



# Une analyse des comportements de recommandation de produits à partir des avis postés sur Amazon

Nessrine Omrani, Tasnime Omrani, Alain Rallet

DANS **INNOVATIONS** 2024/1 N° 73 , PAGES 119 À 152

ÉDITIONS **DE BOECK SUPÉRIEUR**

ISSN 1267-4982

ISBN 9782807380417

DOI 10.3917/inno.pr2.0157

Date de mise en ligne : 22/01/2024

Article disponible en ligne à l'adresse

<https://shs.cairn.info/revue-innovations-2024-1-page-119?lang=fr>



Découvrir le sommaire de ce numéro, suivre la revue par email, s'abonner...  
Scannez ce QR Code pour accéder à la page de ce numéro sur Cairn.info.



**Distribution électronique Cairn.info pour De Boeck Supérieur.**

Vous avez l'autorisation de reproduire cet article dans les limites des conditions d'utilisation de Cairn.info ou, le cas échéant, des conditions générales de la licence souscrite par votre établissement. Détails et conditions sur [cairn.info/copyright](https://shs.cairn.info/copyright).

Sauf dispositions légales contraires, les usages numériques à des fins pédagogiques des présentes ressources sont soumises à l'autorisation de l'Éditeur ou, le cas échéant, de l'organisme de gestion collective habilité à cet effet. Il en est ainsi notamment en France avec le CFC qui est l'organisme agréé en la matière.

---

# Une analyse des comportements de recommandation de produits à partir des avis postés sur Amazon

**Nessrine OMRANI**

*Paris School of Business*  
n.omrani@psbedu.paris

**Tasnime OMRANI**

*Efrei Paris Panthéon-Sorbonne Université*  
tasnim.omrani@gmail.com

**Alain RALLET**

*RITM, Université Paris-Saclay*  
alain.rallet@universite-paris-saclay.fr

## RÉSUMÉ

La mise en place par Amazon d'un système de notation des biens achetés par les consommateurs est une des innovations majeures apportées par la plateforme à l'industrie du e-commerce. Reposant sur le travail des consommateurs, elle a donné naissance à une littérature sur les raisons poussant les consommateurs à faire ce travail, ces raisons étant testées à partir d'enquêtes déclaratives. L'article reprend cette question des comportements contributifs mais en l'examinant à partir des données observables sur le site d'Amazon, *i. e.* les informations laissées par les acheteurs sur leurs avis postés. Le nombre très élevé de données a contraint à des choix (États-Unis, biens culturels hors livres), et la nature des données à un travail d'interprétation visant à tester ou retrouver par des modèles économétriques des déterminants mentionnés dans la littérature tels que

la satisfaction des consommateurs, leur engagement, leur fidélité ainsi que des variables contextuelles.

**MOTS CLÉS :** Avis en ligne, Plateformes, Comportement Consommateur, Co-crédation, E-commerce, Amazon

**CODES JEL :** O35, D12

## ABSTRACT

### **An Analysis of Product Recommendation Behavior Based on Amazon Reviews**

Amazon's implementation of a rating system for goods is one of the major innovations of the e-commerce industry. This has led to new literature on the reasons behind consumers making use of the system, investigated using declarative surveys. This article examines this contributory behavior from data on Amazon's website (*i. e.*, information by consumers in their reviews). The extensive data have influenced choices in the United States (including cultural goods but excluding books), and the nature of the data prompted this interpretive work aimed at identifying and evaluating, by econometric models, the determinants mentioned in the literature, such as consumer satisfaction, engagement, and loyalty, as well as contextual variables.

**KEYWORDS:** Online Reviews, Platforms, Consumer Behavior, Co-creation, E-commerce, Amazon

**JEL CODES:** O35, D12

L'innovation majeure induite par les plateformes digitales est de transformer radicalement l'organisation des secteurs économiques. Au modèle verticalement intégré des industries hérité du XX<sup>e</sup> siècle tendent à succéder des organisations de marché à deux versants, les plateformes en étant les coordinatrices et, en occupant cette place stratégique, les principales bénéficiaires. Le secteur de l'informatique a été lui-même un des premiers à connaître cette transformation radicale dans les années 1990 (Baldwin, 2017). D'autres industries ont suivi de près comme les industries culturelles dans les années 2000 (Thuillas, Wiart, 2023) puis le mouvement de plateformes s'est étendu à l'ensemble des secteurs car, si tous n'ont pas encore succombé à ce type d'organisation industrielle, tous en sont frappés (Béjean *et al.*, 2022).

Le caractère spectaculaire, disruptif, de ce nouveau type d'organisation explique que l'innovation apportée par cette transformation ait été centrée sur l'introduction d'externalités indirectes de réseau, *i. e.* sur

les effets croisés entre les deux versants du marché, comme dynamique motrice de la nouvelle organisation (Bacache-Beauvallet, Bourreau, 2022). C'est pourquoi nombre d'économistes des plateformes, notamment les pionniers, en ont fait un critère distinctif de la définition d'une plateforme (Armstrong, 2006 ; Rochet, Tirole, 2003). Or l'innovation apportée par les plateformes ne se réduit pas à l'introduction d'externalités indirectes de réseau mais aussi, comme nous le soulignons dans cet article, à la création d'effets directs de réseau. De sorte que la définition plus large donnée par Belleflamme et Peitz (2019) des plateformes comme gestionnaires d'effets réseau, directs et indirects, nous convient davantage, l'innovation touchant les deux types d'effets.

La plateformesation du commerce de détail (Moati, 2021) illustre ces deux types d'innovation : la transformation des distributeurs en *marketplace* (externalités indirectes de réseau entre acheteurs et vendeurs), le développement de systèmes de recommandation des produits (effets de réseau directs entre acheteurs). C'est ce dernier aspect qui est ici développé.

Initiée par Amazon à la fin des années 1990, la mise en place d'un système de *Reviews and Ratings* (Notes et Avis, N+A dans cet article) est devenue un standard de la distribution. Elle intervient très tôt dans l'histoire d'Amazon, avant sa mutation en *marketplace*. Bezos l'évoque dès le lancement de l'entreprise en 1995 et la met en place en 1999 (Stone, 2013). Dès cette date, elle transforme Amazon en une plateforme au sens de Belleflamme et Peitz. L'introduction d'avis sur les biens postés sur le site par les acheteurs crée un effet de réseau direct entre les acheteurs réels et potentiels : l'utilité retirée de la consultation du site par l'acheteur potentiel d'un bien dépend du nombre d'avis laissés par les acheteurs existants. C'est au départ une innovation controversée : Bezos l'impose aux éditeurs de livre qui redoutent que la possibilité de laisser un avis à la suite d'un achat n'ait un effet négatif sur les ventes.

La notation de biens et de services par les clients n'est pas nouvelle en soi. Coquaz et Halissat (2020) soulignent que le *word-of-mouth* existait auparavant de manière informelle ou/et limitée mais le numérique permet de systématiser à une vaste échelle « *l'idéal démocratique de la note* » (i. e. « *les consommateurs parlent aux consommateurs* ») qui vient se substituer efficacement sur le Net à l'impact subjectif du vendeur dans le magasin. Cette innovation va tenir une place centrale dans le modèle d'affaires d'Amazon, et par la suite, de tous les e-commerçants, puisque Bezos déclare le fonder non pas sur la vente de biens mais sur l'exploitation des

effets directs de réseau (« *We don't make money when we sell things. We make money when we help customers make purchase decisions* »).

À côté d'autres types de « *user generated contents* », les systèmes de recommandation fondés sur les N+A ont donné corps à l'idée d'une *sharing economy* transformant les consommateurs en co-constructeurs de services dans le cadre d'un « *crowd-based capitalism* » (Sundararajan, 2016). Ces systèmes visent en l'occurrence à réduire une caractéristique structurelle des marchés de biens et de services : l'asymétrie d'information entre acheteurs et vendeurs. Elle ajoute aux dispositifs traditionnels de réduction de cette asymétrie (marques, labels, signes de qualité, normes techniques, messages publicitaires, avis d'experts, guides d'achat, etc.) un mécanisme d'évaluation fondé sur le jugement subjectif des pairs. Elle présente 2 avantages. Le partage d'une expérience vécue de consommation fait tout d'abord contrepoids au marketing de l'offre et aux dires d'experts. La diversité des expériences sensibles présentées permet ensuite aux consommateurs potentiels de s'identifier aux caractéristiques de ceux qui relatent leur propre expérience de consommation (Oliveira *et al.*, 2020 ; Scott, Orlikowski, 2012).

Pour que le dispositif soit efficace, la note (*rating*) doit être associée à une explicitation qualitative du jugement délivrant des indications sensibles sur le bien et la personnalité des commentateurs. La note sans l'avis associé ne suffit pas à traiter l'asymétrie d'information tandis que l'avis sans note laisse l'acheteur potentiel dépourvu d'une règle de décision simple (par exemple ne s'intéresser qu'aux biens dont la note est au moins de 4 sur 5). C'est donc l'association (N+A) de la note et d'un avis à laquelle nous nous intéressons dans cet article.

La littérature sur les N+A recouvre deux champs distincts. Les travaux les plus nombreux analysent leur impact sur les ventes (Vautrot *et al.*, 2018 ; Lemonnier, Spencer, 2016 ; Mahadin, Akroush, 2020). L'article s'inscrit dans le second champ qui s'attache aux motivations des acheteurs à laisser des N+A. Ces motivations conditionnent l'innovation apportée par les N+A. Les analyser est donc essentiel. Ces travaux s'inscrivent dans la littérature sur les ressorts de l'économie contributive et des *User Generated Contents*.

