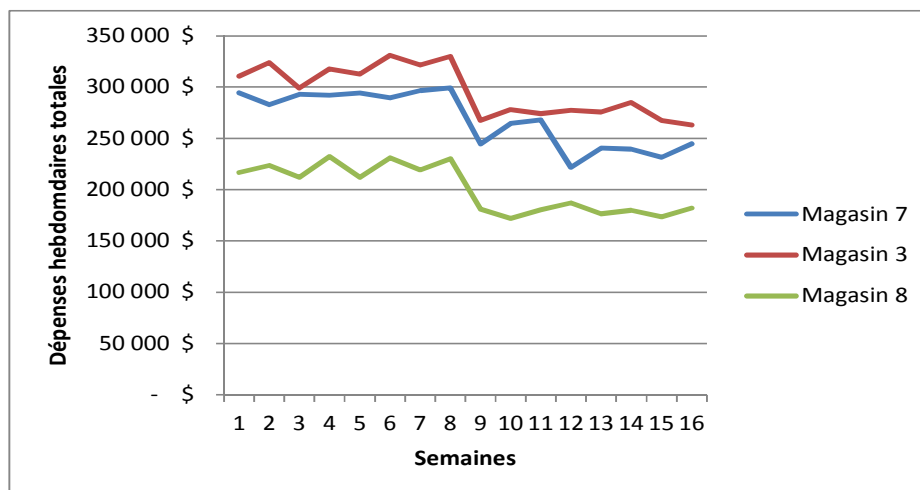


Figure 14: Évolution des dépenses hebdomadaires pour les magasins se trouvant en milieu rural

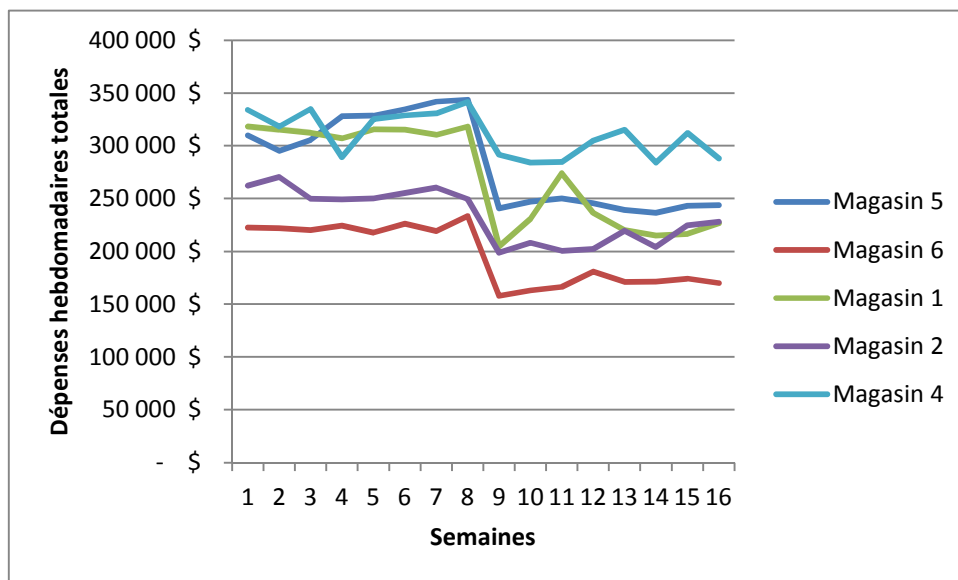


Tableau 16: Moyennes des dépenses hebdomadaires par type de milieu avant/après l'arrivée d'un compétiteur

	Urbain	Rural
Moyenne des 8 semaines pré-ouverture	277 771 \$	280 577 \$
Moyenne des 8 semaines post-ouverture	232 332 \$	229 345 \$
Variation Avant/Après	-16%	-18%

En comparant l'évolution des ventes en milieu rural par rapport à celles en milieu urbain, on observe une légère différence entre avant et après l'arrivée d'un compétiteur. En effet, les magasins en milieu rural semblent un peu plus affectés par l'arrivée d'un nouveau compétiteur que les magasins situés en milieu urbain.

3.1.5 – Effet sur les dépenses hebdomadaires selon la proximité du compétiteur

La figure et le tableau suivants mettent en évidence l'évolution des dépenses hebdomadaires totales des clients en fonction du type de magasin le plus à proximité de leur domicile. Dans ce cas-ci, le type de magasin fait référence au fait que ce soit un magasin de la compétition ou un magasin de la bannière étudiée qui soit situé le plus près des clients.

Figure 15: Évolution des dépenses hebdomadaires en fonction du magasin le plus près des clients étudiés

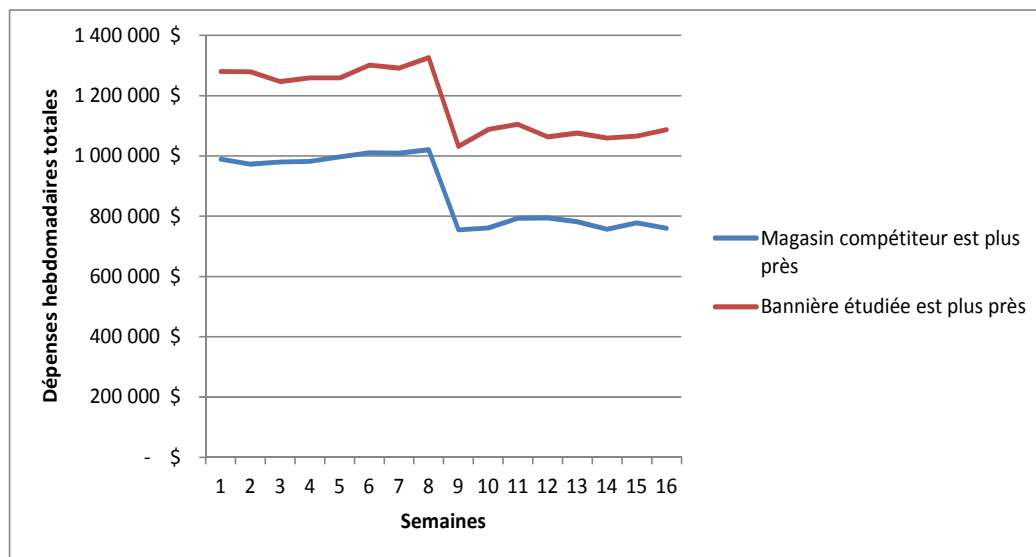


Tableau 17: Moyennes des dépenses hebdomadaires en fonction du type de magasin qui est le plus près des clients étudiés

Statistique sur les dépenses hebdomadaires	Lorsque le compétiteur est le magasin le plus près	Lorsque la bannière étudiée est le magasin le plus près
Moyenne des 8 semaines pré-ouverture	994 846 \$	1 280 058 \$
Moyenne des 8 semaines suivant les 2 premières semaines d'ouverture	772 064 \$	1 071 658 \$
Variation Avant/Après	-22%	-16%

La figure et le tableau précédents permettent de constater qu'il semble y avoir une différence du comportement d'achat des clients en fonction du type de magasin situé le plus près de leur domicile. En effet, l'impact de l'arrivée d'un compétiteur semble moins important chez les clients ayant un magasin de la bannière étudiée comme magasin situé le plus à proximité de leur domicile. Ce phénomène est en accord avec la connaissance qu'a l'entreprise sur ses consommateurs, puisqu'elle sait qu'une des principales raisons de sélection d'un magasin est le critère de proximité.

3.2- Résultats du modèle de régression linéaire sur données corrélées

Tel que mentionné dans le chapitre précédant, le modèle que nous avons choisi afin d'analyser l'influence des différentes variables sur la défection de clientèle est un modèle de régression linéaire. Ce modèle de régression linéaire se distingue d'un modèle habituel par le fait qu'il prend en considération la corrélation existante entre les transactions effectuées par un même client en plus de considérer l'aspect longitudinal des transactions effectuées pendant les seize semaines d'observation. La prochaine section portera un regard sur l'effet qu'ont les variables simples ainsi que les variables d'interaction double et triple sur la variable dépendante (les dépenses hebdomadaires des clients). Les termes d'interaction ont été créés en fonction d'hypothèses de recherche ayant émanées de l'analyse des statistiques descriptives de la section précédente. Les interactions doubles avec la variable Noël ont été écartées du modèle, car cette dernière servait uniquement à contrôler l'effet des ventes de la période des fêtes. De plus, certains termes d'interaction double comme *Compétiteur_Plus_Proche*Semaine* et *Urbain_Rural*Semaine* ont aussi été écartés du modèle car ils ont été jugés comme des effets sur lesquels l'entreprise ne pourra pas prendre action directement. Voici un tableau résumant l'ensemble des effets analysés dans le modèle proposé.

Tableau 18: Ensemble des effets simples et des interactions faisant partie du modèle

Effets simples	Interaction double	Interaction triple
RFV_Code	RFV_Code*Semaine	Semaine*RFV*Compétiteur
Semaine	Semaine*Compétiteur	
Compétiteur	Compétiteur_Plus_Proche*Urbain_Rural	
Noël	Compétiteur*Compétiteur_Plus_Proche	
Compétiteur_Plus_Proche	Ratio_Département*Compétiteur	
Ratio_Département		
Urbain_Rural		

3.2.1 – Effets des variables simples et de leur interaction double

RFV Code : Baesens et al. (2008) avaient trouvé dans leur étude sur le milieu bancaire, que leur variable de type RFV arrivait à expliquer de façon significative la défection de clients. Dans le cadre de ce travail de recherche, on peut aussi conclure que la variable de type RFV, du nom de *RFV_Code*, possède un effet significatif sur la variable dépendante. Avant de s'attarder aux termes d'interaction en lien avec la variable *RFV_Code*, débutons par l'analyse de son effet individuel. Pour cette analyse, le segment *RFV_Code_1* (regroupant les clients les plus fidèles) a été choisi comme segment de référence. De plus, l'effet étudié pour cette variable est celui dans un contexte où les variables d'interaction *Semaine*RFV_Code* et *Compétiteur*RFV_Code* sont dans leurs catégories de référence, c'est-à-dire durant la période précédant l'arrivée de la compétition.

