

Commerce en Ligne

Table des matières

1. Introduction	3
2. Sources de données	3
Dataset Search	3
Kaggle.....	3
3. Modèle des données	3
4. Processus ETL (Extraction, Transformation, Chargement).....	4
5. Tableau de bord envisagé	4
Perfume E-Commerce Dataset 2024	5
Visualisations Actuelles :	5
Visualisations Futures Proposées :	7
Gaming Sales ANALYSYS	8
E-commerce Customer Behavior Dataset	11

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

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

Kaggle

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

- [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](#)
- [E-commerce Customer Behavior Dataset](#) : Comportement des clients du commerce électronique. La licence : [CC0: Public Domain](#)

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

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.

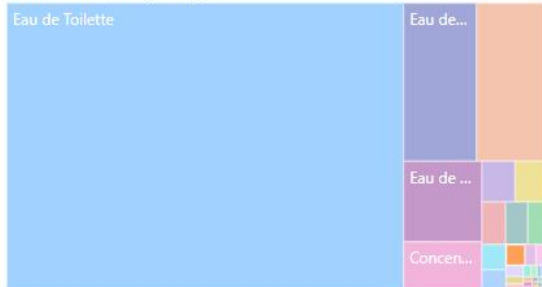
3. Tendances des avis dans le temps :

- **Graphique en courbes** : Suivi des moyennes de ventes pour identifier les tendances.

itemLocation



Somme de sold par type



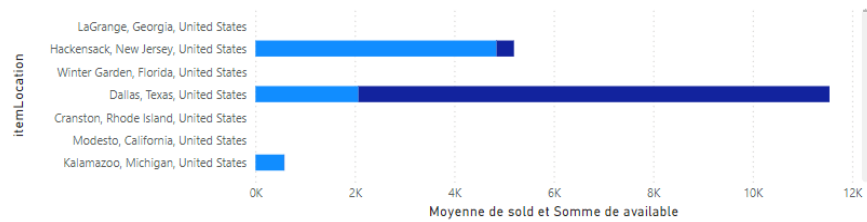
1959
Somme de sold

Gift Sets
Premier type

type	sold	price
Eau de Toilette	89.97	
Eau de Toilette	1 15.39	
Eau de Toilette	1 22	
Eau de Toilette	1 36.68	
Eau de Toilette	1 55	
Eau de Toilette	1 79.99	
Eau de Toilette	2 22.5	
Eau de Toilette	2 34.68	
Eau de Toilette	2 37.99	
Eau de Toilette	3 15	
Eau de Toilette	3 16.99	
Eau de Toilette	3 26.5	
Eau de Toilette	3 32.99	
Eau de Toilette	3 39.99	
Eau de Toilette	3 49.99	
Eau de Toilette	3 53.99	
Eau de Toilette	3 58.99	
Eau de Toilette	3 59.99	
Eau de Toilette	4 135	
Eau de Toilette	4 17.98	

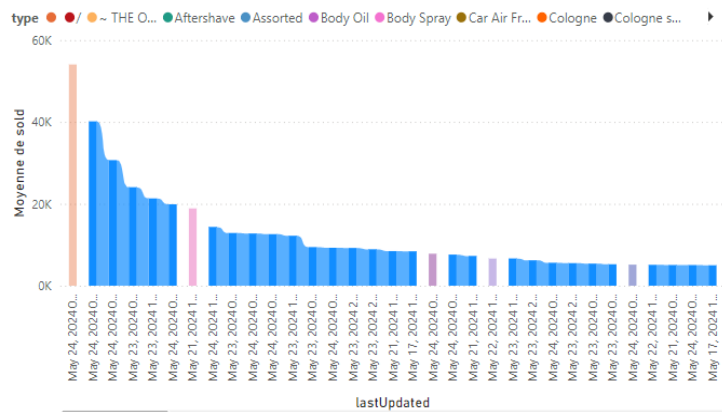
Moyenne de sold et Somme de disponible par itemLocation

● Moyenne de sold ● Somme de disponible



Moyenne de sold par lastUpdated et type

560K
Somme de sold



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

Best Selling Games OAT.csv

File Origin: 65001: Unicode (UTF-8) | Delimiter: Comma | Data Type Detection: Based on first 200 rows

