

Marketing

BOOK

Data Marketing

Statistiques appliquées au marketing avec **Excel** et **R**

Magali TRELOHAN

Fichiers complémentaires
à télécharger



Chapitre 1

Les data, pour quoi faire ?

A. Introduction	11
B. Data mining, big data, smart data, open data : comprendre l'univers des data ...	12
1. L'arrivée des data dans la gestion des entreprises	12
a. L'exemple de la distribution	13
b. Le développement des entreprises d'études de marché	14
2. Comprendre les termes data.	17
a. Le data mining	17
b. Le big data	18
c. Le smart data	22
d. Les open data	22
3. L'intelligence artificielle.	23
a. Qu'est-ce que l'IA est et n'est pas ?	24
b. Application de l'IA : le marketing automation	25
c. Application de l'IA : les chatbots	26
d. Application de l'IA : la reconnaissance faciale	26
4. La réalité virtuelle.	26
C. Les acteurs du data marketing	27
1. Le chief data officer (CDO)	27
2. Le data scientist/analyst	27
3. Le data protection officer (DPO)	28
4. Le data marketer.	28
D. Data et marketing : quelles finalités ?	28
1. Stratégie d'entreprise et data marketing.	29
a. L'itération entre stratégie d'entreprise, stratégie marketing et stratégie data.	29
b. Pas de stratégie data sans objectifs SMART.	30
2. Les data au sein du plan marketing.	31
3. Analyser la demande : comprendre ses clients actuels et potentiels	32
a. Les études de marché traditionnelles	32
b. Les échanges spontanés entre consommateurs.	32
c. Les mesures comportementales.	33
4. Analyser la concurrence.	34
5. Mettre en place une stratégie : segmentation et ciblage.	35

6. Optimiser ses actions marketing : le bon produit, au bon prix, au bon endroit avec la bonne communication et les bonnes actions commerciales	37
a. Rappels sur le marketing mix	37
b. Le produit	39
c. Le prix	40
d. La distribution	41
e. La communication	44

Chapitre 2

Les outils du data marketing

A. Introduction	49
B. Structurer et analyser ses données	50
1. Excel : usages et fonctionnalités	50
a. Visualiser rapidement la moyenne, le nombre et la somme d'une variable	50
b. Filtrer les données	51
c. Trier les données	52
d. Figer les volets	53
e. La mise en forme conditionnelle	55
f. Le tableau de contingence : le tableau croisé	55
g. Convertir	57
2. R : Prise en mains	61
a. L'installation de R	61
b. RStudio	64
c. Définir le répertoire de travail	65
d. Télécharger un fichier de données dans R	66
e. Ouvrir un onglet de script	69
f. Installer et charger des packages	70
g. Le package Rcmdr	71
C. Gérer les big data	74
1. La problématique de l'analyse de logs pour comprendre les comportements des consommateurs en ligne	75
2. L'univers Hadoop	76
a. Naissance et enjeux d'Hadoop	76
b. Le principe MapReduce	77
c. Hadoop, un framework puissant, mais demandant un niveau de technicité élevé	78
d. Les briques Hadoop	78

3. Les big data dans R	79
a. Hadoop dans R	79
b. Traitements statistiques de big data dans R	80

Chapitre 3

Auditer son système marketing

A. Collecter des données	85
1. Le type de données et ce que l'on peut en faire	85
a. Les données structurées	85
b. Les données non structurées	87
2. La richesse de vos données	90
a. Les sources d'information en interne	90
b. Type d'informations disponibles en interne	91
3. Les données externes	91
a. Open data	91
b. Données de panel	92
c. Données de tracking	92
B. Stocker ses données	93
1. Faire un état des lieux de ses données	93
2. Les solutions techniques	93
3. Vision à 360° du client : DMP, data lakes	94
a. Les data management platform (DMP)	94
b. Les data lakes	94
c. Est-ce pour mon entreprise ?	95
C. Nettoyer ses bases de données	95
1. Qu'est-ce qu'une base de données propre ?	96
2. Pourquoi « nettoyer » sa base de données ?	96
a. Éviter la perte de temps	96
b. Ne pas renforcer ou susciter de la résistance à une stratégie data-driven	96
c. Faciliter la communication entre les différents outils de data	97
d. Ne pas nuire à la relation client	97
3. Comment procéder (les données manquantes, la clé d'identification, les données anormales) ?	97
a. Définir l'unité statistique d'analyse	98
b. Adopter une logique de table	99
c. Supprimer les variables inutiles	100
d. Donner une clé unique à vos sujets	100