Dans le contexte décrit précédemment, on remarque que le segment *RFV_Code_2* (regroupant les clients à fidélité intermédiaire) possède une dépense hebdomadaire moyenne inférieure au segment *RFV_Code_1* de l'ordre de 54,83\$. Pour ce qui est du segment *RFV_Code_3* (regroupant les clients les moins fidèles), il possède une dépense hebdomadaire moyenne inférieure de 81,32\$ face au segment de référence des clients les plus fidèles. Ces différences entre les segments confirment donc que moins ils sont fidèles à l'entreprise, moins ils dépensent chez elle. Le tableau 19 illustre la valeur des différents paramètres de la variable dans le modèle. À noter cependant que malgré la significativité de cette variable, il est important d'interpréter ces paramètres avec prudence. Avec une ordonnée à l'origine ayant une valeur de 62,30, les paramètres laissent présager qu'un client du segment *RFV_Code_3* dépense en moyenne 19,02\$ de moins par semaine que le segment de référence *RFV_Code_1* ($62,30\$ - 81,32\$ = 19,02\$$) lorsque toutes les autres variables sont nulles. Il faut cependant garder en tête que l'ordonnée à l'origine utilisée dans ce calcul représente une importante extrapolation, puisqu'elle concerne un client achetant la totalité de ses produits dans le département du prêt-à-manger, durant la période de Noël, dans un milieu rural et ce avant l'arrivée de compétiteur (contexte où l'ensemble des variables sont dans leur état de référence).

Tableau 19: Estimation du paramètre *RFV_Code* (avec *RFV_Code* = 1 comme référence)

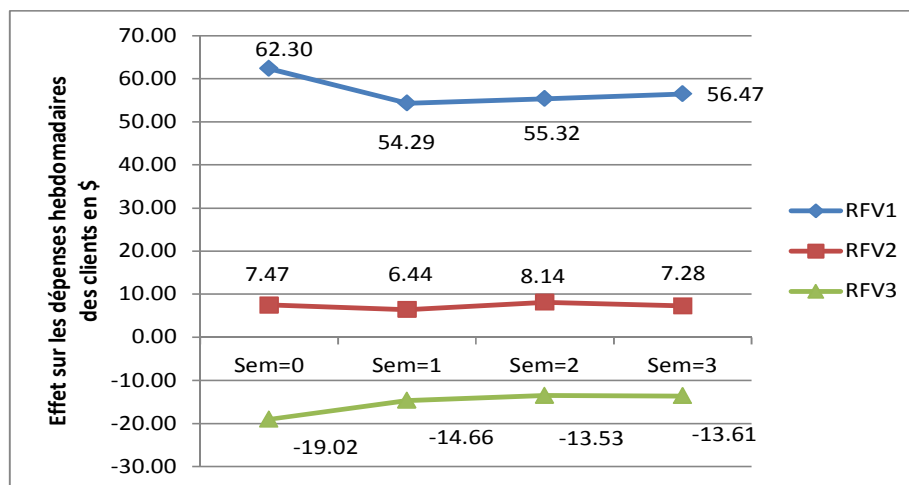
Analyse des valeurs estimées des paramètres GEE							
Paramètre		Valeur estimée	Erreur type	Intervalle de confiance à 95 %		Z	Pr > Z
<i>RFV_Code</i>	3	-81.3214	0.6194	-82.5336	-80.1055	-131.28	<.0001
<i>RFV_Code</i>	2	-54.8306	0.6592	-56.1158	-53.5318	-83.17	<.0001

Semaine: La variable *Semaine* qui devait valider l'existence d'un effet temporel sur les dépenses hebdomadaires des clients s'est aussi révélée significative. Pour analyser son effet, la modalité *Semaine_0* représentant la période pré-ouverture du compétiteur fut utilisée comme période de référence. L'effet analysé est donc celui à la période pré-ouverture lorsque les variables d'interaction *Semaine*RFV_Code* et *Semaine*Compétiteur* sont dans leur catégorie de référence. Dans ce cas-ci, cela signifie en présence du compétiteur A pour les clients les plus fidèles (*RFV_Code_1*). Le tableau 20 permet de voir que la période *Semaine_1* représentant les 2 premières semaines post-ouverture de la compétition est la période qui semble toucher le plus les dépenses hebdomadaires des clients avec une diminution moyenne de 8,01\$. La période *Semaine_2* représentant les 3 semaines suivant la période 1 semble un peu moins touchée que la période précédente avec une diminution moyenne de 6,98\$. On observe le même scénario avec la période *Semaine_3* représentant les 3 semaines suivant la période 2 qui elle aussi semble avoir un impact inférieur à la période précédente avec une diminution moyenne de 5,83\$. Ces résultats démontrent que le temps semble avoir un certain effet sur les dépenses hebdomadaires des clients. Bien qu'on observe une diminution des dépenses hebdomadaires moyennes à partir du moment où la compétition ouvre ses portes (*Semaine=1*), l'effet semble se stabiliser durant les six semaines suivantes (*Semaine=2* & *Semaine=3*). Voici un tableau illustrant la valeur des différents paramètres de cette variable dans le modèle. La représentation graphique de l'effet *Semaine* est représentée à la figure 17 via la ligne *RFV1* représentant la modalité de référence.

Tableau 20: Estimation du paramètre *Semaine* (avec *Semaine* = 0 comme référence)

Analyse des valeurs estimées des paramètres GEE							
Paramètre		Valeur estimée	Erreur type	Intervalle de confiance à 95 %		Z	Pr > Z
Semaine	1	-8.0074	1.4285	-11.2770	-5.6774	-5.93	<.0001
Semaine	2	-6.9777	1.4221	-10.2364	-4.6619	-5.24	<.0001
Semaine	3	-5.8318	1.4139	-9.0722	-3.5299	-4.46	<.0001

*Semaine*RFV Code*: Le tableau 19 montre que la variable individuelle *RFV_Code* est significative et que les clients les plus fidèles (*RFV_Code_1*) dépensent plus que tous les autres type de clients. Afin de voir si cette variable permet d'identifier des clients à risque de quitter l'entreprise après l'arrivée d'un compétiteur, il importe d'analyser son effet combiné avec le temps (variable *Semaine*). Cette interaction s'est révélée significative et la figure suivante permet de visualiser la façon dont les différents segments évoluent dans le temps.

Figure16: Effet du temps sur les différents segments de clients étudiés (le temps est représenté par la variable *Semaine* et les segments par la variable *RFV_Code*)

À noter que sur la figure précédente, la modalité de référence est le segment *RFV_Code_1* (clients les plus fidèles), à la période *Semaine_0* (avant l'arrivée de la compétition). En analysant tout d'abord le segment *RFV_Code_1*, on remarque qu'il semble être le plus touché par l'arrivée d'un compétiteur se produisant à partir de la période *Semaine_1*. On remarque qu'avant l'arrivée du compétiteur (*Semaine_0*), le segment *RFV_Code_1* possède une dépense hebdomadaire moyenne de 62,30\$. Cette moyenne tend cependant à chuter à 54,29\$ lors des 2 premières semaines où le compétiteur arrive (*Semaine_1*), mais remonte période après période jusqu'à 56.47\$. Malgré que le segment *RFV_Code_1* semble être le plus touché des trois, il demeure clairement celui ayant la dépense hebdomadaire la plus élevée à tout moment de l'analyse.

Pour ce qui est du segment *RFV_Code_2* (clients moyennement fidèles), on remarque que par rapport à la modalité de référence, ce dernier affichait une dépense hebdomadaire moyenne de 7,47\$ avant l'arrivée du compétiteur. Une fois le compétiteur arrivé (*Semaine_1*), on remarque que la moyenne des dépenses décroît légèrement à 6,44\$ pour ensuite remonter légèrement à 8,14\$ à la période *Semaine_2* et finalement retomber à 7,28\$ à la période *Semaine_3*. La faible variation dans le temps des dépenses hebdomadaires moyennes du segment *RFV_Code_2* indique que l'arrivée d'un compétiteur n'a pas d'impact significatif sur leurs dépenses hebdomadaires.

Finalement pour le segment *RFV_Code_3* (clients les moins fidèles), on note que contrairement au segment *RFV_Code_1*, une légère augmentation des dépenses hebdomadaires moyennes survient entre la période pré et post arrivée du compétiteur. En effet, la dépense hebdomadaire moyenne passe de -19,02\$ à -14,66\$ entre la période *Semaine_0* et *Semaine_1* pour ensuite augmenter légèrement à -13,53\$ à la *Semaine_2* et à -13,61\$ pour la dernière période. À noter cependant qu'il faut interpréter le résultat de *RFV_Code_3* avec grande précaution. Après avoir roulé à nouveau le modèle avec *RFV_Code_3* comme catégorie de référence, il fut constaté

que l'effet simple de la variable temporelle *Semaine* n'était pas entièrement significatif. En effet les modalités *Semaine1* et *Semaine2* affichaient des p-values respectives de 0,08 et de 0,06 alors que *Semaine3* était de 0,01. Il est donc faux de croire que l'arrivée du compétiteur ait pu faire augmenter la dépense hebdomadaire des clients *RFV_Code_3*.

