

Calcul de l'Uplift drive to store, attention aux erreurs statistiques !

Afin de mesurer la performance de campagnes publicitaires mobile drive-to-store, de nombreux acteurs du marché de l'AdTech proposent d'évaluer l'efficacité d'une campagne en calculant l'« uplift ». Cet indicateur cherche à évaluer le nombre de mobinautes visiteurs du point de vente *qui ne s'y seraient pas rendus s'ils n'avaient pas vu la publicité*.

La population exposée à la campagne est découpée en 3 catégories : les non-visiteurs, les visiteurs « organiques » et les visiteurs « générés » :

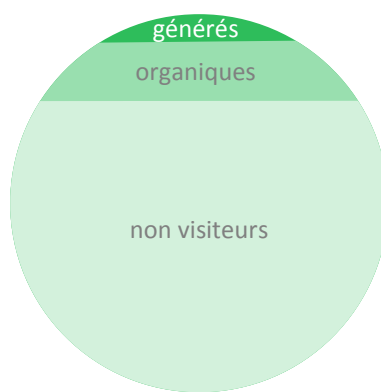


Figure 1 - Répartition des mobinautes touchés par la campagne publicitaire

- Les non visiteurs ont vu la campagne mais ne se sont pas rendus en magasin ;
- Les "organiques" ont vu la publicité et se sont rendus en magasin, mais s'y seraient rendus même s'ils n'avaient pas vu la publicité. Leur visite n'est donc pas attribuable à la campagne publicitaire ;
- Les "générés" ne se sont rendus en magasin que parce qu'ils ont vu la campagne.

L'uplift est défini comme *la variation du taux de visite des mobinautes exposés à la campagne par rapport à la moyenne*. Concrètement cela revient à **calculer le nombre de visiteurs générés parmi la population exposée**.

Cette pratique est notamment issue du web où de nombreux outils de tracking existent et permettent de suivre la navigation de l'internaute, du site média jusqu'à des pages identifiées sur le site de l'annonceur. Obtenir le même niveau d'information dans le monde physique se révèle plus compliqué, et on remarque aujourd'hui que des erreurs existent dans les méthodes de calcul utilisées par certains acteurs du marché.

Les règles d'attribution usuellement utilisées sur le web ne pouvant fonctionner dans la réalité physique, l'uplift drive-to-store dans ce contexte fait l'objet d'un **calcul statistique** : on cherche à estimer, parmi les mobinautes visitant un point de vente, combien y seraient venus « de toute façon », c'est-à-dire même sans avoir vu la publicité. On possède d'un côté le nombre total de mobinautes

exposés, de l'autre le nombre de visiteurs en point de vente, et on cherche à calculer la part de visiteurs organiques parmi les mobinautes touchés pour en déduire les mobinautes générés. Ce calcul est statistique, et il est donc capital de s'assurer de la pertinence du résultat en surveillant deux éléments : l'**absence de biais**, et la **significativité**.

Or la méthode la plus couramment employée aujourd'hui par le marché consiste à séparer **aléatoirement** la base de devices (mobiles) disponibles en deux parties : l'**audience « cible »** à qui on essaiera de diffuser la publicité avec un méthode de geofencing, et l'**audience « témoin »** ou « non exposée » à qui l'on s'interdit de diffuser une publicité. La campagne opérée, on mesure la part de visiteurs parmi l'audience cible, on mesure la part de visiteurs parmi l'audience témoin, et on en déduit l'uplift.

Par exemple : on obtient un taux de visite de 3.27 % pour la population cible, et de 2.23 % pour la population témoin. Ces chiffres indiquent un delta de 1.04%, ce qui correspond à un uplift de 46.64%. Bingo, la campagne a cartonné !

Mais cette méthode est statistiquement erronée ! Le résultat n'est pas pertinent et ne permet pas en réalité d'évaluer la performance de la campagne.