e. Visualiser ses données dans R	100
f. Effectuer quelques tests statistiques basiques pour s'assurer que les variables soient bien définies.	101
g. Supprimer les doublons.	101
h. Nommer les variables de manière efficace	105
i. Coder les variables de manière efficace.	105
j. Traiter les données manquantes	107
k. Traiter les données anormales ou aberrantes (les outliers)	110

Chapitre 4

Que s'est-il passé ? L'analyse descriptive des données et le diagnostic

A. Introduction	115
B. Choisir ses KPI	115
1. Plan : planifier les actions marketing et les objectifs	116
2. Do : la phase d'action.	117
3. Check : la phase d'analyse et de diagnostic.	117
4. Act : ajuster les actions ou poursuivre sur la même voie	117
C. Analyse univariée : étudier les variables une à une	118
1. La tendance centrale	119
a. La moyenne.	119
b. La médiane	120
c. Le mode	122
2. La dispersion.	126
a. La variance	126
b. L'écart-type	128
c. La fréquence	129
d. L'étendue, le minimum et le maximum	131
e. Le fractile (quartile)	132
f. Un outil de visualisation des principaux indices : la boîte à moustaches.	133
3. L'intervalle de confiance	137
D. Analyse bivariable : faire des rapprochements entre deux variables	138
1. Deux variables qualitatives	138
a. Créer un tableau de contingence dans R.	138
b. Test du Khi-deux dans R.	140
2. Deux variables numériques : la corrélation	142
a. La covariance	143
b. Le coefficient de corrélation linéaire	146

c. La significativité du coefficient de corrélation linéaire.....	148
3. Deux variables ordinales.....	149
a. Le coefficient de corrélation de Spearman.....	150
b. La significativité du coefficient de corrélation de Spearman.....	151
4. Test de comparaison de moyennes.....	152
a. Préalable statistique à la comparaison de moyenne.....	153
b. Test de comparaison de moyennes : t de Student.....	154
E. Conclusion.....	157

Chapitre 5

Que peut-il se passer dans le futur ? Le marketing prédictif

A. Introduction.....	161
B. La classification : segmenter ses clients ou ses prospects.....	161
1. La classification, une des méthodes d'analyse multivariée.....	162
2. Méthodologie.....	162
3. Étape 1 : Analyse en composantes principales.....	163
a. La matrice des coefficients de corrélation.....	164
b. Réalisation de l'ACP sous R.....	164
c. Interprétation des dimensions.....	168
4. Étape 2 : la classification ascendante hiérarchique.....	169
a. Regroupement des individus par classes.....	169
b. Réalisation de la Classification Ascendante Hiérarchique (CAH) sous R.....	169
c. Lecture des résultats de la CAH.....	173
C. Les modèles de régression.....	178
1. Méthodologie.....	178
a. Prévoir le prix de vente d'un bien immobilier : valeur continue et régression multiple.....	178
b. Prévoir l'achat de glaces : valeur catégorielle et régression logistique.....	181
2. Applications pratiques.....	184
D. Le scoring.....	184
1. Data mining et scoring.....	185
2. Applications des méthodes de scoring.....	186
a. Score de risque.....	186
b. Score d'appétence.....	186
c. Score d'attrition.....	187
3. Les différents types d'analyses.....	187

4. Le scoring par régression logistique	188
a. La régression logistique binaire	188
b. Mise en œuvre du scoring par la régression logistique binaire dans R ...	189
E. Conclusion.....	198

