

# PROJET FINAL 2

Analyse de Ventes E-commerce

Module BSC-ADD-201 | BDC-ADD-2

Dr. John Benson LOUIS

Décembre 2025

[DEADLINE] Date Limite

**DATE LIMITE : Jeudi 5 Décembre 2025 à 23h59**

## 1 Contexte du Projet

### 1.1 Présentation

Vous êtes data analyst pour une entreprise de e-commerce qui souhaite optimiser ses ventes. L'entreprise dispose de données transactionnelles sur 2 ans (2023-2024) couvrant 10,000 transactions, 500 clients, 100 produits et 20 vendeurs.

### 1.2 Objectifs

Vos objectifs sont de :

1. Analyser les performances de ventes globales
2. Identifier les produits et catégories les plus rentables
3. Segmenter la clientèle et analyser les comportements d'achat
4. Évaluer la performance des vendeurs et régions
5. Optimiser la logistique et les délais de livraison
6. Formuler des prédictions et recommandations stratégiques

### 1.3 Questions de Recherche

Votre analyse doit répondre aux questions suivantes :

- Quelle est l'évolution des ventes sur la période 2023-2024 ?
- Quels sont les produits stars et les moins performants ?
- Quels segments de clients génèrent le plus de revenus ?

- Existe-t-il des différences de performance entre régions ?
- Comment optimiser les délais de livraison ?
- Quelles sont les prédictions pour le trimestre suivant ?

## 2 Description des Données

### 2.1 Localisation

#### [INFO] Information Importante

Les données sont disponibles dans le dossier : lien :

<https://drive.google.com/drive/folders/1dlqVwVjaC0EMeRC37QXUVETWNwICG3Ku?usp=sharing>  
 /data/projet2/

Fichiers disponibles :

- `transactions.csv` : 10,000 transactions
- `clients.csv` : 500 clients
- `vendeurs.csv` : 20 vendeurs
- `produits.json` : 100 produits

### 2.2 Structure des Données

#### 2.2.1 Fichier : `transactions.csv`

**10,000 lignes** représentant les transactions de ventes.

Variable	Type	Description
<code>id_transaction</code>	Character	ID unique (TR001-TR10000)
<code>date</code>	Date	Date de la transaction (2023-2024)
<code>id_client</code>	Character	Référence au client (CLI001-CLI500)
<code>id_produit</code>	Character	Référence au produit (PROD001-PROD100)
<code>id_vendeur</code>	Character	Référence au vendeur (VEN01-VEN20)
<code>quantite</code>	Numeric	Quantité vendue (1-20)
<code>prix_unitaire</code>	Numeric	Prix unitaire du produit
<code>montant_total</code>	Numeric	Montant total de la transaction
<code>remise</code>	Numeric	Remise appliquée (%)
<code>moyen_paiement</code>	Factor	Carte / PayPal / Virement
<code>statut</code>	Factor	Livré / En cours / Annulé
<code>dela_livraison</code>	Numeric	Jours entre commande et livraison

Table 1: Structure de `transactions.csv`

#### 2.2.2 Fichier : `clients.csv`

**500 lignes** avec les informations clients.

#### 2.2.3 Fichier : `vendeurs.csv`

**20 lignes** avec les informations vendeurs.

Variable	Type	Description
id_client	Character	ID unique (CLI001-CLI500)
nom	Character	Nom du client
age	Numeric	Âge (18-70 ans)
ville	Character	Ville de résidence
region	Factor	Nord / Sud / Est / Ouest / Centre
date_inscription	Date	Date d'inscription sur la plateforme
segment	Factor	Premium / Standard / Occasionnel

Table 2: Structure de clients.csv

Variable	Type	Description
id_vendeur	Character	ID unique (VEN01-VEN20)
nom_vendeur	Character	Nom du vendeur
region	Factor	Région d'affectation
anciennete	Numeric	Années d'ancienneté (1-10)

Table 3: Structure de vendeurs.csv

#### 2.2.4 Fichier : produits.json

100 entrées avec les informations produits (format JSON).