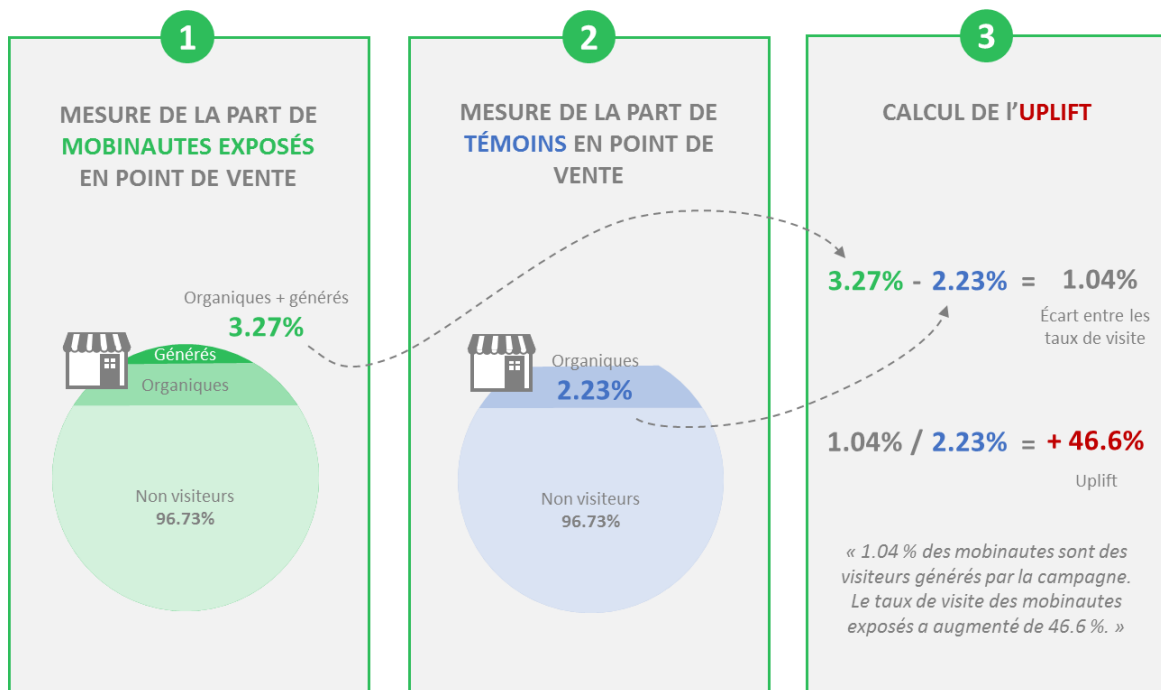


Figure 2 - Méthode de calcul de l'uplift

Voici pourquoi :

L'erreur se situe dans la constitution de la population témoin. En effet, en utilisant une population témoin, l'objectif est de comparer le comportement de la population cible à un « étalon » qui ne diffère pas de celle-ci uniquement sur le facteur « a vu/n'a pas vu la publicité ». Or un **biais statistique** est introduit par la méthodologie employée : on ne compare pas l'audience témoin à l'audience cible, mais l'audience témoin à l'audience cible **exposée**. Et cela introduit un biais très important : la part de visiteurs organiques y est naturellement beaucoup plus élevée !

Quelles sont les principales sources de biais ?

- L'audience totale **peut ne pas être accessible** : supposons par exemple que parmi les 100 000 identifiants publicitaires disponibles, 20 % d'entre eux soient des identifiants « fantômes » n'étant pas ou plus liés à un device. Le choix aléatoire des populations assure une présence de 20 % d'identifiants « fantômes » parmi les témoins. En revanche parmi la population cible exposée, il y a 0 % d'identifiants « fantômes » puisqu'ils ont tous été exposés. Il y aura donc naturellement plus de visiteurs parmi la population cible exposée que parmi la population témoin.
- Le mode de diffusion : dans le cadre de campagnes drive-to-store les impressions sont le plus souvent diffusées en **geofencing**. Dans ce cas on ne diffuse la publicité à la population cible **que s'ils entrent dans un périmètre défini autour du point de vente**. On compare donc la population témoin à une population cible, exposée, **s'étant rendu à proximité du point de vente**. Dire que les cibles se sont rendues dans le point de vente avec un uplift de 46.72 % par rapport à la moyenne est alors une erreur grossière. Imaginons que l'on compare le nombre de visiteurs d'un restaurant parisien parmi les mobinautes s'étant rendu à moins de 400 mètres de celui-ci d'un côté, et parmi l'ensemble des mobinautes parisiens de l'autre... Le taux de visite sera naturellement plus élevé parmi la première population !
- Le créneau temporel : le même phénomène se présente si la diffusion a lieu **sur un créneau temporel précis** (au moment des repas pour des publicités de restauration par exemple).
- Enfin, l'activité même des mobinautes sur leur mobile est un biais en soi. Prenons le cas d'une chaîne de supermarchés alimentaires. La direction marketing décide de diffuser sa publicité sur les utilisateurs de sa propre application (et uniquement celle-ci). En appliquant le protocole ci-dessus, la population cible exposée est constituée **uniquement de mobinautes ayant ouvert l'application pendant la période de campagne** alors que la population témoin comprend alors des utilisateurs de l'application l'ayant ouverte pendant la période de campagne mais surtout une grande majorité ne l'ayant pas ouverte.