Bien que ces moyennes hebdomadaires soient calculées à partir de toutes les modalités de référence (*RFV_Code_1*, à la période *Semaine_0*, hors période de Noël, sans compétiteur), elles permettent de voir que le segment *RFV_Code_1* a ressenti une baisse de ses dépenses avec l'arrivée d'un compétiteur alors que les segments *RFV_Code_2* et *RFV_Code_3* ne semblent pas avoir été affectés par cette arrivée. Voici un tableau illustrant la valeur des différents paramètres de cette variable d'interaction dans le modèle.

Tableau 21: Estimation du paramètre *Semaine*RFV Code* (avec *RFV Code* = 1 et *Semaine=0* comme référence)

Analyse des valeurs estimées des paramètres GEE								
Paramètre			Valeur estimée	Erreur type	Intervalle de confiance à 95 %		Z	Pr > Z
<i>Semaine*RFV_Code</i>	1	2	6.9860	0.6707	5.6248	8.2541	10.35	<.0001
<i>Semaine*RFV_Code</i>	1	3	12.3734	0.5957	11.0473	13.3824	20.51	<.0001
<i>Semaine*RFV_Code</i>	2	2	7.6505	0.6548	6.3205	8.8872	11.61	<.0001
<i>Semaine*RFV_Code</i>	2	3	12.4767	0.5857	11.1704	13.4662	21.03	<.0001
<i>Semaine*RFV_Code</i>	3	2	5.6415	0.6202	4.3793	6.8105	9.02	<.0001
<i>Semaine*RFV_Code</i>	3	3	11.2438	0.5558	9.9961	12.1748	19.94	<.0001

Compétiteur : La variable *Compétiteur* signifiant la présence du compétiteur A ou B ne s'est pas révélée significative dans notre modèle. Cela signifie que l'on ne peut pas conclure que la seule présence d'un type de compétiteur (A ou B) affecte davantage les dépenses hebdomadaires de la bannière étudiée dans un contexte où les variables d'interaction *Compétiteur*RFV_Code* et *Compétiteur*Compétiteur_Plus_Proche* sont dans leurs catégories de référence, c'est-à-dire lorsque le magasin de la bannière

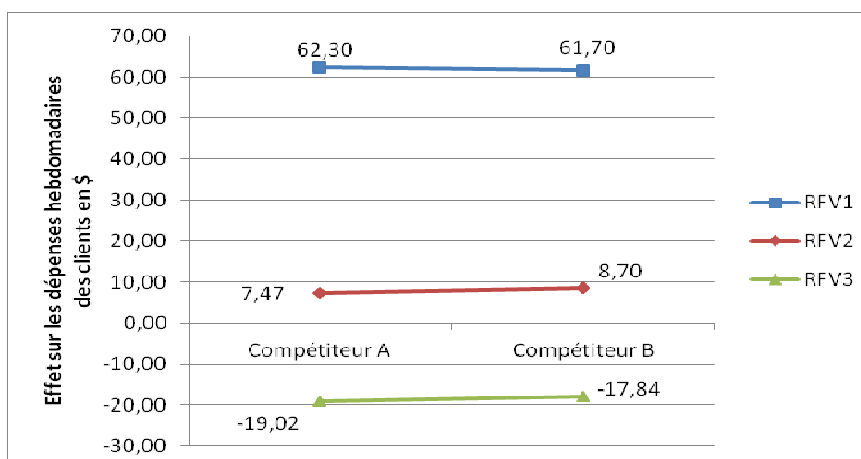
étudiée est le plus près du client et que ce dernier fait parti du segment de clients les plus fidèles. Voici un tableau illustrant la valeur des différents paramètres de cette variable dans le modèle.

Tableau 22: Estimation du paramètre *Compétiteur* (avec *Compétiteur=2* comme référence)

Analyse des valeurs estimées des paramètres GEE							
Paramètre		Valeur estimée	Erreur type	Intervalle de confiance à 95 %		Z	Pr > Z
Compétiteur	0	0.0000	0.0000	0.0000	0.0000	.	.
Compétiteur	1	-0.5938	1.7553	-4.0342	2.8466	-0.34	0.7351

*RFV Code*Compétiteur* : La dernière interaction qui fut mesurée avec la variable de segmentation *RFV_Code* est celle avec les différents types de compétiteurs. Cette interaction fut trouvée comme étant significative. La figure suivante montre l'effet de cette interaction sur les dépenses hebdomadaires des clients.

Figure 17: Effet de la présence du compétiteur A ou B sur le type de segment de clientèle (le type de compétiteur est représenté par la variable *Compétiteur* et le type de segment de clientèle par la variable *RFV_Code*)



Les effets de la variable d'interaction *RFV_Code* Compétiteur* sont présentés en fonction de la modalité de référence représentant le segment de clients les plus fidèles (*RFV_Code_1*) exposé au compétiteur B. La figure précédente permet de constater qu'il semble y avoir une légère différence entre la présence du compétiteur A ou B selon le type de segment. En effet, les segments moyennement et peu fidèles (RFV2 et RFV3), semblent légèrement moins affectés par la présence du compétiteur B que par celle du compétiteur A. À l'inverse, le segment le plus fidèle (RFV1) semble quant à lui plus affecté par le compétiteur B. Malgré la significativité du paramètre d'interaction, les différences entre le type de compétiteur restent assez faibles pour l'ensemble des segments. Il faut donc interpréter ces résultats avec précaution. Voici un tableau illustrant la valeur des différents paramètres de cette variable d'interaction dans le modèle.

Tableau 23: Estimation du paramètre *RFV_Code*Compétiteur* (avec *RFV_Code = 1* et *Compétiteur=2* comme référence)

Analyse des valeurs estimées des paramètres GEE								
Paramètre			Valeur estimée	Erreur type	Intervalle de confiance à 95 %		Z	Pr > Z
<i>RFV_Code*Compétiteur</i>	2	1	1.8230	0.7726	0.3254	3.3540	2.38	0.0173
<i>RFV_Code*Compétiteur</i>	3	1	1.7795	0.6887	0.5237	3.2232	2.72	0.0065

*Semaine*Compétiteur*: Bien que l'effet individuel de la variable *Compétiteur* ne s'est pas révélé significatif, il fut décidé d'explorer l'effet combiné de *Compétiteur* avec la variable temporelle *Semaine*, puisque cette dernière s'était révélée significative. L'effet combiné de ces variables ne s'est cependant pas révélé significatif. Il n'ajoute ainsi aucun effet supplémentaire à ceux des deux variables étudiées séparément. Voici un tableau illustrant la valeur des différents paramètres de cette variable d'interaction dans le modèle. Ce dernier montre que certaines d'interaction ont soit un effet nul sur le modèle (valeur estimée du paramètre = 0) ou un effet non-significatif ($Pr > |z|$ supérieur à 0,05).

Tableau 24: Estimation du paramètre *Semaine*Compétiteur* (avec *Semaine=0* et *Compétiteur=0* comme référence)

Analyse des valeurs estimées des paramètres GEE								
Paramètre			Valeur estimée	Erreur type	Intervalle de confiance à 95 %		Z	Pr > Z
<i>Semaine*Compétiteur</i>	3	1	-0.1307	0.4164	-0.9468	0.6853	-0.31	0.7535
<i>Semaine*Compétiteur</i>	3	2	0.0000	0.0000	0.0000	0.0000	.	.
<i>Semaine*Compétiteur</i>	2	1	-0.0272	0.4547	-0.9183	0.8640	-0.06	0.9524
<i>Semaine*Compétiteur</i>	2	2	0.0000	0.0000	0.0000	0.0000	.	.
<i>Semaine*Compétiteur</i>	1	1	0.0000	0.0000	0.0000	0.0000	.	.
<i>Semaine*Compétiteur</i>	1	2	0.0000	0.0000	0.0000	0.0000	.	.

Noël: Tel que mentionné dans le chapitre 2, la variable *Noël* a été créée afin d'identifier la période de deux semaines précédant la fête de Noël étant reconnues comme des semaines où les ventes sont beaucoup plus élevées que pendant le reste de l'année. Cette variable s'est effectivement révélée significative avec un effet négatif de 2,85\$ sur les ventes hebdomadaires moyennes effectuées hors période *Noël* en comparaison avec celles effectuées pendant la période *Noël* qui fut établie comme période de référence. Voici un tableau illustrant la valeur des différents paramètres de cette variable dans le modèle.

Tableau 25: Estimation du paramètre *Noël* (avec *Noël =1* comme référence)

Analyse des valeurs estimées des paramètres GEE								
Paramètre			Valeur estimée	Erreur type	Intervalle de confiance à 95 %		Z	Pr > Z
<i>Noel</i>	0		-2.8451	0.2670	-3.3684	-2.3219	-10.66	<.0001

Compétiteur Plus Proche: La variable permettant d'identifier qui du compétiteur ou du magasin de la bannière étudiée se situe le plus près du client ne s'est pas révélée significative. Il faut ainsi comprendre que cette variable n'est pas significative dans un contexte où les variables *Compétiteur Plus Proche*Urbain Rural* et

*Compétiteur_Plus_Proche*Compétiteur* sont dans leurs catégories de référence, c'est-à-dire lorsqu'on se trouve en milieu rural et en présence du compétiteur B. Voici un tableau illustrant la valeur des différents paramètres de cette variable dans le modèle. La valeur nulle de la modalité 0 indique qu'il n'y a pas de différence entre la présence du compétiteur B (modalité de référence) et le fait de ne pas avoir de compétition (*Compétiteur_Plus_Proche* = 0).