Structure JSON :

```
{
  "id_produit": "PROD001",
  "nom_produit": "Ordinateur portable XZ",
  "categorie": "Électronique",
  "sous_categorie": "Informatique",
  "prix_de_base": 899.99
}
```

Catégories disponibles :

- Électronique (Informatique, Audio, Photo)
- Vêtements (Homme, Femme, Enfant)
- Maison (Décoration, Cuisine, Jardin)
- Sport (Fitness, Outdoor, Nautique)
- Livres (Fiction, Non-fiction, BD)

### 3 Analyses Requises

#### 3.1 Analyse 1 : Vue d'Ensemble des Ventes (20 points)

##### 3.1.1 Chiffre d'Affaires Global (8 points)

- a) CA total sur la période 2023-2024
- b) CA par année (2023 vs 2024)
- c) CA par mois (tendance mensuelle)

- d) CA par trimestre
- e) Taux de croissance annuel

### 3.1.2 Volume des Transactions (6 points)

- a) Nombre total de transactions
- b) Nombre de transactions par mois
- c) Panier moyen (montant moyen par transaction)
- d) Distribution des montants de transaction

### 3.1.3 Statuts et Moyens de Paiement (6 points)

- a) Répartition par statut (Livré/En cours/Annulé)
- b) Taux d'annulation
- c) Répartition par moyen de paiement
- d) CA par moyen de paiement

## 3.2 Analyse 2 : Performance Produits (20 points)

### 3.2.1 Produits Individuels (10 points)

- a) Top 10 produits par CA
- b) Top 10 produits par volume
- c) 10 produits les moins performants
- d) Prix moyen par produit
- e) Impact des remises sur les ventes

### 3.2.2 Catégories de Produits (10 points)

- a) CA par catégorie principale
- b) CA par sous-catégorie
- c) Catégorie la plus rentable
- d) Évolution par catégorie (2023 vs 2024)
- e) Part de marché de chaque catégorie

## 3.3 Analyse 3 : Analyse Clients (20 points)

### 3.3.1 Segmentation (10 points)

- a) CA par segment client (Premium/Standard/Occasionnel)
- b) Nombre de clients par segment
- c) CA moyen par client et par segment
- d) Fréquence d'achat par segment

**3.3.2 Démographie et Géographie (10 points)**

- a) CA par région
- b) CA par tranche d'âge
- c) Top 10 villes par CA
- d) Profil du client type
- e) Clients les plus actifs (Top 20)

**3.4 Analyse 4 : Performance Vendeurs (15 points)****3.4.1 Performance Individuelle (8 points)**

- a) CA par vendeur
- b) Top 5 et Bottom 5 vendeurs
- c) Nombre de transactions par vendeur
- d) Panier moyen par vendeur

**3.4.2 Performance par Région (7 points)**

- a) CA par région
- b) Comparaison inter-régions
- c) Impact de l'ancienneté sur les performances

**3.5 Analyse 5 : Logistique (15 points)****3.5.1 Délais de Livraison (10 points)**

- a) Délai moyen de livraison
- b) Distribution des délais
- c) Délais par région
- d) Délais par catégorie de produit
- e) Identification des problèmes logistiques

**3.5.2 Qualité de Service (5 points)**

- a) Taux de livraison réussie
- b) Corrélation délai / annulations
- c) Recommandations d'optimisation

**3.6 Analyse 6 : Prédictions (10 points)**

- a) Prédiction CA pour le trimestre suivant
- b) Identification des tendances saisonnières
- c) Prévision par catégorie
- d) Recommandations stratégiques basées sur les tendances

### 3.7 Visualisations (20 points)

Vous devez créer **au minimum 6 graphiques de qualité :**

1. **Évolution temporelle** : CA mensuel sur 2 ans (4 pts)
2. **Barplot** : CA par catégorie de produits (3 pts)
3. **Carte géographique ou barplot** : CA par région (3 pts)
4. **Scatter plot** : Délai vs CA (3 pts)
5. **Heatmap** : Ventes par mois et catégorie (4 pts)
6. **Graphique créatif** : Au choix, pertinent (3 pts)

