

BRIEF PROJET
Analyse des ventes d'une PME

Présentation des livrables et de la solution du projet

Candidate Harley RUEDA OSMA

Paris, le 27 mars 2025



Lettre de motivation – Formation Data Engineer

À l'équipe pédagogique de Simplon,

Passionné par l'univers de la donnée, je souhaite aujourd'hui me reconvertir pleinement vers le métier de Data Engineer. Fort de plus de 6 ans d'expérience dans le secteur de l'énergie en Europe, j'ai toujours évolué dans des environnements techniques complexes où la gestion et l'analyse de données occupaient une place stratégique.

Afin de concrétiser cette reconversion, j'ai réalisé avec rigueur et enthousiasme le projet demandé pour le positionnement à la formation Simplon. Ce travail m'a permis de mettre en œuvre des compétences concrètes : mise en place d'une architecture conteneurisée avec Docker, ingestion et traitement automatisé de fichiers CSV via Python, création d'une base de données SQLite relationnelle, exécution de requêtes SQL et génération de visualisations claires et exploitables. Ce fut un exercice à la fois exigeant et passionnant, qui a renforcé mon envie d'approfondir ces compétences au sein de votre formation.

Je suis convaincu que la formation de Data Engineer proposée par Simplon est le tremplin idéal pour structurer mes acquis, maîtriser les outils Big Data, et m'insérer dans un environnement professionnel où je pourrai allier mes expériences passées à une expertise technique de pointe. Je suis également ouvert aux dispositifs en alternance, POEI ou POEC.

C'est avec un engagement sincère et une réelle motivation que je souhaite rejoindre cette aventure, pour continuer à apprendre, construire, et contribuer à des projets à fort impact.

Dans l'attente de pouvoir échanger avec vous, je vous prie d'agréer, Madame, Monsieur, l'expression de mes salutations distinguées.

Harley RUEDA

Table des matières

Livrables.....	3
1. Le schéma de l'architecture conçu.....	3
Description et application de ressources techniques.....	3
Architecture technique.....	3
Schéma de l'architecture conçu.png	4
2. Schéma de l'architecture de la base de données.....	5
Base de données SQLite – analyses.db	5
Schéma de l'architecture de la base de données.png.....	7
3. Le Dockerfile	8
4. Le fichier yaml du Docker Compose.....	8
5. Le(s) script(s) d'exécution pour la collecte, transformation, et import des données.....	8
6. Une fiche synthèse des résultats d'analyse obtenus (point 4.a, 4.b, 4.c).	9
Fiche synthèse – Résultats de l'analyse	9
7. Une vidéo de démonstration, en enregistrant l'écran de l'ordinateur, du projet réalisé	11

Livrables

1. Le schéma de l'architecture conçu

Description et application de ressources techniques

Ce projet a pour objectif de construire une solution conteneurisée capable d'analyser les ventes d'une PME à partir de fichiers CSV (locaux et en ligne). Le livrable final comprend une base de données, des fichiers exportés (CSV) et des graphiques générés automatiquement.