	Title	Sales	Platform(s)	Initial release date	Developer(s)	Publisher(s)
0	Minecraft	200000000	Multi-platform	November 18, 2011[b]	Mojang Studios	Mojang Studios
1	Grand Theft Auto V	130000000	Multi-platform	September 17, 2013	Rockstar North	Rockstar Games
2	Tetris (EA)	100000000	Mobile	September 12, 2006	EA Mobile	Electronic Arts
3	Wii Sports	82900000	Wii	November 19, 2006	Nintendo EAD	Nintendo
4	PlayerUnknown's Battlegrounds	60000000	Multi-platform	December 20, 2017	PUBG Corporation	PUBG Corporation
5	Super Mario Bros.	48240000	Multi-platform	September 13, 1985	Nintendo	Nintendo
6	Pokémon Red / Green / Blue / Yellow	47520000	Multi-platform	February 27, 1996	Game Freak	Nintendo
7	Wii Fit and Wii Fit Plus	43800000	Wii	December 1, 2007	Nintendo EAD	Nintendo
8	Tetris (Nintendo)	43000000	Game Boy / NES	June 14, 1989	Nintendo R&D1	Nintendo
9	Pac-Man	39098000	Multi-platform	July 1980	Namco	Namco
10	Mario Kart Wii	37320000	Wii	April 10, 2008	Nintendo EAD	Nintendo
11	Mario Kart 8 / Deluxe	33220000	Wii U / Switch	May 29, 2014	Nintendo EAD	Nintendo
12	Wii Sports Resort	33130000	Wii	June 25, 2009	Nintendo EAD	Nintendo
13	Red Dead Redemption 2	31000000	Multi-platform	October 26, 2018	Rockstar Studios	Rockstar Games
14	New Super Mario Bros.	30800000	Nintendo DS	May 15, 2006	Nintendo EAD	Nintendo
15	Terraria	30300000	Multi-platform	May 16, 2011	Re-Logic	Re-Logic / 505 Games
16	New Super Mario Bros. Wii	30300000	Wii	November 11, 2009	Nintendo EAD	Nintendo
17	The Elder Scrolls V: Skyrim	30000000	Multi-platform	November 11, 2011	Bethesda Game Studios	Bethesda Softworks
18	Diablo III and Reaper of Souls	30000000	Multi-platform	May 16, 2012	Blizzard Entertainment	Blizzard Entertainment
19	Pokémon Gold / Silver / Crystal	29490000	Game Boy Color	November 21, 1999	Game Freak	Nintendo

The data in the preview has been truncated due to size limits.

Extract Table Using Examples | Load | Transform Data | Cancel

Untitled - Power Query Editor

Home | Transform | Add Column | View | Tools | Help

Close & Apply | New Source | Recent Sources | Enter Data | Data source settings | Manage Parameters | Refresh Preview | Advanced Editor | Choose Columns | Remove Columns | Keep Rows | Remove Rows | Split Column | Group By | Data Type: Text | Use First Row as Headers | Replace Values | Merge Queries | Append Queries | Text Analytics | Vision | Azure Machine Learning | AI Insights

Queries [1] | Table.TransformColumnTypes(#Promoted Headers,({{"Title", Int64.Type}, {"Sales", Int64.Type}, {"Platform(s)", type text}, {"Initial release date", type text}, {"Developer(s)", type text}, {"Publisher(s)", type text}}))

