JEUX DE DONNEES

# Choix du Jeu de données

Nous avons choisi des données concernant un site e-commerce dans la vente de bijoux et accessoires. On cherchera à prouver que l’étude des statistiques d’un site est un enjeu majeur pour la prise de décisions stratégique pour l’entreprise.

On émettra ainsi la problématique suivante :

*Comment les choix stratégiques d’un pôle d’analyse de comportement utilisateur d’un site e-commerce, révèle les dysfonctionnements du système d’information et détermine les choix stratégiques de l’orientation de ce dernier ?*

Afin de répondre à cette problématique, nous allons analyser l’utilisation du site par un échantillon de 1000 d’utilisateurs .

On trouvera ainsi les variables suivantes :

- 'id' : l'identifiant de l'utilisateur est une variable quantitative discrète permettant d'identifier l'utilisateur.

- 'sexe' : le sexe de l'utilisateur est une variable qualitative nominale permettant d'identifier le sexe de l'utilisateur. Cela nous permettra d'orienter nos choix stratégiques en fonction de la clientèle.

- 'device' : le device de l'utilisateur est une variable qualitative nominale permettant d'identifier par quel appareil l'utilisateur se connecte au site. Cela nous permettra d'orienter nos choix pour améliorer les devices les plus utilisés.

- 'pays' : le pays est une variable qualitative nominale permettant d'identifier le pays dans lequel l'utilisateur se connecte.

- 'age' : l'âge de l'utilisateur est une variable quantitative discrète permettant d'identifier l'âge de l'utilisateur.

- 'temps\_site' : la variable temps\_site est une variable quantitative continue permettant d'identifier le temps de session en seconde de l’utilisateur sur les pages du site. Cela nous donne une indication sur le fait que le site est attractif/séduisant ou non.

- 'nb\_page' : la variable nb\_page est une variable quantitative continue permettant d'identifier le nombre de page que l'utilisateur a visité sur le site. Cela nous donne une indication sur le fait que le site est attractif/séduisant ou non.

- 'moyenne\_chargement’ : la variable temps\_chargement\_page est une variable quantitative continue permettant d'identifier le temps moyen de chargement d'une page en seconde. Cela nous donne une indication sur les performances du site.

# Déroulement de l'analyse

L’analyse est découpée en deux sections :

* Une succession d’analyses univariées pour calculer en autre le nombre de page par utilisateur, le temps moyen d’une session utilisateur etc…
* Une succession d’analyses bivariées pour déterminer l’existence de corrélations pour prouver l’attractivité du site ou déceler les problèmes de performance.
* Une conclusion apportant une réponse à notre problématique

Pour cela, nous allons étudier les corrélations entre le temps passé sur le site et le nombre de page, le device de l'utilisateur et le temps de chargement du site et enfin si le fait d'être une femme ou un homme impact le temps de session.

# Jeu de données

*Tableau de jeux de données*

ANALYSE UNIVARIEES

# Genre

Nous pouvons voir grâce à ce diagramme que nous avons une majorité de femme qui viennent sur le site par rapport aux hommes.

Cela nous permettra de prendre des décisions stratégiques comme par exemple, continuer à attirer un public féminin ou bien au contraire, faire des changements sur le site pour attirer plus d'hommes.

# Age

Nous remarquons ainsi que le principal public touché sont des utilisateurs âgés de 15 à 45 ans représentant 907 personne avec une moyenne d'âge de 35 ans.

Cependant, ce sont les 22-23 ans qui sont le plus touché par le design du site avec plus de 75 utilisateurs sur cette tranche.

Par conséquent, nous devons être en mesure d'adapté le contenu du site pour cette tranche d'âge ciblé.

Nous pouvons également faire le choix de changer de stratégie et ainsi adapté le site pour un tout autre public.

# Devices

Nous pouvons voir grâce à ce schéma que plus de la moitié des équipements utilisés sont des terminaux mobiles, suivi de poste fixe et enfin de tablette tactile.