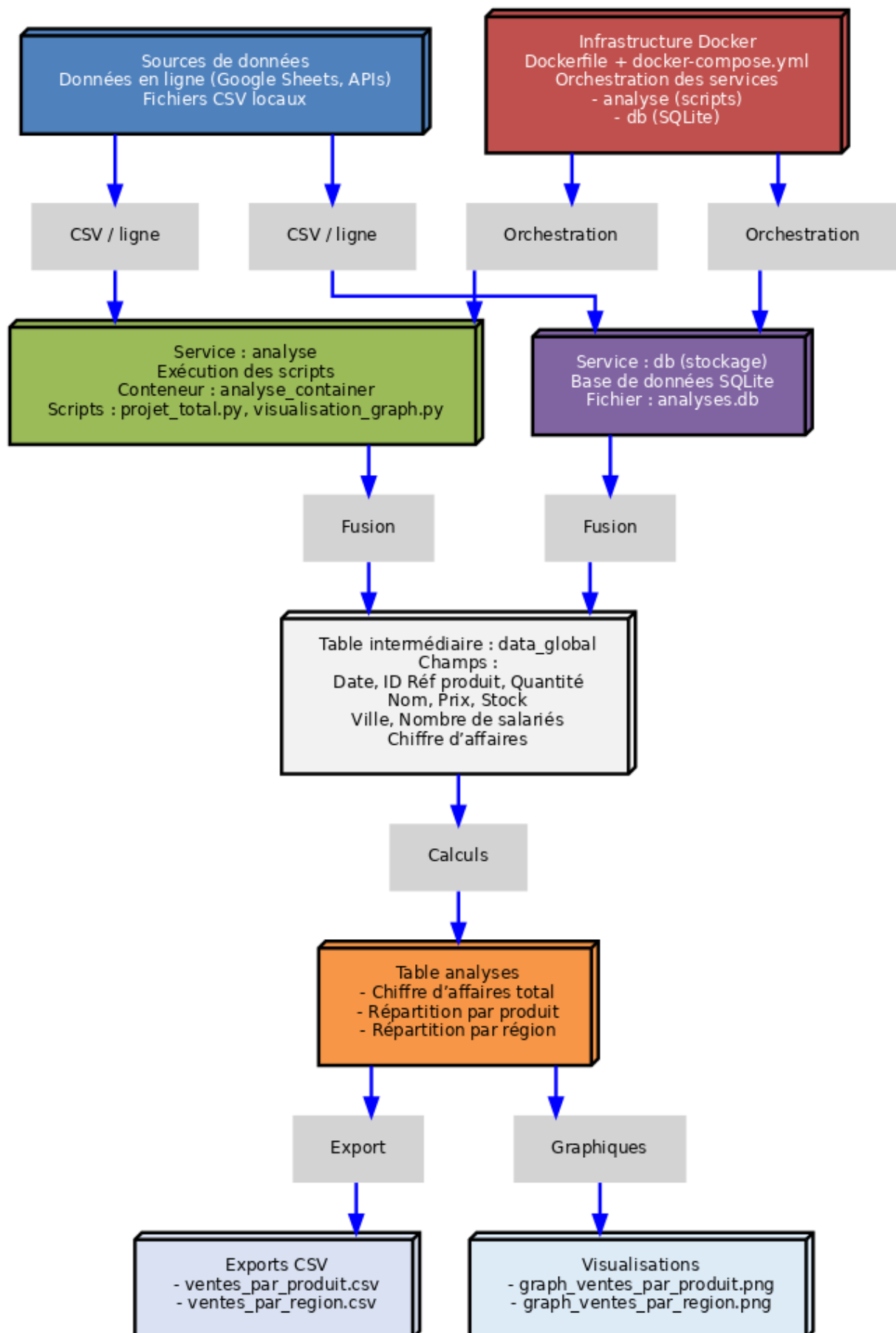
Architecture technique

L'architecture repose sur un conteneur Docker unique appelé analyse_container, orchestré via docker-compose.yml, qui exécute le script principal projet_total.py.

Fichiers principaux :

- Dockerfile : définit l'image de base et installe les dépendances
- docker-compose.yml : configure et lance le conteneur avec les volumes nécessaires
- projet_total.py : script principal d'importation, de transformation et d'analyse des données
- visualisation_graph.py : génère les graphiques
- analyses.db : base de données SQLite contenant les résultats

Schéma de l'architecture conçu.png



2. Schéma de l'architecture de la base de données

Base de données SQLite – analyses.db

La base de données analyses.db est générée par le script projet_total.py, à partir de fichiers CSV (locaux et en ligne). Le script effectue l'importation, le nettoyage, la fusion et l'analyse des données, puis construit une base SQLite relationnelle contenant cinq tables principales.

Structure des tables

Magasins

Colonne	Type
ID Référence produit	TEXT
Nom	TEXT
Prix	REAL
Stock	INTEGER

Clé primaire : ID Référence produit

produits

Colonne	Type
Date	TEXT
ID Référence produit	TEXT
Quantité	INTEGER
ID Magasin	INTEGER

Clé primaire : Date, ID Référence produit, Quantité, ID Magasin

ventes

Colonne	Type
ID Magasin	INTEGER

Ville	TEXT
Nombre de salariés	INTEGER

Clé primaire : ID Magasin

data_global

Colonne	Type
Date	TEXT
ID Référence produit	TEXT
Quantité	INTEGER
ID Magasin	INTEGER
Nom	TEXT
Prix	REAL
Stock	INTEGER
Ville	TEXT
Nombre de salariés	INTEGER
Chiffre d'affaires	REAL