L'originalité du papier est d'analyser les déterminants des N+A sans les référer à des motivations individuelles dégagées d'enquêtes déclaratives. Notre analyse est fondée sur les informations lisibles à partir des avis postés sur le site d'Amazon. Le web offre de nouvelles ressources aux chercheurs, non seulement avec les enquêtes en ligne (Jean, 2015), mais aussi avec les

données captées directement des sites Web (cf. par exemple Balech (2022) pour Twitter). Nous nous inscrivons dans cette dernière démarche. Les informations captées nous ont permis de construire une base de données et de tester un certain nombre de facteurs explicatifs issus de la littérature marketing : le niveau de satisfaction du consommateur, son niveau d'engagement, sa fidélité à l'égard de la plateforme ainsi que des facteurs contextuels liés à sa localisation. Ne pouvant articuler ces informations à des données individuelles sociodémographiques couvertes par le secret commercial, cette méthode éclaire un autre versant de l'explication des comportements contributifs. C'est à la fois son intérêt et sa limite.

Compte tenu du nombre très élevé de données (17 millions d'avis de consommateurs) sur la période considérée (2001-2014), du temps pour les extraire et du coût de leur stockage, nous nous sommes limités à une zone géographique (les États-Unis) et aux produits culturels (hors livres), d'autant que notre approche étant compréhensive et non prescriptive, nous souhaitions rester dans les limites d'un traitement économétrique maîtrisable.

Dans la section 2 nous procédons à une revue de littérature et formulons nos hypothèses. La section 3 présente notre méthodologie et les données utilisées. La section 4 expose les principaux résultats empiriques obtenus. Enfin, la section 5 discute les résultats et conclut.

## Revue de littérature et hypothèses

### **La littérature sur les avis postés sur les plateformes Internet possède deux volets, l'un sur leur impact sur les ventes, l'autre sur leurs déterminants**

La littérature sur les comportements d'achats en ligne est abondante en marketing et en systèmes d'information (Chevalier, Mayzlin, 2006 ; Chen, Xie, 2008 ; Zhu, Zhang, 2010 ; Floyd *et al.*, 2014 ; Seiler *et al.*, 2017) et parfois en sociologie (Beauvisage *et al.*, 2013). Ces travaux analysent en détail les avis (leur longueur, la crédibilité, la réputation ou le degré d'expertise de l'auteur, les biais introduits, etc.) pour évaluer leur utilité pour les consommateurs et leur rôle dans les décisions d'achat (Wu *et al.*, 2015 ; Askalidis *et al.*, 2017 ; Bhole, Hanna, 2017) ou les prédictions de vente (Chong *et al.*, 2015). Mais ce n'est pas cette littérature qui nous intéresse ici.

Celle sur les déterminants des N+A postés est moins prolifique mais néanmoins nombreuse. Elle s'inscrit dans les travaux sur les ressorts de l'économie contributive sur Internet qui s'attachent à comprendre les motivations poussant les individus à fournir des contributions sans être rémunérés directement et à produire des services collectifs (ou biens communs) par effet d'agrégation. Lancés à l'occasion du développement du logiciel libre puis des forums sur Internet et des phénomènes de type Wikipédia, ces travaux se sont partagés entre les économistes tentant de maintenir le principe d'une rémunération monétaire indirecte (Lerner, Tirole, 2002), les sociologues soulignant le rôle du capital social et des effets de groupe et les psychologues utilisant la distinction entre motivations intrinsèques et extrinsèques (Deci, Ryan, 1985). La littérature marketing sur le sujet (Hennig-Thurau, 2004 ; Cheung, Lee, 2012 ; Tong *et al.*, 2013) reprend ces lignes de force tout en prolongeant les travaux antérieurs sur le rôle du *Word-of Mouth* (Dichter, 1966). Les auteurs s'interrogent notamment sur la part des motivations intéressées et désintéressées : le souci d'aider son prochain (les autres consommateurs), la volonté de soutenir une firme (cas fréquent pour les restaurants) ou à l'inverse celle de nuire, le *warm glow* (valorisation intrinsèque de soi ou extrinsèque via la reconnaissance d'un groupe), l'attente de bénéfices sociaux, le résultat de pratiques marketing (prix et qualité des produits, Chen *et al.*, 2011), ou de rémunérations directes ou indirectes par les vendeurs et enfin la production d'effets de « *crowding out* » par les interactions entre ces motivations, les motivations extrinsèques pouvant chasser les motivations intrinsèques (Sun *et al.*, 2017 ; Wu, 2019 ; Chang *et al.*, 2020).

Les études empiriques s'efforcent de démêler ces motivations variées dont le poids relatif sert à dresser des typologies de rédacteurs d'avis (Mathwick, Mosteller, 2017). Elles s'appuient parfois sur des études qualitatives (Wu, 2019), mais le plus souvent sur des études quantitatives utilisant des enquêtes par questionnaire sur les motivations des *reviewers*.

## Démarche et hypothèses

Un nombre plus restreint d'études utilisent des données extraites sur Internet en en inférant des comportements. Nous nous sommes inscrits dans cette voie en collectant des données sur le site d'Amazon et les avis qui y ont été postés sur une longue période (2000-2014). L'intérêt est de disposer de données massives qui reflètent des comportements, l'inconvénient est de ne pouvoir les relier aux caractéristiques sociodémographiques

et comportementales des *reviewers*, ces données relevant de la base commerciale d'Amazon. Les données extraites permettent de construire des variables qui sont utilisées comme proxys de variables comportementales au moyen d'un travail d'interprétation fondé sur les enquêtes traditionnelles, qualitatives et quantitatives.

Dans la base construite à partir des données extraites d'Amazon, nous avons sélectionné des données qui permettent d'expliquer les comportements de contribution. La propension à poster des avis qui est notre variable expliquée est estimée par le nombre d'avis laissés sur la plateforme pendant toute la période par un individu. Plus ce nombre est élevé, plus la propension à poster des recommandations est forte. Les données permettent par ailleurs d'identifier un certain nombre de déterminants qui constituent nos variables explicatives.

Nous disposons tout d'abord de la note qui accompagne la partie textuelle de l'avis. C'est un indicateur de la satisfaction du consommateur : un niveau de satisfaction élevé l'incite à poster un avis (Ranaweera, Prabhu, 2003). Gvili et Levy (2016) montrent qu'écrire un commentaire en ligne est une pratique des consommateurs les plus satisfaits. La satisfaction est indexée sur la performance du produit du vendeur ou/et du service rendu par la *marketplace*, en particulier quand les qualités du produit ou/et du service expérimentées par le consommateur à la réception ou à l'usage dépassent ses attentes (Anaza, Zhao, 2013), ce qui ne fait que confirmer des observations antérieures (Tse, Wilton, 1988). L'écart perçu génère un bouche-à-oreille positif et pousse les consommateurs à recourir au même site web dans le futur (Chiou, 2003). On peut donc s'attendre à ce qu'une note élevée traduisant une forte satisfaction ait un effet positif sur le nombre d'avis laissés. C'est d'autant plus vrai que la note cristallise un effet d'auto-renforcement : à la satisfaction propre du consommateur s'ajoute le signal qu'il envoie aux autres consommateurs. Fondé sur le sentiment de satisfaction éprouvé par l'acheteur et qu'il désire partager, ce signal ne peut, pour être efficace, s'embarasser de trop grandes nuances (Schoenmueller *et al.*, 2020). Même s'il n'est pas totalement convaincu par le produit ou le service, le consommateur sait que son avis ne produira l'effet désiré (convaincre les autres d'acheter le produit) que si la note dépasse un certain seuil (variable selon les conventions propres à chaque site). Le signal ne peut fonctionner que dans les marges de la note la plus élevée (entre 4 et 5 dans le cas d'Amazon). À l'autre extrémité du spectre, les notes négatives sont aussi fortement tassées vers le bas (vers le 1, le 0 n'étant pas possible). On s'explique ainsi la forte polarisation des notes

postées sur les sites de e-commerce (les notes comprises entre 4 et 5 représentent 73,68 % dans notre base et les notes inférieures à 2 représentent 2,11 %).

Si le sentiment de satisfaction influence favorablement la propension à laisser un avis (Ranaweera, Prabhu, 2003), qu'en est-il de l'effet produit par son contraire, le sentiment d'insatisfaction ? On sait que les avis négatifs diminuent plus fortement les ventes que les avis positifs ne les augmentent, en raison du comportement dominant d'aversion aux pertes (Weisstein *et al.*, 2017). Ils sont à ce titre redoutés par les vendeurs mais ils sont peu fréquents dans l'ensemble des avis laissés par les consommateurs (dans notre échantillon, la note 1 représente 1,47 % des avis). Hennig-Thurau *et al.* (2004) notent que les consommateurs postant des avis négatifs cherchent moins à partager une expérience négative, ce qui est en soi peu valorisant, qu'à diminuer leur insatisfaction propre. La fonction signal de l'avis négatif a l'effet inverse de celle des avis positifs sur la propension à poster des avis. Elle tend à les réduire tandis que l'autre les amplifie en cumulant satisfaction intérieure et partage d'une expérience positive. L'hypothèse suivante est dès lors justifiée :

H1 : La satisfaction du consommateur mesurée par la note postée a une influence positive sur le nombre d'avis laissés en ligne.

Un second déterminant du nombre d'avis postés est lisible dans nos données : l'engagement du consommateur. La littérature marketing fait un usage extensif et multidimensionnel de cette notion (Van Doorn *et al.*, 2013 ; Roderick *et al.*, 2013). Il y a engagement dès lors que le comportement du consommateur à l'égard d'une firme ou d'une marque excède le cadre monétaire ponctuel d'une transaction. Thakur (2018) montre que l'engagement du consommateur a un impact positif sur l'intention d'écrire un commentaire en ligne.

En croisant cette notion avec la littérature citée plus haut sur les motivations, nous entendrons par engagement non un rapport à la firme mais un rapport à la communauté des acheteurs potentiels visitant le site. C'est la manifestation d'un souci des autres. L'attention portée à la rédaction du message matérialise ce souci. Plus le message développe l'expérience vécue, par l'acheteur, plus il apporte d'informations pertinentes aux consommateurs potentiels et plus il leur apporte de valeur. L'attention se mesure ainsi par l'effort consacré à la rédaction du message, effort qui se reflète dans la longueur du message utilisée comme proxy du souci des autres.