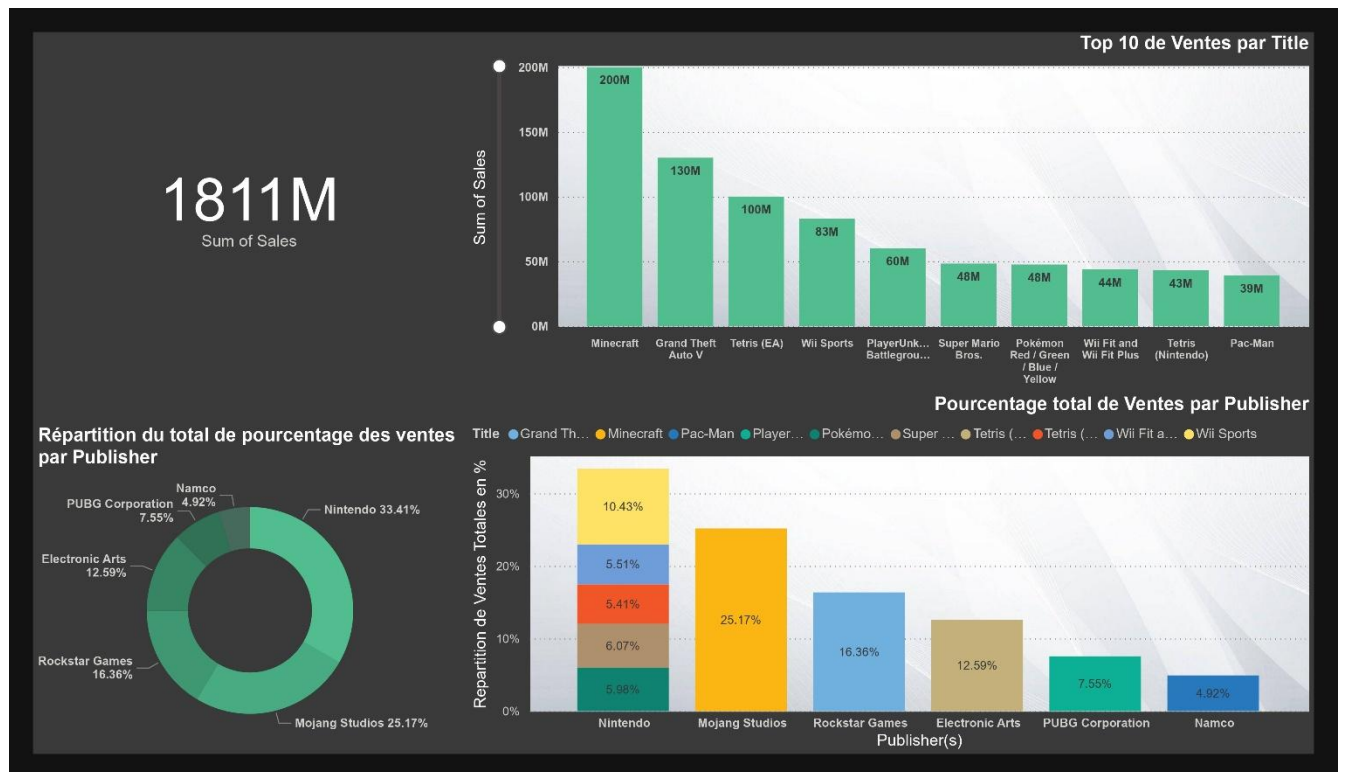
	Title	Sales	Platform(s)	Initial release date	Developer(s)	Publisher(s)
1	0 Minecraft	200000000	Multi-platform	November 18, 2011[b]	Mojang Studios	Mojang Studios
2	1 Grand Theft Auto V	130000000	Multi-platform	September 17, 2013	Rockstar North	Rockstar Games
3	2 Tetris (EA)	100000000	Mobile	September 12, 2006	EA Mobile	Electronic Arts
4	3 Wii Sports	82900000	Wii	November 19, 2006	Nintendo EAD	Nintendo
5	4 PlayerUnknown's Battlegrounds	60000000	Multi-platform	December 20, 2017	PUBG Corporation	PUBG Corporation
6	5 Super Mario Bros.	48240000	Multi-platform	September 13, 1985	Nintendo	Nintendo
7	6 Pokémon Red / Green / Blue / Yellow	47520000	Multi-platform	February 27, 1996	Game Freak	Nintendo
8	7 Wii Fit and Wii Fit Plus	43800000	Wii	December 1, 2007	Nintendo EAD	Nintendo
9	8 Tetris (Nintendo)	43000000	Game Boy / NES	June 14, 1989	Nintendo R&D1	Nintendo
10	9 Pac-Man	39098000	Multi-platform	July 1980	Namco	Namco
11	10 Mario Kart Wii	37320000	Wii	April 10, 2008	Nintendo EAD	Nintendo
12	11 Mario Kart 8 / Deluxe	33220000	Wii U / Switch	May 29, 2014	Nintendo EAD	Nintendo
13	12 Wii Sports Resort	33130000	Wii	June 25, 2009	Nintendo EAD	Nintendo
14	13 Red Dead Redemption 2	31000000	Multi-platform	October 26, 2018	Rockstar Studios	Rockstar Games
15	14 New Super Mario Bros.	30800000	Nintendo DS	May 15, 2006	Nintendo EAD	Nintendo
16	15 Terraria	30300000	Multi-platform	May 16, 2011	Re-Logic	Re-Logic / 505 Games
17	16 New Super Mario Bros. Wii	30300000	Wii	November 11, 2009	Nintendo EAD	Nintendo
18	17 The Elder Scrolls V: Skyrim	30000000	Multi-platform	November 11, 2011	Bethesda Game Studios	Bethesda Softworks
19	18 Diablo III and Reaper of Souls	30000000	Multi-platform	May 16, 2012	Blizzard Entertainment	Blizzard Entertainment
20	19 Pokémon Gold / Silver / Crystal	29490000	Game Boy Color	November 21, 1999	Game Freak	Nintendo
21	20 Duck Hunt	28000000	NES	April 21, 1984	Nintendo R&D1	Nintendo
22	21 Wii Play	28000000	Wii	December 2, 2006	Nintendo EAD	Nintendo
23	22 The Witcher 3 / Hearts of Stone / Blood and Wine	28000000	Multi-platform	May 19, 2015	CD Projekt Red	CD Projekt
24	23 Grand Theft Auto: San Andreas	27500000	Multi-platform	October 26, 2004	Rockstar North	Rockstar Games