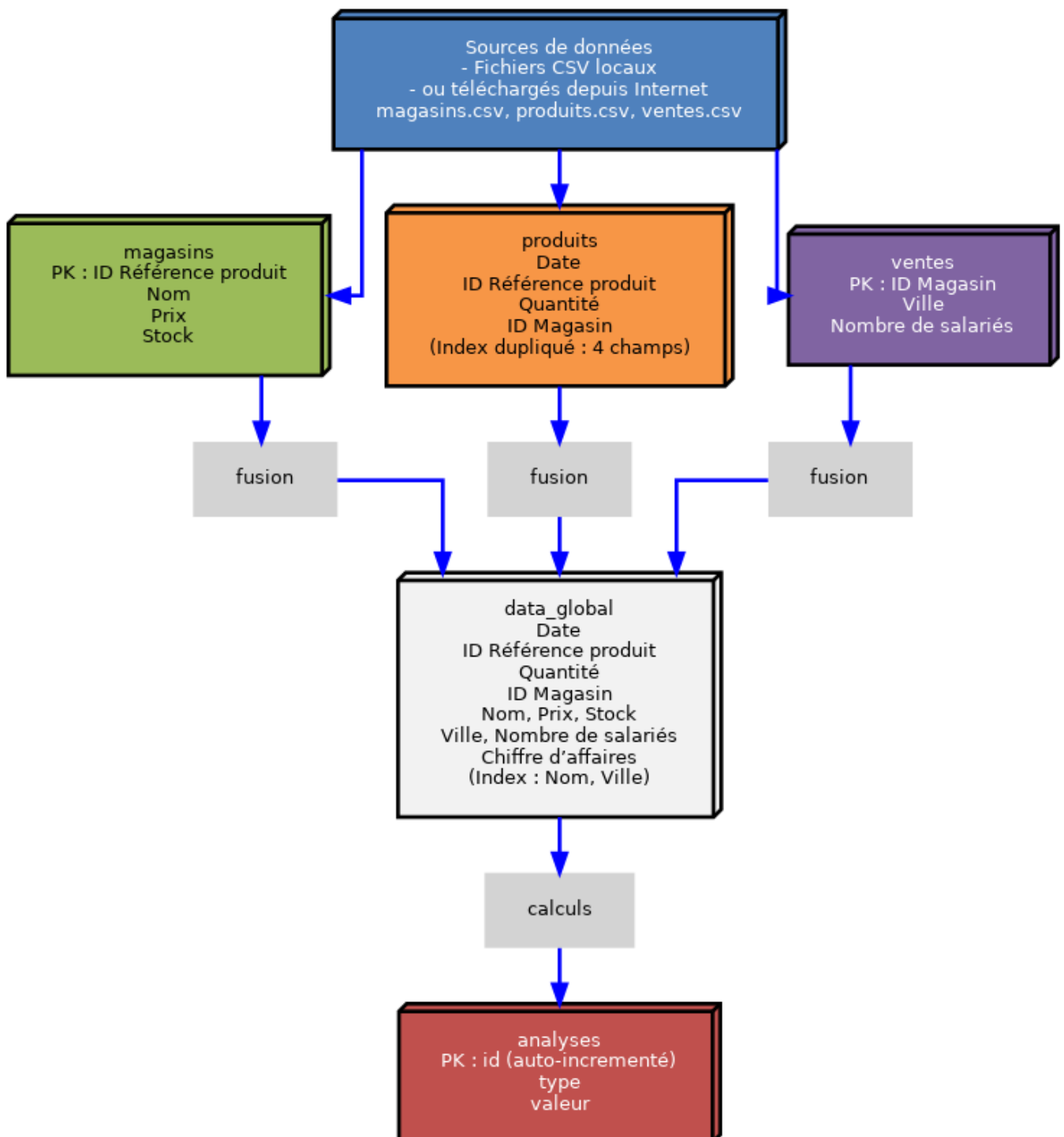
Clé primaire : Nom, Ville

analyses

Colonne	Type
id	INTEGER
type	TEXT
valeur	TEXT

Clé primaire : id (auto-incrémentée)

Schéma de l'architecture de la base de données.png



3. Le Dockerfile

Le Dockerfile est essentiel pour garantir que l'environnement d'exécution du projet soit cohérent sur toutes les machines. Ce fichier :

- Utilise une image légère de Python (python:3.10-slim), ce qui permet de minimiser la taille de l'image Docker.
- Installe toutes les dépendances nécessaires listées dans le fichier requirements.txt, garantissant ainsi une configuration uniforme de l'environnement Python.
- Configure l'environnement d'exécution du projet et l'accès aux bibliothèques et outils nécessaires à l'exécution du code.

