

Compte Rendu d' Analyse de Données Marketing Client

Date: 10 Décembre 2025

1. Introduction et Contexte de l' Analyse

Ce rapport synthétise l' analyse exploratoire et la préparation des données effectuées sur un jeu de données de marketing client. L' objectif principal de cette démarche est de transformer les données brutes en un ensemble de données **robuste et enrichi**, prêt pour une modélisation avancée, notamment la **segmentation client**. Une segmentation efficace est essentielle pour optimiser les futures campagnes marketing en ciblant des groupes de clients aux profils et comportements d' achat distincts.

Le jeu de données initial comprend des informations détaillées sur les clients, couvrant leurs caractéristiques démographiques (âge, éducation, statut marital, revenu), leur historique de dépenses par catégorie de produit, et leur engagement dans les campagnes promotionnelles passées.

2. Préparation et Assainissement des Données

La qualité des données est la fondation de toute analyse fiable. Les étapes suivantes ont été mises en œuvre pour nettoyer et enrichir l' ensemble de données.

2.1. Nettoyage et Gestion des Anomalies

Le nettoyage a ciblé les valeurs manquantes et les anomalies statistiques:

- Revenu (Income)**: Les 24 enregistrements sans revenu renseigné ont été retirés. De plus, pour garantir la robustesse des modèles statistiques, les valeurs extrêmes (outliers) ont été traitées. Les clients avec un revenu supérieur à 120 000 \$ et ceux dont l' âge calculé dépassait 100 ans ont été exclus de l' analyse.

- **Dupliqués et Identifiants:** Aucune ligne dupliquée n’ a été trouvée. La colonne d’ identifiant unique (`id`) a été retirée.

2.2. Ingénierie des Caractéristiques Clés

Pour enrichir le jeu de données et faciliter la segmentation, plusieurs variables composites ont été créées:

Caractéristique Créeée	Définition	Pertinence pour la Segmentation
Âge	Âge du client (calculé à partir de l’ année de naissance).	Ciblage générationnel et cycle de vie.
Dépense Totale (<code>Total_Spend</code>)	Somme des dépenses dans toutes les catégories de produits.	Mesure directe de la valeur client (Customer Value).
Enfants (<code>children</code>)	Nombre total d’ enfants à la maison (enfants et adolescents).	Indicateur des besoins du foyer et du pouvoir d’ achat disponible.
Statut Marital Groupé	Simplification des statuts en <code>Relationship</code> ou <code>Single</code> .	Simplification des variables catégorielles pour la modélisation.

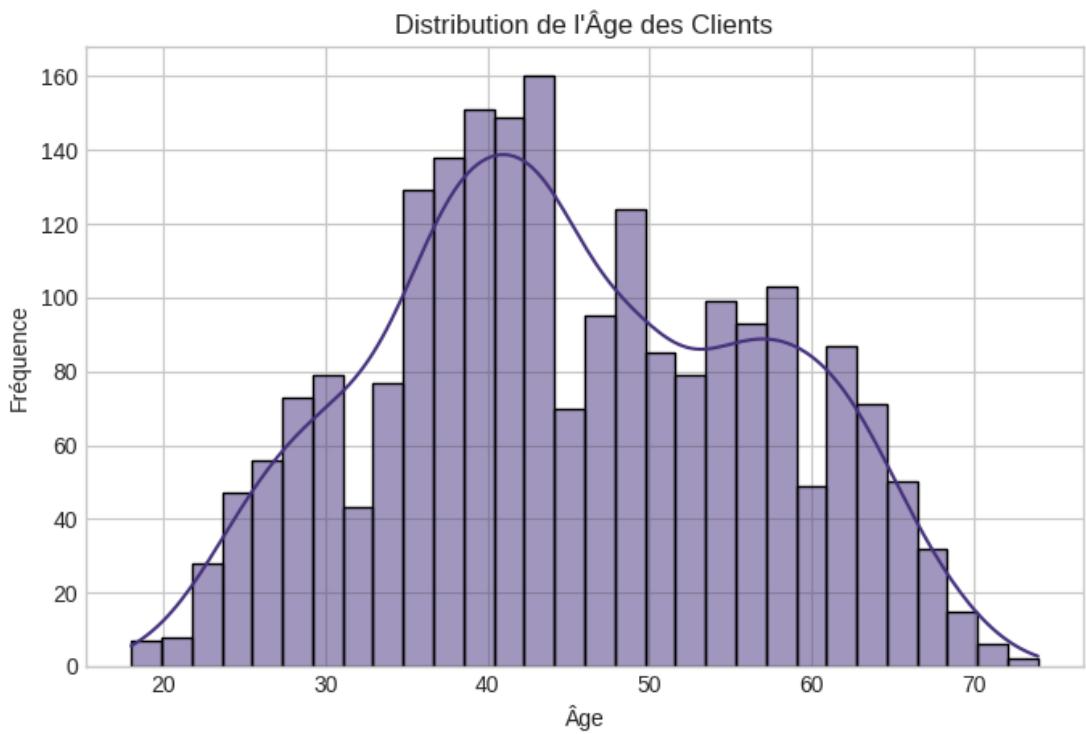
3. Analyse Exploratoire des Données (EDA) et Résultats