November 18, 2011[b]

Query Settings: Name: Best Selling Games OAT | Source: Promoted Headers | Applied Steps: Changed Type

Sur les données recollectées sur Kaggle, on observe des différentes variables (Title, Sales,

Platform(s), Initial release Date, Développeur(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 :

- Minecraft* est le jeu le plus vendu avec 200 millions de ventes.
- Grand Theft Auto V* suit avec 130 millions.
- Tetris* (version EA) a atteint 100 millions de ventes.
- Wii Sports* et *PlayerUnknown's Battlegrounds* enregistrent respectivement 83 et 60 millions de ventes.
- 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 :

- Nintendo* détient la plus grande part de marché avec 33,41 % des ventes totales.
- Mojang Studios*, grâce à *Minecraft*, a une part significative de 25,17 %.
- Rockstar Games* représente 16,36 % des ventes grâce à *Grand Theft Auto V*.

- d. *Electronic Arts* possède 12,59 %, notamment avec *Tetris (EA)*.
- e. *PUBG Corporation* et *Namco* suivent avec respectivement 7,55 % et 4,92 % des parts de marché.

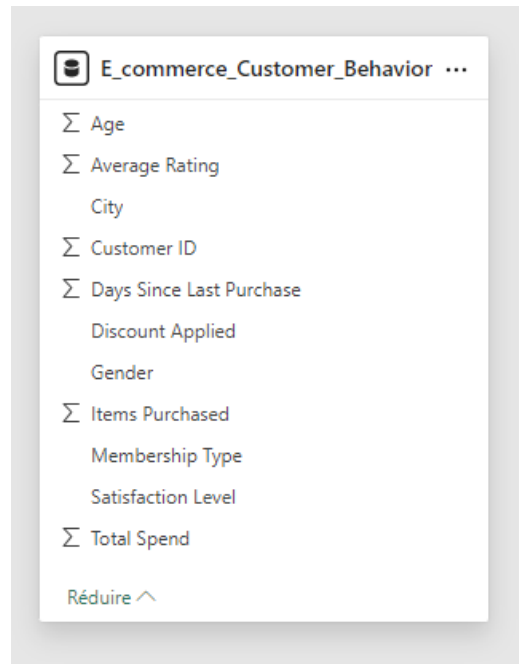
3. Observations générales :

- a. Les jeux de *Nintendo* dominent le classement, représentant une proportion importante des ventes globales.
- b. *Minecraft*, bien qu'étant un seul jeu, représente une part très importante grâce à sa popularité mondiale.
- c. 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 :