H2 : L'engagement du consommateur a une influence positive sur le nombre d'avis laissés en ligne.

Les données nous permettent de considérer un troisième déterminant : la fidélité à l'égard du site. Initialement, la fidélité à une firme ou à une marque est liée au comportement d'achat (Cunningham, 1996) ou de réachat (Tucker, 1964 ; Filser, 1994). Dans le marketing relationnel, la fidélité ou loyauté (McKnight *et al.*, 2002 ; Kim, Benbasat, 2003) s'entend plus largement comme le maintien d'une relation à long terme ayant deux aspects : la préférence pour un prestataire (Gremler, Brown, 1996) et l'envie de poursuivre la relation (Hikkerova *et al.*, 2015). Elle incarne la confiance dans un site (Morgan, Hunt, 1994 ; Berry, 1995) et influence la fréquence des visites (Jevons, Gabbott, 2000). Dans notre cas, la fidélité est mesurée par le temps écoulé entre le premier et le dernier avis posté sur le site d'Amazon dans la période considérée. La durée de cette période reflète le niveau de confiance des *reviewers* vis-à-vis de la plateforme. Plus le consommateur a confiance dans le site, plus il lui sera fidèle et laissera des commentaires sur une plus longue période de temps (Kim *et al.*, 2009 ; Gremler *et al.*, 2001). D'où notre troisième hypothèse :

H3 : La fidélité des consommateurs à l'égard du site influence positivement le nombre d'avis qu'ils laissent en ligne.

Nous ne disposons pas des variables sociodémographiques qui nous auraient permis de contrôler les effets de nos variables explicatives. Nous avons en revanche introduit des facteurs contextuels pouvant influencer les comportements en ligne, ce que ne fait pas la littérature sur la publication des avis en ligne qui s'est surtout focalisée sur des déterminants individuels. Par facteurs contextuels, nous entendons des variables d'environnement socio-économiques caractérisant les zones de résidence des consommateurs ayant posté un avis sur Amazon. Nous connaissons en effet ces zones à une échelle hélas large (les États américains).

Nous avons alors extrait de bases de données américaines des variables d'environnement par État tels que l'âge moyen, le niveau de revenu, le niveau d'études, le mode d'utilisation d'Internet, le temps qui y est consacré, l'usage du mobile, le temps de trajet domicile-travail qui sont réputés influencer le comportement en ligne des consommateurs résidants dans l'État.

## Comment peuvent-elles l'influencer ?

Dans la période considérée, l'Internet a connu à la fois un accroissement considérable de l'offre marchande de biens culturels en ligne et une démocratisation sociale, générationnelle et géographique de l'Internet. Nous ne disposons pas de données sur l'impact des variables d'âge, de revenu et de localisation sur les avis postés dans le domaine des biens culturels. Mais nous pouvons supposer qu'elles agissent dans le même sens que celui de leur impact sur l'achat de ces biens en ligne. La démocratisation de l'Internet a réduit les écarts observés en 2010 mais l'âge, le revenu (cadres et professions intellectuelles), le niveau de diplôme influencent positivement la probabilité d'effectuer un achat de produit culturel en ligne, mais pas le genre (Thoumelin, 2020 pour la période 2000-2018 en France). L'impact positif de la variable d'âge a été amplifié par la pratique très répandue du piratage des biens culturels par les plus jeunes dans les années 2000 et 2010. La disponibilité d'Internet et l'expérience de navigation (temps consacré, usage du mobile) sont aussi connues comme des facteurs influençant les achats en ligne (Le Guel *et al.*, 2005) et donc a fortiori le nombre d'avis postés.

Nous ajoutons à ces variables le temps de trajet domicile-travail car celui-ci est utilisé, parmi d'autres activités (Aguilera, Rallet, 2016), pour le shopping en ligne (acheter mais aussi rechercher et produire de l'information commerciale) pendant les déplacements dans le cas des transports en commun (Adoue, 2017 ; Ben-Elia *et al.*, 2018). Il est intéressant de voir si cette variable a un impact sur le nombre d'avis postés qui est une des composantes du shopping. Combinés à la généralisation du mobile au sein de la population, les déplacements domicile-travail qui représentent une partie importante du temps quotidien pour certaines populations sont a priori favorables au développement du shopping en ligne. Il faut toutefois nuancer cette attente car une partie importante des déplacements s'opère en voiture qui est un cadre moins propice au shopping en ligne que les transports en commun ! Aux États-Unis, le temps moyen du *commuting* a légèrement augmenté lors de la période étudiée de 2006 à 2019 (de 25 à 28 minutes quotidiennement) mais avec 10 % de salariés consacrant plus de 60 minutes par jour à cette activité<sup>1</sup>. Plus le trajet est long, plus les *commuters* utilisent les transports en commun (25 % de ceux dont le temps de *commuting* dépasse 60mn par jour contre 5 % en moyenne pour l'ensemble des *commuters*). On doit s'attendre à une progression du

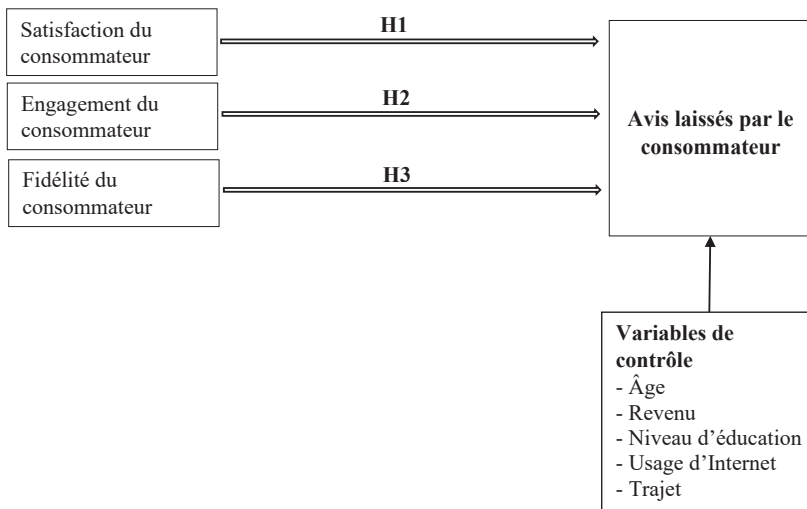
---

1. U.S. CENSUS BUREAU, 2006 to 2021 American Community Survey.

shopping (donc des avis postés) effectué lors des déplacements domicile-travail, mais cette progression est freinée par la part de l'autosolisme dans ces déplacements, notamment aux États-Unis où 75 % des travailleurs se rendent seuls au travail en voiture<sup>2</sup>.

Résumons l'effet attendu des variables de contexte sur le nombre d'avis postés : positif pour l'âge, le revenu, le niveau de formation, la disponibilité d'Internet, l'expérience d'utilisation. L'effet attendu des trajets-domicile-travail est plus modéré. Répétons que, ne disposant pas des données à l'échelle individuelle mais agrégée par État américain, nous les prenons comme des variables contextuelles susceptibles d'influencer le comportement des résidents et non comme des variables individualisées déterminant directement ce comportement.

**Figure 1 - Modèle conceptuel**



## Méthodologie et données

### Données

Les données sont collectées à partir du site Amazon.com et visibles sur ce site pour la période 2000-2014. Ces données concernent 17 166 988 avis

2. U.S. CENSUS BUREAU, 2006 to 2021 American Community Survey.

laissés par les utilisateurs du site. Notre objectif n'étant pas d'élaborer un modèle prédictif des comportements justifiant une accumulation maximale de données mais de les comprendre, nous avons prélevé dans cette base un échantillon de données limité aux biens culturels (hors livres, car les livres contiennent aussi un très grand nombre de données : 8 898 041). Les données sur les biens culturels hors livres portent sur 3 101 872 avis. Cinq catégories de biens culturels sont donc analysées : instrument de musique, musique numérique, jeux vidéo, CD et vinyles et enfin films et TV. Ces 5 types de biens peuvent être tous considérés comme des biens d'expérience. Ils sont donc homogènes au regard de ce critère, nous avons introduit la catégorie du produit comme variable de contrôle. Nous nous intéressons par ailleurs aux facteurs contextuels géographiques par état, or les données relatives à la localisation géographique des consommateurs ne sont pas disponibles dans la base Amazon. Afin de procéder à la collecte de ces informations, nous avons utilisé la technique du « *Web Scraping* » (Julian, Natalia, 2015).

Désireux d'introduire le cadre contextuel des rédacteurs d'avis, *i. e.* des données de contexte liées à leur localisation (États américains), nous avons gardé uniquement les consommateurs dont nous pouvions certifier la localisation postée sur le profil Amazon. La localisation est parfois non fournie ou bien erronée, Amazon ne contrôlant pas la validité des informations introduites par le consommateur. Pour cela, nous avons développé une application informatique permettant de détecter et de supprimer les informations représentant une localisation invalide ou inexistante dans le monde réel. Nous avons retenu les consommateurs dont la localisation se trouve aux États-Unis et ayant au moins un avis. Nous aurions souhaité utiliser la localisation par ville, ce qui aurait permis d'introduire une variable taille urbaine dont le pouvoir d'interprétation est plus important que celui d'une grande région ou, comme aux États-Unis, des états. Mais l'indication de localisation étant renseignée par les consommateurs eux-mêmes est très variable. De ce fait, l'indication selon l'État s'avère plus fiable et surtout plus complète. Cela représente 92 888 consommateurs répartis dans 51 états.

Nous avons ensuite collecté les données relatives aux États des États-Unis et nous les avons introduits dans la base de données afin de traiter les variables contextuels (par état)<sup>3</sup>.