L’ analyse des distributions des variables clés révèle des informations cruciales sur la clientèle.

3.1. Distribution Démographique

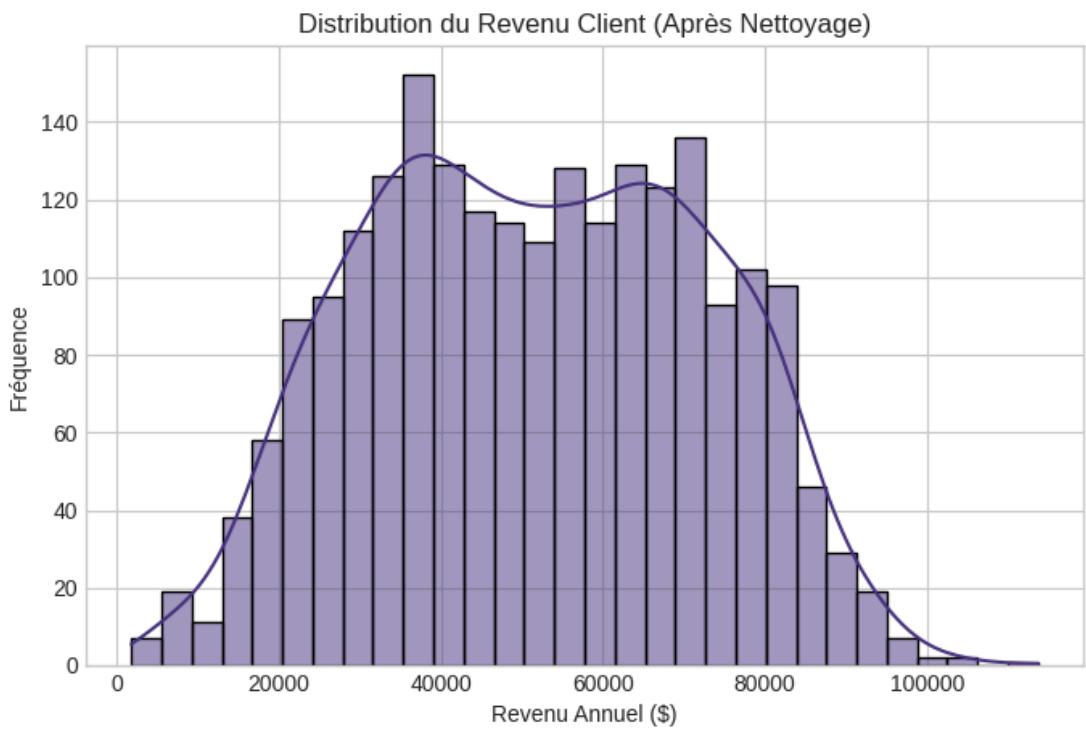
L’ analyse de l’ âge et du revenu, après le traitement des outliers, montre une clientèle concentrée dans une tranche d’ âge mature et un revenu modéré.

Distribution de l’ Âge des Clients



La majorité des clients se situent entre 40 et 60 ans, suggérant une cible potentielle de professionnels établis ou de familles matures.

