Une image contenant Police, texte, Graphique, graphisme

Description générée automatiquement

Commerce en Ligne

Table des matières

[1. Introduction 3](#_Toc182594775)

[2. Sources de données 3](#_Toc182594776)

[Dataset Search 3](#_Toc182594777)

[Kaggle 3](#_Toc182594778)

[3. Modèle des données 3](#_Toc182594779)

[4. Processus ETL (Extraction, Transformation, Chargement) 4](#_Toc182594780)

[5. Tableau de bord envisagé 4](#_Toc182594781)

[Perfume E-Commerce Dataset 2024 5](#_Toc182594782)

[Visualisations Actuelles : 5](#_Toc182594783)

[Visualisations Futures Proposées : 7](#_Toc182594784)

[Gaming Sales ANALYSYS 8](#_Toc182594785)

[E-commerce Customer Behavior Dataset 11](#_Toc182594786)

# 1. Introduction

Dans le domaine du commerce en ligne, l’analyse des comportements d’achat permet de mieux comprendre les habitudes des consommateurs et d’optimiser les stratégies de vente. La problématique que nous avons choisie pour ce projet d’intégration de données est la suivante :

”**Quels sont les plus grands consommateurs de produits en ligne ?**”

Pour répondre à cette question, nous allons exploiter diverses sources de données indépendantes pour obtenir une vue d’ensemble complète du profil des consommateurs et des produits les plus prisés.

# 2. Sources de données

Les données nécessaires pour cette analyse proviendront de trois sources principales :

## [Dataset Search](https://datasetsearch.research.google.com/)

Outil de recherche de jeux de données qui peut offrir une diversité de fichiers de données externes pertinents.

## [Kaggle](https://www.kaggle.com/)

Plateforme de partage de jeux de données, de laquelle nous extrairons :

* [Perfume E-Commerce Dataset 2024](https://www.kaggle.com/datasets/kanchana1990/perfume-e-commerce-dataset-2024) : Le marché des parfums en ligne est en pleine expansion. Le dataset Perfume E-Commerce Dataset 2024 est une ressource clé pour analyser les tendances de consommation, les niveaux de satisfaction client et les performances des différentes catégories de produits.
* <https://www.kaggle.com/datasets/aayushmishra1512/best-selling-video-games-in-the-world>: Avis de consommateurs d’Amazon, offrant un aperçu de la satisfaction et des produits les plus populaires. La licence : [MIT](https://www.mit.edu/~amini/LICENSE.md)
* [E-commerce Customer Behavior Dataset](https://www.kaggle.com/datasets/uom190346a/e-commerce-customer-behavior-dataset/data) : Comportement des clients du commerce électronique. La licence : [CC0: Public Domain](https://creativecommons.org/publicdomain/zero/1.0/)

# 3. Modèle des données

Pour intégrer efficacement ces différentes sources de données, un modèle de données structuré est essentiel.

* Organisation des tables par source : Chaque source de données sera représentée par une ou plusieurs tables distinctes dans le datawarehouse. Cela permet de conserver la granularité et la spécificité des informations propres à chaque plateforme (par exemple, les champs d’avis peuvent varier entre Shopee et Amazon).
* Liens entre les sources : Les tables seront reliées par des clés communes (si disponibles), telles que l’ID de produit, la catégorie de produit, ou encore les régions géographiques. Ces liens permettront de regrouper les données selon des dimensions d’analyse homogènes et pertinentes.

# 4. Processus ETL (Extraction, Transformation, Chargement)

Pour passer des sources de données brutes à un datawarehouse structuré (Système d’Information Décisionnel (SID)), un processus ETL sera mis en place.

* Granularité fine : La plupart des données issues de chaque source sont déjà détaillées, avec un niveau de précision adapté pour une analyse approfondie.
* Granularité supérieure : Les avis des clients incluront uniquement des données générales de satisfaction, ce qui limite la profondeur d’analyse mais reste pertinent pour identifier les tendances de consommation.
* Granularité instantanée périodique : Nous définirons des intervalles de mise à jour (par exemple, mensuels) pour s’assurer que le SID est alimenté avec les dernières données disponibles.