---

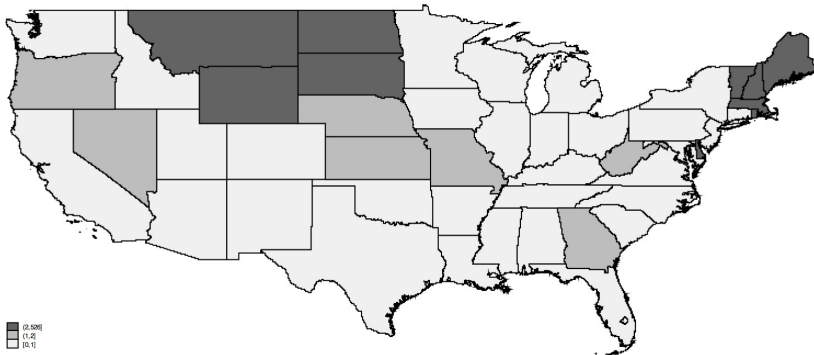
3. Source des données : [statista.com](https://www.statista.com) et [priceonomics.com](https://www.priceonomics.com)

## Statistiques descriptives

Notre intérêt porte sur la répartition du nombre d'avis laissés par les contributeurs.

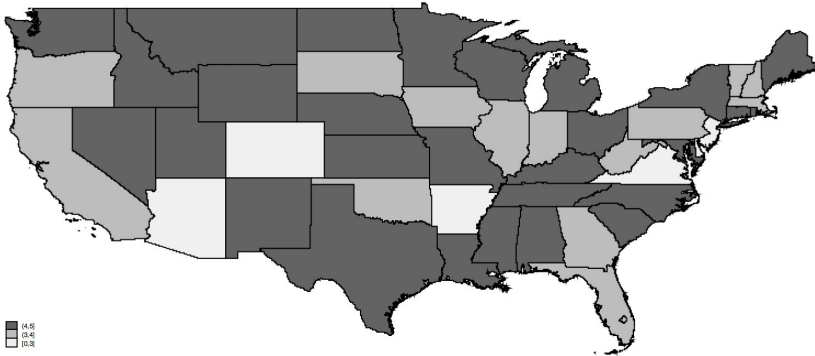
Le graphique 1 représente la répartition du nombre d'avis pondéré par les populations des États (nombre d'avis par million d'habitants). Le graphique montre que le nombre moyen d'avis laissés par les contributeurs est plus élevé (supérieur à 2) dans 10 États, à savoir : Montana, Wyoming, Dakota du nord, Dakota du sud, Vermont, New Hampshire, Maine, Rhode Island, Massachusetts et Delaware. Il se situe de manière intermédiaire entre 1 et 2 avis (compris) dans l'ensemble formé par l'Oregon, le Nevada, le Nebraska, le Kansas, le Missouri, la West Virginie et la Géorgie. Enfin, il est faible (inférieur ou égal à 1) dans le reste des États, soit 34 États. Le nombre d'avis est important principalement à l'ouest et le nord-est. Cela peut s'expliquer par les habitudes et les cultures différentes d'un État à un autre ainsi que par l'application des différentes lois telle que la loi américaine du Consumer Review Fairness Act.

**Graphique 1 – Répartition du nombre d'avis par État**



Le graphique 2 représente la répartition de la moyenne des notes attribuée par les contributeurs par état. Le graphique montre que dans la majorité des États les contributeurs ont donné une moyenne supérieure à 4 sur 5. Ce résultat montre que les consommateurs ont un niveau de satisfaction élevé pour ce type de bien. Cela représente 57 713 contributeurs (62,13 %). La moyenne des notes est inférieure ou égale à 3 dans 5 États (Arizona, Colorado, Arkansas, Virginia et New Jersey, soit 9,80 % du nombre total d'états). Dans le reste des États, soit 28,07 %, la moyenne est comprise 3 et 4 compris.

**Graphique 2 - Répartition de la moyenne des notes par État**



Les statistiques descriptives par État sont données en annexe en tableau A.

Le tableau 1 décrit les variables (endogène et exogènes) de notre modèle explicatif du nombre d'avis laissés par le contributeur.

Les matrices de corrélations aux niveaux des consommateurs et des États sont données en annexe (voir tableau B, C et D). La matrice de corrélation montre un lien positif entre le nombre d'avis d'une part et l'engagement, la fidélité et la satisfaction du consommateur d'autre part. Le graphique A en annexe représente la répartition de la longueur des commentaires laissés par le consommateur.

## Résultats

Afin de valider nos hypothèses, un modèle a été testé. Le nombre d'avis laissés par chaque consommateur est la variable expliquée. La mesure du nombre d'avis étant une variable continue, nous utilisons une régression par la méthode des moindres carrés ordinaires pour estimer le lien entre le nombre d'avis et les variables explicatives que nous avons déjà présentées.

Le tableau 2 représente les résultats des modèles relatifs aux déterminants du nombre d'avis laissés par le contributeur (le Tableau E décrivant le modèle détaillé avec les *dummies* « État » est en annexe).

Le modèle (1) donne les résultats relatifs au comportement du consommateur pour laisser un avis en ligne en prenant en compte les caractéristiques individuelles en termes de notation pour exprimer sa satisfaction vis-à-vis du site, du délai entre le premier avis posté et le dernier avis posté pour exprimer sa fidélité au site, la longueur du message posté en guide

Tableau 1 – Statistiques descriptives

Variable	Description	Moyenne	Écart type	Min,	Max,
Variable endogène					
Avis	Le nombre d'avis laissés par le consommateur	5 988	17,117	1	1 003
Variables exogènes					
Satisfaction	Moyenne des notes attribuées par le consommateur	4 214	0,865	1	5
Fidélité	Délai entre le premier et le dernier avis laissé par un consommateur (en jours)	546 466	825,781	0	5 787
Engagement	Longueur des messages laissés par le consommateur (en nombre de caractères)	747 196	789,082	0	32 766
Localisation	La localisation du consommateur par État			1	51
Catégorie de produits	Instrument de musique	0,006	0,077	0	1
	Musique numérique	0,030	0,170	0	1
	Jeux vidéo	0,111	0,315	0	1
	CD et Vinyles	0,441	0,496	0	1
	Films et TV	0,582	0,493	0	1
Variables de contrôle					
Âge	L'âge moyen par état	38,041	2,076	30,7	44,5
Revenu	Le revenu moyen par État par an en dollar	60 021,23	8 458,207	20 078	78 945
Niveau d'éducation	Le nombre de diplômés du supérieur divisé par le nombre d'habitants par état	0,015	0,002	0,008	0,031
Usage d'Internet	Le taux d'utilisation d'Internet par État (en %)	82,128	3,165	70,7	87,4
Usage d'Internet mobile	Le taux d'utilisation d'Internet sur mobile (en %)	60,957	5,677	40,1	72,2
Trajet	Le temps de trajet entre le domicile et le travail (en minutes)	25,575	3,324	15,7	32

Tableau 2 - Résultats du nombre d'avis laissés par les consommateurs

	(1)	(2)	(3)	(4)	(5)	(6)
Nombre d'avis						
Satisfaction	1,169*** (0,100)	1,169*** (0,1000)	1,164*** (0,100)	1,163*** (0,100)	1,161*** (0,100)	1,165*** (0,099)
Fidélité	0,001*** (0,000)	0,001*** (0,000)	0,001*** (0,000)	0,001*** (0,000)	0,001*** (0,000)	0,000*** (0,000)
Engagement	0,002*** (0,000)	0,002*** (0,000)	0,002*** (0,000)	0,002*** (0,000)	0,002*** (0,000)	0,002*** (0,000)
<i>Catégorie de produits</i>						
CD et Vinyles	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
Musique num,	34,270*** (1,290)	34,267*** (1,289)	34,262*** (1,289)	34,263*** (1,289)	34,259*** (1,289)	34,261*** (1,287)
Films et TV	4,735*** (0,139)	4,739*** (0,140)	4,743*** (0,140)	4,743*** (0,140)	4,743*** (0,140)	4,744*** (0,140)
Jeux vidéo	4,852*** (0,285)	4,857*** (0,285)	4,863*** (0,285)	4,863*** (0,285)	4,862*** (0,285)	4,867*** (0,285)
Instrument de musique	4,520*** (0,291)	4,520*** (0,291)	4,528*** (0,291)	4,528*** (0,291)	4,528*** (0,291)	4,530*** (0,291)
<b>Variable de contrôle</b>						
Âge		0,062** (0,026)	0,050* (0,028)	0,051* (0,029)	0,051* (0,029)	0,830 (0,974)
Revenu		0,000	0,000**	0,000**	0,000**	0,000

	(1)	(2)	(3)	(4)	(5)	(6)
		(0,000)	(0,000)	(0,000)	(0,000)	(0,000)
Niveau d'études		4,922 (15,892)	9,137 (16,502)	8,789 (16,573)	13,455 (16,605)	-78,873 (224,024)
Usage d'Internet			-0,074** (0,032)	-0,075** (0,033)	-0,080** (0,033)	-0,431 (0,439)
Usage d'Internet mobile			-0,008	-0,008	-0,134**	-1,467*
Trajet			(0,011)	(0,011)	(0,068)	(0,764)
Trajet*mobile				-0,003 (0,018)		
Trajet*voiture					0,005* (0,003)	0,061* (0,034)
Etats	<i>Non</i>	<i>Non</i>	<i>Non</i>	<i>Non</i>	<i>Non</i>	<i>Oui</i>
_cons	-2,160*** (0,236)	-4,766*** (1,131)	0,794 (2,247)	0,874 (2,272)	8,384* (4,696)	83,590 (59,690)
R <sup>2</sup>	0,15	0,15	0,15	0,15	0,15	0,15
N	92,888	92,888	92,888	92,888	92,888	92,888

\*  $p < 0.1$ ; \*\*  $p < 0.05$ ; \*\*\*  $p < 0.01$

d'engagement vis-à-vis du site et en prenant en compte la localisation et la catégorie du produit acheté.