Distribution du Revenu Client

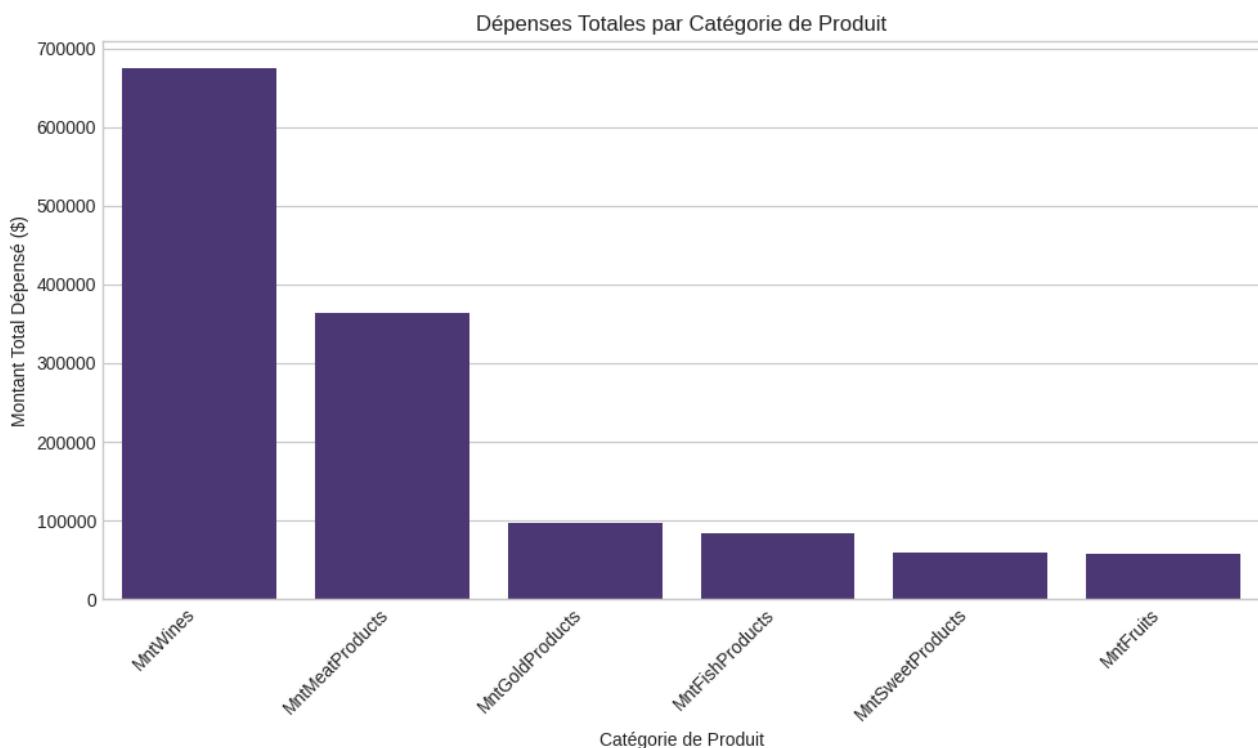


La distribution du revenu est relativement normale, avec une concentration autour de la moyenne, ce qui indique une base de clients de classe moyenne.

3.2. Comportement d' Achat et Dépenses

L' examen des dépenses par catégorie de produit est essentiel pour comprendre les préférences des clients et identifier les leviers de croissance.