Chapitre 6

Que devrions-nous faire ? Le marketing prescriptif

A. Introduction	201
B. À quelles problématiques répond le marketing prescriptif ?	202
C. Mettre en place une expérimentation.....	204
1. Conditions réelles ou en laboratoire	204
a. L'expérimentation en laboratoire	204
b. L'expérimentation en conditions réelles.....	205
2. Les éléments à considérer pour mettre en place une expérimentation	205
a. Les traitements à comparer	206
b. Les critères d'évaluation de ces traitements.....	206
c. Les variables externes	207
d. Les sujets.....	207
3. La validité de l'étude	208
a. La validité interne et externe	208
4. Les biais à contrôler	208
a. Les effets des paramètres extérieurs à l'expérimentation	208
b. L'effet de sélection.....	208
c. L'effet d'histoire	209
d. L'effet de maturation.....	209
e. L'effet de test	209
f. L'effet d'instrumentation	209
g. L'effet de mortalité expérimentale	210
h. L'effet de halo	210
5. Les plans d'expérience.....	210
a. Le plan post-traitement	211
b. Le plan pré et post-traitement	212
c. Le plan de Solomon	212
d. Le plan factoriel complet.....	213
e. Plan factoriel à blocs aléatoires	213
f. Plan en carré latin ou en carré gréco-latin	215

D. Analyser les résultats	215
1. Quelques préalables	215
a. Qu'est-ce qu'un résultat significatif ?	215
b. Mes données suivent-elles une loi normale ?	216
c. Les variances sont-elles égales entre les populations ?	219
2. Comparaison de deux moyennes	220
3. Analyse de variance (ANOVA)	222
a. ANOVA à un facteur	222
b. ANOVA à plusieurs facteurs	225
c. MANOVA	229
E. Conclusion	230

Chapitre 7

Comment communiquer ? Règles de base, data visualisation et data storytelling

A. Introduction	233
B. Quelques règles basiques pour bien présenter sa stratégie data-driven	234
1. Communiquer sur ses données : quels éléments-clé prendre en compte ?	234
a. L'émetteur du message	234
b. Le message	235
c. Les interlocuteurs	235
d. Les canaux	235
e. Quels effets ?	236
2. Le traitement des messages : que se passe-t-il chez mon interlocuteur quand je communique ?	236
C. Les outils de la data visualisation	238
1. Quelle est l'idée-clé de mon graphique ?	240
2. Quelles données sont les plus pertinentes pour démontrer mon idée ?	240
3. Quelle représentation graphique est la plus adaptée ?	240
4. Applications dans Excel	241
D. Le data storytelling	244
1. Écrivez votre fil conducteur	244
2. Accrocher votre public	244
3. Plantez votre décor	245
4. Ne faites pas trop de teasing	245
5. Ne vous perdez pas dans les détails	245
6. Ayez une chute efficace	245
7. Inventez un personnage	245

E. Conclusion.	246
---------------------	-----

Chapitre 8

Conclusion

A. 6 éléments-clé de la réussite de votre stratégie data-driven	249
1. Construire une culture data-driven	249
2. Travailler en équipe	249
3. Le pragmatisme et le bon sens orienté client.	250
4. La plus-value des data	250
5. Commencer simple mais efficace.	250
6. Avoir une attitude responsable vis-à-vis des données	251

Bibliographie.	253
---------------------	-----

Index	257
-------------	-----



A. Introduction

Les chapitres précédents nous ont permis de comprendre l'intérêt des data dans la démarche marketing (chapitre Les data, pour quoi faire ?), de découvrir les outils de la data (chapitre Les outils du data marketing) et d'auditer son système data (chapitre Auditer son système marketing). Dans ce quatrième chapitre, nous allons rentrer concrètement dans la manipulation des données. Pour cela, nous allons aborder le premier niveau d'analyse des données, l'analyse descriptive. Nous verrons également certains tests qui relèvent des statistiques inférentielles.