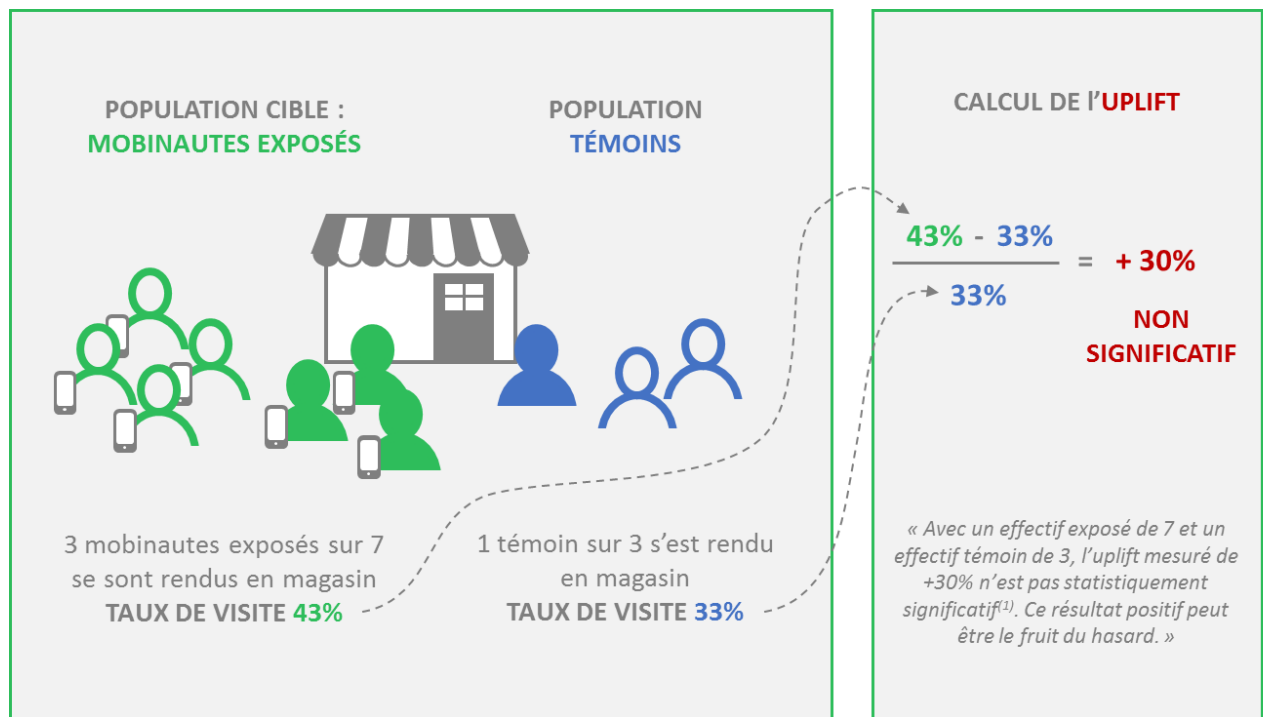
La conséquence de ces biais est que les acteurs n'annoncent pas d'uplifts nuls ou négatifs, que la campagne est effectivement bien fonctionnée ou non.