Tableau 26: Estimation du paramètre *Compétiteur Plus Proche* (avec *Compétiteur Plus Proche* =2 comme référence)

Analyse des valeurs estimées des paramètres GEE							
Paramètre		Valeur estimée	Erreur type	Intervalle de confiance à 95 %		Z	Pr > Z
<i>Compétiteur_plus_proche</i>	0	0.0000	0.0000	0.0000	0.0000	.	.
<i>Compétiteur_plus_proche</i>	1	0.3995	0.4917	-0.5643	1.3633	0.81	0.4166

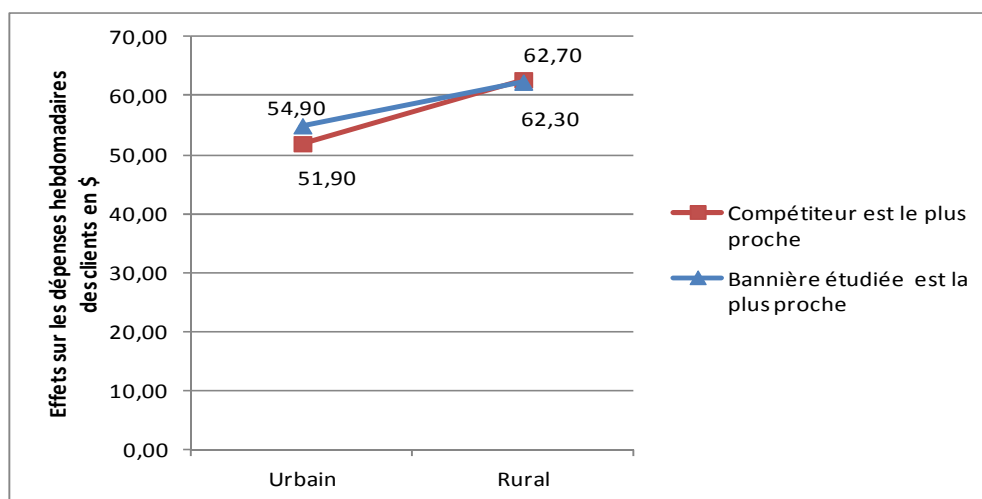
Urbain_Rural: La variable permettant de déterminer si les magasins étudiés se trouvaient dans un milieu rural ou urbain s'est révélée significative. Plus précisément, on remarque que le fait qu'un magasin se trouve en milieu urbain impacte davantage la dépense hebdomadaire des clients (-7,41\$) qu'un magasin se trouvant en milieu rural. Cette significativité est valable dans un contexte où la variable d'interaction *Compétiteur_Plus_Proche*Urbain_Rural* se trouve dans sa catégorie de référence, c'est-à dire lorsque le magasin de la bannière étudiée est le plus près du client. Voici un tableau illustrant la valeur des différents paramètres de cette variable dans le modèle.

Tableau 27: Estimation du paramètre *Urbain Rural* (avec *Urbain Rural* =2 comme référence)

Analyse des valeurs estimées des paramètres GEE							
Paramètre		Valeur estimée	Erreur type	Intervalle de confiance à 95 %		Z	Pr > Z
URBAIN_RURAL	1	-7.4050	0.9768	-9.3194	-5.4905	-7.58	<.0001

Compétiteur Plus Proche*Urbain Rural : La variable *Compétiteur Plus Proche* a été mise en interaction avec la variable *Urbain_Rural* afin d'analyser l'effet de la plus grande proximité d'un magasin face à un client dans un contexte urbain et rural. Contrairement à l'effet individuel de *Compétiteur Plus Proche*, l'interaction avec *Urbain_Rural* s'est plutôt révélée significative. La figure suivante illustre l'effet de cette interaction sur les ventes hebdomadaires des clients.

Figure 18: Effet de la proximité du magasin sur le type de ville où il se trouve (la proximité du magasin est représentée par la variable *Compétiteur Plus Proche* et le type de ville par la variable *Urbain_Rural*)



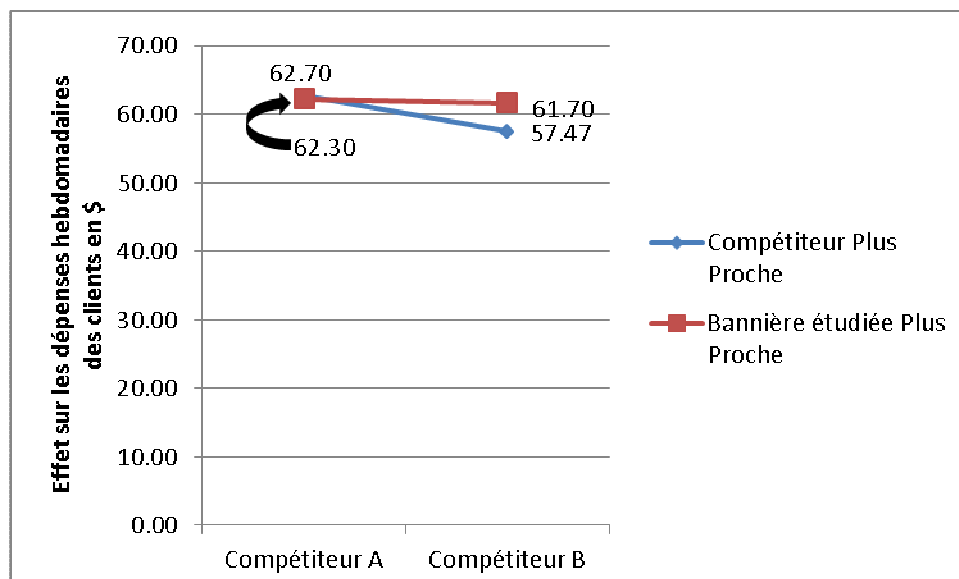
La modalité de référence choisie pour la figure 19 est celle où le magasin de la bannière étudiée est le plus près du client et se trouve en milieu rural. Dans ce milieu, on remarque que le type de magasin étant le plus rapproché des clients n'impacte pas différemment ces derniers (62,30\$ vs 62,70\$). En milieu urbain, l'impact semble cependant un peu plus négatif lorsque c'est le compétiteur qui est situé plus près des consommateurs (51,90\$ vs 54,90\$). Voici un tableau illustrant la valeur des différents paramètres de cette variable dans le modèle.

Tableau 28: Estimation du paramètre *Compétiteur Plus Proche*Urbain Rural* (avec *Compétiteur Plus Proche = 2* et *Urbain Rural=2* comme référence)

Analyse des valeurs estimées des paramètres GEE								
Paramètre			Valeur estimée	Erreur type	Intervalle de confiance à 95 %		Z	Pr > Z
COMP_PLUS*URBAIN_RUR	1	1	-3.3896	0.5491	-3.2400	-1.0876	-3.94	<.0001

*Compétiteur*Compétiteur Plus Proche*: Il fut observé précédemment que les variables individuelles *Compétiteur* et *Compétiteur Plus Proche* ne se sont pas avérées significatives. Après s'être attardé à leur effet combiné, on remarqua que ce dernier était quant à lui significatif. La figure suivante illustre l'effet de cette interaction sur les ventes hebdomadaires des clients.

Figure 19: Effet de la proximité du magasin sur le type de compétiteur présent dans le marché (la proximité du magasin est représentée par la variable *Compétiteur Plus Proche* et le type de compétiteur par la variable *Compétiteur*)



Pour la figure précédente, la modalité de référence est celle où le magasin de la bannière étudiée est le plus près du consommateur dans un contexte où le compétiteur A a déjà fait son apparition sur le marché. Dans ce contexte, la variable d'interaction

permet d'observer que lorsqu'un magasin de la compétition est situé plus près du client que celui de la bannière étudiée, l'effet sur les dépenses hebdomadaires des clients est plus dommageable lorsqu'il s'agit du compétiteur B plutôt que le compétiteur A (57,47\$ vs 62,70\$). Dans le cas où il s'agit de la bannière étudiée qui se situe le plus près du client, on constate qu'il n'y a pratiquement aucune différence entre le fait qu'elle soit en présence du compétiteur A ou du compétiteur B (62,30\$ vs 61,70\$). Voici un tableau illustrant la valeur des différents paramètres de cette variable d'interaction dans le modèle. La valeur nulle de la première interaction (2,1) montre que dans un contexte où l'on retrouve le compétiteur A, il n'y a pas de différence entre le fait que la bannière étudiée ou celle du compétiteur soit la plus près du consommateur. La valeur nulle du second paramètre (2,1) démontre qu'il n'y a pas de différence entre le fait avoir la présence du compétiteur A ou B lorsque la bannière étudiée est la plus près du consommateur.

Tableau 29: Estimation du paramètre *Compétiteur*Compétiteur Plus Proche* (avec *Compétiteur Plus Proche = 2* et *Compétiteur =2* comme référence)

Analyse des valeurs estimées des paramètres GEE								
Paramètre			Valeur estimée	Erreur type	Intervalle de confiance à 95 %		Z	Pr > Z
Compétiteur*Compétiteur_ plus_proche	1	1	-4.6326	0.5655	-5.7409	-3.5243	-8.19	<.0001
Compétiteur*Compétiteur_ plus_proche	1	2	0.0000	0.0000	0.0000	0.0000	.	.
Compétiteur*Compétiteur_ plus_proche	2	1	0.0000	0.0000	0.0000	0.0000	.	.

