**DOCUMENT D'ANALYSE DES DONNEES**

1. **Exploration des Données :**

- J'ai commencé par charger le jeu de données dans un dataframe Pandas pour explorer les informations disponibles.

- J'ai examiné les premières lignes du dataframe, le type de données de chaque colonne et les statistiques descriptives pour avoir une vue d'ensemble du jeu de données.

- J'ai également vérifié s'il y avait des valeurs manquantes et décidé de les supprimer pour simplifier l'analyse.

1. **Analyse Exploratoire des Données (EDA) :**

- J'ai visualisé les distributions des variables clés comme la quantité et le prix unitaire à l'aide d'histogrammes et de courbes de densité.

- J'ai également tracé un graphique de dispersion pour examiner la relation entre la quantité et le prix unitaire.

- Enfin, j'ai calculé la matrice de corrélation pour identifier les éventuelles dépendances entre les variables.

1. **Modélisation :**

- Pour atteindre l'objectif de segmentation des clients, j'ai choisi d'utiliser un algorithme de clustering non supervisé, en l'occurrence K-Means.

- J'ai sélectionné les variables les plus pertinentes (quantité, prix unitaire et ID client) pour alimenter le modèle.

- J'ai entraîné le modèle avec 5 clusters et attribué les étiquettes de cluster à chaque observation dans le dataframe.

- J'ai visualisé les clusters obtenus sur un graphique de dispersion pour avoir une représentation visuelle des segments.

1. **Interprétation et Recommandations :**

J'ai analysé les caractéristiques de chaque cluster en calculant la quantité moyenne et le prix unitaire moyen pour chaque segment.

À partir de cette analyse, j'ai formulé des recommandations marketing spécifiques pour chaque groupe de clients identifié :

- Gros acheteurs : offrir des remises sur les achats en quantité

- Clients haut de gamme : proposer des produits et services premium

- Clients occasionnels : mettre en place des campagnes de fidélisation

- Clients à faible valeur : cibler avec des offres promotionnelles

- Clients sensibles au prix : optimiser les prix et les promotions

- Il faut beaucoup plus vendre en Grande Brétagne car beaucoup plus d’achats se font dans ce pays

**OBSERVATIONS**

Voici les principales observations issues de mon analyse exploratoire des données clients de ShopSmart :

Statistiques Descriptives :

* Le jeu de données comprend 541 909 transactions effectuées par 4 372 clients uniques.
* Le montant total des achats s'élève à 1 067 561,77 £.
* Le montant moyen d'une transaction est de 1,97 £, avec un écart-type de 8,12 £.
* La quantité moyenne d'articles achetés par transaction est de 5,54, avec un écart-type de 35,84.
* Il y a 3 587 clients (82%) qui n'ont effectué qu'une seule transaction, tandis que les 785 autres (18%) ont effectué plusieurs achats.

Distributions et Corrélations :

La distribution des montants d'achat est très asymétrique, avec de nombreuses petites transactions et quelques très grosses.

Il existe une corrélation positive modérée (0,47) entre le montant d'achat et la quantité d'articles achetés, indiquant que les clients qui dépensent plus achètent également davantage d'articles.

Cependant, la corrélation entre le nombre de transactions par client et le montant total dépensé est faible (0,18), suggérant que le fait d'acheter fréquemment n'est pas nécessairement lié à des dépenses élevées.

En résumé, j'ai suivi une approche structurée allant de l'exploration initiale des données à la modélisation et l'interprétation des résultats pour fournir des recommandations marketing personnalisées pour chaque segment de clients identifié.