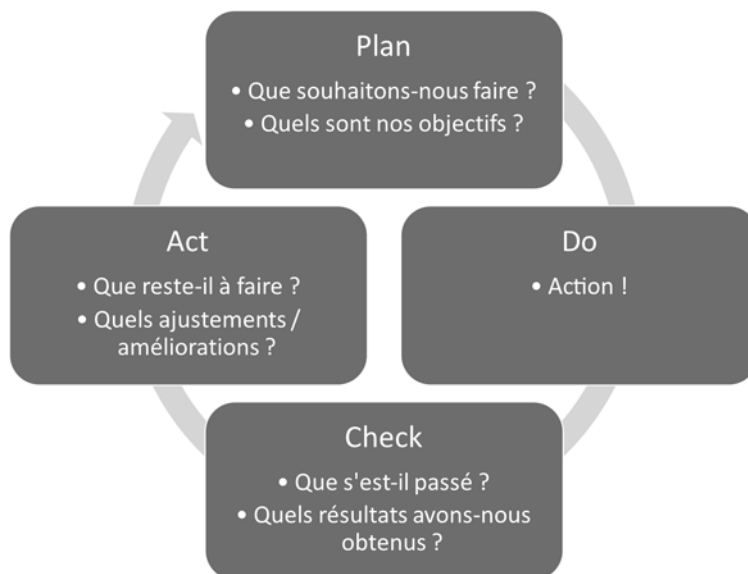
Les statistiques inférentielles permettent, à partir d'un échantillon, d'extrapoler des comportements sur la population totale.

Les outils utilisés seront Excel ou R. Ils nous permettront très simplement d'avoir une première approche des données.

B. Choisir ses KPI

Les KPI (*Key Performance Indicators*) sont les indicateurs qui vont vous permettre de suivre votre activité marketing et commerciale. Ils vous permettent de donner un cap à vos actions (objectifs), de mesurer la réussite des actions lorsqu'elles ont été mises en œuvre et, éventuellement, de réajuster ces actions dans le temps en fonction de l'analyse que vous aurez faite des résultats.

Il s'agit d'une démarche PDCA : Plan, Do, Act, Check qui doit guider vos actions et votre stratégie.



Les data sont au service de cette démarche d'amélioration. Elles vont permettre de planifier l'activité et ses objectifs, de mesurer a posteriori les résultats obtenus en les comparant aux objectifs et aux résultats des périodes précédentes. Enfin, elles permettront de réajuster les actions futures.

1. Plan : planifier les actions marketing et les objectifs

Les KPI vont donc être l'outil de pilotage de votre stratégie marketing et de chacune de vos actions. Ils vont également permettre de comparer les résultats entre eux. Par exemple, par canal d'achat, par établissement, par commercial, etc.

Le nombre de KPI pouvant être suivi est infini. Ils dépendent de votre activité (online ou offline, B to C ou B to B, du secteur d'activité), de votre position sur le marché (outsider, suiveur, leader) et de votre stratégie. Il convient de retenir ceux qui vous correspondent le mieux. Il est important d'intégrer l'ensemble de l'équipe dans le choix et la définition des KPI. En effet, si vous avez une vision stratégique, vos collaborateurs ont, eux, une vision plus opérationnelle et pourront vous aiguiller sur la pertinence de tel ou tel indicateur. Également, ils pourront vous indiquer si la donnée correspondante peut réellement être collectée. Enfin, dans la mesure où les KPI peuvent servir de base à l'évaluation de vos équipes, il est préférable de les intégrer en amont dans leur définition pour en faciliter leur compréhension, leur acceptation et l'implication de chacun dans l'atteinte des objectifs.

Vous définirez également la temporalité pertinente pour mesurer chacun de ces indicateurs : par mois, semaine, jour, heure.

2. Do : la phase d'action

La phase d'action est celle où les actions sont mises en œuvre. Il peut s'agir de faire évoluer le site web, d'envoyer une newsletter ou un mailing. Cela peut également être le nombre de leads générés par une nouvelle action marketing.

