

# Commerce en Ligne

MATTEO CAI

DIEGO CASAS

LEO JEAN UNITE

## Table des matières

Table des matières .....	2
1. Introduction .....	3
2. Sources de données .....	3
2.1. Dataset Search .....	3
2.2. Kaggle .....	3
3. Modèle des données .....	4
4. Processus ETL (Extraction, Transformation, Chargement) .....	4
5. Tableau de bord envisagé .....	4
1. Modélisation des SID .....	5
1.1. Choix du processus métier à modéliser .....	5
1.2. Définition de la granularité .....	5
1.3. Choix des dimensions (Par rapport à...) .....	5
1.4. Identification des faits (Quelque chose...) .....	5
2. Intégration des données .....	6
2.1. Video Game Sales Dataset Updated -Extra Feat. ....	6
2.2. Top 1500 games on steam by revenue 09-09-2024 .....	6
2.3. Epic Games Store .....	6
3. Restitution .....	7
3.1. Reporting (Rapport) .....	7
3.2. Dashboard (Tableau de bord) .....	7
3.3. Carte de performance (Scorecard) .....	7

# 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 :

### 2.1. [Dataset Search](#)

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

### 2.2. [Kaggle](#)

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

- [Video Game Sales Dataset Updated -Extra Feat](#) : 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.
- [Top 1500 games on steam by revenue 09-09-2024](#) : Avis de consommateurs d'Amazon, offrant un aperçu de la satisfaction et des produits les plus populaires. La licence : [MIT](#)
- [Metacritic video-games data](#) : un système simple de recommandation. CC0: Public Domain

<https://www.kaggle.com/datasets/ramjasmaurya/epic-games-store> matteo

<https://www.kaggle.com/datasets/kanchana1990/steam-store-2024-hot-picks-and-reviews>

### 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.

# Changement des données

## 1. Modélisation des SID

### 1.1. Choix du processus métier à modéliser

- Vente de jeu
- Notation du jeu

### 1.2. Définition de la granularité

- Une ligne dans le fait représente une vente unique de jeu

Granularité : Transaction

Periode de temp

### 1.3. Choix des dimensions (Par rapport à...)

Processus	Genre	Date	Prix	Score	Region	Plateforme
Vente de jeu	☐—	☐—	☐	☐	☐—	☐—
Notation du jeu	☐—	☐—	☐	☐—	☐	☐—

### 1.4. Identification des faits (Quelque chose...)

- Nombre de ventes (Quantité)
- Qualification
- Revenue

## 2. Intégration des données

Données communes :

- Titre
- Platform
- Sales
- Genre
- Developper
- Publisher

2.1. [Video Game Sales Dataset Updated -Extra Feat](#)

2.2. [Top 1500 games on steam by revenue 09-09-2024](#)

2.3. [Epic Games Store](#)

### 3. Restitution

- 3.1. Reporting (Rapport)
  -
- 3.2. Dashboard (Tableau de bord)
  -
- 3.3. Carte de performance (Scorecard)
  -