Nous savons donc qu'il faut axer nos stratégies d'évolution sur le site en fonction de l’utilisation des smartphones dans notre société.

# Chargement

On observe que :

* Le temps maximum enregistré d’une session active sur le site est de 2500 secondes soit environ 40 minutes.
* Le temps minimum est de 3 secondes.
* Le temps moyen est de 777 secondes soit environ 12 minutes.

On observe notamment que 50% des utilisateurs ont une session supérieure à 500 secondes (8 min et 30 secondes)

# Pages visités

On observe que :

* 50 % des utilisateurs enregistrés visite plus de 6 pages avec un maximum à 12.

Il serait intéressant d’observer ces pages afin de comprendre pourquoi le taux de rétention est plus élevé.

Les premières hypothèses peuvent-être :

* Ce sont des pages d’achat. Ainsi, l’utilisateur est obligé de passer par ces pages.
* Ce sont des pages « catalogues », cela indiquerait que l’utilisateur apprécie l’action de « cataloguer » sur le site.
* Ce sont des pages d’information (évènement, promotion etc…). Et alors la fonctionnalité primaire du site n’est pas la vente en e-commerce mais la communication externe.

ANALYSE BIVARIEES

# Le temps utilisateur

## NB PAGE visitées

On observe à travers cette droite une corrélation positive (0,8) entre le temps de session et le nombre de page visité.

Cela peut être interprété en plusieurs points :

* Les pages ne sont pas assez attractives car leur taux de rétention est trop faible et ainsi l’utilisateur change de page plus souvent.
* Le catalogue produit est bien trop grand et aucun produit phare n’attire principalement les utilisateurs.
* Le site est très attractif et capturant, pour étayer cette hypothèse il faudrait comparer cette donnée avec le panier de l’utilisateur et la somme dépensée.

## Moyenne de chargement

On détermine qu’il n’existe aucune corrélation significative (-0.2) entre le temps d’une session et le temps de chargement moyen entre chaque page.

On en déduit un problème de performances/d’architecture car l’utilisateur à toujours le même temps de chargement quel que soit son temps de session active.

Or, dans l’idéal, il faudrait obtenir une corrélation négative et tendre à augmenter cette corrélation.

En effet, l’utilisateur ne devrait pas recharger les éléments du site a chaque page (comme le *header*, le logo, le *footer* etc…) ou des informations clients cotés *back-office* (identifiant du visiteur, panier etc…·).

On pourrait résoudre cela par la mise en place d’un cache utilisateur et la réduction d’appel serveur.

# L’utilisateur

## Sexe

Par ce diagramme on observe que le temps d’une session est plus importante lorsque l’utilisateur est une femme.

Deux axes stratégiques peuvent ainsi se dégager :

* Privilégier les utilisateurs « Femme » car il y a plus de chance qu’elle achète un produit.
* Privilégier les utilisateurs « Homme » afin d’accéder à une nouvelle clientèle et d’influencer la prise de décision d’achat.

## Device

On observe que le device ayant le plus de problèmes de performance est l’équipement mobiles/smartphones.

On détecte ainsi un réel problème de performance sur cet équipement. Il faudrait réorienter le travail de l’équipe IT sur cette plateforme.

Conclusion :

Grâce à ces analyses ont en déduit deux points importants :

* Le site actuel possède un design féminin sur la tranche d’âge 20-30 ans.
* Un réel problème de performance existe sur le temps de chargement entre les pages et notamment sur les terminaux mobiles.

Or,

* L’utilisation du téléphone mobile est le premier équipement enregistré. Il est ainsi primordial d’améliorer l’expérience utilisateur sur cet équipement afin d’améliorer le taux de rétention sur les pages.
* Le site soit s’orienter sur une stratégie d’accentuation sur le publique féminin car elle représente plus de 60% de ses utilisateurs.
* Le site doit privilégier la catégorie d’âge 20-30 pour maximiser son utilisation.