Enfin, la **significativité des résultats est trop souvent négligée**. Bien souvent, les campagnes mobiles drive-to-store ne présentent pas de volumes suffisamment élevés pour obtenir un résultat significatif. Les taux de visites des différents ciblage sont généralement faibles (moins de 1%), les volumes de diffusion doivent donc être d'autant plus gros que l'écart de taux de visite est faible.

Prenons une campagne de 500 000 impressions, avec un capping de 2 : on touche donc 250 000 mobinautes qui constituent la population cible exposée. La population témoin est constituée de 150 000 identifiants. Les taux de visite sont respectivement de 0.93% et 0.87%, soit un uplift de 7%. Au vu des volumes, ce résultat n'est pas significatif¹ : la différence de taux de visite peut complètement **être le fruit du hasard**, et l'on ne peut pas raisonnablement conclure que la population cible exposée s'est plus rendue en point de vente que la population témoin. Dans cet exemple précis, augmenter la

¹ La significativité est ici calculée avec le test de Student à 95%.

population témoin à 200 000 identifiants (tous paramètres égaux par ailleurs) aurait rendu le résultat significatif.



(1) Statistiquement significatif = satisfait le test de Student à 95%.

Figure 3 - Exemple d'uplift non significatif

Alors, comment calculer un uplift statistiquement significatif ?

Pour éviter le maximum de biais, Retency propose de cibler une population avec des caractéristiques rigoureusement identiques à la population cible exposée. Cette population témoin doit donc être exposée **exactement de la même façon** (c'est-à-dire avec les mêmes règles de diffusion) que la population cible, avec une publicité la plus neutre possible. Ainsi en comparant la population cible exposée à la population témoin exposée, on s'assure d'annuler tous les biais possibles.

Et pour éviter le risque lié à la significativité du résultat, nous estimons en amont les volumes nécessaires pour la mesure d'un uplift en fonction de la précision de la mesure souhaitée. Ainsi la taille de la population témoin exposée peut atteindre jusqu'à 40% de la taille de la population des mobinautes touchés.

En conclusion : la mesure et le calcul d'un uplift pour une campagne mobile drive-to-store est un procédé complexe, qui nécessite de prendre un certain nombre de précautions en amont de la campagne pour garantir sa pertinence.

Attention à ne pas se laisser confondre par un beau chiffre bien présenté.

Définitions :

Un **biais statistique** apparaît lorsque l'on cherche à comparer deux éléments (ici des populations) qui ne sont pas comparables. Imaginons que je veuille savoir si les employés d'une entreprise apprécient manger à la cantine proposée par celle-ci. Je me rends dans la cantine et interroge les employés présents. J'obtiens 67% d'avis positifs, 28% d'avis négatifs et 5% d'avis neutres. Puis-je conclure que 67% des employés de l'entreprise apprécient la cantine ? La réponse est non : en procédant ainsi, j'ignore complètement les employés qui ne mangent pas à la cantine (peut-être justement parce qu'ils n'apprécient pas y manger...). J'ai introduit un biais statistique : la présence à la cantine de mes interrogés. La seule conclusion que je puisse tirer est donc : 67% des employés de l'entreprise **qui mangent à la cantine** apprécient y manger.

La **significativité statistique** est l'évaluation d'un résultat statistique indiquant si celui-ci reflète ou non la réalité. Je m'intéresse à la taille des garçons de café : j'en mesure 100 et j'obtiens une taille moyenne de 1.754 m. La taille moyenne des hommes français est de 1.774 m. Puis-je extrapoler et conclure que les garçons de café sont plus petits que la moyenne des français ? La réponse est de nouveau négative : en interrogeant « seulement » 100 garçons de café, je ne peux pas dire que l'écart de 2.0 cm soit significatif¹. Pour obtenir un résultat significatif, il aurait fallu soit que l'écart entre mon échantillon de garçons de café et la moyenne nationale soit plus grand (3 cm par exemple, c'est-à-dire des garçons de café de 1.744 cm), soit que mon échantillon soit plus grand.

Equipe Data Retency