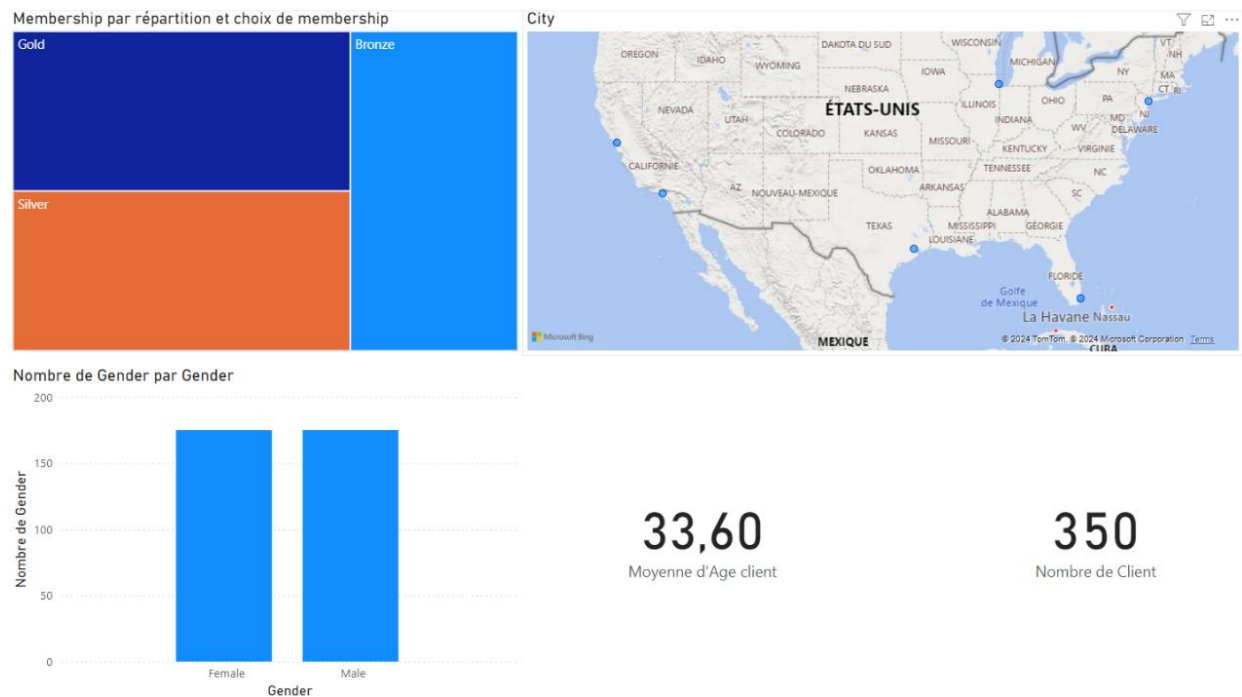


- Age : âge des clients
- Average Rating : Note moyenne attribuée par les clients
- City : Ville de résidence des clients
- Customer ID : Identifiant unique de chaque client
- Days Since Last Purchase : Nombre de jours depuis le dernier achat
- Discount Applied : Indication si un rabais a été appliqué
- 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 :