## 4 Livrables Attendus

### 4.1 Structure du Projet

Votre projet doit avoir cette structure :

```
mon_projet_2/
|-- data/
|   |-- raw/
|   |   |-- transactions.csv
|   |   |-- clients.csv
|   |   |-- vendeurs.csv
|   |   '-- produits.json
|   '-- processed/
|       '-- donnees_fusionnees.csv
|-- scripts/
|   |-- 00_fonctions.R
|   |-- 01_import_nettoyage.R
|   |-- 02_analyse_ventes.R
|   |-- 03_analyse_produits.R
|   |-- 04_analyse_clients.R
|   |-- 05_visualisations.R
|   '-- main.R
|-- output/
|   |-- figures/
|   |   |-- 01_evolution_ca.png
|   |   |-- 02_ca_categories.png
|   |   '-- ...
|   '-- tables/
|       |-- top_produits.csv
|       |-- ca_par_region.csv
|       '-- ...
|-- rapport.pdf (ou rapport.docx)
 '-- README.txt
```

### 4.2 Fichiers à Soumettre

1. **Scripts R** : Tous les fichiers .R organisés

2. **Rapport** : PDF ou DOCX (3 à 6 pages)
3. **Graphiques** : Tous les graphiques en PNG haute résolution
4. **Tableaux** : Résultats exportés en CSV
5. **README.txt** : Instructions d'exécution

### 4.3 Contenu du Rapport

Votre rapport doit contenir :

1. **Page de garde** : Titre, nom, date
2. **Résumé exécutif**
3. **Introduction** : Contexte et objectifs
4. **Méthodologie** : Description des données et méthodes
5. **Résultats** : 6 sections d'analyse avec tableaux et graphiques
6. **Discussion** : Interprétation des résultats
7. **Recommandations** : Au moins 5 recommandations stratégiques
8. **Conclusion** : Synthèse et perspectives
9. **Annexes** : Tableaux détaillés si nécessaire

## 5 Grille d'Évaluation

### 5.1 Barème Détaillé (100 points)

### 5.2 Import des Données

```
# CSV
transactions <- read.csv("/data/projet2/transactions.csv")
clients <- read.csv("/data/projet2/clients.csv")
vendeurs <- read.csv("/data/projet2/vendeurs.csv")

# JSON
library(jsonlite)
produits <- fromJSON("/data/projet2/produits.json")

# Fusion
library(dplyr)
donnees_completes <- transactions %>%
  left_join(clients, by = "id_client") %>%
  left_join(vendeurs, by = "id_vendeur") %>%
  left_join(produits, by = "id_produit")
```

### 5.3 Gestion des Dates

[!] Attention

**Important :** Convertissez les dates en format Date R

```
# Conversion
donnees$date <- as.Date(donnees$date)

# Extraction
donnees$annee <- format(donnees$date, "%Y")
donnees$mois <- format(donnees$date, "%m")
donnees$trimestre <- quarters(donnees$date)
```

### 5.4 Analyses Temporelles

```
# CA par mois
library(dplyr)
ca_mensuel <- donnees %>%
  group_by(annee, mois) %>%
  summarise(ca = sum(montant_total, na.rm = TRUE))

# Tendance
library(ggplot2)
ggplot(ca_mensuel, aes(x = interaction(annee, mois), y = ca)) +
  geom_line(group = 1) +
  geom_point()
```

## 6 Checklist Finale

Avant de soumettre, vérifiez :

### 6.1 Code

Structure de dossiers correcte

Import CSV et JSON fonctionnel

Fusion des 4 fichiers réussie

Au moins 4 fonctions créées

Code bien commenté

`main.R` s'exécute sans erreur

Gestion correcte des dates

### 6.2 Analyses

Les 6 analyses requises effectuées

Résultats exportés (CSV)

Calculs vérifiés

Prédictions formulées

### 6.3 Visualisations

- Au moins 6 graphiques
- Évolution temporelle présente
- Haute résolution ( $>= 300$  DPI)
- Titres et labels présents
- Heatmap ou graphique avancé inclus

### 6.4 Soumission

- Tous les fichiers présents
  - Archive ZIP créée
  - Nom correct : `NOM_Prenom_Projet2.zip`
  - Soumis avant le 5 Décembre 2025 23h59
  - lien de soumission :
- <https://drive.google.com/drive/folders/1RGY5lnLwNtzXJS8yz8C56OdhaVb07AR2?usp=sharing>*

## 7 Contact et Support

### 7.1 Enseignant

**Dr. John Benson LOUIS**

Email : john\_benson.louis@ueh.edu.ht

**Bon courage et excellent travail !**

*“Les données racontent une histoire. À vous de la découvrir.”*