Ratio_Département : L'analyse de la variable *Ratio_Département* a révélé que cette dernière était significative. Cependant, il faut garder à l'esprit que cette variable reflète uniquement le comportement du client avant l'arrivée du compétiteur et ne nous permet donc pas de déduire quoique ce soit par rapport à l'impact de l'arrivée de la compétition. La modalité de référence est celle du ratio du prêt-à-manger (PAM) et les paramètres du tableau 30 permettent de quantifier l'augmentation en dollars

(tout autre facteur demeurant constant) de la dépense hebdomadaire d'un client, lorsque celui-ci augmente de 1% ses achats dans un département quelconque, au détriment de celui du PAM. On remarque que pour la majorité des départements (viandes, déli, dairy et fruits et légumes), une augmentation de 1% des dépenses dans ces départements au détriment du PAM résulte en une augmentation de la dépense hebdomadaire d'environ 0,30\$. C'est pour les départements du non-alimentaire et de l'épicerie que l'impact est plus important. En effet, une augmentation de 1% des dépenses dans ces départements au détriment du PAM se traduit par des augmentations respectives de 0,56\$ et de 0,40\$ de la dépense hebdomadaire des clients. Ce sont les dépenses faites dans le département de la boulange au détriment du PAM qui touchent le moins les dépenses hebdomadaires du client. Pour chaque augmentation de 1% des ce type de dépenses, c'est une augmentation de 0,17\$ que l'on observe sur la dépense hebdomadaire. Voici le tableau illustrant la valeur des différents paramètres de cette variable dans le modèle.

Tableau 30: Estimation du paramètre *Ratio_Département* (avec *Département = Ratio_PAM* comme référence)

Analyse des valeurs estimées des paramètres GEE							
Paramètre		Valeur estimée	Erreur type	Intervalle de confiance à 95 %		Z	Pr > Z
Ratio_Épicerie		0.3995	0.0235	0.3534	0.4456	16.98	<.0001
Ratio_Fruits_&_Légumes		0.3457	0.0270	0.2928	0.3987	12.79	<.0001
Ratio_Viandes		0.2979	0.0224	0.2540	0.3418	13.29	<.0001
Ratio_Déli		0.2783	0.0342	0.2112	0.3454	8.13	<.0001
Ratio_Boulanges		0.1769	0.0331	0.1120	0.2417	5.34	<.0001
Ratio_Non_Alimentaire		0.5603	0.0905	0.3829	0.7377	6.19	<.0001
Ratio_Dairy		0.3282	0.0275	0.2743	0.3822	11.92	<.0001
Ratio_Cig_Vin_Biere		0.2418	0.0276	0.1878	0.2959	8.77	<.0001

*Ratio_Département*Compétiteur*: L'analyse de l'interaction entre les ratios des dépenses par département et du type de compétiteur avait pour but de déterminer si

certaines clients dépensant davantage dans certains départements pouvaient être plus affectés par l'arrivée d'un nouveau type de compétiteur. Tel fut le cas, car les interactions *Ratio Département*Compétiteur* se sont révélées significatives pour les départements du déli, des fruits et légumes et du non-alimentaires. Les sections suivantes expliqueront tour à tour les effets sur les acheteurs de ces trois départements. À noter que les paramètres sont exprimés en fonction de la modalité de référence qui dans ce cas-ci, représente le département de prêt-à-manger chez le compétiteur A.

Déli : Le tableau 31 permet de constater que pour une augmentation de 1% des dépenses hebdomadaires faites dans le département du déli au détriment du département de prêt-à-manger (en gardant tout autre facteur constant), les clients en présence du compétiteur B (*Compétiteur =1*) dépensent en moyenne 0,09\$ de moins par semaine que lorsqu'ils sont en présence du compétiteur A.

Tableau 31: Estimation du paramètre *Ratio Déli*Compétiteur* (avec *Ratio Département =PAM* et *Compétiteur =2* comme référence)

Analyse des valeurs estimées des paramètres GEE							
Paramètre		Valeur estimée	Erreur type	Intervalle de confiance à 95 %		Z	Pr > Z
Ratio_Déli*Compétiteur	1	-0.0842	0.0321	-0.1472	-0.0212	-2.62	0.0088

Fruits & Légumes : Le tableau 32 permet de constater que pour une augmentation de 1% des dépenses hebdomadaires faites dans le département des fruits et légumes au détriment du département de prêt-à-manger (en gardant tout autre facteur constant), les clients en présence du compétiteur B (*Compétiteur =1*) dépensent en moyenne 0,11\$ de moins par semaine que lorsqu'ils sont en présence du compétiteur A.

Tableau 32: Estimation du paramètre *Ratio_Fruits&Légumes*Compétiteur* (avec *Ratio Département =PAM* et *Compétiteur =2* comme référence)

Analyse des valeurs estimées des paramètres GEE							
Paramètre		Valeur estimée	Erreur type	Intervalle de confiance à 95 %		Z	Pr > Z
Ratio_Fruits*Compétiteur	1	-0.1088	0.0257	-0.1592	-0.0584	-4.23	<.0001

Non-alimentaire : Le tableau 33 permet de constater que pour une augmentation de 1% des dépenses hebdomadaires faites dans le département du non-alimentaire au détriment du département de prêt-à-manger (en gardant tout autre facteur constant), les clients en présence du compétiteur B (*Compétiteur =1*) dépensent en moyenne 0,21\$ de moins par semaine que lorsqu'ils sont en présence du compétiteur A.

Tableau 33: Estimation du paramètre *Ratio Non Alimentaire*Compétiteur* (avec *Ratio Département =PAM* et *Compétiteur =2* comme référence)

Analyse des valeurs estimées des paramètres GEE							
Paramètre		Valeur estimée	Erreur type	Intervalle de confiance à 95 %		Z	Pr > Z
Ratio_Non_Alimentaire*Compétiteur	1	-0.2092	0.0830	-0.3719	-0.0466	-2.52	0.0117

3.2.2 – Effets des variables d'interaction triple

*Semaine*RFV Code*Compétiteur* : La seule variable d'interaction triple qui fut analysée dans le cadre de ce travail est la variable qui croisait simultanément la variable temporelle *Semaine* avec la variable de segmentation de clientèle *RFV_Code* et la variable *Compétiteur* permettant de déterminer le type de compétiteur présent dans l'environnement. Un tel effet permettrait de voir si à travers le temps, les segments de clientèles étaient impactés différemment en fonction du type de compétiteur qui était sur place. Cette interaction triple n'a finalement pas été intégrée au modèle final, car elle ne s'était pas révélée comme significative au départ.

À noter qu'un tableau récapitulatif regroupant l'ensemble des effets significatifs et non-significatifs couverts dans ce chapitre se trouve en annexe 1.

CHAPITRE 4 : DISCUSSION ET RECOMMANDATIONS

4.1 – Discussion

Dans le cadre de ce mémoire, un modèle de régression linéaire mixte pour données corrélées a été proposé afin d'expliquer une problématique de défection de clientèle dans un contexte où un détaillant alimentaire subissait l'arrivée d'un compétiteur dans son environnement. Ce modèle a été appliqué sur les transactions de 43 071 clients effectuées pendant une période de huit semaines avant l'arrivée d'un compétiteur ainsi que huit semaines après son arrivée. Cette modélisation aura permis d'identifier des facteurs permettant de repérer un client à risque de défection. Ces derniers donneront des orientations tangibles aux marketers désirant mieux comprendre les consommateurs à risque de quitter leur entreprise et fourniront des pistes pouvant orienter une stratégie de rétention dans un contexte futur similaire.

Tout d'abord, le modèle a établi que la variable de segmentation *RFV_Code* avait un effet explicatif sur la défection de clientèle. Comme pour les travaux de Buckinx et Van den Poel (2004), cette variable de type RFV s'est elle aussi avérée un bon prédicteur de la variable dépendante. Cependant, puisque les huit semaines de transactions effectuées avant l'arrivée du compétiteur ont servi à construire la variable *RFV_Code*, il était normal que cette dernière arrive à bien prédire la dépense hebdomadaire d'un client durant les seize semaines d'analyse. L'apprentissage sur cette variable de segmentation provient donc davantage de son interaction avec la variable temporelle *Semaine* qui elle aussi, avait un effet prédictif significatif. En effet, l'interaction de ces deux variables a permis de constater que les différents segments de clients n'étaient pas affectés de la même façon à partir du moment où un compétiteur arrivait sur le marché. Ce sont les clients les plus fidèles avant l'arrivée du compétiteur qui ont diminué le plus leur dépense hebdomadaire après l'arrivée d'un compétiteur. Les deux autres segments ont conservé un comportement plutôt stable dans le temps malgré l'arrivée du compétiteur.

Toujours selon le modèle proposé, la présence du compétiteur B dans le marché semblait avoir un impact légèrement plus négatif sur les clients les plus fidèles (RFV1) alors que le compétiteur A semblait toucher un peu plus à la baisse les clients moyennement et peu fidèles (RFV2 et RFV3). Tel que mentionné dans le premier chapitre, l'arrivée d'un compétiteur sur le marché ainsi que sa distance par rapport au sujet étudié n'avaient à notre connaissance pas été couverts dans la littérature de l'analyse de la défection. Pourtant, ces deux variables auront aussi permis d'identifier des comportements plus à risque de délaisser l'entreprise. Plus précisément, l'interaction de ces deux variables montre que lorsqu'un compétiteur était situé plus près du client que le magasin de la bannière étudiée, le client devenait plus susceptible de réduire ses achats chez ce dernier si ce compétiteur était de type B. Le modèle a aussi pu déterminer que la proximité d'un compétiteur par rapport au client avait davantage un impact négatif sur sa dépense lorsque ce dernier habitait en milieu urbain plutôt qu'en milieu rural.

En plus de la variable de type RFV et de l'information sur la présence et la distance du compétiteur dans l'environnement, le dernier facteur de risque ayant été détecté par le modèle est en lien avec le type de produits achetés par les consommateurs avant l'arrivée du compétiteur. En effet, l'interaction de la variable *Ratio_Département* (indiquant la proportion de produits achetés dans un département) avec la variable *Compétiteur* démontre que les clients achetant davantage de leurs produits dans les départements de fruits et légumes, du non-alimentaire et du déli (charcuteries et fromages fins) étaient plus susceptibles de réduire leurs achats après l'arrivée du compétiteur B. Le modèle proposé a aussi tenté de comprendre si ce phénomène pouvait varier en fonction du type de segment de client, mais l'interaction triple entre les variables précédentes et la variable *RFV_Code* s'est avérée non-significative.