# 5. Tableau de bord envisagé

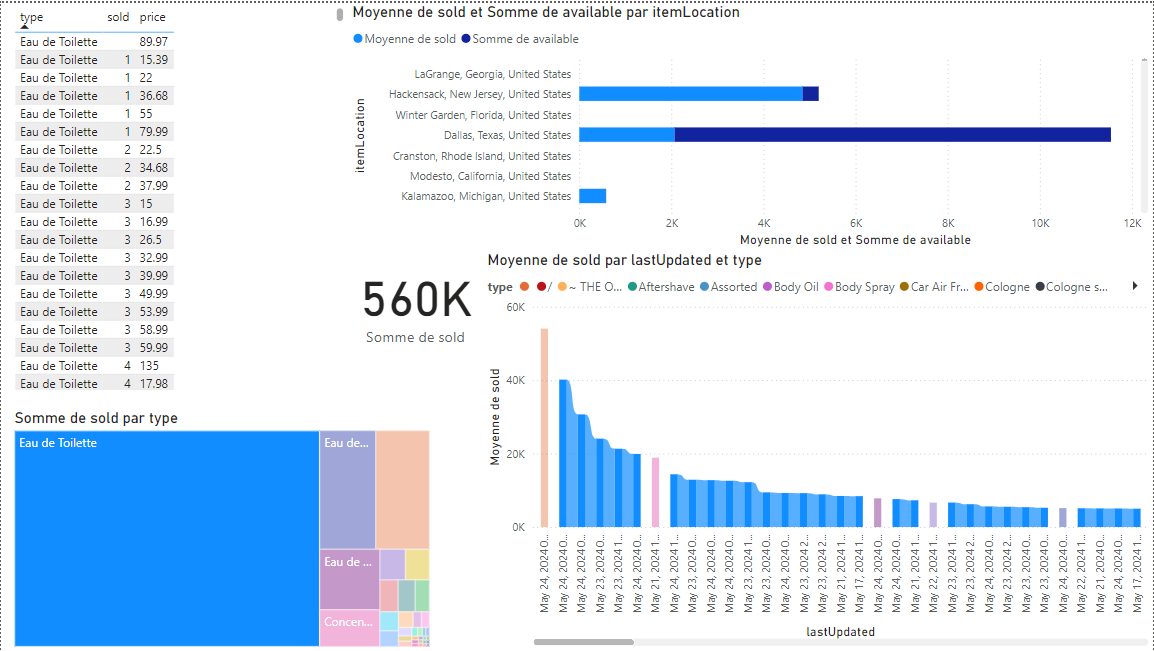
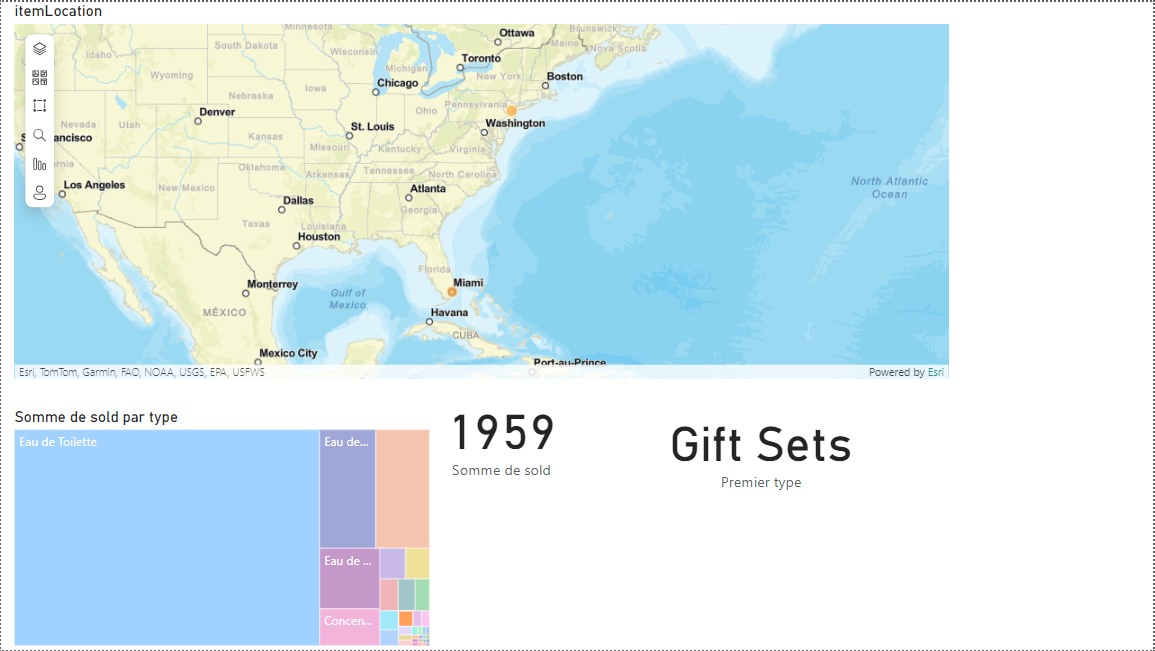
À partir des différentes bases de données centrées sur le commerce en ligne, il est possible d'analyser plusieurs valeurs numériques, telles que la somme d'argent générée par les ventes des différents produits, la qualité des produits, la quantité des ventes d’un produit particulier, ainsi qu'une étude sur l'âge des consommateurs (moyenne d'âge, répartition par tranches d’âge). Ces analyses peuvent être organisées et présentées sur différents onglets pour une vue d'ensemble claire et structurée.