Cette phase est plus ou moins longue selon la temporalité définie précédemment. Cela peut être une journée, une semaine, un mois ou une année de ventes dans un point de vente.

Si sa durée est longue, il sera préférable d'inscrire des points d'étape pour éventuellement corriger en cours de route les actions.

3. Check : la phase d'analyse et de diagnostic

Une fois la fin de la période définie atteinte, il est temps de passer à l'analyse. Avec les données collectées, vous pourrez procéder au premier niveau d'analyse de vos données : l'analyse descriptive. L'**analyse descriptive** correspond à une phase d'analyse du passé pour mesurer ce qu'il s'est passé et en tirer des enseignements.

Elle est trop souvent négligée au profit de l'action et cette négligence entraîne des ajustements tactiques parfois infondés ou mal orientés. Comme chez le médecin, il faut prendre le temps de discuter, de lire les données (que ce soit un bilan sanguin ou un tableau de bord de résultats) pour poser le bon diagnostic, car c'est ce bon diagnostic qui permettra d'établir la bonne prescription.

Dans cette phase, les données chiffrées seront essentielles, mais il ne faudra pas oublier de discuter avec ceux qui ont participé à l'action pour avoir des feedbacks qualitatifs. C'est votre capacité à confronter ces deux éléments d'analyse qui vous permettra de poser le bon diagnostic.

4. Act : ajuster les actions ou poursuivre sur la même voie

En fonction des résultats obtenus, vous pourrez prendre deux grands types de décisions : ajuster les actions ou poursuivre les actions mises en place. Si des résultats mauvais orientent sur la première voie et des résultats excellents sur la seconde, lorsque les résultats sont mitigés, c'est votre finesse d'analyse qui vous guidera. Plus les résultats sont entre deux eaux, plus il faudra faire de tests pour comprendre en profondeur la situation.

Voyons maintenant concrètement les tests qui peuvent être menés en ce sens.

C. Analyse univariée : étudier les variables une à une

La première étape de l’analyse descriptive consiste à analyser les variables une à une pour avoir une première compréhension des données disponibles. On appelle cette analyse l’analyse univariée en opposition avec les analyses multivariées qui traitent de plusieurs variables à la fois. L’analyse descriptive univariée comprend des mesures souvent connues, mais parfois mal interprétées, nous allons donc passer en revue les indices statistiques présentant les tendances centrales (la moyenne, la médiane, le mode), la dispersion (la variance, l’écart-type, les minimum et maximum, la fréquence et les fractiles) et la forme des données (asymétrie et aplatissement). Enfin, nous aborderons en complément l’intervalle de confiance.

L’objectif ici est de résumer au mieux chacune des variables afin de faciliter la compréhension d’un phénomène (achat de produits, visites en magasin, visites sur un site Internet par exemple). Nous allons donc passer d’une base de données comprenant de nombreuses observations à quelques indicateurs-clés. Pour illustrer ce type d’analyse, nous prendrons l’exemple d’un hôtel qui propose depuis peu ses chambres à différents prix en fonction de la date de réservation et du support de réservation (site Internet, référenteur, etc.).

Il souhaite s’assurer de la rentabilité de cette nouvelle stratégie de prix. La base de données que vous pouvez retrouver dans le classeur `BDD_exemples_chapitre_4.xlsx`, comprend 200 réservations et se présente comme suit :

Client	Canal achat	Type de chambre	Délai réservation (en jours)	Nombre de nuits	Tarif _nuit	CA	satisfaction
927	téléphone	supérieure	4	2	101	202,00	9
549	booking.com	suite familiale	16	2	177	354,00	10
529	mail	suite familiale	6	5	173	865,00	10
885	trivago.fr	standard	6	2	83	166,00	8
883	booking.com	suite familiale	10	4	161	644,00	4
999	booking.com	suite familiale	9	6	170	1020,00	9
960	site internet	standard	8	5	110	550,00	7
419	téléphone	supérieure	8	4	135	540,00	10
434	tripadvisor.fr	standard	7	4	80	320,00	6
517	téléphone	supérieure	5	6	123	738,00	8

À partir de cette base de données, nous allons calculer les indices de tendance centrale et de dispersion.