Bien que le modèle proposé ait permis d'établir les facteurs représentant un risque de défection pour les clients de l'entreprise étudiée, ce dernier comporte certains inconvénients et impose quelques limites quant à l'interprétation des résultats. Tout d'abord, il est vrai que puisque ce modèle se veut de type explicatif plutôt que prédictif, il ne permet pas d'identifier directement les individus potentiellement à risque de quitter l'entreprise. Pour ce faire, un modèle de type régression logistique pour données corrélées pourrait être envisagé. Cependant, l'entreprise devra d'abord creuser les données transactionnelles de ses clients afin d'établir ce qu'elle considérera comme un état de défection.

Deuxièmement, il est important de garder à l'esprit que l'échantillon utilisé comporte certaines limites en ce qui a trait à l'étude de la défection. En effet, bien que la variable *RFV_Code* permette de départager certains niveaux de loyauté entre les membres du programme de fidélité, il n'en demeure pas moins que l'ensemble des clients étudiés étaient des membres actifs au sein d'un programme de fidélité. Bien qu'il soit probable que certains clients fidèles fassent le choix de ne pas adhérer au programme de fidélité, il est légitime de croire que les clients non-membres soient des individus moins fidèles. Le fait de ne pas avoir pu les inclure dans l'échantillon introduit un biais potentiel dans nos résultats et prive l'analyse d'individus potentiellement très pertinents au but fixé par ce travail. Une avenue intéressante permettant de contourner partiellement cette limite est celle de l'utilisation du numéro de carte de crédit des non-membres tel que proposé par Angel et Hadary (1998). Comme le mentionne un rapport d'Industrie Canada paru en 2011, l'utilisation de la carte de crédit est en augmentation au pays depuis les dernières années et cette tendance est bien connue par l'entreprise étudiée. En intégrant à la base de données de l'entreprise (Angel et Hadary 1998) le numéro de la carte de crédit utilisée, l'entreprise bénéficierait d'un moyen efficace de suivre dans le temps des clients ne faisant pas partie de leur programme de fidélité.

La troisième limite observée dans les résultats proposés par ce modèle réside dans la robustesse de certaines variables. Bien que l'ensemble des variables aient été attribuées à l'ensemble des 43 071 clients étudiés, elles n'ont pas toutes été construites à partir du même nombre d'observations. Par exemple, les variables *RFV_Code*, *Compétiteur_Plus_Proche*, *Ratio_Département*, *Semaine* ainsi que toutes les variables d'interaction impliquant ces variables ont été construites à partir de caractéristiques uniques provenant de 43 071 clients différents. À l'opposé, les variables *Compétiteur* et *Urbain_Rural* ont été construites à partir de seulement huit observations, c'est-à-dire les caractéristiques provenant des huit magasins étudiés. Puisque les variables sur la compétition et sur l'environnement ont été construites avec beaucoup moins d'observations que les variables caractérisant les clients, il est important d'interpréter leurs résultats avec davantage de précautions. Cette faiblesse pourra être compensée dans l'avenir en réutilisant ce modèle sur un plus grand nombre de compétiteurs et d'environnements lorsque de nouveaux cas d'ouverture de compétiteurs se présenteront dans le futur.

Quatrièmement, il faut comprendre que de la façon dont le modèle a été construit, il considérait que les huit magasins de la bannière étaient de qualité équivalente et proposaient tous la même offre et la même expérience. Dans les faits, la connaissance du terrain amène à reconnaître que l'offre et l'expérience reçues par les consommateurs sont fortement tributaires de la façon dont le magasin est opéré par son équipe de gestion. Il est donc légitime de croire qu'un magasin dégageant davantage d'insatisfaction auprès de sa clientèle puisse être plus vulnérable face à la défection de ses clients. Dans une modélisation future de la même problématique d'affaires, il pourrait être intéressant d'intégrer des données provenant des sondages biannuels de satisfaction de la clientèle face aux magasins étudiés. L'ajout de ce genre de variables permettrait de capturer une réalité du terrain qui ne transparait pas dans les données transactionnelles des consommateurs.

Pour ce qui est des avantages proposés par ce modèle, certaines caractéristiques sont dignes de mention. Tout d'abord, la facilité avec laquelle il est possible d'interpréter

l'effet des différentes variables sur la variable d'intérêt (la dépense hebdomadaire) représente un avantage certain par rapport à d'autres modèles plus complexes couverts dans la revue de littérature. En effet, les paramètres estimés trouvés significatifs donnent directement l'effet en dollars sur la dépense hebdomadaire du client.

Un second avantage de la méthodologie utilisée réside en sa rapidité computationnelle sur les jeux de données de grande taille. En effet, des méthodologies abordées dans la revue de littérature comme la régression logistique ou les réseaux de neurones peuvent prendre plusieurs heures lorsque roulées sur d'imposant jeux de données. La procédure de régression linéaire pour données corrélées roulée via le logiciel SAS a obtenu des résultats en une dizaine de minutes. Dans le contexte actuel de l'entreprise, cela s'avère un avantage non négligeable puisque le rafraîchissement quotidien des serveurs de données interrompt souvent les requêtes roulant pendant un long moment hors des périodes traditionnelles de bureau.

Les résultats proposés par ce modèle auront permis à l'entreprise de constater pour la première fois les facteurs de risque de défection auprès de ses clients dans un contexte d'arrivée d'un nouveau compétiteur. Afin que le modèle proposé soit le plus utile possible pour l'entreprise étudiée, il serait judicieux d'appliquer à nouveau ce modèle sur l'ensemble des magasins ayant subi l'ouverture de compétiteurs en date d'aujourd'hui. En fonction des résultats obtenus, une stratégie de rétention pourrait ainsi être mise en place auprès des membres possédant les « traits à risque » de défection avant l'arrivée d'un prochain compétiteur. Au fur et à mesure que d'autres compétiteurs arriveront sur le marché, le modèle pourra être rafraîchi et ainsi donner des informations de plus en plus précises.

Une autre idée intéressante pour l'entreprise étudiée serait d'intégrer au modèle davantage de variables portant sur le profil de ses clients. Par le biais de l'adhésion de nouveaux membres au programme de fidélité, le simple fait de poser des questions de base sur le sexe, l'âge, l'occupation et le revenu ajouterait de l'information précieuse

sur les clients de l'entreprise. De plus, le fait de noter la date à partir de laquelle un client commence à être membre du programme de fidélité servirait à quantifier la durée de la relation entre ce dernier et l'entreprise, une variable s'étant avérée significative dans l'analyse de la défection par Buckinx et Van den Poel (2004).

Cependant, cette collecte de nouvelles informations via le processus d'adhésion au programme de fidélité prendrait un certain temps avant d'atteindre un nombre appréciable d'individus. Un moyen plus rapide d'ajouter des renseignements sociodémographiques à la base de données actuelle serait d'utiliser de l'information déjà disponible. En s'appuyant sur la prémisse que les habitants d'un même quartier possèdent des styles de vie et des habitudes de consommation similaires, il serait intéressant d'ajouter de l'information sociodémographique provenant des données publiques de recensement de Statistique Canada. Puisque la base de données actuelle de l'entreprise étudiée possède déjà l'information relative aux codes postaux des membres du programme de fidélité, il serait relativement aisé de lier les données de recensement que Statistique Canada compile par code postal à celle des membres de l'entreprise et d'ainsi augmenter la connaissance sur les clients de l'entreprise.

4.2 – Recommandations managériales

La section précédente avait pour objectif de mettre en évidence les forces et les faiblesses du modèle proposé. Cette partie-ci aura plutôt comme but de donner des solutions tangibles aux *marketers* souhaitant utiliser directement les résultats mis en évidence par le modèle. Ils pourront ainsi orienter leurs tactiques dans un contexte où un nouveau compétiteur viendrait à s'implanter à proximité d'un magasin de la bannière étudiée. Nous passerons donc en revue les variables significatives et nous formulerons différentes recommandations par rapport à ces dernières.

1-Semaine : L'analyse de cette variable a permis de constater que les deux premières semaines (*Semaine=1*) suivant les deux premières semaines d'ouverture de la compétition représentaient la période où la dépense hebdomadaire moyenne des

clients chutait le plus. Il serait donc possible de combattre cet effet en envoyant pour cette période un encart circulaire spécial aux clients vivant dans la zone commerciale du magasin touché. En proposant un encart bonifiant l'offre de la circulaire hebdomadaire avec des produits intéressants à prix concurrentiels, l'effet dommageable de la période *Semaine=1* pourrait se voir réduit. De façon alternative, une offre spéciale pourrait aussi être envoyée par courriel auprès des clients du magasin (faisant partie du programme de fidélité) une semaine avant la période *Semaine=1*. Cette offre pourrait proposer un produit gratuit ou une remise en argent à la condition que le client dépense un minimum le montant d'une épicerie complète au magasin visé.