Voir le fichier : **Dockerfile**

4. Le fichier yaml du Docker Compose

Le fichier docker-compose.yml sert à orchestrer les différents services et composants nécessaires au projet. Il configure plusieurs éléments :

- Volume pour persister les données : Ce mécanisme permet de garantir que les données traitées ou générées par l'application (comme les résultats de calculs) ne sont pas perdues lors de la fermeture du conteneur.
- Environnement Python : Il définit l'environnement Python nécessaire pour l'exécution du projet.
- Point d'entrée du script principal : Il configure le point d'entrée du conteneur afin d'exécuter le script projet_total.py.

Voir le fichier : **docker-compose.yml**

5. Le(s) script(s) d'exécution pour la collecte, transformation, et import des données

Les scripts d'exécution automatisent plusieurs étapes clés du processus de gestion des données:

- projet_total.py : Ce script est le cœur du projet. Il s'occupe de la collecte des fichiers CSV (soit depuis des sources locales, soit depuis des sources en ligne), du nettoyage et de la fusion des données, des calculs analytiques (comme le chiffre d'affaires), et de l'insertion des résultats dans la base de données analyses.db.
- visualisation_graph.py : Ce script génère des graphiques à partir des données traitées. Il permet de visualiser les résultats analytiques sous forme de graphiques exploitables pour l'analyse des données.
- **Voir le fichier : projet_total.py**
- **Voir le fichier : visualisation_graph.py**

6. Une fiche synthèse des résultats d'analyse obtenus (point 4.a, 4.b, 4.c).

Fiche synthèse – Résultats de l'analyse

1. Chiffre d'affaires total

Total : 5 268,78 €

2. Répartition du chiffre d'affaires par produit

Produit	Chiffre d'affaires (€)
Produit D	1 690,00 €
Produit E	1 400,00 €
Produit A	1 200,00 €
Produit B	530,00 €
Produit C	448,78 €

3. Répartition du chiffre d'affaires par region

Ville	Chiffre d'affaires (€)
Lyon	1 059,59 €
Marseille	1 011,62 €
Bordeaux	828,60 €
Paris	799,86 €
Nantes	737,63 €
Strasbourg	578,55 €
Lille	253,93 €

4. Fichiers générés

Exports CSV :

- *ventes_par_produit.csv*

- *ventes_par_region.csv*

Graphiques PNG :

- *graph_ventes_par_produit.png*

- *graph_ventes_par_region.png*

5. Conclusion

L'analyse du chiffre d'affaires montre une concentration des ventes sur certains produits clés, notamment Produit D et Produit E, qui représentent à eux seuls plus de 58 % du total.

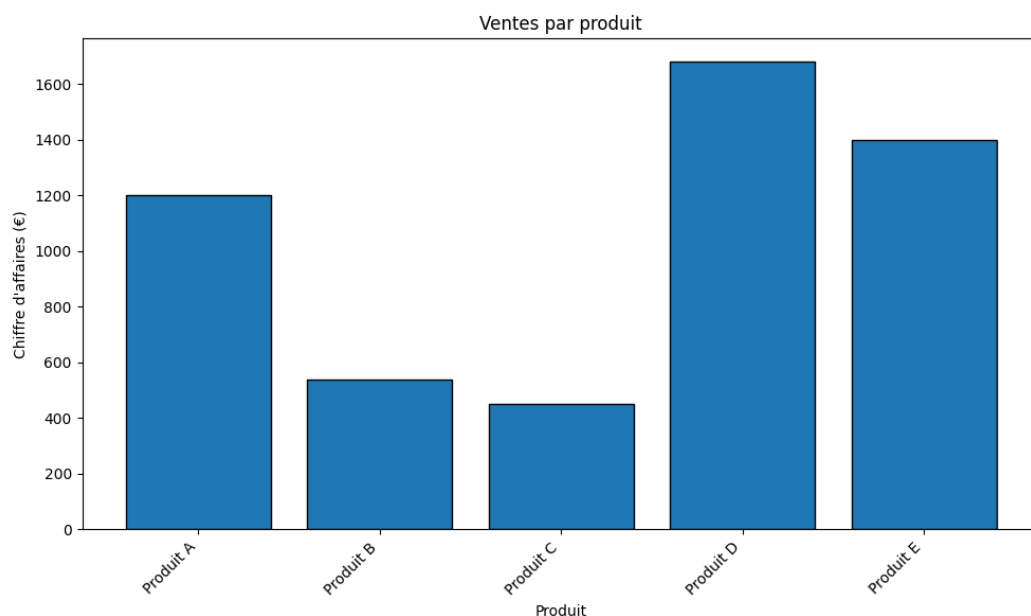
Au niveau régionale, la ville de Lyon génère le chiffre d'affaires le plus élevé, suivie de près par Marseille et Bordeaux.