Le modèle (2) intègre, en plus des caractéristiques individuelles, des variables contextuelles par État relatives à l'âge moyen, le revenu moyen et le niveau d'études.

Le modèle (3) ajoute les variables contextuelles relatives aux usages d'Internet (usage d'Internet et usage d'Internet mobile).

Le modèle (4) s'intéresse en plus à la variable contextuelle « trajet » qui représente le temps de trajet entre le domicile et le travail. Enfin, le modèle (5) détaille la variable trajet en la remplaçant par la variable « trajet\*mobile » qui est le croisement entre le trajet entre domicile et travail et l'usage d'Internet sur mobile et la variable « trajet\*voiture » qui mesure le trajet entre domicile et travail et le nombre de voitures moyen par consommateur. Le modèle (6) intègre les États en *dummies*.

De façon attendue, le modèle (1) montre qu'une note élevée a un lien positif et significatif à 1 % avec le nombre d'avis, ce qui nous permet de valider notre première hypothèse (H1). Une note élevée reflète un niveau de satisfaction élevé vis-à-vis du site et du produit et donne donc lieu à un plus grand nombre d'avis laissés par le contributeur. Un niveau de satisfaction élevé incite le contributeur à poster un avis envoyant un signal aux autres consommateurs et générant un bouche-à-oreille positif. Ce résultat pourrait être faussé par une manipulation de l'information par les vendeurs de produits, mais elle est peu vraisemblable sur une plateforme qui réunit autant d'avis qu'Amazon et par la chasse systématique que fait cette plateforme aux *fake reviews* (1/3 des avis postés seraient supprimés *ex post* par Amazon selon He *et al.*, 2020).

Les résultats montrent que l'engagement mesuré par le soin apporté à la rédaction d'avis détaillés (longueur des messages) pousse à la rédaction d'avis. En effet, le lien entre la longueur du message et le nombre d'avis postés est positif et significatif (à 5 % dans le premier modèle et à 1 % dans les modèles de 2 à 6).

La deuxième hypothèse (H2) est donc validée. Ce lien peut s'expliquer par un niveau d'engagement plus important du consommateur à l'égard de la qualité du service de recommandation offert par la plateforme. Nous avons soutenu dans la section précédente que la longueur d'un message pouvait refléter la complexité du jugement inférée par la nature du produit évalué et le degré d'engagement du consommateur à l'égard de la qualité du service de recommandation offert à la communauté des acheteurs de

la plateforme. Or tous les produits retenus dans le modèle sont des biens d'expérience, qui sont des biens complexes. L'effet complexité est donc neutralisé, quelle que soit la catégorie du produit, la longueur du message est corrélée positivement au nombre d'avis. Un message plus long reflète donc un niveau d'engagement du consommateur plus élevé à l'égard de la qualité du service de recommandation offert par la plateforme. Le rapport du contributeur à la communauté des acheteurs potentiels le pousse à leur apporter de la valeur en postant plus d'avis longs.

En ce qui concerne le délai, il y a un lien positif et significatif à 1 % entre délai et nombre d'avis. Cela veut dire que plus le délai entre la date du premier avis laissé par le contributeur et la date du dernier avis est long plus il y a d'avis sur le site. Cela peut s'expliquer de deux manières. Il peut s'agir d'acheteurs récents (par rapport à 2014, date finale de la base) qui n'ont donc pas eu le temps d'espacer leurs avis. Ou alors le délai est lié à l'opportunité d'un achat ou à une séquence rapprochée d'achats tandis qu'un délai long manifeste une permanence d'intérêt et de fidélité à l'égard de la plateforme, ce qui conduit le consommateur à laisser plus d'avis répartis dans le temps. Dans ce cas, celui où la distanciation temporelle des avis peut être interprétée comme une marque de fidélité à la plateforme, notre troisième hypothèse serait validée (H3).

La catégorie du bien a un lien significatif avec le nombre d'avis, ce qui est cohérent avec leur forte homogénéité (ce sont des produits culturels). Cela peut nous conduire à poursuivre cette étude par la prise en compte de produits hétérogènes (*search goods versus experience goods*) et de voir si on observe des effets différents sur la propension à laisser des avis. Les résultats montrent des liens significatifs avec un  $R^2$  de 15 %. Les données ne nous ont pas permis de prendre en compte des données individuelles qui permettraient d'apporter plus de compréhension aux modèles testés. Intégrer d'autres variables avec d'autres types de produits améliorer la saisie des effets de localisation, traiter la dimension subjective des messages ou encore prendre en compte les *fake reviews* permettrait d'améliorer la valeur du R carré.

Les résultats relatifs aux caractéristiques sociodémographiques révèlent que l'âge moyen par État un lien positif et significatif à 5 % (modèle 2) et à 10 % (modèle 3 à 5) avec le nombre d'avis postés. Le revenu moyen par État a un lien significatif à 5 % (modèle 3 à 5) avec le nombre d'avis postés. L'âge et le revenu influencent le comportement en ligne des consommateurs. C'est dans les États ayant l'âge moyen et le revenu moyen les plus élevés que les contributeurs postent le plus d'avis. Le taux d'utilisation

d'Internet par État ainsi que le taux d'utilisation d'Internet sur mobile ont un lien négatif et significatif avec le nombre d'avis postés. Plus les contributeurs utilisent Internet moins ils vont poster des avis en ligne, le temps d'usage d'Internet peut être corrélé à une diversité d'usage ce qui peut expliquer le temps réduit consacré au partage d'expérience à travers le partage d'avis. Lorsque l'on s'intéresse au temps de trajet entre le domicile et le travail, le lien avec le nombre d'avis postés n'est pas significatif. Par contre ce lien devient négatif et significatif lorsque l'on croise cette variable avec le nombre moyen de voitures utilisées par ménage. Il devient enfin positif et significatif lorsqu'on croise cette variable avec l'usage du mobile (l'utilisation du mobile pendant le temps de trajet). Il y a bien un effet positif attendu des déplacements domicile-travail via l'usage du mobile pendant ces déplacements sur le nombre d'avis postés mais cet effet est neutralisé à l'échelle sociale par la forte utilisation de la voiture pour le *commuting* aux États-Unis. Nous trouvons là un résultat conforme à la discussion introduite sur cet effet lors de la présentation des hypothèses sur les variables contextuelles.

Ces résultats confirment l'existence de lien entre les facteurs contextuels et le nombre d'avis postés en ligne par les consommateurs. Les facteurs contextuels représentés par des facteurs d'environnement numérique et sociétal influencent le comportement contributif en ligne.

## Discussion et conclusion

L'objectif de l'article était de s'interroger sur un usage possible de données extraites d'Internet (les avis postés sur le site d'Amazon suite à l'achat de biens culturels sur une longue période) pour comprendre des comportements, en l'occurrence ceux des consommateurs postant des avis sur des sites d'e-commerce, à côté d'autres méthodes qualitatives par entretiens ou quantitatives par questionnaire. Les données extraites présentent des avantages importants sur les autres méthodes, ceux de l'échelle (volume de données) et de la temporalité (longue période), mais aussi une limite, celle de ne pas disposer de certaines données individualisées pouvant éclairer les comportements observés. C'est particulièrement le cas dans le domaine du e-commerce en raison du secret commercial qui ne permet pas de lier un comportement observé (d'achat, d'avis) à la connaissance des caractéristiques des individus concernés. On observera que la difficulté est moindre dans le domaine des réseaux sociaux car l'individu y affiche nombre de caractéristiques comportementales. L'article s'est

attaché à traiter cette difficulté : comment expliquer des comportements observés sur une vaste échelle à l'aide de données extraites massivement ?

Nous avons fait pour cela l'hypothèse que les données extraites comportaient aussi des éléments d'explication des comportements observés à grande échelle, bien qu'en nombre réduit. Mais il n'est pas possible de les lire directement dans ces données comme dans une enquête par questionnaire. Ils sont présents dans l'existence de variables observées pouvant jouer le rôle de *proxy* de facteurs d'explication des comportements. Pour cela, il faut interpréter les variables issues des données extraites d'Internet comme des proxys de déterminants des comportements observés, agissant positivement ou négativement sur le nombre d'avis postés. Ce travail d'interprétation doit nécessairement s'appuyer sur les enquêtes traditionnelles, qualitatives ou quantitatives, qui nous renseignent sur la nature des déterminants des comportements contributifs qui s'approchent de nos données. Nous avons ainsi pu, à partir des données extraites, construire des variables explicatives d'un modèle économétrique expliquant les comportements contributifs par des variables approximant les notions de satisfaction (la note donnée par les consommateurs), d'engagement (la longueur des messages mesurant l'attention portée à la communauté des acheteurs potentiels) et de fidélité que l'on trouve dans la littérature. Notre objectif n'était donc pas de produire des résultats différents de ceux qu'on trouve dans la littérature mais d'explorer, de manière complémentaire aux enquêtes traditionnelles, la possibilité d'utiliser des données extraites massives, tout en restant dans le cadre d'un schéma de causalité (ce qui n'est pas le cas du Big Data).

La complémentarité tient à ce que la construction de proxys à partir des données massivement extraites d'Internet s'appuie sur les résultats produits par les enquêtes par questionnaire et par entretiens. Notre démarche peut se résumer ainsi : la construction de nos hypothèses s'opère par une transformation interprétative de certaines des données extraites en variables proxy explicatives (satisfaction, engagement, fidélité) des comportements contributifs en s'appuyant sur les travaux classiques, ces hypothèses sont ensuite économétriquement testées sur le jeu de données massives. Cette méthode revient à mettre les résultats des études classiques à l'épreuve d'un jeu de données entièrement différent de ceux qui les ont constitués. Nos résultats confortent le rôle de la satisfaction, de l'engagement et de la fidélité des consommateurs dans l'explication des comportements contributifs, trois facteurs solidement établis par la littérature.