2-RFV Code*Semaine: Puisque l'effet individuel de la variable *RFV_Code* ne nous permet pas d'analyser le comportement des segments de clientèle après l'arrivée de concurrents, nous allons plutôt émettre des recommandations en fonction de sa variable d'interaction avec la variable semaine. Tel que l'illustre la figure 17, les différents segments ne sont pas impactés de la même façon par l'arrivée d'un concurrent. En effet les deux segments de clients moins fidèles (RFV2 et RFV3) semblent peu impactés par l'arrivée de la compétition contrairement au segment de clients le plus fidèles (RFV1). Ce dernier a réduit en moyenne près de 8,00\$ sa dépense hebdomadaire à la période *Semaine=1* pour ensuite garder le même niveau de dépenses durant les deux périodes suivantes (*Semaine=2* et *Semaine=3*). Pour cette raison, nous proposons d'abord de concentrer les efforts de rétention sur ces derniers. Rappelons-nous que les clients RFV1 visitaient la bannière étudiée régulièrement et y dépensaient un montant élevé à chaque visite avant l'arrivée du concurrent. Nous proposons donc de leur donner un incitatif visant à s'assurer qu'ils visitent la bannière étudiée durant les huit semaines suivant l'arrivée du concurrent (durant les périodes : *Semaine=1*, *Semaine=2*, *Semaine=3*) tout en ne réduisant pas leurs dépenses. Une promotion électronique pourrait ainsi être envoyée directement aux clients RFV1 afin de leur proposer de dépenser un montant minimal substantiel par semaine pendant les huit semaines suivant l'arrivée du concurrent. Lorsque les clients auraient atteint ce niveau de dépenses pendant huit semaines consécutives, une

remise significative en argent ou en points de récompenses du programme de fidélité pourrait alors leur être octroyée.

Si le budget le permet, il serait aussi possible d'adopter une tactique semblable auprès des deux autres segments de clients (RFV2 et RFV3). Puisque ces derniers ont des niveaux de dépenses inférieurs au segment RFV1, nous proposons une initiative similaire mais qui exigerait une dépense hebdomadaire de 60\$ pour les RFV2 et de 40\$ pour les RFV3. En ayant trois différentes offres pour les trois différents segments, il faudrait cependant s'assurer que les segments supérieurs ne puissent pas profiter des offres dédiées aux segments inférieurs. Un tel oubli permettrait aux segments supérieurs de diminuer leurs dépenses tout en profitant de l'offre spéciale.

3-RFV Code*Compétiteur: L'interaction entre la variable de segmentation de clientèle et celle définissant le type de compétiteur ouvrant ses portes met en évidence certaines différences entre les segments de clientèle. En effet, le segment le plus fidèle (RFV1) semblait un peu plus touché par la présence du compétiteur B alors que les deux autres segments (RFV2 et RFV3) étaient plus touchés par la présence du compétiteur A. En sachant cela, la promotion proposée au point précédent pourrait être bonifiée en offrant aux clients RFV1 la possibilité d'obtenir aux mêmes prix, les produits offerts dans la circulaire du compétiteur B durant la durée de la promotion. La même chose pourrait aussi être offerte aux clients RFV2 et RFV3 par rapport aux produits offerts dans la circulaire du compétiteur A. De cette façon, les fuites des clients plus à risque de quitter l'entreprise pourraient potentiellement être réduites.

4-Ratio Département*Compétiteur: L'interaction entre les ratios de départements et le type de compétiteur ouvrant ses portes a permis de réaliser que les clients achetant certaines catégories de produits avant l'arrivée des compétiteurs sont davantage touchés lors de l'ouverture de ces derniers. Plus précisément, les clients fervents de produits provenant des départements de déli, de fruits et légumes et du non-alimentaire ont été plus touchés par l'arrivée du compétiteur A. Dans le futur, on pourrait ainsi offrir des promotions électroniques ciblées donnant des réductions

monétaires ou des points de récompenses additionnels sur les produits les plus achetés dans les départements de déli, de fruits et légumes et de non-alimentaires. Cette promotion ciblée pourrait contenir deux ou trois offres par département pouvant être utilisées uniquement durant les deux mois suivant l'ouverture du compétiteur A.

5-Compétiteur*Compétiteur Plus Proche: L'interaction entre le type de compétiteur présent et sa plus grande proximité face au client a permis de constater que dans un contexte où il y a présence du compétiteur B, le fait qu'un compétiteur soit situé plus près du client a un impact plus négatif que lorsqu'il s'agit d'un magasin de la bannière étudiée ou du compétiteur A. Dans le cas où le compétiteur B ouvre ses portes, l'entreprise pourrait envoyer ces offres électroniques personnalisées (recommandations 2&3) uniquement aux clients demeurant plus près du compétiteur B ou bien tenter de leur envoyer une communication additionnelle. La communication additionnelle pourrait prendre la forme d'un message envoyé par la poste rappelant la présence et l'implication du propriétaire dans sa communauté. Cet aspect est un avantage dont ne peuvent se vanter les nouveaux magasins du compétiteur B qui émanent davantage d'une structure corporative. En effet, ces nouveaux magasins sont de plus en plus la propriété de la compagnie mère et non plus de marchands affiliés.

6- Compétiteur Plus Proche*Urbain Rural: Puisque l'effet individuel de la variable *Urbain_Rural* englobe autant l'effet de l'avant que de l'après compétition, nous allons plutôt émettre une recommandation par rapport à sa variable d'interaction *Compétiteur Plus Proche*Urbain_Rural* qui permet d'isoler l'effet de l'arrivée de la compétition. L'effet de cette variable d'interaction indique que la proximité d'un compétiteur face à un client est plus dommageable pour l'entreprise lorsqu'on se trouve en milieu urbain. Ceci n'est pas entièrement surprenant, car on peut comprendre qu'en milieu urbain, la proximité d'une épicerie par rapport à la résidence peut être un critère de choix d'épicerie plus important que lorsqu'on se trouve en milieu rural. Afin d'inciter les clients vivant plus près d'un magasin de la compétition en milieu urbain, nous recommandons d'envoyer des offres

électroniques ciblées sur des produits de différenciation exclusifs à la bannière étudiée. En agissant de la sorte, ces offres inciteraient les clients à se procurer des items qu'ils ne peuvent trouver chez la compétition. Cela aurait comme effet potentiel de réduire l'importance de la proximité comme critère de sélection d'épicerie auprès de ceux qui demeurent plus près d'un magasin de la compétition en milieu rural.

ANNEXE

Annexe 1 : Estimation des paramètres du modèle

Analyse des valeurs estimées des paramètres GEE							
Estimations de l'erreur type empirique							
Paramètre		Valeur estimée	Erreur type	Intervalle de confiance à 95 %		Z	Pr > Z
Intercept		62.2968	1.8873	58.6000	65.9982	33.01	<.0001
RFV_Code	3	-81.3214	0.6194	-82.5336	-80.1055	-131.28	<.0001
RFV_Code	2	-54.8306	0.6592	-56.1158	-53.5318	-83.17	<.0001
RFV_Code	1	0.0000	0.0000	0.0000	0.0000	.	.
Semaine	3	-5.8318	1.4139	-9.0722	-3.5299	-4.46	<.0001
Semaine	2	-6.9777	1.4221	-10.2364	-4.6619	-5.24	<.0001
Semaine	1	-8.0074	1.4285	-11.2770	-5.6774	-5.93	<.0001
Semaine	0	0.0000	0.0000	0.0000	0.0000	.	.
Compétiteur	0	0.0000	0.0000	0.0000	0.0000	.	.
Compétiteur	1	-0.5938	1.7553	-4.0342	2.8466	-0.34	0.7351
Compétiteur	2	0.0000	0.0000	0.0000	0.0000	.	.
Noel	0	-2.8451	0.2670	-3.3684	-2.3219	-10.66	<.0001
Noel	1	0.0000	0.0000	0.0000	0.0000	.	.
Compétiteur_plus_proche	0	0.0000	0.0000	0.0000	0.0000	.	.
Compétiteur_plus_proche	1	0.3995	0.4917	-0.5643	1.3633	0.81	0.4166
Compétiteur_plus_proche	2	0.0000	0.0000	0.0000	0.0000	.	.
URBAIN_RURAL	1	-7.4050	0.9768	-9.3194	-5.4905	-7.58	<.0001
URBAIN_RURAL	2	0.0000	0.0000	0.0000	0.0000	.	.
Ratio_Épicerie		0.3995	0.0235	0.3534	0.4456	16.98	<.0001
Ratio_Fruits_&_Légumes		0.3457	0.0270	0.2928	0.3987	12.79	<.0001
Ratio_Viandes		0.2979	0.0224	0.2540	0.3418	13.29	<.0001
Ratio_Déli		0.2783	0.0342	0.2112	0.3454	8.13	<.0001
Ratio_Boulanges		0.1769	0.0331	0.1120	0.2417	5.34	<.0001
Ratio_Non_Alimentaire		0.5603	0.0905	0.3829	0.7377	6.19	<.0001
Ratio_Dairy		0.3282	0.0275	0.2743	0.3822	11.92	<.0001
Ratio_Cig_Vin_Biere		0.2418	0.0276	0.1878	0.2959	8.77	<.0001
Ratio_PAM		0.0000	0.0000	0.0000	0.0000	.	.