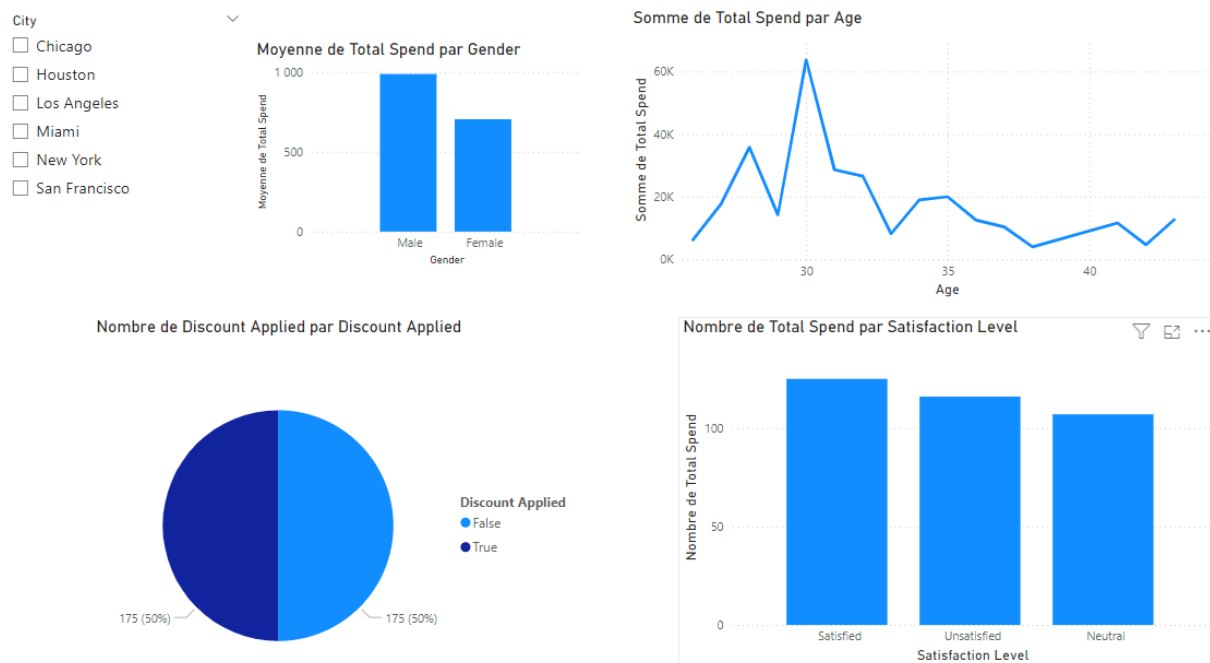
Customer ID	Gender	Age	City	Membership Type	Total Spend	Items Purchased	Average Rating	Discount Applied	Days Since Last Purchase	Satisfaction Level
101	Female	29	New York	Gold	1120,2	14	4,6	True	25	Satisfied
102	Male	34	Los Angeles	Silver	780,5	11	4,1	False	18	Neutral
103	Female	43	Chicago	Bronze	510,75	9	3,4	True	42	Unsatisfied
104	Male	30	San Francisco	Gold	1480,3	19	4,7	False	12	Satisfied
105	Male	27	Miami	Silver	720,4	13	4	True	55	Unsatisfied
106	Female	37	Houston	Bronze	440,8	8	3,1	False	22	Neutral
107	Female	31	New York	Gold	1150,6	15	4,5	True	28	Satisfied
108	Male	35	Los Angeles	Silver	800,9	12	4,2	False	14	Neutral
109	Female	41	Chicago	Bronze	495,25	10	3,6	True	40	Unsatisfied
110	Male	28	San Francisco	Gold	1520,1	21	4,8	False	9	Satisfied
111	Male	32	Miami	Silver	690,3	11	3,8	True	34	Unsatisfied
112	Female	36	Houston	Bronze	470,5	7	3,2	False	20	Neutral
113	Female	30	New York	Gold	1200,8	16	4,3	True	21	Satisfied
114	Male	33	Los Angeles	Silver	820,75	13	4,4	False	15	Satisfied
115	Female	42	Chicago	Bronze	530,4	9	3,5	True	38	Unsatisfied
116	Male	29	San Francisco	Gold	1360,2	18	4,9	False	11	Satisfied
117	Male	26	Miami	Silver	700,6	12	3,7	True	48	Unsatisfied
118	Female	38	Houston	Bronze	450,9	8	3	False	25	Neutral
119	Female	32	New York	Gold	1170,3	14	4,7	True	29	Satisfied
120	Male	34	Los Angeles	Silver	790,2	11	4	False	16	Neutral
121	Female	43	Chicago	Bronze	505,75	10	3,3	True	41	Unsatisfied
122	Male	30	San Francisco	Gold	1470,5	20	4,8	False	13	Satisfied
123	Male	27	Miami	Silver	710,4	13	4,1	True	54	Unsatisfied
124	Female	37	Houston	Bronze	430,8	7	3,4	False	23	Neutral
125	Female	31	New York	Gold	1140,6	15	4,6	True	27	Satisfied
126	Male	35	Los Angeles	Silver	810,9	12	4,3	False	13	Neutral
127	Female	41	Chicago	Bronze	485,25	9	3,6	True	39	Unsatisfied
128	Male	28	San Francisco	Gold	1500,1	21	4,9	False	10	Satisfied
129	Male	32	Miami	Silver	670,3	10	3,8	True	33	Unsatisfied
130	Female	36	Houston	Bronze	460,5	8	3,1	False	19	Neutral
131	Female	30	New York	Gold	1190,8	16	4,5	True	20	Satisfied
132	Male	33	Los Angeles	Silver	830,75	13	4,2	False	14	Satisfied
133	Female	42	Chicago	Bronze	520,4	9	3,5	True	37	Unsatisfied
134	Male	29	San Francisco	Gold	1370,2	18	4,7	False	10	Satisfied
135	Male	26	Miami	Silver	690,6	12	3,9	True	47	Unsatisfied
136	Female	38	Houston	Bronze	440,9	8	3,2	False	24	Neutral
137	Female	32	New York	Gold	1160,3	14	4,4	True	22	Satisfied

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.



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.



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.