L'intérêt serait de tester des facteurs moins triviaux mais nous touchons à une première limite de notre démarche, le jeu de données imposant d'importantes limites quant au nombre et à la nature des motivations possibles. Mais c'est souvent aussi le cas dans la littérature empirique, nombre d'articles se focalisant sur un petit nombre de motivations en raison d'une limitation soit des données, soit de l'objectif que s'assigne l'article (montrer par exemple qu'il peut y avoir des effets de « *crowding out* » entre types de motivation). Notre méthode suppose par ailleurs un travail d'interprétation des données (élaboration de *proxys*) qui est plus important que dans les travaux fondés sur des enquêtes classiques. Le biais déclaratif est en quelque sorte remplacé par un biais d'interprétation. Nous avons cherché à justifier nos interprétations mais nous sommes bien conscients qu'elles sont encore contestables et qu'il reste tout un travail à faire pour les fonder plus solidement.

Ajoutons que, comme nous venons de le souligner, notre démarche ne se substitue pas aux analyses exploitant des données d'observation directe des motivations des contributeurs. Il serait évidemment souhaitable de coupler les deux démarches en appariant les données observées sur Internet avec les données individuelles procédant d'enquêtes sur les motivations. Mais, outre la complexité et le coût des montages requis, l'appariement des deux types de jeux de données poserait de sérieux problèmes de *privacy* que la pseudonymisation ne permet pas de résoudre.

Comment améliorer nos modèles explicatifs à partir du même type de jeux de données, des données observables directement sur Internet. Nous indiquons 5 directions possibles en allant du plus facile au plus difficile : considérer des types de produits différents, utiliser plus avant les données de profil des contributeurs, améliorer la saisie des effets de localisation, traiter la dimension subjective des messages, prendre en considération les *fake reviews*.

Nous n'avons pris en compte que des biens homogènes or il serait intéressant de sélectionner des produits de nature différente, *search goods versus experience goods* ou produits de faible ou de haute qualité, pour examiner si leurs caractéristiques influencent la propension à laisser des avis. Des travaux traitent déjà des rapports des caractéristiques de ces biens à la nature des messages, notamment leur longueur. Il s'agirait d'examiner leurs différences d'impact sur le comportement de contribution.

L'exploitation des données de profil de contributeurs fournit une seconde piste. Les données fournies par les consommateurs sur eux-mêmes sont lacunaires mais d'autres sont fournies automatiquement par

le site pour chaque contributeur : le nombre de votes utiles (feedback des lecteurs des avis) et le nombre de commentaires laissés. Le ratio votes utiles/nombre de commentaires exprime un retour des lecteurs aux contributeurs et produit sur ces derniers un effet de « *warm glow* ». On ajouterait ainsi une motivation différente de l'« altruisme pur » représenté par le souci des autres dans notre modèle.

Il conviendrait ensuite d'améliorer la captation des effets de contexte liés à la localisation des contributeurs. Il serait en particulier intéressant de resserrer l'échelle spatiale pour tester l'impact de la taille urbaine ou de la distinction rural/urbain. Mais les données de localisation fournies par les consommateurs sur les plateformes sont manquantes ou peu fiables. Dans le cas d'Amazon, la donnée de localisation de l'acheteur est fournie par la plateforme au vendeur mais pas au lecteur.

Les deux dernières directions passent par une analyse du contenu des messages et donc l'emploi de techniques de *text mining*. En premier lieu, il serait intéressant d'analyser le rôle des émotions dans la propension à laisser des avis. Les émotions ont été introduites dans la littérature pour analyser leur impact sur la réception des messages par les lecteurs-consommateurs (Kim, Gupta, 2012), en particulier les effets des émotions négatives *versus* ceux des émotions positives. L'analyse du lexique des avis postés permettrait de situer leur impact sur la propension à laisser un avis. La prise en compte du jugement subjectif éclairerait par ailleurs la note accordée.

Il serait en second lieu nécessaire de prendre en considération l'existence des *fake reviews*. Instrumentés par des vendeurs, ces avis biaisés sont en nombre croissant et lèsent tant les consommateurs que la crédibilité des plateformes. Les rédacteurs de ces avis sont recrutés en nombre par des vendeurs sur des réseaux sociaux (He *et al.*, 2020). C'est devenu un sujet majeur et l'objet d'une littérature émergente (Mayzlin, Chevalier, 2014 ; Luca, Zervas, 2016). Or le développement des *fake reviews* peut être un facteur de la propension à laisser des avis. Le plus difficile est de les identifier, d'autant qu'un site comme Amazon s'évertue à les supprimer en masse pour préserver la crédibilité des recommandations. Il est possible de les identifier en croisant un certain nombre de critères (avis non vérifiés, messages courts, utilisation de certaines formules de phrases, fréquence des avis...), voir par exemple les critères utilisés par la plateforme *ReviewMeta*<sup>4</sup> pour filtrer les faux avis) mais ces critères d'identification

---

4. <https://reviewmeta.com/>

sont controversés, les avis identifiés par Amazon comme étant des *fake reviews* semblant posséder de manière retorse des caractéristiques opposées à celles que nous venons de citer (He *et al.*, 2020).

Enfin soulignons les perspectives offertes par l'ouverture aux chercheurs des données des grandes plateformes, ce qui permettrait de dépasser les limites du scrapping de données. Jusqu'ici cette ouverture était à la discrétion des plateformes entretenant des relations avec certains chercheurs ou fournissant des API d'accès à certaines de leurs données. La régulation aujourd'hui s'en mêle soit parce que le régulateur considère que cela fait partie de ses missions (voir la consultation publique récente organisée par l'ARCOM en France, 2023), soit parce qu'une réglementation européenne (le Digital Service Act entré en vigueur en août 2023) en encadre la possibilité pour comprendre l'évolution des risques en ligne entraînés par les plus grandes plateformes.

## RÉFÉRENCES

- ADOUÉ, F. (2017), Mobilité connectée et continuité spatio-temporelle des activités, *Réseaux*, 200(6), 87-115.
- AGUILERA, A., RALLET, A. (2016), Mobilité connectée et changements dans les pratiques de déplacement, *Réseaux*, 200(6), 17-59.
- ARMSTRONG, M. (2006), Competition in Two-sided Markets, *The RAND Journal of Economics*, 37(3), 668-691.
- ANAZA, N. A., ZHAO J. (2013), Encounter-based Antecedents of E-customer Citizenship Behaviors, *Journal of Services Marketing*, 27(2), 130-140.
- ASKALIDIS, G., KIM, S. J., MALTHOUSE, E. C. (2017), Understanding and Overcoming Biases in Online Review Systems, *Decision Support Systems*, 97, 23-30.
- BACACHE-BEAUVALLET, M., BOURREAU, M. (2022), *Économie des plateformes*, Repères économie, Paris, La Découverte.
- BALDWIN, C. Y. (2017), *Explaining the Vertical-to-Horizontal Transition in the Computer Industry*, Working Paper, 17-084, Harvard Business School.
- BALECH, S. (2022), Une application du modèle ELM (Elaboration Likelihood Model) au partage d'information sur Twitter : étude du rôle de la forme du message et du profil de l'émetteur, *Innovations*, 69, 129-161.
- BEJEAN, M., BENAVENT, C., DE PECHPEYROU, P. (2022), Ouvrir la boîte noire des plateformes : structures, acteurs, opérations, *Innovations*, 69, 5-14.
- BEAUVISAGE, T., BEUSCART, J. S., CARDON, V., MELET, K., TRESPÉUCH, M. (2013), Notes et avis des consommateurs sur le web. Les marchés à l'épreuve de l'évaluation profane, *Réseaux*, 177, 131-161.

- BELLEFLAMME, P., PEITZ, M. (2019), Managing Competition on a Two-sided Platform, *Journal of Economics & Management Strategy*, 28(1), 5-22.
- BERRY, L. (1995), Relationship Marketing of Service – Growing Interest, Emerging Perspectives, *Journal of the Academy of Marketing Science*, 23(4), 236-245.
- BHOLE, B., HANNA, B. (2017), The Effectiveness of Online Reviews in the Presence of Self-Selection Bias, *Simulation Modelling Practice and Theory*, 77, 108-123.
- BRODIE, R. J., ILIC, A., JURIC, B., HOLLEBEEK, L. (2013), Consumer Engagement in a Virtual Brand Community: An Exploratory Analysis, *Journal of Business Research*, 66(1), 105-114.
- CHANG, Y., HOU, R. J., WANG, K., CUI, A. P., ZHANG, C. B. (2020), Effects of Intrinsic and Extrinsic Motivation on Social Loafing in Online Travel Communities, *Computers in Human Behavior*, 109, 106360.
- CHEN, Y., FAY, S., WANG, Q. (2011), The Role of Marketing in Social Media: How Online Consumer Reviews Evolve, *Journal of Interactive Marketing*, 25(2), 85-94.
- CHEN, Y., XIE, J. (2008), Online Consumer Review: Word-of-mouth as a New Element of Marketing Communication Mix, *Management Science*, 54(3), 477-491.
- CHEUNG, C., LEE, M. (2012), What Drives Consumers to Spread Electronic Word of Mouth in Online Consumer-Opinion Platforms, *Decision Support Systems*, 53(1), 218-225.
- CHEVALIER, J. A., MAYZLIN, D. (2006), The Effect of Word of Mouth on Sales: Online Book Reviews, *Journal of Marketing Research*, 43(3), 345-354.
- CHIOU, J. S. (2003), The Antecedents of Consumers' Loyalty toward Internet Service Providers, *Information & Management*, 41(6), 685-695.
- CHONG, A., CH'NG, E., LIU, M., LI, B. (2015), Predicting Consumer Product Demands via Big Data: The Roles of Online Promotional Marketing and Online Reviews, *International Journal of Production Research*, 55(17), 5142-5156.
- COQUAZ, V., HALISSAT, I. (2020), *La nouvelle guerre des étoiles : Enquête : nous sommes tous notés*, Paris, Kero.
- CUNNINGHAM, R. (1996), Brand Loyalty: What, Where, How Much?, *Harvard Business Review*, 34(1), 116-128.
- DECI, E. L., RYAN, R. M. (1985), *Conceptualizations of intrinsic motivation and self-determination*, Intrinsic motivation and self-determination in human behavior, Perspectives in Social Psychology book series, 11-40.
- DICHTER, E. (1966), How Word-of-Mouth Advertising Works, *Harvard Business Review*, 44, 147-166.
- BEN-ELIA, E., LYONS, G., MOKHTARIAN, P. (2018), Epilogue: The New Frontiers of Behavioral Research on the Interrelationships between ICT, Activities, Time Use and Mobility, *Transportation*, 45, 479-497.
- FILSER, M. (1994), *Le comportement du consommateur*, Paris, Dalloz.
- FLOYD, K., FRELING, R., ALHOQAIL, S., CHO, H. Y., FRELING, T. (2014), How Online Product Reviews Affect Retail Sales: A Meta-Analysis, *Journal of Retailing*, 90(2), 217-232.