# [Perfume E-Commerce Dataset 2024](https://www.kaggle.com/datasets/kanchana1990/perfume-e-commerce-dataset-2024)

À partir des données, il est possible d'analyser des valeurs clés telles que la somme d'argent générée par les ventes, la qualité des produits, et la quantité des ventes. Ces analyses peuvent être organisées sur différents onglets pour offrir une vue d'ensemble claire et structurée.

## Visualisations Actuelles :

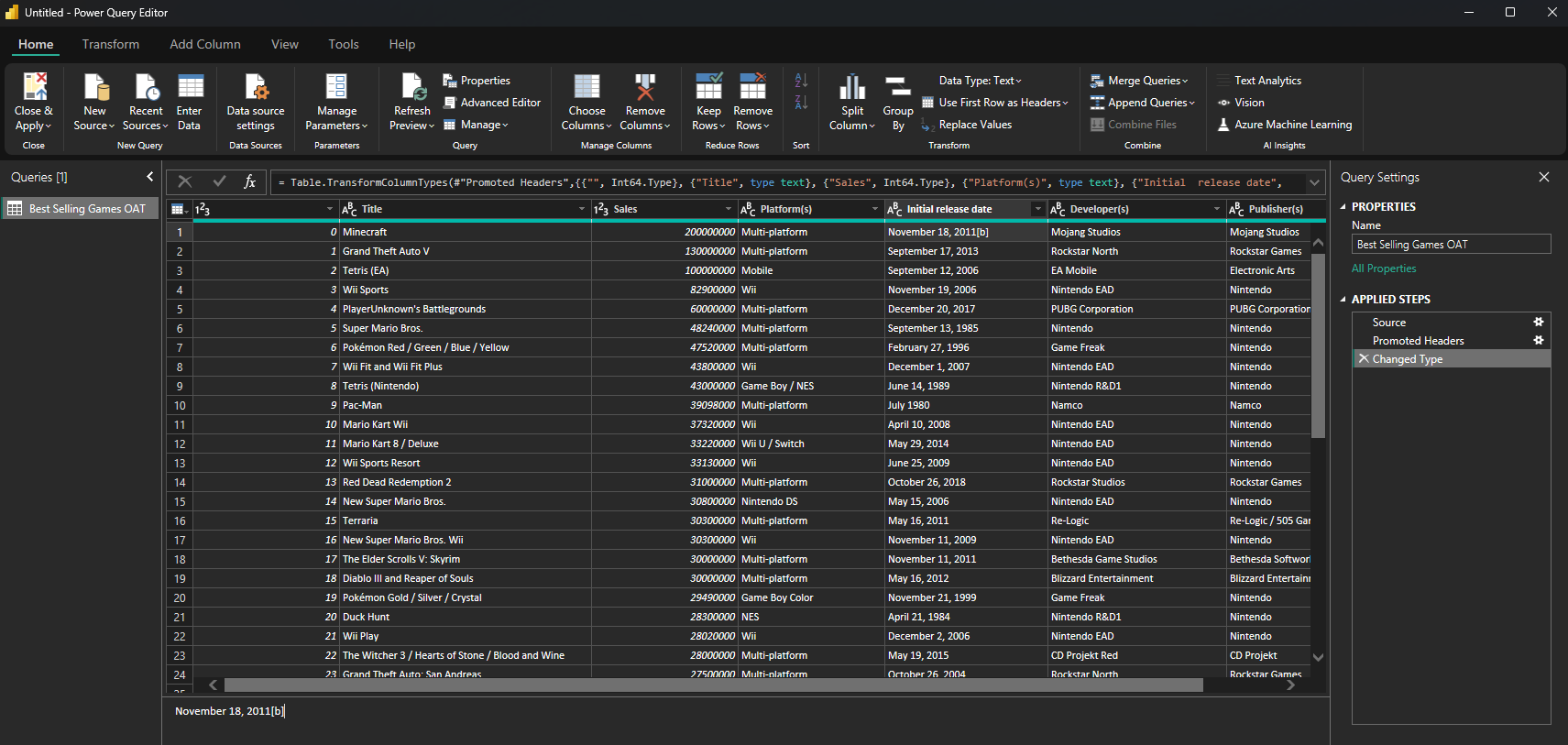
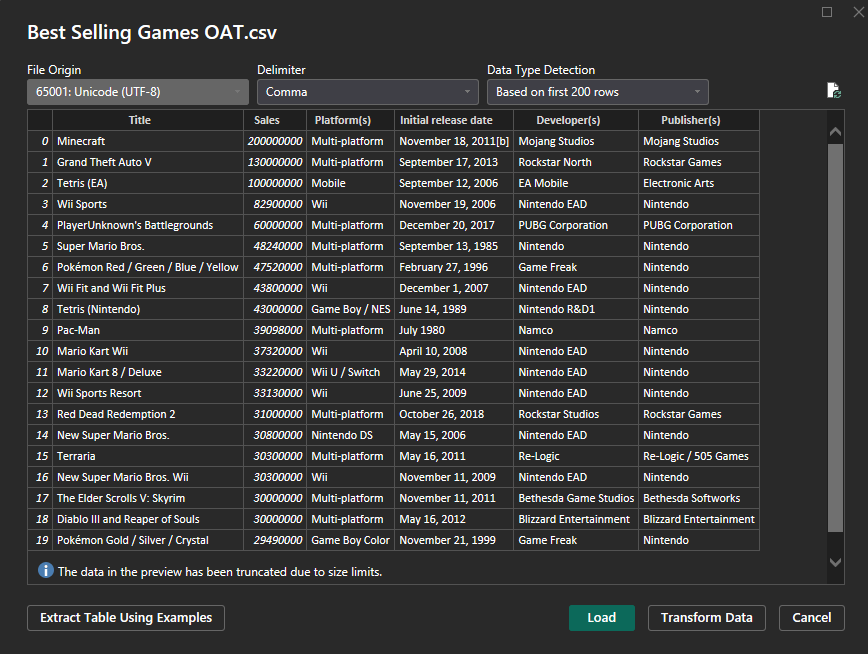
1. **Vue d’ensemble des ventes et des avis :**
   * **Graphique en barres empilées** : Classement des catégories de parfums (eau de parfum, eau de toilette, etc.) selon le volume des ventes et la note moyenne.
   * **Indicateurs principaux (KPI)** : Nombre total de ventes.
2. **Répartition géographique des consommateurs :**
   * **Carte interactive** :
     + Affichage des ventes et des notes moyennes par région géographique.
     + Taille des cercles proportionnelle au nombre de ventes, avec un code couleur pour représenter les niveaux de satisfaction.
3. **Tendances des avis dans le temps :**
   * **Graphique en courbes** : Suivi des moyennes de ventes pour identifier les tendances.