Dépenses Totales par Catégorie de Produit



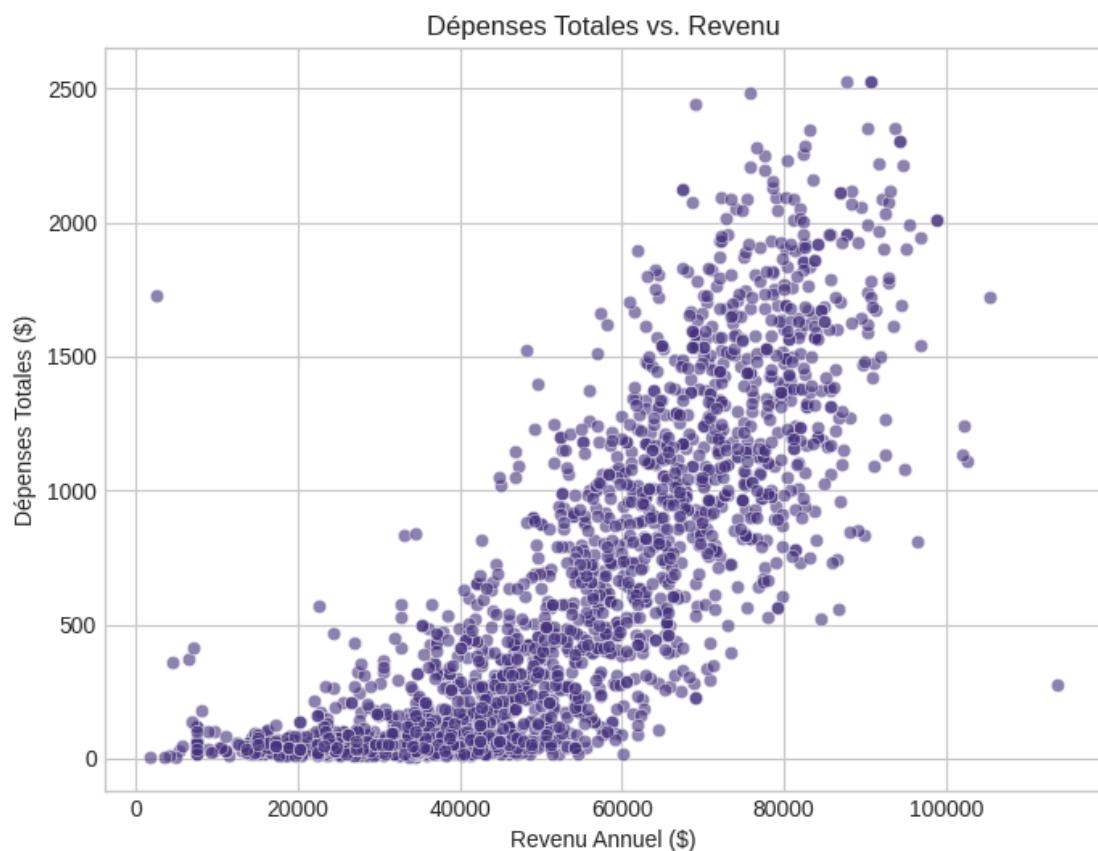
Le graphique ci-dessus met en évidence une forte disparité dans les habitudes de dépenses:

- Le **Vin (MntWines)** est de loin la catégorie de produit la plus populaire, représentant la majorité des dépenses totales.
- Les **Produits Carnés (MntMeatProducts)** arrivent en deuxième position, mais avec un écart significatif.
- Les autres catégories (Fruits, Poissons, Sucreries, Or) affichent des montants de dépenses nettement inférieurs.

3.3. Relation entre Revenu et Dépenses

L' analyse de la corrélation entre le revenu et la dépense totale permet de visualiser l' impact du pouvoir d' achat sur la valeur client.

Dépenses Totales vs. Revenu



Ce nuage de points illustre une **corrélation positive** claire entre le revenu et la dépense totale. Les clients à revenu élevé sont systématiquement ceux qui dépensent le plus. Cependant, la dispersion des points montre qu' à revenu égal, il existe une grande variabilité dans les dépenses, ce qui justifie l' utilisation d' une segmentation plus fine que le simple revenu.

4. Conclusion et Perspectives

L' analyse a permis de valider un ensemble de données nettoyé et enrichi, prêt pour la modélisation. Les conclusions tirées de l' EDA sont les suivantes:

- **Le Vin est le Moteur de Revenu:** La stratégie marketing doit continuer à capitaliser sur la catégorie Vin, tout en explorant des opportunités de vente croisée pour les autres catégories de produits.
- **Potentiel de Vente Croisée:** Les faibles dépenses dans les catégories comme les fruits et les poissons représentent un potentiel inexploité. Des offres personnalisées basées sur les segments de clients pourraient stimuler ces achats.
- **Nécessité de Segmentation:** La relation entre le revenu et les dépenses, bien que positive, n'est pas linéaire, et les profils démographiques sont variés. Une segmentation client est indispensable pour identifier les groupes de clients (par exemple, les "gros dépensiers en vin" vs. les "petits dépensiers à revenu élevé") et adapter les messages marketing.

Prochaine Étape Recommandée: Procéder à l'application d'un algorithme de clustering (tel que K-Means ou DBSCAN) sur les variables clés transformées (`Total_Spend`, `Income`, `Age`, `Recency`) pour définir les segments de clientèle et élaborer des stratégies marketing spécifiques à chaque groupe.