- GREMLER, D. D., BROWN, S. W. (1996), Service Loyalty: Its Nature, Importance, and Implications, in Edvardsson, B., Brown, S. W., Johnston, R., Scheuing, E. E. (eds), *Advancing Service Quality: A Global Perspective*, International Service Quality Association, 171-180.
- GREMLER, D. D., GWINNER, K. P., BROWN, S. W. (2001), Generating Positive Word-Of-Mouth Communication through Customer-Employee Relationships, *International Journal of Service Industry Management*, 12(1), 44-59.
- GVILL, Y., LEVY, S. (2016), Antecedents of Attitudes toward EWOM Communication: Differences across Channels, *Internet Research*, 26(5), 1030-1051.
- HE, S., HOLLENBECK, B., PROSERPIO, D. (2020), The Market for Fake Reviews, *Marketing Science*, 41(5), 896-921.
- HENNIG-THURAU, T., GWINNER, K. P., WALSH, G., GREMLER, D. D. (2004), Electronic Word-of-Mouth via Consumer-Opinion Platforms: What Motivates Consumers to Articulate Themselves on the Internet?, *Journal of Interactive Marketing*, 18(1), 38-52.
- HIKKEROVA, L., PUPION, P. C., SAHUT, J. M. (2015), Fidélisation et fidélité dans l'hôtellerie : une comparaison franco-slovaque, *Management et Avenir*, 78(4), 161-181.
- JEAN, E. (2015), Les enjeux liés à la collecte de données en ligne : le cas d'une recherche auprès de gestionnaires, *La revue des sciences de gestion*, 272, 13-21.
- JEVONS, C., GABBOTT, M. (2000), Trust, Brand Equity, and Brand Reality in Internet Business Relationships: An Interdisciplinary Approach, *Journal of Marketing Management*, 16(6), 619-634.
- JULIAN, L. R., NATALIA, F. (2015), The Use of Web Scraping in Computer Parts and Assembly Price Comparison, *2015 3rd International Conference on New Media (CONMEDIA)*, IEEE, 1-6
- KIM, D., BENBASAT, I. (2003), Trust-Related Arguments in Internet Stores: A Framework for Evaluation, *Journal of Electronic Commerce Research*, 4(2), 49-64.
- KIM, J., JIN, B., SWINNEY, J. L. (2009), The Role of Etail Quality, E-Satisfaction and E-Trust in Online Loyalty Development Process, *Journal of Retailing and Consumer Services*, 16(4), 239-247.
- KIM, J., GUPTA, P. (2012), Emotional Expressions in Online User Reviews: How They Influence Consumers' Product Evaluations, *Journal of Business Research*, 65(7), 985-992.
- LEGUEL, F., PENARD, T., SUIRE, R. (2005), Adoption et usage marchand de l'Internet : une étude économétrique sur données bretonnes, *Economie & Prévisions*, 167, 67-84.
- LEMONNIER, N., SPENCER, T. (2016), Chapitre 6. Altération des avis en ligne : quels moyens pour protéger le consommateur ?, dans Roux, D. (éd.), *Protection des consommateurs. Les nouveaux enjeux du consumérisme*, Caen, EMS Editions, « Versus ».
- LERNER, J., TIROLE, J. (2002), Some Simple Economics of Open Source, *The Journal of Industrial Economics*, 50(2), 197-234.

- LUCA, M., ZERVAS, G. (2016), Fake It till You Make It: Reputation, Competition, and Yelp Review Fraud, *Management Science*, 62(12), 3412-3427.
- MAHADIN, B., AKROUSH, M. N., BATA, H. (2020), The Effects of Tourism Websites' Attributes On E-Satisfaction And E-Loyalty: A Case of American Travelers' to Jordan, *International Journal of Web Based Communities*, 16(1), 4-33.
- MATHWICK, C., MOSTELLER, J. (2017), Online Reviewer Engagement: A Typology Based on Reviewer Motivations, *Journal of Service Research*, 20(2), 204-218.
- MAYZLIN, D., DOVER, Y., CHEVALIER, J. (2014), Promotional Reviews: An Empirical Investigation of Online Review Manipulation, *American Economic Review*, 104, 2421-2455.
- MCKNIGHT, D. H., CHOUDHURY, V., KACMAR, C. (2002), Developing and Validating Trust Measures for E-Commerce: An Integrative Typology, *Information Systems Research*, 13(3), 334-359.
- MOATI, P. (2021), *La plateforme de la consommation*, Paris, Gallimard.
- MORGAN, R. M., HUNT, S. D. (1994), The Commitment-Trust Theory of Relationship Marketing, *Journal of Marketing*, 58(3), 20-38.
- OLIVEIRA, T., ARAUJO, B., TAM, C. (2020), Dataset for Understanding Why People Share Their Travel Experiences on Social Media: Structural Equation Model Analysis, *Data in Brief*, 30, 105447.
- RANAWEERA, C., PRABHU, J. (2003), The Influence of Satisfaction, Trust and Switching Barriers on Customer Retention in a Continuous Purchasing Setting, *International Journal of Service Industry Management*, 14(4), 374-395.
- ROCHET, J.-C., TIROLE, J. (2003), Platform Competition in Two-Sided Markets, *Journal of the European Economic Association*, 1(4), 990-1029.
- THUILLAS, O., WIART, L. (2023), *Les plateformes à la conquête des industries culturelles*, Grenoble, PUG.
- SCHOENMUELLER, V., NETZER, O., STAHL, F. (2020), The Polarity of Online Reviews: Prevalence, Drivers and Implications, *Journal of Marketing Research*, 57(5), 853-877.
- SCOTT, S., ORLIKOWSKI, W. (2012), Reconfiguring Relations of Accountability: Materialization of Social Media in the Travel Sector, *Accounting, Organizations and Society*, 37, 26-40.
- SEILER, S., YAO, S., WANG, W. (2017), Does Online Word of Mouth Increase Demand? (And How?). Evidence from a Natural Experiment, *Marketing Science*, 36(6), 838-861.
- STONE, B. (2013), *The Everything Store: Jeff Bezos and the Age of Amazon*, Little, Brown and Company.
- SUN, Y., DONG, X., MCINTYRE, S. (2017), Motivation of User-Generated Content: Social Connectedness Moderates the Effects of Monetary Rewards, *Marketing Science*, 36(3), 327-470.
- SUNDARARAJAN, A. (2016), *The Sharing Economy: The End of Employment and the Rise of Crowd-Based Capitalism*, Cambridge, MIT Press.

- THAKUR, R. (2018), Customer Engagement and Online Reviews, *Journal of Retailing and Consumer Services*, 41, 48-59.
- THOUMELIN, C. (2020), Acheter en ligne des produits culturels, 2010-2018, *Culture chiffres*, 1, 1-13.
- TONG, Y., WANG, X., TAN, C. H., TEO, H. H. (2013), An Empirical Study of Information Contribution to Online Feedback Systems: A Motivation Perspective, *Information & Management*, 50(7), 562-570.
- TSE, D., WILTON, P. (1988), Models of Consumer Satisfaction Formation: An Extension, *Journal of Marketing Research*, 25, 204-2012.
- TUCKER, W. (1964), The Development of Brand Loyalty, *Journal of Marketing Research*, 1(3), 32-35.
- VAN DOORN, J., LEMON, K. N., MITTAL, V., NASS, S., PICK, D., PIRNER, P., VERHOEF, P. C. (2013), Customer Engagement Behavior: Theoretical Foundations and Research Directions, *Journal of Service Research*, 13(3), 253-266.
- VAUTROT, P., BENAVENT, C., DE PECHPEYROU, P. (2018), Satisfaction à l'égard des moteurs de recommandation : la transparence optimale, *Management & Datascience*, 2(1), 9-17.
- WEISSTEIN, F. L., SONG, L., ANDERSON, P., ZHU, Y. (2017), Examining Impacts of Negative Reviews and Purchase Goals on Consumer Purchase Decision, *Journal of Retailing and Consumer Services*, 39, 201-207.
- WU, P. F. (2019), Motivation Crowding in Online Product Reviewing: A Qualitative Study of Amazon Reviewers, *Information & Management*, 56(8), 103163.
- WU, C., CHE, H., CHAN T. Y., LU, W. (2015), The Economic Value of Online Reviews, *Marketing Science*, 34(5), 739-754.
- ZHU, F., ZHANG, X. (2010), Impact of Online Consumer Reviews on Sales: The Moderating Role of Product and Consumer Characteristics, *Journal of Marketing*, 74(2), 133-148.