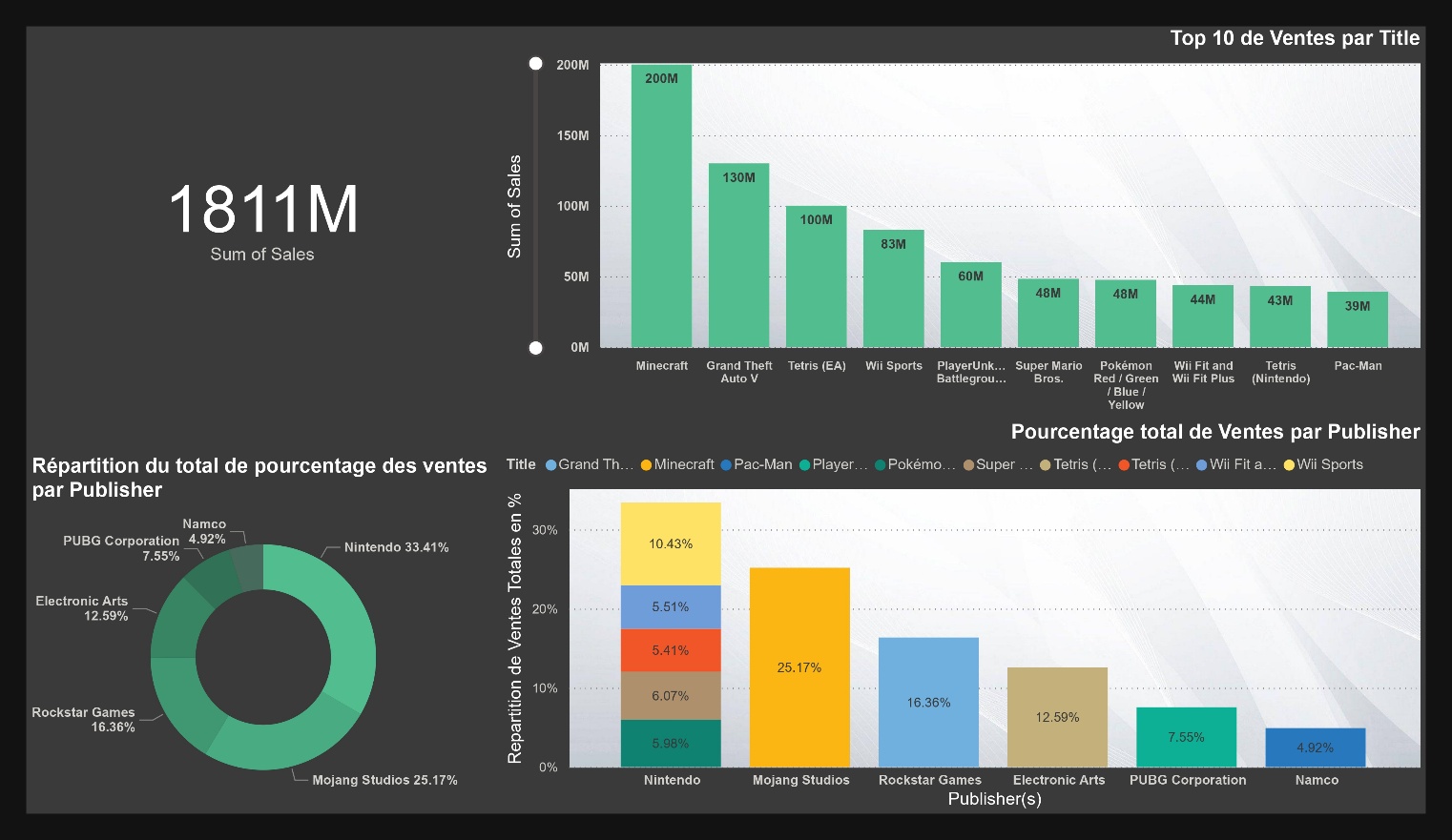


## Visualisations Futures Proposées :

1. **Analyse des segments de consommateurs :**
   * **Diagramme en radar** : Comparaison des préférences par groupe démographique (âge, genre) pour identifier des segments clés.
2. **Performance des produits par vendeur :**
   * **Heatmap (carte de chaleur)** : Croisement entre les notes moyennes et les ventes pour chaque vendeur, permettant de repérer rapidement les vendeurs performants ou sous-performants.
3. **Corrélation Prix-Satisfaction :**
   * **Graphique de dispersion** : Analyse de la relation entre le prix des produits et les notes des clients, pour explorer l'effet du coût sur la satisfaction.
4. **Analyse prédictive des tendances :**
   * **Graphique en courbes projetées** : Utilisation de modèles prédictifs pour estimer l’évolution des ventes et des notes dans les prochains mois.
5. **Taux de retour ou de fidélité des clients :**
   * **Histogramme** : Répartition des clients selon le nombre de commandes passées.
   * **Graphique circulaire** : Part des consommateurs ayant effectué plusieurs achats par rapport aux nouveaux acheteurs.

# Gaming Sales ANALYSYS

Sur les données recollectées sur Kaggle, on observe des différentes variables (Title, Sales, Platform(s), Initial release Date, Développer(s) et Publisher(s)), a l’aide de l’outil Power Query, on transforme le dataset pour pouvoir travailler avec la base de données.



Cette page de l'étude analyse le top 10 des ventes de jeux vidéo jusqu'en 2018. On y apprend que le total des ventes combinées de ces jeux dépasse 1,81 milliard de copies.

Les graphiques présentent les ventes totales des jeux vidéo les plus vendus, ainsi que la répartition des parts de marché par éditeur.