1. La tendance centrale

Dans le chapitre Auditer son système marketing, nous avons évoqué les différents types de variables : numériques, ordinales ou nominales. Lorsque nous souhaitons connaître la tendance centrale de données, plusieurs indices statistiques existent. Le choix d'un indice dépend du type de variable. Ainsi pour des données numériques (par exemple, le CA) nous pouvons calculer la moyenne, la médiane et le mode. Pour des données ordinales (évaluation de l'image d'une marque), nous pourrions calculer la médiane et le mode. Enfin, pour des données nominales, nous pourrions calculer le mode (canal d'achat, par exemple).

Type de variables	Indice de tendance centrale
Numérique	Moyenne, Médiane, Mode
Ordinales	Médiane, Mode
Nominale	Mode

a. La moyenne

La moyenne est un indice de tendance centrale qui permet de résumer les valeurs d'une variable numérique. Dans l'exemple de l'hôtel, nous souhaitons dans un premier temps calculer le prix moyen payé par les clients. La formule de la moyenne est la suivante :

$$\overline{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

☞ Dans la cellule F202 d'Excel, saisissez =moyenne (F2 : F201) .

Cette formule demande à Excel de calculer la moyenne des valeurs allant de la cellule F2 à la cellule F201.

Accueil Insertion Mise en page Formules Données Révision Affichage								
PERSONNALISÉE								
Mise en forme conditionnelle Mettre sous forme de tableau Styles de cellule								
ALEA.EN... X fx =MOYENNE(F2:F201)								
	A	B	C	D	E	F	G	H
1	Client	Canal achat	Type de chambre	Délai réservation (en jours)	Nombre de nuits	Tarif_nuit	CA	satisfaction
192	882	trivago.fr	suite familiale	3	4	186	744,00	10
193	819	mail	suite familiale	3	4	186	744,00	9
194	781	booking.com	suite familiale	9	1	186	186,00	5
195	664	mail	suite familiale	14	3	186	558,00	7
196	555	trivago.fr	suite familiale	8	6	187	1122,00	8
197	323	site internet	suite familiale	3	2	187	374,00	8
198	136	mail	suite familiale	7	2	189	378,00	8
199	627	tripadvisor.fr	suite familiale	1	5	190	950,00	10
200	891	site internet	suite familiale	5	4	190	760,00	10
201	217	booking.com	suite familiale	7	1	190	190,00	6
202	=MOYENNE(F2:F201)							

On obtient pour cette base de données une moyenne de 136,89 €. Nous verrons dans la section Analyse bivariée : faire des rapprochements entre deux variables de ce chapitre comment aller plus loin en comparant les moyennes selon une deuxième variable (canal d'achat, type de chambre, etc.).

b. La médiane

La médiane est un indice de tendance centrale pouvant être utilisée pour les données numériques et ordinales. Elle indique la valeur de la variable qui partage l'échantillon en deux parts égales. Pour la calculer, nous trions les valeurs de la variable par ordre croissant. La valeur partageant l'échantillon en deux parts égales est la médiane.

Si l'échantillon comprend un nombre impair d'observations, elle est facile à repérer. Par exemple, s'il existe 201 clients dans un fichier. Le 101e client (après le tri de la variable en ordre croissant) portera la valeur médiane, car il y aura 100 clients portant une valeur inférieure (clients 1 à 100) et 100 clients portant une valeur supérieure (clients 102 à 201).

Si l'échantillon comprend un nombre pair d'observations, il s'agira alors de toutes valeurs se situant entre les deux valeurs centrales. Ainsi, dans notre exemple de l'hôtel, il y a 200 clients. Comme il s'agit d'un nombre pair, la médiane peut être toute valeur se situant entre celle du client 100 et celle du client 101 (toujours après tri en ordre croissant des valeurs). Pour donner une valeur unique, il est admis d'utiliser la moyenne des valeurs des clients 100 et 101.