Analyse des valeurs estimées des paramètres GEE								
Estimations de l'erreur type empirique								
Paramètre			Valeur estimée	Erreur type	Intervalle de confiance à 95 %		Z	Pr > Z
Compétiteur_plus_proche* Urbain_Rural	0	1	-5.7947	0.3712	-5.9504	-4.4954	-14.07	<.0001
Compétiteur_plus_proche* Urbain_Rural	0	2	0.0000	0.0000	0.0000	0.0000	.	.
Compétiteur_plus_proche* Urbain_Rural	1	1	-3.3896	0.5491	-3.2400	-1.0876	-3.94	<.0001
Compétiteur_plus_proche* Urbain_Rural	1	2	0.0000	0.0000	0.0000	0.0000	.	.
Compétiteur_plus_proche* Urbain_Rural	2	1	0.0000	0.0000	0.0000	0.0000	.	.
Compétiteur_plus_proche* Urbain_Rural	2	2	0.0000	0.0000	0.0000	0.0000	.	.
Semaine*Compétiteur	3	1	-0.1307	0.4164	-0.9468	0.6853	-0.31	0.7535
Semaine*Compétiteur	3	2	0.0000	0.0000	0.0000	0.0000	.	.
Semaine*Compétiteur	2	1	-0.0272	0.4547	-0.9183	0.8640	-0.06	0.9524
Semaine*Compétiteur	2	2	0.0000	0.0000	0.0000	0.0000	.	.
Semaine*Compétiteur	1	1	0.0000	0.0000	0.0000	0.0000	.	.
Semaine*Compétiteur	1	2	0.0000	0.0000	0.0000	0.0000	.	.
Semaine*Compétiteur	0	0	0.0000	0.0000	0.0000	0.0000	.	.
Semaine*RFV_Code	3	3	11.2438	0.5558	9.9961	12.1748	19.94	<.0001
Semaine*RFV_Code	3	2	5.6415	0.6202	4.3793	6.8105	9.02	<.0001
Semaine*RFV_Code	3	1	0.0000	0.0000	0.0000	0.0000	.	.
Semaine*RFV_Code	2	3	12.4767	0.5857	11.1704	13.4662	21.03	<.0001
Semaine*RFV_Code	2	2	7.6505	0.6548	6.3205	8.8872	11.61	<.0001
Semaine*RFV_Code	2	1	0.0000	0.0000	0.0000	0.0000	.	.
Semaine*RFV_Code	1	3	12.3734	0.5957	11.0473	13.3824	20.51	<.0001
Semaine*RFV_Code	1	2	6.9860	0.6707	5.6248	8.2541	10.35	<.0001
Semaine*RFV_Code	1	1	0.0000	0.0000	0.0000	0.0000	.	.
Semaine*RFV_Code	0	3	0.0000	0.0000	0.0000	0.0000	.	.
Semaine*RFV_Code	0	2	0.0000	0.0000	0.0000	0.0000	.	.

Analyse des valeurs estimées des paramètres GEE								
Estimations de l'erreur type empirique								
Paramètre			Valeur estimée	Erreur type	Intervalle de confiance à 95 %		Z	Pr > Z
Semaine*RFV_Code	0	1	0.0000	0.0000	0.0000	0.0000	.	.
Compétiteur*Compétiteur_ plus_proche	0	0	0.0000	0.0000	0.0000	0.0000	.	.
Compétiteur*Compétiteur_ plus_proche	1	1	-4.6326	0.5655	-5.7409	-3.5243	-8.19	<.0001
Compétiteur*Compétiteur_ plus_proche	1	2	0.0000	0.0000	0.0000	0.0000	.	.
Compétiteur*Compétiteur_ plus_proche	2	1	0.0000	0.0000	0.0000	0.0000	.	.
Compétiteur*Compétiteur_ plus_proche	2	2	0.0000	0.0000	0.0000	0.0000	.	.
RFV_Code*Compétiteur	3	0	0.0000	0.0000	0.0000	0.0000	.	.
RFV_Code*Compétiteur	3	1	1.7795	0.6887	0.5237	3.2232	2.72	0.0065
RFV_Code*Compétiteur	3	2	0.0000	0.0000	0.0000	0.0000	.	.
RFV_Code*Compétiteur	2	0	0.0000	0.0000	0.0000	0.0000	.	.
RFV_Code*Compétiteur	2	1	1.8230	0.7726	0.3254	3.3540	2.38	0.0173
RFV_Code*Compétiteur	2	2	0.0000	0.0000	0.0000	0.0000	.	.
RFV_Code*Compétiteur	1	0	0.0000	0.0000	0.0000	0.0000	.	.
RFV_Code*Compétiteur	1	1	0.0000	0.0000	0.0000	0.0000	.	.
RFV_Code*Compétiteur	1	2	0.0000	0.0000	0.0000	0.0000	.	.
Ratio_Épicerie*Compétiteur	0		0.1962	0.0182	0.1605	0.2320	10.76	<.0001
Ratio_Épicerie*Compétiteur	1		0.0320	0.0220	-0.0111	0.0750	1.45	0.1459
Ratio_Épicerie*Compétiteur	2		0.0000	0.0000	0.0000	0.0000	.	.
Ratio_Fruits&_Légumes* Compétiteur	0		-0.0554	0.0210	-0.0966	-0.0141	-2.63	0.0085
Ratio_Fruits&_Légumes* Compétiteur	1		-0.1088	0.0257	-0.1592	-0.0584	-4.23	<.0001
Ratio_Fruits&_Légumes* Compétiteur	2		0.0000	0.0000	0.0000	0.0000	.	.
Ratio_Viandes*Compétiteur	0		0.1342	0.0167	0.1016	0.1668	8.06	<.0001

Analyse des valeurs estimées des paramètres GEE							
Estimations de l'erreur type empirique							
Paramètre		Valeur estimée	Erreur type	Intervalle de confiance à 95 %		Z	Pr > Z
Ratio_Viandes*Compétiteur	1	-0.0173	0.0205	-0.0575	0.0229	-0.84	0.3986
Ratio_Viandes*Compétiteur	2	0.0000	0.0000	0.0000	0.0000	.	.
Ratio_Déli*Compétiteur	0	0.0423	0.0270	-0.0107	0.0952	1.56	0.1176
Ratio_Déli*Compétiteur	1	-0.0842	0.0321	-0.1472	-0.0212	-2.62	0.0088
Ratio_Déli*Compétiteur	2	0.0000	0.0000	0.0000	0.0000	.	.
Ratio_Boulanges*Compétiteur	0	0.0614	0.0255	0.0115	0.1113	2.41	0.0158
Ratio_Boulanges*Compétiteur	1	0.0241	0.0309	-0.0366	0.0848	0.78	0.4360
Ratio_Boulanges*Compétiteur	2	0.0000	0.0000	0.0000	0.0000	.	.
Ratio_Non-Alimentaire* Compétiteur	0	-0.1813	0.0714	-0.3213	-0.0413	-2.54	0.0112
Ratio_Non-Alimentaire* Compétiteur	1	-0.2092	0.0830	-0.3719	-0.0466	-2.52	0.0117
Ratio_Non-Alimentaire* Compétiteur	2	0.0000	0.0000	0.0000	0.0000	.	.
Ratio_Dairy*Compétiteur	0	0.1081	0.0211	0.0667	0.1494	5.12	<.0001
Ratio_Dairy*Compétiteur	1	0.0097	0.0257	-0.0406	0.0601	0.38	0.7047
Ratio_Dairy*Compétiteur	2	0.0000	0.0000	0.0000	0.0000	.	.
Ratio_Cig_Vin_Biere* Compétiteur	0	0.1463	0.0214	0.1043	0.1882	6.83	<.0001
Ratio_Cig_Vin_Biere* Compétiteur	1	0.0301	0.0262	-0.0214	0.0815	1.15	0.2517
Ratio_Cig_Vin_Biere* Compétiteur	2	0.0000	0.0000	0.0000	0.0000	.	.
Ratio_PAM*Compétiteur	0	0.0000	0.0000	0.0000	0.0000	.	.
Ratio_PAM*Compétiteur	1	0.0000	0.0000	0.0000	0.0000	.	.
Ratio_PAM*Compétiteur	2	0.0000	0.0000	0.0000	0.0000	.	.

BIBLIOGRAPHIE

ANGEL, G., HADARY, J.(1998). « Using Card Transaction Data», *American Demographics* 20, 8, 1998, 38-41.

BAESENS, B., GLADY, N. et CROUX, C. (2008). « Modeling Churn using customer lifetime value », *European Journal of Operational Research* 197, 2009, 402-411.

BREIMAN, L. (1996). « Bagging predictors », *Journal Machine Learning* Volume 24 Édition #2, Août 1996, 123-140.

BUCKINX, W. et VAN DEN POEL, D. (2004). « Customer base analysis: partial defection of behaviourally loyal clients in a non-contractual FMCG retail setting », *European Journal of Operational Research* 164, 2005, 252-268.

DUKE TERADATA FOR CUSTOMER RELATIONSHIP MANAGEMENT, « US Wireless Industry » [En ligne], Durnham, http://www.teradata.com/news_t_2.html, (Ref. du 7 Novembre 2002)

FADER, P.S., HARDIE, B.G.S. et LEE, Ka Lok, (2005). « RFM and CLV: Using iso-value curves for customer base analysis », *Journal of Marketing Research* 42 (4), 415–430.

FREUND, Y. et SCHAPIRE, R. (1999), « A short introduction to boosting », *Journal of Japanese Society for Artificial Intelligence* 14(5), Septembre 1999, 771-780.

GROUPE AIMIA (2012). « À Propos d' Aimia », Portait de la compagnie Aeroplan, [En ligne], Montréal, <http://www.aimia.com/French/A-propos-dAimia/Portefeuille/Aeroplan/default.aspx>, (Ref. du 23 Août 2012)

INDUSTRIE CANADA (2011), « Rapport sur les tendances en consommation », Chapitre 7 sur l'endettement des consommateurs, [En ligne], Montréal, <http://www.ic.gc.ca/eic/site/oca-bc.nsf/fra/ca02111.html> (Ref. du 2 Décembre 2012)

LAROCQUE, D. (2008). « Analyse de données longitudinales et de survie », notes de cours, Montréal, HEC Montréal, 242p.

LEMMENS, A. et CROUX, C. (2006). « Bagging and boosting classification trees to predict churn », *Journal of Marketing Research* Vol. XLIII, Mai 2006, 276-286.

NATH, S. et BEHARA, R. (2002). « Customer churn analysis in the wireless industry: A data mining approach », BOCA RATON, Florida Atlantic University, 20p.

SHAW, R. (1988), *Database Marketing*, Londres, Royaume-Uni, Gower Pub Co, 224p.

