

Analyse des données de ShopSmart

Etape 1 : Etude du besoin

Avant de commencer de notre analyse, nous devons tout d'abord étudier les besoins de notre cible et comprendre quels sont les objectifs attendus et pouvoir discerner les éléments qui vont nous aider à répondre aux besoins . Ainsi, il est primordial de se poser les questions suivantes :

- Quel est le besoin ?
- Pourquoi je fais cette analyse ?
- Comment cela va aider l'entreprise ?
- Est-ce que ce je fais répond précisément aux attentes de l'entreprise ?
- Comment je peux interpréter les insights
- Le KPI là répond à quel besoin ?
- Quels indicateurs peuvent vous aider à analyser vos performances en la matière

De ce fait, en renvoyant la situation dans notre contexte nous pouvons constater que les questions à se poser ici sont les suivantes :

- Qu'est ce qu'on entend par comportement d'achat ?
- Qu'est cela implique ?
- Que faut-il observer pour proposer de l'amélioration du marketing ?
- Comment les clients interagissent-ils avec les produits ?
- Comment augmenter la productivité ?
- Comment savoir si les efforts en marketing sont efficaces ?

Etape 2 : Prise de de connaissance du dataset et étude

Il est important d'effectuer une lecture et une exploration globale des données que nous avons sous la main car cela nous permet de comprendre les différentes informations qui sont à notre portée et observer les relations entre elles car en fonction de cela nous pouvons effectuer le nettoyage et le traitement de la base de données . Dans notre cas nous avons pu observer que les différentes colonnes nous proposaient ceci :

- **InvoiceNo** : le numéro de facture : chaque transaction (achat) est associée à un numéro de facture unique.
- **StockCode** : le code produit : l'identifiant unique attribué à chaque produit vendu par ShopSmart.
- **Description** : la description du produit vendu : cela donne des détails sur le produit acheté.
- **Quantity** : la quantité achetée de chaque produit pour cette transaction spécifique.

- **InvoiceDate** : la date et heure de la transaction : ceci indique quand la vente a été enregistrée
- **UnitPrice** : le prix unitaire du produit : c'est le prix d'une seule unité du produit vendu
- **CustomerID** : l'identifiant du client : chaque client est identifié par un numéro unique
- **Country** : le pays où la transaction a eu lieu

Etape 3 : Traitement du dataset

Ici nous avons d'abord transformer notre jeu de données (le jeu de données a été converti en dataframe dans notre code afin d'être facilement exploitable). Nous avons vérifié les valeurs aberrantes, manquantes dans notre dataframe et nous en avons trouvé au niveau de la colonne « description » et « customerId » donc nous avons décidé de supprimer les valeurs nulles sauf dans « customerId » car elle en remplaçant les valeurs manquantes par une catégorie comme "Inconnu". Cela nous permet de conserver les transactions sans identifiant client tout en les distinguant.

De plus, nous avons vérifié la redondance et nous avons trouvé qu'elle était normale car nous avons vu qu'ici les doublons avaient une valeur significatives notamment avec :

- **Les transactions multiples** : Il est fréquent d'avoir plusieurs lignes dans un jeu de données avec le même InvoiceNo, chaque ligne représentant un article différent acheté dans la même transaction. Donc, plusieurs lignes avec le même InvoiceNo sont normales car elles représentent des items distincts achetés simultanément.
- **Un produit acheté par différents clients** : Un même StockCode peut être acheté par plusieurs clients différents. Donc, plusieurs lignes avec le même StockCode mais des CustomerID différents peuvent également être normales
- **Le partage de factures, commandes** : Dans certaines situations, une même facture peut être partagée entre plusieurs clients (CustomerID)

Etape 4 : Explication de l'analyse exploratoire

I