## Annexes

**Tableau A - Statistiques descriptives par État**

État	Âge moyen	Diplômés du supérieur	Revenu moyen
AK	33.5	7346	76440
AL	39	69329	46257
AR	38	45573	44334
AZ	37.5	173003	53558
CA	36.4	565212	67739
CO	36.7	89945	65685
CT	40.9	54430	73433

État	Âge moyen	Diplômés du supérieur	Revenu moyen
DE	40.6	14675	61757
FL	42.1	323778	50860
GA	36.5	140578	53559
HI	38.9	15465	74511
IA	38	85954	56247
ID	36.1	22041	51807
IL	37.9	220500	60960
IN	37.6	105518	52314
KS	36.5	51858	54935
KY	38.7	73083	46659
LA	36.5	73938	45146
MA	39.5	138468	75297
MD	38.5	87213	78945
ME	44.5	16057	53079
MI	39.7	146637	52492
MN	37.9	106122	65599
MO	38.4	105678	51746
MS	37.2	42318	41754
MT	40.1	11991	50027
NC	38.7	142854	50584
ND	35	12646	60656
NE	36.3	31528	56927
NH	42.7	24490	70936
NJ	39.5	102305	76126
NM	37.7	32910	46748
NV	37.9	26610	55180
NY	38.4	335061	62909
OH	39.3	170892	52334
OK	36.4	56546	49176
OR	39.2	61400	57532
PA	40.6	206164	56907
PR	40.7	60657	20078
RI	40.2	21309	60596
SC	39.1	57552	49501
SD	36.8	11761	54467
TN	38.6	84665	48547
TX	34.5	338219	56565
UT	30.7	69068	65977
VA	38.2	138122	68114
VT	43.1	11206	57677
WA	37.7	108019	67106
WI	39.4	90700	56811

État	Âge moyen	Diplômés du supérieur	Revenu moyen
WV	42.3	35161	43385
WY	37.2	7904	59882

**Tableau B - Matrice de corrélations - consommateurs**

	Âge	Revenu	Éducation	Usage Internet	Usage mobile	Trajet mobile	Trajet voiture
Âge	1						
Revenu	-0,1896	1					
Éducation	-0,2931	0,2815	1				
Usage Internet	-0,1705	0,8419	0,3459	1			
Usage mobile	-0,4722	0,1462	0,1273	0,2593	1		
Trajet mobile	-0,2612	0,5528	0,5237	0,4773	0,5582	1	
Trajet voiture	0,0396	0,54	0,5232	0,3641	-0,0666	0,7864	1

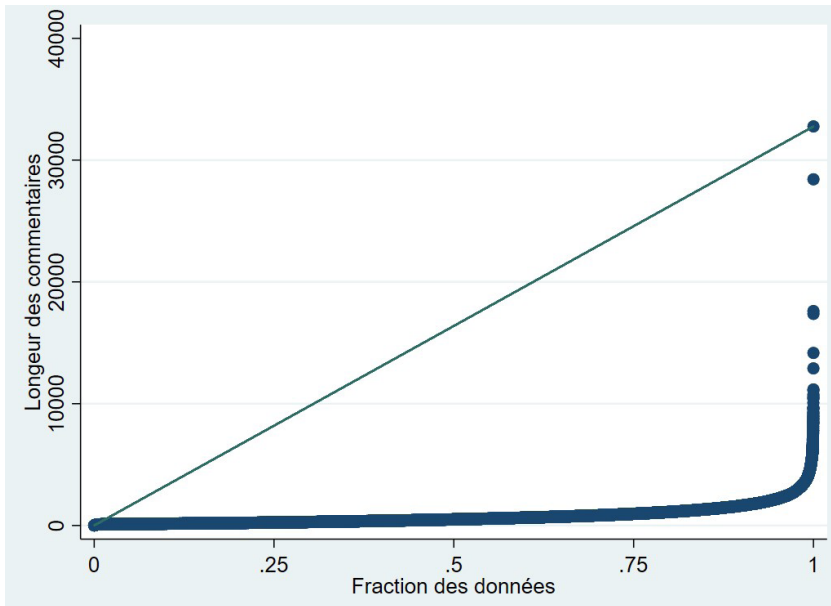
**Tableau C - Matrice de corrélations - États**

	Âge	Revenu	Éducation	Usage Internet	Usage mobile	Trajet mobile	Trajet voiture
Âge	1						
Revenu	-0,0801	1					
Éducation	-0,0872	0,101	1				
Usage Internet	-0,0548	0,8594	0,1867	1			
Usage mobile	-0,3287	0,117	0,1804	0,1134	1		
Trajet mobile	0,045	0,3301	0,501	0,2417	0,6601	1	
Trajet voiture	0,2901	0,3493	0,5281	0,2298	0,2027	0,8644	1

Tableau D – Matrice de corrélations

	Avis	Satisfaction	Fidélité	Engagement	Digital	Film	Jeux	Instruments	Âge	Revenu	Études	Internet	Mobile	Trajet/mobile	Trajet/voiture
Avis	1.0000														
Satisfaction	0.0029	1.0000													
Fidélité	0.0475	-0.0026	1.0000												
Engagement	0.1349	-0.0768	0.0831	1.0000											
Digital	0.3425	0.0103	0.0251	0.0653	1.0000										
Film	0.0878	-0.0484	0.0071	-0.0915	-0.0456	1.0000									
Jeux	0.0676	-0.0502	0.0151	0.1292	-0.0119	-0.2316	1.0000								
Instruments	0.0068	0.0198	-0.0113	-0.0194	-0.0073	-0.0629	-0.0059	1.0000							
Âge	0.0076	0.0049	-0.0034	0.0019	0.0056	-0.0104	-0.0057	0.0026	1.0000						
Revenu	0.0044	-0.0249	0.0300	0.0529	0.0016	-0.0165	-0.0015	-0.0072	-0.1896	1.0000					
Études	0.0028	-0.0054	-0.0017	0.0108	-0.0018	-0.0022	0.0110	-0.0061	0.1261	0.0501	1.0000				
Internet	-0.0006	-0.0287	0.0255	0.0443	-0.0021	-0.0092	0.0067	-0.0046	-0.1705	0.8419	0.0940	1.0000			
Mobile	-0.0092	-0.0017	-0.0060	-0.0042	-0.0104	0.0163	-0.0003	-0.0003	-0.4722	0.1462	-0.0814	0.2593	1.0000		
Trajet/mobile	0.0014	-0.0212	0.0173	0.0334	0.0085	-0.0157	-0.0089	-0.0015	-0.2612	0.5528	-0.1043	0.4773	0.5582	1.0000	
Trajet/voiture	0.0078	-0.0247	0.0254	0.0430	0.0175	-0.0310	-0.0107	-0.0016	0.0395	0.5401	-0.0564	0.3641	-0.0666	0.7864	1.0000

**Graphique A - Longueur des commentaires**



**Tableau E - Résultats du modèle (6) avec les détails des dummies « État »**

		<b>(6)</b>	
<b>Nombre d'avis</b>			
	Satisfaction	1,165***	(0,099)
	Fidélité	0,000***	(0,000)
	Engagement	0,002***	(0,000)
	<i>Catégorie de produits</i>		
	CD et Vinyles	<i>Ref.</i>	
	Musique num,	34,261***	(1,287)
	Films et TV	4,744***	(0,140)
	Jeux vidéo	4,867***	(0,285)
	Instrument de musique	4,530***	(0,291)
	<b>Variable de contrôle</b>		
	Âge	0,830	(0,974)
	Revenu	0,000	(0,000)
	Niveau d'études	-78,873	(224,024)
	Usage d'Internet	-0,431	(0,439)
	Usage d'Internet mobile	-1,467*	(0,764)

	<b>(6)</b>	
Trajet		
Trajet*mobile	0,061*	(0,034)
Trajet*voiture	-1,927*	(1,095)
<i>Etats</i>		
Alaska	Ref.	
Alabama	-3,694	(3,228)
Arkansas	-3,185	(3,389)
Arizona	0,454	(0,470)
California	-1,234	(3,182)
Colorado	0,004	(1,024)
Connecticut	-5,933	(4,919)
Delaware	-4,734	(3,792)
Florida	-3,840	(2,905)
Georgia	-0,264	(4,526)
Hawaii	-6,365	(4,664)
Iowa	-1,951	(4,020)
Idaho	2,187	(1,707)
Illinois	-1,888	(2,644)
Indiana	-1,814	(2,263)
Kansas	-3,201	(2,051)
Kentucky	-2,221	(2,040)
Louisiana	-0,492	(3,759)
Massachusetts	-6,702	(5,704)
Maine	-5,690	(5,804)
Michigan	-2,597	(2,426)
Minnesota	-2,089	(2,577)
Mississippi	-2,167	(4,391)
Montana	-7,692*	(4,644)
North Carolina	-2,722	(3,474)
North Dakota	-1,276	(2,649)
Nebraska	-7,681	(4,871)
New Hampshire	-5,747	(5,068)
New Jersey	-2,800	(3,313)
New Mexico	-0,581	(3,498)
Nevada	-2,524	(3,161)
New York	1,401	(4,465)
Ohio	-2,144	(1,987)
Oklahoma	-2,819	(3,320)
Oregon	-1,558	(1,850)
Pennsylvania	-3,153	(2,636)
Puerto Rico	0,000	(2,636)
Rhode Island	-4,400	(2,687)

		<b>(6)</b>	
	South Dakota	-10,243*	(4,812)
	Tennessee	-3,092	(3,288)
	Texas	-0,553	(4,334)
	Utah	3,258	(4,364)
	Virginia	-3,043	(2,945)
	Vermont	-7,091	(6,008)
	Washington	-2,270	(2,009)
	Wisconsin	-3,611	(2,741)
	West Virginia	-2,312	(4,553)
	Wyoming	-4,638**	(2,353)
	Missouri	-2,099	(1,918)
	South Carolina	-3,670	(3,215)
	Maryland	-4,726	(3,921)
	_cons	83,590	(59,690)
	R <sup>2</sup>	0,15	
	N	92,888	