Ces résultats peuvent aider l'entreprise à :

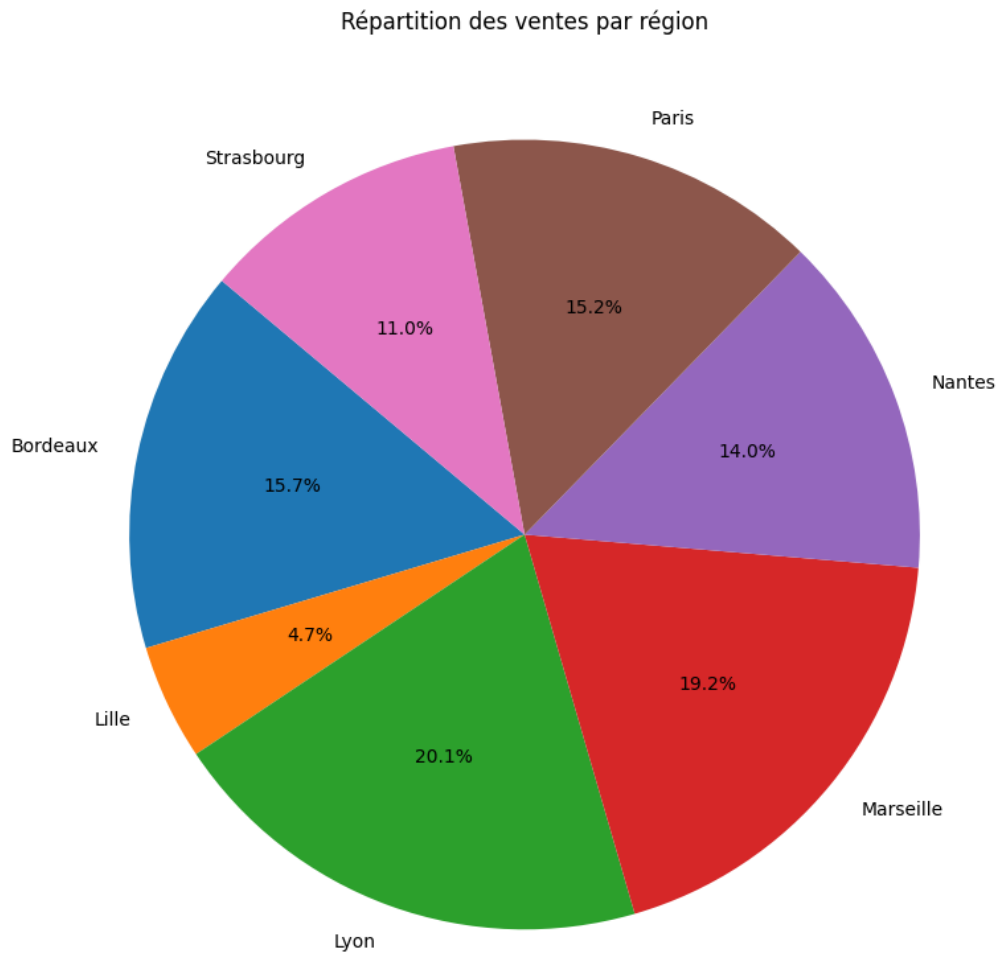
- Renforcer l'approvisionnement des produits les plus rentables
- Focaliser ses efforts marketing dans les régions les plus performantes
- Identifier des axes d'amélioration pour les villes comme Lille, où les ventes restent faibles

6. Graphiques générés

Graphique – Chiffre d'affaires par produit :



Graphique – Répartition par région :



7. Une vidéo de démonstration, en enregistrant l'écran de l'ordinateur, du projet réalisé

La vidéo doit commencer en montrant la base de données créée mais vide, pour ensuite aller vers le terminal, lancer les services et observer les logs d'exécution. Finalement la vidéo doit montrer la base de données avec les données importées. La vidéo doit enregistrer l'intégralité de l'écran afin que la date et l'heure de l'ordinateur soient incluses dans l'enregistrement.