1. **Top 10 des ventes par titre :**
   1. *Minecraft* est le jeu le plus vendu avec 200 millions de ventes.
   2. *Grand Theft Auto V* suit avec 130 millions.
   3. *Tetris* (version EA) a atteint 100 millions de ventes.
   4. *Wii Sports* et *PlayerUnknown's Battlegrounds* enregistrent respectivement 83 et 60 millions de ventes.
   5. Les autres titres du top 10 incluent *Super Mario Bros.*, *Pokémon Red/Green/Blue/Yellow*, *Wii Fit and Wii Fit Plus*, *Tetris* (Nintendo) et *Pac-Man*.
2. **Répartition des ventes par éditeur :**
   1. *Nintendo* détient la plus grande part de marché avec 33,41 % des ventes totales.
   2. *Mojang Studios*, grâce à *Minecraft*, a une part significative de 25,17 %.
   3. *Rockstar Games* représente 16,36 % des ventes grâce à *Grand Theft Auto V*.
   4. *Electronic Arts* possède 12,59 %, notamment avec *Tetris (EA)*.
   5. *PUBG Corporation* et *Namco* suivent avec respectivement 7,55 % et 4,92 % des parts de marché.
3. **Observations générales :**
   1. Les jeux de *Nintendo* dominent le classement, représentant une proportion importante des ventes globales.
   2. *Minecraft*, bien qu'étant un seul jeu, représente une part très importante grâce à sa popularité mondiale.
   3. Les éditeurs de jeux ayant plusieurs titres à succès, tels que *Nintendo*, bénéficient d'une présence diversifiée sur le marché.

Cela montre que les éditeurs avec des franchises emblématiques et des titres uniques très populaires capturent une grande part du marché des ventes de jeux vidéo.

# E-commerce Customer Behavior Dataset

Voici d’abord les variables présentes dans la table du jeu de données :

Une image contenant texte, capture d’écran, Police

Description générée automatiquement

* Age : âge des clients
* Average Rating : Note moyenne attribuée par les clients
* City : Ville de résidence des clients
* Customer ID : Identifiant unique de chaque client
* Days Since Last Purchase : Nombre de jours depuis le dernier achat
* Discount Applied : Indication si un rabais a été appliqué
* Gender : Genre des clients
* Items Purchased : Nombre d'articles achetés
* Membership Type : Type d'adhésion des clients
* Satisfaction Level : Niveau de satisfaction des clients
* Total Spend : Dépenses totales réalisées par chaque client

Nous allons tout d'abord charger des données récupérées sur Kaggle et importées dans Power BI. Ensuite, nous passerons à l'étape de nettoyage des données. Après analyse, nous avons constaté que les données ne nécessitaient pas de nettoyage approfondi. Cependant, nous avons remplacé les chiffres contenant un point (.) par des virgules (,). Voici un aperçu de ce que représentent nos données :

Une image contenant texte, nombre, menu

Description générée automatiquement

Nous passons ensuite à une analyse visuelle des données client. Deux visualisations principales ont été créées. La première présente des informations essentielles sur les clients.

Une image contenant texte, diagramme, capture d’écran, carte

Description générée automatiquement

Ces visualisations nous permettent de voir la répartition des clients selon leur type d'adhésion, leur genre, et leur âge moyen, ainsi que le nombre total de clients dans notre base de données. De plus, nous avons accès à une représentation géographique des clients, qui nous montre leur répartition par localisation. Il est aussi possible de visualiser la localisation, en fonction le type d'adhésion de ainsi que la localisation des clients par genre. Grâce à ces éléments, nous pouvons obtenir une vue d'ensemble du nombre de clients en fonction de divers critères contextuels, ainsi que de leur âge moyen.

La seconde visualisation se concentre sur les dépenses des clients.

Une image contenant texte, capture d’écran, diagramme, Police

Description générée automatiquement

Elle permet d'analyser les dépenses par tranche d'âge ainsi que les dépenses moyennes des clients. On y trouve également la proportion de rabais appliqués à la population étudiée, ce qui donne un aperçu des réductions utilisées. La satisfaction des clients est aussi mise en évidence, en fonction de leur niveau d'adhésion (Bronze, Silver, Gold). Enfin, toutes ces données peuvent être filtrées par ville grâce à un filtre situé en haut à gauche, ce qui permet d'affiner l’analyse en fonction de la localisation.